

ON THE FAILURE OF CONCENTRATION FOR THE ℓ_∞ -BALL

BY

TIM AUSTIN*

*Courant Institute, New York University
251 Mercer St, New York, NY 10012, USA
e-mail: tim@cims.nyu.edu*

ABSTRACT

Let (X, d) be a compact metric space and μ a Borel probability on X . For each $N \geq 1$ let d_∞^N be the ℓ_∞ -product on X^N of copies of d , and consider 1-Lipschitz functions $X^N \rightarrow \mathbb{R}$ for d_∞^N .

If the support of μ is connected and locally connected, then all such functions are close in probability to juntas: that is, functions that depend on only a few coordinates of X^N . This describes the failure of measure concentration for these product spaces, and can be seen as a Lipschitz-function counterpart of the celebrated result of Friedgut that Boolean functions with small influences are close to juntas.

1. Introduction

In this paper, a **metric probability space** will be a triple (X, d, μ) in which d is a compact metric on X and μ is a probability measure on the Borel σ -algebra of (X, d) .

Given a metric space (X, d) and an integer $N \geq 1$, we shall write d_∞^N for the ℓ_∞ -product of N copies of d : that is,

$$d_\infty^N((x_n)_{n \leq N}, (x'_n)_{n \leq N}) := \max_{n \leq N} d(x_n, x'_n).$$

If μ is a probability measure on X , then μ^N denotes its N -fold product on X^N .

* Research supported by a fellowship from the Clay Mathematics Institute.

Received October 22, 2013 and in revised form June 23, 2014

Recall that a sequence of metric probability spaces (X_N, d_N, μ_N) exhibits **concentration of measure** if

$$(1) \quad \sup_{\substack{f : X_N \rightarrow \mathbb{R} \\ \text{Lip}(f) \leq 1}} \mu_N\{|f - \int f d\mu_N| \geq \delta\} \rightarrow 0 \quad \forall \delta > 0$$

as $N \rightarrow \infty$ (see, for instance, [Led01] and the many references there).

Unless μ is a Dirac mass, the sequence of metric probability spaces

$$(X^N, d_\infty^N, \mu^N)$$

does not exhibit concentration of measure. Each coordinate projection $\pi_n : X^N \rightarrow X$ is 1-Lipschitz and pushes μ^N to μ . Composing one of these with any non-constant 1-Lipschitz function $f : X \rightarrow \mathbb{R}$ witnesses that (1) is violated, because for small enough $\delta > 0$ the supremum on the left is bounded below by the positive constant $\mu\{|f - \int f d\mu| \geq \delta\}$, independently of N .

In light of this, one can ask for a rough description of all functions $X^N \rightarrow \mathbb{R}$ that have a fixed Lipschitz constant but are not highly concentrated for the measure μ^N . If the support of μ , $\text{spt } \mu$, is disconnected, then there are many such functions, and no special description is conceivable: for instance, if $X = \{0, 1\}$, $d(0, 1) = 1$ and $\mu = \frac{1}{2}(\delta_0 + \delta_1)$, then all functions $X^N \rightarrow [0, 1]$ are 1-Lipschitz for the metric d_∞^N . This observation may easily be extended to other examples with disconnected $\text{spt } \mu$.

However, the situation changes in case $\text{spt } \mu$ has some connectedness. This will be made precise in Theorem 1.1 below, which will involve the following notions.

First, (X, d, μ) is **(locally) connected** if $\text{spt } \mu$ is (locally) connected.

Next, when a probability space (X, μ) is understood and $n \in [N]$, let $E_{[N] \setminus \{n\}}$ denote the conditional expectation operator on $L_1(\mu^N)$ defined by integrating out the n^{th} coordinate:

$$E_{[N] \setminus \{n\}} f(x_1, \dots, x_N) := \int_X f(x_1, \dots, x_{n-1}, y, x_{n+1}, \dots, x_N) \mu(dy).$$

More generally, for $S = [N] \setminus \{n_1, \dots, n_k\} \subseteq [N]$ let

$$E_S := E_{[N] \setminus \{n_1\}} \circ \dots \circ E_{[N] \setminus \{n_k\}},$$

where clearly the order of composition is unimportant. In this notation, S records those coordinates in $[N]$ on which the dependence of f is retained.

Lastly, for any metric space (X, d) and function $f : X \rightarrow \mathbb{R}$, the **non-decreasing modulus of continuity** of f is the function $\omega : [0, \infty) \rightarrow [0, \infty]$ defined by

$$\omega(r) := \sup\{|f(x) - f(y)| \mid d(x, y) \leq r\}.$$

Clearly $\omega(0) = 0$, and uniform continuity of f is equivalent to continuity of ω at 0.

THEOREM 1.1: *Let (X, d, μ) be a connected and locally connected metric probability space, and let $\omega : [0, \infty) \rightarrow [0, \infty]$ be non-decreasing. For every $\varepsilon > 0$ there is some integer $p \geq 1$, depending on X, ε and ω , with the following property. For every $N \geq 1$, if $f : X^N \rightarrow \mathbb{R}$ has modulus of continuity at most ω for the metric d_∞^N , then there is $S \subseteq [N]$ with $|S| \leq p$ such that*

$$\|f - E_S f\|_{L_1(\mu^N)} < \varepsilon.$$

I suspect that this theorem needs only the assumption of connectedness, not local connectedness.

Approximation in $L_1(\mu^N)$ is convenient for the proof, but it implies an approximation in $L_p(\mu^N)$ for any $p < \infty$: this is because

$$\text{diam}(X^N, d_\infty^N) = \text{diam}(X, d) < \infty$$

for all N , and so the modulus-of-continuity bound implies also a uniform bound on $|f - \int f|$. However, Theorem 1.1 cannot be tightened to an approximation in $L_\infty(\mu^N)$: for instance, on $[0, 1]^N$ with Lebesgue measure, the function

$$f(x_1, \dots, x_N) := \max_{n \leq N} x_n$$

is 1-Lipschitz, takes values in $(1 - \varepsilon, 1]$ with probability $1 - (1 - \varepsilon)^N$, but is not uniformly close to any function depending on only $N - 1$ coordinates.

Let $\mathbb{T} := \mathbb{R}/\mathbb{Z}$, endowed with the metric inherited from the usual distance on \mathbb{R} , which we still denote by $|\cdot|$. Most of the work in proving Theorem 1.1 will go towards the following special case:

THEOREM 1.2: *For every $\varepsilon > 0$ there is a $q \geq 1$ such that if $f : \mathbb{T}^N \rightarrow \mathbb{R}$ is 1-Lipschitz for the metric $|\cdot|_\infty^N$, then there is some $S \subseteq [N]$ with $|S| \leq q$ such that*

$$\|f - E_S f\|_{L_1(\mathbb{T}^N)} < \varepsilon.$$

An easy exercise shows that with the above metrics on product spaces, any conditional expectation of the form $E_S f$ has modulus of continuity not greater than that of f itself. Therefore both of these theorems give approximations in measure by juntas that are ‘as continuous’ as the original function f .

In fact, the method below will give a stronger variant of Theorem 1.2, formulated in terms of a kind of Sobolev norm for f : see Theorem 2.1. This stronger version also has a consequence for maps between Hamming metrics, rather than ℓ_∞ -metrics. Recall that given (X, d) , the associated **Hamming metric** d_1^N on X^N is defined by

$$d_1^N((x_n)_{n \leq N}, (x'_n)_{n \leq N}) := \frac{1}{N} \sum_{n=1}^N d(x_n, x'_n).$$

Clearly $d_1^N \leq d_\infty^N$. The additional consequence of Theorem 2.1 is as follows.

THEOREM 1.3: *Let $N, M \geq 1$ with $M/N =: \alpha$. Then for every $\varepsilon, L > 0$ there is a $q \geq 1$, depending only on ε, α and L , with the following property. If $F : [0, 1]^N \rightarrow [0, 1]^M$ is L -Lipschitz for the Hamming metrics on domain and target, then there is another L -Lipschitz function*

$$G = (G_1, \dots, G_M) : [0, 1]^N \rightarrow [0, 1]^M$$

such that

$$\int_{[0,1]^N} |F(x) - G(x)|_1^M dx < \varepsilon,$$

and such that each G_m depends on only q coordinates of $[0, 1]^N$.

The proof of Theorem 1.2 is a close relative of arguments of Kahn, Kalai and Linial [KKL], Bourgain, Kahn, Kalai, Katznelson and Linial [BKK⁺92], Friedgut and Kalai [FK96], and Friedgut [Fri98, Fri04], used for the analysis of Boolean functions on product spaces under various conditions. That connection will be discussed further during the course of the proof. Perhaps the closest predecessor of Theorems 1.1 and 1.2 is the theorem of Friedgut that a Boolean function on $\{0, 1\}^N$ with controlled total influences is close, in the uniform measure, to a Boolean function depending on a controlled number of coordinates: see [Fri98].

Theorems 1.1 and 1.2 also have relatives in other studies of concentration for product measures and ℓ_∞ -metrics. Several earlier works have sought conditions on a metric probability space (X, μ) under which the isoperimetric function of

(X^N, μ^N, d_∞^N) may be either estimated or determined exactly, and some quite general results are now known: see, for instance, Barthe [Bar04] and the reference given there. However, those results generally concern the strictly extremal behaviour of the isoperimetric problem in these spaces, for situations in which the minimizing sets can be described exactly, or in which that problem can be shown to behave well under tensorizing. By contrast, the isoperimetric result descending from Theorem 1.1 is very crude:

THEOREM: *For any connected and locally connected (X, d, μ) , and any $\delta, \varepsilon > 0$, there is some q , depending on the space and on δ and ε , such that the following holds. If $A, B \subseteq X^N$ are such that*

$$\inf\{d_\infty^N(x, y) \mid x \in A, y \in B\} \geq \delta,$$

then there are a set $S \subseteq [N]$ with $|S| \leq q$, and subsets $A', B' \subseteq X^N$ that depend only on coordinates in S , such that $\mu^N(A \setminus A'), \mu^N(B \setminus B') < \varepsilon$ and

$$\inf\{d_\infty^N(x, y) \mid x \in A', y \in B'\} \geq \delta - \varepsilon.$$

Thus, in a sense, the approximate isoperimetric problem for ℓ_∞ -product spaces may be confined to low-dimensional products.

Sketch proof. The 1-Lipschitz function $f(x) := \min\{\delta, \text{dist}(x, A)\}$ has the property that $f|_A = 0$ and $f|_B = \delta$. Applying Theorem 1.1 to f gives a set S of controlled size such that $f \approx E_S f$, and now one can read off A' and B' as suitable level sets of $E_S f$. ■

This argument is likely to give a very poor quantitative dependence, and the further details are routine, so they will be omitted. It is also similar to part of the argument of Dinur, Friedgut and Regev in [DFR08] concerning the structure of independent sets in graph powers.

ACKNOWLEDGEMENT. An earlier version of this paper was significantly strengthened following suggestions by the anonymous referee.

2. Proof of the special case

For a differentiable function $f : \mathbb{T}^N \rightarrow \mathbb{R}$ and $n \leq N$, let $\partial_n f$ be the partial derivative in the n^{th} coordinate, and let

$$\vec{\partial} f := (\partial_n f)_{n=1}^N : \mathbb{T}^N \rightarrow \mathbb{R}^N.$$

At any point $x \in \mathbb{T}^N$ one may identify the tangent space with \mathbb{R}^N in the obvious way. As one zooms in on that point the metric d_∞^N converges to the norm of ℓ_∞^N . From this it follows easily that

$$(2) \quad \text{Lip}(f) = \sup_{x \in \mathbb{T}^N} \|\vec{\partial}f(x)\|_{\ell_1^N} = \sup_{x \in \mathbb{T}^N} \left(\sum_{n=1}^N |\partial_n f(x)| \right)$$

for all $f \in \mathcal{C}^1(\mathbb{T}^N)$.

The method below actually gives a slightly stronger result than Theorem 1.2. For $p \in [1, \infty)$ and a suitably integrable function $F = (F_1, \dots, F_N) : \mathbb{T}^N \rightarrow \mathbb{R}^N$, define the norm

$$\|F\|_{L_p \ell_p^N} := \left(\int_{\mathbb{T}^N} \|F(x)\|_{\ell_p^N}^p dx \right)^{1/p} = \left(\sum_{n=1}^N \int_{\mathbb{T}^N} |F_n(x)|^p dx \right)^{1/p},$$

where the integrals are with respect to the Haar probability measure on \mathbb{T}^N . This norm will be used with p equal to 1 or 2.

From each norm $\|\cdot\|_{L_p \ell_p^N}$, one may define a seminorm on $\mathcal{C}^1(\mathbb{T}^N)$ by

$$f \mapsto \|\vec{\partial}f\|_{L_p \ell_p^N}.$$

When $p = 2$, this is the homogeneous Sobolev seminorm $\|f\|_{W^{0,1,2}(\mathbb{T}^N)}$.

Since $\|\cdot\|_{\ell_p^N} \leq \|\cdot\|_{\ell_1^N}$ for all p , any $f \in \mathcal{C}^1(\mathbb{T}^N)$ satisfies

$$\|\vec{\partial}f\|_{L_p \ell_p^N} \leq \text{Lip}(f) \quad \forall p \in [1, \infty).$$

From this and the fact that any 1-Lipschitz function on \mathbb{T}^N may be uniformly approximated by smooth 1-Lipschitz functions, the following immediately implies Theorem 1.2.

THEOREM 2.1: *For every $\varepsilon > 0$ there is a $q \geq 1$ such that if $f \in \mathcal{C}^1(\mathbb{T}^N)$ satisfies $\|\vec{\partial}f\|_{L_1 \ell_1^N} \leq 1$ and $\|f - \int f\|_2 \leq 1$, then there is some $S \subseteq [N]$ with $|S| \leq q$ such that*

$$\|f - E_S f\|_1 < \varepsilon.$$

The same result holds for functions on $[0, 1]^N$, possibly with a different dependence of q on ε .

Most of the work below will go into the first part of Theorem 2.1, and that is also the case that will imply Theorem 1.1. The result for $[0, 1]^N$ is an easy corollary, but is included here because it is needed for Theorem 1.3.

Now let $\Delta := \sum_{n \leq N} \partial_n^2$, the non-positive definite Laplacian. Let $P_t := \exp(t\Delta)$, $t \geq 0$, be the resulting heat semigroup, which is well-defined and self-adjoint on $L_2(\mathbb{T}^N)$. As is standard, this is the semigroup corresponding to Brownian motion on \mathbb{T}^N . For each $t > 0$ it may be written explicitly in terms of the Gaussian measure γ_t on \mathbb{R} with variance t :

$$(3) \quad P_t f(x) = \int_{\mathbb{R}^N} f(x + \bar{y}) \gamma_t^{\otimes N}(dy),$$

where \bar{y} denotes the image of y under the obvious quotient map $\mathbb{R}^N \rightarrow \mathbb{T}^N$.

Every P_t is a contraction for the norm $\|\cdot\|_{L_p \ell_p^N}$ for any $p \in [1, \infty]$. A simple calculation gives $\partial_n P_t f = P_t \partial_n f$ for all t, n and differentiable f , and hence

$$(4) \quad \|\vec{\partial} P_t f\|_{L_p \ell_p^N} \leq \|\vec{\partial} f\|_{L_p \ell_p^N}$$

for any p, f and $t \geq 0$.

The proof of Theorem 2.1 follows similar lines to the proofs of the main results in [KKL, BKK⁺92, FK96, Fri98, Fri04], concerning the structure of Boolean functions with low total influences on $\{0, 1\}^N$ and $[0, 1]^N$. Note, however, that the obvious analog of Theorem 1.2 for Boolean functions — that Boolean functions with small total influences are close to juntas — is false for functions on $[0, 1]^N$: see Hatami [Hat09, Hat12] for the precise structure in that case.

- On the one hand, if $\|\vec{\partial} f\|_{L_1 \ell_1^N} \leq 1$, then the evolution $t \mapsto P_t f$ is not too fast in the norm $\|\cdot\|_1$, so that for small t one has $P_t f \approx f$.
- On the other hand, a hypercontractivity inequality for the semigroup $(P_t)_{t \geq 0}$ implies that if $\|\partial_n f\|_1 \ll 1$ for some n , then $\|\partial_n P_t f\|_2$ rapidly decays towards zero. This will occur so fast that even for small t , the derivatives $\partial_n P_t f$ are extremely small in $\|\cdot\|_2$ for all but a few choices of n , so that $P_t f$ is close to a function that depends on only those few coordinates.

Thus, one may first approximate f by $P_t f$ for some small $t > 0$, and then prove that $P_t f$ is close to a function of only a few coordinates.

The first of these two steps results from the following.

LEMMA 2.2: *If $f \in C^1(\mathbb{T}^N)$ then*

$$\|f - P_t f\|_1 \leq \sqrt{t} \|\vec{\partial} f\|_{L_1 \ell_1^N} \quad \forall t \geq 0.$$

Proof. This is easily deduced from the representation (3), which gives

$$\|f - P_t f\|_1 \leq \int_{\mathbb{T}^N} \int_{\mathbb{R}^N} |f(x) - f(x + \bar{y})| \gamma_t^{\otimes N}(dy) dx.$$

By the Fundamental Theorem of Calculus and a change of variables, this last bound equals

$$\begin{aligned} & \int_{\mathbb{T}^N} \int_{\mathbb{R}^N} \left| \int_0^1 \frac{d}{ds} f(x + s\bar{y}) ds \right| \gamma_t^{\otimes N}(dy) dx \\ & \leq \int_{\mathbb{T}^N} \int_0^1 \int_{\mathbb{R}^N} \left| \sum_{n=1}^N y_n \partial_n f(x + s\bar{y}) \right| \gamma_t^{\otimes N}(dy) ds dx \\ & = \int_{\mathbb{T}^N} \int_{\mathbb{R}^N} \left| \sum_{n=1}^N y_n \partial_n f(x') \right| \gamma_t^{\otimes N}(dy) dx'. \end{aligned}$$

By the monotonicity of Lebesgue norms, this is at most

$$\int_{\mathbb{T}^N} \left(\int_{\mathbb{R}^N} \left(\sum_{n=1}^N y_n \partial_n f(x') \right)^2 \gamma_t^{\otimes N}(dy) \right)^{1/2} dx',$$

and now, since the coordinates y_n of y are independent under $\gamma_t^{\otimes N}$, this bound equals

$$\left(\int_{\mathbb{R}^N} y_1^2 \gamma_t(dy_1) \right)^{1/2} \int_{\mathbb{T}^N} \left(\sum_{n=1}^N |\partial_n f(x')|^2 \right)^{1/2} dx' = \sqrt{t} \int_{\mathbb{T}^N} \|\bar{\partial} f(x')\|_{\ell_2^N} dx'.$$

Finally, this is at most $\sqrt{t} \|f\|_{L_1 \ell_2^N}$, since $\|\cdot\|_{\ell_2^N} \leq \|\cdot\|_{\ell_1^N}$. ■

For the second part of the proof, the key ingredients are a reverse Poincaré inequality and a hypercontractive estimate for the heat semigroup on a torus.

PROPOSITION 2.3 (Reverse Poincaré inequality): *If $f \in \mathcal{C}^1(\mathbb{T}^N)$ then*

$$\|\bar{\partial} P_t f\|_{L_2 \ell_2^N} \leq \|f\|_2 / \sqrt{t} \quad \forall t > 0.$$

Proof. For any $f \in \mathcal{C}^1(\mathbb{T}^N)$, the self-adjointness of P_t and integration by parts give

$$\begin{aligned} \|\bar{\partial} P_t f\|_{L_2 \ell_2^N}^2 &= \sum_{n=1}^N \int_{\mathbb{T}^N} (\partial_n P_t f(x))^2 dx = - \sum_{n=1}^N \int_{\mathbb{T}^N} P_t f(x) \cdot \partial_n^2 P_t f(x) dx \\ &= - \sum_{n=1}^N \int_{\mathbb{T}^N} e^{t\Delta} f(x) \cdot \partial_n^2 e^{t\Delta} f(x) dx = \int_{\mathbb{T}^n} f(x) \cdot (-\Delta) e^{2t\Delta} f(x) dx. \end{aligned}$$

Since $-\Delta$ has spectrum contained in $[0, \infty)$, the Spectral Theorem gives

$$\|(-\Delta)e^{2t\Delta}f\|_2 \leq \left(\max_{x \in [0, \infty)} xe^{-2tx}\right)\|f\|_2 \leq \|f\|_2/t.$$

Substituting above, the Cauchy–Bunyakovski–Schwartz inequality gives

$$\|\vec{\partial}P_t f\|_{L_2 \ell_2^N}^2 \leq \|f\|_2 \|(-\Delta)e^{2t\Delta}f\|_2 \leq \|f\|_2^2/t. \quad \blacksquare$$

PROPOSITION 2.4 (Hypercontractivity on tori): *Let*

$$t > 0 \quad \text{and} \quad p := 1 + e^{-2t} \in (1, 2).$$

If $f \in L_p(\mathbb{T}^N)$, then

$$\|P_t f\|_2 \leq \|f\|_p.$$

Proof. In case $N = 1$, this is a special case of Weissler’s hypercontractive estimates in [Wei80, Theorem 2]. Given this, the result for general N follows because the operator P_t is a tensor product of one-dimensional operators, and all the norms $\|\cdot\|_p$ also tensorize. \blacksquare

Weissler’s proof of the case $N = 1$ closely follows Gross’ famous work [Gro75] on the Ornstein–Uhlenbeck semigroup and Nelson’s hypercontractive estimates. The main task is to prove a logarithmic Sobolev inequality for the Laplacian on the circle; this can then be integrated over a time interval to give hypercontractivity. Weissler’s proof of the logarithmic Sobolev inequality uses Fourier analysis, but a more direct argument from standard properties of the heat semigroup is also possible: cf. [Led01, Theorem 5.1].

Lastly, the proof will also need the following simple Poincaré inequality.

LEMMA 2.5: *For any $S \subseteq [N]$ and $f \in \mathcal{C}^1(\mathbb{T}^N)$ one has*

$$\|f - E_S f\|_2^2 \leq \sum_{n \in [N] \setminus S} \|\partial_n f\|_2^2.$$

Proof. If $S = [N] \setminus \{n\}$, then the desired inequality is

$$\|f - E_S f\|_2^2 \leq \|\partial_n f\|_2^2.$$

This follows by applying the Poincaré inequality on \mathbb{T} to each of the one-dimensional slices

$$f(x_1, \dots, x_{n-1}, \cdot, x_{n+1}, \dots, x_N), \quad (x_1, \dots, x_{n-1}, x_{n+1}, \dots, x_N) \in \mathbb{T}^{N-1},$$

and then integrating over \mathbb{T}^{N-1} .

For the general case, observe that for any n and S , the operators ∂_n and E_S commute, and so one has

$$\|\partial_n E_S f\|_2 = \|E_S(\partial_n f)\|_2 \leq \|\partial_n f\|_2.$$

Now, if $S=[N]\setminus\{n_1, \dots, n_M\}$ and one defines $f_0:=f$ and then $f_k := E_{[N]\setminus\{n_k\}} f_{k-1}$ for $k = 1, \dots, M$, then this sequence is a reverse martingale, so we have

$$\|f - E_S f\|_2^2 = \sum_{k=1}^M \|f_{k-1} - f_k\|_2^2 \leq \sum_{k=1}^M \|\partial_{n_k} f_{k-1}\|_2^2 \leq \sum_{n \in [N]\setminus S} \|\partial_n f\|_2^2. \quad \blacksquare$$

LEMMA 2.6: *Suppose that $f \in C^1(\mathbb{T}^N)$ has $\|\vec{\partial} f\|_{L^1 \ell^N_1} \leq 1$ and $\|f\|_2 \leq 1$. Fix $t > 0$ and $\eta > 0$, and let*

$$S := \{n \in [N] \mid \|\partial_n f\|_1 \geq \eta\}.$$

Then

$$\|P_{2t} f - E_S P_{2t} f\|_2 < t^{-\frac{e^{-2t}}{1+e^{-2t}}} \cdot \eta^{\frac{1-e^{-2t}}{2(1+e^{-2t})}}.$$

Proof. Proposition 2.4 and the log-convexity of the Lebesgue norms give

$$\begin{aligned} \|P_{2t}(\partial_n f)\|_2 &= \|P_t(\partial_n P_t f)\|_2 \\ &\leq \|\partial_n P_t f\|_{1+e^{-2t}} \\ &\leq \|\partial_n P_t f\|_2^{\frac{2e^{-2t}}{1+e^{-2t}}} \|\partial_n P_t f\|_1^{\frac{1-e^{-2t}}{1+e^{-2t}}}. \end{aligned}$$

Substituting this into Lemma 2.5 gives

$$\begin{aligned} \|P_{2t} f - E_S P_{2t} f\|_2^2 &\leq \sum_{n \in [N]\setminus S} \|P_{2t}(\partial_n f)\|_2^2 \\ &\leq \sum_{n \in [N]\setminus S} \|\partial_n P_t f\|_2^{\frac{4e^{-2t}}{1+e^{-2t}}} \|\partial_n P_t f\|_1^{\frac{2(1-e^{-2t})}{1+e^{-2t}}} \\ &\leq \left(\sum_{n \in [N]\setminus S} \|\partial_n P_t f\|_2^2 \right)^{\frac{2e^{-2t}}{1+e^{-2t}}} \left(\sum_{n \in [N]\setminus S} \|\partial_n P_t f\|_1^2 \right)^{\frac{1-e^{-2t}}{1+e^{-2t}}}, \end{aligned}$$

where the last bound follows from Hölder’s Inequality with exponents

$$\left(\frac{1 + e^{-2t}}{2e^{-2t}}, \frac{1 + e^{-2t}}{1 - e^{-2t}} \right).$$

Applying the contractivity of P_t and then the definition of S to the second factor, we may now bound this by

$$\begin{aligned} & \left(\sum_{n \in [N] \setminus S} \|\partial_n P_t f\|_2^2 \right)^{\frac{2e^{-2t}}{1+e^{-2t}}} \left(\sum_{n \in [N] \setminus S} \|\partial_n f\|_1^2 \right)^{\frac{1-e^{-2t}}{1+e^{-2t}}} \\ & \leq \left(\sum_{n \in [N] \setminus S} \|\partial_n P_t f\|_2^2 \right)^{\frac{2e^{-2t}}{1+e^{-2t}}} \left(\eta \sum_{n \in [N] \setminus S} \|\partial_n f\|_1 \right)^{\frac{1-e^{-2t}}{1+e^{-2t}}} \\ & \leq \eta^{\frac{1-e^{-2t}}{1+e^{-2t}}} (\|\vec{\partial} P_t f\|_{L_2 \ell_2^N}^2)^{\frac{2e^{-2t}}{1+e^{-2t}}} (\|\vec{\partial} f\|_{L_1 \ell_1^N})^{\frac{1-e^{-2t}}{1+e^{-2t}}}. \end{aligned}$$

Finally, our assumptions on f and Proposition 2.3 bound this by

$$\eta^{\frac{1-e^{-2t}}{1+e^{-2t}}} (t^{-1} \|f\|_2^2)^{\frac{2e^{-2t}}{1+e^{-2t}}} (\|\vec{\partial} f\|_{L_1 \ell_1^N})^{\frac{1-e^{-2t}}{1+e^{-2t}}} \leq t^{-\frac{2e^{-2t}}{1+e^{-2t}}} \cdot \eta^{\frac{1-e^{-2t}}{1+e^{-2t}}}.$$

Taking square roots completes the proof. ■

Proof of Theorem 2.1. Consider $f \in C^1(\mathbb{T}^N)$. Replacing f with $f - \int f$ if necessary, we may assume that $\int f = 0$. Let $t, \eta > 0$, and let S be as in Lemma 2.6. Combining Lemmas 2.2 and 2.6 gives

$$\begin{aligned} \|f - E_S f\|_1 & \leq \|f - P_{2t} f\|_1 + \|P_{2t} f - E_S P_{2t} f\|_1 + \|E_S(f - P_{2t} f)\|_1 \\ & \leq 2\|f - P_{2t} f\|_1 + \|P_{2t} f - E_S P_{2t} f\|_2 \\ & \leq 2\sqrt{2t} + t^{-\frac{e^{-2t}}{1+e^{-2t}}} \cdot \eta^{\frac{1-e^{-2t}}{2(1+e^{-2t})}}. \end{aligned}$$

For any $\varepsilon > 0$, choose $t \in (0, \varepsilon^2/32)$, so that

$$2\sqrt{2t} < \varepsilon/2;$$

then choose $\eta > 0$ so small that

$$t^{-\frac{e^{-2t}}{1+e^{-2t}}} \cdot \eta^{\frac{1-e^{-2t}}{2(1+e^{-2t})}} < \varepsilon/2.$$

For this choice of t and η , one obtains

$$\|f - E_S f\|_1 < \varepsilon.$$

On the other hand, the definition of S gives

$$\eta|S| \leq \sum_{n \in S} \|\partial_n f\|_1 = \|\vec{\partial} f\|_{L_1 \ell_1^N} \leq 1,$$

so $|S| \leq 1/\eta$, which is bounded only in terms of ε .

Lastly, suppose instead that $f \in \mathcal{C}^1([0, 1]^N)$. Let $F : \mathbb{T} \rightarrow [0, 1]$ be the map defined by

$$F(\theta + \mathbb{Z}) = \begin{cases} 2\theta & \text{if } \theta \in [0, 1/2), \\ 2 - 2\theta & \text{if } \theta \in [1/2, 1). \end{cases}$$

Then $F^{\times N} : \mathbb{T}^N \rightarrow [0, 1]^N$ is 2-Lipschitz, differentiable almost everywhere and measure preserving, so $f \circ F^{\times N}$ is differentiable almost everywhere and

$$\|\vec{\partial}(f \circ F^{\times N})\|_{L_p \ell_p^N} \leq 2\|\vec{\partial}f\|_{L_p \ell_p^N} \quad \forall p \in [1, \infty).$$

Therefore $f \circ F^{\times N}$ may be uniformly approximated by continuously differentiable functions satisfying the same inequalities, for example by convolving with a mollifier. Since also $(E_S f) \circ F^{\times N} = E_S(f \circ F^{\times N})$, the proof is completed by applying the first part of the result to $f \circ F^{\times N}$. ■

Theorem 2.1 begs the following question. I suspect the answer is negative, but have not been able to find a counterexample.

Question 2.7: In the statement of Theorem 2.1, is it enough to assume that $\|\vec{\partial}f\|_{L_1 \ell_1^N} \leq 1$, without the bound on $\|f - \int f\|_2$? (Of course one should expect a worse dependence of q on ε .)

Remarks: 1. The argument above can also be used, essentially without change, to prove an analog of Theorem 2.1 for the standard Gaussian measure γ on \mathbb{R}^N : a function $f \in \mathcal{C}^1(\mathbb{R}^N)$ for which $\int \|\vec{\partial}f\|_{\ell_1^N} d\gamma \leq 1$ and $\|f\|_{L_2(\gamma)} \leq 1$ may be approximated in $L_1(\gamma)$ by functions of boundedly many coordinates.

Indeed, it seems likely that these methods, and Theorem 2.1, generalize to a large class of Markov diffusion semigroups subject to suitable ‘curvature’ conditions, as studied in [Led00, BGL14]. More general reverse Poincaré inequalities, for instance, can be found in [BGL14, Section 4.7], and hypercontractivity estimates in [BGL14, Section 5.2].

2. Another connection worth remarking is with the recent works [KMS12] and [CEL12]. These establish versions of Talagrand’s famous variance-bound for functions on $\{\pm 1\}^N$ (see [Tal94]) in various new settings, including some product and non-product measures on \mathbb{R}^N , using a similar strategy to that above. It should also be possible to deduce Theorem 2.1 directly from one of those Talagrand-type inequalities, such as [CEL12, Theorem 1].

3. Proof of the general case

LEMMA 3.1: *Let (X, d) be a compact metric space and let $f : X \rightarrow \mathbb{R}$ be continuous with a non-decreasing modulus of continuity ω . Let $\varepsilon > 0$, and let $K_1 := \sup_{r \geq \varepsilon} r^{-1}\omega(r) \in [0, \infty]$ and $K = \max\{K_1, 1\}$. Then there is a K -Lipschitz function $h : X \rightarrow \mathbb{R}$ such that $\|f - h\|_\infty \leq K\varepsilon$.*

Proof. This is vacuous if $K = \infty$, so assume K is finite. For each $n \in \mathbb{Z}$ let

$$X_n := \{x \in X \mid f(x) \leq \varepsilon n\},$$

and define $h : X \rightarrow \mathbb{R}$ by

$$h(x) := \inf_{n \in \mathbb{Z}} (\varepsilon n + Kd(x, X_n)).$$

As a pointwise infimum of K -Lipschitz functions, h is K -Lipschitz. If $f(x) \leq \varepsilon n$, then $x \in X_n$, and so $h(x) \leq \varepsilon n + Kd(x, X_n) = \varepsilon n$. Infimizing over n , this gives $h(x) \leq f(x) + \varepsilon$.

On the other hand, for any $x \in X$, $n \in \mathbb{Z}$ and $y \in X_n$, one has

$$f(x) \leq f(y) + \omega(d(x, y)) \leq f(y) + \omega(d(x, y) + \varepsilon) \leq f(y) + Kd(x, y) + K\varepsilon.$$

Using that $f(y) \leq \varepsilon n$ and then infimizing over $y \in X_n$, this gives

$$f(x) \leq \varepsilon n + Kd(x, X_n) + K\varepsilon.$$

Now infimizing over n gives $f(x) \leq h(x) + K\varepsilon$, as required. ■

Proof of Theorem 1.1. As Gromov reminds us in the proof of the Non-Dissipation Theorem in [Gro01, Section 3 $\frac{1}{2}$.62], since (X, d, μ) is connected and locally connected, there is a uniformly continuous map $F : \mathbb{T} \rightarrow X$ such that $F_*m = \mu$. Let σ be a non-decreasing modulus of continuity for F . For any $(x_n)_n, (y_n)_n \in X^{\mathbb{N}}$, one has

$$\max_{n \leq N} d(F(x_n), F(y_n)) \leq \max_{n \leq N} \sigma(|x_n - y_n|) = \sigma(|(x_n)_n - (y_n)_n|_\infty^N),$$

so $F^{\times N} : \mathbb{T}^{\mathbb{N}} \rightarrow X^{\mathbb{N}}$ has the same modulus of continuity σ for all N for the metrics $|\cdot|_\infty^N$ and d_∞^N .

Now let $f : X^{\mathbb{N}} \rightarrow \mathbb{R}$ have non-decreasing modulus of continuity ω for the metric d_∞^N . By the above, $f \circ F^{\times N} : \mathbb{T}^{\mathbb{N}} \rightarrow \mathbb{R}$ has non-decreasing modulus of continuity $\omega \circ \sigma$. Let $\varepsilon > 0$, and let $K_\varepsilon := \max\{\sup_{r \geq \varepsilon} r^{-1}\omega(\sigma(r)), 1\}$, as in Lemma 3.1 for the function f . That lemma gives a K_ε -Lipschitz function $h : \mathbb{T}^{\mathbb{N}} \rightarrow \mathbb{R}$ for which $\|f \circ F^{\times N} - h\|_\infty \leq K_\varepsilon \varepsilon$.

Having found this h , Theorem 1.2 gives a set $S \subseteq [N]$ with $|S|$ bounded in terms of ε and K_ε (hence, in terms of ε, ω and σ) such that

$$\|h - E_S h\|_1 \leq \varepsilon,$$

and combining this with the previous inequalities gives

$$\|f - E_S f\|_1 \leq 2\|f \circ F^{\times N} - h\|_1 + \|h - E_S h\|_1 \leq 2K_\varepsilon \varepsilon + \varepsilon.$$

An easy exercise shows that $K_\varepsilon \varepsilon \rightarrow 0$ as $\varepsilon \rightarrow 0$, so this completes the proof. ■

4. Lipschitz maps between Hamming cubes

Proof of Theorem 1.3. Suppose that $F = (F_1, \dots, F_M) : [0, 1]^N \rightarrow [0, 1]^M$ is L -Lipschitz between the Hamming metrics. It may be uniformly approximated by smooth L -Lipschitz functions, so assume smoothness also. The L -Lipschitz bound implies that

$$\sum_{m=1}^M |F_m(x + dx) - F_m(x)| \leq L \sum_{n=1}^N |dx_n|$$

for all $x \in [0, 1]^N$ and all perturbations $dx = (dx_1, \dots, dx_N)$.

For fixed $n \leq N$, this inequality may be applied with

$$dx = (0, \dots, 0, dx_n, 0, \dots, 0)$$

to obtain

$$\max_{m \leq M} |F_m(x + dx) - F_m(x)| \leq \sum_{m=1}^M |F_m(x + dx) - F_m(x)| \leq L|dx_n|.$$

Normalizing and letting $|dx_n| \rightarrow 0$, this implies firstly that $\|\partial_n F_m\|_{L_\infty} \leq L$ for all n and m , and secondly that

$$\sum_{m=1}^M \sum_{n=1}^N |\partial_n F_m(x)| \leq LN = LM/\alpha \implies \frac{1}{M} \sum_{m=1}^M \|\vec{\partial} F_m\|_{L_1 \ell_1^N} \leq L/\alpha.$$

Given $\varepsilon > 0$, let

$$I := \{m \leq M \mid \|\vec{\partial} F_m\|_{L_1 \ell_1^N} \leq 2L/\alpha\varepsilon\}.$$

Applying Chebyshev's Inequality, the last inequality above implies that $|I| \geq (1 - \varepsilon/2)M$.

For each $m \in I$, one has $\|\vec{\partial}F_m\|_{L_1 \ell_1^N} \leq 2L/a\varepsilon$, and of course F_m is uniformly bounded by 1. Therefore, for each $m \in I$, the second part of Theorem 2.1 gives a subset $S_m \subseteq [N]$ of size bounded in terms of L , ε and α such that

$$\|F_m - E_{S_m}F_m\|_1 \leq \varepsilon/2.$$

Let $G_m := E_{S_m}F_m$ for $m \in I$, and let G_m be any constant function in case $m \in [M] \setminus I$. Then the above inequalities combine to give

$$\begin{aligned} \int_{[0,1]^N} |F(x) - G(x)|_1^M dx &\leq \frac{1}{M} \sum_{m \in I} \|F_m - E_{S_m}F_m\|_1 + \frac{M - |I|}{M} \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \quad \blacksquare \end{aligned}$$

Equivalently, Theorem 1.3 asserts that most of the functions F_m are individually close to functions that depend on only small sets of coordinates. One cannot tighten this conclusion to apply to strictly every F_m . For example, the function

$$[0, 1]^N \longrightarrow [0, 1]^N : (x_1, \dots, x_N) \mapsto (x_1, \dots, x_{N-1}, \sin(2\pi(x_1 + \dots + x_N)))$$

is easily checked to be continuously differentiable and 2-Lipschitz for the Hamming metrics, but its last coordinate is not close to any function depending on fewer than N coordinates. By concatenating several such examples, for any divergent positive sequence $a_N = o(N)$, one may construct a sequence of 2-Lipschitz maps in which roughly a_N of the output coordinates each depend on roughly N/a_N of the input coordinates.

A similar argument to the proof of Theorem 1.3, using Friedgut’s Theorem [Fri98] on Boolean functions with small influences in place of Theorem 2.1, shows that the analog of Theorem 1.3 also holds for functions

$$\{0, 1\}^N \longrightarrow \{0, 1\}^M,$$

where domain and target are again given the Hamming metrics. This argument appeared recently in Subsection 1.1 of [BCS], where it was used to analyze a particular bi-Lipschitz bijection between $\{0, 1\}^N$ and a radius- $(N/2)$ Hamming ball in $\{0, 1\}^{N+1}$ (although those authors were also able to give a more precise bespoke analysis for that function). Actually, the argument using Friedgut’s Theorem in [BCS] requires only ‘bounded average stretch’ for the function $\{0, 1\}^N \longrightarrow \{0, 1\}^M$. In our setting of a map $F = (F_1, \dots, F_m) : [0, 1]^N \longrightarrow [0, 1]^M$,

this corresponds to assuming that

$$\sum_{m=1}^M \|\vec{\partial}F_m\|_{L_1\ell_1^N} \leq LM$$

for some fixed constant L . One can therefore obtain a closer analog of that conclusion from [BCS] by using the full strength of Theorem 2.1.

By contrast, no obvious analog of Theorem 1.3 holds between Euclidean spheres with the Euclidean metrics. To see this, consider the following two natural classes of Lipschitz map $S^{3N-1} \rightarrow S^{3N-1}$:

- maps of the form

$$(x_1, \dots, x_{3N}) \mapsto (F_1(x_1, x_2, x_3), \dots, F_N(x_{3N-2}, x_{3N-1}, x_{3N})),$$

where $F_i : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ for $i \leq N$ are 10-Lipschitz functions such that $\|F_i(u, v, w)\|_{\ell_2^3} = \|F_i(u, v, w)\|_{\ell_2^3}$ (for example, one may choose a sequence of 10-Lipschitz maps $S^2 \rightarrow S^2$ and extend them radially to \mathbb{R}^3);

- orthogonal rotations.

Both of these classes are individually somewhat simple. However, one may now form a composition of a few maps drawn alternately from these two classes, and easily produce a map with the property that no output coordinate is close to any function that depends on only a low-dimensional projection of the domain sphere. It is also easy to see that such compositions can be chosen to be $O(1)$ -bi-Lipschitz and preserve Lebesgue measure on S^{3N-1} , by a suitable choice of the functions F_i appearing in the examples of the first kind.

5. Some questions about non-product measures

Several important developments in the theory of concentration have come from extensions of concentration inequalities from product to non-product measures. Similarly, it would be interesting to know whether any other natural measures on product spaces enjoy an analog of Theorem 1.1.

One rich source of examples is dynamics. Suppose that (X, d, μ) is a connected and locally connected metric probability space and that $T : X \rightarrow X$ is a μ -preserving homeomorphism. Then for each N one may form the topological embeddings

$$T^{[0:N-1]} := (\text{id}, T, T^2, \dots, T^{N-1}) : X \rightarrow X^N,$$

and consider the image $T^{[0;N-1]}(X)$ and pushforward measure $T_*^{[0;N-1]}\mu$. Equivalently, one may consider the original measure space (X, μ) with the pullback of the metric d_∞^N , giving the new metric

$$D_\infty^N(x, y) := \max_{0 \leq n \leq N-1} d(T^n x, T^n y).$$

This construction is important in the theory of dynamical systems: the exponential growth rate of the D_∞^N -covering numbers of X gives the Bowen–Dinaburg definition of topological entropy.

One may obtain examples similar to ℓ_∞ -product spaces with product measures in this way. For instance, if (X, μ, T) is the shift on $\mathbb{T}^{\mathbb{Z}}$ with its Haar measure and some sensible choice of compact metric, then the above metrics D_∞^N behave increasingly like the metrics $|\cdot|_\infty^N$ on finite marginals of $\mathbb{T}^{\mathbb{Z}}$. This suggests the following question.

Question 5.1: For which systems (X, μ, T) is it the case that $\forall \varepsilon > 0 \exists p, K \geq 1$ such that $\forall N \geq 1$, if $f : X \rightarrow \mathbb{R}$ is 1-Lipschitz for the metric D_∞^N , then it is ε -close in $\|\cdot\|_{L_2(\mu)}$ to a function of the form

$$F(T^{n_1} x, T^{n_2} x, \dots, T^{n_p} x)$$

for some $n_1, \dots, n_p \in \{0, 1, \dots, N - 1\}$, where $F : X^p \rightarrow \mathbb{R}$ is K -Lipschitz for the metric d_∞^p ?

Another instructive example is the map $\theta \mapsto 2\theta$ on $(\mathbb{T}, |\cdot|, m_{\mathbb{T}})$. An easy exercise shows that for this map, D_∞^N is equivalent up to constants to the metric

$$\rho^N(\theta, \theta') := \min\{2^N|\theta - \theta'|, 1\}.$$

Therefore any function $\mathbb{T}^N \rightarrow [0, 1]$ which is 2^N -Lipschitz for the usual metric $|\cdot|$ becomes Lipschitz with bounded Lipschitz constant for the metric D_∞^N . This certainly includes many maps that cannot be approximated in measure by ‘low-dimensional’ functions F as above, so the $(\times 2)$ -map is not an example. Similar reasoning seems to apply to Anosov diffeomorphisms, so these in general do not give examples.

Question 5.2: Are there any positive examples for the previous question in which (X, T) has finite topological entropy?

References

- [Bar04] F. Barthe, *Infinite dimensional isoperimetric inequalities in product spaces with the supremum distance*, Journal of Theoretical Probability **17** (2004), 293–308.
- [BCS] I. Benjamini, G. Cohen and I. Shinkar, *Bi-Lipschitz bijection between the Boolean cube and the Hamming ball*, preprint, available online at [arXiv.org](https://arxiv.org/abs/1310.2017): 1310.2017.
- [BGL14] D. Bakry, I. Gentil and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, Grundlehren der Mathematischen Wissenschaften, Vol. 348, Springer, Cham, 2014.
- [BKK⁺92] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson and N. Linial, *The influence of variables in product spaces*, Israel Journal of Mathematics **77** (1992), 55–64.
- [CEL12] D. Cordero-Erausquin and M. Ledoux, *Hypercontractive measures, Talagrand’s inequality, and influences*, in *Geometric Aspects of Functional Analysis*, Lecture Notes in Mathematics, Vol. 2050, Springer, Heidelberg, 2012, pp. 169–189.
- [DFR08] I. Dinur, E. Friedgut and O. Regev, *Independent sets in graph powers are almost contained in juntas*, Geometric and Functional Analysis **18** (2008), 77–97.
- [FK96] E. Friedgut and G. Kalai, *Every monotone graph property has a sharp threshold*, Proceedings of the American Mathematical Society **124** (1996), 2993–3002.
- [Fri98] E. Friedgut, *Boolean functions with low average sensitivity depend on few coordinates*, Combinatorica **18** (1998), 27–35.
- [Fri04] E. Friedgut, *Influences in product spaces: KKL and BKKKL revisited*, Combinatorics, Probability and Computing **13** (2004), 17–29.
- [Gro75] L. Gross, *Logarithmic Sobolev inequalities*, American Journal of Mathematics **97** (1975), 1061–1083.
- [Gro01] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*, second edition, Modern Birkhäuser Classics, Birkhäuser Boston, Boston, MA, 2001.
- [Hat09] H. Hatami, *Decision trees and influences of variables over product probability spaces*, Combinatorics, Probability and Computing **18** (2009), 357–369.
- [Hat12] H. Hatami, *A structure theorem for Boolean functions with small total influences*, Annals of Mathematics **176**(2012), 509–533.
- [KKL] J. Kahn, G. Kalai and N. Linial, *The influence of variables on Boolean functions*, in *Proceedings of the 29th Annual Symposium on Foundations of Computer Science* IEEE Computer Society Press, Los Alamitos, CA, 1988, pp. 68–80.
- [KMS12] N. Keller, E. Mossel and A. Sen, *Geometric influences*, Annals of Probability **40** (2012), 1135–1166.
- [Led00] M. Ledoux, *The geometry of Markov diffusion generators*, Annales de la Faculté des Sciences de Toulouse. Mathématique. Série 6 **9** (2000), 305–366. Probability theory.
- [Led01] M. Ledoux, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs, Vol. 89, American Mathematical Society, Providence, RI, 2001.
- [Tal94] M. Talagrand, *On Russo’s approximate zero-one law*, Annals of Probability **22** (1994), 1576–1587.
- [Wei80] F. B. Weissler, *Logarithmic Sobolev inequalities and hypercontractive estimates on the circle*, Journal of Functional Analysis **37** (1980), 218–234.