CS229r: Duality between Brascamp-Lieb Inequality & Subadditivity of Entropy

Juspreet Singh Sandhu and Prayaag Venkat

1 Introduction

The Brascamp-Lieb (BL) inequality is a general inequality which captures many well-known inequalities as special cases.

Theorem 1 (Brascamp-Lieb Inequality, general form [BCCT08]). *Let $\{B_i\}_{i=1}^m$ be a collection of surjective linear maps $B_i : \mathbb{R}^n \to \mathbb{R}^{n_i}$ and $\{c_i\}_{i=1}^m$ be a collection of positive scalars such that:*

$$\sum_{i=1}^{m} c_i n_i = n. \tag{1}$$

Then for any collection $\{f_i\}_{i=1}^m$ of non-negative functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}_{\geq 0}$ such that $\int_{x \in \mathbb{R}^{n_i}} |f_i(x)| dx < \infty$, the following inequality holds

$$\int_{x \in \mathbb{R}^n} \prod_{i=1}^{m} f_i(B_i x)^{c_i} dx \leq D \prod_{i=1}^{m} \left(\int_{x \in \mathbb{R}^{n_i}} f_i(x) dx \right)^{c_i}, \tag{2}$$

where D is a constant that depends only on $\{B_i\}_{i=1}^m$ and $\{c_i\}_{i=1}^m$.

In Theorem 1, $\{B_i\}_{i=1}^m$ are to be thought of as projections of $\mathbb{R}^n$ onto lower dimensional spaces. Roughly, this inequality allows one to upper bound the integral of a product of functions on lower-dimensional spaces by the product of integrals. Condition 1 is a condition which ensures that the "dimensions" on either side of the inequality "match". In Inequality 2, the constant D has been explictly characterized as the solution of a certain optimization problem involving $\{B_i\}_{i=1}^m$ and $\{c_i\}_{i=1}^m$. In the following, we will give several examples to provide intuition and illustrate the key aspects of Theorem 1.

Hölder's Inequality Pick $m = 2$, $c_1 = \frac{1}{p_1}, c_2 = \frac{1}{p_2}$ for $p_1, p_2 \geq 1$ and let $u, v : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ be functions such that $\int_{x \in \mathbb{R}^n} u(x)^{p_1} dx, \int_{x \in \mathbb{R}^n} v(x)^{p_2} dx < \infty$. Set $f_1 = u^{p_1}$ and $f_2 = v^{p_2}$ and take both the projections B_1, B_2 to be the identity map. Then, Condition 1 is easily seen to be satisfied when

$$\frac{1}{p_1} + \frac{1}{p_2} = 1$$

and Inequality 2 simplifies to Hölder's inequality:

$$|\langle f, g \rangle| \leq \|f\|_{p_1} \|g\|_{p_2}.$$

This example illustrates the flexibility of Condition 1; it allows us to choose c_i in such a way so as to control the left-hand side of Inequality 2 by different norms of the functions. A well-known special-case of Holder's inequality is the *Cauchy-Schwarz Inequality*, which can be retreived by setting $p_1 = p_2 = 2$.

Loomis-Whitney Inequality Pick $m = n$, $c_1 = \dots c_n = \frac{1}{n-1}$, and $B_i : \mathbb{R}^n \to \mathbb{R}^{n-1}$ be the projection onto the ith coordinate hyperplane (meaning $B_i(x_1, \dots, x_n) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$), for $i = 1, \dots, n$. Again, Condition 1 is straightforward to check. For $i = 1, \dots, n$, take any $f_i : \mathbb{R}^{n-1} \to \mathbb{R}_{\geq 0}$ such that $\int_{x \in \mathbb{R}^{n-1}} f_i(x)^{n-1} dx < \infty$. Then, Inequality 2 applied to the functions f_i^{n-1} simplifies to the Loomis-Whitney inequality:

$$\int_{x \in \mathbb{R}^n} \prod_{i=1}^{m} f_i(B_i(x)) dx \leq \prod_{i=1}^{m} \|f_i\|_{n-1}.$$

We remark that this is a (continuous version of) a special case of Shearer's Lemma, see 7. Shearer's Lemma itself can be recovered through different choices of $\{B_i\}$ and $\{c_i\}$. This example illustrates two important features of Theorem 1. First, we prove statements about a collection $\{f_1, \dots, f_n\}$ of many functions (instead of just $m = 2$, as in Hölder). Second, we can choose the projections $\{B_i\}_{i=1}^m$ to replace an integral over $\mathbb{R}^n$ with integrals over $\mathbb{R}^{n_i}$, the latter of which may be easier to evaluate in applications. Moreover, the $\{B_i\}$ and $\{c_i\}$ do not have to all be the same.

To get a better understanding of what the Loomis-Whitney inequality is saying, let $S \subset \mathbb{R}^n$ be a measurable set and take $f_i = \mathbb{I}_{B_i(S)}^{1/(n-1)}$ to be the indicator function (raised to the $1/(n-1)$ power) for the projection of S onto the ith coordinate plane. Note that if $x \in S$ then $\prod_{i=1}^n f_i(B_i(x)) = 1$. That is, if a point is in the set then all of its $(n-1)$-dimensional projections are in the $(n-1)$-dimensional projections of the set. This means, that the left-hand side of the Loomis-Whitney inequality is an upper bound on $vol(S)$, where $vol(\cdot)$ denotes the volume of a set in $\mathbb{R}^n$. Along the same vein, considering each of the terms on the right-hand side, we have that

$$\|f_i\|_{n-1} = \left(\int_{x \in \mathbb{R}^{n-1}} \mathbb{I}_{B_i(S)} \right)^{1/(n-1)} = (vol(B_i(S)))^{1/(n-1)}.$$

Substituting, the Loomis-Whitney inequality then states that:

$$vol(S) \leq \prod_{i=1}^{n} vol(B_i(S))^{1/(d-1)}.$$

In other words, we obtain an upper bound on the volume of a set in terms of the volumes of its lower-dimensional projections. For example, when $n = 2$, this says that the area of a set is at most its length (along the vertical y-axis) times its width (along the horizontal x-axis). The above inequality generalizes this to higher dimensions.

This example illustrates how Theorem 1 is often applied; the functions $f_i \circ B_i$ are taken to be indicator functions of lower dimensional projections a set, so that the left-hand side of Inequality 2 represents the volume of the set and the right-hand side involves volumes of the lower-dimensional projections. See the lecture notes of Ball [B+97] for further applications of this technique in convex geometry.

Application: Counting triangles in large random graphs We now present another special case of the BL inequality that was recently used by Lubetzy and Zhao [LZ15] to analyze the number of triangles appearing in large random graphs. Specifically, one of the technical lemmas they proved states that if $f : [0,1]^2 \to [0,1]$ is a measurable function such that $f(x,y) = f(y,x)$, then:

$$\left| \int_{(x,y,z) \in [0,1]^3} f(x,y) f(y,z) f(z,x) dx dy dz \right| \leq \left(\int_{(x,y) \in [0,1]^2} f(x,y)^2 dx dy \right)^{3/2}. \tag{3}$$

Roughly speaking, the triple of pairs $(x, y), (y, z), (z, x)$ corresponds to a triangle in a graph and the quantity on the left-hand side of 3 represents the density of triangles in a certain random graph associated with the function f. This inequality allows one to estimate the number of triangles in this graph based on the number of edges, which is a simpler quantity to understand in the setting of [LZ15].

To solve the problem of counting larger subgraphs than triangles, they derive a generalized version of inequality 3. However, it is straightforward to see that inequality 3 is just a special case of the Loomis-Whitney inequality and the generalized version is just an extension of Hölder's Inequality, which is again a special case of Theorem 1.

1.1 Past work on the Brascamp-Lieb inequalities

In this section, we provide a brief summary of the history of the BL inequality. Originally proven in various special cases by Lieb [Lie02], Brascamp, Lieb and Luttinger [BLL74] and Brascamp and Lieb [BL76], the general form of the BL-inequality given in Theorem 1 is due to Bennett et al. [BCCT08]. Inequalities of this type have application in mathematical physics [CLL05], convex geometry [B$^+$97] and beyond.

We briefly mention an interesting connection between the BL inequality and theoretical computer science. As discussed previously, the constant D in Inequality 2 was characterized in [BCCT08] as the solution of an optimization problem. Recently, Garg et al. [GGOW18] studied the algorithmic task of computing D, given the projections $\{B_i\}_{i=1}^m$ and constants $\{c_i\}_{i=1}^m$. Although all special cases of Theorem 1 considered in this report satisfy the inequality with $D = 1$, until the work of Garg et al., it was not known how to compute D in general in polynomial time.

2 Subadditivity of entropy

We begin by introducing the basic subadditivity of entropy inequality and then stating a slight generalization of it (Han's Inequality (5)), yielding a connection to Shearer's Lemma (7). We end with a general statement about the subadditivity of entropy (8).

Theorem 2 (Basic Subadditivity of Entropy). *For random variables $X_1, .., X_n$, we have the following:*

$$H(X_1, .., X_n) \leq \sum_{i=1}^{n} H(X_i) \tag{4}$$

This statement is a simple result that can be derived from the chain rule of entropy and the fact that conditioning can only descrease entropy. Equality in 4 holds iff $X_i \perp X_j$, $\forall i, j \in [n]$. This statement hints at a connection between a **global** quantity (*Joint Entropy*) and a **local** quantity (*projected entropies*).

We now state Han's Inequality (5), which is a slight generalization of the statement above and is the key ingredient in the special case of a version of Shearer's Lemma (7).

Hanś Inequality Han's inequality allows us to control the joint entropy of a system via the marginal entropies:

$$H(X_1, .., X_n) \leq \frac{1}{n-1} \sum_{i=1}^{n} H(X_1, .., X_{i-1}, X_{i+1}, .., X_n) \tag{5}$$

- Han's inequality can be proved by rewriting the joint entropy as a conditional sum of entropies (for some $i \in [n]$), dropping conditioning on all coordinates $> i$, repeating this $\forall i$, and then summing up the terms and applying the chain-rule for entropy.

- In the spirit of generalizing further, Han's inequality can be applied to relative entropies as well. Given a finite alphabet $\mathcal{X}$, a product distribution $P = \prod_{i=1}^{n} p_i$ on $\mathcal{X}^n$, and a joint distribution Q on $\mathcal{X}^n$, it holds that:

$$D_{KL}(Q \| P) \geq \frac{1}{n-1} \sum_{i=1}^{n} D_{KL}(Q^{(i)} \| P^{(i)}) \tag{6}$$

  where $P^{(i)}, Q^{(i)}$ denote the marginal distributions for P, Q respectively.

Shearer's Lemma & Han's Inequality *Shearer's Lemma* is a statement that allows us to bound the volume of a discrete high-dimensional object by looking at its low-dimensional projections.

Theorem 3 (**Shearer's Lemma**). *Given some n, k, $N \in \mathbb{Z}^+$ where $k \leq n$, an object $F = (f_1, .., f_n) \subseteq [N]^n$, and the k-dimensional projections $F_S = \{(f_{i_1}, .., f_{i_k}) \mid i_j \in S,\ S \subset [n]\}$, the following inequality holds:*

$$|F|^{\binom{n-1}{k-1}} \leq \prod_{S \subset [n], |S| = k} |F_S| \tag{7}$$

If we reformulate Shearer's Lemma by replacing the volume terms $|\cdot|$ in 7 by their entropic counterparts, then we recover Han's Inequality 5 when $k = n-1$, and Basic Subadditivity of Entropy 4 when $k = 1$. This hints at a connection between subadditivity of entropy and problems in geometry that involve estimating high-dimensional volumes by looking at low-dimensional projections. Since the statement of the *Brascamp-Lieb inequality* (2) can be inuitively interpreted as trying to bound the volume of a high-dimensional continuous object via its low-dimensional projections, it is natural to question whether subadditivity of entropy (in its general form) is related to 2. A general form of subadditivity of entropy, which subsumes 4, 5 and 6 is:

Theorem 4 (**Subadditivity of Entropy**, [BLM13]). *Given independent random variables $X_1, .., X_n$ taking values in a countable set $\mathcal{X}$ with the product distribution:*

$$P = \prod_{i=1}^{n} p_i$$

and $Z = f : \mathcal{X}^n \to [0, \infty)$ a positive, bounded random variable, the following inequality holds:

$$Ent[Z] \leq \mathbb{E}_P[\sum_{i=1}^{n} Ent^{(i)}[Z]]^{1} \tag{8}$$

This equation has a similar skeletal structure to 4, and basically asserts that the joint entropy of the system is bounded from above by the average of the sum of the marginal entropies. The generalized $Ent[Z]$ term allows us to control higher moments, forming the basis of the *Entropy Method* for proving concentration inequalities.[2]

[1] $Ent[Z]$ is a generalization of entropy and $Ent^{(i)}[Z]$ is a marginal of the generalization. We define $Ent[Z] = \mathbb{E}[Z \log(Z)] - \mathbb{E}[Z] \log(\mathbb{E}[Z])$, and the corresponding marginal: $Ent^{(i)}[Z] = \mathbb{E}^{(i)}[Z \log(Z)] - \mathbb{E}^{(i)}[Z] \log(\mathbb{E}^{(i)}[Z])$. Refer to [BLM13] for a complete treatment.

[2] To see how 8 can tigthen a concentration inequality, refer to Chapter-4 of [BLM13]. For the extension of subadditivity to more general entropies, refer to Chapter-14 in [BLM13].

3 Duality of subadditivity of entropy and Brascamp-Lieb Inequality

While there are several proofs of Theorem 1, the first of which was due to Bennett et al. [BCCT08], we now outline an alternative approach due to Carlen and Cordero-Erausquin [CCE09]. Among other things, they show that Inequality 2 is dual (in a precise sense to be explained below) to a continuous version of 8 (see 9). This exactly quantifies and generalizes the connection we briefly saw in Section 2.

3.1 Statement of the main result

We now give one of the main results of Carlen and Cordero-Erausquin, which establishes a duality between the BL inequality and subadditivity of entropy. For a probability density function $f : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with $\int_{x \in \mathbb{R}^n} f(x)dx = 1$, we can define the entropy as

$$S(f) = \int_{x \in \mathbb{R}^n} f(x) \log f(x) dx.$$

In fact, this is actually the *negative* entropy, but it will be simpler to work with this version. So, as we will see, "subadditivity" of this definition of entropy really means superadditivity. For a projection $B_i : \mathbb{R}^n \to \mathbb{R}^{n_i}$, we use f_{B_i} to denote the marginal probability density function of f under the projection B_i[3]. With these definitions, we can state the main result.

Theorem 5 (Theorem 2.1 of [CCE09])**.** *Let $\{B_i\}_{i=1}^m$ and $\{c_i\}_{i=1}^m$ be as in Theorem 1. Then the following statements are equivalent:*

1. *(BL inequality) For any collection $\{f_i\}_{i=1}^m$ of non-negative functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}_{\geq 0}$ such that $\int_{x \in \mathbb{R}^{n_i}} |f_i(x)|dx < \infty$, Inequality 2 holds.*
2. *(Subadditivity of entropy) For any probability density function $f : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with $\int_{x \in \mathbb{R}^n} f(x)dx = 1$ and $S(f) < \infty$, the following inequality holds:*

$$\sum_{i=1}^{m} c_i S(f_{B_i}) \leq S(f) + \log D. \tag{9}$$

Having stated Theorem 5, we can see that both Inequalities 2 and 9 are statements that relate integrals of a function to integrals of its marginals. Moreover, note that Theorem 5 preserves the constant D when moving between 2 and 9. Carlen and Cordero-Erausquin [CCE09] exploit this fact to characterize the functions which make the BL inequality tight.

3.2 Dual relationship

We can connect entropy and the BL inequality through the following formula from [CCE09].

Theorem 6. *Let f be a probability density function on $\mathbb{R}^n$ and $\phi : \mathbb{R}^n \to \mathbb{R}$ be any function such that $\int_{x \in \mathbb{R}^n} e^{\phi(x)}dx < \infty$. Then, the following holds:*

$$\log \left(\int_{x \in \mathbb{R}^n} e^{\phi(x)} dx \right) \geq \langle f, \phi \rangle - S(f). \tag{10}$$

The quantity on the left-hand side of Inequality 10 is known in other contexts as the logarithm of the moment generating function (or the cumulant generating function) of ϕ. Inequality 10 simply says that the dual[4] of entropy $S(f)$ is the logarithm of the moment generating function.

[3] This means that if the random variable X is distributed according to f, then $B_i X$ is distributed according to f_{B_i}.

[4] Formally, this is called *Fenchel-Legendre duality*, a well-studied notion on its own right. See [Roc15] for a reference.

Proof. Our starting point is the relative entropy between ϕ and f, which can be simplified to get the right-hand side of the inequality we want to prove:

$$\int_{x\in\mathbb{R}^n} f(x)\log\left(\frac{e^{\phi(x)}}{f(x)}\right)dx = \langle f,\phi\rangle - S(f),$$

where we have used the inner product $\langle f,\phi\rangle = \int_{x\in\mathbb{R}^n} f(x)g(x)dx$. By the concavity of the logarithm and Jensen's inequality, we can upper bound the relative entropy as:

$$\int_{x\in\mathbb{R}^n} f(x)\log\left(\frac{e^{\phi(x)}}{f(x)}\right)dx \le \log\left(\int_{x\in\mathbb{R}^n} f(x)\frac{e^{\phi(x)}}{f(x)}dx\right) = \log\left(\int_{x\in\mathbb{R}^n} e^{\phi(x)}dx\right).$$

Combining the previous inequality with the original equality gives the result. □

3.3 Proof of Theorem 5

The proof strategy for Theorem 5 is simple. We will associate the BL inequality with the left-hand side of Formula 10 and Inequality 9 with the right-hand side, allowing us to pass back and forth between two.

Proof. For the first direction, assume that Inequality 2 holds and let f be any probability density function on $\mathbb{R}^n$ with $S(f) < \infty$. Instantiating the duality formula 10 with $\phi = \sum_{i=1}^m c_i \log f_{B_i} \circ B_i$ and rearranging, we have that:

$$S(f) \ge \sum_{i=1}^m c_i \left(\int_{x\in\mathbb{R}^{n_i}} f_{B_i}(x)\log f_{B_i}(x)dx\right) - \log\left(\int_{x\in\mathbb{R}^n}\prod_{i=1}^m f_{B_i}(B_i x)^{c_i}\right).$$

On the right-hand side of the above inequality, we apply the BL inequality to control the integral of products so that we get

$$\int_{x\in\mathbb{R}^n}\prod_{i=1}^m f_{B_i}(B_i x)^{c_i} \le D\prod_{i=1}^m\left(\int_{x\in\mathbb{R}^{n_i}} f_{B_i}(x)dx\right)^{c_i} = D,$$

where we used the fact that each f_{B_i} is a density function which integrates to 1. Substituting this into the above inequality, we have:

$$S(f) \ge \sum_{i=1}^m c_i \left(\int_{x\in\mathbb{R}^{n_i}} f_{B_i}(x)\log f_{B_i}(x)dx\right) - \log D = \sum_{i=1}^m c_i S(f_{B_i}) - \log D,$$

completing the first direction of the proof. The reverse direction is similar, so we omit the details. □

The proof illustrates that the logarithm of the moment generating function makes it easier to work with marginals, since the exponential of a sum factors into a product of exponentials.

4 Conclusion

In this report, we saw that the BL inequality is a general tool for bounding integrals of products of functions by products of integrals, which arise in various applications.

The main takeaway of Theorem 5 is that we can now prove BL-type inequalities by making use of tools from information theory to prove the corresponding subadditivity of entropy inequalities. In some cases, the latter task is significantly easier (as demonstrated by Carlen and Cordero-Erausquin [CCE09]). As a simple example of this, instantiate Theorem 5 with $\{B_i\}_{i=1}^n$ as the projections onto the n coordinate hyperplanes. Then Inequality 9 becomes Han's Inequality 5 and Inequality 2 becomes the Loomis-Whitney Inequality.

Acknowledgements

The authors would like to thank Professor Madhu Sudan for his feedback on this report and corresponding presentation.

References

[B+97] Keith Ball et al. An elementary introduction to modern convex geometry. *Flavors of geometry*, 31:1–58, 1997.

[BCCT08] Jonathan Bennett, Anthony Carbery, Michael Christ, and Terence Tao. The brascamp–lieb inequalities: finiteness, structure and extremals. *Geometric and Functional Analysis*, 17(5):1343–1415, 2008.

[BL76] Herm Jan Brascamp and Elliott H Lieb. Best constants in young's inequality, its converse, and its generalization to more than three functions. *Advances in Mathematics*, 20(2):151–173, 1976.

[BLL74] Herm J Brascamp, Elliott H Lieb, and JM Luttinger. A general rearrangement inequality for multiple integrals. *Journal of functional analysis*, 17(2):227–237, 1974.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[CCE09] Eric A Carlen and Dario Cordero-Erausquin. Subadditivity of the entropy and its relation to brascamp–lieb type inequalities. *Geometric and Functional Analysis*, 19(2):373–405, 2009.

[CLL05] Eric Carlen, Elliott H Lieb, and Michael Loss. An inequality of hadamard type for permanents. *arXiv preprint math/0508096*, 2005.

[GGOW18] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. Algorithmic and optimization aspects of brascamp-lieb inequalities, via operator scaling. *Geometric and Functional Analysis*, 28(1):100–145, 2018.

[Lie02] Elliott H Lieb. Gaussian kernels have only gaussian maximizers. In *Inequalities*, pages 595–624. Springer, 2002.

[LZ15] Eyal Lubetzky and Yufei Zhao. On replica symmetry of large deviations in random graphs. *Random Structures & Algorithms*, 47(1):109–146, 2015.

[Roc15] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.