On the nonlinear Brascamp-Lieb inequality

Bennett, Jonathan; Bez, Neal; Buschenhenke, Stefan; Cowling, Michael G.; Flock, Taryn C.

Document Version
Peer reviewed version

Link to publication on Research at Birmingham portal

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. 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. Instances of this inequality are quite prevalent in mathematics, and we illustrate this with some applications in harmonic analysis.

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

$$(1.1) \qquad \int_{\mathbb{R}^n} \prod_{j=1}^m f_j^{p_j}(L_j x)\, dx \le \mathrm{BL}(\mathbf{L}, \mathbf{p}) \prod_{j=1}^m \left(\int_{\mathbb{R}^{n_j}} f_j(x_j)\, dx_j \right)^{p_j}$$

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 $\mathrm{BL}(\mathbf{L}, \mathbf{p})$ denotes the smallest constant (which may be infinite). We refer to $(\mathbf{L}, \mathbf{p}) = ((L_j)_{j=1}^m, (p_j)_{j=1}^m)$ as the *Brascamp–Lieb datum*, and $\mathrm{BL}(\mathbf{L}, \mathbf{p})$ as the *Brascamp–Lieb constant*. This celebrated inequality was first formulated in [29] and has given rise to an elegant theory with wide-ranging applications across the mathematical sciences, from convex geometry, to information theory, harmonic analysis, partial differential equations, number theory and theoretical computer science. We refer to [19], [2], [3], [32], [11], [30], [26], [41] and the references there for some further discussion of these and other connections.

A natural question that has arisen in harmonic analysis and partial differential equations in recent years, beginning with [22], is whether one may replace the linear surjections L_j in (1.1) 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.

Theorem 1.1. *Let* $(\mathbf{L}, \mathbf{p})$ *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* $1 \le j \le m$. *Then, for each* $\varepsilon > 0$ *there exists a neighbourhood* U *of* x_0 *such that*

$$(1.2) \qquad \int_U \prod_{j=1}^m f_j^{p_j}(B_j(x))\, dx \le (1 + \varepsilon)\, \mathrm{BL}(\mathbf{L}, \mathbf{p}) \prod_{j=1}^m \left(\int_{\mathbb{R}^{n_j}} f_j(x_j)\, dx_j \right)^{p_j}.$$

Theorem 1.1 has some significant applications, for which we refer to Section 2.

As may be expected, 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.

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 $\mathbf{B}$. This began with [22] and [12], where the underlying linear maps $\mathbf{L}$ in the statement of Theorem 1.1 were required to satisfy the highly-restrictive "basis condition"

$$\bigoplus_{j=1}^{m} \ker L_j = \mathbb{R}^n, \tag{1.3}$$

which keeps the complexity at the level of the classical Loomis–Whitney inequality; see also [10], [48], [31], and applications in [9], [8], [36], [44], [43], and [46]. We note that the results in [31] also feature the best-possible $(1 + \varepsilon)$ factor present in (1.2), and cater for a wider class of data originating in work of Finner [37]. Further examples beyond the scope of (1.3) may also be found in unpublished work of Stovall. Very different arguments in [57] allowed for polynomial $\mathbf{B}$, but generated bounds that depend strongly on the degree. The only nonlinear statements that were known for completely general data involved an ε-loss in the scale of Sobolev spaces; see [16].

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 throughout. A particularly important example of linearity at work is in a fundamental theorem of Lieb [49], which states that the Brascamp–Lieb constant $\mathrm{BL}(\mathbf{L}, \mathbf{p})$ is determined by centred gaussian inputs f_j; that is

$$f_j(x) = c_j \exp(-\pi \langle A_j x, x \rangle), \tag{1.4}$$

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

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

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

Theorem 1.2 (Lieb)**.**

$$\mathrm{BL}(\mathbf{L}, \mathbf{p}) = \sup_{\mathbf{A}} \mathrm{BL_g}(\mathbf{L}, \mathbf{p}; \mathbf{A}), \tag{1.5}$$

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

$$\mathrm{BL_g}(\mathbf{L}, \mathbf{p}; \mathbf{A}) = \frac{\prod_{j=1}^{m} \det(A_j)^{p_j/2}}{\det \left(\sum_{j=1}^{m} p_j L_j^* A_j L_j \right)^{1/2}}.$$

Lieb's theorem, which clearly only makes sense for linear L_j, has been central to much of our understanding of the Brascamp–Lieb inequality. For instance, it was used in [19] to show that $\mathrm{BL}(\mathbf{L}, \mathbf{p})$ is finite if and only if the transversality condition

$$\dim(V) \leq \sum_{j=1}^{m} p_j \dim(L_j V) \tag{1.6}$$

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

$$n = \sum_{j=1}^{m} p_j n_j. \tag{1.7}$$

In the setting of *nonlinear* data, three viable approaches have emerged: (i) the method of *refinements* of Christ [34], (ii) the *factorisation* method of Carbery, Hänninen and Valdimarsson [31], and (iii) the method of *induction-on-scales*, introduced in this setting by Bejenaru, Herr and Tataru [10]. All of these methods are at least effective under the rather rigid Loomis–Whitney-type structural condition (1.3), as may be seen in [22], [31] and [12] respectively.

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 [24] 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 [53] and [10]. 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 "δ-near-extremisers", also known as "δ-optimal solutions" in the optimisation theory literature (see, for example, [23]). We refer to an input $\mathbf{f} = (f_j)_{j=1}^m$ as a δ-near-extremiser for (1.1) if

$$(1.8) \qquad \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{f}) := \frac{\int_{\mathbb{R}^n} \prod_{j=1}^m f_j^{p_j}(L_j x)\, dx}{\prod_{j=1}^m \left(\int_{\mathbb{R}^{n_j}} f_j(x_j)\, dx_j \right)^{p_j}} \geq (1 - \delta)\, \mathrm{BL}(\mathbf{L}, \mathbf{p}).$$

Of course, Lieb's Theorem guarantees that every Brascamp–Lieb functional $\mathbf{f} \mapsto \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{f})$ has *gaussian* δ-near-extremisers, for all $\delta > 0$. What it doesn't tell us is anything quantitative about the set of gaussian δ-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 $\mathbf{p}$ and underlying dimensions $n, n_1, \ldots, n_m$, and refer to an m-tuple of linear maps $\mathbf{L}$ as *admissible* if each $L_j : \mathbb{R}^n \to \mathbb{R}^{n_j}$ is a surjection (in other words, $(\mathbf{L}, \mathbf{p})$ 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 $\mathbf{L}$,*

$$(1.9) \qquad \inf_{\substack{\|C_j^* A_j C_j\| \leq \delta^{-N} \\ \|(C_j^* A_j C_j)^{-1}\| \leq \delta^{-N}}} \mathrm{BL_g}(\mathbf{L}, \mathbf{p}; \mathbf{A})^{-2} \leq \mathrm{BL}(\mathbf{L}, \mathbf{p})^{-2} + \delta \prod_{j=1}^m \det(L_j L_j^*)^{p_j}.$$

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

We remark that whilst $\mathrm{BL_g}(\mathbf{L}, \mathbf{p}; \mathbf{A})$ is invariant under the isotropic rescaling $\mathbf{A} \mapsto \lambda \mathbf{A}$, 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 $(\mathbf{L}_0, \mathbf{p})$ 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 $\delta \in (0, \delta_0)$,*

$$\sup_{\|A_j\|, \|A_j^{-1}\| < \delta^{-N}} \mathrm{BL_g}(\mathbf{L}, \mathbf{p}; \mathbf{A}) \geq (1 - \delta)\, \mathrm{BL}(\mathbf{L}, \mathbf{p})$$

for all $\mathbf{L}$ sufficiently close to $\mathbf{L}_0$.

In brief, both of these statements tell us that we may find gaussian δ-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 $\mathbf{L}$. 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 δ_0 are entirely independent of $\mathbf{L}$, given fixed $\mathbf{p}$ 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 $\mathrm{BL}(\cdot, \mathbf{p})$, established using Lieb's theorem in [15] and [41], 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 "δ-near-minimisers" for functions of the form

$$(1.10) \qquad \mathbb{R}^n \ni y \mapsto \sum_{j \in J} d_j \exp(\langle u_j, y \rangle),$$

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_δ as a δ-*near-minimiser* for the function in (1.10) if

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

Notice that this is a near-minimiser in an "additive" sense; this differs somewhat from our earlier definition of a δ-near-extremiser in (1.8) for the Brascamp–Lieb functional. An elementary example of a function of the form (1.10) 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 δ-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 Corollaries 2.2 and 2.4. 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 (1.10), 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.

2. Applications

From the point of view of applications, the Brascamp–Lieb inequality (1.1) is often more usefully expressed as a bound on a multilinear form, namely

$$(2.1) \qquad \int_{\mathbb{R}^n} \prod_{j=1}^m f_j(L_j x) \, dx \leq \mathrm{BL}(\mathbf{L}, \mathbf{p}) \prod_{j=1}^m \|f_j\|_{L^{q_j}(\mathbb{R}^{n_j})};$$

here we have simply replaced $f_j^{p_j}$ with f_j, and p_j with $1/q_j$. This is a parametrised version of the more intrinsic inequality

$$(2.2) \qquad \int_H f_1 \otimes \cdots \otimes f_m \, d\mu_H \leq \mathrm{BL}(H, \mathbf{q}) \prod_{j=1}^m \|f_j\|_{L^{q_j}(\mathbb{R}^{n_j})},$$

where H is a subspace of $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}$, and by a minor abuse of notation, we use $\mathrm{BL}(H, \mathbf{q})$ to denote the best constant. Here, integration in (2.2) is with respect to Lebesgue measure on H, and the connection between (2.1) and (2.2) of course comes from expressing H as the range of the linear map $x \mapsto (L_1 x, \ldots, L_m x)$; we caution that the constants $\mathrm{BL}(H, \mathbf{q})$ and $\mathrm{BL}(\mathbf{L}, \mathbf{p})$ will not in general coincide under this identification, although their finiteness is certainly equivalent. From this perspective Theorem 1.1 acquires the following useful form:

Theorem 2.1. *Suppose M is a C^2 submanifold of $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}$ and $x \in M$. Then given any $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$(2.3) \qquad \int_{M \cap B(x,\delta)} f_1 \otimes \cdots \otimes f_m \, d\mu_M \leq (1+\varepsilon) \, \mathrm{BL}(T_x M, \mathbf{q}) \prod_{j=1}^m \|f_j\|_{L^{q_j}(\mathbb{R}^{n_j})}.$$

Here $d\mu_M$ denotes Lebesgue measure on M, and $T_x M$ denotes the tangent space to M at x.

Multilinear forms of the type handled by Theorem 2.1 arise often in analysis. In this section we illustrate this with two substantial applications. The first provides a broad family of multilinear convolution inequalities that arise naturally in dispersive PDE, where the generality of Theorem 2.1 allows for rather minimal structural hypotheses. The second 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.

Remark. Theorem 2.1 belongs to the theory of Radon-like transforms, and provides a sharp multilinear estimate under minimal first order structural (as opposed to regularity) hypotheses on the manifold M. It is not difficult to show that imposing additional higher order (curvature-related) conditions does not lead to improved bounds; see [17, Lemma 6] for an explicit statement of this fact. 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 M — see [55] and [51]. Obtaining such a higher order result at the level of generality of Theorem 2.1 would be of considerable interest.

2.1. **Euclidean harmonic analysis and partial differential equations.** The regularising effect of convolution on functions, and more generally measures, has long been a matter at the heart of harmonic analysis and its applications. A particularly natural question that arises in both the restriction theory of the Fourier transform and nonlinear dispersive PDE, concerns convolutions of singular measures supported on submanifolds of $\mathbb{R}^n$, and seeks to understand the conditions under which such convolutions may give rise to bounded or L^p densities – see for example, [54], [52]. Such phenomena have been important, for example, in recent breakthroughs in the low-regularity theory of the Zakharov system of [9] and [8], and rest on certain "transversality" properties of the families of submanifolds involved. In these works the underlying estimates ultimately rely on nonlinear perturbations of the Brascamp–Lieb inequality for data satisfying the Loomis–Whitney type condition (1.3). There are further manifestations of the nonlinear Brascamp–Lieb inequality in PDE-related problems – see for example, [36], [46], [44], [43], [18].

In this section we use our general nonlinear Brascamp–Lieb inequality (in the form of Theorem 2.1) to establish a much broader class of multilinear convolution estimates, where the manifolds involved satisfy minimal structural (or "transversality") conditions.

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 $\mathbb{R}^n$. We equip S_j with the measure $d\sigma_j$, 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 the differential $d\Sigma_j(0)$ is the tangent space to S_j at the (arbitrary) point $\Sigma_j(0)$. Finally, we denote by $d\mathbf{\Sigma}(0)^*$ the m-tuple of linear maps $(d\Sigma_1(0)^*, \ldots, d\Sigma_m(0)^*)$.

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

$$(2.4) \qquad \|g_1 d\sigma_1 * \cdots * g_m d\sigma_m\|_{L^\infty(\mathbb{R}^n)} \leq C \prod_{j=1}^{m} \|g_j\|_{L^{q_j}(S_j)}$$

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

Some remarks are in order. The first thing to observe is that the condition $\mathrm{BL}(d\mathbf{\Sigma}(0)^*, \mathbf{p}) < \infty$ may be replaced with the manifestly geometric condition

$$(2.5) \qquad \sum_{j=1}^{m} \frac{\dim(V \cap (T_{\Sigma_j(0)} S_j)^\perp)}{q_j'} \leq (\rho - 1) \dim(V);$$

here $\rho = \sum_{j=1}^{m} 1/q_j'$, and (2.5) should hold for all subspaces V of $\mathbb{R}^n$, with equality when $V = \mathbb{R}^n$. This is simply an interpretation of the finiteness condition (1.6) in this context, and we clarify that $T_{\Sigma_j(0)} S_j$ denotes the tangent space to S_j at $\Sigma_j(0)$.

Corollary 2.2 is closely related to the conjectural "restriction Brascamp–Lieb inequality" studied in [16] and [57]. This very general conjectural inequality states that, under the hypotheses of

Corollary 2.2,

$$(2.6) \qquad \int_{\mathbb{R}^n} \prod_{j=1}^m |\widehat{g_j d\sigma_j}(x)|^{2p_j} \, dx \le C \prod_{j=1}^m \|g_j\|_{L^2(S_j)}^{2p_j};$$

see [16] for a weaker local form of this inequality, which generalises the multilinear restriction inequality of [21], along with those used in the recent theory of decouplings in, for example, [26] and [27]. For instance, if $p_j \le 1/2$ (or equivalently $1 \le q_j \le 2$), for all j, then (2.6) very quickly implies (2.4), at least when the g_j are characteristic functions of sets. To see this, we observe that

$$\|g_1 d\sigma_1 * \cdots * g_m d\sigma_m\|_{L^\infty(\mathbb{R}^n)} \le \int_{\mathbb{R}^n} \prod_{j=1}^m |\widehat{g_j d\sigma_j}(x)| \, dx$$

$$\le \prod_{j=1}^m \|\widehat{g_j d\sigma_j}\|_{L^\infty(\mathbb{R}^n)}^{1-2p_j} \int_{\mathbb{R}^n} \prod_{k=1}^m |\widehat{g_k d\sigma_k}(x)|^{2p_k} \, dx$$

$$\le C \prod_{j=1}^m \|g_j\|_{L^1(S_j)}^{1-2p_j} \prod_{k=1}^m \|g_k\|_{L^2(S_k)}^{2p_k} = C \prod_{j=1}^m \|g_j\|_{L^{q_j}(S_j)}.$$

Here we have used the trivial $L^1 \to L^\infty$ bound for the extension operators, along with the conjectural inequality (2.6), 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 [10] and [12]. We refer to [13] for a detailed analysis of the relationship between (2.4) and (2.6).

Corollary 2.2 contains subtleties which are related to the appearance of the dual exponent q_j' in the transversality assumption (2.5). In order to deal with these we first establish the following general (Fourier) duality principle for Brascamp–Lieb data, phrased in terms of the Brascamp–Lieb constant appearing in (2.2).

Proposition 2.3. *Suppose* $\mathbf{q} = (q_j)_{j=1}^m \in [1, \infty]^m$. *Then* $\mathrm{BL}(H, \mathbf{q}) = B_\mathbf{q} \mathrm{BL}(H^\perp, \mathbf{q}')$, *where*

$$B_\mathbf{q} = \prod_{j=1}^m A_{q_j}^{n_j}$$

and

$$(2.7) \qquad A_r = \left(\frac{r^{1/r}}{r'^{1/r'}} \right)^{1/2}.$$

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

$$(2.8) \qquad \mathrm{BL}(H, \mathbf{q}) \le B_\mathbf{q} \mathrm{BL}(H^\perp, \mathbf{q}').$$

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

$$\mathrm{BL}(H, \mathbf{q}) = \lim_{\nu \to 0} \mathrm{BL}(H, \mathbf{q}; \mathbf{f}_\nu),$$

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,

$$(2.9) \qquad \int_H f_1 \otimes \cdots \otimes f_m \, d\mu_H = \int_{H^\perp} \widehat{f}_1 \otimes \cdots \otimes \widehat{f}_m \, d\mu_{H^\perp}$$

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

$$\mathrm{BL}(H, \mathbf{q}; \mathbf{f}_\nu) = \mathrm{BL}(H^\perp, \mathbf{q}'; \widehat{\mathbf{f}}_\nu) \prod_{j=1}^m \frac{\|\widehat{f}_{\nu,j}\|_{q_j'}}{\|f_{\nu,j}\|_{q_j}}$$

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

$$\prod_{j=1}^m \frac{\|\widehat{f}_{\nu,j}\|_{q_j'}}{\|f_{\nu,j}\|_{q_j}} = B_\mathbf{q}$$

for each $\nu > 0$, and thus (2.8) follows. $\square$

Proposition 2.3 appears to be new, and so merits some brief remarks. First of all it implies the "Fourier–Brascamp–Lieb inequality"

$$\int_H f_1 \otimes \cdots \otimes f_m \, d\mu_H \leq \mathrm{BL}(H, \mathbf{q}) \prod_{j=1}^m A_{q_j}^{-n_j} \|\widehat{f}_j\|_{L^{q_j'}(\mathbb{R}^{n_j})},$$

which may, depending on the size of the exponents q_j relative to 2, be a stronger bound than the usual Brascamp–Lieb inequality (2.1). An early manifestation of this idea may be found in [2].

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

$$\sum_{j=1}^m L_j^* \widetilde{L}_j = 0;$$

here $\widetilde{n} = n_1 + \cdots + n_m - n$. 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 $\widetilde{\mathbf{L}}$ is rather more manageable than the data $\mathbf{L}$.

Proof of Corollary 2.2. We begin by observing that

$$(2.10) \qquad g_1 d\sigma_1 * \cdots * g_m d\sigma_m(x) = \int_{U_1 \times \cdots \times U_m} f_1 \otimes \cdots \otimes f_m(y) \delta(F(y, x)) \, dy,$$

where $f_j = g_j \circ \Sigma_j$, and

$$(2.11) \qquad F(y, x) = \Sigma_1(y_1) + \cdots + \Sigma_m(y_m) - x.$$

By considering input functions with sufficiently small supports, it suffices to prove that

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

Hence by (2.10), it is enough 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

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

μ is a suitable measure on M, and F is given by (2.11). By Theorem 2.1 it suffices to prove that $\mathrm{BL}(T_0 M, \mathbf{q}) < \infty$ 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 $\mathrm{BL}(d\mathbf{\Sigma}(0)^*, \mathbf{p}) < \infty$, Corollary 2.2 now follows from Proposition 2.3. $\square$

2.2. **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 [1], Beckner [6, 7] and Brascamp–Lieb [29] which covered the euclidean space $\mathbb{R}^n$. For example, in the 1970s, Fournier [38] showed that the optimal constant in both inequalities is less than 1 for nonabelian groups with no compact open subgroups, and Klein and Russo [47] 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 [39, 40] 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 [35] computed the best constant for the Young convolution inequality and Hausdorff–Young inequality for a more extensive class of non-abelian 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 $Y_{\mathbf{q}}(G)$ 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 $Y_{\mathbf{q}}(G; U)$ 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 $Y_{\mathbf{q}}(G; U)$ get smaller as U gets smaller, and so it makes sense to define the constant $Y_{\mathbf{q}}^{\mathrm{loc}}(G)$ to be the infimum of the $Y_{\mathbf{q}}(G; U)$ as U shrinks to the identity. It is easy to see that

$$(2.12) \qquad\qquad Y_{\mathbf{q}}^{\mathrm{loc}}(G) \leq Y_{\mathbf{q}}(G) \leq 1;$$

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 function becomes closer to 1. As a result, one might also define $Y_{\mathbf{q}}^{\mathrm{loc}}(G)$ 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 $Y_{\mathbf{q}}(G)$ and $Y_{\mathbf{q}}^{\mathrm{loc}}(G)$ 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 $Y_{\mathbf{q}}(G)$ or $Y_{\mathbf{q}}^{\mathrm{loc}}(G)$ before this work.

By using a contraction argument, it was shown in [35, Proposition 2.4] that

$$(2.13) \qquad\qquad Y_{\mathbf{q}}^{\mathrm{loc}}(G) \geq (A_{q_1} A_{q_2} A_{q_3})^{\dim(G)},$$

where, for any $q \in [1, \infty]$, the constant A_q is given by (2.7).

Combined with the inequality (2.12) and the estimates for solvable groups, it followed that $Y_{\mathbf{q}}(G)$, for a connected simply connected unimodular solvable group G, is exactly the right-hand side of (2.13). In particular, $Y_{\mathbf{q}}(G)$ 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 ($Y_{\mathbf{q}}(G)$ or $Y_{\mathbf{q}}^{\mathrm{loc}}(G)$) was the trivial bound (2.12). The following corollary of the nonlinear Brascamp–Lieb inequality (Theorem 1.1) identifies the local constant $Y_{\mathbf{q}}^{\mathrm{loc}}(G)$ in full generality.

Corollary 2.4. *Suppose that G is a connected Lie group. Then*

$$Y_{\mathbf{q}}^{\mathrm{loc}}(G) = (A_{q_1} A_{q_2} A_{q_3})^{\dim(G)}.$$

It is straightforward to verify that Corollary 2.4 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.

Young's convolution theorem shows that sharp local nonlinear Brascamp–Lieb inequalities do not in general give rise to sharp global inequalities. For example, for a compact Lie group G, we have shown that $Y_{\mathbf{q}}^{\mathrm{loc}}(G) < 1$, but $Y_{\mathbf{q}}(G) = 1$, so the global constant is larger than the local constant. This is in contrast with the case of homogeneous groups, such as the Heisenberg group, where the local and global constants are easily seen to be the same by scaling. We refer the reader to [35] for much more on this topic, including many references and its connection with the best constant in the Hausdorff–Young inequality. It should be remarked here that there are other examples of global nonlinear Brascamp–Lieb inequalities established with sharp constants – see for instance Carlen–Lieb–Loss [33], Barthe–Cordero-Erausquin–Maurey [5], Barthe–Cordero-Erausquin–Ledoux–Maurey [4], and Bramati [28].

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 [2]. This inequality captures an important self-similarity property of the *Brascamp–Lieb functional*

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

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. Indeed the method of induction-on-scales has been fruitful on several occasions in this wider context — see [11, 21, 10, 12, 42, 16], along with closely-related papers in Fourier restriction theory beginning with [24].

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

$$
\begin{aligned}
\mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{f})\, \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{g}) &= \int_{\mathbb{R}^n} \left(\prod_{j=1}^m (f_j^{p_j} \circ L_j) \right) * \left(\prod_{k=1}^m (g_k^{p_k} \circ L_k) \right)(x)\, dx \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \prod_{j=1}^m [f_j(L_j y) g_j(L_j(x - y))]^{p_j}\, dy \right) dx \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \prod_{j=1}^m [f_j(L_j y) g_j(L_j x - L_j y))]^{p_j}\, dy \right) dx \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \prod_{j=1}^m [h_j^x(L_j y)]^{p_j}\, dy \right) dx,
\end{aligned}
$$

(3.1)

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

$$
\begin{aligned}
\mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{f})\, \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{g}) &\leq \int_{\mathbb{R}^n} \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{h}^x) \prod_{j=1}^m \left(\int_{\mathbb{R}^{n_j}} h_j^x(z_j)\, dz_j \right)^{p_j} dx \\
&= \int_{\mathbb{R}^n} \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{h}^x) \prod_{j=1}^m (f_j * g_j(L_j x))^{p_j}\, dx \\
&\leq \sup_x \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{h}^x)\, \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{f} * \mathbf{g}).
\end{aligned}
$$

This we refer to as Ball's inequality. It should be noted that the L^1 normalisation hypotheses on $\mathbf{f}$ and $\mathbf{g}$ may be dropped by homogeneity considerations.

Now, if $\mathbf{g}$ is a δ-*near-extremiser* for (1.1), in the sense that (1.8) holds, then we may immediately deduce that

(3.2) $$\mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{f}) \leq (1 + O(\delta))\, \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{f} * \mathbf{g})$$

and

(3.3) $$\mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{f}) \leq (1 + O(\delta)) \sup_x \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{h}^x).$$

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 [19]), 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 [33] and [19] 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 δ-near-extremising input $\mathbf{g}$ *which resembles a bump function with small support*, then the function

$$h_j^x(z) = f_j(z)g_j(L_jx - z)$$

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 (δ 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 $(\mathbf{L}, \mathbf{p})$ 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 δ-near-extremisers. In particular, it does not tell us whether there exist δ-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 δ is taken from a sufficiently lacunary sequence (δ_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.

It should be remarked that for Brascamp–Lieb data where gaussian extremisers exist, such as for the class of so-called *simple data*, there is no need for our effective version of Lieb's theorem. This simplification was helpful in devising the forthcoming induction-on-scales argument – see the unpublished work [14] which this paper supercedes.

The basic inductive scheme described here should be compared with that in [10] and [12] in the setting of Loomis–Whitney-type data (data satisfying the structural condition (1.3)). 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 [33], [19] and [21].

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

$$(4.1) \qquad\qquad f_I(y, d) = \sum_{j \in I} d_j \exp(\langle u_j, y \rangle),$$

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

$$(4.2) \qquad\qquad g_I(d) = \inf_{y \in \mathbb{R}^n} f_I(y, d).$$

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_δ? Minimisers are not necessarily unique and can vary on an affine subspace, which means that y_δ need not be bounded in terms of δ. Nevertheless, as we will show, we can always find a δ-near-minimiser that is bounded logarithmically in δ.

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

$$(4.3) \qquad g_{I_1}(d) \leq g_{I_2}(d).$$

For any subset I of J, we denote by $K(I)$ the convex hull of the set $\{u_j : j \in I\}$. This is of course a convex polytope in $\mathbb{R}^n$. We denote by $K(I)^\circ$ the relative interior of $K(I)$, that is, the interior of $K(I)$ relative to its affine hull, and denote by $\partial K(I) = K(I) \setminus K(I)^\circ$ its relative boundary. The polytope $K(I)$ 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 $f_I(\cdot, d)$ 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 $f_I(\cdot, d)$, 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 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[1]

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

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

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

for all subsets I of J such that $0 \in K(I)^\circ$. We note that since J is a finite set, 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.

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*

$$(4.4) \qquad \langle v, u_j \rangle = 0 \quad \text{if } j \in I_1,$$

$$(4.5) \qquad \langle v, u_j \rangle \geq c_1 \quad \text{if } j \in I \setminus I_1,$$

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 maximal 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

$$\langle v, y \rangle = 0 \qquad \text{if } y \in K(I_1),$$

$$\langle v, y \rangle > 0 \qquad \text{if } y \in K(I) \setminus K(I_1),$$

and thus (4.5) holds as long as we choose $c_1 \in (0, 1)$ such that

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

---

[1] For technical reasons, it is more convenient if $C_0 \geq 1$, though it is not crucial for the argument.

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 $0 \notin K(I)$. □

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

$$(4.6) \qquad \Delta = \Delta(I, d) = \min_{j \in I} d_j.$$

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*

$$(4.7) \qquad f_I(y, d) = g_I(d)$$

and

$$(4.8) \qquad |y| \leq \frac{1}{c_0} \log \frac{|I|}{\Delta}.$$

Proof. After restriction to 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 $v \in \text{span } K(I)$.[2]

From the definition of c_0, it follows that $c_0 y/|y|$ is contained in $K(I)$, so we can write

$$(4.9) \qquad c_0 y = |y| \sum_{j \in I} \lambda_j u_j,$$

where $\lambda_j \in [0, 1]$ and $\sum_{j \in I} \lambda_j = 1$. Then

$$(4.10) \qquad c_0 |y| = \sum_{j \in I} \lambda_j \langle u_j, y \rangle,$$

so there exists at least one j (depending on y) such that $\langle u_j, y \rangle \geq c_0 |y|$. We conclude that

$$(4.11) \qquad \Delta \exp(c_0 |y|) \leq d_j \exp(\langle u_j, y \rangle) \leq f_I(y, d) = g_I(d) \leq f_I(0, d) \leq |I|,$$

which immediately gives the bound (4.8). □

Remark. A simple inequality that will be useful below is

$$(4.12) \qquad f_I(y, d) \leq |J| D \exp(C_0 R)$$

whenever $I \subseteq J$, $|y| \leq R$ and $d \in [0, D]^I$.

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 δ-near-minimiser of size comparable to $\log(1/(\delta\Delta))$.

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 $d \in (0, 1]^I$, there exists $y \in \mathbb{R}^n$ such that*

$$f_I(y, d) \leq g_I(d) + \delta$$

and

$$(4.13) \qquad |y| \leq C_1 \log \frac{1}{\delta\Delta}.$$

---

[2]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.

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 4.3 (if $I_1 = \emptyset$, then $y_0 = 0$ will trivially fulfil (4.7) and (4.8)). We will modify y_0 by moving along v, which keeps f_{I_1} stable and shrinks the remainder term $f_I - f_{I_1} = f_{I \setminus I_1}$. Thus we set $y = y_0 - sv$, where s will be chosen later, and use (4.4) and (4.5) to estimate

$$
\begin{aligned}
f_I(y, d) &= \sum_{j \in I_1} d_j \exp(\langle u_j, y_0 \rangle) + \sum_{j \in I \setminus I_1} d_j \exp(\langle u_j, y_0 \rangle - s \langle u_j, v \rangle) \\
&\leq f_{I_1}(y_0, d) + f_{I \setminus I_1}(y_0, d) \exp(-s c_1) \\
&\leq g_{I_1}(d) + |J| \left(\frac{|J|}{\Delta} \right)^{C_0/c_0} \exp(-s c_1).
\end{aligned}
$$

For the final inequality we used the trivial estimate (4.12). Therefore, from the monotonicity property (4.3) and the choice

$$
s = \frac{1}{c_1} \left(\log \frac{1}{\delta} + \log |J| + \frac{C_0}{c_0} \log \frac{|J|}{\Delta} \right),
$$

it follows that

(4.14) $$ f_I(y, d) \leq g_{I_1}(d) + \delta \leq g_I(d) + \delta. $$

We note the above choice of s is strictly positive.

To complete the proof, we choose

(4.15) $$ \delta_1 = \frac{1}{|J|} \qquad \text{and} \qquad C_1 = \frac{4C_0}{c_0 c_1}, $$

and may easily verify that

$$
s \leq \frac{3C_0}{c_0 c_1} \log \frac{1}{\delta \Delta},
$$

whence (4.13) holds. $\qquad\square$

Before proving Theorem 1.5, we establish Proposition 4.1.

Proof of Proposition 4.1. First of all, if $0 \in K(I)^\circ$, then the existence of a minimiser follows from Lemma 4.3, and it is easy to see that then $g_I(d) > 0$.

Now suppose that $0 \notin K(I)^\circ$. By (4.14),

(4.16) $$ g_I(d) = \inf_{y \in \mathbb{R}^n} f_I(y, d) = g_{I_1}(d), $$

where either $0 \in K(I_1)^\circ$ or $I_1 = \emptyset$. Now, from the proof of Lemma 4.2, if $0 \in \partial K(I)$ then $0 \in K(I_1)^\circ$ (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 $0 \in K(I)^\circ$. To this end, assume that a minimiser for $f_I(\cdot, d)$ exists and $0 \notin K(I)^\circ$. 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 $0 \in K(I)^\circ$ is necessary for extremisability, and this concludes the proof of Proposition 4.1. $\qquad\square$

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 δ, 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 δ 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 4.4. For nonnegative integers k, we define the disjoint subsets I_k of J by

$$I_k = \{j \in J : \delta^{N_1^{k+1}} \leq d_j < \delta^{N_1^k}\}.$$

Then, by the pigeonhole principle, there exists $k = k_\delta \in \{0, \ldots, |J|\}$ such that I_k is empty. For this k, we introduce

$$I' = \{j \in J : d_j \geq \delta^{N_1^k}\}$$

and observe that, since I_k is empty,

$$J \setminus I' = \{j \in J : d_j \leq \delta^{N_1^{k+1}}\}.$$

Now consider the function $f_{I'}$, where

$$\Delta = \Delta(d, I') = \min_{I \in I'} d_j \geq \delta^{N_1^k}.$$

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

$$(4.17) \qquad |y| \leq C_1(N_1^{|J|} + 2) \log \frac{1}{\delta},$$

and we will show that by choosing N_1 and δ_0 appropriately, y is a δ-near-minimiser of f_J. To see this, we use the trivial estimate (4.12) to obtain

$$f_{J \setminus I'}(y) \leq |J| \delta^{N_1^{k+1}} \delta^{-C_0 C_1 (N_1^k + 2)}$$

and thus, if we choose $N_1 = 4C_0 C_1$, then

$$f_{J \setminus I'}(y) \leq |J| \delta^2.$$

This immediately gives

$$f_J(y, d) \leq g_{I'}(d) + \delta^2 + |J| \delta^2 \leq g_J(d) + \delta,$$

provided that

$$(4.18) \qquad \delta_0 \leq (1 + |J|)^{-1}.$$

Finally, we note that if we choose

$$(4.19) \qquad N = (4C_0 C_1)^{|J|+1},$$

then $|y| \leq N \log(1/\delta)$ from our choice of N_1 and (4.17). $\square$

Remark. A review of the proofs of Lemma 4.4 and Theorem 1.5 reveals that an explicit choice of N and δ_0 for which the claim in Theorem 1.5 holds is given by

$$(4.20) \qquad N = \left(\frac{16 C_0^2}{c_0 c_1}\right)^{|J|+1} \qquad \text{and} \qquad \delta_0 = \frac{1}{|J| + 1},$$

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 $I \subseteq J$,*

$$\text{(4.21)} \qquad |g_I(d) - g_I(d')| \leq (D + |J| + 2D|J|\delta_0^{-1})\|d - d'\|_\infty^\alpha$$

for all $d, d' \in [0, D]^I$. Here δ_0 is given by Theorem 1.5.

Proof. We shall show that the claimed local Hölder continuity holds for $\alpha = (1 + C_0 N)^{-1}$. We begin by observing that (4.21) holds easily if $\|d - d'\|_\infty^\alpha \geq \delta_0$ thanks to the elementary bound $g_I \leq D|J|$ on $[0, D]^I$. Thus it suffices to show that

$$\text{(4.22)} \qquad |g_I(d) - g_I(d')| \leq (D + |J|)\|d - d'\|_\infty^\alpha$$

whenever $\|d - d'\|_\infty^\alpha < \delta_0$.

To prove (4.22), suppose that $0 < \|d - d'\|_\infty < \delta_0^{1 + C_0 N}$ and define $\delta \in (0, \delta_0)$ by

$$\delta = \|d - d'\|_\infty^{1/(1 + C_0 N)}.$$

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

$$|f_I(y, d) - f_I(y, d')| \leq |J|\|d - d'\|_\infty e^{C_0|y|}$$

and our choice of δ, we thus see that

$$\begin{aligned}
g_I(d') - g_I(d) &\leq f_I(y, d') - f_I(y, d) + D\delta \\
&\leq |J|\|d - d'\|_\infty \delta^{-C_0 N} + D\delta \\
&= (D + |J|)\|d - d'\|_\infty^{1/(1 + C_0 N)}.
\end{aligned}$$

By symmetry, (4.22) follows. $\qquad\square$

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 4.5. Naturally, both of these deductions are based on having a suitable expression for $\mathrm{BL}_g(\mathbf{L}, \mathbf{p}; \mathbf{A})$ 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 [15, Theorem 3.1]. To state the result, we need to set up some more notation.

First, fix an admissible Brascamp–Lieb datum $(\mathbf{L}, \mathbf{p})$. Next, given an $n_j \times n_j$ positive definite matrix A_j, we write

$$A_j = R_j^* X_j R_j,$$

where R_j is an $n_j \times n_j$ 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 $\mathcal{I}$ is given by

$$\mathcal{I} = \big\{ I \subseteq \{1, \ldots, M\} : |I| = n \big\}$$

and, for each $I \in \mathcal{I}$, the vector $u_I \in \mathbb{R}^M$ is given by

$$u_I = \mathbb{1}_I - q.$$

Here, $\mathbb{1}_I$ denotes the vector $(\mathbb{1}_I(j))_{j=1}^M \in \mathbb{R}^M$, where $\mathbb{1}_I(j) = 1$ if $j \in I$, and $\mathbb{1}_I(j) = 0$ 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 $I \in \mathcal{I}$, we define

$$\text{(5.1)} \qquad d_I(\mathbf{L}, \mathbf{p}; \mathbf{R}) = q_I \det((v_k)_{k \in I})^2$$

where $v_k = L_j^* R_j^* e_{(j,\ell)}$, $\mathbf{R} = (R_j)_{j=1}^m$ and $q_I = \prod_{k \in I} q_k$. 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, ℓ) via $k = n_0 + \cdots + n_{j-1} + \ell$, where $j \in \{1, \ldots, m\}$, $\ell \in \{1, \ldots, n_j\}$ and $n_0 = 0$.

Proposition 5.1. *With the notation as above,*

$$(5.2) \qquad \mathrm{BL_g}(\mathbf{L}, \mathbf{p}; \mathbf{A})^{-2} = \sum_{I \in \mathcal{I}} d_I(\mathbf{L}, \mathbf{p}; \mathbf{R}) \exp(\langle u_I, y \rangle).$$

Proof. To see (5.2), we follow the proof of [15, 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 kth column is $q_k e^{y_k} v_k$ and the matrix whose kth row is v_k^*, and thus

$$\det \left(\sum_{j=1}^m p_j L_j^* A_j L_j \right) = \sum_{I \in \mathcal{I}} d_I(\mathbf{L}, \mathbf{p}; \mathbf{R}) \exp(\langle \mathbb{1}_I, y \rangle)$$

follows from the Cauchy–Binet formula. The identity (5.2) is now immediate. $\qquad\square$

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

$$f_{\mathcal{I}}(y, d) = \sum_{I \in \mathcal{I}} d_I \exp(\langle u_I, y \rangle)$$

where $y \in \mathbb{R}^M$ and $d = (d_I)_{I \in \mathcal{I}}$ 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 (1.9) holds when the L_j are projections. Then, given any admissible $\mathbf{L}$, 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 $\mathbf{L}'$ is admissible and each L_j' is a projection, and thus (1.9) holds for $\mathbf{L}'$. It may be deduced that (1.9) holds for $\mathbf{L}$ using the elementary fact

$$\mathrm{BL_g}(\mathbf{L}, \mathbf{p}; \mathbf{A})^{-2} = \mathrm{BL_g}(\mathbf{L}', \mathbf{p}; \mathbf{A}')^{-2} \prod_{j=1}^m \det(L_j L_j^*)^{p_j},$$

whenever $\mathbf{A}$ and $\mathbf{A}'$ satisfy $A_j = (C_j^{-1})^* A_j' C_j^{-1}$.

We now prove (1.9) under the assumption that $L_j L_j^* = I_{n_j}$. From Proposition 5.1

$$\mathrm{BL_g}(\mathbf{L}, \mathbf{p}; \mathbf{A})^{-2} = f_{\mathcal{I}}(y, d(\mathbf{L}, \mathbf{p}; \mathbf{R})),$$

where $d(\mathbf{L}, \mathbf{p}; \mathbf{R}) = (d_I(\mathbf{L}, \mathbf{p}; \mathbf{R}))_{I \in \mathcal{I}}$ and therefore, from Lieb's theorem (Theorem 1.2),

$$(5.3) \qquad \mathrm{BL}(\mathbf{L}, \mathbf{p})^{-2} = \inf_{\mathbf{R}} \inf_{y \in \mathbb{R}^M} f_{\mathcal{I}}(y, d(\mathbf{L}, \mathbf{p}; \mathbf{R})) = \inf_{\mathbf{R}} g_{\mathcal{I}}(d(\mathbf{L}, \mathbf{p}; \mathbf{R})).$$

Since $d(\mathbf{L}, \mathbf{p}; \mathbf{R})$ is a continuous function of $\mathbf{R}$ and $g_{\mathcal{I}}(d)$ is a continuous function of d by Theorem 4.5 (this is tantamount to the continuity of the Brascamp–Lieb constant observed in [15]), there exists some $\mathbf{R}$ (depending on $\mathbf{L}$ and $\mathbf{p}$) such that

$$(5.4) \qquad \mathrm{BL}(\mathbf{L}, \mathbf{p})^{-2} = g_{\mathcal{I}}(d(\mathbf{L}, \mathbf{p}; \mathbf{R})).$$

For the remainder of the proof, $\mathbf{R} = (R_j)_{j=1}^m$ 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 $\mathbf{p}$, such that, for all $\delta \in (0, \delta_0)$,

$$\inf_{|y| \leq N \log(1/\delta)} f_{\mathcal{I}}(y, d(\mathbf{L}, \mathbf{p}; \mathbf{R})) \leq g_{\mathcal{I}}(d(\mathbf{L}, \mathbf{p}; \mathbf{R})) + \delta \max_I d_I(\mathbf{L}, \mathbf{p}; \mathbf{R}).$$

From the assumption $L_j L_j^* = I_{n_j}$, it follows that $d_I(\mathbf{L}, \mathbf{p}; \mathbf{R}) \leq 1$ and therefore there exists some $y = y(\delta, \mathbf{L}, \mathbf{p}) \in \mathbb{R}^M$ satisfying

$$(5.5) \qquad\qquad |y| \leq N \log \frac{1}{\delta}$$

and

$$f_{\mathcal{I}}(y, d(\mathbf{L}, \mathbf{p}; \mathbf{R})) \leq g_{\mathcal{I}}(d(\mathbf{L}, \mathbf{p}; \mathbf{R})) + \delta.$$

Hence

$$(5.6) \qquad \mathrm{BL}_g(\mathbf{L}, \mathbf{p}; \mathbf{A})^{-2} = f_{\mathcal{I}}(y, d(\mathbf{L}, \mathbf{p}; \mathbf{R})) \leq g_{\mathcal{I}}(d(\mathbf{L}, \mathbf{p}; \mathbf{R})) + \delta = \mathrm{BL}(\mathbf{L}, \mathbf{p})^{-2} + \delta.$$

Here, $A_j = R_j^* X_j R_j$, 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 $e^{y_{n_1+1}}, \ldots, e^{y_{n_1+n_2}}$, and so forth. From (5.5), it is easily checked that $\|A_j\| \leq (1/\delta)^N$ and $\|A_j^{-1}\| \leq (1/\delta)^N$. Hence (1.9) holds when $L_j L_j^* = I_{n_j}$, and this completes our proof of Theorem 1.3. $\qquad\square$

To end this section, we improve the result in [15], which established the continuity of the Brascamp–Lieb constant, to local Hölder continuity. Given that differentiability of the Brascamp–Lieb constant may fail (see [15]), it seems to be a challenging problem to improve upon the following result.

Theorem 5.2. *For each $\mathbf{p}$, the Brascamp–Lieb constant $\mathrm{BL}(\cdot, \mathbf{p})$ is locally Hölder continuous. More precisely, there exists a number $\alpha \in (0, 1)$ and a constant $C_0 > 0$, depending only on $(n_j)_{j=1}^m$ and $(p_j)_{j=1}^m$, such that the following holds. For any $C > 0$,*

$$(5.7) \qquad |\mathrm{BL}(\mathbf{L}, \mathbf{p})^{-2} - \mathrm{BL}(\mathbf{L}', \mathbf{p})^{-2}| \leq C_0 C^{n+\alpha(n-1)} \|\mathbf{L} - \mathbf{L}'\|^\alpha$$

for all data $\mathbf{L}, \mathbf{L}'$ satisfying the bound $\|\mathbf{L}\|, \|\mathbf{L}'\| \leq C$. In particular, if additionally the Brascamp-Lieb constants $\mathrm{BL}(\mathbf{L}, \mathbf{p})$, $\mathrm{BL}(\mathbf{L}', \mathbf{p})$ are bounded above by a constant C_1, we have

$$(5.8) \qquad |\mathrm{BL}(\mathbf{L}, \mathbf{p}) - \mathrm{BL}(\mathbf{L}', \mathbf{p})| \leq C_0 C_1^3 C^{n+\alpha(n-1)} \|\mathbf{L} - \mathbf{L}'\|^\alpha.$$

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

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

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

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

$$f(x, y, a, b) = a e^x + e^{-x} + e^{x-y} + b e^y \quad \text{and} \quad g(a, b) = \inf\{f(x, y, a, b) : (x, y) \in \mathbb{R}^2\},$$

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 $|y_\delta| \leq 2\log(2/\delta)$.

Case 2: $a \neq 0$ and $b = 0$. In this case,

$$f(x, y, a, 0) = ae^x + e^{-x} + e^{x-y},$$

and evidently

$$g(a, 0) = \min\{ae^x + e^{-x} : x \in \mathbb{R}\} = 2a^{1/2}.$$

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 $x_\delta = -\frac{1}{2}\log(a + \delta^2/4)$ and $y_\delta = -2\log(\delta/2)$. 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

$$(x_\delta, y_\delta) = \begin{cases} (\frac{1}{3}\log(1/b), \frac{2}{3}\log(1/b)) & \text{if } b > (\delta/3)^3 \\ (\log(3/\delta), 2\log(3/\delta)) & \text{if } b \leq (\delta/3)^3. \end{cases}$$

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

$$f(x_\delta, y_\delta, 0, b) - g(0, b) = 2\delta/3 + b(3/\delta)^2 - 3b^{1/3} \leq \delta - 3b^{1/3} < \delta.$$

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_*, y_*), where

$$e^{y_*} = b^{-2/3}\tau(ab^{-2/3})^{-1}$$
$$e^{x_*} = b^{-1/3}\tau(ab^{-2/3})^{-2};$$

here $\tau(c)$ 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_δ, y_δ) 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 $(\mathbf{L}, \mathbf{p})$, as may be seen by a slight modification of the argument in Lemma 6 from [17]; see also [35]. This argument is related to the so-called "transplantation" method — see for example [45]. As may be expected, it is not in general possible to take $\varepsilon = 0$ either, as $\mathrm{BL}(d\mathbf{B}(x), \mathbf{p})$ 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 $(\mathbf{L}, \mathbf{p})$ 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 $\mathbf{p}$, the neighbourhood U may be taken to be locally uniform in $\mathbf{L}$ — that is, given $\mathbf{L}_0$ and $A > 0$, there exists a neighbourhood U of

x_0 such that (1.2) 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 $\mathbf{L}$ is sufficiently close to $\mathbf{L}_0$. 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 $\mathbf{f}$. In order for there to be a base case for this induction it will be convenient to suppose that $\mathbf{f}$ satisfies a certain auxiliary regularity condition. The passage to general integrable $\mathbf{f}$ will consist of an elementary limiting argument.

Definition 6.1. Suppose that $\kappa > 1, \mu > 0$ and Ω is a measurable subset of $\mathbb{R}^d$. A nonnegative function f is κ-constant at scale μ on Ω 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 μ from Ω. 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 μ is taken sufficiently small. One way to see this is to observe that the d-dimensional Poisson kernel $P_t(x)$ is κ-constant at scale μ on the whole of $\mathbb{R}^d$ provided that μ 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 κ to vary in a controlled manner through the induction. This flexibility in κ 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

$$(6.1) \qquad\qquad U_\delta(y) = \{x \in \mathbb{R}^n : |x - y| \leq \delta\}.$$

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

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

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

We think of $\mathcal{C}(u, \delta, \mu, \kappa)$ as a regularised and localised nonlinear Brascamp–Lieb constant; here we are of course suppressing the dependence on $\mathbf{B}$ and $\mathbf{p}$ 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 δ is below a certain threshold in terms of μ, 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_j^u at u, given by

$$L_j^u := B_j(u) + dB_j(u)(\cdot - u).$$

This allows us to reduce the estimation of $\mathcal{C}(u, \delta, \mu, \kappa)$ 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 $\alpha, \beta, \gamma, \tau$ and σ, which should be regarded as fixed, are chosen in this order as follows. First, we fix α satisfying $1 < \alpha < 2$, and then β such that $0 < \beta < 2 - \alpha$. Next, γ is chosen to satisfy

$$(6.2) \qquad\qquad 0 < \gamma < \min\left\{ \frac{2 - \alpha - \beta}{(n+2)\alpha}, \frac{\beta}{2\alpha} \right\},$$

and then τ such that

$$(6.3) \qquad\qquad 0 < \tau < \frac{\gamma}{N},$$

where $N \in \mathbb{N}$ is given by Corollary 1.4. Finally, σ is chosen to satisfy $0 < \sigma < \min\{\beta - 2\alpha\gamma, \alpha\tau\}$. The reader is encouraged not to dwell on these restrictions at this stage. We simply note here that the crucial recursive inequality stated below in Proposition 6.4 compares the function $\mathcal{C}$ at scales δ and δ^α; the roles of the other exponents are of a more technical nature, so we postpone further remarks of this type.

Notation. In this section, we refer to a positive real number c as a *constant* if it depends on at most $(\mathbf{B}, \mathbf{p})$, and the parameters α, β, γ, τ and σ. For $\varepsilon > 0$, we say that $c > 0$ is a *constant depending on ε* if it is a constant that may depend additionally on ε. Finally, we write $A \lesssim B$ and $B \gtrsim A$ to mean that $A \leq cB$, where $c > 0$ is a constant, and $A \sim B$ means that both $A \lesssim B$ and $A \gtrsim B$. The reader may wish to observe that in all instances where such constants occur, they may in fact be chosen to depend locally uniformly on the nonlinear maps $\mathbf{B}$, in line with our remark at the beginning of this section.

Before stating the base case for the induction, we note that by the C^2 regularity[3] of the B_j,

$$(6.4) \qquad |B_j(x) - L_j^u x| \lesssim |x - u|^2$$

for all $x, u \in U_\nu(0)$ (with an appropriately small constant ν). 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*

$$\mathcal{C}(u, \delta, \mu, \kappa) \leq \kappa^\rho \, \mathrm{BL}(d\mathbf{B}(u), \mathbf{p}),$$

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),

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

for all $y \in U_\delta(u)$. Together with our choices that $\alpha + \beta < 2$ and $\delta^{\alpha+\beta} \leq \mu$, this implies that $|B_j(y) - L_j^u y| \leq \mu$, for a possibly smaller choice of the constant ν. Now $f_j \in L^1(B_j(U_{2\delta}(u)); \mu, \kappa)$, and so $f_j(B_j(y)) \leq \kappa f_j(L_j^u y)$ for all $y \in U_\delta(u)$, and so

$$\int_{U_\delta(u)} \prod_{j=1}^m f_j^{p_j}(B_j(y)) \, dy \leq \kappa^\rho \int_{U_\delta(u)} \prod_{j=1}^m f_j^{p_j}(L_j^u y) \, dy$$

$$\leq \kappa^\rho \, \mathrm{BL}(d\mathbf{B}(u), \mathbf{p}) \prod_{j=1}^m \left(\int_{\mathbb{R}^{n_j}} f_j(x_j) \, dx_j \right)^{p_j},$$

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. $\qquad\square$

Our objective is to prove a suitable recursive inequality involving the function $\mathcal{C}$, which upon iteration will establish Theorem 1.1. This inequality, which we now state, may be viewed as a certain nonlinear version of (3.3).

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*

$$(6.5) \qquad \mathcal{C}(u, \delta, \mu, \kappa) \leq (1 + \delta^\sigma) \max_{x \in U_{2\delta}(u)} \mathcal{C}(x, \delta^\alpha, \mu, \kappa \exp(\delta^\sigma)).$$

The reader may wish to view (6.5) as a certain "near-monotonicity property" of the function $\mathcal{C}$ in the parameter δ, since the multiplicative factors $(1 + \delta^\sigma)$ and $\exp(\delta^\sigma)$ are close to 1 for small δ. It is interesting to compare (6.5) with the related near-monotonicity results in [21], [10], [12] and [53].

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

---

[3]This is the main use of the C^2 regularity hypothesis on $\mathbf{B}$, and where the reader may wish to weaken it to $C^{1,\theta}$ for some $\theta > 0$. This would require that the condition $\alpha + \beta < 2$ be tightened to $\alpha + \beta < 1 + \theta$, and the relation $\sigma < \beta < 2 - \alpha$ be replaced with $\sigma < \beta < 1 + \theta - \alpha$. Finally, γ must also satisfy $\gamma < (1 + \theta - \alpha - \beta)/(\alpha(n+2))$.

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 δ_0 depending on ε and a constant $C > 0$ such that

$$(6.6) \qquad \mathcal{C}(0, \delta_0, \mu, 1 + \varepsilon) \leq (1 + C\varepsilon) \operatorname{BL}(d\mathbf{B}(0), \mathbf{p})$$

uniformly in $\mu > 0$. As we observed above, for each $t > 0$, and $f_j \in L^1(2B_j(U_{\delta_0}(0)))$, 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 μ 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

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

uniformly in $t > 0$. Since the nonnegative functions $f_j * P_t^{(n_j)}$ converges 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 δ_0, depending on ε, to satisfy a number of constraints. The first of these is $\delta_0 \leq \nu/3$, where ν 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 ν 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 μ 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 $\mathcal{C}(0, \delta_0, \mu, 1 + \varepsilon)$ with the base case.

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

$$(6.7) \qquad \mathcal{C}(u, \delta_k, \mu, \kappa) \leq (1 + \delta_k^\sigma) \max_{x \in U_{2\delta_k}(u)} \mathcal{C}(x, \delta_{k+1}, \mu, \kappa \exp(\delta_k^\sigma))$$

for all u in the algebraic sum $\widetilde{U}_k := U_{2\delta_0}(0) + U_{2\delta_1}(0) + \cdots + U_{2\delta_{k-1}}(0)$ 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 $\delta_0 \leq \nu$): (6.7) requires that $\widetilde{U}_k \subseteq U_\nu(0)$, which is indeed the case for sufficiently small δ_0 independent of k.

Choosing k_* to be the smallest natural number k for which $\delta_k^{\alpha+\beta} \leq \mu$, we may iterate (6.7) k_* times to obtain

$$\mathcal{C}(0, \delta_0, \mu, 1 + \varepsilon) \leq \prod_{k=0}^{k_*-1} (1 + \delta_k^\sigma) \max_{u \in \widetilde{U}_{k_*}} \mathcal{C}\left(u, \delta_{k_*}, \mu, (1 + \varepsilon) \prod_{k=0}^{k_*-1} \exp(\delta_k^\sigma) \right).$$

Since $\delta_{k_*}^{\alpha+\beta} \leq \mu$ we may now deduce from Proposition 6.3 that

$$(6.8) \qquad \mathcal{C}(0, \delta_0, \mu, 1 + \varepsilon) \leq (1 + \varepsilon)^\rho \max_{u \in \widetilde{U}_{k_*}} \operatorname{BL}(d\mathbf{B}(u), \mathbf{p}) \prod_{k=0}^{k_*-1} (1 + \delta_k^\sigma) \exp(\rho \delta_k^\sigma).$$

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 [15] and [41] is sufficient here, and ensures that

$$\max_{u \in \widetilde{U}_{k_*}} \operatorname{BL}(d\mathbf{B}(u), \mathbf{p}) \leq (1 + \varepsilon) \operatorname{BL}(d\mathbf{B}(0), \mathbf{p})$$

for an appropriate choice of a constant δ_0 (depending on ε).

For the product term in (6.8), note that

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

and so $\prod_{k=0}^{k_*-1} (1 + \delta_k^\sigma) \exp(\rho \delta_k^\sigma) \leq 1 + \varepsilon$, for an appropriate ε-dependent choice of constant δ_0. Putting together the above, we have shown that

$$\mathcal{C}(0, \delta_0, \mu, 1 + \varepsilon) \leq (1 + \varepsilon)^{\rho+3} \operatorname{BL}(d\mathbf{B}(0), \mathbf{p}),$$

which establishes (6.6), and hence Theorem 1.1. $\qquad \square$

It remains to prove Proposition 6.4.

6.2. Gaussian preliminaries.
Before embarking on our proof of Proposition 6.4, we introduce a family of m-tuples of gaussians that will be pivotal in our argument.

As described in Section 3, our proof of Proposition 6.4 may be interpreted as a certain nonlinear version of the proof of the inequality (3.3) for the linear Brascamp–Lieb constant. As the nonlinear Brascamp–Lieb constant $\mathcal{C}(u, \delta, \mu, k)$ from Definition 6.2 involves a localisation to the δ-ball $U_\delta(u)$, the gaussians that we introduce here will be near-extremisers for the underlying linear datum $(d\mathbf{B}(u), \mathbf{p})$, 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 γ satisfying (6.2)). Corollary 1.4 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 δ^τ-near extremisers, where the small exponent τ satisfies (6.3). This subsection clarifies these heuristic ideas, and establishes several key lemmas that will be used in the forthcoming proof of Proposition 6.4.

Note the dual role of δ 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 $d\mathbf{B}(0)$ satisfies the Loomis–Whitney-type kernel condition (1.3). Another special case is when $(d\mathbf{B}(0), \mathbf{p})$ is *simple* — that is, when (1.6) 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 [56]. We refer to [14] for the details of this.

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 δ-near extremisers, for a given datum $(\mathbf{L}, \mathbf{p})$, is invariant under common isotropic scalings of their associated quadratic forms. To be precise, the family of all positive-definite $\mathbf{A} = (A_j)$ satisfying

$$\mathrm{BL_g}(\mathbf{L}, \mathbf{p}; \mathbf{A}) \geq (1 - \delta)\,\mathrm{BL}(\mathbf{L}, \mathbf{p})$$

is invariant under the scaling $\mathbf{A} \mapsto \lambda\mathbf{A}$ for all $\lambda > 0$. Using the elementary fact that

$$(6.9) \qquad\qquad\qquad \mathrm{BL}(\mathbf{L}, \mathbf{p}; \mathbf{f}) = \mathrm{BL_g}(\mathbf{L}, \mathbf{p}; \mathbf{A})$$

for $\mathbf{f}$ satisfying (1.4), we are therefore at liberty to rescale our gaussian near-extremisers without penalty.

By Corollary 1.4, there exists a constant $\nu > 0$ such that for each $u \in U_\nu(0)$ and $\delta \in (0, \nu)$, there exists an m-tuple of positive definite matrices $(A_{u,\delta,j})_{j=1}^m$ satisfying

$$(6.10) \qquad\qquad\qquad\qquad \|A_{u,\delta,j}^{-1}\| < \delta^{-2\tau N},$$

$$(6.11) \qquad\qquad\qquad\qquad \|A_{u,\delta,j}\| \lesssim 1,$$

$$(6.12) \qquad\qquad \mathrm{BL_g}(d\mathbf{B}(u), \mathbf{p}; \mathbf{A}_{u,\delta}) \geq (1 - \delta^\tau)\,\mathrm{BL}(d\mathbf{B}(u), \mathbf{p}).$$

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 γ 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

$$(6.13) \qquad \prod_{j=1}^{m} \exp(-\pi p_j \langle A_{u,\delta,j} dB_j(u)x, dB_j(u)x \rangle) = \exp(-\pi \langle M_{u,\delta} x, x \rangle).$$

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 δ^α to scale δ in Proposition 6.4. 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

$$(6.14) \qquad \sup_{u \in U_\nu(0)} \|M_{u,\delta}^{-1/2}\| \lesssim \delta^{-\tau N}$$

for all $\delta \in (0, \nu)$. To see this, note that if λ is any eigenvector of $M_{u,\delta}$ with corresponding unit eigenvector x in $\mathbb{R}^n$, then

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

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 $\mathrm{BL}(\mathbf{L}, \mathbf{p})$ 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 (6.14) now follows.

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

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

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

$$(6.15) \qquad \mathrm{BL}(d\mathbf{B}(u), \mathbf{p}; \mathbf{g}_{u,\delta}) \geq (1 - \delta^\tau) \, \mathrm{BL}(d\mathbf{B}(u), \mathbf{p}).$$

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 $\mathbf{g}_{u,\delta}$, 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 δ-balls without losing too much.

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

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

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

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

$$(6.16) \qquad \int_{\mathbb{R}^n \setminus U_\delta(0)} \prod_{j=1}^{m} g_{u,\delta,j}^{p_j}(dB_j(u)x) \, dx \lesssim \delta^{2\eta} \int_{\mathbb{R}^n} \prod_{j=1}^{m} g_{u,\delta,j}^{p_j}(dB_j(u)x) \, dx$$

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

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

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

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

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

$$(6.17) \qquad \int_{|x| \gtrsim \delta^{-\gamma + \tau N}} e^{-\pi |x|^2} \, dx \lesssim \delta^{2\eta}.$$

for sufficiently small δ. To see this, we simply observe that

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

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

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 $L_j^{u,y} : \mathbb{R}^n \to \mathbb{R}^{n_j}$ given by

$$L_j^{u,y} x = L_j^u x - B_j(y) = B_j(u) + dB_j(u)(x-u) - B_j(y).$$

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

$$\int_{U_{\delta^\alpha}(y)} \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(dB_j(u)(x-y)) \, dx \le (1 + \delta^\beta) \int_{U_{\delta^\alpha}(y)} \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(L_j^{u,y} x) \, dx.$$

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

$$(6.18) \qquad \int_{U_{\delta^\alpha}(y)} \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(dB_j(u)(x-y)) \, dx - \delta^{\beta'} \le \int_{U_{\delta^\alpha}(y)} \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(L_j^{u,y} x) \, dx,$$

as we may then obtain the desired estimate by a bootstrapping argument. Indeed, by continuity of the Brascamp–Lieb constant [15], (6.12) and Lemma 6.5,

$$\mathrm{BL}(d\mathbf{B}(0),\mathbf{p}) \le 2\mathrm{BL}(d\mathbf{B}(u),\mathbf{p}) \le 4 \int_{U_{\delta^\alpha}(y)} \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(dB_j(u)(x-y)) \, dx,$$

provided that ν is sufficiently small. Therefore (6.18) implies that

$$\left(1 - \frac{4\delta^{\beta'}}{\mathrm{BL}(d\mathbf{B}(0),\mathbf{p})}\right) \int_{U_{\delta^\alpha}(y)} \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(dB_j(u)(x-y)) \, dx \le \int_{U_{\delta^\alpha}(y)} \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(L_j^{u,y}(x)) \, dx,$$

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

In order to show (6.18), we define

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

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

$$(6.19) \qquad \|G \circ L - G \circ \widetilde{L}\|_{L^1(U_{\delta^\alpha}(y))} \le \delta^{\beta'},$$

where $L_j = L_j^{u,y}$ and $\widetilde{L}_j = dB_j(u)(\cdot - y)$, 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 (1.7) to collect powers of δ. Since moreover

$$|L_j x - \widetilde{L}_j x| \lesssim \delta^2$$

from (6.4), the mean value theorem implies that

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

and hence

$$\|G \circ L - G \circ \widetilde{L}\|_{L^1(U_{\delta^\alpha}(y))} \lesssim \delta^{2-\alpha} \left(\delta^{-\gamma}\right)^{\alpha(2+n)}.$$

Our choice of γ ensures that the exponent on the right-hand side of the above estimate is strictly larger than β. Provided the constant $\nu > 0$ is sufficiently small, this implies the existence of $\beta' > \beta$ such that (6.19) holds, and completes the proof. $\qquad\square$

Our final technical lemma shows that appropriately *truncated* gaussians enjoy a local constancy property. We note that gaussians are not κ-constant at scale μ on $\mathbb{R}^n$ for any choice of κ and μ — 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 $x \in U_{2\delta}(u)$, the function $g_{u,\delta^\alpha,j}(L_j^u x - \cdot)$ is $\exp(\delta^\sigma)$-constant at scale μ on $B_j(U_{2\delta^\alpha}(x))$.*

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 μ 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, $g_{u,\delta^\alpha,j}(L_j^u x - \cdot)$ is simply a translation of $g_{u,\delta^\alpha,j}$, so to establish the desired claim, it is only necessary to check that $L_j^u x - w \in U_{C\delta^\alpha}(0)$ for some constant $C > 0$ whenever $w \in B_j(U_{2\delta^\alpha}(x))$. This follows since, given such $w \in B_j(U_{2\delta^\alpha}(x))$, we write $w = B_j(y)$ for some $y \in U_{2\delta^\alpha}(x)$ and use (6.4) to deduce that

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

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

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

by elementary considerations. Applying this fact when $A = A_{u,\delta^\alpha,j}$, $R = C\delta^{-\alpha\gamma}$ and $\widetilde{\mu} = \delta^{-\alpha(1+\gamma)}\mu$, we see that $g_{u,\delta^\alpha,j}$ is $\exp(\eta)$-constant at scale μ 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 $\widetilde{\mu} \leq R$; 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 σ, 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 μ on $U_{C\delta^\alpha}(0)$, for any $u \in U_\nu(0)$, and thus completes the proof of the lemma. $\qquad\square$

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

$$(6.20) \qquad \mathcal{C}(u, \delta, \mu, \kappa) \leq (1 + \delta^\beta)(1 + 2\delta^{\alpha\tau})^2 \max_{x \in U_{2\delta}(u)} \mathcal{C}(x, \delta^\alpha, \mu, \kappa \exp(\delta^\sigma)).$$

Since $\sigma < \min\{\beta, \alpha\tau\}$, 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 implicitly imposing its own smallness requirement on ν.

Proof of (6.20). We fix $u \in U_\nu(0)$, and assume that $\delta \in (0, \nu)$ and μ satisfy the threshold condition $\delta^{\alpha+\beta} > \mu$. We also fix $f_j \in L^1(B_j(U_{2\delta}(u)); \mu, \kappa)$ and, to shorten some forthcoming expressions, we write $F = \prod_{j=1}^m f_j^{p_j} \circ B_j$.

By (6.15),

$$(6.21) \qquad \mathrm{BL}(\mathrm{d}\mathbf{B}(u), \mathbf{p}; \mathbf{g}_{u,\delta^\alpha}) \geq (1 - \delta^{\alpha\tau})\,\mathrm{BL}(\mathrm{d}\mathbf{B}(u), \mathbf{p}),$$

and so, for a sufficiently small constant $\nu > 0$,

$$\int_{U_\delta(u)} F(y)\, dy \leq \frac{1 + 2\delta^{\tau\alpha}}{\mathrm{BL}(\mathrm{d}\mathbf{B}(u), \mathbf{p})} \int_{U_\delta(u)} F(y) \int_{\mathbb{R}^n} \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(dB_j(u)x)\, dx\, dy.$$

By Lemma 6.5 with $\eta = \tau$, followed by a translation change of variables in x,

$$\int_{U_\delta(u)} F(y)\,dy \leq \frac{(1+2\delta^{\tau\alpha})^2}{\mathrm{BL}(\mathrm{d}\mathbf{B}(u),\mathbf{p})} \int_{U_\delta(u)} F(y) \int_{U_{\delta^\alpha}(y)} \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(dB_j(u)(x-y))\,dx\,dy.$$

Now we may apply Lemma 6.6, and then Fubini's theorem, to deduce that

$$\int_{U_\delta(u)} F(y)\,dy \leq \frac{(1+\delta^\beta)(1+2\delta^{\tau\alpha})^2}{\mathrm{BL}(\mathrm{d}\mathbf{B}(u),\mathbf{p})} \int_{U_\delta(u)} \int_{U_{\delta^\alpha}(y)} F(y) \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(L_j^{u,y}x)\,dx\,dy$$

$$\leq \frac{(1+\delta^\beta)(1+2\delta^{\tau\alpha})^2}{\mathrm{BL}(\mathrm{d}\mathbf{B}(u),\mathbf{p})} \int_{U_\delta(u)+U_{\delta^\alpha}(0)} \int_{U_{\delta^\alpha}(x)} F(y) \prod_{j=1}^m g_{u,\delta^\alpha,j}^{p_j}(L_j^{u,y}x)\,dy\,dx.$$

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

$$\int_{U_{\delta^\alpha}(x)} \prod_{j=1}^m \left(h_j^x\right)^{p_j}(B_j(y))\,dy$$

where

$$h_j^x(w) = f_j(w)g_{u,\delta^\alpha,j}(L_j^u x - w).$$

By assumption, $f_j \in L^1(B_j(U_{2\delta}(u)); \mu, \kappa)$. Further, if $x \in U_\delta(u) + U_{\delta^\alpha}(0)$, then $U_{2\delta^\alpha}(x) \subseteq U_{2\delta}(u)$, whence (after restriction) $f_j \in L^1(B_j(U_{2\delta^\alpha}(x)); \mu, \kappa)$. Moreover, from Lemma 6.7,

$$g_{u,\delta^\alpha,j}(L_j^u x - \cdot) \in L^1(B_j(U_{2\delta^\alpha}(x)); \mu, \exp(\delta^\sigma)),$$

and therefore

$$h_j^x \in L^1(B_j(U_{2\delta^\alpha}(x)); \mu, \kappa\exp(\delta^\sigma))$$

whenever $x \in U_\delta(u) + U_{\delta^\alpha}(0)$.

Using the definition of $\mathcal{C}(x, \delta^\alpha, \mu, \kappa\exp(\delta^\sigma))$, and slightly enlarging the domain of integration in x, we see that

$$\int_{U_\delta(u)} F(x)\,dx \leq \frac{(1+\delta^\beta)(1+2\delta^{\tau\alpha})^2 M}{\mathrm{BL}(\mathrm{d}\mathbf{B}(u),\mathbf{p})} \int_{U_{2\delta}(u)} \prod_{j=1}^m \left(\int_{\mathbb{R}^{n_j}} h_j^x(y)\,dy\right)^{p_j}\,dx$$

$$= \frac{(1+\delta^\beta)(1+2\delta^{\tau\alpha})^2 M}{\mathrm{BL}(\mathrm{d}\mathbf{B}(u),\mathbf{p})} \int_{U_{2\delta}(u)} \prod_{j=1}^m (f_j * g_{u,\delta^\alpha,j})^{p_j}(L_j^u x)\,dx,$$

where $M = \max_{x \in U_{2\delta}(u)} \mathcal{C}(x, \delta^\alpha, \mu, \kappa\exp(\delta^\sigma))$. To conclude, we simply apply the linear inequality and the fact that the gaussians $g_{u,\delta^\alpha,j}$ are normalised in L^1 to obtain

$$\int_{U_\delta(u)} F(x)\,dx \leq (1+\delta^\beta)(1+2\delta^{\tau\alpha})^2 M \prod_{j=1}^m \left(\int_{\mathbb{R}^{n_j}} f_j * g_{u,\delta^\alpha,j}(x + B_j(u))\,dx\right)^{p_j}$$

$$= (1+\delta^\beta)(1+2\delta^{\tau\alpha})^2 M \prod_{j=1}^m \left(\int_{\mathbb{R}^{n_j}} f_j(x_j)\,dx_j\right)^{p_j}$$

and (6.20) follows. $\square$

REFERENCES

[1] 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.

[2] 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.

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

[4] 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.

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

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

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

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

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

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

[11] 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.

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

[13] J. Bennett and N. Bez, *Some manifestations of transversality in multilinear harmonic analysis*, to appear.

[14] J. Bennett, N. Bez, S. Buschenhenke and T. C. Flock, *The nonlinear Brascamp–Lieb inequality for simple data*, arXiv:1801.05214.

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

[16] 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.

[17] J. Bennett, N. Bez and S. Gutiérrez, *Global nonlinear Brascamp–Lieb inequalities*, J. Geom. Anal. **23** (2013), 1806–1817.

[18] J. Bennett, N. Bez and S. Gutiérrez, *Transversal multilinear Radon-like transforms: local and global estimates*, Rev. Mat. Iberoam. **29** (2013), 765–788.

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

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

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

[22] J. Bennett, A. Carbery and J. Wright, *A nonlinear generalisation of the Loomis–Whitney inequality and applications*, Math. Res. Lett. **12** (2005), 10001–10015.

[23] J. F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, Springer Series in Operations Research, 2000.

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

[25] 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.

[26] J. Bourgain, C. Demeter 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.

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

[28] R. Bramati, *Brascamp–Lieb inequalities on compact homogeneous spaces*, Anal. Geom. Metr. Spaces **7** (2019), 130–157.

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

[30] R. Brown, *Estimates for the scattering map associated with a two-dimensional first-order system*, J. Nonlinear Sci. **11** (2001), 459–471.

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

[32] 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.

[33] 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.

[34] M. Christ, *Convolution, curvature, and combinatorics: a case study*, Int. Math. Res. Not. **19** (1998), 1033–1048.

[35] M. G. Cowling, A. Martini, D. Müller and J. Parcet, *The Hausdorff-Young inequality on Lie groups*, Math. Ann. **375** (2019), no. 1-2, 93–131.

[36] 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.

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

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

[39] 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.

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

[41] 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.

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

[43] S. Herr and S. Kinoshita, *Subcritical well-posedness results for the Zakharov–Kuznetsov equation in dimension three and higher*, arXiv:2001.09047.

[44] H. Hirayama and S. Kinoshita, *Sharp bilinear estimates and its application to a system of quadratic derivative nonlinear Schrödinger equations*, arXiv:1802:06563.

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

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

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

[48] 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.

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

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

[51] B. Stovall, *L^p-improving multilinear Radon-like transforms*, Rev. Mat. Iberoam. **27** (2011), 1059–1085.

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

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

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

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

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

[57] R. Zhang, *The endpoint perturbed Brascamp–Lieb inequality with examples*, Anal. PDE **11** (2018), 555–581.

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.
    Current address: Department of Mathematics, Statistics, and Computer Science, Macalester College, 1600 Grand Ave, St Paul MN 55105 USA.
    E-mail address: tflock@macalester.edu