An algebraic approach to information theory

Manas K. Patra and Samuel L. Braunstein
Department of Computer Science, University of York, York YO10 5DD, UK

***Abstract*—This work proposes an algebraic model for classical information theory. We first give an algebraic model of probability theory. Information theoretic constructs are based on this model. In addition to theoretical insights provided by our model one obtains new computational and analytical tools. Several important theorems of classical probability and information theory are presented in the algebraic framework.**

I. Introduction

The present paper reports a brief synopsis of our work on an algebraic model of *classical* information theory based on operator algebras. Let us recall a simple model of a communication system proposed by Shanon [Sha48]. This model has essentially four components: source, channel, encoder/decoder and receiver. Some amount of noise affects every stage of the operation and the behavior of components are generally modeled as stochastic processes. In this work our primary focus will be on discrete processes. A discrete source can be viewed as a generator of a countable set of random variables. In a communication process the source generates sequence of random variables. Then it is sent through the channel (with encoding/decoding) and the output at the receiver is another sequence of random variables. Thus, the concrete objects or *observables*, to use the language of quantum theory, are modeled as random variables. The underlying probability space is primarily used to define probability distributions or *states* associated with the relevant random variables. In the algebraic approach we directly model the observables. Since random variables can be added and multiplied [1]they constitute an *algebra*. This is our starting point. In fact, the algebra of random variables have a richer structure called a C^* algebra. Starting with a C^* algebra of observables we can define most important concepts in probability theory in general and information theory in particular. A natural question is: why should we adopt this algebraic approach? We discuss the reasons below.

First, it seems more appropriate to deal with the "concrete" quantities, *viz*. observables and their intrinsic structure. The choice of underlying probability space is somewhat arbitrary as a comparison of standard textbooks on information theory [CT99], [CK81] reveals. Moreover, from the algebra of observables we can recover particular probability spaces from representations of the algebra. Second, some constraints, may have to be imposed on the set of random variables. In security protocols different participants have access to different sets of observables and may assign different probability structures. In this case, the algebraic approach seems more natural: we have to study different subalgebras. Third, the algebraic approach gives us new theoretical insights and computational tools. This will be justified in the following sections. Finally, and this was our original motivation, the algebraic approach provides the basic framework for a unified approach to classical *and* quantum information. All quantum protocols have some classical components, e.g. classical communication, "coin-tosses" etc. But the language of the two processes, classical and quantum, seem quite different. In the former we are dealing with random variables defined on one or more probability spaces where as in the latter we are processing quantum states which also give complete information about the measurement statistics of *quantum* observables. The algebraic framework is eminently suitable for bringing together these somewhat disparate viewpoints. Classical observables are simply elements that commute with every element in the algebra.

The connection between operator algebras and information theory—classical *and* quantum—have appeared in the scientific literature since the beginnings of information theory and operator algebras—both classical and quantum (see e.g. [Ume62], [Seg60], [Ara75], [Key02], [BKK07], [KW06]). Most previous work focus on some aspects of information theory like the noncommutative generalizations of the concepts of entropy. There does not appear to be a unified and coherent approach based on intrinsically algebraic notions. The construction of such a model is one of the goals of the paper. As probabilistic concepts play such an important role in the development of information theory we first present an algebraic approach to probability. I. E. Segal [Seg54] first proposed such an algebraic approach model of probability theory. Later Voiculescu [VDN92] developed noncommutative or "free probability" theory. We believe several aspects of our approach are novel and yield deeper insights to information processes. In this summary, we have omitted most proofs or give only brief outlines. The full proofs can be found in our arXiv submission [PB]. A brief outline of the paper follows.

In Section II we give the basic definitions of the C^* algebras. This is followed by an account of probabilistic concepts from an algebraic perspective. In particular, we investigate the fundamental notion of independence and demonstrate how it relates to the algebraic structure. One important aspect in which our approach seems novel is the treatment of probability distribution functions. In Section III we give a precise algebraic model of information/communication system. The fundamental concept of entropy is introduced. We also define and study the crucial notion of a channel as a (completely) positive map. In particular, the *channel coding theorem* is presented as an approximation result. Stated informally: *Every channel other than the useless ones can be approximated by a lossless channel under appropriate coding*. We conclude the paper with some comments and discussions.

[1]We assume that they are real or complex valued.

II. C^* Algebras and Probability

A Banach algebra A is a complete normed algebra [Rud87], [KR97]. That is, A is an algebra over real ($\mathbb{R}$) or complex numbers ($\mathbb{C}$), for every $x \in A$ the norm $\|x\| \geqslant 0$ is defined satisfying the usual properties and every Cauchy sequence converges in the norm. A C^* algebra B is a Banach algebra[KR97] with an anti-linear involution $*$ ($x^{**} = x$ and $(x + cy)^* = x^* + \bar{c}y^*$, $x, y \in B$ and $c \in \mathbb{C}$) such that $\|xx^*\| = \|x\|^2$ and $(xy)^* = y^*x^*\forall x, y \in B$. This implies that $\|x\| = \|x^*\|$. We often assume that the unit $I \in B$. The fundamental Gelfand-Naimark-Segal (**GNS**) theorem states that every C^* algebra can be isometrically embedded in some $\mathcal{L}(H)$, the set of bounded operators on a Hilbert space of H. The spectrum of an element $x \in B$ is defined by $\mathtt{sp}(x) = \{c \in \mathbb{C} : x - cI$ invertible $\}$. The spectrum is a nonempty closed and bounded set and hence compact. An element x is self-adjoint if $x = x^*$, normal if $x^*x = xx^*$ and positive (strictly positive) if x is self-adjoint and $\mathtt{sp}(x) \subset [0, \infty)((0, \infty))$. A self-adjoint element has a real spectrum and conversely. Since $x = x_1 + ix_2$ with $x_1 = (x + x^*)/2$ and $x_1 = (x + x^*)/2i$ any element of a C^* algebra can be decomposed into self-adjoint "real" and "imaginary" parts. The positive elements define a partial order on A: $x \leqslant y$ iff $y - x \geqslant 0$ (positive). A positive element a has a unique square-root $\sqrt{a}$ such that $\sqrt{a} \geqslant 0$ and $(\sqrt{a})^2 = a$. If x is self-adjoint, $x^2 \geqslant 0$ and $|x| = \sqrt{x^2}$. A self-adjoint element x has a decomposition $x = x_+ - x_-$ into positive and negative parts where $x_+ = (|x| + x)/2$ and $x_- = (|x| - x)/2)$ are positive. An element $p \in B$ is a projection if p is self-adjoint and $p^2 = p$. Given two C^*-algebras A and B a homomorphism F is a linear map preserving the product and $*$ structures. A homomorphism is positive if it maps positive elements to positive elements. A (linear) functional on A is a linear map $A \to \mathbb{C}$. A positive functional ω such that $\omega(\mathbb{1}) = 1$ is called a *state*. The set of states G is convex. The extreme points are called *pure states* and G is the convex closure of pure states (Krein-Millman theorem). A set $B \subset A$ is called a subalgebra if it is a C^* algebra with the inherited product. A subalgebra is called unital if it contains the identity of A. Our primary interest will be on *abelian* or commutative algebras. The basic representation theorem (Gelfand-Naimark) [KR97] states that: *An abelian C^* algebra with unity is isomorphic to the algebra $C(X)$ continuous complex-valued functions on a compact Hausdorff space X.*

Now let $X = \{a_1, \ldots, a_n\}$ be a finite set with discreet topology. Then $A = C(X)$ is the set of all functions $X \to \mathbb{C}$. The algebra $C(X)$ can be considered as the algebra of (complex) random variables on the finite probability space X. Let $x_i(a_j) = \delta_{ij},\ i, j = 1, \ldots, n$. Here $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. The functions $x_i \in A$ form a basis for A. Their multiplication table is particularly simple: $x_ix_j = \delta_{ij}x_i$. They also satisfy $\sum_i x_i = \mathbb{1}$. These are projections in A. They are orthogonal in the sense that $x_ix_j = 0$ for $i \neq j$. We call any basis consisting of elements of norm 1 with distinct elements orthogonal *atomic*. A set of linearly independent elements $\{y_i\}$ satisfying $\sum_i y_i = \mathbb{1}$ is said to be complete. The next theorem gives us the general structure of any finite-dimensional algebra.

Theorem 1. *Let A be a finite-dimensional abelian C^* algebra. Then there is a unique (up to permutations) complete atomic basis $\mathcal{B} = \{x_1, \ldots, x_n\}$. That is, the basis elements satisfy*

$$x_i^* = x_i,\ x_ix_j = \delta_{ij}x_i,\ \|x_i\| = 1 \text{ and } \sum_i x_i = \mathbb{1}, \qquad (1)$$

Let $x = \sum_i a_ix_i \in A$. Then $\mathtt{sp}(x) = \{a_i\}$ and hence $\|x\| = \max_i\{|a_i|\}$.

We next describe an important construction for C^* algebras. Given two C^* algebras A and B, the tensor product $A \otimes B$ is defined as follows. As a set it consists of all finite linear combinations of symbols of the form $\{x \otimes y : x \in A, y \in B\}$ subject to the conditions that the map $(x, y) \to x \otimes y$ is bilinear in each variable. Hence, if $\{x_i\}$ and $\{y_j\}$ are bases for A and B respectively then $\{x_i \otimes y_j\}$ is a basis for $A \otimes B$. The linear space $A \otimes B$ becomes an algebra by defining $(x \otimes y)(u \otimes z) = xu \otimes yz$ and extending by bilinearity. The $*$ is defined by $(x \otimes y)^* = x^* \otimes y^*$ and extending *anti-linearly*. We will define the norm in a more general setting. Our basic model will be an *infinite* tensor product of finite dimensional C^* algebras which we present next.

Let $A_k,\ k = 1, 2, \ldots,$ be finite dimensional abelian C^* algebras with atomic basis $B_k = \{x_{k1}, \ldots, x_{kn_k}\}$. Let B^∞ be the set consisting of all infinite strings of the form $z_{i_1} \otimes z_{i_2} \otimes \cdots$ where all but a finite number (> 0) of z_{i_k}s are equal to $\mathbb{1}$ and if some $z_{i_k} \neq \mathbb{1}$ then $z_{i_k} \in B_k$. Let $\tilde{\mathfrak{A}} = \otimes_{i=1}^\infty A_i$ be the vector space with basis B^∞ such that $z_{i_1} \otimes z_{i_2} \otimes \cdots \otimes z_{i_k} \otimes \cdots$ is linear in each factor separately. We define a product in $\tilde{\mathfrak{A}}$ as follows. First, for elements of B^∞: $(z_{i_1} \otimes z_{i_2} \otimes \cdots)(z'_{i_1} \otimes z'_{i_2} \otimes \cdots) = (z_{i_1}z'_{i_1} \otimes z_{i_2}z'_{i_2} \otimes \cdots)$ We extend the product to whole of $\tilde{\mathfrak{A}}$ by linearity. Next define a norm by:

$$\| \sum_{i_1, i_2, \ldots} a_{i_1 i_2 \ldots} z_{i_1} \otimes z_{i_2} \otimes \cdots \| = \sup\{|a_{i_1 i_2 \ldots}|\}$$

B^∞ is an atomic basis. It follows that $\tilde{\mathfrak{A}}$ is an abelian normed algebra. We define $*$-operation by $\left(\sum_{i_1, i_2, \ldots} a_{i_1 i_2 \ldots} z_{i_1} \otimes z_{i_2} \otimes \cdots\right)^* = \sum_{i_1, i_2, \ldots} \overline{a_{i_1 i_2 \ldots}} z_{i_1} \otimes z_{i_2} \otimes \cdots$ It follows that for $x \in \tilde{\mathfrak{A}}$, $\|xx^*\| = \|x\|^2$. Finally, we complete the norm [KR97] and call the resulting C^* algebra $\mathfrak{A}$. With these definitions $\mathfrak{A}$ is a C^* algebra. We call a C^* algebra B of **finite type** if it is either finite dimensional or infinite tensor product of finite-dimensional algebras. An important special case is when all the factor algebras $A_i = A$. We then write the infinite tensor product C^* algebra as $\bigotimes^\infty A$. Intuitively, the elements of an atomic basis B^∞ of $\bigotimes^\infty A$ correspond to strings from an alphabet (represented by the basis B). Of particular interest is the 2-dimensional algebra D corresponding to a binary alphabet.

The next step is to describe the state space. Given a C^* subalgebra $V \subset A$ the set of states of V will be denoted by $\mathscr{S}(V)$. Let $\mathfrak{A} = \otimes_{i=1}^\infty A_i$ denote the infinite tensor product of finite-dimensional algebras A_i. An infinite product state of $\mathfrak{A}$ is a functional of the form $\Omega = \omega_1 \otimes \omega_2 \otimes \cdots$ such that $\omega_i \in \mathscr{S}(A_i)$ This is indeed a state of $\mathfrak{A}$ for if $\alpha_k = z_1 \otimes z_2 \otimes \cdots \otimes z_k \otimes \mathbb{1} \otimes \mathbb{1} \cdots \in \mathfrak{A}$ then $\Omega(\alpha) = \omega_1(z_1)\omega_2(z_2) \cdots \omega_k(z_k)$, a *finite* product. A general state on $\mathfrak{A}$ is a convex combination of product states like Ω. Finally, we discuss another useful construction in a C^* algebra A. If $f(z)$ is an analytic function whose Taylor series $\sum_{n=0}^{\infty} a_n(z-c)^n$ converges in a region $|z-c| < R$. Then the series $\sum_{n=0}^{\infty}(x-c\mathbb{1})^n$ converges and it makes sense to talk of analytic functions on a C^* algebra. If we have an atomic basis $\{x_1, x_2, \dots\}$ in an abelian C^* algebra then the functions are particularly simple in this basis. Thus if $x = \sum_i a_i x_i$ then $f(x) = \sum_i f(a_i)x_i$ provided that $f(a_i)$ are defined in an appropriate domain.

We gave a brief description of C^* algebras. We now introduce an algebraic model of probability which is used later to model communication processes. In this model we treat random variables as elements of a C^* algebra. The probabilities are introduced via states. A classical observable algebra is a complex abelian C^* algebra A. We can restrict our attention to real algebras whenever necessary. The Riesz representation theorem [Rud87] makes it possible identify ω with some *probability measure*. A *probability algebra* is a pair (A, S) where A is an observable algebra and $S \subset \mathscr{S}(A)$ is a set of states. A probability algebra is defined to be *fixed* if S contains only one state.

Let ω be a state on an abelian C^* algebra A. Call two elements $x, y \in A$ *uncorrelated in the state* ω if $\omega(xy) = \omega(x)\omega(y)$. This definition depends on the state: two uncorrelated elements can be correlated in some other state ω'. A state ω is called multiplicative if $\omega(xy) = \omega(x)\omega(y)$ for all $x, y \in A$. The set of states, $\mathscr{S}$, is convex. The extreme points of $\mathscr{S}$ are called *pure* states. In the case of abelian C^* algebras a state is pure if and only of it is multiplicative [KR97]. Thus, in a pure state any two observables are uncorrelated. This is not generally true in the non-abelian quantum case. Now we can introduce the important notion of *independence*. Given $S \subset A$ let $A(S)$ denote the subalgebra generated by S (the smallest subalgebra of A containing S). Two subsets $S_1, S_2 \subset A$ are defined to be *independent* if all the pairs $\{(x_1, x_2) : x_1 \in A(S_1), x_2 \in A(S_2)\}$ are uncorrelated. As independence and correlation depend on the state we sometimes write ω-independent/uncorrelated. Independence is a much stronger condition than being uncorrelated. The next theorem states the structural implications of independence.

Theorem 2. *Two sets of observables S_1, S_2 in a finite dimensional abelian C^* algebra A are independent in a state ω if and only if for the subalgebras $A(S_1)$ and $A(S_2)$ generated by S_1 and S_2 respectively there exist states $\omega_1 \in \mathscr{S}(A(S_1))$, $\omega_2 \in \mathscr{S}(A(S_2))$ such that $(A(S_1) \otimes A(S_2), \{\omega_1 \otimes \omega_2\})$ is a cover of $(A(S_1S_2), \omega')$ where $A(S_1S_2)$ is the subalgebra generated by $\{S_1, S_2\}$ and ω' is the restriction of ω to $A(S_1S_2)$.*

We thus see the relation between independence and (tensor) product states in the classical theory. Next we show how one can formulate another important concept, *distribution function* (**d.f**) in the algebraic framework. We restrict our analysis to C^* algebras of finite type. The general case is more delicate and is defined using approximate identities in subalgebras in [PB]. The idea is that we approximate indicator functions of sets by a sequence of elements in the algebra. In the case of finite type algebras the sequence converges to a projection operator J_S. Thus, if we consider a representation where the elements of A are functions on some finite set F then J_S is precisely the indicator function of the set $S' = \{c : x_i(c) - t_i = 0 : c \in F \text{ and } i = 1, \dots, n\}$. The set S' corresponds to the subalgebra $(S_{\mathtt{t}})_a$ and J_S, a projection in A, acts as identity in $(S_{\mathtt{t}})_a$. From the notion of distribution functions we can define now probabilities $Pr(a \leqslant x \leqslant b)$ in the algebraic context. We can now formulate problems in any discrete stochastic process in finite dimensions. The algebraic method actually provides practical tools besides theoretical insights as the example of "waiting time" shows [PB]. Now we consider the algebraic formulation of a basic limit theorem of probability theory: the **weak law of large numbers**. From information theory perspective it is perhaps the most useful limit theorem. Let $X_1, X_2, \cdots, X_n$ be independent, identically distributed (i.i.d) bounded random variables on a probability space Ω with probability measure P. Let μ be the mean of X_1. Recall the *Weak law of large numbers*. Given $\epsilon > 0$

$$\lim_{n\to\infty} P(|S_n = \frac{X_1 + \cdots + X_N}{n} - \mu| > \epsilon) = 0$$

We have an algebraic version of this important result.

Theorem 3 (Law of large numbers (weak))**.** *If $x_1, \dots, x_n, \dots$ are ω-independent self-adjoint elements in an observable algebra and $\omega(x_i^k) = \omega(x_j^k)$ for all positive integers i, j and k (identically distributed) then*

$$\lim_{n\to\infty} \omega(|\frac{x_1 + \cdots + x_n}{n} - \mu|^k) = 0 \text{ where } \mu = \omega(x_1) \text{ and } k > 0$$

Using the algebraic version of Chebysev inequality the above result implies the following. Let $x_1, \dots, x_n$ and μ be as in the Theorem and set $s_n = (x_1 + \cdots + x_n)/n$. Then for any $\epsilon > 0$ there exist n_0 such that for all $n > n_0$ $P(|s_n - \mu| > \epsilon) < \epsilon$

III. Communication and Information

We now come to our original theme: an algebraic framework for communication and information processes. Since our primary goal is the modeling of information processes we refer to the simple model of communication in the Introduction and model different aspects of it. In this work we will only deal with sources with a finite alphabet.

Definition. *A source is a pair $\mathscr{S} = (B, \Omega)$ where B is an atomic basis of a finite-dimensional abelian C^* algebra A and Ω is a state in $\bigotimes^{\infty} A$.*

This definition abstracts the essential properties of a source. The basis B is called the *alphabet*. A typical output of the source is of the form $x_1 \otimes x_2 \otimes \cdots \otimes x_k \otimes \mathbb{1} \otimes \cdots \in B^{\infty}$, the infinite product basis of $\bigotimes^{\infty} A$. We identify $\hat{x}_k = \mathbb{1} \otimes \cdots \otimes \mathbb{1} \otimes x_k \otimes \mathbb{1} \otimes \cdots$ with the kth signal. If these are independent then Theorem 2 tells us that Ω must be product state. Further, if the state of the source does not change then $\Omega = \omega \otimes \omega \otimes \cdots$ where ω is a state in A. For a such state ω define: $\mathcal{O}_\omega = \sum_{i=1}^n \omega(x_i)x_i,\ \{x_1, \dots, x_n\},\ x_i \in B$ We say that $\mathcal{O}_\omega$ is the "instantaneous" output of the source in state ω. Let A' be another finite-dimensional C^* algebra with atomic basis B' A source coding is a linear map $f : B \to T = \sum_{k=1}^m \otimes^k A'$. Such that for $x \in B$, $f(x) = x'_{i_1} \otimes x'_{i_2} \otimes \cdots \otimes x'_{i_r},\ r \leqslant k$

with $x'_{i_j} \in B'$. Thus each "letter" in the alphabet B is coded by "words" of maximum length k from B'.

A code $f : B \to T$ is defined to be prefix-free if for distinct members x_1, x_2 in an atomic basis of B, $f'(x_1)f'(x_2) = 0$ where f' is the map $f' : B \to \bigotimes^{\infty} B'$ induced by f. That is, distinct elements of the atomic basis of B are mapped to *orthogonal* elements. Thus the "code-word" $z_1 \otimes z_1 \otimes \cdots \otimes z_k \otimes \mathbb{1} \otimes \mathbb{1} \otimes \cdots$ is not orthogonal to another $z'_1 \otimes z'_1 \otimes \cdots \otimes z'_m \otimes \mathbb{1} \otimes \mathbb{1} \otimes \cdots$ with $k \leqslant m$ if and only if $z_1 = z'_1, \ldots, z_k = z'_k$. The useful Kraft inequality can be proved using algebraic techniques. Corresponding to a finite sequence $k_1 \leqslant k_2 \leqslant \cdots \leqslant k_m$ of positive integers let $\alpha_1, \ldots, \alpha_m$ be a set of prefix-free elements in $\sum_{i \geqslant 1} \otimes^i A'$ such that $\alpha_i \in \otimes^{k_i} A'$. Further, suppose that each α_i is a tensor product of elements from B'. Then

$$\sum_{i=1}^{m} n^{k_m - k_i} \leqslant n^{k_m} \tag{2}$$

This inequality is proved by looking at bounds on dimensions of a sequence of orthogonal subspaces. In the following, we restrict ourselves to prefix-free codes. Using convexity function $f(x) = -\log x$ and the Kraft inequality 2 we deduce the following.

Proposition 1 (Noiseless coding)**.** *Let $\mathscr{S}$ be a source with output $\mathcal{O}_\omega \in A$, a finite-dimensional C^* algebra with atomic basis $\{x_1, \ldots, x_n\}$ (the alphabet). Let g be prefix-free code such that $g(x_i)$ is a tensor product of k_i members of the code basis. Then $\omega(\sum_i k_i x_i + \log \mathcal{O}_\omega) \geqslant 0$*

Next we give a simple application of the law of large numbers. First define a positive functional $\mathtt{Tr}$ on a finite dimensional abelian C^* algebra A with an atomic basis $\{x_1, \ldots, x_d\}$ by $\mathtt{Tr} = \omega_1 + \cdots + \omega_d$ where ω_i are the dual functionals. It is clear that $\mathtt{Tr}$ is independent of the choice of atomic basis.

Theorem 4 (Asymptotic Equipartition Property (AEP))**.** *Let $\mathscr{S}$ be a source with output $\mathcal{O}_\omega = \sum_{i=1}^{d} \omega(x_i)x_i$ where ω is a state on the finite dimensional algebra with atomic basis $\{x_i\}$. Then given $\epsilon > 0$ there is a positive integer n_0 such that for all $n > n_0$*

$$P(2^{n(H(\omega)-\epsilon)} \leqslant \otimes^n \mathcal{O}_\omega \leqslant 2^{n(H(\omega)+\epsilon)}) > 1 - \epsilon$$

where $H = \omega(\log_2(\mathcal{O}_\omega))$ is the entropy *of the source and the probability distribution is calculated with respect to the state $\Omega_n = \omega \otimes \cdots \otimes \omega$ (n factors) of $\otimes^n A$. If Q denotes the identity in the subalgebra generated by $(\epsilon I - |\log_2(\otimes^n \mathcal{O}_\omega) + nH|)_+$ then*

$$(1-\epsilon)2^{n(H(\omega)-\epsilon)} \leqslant \mathtt{Tr}(Q) \leqslant 2^{n(H(\omega)+\epsilon)}$$

Note that the element Q is a projection on the subalgebra generated by $(\epsilon I - |\log_2(\otimes^n \mathcal{O}_\omega) - nH|)_+$. It corresponds to the set of strings whose probabilities are between $2^{-nH-\epsilon}$ and $2^{-nH+\epsilon}$. The integer $\mathtt{Tr}(Q)$ is simply the cardinality of this set.

We now come to the most important part of the communication model: the *channel*. The original paper of Shannon characterized channels by a transition probability function. We will consider only (discrete) memoryless channel (DMS). A DMS channel has an input alphabet X and output alphabet Y and a *channel transformation matrix* $C(y_j|x_i)$ with $y_j \in Y$ and $x_i \in X$. Since the matrix $C(y_j|x_i)$ represents the probability that the channel outputs y_j on input x_i we have $\sum_j C(y_j|x_i) = 1$ for all i: $C(ij) = C(y_j|x_i)$ is *row stochastic*. This is the standard formulation. [CK81], [CT99]. We now turn to the algebraic formulation.

Definition. A DMS channel $\mathcal{C} = \{X, Y, C\}$ where X and Y are abelian C^* algebras of dimension m and n respectively and $C : Y \to X$ is a unital positive map. The algebras X and Y will be called the input and output algebras of the channel respectively. Given a state ω on X we say that (X, ω) is the input source for the channel.

Sometimes we write the entries of C in the more suggestive form $C_{ij} = C(y_j|x_i)$ where $\{y_j\}$ and $\{x_i\}$ are atomic bases for Y and X respectively. Thus $C(y_j) = \sum_i C_{ij} x_i = \sum_i C(y_j|x_i)x_i$. Note that in our notation C is an $m \times n$ matrix. Its transpose $C^T_{ji} = C(y_j|x_i)$ is the channel matrix in the standard formulation. We have to deal with the transpose because the channel is a map *from* the output alphabet to the input alphabet. This may be counterintuitive but observe that any map $Y \to X$ defines a unique dual map $\mathcal{S}(X) \to \mathcal{S}(Y)$, on the respective state spaces. Informally, a channel transforms a probability distribution on the input alphabet to a distribution on the output. We characterize a channel by input/output algebras (of observables) and a positive map. Like the source output we now define a useful quantity called *channel output*. Corresponding to the atomic basis $\{y_i\}$ of Y let $\otimes^k y_{i(k)}$ be an atomic basis in $\otimes^n Y$. Here $i(k) = (i_1 i_2 \ldots i_k)$ is a multi-index. Similarly we have an atomic basis $\{\otimes^k x_{j(k)}\}$ for $\otimes^k X$. The level-k channel output is defined to be $O^k_C = \sum_{i(k)} y_{i(k)} \otimes C^{(k)}(y_{i(k)})$. Here $C^{(k)}$ represents the channel transition probability matrix on the k-fold tensor product corresponding to strings of length k. In the DMS case it is simply the k-fold tensor product of the matrix C. The channel output defined here encodes most important features of the communication process. First, given the input source function $\mathcal{I}_{\omega^k} = \sum_i \omega^k(x_{i(k)})x_{i(k)}$ the output source function is defined by $\mathcal{O}_{\tilde{\omega}^k} = I \otimes \mathtt{Tr}_{\otimes^k X}((\mathbb{1} \otimes \mathcal{I}_{\omega^k})O^k_c) = \sum_i \sum_j C(y_{i(k)}|x_{j(k)})\omega^k(x_{j(k)})y_{i(k)}$. Here, the state $\tilde{\omega}^k$ on the output space $\otimes^k Y$ can be obtained via the dual $\tilde{\omega}^k(y) = \tilde{C}^k(\omega^k)(y) = \omega^k(C^k(y))$. The formula above is an alternative representation which is very similar to the quantum case. The *joint output* of the channel can be considered as the combined output of the two terminals of the channel. Thus the joint output

$$\begin{aligned} &\mathcal{J}_{\tilde{\Omega}^k} = (\mathbb{1} \otimes \mathcal{I}_{\omega^k})O^k_C = \sum_{ij} \Omega^k(y_{i(k)} \otimes x_{j(k)})y_{i(k)} \otimes x_{j(k)}, \\ &\Omega^k(y_{i(k)} \otimes x_{j(k)}) \equiv C(y_{i(k)}|x_{j(k)})\omega(x_{j(k)}) \end{aligned} \tag{3}$$

Let us analyze the algebraic definition of channel given above. For simplicity of notation, we restrict ourselves to level 1. The explicit representation of channel output is $\sum_i y_i \otimes \sum_j C(y_i|x_j)x_j$ We interpreted this as follows: if on the channel's out-terminal y_i is observed then the input could be x_j with probability $C(y_i|x_j)\omega(x_j)/\sum_j C(y_i|x_j)\omega(x_j)$. Now suppose that for a fixed i $C(y_i|x_j) = 0$ for all j except one say, j_i. Then on observing y_i at the output we are certain that the the input is x_{j_i}. If this is true for all values of y then we have an instance of a lossless channel. Given $1 \leqslant j \leqslant n$ let d_j be the set of integers i for which $C(y_i|x_j) > 0$. If the channel is lossless then $\{d_j\}$ form a partition of the set $\{1, \ldots, m\}$. The corresponding channel output is $O_C = \sum_j \left(\sum_{i \in d_j} C(y_i|x_j) y_i \right) \otimes x_j$. At the other extreme is the *useless* channel in which there is no correlation between the input and the output. To define it formally, consider a channel $\mathcal{C} = \{X, Y, C\}$ as above. The map C induces a map $C' : Y \otimes X \to X$ defined by $C'(y \otimes x) = xC(y)$. Given a state ω on X the dual of the map C' defines a state Ω_C on $Y \otimes X$: $\Omega_C(y \otimes x) = \omega(C'(y \otimes x)) = C(y|x)\omega(x)$. We call Ω_C the joint (input-output) state of the channel. A channel is useless if Y and X (identified as $Y \otimes \mathbb{1}$ and $\mathbb{1} \otimes X$ resp.) are Ω_C-independent. It is easily shown that: *a channel $\mathcal{C} = \{X, Y, C\}$ with input source (X, ω) is useless iff the matrix $C_{ij} = C(y_j|x_i)$ is of rank 1*. The algebraic version of the channel coding theorem assures that it is possible to approximate, in the long run, an arbitrary channel (excepting the useless case) by a lossless one.

Theorem 5 (Channel coding)**.** *Let $\mathcal{C}$ be a channel with input algebra X and output algebra Y. Let $\{x_i\}_{i=1}^n$ and $\{y_j\}_{j=1}^m$ be atomic bases for X and Y resp. Given a state ω on X, if the channel is not useless then for each k there are subalgebras $Y_k \subset \otimes^k Y, X_k \subset \otimes^k X$, a map $C_k : Y_k \to X_k$ induced by C and a lossless channel $L_k : Y_k \to X_k$ such that*

$$\lim_{k \to \infty} \Omega(|O_{C_k} - O_{L_k}|) = 0 \text{ on } T_k = Y_k \otimes X_k$$

Here $\Omega = \otimes^\infty \Omega_C$ and on $\otimes^k Y \otimes \otimes^k Y$ it acts as $\Omega^k = \otimes^k \Omega_C$ where Ω_C is the state induced by the channel and a given input state ω. Moreover, if $r_k = dim(X_k)$ then $R = \frac{\log r_k}{k}$, called transmission rate, is independent of k.

Let us clarify the meaning of the above statements. The theorem simply states that on the chosen set of codewords the channel output of C_k induced by the given channel can be made arbitrarily close to that of a lossless channel L_k. Since a lossless channel has a definite decision scheme for decoding the choice of L_k is effectively a decision scheme for decoding the original channel's output when the input is restricted to our "code-book". This implies probability of error tends to 0. Hence, it is possible to choose a set of "codewords" which can be transmitted with high reliability. The proof of the theorem [PB] uses algebraic arguments only. The theorem guarantees "convergence in the mean" in the appropriate subspace which implies convergence in probability. For a lossless channel the input entropy $H(X)$ is equal to the mutual information. We may think of this as conservation of entropy or information which justifies the term "lossless". Since it is always the case that $H(X) - H(X|Y) = I(X, Y)$ the quantity $H(X|Y)$ can be considered the loss due to the channel. The algebraic version of the theorem serves two primary purposes. It gives us the abelian perspective from which we will seek possible extensions to the non-commutative case. Secondly, the channel map L can be used for a decoding scheme. Thus we may think of a coding-decoding scheme for a given channel as a sequence of pairs (X_k, L_k) as above.

The coding theorems can be extended to more complicated scenarios like ergodic sources and channels with finite memory. We will not pursue these issues further here. But we are confident that these generalizations can be appropriately formulated and proved in the algebraic framework. In the preceding sections we have laid the basic algebraic framework for classical information theory. Although, we often confined our discussion to finite-dimensional algebras corresponding to finite sample spaces it is possible to extend it to infinite-dimensional algebras of continuous sample spaces. These topics will be investigated in the future in the non-commutative setting. We will delve deeper into these analogies and aim to throw light on some basic issues like quantum Huffman coding [BFGL00], channel capacities and general no-go theorems among others, once we formulate the appropriate models.

References

[Ara75] H. Araki. Relative entropy of states of von Neumann algebras. *Publications of the Research Institute for Mathematical Sciences*, 11:173–192, 1975.

[BFGL00] S. L. Braunstein, C. A. Fuchs, D. Gottesman, and H-K. Lo. A quantum analog of Huffman coding. *IEEE Trans. Inf. Th.*, 46:1545, 2000.

[BKK07] C. Bêny, A. Kempf, and D. W. Kribs. Quantum error correction of observables. *Phys. Rev. A.*, 76:042303, 2007.

[CK81] I. Csiszâr and K'orner. *Information Theory*. Academic Press, 1981.

[CT99] T. M. Cover and Joy. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1999.

[Key02] M. Keyl. Fundamentals of quantum information theory. *Phys. Rep.*, 369:531–548, 2002.

[KR97] R. V. Kadison and J. R. Ringrose. *Fundamentals of the Theory of Operator Algebras Vol. I.* American Mathematical Society, 1997.

[KW06] D. Kretschmann and R. F. Werner. Quantum channels with memory. *Phys. Rev. A.*, 72:062323, 2006.

[PB] M. K. Patra and S. L. Braunstein. An algebraic framework for information theory: Classical information. arXiv:0910.1536.

[Rud87] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 3 edition, 1987.

[Seg54] I. E. Segal. Abstract probability spaces and a theorem of Kolmogoroff. *Am. J. Math*, 76(3):721–732, 1954.

[Seg60] I. E. Segal. A note on the concept of entropy. *J. Math. Mech.*, 9:623–629, 1960.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:379–423,623–656, 1948.

[Ume62] H. Umegaki. Conditional expectation in an operator algebra, IV (Entropy and Information). *Kodai Mathematical Seminar Reports*, 14:59–85, 1962.

[VDN92] D. Voiculescu, K. Dykema, and A. Nica. *Free random variables*. CRM monograph series. AMS, 1992.