# Fourier Entropy-Influence Conjecture for Random Linear Threshold Functions

Sourav Chakraborty[1], Sushrut Karmalkar[2], Srijita Kundu[3], Satyanarayana V. Lokam[4], and Nitin Saurabh*[5]

[1] Chennai Mathematical Institute, Chennai, India
Centrum Wiskunde Informatika, Amsterdam, Netherlands, sourav@cmi.ac.in
[2] University of Texas, Austin, USA, sushrutk@cs.utexas.edu
[3] Centre for Quantum Technologies, Singapore, srijita.kundu@u.nus.edu
[4] Microsoft Research, Bangalore, India, Satya.Lokam@microsoft.com
[5] Charles University, Prague, Czech Republic, nitin@iuuk.mff.cuni.cz

**Abstract.** The Fourier-Entropy Influence (FEI) Conjecture states that for any Boolean function $f : \{+1, -1\}^n \to \{+1, -1\}$, the Fourier entropy of $f$ is at most its influence up to a universal constant factor. While the FEI conjecture has been proved for many classes of Boolean functions, it is still not known whether it holds for the class of Linear Threshold Functions. A natural question is: Does the FEI conjecture hold for a "random" linear threshold function? In this paper, we answer this question in the affirmative. We consider two natural distributions on the weights defining a linear threshold function, namely uniform distribution on $[-1, 1]$ and Normal distribution.

## 1 Introduction

Boolean functions are the most basic object of study in Theoretical Computer Science. There are many complexity measures associated with a Boolean function $f\colon \{+1, -1\}^n \to \{+1, -1\}$, e.g., *degree*, *sensitivity*, *certificate complexity*, etc. These measures provide lower bound on the complexity of a Boolean function in different models of computation (e.g, PRAM, Circuits/Formulas, Decision trees, etc.).

A particularly interesting one is the *average sensitivity* of a Boolean function. Given $f\colon \{+1, -1\}^n \to \{+1, -1\}$, and an input $x \in \{+1, -1\}^n$, sensitivity of $f$ at $x$, denoted $\mathsf{s}(f, x)$, is defined as the number of neighbours of $x$ in the Hamming cube where $f$ takes different value than at $x$. The *average sensitivity* of $f$, denoted $\mathsf{as}(f)$, is the average of $\mathsf{s}(f, x)$, $\mathbb{E}_x[\mathsf{s}(f, x)]$, under the uniform distribution. In the definition above, we fixed an $x \in \{+1, -1\}^n$ and looked at the variables the value of $f(x)$ depends on. Similarly, we can fix a variable $i \in [n]$, and look at the number of inputs $x$ such that the value $f(x)$ depends on this variable. This leads

* The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 616787.

us to the notion of *influence*. The *influence* of the $i$-th variable, denoted $\mathsf{Inf}_i(f)$, is defined to be $\Pr_x[f(x) \neq f(x^i)]$, where $x$ is chosen uniformly at random and $x^i$ denotes $x$ with $i$-th variable negated. The *influence* of $f$, $\mathsf{Inf}(f)$, is defined as $\sum_{i=1}^{n} \mathsf{Inf}_i(f)$. It is easily seen that $\mathsf{Inf}(f)$ equals $\mathsf{as}(f)$. Influence is related to many complexity measures, and many of these relations can be established via Fourier analysis. For example, circuit size [1, 2], formula size [3], decision trees [4], etc. For an in-depth treatment of Fourier analysis in Boolean functions we refer to [5].

Every Boolean function $f\colon \{+1,-1\}^n \to \{+1,-1\}$ has a unique representation in the Fourier basis :

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \prod_{i \in S} x_i \ .$$

The set of Fourier coefficients is given by $\{\widehat{f}(S)\}_{S\subseteq[n]}$. It follows from Parseval's identity that $\sum_{S\subseteq[n]} \widehat{f}(S)^2 = \mathbb{E}_x[f(x)^2] = 1$. Thus, the squared Fourier coefficients naturally define a distribution over the subsets of $\{1, 2, \ldots, n\}$. The Fourier-entropy of $f$, denoted $\mathbb{H}(f)$, is defined to be the Shannon entropy of this distribution. That is,

$$\mathbb{H}(f) := \sum_{S \subseteq [n]} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2} \ .$$

A longstanding and important conjecture in Analysis of Boolean functions states that the Fourier entropy of a Boolean function is bounded above by the total influence of the function up to a constant factor. More formally,

*Fourier-Entropy Influence (FEI) Conjecture* : There exists a universal constant $C > 0$ such that for all $f\colon \{+1,-1\}^n \to \{+1,-1\}$, $\mathbb{H}(f) \leqslant C \cdot \mathsf{Inf}(f)$. That is,

$$\sum_{S \subseteq [n]} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2} \leqslant C \cdot \sum_{S \subseteq [n]} |S| \widehat{f}(S)^2 \ . \tag{1}$$

The conjecture was made by Friedgut and Kalai [6] in 1996. The genesis of the conjecture is in the study of threshold phenomena in random graphs [6]. For example, it implies a lower bound of $\Omega(\log^2 n)$ on the influence of any $n$-vertex monotone graph property. The current best lower bound, by Bourgain and Kalai [7], is $\Omega(\log^{2-\epsilon} n)$, for any constant $\epsilon > 0$. However, over time, many non-trivial implications have been observed. In particular, it implies a variant of *Mansour's Conjecture* [8] that is sufficient to imply a polynomial time *agnostic* learning algorithm for DNFs [9], for any constant error parameter, resolving a major open problem [10] in computational learning theory.

*Mansour's Conjecture (variant)* : Let $f\colon \{+1,-1\}^n \to \{+1,-1\}$ be any Boolean function computable by a $t$-term DNF formula. Then, for any $\epsilon > 0$, there exists a polynomial $p$ with $t^{O(1/\epsilon)}$ terms such that $\mathbb{E}[(f-p)^2] \leqslant \epsilon$. (Mansour conjectured the exponent of $t$ to be $O(\log \frac{1}{\epsilon})$ [8].)

In general, the FEI conjecture implies *sparse $L_2$-approximations* for Boolean functions. That is, it implies the existence of a polynomial $p$ with $2^{O(\mathsf{Inf}(f)/\epsilon)}$ terms such that $\mathbb{E}[(f-p)^2] \leqslant \epsilon$, for any $\epsilon > 0$. Presently, the best known construction [11] yields a bound of $2^{O((\mathsf{Inf}(f)/\epsilon)^2)}$ on the number of terms.

Finally, the FEI inequality (1) is also known to imply the famous Kahn-Kalai-Linial theorem [12].

*KKL Theorem* : For any Boolean function $f\colon \{+1,-1\}^n \to \{+1,-1\}$, there exists an $i \in [n]$ such that $\mathsf{Inf}_i(f) = \Omega\left(\mathrm{Var}(f)\frac{\log n}{n}\right)$.

In fact, as observed by [13], the following weakening of the FEI conjecture suffices to imply the KKL theorem.

*Fourier Min-Entropy-Influence (FMEI) Conjecture* : There exists a universal constant $C > 0$ such that for all $f\colon \{+1,-1\}^n \to \{+1,-1\}$,

$$\min_{S\subseteq[n]} \log \frac{1}{\widehat{f}(S)^2} \leqslant C \cdot \mathsf{Inf}(f) \ .$$

In other words, it conjectures how large the maximum Fourier coefficient must be. That is, there exists a set $S \subseteq [n]$ such that $\widehat{f}(S)^2 \geqslant 2^{-C\cdot\mathsf{Inf}(f)}$.

For more on the FEI conjecture we refer the interested reader to Gil Kalai's blog post [14].

While the FEI and the FMEI conjectures have resisted solutions for long, it is easy to verify the conjectures for simple functions such as OR, AND, Tribes, Majority, etc. Thus, researchers have tried to establish the conjectures for special classes of Boolean functions. In particular, it has been shown that FEI conjecture holds for random DNFs [15], symmetric functions and read-once decision trees [13], random functions [16], read-once formulas [17, 18], decision trees with bounded average depth [18, 19], and bounded read decision trees [19]. In [13] it was also observed that FMEI conjecture holds for monotone functions.

In this paper, we study the FEI conjecture for the class of *linear threshold functions* (LTF). A Boolean function $f\colon \{+1,-1\}^n \to \{+1,-1\}$ is said to be an LTF if there exists $w_0, w_1, \ldots, w_n \in \mathbb{R}$ such that for all $x \in \{+1,-1\}^n$, $f(x) = \mathsf{sign}(w_0 + w_1x_1 + \cdots + w_nx_n)$, where, the sign is the function that maps positive values to $+1$ and negative values to $-1$. We assume, without loss of generality, $\mathsf{sign}(0) = -1$.

We observe that the KKL theorem implies Fourier Min-Entropy-Influence conjecture holds for linear threshold functions using the same idea from [13]. But for the case of the FEI conjecture we still cannot show that it holds for linear threshold functions in general. So a natural question to ask is:

Does FEI conjecture hold for a random LTF?

Similar question has been asked and answered for other classes of functions (for random DNFs [15] and for random Boolean functions [16]). A natural way of defining a random LTF is when the coefficients $w_0, \ldots, w_n$ are drawn from a distribution $\mathcal{D}$. Choosing $w_0, \ldots, w_n$ is equivalent to choosing a direction in

$(n+1)$-dimensions. Thus, a natural way of sampling an LTF is to sample a unit vector in $(n+1)$-dimensions. There are many ways to sample a unit vector uniformly [20–23]. It is well known that a simple way to sample a unit vector uniformly in $(n+1)$-dimensions is to sample each coordinate from standard normal distribution and then normalize the vector. Specifically, we consider for $\mathcal{D}$ the standard normal distribution and the uniform distribution over $[-1,1]$. To establish our main result, Theorem 1, we prove a generic technical result that says that the FEI conjecture holds with very high probability as long as $\mathcal{D}$ possesses some "nice" properties. Informally, these properties say that a centered random variable with variance 1 and bounded third absolute moment has at least a constant probability mass above the third absolute moment. Our main results are stated below for two specific natural distributions.

**Theorem 1 (Main)** *[Informal]*

1. *If $w_0, \ldots, w_n$ are drawn from the normal distribution $N(0,1)$ and $f(x) = \mathsf{sign}(w_0 + w_1 x_1 + \cdots + w_n x_n)$ then with high probability the FEI Conjecture holds for $f$.*
2. *If $w_0, \ldots, w_n$ are drawn from the uniform distribution over $[-1,1]$ and $f(x) = \mathsf{sign}(w_0 + w_1 x_1 + \cdots + w_n x_n)$ then with high probability the FEI Conjecture holds for $f$.*

We also identify (Corollary 10) certain subclasses of linear threshold functions for which FEI conjecture holds.

### 1.1 Our Proof Technique

To prove the FEI conjecture for a function $f$ (or, any class of Boolean functions) one needs to provide an upper bound on the Fourier entropy of $f$ and a matching (up to a constant factor) lower bound on the influence of $f$. For upper bounding the Fourier entropy of a LTF we crucially use the following theorem proved in [18].

**Theorem 2** *[18] If $f : \{+1,-1\}^n \to \{+1,-1\}$ is* any *linear threshold function, then $\mathbb{H}(f) \leqslant C \cdot \sqrt{n}$, where, $C$ is a universal constant.*

Thus if $f$ is an LTF with influence lower bounded by $\Omega(\sqrt{n})$, then using Theorem 2, we conclude that FEI conjecture holds for $f$. Our main contribution in the paper is to prove that a "random" LTF has influence $\Omega(\sqrt{n})$.

We provide two techniques of proving a lower bound on the influence of an LTF $f$. The first technique (Theorem 4) is a simple application of the Khintchine Inequality [24]. While this lower bound technique is easy and simple, this does not yield a high probability statement about the event $\mathsf{Inf}(f) = \Omega(\sqrt{n})$ when the coefficients are distributed according to the normal distribution (see Section 3.2). For this, we need the second technique.

The starting point in this case is the following basic inequality for LTF's.

$$\mathsf{Inf}_i(f) = \Pr_{x\in\{+1,-1\}^n}\left[-|w_i| < w_0 + \sum_{j\leqslant n:\ j\neq i} w_j x_j \leqslant |w_i|\right] \ . \tag{2}$$

We bound this probability using concentration inequalities. In particular, we crucially use an optimal version of the Berry-Esseen Theorem (Theorem 7) that was proved recently by Shevtsova [25]. Using the Berry-Esseen Theorem we show, informally, if a variable has a "high" weight then it has "high" influence. More precisely, we prove the following technical lemma en route to establishing the main theorem. Consider a symmetric distribution $\mathcal{D}$ around 0 with variance 1.

**Lemma 3** *Let $w_j \sim \mathcal{D}$ for $1 \leqslant j \leqslant n$. For all $i \in [n]$, $\alpha \in \mathbb{R}^+$, and $\delta > 0$, with probability at least $1 - e^{-\Omega(n\mu_3^2)}$ over the choices of $w_j$'s,*

$$\Pr_{x\in\{+1,-1\}^n}\left[\left|\sum_{1\leqslant j\leqslant n:\ j\neq i} w_j x_j\right| \leqslant \alpha\right] \geqslant \frac{\theta}{\sqrt{n}} - O\left(\frac{1}{n^{2/3}}\right),$$

*where $\theta = (\alpha - \mu_3(1 + \frac{2\delta}{1-\delta}))/\sqrt{2\pi(1+\delta)}$, and $\mu_3 = \mathbb{E}_{w\sim\mathcal{D}}[|w|^3]$.*

Observe that for the statement to be non-trivial $\alpha > \mu_3 + \delta'$ for some $\delta' > 0$. The optimality of Shevtsova's theorem (Theorem 7) is crucial to the fact that $\alpha$ can be chosen to be only slightly larger than $\mu_3$, by an *additive* constant rather than $C \cdot \mu_3$ for some large constant $C > 1$. This difference is significant when $\mathcal{D}$ is a *truncated* distribution. A truncated distribution is a conditional distribution that results from restricting the support of some other probability distribution.

*Organization of the paper:* In Section 2 we present the two techniques to lower bound the influence of an LTF. We then show that the FEI conjecture holds for a 'random' LTF in Section 3, and conclude with some observations in Section 4.

## 2 Lower Bounds on Influence

Recall that $f\colon \{+1,-1\}^n \to \{+1,-1\}$ is a linear threshold function if there exists $n+1$ real numbers, $w_0, w_1, \ldots, w_n$, such that

$$\forall x = x_1 \ldots x_n \in \{+1,-1\}^n\colon f(x) = \mathsf{sign}(w_0 + w_1 x_1 + \cdots + w_n x_n) \ ,$$

where $\mathsf{sign}(z) = 1$ if $z > 0$, and $-1$ if $z \leqslant 0$. It is clear that there are infinitely many $(n+1)$-tuples that define the same Boolean function $f$. Nevertheless, relationships among $w_i$'s characterize many properties of the function. In fact, the influence of $f$ is a function on the $w_i$'s. Our goal is to obtain lower bounds for this function using various properties of $w_i$'s. In this section we present two such lower bounds.

The first lower bound that we obtain is a simple observation and the lower bound is in terms of the $\ell_2$-norm of the $w_i$'s and the maximum value of the $w_i$'s.

**Theorem 4** *Let* $f\colon \{+1,-1\}^n \to \{+1,-1\}$ *be such that* $f(x) = \mathsf{sign}(w_0 + \sum_{i=1}^n x_i w_i)$ *for some weights* $w_0, w_1, \ldots, w_n \in \mathbb{R}$*. Then,*

$$\mathsf{Inf}(f) \geqslant \frac{\sqrt{\sum_{i=0}^n w_i^2}}{2\sqrt{2}\max_i |w_i|} - |\hat{f}(\emptyset)| \geqslant \frac{\sqrt{\sum_{i=0}^n w_i^2}}{2\sqrt{2}\max_i |w_i|} - 1 \ .$$

The proof is omitted. It follows from an easy application of the well-known Khintchine Inequality.

**Theorem 5 (Khintchine Inequality)** *[24] Let* $w_1, w_2, \ldots, w_n \in \mathbb{R}$ *be such that* $\sum_i w_i^2 = 1$*. Then,* $\mathbb{E}_{x\in\{+1,-1\}^n}[|w_1x_1 + \cdots + w_nx_n|] \geqslant \frac{1}{\sqrt{2}}$.

The second lower bound we obtain is more sophisticated lower bound. Recall that the $i$-th influence of $f$ is given by $\Pr_{x\in\{+1,-1\}^n}[f(x) \neq f(x^i)]$. Using the fact that $f$ is an LTF, it is easily seen to be equivalent to Eq. (2). We bound the expression in (2) from below to establish the lower bound.

**Theorem 6** *Let* $w_0, w_1, \ldots, w_n \in \mathbb{R}$ *and let* $f : \{+1,-1\}^n \to \{+1,-1\}$ *be defined by* $f(x) = \mathsf{sign}(w_0 + \sum_{i=1}^n x_i w_i)$*. Then, for* $1 \leqslant i \leqslant n$*,*

$$\mathsf{Inf}_i(f) \geqslant \frac{1}{2\sqrt{2\pi(n-1)B_i}}\left(|w_i| - \frac{A_i}{B_i}\right) - \frac{|w_i|^3}{6\sqrt{2\pi}((n-1)B_i)^{3/2}} - \frac{3.4106\ A_i^{4/3}}{(n-1)^{2/3}B_i^2},$$

*where* $A_i := \frac{1}{n-1}\sum_{0\leqslant j\neq i\leqslant n}|w_j|^3$ *and* $B_i := \frac{1}{n-1}\sum_{0\leqslant j\neq i\leqslant n}|w_j|^2$.

A crucial ingredient in the proof is the following optimal version of Berry-Esseen Theorem that bounds the uniform distance between the cumulative distribution function of the standard normal distribution $N(0,1)$ and the standardized sum of independent symmetric Bernoulli random variables.

**Theorem 7 (Corollary 4.19 in [25])** *Let* $X_1, \ldots, X_n$ *be independent symmetric Bernoulli random variables such that for all* $1 \leqslant j \leqslant n$ *there exist* $a_j$ *with* $\Pr[X_j = a_j] = \Pr[X_j = -a_j] = \frac{1}{2}$*. If* $\Phi(x)$ *is the cumulative distribution function of the standard normal distribution, then,*

$$\sup_x \left|\Pr\left[\frac{X_1 + \cdots + X_n}{\sqrt{\sum_{j=1}^n a_j^2}} < x\right] - \Phi(x)\right| \leqslant \frac{\ell_n}{\sqrt{2\pi}} + 3.4106 \cdot \ell_n^{4/3} \ ,$$

*where* $\ell_n = (\sum_j |a_j|^3)/(\sum_j a_j^2)^{3/2}$.

The optimality is in the fact that the constant of $1/\sqrt{2\pi}$ (as a coefficient of $\ell_n$) is the best possible.

We will prove Theorem 6 in two stages. First we establish it under the assumption that $w_0 = 0$. Then, in Section A, we reduce the non-homogeneous case ($w_0 \neq 0$) to the homogeneous case ($w_0 = 0$) to complete the proof.

### 2.1 Proof of Theorem 6 (assuming $w_0 = 0$)

As mentioned earlier, the full proof of Theorem 6 will follow from the discussions in Section A. Thus, for some $w_1, \ldots, w_n \in \mathbb{R}$, we have $f(x) = \mathsf{sign}(\sum_{i=1}^{n} w_i x_i)$ for all $x \in \{+1, -1\}^n$. Relaxing the estimate in Eq. (2), we have

$$\mathsf{Inf}_i(f) \geqslant \frac{1}{2} \Pr_{x \in \{+1,-1\}^n} \left[ \left| \sum_{j \leqslant n:\ j \neq i} w_j x_j \right| \leqslant |w_i| \right] \ .$$

To obtain the claimed lower bound on influence, we further lower bound the above expression. We start with some definitions required for the proof.

For any $1 \leqslant i \leqslant n$, we define

$$A_i := \frac{1}{n-1} \sum_{j \neq i} |w_j|^3, \quad \text{and} \quad B_i := \frac{1}{n-1} \sum_{j \neq i} |w_j|^2.$$

Note that the proof of Theorem 6, for the case when $w_0 = 0$, follows by substituting $\alpha = |w_i|$ in the following lemma.

**Lemma 8** *For all $i \in [n]$ and any $\alpha \in \mathbb{R}^+$,*

$$\frac{1}{2} \Pr_{x \in \{+1,-1\}^n} \left[ \left| \sum_{j \leqslant n:\ j \neq i} w_j x_j \right| \leqslant \alpha \right] \geqslant \frac{1}{\sqrt{2\pi(n-1)B_i}} \left( \alpha - \frac{A_i}{B_i} \right) - \frac{\alpha^3}{6\sqrt{2\pi}((n-1)B_i)^{3/2}} - \frac{3.4106\ A_i^{4/3}}{(n-1)^{2/3} B_i^2} \ .$$

*Proof.* For any $\alpha > 0$, by the symmetry of distribution over $\sum w_j x_j$, we have

$$\begin{aligned} \frac{1}{2} \Pr_x \left[ \left| \sum_{j \leqslant n:\ j \neq i} w_j x_j \right| \leqslant \alpha \right] &= \Pr_x \left[ 0 \leqslant \sum_{j \leqslant n:\ j \neq i} w_j x_j \leqslant \alpha \right] \\ &= \Pr_x \left[ \frac{\sum_{j \leqslant n:\ j \neq i} w_j x_j}{\sqrt{\sum_{j \leqslant n:\ j \neq i} w_j^2}} \leqslant \frac{\alpha}{\sqrt{\sum_{j \leqslant n:\ j \neq i} w_j^2}} \right] - \frac{1}{2} \ . \end{aligned}$$

Using Theorem 7 on random variables $w_j x_j$'s, we obtain the following lower bound

$$\frac{1}{2} \Pr_x \left[ \left| \sum_{j \leqslant n:\ j \neq i} w_j x_j \right| \leqslant \alpha \right] \geqslant \Phi \left( \frac{\alpha}{\sqrt{\sum_{j \leqslant n:\ j \neq i} w_j^2}} \right) - \left( \frac{1}{2} \right) - \frac{L_{(n-1)}}{\sqrt{2\pi}} - 3.4106 \cdot L_{(n-1)}^{4/3} \ , \tag{3}$$

where

$$L_{(n-1)} = \frac{\sum_{j \leqslant n:\ j \neq i} |w_j|^3}{(\sum_{j \leqslant n:\ j \neq i} |w_j|^2)^{3/2}} = \frac{(n-1)A_i}{((n-1)B_i)^{3/2}} \ .$$

Thus we have

$$-\frac{L_{(n-1)}}{\sqrt{2\pi}} - 3.4106 \cdot L_{(n-1)}^{4/3} = -\frac{1}{\sqrt{2\pi}} \cdot \frac{A_i}{\sqrt{(n-1)}B_i^{3/2}} - 3.4106\frac{A_i^{4/3}}{(n-1)^{2/3}B_i^2} \ . \tag{4}$$

Also, using the following Maclaurin series

$$\Phi(x) - \frac{1}{2} = \frac{1}{\sqrt{2\pi}}\left(x - \frac{1}{6}x^3 + \frac{1}{40}x^5 - \cdots\right) \ ,$$

we have

$$\Phi\left(\frac{\alpha}{\sqrt{\sum_{j\leqslant n:\ j\neq i} w_j^2}}\right) - \frac{1}{2} \geqslant \frac{\alpha}{\sqrt{2\pi B_i(n-1)}} - \frac{\alpha^3}{6\sqrt{2\pi}((n-1)B_i)^{3/2}} \ . \tag{5}$$

Using equations (5) and (4) in Equation (3), we obtain

$$\frac{1}{2}\Pr_x\left[\left|\sum_{j\leqslant n:\ j\neq i} w_j x_j\right| \leqslant \alpha\right] \geqslant \frac{1}{\sqrt{2\pi(n-1)B_i}}\left(\alpha - \frac{A_i}{B_i}\right) - \frac{\alpha^3}{6\sqrt{2\pi}((n-1)B_i)^{3/2}} - \frac{3.4106\ A_i^{4/3}}{(n-1)^{2/3}B_i^2} \ .$$

□

In our applications of the lemma, the second and third terms will be negligible compared to the first term.

### 2.2 Fourier Entropy Influence Conjecture for a class of Linear Threshold Functions

In this section we identify a class of LTFs for which we prove the FEI conjecture. The class of functions is $\tau$-regular functions.

**Definition 9** *Suppose $w_0, w_1, \ldots, w_n$ are real numbers such that $\sum_{i=0}^n w_i^2 = 1$. If for all $0 \leqslant i \leqslant n$, $|w_i| \leqslant \tau$ then the linear threshold function, $f$ defined as $f(x_1, \ldots, x_n) = \mathsf{sign}(w_0 + \sum_{i=1}^n w_i x_i)$ is called a $\tau$-regular LTF.*

From Theorem 4 and Theorem 2 we can show that $\tau$-regular LTFs satisfy the FEI conjecture.

**Corollary 10** *If $\tau \leqslant c/\sqrt{n}$ and if $f$ is a $\tau$-regular function then*

$$\mathbb{H}(f) \leqslant O(\mathsf{Inf}(f)),$$

*where the constant in the Big-oh notation depends on $c$.*

# 3 Lower Bounds on Influence for Random Linear Threshold Functions

Given that we still cannot prove the FEI conjecture for all Linear Threshold Functions a natural question is whether FEI conjecture holds for a "random" LTF.

Suppose $w_0, w_1, \ldots, w_n$ are drawn independently from a distribution $\mathcal{D}$. Consider the function $f(x)$ where

$$\forall x = x_1, \ldots, x_n \in \{+1, -1\}^n \colon f(x) = \mathsf{sign}(w_0 + w_1 x_1 + \cdots + w_n x_n).$$

Without loss of generality we can assume that the distribution is symmetric around the origin and thus the mean of the distribution $\mu(\mathcal{D})$ is 0. Also since the function remains same even if we scale the $w_i$'s by any value, so we can also assume that the variance of the distribution $\sigma(\mathcal{D})$ is 1.

As a step towards proving the FEI conjecture for $f$ we need to lower bound the influence of $f$. Note that Theorem 4 and Theorem 6 give lower bounds on the influence of $f$ in terms of $w_i$'s. Using similar arguments we can obtain a lower bound on the influence that holds for a random $f$ (that is, when $w_i$'s are drawn from a distribution $\mathcal{D}$) with high probability.

For any $n \in \mathbb{N}$, let $w_1, \ldots, w_n \sim \mathcal{D}$ be the outcome of $n$ independent samples according to $\mathcal{D}$. We define, for any $\alpha \in \mathbb{R}^+$, $p_{\mathcal{D},n}(\alpha) = \Pr[\max_{i=1}^{n}\{|w_i|\} \geqslant \alpha]$.

**Corollary 11 (Corollary of Theorem 4)** *For any $\alpha \in \mathbb{R}^+$, with probability at least $1 - p_{\mathcal{D},n+1}(\alpha) - o(1)$,*

$$\mathsf{Inf}(f) \geqslant \Omega(\sqrt{n}/\alpha).$$

Using Bernstein's inequality it is easily seen that $\sum_i w_i^2 = \Omega(n)$ with probability $1 - o(1)$.

**Theorem 12** *For any $\alpha \in \mathbb{R}^+$ and any $\delta > 0$, with probability at least $1 - e^{-\delta^2 n \mu_3^2/\sigma_3} - 2e^{-\delta^2 n/\sigma_2} - 2e^{-(n/4)p_{\mathcal{D},1}(\alpha)^2}$ over the choices of $w_j$'s*

$$\mathsf{Inf}(f) \geqslant p_{\mathcal{D},1}(\alpha) \cdot \Theta(\alpha) \cdot \sqrt{n},$$

*where $\Theta(\alpha) = \left(\alpha - \mu_3(1 + \frac{2\delta}{1-\delta})\right)/\sqrt{2\pi(1+\delta)}$, $\mu_3 = \mathbb{E}_{w\sim\mathcal{D}}[|w|^3]$, and $\sigma_2$ and $\sigma_3$ are the standard deviations of $w^2$ and $|w|^3$, respectively.*

## 3.1 Proof of Theorem 12

As argued in Section A, it suffices to establish the lower bound on influence when $w_0 = 0$. To establish the theorem we argue similarly as in the proof of Lemma 8. We consider the case when the distribution $\mathcal{D}$ over $\mathbb{R}$ is symmetric around 0 and with variance 1. That is, $\mathbb{E}_{w\sim\mathcal{D}}[w^2] = 1$. Let $\mu_3 := \mathbb{E}_{w\sim\mathcal{D}}[|w|^3]$. Further, $\sigma_2$ and $\sigma_3$ denote the standard deviation of $w^2$ and $|w|^3$, respectively. That is, $\sigma_2^2 = \mathbb{E}[(w^2 - \mathbb{E}[w^2])^2]$ and $\sigma_3^2 = \mathbb{E}[(|w|^3 - \mathbb{E}[|w|^3])^2]$.

The following lemma gives a lower bound on the $i$-th influence.

**Lemma 13** *Let $w_j \sim \mathcal{D}$ for $1 \leqslant j \leqslant n$. For all $i \in [n]$, $\alpha \in \mathbb{R}^+$, and $\delta > 0$, with probability at least $1 - e^{-\delta^2 n \mu_3^2/\sigma_3} - 2e^{-\delta^2 n/\sigma_2}$ over the choices of $w_j$'s,*

$$\Pr_{x \in \{+1,-1\}^n}\left[\left|\sum_{1 \leqslant j \leqslant n:\ j \neq i} w_j x_j\right| \leqslant \alpha\right] \geqslant \frac{\theta}{\sqrt{n}} - O\left(\frac{1}{n^{2/3}}\right),$$

*where $\theta = (\alpha - \mu_3(1 + \frac{2\delta}{1-\delta}))/\sqrt{2\pi(1+\delta)}$.*

*Proof.* To start we have $w_1, \ldots, w_n$ that are independently and identically distributed according to $\mathcal{D}$. Define random variables $A_i = (\sum_{j \neq i} |w_j|^3)/(n-1)$ and $B_i = (\sum_{j \neq i} w_j^2)/(n-1)$. Thus $\mathbb{E}[A_i] = \mu_3$ and $\mathbb{E}[B_i] = 1$. Using Bernstein's inequality (see Corollary 2.11 in Chapter 2 of [26]) on independent symmetric random variables $w_j$'s we obtain

$$(1-\delta)(n-1) \leqslant \sum_{1 \leqslant j \leqslant n:\ j \neq i} w_j^2 \leqslant (1+\delta)(n-1)\ , \tag{6}$$

with probability at least $1 - 2e^{-\delta^2 n/\sigma_2}$, where $\sigma_2^2 = \mathrm{Var}_{w \sim \mathcal{D}}(w^2)$. Similarly, we also have

$$\sum_{1 \leqslant j \leqslant n:\ j \neq i} w_j^3 \leqslant (1+\delta)C(n-1)\ , \tag{7}$$

with probability at least $1 - e^{-2\delta^2 n \mu_3^2/\sigma_3}$, where $\sigma_3^2 = \mathrm{Var}_{w \sim \mathcal{D}}(|w|^3)$. Thus, from the union bound, it follows that both equations, (6) and (7), holds with probability at least $1 - e^{-\delta^2 n \mu_3^2/\sigma_3} - 2e^{-\delta^2 n/\sigma_2}$, and thus with this probabilty we have $(A_i/B_i) \leqslant \left(1 + \frac{2\delta}{1-\delta}\right)\mu_3$. Recall from Lemma 8 we know

$$\Pr_{x \in \{+1,-1\}^n}\left[\left|\sum_{j \leqslant n : j \neq i} w_j x_j\right| \leqslant \alpha\right] \geqslant \frac{1}{\sqrt{2\pi(n-1)B_i}}\left(\alpha - \frac{A_i}{B_i}\right) - O\left(\frac{1}{n^{2/3}}\right)\ .$$

Thus plugging in the bound on $(A_i/B_i)$ in the above inequality we obtain the lemma,

$$\Pr_{x \in \{+1,-1\}^n}\left[\left|\sum_{j \leqslant n : j \neq i} w_j x_j\right| \leqslant \alpha\right] \geqslant \frac{\alpha - (1 + \frac{2\delta}{1-\delta})\mu_3}{\sqrt{2\pi(1+\delta)(n-1)}} - O\left(\frac{1}{n^{2/3}}\right)\ .$$

□

We now proceed to complete the proof of the theorem. For any $\alpha \in \mathbb{R}^+$,

$$\mathsf{Inf}(f) = \sum_i \mathsf{Inf}_i(f) \geqslant \sum_{i : w_i \geqslant \alpha} \mathsf{Inf}_i(f) \geqslant \sum_{i : w_i \geqslant \alpha} \Pr_{x \in \{+1,-1\}^n}\left[\left|\sum_{j \leqslant n:\ j \neq i} w_j x_j\right| \leqslant \alpha\right]\ .$$

Thus, from Lemma 13, we have with probabiltity at least $1 - e^{-\delta^2 n\mu_3^2/\sigma_3} - 2e^{-\delta^2 n/\sigma_2}$ over the choices of $w_j$'s,

$$\mathsf{Inf}(f) \geqslant N_\alpha \cdot \left( \frac{\alpha - \mu_3\left(1 + \frac{2\delta}{1-\delta}\right)}{\sqrt{2(n-1)\pi(1+\delta)}} - O\left(\frac{1}{n^{2/3}}\right)\right) \ ,$$

where $N_\alpha$ is the number of $w_i$'s that are bigger than $\alpha$. Now the following lemma shows that with probability at least $1 - 2e^{-\frac{n}{4}p_{\mathcal{D},1}(\alpha)^2}$ we have $N_\alpha \geqslant \frac{n}{2} \cdot p_{\mathcal{D},1}(\alpha)$, and that establishes the theorem.

**Lemma 14** *Fix an $\alpha \in \mathbb{R}^+$. If we sample $n$ points according to $\mathcal{D}$ then with probability at least $1 - 2e^{-\frac{n}{4}p_{\mathcal{D},1}(\alpha)^2}$ the number of points in the interval $[\alpha, \infty)$ is greater than $\frac{n \cdot p_{\mathcal{D},1}(\alpha)}{2}$ and less than $\frac{3n \cdot p_{\mathcal{D},1}(\alpha)}{2}$.*

*Proof.* Let $\mathbf{1}_i$ be the indicator function of the event that the $i$-th sample lies in the interval $[\alpha, \infty)$. Define $X = \sum_{j=1}^n \mathbf{1}_j$. Note that for each $j$, $\mathbb{E}[I_j] = p_{\mathcal{D},1}(\alpha)$, so using Chernoff bound on $X$ we get,

$$\Pr\left[\frac{n \cdot p_{\mathcal{D},1}(\alpha)}{2} \leqslant X \leqslant \frac{3n \cdot p_{\mathcal{D},1}(\alpha)}{2}\right] \geqslant 1 - 2e^{-\frac{1}{4}n(p_{\mathcal{D},1}(\alpha))^2}.$$

□

### 3.2 Fourier Entropy Conjecture for a Random Linear Threshold Function

In this section, we give two natural examples of distributions on the $w_i$'s under which the FEI conjecture holds with high probability. First we consider the uniform distribution on the closed interval $[-1, 1]$, and then the normal distribution $N(0, 1)$. Together this completes the proof of Theorem 1.

#### Uniform distribution on $[-1, 1]$: $\mathcal{U}(-1, 1)$

**Corollary 15** *If $w_i \sim \mathcal{U}(-1, 1)$ for $i \in \{0, \ldots, n\}$, then the FEI holds with high probability.*

*Proof.* Since $\mathbb{E}_{x\sim\mathcal{U}(-1,1)}[x^2] = 1/3$, when $w_0, \ldots, w_n$ are drawn independently from $\mathcal{U}[-1, 1]$, by Chernoff bound we have with high probability $\sum_{i=0}^n \mathbb{E}[w_i^2] = \Omega(n)$. Therefore, from Corollary 11, we have with high probability $\mathsf{Inf}(f) = \Omega(\sqrt{n})$. Using the upper bound on the Fourier entropy from Theorem 2 we have our result. □

**Normal distribution : $N(0, 1)$** We note that for the normal distribution, unlike the case of uniform distribution, the FEI conjecture does not follow from Khintchine's inequality. Indeed, Corollary 11 does not give us what we want, i.e., a high probability on the event $\mathsf{Inf}(f) = \Omega(\sqrt{n})$. On the other hand, if we try to boost the probability we end up with a weaker lower bound on the influence. To see this, observe that when $w_1, \ldots, w_n$ are drawn independently from $N(0, 1)$,

1. $\mathbb{E}[\sum_i w_i^2] = n$ (since $\mathbb{E}[X^2] = 1$ for $X \sim N(0,1)$). This implies that with high probability $\sum w_i^2 = \Omega(n)$, and hence $\sqrt{\sum w_i^2} = \Omega(\sqrt{n})$.
2. The probability that $\max_{i=1}^n |w_i| = O(\sqrt{\log n})$ is $1 - o(1)$. Using the fact $\Pr[X > x] \leqslant e^{-x^2/2}$ when $X \sim N(0,1)$, we have $\Pr[w_i > \sqrt{2c \log n}] \leqslant \frac{1}{n^c}$. Since $w_i$'s are drawn independently, the probability that all the $w_i$'s are less than $\sqrt{2c \log n}$ is lower bounded by $1 - \frac{1}{n^{c-1}}$, where $c > 1$ is a constant.

Applying Corollary 11 to this now gives us $\mathsf{Inf}(f) \geqslant \Omega\left(\sqrt{\frac{n}{\log n}}\right)$ with $1 - o(1)$ probability. Moreover, the max-central limit theorem implies that for large enough $n$, with high probability, $\max_i |w_i| = \Theta(\sqrt{\log n})$ (see Example 10.5.3 in [27]). However using the other technique, namely Theorem 12, we can obtain our desired result.

**Corollary 16** *If $w_i \sim N(0,1)$ for $i \in \{0, \ldots, n\}$, then the FEI holds with probability at least $1 - e^{-\Omega(n)}$.*

*Proof.* We use the following well known bounds on moments of $N(0,1)$.

$$\mathbb{E}_{w \sim N(0,1)}[|w|^r] = \begin{cases} \sqrt{\frac{2}{\pi}}\,(r-1)!! & \text{if } r \text{ is odd,} \\ (r-1)!! & \text{if } r \text{ is even.} \end{cases}$$

Therefore, we have $\mu_3 = \frac{2\sqrt{2}}{\sqrt{\pi}}$, $\sigma_2^2 = \mathbb{E}[w^4] - (\mathbb{E}[w^2])^2 = 2$, and $\sigma_3^2 = \mathbb{E}[|w|^6] - (\mathbb{E}[|w|^3])^2 = (15 - \frac{8}{\pi})$. Further, we set $\alpha = \mu_3(2 + \frac{2\delta}{1-\delta})$, for any $\delta > 0$. Hence,

$$p_{\mathcal{D},1}(\alpha) \geqslant 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}} \left(\frac{1}{\alpha} - \frac{1}{\alpha^3}\right) = \Omega(1) \ .$$

The first inequality is easily established; for example, see Exercise 1, Chapter 7 in [28]. Thus, from Theorem 12 we have that $\mathsf{Inf}(f) = \Omega(\sqrt{n})$ and now using Theorem 2 we have our result. □

*Remark 1.* We note that the proof of Theorem 12 shows that the FEI conejcture holds with high probability, as long as $\mathcal{D}$ (with $\mu = 0$ and $\sigma^2 = 1$) statisfies the following properties: $(i)$ $\mathbb{E}_{w \sim \mathcal{D}}[|w|^3]$ is finite, and $(ii)$ $\Pr_{w \sim \mathcal{D}}[|w| > \mathbb{E}_{w \sim \mathcal{D}}[|w|^3]] = \Omega(1)$. In particular, our proof holds for *truncated* distributions, and the optimality of the constant in Shevtsova's theorem (Theorem 7) is crucial in such cases to establish property $(ii)$.

## 4 Conclusion and Open Problems

We proved in this paper that for a random linear threshold function $f(x) := \mathsf{sign}(w_0 + \sum_i w_i x_i)$, where the $w_i$'s are drawn from a distribution that has some 'nice' properties (Remark 1), the FEI conjecture holds with high probability. Moreover, we show that the uniform distribution over an interval, say $[-1, 1]$, and the normal distribution $N(0,1)$ satisfy these 'nice' properties. Indeed, we

established that a random LTF sampled according to these distributions has influence $\Omega(\sqrt{n})$. When combined with the $O(\sqrt{n})$ upper bound (Theorem 2) from previous work on the Fourier entropy of *all* LTF's, we conclude that the FEI conjecture holds for the random LTF's sampled as above. In the process, we obtain non-trivial lower bounds on the influence of $f$ in terms of the $w_i$'s.

An obvious open question is to prove that the FEI conjecture holds for all LTF's. While our current techniques seem far from sufficient to achieve this goal, a natural question is to prove a generalization of Theorem 2, and give an upper bound on the Fourier entropy in terms of $w_i$'s. We believe it is possible to obtain such a general upper bound as a function of a suitable notion of "skewness" of the weights $w_i$.

Another natural question is whether one can show that FEI conjecture holds for a random *polynomial* threshold function of degree $d$. A polynomial threshold function of degree $d$ is defined as the sign of a degree-$d$ polynomial. For a degree-$d$ PTF $f(x) = \mathsf{sign}(p(x_1, \ldots, x_n))$, observe that $\mathsf{Inf}_i(f) = \Pr_x[\mathsf{sign}(p(x)) \neq \mathsf{sign}(p(x^i))] = \Pr_x[|p(x)| < 2|D_i p(x)|]$, where $D_i p(x)$ is the partial derivative of $p(x)$ with respect to $x_i$. Hence, to lower bound the influence, a standard approach would be to show that $p(x)$ has certain *concentration* properties and $D_i p(x)$ has certain *anti-concentration* properties, for a random $p$. However, we currently do not know such strong enough concentration/anti-concentration bounds. Our techniques do not seem to generalize even to polynomial threshold functions of degree 2. Such generalizations seem to require more powerful tools. We also note that currently the best bound on the Fourier entropy of a degree-$d$ PTF is $O(\sqrt{n}(\log n)^{O(d \log d)})$ where the constant in $O(\cdot)$ depends on the degree $d$. (It follows from Kane's bound on the average sensitivity of a PTF [29].)

*Acknowledgments.* We thank the reviewers for helpful comments that improved the presentation of the paper.

## References

1. Linial, N., Mansour, Y., Nisan, N.: Constant depth circuits, fourier transform, and learnability. J. ACM **40**(3) (July 1993) 607–620
2. Boppana, R.B.: The average sensitivity of bounded-depth circuits. Inf. Process. Lett. **63**(5) (1997) 257–261
3. Ganor, A., Komargodski, I., Lee, T., Raz, R.: On the noise stability of small demorgan formulas. Technical report, Electronic Colloquium on Computational Complexity (ECCC) TR 12-174 (2012)
4. O'Donnell, R., Saks, M., Schramm, O.: Every decision tree has an in.uential variable. In: Proc. of the 46th Annual IEEE Symposium on Foundations of Computer Science. FOCS '05, IEEE Computer Society (2005) 31–39
5. O'Donnell, R.: Analysis of Boolean Functions. Cambridge University Press (2014)
6. Friedgut, E., Kalai, G.: Every monotone graph property has a sharp threshold. Proceedings of the American Mathematical Society **124**(10) (1996) 2993–3002
7. Bourgain, J., Kalai, G.: Influences of variables and threshold intervals under group symmetries. Geometric and Functional Analysis (GAFA) **7**(3) (1997) 438–461

8. Mansour, Y.: An $\mathcal{O}(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. Journal of Computer and System Sciences **50**(3) (1995) 543–550
9. Gopalan, P., Kalai, A.T., Klivans, A.: Agnostically learning decision trees. In: Proc. of the 40th annual ACM symposium on Theory of computing. STOC '08 (2008) 527–536
10. Gopalan, P., Kalai, A., Klivans, A.R.: A query algorithm for agnostically learning dnf? In: 21st Annual Conference on Learning Theory - COLT 2008, Helsinki, Finland, July 9-12, 2008. (2008) 515–516
11. Friedgut, E.: Boolean functions with low average sensitivity depend on few coordinates. Combinatorica **18**(1) (1998) 27–35
12. Kahn, J., Kalai, G., Linial, N.: The influence of variables on Boolean functions. In: Proc. of the 29th Annual IEEE Symposium on Foundations of Computer Science. (1988) 68–80
13. O'Donnell, R., Wright, J., Zhou, Y.: The Fourier entropy-influence conjecture for certain classes of Boolean functions. In: Proceedings of Automata, Languages and Programming - 38th International Colloquium. (2011) 330–341
14. Kalai, G.: The entropy/influence conjecture. Terence Tao's blog: https://terrytao.wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjecture/
15. Klivans, A., Lee, H., Wan, A.: Mansour's conjecture is true for random DNF formulas. In: Proc. of the 23rd Conference on Learning Theory. (2010) 368–380
16. Das, B., Pal, M., Visavaliya, V.: The entropy influence conjecture revisited. Technical report, arXiv:1110.4301 (2011)
17. O'Donnell, R., Tan, L.Y.: A composition theorem for the Fourier entropy-influence conjecture. In: Proceedings of Automata, Languages and Programming - 40th International Colloquium. (2013) 780–791
18. Chakraborty, S., Kulkarni, R., Lokam, S.V., Saurabh, N.: Upper bounds on fourier entropy. Theoretical Computer Science **654** (2016) 92–112
19. Wan, A., Wright, J., Wu, C.: Decision trees, protocols and the entropy-influence conjecture. In: Innovations in Theoretical Computer Science. (2014) 67–80
20. Hicks, J.S., Wheeling, R.F.: An efficient method for generating uniformly distributed points on the surface of an n-dimensional sphere. Commun. ACM **2**(4) (April 1959) 17–19
21. Muller, M.E.: A note on a method for generating points uniformly on n-dimensional spheres. Commun. ACM **2**(4) (April 1959) 19–20
22. Marsaglia, G.: Choosing a point from the surface of a sphere. Ann. Math. Statist. **43**(2) (04 1972) 645–646
23. Petersen, W.P., Bernasconi, A.: Uniform sampling from an $n$-sphere. Technical report, Swiss Center for Scientific Computing (1997)
24. Szarek, S.J.: On the best constants in the khinchine inequality. Studia Math. **58** (1976) 197–208
25. Shevtsova, I.: Moment-type estimates with asymptotically optimal structure for the accuracy of the normal approximation. Annales Mathematicae Et Informaticae **39** (2012) 241–307
26. Boucheron, S., Lugosi, G., Massart, P.: Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford University Press (2013)
27. David, H.A., Nagaraja, H.N.: Order Statistics. 3 edn. John Wiley & Sons (2003)
28. Feller, W.: An Introduction to Probability Theory and Its Applications, Volume 1. 3 edn. John Wiley & Sons (1968)
29. Kane, D.M.: The correct exponent for the gotsman–linial conjecture. computational complexity **23**(2) (Jun 2014) 151–175

## A Reducing Non-Homogeneous to Homogeneous case

Let $f(x) = \mathsf{sign}(w_0 + \sum_{i=1}^{n} w_i x_i)$ for all $x \in \{+1, -1\}^n$. Recall from Eq. (2) we have an exact expression for the $i$-th influence for all $1 \leqslant i \leqslant n$. We can relax the probability estimate to lower bound the influence as follows,

$$\mathsf{Inf}_i(f) \geqslant \frac{1}{2} \Pr_{x \in \{+1,-1\}^n} \left[ \left| w_0 + \sum_{j \leqslant n:\ j \neq i} w_j x_j \right| \leqslant |w_i| \right] \ . \tag{8}$$

Now consider the function $g(x_0, x_1, \ldots, x_n) = \mathsf{sign}(\sum_{i=0}^{n} w_i x_i)$ by adding the extra variable $x_0$. We claim that $\mathsf{Inf}_i(g) \leqslant 2\mathsf{Inf}_i(f)$, for all $1 \leqslant i \leqslant n$. Fix an $i \in [n]$. From Eq. (2) we know that $\mathsf{Inf}_i(g)$ equals

$$\frac{1}{2} \left( \Pr \left[ -|w_i| < w_0 + \sum_{1 \leqslant j \leqslant n:\ j \neq i} w_j x_j \leqslant |w_i| \right] + \Pr \left[ -|w_i| < -w_0 + \sum_{1 \leqslant j \leqslant n:\ j \neq i} w_j x_j \leqslant |w_i| \right] \right),$$

where the probabilities are uniform distribution over $x_1, \ldots, x_n \in \{+1, -1\}^n$. By relaxing the event in each case we have

$$\mathsf{Inf}_i(g) \leqslant \frac{1}{2} \left( \Pr \left[ -|w_i| \leqslant w_0 + \sum_{1 \leqslant j \leqslant n:\ j \neq i} w_j x_j \leqslant |w_i| \right] + \Pr \left[ -|w_i| \leqslant -w_0 + \sum_{1 \leqslant j \leqslant n:\ j \neq i} w_j x_j \leqslant |w_i| \right] \right).$$

It is easily seen that,

$$\Pr_{x \in \{+1,-1\}^n} \left[ -|w_i| \leqslant w_0 + \sum_{1 \leqslant j \leqslant n:\ j \neq i} w_j x_j \leqslant |w_i| \right] = \Pr_{x \in \{+1,-1\}^n} \left[ -|w_i| \leqslant -w_0 + \sum_{1 \leqslant j \leqslant n:\ j \neq i} w_j x_j \leqslant |w_i| \right] \ .$$

Indeed, there exists a 1-1 correspondence between $n$-bit strings satisfying the left hand side event and the right hand side event. Thus,

$$\mathsf{Inf}_i(g) \leqslant \Pr_{x \in \{+1,-1\}^n} \left[ -|w_i| \leqslant w_0 + \sum_{1 \leqslant j \leqslant n:\ j \neq i} w_j x_j \leqslant |w_i| \right] \leqslant 2 \cdot \mathsf{Inf}_i(f) \ ,$$

where the second inequality follows from (8).

Therefore, we have

$$\mathsf{Inf}(f) \geqslant \sum_{i=1}^{n} \frac{\mathsf{Inf}_i(g)}{2} = \frac{\mathsf{Inf}(g) - \mathsf{Inf}_0(g)}{2} \geqslant \frac{\mathsf{Inf}(g) - 1}{2} \ .$$

Thus we can translate a lower bound on influences in the homogeneous case to the non-homogeneous case.