Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in Theoretical Computer Science 218 (2008) 111–129

Electronic Notes in Theoretical Computer Science

www.elsevier.com/locate/entcs

A Monotonicity Principle for Information Theory

Konstantinos Chatzikokolakis[1]

Oxford Computing Lab and LIX École Polytechnique

Keye Martin[2]

Naval Research Laboratory

Abstract

We establish a monotonicity principle for convex functions that enables high-level reasoning about capacity in information theory. Despite its simplicity, this single idea is remarkably applicable. It leads to a significant extension of algebraic information theory, a solution of the capacity reduction problem, intuitive graphical methods for comparing channels, new inequalities that provide useful estimates on the information transmitting capability of a channel operating in an unknown environment, further explication of the fascinating relationship between capacity and Euclidean distance, and the solution of an open problem in quantum steganography.

Keywords: binary channels, capacity, euclidean distance, information theory, quantum steganography

1 Introduction

A notion of particular interest in the field of information theory is that of a *channel*. A channel is a device that transforms symbols of an input alphabet $\mathcal{X}$ to symbols of an output alphabet $\mathcal{Y}$ in a probabilistic way: when $x \in \mathcal{X}$ is the input, $p(y|x)$ gives the probability of getting $y \in \mathcal{Y}$ in the output. For such channels, *capacity* is a measure of the maximum correlation between the input and the output. It is 0 when they are independent (all inputs produce the same output with the same probability) and takes its maximum value when there are no transmission errors (each output is produced by exactly one input). The importance of capacity rises from Shannon's theorem: capacity gives the maximum achievable rate at which

[1] kostas@lix.polytechnique.fr

[2] kmartin@itd.nrl.navy.mil

doi:10.1016/j.entcs.2008.10.008

we can transmit information using the channel, with arbitrarily low probability of error.

Apart from their use in information theory, the notions of channel and capacity have been proven quite useful in the area of security. It has been shown that in many scenarios, systems or protocols can be fruitfully viewed as channels, and the capacity of these channels can be regarded as a measure of the security guarantees of the system. Techniques from information theory have been applied to a broad range of security fields, including those of information flow ([12,4]), quantum cryptography ([15]), anonymity ([13,3]) and trust ([14]).

However, an important drawback of capacity is its complexity. Despite its simple definition, there is no analytical formula that gives the capacity of a discrete channel in the general case. It can be only computed approximately using numerical algorithms such as the Arimoto-Blahut algorithm ([5]). And even in simple cases where an analytical formula does exist, for example in the case of *binary channels* having only two inputs and outputs, it is complicated and difficult to use in practice. For example, in many problems we need to be able to *predict* how a channel will perform, but find that its noise matrix varies with several parameters that depend on random aspects of the environment which arise *during* the transmission. In such cases we cannot compute the capacity of the channel, but we would still like to obtain bounds on it or compare the performance of different classes of channels. The resulting formulae, however, make this goal very challenging.

It is thus natural to seek tools that allow high-level reasoning about capacity. Developed in a recent line of work, *algebraic information theory* ([11]) offers such tools for binary channels. In that work, studying the relation of order, algebra and topology, a domain of binary channels is considered and it is shown that capacity is Scott-continuous, providing a tool to compare channels. Moreover, in [9] it is shown that capacity is a measurement on this domain.

In this paper we exploit convexity, a property of capacity that has been in general underused in the literature, but which turns out to be a fundamental property on which a lot of results can be based. Convexity provides us with a monotonicity principle that we use to give simpler and more general proofs of results from [11,9], as well as numerous news ones, outlined in the next section.

Contribution

We establish a monotonicity principle for convex functions: a convex function decreases on a line segment iff it assumes its minimum value at the end of that line segment. Though quite simple, this single idea has an unusual number of important consequences for information theory and the areas which benefit from it (e.g. information hiding, security, quantum information).

The first of these it that it offers a significant extension of algebraic information theory: a new partial order is introduced on binary channels with respect to which capacity is monotone. This new order is much larger than the interval order considered in [11], and can be characterized in at least three different ways, each of which has its own value: by means of a simple formula, which makes it easy to apply in practice; geometrically, which makes it easy to understand and reason about; and algebraically, which establishes its canonical nature, mathematically speaking.

These results provide graphical methods for reasoning about the capacity of channels, which are of value to practitioners in information theory and security. The mathematics of information theory often prevent reviewers of high assurance devices from using it. The graphical methods we introduce avoid this problem. There is a 'geometry of binary channels', in which, roughly speaking, a line of channels either hits the diagonal, or is parallel to it. We determine the behavior of capacity in both these cases, which allows one to answer most (but not all) questions when it comes to comparing channel behavior.

Another use of the monotonicity principle is in establishing inequalities relating different measurements on the domain of channels. These inequalities have several uses. As already explained, the channel matrix often depends on external parameters. While determining these parameters precisely in advance is usually not possible, one can surprisingly often obtain useful bounds on them. Thus, we need methods for estimating the capacity of a channel given only *partial information* about its noise matrix. Specifically, methods that provide estimates of capacity from estimates of a channel's noise matrix which themselves are derived from estimates of underlying experimental parameters. Our inequalities provide exactly these kinds of methods.

As a case in point, we derive a lower bound on the capacity of a hidden channel within quantum key distribution in the presence of noise. Previously, a bound was known ([8]) only for the case where the noise is due solely to eavesedropping, in which the noise matrix has a simple symmetric form. We establish a fundamental theoretical limit of $1 - H(1/4) \approx 0.18$ for the case of arbitrary noise caused by any combination of eavesdropping and/or environment. Our results also make it clear that there is a best way to interrupt this hidden communication, and they even tell us what this way is.

Note that, even though most of these results are limited to binary channels, the monotonicity principle itself holds in general.

The proofs of all results can be found in the report version of this paper ([2]).

2 Channels

Nearly all the results in this paper concern binary channels, so we devote the majority of this section to discussing their specifics. A binary channel has two inputs ("0" and "1") and two outputs ("0" and "1"). An input is sent through the channel to a receiver. Because of noise in the channel, what arrives may not necessarily be what the sender intended. The effect of noise on input data is modeled by a noise matrix u. If data is sent through the channel according to the distribution x, then the output is distributed as $y = x \cdot u$. The noise matrix u is given by

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

where $a = P(0|0)$ is the probability of receiving 0 when 0 is sent and $b = P(0|1)$ is the probability of receiving 0 when 1 is sent and $\bar{x} := 1 - x$ for $x \in [0,1]$.

Thus, the noise matrix of a binary channel can be represented by a point (a,b) in the unit square $[0,1]^2$ and all points in the unit square represent the noise matrix of some binary channel.

The *composition* of two binary channels x and y is the channel whose noise matrix is the usual product of matrices $x \cdot y = xy$. The multiplication of two noise matrices $x = (a,b)$ and $y = (c,d)$ in the unit square representation is

$$(a,b) \cdot (c,d) = (\ a(c-d) + d,\ b(c-d) + d\ ) = c(a,b) + d(\bar{a},\bar{b})$$

where the expression to the right uses scalar multiplication and addition of vectors. By contrast, the representation for the sum of noise matrices is simply the sum of each representing vector.[3]

A *monoid* is a set with an associative binary operation that has an identity. The set of binary channels is a monoid under the operation of multiplication whose identity is the noiseless channel $1 := (1,0)$. A binary channel can be classified according to the sign of its determinant, $\det(a,b) = a - b$, which defines a homomorphism $\det : ([0,1]^2, \cdot) \to ([-1,1], \cdot)$ between monoids.

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

Notice that $\det(x) \in (0,1]$ for positive channels, and that $\det(x) \in [-1,0)$ for negative channels. Thus, the set of positive channels is a submonoid of $[0,1]^2$ as is the set of nonnegative channels; the determinant is a homomorphism from the nonnegative channels into $([0,1], \cdot)$.

Notation 2.2 The set of nonnegative binary channels is denoted $\mathbb{N}$. The set of positive binary channels is denoted $\mathbb{P}$.

A nice property of positive channels is that composition can be inverted (even though the "inverse" of a channel is not a channel).

Lemma 2.3 *For $a \in \mathbb{P}$, $x, y \in \mathbb{N}$ we have $ax = ay$ iff $x = y$ iff $xa = ya$.*

The amount of information that may be sent through a channel (a,b) is given

[3] More specifically we mean the convex sum $tM_1 + \bar{t}M_2, t \in [0,1]$ of noise matrices M_1, M_2 which itself is a noise matrix

by its capacity:

$$c(a,b) = \log_2 \left(2^{\frac{\bar{a}H(b)-\bar{b}H(a)}{a-b}} + 2^{\frac{bH(a)-aH(b)}{a-b}} \right)$$

where $H(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the base two entropy. Capacity is continuous on the unit square ([9]).

The general case: (m,n) channels.

In the general case, we have m inputs and n outputs, so the noise matrix u has m rows and n columns, each entry is a probability, and as before, each row sums to one. Noise matrices for (m,n) channels are closed under finite convex sums, but are only closed under composition when the associated matrix multiplication is defined.

If we write the noise matrix u of a channel as a list of rows $u = (u_1, \ldots, u_m)$ and p is a distribution over the channel's inputs, then the *mutual information* between the input and output of channel is

$$I_p(u) = H(p \cdot u) - \sum_{i=1}^{m} p_i H(u_i).$$

The capacity of the channel u is then $c(u) = \sup_p I_p(u)$.

3 The monotonicity principle

We will denote $1-t$ by $\bar{t}$. A subset S of a vector space is *convex* iff $tx_1 + \bar{t}x_2 \in S$ for all $x_1, x_2 \in S, t \in [0,1]$. A function $f : S \to \mathbb{R}$ is *convex* iff

$$tf(x_1) + \bar{t}f(x_2) \geq f(tx_1 + \bar{t}x_2) \quad \forall x_1, x_2 \in S, \forall t \in [0,1]$$

A function f is *strictly convex* if the equality (in the above inequality) holds iff $x_1 = x_2$ or $t \in \{0,1\}$. We now come to *the monotonicity principle*: a convex function decreases along a line segment iff it assumes its minimum value at the end of that line segment.

Theorem 3.1 *If S is a set of vectors, $x, y \in S$, $\pi(t) = ty + \bar{t}x$ is the line from x to y and $c : S \to \mathbb{R}$ is a function (strictly) convex on $\pi[0,1]$, then the following are equivalent:*

(i) *The function $c \circ \pi : [0,1] \to \mathbb{R}$ is (strictly) monotone decreasing,*

(ii) *The minimum value of $c \circ \pi$ on $[0,1]$ is $c(\pi(1)) = c(y)$.*

It is by no means obvious that the monotonicity principle is of any value in problem solving. However, as we will see shortly, there are many situations in information theory where it is far easier to establish a minimum value along a line than it is to establish monotonicity itself.

It is well-known ([5]) that mutual information I_p is a convex function of u for a fixed p. An important consequence of this, first observed by Shannon ([16]), though not particularly well-known, is that capacity itself is convex:

Theorem 3.2 *Capacity $c(u)$ is a convex function of u.*

In the case of binary channels we make the previous result more precise by showing that the capacity is *strictly* convex everywhere, except on the zero channels.

Theorem 3.3 *The capacity on binary channels is strictly convex everywhere except on the zero channels. That is, given $u_1, u_2 \in [0,1]^2$, $u_1 \neq u_2$ and $t \in (0,1)$, we have $c(tu_1 + \bar{t}u_2) \leq tc(u_1) + \bar{t}c(u_2)$, with equality if and only if both u_1, u_2 are zero channels.*

Because Theorem 3.1 can be applied to *any* line that ends on a minimum capacity channel, it provides a powerful technique for comparing the capacity of channels. One immediate application of it is that we can solve the *capacity reduction problem* for arbitrary (m, n) channels. In the capacity reduction problem, we have an (m, n) channel x and would like to systematically obtain a channel whose capacity is smaller by some specified amount. The monotonicity principle offers a solution:

Proposition 3.4 *Let x be any (m, n) channel, y be any (m, n) channel with zero capacity and π denote the line from x to y. Then $c(\pi[0,1]) = [0, c(x)]$ and the function $c \circ \pi$ is monotone decreasing.*

Thus, given any $0 < r < c(x)$, we need only solve the equation $c(\pi(t)) = r$ for t. This equation can be solved numerically since $c \circ \pi - r$ changes sign on $[0,1]$. Notice that this enables us to systematically solve a problem that otherwise would have $m(n-1)$ unknowns but only a single equation. Moreover, the channel obtained is a linear degradation of the original. Similarly, we can systematically increase the capacity using the line from x to a maximum capacity channel. We now turn to another use of the monotonicity principle.

4 Relations between channels

In this section, we consider partial orders on binary channels with respect to which capacity is monotone. Their importance stems from the fact that a statement like "$x \leq y$" is much easier to verify than a statement like "$c(x) \leq c(y)$". In situations where the noise matrix of a channel is only partially specified, by means of bounds on experimental parameters for instance, or where it is known but varies with a parameter like time or the probability of losing a photon [9], their usefulness is especially apparent.

4.1 Algebraic information theory

Algebraic information theory uses the interplay of order, algebra and topology to study communication.

Definition 4.1 The *interval domain* is the set of nonnegative binary channels $(\mathbb{N}, \sqsubseteq)$ together with the relation $\sqsubseteq$ defined by $x \sqsubseteq y$ iff $b \le d$ and $c \le a$, for $x = (a, b), y = (c, d) \in \mathbb{N}$. The natural measurement $\mu : \mathbb{N} \to [0, 1]^*$ is given by

$$\mu x = \det(x) = a - b \quad x \in \mathbb{N}$$

This is not the usual notation in domain theory for the interval domain, but experience has taught us that this is the simplest way of handling things in the context of information theory. The following result is proven in [11]:

Theorem 4.2 *Let $(\mathbb{N}, \cdot)$ denote the monoid of nonnegative channels. The right zero elements of $\mathbb{N}$ are precisely the zero channels. The maximally commutative submonoids of $\mathbb{N}$ are precisely the lines which join the identity to a zero channel. For any maximally commutative submonoid $\pi \subseteq \mathbb{N}$:*

$$(\forall x, y \in \pi)\ x \sqsubseteq y \ \Leftrightarrow\ \mu x \ge \mu y \ \Leftrightarrow\ cx \ge cy$$

Capacity on $\mathbb{N}$ is monotone: $x \sqsubseteq y \Rightarrow c(x) \ge c(y)$.

We will now see that the monotonicity principle offers a new order $\le$ on channels that leads to a clear and significant extension of algebraic information theory.

4.2 A new partial order on binary channels

By the monotonicity principle, capacity decreases along any line that ends on a zero capacity channel. This suggests a new way of ordering positive channels:

Definition 4.3 *Let $x = (a, b)$ and $y = (c, d) \in \mathbb{P}$,*

$$x \le y \ \equiv\ c \cdot \mu x \ge a \cdot \mu y \quad \textit{and} \quad \bar{c} \cdot \mu x \ge \bar{a} \cdot \mu y$$

Recall that a *partial order* on a set is a relation which is reflexive, transitive and antisymmetric.

Proposition 4.4

(i) *The relation $\le$ is a partial order on the set $\mathbb{P}$ of positive channels,*

(ii) *For $x, y \in \mathbb{P}$, if $x \sqsubseteq y$, then $x \le y$. In particular, the least element of $(\mathbb{P}, \le)$ is the identity channel $\perp = (1, 0)$,*

(iii) *For $x, y \in \mathbb{P}$, we have $x \le y$ iff there is a line segment that begins at x, passes through y and ends at some point of $\{(t, t) : t \in [0, 1]\}$,*

(iv) *Capacity $c : \mathbb{P} \to [0, 1]^*$ is strictly monotone: if $x \le y$, then $c(x) \ge c(y)$ with equality iff $x = y$.*

Notice that the monotonicity of capacity on $(\mathbb{N}, \sqsubseteq)$, given in Theorem 4.2, is now a trivial consequence of (ii) and (iv) in Proposition 4.4, showing also that capacity is *strictly* monotone wrt $\sqsubseteq$.

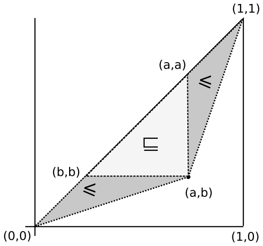

Fig. 1. Geometric representation of $\sqsubseteq, \leq$.

4.3 The coincidence of algebra, order and geometry

Each order is given by a simple formula (Definitions 4.1 and 4.3) that is easy to verify in practice. Each also has a clear geometric significance which makes it easy to reason about: for $x = (a, b) \in \mathbb{P}$ and $y \in \mathbb{P}$,

- $x \sqsubseteq y$ iff y is contained in the triangle with vertices $\{(a, a), x, (b, b)\}$ iff there is a line segment from x to a point of $\{(t, t) : t \in [b, a]\}$ that passes through y.
- $x \leq y$ iff y is contained in the triangle with vertices $\{(0, 0), x, (1, 1)\}$ iff there is a line segment from x to a point of $\{(t, t) : t \in [0, 1]\}$ that passes through y.

A geometric interpretation of these orders is shown in Figure 1. Remarkably, each of these orders can also be characterized algebraically:

Lemma 4.5 *For $x, y \in \mathbb{P}$,*

(i) $x \sqsubseteq y$ *iff* $(\exists z \in \mathbb{P})\, zx = y$,

(ii) $x \leq y$ *iff* $(\exists z \in \mathbb{P})\, xz = y$.

Thus, despite the somewhat awkward formulation of $\leq$ given in Definition 4.3, we see that $\leq$ is nevertheless quite natural. In fact, from the point of view of information theory, it is more natural than $\sqsubseteq$:

Theorem 4.6 *Let $(\mathbb{P}, \cdot, 1)$ denote the monoid of positive binary channels.*

(i) *The relation $x \leq y \equiv (\exists z \in \mathbb{P})\, xz = y$ defines a partial order on $\mathbb{P}$ with respect to which capacity $c : \mathbb{P} \to [0, 1]^*$ is strictly monotone,*

(ii) *For all $a, x \in \mathbb{P}$ there exists $y \in \mathbb{P}$ s.t. $xa = ay$,*

(iii) *The operator $l_x : \mathbb{P} \to \mathbb{P} :: l_x(y) = xy$ is monotone with respect to $\leq$,*

(iv) *The operator $r_x : \mathbb{P} \to \mathbb{P} :: r_x(y) = yx$ is monotone with respect to $\leq$.*

By contrast, r_x is monotone with respect to $\sqsubseteq$, but l_x is not. The reason for this difference is that $\mathbb{P} \cdot x \subseteq x \cdot \mathbb{P}$ holds for all $x \in \mathbb{P}$, and this inclusion is strict. So even though $\mathbb{P}$ is not commutative, it has a special property commutative monoids have which ensures that both l_x and r_x are monotone with respect to $\leq$. The monotonicity of l_x and r_x implies that

$$(\forall\, a, b, x, y \in \mathbb{P})\, x \leq y \Rightarrow c(axb) \geq c(ayb)$$

with equality iff $x = y$ since $axb \leq ayb$ and $c(axb) = c(ayb)$ implies $axb = ayb$ which from Lemma 2.3 implies that $x = y$. The above inequality, in turn, has an important and new consequence for information theory:

Corollary 4.7 *For all $a, b, x, y \in \mathbb{P}$,*

$$c(axyb) \leq \min\{c(axb), c(ayb)\}$$

with equality iff $x = 1$ or $y = 1$.

In particular, for $a = b = 1$, the well-known inequality $c(xy) \leq \min\{c(x), c(y)\}$ follows. It is interesting indeed that it may be derived from an order which itself may be derived from algebraic structure. This illustrates the value of knowing about the coincidence of algebra, order and geometry.

5 Relations between monotone mappings on channels

Having just considered relations between binary channels, we now turn to relations between monotone mappings on binary channels. Of particular interest is the fascinating relationship between capacity and Euclidean distance.

5.1 Algebraic relations

Both capacity and Euclidean distance are invariant under multiplication by the idempotent $e = (0, 1)$:

Lemma 5.1 *Let $e := (0, 1)$. For any $x = (a, b) \in [0, 1]^2$,*

(i) $e \cdot (a, b) = (b, a) \quad \& \quad (a, b) \cdot e = (\bar{a}, \bar{b})$

(ii) $c(ex) = c(xe) = c(x)$

(iii) $|\det(ex)| = |\det(xe)| = |\det(x)|$

We now establish our first result which relates capacity to distance:

Theorem 5.2 *For two binary channels $x, y \in [0, 1]^2$,*

$$c(xy) \leq \min\{\ c(x)|\det(y)|, |\det(x)|c(y)\ \}.$$

with equality iff x (or y) is 1, e or a zero channel.

The last result extends to *any* convex function on $\mathbb{N}$. It gives a new proof of a well-known result in information theory,

Corollary 5.3 *For $x, y \in [0, 1]^2$, $c(xy) \leq \min\{c(x), c(y)\}$ with equality iff x (or y) is 1, e or a zero channel.*

and also sheds light on the relation between Euclidean distance and capacity:

Corollary 5.4 *For a binary channel $x \in [0, 1]^2$, $c(x) \leq |\det(x)|$ with equality iff x is 1, e or a zero channel.*

Intuitively, the Euclidean distance $|\det|$ is a canonical upper bound on capacity. Our goal now is to prove this. First, $|\det|$ is determined by its value on the set $\mathbb{N}$ of nonnegative channels. Next, as a function on $\mathbb{N}$, it preserves multiplication, convex sum and identity. There are only two functions like this in existence:

Theorem 5.5 *If $f : \mathbb{N} \to [0, 1]$ is a function such that*

- $f(1) = 1$
- $f(xy) = f(x)f(y)$
- $f(px + \bar{p}y) = pf(x) + \bar{p}f(y)$

then either $f \equiv 1$ or $f = \det$.

Thus, there is only one nontrivial convex-linear homomorphism above capacity: the determinant. This raises the question of how close in value the two are.

5.2 Inequalities

In the formulation of $\leq$ given in Definition 4.3, the case $\mu x = \mu y$ is specifically excluded i.e. channels that lie on a line of constant determinant do not compare with respect to $\leq$ unless they are equal. The behavior of capacity on such lines is more involved than it is for lines that hit the diagonal. We now turn to this important special case, and once again, find the monotonicity principle indispensable.

Consider a line in $\mathbb{N}$ of fixed determinant, that is, a line joining the Z-channels $(d, 0)$ and $(1, 1-d)$:

$$\pi_d(t) = t(1, 1-d) + \bar{t}(d, 0)$$

Let $c(t)$ denote the capacity of the channel $\pi_d(t)$.

Theorem 5.6 *The function $c \circ \pi_d$ for $d > 0$ is strictly monotonically decreasing on $[0, \frac{1}{2}]$ and strictly monotonically increasing on $[\frac{1}{2}, 1]$. For $d = 0$, it is identically zero.*

We have derived the following lower and upper bounds on the capacity:

Corollary 5.7 *For any binary channel $x \in [0, 1]^2$,*

$$1 - H\left(\frac{1 - |\det(x)|}{2}\right) \leq c(x) \leq \log_2\left(1 + 2^{\frac{-H(|\det(x)|)}{|\det(x)|}}\right)$$

with the understanding that the expression on the right is zero when $\det(x) = 0$.

The bounds in Corollary 5.7 are canonical:

Definition 5.8 A function $f : [0, 1]^2 \to \mathbb{R}$ is called *det-invariant* if $|\det(x)| = |\det(y)|$ implies $f(x) = f(y)$ for all $x, y \in \mathbb{N}$.

Thus, a det-invariant function is one whose value depends only on the magnitude of the channel's determinant – in particular, such functions are symmetric.

Corollary 5.9 *The supremum $a(x)$ and infimum $b(x)$ of all det-invariant lower bounds on capacity are equal to*

$$a(x) = 1 - H\left(\frac{1 - |\det(x)|}{2}\right) \qquad\qquad b(x) = \log_2\left(1 + 2^{\frac{-H(|\det(x)|)}{|\det(x)|}}\right)$$

The best det-invariant lower bound in Corollary 5.7 is the key idea in determining how close in value $|\det|$ is to c:

Theorem 5.10

$$\sup_{(a,b)\in[0,1]^2} |\det(a,b)| - c(a,b) = \log_2(5/4)$$

which is attained by the channels $(4/5, 1/5), (1/5, 4/5)$.

The number $\log_2(5/4)$ is approximately equal to 0.3219. Because $|\det|$ itself is a det-invariant upper bound on capacity, $b(x) \leq |\det(x)|$ by Corollary 5.9, and we have the following chain of inequalities:

$$a(x) \leq c(x) \leq b(x) \leq |\det(x)| \leq c(x) + \log_2\left(\frac{5}{4}\right)$$

Results like these can be applied to the difficult problem of bounding the capacity of a *timing channel*: a channel where each output symbol has an associated cost $t_i > 0$ and one seeks to determine *capacity per unit time* or put simply, timed capacity. In this problem, bounds are especially useful because there are no formulae available for computing timed capacity – even in the case of a noiseless binary timing channel, one must resort to numerical methods [10]. However, any lower and upper bound on capacity c leads to one on timed capacity c_t because

$$\frac{c}{\max t_i} \leq c_t \leq \frac{c}{\min t_i}$$

so we obtain a bound on the capacity of a binary timing channel that in some cases may be quite useful.

5.3 A topological relation

Earlier we studied partial orders on binary channels, each offers a different way of relating a pair of channels to one another. We then jumped up a level of abstraction and studied relations that exist between fundamental monotone mappings on binary channels. However, the partial orders are not merely partial orders, and the monotone mappings are not merely monotone. In each case, more mathematical structure is present, and by taking this additional structure into account, a new relation between capacity and distance emerges. This one is topological.

The extra structure that the poset $\mathbb{N}$ has is that it is a *domain*: a partially ordered set with intrinsic notions of completeness and approximation defined by the order. The extra structure that capacity has is that it is a *measurement*: a function μ that to each informative object x assigns a number μx which measures the information content of the object x. We now define each of these terms precisely before discussing them further.

The intrinsic notion of completeness that a domain has is that it forms a dcpo:

Definition 5.11 Let $(P, \sqsubseteq)$ be a partially ordered set or *poset*. 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.

The intrinsic notion of approximation possessed by a domain is formalized by continuity:

Definition 5.12 Let $(D, \sqsubseteq)$ be a dcpo. For elements $x, y \in D$, we write $x \ll y$ iff for every directed subset S with $y \sqsubseteq \bigsqcup S$, we have $x \sqsubseteq s$, for some $s \in S$. We set

$$\mathord{\downarrow\!\!\downarrow} x := \{y \in D : y \ll x\} \text{ and } \mathord{\uparrow\!\!\uparrow} x := \{y \in D : x \ll y\}$$

and say D is *continuous* if $\mathord{\downarrow\!\!\downarrow} x$ is directed with supremum x for each $x \in D$.

Definition 5.13 A *domain* is a continuous dcpo.

The poset of nonnegative channels $\mathbb{N}$ is order isomorphic to the compact subintervals of the unit interval

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

ordered by reverse inclusion with an explicit isomorphism given by $\mathbb{N} \to \mathbf{I}[0,1] :: (a,b) \mapsto [b,a]$. This correspondence implies that $\mathbb{N}$ forms a domain, called the *interval domain*, where $\bigsqcup S = \bigcap S$, for directed $S \subseteq \mathbf{I}[0,1]$ and $x \ll y$ iff $y \subseteq \mathrm{int}(x)$. Notice that $\mathrm{int}(x)$ refers to the interior of the interval x in its relative Euclidean topology.

Definition 5.14 The *Scott topology* on a continuous dcpo D has as a basis all sets of the form $\mathord{\uparrow\!\!\uparrow} x$ for $x \in D$.

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. Let $[0, \infty)^*$ denote the poset of nonnegative reals in their dual order: $x \sqsubseteq y \equiv y \leq x$.

Definition 5.15 A Scott continuous $\mu : D \to [0, \infty)^*$ is said to *measure the content* of $x \in D$ if for all Scott open sets $U \subseteq D$,

$$x \in U \Rightarrow (\exists \varepsilon > 0)\, x \in \mu_\varepsilon(x) \subseteq U$$

where $\mu_\varepsilon(x) := \{y \in D : y \sqsubseteq x \ \&\ |\mu x - \mu y| < \varepsilon\}$.

Definition 5.16 A *measurement* $\mu : D \to [0, \infty)^*$ is a Scott continuous map that measures the content of $\ker(\mu) := \{x \in D : \mu x = 0\}$.

The order on a domain D defines a clear sense in which one object has 'more information' than another: a *qualitative* view of information content. The definition of measurement attempts to identify those monotone mappings μ which offer a *quantitative* measure of information content in the sense specified by the order. The essential point in the definition of measurement is that μ measure content in a manner that is consistent with the particular view offered by the order. There are plenty of monotone mappings that are not measurements – and while some of them may measure information content in *some other sense*, each sense must first be specified by a different information order. The definition of measurement is then a minimal test that a function μ must pass if we are to regard it as providing a measure of information content.

We now consider a few properties that measures of information content have which arbitrary monotone mappings in general need not have: qualities that make them 'different' from maps that are simply monotone. Other such properties may be found in [7]. Define $d : D^2 \to [0, \infty)^*$ by

$$d(x, y) = \inf\{\mu z : z \sqsubseteq x, y\}$$

where we assume that D has either a least element or more generally is 'filtered'. Denote the ε balls with respect to d by $B_\varepsilon(x) := \{y \in D : d(x, y) < \varepsilon\}$.

Theorem 5.17 (Martin[7]) *Let $\mu : D \to [0, \infty)^*$ be a measurement.*

(i) $x \in \ker(\mu) \Rightarrow x \in \max(D) = \{x \in D : \uparrow\!x = \{x\}\}$.

(ii) *If μ measures the content of $y \in D$, then*

$$(\forall x \in D)\ x \sqsubseteq y \ \&\ \mu x = \mu y \Rightarrow x = y.$$

(iii) *If μ measures $X \subseteq D$, then $\{B_\varepsilon(x) \cap X : x \in X, \varepsilon > 0\}$ is a basis for the Scott topology on X.*

Theorem 5.17 says that *any measurement* on $\mathbb{N}$ induces the Euclidean topology on its kernel.

Theorem 5.18 (Martin[9]) *Capacity $c : \mathbb{N} \to [0, 1]^*$ is a measurement.*

In the case of capacity $c : \mathbb{N} \to [0, 1]^*$, the associated distance function on $\ker(c) = \max(\mathbb{N})$ works out to be $\rho([a], [b]) = c(a, b) = c(b, a)$. Then, just like Euclidean distance, capacity also has the following three properties: (i) $c(a, b) = c(b, a)$, (ii) $c(a, b) = 0$ iff $a = b$, (iii) the sets $\{y \in [0, 1] : c(x, y) < \varepsilon\}$ for $\varepsilon > 0$ form a basis for the Euclidean topology on $[0, 1]$.

Capacity does not satisfy the triangle inequality, so it is not a priori obvious that the sets in (iii) form a basis for any topology, let alone the Euclidean topology. Thus, this is another relationship between capacity and the determinant: each of them is a measure of distance that induces the Euclidean topology. What we now seek is a better understanding of *why* this happens. For this, we need to think about how this result is proved.

The proof that capacity is a measurement given in [9] depends on the result of Majani and Rumsey [6], and is interesting since it connects the study of measurement in domain theory to a fundamental and beautiful result in information theory. Specifically, it is shown that $c(a,b) \geq (\det(a,b))^2/(e^2\ln(2))$. This lower bound has the form $\nu \circ \det$ where $\nu : [0,1] \to [0, 1/(e^2\ln(2))]$ is the order isomorphism $\nu(t) = t^2/(e^2\ln(2))$. The proof in [9] relies heavily on specific results only known to hold for binary channels as well as intricate arguments from analysis. We now give a new proof of this result which has several advantages over the one in [9].

Proposition 5.19 *Let $\varphi : [0,1] \to [0,1]$ be the function $\varphi(t) = 1 - H((1-t)/2)$. If $\lambda : \mathbb{N} \to [0,\infty)^*$ is Scott continuous and $\lambda \geq \varphi \circ \det$, then λ is a measurement. In particular, capacity is a measurement.*

In fact, *every* function in the string of inequalities

$$1 - H\left(\frac{1-|\det(x)|}{2}\right) \leq c(x) \leq \log_2\left(1 + 2^{\frac{-H(|\det(x)|)}{|\det(x)|}}\right)$$

is a measurement and has properties (i), (ii) and (iii) discussed earlier. This new proof is an improvement over the one in [9]:

- $\varphi \circ \det$ is a better lower bound than $\nu \circ \det$: by Corollary 5.9, $\nu \circ \det \leq \varphi \circ \det$,
- The inequality $\varphi \circ \det \leq c$ is derived using only the monotonicity principle, a fact known to hold for arbitrary channels.

Moreover, the measurement $\varphi \circ \det$ has profound applied significance, as we now demonstrate by using it to solve an open problem in quantum steganography.

6 Quantum steganography

In this section, we will learn a few of the fascinating implications the results in this paper have within the realm of communication. We first review the basic protocol for quantum key distribution. Because we intend for this paper to be readable by someone with no prior knowledge of quantum mechanics, we discuss only the minimal background needed to understand quantum key distribution. The few ideas we make use of are very simple.

6.1 Quantum information

Like all systems, a quantum system has *state.* The state of a quantum system is represented by a unit vector in a vector space that has a lot more structure than most, known as a Hilbert space. The state of a quantum system is also called a *ket.* Here are two examples of kets: $|0\rangle$ and $|1\rangle$. It is useful to think about these two particular kets as being quantum realizations of the classical bits 0 and 1. Each refers to a legitimate state of a quantum system. A photon is an example of a quantum system and its polarization (state) is something we need kets to describe.

One of the neat things about a quantum system is that it can also be in any state ‘in between’ $|0\rangle$ and $|1\rangle$, such as $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ or $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, which we also think of as representing the classical bits 1 and 0 respectively. Any ket $|\psi\rangle$ that can be written as $|\psi\rangle = a|0\rangle + b|1\rangle$, for $|a|^2 + |b|^2 = 1$, is called a *qubit.* There are only four qubits that we care about in this paper: $|0\rangle$, $|1\rangle$, $|+\rangle$ and $|-\rangle$.

Like all systems, one would like to extract information from a quantum system. One way to extract information from a quantum system is to perform a measurement on it. Before an observer can perform a measurement on a quantum system, they must say what they want to measure. One way an observer can specify what they want to measure is by specifying a *basis* and then “performing a measurement in the specified basis.” Two examples of bases are $X = \{|+\rangle, |-\rangle\}$ and $Z = \{|0\rangle, |1\rangle\}$. They are the only bases we will use in this paper [4] . What happens when we measure a quantum system?

If the state of a quantum system is described by the qubit $|\psi\rangle = a|0\rangle + b|1\rangle$, then a measurement in the Z basis will yield the result $|0\rangle$ with probability $|a|^2$ and the result $|1\rangle$ with probability $|b|^2$. Notice that these are the only possible outcomes of this measurement because qubits satisfy $|a|^2 + |b|^2 = 1$, a property they have because they are unit vectors. In this paper, we only care about measuring the following four states in the Z basis: $|0\rangle$, $|1\rangle$, $|+\rangle$ and $|-\rangle$. If we measure a system with state $|0\rangle$ in the Z basis, we get $|0\rangle$ with probability 1; the same is true of the state $|1\rangle$. If we measure either $|+\rangle$ or $|-\rangle$ in the Z basis, we obtain $|0\rangle$ with probability 1/2 and $|1\rangle$ with probability 1/2.

It is also possible to measure a system in the X basis. If a system is in the state $|+\rangle$ and we measure it in the X basis, we get $|+\rangle$ with probability 1, similarly for $|-\rangle$. But what happens when we measure a system with state $|0\rangle$ or $|1\rangle$ in the X basis? Well, first we have to express these states as sums of states in the X basis:

$$|0\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle) \qquad\qquad |1\rangle = \frac{1}{\sqrt{2}}(|+\rangle - |-\rangle)$$

Now we can see that if we measure $|0\rangle$ in the X basis, we get $|+\rangle$ with probability 1/2 and $|-\rangle$ with probability 1/2, similarly for $|1\rangle$.

6.2 Quantum key distribution

We now recall one of the standard accounts of quantum key distribution (QKD), the BB84 protocol [15]: (1) Alice chooses a random string k of about $4n$ bits containing the eventual key. (2) Alice randomly codes each bit of k in either the $X = \{|+\rangle, |-\rangle\}$ or $Z = \{|0\rangle, |1\rangle\}$ bases. (3) Alice sends each resulting qubit to Bob. (4) Bob receives the $4n$ qubits, randomly measuring each in either the X or the Z basis. (5) Alice announces in which basis she originally coded each bit of k. (6) Bob tells Alice which bits he received correctly; they now share about $2n$ bits. (7) Alice selects a subset of n bits from the group she formed in step (6) that will be used. to check on Eve’s

[4] Many bases are possible, and each offers a *different* way of representing the classical bits 0 and 1. The ability to alternate between such representations helps prevent eavesdropping in QKD.

interference, and tells Bob which bits she selected. (8) Alice and Bob compare their values of the n check bits; if more than an acceptable number disagree, they abort the protocol (eavesdropping). (9) Alice and Bob perform information reconciliation and privacy amplification to select a smaller m-bit key from the remaining n bits.

The bits in step (6) are called the *sifted bits.* If Alice has coded a classical bit in either of the X or Z bases, and later Bob measures in the same basis, he will receive the bit sent by Alice with probability 1. Such a bit will be one of the sifted bits. But now suppose that an eavesdropper wishes to know the bit Alice is sending Bob. Well the eavesdropper, named Eve, has to guess which basis Alice coded the bit in, and then measure it herself. When Eve guesses, she introduces an error into the sifted bits with probability 1/4 – but an error that Alice and Bob will know about, and this is the reason they are able to detect the presence of an eavesdropper.

It is fundamental in QKD that Alice and Bob insist on an error rate within the sifted bits that is less than 1/4 to defend themselves from precisely this type of attack, or else the security of QKD cannot be guaranteed [1]. For instance, assuming errors only due to Eve, if Eve has measured all the qubits sent from Alice to Bob, then Eve knows which of the sifted bits Bob and Alice share, and which of the sifted bits they may not share[5]. This is something that Bob himself does not even know. With an error rate beyond 1/4, Bob cannot have more information than Eve about any key generated – remember that after the sifted bits are identified, Eve can listen in on the rest of the protocol, since it takes place over a public channel.

6.3 Analysis of hidden channels within quantum key distribution

As explained in [8], QKD is not communication, for the simple fact that neither Alice nor Bob has any control over the sifted bits, or the key their interaction ultimately produces. However, as first shown in [8], it can be modified so that communication is possible: a quantum protocol can be used to obtain a new protocol which is physically indistinguishable from the original, but which also contains a channel whose existence is undetectable by any currently known technology. The potential of such 'hidden channels' is discussed in [8].

Let us give a simple illustration of how such hidden channels arise. Assume Alice would like to send Bob a single bit of information. All we have to do is make a simple change to step (7) in QKD: "(7) Alice randomly selects a bit from the group of $2n$ whose value is the information she wants to transmit. Then she randomly selects $n-1$ check bits from the remaining $2n-1$. The n^{th} check bit is chosen from the remaining $n+1$ as being the bit to the immediate left of the information." Bob now has the information Alice sent: he knows its relation to the last check bit, because the two parties have agreed on this scheme in advance. They have agreed that Alice will covertly send Bob a 'pointer' to the information.

Here is an example: Alice and Bob share the $2n$ bits, Alice selects the information bit

$$0\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ \overline{1}\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 1\ 0\ 0$$

[5] Eve knows when she guessed the right basis and when she did not.

Now she selects $n-1$ check bits at random, which leaves Alice and Bob with $n+1$ remaining bits

$$0 * 1\ 0 * 0\ 1 * 1 * 0 * 0 * 0 * * \bar{1} * 1\ 0 * 1 * 1\ 1 * *$$

Alice now selects the last check bit as being the pointer to the information i.e. the bit to the immediate left of the information bit:

$$0 * 1\ 0 * 0\ 1 * 1 * 0 * 0 * \overrightarrow{\mathbf{0}} * * \bar{1} * 1\ 0 * 1 * 1\ 1 * *$$

Is Bob *guaranteed* to receive the information sent by Alice? No. There are many reasons why. Suppose an eavesdropper *just happens* to measure only the qubit that holds the information in the wrong basis, then there is a 50 percent chance that Bob has the wrong bit, even though he believes he has the right bit. Or suppose that background light acts as noise which causes the information bit to flip. In either case, Bob would have no idea, and neither would Alice. But chances are good that such errors, whether caused by the environment or an eavesdropper, would also manifest themselves in the check bits as well, which would then enable them to estimate the likelihood that their attempt to communicate will succeed. Alice and Bob always have the option of aborting the protocol if their chances of success are not deemed high enough. The question we want to answer is: what is the capacity of this channel?

Theorem 6.1 (Martin [8]) *If the error rate $\alpha \in [0, 1/4)$ is due solely to interference caused by Eve, then the capacity of the Alice-Bob steganographic channel is $1 - H(\alpha)$. In particular, the capacity of the Alice-Bob channel is never smaller than $1 - H(1/4) \approx 0.1887$.*

It is important to understand in the last result that the phrase "interference caused by Eve" means not only that Eve causes all errors, but that she causes these errors by essentially performing random combinations of X and Z measurements. This point is fundamental, since it implies that the probability of a 0 flipping to a 1 is the same as the probability that a 1 flips to a 0. The reason is that Eve cannot tell which classical bit a qubit represents before she performs a measurement and that Alice sends bits with equal frequency. This means the hidden channel is binary symmetric where the probability of a flip is α: the noise matrix of the hidden channel is $(\bar{\alpha}, \alpha)$, so its capacity is $1 - H(\alpha)$. Moreover, because the parameter α is an experimentally determined quantity that *must* be calculated in any realization of QKD to check for the presence of an eavesdropper, the last result allows us to calculate the capacity of any hidden channel based on sifted bits any time that a QKD experiment is performed [8].

But now suppose Eve does something other than perform measurements in the X and Z bases at random. Or suppose some of the noise is caused by the environment. Then it is no longer necessarily the case that 1 and 0 flip with the same probability, so whatever the noise matrix of the hidden channel is, we know that it is not necessarily binary symmetric. For instance, a well-known effect like amplitude damping [15] does not affect $|0\rangle$ but does flip $|1\rangle$. Thus, in the case of general noise, the matrix for the channel is unknown, in the sense that it cannot be determined from *experimentally necessary quantities.* We can nevertheless establish the following important lower bound on its capacity:

Theorem 6.2 *Let the error rate be $\alpha \in [0, 1/4)$. Then the capacity of the Alice-Bob steganographic channel is at least $1 - H(\alpha)$. In particular, the capacity of the Alice-Bob channel is never smaller than $1 - H(1/4) \approx 0.1887$.*

The principle underlying this last result is clear as well as surprising: from the point of view of Alice and Bob, noise caused by an arbitrary combination of environment and eavesdropper is preferable to noise caused by an eavesdropper alone. The reason is that an eavesdropper causes bits to flip with equal probability, leading to a binary symmetric channel, whereas the environment may not.

Put another way, any attempt to interrupt the hidden channel should *necessarily* employ a random combination of X and Z measurements. Even for error rates α arbitrarily close to $1/4$, any scheme that does not flip bits with equal probability will permit a capacity higher than the theoretical limit of $1 - H(1/4)$, up to a possible maximum of $c(1, 1/2) = \log_2(5/4) \approx 0.32$.

As these results make clear, det-invariant bounds on capacity provide a valuable way to approximate the capacity of a channel when its determinant is known, but its noise matrix is not. As we have seen, such channels arise naturally in experimental situations.

7 Acknowledgments

As far as we aware, Ira S. Moskowitz of NRL was the first to conjecture the behavior of capacity along a line of constant determinant, in a private communication to the authors. We are grateful to him for this important insight. We thank Catuscia Palamidessi for a series of valuable discussions on the work contained herein, and Prakash Panangaden, who introduced the authors to each other at last year's Bellairs workshop on information theory and security.

References

[1] Brassard, G., N. Lütkenhaus, T. Mor and B. C. Sanders, *Limitations on practical quantum cryptography*, Phys. Rev. Lett. **85** (2000), pp. 1330–1333.

[2] Chatzikokolakis, K. and K. Martin, *A mononoticity principle for information theory*, report version. Available at http://www.lix.polytechnique.fr/~kostas/papers/mfps08.pdf.

[3] Chatzikokolakis, K., C. Palamidessi and P. Panangaden, *Anonymity protocols as noisy channels*, Information and Computation (2007), to appear.

[4] Clark, D., S. Hunt and P. Malacaria, *Quantified interference for a while language*, in: *Proc. of QAPL 2004*, ENTCS **112** (2005), pp. 149–166.

[5] Cover, T. M. and J. A. Thomas, "Elements of Information Theory," John Wiley & Sons, Inc., 1991.

[6] Majani, E. and H. Rumsey, *Two results on binary-input discrete memoryless channels*, in: *Proceedings of the IEEE International Symposium on Information Theory*, 1991, pp. 104–104.

[7] Martin, K., “A foundation for computation,” Ph.D. thesis, Tulane University, Department of Mathematics (2000).

[8] Martin, K., *Steganographic communication with quantum information*, in: *Proceedings of Information Hiding '07*, LNCS (2007), to appear.

[9] Martin, K., *Topology in information theory in topology*, Theoretical Computer Science (2007), to appear.

[10] Martin, K. and I. S. Moskowitz, *Noisy timing channels with binary inputs and outputs*, in: *Proceedings Information Hiding '06*, LNCS **4437** (2006), pp. 124–144.

[11] Martin, K., I. S. Moskowitz and G. Allwein, *Algebraic information theory for binary channels*, ENTCS **158** (2006), pp. 289–306.

[12] Millen, J., *Covert channel capacity*, in: *Proceedings of the 1987 IEEE Symp. on Computer Security and Privacy*, 1987.

[13] Moskowitz, I. S., R. E. Newman and P. F. Syverson, *Quasi-anonymous channels*, in: *IASTED CNIS*, 2003, pp. 126–131.

[14] Nielsen, M., K. Krukow and V. Sassone, *A bayesian model for event-based trust*, ENTCS **172** (2007), pp. 499–521.

[15] Nielsen, M. A. and I. L. Chuang, “Quantum Computation and Quantum Information,” Cambridge University Press, 2000.

[16] Shannon, C. E., *Some geometrical results in channel capacity*, in: *Collected Papers of C.E. Shannon*, IEEE Press, 1993 pp. 259–265.