Available online at www.sciencedirect.com
SCIENCE DIRECT®
Electronic Notes in Theoretical Computer Science 158 (2006) 289–306
Electronic Notes in Theoretical Computer Science
www.elsevier.com/locate/entcs

Algebraic Information Theory For Binary Channels

Keye Martin[1] Ira S. Moskowitz[2] and Gerard Allwein[3]

Center for High Assurance Computer Systems, Code 5540
U.S. Naval Research Laboratory
Washington D.C. 20375

Abstract

We study the algebraic structure of the monoid of binary channels and show that it is dually isomorphic to the interval domain over the unit interval with the operation from [3]. We show that the capacity of a binary channel is Scott continuous as a map on the interval domain and that its restriction to any maximally commutative submonoid of binary channels is an order isomorphism onto the unit interval. These results allows us to solve an important open problem in the analysis of covert channels: a provably correct method for injecting noise into a covert channel which will reduce its capacity to any level desired in such a way that the practitioner is free to insert the noise at any point in the system.

Keywords: information theory, domain theory, covert channel, monoid

1 Introduction

By a *channel* in this paper, we mean a discrete, memoryless untimed channel [5]. In a binary channel, a sender attempts to transmit bits (either a '0' or a '1') to a receiver. However, because of noise, sometimes a '0' arrives as a '1' and conversely. This noise is modelled in information theory by a *noise matrix* which is entirely determined by two probabilities: $a = P(0|0)$, the probability that '0' is received when '0' is sent and $b = P(0|1)$, the probability that '0' is received when '1' is sent. Thus, the noise matrix of a channel can be written

[1] Email: kmartin@itd.nrl.navy.mil
[2] Email: moskowitz@itd.nrl.navy.mil
[3] Email: allwein@itd.nrl.navy.mil

doi:10.1016/j.entcs.2006.04.015

as a pair $(a, b) \in [0, 1]^2$. It seems to have gone unnoticed in the information theory literature that the set of noise matrices form a monoid under matrix multiplication with the identity matrix $(1, 0)$ as the identity of the monoid. But this monoid structure is nevertheless quite interesting. In this paper, we will study it and use our results to develop methods for reducing the threat posed by covert channels.

As we will show, there is no loss of generality in restricting attention to what we call *nonnegative channels* i.e. channels whose noise matrices satisfy $a \geq b$. The resulting monoid $\mathbb{N}$ of nonnegative channels has the beautiful property that it is dually isomorphic to the interval domain $\mathbf{I}[0, 1]$ with the binary operation discovered independently in both [2] and [3]. By interpreting the capacity of a binary channel as a function on $\mathbf{I}[0, 1]$, we discover the 'surprising yet intuitive' result that capacity is Scott continuous. This result alone provides formal justification for several intuitions often used in information theory but whose actual proofs are often either difficult or omitted due to inequalities and formulae that even in the binary case are simply too complex to efficiently manipulate. But this result also allows us to solve an important open problem in the analysis of covert channels.

A covert channel is a channel in which two parties communicate by using certain elements of a system in a way other than they were originally designed for. When a covert channel is discovered within a high assurance device or system, the capacity of that channel is a measure of the threat it poses. The greater the capacity, the greater the threat. Ideally, one would like to simply eliminate a covert channel altogether. However, this is usually not possible since it normally requires degrading system performance to an unacceptable level. Thus, if we encounter a covert channel whose capacity is too high, the most we can hope for in general is a method for reducing its capacity to some level where system performance is preserved but the threat posed by the channel is sufficiently reduced. Specifically, given a covert channel whose capacity needs to be reduced to some lower level r, how can we canonically calculate a noise matrix (a, b) whose injection into the given covert channel reduces the capacity of the original channel to r? Notice that a priori the problem has no canonical solution since we have one equation $\mathrm{C}(a, b) = r$ but two unknowns (a, b). However, the algebraic and domain theoretic structure of noise matrices provides exactly such a method, an 'extra equation' if you like.

First, each nonnegative channel different from the identity lies in a unique maximal commutative submonoid. Though there are infinite number of noise matrices with capacity r, there is *only one* within the maximal commutative submonoid determined by a channel. The reason is that we can show that

the restriction of capacity to the maximal commutative submonoid determined by a channel is an order isomorphism of $[0,1]$. We also give a provably correct algorithm for calculating this unique channel. Because this unique channel commutes with the original, the practitioner then has the option of injecting noise into a covert channel either at the beginning or at the end of a covert channel, granting the reviewers of high assurance devices the maximum amount of freedom when injecting noise into a covert channel.

2 The monoid of noise matrices

The noise matrix M of a binary channel models the effect that noise has on data sent through the channel. If data is sent through the channel according to the distribution x, then the output is distributed as $y = x \cdot M$. This noise matrix is given by

$$M = \begin{pmatrix} a & \bar{a} \\ b & \bar{b} \end{pmatrix}$$

where a is the probability of receiving the first symbol when the first symbol is sent and b is the probability of receiving the first symbol when the second symbol is sent and $\bar{x} := 1 - x$ for $x \in [0,1]$. We denote this noise matrix by (a, b).

The set of noise matrices is described by the unit square $[0,1]^2$. The multiplication of two noise matrices is

$$(a,b) \cdot (\alpha,\beta) = (\ a(\alpha-\beta)+\beta,\ b(\alpha-\beta)+\beta\ ) = \alpha(a,b) + \beta(\bar{a},\bar{b})$$

where the expression to the right uses scalar multiplication and addition of vectors. The identity on $[0,1]^2$ is $1 := (1,0)$. The determinant is a function of type $\det : ([0,1]^2, \cdot) \to ([-1,1], \cdot)$ defines a homomorphism between monoids; happily, we find $\det(a,b) = a - b$ for any noise matrix $(a,b) \in [0,1]^2$.

Definition 2.1 A channel (a,b) is called *positive* when $\det(a,b) > 0$, *negative* when $\det(a,b) < 0$ and a *zero channel* when $\det(a,b) = 0$. A channel is *nonnegative* if it is either positive or zero.

Notice that $\det(a,b) \in (0,1]$ for positive channels, and that $\det(a,b) \in [-1,0)$ for negative channels. The set of positive channels is a submonoid as is the set of nonnegative channels; the determinant function is a homomorphism from the nonnegative channels into $([0,1], \cdot)$. For our purposes, all channels may be assumed nonnegative, as follows. The amount of information that

may be sent through a channel (a, b) is given by its capacity

$$\mathrm{C}(a,b) = \sup_{x\in[0,1]} H((a-b)x+b) - xH(a) - (1-x)H(b)$$

It can be shown [4] that this defines a continuous function on the unit square $[0,1]^2$ given by

$$\begin{aligned}\mathrm{C}(a,b) &= \tfrac{\bar{a}H(b)-\bar{b}H(a)}{a-b} + \log_2\left(1+2^{\frac{H(a)-H(b)}{a-b}}\right)\\ &= \log_2\left(2^{\frac{\bar{a}H(b)-\bar{b}H(a)}{a-b}} + 2^{\frac{bH(a)-aH(b)}{a-b}}\right)\end{aligned}$$

where $\mathrm{C}(a,a) := 0$ and $H(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the base two entropy. But any negative channel may be easily converted into a positive one with the same capacity using the map $\mathrm{twist}(a,b) := (b,a)$:

Lemma 2.2 *The twist map turns negative channels into positive channels and preserves capacity.*

Proof. To show that the twist map preserves capacity requires only the symmetric nature of the untimed capacity $\mathrm{C}(a,b) = \mathrm{C}(b,a)$. □

There are plenty of mappings which map negative channels into positive channels and preserve capacity. What distinguishes the twist map, though, is that it corresponds to the trivial act of renaming 0 to 1 and 1 to 0. For this reason, we can assume without loss of generality that a channel is nonnegative. Thus, we study the monoid of *nonnegative channels.*

Definition 2.3 The monoid of nonnegative channels is denoted $(\mathbb{N}, \cdot, 1)$.

3 Commutativity

Each channel $x = (a,b) \in \mathbb{N}$ different from the identity lies on a unique line which joins the identity 1 to the diagonal. In parametric form, this line $\pi_x : [0,1] \to \mathbb{N}$ is given by

$$\pi_x(t) = (1-t)\cdot(0_x, 0_x) + t\cdot(1,0)$$

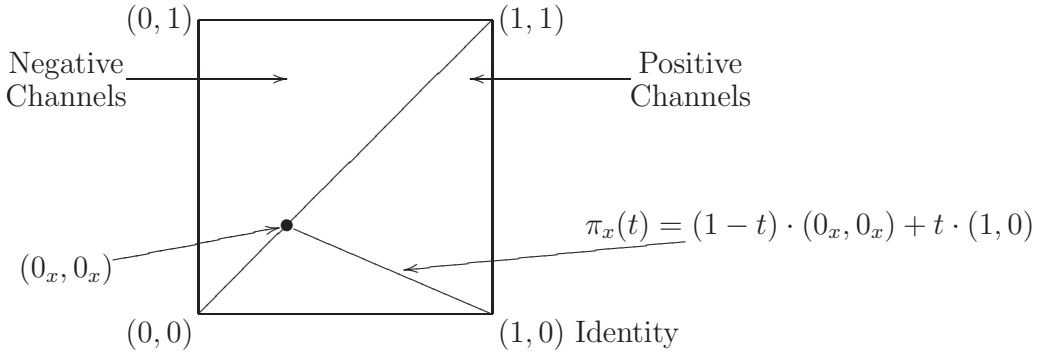
for $t \in [0,1]$, where $(0_x, 0_x)$ is the point on the diagonal given by

$$0_x := \frac{b}{1-\det(x)}$$

Notice that for the particular x used to define 0_x, we have

$$(\pi_x \circ \det)(x) = x$$

so that this line travels from $\pi(0) = (0_x, 0_x)$ to $\pi(a - b) = x$ and then on to the identity $\pi(1) = 1$. Here is a picture of the situation so far:



Definition 3.1 The line through the identity which joins $x \in \mathbb{N} \setminus \{1\}$ to the diagonal is denoted by $\pi_x : [0,1] \to \mathbb{N}$. We define $\pi_1 : [0,1] \to \mathbb{N}$ to be the line which joins the identity to $(1/2, 1/2)$.

Lemma 3.2 *For any $x \in \mathbb{N}$ and $s, t \in [0,1]$, we have $\pi_x(s) \cdot \pi_x(t) = \pi_x(st)$.*

Proof. Let $(a,b) = \pi_x(s) = ((1-s)0_x + s, (1-s)0_x)$ and $(\alpha, \beta) = \pi_x(t) = ((1-t)0_x + t, (1-t)0_x)$. Then

$$\begin{aligned}
\pi_x(s) \cdot \pi_x(t) &= (a,b) \cdot (\alpha, \beta) \\
&= (\ a(\alpha - \beta) + \beta,\ b(\alpha - \beta) + \beta\ ) \\
&= (\ ((1-s)0_x + s) \cdot t + (1-t)0_x, (1-s)0_x \cdot t + (1-t)0_x\ ) \\
&= ((1-st)0_x + st, (1-st)0_x) \\
&= \pi_x(st)
\end{aligned}$$

for any $s, t \in [0,1]$. □

Lemma 3.3 *Two elements $(a,b), (\alpha, \beta) \in \mathbb{N}$ commute iff $\beta\bar{a} = b\bar{\alpha}$.*

Proof. If we write out each product

$$(a,b) \cdot (\alpha, \beta) = (\ a(\alpha - \beta) + \beta,\ b(\alpha - \beta) + \beta\ )$$

$$(\alpha, \beta) \cdot (a,b) = (\ \alpha(a-b) + b,\ \beta(a-b) + b\ )$$

then we can see that (a,b) and (α, β) commute if and only if

$$\beta - \beta a = b - b\alpha \ \text{ and } \ b\alpha + \beta = \beta a + b \quad \Leftrightarrow \quad \beta\bar{a} = b\bar{\alpha} \ \text{ and } \ \beta\bar{a} = b\bar{\alpha}$$

which finishes the proof. □

A commutative submonoid S in a monoid M is *maximal* if for all commutative submonoids T with $S \subseteq T$, we have $S = T$.

Theorem 3.4

(i) *Two channels $x, y \in \mathbb{N}$ commute iff there is a line which passes through x and y and joins the identity to the diagonal.*

(ii) *The maximal commutative submonoids of $\mathbb{N}$ are the lines joining the diagonal to the identity*[4] *.*

(iii) *The determinant is an isomorphism between a maximal commutative submonoid and $([0,1], \cdot)$.*

Proof. (i) Let $x = (a, b)$ and $y = (\alpha, \beta)$. Without loss of generality we can assume that neither is the identity. Then we have

$$\pi_x = \pi_y \;\Leftrightarrow\; 0_x = 0_y \;\Leftrightarrow\; \beta\bar{a} = b\bar{\alpha} \;\Leftrightarrow\; x \cdot y = y \cdot x$$

which means x and y commute iff they are connected by a line joining the identity to the diagonal. Notice that the first equivalence holds because a line is uniquely determined by two points, the second holds by straightforward arithmetic, and the last holds by Lemma 3.3.

(ii) By Lemma 3.2, $\pi_x[0,1]$ is closed under multiplication. By (i), it is commutative submonoid. Let S be a commutative submonoid with $\pi_x[0,1] \subseteq S$. If $y \in S$, then it commutes with x, and so by (i), we must have $y \in \pi_x[0,1]$. Thus, $\pi_x[0,1]$ is a maximal commutative submonoid. Conversely, any commutative submonoid must be contained in some $\pi_x[0,1]$, so one that is maximal must be equal to $\pi_x[0,1]$.

(iii) The determinant maps $\pi_x[0,1]$ surjectively onto $[0,1]$ since $\det(\pi_x(t)) = t$ for all $t \in [0,1]$. However, this also implies it is injective when restricted to $\pi_x[0,1]$:

$$\det(\pi_x(s)) = \det(\pi_x(t)) \Rightarrow s = t \Rightarrow \pi_x(s) = \pi_x(t).$$

This finishes the proof. □

By (ii) then, the maximal commutative submonoids are $\{\pi_x[0,1] : x = (p,p), p \in [0,1]\}$. And by (iii), any two maximal commutative submonoids are isomorphic.

Example 3.5 The Z channels $\{(p,0) : p \in [0,1]\}$ and $\{(1,p) : p \in [0,1]\}$ are maximal commutative submonoids. The binary symmetric channels $\{(\bar{p}, p) : p \in [0, 1/2]\}$ also form a maximal commutative submonoid.

Let us give a purely algebraic characterization of the maximal commutative submonoids of $\mathbb{N}$.

[4] Much more is true. The maximal commutative submonoids are also 'largest' in the sense that they contain any commutative submonoid they intersect at a point other than the identity.

Definition 3.6 A *right zero* in a monoid M is an element 0 such that $x \cdot 0 = 0$ for all $x \in M$. The set of right zero elements in M is denoted $0(M)$.

Proposition 3.7

(i) *The set of right zero elements in* $\mathbb{N}$ *is* $0(\mathbb{N}) = \{(a, b) : \det(a, b) = 0\}$.

(ii) *For any* $x \in \mathbb{N} \setminus \{1\}$, *there is a unique* $0_x \in 0(\mathbb{N})$ *which commutes with* x, *given explicitly by*

$$0_x = \frac{b}{1 - \det(x)}$$

where $x = (a, b)$.

(iii) *The maximal commutative submonoids of* $\mathbb{N}$ *are in bijective correspondence with*

$$\{Z(x) : x \in 0(\mathbb{N})\}$$

where $Z(x) = \{y \in \mathbb{N} : x \cdot y = y \cdot x\}$ *is the set of elements that commute with* x.

Proof. (i) If $\det(\alpha, \beta) = 0$, then $\alpha = \beta$, so for an element $x = (a, b)$, we have

$$x \cdot (\alpha, \beta) = (a, b) \cdot (\alpha, \alpha) = (a \cdot 0 + \alpha, b \cdot 0 + \alpha) = (\alpha, \alpha)$$

which means that $(\alpha, \alpha) \in 0(\mathbb{N})$. Conversely, suppose that $x = (\alpha, \beta) \in 0(\mathbb{N})$, then $(1, 1) \cdot x = x$. But $(1, 1) \cdot x = (\alpha, \alpha)$, which means that $\det(x) = 0$.

(ii) By Theorem 3.4(i), we already know that 0_x commutes with x since it lies on the line π_x which passes through x and joins the identity to the diagonal. If there is another $0_y \in 0(\mathbb{N})$ which commutes with x, then by Theorem 3.4, 0_y also lies on the line π_x, which means that 0_y and 0_x commute. But then

$$0_x = 0_y \cdot 0_x = 0_x \cdot 0_y = 0_y$$

which proves uniqueness. (iii) Immediate. □

Geometrically, the set $0(\mathbb{N})$ is the diagonal in the unit square, while the set $Z(x)$ is the line that passes through x joining the identity to the diagonal. Notice the following beautiful connection between algebra and information theory: a binary channel x has positive capacity iff x is not a right zero element in the monoid of all noise matrices.

4 Domains

Let $(P, \sqsubseteq)$ be a partially ordered set or *poset* [1]. A nonempty subset $S \subseteq P$ is *directed* if $(\forall x, y \in S)(\exists z \in S)\, x, y \sqsubseteq z$. The *supremum* $\bigsqcup S$ of $S \subseteq P$ is the

least of its upper bounds when it exists. A *dcpo* is a poset in which every directed set has a supremum.

For elements x, y of a dcpo D, we write $x \ll y$ iff for every directed subset S with $y \sqsubseteq \bigsqcup S$, we have $x \sqsubseteq s$, for some $s \in S$.

Definition 4.1 Let $(D, \sqsubseteq)$ be a dcpo. We set

- $\Downarrow x := \{y \in D : y \ll x\}$ and $\Uparrow x := \{y \in D : x \ll y\}$
- $\downarrow x := \{y \in D : y \sqsubseteq x\}$ and $\uparrow x := \{y \in D : x \sqsubseteq y\}$

and say D is *continuous* if $\Downarrow x$ is directed with supremum x for each $x \in D$.

Definition 4.2 A *basis* for a domain D is a subset $B \subseteq D$ such that $B \cap \Downarrow x$ is directed with supremum x, for all $x \in D$. A domain is ω*-continuous* if it has a countable basis.

Example 4.3 The collection of compact subintervals of the unit interval

$$\mathbf{I}[0,1] = \{[a,b] : a, b \in [0,1] \ \&\ a \leq b\}$$

ordered under reverse inclusion

$$[a,b] \sqsubseteq [c,d] \Leftrightarrow [c,d] \subseteq [a,b]$$

is an ω-continuous dcpo:

- For directed $S \subseteq \mathbf{I}[0,1]$, $\bigsqcup S = \bigcap S$,
- $I \ll J \Leftrightarrow J \subseteq \mathrm{int}(I)$, and
- $\{[p,q] : p, q \in \mathbb{Q} \cap [0,1] \ \&\ p \leq q\}$ is a countable basis for $\mathbf{I}\mathbb{R}$.

The domain $\mathbf{I}[0,1]$ is called the *interval domain.*

Notice that $\mathrm{int}(I)$ refers to the interior of the interval I in its relative Euclidean topology, so that $\mathrm{int}[a,b] = (a,b)$ for $a > 0$ and $b < 1$, while $\mathrm{int}[0,b] = [0,b)$ for $b < 1$ and $\mathrm{int}[a,1] = (a,1]$ for $a > 0$.

The interval domain $\mathbf{I}[0,1]$ has a natural monoid structure that was discovered independently on at least two separate occasions: by Escardo in [2], while studying integration in real PCF, and by the first author in [3], while studying entropy in quantum mechanics.

Example 4.4 The binary operation on $\mathbf{I}[0,1]$

$$[a,b] \otimes [c,d] = [a + c \cdot (b-a), a + d \cdot (b-a)]$$

is associative, has the unit interval $[0,1]$ as an identity and many other interesting properties. The measurement $\mu[a,b] = b - a$ is a homomorphism from $(\mathbf{I}[0,1], \otimes)$ into $([0,1], \cdot)$.

The connection between $\mathbf{I}[0,1]$ and binary channels is easy to prove but nevertheless eye opening:

Theorem 4.5 *The natural mapping* $\varphi : (\mathbb{N}, \cdot, 1) \to (\mathbf{I}[0,1], \otimes, \bot)$ *given by*

$$\varphi(a,b) = [b,a]$$

is a dual isomorphism of monoids with $\varphi(1) = \bot$ *and* $\mu \circ \varphi = \det$.

Proof. The map φ sends nonnegative channels into compact intervals of $[0,1]$. For $(a,b),(c,d) \in \mathbb{N}$,

$$\begin{aligned}\varphi((a,b)\cdot(c,d)) &= \varphi(\ a(c-d)+d,\ b(c-d)+d\ ) \\ &= [b(c-d)+d, a(c-d)+d] \\ &= [d,c] \otimes [b,a] \\ &= \varphi(c,d) \otimes \varphi(a,b)\end{aligned}$$

and since φ takes 1 to $[0,1]$ and has an inverse given by $\varphi^{-1}[a,b] = (b,a)$, the proof is finished. □

That is, each nonnegative channel determines a unique interval under the twist map φ, multiplication of channel matrices is exactly the dual operation on $\mathbf{I}[0,1]$ and the determinant of a channel matrix is the length of its associated interval. Can we learn anything from this alone?

We think of domains as spaces of informative objects that come, roughly speaking, in two forms: partial and total. The set of *maximal elements* in a domain D is

$$\max(D) := \{x \in D : \uparrow x = \{x\}\}$$

are examples of total elements, while the quintessential example of a partial element in a domain D is its least element, when it exists. The *least element* in a domain D is the unique element $\bot \in D$ with $\bot \sqsubseteq x$ for all $x \in D$. In the case of $\mathbf{I}[0,1]$, the maximal elements are

$$\max(\mathbf{I}[0,1]) = \{[a,a] : a \in [0,1]\}$$

while the least element is $\bot = [0,1]$. In information theoretic terms, the maximal elements are the *zero channels* while the least element is the *noiseless channel.* This makes perfect sense from the viewpoint of the security of a system.

In domain theory, partial elements come in varying degrees of uncertainty, with maximal elements being completely certain, and the least element being maximally uncertain. Imagine a system with a single covert channel in it, perhaps a subsystem sufficiently restricted. Then the capacity of the channel is a measure of how secure the system is i.e. it measures our uncertainty that

information flows in only known ways. From a security standpoint, we are most certain the system is secure when the covert channel is a zero channel; we are least certain the system is secure when the covert channel is $\perp$.

With this connection between $\mathbb{N}$ and $\mathbf{I}[0,1]$ in mind, we then regard the order on $\mathrm{I}[0,1]$ as also defining an order on $\mathbb{N}$:

Definition 4.6 For $(a,b),(c,d) \in \mathbb{N}$,

$$(a,b) \sqsubseteq (c,d) \equiv \varphi(a,b) \sqsubseteq \varphi(c,d)$$

By acknowledging this new aspect of binary channels, their domain theoretic aspect, deeper connections between the disciplines emerge which, as we will see shortly, allow one to solve problems in information theory whose solutions are currently unknown. One such connection is the following: the capacity of a binary channel is a *Scott continuous* function of the noise matrix.

Definition 4.7 The *Scott topology* on a continuous dcpo D has as a basis all sets of the form $\Uparrow x$ for $x \in D$.

Example 4.8 A basic Scott open set in $\mathbf{I}[0,1]$ is

$$\Uparrow [a,b] = \{x \in \mathbf{I}[0,1] : x \subseteq \mathrm{int}([a,b])\}.$$

In $\mathbb{N}$, such a set forms a right triangle whose hypotenuse lies along the diagonal, but whose other two sides are removed.

A function $f : D \to E$ between domains is *Scott continuous* if the inverse image of a Scott open set in E is Scott open in D. This is equivalent [1] to saying that f is *monotone,*

$$(\forall x, y \in D)\, x \sqsubseteq y \Rightarrow f(x) \sqsubseteq f(y),$$

and that it *preserves directed suprema*:

$$f(\bigsqcup S) = \bigsqcup f(S),$$

for all directed $S \subseteq D$.

Before proving the Scott continuity of capacity, let us think about this from the perspective of an information theorist. Suppose we have two binary channels with $(a,b) \sqsubseteq (c,d)$. Then this means that

$$\varphi(a,b) = [b,a] \sqsubseteq \varphi(c,d) = [d,c] \;\Rightarrow\; b \leq d \leq c \leq a$$

¿From $c \leq a$, we can see that the probability of receiving a '0' when it is sent decreases i.e. the probability of a '0' being flipped into a '1' increases; from

$b \leq d$, we can see that the probability of receiving a '0' when a '1' is sent increases i.e. the probability of a '1' being flipped also increases. Thus, no matter what symbol is sent, there is a greater chance that it will be flipped if it is sent through the channel (c, d) than if it is sent through the channel (a, b). To put it another way, the channel (c, d) has more noise in it than (a, b). For this reason, we expect that it has less capacity. In symbols,

$$\mathrm{C}(a, b) \geq \mathrm{C}(c, d)$$

which is to say that we expect the function C to be monotone as a function from $\mathbf{I}[0, 1]$ into the unit interval with its dual order $([0, 1]^*, \sqsubseteq)$ i.e. the order $(\forall x, y \in [0, 1])\, x \sqsubseteq y \equiv y \leq x$. However, because C is already known to be Euclidean continuous, we are asserting exactly that capacity is a Scott continuous function from the domain $\mathbf{I}[0, 1]$ to the domain $[0, 1]^*$.

Theorem 4.9 *The capacity* $\mathrm{C} : \mathbf{I}[0, 1] \to [0, 1]^*$ *is Scott continuous and strictly monotone:*

$$x \sqsubseteq y \;\&\; \mathrm{C}(x) = \mathrm{C}(y) \Rightarrow x = y$$

for all $x, y \in \mathbf{I}[0, 1]$.

Proof. Because capacity is Euclidean continuous, we only have to prove its monotonicity into $[0, 1]$ with the dual order. First consider the base e mutual information of a positive channel (a, b)

$$I(x, a, b) = h((a - b)x + b) - xh(a) - (1 - x)h(b),$$

where x is the probability that the first input symbol is sent. Assume that $a, b, x \in (0, 1)$. Then $(a - b)x + b \in (0, 1)$ since $0 < b < (a - b)x + b < a < 1$. The partial derivatives of I with respect to a and b are

$$\frac{\partial I}{\partial a} = x \cdot \left(\ln \left(\frac{1}{(a - b)x + b} - 1 \right) - \ln \left(\frac{1}{a} - 1 \right) \right) > 0$$

$$\frac{\partial I}{\partial b} = (1 - x) \cdot \left(\ln \left(\frac{1}{(a - b)x + b} - 1 \right) - \ln \left(\frac{1}{b} - 1 \right) \right) < 0$$

Lastly, let us point out that these results also hold when $a = 1$ and $b = 0$. For $a = 1$,

$$\frac{\partial I}{\partial b} = (1 - x) \cdot \left(\ln \left(\frac{1}{(1 - b)x + b} - 1 \right) - \ln \left(\frac{1}{b} - 1 \right) \right) < 0$$

and for $b = 0$,

$$\frac{\partial I}{\partial a} = x \cdot \left(\ln \left(\frac{1}{ax} - 1 \right) - \ln \left(\frac{1}{a} - 1 \right) \right) > 0$$

Now suppose $[b, a] \sqsubseteq [d, c]$. If $b = 1$, $a = 0$ or $a = b$, there is nothing to prove, since the two are equal. Then $a \in (0, 1]$, $b \in [0, 1)$, $a \neq b$. But since $[b, a]$ must now be a channel with positive capacity, we can also assume $d \in [0, 1)$, $c \in (0, 1]$ and $c \neq d$. Let $x \in (0, 1)$. If $a = 1$ and $b = 0$, there is nothing to prove, so one of these must fail. Assume $a < 1$.

Then $a \in (0, 1)$. For fixed $b \in [0, 1)$, we know that $I(x, a, b)$ increases with a. Then since $0 < c \leq a < 1$,

$$I(x, a, b) \geq I(x, c, b)$$

while for fixed $c \in (0, 1]$, $I(x, c, b)$ decreases with increasing b, so from $0 < b \leq d < 1$ we have

$$I(x, c, b) \geq I(x, c, d)$$

Further, because $[b, a] \neq [d, c]$, one of these inequalities must be strict, proving $I(x, a, b) > I(x, c, d)$. Since $c \neq d$, $[d, c]$ is a positive capacity channel, there is a unique $x^* \in (0, 1)$ with $\mathrm{C}[d, c] = I(x^*, c, d)$. This gives $C[b, a] \geq I(x^*, a, b) > C[d, c]$.

Suppose instead that $b > 0$ and $a \in (0, 1]$. Then consider the channels $x = [1 - a, 1 - b]$ and $y = [1 - c, 1 - d]$ which satisfy $x \sqsubseteq y$. Since $b > 0$, $1 - b < 1$, so the result just proven applies to give

$$\mathrm{C}[b, a] = \mathrm{C}(x) > \mathrm{C}(y) = C[d, c]$$

which finishes the proof. □

Two channels contained in a commutative submonoid must lie on a line that travels from the diagonal to the identity. But we can get from the point closest to the diagonal to the point closest the identity, by first moving right (which means a increases) and then moving down (which means b decreases). By the last theorem then, capacity can only increase during such a motion. Here is the formal proof, which reveals something quite surprising: within the commutative submonoid determined by a channel, the determinant is an isomorphism that qualitatively reflects capacity.

Proposition 4.10 *Let π be a maximal commutative submonoid of $\mathbb{N}$. Then*

$$x \sqsubseteq y \;\Leftrightarrow\; \det(x) \geq \det(y) \;\Leftrightarrow\; \mathrm{C}(x) \geq \mathrm{C}(y)$$

for any $x, y \in \pi$.

Proof. Let $\pi(t)$ be a line from the identity to a zero channel $(\alpha, \alpha) \in 0(\mathbb{N})$, given by

$$\pi(t) = (t\alpha + \bar{t}, t\alpha)$$

First notice that $(\forall s, t \in [0,1])\ s \leq t \Leftrightarrow \pi(s) \sqsubseteq \pi(t)$ as follows:

$$\begin{aligned}
\pi(s) \sqsubseteq \pi(t) &\Leftrightarrow [s\alpha, s\alpha + \bar{s}] \sqsubseteq [t\alpha, t\alpha + \bar{t}] \\
&\Leftrightarrow (s\alpha \leq t\alpha)\ \&\ (t\alpha + \bar{t} \leq s\alpha + \bar{s}) \\
&\Leftrightarrow (s \leq t\ \vee\ \alpha = 0)\ \&\ (s - t)(1 - \alpha) \leq 0 \\
&\Leftrightarrow (s \leq t\ \vee\ \alpha = 0)\ \&\ (s \leq t\ \vee\ \alpha = 1) \\
&\Leftrightarrow s \leq t
\end{aligned}$$

In particular, π is injective.

Now let us prove that $x \sqsubseteq y \Leftrightarrow \mathrm{C}(y) \leq \mathrm{C}(x)$. The direction $(\Rightarrow)$ is clear from Theorem 4.9. Assume that $\mathrm{C}(y) \leq \mathrm{C}(x)$. Since x and y belong to a maximal commutative submonoid, $x = \pi(s)$ and $y = \pi(t)$. If $t < s$, then $x = \pi(t) \sqsubset \pi(s) = y$ by the injectivity of π. By Theorem 4.9, $\mathrm{C}(x) > \mathrm{C}(y)$, which is a contradiction. Thus, $s \leq t$, which gives $y \sqsubseteq x$.

To prove $x \sqsubseteq y \Leftrightarrow \det(y) \leq \det(x)$, we consider the $(\Leftarrow)$ direction since the other is clear. Since x and y belong to a maximal commutative submonoid, $x = \pi(s)$ and $y = \pi(t)$, and since $\det(\pi(t)) = t = \det(y) \leq \det(x) = s = \det(\pi(s))$, we have $y = \pi(t) \sqsubseteq \pi(s) = x$ by the monotonicity of π. □

The qualitative equivalence between length and capacity on a commutative submonoid of $\mathbb{N}$ is established using order theoretic techniques. We are not aware of a direct proof of this fact based on inequalities.

Corollary 4.11 $\mathrm{C} \circ \pi_x : [0,1] \to [0,1]$ *is an order isomorphism.*

Proof. The line π_x travels from the diagonal to the identity, instead of from the identity to the diagonal, which reverses the inequalities in Prop. 4.10. Thus, $s \leq t \Leftrightarrow \mathrm{C}(\pi_x(s)) \leq \mathrm{C}(\pi_x(t))$. □

There are some surprising applications of the results in this section.

Example 4.12 Consider two binary channels with respective noise matrices

$$A = \begin{pmatrix} 7/16 & 9/16 \\ 3/16 & 13/16 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 11/32 & 21/32 \\ 7/32 & 25/32 \end{pmatrix}$$

Which has larger capacity? We can answer this question without calculating the capacity of either.

One way is to notice that $A = (14/32, 6/32)$ while $B = (11/32, 7/32)$. Since $A \sqsubseteq B$, we have $\mathrm{C}(A) > \mathrm{C}(B)$, which means A has larger capacity. Another way in this case is to use the determinant.

By Lemma 3.3, these two noise matrices commute, so by Prop. 4.10, we can determine the channel with larger capacity by simply comparing determinants:

$$\det(A) = 4/16 = 1/4 > 1/8 = 4/32 = \det(B)$$

so again we see that A is the channel with larger capacity.

One could easily look at this example and intuitively reason that A has larger capacity than B since $P(0|0) = 14/32$ and $P(1|1) = 26/32$ for A while $P(0|0) = 11/32$ and $P(1|1) = 25/32$ for B. But that kind of intuitive reasoning exactly amounts to the statement that $A \sqsubseteq B$. The question is: why does it then follow that $\mathrm{C}(A) > \mathrm{C}(B)$? The answer is that capacity is Scott continuous and strictly monotone.

Example 4.13 Each channel (a, b) is the product of Z channels

$$(a, b) = (1, b/a) \cdot (a, 0)$$

when $a \neq 0$ and

$$(0, 0) = (1, 0) \cdot (0, 0)$$

when $a = 0 \geq b \geq 0$. Let us suppose that $a > 0$. Then because $x \sqsubseteq x \otimes y$ on $\mathbf{I}[0, 1]$ and the capacity C on $\mathbf{I}[0, 1]$ is Scott continuous, we have

$$\mathrm{C}(\varphi(a, 0)) \geq \mathrm{C}(\varphi(a, 0) \otimes \varphi(1, b/a)) = \mathrm{C}(\varphi(a, b))$$

This result is in accord with intuition. However, if we look at it in terms of inequalities, we see that we have just proven

$$\mathrm{C}(a, 0) = \log_2\left(1 + 2^{-\frac{H(a)}{a}}\right) \geq \log_2\left(2^{\frac{\bar{a}H(b) - \bar{b}H(a)}{a-b}} + 2^{\frac{bH(a) - aH(b)}{a-b}}\right) = \mathrm{C}(a, b)$$

We are not aware of an analytic proof of this result. However, the domain theoretic techniques developed in this paper allow one to give a simple proof of a result whose proof should be simple. In short, the Scott continuity of capacity formally justifies a number of valid intuitions one often relies on when reasoning about binary channels.

We have stressed in this section how domain theory benefits information theory. It is worth pointing out, though, that information theory also offers some interesting interpretations of domain theoretic ideas. The monoid $(\mathbf{I}[0, 1], \otimes, \bot)$ is a continuous version of the domain of bit streams $(\Sigma^\infty, \cdot, \varepsilon)$ with concatenation and the empty string as identity. Let us make this explicit: the map

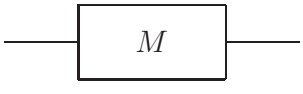
$$\phi(\varepsilon) = \bot, \quad \phi(0) = [0, 1/2], \quad \phi(1) = [1/2, 1]$$

extends to a Scott continuous homomorphism. Its restriction to the set of maximal elements is a continuous surjection from the Cantor set (a Stone space) to the unit interval. In fact, this mapping is a very important numerical method in disguise: the bisection method arises from repeatedly multiplying the two elements $\text{left}(\perp) = [0, 1/2]$ and $\text{right}(\perp) = [1/2, 1]$. But notice what this says: if concatenation of strings is multiplication of intervals, but multiplication of intervals is matrix multiplication, then concatenation of binary strings can be understood as matrix multiplication. Even better: each finite string is represented by an invertible matrix!

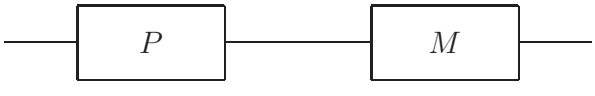
5 The injection of noise into a covert channel

Increasing the amount of noise in a covert channel reduces its capacity. Shannon [5] has shown that one cannot transmit at a rate greater than capacity. Thus, the ability to lower the capacity of a covert channel provides a method for reducing the threat it poses. The algebraic and domain theoretic structure of binary channels provides an elegant solution of the capacity reduction problem as follows.

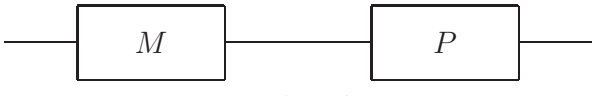
Suppose we have a covert channel with noise matrix M whose capacity we would like to reduce to r:



To do so, we introduce a new 'component' into the system with noise matrix P. There are two places we can introduce the noise, either



or



The capacity of the first system is $\mathrm{C}(PM)$ and the capacity of the second is $\mathrm{C}(MP)$. However, in some cases, it may not be physically possible to introduce P at certain points in the system. For instance, if the receiver at the end of channel M is an eavesdropper, then because we cannot necessarily know their exact location or even that they exist at all, the system MP is impossible to reliably build; the only possibility is PM. On the other hand, in cases where it *is* possible to introduce noise at any point in the system, we would like the practitioner to have the freedom of choosing the most inexpensive way

possible. Both of these problems can be solved if it is possible to find a matrix P which commutes with M such that $\mathrm{C}(MP) = \mathrm{C}(PM) = r$.

The value of M and P commuting is that we are free to insert this new component either before the channel M or after the channel M since both of the modifications above are then statistically identical. Because of this, the practitioner is also free to place the component in the most inexpensive way possible thereby avoiding the impossible situation of having a noise matrix that must be inserted at the beginning of the system despite the fact that capacity r is only achieved by placing P at the end of the system.

We now use the algebraic and domain theoretic techniques developed in the previous sections to prove that such a matrix exists *uniquely* and how to compute it:

Theorem 5.1 *Let $x \in \mathbb{N} \setminus \{1\}$ be a positive channel* [5] *and $0 \leq r \leq \mathrm{C}(x)$.*

(i) *There is a unique y which commutes with x such that $\mathrm{C}(xy) = \mathrm{C}(yx) = r$.*

(ii) *If $\pi_x[0,1]$ is the maximal commutative submonoid joining 0_x to 1, then $f : [0,1] \to \mathbb{R}$ given by $f(t) = \mathrm{C}(\pi_x(t)) - r$ is continuous and changes sign on $[0, \det(x)]$.*

Proof. (i) Let $\pi_x : [0,1] \to \mathbb{N}$ be the line through x which joins the identity to the diagonal. By Corollary 4.11, the function $\mathrm{C} \circ \pi_x : [0,1] \to [0,1]$ is strictly increasing, i.e.

$$(\forall s, t \in [0,1])\ s < t \Leftrightarrow \mathrm{C}(\pi_x(s)) < \mathrm{C}(\pi_x(t)).$$

Because it is Euclidean continuous with a connected domain, it has to surjective, since it must assume all values between $\mathrm{C}(\pi_x(0)) = 0$ and $\mathrm{C}(\pi_x(1))) = 1$. However, because it is strictly increasing, it is also injective.

Let $x = \pi_x(t)$ and $z = \pi_x(s)$ be the unique channel in $\pi_x[0,1]$ with $\mathrm{C}(z) = r$. Then $0 \leq s \leq t$ since $r \leq \mathrm{C}(x)$. Since $t > 0$, we can define $y = \pi_x(s/t)$. Then $xy = yx$ and we have

$$\mathrm{C}(yx) = \mathrm{C}(\pi_x(s/t) \cdot \pi_x(t)) = \mathrm{C}(\pi_x((s/t) \cdot t))) = \mathrm{C}(\pi_x(s)) = r$$

For the uniqueness of y, let $u = \pi_x(v)$ be another such element, then

$$\mathrm{C}(ux) = \mathrm{C}(\pi_x(v) \cdot \pi_x(t)) = \mathrm{C}(\pi_x(vt)) = r$$

and by the injectivity of $\mathrm{C} \circ \pi_x$, $vt = s$, which means $v = s/t$, and hence that $u = y$.

[5] By Theorem 4.9, the positive channels are exactly the nonnegative channels which have positive capacity.

(ii) This is immediate since $f(0) = \mathrm{C}(0_x) - r = 0 - r \leq 0$ and the fact that $f(\det(x)) = \mathrm{C}(x) - r \geq 0$. □

We now have an algorithm for reducing the capacity of a covert channel $M = x$ to any level r desired. First, we solve the equation $f(t) = 0$. The noise matrix $\pi_x(t)$ then has capacity r. If it is possible to replace M with $\pi_x(t)$ in the system, then we can do so, and the remaining covert channel will have capacity r. On the other hand, if we can only add components to the existing system in order to reduce capacity, then adding the component $P = M^{-1}\pi_x(t)$ to either side of M will reduce the covert channel's capacity to r, where M^{-1} exists because $\det(M) > 0$.

Notice that the algorithm also applies to the identity channel, provided we also choose a particular commutative submonoid ('path') along which we want to achieve capacity r. A canonical choice of path seems to be the line joining 1 to $(1/2, 1/2)$, the commutative submonoid of binary symmetric channels. The function f always changes sign on $[0, 1]$.

6 Future Work

We would like to extend our results to binary *timing* channels. Only recently has the capacity problem for binary timing channels been solved [4]; we are optimistic about the adaptability of the algebraic and domain theoretic techniques introduced here to the timed setting.

The measurement property is particularly interesting in this setting. Recall that often measurements are used to help us determine the degree to which a given element in a domain approximates a maximal element. That is precisely what we are doing when we attempt to reduce the capacity of a covert channel x: we are viewing x as an approximation of a zero channel (α, α) and we are saying that we would like to compute a 'better' approximation of the maximal element (α, α).

There are other approaches one can take to reduce the capacity of a covert channel. For instance, follow the gradient on the surface $(a, b, \mathrm{C}(a, b))$. While this will certainly reduce capacity 'more rapidly,' it has the unpragmatic effect of restricting the practitioner's ability to introduce noise into a system, since the matrix obtained in general will not commute with the matrix of the channel. Nevertheless, there may be other uses for such a method. This is something we are interested in.

Finally, while our emphasis here has been largely pragmatic in nature, wanting to know how to reduce capacity in an algorithmic way that practitioners can use, it has not escaped the attention of the authors that $\mathbf{I}[0, 1]$ is a compact monoid with a number of interesting properties that relate order

and an algebra in an intensely exciting way. It would be good to discover the properties that monoids like $\mathbf{I}[0,1]$ and Σ^∞ have so that many of the results in this paper can be proven abstractly. For instance, it should follow from axioms of domain theoretic monoids that commutative elements always compare, or that the Lawson topology on a domain theoretic monoid is always compact.

7 Acknowledgement

The first author thanks the National Academy of Sciences and all three authors thank the Naval Research Laboratory.

We thank the first two referees for valuable suggestions that improved the quality of this paper and especially for their encouraging remarks. Finally, we were amused by the third referee, who wrote "Your use of $\mathbb{N}$ is just brilliant; it was high time that this underused symbol got overloaded." We are quite pleased that you think so highly of our $\mathbb{N}$otation.

References

[1] S. Abramsky and A. Jung. *Domain theory.* In S. Abramsky, D. M. Gabbay, T. S. E. Maibaum, editors, Handbook of Logic in Computer Science, vol. III. Oxford University Press, 1994.

[2] A. Edalat and M. Escardo. *Integration in real PCF.* Information and Computation, Volume 160, Numbers 1-2, p. 128–166, 2000.

[3] K. Martin. *Entropy as a fixed point.* ICALP 2004. Theoretical Computer Science, Volume 350, Issues 2-3, p. 292–324, 2006.

[4] K. Martin and I. S. Moskowitz. *Noisy timing channels with binary inputs and outputs.* Information Hiding, Lecture Notes in Computer Science, Springer-Verlag, 2006, to appear.

[5] C. E. Shannon. *A mathematical theory of communication.* Bell Systems Technical Journal, 27:379–423, 623–656, 1948.