INFORMATION AND CONTROL 39, 55–72 (1978)

Polymatroidal Dependence Structure of a Set of Random Variables

SATORU FUJISHIGE

Department of Mathematical Engineering and Instrumentation Physics, Faculty of Engineering, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Given a finite set E of random variables, the entropy function h on E is a mapping from the set of all subsets of E into the set of all nonnegative real numbers such that for each $A \subseteq E$ $h(A)$ is the entropy of A. The present paper points out that the entropy function h is a β-function, i.e., a monotone nondecreasing and submodular function with $h(\varnothing) = 0$ and that the pair (E, h) is a polymatroid. The polymatroidal structure of a set of random variables induced by the entropy function is fundamental when we deal with the interdependence analysis of random variables such as the information-theoretic correlative analysis, the analysis of multiple-user communication networks, etc. Also, we introduce the notion of the principal partition of a set of random variables by transferring some results in the theory of matroids.

1. INTRODUCTION

The notion of "entropy" is fundamental in the Shannon theory of communication networks (cf. Gallager, 1968). By definition, the entropy of a set of discrete random variables X_k $(k = 1,..., n)$ is given by

$$h(\{X_1,..., X_n\}) = -\sum_{i_1=1}^{N_1} \cdots \sum_{i_n=1}^{N_n} P(X_1 = i_1,..., X_n = i_n) \log P(X_1 = i_1,..., X_n = i_n), \quad (1.1)$$

where X_k is assumed to take on values of $1,..., N_k$ $(k = 1,..., n)$ and $P(X_1 = i_1,..., X_n = i_n)$ denotes the probability that the values of the random variables X_k $(k = 1,..., n)$ are, respectively, equal to i_k $(k = 1,..., n)$. The value of the entropy $h(\{X_1,..., X_n\})$ is a functional of the probability distribution for random variables $X_1,..., X_n$. Properties of the entropy h as a functional of the probability distribution are well understood and there are several axiomatic definitions of the entropy h.

On the other hand, once the probability distribution for random variables $X_1,..., X_n$ is given and fixed, we have 2^n entropies $h(A)$ $(A \subseteq E \equiv \{X_1,..., X_n\})$

0019-9958/78/0391-0055$02.00/0

and we can thus regard h as a function from the set 2^E of all subsets of E into the set R_+ of all nonnegative real numbers, where, for each $A \subseteq E$, $h(A)$ is given similarly as (1.1) in terms of the marginal probability distribution for A and we put $h(\varnothing) = 0$. We call $h(:2^E \to R_+)$ the entropy function on $E = \{X_1, ..., X_n\}$.

Properties of the thus defined entropy function have not been fully understood, though the nonnegativity of the (conditional) mutual information in Shannon's sense and the subadditivity of the entropy function have been recognized (see Watanabe, 1960).

The present paper points out that the entropy function h on a finite set E of random variables is a β-function, i.e., a monotone nondecreasing and submodular function with $h(\varnothing) = 0$ and that the pair (E, h) is a polymatroid with the ground set E and the ground-set rank function h (see Sections 2 and 3). There are many information-theoretic problems of the polymatroidal type, where the entropy function plays a crucial role in analyzing the structural properties. As information-theoretic problems of the polymatroidal type we can take, for example, the information-theoretic correlative analysis (cf. McGill, 1954; Watanabe, 1960; and Han, 1975, 1978; and the analysis of multiple-user communication networks (cf. Wyner, 1974 and van der Meulen, 1977), which will be discussed in Section 4. It will be important to distinguish problems of the polymatroidal type from those of the nonpolymatroidal type arising in the Shannon theory of information.

Applications of the matroid and polymatroid theory have been recently made to engineering sciences such as electric-network analysis (Iri and Tomizawa, 1975; Recski, 1976), operations research (Iri and Tomizawa, 1976; Fujishige, 1976, 1977a, b), control systems analysis, etc. Major advantages of viewing problems from the matroid and polymatroid standpoints are that we can clearly understand the underlying structures which are essential to the problems under consideration and that some results concerning a specific problem of the (poly)matroidal type can be applied to a seemingly unrelated problem if it turns out to be of the (poly)matroidal type.

Since a polymatroid is a generalization of a matroid which is an abstraction of some aspect of graphs and matrices, we can develop a structural analysis of information-theoretic problems by extending notions and results related to matroids and graphs to the Shannon theory of information. In Section 5, we will transfer some results already obtained in the matroid theory to the interdependence analysis of random variables and introduce the notion of the principal partition of a set of random variables.

The present paper will throw a new light on the Shannon theory of information and provide a unifying approach to the structural analysis of random variables, communication networks, etc., from the point of view of polymatroids.

2. Preliminaries from Polymatroids

In this section we shall give several definitions with regard to polymatroids, which will be used in the following sections.

Denote by R_+ the set of all nonnegative real numbers. For a finite set A, denote by $|A|$ the cardinality of A.

Let E be a nonempty finite set and ρ be a function from the set 2^E of all subsets of E into the set R_+ such that

$$\text{(i)} \quad \rho(\varnothing) = 0, \tag{2.1}$$

$$\text{(ii)} \quad \rho(A) \leqslant \rho(B) \qquad (A \subseteq B \subseteq E), \tag{2.2}$$

$$\text{(iii)} \quad \rho(A) + \rho(B) \geqslant \rho(A \cup B) + \rho(A \cap B) \qquad (A, B \subseteq E). \tag{2.3}$$

Condition (2.2) together with (2.1) means that the set function ρ is nonnegative and monotone nondecreasing. A set function satisfying (2.3) is called a *submodular function*. Subadditivity of ρ follows from (2.1) and (2.3) by taking sets A and B in (2.3) such as $A \cap B = \varnothing$. The set function ρ is called a *β-function*. The pair (E, ρ) is called a *polymatroid*, where E is called the *ground set* and ρ the *ground-set rank function* or simply the *rank function* of the polymatroid.

Denote by R_+^E the set of all mappings from E into R_+. A mapping $\mathbf{x}$ $(\in R_+^E)$ will be regarded as a vector $(x(e))_{e \in E}$ with coordinates indexed by E, where, for each $e \in E$, $x(e)$ is the image of e by $\mathbf{x}$. A vector $\mathbf{x}$ $(\in R_+^E)$ is called an *independent vector* of polymatroid (E, ρ) if $\mathbf{x}$ satisfies

$$x(A) \leqslant \rho(A) \qquad (A \subseteq E), \tag{2.4}$$

where

$$x(A) = \sum_{e \in A} x(e). \tag{2.5}$$

We shall adopt an abbreviation similar to (2.5) for any vector in R_+^E in the following. The set of all independent vectors of polymatroid (E, ρ) is a convex polyhedron determined by the linear inequalities (2.4) and the nonnegativity of all the components of independent vectors.

Let us define an order relation $\preccurlyeq$ on R_+^E as follows. For vectors $\mathbf{x}$ and $\mathbf{y}$ in R_+^E,

$$\mathbf{x} \preccurlyeq \mathbf{y} \Leftrightarrow x(e) \leqslant y(e) \qquad (e \in E). \tag{2.6}$$

An independent vector of (E, ρ) which is maximal in the sense of the order relation $\preccurlyeq$ is called a *base of* (E, ρ). A vector $\mathbf{v}$ $(\in R_+^E)$ is called a *dominating vector of* (E, ρ) if for any independent vector $\mathbf{x}$ of (E, ρ) there holds

$$\mathbf{x} \preccurlyeq \mathbf{v}, \tag{2.7}$$

or if

$$\rho(\{e\}) \leqslant v(e) \qquad (e \in E). \tag{2.8}$$

For a dominating vector $\mathbf{v}$ of (E, ρ), a set function $\rho^*_{(\mathbf{v})}: 2^E \to R_+$ defined by

$$\rho^*_{(\mathbf{v})}(A) = v(A) + \rho(E - A) - \rho(E) \qquad (A \subseteq E) \tag{2.9}$$

is a β-function and $(E, \rho^*_{(\mathbf{v})})$ is called a *dual polymatroid of* (E, ρ) *with respect to* $\mathbf{v}$ or a $\mathbf{v}$-*dual polymatroid of* (E, ρ). (See Fig. 1.) A set function $\nu_{(\mathbf{v})}: 2^E \to R_+$ defined by

$$\nu_{(\mathbf{v})}(A) = v(A) - \rho(A) \qquad (A \subseteq E) \tag{2.10}$$

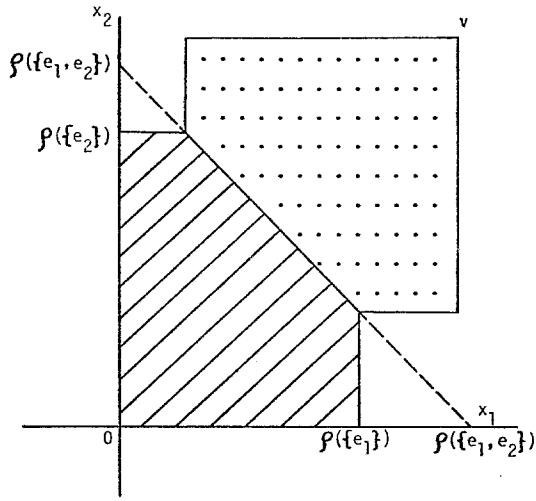

FIG. 1. A polymatroid and its dual: the shaded area and the dotted area, respectively, represent the sets of all independent vectors of a polymatroid (E, ρ) and its $\mathbf{v}$-dual polymatroid, when $E = \{e_1, e_2\}$.

with respect to polymatroid (E, ρ) and its dominating vector $\mathbf{v}$ is called the *nullity function of* (E, ρ) *with respect to dominating vector* $\mathbf{v}$. The $\rho^*_{(\mathbf{v})}$ of (2.9) can be expressed in terms of $\nu_{(\mathbf{v})}$ as

$$\rho^*_{(\mathbf{v})}(A) = \nu_{(\mathbf{v})}(E) - \nu_{(\mathbf{v})}(E - A) \qquad (A \subseteq E). \tag{2.11}$$

For any vector $\mathbf{x}$ $(\in R_+^E)$, the *reduction of* (E, ρ) *with respect to* $\mathbf{x}$ is a polymatroid $(E, \rho_{\mathbf{x}})$ with the ground-set rank function $\rho_{\mathbf{x}}$ defined by

$$\rho_{\mathbf{x}}(A) = \min_{D \subseteq A} \{\rho(D) + x(A - D)\} \qquad (A \subseteq E). \tag{2.12}$$

The set $P_{\mathbf{x}}$ of all independent vectors of $(E, \rho_{\mathbf{x}})$ is given by

$$P_{\mathbf{x}} = \{\mathbf{y} \mid \mathbf{0} \leqslant \mathbf{y} \leqslant \mathbf{x};\ \mathbf{y} \text{ is an independent vector of } (E, \rho)\}.$$

(See Fig. 2.) A base of $(E, \rho_{\mathbf{x}})$ is called a *base of* $\mathbf{x}$ *with respect to* (E, ρ).

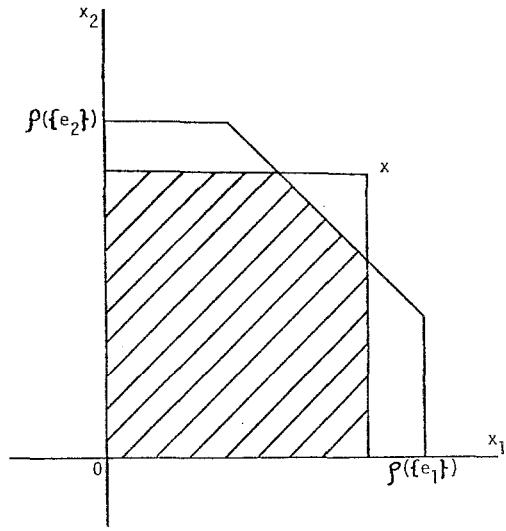

FIG. 2. A reduction: the shaded area represents the set of all independent vectors of the reduction of (E, ρ) with respect to $\mathbf{x}$, when $E = \{e_1, e_2\}$.

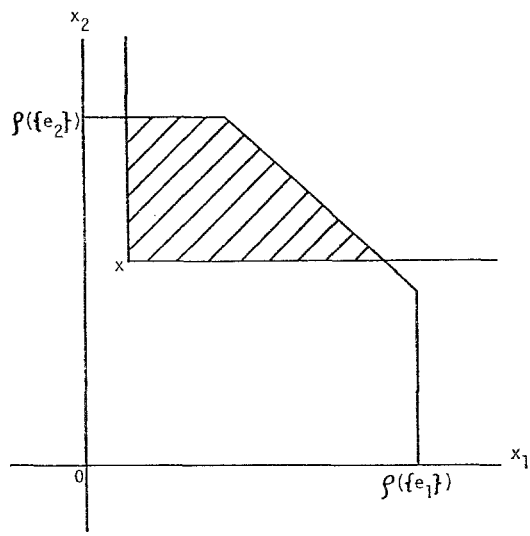

FIG. 3. A contraction: the shaded area represents the set of all independent vectors of the contraction of (E, ρ) with respect to $\mathbf{x}$, when $E = \{e_1, e_2\}$.

For any independent vector $\mathbf{x}$ of (E, ρ), the *contraction of* (E, ρ) *with respect to* $\mathbf{x}$ is a polymatroid $(E, \rho^{\mathbf{x}})$ with the ground-set rank function $\rho^{\mathbf{x}}$ defined by

$$\rho^{\mathbf{x}}(A) = \min_{D \subseteq E-A} \{\rho(A \cup D) - x(A \cup D)\} \qquad (A \subseteq E). \tag{2.13}$$

The set $P^{\mathbf{x}}$ of all independent vectors of $(E, \rho^{\mathbf{x}})$ is given by

$$P^{\mathbf{x}} = \{\mathbf{y} \mid \mathbf{y} \in R_+^E; \mathbf{y} + \mathbf{x} \text{ is an independent vector of } (E, \rho)\}.$$

(See Fig. 3.)

In particular, when $\mathbf{x}$ is a vector such that

$$x(e) = 0 \qquad (e \in E - S), \tag{2.14}_1$$

$$x(e) \geqslant \rho(\{e\}) \qquad (e \in S) \tag{2.14}_2$$

for some $S \subseteq E$, then the reduction of $\mathbf{P} = (E, \rho)$ with respect to $\mathbf{x}$ is denoted by $\mathbf{P} \cdot S$, which is independent of the values of $x(e)$ $(e \in S \subseteq E)$ as far as inequalities $(2.14)_2$ hold. The polymatroid $\mathbf{P} \cdot S$ is called the *reduction of* $\mathbf{P}$ *onto* S. The rank function ρ_S of $\mathbf{P} \cdot S$ is given by

$$\rho_S(A) = \rho(A \cap S) \qquad (A \subseteq E). \tag{2.15}$$

Note that by definition the ground set of $\mathbf{P} \cdot S$ is not S but E. Furthermore, when $\mathbf{y}$ is a base of the reduction $\mathbf{P} \cdot (E - S)$ of $\mathbf{P} = (E, \rho)$ onto $E - S$, then the contraction of $\mathbf{P}$ with respect to $\mathbf{y}$ is denoted by $\mathbf{P} \times S$, which is independent of the choice of a base $\mathbf{y}$ of $\mathbf{P} \cdot (E - S)$. The polymatroid $\mathbf{P} \times S$ is called the *contraction of* $\mathbf{P}$ *onto* S. The rank function ρ^S of $\mathbf{P} \times S$ is given by

$$\rho^S(A) = \rho(A \cup (E - S)) - \rho(E - S) \qquad (A \subseteq E). \tag{2.16}$$

If the function ρ of polymatroid (E, ρ) is expressed as the sum

$$\rho = \rho_S + \rho_{S-E}$$

for some nonempty subset S of E such that $S \neq E$, then polymatroid (E, ρ) is called *separable*.

Finally, it should be noted that conditions (2.1), (2.2), and (2.3) are equivalent to the apparently weaker ones which are given by (2.1), (2.2), and (2.17):

$$\rho(A \cup \{a\}) + \rho(A \cup \{b\}) \geqslant \rho(A \cup \{a, b\}) + \rho(A) \qquad (A \subseteq E; a, b \in E - A; a \neq b). \tag{2.17}$$

Inequalities (2.17) express the "local" submodularity of ρ. (For matroids and/or polymatroids, also see Welsh, 1976; Edmonds, 1970; and McDiarmid, 1975.)

3. Properties of the Entropy Function

Let X_k $(k = 1,..., n)$ be random variables taking on values of $1,..., N_k$ $(k = 1,..., n)$, respectively, where n and N_k are positive integers. Denote by E the set $\{X_1,..., X_n\}$ of all the random variables. For an arbitrary nonempty subset $A = \{X_{a_1},..., X_{a_l}\}$ of E, the entropy of A in Shannon's sense is given by

$$h(A) = -\sum_{i_1=1}^{N_{a_1}} \cdots \sum_{i_l=1}^{N_{a_l}} P(X_{a_1} = i_1,..., X_{a_l} = i_l) \log P(X_{a_1} = i_1,..., X_{a_l} = i_l), \quad (3.1)$$

where the logarithms are to the base 2 and $P(X_{a_1} = i_1,..., X_{a_l} = i_l)$ denotes the probability that the random variables X_{a_j} $(j = 1,..., l)$ take the values i_j, respectively. For the empty set $\varnothing \subseteq E$, we define its entropy as

$$h(\varnothing) = 0. \quad (3.2)$$

The function h $(:2^E \to R_+)$ is called the *entropy function on* E.

We can easily show that

$$h(A) \leqslant h(B) \qquad (A \subseteq B \subseteq E). \quad (3.3)$$

Also, it is well known that the (conditional) mutual information in Shannon's sense is nonnegative, from which follows (2.17) with ρ replaced by h. Therefore, as was noted in the last paragraph of the preceding section, the pair (E, h) is a polymatroid. We can also directly show the submodularity

$$h(A) + h(B) \geqslant h(A \cup B) + h(A \cap B) \qquad (A, B \subseteq E). \quad (3.4)$$

It should be also noted that when random variables $X_1,..., X_n$ are stochastically independent, we have

$$h(A) = \sum_{e \in A} h(\{e\}) \qquad (A \subseteq E). \quad (3.5)$$

A set function h $(:2^E \to R_+)$ satisfying (3.5) is called a modular function, since such h satisfies (3.4) with equality.

The entropy function is thus intimately related to a polymatroid. To the author's knowledge, the relationship between entropy functions and polymatroids is not fully understood by researchers in the field of information theory, though the "local" submodularity (2.17) has been frequently employed in the argument of information-theoretic problems. To consider entropy functions as rank functions of polymatroids is extremely useful for studying information-theoretic problems where a fundamental role is played by entropy functions or (conditional) mutual informations, as will be seen in the following section.

4. Information-Theoretic Problems Viewed from Polymatroids

In this section we shall consider some information-theoretic problems for which a fundamental role is played by the polymatroidal structure induced by the entropy function on a set of random variables.

4.1. *Information-Theoretic Correlative Analysis*

Let E be a set of n random variables $X_1,\ldots, X_n$ and h $(: 2^E \to R_+)$ be the entropy function on E.

An information-theoretic measure of "correlation" among random variables $X_1,\ldots, X_n$ was proposed by Watanabe (1960):

$$S(E) = \sum_{k=1}^{n} h(\{X_k\}) - h(E). \tag{4.1}$$

The $S(E)$ is called the total correlation of E. Based on the lattice-theoretic duality, Han (1975) proposed the dual total correlation $D(E)$:

$$D(E) = \sum_{k=1}^{n} h(E - \{X_k\}) - (n-1)\, h(E). \tag{4.2}$$

Noting that the entropy function h is a β-function (see (2.1), (2.2), and (2.3)), we can easily show that $S(E)$ and $D(E)$ are nonnegative and vanish when $X_1,\ldots, X_n$ are stochastically independent. Here, it may be noted that the total correlation $S(E)$ is the nullity of E with respect to the polymatroid (E, h) and the dominating vector $(h(\{e\}))_{e\in E}$.

Note that for $S(E)$ of (4.1) the family of sets $\{X_k\}$ $(k = 1,\ldots, n)$ of cardinality 1 covers each element of E once and that for $D(E)$ of (4.2) the family of sets $E - \{X_k\}$ $(k = 1,\ldots, n)$ of cardinality $n-1$ covers each element of E $n-1$ times. The last term of (4.1) (resp. (4.2)) can be considered as a normalizing quantity which ensures that $S(E)$ (resp. $D(E)$) is nonnegative and vanishes when $X_1,\ldots, X_n$ are stochastically independent.

Now, for each $i = 1,\ldots, n-1$ we can propose a "total correlation" of the following type.

$$F_i(E) = \sum_{A\in\mathscr{E}_i} h(A) - \binom{n-1}{i-1} h(E) \qquad (i = 1,\ldots, n-1), \tag{4.3}$$

where

$$\mathscr{E}_i = \{B \mid B \subseteq E, \mid B \mid = i\}. \tag{4.4}$$

Here, $F_1(E) = S(E)$ and $F_{n-1}(E) = D(E)$. Note that $\mathscr{E}_i$ is the family of all subsets of E of cardinality i which covers each element of E $\binom{n-1}{i-1}$ times. By repeatedly applying submodularity inequality (3.4) to the first term of (4.3),

we can easily show that $F_i(E)$ $(i = 1,..., n)$ of (4.3) are nonnegative and vanish when $X_1,..., X_n$ are stochastically independent.

For a dominating vector $\mathbf{v}$ of polymatroid (E, h), let (E, h^*) be the $\mathbf{v}$-dual polymatroid of (E, h), i.e., from (2.9)

$$h^*(A) = v(A) + h(E - A) - h(E) \qquad (A \subseteq E). \tag{4.5}$$

Using h^* in place of h, $F_i(E)$ of (4.3) becomes

$$\begin{aligned} F_i^*(E) &\equiv \sum_{A\in\mathscr{C}_i} h^*(A) - \binom{n-1}{i-1} h^*(E) \\ &= \sum_{A\in\mathscr{C}_i} h(E-A) - \binom{n-1}{i} h(E) \\ &= \sum_{A\in\mathscr{C}_{n-i}} h(A) - \binom{n-1}{(n-i)-1} h(E) \\ &= F_{n-i}(E). \end{aligned} \tag{4.6}$$

Therefore, for each $i = 1,..., n-1$, $F_i(E)$ and $F_{n-i}(E)$ are dual to each other by the replacement of h by h^* of (4.5). In particular, Watanabe's total correlation $S(E)$ and Han's dual total correlation $D(E)$ are dual to each other in the sense of polymatroids as well. Furthermore, since

$$\begin{aligned} &(n-i)\left\{\sum_{A\in\mathscr{C}_i} h(A) - \binom{n-1}{i-1} h(E)\right\} - i\left\{\sum_{A\in\mathscr{C}_{i+1}} h(A) - \binom{n-1}{i} h(E)\right\} \\ &\quad = \sum_{A\in\mathscr{C}_{i+1}} \left\{\sum_{\substack{B\subseteq A \\ B\in\mathscr{C}_i}} h(B) - ih(A)\right\} \geqslant 0, \end{aligned}$$

we get

$$(n-i)F_i(E) \geqslant iF_{i+1}(E) \qquad (i = 1,..., n-2). \tag{4.7}$$

Since inequalities (4.7) hold for any β-function instead of h, replacing h in (4.7) by h^* of (4.5) yields

$$(n-i)F_i^*(E) \geqslant iF_{i+1}^*(E) \qquad (i = 1,..., n-2),$$

that is,

$$(n-i)F_{n-i}(E) \geqslant iF_{n-i-1}(E) \qquad (i = 1,..., n-2). \tag{4.8}$$

It follows from (4.7) and (4.8) that if one of $F_i(E)$ $(i = 1,..., n-1)$ vanishes, then all $F_i(E)$ $(I = 1,..., n-1)$ vanish. For each $i = 2,..., n-2$, $F_i(E)$ seems to be a fundamental index in the correlative analysis and a good candidate for a

"total correlation" of random variables $X_1, \ldots, X_n$ as well as $F_1(E) = S(E)$ and $F_{n-1}(E) = D(E)$.

Finally, it should be noted that the submodularity (3.4) of the entropy function h plays a crucial role in the argument of Han (1975, 1978), where use is, however, made only of the "local" submodularity (2.17), i.e., the nonnegativity of Shannon's (conditional) mutual information. Han (1978) provides a theorem which completely determines the class of all (symmetric) linear inequalities among entropy functions derivable from the polymatroidal properties alone.

4.2. *Data Compression Theorem of Slepian, Wolf and Cover*

Let $(X_1, \ldots, X_n)$ be an n-tuple of correlated random variables and $\{(X_1^{(i)}, \ldots, X_n^{(i)})\}_{i=1}^{\infty}$ be a sequence of independent random vectors $(X_1^{(i)}, \ldots, X_n^{(i)})$ $(i = 1, 2, \ldots)$, where for each $i = 1, 2, \ldots$ the probability distribution for $(X_1^{(i)}, \ldots, X_n^{(i)})$ is exactly the same as that for $(X_1, \ldots, X_n)$. The random variables X_k $(k = 1, \ldots, n)$ are supposed to take on values in the at most countable set.

Suppose that there are n encoders to which n information sequences $\{X_k^{(i)}\}_{i=1}^{\infty}$ $(k = 1, \ldots, n)$ are sent separately and that a single decoder has available all the encoded measages. Then we have the following theorem due to Slepian and Wolf (1973) and Cover (1975).

THEOREM. *The n information sequences $\{X_k^{(i)}\}_{i=1}^{\infty}$ $(k = 1, \ldots, n)$ can be sent separately at rates $r(X_k)$ $(k = 1, \ldots, n)$, respectively, to a common receiver with arbitrary small probability of error events if and only if*

$$\sum_{e \in A} r(e) > h(E) - h(E - A) \qquad (A \subseteq E), \tag{4.9}$$

where $E = \{X_1, \ldots, X_n\}$ and h is the entropy function on E.

Here, it should be noted that X_k of $r(X_k)$ is merely an index so that $r(X_k)$ is not a random variable. Rates $r(e)$ $(e \in E)$ are called *admissible* if they satisfy

$$r(A) \equiv \sum_{e \in A} r(e) \geqslant h(E) - h(E - A) \qquad (A \subseteq E). \tag{4.10}$$

As is seen from (4.10) and (4.5), admissible rates $r(e)$ $(e \in E)$ are closely related to a dual polymatroid of (E, h). Suppose that because of physical constraints rates $r(e)$ $(e \in E)$ are bounded by $v(e)$, respectively, and that $(v(e))_{e \in E}$ is a dominating vector of (E, h). Then by setting

$$r^*(e) = v(e) - r(e) \qquad (e \in E), \tag{4.11}$$

$$h^*(A) = v(A) + h(E - A) - h(E) \qquad (A \subseteq E), \tag{4.12}$$

we get from (4.10)

$$r^*(A) \leqslant h^*(A) \qquad (A \subseteq E), \tag{4.13}$$

where h^* is the rank function of the **v**-dual polymatroid of (E, h) and the vector $(r^*(e))_{e\in E}$ belongs to R_+^E. Therefore, rates $r(e)$ $(e \in E)$ are admissible if and only if the vector $(r^*(e))_{e\in E}$ defined by (4.11) is an independent vector of the dual polymatroid (E, h^*). It is interesting that an independent vector of a dual polymatroid of (E, h) has a definite physical meaning as above.

Moreover, when admissible rates $r(e)$ $(e \in E)$ which attain the minimum of the linear form

$$\sum_{e\in E} \alpha(e)\, r(e) \tag{4.14}$$

with nonnegative coefficients $\alpha(e)$ $(e \in E)$ are to be found, optimal admissible rates can be obtained efficiently by employing the greedy algorithm (Edmonds, 1970) with an obvious modification. The linear form (4.14) may express the total cost per unit time of encoding at rates $r(e)$ $(e \in E)$.

Similarly as the coding problem of correlated information sources, problems of multiple-user channels are closely related to polymatroids induced by entropy functions defined on input and output random variables (cf. Wyner, 1974 and van der Meulen, 1977).

4.3. *Common Information and Mutual Information*

Let X and Y be discrete random variables taking values in a finite set. The common information (or the common randomness) $C(X, Y)$ defined by Wyner (1975) is given by

$$C(X, Y) = \inf I(\tilde{X}, \tilde{Y}; W), \tag{4.15}$$

where the infimum is taken over all triples of random variables $\tilde{X}$, $\tilde{Y}$ and W such that W takes on values in a finite set and that

(i) the marginal probability distribution for $\tilde{X}$ and $\tilde{Y}$ coincides with that for the original random variables X and Y; (4.16)

(ii) $\tilde{X}$ and $\tilde{Y}$ are conditionally independent given W. (4.17)

Here, $I(\tilde{X}, \tilde{Y}; W)$ is the mutual information between $\{\tilde{X}, \tilde{Y}\}$ and $\{W\}$ in Shannon's sense.

Let h be the entropy function defined on $E \equiv \{\tilde{X}, \tilde{Y}, W\}$. Note that the entropy function h is also a functional of the probability distribution for $\tilde{X}$, $\tilde{Y}$, and W. By the use of h, condition (4.17) is equivalent to

$$h(\{\tilde{X}, \tilde{Y}\} \mid \{W\}) = h(\{\tilde{X}\} \mid \{W\}) + h(\{\tilde{Y}\} \mid \{W\}), \tag{4.18}_1$$

or

$$h(\{\tilde{X}, \tilde{Y}, W\}) - h(\{W\}) = h(\{\tilde{X}, W\}) - h(\{W\}) + h(\{\tilde{Y}, W\}) - h(\{W\}), \tag{4.18}_2$$

where $h(A \mid B)$ is the averaged conditional entropy of A given B in Shannon's sense.

Moreover, let $\bar{h}$ be the rank function of the contraction of polymatroid (E, h) onto $\{\tilde{X}, \tilde{Y}\}$, i.e.,

$$\bar{h}(A) = h(A \cup \{W\}) - h(\{W\}) \qquad (A \subseteq E) \tag{4.19}$$

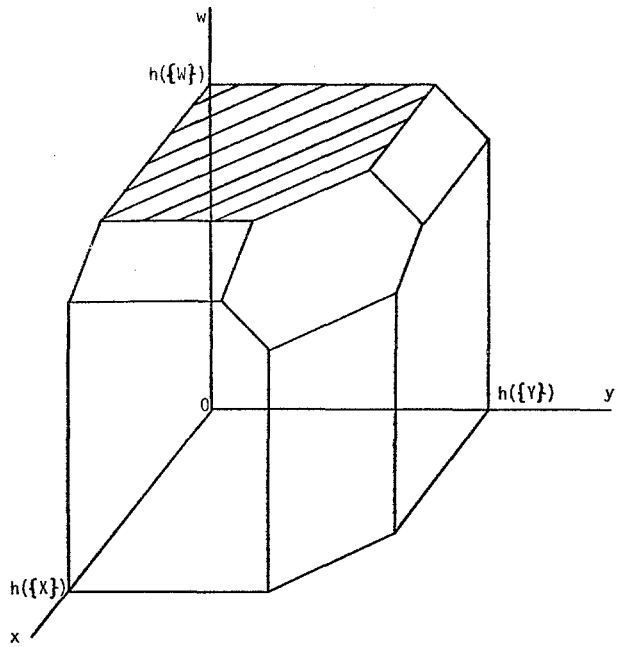

FIG. 4. The convex polyhedron which represents the set of all independent vectors of polymatroid (E, h) with $E = \{X, Y, W\}$.

by (2.16). Figure 4 shows the polyhedron consisting of all independent vectors of polymatroid (E, h). In Fig. 4, polymatroid $(E, \bar{h})$ corresponds to the shaded face lying on the plane: $w = h(\{W\})$. Condition (4.18) means that the shaded face corresponding to the contraction polymatroid $(E, \bar{h})$ should be rectangular, i.e., polymatroid $(E, \bar{h})$ should be separable.

Now, the mutual information $I(\tilde{X}, \tilde{Y}; W)$ in (4.15) is rewritten as

$$I(\tilde{X}, \tilde{Y}; W) = h(\{\tilde{X}, \tilde{Y}\}) + h(\{W\}) - h(\{\tilde{X}, \tilde{Y}, W\}). \tag{4.20}$$

Therefore, by relaxing the constraints with regard to (4.15), let us consider the problem of finding the minimum, with respect to ρ, of

$$\Sigma_\rho(x, y; w) \equiv \rho(\{x, y\}) + \rho(\{w\}) - \rho(\{x, y, w\}) \tag{4.21}$$

under the constraints:

$$\text{(0)}\quad \rho \text{ is a } \beta\text{-function on } \{x, y, w\}; \tag{4.22}$$

$$\text{(1)}\quad \rho(\{x\}) = h(\{X\}),\quad \rho(\{y\}) = h(\{Y\}),\quad \rho(\{x, y\}) = h(\{X, Y\}); \tag{4.23}$$

$$\text{(2)}\quad \rho(\{x, y, w\}) - \rho(\{w\}) = \rho(\{x, w\}) - \rho(\{w\}) + \rho(\{y, w\}) - \rho(\{w\}). \tag{4.24}$$

Note that (4.21) corresponds to (4.20), (4.23) to (4.16), and (4.24) to (4.17) (or (4.18)). Noting that the ramark made below (4.19) is valid for ρ satisfying (4.24) and that the polyhedron consisting of all independent vectors of $(\{x, y, w\}, \rho)$ is as shown in Fig. 4, we can easily show that

$$I(X;\ Y) = \min_{\rho} \Sigma_{\rho}(x, y;\ w),$$

where the minimum is taken over all ρ's satisfying (4.22), (4.23), and (4.24). This reveals that the mutual information $I(X;\ Y)$ is equal to the infimum, with respect to the probability distribution for $\tilde{X}$ and $\tilde{Y}$, of the common information $C(\tilde{X}, \tilde{Y})$ of random variables $\tilde{X}$ and $\tilde{Y}$ such that $h(\{\tilde{X}\}) = h(\{X\})$, $h(\{\tilde{Y}\}) = h(\{Y\})$, and $h(\{\tilde{X}, \tilde{Y}\}) = h(\{X, Y\})$, if the β-function ρ which attains the minimum of (4.21) can be realized by a probability distribution for $\tilde{X}$, $\tilde{Y}$, and W as an entropy function*.

5. Principal Partition of a Set of Random Variables

The principal partition of a graph was first considered by Kishi and Kajitani (1967) based on the notion of a maximally distinct pair of trees, which was examined in more detail by Iri (1971). The result with regard to graphs was extended to matroids by Bruno and Weinberg (1971) and more completely to matroids by Tomizawa (1976). The principal partition of a matroid treated by Tomizawa (1976) can be easily extended to that of a polymatroid.

We shall describe the principal partition of a polymatroid in a form different from that of a matroid due to Tomizawa (1976). Let $\mathbf{P} = (E, \rho)$ be a polymatroid and, for each a $(\in R_+)$, define a vector $\mathbf{v}_a$ $(\in R_+^E)$ by

$$v_a(e) = a \qquad (e \in E). \tag{5.1}$$

Also, define vectors $\mathbf{u}_a$ $(a \in R_+)$ as follows:

$$\text{(i)}\quad \text{for each } a\ (\in R_+)\ \mathbf{u}_a \text{ is a base of } \mathbf{v}_a; \tag{5.2}$$

$$\text{(ii)}\quad \mathbf{u}_a \preccurlyeq \mathbf{u}_b \qquad (a \leqslant b). \tag{5.3}$$

Conditions (5.2) and (5.3) uniquely determine $\mathbf{u}_a$ for each a $(\in R_+)$. Since the set

* Through a private communication from Dr. A. D. Wyner, Dr. H. S. Witsenhausen has pointed out that the infimum is in fact attained by an entropy function.

of all independent vectors of $\mathbf{P}$ is bounded, for each e ($\in E$) there exists a non-negative real number $c(e)$ such that

$$u_a(e) = a \qquad \text{for} \quad 0 \leqslant a \leqslant c(e), \tag{5.4}_1$$

$$u_a(e) = c(e) \qquad \text{for} \quad c(e) \leqslant a. \tag{5.4}_2$$

Let the distinct values of $c(e)$ ($e \in E$) be given by c_i ($i = 1,..., p$) such that $c_i < c_{i+1}$ ($i = 1,..., p-1$), where p is a positive integer not more than $|E|$.

Let us define

$$S_i = \{e \mid e \in E,\ c(e) \leqslant c_i\} \qquad (i = 1,..., p). \tag{5.5}$$

Consequently, we have

$$\varnothing \equiv S_0 \subsetneq S_1 \subsetneq \cdots \subsetneq S_p = E. \tag{5.6}$$

It then follows that vector $(c(e))_{e \in E}$ is a base of (E, ρ) such that

$$c(S_i) = \rho(S_i) \qquad (i = 0, 1,..., p). \tag{5.7}$$

For each $i = 1,..., p$, polymatroid $(\mathbf{P} \cdot S_i) \times (S_i - S_{i-1})$ is called an irreducible minor of $\mathbf{P} = (E, \rho)$, following Tomizawa (1976).

Furthermore, for each $i = 0, 1,..., p-1$, let us define a family $\mathscr{S}_i$ of subsets of S_{i+1} by

$$\mathscr{S}_i = \{S \mid S_i \subseteq S \subseteq S_{i+1},\ c(S) = \rho(S)\}. \tag{5.8}$$

From (5.7), S_i and S_{i+1} belong to $\mathscr{S}_i$. Moreover, by the submodularity of ρ and the definition of $c(e)$ ($e \in E$) we can easily show that the family $\mathscr{S}_i$ is closed under the operations of intersection and union and that $\mathscr{S}_i$ forms a distributive lattice. Let

$$S_i = S_i^0 \subsetneq S_i^1 \subsetneq \cdots \subsetneq S_i^{q_i} = S_{i+1} \tag{5.9}$$

be a composition series of the distributive lattice $\mathscr{S}_i$. For each $j = 1,..., q_i$, polymatroid $(\mathbf{P} \cdot S_i^j) \times (S_i^j - S_i^{j-1})$ is called a strongly irreducible minor of $\mathbf{P} = (E, \rho)$. As is well known, polymatroids $(\mathbf{P} \cdot S_i^j) \times (S_i^j - S_i^{j-1})$ $(j = 1,..., q_i)$ do not depend on the choice of a composition series (5.9) and are uniquely determined by $\mathbf{P}$ and $\mathscr{S}_i$.

The partition of $\mathbf{P}$ into the minors

$$(\mathbf{P} \cdot S_i^j) \times (S_i^j - S_i^{j-1}) \qquad (i = 0, 1,..., p;\ j = 1,..., q_i) \tag{5.10}$$

is called the *principal partition of* $\mathbf{P}$ and the partition of E into the sets

$$T_i^j = S_i^j - S_i^{j-1} \qquad (i = 0, 1,..., p;\ j = 1,..., q_i) \tag{5.11}$$

is called the *principal partition of the ground set* E *of* $\mathbf{P}$.

Now, let us consider random variables X_k $(k = 1,..., n)$ and the entropy function h on $E = \{X_1,..., X_n\}$. Suppose that the principal partition of the ground set E of (E, h) is given by (5.11). From (5.4) $\sim$ (5.11), we see that for each $i = 0, 1,... p$ and $j = 1,..., q_i$

$$h(S_i) = \sum_{l=1}^{i} c_l \mid S_l - S_{l-1} \mid, \tag{5.12}_1$$

$$h(S_i^j) = h(S_i) + c_i \mid S_i^j - S_i \mid. \tag{5.12}_2$$

It follows from (5.12) that, if $i_0 < i_1$, then the averaged conditional entropy of $S_{i_1}^j - S_{i_1}$ $(1 \leqslant j \leqslant q_{i_1})$ per variable given S_{i_1} is greater than that of $S_{i_0}^{j'} - S_{i_0}$ $(1 \leqslant j' \leqslant q_{i_0})$ per variable given S_{i_0}, i.e.,

$$c_{i_1} = \{h(S_{i_1}^j) - h(S_{i_1})\}/\mid S_{i_1}^j - S_{i_1} \mid$$

$$> \{h(S_{i_0}^{j'}) - h(S_{i_0})\}/\mid S_{i_0}^{j'} - S_{i_0} \mid = c_{i_0}.$$

Thus the principal partition of a set of random variables and the values of c_i $(i = 1,..., p)$ provide us information on the interdependence structure of the random variables and may be useful for grouping the random variables in a meaningful way. In general, because the entropy function takes on values of real numbers, the ground set E of random variables may be decomposed into $\mid E \mid$ parts, each composed of a single random variable, by the principal partition. In this case, partitioning E based on the distribution of the values of $c(e)$ $(e \in E)$ may be recommended.

EXAMPLE. The present example is taken from Watanabe (1960). Consider random variables X_k $(k = 1, 2, 3, 4)$ which take on values 0 and 1. Suppose that the probability distribution for X_k $(k = 1, 2, 3, 4)$ is given by

$$\begin{aligned} &P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0) \\ &\quad = P(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0) \\ &\quad = P(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 1) \\ &\quad = P(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1) \\ &\quad = \tfrac{1}{4} \end{aligned} \tag{5.13}$$

and the probability of the other events is equal to zero.

Then we have

$$\begin{gathered} h(\{X_k\}) = 1 \qquad (k = 1, 2, 3, 4), \\ h(\{X_1, X_4\}) = h(\{X_2, X_3\}) = 1, \\ h(\{X_1, X_2\}) = h(\{X_3, X_4\}) = h(\{X_1, X_3\}) = h(\{X_2, X_4\}) = 2, \\ h(\{X_1, X_2, X_3, X_4\} - \{X_k\}) = 2 \qquad (k = 1, 2, 3, 4), \\ h(\{X_1, X_2, X_3, X_4\}) = 2. \end{gathered} \tag{5.14}$$

Consequently, we get

$$c_1 = \tfrac{1}{2}, \qquad S_0 = \varnothing, \qquad S_1 = \{X_1, X_2, X_3, X_4\},$$

and

$$\mathscr{S}_0 = \{\varnothing, \{X_1, X_4\}, \{X_2, X_3\}, \{X_1, X_2, X_3, X_4\}\}.$$

Therefore, we have the principal partition of $\{X_1, X_2, X_3, X_4\}$ as

$$\{X_1, X_4\}, \qquad \{X_2, X_3\}. \tag{5.15}$$

We see from (5.13) that the value of X_1 (resp. X_2) is completely determined by the value of X_4 (resp. X_3) and vice versa, which reflects on the principal partition (5.15).

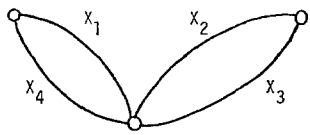

FIG. 5. A graph which represents the entropy function h of (5.14).

It may be noted that the entropy function (5.14) is graphic, i.e., h defined by (5.14) is representable as a rank function of a graph, which is in fact given by the one shown in Fig. 5. We see that Watanabe's total correlation of $\{X_1, X_2, X_3, X_4\}$ coincides with the nullity of the graph. Finally, it should be noted that finding the principal partition of the ground set E of polymatroid (E, h) is not an easy task in general except for the case when all $h(A)$ $(A \subseteq E)$ are integral multiples of some positive real number (such a polymatroid is called an integral polymatroid). An algorithm for finding the principal partition of a matroid is provided by Tomizawa (1976), which is also applicable to an integral polymatroid.

6. CONCLUDING REMARKS

We have shown that the entropy function h on a set E of random variables is a β-function and thus that the pair (E, h) is a polymatroid. The polymatroidal structure induced by the entropy function is fundamental and plays a crucial role in the information-theoretic analysis of a set of random variables such as the information-theoretic correlative analysis, the analysis of multiple-user communication networks, etc. Keeping the polymatroidal structure in mind, we can get much understanding of information-theoretic problems in which entropy functions play a basic role. Furthermore, there is a possibility that the theory of polymatroids and/or matroids will contribute to the Shannon theory of information.

ACKNOWLEDGMENTS

The author is deeply indebted to Professor Masao Iri of the University of Tokyo and Dr. Te Sun Han of the Sagami Institute of Technology for their valuable comments on the present paper. Thanks are also due to the referee for his useful comments.

This work was supported by the Grant in Aid for Scientific Research of the Ministry of Education, Science and Culture of Japan under Grant: Cooperative Research (A) 135017 (1976–1977). The author is also supported by the Sakkokai Foundation.

RECEIVED: July 15, 1977; REVISED: November 18, 1977

REFERENCES

BRUNO, J., AND WEINBERG, L. (1971), The principal minors of a matroid, *Linear Algebra and Appl.* **4**, 17–54.

COVER, T. M. (1975), A proof of the data compression theorem of Slepian and Wolf for ergodic sources, *IEEE Trans. Information Theory* **IT-21**, 226–228.

EDMONDS, J. (1970), Submodular functions, matroids, and certain polyhedra, *in* "Proceedings of the International Conference on Combinatorial Structures and Their Applications," pp. 67–87, Gordon and Breach, New York.

FUJISHIGE, S. (1976), "Algorithms for Solving the Independent-Flow Problems," Papers of the Technical Group on Circuit and System Theory, the Institute of Electronics and Communication Engineers of Japan, CST76-120 (in Japanese); also, *J. Operations Research Soc. Japan* **21**, 189–204 (1978).

FUJISHIGE, S. (1977a), A primal approach to the independent assignment problem, *J. Operations Research Soc. Japan* **20**, 1–15.

FUJISHIGE, S. (1977b), An algorithm for finding a noptimal independent linkage, *J. Operations Research Soc. Japan* **20**, 59–75.

GALLAGER, R. G. (1968), "Information Theory and Reliable Communication," Wiley, New York.

HAN, T. S. (1975), Linear dependence structure of the entropy space, *Inform. Contr.* **29**, 337–368.

HAN, T. S. (1978), Nonnegative entropy measures of multivariate symmetric correlations, *Inform. Contr.* **36**, 133–156.

IRI, M. (1971), Combinatorial canonical form of a matrix with applications to the principal partition of a graph, *Trans. Inst. of Electronics and Communication Engineers of Japan* **54**-A, 30–37 (in Japanese) (English translation: *Electronics and Communications in Japan*).

IRI, M., AND TOMIZAWA, N. (1976), An algorithm for finding an optimal"independent assignment," *J. Operations Research Soc. of Japan* **19**, 32–57.

KISHI, G., AND KAJITANI, Y. (1967), On maximally distinct trees, *in* "Proceedings of the Fifth Annual Allerton Conference on Circuit and System Theory."

MCDIARMID, C. J. H. (1975), Rado's theorem for polymatroids, *Math. Proc. Cambridge Philos. Soc.* **78**, 263–281.

MCGILL, W. J. (1954), Multivariate information transmission, *Psychometrika* **19**, 97–116.

RECSKI, A. (1976), Matroids and independent state variables, *in* "Proceedings of the Second European Conference on Circuit Theory and Design."

SLEPIAN, D., AND WOLF, J. (1973), Noiseless coding of correlated information sources, *IEEE Trans. Information Theory* **IT-19**, 471–480.

TOMIZAWA, N. (1976), Strongly irreducible matroids and principal partitions of a matroid into strongly irreducible minors, *Trans. Inst. of Electronics and Communication Engineers of Japan* **J59**-A, 83–91 (in Japanese) (English translation: *Electronics and Communications in Japan*).

VAN DER MEULEN, E. C. (1977), A survey of multi-way channels in information theory: 1961–1976, *IEEE Trans. Information Theory* **IT-23**, 1–37.

WATANABE, S. (1960), Information theoretical analysis of multivariate correlation, *IBM J.* 66–81.

WELSH, D. J. A. (1976), "Matroid Theory," Academic Press, London.

WYNER, A. D. (1974), Recent results in the Shannon theory, *IEEE Trans. Information Theory* **IT-20**, 2–10.

WYNER, A. D. (1975), The common information of two dependent random variables, *IEEE Trans. Information Theory* **IT-21**, 163–179.