Upper concave envelopes and auxiliary random variables

Chandra Nair

Abstract

We propose a new characterization of inner and outer bounds of some network information theoretic regions in terms of upper concave envelopes of certain functions of mutual information. While this characterization is related to the characterization using auxiliary random variables, it is shown that the new characterization can make computations of boundary points much simpler. Further this representation also leads to some new kinds of factorization inequalities concerning information theoretic quantities. It also provides some new pathways into proving optimality of certain achievable rate regions.

1 Introduction

In multiterminal information theory problems one wishes to compute optimal communication rates over a noisy network subject to some decoding constraints. A lot of the basic problems in this field has been open for more than three decades. In this paper we mainly focus on the *broadcast channel* [1], though some of the ideas presented here have also been applied to obtain new results in the *interference channel*, and can be applied to other settings as well. A focus of this paper is to formulate some of the inner and outer bounds using the language of concave envelopes, and then show that this formulation has some advantages over the traditional approach, i.e. using auxiliary random variables. Indeed this work is motivated by recent results [2, 3, 4, 5] regarding explicit computation of inner and outer bounds in some examples.

A broadcast channel refers to a communication scenario where a single sender wishes to communicate different messages to multiple receivers over a noisy medium. Let us consider the simple case of a two-receiver broadcast channel. In this setting a sender X wishes to communicate two messages, say M_1, M_2, to two receivers Y, Z over a discrete memoryless channel $\mathfrak{q}(y, z|x)$ so that receiver Y can decode message M_1 with high probability and receiver Z can decode M_2 with high probability.

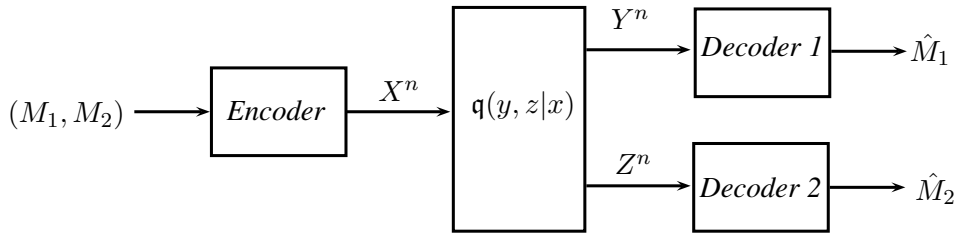

Figure 1: A broadcast channel

Let the channel have an input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}, \mathcal{Z}$ respectively. We assume that $|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}| < \infty$. A rate pair (R_1, R_2) is said to be achievable if there exists a sequence of codes consisting of

- An encoder which maps a message pair (m_1, m_2) according to $\Theta : [1 : 2^{nR_1}] \times [1 : 2^{nR_2}] \mapsto \mathcal{X}^n$,
- A decoder (at receiver Y) that maps a received sequence Y^n according to $\Psi : \mathcal{Y}^n \mapsto [1 : 2^{nR_1}]$,
- A decoder (at receiver Z) that maps a received sequence Z^n according to $\Phi : \mathcal{Z}^n \mapsto [1 : 2^{nR_2}]$, so that the probability of error defined as

$$P_e^{(n)} = \max_{\substack{m_1 \in [1:2^{nR_1}], \\ m_2 \in [1:2^{nR_2}]}} \mathrm{P}((\Psi(Y^n), \Phi(Z^n)) \neq (m_1, m_2) | X^n = \Theta(m_1, m_2))$$

  goes to zero as n goes to infinity.

Remark 1. Here we require that the maximal probability of error goes to zero. However it is known[6] that the region corresponding to average probability of error, i.e. the decoding error averaged over a uniform distribution of the message pairs, matches the region corresponding to the maximal probability of error.

The closure of the set of all achievable rate pairs is known as the *capacity region*. A computable[1] characterization of the capacity region for the two-receiver broadcast channel is still unknown, and is one of the most fundamental open problems in the field of multi-user information theory. In particular there are computable inner bounds and outer bounds to the capacity region of the two-receiver broadcast channel.

Motivation

The motivation for this work comes from the following observation: recently established capacity regions [7, 8] in the broadcast channels rely on the properties of the *extremal* auxiliary random variables, i.e. auxiliary random variables that result in points on the boundary of the capacity region. These auxiliary random variables were shown to possess additional properties and the establishment of the capacity regions utilized these properties. Similarly, in a few other cases as well, the extremal auxiliary random variables were observed to have a simple structure (examples will be given in later sections) while the proofs of these were quite involved. In this work we present a different representation using the idea of concave envelopes. While concave envelopes (or convex hulls) and auxiliary random variables were always known to be connected (for example, the use of Caratheodory's theorem in bounding the cardinality of the auxiliary random variables); the principle advantage of this representation is that the only auxiliary random variables that show up in computation of the concave envelopes are the extremal ones. In some sense this is an interchange of operations: traditionally one writes a region in terms of auxiliary random variables and then solves an optimization problem if one is interested in computing the boundary points; in the current form, we first obtain the concave envelopes (indirectly an optimization over the auxiliary random variables) and then write the region in terms of concave envelopes. As we shall see below (an interested reader may also refer to the related work in [9, 10, 11]) this representation has some definite advantages.

Outline

In the next section we will revisit superposition coding region, one of the earliest examples to use the auxiliary random variables idea, and discuss some properties of *extremal* distributions in some

[1] A rate region is said to be computable if for every $\epsilon > 0$, one can find a bounded time T_ϵ such that the region can be approximated to within ϵ in time T_ϵ.

well-known examples here. Then we will give an equivalent representation using concave envelopes and show how this representation yields simple proofs of the previously mentioned extremal distributions. In the next section, we will briefly mention other settings where the concave envelope representations naturally arise. Subsequently we will introduce the factorization property of some of the concave envelopes discussed previously and show the role of these factorization inequalities in establishing the optimality of the regions involved. We also present a conjecture, which if established, would yield the capacity region of a degraded message setting with three receivers, a problem that has been open for some time.

2 Superposition coding region

The superposition coding strategy [1] was introduced for the degraded broadcast channel, i.e. the receiver Z is a noisier version of receiver Y or mathematically $X \to Y \to Z$ forms a Markov chain.

2.1 Superposition coding region for degraded broadcast channel

Superposition coding region, $\mathcal{S}$, for a degraded broadcast channel is the union of all non-negative rate pairs (R_1, R_2) satisfying:

$$\begin{aligned} R_2 &\leq I(V;Z) \\ R_1 &\leq I(X;Y|V), \end{aligned}$$

over all (V, X) such that $V \to X \to Y \to Z$ forms a Markov chain. Note that by the data-processing inequality we have $I(V;Y) \geq I(V;Z)$. (This inequality allows receiver Y to decode the message M_2.) In the absence of such an ordering one usually requires the constraint $R_1 + R_2 \leq I(X;Y)$ as well.

The superposition region for degraded broadcast channels satisfies certain properties:

1. $\mathcal{S}$ is a convex, closed and bounded set belonging to the positive quadrant.
2. For a fixed μ let
$$V_\mu = \max_{(R_1,R_2)\in\mathcal{S}} R_1 + \mu R_2.$$
Then we can recast $\mathcal{S}$ as
$$\mathcal{S} = \cap_{\mu\geq 1}\{(R_1, R_2) \in \mathbb{R}_+^2 : R_1 + \mu R_2 \leq V_\mu\}.$$

The second condition says that the region can be considered as the intersection of the supporting hyperplanes, which follows from the convexity.

Remark 2. The reason for not considering $\mu < 1$ is the following: due to the degraded nature of the receivers, note that if $(R_1, R_2) \in \mathcal{S}$ then $(R_1 + R_2, 0) \in \mathcal{S}$. When $\mu < 1$ note that $R_1 + \mu R_2 \leq (R_1 + R_2) + \mu \cdot 0 = R_1 + 1 \cdot R_2$. Thus there will not be any new boundary point obtained by considering $\mu < 1$. Geometrically if R_1 is plotted along the X-axis and R_2 along the Y-axis, then the slope of the tangent to the capacity region with the X-axis will be at least 135 degrees. Hence we only need to consider supporting hyperplanes of the form $R_1 + \mu R_2$ with $\mu \geq 1$.

2.2 Superposition coding region for some examples

In this section, we exhibit the superposition coding region for some known examples. Note the simplicity of the structure of the optimal (V, X) in all of these examples. In the following examples we are interested in computing V_μ which in turn characterizes $\mathcal{S}$ as mentioned earlier.

Remark: The previously known rigorous proofs of the optimality of the (V, X) in each of these cases were non-trivial exercises.

Example 1: The degraded BSC channel

Consider the following degraded broadcast channel depicted in the Figure 2. The superposition coding region can be obtained by considering just $V \to X \sim BSC(s)$, shown by Wyner and Ziv [12] as predicted by Cover [1]. This in particular implies that

$$V_\mu = \max_{(R_1,R_2)\in\mathcal{S}} R_1 + \mu R_2 = \max_{s\in[0,\frac{1}{2}]} h(s*p) - h(p) + \mu\big(1 - h(s*p*q)\big),$$

where "$*$" refers to binary convolution defined as $a*b = a(1-b)+b(1-a)$, and $h(x) := -x\log_2 a - (1-x)\log_2(1-x)$ refers to the binary entropy function.

The proof of this fact turned out to be non-trivial and involved showing the convexity of the following function: $h(p*h^{-1}(x))$ in x. Further more, this technique could not be extended towards other symmetric channels.

Recently [8, 13], a new *symmetrization* based argument, proved that to maximize $R_1 + \mu R_2$ for any $\mu > 1$ it suffices to consider $V \to X$ to be $BSC(s)$. Further, this proof could also be generalized to binary input symmetric output channels. In the next section, we will use a similar idea but give an even simpler proof of this same result.

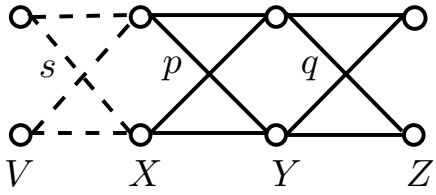

Figure 2: A degraded BSC

Example 2: The degraded Z channel

Consider the degraded Z-channel shown in Figure 3. It turns out that $\max R_1 + \mu R_2$ can be computed [14] by just considering $V \to X$ being another Z-channel. This is again a very arduous proof. We will show a very simple proof of this fact in the next section.

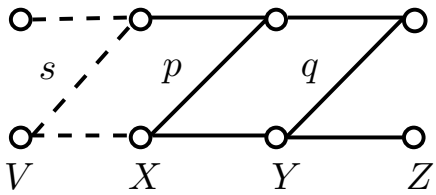

Figure 3: A degraded Z-channel

That is, in particular,

$$V_\mu = \max_{s,v\in[0,1]} vh((1-s)(1-p)) - v(1-s)h(1-p) + \mu\left(h(v(1-s)(1-p)(1-q)) - vh\big((1-s)(1-p)(1-q)\big)\right),$$

where $v = \mathrm{P}(V=1)$.

Example 3: Degraded Gaussian channel

Consider the degraded additive Gaussian channel shown in Figure 4. Here $Y = X+N_1, Z = Y+N_2$ where N_1, N_2 are independent Gaussian random variables. To compute the superposition coding region it suffices [15] to consider $X = U + V$ where U and V are independent Gaussian random variables. The proof of this result uses entropy power inequality, a highly non-trivial result.

The technique we employ can again be used to show this result without needing to resort to the Entropy Power Inequality. This argument can be inferred from the results in [11].

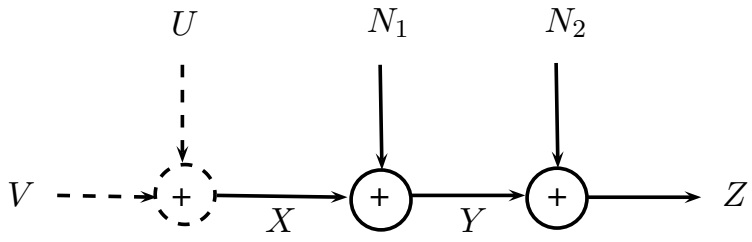

Figure 4: A Gaussian degraded broadcast channel

2.3 A representation of the region using concave envelopes

2.3.1 On upper concave envelopes

Let $f(x)$ be a function defined on a convex subset $\mathcal{D}$ of some Hilbert space. The notation $\mathfrak{C}[f(x)]$ refers to the *upper concave envelope* of a function $f(x)$ (on the domain $\mathcal{D}$). The upper concave envelope, $g(x)$, of the function $f(x)$ can be expressed in many equivalent ways:

- $g(x)$ is the smallest concave function such that $g(x) \geq f(x),\ \forall x \in \mathcal{D}$. Here smallest is defined in a point-wise sense.
- $g(x) = \sup_{\mu_x} \int f d\mu_x$, where μ_x is the set of all probability measures on $(\mathcal{D}, \Sigma_{\mathcal{D}})$ (where Σ_D is the Borel σ-field) with mean value x. If $\mathcal{D}$ is a subset of some finite d-dimensional Euclidean space, then by Caratheodory's theorem

$$g(x) = \sup_{x_i, p_i} \sum_{i=1}^{d+1} p_i f(x_i)$$

  where $\{p_i\}$ is a $(d+1)$-dimensional probability vector and $\sum_{i=1}^{d+1} p_i x_i = x$.

In general, computing the upper concave envelope is a global operation and hence it is probably best represented by the supporting hyperplanes to the original function.

2.3.2 Superposition coding region using concave envelopes

Given a broadcast channel and $\mu > 1$ consider the following function of $p(x)$ defined as

$$S(X) := \mathfrak{C}[I(X;Y) - \mu I(X;Z)].$$

Note, the set $\mathcal{D}$ used here is the set of all probability distributions on $\mathcal{X}$.

Claim 1. *For a degraded broadcast channel $X \to Y \to Z$ and $\mu > 1$*

$$\max_{(R_1,R_2)\in\mathcal{C}} R_1 + \mu R_2 = \max_{p(x)} \left(\mu I(X;Z) + \mathfrak{C}[I(X;Y) - \mu I(X;Z)]\right).$$

A standalone proof (i.e. not dependent on the current proofs by explicitly identifying the auxiliaries in the usual representation) is not known, and is a very interesting question to explore.

Proof. This proof directly follows from our explicit knowledge of the capacity region characterization using auxiliary random variables.

$$\begin{aligned}
&\max_{(R_1,R_2)\in\mathcal{C}} R_1 + \mu R_2 \\
&= \max_{\substack{p(v,x) \\ V\to X\to(Y,Z)}} \mu I(V;Z) + I(X;Y|V) \\
&= \max_{\substack{p(v,x) \\ V\to X\to(Y,Z)}} \mu I(X;Z) + I(X;Y|V) - \mu I(X;Z|V) \\
&= \max_{p(x)} \left(\mu I(X;Z) + \max_{p(v|x)} (I(X;Y|V) - \mu I(X;Z|V))\right) \\
&= \max_{p(x)} \left(I(X;Z) + \mathfrak{C}[I(X;Y) - \mu I(X;Z)]\right).
\end{aligned}$$

□

Thus we have expressed V_μ in an alternate form according to

$$V_\mu = \max_{p(x)} \left(\mu I(X;Z) + \mathfrak{C}[I(X;Y) - \mu I(X;Z)]\right).$$

Remark 3. As said in the motivation, it is clear that the only auxiliary random variables that show up in the computation of the concave envelopes are the extremal ones. As mentioned previously, some of the recent capacity results that were established for broadcast channels [7, 8] exploited the properties of the extremal auxiliary random variables to fashion a converse. Hence focusing on extremal random variables, we hope to extract more insight into the finer properties of the boundary points of the various regions.

2.4 Evaluation of the superposition coding using the concave envelope representation

In all of these examples, we wish to compute for $\mu > 1$,

$$\max_{p(x)} \mu I(X;Z) + \mathfrak{C}[I(X;Y) - \mu I(X;Z)].$$

Further we also wish to compute the optimal (V^*, X^*) that yield $\mathfrak{C}[I(X;Y) - \mu I(X;Z)]$ at the optimal input distribution.

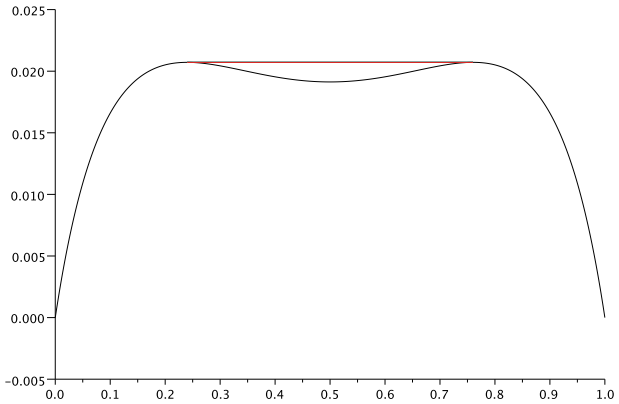

Figure 5: $I(X;Y) - \mu I(X;Z)$: Degraded BSC with $p = 0.1, q = 0.1, \mu = 1.6$

Example 1: Degraded BSC

Consider the channel in Figure 2. Observe that

$$\begin{aligned}&\max_{p(x)} \big(\mu I(X;Z) + \mathfrak{C}[I(X;Y) - \mu I(X;Z)]\big)\\ &\quad \le \max_{p(x)} \mu I(X;Z) + \max_{p(x)} \mathfrak{C}[I(X;Y) - \mu I(X;Z)].\end{aligned}$$

Now observe that $I(X;Z)$ attains its global maximum at $\mathrm{P}(X = 0) = \frac{1}{2}$. Further, note that $\mathfrak{C}[I(X;Y) - \mu I(X;Z)]$ also attains the global maximum at $\mathrm{P}(X = 0) = \frac{1}{2}$, due to symmetry of $I(X;Y) - \mu I(X;Z)$ about $\mathrm{P}(X = 0) = \frac{1}{2}$. This is because the function $I(X;Y) - \mu I(X;Z)$ attains a global maximum at points $\mathrm{P}(X = 0) = s, \mathrm{P}(X = 0) = 1 - s$ for some $0 \le s \le \frac{1}{2}$. Further (V, X) is a doubly symmetric binary source. Hence consider a binary and uniform V^* such that $\mathrm{P}(X = 0|V^* = 0) = s$ and $\mathrm{P}(X = 0|V^* = 1) = 1 - s$. Note that with this choice we attain the global maximum of $\mathfrak{C}[I(X;Y) - \mu I(X;Z)]$ at $\mathrm{P}(X = 0) = \frac{1}{2}$.

In Figure 2.4, we plot the function $I(X;Y) - \mu I(X;Z)$ and its concave envelope (*in red*) for the choice of parameters $p = 0.1, q = 0.1, \mu = 1.6$. Notice the symmetry about $\mathrm{P}(X = 0) = \frac{1}{2}$.

Thus, we obtain in a very simple fashion, the well-known result

$$V_\mu = \max_{(R_1,R_2)\in\mathcal{S}} R_1 + \mu R_2 = \max_{s\in[0,\frac{1}{2}]} h(s * p) - h(p) + \mu\big(1 - h(s * p * q)\big).$$

Example 2: Degraded Z-channel

Consider the channel in Figure 3. Observe that if $\mathrm{P}(X = 1) = x$ then

$$f(x) := I(X;Y) - \mu I(X;Z) = h(x\bar{p}) - xh(\bar{p}) - \mu\big(h(x\bar{p}\bar{q}) - xh(\bar{p}\bar{q})\big).$$

Here $\bar{p} = 1 - p$, $\bar{q} = 1 - q$. Consider the second derivative of $f(x)$ and we obtain that

$$f''(x) \ln 2 = -\frac{\bar{p}}{x(1 - \bar{p}x)} + \mu \frac{\bar{p}\bar{q}}{x(1 - \bar{p}\bar{q}x)}.$$

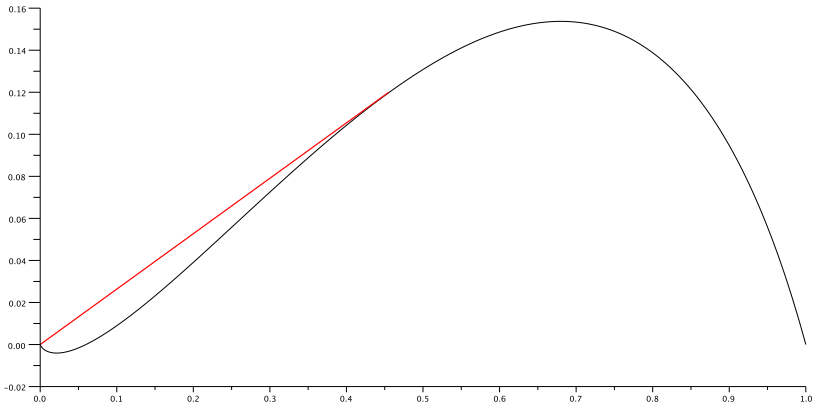

Figure 6: $I(X;Y) - \mu I(X;Z)$: Degraded Z with $p = 0.1, q = 0.5, \mu = 2.3$

Observe that this function has at most one zero in $x \in [0,1]$, and this zero occurs when $0 \leq \mu\bar{q} - 1 \leq \bar{p}\bar{q}(\mu - 1)$. For this regime of parameters observe that whenever $0 \leq x \leq x_* := \frac{\mu\bar{q}-1}{\bar{p}\bar{q}(\mu-1)}$ the function $f(x)$ is convex and beyond that it is concave. Therefore its concave envelope will be of the following form: for $0 \leq x \leq x_t$, the concave envelope will be a straight line, tangential to $f(x)$ at $x_t (\geq x_*)$ and after that it will follow the curve.

Hence for any $\mathrm{P}(X = 1) \in [0,1]$, the concave envelope is either the function $f(x)$ at $\mathrm{P}(X = 1) = x$ or it is a convex combination of two points $x_1 = 0, x_2 = x_t$, i.e. $\mathrm{P}(X = 1|V = 0) = 0, \mathrm{P}(X = 1|V = 1) = x_t$ and the distribution on V satisfies $x_t \,\mathrm{P}(V = 1) = x$. Thus the optimal $V \to X$ is also a Z-channel.

In Figure 6, we plot the function $I(X;Y) - \mu I(X;Z)$ and its concave envelope (*in red*) for the choice of parameters $p = 0.1, q = 0.5, \mu = 2.3$.

Remark 4. The degraded Gaussian example has been be dealt with in [11] for a much more general setting. The key idea of the proof is to show the optimality of Gaussian using a *factorization of concave envelopes*. Notice that the proofs of the optimality we developed in this section using concave envelopes is much simpler than the original proofs. Further the concave envelopes have a reasonably simple characterization. This is a key and continuing observation across other examples.

3 Concave envelopes arising in other settings

Here we consider a couple of other network information theory settings and we define the appropriate concave envelopes that characterize the various regions. Please see [16] for the various definitions of the settings and proofs of their optimality.

3.1 Channels with state known non-causally at the encoder

This setting, shown in Figure 7, represents a communication scenario over a state-dependent channel where the state distribution varies i.i.d. according to a fixed distribution $p(s)$. The goal is for a single sender X to maximize the reliable transmission rate to a receiver Y over this state dependent

channel. One also assumes that the entire state sequence S^n is known non-causally at the encoder, like in a scenario of coding over memory with defects.

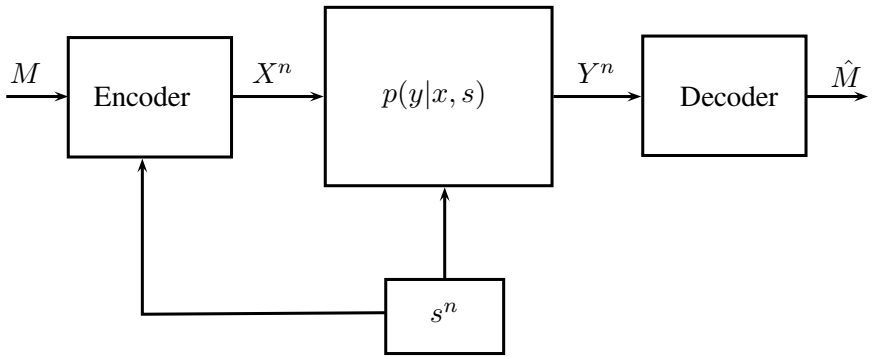

Figure 7: Point-to-point system with state information at encoder

Here the capacity[17] or maximum achievable rate is given by

$$\begin{aligned}
&\max_{p(u,x|s)} I(U;Y) - I(U;S) \\
&= \max_{p(u,x|s)} H(Y) - H(S) + (H(S|U) - H(Y|U)) \\
&= \max_{p(x|s)} H(Y) - H(S) + \mathfrak{C}[H(S) - H(Y)].
\end{aligned}$$

Thus $\max_{p(x|s)} H(Y)-H(S)+\mathfrak{C}[H(S)-H(Y)]$ yields an alternate representation of the capacity.

Given a $p(s)$ the function $H(S) - H(Y)$ is a function of $p(x|s)$, i.e. it is a function of $|\mathcal{S}|$ probability vectors of size $|\mathcal{X}|$. Hence the concave envelope is taken over this domain. Further $H(S) - H(Y)$ is a convex function of $p(x|s)$ and hence the concave envelope of $(H(S) - H(Y))$ is supported at its extreme points. This implies that it suffices to consider $p(x|u, s)$ to be an vertex of the probability simplex, or in other words $x = f(u, s)$, a deterministic function of u and s.

3.2 Wiretap Channel

This setting first considered in a seminal paper by Wyner [18], shown in Figure 8, represents a scenario where a sender X wishes to communicate a message M reliably to a receiver Y while keeping it secret from an eavesdropper Z.

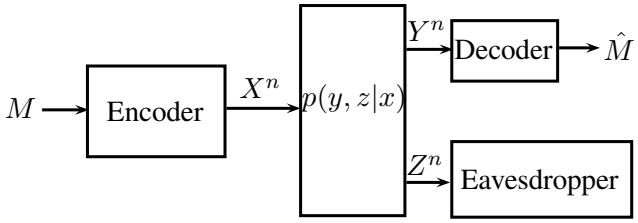

Figure 8: The wiretap channel

The maximal achievable secrecy rate [19] is given by

$$\begin{aligned}
&\max_{\substack{p(u,x):\\ U\to X\to(Y,Z)}} I(U;Y)-I(U;Z)\\
&=\max_{\substack{p(q,u,x):\\ (Q,U)\to X\to(Y,Z)}} I(U;Y|Q)-I(U;Z|Q)\\
&=\max_{p(x)} \mathfrak{C}\big[I(X;Y)-I(X;Z)+\mathfrak{C}[I(X;Z)-I(X;Y)]\big].
\end{aligned}$$

Thus $\max_{p(x)} \mathfrak{C}\big[I(X;Y)-I(X;Z)+\mathfrak{C}[I(X;Z)-I(X;Y)]\big]$ yields an alternate expression to the secrecy rate.

Remark 5. It is worth noting that although the expressions for the capacity rate look similar in terms of auxiliary random variables for both the above settings, the concave envelope representations are different and so is the domain on which the concave envelope is computed.

In the previous sections we described how one may obtain alternate representations of the various capacity regions and rates using the concave envelope representation. In the next section we will describe how one may verify the optimality of such representations using the *factorization* approach[2].

4 Factorization of concave envelopes

Testing of the optimality of expressions using concave envelopes follows a reasonably straightforward path in general. The first step is to verify that the n-letter forms of the expressions approach capacity (this is the easy part, which usually follows from Fano's inequality). The second step is to verify that normalized two-letter form of the concave envelope expression does not improve on the single-letter expression rate. Then by repeated application of this doubling idea, we show that there is not increase along the dyadic powers and hence combining this with the first observation, we have the required optimality.

We generalize the second step to say that a particular concave envelope expression $F_{\mathfrak{q}}(X)$ defined over the input distributions of a channel possesses the *factorization property* if the corresponding expression defined over the joint input distributions of a product channel $\mathfrak{q}_1\times\mathfrak{q}_2$ satisfy

$$F_{\mathfrak{q}_1\times\mathfrak{q}_2}(X_1,X_2)\le F_{\mathfrak{q}_1}(X_1)+F_{\mathfrak{q}_2}(X_2),$$

where the first function $F_{\mathfrak{q}_1\times\mathfrak{q}_2}(X_1,X_2)$ is evaluated at $p(x_1,x_2)$ and the functions $F_{\mathfrak{q}_1}(X_1), F_{\mathfrak{q}_2}(X_2)$ are evaluated at the marginal distributions $p_1(x_1):=\sum_{x_2}p(x_1,x_2), p_2(x_2):=\sum_{x_1}p(x_1,x_2)$ respectively.

Remark 6. Note that factorization property in particular implies that the normalized two-letter form of the concave envelope expression does not improve on the single-letter expression. However, it is stronger than that, as it establishes the sub-additivity with respect to any product channel and for any pair of distributions and not just the maximizers.

Remark 7. The *factorization* inequalities have, to the best of the knowledge of the author, not been observed earlier[3] for the functions that we describe in the sections below. These are inequalities

[2]The factorization approach was briefly developed in [9] and [10]; and it was exploited further in [11].

[3]One of these inequalities is mentioned in a related work of the author[11] while different factorization inequalities have also been proved for special classes of channels in [9], [10].

observed by the concave envelopes and not by the underlying functions. The proofs of the factorization inequalities here are motivated by the usual converses and in particular the Csiszar-sum lemma. However the statement of the inequalities themselves suggest the existence of alternate proof forms.

4.1 Illustrations of optimality

In this section we will illustrate the above general technique for the three settings for which we derived the concave envelope representations earlier.

4.1.1 Degraded broadcast channel

We wish to show that, when $\mu \geq 1$,

$$R_1 + \mu R_2 \leq \mu I(X;Z) + \mathfrak{C}[I(X;Y) - \mu I(X;Z)], \tag{1}$$

for any rate pair (R_1, R_2) that belongs to the capacity region.

Remark 8. The above statement, in fact, holds for any broadcast channel and it is only for the achievability of the rate pairs that we require that the broadcast channel be degraded.

First note the following from Fano's inequality:

$$\begin{aligned}
R_1 + \mu R_2 &\leq \frac{1}{n}\big(\mu I(M_2;Z^n) + I(M_1;Y^n|M_2)\big) + \epsilon_n \\
&= \frac{1}{n}\big(\mu I(M_1,M_2;Z^n) + I(M_1;Y^n|M_2) - \mu I(M_1;Z^n|M_2)\big) + \epsilon_n \\
&\leq \frac{1}{n}\big(\mu I(X^n;Z^n) + I(X^n;Y^n|M_2) - \mu I(X^n;Z^n|M_2)\big) + \epsilon_n \\
&\leq \frac{1}{n}\big(\mu I(X^n;Z^n) + \mathfrak{C}[I(X^n;Y^n) - \mu I(X^n;Z^n)]\big) + \epsilon_n,
\end{aligned}$$

where the last inequality follows from the definition of the concave envelope. Since $\epsilon_n \to 0$ as $n \to \infty$ we have that the normalized n-letter version is indeed an outer bound.

Thus, the single-letter inequality in (1) will follow (from 2-letter to 2^k-letter is just a repeated application to larger channels) if we show that for any product broadcast channel we have

$$\begin{aligned}
&\mu I(X_1,X_2;Z_1,Z_2) + \mathfrak{C}[I(X_1,X_2;Y_1,Y_2) - \mu I(X_1,X_2;Z_1,Z_2)] \\
&\quad \leq \mu I(X_1;Z_1) + \mathfrak{C}[I(X_1;Y_1) - \mu I(X_1;Z_1)] + \mu I(X_2;Z_2) + \mathfrak{C}[I(X_2;Y_2) - \mu I(X_2;Z_2)].
\end{aligned}$$

Since the channel is a product channel we have $Z_1 \to X_1 \to X_2 \to Z_2$ forming a Markov chain, implying that

$$I(X_1,X_2;Z_1,Z_2) = I(X_1;Z_1) + I(X_2;Z_2) - I(Z_1;Z_2) \leq I(X_1;Z_1) + I(X_2;Z_2).$$

The proof is completed if we show that

$$\begin{aligned}
&\mathfrak{C}[I(X_1,X_2;Y_1,Y_2) - \mu I(X_1,X_2;Z_1,Z_2)] \\
&\quad \leq \mathfrak{C}[I(X_1;Y_1) - \mu I(X_1;Z_1)] + \mathfrak{C}[I(X_2;Y_2) - \mu I(X_2;Z_2)].
\end{aligned}$$

Observe that for any $U \to (X_1, X_2) \to (Y_1, Y_2, Z_1, Z_2)$ where $(Y_1, Z_1) \to X_1 \to X_2 \to (Y_2, Z_2)$ is Markov (i.e. the underlying channel is a product channel) we have by routine manipulations

$$\begin{aligned}
&I(X_1, X_2; Y_1, Y_2|U) - \mu I(X_1, X_2; Z_1, Z_2|U) \\
&\quad \overset{(a)}{=} I(X_1; Y_1|U, Y_2) + I(X_2; Y_2|U) - \mu I(X_1; Z_1|U) - \mu I(X_2; Z_2|U, Z_1) \\
&\quad \overset{(b)}{=} I(X_1; Y_1|U, Y_2) - \mu I(X_1; Z_1|U, Y_2) \\
&\qquad + I(X_2; Y_2|U, Z_1) - \mu I(X_2; Z_2|U, Z_1) - (\mu - 1) I(Z_1; Y_2|U) \\
&\quad \leq \mathfrak{C}[I(X_1; Y_1) - \mu I(X_1; Z_1)] + \mathfrak{C}[I(X_2; Y_2) - \mu I(X_2; Z_2)],
\end{aligned}$$

where the last inequality follows from the definition of the concave envelope and that $(U, Y_2) \to X_1 \to (Y_1, Z_1)$ and $(U, Z_1) \to X_2 \to (Y_2, Z_2)$ form Markov chains. The equality (a) follows from chain rule and the following Markov chains: $(U, X_2, Y_2) \to X_1 \to Y_1$, $(U, X_1) \to X_2 \to Y_2$, $(U, X_1, Z_1) \to X_2 \to Z_2$, and $(U, X_2) \to X_1 \to Z_1$. These Markov chains are in turn a consequence of the following two Markov chains: $U \to (X_1, X_2) \to (Y_1, Y_2, Z_1, Z_2)$ and $(Y_1, Z_1) \to X_1 \to X_2 \to (Y_2, Z_2)$. For equality (b) as $(U, Z_1) \to X_2 \to Y_2$ is Markov we have $I(X_2; Y_2|U) = I(X_2; Y_2|U, Z_1) + I(Z_1; Y_2|U)$. Similarly $I(X_1; Z_1|U) = I(X_1; Z_1|U, Y_2) + I(Y_2; Z_1|U)$.

Maximizing the left hand side over the choices of U we obtain that

$$\begin{aligned}
&\mathfrak{C}[I(X_1, X_2; Y_1, Y_2) - \mu I(X_1, X_2; Z_1, Z_2)] \\
&\quad \leq \mathfrak{C}[I(X_1; Y_1) - \mu I(X_1; Z_1)] + \mathfrak{C}[I(X_2; Y_2) - \mu I(X_2; Z_2)],
\end{aligned}$$

as desired.

4.1.2 Channels with state known non-causally at the encoder

Here we wish to show that the maximum sum-rate is bounded above by

$$\max_{p(x|s)} H(Y) - H(S) + \mathfrak{C}[H(S) - H(Y)].$$

Clearly from Fano's inequality

$$\begin{aligned}
R &\leq \frac{1}{n}\Big(I(M; Y^n) - I(M; S^n)\Big) + \epsilon_n \\
&= \frac{1}{n}\Big(H(Y^n) - H(S^n) + \big(H(S^n|M) - H(Y^n|M)\big)\Big) + \epsilon_n \\
&\leq \frac{1}{n}\Big(H(Y^n) - H(S^n) + \mathfrak{C}[H(S^n) - H(Y^n)]\Big) + \epsilon_n.
\end{aligned}$$

The proof is completed if we show the factorization over product channelss. Towards this end first observe that $H(Y_1, Y_2) - H(S_1, S_2) \leq H(Y_1) - H(S_1) + H(Y_2) - H(S_2)$, since independence of state realizations implies $H(S_1, S_2) = H(S_1) + H(S_2)$. Further observe that

$$\begin{aligned}
&H(S_1, S_2|U) - H(Y_1, Y_2|U) \\
&\quad = H(S_1|U, Y_2) - H(Y_1|U, Y_2) + H(S_2|U, S_1) - H(Y_2|U, S_1) \\
&\quad \leq \mathfrak{C}[H(S_1) - H(Y_1)] + \mathfrak{C}[H(S_2) - H(Y_2)],
\end{aligned}$$

and by taking the maximum over all choices of $U \to (S_1, S_2, X_1, X_2) \to (Y_1, Y_2)$ we get the desired inequality. The equality above follows from $H(S_1|U) - H(Y_2|U) = H(S_1|U, Y_2) - H(Y_2|U, S_1)$.

4.1.3 Wiretap Channel

In this case we wish to show that

$$C \leq \max_{p(x)} \mathfrak{C}\big[I(X;Y) - I(X;Z) + \mathfrak{C}[I(X;Z) - I(X;Y)]\big].$$

Again from Fano's inequality and the secrecy constraints we have that

$$\begin{aligned} C &\leq \frac{1}{n}\big(I(M_1;Y^n) - I(M_1;Z^n)\big) + \epsilon_n \\ &\leq \frac{1}{n}\big(I(X^n;Y^n) - I(X^n;Z^n) + I(X^n;Z^n|M_1) - I(X^n;Y^n|M_1)\big) + \epsilon_n \\ &\leq \frac{1}{n}\mathfrak{C}\big[I(X^n;Y^n) - I(X^n;Z^n) + \mathfrak{C}[I(X^n;Z^n) - I(X^n;Y^n)]\big] + \epsilon_n. \end{aligned}$$

To show the factorization, i.e.

$$\begin{aligned} &\mathfrak{C}\big[I(X_1,X_2;Y_1,Y_2) - I(X_1,X_2;Z_1,Z_2) + \mathfrak{C}[I(X_1,X_2;Z_1,Z_2) - I(X_1,X_2;Y_1,Y_2)]\big] \\ &\quad \leq \mathfrak{C}\big[I(X_1;Y_1) - I(X_1;Z_1) + \mathfrak{C}[I(X_1;Z_1) - I(X_1;Y_1)]\big] \\ &\qquad + \mathfrak{C}\big[I(X_2;Y_2) - I(X_2;Z_2) + \mathfrak{C}[I(X_2;Z_2) - I(X_2;Y_2)]\big], \end{aligned}$$

observe that for any $(Q,U) \to (X_1,X_2) \to (Y_1,Y_2,Z_1,Z_2)$ where $(Y_1,Z_1) \to X_1 \to X_2 \to (Y_2,Z_2)$ is Markov (i.e. the underlying channel is a product channel) we have by routine manipulations

$$\begin{aligned} &I(X_1,X_2;Y_1,Y_2|Q) - I(X_1,X_2;Z_1,Z_2|Q) + I(X_1,X_2;Z_1,Z_2|U,Q) - I(X_1,X_2;Y_1,Y_2|U,Q) \\ &\quad = I(X_1;Y_1|Q,Y_2) - I(X_1;Z_1|Q) + I(X_1;Z_1|U,Q) - I(X_1;Y_1|U,Q,Y_2) \\ &\qquad + I(X_2;Y_2|Q) - I(X_2;Z_2|Q,Z_1) + I(X_2;Z_2|U,Q,Z_1) - I(X_2;Y_2|U,Q) \\ &\quad = I(X_1;Y_1|Q,Y_2) - I(X_1;Z_1|Q,Y_2) + I(X_1;Z_1|U,Q,Y_2) - I(X_1;Y_1|U,Q,Y_2) \\ &\qquad + I(X_2;Y_2|Q,Z_1) - I(X_2;Z_2|Q,Z_1) + I(X_2;Z_2|U,Q,Z_1) - I(X_2;Y_2|U,Q,Z_1) \\ &\quad \leq \mathfrak{C}\big[I(X_1;Y_1) - I(X_1;Z_1) + \mathfrak{C}[I(X_1;Z_1) - I(X_1;Y_1)]\big] \\ &\qquad + \mathfrak{C}\big[I(X_2;Y_2) - I(X_2;Z_2) + \mathfrak{C}[I(X_2;Z_2) - I(X_2;Y_2)]\big]. \end{aligned}$$

Thus taking maximum over Q,U we obtain the desired factorization inequality. The first equality above follow from our Markov chain assumptions similar to the equality labeled (a) in Section 4.1.1, while the second equality follows since the following two inequalities

$$\begin{aligned} I(X_2;Y_2|Q) - I(X_1;Z_1|Q) &= I(X_2;Y_2|Q,Z_1) - I(X_1;Z_1|Q,Y_2), \\ I(X_2;Y_2|U,Q) - I(X_1;Z_1|U,Q) &= I(X_2;Y_2|U,Q,Z_1) - I(X_1;Z_1|U,Q,Y_2), \end{aligned}$$

hold from our Markov chain assumptions similar to the equality labeled (b) in Section 4.1.1.

In the next section we present a factorization inequality that appears to be true from numerical simulations and whose establishment would solve an interesting problem that has been open for about three decades.

4.2 A conjecture

We now present a conjecture that is related to showing the optimality of superposition coding for a three receiver broadcast channel with two degraded message sets. In this communication setting, message M_0 is required to be decoded by all three receivers Y_1, Y_2, Y_3 while the message M_1 is only

required to be decoded by receivers Y_1, Y_2. The best known achievable region in this scenario is given by the following: the union of rate pairs R_0, R_1 satisfying the following constraints

$$\begin{aligned} R_0 &\leq I(U;Y_3) \\ R_0 + R_1 &\leq I(U;Y_3) + I(X;Y_1|U) \\ R_0 + R_1 &\leq I(U;Y_3) + I(X;Y_2|U) \\ R_0 + R_1 &\leq I(X;Y_1) \\ R_0 + R_1 &\leq I(X;Y_2) \end{aligned} \tag{2}$$

over all choices of random variables $U \to X \to (Y_1, Y_2, Y_3)$ is achievable. The achievability follows from superposition coding strategy [1]. Unfortunately the optimality of this scheme has remained unsolved for over three decades.

Conjecture 1. *Let $X \to (Y_1, Y_2, Y_3)$ be a discrete memoryless broadcast channel whose transition probability is given by $\mathfrak{q}(y_1, y_2, y_3|x)$. For $0 \leq \lambda \leq 1, \mu \geq 1$ consider the function (on $p(x)$) defined by*

$$T^{\mathfrak{q}}_{\lambda,\mu}(X) := \mathfrak{C}[\lambda I(X;Y_1) + (1-\lambda) I(X;Y_2) - \mu I(X;Y_3)].$$

For a product broadcast channel $\mathfrak{q}_1(y_{11}, y_{21}, y_{31}|x_1) \times \mathfrak{q}_2(y_{12}, y_{22}, y_{32}|x_2)$, let $T^{\mathfrak{q}_1 \times \mathfrak{q}_2}_{\lambda,\mu}(X_1, X_2)$ be defined by

$$T^{\mathfrak{q}_1 \times \mathfrak{q}_2}_{\lambda,\mu}(X_1, X_2) := \mathfrak{C}[\lambda I(X_1, X_2; Y_{11}, Y_{12}) + (1-\lambda) I(X_1, X_2; Y_{21}, Y_{22}) - \mu I(X_1, X_2; Y_{31}, Y_{32})].$$

We claim that the following factorization inequality holds:

$$T^{\mathfrak{q}_1 \times \mathfrak{q}_2}_{\lambda,\mu}(X_1, X_2) \leq T^{\mathfrak{q}_1}_{\lambda,\mu}(X_1) + T^{\mathfrak{q}_2}_{\lambda,\mu}(X_2).$$

If Conjecture 1 is valid then it is immediate that the following function

$$(\alpha_1+\alpha_2+\alpha_3) I(X;Y_3) + \mathfrak{C}[\alpha_2 I(X;Y_1) + \alpha_3 I(X;Y_2) - (\alpha_1+\alpha_2+\alpha_3) I(X;Y_3)] + \alpha_4 I(X;Y_1) + \alpha_5 I(X;Y_2)$$

factorizes for any $\alpha_i \geq 0, i = 1, \ldots, 5$. This in turn implies that the normalized two-letter version of the region in equations (2) reduces to the single-letter form. Since the normalized n-letter form of the region given by (2) approaches capacity (easy to see using Fano's inequality), we obtain the optimality of superposition coding region.

Thus showing the optimality of the superposition coding region has been reduced to establishing the veracity of an information inequality involving concave envelopes of linear combination of mutual information terms. Similar statements can also be made for some other instances in network information theoretic settings.

5 Conclusion

In this article, we describe a novel way of representing achievable regions using concave envelopes. This representation has been shown to vastly simplify explicit calculations of the regions in several instances. One of the major achievements of this line of work, going beyond existing known facts, is in establishing the capacity region of the two receiver MIMO Gaussian broadcast channel in [11]. In this work, we present the preliminary observations that indeed precede the above mentioned result. Further we also propose a conjecture which would help resolve an important open problem.

Acknowledgments

The author wishes to thank several of his collaborators for various discussions that eventually led to some concrete observations and connections between various problems. In particular, the author is thankful to Venkat Anantharam, Abbas El Gamal, Yanlin Geng, and Amin Gohari for the various interesting discussions related to this topic. The author is also thankful to the anonymous referees for various suggestions that improved the presentation of this manuscript.

The work of Chandra Nair was partially supported by the following: an area of excellence grant (Project No. AoE/E-02/08) and two GRF grants (Project Nos. 415810 and 415612) from the University Grants Committee of the Hong Kong Special Administrative Region, China.

References

[1] T. Cover, "Broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-18, pp. 2–14, January, 1972.

[2] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Info. Theory*, vol. IT-53, pp. 350–355, January, 2007.

[3] C. Nair and Z. V. Wang, "On the inner and outer bounds for 2-receiver discrete memoryless broadcast channels," *Proceedings of the ITA Workshop*, 2008, cs.IT/0804.3825.

[4] C. Nair and Z. V. Wang, "On the inner and outer bounds of 3-receiver broadcast channels with 2-degraded message sets," *International Symposium on Information Theory*, pp. 1844–1848, 2009, http://arXiv.org/abs/0806.4415.

[5] C. Nair, Z. V. Wang, and Y. Geng, "An information inequality and evaluation of Marton's inner bound for binary input broadcast channels," *International Symposium on Information Theory*, 2010.

[6] F. M. J. Willems, "The maximal-error and average-error capacity region of the broadcast channel are identical," *Problems of Control and Information Theory*, vol. 19, no. 4, pp. 339–347, 1990.

[7] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *Information Theory, IEEE Transactions on*, vol. 52, pp. 3936 –3964, sept. 2006.

[8] C. Nair, "Capacity regions of two new classes of two-receiver broadcast channels," *Information Theory, IEEE Transactions on*, vol. 56, pp. 4207–4214, sep. 2010.

[9] Y. Geng, A. Gohari, C. Nair, and Y. Yu, "On Marton's inner bound for two receiver broadcast channels," *Presented at ITA Workshop*, 2011.

[10] Y. Geng, A. Gohari, C. Nair, and Y. Yu, "The capacity region of classes of product broadcast channels," *Proceedings of IEEE International Symposium on Information Theory*, pp. 1549–1553, 2011.

[11] Y. Geng and C. Nair, "The capacity region of the two-receiver vector gaussian broadcast channel with private and common messages," Feb. 2012, 1202.0097.

[12] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 769–772, Nov 1973.

[13] Y. Geng, C. Nair, S. Shamai, and Z. V. Wang, "On broadcast channels with binary inputs and symmetric outputs," *International Symposium on Information Theory*, 2010.

[14] B. Xie, M. Griot, A. Casado, and R. Wesel, "Optimal transmission strategy and explicit capacity region for broadcast z channels," *Information Theory, IEEE Transactions on*, vol. 54, pp. 4296 –4304, sept. 2008.

[15] P. F. Bergmans, "Coding theorem for broadcast channels with degraded components," *IEEE Trans. Info. Theory*, vol. IT-15, pp. 197–207, March, 1973.

[16] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.

[17] S. I. Gel'fand and M. S. Pinsker, "Coding for Channel with Random Parameters," *Probl. Pered. Inform.*, vol. 9, pp. 19–31, 1980.

[18] A. D. Wyner, "The Wire-tap Channel," *Bell Systems Technical Journal*, vol. 54, pp. 1355–1387, Jan. 1975.

[19] I. Csizár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Info. Theory*, vol. IT-24, pp. 339–348, May, 1978.