Gaussian Noise Sensitivity and Fourier Tails

Guy Kindler[*]   Ryan O'Donnell[†]

December 8, 2011

Abstract

We observe a subadditivity property for the noise sensitivity of subsets of Gaussian space. For subsets of volume $\frac{1}{2}$, this leads to an almost trivial proof of Borell's Isoperimetric Inequality for $\rho = \cos(\frac{\pi}{2\ell})$, $\ell \in \mathbb{N}$. In turn this can be used to obtain the Gaussian Isoperimetric Inequality for volume-$\frac{1}{2}$ sets and also .8787-factor UG-hardness for Max-Cut (within 10^{-4} of the optimum). As another corollary we show the Hermite tail bound $\|f^{\geq k}\|_2^2 \geq \Omega(\mathbf{Var}[f]) \cdot \frac{1}{\sqrt{k}}$ for $f : \mathbb{R}^n \to \{-1, 1\}$. Combining this with the Invariance Principle shows the same Fourier tail bound for any Boolean $f : \{-1, 1\}^n \to \{-1, 1\}$ with all its noisy-influences small. This improves on a result of Bourgain in the Boolean setting, which only had $\frac{1}{k^{1/2+o(1)}}$. Without using Invariance, we also show how to simplify and improve Bourgain's proof to obtain the bound $\frac{1}{\sqrt{k}\log^{1.5}k}$.

1 Our results

In this paper we study noise sensitivity and Fourier and Hermite tail bounds for Boolean-valued functions, defined either on the discrete Boolean cube or on multidimensional Gaussian space.

For functions on Gaussian space we define "rotation sensitivity", $\mathbf{RS}_f(\epsilon)$, a slightly different parametrization of noise sensitivity. With a very simple proof (reminiscent of [9]) we show that rotation sensitivity is subadditive. In particular:

Theorem 1.1. *Let* $f : \mathbb{R}^d \to \{-1, 1\}$. *Then* $\mathbf{RS}_f(\epsilon) \geq \frac{1}{\ell} \mathbf{RS}_f(\ell \epsilon)$ *for any* $\epsilon \in \mathbb{R}$, $\ell \in \mathbb{N}$.

As a direct corollary we obtain that the ϵ-rotation sensitivity of an unbiased function f is bounded from below by $\frac{\epsilon}{\pi}$ for any ϵ of the form $\frac{\pi}{2\ell}$ where $\ell \in \mathbb{N}$. This we state as follows:

Theorem 1.2. *Let* $A \subseteq \mathbb{R}^d$ *have Gaussian measure* $\frac{1}{2}$. *Then* $\mathbf{RS}_A(\epsilon) \geq \frac{\epsilon}{\pi}$ *for any* $\epsilon = \frac{\pi}{2\ell}$, $\ell \in \mathbb{N}$.

Theorem 1.2 is an isoperimetric bound for Gaussian space; it was proved by Borell [5] for *all* $\epsilon \in [0, \frac{\pi}{2}]$. Taking $\epsilon \to 0$ in Borell's Theorem yields the Gaussian Isoperimetric Inequality (GII); we obtain a simple proof of the GII for volume-$\frac{1}{2}$ sets by taking $\ell \to \infty$ in our Theorem 1.2. Taking $\epsilon = \epsilon_0 \approx \frac{\pi}{3.876}$ in Borell's Theorem and combining with the Invariance Principle [19] yields the optimal .8786-factor UG-hardness result for Max-Cut [12]. Theorem 1.2 gives a simple proof of Borell's Theorem for $\epsilon = \frac{\pi}{4}$, and using this in place of ϵ_0 gives .8787-factor hardness.

Theorem 1.1 is also the key to a very simple proof of the following "Hermite tail bound", which is optimal up to constant factors:

Theorem 1.3. *Let* $f : \mathbb{R}^d \to \{-1, 1\}$ *be a function with constant variance, i.e.* $\mathbf{Var}[f] \geq \Omega(1)$. *Then* $\sum_{|\beta| \geq k} \widehat{f}(\beta)^2 \geq \Omega(\frac{1}{\sqrt{k}})$ *for any* $k \in \mathbb{N}$, *where the* $\widehat{f}(\beta)$*'s are* f*'s Hermite coefficients.*

Theorem 1.3 is a Gaussian version of the following well-known result of Bourgain [6] on the Fourier tail of Boolean functions:

Bourgain's Theorem. *Let* $f : \{-1, 1\}^n \to \{-1, 1\}$ *satisfy* $\mathbf{Var}[f] \geq \Omega(1)$ *and* $\mathbf{Inf}_i^{\leq k}[f] \leq k^{-O(k)}$ *for all* $i \in [n]$. *Then* $\sum_{|S| \geq k} \widehat{f}(S)^2 \geq \frac{1}{\sqrt{k}} \cdot \frac{1}{2^{\tilde{O}(\sqrt{\log k})}}$ *for any* $k \in \mathbb{N}$.

(Bourgain's paper wrote simply $\frac{1}{k^{1/2+o(1)}}$; the more precise bound above is from [13].)

By a Central Limit Theorem argument, Bourgain's Theorem implies Theorem 1.3 with the bound $\frac{1}{\sqrt{k}} \cdot \frac{1}{2^{\tilde{O}(\sqrt{\log k})}}$; prior to our work, this was best known bound for Hermite tails. Combining our Theorem 1.3 with the Invariance Principle from [19] improves Bourgain's Theorem on Boolean Fourier tails to the optimal $\Omega(\frac{1}{\sqrt{k}})$, even assuming only $\mathbf{Inf}_i^{\leq k}[f] \leq 2^{-O(k)}$ for all i.

Finally, we give a simplified and improved version of Bourgain's proof which avoids the use of Invariance and yields the nearly optimal Fourier tail bound $\sum_{|S| \geq k} \widehat{f}(S)^2 \geq \frac{1}{\sqrt{k}} \cdot \frac{1}{O(\log^{1.5} k)}$. We state here a version of the key intermediate theorem, a lower bound on noise sensitivity which has an "elementary" (i.e., hypercontractivity-free) proof:

Theorem 1.4. *Let* $f : \{-1, 1\}^n \to \{-1, 1\}$ *and let* $\delta \in [0, \frac{1}{2}]$. *Suppose* $\sum_{i=1}^n \widehat{f}(i)^2 \geq \Omega(1)$ *and* $|\widehat{f}(i)| \ll \sqrt{\delta}/\ln(1/\delta)$ *for all* $i \in [n]$. *Then* $\mathbf{NS}_f(\delta) \geq \Omega(\sqrt{\delta/\ln(1/\delta)})$.

2 Gaussian sensitivity and tails

Throughout this section we think of $\mathbb{R}^d$ as being endowed with the standard Gaussian distribution. We will mainly be concerned with subsets $A \subseteq \mathbb{R}^d$ and their ± 1-indicator functions, $f : \mathbb{R}^d \to \{-1, 1\}$ (we call such functions *Boolean-valued*). We will use notation such as

$$\mathrm{vol}(A) = \Pr_{\mathbf{X} \sim \mathrm{N}(0,1)^d}[\mathbf{X} \in A] = \mathbf{E}_{\mathbf{X} \sim \mathrm{N}(0,1)^d}[\tfrac{1}{2} + \tfrac{1}{2}f(\mathbf{X})] = \mathbf{E}[\tfrac{1}{2} + \tfrac{1}{2}f].$$

We assume all functions $f : \mathbb{R}^d \to \mathbb{R}$ mentioned are measurable.

2.1 Rotation sensitivity

Definition 2.1. Let $\rho \in [-1, 1]$. We say that the $\mathbb{R}^d$-valued random variables $\mathbf{X}, \mathbf{Y}$ are ρ-*correlated Gaussians* if they are jointly Gaussian with mean zero, $\mathbf{E}[\mathbf{X}_i \mathbf{Y}_i] = \rho$ for $i \in [n]$, and $\mathbf{E}[\mathbf{X}_i \mathbf{Y}_j] = 0$ for $i \neq j$.

Definition 2.2. Let $f : \mathbb{R}^d \to \{-1, 1\}$ and let $\epsilon \in \mathbb{R}$. The *rotation sensitivity of f at ϵ* is defined to be

$$\mathbf{RS}_f(\epsilon) = \Pr[f(\mathbf{X}) \neq f(\mathbf{Y})], \text{ where } \mathbf{X}, \mathbf{Y} \text{ are } \cos(\epsilon)\text{-correlated Gaussians.}$$

Usually we assume $\epsilon \in [0, \pi]$. If f is the indicator of $A \subseteq \mathbb{R}^d$ we also write $\mathbf{RS}_A(\epsilon)$, the probability that one of $\mathbf{X}, \mathbf{Y}$ is in A and the other is out.

Fact 2.3. *Let $f : \mathbb{R}^d \to \{-1, 1\}$ be the indicator of $A \subseteq \mathbb{R}^d$. Then:*

1. $\mathbf{RS}_f(0) = 0$ *since 1-correlated Gaussians are identical almost surely.*

2. $\mathbf{RS}_f(\frac{\pi}{2}) = 2\mathrm{vol}(A)(1 - \mathrm{vol}(A)) = \frac{1}{2}\mathbf{Var}[f]$, *since 0-correlated Gaussians are independent. In particular, $\mathbf{RS}_f(\frac{\pi}{2}) = \frac{1}{2}$ when $\mathrm{vol}(A) = \frac{1}{2}$ (i.e $\mathbf{E}[f] = 0$).*

3. $\mathbf{RS}_f(\pi) = 1$ *if f is odd (i.e. $f(-x) = -f(x)$ or $-A = A^c$).*

The rotation sensitivity of a set A is a kind of measure of its boundary size. The associated isoperimetric problem was solved by Borell [5]:

Borell's Theorem. *Fix $\epsilon \in [0, \frac{\pi}{2}]$ and $\alpha \in [0, 1]$. Then for any $A \subseteq \mathbb{R}^d$ satisfying $\mathrm{vol}(A) = \alpha$ it holds that $\mathbf{RS}_A(\epsilon) \geq \mathbf{RS}_H(\epsilon)$, where H is a halfspace of volume α.*

Borell's proof uses Ehrhard symmetrization and differential equations; in fact it gives a much stronger statement than the above. For another proof using two-point symmetrization on the sphere and Poincaré's lemma, see [1, 7]. When $\alpha = \frac{1}{2}$, the minimizing halfspaces in Borell's Theorem pass through the origin and there is a closed form for their rotation sensitivity. This easy result is known as Sheppard's Formula [20]:

Sheppard's Formula. *Let $H \subseteq \mathbb{R}^d$ be a halfspace through the origin. Then $\mathbf{RS}_H(\epsilon) = \frac{\epsilon}{\pi}$.*

2.2 Subadditivity of rotation sensitivity

We now prove that rotation sensitivity is subadditive.

Theorem 2.4. *Let $f : \mathbb{R}^d \to \{-1, 1\}$. Then $\mathbf{RS}_f(\epsilon_1 + \cdots + \epsilon_\ell) \leq \mathbf{RS}_f(\epsilon_1) + \cdots + \mathbf{RS}_f(\epsilon_\ell)$ for $\epsilon_1, \ldots, \epsilon_\ell \in \mathbb{R}$.*

Proof. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be independent standard d-dimensional Gaussians. For each $\theta \in \mathbb{R}$, define $\boldsymbol{Z}(\theta) = \cos(\theta)\boldsymbol{X} + \sin(\theta)\boldsymbol{Y}$, which is also a standard Gaussian. Note that $\boldsymbol{Z}(\theta)$ and $\boldsymbol{Z}(\theta + \epsilon)$ are correlated d-dimensional Gaussians with correlation $\cos(\theta)\cos(\theta + \epsilon) + \sin(\theta)\sin(\theta + \epsilon) = \cos(\epsilon)$. Thus

$$
\begin{aligned}
\mathbf{RS}_f(\epsilon_1 + \cdots + \epsilon_\ell) &= \mathbf{Pr}[f(\boldsymbol{Z}(0)) \neq f(\boldsymbol{Z}(\epsilon_1 + \cdots + \epsilon_\ell))] \\
&\leq \mathbf{Pr}[f(\boldsymbol{Z}(0)) \neq f(\boldsymbol{Z}(\epsilon_1))] + \mathbf{Pr}[f(\boldsymbol{Z}(\epsilon_1) \neq f(\boldsymbol{Z}(\epsilon_1 + \epsilon_2))] + \cdots \\
&\qquad \cdots + \mathbf{Pr}[f(\boldsymbol{Z}(\epsilon_1 + \cdots + \epsilon_{\ell-1})) \neq f(\boldsymbol{Z}(\epsilon_1 + \cdots + \epsilon_\ell))] \\
&= \mathbf{RS}_f(\epsilon_1) + \cdots + \mathbf{RS}_f(\epsilon_\ell),
\end{aligned}
$$

where the inequality is the union bound. $\qquad\square$

Theorem 1.1 is an immediate consequence, and combining it with Fact 2.3 yields the following generalization of Theorem 1.2:

Theorem 2.5. *Let* $f : \mathbb{R}^d \to \{-1,1\}$ *and let* $\epsilon = \frac{\pi}{2\ell}$, $\ell \in \mathbb{N}$. *Then* $\mathbf{RS}_f(\epsilon) \geq \frac{\epsilon}{\pi} \cdot \mathbf{Var}[f]$. *In particular, if* $\mathrm{vol}(A) = \frac{1}{2}$ *then* $\mathbf{RS}_A(\epsilon) \geq \frac{\epsilon}{\pi}$. *If* f *is odd then we obtain the inequality for any* $\epsilon = \frac{\pi}{\ell}$.

Thus we have recovered the $\alpha = \frac{1}{2}$ case of Borell's Theorem for any $\epsilon = \frac{\pi}{2\ell}$, $\ell \in \mathbb{N}$.

Remark 2.6. It is easy to see that our proof of $\mathbf{RS}_A(\epsilon) \geq \frac{\epsilon}{\pi}$ indeed has equality when A is a halfspace through the origin. When we take $\epsilon_1 = \cdots = \epsilon_\ell = \frac{\pi}{2\ell}$ in Theorem 2.4, the points $\boldsymbol{Z}(j\frac{\pi}{2\ell})$, $j = 0\ldots\ell$ in the proof go from $\boldsymbol{X}$ to $-\boldsymbol{X}$ along the ellipse containing $\pm\boldsymbol{X}$, $\pm\boldsymbol{Y}$. With probability 1, any half-ellipse thus defined will cross A's boundary (a hyperplane through the origin) exactly once. Thus the union bound in the proof of Theorem 2.4 will always have equality.

2.3 Applications

We now mention some applications of Borell's Theorem which our proof is sufficient to obtain. The first is the Gaussian Isoperimetric Inequality (GII) for sets of volume $\frac{1}{2}$.

Gaussian Isoperimetric Inequality [21, 3]. *Fix* $\alpha \in [0,1]$. *Then for any* $A \subseteq \mathbb{R}^d$ *satisfying* $\mathrm{vol}(A) = \alpha$ *it holds that* $\mathrm{surf}(A) \geq \mathrm{surf}(H)$, *where* H *is a halfspace of volume* α. *Here* $\mathrm{surf}(A)$ *denotes Gaussian surface area. In particular, for* $\alpha = \frac{1}{2}$ *we have* $\mathrm{surf}(A) \geq \frac{1}{\sqrt{2\pi}}$.

To make this theorem precise we need to define Gaussian surface area. There are several possible choices here, all of which coincide for sufficiently "nice" sets A. The traditional definition of Gaussian surface area (for Borel sets A) is:

$$
\gamma^+(A) = \liminf_{\delta \to 0^+} \frac{\gamma(A^{+\delta}) - \gamma(A)}{\delta} \in [0, \infty]. \tag{1}
$$

In the context of Gaussian space using $A^{+\delta}$ is perhaps not completely natural; we propose the following alternative definition (for any measurable $A \subseteq \mathbb{R}^d$):

$$
\mathrm{surf}(A) = \sqrt{\frac{\pi}{2}} \cdot \limsup_{\epsilon \to 0^+} \frac{\mathbf{RS}_f(\epsilon)}{\epsilon} \in [0, \infty]. \tag{2}
$$

In Appendix A we discuss why the definitions coincide for nice enough sets. We also mention that Ledoux [17] has shown $\mathrm{surf}(A) \leq \gamma^+(A)$ always and hence using $\mathrm{surf}(A)$ in the GII is formally stronger.

Accepting surf(A) as the definition of Gaussian surface area, it is immediate that the GII follows from Borell's Theorem. Further, it is immediate that the GII for sets of Gaussian volume $\frac{1}{2}$ follows from our simpler Theorem 2.5, by taking $\ell \to \infty$. We remark that Ledoux [17, (8.25)] has also given a simpler proof of the GII in the volume-$\frac{1}{2}$ case.

The next application of Borell's Theorem we mention is in the area of hardness for approximation algorithms. We have the following result from [12, 19]:

Theorem 2.7. *([12, 19]) Fix $\epsilon \in (0, \frac{\pi}{2})$. Then for any $\eta > 0$ it is "Unique Games-hard" to $(\frac{1}{2} + \frac{1}{2}\cos(\epsilon), 1 - s(\epsilon) + \eta)$-approximate the Max-Cut problem. Here $s(\epsilon) = \inf\{\mathbf{RS}_A(\epsilon) : A \subseteq \mathbb{R}^d, \mathrm{vol}(A) = \frac{1}{2}\}$.*

By applying Borell's Theorem, we get hardness of $(\frac{1}{2} + \frac{1}{2}\cos(\epsilon), 1 - \frac{\epsilon}{\pi} + \delta)$-approximating Max-Cut. Taking $\epsilon \sim 2\sqrt{\delta}$ for $\delta \to 0$ yields $(1 - \delta, 1 - (\frac{2}{\pi} + o(1))\sqrt{\delta})$-hardness. Taking $\epsilon = \epsilon_0 \approx \frac{\pi}{3.876}$ to minimize the ratio yields factor-.8786 hardness. Both results are optimal, by the Goemans–Williamson algorithm [8].

Using our Theorem 2.5 in place of Borell's Theorem, we can also obtain the $(1 - \delta, 1 - (\frac{2}{\pi} + o(1))\sqrt{\delta})$-hardness for Max-Cut. For the ratio result, we can take $\epsilon = \frac{\pi}{4}$ in place of ϵ_0 and obtain hardness of approximating Max-Cut to any factor exceeding $\frac{3}{2+\sqrt{2}} \approx .8787$.

2.4 Hermite tail bounds

We next describe an application of the subadditivity of rotation sensitivity to the study of Hermite tail bounds. Recall (see, e.g., [17]) that any square-integrable $f : \mathbb{R}^d \to \mathbb{R}$ can be expressed as

$$f(x) = \sum_{\beta \in \mathbb{N}^d} \widehat{f}(\beta) \prod_{i=1}^d h_{\beta_i}(x_i),$$

where $h_j(t)$ denotes the (normalized) Hermite polynomial of degree-j. The numbers $\widehat{f}(\beta)$ are called the Hermite coefficients of f. We often stratify them according to $|\beta| = \sum_{i=1}^d \beta_i$, making the following definition:

Definition 2.8. For square-integrable $f : \mathbb{R}^d \to \mathbb{R}$ and $k \in \mathbb{N} \cup \{0\}$ we define $\mathbf{W}^k[f] = \sum_{|\beta|=k} \widehat{f}(\beta)^2$, the *Hermite weight of f at degree k*, and $\mathbf{W}^{>k}[f] = \sum_{j>k} \mathbf{W}^j[f]$, *the tail weight of f beyond degree k.*

Hermite tail weights are of interest in, e.g., approximation theory and learning theory [15], since

$$\mathbf{W}^{>k}[f] = \min\{\mathbf{E}[(f - p)^2] : p \text{ is a polynomial of degree at most } k\}.$$

In particular, $\mathbf{Var}[f] = \mathbf{W}^{>0}[f]$. We have the following well-known connection between rotation sensitivity and Hermite weights:

$$\mathbf{RS}_f(\epsilon) = \frac{1}{2} \sum_{j=0}^{\infty} (1 - \cos^j(\epsilon)) \cdot \mathbf{W}^j[f]. \tag{3}$$

Using this formula we get a tail bound for Boolean valued functions over Gaussian domains. This is the Gaussian space analogue of Bourgain's theorem, and is a generalization of Theorem 1.3. It quatifies the extent to which indicator functions can be approximated by low-degree polynomials:

Theorem 2.9. *Let $f : \mathbb{R}^d \to \{-1, 1\}$. Then $\mathbf{W}^{>k}[f] \geq \Omega(\mathbf{Var}[f]) \cdot \frac{1}{\sqrt{k}}$ for any $k \in \mathbb{N}$.*

This theorem is sharp up to the constant in the case when f is the indicator of a halfspace through the origin. As mentioned, the previous best lower bound was $\frac{1}{\sqrt{k}} \cdot \frac{1}{2^{\tilde{O}(\sqrt{\log k})}}$, following from the somewhat technical work of Bourgain [6, 13].

4

Proof. Assume $\epsilon \in [0, \frac{\pi}{4}]$ and apply Theorem 1.1 with $\ell = 2$. Using (3) we deduce

$$\sum_{j=0}^{\infty} (1 - \cos^j(\epsilon)) \cdot \mathbf{W}^j[f] \geq \sum_{j=0}^{\infty} \tfrac{1}{2}(1 - \cos^j(2\epsilon)) \cdot \mathbf{W}^j[f]. \tag{4}$$

We consider splitting the sums according to whether $j \leq \frac{1}{4\epsilon^2}$. On the right side of (4) we use

$$\tfrac{1}{2}(1 - \cos^j(2\epsilon)) \geq \tfrac{1}{2}(1 - \exp(-2\epsilon^2)^j) \geq \tfrac{1}{2}(2j\epsilon^2 - 2j^2\epsilon^4) \geq \tfrac{3}{4}j\epsilon^2 \tag{5}$$

when $j \leq \frac{1}{4\epsilon^2}$. On the left side of (4) we may use

$$1 - \cos^j(\epsilon) \leq 1 - (1 - \tfrac{\epsilon^2}{2})^j \leq \tfrac{1}{2}j\epsilon^2. \tag{6}$$

In light of (5) and (6), if we subtract $\sum_{j \leq \frac{1}{4\epsilon^2}} (1 - \cos^j(\epsilon)) \cdot \mathbf{W}^j[f]$ from both sides of (4), the right side shrinks by a multiplicative factor no smaller than $(\tfrac{3}{4} - \tfrac{1}{2})/(\tfrac{3}{4}) = \tfrac{1}{3}$. Hence we obtain

$$\sum_{j > \frac{1}{4\epsilon^2}} (1 - \cos^j(\epsilon)) \cdot \mathbf{W}^j[f] \geq \tfrac{1}{3} \sum_{j=0}^{\infty} \tfrac{1}{2}(1 - \cos^j(2\epsilon)) \cdot \mathbf{W}^j[f] = \tfrac{1}{3}\mathbf{RS}_f(2\epsilon).$$

The left side above is clearly at most $\mathbf{W}^{>4/\epsilon^2}[f]$. By Theorem 2.5 the right side is at least $\frac{2}{3\pi}\mathbf{Var}[f] \cdot \epsilon$ whenever ϵ is of the form $\frac{\pi}{4m}$, $m \in \mathbb{N}$. Thus we have $\mathbf{W}^{>4/\epsilon^2}[f] \geq \frac{2}{3\pi}\mathbf{Var}[f] \cdot \epsilon$ for ϵ of the appropriate form and the proof is completed by suitably choosing $\epsilon = \Theta(\frac{1}{\sqrt{k}})$. $\qquad\square$

3 Boolean sensitivity and tails

In this section we describe how the results on Gaussian sensitivity and tails translate to the setting of Boolean-valued functions on the discrete cube, $f : \{-1, 1\}^n \to \{-1, 1\}$.

3.1 Notation and definitions

Let us recall the necessary definitions. We think of $\{-1, 1\}^n$ as having the uniform probability distribution. We say that $\boldsymbol{x}, \boldsymbol{y}$ are *ρ-correlated random strings*, $\rho \in [-1, 1]$, if the n random pairs $(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \{-1, 1\} \times \{-1, 1\}$ are mutually independent and satisfy $\mathbf{E}[\boldsymbol{x}_i] = \mathbf{E}[\boldsymbol{y}_i] = 0$, $\mathbf{E}[\boldsymbol{x}_i\boldsymbol{y}_i] = \rho$ for each $i \in [n]$. We may still define rotation sensitivity by $\mathbf{RS}_f(\epsilon) = \mathbf{Pr}[f(\boldsymbol{x}) \neq f(\boldsymbol{y})]$, where $\boldsymbol{x}, \boldsymbol{y}$ are $\cos(\epsilon)$-correlated random strings; however in the Boolean setting it is more usual to define the *noise sensitivity of f at $\delta \in [0, 1]$*,

$$\mathbf{NS}_f(\delta) = \mathbf{Pr}[f(\boldsymbol{x}) \neq f(\boldsymbol{y})], \text{where } \boldsymbol{x}, \boldsymbol{y} \text{ are } (1 - 2\delta)\text{-correlated strings}.$$

Equivalently, we may say that $\boldsymbol{x} \sim \{-1, 1\}^n$ is uniformly random and $\boldsymbol{y}$ is formed from $\boldsymbol{x}$ by negating each coordinate independently with probability δ. Since $\cos(2\sqrt{\delta}) = 1 - 2\delta + O(\delta^2)$ as $\delta \to 0$, it follows that $\mathbf{RS}_f(2\sqrt{\delta}) \sim \mathbf{NS}_\delta[f]$.

Any $f : \{-1, 1\}^n \to \mathbb{R}$ can be expressed as $f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \prod_{i \in S} x_i$, where the real numbers $\widehat{f}(S)$ are called the Fourier coefficients of f. As with the Hermite expansion we stratify them according to $|S|$, defining $\mathbf{W}^k[f] = \sum_{|S|=j} \widehat{f}(S)^2$ and $\mathbf{W}^{>k}[f]$ as before. We have the formula analogous to (3),

$$\mathbf{NS}_f(\delta) = \tfrac{1}{2} \sum_{j=0}^{n} (1 - (1 - 2\delta)^j) \cdot \mathbf{W}^j[f]. \tag{7}$$

5

This formula extends the definition of noise sensitivity to functions $f : \{-1,1\}^n \to \mathbb{R}$ which are not necessarily Boolean-valued. We will use two more related quantities. The first is the derivative of the noise sensitivity: $\mathbf{NS}'_f(\delta) = \sum_{j=1}^n (1-2\delta)^{j-1} \cdot \mathbf{W}^j[f]$. The second is the *noise stability* of f at $\rho \in [-1,1]$, $\mathbf{Stab}_\rho[f] = \sum_{j=0}^n \rho^j \cdot \mathbf{W}^j[f]$.

We will require the notion of discrete derivatives; for $i \in [n]$,

$$D_i f(x) = \frac{f(x^{i \to 1}) - f(x^{i \to -1})}{2} = \sum_{S \ni i} \widehat{f}(S) \prod_{j \in S} x_j.$$

Relatedly, we need the notion of low-degree influences and noisy influences:

$$\mathbf{Inf}_i^{\leq k}[f] = \sum_{|S| \leq k,\ S \ni i} \widehat{f}(S)^2, \qquad \mathbf{Inf}_i^{(1-\gamma)}[f] = \sum_{S \ni i} (1-\gamma)^{|S|-1} \widehat{f}(S)^2.$$

We also recall the notion of *restrictions*. For a function $f : \{-1,1\}^n \to \mathbb{R}$, a subset of coordinates $R \subseteq [n]$ and a setting for the coordinates $z \in \{-1,1\}^{\overline{R}}$ (where $\overline{R} = [n] \setminus R$), the restriction (subfunction) $f_{R|z} : \{-1,1\}^R \to \mathbb{R}$ is naturally defined. For a fixed set $R \subseteq [n]$ of coordinates and a uniformly random setting $\boldsymbol{z} \in \{-1,1\}^{\overline{R}}$ of the other coordinates, we define the *variation of f on R* by

$$\mathbf{Vr}_R[f] = \mathbf{E}[\mathbf{Var}[f_{R|\boldsymbol{z}}]] = \sum_{S \cap R \neq \emptyset} \widehat{f}(S)^2.$$

When $\boldsymbol{R}$ is also random, independently containing each coordinate with probability δ, then $(\boldsymbol{R}, \boldsymbol{z})$ is called a δ-*random restriction*. We have the following easy facts concerning random restrictions:

Fact 3.1. *Let $f : \{-1,1\}^n \to \{-1,1\}$ and let $(\boldsymbol{R}, \boldsymbol{z})$ be a δ-random restriction. Then*

$$\mathbf{E}[\mathbf{Var}[f_{\boldsymbol{R}|\boldsymbol{z}}]] = 2\mathbf{NS}_f(\delta/2), \qquad \mathbf{E}[\mathbf{W}^1[f_{\boldsymbol{R}|\boldsymbol{z}}]] = \delta \mathbf{NS}'_f(\tfrac{\delta}{2}).$$

3.2   An optimal result on Fourier tails

The results from Section 2 — e.g., Borell's Theorem, our Hermite tail bound Theorem 2.9 — cannot hold as stated for functions on the discrete cube. This is because of functions like "dictators", $f(x) = x_i$, which have $\mathbf{RS}_f(\epsilon) \sim \frac{1}{4}\epsilon^2 \ll \frac{\epsilon}{\pi}$ (i.e., $\mathbf{NS}_f(\delta) = \delta \ll \sqrt{\delta}$) and $\mathbf{W}^{>1}[f] = 0$. However it is standard to restrict attention to functions which have all of their noisy/low-degree influences smaller than some κ. In this case, one can often use the Invariance Principle technology from [19] to obtain Gaussian-like bounds. For example, [19] established:

Majority Is Stablest Theorem. *Fix $\epsilon \in (0, \frac{\pi}{2})$ and let $f : \{-1,1\}^n \to \{-1,1\}$ satisfy $\mathbf{E}[f] = 0$. Further assume $\mathbf{Inf}_i^{\leq 1/\kappa}[f] \leq C^{-\kappa}$ for all $i \in [n]$, where C is a large universal constant. Then $\mathbf{RS}_f(\epsilon) \geq \frac{\epsilon}{\pi} - \widetilde{O}(\kappa)$.*

For the case of Fourier tails, we can transfer our Hermite tail bound Theorem 2.9 to the Boolean setting very easily using Invariance:

Theorem 3.2. *Fix an integer $k \geq 2$. Let $f : \{-1,1\}^n \to \{-1,1\}$ satisfy $\mathbf{Inf}_i^{\leq k}[f] \leq \kappa = \mathbf{Var}[f]^5 \cdot C^{-k}$ for all $i \in [n]$, where C is a large universal constant. Then $\mathbf{W}^{>k}[f] \geq \Omega(\mathbf{Var}[f]) \cdot \frac{1}{\sqrt{k}}$.*

Proof. Write $g = f^{\leq k}$, so $\mathbf{Inf}_i[g] \leq \kappa$ for all $i \in [n]$. Let $\xi(t) = (t - \mathrm{sgn}(t))^2$, measuring the ℓ_2^2-distance of t from being Boolean-valued. Clearly

$$\mathbf{W}^{>k}[f] = \mathbf{E}[(f-g)^2] \geq \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \{-1,1\}^n} [\xi(g(\boldsymbol{x}))].$$

We claim that

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \{-1,1\}^n} [\xi(g(\boldsymbol{x}))] \quad - \mathop{\mathbf{E}}_{\boldsymbol{X} \sim \mathrm{N}(0,1)^n} [\xi(g(\boldsymbol{X}))] \right| \le 2 \cdot O(10^{k/4} \kappa^{1/4}).$$

This follows immediately from [19, Theorem 3.19, (27)], the Invariance Principle with absolute-value functional, because $\xi(t) = 1 + t^2 - 2|t|$ and because $\mathbf{E}[g(\boldsymbol{x})^2] = \mathbf{E}[g(\boldsymbol{X})^2]$ by Parseval. Hence

$$\mathbf{W}^{>k}[f] \ge \mathop{\mathbf{E}}_{\boldsymbol{X} \sim \mathrm{N}(0,1)^n} [\xi(g(\boldsymbol{X}))] - O(10^{k/4} \kappa^{1/4}) = \mathbf{E}[(h(\boldsymbol{X}) - g(\boldsymbol{X}))^2] - O(10^{k/4} \kappa^{1/4}),$$

where $h : \mathbb{R}^n \to \{-1,1\}$ is defined by $h(X) = \mathrm{sgn}(g(X))$. We have $\mathbf{E}[(h(\boldsymbol{X}) - g(\boldsymbol{X}))^2] \ge \mathbf{W}^{>k}[h]$, the Hermite tail of h above degree k. Hence using Theorem 2.9 we deduce

$$\mathbf{W}^{>k}[f] \ge \Omega(\mathbf{Var}[h]) \cdot \frac{1}{\sqrt{k}} - O(10^{k/4} \kappa^{1/4}). \tag{8}$$

The remainder of the proof is devoted to showing $\mathbf{Var}[h] \ge \Omega(\mathbf{Var}[f])$; once we have this, the proof follows from (8) once C is chosen appropriately.

To compare $\mathbf{Var}[f]$ with $\mathbf{Var}[f]$, first note that $\mathbf{Pr}[f(\boldsymbol{x}) = 1], \mathbf{Pr}[f(\boldsymbol{x}) = -1] \ge \mathbf{Var}[f]/2$. We may further assume $\mathbf{Pr}[g(\boldsymbol{x}) \ge \frac{2}{3}], \mathbf{Pr}[g(\boldsymbol{x}) \le -\frac{2}{3}] \ge \mathbf{Var}[f]/4$, as otherwise $\mathbf{W}^{>k}[f] = \mathbf{E}[(f-g)^2] \ge (\mathbf{Var}[f]/4)(\frac{1}{3})^2 = \mathbf{Var}[f]/36$ and we have the desired conclusion of the theorem. Finally, we apply [19, Theorem 3.19, (28)], the Invariance Principle with Lévy distance. Assuming C is chosen small enough, this lets us deduce $\mathbf{Pr}[g(\boldsymbol{X}) \ge \frac{1}{3}], \mathbf{Pr}[g(\boldsymbol{X}) \le -\frac{1}{3}] \ge \mathbf{Var}[f]/8$. It follows that $\mathbf{Var}[h] \ge \Omega(\mathbf{Var}[f])$, as required. $\qquad \square$

3.3 On Bourgain's proof

In the remainder of the paper we repeat Bourgain's proof of his theorem, but with some variations. We believe these make the proof cleaner; they also improve it quantitatively, achieving the Fourier tail bound $\frac{1}{\sqrt{k}} \cdot \frac{1}{O(\log^{1.5} k)}$. Also this is a polylogarithmic factor away from the optimal, we feel it is still worth presenting. One reason is that it avoids the use of Invariance, and thus could be considered simpler than the proof of Theorem 3.2. Another reason is that the key intermediate result Theorem 1.4 has an elementary proof and could be of independent interest, since it has only polynomial (not exponential) dependence on the "low-influence parameter".

In the proof we present, the main aim is obtaining an analogue of Borell's Theorem; i.e., $\mathbf{NS}_f(\delta) \ge \Omega(\mathbf{Var}[f]) \cdot \sqrt{\delta}$. (In fact, the inapproximability applications of Bourgain's Theorem [11, 14] only need such noise sensitivity bounds, not Fourier tail bounds.) Unfortunately the subadditivity of Gaussian rotation sensitivity — the source of the "square-root gain" for noise sensitivity — does not have an obvious Boolean analogue. However there is one regime in which Boolean and Gaussian domains do have similar isoperimetric properties: that of "small sets", i.e., ± 1-valued functions with small variance. One manifestation of this is a basic lemma in analysis of Boolean functions:

Level 1 Inequality. *Let $f : \{-1,1\}^n \to \{-1,1\}$. Then $\mathbf{W}^1[f] \le \mathbf{Var}[f]^2 \cdot O(\log(2/\mathbf{Var}[f]))$.*

The (nearly) quadratic dependence on $\mathbf{Var}[f]$ here will be the ultimate source of the square-root in our Boolean noise sensitivity and Fourier tail bounds. The Level 1 Inequality has a 5-line proof using nothing more than the Chernoff bound; see, e.g., [22, Proposition 2.2]. One may check that the constant in the $O(\cdot)$ may be taken to be 9. We add on additional remark:

Remark 3.3. The Level 1 Inequality holds for any function $f + c$ where $f : \{-1,1\}^n \to \{-1,1\}$ and c is a constant. This is because neither $\mathbf{W}^1[f]$ nor $\mathbf{Var}[f]$ changes when f is replaced by $f + c$.

We now begin the proof. We will need some slight generalizations of the Level 1 Inequality.

Proposition 3.4. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *and let* $L : \{-1,1\}^n \to \mathbb{R}$ *be linear; i.e.,* $L^{=1} = L$. *Then*

$$\langle f, L \rangle \le \|L\|_2 \cdot \psi(\mathbf{Var}[f]), \quad \text{where } \psi(t) = 3t\sqrt{\log(2/t)}.$$

Proof. We have $\langle f, L \rangle = \langle f^{=1}, L \rangle \le \|f^{=1}\|_2 \cdot \|L\|_2 = \sqrt{\mathbf{W}^1[f]} \cdot \|L\|_2$ by Cauchy–Schwarz. The result now follows by taking the square-root of the Level 1 Inequality. $\square$

Proposition 3.5. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *and let* $R \subseteq [n]$. *Then* $\sum_{j \in R} \widehat{f}(j)^2 \le \psi^2(\mathbf{Vr}_R[f])$.

Proof. Fix $L(x) = \sum_{j \in R} \widehat{f}(j) x_j$. Apply Proposition 3.4 to $f_{R|\boldsymbol{z}}$ and L, where $\boldsymbol{z} \sim \{-1,1\}^{\overline{R}}$ is uniformly random. Taking expectations, we get

$$\mathop{\mathbf{E}}_{\boldsymbol{z}}[\langle f_{R|\boldsymbol{z}}, L \rangle] \le \|L\|_2 \cdot \mathop{\mathbf{E}}_{\boldsymbol{z}}[\psi(\mathbf{Var}[f_{R|\boldsymbol{z}}])] \le \|L\|_2 \cdot \psi\left(\mathop{\mathbf{E}}_{\boldsymbol{z}}[\mathbf{Var}[f_{R|\boldsymbol{z}}]]\right), \tag{9}$$

where we used the fact that ψ is concave. We have $\mathbf{E}_{\boldsymbol{z}}[\langle f_{R|\boldsymbol{z}}, L \rangle] = \|L\|_2^2$ and $\mathbf{E}_{\boldsymbol{z}}[\mathbf{Var}[f_{R|\boldsymbol{z}}]] = \mathbf{Vr}_R[f]$. Thus (9) is equivalent to $\|L\|_2^2 \le \|L\|_2 \cdot \psi(\mathbf{Vr}_R[f])$ and the result follows. $\square$

We remark that the Level-1 Inequality is the special case of Proposition 3.5 when $J = [n]$.

Since $\psi^2 = t \mapsto 9t^2 \log(2/t)$ is increasing on $[0,1]$ it has an inverse function, $(\psi^2)^{-1}$. We introduce the function $\mathscr{Q}$ which is a lower bound on this inverse.

Definition 3.6. Define $\mathscr{Q} : [0,1] \to \mathbb{R}$ by $\mathscr{Q}(t) = \frac{1}{4}\sqrt{t/\log(8/t)}$.

Fact 3.7. *We have* $\mathscr{Q} \le (\psi^2)^{-1}$ *pointwise on* $[0,1]$. *Also,* $\mathscr{Q}$ *is increasing, concave, and satisfies* $\mathscr{Q}(0) = 0$. *In particular,* $\mathscr{Q}$ *is subadditive.*

Fact 3.8. *Let* $s \ge 0$ *and* $t \ge -s$. *Then* $\mathscr{Q}(s+t) \ge \mathscr{Q}(s) - \mathscr{Q}(|t|)$.

Proof. If t is nonnegative this follows trivially from $\mathscr{Q}(s+t) \ge \mathscr{Q}(s)$. Otherwise we have

$$\mathscr{Q}(s+t) = \mathscr{Q}(s - |t|) = \mathscr{Q}(s - |t|) + \mathscr{Q}(|t|) - \mathscr{Q}(|t|) \ge \mathscr{Q}(s) - \mathscr{Q}(|t|)$$

by subadditivity. $\square$

Finally, Proposition 3.5 is equivalent to the following main lemma:

Lemma 3.9. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *and let* $R \subseteq [n]$. *Then* $\mathbf{Vr}_R[f] \ge \mathscr{Q}\left(\sum_{j \in R} \widehat{f}(j)^2\right)$.

3.4 The intermediate theorem

We can now prove (a strengthening of) Theorem 1.4:

Theorem 3.10. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *and let* $\delta \in [0, \frac{1}{2}]$. *Write* $\omega[f] = \sum_{i=1}^{n} \widehat{f}(i)^4$. *Then*

$$\mathbf{NS}_f(\delta) \ge \frac{1}{2}\mathscr{Q}\left(\mathbf{W}^1[f] \cdot \delta\right) - \frac{1}{2}\mathscr{Q}\left(\sqrt{\delta \cdot \omega[f]}\right).$$

Proof. Apply Lemma 3.9 with a 2δ-random subset $\boldsymbol{R} \subseteq [n]$. Taking expectations we get

$$2\mathbf{NS}_f(\delta) \ge \mathop{\mathbf{E}}_{\boldsymbol{R}}\left[\mathscr{Q}\left(\sum_{j \in \boldsymbol{R}} \widehat{f}(i)^2\right)\right] = \mathbf{E}\left[\mathscr{Q}\left(\sum_{i=1}^{n} \widehat{f}(i)^2 \boldsymbol{R}_i\right)\right],$$

where we've written $\boldsymbol{R}_i$ for the 0-1 indicator of "$i \in \boldsymbol{R}$". On the right side above, roughly speaking we have the (1/2)th moment of the random variable $\widehat{f}(1)^2\boldsymbol{R}_1 + \cdots + \widehat{f}(n)^2\boldsymbol{R}_n$; we can estimate this in a straightforward manner. Isolating the mean, $\mathbf{W}^1[f]\cdot\delta$, we get

$$\mathbf{E}\left[\mathcal{Q}\left(\sum_{i=1}^{n}\widehat{f}(i)^2\boldsymbol{R}_i\right)\right] = \mathbf{E}\left[\mathcal{Q}\left(\mathbf{W}^1[f]\cdot\delta + \sum_{i=1}^{n}\widehat{f}(i)^2(\boldsymbol{R}_i - \delta)\right)\right]$$

$$\geq \mathcal{Q}\left(\mathbf{W}^1[f]\cdot\delta\right) - \mathbf{E}\left[\mathcal{Q}\left(\left|\sum_{i=1}^{n}\widehat{f}(i)^2(\boldsymbol{R}_i - \delta)\right|\right)\right] \geq \mathcal{Q}\left(\mathbf{W}^1[f]\cdot\delta\right) - \mathcal{Q}\left(\mathbf{E}\left[\left|\sum_{i=1}^{n}\widehat{f}(i)^2(\boldsymbol{R}_i - \delta)\right|\right]\right)$$

where we used Fact 3.8 and concavity of $\mathcal{Q}$. Recalling that $\mathcal{Q}$ is increasing it remains to upper-bound $\mathbf{E}\left[\left|\sum \widehat{f}(i)^2(\boldsymbol{R}_i - \delta)\right|\right]$. This is easy to do with Cauchy–Schwarz; since the random variables $\widehat{f}(i)^2(\boldsymbol{R}_i - \delta)$ are independent, mean-zero, and have second moment $\widehat{f}(i)^4 \cdot \delta(1-\delta) \leq \widehat{f}(i)^4 \cdot \delta$, we conclude

$$\mathbf{E}\left[\left|\sum_{i=1}^{n}\widehat{f}(i)^2(\boldsymbol{R}_i - \delta)\right|\right] \leq \sqrt{\sum_{i=1}^{n}\widehat{f}(i)^4 \cdot \delta} = \sqrt{\delta \cdot \omega[f]}.$$

This completes the proof. $\qquad\square$

Remark 3.11. Theorem 3.10 is almost sharp; e.g., for the majority function Maj_n we have $\mathbf{NS}_{\mathrm{Maj}_n}(\delta) = \Theta(\sqrt{\delta})$ (when $\delta \gg 1/\sqrt{n}$), $\mathbf{W}^1[\mathrm{Maj}_n] = \Theta(1)$, and $\omega[\mathrm{Maj}_n] = \Theta(1/n)$.

3.5 A stronger noise sensitivity lower bound

We begin with a corollary of Theorem 3.10.

Corollary 3.12. *Let $f : \{-1,1\}^n \to \{-1,1\}$ and let $\delta \in [0,\frac{1}{2}]$, $\gamma \in [0,1]$. Then*

$$\mathbf{NS}_f(\gamma\delta) \geq \tfrac{\gamma}{2}\mathbf{NS}'_f(\tfrac{\gamma}{2})\cdot\mathcal{Q}(\delta) - \delta^{1/4}\cdot\mathop{\mathbf{E}}_{\boldsymbol{R},\boldsymbol{z}}[\omega[f_{\boldsymbol{R}|\boldsymbol{z}}]]^{1/4},$$

where $(\boldsymbol{R},\boldsymbol{z})$ is a γ-random restriction.

Proof. We first weaken the statement of Theorem 3.10 by using $\mathcal{Q}\left(\mathbf{W}^1[f]\cdot\delta\right) \geq \mathbf{W}^1[f]\cdot\mathcal{Q}(\delta)$ and $\frac{1}{2}\mathcal{Q}\left(\sqrt{\delta\cdot\omega[f]}\right) \leq \delta^{1/4}\cdot\omega[f]^{1/4}$. (The former holds because $\mathcal{Q}$ is concave and $\mathcal{Q}(0) = 0$.) Now applying this weakening to $f_{\boldsymbol{R}|\boldsymbol{z}}$ and taking expectations yields

$$\mathbf{NS}_f(\gamma\delta) \geq \tfrac{1}{2}\mathbf{E}[\mathbf{W}^1[f_{\boldsymbol{R}|\boldsymbol{z}}]]\cdot\mathcal{Q}(\delta) - \delta^{1/4}\cdot\mathop{\mathbf{E}}_{\boldsymbol{R},\boldsymbol{z}}[\omega[f_{\boldsymbol{R}|\boldsymbol{z}}]^{1/4}].$$

The proof is completed using the concavity of $t \mapsto t^{1/4}$ and Fact 3.1. $\qquad\square$

We now want to estimate $\mathbf{E}[\omega[f_{\boldsymbol{R}|\boldsymbol{z}}]]$; this is slightly technical and requires hypercontractivity.

Proposition 3.13. *Fix $\gamma \in [0,1]$ and let $\boldsymbol{w}_0$ be the random variable which is 0 with probability γ and ± 1 with probability $(1-\gamma)/2$ each. Then $\boldsymbol{w}_0$ is $(2, 2 + c\gamma, (1-\gamma)^{1/4})$-hypercontractive for some universal $c > 0$. In fact, $c = 1/2$ is acceptable.*

Proof. Since $\boldsymbol{w}_0$ is a symmetric random variable, a simple reduction to the hypercontractivity of standard Rademachers ([4, Lemma 3.3] or [16, Lemma 4.1]) shows that $\boldsymbol{w}_0$ is $(2, q, \frac{1}{\sqrt{q-1}}\frac{\|\boldsymbol{w}_0\|_2}{\|\boldsymbol{w}_0\|_q})$-hypercontractive for any $q \geq 2$. The result follows because

$$\frac{1}{\sqrt{1+c\gamma}}\frac{\|\boldsymbol{w}_0\|_2}{\|\boldsymbol{w}_0\|_{2+c\gamma}} = \frac{1}{\sqrt{1+c\gamma}}(1-\gamma)^{\frac{1}{2}-\frac{1}{2+c\gamma}} \geq (1-\gamma)^{1/4}$$

for $c > 0$ sufficiently small. In fact, the inequality holds for $c = 1/2$. $\qquad\square$

9

Lemma 3.14. *Let $g : \{-1,1\}^n \to \mathbb{R}$ and let $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ be independent copies of $\boldsymbol{w}_0$ from Proposition 3.13. Then*

$$\mathbf{E}[|g(\boldsymbol{w})|^{2+\gamma/2}] \le \mathbf{Stab}_{\sqrt{1-\gamma}}[g]^{1+\gamma/4}.$$

(Regarding the meaning of $g(\boldsymbol{w})$, we identify g with its real multilinear expansion.)

Proof. Define $h = T_{(1-\gamma)^{-1/4}}g$; then

$$\mathbf{E}[|g(\boldsymbol{w})|^{2+\gamma/2}] = \mathbf{E}[|T_{(1-\gamma)^{1/4}}h(\boldsymbol{w})|^{2+\gamma/2}] = \|T_{(1-\gamma)^{1/4}}h(\boldsymbol{w})\|_{2+\gamma/2}^{2+\gamma/2} \le \|h(\boldsymbol{w})\|_2^{2+\gamma/2} = \mathbf{E}[h(\boldsymbol{w})^2]^{1+\gamma/4},$$

where the inequality uses the hypercontractivity result Proposition 3.13. But since the $\boldsymbol{w}_i$'s are independent, mean-zero, and satisfy $\mathbf{E}[\boldsymbol{w}_i^2] = 1-\gamma$, it's easy to compute

$$\mathbf{E}[h(\boldsymbol{w})^2] = \sum_{S \subseteq [n]} \widehat{h}(S)^2 (1-\gamma)^{|S|} = \sum_{S \subseteq [n]} \widehat{g}(S)^2 \sqrt{1-\gamma}^{|S|} = \mathbf{Stab}_{\sqrt{1-\gamma}}[g],$$

as needed. $\qquad\square$

Corollary 3.15. *Let $f : \{-1,1\}^n \to \{-1,1\}$ and let $(\boldsymbol{R}, \boldsymbol{z})$ be a γ-random restriction. Then $\mathbf{E}[\widehat{f_{\boldsymbol{R}|\boldsymbol{z}}}(i)^4] \le \gamma \mathbf{Inf}_i^{(\sqrt{1-\gamma})}[f]^{1+\gamma/4}$ for any $i \in [n]$.*

Proof. Clearly $\widehat{f_{\boldsymbol{R}|\boldsymbol{z}}}(i) = 0$ unless $i \in \boldsymbol{R}$, an event that happens with probability γ. Conditioned on this event, $\widehat{f_{\boldsymbol{R}|\boldsymbol{z}}}(i)$ is distributed as $D_i f(\boldsymbol{w})$, where $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ is as in Lemma 3.14. Thus

$$\mathbf{E}[\widehat{f_{\boldsymbol{R}|\boldsymbol{z}}}(i)^4] = \gamma \mathbf{E}[D_i f(\boldsymbol{w})^4] \le \gamma \mathbf{E}[|D_i f(\boldsymbol{w})|^{2+\gamma/2}] \le \gamma \mathbf{Inf}_i^{(\sqrt{1-\gamma})}[f]^{1+\gamma/4},$$

where the first inequality is because $|D_i f| \le 1$ and the second inequality is Lemma 3.14. $\qquad\square$

We can now combine our results to obtain:

Theorem 3.16. *Let $f : \{-1,1\}^n \to \{-1,1\}$ and let $\delta \in [0, \frac{1}{2}]$, $\gamma \in [0,1]$. Assume $\mathbf{Inf}_i^{(\sqrt{1-\gamma})}[f] \le \kappa$ for all $i \in [n]$. Then*

$$\mathbf{NS}_f(\gamma\delta) \ge \frac{\gamma}{2}\mathbf{NS}_f'(\frac{\gamma}{2}) \cdot \mathscr{Q}(\delta) - \delta^{1/4} \cdot \kappa^{\gamma/16}.$$

Proof. Given Corollary 3.12 we only need to prove that $\mathbf{E}_{\boldsymbol{R},\boldsymbol{z}}[\omega[f_{\boldsymbol{R}|\boldsymbol{z}}]] \le \kappa^{\gamma/4}$ for $(\boldsymbol{R}, \boldsymbol{z})$ a γ-random restriction. From Corollary 3.15,

$$\mathop{\mathbf{E}}_{\boldsymbol{R},\boldsymbol{z}}[\omega[f_{\boldsymbol{R}|\boldsymbol{z}}]] \le \sum_{i=1}^n \gamma \mathbf{Inf}_i^{(\sqrt{1-\gamma})}[f]^{1+\gamma/4} \le \gamma \mathbf{I}^{(\sqrt{1-\gamma})}[f] \cdot \kappa^{\gamma/4},$$

and the result follows from $\gamma \mathbf{I}^{(\sqrt{1-\gamma})}[f] \le 1$. $\qquad\square$

This result will be our Boolean replacement for Theorem 1.1, the subadditivity of rotation sensitivity. If we had $\mathbf{NS}_f(\frac{\gamma}{2})$ in place of $\frac{\gamma}{2}\mathbf{NS}_f'(\frac{\gamma}{2})$ then it would be essentially the same result — ignoring the term involving κ, the only difference would be an $O(\sqrt{\log(1/\delta)})$ factor. However Theorem 3.16 is strictly weaker because $\mathbf{NS}_f(\frac{\gamma}{2}) \ge \frac{\gamma}{2}\mathbf{NS}_f'(\frac{\gamma}{2})$ by the concavity of $\mathbf{NS}$. But as we'll see, the difference between the two quantities is comparable to $\mathbf{W}^{>1/\gamma}[f]$. So if this difference is large then we obtain a tail lower-bound anyway.

3.6 Completing the noise sensitivity and Fourier tail bounds

Towards the proof of the Fourier tail bound, we begin by establishing an analogue of Borell's Theorem in the Boolean setting:

Theorem 3.17. *Let $f : \{-1,1\}^n \to \{-1,1\}$ and let $\delta \in [0, \frac{1}{2}]$. Assume $\mathbf{Inf}_i^{(\sqrt{1-\delta})}[f] \le \kappa$ for all $i \in [n]$. Then*

$$\mathbf{NS}_f(\delta) \ge \frac{\sqrt{\delta}}{10 \log^{1.5}(8/\delta)} \cdot \mathbf{Var}[f] - \delta^{1/4} \kappa^{\delta/16}.$$

Proof. Consider the quantities $\mathbf{NS}_f(\frac{1}{2}) - \mathbf{NS}_f(\frac{1}{4})$, $\mathbf{NS}_f(\frac{1}{4}) - \mathbf{NS}_f(\frac{1}{8})$, ..., $\mathbf{NS}_f(2^{-r}) - \mathbf{NS}_f(2^{-(r+1)})$, where $r = \lfloor \log(1/\delta) \rfloor$. If any of these, say the jth, is at least $\frac{1}{2.5 \log(8/\delta)} \mathbf{Var}[f]$ then we apply Theorem 3.16 with $\gamma = 2^{-j}$ to deduce

$$\mathbf{NS}_f(2^{-j}\delta) \ge 2^{-(j+1)} \mathbf{NS}_f'(2^{-(j+1)}) \cdot \mathcal{Q}(\delta) - \delta^{1/4} \cdot \kappa^{\delta/16}.$$

(Here we used $\gamma \ge \delta$.) But concavity of $\mathbf{NS}$ implies $2^{-(j+1)} \mathbf{NS}_f'(2^{-(j+1)}) \ge \mathbf{NS}_f(2^{-j}) - \mathbf{NS}_f(2^{-(j+1)}) \ge \frac{1}{2.5 \log(8/\delta)} \mathbf{Var}[f]$ by assumption; the claimed result thus follows.

On the other hand, if no such j exists then we conclude

$$\mathbf{NS}_f(\tfrac{1}{2}) - \mathbf{NS}_f(2^{-(r+1)}) \le \frac{r}{2.5 \log(8/\delta)} \mathbf{Var}[f]$$
$$\Rightarrow \tfrac{1}{2} \mathbf{Var}[f] - \mathbf{NS}_f(\delta) \le .4 \mathbf{Var}[f]$$
$$\Rightarrow \mathbf{NS}_f(\delta) \ge .1 \mathbf{Var}[f],$$

which is stronger than what we need. $\qquad\square$

As mentioned earlier, the difference between $\frac{\gamma}{2} \mathbf{NS}_f'(\frac{\gamma}{2})$ and $\mathbf{NS}_f(\frac{\gamma}{2})$ is comparable to $\mathbf{W}^{>1/\gamma}[f]$:

Lemma 3.18. *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $\gamma \in [0,1]$. Then $\frac{\gamma}{2} \mathbf{NS}_{\frac{\gamma}{2}}'(f) \ge \frac{1}{2e} \mathbf{NS}_f(\frac{\gamma}{2}) - \frac{1}{4e} \mathbf{W}^{>1/\gamma}[f]$.*

Proof. We have

$$\frac{1}{2e} \mathbf{NS}_{\frac{\gamma}{2}}(f) - \frac{1}{4e} \mathbf{W}^{>1/\gamma}[f] = \frac{1}{2e} \left(\frac{1}{2} \sum_{j=1}^n (1 - (1-\gamma)^j) \mathbf{W}^j[f] - \frac{1}{2} \sum_{j > 1/\gamma} \mathbf{W}^j[f] \right)$$
$$\le \frac{1}{4e} \sum_{1 \le j \le 1/\gamma} (1 - (1-\gamma)^j) \mathbf{W}^j[f]$$
$$\le \frac{\gamma}{4e} \sum_{1 \le j \le 1/\gamma} j \mathbf{W}^j[f]$$
$$\le \frac{\gamma}{4} \sum_{1 \le j \le 1/\gamma} j (1-\gamma)^{j-1} \mathbf{W}^j[f]$$
$$\le \frac{\gamma}{4} \sum_{j=1}^n j (1-\gamma)^{j-1} \mathbf{W}^j[f] = \frac{\gamma}{2} \mathbf{NS}_f'(\frac{\gamma}{2}). \qquad\square$$

We can now establish our improved version of Bourgain's Theorem.

Theorem 3.19. *There is a universal constant C such that the following holds. If $k \ge 2$, $f : \{-1,1\}^n \to \{-1,1\}$, and $\mathbf{Inf}_i^{(1-\frac{1}{Ck})}[f] \le \kappa \le \frac{1}{2}$ for all $i \in [n]$, then*

$$\mathbf{W}^{>k}[f] \ge \frac{1}{\sqrt{k}} \cdot \frac{1}{C \log^{1.5} k} \cdot \mathbf{Var}[f] - \kappa^{1/(Ck)}.$$

Proof. The proof is very similar to that of Theorem 2.9. Since $\mathcal{Q}(\delta) = \widetilde{\Theta}(\sqrt{\delta})$ we may select some absolute constant $0 < \delta_0 < \frac{1}{2}$ such that $\delta_0 \le \frac{1}{2} \cdot \frac{1-e^{-1}}{4e} \cdot \mathcal{Q}(\delta_0)$. Let us apply Theorem 3.16 with $\delta = \delta_0$ and $\gamma = \frac{1}{k}$. Using also Lemma 3.18 we get

$$\mathbf{NS}_f(\tfrac{\delta_0}{k}) \ge \left(\tfrac{1}{2e}\mathbf{NS}_f(\tfrac{1}{2k}) - \tfrac{1}{4e}\mathbf{W}^{>k}[f] \right) \cdot \mathcal{Q}(\delta_0) - \delta_0^{1/4}\kappa^{1/(16k)},$$

and hence

$$
\begin{aligned}
\tfrac{1}{2e} \cdot \mathcal{Q}(\delta_0) \cdot \mathbf{NS}_f(\tfrac{1}{2k}) &\le \mathbf{NS}_f(\tfrac{\delta_0}{k}) + \tfrac{\mathcal{Q}(\delta_0)}{4e} \cdot \mathbf{W}^{>k}[f] + O(\kappa^{1/(16k)}) \\
&\le \tfrac{1}{2}\sum_{1 \le j \le k}(1 - (1 - 2\tfrac{\delta_0}{k})^j) \cdot \mathbf{W}^j[f] + (\tfrac{1}{2} + \tfrac{\mathcal{Q}(\delta_0)}{4e})\mathbf{W}^{>k}[f] + O(\kappa^{1/(16k)}) \\
&\le \tfrac{\delta_0}{k}\sum_{1 \le j \le k} j \cdot \mathbf{W}^j[f] + \mathbf{W}^{>k}[f] + O(\kappa^{1/(16k)}) \quad\quad\quad\quad (10)
\end{aligned}
$$

Observe that

$$
\begin{aligned}
\tfrac{1}{2e} \cdot \mathcal{Q}(\delta_0) \cdot \mathbf{NS}_f(\tfrac{1}{2k}) &= \tfrac{1}{2e} \cdot \mathcal{Q}(\delta_0) \cdot \left(\tfrac{1}{2}\sum_{1 \le j \le k}(1 - (1 - \tfrac{1}{k})^j) \cdot \mathbf{W}^j[f] + \tfrac{1}{2}\sum_{j > k}(1 - (1 - \tfrac{1}{k})^j) \cdot \mathbf{W}^j[f] \right), \\
&\ge \tfrac{1}{2e} \cdot \mathcal{Q}(\delta_0) \cdot \left(\tfrac{1-e^{-1}}{2k}\sum_{1 \le j \le k} j \cdot \mathbf{W}^j[f] + \tfrac{1}{2}\sum_{j > k}(1 - (1 - \tfrac{1}{k})^j) \cdot \mathbf{W}^j[f] \right).
\end{aligned}
$$

Thus if we subtract $\frac{\delta_0}{k}\sum_{1 \le j \le 1/\gamma} j \cdot \mathbf{W}^j[f]$ from both sides of (10), the left side shrinks by a factor of at most $\frac{1}{2}$, by our choice of δ_0. We therefore conclude

$$\Omega(\mathbf{NS}_f(\tfrac{1}{2k})) \le \mathbf{W}^{>k}[f] + O(\kappa^{1/(16k)}).$$

We now apply Theorem 3.17 with $\delta = \frac{1}{2k}$, deducing

$$\frac{1}{\sqrt{k}} \cdot \frac{1}{O(\log^{1.5}k)} \cdot \mathbf{Var}[f] - (\tfrac{1}{2k})^{1/4}\kappa^{1/(32k)} \le \mathbf{W}^{>k}[f] + O(\kappa^{1/(16k)}).$$

The result follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.7 The junta conclusion

Theorem 3.19 shows that if $f : \{-1, 1\}^n \to \{-1, 1\}$ has $\mathbf{W}^{>k}[f] \ll \frac{1}{\sqrt{k}}$ then f must have at least one coordinate with large noisy-influence. As in Bourgain's paper, it is not hard to conclude that in fact f must be close to a $k^{O(k)}$-junta (i.e., a function depending on at most $k^{O(k)}$ coordinates). Herein we establish that stronger deduction.

In Theorem 3.19 the hypothesis is that all coordinates $i \in [n]$ satisfy $\mathbf{Inf}_i^{(1-\frac{1}{Ck})}[f] \le \kappa$. We can adapt the theorem to the case when this holds only for some of the coordinates:

Theorem 3.20. *Let C be the constant from Theorem 3.19. If $k \ge 2$, $f : \{-1, 1\}^n \to \{-1, 1\}$, $J \subseteq [n]$, and $\mathbf{Inf}_i^{(1-\frac{1}{Ck})}[f] \le \kappa \le \frac{1}{2}$ for all $i \in \overline{J}$, then*

$$\mathbf{W}^{>k}[f] \ge \frac{1}{\sqrt{k}} \cdot \frac{1}{C\log^{1.5}k} \cdot \mathbf{Vr}_{\overline{J}}[f] - \kappa^{1/(Ck)}.$$

Before explaining how to obtain Theorem 3.20, we use it to deduce the desired conclusion:

Corollary 3.21. *Let C be the constant from Theorem [3.20](). Suppose $f : \{-1,1\}^n \to \{-1,1\}$, $k \geq 2$, $\epsilon \in (0,1)$, and*

$$\mathbf{W}^{>k}[f] \leq \frac{\epsilon}{2C\sqrt{k} \cdot \log^{1.5} k}.$$

Then f is ϵ-close to some r-junta, where $r = (k/\epsilon)^{O(k)}$.

Proof. Set $\kappa = (\frac{\epsilon}{2C\sqrt{k} \cdot \log^{1.5} k})^{Ck}$, and let $J = \{i \in [n] : \mathbf{Inf}_i^{(1-\frac{1}{Ck})}[f] \leq \kappa\}$. A well known and simple calculation shows that $|J| \leq \frac{Ck}{\kappa} = r$. Now let $g = \mathrm{Avg}_{\overline{J}} f = \sum_{S \subseteq J} \widehat{f}(S) \chi_S$, a real-valued function depending only on the coordinates in J. If we can show that $\mathbf{E}[(f-g)^2] \leq \epsilon$, it follows that f is ϵ-close to the Boolean-valued r-junta $\mathrm{sgn}(g)$. But by Theorem [3.20](),

$$\mathbf{E}[(f-g)^2] = \mathbf{Vr}_{\overline{J}}[f] \leq (C\sqrt{k}\log^{1.5} k) \cdot (\mathbf{W}^{>k}[f] + \kappa^{1/(Ck)}) \leq \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2},$$

as required. $\qquad\square$

We now sketch how to obtain the adaptation Theorem [3.20]() of Theorem [3.19](). The key idea is that given the hypotheses of Theorem [3.20](), we define $g = f - \mathrm{Avg}_{\overline{J}} f$ and then prove Theorem [3.19]() for g. Note that g is not a Boolean-valued function; nevertheless, the proof of Theorem [3.19]() can be carried out. We now shortly explain how this is done, beginning with the crucial Proposition [3.5]().

Proposition 3.5. Notice that it suffices to prove Proposition [3.5]() for g assuming $R \subseteq \overline{J}$; this is because $\widehat{g}(i) = 0$ for $i \in J$. Now Proposition [3.5]() relies on Proposition [3.4](), which is stated only for Boolean-valued f. However, just as in Remark [3.3](), Proposition [3.4]() also holds for "shifted" Boolean-valued functions; i.e., those with range $\{-1+c, 1+c\}$ for some real c. Since $R \subseteq \overline{J}$, each restriction employed in the proof of Proposition [3.5]() fixes all of the coordinates J. And for each such restriction, our $g = f - \mathrm{Avg}_{\overline{J}} f$ becomes a "shifted" function, since $\mathrm{Avg}_{\overline{J}} f$ only depends on the coordinates in J. Thus we may use Proposition [3.4]() and the proof goes through.

Completing the proof. After Proposition [3.5](), all subsequent steps of the proof work for any $g : \{-1,1\}^n \to \mathbb{R}$ with $\|g\|_2^2 \leq 1$. In the end, Theorem [3.20]() immediately follows from applying Theorem [3.19]() to g, because $\mathbf{Var}[g] = \mathbf{Vr}_{\overline{J}}[f]$ and $\mathbf{W}^{>k}[f] \geq \mathbf{W}^{>k}[g]$.

References

[1] William Beckner. Sobolev inequalities, the Poisson semigroup, and analysis on the sphere s^n. *Proceedings of the National Academy of Sciences*, 89(11):4816–4819, 1992. [2.1]()

[2] Sergey Bobkov and Christian Houdré. *Some Connections between Isoperimetric and Sobolev-type Inequalities*. Memoirs of the American Mathematical Society, 1997. [A]()

[3] Christer Borell. The Brunn–Minkowski inequality in Gauss space. *Inventiones Mathematicae*, 30(2):207–216, 1975. [2.3]()

[4] Christer Borell. On polynomial chaos and integrability. *Probabability and Mathematical Statistics*, 3(2):191–203, 1984. [3.5]()

[5] Christer Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Probability Theory and Related Fields*, 70(1):1–13, 1985. [1](), [2.1]()

[6] Jean Bourgain. On the distribution of the Fourier spectrum of Boolean functions. *Israel Journal of Mathematics*, 131(1):269–276, 2002. 1, 2.4

[7] Eric Carlen and Michael Loss. Extremals of functionals with competing symmetries. *Journal of Functional Analysis*, 88(2):437–456, 1990. 2.1

[8] Michel Goemans and David Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995. 2.3

[9] Daniel Kane. The Gaussian surface area and noise sensitivity of degree-d polynomial threshold functions. *Computational Complexity*, 20(2):389–412, 2011. 1

[10] Daniel Kane. *On Elliptic Curves, the ABC Conjecture, and Polynomial Threshold Functions*. PhD thesis, Harvard University, 2011. A

[11] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symposium on Theory of Computing*, pages 767–775, 2002. 3.3

[12] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for Max-Cut and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007. 1, 2.3, 2.7

[13] Subhash Khot and Assaf Naor. Nonembeddability theorems via Fourier analysis. *Mathematische Annalen*, 334(4):821–852, 2006. 1, 2.4

[14] Subhash Khot and Nisheeth Vishnoi. The Unique Games Conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 53–62, 2005. 3.3

[15] Adam Klivans, Ryan O'Donnell, and Rocco Servedio. Learning geometric concepts via Gaussian surface area. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 541–550, 2008. 2.4

[16] Wieslaw Krakowiak and Jerzy Szulga. Hypercontraction principle and random multilinear forms. *Probability Theory and Related Fields*, 77(3):325–342, 1988. 3.5

[17] Michel Ledoux. Isoperimetry and Gaussian analysis. In Pierre Bernard, editor, *Lectures on Probability Theory and Statistics*, volume XXIV of *Lecture Notes in Mathematics 1648*, pages 165–294. Springer, 1996. 2.3, 2.4, A

[18] Michel Ledoux. Personal communication, 2006. A

[19] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171(1), 2010. 1, 1, 2.3, 2.7, 3.2, 3.2, 3.2

[20] William Sheppard. On the application of the theory of error to cases of normal distribution and normal correlation. *Philosophical Transactions of the Royal Society of London, Series A*, 192:101–167, 531, 1899. 2.1

[21] Vladimir Sudakov and Boris Tsirel'son. Extremal properties of half-spaces for spherically invariant measures. *Journal of Soviet Mathematics*, 9(1):9–18, 1978. Originally published in *Zap. Nauchn. Sem. Leningrad. Otdel. Math. Inst. Steklova.*, 41:14–21, 1974. 2.3

[22] Michel Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996. 3.3

A Definitions of Gaussian surface area

Here we make some comments on the definition of the Gaussian surface area of a set $A \subseteq \mathbb{R}^d$. Recall the two definitions we have discussed so far, $\gamma^+(A)$ from (1) and surf(A) from (2). There is also a third natural definition, namely the integral of the Gaussian density over the boundary of A. All of these definitions "should" coincide for "nice enough" sets A. In particular, let us explain heuristically why surf(A) = $\gamma^+(A)$ holds when, say, A has smooth boundary. (A similar explanation is given in [10, Chap. G.1.2].)

For $\cos(\epsilon)$-correlated (X, Y) we can think of $Y = \cos(\epsilon)X + \sin(\epsilon)X'$, where X' is an independent standard d-dimensional Gaussian. From the definition (1) we should have $\mathbf{Pr}[t \leq \text{dist}(X, \partial A) < t + dt] \approx 2\gamma^+(A)dt$. Conditioned on this event, we are concerned with the probability that

$$Y = \cos(\epsilon)X + \sin(\epsilon)X' \approx (1 - \tfrac{\epsilon^2}{2})X + \epsilon X'$$

lands on the "other side" of ∂A. For small ϵ we neglect the ϵ^2 and think of ∂A as being locally flat near X. Thus we would like the probability that $X + \epsilon X'$ lands on the other side of a halfspace at distance roughly t from X. Projecting X' onto the line through X perpendicular to the halfspace, we see that the probability is $\mathbf{Pr}[\epsilon N(0, 1) > t]$. Thus we conclude that

$$\mathbf{RS}_f(\epsilon) = \mathbf{Pr}[f(X) \neq f(Y)] \approx \int_0^\infty \mathbf{Pr}[\epsilon N(0, 1) > t] \cdot 2\gamma^+(A)dt = \gamma^+(A)\mathbf{E}[|\epsilon N(0, 1)|] = \sqrt{\tfrac{2}{\pi}}\gamma^+(A)\epsilon.$$

This heuristically justifies the claim that surf(A) = $\gamma^+(A)$ for "nice" sets.

Even though this heuristic description is not rigorous, Ledoux [17, Proposition 8.5] has shown:

Theorem A.1. *(Ledoux.) For any $A \subseteq \mathbb{R}^d$ with smooth boundary and any $\epsilon \in (0, \frac{\pi}{2})$ it holds that* $\sqrt{\tfrac{\pi}{2}} \cdot \frac{\mathbf{RS}_f(\epsilon)}{\epsilon} \leq \gamma^+(A)$. *Hence* surf($A$) $\leq \gamma^+(A)$.

Further, this inequality also extends [18] to the case of all Borel sets A using the techniques in [2, Theorem 1.1]. Thus our proof of the GII for volume-$\frac{1}{2}$ sets (using definition surf(A)) implies the standard statement of the GII.

A final technical note: using the fact that $\mathbf{RS}_f(\epsilon)$ is increasing for $\epsilon \in [0, \frac{\pi}{2}]$, it is easy to show that the lim sup in our definition of surf(A) may be equivalently replaced by a lim. We omit the details.