

Improved log-Sobolev inequalities, hypercontractivity and uncertainty principle on the hypercube

Yury Polyanskiy^a, Alex Samorodnitsky^b

^a*Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139 USA.*

^b*School of Engineering and Computer Science, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.*

Abstract

Log-Sobolev inequalities (LSIs) upper-bound entropy via a multiple of the Dirichlet form (i.e. norm of a gradient). In this paper we prove a family of entropy-energy inequalities for the binary hypercube which provide a non-linear comparison between the entropy and the Dirichlet form and improve on the usual LSIs for functions with small support. These non-linear LSIs, in turn, imply a new version of the hypercontractivity for such functions. As another consequence, we derive a sharp form of the uncertainty principle for the hypercube: a function whose energy is concentrated on a set of small size, and whose Fourier energy is concentrated on a small Hamming ball must be zero. The tradeoff between the sizes that we derive is asymptotically optimal. This new uncertainty principle implies a new estimate on the size of Fourier coefficients of sparse Boolean functions. We observe that an analogous (asymptotically optimal) uncertainty principle in the Euclidean space follows from the sharp form of Young's inequality due to Beckner. This hints that non-linear LSIs augment Young's inequality (which itself is sharp for finite groups).

Keywords: Hamming space, log-Sobolev inequality, hypercontractivity, Fourier transform on the hypercube, uncertainty principle, coding theory, Boolean functions

Contents

1	Introduction	2
1.1	Definitions, background	2
1.2	Motivation and Organization	4
2	Non-linear log-Sobolev inequalities	5

Email addresses: yp@mit.edu (Yury Polyanskiy), salex@cs.huji.ac.il (Alex Samorodnitsky)

3	New LSIs and hypercontractivity for the hypercube	11
4	Uncertainty principle on the hypercube	15
4.1	Background	15
4.2	Sharp uncertainty principle on \mathbb{F}_2^n	16
4.3	Discussion	19
4.4	A similar result for Euclidean space	21
5	Application: lower bound on spectrum of sparse Boolean functions	24
5.1	Restatement as a property of linear maps (coding theory)	25
5.2	Method 1 – graph covers	26
5.3	Method 2 – analytic	27

1. Introduction

1.1. Definitions, background

We introduce some standard notions for continuous-time semigroups on finite state spaces, e.g. [4, Section 1.7.1]. Consider a finite alphabet \mathcal{X} and a matrix $(L_{x,y})_{x,y \in \mathcal{X}}$ such that 1) $L_{x,y} \geq 0$ for $x \neq y$; and 2) $\sum_{y \in \mathcal{X}} L_{x,y} = 0$ for all x . Then $T_t = e^{tL}$ is a stochastic semigroup, for which we assume that π is a stationary measure. We define $\|f\|_p \triangleq \mathbb{E}^{\frac{1}{p}}[|f|^p]$ and $(f, g) = \mathbb{E}[fg]$ with expectation over π . The Dirichlet form of semigroup T_t is

$$\mathcal{E}(f, g) \triangleq - \sum_{x,y} L_{x,y} f(y)g(x)\pi(x) = \mathbb{E}_\pi[(-Lf)g].$$

We also define $T_t^{\otimes n}$ – a product semigroup on \mathcal{X}^n – and notice that its Dirichlet form is given by

$$\mathcal{E}_n(f, g) \triangleq \sum_{k=1}^n \sum_{x_{\hat{k}} \in \mathcal{X}^{n-1}} h(x_{\hat{k}}) \prod_{j \neq k} \pi(x_j), \quad (1)$$

where $x_{\hat{k}} = (x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$ and $h(x_{\hat{k}}) = \mathcal{E}(f(x_{\hat{k}}, \cdot), g(x_{\hat{k}}, \cdot))$ is the action of Dirichlet form \mathcal{E} on k -th coordinate of f and g with other coordinates held frozen.

We will be interested in understanding evolution of $\|f_t\|_p$, where $f_t = T_t f$. Notice that derivative of this quantity in t gives rise to $\mathcal{E}(f_t, f_t^{p-1})$, whereas derivative in p leads to $\text{Ent}_\pi(|f_t|^p)$, where for any $g \geq 0$ we define

$$\text{Ent}_\pi(g) \triangleq \mathbb{E}_\pi \left[g(X) \ln \frac{g}{\mathbb{E}[g]} \right] = \mathbb{E}[g] D(\pi^{(g)} \| \pi),$$

with $\pi^{(g)}(x) \triangleq \frac{g(x)\pi(x)}{\mathbb{E}[g]}$ and $D(\cdot \| \cdot)$ – the Kullback-Leibler divergence. The idea of bounding these two derivatives (in t and p) in terms of one another was introduced in [18]. This explains introduction of the following concept.

We say that a semigroup admits a p -logarithmic Sobolev inequality (p -LSI for short), see [3, Section 3], if for some constant α_p

$$\text{Ent}_\pi(f^p) \leq \frac{1}{\alpha_p} \mathcal{E}(f, f^{p-1}), \quad (2)$$

We note that $\mathcal{E}(f, f^{p-1}) \geq 0$ for $p > 1$ and $\mathcal{E}(f, f^{p-1}) \leq 0$ for $p < 1$ and this implies corresponding signs for constants α_p . As $p \rightarrow 1$ we have $\mathcal{E}(f, f^{p-1}) \rightarrow 0$ and so we need to renormalize by $\frac{1}{p-1}$ in this limit. Consequently, we define 1-LSI as

$$\text{Ent}_\pi(f) \leq \frac{1}{\alpha_1} \mathcal{E}(f, \ln f),$$

which is required to hold for all $f > 0$ on \mathcal{X} .

We do not discuss full history of LSI, only mentioning that $p = 2$ case originated in [18] and $p = 1$ in [8]; for more detailed history see [9, 28]. The $p = 1$ case is also known as modified LSI and connects to (one version of) discrete Ricci curvature [15].

We will mostly deal in this paper with a special case of a hypercube. Namely, we set $\mathcal{X} = \{0, 1\}$, $L_{x,y} = -1\{x = y\} + 1/2$, $\pi = \text{Bern}(1/2)$ and

$$T_t f(x) = f(x) \frac{1 + e^{-t}}{2} + f(1-x) \frac{1 - e^{-t}}{2}. \quad (3)$$

For this case the best LSI constants are $\alpha_p = \frac{2(p-1)}{p^2}$, see [33, Theorem 2.2.8], and the Dirichlet form takes particularly simple form:

$$\mathcal{E}_n(f, g) = -\frac{1}{2} (\Delta f, g), \quad \Delta f(x) \triangleq \sum_{y: y \sim x} (f(y) - f(x)) \quad (4)$$

$$\mathcal{E}_n(f, f) = \frac{1}{4} 2^{-n} \sum_{(x,y): x \sim y} (f(x) - f(y))^2, \quad (5)$$

where $x \sim y$ means that $x, y \in \{0, 1\}^n$ differ in precisely one coordinate.

We also quote one inequality from information theory, known as Mrs. Gerber's lemma, or MGL [43], which we write in the following form: for any $f \geq 0$

$$\frac{1}{n} \frac{\text{Ent}_{\pi^n}(T_t^{\otimes n} f)}{\mathbb{E}[f]} \leq \ln 2 - m \left(t, \frac{1}{n} \frac{\text{Ent}_{\pi^n}(f)}{\mathbb{E}[f]} \right). \quad (6)$$

Here $m(t, x) = h(h^{-1}(\ln 2 - x) * \frac{1-e^{-t}}{2})$, where

$$h(x) \triangleq -x \ln x - (1-x) \ln(1-x) \quad (7)$$

is the binary entropy function, $h^{-1} : [0, \ln 2] \rightarrow [0, 1/2]$ is its functional inverse and $a * b = (1-a)b + (1-b)a$ is the binary convolution.

A less cryptic restatement of MGL is the following: For all $t \geq 0$

$$\text{Ent}_{\pi^n}(T_t^{\otimes n} f) \leq \text{Ent}_{\pi^n}(T_t^{\otimes n} f_{iid}), \quad (8)$$

where $f_{iid}(x) = \prod_{k=1}^n f_1(x_k)$ with $f_1(\cdot)$ selected so that a) in (8) the equality holds for $t = 0$; b) $\mathbb{E}[f_{iid}] = \mathbb{E}[f]$. In other words, MGL states that among all functions f on the hypercube, Ent decreases slowest for product functions. (Note that for a general product semigroup the statement (8) does not need to hold even if we add an extra constraint that f_1 should be chosen so that, in addition to a) and b), it maximizes $\text{Ent}_\pi(T_t f_1)$; see [42, Theorem 6].)

1.2. Motivation and Organization

We motivate our investigation by the following three questions:

- Log-Sobolev inequality implies an estimate of the form

$$\text{Ent}(T_t^{\otimes n} f) \leq e^{-Ct} \text{Ent}(f).$$

However, for the hypercube a stronger estimate is given by the MGL (6). *Can MGL be derived from some strengthening of LSI?*

Note that by a method of comparison of Dirichlet forms, results derived from log-Sobolev inequalities can then be extended to semigroups other than $T_t^{\otimes n}$. As an illustration, note that [14, Example 3.3] estimates speed of convergence of a Metropolis chain on $\{0, \dots, n\}$ by comparing to $T_t^{\otimes n}$. Our methods allow to show better estimates, similar to (6).

- Hypercontractivity inequality for the hypercube (variously attributed to [31, 10, 5, 18]) says

$$\|T_t^{\otimes n} f\|_{p(t)} \leq \|f\|_{p_0}, \quad p(t) = 1 + (p_0 - 1)e^{2t}, \quad p_0 \geq 1. \quad (9)$$

This is well known to be tight in the sense that for any $q > p(t)$ we can find f s.t. $\|T_t f\|_q > \|f\|_{p_0}$. However, such f will be very close to identity (for this particular semigroup). *Is it possible to improve the range of (p, q) in (9) provided f is far from identity (say in the sense of $|\text{supp } f| \ll 2^n$)?*

For example, it is clear that $\|T_t^{\otimes n}\|_{1 \rightarrow \infty} = (1 + e^{-t})^n$. If f has small support, we have $\|f\|_{p_0} \geq e^{n\rho_0} \|f\|_1$, where $\rho_0 = (1 - \frac{1}{p_0}) \frac{1}{n} \ln \frac{2^n}{|\text{supp } f|}$ and thus

$$\|T_t^{\otimes n} f\|_\infty \leq \|f\|_{p_0}, \quad \forall t \geq \ln \frac{1}{e^{\rho_0} - 1} \quad (10)$$

which is a significant improvement of (9) for large times t .

- Finally, it was noticed in [6] that LSIs on Euclidean space are closely related to a form of uncertainty principle, which connects the tail behavior of the function and its Fourier transform. *We ask whether LSIs on finite groups (e.g. hypercube) imply bounds on the tradeoff between the sizes of supports of the function and its Fourier image.*

All these questions will be answered positively.

The structure of the paper is the following. In Section 2 we describe the main concept of this paper – the non-linear LSIs and prove some of its consequences, such as refined hypercontractivity and general MGL. In Section 3 we

switch from general theory to the particular case of the hypercube. We establish explicit forms of new LSIs and new hypercontractive estimates for functions of small support. In Section 4 we apply the latter to establish a sharp version of the uncertainty principle on the hypercube. Finally, Section 5 applies the uncertainty principle to derive a lower bound on large-degree Fourier coefficients of sparse Boolean functions.

2. Non-linear log-Sobolev inequalities

In this section we introduce a family of non-linear log-Sobolev inequalities (LSI) and prove three implications relevant for this paper. We mention that special case of $p = 2$ (which is the main case, especially for diffusion semigroups) has been known in analysis for a long time under the name of “entropy-energy” inequalities (see below), and thus our generalization is to consider general p . The results we prove are: tensorization (i.e. extension from T_t to $T_t^{\otimes n}$), integrating 1-LSI to get entropy decay, integrating p -LSI to get hypercontractivity. The first two are routine verifications, whereas the third required some new ideas.

Definition 1. For $p \geq 1$ and a concave, continuous, non-negative function $\Phi_p : [0, \infty) \rightarrow [0, \infty)$ with $\Phi_p(0) = 0$, let us define a (p, Φ_p) -LSI as

$$\frac{\text{Ent}(f^p)}{\mathbb{E}[f^p]} \leq \Phi_p \left(\frac{\mathcal{E}(f, f^{p-1})}{\mathbb{E}[f^p]} \right), \quad (11)$$

where for $p = 1$ we understand $\mathcal{E}(f, f^{p-1}) = \mathcal{E}(f, \ln f)$. For $p < 1$ the domain of Φ_p is replaced with $(-\infty, 0]$, and the definition remains the same. When convenient, we will restate (p, Φ_p) -LSI in the form

$$\pm \frac{\mathcal{E}(f, f^{p-1})}{\mathbb{E}[f^p]} \geq b_p \left(\frac{\text{Ent}(f^p)}{\mathbb{E}[f^p]} \right), \quad (12)$$

where $b_p : [0, \infty) \rightarrow [0, \infty]$ is a convex increasing with $b_p(0) = 0$ function defined as $b_p(y) \triangleq \inf\{x : \Phi_p(x) \geq y\}$ (with the usual agreement that $\inf \emptyset = \infty$). The \pm is taken to be $+$ for $p \geq 1$ and $-$ for $p < 1$.

Remark 1. For convenience we define Φ_p and b_p on $[0, \infty)$ even though the arguments in (11) and (12) may be universally bounded by constants $< \infty$. Note also that a concave non-negative function on $[0, \infty)$ must be increasing on $[0, a)$ and then constant on $[a, \infty)$ (either interval could be empty).

It is clear from concavity of Φ_p that the linear-LSIs (2) are obtained by taking $\frac{1}{\alpha_p} = \frac{d}{dx} \Big|_{x=0} \Phi_p$. We briefly review the history of such inequalities:

- For a Lebesgue measure on \mathbb{R}^n and $\mathcal{E}(f, g) = \int (\nabla f, \nabla g)$ the $p = 2$ inequality takes the form:

$$\int_{\mathbb{R}^n} h^2(x) \ln h^2 dx \leq \frac{n}{2} \ln \left(\frac{2}{n\pi e} \int_{\mathbb{R}^n} \|\nabla h(x)\|^2 dx \right), \quad \int h(x)^2 dx = 1. \quad (13)$$

It appeared in information theory [37, (2.3)] and [12] as a solution to the problem of minimizing Fisher information subject to differential entropy constraint (the minimizer is Gaussian density). In analysis, (13) has been used early by [41].

- Inequality (13) is in fact equivalent to a 2-LSI [18] for Ornstein-Uhlenbeck semigroup:

$$\text{Ent}_\gamma(f^2) \leq 2 \int_{\mathbb{R}^n} \|\nabla f\|^2 d\gamma, \quad (14)$$

where $\gamma = \mathcal{N}(0, I_n)$ (to see equivalence, take $f^2(x) = \lambda h^2(\lambda x)(2\pi)^{n/2} e^{-\|x\|^2/2}$ with $\lambda = (4 \int \|\nabla h\|^2)^{-\frac{1}{2}}$ and integrate by parts). It is known [11] that (13) (resp., (14)) is saturated by and only by Gaussian densities (resp., exponentials). In particular, taking $f = e^{ax-a^2}$ in (14) shows that for Ornstein-Uhlenbeck semigroup no improvement of (14), in the sense of Def. 1 is possible (linear LSI is the best one).

- More generally, the $p = 2$ inequalities were introduced into operator theory by Davies and Simon [13] under the name of entropy-energy inequalities; see [3] for a survey.
- A $p = 2$ inequality for the hypercube was proved in [35] for the purpose of showing that the Faber-Krahn problem on the hypercube is asymptotically solved by a Hamming ball. Same reference mentioned [35, paragraph after (11)] a tightening of hypercontractivity (9) for $p_0 = 2$ and functions of large entropy, although no proof was published at the time.
- Miclo [27] proved a class of restricted entropy-energy inequalities: Given continuous $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $\phi(x)/(x \ln x)$ is monotonically increasing for large enough x there exists a continuous increasing ψ and a universal constant $C > 0$ such that

$$\text{Var}_\pi[f^2] \geq C\psi(\text{Ent}_\pi(f^2)) \quad (15)$$

for all π simultaneously but only for functions f satisfying $\mathbb{E}_\pi[f^2] = 1$ and $\mathbb{E}_\pi[\phi(f^2)] \leq K$. Function ψ in (15) depends on ϕ and K roughly via $\psi(8K \frac{x \ln x}{\phi(x)}) = \frac{x}{\phi(x)}$, and in particular $\psi(t) = o(t)$ as $t \rightarrow 0$, so that (15) does not imply standard LSI (2). Here, we are interested in improving upon (2) and also in unrestricted inequalities (without constraint on $\mathbb{E}_\pi[\phi(f^2)]$), but for a fixed known π .

We move on to proving general results about non-linear LSIs.

Theorem 1. (*Tensorization*) *Suppose that (p, Φ_p) -LSI holds for semigroup $(\mathcal{X}, \pi, T_t, \mathcal{E})$. Then for all $n \geq 1$ the $(p, n\Phi_p(\frac{1}{n}))$ -LSI holds for $(\mathcal{X}^n, \pi^n, T_t^{\otimes n}, \mathcal{E}_n)$. In other words, for all $f : \mathcal{X}^n \rightarrow \mathbb{R}_+$ we have*

$$\frac{1}{n} \frac{\text{Ent}_{\pi^n}(f^p)}{\mathbb{E}_{\pi^n}[f^p]} \leq \Phi_p \left(\frac{1}{n} \frac{\mathcal{E}_n(f, f^{p-1})}{\mathbb{E}_{\pi^n}[f^p]} \right), \quad (16)$$

where $\pi^n = \prod_{k=1}^n \pi$ – a product measure on \mathcal{X}^n and \mathcal{E}_n is the Dirichlet form associated to the product semigroup (1).

Theorem 2. (General MGL) Suppose a semigroup T_t admits a $(1, \Phi_1)$ -LSI. Let $b_1 = \Phi_1^{-1}$ be a convex, strictly increasing inverse of Φ_1 and assume that the differential equation

$$\frac{d}{dt} \tilde{\rho}(t) = -b_1(\tilde{\rho}(t))$$

has a C^1 -solution $\tilde{\rho}(t)$ on $[0, t_0)$ with $\tilde{\rho}(0) > 0$. Then for any $f : \mathcal{X}^n \rightarrow \mathbb{R}_+$ with $\frac{1}{n} \frac{\text{Ent}(f)}{\mathbb{E}[f]} \leq \tilde{\rho}(0)$ we have

$$\text{Ent}(T_t^{\otimes n} f) \leq n \tilde{\rho}(t) \mathbb{E}[f] \quad \forall 0 \leq t < t_0.$$

Theorem 3. (Hypercontractivity) Fix a non-constant function $f : \mathcal{X}^n \rightarrow \mathbb{R}_+$ and $p_0 \in (1, \infty)$. Then there is a finite t_0 and a unique function $p(t)$ on $[0, t_0)$ satisfying $\|T_t^{\otimes n} f\|_{p(t)} = \|f\|_{p_0}$. This function is C^∞ -smooth, strictly increasing and surjective onto $[p_0, \infty)$ with $p(0) = p_0$. Furthermore, if a semigroup T_t admits a (p, Φ_p) -LSI for each $p \geq p_0$, then

$$\frac{d}{dt} p(t) \geq \frac{p(t)(p(t) - 1)}{\rho_0} b_{p(t)} \left(\frac{p(t)\rho_0}{p(t) - 1} \right), \quad \rho_0 = \frac{1}{n} \ln \frac{\|f\|_{p_0}}{\|f\|_1}. \quad (17)$$

Proof of Theorem 1. Let us consider the case $n = 2$. For a function $f(x_1, x_2)$ denote by $\text{Ent}_{\pi_i}(f^p)$ the entropy evaluated only along x_i , $i = 1, 2$. Then, from standard chain-rule and convexity of Ent we have

$$\text{Ent}_{\pi_1 \times \pi_2}(f^p) = \mathbb{E}_{X_1}[\text{Ent}_{\pi_2}(f^p)] + \text{Ent}_{\pi_1}(\mathbb{E}_{X_2}[f^p]) \quad (18)$$

$$\leq \mathbb{E}_{X_1}[\text{Ent}_{\pi_2}(f^p)] + \mathbb{E}_{X_2}[\text{Ent}_{\pi_1}(f^p)] \quad (19)$$

Now, we apply Φ_p -LSI to each term (not forgetting appropriate normalization). For example, for the first term we get

$$\mathbb{E}_{X_1}[\text{Ent}_{\pi_2}(f^p)] \leq \mathbb{E}_{X_1} \left[\Phi_p \left(\frac{\mathcal{E}_{\pi_2}(f, f^{p-1})}{\mathbb{E}_{X_2}[f^p]} \right) \mathbb{E}_{X_2}[f^p] \right] \quad (20)$$

$$\leq \Phi_p \left(\frac{\mathbb{E}_{X_1}[\mathcal{E}_{\pi_2}(f, f^{p-1})]}{\mathbb{E}_{X_1, X_2}[f^p]} \right) \mathbb{E}_{X_1, X_2}[f^p], \quad (21)$$

where in the second step we used Jensen's inequality and the fact that

$$(x, y) \mapsto \Phi \left(\frac{x}{y} \right) y$$

is jointly concave for any concave function Φ . Now plugging (21) (and its analog for the second term) into (19) and after applying Jensen's inequality again we get

$$\frac{1}{2} \frac{\text{Ent}_{\pi_1 \times \pi_2}(f^p)}{\mathbb{E}_{X_1, X_2}[f^p]} \leq \Phi_p \left(\frac{1}{2} \frac{\mathbb{E}_{X_1}[\mathcal{E}_{\pi_2}(f, f^{p-1})] + \mathbb{E}_{X_2}[\mathcal{E}_{\pi_1}(f, f^{p-1})]}{\mathbb{E}_{X_1, X_2}[f^p]} \right),$$

which is precisely (16). The $n > 2$ is treated similarly. \square

Proof of Theorem 2. Since the statement is scale-invariant, we assume $\mathbb{E}[f] = 1$. Define $\rho(t) \triangleq \frac{1}{n} \text{Ent}(T_t^{\otimes n} f)$. Consider the identity

$$\frac{d}{dt} \text{Ent}(T_t^{\otimes n} f) = -\mathcal{E}(T_t^{\otimes n} f, \ln T_t^{\otimes n} f).$$

From tensorizing the $(1, \Phi_1)$ -LSI we get

$$\frac{1}{n} \mathcal{E}(T_t^{\otimes n} f, \ln T_t^{\otimes n} f) \geq b_1(\rho(t)),$$

and hence

$$\rho'(t) \leq -b_1(\rho(t)).$$

Let us introduce $\alpha(t) = \ln \rho(t) - \ln \tilde{\rho}(t)$, then we have for $\alpha(t)$ the following

$$\alpha'(t) \leq -\Psi(\tilde{\rho}(t)e^{\alpha(t)}) + \Psi(\tilde{\rho}(t)), \quad (22)$$

where $\Psi(x) = b_1(x)/x$ is a non-decreasing function of $x \geq 0$. We know $\alpha(0) \leq 0$. Suppose that for some $\tilde{t}_0 > 0$ we have $\alpha(\tilde{t}_0) > 0$. Let $t_1 = \sup\{0 \leq t < \tilde{t}_0 : \alpha(t) = 0\}$. From continuity of α we have $t_1 < \tilde{t}_0$, $\alpha(t_1) = 0$ and $\alpha(t) > 0$ for all $t \in (t_1, \tilde{t}_0]$. From mean value theorem, we have for some $t_2 \in (t_1, \tilde{t}_0)$ that $\alpha'(t_2) > 0$. But then from monotonicity of Ψ , we have

$$\Psi(\tilde{\rho}(t_2)e^{\alpha(t_2)}) - \Psi(\tilde{\rho}(t_2)) \geq 0,$$

contradicting (22). Hence $\alpha(\tilde{t}_0) \leq 0$ for all $\tilde{t}_0 \in (0, t_0)$. \square

Proof of Theorem 3. The core idea is to integrate the estimates obtained from a non-linear p -LSI. Integrating entropy-energy inequalities have been done before for establishing ultra-contractivity (i.e. for bounding the kernel function of T_t), see e.g. [3, Theorem 4.4]. However, for $p \rightarrow q$ estimates we will need a new idea (see (32) below).

Since all the statements are scale-invariant, we assume $\mathbb{E}[f] = 1$. To avoid clutter, we will write T_t instead of $T_t^{\otimes n}$. We define the following function on \mathbb{R}_+^2

$$\phi(t, \xi) \triangleq \ln \|T_t f\|_{\frac{1}{\xi}}.$$

It is clear that ϕ is monotonically decreasing in ξ . Steepness of ϕ in ξ encodes information about non-uniformity of $T_t f$. As time progresses, $\xi \mapsto \phi(t, \xi)$ converges to an all-zero function. MGL, LSI and hypercontractivity are estimates on the speed of this relaxation.

We summarize the information we have about $\phi(t, \xi)$ assuming f is non-constant:

- A consequence of Hölder's inequality, cf. [19, Theorems 196-197], implies $\xi \mapsto \ln \|g\|_{\frac{1}{\xi}}$ is strictly convex, unless $g = c1_S$ (a scaled indicator), in which case the function of is linear in ξ . Thus, $\phi(t, \xi)$ is convex in ξ .

- We have

$$\phi(0, \xi_2) \geq \phi(0, \xi_1) + (\xi_1 - \xi_2) \ln \frac{1}{\pi^n [\text{supp } f]} \quad \forall \xi_2 < \xi_1$$

with equality iff $f = c1_S$ (scaled indicator).

- Note that $T_t f = 0$ has only $f = 0$ as solution (indeed, $\det e^{tL} = e^{\text{tr } L} \neq 0$ since \mathcal{X} is finite). So $\phi(t, \xi)$ is finite and infinitely differentiable in (t, ξ) .
- The function $t \mapsto \phi(t, \xi)$ is strictly decreasing from $\phi(0, \xi)$ to 0 for any $\xi < 1$ and strictly increasing from $\phi(0, \xi)$ to 0 for $\xi > 1$. Indeed, $\|T_t f\|_p = \|f\|_p$ implies f is constant. Furthermore, $\|T_t f\|_{\frac{1}{\xi}} \rightarrow \mathbb{E}[f] = 1$ since $T_t f \rightarrow \mathbb{E}[f]$ as $t \rightarrow \infty$.
- Consequently, for each ξ_0 the fiber

$$\{t : \phi(t, \xi_0) = c\} \tag{23}$$

consists of at most one point. Define t_0 as the unique solution of

$$\phi(t_0, 0) = n\rho_0.$$

Solution exists from continuity of ϕ and the fact that $\phi(0, 0) = \ln \|f\|_\infty > \rho_0 > \phi(+\infty, 0) = 0$.

- We have the standard identities:

$$\frac{\partial \phi}{\partial \xi} = - \frac{\text{Ent}((T_t f)^{\frac{1}{\xi}})}{\mathbb{E}[(T_t f)^{\frac{1}{\xi}}]} \tag{24}$$

$$\frac{\partial \phi}{\partial t} = - \frac{\mathcal{E}(T_t f, (T_t f)^{\frac{1}{\xi}-1})}{\mathbb{E}[(T_t f)^{\frac{1}{\xi}}]}. \tag{25}$$

- Since f is non-constant, so is $T_t f$ for all $t \geq 0$ (for otherwise $f - \mathbb{E}[f]$ is in the kernel of T_t). Therefore, $\frac{\partial \phi}{\partial \xi} < 0$ for all (ξ, t) . Thus, for any $t \in [0, t_0]$ there is at most one solution ξ of

$$\phi(t, \xi(t)) = \phi\left(0, \frac{1}{p_0}\right) = n\rho_0. \tag{26}$$

$\xi(t)$ is simply a parametrization of the level-set of ϕ . It is clear that $\xi(t)$ is non-increasing. Since fibers (23) are singletons, we also conclude that $\xi(t)$ is strictly decreasing. From implicit function theorem and $\frac{\partial \phi}{\partial \xi} \neq 0$, we infer that solution $\xi(t)$ of (26) is C^∞ -smooth.

- As we mentioned $\xi \mapsto \phi(t, \xi)$ is convex and strictly decreasing. Furthermore, it is strictly convex for $t > 0$. From this convexity and (24) we infer

the following important consequences:

$$r \mapsto \frac{\text{Ent}(f^r)}{\mathbb{E}[f^r]} \text{ is increasing in } r \in (0, \infty); \text{ strictly unless } f = c1_S \quad (27)$$

$$\frac{\text{Ent}(f^r)}{\mathbb{E}[f^r]} \geq \frac{\ln \|f\|_r - \ln \|f\|_1}{1 - \frac{1}{r}}. \quad r > 1. \quad (28)$$

We now set $p(t) = \frac{1}{\xi(t)}$, where $\xi(t)$ was found from solving (26). From observations after (26) we already know that $t \mapsto p(t)$ is well-defined, strictly increasing and C^∞ -smooth. The fact that $p(t)$ is surjective follows from $\xi(t) \rightarrow 0$ as $t \rightarrow t_0$.

It remains to show (17). This follows from differentiating (26):

$$\xi'(t) = -\frac{\mathcal{E}(t)}{E(t)},$$

where we defined

$$\mathcal{E}(t) \triangleq \frac{1}{n} \frac{\mathcal{E}(T_t f, (T_t f)^{p(t)-1})}{\mathbb{E}[(T_t f)^{p(t)}]} \quad (29)$$

$$E(t) \triangleq \frac{1}{n} \frac{\text{Ent}((T_t f)^{p(t)})}{\mathbb{E}[(T_t f)^{p(t)}]}. \quad (30)$$

From (p, Φ_p) -LSI we get then

$$\xi'(t) \leq -\frac{b_{p(t)}(E(t))}{E(t)}. \quad (31)$$

Here we arrived at a key new step. Note that without further information about $E(t)$ we can only bound (due to convexity of $b_p(\cdot)$) the right-hand side of the above by $-\frac{d}{ds}\big|_{s=0} b_{p(t)}(s)$, which would result in a standard, i.e. ρ_0 -independent, hypercontractivity such as (9). To improve it, we need to lower-bound $E(t)$ away from 0. Note that from (28) we know that $E(0) \geq \frac{\rho_0}{1-p_0}$. To extend this to other times we use (28) coupled with the fact that $\xi(t)$ is precisely the level-set of ϕ . Hence, we get

$$E(t) \geq \frac{1}{n} \frac{\phi(t, \xi(t))}{1 - \xi(t)} = \frac{\rho_0}{1 - \xi(t)} \quad (32)$$

From convexity of $b_p(\cdot)$, the function $\frac{b_p(E)}{E}$ is increasing in E and so we can further upper-bound $\xi'(t)$ via (32) and replacing $\xi(t)$ with $\frac{1}{p(t)}$ as

$$\xi'(t) \leq -\frac{b_{p(t)}\left(\frac{p(t)\rho_0}{p(t)-1}\right)}{\frac{p(t)\rho_0}{p(t)-1}}.$$

Noticing that $\xi'(t) = -\frac{p'(t)}{p^2(t)}$ we get (17). \square

3. New LSIs and hypercontractivity for the hypercube

The fact that we can compare [38, 40] Dirichlet forms $\mathcal{E}(f, f^{p-1})$ with $\mathcal{E}(f^{\frac{p}{2}}, f^{\frac{p}{2}})$ immediately leads to the conclusion that for any reversible semigroup (i.e. $T_t^* = T_t$ in $L_2(\pi)$) we have

$$b_p(x) \geq \frac{4|p-1|}{p^2} b_2(x) \quad \forall x \geq 0 \forall p \in (-\infty, \infty) \setminus \{1\} \quad (33)$$

$$b_p(x) \geq \frac{1-p}{p^2} b_1(x) \quad \forall x \geq 0 \forall p < 1. \quad (34)$$

(see, e.g., [3] for $p = 2$ and [28] for $p = 1$). Thus, we can get non-trivial non-linear p -LSIs by only establishing $p = 1, 2$ cases (of which $p = 2$ was already done in [35]). However, we can also find the sharpest non-linear LSIs for all p explicitly, which is what we proceed to do.

Theorem 4 (1-LSI for the hypercube). *For all $f : \{0, 1\}^n \rightarrow (0, \infty)$ with $\mathbb{E}[f] = 1$ we have*

$$\frac{1}{n} \mathcal{E}(f, \ln f) \geq b_1 \left(\frac{1}{n} \text{Ent}(f) \right), \quad (35)$$

where Dirichlet form is given by (4), all expectations and Ent are with respect to uniform probability measure on $\{0, 1\}^n$ and $b_1 : [0, \ln 2) \rightarrow [0, \infty)$ is a convex increasing function given by

$$b_1(\ln 2 - h(y)) = \left(\frac{1}{2} - y\right) \ln \frac{1-y}{y}, \quad y \in (0, 1/2), \quad (36)$$

where $h(\cdot)$ is the binary entropy function (7).

Proof. This result follows from Theorem 6 (below) upon taking $p \rightarrow 1+$. \square

Corollary 5. *Classical MGL (6) holds.*

Proof. Since $\text{Ent}(f) \leq n\mathbb{E}[f] \ln 2$, we can define $\tilde{\rho}(t) = \ln 2 - m\left(t, \frac{\text{Ent}(f)}{n\mathbb{E}[f]}\right)$, where $m(\cdot, \cdot)$ was defined after (6). A calculation shows that $\tilde{\rho}(t)$ solves $\tilde{\rho}(t)' = -b_1(\tilde{\rho}(t))$ with b_1 from (36). Since $\tilde{\rho}(0) \geq \frac{1}{n} \frac{\text{Ent}(f)}{\mathbb{E}[f]}$, application of Theorem 2 completes the proof. \square

Next, we proceed to LSI's with $p \neq 1$.

Theorem 6 (p -LSI for the hypercube). *Fix $p \in (-\infty, \infty) \setminus \{0, 1\}$. For all $f : \{0, 1\}^n \rightarrow \mathbb{R}_+$ with $\mathbb{E}[f^p] = 1$ (and $f > 0$ if $p < 1$) we have*

$$\frac{1}{n} \text{sgn}(p-1) \mathcal{E}(f, f^{p-1}) \geq b_p \left(\frac{1}{n} \text{Ent}(f^p) \right), \quad (37)$$

where the Dirichlet form is given by (4), all expectations and Ent are with respect to uniform probability measure on $\{0, 1\}^n$ and $b_p : [0, \ln 2] \rightarrow [0, \infty)$ is a convex increasing function given by

$$b_p(\ln 2 - h(y)) = \frac{\text{sgn}(p-1)}{2} \left(1 - y^{\frac{1}{p}}(1-y)^{1-\frac{1}{p}} - y^{1-\frac{1}{p}}(1-y)^{\frac{1}{p}} \right), \quad (38)$$

with $0 < y \leq \frac{1}{2}$, and $h(\cdot)$ being the binary entropy function (7).

Remark 2. Recall that the proof of (14) in [18] for Ornstein-Uhlenbeck semigroup was done by first deriving the LSI for the hypercube and then applying the CLT. Since we derive a better LSI for the hypercube, will we get a better LSI for the Ornstein-Uhlenbeck? The answer is negative since in the CLT limit we would have $\text{Ent}(f^p) = O(1)$ as $n \rightarrow \infty$ and hence the argument of b_p in (37) converges to 0 and we get the linear LSI in the limit. In fact, as noted above, (14) is tight.

Proof. By Theorem 1, we only need to work out the case $n = 1$. Then, the space of all f can be parameterized by $f(0) = (2y)^{\frac{1}{p}}, f(1) = (2-2y)^{\frac{1}{p}}$ with $y \in [0, 1/2]$. Thus we only need verify monotonicity and convexity.

First, consider the case $p > 1$. Let $q = \frac{p}{p-1}$. Taking the first derivative, we get

$$b'_p(\ln 2 - h(y)) = -\frac{1}{2} \frac{1}{\ln \frac{y}{1-y}} \cdot \left(\frac{1}{p} \left(\left(\frac{1-y}{y} \right)^{1/q} - \left(\frac{y}{1-y} \right)^{1/q} \right) + \frac{1}{q} \left(\left(\frac{1-y}{y} \right)^{1/p} - \left(\frac{y}{1-y} \right)^{1/p} \right) \right) \quad (39)$$

From here, monotonicity of b_p follows from the fact that the RHS is positive ($\frac{1-y}{y} > \frac{y}{1-y}$). We proceed to showing convexity. Let $z = \frac{y}{1-y}$. Then $0 < z \leq 1$ and, taking another derivative, we have

$$b''_p(\ln 2 - h(y)) = -\frac{1}{2} \frac{1}{(1-y)^2} \cdot \frac{1}{\ln z} \cdot \frac{d}{dz} \left[\frac{1}{\ln z} \cdot \left(\frac{1}{p} \left(\left(\frac{1}{z} \right)^{1/q} - z^{1/q} \right) + \frac{1}{q} \left(\left(\frac{1}{z} \right)^{1/p} - z^{1/p} \right) \right) \right] \quad (40)$$

Since $\ln z < 0$ for $z < 1$, it would suffice to argue that the derivative w.r.t. z on RHS is nonnegative. Let $r(z) = \frac{1}{p} \left(\left(\frac{1}{z} \right)^{1/q} - z^{1/q} \right) + \frac{1}{q} \left(\left(\frac{1}{z} \right)^{1/p} - z^{1/p} \right)$. We need to show $z \ln \frac{1}{z} \cdot (-r') \geq r$.

Making another substitution of variables, let $w = \ln z$, that is $-\infty < w \leq 0$. Let $t(w) = r(z) = r(e^w)$. Substituting and simplifying, we need to show $wt'(w) \geq t(w)$.

We have $t(w) = \frac{1}{p} (e^{-w/q} - e^{w/q}) + \frac{1}{q} (e^{-w/p} - e^{w/p})$. Hence

$$t'(w) = -\frac{1}{pq} \left(e^{-w/q} + e^{w/q} + e^{-w/p} + e^{w/p} \right) \quad \text{and}$$

$$t''(w) = -\frac{1}{pq} \left(-\frac{1}{q}e^{-w/q} + \frac{1}{q}e^{w/q} - \frac{1}{p}e^{-w/p} + \frac{1}{p}e^{w/p} \right)$$

In particular, t is a decreasing convex function on $(-\infty, 0]$ which vanishes at 0, and $wt'(w) \geq t(w)$ is satisfied.

Next, consider the case $0 < p < 1$. We repeat the computation above, multiplying throughout by $-1 = \text{sgn}(p-1)$. Since in this case $q < 0$, the sign change cancels out, and the convexity argument, with minor changes as needed, goes through. For monotonicity, observe that again the signs of both terms in the RHS of (39) are negative (the front $-$ sign is canceled by $\text{sgn}(p-1)$).

Finally, for the case $p < 0$, observe that we can set $g = f^{p-1}$ and apply the already proven inequality to the pair $(g, \frac{p}{p-1})$ since $\frac{p}{p-1} \in (0, 1)$. \square

Our chief goal is to derive hypercontractivity inequality tighter than (9) for functions with small support. We will replace the constraint on the support size $|\text{supp } f| \leq 2^{nR}$ with an analytical proxy:

$$\|f\|_{p_0} \geq e^{n\rho_0} \|f\|_1, \quad \rho_0 = (1 - p_0^{-1})(1 - R) \ln 2,$$

as discussed in (10). We get the following result:

Theorem 7. *Fix $1 < p_0 < \infty$ and $0 \leq \rho_0 \leq (1 - p_0^{-1}) \ln 2$. Then the differential equation*

$$u'(t) = C \left(\rho_0(1 + e^{-u(t)}) \right), \quad C(\ln 2 - h(y)) = \frac{2 - 4\sqrt{y(1-y)}}{\ln 2 - h(y)} \quad (41)$$

has a unique solution on $[0, \infty)$ with $u(0) = \ln(p_0 - 1)$. Furthermore, for any $f : \{0, 1\}^n \rightarrow \mathbb{R}_+$ with $\|f\|_{p_0} \geq e^{n\rho_0} \|f\|_1$ we have

$$\|T_t^{\otimes n} f\|_{p(t)} \leq \|f\|_{p_0}, \quad p(t) = 1 + e^{u(t)}. \quad (42)$$

Remark 3. *Ref. [35] showed that $C : [0, \ln 2] \rightarrow [2, 2/\ln 2]$ is a smooth, convex and strictly increasing bijection. Consequently, the function $p(t)$ in (42) is smooth and satisfies*

$$p(t) > 1 + (p_0 - 1)e^{2t} \quad \forall t > 0$$

thereby strictly improving the hypercontractivity inequality (9). Furthermore, it satisfies

$$p(t) = p_0 + p'(0)t + \frac{1}{2}p''(0)t^2 + o(t^2), \quad t \rightarrow 0, \quad (43)$$

$$p'(0) = (p_0 - 1)C(x_0) \quad (44)$$

$$p''(0) = (p_0 - 1) \left(C(x_0)^2 - C'(x_0)C(x_0)\frac{x_0}{p_0} \right) \quad (45)$$

$$x_0 = \frac{\rho_0 p_0}{p_0 - 1}. \quad (46)$$

Remark 4. Our estimate is locally optimal at $t = 0$ in the following sense: for every $q(t)$ such that $q(0) = p_0$ and $q'(0) > p'(0)$ there exists a function f with $\|f\|_{p_0} \geq e^{n\rho_0}\|f\|_1$ and $\|T_t f\|_{q(t)} > \|f\|_{p_0}$ for a sequence of $t \rightarrow 0$. This follows from the fact that had a counter-example $q(t)$ existed, it would imply that the second half of the proof of Theorem 9 (see below) could be improved to contradict the first half.

Proof. First, notice that $C(x) = \frac{4b_2(x)}{x}$, where b_2 was defined in Theorem 6. Let $p_1(t)$ be the function defined by

$$\|T_t^{\otimes n} f\|_{p_1(t)} = \|f\|_{p_0}.$$

Theorem 3 showed this function to be smooth and growing faster than (17). From (17) and using (33) to lower-bound $b_p(\cdot)$ via $b_2(\cdot)$ we get that

$$p_1'(t) \geq (p_1(t) - 1)C\left(\frac{p_1(t)\rho_0}{p_1(t) - 1}\right),$$

or introducing $u_1(t) = \ln(p_1(t) - 1)$ that

$$u_1'(t) \geq C\left(\rho_0(1 + e^{-u_1(t)})\right).$$

The case of $\rho_0 = (1 - p_0^{-1}) \ln 2$ corresponds to f supported on a single point and can be dealt with separately. So we assume $\rho_0 < (1 - p_0^{-1}) \ln 2$, in which case the map

$$u \mapsto C\left(\rho_0(1 + e^{-u})\right)$$

is smooth on some interval $(\ln(p_0 - 1) - \epsilon, \infty)$. Consequently, (41) possesses a unique solution with $u(0) = \ln(p_0 - 1)$ and a Chaplygin-type theorem, e.g. [20, Theorem 4.1], implies

$$u_1(t) \geq u(t).$$

□

For $p_0 = 2$, we also prove an alternative estimate on $p(t)$ via a method tailored to the hypercube.

Theorem 8. In the setting of Theorem 7 assume $p_0 = 2$. Then (42) holds with $p(t)$ given as

$$p(t) = 1 + e^{\int_0^t C(\tilde{\rho}(s) \vee 0) ds} \tag{47}$$

$$\tilde{\rho}(s) = 2\rho_0 - \ln\left(\frac{2}{1 + e^{-2s}}\right). \tag{48}$$

Remark 5. Using convexity of C we get $C(x \vee 0) \geq C(x_0) + (x - x_0)C'(x_0)$, where $x_0 = \tilde{\rho}(0) = 2\rho_0$ is from (46). Similarly, $\ln \frac{1+e^{-2t}}{2} \geq -t$. Therefore, altogether we get an explicit estimate:

$$p(t) \geq 1 + e^{C(x_0)t - \frac{C'(x_0)}{2}t^2}, \tag{49}$$

The t^2 term here is, however, worse than that of (43).

Proof. We return to (31). Recalling that $\xi(t) = \frac{1}{p(t)}$ and lower-bounding b_p by b_2 via (33) we get

$$\frac{d}{dt} \ln(p(t) - 1) \geq C(E(t)),$$

(with $E(t)$ from (30)), which implies after integration

$$p(t) \geq 1 + e^{\int_0^t C(E(s)) ds}.$$

Since $E(t) \geq 0$ by definition it only suffices to prove

$$E(t) \geq 2\rho_0 - \ln\left(\frac{2}{1 + e^{-2t}}\right) \quad (50)$$

Next, we obtain a lower bound on $\|T_t^{\otimes n} f\|_2$. To that end introduce a function $\Lambda_t^{\otimes n}$ on $\{0, 1\}^n$ with the property $T_t^{\otimes n} f = \Lambda_t^{\otimes n} * f$. Note that $\Lambda_t^{\otimes n}(x) = (1 - e^{-t})^{|x|} (1 + e^{-t})^{n-|x|}$, where $|x|$ denotes the Hamming weight of x . Clearly $\Lambda_t^{\otimes n} \geq 0$. Since $f \geq 0$ we have

$$\begin{aligned} \|T_t^{\otimes n} f\|_2^2 &= \langle T_t^{\otimes n} f, T_t^{\otimes n} f \rangle = \langle \Lambda_t^{\otimes n} * f, \Lambda_t^{\otimes n} * f \rangle = \\ &\langle \Lambda_t^{\otimes n} * \Lambda_t^{\otimes n}, f * f \rangle = \langle \Lambda_{2t}^{\otimes n}, f * f \rangle \geq \frac{1}{2^n} \Lambda_{2t}^{\otimes n}(0) \cdot (f * f)(0) = \left(\frac{1 + e^{-2t}}{2}\right)^n \cdot \|f\|_2^2 \end{aligned}$$

To prove (50), observe that by (27) and (28) and by the preceding calculation,

$$E(t) \geq \frac{1}{n} \ln \frac{\|T_t^{\otimes n} f\|_2^2}{\|T_t^{\otimes n} f\|_1^2} = \frac{1}{n} \ln \frac{\|T_t^{\otimes n} f\|_2^2}{\|f\|_1^2} \geq \frac{1}{n} \ln \frac{\|f\|_2^2}{\|f\|_1^2} - \ln\left(\frac{2}{1 + e^{-2t}}\right) \geq \tilde{\rho}(t)$$

We remark that the main difference in this proof compared to Theorem 7 is in using a different idea for lower-bounding the entropy $E(t)$. Theorem 7 essentially relied on (32). \square

4. Uncertainty principle on the hypercube

4.1. Background

Uncertainty principle asserts that a function and its Fourier transform cannot be simultaneously narrowly concentrated. There are several approaches to quantifying this statement, and here we adopt the Hilbert space point of view, cf. [21, Chapter 3]. Namely, for a pair of subspaces V_1, V_2 of a Hilbert space with inner product (\cdot, \cdot) and $\|f\|_2^2 \triangleq (f, f)$ we define

$$\cos \angle(V_1, V_2) \triangleq \sup_{f_1 \in V_1, f_2 \in V_2} \frac{|(f_1, f_2)|}{\|f_1\|_2 \|f_2\|_2}.$$

For the uncertainty principle, we will select sets S and Σ and define subspaces

$$V_S \triangleq \{f : \text{supp } f \subset S\} \quad (51)$$

$$\hat{V}_\Sigma \triangleq \{f : \text{supp } \hat{f} \subset \Sigma\}, \quad (52)$$

where \hat{f} denotes the corresponding Fourier transform (we will define it precisely). Uncertainty principle corresponds to bounding $\cos \angle(V_S, \hat{V}_\Sigma)$ away from 1, thus establishing to what extent functions can simultaneously concentrate on (S, Σ) .

There is a number of equivalent ways to think of $\cos \angle(V_1, V_2)$. Letting P_i be an orthogonal projection on V_i and P_i^\perp projection on V_i^\perp , it can be shown [21, Chapter 3]:

$$\cos \angle(V_1, V_2) \leq \theta \iff \forall f \in V_1 : \|P_2 f\|_2 \leq \theta \|f\|_2 \quad (53)$$

$$\iff \forall f \in V_1 : \|P_2^\perp f\|_2 \geq \sqrt{1 - \theta^2} \|f\|_2 \quad (54)$$

$$\iff \lambda_{\max}(P_1 P_2 P_1) \leq \theta \quad (55)$$

$$\iff \|P_1 P_2\|_{2 \rightarrow 2} \leq \sqrt{\theta} \quad (56)$$

$$\iff \forall f : \|f\|_2^2 \leq \frac{1}{1 - \theta} (\|P_1^\perp f\|_2^2 + \|P_2^\perp f\|_2^2) \quad (57)$$

Furthermore, there is also a simple criterion:

$$\cos \angle(V_1, V_2) < 1 \iff V_1 \cap V_2 = \{0\} \text{ and } (V_1 + V_2) \text{ --- closed,}$$

where for finite-dimensional V_i 's the closedness condition is vacuous (but not in general).

Finally, as shown in [17] and [25], knowledge of $\cos \angle(V_1, V_2)$ is sufficient for completely characterizing the two-dimensional region

$$\{(\|P_1 f\|_2^2, \|P_2 f\|_2^2)\}$$

Before proceeding to our own results, we briefly review the history of results for \mathbb{R}^n . First, [36] computed $\cos \angle(V_S, \hat{V}_\Sigma)$ for S, Σ being two balls (in fact they computed $\lambda_{\max}(P_1 P_2 P_1)$ and named eigenfunctions of the latter prolate spheroidal functions). Next, [7] (worked out in 1974, but published much later) showed that

$$\text{vol}(S), \text{vol}(\Sigma) < \infty \implies V_S \cap \hat{V}_\Sigma = \{0\}.$$

Later, [1] strengthened this to

$$\text{vol}(S), \text{vol}(\Sigma) < \infty \implies \cos \angle(V_S, V_\Sigma) < 1.$$

Finally, for $n = 1$ [30] showed

$$\text{vol}(S), \text{vol}(\Sigma) < \infty \implies \cos \angle(V_S, V_\Sigma) < 1 - ce^{-c \text{vol}(S) \text{vol}(\Sigma)}.$$

Lately, there were a number of extensions and improvements of this result for $n > 1$, e.g. [22].

4.2. Sharp uncertainty principle on \mathbb{F}_2^n

Define the characters, indexed by $v \in \mathbb{F}_2^n$,

$$\chi_v(x) \triangleq \prod_{j:v_j=1} \chi_j(x) = (-1)^{\langle v, x \rangle},$$

where $\langle v, x \rangle = \sum_{j=1}^n v_j x_j$ is a non-degenerate bilinear form on \mathbb{F}_2^n . The Fourier transform of $f : \mathbb{F}_2^n \rightarrow \mathbb{C}$ is

$$\hat{f}(\omega) \triangleq \sum_{x \in \mathbb{F}_2^n} \chi_\omega(x) f(x) = 2^n (f, \chi_\omega), \quad \omega \in \mathbb{F}_2^n.$$

We denote by $|x|$ the Hamming weight of $x \in \mathbb{F}_2^n$ and by $B_r = \{x : |x| \leq r\}$ – Hamming ball.

Theorem 9. *For any $\rho_1, \rho_2 \in [0, 1/2]$ satisfying*

$$(1 - 2\rho_1)^2 + (1 - 2\rho_2)^2 > 1, \quad (58)$$

there exist an $\epsilon > 0$ and n_0 such that for any $n \geq n_0$, any $S \subset \mathbb{F}_2^n$ with $|S| \leq e^{nh(\rho_1)}$ and $\Sigma = B_{\rho_2 n}$ we have

$$\cos \angle(V_S, \hat{V}_\Sigma) \leq e^{-n\epsilon}, \quad (59)$$

where V_S, \hat{V}_Σ are defined in (51)-(52).

Conversely, for $\rho_1, \rho_2 \in [0, 1/2]$ satisfying¹

$$(1 - 2\rho_1)^2 + (1 - 2\rho_2)^2 < 1, \quad (60)$$

there exist $\epsilon > 0$ and n_0 such that for all $n \geq n_0$ we have

$$\cos \angle(V_S, \hat{V}_\Sigma) \geq 1 - e^{-n\epsilon}, \quad S = B_{\rho_1 n}, \Sigma = B_{\rho_2 n}. \quad (61)$$

Proof. For the case (60), fix $\alpha \in \mathbb{R}$, $\alpha \neq \pm 1$ and consider the following Fourier pair:

$$f(x) = \alpha^{|x|} \quad (62)$$

$$\hat{f}(\omega) = c \left(\frac{1 - \alpha}{1 + \alpha} \right)^{|\omega|}, \quad c = (1 + \alpha)^n. \quad (63)$$

Then, it is easy to see that the L_2 -norm of f is concentrated around $|x| \approx \frac{1}{1+\alpha^{-2}}n$. Thus, whenever radius $\rho_1 > \frac{1}{1+\alpha^{-2}}$, we have for some $\epsilon > 0$

$$\sum_{x: |x| > \rho_1 n} f(x)^2 \leq e^{-\epsilon n} \sum_{x \in \mathbb{F}_2^n} f(x)^2.$$

Similarly, whenever $\rho_2 > \frac{1}{1+\beta^{-2}}$, where $\beta = \frac{1-\alpha}{1+\alpha}$, we have

$$\sum_{\omega: |\omega| > \rho_2 n} \hat{f}(\omega)^2 \leq e^{-\epsilon n} \sum_{\omega \in \mathbb{F}_2^n} \hat{f}(\omega)^2.$$

¹When $\rho_1 > \frac{1}{2}$ (or $\rho_2 > \frac{1}{2}$), the result (61) also holds by reducing to $\rho_1 = \frac{1}{2}$. This is possible since $\cos \angle(V_S, \hat{V}_\Sigma)$ is monotone in S, Σ .

Whenever, (60) holds, it is not hard to see that there exists a choice of $\alpha \in (0, 1)$ satisfying both $\rho_1 > \frac{1}{1+\alpha-2}$ and $\rho_2 > \frac{1}{1+\beta-2}$. Thus, taking corresponding f and using (57) we get (61).

Next, we assume (58). We define the Fourier projection operators Π_a as

$$\widehat{\Pi_a f}(\omega) \triangleq \hat{f}(\omega) 1\{|\omega| = a\}, \quad a = 0, 1, \dots, n. \quad (64)$$

and set $\Pi_{\leq r} = \sum_{a=0}^r \Pi_a$. By (53) we need to show that for any function f with support $|\text{supp } f| \leq e^{nh(\rho_1)}$ we have

$$\|\Pi_{\leq \rho_2 n} f\|_2 \leq e^{-n\epsilon} \|f\|_2,$$

for some $\epsilon > 0$ independent of n and f .

Note that $\widehat{T_t f}(\omega) = e^{-t|\omega|} \hat{f}(\omega)$. Thus, comparing eigenvalues we have $e^{ta} T_t \succeq \Pi_a$ (in the sense of positive-semidefiniteness). Consequently,

$$\|\Pi_a f\|_2^2 = (\Pi_a f, f) \leq e^{at} (T_t f, f) \leq e^{at} \|f\|_q \|T_t f\|_p, \quad (65)$$

where p and q are Hölder conjugates. Since $|\text{supp } f| \leq e^{nh(\rho_1)}$ we have from Theorem 7 with $p_0 = 2$ and $\rho_0 = \frac{\ln 2 - h(\rho_1)}{2}$:

$$\|T_t f\|_{p(t)} \leq \|f\|_2, \quad p(t) = 2 + p'(0)t + o(t),$$

where the value of $p'(0)$ is given in (44).² Taking $p = p(t) > 2$ in (65) we need upper-bound $\|f\|_q$, which we again do by invoking the bound on support

$$\|f\|_q \leq \|f\|_2 e^{-n(\ln 2 - h(\rho_1))(\frac{1}{q} - \frac{1}{2})}, \quad \forall 1 \leq q \leq 2. \quad (66)$$

Overall, we have shown for all a and t that

$$\|\Pi_a f\|_2^2 \leq e^{at} e^{-n(\ln 2 - h(\rho_1))(\frac{1}{2} - \frac{1}{p(t)})} \|f\|_2^2.$$

Analyzing this inequality for t close to 0 we conclude that whenever

$$\rho_2 < \frac{p'(0)}{4} (\ln 2 - h(\rho_1)) \quad (67)$$

we necessarily have for some $\epsilon > 0$ (depending on the gap in the inequality above and on the local bound for $p''(t)$ at 0) that for all $a \leq \rho_2 n$

$$\|\Pi_a f\|_2 \leq e^{-n\epsilon} \|f\|_2.$$

Using expression for $p'(0)$ in (44), we see that (67) is equivalent to

$$2\rho_2 < 1 - 2\sqrt{\rho_1(1 - \rho_1)}, \quad (68)$$

which is in turn equivalent to (58). □

²For extracting explicit constants, one may invoke (49) instead.

For completeness, we also provide a criterion for when two subspaces have a common element (for the special case of S, Σ being two balls). It demonstrates that there is a “discontinuity” between the regime $\cos \angle(V_S, \hat{V}_\Sigma) \geq 1 - e^{O(n)}$ and $\cos \angle(V_S, \hat{V}_\Sigma) = 1$.

Proposition 10. *Let $S = B_{r_1}$ and $\Sigma = B_{r_2}$ in \mathbb{F}_2^n . Then*

$$V_S \cap \hat{V}_\Sigma \neq \{0\} \iff \cos \angle(V_S, \hat{V}_\Sigma) = 1 \iff r_1 + r_2 \geq n.$$

Proof. If $r_1 + r_2 \geq n$, then take $f(x) = 1\{x_{r_1+1} = \dots = x_n = 0\}$. Its Fourier transform is supported on $\{\omega : \omega_1 = \dots = \omega_{r_1} = 0\}$. Thus $f \in V_S \cap \hat{V}_\Sigma$. On the other hand, suppose there is $f \in V_S \cap \hat{V}_\Sigma$. By averaging over permutations of coordinates (both subspaces are invariant to such), we conclude that $f(x) = f_1(|x|)$. As such, it can be expanded in terms of Krawtchouk polynomials:

$$f_1(|x|) = \sum_{k=0}^n a_k K_k(|x|),$$

where each $K_k(\cdot)$ is a degree k univariate polynomial. Note that $\hat{K}_k(\omega) \neq 0$ iff $|\omega| = k$. Thus, constraint $\text{supp } \hat{f} \subset B_{r_2}$ is equivalent to requiring $a_k = 0$ for $k > r_2$. Thus, we conclude that f_1 on integers inside $[0, n]$ coincides with a degree r_2 polynomial, and hence has $\leq r_2$ zeros. Thus, $r_1 \geq n - r_2$ as claimed. \square

4.3. Discussion

To start the discussion, let us recall the function

$$R_{LP1}(\delta) \triangleq h\left(\frac{1}{2} - \sqrt{\delta(1-\delta)}\right), \quad (69)$$

which is known as the first linear-programming (LP1) bound [26]. Its importance is in that it gives an upper bound $2^{nR_{LP1}(\delta)+o(n)}$ on the number of points in Hamming space $\{0, 1\}^n$ that have pairwise distance exceeding $n\delta$. In the range $\delta \gtrsim 0.28$ this bound is the best known to date, whereas for smaller δ it is superseded by the second linear-programming bound [26].

It is instructive, next, to provide an equivalent statement of Theorem 9.

Theorem 11 (Restatement of the uncertainty principle). *For any $\delta < 1/2$ and $0 < E < R_{LP1}(\delta)$ there is $\epsilon > 0$ with the following property. Let $f(x_1, \dots, x_n)$ be polynomial of total degree at most δn . Then, for any $S \subset \{\pm 1\}^n$ of size $|S| \leq e^{nE}$ we have*

$$\sum_{x \in S} f(x)^2 \leq e^{-n\epsilon} \sum_{x \in \{\pm 1\}^n} f(x)^2.$$

First, we mention that a weaker estimate with $E = (1 - 3\delta) \ln 2$ was shown by [24] by using hypercontractivity (9) similarly to [23]. Their argument can be easily tightened to yield a stronger $E = \ln 2 - 2\delta$. Both of these are weaker than the sharp estimate above.

Theorem 11 shows that any low-degree polynomial (restricted to the hypercube) smears its L_2 -norm so evenly that one needs to sum $e^{nR_{LP1}(\delta)}$ top values in order to obtain a sizable fraction of its overall L_2 -norm. It is interesting to compare this with [34] showing that any f that is a) a degree $\leq \delta n$ polynomial and b) $f \geq 0$ satisfies

$$\max_{x \in \{\pm 1\}^n} |f(x)| \leq e^{-n(\ln 2 - h(\frac{\delta}{2})) + o(n)} \sum_{x \in \{\pm 1\}^n} |f(x)|. \quad (70)$$

We conjecture that (70) holds for all $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ of degree $\leq \delta n$. This could be called an L_1 -version of the uncertainty principle. If true, it would imply that the sum of any $e^{nR_{Ham}(\delta)}$, $R_{Ham}(\delta) = \ln 2 - h(\frac{\delta}{2})$ values of $|f(x)|$ is negligible compared to the sum over all of $\{\pm 1\}^n$.

Finally, we discuss to what extent one can relax condition that $\Sigma = B_{\rho_2 n}$ in Theorem 9. First, notice that clearly the same conclusion holds for Σ which is an image of a ball $B_{\rho_2 n}$ under a linear isomorphism $\mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$. This provides a wealth of examples of Σ that are less “contiguous” than $B_{\rho_2 n}$.

At the same time, we cannot extend Theorem 9 to Σ being an arbitrary subset of the same cardinality as $B_{\rho_2 n}$ (recall that $|B_{\rho_2 n}| = e^{nh(\rho_2) + o(n)}$). Indeed, a simple computation shows that when S and Σ are linear subspaces of \mathbb{F}_2^n we have

$$\cos \angle(V_S, \hat{V}_\Sigma) = \sqrt{\frac{|\Sigma \cap S^\perp|}{|S^\perp|}}, \quad (71)$$

where $S^\perp \triangleq \{x : \langle x, v \rangle = 0, \forall v \in S\}$ is the dual of S . Thus, if we take S to be a linear subspace of dimension αn , $0 < \alpha < 1$, and $\Sigma = S^\perp$ (of dimension $(1 - \alpha)n$) and solve for ρ_1 and ρ_2 in

$$h(\rho_1) = \alpha \ln 2, h(\rho_2) = (1 - \alpha) \ln 2$$

we conclude that these ρ_1 and ρ_2 always satisfy

$$(1 - 2\rho_1)^2 + (1 - 2\rho_2)^2 > 1,$$

while from (71) we have $\cos \angle(V_S, \hat{V}_\Sigma) = 1$.

Consequently, we leave open the question of determining the more general uncertainty principle, i.e. characterizing the best pairs (E_1, E_2) for which one can prove implication

$$|S| \leq e^{nE_1}, |\Sigma| \leq e^{nE_2} \implies \cos \angle(V_S, \hat{V}_\Sigma) \leq \epsilon.$$

A partial result easily follows from the Hausdorff-Young inequality:

Proposition 12. *For any $E_1, E_2 \in (0, \ln 2)$ satisfying $E_1 + E_2 < \ln 2$ there exist $\epsilon > 0$ and n_0 such that for all $n \geq n_0$, all $S, \Sigma \subset \mathbb{F}_2^n$ with $|S| = e^{nE_1}$, $|\Sigma| = e^{nE_2}$ we have*

$$\cos \angle(V_S, \hat{V}_\Sigma) \leq 1 - \epsilon.$$

Conversely, for any positive integers $k_1, k_2 \leq n$ such that $k_1 + k_2 \geq n$ there exist $|S| = 2^{k_1}$ and $|\Sigma| = 2^{k_2}$ such that

$$\cos \angle(V_S, \hat{V}_\Sigma) = 1.$$

Proof. Second part follows from (71). For the first part, let $\theta = \cos \angle(V_S, \hat{V}_\Sigma)$ and $E_1 + E_2 = \ln 2 - \delta$ for $\delta > 0$. We will show that

$$\theta^2 \leq 1 - \frac{\delta - \frac{1}{n} \ln 2}{\ln 2 - \max(E_1, E_2)}. \quad (72)$$

Without loss of generality, suppose $E_2 \geq E_1$. Recall that a simple consequence of the Hausdorff-Young inequality is the Hirschmann (or entropic) uncertainty principle [39, Exercise 4.2.10]: For any $f : \mathbb{F}_2^n \rightarrow \mathbb{R}$ we have

$$\frac{\text{Ent}(f^2)}{\mathbb{E}[f^2]} + \frac{\text{Ent}(\hat{f}^2)}{\mathbb{E}[\hat{f}^2]} \leq \ln |\mathbb{F}_2^n| = n \ln 2. \quad (73)$$

Thus, taking f supported on S we estimate (from Jensen's inequality)

$$\frac{\text{Ent}(f^2)}{\mathbb{E}[f^2]} \geq n \ln 2 - \ln |S| = n(\ln 2 - E_1). \quad (74)$$

Suppose that $\frac{\mathbb{E}[\hat{f}^2 1_\Sigma]}{\mathbb{E}[\hat{f}^2]} = \theta^2$, and introduce a random variable U taking values in \mathbb{F}_2^n with

$$\mathbb{P}[U = u] \triangleq \frac{\hat{f}^2(u)}{\sum_\omega \hat{f}^2(\omega)}.$$

Then, we have $\frac{\text{Ent}(\hat{f}^2)}{\mathbb{E}[\hat{f}^2]} = n \ln 2 - H(U)$, with $H(\cdot)$ denoting the Shannon entropy. Introducing also $T = 1\{U \in \Sigma\}$ we get by the chain rule

$$n \ln 2 - \frac{\text{Ent}(\hat{f}^2)}{\mathbb{E}[\hat{f}^2]} = H(U) = H(U, T) = H(T) + H(U|T) \quad (75)$$

$$\leq \ln 2 + \theta^2 \ln |\Sigma| + (1 - \theta^2) \ln |\Sigma^c| \quad (76)$$

$$\leq \ln 2 + n(\theta^2 E_2 + (1 - \theta^2) \ln 2). \quad (77)$$

Altogether, from (73), (74) and (77) we get (72). \square

4.4. A similar result for Euclidean space

It is interesting to observe that a result analogous to Theorem 9 in \mathbb{R}^n with Lebesgue measure follows from the sharp form of Young's inequality [5]. This provokes us to hypothesize that the refined hypercontractivity result on the hypercube (Theorem 7) could play the role of the sharp Young inequality (or Babenko-Beckner inequality [2, 5]) in \mathbb{R}^n .

Notation: In this section we define $B_r = \{x \in \mathbb{R}^n : \|x\| \leq r\}$, $\|x\|^2 = (x, x)$, $(x, y) = \sum_{k=1}^n x_k y_k$, $|S|$ – the Lebesgue measure of S , $\|f\|_p = (\int_{\mathbb{R}^n} |f(x)|^p dx)^{\frac{1}{p}}$ and for $f \in L_1 \cap L_2$

$$\hat{f}(\omega) = \int_{\mathbb{R}^n} e^{-2\pi i(\omega, x)} f(x) dx, \quad \omega \in \mathbb{R}^n,$$

with the standard extension by continuity to all of $f \in L_2$.

Theorem 13. *For any $\rho_1, \rho_2 > 0$ satisfying*

$$\rho_1 \rho_2 < \frac{1}{4\pi} \tag{78}$$

there exist an $\epsilon > 0$ and n_0 such that for any $n \geq n_0$, any $S \subset \mathbb{R}^n$ with $|S| = |B_{\rho_1 \sqrt{n}}|$ and $\Sigma = B_{\rho_2 \sqrt{n}}$ we have

$$\cos \angle(V_S, \hat{V}_\Sigma) \leq e^{-n\epsilon}, \tag{79}$$

where V_S, \hat{V}_Σ are defined in (51)-(52).

Conversely, for $\rho_1, \rho_2 \geq 0$ satisfying

$$\rho_1 \rho_2 > \frac{1}{4\pi} \tag{80}$$

there exist $\epsilon > 0$ and n_0 such that for all $n \geq n_0$ we have

$$\cos \angle(V_S, \hat{V}_\Sigma) \geq 1 - e^{-n\epsilon}, \quad S = B_{\rho_1 \sqrt{n}}, \Sigma = B_{\rho_2 \sqrt{n}}. \tag{81}$$

Remark 6. *Recall that a standard Heisenberg-Weyl uncertainty (in dimension 1) states that for all $f : \mathbb{R} \rightarrow \mathbb{R}$ with $\int f^2 = 1$ we have*

$$\left(\int x^2 f^2(x) \right) \left(\int \omega^2 \hat{f}^2(\omega) \right) \geq \frac{1}{16\pi^2}.$$

So the product of mean-square widths of f and \hat{f} should exceed $\frac{1}{4\pi}$, in accord with our estimate.

Proof. Since the statement is asymptotic, we will use the standard fact

$$\ln |B_1| = \frac{n}{2} \ln \frac{2\pi e}{n} - \frac{1}{2} \ln(\pi n) + O\left(\frac{1}{n}\right)$$

and thus

$$\ln |B_{\rho \sqrt{n}}| = \frac{n}{2} \ln(2\pi e \rho^2) + O(\ln n). \tag{82}$$

To prove the second part, consider the Fourier pair (for any $\sigma > 0$):

$$f(x) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\|x\|^2}{2\sigma^2}} \tag{83}$$

$$\hat{f}(\omega) = e^{-2\pi^2 \sigma^2 \|\omega\|^2}. \tag{84}$$

Choose $\sigma > 0$ so that $\rho_1 > \frac{\sigma}{\sqrt{2}}$ and $\rho_2 > \frac{1}{\sqrt{8\sigma\pi}}$ (which is possible due to (80)). From concentration of Gaussian measure, it is easy to check that for some $\epsilon > 0$ we have

$$\|f1_{B_{\rho_1\sqrt{n}}}\|_2 \geq (1 - e^{-n\epsilon})\|f\|_2 \quad (85)$$

$$\|\hat{f}1_{B_{\rho_2\sqrt{n}}}\|_2 \geq (1 - e^{-n\epsilon})\|\hat{f}\|_2 \quad (86)$$

and therefore (81) follows from (57).

For the first part, recall a sharp form of the Young inequality on \mathbb{R}^n (from [5])

$$\|f * g\|_r \leq \left(\frac{C_p C_q}{C_r}\right)^n \|f\|_p \|g\|_q, \quad C_s = e^{\frac{1}{2}(\frac{\ln s}{s} + \frac{s-1}{s} \ln(1-s^{-1}))} \quad (87)$$

valid for $1 \leq p, q, r \leq \infty$ and $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. Consider the heat semigroup

$$e^{t\Delta} f \triangleq f * \phi_t, \quad \phi_t(x) = \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{\|x\|^2}{4t}}.$$

Let $\gamma > 0$ be a constant to be specified later, and for a real $t < \frac{1}{2\gamma}$ we set $\frac{1}{p(t)} = \frac{1}{2} - \gamma t$. Then apply (87) with $r = p(t), p = 2$ and $q = q(t)$ given by $\frac{1}{q(t)} = 1 - \gamma t$ to get, after some calculations, a hypercontractive inequality

$$\|e^{t\Delta} f\|_{p(t)} \leq e^{nE(t)} \|f\|_2, \quad (88)$$

where

$$E(t) = \frac{\gamma t}{2} \ln \frac{\gamma}{\pi e^2} + o(t), \quad t \rightarrow 0.$$

Now, we proceed as in the proof of Theorem 9 with (88) replacing the use of the more precise hypercontractivity for the cube.

Namely, we define the ball-multiplier operator

$$\widehat{\Pi_r} f(\omega) = \hat{f}(\omega) 1_{B_r}(\omega).$$

Now consider a function f supported on S and note the chain

$$\|\widehat{\Pi_r} f\|_2^2 = \langle \widehat{\Pi_r} f, f \rangle \leq e^{4\pi^2 r^2 t} \langle e^{t\Delta} f, f \rangle \quad (89)$$

$$\leq e^{4\pi^2 r^2 t} \|f\|_{q(t)} \|e^{t\Delta} f\|_{p(t)} \quad (90)$$

$$\leq e^{4\pi^2 r^2 t + nE(t)} \|f\|_2 \|f\|_{\tilde{q}(t)} \quad (91)$$

$$\leq e^{4\pi^2 r^2 t + nE(t) + \gamma t \ln |S|} \|f\|_2^2, \quad (92)$$

where in (89) we used the fact that $\widehat{e^{t\Delta} f}(\omega) = e^{-4\pi^2 \|\omega\|^2 t} \hat{f}(\omega)$, in (90) we used Hölder's inequality with $\tilde{q}(t)$ denoting the conjugate of $p(t)$, (91) is by (88), and (92) is by invoking the bound on the support of S via Hölder's inequality

$$\|f\|_{\tilde{q}} \leq \|f\|_2 |S|^{\frac{1}{\tilde{q}} - \frac{1}{2}}.$$

Taking $r = \rho_2\sqrt{n}$ and using (82) to estimate $|S|$, we conclude that

$$\|\Pi_r f\|_2 \leq e^{-n\epsilon} \|f\|_2, \quad (93)$$

whenever there is a $\gamma > 0$ such that

$$4\pi^2\rho_2^2 + \frac{\gamma}{2} \ln \frac{2\rho_1^2\gamma}{e} < 0.$$

Since $\min_{\gamma>0} \gamma \ln(a\gamma) = -\frac{1}{ea}$, we get that (93) holds whenever

$$4\pi^2\rho_2^2 - \frac{1}{4\rho_1^2} < 0,$$

which is equivalent to (78). \square

The structure of the proof for \mathbb{R}^n suggests that perhaps it is worthwhile to look for a general inequality on the hypercube that could replace the use of hypercontractivity in the proof of Theorem 9, i.e. play a role similar to that of the sharp Young inequality on \mathbb{R}^n (of course, the Young inequality itself cannot be sharpened on the hypercube, or on any finite group).

To complete the parallel with the hypercube, we also note that Proposition 12 (uncertainty principle for general supports) also has an \mathbb{R}^n -analog.

Proposition 14. *For any $S, \Sigma \subset \mathbb{R}^n$ with $|S| = |B_{\rho_1\sqrt{n}}|$ and $|\Sigma| = |B_{\rho_2\sqrt{n}}|$ with*

$$\rho_1\rho_2 < \frac{1}{2\pi e} \quad (94)$$

we have $\cos(\angle V_S, \hat{V}_\Sigma) \leq e^{-n\epsilon}$.

Proof. Denoting by P_2 the operator of orthogonal projection on \hat{V}_Σ and taking $f \in V_S$ we get

$$\|P_2 f\|_2 = \left(\int_\Sigma |\hat{f}(\omega)|^2 d\omega \right)^{\frac{1}{2}} \leq \|\hat{f}\|_\infty |\Sigma|^{\frac{1}{2}} \leq \|f\|_1 |\Sigma|^{\frac{1}{2}} \leq \|f\|_2 (|S| \cdot |\Sigma|)^{\frac{1}{2}}$$

and the rest follows from (82) and (53). \square

We do not think (94) is sharp. In fact, it is natural to conjecture that the sharp constant in (94) should be $\frac{1}{4\pi}$, that is that the pair of balls present the worst case for the uncertainty principle. For the latter, see also the discussion in [22, Section I].

5. Application: lower bound on spectrum of sparse Boolean functions

In this section we will use base-2 binary entropy defined as

$$h_2(\rho) = -\rho \log_2(\rho) - (1 - \rho) \log_2(1 - \rho),$$

and denote $h_2^{-1} : [0, 1] \mapsto [0, 1/2]$ its functional inverse.

Consider a sparse Boolean function $\Psi : \mathbb{F}_2^k \rightarrow \{0, 1\}$ with $|\text{supp } \Psi| = n$. It is clear that every Fourier coefficient of Ψ satisfies:

$$\hat{\Psi}(\omega) = \sum_{x \in \mathbb{F}_2^k} (-1)^{\langle \omega, x \rangle} \Psi(x) \in [-n, n].$$

What we show below is that Fourier coefficients $\hat{\Psi}(\omega)$ are large ($\Theta(n)$) even for large frequencies ω , i.e. $|\omega| \geq ck$, where $|\cdot|$ denotes Hamming weight.

Theorem 15. *Fix $\tau, \rho'_1, \rho'_2 > 0$ such that*

$$(1 - 2\rho'_1)^2 + (1 - 2\rho'_2)^2 > 1$$

There exists $\delta_k \rightarrow 0$ such that for k , and every Boolean function $\Psi : \mathbb{F}_2^k \rightarrow \{0, 1\}$ with $|\text{supp } \Psi| = n$, $n \in [\tau k, k/\tau]$ we have

$$\max_{\omega: |\omega| \geq \rho'_2 k} \hat{\Psi}(\omega) \geq n(2\sqrt{\rho_1(1 - \rho_1)} + \delta_k),$$

where $\rho_1 = h_2^{-1}(h_2(\rho'_1) \frac{k}{n})$.

Proof. It will turn out to be more convenient to prove this estimate in the language of linear maps, which we will do in the next section. Here we notice how to convert to that statement. Given Ψ define operator A via $Ah = \Psi * h$, with $*$ denoting convolution. Define also numbers d_r via

$$n - 2d_r = \max_{\omega: |\omega| \geq r} \hat{\Psi}(\omega).$$

(see (97) for an equivalent definition). Then the proof of Theorem 16, or more exactly (99), shows the stated bound. \square

5.1. Restatement as a property of linear maps (coding theory)

We now restate the previous result as a curious property of linear maps between binary spaces.

Theorem 16. *For any $0 < R' < R < 1$ there exists $\delta_n \rightarrow 0$ such that for any linear map $f : \mathbb{F}_2^k \rightarrow \mathbb{F}_2^n$ with $\frac{k}{n} = R$ there exists an $x \in \mathbb{F}_2^k$ s.t.*

$$\frac{1}{n}|f(x)| \leq \delta_{LP1}(R') + \delta_n \tag{95}$$

$$\frac{1}{k}|x| \geq \delta_{LP1}\left(\frac{R'}{R}\right) - \delta_n, \tag{96}$$

where $\delta_{LP1}(h_2(\rho)) = \frac{1}{2} - \sqrt{\rho(1 - \rho)}$ is the inverse of the earlier $R_{LP1}(\delta)$ function in (69).

Remark 7. *This estimate significantly outperforms previously best known bounds of this kind [32, Theorem 1], but only applies to linear maps.*

We give two different proofs, in two subsections below. Note that the two proofs take slightly different points of view. The first proof deals with linear maps $f : \mathbb{F}_2^k \rightarrow \mathbb{F}_2^n$, while the second proof looks rather at images of these maps, linear codes in \mathbb{F}_2^n . In particular, in the second proof we assume that the image of f is of dimension k (i.e. f is of full rank).

5.2. Method 1 – graph covers

Proof. To every linear map $f : \mathbb{F}_2^k \rightarrow \mathbb{F}_2^n$ we associate the following increasing sequence of numbers:

$$d_r(f) \triangleq \min\{|f(x)| : |x| \geq r\},$$

where d_1 is just the minimum distance of f . Note that, as in coding theory, we think of elements of \mathbb{F}_2^k and \mathbb{F}_2^n as row-vectors and thus map f can be represented as a binary $k \times n$ matrix, whose columns we denote by c_1, \dots, c_n . Following [16] we also associate to f a Cayley graph Γ with vertices \mathbb{F}_2^k and generators $\{c_i, i = 1, \dots, n\}$. (We will use freely facts from [16], perhaps in a somewhat different formulation, from now on.) Then

$$n - 2d_r = \max \left\{ \frac{(Ah, h)}{\|h\|_2^2} : \hat{h} = 0 \text{ on ball } B(0, r-1) \right\}, \quad (97)$$

where A is the adjacency matrix of Γ . Note that A is also a convolution operator on \mathbb{F}_2^k :

$$Ah = h * \left(\sum_{i=1}^n \delta_{c_i} \right). \quad (98)$$

As in [16], select a covering map $\mathbb{F}_2^n \rightarrow \mathbb{F}_2^k$ and take $B \subset \mathbb{F}_2^n$ to be the Hamming ball of radius $n\rho_1$, with $0 < \rho_1 < \frac{1}{2}$ found as $h_2(\rho_1) = R'$. There exists a function $g_B : \mathbb{F}_2^n \rightarrow \mathbb{R}$, supported on B with the property:

$$(A_C g_B, g_B) \geq \lambda_B \|g_B\|_2^2,$$

where A_C is the adjacency matrix of the n -dimensional hypercube, and $\lambda_B = 2n\sqrt{\rho_1(1-\rho_1)} + o(n)$.

Hence, there exists a function $h_B : \mathbb{F}_2^k \rightarrow \mathbb{R}$ supported on the image of B under the covering map with the property:

$$(Ah_B, h_B) \geq \lambda_B \|h_B\|_2^2,$$

and $|\text{supp } h_B| \leq |B| = 2^{nh_2(\rho_1)+o(n)}$. Then to get a lower bound on (97) we set

$$h = h_B - \Pi_{<r} h_B,$$

where $\Pi_{<r} = \sum_{a<r} \Pi_a$ and Π_a is from (64).

Note that A and $\Pi_{<r}$ commute and eigenvalues of A are bounded by n , so $(A\Pi_{<r} h_B, h_B) = (A\Pi_{<r} h_B, \Pi_{<r} h_B) \leq n(\Pi_{<r} h_B, h_B)$. We then have:

$$(Ah, h) = (Ah_B, h_B) - (A\Pi_{<r} h_B, h_B) \geq \lambda_B \|h_B\|_2^2 - n(\Pi_{<r} h_B, h_B)$$

Thus,

$$n - 2d_r \geq 2n\sqrt{\rho_1(1 - \rho_1)} + o(n) \quad (99)$$

whenever

$$\|\Pi_{<r} h_B\|_2^2 \leq \|h_B\|_2^2 \cdot o(1).$$

Using the uncertainty principle for the k -dimensional cube (Theorem 9) we estimate

$$\|\Pi_{<r} h_B\| \ll \|h_B\|^2,$$

as long as

$$\frac{r}{k} < \frac{1}{2} - \sqrt{\rho'_1(1 - \rho'_1)}, \quad (100)$$

where ρ'_1 is found from $h_2(\rho'_1) = \frac{h_2(\rho_1)}{R} = \frac{R'}{R}$. After simple algebra, we see that (99)-(100) are equivalent to (95)-(96). \square

5.3. Method 2 – analytic

We start an uncertainty-type claim for subspaces of \mathbb{F}_2^n .

Let C be a k -dimensional linear subspace C of \mathbb{F}_2^n . Given a basis $\mathbf{v} = \{v_1, \dots, v_k\}$ of C , denote the length of representation of a vector $x \in C$ in terms of V by $|x|_{\mathbf{v}}$.

Lemma 17. *Let f be a function supported on a subset $A \subseteq \mathbb{F}_2^n$. Let $0 \leq r \leq k \leq n$ be integer parameters such that $k \geq \log_2 |A|$, and, moreover, writing $|A| = 2^{h_2(\rho_1) \cdot k}$, $\binom{k}{r} = 2^{h_2(\rho_2) \cdot k}$, we have $(1 - 2\rho_1)^2 + (1 - 2\rho_2)^2 > 1$.*

Then, for any k -dimensional subspace C of \mathbb{F}_2^n and for any basis \mathbf{v} of C holds

$$\sum_{\omega \in C, |\omega|_{\mathbf{v}} \leq r} \widehat{f}^2(\omega) \ll \sum_{\omega \in C} \widehat{f}^2(\omega)$$

Here the \ll sign means that the LHS is exponentially smaller than the RHS.

Proof. Let $F = f * 1_{C^\perp}$.

Note that F is constant on cosets of C^\perp and that $\widehat{F}(\omega) = \begin{cases} |C^\perp| \widehat{f}(\omega) & \text{if } \omega \in C \\ 0 & \text{otherwise} \end{cases}$.

Let M be a $k \times n$ matrix with rows v_1, \dots, v_k . We define a function g on \mathbb{F}_2^k as follows. For $x \in \mathbb{F}_2^k$, the pre-image $\{y \in \mathbb{F}_2^n, My = x\}$ is a coset of C^\perp , and we set $g(x)$ to be the (fixed) value of F on this coset.

Next, we calculate the Fourier transform of g . Let $\alpha \in \mathbb{F}_2^k$. Let $\omega = \alpha^t M \in C$. We claim that $\widehat{g}(\alpha) = \widehat{f}(\omega)$. To see this, note that for any y such that $My = x$ holds $\langle x, \alpha \rangle = \langle My, \alpha \rangle = \langle y, M^t \alpha \rangle = \langle y, \omega \rangle$. Using this we compute

$$\widehat{g}(\alpha) = \sum_{x \in \mathbb{F}_2^k} g(x) (-1)^{\langle x, \alpha \rangle} = \frac{1}{|C^\perp|} \sum_{x \in \mathbb{F}_2^k} \sum_{y: My=x} F(y) (-1)^{\langle y, \omega \rangle} = \frac{1}{|C^\perp|} \widehat{F}(\omega) = \widehat{f}(\omega)$$

Next, we apply the uncertainty principle for g on \mathbb{F}_2^k . Observe that the cardinality of the support of g is given by the number of cosets of C^\perp intersecting A , which is at most $|A|$. The constraints on $|A|$, k , and r imply $\sum_{|\alpha| \leq r} \widehat{g}^2(\alpha) \ll \sum_\alpha \widehat{g}^2(\alpha)$, which is equivalent to the claim of the lemma. \square

We now prove Theorem 16, first restating it for linear codes rather than for linear maps.

Theorem 18. *Let $0 < R < 1$. Let $C \subseteq \mathbb{F}_2^n$ be a linear code of rate $k = Rn$. Let $\mathbf{v} = \{v_1, \dots, v_k\}$ be a basis of C . Then for any $0 \leq R' < R$ there is a vector $x \in C$ with*

$$\frac{1}{n}|x| \leq \delta_{LP1}(R') + \delta_n \quad \text{and} \quad \frac{1}{k}|x|_{\mathbf{v}} \geq \delta_{LP1}\left(\frac{R'}{R}\right) - \delta_n$$

Proof. Let $r = h_2^{-1}(R') \cdot n$. Let B be the Hamming ball of radius r around zero in \mathbb{F}_2^n . As in [16], let g_B be a function supported on B , with the property:

$$(A_C g_B, g_B) \geq \lambda_B \|g_B\|_2^2,$$

where A_C is the adjacency matrix of the n -dimensional hypercube, and $\lambda_B = 2n\sqrt{\frac{r}{n}(1 - \frac{r}{n})} + o(n)$.

Let $d = \frac{n - \lambda_B + 1}{2}$. Note that $d = \left(\frac{1}{2} - \sqrt{h_2^{-1}(R')(1 - h_2^{-1}(R'))}\right) \cdot n + o(n) = \delta_{LP1}(R') \cdot n + o(n)$.

Note that $|B| = 2^{R' \cdot n} \leq |C| = 2^k$. We introduce two additional parameters with a view towards using Lemma 17. Let ρ_1 be such that $|B| = 2^{h_2(\rho_1) \cdot k}$, and let ρ_2 satisfy $(1 - 2\rho_1)^2 + (1 - 2\rho_2)^2 = 1$. Computing explicitly,

$$\rho_1 = h_2^{-1}\left(\frac{R'}{R}\right) \quad \text{and} \quad \rho_2 = \delta_{LP1}\left(\frac{R'}{R}\right)$$

We proceed with the following computation, as in [29]. Let $F = |C| \cdot g_B * 1_{C^\perp}$. Compute $(A_C F, F)$ in two ways. On one hand, since A_C commutes with convolutions, we have $(A_C F, F) \geq \lambda_B \cdot (F, F) = \frac{\lambda_B}{2^n} \sum_x \widehat{F}^2(x)$. On the other hand, observe that $A_C F = F * (\sum_{i=1}^n \delta_{e_i})$, where e_i is the i^{th} unit vector (compare with (98)). Hence $\widehat{A_C F}(x) = \widehat{F}(x) \cdot \left(\sum_{i=1}^n \widehat{\delta_{e_i}}(x)\right) = \widehat{F}(x) \cdot \left(\sum_{i=1}^n (-1)^{x_i}\right) = (n - 2|x|) \cdot \widehat{F}(x)$. And therefore $(A_C F, F) = \frac{1}{2^n} \sum_x (n - 2|x|) \cdot \widehat{F}^2(x)$. Substituting $\lambda_B = n - 2d + 1$, we get the inequality $\sum_x (n - 2|x|) \cdot \widehat{F}^2(x) \geq (n - 2d + 1) \cdot \sum_x \widehat{F}^2(x)$. Rearranging and simplifying, this implies that $2d \cdot \sum_{|x| \leq d} \widehat{F}^2(x) \geq \sum_x \widehat{F}^2(x)$.

Since $\widehat{F} = |C| \cdot \widehat{g_B} \cdot \widehat{1_{C^\perp}} = 2^n \cdot \widehat{g_B} \cdot 1_C$, we deduce that

$$2d \cdot \sum_{x \in C, |x| \leq d} \widehat{g_B}^2(x) \geq \sum_{x \in C} \widehat{g_B}^2(x).$$

We now apply Lemma 17 for g_B . By the lemma, we can choose a sequence $\delta_n \rightarrow 0$, so that for $r < (\rho_2 - \delta_n) \cdot k$ holds

$$\sum_{x \in C, |x|_{\mathbf{v}} > r} \widehat{g}_B^2(x) > \left(1 - \frac{1}{2d}\right) \cdot \sum_{x \in C} \widehat{g}_B^2(x).$$

Combining these two inequalities, we deduce that $\sum_{x \in C, |x| \leq d, |x|_{\mathbf{v}} > r} \widehat{g}_B^2(x) > 0$, implying that there exists a vector $x \in C$ such that $|x| \leq d$ and $|x|_{\mathbf{v}} > r$, proving the claim of the theorem. \square

Acknowledgement

The work of Y.P. was supported (in part) by the National Science Foundation under Grant No CCF-13-18620, and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370. The work of A. S. was supported (in part) by grants from the US-Israel Binational Science Foundation and from the Israel Science Foundation.

References

- [1] W. Amrein and A. Berthier. On support properties of l_p -functions and their Fourier transforms. *J. Func. Anal.*, 24(3):258–267, 1977.
- [2] K. Babenko. An inequality in the theory of Fourier integrals. *Izv. Akad. Nauk SSSR, Ser. Mat.*, 25:531–542, 1961.
- [3] D. Bakry. L’hypercontractivité et son utilisation en théorie des semi-groupes. In *Lectures on probability theory*, pages 1–114. Springer, 1994.
- [4] D. Bakry. Functional inequalities for Markov semigroups. In *Probability measures on groups*, pages 91–147. Tata Institute of Fundamental Research, Mumbai, 2006.
- [5] W. Beckner. Inequalities in Fourier analysis. *Ann. Math.*, 102(1):159–182, July 1975.
- [6] W. Beckner. Pitt’s inequality and the uncertainty principle. *Proc. Amer. Math. Soc.*, 123(6):1897–1905, 1995.
- [7] M. Benedicks. On Fourier transforms of functions supported on sets of finite Lebesgue measure. *J. Math. Anal. Appl.*, 106(1):180–183, 1985.
- [8] S. G. Bobkov and M. Ledoux. On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures. *Journal of functional analysis*, 156(2):347–365, 1998.

- [9] S. G. Bobkov and P. Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, Jun 2006.
- [10] A. Bonami. Étude des coefficients de Fourier des fonctions de $l_p(g)$. *Ann. Inst. Fourier (Grenoble)*, 20(2):335–402, 1970.
- [11] E. A. Carlen. Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 101(1):194 – 211, 1991.
- [12] M. Costa and T. Cover. On the similarity of the entropy power inequality and the Brunn-Minkowski inequality (corresp.). *IEEE Trans. Inf. Theory*, 30(6):837–839, 1984.
- [13] E. B. Davies and B. Simon. Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. *J. Func. Anal.*, 59(2):335–395, 1984.
- [14] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.
- [15] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Archive for Rational Mechanics and Analysis*, 206(3):997–1038, 2012.
- [16] J. Friedman and J.-P. Tillich. Generalized Alon–Boppana theorems and error-correcting codes. *SIAM Journal on Discrete Mathematics*, 19(3):700–718, 2005.
- [17] W. H. J. Fuchs. On the magnitude of Fourier transforms. In *Proc. Int. Math. Cong., Amsterdam*, pages 106–107, 1954.
- [18] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97:1061–1083, 1975.
- [19] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities*. Cambridge University Press, 1988.
- [20] P. Hartman. *Ordinary differential equation*. John Wiley & Sons, New York, USA, 1964.
- [21] V. Havin and B. Jöricke. *The uncertainty principle in harmonic analysis*. Springer, 1994.
- [22] P. Jaming. Nazarov’s uncertainty principles in higher dimension. *J. Approximation Th.*, 149(1):30–41, 2007.
- [23] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *Proc. 29th Ann. Symp. on Foundations of Comp. Sci.*, pages 68–80, Los Alamitos, CA, 1988.

- [24] J. Kahn and R. Meshulam. Uncertainty inequalities on Hamming cubes. manuscript.
- [25] H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty– II. *Bell Syst. Tech. J.*, 40(1):65–84, 1961.
- [26] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch. New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE Trans. Inf. Theory*, 23(2):157–166, 1977.
- [27] L. Miclo. Une majoration sous-exponentielle pour la convergence de l’entropie des chaînes de Markov à trou spectral. *Ann. Inst. H. Poincaré Probab. Statist.*, 35(3):261–311, 1999.
- [28] E. Mossel, K. Oleszkiewicz, and A. Sen. On reverse hypercontractivity. *Geometric and Functional Analysis*, 23(3):1062–1097, 2013.
- [29] M. Navon and A. Samorodnitsky. Linear programming bounds for codes via a covering argument. *Discrete & Computational Geometry*, 41(2):199–207, 2009.
- [30] F. L. Nazarov. Local estimates for exponential polynomials and their applications to inequalities of the uncertainty principle type. *Algebra i Analiz (in Russian)*, 5(4):3–66, 1993.
- [31] E. Nelson. A quartic interaction in two dimensions. In R. Goodman and I. Segal, editors, *Mathematical Theory of Elementary Particles*, Cambridge, MA, 1966. M.I.T. Press.
- [32] Y. Polyanskiy. On metric properties of maps between Hamming spaces and related graph homomorphisms. *J. Combin. Theory Ser. A*, 145:227–251, 2017.
- [33] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics*, pages 301–413. Springer, 1997.
- [34] A. Samorodnitsky. Extremal properties of solutions for Delsarte’s linear program. *preprint*.
- [35] A. Samorodnitsky. A modified logarithmic Sobolev inequality for the Hamming cube and some applications. *arXiv preprint arXiv:0807.1679*, 2008.
- [36] D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty – I. *Bell Syst. Tech. J.*, 40(1):43–63, 1961.
- [37] A. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Contr.*, 2(2):101–112, 1959.
- [38] D. W. Stroock. *An introduction to the theory of large deviations*. Universitext, Springer-Verlag, New York, 1984.

- [39] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.
- [40] N. T. Varopoulos. Hardy-Littlewood theory for semigroups. *J. Functional Analysis*, 63(2):240–260, 1985.
- [41] F. B. Weissler. Logarithmic Sobolev inequalities for the heat-diffusion semigroup. *Trans. Amer. Math. Soc.*, 237:255–269, 1978.
- [42] H. Witsenhausen. Entropy inequalities for discrete channels. *IEEE Trans. Inf. Theory*, 20(5):610–616, 1974.
- [43] A. D. Wyner and J. Ziv. A theorem on the entropy of certain binary sequences and applications–I. *IEEE Trans. Inf. Theory*, 19(6):769–772, Nov. 1973.