Electronic Journal of Probability

Electron. J. Probab. **23** (2018), no. 16, 1–17.
ISSN: 1083-6489 https://doi.org/10.1214/18-EJP145

Noise stability and correlation with half spaces

Elchanan Mossel* Joe Neeman†

Abstract

Benjamini, Kalai and Schramm showed that a monotone function $f : \{-1,1\}^n \to \{-1,1\}$ is noise stable if and only if it is correlated with a half-space (a set of the form $\{x : \langle x, a \rangle \leq b\}$).

We study noise stability in terms of correlation with half-spaces for general (not necessarily monotone) functions. We show that a function $f : \{-1,1\}^n \to \{-1,1\}$ is noise stable if and only if it becomes correlated with a half-space when we modify f by randomly restricting a constant fraction of its coordinates.

Looking at random restrictions is necessary: we construct noise stable functions whose correlation with any half-space is $o(1)$. The examples further satisfy that different restrictions are correlated with different half-spaces: for any fixed half-space, the probability that a random restriction is correlated with it goes to zero.

We also provide quantitative versions of the above statements, and versions that apply for the Gaussian measure on $\mathbb{R}^n$ instead of the discrete cube. Our work is motivated by questions in learning theory and a recent question of Khot and Moshkovitz.

Keywords: boolean function; noise stability; functional inequality.
AMS MSC 2010: 60.
Submitted to EJP on April 25, 2017, final version accepted on January 25, 2018.
Supersedes arXiv:1603.01799.

1 Introduction

In a seminal paper, Benjamini, Kalai and Schramm [2] related noise stability to correlation with half-spaces by showing that a monotone boolean function is noise stable if and only if it is correlated with a half-space. Our interest in this paper is relating noise stability with correlation with half-spaces for general boolean functions. Our results are motivated by recent work of Khot and Moshkovitz whose goal is to construct a Lasserre integrality gap for the Unique Games problems as well as by natural problems in learning theory.

In the following subsections we introduce the setup and results in the boolean and Gaussian cases and discuss the motivation for our work.

*MIT; E-mail: elmos@mit.edu
†UT Austin; E-mail: joeneeman@gmail.com

1.1 The boolean setting

Let μ_n denote the uniform measure on $\{-1,1\}^n$. For $t \geq 0$, let P_t denote the Bonami-Beckner semigroup, defined by

$$(P_t f)(x) = (P_{t,1} f)(x) = \mathbb{E}f + e^{-t}(f(x) - \mathbb{E}f)$$

in the case $n = 1$ and $P_{t,n} = P_{t,1}^{\otimes n}$ otherwise. The *boolean noise stability* of a set $A \subset \{-1,1\}^n$ is

$$\mathrm{NS}_t(A) = \mathbb{E}[1_A P_t 1_A],$$

where the expectation is taken with respect to μ_n. There is a natural probabilistic interpretation of the noise stability: let X be a uniformly random element of $\{-1,1\}^n$ and let Y be a "noisy" copy of X defined by (independently for every i) setting $Y_i = X_i$ with probability e^{-t} and resampling Y_i uniformly from $\{-1,1\}$ otherwise. Then $\mathrm{NS}_t(A) = \Pr(X \in A \text{ and } Y \in A)$. In this setting, the operator P_t can be understood as a conditional expectation: $(P_t f)(x) = \mathbb{E}[f(Y) \mid X = x]$.

Since $P_t = P_{t/2} P_{t/2}$ and P_t is self-adjoint, we may also write $\mathrm{NS}_t(A) = \mathbb{E}[(P_{t/2} 1_A)^2]$. Then

$$\mathrm{NS}_t(A) - \mu_n(A)^2 = \mathrm{NS}_t(A) - (\mathbb{E}P_{t/2} 1_A)^2 = \mathrm{Var}(P_{t/2} 1_A) \geq 0;$$

the quantity $\mathrm{Var}(P_t 1_A)$ turns out to be a useful re-parametrization of the usual boolean noise sensitivity.

We say that a sequence $A_i \subset \{-1,1\}^{n_i}$ of sets is *noise sensitive* if for every $t > 0$, $\mathrm{Var}(P_t 1_{A_i}) \to 0$ as $i \to \infty$. Otherwise, we say that the sequence A_i is *noise stable*.

A *half-space* is a set of the form $\{x \in \{-1,1\}^n : \langle x, a \rangle \leq b\}$; write $\mathcal{H}_n$ for the set of all half-spaces in $\{-1,1\}^n$. Define

$$M(A) = \sup_{B \in \mathcal{H}_n} \mathrm{Cov}(1_A, 1_B).$$

Clearly $0 \leq M(A) \leq \frac{1}{4}$ for all A.

The set $A \subset \{-1,1\}^n$ is *monotone* if whenever $x \in A$ and $y \geq x$ coordinatewise then $y \in A$. Benjamini, Kalai, and Schramm [2] proved that a sequence A_i of monotone sets is noise sensitive if and only if $M(A_i) \to 0$. In this article, we explore removing the condition of monotonicity. First, we show that one direction of Benjamini et al.'s equivalence fails when the A_i are allowed to be non-monotone. In particular, we construct a sequence of sets $B_i \subset \{-1,1\}^{n_i}$ such that $M(B_i) \to 0$ but $\mathrm{NS}_t(B_i) \not\to 0$; in other words, noise-stable sets are not necessarily correlated with any half-spaces.

Although noise-stable sets may not be correlated with half-spaces, there is a characterization of noise stability in terms of half-spaces; this characterization requires the notion of a *restriction*. For $z \in \{-1,0,1\}$ and $y \in \{-1,1\}$, define $\Pi_z y \in \{-1,1\}$ by

$$\Pi_z y = \begin{cases} y & \text{if } z = 0 \\ z & \text{otherwise.} \end{cases}$$

For $z \in \{-1,0,1\}^n$ and $y \in \{-1,1\}^n$, define $\Pi_z y \in \{-1,1\}^n$ coordinatewise: $(\Pi_z y)_i = \Pi_{z_i} y_i$. We use the notation Π because this operation may be interpreted as the projection of y onto the subcube

$$\Omega_z = \{x \in \{-1,1\}^n : x_i = z_i \text{ whenever } z_i \neq 0\}.$$

For a set $B \subset \{-1,1\}^n$, define the *restriction* of B by z to be

$$B_z = \{x \in \{-1,1\}^n : \Pi_z x \in B\};$$

for a function $f : \{-1,1\}^n \to \mathbb{R}$, define the restriction of f by z to be

$$f_z(y) = f(\Pi_z y).$$

Observe that our two notions of restriction are compatible in the sense that $(1_B)_z = 1_{B_z}$. The term "restriction" comes from the fact that f_z is essentially the same as $f|_{\Omega_z}$, in the sense that all properties that we care about (e.g. correlation with half-spaces) are the same for f_z and $f|_{\Omega_z}$.

Write μ_t for the measure on $\{-1,0,1\}^n$ under which each coordinate is independent, equal to zero with probability e^{-t}, and chosen uniformly from $\{-1,1\}$ otherwise. We remark that the notion of a *random* restriction is closely related to the notion of noise stability: if $Z \sim \mu_t$ and X is uniformly random in $\{-1,1\}^n$ (and independent of Z) then $\Pi_Z X$ is a "noisy" copy of X in the sense above. In particular, $\mathrm{NS}_t(A) = \Pr(X \in A$ and $\Pi_Z X \in A)$. Similarly, note that if $Z_s \sim \mu_s$ then $P_s f = \mathbb{E} f_{Z_s}$

Our main theorem, in its qualitative form (its analogous quantitative versions are Theorem 3.1 and Theorem 3.10), says that a set is noise stable if and only if we can make it correlated with a half-space by randomly restricting a constant fraction of its coordinates.

Theorem 1.1. *The sequence $B^{(i)} \subset \{-1,1\}^{n_i}$ is noise stable if and only if there are some $t, \epsilon > 0$ such that for all sufficiently large i, $M(B_Z^{(i)}) \geq \epsilon$ with probability at least ϵ, where $Z \sim \mu_t$.*

The proof of Theorem 1.1 is not very complicated. In one direction, it is well-known (Theorem 3.11) that every half-space is noise-stable, and it then follows that if a set is correlated with a half-space then it is noise-stable (Proposition 3.12). Finally, if a random restriction of a set is noise-stable then the original set must also be noise-stable. This proves that if a random restriction is correlated with a half-space then the set is noise-stable.

For the other direction, the main idea is to involve the "first-level Fourier weight" of a set, defined by

$$w_1(B) = \sum_{i=1}^{n} \mathbb{E}[X_i 1_B(X)]^2.$$

We then proceed in two steps: first (in Proposition 3.2), we prove that if $w_1(B)$ is large then B is correlated with a half-space. For this step, note that if $w_1(B)$ is large then B is correlated with a linear function, and we can use that linear function to find a correlated half-space. For the second step (in Proposition 3.3), we prove that for a noise-stable function, random restrictions have large first-level Fourier weight.

Since the notion of taking restrictions may seem artificial, it is natural to ask whether taking restrictions in Theorem 1.1 is really necessary. That is, could it be that $B^{(i)}$ noise stable already implies that $M(B^{(i)}) \not\to 0$? In fact, this is not the case. As an example, take $n_m = n^2$ and consider the sets $B^{(m)} \subset \{-1,1\}^{n_m}$ defined by

$$B^{(m)} = \left\{ x : \sum_{i=1}^{m} \left(\frac{1}{\sqrt{m}} \sum_{j=(i-1)m+1}^{im} x_j \right)^2 \leq m \right\}.$$

Proposition 1.2. *The sets $B^{(m)}$ are noise stable, but $M(B^{(m)}) \leq C m^{-1/200}$ for a universal constant C.*

1.2 The Gaussian setting

The preceding results also make sense in a Gaussian setting: Let γ_n denote the standard Gaussian measure on $\mathbb{R}^n$ and write P_t for the Ornstein-Uhlenbeck semigroup, defined by

$$P_t f(x) = \mathbb{E} f(e^{-t}x + \sqrt{1-e^{-2t}}X), \qquad X \sim \gamma_n.$$

(Here and elsewhere we will reuse symbols that we also used in the boolean setting; however, the meaning should always be clear from the context.) The *Gaussian noise stability* of a set $A \subset \mathbb{R}^n$ is

$$\mathrm{NS}_t(A) = \mathbb{E}[1_A P_t 1_A].$$

For a probabilistic definition of noise stability, suppose that $X \sim \gamma_n$ and $Y \sim \gamma_n$ are jointly Gaussian with correlation e^{-t}. Then $\mathrm{NS}_t(A) = \Pr(X \in A \text{ and } Y \in A)$. As in the boolean case, we have $\mathrm{NS}_t(A) - \gamma_n(A)^2 = \mathrm{Var}(P_{t/2} 1_A)$; we say that a sequence A_i of sets is *noise sensitive* if $\mathrm{Var}(P_t 1_{A_i}) \to 0$ for all $t > 0$, and we say that A_i is *noise stable* otherwise. A *half-space* is a set of the form $\{x \in \mathbb{R}^n : \langle x, a\rangle \le b\}$; write $\mathcal{H}_n$ for the set of all half-spaces in $\mathbb{R}^n$ and define

$$M(A) = \sup_{B \in \mathcal{H}_n} \mathrm{Cov}(1_A, 1_B).$$

In the setting above, we prove that a sequence of sets is noise stable if and only if by scaling and randomly shifting it, we make them correlated with half-spaces. Specifically, given $B \subset \mathbb{R}^n$, $t \ge 0$, and $y \in \mathbb{R}^n$, define

$$B_{t,y} = \{x \in \mathbb{R}^n : \sqrt{1-e^{-2t}}c + e^{-t}y \in B\}.$$

Theorem 1.3. *The sequence $B^{(i)} \subset \mathbb{R}^n$ is noise stable if and only if there are some $t, \epsilon > 0$ such that for all sufficiently large i, $M(B^{(i)}_{t,Y}) \ge \epsilon$ with probability at least ϵ, where $Y \sim \gamma_n$.*

The proof of Theorem 1.3 follows the same general outline as the proof of Theorem 1.1. In fact, Theorem 1.3 is a little bit nicer to prove, because the Gaussian measure is particularly easy to work with; therefore, we will prove Theorem 1.3 first.

As in the boolean case, one can find examples showing that Theorem 1.3 would be false if we didn't introduce the scaling and random shifting. In this case, the example is very easy: let $B^{(n)} \subset \mathbb{R}^n$ be the Euclidean ball of radius $\sqrt{n}$.

Proposition 1.4. *The sets $B^{(n)}$ are noise stable, but $M(B^{(n)}) \le n^{-1/2}$.*

One can learn a little more from this example. First, note that any restrictions of $B^{(n)}$ are also Euclidean balls. In the Gaussian setting, therefore, unlike in the boolean one, noise stability does not imply that random restrictions are correlated with half-spaces. Another observation (since $B^{(n)}$ is rotationally invariant) is that noise stable sets do not necessarily "encode" directions. We make this more precise in Proposition 2.4, which says that even though random shifts and scalings of $B^{(n)}$ are correlated with half-spaces, the directions in which those half-spaces point are unpredictable.

1.3 Motivation

Our work is motivated by extending the results of [2] to non-monotone functions, as well by the following motivations:

- In a recent work Khot and Moshkovitz [4], proposed a Lasserre integrality gap for the Unique Games problem. The proposed construction is based on the assumption that in a certain family of functions, the most stable functions are half-spaces. More specifically [4] considers $f : \mathbb{R}^n \to \{-1, +1\}$ which satisfy

$$f(-x) = f(x + e_i) = -f(x),$$

for all x and for the standard basis vectors e_i; they asked whether the most stable functions in this family are of the form $\operatorname{sgn}(\sum_i \sigma_i x_i)$ where $\sigma_i \in \{-1, 1\}^n$, and also whether every function that is almost as noise stable as possible must be correlated with a function of this form.

In this context, it is natural to ask whether every noise stable function is correlated with a half-space. This is the question we address in this paper. However, since our functions are not required to satisfy $f(x + e_i) = -f(x)$, our results and examples do not have direct implications for the proposed Lasserre integrality gap instances.

- It is well known that the class of functions having a constant fraction (resp. most) of their Fourier mass on "low" coefficients can be weakly (resp. strongly) learned under the uniform distribution [7, 5]. In particular, noise stable functions can be weakly learned. On the other hand, the most classical learning algorithms involve learning half-spaces. Thus it is natural to ask if there is more direct relation between the weak learnability of noise stable functions and the learnability of half-spaces. Our examples seem to provide a negative answer to this question.

Remark 1.5. It is natural to ask whether the theorem of [2] can be recovered from our results. For example, by combining Theorem 1.1 with [2] it follows that a monotone set is correlated with a half-space if and only if its random restrictions are correlated with half-spaces. But is this fact obvious from first principles? If it were, it would combine with Theorem 1.1 to give a different proof of [2].

2 The Gaussian case

For this section, let $X \sim \gamma_n$. Recall that the Ornstein-Uhlenbeck semi-group is defined by

$$P_t f(x) = \mathbb{E} f(e^{-t} x + \sqrt{1 - e^{-2t}} X).$$

For $t \in \mathbb{R}$ and $y \in \mathbb{R}^n$, define $f_{t,y}$ by $f_{t,y}(x) = f(\sqrt{1 - e^{-2t}} x + e^{-t} y)$.

Theorem 2.1. *There is a universal constant $c > 0$ such that for any measurable $f : \mathbb{R}^n \to [0, 1]$ and any $t > 0$,*

$$\mathbb{E}\left[M^2(f_{t,Y}) \log \frac{1}{M(f_{t,Y})}\right] \geq c(e^{2t} - 1) \operatorname{Var}(P_t f),$$

where the expectation is with respect to $Y \sim \gamma_n$.

2.1 An example

As discussed in the introduction, a simple example shows that f itself may not be correlated with a half-space: let $B_n \subset \mathbb{R}^n$ be the Euclidean ball of radius $\sqrt{n}$. First, we note that for sufficiently small t, $\operatorname{Var}(P_t 1_{B_n})$ is bounded away from zero as $n \to \infty$. (This is already well-known [3], since B_n is obtained by thresholding a quadratic function, but the computation in our special case is quite easy.)

Proposition 2.2. *For any n and any $t > 0$,*

$$\operatorname{Var}(P_t 1_{B_n}) \geq \frac{1}{4} - \frac{\arccos(e^{-2t})}{\sqrt{2}\pi} - o_n(1).$$

In particular B_n is noise stable.

Proof. For a set of B of smooth boundary, we may define the *Gaussian perimeter* of B as

$$\int_{\partial B} \frac{d\gamma_n}{d\lambda}(x) \, d\mathcal{H}_{n-1}(x),$$

where $\mathcal{H}_{n-1}$ denotes the $(n-1)$-dimensional Hausdorff measure and $\frac{d\gamma_n}{d\lambda}$ denotes the Gaussian density with respect to the Lebesgue measure. Since the Gaussian density restricted to ∂B_n takes the constant value $(2\pi e)^{-n/2}$ and the Euclidean surface area of B_n is $\sqrt{n}^{n-1} \cdot 2\pi^{n/2}/\Gamma(n/2)$, it follows that the Gaussian perimeter of B_n is

$$\frac{2n^{n/2-1/2}}{(2e)^{n/2}\Gamma(n/2)} \sim \frac{1}{\sqrt{\pi}},$$

where the approximation follows from Stirling's formula.

On the other hand, Ledoux [6] proved that if P is the Gaussian perimeter of B then

$$\mathbb{E}[1_B(1_B - P_t 1_B)] \leq \frac{\arccos(e^{-t})P}{\sqrt{2\pi}}.$$

Plugging in our asymptotics for the Gaussian perimeter of B_n, we have

$$\mathbb{E}[1_{B_n}(1_{B_n} - P_t 1_{B_n})] \leq (1 + o_n(1))\frac{\arccos(e^{-t})}{\sqrt{2\pi}}.$$

Since $P_t = P_{t/2}P_{t/2}$ and $P_{t/2}$ is self-adjoint, this may be rearranged into

$$\mathbb{E}[(P_{t/2}1_{B_n})^2] \geq \Pr(B_n) - (1 + o_n(1))\frac{\arccos(e^{-t})}{\sqrt{2\pi}}.$$

Since $\Pr(B_n) = \frac{1}{2} + o_n(1)$, this proves the claim. □

Next, we observe that B_n is not correlated with any half-space:

Proposition 2.3. $M(B_n) \leq n^{-1/2}$.

In particular, Propositions 2.2 and 2.3 together imply that Theorem 2.1 would no longer be true if $f_{t,y}$ were replaced by f.

Proof. Since B_n is rotationally invariant, it suffices to consider half-spaces of the form $A_i := \{x : x_i \leq b\}$. Since $\Pr(A_i) = \Phi(b)$,

$$\operatorname{Cov}(1_{B_n}, 1_{A_i}) = \mathbb{E}[1_{B_n}(1_{A_i} - \Phi(b))].$$

Now let $f_i = 1_{A_i} - \Phi(b)$. Then the f_i are orthogonal and satisfy $\|f_i\|_2 \leq 1$. Hence,

$$1 \geq \|1_{B_n}\|_2^2 \geq \sum_{i=1}^n \mathbb{E}[1_{B_n} f_i]^2 = n\mathbb{E}[1_{B_n} f_1]^2,$$

and so $\mathbb{E}[1_{B_n} f_1] \leq n^{-1/2}$. □

A very similar argument shows that even though shifts of A_n may be correlated with half-spaces, the half-spaces are pointed in unpredictable directions.

Proposition 2.4. *Let $g = 1_{B_n}$ and let $g_{t,y}(x) = g(\sqrt{1-e^{-2t}}x + e^{-t}y)$. For any half-space A,*

$$\mathbb{E}_Y[\operatorname{Cov}(g_{t,Y}, 1_A)^2] \leq \frac{1}{n}.$$

In particular, Chebyshev's inequality implies that for any $u > 0$, with probability at least $1 - u^{-2}$ over $Y \sim \mathcal{N}(0, I_n)$

$$|\operatorname{Cov}(g_{t,Y}, 1_A)| \leq \frac{u}{n}.$$

Proof. Let $A_i = \{x : x_i \leq b\}$ and $f_i = 1_{A_i} - \Phi(b)$. As in the proof of the previous proposition, for any Y and t,

$$1 \geq \sum_{i=1}^n \mathbb{E}[g_{t,Y} f_i]^2 = n\mathbb{E}[g_{t,Y} f_1]^2 = n \operatorname{Cov}(g_{t,Y}, f_1)^2.$$

Taking the expectation over Y completes the proof. □

2.2 Proof of Theorem 2.1

For $f \in L_2(\gamma_n)$, define $w_1(f) = \sum_i \mathbb{E}[X_i f(X)]^2$. Using the integration by parts formula $\mathbb{E}[X_i f(X)] = \mathbb{E}[\frac{\partial f}{\partial x_i}(X)]$, we may also write $w_1(f) = |\mathbb{E}\nabla f|^2$. The proof of Theorem 2.1 goes in two steps: first, we show that if $w_1(f)$ is non-negligible then there exists a half-space correlated with f. Then, we show that for a random $Y \sim \gamma_n$, $w_1(f_{t,Y})$ is non-negligible in expectation.

For the first step, we will make use of the following simple identity:

Lemma 2.5. *Let ν be a probability measure on $\mathbb{R}$ that has a finite mean and is symmetric, in the sense that if $Y \sim \nu$ then $-Y \sim \nu$. Then for any $x \in \mathbb{R}$,*

$$x = \int_{-\infty}^{\infty} 1_{\{x \geq t\}} - \nu[t, \infty)\, dt.$$

Proof. Note that the convergence of the integral follows from the finiteness of the mean. Moreover, if we define $\psi(x) = \int_{-\infty}^{\infty} 1_{\{x \geq t\}} - \nu[t, \infty)\, dt$ then we may write

$$\begin{aligned}\psi(x) &= \int_{-\infty}^{x} 1 - \nu[t, \infty)\, dt - \int_{x}^{\infty} \nu[t, \infty)\, dt \\ &= \int_{-\infty}^{x} \nu(-\infty, t)\, dt - \int_{x}^{\infty} \nu[t, \infty)\, dt.\end{aligned}$$

It follows that ψ is continuous and Lebesgue-a.e. differentiable, and $\psi'(x) = \nu(-\infty, x) + \nu[x, \infty) = 1$ a.e. Then we must have $\psi(x) = x + C$, but the symmetry of ν implies that $\psi(0) = 0$; hence $C = 0$ and so $\psi(x) = x$. □

Proposition 2.6. *There is a universal constant $c > 0$ such that for any $f : \mathbb{R}^n \to [0, 1]$,*

$$M^2(f) \log \frac{1}{M(f)} \geq c w_1(f).$$

Proof. Let $a_i = \mathbb{E}[X_i f(X)]$, and note that

$$\sqrt{w_1(f)} = \left(\sum_i a_i^2 \right)^{1/2} = \max_{|v|=1} \langle a, v \rangle = \max_{|v|=1} \mathbb{E}\left[\langle v, X \rangle f(X) \right],$$

where $|\cdot|$ denotes the Euclidean norm. Choose v to maximize the right hand side above. Since the distribution of $\langle v, X \rangle$ is symmetric, Lemma 2.5 implies that

$$\sqrt{w_1(f)} = \mathbb{E}[\langle v, X \rangle f(X)] = \int_{-\infty}^{\infty} \mathrm{Cov}(f, 1_{\{\langle v, \cdot \rangle \geq t\}})\, dt. \tag{2.1}$$

Next, we will show that the tails of the above integral decay rapidly, and it will follow that there exists some $t \in \mathbb{R}$ for which $\mathrm{Cov}(f, 1_{\{\langle v, \cdot \rangle \geq t\}})$ is large.

Since f takes values in $[0, 1]$, we have $\mathrm{Var}(f) \leq \frac{1}{4}$. Then the Cauchy-Schwarz inequality implies that

$$\mathrm{Cov}(f, 1_{\{\langle v, \cdot \rangle \geq t\}}) \leq \sqrt{\mathrm{Var}(f)\, \mathrm{Var}(1_{\{\langle v, \cdot \rangle \geq t\}})} \leq \frac{1}{2} \sqrt{\Pr(\langle v, X \rangle \geq t) \Pr(\langle v, X \rangle < t)}.$$

Now, $\langle v, X \rangle$ has a standard Gaussian distribution and so at least one out of $\Pr(\langle v, X \rangle \geq t)$ and $\Pr(\langle v, X \rangle < t)$ is bounded by $\exp(-t^2/2)$. Hence,

$$\mathrm{Cov}(f, 1_{\{\langle v, \cdot \rangle \geq t\}}) \leq \frac{1}{2} \exp(-t^2/4).$$

For any $s \geq 1$, it follows that

$$\int_{\mathbb{R}\setminus[-s,s]} \mathrm{Cov}(f, 1_{\{\langle v,\cdot\rangle \geq t\}})\, dt \leq \int_s^\infty \exp(-t^2/4)\, dt \leq 2\exp(-s^2/4),$$

where the second inequality follows from bounding $\exp(-t^2/4)$ by $t\exp(-t^2/4)$ and then integrating by parts. Going back to (2.1), we have (for any $s \geq 1$)

$$\int_{-s}^s \mathrm{Cov}(f, 1_{\{\langle v,\cdot\rangle \geq t\}})\, dt \geq \sqrt{w_1(f)} - 2\exp(-s^2/4).$$

Choosing $s = \sqrt{2\log(16/w_1(f))}$ (which is at least 1 because $w_1(f) \leq 1$), we have $2\exp(-s^2/4) = \frac{1}{2}\sqrt{w_1(f)}$ and so

$$\int_{-s}^s \mathrm{Cov}(f, 1_{\{\langle v,\cdot\rangle \geq t\}})\, dt \geq \frac{1}{2}\sqrt{w_1(f)}.$$

In particular, there is some $t \in [-s, s]$ such that

$$\mathrm{Cov}(f, 1_{\{\langle v,\cdot\rangle \geq t\}}) \geq \frac{1}{4s}\sqrt{w_1(f)}.$$

Plugging in our value for s proves that

$$M(f) \geq \frac{1}{4\sqrt{2\log(16/w_1(f))}}\sqrt{w_1(f)},$$

which (after some rearrangement) implies the claim. □

Remark 2.7. We remark that the proof of Proposition 2.6 does not use the Gaussian setting in a particularly strong way. In particular, the proof is valid whenever X has a symmetric sub-Gaussian distribution, where "sub-Gaussian" means that

$$\Pr(\langle X, v\rangle \geq t) \leq \exp(-t^2/2)$$

for any unit vector v and any $t \geq 0$.

The second step in the proof of Theorem 2.1 is to show that if a function f is noise stable then it has some shifts $f_{t,y}$ with non-negligible $w_1(f_{t,y})$. In order to do this, recall the Gaussian Poincaré inequality (see, e.g. [1]), which states that $\mathrm{Var}(f) \leq \mathbb{E}|\nabla f|^2$ for any f with continuous derivatives.

Proposition 2.8. *For any f and any $t > 0$, if $Y \sim \mathcal{N}(0, I_n)$ then*

$$\mathbb{E}w_1(f_{t,Y}) \geq (e^{2t} - 1)\,\mathrm{Var}(P_t f).$$

Proof. Since smooth functions are dense in $L_2(\gamma_n)$, and since both $w_1(f)$ and $\mathrm{Var}(P_t f)$ are preserved under $L_2(\gamma_n)$ convergence, we may assume that f is smooth. Then $\nabla f_{t,y} = \sqrt{1 - e^{-2t}}(\nabla f)_{t,y}$. Hence,

$$w_1(f_{t,y}) = |\mathbb{E}\nabla f_{t,y}|^2 = (1 - e^{-2t})|\mathbb{E}\nabla f(\sqrt{1 - e^{-2t}}X + e^{-t}y)|^2.$$

Now set Y to be a standard Gaussian vector in $\mathbb{R}^n$, independent of X. Then

$$\begin{aligned}
\mathbb{E}w_1(f_{t,Y}) &= (1 - e^{-2t})\mathbb{E}_Y|\mathbb{E}_X\nabla f(\sqrt{1 - e^{-2t}}X + e^{-t}Y)|^2 \\
&= (1 - e^{-2t})\mathbb{E}_Y|(P_t\nabla f)(Y)|^2 \\
&= (e^{2t} - 1)\mathbb{E}_Y|(\nabla P_t f)(Y)|^2,
\end{aligned}$$

where the last line follows because $P_t\nabla f = e^t\nabla P_t f$. Finally, the Poincaré inequality applied to $P_t f$ yields

$$\mathbb{E}w_1(f_{t,Y}) = (e^{2t} - 1)\mathbb{E}|\nabla P_t f|^2 \geq (e^{2t} - 1)\,\mathrm{Var}(P_t f).$$ □

Proof of Theorem 2.1. By Proposition 2.8, there exists some $y \in \mathbb{R}^n$ such that $w_1(f_{t,y}) \geq \mathrm{Var}(P_t f)$. Applying Proposition 2.6 to $f_{t,y}$ completes the proof. □

2.3 The converse of Theorem 2.1

The following result is a (qualitative) converse of Theorem 2.1. For example, it implies that if $M(f_{s,Y})$ is non-negligible with constant probability then f is noise stable. In particular, together with Theorem 2.1 it implies Theorem 1.3.

Theorem 2.9. *For any $0 < r < s$ and any $f : \mathbb{R}^n \to [0,1]$,*

$$(1 - e^{-2(s-r)}) \operatorname{Var}(P_r f) \geq 4\mathbb{E}_Y M^2(f_{s,Y}) - C \left(\frac{1 - e^{-2r}}{1 - e^{-2s}} \right)^{1/4}.$$

Lemma 2.10. *For any half-space A and any $t > 0$,*

$$\mathbb{E}[(1_A - P_t 1_A)^2] \leq \frac{1}{\pi} \arccos(e^{-t}).$$

Proof. Ledoux's bound gives

$$\mathbb{E}[(1 - 1_A) P_t 1_A] \leq \frac{\arccos(e^{-t})}{2\pi}.$$

Rearranging this,

$$\mathbb{E}[1_A P_t 1_A] \geq \gamma_n(A) - \frac{\arccos(e^{-t})}{2\pi}. \tag{2.2}$$

On the other hand,

$$\mathbb{E}[(1_A - P_t 1_A)^2] = \gamma_n(A) - 2\mathbb{E}[1_A P_t 1_A] + \mathbb{E}[(P_t 1_A)^2] \leq 2\gamma_n(A) - 2\mathbb{E}[1_A P_t 1_A].$$

Applying (2.2) completes the proof. □

Next, we show that any set which is correlated with a half-space must be noise stable (indeed, almost as noise stable as the half-space itself).

Proposition 2.11. *Suppose that $A \subset \mathbb{R}^n$ is a half-space. Then for any $f : \mathbb{R}^n \to [0,1]$ and any $t > 0$,*

$$\begin{aligned} \operatorname{Var}(P_t f) &\geq \frac{\operatorname{Cov}(A, f)^2}{\big(\gamma_n(A)(1 - \gamma_n(A))\big)^2} \operatorname{Var}(P_t 1_A) - \frac{\sqrt{\arccos(e^{-2t})}}{\sqrt{\pi}} \\ &\geq 4 \operatorname{Cov}(A, f)^2 - C t^{1/4} \end{aligned}$$

for a universal constant C.

Proof. Let $g = 1_A - \gamma_n(A)$ and $h = f - \mathbb{E} f$, so that g and h both have mean zero and $\mathbb{E}[gh] = \operatorname{Cov}(1_A, f)$. Write $h = cg + h^\perp$, where $\mathbb{E}[gh^\perp] = 0$; then $c = \mathbb{E}[gh]/\mathbb{E}[g^2] = \operatorname{Cov}(1_A, f)/\operatorname{Var}(1_A)$. Since $P_t f - \mathbb{E} f = P_t h$, we have

$$\operatorname{Var}(P_t f) = \mathbb{E}[(P_t h)^2] = \mathbb{E}[c^2 (P_t g)^2 + (P_t h^\perp)^2 + 2c P_t g P_t h^\perp]. \tag{2.3}$$

Now, $\mathbb{E}[(P_t g)^2] = \operatorname{Var}(P_t 1_A)$ and $\mathbb{E}[(P_t h^\perp)^2] \geq 0$. For the last term, since $\mathbb{E}[gh^\perp] = 0$, the Cauchy-Schwarz inequality implies

$$\mathbb{E}[P_t g P_t h^\perp] = \mathbb{E}[h^\perp P_{2t} g] = \mathbb{E}[h^\perp (P_{2t} g - g)] \geq -\sqrt{\mathbb{E}[(h^\perp)^2] \mathbb{E}[(P_{2t} g - g)^2]}.$$

Since $P_{2t} g - g = P_{2t} 1_A - 1_A$, Lemma 2.10 implies that

$$\mathbb{E}[P_t h P_t g] \geq -\frac{\sqrt{\mathbb{E}[(h^\perp)^2] \arccos(e^{-2t})}}{\sqrt{\pi}}.$$

Going back to (2.3) and using the bound $\mathbb{E}[(h^\perp)^2] \le \mathbb{E}[h^2] \le 1$,

$$\mathrm{Var}(P_t f) \ge c^2 \,\mathrm{Var}(P_t 1_A) - \frac{\sqrt{\arccos(e^{-2t})}}{\sqrt{\pi}}.$$

Recalling that $c = \mathrm{Cov}(A, f)/\mathrm{Var}(1_A)$, this proves the first claimed inequality.

For the second inequality, note that Lemma 2.10 implies that

$$\frac{\mathrm{Var}(P_t 1_A)}{\mathrm{Var}(A)} \ge 1 - \frac{\arccos(e^{-2t})}{2\pi\,\mathrm{Var}(A)}.$$

Combining this with the first claimed inequality,

$$\begin{aligned}\mathrm{Var}(P_t f) &\ge \frac{\mathrm{Cov}(A,f)^2}{\mathrm{Var}(A)}\left(1 - \frac{C\arccos(e^{-2t})}{\mathrm{Var}(A)}\right) - Ct^{1/4} \\ &\ge 4\,\mathrm{Cov}(A,f)^2 - C\frac{\mathrm{Cov}(A,f)^2}{\mathrm{Var}(A)}\arccos(e^{-2t}) - Ct^{1/4}.\end{aligned}$$

Finally, $\mathrm{Cov}(A,f)^2 \le \mathrm{Var}(A)$ and $\arccos(e^{-2t}) \le Ct^{1/4}$, thus proving the second inequality. □

In order to relate the noise stability of f to half-spaces correlated with $f_{t,y}$, note that

$$\mathbb{E}_Y \mathbb{E}[f_{s,Y} P_{2t} f_{s,Y}] = \mathbb{E}[f P_{2r} f]$$

when $e^{-2r} = e^{-2s} + e^{-2t} - e^{-2s-2t}$. Hence,

$$\mathrm{Var}(P_r f) = \mathbb{E}_Y \mathrm{Var}(P_t f_{s,Y}) + \mathrm{Var}(P_s f).$$

Now, the Poincaré inequality implies that $\mathrm{Var}(P_s f) \le e^{-2(s-r)}\,\mathrm{Var}(P_r f)$; hence,

$$(1 - e^{-2(s-r)})\,\mathrm{Var}(P_r f) \ge \mathbb{E}_Y \mathrm{Var}(P_t f_{s,Y}).$$

By Proposition 2.11 applied to $f_{s,Y}$,

$$(1 - e^{-2(s-r)})\,\mathrm{Var}(P_r f) \ge 4\mathbb{E}_Y M^2(f_{s,Y}) - Ct^{1/4}.$$

To prove Theorem 2.9, note that if we fix r and s and solve for t the we obtain $e^{-2t} = 1 - \frac{1-e^{-2r}}{1-e^{-2s}}$. For small t, this gives $t = \Theta(\frac{1-e^{-2r}}{1-e^{-2s}})$ (while for large t the Theorem is vacuous anyway).

3 Boolean functions

For this section, P_t denotes the Bonami-Beckner semigroup defined in Section 1.1. Recall also the definition of f_z for $z \in \{-1,0,1\}^n$ from that section. Let μ_s be the probability distribution $e^{-s}\delta_0 + \frac{1}{2}(1-e^{-s})(\delta_1 + \delta_{-1})$ on $\{-1,0,1\}$ and take $Z_s \sim \mu_s^{\otimes n}$. Then we have the following relationship between P_s and Z_s:

$$(P_s f)(x) = \mathbb{E} f_{Z_s}(x).$$

Theorem 3.1. *There is a universal constant $c > 0$ such that for any $f : \{-1,1\}^n \to [0,1]$ and any $t > 0$,*

$$\mathbb{E}\left[M^2(f_{Z_s}) \log \frac{1}{M(f_{Z_s})}\right] \ge c(e^{2t} - 1)\,\mathrm{Var}(P_t f),$$

where s is defined by $e^{-s} + e^{-t} = 1$, and the expectation is with respect to $Z_s \sim \mu_s$.

Before proceeding with the proof of Theorem 3.1, let us make some remarks about how sharp it is. First of all, it is no longer true if we replace f_{Z_t} by f; that is, noise stable functions are not necessarily correlated with half-spaces. We demonstrate this using a boolean version of the earlier Gaussian example; details are in Section 3.2.

Next, Theorem 3.1 has a qualitative converse, which we will state later as Theorem 3.10. That is, if $M(f_{Z_s})$ is non-negligible on average then f is noise stable. In particular, Theorem 3.1 and Theorem 3.10 imply Theorem 1.1.

Finally, Theorem 3.1 implies that $M(f_{Z_t})$ is large with constant probability over Z_t. It turns out that this probability estimate cannot be substantially improved. As an example, consider the function

$$f(x) = \begin{cases} x_2 & \text{if } x_1 = 1 \\ \prod_{i=3}^n x_i & \text{if } x_1 = -1. \end{cases}$$

Then f is noise-stable, but if $z_1 = -1$ then f_z is noise sensitive and uncorrelated with any half-space. In other words, f_{Z_t} has probability $\frac{1}{2}e^{-t}$ of failing to be correlated with any half-space.

3.1 Proof of Theorem 3.1

The proof of Theorem 3.1 follows the same lines as the proof of Theorem 2.1, but it requires a little background on Fourier analysis of boolean functions: for a set $S \subset \{1, \dots, n\}$, define $\chi_S : \{-1,1\}^n \to \{-1,1\}$ by

$$\chi_S(x) = \prod_{i \in S} x_i.$$

It is well-known (see e.g. [9]) that $\{\chi_S : S \subset \{1, \dots, n\}\}$ is an orthonormal basis of $L_2(\{-1,1\}^n)$; in particular, every $f : \{-1,1\}^n \to [0,1]$ may be expanded in this basis: define $\hat{f}(S)$ as the coefficients of this expansion:

$$f(x) = \sum_{S \subset \{-1,1\}^n} \hat{f}(S)\chi_S(x).$$

Also, we abbreviate $\hat{f}(\{i\})$ by $\hat{f}(i)$, and we define

$$w_1(f) = \sum_{i=1}^n \hat{f}(i)^2.$$

Since $\hat{f}(i) = \mathbb{E}[X_i f(X)]$, we may also write

$$w_1(f) = \sum_{i=1}^n \mathbb{E}[X_i f(X)]^2,$$

as in the Gaussian case.

We will prove Theorem 3.1 in two steps. First, we will show that if $w_1(f)$ is non-negligible then there is a half-space correlated with f. Then we will show that $\mathbb{E}w_1(f_{Z_t})$ is non-negligible. Actually, the first step is already done, thanks to Remark 2.7. Indeed, Hoeffding's inequality implies that the uniform measure on $\{-1,1\}^n$ is sub-Gaussian in the sense of Remark 2.7, and so the proof of Proposition 2.6 applies with no changes to the boolean setting:

Proposition 3.2. *There is a universal constant $c > 0$ such that for every $f : \{-1,1\} \to [0,1]$,*

$$M^2(f) \log \frac{1}{M(f)} \geq c w_1(f).$$

Next, we show that $\mathbb{E}[w_1(f_Z)]$ is substantial if f is noise-stable.

Proposition 3.3. *For any $t > 0$, if $e^{-s} = 1 - e^{-t}$ then*

$$\mathbb{E}[w_1(f_{Z_s})] = (1 - e^{-t}) \sum_S |S| \hat{f}^2(S) e^{-2t(|S|-1)} \geq (e^{2t} - e^t) \operatorname{Var}(P_t f).$$

Proof. Fix t and set $Z = Z_s$. Recalling the definition of w_1, we have

$$\mathbb{E}[w_1(f_Z)] = \sum_{i=1}^n \mathbb{E}[\hat{f}_Z^2(i))].$$

Note that $\hat{f}_Z(i) = 0$ if $Z_i = \pm 1$, which happens with probability $1 - e^{-s} = e^{-t}$. Otherwise $\hat{f}_Z(i)$ is given by

$$\hat{f}_Z(i) = \sum_{S : i \in S} \hat{f}(S) \prod_{j \in S \setminus \{i\}} Z_j. \tag{3.1}$$

Therefore

$$\begin{aligned}
\mathbb{E}[\hat{f}_Z(i)^2] &= (1 - e^{-t}) \sum_{S,T : i \in S, i \in T} \hat{f}(S)\hat{f}(T)\mathbb{E}[\prod_{j \in S \setminus \{i\}} Z_j \prod_{k \in T \setminus \{i\}} Z_k] \\
&= (1 - e^{-t}) \sum_{S : i \in S} \hat{f}^2(S) e^{-2t(|S|-1)}.
\end{aligned}$$

Summing over i proves the first claim; the second follows from the fact that

$$\operatorname{Var}(P_t f) = \sum_{|S| \geq 1} e^{-2t|S|} \hat{f}^2(S) \leq \sum_S |S| e^{-2t|S|} \hat{f}^2(S).$$ □

Proof of Theorem 3.1. Taking s so that $e^{-s} = 1 - e^{-t}$ and applying Proposition 3.3 gives $\mathbb{E} w_1(f_{Z_s}) \geq (e^{2t} - e^t) \operatorname{Var}(P_t f)$. By Proposition 3.2,

$$\mathbb{E}\left[M^2(f_{Z_s}) \log \frac{1}{M(f_{Z_s})}\right] \geq c\mathbb{E} w_1(f_{Z_s}) \geq c(e^{2t} - e^t) \operatorname{Var}(P_t f).$$

Finally, $e^{2t} - e^t = e^t(e^t - 1) \geq \frac{1}{2}(e^t + 1)(e^t - 1) = \frac{1}{2}(e^{2t} - 1)$. □

3.2 An example

Let $n = m^2$, and let $J_i = \{(i-1)m, \dots, im - 1\}$. Let $B_n \subset \{-1, 1\}^n$ be the set

$$\left\{ x : \sum_{i=1}^m \left(\frac{1}{\sqrt{m}} \sum_{j \in J_i} x_j \right)^2 \leq m \right\}.$$

From the central limit theorem, one sees immediately that B_n is noise stable, with the same estimate as its Gaussian analogue in Section 2.1.

Proposition 3.4. *For any n and any $t > 0$,*

$$\operatorname{Var}(P_t 1_{B_n}) \geq \frac{1}{4} - \frac{\arccos(e^{-2t})}{\sqrt{2}\pi} - o_n(1).$$

In particular B_n is noise stable.

Finally, we show that B_n is not correlated with any half-space. This essentially follows from the invariance principle, which says that nice boolean functions have almost the same distribution when their arguments are replaced by Gaussian variables.

Proposition 3.5. $M(B_n) \le Cm^{-1/200}$

For the rest of this section, fix $x \in \mathbb{R}^n$ and $b \in \mathbb{R}$, and suppose that $A = \{x \in \{-1,1\}^n : \sum_i a_i x_i \le b\}$. Let $J^* \subset \{1,\dots,n\}$ be the set containing the indices of the $\lfloor m^{1/3} \rfloor$ largest $|a_i|$. Define a^+ by $a_i^+ = 1_{\{i \in J^*\}} a_i$ and set $a^- = a - a^+$.

We split our proof of Proposition 3.5 into two parts, depending on the decay properties of a. If a^- is unbalanced, it follows that a^+ must contain only large coordinates. We apply the Littlewood-Offord theorem to argue that a^- is essentially irrelevant and A depends only on a few coordinates. Since B_n doesn't depend on any small set of coordinates, this implies that A and B_n are uncorrelated. If a^- is fairly balanced then we condition on $\{X_i : i \in J^*\}$ and apply an invariance principle to $\{X_i : i \notin J^*\}$, replacing boolean variables with Gaussian variables and applying Proposition 2.3.

First, we recall the Littlewood-Offord inequality:

Theorem 3.6. *If X is uniformly distributed in $\{-1,1\}^n$ then*

$$\sup_{c \in \mathbb{R}} \Pr\left(\Big|\sum_i X_i a_i - c\Big| \le t \min_i |a_i| \right) \le Ctn^{-1/2}.$$

Lemma 3.7. *If $\|a^-\|_\infty \ge m^{-1/24}\|a^-\|_2$ then* $\mathrm{Cov}(1_A, 1_{B_n}) \le Cm^{-1/12}$.

Proof. By Theorem 3.6 and since $|a_i| \ge \|a^-\|_\infty$ for all $i \in J^*$,

$$\Pr\left(\Big| \sum_{j \in J^*} a_j X_j - b \Big| \le m^{1/24}\|a^-\|_2 \right) \le Cm^{1/12}|J^*|^{-1/2} \le Cm^{-1/12}.$$

On the other hand, Chebyshev's inequality implies that

$$\Pr\left(\Big| \sum_{j \notin J^*} a_j X_j \Big| \ge m^{1/24}\|a^-\|_2 \right) \le m^{-1/12}.$$

Putting these two inequalities together, we see that with probability at least $1 - Cm^{-1/12}$ over $\{X_j : j \in J^*\}$ we have

$$\Pr(X \in A \mid X_j : j \in J^*) \in [0, m^{-1/12}] \cup [1 - m^{-1/12}, 1]. \tag{3.2}$$

On the other hand, conditioning on $\{X_j : j \in J^*\}$ has little effect on the event B_n: each random variable $Z_i := \big(\sum_{j \in J_i} X_j\big)^2$ has conditional expectation $m \pm O(|J_i \cap J^*|^2)$ and conditional variance $O(m)$; moreover, $\mathbb{E}[|Z_i - \mathbb{E}Z_i|^3] = O(m^{3/2})$. Then

$$\sum_{i=1}^m \Big(\sum_{j \in J_i} X_j \Big)^2$$

has conditional expectation $m^2 \pm O(|J^*|^2) = m^2 \pm O(m^{2/3})$. By the Berry-Esseen theorem,

$$\Pr(X \in B_n \mid X_j : j \in J^*) = \frac{1}{2} \pm O(m^{-1/2}).$$

Combined with (3.2), this implies that

$$\mathbb{E}[(1_{B_n} - \Pr(B_n))1_A \mid X_j : j \in J^*] \le Cm^{-1/12}$$

with probability at least $1 - Cm^{-1/12}$. Integrating over $\{X_j : j \in J^*\}$, this implies the claim. □

Since Lemma 3.7 implies Proposition 3.5 in the case $\|a^-\|_\infty \geq m^{-1/24}\|a^-\|_2$, we may assume from now on that $\|a^-\|_\infty \leq m^{-1/24}\|a^-\|_2$. We will prove the remaining case of Proposition 3.5 in two steps: for the rest of the section, let X be uniform on $\{-1,1\}^n$ and take $Y \sim \gamma_n$; note that A and B_n can be canonically extended to subsets of $\mathbb{R}^n$.

For any $c \in \mathbb{R}$, let $h_c : \mathbb{R} \to [0,1]$ be the function $h_c(x) = 1_{\{x \leq c\}}$. For $\epsilon > 0$, let $h_{c,\epsilon}$ be a function satisfying

- $h_{c,\epsilon}$ takes values in $[0,1]$,
- $h_{c,\epsilon}(x) = h_c(x)$ for all x such that $|x - c| \geq \epsilon$, and
- for $k = 1,2,3$, $h_{c,\epsilon}^{(k)}$ is uniformly bounded by $C\epsilon^{-k}$ for some universal constant C (where $h^{(k)}$ denotes the kth derivative of h).

For $z \in \{-1,1\}^{J^*}$ and let Ω_z be the event $\{X_i = z_i \ \forall i \in J^*\}$. Set $J_i' = J_i \setminus J^*$ and $s_i = \sum_{j \in J_i \cap J^*} z_i$. Next, define the polynomials

$$
\begin{aligned}
p(x) &= \frac{1}{m^2} \sum_i \Big(\sum_{j \in J_i} x_j \Big)^2 \\
p_z(x) &= \frac{1}{m^2} \sum_i \Big(\sum_{j \in J_i'} x_j + s_i \Big)^2 \\
q_z(x) &= \frac{1}{\|a^-\|} \Big(\sum_{j \notin J^*} a_j x_j + \sum_{j \in J^*} a_j z_j \Big).
\end{aligned}
$$

Recalling (from the Berry-Esseen theorem) that $\Pr(X \in B_n) = \frac{1}{2} + O(m^{-1/2})$, our goal is to show that

$$\mathbb{E}\Big[1_A(X)\Big(1_{B_n}(X) - \frac{1}{2}\Big)\Big] \leq C m^{-1/12}.$$

We will achieve this by conditioning on Ω_z: for an arbitrary z, we claim that

$$\mathbb{E}\Big[1_A(X)\Big(1_{B_n}(X) - \frac{1}{2}\Big) \ \Big| \ \Omega_z\Big] \leq C m^{-1/12}.$$

Going back to the definitions of p_z and q_z, this is equivalent to

$$\mathbb{E}\Big[h_{b'}(q_z(X))\Big(h_1(p_z(X)) - \frac{1}{2}\Big)\Big] \leq C m^{-1/12}, \tag{3.3}$$

We divide the proof of (3.3) into several steps: for any $\epsilon > 0$,

$$\mathbb{E}|h_1(p_z(X)) - h_1(p(X))| \leq C m^{-1/6} \tag{3.4}$$
$$\mathbb{E}|h_{1,\epsilon}(p(X)) - h_1(p(X))| \leq C \max\{\epsilon, m^{-1/2}\} \tag{3.5}$$
$$\mathbb{E}|h_{b',\epsilon}(q_z(X)) - h_{b'}(q_z(X))| \leq C \max\{\epsilon, m^{-1/24}\} \tag{3.6}$$
$$|\mathbb{E}[h_{b',\epsilon}(q_z(X))h_{1,\epsilon}(p(X))] - \mathbb{E}[h_{b',\epsilon}(q_z(Y))h_{1,\epsilon}(p(Y))]| \leq C\epsilon^{-3} m^{-1/48} \tag{3.7}$$
$$\mathbb{E}|h_{1,\epsilon}(p(Y)) - h_1(p(Y))| \leq C \max\{\epsilon, m^{-1/2}\} \tag{3.8}$$
$$\mathbb{E}|h_{b',\epsilon}(q_z(Y)) - h_{b'}(q_z(Y))| \leq C \max\{\epsilon, m^{-1/2}\} \tag{3.9}$$
$$\operatorname{Cov}(h_{b'}(q_z(Y)), h_1(p(Y))) \leq C m^{-1/2}. \tag{3.10}$$

Taking $\epsilon = m^{-1/200}$ and combining (3.4) through (3.10) using the triangle inequality yields (3.3).

Fortunately, most of the pieces above are easy: (3.5) follows from the Berry-Esseen theorem, since $h_{1,\epsilon}$ and h_1 are both bounded by one, and agree except on an interval of length 2ϵ. Inequalities (3.6), (3.8), and (3.9) follow by the same argument (the reason for the worse bound in (3.6) is because the error term in the Berry-Esseen theorem depends on $\|a^-\|_\infty/\|a^-\|_2$, which we only know to be bounded by $m^{-1/24}$).

It remains to check (3.4), (3.7), and (3.10); for these, it helps to introduce the notion of influences: for function $f : \{-1,1\}^n \to \mathbb{R}$, we define the influence of the ith coordinate to be

$$\mathrm{Inf}_i(f) = \mathrm{Var}\,\mathbb{E}[f(X) \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n].$$

If the range of f is $\{-1,1\}$ then $\mathrm{Inf}_i(f)$ is just the probability that negating X_i will change the value of $f(X)$.

For (3.4), note that the Berry-Esseen theorem applied to $S_k := \left(\sum_{j\in J_k} X_j\right)$ implies that with probability at least $1-Cm^{-1/6}$, $h_1(p(X))$ falls outside the interval $[1-6m^{-2/3}, 1+6m^{-2/3}]$. Hence, in order to change the value of $h_1(p(X))$, one would need to change the value of $\sum_k S_k^2$ by at least $6m^{4/3}$. On the other hand, Hoeffding's inequality implies that with probability at least $1 - Cm^{-1/6}$, $\max_k |S_k| \le 2m$. On this event, in order to change the value of $\sum_k S_k^2$ by $6m^{4/3}$, one would need to change at least $2m^{1/3}$ of the X_j. Since $p_z(X)$ is obtained from $p(X)$ by changing at most $m^{1/3}$ of the X_j, we see that $h_1(p(X)) = h_1(p_z(X))$ unless one of the two events above fails. This proves (3.4).

Recognizing that $h_1(p(Y)) = 1_{B_n}(Y)$ and $h_{b'}(q_z(Y))$ is the indicator function of some half-space, the following Lemma proves (3.10).

Lemma 3.8. *For any half-space A,* $\mathrm{Cov}(1_A(Y), 1_{B_n}(Y)) \le m^{-1/2}$

Proof. The covariance in question can be written in terms of covariances between half-spaces and m-dimensional balls, which we may then bound using Proposition 2.3. To do this, we break each block of m variables in terms of its contribution in the $(1, \dots, 1)$ direction and the contribution in the orthogonal direction: for each block J of m variables, define

$$x_J = m^{-1/2} \sum_{j\in J} x_j, \quad a_J = \mathbb{E}\big[X_J \sum_{j\in J} a_j x_j\big] = m^{-1/2} \sum_{j\in J} a_j$$

and

$$r_J = \sqrt{\sum_{j\in J} a_j^2 - a_J^2}, \quad r = \sqrt{\sum_J r_J^2}.$$

Now define $A', B' \subset \mathbb{R}^{m+1}$ by

$$A' = \left\{ x \in \mathbb{R}^{m+1} : \sum_{i=1}^m a_{J_i} x_i + r x_{m+1} \le b \right\}$$
$$B' = \left\{ x \in \mathbb{R}^{m+1} : \sum_{i=1}^m x_i^2 \le m \right\}.$$

Note that A' and B' are the push-forwards of $\tilde{A}_n$ and $\tilde{B}$ under a map that preserves the standard Gaussian measure: if $\Pi_m : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $\Pi_m x = (x_{J_1}, \dots, x_{J_m})$ and Π is defined by

$$\Pi x = (\Pi_m x, r^{-1}(\langle a, x\rangle - \langle \Pi_m a, \Pi_m x\rangle))$$

then $x \in A$ (resp. B) if and only if $\Pi x \in A'$ (resp. B'). Since Π pushes forward γ_n onto γ_{m+1}, we have

$$\mathrm{Cov}(1_{\tilde{A}}, 1_{\tilde{B}_n}) = \mathrm{Cov}(1_{A'}, 1_{B'}).$$

On the other hand, $\mathrm{Cov}(1_{A'}, 1_{B'}) \le m^{-1/2}$ by Proposition 2.3. □

Finally, (3.7) follows from the following multivariate invariance principle that was proved by the first author in [8]:

Theorem 3.9. *Suppose $p(x)$ and $q(x)$ are polynomials of degree at most d such that $\mathrm{Inf}_i(p) \le \tau$ and $\mathrm{Inf}_i(q) \le \tau$ for all i. For any $\Psi : \mathbb{R}^2 \to \mathbb{R}$ with third partial derivatives uniformly bounded by B,*

$$|\mathbb{E}\Psi(p(X), q(X)) - \mathbb{E}\Psi(p(Y), q(Y))| \le C^d dB\sqrt{\tau},$$

where $Y \sim \gamma_n$, X is uniform on $\{-1,1\}^n$, and C is a universal constant.

Taking $d = 2$, $\tau = m^{-1/24}$ and $\Psi(x, y) = h_{1,\epsilon}(x)h_{b',\epsilon}(y)$ (which has third derivatives bounded by $C\epsilon^{-3}$) proves (3.7).

3.3 The converse of Theorem 3.1

Here, we state and prove the boolean analogue of Theorem 2.9 (or, the qualitative converse of Theorem 3.1). That is, we show that if $M(f_{s,Y})$ is non-negligible with constant probability then f is noise stable.

Theorem 3.10. *For any $0 < r < s$ and any $f : \{-1,1\}^n \to [-1,1]$,*

$$(1 - e^{-2(s-r)}) \operatorname{Var}(P_r f) \ge 4\mathbb{E}M^2(f_{Z_s}) - C\left(\frac{1 - e^{-2r}}{1 - e^{-2s}}\right)^{1/4},$$

where $Z_s \sim \mu_s$ and C is a universal constant.

The proof of Theorem 3.10 is very much like the proof of Theorem 2.9, so we give only a sketch. As in the proof of Theorem 2.9, the first step is a bound on the noise stability of half-spaces. However, the bound that we used to prove Lemma 2.10 is not known for boolean functions (it would be equivalent to a weak version of the "majority is least stable" conjecture). Instead, we use a weaker (by a constant factor) bound due to Peres [10]:

Theorem 3.11. *For any half-space A and any $t > 0$, $\mathbb{E}[(1_A - P_t 1_A)^2] \le C\sqrt{t}$, where C is a universal constant.*

Next, we show that any set which is correlated with a half-space must be noise stable (indeed, almost as noise stable as the half-space itself).

Proposition 3.12. *Suppose that $A \subset \{-1,1\}^n$ is a half-space. Then for any function $f : \{-1,1\}^n \to [0,1]$ and any $t > 0$,*

$$\operatorname{Var}(P_t f) \ge 4 \operatorname{Cov}(A, f)^2 - Ct^{1/4}$$

for a universal constant C.

The proof of Proposition 3.12 is essentially identical to the proof of Proposition 2.11, so we omit it. The only difference is that we use Theorem 3.11 instead of Lemma 2.10.

Finally, the argument to go from Proposition 3.12 to Theorem 3.10 is also essentially identical to the Gaussian case: the only property of Gaussians that we used in that argument was the Poincaré inequality, which takes the same form in the boolean case.

3.4 Acknowledgement

We thank Dana Moshkovitz, Gil Kalai and Irit Dinur for encouragement to complete this work. E.M acknowledges the support of NSF grant CCF 1320105, DOD ONR grant N00014-14-1-0823, and grant 328025 from the Simons Foundation".

We also thank the anonymous reviewers for many comments that helped us to improve the presentation, most notably a substantial improvement to the proofs of Theorems 2.1 and 3.1.

References

[1] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Springer, 2014.

[2] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[3] D. M. Kane. The Gaussian surface area and noise sensitivity of degree-d polynomial threshold functions. *Computational Complexity*, 20(2):389–412, 2011.

[4] S. Khot and D. Moshkovitz. Candidate hard Unique Game. In *STOC*, 2016. to appear. MR-3536555

[5] A. Klivans, R. O'Donnell, and R. Servedio. Learning intersections and thresholds of halfspaces. *Journal of Computer and System Sciences*, 68(4):808–840, 2004.

[6] M. Ledoux. Semigroup proofs of the isoperimetric inequality in Euclidean and Gauss space. *Bulletin des sciences mathématiques*, 118(6):485–510, 1994. MR-1309086

[7] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

[8] E. Mossel. Gaussian bounds for noise correlation of functions. *GAFA*, 19:1713–1756, 2010.

[9] R. O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.

[10] Y. Peres. Noise stability of weighted majority. arXiv:math/0412377, 2004.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm