A Simpler Proof of the Four Functions Theorem and Some New Variants

Dimitris Achlioptas
Department of Informatics & Telecommunications
University of Athens
Athens, Greece
optas@uoa.gr

Kostas Zampetakis
Department of Computer Science & Engineering
University of California Santa Cruz
Santa Cruz, U.S.A.
kzampeta@ucsc.edu

***Abstract*—The celebrated Four Functions Theorem of Ahlswede and Daykin is a functional correlation inequality on distributive lattices with myriad applications. Ruozzi proved a variant of the inequality and used it to settle a major conjecture in the area of graphical models. We prove a new functional correlation inequality in the same vein which simplifies the proof of both the Four Functions Theorem and of Ruozzi's inequality and suggests a unified picture for correlation inequalities on distributive lattices.**

***Index Terms*—Correlation-type Inequalities, Four Functions Theorem, Super-modular Functions**

I. Introduction

For a function f on a finite domain we denote the sum of f over its entire domain by $\mathcal{Z}(f)$. When f expresses relative probability, $\mathcal{Z}(f)$ is also known as the "partition function" and $f/\mathcal{Z}(f)$ is a probability distribution.

Closely related to the problem of partition function estimation are functional correlation inequalities, such as the FKG inequality [4], the Holley inequality [5], and the Fishburn-Shepp [3], [10] inequality. Remarkably, all of these inequalities are special cases of the celebrated *Four Functions Theorem* of Ahlswede and Daykin [2]. While this theorem is generally stated for distributive lattices and sums over arbitrary subsets of the domain, its heart is the more readily accessible Theorem 1 below, from which one can easily derive even the most general formulation, as we discuss in Section IV.

For $x, y \in \{0,1\}^n$ let us denote by $x \vee y$ and $x \wedge y$ the bitwise OR and AND of x and y, respectively.

Theorem 1 (Four Functions)**.** *If $f_1, f_2, g_1, g_2 : \{0,1\}^n \mapsto \mathbb{R}_{\geq 0}$ are such that for all x, y,*

$$f_1(x)f_2(y) \leq g_1(x \vee y)g_2(x \wedge y) \ , \tag{1}$$

then $\mathcal{Z}(f_1)\mathcal{Z}(f_2) \leq \mathcal{Z}(g_1)\mathcal{Z}(g_2)$.

Rinott and Saks [7] and, independently, Aharoni and Keich [1] generalized Theorem 1 so that there are k functions on each side of (1) instead of two, for any $k \geq 2$. In order to state this generalization and other related results, and to connect these results with partition function estimation, it will be convenient to introduce the following notation.

Given a 0/1 matrix A, we denote by $\uparrow A$ the matrix obtained by sorting each column so that the 1s are on top and by $\overleftarrow{A}$ the matrix obtained by sorting each row so that the 1s are on the left. We write A_{i*}, A_{*j} for the i-th row and j-th column of A, respectively. We say that a function $f : \{0,1\}^{k \times n} \mapsto \mathbb{R}_{\geq 0}$ is *k-product* if $f(X) = \prod_{i \in [k]} f_i(X_{i*})$, for some non-negative functions $f_1, \ldots, f_k$ on $\{0,1\}^n$.

With this notation, the generalization of the Four Functions Theorem to $2k$ functions can be stated as follows.

Theorem 2 ([7], [1])**.** *If $f, g : \{0,1\}^{k \times n} \mapsto \mathbb{R}_{\geq 0}$ are k-products and $f(X) \leq g(\uparrow X)$ for all X, then $\mathcal{Z}(f) \leq \mathcal{Z}(g)$.*

We strongly encourage the reader at this point to take the time and verify the equivalence of Theorem 1 with the $k = 2$ case of Theorem 2, as in the rest of the paper we will be using exclusively the notation of the latter.

Recall that a function f is log-supermodular if $f(x)f(y) \leq f(x \vee y)f(x \wedge y)$, and log-submodular if the reverse inequality holds. (Note that x, y may be binary matrices, or even tensors, instead of binary vectors. Since $\vee, \wedge$ are element-wise operations, the "shape" of x, y is immaterial.) More than twenty years after the generalization of the Four Functions Theorem to $2k$ functions, in a seminal work Ruozzi [8] showed that Theorem 2 continues to hold if the function f being bounded is log-supermodular instead of k-product.

Theorem 3 ([8])**.** *If $f, g : \{0,1\}^{k \times n} \mapsto \mathbb{R}_{\geq 0}$ are such that f is log-supermodular, g is k-product, and $f(X) \leq g(\uparrow X)$ for all X, then $\mathcal{Z}(f) \leq \mathcal{Z}(g)$.*

Ruozzi [8] used Theorem 3 and Vontobel's characterization of the Bethe approximation in terms of covers [11], in order to prove that if a function f can be expressed as a binary, attractive graphical model, then $\mathcal{Z}(f)$ is bounded from below by its Bethe approximation, a major advance in the area of graphical models.

II. Our Contribution

Our contribution is to suggest that all three theorems mentioned so far, i.e., the Four Functions Theorem, its generalization to $2k$ functions, and Ruozzi's variant for log-supermodular functions, may be fragments of a larger picture. To that end we prove a new variant of the Four Functions Theorem involving log-*sub*modular functions. Our theorem greatly simplifies the proof of all aforementioned theorems.

A. Log-submodular Functions as Upper Bounds

Our new inequality asserts that Theorem 3 continues to hold if the *bounding* function g is log-submodular instead of k-product, so that a log-submodular function bounds a log-supermodular function. To emphasize the parallels between the different results, in Theorem 4 we bundle our new inequality (case (a)) with Theorem 2 (case (b)) and Theorem 3 (case (c)).

Definition 1. *If functions $f, g : \{0,1\}^{k\times n} \mapsto \mathbb{R}_{\geq 0}$ are such that $f(X) \leq g(\uparrow X)$ for all X, we write $f \prec_{\uparrow} g$.*

Theorem 4. *If $f \prec_{\uparrow} g$ and*

(a) f is log-supermodular and g is log-submodular, or

(b) f is k-product and g is k-product, or [7], [1]

(c) f is log-supermodular and g is k-product, [8]

then $\mathcal{Z}(f) \leq \mathcal{Z}(g)$.

Besides its inherent interest, our new inequality (case (a)) greatly simplifies both the proof of the $2k$ functions theorem (case (b)) and of Ruozzi's variant for log-supermodular functions (case (c)). Indeed, our combined, self-contained proof of all three cases of Theorem 4 occupies less than one page.

B. Discussion

If we think of the two functions f, g in Theorem 4 as landscapes over the domain $\{0,1\}^{k\times n}$, all three cases of the theorem conclude that the volume under f is at most the volume under g. To understand how this conclusion comes about, it is helpful to think of (i) the domain of f, g as B^n for some arbitrary base set B, instead of the specific $B = \{0,1\}^k$, and (ii) of $\uparrow$ as an arbitrary function $B^n \mapsto B^n$, instead of the specific function that pushes the 1s in each column to the top.

With the above in mind, consider the partition of B^n induced by $\uparrow$, where $X, Y \in B^n$ are in the same part iff $\uparrow X = \uparrow Y$. Since $\uparrow\uparrow X = \uparrow X$, the assumption that $f \prec_{\uparrow} g$, simply asserts that within each part, the highest peak of g is at least as high as the highest peak of f. Of course, this is very far from sufficient to conclude that $\mathcal{Z}(f) \leq \mathcal{Z}(g)$ (except for the trivial case where $\uparrow$ is the identity, i.e., $f(X) \leq g(X)$ for all X). That's where assumptions (a)–(c) come in: by imposing structure on f, g they constrain the relative drop-off rates from the respective peaks, enabling the theorem's conclusion.

A truly amazing property of the Four Functions Theorem is that it allows for functions where the f-volume *strictly exceeds* the g-volume in some parts. The conclusion $\mathcal{Z}(f) \leq \mathcal{Z}(g)$ is remarkably reached by combining information about different parts through projection, i.e., by induction on n. The same is true for Theorems 2,3, i.e., cases (b),(c) of Theorem 4. Naturally, for $n = 1$ the task of favorably combining information about different parts has to be done "by hand." Correspondingly, it is not an accident that the base case of the induction is, by far, the hardest part in each proof.

To get a flavor, we state the base case of Theorem 1, which, as noticed in [1], can be interpreted geometrically. Let L and R be two rectangles, each subdivided into four rectangles as in Fig. 1. For $i \in \{0,1\}$, let $f_1(i) := \alpha_i$, $f_2(i) := \beta_i$, $g_1(i) := \gamma_i$, and $g_2(i) := \delta_i$. The base case of the Four Functions Theorem asserts the following highly non-obvious fact: if the lower left, upper right, and lower right subrectangles of L have areas bounded by the areas of the corresponding subrectangles of R, while the area of the upper left subrectangle of L is bounded by the area of the *lower* right subrectangle of R, then the area of L is bounded by the area of R.

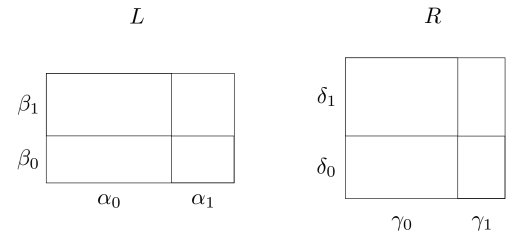

Fig. 1. Geometric interpretation of Theorem 1 for $n = 1$

Our Theorem 4(a), for $n = 1, k = 2$, can also be interpreted geometrically. But now L and R can be *arbitrary* shapes, divided into four pieces each as in Fig. 2, and in order to conclude that the area of L is bounded by the area of R it is enough that: (i) the product of the areas of the two pieces on the "main diagonal" of L is less than or equal than the product of the areas of its other two pieces (reflecting that f is log-supermodular), and (ii) the reverse inequality holds for R (reflecting that g is log-submodular). This holds trivially in the rectangular setting, as for any rectangle subdivided into four, the product of the areas of its subrectangles in the main diagonal equals the product of the areas of the other two subrectangles. Thus, Theorem 4(a), establishes *simultaneously* the base cases of both Theorems 4(b) and 4(c), giving a unified and simplified proof of both results (see Section III).

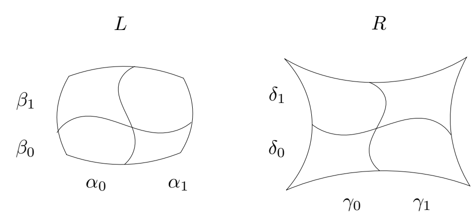

Fig. 2. Geometric interpretation of Theorem 4(a) for $k = 2, n = 1$.

Somewhat surprisingly, the proof of our Theorem 4(a) is *not* inductive. We consider this to be an indication of getting closer to the right "level of abstraction" governing these inequalities. Naturally, deriving a non-inductive proof of cases (b) and (c) would be further illuminating. For example, the fact that Theorem 4 does not cover the mirror image of case (c), wherein f is k-product and g is log-submodular, is not a coincidence. As observed by Ruozzi [8], the inductive proof of these two cases crucially relies on the bounding function being closed under marginalization, a property that does not hold for log-submodular functions.

III. Proof of Theorem 4

A. Case (a)

For any two functions f, g with common domain D, say that f is *weakly log-majorized* by g if for every $A \subseteq D$, there exists $B \subseteq D$ with $|B| = |A|$, such that $\prod_{x\in A} f(x) \le \prod_{x \in B} g(x)$.

Lemma 1 ([6] p. 168, Corollary 5.A.2.b.). *If a function f is weakly log-majorized by a function g, then $\mathcal{Z}(f) \le \mathcal{Z}(g)$.*

Theorem 4(a) follows from Lemma 1 and the following.

Theorem 5. *Under the conditions of Theorem 4(a), f is weakly log-majorized by g.*

Proof. For a $k \times t \times n$ tensor T (think of k as height, t as width, and n as depth) call each $k \times n$ matrix a slice, each $k \times t$ matrix a screen, and each $t \times 1$ vector a rod. We write $\uparrow T$ for the tensor resulting by applying $\uparrow$ to each slice and $\overleftarrow{T}$ for the tensor resulting by applying $\leftarrow$ to each screen.

To prove that f is weakly log-majorized by g it suffices to prove that for every $t \ge 1$, for every t-subset of $\{0,1\}^{k\times n}$, if we arbitrarily stack the t matrices in the subset to form a tensor $T \in \{0,1\}^{k\times t\times n}$, we can find a tensor $U \in \{0,1\}^{k \times t \times n}$ with distinct slices, i.e., another t-subset of $\{0,1\}^{k\times n}$, such that $F(T) := \prod_{i\in[t]} f(T_{*i*}) \le \prod_{i \in [t]} g(U_{*i*}) := G(U)$.

Given T, to find U we let π be any permutation of the rods of T that keeps each rod within its screen while rearranging (sorting) the rods of each screen in order of weight (number of ones). We let $U = \pi(T)$. Observe that U has distinct slices as it is the result of permuting the rods of T, i.e., of a tensor with distinct slices. Crucially, observe that $\uparrow(\overleftarrow{T}) = \pi(\overleftarrow{T}) = \overleftarrow{\pi(T)} = \overleftarrow{U}$. Invoking first the log-supermodularity of f, then the fact $f \prec_{\uparrow} g$, then the fact $\uparrow(\overleftarrow{T}) = \overleftarrow{U}$, and finally the log-submodularity of g, we see that

$$F(T) \le F\left(\overleftarrow{T}\right) \le G\left(\uparrow(\overleftarrow{T})\right) = G\left(\overleftarrow{U}\right) \le G(U) \quad . \tag{2}$$

□

B. Case (b)

Proof. Let $P(t)$ denote the proposition that Theorem 4(b) holds for $n = t$ and all $k \ge 1$. We proceed by induction.

When $n = 1$, the fact that f is k-product is equivalent to it being log-modular, i.e., log-submodular and log-supermodular. Therefore, by hypothesis, $f, g : \{0,1\}^k \mapsto \mathbb{R}_{\ge 0}$ are such that f is log-supermodular, g is log-submodular, and $f \prec_{\uparrow} g$. Thus, Theorem 4(a) applies yielding $\mathcal{Z}(f) \le \mathcal{Z}(g)$, i.e., $P(1)$. We note that establishing the base case by applying our Theorem 4(a) is our only simplification of the proof, as the inductive step is as the original (the same is true fore case (c)).

For $n \ge 2$, let $\overline{f}, \overline{g} : \{0,1\}^{k\times(n-1)} \mapsto \mathbb{R}_{\ge 0}$ denote the sum of f, g, respectively, over all 2^k possible last columns. Since $\mathcal{Z}(\overline{f}) = \mathcal{Z}(f)$ and $\mathcal{Z}(\overline{g}) = \mathcal{Z}(g)$ it suffices to prove that $\mathcal{Z}(\overline{f}) \le \mathcal{Z}(\overline{g})$. For this we observe that, trivially, $\overline{f}, \overline{g}$ are k-product, because f, g, respectively, are k-product. Thus, if can we prove $\overline{f} \prec_{\uparrow} \overline{g}$ we can invoke $P(n-1)$ and conclude.

To prove $\overline{f} \prec_{\uparrow} \overline{g}$ we define for arbitrary $Y \in \{0,1\}^{k\times(n-1)}$ functions $f_*, \widehat{g} : \{0,1\}^k \mapsto \mathbb{R}_{\ge 0}$, as $f_*(\boldsymbol{t}) = f(Y \mid \boldsymbol{t})$ and $\widehat{g}(\boldsymbol{t}) = g(\uparrow Y \mid \boldsymbol{t})$. Since $\mathcal{Z}(f_*) = \overline{f}(Y)$ and $\mathcal{Z}(\widehat{g}) = \overline{g}(\uparrow Y)$, we are left to prove $\mathcal{Z}(f_*) \le \mathcal{Z}(\widehat{g})$. For this we observe that (i) $f_* \prec_{\uparrow} \widehat{g}$ since $f_*(\boldsymbol{t}) = f(Y \mid \boldsymbol{t}) \le g(\uparrow(Y \mid \boldsymbol{t})) = \widehat{g}(\uparrow \boldsymbol{t})$, as $f \prec_{\uparrow} g$, and that (ii) product functions are closed under restriction. Thus, $f_*, \widehat{g}$ satisfy the conditions of $P(1)$. □

C. Case (c)

Proof. The reasoning is identical to that of case (b), except that now $f, \overline{f}, f_*$ are log-supermodular instead of k-product. Specifically, for f, f_*, log-supermodularity is given and trivial, respectively. To prove that $\overline{f}$ is log-supermodular we use the Four Functions Theorem, i.e., case (b) for $k = 2$.

Let $h : \{0,1\}^n \mapsto \mathbb{R}_{\ge 0}$ be log-supermodular and let $S \subseteq [n]$ be arbitrary. For $\boldsymbol{x} \in \{0,1\}^S$, define $h_{\boldsymbol{x}} : \{0,1\}^{n-|S|} \mapsto \mathbb{R}_{\ge 0}$ with $h_{\boldsymbol{x}}(\boldsymbol{t}) := h(\boldsymbol{x}, \boldsymbol{t})$. Given any $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^S$, applying the Four Functions Theorem for $f_1 = h_{\boldsymbol{x}}$, $f_2 = h_{\boldsymbol{y}}$, $g_1 = h_{\boldsymbol{x}\vee\boldsymbol{y}}$, and $g_2 = h_{\boldsymbol{x}\wedge\boldsymbol{y}}$, yields $\mathcal{Z}(h_{\boldsymbol{x}})\mathcal{Z}(h_{\boldsymbol{y}}) \le \mathcal{Z}(h_{\boldsymbol{x}\vee\boldsymbol{y}})\mathcal{Z}(h_{\boldsymbol{y}\wedge\boldsymbol{y}})$, i.e., that the function that results after summing over all variables not in S is log-supermodular. □

IV. Distributive Lattices

Recall that a lattice L is a partially ordered set such that every two elements, x and y, have a unique minimal upper bound, denoted by $x \vee y$, and a unique maximal lower bound, denoted by $x \wedge y$. If $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ for all $x, y, z \in L$, we say that lattice L is *distributive*. The Four Functions Theorem is usually stated in the framework of distributive lattices and with a stronger conclusion, claiming the inequality for arbitrary marginals, as follows.

Theorem 6. *Let L be a finite distributive lattice. If f_1, f_2, g_1, g_2 are non-negative real valued functions on L such that for all $x, y \in L$,*

$$f_1(x) f_2(y) \le g_1(x \vee y) g_2(x \wedge y) \quad , \tag{3}$$

then, for all $X, Y \subseteq L$,

$$f_1(X) f_2(Y) \le g_1(X \vee Y) g_2(X \wedge Y) \quad , \tag{4}$$

where $X \vee Y = \{x \vee y : x \in X, y \in Y\}$, $X \wedge Y = \{x \wedge y : x \in X, y \in Y\}$, and $f(X) = \sum_{x\in X} f(x)$.

Since every distributive lattice can be embedded in the subsets of some set, Theorem 1 readily implies the case of Theorem 6 above, where $X = Y = L$. Moreover, it is not hard to see that this case implies Theorem 6 in its generality as follows: modify f_1, f_2 to equal 0 for x not in X, Y, respectively; it is trivial to check that (3) continues to hold after doing so.

The above observations apply also to all three cases of Theorem 4. In particular, Theorem 4 holds for non-binary alphabet as well, which can be useful for counting problems such as counting graph colorings. Finally, Theorem 4 extends even to the case of continuous variables, with the obvious modifications. Indeed, the continuous version of Theorem 4(c), was exploited by Ruozzi in [9].

ACKNOWLEDGEMENTS

We are grateful to the anonymous reviewers. Their suggestions greatly improved the presentation of the paper.

REFERENCES

[1] Ron Aharoni and Uri Keich. A generalization of the Ahlswede-Daykin inequality. *Discrete Mathematics*, 152(1):1–12, 1996.

[2] Rudolf Ahlswede and David E. Daykin. An inequality for the weights of two families of sets, their unions and intersections. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 43(3):183–185, 1978.

[3] Peter C. Fishburn. A correlational inequality for linear extensions of a poset. *Order*, 1(2):127–137, 1984.

[4] Cornelis M. Fortuin, Pieter W. Kasteleyn, and Jean Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103, 1971.

[5] Richard Holley. Remarks on the FKG inequalities. *Communications in Mathematical Physics*, 36(3):227–231, 1974.

[6] Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. *Inequalities: Theory of Majorization and its Applications*, volume 143. Springer, second edition, 2011.

[7] Yosef Rinott and Michael Saks. Correlation inequalities and a conjecture for permanents. *Combinatorica*, 13(3):269–277, 1993.

[8] Nicholas Ruozzi. The bethe partition function of log-supermodular graphical models. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[9] Nicholas Ruozzi. A Lower Bound on the Partition Function of Attractive Graphical Models in the Continuous Case. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1048–1056. PMLR, 20–22 Apr 2017.

[10] Larry A. Shepp. The XYZ Conjecture and the FKG Inequality. *The Annals of Probability*, 10(3):824 – 827, 1982.

[11] Pascal O. Vontobel. Counting in graph covers: A combinatorial characterization of the bethe entropy function. *IEEE Trans. Inf. Theory*, 59(9):6018–6048, 2013.