Electronic Colloquium on Computational Complexity, Report No. 129 (2015)

On the entropy of a noisy function

Alex Samorodnitsky

Abstract

Let $0 < \epsilon < 1/2$ be a noise parameter, and let T_ϵ be the noise operator acting on functions on the boolean cube $\{0,1\}^n$. Let f be a nonnegative function on $\{0,1\}^n$. We upper bound the entropy of $T_\epsilon f$ by the average entropy of conditional expectations of f, given sets of roughly $(1-2\epsilon)^2 \cdot n$ variables.

As an application, we show that for a boolean function f, which is close to a characteristic function g of a subcube of dimension $n-1$, the entropy of $T_\epsilon f$ is at most that of $T_\epsilon g$.

This, combined with a recent result of Ordentlich, Shayevitz, and Weinstein shows that the "Most informative boolean function" conjecture of Courtade and Kumar holds for balanced boolean functions and high noise $\epsilon \geq 1/2 - \delta$, for some absolute constant $\delta > 0$.

Namely, if X is uniformly distributed in $\{0,1\}^n$ and Y is obtained by flipping each coordinate of X independently with probability ϵ, then, provided $\epsilon \geq 1/2 - \delta$, for any balanced boolean function f holds $I\Big(f(X);Y\Big) \leq 1 - H(\epsilon)$.

1 Introduction

This paper is motivated by the following conjecture of Courtade and Kumar [5].

Let (X,Y) be jointly distributed in $\{0,1\}^n$ such that their marginals are uniform and Y is obtained by flipping each coordinate of X independently with probability ϵ. Let H denote the binary entropy function $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$. The conjecture of [5] is:

Conjecture 1.1: For all boolean functions $f: \{0,1\}^n \to \{0,1\}$,

$$I\Big(f(X);Y\Big) \quad \leq \quad 1 \; - \; H(\epsilon)$$

∎

This inequality holds with equality if f is a characteristic function of a subcube of dimension $n-1$. Hence, the conjecture is that such functions are the "most informative" boolean functions.

We express $I(f(X);Y)$ in terms of the 'value of the entropy functional of the image of f under the noise operator' (all notions will be defined shortly). The question then becomes:

Which boolean functions with are the "stablest" under the action of the noise operator? That is, for which functions the entropy functional decreases the least under noise.

ISSN 1433-8092

One can also consider a more general question of how the noise operator affects the entropy of a nonnegative function.

Our main result is that for a nonnegative function f on $\{0,1\}^n$, the entropy of the image of f under the noise operator with noise parameter ϵ is upper bounded by the average entropy of conditional expectations of f, given sets of roughly $(1-2\epsilon)^2 \cdot n$ variables.

As an application, we show that characteristic functions of $(n-1)$-dimensional subcubes are at least as stable under the noise operator as functions which are close to them.

This, in conjunction with a recent result of [10], and a theorem of [3] that show that for high noise levels $\epsilon \sim 1/2$, balanced boolean functions which are potentially as stable as the characteristic functions of $(n-1)$-dimensional subcubes, have to be close to these functions, implies the validity of Conjecture 1.1 for balanced functions and high noise levels.

1.1 Entropy of nonnegative functions and the noise operator

We introduce some relevant notions.

For a nonnegative function $f:\ \{0,1\}^n \rightarrow \mathbb{R}$, we let the *entropy* of f to be defined as

$$Ent\Big(f\Big) \quad = \quad \mathop{\mathbb{E}}_{x} f(x) \log_2 f(x) \quad - \quad \mathop{\mathbb{E}}_{x} f(x) \cdot \log_2 \Big(\, \mathbb{E}\, f(x)\Big)$$

We note for future use that entropy is nonnegative, homogeneous $Ent\Big(\lambda f\Big) = \lambda \cdot Ent\Big(f\Big)$ and convex in f [6].

Given $0 \le \epsilon \le 1/2$, we define the *noise operator* acting on functions on the boolean cube as follows: for $f:\ \{0,1\}^n \rightarrow \mathbb{R}$, we let $T_\epsilon f$ at a point x be the expected value of f at y, where y is ϵ-correlated with x. That is,

$$\left(T_\epsilon f\right)(x) \quad = \quad \sum_{y \in \{0,1\}^n} \epsilon^{|y-x|} \cdot (1-\epsilon)^{n-|y-x|} \cdot f(y) \tag{1}$$

Here $|\cdot|$ denotes the Hamming distance.

Note that $T_\epsilon f$ is a convex combination of shifted copies of f. Hence, convexity of entropy implies that the noise operator decreases entropy. Our goal is to quantify this statement.

1.1.1 Connection between notions

Clearly, for a boolean function $f:\ \{0,1\}^n \rightarrow \{0,1\}$, and a random variable X uniformly distributed in $\{0,1\}^n$,

$$H\Big(f(X)\Big) \quad = \quad Ent\Big(f\Big) \quad + \quad Ent\Big(1-f\Big)$$

We also have the following simple claim (proved in the Appendix)

Lemma 1.2: *In the notation above, for a boolean function* $f:\ \{0,1\}^n \rightarrow \{0,1\}$,

$$I\Big(f(X);Y\Big) \quad = \quad Ent\Big(T_\epsilon f\Big) \;+\; Ent\Big(T_\epsilon(1-f)\Big)$$

Therefore, the conjecture above translates as follows:

Conjecture 1.3: (An equivalent form of Conjecture 1.1)

For any boolean function $f:\ \{0,1\}^n \rightarrow \{0,1\}$ holds

$$Ent\Big(T_\epsilon f\Big) \;+\; Ent\Big(T_\epsilon(1-f)\Big) \quad \le \quad 1 \;-\; H(\epsilon)$$

∎

1.2 Mrs. Gerber's function and Mrs. Gerber's lemma

We describe a result from information theory, and a related function, which will be important for us [1].

Let f_t be a function on the two-point space $\{0,1\}$, which is t at zero and $2-t$ at one. We have

$$Ent\Big(f_t\Big) \quad = \quad 1 \;-\; H\left(\frac{t}{2}\right)$$

Let $\phi(x,\epsilon)$ be a function on $[0,1]\times[0,1/2]$ defined as follows:

$$\phi(x,\epsilon) \quad = \quad Ent\Big(T_\epsilon f_t\Big) \tag{2}$$

where t is chosen so that $Ent\left(f_t\right) = x$.

This function was introduced in [13]. We will now describe some of its properties.

Note that ϕ is increasing in x, starting from zero at $x=0$.

In fact, it is easy to derive the following explicit expression for ϕ:

$$\phi(x,\epsilon) \quad = \quad 1 \;-\; H\Big((1-2\epsilon)\cdot H^{-1}(1-x) \;+\; \epsilon\Big)$$

A key property of ϕ is its concavity.

Theorem 1.4: *([13]) The function* $\phi(x,\epsilon)$ *is concave in* x *for any* $0 \le \epsilon \le 1/2$.

We mention a simple corollary.

[1] We are grateful to V. Chandar [2] for explaining the relevance of this result in connection to our previous work [11] on the subject.

Corollary 1.5: *For all* $0 \le \epsilon \le 1/2$,

$$\Big(1 - H(\epsilon)\Big) \cdot x \quad \le \quad \phi(x,\epsilon) \quad \le \quad (1-2\epsilon)^2 \cdot x \tag{3}$$

Proof: It's easy to check $\phi(0,\epsilon) = 0$ and $\phi(1,\epsilon) = 1 - H(\epsilon)$. And, it's easy to check that $\frac{\partial \phi}{\partial x}$ at $x = 0$ is $(1-2\epsilon)^2$. ∎

From now on, when the value of ϵ is clear from the context, we omit the second parameter in ϕ and write $\phi(x)$ instead of $\phi(x,\epsilon)$.

We now describe an inequality of [13], which is known as Mrs. Gerber's lemma. Following this usage, we will refer to the function ϕ as Mrs. Gerber's function.

This inequality upperbounds the entropy of the image of a nonnegative function under the action of the noise operator. We present it in terms of the entropy functional and the noise operator[2].

Theorem 1.6: *([13]) Let f be a nonnegative function on* $\{0,1\}^n$. *Then*

$$Ent\Big(T_\epsilon f\Big) \quad \le \quad n\,\mathbb{E}\,f \cdot \phi\left(\frac{Ent(f)}{n\,\mathbb{E}\,f},\ \epsilon\right) \tag{4}$$

1.3 Main results

For $A \subseteq [n]$, and for a nonnegative function $f: \{0,1\}^n \to \mathbb{R}$, we denote

$$Ent\Big(f \mid A\Big) \quad = \quad Ent\Big(\mathbb{E}\left(f \;\Big|\; \{x_i\}_{i \in A}\right)\Big)$$

Here $\mathbb{E}$ is the conditional expectation operator. That is, $\mathbb{E}\Big(f \mid A\Big)$ is the function of the variables $\{x_i\}_{i\in A}$, defined as the expectation of f given the values of $\{x_i\}$.[3]

Our main result states that the entropy of a nonnegative function f under noise is upper bounded by the average entropy of the conditional expectations of f, given subsets of variables of a certain size.

Theorem 1.7: *Let f be a nonnegative function on the cube with expectation* 1.
Let $0 < \epsilon < 1$ *be a noise parameter. Let* $\lambda = (1-2\epsilon)^2$. *Assume* $n \ge 10 \cdot \left(\frac{1}{(1-\lambda^2)} \cdot \ln\left(\frac{1}{1-\lambda}\right)\right)$.
Let $v = \left\lceil \lambda \cdot n + \sqrt{n \ln n} \right\rceil$.

[2] As pointed out to us by Chandar [2], this is equivalent to the standard information-theoretic formulation $H\Big(T_\epsilon f\Big) \ge H\left(\epsilon + (1-2\epsilon)\cdot H^{-1}\left(\frac{H(f)}{n}\right)\right)$.

[3] We also may (and will) view $\mathbb{E}\Big(f \mid A\Big)$ as a function on $\{0,1\}^n$, which depends only on the variables with indices in A.

Then, we have

$$Ent\Big(T_\epsilon f\Big) \quad \leq \quad \mathop{\mathbb{E}}_{|B|=v} \left(Ent\Big(f \mid B\Big) \; - \; \sum_{i \in B} Ent\Big(f \mid \{i\}\Big)\right) \;\; + \;\; \sum_{i=1}^{n} \phi\left(Ent\Big(f \;\Big|\; \{i\}\Big)\right) \;\; +$$

$$O\left(\sqrt{\frac{\ln n}{n}}\right) \cdot Ent\Big(f\Big)$$

The asymptotic notation in the error term hides an absolute constant.

We discuss this claim in the next subsection, where we suggest it might be viewed as a strengthening of Mrs. Gerber's inequality (4), at least for functions with small entropy.

Applying the inequality $\phi(x, \epsilon) \leq (1 - 2\epsilon)^2 \cdot x$ (see (3)) to the claim of the theorem, gives the following, more streamlined corollary. (However, the somewhat stronger claim of the theorem is needed for the applications below.)

Corollary 1.8: *In the notation of Theorem 1.7,*

$$Ent\Big(T_\epsilon f\Big) \quad \leq \quad \mathop{\mathbb{E}}_{|B|=v} Ent\Big(f \mid B\Big) \quad + \quad O\left(\sqrt{\frac{\ln n}{n}}\right) \cdot Ent\Big(f\Big)$$

Specializing to boolean functions, this implies the following claim.

Corollary 1.9: *In the notation of Conjecture 1.1 and of Theorem 1.7, for a boolean function* $f: \{0,1\}^n \rightarrow \{0,1\}$ *holds*

$$I\Big(f(X); Y\Big) \quad \leq \quad \mathop{\mathbb{E}}_{|B|=v} I\Big(f(X);\; \{X_i\}_{i \in B}\Big) \quad + \quad O\left(\sqrt{\frac{\ln n}{n}}\right)$$

Remark 1.10: Corollary 1.9 implies that, roughly speaking,

$$I\Big(f(X); Y\Big) \quad \lesssim \quad \mathop{\mathbb{E}}_{|B|=(1-2\epsilon)^2 \cdot n} I\Big(f(X);\; \{X_i\}_{i \in B}\Big)$$

As pointed out by Or Ordentlich [9], it seems instructive to compare this bound to the weaker bound

$$I\Big(f(X); Y\Big) \quad \lesssim \quad \mathop{\mathbb{E}}_{|B|=(1-2\epsilon) \cdot n} I\Big(f(X);\; \{X_i\}_{i \in B}\Big)$$

which can be obtained by the following information-theoretic argument.

An equivalent way to obtain Y from X is to replace each coordinate of X independently with a random bit, with probability 2ϵ .

Let S be the set of indices where the input bits were replaced with random bits. Using the chain rule of mutual information we have

$$I\Big(f(X);Y\Big) \quad = \quad I\Big(f(X);Y,S\Big) \; - \; I\Big(f(X);S \mid Y\Big) \quad = \quad I\Big(f(X);Y \mid S\Big) \; - \; I\Big(f(X);S \mid Y\Big)$$

where the last equality follows since $I\Big(f(X);S\Big) = 0$.

In particular, by non-negativity of mutual information

$$I\Big(f(X);Y\Big) \quad \le \quad I\Big(f(X);Y \mid S\Big) \quad \approx \quad \mathbb{E}_{|B|=(1-2\epsilon)\cdot n} \; I\Big(f(X);\{X_i\}_{i\in B}\Big)$$

∎

As an application of Theorem 1.7, we prove the following result.

Theorem 1.11: *There exists an absolute constant* $\delta > 0$ *such that for any noise* $\epsilon \ge 0$ *with* $(1-2\epsilon)^2 \le \delta$ *and for any boolean function* $f: \{0,1\}^n \to \{0,1\}$ *such that*

- $\frac{1}{2} - \delta \le \mathbb{E} f \le \frac{1}{2}$;
- *There exists* $1 \le k \le n$ *such that* $|\widehat{f}(\{k\})| \ge (1-\delta) \cdot \mathbb{E} f$

Holds

$$Ent\Big(T_\epsilon f\Big) \quad \le \quad \frac{1}{2} \cdot \Big(1 - H(\epsilon)\Big)$$

A simple corollary of this theorem, taken together with a recent result of [10] and a theorem of [3], is the validity of Conjecture 1.1 for *balanced* boolean functions on the cube, provided the noise parameter is close to 1/2.

Theorem 1.12: *There exists an absolute constant* $\delta > 0$ *such that for any noise* $\epsilon \ge 0$ *with* $(1-2\epsilon)^2 \le \delta$ *and for any boolean function* $f: \{0,1\}^n \to \{0,1\}$ *with expectation* $1/2$ *holds*

$$I\Big(f(X);Y\Big) \quad \le \quad 1 \; - \; H(\epsilon)$$

1.4 More on Theorem 1.7

In this subsection we give a high-level description of the proof of the theorem and argue that its claim may be viewed as a strengthening of Mrs. Gerber's lemma.

Notation: For a direction $1 \le i \le n$ we define the noise operator in direction i as follows:

$$\left(T_{\epsilon_{\{i\}}} f\right)(x) \quad = \quad \epsilon \cdot f\left(x + e_i\right) \quad + \quad (1-\epsilon) \cdot f(x)$$

where e_i is the i^{th} unit vector. The operators $\left\{T_{\epsilon_{\{i\}}}\right\}$ commute and, for $R \subseteq [n]$, we define T_{ϵ_R} to be the composition of $T_{\epsilon_{\{i\}}}$, $i \in R$. Note that the noise operator T_ϵ would be written in this notation as $T_{\epsilon_{[n]}}$.

We start with the proof of (4). Since both sides of the inequality are homogeneous in f, we may assume $\mathbb{E} f = 1$.

By the chain rule for entropy, we have, for any $\sigma \in S_n$ that

$$Ent\left(T_\epsilon f\right) \quad = \quad \sum_{i=1}^n \left(Ent\left(T_\epsilon f \mid \{\sigma(1), \ldots, \sigma(i)\}\right) \; - \; Ent\left(T_\epsilon f \mid \{\sigma(1), \ldots, \sigma(i-1)\}\right)\right) \quad =$$

$$\sum_{i=1}^n \left(Ent\left(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i)\}}} f \mid \{\sigma(1), \ldots, \sigma(i)\}\right) \; - \; Ent\left(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}} f \mid \{\sigma(1), \ldots, \sigma(i-1)\}\right)\right) \quad \le$$

$$\sum_{i=1}^n \phi\left(Ent\left(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}} f \mid \{\sigma(1), \ldots, \sigma(i)\}\right) - Ent\left(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}} f \mid \{\sigma(1), \ldots, \sigma(i-1)\}\right)\right) \qquad (5)$$

Let us explain the last inequality. Let $y \in \{0,1\}^{i-1}$. Let $\tilde{f}_y$ be a function on $\{0,1\}$ defined by the restriction of the function $\mathbb{E}\left(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}} f \mid \{\sigma(1), \ldots, \sigma(i)\}\right)$, which we view as a function on the i-dimensional cube, to the points in which the coordinates $\sigma(k)$, $k = 1, ..., i-1$ are set to be y_k. Then, it is easy to see that

$$Ent\left(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i)\}}} f \mid \{\sigma(1), \ldots, \sigma(i)\}\right) \; - \; Ent\left(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i)\}}} f \mid \{\sigma(1), \ldots, \sigma(i-1)\}\right) \quad =$$

$$\mathop{\mathbb{E}}_{y} Ent\left(T_\epsilon \tilde{f}_y\right) \quad = \quad \mathop{\mathbb{E}}_{y}\left(\mathbb{E}\tilde{f}_y \cdot \phi\left(Ent\left(\frac{\tilde{f}_y}{\mathbb{E}\tilde{f}_y}\right)\right)\right) \quad \le \quad \phi\left(\mathop{\mathbb{E}}_{y} Ent\left(\tilde{f}_y\right)\right) \quad =$$

$$\phi\left(Ent\left(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}} f \mid \{\sigma(1), \ldots, \sigma(i)\}\right) \; - \; Ent\left(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}} f \mid \{\sigma(1), \ldots, \sigma(i-1)\}\right)\right)$$

The first equality in the second row follows from (2) and the linearity of entropy. The inequality follows from concavity of the function ϕ and the fact that $\mathbb{E}_y \mathbb{E}\tilde{f}_y = \mathbb{E}\left(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i)\}}} f \mid \{\sigma(1), \ldots, \sigma(i)\}\right)$ $= \mathbb{E} f = 1$.

We now continue from (5).

For $y \in \{0,1\}^{i-1}$, let f_y be a function on $\{0,1\}$ defined by the restriction of the function $\mathbb{E}\Big(f \mid \{\sigma(1),\ldots,\sigma(i)\}\Big)$ to the points in which the coordinates $\sigma(k)$, $k = 1, ..., i-1$ are set to be y_k.

Since the noise operator $T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}}$ is stochastic, the functions $\Big\{\tilde{f}_y\Big\}$ are a stochastic mixture of the functions $\Big\{f_y\Big\}$. Hence, since the *Ent* functional is convex, for any $0 \le \epsilon \le 1$ holds

$$Ent\Big(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}}f \mid \{\sigma(1),\ldots,\sigma(i)\}\Big) - Ent\Big(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}}f \mid \{\sigma(1),\ldots,\sigma(i-1)\}\Big) =$$

$$\mathop{\mathbb{E}}_{y} Ent\Big(\tilde{f}_y\Big) \quad \le \quad \mathop{\mathbb{E}}_{y} Ent\Big(f_y\Big) \quad = \tag{6}$$

$$Ent\Big(f \mid \{\sigma(1),\ldots,\sigma(i)\}\Big) - Ent\Big(f \mid \{\sigma(1),\ldots,\sigma(i-1)\}\Big)$$

And hence (5) is upper bounded by

$$\sum_{i=1}^{n} \phi\bigg(Ent\Big(f \mid \{\sigma(1),\ldots,\sigma(i)\}\Big) - Ent\Big(f \mid \{\sigma(1),\ldots,\sigma(i-1)\}\Big)\bigg) \le n \cdot \phi\left(\frac{Ent\Big(f\Big)}{n}\right)$$

where in the last inequality the concavity of ϕ is used again.

1.4.1 Our improvement

We attempt to quantify the loss in inequality (6).

Let us introduce some notation. For a nonnegative function g on the cube, for a subset $A \subset [n]$ and for an element $m \notin A$, we define

$$I_g(A, m) \quad = \quad Ent\Big(g \mid A \cup \{m\}\Big) - Ent\Big(g \mid A\Big) - Ent\Big(g \mid \{m\}\Big)$$

By supermodularity of the entropy functional, this quantity is always nonnegative. In fact, it is easily seen to be proportional to the mutual information between $\Big\{X_j\Big\}_{j \in A}$ and X_m, provided the random variable $X = \Big(X_1, ..., X_n\Big)$ is distributed on $\{0,1\}^n$ according to the distribution $P_g = g / \sum g$.

Coming back to (6), observe that $Ent\Big(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}}f \mid \{\sigma(i)\}\Big) = Ent\Big(f \mid \{\sigma(i)\}\Big)$.

Hence, taking $A = \{\sigma(1),\ldots,\sigma(i-1)\}$ and $m = \{\sigma(i)\}$, the decrease in (6) is from $I_f(A, m)$ to $I_{T_{\epsilon_A} f}(A, m)$. Hence, our goal amounts to quantifying the decrease in mutual information in the presence of noise.

In the next two sections we consider a slightly more general question of upper bounding $I_{T_{\epsilon_A} f}(A, m)$, given f, A, and m. In Section 2 we upper bound $I_{T_{\epsilon_A} f}(A, m)$ by the value of a certain linear program. In Section 3 we introduce a symmetric version of this program and a symmetric solution for the symmetric program, and show its value to be at least as large as that of the original program.

We then find the value of the symmetric solution, as a function of f, A, and m. This value provides an upper bound on the noisy mutual information.

In Section 4 we apply this bound in (6), averaging the chain rule for the entropy of $T_\epsilon f$ over all permutations $\sigma \in S_n$. This proves Theorem 1.7.

The improved bound in (6) is the reason we suggest to view this theorem as a strengthening of Mrs. Gerber's lemma.

On the other hand, strictly speaking, this line of argument does not provide a direct improvement of (4), since in the averaging step in Section 4 we have to replace $\phi(x, \epsilon)$ by a larger linear function $(1 - 2\epsilon)^2 \cdot x$, in order to be able to come up with manageable estimates.

With that, the two functions $\phi(x, \epsilon)$ and $(1 - 2\epsilon)^2 \cdot x$ almost coincide for small values of x, and, loosely speaking, if the entropy of f, the function in Theorem 1.7, is not too large, as is the case for boolean functions, say, all the arguments of ϕ should lie very close to zero, meaning not much lost in the linear approximation.

Organization of the paper

This paper is organized as follows. The proof of Theorem 1.7 is given in Sections 2 to 4. Theorem 1.11 is proved in Section 5. The remaining proofs are presented in Section 6.

2 A linear programming bound for noisy mutual information

In this section we upper bound the noisy mutual information $I_{T_{\epsilon_A} f}(A, m)$ by the value of a certain linear program.

Let f be a nonnegative function on the cube. Let A be a subset of $[n]$ and let $m \notin A$.

Let $|A| = k$. We will assume, without loss of generality, that $A = [k]$ and that $m = k + 1$.

Notation: From now on, we write λ for $(1 - 2\epsilon)^2$.

Consider the following linear optimization problem.

Optimization problem

Boundary data: For $S \subseteq [k]$ and for $i \in S$, we write

$$y_{S,i} \;=\; Ent\Big(f \mid S \cup \{k+1\}\Big) \;-\; Ent\Big(f \mid S \setminus \{i\} \cup \{k+1\}\Big) \;-\; Ent\Big(f \mid S\Big) \;+\; Ent\Big(f \mid S \setminus \{i\}\Big)$$

The numbers $\left\{y_{S,i}\right\}$ are the boundary data for this problem [4].

Variables: $x^R_{S,i}$ for $R, S \subseteq [k]$ and $i \in S$.

The optimization problem: Given the boundary data, we want to upper bound μ, where

$$\mu \quad = \quad \text{Max} \quad \sum_{i=1}^{k} x^{[k]}_{\{1,\ldots,i\};\ i} \tag{7}$$

under the following constraints.

Constraints:

1. $$x^{\emptyset}_{S,i} \quad = \quad y_{S,i}$$
2. $$x^{R}_{S,i} \quad = \quad x^{R\cap S}_{S,i}$$
3. For all $\sigma, \tau \in S_k$ holds
$$\sum_{i=1}^{k} x^{R}_{\{\sigma(1),\ldots,\sigma(i)\},\ \sigma(i)} \quad = \quad \sum_{i=1}^{k} x^{R}_{\{\tau(1),\ldots,\tau(i)\},\ \tau(i)}$$
4. If $i \in R$ then
$$x^{R}_{S,i} \quad \leq \quad \lambda \cdot \left(x^{R\setminus i}_{S,i}\right)$$

We then have the following claim.

Theorem 2.1: *The noisy mutual information $I_{T_{\epsilon_{[k]}} f}\Big([k], k+1\Big)$ is upperbounded by the value of the optimization problem (7).*

Proof:

First, consider the boundary data. We claim that for any permutation $\sigma \in S_k$ holds

$$\sum_{i=1}^{k} y_{\{\sigma(1),\ldots,\sigma(i)\},\ \sigma(i)} \quad = \quad I_f\Big([k], k+1\Big) \tag{8}$$

In fact, it is easy to see that the LHS is a telescopic sum, summing to

$$Ent\Big(f \mid [k+1]\Big) \quad - \quad Ent\Big(f \mid [k]\Big) \quad - \quad Ent\Big(f \mid \{k+1\}\Big) \quad = \quad I_f\Big([k], k+1\Big)$$

Next we define a feasible solution for (7) whose value is $I_{T_{\epsilon_{[k]}} f}\Big([k], k+1\Big)$.

[4] Note that $y_{S,i} \geq 0$ for all S and i. This follows from supermodularity of the entropy functional. In fact, the value of $y_{S,i}$ is proportional to the mutual information between i and $k+1$, given $S \setminus \{i\}$.

Fix $R \subseteq [k]$. Write f^R for $T_{\epsilon_R} f$. For $S \subseteq [k]$ and $i \in S$ set

$$x^R_{S,i} \; = \; Ent\Big(f^R \mid S \cup \{k+1\}\Big) - Ent\Big(f^R \mid S \setminus \{i\} \cup \{k+1\}\Big) - Ent\Big(f^R \mid S\Big) + Ent\Big(f^R \mid S \setminus \{i\}\Big)$$

Clearly, $x^{\emptyset}_{S,i} = y_{S,i}$ and hence the first constraint of the program is satisfied.

As above, for any permutation $\sigma \in S_k$ holds

$$\sum_{i=1}^{k} x^R_{\{\sigma(1),\ldots,\sigma(i)\},\ \sigma(i)} \quad = \quad I_{T_{\epsilon_R} f}\Big([k], k+1\Big)$$

Hence, the third constraint is satisfied as well.

In particular,

$$\sum_{i=1}^{k} x^{[k]}_{\{1,\ldots,i\},\ i} \quad = \quad I_{T_{\epsilon_{[k]}} f}\Big([k], k+1\Big)$$

so, the value given by this solution is indeed $I_{T_{\epsilon_{[k]}} f}\Big([k], k+1\Big)$.

We continue to prove its feasibility. We claim that for any $A \subseteq [k]$ holds $Ent\Big(f^R \mid A\Big) \; = \; Ent\Big(f^{R\cap A} \mid A\Big)$.

To see this, note that the noise operators commute with the conditional expectation operators, and hence

$$\mathbb{E}\left(T_{\epsilon_R} f \mid A\right) \; = \; T_{\epsilon_R}\, \mathbb{E}\left(f \mid A\right) \; = \; T_{\epsilon_{R\setminus A}} T_{\epsilon_{R\cap A}}\, \mathbb{E}\left(f \mid A\right) \; = \; T_{\epsilon_{R\cap A}}\, \mathbb{E}\left(f \mid A\right) \; = \; \mathbb{E}\left(T_{\epsilon_{R\cap A}} f \mid A\right)$$

Hence, by definition, $x^R_{S,i} = x^{R\cap S}_{S,i}$ for any $R, S \subseteq [k]$, and the second constraint holds.

To conclude the proof of the theorem, it remains to show that for any $R \subseteq S \subseteq [k]$ and $i \in R$ holds

$$x^R_{S,i} \quad \le \quad \lambda \cdot \left(x^{R\setminus i}_{S,i}\right) \tag{9}$$

This requires a somewhat longer proof, which will be done in a separate subsection.

2.1 Proof of (9)

We start with the following technical claim, which will be proved in Section 6.4

Proposition 2.2: *Let h be a nonnegative function on $\{0,1\}^2$ Then*

$$I_{T_{\epsilon_{\{1\}}} h}\left(\{2\}, 1\right) \quad \le \quad \lambda \cdot I_h\left(\{2\}, 1\right)$$

We observe that this claim, with the appropriate modification of indices, immediately implies (9) in the case S is a singleton.

Indeed, in this case $S = R = \{i\}$, and (9) reduces to

$$Ent\Big(T_{\epsilon_{\{i\}}} f \mid \{i, k+1\}\Big) \; - \; Ent\Big(T_{\epsilon_{\{i\}}} f \mid \{i\}\Big) \; - \; Ent\Big(T_{\epsilon_{\{i\}}} f \mid \{k+1\}\Big) \quad \le$$

$$\lambda \cdot \left(Ent\Big(f \mid \{i, k+1\}\Big) \; - \; Ent\Big(f \mid \{i\}\Big) \; - \; Ent\Big(f \mid \{k+1\}\Big)\right)$$

Set $h = \mathbb{E}\left(f \mid \{i, k+1\}\right)$. This is a function of two variables i and $k+1$, that is on a 2-dimensional cube.

Note that, by definition, $Ent\Big(f \mid \{i,k+1\}\Big) - Ent\Big(f \mid \{i\}\Big) - Ent\Big(f \mid \{k+1\}\Big) = I_h\Big(\{k+1\}, i\Big)$.

Similarly, $Ent\Big(T_{\epsilon_{\{i\}}} f \mid \{i,k+1\}\Big) - Ent\Big(T_{\epsilon_{\{i\}}} f \mid \{i\}\Big) - Ent\Big(T_{\epsilon_{\{i\}}} f \mid \{k+1\}\Big) = I_{T_{\epsilon_{\{i\}}} h}\Big(\{k+1\}, i\Big)$.

Hence the inequality we need to show is equivalent to the claim of the proposition applied to h, and renumbering i with 1 and $k+1$ with 2.

Let now $|S| > 1$ and let $i \in R \subseteq S$. Set $g = \mathbb{E}\left(f^{R \setminus \{i\}} \;\Big|\; S \cup \{k+1\}\right)$. Since g depends only on the coordinates in $S \cup \{k+1\}$, we may (and will) view g as a function on the appropriate cube, which we denote by $\{0,1\}^{S \cup \{k+1\}}$.

For each $y \in \{0,1\}^{S \setminus \{i\}}$, let h_y be the function on the 2-dimensional cube obtained by restricting g to the points whose restriction to $S \setminus \{i\}$ is y. Note that the following three simple identities hold

$$\mathop{\mathbb{E}}_{y} Ent\Big(h_y\Big) \;=\; Ent\Big(f^{R\setminus\{i\}} \;\Big|\; S \cup \{k+1\}\Big) - Ent\Big(f^{R\setminus\{i\}} \;\Big|\; S \setminus \{i\}\Big)$$

$$\mathop{\mathbb{E}}_{y} Ent\Big(h_y \mid \{k+1\}\Big) \;=\; Ent\Big(f^{R\setminus\{i\}} \;\Big|\; S \setminus \{i\} \cup \{k+1\}\Big) \; - \; Ent\Big(f^{R\setminus\{i\}} \;\Big|\; S \setminus \{i\}\Big)$$

$$\mathop{\mathbb{E}}_{y} Ent\Big(h_y \mid \{i\}\Big) \;=\; Ent\Big(f^{R\setminus\{i\}} \;\Big|\; S\Big) \; - \; Ent\Big(f^{R\setminus\{i\}} \;\Big|\; S \setminus \{i\}\Big)$$

Combining these identities gives $x_{S,i}^{R\setminus\{i\}} \;=\; \mathbb{E}_y \; I_{h_y}\Big(\{k+1\}, i\Big)$.

Similarly, $x_{S,i}^{R} \;=\; \mathbb{E}_y \; I_{T_{\epsilon_{\{i\}}} h_y}\Big(\{k+1\}, i\Big)$.

Applying the claim of the proposition to each h_y and averaging over y we obtain

$$x_{S,i}^{R} \quad = \quad \mathop{\mathbb{E}}_{y} \; I_{T_{\epsilon_{\{i\}}} h_y}\Big(\{k+1\}, i\Big) \quad \le \quad \lambda \cdot \mathop{\mathbb{E}}_{y} I_{h_y} \; \Big(\{k+1\}, i\Big) \quad = \quad \lambda \cdot \Big(x_{S,i}^{R\setminus i}\Big)$$

This completes the proof of (9) and of the theorem.

∎

3 The optimization problem and its symmetric version

Let $k \geq 1$. Consider the optimization problem (7). In this section, we introduce a symmetric version of this problem and a specific symmetric feasible solution for the symmetric problem. We will then argue that the value of this solution for the symmetric problem is at least as large as the optimal value for the original problem. Hence this value would provide an upper bound on the noisy mutual information.

3.1 The symmetric problem and solution

Let $\left\{x_{S,i}^R\right\}$ be a feasible solution to the optimization problem (7) with boundary data $\left\{y_{S,i}\right\}$. We define the numbers $y_1, \ldots, y_k$ as follows. For $1 \leq s \leq k$ let

$$y_s \quad = \quad \mathop{\mathbb{E}}_{(S,i)} y_{S,i} \tag{10}$$

where the expectation is taken over all pairs (S, i) such that $|S| = s$ and $i \in S$.

For $0 \leq r < s \leq k$ we define x_s^r recursively in the following manner:

$$x_s^r \quad = \quad \begin{cases} y_s & \text{if} \qquad r = 0 \\ \lambda \cdot x_s^{r-1} \; + \; (1-\lambda) \cdot x_{s-1}^{r-1} & \text{otherwise} \end{cases} \tag{11}$$

We now define the *symmetric version* of (7), by replacing the boundary data by a new, symmetric one. We set, for all $i \in S \subseteq [k]$ with $|S| = s$:

$$\bar{y}_{S,i} \quad = \quad y_s$$

Next we define the *symmetric solution* for the symmetric problem, in the following way. For $R \subseteq S$, we set

$$\bar{x}_{S,i}^R \quad = \quad \begin{cases} \lambda \cdot x_s^{r-1} & \text{if} \qquad i \in R \\ x_s^r & \text{otherwise} \end{cases}$$

and for general R, S we set

$$\bar{x}_{S,i}^R \quad = \quad \bar{x}_{S,i}^{R \cap S}$$

Proposition 3.1: *The solution above is a feasible solution of the symmetric version of (7). Moreover, for any $R \subseteq [k]$ of cardinality r and for any $\tau \in S_k$ holds*

$$\sum_{i=1}^{k} \bar{x}_{\{\tau(1),\ldots,\tau(i)\},\tau(i)}^R \quad = \quad \sum_{j=1}^{k-r} y_j \quad + \quad \lambda \cdot \sum_{t=0}^{r-1} x_{k-r+t+1}^t \tag{12}$$

Proof:

The constraints 1 and 2 of (7) hold, by the definition of $\bar{x}^R_{S,i}$. We pass to constraint 4. Clearly, because of constraint 2, it suffices to prove it for $R \subseteq S$. In this case, taking $i \in R$, we have, by the definition of $\bar{x}^R_{S,i}$

$$\bar{x}^R_{S,i} \quad = \quad \lambda \cdot x_s^{r-1} \quad = \quad \lambda \cdot \bar{x}^{R\setminus\{i\}}_{S,i}$$

Next, we note that (12) will imply validity of constraint 4, since the RHS of (12) does not depend on τ.

It remains to prove (12). Let $i_1 < i_2 < ... < i_r$ be such that $R = \{\tau(i_1), \tau(i_2), ..., \tau(i_r)\}$. Then

$$\sum_{i=1}^{k} \bar{x}^R_{\{\tau(1),...,\tau(i)\},\tau(i)} \quad = \quad \sum_{j=1}^{i_1-1} \; + \; \sum_{j=i_1}^{i_2-1} \; + \; \dots \; + \; \sum_{j=i_r}^{k} \quad =$$

$$\sum_{j=1}^{i_1-1} y_j \quad + \quad \left(\lambda \cdot y_{i_1} + \sum_{j=i_1+1}^{i_2-1} x_j^1\right) \quad + \quad \left(\lambda \cdot x_{i_2}^1 + \sum_{j=i_2+1}^{i_3-1} x_j^2\right) \quad + \dots \left(\lambda \cdot x_{i_r}^{r-1} + \sum_{j=i_r+1}^{k} x_j^r\right)$$

Expanding $x_s^t = \lambda \cdot x_s^{t-1} + (1-\lambda) \cdot x_{s-1}^{t-1}$, we have the following exchange rule:

Two adjacent summands of the form $\lambda \cdot x_j^t + x_{j+1}^{t+1}$ can always be replaced by $x_j^t + \lambda \cdot x_{j+1}^t$. Applying this the appropriate number of times in each bracket, transforms the expression above into

$$\sum_{j=1}^{i_1-1} y_j \quad + \quad \left(\sum_{j=i_1}^{i_2-2} y_j + \lambda \cdot y_{i_2-1}\right) \quad + \quad \left(\sum_{j=i_2}^{i_3-2} x_j^1 + \lambda \cdot x_{i_3-1}^1\right) \quad + \dots \left(\sum_{j=i_r}^{k-1} x_j^{r-1} + \lambda \cdot x_k^{r-1}\right)$$

Next we observe that the following rules apply in the original ordering of the summands: To the right of x_j^t is always either x_{j+1}^t or $\lambda \cdot x_{j+1}^t$. To the right of $\lambda \cdot x_s^r$ is always either x_{s+1}^{r+1} or $\lambda \cdot x_{s+1}^{r+1}$.

Moreover, this is easily verified to be preserved by the exchange rule above, by checking the four arising cases.

This means that applying the exchange rule as many times as needed, we can ensure all the summands multiplied by λ to be on the last r places on the right. Since the first summand is always either y_1 or $\lambda \cdot y_1$, these invariants guarantee that by doing so we obtain (12).

∎

3.2 Optimality of the symmetric solution

Theorem 3.2: *Let* $\left\{x_{S,i}^R\right\}$ *be a feasible set of solutions to the linear optimization problem (7). Let* $\bar{x}_{S,i}^R$ *be the symmetric solution for the symmetric version of this problem.*
Then, for any $0 \le r \le k$ *holds:*

$$\mathop{\mathbb{E}}_{|R|=r} \sum_{i=1}^{k} x_{\{1,\ldots,i\},i}^R \quad \le \quad \mathop{\mathbb{E}}_{|R|=r} \sum_{i=1}^{k} \bar{x}_{\{1,\ldots,i\},i}^R$$

Corollary 3.3: *The optimal value of (7) is upper bounded by the value of the symmetric solution to the symmetric version of the problem. This value is given by*

$$\lambda \cdot \sum_{t=0}^{k-1} x_{t+1}^t$$

Proof: Apply the theorem with $r = k$ and use (12). ∎

Proof: (Of the theorem).

We proceed by double induction - on k and on $0 \le r \le k$. For $k = 1$ the claim is easily seen to be true.

Note also that the claim is true for any k and $r = 0$. This follows from constraints 1 and 3 of the linear program (7) and the definition of the symmetric boundary data. In fact, we have

$$\sum_{j=1}^{k} y_{\{1,\ldots,j\},j} \quad = \quad \mathop{\mathbb{E}}_{\sigma\in S_k} \sum_{j=1}^{k} y_{\{\sigma(1),\ldots,\sigma(j)\},\ \sigma(j)} \quad = \quad \sum_{j=1}^{k} \mathop{\mathbb{E}}_{\sigma\in S_k} y_{\{\sigma(1),\ldots,\sigma(j)\},\ \sigma(j)} \quad =$$

$$\sum_{j=1}^{k} \mathop{\mathbb{E}}_{|S|=j,\ i\in S} y_{S,i} \quad = \quad \sum_{j=1}^{k} y_j \quad = \quad \sum_{j=1}^{k} \bar{y}_{\{1,\ldots,j\},j}$$

Let now numbers r and k, with $0 < r \le k$ be given. Assume the claim holds for $k-1$, and also for k, for all $0 \le t \le r-1$. We will argue it also holds for k and r.

We start with some simple properties of the linear program (7). We assume to be given the boundary data and a specific feasible solution to (7), and the symmetric solution to the symmetric version of (7), as in Theorem 3.2.

Lemma 3.4: *Let* $M \subseteq [k]$*. Let* $\left\{y_{K,i}\right\}_{i\in K\subseteq M}$ *be the restriction of the boundary data to subsets of* M*. For* $R \subseteq M$*, let* $\left\{x_{K,i}^R\right\}_{i\in K\subseteq M}$ *be the restriction of the feasible solution to subsets of* M*. Then* $\left\{x_{K,i}^R\right\}_{i\in K\subseteq M}$ *is a feasible solution to the appropriate (smaller) optimization problem on* M*.*

Proof:

Constraints 1, 2, and 4 are easy to check. As for constraint 3, let σ and τ be two permutations from M to itself. Extend them in the same way to permutations σ' and τ' on $[k]$. It is then easy to see that constraint 3 holds for σ and τ in the smaller problem, since it holds for σ' and τ' in the larger one.

∎

Lemma 3.5: *Let $M \subseteq [k]$, with $|M| = m$ and let $R \subseteq [k]$. Let τ be a bijection from $[m]$ to M. Let*

$$F\Big(M, R, \tau\Big) \quad = \quad \sum_{j=1}^{m} \bar{x}^{R}_{\{\tau(1),\ldots,\tau(j)\},\ \tau(j)}$$

Then $F\Big(M, R, \tau\Big)$ depends only on m and $|R \cap M|$.

Proof:

Since the symmetric solution $\Big\{\bar{x}^{R}_{S,i}\Big\}$ satisfies constraint 2 of (7), we have

$$F\Big(M, R, \tau\Big) \quad = \quad \sum_{j=1}^{m} \bar{x}^{R}_{\{\tau(1),\ldots,\tau(j)\},\ \tau(j)} \quad = \quad \sum_{j=1}^{m} \bar{x}^{R\cap M}_{\{\tau(1),\ldots,\tau(j)\},\ \tau(j)} \quad = \quad F\Big(M, R\cap M, \tau\Big)$$

Let $r = |R \cap M|$.

Proceeding exactly as in the proof of Proposition 3.1, we get that

$$F\Big(M, R\cap M, \tau\Big) \quad = \quad \sum_{j=1}^{m-r} y_j \quad + \quad \lambda \cdot \sum_{t=0}^{r-1} x^{t}_{m-r+t+1}$$

That is, $F\Big(M, R, \tau\Big)$ depends only on m and $r = |R \cap M|$, as claimed. ∎

Next, we introduce some notation.

3.2.1 Notation

1. Let $M \subseteq [k]$. Let $\Big\{y_{K,i}\Big\}_{i\in K\subseteq M}$ be the restriction of the boundary data to the subsets of M.

   We will denote by $\Big\{\mathcal{S}_M\Big[x^{R}_{K,i}\Big]\Big\}$ the symmetric solution to the symmetric version of the smaller problem with this boundary data.

2. Let $L \subseteq [k]$, with $L = \{i_1, ..., i_\ell\}$, so that $i_1 < i_2 < ... < i_\ell$. Let $R \subseteq [k]$. Write

$$\mu^R(L) \quad = \quad \sum_{j=1}^{\ell} x^R_{\{i_1,\dots,i_j\},\ i_j}$$

For $L \subseteq M \subseteq [k]$, and $R \subseteq M$, we denote

$$\mathcal{S}[\mu]^R_M(L) \quad = \quad \sum_{j=1}^{\ell} \mathcal{S}_M\Big[x^R_{\{i_1,\dots,i_j\},\ i_j}\Big]$$

Note that this quantity depends on M. With that, by Lemmas 3.4 and 3.5, given M, it depends only on the cardinalities $|L|$ and $|R \cap L|$.

3. Using the observation in the preceding paragraph, given $R \subseteq L \subseteq M \subseteq [k]$, with $|L| = \ell$, and $|R| = r$, we may also write $\mathcal{S}[\mu]^r_M\Big(\ell\Big)$ for $\mathcal{S}[\mu]^R_M(L)$.

   In particular, note that the proof of Lemma 3.5 gives, in this notation

$$\mathcal{S}[\mu]^r_{[k]}\Big(m\Big) \quad = \quad \sum_{j=1}^{m-r} y_j \quad + \quad \lambda \cdot \sum_{t=0}^{r-1} x^t_{m-r+t+1} \tag{13}$$

4. Finally, for $M \subseteq [k]$ and $0 \leq r \leq |M|$, we write

$$\mu^r_M \quad = \quad \mathop{\mathbb{E}}_{|R|=r, R\subseteq M} \quad \mu^R(M) \quad \text{and} \quad \mathcal{S}[\mu]^r_M \quad = \quad \mathop{\mathbb{E}}_{|R|=r, R\subseteq M} \quad \mathcal{S}[\mu]^R_M(M)$$

We have completed introducing the new notation. In this notation the claim of the theorem amounts to:

$$\mu^r_{[k]} \quad \leq \quad \mathcal{S}[\mu]^r_{[k]} \tag{14}$$

We start with a lemma connecting the value of a solution of the optimization problem to these of smaller problems.

Lemma 3.6:

$$\mu^r_{[k]} \quad \leq \quad \lambda \cdot \mu^{r-1}_{[k]} \quad + \quad (1-\lambda) \cdot \mathop{\mathbb{E}}_{i\in[n]} \ \mu^{r-1}_{[k]\setminus\{i\}} \tag{15}$$

Proof:

Since the feasible solution $\left\{x^R_{S,i}\right\}$ satisfies constraints 2 and 3 of (7), for any $i \in R \subseteq [k]$ holds $\mu^R\Big([k]\Big) \ = \ \mu^{R\setminus\{i\}}\Big([k] \setminus \{i\}\Big) \ + \ x^R_{[k],i}$.

Similarly, $\mu^{R\setminus\{i\}}\Big([k]\Big) \quad = \quad \mu^{R\setminus\{i\}}\Big([k] \setminus \{i\}\Big) \quad + \quad x^{R\setminus\{i\}}_{[k],i}$.

Hence, by constraint 4,

$$x^R_{[k],i} \;\leq\; \lambda \cdot \Big(x^{R\setminus\{i\}}_{[k],i}\Big) \;=\; \lambda \cdot \Big(\mu^{R\setminus\{i\}}\Big([k]\Big) \;-\; \mu^{R\setminus\{i\}}\Big([k]\setminus\{i\}\Big)\Big)$$

Averaging,

$$\mu^r_{[k]} \;=\; \mathop{\mathbb{E}}_{R\subseteq[k],\ |R|=r} \mu^R_{[k]} \;=\; \mathop{\mathbb{E}}_{R,\ i\in R} \Big(\mu^{R\setminus\{i\}}\Big([k]\setminus\{i\}\Big) \;+\; x^R_{[k],i}\Big) \;\leq$$

$$\mathop{\mathbb{E}}_{R,\ i\in R} \mu^{R\setminus\{i\}}\Big([k]\setminus\{i\}\Big) \;+\; \lambda \cdot \mathop{\mathbb{E}}_{R,\ i\in R} \Big(\mu^{R\setminus\{i\}}\Big([k]\Big) \;-\; \mu^{R\setminus\{i\}}\Big([k]\setminus\{i\}\Big)\Big) \;=$$

$$\lambda \cdot \mathop{\mathbb{E}}_{R,\ i\in R} \mu^{R\setminus\{i\}}\Big([k]\Big) \;+\; (1-\lambda)\cdot \mathop{\mathbb{E}}_{R,\ i\in R} \mu^{R\setminus\{i\}}\Big([k]\setminus\{i\}\Big)$$

It remains to note

$$\mathop{\mathbb{E}}_{R,\ i\in R} \mu^{R\setminus\{i\}}\Big([k]\setminus\{i\}\Big) \;=\; \mathop{\mathbb{E}}_{i\in[k]} \mathop{\mathbb{E}}_{|T|=r-1,\ T\subseteq[k]\setminus\{i\}} \mu^T\Big([k]\setminus\{i\}\Big) \;=\; \mathop{\mathbb{E}}_{i\in[k]} \mu^{r-1}_{[k]\setminus\{i\}}$$

and, similarly, $\mathbb{E}_{R,\ i\in R}\ \mu^{R\setminus\{i\}}\Big([k]\Big) \;=\; \mu^{r-1}_{[k]}$.

∎

We now prove (14), starting from (15).

First, note that, by Lemma 3.4 and by the induction hypothesis for $k-1$, we have $\mu^{r-1}_{[k]\setminus\{i\}} \;\leq\; \mathcal{S}[\mu]^{r-1}_{[k]\setminus\{i\}}$, for all $i\in[k]$.

Next, note that, by the induction hypothesis for k and $r-1$, we have $\mu^{r-1}_{[k]} \;\leq\; \mathcal{S}[\mu]^{r-1}_{[k]}$.

This gives

$$\mu^r_{[k]} \;\leq\; \lambda\cdot\mathcal{S}[\mu]^{r-1}_{[k]} \;+\; (1-\lambda)\cdot \mathop{\mathbb{E}}_{i\in[k]} \mathcal{S}[\mu]^{r-1}_{[k]\setminus\{i\}}$$

This implies that to prove (14) it suffices to show the following two identities:

1. $$\mathop{\mathbb{E}}_{i\in[k]} \mathcal{S}[\mu]^{r-1}_{[k]\setminus\{i\}} \;=\; \mathcal{S}[\mu]^{r-1}_{[k]}\Big(k-1\Big)$$

2. $$\mathcal{S}[\mu]^r_{[k]} \;=\; \lambda\cdot\mathcal{S}[\mu]^{r-1}_{[k]} \;+\; (1-\lambda)\cdot\mathcal{S}[\mu]^{r-1}_{[k]}\Big(k-1\Big)$$

Lemma 3.7:

$$\mathop{\mathbb{E}}_{i\in[k]} \mathcal{S}[\mu]^{r-1}_{[k]\setminus\{i\}} \;=\; \mathcal{S}[\mu]^{r-1}_{[k]}\Big(k-1\Big)$$

Proof: We introduce the following notation. For $i = 1, ..., k$ and for $0 \leq r < s \leq k-1$, let

$$y_{s,i} \quad = \quad y_{s,\ [k]\setminus\{i\}} \quad \text{and} \quad x^r_{s,i} \quad = \quad x^r_{s,\ [k]\setminus\{i\}}$$

The values on the RHS of these identities are defined as in (10) and in (11) for the corresponding restricted problems.

We start with observing that $\mathbb{E}_{i\in[k]}\, y_{s,i} \ = \ y_s$. In fact, by definition,

$$\mathop{\mathbb{E}}_{i\in[k]} y_{s,i} \quad = \quad \mathop{\mathbb{E}}_{i\in[k]} \ \mathop{\mathbb{E}}_{|S|=s, S\subseteq[k]\setminus\{i\}, j\in S} \ y_{S,j} \quad = \quad \mathop{\mathbb{E}}_{|S|=s, j\in S} \ y_{S,j} \quad = \quad y_s$$

Next, we claim that for all $0 \leq r < s \leq k-1$ holds $\mathbb{E}_{i\in[k]} \ x^r_{s,i} \ = \ x^r_s$.

This is easy to verify by induction on r. Note that we already know the claim holds for $r = 0$, and the induction step follows directly from the definitions and the induction hypothesis.

We now apply (12) to the restricted problems, to obtain that, for each $1 \leq i \leq k$ holds

$$\mathcal{S}[\mu]^{r-1}_{[k]\setminus\{i\}} \quad = \quad \sum_{j=1}^{k-r} y_{j,\ i} \quad + \quad \lambda \cdot \sum_{t=0}^{r-2} x^t_{k-r+t+1,\ i}$$

Hence, we have:

$$\mathop{\mathbb{E}}_{i\in[k]} \ \mathcal{S}[\mu]^{r-1}_{[k]\setminus\{i\}} \quad = \quad \sum_{j=1}^{k-r} \mathop{\mathbb{E}}_{i\in[k]} \ y_{j,\ i} \quad + \quad \lambda \cdot \sum_{t=0}^{r-2} \mathop{\mathbb{E}}_{i\in[k]} \ x^t_{k-r+t+1,\ i} \quad = \quad \sum_{j=1}^{k-r} y_j \ + \ \lambda \cdot \sum_{t=0}^{r-2} x^t_{k-r+t+1}$$

This, by (13), equals to $\mathcal{S}[\mu]^{r-1}_{[k]}(k-1)$, completing the proof of the lemma.

∎

Lemma 3.8:

$$\mathcal{S}[\mu]^r_{[k]} \quad = \quad \lambda \cdot \mathcal{S}[\mu]^{r-1}_{[k]} \quad + \quad (1-\lambda) \cdot \mathcal{S}[\mu]^{r-1}_{[k]}\Big(k-1\Big)$$

Proof:

The proof of this lemma is similar to that of Lemma 3.6.

Since the symmetric solution $\mathcal{S}_{[k]}\left[x^R_{S,i}\right]$ $\Big($ which is the same as $\left\{\bar{x}^R_{S,i}\right\}\Big)$ satisfies constraints 2 and 3 of (7), for any $i \in R \subseteq [k]$ holds

$$\mathcal{S}[\mu]^R_{[k]}\Big([k]\Big) \quad = \quad \mathcal{S}[\mu]^{R\setminus\{i\}}_{[k]}\Big([k]\setminus\{i\}\Big) \quad + \quad \mathcal{S}_{[k]}\left[x^R_{[k],i}\right]$$

Consider the notation we have introduced above. Using items 3 and 4 in the description of this notation, and recalling $\mathcal{S}_{[k]}\left[x^R_{[k],i}\right] = \lambda \cdot x^{r-1}_k$, we can rewrite this equality as

$$\mathcal{S}[\mu]^r_{[k]} \quad = \quad \mathcal{S}[\mu]^{r-1}_{[k]}\Big(k-1\Big) \quad + \quad \lambda \cdot x^{r-1}_k$$

On the other hand, we have, for $i \in R \subseteq [k]$:

$$\mathcal{S}[\mu]_{[k]}^{R\setminus\{i\}}\Big([k]\Big) \quad = \quad \mathcal{S}[\mu]_{[k]}^{R\setminus\{i\}}\Big([k] \setminus \{i\}\Big) \quad + \quad \mathcal{S}_{[k]}\Big[x_{[k],i}^{R\setminus\{i\}}\Big]$$

which is the same as

$$\mathcal{S}[\mu]_{[k]}^{r-1} \quad = \quad \mathcal{S}[\mu]_{[k]}^{r-1}\Big(k-1\Big) \quad + \quad x_k^{r-1}$$

Combining these two identities immediately implies the claim of the lemma.

▮

This completes the proof of (14) and of the theorem.

▮

3.3 The value of the symmetric optimization problem

Let $\left\{\bar{x}_{S,i}^R\right\}$ be the symmetric solution for the symmetric version of (7). By Corollary 3.3, its value depends linearly on the symmetric boundary data $y_1, ..., y_k$, since $\{x_t^r\}$ are fixed linear functions of $y_1, ..., y_k$. Let us denote this value by $V(y_1, ..., y_k)$.

For $1 \le s \le k$, let e_s be the initial data vector with $y_s = 1$ and all the remaining y_t vanishing. Then $V(y_1, \ldots, y_k) \;=\; \sum_{s=1}^k y_s \cdot V(e_s)$.

Next, we find the values of the parameters x_t^r for initial data given by a unit vector.

Lemma 3.9: *Let the initial data be given by the unit vector e_s, for some $1 \le s \le k$. Then the values of the parameters x_t^r, for $0 \le r < t \le k$, are as follows.*

$$x_t^r \quad = \quad \begin{cases} \binom{r}{t-s} \cdot \lambda^{r-(t-s)} \cdot (1-\lambda)^{t-s} & \textit{if} \qquad s \;\le\; t \;\le\; s+r \\ 0 & \textit{otherwise} \end{cases}$$

(We use the convention $\binom{0}{0} = 1$.)

Proof: The claim of the lemma is easily verifiable by induction on r, or by directly verifying that (11) holds. ▮

Corollary 3.10:

$$V(e_s) \quad = \quad \lambda^s \cdot \sum_{m=0}^{k-s} \binom{s+m-1}{m} \cdot (1-\lambda)^m \quad = \quad 1 \quad - \quad \sum_{j=0}^{s-1} \binom{k}{j} \lambda^j (1-\lambda)^{k-j}$$

Proof: The first equality follows from Corollary 3.3. For the second equality, we proceed as follows

$$V\left(e_s\right) \quad = \quad \frac{\lambda^s}{(s-1)!} \cdot \frac{\partial^{s-1}}{\partial x^{s-1}} \quad \left[(1+x+\ldots+x^{k-1}\right]_{x=1-\lambda} \quad =$$

$$\frac{\lambda^s}{(s-1)!} \cdot \left(\frac{\partial^{s-1}}{\partial x^{s-1}} \quad \left[\frac{1}{1-x}\right]_{x=1-\lambda} \quad - \quad \frac{\partial^{s-1}}{\partial x^{s-1}} \quad \left[\frac{x^k}{1-x}\right]_{x=1-\lambda} \right) \quad =$$

$$1 \quad - \quad \frac{\lambda^s}{(s-1)!} \cdot \frac{\partial^{s-1}}{\partial x^{s-1}} \quad \left[\frac{x^k}{1-x}\right]_{x=1-\lambda}$$

We have

$$\frac{\partial^t}{\partial x^t} \quad \left[\frac{x^k}{1-x}\right] \quad = \quad \sum_{i=0}^{t} \binom{t}{i} \frac{\partial^i}{\partial x^i} \quad \left[\frac{1}{1-x}\right] \cdot \frac{\partial^{t-i}}{\partial x^{t-i}} \quad \left[x^k\right] \quad =$$

$$\sum_{i=0}^{t} \binom{t}{i} \cdot i! \cdot \frac{k!}{(k-t+i)!} \cdot x^{k-t+i} \cdot \frac{1}{(1-x)^{i+1}}$$

Substituting $j = t - i$ and rearranging, this is

$$\frac{t!}{(1-x)^{t+1}} \cdot \sum_{j=0}^{t} \binom{k}{j} (1-x)^j \cdot x^{k-j}$$

Substituting $t = s - 1$, $x = 1 - \lambda$, and simplifying, we get

$$V\left(e_s\right) \quad = \quad 1 \quad - \quad \sum_{j=0}^{s-1} \binom{k}{j} \lambda^j (1-\lambda)^{k-j}$$

∎

Corollary 3.11:

$$V\left(y_1, \ldots, y_k\right) \quad = \quad \sum_{s=1}^{k} \left(1 \ - \ \sum_{j=0}^{s-1} \binom{k}{j} \lambda^j (1-\lambda)^{k-j} \right) \ \cdot \ y_s$$

4 Proof of Theorem 1.7

We start with introducing some more notation.

4.0.1 Notation

- For a subset S of $[n]$ of cardinality at most $n-2$, and for distinct $i,j \notin S$, we set

$$Z_{S;i,j} \quad = \quad Ent\Big(f \mid S \cup \{i,j\}\Big) \;-\; Ent\Big(f \mid S \cup \{i\}\Big) \;-\; Ent\Big(f \mid S \cup \{j\}\Big) \;+\; Ent\Big(f \mid S\Big)$$

- For $s = 1, ..., n-1$, let $t_s \;=\; \mathbb{E}_{S,i,j}\; Z_{S;i,j}$.
  Here the expectation is taken over all subsets S of $[n]$ of cardinality $s-1$, and, given S, over all distinct i,j not in S.

- Let A be a subset of $[n]$ of cardinality $k < n$ and let $m \notin A$. For $1 \le s \le k$, let

$$Y(A,m,s) \quad = \quad \mathop{\mathbb{E}}_{S,i}\; Z_{S;i,m}$$

  where the expectation goes over subsets $S \subseteq A$ of cardinality $s-1$, and over $i \in A \setminus S$.

- For $1 \le s \le k \le n$ let

$$\Lambda(k,s,\lambda) \quad = \quad 1 \;-\; \sum_{j=0}^{s-1} \binom{k}{j} \lambda^j (1-\lambda)^{k-j}$$

Proposition 4.1: *Let f be a nonnegative function on $\{0,1\}^n$. Let A be a subset of $[n]$ of cardinality $k < n$ and let $m \notin A$.*

Then

$$I_{T_{e_A} f}\;(A,m) \quad \le \quad \sum_{s=1}^{k} \;\Lambda(k,s,\lambda)\;\cdot\;Y(A,m,s)$$

Proof:

By Theorem 2.1, the value of $I_{T_{e_A} f}\;(A,m)$ is bounded by the value of the linear optimization problem (7), with appropriate changes of indices.

By Theorem 3.2, this last value is upperbounded by the value of the symmetric version of the problem, which, according to Corollary 3.11, and tracing out the appropriate changes in indices and notation, is given by $\sum_{s=1}^k \;\Lambda(k,s,\lambda) \cdot Y(A,m,s)$.

∎

Proof: (Of the theorem)

The proof relies on several lemmas. We start with a technical claim.

Lemma 4.2: *Let $1 \le s \le n-1$ be integer parameters. Let $0 < \lambda < 1$. Then*

$$\sum_{k=s}^{n-1} \Lambda(k,s,\lambda) \quad = \quad \left(n - \frac{s}{\lambda}\right) \;+\; \frac{1}{\lambda}\cdot \sum_{j=0}^{s-1}\; \sum_{t=0}^{j} \binom{n}{t} \lambda^t (1-\lambda)^{n-t}$$

Proof:

$$\sum_{k=s}^{n-1} \Lambda(k,s,\lambda) \quad = \quad \sum_{k=s}^{n-1} \left(1 \quad - \quad \sum_{j=0}^{s-1} \binom{k}{j} \lambda^j (1-\lambda)^{k-j}\right) \quad =$$

$$\Big(n-s\Big) \quad - \quad \sum_{k=s}^{n-1} \sum_{j=0}^{s-1} \binom{k}{j} \lambda^j (1-\lambda)^{k-j} \quad = \quad \Big(n-s\Big) \quad - \quad \sum_{j=0}^{s-1} \lambda^j \cdot \sum_{k=s}^{n-1} \binom{k}{j} (1-\lambda)^{k-j}$$

A simple calculation, similar to that in the proof of Corollary 3.10, gives

$$\lambda^j \cdot \sum_{k=s}^{n-1} \binom{k}{j} (1-\lambda)^{k-j} \quad = \quad \frac{1}{\lambda} \cdot \left(\sum_{t=0}^{j} \binom{s}{t} \lambda^t (1-\lambda)^{s-t} \quad - \quad \sum_{t=0}^{j} \binom{n}{t} \lambda^t (1-\lambda)^{n-t} \right)$$

The proof of the lemma is completed by summing the RHS over j, and observing

$$\sum_{j=0}^{s-1} \sum_{t=0}^{j} \binom{s}{t} \lambda^t (1-\lambda)^{s-t} \quad = \quad (1-\lambda) \cdot s$$

∎

Lemma 4.3: *Let f be a nonnegative function on $\{0,1\}^n$ with expectation 1. Then*

$$Ent\Big(T_\epsilon f\Big) \quad \le \quad \sum_{i=1}^{n} \phi\Big(Ent\Big(f \mid \{i\}\Big)\Big) \quad + \quad \sum_{s=1}^{n-1} w_s \cdot t_s$$

where

$$w_s \quad = \quad \Big(\lambda n - s\Big) \; + \; \sum_{j=0}^{s-1} \sum_{t=0}^{j} \binom{n}{t} \lambda^t (1-\lambda)^{n-t}$$

Lemma 4.4: *Let f be a nonnegative function on $\{0,1\}^n$. For any $0 \le u \le n-1$ holds*

$$\mathop{\mathbb{E}}_{|B|=u+1} Ent\Big(f \mid B\Big) \; - \; (u+1) \cdot \mathop{\mathbb{E}}_{i \in [n]} Ent\Big(f \mid \{i\}\Big) \quad = \quad \sum_{s=1}^{u} \Big(u-s+1\Big) \cdot t_s$$

Next, we derive the theorem, assuming Lemmas 4.3 and 4.4 to hold.

We are going to apply the Chernoff bound in the following form [1]:

Let $X_k \sim B(k,\lambda)$ be a Bernoulli random variable. Then for any $a \ge 0$ holds $Pr\Big\{\Big|X_k - \lambda k\Big| > a\Big\} \le e^{-2a^2/k}$.

Let us set $d_s = \sum_{j=0}^{s-1} \sum_{t=0}^{j} \binom{n}{t} \lambda^t (1-\lambda)^{n-t}$. Note that $d_s = \sum_{j=0}^{s-1} Pr\Big\{X_n \le j\Big\}$, and that $w_s = \Big(\lambda n - s\Big) + d_s$.

Remark 4.5: We note, for future use, the following two probabilistic interpretations for w_s:

$$w_s \quad = \quad \Big(\lambda n - s\Big) \;+\; \sum_{j=0}^{s-1} Pr\Big\{X_n \le j\Big\} \quad \text{and} \quad w_s \quad = \quad \lambda \cdot \sum_{k=s}^{n-1} \Lambda(k,s,\lambda) \quad = \quad \lambda \cdot \sum_{k=s}^{n-1} Pr\Big\{X_k \ge s\Big\}$$

▮

Using the Chernoff bound for X_n gives that for $s < \lambda n - \sqrt{n \ln n}$ holds

$$d_s \quad = \quad \sum_{j=0}^{s-1} Pr\Big\{X_n \le j\Big\} \quad \le \quad \frac{1}{n}$$

Applying the bound to $X_k \sim B(k,\lambda)$ gives that for $s > \lambda n + \sqrt{n \ln n}$ holds

$$w_s \quad = \quad \lambda \cdot \sum_{k=s}^{n-1} \Lambda(k,s,\lambda) \quad \le \quad \sum_{k=s}^{n-1} Pr\Big\{X_k > s\Big\} \quad < \quad \frac{1}{n}$$

In the first step we used Lemma 4.2.

Finally, applying the bound to X_n again, we have for $\lambda n - \sqrt{n \ln n} \le s \le \lambda n + \sqrt{n \ln n}$

$$d_s \quad = \quad \sum_{j=0}^{s-1} Pr\Big\{X_n \le j\Big\} \quad = \quad \sum_{j=0}^{\lambda n - \sqrt{n \ln n}} Pr\Big\{X_n \le j\Big\} \;+\; \sum_{j=\lambda n - \sqrt{n \ln n}+1}^{s-1} Pr\Big\{X_n \le j\Big\} \quad \le$$

$$O\Big(\sqrt{n \ln n}\Big)$$

Combining the two lemmas, taking $u = \lambda n + \sqrt{n \ln n} - 1$ in Lemma 4.4, and setting $v = u + 1$, we have

$$Ent\Big(T_\epsilon f\Big) \quad \le \quad \mathop{\mathbb{E}}_{|B|=v} Ent\Big(f \mid B\Big) \;-\; \frac{v}{n} \cdot \sum_{i=1}^{n} Ent\Big(f \mid \{i\}\Big) \;+\; \sum_{i=1}^{n} \phi\Big(Ent\Big(f \;\Big|\; \{i\}\Big)\Big) \;+$$

$$\sum_{s=1}^{v} d_s \cdot t_s \;+\; \sum_{s=v+1}^{n-1} w_s \cdot t_s \quad =$$

$$\mathop{\mathbb{E}}_{|B|=v} \Big(Ent\Big(f \mid B\Big) \;-\; \sum_{i \in B} Ent\Big(f \mid \{i\}\Big)\Big) \;+\; \sum_{i=1}^{n} \phi\Big(Ent\Big(f \;\Big|\; \{i\}\Big)\Big) \;+\; E(n),$$

where $E(n)$ is an error term.

We now estimate $E(n)$. Observe that Lemma 4.4 applied with $u = n-1$ implies

$$\sum_{s=1}^{n-1}(n-s)\cdot t_s \quad \leq \quad Ent\Big(f\Big)$$

In particular

$$\sum_{s=\lambda n-\sqrt{n\ln n}}^{\lambda n+\sqrt{n\ln n}} t_s \quad \leq \quad O\left(\frac{1}{n}\right)\cdot Ent\Big(f\Big)$$

Assuming $n \geq 10\cdot\left(\frac{1}{(1-\lambda)^2}\cdot\log\left(\frac{1}{1-\lambda}\right)\right)$, the asymptotic notation in the above expression hides an absolute constant.

Hence, by the above discussion,

$$\sum_{s=1}^{v} d_s\cdot t_s \quad = \quad \sum_{s=1}^{\lambda n-\sqrt{n\ln n}} d_s\cdot t_s \;+\; \sum_{s=\lambda n-\sqrt{n\ln n}+1}^{v} d_s\cdot t_s \quad \leq \quad O\left(\sqrt{\frac{\ln n}{n}}\right)\cdot Ent\Big(f\Big)$$

where both summands in the middle expression are upperbounded by the maximum of d_s on the corresponding domain, times the sum of t_s on that domain.

Similarly,

$$\sum_{s=v+1}^{n-1} w_s\cdot t_s \quad \leq \quad \max_{v+1\leq s\leq n} w_s\cdot\sum_{s=v+1}^{n-1} t_s \quad \leq \quad O\left(\frac{1}{n}\right)\cdot Ent\Big(f\Big)$$

Hence $E(n) \;\leq\; O\left(\sqrt{\frac{\ln n}{n}}\right)\cdot Ent\Big(f\Big)$.

This completes the proof of the theorem, given the lemmas hold.

It remains to prove the lemmas.

Proof: (Of Lemma 4.3)

Recall that, by the chain rule for noisy entropy (5), for any permutation $\sigma \in S_n$ holds that $Ent\Big(T_\epsilon f\Big)$ is bounded from above by

$$\sum_{i=1}^{n}\phi\left(Ent\Big(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}}f \mid \{\sigma(1),\ldots,\sigma(i)\}\Big) \;-\; Ent\Big(T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}}f \mid \{\sigma(1),\ldots,\sigma(i-1)\}\Big)\right)$$

Using the notation introduced in Subsection 1.4.1, we can write this as

$$\sum_{i=1}^{n}\phi\left(Ent\Big(f \mid \{\sigma(i)\}\Big) \;+\; I_{T_{\epsilon_{\{\sigma(1),\ldots,\sigma(i-1)\}}}f}\left(\Big\{\sigma(1),\ldots,\sigma(i-1)\Big\},\ \sigma(i)\right)\right)$$

Observe that the function ϕ is concave, and $\phi(0) = 0$. Hence $\phi(x + y) \leq \phi(x) + \phi(y)$ for any $0 \leq x, y \leq 1$. By this subbaditivity of ϕ, the last expression is at most

$$\sum_{i=1}^{n} \phi\Big(Ent\Big(f \mid \{i\}\Big)\Big) \quad + \quad \sum_{k=2}^{n} \phi\Big(I_{T_{\epsilon_{\{\sigma(1),\ldots,\sigma(k-1)\}}} f} \quad \Big(\Big\{\sigma(1), \ldots, \sigma(k-1)\Big\},\ \sigma(k)\Big)\Big)$$

Averaging this expression over all $\sigma \in S_n$, we obtain

$$Ent\Big(T_\epsilon f\Big) \quad \leq \quad \sum_{i=1}^{n} \phi\Big(Ent\Big(f \mid \{i\}\Big)\Big) \quad + \quad \mu,$$

where

$$\mu = \mathop{\mathbb{E}}_{\sigma} \quad \sum_{k=2}^{n} \phi\Big(I_{T_{\epsilon_{\{\sigma(1),\ldots,\sigma(k-1)\}}} f} \quad \Big(\Big\{\sigma(1), \ldots, \sigma(k-1)\Big\},\ \sigma(k)\Big)\Big)$$

Next, we upper bound μ. By transitivity of action of the symmetric group and by concavity of ϕ we have

$$\mu \quad \leq \quad \sum_{k=1}^{n-1} \phi\Big(b_k\Big) \quad \text{where} \quad b_k \quad = \quad \mathop{\mathbb{E}}_{A,m} \ T_{\epsilon_A f} \ \Big(A, m\Big)$$

where the expectation is over all $A \subseteq [n]$ of cardinality k and $m \notin A$.

Applying Proposition 4.1, we get

$$b_k \quad \leq \quad \mathop{\mathbb{E}}_{A,m} \ \sum_{s=1}^{k} \Lambda\Big(k, s, \lambda\Big) \cdot Y\Big(A, m, s\Big) \quad = \quad \sum_{s=1}^{k} \Lambda\Big(k, s, \lambda\Big) \cdot \mathop{\mathbb{E}}_{A,m} \ Y\Big(A, m, s\Big)$$

By the definition of $Y\Big(A, m, s\Big)$,

$$\mathop{\mathbb{E}}_{A,m} \ Y\Big(A, m, s\Big) \quad = \quad \mathop{\mathbb{E}}_{A,m} \ \mathop{\mathbb{E}}_{S,i} \ Z_{S;i,m} \quad = \quad \mathop{\mathbb{E}}_{S,i,m} \ Z_{S;i,m} \cdot \mathop{\mathbb{E}}_{A} \ 1 \quad = \quad \mathop{\mathbb{E}}_{S,i,m} \ Z_{S;i,m}$$

where in the second expression the first expectation is over k-subsets A of $[n]$ and $m \notin A$, and the second expectation is over $(s-1)$-subsets S of A and over $i \in A \setminus S$. Rearranging, we get the third expression in which the first expectation is over all subsets S of $[n]$ of cardinality $s - 1$ and over all distinct $i, m \notin S$, and the second expectation is over all supersets A of S of cardinality k with $i \in A$ and $m \notin A$.

Recalling the definition of t_s above, we deduce $b_k \quad = \quad \sum_{s=1}^{k} \Lambda\Big(k, s, \lambda\Big) \cdot t_s$.

Using the inequality $\phi(x) \leq \lambda x$, and Lemma 4.2, we have

$$\mu \quad \leq \quad \lambda \cdot \sum_{k=1}^{n-1} b_k \quad = \quad \lambda \cdot \sum_{k=1}^{n-1} \ \sum_{s=1}^{k} \Lambda\Big(k, s, \lambda\Big) \cdot t_s \quad =$$

$$\sum_{s=1}^{n-1} t_s \cdot \left(\lambda \cdot \sum_{k=s}^{n-1} \Lambda\Big(k, s, \lambda\Big)\right) \quad = \quad \sum_{s=1}^{n-1} w_s \cdot t_s$$

▮

Proof: (of Lemma 4.4)

By (8), for any subset A of $[n]$ of cardinality $1 \le k \le n-1$, for any $m \notin A$, and for any bijection $\tau : [k] \rightarrow A$ holds, in the notation of this section,

$$\sum_{s=1}^{k} Z_{\{\tau(1),\ldots,\tau(s-1)\};\tau(s),m} \quad = \quad I_f\Big(A, m\Big)$$

We now average over all the variables, setting

$$c_k \quad = \quad \mathop{\mathbb{E}}_{A,m,\tau} \sum_{s=1}^{k} Z_{\{\tau(1),\ldots,\tau(s-1)\};\tau(s),m}$$

On one hand, we have

$$c_k \quad = \quad \mathop{\mathbb{E}}_{A,m} I_f\Big(A, m\Big) \quad = \quad \mathop{\mathbb{E}}_{A,m} \left(Ent\Big(f \mid A \cup \{m\}\Big) - Ent\Big(f \mid A\Big) - Ent\Big(f \mid \{m\}\Big)\right) \quad =$$

$$\mathop{\mathbb{E}}_{|B|=k+1} Ent\Big(f \mid B\Big) \; - \; \mathop{\mathbb{E}}_{|A|=k} Ent\Big(f \mid A\Big) \; - \; \mathop{\mathbb{E}}_{i \in [n]} Ent\Big(f \mid \{i\}\Big)$$

On the other hand, similarly to the computation in the preceding lemma, we have

$$c_k \quad = \quad \sum_{s=1}^{k} \mathop{\mathbb{E}}_{A,m,\tau} Z_{\{\tau(1),\ldots,\tau(s-1)\};\tau(s),m} \quad = \quad \sum_{s=1}^{k} \mathop{\mathbb{E}}_{A,S,i,m} Z_{S;i,m} \quad = \quad \sum_{s=1}^{k} t_s$$

where the expectation in the third expression is over k-subsets A of $[n]$, over $(s-1)$-subsets S of A, over $m \notin A$ and $i \in A \setminus S$.

Hence, for any $1 \le u \le n-1$ holds

$$\mathop{\mathbb{E}}_{|B|=u+1} Ent\Big(f \mid B\Big) \; - \; (u+1) \cdot \mathop{\mathbb{E}}_{i \in [n]} Ent\Big(f \mid \{i\}\Big) \quad = \quad \sum_{k=1}^{u} c_k \quad = \quad \sum_{s=1}^{u} \Big(u - s + 1\Big) \cdot t_s$$

completing the proof of the lemma and of the theorem.

▮

5 Proof of Theorem 1.11

Let δ be the constant in the theorem. We will assume in the following argument that δ is sufficiently small.

Let $0 < \epsilon < 1/2$ be a noise parameter, such that $(1-2\epsilon)^2 \le \delta$. Let $\lambda = (1-2\epsilon)^2$.

Let $f : \{0,1\}^n \rightarrow \{0,1\}$ be a boolean function, satisfying the constraints of the theorem. Let $1 \le k \le n$ be the coordinate such that $|\widehat{f}(k)|$ is large. W.l.o.g. assume that $k = 1$ and that $\widehat{f}(1)$ is positive.

We introduce some additional notation.

Notation

- Let $0 \le \alpha \le \delta$ be such that $\widehat{f}(1) = (1-\alpha) \cdot \mathbb{E} f$.
- Let $0 \le \beta \le \delta$ be such that $\mathbb{E} f = 1/2 - \beta$. Let $\gamma = \alpha + \beta$.
- If $\alpha \le \lambda$, we define $\tau = \left(\frac{1-\lambda}{1-\alpha}\right)^2$, and define auxiliary noise ϵ_τ, such that $\Big(1 - 2\epsilon_\tau\Big)^2 = \tau$. If $\alpha > \lambda$, we set $\tau = 1$ and $\epsilon_\tau = 0$.
- Let ϵ_1 be such that $T_\epsilon = T_{\epsilon_1} T_{\epsilon_\tau}$. Let $\lambda_1 = (1-2\epsilon_1)^2$. Note that $\lambda = \tau \cdot \lambda_1$.
- Let $h = T_{\epsilon_\tau} f$. Note that $T_\epsilon f = T_{\epsilon_1} h$, and hence $Ent\Big(T_\epsilon f\Big) = Ent\Big(T_{\epsilon_1} h\Big)$.

We will first show that the claim of the theorem holds up to a small error. That is,

$$Ent\Big(T_\epsilon f\Big) \quad \le \quad \frac{1}{2} \cdot \Big(1 - H(\epsilon)\Big) \; + \; e(n) \tag{16}$$

Here and from now on in this section $e(n)$ defines an error term which goes to zero with n. We will use the same notation for different error terms. We justify this abuse of notation by the fact that it is not hard to verify all these error terms go to zero uniformly with n.

This would be the main part of the proof. We will then show, in Subsection 5.2, that the error term may be removed by a direct product argument.

We start with applying Theorem 1.7 to the function h with noise ϵ_1. The theorem is stated for functions with expectation 1. We modify it, using the linearity of entropy, to obtain

$$Ent\Big(T_{\epsilon_1} h\Big) \quad \le \quad \mathop{\mathbb{E}}_{|B|=v} \left(Ent\Big(h \mid B\Big) \; - \; \sum_{i \in B} Ent\Big(h \mid \{i\}\Big)\right) \quad +$$

$$\mathbb{E}\, h \cdot \sum_{i=1}^{n} \phi\left(Ent\left(\frac{h}{\mathbb{E}\, h} \;\Big|\; \{i\}\right), \; \epsilon_1\right) \quad + \quad O\left(\sqrt{\frac{\ln n}{n}}\right) \cdot Ent\Big(h\Big)$$

Here $v = \left\lceil \lambda_1 \cdot n + \sqrt{n \ln n} \right\rceil$.

Note that, since there are several noise parameters involved, we now write the function ϕ with the noise parameter stated explicitly.

Let $\lambda_2 = v/n$. Let ϵ_2 be such that $(1-2\epsilon_2)^2 = \lambda_2$. Then the statement above as implies:

$$Ent\Big(T_{\epsilon_1} h\Big) \quad \le \quad \mathop{\mathbb{E}}_{|B|=\lambda_2 n} \left(Ent\Big(h \mid B\Big) \; - \; \sum_{i \in B} Ent\Big(h \mid \{i\}\Big) \right) \quad +$$

$$\mathbb{E}\, h \cdot \sum_{i=1}^{n} \phi\left(Ent\left(\frac{h}{\mathbb{E}\, h} \;\Big|\; \{i\} \right), \; \epsilon_2 \right) \quad + \quad e(n)$$

To see this, note that $\epsilon_2 \le \epsilon_1$, that $\phi(x, \epsilon)$ decreases in ϵ, and that $Ent\Big(h\Big) \le Ent\Big(f\Big) \le 1$.

Next, note that, by (3), for any $1 \le i \le n$ holds

$$\mathbb{E}\, h \cdot \phi\left(Ent\left(\frac{h}{\mathbb{E}\, h} \;\Big|\; \{i\} \right), \; \epsilon_2 \right) \quad \le \quad \lambda_2 \cdot Ent\Big(h \mid \{i\}\Big)$$

Hence the previous inequality implies

$$Ent\Big(T_{\epsilon_1} h\Big) \; \le \; \lambda_2 \cdot \mathop{\mathbb{E}}_{|B|=\lambda_2 n, 1 \in B} \left(Ent\Big(h \mid B\Big) - Ent\Big(h \mid \{1\}\Big) \right) \quad +$$

$$(1-\lambda_2) \cdot \mathop{\mathbb{E}}_{|B|=\lambda_2 n, 1 \notin B} Ent\Big(h \mid B\Big) \quad + \quad \mathbb{E}\, h \cdot \phi\left(Ent\left(\frac{h}{\mathbb{E}\, h} \;\Big|\; \{1\} \right), \; \epsilon_2 \right) \quad + \quad e(n) \qquad (17)$$

The claim in (16) will be based on three lemmas, which upperbound each of the three significant summands in the RHS of (17).

Lemma 5.1:

$$\mathop{\mathbb{E}}_{|B|=\lambda_2 n, \; 1 \in B} \left(Ent\Big(h \mid B\Big) - Ent\Big(h \mid \{1\}\Big) \right) \quad \le \quad O\left(\lambda_2 \cdot \gamma \; + \; \gamma^2 \ln\left(\frac{1}{\gamma}\right) \right) \; + \; e(n)$$

Lemma 5.2:

$$\mathop{\mathbb{E}}_{|B|=\lambda_2 n, \; 1 \notin B} Ent\Big(h \mid B\Big) \quad \le \quad O\left(\lambda_2^2 \cdot \gamma \; + \; \lambda_2 \cdot \gamma^2 \ln\left(\frac{1}{\gamma}\right) \right)$$

Lemma 5.3:

$$\mathbb{E}\, h \cdot \phi\left(Ent\left(\frac{h}{\mathbb{E}\, h} \;\Big|\; \{1\}\right),\ \epsilon_2\right) \quad \le \quad \frac{1}{2} \cdot \Big(1 - H(\epsilon)\Big) \ - \ \Omega\Big(\lambda \cdot \gamma\Big) \ + \ e(n)$$

The asymptotic notation in each of the lemmas hides absolute constants.

Given the lemmas, (16) is easy to verify. Indeed, it is easy to check that $\lambda \le \lambda_2 \le c \cdot \lambda + e(n)$, for some absolute constant c. Hence, the lemmas and (17) imply that

$$Ent\Big(T_\epsilon f\Big) \quad = \quad Ent\Big(T_{\epsilon_1} h\Big) \quad \le \quad \frac{1}{2} \cdot \Big(1 - H(\epsilon)\Big) \ - \ \Omega\Big(\lambda \cdot \gamma\Big) \ + \ o_{\lambda,\gamma \to 0}\Big(\lambda \cdot \gamma\Big) \ + \ e(n)$$

That is, for a sufficiently small $\delta > 0$, such that $0 \le \alpha, \beta, \lambda \le \delta$, the claim of (16) holds.

It remains to prove the lemmas. For that purpose, we will need the following version of the logarithmic Sobolev inequality for the boolean cube.

Lemma 5.4: *Let g be a nonnegative function on $\{0,1\}^n$. Let $\mathcal{E}(g,g)$ be the Dirichlet form, given by $\mathcal{E}(g,g) = \mathbb{E}_{x \in \{0,1\}^n} \mathbb{E}_{y \sim x} \Big(g(y) - g(x)\Big)^2$. Then*

$$\mathbb{E}(g,g) \quad \ge \quad 2 \ln 2 \cdot \mathbb{E}\, g \cdot Ent\Big(g\Big)$$

Proof:

We start with a simple auxiliary claim.

Let $x_1 \ge x_2 \ge ... \ge x_N$ be nonnegative numbers summing to 1. Then the numbers $y_k = \frac{x_k^2}{\sum_{i=1}^N x_i^2}$, for $k = 1, ..., N$, majorize $\{x_k\}$, that is

$$y_1 \ \ge \ x_1, \quad y_1 + y_2 \ \ge \ x_1 + x_2, \quad \ldots \quad , y_1 + ... + y_N \ = \ 1 \ = \ x_1 + ... + x_N$$

To see this, fix some $1 \le t \le N$. We have to show $\sum_{k=1}^t x_k^2 \ \ge \ \left(\sum_{k=1}^t x_k\right) \cdot \Big(\sum_{k=1}^N x_k^2\Big)$.

We may and will assume that all of the x_k are strictly positive. After some rearrangement, the claim reduces to showing

$$\frac{\sum_{k=1}^t x_k^2}{\sum_{k=1}^t x_k} \quad \ge \quad \frac{\sum_{m=t+1}^N x_m^2}{\sum_{k=t+1}^N x_m}$$

This holds because the LHS is lowerbounded by x_t, and the RHS is upperbounded by x_{t+1}.

A simple corollary of this claim is that for any nonnegative not identically zero function g on a finite domain endowed with uniform measure, holds that $g^2 / \mathbb{E}\, g^2$ majorizes $g / \mathbb{E}\, g$.

This is well-known (see [7]) to imply that $g/\mathbb{E}\, g$ is a convex combination of permuted versions of $g^2/\mathbb{E}\, g^2$. Since the entropy functional is linear and convex, this implies

$$Ent\Big(g^2\Big) \quad \geq \quad \frac{\mathbb{E}\, g^2}{\mathbb{E}\, g} \cdot Ent\Big(g\Big) \quad \geq \quad \mathbb{E}\, g \cdot Ent\Big(g\Big)$$

The claim of the lemma follows from this inequality combined with the logarithmic Sobolev inequality [6]

$$\mathcal{E}(g,g) \quad \geq \quad 2\ln 2 \cdot Ent\Big(g^2\Big)$$

∎

We are going to use the Walsh-Fourier expansion for functions on the boolean cube, writing a function g as $\sum_{S\subseteq[n]} \widehat{g}(S)\cdot W_S$, where $\Big\{W_S\Big\}_{S\subseteq[n]}$ is the Walsh-Fourier basis [8]. In particular, for the Dirichlet form, we have $\mathcal{E}(g,g) = 4\cdot\sum_{S\subseteq[n]} |S|\widehat{g}^2(S)$. Hence the preceding lemma implies

$$Ent\Big(g\Big) \quad \leq \quad \frac{2}{\ln 2}\cdot\frac{1}{\mathbb{E}\, g}\cdot\sum_{S\subseteq[n]} |S|\ \widehat{g}^2(S) \tag{18}$$

We will also need the following precise version of an inequality of [3], due to [4]:

Theorem 5.5: *There exists a universal constant $L > 0$ with the following property. For $g:\ \{0,1\}^n \to \{-1,1\}$, let $\rho = \Big(\sum_{A\subseteq[n]:|A|\geq 2}\widehat{g}^2(A)\Big)^{1/2}$. Then there exists some $B\subseteq[n]$ with $|B|\leq 1$ such that*

$$\sum_{A\subseteq[n]:|A|\leq 1, A\neq B} \widehat{g}^2(A) \quad \leq \quad L\cdot\rho^4 \ln\left(\frac{2}{\rho}\right)$$

and $|\widehat{g}(B)|^2 \geq 1\ -\ \rho^2\ -\ L\cdot\rho^4\ln\left(\frac{2}{\rho}\right)$.

Consider the boolean function f given in Theorem 1.11. Let $g = 2f-1$. Then $g:\ \{0,1\}^n \to \{-1,1\}$. Note that $\widehat{g}(0) = 2\widehat{f}(0) - 1$, and that $\widehat{g}(S) = 2\widehat{f}(S)$, for $|S| > 0$.

In particular, $\widehat{g}(0) = 2\,\mathbb{E}\, f - 1 = -2\beta$, and $\widehat{g}(\{1\}) = 2(1-\alpha)\,\mathbb{E}\, f = (1-\alpha)(1-2\beta)$.

Recall that $0\leq\alpha,\beta\leq\delta$, and that $\gamma = \alpha+\beta$. Hence, assuming δ is sufficiently small, we have

$$\sum_{|A|\geq 2}\widehat{f}^2(A) \quad\leq\quad \sum_{|A|\geq 2}\widehat{g}^2(A) \quad\leq\quad 1\ -\ \widehat{g}^2(\{1\}) \quad\leq\quad L\cdot\gamma, \tag{19}$$

for some absolute constant L.

Applying Theorem 5.5 to the function g, we get, for a sufficiently large constant L_1,

$$\sum_{k=2}^{n}\widehat{f}^2\Big(\{k\}\Big) \quad\leq\quad \sum_{k=2}^{n}\widehat{g}^2\Big(\{k\}\Big) \quad\leq\quad L_1\cdot\gamma^2\ln\left(\frac{1}{\gamma}\right) \tag{20}$$

Proof of Lemma 5.2

Fix $B \subseteq [n]$, with $|B| = \lambda_2 n$. Let $g_B = \mathbb{E}\Big(h \mid B\Big)$.

Note that $g_B = \sum_{S \subseteq B} \widehat{h}(S) \cdot W_S$, and hence, by (18), we have

$$Ent\Big(g_B\Big) \quad \le \quad \frac{2}{\ln 2} \cdot \frac{1}{\mathbb{E}\, g_B} \cdot \sum_{S \subseteq [B]} |S|\ \widehat{h}^2(S) \quad = \quad \frac{2}{\ln 2} \cdot \frac{1}{\mathbb{E}\, h} \cdot \sum_{S \subseteq [B]} |S|\ \widehat{h}^2(S)$$

Hence,

$$\mathop{\mathbb{E}}_{|B|=\lambda_2 n, 1 \notin B} Ent\Big(h \mid B\Big) \quad \le \quad \frac{2}{\ln 2} \cdot \frac{1}{\mathbb{E}\, h} \cdot \mathop{\mathbb{E}}_{|B|=\lambda_2 n, 1 \notin B} \sum_{S \subseteq [B]} |S|\ \widehat{h}^2(S) \quad \le$$

$$\frac{2}{\ln 2} \cdot \frac{1}{\mathbb{E}\, h} \cdot \sum_{S, 1 \notin S} |S| \lambda_2^{|S|}\ \widehat{h}^2(S)$$

Recall that $h = T_{\epsilon_\tau} f$. This means (see [8]) that for any $S \subseteq [n]$, holds $\widehat{h}(S) = \tau^{|S|/2} \cdot \widehat{f}(S)$. In particular, $|\widehat{h}(S)| \le |\widehat{f}(S)|$. Applying (19) and (20), we have that, for a sufficiently large absolute constant L, the last expression is bounded by

$$L \cdot \left(\lambda_2^2 \cdot \gamma \ + \ \lambda_2 \cdot \gamma^2 \ln\left(\frac{1}{\gamma}\right)\right)$$

This concludes the proof of the lemma. ∎

Proof of Lemma 5.3

Let $g = \mathbb{E}\left(\frac{f}{\mathbb{E} f} \;\Big|\; \{1\}\right)$. Then g is a function on a 2-point space $\{0,1\}$, with $g(0) = 2 - \alpha$ and $g(1) = \alpha$.

Observe that the noise operator commutes with the projection operator. Hence, since $h = T_{\epsilon_\tau} f$, we have $g_1 := \mathbb{E}\left(\frac{h}{\mathbb{E} h} \;\Big|\; \{1\}\right) = T_{\epsilon_\tau} g$.

Observe also that, by the definition of Mrs. Gerber's function ϕ, we have

$$\phi\left(Ent\left(\frac{h}{\mathbb{E}\, h} \;\Big|\; \{1\}\right),\ \epsilon_2\right) \quad = \quad Ent\Big(T_{\epsilon_2} g_1\Big) \quad = \quad Ent\Big(T_{\epsilon_2} T_{\epsilon_\tau} g\Big) \quad \le$$

$$Ent\Big(T_{\epsilon_1} T_{\epsilon_\tau} g\Big) \ + \ e(n) \quad = \quad Ent\Big(T_{\epsilon} g\Big) \ + \ e(n)$$

The inequality follows from the definition of ϵ_1 and ϵ_2, and the last equality follows from the definition of ϵ_1 and ϵ_τ.

It is easy to verify that $T_\epsilon g(0) = 1 + (1-\alpha)\cdot\lambda^{1/2}$ and that $T_\epsilon g(1) = 1 - (1-\alpha)\cdot\lambda^{1/2}$.

Hence, $Ent\Big(T_\epsilon g\Big) = 1 - H\left(\frac{1-(1-\alpha)\cdot\lambda^{1/2}}{2}\right)$.

Recall that

$$H\left(\frac{1-x}{2}\right) \quad = \quad 1 \quad - \quad \frac{1}{\ln 2}\cdot\sum_{k=1}^{\infty}\frac{1}{2k(2k-1)}\cdot x^{2k}$$

with the series converging absolutely for $-1 \le x \le 1$.

Let $F(x) = 1 - H\left(\frac{1-\sqrt{x}}{2}\right)$, for $0 \le x \le 1$. Then $F(x) = \frac{1}{\ln 2}\cdot\sum_{k=1}^{\infty}\frac{1}{2k(2k-1)}\cdot x^k$.

It is a convex function on $[0,1]$, and hence for any $0 \le x < y \le 1$ holds $F(y) - F(x) \ge (y-x)\cdot F'(x)$. The derivative F' is given by $F'(x) = \frac{1}{2\ln 2}\cdot\sum_{k=1}^{\infty}\frac{1}{2k-1}\cdot x^{k-1}$, with the series converging absolutely for x bounded away from 1.

Hence $F' \ge \frac{1}{2\ln 2}$ on $(0,1)$, and $F(y) - F(x) \ge \frac{1}{2\ln 2}\cdot(y-x)$. Applying this with $y = \lambda$ and $x = (1-\alpha)^2\cdot\lambda$, we get

$$\Big(1 - H(\epsilon)\Big) \; - \; Ent\Big(T_\epsilon g\Big) \quad = \quad F(\lambda) \; - \; F\Big((1-\alpha)^2\cdot\lambda\Big) \quad \ge \quad c_1\cdot\lambda\cdot\alpha$$

where $c_1 > 0$ is an absolute constant.

To conclude the proof of the lemma, note that, for a sufficiently small α, we have $Ent\Big(T_\epsilon g\Big) \ge c_2\cdot\lambda$, for an absolute constant c_2, and hence

$$\mathbb{E}\,h\cdot\phi\left(Ent\left(\frac{h}{\mathbb{E}\,h}\;\Big|\;\{1\}\right),\;\epsilon_2\right) \quad \le \quad \left(\frac{1}{2}-\beta\right)\cdot Ent\Big(T_\epsilon g\Big) \; + \; e(n) \quad \le$$

$$\frac{1}{2}\cdot\Big(1-H(\epsilon)\Big) \; - \; c\cdot\lambda\cdot(\alpha+\beta) \; + \; e(n) \quad = \quad \frac{1}{2}\cdot\Big(1-H(\epsilon)\Big) \; - \; c\cdot\lambda\cdot\gamma \; + \; e(n)$$

for a sufficiently small absolute constant c.

This completes the proof of the lemma. ∎

The proof of Lemma 5.1 is somewhat harder. We present it in the next subsection.

5.1 Proof of Lemma 5.1

We proceed similarly to the proof of Lemma 5.2, and use the notation introduced in that proof.

Given a function g on the boolean cube, we write $\mathbb{E}\left(g \mid x_1 = 0, x_2, ..., x_k\right)$ for the restriction of $\mathbb{E}\left(g \mid x_1, x_2, ..., x_k\right)$ on the subcube $x_1 = 0$, and similarly for $\mathbb{E}\left(g \mid x_1 = 1, x_2, ..., x_k\right)$.

We note that for $g = \sum_{S \subseteq [n]} \widehat{g}(S) \cdot W_S$, we have

$$\mathbb{E}\Big(g \mid x_1 = 0, x_2, ..., x_n\Big) \quad = \quad \sum_{T \subseteq [n], 1 \notin T} \Big(\widehat{g}(T) \; + \; \widehat{g}(T \cup \{1\})\Big) \cdot W_T$$

and

$$\mathbb{E}\Big(g \mid x_1 = 1, x_2, ..., x_n\Big) \quad = \quad \sum_{T \subseteq [n], 1 \notin T} \Big(\widehat{g}(T) \; - \; \widehat{g}(T \cup \{1\})\Big) \cdot W_T$$

We will also use the following easily verifiable identity, holding for nonnegative functions g:

$$Ent\Big(g\Big) \; - \; Ent\Big(g \mid \{1\}\;\Big) \quad = \quad \frac{1}{2} \cdot Ent\Big(g \mid x_1 = 0, x_2, ..., x_n\Big) \; + \; \frac{1}{2} \cdot Ent\Big(g \mid x_1 = 1, x_2, ..., x_n\Big)$$

As before, let $g_B = \mathbb{E}\Big(h \mid B\Big)$, for a subset $B \subseteq [n]$, with $|B| = \lambda_2 n$. Note that if $1 \in B$, then $\mathbb{E}\Big(g_B \mid \{1\}\Big) = \mathbb{E}\Big(h \mid \{1\}\Big)$.

Hence

$$\mathop{\mathbb{E}}_{|B| = \lambda_2 n, \; 1 \in B} \Big(Ent\Big(h \mid B\Big) - Ent\Big(h \mid \{1\}\Big)\Big) \quad = \quad \mathop{\mathbb{E}}_{|B| = \lambda_2 n, \; 1 \in B} \Big(Ent\Big(g_B\Big) - Ent\Big(g_B \mid \{1\}\Big)\Big) \quad =$$

$$\frac{1}{2} \cdot \mathop{\mathbb{E}}_{|B| = \lambda_2 n, \; 1 \in B} Ent\Big(g_B \mid x_1 = 0, x_2, ..., x_n\Big) \; + \; \frac{1}{2} \cdot \mathop{\mathbb{E}}_{|B| = \lambda_2 n, \; 1 \in B} Ent\Big(g_B \mid x_1 = 1, x_2, ..., x_n\Big)$$

We will prove the lemma by showing that, for a sufficiently large absolute constant L, holds both

$$\mathop{\mathbb{E}}_{|B| = \lambda_2 n, \; 1 \in B} Ent\Big(g_B \mid x_1 = 0, x_2, ..., x_n\Big) \quad \le \quad L \cdot \lambda_2 \cdot \gamma \tag{21}$$

and

$$\mathop{\mathbb{E}}_{|B| = \lambda_2 n, \; 1 \in B} Ent\Big(g_B \mid x_1 = 1, x_2, ..., x_n\Big) \quad \le \quad L \cdot \left(\lambda_2 \cdot \gamma \; + \; \gamma^2 \ln\left(\frac{1}{\gamma}\right)\right) \; + \; e(n) \tag{22}$$

Proof of (21)

Fix a subset $B \subseteq [n]$, with $|B| = \lambda_2 n$, and $1 \in B$. Recall that $g_B = \sum_{S \subseteq B} \widehat{h}(S) \cdot W_S$, and hence

$$\mathbb{E}\Big(g_B \mid x_1 = 0, x_2, ..., x_n\Big) = \sum_{T \subseteq B \setminus \{1\}} \Big(\widehat{h}(T) \ + \ \widehat{h}(T \cup \{1\})\Big) \cdot W_T$$

In particular,

$$\mathbb{E}\Big(g_B \mid x_1 = 0\Big) \quad = \quad \widehat{h}(0) \ + \ \widehat{h}(\{1\}) \quad = \quad \widehat{f}(0) \ + \ \tau^{1/2} \cdot \widehat{f}(\{1\}) \quad \geq \quad \mathbb{E} f$$

Applying (18), we have, for a sufficiently large constant L_1,

$$Ent\Big(g_B \mid x_1 = 0, x_2, ..., x_n\Big) \quad \leq \quad \frac{2}{\ln 2} \cdot \frac{1}{\mathbb{E} f} \cdot \sum_{T \subseteq B \setminus \{1\}} |T| \cdot \Big(\widehat{h}(T) \ + \ \widehat{h}(T \cup \{1\})\Big)^2 \quad \leq$$

$$L_1 \cdot \sum_{T \subseteq B \setminus \{1\}} |T| \cdot \Big(\widehat{h}^2(T) \ + \ \widehat{h}^2(T \cup \{1\})\Big)$$

Averaging over B, we have

$$\mathbb{E}_{|B| = \lambda_2 n, \ 1 \in B} \ Ent\Big(g_B \mid x_1 = 0, x_2, ..., x_n\Big) \ \leq \ L_1 \cdot \left(\sum_{T, 1 \notin T} |T| \lambda_2^{|T|} \widehat{h}^2(T) \ + \ \sum_{T, 1 \notin T} |T| \lambda_2^{|T|} \widehat{h}^2(T \cup \{1\})\right)$$

Using the fact that $|\widehat{h}(S)| \leq |\widehat{f}(S)|$ for all $S \subseteq [n]$, and applying (19) and (20), we have, for a sufficiently large constant L_2,

$$\sum_{T, 1 \notin T} |T| \lambda_2^{|T|} \widehat{h}^2(T) \quad \leq \quad L_2 \cdot \left(\lambda_2 \cdot \gamma^2 \ln\left(\frac{1}{\gamma}\right) \ + \ \lambda_2^2 \cdot \gamma\right)$$

and

$$\sum_{T, 1 \notin T} |T| \lambda_2^{|T|} \widehat{h}^2(T \cup \{1\}) \quad \leq \quad L_2 \cdot \lambda_2 \cdot \gamma$$

Summing up, this gives (21).

Proof of (22)

Similarly to the above,

$$\mathbb{E}\left(g_B \mid x_1 = 1, x_2, ..., x_n\right) = \sum_{T \subseteq B \setminus \{1\}} \left(\widehat{h}(T) \;-\; \widehat{h}(T \cup \{1\})\right) \cdot W_T$$

Which means that

$$\mathbb{E}\left(g_B \mid x_1 = 1\right) \;=\; \widehat{h}(0) \;-\; \widehat{h}(\{1\}) \;=\; \widehat{f}(0) \;-\; \tau^{1/2} \cdot \widehat{f}(\{1\}) \;=\; \mathbb{E}\, f \cdot \left(1 \;-\; \tau^{1/2} \cdot (1-\alpha)\right)$$

Recall that $\tau^{1/2} = 1$ if $\alpha \geq \lambda$ and $\tau^{1/2} = \frac{1-\lambda}{1-\alpha}$ otherwise. In both cases, note that we have $\mathbb{E}\left(g_B \mid x_1 = 1\right) \geq \lambda \cdot \mathbb{E}\, f$.

Applying (18), and averaging over B, we have, for a sufficiently large constant L_1,

$$\mathop{\mathbb{E}}_{|B|=\lambda_2 n,\ 1\in B} Ent\left(g_B \mid x_1 = 1, x_2, ..., x_n\right) \;\leq\; L_1 \cdot \frac{1}{\lambda} \cdot \sum_{T, 1\notin T} |T| \lambda_2^{|T|} \cdot \left(\widehat{h}(T) \;-\; \widehat{h}(T \cup \{1\})\right)^2$$

Let $g = \mathbb{E}\left(h \mid x_1 = 1, x_2, ..., x_n\right)$. Then $g = \sum_{T \subseteq [n], 1\notin T} \left(\widehat{h}(T) \;-\; \widehat{h}(T \cup \{1\})\right) \cdot W_T$. Hence

$$\mathop{\mathbb{E}}_{|B|=\lambda_2 n,\ 1\in B} Ent\left(g_B \mid x_1 = 1, x_2, ..., x_n\right) \;\leq\; L_1 \cdot \frac{1}{\lambda} \cdot \sum_{T, 1\notin T} |T| \lambda_2^{|T|} \cdot \widehat{g}^2(T) \tag{23}$$

Consider the function g. Since $h = T_{\epsilon_\tau} f$, we have

$$g \;=\; \epsilon_\tau \cdot T_{\epsilon_\tau}\left(\mathbb{E}\left(f \mid x_1 = 0, x_2, ..., x_n\right)\right) \;+\; \left(1 \;-\; \epsilon_\tau\right) \cdot T_{\epsilon_\tau}\left(\mathbb{E}\left(f \mid x_1 = 1, x_2, ..., x_n\right)\right)$$

For $i = 0, 1$, let $f_i = \mathbb{E}\left(f \mid x_1 = i, x_2, ..., x_n\right)$, and let $t_i = T_{\epsilon_\tau} f_i$. Note that for $i = 0, 1$ and for any T, $1 \notin T$, holds $|\widehat{t_i}(T)| \;\leq\; |\widehat{f_i}(T)|$.

Therefore, since $g = \epsilon_\tau \cdot t_0 + \left(1 - \epsilon_\tau\right) \cdot t_1$, we have, for any T, $1 \notin T$ that

$$\widehat{g}^2(T) \;\leq\; \epsilon_\tau \cdot \widehat{t_0}^2(T) \;+\; \left(1 \;-\; \epsilon_\tau\right) \cdot \widehat{t_1}^2(T) \;\leq\; \epsilon_\tau \cdot \widehat{f_0}^2(T) \;+\; \left(1 \;-\; \epsilon_\tau\right) \cdot \widehat{f_1}^2(T)$$

Hence,

$$\sum_{T, 1\notin T} |T| \lambda_2^{|T|} \cdot \widehat{g}^2(T) \;\leq\; \epsilon_\tau \cdot \sum_{T, 1\notin T} |T| \lambda_2^{|T|} \widehat{f_0}^2(T) \;+\; \left(1 - \epsilon_\tau\right) \cdot \sum_{T, 1\notin T} |T| \lambda_2^{|T|} \widehat{f_1}^2(T) \tag{24}$$

Exactly as above, we have the following upper bound for the first summand: For a sufficiently large constant L_2 holds

$$\sum_{T, 1 \notin T} |T| \lambda_2^{|T|} \widehat{f_0}^2(T) \;=\; \sum_{T, 1 \notin T} |T| \lambda_2^{|T|} \Big(\widehat{f}(T) \;+\; \widehat{f}(T \cup \{1\}) \Big)^2 \;\leq\; L_2 \cdot \lambda_2 \cdot \gamma$$

Consider the second summand. The function f_1 is a boolean function, whose expectation equals $\widehat{f}(0) - \widehat{f}(\{1\}) = \alpha \cdot \mathbb{E} f \leq \alpha$. Similarly, $\mathbb{E} f_1^2 = \mathbb{E} f_1 \leq \alpha$.

We now apply the inequality of [12], which states that

For a boolean function $g : \{0,1\}^m \to \{0,1\}$ *with expectation* $\mu \leq 1/2$ *holds* $\sum_{k=1}^m \widehat{g}^2(\{k\}) \leq L_3 \cdot \mu^2 \cdot \ln(1/\mu)$, *for a sufficiently large absolute constant* L_3.

In our case, this implies $\sum_{k=2}^n \widehat{f_1}^2 \Big(\{k\} \Big) \leq L_3 \cdot \alpha^2 \cdot \ln\left(\frac{1}{\alpha}\right)$, for a sufficiently large constant L_3.

This means that, for a sufficiently large constant L_4, we can upperbound the second summand in (24) by

$$\sum_{T, 1 \notin T} |T| \lambda_2^{|T|} \widehat{f_1}^2(T) \;\leq\; L_4 \cdot \left(\lambda_2 \cdot \alpha^2 \ln\left(\frac{1}{\alpha}\right) \;+\; \lambda_2^2 \cdot \alpha \right)$$

Recall that for $\alpha < \lambda$, we have $\epsilon_\tau = \frac{1 - \tau^{1/2}}{2} = \frac{1 - (1-\lambda)/(1-\alpha)}{2} \leq L_5 \cdot \lambda$, for an absolute constant L_5; and that for $\alpha \geq \lambda$, we have $\epsilon_\tau = 0$. Plugging these estimates into (24), we have

$$\sum_{T, 1 \notin T} |T| \lambda_2^{|T|} \cdot \widehat{g}^2(T) \;\leq\; L_2 \cdot L_5 \cdot \lambda \cdot \lambda_2 \cdot \gamma \;+\; L_4 \cdot \left(\lambda_2 \cdot \alpha^2 \ln\left(\frac{1}{\alpha}\right) \;+\; \lambda_2^2 \cdot \alpha \right)$$

And hence, coming back to (23), and recalling that $\lambda \leq \lambda_2 \leq c \cdot \lambda + e(n)$, for some absolute constant c, we have, for sufficiently large absolute constants L, L', that

$$\mathop{\mathbb{E}}_{|B| = \lambda_2 n, \; 1 \in B} Ent\Big(g_B \mid x_1 = 1, x_2, ..., x_n \Big) \;\leq\; L' \cdot \left(\lambda_2 \cdot \gamma \;+\; \alpha^2 \ln\left(\frac{1}{\alpha}\right) \;+\; \lambda_2 \cdot \alpha \right) \;+\; e(n) \;\leq$$

$$L \cdot \left(\lambda_2 \cdot \gamma \;+\; \gamma^2 \ln\left(\frac{1}{\gamma}\right) \right) \;+\; e(n)$$

This completes the proof of (22), of Lemma 5.1, and of (16).

∎

5.2 Removing the error in (16)

We show that the error term in (16) can be removed, by considering the claim for direct products of a function f with other functions.

Notation: For $f:\ \{0,1\}^n \to \mathbb{R}$ and $g:\ \{0,1\}^k \to \mathbb{R}$, let the direct product $f \times g:\ \{0,1\}^{n+k} \to \mathbb{R}$ be given by

$$(f \times g)(x, y) = f(x) \cdot g(y)$$

The following properties are easily verifiable (and well-known):

1. $$\mathbb{E}(f \times g) \quad = \quad \mathbb{E}\, f \cdot \mathbb{E}\, g$$
2. More generally, for all $S \subseteq [n]$ and $T \subseteq [k]$ holds
$$\widehat{f \times g}(S, T) \quad = \quad \widehat{f}(S) \cdot \widehat{g}(T)$$
3. $$T_\epsilon(f \times g) \quad = \quad T_\epsilon f \times T_\epsilon g$$
4. $$Ent(f \times g) \quad = \quad \mathbb{E}\, f \cdot Ent(g) + \mathbb{E}\, g \cdot Ent(f)$$

Let now $f:\ \{0,1\}^n \to \{0,1\}$ be a function satisfying the conditions of Theorem 1.11. Let N be a large integer, let g be a constant-1 function on $\{0,1\}^N$, and let $F = f \times g$.

By the properties above, it is easy to see that $\mathbb{E}\, F = \mathbb{E}\, f$ and that for any $S \subseteq [n]$ holds $\widehat{F}(S, 0) = \widehat{f}(S)$. Hence F also satisfies the conditions of Theorem 1.11, and we have, by (16), and by the properties above, that

$$Ent\Big(T_\epsilon f\Big) \quad = \quad Ent\Big(T_\epsilon F\Big) \quad \le \quad \frac{1}{2} \cdot \Big(1 - H(\epsilon)\Big) \;+\; e(N + n),$$

where $e(\cdot)$ is the error term in (16).

Letting N go to infinity shows $Ent\Big(T_\epsilon f\Big) \le \frac{1}{2} \cdot \Big(1 - H(\epsilon)\Big)$, completing the proof of Theorem 1.11.

6 Remaining proofs

6.1 Proof of Lemma 1.2

We have

$$I\Big(f(X); Y\Big) \;=\; H\Big(f(X)\Big) \;-\; H\Big(f(X)|Y\Big) \;=\; H\Big(f(X)\Big) \;-\; \mathop{\mathbb{E}}_{y} H\Big(f(X)|Y = y\Big) \;=$$

$$H\Big(f(X)\Big) \;-\; \mathop{\mathbb{E}}_{y} H\Big(\left(T_\epsilon f\right)(y)\Big)$$

Clearly

$$H\Big(f(X)\Big) \quad = \quad \mathbb{E}\, f \log \frac{1}{\mathbb{E}\, f} \ + \ (1 - \mathbb{E}\, f) \log \frac{1}{1 - \mathbb{E}\, f}$$

We also have (all the logarithms are binary)

$$\mathbb{E}_y H\Big((T_\epsilon f)\,(y)\Big) \quad = \quad \mathbb{E}_y \left((T_\epsilon f)\,(y) \log \frac{1}{(T_\epsilon f)\,(y)} \ + \ \Big(1 - (T_\epsilon f)\,(y)\Big) \log \frac{1}{1 - (T_\epsilon f)\,(y)} \right) \quad =$$

$$-\Big(Ent\Big(T_\epsilon f\Big) \ + \ \mathbb{E}\, T_\epsilon f \log \mathbb{E}\, T_\epsilon f \Big) \ - \ \Big(Ent\Big(T_\epsilon (1-f)\Big) \ + \ \mathbb{E}\, T_\epsilon (1-f) \log \mathbb{E}\, T_\epsilon (1-f) \Big) \quad =$$

$$-\Big(Ent\Big(T_\epsilon f\Big) + Ent\Big(T_\epsilon (1-f)\Big)\Big) \ + \ \mathbb{E}\, f \log \frac{1}{\mathbb{E}\, f} \ + \ (1 - \mathbb{E}\, f) \log \frac{1}{1 - \mathbb{E}\, f}$$

In the last step we have used the fact $\mathbb{E}\, T_\epsilon g \ = \ \mathbb{E}\, g$ for any function g. The claim of the lemma follows. ∎

6.2 Proof of Corollary 1.9

Note that for a boolean function f holds $Ent\Big(f\Big) + Ent\Big(1-f\Big) \le 1$. Hence, applying Corollary 1.8 to the functions f and $1-f$, we obtain, by Lemma 1.2:

$$I\Big(f(X); Y\Big) \quad = \quad Ent\Big(T_\epsilon f\Big) \ + \ Ent\Big(T_\epsilon (1-f)\Big) \quad \le$$

$$\mathbb{E}_{|B|=v} Ent\Big(f \mid B\Big) \ + \ \mathbb{E}_{|B|=v} Ent\Big((1-f) \mid B\Big) \ + \ O\left(\sqrt{\frac{\log n}{n}}\right)$$

To conclude the proof of the corollary, it suffices to show that for any $B \subseteq [n]$ holds

$$Ent\Big(f \mid B\Big) \ + \ Ent\Big((1-f) \mid B\Big) \quad = \quad I\Big(f(X); \ \{X_i\}_{i \in B}\Big)$$

To see this, we proceed exactly as in the proof of Lemma 1.2, observing that, by the definition,

$$Pr\Big\{f(X) = 1 \ \Big| \ \{X_i\}_{i \in B}\Big\} \quad = \quad \mathbb{E}\Big(f \mid B\Big)$$

Here we interpret both sides as functions of $\{X_i\}$, $i \in B$.

∎

6.3 Proof of Theorem 1.12

Let δ be the constant in the theorem. We will assume in the following argument that δ is sufficiently small, and, in particular, is at most as large as the constant in Theorem 1.11,

Let ϵ be a noise parameter, such that $(1-2\epsilon)^2 \le \delta$. Denote $\lambda = (1-2\epsilon)^2$.

Let $f:\ \{0,1\}^n \to \{0,1\}$ be a balanced function, that is $\mathbb{E} f = 1/2$.

Notation: In the following argument L_i, with $i = 1, ..., 3$, denote absolute constants.

Applying the result of [10] (Corollary 1), we have (in our notation) that, for a sufficiently small λ, holds

$$I\Big(f(X);Y\Big) \quad \le \quad \left(\frac{2}{\ln 2} \cdot \sum_{k=1}^{n} \widehat{f}^2(\{k\})\right) \cdot \lambda \quad + \quad L_1 \cdot \lambda^2$$

Next, note that, as in the proof of Lemma 5.3, we have $1 - H(\epsilon) \ = \ 1 - H\left(\frac{1-\lambda^{1/2}}{2}\right) \ge \frac{1}{2\ln 2} \cdot \lambda$.

Therefore, if $\sum_{k=1}^n \widehat{f}^2(\{k\}) \ < \ \frac14 - L_2 \cdot \lambda$, for a sufficiently large L_2, then $I\Big(f(X);Y\Big) < 1 - H(\epsilon)$.

Hence, we may assume $\sum_{k=1}^n \widehat{f}^2(\{k\}) \ > \ \frac14 - L_2 \cdot \lambda$.

Let $g = 2f-1$. Then $g:\ \{0,1\}^n \to \{-1,1\}$, and $\sum_{k=1}^n \widehat{g}^2(\{k\}) \ > \ 1 - 4L_2 \cdot \lambda$.

If λ is sufficiently small, we may apply Theorem 5.5, obtaining that there exists an index $1 \le k \le n$ such that $\widehat{g}^2(\{k\}) \ge 1 - L_3 \cdot \lambda$, for some sufficiently large L_3.

This means that

$$|\widehat{f}(\{k\})| \quad \ge \quad \frac12 \ - \ \frac{L_3}{2} \cdot \lambda \quad = \quad (1-\alpha) \cdot \mathbb{E} f,$$

with $\alpha = L_3 \cdot \lambda$.

Therefore, f satisfies the conditions of Theorem 1.11.

Since $\mathbb{E}(1-f) = 1/2$ and $\widehat{1-f}(S) = -\widehat{f}(S)$, for $|S| > 0$, the function $1-f$ satisfies these conditions as well.

Hence,

$$Ent\Big(T_\epsilon f\Big) \quad \le \quad \frac12 \cdot \Big(1 \ - \ H(\epsilon)\Big) \quad \text{and} \quad Ent\Big(T_\epsilon(1-f)\Big) \quad \le \quad \frac12 \cdot \Big(1 \ - \ H(\epsilon)\Big)$$

This, by Lemma 1.2, gives

$$I\Big(f(X);Y\Big) \quad = \quad Ent\Big(T_\epsilon f\Big) \quad + \quad Ent\Big(T_\epsilon(1-f)\Big) \quad \le \quad 1 \ - \ H(\epsilon),$$

completing the proof.

∎

6.4 Proof of Proposition 2.2

We repeat the statement of the proposition for the reader's convenience.

Proposition: *Let h be a nonnegative function on $\{0,1\}^2$ Then*

$$I_{T_{\epsilon_{\{1\}}}h}\left(\{2\},1\right) \quad \le \quad (1-2\epsilon)^2 \cdot I_h\left(\{2\},1\right)$$

Proof:

By homogeneity, we may and will assume $\mathbb{E}\, h = 1$.

Note that

$$I_h\left(\{2\},1\right) \quad = \quad I_h\left(\{1\},2\right) \quad = \quad Ent(h) \; - \; Ent\Big(h \mid \{1\}\Big) \; - \; Ent\Big(h \mid \{2\}\Big) \quad =$$

$$\frac{1}{2}\cdot Ent\Big(h \mid x_2 = 0\Big) \; + \; \frac{1}{2}\cdot Ent\Big(h \mid x_2 = 1\Big) \; - \; Ent\Big(h \mid \{1\}\Big)$$

Similarly

$$I_{T_{\epsilon_{\{1\}}}h}\left(\{2\},1\right) \quad = \quad \frac{1}{2}\cdot Ent\Big(T_{\epsilon_{\{1\}}}h \mid x_2 = 0\Big) \; + \; \frac{1}{2}\cdot Ent\Big(T_{\epsilon_{\{1\}}}h \mid x_2 = 1\Big) \; - \; Ent\Big(T_{\epsilon_{\{1\}}}h \mid \{1\}\Big)$$

Let us write $E(x) = Ent\Big(x, 2-x\Big)$, for $0 \le x \le 2$.

Set $\theta = \frac{1}{2}\cdot\mathbb{E}\left(h \mid x_2 = 0\right)$. Note that $\mathbb{E}\, h = 1$ implies $1-\theta = \frac{1}{2}\cdot\mathbb{E}\left(h \mid x_2 = 1\right)$.

Finally, take $s = 2\cdot Pr\Big\{x_1 = 0 \mid x_2 = 0\Big\}$ and $t = 2\cdot \Pr\Big\{x_1 = 0 \mid x_2 = 1\Big\}$. Observe that $0 \le s, t \le 2$.

In this new notation,

$$I_h\left(\{2\},1\right) \quad = \quad \theta\cdot E(s) \; + \; (1-\theta)\cdot E(t) \; - \; E\Big(\theta E(s) + (1-\theta)E(t)\Big)$$

And, using (2), we have

$$I_{T_{\epsilon_{\{1\}}}h}\left(\{2\},1\right) \quad = \quad \theta\cdot\phi\Big(E(s)\Big) \; + \; (1-\theta)\cdot\phi\Big(E(t)\Big) \; - \; \phi\Big(E\Big(\theta s + (1-\theta)t\Big)\Big)$$

The statement of the proposition can be now rephrased as follows:

$$\theta\cdot\phi\Big(E(s)\Big) \quad + \quad (1-\theta)\cdot\phi\Big(E(t)\Big) \quad - \quad \phi\Big(E\Big(\theta s + (1-\theta)t\Big)\Big) \quad \le$$

$$(1-2\epsilon)^2\cdot\Big(\theta\cdot E(s) \; + \; (1-\theta)\cdot E(t) \; - \; E\Big(\theta s + (1-\theta)t\Big)\Big) \tag{25}$$

For all $0 \leq s, t \leq 2$, $0 \leq \theta \leq 1$, and $0 \leq \epsilon \leq 1/2$.

Let $F(x) = (1-2\epsilon)^2 \cdot E(x) - \phi(E(x))$. Then this claim is equivalent to

$$\theta \cdot F(s) \ + \ (1-\theta) \cdot F(t) \quad \geq \quad F\Big(\theta s + (1-\theta) t\Big)$$

That is, to the fact that F is a convex function on $[0,2]$. We will verify this in the next lemma.

Lemma 6.1: *The second derivative F'' is nonnegative on* $(0,2)$.

Proof: Recall that $\phi(x) = \phi(x, \epsilon) = 1 - H\Big((1-2\epsilon) \cdot H^{-1}(1-x) + \epsilon\Big)$.

Note also that $E(s) = 1 - H\left(s/2\right)$, for $0 \leq s \leq 2$. Hence $\phi(E(s)) = 1 - H\Big((1/2 - \epsilon) \cdot s + \epsilon\Big)$.

We need to show that $(1-2\epsilon)^2 \cdot E'' \geq \Big(\phi(E)\Big)''$. Computing the second derivatives,

$$\left(\phi(E)\right)''(s) = \frac{(1-2\epsilon)^2}{\ln 2} \cdot \frac{1}{\Big((1-2\epsilon) \cdot s + 2\epsilon\Big) \cdot \Big((2-2\epsilon) - (1-2\epsilon) \cdot s\Big)}$$

And $E'' = \frac{1}{\ln 2} \cdot \frac{1}{s(2-s)}$. Hence we need to check

$$s \cdot (2-s) \quad \leq \quad \Big((1-2\epsilon) \cdot s + 2\epsilon\Big) \cdot \Big((2-2\epsilon) - (1-2\epsilon) \cdot s\Big)$$

and this is easily verifiable.

∎

Acknowledgments

We are grateful to Yuval Kochman for many very helpful conversations and valuable remarks. We also thank Venkat Chandar and Or Ordentlich for valuable remarks.

References

[1] N. Alon and J. Spencer, **The Probabilistic Method**, 3rd ed. Hoboken, NJ: Wiley, 2008.

[2] V. Chandar, personal communication, 2014.

[3] E.Friedgut, G. Kalai, and A. Naor, *Boolean functions whose Fourier transform is concentrated on the first two levels*, Advances in Applied Mathematics, 2002.

[4] J. Jendrej, K. Oleszkiewicz, and J. O. Wojtaszczyk, *On some extensions of the FKN theorem*, preprint, 2013.

[5] T. Courtade and G. Kumar, *Which boolean functions maximize mutual information on noisy inputs?* IEEE Transactions on Information Theory, vol. 60, no. 8, pp. 4515-4525, 2014.

[6] M. Ledoux, **Concentration of measure and logarithmic Sobolev inequaities**, 1997.

[7] A. W. Marshall, and I. Olkin, **Inequalities: Theory of Majorization and Its Applications**, Academic, New York, 1979.

[8] R. O'Donnel, **Analysis of Boolean functions**, Cambridge U.P., 2014.

[9] O. Ordentlich, personal communication, 2015.

[10] O. Ordentlich, O. Shayevitz, and O.Weinstein, *An Improved Upper Bound for the Most Informative Boolean Function Conjecture*, preprint, 2015.

[11] S. Sachdeva, A. Samorodnitsky, and I. Shahaf, *On conjectures of Kumar and Courtade*, manuscript, 2014.

[12] M. Talagrand, *How much are increasing sets positively correlated?*, Combinatorica 16.2, 243-258, 1996.

[13] A. D. Wyner and J. Ziv, *A theorem on the entropy of certain binary sequences and applications: Part I*, IEEE Trans. Inform. Theory, vol. 19, no. 6, pp. 769-772, 1973.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de