Proceedings CWIT 2005, June 5–8 2005, Montréal, Canada, 28–31

INEQUALITIES FOR SHANNON ENTROPIES AND ADHESIVITY OF POLYMATROIDS

*František Matúš**

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
18208 Prague, P.O.Box 18, Czech Republic
e-mail: matus@utia.cas.cz

ABSTRACT

Given a random vector, the collection of Shannon entropies of its subvectors gives rise to an entropy function that can be considered for a polymatroid. The entropy functions admit special pasting which translates to a notion of adhesivity of polymatroids. The classes of polymatroids that have adhesive restrictions or copies are shown to contain the entropy functions. Information inequalities and their applications are discussed as consequences of this containment.

1 INTRODUCTION

For a random vector $\xi = (\xi_i)_{i \in N}$ indexed by a finite set N, the *entropy function* h_ξ of ξ maps each subset I of N to the Shannon entropy of the subvector $\xi_I = (\xi_i)_{i \in I}$. This contribution deals with the problem whether a given function g on the power set $\mathcal{P}(N)$ of N is equal to the entropy function of a random vector that takes a finite number of values. By basic properties of the entropy, obvious necessary conditions for g to be equal to an entropy function require that $g(\emptyset) = 0$, g is nondecreasing, thus $g(I) \leqslant g(J)$ for $I \subseteq J$, and submodular [2]. This means the nonnegativity of $g(I)+g(J)-g(I\cup J)-g(I\cap J)$, denoted by $\varDelta_{I,J}\, g$, for any $I, J \subseteq N$. A pair (N, g) with g satisfying these three requirements is called a *polymatroid* on the ground set N with the rank function g [4, 9], making usually no difference between a polymatroid and its rank function. A polymatroid is *entropic* if it is equal to the entropy function of a vector taking a finite number of values.

A real function ϕ on $\mathbb{R}^{\mathcal{P}(N)}$, the Euclidean space hosting rank functions, generates an *information inequality* if $\phi(g) \geqslant 0$ is valid for all entropic polymatroids g. If the nonnegativity holds even for all polymatroids (N, g) then the inequality is of *Shannon type*, and if it holds only for the entropic polymatroids g satisfying additional constraints then the inequality is *conditional*, following [13, 14]. The information inequalities and related problems are summarized in [13, 12], and earlier in [15], presenting first unconditional non-Shannon type one.

In Section 2, a simple necessary condition for a polymatroid to be entropic is proved in Lemma 1. It is based on the observation that restrictions of an entropic polymatroid can be pasted together in a special way, which gives rise to a notion of adhesivity of polymatroids. Lemma 3 in Section 3 presents even a stronger necessary condition of selfadhesivity, a notion that captures the adhesivity of copies of the polymatroid. Each of the conditions is equivalent to a finite number of unconditional linear information inequalities, comprising all known unconditional non-Shannon type information inequalities and a new one, see Theorem 1 and Remark 1. Finally, Theorem 2 describes the selfadhesive polymatroids with the ground set of cardinality four in terms of the unconditional non-Shannon type information inequality of Zhang and Yeung [15, Theorem 3]. Detailed versions and extensions of these results are in [8] where the focus is on the polymatroids rather than entropy functions.

Section 4 is devoted to applications of the necessary conditions in optimization problems involving entropy functions. The information rates of perfect secret sharing schemes and implication problem for conditional independences are discussed in detail.

2 ADHESIVITY OF RESTRICTIONS OF POLYMATROIDS

If $N \subseteq M$ and two polymatroids (N, g) and (M, h) satisfy $g(I) = h(I)$ for $I \subseteq N$ then the former is the restriction of the latter to N and the latter is an extension of the former to M. Two polymatroids are *adhesive* or *adhere* if a polymatroid $(N \cup M, f)$ extends each of them and $\varDelta_{N,M}\, f = 0$. Any such polymatroid f is called an adhesive extension of g and h.

Lemma 1. *Two restrictions of an entropic polymatroid have an adhesive extension that is entropic.*

Proof. Let I, J be two subsets of N, $M = I \cup J$, and h_ξ the entropy function of a random vector $\xi = (\xi_i)_{i \in N}$. If ξ_i ranges in a finite set X_i let $P(x_K)$ denote the probability that ξ_K attains the

*This work has been supported by the grant A1075104 of Grant Agency of Academy of Sciences of the Czech Republic and by the grant 201/04/0393 of GA CR.

value $x_K = (x_k)_{k \in K}$ with $x_k \in X_k$, $K \subseteq N$. Define $Q(x_M)$ equal to zero if $P(x_{I \cap J}) = 0$ and $P(x_I)P(x_J)/P(x_{I \cap J})$ otherwise; here x_I, x_J and $x_{I \cap J}$ are the coordinate projections of x_M. Then, Q is equal to the distribution of a random vector $\eta = (\eta_i)_{i \in M}$. By the construction, ξ_I and η_I are equidistributed, and hence the restrictions of h_ξ and h_η to I coincide. This holds also when I is replaced by J, by symmetry. Since $\eta_{I \setminus J}$ and $\eta_{J \setminus I}$ are conditionally independent given $\eta_{I \cap J}$ it follows that $\Delta_{I,J}\, h_\eta$ equals 0. Thus, h_η is an adhesive extension of the restrictions of h_ξ to I and J. $\square$

The set of rank functions of polymatroids (N, g) with the same ground set N is defined by a finite number of nonstrict linear homogeneous inequalities [16], and thus is a closed polyhedral cone, denoted by $\boldsymbol{H}_N$. Its subset given by $\Delta_{I,J}\, g = 0$ is obviously a face of $\boldsymbol{H}_N$, denoted in the sequel by $\boldsymbol{H}_N^{I,J}$, $I, J \subseteq N$.

Lemma 2. *The restrictions of a polymatroid (N, g) to subsets I, J of N adhere if and only if the coordinate projection $\Pi_N^{I,J}$ of $\mathbb{R}^{\mathcal{P}(N)}$ onto $\mathbb{R}^{\mathcal{P}(I) \cup \mathcal{P}(J)}$ maps g into the $\Pi_{I \cup J}^{I,J}$-image of $\boldsymbol{H}_{I \cup J}^{I,J}$.*

Proof. The projection $\Pi_N^{I,J}(g)$ belongs to $\Pi_{I \cup J}^{I,J}(\boldsymbol{H}_{I \cup J}^{I,J})$ if and only if a polymatroid f in $\boldsymbol{H}_{I \cup J}^{I,J}$ exists such that $\Pi_N^{I,J}(g)$ equals $\Pi_{I \cup J}^{I,J}(f)$. The latter means coincidence of the restrictions of g and f to I and J for some polymatroid $(I \cup J, f)$ with $\Delta_{I,J}\, f$ equal to 0, thus adhesivity of the restrictions of g. $\square$

For $I, J \subseteq N$ let $\boldsymbol{H}_{N,I,J}^{\mathsf{ar}}$ denote the set of polymatroids from $\boldsymbol{H}_N$ whose restrictions to I and J adhere and $\boldsymbol{H}_N^{\mathsf{ar}}$ the intersection of $\boldsymbol{H}_{N,I,J}^{\mathsf{ar}}$ over all I, J. By Lemma 1, the entropic polymatroids from $\boldsymbol{H}_N$ are contained in $\boldsymbol{H}_N^{\mathsf{ar}}$.

Corollary 1. *The set $\boldsymbol{H}_N^{\mathsf{ar}}$ is a polyhedral cone.*

Proof. By Lemma 2, $\boldsymbol{H}_{N,I,J}^{\mathsf{ar}}$ equals the intersection of $\boldsymbol{H}_N$ and $(\Pi_N^{I,J})^{-1} \Pi_{I \cup J}^{I,J}(\boldsymbol{H}_{I \cup J}^{I,J})$. The assertion follows by mentioning that faces, linear images, their preimages and intersections of polyhedral cones are polyhedral. $\square$

A set of linear inequalities that define the cone $\boldsymbol{H}_{N,I,J}^{\mathsf{ar}}$ is described in [8, Theorem 2]. The following assertion is actually a corollary of this result but preference is given to a simple direct proof.

Theorem 1. *If the cardinality of $N = \{1, \ldots, n, u, v, 0\}$ is at least four and a polymatroid (N, g) has adhesive restrictions to $I = \{1, \ldots, n, u, v\}$ and $J = \{u, v, 0\}$ then*

$$n\big[g_{u0} - g_u + g_{v0} - g_v\big] + \sum_{j=1}^{n} \big[g_{uj} + g_{vj} - g_{uvj}\big] \geqslant \\ (n+1)\big[g_{uv0} - g_{uv}\big] + (n-1)g_0 + g_{1\ldots n}$$

where g_{u0} denotes the value of g on the set $\{u, 0\}$, etc.

Proof. If g has adhesive restrictions to I, J and f is their adhesive extension then the inequality holds if and only if it holds with f replacing g because only the values $g(K)$ with $K \subseteq I$ or $K \subseteq J$ occur. After the replacement it rewrites to

$$\Delta_{uv,0|\{1,\ldots,n\}}\, f + \sum_{j=1}^{n} \big[\Delta_{u,v|j0}\, f + \Delta_{j,0|u}\, f + \Delta_{j,0|v}\, f\big] \\ + \sum_{j=1}^{n-1} \Delta_{\{1,\ldots,j\},\{j+1\}|0}\, f \geqslant \Delta_{\{1,\ldots,n\},0|uv}\, f + \sum_{j=1}^{n} \Delta_{0,j|uv}\, f\,,$$

where $\Delta_{uv,0|\{1,\ldots,n\}}$ abbreviates $\Delta_{\{1,\ldots,n,u,v\},\{1,\ldots,n,0\}}$, etc. It follows from the submodularity of f and $\Delta_{I,J}\, f = 0$ that the left-hand side of this inequality is nonnegative and all terms of the right-hand side vanish. $\square$

In particular, the inequality of Theorem 1 holds for the entropic polymatroids by Lemma 1. With this quantification it appeared in [14, Theorem 4], generalizing the special case [5, Theorem 1] with N of the cardinality 5.

Remark 1. Expressing $\boldsymbol{H}_N^{\mathsf{ar}}$ as the intersection of $\boldsymbol{H}_{N,I,J}^{\mathsf{ar}}$, the terms with I, J in inclusion are equal to $\boldsymbol{H}_N$ and can be ignored. By submodularity, it is not difficult to see that $\boldsymbol{H}_{N,I,J \cup (N \setminus I)}^{\mathsf{ar}}$ is contained in $\boldsymbol{H}_{N,I,J}^{\mathsf{ar}}$, and thus the intersection can run only over I, J covering N. Also the pairs I, J such that $I \cap J$ has at most one element, or both $J \setminus I$ and $I \setminus J$ are singletons are irrelevant in this intersection [8, Section 2]. Hence, $\boldsymbol{H}_N = \boldsymbol{H}_N^{\mathsf{ar}}$ if N has at most four elements. If $N = \{1, 2, u, v, 0\}$ then a computation with the software package PORTA [1], implementing Fourier-Motzkin elimination [16], indicates that $\boldsymbol{H}_N^{\mathsf{ar}}$ consists of the polymatroids g that satisfy the inequalities obtained by permutations of the elements of N from the inequalities

$$\Delta_{uv|1}\, g + \Delta_{uv|2}\, g + \Delta_{1,2}\, g - \Delta_{u,v}\, g \\ + \big[\Delta_{uv|0}\, g + \Delta_{u0|v}\, g + \Delta_{v0|u}\, g\big] \geqslant 0$$

and

$$\Delta_{2v|1}\, g + \Delta_{2v|u}\, g + \Delta_{1,u}\, g - \Delta_{2,v}\, g \\ + \big[\Delta_{uv|0}\, g + \Delta_{u0|v}\, g + \Delta_{v0|u}\, g\big] \geqslant 0\,.$$

The former is equivalent to the inequality of Theorem 1 if N has five elements whereas the latter is a new unconditional information inequality; for a simple proof see [8, Corollary 4].

3 SELFADHESIVITY OF POLYMATROIDS

For a polymatroid (N, g) and bijection $\pi\colon N \to M$ such that $\pi(i) = i$ for $i \in N \cap M$ let

$$g^\pi(K) = g((K \cap N) \cup \pi^{-1}(K \cap M))\,, \qquad K \subseteq N \cup M\,.$$

The polymatroid $(N \cup M, g^\pi)$ obtains by parallel extensions, with elements $\pi(i)$ parallel to i, $i \in N \setminus M$ [4, p. 245]. Note that the restriction of g^π to $\pi(N) = M$ is an isomorphic copy of g. If this copy and g adhere then g is called *selfadhesive at* $N \cap M$. This notion depends on π only through the intersection. A polymatroid is *selfadhesive* if it is selfadhesive at each subset of its ground set.

Lemma 3. *The entropic polymatroids are selfadhesive.*

Proof. If $\xi = (\xi_i)_{i\in N}$ is a random vector and π a bijection between N and M with $\pi(i) = i$ for $i \in N \cap M$ then $(h_\xi)^\pi = h_\eta$ for the vector $\eta = (\eta_i)_{i\in N\cup M}$ consisting of η_i equal to ξ_i for $i \in N$ and $\xi_{\pi^{-1}(i)}$ otherwise. The assertion follows from Lemma 1 applied to the restrictions of h_η to N and M. □

Lemma 4. *If $I, J \subseteq N$ and a polymatroid (N, g) is selfadhesive at $I \cap J$ then the restrictions of g to I and J adhere.*

Proof. By assumption, for a bijection $\pi\colon N \to M$ such that $I \cap J = N \cap M$, a polymatroid $(N \cup M, f)$ with $\varDelta_{N,M}\, f = 0$ extends g and the restriction of g^π to M. Then, the polymatroid given by

$$h(K) = f((K \cap I) \cup \pi(K \cap J)), \qquad K \subseteq I \cup J,$$

extends the restrictions of g to I and J. It suffices to show that $\varDelta_{I,J}\, h = 0$, thus $\varDelta_{I,\pi(J)}\, f = 0$. This follows from $\varDelta_{N,M}\, f = 0$ and the submodularity of f. □

The set of polymatroids from $\boldsymbol{H}_N$ that are selfadhesive at I is denoted by $\boldsymbol{H}^{\mathsf{sa}}_{N,I}$ and the intersection of these sets over $I \subseteq N$ by $\boldsymbol{H}^{\mathsf{sa}}_N$.

Corollary 2. $\boldsymbol{H}^{\mathsf{sa}}_N \subseteq \boldsymbol{H}^{\mathsf{ar}}_N$

Proof. By Lemma 4, $\boldsymbol{H}^{\mathsf{sa}}_{N,I\cap J} \subseteq \boldsymbol{H}^{\mathsf{ar}}_{N,I,J}$. Apply the intersections over $I, J \subseteq N$. □

Lemma 5. *The set $\boldsymbol{H}^{\mathsf{sa}}_N$ is a polyhedral cone.*

Proof. It suffices to show that $\boldsymbol{H}^{\mathsf{sa}}_{N,N\cap M}$ is a polyhedral cone for a bijection $\pi\colon N \to M$. A polymatroid $g \in \boldsymbol{H}_N$ is selfadhesive at $N \cap M$ if and only if g^π belongs to $\boldsymbol{H}^{\mathsf{ar}}_{N\cup M,N,M}$. By Lemma 2, this is a polyhedral cone and, in turn, its intersection with the image $\boldsymbol{H}^\pi_N$ of $\boldsymbol{H}_N$ under the linear mapping $g \mapsto g^\pi$ is polyhedral. It follows that $\boldsymbol{H}^{\mathsf{sa}}_{N,N\cap M}$ consists of the restrictions of the polymatroids from $\boldsymbol{H}^\pi_N \cap \boldsymbol{H}^{\mathsf{ar}}_{N\cup M,N,M}$ to N, thus is a polyhedral cone. □

Remark 1 implies that $\boldsymbol{H}^{\mathsf{sa}}_N = \boldsymbol{H}_N$ if N has at most three elements. If the cardinality of N is four a complete description of $\boldsymbol{H}^{\mathsf{sa}}_N$ is presented in the following theorem. Recall the non-Shannon type information inequality of Zhang and Yeung [15, Theorem 3], rewritten as

$$\begin{aligned} 3\big[g(ik) + g(il) + g(kl)\big] + g(jk) + g(jl) \geqslant \\ g(i) + 2\big[g(k) + g(l)\big] + g(ij) + 4g(ikl) + g(jkl) \end{aligned}$$

where i, j, k, l are different singletons of N and the signs for union are omitted.

Theorem 2. *A polymatroid on a set of cardinality four is selfadhesive if and only if it satisfies the above inequality for all choices of different singletons.*

Proof. The necessity follows from Theorem 1 applied to the parallel extension of the polymatroid by one element. The sufficiency is much more sophisticated, based on analysis of special subcones of $\boldsymbol{H}_N$, see [8, Section 4]. □

4 DISCUSSION

In applications, for a real function $\varPsi$ on $\mathbb{R}^{\mathcal{P}(N)}$ the maximization of $\varPsi(h_\xi)$ over a random vector $\xi = (\xi_i)_{i\in N}$ subject to some constraints on the entropy function h_ξ occurs. Since optimization over distributions of random vectors may be very difficult or unrealistic it is sometimes useful to upper bound the supremum of $\varPsi(h_\xi)$ by the supremum of $\varPsi(g)$ over $g \in \boldsymbol{H}_N$ subject to the constraints. A new better bound can be obtained if g has additionally adhesive restrictions or is selfadhesive. These relaxations work directly if the defining inequalities of the cones g belongs to are known. To find the inequalities is, however not a straightforward task, being likely beyond the scope of existing software packages for manipulations with polyhedra, even for medium cardinalities of N. An alternative way is outlined in the special case when the entropic polymatroids are relaxed to $\boldsymbol{H}^{\mathsf{ar}}_{N,I,J}$, nevertheless the idea applies in general. Namely, the maximization over g in $\boldsymbol{H}^{\mathsf{ar}}_{N,I,J}$ can equivalently run over the pairs (g, f) with $g \in \boldsymbol{H}_N$ and $f \in \boldsymbol{H}^{I,J}_{I\cup J}$ whose restrictions to both I and J coincide. These constraints on (g, f) are in the form of linear inequalities. If $\varPsi$ and the original constraints are also linear then a linear programming problem provides the upper bound.

This idea is illustrated in two settings.

A perfect secret sharing scheme is a random vector $\xi = (\xi_i)_{i\in N}$ satisfying certain simple linear constraints on h_ξ, for references see [3]. The underlying set N has a distinguished element $\mathit{0}$ and the constraints express the independence or functional dependence of the secret ξ_{0} from or on some subvectors of $\xi_{N\setminus\mathit{0}}$, according to an access structure. The average information rate of the scheme is the value $\varPsi(h_\xi)$ of the function defined by

$$\varPsi(g) = \frac{g(\mathit{0})}{\frac{1}{n}\sum_{i\in N\setminus\mathit{0}} g(i)}$$

where n is the cardinality of the set of participants $N \setminus \mathit{0}$. The supremum of the rates $\varPsi(h_\xi)$ over the schemes ξ subject to the constraints (thus with the same access structure) is upper bounded by the supremum of $\varPsi(g)$ over $g \in \boldsymbol{H}^{\mathsf{sa}}_N$ subject to the constraints. By homogeneity, the bound equals the maximum of $g(\mathit{0})$ over $g \in \boldsymbol{H}^{\mathsf{sa}}_N$ subject to $\Sigma_{i\in N\setminus\mathit{0}}\, g(i) = n$ and the constraints. This approach might provide new bounds on rates of the schemes, in particular, improve bounds of the table of rates for access structures with five participants, see [3, p. 285].

In some statistical models it is of interest to decide whether conditional independence constraints on subvectors of a random vector imply another conditional independences [10]. In particular, given a pair (I, J) of subsets of a finite set N and a family $\mathcal{L}$ of such pairs, a question arises whether the conditional independence of $\xi_{I\setminus J}$ and $\xi_{J\setminus I}$ given $\xi_{I\cap J}$ is a consequence of the conditional independences of $\xi_{K\setminus L}$ and $\xi_{L\setminus K}$ given $\xi_{K\cap L}$, $(K, L) \in \mathcal{L}$, whatever the vector ξ taking a finite number of values can be. This implication problem for conditional independences has been solved only when the cardinality of N is at most four [7]. By basic properties of the Shannon entropy, the problem is equivalent to the question whether the supremum of $\varDelta_{I,J}\, h_\xi$ over ξ subject to the constraints $\varDelta_{K,L}\, h_\xi = 0$,

$(K,L) \in \mathcal{L}$, is zero or not. If the supremum m of $\Delta_{I,J}\, g$ over $g \in \boldsymbol{H}_N$ subject to $\Delta_{K,L}\, g = 0$, $(K,L) \in \mathcal{L}$, is zero then the implication for conditional independences is valid, as observed in [12, p. 276] or [6, p. 119]. This sufficient condition for the validity is equivalent to the i-implication of imsets by [11, Section 5], studied in detail in [10]. A stronger sufficient condition for the validity is $m^{\mathsf{ar}} = 0$ where m^{ar} is the supremum of $\Delta_{I,J}\, g$ over $g \in \boldsymbol{H}_N^{\mathsf{ar}}$ subject to the constraints imposed by $\mathcal{L}$. For example, given $N = \{\mathit{1}, \mathit{2}, \mathsf{u}, \mathsf{v}, \mathit{0}\}$, $I = \mathsf{u}$, $J = \mathsf{v}$ and

$$\mathcal{L} = \{(\mathit{1}\mathsf{u}, \mathit{1}\mathsf{v}), (\mathit{2}\mathsf{u}, \mathit{2}\mathsf{v}), (\mathit{1}, \mathit{2}), (\mathsf{u}\mathit{0}, \mathsf{v}\mathit{0}), (\mathsf{uv}, \mathsf{v}\mathit{0}), (\mathsf{uv}, \mathsf{u}\mathit{0})\},$$

it follows from the inequality of Theorem 1 rewritten in Remark 1 that $m^{\mathsf{ar}} = 0$, and the implication problem can be decided in positive. At the same time, it is not difficult to show that $m > 0$ which means that the i-implication of imsets does not decide this instance of the problem.

REFERENCES

[1] T. Christof, PORTA: A Polyhedron Representation Transformation Algorithm, http://www.zib.de/Optimization/Software/Porta, Version 1.3 revised by A. Loebel and M. Stoer.

[2] S. Fujishige (1978) Polymatroidal dependence structure of a set of random variables. *Inform. and Control* **39** 55–72.

[3] W.-A. Jackson and K.M. Martin (1996) Perfect secret sharing schemes on five participants. *Designs, Codes and Cryptography* **9** 267–286.

[4] L. Lovász (1982) Submodular functions and convexity. *Mathematical Programming – The State of the Art* (A. Bachem, M. Grötchel and B. Korte, eds.), Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 235–257.

[5] K. Makarychev, Yu. Makarychev, A. Romashchenko and N. Vereshchagin (2002) A new class of non-Shannon type inequalities for entropies. *Communications in Information and Systems* **2** 147–166.

[6] F. Matúš (1997) Conditional independence structures examined via minors. *Annals of Mathematics and Artificial Intelligence* **21** 99–128.

[7] F. Matúš (1999) Conditional independences among four random variables III: final conclusion. *Comb. Probab. & Computing* **8** 269–276.

[8] F. Matúš (2004) Adhesivity of polymatroids. (submitted to *Discrete Mathematics*)

[9] H. Narayan (1997) *Submodular Functions and Electrical Networks.* Elsevier, Amsterdam.

[10] M. Studený (2005) *Probabilistic Conditional Independence Structures.* Springer-Verlag London Ltd.

[11] M. Studený (2004) Structural Imsets: an algebraic method for describing conditional independence structures. *Proc. of IPMU*, Editrice Università La Sapienza, Italy, 1323–1330.

[12] R.W. Yeung (2002) *A First Course in Information Theory*. Kluwer Academic/Plenum Publ., New York.

[13] R.W. Yeung and Z. Zhang (2001) A class of non-Shannon-type information inequalities and their applications. *Communications in Information and Systems* **1** 87–100.

[14] Z. Zhang (2003) On a new non-Shannon type information inequality. *Communications in Information and Systems* **3** 47–60.

[15] Z. Zhang and R.W. Yeung (1998) On characterization of entropy function via information inequalities. *IEEE Trans. Information Theory* **44** 1440–1452.

[16] G.M. Ziegler (1998). *Lectures on Polytopes.* Graduate Texts in Mathematics **152**, Springer-Verlag, New York, Berlin, Heidelberg.