Review

On a Connection between Information and Group Lattices

Hua Li [1] **and Edwin K. P. Chong** [2,*]

[1] FICO, 3661 Valley Centre Drive, San Diego, CA 92130, USA

[2] Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA

* Author to whom correspondence should be addressed; E-Mail: edwin.chong@colostate.edu; Tel.: +1-970-491-7858; Fax: +1-970-491-2249.

Abstract: In this paper we review a particular connection between information theory and group theory. We formalize the notions of information elements and information lattices, first proposed by Shannon. Exploiting this formalization, we expose a comprehensive parallelism between information lattices and subgroup lattices. Qualitatively, isomorphisms between information lattices and subgroup lattices are demonstrated. Quantitatively, a decisive approximation relation between the entropy structures of information lattices and the log-index structures of the corresponding subgroup lattices, first discovered by Chan and Yeung, is highlighted. This approximation, addressing both joint and common entropies, extends the work of Chan and Yeung on joint entropy. A consequence of this approximation result is that any continuous law holds in general for the entropies of information elements if and only if the same law holds in general for the log-indices of subgroups. As an application, by constructing subgroup counterexamples, we find surprisingly that common information, unlike joint information, obeys neither the submodularity nor the supermodularity law. We emphasize that the notion of information elements is conceptually significant—formalizing it helps to reveal the deep connection between information theory and group theory. The parallelism established in this paper admits an appealing group-action explanation and provides useful insights into the intrinsic structure among information elements from a group-theoretic perspective.

Keywords: information element; entropy; information lattice; subgroup lattice; information inequality; information law; common information; joint information; isomorphism

1. Introduction

Information theory was born with the celebrated entropy formula measuring the *amount* of information for the purpose of communication. However, a suitable mathematical model for *information itself* remained elusive over the last sixty years. It is reasonable to assume that information theorists have had certain intuitive conceptions of information, but in this paper we seek a mathematic model for such a conception. In particular, building on Shannon's work [1], we formalize the notion of *information elements* to capture the syntactical essence of *information*, and identify information elements with σ-algebras and *sample-space-partitions*. As we shall see in the following, by providing such a mathematical framework for information and exposing the lattice structure of information elements, the seemingly surprising connection between information theory and group theory, established by Chan and Yeung [2], can be understood in terms of isomorphism relations between information lattices and subgroup lattices. Consequently, a full-fledged and decisive approximation relation between the entropy structure of information lattices, including both joint and common entropies, and the subgroup-index structure of corresponding subgroup lattices becomes evident.

We first motivate a formal definition for the notion of information elements.

1.1. Informationally Equivalent Random Variables

Recall the profound insight offered by Shannon [3] on the essence of communication: "the fundamental problem of communication is that of reproducing at one point exactly or approximately a message selected at another point." Consider the following motivating example. Suppose a message, *in English*, is delivered from person A to person B. Then, the message is translated and delivered *in German* by person B to person C (perhaps because person C does not know English). Assuming the translation is faithful, person C should receive the message that person A intends to convey. Reflecting upon this example, we see that the message (information) assumes two different "representations" over the process of the entire communication—one in English and the other in German, but the message (information) itself remains the same. Similarly, coders (decoders), essential components of communication systems, perform the similar function of "translating" one representation of the same information to another one. This suggests that "information" itself should be defined in a translation invariant way. This "translation-invariant" quality is precisely how we seek to characterize information.

To introduce the formal definition for information elements to capture the essence of information itself, we note that information theory is built within the probabilistic framework, in which one-time information sources are usually modeled by random variables. Therefore, we start in the following with the concept of *informational equivalence* between random variables and develop the formal concept of information elements from first principles.

Recall that, given a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a measurable space $(\mathrm{S}, \mathcal{S})$, a random variable is a measurable function from Ω to S. The set S is usually called the state space of the random variable, and $\mathcal{S}$ is a σ-algebra on S. The set Ω is usually called the *sample space*; $\mathcal{F}$ is a σ-algebra on Ω, usually called the *event space*; and $\mathbf{P}$ denotes a probability measure on the measurable space $(\Omega, \mathcal{F})$.

To illustrate the idea of *informational equivalence*, consider a random variable $X : \Omega \to \mathrm{S}$ and another random variable $X' = f(X)$, where the function $f : \mathrm{S} \to \mathrm{S}'$ is bijective and measurable (f^{-1}

is assumed to be measurable as well). Certainly, the two random variables X and X' are *technically different* for they have different codomains. However, it is intuitively clear that that they are "equivalent" in some sense. In particular, one can infer the exact state of X by observing that of X', and *vice versa*. For this reason, we may say that the two random variables X and X' carry the same piece of information. Note that the σ-algebras induced by X and X' coincide with each other. In fact, two random variables such that the state of one can be inferred from that of the other induce the same σ-algebra. This leads to the following definition for *information equivalence*.

Definition 1. *We say that two random variables X and X' are* informationally equivalent*, denoted $X \cong X'$, if the σ-algebras induced by X and X' coincide.*

It is easy to verify that the "being-informational-equivalent" relation is an equivalence relation. The definition reflects our intuition, as demonstrated in the previous motivating examples, that two random variables carry the same piece information if and only if they induce the same σ-algebra. This motivates the following definition for *information elements* to capture the syntactical essence of information itself.

Definition 2. *An* information element *is an equivalence class of random variables with respect to the "being-informationally-equivalent" relation.*

We call the random variables in the equivalent class of an information element m *representing random variables* of m. Or, we say that a random variable X *represents* m.

We believe that our definition of information elements reflects exactly Shannon's original intention [1]:

> Thus we are led to define the actual information of a stochastic process as that which is common to all stochastic processes which may be obtained from the original by reversible encoding operations.

Intuitive (also informal) discussion on identifying "information" with σ-algebras surfaces often in probability theory, martingale theory, and mathematical finance. In probability theory, see for example [4], the concept of conditional probability is usually introduced with discussion of treating the σ-algebras conditioned on as the "partial information" available to "observers". In martingale theory and mathematical finance, see for example [5,6], *filtrations*—increasing sequences of σ-algebras—are often interpreted as records of the information available over time.

A Few Observations

Proposition 1. *If $X \cong X'$, then $H(X) = H(X')$.*

(Throughout the paper, we use $H(X)$ to denote the entropy of random variable X.)

The converse to Proposition 1 clearly fails—two random variables with a same entropy do not necessarily carry the same information. For example, consider two binary random variables $X, Y : \Omega \to \{0, 1\}$, where $\Omega = \{a, b, c, d\}$ and $\mathbf{P}$ is uniform on Ω. Suppose $X(\omega) = 0$ if $\omega = a, b$ and 1 otherwise, and $Y(\omega) = 0$ if $\omega = a, c$ and 1 otherwise. Clearly, we have $H(X) = H(Y) = 1$, but one can readily agree that X and Y do *not* carry the same information. Therefore, the notion of "informationally-equivalent" is stronger than that of "identically-distributed."

On the other hand, we see that the notion of "informationally-equivalent" is weaker than that of "being-equal."

Proposition 2. *If $X = X'$, then $X \cong X'$.*

The converse to Proposition 2 fails as well, since two informationally equivalent random variable X and X' may have totally different state spaces, so that it does not even make sense to say $X = X'$.

As shown in the following proposition, the notion of "informational equivalence" characterizes a kind of state space invariant "equalness."

Proposition 3. *Two random variables X and Y with state spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, are informationally equivalent if and only if there exists a one-to-one correspondence $f : \mathcal{X} \rightarrow \mathcal{Y}$ such that $Y = f(X)$.*

Remark: Throughout the paper, we fix a probability space unless otherwise stated. For ease of presentation, we confine ourselves in the following to finite discrete random variables. However, most of the definitions and results can be applied to more general settings without significant difficulties.

1.2. *Identifying Information Elements* via σ-algebras and Sample-Space-Partitions

Since the σ-algebras induced by informationally equivalent random variables are the same, we can unambiguously identify information elements with σ-algebras. Moreover, because we deal with finite discrete random variables exclusively in this paper, we can afford to discuss σ-algebras more explicitly as follows.

Recall that a *partition* Π of a set A is a collection $\{\pi_i : i \in [k]\}$ of disjoint subsets of A such that $\cup_{i \in [k]} \pi_i = A$. (Throughout the paper, we use the bracket notation $[k]$ to denote the generic index set $\{1, 2, \cdots, k\}$.) The elements of a partition Π are usually called the *parts* of Π. It is well known that there is a natural one-to-one correspondence between partitions of the sample space and the σ-algebras—any given σ-algebra of a sample space can be generated uniquely, via union operation, from the atomic events of the σ-algebra, while the collection of the atomic events forms a partition of the sample space. For example, for a random variable $X : \Omega \rightarrow \mathcal{X}$, the atomic events of the σ-algebra induced by X are $X^{-1}(\{x\}), x \in \mathcal{X}$. For this reason, from now on, we shall identify an information element by either its σ-algebra or its corresponding sample space partition.

It is well known that the number of distinct partitions of a set of size n is the nth Bell number and that the Stirling number of the second kind $S(n, k)$ counts the number of ways to partition a set of n elements into k nonempty parts. These two numbers, important to the remarkable results obtained by Orlitsky *et al.* in [7], suggest a possibly interesting connection between the notion of information elements discussed in this paper and the "patterns" studied in [7].

1.3. *Shannon's Legacy*

As mentioned before, the notion of *information elements* was originally proposed by Shannon in [1]. In the same paper, Shannon also proposed a partial order for information elements and a lattice structure for collections of information elements. We follow Shannon and call such lattices *information lattices* in the following.

Abstracting the notion of information elements out of their representations—random variables—is a conceptual leap, analogous to the leap from the concrete calculation with matrices to the study of abstract vector spaces. To this end, we formalize both the ideas of information elements and information lattices. By identifying information elements with sample-space-partitions, we are equipped to establish a comprehensive parallelism between information lattices and subgroup lattices. Qualitatively, we demonstrate isomorphisms between information lattices and certain subgroup lattices. On such isomorphisms, quantitatively, we establish an approximation for the entropy structure of information lattices, consisting of joint, common, and many other information elements, using the log-index structures of their counterpart subgroup lattices. Our approximation builds on the construction of joint information elements by Chan and Yeung [2]. Further reinforcing the view of [2], the parallelism identified here reveals an intimate connection between information theory and group theory and suggests that group theory may provide a suitable mathematical language to describe and study laws of information.

The full-fledged parallelism between information lattices and subgroup lattices established in paper is one of our main contributions of this review. With this intrinsic mathematical structure among multiple information elements being uncovered, we anticipate more systematic attacks on certain network information problems, where a better understanding of intricate internal structures among multiple information elements is in urgent need. Indeed, the ideas of information elements and information lattices were originally motivated by network communication problems—in [1]. Shannon wrote:

> The present note outlines a new approach to information theory which is aimed specifically at the analysis of certain communication problems in which there exist a number of sources simultaneously in operation.

and

> Another more general problem is that of a communication system consisting of a large number of transmitting and receiving points with some type of interconnecting network between the various points. The problem here is to formulate the best system design whereby, in some sense, the best overall use of the available facilities is made.

It is interesting to note that current research of information inequalities are mostly motivated by network coding capacity problems.

Certainly, we do not claim that all the ideas in this paper are our own. For example, as we pointed out previously, the notions of information elements and information lattices were proposed as early as the 1950s by Shannon [1]. However, this paper of Shannon's is not well recognized, perhaps owing to the abstruseness of the ideas. Formalizing these ideas and connecting them to current research by re-exposing internal structure relation as in Theorems 2 and 3 is one of the primary goals of this paper. For all other results and ideas that have been previously published, we separate them from those of our own by giving detailed references to their original sources.

Our review focuses on one possible connection between group theory and information theory: the parallelism between information lattices and subgroup lattices. Over the years, other connections have been made between group theory and information theory, which are beyond the scope of our focus.

Johnson and Suhov [8] have studied the behavior of the entropy of convolutions of independent random variables on compact groups, providing an explicit exponential bound on the rate of convergence of entropy to its maximum. Harremoës [9] presented a simplified proof of convergence. Chirikjian [10,11] has shown how classical inequalities used in information theory, such as those of de Bruijn, Fisher, Cramér, Rao, and Kullback, carry over in a natural way from Euclidean space to unimodular Lie groups. That connection is motivated by problems relating to information gathering in mobile robotics, satellite attitude control, tomographic image reconstruction, biomolecular structure determination, and quantum information theory. Indeed, Chirikjian [11] points out the following in the context of connecting "Shannon's brand of information theory" and Lie groups:

> Despite their relatively long and roughly parallel history, surprisingly few connections appear to have been made between these two vast fields. The only attempts to do so known to the author include those of Johnson and Suhov from an information-theoretic perspective, Willsky from an estimation and controls perspective, and Maksimov and Roy from a probability perspective.

(The references cited above are Johnson and Suhov [8,12], Willsky [13], Maksimov [14], and Roy [15].)

The connections between group theory and information theory identified above are concerned with information theoretic analysis of random objects that take values in groups—the group is the state space of the random variable being studied. In contrast, the connection between group theory and information theory that we draw in this paper is on the group-theoretic nature of information itself (in terms of information elements), quite apart from the state spaces of random objects. We hope to make this point clear in our development.

1.4. Organization

The paper is organized as follows. In Section 2, we introduce a "being-richer-than" partial order between information elements and study the information lattices induced by this partial order. In Section 3, we formally establish isomorphisms between information lattices and subgroup lattices. Section 4 is devoted to the quantitative aspects of information lattices. We show that the entropy structure of information lattices can be approximated by the log-index structure of their corresponding subgroup lattices. As a consequence of this approximation result, in Section 5, we show that any continuous law holds for the entropies of common and joint information if and only if the same law holds for the log-indices of subgroups. As an application of this result, we show a result, which is rather surprising, that unlike joint information neither the submodularity nor the supermodularity law holds for common information in general. We conclude the paper with a discussion in Section 6.

2. Information Lattices

2.1. "Being-Richer-Than" Partial Order

Recall that every information element can be identified with its corresponding sample-space-partition. Consider two sample-space-partitions Π and Π'. We say that Π is *finer than* Π', or Π' is *coarser than* Π, if each part of Π is contained in some part of Π'.

Definition 3. *For two information elements m_1 and m_2, we say that m_1 is richer than m_2, or m_2 is poorer than m_1, if the sample-space-partition of m_1 is finer than that of m_2. In this case, we write $m_1 \geq m_2$.*

It is easy to verify that the above defined "being-richer-than" relation is a partial order.

We have the following immediate observations:

Proposition 4. $m_1 \geq m_2$ *if and only if* $H(m_2|m_1) = 0$.

As a corollary to the above proposition, we have

Proposition 5. *If* $m_1 \geq m_2$, *then* $H(m_1) \geq H(m_2)$.

The converse of Proposition 5 does not hold in general.

With respect to representative random variables of information elements, we have

Proposition 6. *Suppose random variables X_1 and X_2 represent information elements m_1 and m_2 respectively. Then, $m_1 \geq m_2$ if and only if $X_2 = f(X_1)$ for some function f.*

A similar result to Proposition 6 was previously observed by Renyi [16] as well.

The "being-richer-than" relation is very important to information theory, because it characterizes the only universal information-theoretic constraint put on all deterministic coders (decoders)—the input information element of any coder is always richer than the output information element. For example, partially via this principle, Yan *et al.* recently characterized the capacity region of general acyclic multi-source multi-sink networks [17]. Harvey *et al.* [18] obtained an improved computable outer bound for general network coding capacity regions by applying this same principle under a different name called *information dominance*—the authors of the paper acknowledged: "...information dominance plays a key role in our investigation of network capacity."

2.2. Information Lattices

Recall that a lattice is a set endowed with a partial order in which any two elements have a unique supremum and a unique infimum with respect to the partial order. Conventionally, the supremum of two lattice elements x and y is also called the *join* of x and y; the infimum is also called the *meet*. In our case, with respect to the "being-richer-than" partial order, the supremum of two information elements m_1 and m_2, denoted $m_1 \vee m_2$, is the poorest among all the information elements that are richer than both m_1 and m_2. Conversely, the infimum of m_1 and m_2, denoted $m_1 \wedge m_2$, is the richest among all the information elements that are poorer than both m_1 and m_2. In the following, we also use m^{12} to denote the join of m_1 and m_2, and m_{12} the meet.

Definition 4. *An* information lattice *is a set of information elements that is closed under the join $\vee$ and meet $\wedge$ operations.*

Recall the one-to-one correspondence between information elements and sample-space-partitions. Consequently, each information lattice corresponds to a partition lattice (with respect to the "being-finer-than" partial order on partitions), and *vice versa*. This formally confirms the assertions made in [1]: "they (information lattices) are at least as general as the class of finite partition lattices."

Since the collection of information lattices could be as general as that of partition lattices, we should not expect any special lattice properties to hold generally for all information lattices, because it is well-known that any finite lattice can be embedded in a finite partition lattice [19]. Therefore, it is not surprising to learn that information lattices are in general not distributive, not even modular.

2.3. Joint Information Element

The *join* of two information elements is straightforward. Consider two information elements m_1 and m_2 represented respectively by two random variables X_1 and X_2. It is easy to check that the joint random variable (X_1, X_2) represents the join m^{12}. For this reason, we also call m^{12} (or $m_1 \vee m_2$) the *joint information element* of m_1 and m_2. It is worth pointing out that the joint random variable (X_2, X_1) represents m^{12} equally well.

2.4. Common Information Element

In [1], the meet of two information elements is called *common information*. More than twenties years later, the same notion of common information was independently proposed and first studied in detail by Gács and Körner [20]. For the first time, it was demonstrated that common information could be far less than mutual information. ("Mutual information" is rather a misnomer because it does not correspond naturally to any information element [20].) Unlike the case of joint information elements, characterizing common information element via their representing random variables is much more complicated. See [20,21] for details.

In contrast to the all-familiar joint information, common information receives far less attention. Nonetheless, it has been shown to be important to cryptography [22–25], indispensable for characterizing of the capacity region of multi-access channels with correlated sources [26], useful in studying information inequalities [27,28], and relevant to network coding problems [29].

2.5. Previously Studied Lattices in Information Theory

Historically, at least three other lattices [30–32] have been considered in attempts to characterize certain ordering relations between information elements. Two of them, studied respectively in [30] and [32], are subsumed by the information lattices considered in this paper.

3. Isomorphisms between Information Lattices and Subgroup Lattices

In this section, we discuss the qualitative aspects of the parallelism between information lattices generated from sets of information elements and subgroup lattices generated from sets of subgroups. In particularly, we establish isomorphism relations between them.

### 3.1. Information Lattices Generated by Information Element Sets

It is easy to verify that both the binary operations "$\vee$" and "$\wedge$" are *associative* and *commutative*. Thus, we can readily extend them to cases of more than two information elements. Accordingly, for a

given set $\{m_i : i \in [n]\}$ of information elements, we denote the joint information element of the subset $\{m_i : i \in \alpha\}$, $\alpha \subseteq [n]$, of information elements by m^α and the common information element by m_α.

Definition 5. *Given a set* $\mathbf{M} = \{m_i : i \in [n]\}$ *of information elements, the* information lattice generated by $\mathbf{M}$, *denoted* $L_{\mathbf{M}}$, *is the smallest information lattice that contains* $\mathbf{M}$. *We call* $\mathbf{M}$ *the* generating set *of the lattice* $L_{\mathbf{M}}$.

It is easy to see that each information element in $L_{\mathbf{M}}$ can be obtained from the information elements in the generating set $\mathbf{M}$ via a sequence of join and meet operations. Note that the set $\{m_\alpha : \alpha \subseteq [n]\}$ of information elements forms a meet semi-lattice and the set $\{m^\beta : \beta \subseteq [n]\}$ forms a join semi-lattice. However, the union $\{m_\alpha, m^\beta : \alpha, \beta \subseteq [n]\}$ of these two semi-lattices does *not* necessarily form a lattice. To see this, consider the following example constructed with partitions (since partitions are in one-to-one correspondence with information elements). Let $\{\pi_i : i = [4]\}$ be a collection of partitions on the set $\{1, 2, 3, 4\}$ where $\pi_1 = 12|3|4$, $\pi_2 = 14|2|3$, $\pi_3 = 23|1|4$, and $\pi_4 = 34|1|2$. See Figure 1 for the Hasse diagram of the lattice generated by the collection $\{\pi_i : i = [4]\}$. It is easy to see that $(\pi_1 \wedge \pi_2) \vee (\pi_3 \wedge \pi_4) = 124|3 \vee 234|1 = 24|1|3$, but $24|1|3 \notin \{\pi_\alpha, \pi^\beta : \alpha, \beta \in [4]\}$. Similarly, we have $(\pi_1 \wedge \pi_3) \vee (\pi_2 \wedge \pi_4) = 13|2|4 \notin \{\pi_\alpha, \pi^\beta : \alpha, \beta \in [4]\}$.

Figure 1. Lattice Generated by $\{\pi_i : i = [4]\}$.
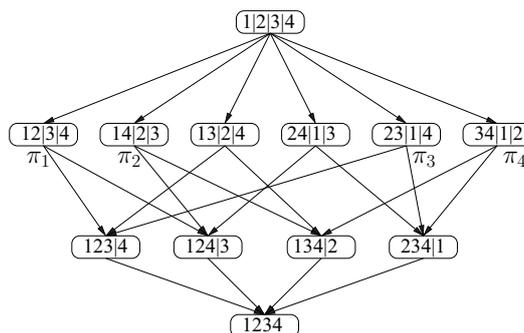
3.2. Subgroup Lattices

Consider the binary operations on subgroups—intersection and union. We know that the intersection $G_1 \cap G_2$ of two subgroups is again a subgroup. However, the union $G_1 \cup G_2$ does *not* necessarily form a subgroup. Therefore, we consider the subgroup *generated* from the union $G_1 \cup G_2$, denoted G^{12} (or $G_1 \vee G_2$). Similar to the case of information elements, the intersection and "$\vee$" operations on subgroups are both associative and commutative. Therefore, we readily extend the two operations to the cases with more than two subgroups and, accordingly, denote the intersection $\cap_{i \in [n]} G_i$ of a set of subgroups $\{G_i : i \in [n]\}$ by $G_{[n]}$ and the subgroup generated from the union by $G^{[n]}$. It is easy to verify that the subgroups $G_{[n]}$ and $G^{[n]}$ are the infimum and the supremum of the set $\{G_i : i \in [n]\}$ with respect to the "being-a-subgroup-of" partial order. For notation consistency, we also use "$\wedge$" to denote the intersection operation.

Note that, to keep the notation simple, we "overload" the symbols "$\vee$" and "$\wedge$" for both the join and the meet operations with information elements and the intersection and the "union-generating" operations with subgroups. Their actual meaning should be clear within context.

Definition 6. *A subgroup lattice is a set of subgroups that is closed under the $\wedge$ and $\vee$ operations.*

For example, the set of all the subgroups of a group forms a lattice.

Similar to the case of information lattices generated by sets of information elements, we consider in the following *subgroup lattices generated by a set of subgroups*.

Definition 7. *Given a set* $\mathbf{G} = \{G_i : i \in [n]\}$ *of subgroups, the* subgroup lattice generated by $\mathbf{G}$, *denoted* $\mathrm{L}_\mathbf{G}$, *is the smallest lattices that contains* $\mathbf{G}$. *We call* $\mathbf{G}$ *the* generating set *of* $\mathrm{L}_\mathbf{G}$.

Note that the set $\{G_\alpha : \alpha \subseteq [n]\}$ forms a semilattice under the meet $\wedge$ operation and the set $\{G^\beta : \beta \subseteq [n]\}$ forms a semilattice under the join $\vee$ operation. However, as in the case of information lattices, the union $\{G_\alpha, G^\beta : \alpha, \beta \subseteq [n]\}$ of the two semilattices does *not* necessarily form a lattice.

In the remainder of this section, we relate information lattices generated by sets of information elements and subgroup lattices generated by collections of subgroups and demonstrate isomorphism relations between them. For ease of presentation, as a special case we first introduce an isomorphism between information lattices generated by sets of coset-partition information elements and their corresponding subgroup lattices.

3.3. Special Isomorphism Theorem

We endow the sample space with a *group* structure—the sample space in question is taken to be a group G. For any subgroup of G, by Lagrange's theorem [33], the collection of its cosets forms a partition of G. Certainly, the coset-partition, as a sample-space-partition, uniquely defines an information element. A collection $\mathbf{G} = \{G_i : i \in [n]\}$ of subgroups of G, in the same spirit, identifies a set $\mathbf{M} = \{m_i : i \in [n]\}$ of information elements via this subgroup–coset-partition correspondence.

Remark: throughout the paper, groups are taken to be multiplicative, and cosets are taken to be right cosets.

It is clear that, by our construction, the information elements in $\mathbf{M}$ and the subgroups in $\mathbf{G}$ are in one-to-one correspondence via the subgroup–coset-partition relation. It turns out that the information elements on the entire information lattice $\mathrm{L}_\mathbf{M}$ and the subgroups on the subgroup lattice $\mathrm{L}_\mathbf{G}$ are in one-to-one correspondence as well via the same subgroup–coset-partition relation. In other words, both the join and meet operations on information lattices are faithfully "mirrored" by the join and meet operations on subgroup lattices.

Theorem 1. *(Special Isomorphism Theorem) Given a set* $\mathbf{G} = \{G_i : i \in [n]\}$ *of subgroups, the subgroup lattice* $\mathrm{L}_\mathbf{G}$ *is isomorphic to the information lattice* $\mathrm{L}_\mathbf{M}$ *generated by the set* $\mathbf{M} = \{m_i : i \in [n]\}$ *of information elements, where* m_i, $i \in [n]$, *are accordingly identified via the coset-partitions of the subgroups* G_i, $i \in [n]$.

The theorem is shown by demonstrating a mapping, from the subgroup lattice $\mathrm{L}_\mathbf{G}$ to the information lattice $\mathrm{L}_\mathbf{M}$, such that it is a lattice-morphism, *i.e.*, it honors both join and meet operations, and is bijective as well. Naturally, the mapping $\phi : \mathrm{L}_\mathbf{G} \to \mathrm{L}_\mathbf{M}$ assigning to each subgroup $G_i \in \mathrm{L}_\mathbf{G}$ the information element identified by the coset-partition of the subgroup G_i is such a morphism. Since this theorem and its general version, Theorem 2, are crucial to our later results—Theorems 3 and 5—and certain aspects of the reasoning are novel, we include a detailed proof for it in Appendix A.

3.4. General Isomorphism Theorem

The information lattices considered in Section 3.3 is rather limited—by Lagrange's theorem, coset-partitions are all equal partitions. In this subsection, we consider arbitrary information lattices—we do not require the sample space to be a group. Instead, we treat a general sample-space-partition as an *orbit-partition* resulting from some group-action on the sample space.

3.4.1. Group-Actions and Permutation Groups

Definition 8. *Given a group G and a set A, a group-action of G on A is a function $(g, a) \mapsto g(a)$, $g \in G$, $a \in A$, that satisfies the following two conditions:*

- $(g_1 g_2)(a) = \big(g_1(g_2(a))\big)$ *for all $g_1, g_2 \in G$ and $a \in A$;*

- $e(a) = a$ *for all $a \in A$, where e is the identity of G.*

We write (G, A) to denote the group-action.

Now, we turn to the notions of *orbits* and *orbit-partitions*. We shall see that every group-action (G, A) induces unambiguously an equivalence relation as follows. We say that x_1 and x_2 are *connected* under a group-action (G, A) if there exists a $g \in G$ such that $x_2 = g(x_1)$. We write $x_1 \overset{\mathrm{G}}{\sim} x_2$. It is easy to check that this "being-connected" relation $\overset{\mathrm{G}}{\sim}$ is an equivalence relation on A. By the fundamental theorem of equivalence relations, it defines a partition on A.

Definition 9. *Given a group-action (G, A), we call the equivalence classes with respect to the equivalence relation $\overset{\mathrm{G}}{\sim}$, or the parts of the induced partition of A, the* orbits *of the group-action. Accordingly, we call the induced partition the* orbit-partition *of (G, A).*

3.4.2. Sample-Space-Partition as Orbit-Partition

In fact, starting with a partition Π of a set A, we can go in the other direction and unambiguously define a group action (G, A) such that the orbit-partition of (G, A) is exactly the given partition Π. To see this, note the following salient feature of group-actions: For any given group-action (G, A), associated with every element g in the group is a mapping from A to itself and any such mappings must be bijective. This feature is the direct consequence of the group axioms. To see this, note that every group element g has a unique inverse g^{-1}. According to the first defining property of group-actions, we have $(gg^{-1})(x) = g\left(g^{-1}(x)\right) = e(x) = x$ for all $x \in A$. This requires that the mappings associated with g and g^{-1} to be invertible. Clearly, the identity e of the group corresponds to the identity map from A to A.

With the observation that under group-action (G, A) every group element corresponds to a permutation of A, we can treat every group as a collection of permutations that is closed under permutation composition. Specifically, for a given partition Π of a set A, it is easy to check that all the permutations of A that permute the elements of the parts of Π *only* to the elements of the same parts form a group. These permutations altogether form the so-called *permutation representation* of G (with respect to A). For this reason in the following, without loss of generality, we treat all groups

as permutation groups. We denote by G_Π the permutation group corresponding as above to a partition Π—G_Π acts naturally on the set A by permutation, and the orbit partition of (G_Π, A) is exactly Π.

From group theory, we know that this orbit-partition–permutation-group-action relation is a one-to-one correspondence. Since every information element corresponds definitively to a sample-space-partition, we can identify every information element by a permutation group. Given a set $\mathbf{M} = \{m_i : i \in [n]\}$ of information elements, denote the set of the corresponding permutation groups by $\mathbf{G} = \{G_i : i \in [n]\}$. Note that all the permutations in the permutation groups G_i, $i \in [n]$, are permutations of the same set, namely the sample space. Hence, all the permutation groups G_i, $i \in [n]$, are subgroups of the symmetric group $S_{|\Omega|}$, which has order $|\Omega|!$. Therefore, it makes sense to take intersection and union of groups from the collection $\mathbf{G}$.

3.4.3. From Coset-Partition to Orbit-Partition—From Equal Partition to General Partition

In fact, the previously studied coset-partitions are a special kind of orbit-partitions. They are orbit-partitions of group-actions defined by the native group multiplication. Specifically, given a subgroup G_1 of G, a group-action (G_1, G) is defined such that $g_1(a) = g_1 \circ a$ for all $g_1 \in G_1$ and $a \in G$, where "$\circ$" denotes the native binary operation of the group G. The orbit-partition of such a group-action is exactly the coset-partition of the subgroup G_1. Therefore, by taking a different kind of group-action—permutation rather than group multiplication—we are freed from the "equal-partition" restriction so that we can correspond arbitrary information elements identified with arbitrary sample-space-partitions to subgroups. It turns out information lattices generated by sets of information elements and subgroup lattices generated by the corresponding sets of permutation groups remain isomorphic to each other. Thus, the isomorphism relation between information lattices and subgroup lattices holds in full generality.

3.4.4. Isomorphism Relation Remains Between Information Lattices and Subgroup Lattices

Similar to Section 3.3, we consider a set $\mathbf{M} = \{m_i, i \in [n]\}$ of information element. Unlike in Section 3.3, the information elements m_i, $i \in [n]$ considered here are arbitrary. As we discussed in the above, with each information element m_i we associate a permutation group G_i according to the orbit-partition–permutation-group-action correspondence. Denote the set of corresponding permutation groups by $\mathbf{G} = \{G_i, i \in [n]\}$.

Theorem 2. *(General Isomorphism Theorem) The information lattice* $\mathrm{L}_\mathbf{M}$ *is isomorphic to the subgroup lattice* $\mathrm{L}_\mathbf{G}$.

The arguments for Theorem 2 are similar to those for Theorem 1—we demonstrate that the orbit-partition–permutation-group-action correspondence is a lattice isomorphism between $\mathrm{L}_\mathbf{M}$ and $\mathrm{L}_\mathbf{G}$.

4. An Approximation Theorem

From this section on, we shift our focus to the quantitative aspects of the parallelism between information lattices and subgroup lattices. In the previous section, by generalizing from coset-partitions to orbit-partitions, we successfully established an isomorphism between general information lattices and

subgroup lattices. In this section, we shall see that not only is the qualitative structure preserved, but also the quantitative structure—the entropy structure of information lattices—is essentially captured by their isomorphic subgroup lattices.

4.1. Entropies of Coset-partition Information Elements

We start with a simple and straightforward observation for the entropies of coset-partition information elements on information lattices.

Proposition 7. *Let $\{G_i : i \in [n]\}$ be a set of subgroups of group G and $\{m_i : i \in [n]\}$ be the set of corresponding coset-partition information elements. The entropies of the joint and common information elements on the information lattice, generated from $\{m_i : i \in [n]\}$, can be calculated from the subgroup-lattice, generated from $\{G_i : i \in [n]\}$, as follows*

$$H(m^{[n]}) = \log \frac{|G|}{|\wedge_{i \in [n]} G_i|} \tag{1}$$

and

$$H(m_{[n]}) = \log \frac{|G|}{|\vee_{i \in [n]} G_i|} \tag{2}$$

Proposition 7 follows easily from the isomorphism relation established by Theorem 2.

Note that the right hand sides of both Equations (1) and (2) are the logarithms of the indices of subgroups. In the following, we shall call them, in short, *log-indices*.

Proposition 7 establishes a quantitative relation between the entropies of the information elements on coset-partition information lattices and the log-indices of the subgroups on the isomorphic subgroup lattices. This quantitative relation is *exact*. However, the scope of Proposition 7 is rather restrictive—it applies only to certain special kind of "uniform" information elements, because, by Lagrange's theorem, all coset-partitions are equal partitions.

In Section 3, by generalizing from coset-partitions to orbit-partitions we successfully removed the "uniformness" restriction imposed by the coset-partition structure. At the same time, we established a new isomorphism relation, namely orbit-partition–permutation-group-action correspondence, between information lattices and subgroup lattices. It turns out that this generalization maintains an "rough" version of the quantitative relation established in Proposition 7 between the entropies of information lattices and the log-indices of their isomorphic permutation-subgroup lattices. As we shall see in the next section, the entropies of the information elements on information lattices can be approximated, up to arbitrary precision, by the log-indices of the permutation groups on their isomorphic subgroup lattices.

4.2. Subgroup Approximation Theorem

To discuss the approximation formally, we introduce two definitions as follows.

Definition 10. *Given an information lattice $L_{\mathbf{M}}$ generated from a set $\mathbf{M} = \{m_i, i \in [n]\}$ of information elements, we call the real vector*

$$\big(H(m) : m \in L_{\mathbf{M}} \big)$$

whose components are the entropies of the information elements on the information lattice $L_{\mathbf{M}}$ generated by $\mathbf{M}$, listed according to a certain prescribed order, the entropy vector *of $L_{\mathbf{M}}$, denoted $h(L_{\mathbf{M}})$.*

The entropy vector $h(\mathrm{L_M})$ captures the informational structure among the information elements of **M**.

Definition 11. *Given a subgroup lattice* $\mathrm{L_G}$ *generated from a set* $\mathbf{G} = \{G_i, i \in [n]\}$ *of subgroups of a group* G, *we call the real vector*

$$\frac{1}{|G|} \left(\log \frac{|G|}{|G'|} : G' \in \mathrm{L_G} \right)$$

whose components are the normalized log-indices of the subgroups on the subgroup lattice $\mathrm{L_G}$ *generated by* **G**, *listed according to a certain prescribed order, the* normalized log-index vector of $\mathrm{L_G}$, *denoted* $l(\mathrm{L_G})$.

In the following, we assume that $l(\mathrm{L_G})$ and $h(\mathrm{L_M})$ are accordingly aligned.

Theorem 3. *Let* $\mathbf{M} = \{m_i, i \in [n]\}$ *be a set of information elements. For any* $\epsilon > 0$ *there exists an* $N > 0$ *and a set* $\mathbf{G}^N = \{G_i : i \in [n]\}$ *of subgroups of the symmetry group* S_N *of order* $N!$ *such that*

$$\|h(\mathrm{L_M}) - l(\mathrm{L_{G^N}})\| < \epsilon \tag{3}$$

where "$\|\cdot\|$" *denotes the norm of real vectors.*

Theorem 3 extends the approximation carried out by Chan and Yeung in [2], which is limited to joint entropies. The approximation procedure we carried out to prove Theorem 3 is similar to that of Chan and Yeung [2]—both use Stirling's approximation formula for factorials. But, with the group-action relation between information elements and permutation groups being exposed, and the isomorphism between information lattices and subgroup lattices being revealed, the approximation procedure becomes transparent and the seemingly surprising connection between information theory and group theory becomes mathematically natural. For these reasons, we included a detailed proof in Appendix B.

5. Parallelism between Continuous Laws of Information Elements and those of Subgroups

As a consequence of Theorem 3, we shall see in the following that if a continuous law holds in general for information elements, then the same law must hold for the log-indices of subgroups, and *vice versa*.

In the following, for reference and comparison purposes, we first review the known laws concerning the entropies of joint and common information elements. These laws, usually expressed in the form of *information inequalities*, are deemed to be fundamental to information theory [34].

5.1. Laws for Information Elements

5.5.1. Non-Negativity of Entropy

Proposition 8. *For any information element* m, *we have* $H(m) \geq 0$.

5.1.2. Laws for Joint Information

Proposition 9. *Given a set* $\{m_i, i \in [n]\}$ *of information elements, if* $\alpha \subseteq \beta$, $\alpha, \beta \subseteq [n]$, *then* $H(m^\alpha) \leq H(m^\beta)$.

Proposition 10. *For any two sets of information elements $\{m_i : i \in \alpha\}$ and $\{m_j : j \in \beta\}$, the following inequality holds:*

$$H(m^\alpha) + H(m^\beta) \geq H(m^{\alpha \cup \beta}) + H(m^{\alpha \cap \beta})$$

This proposition is mathematically equivalent to the following one.

Proposition 11. *For any three information elements m_1, m_2, and m_3, the following inequality holds:*

$$H(m^{12}) + H(m^{23}) \geq H(m^{123}) + H(m^2)$$

Note that $H(m^2) = H(m_2)$.

Proposition 10 (or equivalently 11) is usually called the submodularity law for entropy function. Propositions 8, 9, and 10 are known, collectively, as the *polymatroidal axioms* [35,36]. Up until very recently, these are the only known laws for entropies of joint information elements.

In 1998, Zhang and Yeung discovered a new information inequality, involving four information elements [36].

Proposition 12. *(Zhang-Yeung Inequality) For any four information elements m_i, $i = 1, 2, 3$, and 4, the following inequality holds:*

$$3H(m^{13}) + 3H(m^{14}) + H(m^{23}) + H(m^{24}) + 3H(m^{34})$$
$$\geq H(m^1) + 2H(m^3) + 2H(m^4) + H(m^{12}) + 4H(m^{134}) + H(m^{234}) \quad (4)$$

This newly discovered inequality, classified as a *non-Shannon type information inequality* [34], proved that our understanding on laws governing the quantitative relations between information elements is incomplete. Recently, six more new four-variable information inequalities were discovered by Dougherty *et al.* [37].

Information inequalities such as those presented above were called "laws of information" [34,38]. Seeking new information inequalities is currently an active research topic [27,36,39,40]; see Chan's review of recent progresses [41]. In fact, they should be more accurately called "laws of joint information", since these inequalities involves only joint information only. We shall see below laws involving common information. Note that recent work [42,43] have started to explore group theoretic approaches to network coding.

5.1.3. Common Information *v.s.* Mutual Information

In contrast to joint information, little research has been done to laws involving common information. So far, the only known non-trivial law involving both joint information and common information is stated in the following proposition, discovered by Gács and Körner [20].

Proposition 13. *For any two information element m_1 and m_2, the following inequality holds:*

$$H(m_{12}) \leq I(m_1; m_2) = H(m^1) + H(m^2) - H(m^{12})$$

Note that $m^1 = m_1$ and $m^2 = m_2$.

5.1.4. Laws for Common Information

Dual to the non-decreasing property of joint information, it is immediately clear that entropies of common information are non-increasing.

Proposition 14. *Given a set $\{m_i, i \in [n]\}$ of information elements, if $\alpha \subseteq \beta$ $\alpha, \beta \subseteq [n]$, then $H(m_\alpha) \geq H(m_\beta)$.*

Comparing to the case of joint information, one might naturally expect a *supermodularity* law to hold for *common* information, as a dual counterpart to the *submodularity* law of *joint* information. In other words, it is natural to conjecture:

Conjecture 1. *For any three information elements m_1, m_2, and m_3, the following inequality holds:*

$$H(m_{12}) + H(m_{23}) \leq H(m_{123}) + H(m_2) \tag{5}$$

This conjecture is natural because of the intrinsic duality between the join and meet operations of information lattices. Due to the combinatorial nature of common information [20], it is not immediately obvious whether this conjecture holds. With the help of our approximation results established in Theorems 3 and 5, we find that neither the conjecture nor its converse holds. In other words, common information observes neither the submodularity nor the supermodularity law.

5.2. Continuous Laws for Joint and Common Information

As a consequence of Theorem 3, we shall see in the following that if a continuous law holds for information elements, then the same law must hold for the log-indices of subgroups, and *vice versa*. To convey this idea, we first present the simpler case involving only joint and common information elements. To state our result formally, we first introduce two definitions.

Definition 12. *Given a set $\mathbf{M} = \{m_i : i \in [n]\}$ of information elements, consider the collection $\mathcal{M} = \{m_\alpha, m^\beta : \alpha, \beta \subseteq [n]\}$ of join and meet information elements generated from $\mathbf{M}$. We call the real vector*

$$\left(H(m^\alpha), H(m_\beta) : \alpha, \beta \subseteq [n], \alpha, \beta \neq \Phi \right)$$

whose components are the entropies of the information elements of $\mathcal{M}$, the entropy vector of $\mathcal{M}$, *denoted by $h_\mathcal{M}$.*

Definition 13. *Given a set $\mathbf{G} = \{G_i : i \in [n]\}$ of subgroups of a group G, consider the set $\mathcal{G} = \{G_\alpha, G^\beta : \alpha, \beta \subseteq [n]\}$ of the subgroups generated from $\mathbf{G}$. We call the real vector*

$$\frac{1}{|G|} \left(\log \frac{|G|}{|G^\alpha|}, \log \frac{|G|}{|G_\beta|} : \alpha, \beta \subseteq [n], \alpha, \beta \neq \Phi \right)$$

whose components are the normalized log-indices of the subgroups in $\mathcal{M}$, the normalized log-index vector of $\mathcal{G}$, *denoted by $l_\mathcal{G}$.*

In this context, we assume that the components of both $l_\mathcal{G}$ and $h_\mathcal{M}$ are listed according to a common fixed order. Moreover, we note that both the vectors $h_\mathcal{M}$ and $l_\mathcal{G}$ have dimension $2^{n+1} - n - 2$.

Theorem 4. *Let $f : \mathbb{R}^{2^{n+1}-n-2} \to \mathbb{R}$ be a continuous function. Then, $f(h_{\mathcal{M}}) \geq 0$ holds for all sets* **M** *of n information elements if and only if $f(l_{\mathcal{G}}) \geq 0$ holds for all sets* **G** *of n subgroups of any group.*

Theorem 4 is a special case of Theorem 5.

Theorem 4 and its generalization—Theorem 5—extend the result obtained by Chan and Yeung in [2] in the following two ways. First, Theorems 4 and 5 apply to all continuous laws, while only linear laws were considered in [2]. Even though so far we have not yet encountered any nonlinear law for entropies, it is highly plausible that nonlinear information laws may exist given the recent discovery that at least certain part of the boundary of the entropy cones involving at least four information elements are curved [44]. Second, our theorems encompass both common information and joint information, while only joint entropies were considered in [2]. For example, laws such as Propositions 13 and 14 cannot even be expressed in the setting of [2]. In fact, as we shall see later in Section 5.4, the laws of common information depart from those of joint information very early—unlike joint information, which obeys the submodularity law, common information admits neither submodularity nor supermodularity. For these reasons, we believe that our extending the subgroup approximation to common information is of interest in its own right.

5.3. Continuous Laws for General Lattice Information Elements

In this section, we extend Theorem 4 to all the information elements in information lattices, not limited to the "pure" joint and common information elements. In the following, we introduce some necessary machinery to formally present the result in full generality.

Note that an element from the lattice generated from a set X has its expression built from the generating elements of the lattice in the similar way that *terms* are built from *literals* in mathematical logic. In particular, we define *lattice-terms* as follows:

Definition 14. *An expression E is called a* lattice-term *formed from a set* X *of literals if either E is a literal from* X *or E is formed from two lattice-terms with either the* join *or the* meet *symbols: $E = x \circ y$, where x and y are lattice-terms and $\circ$ is either the join symbol $\vee$ or the meet symbol $\wedge$.*

Definition 15. *Suppose that E_i, $i \in [k]$, are lattice-terms generated from a literal set of size n:* $X = \{x_1, \cdots, x_n\}$. *We call an expression of the form*

$$f\big(H(E_1), \cdots, H(E_k)\big)$$

where f represents a function from $\mathbb{R}^k$ to $\mathbb{R}$ and H represents the entropy function, an n-variable generalized information expression.

We *evaluate* an n-variable generalized information expression $f\big(H(E_1), \cdots, H(E_k)\big)$ against a set $M = \{m_i : i \in [n]\}$ of information elements by substituting x_i with m_i respectively, calculating the entropy of the information elements obtained by evaluating the lattice-terms E_i according to the semantics of the join and meet operations on information elements, and then obtaining the corresponding function value. We denote this value by

$$f\big(H(E_1), \cdots, H(E_k)\big)\big|_{\text{M}}$$

Definition 16. *If an n-variable generalized information expression $f\big(H(E_1),\cdots,H(E_k)\big)$ is evaluated non-negatively for any set of n information elements,* i.e.,

$$f\big(H(E_1),\cdots,H(E_k)\big)\big|_{\mathrm{M}} \geq 0,\ \text{for all } \mathrm{M}$$

then we call

$$f\big(H(E_1),\cdots,H(E_k)\big) \geq 0$$

an n-variable information law.

Similar to generalized information expressions, we define *generalized log-index expression* as follows.

Definition 17. *we call an expression of the form*

$$f\big(L(E_1),\cdots,L(E_k)\big)$$

where f represents a function from $\mathbb{R}^k$ to $\mathbb{R}$ and L represents the normalized log-index function of subgroups, an n-variable generalized log-index expression.

We *evaluate* an n-variable generalized log-index expression $f\big(L(E_1),\cdots,L(E_k)\big)$ against a set $\mathbf{G} = \{G_i : i \in [n]\}$ of subgroups of a group G by substituting x_i with G_i respectively, calculating the log-index of the subgroups obtained by evaluating the lattice-terms E_i according to the semantics of the join and meet operations on subgroups, and then obtaining the corresponding function value. We denote this value by

$$f\big(L(E_1),\cdots,L(E_k)\big)\big|_{\mathbf{G}}$$

Definition 18. *If an n-variable generalized log-index expression $f\big(H(E_1),\cdots,H(E_k)\big)$ is evaluated non-negatively for any set of n subgroups of any group,* i.e.,

$$f\big(L(E_1),\cdots,L(E_k)\big)\big|_{\mathbf{G}} \geq 0,\ \text{for all } \mathbf{G}$$

then we call

$$f\big(L(E_1),\cdots,L(E_k)\big) \geq 0$$

an n-variable subgroup log-index law.

With the above formalism and corresponding notations, we are ready to state our equivalence result concerning the generalized information laws.

Theorem 5. *Suppose that f is continuous. Then an n-variable information law*

$$f\big(H(E_1),\cdots,H(E_k)\big) \geq 0$$

holds if and only if the corresponding n-variable subgroup log-index law

$$f\big(L(E_1),\cdots,L(E_k)\big) \geq 0$$

holds.

Proof. To see one direction, namely that $f\big(L(E_1), \cdots, L(E_k)\big) \geq 0$ implies that $f\big(H(E_1), \cdots, H(E_k)\big) \geq 0$, assume that there exists a set **M** of information elements such that $f\big(H(E_1), \cdots, H(E_k)\big)\big|_\mathbf{M} = a$ for some $a < 0$. By the continuity of the function f and Theorem 3, we are guaranteed to be able to construct, from the information lattice generated from **M**, some subgroup lattice $\mathbf{L_G}$ such that the value of the function f at the normalized log-indices of the correspondingly constructed subgroups is arbitrarily close to $a < 0$. This contradicts the assumption that $f\big(L(E_1), \cdots, L(E_k)\big)\big|_\mathbf{G} \geq 0$ holds for all sets **G** of n subgroups of any group.

On the other hand, for any normalized log-indices of the subgroups from subgroup lattices, it can be readily interpreted as the entropies of information elements by taking permutation representation for the subgroups on the subgroup lattice and then producing an information lattice, according to the orbit-partition-permutation-group-action correspondence. Therefore, that $f\big(H(E_1), \cdots, H(E_k)\big)\big|_\mathbf{M} \geq 0$ holds for all sets **M** implies that $f\big(L(E_1), \cdots, L(E_k)\big)\big|_\mathbf{G} \geq 0$ holds for all sets **G**. $\qquad\square$

5.4. *Common Information Observes Neither Submodularity Nor Supermodularity Laws*

As discussed in the above, appealing to the duality between the join and the meet operations, one might conjecture, dual to the well-known submodularity of joint information, that common information would observe the supermodularity law. It turns out that common information observes neither the submodularity (6) nor the supermodularity (7) law—neither of the following two inequalities holds in general:

$$H(m_{12}) + H(m_{23}) \geq H(m_{123}) + H(m_2) \tag{6}$$

$$H(m_{12}) + H(m_{23}) \leq H(m_{123}) + H(m_2) \tag{7}$$

Because common information is combinatorial in flavor—it depends on the "zero pattern" of joint probability matrices [20]—it is hard to directly verify the validity of (6) and (7). However, thanks to Theorem 5, we are able to construct subgroup counterexamples to invalidate (6) and (7) indirectly.

To show that (7) fails, it suffices to find three subgroups G_1, G_2, and G_3 such that

$$|G_1 \vee G_2||G_2 \vee G_3| < |G_1 \vee G_2 \vee G_3||G_2| \tag{8}$$

Consider $G = S_5$, the symmetry group of order $5!$, and its subgroups $G_1 = \langle(12345)\rangle$, $G_2 = \langle(12)(45)\rangle$, and $G_3 = \langle(12543)\rangle$. The subgroup G_1 is the permutation group generated by permutation (12345), G_2 by $(12)(45)$, and G_3 by (12543). (Here, we use the standard cycle notation to represent permutations.) Consequently, we have $G_1 \vee G_2 = \langle(12345), (12)(45)\rangle$, $G_2 \vee G_3 = \langle(12543), (12)(45)\rangle$, and $G_1 \vee G_2 \vee G_3 = \langle(12345), (12)(45), (12543)\rangle$. It is easy to see that both $G_1 \vee G_2$ and $G_2 \vee G_3$ are dihedral groups of order 10 and that $G_1 \vee G_2 \vee G_3$ is the alternative group A_5, hence of order 60. The order of G_2 is 2. Therefore, we see that the subgroups G_1, G_2, and G_3 satisfy (8). By Theorem 5, the supermodularity law (7) does not hold in general for common information.

Similar to the case of supermodularity, the example with $G_2 = \{e\}$ and $G_1 = G_3 = G$, $|G| \neq 1$, invalidates the group version of (6). Therefore, according to Theorem 5, the submodularity law (6) does not hold in general for common information either.

6. Discussion

This paper builds on some of Shannon's little-recognized legacy and adopts his interesting concepts of information elements and information lattices. We formalize all these concepts and clarify the relations between random variables and information elements, information elements and σ-algebras, and, especially, the one-to-one correspondence between information elements and sample-space-partitions. We emphasize that such formalization is conceptually significant. As demonstrated in this paper, beneficial to the formalization carried out, we are able to establish a comprehensive parallelism between information lattices and subgroup lattices. This parallelism is mathematically natural and admits intuitive group-action explanations. It reveals an intimate connection, both structural and quantitative, between information theory and group theory. This suggests that group theory might serve a promising role as a suitable mathematical language in studying deep laws governing information.

Network information theory in general, and capacity problems for network coding specifically, depend crucially on our understanding of intricate structures among multiple information elements. By building a bridge from information theory to group theory, we can now access the set of well-developed tools from group theory. These tools can be brought to bear on certain formidable problems in areas such as network information theory and network coding. Along these lines, by constructing subgroup counterexamples we show that neither the submodularity nor the supermodularity law holds for common information, neither of which is obvious from traditional information theoretic perspectives.

Acknowledgements

References

1. Shannon, C.E. The lattice theory of information. *IEEE Trans. Inform. Theory* **1953**, *1*, 105–107.
2. Chan, T.H.; Yeung, R.W. On a relation between information inequalities and group theory. *IEEE Trans. Inform. Theory* **2002**, *48*, 1992–1995.
3. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.
4. Billingsley, P. *Probability and Measure*, 3rd ed.; Wiley-Interscience: New York, NY, USA, 1995.
5. Shreve, S.E. *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*; Springer: New York, NY, USA, 2005.
6. Ankirchner, S.; Dereich, S.; Imkeller, P. The Shannon information of filtrations and the additional logarithmic utility of insiders. *Ann. Probab.* **2006**, *34*, 743–778.
7. Orlitsky, A.; Santhanam, N.P.; Zhang, J. Universal compression of memoryless sources over unknown alphabets. *IEEE Trans. Inform. Theory* **2004**, *50*, 1469–1481.
8. Johnson, O.; Suhov Y. Entropy and convergence on compact groups. *J. Theor. Probability* **2000**, *13*, 843–857.
9. Harremoës, P. Maximum entropy on compact groups. *Entropy* **2009**, *11*, 222–237.

10. Chirikjian, G.S. *Stochastic Models, Information Theory, and Lie Groups: Volume 1 Classical Results and Geometric Methods*; Birkhauser: Boston, MA, USA, 2009.

11. Chirikjian, G.S. Information-Theoretic Inequalities on Unimodular Lie Groups. *J. Geom. Mech.* **2010**, *2*, 119–158.

12. Johnson, O. *Information Theory and the Central Limit Theorem*; Imperial College Press: London, UK, 2004.

13. Willsky, A.S. Dynamical Systems Defined on Groups: Structural Properties and Estimation. Ph.D. dissertation, Dept. Aeronautics and Astronautics, MIT; Cambridge, MA, USA, 1973.

14. Maksimov, V.M. Necessary and sufficient statistics for the family of shifts of probability distributions on continuous bicompact groups. *Theor. Probab. Appl.* **1967**, *12*, 267–280.

15. Roy, K.K. Exponential families of densities on an analytic group and sufficient statistics. *Sankhy A* **1975**, *37*, 82–92.

16. Renyi, A. *Foundations of Probability*; Holden-Day Inc.: San Francisco, CA, USA, 1970.

17. Yan, X.; Yeung, R.W.; Zhang, Z. The Capacity region for multi-source multi-sink network coding. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; pp. 116–120.

18. Harvey, N.J.A.; Kleinberg, R.; Lehman, A.R. On the capacity of information networks. *IEEE Trans. Inform. Theory* **2006**, *52*, 2345–2364.

19. Pudlák, P.; Tůma, J. Every finite lattice can be embedded in a finite partition lattice. *Algebra Univ.* **1980**, *10*, 74–95.

20. Gács, P.; Körner, J. Common information is far less than mutual information. *Probl. Control Inform. Theory* **1973**, *2*, 149–162.

21. Witsenhausen, H.S. On sequences of pairs of dependent random variables. *SIAM J. Appl. Math.* **1975**, *28*, 100–113.

22. Ahlswede, R.; Csiszàr, I. Common randomness in information theory and cryptography—Part I: Secret sharing. *IEEE Trans. Inform. Theory* **1993**, *39*, 1121–1132.

23. Ahlswede, R.; Csiszàr, I. Common randomness in information cryptography—Part II: CR capacity. *IEEE Trans. Inform. Theory* **1998**, *44*, 225–240.

24. Csiszàr, I.; Narayan, P. Common randomness and secret key generation with a helper. *IEEE Trans. Inform. Theory* **2000**, *46*, 344–366.

25. Wolf, S.; Wullschleger, J. Zero-error information and application in cryptography. In Proceedings of the 2004 IEEE Information Theory Workshop (ITW 2004), San Antonio, TX, USA, 24–29 October 2004.

26. Cover, T.; Gamal, A.E.; Salehi, M. Multiple access channels with arbitrarily correlated sources. *IEEE Trans. Inform. Theory* **1980**, *26*, 648– 657.

27. Zhang, Z. On a new non-Shannon type information inequality. *Comm. Inform. Syst.* **2003**, *3*, 47–60.

28. Hammer, D.; Romashchenko, A.; Shen, A.; Vereshchagin, N. Inequalities for Shannon entropy and Kolmogorov complexity. *J. Comput. Syst. Sci.* **2000**, *60*, 442–464.

29. Niesen, U.; Fragouli, C.; Tuninetti, D. On capacity of line networks. *IEEE Trans. Inform. Theory* **2011**, submitted for publication.

30. Fujishige, S. Polymatroidal dependence structure of a set of random variables. *Inform. Contr.* **1978**, *39*, 55–72.

31. Cicalese, F.; Vaccaro, U. Supermodularity and subadditivity properties of entropy on the majorization lattice. *IEEE Trans. Inform. Theory* **2002**, *48*, 933–938.

32. Chernov, A.; Muchnik, A.; Romashchenko, A.; Shen, A.; Vereshchagin, N. Upper semilattice of binary strings with the relation 'x is simple conditional to y'. *Theor. Comput. Sci.* **2002**, *271*, 69–95.

33. Dummit, D.S.; Foote, R.M. *Abstract Algebra*, 3rd ed.; Wiley, New York, NY, USA, 2003.

34. Yeung, R.W. *A First Course in Information Theory*; Kluwer Academic/Plenum Publishers: New York, NY, USA, 2002.

35. Oxley, J.G. *Matroid Theory*; Oxford University Press: New York, NY, USA, 1992.

36. Zhang, Z.; Yeung, R.W. On characterization of entropy function via information inequalities. *IEEE Trans. Inform. Theory* **1998**, *44*, 1440–1452.

37. Dougherty, R.; Freiling, C.; Zeger, K. Six New Non-Shannon information inequalities. In Proceedings of the 2006 IEEE International Symposium on Information Theory, Seattle, WA, USA, 9–14 July 2006; pp. 233–236.

38. Pippenger, N. What are the laws of information theory. In Proceedings of the 1986 Special Problems on Communication and Computation Conference, Palo Alto, CA, USA, 3–5 September 1986.

39. Matúš, F. Piecewise linear conditional information inequality. *IEEE Trans. Inform. Theory* **2006**, *52*, 236–238.

40. Li, H.; Chong, E.K.P. On connections between group homomorphisms and the Ingleton inequality. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; pp. 1996–2000.

41. Chan, T. Recent progresses in characterising information inequalities. *Entropy* **2011**, *13*, 379–401.

42. Chan, T. On the optimality of group network codes. In Proceedings of the International Symposium on Information Theory, Adelaide, Australia, 4–9 September 2005.

43. Mao, W.; Thill, M.; Hassibi, B. On group network codes: Ingleton-bound violations and independent sources. In Proceedings of the 2010 IEEE International Symposium on Information Theory, Austin, TX, USA, 13–18 June 2010; pp. 2388–2392.

44. Matúš, F. Infinitely many information inequalities. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; pp. 41–44.

Appendix

A.   Proof of Theorem 1

Proof. To show two lattices are isomorphic, we need to demonstrate a mapping, from one lattice to the other, such that it is a lattice-morphism—it honors both join and meet operations—and bijective as well. Instead of proving that $L_\mathbf{G}$ is isomorphic to L_G directly, we show that the dual of $L_\mathbf{G}$ is isomorphic to $L_\mathbf{M}$. Figuratively speaking, the dual of a lattice L is the lattice obtained by flipping L upside down. Formally, the dual lattice L' of a lattice L is the lattice defined on the same set with the partial order

reversed. Accordingly, the join operation of the prime lattice L corresponds to the meet operation for the dual lattice L$'$ and the meet operation of L to the join operation for L$'$. In the other words, we show that L$_G$ is isomorphic to L$_\mathbf{M}$ by demonstrating a bijective mapping $\phi : L_G \to L_\mathbf{M}$ such that

$$\phi(G \vee G') = \phi(G) \wedge \phi(G') \tag{9}$$

and

$$\phi(G \wedge G') = \phi(G) \vee \phi(G') \tag{10}$$

hold for all $G, G' \in L_\mathbf{G}$.

Note that each subgroups on the subgroup lattice L$_\mathbf{G}$ is obtained from the set $\mathbf{G} = \{G_i : i \in [n]\}$ via a sequence of join and meet operations and each information element on the information lattice L$_\mathbf{M}$ is obtained similarly from the set $\mathbf{M} = \{m_i : i \in [n]\}$. Therefore, to show that L$_\mathbf{G}$ is isomorphic to L$_\mathbf{M}$, according to the induction principle, it is enough to demonstrate a bijective mapping ϕ such that

- $\phi(G_i) = m_i$, for all $G_i \in \mathbf{G}$ and $m_i \in \mathbf{M}$;

- For any $G, G' \in L_\mathbf{G}$, if $\phi(G) = m$ and $\phi(G') = m'$, then

$$\phi(G \vee G') = m \wedge m' \tag{11}$$

and

$$\phi(G \wedge G') = m \vee m' \tag{12}$$

Naturally, we take $\phi : L_\mathbf{G} \to L_\mathbf{M}$ to be the mapping that assigns to each subgroup $G \in L_\mathbf{G}$ the information element identified by the coset-partition of the subgroup G. Thus, the initial step of the induction holds by assumption. On the other hand, it is easy to see that the mapping ϕ so defined is bijective simply because different subgroups always produce different coset-partitions and *vice versa*. Therefore, we are left to show that Equations (11) and (12) holds.

We first show that ϕ satisfies Equation (11). In other words, we show that the coset-partition of the intersection subgroup $G \cap G'$ is the coarsest among all the sample-space-partitions that are finer than both the coset-partitions of G and G'. To see this, let Π be a sample-space-partition that is finer than both the coset-partitions of G and G' and π be a part of Π. Since Π is finer than the coset-partitions of G, π must be contained in some coset C of G. For the same reason, π must be contained in some coset C' of G' as well. Consequently, $\pi \subseteq C \cap C'$ hold. Realizing that $C \cap C'$ is a coset of $G \cap G'$, we conclude that the coset-partition of $G \cap G'$ is coarser than Π. Since Π is chosen arbitrary, this proves that the coset-partition of the intersection subgroup $G \cap G'$ is the coarsest among all the sample space partitions that are finer than both the coset-partitions of G and G'. Therefore, Equation (11) holds for ϕ.

The proof for Equation (12) is more complicated. We use an idea called "transitive closure". Similarly, we need to show that the coset-partition of the subgroup $G \vee G'$ generated from the union of G and G' is the finest among all the sample-space-partitions that are coarser than both the coset-partitions of G and G'. Let Π be a sample-space-partition that is coarser than both the coset-partitions of G and G'. Denote the coset partition of the subgroup $G \vee G'$ by $\bar{\Pi}$. Let $\bar{\pi}$ be a part of $\bar{\Pi}$. It suffices to show that $\bar{\pi}$ is contained in some part of Π. Pick an element x from $\bar{\pi}$. This element x must belong to some part π of Π. It remains to show $\bar{\pi} \subseteq \pi$. In other words, we need to show that $y \in \pi$ for any $y \neq x, y \in \pi^{ij}$. Note

that π is a part of the coset-partition of the subgroup $G_i \vee G_j$. In other words, π is a coset of $G_i \vee G_j$. The following reasoning depends on the following fact from group theory [33].

Proposition 15. *Two elements g_1 and g_2 belong to a same (right) coset of a subgroup if and only if $g_1 g_2^{-1}$ belongs to the subgroup.*

Since x and y belong to a same coset $\bar{\pi}$ of the subgroup $G \vee G'$, we have $yx^{-1} \in G \vee G'$. Note that any element g from $G \vee G'$ can be written in the form of $g = a_1 b_1 a_1 b_2 \cdots a_K b_K$ where $a_k \in G$ and $b_k \in G'$ for all $k \in [K]$. Suppose $yx^{-1} = g = a_1 b_1 a_1 b_2 \cdots a_K b_K$. We have

$$y = a_1 b_1 a_2 b_2 \cdots a_K b_K x$$

In the following we shall show that y belongs to π by induction on the sequence $a_1 b_1 \cdots a_K b_K$.

First, we claim $b_K x \in \pi$. To see this, note that $x \in \pi$. Since $(b_K x)x^{-1} = b_K \in G'$, by Proposition 15, we know that $b_K x$ and x belong to a same coset C_K of G'. By assumption, the partition Π is coarser than the coset-partition of G', the coset C_K must be contained in π, since it already contains an element x of C_K.

For the same reason, with $b_K x \in \bar{\pi}$ showed, we can see that $a_K b_K x$ belongs to π as well, because $(a_K b_K x)(b_K x)^{-1} = a_K \in G$ implies $a_K b_K x$ and $b_K x$ belong to a same coset of G.

Continuing the above argument inductively on the sequence $a_1 b_1 \cdots a_K b_K$, we can finally have $a_1 b_1 \cdots a_K b_K x \in \pi$. Therefore, we have $y \in \pi$. This concludes the proof. $\square$

B. Proof of Theorem 3

Proof. The approximation process is decomposed into three steps. The first step is to "dilate" the sample space such that we can turn a non-uniform probability space into a uniform probability space. The sample space partitions of the information elements are accordingly "dilated" as well. After dilating the sample space, depending on the approximation error tolerance, *i.e.*, ϵ, we may need to further "amplify" the sample space. Then, we follow the same procedure as in Section 3.4 and construct a subgroup lattice using the orbit-partition–permutation-group-action correspondence.

We assume the probability measure $\mathbf{P}$ on the sample space are rational. In other words, the probabilities of the elementary event $p_i = \Pr\{\omega_i\}$, $\omega_i \in \Omega$ are all rational numbers, namely $p_i = \frac{p_i}{q_i}$ for some $p_i, q_i \in \mathbb{N}$. This assumption is reasonable, because any finite dimensional real vector can be approximated, up to an arbitrary precision, by some rational vector.

Let M be the least common multiple of the set $\{q_i\}$ of denominators. We "split" each sample point in Ω into $\frac{M p_i}{q_i}$ points. Note that $\frac{M p_i}{q_i}$ is integral. We need to accordingly "dilate" the sample space partitions of the information elements. Specifically, for each part π of the partition of every information element m_i, its "dilated" partition π', in the dilated sample space $\hat{\Omega}$, contains exactly all the sample points that are "split" from the sample points in π. The dilated sample space $\hat{\Omega}$ has size of $\sum_{\omega_i \in \Omega} \frac{M p_i}{q_i}$. To maintain the probability structure, we assign to each sample point in the dilated sample space $\hat{\Omega}$ probability $\frac{1}{|\hat{\Omega}|}$. In other words, we equip the dilated sample space with a uniform probability measure. It is easy to check that the entire (quantitative) probability structure remains the same. Thus, we can consider all the information elements as if defined on the dilated probability space.

If necessary, depending on the approximation error tolerance ϵ, we may further "amplify" the dilated sample space $\hat{\Omega}$ by K times by "splitting" each of its sample points into to K points. At the same time, we scale the probability of each sample point in the post-amplification sample space down by K times to $\frac{1}{K|\hat{\Omega}|}$. By abusing of notation, we still use $\hat{\Omega}$ to denote the post-amplification sample space. Similar to the "dilating" process, all the partitions are accordingly amplified.

Before we move to the third step, we compute entropies for information elements in terms of the cardinality of the parts of its dilated sample space partition. Consider an information element m_i. Denote its pre-dilation sample space partition by $\Pi_i = \{\pi_i^j, j \in [J]\}$ and its post-amplification sample space partition by $\hat{\Pi}_i = \{\hat{\pi}_i^j, j \in [J]\}$. It is easy to see that the entropy $H(m_i)$ can be calculated as follows:

$$
\begin{aligned}
H(m_i) &= -\sum_{j \in [J]} \Pr\{\pi_i^j\} \log \Pr\{\pi_i^j\} \\
&= -\sum_{j \in [J]} \Pr\{\hat{\pi}_i^j\} \log \Pr\{\hat{\pi}_i^j\} \\
&= -\sum_{j \in [J]} \frac{|\hat{\pi}_j^i|}{|\hat{\Omega}|} \log \frac{|\hat{\pi}_j^i|}{|\hat{\Omega}|}
\end{aligned}
\tag{13}
$$

All the entropies of the other information elements, including the joint and common information elements, on the entire information lattices can be computed in the exactly same way in terms of the cardinalities of the parts of their dilated sample space partitions.

In the third step, we follow the same procedure as in Section 3.4, and construct, based on the orbit-partition–permutation-group-action correspondence, a subgroup lattice that isomorphic to the information lattice generated by the set of information elements $\{m_i : i \in [n]\}$. More specifically, the subgroup lattice is constructed according to their "post-amplification" sample space partitions. Suppose, on the constructed subgroup lattice, the permutation groups G_i corresponds to the information element m_i. As in the above, the "post-amplification" sample space partition of m_i is $\hat{\Pi}_i = \{\hat{\pi}_i^j, j \in [J]\}$. Then, the cardinality of the permutation group is simply

$$
|G_i| = \prod_{j \in J} \hat{\pi}_i^j!
$$

According to the isomorphism relation established in Theorem 2, the above calculations remain valid for all the subgroups on the subgroup lattices.

Recall that all the groups on the subgroup lattice are permutation groups and are all subgroups of the symmetry group of order $|\hat{\Omega}|!$. So the log-index of G_i, corresponding to m_i, is

$$
\log \frac{|\hat{\Omega}|!}{|G_i|} = \log \frac{|\hat{\Omega}|!}{\prod_{j \in J} \hat{\pi}_i^j!}
\tag{14}
$$

As we see from Equations (1) and (2) of Proposition 7, the entropies of the coset-partition information elements on information lattices equal *exactly* the log-indices of their subgroups on subgroup lattices. However, for the information lattice generated from general information elements, namely information elements with non-equal sample space partitions, as we see from Equations (13) and (14), the entropies of the information elements on the information lattice does not equal the log-indices of their corresponding

permutation groups on the subgroup lattices exactly any more. But, as we can shall see, the entropies of the information elements are well *approximated* by the log-indices of their corresponding permutation groups. Recall the following Stirling's approximation formula for factorials:

$$\log n! = n \log n - n + o(n) \tag{15}$$

"Normalizing" the log-index in Equation (14) by a factor $\frac{1}{|\hat{\Omega}|}$ and then substituting the factorials with the above Stirling approximation formula, we get

$$\frac{1}{|\hat{\Omega}|} \log \frac{|\hat{\Omega}|}{|G_i|} = \frac{1}{|\hat{\Omega}|} \left(|\hat{\Omega}| \log |\hat{\Omega}| - |\hat{\Omega}| - \left(\sum_{j \in [J]} |\hat{\pi}_i^j| \log |\hat{\pi}_i^j| - |\hat{\pi}_i^j| \right) + o(|\hat{\Omega}|) \right)$$

Note that in the above substitution process, we combined some finite $o(|\hat{\Omega}|)$ terms "into" one $o(|\hat{\Omega}|)$ term. It is clear that $\sum_{j \in [J]} |\hat{\pi}_i^j| = |\hat{\Omega}|$, since $\{\hat{\pi}_i^j : j \in [J]\}$ forms a partition of $\hat{\Omega}$. Therefore, we get

$$\frac{1}{|\hat{\Omega}|} \log \frac{|\hat{\Omega}|}{|G_i|} = \frac{1}{|\hat{\Omega}|} \left(|\hat{\Omega}| \log |\hat{\Omega}| - \sum_{j \in [J]} |\hat{\pi}_i^j| \log |\hat{\pi}_i^j| + o(\hat{\Omega}) \right) = H(m_i) + \frac{o(|\hat{\Omega}|)}{|\hat{\Omega}|}$$

So, the difference between the entropy $H(m_i)$ and the normalized log-index of its corresponding permutation subgroup G_i diminishes for $\hat{\Omega}$ large.

Since both the entropy vector $h_{\mathbf{M}}$ and the log-index vector $l_{\mathbf{G}^N}$ are of finite dimension, it follows easily

$$\left\| h_{\mathbf{M}} - \frac{l_{\mathbf{G}^N}}{N} \right\|_1 = \frac{o(|\hat{\Omega}|)}{|\hat{\Omega}|} \to 0$$

with

$$N = |\hat{\Omega}| = K \sum_{\omega_i \in \Omega} \frac{M p_i}{q_i} \to \infty, \text{ by taking } K \to \infty.$$

This concludes the proof. $\square$