An Approximation of the Capacity of a Simple Channel

Ira S. Moskowitz
CHACS — 5540, Naval Research Laboratory, Washington, DC 20375

Abstract

We discuss the bounds for capacity of a binary-input binary-output discrete memoryless communication channel. We introduce a new lower bound that gives a very good and elementary approximation to the capacity.

1. Introduction

In this paper we study only "simple channels," whereby "simple channel" we mean a binary-input binary-output discrete memoryless channel [11]. The capacity (C) of a simple channel is a function of two variables: the "noise" terms a and b. The graph of C is a surface in $\mathbb{R}^3$. Analysis of this surface and its level sets gives us guidance as to the best ways to manipulate capacity by adjusting the noise terms, the idea being to manipulate capacity while having minimal impact upon system performance. To simplify terminology for the rest of the paper unless stated otherwise, "channel" will mean "simple channel."

Capacity is expressed as a function of the noise terms a and b. Unfortunately, this expression is non-linear and logarithmic in nature and thus does not readily lend itself to obvious "rules of thumb" describing how the noise terms affect the capacity. We seek a simple approximation for capacity in terms of noise that can guide attempts to alter noise on channels, and hence, capacity.

We bound capacity from above and below by uncomplicated functions of the determinant of the channel matrix. We use our lower bound as an approximation to the capacity. This simple formulation of the approximate capacity is readily useful for analysis of channel behavior and holds promise for approximations to multiple bit channels.

The lower bound for capacity developed in this paper is approximately $.72(a-b)^2$, a vast simplification of the actual capacity, which is $\log_2\left(2^{\frac{\bar{a}h(b)-\bar{b}h(a)}{a-b}}+2^{\frac{bh(a)-ah(b)}{a-b}}\right)$, where $\bar{a}=1-a$, and $\bar{b}=1-b$. Our lower bound gives small error for most values of a and b. We also give a simple proof of the known [2] result that $|a-b|$ is an upper bound for capacity and discuss the implications of our approximations to capacity.

$$M=\begin{pmatrix} a & 1-a \\ b & 1-b \end{pmatrix} \qquad (1)$$

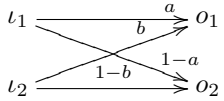

Figure 1. Channel matrix , $\det M = a-b$**.**

2. Mutual Information and Capacity

The input symbols to the channel are $\{\iota_1,\iota_2\}$, and the output symbols are $\{o_1,o_2\}$. Since all symbols take the same time to go through the channel, all information theoretic measurements are in units of bits per channel usage (symbol).

The channel matrix represents the conditional probability relationships between the input and output symbols. That is, $a=P(o_1|\iota_1), 1-a=P(o_2|\iota_1), b=P(o_1|\iota_2)$, and $1-b=P(o_2|\iota_2)$. The input probabilities are represented by the random variable X, $P(X=\iota_i)=x_i, i=1,2$, which we simplify to $P(\iota_1)=x_1$ and $P(\iota_2)=x_2$. Similarly, we have the random variable Y such that $P(o_1)=y_1$ and $P(o_2)=y_2$. This is illustrated in Fig. 1 . We summarize the probabilities as follows: $\vec{y}=\vec{x}\cdot\begin{pmatrix} a & 1-a \\ b & 1-b \end{pmatrix}$. Letting $x=x_1$ and using the fact that $x_1+x_2=1=y_1+y_2$, we have that $y_1=(a-b)x+b$ and $y_2=1-y_1$. To calculate the capacity we want to maximize the mutual information: $I=H(Y)-H(Y|X)$, over all possible distributions (x_1,x_2).

Since with a and b fixed, we can view I as a function of one variable x, the maximization problem reduces to maximizing the function $I:[0,1]\to\mathbb{R}$ given by:
$I(x)=h\big(f(x)\big)-xh(a)-(1-x)h(b)$, where $h:[0,1]\to\mathbb{R}$ is the binary entropy function[1]

$$h(x)=-x\log x-(1-x)\log(1-x) \qquad (2)$$

and $f:[0,1]\to[0,1]\subseteq\mathbb{R}$ is $f(x)=(a-b)x+b$. Of course, we can also let a and b vary and view $I(x)$ as a function of three variables $I_x(a,b)$. For fixed a and b, $C=C(a,b)=\max_x I(x)=\max_x I_x(a,b)$, where C is the capacity ([11]) as a function of a and b.

[1]The use of "log" is for the base two logarithm, whereas as "ln" is of course the natural logarithm. The binary entropy function is defined for $x\in[0,1]$, with $h(0)=h(1):=0$.

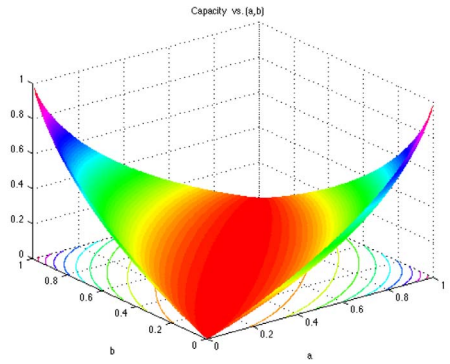

Figure 2. Capacity as a function of a and b.

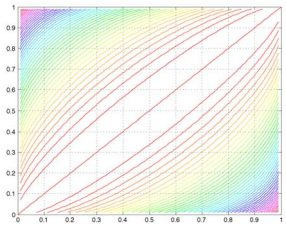

Figure 3. Capacity level sets a against b.

It can be shown (see Silverman [13, Eq. 5], Ash [1, Eq. 3.3.5], or [9, 10]) that capacity $C : I^2 \rightarrow [0,1]$, where $I^2 = [0,1] \times [0,1]$, as function of a and b is:

$$\begin{aligned} \mathrm{C}(a,b) &= \frac{\bar{a}h(b) - \bar{b}h(a)}{a-b} + \log\left(1 + 2^{\frac{h(a)-h(b)}{a-b}}\right) \\ &= \log\left(2^{\frac{\bar{a}h(b)-\bar{b}h(a)}{a-b}} + 2^{\frac{bh(a)-ah(b)}{a-b}}\right) \end{aligned} \tag{3}$$

where $\mathrm{C}(a,a) := 0$.

The above Eq. 3 shows that capacity is symmetric [13] about the diagonal line $\{a,a\} \subset I^2$ and the anti-diagonal line $\{a, 1-a\} \subset I^2$. We easily see, except for the points $\{a,a\}$, that C is a smooth function of a and b since it is made up of elementary functions. To show that C is continuous on all of I^2 is not difficult.

We divide the unit square I^2, with the line $b = a$ removed, into four regions as shown in Fig. 4. Note, behavior around (*e.g.* $(.8, .1)$) a point depends on how the neighborhood intersects different regions. Region 1 consists of the points $\{(a,b) : b < a, b < 1-a\}$, Region 2 is $\{(a,b) : b < a, b > 1-a\}$, Region 3 is $\{(a,b) : b > a, b > 1-a\}$, and Region 4 is $\{(a,b) : b > a, b < 1-a\}$. As noted above by using Eq. 3 it is obvious that capacity is symmetric in the four different regions; that is $C(a,b) = C(1-b, 1-a) = C(1-a, 1-b) = C(b,a)$.

As in [10] we call the union of Region 1, Region 2, and $\{(1, 1-a) : a \in (.5, 1]\}$, the *positive channels* ($b < a$). That is, the positive channels are all channels "under" the main diagonal $a = b$. The union of Region 3, Region 4, and $\{(1, 1-a) : a \in [0, .5)\}$, is called the *negative channels* ($b > a$), that is; all channels "above" the main diagonal. The channels of the form $\{(a,b) : a = b\}$ are called the *zero channels*. As discussed in [10], this terminology is also used to keep the behavior of the determinant (which is $a-b$) of the channel matrix Eq. 1 in mind.

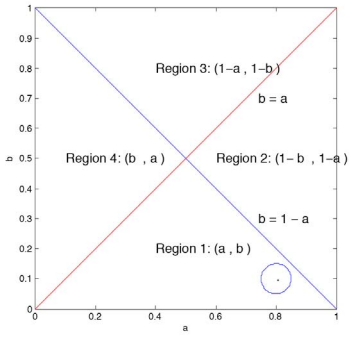

Figure 4. Capacity quadrants of I^2.

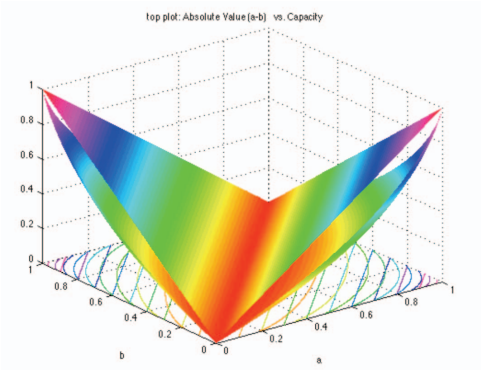

Figure 5. Top $|a-b|$, bottom is capacity.

3. Capacity Upper Bound

$C(a,b)$ denotes the capacity of the channel (a,b) and it was first shown in [2, Cor. 5.4], via convexity arguments, that $C(a,b) \leq |a-b|$, with equality only for zero channels and the channels $(1,0)$ and $(0,1)$. From Helgert (1967) [3] we can trivially show the weaker result $C(a,b) \leq |a-b|$. To show that $C(a,b) < |a-b|$, except for zero channels and the channels $(1,0)$ and $(0,1)$, requires [2, Cor. 5.4][2].

Theorem 3.1 *: For a (2,2) channel with transition matrix M, $C \leq |det(M)|$.*

This result was first conjectured by the authors of [10]. As noted above, we offer an alternate proof to the version given in [2] that follows as a trivial corollary from Helgert's earlier result. Of course, it is not surprising that both [3]

[2]The proofs for many of the stated results in [2] are in the appendix of a report version of the paper available at `http://www.win.tue.nl/~kostas/publications.html`.

and [2] rely upon convexity arguments. Also keep in mind that [2] deals with much more than Thm. 3.1. To prove Thm. 3.1, we use the following result of Helgert (in the statement of Helgert's theorem the term channel no longer means simple channel):

Theorem 3.2 *(Helgert [3]): Let P be a $m \times n$ stochastic matrix with entries $p_{i,j}$ of rank ρ characterizing a discrete, memoryless channel whose capacity is C_P. Then,* $C_P \leq (1-\Delta)\log_2 \rho$, *where* $\Delta = \sum_{j=1}^{n} \min_i(p_{i,j})$.

Proof sketch of Thm. 3.2: Details in [3]. The proof uses Shannon's [12] result that if a channel matrix P can be written as the convex sum of other channel matrices, then the capacity of P is less than, or equal to, the convex sum of the capacity of the other channel matrices. ■

Proof of Thm. 3.1: If the (2,2) channel is a zero channel, then $\rho = 1$; therefore $\log_2 \rho = 0$. Since $a = b$, the mutual information $I = H(Y) - H(Y|X)$ is easily seen to be identically equal to zero (independent of the probability mass function of X); therefore $C = 0$ and $C = \det(M)$. Now we use Helgert's result —

If the (2,2) channel is positive then $\Delta = b + (1-a)$ and $\rho = 2$. Therefore $C \leq a - b = \det(M)$.

If the (2,2) channel is negative then $\Delta = a + (1-b)$ and $\rho = 2$. Therefore $C \leq b - a = -\det(M)$. ■

In Fig. 5 we see how $|\det M|$ graphically behaves as a strict upper bound for capacity, except for zero channels and the two channels with capacity one. Looking at Fig. 5 it seems as if $|\det M|$ would be a good approximation to capacity. More careful scrutiny of Fig. 5 (also see Fig. 6) reveals that for a given capacity, the level sets of the approximation and the actual capacity are far apart. In the section that follows, we present a lower bound for capacity that also acts as a good approximation of capacity.

4. Taylor Approximations and Lower Bound

If the level sets of capacity were straight lines, then the analysis of how noise affects capacity would be easier, and we could easily make accurate approximations to the behavior of capacity as a function of the noise terms a and b. But, as we see in Fig. 3, they are not straight lines. The level sets (or curves) are slightly curved and symmetrical about both the main diagonal (the line $b = a$) and the anti-diagonal (the line $b = 1 - a$). Attempting to calculate the parameterization of the level sets, or even derivatives of the capacity function itself, is quite difficult. There does not seem to be any nice closed form results for the various geometric terms of interest. We cannot rely on pictures and say, "Well the level sets are almost straight lines, so let us approximate them and the capacity function by something easy!" However, we can attempt to find an approximation of

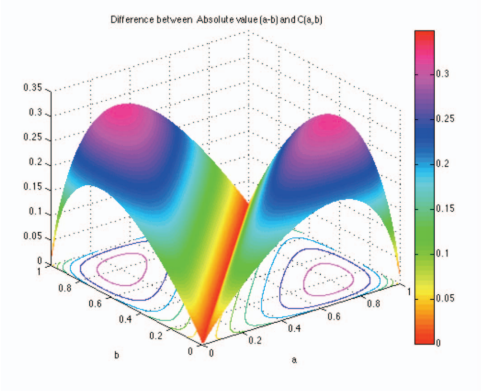

Figure 6. Plot of $|\det M| - C$, with level sets

the capacity that has straight line level sets, based on sound theoretical reasoning.

We note that Majani [6] has done work on approximations to capacity, and in fact we are motivated by his use of the Taylor approximation of the binary entropy function. However, the emphasis in [6] was viewing capacity as a function of the input probability x. We wish to emphasize the effects of the noise terms a and b, rather than x. Since the binary entropy function (Eq. (2)) is symmetric about $1/2$ we define a new function $t(x) = h(x + .5), x \in [-.5, .5]$. We have the Maclaurin series of $t(x)$

$$t(x) = t(0) + t'(0)(x) + t''(0)\frac{(x)^2}{2!} + \cdots \tag{4}$$

For simplicity, we switch to natural logarithms with

$$h_e(x) = -x\ln x - (1-x)\ln(1-x) \text{ and } t_e(x) = h_e(x+.5) \tag{5}$$

We easily see that $h'_e(x) = \ln\left(\frac{1-x}{x}\right)$ and $h''_e(x) = \frac{-1}{x(1-x)}$. Therefore, $t'_e(x) = \ln\left(\frac{.5-x}{.5+x}\right)$ and $t''_e(x) = \frac{-1}{(.5+x)(.5-x)}$, and $t'_e(0) = 0, t''_e(0) = -4$.

Approximate $t_e(x)$ by 2^{nd} order Maclaurin polynomial

$$h_e(x+.5) = t_e(x) \approx \ln(2) - 2x^2 \tag{6}$$

Switching to the base two logarithm gives us

$$h(x+.5) = t(x) \approx 1 - \frac{2}{\ln(2)}x^2 = 1 - 2\log(e)x^2 \tag{7}$$

Combining the fact that, by definition of capacity as a supremum, $I(.5) \leq C$, with the lower bound given below of Majani [6, 7] we have that

$$.9421C \approx \frac{.5}{e^{-1}\log(e)}C \leq I(.5) \leq C \tag{8}$$

Therefore, we can reasonably approximate C by $I(.5) = h\left(\frac{a+b}{2}\right) - \frac{1}{2}(h(a) + h(b))$.

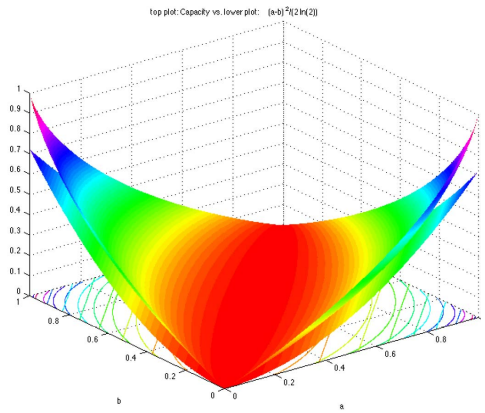

Figure 7. Top is capacity, bottom QT approx.

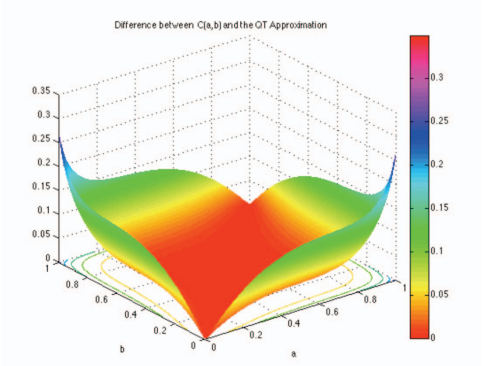

Figure 8. Scaled C-QT approx. with level sets

For any fixed point $(a_0, b_0) \in I^2$, we see that there is a unique straight line of slope 1 through (a_0, b_0) given by $b = a - k$, where $k = a_0 - b_0$. We see that I^2 is foliated (degeneracies at $(0,1)$ and $(1,0)$) by the intersection of the straight lines $b = a - k, k \in [-1, 1]$ with I^2. For a fixed k, consider the leaf given by the points $(a, b) \in I^2$ that satisfy $b = a - k$. We now look at $C(a, b) = C(a, a - k)$, which we approximate by $I_{.5}(a, a - k) = h\left(a - \frac{k}{2}\right) - \frac{1}{2}\left(h(a) + h(a - k)\right) = h\left[a - \left(\frac{1+k}{2}\right) + \frac{1}{2}\right] - \frac{1}{2}\left\{h\left[\left(a - \frac{1}{2}\right) + \frac{1}{2}\right] + h\left[\left(a - \left(\frac{1}{2} + k\right) + \frac{1}{2}\right]\right\}\right.$.
We now use our approximation Eq. (7). $h\left[a - \left(\frac{1+k}{2}\right) + \frac{1}{2}\right] \approx 1 - 2\log(e)\left[a - \left(\frac{1+k}{2}\right)\right]^2$, $h\left[\left(a - \frac{1}{2}\right) + \frac{1}{2}\right] \approx 1 - 2\log(e)\left[a - \frac{1}{2}\right]^2$, $h\left[\left(a - \left(\frac{1}{2} + k\right)\right) + \frac{1}{2}\right] \approx 1 - 2\log(e)\left[a - \left(\frac{1}{2} + k\right)\right]^2$.
So now approximate $I_{.5}(a, a - k) \approx \log(e)\left\{-2\left[a - \left(\frac{1+k}{2}\right)\right]^2 + \left[a - \frac{1}{2}\right]^2 + \left[a - \left(\frac{1}{2} + k\right)\right]^2\right\}$.
Since the term in braces reduces to $\frac{k^2}{2}$ we have that:

$$I_{.5}(a, a - k) \approx \frac{\log(e)}{2}k^2 = \frac{k^2}{2\ln(2)} \approx .72\ k^2 \qquad (9)$$

Given any point (a, b), we can always write it as $(a, a - (a - b))$, so $I_{.5}(a, b) \approx .72(a - b)^2$, and

$$C(a, b) \approx \frac{(a-b)^2}{2\ln(2)} \approx .72(a - b)^2 \qquad (10)$$

Recall that the determinant of the channel matrix is $a - b$.

We call the above $\frac{(a-b)^2}{2\ln(2)} = \frac{\log(e)}{2}(a - b)^2$ the *Quasi-Taylor approximation*, or simply the *QT approximation*, of capacity $C(a, b)$.

Graphically in Fig. 7, we can *see* that $\frac{(a-b)^2}{2\ln(2)}$ is a lower bound for $C(a, b)$, but we need to *prove* it. Fig. 8 shows how small the difference $C(a, b) - \frac{(a-b)^2}{2\ln(2)}$ is for most values of (a, b). We start by considering how the QT approximation compares to the capacity on the anti-diagonal (the binary symmetric channels, $(a, 1 - a), a \in [0, 1]$).

Lemma 4.1 *On the anti-diagonal* $\{(a, 1 - a) : a \in [0, 1]\}$ *the capacity is greater than or equal to the QT approximation with equality only at the point* $(.5, .5)$*, where the capacity is zero.*

Proof: First we consider the lower right hand part of the anti-diagonal: $\{(a, 1 - a) : a \in [.5, 1]\}$. Due to the symmetry of the channel matrix the capacity on the anti-diagonal is achieved ([1]) at $I_{.5}$ and the capacity is $C(a, 1-a) = 1 - h(a)$. We will compare $1 - h(a)$ to $\frac{(a-(1-a))^2}{2\ln(2)} = \frac{(2a-1)^2}{2\ln(2)}$. Let $f(a) = [1 - h(a)] - [\frac{(2a-1)^2}{2\ln(2)}] = [1 - \frac{h_e(a)}{\ln(2)}] - [\frac{(2a-1)^2}{2\ln(2)}]$, then $f'(a) = \frac{1}{\ln(2)}\left\{-\ln(\frac{1-a}{a}) - 4a + 2\right\}$. Since $f(.5) = 0$ we wish to show that $f'(a) > 0$ for $a > .5$. If we can show that $f'(a) \neq 0, a \in (.5, 1)$, that will suffice to show that $f'(a) > 0$ for $a > .5$, since $f(1) \approx .28$. If $f'(a) = 0, a \in (.5, 1)$, then there is some $a_0 \in (.5, 1)$ such that $\ln(\frac{1-a_0}{a_0}) = 2 - 4a_o$. We claim this is not possible since $2 - 4a > \ln(\frac{1-a}{a}), a \in (.5, 1)$. In the interval $[.5, 1]$, $2 - 4a$ ranges from 0 down to -2, via a straight line. On the other hand $\ln(\frac{1-a}{a})$ starts at 0 and logarithmically descends to $-\infty$ as a ranges over $[.5, 1)$. Therefore, we will show that $2 - 4a$ and $\ln(\frac{1-a}{a})$ do not intersect on $(.5, 1)$, and if that is true, we will have shown that $2 - 4a > \ln(\frac{1-a}{a}), a \in (.5, 1)$. We show that $2 - 4a$ and $\ln(\frac{1-a}{a})$ do not intersect by considering $\eta(a) = 2 - 4a - \ln(\frac{1-a}{a})$. We show that $\eta(x)$ is increasing and since $\eta(.5) = 0$ and $\lim_{a\to 1} \eta(a) = \infty$ that suffices to show that $2 - 4a$ and $\ln(\frac{1-a}{a})$ do not intersect in the region in question. We have that $\eta'(a) = -4 + \frac{1}{a(1-a)}$. If $\eta'(a) = 0$, we have that $4a^2 - 4a + 1 = 0$, which has a double root at $a = .5$. Therefore, $\eta'(a) \neq 0$ for $a \in (.5, 1)$, therefore $\eta'(a)$ is positive, so $\eta(x)$ is increasing in the region in question.

Now, by the symmetry of capacity with respect to positive and negative channels (swap a with $1 - a$), we can extend this argument to the other part of the anti-diagonal $\{(a, 1 - a) : a \in [0, .5]\}$. ■

Corollary 4.1 *The lower bound $\frac{(a-b)^2}{2\ln(2)}$ for capacity along the anti-diagonal is the best lower bound that is a constant multiple of $(a-b)^2$ in the sense that the lower bound agrees for the zero capacity channel $(.5,.5)$ and that the lower bound is strictly less than capacity for non-zero channels along the anti-diagonal.*

Proof: If there is a better lower bound it can be written as $(.5+\epsilon)\frac{(a-b)^2}{\ln(2)}, \epsilon > 0$. If we were to form $\eta'(a)$ as in the proof of the above lemma it would now be $-4+\delta+\frac{1}{a(1-a)}, \delta > 0$ which would have a root greater than .5. This cannot be, so we cannot allow $\epsilon > 0$. (Note: It is interesting to plot the difference of $1-h(a)$ and $(.5+\epsilon)\frac{(a-b)^2}{\ln(2)}, \epsilon > 0$ to see how small ϵ will subtly cause the difference to be negative near .5.) ■

Now we look at the leaves (lines) of constant channel matrix determinant that intersect with Regions 1 & 2, the positive channels. So we consider points of the form $\{(a, a-k)\}$. Since $C(a,a-k) \geq I_{.5}(a,a-k)$, if we can show that $I_{.5}(a,a-k)$ is greater than the QT approximation on lines of constant determinant (that intersect Regions 1& 2) then we will have also shown that capacity is greater than the QT approximation on the leaves of constant capacity. From before we know that $I_{.5}(a,a-k) = h\left(a-\frac{k}{2}\right) - \frac{1}{2}\left(h(a)+h(a-k)\right)$. The leaves of constant capacity that interesect Regions 1 & 2 are given by $\{(a,a-k) : a \in [k,1], k \in (0,1)\}$. Note, for $k=0$, which is not a positive channel, we are on the line of zero capacity and capacity agrees with the QT approximation. For $k=1$, which is a positive channel, we are on the degenerate leaf consisting of the single point $(1,0)$, here the capacity is one, which is obviously greater than the QT approximation. What happens on $\{(a,a-k) : a \in [k,1], k \in (0,1)\}$? In fact we restrict ourselves to Region 2 since capacity is symmetric on the leaves about the anti-diagonal. So we only look at $\{(a,a-k) : a \in [\frac{1+k}{2},1], k \in (0,1)\}$. (By our discussion of symmetry before we know for a point in Region 2, $C(\frac{1+k}{2}+\epsilon, \frac{1+k}{2}+\epsilon-k) = C(\frac{1+k}{2}+\epsilon, \frac{1-k}{2}+\epsilon) = C(1-(\frac{1-k}{2}+\epsilon), 1-(\frac{1+k}{2}+\epsilon)) = C(\frac{1+k}{2}-\epsilon, \frac{1-k}{2}-\epsilon) = C(\frac{1+k}{2}-\epsilon, \frac{1+k}{2}-\epsilon-k)$, which is the symmetric point in Region 1, and both are on the same leaf of constant k.)

Theorem 4.1 *The QT approximation is a lower bound for the capacity, which is strict for non-zero capacity channels.*

Corollary 4.2 $\frac{(a-b)^2}{2\ln(2)} \leq C(a,b) \leq |a-b|$

Proof (Thm. 4.1 and Cor. 4.2)**:** We fix k and restrict ourselves to Region 2, that is the segment $\{(a,a-k) : a \in [\frac{1+k}{2},1], k \in (0,1)\}$. Consider $g(a) := I_{.5}(a,a-k)$. We have $g'(a) = \log(e)\left\{\ln\left(\frac{1-a+k/2}{a-k/2}\right) - \frac{1}{2}\left[\ln\left(\frac{1-a}{a}\right)+\ln\left(\frac{1-a+k}{a-k}\right)\right]\right\} =$

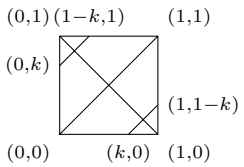

Figure 9. Upper line of slope 1: See Cor. 4.2

$\frac{1}{\ln(2)}\left\{\frac{1}{2}\ln\left(\frac{1-a+k/2}{a-k/2}\right)^2 - \frac{1}{2}\left[\ln\left(\frac{1-a}{a}\right)+\ln\left(\frac{1-a+k}{a-k}\right)\right]\right\}$

So $2\ln(2)\cdot g'(a) = \ln\left(\frac{1-a+k/2}{a-k/2}\right)^2 - \ln\left(\frac{(1-a)(1-a+k)}{a(a-k)}\right) = \ln\frac{\left(\frac{1-a+k/2}{a-k/2}\right)^2}{\left(\frac{(1-a)(1-a+k)}{a(a-k)}\right)} = \ln\frac{\left(\frac{1-2a+a^2-ak+k+(k^2/4)}{a^2-ak+(k^2/4)}\right)}{\left(\frac{1-2a+a^2-ak+k}{a^2-ak}\right)}$.

Let $A = 1-2a+a^2-ak+k$, $B = a^2-ak$, and $C = k^2/4$. So,

$$2\ln(2)\cdot g'(a) = \ln\frac{\left(\frac{A+C}{B+C}\right)}{\left(\frac{A}{B}\right)} = \ln\left(\frac{AB+BC}{AB+AC}\right)$$

Consider $B-A = \{2a-(1+k) : a \in (\frac{1+k}{2},1], k \in (0,1)\}$. We see that $B > A$, hence $\left(\frac{AB+BC}{AB+AC}\right) > 1$, and hence $g'(a) > 0$, for $\{a \in (\frac{1+k}{2},1] : k \in (0,1)\}$. So, on a line of constant determinant k, restricted to Region 2, we see that $I_{.5}$ is increasing from its value on the main diagonal $(\frac{1+k}{2}, \frac{1-k}{2})$ to its value at $(1,1-k)$. Similarly, we also have that it is decreasing along the same line restricted to Region 1 from its value at $(k,0)$ to its value on the main diagonal $(\frac{1+k}{2}, \frac{1-k}{2})$. Therefore, $I_{.5}$, restricted to a line of positive determinant, achieves a minimum along the anti-diagonal. Since, by Lemma 3.1, $I_{.5}$ is greater than the QT approximation on the anti-diagonal, we have that $I_{.5}$ is greater than the QT approximation (which is constant) on a line on positive capacity. Of course, on the line of zero determinant the capacity and QT approximation are both equal to zero.

Now consider line $L_n(k)$ of negative channels with determinant $-k = a-b < 0$. Let $a-b = -k$ and consider the line $L_p(k)$ positive channels with determinant k. These positive channels are of the form $(a,a-k), a \in [k,1]$ and are the locus of points in I^2 such that $b = a-k$, whereas the associated negative channels are of the form $(a,a+k), a \in [0,1-k]$ and are the locus of points in I^2 such that $b = a-(-k)$.

We can express $L_p(k)$ as $(\frac{1+k}{2}\pm\epsilon, \frac{1-k}{2}\pm\epsilon), \epsilon \in [0, \frac{1-k}{2}]$. Similarly, $L_n(k)$ is $(\frac{1-k}{2}\pm\epsilon, \frac{1+k}{2}\pm\epsilon), \epsilon \in [0, \frac{1-k}{2}]$.

By symmetry between Regions 1 and 3, and Regions 2 and 4, we have that the $C(\frac{1+k}{2}\pm\epsilon, \frac{1-k}{2}\pm\epsilon) = C(\frac{1-k}{2}\pm\epsilon, \frac{1+k}{2}\pm\epsilon)$. Since $k^2 = (-k)^2$ the QT approximation on $L_n(k)$ is the same as the QT approximation on $L_p(k)$ in a point by point manner with respect to ϵ. ■

It is important to note that in [8, Lemma 3.7] it was

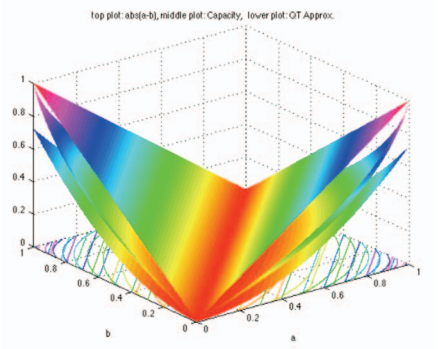

Figure 10. Capacity & two bounds

shown that capacity is lower bounded by $\frac{(a-b)^2}{e^2 \ln(2)} \approx .2(a-b)^2$, so [8] was the first to use a multiple of $(a-b)^2$ as a lower bound for capacity. Our result improves upon this existing result, providing the "best" possible constant multiple of the square of the determinant; even better from a practical standpoint, our result closely approximates capacity for moderate values of a and b.

Note, in this section we could have presented a "slicker" way by just looking at the Taylor series of the binary entropy function about $1/2$. However, in this conference paper we chose to present a more intuitive development, more in line with how we came about the result. As much as it pains the author, we put the mantra of pure mathematics on the back burner in this present exposition of the result.

5. Applications and Future Work

When a covert channel exists in a computer system, one may wish to introduce noise to decrease the capacity of the covert channel. However, noise should be introduced in a pragmatic manner so as to have as little ill-effect upon system performance as possible (this philosophy was demonstrated in [4, 5]). Alternatively, we may be faced with a situation in which noise may be abated at some cost in order to improve a channel. If the channel is too weak for our purposes, we may wish to reduce noise in order to improve performance. The approximation presented here may be used to guide us to a low cost improvement to produce an adequate channel.

Given a channel (a, b) we may wish to disturb this channel to a new channel (a', b') with a lesser (or greater) capacity while minimizing some metric between the two channels. For the standard L^2 metric the behavior of the level sets influences the choice of the modified channel (a', b'). The level sets of the capacity function are curved and the optimal modification is not obvious and, for us at least, can only be determined by numerical means. However, if we replace capacity with our QT bound, we see that we are now dealing with level sets that are parallel straight lines of slope one. The most "bang for the buck" in changing capacity is obtained by following straight lines (slope negative one) that are orthogonal to the level sets. It is interesting to compare this to the algebraic structures developed in [10].

We also plan to extend our analysis to more complicated channels and use the results in a geometric analysis of channel capacity and behavior. We have hope for this approach since Majani's result can be extended somewhat to more complicated channels than what we used in this paper.

6. Acknowledgements

We thank Gerard Allwein, Tanner Crowder, Ruth Irene, Myong Kang, and Richard Newman.

References

[1] Robert B. Ash. *Information Theory*. Dover, 1990.

[2] Konstantinos Chatzikokolakis and Keye Martin. A monotonicity principle for information theory. In *Proc. MFPS XXIV*, 2008. ENTCS, 218, pp. 111–129.

[3] Hermann J. Helgert. On a bound for channel capacity. *IEEE Trans. on Info. Th.*, 13:124–126, January 1967.

[4] Myong H. Kang and Ira S. Moskowitz. A pump for rapid, reliable, secure communication. In *Proc. ACM Conf. Comp. & Comm. Security'93*, pages 119–129.

[5] Myong H. Kang, Ira S. Moskowitz, and Daniel C. Lee. A network Pump. *IEEE Trans. Soft. Eng.*, 22:329–338, 1996.

[6] E. E. Majani. *A Model for the Study of Very Noisy Channels & Applications, 1988*. PhD thesis, Cal Tech.

[7] E.E. Majani and H. Rumsey. Two results on binary-input discrete memoryless channels. In *Proceedings Int. Symp. on Information Theory*, page 104, 1991.

[8] Keye Martin. Topology in information theory in topology. *Theoretical Computer Science*, 405:75–87, 2008.

[9] Keye Martin and Ira S. Moskowitz. Noisy timing channels with binary inputs and outputs. In *Information Hiding*, pages 124–144. LNCS 4437, July 2006.

[10] Keye Martin, Ira S. Moskowitz, and Gerard Allwein. Algebraic information theory for binary channels. In *Proc. MFPS XXII*, 2006. ENTCS, 158, pp. 289–306.

[11] Claude E. Shannon. A mathematical theory of communication. *BSTC*, 27:379–423,623–656, 1948.

[12] Claude E. Shannon. Some geometrical results in channel capacity. *Verband Deutsche Elektrotechniker Fachber*, 19(2):13–15, 1956.

[13] Richard A. Silverman. On binary channels and their cascades. In *IRE Trans. on Info. Th., pp. 19–27*, 1955.