ON THE NONLINEAR BRASCAMP–LIEB INEQUALITY

JONATHAN BENNETT, NEAL BEZ, STEFAN BUSCHENHENKE, MICHAEL G. COWLING, AND TARYN C. FLOCK

Abstract. We prove a nonlinear variant of the general Brascamp–Lieb inequality. Instances of this inequality are quite prevalent in analysis, and we illustrate this with substantial applications in harmonic analysis and partial differential equations. Our proof consists of running an efficient, or “tight”, induction on scales argument, which uses the existence of gaussian near-extremisers to the underlying linear Brascamp–Lieb inequality (Lieb’s theorem) in a fundamental way. A key ingredient is an effective version of Lieb’s theorem, which we establish via a careful analysis of near-minimisers of weighted sums of exponential functions.

1. Introduction

The Brascamp–Lieb inequality is a well-known and far-reaching generalisation of a wide range of sharp functional inequalities in analysis, including the multilinear Hölder, Loomis–Whitney and Young convolution inequalities. It takes the form

\[
\int_{\mathbb{R}^n} \prod_{j=1}^{m} f_j^{p_j}(L_j x) \, dx \leq BL(L, p) \prod_{j=1}^{m} \left( \int_{\mathbb{R}^{n_j}} f_j(x_j) \, dx_j \right)^{p_j}
\]

over all nonnegative integrable functions \( f_j \), where the mappings \( L_j : \mathbb{R}^n \to \mathbb{R}^{n_j} \) are linear surjections, the exponents \( p_j \in [0, 1] \), and \( BL(L, p) \) denotes the smallest constant (which may be infinite). We refer to \((L, p) = ((L_j)_{j=1}^{m}, (p_j)_{j=1}^{m})\) as the Brascamp–Lieb datum, and \( BL(L, p) \) as the Brascamp–Lieb constant. This celebrated inequality was first formulated by Brascamp and Lieb in [38] and gives rise to an elegant and powerful theory; further notable contributions in this direction include those of Ball [5], Lieb [78], Barthe [6], Carlen, Lieb and Loss [46], as well as Bennett, Carbery, Christ and Tao [25] and Barthe, Cordero-Erausquin, Ledoux and Maurey [8]. Ball [2] derived a particular class of rank-one Brascamp–Lieb inequalities and pioneered their use in convex geometry, solving several problems on volumes of sections of convex bodies. Barthe [6, 7] extended such geometric Brascamp–Lieb inequalities to the general rank case, and introduced the use of optimal transport methods in order to advance the general theory of the Brascamp–Lieb inequality and its dual form. Applications and perspectives on the Brascamp–Lieb inequality may be found even more widely, including in convex geometry [e.g. 10, 39, 40, 63], probability, stochastic processes and statistics [e.g. 2, 3, 55, 77], information theory [e.g. 4, 45, 49, 57, 70], scattering theory [e.g. 1, 41, 42, 82], combinatorics [e.g. 20, 56], group theory [e.g. 47, 53], and theoretical computer science [e.g. 49, 50, 54, 64, 94]. For example, the information-theoretic connection reveals that the Brascamp–Lieb inequality has an equivalent formulation in terms of a generalised notion of subadditivity of the entropy (see [45]).
The Brascamp–Lieb inequality (1.1) is often referred to as a multilinear inequality, since it is equivalent to

\[
\int_{\mathbb{R}^n} \prod_{j=1}^m f_j(L_j x) \, dx \leq \text{BL}(L, \mathbf{p}) \prod_{j=1}^m \|f_j\|_{L^{p_j}((\mathbb{R}^n))};
\]

here we have simply replaced \( f_j^{p_j} \) with \( f_j \), and \( p_j \) with \( 1/q_j \). This allows (1.1) to be interpreted as a bound on the multilinear form

\[
(f_1, \ldots, f_m) \mapsto \int_H f_1 \otimes \cdots \otimes f_m \, d\mu_H.
\]

Here, \( L \) is encoded in \( H \), the range of \( x \mapsto (L_1 x, \ldots, L_m x) \), and integration in (1.3) is with respect to Lebesgue measure on \( H \). As one may expect, quite how this subspace \( H \) sits relative to the ranges of the various \( L_j \) determines the nature of the Brascamp–Lieb inequality. A well-known example, and indeed the historical motivation for the Brascamp–Lieb inequality, is the celebrated sharp version of Young’s convolution inequality, first proved by Beckner [13, 12] and Brascamp–Lieb [38]. In the framework of (1.1), this may be stated as

\[
\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f_1(y)^{p_1} f_2(x-y)^{p_2} f_3(x)^{p_3} \, dy \, dx
\]

\[
\leq (C_{p_1} C_{p_2} C_{p_3})^d \left( \int_{\mathbb{R}^d} f_1(x) \, dx \right)^{p_1} \left( \int_{\mathbb{R}^d} f_2(x) \, dx \right)^{p_2} \left( \int_{\mathbb{R}^d} f_3(x) \, dx \right)^{p_3},
\]

where \( p_1, p_2, p_3 \in [0, 1] \), \( p_1 + p_2 + p_3 = 2 \), and \( C_r = ((1-r)^{1-r}/r^r)^{1/2} \). It is pertinent here that on the relative interior of the set of admissible exponents, the optimal constant \( (C_{p_1} C_{p_2} C_{p_3})^d \) is strictly less than 1, and is uniquely attained on suitably scaled isotropic centred gaussian inputs \( f_j \). We shall return to the distinguished role of gaussians in the general setting of (1.1) later in this introduction.

In the last fifteen years a number of variants of the Brascamp–Lieb inequality have emerged in harmonic analysis, with significant applications in diverse areas of mathematics, including Fourier analysis, combinatorial geometry, analytic number theory and partial differential equations. This extraordinary effectiveness reflects an improved understanding of the ubiquitous role of curvature in harmonic analysis, and in particular its relation to multilinear manifestations of transversality. (The reader may have already noticed that some notion of transversality should be decisive to the behaviour of the multilinear form (1.3), and this will be clarified further shortly). In particular, the multilinear variants of the celebrated Fourier restriction and Kakeya conjectures developed in Bennett, Carbery and Tao [27], Guth [57], Bourgain and Guth [37], Carbery and Valdimarsson [14], Bennett, Bez, Flock and Lee [21], Zhang [95] and Zorin–Kranich [96], may be interpreted as certain oscillatory and Kakeya-type Brascamp–Lieb inequalities, and these underpin the recent theory of decouplings (also known as Wolff inequalities), developed and utilised to great effect by many authors — see, in particular, work of Bourgain and Demeter, Bourgain, Demeter and Guth, and Bourgain, Demeter and Guo [34, 37, 36]. Another oscillatory form of the Brascamp–Lieb inequality was introduced by Christ, Li, Thiele and Tao [51], this time exhibiting connections with questions related to Szemerédi’s theorem from additive combinatorics. There is also a natural singular Brascamp–Lieb inequality, which notably includes the multilinear Hilbert transform — see Durcik and Thiele [55]. The Brascamp–Lieb inequality is also naturally formulated in discrete settings — see, for example Christ, Demmel, Knight, Scanlon and Yelick [50], where connections are made with Hilbert’s tenth problem over the rational numbers. A further significant variant is the so-called nonlinear Brascamp–Lieb inequality, introduced by Bennett, Carbery and Wright [28], which asks whether one may replace the linear surjections \( L_j \) by smooth submersions, at least in a
neighbourhood of a point. The main purpose of this paper is to give an affirmative answer to this question. Our main theorem is the following.

**Theorem 1.1.** Let \((L, p)\) be a Brascamp–Lieb datum. Suppose that \(B_j : \mathbb{R}^n \to \mathbb{R}^{n_j}\) are \(C^2\) submersions in a neighbourhood of a point \(x_0\) and \(dB_j(x_0) = L_j\), where \(1 \leq j \leq m\). Then, for each \(\varepsilon > 0\) there exists a neighbourhood \(U\) of \(x_0\) such that

\[
\int_U \prod_{j=1}^m f_j^{p_j}(B_j(x)) \, dx \leq (1 + \varepsilon) \text{BL}(L, p) \prod_{j=1}^m \left( \int_{\mathbb{R}^{n_j}} f_j(x_j) \, dx_j \right)^{p_j}.
\]

Theorem 1.1 has some significant applications, which we present in Section 2; see, in particular, the forthcoming Corollary 2.1 and Theorem 2.2. This applicability is perhaps to be expected, as Theorem 1.1 provides bounds on a broad class of quite general positive multilinear forms. Notice that (1.5) is equivalent to the multilinear form bound

\[
\int_U \prod_{j=1}^m f_j(B_j(x)) \, dx \leq (1 + \varepsilon) \text{BL}(L, p) \prod_{j=1}^m \|f_j\|_{L^{p_j}(\mathbb{R}^{n_j})},
\]

which may be interpreted as a Lebesgue space bound on

\[
(f_1, \ldots, f_m) \mapsto \int_{\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}} f_1 \otimes \cdots \otimes f_m \, d\mu_M,
\]

where \(d\mu_M\) is a smooth density supported on a smooth submanifold \(M\) of \(\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}\). Here \(B = (B_j)_{j=1}^m\) is encoded in \(M\) by

\[
M = \{(B_1(x), \ldots, B_m(x)) : x \in U\}.
\]

By the (multilinear) Schwartz kernel theorem, (1.7) takes the form of a quite general positive multilinear form,

\[
(f_1, \ldots, f_m) \mapsto \int_{\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}} f_1 \otimes \cdots \otimes f_m \, d\mu,
\]

whose measure \(\mu\), given by the Riesz representation theorem, is a smooth density supported on a smooth submanifold of euclidean space. Multilinear forms of this type, which are sometimes referred to as multilinear Radon-like transforms, abound — see Section 2 for examples and further discussion. Theorem 1.1 has only first order structural conditions on the maps \(B_j\) (and thus the manifold \(M\)), and it is not difficult to show that imposing additional higher order conditions on the \(B_j\), which may be interpreted as curvature conditions on \(M\), does not lead to improved bounds (see Bennett, Bez and Gutiérrez [22, Lemma 6]). However, it is important to point out that multilinear Radon-like transforms often arise with no underlying finite Brascamp–Lieb constant, yet estimates are possible under alternative (higher order) nondegeneracy conditions on the \(B_j\) — see Tao and Wright [90] and Stovall [85]. Obtaining such a higher order result at the level of generality of Theorem 1.1 would be of considerable interest. It should also be noted that Theorem 1.1 is local in nature, and it would be interesting to try to formulate a global version, possibly involving a suitable weight factor — see Bennett, Bez and Gutiérrez [22], Bennett, Carbery and Wright [28], Koch and Steinberger [76] and Carbery, Hämminen and Valdimarsson [13] for further discussion.

We remark that Theorem 1.1 is equivalent to a superficially stronger statement where the underlying euclidean spaces are replaced by smooth manifolds. This is a reflection of the (essential) diffeomorphism-invariance of its statement. While this is arguably the natural context for such nonlinear Brascamp–Lieb inequalities, we choose to work in the euclidean setting for concreteness.
We also note that Theorem 1.1 may be reformulated, in an equivalent way, in terms of a
generalised subadditivity for the entropy functional; this follows from Carlen and Cordero-
Erausquin [45, Theorem 2.1], and this perspective may well be of wider independent
interest.

Theorem 1.1 was previously unknown with any finite constant in place of the optimal
$1 + \varepsilon$ factor, and we do not know how to prove such a statement without the full argument
of this paper.

Prior to Theorem 1.1, such nonlinear statements were only known under additional
structural constraints on the nonlinear maps $B$. This began with Bennett, Carbery and
Wright [28] and Bennett and Bez [19], where the underlying linear maps $L$ in the statement
of Theorem 1.1 were required to satisfy the highly-restrictive “basis condition”

$$\bigoplus_{j=1}^m \ker L_j = \mathbb{R}^n,$$

which keeps the complexity at the level of the classical Loomis–Whitney inequality; see also
Bejenaru, Herr and Tataru [17], Koch and Steinberger [76], Carbery, Hänninen and Valdimarsson [43],
and PDE applications in Bejenaru, Herr and Tataru [16], Bejenaru and Herr [15], Faou, Germain and Hani [58],
Hirayama and Kinoshita [70], and Kinoshita [73] — see Section 2.2 for further discussion. We note that the results in [43],
which predate this work, also feature the best-possible $(1 + \varepsilon)$ factor present in (1.5).

Some isolated examples beyond the scope of (1.9) may also be found in unpublished work
of Stovall. Very different arguments of Zhang [95] and Zorin–Kranich [96], building on
work of Guth [67], Bourgain and Guth [37], and Carbery and Valdimarsson [44], allowed
for polynomial $B$, but generated bounds that depend strongly on the degrees. The only
nonlinear statements that were known for completely general data — see Bennett, Bez,
Flock and Lee [21] — involved an $\varepsilon$-loss in the scale of Sobolev spaces, and were of the

$$\left| \int \prod_{j=1}^m f_j(B_j(x)) \, dx \right| \leq C_\varepsilon \prod_{j=1}^m \|f_j\|_{L_p^\varepsilon(\mathbb{R}^n)}.$$ 

Theorem 1.1 has proved quite elusive. This is perhaps to be expected, as the analysis
of the classical inequality (1.1) is delicate and relies heavily on linear-algebraic arguments.
A particularly important example of linearity at work is in a fundamental theorem of
Lieb [78], which states that the Brascamp–Lieb constant $BL(L, p)$ is exhausted by centred
gaussian inputs $f_j$; that is

$$f_j(x) = c_j \exp(-\pi (A_j x, x)),$$

where $A_j$ is a positive definite $n_j \times n_j$ matrix, and $c_j$ is a positive constant. For such inputs,

$$\int_{\mathbb{R}^n} \prod_{j=1}^m f_j(L_j x) \, dx \quad \geq \quad \prod_{j=1}^m \frac{\det(A_j)^{p_j/2}}{\det \left( \sum_{j=1}^m p_j L_j^* A_j L_j \right)^{1/2}},$$

and so Lieb’s theorem may be stated as follows:

Theorem 1.2 (Lieb).

$$BL(L, p) = \sup_{A_j} BL_g(L, p; A),$$

where $A = (A_j)_{j=1}^m$, and

$$BL_g(L, p; A) = \frac{\prod_{j=1}^m \det(A_j)^{p_j/2}}{\det \left( \sum_{j=1}^m p_j L_j^* A_j L_j \right)^{1/2}}.$$
Lieb’s theorem, which of course relies heavily on the linearity of the maps $L_j$, has been central to much of our understanding of the Brascamp–Lieb inequality. For instance, Bennett, Carbery, Christ and Tao [25] used it to characterise the data $(L, p)$ for which $\text{BL}(L, p) < \infty$. They show that $\text{BL}(L, p) < \infty$ if and only if the “transversality condition”

\begin{equation}
\dim(V) \leq \sum_{j=1}^{m} p_j \dim(L_j V)
\end{equation}

holds for all subspaces $V$ of $\mathbb{R}^n$, along with the scaling condition

\begin{equation}
n = \sum_{j=1}^{m} p_j n_j.
\end{equation}

An alternative derivation of this basic result, which also pivots on linear algebraic arguments (in this case by passing to subspaces and using induction on $n$) was given by Bennett, Carbery, Christ and Tao [26]. Further refinements include a characterisation of the data for which extremisers exist, and for which they are unique (up to the natural scale-invariance of the inequality) — see, for example, [25], or Carlen, Lieb and Loss [46], Valdimarsson [91, 92] for further discussion and references. Again, none of the above linear-algebraic considerations continue to be directly relevant if the mappings $L_j$ cease to be linear. In the setting of nonlinear data, there appear to be essentially three viable approaches available: (i) the method of refinements of Christ [48], (ii) the factorisation method of Carbery, Hänninen and Valdimarsson [43], and (iii) the method of induction-on-scales, introduced in this setting by Bejenaru, Herr and Tataru [17]. All of these methods are effective under the rather rigid Loomis–Whitney-type structural condition (1.9), as may be seen in Bennett, Carbery and Wright [28], Carbery, Hänninen and Valdimarsson [43] and Bennett and Bez [19] respectively. Beyond Loomis–Whitney-type data, the factorisation method has been effective in dealing with nonlinear perturbations of a class of data originating in work of Finner [59] (see Carbery, Hänninen and Valdimarsson [43]), and the method of refinements is behind the aforementioned examples of Stovall.

Our approach is based on the method of induction-on-scales. This general method has proved to be extraordinarily effective in harmonic analysis since its introduction by Bourgain [31] in the 1990s. The basic idea is to introduce an auxiliary scale to the inequality in hand — in our case, the diameter of the neighbourhood $U$ appearing in the statement of Theorem 1.1 — and show that for some suitable increasing sequence of scales, the relevant bounds are suitably stable as one passes from one scale to the next. Usually there are many inefficiencies in passing between scales, and losses are incurred. As a result the method rarely leads to endpoint estimates, although there are some notable exceptions, such as Tao [88] and Bejenaru, Herr and Tataru [17]. Our proof of Theorem 1.1 is another such exception, also requiring us to find an efficient passage between scales. This turns out to be rather delicate, and in particular, requires us to find a suitably effective version of Lieb’s theorem (Theorem 1.2). This ultimately requires us to address some natural questions in optimisation theory, which appear to be largely absent in the literature, and have some independent interest. Our effective version — the forthcoming Theorem 1.3 — is phrased in terms of “$\delta$-near-extremisers”, also known as “$\delta$-optimal solutions” in the optimisation theory literature (see, for example, Bonnans and Shapiro [30]). We refer to an input $f = (f_j)_{j=1}^{m}$ as a $\delta$-near-extremiser for (1.1) if

\begin{equation}
\text{BL}(L, p; f) := \frac{\int_{\mathbb{R}^n} \prod_{j=1}^{m} f_j^{p_j}(L_j x) \, dx}{\prod_{j=1}^{m} \left( \int_{\mathbb{R}^n} f_j(x) \, dx \right)^{p_j}} \geq (1 - \delta) \text{BL}(L, p).
\end{equation}

Of course, Lieb’s Theorem guarantees that every Brascamp–Lieb functional $f \mapsto \text{BL}(L, p; f)$ has gaussian $\delta$-near-extremisers, for all $\delta > 0$. What it doesn’t tell us is anything quantitative about the set of gaussian $\delta$-near-extremisers, and this is what our effective version
of Lieb’s theorem will do. In order to state this quantified form of Lieb’s theorem, we shall fix \( p \) and underlying dimensions \( n, n_1, \ldots, n_m \), and refer to an \( m \)-tuple of linear maps \( L \) as admissible if each \( L_j : \mathbb{R}^n \to \mathbb{R}^{n_j} \) is a surjection (in other words, \((L, p)\) is a Brascamp–Lieb datum).

**Theorem 1.3** (Effective version of Lieb’s theorem). There exist \( N \in \mathbb{N} \) and \( \delta_0 > 0 \) such that for every \( \delta \in (0, \delta_0) \) and admissible \( L \),

\[
\inf_{\|C_j^* A_j C_j\| \leq \delta^{-N}} \text{BL}_k(L, p; A)^{-2} \leq \text{BL}(L, p)^{-2} + \delta \prod_{j=1}^m \det(L_j L_j^*)^{p_j}.
\]

Here \( C_j = (L_j L_j^*)^{1/2} \).

We remark that whilst \( \text{BL}_k(L, p; A) \) is invariant under the isotropic rescaling \( A \mapsto \lambda A \), the conditions that \( \|C_j^* A_j C_j\| \leq \delta^{-N} \) and \( \|(C_j^* A_j C_j)^{-1}\| \leq \delta^{-N} \) are not (this point will re-emerge in the course of our proof of Theorem 1.1). Also, on first sight it is perhaps a little unclear that Theorem 1.3 is a statement about near-extremisers. However, it is straightforward to conclude the following.

**Corollary 1.4.** Suppose that \((L_0, p)\) is a Brascamp–Lieb datum with finite Brascamp–Lieb constant. Then there exists \( N \in \mathbb{N} \) and \( \delta_0 > 0 \) such that for every \( \delta \in (0, \delta_0) \),

\[
\sup_{\|A_j\|,\|A_j^{-1}\| < \delta^{-N}} \text{BL}_k(L, p; A) \geq (1 - \delta) \text{BL}(L, p)
\]

for all \( L \) sufficiently close to \( L_0 \).

In brief, both of these statements tell us that we may find gaussian \( \delta \)-near-extremisers for the Brascamp–Lieb inequality, whose eccentricity grows at most polynomially in \( 1/\delta \), and that this growth rate can be taken to be locally uniform in the linear maps \( L \). It is perhaps interesting to note that Theorem 1.3 is considerably stronger that Corollary 1.4; in particular, the former is universal in the sense that the constants \( N \) and \( \delta_0 \) are entirely independent of \( L \), given fixed \( p \) and \( n, n_1, \ldots, n_m \).

Theorem 1.3 (or Corollary 1.4) has the potential for application wherever Lieb’s theorem is of use. For instance, the continuity of the classical Brascamp–Lieb constant \( \text{BL}(\cdot, p) \), established using Lieb’s theorem by Bennett, Bez, Cowling and Flock [20], and Garg, Gurvits, Oliveira and Wigderson [64], may easily be improved to local Hölder continuity using Corollary 1.4 — see Section 5 for a short proof of this fact.

Our approach to Theorem 1.3 is to reduce it to a similar statement involving “\( \delta \)-near-minimisers” for functions of the form

\[
\mathbb{R}^n \ni y \mapsto \sum_{j \in J} d_j \exp((u_j, y)),
\]

where each \( d_j \) is a nonnegative real number (coefficient), and each \( u_j \) is a vector in \( \mathbb{R}^n \). Here, we are referring to a vector \( y_\delta \) as a \( \delta \)-near-minimiser for the function in (1.17) if

\[
\sum_{j \in J} d_j \exp((u_j, y_\delta)) \leq \inf_{y \in \mathbb{R}^n} \sum_{j \in J} d_j \exp((u_j, y)) + \delta \max_{j \in J} d_j.
\]

Notice that this is a near-minimiser in an “additive” sense; this differs somewhat from our earlier definition of a \( \delta \)-near-extremiser in (1.15) for the Brascamp–Lieb functional. An elementary example of a function of the form (1.17) is the hyperbolic cosine, which of course attains its minimum at the origin, although minimisers do not need to exist in general — for instance if \( J \) contains only one element. Our main result in this setting is the following.
Theorem 1.5. There exist $N \in \mathbb{N}$ and $\delta_0 > 0$, depending only on $(u_j)_{j \in J}$, such that for every $\delta \in (0, \delta_0)$ and nonnegative coefficients $d_j$,

$$\inf_{|y| \leq N \log(1/\delta)} \sum_{j \in J} d_j \exp(\langle u_j, y \rangle) \leq \inf_{y \in \mathbb{R}^n} \sum_{j \in J} d_j \exp(\langle u_j, y \rangle) + \delta \max_{j \in J} d_j.$$

An important feature of Theorem 1.5 is the uniformity of the conclusion in the coefficients $d_j$. Theorem 1.5 tells us that we can always find a $\delta$-near-minimiser $y$ within a distance $N \log(1/\delta)$ of the origin, provided that $N$ and $1/\delta$ are sufficiently large (depending only on the $u_j$). We briefly describe a simple example that illustrates the nature and essential optimality of this result in Section 5.

Structure of the paper. In Section 2 we present our applications of the nonlinear Brascamp–Lieb inequality (Theorem 1.1) — see the forthcoming Corollary 2.1 and Theorem 2.2. In Section 3 we describe the strategy behind our proof of Theorem 1.1 with emphasis on the inductive process. In Section 4 we develop a near-optimisation theory of functions of the form (1.17), proving Theorem 1.5. In Section 5 we deduce Theorem 1.3 from Theorem 1.5 and in Section 6 we prove Theorem 1.1.

Acknowledgments. The authors wish to thank Tony Carbery for many in-depth discussions on the subject of this paper. They also thank Alessio Martini for discussions surrounding the local sharp version of Young’s convolution inequality.

2. Applications

In this section we describe three settings in which the nonlinear Brascamp–Lieb inequality (Theorem 1.1) has significant applications. This also serves as an illustration of the prevalence of nonlinear Brascamp–Lieb forms in analysis.

The first application concerns best constants in Young’s convolution inequality on an abstract Lie group, and capitalises on the optimality of the nonlinear Brascamp–Lieb constant in Theorem 1.1. The second provides a broad family of multilinear convolution inequalities that arise naturally in dispersive PDE, where the generality of Theorem 1.1 allows for rather minimal structural hypotheses. The third interprets Theorem 1.1 in the context of the general multilinear theory of oscillatory integral operators of Hörmander type.

2.1. Abstract harmonic analysis. In abstract harmonic analysis, the question of determining the best constants in the Young convolution inequality and its cousin, the Hausdorff–Young inequality, has often been discussed, particularly following the work of Babenko [4], Becker [12, 13] and Brascamp–Lieb [38] which covered the euclidean space $\mathbb{R}^n$. For example, in the 1970s, Fournier [60] showed that the optimal constant in both inequalities is less than 1 for nonabelian groups with no compact open subgroups, and Klein and Russo [74] calculated them in terms of the euclidean constants for some nonabelian Lie groups including the Heisenberg groups. Later, García–Cuerva, Marco and Parcet and then García–Cuerva and Parcet [61, 62] considered this and related constants in the context of Banach space structure, but without finding any numerical values.

In recent work, Cowling, Martini, Müller and Parcet [52] computed the best constant for the Young convolution inequality and Hausdorff–Young inequality for a more extensive class of nonabelian Lie groups, and established a lower bound for general Lie groups when the functions are restricted to lie in small neighbourhoods of the identity. More precisely, in the case of the Young convolution inequality, for a connected unimodular Lie group $G$ and triple of exponents $\mathbf{q} = (q_1, q_2, q_3) \in [1, \infty]^3$ satisfying $\sum_{j=1}^3 1/q_j = 2$, they define the
constant $Y_q(G)$ to be the best value of the constant $C$ in the inequality
\[
\int_G \int_G f_1(y) f_2(y^{-1} x) f_3(x) \, dy \, dx \leq C \prod_{j=1}^{3} \|f_j\|_{q_j},
\]
where $f_j$ are nonnegative functions in $L^{q_j}(G)$, for $j = 1, 2, 3$. The constant $Y_q(G; U)$ is defined similarly, but the functions $f_1, f_2, f_3$ are constrained to be supported in a relatively compact neighbourhood $U$ of the identity in $G$. Evidently the constants $Y_q(G; U)$ get smaller as $U$ gets smaller, and so it makes sense to define the constant $Y_q^{\text{loc}}(G)$ to be the infimum of the $Y_q(G; U)$ as $U$ shrinks to the identity. It is easy to see that
\[
Y_q^{\text{loc}}(G) \leq Y_q(G) \leq 1;
\]
the right-hand inequality follows from standard interpolation arguments.

In this definition, the Lebesgue spaces are defined in terms of a left-invariant Haar measure on $G$. The natural version of Young’s inequality for convolution in the context of right-invariant Haar measures on general nonunimodular groups involves powers of the modular function as well; however, the modular function is bounded and bounded away from 0 on relatively compact neighbourhoods of the identity in $G$, and as the neighbourhood shrinks, the modular functions becomes closer to 1. As a result, one might also define $Y_q^{\text{loc}}(G)$ with a right-invariant measure, and the value of the constant would be the same as for a left-invariant measure.

In the previous papers on the subject, estimates were obtained for $Y_q(G)$ and $Y_q^{\text{loc}}$ for various types of Lie groups. In particular, upper bounds were found for both constants on solvable Lie groups, and for similarly defined constants for central functions on compact Lie groups, but for general functions on semisimple Lie groups, say, there were no nontrivial upper bounds for $Y_q(G)$ or $Y_q^{\text{loc}}(G)$ before this work.

By using a contraction argument, it was shown in [52, Proposition 2.4] that
\[
Y_q^{\text{loc}}(G) \geq (A_{q_1} A_{q_2} A_{q_3})^{\dim(G)},
\]
where, for any $q \in [1, \infty]$, the constant $A_q$ is given by
\[
A_q = \left( \frac{q^{1/q}}{q^{1/q}} \right)^{1/2}.
\]
Naturally this constant may be reconciled with that in [14].

Combined with the inequality (2.1) and the estimates for solvable groups, it followed that $Y_q(G)$, for a connected simply connected unimodular solvable group $G$, is exactly the right-hand side of (2.2). In particular, $Y_q(G)$ was known for many interesting groups, such as the Heisenberg groups that appear in several areas of complex analysis and partial differential equations. However, for many Lie groups, it remained the case that the only upper bound for either constant ($Y_q(G)$ or $Y_q^{\text{loc}}(G)$) was the trivial bound (2.1). The following corollary of the nonlinear Brascamp–Lieb inequality (Theorem 1.1) identifies the local constant $Y_q^{\text{loc}}(G)$ in full generality.

**Corollary 2.1.** Suppose that $G$ is a connected Lie group. Then
\[
Y_q^{\text{loc}}(G) = (A_{q_1} A_{q_2} A_{q_3})^{\dim(G)}.
\]

It is easy to see that this corollary follows from Theorem 1.1, as the differentials of the underlying group multiplication and inversion mappings at the identity correspond to the linear mappings in the classical version of Young’s convolution inequality, via the Baker–Campbell–Hausdorff formula.

By the way, Young’s convolution theorem shows that in general, local Brascamp–Lieb inequalities do not give rise to global inequalities. For example, for a compact Lie group $G$, we have shown that $Y_q^{\text{loc}}(G) < 1$, but $Y_q(G) = 1$, so the global constant is larger than the
local constant. We refer the reader to Cowling, Martini, Müller and Parcet [52] for much more on this topic, including many references and its connection with the best constant in the Hausdorff–Young inequality.

2.2. Euclidean harmonic analysis and partial differential equations. In the analysis of nonlinear dispersive equations, estimates involving iterated convolutions frequently arise. A widely used technique (see, for example, Beals [11], Bourgain [32], Klainerman and Machedon [74]) in the study of the local well-posedness of subcritical problems is the use of $X^s_{S,S}^b$ spaces, associated with an underlying surface $S$, in an iteration argument. In this way, it is desirable to establish control on the nonlinearity in the setting of these function spaces and, under certain structural assumptions on the nonlinearity, it is often possible to reduce estimates of this type to multilinear convolution estimates. Naturally it is desirable to have a general framework for such estimates, and a systematic study of weighted convolution estimates of the form

\begin{equation}
\int_H f_1 \otimes \cdots \otimes f_m(y) \, w(y) \, d\mu_H(y) \leq C \prod_{j=1}^m \|f_j\|_{L^2(Z)}
\end{equation}

for $L^2$ functions was carried out by Tao [87]. Here, $Z$ is an abelian group, written additively, $H$ is the subgroup $\{y \in Z^m : y_1 + \cdots + y_m = 0\}$ of $Z^m$, and $\mu_H$ is a Haar measure and $w$ is a (weight) function on $H$. We refer the reader to [87] for further details concerning the motivation for estimates of the type (2.4), along with applications to bilinear estimates associated with the KdV, wave and Schrödinger equations.

Multilinear convolution estimates of a more singular nature, which fall outside the scope of (2.4), have appeared in recent breakthroughs in the low-regularity theory of the Zakharov system by Bejenaru–Herr–Holmer–Tataru [16] and Bejenaru–Herr [15]. These estimates provide bounds on multiple convolutions of Lebesgue space densities supported on certain transversal submanifolds of euclidean space, and ultimately rely on the nonlinear Brascamp–Lieb inequality with Loomis–Whitney-type data, described following (1.9). In the spirit of Tao’s work [87] in the framework (2.4), here we use the general nonlinear Brascamp–Lieb inequality (Theorem 1.1) to establish a much broader class of multilinear singular convolution estimates, where the transversality condition on the family of submanifolds is in some sense minimal. After introducing some notation, we state this application, and then elaborate further on the manner in which such estimates arise in the analysis of dispersive partial differential equations.

For $1 \leq j \leq m$, we let $\Sigma_j : U_j \to \mathbb{R}^n$ be a parametrisation of a $C^2$, bounded and $n_j$-dimensional submanifold $S_j$ of $\mathbb{R}^n$, and write $d\Sigma_j : \mathbb{R}^{n_j} \to \mathbb{R}^n$ for its differential. We equip $S_j$ with the measure $\sigma_j$, which is the push-forward of Lebesgue measure on $U_j$, that is,

$$\int_{S_j} g(y) \, d\sigma_j(y) = \int_{U_j} g(\Sigma_j(x)) \, dx$$

for any continuous function $g$ on $S_j$. For notational convenience, and without loss of generality, we may suppose that these surfaces are parametrised in such a way that $0 \in U_j$ for each $j$, so that the image of $d\Sigma_j(0)$ is the tangent space to $S_j$ at the (arbitrary) point $\Sigma_j(0)$. Finally, we denote by $d\Sigma(0)^*$ the vector of transposes of the differentials $(d\Sigma_1(0)^*, \ldots, d\Sigma_m(0)^*)$.

**Theorem 2.2.** Suppose that $\text{BL}(d\Sigma(0)^*, p) < \infty$. Then

\begin{equation}
\|g_1 \, d\sigma_1 \ast \cdots \ast g_m \, d\sigma_m\|_{L^\infty(\mathbb{R}^n)} \leq C \prod_{j=1}^m \|g_j\|_{L^{q_j}(S_j)}
\end{equation}

for all $g_j \in L^{q_j}(S_j)$ with sufficiently small support, where $q_j = (1/p_j)' = 1/(1 - p_j)$. 

The first thing to observe is that the condition $BL(d\Sigma(0)^*, p) < \infty$ may be replaced with the manifestly geometric condition
\begin{equation}
\sum_{j=1}^{m} \frac{\dim(V \cap (T_{\Sigma_j(0)} S_j)^\perp)}{q_j} \leq (\rho - 1) \dim(V);
\end{equation}
here $\rho = \sum_{j=1}^{m} 1/q_j$, and (2.6) should hold for all subspaces $V$ of $\mathbb{R}^n$. This is simply an interpretation of the finiteness condition (1.13) in this context, and we clarify that $T_{\Sigma_j(0)} S_j$ denotes the tangent space to $S_j$ at $\Sigma_j(0)$. We note that the scaling condition (1.14), which corresponds to equality in (2.6) when $V = \mathbb{R}^n$, ceases to be relevant as the surfaces $S_j$ are compact, and so Lebesgue exponents may always be raised by Hölder’s inequality.

Of course (2.5) may be interpolated with the elementary bound
$$
\|g_1 d\sigma_1 \ast \cdots \ast g_m d\sigma_m\|_{L^1(\mathbb{R}^n)} \leq \prod_{j=1}^{m} \|g_j\|_{L^1(S_j)};
$$
to provide $L^p$ estimates on $g_1 d\sigma_1 \ast \cdots \ast g_m d\sigma_m$ for all $r \in [1, \infty]$.

Theorem 2.2 is closely-related to the conjectural “oscillatory Brascamp–Lieb inequality” stated in [21]. This very general conjectural inequality states that, under the hypotheses of Theorem 2.2,
\begin{equation}
\int_{\mathbb{R}^n} \prod_{j=1}^{m} |g_j d\sigma_j(x)|^{2p_j} \, dx \leq C \prod_{j=1}^{m} \|g_j\|_{L^{p_j}(S_j)}^{2p_j};
\end{equation}
see [21] for a weaker version of this inequality, where the integration on the left-hand side is restricted to a ball of radius $R$, and an additional factor of $R^d$ appears on the right-hand side. This inequality, which we mentioned briefly in the introduction, generalises the multilinear restriction inequality [27], along with those used in the recent theory of decouplings; see, for example [34], [35] and [36].

In particular, if $p_j \leq 1/2$ (which equates to $1 \leq q_j \leq 2$), for all $j$, then (2.7) quickly implies (2.5), at least when the $g_j$ are characteristic functions of sets. To see this, we observe that
\begin{align*}
\|g_1 d\sigma_1 \ast \cdots \ast g_m d\sigma_m\|_{L^\infty(\mathbb{R}^n)} &\leq \int_{\mathbb{R}^n} \prod_{j=1}^{m} \|g_j d\sigma_j(x)\| \, dx \\
&\leq \prod_{j=1}^{m} \|g_j d\sigma_j\|_{L^{\infty}(\mathbb{R}^n)}^{1-2p_j} \int_{\mathbb{R}^n} \prod_{k=1}^{m} \|g_k d\sigma_k(x)\|^{2p_k} \, dx \\
&\leq C \prod_{j=1}^{m} \|g_j\|_{L^1(S_j)} \prod_{k=1}^{m} \|g_k\|_{L^{2p_k}(S_k)}^{2p_k} = C \prod_{j=1}^{m} \|g_j\|_{L^{p_j}(S_j)}^{2p_j}.
\end{align*}
Here we have used the trivial $L^1 \rightarrow L^\infty$ bound for the extension operators, along with the conjectural inequality (2.7), and the assumption that the $g_j$ are characteristic functions of sets. Similar remarks in the special case of the Loomis–Whitney inequality were made in [13] and [19].

Our proof of Theorem 2.2 will appear at the very end of this section. Before that, we provide a number of additional comments, the first of which is that the previously known estimates subsumed by Theorem 2.2 are restricted to submanifolds that satisfy restrictive transversality assumptions. This is to be expected given that estimates of the type (2.6) prior to this work all relied upon nonlinear Loomis–Whitney-type inequalities, as discussed following (1.9).

Additionally, our proof of Theorem 2.2 is more revealing and hinges on Proposition 2.3 below, which we believe to be of independent interest in the theory of the standard Brascamp–Lieb inequality. As a prelude to this, we note that a superficial connection...
between estimates of the form (2.5) to nonlinear Brascamp–Lieb inequalities can be seen rather immediately since pointwise evaluation of the singular convolution on the left-hand side of (2.3) is given by

\[ g_1 d\sigma_1 \ast \cdots \ast g_m d\sigma_m(x) = \int_{U_1 \times \cdots \times U_m} f_1 \otimes \cdots \otimes f_m(y) \delta(F(y, x)) \, dy, \]

where \( f_j = g_j \circ \Sigma_j \), and

\[ F(y, x) = \Sigma_1(y_1) + \cdots + \Sigma_m(y_m) - x. \]

However, the proof of Theorem 2.2 contains subtleties which are related to the appearance of the dual exponent \( q_j' \) in the transversality assumption. Indeed, an interesting feature of our proof of Theorem 2.2 is the use of the relationship between the best constants \( BL(H, \mathbf{q}) \) and \( BL(H^\perp, \mathbf{q}') \) given in Proposition 2.3 below. For a given subspace \( H \) of \( \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m} \) and \( m \)-tuple of exponents \( \mathbf{q} = (q_j)_{j=1}^m \in [1, \infty]^m \), \( BL(H, \mathbf{q}) \) denotes the smallest constant for which the inequality

\[ \int_H f_1 \otimes \cdots \otimes f_m \, d\mu_H \leq BL(H, \mathbf{q}) \prod_{j=1}^m \| f_j \|_{L^{n_j}(\mathbb{R}^{n_j})} \]

holds over all (nonnegative integrable) functions \( f_j \), where integration on the left-hand side is taken with respect to the Lebesgue measure on \( H \). Here, we are abusing notation a little, given our earlier definition of \( BL(L, \mathbf{p}) \) as the smallest constant in the Brascamp–Lieb inequality (1.1). However, in light of the correspondence between (1.1) (or (1.2)) and (2.10), this dual role of \( BL \) is arguably quite natural.

In order to proceed, we also define the constant

\[ B_{\mathbf{q}} = \prod_{j=1}^m A_{q_j} \]

for \( \mathbf{q} = (q_j)_{j=1}^m \in [1, \infty]^m \), where the notation \( A_{\mathbf{q}} \) was defined earlier in (2.3). The constants \( A_{\mathbf{q}} \) arise in the best constant for the Young convolution inequality and, as will be discussed after the proof of the proposition, this is not a coincidence.

**Proposition 2.3.** With the notation of (2.10) and (2.3), \( BL(H, \mathbf{q}) = B_{\mathbf{q}} BL(H^\perp, \mathbf{q}') \).

**Proof.** Since \( B_{\mathbf{q}}^{-1} = B_{\mathbf{q}'} \), it suffices to prove the inequality

\[ BL(H, \mathbf{q}) \leq B_{\mathbf{q}} BL(H^\perp, \mathbf{q}'). \]

To see this, we use Lieb’s theorem (Theorem 1.2) to write

\[ BL(H, \mathbf{q}) = \lim_{\nu \to 0} BL(H, \mathbf{q}; f_{\nu}), \]

where \( f_{\nu,j}(x) = \exp(-\langle A_{\nu,j} x, x \rangle) \) for a suitable family of positive definite matrices \( A_{\nu,j} \). By a generalised form of Parseval’s identity,

\[ \int_H f_1 \otimes \cdots \otimes f_m \, d\mu_H = \int_{H^\perp} \hat{f}_1 \otimes \cdots \otimes \hat{f}_m \, d\mu_{H^\perp} \]

for all sufficiently nice \( f_j \) (the integrals involve the Lebesgue measures on \( H \) and \( H^\perp \)), and therefore it follows that

\[ BL(H, \mathbf{q}; f_{\nu}) = BL(H^\perp, \mathbf{q}; \hat{f}_{\nu}) \prod_{j=1}^m \| \hat{f}_{\nu,j} \|_{q_j'} \]

for each \( \nu > 0 \). An explicit computation reveals that

\[ \prod_{j=1}^m \| \hat{f}_{\nu,j} \|_{q_j'} = B_{\mathbf{q}} \]
for each \( \nu > 0 \), and thus (2.11) follows. \( \square \)

Derived from the involution \((H, q) \mapsto (H^\perp, q')\), there is an involution acting on Brascamp–Lieb data \((L, p)\), which gives rise to linear mappings \(\widetilde{L}_j : \mathbb{R}^n \to \mathbb{R}^{n_j}\) determined (not uniquely) by

\[
\sum_{j=1}^m L_j^\ast \widetilde{L}_j = 0.
\]

As a concrete instance of this, one may easily check that data associated with Young’s convolution inequality is transformed to data associated with the trilinear Hölder inequality; the best constant in the latter case is equal to one and thus Proposition 2.3 generates the best constant in Young’s convolution inequality. Note that we are not providing any kind of simplification to the derivation of the best constant in Young’s convolution inequality since, in the course of our proof of Proposition 2.3, we made use of Lieb’s theorem. However, it provides an interesting perspective which may be fruitful in similar contexts where, for example, the data \(\widetilde{L}_j\) is rather more manageable than the data \(L_j\).

Returning to the discussion on Theorem 2.2, as will be obvious from the proof below, we remark that the implicit constant in (2.5) may be quantified. For any given \(\varepsilon > 0\), there exist a sufficiently small neighbourhood of the origin such that (2.5) holds with constant

\[(1 + \varepsilon)B_q BL(d\Sigma(0)^\ast, p).
\]

As we have already mentioned, particular instances of the estimate (2.5) have arisen in [16] and [15] concerning the Zakharov system:

\[
(i\partial_t + \Delta) u = nu
\]

\[
(\partial_t^2 - \Delta) n = \Delta |u|^2,
\]

for \(u : \mathbb{R}^{d+1} \to \mathbb{C}\) and \(n : \mathbb{R}^{d+1} \to \mathbb{R}\) with initial data

\[(u(0), n(0), \partial_t n(0)) = (u_0, n_0, n_1).
\]

Formulated in [93], this arises as a model in plasma physics, and much effort has been spent developing the well-posedness theory of this system (see, for example, [15, 16, 33, 65, 81, 86]). For initial data in \(L^2\)-based Sobolev spaces, the papers [16] and [15] provided definitive results on the local well-posedness of the system in two and three spatial dimensions, respectively, in the subcritical regime. The iteration argument in [16] involves multilinear estimates in terms of certain \(X_{s,b}^{1}\) spaces, where \(S\) is either a paraboloid or a cone in \(\mathbb{R}^3\). The underlying qualitative phenomenon is the fact that, given three \(C^2\), bounded, and transversal \(S_1, S_2\) and \(S_3\) in \(\mathbb{R}^3\), the convolution of two \(L^2\) functions supported on two of the surfaces has a well-defined restriction, as an \(L^2\) function, to the third surface \(S_3\). This phenomenon is captured by the estimate

\[
\|g_1 \sigma_1 \ast g_2 \sigma_2\|_{L^2(S_1)} \leq C \|g_1\|_{L^2(S_1)} \|g_2\|_{L^2(S_2)},
\]

or equivalently, by duality,

\[
(2.13) \quad |g_1 \sigma_1 \ast g_2 \sigma_2 \ast g_3 \sigma_3(0)| \leq C \prod_{j=1}^3 \|g_j\|_{L^2(S_j)}.
\]

The transversality assumption on the hypersurfaces means that the underlying nonlinear Brascamp–Lieb inequality is the nonlinear Loomis–Whitney inequality, discussed earlier. The specific bound (2.13), in the quantitative form of relevance to applications, was established in [17] by an induction-on-scales argument.

\footnote{In this context, if \(\tilde{n}_1, \tilde{n}_2\) and \(\tilde{n}_3\) are linearly independent for any choice of normal vector \(\tilde{n}_j\) to \(S_j\), we say that the \(S_j\) are transversal.}
There are further, quite varied applications of the nonlinear Brascamp–Lieb inequality in PDE-related problems. For example, in their study of the weakly nonlinear large-box limit of the two-dimensional cubic nonlinear Schrödinger equation, Faou, Germain and Hani [58] derive a new nonlinear integro-differential equation governed by a trilinear form that draws on the nonlinear Loomis–Whitney inequality from [28]. Also, in the spirit of [16] and [15], nonlinear Loomis–Whitney-type inequalities have been used very recently to push forward the mathematical theory of the Klein–Gordon–Zakharov system in two dimensions [73] and a system of quadratic derivative nonlinear Schrödinger equations [70].

In somewhat different territory, explicit examples of multilinear Radon-like transforms of the form (1.7) have also appeared in obstacle scattering, in particular, in the recovery of singularities of a potential \( q \) by its so-called Born series \( q_B \); we refer the reader to [83] (see also [23]) for further background and discussion on the manner in which bounds of the form (1.7) naturally arise in this context. It seems reasonable to expect further such applications arising from Theorems 1.1 and 2.2.

**Proof of Theorem 2.2.** By considering input functions with sufficiently small supports, it suffices to prove that

\[
|g_1 d\sigma_1 \ast \cdots \ast g_m d\sigma_m(0)| \leq C \prod_{j=1}^m \|g_j\|_{L^{q_j}(S_j)}
\]

and hence, from (2.8), it suffices to prove that

\[
\int_M f_1 \otimes \cdots \otimes f_m d\mu \leq C \prod_{j=1}^m \|f_j\|_{L^{q_j}(\mathbb{R}^n_j)}
\]

for \( f_j \) supported in a sufficiently small neighbourhood of the origin, where

\[
M = \{(y_1, \ldots, y_m) \in U_1 \times \cdots \times U_m : F(y, 0) = 0\},
\]

\( \mu \) is a suitable measure on \( M \), and \( F \) is given by (2.9). By Theorem 1.1 (interpreted in terms of the multilinear form (1.7)), it suffices to prove that \( \text{BL}(T_0 M, q) < \infty \) where \( T_0 M \) is the tangent space to \( M \) at the origin. Since

\[
(T_0 M)^\perp = \{(d\Sigma_1(0)^* x, \ldots, d\Sigma_m(0)^* x) : x \in \mathbb{R}^n\}
\]

and given the assumption \( \text{BL}(d\Sigma(0)^* x, p) < \infty \), Theorem 2.2 now follows from Proposition 2.3. \( \square \)

**2.3. Multilinear oscillatory integrals.** The purpose of this section is to interpret Theorem 1.1 in the wider context of oscillatory integral operators, where it provides some concrete evidence for the viability of some far-reaching conjectural inequalities. These inequalities are a further generalisation of the oscillatory Brascamp–Lieb inequality (2.7) (which is also open for all but the most elementary Brascamp–Lieb data) to the setting of oscillatory integral operators of Hörmander type. These operators take the form

\[
T_\lambda f(\xi) = \int_{U'} e^{i\lambda \Phi(x, \xi)} f(x) \, dx, \quad \xi \in U,
\]

where \( \Phi \) is a smooth real-valued phase function on \( U' \times U \). Here \( U' \) and \( U \) are compact subsets of \( \mathbb{R}^{n'} \) and \( \mathbb{R}^n \) respectively, where \( n' \leq n \), and \( \lambda \) is a large real parameter.

**Theorem 2.4 (Hörmander [71]).** If \( n' = n \) and

\[
(2.14) \quad \det(\nabla_x \nabla_\xi \Phi(x, \xi)) \neq 0
\]

for all \((x, \xi) \in U' \times U\), then

\[
\|T_\lambda f\|_{L^2(\mathbb{R}^n)} \leq C \lambda^{-n/2} \|f\|_{L^2(\mathbb{R}^n)}.
\]
We remark that Hörmander’s theorem is usually stated for operators of the form

\[ \tilde{T}_\lambda f(\xi) = \int_{\mathbb{R}^n} e^{i\lambda \Phi(x,\xi)} \psi(x,\xi) f(x) \, dx, \]

where \( \psi \) is a smooth function of compact support on \( \mathbb{R}^n \times \mathbb{R}^n \) — see, for example, Stein [84]. However, it is easily seen that such bounds on \( T_\lambda \) and \( \tilde{T}_\lambda \) are interchangeable by writing \( \psi(x,\xi) \) in terms of its Fourier transform \( \widehat{\psi} \) via the inversion formula, and using the integrability of \( \widehat{\psi} \). A similar remark applies to Conjecture 2.5 below.

In order to state a Brascamp–Lieb-type generalisation of Hörmander’s theorem we introduce Hörmander operators \( T_{j,\lambda} \) associated with smooth phase functions \( \Phi_j : U_j \times U \to \mathbb{R} \). Here \( U_j \) and \( U \) are compact subsets of \( \mathbb{R}^n \) and \( \mathbb{R}^n \) respectively, and \( 1 \leq j \leq m \).

**Conjecture 2.5.** If

\[ \dim(V) \leq \sum_{j=1}^m p_j \dim(\nabla_x \nabla_\xi \Phi_j(x_j,\xi)^t V) \]

for all subspaces \( V \) of \( \mathbb{R}^n \), with equality when \( V = \mathbb{R}^n \), for all \( (x_j,\xi) \in U_j \times U \), \( 1 \leq j \leq m \), then

\[ \int_{\mathbb{R}^n} \prod_{j=1}^m |T_{j,\lambda} f_j(x)|^{2p_j} \, dx \leq C_{\lambda}^{-n} \prod_{j=1}^m \|f_j\|_{L^2(\mathbb{R}^n)}^{2p_j}. \]

It should be noticed that the hypothesis (2.15) (with equality when \( V = \mathbb{R}^n \)) is equivalent to the statement that \( BL(L^p,\mathbb{R}^n) < \infty \) for all \( (x_j,\xi) \in U_j \times U \), where \( L^{p,\xi}_j = \nabla_x \nabla_\xi \Phi_j(x_j,\xi)^t : \mathbb{R}^n \to \mathbb{R}^{n_j} \) for each \( 1 \leq j \leq m \). This is an immediate consequence of the finiteness characterisation given by (1.13) and (1.14).

Conjecture 2.5 has its origins in [27], and is a generalisation of (2.7) to phases that may be nonlinear in the \( \xi \)-variable; see also [18]. In [27] the inequality (2.10) is obtained with an additional factor of \( \lambda^c \) on the right-hand side in the particular case of Loomis–Whitney data, and this weak form of (2.10) may be verified for general data by combining the wavepacket analysis methods of [27] and [21] in a by now routine manner. Specifically we have the following.

**Theorem 2.6.** Let \( \varepsilon > 0 \). Under the hypotheses of Conjecture 2.5

\[ \int_{\mathbb{R}^n} \prod_{j=1}^m |T_{j,\lambda} f_j(x)|^{2p_j} \, dx \leq C_\varepsilon \lambda^{-n-\varepsilon} \prod_{j=1}^m \|f_j\|_{L^2(\mathbb{R}^n)}^{2p_j}. \]

We refer the reader to Bourgain and Guth [37] and Guth, Hickman and Iliopoulou [69] for applications of these and related inequalities.

Conjecture 2.5 is known only in extremely elementary situations, such as when it may be deduced from Hörmander’s theorem via a use of Hölder’s inequality (the underlying Brascamp–Lieb inequality being an instance of Fubini’s theorem). However, Theorem 2.6 constitutes progress in two concrete respects: (i) it allows us to verify Conjecture 2.5 for phases \( \Phi_j \) that are linear in the \( x \)-variable, and (ii) it provides us with a simplified proof of Theorem 2.6 avoiding the use of Kakeya information, and even avoiding the use of wavepacket decompositions, appealing instead to more classical frequency decompositions.

For (ii), which has some similarities with Bejenaru [14], we refer the reader to [24]. To see (i), we simply observe that if \( \Phi_j(x_j,\xi) = x_j \cdot B_j(\xi) \) for some smooth submersion \( B_j : U \to \mathbb{R}^{n_j} \), then

\[ T_{j,\lambda} f_j(\xi) = \hat{f}_j(B_j(\xi)) \]

whenever \( \text{supp}(f_j) \subseteq U_j \), \( 1 \leq j \leq m \). The claim then follows by applying Theorem 1.1 to the functions \( |\hat{f}_j|^2 \), and applying Plancherel’s theorem.
Establishing Conjecture 2.5 would appear to be very difficult, even in some relatively simple cases. For example, the endpoint multilinear restriction inequality of [27] is a particular case, and is very much still open. It is hoped that the perspectives of this paper may eventually lead to further developments in this interesting direction.

3. The proof strategy

As we describe in the introduction, a key element in our proof of Theorem 1.1 is Theorem 1.3 — our effective version of Lieb’s theorem. Our mechanism for applying Theorem 1.3 is a rather delicate instance of the method of induction-on-scales. The purpose of this short section is to explain our general strategy for achieving this, avoiding reference to technical aspects.

The basic applicability of induction-on-scales in the setting of the Brascamp–Lieb inequality is easily seen in a fundamental inequality originating in work of Ball [5]. This inequality captures an important self-similarity property of the Brascamp–Lieb functional

\[ f \mapsto \text{BL}(L, p; f) = \frac{\int_{\mathbb{R}^n} \prod_{j=1}^m f_j^{p_j}(L_j x) \, dx}{\prod_{j=1}^m \left( \int_{\mathbb{R}^n} f_j(x_j) \, dx_j \right)^{p_j}}. \]

As we shall now see, this property strongly suggests approaching any sort of “perturbation” of the Brascamp–Lieb inequality by the method of induction-on-scales. This has been fruitful on several occasions — see [18, 27, 17, 19, 68, 21], along with closely-related papers in Fourier restriction theory beginning with [31].

We begin by observing that for two \( L^1 \)-normalised inputs \( f \) and \( g \) (in the sense that each \( f_j \) and \( g_j \) is a probability density),

\[
\text{BL}(L, p; f) \, \text{BL}(L, p; g) = \int_{\mathbb{R}^n} \left( \prod_{j=1}^m (f_j^{p_j} \circ L_j) \right) \ast \left( \prod_{k=1}^m (g_k^{p_k} \circ L_k) \right) (x) \, dx
\]

\[
= \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} \prod_{j=1}^m [f_j(L_j y) g_j(L_j(x - y))]^{p_j} \, dy \right) \, dx
\]

\[
= \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} \prod_{j=1}^m [f_j(L_j y) g_j(L_j x - L_j y)]^{p_j} \, dy \right) \, dx
\]

\[
= \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} \prod_{j=1}^m [h_j^x(L_j y)]^{p_j} \, dy \right) \, dx,
\]

where \( h_j^x(z) = f_j(z) g_j(L_j x - z) \) for each \( j \). Writing \( h^x = (h_j^x)_{j=1}^m \) and \( f \ast g = (f_j \ast g_j)_{j=1}^m \), we deduce that

\[
\text{BL}(L, p; f) \, \text{BL}(L, p; g) \leq \int_{\mathbb{R}^n} \text{BL}(L, p; h^x) \prod_{j=1}^m \left( \int_{\mathbb{R}^n} h_j^x(z_j) \, dz_j \right)^{p_j} \, dx
\]

\[
= \int_{\mathbb{R}^n} \text{BL}(L, p; h^x) \prod_{j=1}^m (f_j \ast g_j(L_j x))^{p_j} \, dx
\]

\[
\leq \sup_x \text{BL}(L, p; h^x) \, \text{BL}(L, p; f \ast g).
\]

This we refer to as Ball’s inequality. It should be noted that the \( L^1 \) normalisation hypotheses on \( f \) and \( g \) may be dropped by homogeneity considerations.

Now, if \( g \) is a \( \delta \)-near-extremiser for (1.1), in the sense that (1.15) holds, then we may immediately deduce that

\[
\text{BL}(L, p; f) \leq (1 + O(\delta)) \, \text{BL}(L, p; f \ast g)
\]
and
\[
(3.3) \quad \text{BL}(L, p; f) \leq (1 + O(\delta)) \sup_x \text{BL}(L, p; h^x).
\]

The inequalities (3.2) and (3.3) contain a surprising amount of information. For example, (3.2) tells us that the set of extremisers for (1.1) is closed under convolution — a fact that may be used along with the central limit theorem to deduce the existence of gaussian extremisers (see [25]), provided that the set of extremisers is nonempty of course. In the presence of a gaussian extremiser, inequality (3.2) may be used again to deduce that the Brascamp–Lieb functional is nondecreasing as the inputs evolve under a suitable heat flow — see Carlen, Lieb and Loss [46] and Bennett, Carbery, Christ and Tao [25] for more information on this heat-flow perspective on the Brascamp–Lieb inequality.

The closely-related inequality (3.3), while more complicated, will be more important for our purposes. Informally, if we happen to have a $\delta$-near-extremising input $g$ which resembles a bump function with small support, then the function
\[
h_j^x(z) = f_j(z)(L_jx - z)
\]
is simply the function $f_j$ localised to a small neighbourhood of the point $L_jx$. The inequality (3.3) then tells us that the Brascamp–Lieb functional is close to increasing ($\delta$ being small) as the general input functions $f_j$ are localised in this way. Our approach to proving Theorem 1.1 is motivated by this, and amounts to proving a suitable nonlinear variant of (3.3). This approach is particularly natural, as input functions $f_j$ with “shrinking support” will be increasingly unable to detect nonlinear structure in $B_j$, ultimately allowing us to reduce matters to a linear Brascamp–Lieb inequality. Proving such a nonlinear variant of (3.3) presents several difficulties. Perhaps the most obvious difficulty is in finding a substitute for the explicit use of linearity in (3.1). Another difficulty relates to the fact that $(L, p)$ may only have gaussian near-extremisers, which needless to say, are not compactly supported. It turns out to be relatively straightforward to overcome this by using suitably truncated gaussians. The main difficulty we encounter, however, stems from the fact that Lieb’s theorem (Theorem 1.1) does not provide any quantitative information about the set of gaussian $\delta$-near-extremisers. In particular, it does not tell us whether there exist $\delta$-near-extremisers that are sufficiently “localised” to allow the inductive process to run. What is needed is the availability of gaussian near-extremisers at all scales, whose “eccentricities” are suitably controlled. It is here where our effective version of Lieb’s theorem (Theorem 1.3) plays a crucial role. Once a suitable nonlinear variant of (3.3) has been obtained (the forthcoming Proposition 6.4), then provided the scale $\delta$ is taken from a sufficiently lacunary sequence $(\delta_k)$ converging to zero, the factors of $1 + O(\delta)$ may be tolerated, as upon iteration they ultimately lead to a convergent product $\prod_k(1 + O(\delta_k))$, which can be made as close to 1 as we please.

The basic inductive scheme described here should be compared with that in [17] and [19] in the setting of Loomis–Whitney-type data (data satisfying the structural condition (1.9)). In that case, suitable compactly supported extremisers are available, and an analogous nonlinear version of (3.3) is obtained. The inductive proof that we present here is necessarily different in several important respects. In particular, and unlike in most induction-on-scales arguments, we completely avoid discrete partitions of unity, and make fundamental use of certain special functions — in this instance, gaussians. This makes our approach somewhat closer to that of the heat-flow monotonicity arguments of [46], [25] and [27].

4. Exponential optimisation and a proof of Theorem 1.5

As preparation for the proof of Theorem 1.5 we begin in Section 4.1 by introducing some notation and preliminary observations, after which we state and prove the key lemmas in Section 4.2. Finally, in Section 4.3 we prove Theorem 1.5 and, in addition, establish that
the infimum of weighted sums of exponential functions is locally Hölder continuous with respect to the weights.

4.1. Preliminaries. Fix a finite index set $J$ and let $(u_j)_{j \in J}$ be a family of pairwise distinct vectors in $\mathbb{R}^n$. For each subset $I$ of $J$, we are interested in the functions $f_I$ given by

$$f_I(y, d) = \sum_{j \in I} d_j \exp(\langle u_j, y \rangle),$$

where $d = (d_j)_{j \in J}$, $d_j \geq 0$, $y \in \mathbb{R}^n$, and their infimum

$$g_I(d) = \inf_{y \in \mathbb{R}^n} f_I(y, d).$$

In studying the infimum of $f_J(\cdot, d)$, since minimisers do not always exist, it is natural to consider near-minimisers of $f_J(\cdot, d)$ and to what extent we can control their size; if

$$f_J(y_\delta, d) \leq g_J(d) + \delta \max_{j \in J} d_j$$

for $\delta > 0$, then what can be said about $y_\delta$? Minimisers are not necessarily unique and can vary on an affine subspace, which means that $y_\delta$ need not be bounded in terms of $\delta$. Nevertheless, as we will show, we can always find a $\delta$-near-minimiser that is bounded logarithmically in $\delta$.

To study $f_J$, it is very helpful to study functions $f_I$ generated by subsets $I$ of $J$, and we make the trivial observation that if $I_1 \subseteq I_2 \subseteq J$, then $f_{I_1}(y, d) \leq f_{I_2}(y, d)$ and therefore

$$g_{I_1}(d) \leq g_{I_2}(d).$$

Also, for any subset $I$ of $J$, denote by $K(I)$ the convex polytope generated by $I$ and by $K(I)^\circ$ its relative interior; this polytope plays a key role in the proof of Theorem 1.5.

If the $u_j$ for which $d_j > 0$ do not span all of $\mathbb{R}^n$, then $f_J(\cdot, d)$ is constant along the cosets of the subspace $K(I)^\perp$. We may, if we wish, restrict attention to the span of $K(I)$.

Clearly, any $j \in J$ for which $d_j = 0$ gives no contribution, so it is natural to introduce

$$J_+ = \{ j \in J : d_j > 0 \}$$

and study the function $f_{J_+}$, since clearly $f_{J_+}(y, d|_{J_+}) = f_J(y, d)$. In fact, we will first study $f_J(\cdot, d)$, where $I \subseteq J$ and $d \in (0, 1]^I$. There is the trivial case in which $I = \emptyset$, that is, $f$ is identically 0. It is easy to see that this is the only case where a minimiser exists and $g_I(d) = 0$. Excluding this pathological case, we have the following trichotomy.

**Proposition 4.1.** Let $I \neq \emptyset$. Then

- $0 \in K(I)^\circ$ if and only if a minimiser for $f_I(\cdot, d)$ exists and $g_I(d) > 0$.
- $0 \in \partial K(I)$ if and only if a minimiser for $f_I(\cdot, d)$ does not exist and $g_I(d) > 0$.
- $0 \notin K(I)$ if and only if a minimiser for $f_I(\cdot, d)$ does not exist and $g_I(d) = 0$.

We will give a proof of this proposition after proving Lemmas 4.3 and 4.4 below. Before we come to these key lemmas, we introduce three positive constants, $C_0$, $c_0$ and $c_1$, which depend on the family $(u_j)_{j \in J}$. The first of these is simply given by

$$C_0 = 1 \vee \max_{j \in J} |u_j|.$$  

The constant $c_0 \in (0, 1)$ is chosen such that

$$\text{span}(K(I)) \cap B(0, c_0) \subseteq K(I)$$

for all subsets $I$ of $J$ such that $0 \in K(I)^\circ$. We note that since $J$ is a finite set, then it is certainly possible to choose such $c_0 \in (0, 1)$ that depends only on $(u_j)_{j \in J}$.

The constant $c_1$ is given by the following lemma.

---

2For technical reasons, it is more convenient if $C_0 \geq 1$, though it is not crucial for the argument.
Lemma 4.2. There exists a constant $c_1 \in (0, 1)$ depending only on $(u_j)_{j \in J}$ with the following property: for each subset $I$ of $J$ such that $0 \notin K(I)^\circ$, there exist a subset $I_1$ of $I$ and a unit vector $v \in \mathbb{R}^n$ such that

\begin{align}
\langle v, u_j \rangle &= 0 \quad \text{if } j \in I_1, \\
\langle v, u_j \rangle &\geq c_1 \quad \text{if } j \in I \setminus I_1,
\end{align}

and either $0 \in K(I_1)^\circ$ or $0 \notin K(I)$ and $I_1 = \emptyset$.

Proof. Suppose that $0 \notin K(I)^\circ$. Then either $0$ is contained in a face of the polytope $K(I)$, or is not contained in $K(I)$ at all. In the first case, there exist a minimal subset $I_1$ of $I$ such that $0 \in K(I_1)^\circ$ and $K(I_1)$ is a face of $K(I)$. Therefore there is a unit vector $v$ that (interpreted as a functional) separates $K(I_1)$ from the rest of $K(I)$, in the sense that

\begin{align}
\langle v, y \rangle &= 0 \quad \text{if } y \in K(I_1), \\
\langle v, y \rangle &> 0 \quad \text{if } y \in K(I) \setminus K(I_1),
\end{align}

and thus (4.3) holds as long as we choose $c_1 \in (0, 1)$ such that

$$c_1 \leq \min_I \min_{y \in K(I \setminus I_1)} \langle v, y \rangle,$$

where the minimum in $I$ is taken over all subsets of $J$ such that $0$ is contained in a face of $K(I)$. Such $c_1 \in (0, 1)$ exists since $J$ is a finite set.

In the second case, when $0 \notin K(I)$, we proceed similarly and separate $0$ from $K(I)$ by a unit vector $v$, in which case (4.5) holds, provided that we choose $c_1 \in (0, 1)$ such that

$$c_1 \leq \min_I \min_{y \in K(I)} \langle v, y \rangle,$$

where the minimum in $I$ is taken over all subsets of $J$ such that $0 \notin K(I)$. \qed

4.2. Key lemmas. A major problem that will arise in the study of $\delta$-near-minimisers is when the coefficients $d_j$ are very small compared to the parameter $\delta$. Initially we shall proceed by considering strictly positive coefficients $d \in (0, 1]^J$ (with the harmless normalisation $d_j \leq 1$) and obtain bounds on certain $\delta$-near-minimisers in terms of the quantity

$$\Delta = \Delta(I, d) = \min_{j \in I} d_j.$$

We will later decompose our function $f_I$ as $f_I' + f_I''$, where $I'$ corresponds to “big” $d_j$ and $I''$ to “small” $d_j$, in a sense to be made precise later.

When $0 \in K(I)^\circ$, the following lemma says that it is possible to find a minimiser of size comparable to $\log(|I|/|\Delta|)$.

Lemma 4.3. Assume that $0 \in K(I)^\circ$, and let $d \in (0, 1]^I$. Then there exists $y \in \mathbb{R}^n$ such that

$$f_I(y, d) = g_I(d)$$

and

$$|y| \leq \frac{1}{c_0} \log \frac{|I|}{\Delta}.$$

Proof. After restriction to $\text{span}K(I)$, the function $f_I$ has a unique minimum $y$, as it is strictly convex and $\lim_{s \to +\infty} f_I(sv, d) = +\infty$ for all nonzero vectors $v \in \text{span}K(I)$.\footnote{We know much more about this case; for instance, the minimum varies real analytically in the $d_i$ and hence $g_I(d)$ is analytic.}
From the definition of $c_0$, it follows that $c_0 y/|y|$ is contained in $K(I)$, so we can write
\begin{equation}
(4.9) \quad c_0 y = |y| \sum_{j \in I} \lambda_j u_j,
\end{equation}
where $\lambda_j \in [0,1]$ and $\sum_{j \in I} \lambda_j = 1$. Then
\begin{equation}
(4.10) \quad c_0 |y| = \sum_{j \in I} \lambda_j \langle u_j, y \rangle,
\end{equation}
so there exists at least one $j$ (depending on $y$) such that $\langle u_j, y \rangle \geq c_0 |y|$. We conclude that
\begin{equation}
(4.11) \quad \Delta \exp(c_0 |y|) \leq d_j \exp(\langle u_j, y \rangle) \leq f_I(y,d) = g_I(d) \leq f_I(0, d) \leq |I|,
\end{equation}
which immediately gives the bound [4.8].

\textbf{Remark.} A simple inequality that will be useful below is
\begin{equation}
(4.12) \quad f_I(y,d) \leq |J| D \exp(C_0 R)
\end{equation}
whenever $I \subseteq J$, $|y| \leq R$ and $d \in [0, D]^J$.

We now consider the general situation, in which case minimisers may not exist. However, in this case, the following lemma tells us that we can find a $\delta$-near-minimiser of size comparable to $\log(1/(\delta \Delta))$.

\textbf{Lemma 4.4.} There exist $C_1 > 0$ and $\delta_1 > 0$ with the following property: for all nonempty subsets $I$ of $J$, $\delta \in (0, \delta_1)$ and $d \in [0, 1]^J$, there exists $y \in \mathbb{R}^n$ such that
\begin{equation}
(4.13) \quad |y| \leq C_1 \log \frac{1}{\delta \Delta}.
\end{equation}

\textbf{Proof.} If $0 \in K(I)^\circ$, then we may find a minimiser by Lemma [4.3] with the desired bound [4.13], so it suffices to consider the case in which $0 \notin K(I)^\circ$.

If $0 \notin K(I)^\circ$, then we use Lemma [4.2] to find a subset $I_1$ of $I$ and a unit vector $v \in \mathbb{R}^n$ such that [4.4] and [4.5] hold. Let $y_0$ be a minimiser of $f_{I_1}$ chosen according to Lemma [4.4] (if $I_1 = \emptyset$, then $y_0 = 0$ will trivially fulfil [4.1] and [4.2]). We will modify $y_0$ by moving along $v$, which keeps $f_{I_1}$ stable and shrinks the remainder term $f_{I_1} - f_{I_1 \setminus I_1} = f_{I_1 \setminus I_1}$. Thus we set $y = y_0 - sv$, where $s$ will be chosen later, and use [4.3] and [4.5] to estimate
\begin{align*}
f_I(y,d) &= \sum_{j \in I_1} d_j \exp(\langle u_j, y_0 \rangle) + \sum_{j \in I \setminus I_1} d_j \exp(\langle u_j, y_0 \rangle - s \langle u_j, v \rangle) \\
&\leq f_{I_1}(y_0, d) + f_{I_1 \setminus I_1}(y_0, d) \exp(-sc_1) \\
&\leq g_{I_1}(d) + |J| \left( \frac{|J|}{\Delta} \right) \frac{C_0}{c_0} \exp(-sc_1).
\end{align*}

For the final inequality we used the trivial estimate [4.12]. Therefore, from the monotonicity property [4.3] and the choice
\begin{equation}
(4.14) \quad s = \frac{1}{c_1} \left( \log \frac{1}{\delta} + \log |J| + \frac{C_0}{c_0} \log \frac{|J|}{\Delta} \right),
\end{equation}
it follows that
\begin{equation}
(4.15) \quad f_I(y,d) \leq g_{I_1}(d) + \delta \leq g_I(d) + \delta.
\end{equation}

We note the above choice of $s$ is strictly positive.

To complete the proof, we choose
\begin{equation}
(4.16) \quad \delta_1 = \frac{1}{|J|} \quad \text{and} \quad C_1 = \frac{4C_0}{c_0 c_1}.
\end{equation}
and may easily verify that
\[ s \leq \frac{3C_0}{c_0c_1} \log \frac{1}{\delta \Delta}, \]
whence (4.13) holds. \[\square\]

Before proving Theorem 1.5, we establish Proposition 4.4.

**Proof of Proposition 4.4.** First of all, if \( 0 \in K(I)^c \), then the existence of a minimiser follows from Lemma 4.3, and it is easy to see that then \( g_I(d) \neq 0 \).

Now suppose that \( 0 \notin K(I) \). Without loss of generality, we may assume that \( K(I)^c \neq \emptyset \). By (4.13),
\[
    g_I(d) = \inf_{y \in \mathbb{R}^n} f_I(y, d) = g_I_1(d),
\]
where either \( 0 \notin K(I_1)^c \) or \( f_I_1 = 0 \). Now, from the proof of Lemma 2.2, if \( 0 \notin K(I) \) then \( 0 \notin K(I_1)^c \) (and \( I_1 \neq \emptyset \)). This means that the infimum of \( g_I_1 \) is attained and
\[
    g_I(d) = g_I_1(d) = \min_{y \in \mathbb{R}^n} f_I_1(y, d) > 0.
\]
On the other hand, if \( 0 \notin K(I) \), then \( I_1 = \emptyset \) and hence \( g_I(d) = g_I_1(d) = 0 \).

It remains to show that extremisability implies that \( 0 \in K(I)^c \). To this end, assume that a minimiser for \( f_I(\cdot, d) \) exists and \( 0 \notin K(I)^c \). If the infimum of \( f_I(\cdot, d) \) is attained at \( y_0 \), then
\[
    f_{I \setminus I_1}(y_0, d) = f_I(y_0, d) - f_{I_1}(y_0, d) \leq g_I(d) - g_{I_1}(d) = 0,
\]
which means \( I = I_1 \), and thus contradicts properties of \( I_1 \). We deduce that \( 0 \in K(I)^c \) is necessary for extremisability, and this concludes the proof of Proposition 4.4. \[\square\]

**4.3. Proof of Theorem 1.5 and local Hölder continuity.**

**Proof of Theorem 1.5.** Without loss of generality, we may assume that \( d_j \leq 1 \) for all \( j \in J \). The proof rests on a partition, depending on \( \delta \), of the coefficients \( d_j \) into two parts, one for “big” coefficients and one for “small” coefficients. However, it will be important that there is a quantitative “gap” at the boundary of this partition and therefore our argument is more subtle than simply choosing an appropriate power of \( \delta \) as the threshold for the partition.

Let \( N_1 \geq C_1 \) be a parameter that we will specify later, where \( C_1 \) is the constant arising from Lemma 4.3. For nonnegative integers \( k \), we define the disjoint subsets \( I_k \) of \( J \) by
\[
    I_k = \{ j \in J : \delta^{N_k+1} \leq d_j < \delta^{N_k} \}.
\]
Then, by the pigeonhole principle, there exists \( k = k_\delta \in \{1, \ldots, |J|\} \) such that \( I_k \) is empty. For this \( k \), we introduce
\[
    I' = \{ j \in J : d_j \geq \delta^{N_k} \}
\]
and observe that, since \( I_k \) is empty,
\[
    J \setminus I' = \{ j \in J : d_j \leq \delta^{N_k+1} \}.
\]

Now consider the function \( f_{I'} \), where
\[
    \Delta = \Delta(d, I') = \min_{I \in I'} d_j \geq \delta^{N_k}.
\]
According to Lemma 4.4, there exists \( y \in \mathbb{R}^n \) such that
\[
    f_{I'}(y, d) \leq g_{I'}(d) + \delta^2
\]
and
\[
    |y| \leq C_1 \log \frac{1}{\delta^2 \Delta}.
\]
Therefore, we have the upper bound
\[
|y| \leq C_1(N_1^{|J|} + 2) \log \frac{1}{\delta},
\]
and we will show that by choosing \(N_1\) and \(\delta_0\) appropriately, \(y\) is a \(\delta\)-near-minimiser of \(f_J\).

To see this, we use the trivial estimate (4.12) to obtain
\[
f_{J \setminus P}(y) \leq |J|^N \delta^{N+1} - C_0 C_1 (N_1^2 + 2)
\]
and thus, if we choose \(N_1 = 4C_0 C_1\), then
\[
f_{J \setminus P}(y) \leq |J| \delta^2.
\]
This immediately gives
\[
f_J(y, d) \leq g_I(d) + \delta^2 + |J| \delta^2 \leq g_J(d) + \delta,
\]
provided that
\[
(4.18) \quad \delta_0 \leq (1 + |J|)^{-1}.
\]
Finally, we note that if we choose
\[
(4.19) \quad N = 3(C_0 C_1)^{|J|},
\]
then \(|y| \leq N \log(1/\delta)\) from our choice of \(N_1\) and (4.17).

**Remark.** A review of the proofs of Lemma 4.4 and Theorem 1.5 reveals that an explicit choice of \(N\) and \(\delta_0\) for which the claim in Theorem 1.5 holds is given by
\[
(4.20) \quad N = 3 \left( \frac{4C_0^2}{c_0 c_1} \right)^{|J|} \quad \text{and} \quad \delta_0 = \frac{1}{|J| + 1},
\]
where the constants \(C_0\), \(c_0\) and \(c_1\) are those defined at the start of Section 4.1.

As an immediate application of Theorem 1.5, we show that the infimum \(g_I\) is a locally Hölder-continuous function.

**Theorem 4.5** (Local Hölder continuity). There exists a number \(\alpha \in (0, 1)\), depending only on \((u_j)_{j \in J}\), such that for any \(D > 0\) and \(I \subseteq J\),
\[
(4.21) \quad |g_I(d) - g_I(d')| \leq (D + |J|) \sup_j |d_j - d'_j|^\alpha
\]
for all \(d, d' \in [0, D]^I\).

**Proof.** We shall show the local Hölder continuity claimed holds when \(\alpha = (1 + C_0 N)^{-1}\), where \(N\) is given by Theorem 1.5 and for this it suffices to show
\[
(4.22) \quad |g_I(d) - g_I(d')| \leq (D + |J|) \|d - d'\|_\infty^\alpha
\]
whenever \(\|d - d'\|_\infty < \delta_0^{1+C_0 N}\). Here, \(\delta_0\) is given by Theorem 1.5. Indeed, a simple rescaling argument and equivalence of norms on finite-dimensional vector spaces will yield (4.21) for all \(d, d' \in [0, D]^I\) from (4.22) (with a different constant).

To prove (4.22), suppose that \(0 < \|d - d'\|_\infty < \delta_0^{1+C_0 N}\) and define \(\delta \in (0, \delta_0)\) by
\[
\delta = \|d - d'\|_\infty^{\frac{1}{1+C_0 N}}.
\]
By Theorem 1.5 there exists \(y \in \mathbb{R}^n\) such that \(|y| \leq N \log(1/\delta)\) and
\[
f_I(y, d) \leq g_I(d) + D\delta.
\]
Using the elementary bound
\[
|f_I(y, d) - f_I(y, d')| \leq |J| \|d - d'\|_\infty c_0 |y|
\]
and our choice of \( \delta \), we thus see that
\[
g(I(d')) - g(I(d)) \leq f(I(y, d')) - f(I(y, d)) + D\delta
\leq |J||d - d'||\infty \delta^{-\phi}\n + D\delta
= (D + |J|)||d - d'||\infty(1 + \phi)\n.
\]

By symmetry, (5.2) follows. \( \square \)

5. Proof of Theorem 1.3 and local Hölder continuity of the Brascamp–Lieb constant

In this section, we shall see how to prove Theorem 1.3, our effective form of Lieb’s theorem, as an application of our abstract near-optimisation result in Theorem 1.5. We shall also establish the local Hölder continuity of the Brascamp–Lieb constant as an application of Theorem 1.5. Naturally, both of these deductions are based on having a suitable expression for \( BL_g(L, p, A) \) in terms of sums of exponential functions. The expression we use is contained in Proposition 5.1 below and is based on ideas from the proof of [20. Theorem 3.1]. To state the result, we need to set up some more notation.

First, fix an admissible Brascamp–Lieb datum \((L, p, A)\). Next, given an \( n \times n \) positive definite matrix \( A_j \), we write
\[
A_j = R_j^* X_j R_j,
\]
where \( R_j \) is an \( n \times n \) rotation matrix and \( X_j \) is a diagonal matrix with positive diagonal entries. It is convenient to list the diagonal entries of \( X_1, \ldots, X_m \) as \( e^{y_1}, \ldots, e^{y_m} \), where each \( y_j \) is a real number and \( M = \sum_{j=1}^m n_j \).

The index set \( I \) is given by
\[
I = \{ I \subseteq \{1, \ldots, M\} : |I| = n \}
\]
and, for each \( I \in I \), the vector \( u_I \in \mathbb{R}^M \) is given by
\[
u_I = 1_I - q.
\]

Here, \( 1_I \) denotes the vector \((1_I(j))_{j=1}^M \in \mathbb{R}^M \), where \( 1_I(j) = 1 \) if \( j \in I \), and \( 1_I(j) = 0 \) if \( j \notin I \). The vector \( q \in \mathbb{R}^M \) takes the form
\[
q = (p_1, \ldots, p_1, p_2, \ldots, p_2, \ldots, p_m, \ldots, p_m),
\]
where the first \( n_1 \) components coincide with \( p_1 \), the next \( n_2 \) components coincide with \( p_2 \), and so on. The vector \( y \in \mathbb{R}^M \) is simply given by \( y = (y_j)_{j=1}^M \), where \( y_1, \ldots, y_M \) were introduced above. Finally, for each \( I \in I \), we define
\[
d_I(L, R) = \det((v_k)_{k \in I})^2
\]
where \( v_k = L_j^* R_j^* e_{(j, \ell)} \) and \( R = (R_j)_{j=1}^m \). Here \( \{e_{(j, \ell)} : \ell = 1, \ldots, n_j\} \) is the standard basis for \( \mathbb{R}^{n_j} \) and we are identifying \( k \in \{1, \ldots, M\} \) with the pair \((j, \ell)\) via \( k = n_0 + \ldots + n_{j-1} + \ell \), where \( j \in \{1, \ldots, m\} \), \( \ell \in \{1, \ldots, n_j\} \) and \( n_0 = 0 \).

**Proposition 5.1.** With the notation as above,
\[
BL_g(L, p, A)^{-2} = \sum_{I \in I} d_I(L, R) \exp(\langle u_I, y \rangle).
\]

**Proof.** To see (5.2), we follow the proof of [21. Theorem 3.1] and write
\[
\sum_{j=1}^m p_j L_j^* A_j L_j = \sum_{k=1}^M q_k e^{y_k} v_k v_k^*.
\]
The expression on the right-hand side coincides with the product of the matrix whose $k$th column is $g_k e^{y_k v_k}$ and the matrix whose $k$th row is $v_k^*$, and thus
\[
\det \left( \sum_{j=1}^{m} p_j L_j^* A_j L_j \right) = \sum_{I \subseteq \mathcal{I}} d_I(L, R) \exp(\langle 1, y \rangle)
\]
follows from the Cauchy–Binet formula. The identity (5.2) is now immediate. \qed

Before beginning the proof of Theorem 1.3, we recall that Theorem 1.5 dealt with minimisers and near minimisers for functions of the form
\[
f_{\mathcal{I}}(y, d) = \sum_{I \subseteq \mathcal{I}} d_I \exp(\langle u_I, y \rangle)
\]
where $y \in \mathbb{R}^M$ and $d = (d_I)_{I \subseteq \mathcal{I}}$ is a family of positive coefficients, as well as
\[
g_{\mathcal{I}}(d) = \inf_{y \in \mathbb{R}^M} f_{\mathcal{I}}(y, d).
\]

**Proof of Theorem 1.3.** As a first step to proving our effective version of Lieb’s theorem, we observe that it suffices to consider the case where each $L_j$ is a projection; that is, $L_j L_j^*$ coincides with the identity transformation $I_{n_j}$ on $\mathbb{R}^{n_j}$ for each $j = 1, \ldots, m$. To see this, assume that (1.16) holds when the $L_j$ are projections. Then, given any admissible $L$, we define
\[
L_j' = C_j^{-1} L_j
\]
with $C_j = (L_j L_j^*)^{1/2}$. Then it is easy to check that $L'$ is admissible and each $L_j'$ is a projection, and thus (1.16) holds for $L'$. It may be deduced that (1.16) holds for $L$ using the elementary fact
\[
\text{BL}_g(L, p; A)^{-2} = \text{BL}_g(L', p; A')^{-2} \prod_{j=1}^{m} \det(L_j L_j^*)^{p_j},
\]
whenever $A$ and $A'$ satisfy $A_j = (C_j^{-1})^* A_j' C_j^{-1}$.

We now prove (1.16) under the assumption that $L_j L_j^* = I_{n_j}$. From Proposition 5.1
\[
\text{BL}_g(L, p; A)^{-2} = f_{\mathcal{I}}(y, d(L, R)),
\]
where $d(L, R) = (d_I(L, R))_{I \subseteq \mathcal{I}}$ and therefore, from Lieb’s theorem (Theorem 1.2),
\[
\text{BL}(L, p)^{-2} = \inf_{R \in \mathbb{R}^M} \inf_{y \in \mathbb{R}^M} f_{\mathcal{I}}(y, d(L, R)) = \inf_{R \in \mathbb{R}^M} g_{\mathcal{I}}(d(L, R)).
\]
Since $d(L, R)$ is a continuous function of $R$ and $g_{\mathcal{I}}(d)$ is a continuous function of $d$ by Theorem 1.5 (this is tantamount to the continuity of the Brascamp–Lieb constant observed in [20]), there exists some $R$ (depending on $L$) such that
\[
\text{BL}(L, p)^{-2} = g_{\mathcal{I}}(d(L, R)).
\]
For the remainder of the proof, $R = (R_j)_{j=1}^{m}$ denotes this minimising collection of rotations.

To complete the proof of Theorem 1.3 we apply Theorem 1.5. Thus, there exist $N \in \mathbb{N}$ and $\delta_0 > 0$, depending only on $p$, such that, for all $\delta \in (0, \delta_0),$
\[
\inf_{|y| \leq N \log(1/\delta)} f_{\mathcal{I}}(y, d(L, R)) \leq g_{\mathcal{I}}(d(L, R)) + \delta \max_I d_I(L, R).
\]
From the assumption $L_j L_j^* = I_{n_j}$, it follows that $d_I(L, R) \leq 1$ and therefore there exists some $y = y(\delta, L) \in \mathbb{R}^M$ satisfying
\[
|y| \leq N \log \frac{1}{\delta}
\]
and
\[
f_{\mathcal{I}}(y, d(L, R)) \leq g_{\mathcal{I}}(d(L, R)) + \delta \max_I d_I(L, R).
\]
and
\[ f_I(y, d(L, R)) \leq g_I(d(L, R)) + \delta. \]
Hence
\[ (5.6) \quad BL(L, p; A)^{-2} = f_I(y, d(L, R)) \leq g_I(d(L, R)) + \delta = BL(L, p)^{-2} + \delta. \]
Here, \( A_j = R_j^*X_jR_j \), where \( X_1 \) is the diagonal matrix with entries \( e^{y_1}, \ldots, e^{y_{n_1}} \), \( X_2 \) is the diagonal matrix with entries \( e^{y_{n_1+1}}, \ldots, e^{y_{n_2+n}} \), and so forth. From (5.6), it is easily checked that \( \|A_j\| \leq (1/\delta)^N \) and \( \|A_j^{-1}\| \leq (1/\delta)^N \). Hence (1.16) holds when \( L_jL_j^* = I_n_j \), and this completes our proof of Theorem 1.3. \( \square \)

To end this section, we improve the result in [20], which established the continuity of the Brascamp–Lieb constant, to local Hölder continuity. Given that differentiability of the Brascamp–Lieb constant may fail (see [20]), it seems to be a challenging problem to improve upon the following result in such generality.

**Theorem 5.2.** For each \( p \), the Brascamp–Lieb constant \( BL(\cdot, p) \) is locally Hölder continuous. More precisely, there exists a number \( \alpha \in (0, 1) \) and a constant \( C_0 > 0 \), depending only on \((n_j)_{j=1}^{m_a} \) and \((p)_{j=1}^{m_i} \), such that the following holds. For any \( C > 0 \),
\[ (5.7) \quad |BL(L, p)^{-2} - BL(L', p)^{-2}| \leq C_0 C^{n+\alpha(n-1)} \|L - L'\|^\alpha \]
for all data \( L, L' \) satisfying the bound \( \|L\|, \|L'\| \leq C \). In particular, if additionally the Brascamp–Lieb constants \( BL(L, p), BL(L', p) \) are bounded above by a constant \( C_1 \), we have
\[ (5.8) \quad |BL(L, p) - BL(L', p)| \leq C_0 C^3 C^{n+\alpha(n-1)} \|L - L'\|^\alpha. \]

We remark that Theorem 5.2 establishes global Hölder continuity of the function \( BL(\cdot, p)^{-2} \) when restricted to projection data \( L \), for example.

**Proof.** First, we note that it is straightforward to verify that (5.7) implies (5.8).

Next, we observe that each \( d_I(L, R) \), defined in (5.1), is locally Lipschitz in \( L \) uniformly in \( R \) with Lipschitz constant \( c_n C^{n-1} \), where \( c_n \) depends only on the dimension \( n \). Also, \( g_I \) is locally Hölder continuous by Theorem 4.5 and thus it obviously follows that \( g_I(d(\cdot, R)) \) is locally Hölder continuous uniformly in \( R \) with constant \( C_0 C^{n+\alpha(n-1)} \). By elementary considerations, it follows that \( \inf_R g_I(d(\cdot, R)) \) is locally Hölder continuous too, with the same bound. However, by (5.3), it follows that \( BL(\cdot, p)^{-2} = \inf_R g_I d(\cdot, R) \). \( \square \)

The next simple example may cast a little light on our previous results, and in particular show that sharper results, such as Lipschitz continuity, cannot hold.

**Example 5.3.** Define \( f : \mathbb{R}^4 \to \mathbb{R}^+ \) and \( g(a, b) \) by
\[ f(x, y, a, b) = ae^x + e^{-x} + e^{x-y} + be^y \quad \text{and} \quad g(a, b) = \inf \{ f(x, y, a, b) : (x, y) \in \mathbb{R}^2 \}, \]
where \( a, b \in [0, 1] \). We will find when the infimum is attained and how it depends on \( a \) and \( b \), as well as examine approximate minimisers where we impose the constraint \( |x| \leq R \) and \( |y| \leq R \).

There are various cases to consider.

**Case 1:** \( a = b = 0 \). Evidently, \( f(x, y, 0, 0) = e^{-x} + e^{x-y} \), and \( g(0, 0) = 0 \). For \( \delta \in (0, 1) \), we take \( (x_\delta, y_\delta) = (\log(2/\delta), 2 \log(2/\delta)) \); then
\[ f(x_\delta, y_\delta, 0, 0) = \delta \leq g(0, 0) + \delta \]
and \( |x_\delta| \leq 2 \log(2/\delta) \) and \( |y_\delta| \leq 2 \log(2/\delta) \).
Moreover, in both cases, \( g(a, 0) = \min\{ae^x + e^{-x} : x \in \mathbb{R}\} = 2a^{1/2} \).

Observe that
\[
g(a, 0) - g(0, 0) = 2a^{1/2},
\]
so \( g(\cdot, 0) \) is Hölder continuous of exponent 1/2, but no more. Given \( \delta \in (0, 1) \), we take
\[
x_\delta = -\frac{1}{2}\log(a + \delta^2/4) \quad \text{and} \quad y_\delta = -2\log(\delta/2).
\]
Then
\[
f(x_\delta, y_\delta) = \frac{2a + \delta^2/2}{(a + \delta^2/4)^{1/2}} \leq g(a, 0) + \delta
\]
and \( |x_\delta| \leq \log(2/\delta) \) and \( |y_\delta| \leq 2\log(2/\delta) \).

Case 3: \( a = 0 \) and \( b \neq 0 \). In this case,
\[
f(x, y) = e^{-x} + e^{x-y} + be^y.
\]
The function \( f(x, y) \) attains its minimum at \((\frac{1}{3}\log(1/b), \frac{2}{3}\log(1/b))\), and
\[
g(0, b) = 3b^{1/3}.
\]
Again \( g(0, \cdot) \) is Hölder continuous of exponent 1/3, but no more.

Given \( \delta \in (0, 1) \), define
\[
(x_\delta, y_\delta) = \begin{cases} \left(\frac{1}{3}\log(1/b), \frac{2}{3}\log(1/b)\right) & \text{if } b > (\delta/3)^3 \\ (\log(3/\delta), 2\log(3/\delta)) & \text{if } b \leq (\delta/3)^3. \end{cases}
\]
If \( b > (\delta/3)^3 \), then \( f(-\frac{1}{3}\log(b), -\frac{2}{3}\log(b), 0, b) - g(0, b) = 0 \), and otherwise
\[
f(x_\delta, y_\delta, 0, b) - g(0, b) = 2\delta/3 + b(3/\delta)^2 - 3b^{1/3} \leq (3b^{1/3} < \delta.
\]
Moreover, in both cases, \( |x_\delta| \leq \log(3/\delta) \) and \( |y_\delta| \leq 2\log(3/\delta) \).

Case 4: \( a > 0 \) and \( b > 0 \). This case is more complex and we cannot find formulae that are as precise as those in the previous cases. Nevertheless, we can show that \( f(x, y, a, b) \) achieves its minimum at \((x_\tau, y_\tau)\), where
\[
e^{y_\tau} = b^{-2/3}r(ab^{-2/3})^{-1}
\]
\[
e^{x_\tau} = b^{-1/3}r(ab^{-2/3})^{-2};
\]
here \( r(c) \) is the unique positive solution of the equation \( t^4 - t = c \). With somewhat more effort, it is possible to show that if \( \delta \in (0, 1/e) \), then there exists \((x_\delta, y_\delta)\) such that
\[
f(x_\delta, y_\delta, a, b) - g(a, b) \leq \delta
\]
and \( |x_\delta| \leq 4\log(1/\delta) \) and \( |y_\delta| \leq 4\log(1/\delta) \).

6. The Nonlinear Brascamp–Lieb Inequality

Before we embark on the proof of Theorem 1.1 we offer further clarifying remarks.

First, it is not possible to take \( \varepsilon < 0 \) in the statement of Theorem 1.1 for any datum \((L, p)\), as may be seen by a slight modification of the argument in Lemma 6 from [22]; see also [52]. This argument is related to the so-called “transplantation” method — see for example [72] and [80]. As may be expected, it is not in general possible to take \( \varepsilon = 0 \) either, as BL(dB(x), p) varies with \( x \) in all but very special situations.

Theorem 1.1 is stated for \( C^2 \) maps merely for simplicity. A careful inspection of our proof will reveal that a \( C^{1+\theta} \) condition suffices for any \( \theta > 0 \).

As our proof also reveals, and as may be expected, the neighbourhood \( U \) appearing in the statement of Theorem 1.1 depends only on the underlying linear datum \((L, p)\).
and regularity bounds on the maps $B_j$ (be they $C^2$ or $C^{1+\theta}$ bounds). From the point of view of potential applications, it is also appropriate to point out that, for fixed $p$, the neighbourhood $U$ may be taken to be locally uniform in $L$ — that is, given $L_0$ and $A > 0$, there exists a neighbourhood $U$ of $x_0$ such that (1.5) holds for all $B_j$ such that $\|B_j\|_{C^2(U)} \leq A$ (or $\|B_j\|_{C^{1+\theta}} \leq A$) and $dB_j(x_0) = L_j$, provided $L$ is sufficiently close to $L_0$. Again, this follows from a close inspection of our arguments.

### 6.1. Setting up the induction.

By translation invariance, we may assume that $x_0 = 0$ in the statement of Theorem 1.1. As discussed in Section 3, our proof of Theorem 1.1 will proceed by induction on the size of the supports of the input functions $f$. In order for there to be a base case for this induction it will be convenient to suppose that $f$ satisfies a certain auxiliary regularity condition, upon which our estimates will be essentially independent. The passage to general integrable $f$ will consist of an elementary limiting argument.

**Definition 6.1.** Suppose that $\kappa > 1, \mu > 0$ and $\Omega$ is a measurable subset of $\mathbb{R}^d$. A nonnegative function $f$ is $\kappa$-constant at scale $\mu$ on $\Omega$ if $f(x) \leq \kappa f(y)$ whenever $x \in \Omega$ and $y \in \mathbb{R}^d$ are such that $|x - y| \leq \mu$. We denote by $L^1(\Omega; \mu, \kappa)$ the subset of $L^1(\mathbb{R}^d)$ with this property.

The reader will observe that the definition above lacks symmetry in $x$ and $y$: we stipulate that $x \in \Omega$, while $y$ may extend a distance $\mu$ from $\Omega$. The reason for this technicality will become clear shortly. At this stage it is worth noting that for any fixed $\kappa > 1$, a nonnegative function $f \in L^1(\Omega)$ may be approximated almost everywhere by functions in $L^1(\Omega; \mu, \kappa)$, provided that $\mu$ is taken sufficiently small. One way to see this is to observe that the $d$-dimensional Poisson kernel $P_t(x)$ is $\kappa$-constant at scale $\mu$ on the whole of $\mathbb{R}^d$ provided that $\mu$ is small enough depending on $t > 0$ and $\kappa > 1$. By linearity, the convolution $f * P_t$ is easily seen to inherit this regularity property for any nonnegative $f \in L^1(\mathbb{R}^d)$, and so the almost everywhere approximation claimed follows from the Lebesgue differentiation theorem.

We will allow $\kappa$ to vary in a controlled manner through the induction. This flexibility in $\kappa$ is required when considering products of such “locally constant” functions. In particular, we will need to appeal to the elementary fact that if $f \in L^1(\Omega; \mu, \kappa)$ and $g \in L^1(\Omega; \mu, \lambda)$, then $fg \in L^1(\Omega; \mu, \kappa \lambda)$. Considerations of this type naturally arise when introducing partitions of unity, as we shall to pass between scales.

We now set up the induction. For each $\delta \in (0, 1)$ and $y \in \mathbb{R}^n$, let

$$U_\delta(y) = \{x \in \mathbb{R}^n : |x - y| \leq \delta\}.$$  

**Definition 6.2.** For $u \in \mathbb{R}^n$, $\delta > 0$, $\mu > 0$ and $\kappa > 1$, let $C(u, \delta, \mu, \kappa)$ denote the best constant $C$ in the inequality

$$\int_{U_\delta(u)} \prod_{j=1}^m f_j^{p_j}(B_j(y)) \, dy \leq C \prod_{j=1}^m \left( \int_{\mathbb{R}^n} f_j(x_j) \, dx_j \right)^{p_j}$$

over all inputs $f_j \in L^1(B_j(U_{2\delta}(u)); \mu, \kappa)$.

We think of $C(u, \delta, \mu, \kappa)$ as a regularised and localised nonlinear Brascamp–Lieb constant; here we are of course suppressing the dependence on $B$ and $p$ in our notation.

The requirement that $f_j \in L^1(B_j(U_{2\delta}(u)); \mu, \kappa)$ may seem unusual as the integral on the left-hand side does not see the part of $f_j$ supported outside of $B_j(U_\delta(u))$, whereas the right-hand side does. This merely technical feature will be important for closing the induction.

If $\delta$ is below a certain threshold in terms of $\mu$, then each $f_j \in L^1(B_j(U_{2\delta}(u)); \mu, \kappa)$ will (effectively) cease to distinguish between $B_j$ and its best affine approximation $L^n_j$ at $u$, given by

$$L^n_j := B_j(u) + dB_j(u)(\cdot - u).$$
This allows us to reduce the estimation of $C(u, \delta, \mu, \kappa)$ to an application of the linear Brascamp–Lieb inequality, and provides us with an effective “base case” for the inductive argument. This base case is contained in the first of our two propositions below, and the second contains the recursive inequality for the main step of the induction.

The statements of the two subsequent propositions and their proofs contain certain indices which we now introduce. The indices are $\alpha, \beta, \gamma, \tau$ satisfying $0 < \gamma < \beta < \tau < \alpha$ and the relation $\sigma < \beta < C$. It allows us to reduce the estimation of $C$ to $C^\alpha \delta^{\alpha + \beta}$, which depends only on $\delta, \mu, \kappa$, and provides us with an effective “base case” for the inductive argument. This base case is contained in the first of our two propositions below, and the second contains the recursive inequality for the main step of the induction.

The parameters $\delta, \mu, \kappa$ are of a more technical nature, so we postpone further remarks of this type.

Before stating the base case for the induction, we note that by the $C^2$ regularity hypothesis on the nonlinear maps $B_j$, we have

$$|B_j(x) - L_j^y x| \lesssim |x - u|^2$$

for all $x, u \in U_\nu(0)$ (with an appropriately small constant $\nu$). We make frequent use of this fact.

**Proposition 6.3** (Base case). *There exists a constant $\nu > 0$ with the following property: if $u \in U_\nu(0)$, $\kappa > 1$, and if $\delta \in (0, \nu)$ and $\mu > 0$ satisfy $\delta^{\alpha + \beta} \leq \mu$, then

$$C(u, \delta, \mu, \kappa) \leq \kappa^\rho \textrm{BL}(dB(u), p),$$

where $\rho = \sum_{j=1}^m p_j$.***

*Proof.* Let $\nu > 0$ be a small constant, to be chosen below. We fix $u \in U_\nu(0)$, and assume that $\delta \in (0, \nu)$ and $\mu \geq \delta^{\alpha + \beta}$. We also fix $f_j \in L^1(B_j(U_{2\delta}(u)); \mu, \kappa)$.

By choosing the constant $\nu > 0$ sufficiently small and applying (6.4),

$$|B_j(y) - L_j^y y| \lesssim \delta^2$$

for all $y \in U_\delta(u)$. Together with our choices that $\alpha + \beta < 2$ and $\delta^{\alpha + \beta} \leq \mu$, this implies that $|B_j(y) - L_j^y y| \leq \mu$, for a possibly smaller choice of the constant $\nu$. Now $f_j \in$
and so $f_j(B_j(y)) \leq \kappa f_j(L_j^n y)$ for all $y \in U_\delta(u)$, and so

$$
\int_{U_\delta(u)} \prod_{j=1}^m f_j^\nu(B_j(y)) \, dy \leq \kappa^\rho \int_{U_\delta(u)} \prod_{j=1}^m f_j^\nu(L_j^n y) \, dy
\leq \kappa^\rho \text{BL}(d\mathbf{B}(u), p) \prod_{j=1}^m \left( \int_{\mathbb{R}^n_j} f_j(x_j) \, dx_j \right)^{p_j},
$$

as required. Here we have used the translation-invariance of the linear Brascamp–Lieb inequality in the $f_j$ to remove the translations in the affine approximation above. \qed

Our objective is to prove a suitable recursive inequality involving the function $C$, which upon iteration will establish Theorem 1.1. This inequality, which we now state, may be viewed as a certain nonlinear version of (6.3).

**Proposition 6.4** (Recursive inequality). There exists a constant $\nu > 0$ with the following property: if $u \in U_\nu(0)$, $\kappa > 1$, and if $\delta \in (0, \nu)$ and $\mu > 0$ satisfy $\delta^{\alpha + \beta} > \mu$, then

$$
C(u, \delta, \mu, \kappa) \leq \left(1 + \delta^\alpha \right) \max_{x \in (L_\delta(u)} C(x, \delta^{\alpha}, \mu, \kappa \exp(\delta^\beta)).
$$

The reader may wish to view (6.5) as a certain “near-monotonicity property” of the function $C$ in the parameter $\delta$, since the multiplicative factors $(1 + \delta^\alpha)$ and $\exp(\delta^\beta)$ are close to 1 for small $\delta$. It is interesting to compare (6.5) with the related near-monotonicity results in [27], [17], [19] and [88].

We conclude this section by showing how Theorem 1.1 follows from Propositions 6.4 and 6.3.

**Proof of Theorem 1.1**. Fix $\varepsilon > 0$ and set $\kappa = 1 + \varepsilon$. By an elementary density argument, which we now sketch, it will be enough to show that there exist a constant $\delta_0$ depending on $\varepsilon$ and a constant $C > 0$ such that

$$
C(0, \delta_0, \mu, 1 + \varepsilon) \leq \left(1 + C\varepsilon \right) \text{BL}(d\mathbf{B}(0), p)
$$

uniformly in $\mu > 0$. As we observed above, for each $t > 0$, and $f_j \in L^1(B_j(U_{\delta_0}(0)))$, the Poisson extension satisfies $f_j * P_t^{(n_j)} \in L^1(\mathbb{R}^{n_j}; \mu, \kappa)$, and so after restriction $f_j * P_t^{(n_j)} \in L^1(2B_j(U_{\delta_0}(0)); \mu, \kappa)$, provided that $\mu$ is chosen to be sufficiently small. Hence (6.6), combined with the $L^1$ normalisation of the $n_j$-dimensional Poisson kernel $P_t^{(n_j)}$, imply that

$$
\int_{U_{\delta_0}(0)} \prod_{j=1}^m \left( f_j * P_t^{(n_j)}(B_j(x)) \right)^{p_j} \, dx \leq \left(1 + C\varepsilon \right) \text{BL}(d\mathbf{B}(0), p) \prod_{j=1}^m \left( \int_{\mathbb{R}^{n_j}} f_j(x_j) \, dx_j \right)^{p_j}
$$

uniformly in $t > 0$. Since the nonnegative functions $f_j * P_t^{(n_j)}$ converge to $f_j$ almost everywhere as $t \to 0$, Theorem 1.1 follows by Fatou’s lemma. We now turn to the proof of (6.6).

We choose a sufficiently small constant $\delta_0$, depending on $\varepsilon$, to satisfy a number of constraints. The first of these is $\delta_0 \leq \nu/3$, where $\nu$ is sufficiently small that both Propositions 6.3 and 6.4 hold (that this is possible follows from the fact that both propositions implicitly require that $\nu$ is sufficiently small).

Given $\mu > 0$, either the base case provided by Proposition 6.3 holds directly, that is, $\delta_0^{\alpha + \beta} \leq \mu$, or $\mu$ satisfies the threshold condition that $\delta_0^{\alpha + \beta} > \mu$. In the former case there is nothing more to prove. In the latter, we shall use the recursive inequality (6.5) of Proposition 6.4 to connect $C(0, \delta_0, \mu, 1 + \varepsilon)$ with the base case.

We define $\delta_k = \delta_0^k$, where $k \in \mathbb{N}$, and apply Proposition 6.4 to see that

$$
C(u, \delta_k, \mu, \kappa) \leq \left(1 + \delta_k^\alpha \right) \max_{x \in U_{2\delta_k}(u)} C(x, \delta_{k+1}, \mu, \kappa \exp(\delta_k^\beta))
$$

(6.7)
for all $u$ in the algebraic sum $\tilde{U}_k := U_{2\delta_0}(0) + U_{2\delta_1}(0) + \cdots + U_{2\delta_{k-1}}(0)$ and $k$ such that $\delta_k^{\alpha+\beta} > \mu$. Here we see the reason for the constraint $\delta_0 \leq \nu/3$ (rather than the seemingly more natural $\delta_0 \leq \nu$): \eqref{eq:C0} requires that $\tilde{U}_k \subseteq U_\nu(0)$, which is indeed the case for sufficiently small $\delta_0$ independent of $k$.

Choosing $k_\ast$ to be the smallest natural number $k$ for which $\delta_k^{\alpha+\beta} \leq \mu$, we may iterate \eqref{eq:C0} $k_\ast$ times to obtain

$$C(0, \delta_0, \mu, 1 + \varepsilon) \leq \prod_{k=0}^{k_\ast-1} (1 + \delta_k^\nu) \max_u C\left(u, \delta_{k_\ast}, \mu, (1 + \varepsilon) \prod_{k=0}^{k_\ast-1} \exp(\delta_k^\nu)\right).$$

Since $\delta_k^{\alpha+\beta} \leq \mu$ we may now deduce from Proposition \ref{prop:linear Brams-Lieb} that

\begin{equation}
C(0, \delta_0, \mu, 1 + \varepsilon) \leq (1 + \varepsilon)^\rho \max_u \max_{u \in \tilde{U}_{k_\ast}} BL(dB(u), p) \prod_{k=0}^{k_\ast-1} (1 + \delta_k^\nu) \exp(\rho \delta_k^\nu).
\end{equation}

At this point, it is clear that we need the Brascamp–Lieb constant to behave sufficiently nicely with respect to the linear mappings. Fortunately, the continuity result for general data proven recently in \cite{20} and \cite{64} is sufficient here, and ensures that

$$\max_{u \in \tilde{U}_{k_\ast}} BL(dB(u), p) \leq (1 + \varepsilon) \max_{u \in U_{k_\ast}} BL(dB(0), p)$$

for an appropriate choice of a constant $\delta_0$ (depending on $\varepsilon$).

For the product term in \eqref{eq:C0}, note that

$$\log \left( \prod_{k=0}^{k_\ast-1} (1 + \delta_k^\nu) \exp(\rho \delta_k^\nu) \right) \leq (1 + \rho) \sum_{k=0}^{\infty} \delta_k^\nu = (1 + \rho) \sum_{k=0}^{\infty} \delta_0^{\rho k \gamma} \lesssim \delta_0^\nu$$

and so $\prod_{k=0}^{k_\ast-1} (1 + \delta_k^\nu) \exp(\rho \delta_k^\nu) \leq 1 + \varepsilon$, for an appropriate $\varepsilon$-dependent choice of constant $\delta_0$. Putting together the above, we have shown that

$$C(0, \delta_0, \mu, 1 + \varepsilon) \leq (1 + \varepsilon)^{\rho+3} BL(dB(0), p),$$

which establishes \eqref{eq:main estimate}, and hence Theorem \ref{thm:main theorem}.

It remains to prove Proposition \ref{prop:gaussian preliminaries}.

### 6.2. Gaussian preliminaries.

Before embarking on our proof of Proposition \ref{prop:gaussian preliminaries}, we introduce a family of $m$-tuples of gaussians that will be pivotal in our argument.

As described in Section \ref{sec:general preliminaries} our proof of Proposition \ref{prop:gaussian preliminaries} may be interpreted as a certain nonlinear version of the proof of the inequality \eqref{eq:linear Brams-Lieb} for the linear Brascamp–Lieb constant. As the nonlinear Brascamp–Lieb constant $C(u, \delta, \mu, k)$ from Definition \ref{def:Brascamp-Lieb} involves a localisation to the $\delta$-ball $U_{\delta}(u)$, the gaussians that we introduce here will be near-extremisers for the underlying linear datum $(dB(u), p)$, and furthermore, will be such that they remain near-extremisers when truncated to fit $U_{\delta}(u)$. This requires that we choose the gaussian near-extremisers to be “nearly isotropic” with variance “close to” $\delta^{1+\gamma}$ for some suitable $\gamma > 0$ (as will be become apparent, the various constraints of our inductive scheme demand that we choose $\gamma$ satisfying \eqref{eq:gamma_condition}). Corollary \ref{cor:isotropy} will guarantee the existence of such gaussian near-extremisers. Of course, in order for the resulting gaussian near-extremisers to be “sufficiently close to isotropic”, they should correspond to positive definite matrices that are, modulo scalings, “sufficiently close to the identity”. This requires us to use $\delta^\tau$-near extremisers, where the small exponent $\tau$ satisfies \eqref{eq:small exponent}. This subsection clarifies these heuristic ideas, and establishes several key lemmas that will be used in the forthcoming proof of Proposition \ref{prop:gaussian preliminaries}.

Note the dual role of $\delta$ in this set-up: it determines both the accuracy and the scale of the near-extremiser.
Potentially many gaussians are in play here. Not only do we typically have a different gaussian at every point \( u \), but we may also have a different gaussian at every scale, at every point. We note that this second layer of complexity in our family of gaussians may sometimes be removed in cases where extremisers exist. One such special case is when the family of linear maps \( dB(0) \) satisfies the Loomis–Whitney-type kernel condition (1.9). Another special case is when \((dB(0), p)\) is simple — that is, when (1.13) holds with strict inequality for all nontrivial proper subspaces \( V \). In each of these situations our proof of Theorem 1.1 admits some simplification. For example, in the case of simple data, the role of Theorem 1.3 may be replaced with certain regularity properties of gaussian extremisers due to Valdimarsson [93].

To select these \( m \)-tuples of gaussians we shall first choose gaussians that are suitably near-extremising, and then scale them appropriately. We pause to observe that the family of gaussian \( \delta \)-near extremisers, for a given datum \((L, p)\), is invariant under common isotropic scalings of their associated quadratic forms. To be precise, the family of all positive-definite \( A = (A_{j}) \) satisfying

\[
BL_{g}(L, p; A) \geq (1 - \delta) BL(L, p)
\]

is invariant under the scaling \( A \mapsto \lambda A \) for all \( \lambda > 0 \). Using the elementary fact that

\[
BL(L, p; f) = BL_{g}(dB(u), p; A_{u,\delta}) \geq (1 - \delta) BL(dB(u), p).
\]

Here, merely for convenience, we have made a renormalisation to obtain (6.10) and (6.11), rather than the symmetric bounds which arise directly from Corollary 1.4. These bounds formalise our requirement that the \( A_{u,\delta,j} \) are sufficiently “close to the identity”, since \( \tau N < \gamma \), and \( \gamma \) is chosen to be sufficiently small — see (6.3) and (6.2).

In what follows the \( n \times n \) matrix

\[
M_{u,\delta} := \sum_{j=1}^{m} p_{j} dB_{j}(u)^{*} A_{u,\delta,j} dB_{j}(u)
\]

will play a crucial role. It naturally arises via the identity

\[
\prod_{j=1}^{m} \exp(-\pi p_{j} \langle A_{u,\delta,j} dB_{j}(u)x, dB_{j}(u)x \rangle) = \exp(-\pi \langle M_{u,\delta} x, x \rangle).
\]

As we shall see, the family of gaussians

\[
\exp(-\pi \langle M_{u,\delta} x, x \rangle)
\]

(after rescaling and suitable truncations) will be our main tool in passing from scale \( \delta^{\alpha} \) to scale \( \delta \) in Proposition 6.3. It will be important to observe that \( M_{u,\delta} \) is also, in some sense, close to the identity matrix. In particular, there exists a constant \( \nu > 0 \) such that

\[
\sup_{u \in U_{\nu}(0)} \| M_{u,\delta}^{-1/2} \| \lesssim \delta^{-\tau N}
\]
for all $\delta \in (0, \nu)$. To see this, note that if $\lambda$ is any eigenvector of $M_{u, \delta}$ with corresponding unit eigenvector $x$ in $\mathbb{R}^n$, then

$$
\lambda = \sum_{j=1}^{m} p_j \langle A_j \delta B_j(u)x, \delta B_j(u)x \rangle \geq \delta^{2\gamma N} \sum_{j=1}^{m} |\delta B_j(u)x|^2 \geq \delta^{2\gamma N}
$$

for a suitably small constant $\nu > 0$, $u \in U_\nu(0)$ and $\delta \in (0, \nu)$. Here, the first lower bound follows from (6.10). The second lower bound at the origin follows from the fact that finiteness of $\text{BL}(\mathbf{L}, \mathbf{p})$ ensures that the intersection of the kernels of the $L_j$ is trivial, and a suitably small choice of the constant $\nu > 0$ ensures that such a bound extends to all $u \in U_\nu(0)$. The estimate in (6.11) now follows.

Finally, we define a family of $L^1$-normalised, scaled gaussians, denoted $g_{u, \delta} = (g_{u, \delta, j})_{j=1}^{m}$, by

$$
g_{u, \delta, j}(x) = (\delta^{1+\gamma})^{-n_j} \det(A_{u, \delta, j})^{1/2} \exp \left( -\pi \left( A_{u, \delta, j} x, x \right) \right).
$$

By (6.9), (6.12) and the scale-invariance of near-extremisers,

$$
\text{BL}(d\mathbf{B}(u), \mathbf{p}; g_{u, \delta}) \geq (1 - \delta^\gamma) \text{BL}(d\mathbf{B}(u), \mathbf{p}).
$$

These gaussians will be used to achieve localisation in the proof of Proposition 6.4.

We conclude this section with three technical lemmas concerning the gaussians $g_{u, \delta}$, which feed into the proof of Proposition 6.4 in the next section. The first of the lemmas essentially states that we may truncate our gaussian near-extremisers to certain $\delta$-balls without losing too much.

**Lemma 6.5.** For each $\eta > 0$ there exists a constant $\nu > 0$ such that

$$
\int_{\mathbb{R}^n} \prod_{j=1}^{m} g_{u, \delta, j}^p(dB_j(u)x) \, dx \leq (1 + \delta^\eta) \int_{U_\nu(0)} \prod_{j=1}^{m} g_{u, \delta, j}^p(dB_j(u)x) \, dx,
$$

for all $u \in U_\nu(0)$ and $\delta \in (0, \nu)$.

**Proof.** Fix $\eta > 0$. By elementary considerations it suffices to show that

$$
\int_{\mathbb{R}^n \setminus U_\delta(0)} \prod_{j=1}^{m} g_{u, \delta, j}^p(dB_j(u)x) \, dx \leq \delta^{2\eta} \int_{\mathbb{R}^n} \prod_{j=1}^{m} g_{u, \delta, j}^p(dB_j(u)x) \, dx
$$

for sufficiently small $\delta > 0$, depending on $\eta$. To establish (6.16) we first observe that, by (6.13), the integrand is expressible as

$$
\prod_{j=1}^{m} g_{u, \delta, j}^p(dB_j(u)x) = (\delta^{1+\gamma})^{-n} \prod_{j=1}^{m} (\det A_{u, \delta, j})^{p_j/2} \exp \left( -\pi \left( M_{u, \delta, j} x, x \right) \right).
$$

After rescaling and a change of variables (6.16) is equivalent to

$$
\int_{|M_{u, \delta}^{-1/2} x| \geq \delta^{-\gamma}} e^{-\pi |x|^2} \, dx \leq \delta^{2\eta}
$$

and, since $|M_{u, \delta}^{-1/2} x| \leq \delta^{-\tau N} |x|$ (for an appropriate choice of the constant $\nu > 0$) from (6.14), it suffices to show that

$$
\int_{|x| \geq \delta^{-\gamma+\tau N} \delta^{-\gamma N}} e^{-\pi |x|^2} \, dx \leq \delta^{2\eta}
$$

for sufficiently small $\delta$. To see this, we simply observe that

$$
\int_{|x| \geq \delta^{-\gamma+\tau N}} e^{-\pi |x|^2} \, dx \lesssim \int_{|x| \geq \delta^{-\gamma+\tau N}} e^{-\pi |x|^2} \, dx \lesssim (\delta^{-\gamma+\tau N})^n e^{-c(\delta^{-\gamma+\tau N})^2}
$$

for some constant $c > 0$, and therefore (6.17) holds, since $\tau < \gamma/N$, provided that $\delta$ is sufficiently small.

□
Our next technical lemma is the following, which captures the fact that gaussian near-extremisers are stable under perturbations of their centres by amounts which are small relative to their scale. This lemma will come into play as we transition between scales in the proof of the recursive inequality \((6.5)\), and thus refers to the more localised gaussians \(g_{u,\delta\alpha,j}\). Before providing the precise statement, we introduce the affine mappings \(L_{u,y}^{\delta}\) : \(\mathbb{R}^n \to \mathbb{R}^{n_j}\) given by

\[L_{u,y}^{\delta} x = L_{u}^y x - B_{j}(y) = B_{j}(u) + dB_{j}(u)(x - u) - B_{j}(y).\]

**Lemma 6.6.** There exists a constant \(\nu > 0\) such that if \(\delta \in (0,\nu)\), then for \(u \in U_\nu(0)\) and \(y \in U_\delta(u)\),

\[\int_{U_{\nu}(y)} \prod_{j=1}^{m} g_{u,\delta\alpha,j}(dB_{j}(u)(x - y)) \, dx \leq (1 + \delta^\beta) \int_{U_{\nu}(y)} \prod_{j=1}^{m} g_{u,\delta\alpha,j}(L_{u,y}^{\delta} x) \, dx.\]

**Proof.** Let \(\nu > 0\) be a small constant to be chosen below. Fix \(\delta \in (0,\nu)\), \(u \in U_\nu(0)\), and \(y \in U_\delta(u)\). It then suffices to prove, for some \(\beta'\) satisfying \(\beta < \beta'\), that

\[\int_{U_{\nu}(y)} \prod_{j=1}^{m} g_{u,\delta\alpha,j}(dB_{j}(u)(x - y)) \, dx - \delta^{\beta'} \leq \int_{U_{\nu}(y)} \prod_{j=1}^{m} g_{u,\delta\alpha,j}(L_{u,y}^{\delta} x) \, dx,
\]

as we may then obtain the desired estimate by a bootstrapping argument. Indeed, by continuity of the Brascamp–Lieb constant \([21], [12]\) and Lemma 6.5

\[\text{BL}(dB(0), p) \leq 2\text{BL}(dB(u), p) \leq 4 \int_{U_{\nu}(y)} \prod_{j=1}^{m} g_{u,\delta\alpha,j}(dB_{j}(u)(x - y)) \, dx,\]

provided that \(\nu\) is sufficiently small. Therefore \((6.18)\) implies that

\[\left(1 - \frac{4\delta^{\beta'}}{\text{BL}(dB(0), p)}\right) \int_{U_{\nu}(y)} \prod_{j=1}^{m} g_{u,\delta\alpha,j}(dB_{j}(u)(x - y)) \, dx \leq \int_{U_{\nu}(y)} \prod_{j=1}^{m} g_{u,\delta\alpha,j}(L_{u,y}^{\delta} x) \, dx,
\]

and the result quickly follows, provided that the constant \(\nu > 0\) is sufficiently small.

In order to show \((6.18)\), we define

\[G(w) = \prod_{j=1}^{m} g_{u,\delta\alpha,j}(w_j)\]

for \(w = (w_1, \ldots, w_m) \in \mathbb{R}^{n_1 + \cdots + n_m}\). Here, and for the remainder of the proof, we suppress the dependence on \(u\) and \(y\). Thus, \((6.18)\) follows once we show

\[\|G \circ L - G \circ \tilde{L}\|_{L^1(U_{\nu}(y))} \leq \delta^{\beta'},\]

where \(L_j = L_{u,y}^{\delta}\) and \(\tilde{L}_j = dB_{j}(u)(\cdot - y)\), and we shall show this using an appropriate uniform bound on the integrand. For this, first we observe the estimate

\[|\nabla G(w)| \lesssim (\delta^{1+\gamma})^{-\alpha(n+2)}|w|,\]

which is a consequence of \((6.11)\) and a use of the scaling condition \((1.14)\) to collect powers of \(\delta\). Since moreover

\[|L_j x - \tilde{L}_j x| \lesssim \delta^2\]

from \((6.4)\), the mean value theorem implies that

\[\|G \circ L - G \circ \tilde{L}\|_{L^\infty(U_{\nu}(y))} \lesssim (\delta^{1+\gamma})^{-\alpha(n+2)}\delta^{2+\alpha},\]

and hence

\[\|G \circ L - G \circ \tilde{L}\|_{L^1(U_{\nu}(y))} \lesssim \delta^{2-\alpha} (\delta^{-\gamma})^{\alpha(2+n)}.\]
Our choice of $\gamma$ ensures that the exponent on the right-hand side of the above estimate is strictly larger than $\beta$. Provided the constant $\nu > 0$ is sufficiently small, this implies the existence of $\beta' > \beta$ such that (6.19) holds, and completes the proof.

Our final technical lemma shows that appropriately truncated gaussians enjoy a local constancy property. We note that gaussians are not $\kappa$-constant at scale $\mu$ on $\mathbb{R}^3$ for any choice of $\kappa$ and $\mu$ — an observation that reflects the absence of a fixed-time parabolic Harnack principle.

Lemma 6.7. There exists a constant $\nu > 0$ with the following property: for all $\delta \in (0, \nu)$ and $\mu > 0$ satisfying the threshold condition $\delta^{\alpha+\beta} > \mu$, and for all $u \in U_\nu(0)$ and $x \in U_{25}(u)$, the function $g_{u,\delta^{\alpha},j}(L_j^n x - \cdot)$ is $\exp(\delta^\sigma)$-constant at scale $\mu$ on $B_j(U_{25\sigma}(x))$.

Proof. It suffices to show that, given any constant $C > 0$, the function $g_{u,\delta^{\alpha},j}$ is $\exp(\delta^\sigma)$-constant at scale $\mu$ on $U_{C\delta^{\alpha}}(0)$, where $u \in U_\nu(0)$, for an appropriate choice of the constant $\nu > 0$, and $\delta \in (0, \nu)$ and $\mu > 0$ satisfying the threshold condition $\delta^{\alpha+\beta} > \mu$. Indeed, $g_{u,\delta^{\alpha},j}(L_j^n x - \cdot)$ is simply a translation of $g_{u,\delta^{\alpha},j}$, so to establish the desired claim, it is only necessary to check that $L_j^n x - w \in U_{C\delta^{\alpha}}(0)$ for some constant $C > 0$ whenever $w \in B_j(U_{25\sigma}(x))$. This follows since, given such $w \in B_j(U_{25\sigma}(x))$, we write $w = B_j(y)$ for some $y \in U_{25\sigma}(x)$ and use (6.4) to deduce that

$$|L_j^n x - w| \lesssim |x - w|^2 + |B_j(x) - B_j(y)| \lesssim \delta^2 + \delta^\sigma \sim \delta^\sigma.$$

To establish that $g_{u,\delta^{\alpha},j}$ is $\exp(\delta^\sigma)$-constant at scale $\mu$ on $U_{C\delta^{\alpha}}(0)$, we first observe that, for any $R \geq 1$, $\tilde{\mu} \leq R$, $|x| \leq R$ and $|y| \leq \tilde{\mu}$, the gaussian $g_A(x) = \det(A)^{1/2} \exp(-\pi i (Ax, x))$ satisfies

$$g_A(x) \leq \exp(3\pi R \|A\| \tilde{\mu}) g_A(x + y)$$

by elementary considerations. Applying this fact when $A = A_{u,\delta^{\alpha},j}$, $R = C\delta^{-\alpha\gamma}$ and $\tilde{\mu} = \delta^{-\alpha(1+\gamma)} \mu$, we see that $g_{u,\delta^{\alpha},j}$ is $\exp(\eta)$-constant at scale $\mu$ on $CU_{\delta^{\alpha}}(0)$, where

$$\eta = 3\pi C \|A_{u,\delta^{\alpha},j}\| \mu \delta^{-\alpha\gamma} \delta^{-\alpha(1+\gamma)}.$$

In order to use the preceding observation, we should ensure $\tilde{\mu} \leq R$; this is simply a consequence of the fact that $\alpha > 1$ and $\beta, \gamma > 0$ (and provided that the constant $\nu > 0$ is sufficiently small). Moreover, under the threshold condition $\delta^{\alpha+\beta} > \mu$, and for $u \in U_\nu(0)$ with the constant $\nu > 0$ sufficiently small,

$$\eta \lesssim \delta^{\beta-2\alpha\gamma} = \delta^{\sigma} \delta^{\beta-2\alpha\gamma-\sigma}$$

by (6.11). From our choice of $\sigma$, it follows that $\eta \leq \delta^\sigma$ (again, provided that the constant $\nu > 0$ is sufficiently small). This establishes that $g_{u,\delta^{\alpha},j}$ is $\exp(\delta^\sigma)$-constant at scale $\mu$ on $U_{C\delta^{\alpha}}(0)$, for any $u \in U_\nu(0)$, and thus completes the proof of the lemma.

6.3. Proof of the recursive inequality. The main idea in our proof of Proposition 6.4 is that it is possible to run a nonlinear variant of the argument leading to (3.3). In doing so, we first claim that it is enough to prove that there exists a constant $\nu > 0$ such that if $u \in U_\nu(0)$ and $\delta \in (0, \nu)$ satisfies $\delta^{\alpha+\beta} > \mu$, then

$$C(u, \delta, \mu, \kappa) \leq (1 + \delta^\beta)(1 + 2\delta^{\alpha\gamma})^2 \max_{x \in U_{25}(u)} C(x, \delta^{\alpha}, \mu, \kappa \exp(\delta^\sigma)).$$

Since $\sigma < \min\{\beta, \alpha\gamma\}$, the desired inequality in (6.5) follows from (6.20) by choosing the constant $\nu > 0$ sufficiently small. Our proof of (6.20) appeals to Lemmas 6.5 to 6.7 in sequence, with each explicitly imposing its own smallness requirement on $\nu$.

Proof of (6.20). We fix $u \in U_\nu(0)$, and assume that $\delta \in (0, \nu)$ and $\mu$ satisfy the threshold condition $\delta^{\alpha+\beta} > \mu$. We also fix $f_j \in L^1(B_j(U_{25}(u)) \mu, \kappa)$ and, to shorten some forthcoming expressions, we write $F = \prod_{j=1}^n f_j^{\mu_j} \circ B_j$. 

By (6.15),
(6.21) \[ BL(d\mathbf{B}(u), \mathbf{p}; g_{\alpha, \delta}) \geq (1 - \delta^{\alpha}) \, BL(d\mathbf{B}(u), \mathbf{p}), \]
and so, for a sufficiently small constant \( \nu > 0 \),
\[ \int_{U_{\delta}(u)} F(y) \, dy \leq \frac{1 + 2 \delta^{\alpha}}{BL(d\mathbf{B}(u), \mathbf{p})} \int_{U_{\delta}(u)} F(y) \, \prod_{j=1}^{m} g_{\alpha, \delta, j}^{p_{j}}(dB_{j}(u) \, dx \, dy. \]
By Lemma 6.5 with \( \eta = \tau \), followed by a translation change of variables in \( x \),
\[ \int_{U_{\delta}(u)} F(y) \, dy \leq \frac{(1 + \delta^{\alpha})^{2}}{BL(d\mathbf{B}(u), \mathbf{p})} \int_{U_{\delta}(u)} F(y) \, \prod_{j=1}^{m} g_{\alpha, \delta, j}^{p_{j}}(dB_{j}(u) \, (x - y) \, dx \, dy. \]
Now we may apply Lemma 6.6 and then Fubini's theorem, to deduce that
\[ \int_{U_{\delta}(u)} F(y) \, dy \leq \frac{(1 + \delta^{\alpha})^{2}}{BL(d\mathbf{B}(u), \mathbf{p})} \int_{U_{\delta}(u)} F(y) \, \prod_{j=1}^{m} g_{\alpha, \delta, j}^{p_{j}}(L_{j}^{u,y} \, x \, dx \, dy \]
\[ \leq \frac{(1 + \delta^{\alpha})^{2}}{BL(d\mathbf{B}(u), \mathbf{p})} \int_{U_{\delta}(u)} F(y) \, \prod_{j=1}^{m} g_{\alpha, \delta, j}^{p_{j}}(L_{j}^{u,y} \, x \, dx \, dy. \]
We recall that \( L_{j}^{u,y} \) is the affine map given by
\[ L_{j}^{u,y} \, x = L_{j}^{u} \, x - B_{j}(y) = B_{j}(u) + dB_{j}(u)(x - u) - B_{j}(y). \]
The inner integral is thus
\[ \int_{U_{\delta}(u)} \prod_{j=1}^{m} (h_{j}^{x})^{p_{j}}(B_{j}(y)) \, dy \]
where
\[ h_{j}^{x}(w) = f_{j}(w)g_{\alpha, \delta, j}(L_{j}^{u} \, x - w). \]
By assumption, \( f_{j} \in L^{1}(B_{j}(U_{2\delta}(u)); \mu, \kappa) \). Further, if \( x \in U_{\delta}(u) + U_{\delta}(0) \), then \( U_{2\delta}(x) \subseteq U_{2\delta}(u) \), whence (after restriction) \( f_{j} \in L^{1}(B_{j}(U_{2\delta}(x)); \mu, \kappa) \). Moreover, from Lemma 6.7
\[ g_{\alpha, \delta, j}(L_{j}^{u} \, x - \cdot) \in L^{1}(B_{j}(U_{2\delta}(x)); \mu, \exp(\delta^{\alpha})), \]
and therefore
\[ h_{j}^{x} \in L^{1}(B_{j}(U_{2\delta}(x)); \mu, \kappa \exp(\delta^{\alpha})) \]
whenever \( x \in U_{\delta}(u) + U_{\delta}(0) \).
Using the definition of \( C(x, \delta^{\alpha}, \mu, \kappa \exp(\delta^{\alpha})) \), and slightly enlarging the domain of integration in \( x \), we see that
\[ \int_{U_{\delta}(u)} F(x) \, dx \leq \frac{(1 + \delta^{\alpha})^{2}M}{BL(d\mathbf{B}(u), \mathbf{p})} \int_{U_{2\delta}(u)} \prod_{j=1}^{m} \left( \int_{\mathbb{R}^{n}} h_{j}^{x}(y) \, dy \right)^{p_{j}} \, dx \]
\[ = \frac{(1 + \delta^{\alpha})^{2}M}{BL(d\mathbf{B}(u), \mathbf{p})} \int_{U_{2\delta}(u)} \prod_{j=1}^{m} \left( \int_{\mathbb{R}^{n}} f_{j} \ast g_{\alpha, \delta, j}(x + B_{j}(u)) \, dx \right)^{p_{j}}\]
where \( M = \max_{x \in U_{2\delta}(u)} C(x, \delta^{\alpha}, \mu, \kappa \exp(\delta^{\alpha})) \). To conclude, we simply apply the linear inequality and the fact that the gaussians \( g_{\alpha, \delta, j} \) are normalised in \( L^{1} \) to obtain
\[ \int_{U_{\delta}(u)} F(x) \, dx \leq (1 + \delta^{\alpha})^{2}M \prod_{j=1}^{m} \left( \int_{\mathbb{R}^{n}} f_{j} \ast g_{\alpha, \delta, j}(x + B_{j}(u)) \, dx \right)^{p_{j}} \]
\[ = (1 + \delta^{\alpha})^{2}M \prod_{j=1}^{m} \left( \int_{\mathbb{R}^{n}} f_{j}(x_{j}) \, dx_{j} \right)^{p_{j}} \]
and (6.20) follows. □
References

[1] K. Astala, D. Faraco and K. Rogers, On Plancherel’s identity for a two-dimensional scattering transform, Nonlinearity 28 (2015), 2721–2729.

[2] F. Avram, N. Leonenko and L. Sakhno, Harmonic analysis tools for statistical inference in the spectral domain, Dependence in Probability and Statistics, Lecture Notes in Statistics, Springer, vol. 200 (2010), 59–70.

[3] F. Avram, N. Leonenko and L. Sakhno, On a Szegő type limit theorem, the Hölder–Young–Brascamp–Lieb inequality, and the asymptotic theory of integrals and quadratic forms of stationary fields, ESAIM Probab. Stat. 14 (2010), 210–255.

[4] K. I. Babenko, An inequality in the theory of Fourier integrals, Izv. Akad. Nauk SSSR Ser. Mat. 25 (1961), 531–542 (Russian); translated as Amer. Math. Soc. Transl. Ser. 2, 44 (1962), 115–128.

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

[6] F. Barthe, On a reverse form of the Brascamp–Lieb inequality, Invent. Math. 134 (1998), 335–361.

[7] F. Barthe, The Brunn–Minkowski theorem and related geometric and functional inequalities, Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006, Vol. II, 1529–1562.

[8] F. Barthe, D. Cordero-Erausquin, M. Ledoux and B. Maurey, Correlation and Brascamp–Lieb inequalities for Markov semigroups, Int. Math. Res. Not., 10 (2011), 2177–2216.

[9] F. Barthe, D. Cordero-Erausquin and B. Maurey, Entropy of spherical marginals and related inequalities, J. Math. Pures Appl., 86 (2006), 89–99.

[10] F. Barthe, O. Guédon, S. Mendelson and A. Naor, A probabilistic approach to the geometry of the $l^n_p$-ball, Ann. Probab. 33 (2005), 480–513.

[11] M. Beals, Self-spreading and strength of singularities for solutions to semilinear wave equations, Ann. of Math. 118 (1983), 187–214.

[12] W. Beckner, Inequalities in Fourier analysis on $\mathbb{R}^n$, Proc. Nat. Acad. Sci. U.S.A. 72 (1975), 638–641.

[13] W. Beckner, Inequalities in Fourier analysis, Ann. of Math. 102 (1975), 159–182.

[14] I. Bejenaru, The multilinear restriction estimate: a short proof and a refinement, Math. Res. Lett. 24 (2017), 1585–1603.

[15] I. Bejenaru and S. Herr, Convolutions of singular measures and applications to the Zakharov system, J. Funct. Anal. 261 (2011), 478–506.

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

[17] I. Bejenaru, S. Herr and D. Tataru, A convolution estimate for two-dimensional hypersurfaces, Rev. Mat. Iberoamericana 26 (2010), 707–728.

[18] J. Bennett, Aspects of multilinear harmonic analysis related to transversality, in: Harmonic Analysis and Partial Differential Equations, Contemp. Math., 612, Amer. Math. Soc., Providence, RI, 2014, 1–28.

[19] J. Bennett and N. Bez, Some nonlinear Brascamp–Lieb inequalities and applications to harmonic analysis, J. Funct. Anal. 259 (2010), 2520–2556.

[20] J. Bennett, N. Bez, M. G. Cowling and T. C. Flock, Behaviour of the Brascamp–Lieb constant, Bull. Lond. Math. Soc. 49 (2017), 512–518.

[21] J. Bennett, N. Bez, T. C. Flock and S. Lee, Stability of the Brascamp–Lieb constant and applications, Amer. J. Math. 140 (2018), 543–569.

[22] J. Bennett, N. Bez and S. Gutiérrez, Global nonlinear Brascamp–Lieb inequalities, J. Geom. Anal. 23 (2013), 1806–1817.
[23] J. Bennett, N. Bez and S. Gutiérrez, *Transversal multilinear Radon-like transforms: local and global estimates*, Rev. Mat. Iberoam. 29 (2013), 765–788.

[24] J. Bennett and S. Buschenhenke, in preparation.

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

[26] J. Bennett, A. Carbery, M. Christ and T. Tao, *Finite bounds for Hörmander–Brascamp–Lieb multilinear inequalities*, Math. Res. Lett. 17 (2010), 647–666.

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

[28] J. Bennett, A. Carbery and J. Wright, *A nonlinear generalisation of the Loomis–Whitney inequality and applications*, Math. Res. Lett. 12 (2005), 1001–10015.

[29] J. Bourgain, *Besicovitch-type maximal operators and applications to Fourier analysis*, Geom. Funct. Anal. 1 (1991), 147–187.

[30] J. Bourgain, *Fourier transform restriction phenomena for lattice subsets and applications to nonlinear evolution equations I, II*, Geom. Funct. Anal. 3 (1993), 107–156, 209–262.

[31] J. Bourgain and J. E. Colliander, *On wellposedness of the Zakharov system*, Int. Math. Res. Not. 1996 (1996), 515–546.

[32] J. Bourgain and L. Guth, *Bounds on oscillatory integral operators based on multilinear estimates*, Geom. Funct. Anal. 21 (2011), 1239–1295.

[33] J. Bourgain and L. Guth, *Proof of the main conjecture in Vinogradov’s Mean Value Theorem for degrees higher than three*, Ann. of Math. 184 (2016), 633–682.

[34] J. Bourgain, C. Demeter, *Sharp bounds for the cubic Parsell–Vinogradov system in two dimensions*, Adv. Math. 320 (2017), 827–875.

[35] J. Bourgain and T. Hänninen and S. Valdimarsson, *Multilinear duality and factorisation for Brascamp–Lieb-type inequalities with applications*, arXiv:1809.02449.

[36] J. Brascamp and E. H. Lieb, *Best constants in Young’s inequality, its converse, and its generalisation to more than three functions*, Adv. Math. 20 (1976), 151–173.

[37] S. Brazitikos, *Brascamp–Lieb inequality and quantitative versions of Helly’s theorem*, Mathematika 63 (2017), 272–291.

[38] S. Brazitikos and A. Giannopoulos, *Continuous version of the approximate geometric Brascamp–Lieb inequalities*, preprint.

[39] A. Carbery, T. S. Hänninen and S. Valdimarsson, *Multilinear duality and factorisation for Brascamp–Lieb-type inequalities with applications*, arXiv:1809.02449.

[40] A. Carbery and S. Valdimarsson, *The multilinear Kakeya theorem via the Borsuk–Ulam theorem*, J. Funct. Anal. 364 (2013), 1643–1663.

[41] E. A. Carlen and D. Cordero-Erausquin, *Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities*, Geom. Funct. Anal. 19 (2009), 373–405.

[42] E. A. Carlen, E. H. Lieb and M. Loss, *A sharp analog of Young’s inequality on $S^N$ and related entropy inequalities*, J. Geom. Anal. 14 (2004), 487–520.
[47] E. A. Carlen, E. H. Lieb and M. Loss, An inequality of Hadamard type for permanents, Methods Appl. Anal. 13 (2006), 1–17.

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

[49] M. Christ, J. Demmel, N. Knight, T. Scanlon and K. Yelick, Communication lower bounds and optimal algorithms for programs that reference arrays — Part 1, arXiv:1308.0068.

[50] M. Christ, J. Demmel, N. Knight, T. Scanlon and K. Yelick, On Hölder–Brascamp–Lieb inequalities for torsion-free discrete abelian groups, arXiv:1510.04190.

[51] M. Christ, X. Li, C. Thiele and T. Tao, On multilinear oscillatory integrals, nonsingular and singular, Duke Math. J. 130 (2005), no. 2, 321–351.

[52] M. G. Cowling, A. Martini, D. Müller and J. Parcet, The Hausdorff-Young inequality on Lie groups, arXiv:1807.04670.

[53] P. Durcik and C. Thiele, Singular Brascamp–Lieb inequalities, arXiv:1809.08688.

[54] Z. Dvir, A. Garg, R. Oliveira and J. Solymosi, Rank bounds for design matrices with block entries and geometric applications, Discrete Anal. 2018:5, 24 pp.

[55] S. Eberhard, Product mixing in the alternating group, Discrete Anal. 2016:2, 19 pp.

[56] T. Eisner and T. Tao, Large values of the Gowers–Host–Kra seminorms, J. Anal. Math. 117 (2012), 133–186.

[57] D. Ellis, E. Friedgut, G. Kindler and A. Yehudayoff, Geometric stability via information theory, Discrete Anal. 2016:10, 28 pp.

[58] E. Faou, P. Germain and Z. Hani, The weakly nonlinear large-box limit of the 2D cubic nonlinear Schrödinger equation, J. Amer. Math. Soc. 29 (2016), 915–982.

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

[60] J. J. F. Fournier, Sharpness in Young’s inequality for convolution, Pacific J. Math. 72 (1977), 383–397.

[61] J. García-Cuerva, J. M. Marco and J. Parcet, Sharp Fourier type and cotype with respect to compact semisimple Lie groups, Trans. Amer. Math. Soc. 355 (2003), 3591–3609.

[62] J. García-Cuerva and J. Parcet, Vector-valued Hausdorff–Young inequality on compact groups, Proc. Lond. Math. Soc. 88 (2004), 796–816.

[63] R. J. Gardner, The Brunn–Minkowski inequality, Bull. Amer. Math. Soc. 39 (2002), 355–405.

[64] A. Garg, L. Gurvits, R. Oliveira and A. Wigderson, Algorithmic and optimization aspects of Brascamp–Lieb inequalities, via Operator Scaling, Geom. Funct. Anal. 28 (2018), 100–145.

[65] J. Ginibre, Y. Tsutsumi and G. Velo, On the Cauchy problem for the Zakharov system, J. Funct. Anal. 151 (1997), 384–436.

[66] S. Guo and R. Zhang, On integer solutions of Parsell–Vinogradov systems, arXiv:1804.02488.

[67] L. Guth, The endpoint case of the Bennett–Carbery–Tao multilinear Kakeya conjecture, Acta Math. 205 (2010), 263–286.

[68] L. Guth, A short proof of the multilinear Kakeya inequality, Math. Proc. Cambridge Philos. Soc. 158 (2015), 147–153.

[69] L. Guth, J. Hickman and M. Iliopoulou, Sharp estimates for oscillatory integral operators via polynomial partitioning, arXiv:1710.10349.

[70] H. Hirayama and S. Kinoshita, Sharp bilinear estimates and its application to a system of quadratic derivative nonlinear Schrödinger equations, arXiv:1802:06563.
L. Hörmander, *Oscillatory integrals and multipliers on FLp*, Ark. Mat. **11** (1973), 1–11.

C. E. Kenig, R. J. Stanton and P. A. Tomas, *Divergence of eigenfunction expansions*, J. Funct. Anal. **46** (1982), 28–44.

S. Kinoshita, *Well-posedness for the Cauchy problem of the Klein–Gordon–Zakharov system in 2D*, Discrete Contin. Dyn. Syst. **38** (2018), 1479–1504.

S. Klainerman and M. Machedon, *Space-time estimates for null forms and the local existence theorem*, Comm. Pure Appl. Math. **46** (1993), 1221–1268.

A. Klein and B. Russo, *Sharp inequalities for Weyl operators and Heisenberg groups*, Math. Ann. **235** (1978), 175–194.

H. Koch and S. Steinerberger, *Convolution estimates for singular measures and some global nonlinear Brascamp–Lieb inequalities*, Proc. Roy. Soc. Edinburgh Sect. A **145** (2015), 1223–1237.

J. Lehec, *Short probabilistic proof of the Brascamp–Lieb and Barthe theorems*, Canad. Math. Bull. **57** (2014), 585–597.

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

J. Liu, T. A. Courtade, P. W. Cuff and S. Verdú, *A Forward-Reverse Brascamp–Lieb Inequality: Entropic Duality and Gaussian Optimality*, Entropy **20** (2018), 418.

A. Martini, *Joint functional calculi and a sharp multiplier theorem for the Kohn Laplacian on spheres*, Math. Z. **286** (2017), 1539–1574.

T. Ozawa and Y. Tsutsumi, *Existence and smoothing effect of solutions for the Zakharov equations*, Publ. Res. Inst. Math. Sci. **28** (1992), 329–361.

P. Perry, *Global well-posedness and long-time asymptotics for the defocussing Davey-Stewartson II equation in H^1(C)*, with an appendix by M. Christ, J. Spectral Theory **6** (2016), 429–481.

A. Ruiz and A. Vargas, *Partial recovery of a potential from backscattering data*, Comm. Partial Differential Equations **30** (2005), 67–96.

E. M. Stein, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton University Press, 1993.

B. Stovall, *Lp-improving multilinear Radon-like transforms*, Rev. Mat. Iberoam. **27** (2011), 1059–1085.

C. Sulum and P. L. Sulum, *Quelques résultats de régularité pour les équations de la turbulence de Langmuir*, C. R. Acad. Sci. Paris Sér A–B **289** (1979), 173–176.

T. Tao, *Multilinear weighted convolution of L2-functions, and applications to nonlinear dispersive equations*, Amer. J. Math. **123** (2001), 839–908.

T. Tao, *Endpoint bilinear restriction theorems for the cone, and some sharp null form estimates*, Math. Z. **238** (2001), 215–268.

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

T. Tao and J. Wright, *Lp-improving for averages along curves*, J. Amer. Math. Soc. **16** (2003), 605–638.

S. Valdimarsson, *Optimisers for the Brascamp–Lieb inequality*, Israel J. Math. **168** (2008), 253–274.

S. Valdimarsson, *The Brascamp–Lieb polyhedron*, Canad. J. Math. **62** (2010), 870–888.

S. Valdimarsson, *Geometric Brascamp–Lieb has the optimal best constant*, J. Geom. Anal. **21** (2011), 1036–1043.

N. Vishnoi and O. Yıldız, *On geodesically convex formulations for the Brascamp–Lieb constant*, arXiv:1804:04051.
[95] R. Zhang, *The endpoint perturbed Brascamp–Lieb inequality with examples*, Anal. PDE 11 (2018), 555–581.

[96] P. Zorin–Kranich, *Kakeya–Brascamp–Lieb inequalities*, arXiv:1807.09604v2.

School of Mathematics, The Watson Building, University of Birmingham, Edgbaston, Birmingham, B15 2TT, England.
E-mail address: J.Bennett@bham.ac.uk

Department of Mathematics, Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan.
E-mail address: nealbez@mail.saitama-u.ac.jp

Mathematisches Seminar, Christian-Albrechts-Universität zu Kiel, 24116 Kiel, Germany.
E-mail address: buschenhenke@math.uni-kiel.de

School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia.
E-mail address: m.cowling@unsw.edu.au

Department of Mathematics and Statistics, Lederle Graduate Research Tower, University of Massachusetts, 710 N. Pleasant Street, Amherst, MA 01003-9305, USA.
E-mail address: flock@math.umass.edu