Analysis of Boolean Functions

Notes from a series of lectures by

Ryan O'Donnell

Barbados Workshop on Computational Complexity

February 26th – March 4th, 2012

Scribe notes by Li-Yang Tan

Contents

1 Linearity testing and Arrow's theorem

Monday, 27th February 2012

- Open Problem [Guy86, HK92]: Let $a \in \mathbb{R}^n$ with $\|a\|_2 = 1$. Prove $\mathbf{Pr}_{x \in \{-1,1\}^n}[|\langle a, x\rangle| \leq 1] \geq 1/2$.
- Open Problem (S. Srinivasan): Suppose $g : \{-1,1\}^n \to \pm[2/3, 1]$ where $g(x) \in [2/3, 1]$ if $\sum_{i=1}^n x_i \geq n/2$ and $g(x) \in [-1, -2/3]$ if $\sum_{i=1}^n x_i \leq -n/2$. Prove $\deg(f) = \Omega(n)$.

In this workshop we will study the analysis of boolean functions and its applications to topics such as property testing, voting, pseudorandomness, Gaussian geometry and the hardness of approximation. Two recurring themes that we will see throughout the week are:

- The noisy hypercube graph is a small set expander.
- Every boolean function has a "junta part" and a "Gaussian part".

1.1 The Fourier expansion

Broadly speaking, the analysis of boolean functions is concerned with properties of boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ viewed as multilinear polynomials over $\mathbb{R}$. Consider the majority function over 3 variables $\mathsf{MAJ}_3(x) = \mathrm{sgn}(x_1 + x_2 + x_3)$. It is easy to check that $\mathsf{MAJ}_3(x) = \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 - \frac{1}{2}x_1x_2x_3$, and this can be derived by summing $2^3 = 8$ polynomials $p_y : \{-1,1\}^3 \to \{-1,0,1\}$, one for each $y \in \{-1,1\}^3$, where $p_y(x)$ takes value $\mathsf{MAJ}_3(x)$ when $y = x$ and 0 otherwise. For example,

$$p_{(-1,1,-1)}(x) = \left(\frac{1-x_1}{2}\right)\left(\frac{1+x_2}{2}\right)\left(\frac{1-x_3}{2}\right) \cdot \mathsf{MAJ}_3(-1,1,-1).$$

Note that the final polynomial that results from expanding and simplifying the sum of p_y's is indeed always multilinear (*i.e.* no variable x_i occurs squared, or cubed, *etc.*) since $x_i^2 = 1$ for bits $x_i \in \{-1,1\}$. The same interpolation procedure can be carried out for any $f : \{-1,1\}^n \to \mathbb{R}$:

Theorem 1 (Fourier expansion) *Every $f : \{-1,1\}^n \to \mathbb{R}$ can be uniquely expressed as a multilinear polynomial $\mathbb{R}$,*

$$f(x) = \sum_{S \subseteq [n]} c_S \prod_{i \in S} x_i, \quad \textit{where each } c_S \in \mathbb{R}.$$

We will write $\hat{f}(S)$ to denote the coefficient c_S and $\chi_S(x)$ for the function $\prod_{i\in S} x_i$, and call $f(x) = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S(x)$ the Fourier expansion of f. We adopt the convention that $\chi_\emptyset \equiv 1$, the identically 1 function. We will write $\deg(f)$ to denote $\max_{S\subseteq[n]}\{|S| : \hat{f}(S) \neq 0\}$, and call this quantity the Fourier degree of f.

We will sometimes refer to $\chi_S(x) : \{-1,1\}^n \to \{-1,1\}$ as the "parity-on-S" function, since it takes value 1 if there are an even number of -1 coordinates in x and -1 otherwise. Using the notation of Theorem 1, we have that $\widehat{\mathsf{MAJ}_3}(\{1\}) = \frac{1}{2}$, $\widehat{\mathsf{MAJ}_3}(\{1,2,3\}) = -\frac{1}{2}$, $\widehat{\mathsf{MAJ}_3}(\{1,2\}) = 0$, and $\deg(\mathsf{MAJ}_3) = 3$.

We have already seen that every function $f : \{-1,1\}^n \to \mathbb{R}$ can be expressed as a multilinear polynomial over $\mathbb{R}$ (via the interpolation procedure described for MAJ_3); to complete the proof of Theorem 1 it remains to show uniqueness. Let V be the vector space of all functions $f : \{-1,1\}^n \to \mathbb{R}$. Here we are viewing f as a 2^n-dimensional vector in $\mathbb{R}^{2^n}$, with each coordinate being the value of f on some input $x \in \{-1,1\}^n$; if f is a boolean function this is simply the truth table of f. Note that the parity functions $\chi_S(x)$ are all elements of V, and furthermore every $f \in V$ can be expressed as a linear combination of them (*i.e.* the $\chi_S(x)$'s are a spanning set for V). Since there are 2^n parity functions and $\dim(V) = 2^n$, it follows that $\{\chi_S : S \subseteq [n]\}$ is a basis for V, and this establishes the uniqueness of the Fourier expansion.

Definition 2 (inner product) *Let $f, g : \{-1,1\}^n \to \{-1,1\}$. We define the inner product between f and g as*

$$\langle f, g \rangle := \sum_{x \in \{-1,1\}^n} \frac{f(x) \cdot g(x)}{2^n} = \mathop{\mathbf{E}}_{x \in \{-1,1\}^n}[f(x)g(x)].$$

Note that this is simply the dot product between f and g viewed as vectors in $\mathbb{R}^{2^n}$, normalized by a factor of 2^{-n}. Given this definition, every boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is a unit vector in $\mathbb{R}^{2^n}$ since $\langle f, f \rangle = 1$. We will also write $\|f\|_2^2$ to denote $\langle f, f \rangle$, and more generally, $\|f\|_p := \mathbf{E}[|f(x)|^p]^{1/p}$.

Theorem 3 (orthonormality) *The set of parity functions $\{\chi_S(x) : S \subseteq [n]\}$ is an orthonormal basis for $\mathbb{R}^{2^n}$. That is, for every $S, T \subseteq [n]$,*

$$\langle \chi_S, \chi_T \rangle = \begin{cases} 1 & \text{if } S = T \\ 0 & \text{otherwise.} \end{cases}$$

Proof. First note that $\chi_S \cdot \chi_T = \chi_{S\Delta T}$ since $\prod_{i \in S} x_i \prod_{j \in T} x_j = \prod_{i \in S\Delta T} x_i \prod_{j \in S \cap T} x_j^2 = \prod_{i \in S\Delta T} x_i$, where the final equality uses the fact that $x_i^2 = 1$ for $x_i \in \{-1,1\}$. Next, we claim that

$$\mathbf{E}[\chi_U] = \begin{cases} 1 & \text{if } U = \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

noting that this implies the theorem since $S\Delta T = \emptyset$ iff $S = T$. Recall that we have defined $\chi_\emptyset$ to be the identically 1 function, and if $U \neq \emptyset$ then exactly half the inputs $x \in \{-1,1\}^n$ have $\chi_U(x) = 1$ and the other half $\chi_U(x) = -1$. ■

Proposition 4 (Fourier coefficient) *Let $f : \{-1,1\}^n \to \mathbb{R}$. Then $\hat{f}(S) = \langle f, \chi_S \rangle = \mathbf{E}[f(x)\chi_S(x)]$.*

Proof. To see that this holds, we check that

$$\langle f, \chi_S \rangle = \Big\langle \sum_{T \subseteq [n]} \hat{f}(T)\chi_T, \chi_S \Big\rangle = \sum_{T \subseteq [n]} \hat{f}(T) \cdot \langle \chi_T, \chi_S \rangle = \hat{f}(S).$$

Here we have used the Fourier expansion of f for the first equality, linearity of the inner product for the second, and orthonormality of parity functions (Theorem 3) for the last. ■

Next we have Plancherel's theorem, which states that the inner product of f and g is precisely the dot product of their vectors of Fourier coefficients.

Theorem 5 (Plancherel) *Let* $f, g : \{-1,1\}^n \to \mathbb{R}$*. Then,* $\langle f, g \rangle = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S)$*.*

Proof. Again we use the Fourier expansions of f and g to check that

$$\langle f, g \rangle = \Big\langle \sum_{S \subseteq [n]} \hat{f}(S)\chi_S, \sum_{T \subseteq [n]} \hat{g}(T)\chi_T \Big\rangle = \sum_{S}\sum_{T} \hat{f}(S)\hat{g}(T) \cdot \langle \chi_S, \chi_T \rangle = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S).$$

The second equality holds by linearity of inner product, and the last by orthonormality. ■

An important corollary of Plancherel's theorem is Parseval's identity: if $f : \{-1,1\}^n \to \mathbb{R}$, then $\|f\|_2^2 = \langle f, f \rangle = \sum_{S \subseteq [n]} \hat{f}(S)^2$ (*i.e.* the Fourier transform preserves L_2-norm). In particular, if f is a boolean function then $\sum_{S \subseteq [n]} \hat{f}(S)^2 = \mathbf{E}[f(x)^2] = 1$, which we may view as a probability distribution over the 2^n possible subsets S of $[n]$. Note that if f and g are boolean functions then $f(x)g(x) = 1$ iff $f(x) = g(x)$, and so $\mathbf{E}[f(x)g(x)] = 1 - 2 \cdot \mathrm{dist}(f, g)$, where $\mathrm{dist}(f, g) = \mathbf{Pr}[f(x) \neq g(x)]$ is the normalized Hamming distance between f and g.

One of the advantages of analyzing f via its Fourier expansion is that this polynomial encodes a lot of combinatorial information about f, and these combinatorial quantities can be "read off" its Fourier coefficients easily. We give two basic examples now. Recall that for functions $f : \{-1,1\}^n \to \mathbb{R}$, the mean of f is $\mathbf{E}[f(x)]$ and its variance is $\mathbf{Var}(f) := \mathbf{E}[f(x)^2] - \mathbf{E}[f(x)]^2$. Note that if f is a boolean function then $\mathbf{E}[f(x)]$ measures the bias of f towards 1 or -1, and $\mathbf{Var}(f) = 4 \cdot \mathbf{Pr}[f(x) = 1] \cdot \mathbf{Pr}[f(x) = -1]$. If f has mean 0 and variance 1 we say that f is balanced, or unbiased.

Proposition 6 (expectation and variance) *Let* $f : \{-1,1\}^n \to \mathbb{R}$*. Then* $\mathbf{E}[f(x)] = \hat{f}(\emptyset)$*, and* $\mathbf{Var}(f) = \sum_{S \neq \emptyset} \hat{f}(S)^2$*.*

Proof. For the first equality, we check that $\hat{f}(\emptyset) = \mathbf{E}[f(x)\chi_\emptyset(x)] = \mathbf{E}[f(x)]$. The second equality holds because

$$\mathbf{Var}(f) = \mathbf{E}[f(x)^2] - \mathbf{E}[f(x)]^2 = \Big(\sum_{S \subseteq [n]} \hat{f}(S)^2 \Big) - \hat{f}(\emptyset)^2 = \sum_{S \neq \emptyset} \hat{f}(S)^2.$$

Here the second equality uses an application of Parseval's identity. ■

It is nice to think of $\hat{f}(S)^2$ as the "weight" of f on S, with the sum of weights of f on all 2^n subsets S of $[n]$ being 1 by Parseval's. Often it will also be convenient to stratify these weights according to the cardinality of the set S.

Definition 7 (level weights) *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $k \in \{0, \ldots, n\}$. The weight of f at level k, or the degree-k weight of f, is defined to be $\mathbf{W}^k(f) := \sum_{|S|=k} \hat{f}(S)^2$.*

For example, in this notation we have $\mathbf{W}^0(\mathsf{MAJ}_3) = \mathbf{W}^2(\mathsf{MAJ}_3) = 0$ and $\mathbf{W}^1(\mathsf{MAJ}_3) = \mathbf{W}^3(\mathsf{MAJ}_3) = 1/2$.

1.1.1 Density functions and convolutions

So far we have been viewing domain $\{-1,1\}^n$ of our functions simply as strings of bits represented by real numbers ± 1. Often we would like to be able to "add" two inputs, in which case we will view our functions as $f : \mathbb{F}_2^n \to \mathbb{R}$ instead. The mapping from $\mathbb{F}_2$ to $\{-1,1\}$ is given by $(-1)^b$, sending $0 \in \mathbb{F}_2$ to $1 \in \mathbb{R}$ and $1 \in \mathbb{F}_2$ to $-1 \in \mathbb{R}$. We will sometimes also associate a boolean function $\mathbb{F}_2^n \to \{-1,1\}$ with its corresponding $\mathbb{F}_2$ polynomial $\mathbb{F}_2^n \to \mathbb{F}_2$. For example, the different representations of the parity function χ_S are given in Table 1.

$\chi_S(x) : \{-1,1\}^n \to \{-1,1\}$	$x \mapsto \prod_{i \in S} x_i$
$\chi_S(x) : \mathbb{F}_2^n \to \{-1,1\}$	$x \mapsto \prod_{i \in S} (-1)^{x_i}$
$\chi_S(x) : \mathbb{F}_2^n \to \mathbb{F}_2$	$x \mapsto \sum_{i \in S} x_i$

Table 1: Different representations of $\chi_S(x)$

The $\mathbb{F}_2$ degree of a boolean function $f : \mathbb{F}_2^n \to \{-1,1\}$, denoted $\deg_{\mathbb{F}_2}(f)$, is its degree as an $\mathbb{F}_2$ polynomial $\mathbb{F}_2^n \to \mathbb{F}_2$. For example, the parity functions are degree-1 $\mathbb{F}_2$ polynomials; in contrast, recall that $\deg(\chi_S)$, the Fourier degree of χ_S, is $|S|$. In general we have the inequality $\deg_{\mathbb{F}_2}(f) \le \deg(f)$ for all boolean functions $f : \mathbb{F}_2^n \to \{-1,1\}$. We remark that unlike Fourier degree, it is not known how to infer the $\mathbb{F}_2$ degree of a boolean function from its Fourier expansion. The following useful fact can be easily verified:

Fact 8 *Let $\chi_S : \mathbb{F}_2^n \to \{-1,1\}$. Then $\chi_S(x+y) = \chi_S(x) \cdot \chi_S(y)$.*

Definition 9 (probability density function) *$\varphi : \mathbb{F}_2^n \to \mathbb{R}^{\ge 0}$ is a probability density function of $\mathbf{E}_{x \in \mathbb{F}_2^n}[\varphi(x)] = 1$.*

Note that a probability density function $\varphi : \mathbb{F}_2^n \to \mathbb{R}^{\ge 0}$ corresponds to the probability distribution over $\mathbb{F}_2^n$ where $\mathbf{Pr}[x] = \varphi(x) \cdot 2^n$. For example, the constant function $\varphi \equiv 1$ corresponds to the uniform distribution over $\mathbb{F}_2^n$. For any $a \in \mathbb{F}_2^n$, the density function $\varphi_a(x)$ that takes value 2^n if $x = a$ and 0 otherwise corresponds to the distribution that puts all its weight on a single point $a \in \mathbb{F}_2^n$.

Definition 10 (convolution) *Let $f, g : \mathbb{F}_2^n \to \mathbb{R}$. The convolution of f and g is the function $f * g : \mathbb{F}_2^n \to \mathbb{R}$ defined by $(f * g)(x) := \mathbf{E}_{y \in \mathbb{F}_2^n}[f(y)g(x+y)]$.*

Note that $(f * g)(x) = (g * f)(x)$, since $(\boldsymbol{y}, \boldsymbol{y} + x)$ is just a uniformly random pair of inputs with distance x and therefore has the same distribution as $(\boldsymbol{y} + x, \boldsymbol{y})$. Similarly it can be checked that the convolution operator is commutative: $(f * g) * h = f * (g * h)$. The following facts also follow easily from definitions:

Fact 11 *Let $f : \mathbb{F}_2^n \to \mathbb{R}$ and φ_2, φ_2 be density functions. Then*

1. $\langle \varphi, f \rangle = \mathbf{E}_{y \sim \varphi}[f(y)]$.
2. $(\varphi * f)(x) = \mathbf{E}_{y \sim \varphi}[f(x + y)]$.
3. *The density for $z = y_1 + y_2$, where $y_1 \sim \varphi_1$ and $y_2 \sim \varphi_2$, is $\varphi_1 * \varphi_2$.*

Theorem 12 (Fourier coefficients of convolutions) *Let $f, g : \mathbb{F}_2^n \to \mathbb{R}$. Then $\widehat{f * g}(S) = \hat{f}(S) \cdot \hat{g}(S)$.*

Proof. We check that

$$\begin{aligned}
\widehat{f * g}(S) = \mathop{\mathbf{E}}_x[(f * g)(x)\chi_S(x)] &= \mathop{\mathbf{E}}_x\left[\mathop{\mathbf{E}}_y[f(y)g(x + y)] \cdot \chi_S(x)\right] \\
&= \mathop{\mathbf{E}}_y\left[f(y) \cdot \mathop{\mathbf{E}}_x[g(x + y)\chi_S(x)]\right] \\
&= \mathop{\mathbf{E}}_y\left[f(y) \cdot \mathop{\mathbf{E}}_z[g(z)\chi_S(z + y)]\right] \qquad (1) \\
&= \mathop{\mathbf{E}}_y\left[f(y) \cdot \mathop{\mathbf{E}}_z[g(z)\chi_S(z)\chi_S(y)]\right] \qquad (2) \\
&= \mathop{\mathbf{E}}_y[f(y)\chi_S(y)] \cdot \mathbf{E}[g(z)\chi_S(z)] = \hat{f}(S) \cdot \hat{g}(S).
\end{aligned}$$

Here (1) uses the fact that $z - y = z + y$ for $y, z \in \mathbb{F}_2^n$, and (2) is an application of Fact 8. ■

Theorem 13 *Let $f, g, h : \mathbb{F}_2^n \to \mathbb{R}$. Then $\langle f * g, h \rangle = \langle f, g * h \rangle$.*

Proof. By Theorem 12 and Plancherel, both sides of the identity equal $\sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S)\hat{h}(S)$. ■

1.2 Blum-Luby-Rubinfeld

We begin by considering two notions of what it means for a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ to be linear.

Definition 14 (linear #1) *A boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is linear if $f(x + y) = f(x) + f(y)$ for all $x, y \in \mathbb{F}_2^n$.*

Definition 15 (linear #2) *A boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is linear if there exists $a_1, \ldots, a_n \in \mathbb{F}_2$ such that $f(x) = a_1x_1 + \ldots + a_nx_n$. Equivalently, there exists some $S \subseteq [n]$ such that $f(x) = \sum_{i \in S} x_i$.*

Proposition 16 **($\#1 \iff \#2$)** *These two definitions are equivalent.*

Proof. Suppose f satisfies $f(x+y) = f(x) + f(y)$ for all $x, y \in \mathbb{F}_2^n$. Let $\alpha_i = f(e_i) \in \mathbb{F}_2$, where e_i is the i-th canonical basis vector for $\mathbb{F}_2^n$. It follows that $f(x) = f(\sum_{i=1}^n x_i e_i) = \sum_{i=1}^n x_i f(e_i) = \sum_{i=1}^n \alpha_i x_i$, where the second equality uses Definition 14 repeatedly, along with the fact that $f(x_i e_i) = x_i f(e_i)$. For the reverse implication, note that $f(x+y) = \sum_{i\in S}(x+y)_i = \sum_{i\in S} x_i + \sum_{i\in S} y_i = f(x) + f(y)$, where the first and final equalities uses Definition 15. ∎

It is natural to consider analogous notions for approximate linearity.

Definition 17 **(approximately linear $\#1$)** *A boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is approximately linear if $f(x+y) = f(x) + f(y)$ for most pairs $x, y \in \mathbb{F}_2^n$.*

Definition 18 **(approximately linear $\#2$)** *A boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is approximately linear if there exists some $S \subseteq [n]$ such that $f(x) = \sum_{i\in S} x_i$ for most $x \in \mathbb{F}_2^n$. Equivalently, there exists an $S \subseteq [n]$ such that f is close in Hamming distance to $g(x) = \sum_{i\in S} x_i$.*

A straightforward generalization of argument given in the proof of Proposition 16 shows that Definition 18 (approximately linear #2) implies Definition 17 (approximately linear #1). However, the argument for the reverse implication no longer holds. We will adopt Definition 18 as our notion of approximate linearity for now, and we will see that the linearity test of Blum, Luby, and Rubinfeld [BLR93] implies that both definitions are in fact equivalent. The Fourier-analytic proof we present here is due to Bellare *et. al* [BCH+96].

Definition 19 **(BLR linearity test)** *Given blackbox access to a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$,*

1. *Pick $x, y \in \mathbb{F}_2^n$ independently and uniformly.*
2. *Query f on x, y and $x+y$.*
3. *Accept iff $f(x) + f(y) = f(x+y)$.*

Theorem 20 **(soundness of BLR)** *If $\mathbf{Pr}[\textsf{BLR accepts } f] \geq 1-\varepsilon$ then f is ε-close to being linear (in the sense of Definition 18).*

Proof. It will be convenient to think of f as $\mathbb{F}_2^n \to \{-1, 1\}$, and so the acceptance criterion (*i.e.* step 3) becomes $f(x)f(y) = f(x+y)$. Viewing f this way, now note that

$$\begin{aligned}
\mathbf{Pr}[\textsf{BLR accepts } f] &= \mathop{\mathbf{E}}_{x,y}\left[\mathbf{1}(f(x)\cdot f(y) = f(x+y))\right] \\
&= \mathop{\mathbf{E}}_{x,y}\left[\tfrac{1}{2} + \tfrac{1}{2}\cdot f(x)f(y)f(x+y)\right] \\
&= \tfrac{1}{2} + \tfrac{1}{2}\mathop{\mathbf{E}}_{x}[f(x)\cdot(f*f)(x)] \\
&= \tfrac{1}{2} + \tfrac{1}{2}\sum_{S\subseteq[n]} \hat{f}(S)\cdot\widehat{f*f}(S) \qquad (3) \\
&= \tfrac{1}{2} + \tfrac{1}{2}\sum_{S\subseteq[n]} \hat{f}(S)^3. \qquad (4)
\end{aligned}$$

Here (3) uses Parseval's identity and (4) uses Theorem 12. Therefore, if $\mathbf{Pr}[\mathsf{BLR} \text{ accepts } f] \geq 1-\varepsilon$ then $1-2\varepsilon \leq \sum_{S\subseteq[n]} \hat{f}(S)^3 \leq \max\{\hat{f}(S)\} \cdot \sum_{S\subseteq[n]} \hat{f}(S)^2 = \max\{\hat{f}(S)\}$, or equivalently, there exists an $S^* \subseteq [n]$ such that $\hat{f}(S^*) \geq 1 - 2\varepsilon$. Since $\hat{f}(S^*) = \mathbf{E}[f(x)\chi_{S^*}(x)] = 1 - 2 \cdot \mathrm{dist}(f, \chi_{S^*})$, we have shown that $\mathrm{dist}(f, \chi_{S^*}) \leq \varepsilon$ and the proof is complete. ■

Theorem 20 says that if $\mathbf{Pr}[\mathsf{BLR} \text{ accepts } f] \geq 1-\varepsilon$ then f is ε-close to some linear function χ_{S^*}; however, we do not know *which* of the 2^n possible linear functions χ_{S^*} this is. The following theorem tells us that we can nevertheless obtain the correct value of $\chi_{S^*}(x)$ with high probability for all $x \in \mathbb{F}_2^n$.

Theorem 21 (local decodability of linear functions) *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be ε-close to some linear function χ_{S^*}, and let $x \in \mathbb{F}_2^n$. The following algorithm outputs $\chi_{S^*}(x)$ with probability at least $1 - 2\varepsilon$:*

1. *Pick $y \in \mathbb{F}_2^n$ uniformly.*
2. *Output $f(y) + f(x + y)$.*

Proof. Since x and $x + y$ are both uniform (though not independent), with probability at least $1 - 2\varepsilon$ we have $f(y) = \chi_{S^*}(y)$ and $f(x+y) = \chi_{S^*}(x+y)$. The claim follows by noting that $\chi_{S^*}(y) + \chi_{S^*}(x + y) = \chi_{S^*}(y + (x + y)) = \chi_{S^*}(x)$, where we have used the linearity of χ_{S^*} along with the fact that $x + y = x - y$ for $x, y \in \mathbb{F}_2^n$. ■

1.3 Voting and influence

- Puzzle: Is it possible for $f : \{-1, 1\}^n \to \{-1, 1\}$ to have exactly k non-zero Fourier coefficients, for $k = 0, 1, 2, 3, 4, 5, 6, 7$? Classify all functions with 2 non-zero Fourier coefficients.
- Puzzle: Find all $f : \{-1, 1\}^n \to \{-1, 1\}$ with $\mathbf{W}^1(f) = 1$.

We may think of a boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ as a voting scheme for an election with 2 candidates (±1) and n voters $(x_1, \ldots, x_n)$. Many boolean functions are named after the voting schemes they correspond to: the i-th dictator $\mathsf{DICT}_i(x) = x_i$ (*i.e.* $\mathsf{DICT}_i \equiv \chi_i$); k-juntas (functions that depend only on k of its n variables, where we think of k as $\ll n$, or even a constant); the majority function $\mathsf{MAJ}(x) = \mathrm{sgn}(x_1 + \ldots + x_n)$. The majority function is special instance of linear threshold functions $f(x) = \mathrm{sgn}(a_0 + a_1x_1 + \ldots + a_nx_n)$, $a_i \in \mathbb{R}$, also known as weighted-majority functions, or halfspaces. Another important voting scheme in boolean function analysis is $\mathsf{TRIBES}_{w,s} : \{-1, 1\}^{ws} \to \{-1, 1\}$, the s-way OR of w-way AND's of disjoint sets of variables (where we think of -1 as true and 1 as false). In $\mathsf{TRIBES}_{w,s}$, the candidate -1 is elected iff at least one member of each of the s disjoint tribes of w members votes for -1.

The following are a few reasonable properties one may expect of a voting scheme:

- Monotone: if $x_i \leq y_i$ for all $i \in [n]$ then $f(x) \leq f(y)$.
- Symmetric: $f(\pi(x)) = f(x)$ for all permutations $\pi \in S_n$ and $x \in \{-1, 1\}^n$.

- Transitive-symmetric (weaker than symmetric): for all $i, j \in [n]$ there exists a permutation $\pi \in S_n$ such that $\pi(i) = j$ and $f(x) = f(\pi(x))$ for all $x \in \{-1,1\}^n$.

Later in this section (for the proof of Arrow's theorem) we will also assume that voters vote independently and uniformly; this is known as the impartial culture assumption in social choice theory.

Definition 22 (influence) *Let $f : \{-1,1\}^n \to \{-1,1\}$. We say that variable $i \in [n]$ is pivotal for $x \in \{-1,1\}^n$ if $f(x) \neq f(x^{\oplus i})$, where $x^{\oplus i}$ is the string x with its i-th bit flipped. The influence of variable i on f, denoted $\mathrm{Inf}_i(f)$, is the fraction of inputs for which i is pivotal. That is, $\mathrm{Inf}_i(f) := \mathbf{Pr}[f(x) \neq f(x^{\oplus i})]$.*

For example, $\mathrm{Inf}_i(\mathsf{DICT}_j)$ is 1 if $i = j$ and 0 otherwise. For the majority function over an odd number n of variables, $\mathrm{Inf}_i(\mathsf{MAJ}) = \binom{n-1}{(n-1)/2} \cdot 2^{-(n-1)}$ since voter i is pivotal iff the votes are split evenly among the other $n-1$ voters. By Stirling's approximation, this quantity is $\sim \sqrt{2/\pi n} = \Theta(1/\sqrt{n})$.

Definition 23 (derivative) *Let $f : \{-1,1\}^n \to \mathbb{R}$. The i-th derivative of f is the function $(D_i f)(x) := \frac{1}{2}(f(x^{i\leftarrow 1}) - f(x^{i\leftarrow -1}))$, where $x^{i\leftarrow b}$ is the string x with its i-th bit set to b.*

Note that if f is a boolean function then $(D_i f)(x) = \pm 1$ if i is pivotal for f at x, and 0 otherwise, and therefore $\mathbf{E}[(D_i f)(x)^2] = \mathrm{Inf}_i(f)$. We will adopt $\mathbf{E}[(D_i f)(x)^2]$ as the generalized definition of the influence of variable i on f for real-valued functions $f : \{-1,1\}^n \to \mathbb{R}$.

Theorem 24 (Fourier expressions for derivatives and influence) *Let $f : \{-1,1\}^n \to \mathbb{R}$. Then*

1. $(D_i f)(x) = \sum_{S \ni i} \hat{f}(S)\chi_{S\setminus i}(x)$.
2. $\mathrm{Inf}_i(f) = \sum_{S \ni i} \hat{f}(S)^2$.

Proof. The first identity holds by noting that $(D_i \chi_S)(x) = \chi_{S\setminus i}(x)$ if $i \in S$ and 0 otherwise, along with the fact that D_i is a linear operator, *i.e.* $D_i(\alpha f + g) = \alpha(D_i f) + (D_i g)$. The second identity then follows by applying Parseval to $\mathrm{Inf}_i(f) = \mathbf{E}[(D_i f)(x)^2]$. ∎

Proposition 25 (influence of monotone functions) *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a monotone function. Then $\mathrm{Inf}_i(f) = \hat{f}(i)$.*

Proof. If f is monotone then $(D_i f)(x) \in \{0,1\}$ and so $\mathrm{Inf}_i(f) = \mathbf{E}[(D_i f)(x)^2] = \mathbf{E}[(D_i f)(x)] = \widehat{D_i f}(\emptyset) = \hat{f}(i)$. Here the final equality uses the Fourier expansion of $D_i f$ given by Theorem 24. ∎

An immediate corollary of Proposition 25 is that monotone, transitive-symmetric functions f have $\mathrm{Inf}_i(f) \leq 1/\sqrt{n}$ for all $i \in [n]$. This follows from the fact that transitive-symmetric functions satisfy $\hat{f}(i) = \hat{f}(j)$ for all $i, j \in [n]$, along with the bound $\sum_{i=1}^n \hat{f}(i)^2 \leq \sum_{S \subseteq [n]} \hat{f}(S)^2 = 1$.

Definition 26 (total influence) *Let* $f : \{-1,1\}^n \to \mathbb{R}$*. The total influence of* f *is* $\mathrm{Inf}(f) := \sum_{i=1}^n \mathrm{Inf}_i(f)$*.*

If f is a boolean function, then

$$\mathrm{Inf}(f) = \sum_{i=1}^n \mathbf{Pr}[f(x) \neq f(x^{\oplus i})] = \sum_{i=1}^n \mathbf{E}[\mathbf{1}(f(x) \neq f(x^{\oplus i}))] = \mathbf{E}\Big[\sum_{i=1}^n \mathbf{1}(f(x) \neq f(x^{\oplus i}))\Big].$$

The quantity $\sum_{i=1}^n \mathbf{1}(f(x) \neq f(x^{\oplus i}))$ is known as the sensitivity of f at x, and so the total influence of a boolean function is also known as its average sensitivity. If f is viewed as a 2-coloring of the boolean hypercube, the total influence can also be seen to be equal to n times the fraction of bichromatic edges.

If f is a monotone boolean function, we see that $\mathrm{Inf}(f) = \sum_{i=1}^n \hat{f}(i) = \sum_{i=1}^n \mathbf{E}[f(x)x_i] = \mathbf{E}[f(x)(x_1 + \ldots + x_n)]$. Recall that if f is boolean then $f(x)(x_1 + \ldots + x_n) = 1$ if $f(x) = \mathrm{sgn}(x_1 + \ldots + x_n)$ and -1 otherwise. Therefore the total influence a monotone boolean function, when viewed as a voting scheme, measures the expected difference between the number of voters whose vote agrees with the outcome of the election and the number whose vote disagrees. It is reasonable to expect this quantity to large, and the next proposition states that (if n is odd) it is maximized by the majority function.

Proposition 27 (MAJ maximizes sum of linear coefficients) *Let* n *be odd. Among all boolean functions* $f : \{-1,1\}^n \to \{-1,1\}$*, the quantity* $\sum_{i=1}^n \hat{f}(i)$ *is maximized by* $\mathsf{MAJ}(x) = \mathrm{sgn}(x_1 + \ldots + x_n)$*. Consequently, if* f *is monotone then* $\mathrm{Inf}(f) \leq \mathrm{Inf}(\mathsf{MAJ}) \sim \sqrt{2n/\pi}$*.*

Proof. Note that $\sum_{i=1}^n \hat{f}(i) = \mathbf{E}[f(x)(x_1 + \ldots + x_n)] \leq \mathbf{E}[|x_1 + \ldots + x_n|]$ since f is (± 1)-valued, where the inequality is tight iff $f(x) = \mathrm{sgn}(x_1 + \ldots + x_n) = \mathsf{MAJ}(x)$. For the second claim recall that $\mathrm{Inf}_i(f) = \hat{f}(i)$ if f is monotone (Proposition 25), and $\mathrm{Inf}_i(\mathsf{MAJ}) \sim \sqrt{2/\pi n}$ for all $i \in [n]$. ■

Proposition 28 (Fourier expression for total influence) *Let* $f : \{-1,1\}^n \to \mathbb{R}$*. Then* $\mathrm{Inf}_i(f) = \sum_{S \subseteq [n]} |S| \cdot \hat{f}(S)^2 = \sum_{k=1}^n k \cdot \mathbf{W}^k(f)$*.*

The proof of this proposition follows immediately from the Fourier expression for variable influence given by Theorem 24. Notice that each Fourier coefficient is weighted by its cardinality in the sum, and so total influence may also be viewed as a measure of the "average degree" of f's Fourier expansion.

Recall that for functions $f : \{-1,1\}^n \to \mathbb{R}$ we have $\mathbf{Var}(f) = \sum_{S \neq \emptyset} \hat{f}(S)^2$ (Proposition 6), and comparing this quantity with the Fourier expression for total influence yields $\mathbf{Var}(f) \leq \mathrm{Inf}(f)$, the Poincaré inequality for the boolean hypercube. The inequality is tight iff $\mathbf{W}^1(f) = 1$, which for boolean functions implies that $f = \pm\mathsf{DICT}_i(f)$ for some $i \in [n]$. If f is a boolean function and $p := \mathbf{Pr}[f(x) = 1]$, it is easy to check that $\mathbf{Var}(p) = 4p(1-p)$, and so the Poincaré inequality can be equivalently stated as

$4p(1-p) \le \mathrm{Inf}(f) = n \cdot$ (fraction of bichromatic edges). The Poincaré inequality is therefore an edge-isoperimetric inequality for the boolean hypercube (where we view boolean functions as indicators of subsets of $\{-1,1\}^n$): for any p, it gives a lower bound on the number of boundary edges between A and $\overline{A}$ where $A \subseteq \{-1,1\}^n$ has density p. This is a sharp bound when $p = 1/2$, but not when p is small. For smaller densities we have the bound $2\alpha \log_2(1/\alpha) \le \mathrm{Inf}(f)$, where $\alpha := \min\{\mathbf{Pr}[f(x) = 1], \mathbf{Pr}[f(x) = -1]\}$. This in turn is sharp whenever $\alpha = 2^k$, achieved by the AND of k coordinates.

1.4 Noise stability and Arrow's theorem

Let $\rho \in [0,1]$ and fix $x \in \{-1,1\}^n$. Let $N_\rho(x)$ be the distribution on $\{-1,1\}^n$ where $y \sim N_\rho(x)$ if for all $i \in [n]$, $y_i = x_i$ with probability ρ, and y_i is uniformly random ± 1 with probability $1-\rho$. More generally, for $\rho \in [-1,1]$, we have that $N_\rho(x)$ is the distribution on strings y where

$$y_i = \begin{cases} x_i & \text{with probability } \frac{1}{2} + \frac{1}{2}\rho \\ -x_i & \text{with probability } \frac{1}{2} - \frac{1}{2}\rho. \end{cases}$$

If $x \sim \{-1,1\}^n$ is uniformly random and $y \sim N_\rho(x)$, we say that x and y are ρ-correlated strings; equivalently, x and y are ρ-correlated if they are both uniformly random and $\mathbf{E}[x_i y_i] = \rho$ for all $i \in [n]$.

Definition 29 (noise stability) *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $\rho \in [-1,1]$. The noise stability of f at noise rate ρ is*

$$\mathrm{Stab}_\rho(f) := \mathbf{E}[f(x)f(y)], \quad \text{where } x, y \text{ are } \rho\text{-correlated strings.}$$

For example $\mathrm{Stab}_\rho(\pm 1) = 1$, $\mathrm{Stab}_\rho(\mathsf{DICT}_i) = \rho$, and $\mathrm{Stab}_\rho(\chi_S) = \rho^{|S|}$. Tomorrow we will prove Sheppard's formula: $\lim_{n\to\infty} \mathrm{Stab}_\rho(\mathsf{MAJ}) = 1 - \frac{2}{\pi}\arccos(\rho)$. In particular, if $\rho = 1-\delta$ we have $\mathrm{Stab}_\rho(\mathsf{MAJ}) = \Theta(\sqrt{\delta})$.

Definition 30 (noise operator) *Let $\rho \in [-1,1]$. The noise operator T_ρ on functions $f : \{-1,1\}^n \to \mathbb{R}$ acts as follows: $(T_\rho f)(x) := \mathbf{E}_{y \sim N_\rho(x)}[f(y)]$.*

Proposition 31 (Fourier expressions for noise operator and stability) *Let $\rho \in [-1,1]$ and $f : \{-1,1\}^n \to \mathbb{R}$. Then*

1. $(T_\rho f)(x) = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \chi_S(x)$.
2. $\mathrm{Stab}_\rho(f) = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S)^2$.

Proof. The first identity follows from the linearity of the noise operator, along with the observation that $(T_\rho \chi_S)(x) = \rho^{|S|}\chi_S(x)$. The second holds by noting that

$$\mathrm{Stab}_\rho(f) = \mathop{\mathbf{E}}_{(x,y)\ \rho\text{-corr}}[f(x)f(y)] = \mathop{\mathbf{E}}_{x}[f(x)(T_\rho f)(x)] = \sum_{S \subseteq [n]} \hat{f}(S)\widehat{T_\rho f}(S) = \sum_{S \subseteq [n]} \rho^{|S|}\hat{f}(S)^2.$$

■

Suppose there is an election with n voters and three candidates: A, B and C. Each voter ranks the candidates by submitting three bits indicating her preferences: whether the prefers A to B (say, -1 if so and 1 otherwise), and similarly for B versus C and C versus A. Clearly a rational voter cannot simultaneously prefer A to B, B to C and C to A; her ordering of the candidates must be non-cyclic.

Definition 32 (rational) *A triple $(a, b, c) \in \{-1, 1\}^3$ is rational if not all three bits are equal (i.e., (a, b, c) defines a total ordering, and is a valid preference profile). We define the function $\mathsf{NAE} : \{-1, 1\}^3 \to \{1, 0\}$ to be 1 iff not all three bits are equal.*

Now suppose the preferences of the n voters are aggregated into three n-bit strings x, y and z, and the aggregate preference of the electorate is represented by the triple $(f(x), f(y), f(z))$ for some boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$. Clearly we would like for the the outcome of the election to be rational; that is, $\mathsf{NAE}(f(x), f(y), f(z)) = 1$.

Fact 33 (Condorcet's paradox [Con85]) *With MAJ as the aggregating function it is possible that all voters submit rational preferences and yet the aggregated preference string is irrational.*

	v_1	v_2	v_3	MAJ
$A > B\,?$	$+1$	$+1$	-1	$+1$
$B > C\,?$	$+1$	-1	$+1$	$+1$
$C > A\,?$	-1	$+1$	$+1$	$+1$

Figure 1: An instance of Condorcet's paradox

Theorem 34 (Arrow's impossibility theorem [Arr50]) *Suppose f is an aggregating function that always produces a rational outcome if all voters vote rationally. Then $f = \pm\mathsf{DICT}_i$ for some $i \in [n]$. If f is further restricted to be unanimous (i.e. $f(1, \dots, 1) = 1$ and $f(-1, \dots, -1) = -1$; certainly a very reasonable assumption) then f must be a dictator.*

The main result of this section is a robust version of Arrow's impossibility theory due Gil Kalai [Kal02]. It expresses the probability that an aggregating function f produces a rational outcome in terms of the noise stability of f, under the impartial culture assumption (each voter selects an NAE-triple (x_i, y_i, z_i) uniformly and independently).

Theorem 35 (Kalai) $\mathbf{E}[\mathsf{NAE}(f(x), f(y), f(z))] = \frac{3}{4} - \frac{3}{4} \cdot \mathrm{Stab}_{-1/3}(f)$*, where the expectation is taken with respect to the impartial culture assumption.*

Proof. Using the arithmetization $\mathsf{NAE}(a, b, c) = \frac{3}{4} - \frac{1}{4}(ab + bc + ac)$, we first note that

$$\begin{aligned}
\mathbf{E}[\mathsf{NAE}(f(x), f(y), f(z))] &= \tfrac{3}{4} - \tfrac{1}{4}\big(\mathbf{E}[f(x)f(y)] + \mathbf{E}[f(y)f(z)] + \mathbf{E}[f(x)f(z)]\big) \\
&= \tfrac{3}{4} - \tfrac{3}{4}\,\mathbf{E}[f(x)f(y)],
\end{aligned}$$

where again, all expectations are taken with respect to the impartial culture assumption. Since $\mathbf{E}[x_i y_i] = -1/3$ if (x_i, y_i, z_i) is an NAE-triple, the quantity $\mathbf{E}[f(x)f(y)]$ is exactly $\mathrm{Stab}_{-1/3}(f)$ and the proof is complete. ■

Theorem 35 does indeed imply Arrow's impossibility theorem since

$$\tfrac{3}{4} - \tfrac{3}{4} \cdot \mathrm{Stab}_{-1/3}(f) = \tfrac{3}{4} - \tfrac{3}{4} \sum_{k=0}^{n} (-\tfrac{1}{3})^k \cdot \mathbf{W}^k(f) \le \tfrac{7}{9} + \tfrac{2}{9} \cdot \mathbf{W}^1(f),$$

and so if $\mathbf{E}[\mathsf{NAE}(f(x), f(y), f(z))] = 1$ then $\mathbf{W}^1(f) \ge 1$. Furthermore note that the probability of an irrational outcome is at least $1 - \varepsilon$ then $\mathbf{W}^1 \ge 1 - O(\varepsilon)$. By a theorem of E. Friedgut, G. Kalai and A. Naor [FKN02], if $\mathbf{W}^1(f) \ge 1 - \varepsilon$ then f is $O(\varepsilon)$-close to $\pm\mathsf{DICT}_i$ for some $i \in [n]$. Therefore Kalai's theorem is in fact a robust version of Arrow's impossibility theorem: if most rational voter preference profiles aggregate to a rational outcome, then the aggregating function must be close to a dictator or anti-dictator.

We conclude by giving an upper bound on level-1 Fourier weight of transitive-symmetric functions. By Theorem 35, this gives an upper bound on the probability that such functions aggregate rational voter preference profiles to a rational outcome. We will also prove a generalization of this fact (Proposition 42) using the Berry-Esséen theorem tomorrow.

Proposition 36 ($\mathbf{W}^1(f)$ of transitive-symmetric functions) *Suppose $\hat{f}(i) = \hat{f}(j)$ for all $i, j \in [n]$. Then $\mathbf{W}^1(f) \le \frac{2}{\pi} + o_n(1)$.*

Proof. First note that $\sum_{i=1}^n \hat{f}(i)^2 = n \cdot \hat{f}(1)^2 = n \cdot \big(\frac{1}{n}\sum_{i=1}^n \hat{f}(i)\big)^2 = \frac{1}{n}\big(\sum_{i=1}^n \hat{f}(i)\big)^2$. The claim then follows since we have seen that $\sum_{i=1}^n \hat{f}(i) \le \sum_{i=1}^n \widehat{\mathsf{MAJ}}(i) \sim \sqrt{2n/\pi}$ (Proposition 27). ■

2 Noise stability and small set expansion

Tuesday, 28th February 2012

- Puzzle: Compute the Fourier expansion of MAJ_n. Hint: consider $T_\rho D_i \mathsf{MAJ}(1,\ldots,1)$.

Roughly speaking, the central limit theorem states that if $X_1,\ldots,X_n$ are independent random variables where none of the X_i's are "too dominant", then $S=\sum_{i=1}^n X_i$ is distributed like a Gaussian as $n\to\infty$. As a warm-up, we begin by giving another proof of the fact that $\mathrm{Inf}(\mathsf{MAJ})\sim\sqrt{2n/\pi}$, this time using the central limit theorem. First note that

$$\mathrm{Inf}(\mathsf{MAJ}) = \mathbf{E}\left[\mathsf{MAJ}(x)\sum_{i=1}^n x_i\right] = \mathbf{E}\left[\left|\sum_{i=1}^n x_i\right|\right] = \sqrt{n}\cdot\mathbf{E}\left[\left|\sum_{i=1}^n \frac{x_i}{\sqrt{n}}\right|\right]$$

Now by the central limit theorem we know that $\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i \to \mathcal{G}\sim N(0,1)$ as $n\to\infty$, and since $\mathbf{E}[|\mathcal{G}|]=\sqrt{2/\pi}$, we conclude that $\mathrm{Inf}(\mathsf{MAJ})\sim\sqrt{2/\pi}\cdot\sqrt{n}$.

Definition 37 (reasonable r.v.) *Let $B\ge 1$. We say that a random variable X is B-reasonable if $\mathbf{E}[X^4]\le B\cdot\mathbf{E}[X^2]^2$. Equivalently, $\|X\|_4\le B^{1/4}\cdot\|X\|_2$.*

For example, a uniformly random ±1 bit (*i.e.* a Rademacher random variable) is 1-reasonable, and a standard Gaussian is 3-reasonable. The Berry-Esséen theorem [Ber41, Ess42] is a finitary version of the central limit theorem, giving explicit bounds on the rate at which reasonable random variables converge towards the Gaussian distribution.

Theorem 38 (Berry-Esséen) *Let $X_1,\ldots,X_n$ be independent, B-reasonable random variables satisfying $\mathbf{E}[X_i]=0$. Let $\sigma_i^2:=\mathbf{E}[X_i^2]$ and suppose $\sum_{i=1}^n\sigma_i^2=1$. Let $S=X_1+\ldots+X_n$ and $\mathcal{G}\sim N(0,1)$. For all $t\in\mathbb{R}$,*

$$|\,\mathbf{Pr}[S\le t]-\mathbf{Pr}[\mathcal{G}\le t]|\le O(\varepsilon)$$

where $\varepsilon=\left(B\cdot\sum_{i=1}^n\sigma_i^4\right)^{1/2}\le\sqrt{B}\cdot\max\{|\sigma_i|\}$.

We prove the Berry-Esséen theorem with a weaker bound of $\varepsilon=\left(B\cdot\sum_{i=1}^n\sigma_i^4\right)^{1/5}$ in Section 4.2.

2.1 Sheppard's formula and $\mathrm{Stab}_\rho(\mathsf{MAJ})$

Definition 39 (ρ-correlated Gaussians) *Let $\mathcal{G},\mathcal{G}'\sim N(0,1)$ be independent standard Gaussians. Set $\mathcal{H}=(\mathcal{G},\mathcal{G}')\cdot(\rho,\sqrt{1-\rho^2}):=\rho\cdot\mathcal{G}+\sqrt{1-\rho^2}\cdot\mathcal{G}'$. Then $\mathcal{G}$ and $\mathcal{H}$ are ρ-correlated Gaussians. Note that if $\mathcal{G}$ and $\mathcal{H}$ are ρ-correlated Gaussians then $\mathbf{E}[\mathcal{G}\mathcal{H}]=\rho\cdot\mathbf{E}[\mathcal{G}^2]+\sqrt{1-\rho^2}\cdot\mathbf{E}[\mathcal{G}]\,\mathbf{E}[\mathcal{G}']=\rho$.*

Theorem 40 ([She99]) *Let $\mathcal{G}$ and $\mathcal{H}$ be ρ-correlated Gaussians. Then $\mathbf{Pr}[\mathrm{sgn}(\mathcal{G})\ne\mathrm{sgn}(\mathcal{H})]=\arccos(\rho)/\pi$.*

Proof. First recall that $\operatorname{sgn}(\vec{u} \cdot \vec{v})$ is determined by which side of the halfspace normal to $\vec{u}$ the vector $\vec{v}$ falls on. Since $\mathcal{H} = (\rho, \sqrt{1-\rho^2}) \cdot (\mathcal{G}, \mathcal{G}')$, and $\mathcal{G} = (1,0) \cdot (\mathcal{G}, \mathcal{G}')$, it follows that $\mathbf{Pr}[\operatorname{sgn}(\mathcal{G}) \neq \operatorname{sgn}(\mathcal{H})]$ is precisely the probability that the halfspace normal to $(\mathcal{G}, \mathcal{G}')$ splits the vectors $(\rho, \sqrt{1-\rho^2})$ and $(1,0)$. Therefore,

$$\mathbf{Pr}[\operatorname{sgn}(\mathcal{G}) \neq \operatorname{sgn}(\mathcal{H})] = \tfrac{1}{\pi} \cdot (\text{angle between } (\rho, \sqrt{1-\rho^2}) \text{ and } (1,0)) = \tfrac{1}{\pi} \arccos(\rho).$$

■

Next, we use Sheppard's formula to prove that $\mathrm{Stab}_\rho(\mathsf{MAJ}) \to 1 - \frac{2}{\pi}\arccos(\rho)$ as $n \to \infty$. First recall that

$$\mathrm{Stab}_\rho(\mathsf{MAJ}) = \mathop{\mathbf{E}}_{(x,y)\ \rho\text{-corr}}[\mathsf{MAJ}(x)\mathsf{MAJ}(y)] = 1 - 2\,\mathbf{Pr}[\mathsf{MAJ}(x) \neq \mathsf{MAJ}(y)],$$

and so it suffices to argue that $\mathbf{Pr}[\mathsf{MAJ}(x) \neq \mathsf{MAJ}(y)] \to \frac{1}{\pi}\arccos(\rho)$. Next, we view

$$\mathsf{MAJ}(x) = \operatorname{sgn}\Big(\frac{x_1 + \ldots + x_n}{\sqrt{n}}\Big), \quad \mathsf{MAJ}(y) = \operatorname{sgn}\Big(\frac{x_1 + \ldots + x_n}{\sqrt{n}}\Big),$$

and note that

$$\mathbf{E}\left[\Big(\sum_{i=1}^n \frac{x_i}{\sqrt{n}}\Big) \cdot \Big(\sum_{i=1}^n \frac{y_i}{\sqrt{n}}\Big)\right] = \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbf{E}[x_i y_i] = \rho.$$

While the standard central limit theorem tells us that $\vec{X} = (x_1 + \ldots + x_n)/\sqrt{n}$ and $\vec{Y} = (y_1 + \ldots + y_n)/\sqrt{n}$ each individually converges towards the standard Gaussian $\mathcal{G} \sim N(0,1)$, the two-dimensional central limit theorem states that $(\vec{X}, \vec{Y})$ actually converge to ρ-correlated Gaussians $(\mathcal{G}, \mathcal{H})$ as $n \to \infty$. In fact, the two-dimensional Berry-Esséen theorem quantifies this rate of convergence, bounding the error by $\pm O(1/\sqrt{n})$ as long as ρ is bounded away from ± 1. Combining this with Sheppard's formula, we conclude that

$$\mathbf{Pr}[\mathsf{MAJ}(x) \neq \mathsf{MAJ}(y)] \to \mathbf{Pr}[\operatorname{sgn}(\mathcal{G}) \neq \operatorname{sgn}(\mathcal{H})] = \tfrac{1}{\pi}\arccos(\rho) \quad \text{as } n \to \infty.$$

Since

$$\mathrm{Stab}_\rho(\mathsf{MAJ}) \to 1 - \tfrac{2}{\pi}\arccos(\rho) = \tfrac{2}{\pi} + \tfrac{3}{\pi}\rho^3 + \ldots + \tbinom{k-1}{(k-1)/2}\tfrac{4}{\pi k 2^k} \cdot \rho^k + \ldots$$

and $\mathrm{Stab}_\rho(\mathsf{MAJ}) = \sum_{k=0}^n \rho^k \cdot \mathbf{W}^k(\mathsf{MAJ})$, we have that $\mathbf{W}^k(\mathsf{MAJ}) \to \binom{k-1}{(k-1)/2}\frac{4}{\pi k 2^k}$ as $n \to \infty$.

2.2 The noisy hypercube graph

Let $\rho \in [-1,1]$. The ρ-noisy hypercube graph is a complete weighted graph on the vertex set $\{-1,1\}^n$, where the weight on an edge (x,y) is the probability of getting x and y when drawing ρ-correlated strings from $\{-1,1\}^n$. Equivalently, $\mathrm{wt}(x,y) = \mathbf{Pr}[x \leftarrow \mathcal{U}]\,\mathbf{Pr}[y \leftarrow N_\rho(x)] = 2^{-n} \cdot (\frac{1}{2} - \frac{1}{2}\rho)^{\Delta(x,y)}(\frac{1}{2} + \frac{1}{2}\rho)^{n-\Delta(x,y)}$, and the sum of weights of all edges incident to any $x \in \{-1,1\}^n$ is exactly 2^{-n}.

Recall that if f is a boolean function, then $\mathrm{Stab}_\rho(f) = \mathbf{E}_{(x,y)\ \rho\text{-corr}}[f(x)f(y)] = 1-2\,\mathbf{Pr}[f(x) \neq f(y)]$, or equivalently, $\mathbf{Pr}_{(x,y)\ \rho\text{-corr}}[f(x) \neq f(y)] = \frac{1}{2} - \frac{1}{2} \cdot \mathrm{Stab}_\rho(f)$. Viewing f as the indicator of a subset $A_f \subseteq \{-1,1\}^n$, the quantity $\mathbf{Pr}[f(x) \neq f(y)]$ is the sum of weights of edges going from A_f to its complement $\overline{A_f}$. Therefore, roughly speaking, a function f is noise stable iff the sum of weights of edges contained within A_f and $\overline{A_f}$ is large. On Friday we will prove the Majority Is Stablest theorem due to E. Mossel, R. O'Donnell, and K. Olezkiewicz [MOO10]:

> **Majority Is Stablest.** Fix a constant $0 < \rho < 1$, and let $f : \{-1,1\}^n \to \{-1,1\}$ be a balanced function. It is easy to see that $\mathrm{Stab}_\rho(f)$ is maximized when $\mathbf{W}^1(f) = 1$, in which case $f = \pm\mathsf{DICT}_i$ and $\mathrm{Stab}_\rho(\mathsf{DICT}_i) = \rho$. However, if we additionally require that $\mathrm{Inf}_i(f) \leq \tau$ for all $i \in [n]$, then the theorem states that $\mathrm{Stab}_\rho(f) \leq \mathrm{Stab}_\rho(\mathsf{MAJ}) + o_\tau(1)$; *i.e.* the set A of density $1/2$ with all influences small that maximizes sum of weights of edges within A and $\overline{A}$ is A_{MAJ}.

Let $A \subseteq \{-1,1\}^n$ and $\mathbf{1}_A(x) : \{-1,1\}^n \to \{0,1\}$ be its indicator function, and let $\alpha := \mathbf{E}[\mathbf{1}_A(x)] = |A| \cdot 2^{-n}$ denote its density. Recall that $\mathrm{Stab}_\rho(\mathbf{1}_A) = \mathbf{E}[\mathbf{1}_A(x)\mathbf{1}_A(y)] = \mathbf{Pr}[x \in A\ \&\ y \in A]$, or equivalently, $\mathbf{Pr}[y \in A \,|\, x \in A] = \frac{1}{\alpha} \cdot \mathrm{Stab}_\rho(\mathbf{1}_A)$, and so $\mathrm{Stab}_\rho(\mathbf{1}_A)$ is the probability that a random walk (where each coordinate is independently flipped with probability $(\frac{1}{2} - \frac{1}{2}\rho)$) starting at a point $x \in A$ remains in A, normalized by the density of A. Tomorrow we will prove (a specific instance of) the small set expansion theorem:

> **Small Set Expansion.** $\mathrm{Stab}_\rho(\mathbf{1}_A) \leq \alpha^{2/(1+\rho)}$, or equivalently, $\mathbf{Pr}[y \in A \,|\, x \in A] \leq \alpha^{(1-\rho)/(1+\rho)}$. In particular, if α is small, the probability that a random walk starting in A landing outside A is very high.

Recall that $\mathrm{Stab}_\rho(\mathbf{1}_A) = \mathbf{W}^0(\mathbf{1}_A) + \rho\cdot\mathbf{W}^1(\mathbf{1}_A) + \rho^2\cdot\mathbf{W}^2(\mathbf{1}_A) + \ldots$, and so $\frac{d}{d\rho}\mathrm{Stab}_\rho(\mathbf{1}_A)\big|_{\rho=0} = \mathbf{W}^1(\mathbf{1}_A)$. As a direct corollary of the small set expansion theorem we have the following bound on $\mathbf{W}^1(\mathbf{1}_A)$:

$$\mathbf{W}^1(\mathbf{1}_A) = \frac{d}{d\rho}\mathrm{Stab}_\rho(\mathbf{1}_A)\Big|_{\rho=0} \leq \frac{d}{d\rho}\alpha^{2/(1+\rho)}\Big|_{\rho=0} = 2\alpha^2\ln(1/\alpha).$$

We now give a self-contained proof this fact, due to Talagrand [Tal96]:

Theorem 41 (level-1 inequality) *Let $f : \{-1,1\}^n \to \{0,1\}$ and $\alpha = \mathbf{E}[f]$. Then $\mathcal{W}_1(f) = O(\alpha^2\ln(1/\alpha))$.*

Proof. First consider an arbitrary linear form $\ell(x) = \sum_{i=1}^n a_i x_i$ normalized to satisfy $\sum a_i^2 = 1$. For any $t_0 \geq 1$, we partition $x \in \{-1,1\}^n$ according to whether $|\ell(x)| < t_0$ or $|\ell(x)| \geq t_0$, and note that

$$\mathbf{E}[f(x)\ell(x)] = \mathbf{E}\left[\mathbf{1}(|\ell(x)| < t_0) \cdot f(x)\ell(x)\right] + \mathbf{E}\left[\mathbf{1}(|\ell(x)| \geq t_0) \cdot f(x)\ell(x)\right].$$

The first summand is at most $\alpha \cdot t_0$, and the second is at most

$$\int_{t_0}^{\infty} 2e^{-t^2/2}\, dt \le 2\int_{t_0}^{\infty} te^{-t^2/2}\, dt = \left[-2e^{-t^2/2}\right]_{t_0}^{\infty} = 2 \cdot e^{-t_0^2/2}$$

by Hoeffding, where the inequality holds since $t_0 \ge 1$. Choosing $t_0 = (2\ln(1/\alpha))^{1/2} \ge 1$, we get

$$\mathbf{E}[f(x)\ell(x)] \le O(\alpha\sqrt{\ln(1/\alpha)}). \tag{5}$$

Now let $\ell(x) = \frac{1}{\sigma}\sum_{i=1}^n \hat{f}(i)x_i$ where $\sigma = \sqrt{\mathbf{W}_1(f)}$ (if $\sigma = 0$ we are done). Note that

$$\mathbf{E}[f(x)\ell(x)] = \sum_{i=1}^n \hat{f}(i)\hat{\ell}(i) = \frac{1}{\sigma}\sum_{i=1}^n \hat{f}(i)^2 = \sqrt{\mathbf{W}_1(f)}.$$

The claimed inequality then follows by applying (5). ■

Proposition 42 ($\frac{2}{\pi}$ theorem) *Let $f : \{-1,1\}^n \to \{-1,1\}$ with $|\hat{f}(i)| \le \varepsilon$ for all $i \in [n]$. Then $\mathbf{W}^1(f) \le \frac{2}{\pi} + O(\varepsilon)$.*

Proof. Let $\sigma = \sqrt{\mathbf{W}^1(f)}$ and assume without loss of generality that $\sigma \ge 1/2$ (otherwise $\mathbf{W}^1(f) < 1/4 < \frac{2}{\pi}$ and we are done). Let $\ell(x) = \frac{1}{\sigma}\sum_{i=1}^n \hat{f}(i)x_i$ where $|\hat{\ell}(i)| \le 2\varepsilon$ for all $i \in [n]$. Note that $\mathbf{E}[f(x)\ell(x)] = \sigma$ and $\mathbf{E}[\ell(x)f(x)] \le \mathbf{E}[|\ell(x)|]$. Applying Berry-Esséen gives us $\mathbf{E}[|\ell(x)|] \approx_{O(\varepsilon)} \mathbf{E}[|\mathcal{G}|] = \sqrt{2/\pi}$ and this completes the proof (technically Berry-Esséen only yields a bound on closeness in cdf-distance, but this can be translated into a bound on closeness in first moments). ■

A dual to Proposition 42 holds as well: if $\mathbf{W}^1(f) \ge \frac{2}{\pi} - \varepsilon$, then f is $O(\sqrt{\varepsilon})$-close to a linear threshold function (in fact the LTF is simply $\mathrm{sgn}\big(\sum_{i=1}^n \hat{f}(i)x_i\big)$). This is the crux of the result that the class of linear threshold functions is testable with $\mathrm{poly}(1/\varepsilon)$-queries [MORS10].

2.3 Bonami's lemma

The next theorem, due to Bonami [Bon70], states that low degree multilinear polynomials of Rademachers are reasonable random variables (this is sometimes known as $(4,2)$-hypercontractivity).

Theorem 43 (Bonami) *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a multilinear polynomial of degree at most d. Then $\|f\|_4 \le \sqrt{3}^d\|f\|_2$. Equivalently, $\mathbf{E}[f(x)^4] \le 9^d \cdot \mathbf{E}[f(x)^2]^2$*

Proof. We proceed by induction on n. If $n = 0$ then f is the constant and the inequality holds trivially for all d. For the inductive step, let

$$f(x_1,\ldots,x_n) = g(x_1,\ldots,x_{n-1}) + x_n h(x_1,\ldots,x_{n-1})$$

Notice that g has degree at most d, h has degree at most $d-1$, and both are polynomials in $n-1$ variables. Notice also that the random variable x_n is independent of both g and h. Therefore, we have:

$$\begin{aligned}\mathbf{E}[f^4] &= \mathbf{E}[(g+x_nh)^4] \\ &= \mathbf{E}[g^4] + 3\,\mathbf{E}[x_n]\,\mathbf{E}[g^3h] + 6\,\mathbf{E}[x_n^2]\,\mathbf{E}[g^2h^2] + 3\,\mathbf{E}[x_n^3]\,\mathbf{E}[gh^3] + \mathbf{E}[x_n^4]\,\mathbf{E}[h^4]\end{aligned}$$

where we used independence for the second equality. Now note that $\mathbf{E}[x_n] = \mathbf{E}[x_n^3] = 0$, $\mathbf{E}[x_n^2] = \mathbf{E}[x_n^4] = 1$, and $\mathbf{E}[g^2h^2] \leq \sqrt{\mathbf{E}[g^4]}\sqrt{\mathbf{E}[h^4]}$ by Cauchy-Schwarz. Therefore,

$$\begin{aligned}\mathbf{E}[f^4] &\leq \mathbf{E}[g^4] + 6\sqrt{\mathbf{E}[g^4]}\sqrt{\mathbf{E}[h^4]} + \mathbf{E}[h^4] \\ &\leq 9^d\,\mathbf{E}[g^2]^2 + 6\sqrt{9^d\,\mathbf{E}[g^2]^2}\sqrt{9^{d-1}\,\mathbf{E}[h^2]^2} + 9^{d-1}\,\mathbf{E}[h^2]^2 \\ &= 9^d \cdot (\mathbf{E}[g^2]^2 + 2\,\mathbf{E}[g^2]\,\mathbf{E}[h^2] + \tfrac{1}{9}\,\mathbf{E}[h^2]^2) \\ &\leq 9^d \cdot (\mathbf{E}[g^2] + \mathbf{E}[h^2])^2\end{aligned}$$

To complete the proof, notice that:

$$\begin{aligned}\mathbf{E}[f^2] &= \mathbf{E}[(g+x_nh)^2] \\ &= \mathbf{E}[g^2] + 2\,\mathbf{E}[x_n]\,\mathbf{E}[gh] + \mathbf{E}[x_n^2]\,\mathbf{E}[h^2] \\ &= \mathbf{E}[g^2] + \mathbf{E}[h^2]\end{aligned}$$

and so we have shown that $\mathbf{E}[f^4] \leq 9^d \cdot \mathbf{E}[f^2]^2$. ■

3 KKL and quasirandomness

Wednesday, 29th February 2012

- Open Problem: Prove that among all functions $f : \{-1,1\}^n \to \{-1,1\}$ with $\deg(f) \leq d$, the quantity $\sum_{i=1}^n \hat{f}(i)$ is maximized by MAJ_d. Less ambitiously, show $\sum_{i=1}^n = O(\sqrt{\deg(f)})$.

3.1 Small set expansion

We begin by proving the $\rho = 1/3$ case of the small set expansion theorem: let $A \subseteq \{-1,1\}^n$ be a set of density $\alpha = |A| \cdot 2^{-n}$, and $\mathbf{1}_A : \{-1,1\}^n \to \{0,1\}$ be its indicator function. We will need the following variant of Bonami's lemma; its proof is identical to that of Theorem 43.

Theorem 44 (Bonami') *Let $f : \{-1,1\}^n \to \mathbb{R}$. Then $\|T_{1/\sqrt{3}} f\|_4 \leq \|f\|_2$.*

Theorem 44 is a special case of the hypercontractivity inequality [Bon70, Gro75, Bec75]: if $1 \leq p \leq q \leq \infty$ and $\rho \leq \sqrt{\frac{p-1}{q-1}}$, then $\|T_\rho f\|_q \leq \|f\|_p$.

Theorem 45 (SSE for $\rho = 1/3$) *Let $A \subseteq \{-1,1\}^n$. Then $\mathrm{Stab}_{1/3}(\mathbf{1}_A) \leq \alpha^{3/2}$, where α is the density of A.*

Proof. We will need a corollary of Theorem 44 that will also be useful for us when proving the KKL theorem in the next section: $\|T_{1\sqrt{3}} f\|_2^2 \leq \|f\|_{4/3}^2$. To see that this holds, we check that

$$\begin{aligned} \|T_{1/\sqrt{3}} f\|_2^2 &= \mathbf{E}[f(x)(T_{1/3}f)(x)] \\ &\leq \|f\|_{4/3} \cdot \|T_{1/3} f\|_4 \qquad (6) \\ &= \|f\|_{4/3} \cdot \|T_{1/\sqrt{3}}(T_{1/\sqrt{3}} f)\|_4 \\ &\leq \|f\|_{4/3} \cdot \|T_{1/\sqrt{3}} f\|_2. \qquad (7) \end{aligned}$$

Here (6) is by Hölder's inequality and (7) by applying Theorem 44 to $T_{1/\sqrt{3}} f$; dividing both sides by $\|T_{1/\sqrt{3}} f\|_2$ yields the claim. Applying this corollary to $f = \mathbf{1}_A$ completes the proof:

$$\mathrm{Stab}_{1/3}(\mathbf{1}_A) = \sum_{S \subseteq [n]} \left(\tfrac{1}{3}\right)^{|S|} \hat{f}(S)^2 = \left\|T_{1/\sqrt{3}} \mathbf{1}_A\right\|_2^2 \leq \mathbf{E}\left[\mathbf{1}_A(x)^{4/3}\right]^{3/2} = \alpha^{3/2}.$$

Here the second equality is an application of Parseval's, and the final uses the fact that $\mathbf{1}_A$ is $\{0,1\}$-valued. ∎

3.2 Kahn-Kalai-Linial

Theorem 46 **([KKL98])** *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *with* $\mathbf{E}[f] = 0$*, and set* $\alpha := \max_{i \in [n]} \{\mathrm{Inf}_i(f)\}$*. Then* $\mathrm{Inf}(f) = \Omega(\log(1/\alpha))$*.*

Proof. Recall that variable i is pivotal for x in f iff $(D_i f)(x) = \pm 1$, and so $\mathbf{E}[|D_i f|] = \mathrm{Inf}_i(f) \le \alpha$; the plan is to apply the small set expansion theorem to $D_i f$ and sum over all $i \in [n]$. We have shown in the proof of Theorem 45 that $\|T_{1/\sqrt{3}} g\|_2^2 \le \|g\|_{4/3}^2$, and applying it to $g = D_i f$ gives us

$$\mathrm{Stab}_{1/3}(D_i f) \le \|D_i f\|_{4/3}^2 = \mathbf{E}\left[|D_i f|^{4/3}\right]^{3/2} = \mathrm{Inf}_i(f)^{3/2}. \tag{8}$$

We sum both sides of the inequality over $i \in [n]$, starting with the left hand side. Recall that $\mathrm{Stab}_{1/3}(D_i f) = \sum_{S \subseteq [n]} \left(\frac{1}{3}\right)^{|S|} \widehat{D_i f}(S)^2 = \sum_{S \ni i} \left(\frac{1}{3}\right)^{|S|-1} \hat{f}(S)^2$, and so:

$$\begin{aligned}
\sum_{i=1}^{n} \mathrm{Stab}_{1/3}(D_i f) &= \sum_{i=1}^{n} \sum_{S \ni i} \left(\tfrac{1}{3}\right)^{|S|-1} \hat{f}(S)^2 \\
&= \sum_{|S| \ge 1} |S| \left(\tfrac{1}{3}\right)^{|S|-1} \hat{f}(S)^2 \\
&\ge \sum_{1 \le |S| \le 2\,\mathrm{Inf}(f)} |S| \left(\tfrac{1}{3}\right)^{|S|-1} \hat{f}(S)^2 \\
&\ge \sum_{1 \le |S| \le 2\,\mathrm{Inf}(f)} 2\,\mathrm{Inf}(f) \cdot \left(\tfrac{1}{3}\right)^{2\,\mathrm{Inf}(f)-1} \hat{f}(S)^2 \qquad (9) \\
&\ge 3\,\mathrm{Inf}(f) \cdot \left(\tfrac{1}{9}\right)^{\mathrm{Inf}(f)}. \qquad (10)
\end{aligned}$$

Here (9) uses the fact that $x \cdot 3^{-(x-1)}$ is a decreasing function when $x \ge 1$, and (10) uses Markov's inequality $\sum_{|S| \ge 2\,\mathrm{Inf}(f)} \hat{f}(S)^2 \le \frac{1}{2}$ along with the assumption that f is balanced. Now summing the right hand side of (8) gives us $\sum_{i=1}^{n} \mathrm{Inf}_i(f)^{3/2} \le \alpha \cdot \mathrm{Inf}(f)$. Combining both inequalities yields $3 \cdot \left(\frac{1}{9}\right)^{\mathrm{Inf}(f)} \le \alpha^{1/2}$, and therefore $\mathrm{Inf}(f) = \Omega(\log(1/\alpha))$ as claimed. ■

Corollary 47 *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *with* $\mathbf{E}[f] = 0$*. Then* $\max_{i \in [n]} \{\mathrm{Inf}_i(f)\} = \Omega\left(\frac{\log(n)}{n}\right)$*.*

This bound on the maximum influence is tight for the Ben-Or Linial TRIBES function [BL89]: the 2^k-way OR of k-way AND's of disjoint sets of variables (so $n = k \cdot 2^k$). We remark that while Corollary 47 only gives a $\log(n)$ improvement over the $1/n$ bound that follows directly from the Poincaré inequality, this factor makes a crucial difference in many applications (*e.g.* it is the crux of Khot and Vishnoi's [KV05] counter-example to the Goemans-Linial conjecture [Goe97, Lin02]).

Corollary 48 *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *be a balanced, monotone function viewed as a voting scheme. Both candidates can bias the outcome of the election in their favor to 99% probability by bribing a* $O\left(\frac{1}{\log(n)}\right)$ *fraction of voters.*

3.3 Dictator versus Quasirandom tests

Definition 49 (noisy influence) *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $\rho \in [0,1]$. The i-th ρ-noisy influence of f is* $\mathrm{Inf}_i^{(\rho)}(f) := \mathrm{Stab}_\rho(D_i f) = \sum_{S \ni i} \rho^{|S|-1} \hat{f}(S)^2$.

Note that $\mathrm{Inf}_i^{(1)}(f) = \mathrm{Inf}_i(f)$ and $\mathrm{Inf}_i^{(0)}(f) = \hat{f}(i)^2$, and intermediate values of $\rho \in (0,1)$ interpolates between these two extremes – the larger the value of ρ is, the more the weight $\hat{f}(S)^2$ on larger sets S is dampened by the attenuating factor $\rho^{|S|-1}$.

Definition 50 (quasirandom) *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $\varepsilon, \delta \in [0,1]$. We say that f is (ε,δ)-quasirandom, or f has (ε,δ)-small noisy influences, if $\mathrm{Inf}_i^{(1-\delta)}(f) \le \varepsilon$ for all $i \in [n]$.*

A few prototypical quasirandom functions are the constants ± 1 (these are $(0,0)$-quasirandom), the majority function $((O(\frac{1}{\sqrt{n}}), 0)$-quasirandom), and large parities χ_S $((1-\delta)^{|S|-1}, 0)$-quasirandom). Unbiased juntas, and dictators in particular, are prototypical examples of functions far from quasirandom. The next proposition states that even functions far from being quasirandom can only have a small number of variables with large noisy influence:

Proposition 51 *Let $f : \{-1,1\}^n \to \mathbb{R}$ with $\mathbf{Var}(f) \le 1$, and let $J = \{i \in [n] : \mathrm{Inf}_i^{(1-\delta)}(f) \ge \varepsilon\}$ be the set of coordinates with large noisy influences. Then $|J| \le 1/\varepsilon\delta$.*

Proof. We first note that

$$\varepsilon \cdot |J| \le \sum_{i \in J} \mathrm{Inf}_i^{(1-\delta)}(f) \le \sum_{i=1}^{n} \mathrm{Inf}_i^{(1-\delta)}(f) = \sum_{|S| \ge 1} |S| \cdot (1-\delta)^{|S|-1} \hat{f}(S)^2.$$

It remains to check that $|S| \cdot (1-\delta)^{|S|-1} \le 1/\delta$ for any $S \subseteq [n]$: to see this holds, note that $(1-\delta)^{|S|-1} \le (1-\delta)^{i-1}$ for any $i \le |S|$, and so summing over i from 1 to $|S|$ gives us $|S| \cdot (1-\delta)^{|S|-1} \le \sum_{i=1}^{|S|} (1-\delta)^{i-1} \le \sum_{i=1}^{\infty} (1-\delta)^{i-1} = 1/\delta$. We have shown that $\varepsilon \cdot |J| \le (1/\delta) \cdot \mathbf{Var}(f) \le 1/\delta$, and the proof is complete. ∎

Consider the problem of testing dictators: given blackbox access to a boolean function f, if f is a dictator the test accepts with probability 1, and if f is ε-far from any of the n dictators it accepts with probability $1 - \Omega(\varepsilon)$. Implicit in Kalai's proof of Arrow's impossibility theorem (Theorem 35) is a 3-query test that comes close to achieving this:

3-query NAE test

1. For each $i \in [n]$, pick (x_i, y_i, z_i) uniformly from the 6 possible NAE triples.
2. Query f on x, y, and z.
3. Accept iff $\mathsf{NAE}(f(x), f(y), f(z)) = 1$.

Recall that the NAE test accepts f with probability $\frac{3}{4} - \frac{3}{4} \cdot \mathrm{Stab}_{-1/3}(f)$, and so if $f = \mathsf{DICT}_i$ for some $i \in [n]$ (in particular, $\mathrm{Stab}_{-1/3}(f) = -1/3$) the test accepts with probability 1 (*i.e.* we have perfect completeness). However, the NAE test does not quite satisfy the soundness criterion. We saw that if the test accepts f with probability $1 - \varepsilon$ then $\mathbf{W}^1(f) \geq 1 - O(\varepsilon)$, and by a theorem of Friedgut, Kalai and Naor, f has to $O(\varepsilon)$-close to a dictator or an anti-dictator; the soundness criterion requires f to be $O(\varepsilon)$-close to a dictator. It therefore remains to rule out functions that are close to anti-dictators, and we will do this using the Blum-Luby-Rubinfeld linearity test. Recall that the BLR test is a 3-query test that accepts f with probability 1 if $f = \chi_S$ for some $S \subseteq [n]$, and with probability $1 - \Omega(\varepsilon)$ if f is ε-far from all the parity functions. Combining the BLR and NAE tests, we have a 6-query test for dictatorship with perfect completeness and soundness $1 - \Omega(\varepsilon)$. In fact it is easy to show that we can perform just one of the two tests, each with probability $\frac{1}{2}$, and reduce the query complexity to 3 while only incurring a constant factor in the rejection probability.

As we will see on Saturday, for applications to hardness of approximation (UGC-hardness in particular) it suffices to design a test that distinguishes dictators from $(\varepsilon, \varepsilon)$-quasirandom functions, instead of one that distinguishes dictators from functions ε-far from dictators.

Definition 52 (DICT vs QRAND) *Let $0 \leq s < c \leq 1$. A (c, s) dictator versus quasirandom test is defined as follows. Given blackbox access to a function $f : \{-1,1\}^n \to \{-1,1\}$,*

1. *The test makes $O(1)$ non-adaptive queries to f.*
2. *If f is a dictator, it accepts with probability at least c.*
3. *If f is $(\varepsilon, \varepsilon)$-quasirandom, it accepts with probability at most $s + o_\varepsilon(1)$.*

As we will see, often we will need to assume that f is odd (*i.e.* $f(-x) = -f(x)$ for all $x \in \{-1,1\}^n$, or equivalently, $\hat{f}(S) = 0$ for all even $|S|$). Let us consider the NAE test as a dictator versus quasirandom test, under the promise that f is odd. We have seen that the test has perfect completeness (*i.e.* $c = 1$), and now we determine the value of s. First note that since f is odd,

$$\begin{aligned}
\mathbf{Pr}[\mathsf{NAE} \text{ accepts } f] &= \tfrac{3}{4} - \tfrac{3}{4} \cdot \big(\mathbf{W}^0(f) - \tfrac{1}{3}\mathbf{W}^1(f) + \tfrac{1}{9}\mathbf{W}^2(f) - \tfrac{1}{27}\mathbf{W}^3(f) + \dots\big) \\
&= \tfrac{3}{4} - \tfrac{3}{4} \cdot \big(- \tfrac{1}{3}\mathbf{W}^1(f) - \tfrac{1}{27}\mathbf{W}^3(f) - \tfrac{1}{243}\mathbf{W}^5(f) - \dots\big) \\
&= \tfrac{3}{4} + \tfrac{3}{4} \cdot \big(\tfrac{1}{3}\mathbf{W}^1(f) + \tfrac{1}{27}\mathbf{W}^3(f) + \tfrac{1}{243}\mathbf{W}^5(f) + \dots\big) \\
&= \tfrac{3}{4} + \tfrac{3}{4} \cdot \mathrm{Stab}_{1/3}(f).
\end{aligned}$$

Now let f be a $(\varepsilon, \varepsilon)$-quasirandom function. Applying the Majority Is Stablest theorem (we will need a statement of it for functions with ε-small ε-noisy influences instead of ε-small regular influences), we have

$$\begin{aligned}
\mathbf{Pr}[\mathsf{NAE} \text{ accepts } f] &= \tfrac{3}{4} + \tfrac{3}{4} \cdot \mathrm{Stab}_{1/3}(f) \\
&\leq \tfrac{3}{4} + \tfrac{3}{4} \cdot \mathrm{Stab}_{1/3}(\mathsf{MAJ}) + o_\varepsilon(1) \\
&\to \tfrac{3}{4} + \tfrac{3}{4} \cdot \big(1 - \tfrac{2}{\pi}\arccos(\tfrac{1}{3})\big) + o_\varepsilon(1) \qquad (11) \\
&= 0.91226\ldots + o_\varepsilon(1),
\end{aligned}$$

where (11) uses the estimate we proved for $\mathrm{Stab}_\rho(\mathsf{MAJ})$ using Sheppard's formula and the central limit theorem in Section 2.1. Therefore we have shown that the NAE test is a $(1, 0.91226\ldots)$ dictator versus quasirandom test.

We consider two more examples of dictator versus quasirandom tests and compute their c and s values: the ρ-noise test of S. Khot, G. Kindler, E. Mossel and R. O'Donnell [KKMO07], and J. Håstad's $\mathsf{3XOR}_\delta$ test [Hås01].

> KKMO 2-query ρ-noise test
>
> 1. Let $\rho \in [0,1]$. Pick $x \in \{-1,1\}^n$ uniformly, and $y \sim N_\rho(x)$.
> 2. Query f on x and y.
> 3. Accept iff $f(x) = f(y)$.

First note the ρ-noise test accepts $f = \mathsf{DICT}_i$ with probability $\frac{1}{2} + \frac{1}{2}\rho$, the probability that x_i is not flipped in y. For soundness, let f be an odd $(\varepsilon, \varepsilon)$-quasirandom function and note that

$$\begin{aligned}
\mathbf{Pr}[\mathsf{KKMO} \text{ accepts } f] &= \mathbf{E}\left[\tfrac{1}{2} + \tfrac{1}{2} f(x) f(y)\right] \\
&= \tfrac{1}{2} + \tfrac{1}{2} \cdot \mathrm{Stab}_\rho(f) \\
&\leq \tfrac{1}{2} + \tfrac{1}{2} \cdot \mathrm{Stab}_\rho(\mathsf{MAJ}) + o_\varepsilon(1) \\
&\to \tfrac{1}{2} + \tfrac{1}{2} \cdot \left(1 - \tfrac{2}{\pi} \arccos(\rho)\right) + o_\varepsilon(1),
\end{aligned}$$

where once again we have used the Majority Is Stablest theorem along with Sheppard's formula (the assumption that f is odd is used in the application of the Majority Is Stablest theorem, which requires need $\mathbf{E}[f] = 0$). Different values of ρ result in different c versus s ratios; for example, if $\rho = 1/\sqrt{2}$ then $c = 0.85$ and $s = 0.75$.

For Håstad's test it will be convenient for us to view our functions as $f : \mathbb{F}_2^n \to \{-1, 1\}$. Like in the analyses of the NAE and ρ-noise tests, we will have to assume that f is odd.

> Håstad's 3-query $\mathsf{3XOR}_\delta$ test
>
> 1. Let $\delta \in [0,1]$. Pick $x, y \in \mathbb{F}_2^n$ uniformly and independently, and set $z = x + y$.
> 2. Pick $x' \sim N_{1-\delta}(x)$.
> 3. Query f on x', y, and z.
> 4. Accept iff $f(x')f(y)f(z) = 1$.

Note that this is identical to the BLR linearity test, except with the noisy x' instead of x. Once again it is easy to see that dictators pass with probability $\frac{1}{2} + \frac{1}{2}(1-\delta) = 1 - \frac{\delta}{2}$, and it remains to analyze soundness:

$$
\begin{aligned}
\mathbf{Pr}[\mathsf{3XOR}_\delta \text{ accepts } f] &= \mathop{\mathbf{E}}_{x,y,x'}\left[\tfrac{1}{2}+\tfrac{1}{2}f(x')f(y)f(z)\right] \\
&= \tfrac{1}{2}+\tfrac{1}{2}\mathop{\mathbf{E}}_{x,y}\left[\mathop{\mathbf{E}}_{x'}[f(x')]f(y)f(z)\right] \\
&= \tfrac{1}{2}+\tfrac{1}{2}\mathop{\mathbf{E}}_{x,y}\left[(T_{1-\delta}f)(x)f(y)f(z)\right] \\
&= \tfrac{1}{2}+\tfrac{1}{2}\mathop{\mathbf{E}}_{x}\left[(T_{1-\delta}f)(x)\mathop{\mathbf{E}}_{y}[f(y)f(x+y)]\right] \\
&= \tfrac{1}{2}+\tfrac{1}{2}\mathop{\mathbf{E}}_{x}\left[(T_{1-\delta}f)(x)\widehat{f*f}(x)\right] \\
&= \tfrac{1}{2}+\tfrac{1}{2}\sum_{S\subseteq[n]}\widehat{T_{1-\delta}f}(S)\widehat{f*f}(S) \;=\; \tfrac{1}{2}+\tfrac{1}{2}\sum_{S\subseteq[n]}(1-\delta)^{|S|}\hat{f}(S)^3.
\end{aligned}
$$

Note that $\sum_{S\subseteq[n]}(1-\delta)^{|S|}\hat{f}(S)^3 \le \max_{S\subseteq[n]}\{(1-\delta)^{|S|}|\hat{f}(S)|\}$ by Parseval's. Next we claim that for all $\varepsilon<\delta$, if f is $(\varepsilon,\varepsilon)$-quasirandom then $(1-\delta)^{|S|}|\hat{f}(S)|\le\sqrt{\varepsilon}$ for all S with odd cardinality (in particular, $S\neq\emptyset$). To see this, assume for the sake of contradiction that there exist an S of odd cardinality for which this inequality does not hold. Then $\sqrt{\varepsilon} < (1-\delta)^{|S|}|\hat{f}(S)| \le (1-\varepsilon)^{|S|}|\hat{f}(S)|$, and squaring both sides gives us $\varepsilon<(1-\varepsilon)^{2|S|}\hat{f}(S)^2 \le (1-\varepsilon)^{|S|-1}\hat{f}(S)^2 \le \mathrm{Inf}_i^{(1-\varepsilon)}(f)$ for all $i\in S$. Since $S\neq\emptyset$, this contradicts the assumption that f is $(\varepsilon,\varepsilon)$-quasirandom.

Since the $\mathsf{3XOR}_\delta$ test accepts $(\varepsilon,\varepsilon)$-quasirandom functions with probability at most $\frac{1}{2}+\frac{1}{2}\sqrt{\varepsilon}$ (*i.e.* $s=\frac{1}{2}$), we have shown that it is a $(1-\frac{\delta}{2},\frac{1}{2})$ dictator versus quasirandom test.

4 CSPs and hardness of approximation

Thursday, 1st March 2012

4.1 Constraint satisfaction problems

We begin by noting that function testers can be viewed more generally as string testers: the tester is given blackbox access to a string $w \in \{-1,1\}^N$ (*i.e.* the truth-table of f, so $N = 2^n$), and if w satisfies some property $P_1 \subseteq \{-1,1\}^N$ (*e.g.* dictatorship) the tester accepts with probability say at least $\frac{2}{3}$, and if w satisfies some other property $P_2 \subseteq \{-1,1\}^N$ (*e.g.* quasirandomness, far from dictatorship, *etc.*) it rejects with probability at least $\frac{2}{3}$.

We may view (non-adaptive) string testers simply as a list of instructions. For example,

$$\begin{array}{l} \text{with probability } p_1 \text{ query } w_1, w_5, w_{10} \text{ and accept iff } \phi_2(w_1, w_5, w_{10}) \\ \text{with probability } p_2 \text{ query } w_{17}, w_4, w_3 \text{ and accept iff } \phi_8(w_{17}, w_4, w_3) \\ \text{with probability } p_3 \text{ query } w_2, w_{12}, w_7 \text{ and accept iff } \phi_4(w_2, w_{12}, w_7) \\ \qquad\qquad \vdots \end{array}$$

Here $\phi_1, \phi_2, \ldots$ are predicates $\{-1,1\}^k \to \{\mathsf{T}, \mathsf{F}\}$. From this point-of-view, we see that a string tester naturally defines a weighted constraint satisfaction problem (CSP) over a domain of N boolean variables, with the predicates ϕ_i's as constraints and the associated p_i's as weights. The question of determining which string $w \in \{-1,1\}^N$ passes the test with highest probability is then equivalent to the question of finding an optimal assignment that satisfies the largest weighted fraction of predicates.

Under this correspondence, an explicit (c, s) dictator versus quasirandom test for functions $f : \{-1,1\}^\ell \to \{-1,1\}$ defines an explicit instance of a weighted CSP over $L = 2^\ell$ boolean variables. Since all ℓ dictators pass with probability at least c, there are ℓ special assignments each of which satisfy at least a c weighted fraction of constraints. Furthermore, since all quasirandom functions pass the test with probability at most $s + o(1)$, any CSP assignment which is, roughly speaking, "very unlike" the ℓ special assignments will satisfy at most a $s + o(1)$ fraction of constraints. In other words, any CSP assignment that satisfies at least an $s + \Omega(1)$ fraction of constraints must be at least "slightly suggestive" of at least one of the ℓ special assignments (we say that f is suggestive of the i-th coordinate if $\mathrm{Inf}_i^{(1-\varepsilon)}(f) > \varepsilon$, and in particular, an $(\varepsilon, \varepsilon)$-quasirandom function suggests none of its coordinates).

On Saturday Per will prove the following theorem establishing a formal connection between dictator versus quasirandom tests, the UNIQUE-LABEL-COVER problem, and the hardness of approximating certain CSPs [Kho02, KR03, KKMO07, Aus08]:

Theorem 53 *Suppose there is an explicit (c, s) dictator versus quasirandom test that uses predicates $\phi_1, \ldots, \phi_r$. For every $\varepsilon > 0$ there exists a polynomial-time reduction where:*

UNIQUE-LABEL-COVER	$\longrightarrow$	*CSP with constraints* $\phi_1, \ldots, \phi_r$
YES *instance*	$\longrightarrow$	*there exists an assignment that satisfies a* $(c - \varepsilon)$ *fraction of constraints*
NO *instance*	$\longrightarrow$	*every assignment satisfies at most an* $(s + \varepsilon)$ *fraction of constraints*

The Unique Games Conjecture [Kho02] asserts that approximating the UNIQUE-LABEL-COVER problem is NP-hard. Theorem 53 therefore says that assuming the UGC, for any constant $\varepsilon > 0$ an explicit (c, s) dictator versus quasirandom test implies the NP-hardness of $((s/c) + \varepsilon)$-factor approximating CSPs with constraints corresponding to the predicates used by the test.

Recall that our analyses of all three dictator versus quasirandom tests we have seen so far depend on the promise that f is odd (in particular, note that the $(0, 0)$-quasirandom constant function $f \equiv 1$ pass both the KKMO ρ-noise test and Håstad's $3\mathsf{XOR}_\delta$ test with probability 1). One way to elide this assumption is to view them as testers for general functions $g : \{-1, 1\}^{\ell-1} \to \{-1, 1\}$ (corresponding to half of the truth table of an odd function $f : \{-1, 1\}^\ell \to \{-1, 1\}$, say for the inputs with $x_i = 1$) where the respective predicates allow literals instead of just variables. Although the string $w = 1^N$ (*i.e.* $f \equiv 1$) trivially satisfies any CSP with constraints of the form $w_i = w_j$ (*i.e.* the KKMO ρ-noise test) or $w_i w_j w_k = 1$ (*i.e.* the Håstad $3\mathsf{XOR}_\delta$ test), the same is no longer true if literals are allowed in the constraints. Given this, we may apply Theorem 53 to the NAE test, KKMO ρ-noise test, and Håstad's $3\mathsf{XOR}_\delta$ test to conclude that under the UGC,

- Approximating MAX-3NAE-SAT to a factor of $0.91226\ldots + \varepsilon$ is NP-hard.
- Approximating MAX-2LIN to a factor of $(1 - \frac{1}{\pi}\arccos(\rho))/(\frac{1}{2} + \frac{1}{2}\rho) + \varepsilon$ is NP-hard.
- Approximating MAX-3LIN to a factor of $(\frac{1}{2} + \varepsilon)$ is NP-hard.

4.2 Berry-Esséen

In this section we prove the Berry-Esséen theorem [Ber41, Ess42], a finitary version of the central limit theorem with explicit error bounds. Actually we will give a proof that only yields a polynomially weaker error bound, the upshot being that the proof is relatively simple and can be easily generalized to other settings (as we will see tomorrow, the Mossel-O'Donnell-Olezkiewicz proof of the invariance principle, an extension of the Berry-Esséen theorem to low-degree polynomials, is very similar in spirit). We will need Taylor's theorem:

Lemma 54 (Taylor) *Let ψ be a smooth function and $r \in \mathbb{N}$. For all $x \in \mathbb{R}$ and $\varepsilon > 0$ there exists a $y \in [x, x + \varepsilon]$ such that*

$$\psi(x+\varepsilon) = \psi(x) + \varepsilon\psi^{(1)}(x) + \tfrac{1}{2!} \cdot \varepsilon^2\psi^{(2)}(x) + \ldots + \tfrac{1}{(r-1)!} \cdot \varepsilon^{r-1}\psi^{(r-1)}(x) + \tfrac{1}{r!} \cdot \varepsilon^r\psi^{(r)}(y).$$

In particular, $\psi(x+\varepsilon) = \psi(x) + \varepsilon \cdot \psi^{(1)}(x) + \frac{1}{2} \cdot \varepsilon^2\psi^{(2)}(x) + \frac{1}{6} \cdot \varepsilon^3\psi^{(3)}(x) + \mathrm{error}$*, where the error term has magnitude at most* $\|\psi^{(4)}\|_\infty \cdot \varepsilon^4/24$.

Proposition 55 (hybrid argument) *Let $X_1, \ldots, X_n$ be independent random variables satisfying $\mathbf{E}[X_i] = 0$. Let $\sigma_i^2 := \mathbf{E}[X_i^2]$ and suppose $\sum_{i=1}^n \sigma_i^2 = 1$. Let $\mathbf{X} = \sum_{i=1}^n X_i$, $\mathcal{G} \sim N(0,1)$ and $\psi : \mathbb{R} \to \mathbb{R}$. Then*

$$|\,\mathbf{E}[\psi(\mathbf{X})] - \mathbf{E}[\psi(\mathcal{G})]| = O\Big(\|\psi^{(4)}\|_\infty \sum_{i=1}^n \mathbf{E}[X_i^4]\Big).$$

Note that if each X_i is B-reasonable then $\sum_{i=1}^n \mathbf{E}[X_i^4] \leq B \cdot \sum_{i=1}^n \sigma_i^4 \leq B \cdot \max\left\{\sigma_i^2\right\}$.

Proof. We will view $\mathcal{G}$ as the sum of independent Gaussians $\mathcal{G}_1 + \ldots + \mathcal{G}_n$, where each $\mathcal{G}_i \sim N(0, \sigma_i^2)$. The proof proceeds by a hybrid argument, showing that only a small error is introduced whenever each X_i in $\mathbf{X}$ is replaced by the corresponding Gaussian. More precisely, for each $i = 0, \ldots, n$, we define the random variable $\mathbf{Z}_i := \mathcal{G}_1 + \ldots + \mathcal{G}_i + X_{i+1} + \ldots + X_n$; these $n+1$ random variables interpolate between $\mathbf{Z}_0 = \mathbf{X}$ and $\mathbf{Z}_n = \mathcal{G}$. We will prove the inequality

$$|\,\mathbf{E}[\psi(\mathbf{Z}_{i-1})] - \mathbf{E}[\psi(\mathbf{Z}_i)]| = O(\|\psi^{(4)}\|_\infty \cdot \mathbf{E}[X_i^4]).$$

for all $i \in [n]$, noting that this implies the theorem by the triangle inequality. Fix $i \in [n]$ and define the random variable $\mathbf{R} := \mathcal{G}_1 + \ldots + \mathcal{G}_{i-1} + X_{i+1} + \ldots + X_n$, so $\mathbf{Z}_{i-1} = \mathbf{R} + X_i$ and $\mathbf{Z}_i = \mathbf{R} + \mathcal{G}_i$. Our goal is therefore to bound $|\,\mathbf{E}[\psi(\mathbf{R} + \sigma_i \cdot X_i)] - \mathbf{E}[\psi(\mathbf{R} + \sigma_i \cdot \mathcal{G}_i)]|$. Applying Taylor's theorem twice, we get

$$\begin{aligned}
|\,\mathbf{E}[\psi(\mathbf{Z}_{i-1})] - \mathbf{E}[\psi(\mathbf{Z}_i)]| &= \Big|\mathbf{E}\left[\psi(\mathbf{R}) + X_i \cdot \psi^{(1)}(\mathbf{R}) + \tfrac{1}{2} X_i^2 \cdot \psi^{(2)}(\mathbf{R}) + \tfrac{1}{6} X_i^3 \cdot \psi^{(3)}(\mathbf{R}) + \mathrm{error}_1\right] \\
&- \mathbf{E}\left[\psi(\mathbf{R}) + \mathcal{G}_i \cdot \psi^{(1)}(\mathbf{R}) + \tfrac{1}{2} \mathcal{G}_i^2 \cdot \psi^{(2)}(\mathbf{R}) + \tfrac{1}{6} \mathcal{G}_i^3 \cdot \psi^{(3)}(\mathbf{R}) + \mathrm{error}_2\right]\Big| \\
&= |\,\mathbf{E}[\mathrm{error}_1 - \mathrm{error}_2]|.
\end{aligned}$$

Here we have used that fact that $\mathbf{R}$ is independent of X_i and $\mathcal{G}_i$, along with the assumption that X_i and $\mathcal{G}_i$ have matching first and second moments. Substituting bounds on the error terms error_1 and error_2 given by Taylor's theorem, we complete the proof:

$$|\,\mathbf{E}[\mathrm{error}_1 - \mathrm{error}_2]| \leq \mathbf{E}\left[\frac{\|\psi^{(4)}\|_\infty \cdot X_i^4}{24} + \frac{\|\psi^{(4)}\|_\infty \cdot \mathcal{G}_i^4}{24}\right] = O(\|\psi^{(4)}\|_\infty \cdot \mathbf{E}[X_i^4]).$$

■

The same proof can be rewritten to show that if $\mathbf{Y} = Y_1 + \ldots + Y_n$ is the sum of independent random variables satisfying $\mathbf{E}[X_i] = \mathbf{E}[Y_i]$, $\mathbf{E}[X_i^2] = \mathbf{E}[X_i^2]$, and $\mathbf{E}[X_i^3] = \mathbf{E}[X_i^3]$ (the matching moments property), then $|\,\mathbf{E}[\psi(\mathbf{X}) - \psi(\mathbf{Y})]| \leq \|\psi^{(4)}\|_\infty \cdot \sum_{i=1}^n \mathbf{E}[X_i^4] + \mathbf{E}[Y_i^4]$.

Let $\psi_t : \mathbb{R} \to \mathbb{R}$ be the threshold function that takes value 1 if $x < t$, and 0 otherwise. Of course the 4^{th} derivative of this function is not uniformly bounded, but note that if it were the Berry-Esséen theorem would follow as an immediate corollary of Proposition 55. Instead, we will use the fact that ψ_t is well-approximated by a function which does have a uniformly bounded 4^{th} derivative, which then implies a slightly weaker version of the Berry-Esséen theorem.

Lemma 56 (smooth approximators of thresholds) *Let $t \in \mathbb{R}$ and $0 < \lambda < 1$. There exists a function $\psi_{t,\lambda} : \mathbb{R} \to \mathbb{R}$ with $\|\psi^{(4)}_{t,\lambda}\|_\infty = O(1/\lambda^4)$ that approximates ψ_t in the following sense:*

1. *$\psi_{t,\lambda}(x) = \psi_t(x) = 1$ if $x < t - \lambda$.*
2. *$\psi_{t,\lambda}(x) \in [0,1]$ if $x \in [t-\lambda, t+\lambda]$.*
3. *$\psi_{t,\lambda}(x) = \psi_t(x) = 0$ if $x > t + \lambda$.*

We are now ready to prove a weak version of the Berry-Esséen theorem.

Proposition 57 (weak Berry-Esséen) *Let $X_1, \dots, X_n$ be independent, B-reasonable random variables satisfying $\mathbf{E}[X_i] = 0$. Let $\sigma_i^2 := \mathbf{E}[X_i^2]$, $\tau := \max\left\{\sigma_i^2\right\}$, and suppose $\sum_{i=1}^n \sigma_i^2 = 1$. Let $S = X_1 + \ldots + X_n$ and $\mathcal{G} \sim N(0,1)$. For all $t \in \mathbb{R}$,*

$$|\,\mathbf{Pr}[S \le t] - \mathbf{Pr}[\mathcal{G} \le t]| \le O((B\tau)^{1/5}).$$

Proof. Since $\psi_{t+\lambda,\lambda}(x) = 1$ for all $x < t$ we have $\mathbf{Pr}[S \le t] \le \mathbf{E}[\psi_{t+\lambda,\lambda}(S)]$. Now using the fact that $\|\psi^{(4)}_{t+\lambda,\lambda}\|_\infty = O(1/\lambda^4)$, we apply Proposition 55 to get

$$\mathbf{E}[\psi_{t+\lambda,\lambda}(S)] = \mathbf{E}[\psi_{t+\lambda,\lambda}(\mathcal{G})] \pm O(B\tau/\lambda^4)$$

Since $\psi_{t+\lambda,\lambda}(x)$ is at most 1 for all $x \le t + 2\lambda$ and 0 otherwise, we have

$$\mathbf{E}[\psi_{t+\lambda,\lambda}(\mathcal{G})] \le \mathbf{Pr}[\mathcal{G} < t + 2\lambda] = \mathbf{Pr}[\mathcal{G} < t] + O(\lambda),$$

and so combining both error bounds gives us $\mathbf{E}[\psi_{t+\lambda,\lambda}(S)] \le \mathbf{Pr}[\mathcal{G} < t] + O(B\tau/\lambda^4) + O(\lambda)$. Arguing symmetrically for $\psi_{t-\lambda,\lambda}(x)$ gives us $|\,\mathbf{Pr}[S < t] - \mathbf{Pr}[\mathcal{G} < t]| = O(B\tau/\lambda^4) + O(\lambda)$, and taking $\lambda = (B\tau)^{1/5}$ yields the claim. ∎

5 Majority Is Stablest

Friday, 2nd March 2012

Our definition of ρ-correlated Gaussians (Definition 39) extend naturally to higher dimensions: let $\vec{\mathcal{G}}$ and $\vec{\mathcal{G}}'$ be independent standard n-dimensional Gaussians (*i.e.* $\vec{\mathcal{G}} = (\mathcal{G}_1, \ldots, \mathcal{G}_n)$ where each $\mathcal{G}_i \sim N(0,1)$ is an independent standard Gaussian, and similarly for $\vec{\mathcal{G}}'$). Then $\vec{\mathcal{G}}$ and $\vec{\mathcal{H}} := \rho \cdot \vec{\mathcal{G}} + \sqrt{1-\rho^2} \cdot \vec{\mathcal{G}}'$ are ρ-correlated Gaussians. Just like in the one-dimension case, we have $\mathbf{E}[\mathcal{G}_i \mathcal{H}_i] = \rho$ for all $i \in [n]$.

Definition 58 (Gaussian noise stability) *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\rho \in [-1,1]$. The Gaussian noise stability of f at noise rate ρ is*

$$\mathrm{GStab}_\rho(f) := \mathbf{E}[f(\vec{\mathcal{G}})f(\vec{\mathcal{H}})], \quad \textit{where } \vec{\mathcal{G}}, \vec{\mathcal{H}} \textit{ are } \rho\textit{-correlated Gaussians.}$$

We begin by showing that $\mathrm{GStab}_\rho(f) = \mathrm{Stab}_\rho(f)$ for multilinear polynomials $f : \mathbb{R}^n \to \mathbb{R}$. To see this, let $f(x) = \sum_{S \subseteq [n]} c_S \prod_{i \in S} x_i$ and note that

$$\mathbf{E}\Big[\Big(\sum_{S\subseteq[n]} c_S \prod_{i\in S} \vec{\mathcal{G}}_i\Big)\Big(\sum_{T\subseteq[n]} c_T \prod_{i\in T} \vec{\mathcal{H}}_i\Big)\Big] = \sum_{S,T\subseteq[n]} c_S c_T \,\mathbf{E}\Big[\prod_{i\in S} \mathcal{G}_i \prod_{i \in T} \mathcal{H}_i\Big] = \sum_{S\subseteq[n]} c_S^2 \,\mathbf{E}\Big[\prod_{i\in S} \mathcal{G}_i \mathcal{H}_i\Big],$$

where we have used the independence of $\mathcal{G}_i$ from $\mathcal{G}_j, \mathcal{H}_j$ for $j \neq i$, along with the fact that $\mathbf{E}[\mathcal{H}_i] = \mathbf{E}[\mathcal{G}_i] = 0$. Now again by independence and the fact that $\mathbf{E}[\mathcal{G}_i \mathcal{H}_i] = \rho$ we conclude that $\mathrm{GStab}_\rho(f) = \sum_{S\subseteq[n]} \rho^{|S|} c_S^2$, which agrees with the formula for $\mathrm{Stab}_\rho(f)$ we derived in Proposition 31.

5.1 Borell's isoperimetric inequality

Theorem 59 ([Bor85]) *Let $f : \mathbb{R}^n \to \{-1,1\}$ with $\mathbf{E}[f(\vec{\mathcal{G}})] = 0$, where $\vec{\mathcal{G}}$ is a standard n-dimensional Gaussian. Let $0 \le \rho \le 1$. Then $\mathrm{GStab}_\rho(f) \le 1 - \frac{2}{\pi}\arccos(\rho)$.*

In this section we present Kindler and O'Donnell's recent simple proof of (a special case of) Borell's theorem [KO12]. We first introduce a few definitions and give a geometric interpretation of the theorem as an isoperimetric inequality in multidimensional Gaussian space.

Definition 60 (rotation sensitivity) *Let $f : \mathbb{R}^n \to \{-1,1\}$ and $\delta \in [0,\pi]$. The rotation sensitivity of f at δ is defined to be $\mathrm{RS}_f(\delta) := \mathbf{Pr}[f(\vec{\mathcal{G}}) \neq f(\vec{\mathcal{H}})]$, where $\vec{\mathcal{G}}$ and $\vec{\mathcal{H}}$ are $\cos(\delta)$-correlated Gaussians.*

Recall that $\vec{\mathcal{G}}$ and $\vec{\mathcal{H}}$ are $\cos(\delta)$-correlated if $\mathcal{H} = \cos(\delta) \cdot \vec{\mathcal{G}} + \sin(\delta) \cdot \vec{\mathcal{G}}'$ where $\vec{\mathcal{G}}'$ is a standard n-dimensional Gaussian independent of $\vec{\mathcal{G}}$; we typically we think of δ as small, so $\cos(\delta) \approx 1 - \frac{1}{2}\delta^2$ is close to 1 and $\sin(\delta) \approx \delta$ is a small quantity. If we view $f : \mathbb{R}^n \to \{-1,1\}$ as the indicator of a subset $\mathbf{1}_f$ of $\mathbb{R}^n$, the quantity $\mathrm{RS}_f(\delta)$ measures the probability that the set $\mathbf{1}_f$ separates a random Gaussian vector $\mathcal{G}$ from a noisy copy of it (roughly speaking,

$\mathcal{G}$ with $\sin(\delta) \cdot \vec{\mathcal{G}}'$ of noise added). Therefore we may think of the rotation sensitivity of f as a measure of the boundary size of $\mathbf{1}_f$. Indeed, Kindler and O'Donnell show that $\limsup_{\delta \to 0^+} \mathrm{RS}_f(\delta)/\delta$ is within a constant factor of the traditional definition of the Guassian surface area of $\mathbf{1}_f$ for "sufficiently nice" sets $\mathbf{1}_f$.

Since $\mathrm{RS}_f(\delta) = \mathbf{Pr}[f(\vec{\mathcal{G}}) \neq f(\vec{\mathcal{H}})] = \frac{1}{2} - \frac{1}{2} \cdot \mathrm{GStab}_{\cos(\delta)}(f)$, Borell's theorem can be equivalently stated as $\mathrm{RS}_f(\delta) \geq \frac{\delta}{\pi}$ for functions f with $\mathbf{E}[f(\vec{\mathcal{G}})] = 0$; it gives a lower bound on the boundary size of sets $\mathbf{1}_f$ with Gaussian volume $\frac{1}{2}$. Sheppard's formula tells us that if $f = \mathrm{sgn}(a_1 x_1 + \ldots + a_n x_n)$ then $\mathbf{Pr}[f(\vec{\mathcal{G}}) \neq f(\vec{\mathcal{H}})] = \frac{1}{\pi} \arccos(\cos(\delta)) = \frac{\delta}{\pi}$, and so Borell's inequality is tight when $\mathbf{1}_f$ is any halfspace defined by a hyperplane passing through the origin.

Kindler and O'Donnell prove Borell's theorem for $\delta = \frac{\pi}{2\ell}$ where $\ell \in \mathbb{N}$ (we may assume $\ell \geq 2$ since the inequality is trivially true for uncorrelated Gaussians). As a warm-up, we consider the case of $\delta = \frac{\pi}{4}$ (*i.e.* $\ell = 2$). Let $f : \mathbb{R}^n \to \{-1, 1\}$ with $\mathbf{E}[f(\vec{\mathcal{G}})] = 0$ and note that our goal is to prove $\mathrm{RS}_f(\frac{\pi}{4}) \geq \frac{1}{4}$. Let $\vec{\mathcal{G}}$ and $\vec{\mathcal{G}}'$ be independent standard n-dimensional Gaussians and set $\vec{\mathcal{H}} = \frac{1}{\sqrt{2}} \cdot \vec{\mathcal{G}} + \frac{1}{\sqrt{2}} \cdot \vec{\mathcal{G}}'$. Note that $(\vec{\mathcal{G}}, \vec{\mathcal{H}})$ and $(\vec{\mathcal{G}}', \vec{\mathcal{H}})$ are both $(\frac{1}{\sqrt{2}})$-correlated Gaussians, and so we have

$$\mathbf{Pr}[f(\vec{\mathcal{G}}) \neq f(\vec{\mathcal{H}})] + \mathbf{Pr}[f(\vec{\mathcal{H}}) + f(\vec{\mathcal{G}}')] = 2 \cdot \mathrm{RS}_f(\tfrac{\pi}{4}).$$

By a union bound, the quantity on the left hand side is at most $\mathbf{Pr}[f(\vec{\mathcal{G}}) \neq f(\vec{\mathcal{G}}')]$, and since f is balanced this probability is exactly $\frac{1}{2}$ and the proof is complete. We remark that an identical proof can be carried out for bounded functions $f : \mathbb{R}^n \to [-1, 1]$, with $\mathrm{RS}_f(\delta) := \mathbf{E}[\frac{1}{2} - \frac{1}{2} f(\vec{\mathcal{G}}) f(\vec{\mathcal{H}})]$ as the generalized definition of rotation sensitivity; all the steps remain the same, except that the inequality $(\frac{1}{2} - \frac{1}{2}ac) \leq (\frac{1}{2} - \frac{1}{2}ab) + (\frac{1}{2} - \frac{1}{2}bc)$ for all $a, b, c \in [-1, 1]$ will be used in place of the union bound.

Theorem 61 ([KO12]) *Let $f : \mathbb{R}^n \to \{-1, 1\}$ with $\mathbf{E}[f(\vec{\mathcal{G}})] = 0$, where $\vec{\mathcal{G}}$ is a standard n-dimensional Gaussian. Let $\delta = \frac{\pi}{2\ell}$ for some $\ell \in \mathbb{N}$, $\ell \geq 2$. Then $\mathrm{RS}_f(\delta) \geq \frac{1}{2\ell}$.*

Proof. Let $\vec{\mathcal{G}}$ and $\vec{\mathcal{G}}'$ be independent standard n-dimensional Gaussians. For $j = 0, \ldots, \ell$, we define the hybrid random variable $\vec{\mathcal{G}}^{(j)} := \cos(j\delta) \cdot \vec{\mathcal{G}} + \sin(j\delta) \cdot \vec{\mathcal{G}}'$, and note that they interpolate between $\vec{\mathcal{G}}^{(0)} = \vec{\mathcal{G}}$ and $\vec{\mathcal{G}}^{(\ell)} = \vec{\mathcal{G}}'$. Next, we claim that $\vec{\mathcal{G}}^{(j-1)}$ and $\vec{\mathcal{G}}^{(j)}$ are $\cos(\delta)$-correlated for every $j \in [\ell]$. Since the n coordinates are independent, it suffices to consider the first coordinates of $\vec{\mathcal{G}}_1^{(j-1)}$ and $\vec{\mathcal{G}}_1^{(j)}$ and show that they are $\cos(\delta)$-correlated. We write $\mathcal{G}^{(j-1)}$ for $\vec{\mathcal{G}}_1^{(j-1)} = \cos((j-1)\delta) \cdot \mathcal{G} + \sin((j-1)\delta) \cdot \mathcal{G}'$ and $\mathcal{G}^{(j)}$ for $\vec{\mathcal{G}}_1^{(j)} = \cos(j\delta) \cdot \mathcal{G} + \sin(j\delta) \cdot \mathcal{G}'$, and expand the expectation of their product to check that

$$\begin{aligned}
\mathbf{E}\left[\mathcal{G}^{(j-1)}\mathcal{G}^{(j)}\right] &= \cos((j-1)\delta)\cos(j\delta)\,\mathbf{E}[\mathcal{G}\mathcal{G}] \\
&\quad + \cos((j-1)\delta)\sin(j\delta)\,\mathbf{E}[\mathcal{G}\mathcal{G}'] \\
&\quad + \sin((j-1)\delta)\cos(j\delta)\,\mathbf{E}[\mathcal{G}'\mathcal{G}] \\
&\quad + \sin((j-1)\delta)\sin(j\delta)\,\mathbf{E}[\mathcal{G}'\mathcal{G}'] \\
&= \cos((j-1)\delta)\cos(j\delta) + \sin((j-1)\delta)\sin(j\delta) \\
&= \cos((j-1)\delta - j\delta) \;=\; \cos(\delta).
\end{aligned}$$

Here we have used the trigonometric identity $\cos(\theta)cos(\theta') + \sin(\theta)\sin(\theta') = \cos(\theta - \theta')$. Since $\vec{\mathcal{G}}^{(j-1)}$ and $\vec{\mathcal{G}}^{(j)}$ are $\cos(\delta)$-correlated, we apply the union bound to get $\frac{1}{2} = \mathbf{Pr}[f(\vec{\mathcal{G}}) \neq f(\vec{\mathcal{G}}')] \leq \sum_{i=1}^{\ell} \mathbf{Pr}[f(\vec{\mathcal{G}}^{(\ell-1)}) \neq f(\vec{\mathcal{G}}^{(\ell)})] = \ell \cdot \mathrm{RS}_f(\delta)$, and the proof is complete. ■

5.2 Proof outline of MIST

In this section we sketch the proof of the Majority Is Stablest theorem (MIST):

Theorem 62 ([KKMO07, MOO10]) *Let $f : \{-1,1\}^n \to \{-1,1\}$ and $\varepsilon, \rho > 0$. Suppose $\mathbf{E}[f] = 0$ and $\mathrm{Inf}_i(f) \leq \varepsilon$ for all $i \in [n]$. Then $\mathrm{Stab}_\rho(f) \leq 1 - \frac{2}{\pi}\arccos(\rho) + o_\varepsilon(1)$.*

Step 1. First consider $T_{1-\gamma}f$ for some small $\gamma > 0$. Note that

(a) $\mathrm{Inf}_i(T_{1-\gamma}f) \leq \mathrm{Inf}_i(f) \leq \varepsilon$ for all $i \in [n]$.

(b) $T_{1-\gamma}f$ is bounded since $T_{1-\gamma}$ is an averaging operator.

(c) $\mathrm{Stab}_\rho(f) = \mathrm{Stab}_{\rho'}(T_{1-\gamma}f)$ where $\rho' := \frac{\rho}{(1-\gamma)^2}$.

Step 2. Next we truncate $T_{1-\gamma}f$ to get $g := (T_{1-\gamma}f)^{\leq k} = \sum_{|S|\leq k}(1-\gamma)^{|S|}\hat{f}(S)\chi_S$, where $k := \mathrm{poly}(\frac{1}{\gamma})$. Note that g has low-degree but it may no longer bounded. Nevertheless we can say that

$$\|(T_{1-\gamma}f) - g\|_2^2 = \sum_{|S|>k}(1-\gamma)^{2\cdot|S|}\hat{f}(S)^2 \leq (1-\gamma)^{2k}\sum_{|S|>k}\hat{f}(S)^2 \leq \gamma,$$

and since g is bounded this implies $\mathbf{E}[\mathsf{sqdist}_{[-1,1]}(g(X_1,\ldots,X_n))] \leq \gamma$. For the same reason, $\mathrm{Stab}_{\rho'}(T_{1-\gamma}f) \leq \mathrm{Stab}_{\rho'}(g) + \gamma$. Here $\mathsf{sqdist}_{[-1,1]} : \mathbb{R} \to \mathbb{R}^{\geq 0}$ is the function that gives the squared distance to the interval $[-1,1]$; *i.e.* $\mathsf{sqdist}_{[-1,1]}(t) = 0$ if $t \in [-1,1]$, and $(|t|-1)^2$ otherwise.

Step 3. We apply the invariance principle (an extension of the Berry-Esséen theorem to low-degree multilinear polynomials, proved in the next section) to g and the test function $\mathsf{sqdist}_{[-1,1]}$ to bound

$$\mathbf{E}[\mathsf{sqdist}_{[-1,1]}(g(\vec{\mathcal{G}}))] \leq \mathbf{E}[\mathsf{sqdist}_{[-1,1]}(g(\vec{X}))] + \mathrm{poly}(2^k, \varepsilon) = \gamma + \mathrm{poly}(2^k, \varepsilon).$$

We are omitting a few details here since the invariance principle requires test functions to have uniformly bounded 4^{th} derivatives, just like in Berry-Esséen, so we actually need a smooth approximation of $\mathsf{sqdist}_{[-1,1]}$.

Step 4. Finally we consider $g' : \mathbb{R}^n \to [-1,1]$, the truncation of g to the interval $[-1,1]$; *i.e.* $g'(u) = g(u)$ if $g(u) \in [-1,1]$, and $\mathrm{sgn}(g(u))$ otherwise. We have $\mathrm{Stab}_{\rho'}(g) = \mathrm{GStab}_{\rho'}(g)$ since g is multilinear, and $\mathrm{GStab}_{\rho'}(g') \leq 1 - \frac{2}{\pi}\arccos(\rho')$ by Borell's theorem (again we are eliding some details here since g' may not satisfy $\mathbf{E}[g'(\mathcal{G})] = 0$). It remains to argue that $\mathrm{GStab}_{\rho'}(g) \approx \mathrm{GStab}_{\rho'}(g')$, and for this we need to define the Ornstein-Uhlenbeck operator U_ρ, the Gaussian analogue of the noise operator T_ρ:

Definition 63 (Ornstein-Uhlenbeck) *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\rho \in [-1, 1]$. The Ornstein-Uhlenbeck operator U_ρ acts of f as follows: $(U_\rho f)(x) := \mathbf{E}[f(\rho \cdot x + \sqrt{1-\rho^2} \cdot \vec{\mathcal{G}})]$, where $\vec{\mathcal{G}}$ is a standard n-dimensional Gaussian.*

With this definition, we may express Gaussian noise stability as $\mathrm{GStab}_\rho(f) = \mathbf{E}[f(\vec{\mathcal{G}})U_\rho(\vec{\mathcal{G}})]$ and compute

$$
\begin{aligned}
|\mathrm{GStab}_{\rho'}(g) - \mathrm{GStab}_{\rho'}(g')| &= |\,\mathbf{E}[g \cdot U_\rho g - g' \cdot U_\rho g']| \\
&\leq |\,\mathbf{E}[g \cdot U_\rho g - g' \cdot U_\rho g]| + |\,\mathbf{E}[g' \cdot U_\rho g - g' \cdot U_\rho g']| \\
&= |\,\mathbf{E}[(g - g') \cdot U_\rho g]| + |\,\mathbf{E}[(g - g') \cdot U_\rho g']| && (12) \\
&= \big(\,\mathbf{E}[(g-g')^2]\big)^{1/2}\big(\,\mathbf{E}[(U_\rho g)^2]\big)^{1/2} \\
&\quad + \big(\,\mathbf{E}[(g-g')^2]\big)^{1/2}\big(\,\mathbf{E}[(U_\rho g')^2]\big)^{1/2} && (13) \\
&\leq 2\,\big(\,\mathbf{E}[(g-g')^2]\big)^{1/2}. && (14)
\end{aligned}
$$

Here (12) holds since $\mathbf{E}[g' \cdot U_\rho g] = \mathbf{E}[U_\rho g' \cdot g]$, (13) is an application of Cauchy-Schwarz, and (14) uses the fact that U_ρ is a contraction on L^2. Finally we note that $\mathbf{E}[(g - g')^2]$ is simply $\mathbf{E}[\mathsf{sqdist}_{[-1,1]}(g)]$ and the proof is complete.

5.3 The invariance principle

In this section we prove (a special case of) the Mossel-O'Donnell-Olezkiewicz invariance principle [MOO10] for multilinear polynomials with low influences and bounded degree; in full generality the principle states that the distribution of such polynomials is essentially invariant for all product spaces. The crux of the proof is a low-degree analogue of Proposition 55; once again we proceed by a hybrid argument, showing that a small error is introduced whenever we replace a Rademacher random variable with a standard Gaussian. This is sometimes known as the Lindeberg replacement trick, first appearing in Lindeberg's proof of the central limit theorem [Lin22]. There has been other work generalizing Lindeberg's argument to the non-linear case [Rot75, Rot79, Cha06], but these results either yield weaker error bounds or require stronger conditions (*e.g.* worst-case influences rather than average-case).

Proposition 64 (hybrid argument) *Let Q be a degree-d multilinear polynomial $Q(u) = \sum_{|S|\leq d} c_S \prod_{i\in S} u_i$ and assume:*

1. *The coefficients $c_S \in \mathbb{R}$ are normalized to satisfy $\sum_{S\neq\emptyset} c_S^2 = 1$.*
2. *We write τ_i to denote $\mathrm{Inf}_i(Q) = \sum_{S\ni i} c_S^2$, and let $\tau = \max_{i\in[n]} \mathrm{Inf}_i(Q)$.*
3. *$\psi : \mathbb{R} \to \mathbb{R}$ is a function satisfying $|\psi^{(4)}(x)| \leq C$ for all $x \in \mathbb{R}$.*
4. *$\mathbf{X} = Q(X_1, \ldots, X_n)$ and $\mathbf{Y} = Q(\mathcal{G}_1, \ldots, \mathcal{G}_n)$ where $X_1, \ldots, X_n$ are independent Rademachers and $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are independent standard Gaussians.*

Then $|\,\mathbf{E}[\psi(\mathbf{X})] - \mathbf{E}[\psi(\mathbf{Y})]| \leq d \cdot 9^d \cdot C \cdot \tau$.

Proof. We first define a sequence of hybrid random variables that interpolate between $\mathbf{X}$ and $\mathbf{Y}$. For each $i = 0, \ldots, n$ we define the random variable $\mathbf{Z}_i = Q(\mathcal{G}_1, \ldots, \mathcal{G}_i, X_{i+1}, \ldots, X_n)$, and note that $\mathbf{Z}_0 = \mathbf{X}$ and $\mathbf{Z}_n = \mathbf{Y}$. As before, it suffices to prove

$$|\,\mathbf{E}[\psi(\mathbf{Z}_{i-1})] - \mathbf{E}[\psi(\mathbf{Z}_i)]| \leq C \cdot 9^d \cdot \tau_i^2 \tag{15}$$

for all $i \in [n]$. Note that the overall claim follows from the above by telescoping, the triangle inequality, and the fact that

$$\sum_{i=1}^{n} \tau_i^2 \leq \tau \cdot \sum_{i=1}^{n} \tau_i = \tau \cdot \sum_{i=1}^{n} \sum_{S \ni i} c_S^2 = \tau \cdot \sum_{|S| \leq d} |S| \cdot c_S^2 \leq \tau \cdot d.$$

Here in the final inequality we have used our assumption that the coefficients are normalized to satisfy $\sum_{S \neq \emptyset} c_S^2 = \mathbf{Var}(\mathbf{X}) = \mathbf{Var}(\mathbf{Y}) = 1$. It remains to prove (15). Fix $i \in [n]$ and first express $Q(u_1, \ldots, u_n)$ as the sum of two polynomials R and S, the former comprising all terms not containing u_i, and the latter the rest with u_i factored out. That is,

$$Q(u_1, \ldots, u_n) = R(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n) + u_i \cdot S(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n),$$

where R has degree at most d, and S at most $d-1$ (note that if $d = 1$ then S is simply the coefficient α_i of u_i in the linear polynomial L). Next define the random variables

$$\begin{aligned}
\mathbf{R} &= R(\mathcal{G}_1, \ldots, \mathcal{G}_{i-1}, X_{i+1}, \ldots, X_n) \\
\mathbf{S} &= S(\mathcal{G}_1, \ldots, \mathcal{G}_{i-1}, X_{i+1}, \ldots, X_n),
\end{aligned}$$

and note that $\mathbf{Z}_{i-1} = \mathbf{R} + X_i \cdot \mathbf{S}$ and $\mathbf{Z}_i = \mathbf{R} + \mathcal{G}_i \cdot \mathbf{S}$. We bound $|\,\mathbf{E}[\psi(\mathbf{R} + X_i \cdot \mathcal{S})] - \mathbf{E}[\psi(\mathbf{R} + \mathcal{G}_i \cdot \mathbf{S})]|$ by considering their Taylor expansions:

$$\begin{aligned}
&|\,\mathbf{E}[\psi(\mathbf{Z}_{i-1})] - \mathbf{E}[\psi(\mathbf{Z}_i)]| \\
&= \Big|\mathbf{E}\left[\psi(\mathbf{R}) + (X_i \cdot \mathbf{S})\psi^{(1)}(\mathbf{R}) + \tfrac{1}{2}(X_i \cdot \mathbf{S})^2\psi^{(2)}(\mathbf{R}) + \tfrac{1}{6}(X_i \cdot \mathbf{S})^3\psi^{(3)}(\mathbf{R}) + \mathrm{error}_1\right] \\
&\quad - \mathbf{E}\left[\psi(\mathbf{R}) + (\mathcal{G}_i \cdot \mathbf{S})\psi^{(1)}(\mathbf{R}) + \tfrac{1}{2}(\mathcal{G}_i \cdot \mathbf{S})^2\psi^{(2)}(\mathbf{R}) + \tfrac{1}{6}(\mathcal{G}_i \cdot \mathbf{S})^3\psi^{(3)}(\mathbf{R}) + \mathrm{error}_2\right]\Big|.
\end{aligned}$$

Note that the first four terms cancel out since X_i and $\mathcal{G}_i$ are independent of $\mathbf{S}$ and $\mathbf{R}$, and the random variables X_i and $\mathcal{G}_i$ have matching first, second and third moments. Applying the bounds on the error terms given by Taylor's theorem, we see that

$$\begin{aligned}
|\,\mathbf{E}[\mathrm{error}_1 - \mathrm{error}_2]| &\leq \tfrac{1}{24} \cdot C\,\mathbf{E}[X_i^4 \cdot \mathbf{S}^4] + \tfrac{1}{24} \cdot C\,\mathbf{E}[\mathcal{G}_i^4 \cdot \mathbf{S}^4] && (16) \\
&= \tfrac{1}{24} \cdot C\,\mathbf{E}[\mathbf{S}^4] + \tfrac{3}{24} \cdot C\,\mathbf{E}[\mathbf{S}^4] && (17) \\
&< C \cdot \mathbf{E}[\mathbf{S}^4].
\end{aligned}$$

Here (16) uses our assumption that $\psi^{(4)}$ is uniformly bounded by C, and (17) is by independence along with the fact that $\mathbf{E}[\mathcal{G}_i^4] = 3$. Next, since $\mathbf{S}$ is a degree-d polynomial, we may apply Bonami's lemma (Theorem 43) to get $C \cdot \mathbf{E}[\mathbf{S}^4] \leq C \cdot 9^d \cdot \mathbf{E}[\mathbf{S}^2]^2$ and it remains to argue that $\mathbf{E}[\mathbf{S}^2]$ upper bounded by $\tau_i = \mathrm{Inf}_i(Q) = \sum_{S \ni i} c_S^2$. Indeed, recall that S is the polynomial comprising all terms of Q containing u_i with u_i factored out, and so

$$\mathbf{E}[\mathbf{S}^2] = \mathbf{E}\Big[\big(\sum_{S \ni i} c_S \prod_{j \in S \setminus i} Y_j\big)^2\Big] = \sum_{S \ni i} c_S^2 = \tau_i,$$

where each Y_j is either a Rademacher or standard Gaussian random variable, depending on whether $j < i$. We have shown that $|\,\mathbf{E}[\psi(\mathbf{Z}_{i-1})] - \mathbf{E}[\psi(\mathbf{Z}_i)]| \leq C \cdot 9^d \cdot \tau_i^2$, and the proof is complete. ■

In the proof of the Berry-Esséen theorem we needed the fact that the anti-concentration of a standard Gaussian $\mathcal{G}$ at radius ε is $O(\varepsilon)$. That is, for all $t \in \mathbb{R}$ we have $\mathbf{Pr}[|\mathcal{G} - t| < \varepsilon] = O(\varepsilon)$. The low-degree analogue of this small ball probability is given by the following proposition due to Carbery and Wright [CW01, Kan11].

Lemma 65 (Carbery-Wright) *There exists a universal constant C such that the following holds. Let Q be a multilinear polynomial of degree d over $\mathcal{G}_1, \ldots, \mathcal{G}_n$, a sequence of independent standard Gaussians, and $\varepsilon > 0$. Then*

$$\mathbf{Pr}[|Q(\mathcal{G}_1, \ldots, \mathcal{G}_n)| \leq \varepsilon] \leq C \cdot d \cdot (\varepsilon / \|Q(\mathcal{G})\|_2)^{1/d}.$$

In particular, if the coefficients of Q are normalized to satisfy $\mathbf{Var}(Q) = 1$ then for all $t \in \mathbb{R}$ and $\varepsilon > 0$ we have $\mathbf{Pr}[|Q(\mathcal{G}_1, \ldots, \mathcal{G}_n) - t| \leq \varepsilon] = O(d \cdot \varepsilon^{1/d})$.

Lemma 66 (smooth test functions) *Let $r \geq 2$ be an integer. There exists a constant B_r such that for all $0 < \lambda \leq 1/2$ and $t \in \mathbb{R}$ there exists a function $\Delta_{\lambda,t} : \mathbb{R} \to \mathbb{R}$ satisfying the following:*

1. *$\Delta_{\lambda,t}$ is smooth and $\|(\Delta_{\lambda,t})^{(r)}\|_\infty \leq B_r \cdot \lambda^{-r}$.*
2. *$\Delta_{\lambda,t}(x) = 1$ for all $x \leq t - 2\lambda$.*
3. *$\Delta_{\lambda,t}(x) \in [0,1]$ for all $x \in (t - 2\lambda, t + 2\lambda)$.*
4. *$\Delta_{\lambda,t}(x) = 0$ for all $x \geq t + 2\lambda$.*

We are now ready to prove the invariance principle.

Theorem 67 (invariance) *Let $Q(u_1, \ldots, u_n) = \sum_{S \subseteq [n]} c_S \prod_{i \in S} u_i$ be a degree-d multilinear polynomial and assume*

1. *The coefficients $c_S \in \mathbb{R}$ are normalized to satisfy $\sum_{S \neq \emptyset} c_S^2 = 1$.*
2. *We write τ_i to denote $\mathrm{Inf}_i(Q) = \sum_{S \ni i} c_S^2$, and let $\tau = \max_{i \in [n]} \mathrm{Inf}_i(Q)$.*
3. *$\mathbf{X} = Q(X_1, \ldots, X_n)$ and $\mathbf{Y} = Q(\mathcal{G}_1, \ldots, \mathcal{G}_n)$ where $X_1, \ldots, X_n$ are independent Rademachers and $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are independent standard Gaussians.*

Then for all $t \in \mathbb{R}$, $|\,\mathbf{Pr}[Q(X_1, \ldots, X_n) \leq t] - \mathbf{Pr}[Q(\mathcal{G}_1, \ldots, \mathcal{G}_n) \leq t]| = O\big(d \cdot (10^d \cdot \tau)^{1/(4d+1)}\big)$.

Proof. Let $t \in \mathbb{R}$. We will write $Q(X)$ to denote $Q(X_1, \ldots, X_n)$, $Q(\mathcal{G})$ for $Q(\mathcal{G}_1, \ldots, \mathcal{G}_n)$, and ψ for $\Delta_{\lambda, t+2\lambda}$ $(r = 4)$, for some value of $\lambda > 0$ to be determined later. Recall that $\psi(x) = 1$ for all $x \geq (t + 2\lambda) - 2\lambda = t$ and so $\mathbf{Pr}[Q(X) \leq t] \leq \mathbf{E}[\psi(Q(X))]$. Since $\psi^{(4)}$ is uniformly bounded by $O(1/\lambda^4)$, we apply Proposition 64 to get that

$$
\begin{aligned}
\mathbf{E}[\psi(Q(X))] &\leq \mathbf{E}[\psi(Q(\mathcal{G}))] + O(10^d \cdot \tau \cdot \lambda^{-4}). \\
&\leq \mathbf{Pr}[Q(\mathcal{G}) \leq t + 4\lambda] + O(10^d \cdot \tau \cdot \lambda^{-4}). && (18) \\
&= \mathbf{Pr}[Q(\mathcal{G}) \leq t] + \mathbf{Pr}[Q(\mathcal{G}) \in (t, t + 4\lambda)] + O(10^d \cdot \tau \cdot \lambda^{-4}) \\
&= \mathbf{Pr}[Q(\mathcal{G}) \leq t] + O(d \cdot (4\lambda)^{1/d}) + O(10^d \cdot \tau \cdot \lambda^{-4}). && (19)
\end{aligned}
$$

Here (18) is again by the properties of ψ, this time using the fact that $\psi(x) = 0$ for all $x \geq (t + 2\lambda) + 2\lambda$, and (19) is by Carbery-Wright. Choosing $\lambda = (10^d \cdot \tau)^{d/(4d+1)}$, we have shown that

$$\mathbf{E}[\psi(Q(X))] \leq \mathbf{Pr}[Q(\mathcal{G}) \leq t] + O\big(d \cdot (10^d \cdot \tau)^{1/(4d+1)}\big).$$

A symmetric argument establishes the analogous lower bound on $\mathbf{E}[\psi(Q(X))]$, and this completes the proof. ∎

6 Testing dictators and UGC-hardness

Saturday, 3rd March 2012

Guest lecture by Per Austrin

Definition 68 (unique label cover) *Let L be a positive integer. An instance Ψ of the L-*UNIQUE-LABEL-COVER *problem is a graph $G = (V, E)$ where each edge $e \in E$ has an associated constraint that is a permutation $\pi_e : [L] \to [L]$. A labelling of Ψ is a assignment to the vertices $\ell : V \to [L]$. We say that an edge (x, y) is satisfied by ℓ if $\ell(x) = \pi_e(\ell(y))$, and the value of ℓ is the fraction of edges satisfied by ℓ. The optimum of Ψ, denoted* $\mathsf{opt}(\Psi)$, *is the maximum value of the optimal assignment ℓ.*

The UNIQUE-LABEL-COVER problem is a special case of the LABEL-COVER problem where the constraints $\pi_e : [L] \to [L]$ are not required to be permutations. In particular, in the UNIQUE-LABEL-COVER problem assigning a label to a vertex necessarily determines the labels of all its neighbors, whereas this is not the case for the LABEL-COVER problem. Consequently, for UNIQUE-LABEL-COVER the task of deciding whether there is an assignment that satisfies all the edges (*i.e.* distinguishing $\mathsf{opt}(\Psi) = 1$ versus $\mathsf{opt}(\Psi) < 1$) is easy: assume a label for a vertex v and deduce the labels for the remaining vertices in a breadth-first fashion. If there is a conflict at some vertex we choose another label for v and repeat the same process. If no consistent labeling can be found after iterating through all L possible labels for v then $\mathsf{opt}(\Psi) < 1$; otherwise $\mathsf{opt}(\Psi) = 1$. This is in sharp contrast to the situation for LABEL-COVER: it is known that for every $\varepsilon > 0$ there is an L such that it is NP-hard to distinguish between $\mathsf{opt}(\Psi) = 1$ versus $\mathsf{opt}(\Psi) < \varepsilon$ where Ψ is an instance of L-LABEL-COVER; we sometimes refer to this as the $(1, \varepsilon)$-hardness of LABEL-COVER.

The Unique Games Conjecture of S. Khot [Kho02] asserts that the UNIQUE-LABEL-COVER problem is nevertheless very hard to approximate as soon as we move to almost-satisfiable instances.

Conjecture 69 (unique games) *For every $\varepsilon > 0$ there exists an L such that the it is* NP*-hard to distinguish between* $\mathsf{opt}(\Psi) \geq 1 - \varepsilon$ *versus* $\mathsf{opt}(\Psi) < \varepsilon$*, where Ψ is an instance of L-*UNIQUE-LABEL-COVER*. Equivalently, for every $\varepsilon > 0$ there exists an L such that the L-*UNIQUE-LABEL-COVER *problem is $(1 - \varepsilon, \varepsilon)$-hard.*

Recent work of S. Arora, B. Barak and D. Steurer [ABS10] gives an algorithm for UNIQUE-LABEL-COVER running in time $\exp(n^{\mathrm{poly}(\varepsilon)})$.

Today we will prove the following theorem showing how explicit (c, s) dictator versus quasirandom tests yield Unique Games-based hardness results for certain constraint satisfaction problems [Kho02, KR03, KKMO07, Aus08]:

Theorem 70 *Suppose we have a (c, s) dictator versus quasirandom using predicates from a set T. For every L there exist a polynomial time reduction R from L-*UNIQUE-LABEL-COVER *to* $\mathsf{MAX\text{-}CSP}(T)$ *such that for every instance Ψ of L-*UNIQUE-LABEL-COVER *and every $\varepsilon > 0$, there exists a $\delta > 0$ satisfying*

1. *(Completeness) If* $\mathsf{opt}(\Psi) \geq 1 - \delta$ *then* $\mathsf{opt}(R(\Psi)) \geq c - \varepsilon$.
2. *(Soundness) If* $\mathsf{opt}(\Psi) < \delta$ *then* $\mathsf{opt}(R(\Psi)) < s + \varepsilon$.

First, a small catch: the dictator versus quasirandom test have to work not only for boolean functions but also for bounded functions $f : \{-1,1\}^n \to [-1,1]$. We may view any predicate $\phi : \{-1,1\}^k \to \{0,1\}$ as $\phi^* : [-1,1]^k \to [0,1]$, where $\phi^*(y_1,\ldots,y_k) := \mathbf{E}[\phi(\boldsymbol{x_1},\ldots,\boldsymbol{x_k})]$, the expectation taken with respect to $\{-1,1\}$-valued random variables $\boldsymbol{x_i}$ satisfying $\mathbf{E}[\boldsymbol{x_i}] = y_i$. It is easy to check that $\phi^*(x) = \phi(x)$ for all $x \in \{-1,1\}^k$, and in fact we have $\phi^*(y_1,\ldots,y_k) = \sum_{S\subseteq[k]} \hat{\phi}(S) \prod_{i\in S} y_i$. With this observation any tester for boolean functions using predicate ϕ can be extended to one for all bounded functions: instead of accepting iff $\phi(f(x_1),\ldots,f(x_k)) = 1$, the tester accepts with probability $\phi^*(f(x_1),\ldots,f(x_k)) \in [0,1]$.

The reduction from Unique-Label-Cover to MAX-CSP

With this caveat out of the way, we are now ready to describe the reduction R:

> Let Ψ be an instance of L-Unique-Label-Cover defined over a graph $G = (V,E)$. Suppose we have a (c,s) dictator versus quasirandom test for functions $\{-1,1\}^L \to [-1,1]$ using k-ary predicates from a set T. Consider the following instance $R(\Psi)$ of $\mathsf{MAX\text{-}CSP}(T)$:
>
> **Variables.** There will be $|V| \cdot 2^L$ variables: for each $u \in V$ we define 2^L boolean variables $\{Z_{u,x} : x \in \{-1,1\}^L\}$.
>
> **Constraints.** A random constraint will be sampled as follows:
>
> 1. Pick $u \in V$ uniformly.
> 2. Pick k neighbors $v_1,\ldots,v_k \in N(u)$ of u uniformly independently.
> 3. Define $\widetilde{f_{u,v_i}} := f_{v_i}(x \circ \pi_{u,v_i})$.
> 4. Pick $x_1,\ldots,x_k \in \{-1,1\}^L$ according to the distribution over k-tuples induced by the tester, and set $y_i := \widetilde{f_{u,v_i}}(x_i)$.
> 5. Return the constraint $\phi(y_1,\ldots,y_k) = 1$.

In step 3, π_{u,v_i} is the permutation associated with the edge $(u,v_i) \in E$, and $x \circ \pi_{u,v_i}$ is the string x with its coordinates permuted according to π_{u,v_i}. For each $u \in V$, it will be convenient for us to think of an assignment to the corresponding 2^L variables of the CSP as a boolean function $f_u : \{-1,1\}^L \to \{-1,1\}$, where $Z_{u,x} \leftarrow f_u(x)$; an assignment to all $|V| \cdot 2^L$ variables can then be defined as a set of $|V|$ boolean functions $\{f_u : \{-1,1\}^L \to \{-1,1\}\}_{u\in V}$. We will assume that G is regular; this is without loss of generality by a result of Khot and Regev [KR03].

Completeness

Suppose $\mathsf{opt}(\Psi) \geq 1 - \delta$, and let $\ell : V \to [L]$ be a labelling achieving this. Our goal is to exhibit an assignment to the variables of the CSP $R(\Psi)$ that satisfies at least a $c - \varepsilon$ fraction of constraints. We consider the fraction satisfied by the assignment $f_u(x) := \mathsf{DICT}_{\ell(u)}(x) = x_{\ell(u)}$ for all $u \in V$.

Consider an edge $(u, v_i) \in E$ satisfied by ℓ (*i.e.* $\ell(u) = \pi_{u,v_i}(\ell(v_i))$, and note that

$$\widetilde{f_{u,v_i}}(x) = f_{v_i}(x \circ \pi_{u,v_i}) = (x \circ \pi_{u,v_i})_{\ell(v_i)} = x_{\pi_{u,v_i}(\ell(v_i))} = x_{\ell(u)} = \mathsf{DICT}_{\ell(u)}.$$

Therefore if the k edges incident to u (chosen in step 2 above) are all satisfied by ℓ then $\mathbf{E}[\phi(\boldsymbol{y_1}, \ldots, \boldsymbol{y_k})]$ is the probability that the test accepts $\mathsf{DICT}_{\ell(u)}$, at least c by our assumption. Since G is regular and ℓ satisfies a $1 - \delta$ fraction of all edges, the probability that k uniformly random edges incident to a random $u \in V$ is satisfied by ℓ is at least $1 - k\delta$, and so we have $\mathsf{opt}(R(\Psi)) \geq c \cdot (1 - k\delta) \geq c - \varepsilon$ (for sufficiently small δ).

Soundness

We will assume that $\mathsf{opt}(R(\Psi)) \geq s + \varepsilon$ and prove $\mathsf{opt}(\Psi) = \Omega_\varepsilon(1)$. We first express the fraction of satisfied constraints as

$$s + \varepsilon \leq \mathop{\mathbf{E}}_{\substack{u, v_1, \ldots, v_k \\ x_1, \ldots, x_k}} \left[\phi\big(\widetilde{f_{u,v_1}}(x_1), \ldots, \widetilde{f_{u,v_k}}(x_j)\big)\right] = \mathop{\mathbf{E}}_{u, x_1, \ldots, x_k} \left[\phi\big(\mathop{\mathbf{E}}_{v_1}[\widetilde{f_{u,v_1}}(x_1)], \ldots, \mathop{\mathbf{E}}_{v_k}[\widetilde{f_{u,v_k}}(x_k)]\big)\right]$$

For each $u \in V$, let $g_u\{-1,1\}^L \to [-1,1]$ be the function defined by $g_u(x) := \mathbf{E}_{v \in N(u)}[\widetilde{f_{u,v}}(x)]$, and so the above can be rewritten as $\mathbf{E}_{u, x_1, \ldots, x_n}[\phi(g_u(x_1), \ldots, g_u(x_k))] \geq s + \varepsilon$. By an averaging argument, at least an $\frac{\varepsilon}{2}$ fraction of all $u \in V$ satisfy $\mathbf{E}_{x_1, \ldots, x_k}[\phi(g_u(x_1), \ldots, g_u(x_k))] \geq s + \frac{\varepsilon}{2}$; call these $u \in V$ "good". By the soundness condition of a dictator versus quasirandom test, it follows that if u is good g_u cannot be (γ, γ)-quasirandom for some $\gamma = \Omega_\varepsilon(1)$; in particular, there must exist at least one $i \in [L]$ such that $\mathrm{Inf}_i^{(1-\gamma)}(g_u) \geq \gamma$.

Let $J_u = \{i \in [L] : \mathrm{Inf}_i^{(1-\gamma)}(g_u) \geq \gamma\}$, and note that $|J_u| \leq \frac{1}{\gamma^2}$ by Proposition 51. We claim that every $i \in J_u$ satisfies $\mathbf{Pr}_{v \in N(u)}\left[\mathrm{Inf}^{(1-\gamma)}_{\pi^{-1}_{u,v}(i)}(f_v) > \frac{\gamma}{2}\right] > \frac{\gamma}{2}$ (once again, at least one such i exists if u is good). To see this, it suffices to check that

$$\mathop{\mathbf{E}}_{v \in N(u)}\left[\mathrm{Inf}^{(1-\gamma)}_{\pi_{u,v}(i)}(f_v)\right] = \mathop{\mathbf{E}}_{v}\left[\mathrm{Inf}_i^{(1-\gamma)}\big(\widetilde{f_{u,v}}\big)\right] \geq \mathrm{Inf}_i^{(1-\gamma)}\big(\mathop{\mathbf{E}}_{v}\left[\widetilde{f_{u,v}}\right]\big) = \mathrm{Inf}_i^{(1-\gamma)}(g_u) \geq \gamma;$$

the claim then follows by Markov's inequality. For every $u \in V$ we also define $J'_u = \{j \in [L] : \mathrm{Inf}_j^{(1-\gamma)}(f_u) \geq \frac{\gamma}{2}\}$, noting that $|J'_u| \leq \frac{2}{\gamma^2}$.

Consider the following labelling $\ell : V \to [L]$: for each $u \in V$, if $J_u \cup J'_u$ is non-empty we assign u a uniformly random label in $J_u \cup J'_u$, otherwise we assign u an arbitrary label. It remains to prove that ℓ satisfies an $\Omega_\varepsilon(1)$ fraction of edges. We have shown that for at least an $\frac{\varepsilon}{2} \cdot \frac{\gamma}{2} = \frac{\varepsilon\gamma}{4}$ fraction of edges (u, v) there exists an $i \in [L]$ such that $i \in J_u \cup J'_u$ and $\pi^{-1}_{u,v}(i) \in J_v \cup J'_v$. Conditioned on the existence of such an i, the edge is satisfied if

$\ell(u) = i$ and $\ell(v) = \pi_{u,v}^{-1}(i)$ (recall that (u, v) is satisfied if $\ell(u) = \pi_{u,v}(\ell(v))$), and this happens with probability at least $(|J_u \cup J'_u||J_v \cup J'_v|)^{-1} \geq \left(\frac{1}{\gamma^2} + \frac{2}{\gamma^2}\right)^2 = \frac{\gamma^4}{9}$. We conclude that $\mathsf{opt}(\Psi) = \Omega(\varepsilon\gamma \cdot \gamma^4) = \Omega(\varepsilon\gamma^5) = \Omega_\varepsilon(1)$, and the proof is complete.

For further details see [Aus08].

References

[ABS10] S. Arora, B. Barak, and D. Steurer. Subexponential algorithms for unique games and related problems. In *Proceedings of the 51st IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 563–572, 2010.

[Arr50] K. Arrow. A difficulty in the concept of social welfare. *J. of Political Economy*, **58** (4), pp. 328–346, 1950.

[Aus08] P. Austrin. *Conditional Inapproximability and Limited Independence*, Ph.D. thesis, Royal Institute of Technology (KTH), 2008.

[Bec75] W. Beckner. Inequalities in Fourier Analysis. *Annals of Mathematics*, **102**, pp. 159–182, 1975.

[BCH+96] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi, and M. Sudan. Linearity testing in characteristic two. *IEEE Transactions on Information Theory*, **42** (6), pp. 1781–1795, 1996.

[Ber41] A. Berry. The accuracy of Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, **49**(1), pp. 122–136, 1941.

[BL89] M. Ben-Or and N. Linial. Collective coin flipping. In S. Micali, editor, *Randomness and Computation.* Academic Press, 1989.

[BLR93] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. Syst. Sci.*, **47**, pp. 549–595, 1993. Preliminary version in STOC 1990.

[Bon70] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier*, **20** (2), pp. 335–402, 1970.

[Bor85] C. Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Probability Theory and Related Fields*, **70** (1), pp. 1–13, 1985.

[CW01] A. Carbery and J. Wright. Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Mathematical Research Letters*, **8** 3, pp. 233–248, 2001.

[Cha06] S. Chatterjee. A generalization of the Lindeberg principle. *Annals Probab.*, **24** (6), pp. 2061–2076, 2006.

[Con85] N. de Condorcet. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. *Imprimerie Royale, Paris*, 1785.

[Ess42] C.-G. Esséen. On the Liapunoff limit of error in the theory of probability. *Arkiv för matematik, astronomi och fysik*, **A28**, pp. 1–19, 1942.

[FKN02] E. Friedgut, G. Kalai, and A. Naor. Boolean functions whose Fourier transform is concentrated on the first two levels. *Adv. in Appl. Math.*, **29** (3), pp. 427–437, 2002.

[Goe97] M. X. Goemans. Semidefinite programming in comibinatorial optimization. *Mathematical Programming*, **79**, pp. 143–161, 1997.

[Gro75] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, **97**, pp. 1061–1083, 1975.

[Guy86] R. K. Guy. Any answers anent these anytical enigmas?, *Amer. Math Monthly*, **93**, pp. 279–281, 1986.

[Hås01] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, **48** 4, pp. 798–859, 2001.

[HK92] R. Holzman and D. Kleitman, On the product of sign vectors and unit vectors, *Combinatorica*, **12** (3), pp. 303–316, 1992

[KKL98] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science (FOCS),* pp. 68–80, 1988.

[Kal02] G. Kalai. A Fourier-theoretic perspective on the Condorcet paradox and Arrow's theorem. *Adv. in Appl. Math.*, **29** (3), pp. 412–426, 2002.

[Kan11] D. Kane. *On Elliptic curves, the ABC Conjecture, and Polynomial Threshold Functions*, Ph.D. thesis, Harvard University, 2011.

[Kho02] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedsings of the 24th ACM Symposium on Theory of Computing (STOC)*, pp. 767–775, 2002.

[KR03] S. Khot and O. Regev. Vertex cover might be hard to approximate to within $2-\varepsilon$. *Journal of Computer and System Sciences*, **74** (3), pp. 335–349, 2003.

[KV05] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems and the embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 53–63, 2005.

[KKMO07] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for Max-Cut and other 2-variable CSPs? *SIAM Journal on Computing*, **37** (1), pp. 319–357, 2007.

[KO12] G. Kindler and R. O'Donnell. Gaussian noise sensitivity and Fourier tails. In *Proceedings of the 27th IEEE Conference on Computational Complexity (CCC)*, 2012.

[Lin22] J. W. Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, *Math Z.*, **15**, pp. 211–225, 1922.

[Lin02] N. Linial. Squared ℓ_2 metrics into ℓ_1. In J. Matoušek, editor, *Open Problems, Workshop on Discrete Metric Spaces and their Algorithmic Applications*. Haifa, March 2002.

[MORS10] K. Matulef, R. O'Donnell, R. Rubinfeld, and R. Servedio. Testing halfspaces. *SIAM Journal on Computing*, **39** (5), pp. 2004–2047, 2010. Preliminary version in SODA 2009.

[MOO10] E. Mossel, R. O'Donnell, and K. Olezkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, **171** 1, pp. 295–341, 2010.

[Rot75] V. I. Rotar'. Limit theorems for multilinear forms and quasipolynomial functions. *Teor. Verojatnost. i Primenen.*, **20**, pp. 527-546, 1975.

[Rot79] V. I. Rotar'. Limit theorems for polylinear forms. *J. Multivariate Anal.*, **9**, pp. 511–530, 1979.

[She99] W. Sheppard. On the application of the theory of error to cases of normal distribution and normal correlation. *Phil. Trans. Royal Soc. London, Series A,* **192**, pp. 101–531, 1899.

[Tal96] Michel Talagrand. How much are increasing sets positively correlated? *Combinatorica*, **16** (2), pp. 243–258, 1996.