The capacity region of the two-receiver vector Gaussian broadcast channel with private and common messages*

Yanlin Geng, Chandra Nair

June 4, 2012

Abstract

We develop a new method for showing the optimality of the Gaussian distribution in multiterminal information theory problems. As an application of this method we show that Marton's inner bound achieves the capacity of the two-receiver vector Gaussian broadcast channels with private and common messages, solving an open problem of considerable interest.

1 Introduction

Channels with additive Gaussian noise are a commonly used model for wireless communications. Hence computing the capacity regions or bounds on the capacity regions for these classes of channels is of wide interest. Usually these bounds or capacity regions are represented using auxiliary random variables and distributions on these auxiliary random variables. Evaluations of these bounds then become an optimization problem of computing the extremal auxiliary random variables. In several instances involving Gaussian noise channels, it turns out that the optimal auxiliaries and the inputs are Gaussian. However proving the optimality of Gaussian distributions is usually very cumbersome and involves certain non-trivial applications of the entropy-power-inequality(EPI), and the perturbation ideas behind its proof.

For the two-receiver vector Gaussian broadcast channel with private messages, the capacity region was established [16] by showing that certain inner and outer bounds match. This argument was indirect, and hence the approach has been hard to generalize to other situations. In the following sections we develop a novel way of proving the optimality of Gaussian input distribution for additive Gaussian noise channels. There are many potential straightforward applications of this new approach which will yield new results as well as recover the earlier results in a simple manner. For the purpose of this article, we will restrict ourselves to two-receiver vector Gaussian channels. We will recover the known results for the private messages case and obtain the capacity region in the presence of a common message as well.

1.1 Definitions

Broadcast channel [4] refers to a communication scenario where a single sender, usually denoted by X, wishes to communicate independent messages (M_0, M_1, M_2) to two receivers Y_1, Y_2. The goal of the communication scheme is to enable receiver Y_1 to recover messages (M_0, M_1) and receiver Y_2 to recover messages (M_0, M_2); both events being required to occur with high probability. For introduction to the broadcast channel problem and a summary of known work one may refer to Chapters 5, 8, and 9 in [6].

A *broadcast channel* is characterized by a probability transition matrix $\mathfrak{q}(y_1, y_2|x)$. The following broadcast channel is referred to as the vector additive Gaussian broadcast channel

$$\begin{aligned} \mathbf{Y}_1 &= G_1\mathbf{X} + \mathbf{Z}_1 \\ \mathbf{Y}_2 &= G_2\mathbf{X} + \mathbf{Z}_2. \end{aligned}$$

In the above $\mathbf{X} \in \mathbb{R}^t$, G_1, G_2 are $t \times t$ matrices, and $\mathbf{Z}_1, \mathbf{Z}_2$ are Gaussian vectors independent of $\mathbf{X}$.

*The paper is in part selected to be presented at the International Symposium on Information Theory, 2012.

Remark 1. We assume, w.l.o.g. that $Z_1, Z_2 \sim \mathcal{N}(0, I)$.

A *product broadcast channel* is a broadcast channel whose transition probability has the form $\mathfrak{q}_1(\mathbf{y}_{11}, \mathbf{y}_{21}|\mathbf{x}_1) \times \mathfrak{q}_2(\mathbf{y}_{12}, \mathbf{y}_{22}|\mathbf{x}_2)$. A vector additve Gaussian product broadcast channel can be represented as

$$\begin{bmatrix} \mathbf{Y}_{11} \\ \mathbf{Y}_{12} \end{bmatrix} = \begin{bmatrix} G_{11} & 0 \\ 0 & G_{12} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{11} \\ \mathbf{Z}_{12} \end{bmatrix}$$
$$\begin{bmatrix} \mathbf{Y}_{21} \\ \mathbf{Y}_{22} \end{bmatrix} = \begin{bmatrix} G_{21} & 0 \\ 0 & G_{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{21} \\ \mathbf{Z}_{22} \end{bmatrix}.$$

In the above $\mathbf{Z}_{11}, \mathbf{Z}_{12}, \mathbf{Z}_{21}, \mathbf{Z}_{22}$ are independent Gaussian vectors, also independent of $\mathbf{X}_1, \mathbf{X}_2$.

Remark 2. We assume that all our channel gain matrices are invertible. Since the set of all matrices is dense (with respect to say, Frobenius norm) by continuity, our capacity results extend to non-invertible cases.

Organization of the paper

In the remainder of this section, we will establish some elementary mathematical results that we will call upon in the rest of the paper. In general there are three main ideas employed in this paper: (*i*) the use of a two-letter expression to identify that the optimizing distribution is Gaussian, (*ii*) the factorization of concave envelopes that relates the two-letter expression to single-letter ones, and (*iii*) use of a max-min interchange to deal with a linearized expression. The second and third ideas had been developed in the context of discrete memoryless broadcast channels [8]. We present our first idea in a very simple instance of maximizing mutual information in subsection 2.1. We then present the second idea in subsection 2.2 for the private messages case and generalize it to the case with private messages in section 2.3. We then show how our earlier results can be used to establish the capacity region in section 3. In this section, we will also use the min-max idea we alluded to earlier. Since, we are working with continuous alphabets, our approach involves some mathematical technicalities that need to be taken care of, and we present these arguments in the Appendix. The arguments in the appendix are presented in a general manner so as to enable future applications of these ideas avoid these technical discussions.

1.2 A couple of mathematical preliminaries

We present some simple claims regarding additive Gaussian channels which will be useful later.

Claim 1. *Consider the following vector additive Gaussian product channel with identical components*

$$\mathbf{Y}_1 = G\mathbf{X}_1 + \mathbf{Z}_1$$
$$\mathbf{Y}_2 = G\mathbf{X}_2 + \mathbf{Z}_2$$

Further let Z_1, Z_2 be independent and distributed as $\mathcal{N}(0, I)$. Define

$$\tilde{\mathbf{X}} = \frac{1}{\sqrt{2}}(\mathbf{X}_1 + \mathbf{X}_2), \quad \mathbf{X}' = \frac{1}{\sqrt{2}}(\mathbf{X}_1 - \mathbf{X}_2), \quad \tilde{\mathbf{Y}} = \frac{1}{\sqrt{2}}(\mathbf{Y}_1 + \mathbf{Y}_2), \quad \mathbf{Y}' = \frac{1}{\sqrt{2}}(\mathbf{Y}_1 - \mathbf{Y}_2).$$

Then $I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2) = I(\tilde{\mathbf{X}}, \mathbf{X}'; \tilde{\mathbf{Y}}, \mathbf{Y}')$.

Proof. The proof is a trivial consequence of the fact that $h(\tilde{\mathbf{Y}}, \mathbf{Y}') = h(\mathbf{Y}_1, \mathbf{Y}_2)$ and $h(\tilde{\mathbf{Y}}, \mathbf{Y}'|\tilde{\mathbf{X}}, \mathbf{X}') = h(\tilde{\mathbf{Z}}, \mathbf{Z}') = h(\mathbf{Z}_1, \mathbf{Z}_2) = h(\mathbf{Y}_1, \mathbf{Y}_2|\mathbf{X}_1, \mathbf{X}_2)$ where $\tilde{\mathbf{Z}} = \frac{1}{\sqrt{2}}(\mathbf{Z}_1 + \mathbf{Z}_2), \mathbf{Z}' = \frac{1}{\sqrt{2}}(\mathbf{Z}_1 - \mathbf{Z}_2)$. $\square$

Remark 3. An interesting consequence of Gaussian noise is that $\tilde{\mathbf{Z}}$ and $\mathbf{Z}'$ are again independent and distributed according to $\mathcal{N}(0, I)$. Hence $\tilde{\mathbf{Y}}, \mathbf{Y}'$ can be regarded as the outputs of the Gaussian channel when the inputs are distributed according to $\tilde{\mathbf{X}}, \mathbf{X}'$. This observation is peculiar to additive Gaussian channels.

Claim 2. *In vector additive Gaussian product broadcast channels with invertible channel gain matrices, the random variables $\mathbf{Y}_{11}$ and $\mathbf{Y}_{22}$ are independent* if and only if *$\mathbf{X}_1$ and $\mathbf{X}_2$ are independent.*

Proof. Here we prove the non-trivial direction. Suppose $\mathbf{Y}_{11}$ and $\mathbf{Y}_{22}$ are independent. We know that $\mathbf{Y}_{11} = G_{11}\mathbf{X}_1 + \mathbf{Z}_{11}$ and $\mathbf{Y}_{22} = G_{22}\mathbf{X}_2 + \mathbf{Z}_{22}$ where $\mathbf{Z}_{11}, \mathbf{Z}_{22}$ are mutually independent and independent of the pair $\mathbf{X}_1, \mathbf{X}_2$. Taking characteristic functions we see that

$$\mathrm{E}\left(e^{i(\mathbf{t}_1\cdot\mathbf{Y}_{11}+\mathbf{t}_2\cdot\mathbf{Y}_{22})}\right) = \mathrm{E}\left(e^{i\mathbf{t}_1\cdot\mathbf{Y}_{11}}\right)\mathrm{E}\left(e^{i\mathbf{t}_2\cdot\mathbf{Y}_{22}}\right) = \mathrm{E}\left(e^{i\mathbf{t}_1\cdot\mathbf{Z}_{11}}\right)\mathrm{E}\left(e^{i\mathbf{t}_1\cdot G_{11}\mathbf{X}_1}\right)\mathrm{E}\left(e^{i\mathbf{t}_2\cdot G_{22}\mathbf{X}_2}\right)\mathrm{E}\left(e^{i\mathbf{t}_2\cdot\mathbf{Z}_{22}}\right).$$

On the other hand

$$\mathrm{E}\left(e^{i(\mathbf{t}_1\cdot\mathbf{Y}_{11}+\mathbf{t}_2\cdot\mathbf{Y}_{22})}\right) = \mathrm{E}\left(e^{i\mathbf{t}_1\cdot\mathbf{Z}_{11}}\right)\mathrm{E}\left(e^{i(\mathbf{t}_1\cdot G_{11}\mathbf{X}_1+\mathbf{t}_2\cdot G_{22}\mathbf{X}_2)}\right)\mathrm{E}\left(e^{i\mathbf{t}_2\cdot\mathbf{Z}_{22}}\right).$$

Since $\mathrm{E}\left(e^{i\mathbf{t}_1\cdot\mathbf{Z}_{11}}\right), \mathrm{E}\left(e^{i\mathbf{t}_2\cdot\mathbf{Z}_{22}}\right) > 0 \ \forall \mathbf{t}_1, \mathbf{t}_2$ we have that

$$\mathrm{E}\left(e^{i(\mathbf{t}_1\cdot G_{11}\mathbf{X}_1+\mathbf{t}_2\cdot G_{22}\mathbf{X}_2)}\right) = \mathrm{E}\left(e^{i\mathbf{t}_1\cdot G_{11}\mathbf{X}_1}\right)\mathrm{E}\left(e^{i\mathbf{t}_2\cdot G_{22}\mathbf{X}_2}\right), \ \forall \mathbf{t}_1, \mathbf{t}_2.$$

Hence $G_{11}\mathbf{X}_1$ and $G_{22}\mathbf{X}_2$ are independent; since G_{11} and G_{22} are invertible, $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent. □

2 Optimality of Gaussian via factorization of concave envelopes

We devise a new technique to show that Gaussian distribution achieves the maximum value of an optimization problem, subject to a covariance constraint. Though some of the results have been known earlier [11], the technique presented here allows us to obtain much broader results. We are also aware[1] that this approach can lead to establishment of a wide variety of new results in the additive Gaussian setting.

The main idea behind the approach is to show that if a certain X (say zero mean) achieves the maximum value of an optimization problem, then so does $\frac{1}{\sqrt{2}}(X_1 + X_2)$ and $\frac{1}{\sqrt{2}}(X_1 - X_2)$; where X_1, X_2 are two i.i.d. copies of X. Further we will show that $\frac{1}{\sqrt{2}}(X_1+X_2)$ and $\frac{1}{\sqrt{2}}(X_1-X_2)$ have to be independent as well, which forces the initial distribution to be Gaussian, see Theorem 3 and Corollary 3 in Appendix A.1. Alternately, one can repeat averaging procedure inductively and use central limit theorem to conclude that Gaussian distribution achieves the maximum. To show the first step we go to the two-letter version[2] of the channel, use a *factorization property* of the function involved and then use Claim 1 to move from the pair X_1, X_2 to $\frac{1}{\sqrt{2}}(X_1 + X_2)$.

Remark 4. It is worth noting the remarkable similarity of the structure of the arguments that follow for the three optimization problems below for which we show the optimality of Gaussian. In particular the first example, though trivial, contains most of the key intuitive elements.

2.1 Example 1: Mutual information

Let $\mathbf{Y} = G\mathbf{X} + \mathbf{Z}$ represent a point-to-point channel, where $\mathbf{Z} \sim \mathcal{N}(0, I)$ and G is invertible. Given $K \succeq 0$, consider the following optimization problem:

$$\mathrm{V}(K) = \max_{\mathbf{X}:\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K} I(\mathbf{X};\mathbf{Y}).$$

Remark 5. By writing max instead of sup we are indeed claiming the existence of a maximizing distribution. This is a non-trivial technical issue that we will deal with (in the Appendix) for the newer functions that we consider. The same arguments used for establishing Claim 6 can be used (essentially verbatim) to imply the existence of a maximizing distribution here. Furthermore for the above optimization problem it is well-known that $\mathbf{X} \sim \mathcal{N}(0, K)$ achieves $\mathrm{V}(K)$ and the aim here is to give a simple illustration of our approach.

Consider a product channel consisting of two identical components of the point-to-point channel described above: $\mathfrak{q}(\mathbf{y}_1|\mathbf{x}_1) \times \mathfrak{q}(\mathbf{y}_2|\mathbf{x}_2)$. We call the below claim as the *factorization property* of mutual information.

[1] Between the time of the appearance of an earlier version of the manuscript in arXiv and polishing of this article, we have already been communicated new results that make use of our proposed approach.

[2] A two-letter version of a channel $\mathfrak{q}(y|x)$ is a product channel consisting of identical components $\mathfrak{q}(y_1|x_1) \times \mathfrak{q}(y_2|x_2)$.

Claim 3. *The following inequality holds for the product channel*

$$I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2) \leq I(\mathbf{X}_1; \mathbf{Y}_1) + I(\mathbf{X}_2; \mathbf{Y}_2).$$

Further if equality is achieved at some $p(\mathbf{x}_1, \mathbf{x}_2)$ then $\mathbf{X}_1, \mathbf{X}_2$ must be independent.

Proof. The proof is essentially a consequence of the following equality for product channels

$$I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2) = I(\mathbf{X}_1; \mathbf{Y}_1) + I(\mathbf{X}_2; \mathbf{Y}_2) - I(\mathbf{Y}_1; \mathbf{Y}_2).$$

Further, if equality holds then $\mathbf{Y}_1, \mathbf{Y}_2$ must be independent, which from Claim 2 implies that $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent. □

Let $p_*(\mathbf{x})$ be a zero mean distribution that achieves $\mathrm{V}(K)$.

Claim 4. *Let $(\mathbf{X}_1, \mathbf{X}_2) \sim p_*(\mathbf{x}_1)p_*(\mathbf{x}_2)$ be two i.i.d. copies of $p_*(\mathbf{x})$. Then the following distributions $\tilde{\mathbf{X}} = \frac{1}{\sqrt{2}}(\mathbf{X}_1 + \mathbf{X}_2)$, $\mathbf{X}' = \frac{1}{\sqrt{2}}(\mathbf{X}_1 - \mathbf{X}_2)$ also achieve $\mathrm{V}(K)$. Further the random variables $\tilde{\mathbf{X}}, \mathbf{X}'$ are independent.*

Proof. Let $\tilde{\mathbf{Y}} = \frac{1}{\sqrt{2}}(\mathbf{Y}_1 + \mathbf{Y}_2)$, $\mathbf{Y}' = \frac{1}{\sqrt{2}}(\mathbf{Y}_1 - \mathbf{Y}_2)$. The claim is a consequence of Claim 3 and the following observations:

$$\begin{aligned} 2\mathrm{V}(K) &= I(\mathbf{X}_1; \mathbf{Y}_1) + I(\mathbf{X}_2; \mathbf{Y}_2) \\ &= I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2) \\ &\stackrel{(a)}{=} I(\tilde{\mathbf{X}}, \mathbf{X}'; \tilde{\mathbf{Y}}, \mathbf{Y}') \\ &\stackrel{(b)}{\leq} I(\tilde{\mathbf{X}}; \tilde{\mathbf{Y}}) + I(\mathbf{X}'; \mathbf{Y}') \\ &\leq \mathrm{V}(K) + \mathrm{V}(K) = 2\mathrm{V}(K). \end{aligned}$$

Here the first equality comes because $p_*(\mathbf{x})$ achieves $\mathrm{V}(K)$, the second one because $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent. Equality (a) is a consequence of Claim 1, Inequality (b) is a consequence of Claim 3, and the last inequality follows from the following:

$$\mathrm{E}(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T) = \mathrm{E}(\mathbf{X}'\mathbf{X}'^T) = \frac{1}{2}\left(\mathrm{E}(\mathbf{X}_1\mathbf{X}_1^T) + \mathrm{E}(\mathbf{X}_2\mathbf{X}_2^T)\right) \preceq K,$$

and the definition of V(K). Since the extremes match, all inequalities must be equalities. Hence (b) must be an equality, which implies from Claim 3 that $\tilde{\mathbf{X}}, \mathbf{X}'$ are independent. Similarly we require $I(\tilde{\mathbf{X}}; \tilde{\mathbf{Y}}) = I(\mathbf{X}'; \mathbf{Y}') = \mathrm{V}(K)$ as desired. □

Hence we have shown that $\mathbf{X} \sim p^*(\mathbf{x})$ that achieves a maximum has the following property: If $(\mathbf{X}_1, \mathbf{X}_2)$ are i.i.d. copies each distributed according to $p_*(\mathbf{x})$, then $\mathbf{X}_1 + \mathbf{X}_2$ and $\mathbf{X}_1 - \mathbf{X}_2$ are also independent. Thus from Theorem 3 and Corollary 3(Appendix A.1) we have that $\mathbf{X} \sim \mathcal{N}(0, K')$ for some $K' \preceq K$. Alternately, one could also use the following approach: For any $\mathbf{X} \sim p^*(\mathbf{x})$ (assume zero mean) that achieves the maximum, we know that $\frac{1}{\sqrt{2}}(\mathbf{X}_1 + \mathbf{X}_2)$ also achieves the maximum. Hence proceeding by induction, we can use Central limit theorem to deduce that the Gaussian distribution also achieves the maximum. This alternate approach is elaborated for the next example in Appendix C. There is a subtle difference between the arguments however; the former one ensures the uniqueness of the maximizer to be Gaussian while the one in Appendix C only yields that Gaussian is a maximizer.

Remark 6. For this example, we can use the monotonicity of the $\log|\cdot|$ function to deduce that $K' = K$. In the examples that follow below we do not have any such monotonicity. Hence, we will only establish that the optimizing distribution is a Gaussian, which is sufficient for our purposes.

From the above discussion we get an alternate argument of this well known result:

Lemma 1. *$V(K)$ is obtained when (and only when) the input $\mathbf{X}$ is distributed as $\mathcal{N}(0, K)$.*

2.2 Example 2: Difference of mutual informations

Consider a vector additive Gaussian broadcast channel. For $\lambda > 1$ let the following function of $p(x)$ be defined by

$$\mathsf{s}_\lambda(\mathbf{X}) := I(\mathbf{X};\mathbf{Y}_1) - \lambda I(\mathbf{X};\mathbf{Y}_2).$$

Let $\mathsf{s}_\lambda(\mathbf{X}|V) := I(\mathbf{X};\mathbf{Y}_1|V) - \lambda I(\mathbf{X};\mathbf{Y}_2|V)$.

Further define

$$S_\lambda(\mathbf{X}) := \mathfrak{C}(\mathsf{s}_\lambda(\mathbf{X}))$$

as the *upper concave envelope*[3] of $\mathsf{s}_\lambda(\mathbf{X})$. It is a straightforward exercise to see that

$$\mathfrak{C}(\mathsf{s}_\lambda(\mathbf{X})) = \sup_{\substack{p(v|\mathbf{x}):\\ V\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)}} I(\mathbf{X};\mathbf{Y}_1|V) - \lambda I(\mathbf{X};\mathbf{Y}_2|V) = \sup_{p(v|\mathbf{x})} \mathsf{s}_\lambda(\mathbf{X}|V).$$

We also define $S_\lambda(\mathbf{X}|V) := \sum_v p(v) S_\lambda(\mathbf{X}|V=v)$ for finite V and its natural extension for arbitrary V.

Remark 7. We will try to keep the language simple in the main body of this paper. In the Appendix we will deal with the various technical issues with due diligence.

For a product broadcast channel $\mathfrak{q}_1(\mathbf{y}_{11},\mathbf{y}_{21}|\mathbf{x}_1)\times\mathfrak{q}_2(\mathbf{y}_{12},\mathbf{y}_{22}|\mathbf{x}_2)$ let $S_\lambda(\mathbf{X}_1,\mathbf{X}_2)$ denote the corresponding upper concave envelope. The following claim is referred to as the "factorization of $S_\lambda(\mathbf{X}_1,\mathbf{X}_2)$".

Claim 5. *The following inequality holds for product broadcast channels*

$$S_\lambda(\mathbf{X}_1,\mathbf{X}_2) \le S_\lambda(\mathbf{X}_1|\mathbf{Y}_{22}) + S_\lambda(\mathbf{X}_2|\mathbf{Y}_{11}) \le S_\lambda(\mathbf{X}_1) + S_\lambda(\mathbf{X}_2).$$

For additive Gaussian noise broadcast channels if $p(v|\mathbf{x}_1,\mathbf{x}_2)$ realizes $S_\lambda(\mathbf{X}_1,\mathbf{X}_2)$, i.e. $S_\lambda(\mathbf{X}_1,\mathbf{X}_2) = \mathsf{s}_\lambda(\mathbf{X}_1,\mathbf{X}_2|V)$, and equality is achieved above i.e. $S_\lambda(\mathbf{X}_1,\mathbf{X}_2) = S_\lambda(\mathbf{X}_1) + S_\lambda(\mathbf{X}_2)$, then all of the following must be true

1. *$\mathbf{X}_1$ and $\mathbf{X}_2$ are conditionally independent of V*
2. *$V,\mathbf{X}_1$ achieves $S_\lambda(\mathbf{X}_1)$*
3. *$V,\mathbf{X}_2$ achieves $S_\lambda(\mathbf{X}_2)$.*

Proof. For any $p(v|\mathbf{x}_1,\mathbf{x}_2)$ observe the following

$$\begin{aligned}
&I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{11},\mathbf{Y}_{12}|V) - \lambda I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{21},\mathbf{Y}_{22}|V)\\
&\quad = I(\mathbf{X}_1;\mathbf{Y}_{11}|V) + I(\mathbf{X}_2;\mathbf{Y}_{12}|V,\mathbf{Y}_{11}) - \lambda I(\mathbf{X}_2;\mathbf{Y}_{22}|V) - \lambda I(\mathbf{X}_1;\mathbf{Y}_{21}|V,\mathbf{Y}_{22})\\
&\quad = I(\mathbf{X}_1;\mathbf{Y}_{11}|V,\mathbf{Y}_{22}) + I(\mathbf{X}_2;\mathbf{Y}_{12}|V,\mathbf{Y}_{11}) - \lambda I(\mathbf{X}_2;\mathbf{Y}_{22}|V,\mathbf{Y}_{11}) - \lambda I(\mathbf{X}_1;\mathbf{Y}_{21}|V,\mathbf{Y}_{22}) - (\lambda-1)I(\mathbf{Y}_{11};\mathbf{Y}_{22}|V)\\
&\quad \le S_\lambda(\mathbf{X}_1|\mathbf{Y}_{22}) + S_\lambda(\mathbf{X}_2|\mathbf{Y}_{11}) - (\lambda-1)I(\mathbf{Y}_{11};\mathbf{Y}_{22}|V)\\
&\quad \le S_\lambda(\mathbf{X}_1) + S_\lambda(\mathbf{X}_2) - (\lambda-1)I(\mathbf{Y}_{11};\mathbf{Y}_{22}|V)\\
&\quad \le S_\lambda(\mathbf{X}_1) + S_\lambda(\mathbf{X}_2).
\end{aligned}$$

Since equality holds all inequalities are tight. Hence $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are conditionally independent of V implying that $\mathbf{X}_1$ and $\mathbf{X}_2$ are conditionally independent of V (Claim 1). Hence

$$I(\mathbf{X}_1;\mathbf{Y}_{11}|V,\mathbf{Y}_{22}) - \lambda I(\mathbf{X}_1;\mathbf{Y}_{21}|V,\mathbf{Y}_{22}) = I(\mathbf{X}_1;\mathbf{Y}_{11}|V) - \lambda I(\mathbf{X}_1;\mathbf{Y}_{21}|V) = S_\lambda(\mathbf{X}_1),$$

$$I(\mathbf{X}_2;\mathbf{Y}_{12}|V,\mathbf{Y}_{11}) - \lambda I(\mathbf{X}_2;\mathbf{Y}_{22}|V,\mathbf{Y}_{11}) = I(\mathbf{X}_2;\mathbf{Y}_{12}|V) - \lambda I(\mathbf{X}_2;\mathbf{Y}_{22}|V) = S_\lambda(\mathbf{X}_2).$$

This completes the proof. ∎

[3] The upper concave envelope of a function $f(x)$ is the smallest concave function $g(x)$ such that $g(x) \ge f(x), \forall x$.

2.2.1 Maximizing the concave envelope subject to a covariance constraint

Consider an Additive Gaussian Noise broadcast channel $\mathfrak{q}(y_1, y_2|x)$. For $K \succeq 0$, define

$$\mathrm{V}_\lambda(K) = \sup_{\mathbf{X}:\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K} S_\lambda(\mathbf{X}).$$

Claim 6. *There is a pair of random variables $(V_*, \mathbf{X}_*)$ with $|V_*| \leq \frac{t(t+1)}{2} + 1$ such that*

$$\mathrm{V}_\lambda(K) = \mathsf{s}_\lambda(\mathbf{X}_*|V_*).$$

Proof. This is a technical claim that shows that the supremum is indeed attained. The details are present in the Appendix B. □

The goal of this section is to show that a single Gaussian distribution achieves $\mathrm{V}_\lambda(K)$, i.e. we can take V to be trivial and $X \sim \mathcal{N}(0, K'), K' \preceq K$. (This result is known and was first shown by Liu and Vishwanath[11] using perturbation based techniques. We use this here as a non-trivial illustration of our technique and then our final result in the next section is new.)

Consider a product channel consisting of two identical components $\mathfrak{q}(\mathbf{y}_{11}, \mathbf{y}_{21}|\mathbf{x}_1) \times \mathfrak{q}(\mathbf{y}_{12}, \mathbf{y}_{22}|\mathbf{x}_2)$.

Notation: In the remainder of the section we assume that $p_*(v, \mathbf{x})$ achieves $\mathrm{V}_\lambda(K)$, $|V| = m \leq \frac{t(t+1)}{2} + 1$ and $\mathbf{X}_v$ be a centered random variable (zero-mean) distributed according to $p(\mathbf{x}|V = v)$. Further let $K_v = \mathrm{E}(\mathbf{X}_v\mathbf{X}_v^T)$. Then we have $\sum_{v=1}^m p_*(v)K_v \preceq K$ and in particular that K_v's are bounded.

Claim 7. *Let $(V_1, V_2, \mathbf{X}_1, \mathbf{X}_2) \sim p_*(v_1, \mathbf{x}_1)p_*(v_2, \mathbf{x}_2)$ be two i.i.d. copies of $p_*(v, \mathbf{x})$. We assume that $|V| \leq \frac{t(t+1)}{2} + 1$. Let*

$$\tilde{V} = (V_1, V_2), \quad \tilde{\mathbf{X}}|\big(\tilde{V} = (v_1, v_2)\big) \sim \frac{1}{\sqrt{2}}\left(\mathbf{X}_{v_1} + \mathbf{X}_{v_2}\right), \quad \mathbf{X}'|\big(\tilde{V} = (v_1, v_2)\big) \sim \frac{1}{\sqrt{2}}\left(\mathbf{X}_{v_1} - \mathbf{X}_{v_2}\right).$$

In the above we take $\mathbf{X}_{v_1}$ and $\mathbf{X}_{v_2}$ to be independent random variables. Then the following hold:

1. *$\tilde{\mathbf{X}}, \mathbf{X}'$ are conditionally independent given $\tilde{V}$.*
2. *$\tilde{V}, \tilde{\mathbf{X}}$ achieves $\mathrm{V}_\lambda(K)$.*
3. *$\tilde{V}, \mathbf{X}'$ achieves $\mathrm{V}_\lambda(K)$.*

Proof.

$$\begin{aligned}
2\mathrm{V}_\lambda(K) &= \mathsf{s}_\lambda(\mathbf{X}_1|V_1) + \mathsf{s}_\lambda(\mathbf{X}_2|V_2)\\
&= \mathsf{s}_\lambda(\mathbf{X}_1, \mathbf{X}_2|V_1, V_2)\\
&\overset{(a)}{=} \mathsf{s}_\lambda(\tilde{\mathbf{X}}, \mathbf{X}'|\tilde{V})\\
&\overset{(b)}{\leq} S_\lambda(\tilde{\mathbf{X}}, \mathbf{X}')\\
&\overset{(c)}{\leq} S_\lambda(\tilde{\mathbf{X}}) + S_\lambda(\mathbf{X}')\\
&\leq \mathrm{V}_\lambda(K) + \mathrm{V}_\lambda(K) = 2\mathrm{V}_\lambda(K).
\end{aligned}$$

Here the first equality comes because $p_*(v, \mathbf{x})$ achieves $\mathrm{V}_\lambda(K)$, the second one because $(V_1, \mathbf{X}_1)$ and $(V_2, \mathbf{X}_2)$ are independent. Equality (a) is a consequence of Claim 1, inequality (c) is a consequence of Claim 5, and the last inequality follows from the following:

$$\mathrm{E}(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T) = \mathrm{E}(\mathbf{X}'\mathbf{X}'^T) = \sum_{v_1,v_2} p_*(v_1)p_*(v_2)\frac{(K_{v_1} + K_{v_2})}{2} = \sum_{v=1}^{m} p_*(v)K_v \preceq K,$$

and the definition of $\mathrm{V}_\lambda(K)$. Since the extremes match, all inequalities must be equalities. Hence (b) must be an equality, $p(\tilde{v}, \tilde{\mathbf{x}}, \mathbf{x}')$ achieves $S_\lambda(\tilde{\mathbf{X}}, \mathbf{X}')$; and since (c) is also equality from Claim 5 we conclude that $\tilde{\mathbf{X}}, \mathbf{X}'$ are conditionally independent of $\tilde{V}$. Furthermore, we also obtain that $p(\tilde{v}|\tilde{\mathbf{x}})$ achieves $S_\lambda(\tilde{\mathbf{X}})$, which from the last inequality matches $\mathrm{V}_\lambda(K)$. Similarly for $p(\tilde{v}|\mathbf{x}')$. □

As a consequence, $\mathbf{X}_{v_1}$, $\mathbf{X}_{v_2}$ are independent random variables and $(\mathbf{X}_{v_1}+\mathbf{X}_{v_2}),(\mathbf{X}_{v_1}-\mathbf{X}_{v_2})$ are also independent random variables. Thus from Corollary 3 (in Appendix A.1) $\mathbf{X}_{v_1}, \mathbf{X}_{v_2}$ are Gaussians, say having the same distribution as $\mathbf{X}_v \sim \mathcal{N}(0, K')$. Since v_1, v_2 are arbitrary, all $\mathbf{X}_{v_i}$ are Gaussians, having the same distribution as $\mathbf{X}_v$. Then

$$\mathrm{V}_\lambda(K) = \sum_{i=1}^{m} p_*(v_i)\mathsf{s}_\lambda(\mathbf{X}_{v_i}) = \sum_{i=1}^{m} p_*(v_i)\mathsf{s}_\lambda(\mathbf{X}_v) = \mathsf{s}_\lambda(\mathbf{X}_v).$$

Hence we obtain the following theorem (originally established in [11]).

Theorem 1. *There exists $\mathbf{X}_* \sim \mathcal{N}(0, K'), K' \preceq K$ such that $\mathrm{V}_\lambda(K) = \mathsf{s}_\lambda(\mathbf{X}_*)$. Further the maximizer is unique.*

Proof. Here, we only comment on uniqueness. The above argument implies that if $\mathbf{X}_1$, $\mathbf{X}_2$ are independent random variables and identically distributed according to a maximizing distribution then one would have that $\frac{1}{\sqrt{2}}(\mathbf{X}_1+\mathbf{X}_2), \frac{1}{\sqrt{2}}(\mathbf{X}_1-\mathbf{X}_2)$ is also independent which then implies, as stated before, that the maximizing distribution was a Gaussian. Suppose there were two maximizing Gaussians with covariances $K_1, K_2 \preceq K$, then taking a mixture according to $X|(V=0) \sim \mathcal{N}(0,K_1)$ and $X|(V=1) \sim \mathcal{N}(0,K_2)$ would also yield a maximum, which would then imply that a mixture of Gaussians is also a maximizer, a contradiction. $\square$

Remark: Notice that we never used the precise form of $S_\lambda(X)$ but just used that the implications of Claim 5. In the next section we will define a new concave envelope that will also satisfy a condition similar to Claim 5, and then establish the optimality of Gaussian.

Corollary 1. *If $\mathbf{X} \sim \mathcal{N}(0, K)$ then there exists $\mathbf{X}_* \sim \mathcal{N}(0, K'), K' \preceq K$ such that $S_\lambda(\mathbf{X}) = \mathsf{s}_\lambda(\mathbf{X}_*) = \mathrm{V}_\lambda(K)$.*

Proof. Clearly from Theorem 1 and definition of $\mathrm{V}_\lambda(K)$ we have

$$S_\lambda(\mathbf{X}) \leq \mathrm{V}_\lambda(K) = \mathsf{s}_\lambda(\mathbf{X}_*).$$

On the other hand let $\mathbf{X}' \sim \mathcal{N}(0, K-K')$ be independent of $\mathbf{X}_*$. Note that $\mathbf{X} \sim \mathbf{X}' + \mathbf{X}_*$ and

$$S_\lambda(\mathbf{X}) = \sup_V \mathsf{s}_\lambda(\mathbf{X}|V) \geq \mathsf{s}_\lambda(\mathbf{X}|\mathbf{X}') = \mathsf{s}_\lambda(\mathbf{X}_*).$$ $\square$

2.3 Example 3: A more complicated example

The function we considered in the previous section can be used to determine the capacity region of vector Gaussian broadcast channel with only private messages (see Section 3.1). The function we consider in this section will enable us to determine the capacity region of vector Gaussian broadcast channel with common message as well (see Section 3.2).

For $\lambda_0, \lambda_1, \lambda_2 > 0$ and for $\alpha \in [0,1]$ (and $\bar{\alpha} := 1-\alpha$) consider the following function of $p(\mathbf{x})$ defined by

$$\mathsf{t}_{\vec{\lambda}}(\mathbf{X}) := -\lambda_0 \alpha I(\mathbf{X};\mathbf{Y}_1) - \lambda_0\bar{\alpha} I(\mathbf{X};\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(\mathbf{X};\mathbf{Y}_2) + \lambda_1 S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}).$$

Further let

$$T_{\vec{\lambda}}(\mathbf{X}) := \mathfrak{C}(\mathsf{t}_{\vec{\lambda}}(\mathbf{X}))$$

denote the upper concave envelope of $\mathsf{t}_{\vec{\lambda}}(\mathbf{X})$. It is easy to see that

$$\mathfrak{C}(\mathsf{t}_{\vec{\lambda}}(\mathbf{X})) = \sup_{p(w|\mathbf{x})} -\lambda_0\alpha I(\mathbf{X};\mathbf{Y}_1|W) - \lambda_0\bar{\alpha} I(\mathbf{X};\mathbf{Y}_2|W) + (\lambda_1+\lambda_2) I(\mathbf{X};\mathbf{Y}_2|W) + \lambda_1 S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}|W).$$

For a product broadcast channel $\mathfrak{q}_1(\mathbf{y}_{11},\mathbf{y}_{21}|\mathbf{x}_1)\times\mathfrak{q}_2(\mathbf{y}_{12},\mathbf{y}_{22}|\mathbf{x}_2)$ let $T_{\vec{\lambda}}(\mathbf{X}_1,\mathbf{X}_2)$ denote the corresponding upper concave envelope. The following claim is referred to as the "factorization of $T_{\vec{\lambda}}(\mathbf{X}_1,\mathbf{X}_2)$".

Claim 8. *When $\lambda_0 > \lambda_1 + \lambda_2$ the following inequality holds for product broadcast channels*

$$T_{\vec{\lambda}}(\mathbf{X}_1, \mathbf{X}_2) \leq T_{\vec{\lambda}}(\mathbf{X}_1|\mathbf{Y}_{22}) + T_{\vec{\lambda}}(\mathbf{X}_2|\mathbf{Y}_{11}) \leq T_{\vec{\lambda}}(\mathbf{X}_1) + T_{\vec{\lambda}}(\mathbf{X}_2).$$

For additive Gaussian noise broadcast channels if $p(w, \mathbf{x}_1, \mathbf{x}_2)$ realizes $T_{\vec{\lambda}}(\mathbf{X}_1, \mathbf{X}_2)$ and equality is achieved above then all of the following must be true

1. *$\mathbf{X}_1$ and $\mathbf{X}_2$ are conditionally independent of W*
2. *$W, \mathbf{X}_1$ achieves $T_{\vec{\lambda}}(\mathbf{X}_1)$*
3. *$W, \mathbf{X}_2$ achieves $T_{\vec{\lambda}}(\mathbf{X}_2)$.*

Proof. Observe the following

$$\begin{aligned}
&- \lambda_0 \alpha I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_{11}, \mathbf{Y}_{12}|W) - \lambda_0 \bar{\alpha} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_{21}, \mathbf{Y}_{22}|W) + (\lambda_1 + \lambda_2) I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_{21}, \mathbf{Y}_{22}|W) \\
&\quad + \lambda_1 S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_1, \mathbf{X}_2|W) \\
&\leq -\lambda_0 \alpha I(\mathbf{X}_1; \mathbf{Y}_{11}|W) - \lambda_0 \alpha I(\mathbf{X}_2; \mathbf{Y}_{12}|W, \mathbf{Y}_{11}) - \lambda_0 \bar{\alpha} I(\mathbf{X}_2; \mathbf{Y}_{22}|W) - \lambda_0 \bar{\alpha} I(\mathbf{X}_1; \mathbf{Y}_{21}|W, \mathbf{Y}_{22}) \\
&\qquad + (\lambda_1 + \lambda_2) I(\mathbf{X}_2; \mathbf{Y}_{22}|W) + (\lambda_1 + \lambda_2) I(\mathbf{X}_1; \mathbf{Y}_{21}|W, \mathbf{Y}_{22}) + \lambda_1 S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_1|W, \mathbf{Y}_{22}) + \lambda_1 S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_2|W, \mathbf{Y}_{11}) \\
&\leq -\lambda_0 \alpha I(\mathbf{X}_1; \mathbf{Y}_{11}|W, \mathbf{Y}_{22}) - \lambda_0 \alpha I(\mathbf{X}_2; \mathbf{Y}_{12}|W, \mathbf{Y}_{11}) - \lambda_0 \bar{\alpha} I(\mathbf{X}_2; \mathbf{Y}_{22}|W, \mathbf{Y}_{11}) - \lambda_0 \bar{\alpha} I(\mathbf{X}_1; \mathbf{Y}_{21}|W, \mathbf{Y}_{22}) \\
&\qquad + (\lambda_1 + \lambda_2) I(\mathbf{X}_2; \mathbf{Y}_{22}|W, \mathbf{Y}_{11}) + (\lambda_1 + \lambda_2) I(\mathbf{X}_1; \mathbf{Y}_{21}|W, \mathbf{Y}_{22}) + \lambda_1 S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_1|W, \mathbf{Y}_{22}) \\
&\qquad + \lambda_1 S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_2|W, \mathbf{Y}_{11}) - (\lambda_0 - \lambda_1 - \lambda_2) I(\mathbf{Y}_{11}; \mathbf{Y}_{22}|W) \\
&\leq T_{\vec{\lambda}}(\mathbf{X}_1|\mathbf{Y}_{22}) + T_{\vec{\lambda}}(\mathbf{X}_2|\mathbf{Y}_{11}) - (\lambda_0 - \lambda_1 - \lambda_2) I(\mathbf{Y}_{11}; \mathbf{Y}_{22}|W) \\
&\leq T_{\vec{\lambda}}(\mathbf{X}_1) + T_{\vec{\lambda}}(\mathbf{X}_2) - (\lambda_0 - \lambda_1 - \lambda_2) I(\mathbf{Y}_{11}; \mathbf{Y}_{22}|W).
\end{aligned}$$

Since equality holds, using Claim 1 we have $\mathbf{X}_1$ and $\mathbf{X}_2$ are conditionally independent of W. Further using this and the equality observe that

$$\begin{aligned}
&- \lambda_0 \alpha I(\mathbf{X}_1; \mathbf{Y}_{11}|W, \mathbf{Y}_{22}) - \lambda_0 \bar{\alpha} I(\mathbf{X}_1; \mathbf{Y}_{21}|W, \mathbf{Y}_{22}) + (\lambda_1 + \lambda_2) I(\mathbf{X}_1; \mathbf{Y}_{21}|W, \mathbf{Y}_{22}) + \lambda_1 S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_1|W, \mathbf{Y}_{22}) \\
&\quad = -\lambda_0 \alpha I(\mathbf{X}_1; \mathbf{Y}_{11}|W) - \lambda_0 \bar{\alpha} I(\mathbf{X}_1; \mathbf{Y}_{21}|W) + (\lambda_1 + \lambda_2) I(\mathbf{X}_1; \mathbf{Y}_{21}|W) + \lambda_1 S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_1|W) \\
&\quad = T_{\vec{\lambda}}(\mathbf{X}_1).
\end{aligned}$$

Similarly for $\mathbf{X}_2$. This completes the proof. ☐

Remark: The above claim is the equivalent of Claim 5.

For $K \succeq 0$, define

$$\hat{\mathrm{V}}_{\vec{\lambda}}(K) = \sup_{\mathbf{X}: E(\mathbf{X}\mathbf{X}^T) \preceq K} T_{\vec{\lambda}}(\mathbf{X}).$$

Claim 9. *There exists a pair $(W_*, \mathbf{X}_*)$ with $|W_*| \leq \frac{t(t+1)}{2} + 1$ such that $\hat{\mathrm{V}}_{\vec{\lambda}}(K) = \mathsf{t}_{\vec{\lambda}}(\mathbf{X}_*|W_*)$.*

Proof. The proof of this claim is relegated to the Appendix. ☐

Notation: In the remainder of the section we assume that $p_*(w, \mathbf{x})$ achieves $\hat{\mathrm{V}}_{\vec{\lambda}}(K)$, $|W| = m \leq \frac{t(t+1)}{2} + 1$ and $\mathbf{X}_w$ be a centered random variable (zero-mean) distributed according to $p(\mathbf{x}|W = w)$. Further let $K_w = \mathrm{E}(\mathbf{X}_w \mathbf{X}_w^T)$. Then we have $\sum_{w=1}^m p_*(w) K_w \preceq K$ and in particular that K_w's are bounded.

Claim 10. *Let $(W_1, W_2, \mathbf{X}_1, \mathbf{X}_2) \sim p_*(w_1, \mathbf{x}_1) p_*(w_2, \mathbf{x}_2)$ be two i.i.d. copies of $p_*(w, x)$. We assume that $|W| \leq \frac{t(t+1)}{2} + 1$. Let*

$$\tilde{W} = (W_1, W_2), \quad \tilde{\mathbf{X}}|\big(\tilde{W} = (w_1, w_2)\big) \sim \frac{1}{\sqrt{2}} \left(\mathbf{X}_{w_1} + \mathbf{X}_{w_2}\right), \quad \mathbf{X}'|\big(\tilde{W} = (w_1, w_2)\big) \sim \frac{1}{\sqrt{2}} \left(\mathbf{X}_{w_1} - \mathbf{X}_{w_2}\right).$$

In the above we take $\mathbf{X}_{w_1}$ and $\mathbf{X}_{w_2}$ to be independent random variables. Then the following hold:

1. $\tilde{\mathbf{X}}, \mathbf{X}'$ *are conditionally independent given* $\tilde{W}$.
2. $\tilde{W}, \tilde{\mathbf{X}}$ *achieves* $\hat{\mathrm{V}}_{\vec{\lambda}}(K)$.
3. $\tilde{W}, \mathbf{X}'$ *achieves* $\hat{\mathrm{V}}_{\vec{\lambda}}(K)$.

Proof.

$$\begin{aligned}
2\hat{\mathrm{V}}_{\vec{\lambda}}(K) &= \mathsf{t}_{\vec{\lambda}}(\mathbf{X}_1|W_1) + \mathsf{t}_{\vec{\lambda}}(\mathbf{X}_2|W_2) \\
&= \mathsf{t}_{\vec{\lambda}}(\mathbf{X}_1, \mathbf{X}_2|W_1, W_2) \\
&\overset{(a)}{=} \mathsf{t}_{\vec{\lambda}}(\tilde{\mathbf{X}}, \mathbf{X}'|\tilde{W}) \\
&\overset{(b)}{\leq} T_{\vec{\lambda}}(\tilde{\mathbf{X}}, \mathbf{X}') \\
&\overset{(c)}{\leq} T_{\vec{\lambda}}(\tilde{\mathbf{X}}) + T_{\vec{\lambda}}(\mathbf{X}') \\
&\leq \hat{\mathrm{V}}_{\vec{\lambda}}(K) + \hat{\mathrm{V}}_{\vec{\lambda}}(K) = 2\hat{\mathrm{V}}_{\vec{\lambda}}(K).
\end{aligned}$$

The proof mirrors that of Claim 7. Here the first equality comes because $p_*(w, \mathbf{x})$ achieves $\hat{\mathrm{V}}_{\vec{\lambda}}(K)$, the second one because $(W_1, \mathbf{X}_1)$ and $(W_2, \mathbf{X}_2)$ are independent. Equality (a) is a consequence of Claim 1, inequality (c) is a consequence of Claim 8, and the last inequality follows from the following:

$$\mathrm{E}(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T) = \mathrm{E}(\mathbf{X}'\mathbf{X}'^T) = \sum_{w_1, w_2} p_*(w_1)p_*(w_2)\frac{(K_{w_1} + K_{w_2})}{2} = \sum_{w=1}^{m} p_*(w)K_w \preceq K,$$

and the definition of $\hat{\mathrm{V}}_{\vec{\lambda}}(K)$. Since the extremes match, all inequalities must be equalities. Hence (b) must be an equality, $p(\tilde{w}, \tilde{\mathbf{x}}, \mathbf{x}')$ achieves $T_{\vec{\lambda}}(\tilde{\mathbf{X}}, \mathbf{X}')$; and since (c) is also equality from Claim 8 we conclude that $\tilde{\mathbf{X}}, \mathbf{X}'$ are conditionally independent of $\tilde{W}$. Furthermore, we also obtain that $p(\tilde{w}|\tilde{\mathbf{x}})$ achieves $T_{\vec{\lambda}}(\tilde{\mathbf{X}})$, which from the last inequality matches $\hat{\mathrm{V}}_{\vec{\lambda}}(K)$. Similarly for $p(\tilde{w}|\mathbf{x}')$. ☐

As a consequence, $\mathbf{X}_{w_1}$, $\mathbf{X}_{w_2}$ are independent random variables and $(\mathbf{X}_{w_1} + \mathbf{X}_{w_2}), (\mathbf{X}_{w_1} - \mathbf{X}_{w_2})$ are also independent random variables. Thus from Corollary 3 (in Appendix A.1) $\mathbf{X}_{w_1}, \mathbf{X}_{w_2}$ are Gaussians, say having the same distribution as $\mathbf{X}_w \sim \mathcal{N}(0, K')$. Since w_1, w_2 are arbitrary, all $\mathbf{X}_{w_i}$ are Gaussians, having the same distribution as $\mathbf{X}_w$. Then

$$\hat{\mathrm{V}}_{\vec{\lambda}}(K) = \sum_{i=1}^{m} p_*(w_i)\mathsf{t}_{\vec{\lambda}}(\mathbf{X}_{w_i}) = \sum_{i=1}^{m} p_*(w_i)\mathsf{t}_{\vec{\lambda}}(\mathbf{X}_w) = \mathsf{t}_{\vec{\lambda}}(\mathbf{X}_w).$$

Hence we obtain the following theorem. The proof of uniqueness is just as in Theorem 1.

Theorem 2. *There exists* $\mathbf{X}_* \sim \mathcal{N}(0, K'), K' \preceq K$ *such that* $\hat{\mathrm{V}}_{\vec{\lambda}}(K) = \mathsf{t}_{\vec{\lambda}}(\mathbf{X}_*)$. *Further the maximizing distribution is unique.*

Corollary 2. *If* $\mathbf{X} \sim \mathcal{N}(0, K)$ *then there exists* $\mathbf{X}_{1*} \sim \mathcal{N}(0, K_1)$ *and an independent random variable* $\mathbf{X}_{2*} \sim \mathcal{N}(0, K_2), K_1 + K_2 = K' \preceq K$ *such that* $T_{\vec{\lambda}}(\mathbf{X}) = \mathsf{t}_{\vec{\lambda}}(\mathbf{X}_{1*} + \mathbf{X}_{2*}) = \hat{\mathrm{V}}_{\vec{\lambda}}(K)$ *and* $S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_{1*} + \mathbf{X}_{2*}) = \mathsf{s}_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_{1*}) = \mathrm{V}_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(K_1 + K_2)$.

Proof. Clearly from Theorem 2 and definition of $\hat{\mathrm{V}}_{\vec{\lambda}}(K)$ we have

$$T_{\vec{\lambda}}(\mathbf{X}) \leq \hat{\mathrm{V}}_{\vec{\lambda}}(K) = \mathsf{t}_{\vec{\lambda}}(\mathbf{X}_*).$$

On the other hand let $\mathbf{X}' \sim \mathcal{N}(0, K - K')$ be independent of $\mathbf{X}_*$. Note that $\mathbf{X} \sim \mathbf{X}' + \mathbf{X}_*$ and

$$T_{\vec{\lambda}}(\mathbf{X}) = \sup_{W} \mathsf{t}_{\vec{\lambda}}(\mathbf{X}|W) \geq \mathsf{t}_{\vec{\lambda}}(\mathbf{X}|\mathbf{X}') = \mathsf{t}_{\vec{\lambda}}(\mathbf{X}_*).$$

Now splitting of $\mathbf{X}_*$ into $\mathbf{X}_{1*}$, $\mathbf{X}_{2*}$ is possible by Corollary 1. ☐

3 Two capacity regions

3.1 Vector Gaussian Broadcast channel with private messages

Consider a vector Gaussian broadcast channel with only private message requirements. Let $\mathcal{C}$ be the capacity region. For $\lambda > 1$ we will seek to maximize the following expression

$$\max_{(R_1,R_2)\in\mathcal{C}} R_1 + \lambda R_2.$$

The case for $\lambda < 1$ is dealt with similarly (with roles of (Y_1, Y_2) interchanged). The case for $\lambda = 1$ follows by continuity.

Remark 8. To show the $\lambda = 1$ case, observe that the function

$$C(\lambda) := \max_{(R_1,R_2)\in\mathcal{C}} \lambda R_1 + R_2$$

is convex and bounded in λ when $\lambda \in [0, 2]$ (more generally, any closed and bounded interval). In the above definition $\mathcal{C}$ refers to the capacity region. Hence $C(\lambda)$ is continuous in λ at $\lambda = 1$.

Here we consider the Korner-Marton outer bound and Marton's inner bound (both from [12]) to the capacity region of the broadcast channel.

Bound 1. *The union of rate pairs (R_1, R_2) satisfying*

$$\begin{aligned} R_2 &\le I(V;Y_2)\\ R_1 &\le I(X;Y_1)\\ R_1 + R_2 &\le I(V;Y_2) + I(X;Y_1|V) \end{aligned}$$

over all $V \to X \to (Y_1, Y_2)$ forms an outer bound to the broadcast channel.

Denote this region as $\mathcal{O}$.

Bound 2. *The union of rate pairs (R_1, R_2) satisfying*

$$\begin{aligned} R_2 &\le I(V;Y_2)\\ R_1 &\le I(U;Y_1)\\ R_1 + R_2 &\le I(U;Y_1) + I(V;Y_2) - I(U;V) \end{aligned}$$

over all $(U, V) \to X \to (Y_1, Y_2)$ forms an inner bound to the broadcast channel.

Denote this region as $\mathcal{I}$.

One can adapt these inner and outer bounds to additive Gaussian setting by introducing a power constraint, i.e. an upper bound on the trace of the covariance matrix, $\mathrm{tr}(K)$. However let us put a covariance constraint on $\mathbf{X}$ and denote $\mathcal{I}_K, \mathcal{C}_K, \mathcal{O}_K$ to be the corresponding inner bound, capacity region, and the outer bound.

Clearly we have

$$\max_{(R_1,R_2)\in\mathcal{I}_K} R_1 + \lambda R_2 \le \max_{(R_1,R_2)\in\mathcal{C}_K} R_1 + \lambda R_2 \le \max_{(R_1,R_2)\in\mathcal{O}_K} R_1 + \lambda R_2.$$

To exhibit the capacity region we will show that

$$\max_{(R_1,R_2)\in\mathcal{O}_K} R_1 + \lambda R_2 \le \max_{(R_1,R_2)\in\mathcal{I}_K} R_1 + \lambda R_2.$$

Thus Marton's inner bound and Korner-Marton's outer bound will match in this setting, and therefore also with the usual trace constraint.

Observe that

$$\begin{aligned}\max_{(R_1,R_2)\in\mathcal{O}_K} R_1+\lambda R_2 &\le \sup_{\substack{V\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \lambda I(V;\mathbf{Y}_2)+I(\mathbf{X};\mathbf{Y}_1|V)\\ &= \sup_{\substack{V\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \lambda I(\mathbf{X};\mathbf{Y}_2)+I(\mathbf{X};\mathbf{Y}_1|V)-\lambda I(\mathbf{X};\mathbf{Y}_2|V)\\ &\le \max_{\mathbf{X}:\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K} \lambda I(\mathbf{X};\mathbf{Y}_2)+\sup_{\substack{V\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} I(\mathbf{X};\mathbf{Y}_1|V)-\lambda I(\mathbf{X};\mathbf{Y}_2|V)\\ &\le \max_{\mathbf{X}:\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K} \lambda I(\mathbf{X};\mathbf{Y}_2)+\mathrm{V}_\lambda(K).\end{aligned}$$

We know that the first term is maximized (Section 2.1) when $\mathbf{X}\sim\mathcal{N}(0,K)$ and $\mathrm{V}_\lambda(K)$ is achieved by $\mathsf{s}_\lambda(\mathbf{X}_*)$ where $\mathbf{X}_*\sim\mathcal{N}(0,K'),K'\preceq K$. Now let $V_*\sim\mathcal{N}(0,K-K')$ be independent of $\mathbf{X}_*$ and let $\mathbf{X}=V_*+\mathbf{X}_*$. Observe that this choice attains both maxima simultaneously. Hence

$$\max_{(R_1,R_2)\in\mathcal{O}_K} R_1+\lambda R_2 \le \lambda I(V_*;\mathbf{Y}_2)+I(\mathbf{X};\mathbf{Y}_1|V_*)=\lambda I(V_*;\mathbf{Y}_2)+I(\mathbf{X}_*;\mathbf{Y}_1|V_*).$$

Lemma 2 (Dirty paper coding)**.** *Let $\mathbf{X}=V_*+\mathbf{X}_*$ and V_*, $\mathbf{X}_*$ be independent Gaussians with covariances $K-K',K'$ respectively. Then there exists U_* jointly Gaussian with V_* such that*

$$I(\mathbf{X};\mathbf{Y}_1|V_*)=I(U_*;\mathbf{Y}_1)-I(U_*;V_*).$$

Here $\mathbf{Y}_1=G\mathbf{X}+\mathbf{Z}$, where $\mathbf{Z}\sim\mathcal{N}(0,I)$ is independent of $V_*,\mathbf{X}_*$.

Proof. This well-known identification stems from the celebrated paper[3]. Set $U_*=\mathbf{X}_*+AV_*$ where $A=K'G^T(GK'G^T+I)^{-1}$ and this works (see Chapter 9.5 of [6]). □

Now using U_* as in the above lemma, we obtain

$$\begin{aligned}\max_{(R_1,R_2)\in\mathcal{O}_K} R_1+\lambda R_2 &\le \lambda I(V_*;\mathbf{Y}_2)+I(\mathbf{X}_*;\mathbf{Y}_1|V_*)\\ &=\lambda I(V_*;\mathbf{Y}_2)+I(U_*;\mathbf{Y}_1)-I(U_*;V_*).\end{aligned}$$

However using Marton's inner bound any rate pair satisfying $R_2=I(V;\mathbf{Y}_2)$, $R_1=I(U;\mathbf{Y}_1)-I(U;V)$ such that $\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K$ belongs to $\mathcal{I}_K$. Hence

$$\max_{(R_1,R_2)\in\mathcal{O}_K} R_1+\lambda R_2\le\lambda I(V_*;\mathbf{Y}_2)+I(U_*;\mathbf{Y}_1)-I(U_*;V_*)\le\max_{(R_1,R_2)\in\mathcal{I}_K} R_1+\lambda R_2.$$

Thus the inner and outer bounds match for vector Gaussian product channels establishing its capacity region.

3.2 Vector Gaussian Broadcast channel with common message

Consider a vector Gaussian broadcast channel with common and private message requirements. Let $\mathcal{C}$ be the capacity region. Assume $\lambda_0>\lambda_1+\lambda_2$. We will seek to maximize the following expression

$$\max_{(R_0,R_1,R_2)\in\mathcal{C}} \lambda_0R_0+\lambda_1R_1+(\lambda_1+\lambda_2)R_2.$$

Remark 9. The case of maximizing $\lambda_0R_0+(\lambda_1+\lambda_2)R_1+\lambda_2R_2$ can be dealt with similarly. On the other hand if $\lambda_0\le(\lambda_1+\lambda_2)$ then it suffices to consider the private messages capacity region. Actually the setting $\lambda_0\ge 2\lambda_1+\lambda_2$ can be deduced from the degraded message sets capacity region and this is also known; however this will be subsumed in our treatment. Hence the setting we are considering is the only interesting unestablished case.

In this section we consider the UVW outer bound[13] and Marton's inner bound[12] to the capacity region of the broadcast channel with private and common messages.

Bound 3 (UVW outer bound)**.** *The union of rate triples* (R_0, R_1, R_2) *satisfying*

$$\begin{aligned} R_0 &\le \min\{I(W;Y_1), I(W;Y_2)\} \\ R_0 + R_1 &\le \min\{I(W;Y_1), I(W;Y_2)\} + I(U;Y_1|W) \\ R_0 + R_2 &\le \min\{I(W;Y_1), I(W;Y_2)\} + I(V;Y_2|W) \\ R_0 + R_1 + R_2 &\le \min\{I(W;Y_1), I(W;Y_2)\} + I(V;Y_2|W) + I(X;Y_1|V,W) \\ R_0 + R_1 + R_2 &\le \min\{I(W;Y_1), I(W;Y_2)\} + I(U;Y_1|W) + I(X;Y_2|U,W) \end{aligned}$$

over all $(U,V,W) \to X \to (Y_1, Y_2)$ *forms an outer bound to the broadcast channel.*

As before, denote this region as $\mathcal{O}$.

Bound 4 (Marton's inner bound)**.** *The union of rate pairs* (R_1, R_2) *satisfying*

$$\begin{aligned} R_0 &\le \min\{I(W;Y_1), I(W;Y_2)\} \\ R_0 + R_1 &\le I(U,W;Y_1) \\ R_0 + R_2 &\le I(V,W;Y_2) \\ R_0 + R_1 + R_2 &\le \min\{I(W;Y_1), I(W;Y_2)\} + I(U;Y_1|W) + I(V;Y_2|W) - I(U;V|W) \end{aligned}$$

over all $(U,V) \to X \to (Y_1, Y_2)$ *forms an inner bound to the broadcast channel.*

Denote this region as $\mathcal{I}$.

Impose a covariance constraint K on $\mathbf{X}$ and denote $\mathcal{I}_K, \mathcal{C}_K, \mathcal{O}_K$ to be the corresponding inner bound, capacity region, and the outer bound respectively. Trivially we have

$$\begin{aligned} \max_{(R_0,R_1,R_2)\in\mathcal{I}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2) R_2 &\le \max_{(R_0,R_1,R_2)\in\mathcal{C}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2) R_2 \\ &\le \max_{(R_0,R_1,R_2)\in\mathcal{O}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2) R_2. \end{aligned}$$

For any $\alpha \in [0,1]$ observe that (from first, third, and fourth constraints of UVW outer bound)

$$\begin{aligned} &\max_{(R_0,R_1,R_2)\in\mathcal{O}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2) R_2 \\ &\le \sup_{\substack{(V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2) \\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V,W) \\ &= \sup_{\substack{(V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2) \\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \alpha\lambda_0 I(X;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(X;\mathbf{Y}_2) - \alpha\lambda_0 I(\mathbf{X};\mathbf{Y}_1|W) - \bar{\alpha}\lambda_0 I(\mathbf{X};\mathbf{Y}_2|W) \\ &\qquad\qquad + (\lambda_1+\lambda_2) I(\mathbf{X};\mathbf{Y}_2|W) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V,W) - (\lambda_1+\lambda_2) I(\mathbf{X};\mathbf{Y}_2|V,W) \\ &\le \sup_{\substack{W\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2) \\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \alpha\lambda_0 I(X;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(X;\mathbf{Y}_2) - \alpha\lambda_0 I(\mathbf{X};\mathbf{Y}_1|W) - \bar{\alpha}\lambda_0 I(\mathbf{X};\mathbf{Y}_2|W) \\ &\qquad\qquad + (\lambda_1+\lambda_2) I(\mathbf{X};\mathbf{Y}_2|W) + \lambda_1 S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}|W) \\ &\le \max_{\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K} (\alpha\lambda_0 I(X;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(X;\mathbf{Y}_2)) + \max_{\substack{W\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2) \\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \mathsf{t}_{\vec{\lambda}}(\mathbf{X}|W) \\ &\le \max_{\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K} (\alpha\lambda_0 I(X;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(X;\mathbf{Y}_2)) + \hat{\mathrm{V}}_{\vec{\lambda}}(K). \end{aligned}$$

We know that the first term is maximized (Section 2.3) when $\mathbf{X} \sim \mathcal{N}(0,K)$ and $\hat{\mathrm{V}}_{\vec{\lambda}}(K)$ is achieved by $\mathsf{t}_{\vec{\lambda}}(\mathbf{X}_{1*} + \mathbf{X}_{2*})$ where $\mathbf{X}_{1*}, \mathbf{X}_{2*}$ are independent and $\mathbf{X}_{1*} \sim \mathcal{N}(0,K_1), \mathbf{X}_{2*} \sim \mathcal{N}(0,K_2), K_1 + K_2 \preceq K$, and $S_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_{1*} + \mathbf{X}_{2*}) = \mathsf{s}_{\frac{\lambda_1+\lambda_2}{\lambda_1}}(\mathbf{X}_{1*})$. See Theorem 2 and Corollary 2. Now let $W_* \sim \mathcal{N}(0, K - (K_1 + K_2))$ be independent of $\mathbf{X}_{1*}, \mathbf{X}_{2*}$ and let $\mathbf{X} = W_* + \mathbf{X}_{1*} + \mathbf{X}_{2*}$. Observe that this choice attains both maxima simultaneously. For conforming to more standard notation, let us call $V_* = \mathbf{X}_{2*}$, thus $\mathbf{X} = W_* + \mathbf{X}_{1*} + V_*$.

Thus

$$\begin{aligned}
&\max_{(R_0,R_1,R_2)\in\mathcal{O}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2) R_2 \\
&\le \alpha\lambda_0 I(X;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(X;\mathbf{Y}_2) - \alpha\lambda_0 I(\mathbf{X};\mathbf{Y}_1|W_*) - \bar{\alpha}\lambda_0 I(\mathbf{X};\mathbf{Y}_2|W_*) \\
&\qquad + (\lambda_1+\lambda_2) I(\mathbf{X};\mathbf{Y}_2|W_*) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V_*,W_*) - (\lambda_1+\lambda_2) I(\mathbf{X};\mathbf{Y}_2|V_*,W_*) \\
&= \alpha\lambda_0 I(W_*;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W_*;\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(V_*;\mathbf{Y}_2|W_*) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V_*,W_*) \\
&= \alpha\lambda_0 I(W_*;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W_*;\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(V_*;\mathbf{Y}_2|W_*) + \lambda_1 I(\mathbf{X}_{1*};\mathbf{Y}_1|V_*,W_*)
\end{aligned}$$

Now using Lemma 2 choose $U_* = X_{1*} + \tilde{A}V_*$ as before to have

$$I(\mathbf{X}_{1*};\mathbf{Y}_1|V_*,W_*) = I(U_*;\mathbf{Y}_1|W_*) - I(U_*;V_*|W_*).$$

Hence

$$\begin{aligned}
&\max_{(R_0,R_1,R_2)\in\mathcal{O}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2) R_2 \\
&\le \alpha\lambda_0 I(W_*;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W_*;\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(V_*;\mathbf{Y}_2|W_*) + \lambda_1 (I(U_*;\mathbf{Y}_1|W_*) - I(U_*;V_*|W_*)) \\
&\le \sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) + \lambda_1 (I(U;\mathbf{Y}_1|W) - I(U;V|W))
\end{aligned}$$

Since the above holds for all $\alpha \in [0,1]$, we have

$$\begin{aligned}
&\max_{(R_0,R_1,R_2)\in\mathcal{O}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2) R_2 \\
&\le \min_{\alpha\in[0,1]} \sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) + \lambda_1 I(U;\mathbf{Y}_1|W) - \lambda_1 I(U;V|W).
\end{aligned}$$

To complete the proof that the inner and outer bounds match we present the following Claim 11 (essentially established in [7]). We will defer the proof of this claim to the Appendix A.2.

Claim 11. *We claim that*

$$\begin{aligned}
&\min_{\alpha\in[0,1]} \sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) + \lambda_1 I(U;\mathbf{Y}_1|W) - \lambda_1 I(U;V|W) \\
&= \sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \min_{\alpha\in[0,1]} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) + \lambda_1 I(U;\mathbf{Y}_1|W) - \lambda_1 I(U;V|W) \\
&= \sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \lambda_0 \min\{I(W;\mathbf{Y}_1), I(W;\mathbf{Y}_2)\} + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) + \lambda_1 I(U;\mathbf{Y}_1|W) - \lambda_1 I(U;V|W).
\end{aligned}$$

Now using Marton's inner bound we can always achieve the following triples: $R_0 = \min\{I(W;\mathbf{Y}_1), I(W;\mathbf{Y}_2)\}$, $R_2 = I(V;\mathbf{Y}_2|W)$, $R_1 = I(U;\mathbf{Y}_1|W) - I(U;V|W)$. Hence

$$\begin{aligned}
&\max_{(R_0,R_1,R_2)\in\mathcal{O}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2) R_2 \\
&\le \sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \lambda_0 \min\{I(W;\mathbf{Y}_1), I(W;\mathbf{Y}_2)\} + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) + \lambda_1 I(U;\mathbf{Y}_1|W) - \lambda_1 I(U;V|W) \\
&\le \max_{(R_0,R_1,R_2)\in\mathcal{I}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2) R_2.
\end{aligned}$$

Hence Marton's inner bound and UVW outer bound match.

3.3 An explicit representation

The boundary is achieved via Gaussian signaling. We will show here that the capacity region established here matches the region given by equations (2) – (4) in [17]. What we have established in the above arguments can be phrased as

$$\begin{aligned}&\max_{(R_0,R_1,R_2)\in\mathcal{C}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2)R_2\\ &\quad = \min_{\alpha\in[0,1]} \max_{\substack{K_w,K_v\succeq 0\\ K_w+K_v\preceq K}} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2)I(V;\mathbf{Y}_2|W) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V,W),\end{aligned}$$

where W,V are independent zero-mean Gaussians with covariances K_w, K_v respectively, and $X \sim U+V+W$, where U is another zero-mean Gaussian independent of W,V having covariance $K-K_w-K_v$.

The region, $\mathcal{R}_K$, implied by the equations (2) – (4) in [17] can be cast as

$$\begin{aligned}&\max_{(R_0,R_1,R_2)\in\mathcal{R}_K} \lambda_0 R_0 + \lambda_1 R_1 + (\lambda_1+\lambda_2)R_2\\ &\quad = \max_{\substack{K_w,K_v\succeq 0\\ K_w+K_v\preceq K}} \lambda_0 \min\{I(W;\mathbf{Y}_1), I(W;\mathbf{Y}_2)\} + (\lambda_1+\lambda_2)I(V;\mathbf{Y}_2|W) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V,W),\end{aligned}$$

where (as before) W,V are independent zero-mean Gaussians with covariances K_w, K_v respectively, and $X \sim U+V+W$, where U is another zero-mean Gaussian independent of W,V having covariance $K-K_w-K_v$.

The main result of this section is to show that $\mathcal{R}_K = \mathcal{C}_K$, i.e. in particular the following claim.

Claim 12. *The following max-min interchange holds:*

$$\begin{aligned}&\min_{\alpha\in[0,1]} \max_{\substack{K_w,K_v\succeq 0\\ K_w+K_v\preceq K}} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2)I(V;\mathbf{Y}_2|W) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V,W)\\ &\quad = \max_{\substack{K_w,K_v\succeq 0\\ K_w+K_v\preceq K}} \lambda_0 \min\{I(W;\mathbf{Y}_1), I(W;\mathbf{Y}_2)\} + (\lambda_1+\lambda_2)I(V;\mathbf{Y}_2|W) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V,W),\end{aligned}$$

where the distributions of W,V,X are as described earlier.

Proof. The non-trivial direction is to establish that

$$\begin{aligned}&\min_{\alpha\in[0,1]} \max_{\substack{K_w,K_v\succeq 0\\ K_w+K_v\preceq K}} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2)I(V;\mathbf{Y}_2|W) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V,W)\\ &\quad \le \max_{\substack{K_w,K_v\succeq 0\\ K_w+K_v\preceq K}} \lambda_0 \min\{I(W;\mathbf{Y}_1), I(W;\mathbf{Y}_2)\} + (\lambda_1+\lambda_2)I(V;\mathbf{Y}_2|W) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V,W).\end{aligned}$$

Let us define

$$SR_{\vec{\lambda}}(\alpha) := \max_{\substack{K_w,K_v\succeq 0\\ K_w+K_v\preceq K}} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2)I(V;\mathbf{Y}_2|W) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V,W),$$

and let $W_\alpha, V_\alpha, \mathbf{X}_\alpha$ be the unique maximizing Gaussian distributions that achieve the maximum value. These maximizers exist because we are on a compact set of covariance matrices. The uniqueness is also stated in Theorem 2.

Now let $\alpha^* \in [0,1]$ be the minimizer of $SR_{\vec{\lambda}}(\alpha)$. The following inequalities hold for any $\beta \in [0,1]$ and are clear from the definitions:

$$\begin{aligned}&\alpha^*\lambda_0 I(W_\beta;\mathbf{Y}_1) + \bar{\alpha^*}\lambda_0 I(W_\beta;\mathbf{Y}_2) + (\lambda_1+\lambda_2)I(V_\beta;\mathbf{Y}_2|W_\beta) + \lambda_1 I(\mathbf{X}_\beta;\mathbf{Y}_1|V_\beta,W_\beta)\\ &\le \alpha^*\lambda_0 I(W_{\alpha^*};\mathbf{Y}_1) + \bar{\alpha^*}\lambda_0 I(W_{\alpha^*};\mathbf{Y}_2) + (\lambda_1+\lambda_2)I(V_{\alpha^*};\mathbf{Y}_2|W_{\alpha^*}) + \lambda_1 I(\mathbf{X}_{\alpha^*};\mathbf{Y}_1|V_{\alpha^*},W_{\alpha^*})\\ &\le \beta\lambda_0 I(W_\beta;\mathbf{Y}_1) + \bar{\beta}\lambda_0 I(W_\beta;\mathbf{Y}_2) + (\lambda_1+\lambda_2)I(V_\beta;\mathbf{Y}_2|W_\beta) + \lambda_1 I(\mathbf{X}_\beta;\mathbf{Y}_1|V_\beta,W_\beta).\end{aligned}$$

Comparing the first and last term, we obtain that

$$(\alpha^*-\beta)I(W_\beta;\mathbf{Y}_1) \le (\alpha^*-\beta)I(W_\beta;\mathbf{Y}_2).$$

Case 1: If $\alpha^* \in (0,1)$ we have that $I(W_\beta;\mathbf{Y}_1) \geq I(W_\beta;\mathbf{Y}_2)$ whenever $\beta \leq \alpha^*$ and $I(W_\beta;\mathbf{Y}_1) \leq I(W_\beta;\mathbf{Y}_2)$ whenever $\beta \geq \alpha^*$. Thus by continuity of $SR_{\vec{\lambda}}(\alpha)$ (in α), by compactness of the space of all possible $K_v, K_w : K_v, K_w \succeq 0, K_v + K_w \preceq K$, and by the uniqueness of the maximizer at any $\alpha \in [0,1]$ we have that $I(W_{\alpha^*};\mathbf{Y}_1) = I(W_{\alpha^*};\mathbf{Y}_2)$. This then implies that

$$SR_{\vec{\lambda}}(\alpha^*) = \lambda_0 \min\{I(W_{\alpha^*};\mathbf{Y}_1), I(W_{\alpha^*};\mathbf{Y}_2)\} + (\lambda_1+\lambda_2) I(V_{\alpha^*};\mathbf{Y}_2|W_{\alpha^*}) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V_{\alpha^*}, W_{\alpha^*}),$$

and yields the desired result.

Case 2: If $\alpha^* = 0$ then a similar arguments using $\beta \geq 0$ yields that $I(W_0;\mathbf{Y}_1) \geq I(W_0;\mathbf{Y}_2)$, which then yields

$$SR_{\vec{\lambda}}(0) = \lambda_0 \min\{I(W_0;\mathbf{Y}_1), I(W_0;\mathbf{Y}_2)\} + (\lambda_1+\lambda_2) I(V_{\alpha^*};\mathbf{Y}_2|W_{\alpha^*}) + \lambda_1 I(\mathbf{X};\mathbf{Y}_1|V_{\alpha^*}, W_{\alpha^*}),$$

as desired. Case $\alpha^* = 1$ follows similarly.

Thus we have established the required max-min interchanged as claimed in Claim 12. □

Remark 10. The authors are grateful to Hon-Fah Chong and Yeow-Khiang Chia who alerted the authors that the earlier version did not contain a proof for this equivalence.

4 Conclusion

We developed a new method to show the optimality of Gaussian distributions. We illustrated this technique for three examples and computed the capacity region of the two-receiver vector Gaussian broadcast channel with common and private messages. We can see several other problems where this technique can have immediate impact. Some of the mathematical tools and results in the Appendix can also be of independent interest.

Acknowledgement

A lot of this work was motivated by the work on the discrete memoryless broadcast channel particularly the work by the authors in collaboration with Amin Gohari. The authors are also grateful to Venkat Anantharam, Abbas El Gamal, Amin Gohari, and Young-Han Kim for their comments on early drafts and suggestions on improving the presentation.

The work of Chandra Nair was partially supported by the following grants from the University Grants Committee of the Hong Kong Special Administrative Region, China: a) (Project No. AoE/E-02/08), b) GRF Project 415810. He also acknowledges the support from the Institute of Theoretical Computer Science and Communications (ITCSC) at the Chinese University of Hong Kong.

References

[1] Dennis D Boos, *A converse to scheffe's theorem*, Annals of Statistics **13** (1985), no. 1, 423–427.

[2] L.N.H. Bunt, *Bijdrage tot de theorie der convexe puntverzamelingen*, Ph.D. thesis, Univ. Groningne, Amsterdam, 1934.

[3] M. Costa, *Writing on dirty paper (corresp.)*, Information Theory, IEEE Transactions on **29** (1983), no. 3, 439 – 441.

[4] T Cover, *Broadcast channels*, IEEE Trans. Info. Theory **IT-18** (January, 1972), 2–14.

[5] R. Durrett, *Probability: Theory and examples*, second ed., Duxbury Press, 1996.

[6] Abbas El Gamal and Young-Han Kim, *Network information theory*, Cambridge University Press, 2012.

[7] 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 (2011), 1549–1553.

[8] Yanlin Geng, Amin Gohari, Chandra Nair, and Yuanming Yu, *The capacity region of classes of product broadcast channels*, 2011.

[9] S. G. Ghurye and Ingram Olkin, *A characterization of the multivariate normal distribution*, The Annals of Mathematical Statistics **33** (1962), no. 2, pp. 533–541 (English).

[10] Mahesh Godavarti and Alfred O. Hero, *Convergence of differential entropies*, IEEE Transactions on Information Theory **50** (2004), no. 1, 171–176.

[11] Tie Liu and P. Viswanath, *An extremal inequality motivated by multiterminal information-theoretic problems*, Information Theory, IEEE Transactions on **53** (2007), no. 5, 1839 –1851.

[12] K Marton, *A coding theorem for the discrete memoryless broadcast channel*, IEEE Trans. Info. Theory **IT-25** (May, 1979), 306–311.

[13] C Nair, *A note on outer bounds for broadcast channel*, Presented at International Zurich Seminar (2010).

[14] V. Yu Protasov and M. E. Shirokov, *Generalized compactness in linear spaces and its applications*, MATHEMATICS **200** (2009), no. 5, 697–722.

[15] V P Skitovic, *Linear combinations of independent random variables and the normal distribution law*, Select Transl Math Stat Probab **2** (1962), 211–228.

[16] H. Weingarten, Y. Steinberg, and S. Shamai, *The capacity region of the gaussian multiple-input multiple-output broadcast channel*, Information Theory, IEEE Transactions on **52** (2006), no. 9, 3936 –3964.

[17] ______, *On the capacity region of the multi-antenna broadcast channel with common messages*, Information Theory, 2006 IEEE International Symposium on, july 2006, pp. 2195 –2199.

A Some known results

A.1 A characterization of Gaussian distribution

Theorem 3 (Theorem 1 in [9]). *Let $\mathbf{X}_1, .., \mathbf{X}_n$ be n mutually independent t-dimensional random column vectors, and let $A_1, .., A_n$ and $B_1, ..., B_n$ be non-singular $t \times t$ matrices. If $\sum_{i=1}^n A_i \mathbf{X}_i$ is independent of $\sum_{i=1}^n B_i \mathbf{X}_i$, then the $\mathbf{X}_i$ are normally distributed.*

Remark 11. In this paper we only use A_i, B_i as multiples of I. In this case, the theorem follows from an earlier result of Skitovic [15]. There were scalar versions of this known since the 30s, including Bernstein's theorem. The proof relies on solving the functional equations satisfied by the characteristic functions.

Corollary 3. *If $\mathbf{X}_1$ and $\mathbf{X}_2$ are zero-mean independent t-dimensional random column vectors, and if $\mathbf{X}_1 + \mathbf{X}_2$ and $\mathbf{X}_1 - \mathbf{X}_2$ are independent then $\mathbf{X}_1, \mathbf{X}_2$ are normally distributed with identical covariances.*

Proof. The fact that $\mathbf{X}_1, \mathbf{X}_2$ are normally distributed follows from Theorem 3. Now observe that $\mathrm{E}((\mathbf{X}_1 + \mathbf{X}_2)(\mathbf{X}_1 - \mathbf{X}_2)^T) = \mathrm{E}(\mathbf{X}_1 + \mathbf{X}_2)\,\mathrm{E}(\mathbf{X}_1 - \mathbf{X}_2)^T = \mathbf{0}$. On the other hand

$$\mathrm{E}((\mathbf{X}_1 + \mathbf{X}_2)(\mathbf{X}_1 - \mathbf{X}_2)^T) = \mathrm{E}(\mathbf{X}_1 \mathbf{X}_1^T) - \mathrm{E}(\mathbf{X}_2 \mathbf{X}_2^T). \qquad \square$$

A.2 Min-max theorem

We reproduce the following Corollary from the Appendix of [7] (full version of this paper can be found in arXiv).

Corollary 4 (Corollary 2 in arXiv version of [7]). *Let Λ_d be the d-dimensional simplex, i.e. $\alpha_i \geq 0$ and $\sum_{i=1}^d \alpha_i = 1$. Let $\mathcal{P}$ be a set of probability distributions $p(u)$. Let $T_i(p(u)), i = 1, .., d$ be a set of functions such that the set $\mathcal{A}$, defined by*

$$\mathcal{A} = \{(a_1, a_2, ..., a_d) \in \mathbb{R}^d : a_i \leq T_i(p(u)) \text{ for some } p(u) \in \mathcal{P}\},$$

is a convex set.

Then

$$\sup_{p(u)\in\mathcal{P}} \min_{\alpha\in\Lambda_d} \sum_{i=1}^{d} \alpha_i T_i(p(u)) = \min_{\alpha\in\Lambda_d} \sup_{p(u)\in\mathcal{P}} \sum_{i=1}^{d} \alpha_i T_i(p(u)).$$

We will now show how one can use the Corollary 4 to establish Claim 11.

Proof of Claim 11

Proof. We take $\mathcal{P}$ as the set of $p(u,v,w,\mathbf{x})$ that satisfy the covariance constraint. Here we take $d=2$ and set

$$\begin{aligned} T_1(p(u,v,w,x)) &= \lambda_0 I(W;\mathbf{Y}_1) + \lambda_1 I(U;\mathbf{Y}_1|W) + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) - \lambda_1 I(U;V|W) \\ T_2(p(u,v,w,x)) &= \lambda_0 I(W;\mathbf{Y}_2) + \lambda_1 I(U;\mathbf{Y}_1|W) + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) - \lambda_1 I(U;V|W) \end{aligned}$$

It is clear that the set

$$\mathcal{A} = \{(a_1,a_2) : a_1 \le T_1(p(u,v,w,\mathbf{x})), a_2 \le T_2(p(u,v,w,\mathbf{x}))\}$$

is a convex set. (In the standard manner, choose $\tilde{W} = (W,Q)$, and when $Q=0$ choose $(U,V,W,\mathbf{X}) \sim p_1(u,v,w,\mathbf{x})$ and $Q=1$ choose $(U,V,W,\mathbf{X}) \sim p_2(u,v,w,\mathbf{x})$). Hence from Corollary 4, we have

$$\begin{aligned} &\min_{\alpha\in[0,1]} \sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) + \lambda_1 I(U;\mathbf{Y}_1|W) - \lambda_1 I(U;V|W) \\ &= \sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \min_{\alpha\in[0,1]} \alpha\lambda_0 I(W;\mathbf{Y}_1) + \bar{\alpha}\lambda_0 I(W;\mathbf{Y}_2) + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) + \lambda_1 I(U;\mathbf{Y}_1|W) - \lambda_1 I(U;V|W) \\ &= \sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}} \lambda_0 \min\{I(W;\mathbf{Y}_1), I(W;\mathbf{Y}_2)\} + (\lambda_1+\lambda_2) I(V;\mathbf{Y}_2|W) + \lambda_1 I(U;\mathbf{Y}_1|W) - \lambda_1 I(U;V|W). \end{aligned}$$

□

B Existence of maximizing distributions

The aim of this section is to give formal proofs of Claims 6 and 9 as our arguments critically hinge on proving properties of maximizing distributions. Our basic topological space consists of Borel probability measures on $\mathbb{R}^t$ endowed with the weak-convergence topology. This is a metric space with the Levy-Prokhorov metric defining the distance between two probability measures.

Remark 12. For the proofs in this section, it is not necessary to know the precise definition of the Levy-Prokhorov metric; but just that the topological space is a metric space and hence normal. Notation wise, most of the time we use random variables $\mathbf{X}$ instead of the induced probability measure to represent points on this space. We will also try to state the various theorems that we employ in this section as and when we use them.

B.1 Properties of Additive Gaussian noise

In this section, we will establish certain properties of distributions obtained according to $Y = X + Z$, where X and Z are independent and $Z \sim \mathcal{N}(0, I)$. For simplicity of notation, we consider the scalar case. Stronger forms of such smoothness results are very well known in certain mathematical circles and are used widely in the study of the heat equation. Here we present the results for completeness.

Let $\tilde{F}(x) = \mathrm{P}(X \le x)$ (where the inequality is co-ordinate wise. Note that $0 \le \tilde{F}(x) \le 1$. Then we see that since $f_z(z)$ has a density, we have

$$\mathrm{P}(Y \le y) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tilde{F}(y-z) dz.$$

Thus we have

$$\mathrm{P}(Y \le y+\delta) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tilde{F}(y+\delta-z)dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z+\delta)^2/2} \tilde{F}(y-z)dz.$$

By Dominated convergence theorem stated below (to justify interchange of derivative and integration) Y has a density given by

$$f_Y(y) = \lim_{\delta \to 0} \frac{1}{\delta}(\mathrm{P}(Y \le y+\delta) - \mathrm{P}(Y \le y)) = \int_{-\infty}^{\infty} \frac{-z}{\sqrt{2\pi}} e^{-z^2/2} \tilde{F}(y-z)dz.$$

Hence

$$|f_Y(y)| \le \int_{-\infty}^{\infty} \frac{|z|}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}}.$$

Again by Dominated convergence theorem we have

$$f_Y'(y) = \int_{-\infty}^{\infty} \frac{z^2-1}{\sqrt{2\pi}} e^{-z^2/2} \tilde{F}(y-z)dz.$$

Thus

$$|f_Y'(y)| \le \int_{-\infty}^{\infty} \frac{|z^2-1|}{\sqrt{2\pi}} e^{-z^2/2} dz \le 2.$$

Remark 13. Thus Y has a bounded density and a bounded first derivative of the density. In the vector case, similarly we have a bounded density and a uniformly bounded L_1 norm for $\nabla f_{\mathbf{Y}}(\mathbf{y})$.

Next, we state a general lemma which relates weak convergence to convergence of densities.

Lemma 3 (Lemma 1 in [1]). *Suppose that $\mathbf{Y}_n$ and $\mathbf{Y}$ have continuous densities $f_n(\mathbf{y}), f(\mathbf{y})$ with respect to the Lebesgue measure on $\mathbb{R}^t$. If $\mathbf{Y}_n \overset{w}{\Rightarrow} \mathbf{Y}$ and*

$$\sup_n |f_n(\mathbf{y})| \le M(\mathbf{y}) < \infty, \ \forall \mathbf{y} \in \mathbb{R}^t$$

and

$$f_n \text{ is equicontinuous, i.e. } \forall\ \mathbf{y}, \epsilon > 0,\ \exists\ \delta(\mathbf{y}, \epsilon), n(\mathbf{y}, \epsilon)$$

such that $\|\mathbf{y} - \mathbf{y}_1\| < \delta(\mathbf{y}, \epsilon)$ implies that $|f_n(\mathbf{y}) - f_n(\mathbf{y}_1)| < \epsilon \ \forall n \ge n(\mathbf{y}, \epsilon)$, then for any compact subset C of $\mathbb{R}^t$

$$\sup_{\mathbf{y} \in C} |f_n(\mathbf{y}) - f(\mathbf{y})| \to 0 \text{ as } n \to \infty.$$

If $\{f_n\}$ is uniformly equicontinuous, i.e. $\delta(\mathbf{y}, \epsilon)$, $n(\mathbf{y}, \epsilon)$ do not depend on $\mathbf{y}$ and $f(\mathbf{y}_n) \to 0$ whenever $\|\mathbf{y}_n\| \to \infty$ then

$$\sup_{\mathbf{y} \in \mathbb{R}^t} |f_n(\mathbf{y}) - f(\mathbf{y})| = \|f_n(\mathbf{y}) - f(\mathbf{y})\|_\infty \to 0 \text{ as } n \to \infty.$$

Claim 13. *Let $\{\mathbf{X}_n\}$ be any sequence of random variables and let $\mathbf{Y}_n = \mathbf{X}_n + \mathbf{Z}$. Let $f_n(\mathbf{y})$ represent the density of $\mathbf{Y}_n$. Then the collection of functions $\{f_n(\mathbf{y})\}$ is uniformly bounded and uniformly equicontinuous.*

Proof. The uniform bound on density is clear from Remark 13. To see the uniform equicontinuity observe that by mean value theorem

$$|f_n(\mathbf{y} + \mathbf{\Delta}) - f_n(\mathbf{y})| = |\nabla f_n(\mathbf{y}') \cdot \mathbf{\Delta}| \overset{(a)}{\le} \|\nabla f_n(\mathbf{y}')\|_1 \|\mathbf{\Delta}\|_\infty \le \|\nabla f_n(\mathbf{y}')\|_1 \|\mathbf{\Delta}\|_2$$

where (a) follows from Holder's inequality. Now the uniform bound on L_1 norm of $\nabla f_{\mathbf{Y}}(\mathbf{y})$ from Remark 13 yields the desired equicontinuity. □

Definition 1. A collection of random variables $\mathbf{X}_n$ on $\mathbb{R}^t$ is said to be *tight* if for every $\epsilon > 0$ there is a compact set $C_\epsilon \subset \mathbb{R}^t$ such that $\mathrm{P}(\mathbf{X}_n \notin C_\epsilon) \le \epsilon, \ \forall n$.

Lemma 4. *Consider a sequence of random variables $\{\mathbf{X}_n\}$ such that $\mathrm{E}(\mathbf{X}_n\mathbf{X}_n^T) \preceq K$, $\forall n$. Then the sequence is tight.*

Proof. Define $C_\epsilon = \{\mathbf{x} : \|\mathbf{x}\|_2^2 \le \frac{tr(K)}{\epsilon}\}$. By Markov's inequality $\mathrm{P}(\|\mathbf{X}_n\|^2 > \frac{tr(K)}{\epsilon}) \le \frac{\epsilon\,\mathrm{E}(\|\mathbf{X}_n\|^2)}{tr(K)} \le \epsilon$, $\forall n$. □

Theorem 4 (Prokhorov)**.** *If $\{\mathbf{X}_n\}$ is a tight sequence of random variables in $\mathbb{R}^t$ then there exists a subsequence $\{\mathbf{X}_{n_i}\}$ and a limiting probability distribution $\mathbf{X}_*$ such that $\mathbf{X}_{n_i} \overset{w}{\Rightarrow} \mathbf{X}_*$.*

Lemma 5. *Let $\mathbf{X}_n \overset{w}{\Rightarrow} \mathbf{X}_*$ and let $\mathbf{Z} \sim \mathcal{N}(0, I)$ be pairwise independent of $\{\mathbf{X}_n\}, \mathbf{X}_*$. Let $\mathbf{Y}_n = \mathbf{X}_n + \mathbf{Z}$, $\mathbf{Y}_* = \mathbf{X}_* + \mathbf{Z}$. Further let $\mathrm{E}(\mathbf{X}_n\mathbf{X}_n^T) \preceq K, \mathrm{E}(\mathbf{X}_*\mathbf{X}_*^T) \preceq K$. Let $f_n(\mathbf{y})$ denote the density of $\mathbf{Y}_n$ and $f_*(\mathbf{y})$ denote the density of $\mathbf{Y}_*$. Then*

1. $\mathbf{Y}_n \overset{w}{\Rightarrow} \mathbf{Y}$
2. $f_n(\mathbf{y}) \to f_*(\mathbf{y})$ *for all* $\mathbf{y}$
3. $h(\mathbf{Y}_n) \to h(\mathbf{Y})$.

Proof. The first part follows from pointwise convergence of characteristic functions (which is equivalent to weak convergence) since $\Phi_{\mathbf{Y}_n}(\mathbf{t}) = \Phi_{\mathbf{X}_n}(\mathbf{t})e^{-\|\mathbf{t}\|^2/2}$. The second part (a stronger claim that weak convergence) comes from Lemma 3. We have uniform equicontinuity since $\nabla f_n(\mathbf{y})$ has a uniformly bounded L_1 norm (see Remark 13). Bounded L_1 norm of $\nabla f_n(\mathbf{y})$ also implies that $f_*(\mathbf{y}_n) \to 0$ whenever $\|\mathbf{y}_n\| \to \infty$ (A reason: if a point has density $> \epsilon$ then it has a neighbourhood depending only on ϵ where the density is bigger than $\frac{\epsilon}{2}$, hence this implies that this neighbourhood has a lower bounded probability measure depending only on ϵ. This cannot happen at infinitely many points of a sequence $\mathbf{y}_n$ such that $|\mathbf{y}_n| \to \infty$). The third part comes from Theorem 5(below) in a direct manner as the densities are uniformly bounded, the second moment($\kappa = 2$) is uniformly bounded by $tr(K)$, and the pointwise convergence from the second part. □

Theorem 5 (Theorem 1 in [10])**.** *Let $\{\mathbf{Y}_i \in \mathbb{C}^t\}$ be a sequence of continuous random variables with pdf's $\{f_i\}$ and $\mathbf{Y}_*$ be a continuous random variable with pdf f_* such that $f_i \to f_*$ pointwise. Let $\|\mathbf{y}\| = \sqrt{\mathbf{y}^\dagger\mathbf{y}}$ denote the Euclidean norm of $\mathbf{y} \in \mathbb{C}^t$. If the conditions* 1) $\max\{\sup_{\mathbf{y}} f_i(\mathbf{y}), \sup_{\mathbf{y}} f_*(\mathbf{y})\} \le F$ $\forall i$ *and* 2) $\max\{\int \|\mathbf{y}\|^\kappa f_i(\mathbf{y})d\mathbf{y}, \int \|\mathbf{y}\|^\kappa f_*(\mathbf{y})d\mathbf{y}\} \le L$ *hold for some $\kappa > 1$ and for all i then $h(\mathbf{Y}_i) \to h(\mathbf{Y}_*)$.*

Remark 14. This theorem is relatively straightforward. One gets $\liminf h(\mathbf{Y}_i) \ge h(\mathbf{Y}_*)$ coming due to upper bound on densities and $\limsup h(\mathbf{Y}_i) \le h(\mathbf{Y}_*)$ due to the moment constraints. Similar kind of result can be found in Appendix 3A of [6].

We now have the tools to prove Claim 6.

Proof of Claim 6

Proof. Define

$$\mathsf{v}_\lambda(K) = \sup_{\mathbf{X}:\mathrm{E}(\mathbf{X}\mathbf{X}^T)=K} \mathsf{s}_\lambda(\mathbf{X}).$$

Let $\mathbf{X}_n$ be a sequence of random variables such that $\mathrm{E}(\mathbf{X}_n\mathbf{X}_n^T) = K$ and $\mathsf{s}_\lambda(\mathbf{X}_n) \uparrow \mathsf{v}_\lambda(K)$. By the covariance constraint (Lemma 4) we know that the sequence of random variables $\mathbf{X}_n$ forms a tight sequence and by Theorem 4 there exists X_K^* and a convergent subsequence such that $\mathbf{X}_{n_i} \overset{w}{\Rightarrow} \mathbf{X}_K^*$. From Lemma 5 we have that $h(\mathbf{Y}_{1n_i}), h(\mathbf{Y}_{2n_i}) \to h(\mathbf{Y}_{1K}^*), h(\mathbf{Y}_{2K}^*)$ and hence $\mathsf{s}_\lambda(\mathbf{X}_K^*) = \mathsf{v}_\lambda(K)$. Thus $\mathrm{V}_\lambda(K)$ can be obtained as a convex combination of $\mathsf{s}_\lambda(X_K^*)$ subject to the covariance constraint.

It takes $\frac{t(t+1)}{2}$ constraints to preserve the covariance matrix and one constraint to preserve $s_\lambda(\mathbf{X}|V)$. Hence by Bunt-Carathedory's theorem[4] we can find a pair of random variables $(V_*, \mathbf{X}_*)$ with $|V_*| \le \frac{t(t+1)}{2} + 1$ such that $\mathrm{V}_\lambda(K) = \mathsf{s}_\lambda(\mathbf{X}_*|V_*)$. □

[4]We need to use Bunt's extension[2] of Caratheodory's theorem as we no longer have compactness of the set required for the usually referred extension due to Fenchel. We can also use vanilla Caratheordory at the expense of one extra cardinality.

B.2 Continuity in a pathwise sense on concave envelopes

In this section we will establish the validity of Claim 9. For this we need more tools and results from analysis.

Claim 14. *For $\lambda > 1$, there exists C_λ such that $\mathsf{s}_\lambda(\mathbf{X}) \le C_\lambda$.*

Proof. We know from Theorem 1 that if $\mathrm{E}(\mathbf{X}\mathbf{X}^T) \preceq K$ then

$$\mathsf{s}_\lambda(\mathbf{X}) \le S_\lambda(\mathbf{X}) \le \mathrm{V}_\lambda(K) \le \mathsf{s}_\lambda(\mathbf{X}_K^*)$$

for some $\mathbf{X}_K^* \sim \mathcal{N}(0, K'), K' \preceq K$. This implies that

$$\sup_{\mathbf{X}} \mathsf{s}_\lambda(\mathbf{X}) \le \sup_{K \succeq 0: \mathbf{X} \sim \mathcal{N}(0,K)} I(\mathbf{X}; \mathbf{Y}_1) - \lambda I(\mathbf{X}; \mathbf{Y}_2).$$

Let $\Sigma_i = (G_i^T G_i)^{-1}$, $i = 1, 2$. For $\mathbf{X} \sim \mathcal{N}(0, K)$, we have

$$\begin{aligned} 2I(\mathbf{X}; \mathbf{Y}_1) - 2\lambda I(\mathbf{X}; \mathbf{Y}_2) &= \log |I + G_1 K G_1^T| - \lambda \log |I + G_2 K G_2^T| \\ &= \log |I + K G_1^T G_1| - \lambda \log |I + K G_2^T G_2| \\ &= -\log |\Sigma_1| + \lambda \log |\Sigma_2| + \log |\Sigma_1 + K| - \lambda \log |\Sigma_2 + K|. \end{aligned}$$

To bound the last two terms, we use the min-max theorem on eigenvalues: Let $\mu_j(A)$ be the j-th smallest eigenvalue of symmetric matrix $A \in \mathbb{R}^{t \times t}$, we have

$$\mu_j(A) = \min_{L_j} \max_{0 \ne u \in L_j} \frac{u^T A u}{u^T u} = \max_{L_{t+1-j}} \min_{0 \ne u \in L_{t+1-j}} \frac{u^T A u}{u^T u},$$

where L_j is a j dimensional subspace of $\mathbb{R}^t$. From this theorem we have

$$\mu_j(K) + \mu_1(\Sigma) \le \mu_j(K + \Sigma) \le \mu_j(K) + \mu_t(\Sigma), \quad j = 1, 2, \ldots, t.$$

Hence

$$\begin{aligned} \log |\Sigma_1 + K| - \lambda \log |\Sigma_2 + K| &= \sum_{j=1}^t \log \frac{\mu_j(K + \Sigma_1)}{(\mu_j(K + \Sigma_2))^\lambda} \\ &\le \sum_{j=1}^t \log \frac{\mu_j(K) + \mu_t(\Sigma_1)}{(\mu_j(K) + \mu_1(\Sigma_2))^\lambda} \\ &\le t \cdot \log \frac{\mu^* + \mu_t(\Sigma_1)}{(\mu^* + \mu_1(\Sigma_2))^\lambda}, \end{aligned}$$

where $\mu^* = \max\{0, \frac{1}{\lambda - 1}(\mu_1(\Sigma_2) - \lambda \mu_t(\Sigma_1))\}$. □

For $m \in \mathbb{N}$ the set $\mathcal{A}_m := \{\mathbf{X} : \mathrm{E}(\|\mathbf{X}\|^2) \le m\}$ is a closed subset of the topology space. This is because if $\mathbf{X}_n \overset{w}{\Rightarrow} \mathbf{X}_*$ then $\mathrm{E}(\|\mathbf{X}_*\|^2) \le \liminf_n \mathrm{E}(\|\mathbf{X}_n\|^2)$ (by definition of weak convergence and monotone convergence theorem by considering continuous and bounded functions $f_n(x) = \min\{x^2, n\}$.).

We defined $S_\lambda(\mathbf{X}) = \mathfrak{C}(\mathsf{s}_\lambda(\mathbf{X})) = \sup_{V \to \mathbf{X} \to (\mathbf{Y}_1, \mathbf{Y}_2)} \mathsf{s}_\lambda(\mathbf{X}|V)$. Taking $V = \mathbf{X}$ we observe that $S_\lambda(\mathbf{X}) \ge 0$. Define $\bar{\mathsf{s}}_\lambda(\mathbf{X}) = \max\{\mathsf{s}_\lambda(\mathbf{X}), 0\}$. Now note that $S_\lambda(\mathbf{X}) = \mathfrak{C}(\bar{\mathsf{s}}_\lambda(\mathbf{X}))$, since $S_\lambda(\mathbf{X}) \ge 0$.

Let $\bar{\mathsf{s}}_\lambda^m(\mathbf{X})$ be $\bar{\mathsf{s}}_\lambda(\mathbf{X})$ restricted to $\mathcal{A}_m$. Let $\mathsf{s}_\lambda^m(\mathbf{X})$ be the continuous extension of $\bar{\mathsf{s}}_\lambda^m(\mathbf{X})$ from $\mathcal{A}_m$ on to $\mathcal{P}$. This exists due to Tietze Extension Theorem (produced below).

Theorem 6 (Tietze Extension Theorem)**.** *Let A be a closed subset in a normal topological space, then every continuous map $f : A \to \mathbb{R}$ can be extended to a continuous map on the whole space.*

Consider a sequence $\mathbf{X}_n \in \mathcal{A}_m$ such that $\mathbf{X}_n \overset{w}{\Rightarrow} X_*$. Since the second moments are uniformly bounded, similar arguments as in Claim 6 will imply that $\bar{\mathsf{s}}_\lambda^m(\mathbf{X}_n) \to \bar{\mathsf{s}}_\lambda^m(\mathbf{X}_*)$. Further observe that the function $\mathsf{s}_\lambda^m(\mathbf{X})$ is bounded and non-negative since $\bar{\mathsf{s}}_\lambda^m(\mathbf{X})$ is bounded (above by C_λ) and non-negative.

The following result follows from a recent result in [14]. The convex hull of a function $f(\mathbf{X})$ is the lower convex envelope, or equivalently $-\mathfrak{C}(-f(\mathbf{X}))$, where $\mathfrak{C}(\cdot)$ is the upper concave envelope used in this article.

Theorem 7. *For the set of Borel probability measures on $\mathbb{R}^t$ endowed with the weak-convergence topology, the convex hull of an arbitrary bounded and continuous function is continuous.*

Proof. This theorem is obtained directly from Corollary 5 and Theorem 1 in [14]. $\square$

An immediate corollary which follows from the fact that convex hull of $f(\mathbf{X}) \equiv -\mathfrak{C}(-f(\mathbf{X}))$ is the following:

Corollary 5. *For the set of Borel probability measures on $\mathbb{R}^t$ endowed with the weak-convergence topology, the upper concave envelope of an arbitrary bounded and continuous function is continuous.*

Now define $S_\lambda^m(\mathbf{X})$ to be concave envelope of $\mathsf{s}_\lambda^m(\mathbf{X})$. From Corollary 5 we have that $S_\lambda^m(\mathbf{X})$ is continuous; Further since $\mathsf{s}_\lambda^m(\mathbf{X})$ is bounded, and non-negative, so is $S_\lambda^m(\mathbf{X})$. Continuity in particular implies that

$$\text{if } \mathbf{X}_n \overset{w}{\Rightarrow} \mathbf{X}_*, \text{ then } S_\lambda^m(\mathbf{X}_n) \to S_\lambda^m(\mathbf{X}_*). \tag{1}$$

Claim 15 (Continuity in a pathwise sense)**.** *If $\mathbf{X}_n \overset{w}{\Rightarrow} \mathbf{X}_*$ and $\mathrm{E}(\mathbf{X}_n\mathbf{X}_n^T), \mathrm{E}(\mathbf{X}_*\mathbf{X}_*^T) \preceq K$, then $S_\lambda(\mathbf{X}_n) \to S_\lambda(\mathbf{X}_*)$.*

Proof. The proof is essentially validating the interchange of limits between m, n in (1). We show a uniform convergence (in m) of $S_\lambda^m(\mathbf{X}_n) \to S_\lambda(\mathbf{X}_n)$ and this suffices to justify the interchange as follows: Given $\epsilon > 0$ choose $M_\epsilon > 0$ such that $|S_\lambda(\mathbf{X}_n) - S_\lambda^m(\mathbf{X}_n)| < \epsilon \ \forall n$ whenever $m > M_\epsilon$ (such an M_ϵ exists by uniform convergence). This implies that $\forall m > M_\epsilon$ we have

$$S_\lambda(\mathbf{X}_n) \leq S_\lambda^m(\mathbf{X}_n) + \epsilon, \overset{n\to\infty}{\Longrightarrow} \limsup_n S_\lambda(\mathbf{X}_n) \leq S_\lambda^m(\mathbf{X}_*) + \epsilon, \overset{m\to\infty}{\Longrightarrow} \limsup_n S_\lambda(\mathbf{X}_n) \leq S_\lambda(\mathbf{X}_*) + \epsilon.$$

Similarly $\forall m > M_\epsilon$

$$S_\lambda(\mathbf{X}_n) \geq S_\lambda^m(\mathbf{X}_n) - \epsilon, \overset{n\to\infty}{\Longrightarrow} \liminf_n S_\lambda(\mathbf{X}_n) \geq S_\lambda^m(\mathbf{X}_*) - \epsilon, \overset{m\to\infty}{\Longrightarrow} \liminf_n S_\lambda(\mathbf{X}_n) \geq S_\lambda(\mathbf{X}_*) - \epsilon.$$

Hence $S_\lambda(\mathbf{X}_n) \to S_\lambda(\mathbf{X}_*)$ provided we show the uniform convergence (in m) of $S_\lambda^m(\mathbf{X}_n) \to S_\lambda(\mathbf{X}_n)$. Given $\epsilon > 0$ consider a V such that $S_\lambda(\mathbf{X}_n) \leq s_\lambda(\mathbf{X}_n|V) + \frac{\epsilon}{4}$. Observe that V induces a probability measure on the space of all probability measures. We now bound the induced probability measure on distributions such that $\mathrm{E}(\|\mathbf{X}\|^2) \geq m$. Since $\mathrm{E}(\|\mathbf{X}_n\|^2) \leq tr(K)$, from Markov's inequality the mass of the induced measure on the probability measures such that $\mathrm{E}(\|\mathbf{X}\|^2) \geq m$ is at most $\frac{tr(K)}{m}$. Hence their contribution to $s_\lambda(\mathbf{X}_n|V)$ is at most $\frac{C_\lambda tr(K)}{m}$, where C_λ is the global upper bound on $\mathsf{s}_\lambda(\mathbf{X})$. Thus by taking m large enough we can make this smaller than $\frac{\epsilon}{4}$. Hence

$$S_\lambda^m(\mathbf{X}_n) \geq s_\lambda^m(\mathbf{X}_n|V) \geq s_\lambda(\mathbf{X}_n|V) - \frac{\epsilon}{4} \geq S_\lambda(\mathbf{X}_n) - \frac{\epsilon}{2}.$$

Similar argument (taking V' such that $S_\lambda^m(\mathbf{X}_n) \leq s_\lambda^m(\mathbf{X}_n|V') + \frac{\epsilon}{4}$) also shows that $S_\lambda(\mathbf{X}_n) \geq S_\lambda^m(\mathbf{X}_n) - \frac{\epsilon}{2}$. Hence for all $m > \frac{4C_\lambda tr(K)}{\epsilon}$ we have that $|S_\lambda(\mathbf{X}_n) - S_\lambda^m(\mathbf{X}_n)| \leq \epsilon$ uniformly in n as desired. $\square$

We now have the tools to prove Claim 9.

Proof of Claim 9

Proof. From Claim 15 and using similar arguments as in the proof of Claim 6 we see that $\hat{\mathrm{V}}_{\vec{\lambda}}(K)$ can be obtained as a convex combination of $\mathsf{t}_{\vec{\lambda}}(X_K^*)$ subject to the covariance constraint. It takes $\frac{t(t+1)}{2}$ constraints to preserve the covariance matrix and one constraint to preserve $\mathsf{t}_{\vec{\lambda}}(\mathbf{X}|W)$. Hence by Bunt-Carathedory's theorem we can find a pair of random variables $(W_*, \mathbf{X}_*)$ with $|W_*| \leq \frac{t(t+1)}{2} + 1$ such that $\hat{\mathrm{V}}_{\vec{\lambda}}(K) = \mathsf{t}_{\vec{\lambda}}(\mathbf{X}_*|W_*)$. $\square$

Indeed the proof technique we used carries over almost verbatim to establish this general lemma, which could be useful in other multi-terminal situations..

Lemma 6. *Consider the space of all Borel probability distributions on $\mathbb{R}^t$ endowed with the topology induced by weak convergence. If $f(\mathbf{X})$ is a bounded real-valued function with the following property, P: for any sequence $\{\mathbf{X}_n\}$ that satisfies the two properties (i) $\exists\ \kappa > 1$, s/t $\mathrm{E}(\|\mathbf{X}_n\|^\kappa) \leq B\ \forall n$ (i.e. sequence has a uniformly bounded κ-th moment) and (ii) $\mathbf{X}_n \overset{w}{\Rightarrow} \mathbf{X}_*$, we have $f(\mathbf{X}_n) \to f(\mathbf{X}_*)$; then the same properties holds for $F(\mathbf{X}) = \mathfrak{C}(f(\mathbf{X}))$, its upper concave envelope; i.e. $F(\mathbf{X})$ is bounded and satisfies P.*

Proof. The boundedness of $F(\mathbf{X})$ is immediate. To show that $F(\mathbf{X})$ satisfies property P, we use the same argument as earlier. Consider a sequence $\{\mathbf{X}_n\}$ with a uniformly bounded κ-th moment such that $\mathbf{X}_n \overset{w}{\Rightarrow} \mathbf{X}_*$. First, restrict f to $\mathcal{A}_m$ (set of all distributions whose κ-th moment is upper bounded by m) and observe that this induces a continuous (by property P of f) and bounded function (on the topology induced by weak convergence) from this closed set, $\mathcal{A}_m$, to reals. Now we extend this restricted function by the Tietze extension theorem to obtain $f^m(\mathbf{X})$, a continuous and bounded function on the whole space. Then from Corollary 5 we see that the concave envelope of $f^m(\mathbf{X})$, denoted by $F^m(\mathbf{X})$ is bounded and continuous. Finally, in a similar fashion as above, one can establish a uniform convergence (in n) of $F^m(\mathbf{X}_n) \to F(\mathbf{X}_n)$ and hence conclude that $F(\mathbf{X}_n) \to F(\mathbf{X}_*)$. ☐

Remark 15. This lemma can be used to establish the existence of the maximizing distributions in other network information theory settings, without having to repeat the arguments or the machinery we used in this paper.

C Alternate path to Theorem 1

Below, we will give an elementary proof of Theorem 1 without invoking Corollary 3.

Corollary 6. *For every $l \in \mathbb{N}, n = 2^l$, let $(V^n, \mathbf{X}_n) \sim \prod_{i=1}^n p_*(V_i, \mathbf{X}_i)$. Then $\tilde{V}, \tilde{\mathbf{X}}_n$ achieves $\mathrm{V}_\lambda(K)$ where $\tilde{V} = (V_1, V_2, .., V_n)$ and $\tilde{\mathbf{X}}_n | \big(\tilde{V} = (v_1, v_2, .., v_n)\big) \sim \frac{1}{\sqrt{n}}\left(\mathbf{X}_{v_1} + \mathbf{X}_{v_2} + \cdots + \mathbf{X}_{v_n}\right)$. We take $\mathbf{X}_{v_1}, \mathbf{X}_{v_2}, \ldots, \mathbf{X}_{v_n}$ to be independent random variables here.*

Proof. The proof follows from induction using Claim 7. ☐

Consider $(V^n, \mathbf{X}^n) \sim \prod_{i=1}^n p_*(V_i, \mathbf{X}_i)$, where $p_*(v, \mathbf{x})$ achieves $\mathrm{V}_\lambda(K)$. Let $\mathcal{V} = \{1, .., m\}$ where $m \leq \frac{t(t+1)}{2} + 1$. Now consider $(V^n, \tilde{\mathbf{X}}_n)$ where $\tilde{\mathbf{X}}_n | \big(V^n = (v_1, v_2, .., v_n)\big) \sim \frac{1}{\sqrt{n}}\left(\mathbf{X}_{v_1} + \mathbf{X}_{v_2} + \cdots + \mathbf{X}_{v_n}\right)$. Again we take $\mathbf{X}_{v_1}, \mathbf{X}_{v_2}, \ldots, \mathbf{X}_{v_n}$ to be independent random variables.

As is common in information theoretic arguments, we are going to consider typical sequences and atypical sequences. Let us define typical sequences in the following fashion:

$$\mathcal{T}^{(n)}(V) := \{v^n : \big||\{i : v_i = v\}| - np_*(v)\big| \leq n\omega_n p_*(v),\ \forall v \in [1:m].\}$$

where ω_n is any sequence such that $\omega_n \to 0$ as $n \to \infty$ and $\omega_n\sqrt{n} \to \infty$ as $n \to \infty$. For instance $\omega_n = \frac{\log n}{\sqrt{n}}$.

Note that (using Chebychev's inequality)

$$\mathrm{P}(\big||\{i : v_i = v\}| - np_*(v)\big| > n\omega_n p_*(v)) \leq \frac{1 - p_*(v)}{p_*(v)\omega_n^2 n}.$$

Hence $\mathrm{P}(v^n \notin \mathcal{T}^{(n)}(V)) \to 0$ as $n \to \infty$.

Consider any sequence of typical sequences $v^n \in \mathcal{T}^{(n)}(V)$. Consider a sequence of induced distributions $\hat{\mathbf{X}}_n \sim \tilde{\mathbf{X}}_n | v^n$.

Claim 16. $\hat{\mathbf{X}}_n \Rightarrow \mathcal{N}(0, \sum_{v=1}^m p_*(v) K_v)$

Proof. For given v^n, let $A_n(v) = |\{i : v_i = v\}|$. We know that $A_n(v) \in np_*(v)(1 \pm w_n), \forall v$. Consider a $\mathbf{c} \in \mathbb{R}^t$ with $\|\mathbf{c}\| = 1$. Let $\hat{\mathbf{X}}^{\mathbf{c}}_{n,i} \sim \frac{1}{\sqrt{n}}\mathbf{c}^T \cdot \mathbf{X}_{v_i}$ and $\hat{\mathbf{X}}^{\mathbf{c}}_{n,i}$ be independent random variables. Note that $\sum_{i=1}^n \hat{\mathbf{X}}^{\mathbf{c}}_{n,i} \sim \mathbf{c}^T \hat{\mathbf{X}}_n$.

Note that

$$\sum_{i=1}^{n} E((\hat{\mathbf{X}}^{\mathbf{c}}_{n,i})^2) = \frac{1}{n}\sum_v A_n(v)\mathbf{c}^T K_v \mathbf{c} \to \mathbf{c}^T \left(\sum_v p_*(v) K_v\right)\mathbf{c}.$$

$$\begin{aligned}\sum_{i=1}^{n} E((\hat{\mathbf{X}}^{\mathbf{c}}_{n,i})^2; |\hat{\mathbf{X}}^{c}_{n,i}| > \epsilon_1) &= \frac{1}{n}\sum_v A_n(v)\,\mathrm{E}(\mathbf{c}^T\mathbf{X}_v\mathbf{X}_v^T\mathbf{c}; \mathbf{c}^T\mathbf{X}_v\mathbf{X}_v^T\mathbf{c} \geq n\epsilon_1^2) \\ &\leq \sum_v p_*(v)(1+\omega_n)\,\mathrm{E}(\mathbf{c}^T X_v X_v^T\mathbf{c}; \mathbf{c}^T\mathbf{X}_v\mathbf{X}_v^T\mathbf{c} \geq n\epsilon_1^2) \to 0.\end{aligned}$$

In the last convergence we use that K_v's are bounded, and hence $\mathbf{c}^T\mathbf{X}_v$ has a bounded seconded moment. Hence from Lindeberg-Feller CLT[5] we have $\sum_{i=1}^n \hat{\mathbf{X}}^{\mathbf{c}}_{n,i} \Rightarrow \mathcal{N}(0, \mathbf{c}^T\sum_v p_*(v)K_v\mathbf{c})$. Hence $\hat{\mathbf{X}}_n \Rightarrow \mathcal{N}(0, \sum_v p_*(v)K_v)$ (Cramer-Wold device). □

The next claim shows a uniform convergence of the conditional laws to the Gaussian.

Claim 17. *Given any $\delta > 0$, there exists N_0 such that $\forall n > N_0$ we have for all $v^n \in \mathcal{T}^{(n)}(V)$*

$$\mathsf{s}_\lambda(\tilde{\mathbf{X}}_n | v^n) - \mathsf{s}_\lambda(\mathbf{X}^*) \leq \delta,$$

where $\mathbf{X}^* \sim \mathcal{N}(0, \sum_v p_*(v)K_v)$.

Proof. Assume not. Then we have a subsequence $v^{n_k} \in \mathcal{T}^{(n_k)}(V)$ and distributions $\tilde{\mathbf{X}}_{n_k}|v^{n_k}$ such that

$$\mathsf{s}_\lambda(\tilde{\mathbf{X}}_{n_k}|v^{n_k}) > \mathsf{s}_\lambda(X^*) + \delta, \forall k.$$

However from Claim 16 we know that $\tilde{\mathbf{X}}_{n_k}|v^{n_k} \overset{w}{\Rightarrow} X^*$ and from Lemma 5 we have $\mathsf{s}_\lambda(\tilde{\mathbf{X}}_{n_k}|v^{n_k}) \to \mathsf{s}_\lambda(X^*)$, a contradiction. □

Theorem 8. *There is a single Gaussian distribution (i.e. no mixture is required) that achieves $\mathrm{V}_\lambda(K)$.*

Proof. We know from Corollary 6 that For every $l \in \mathbb{N}, n = 2^l$, the pair $V^n, \tilde{\mathbf{X}}_n$ achieves $\mathrm{V}_\lambda(K)$. Hence

$$\mathrm{V}_\lambda(K) = \sum_{v^n} p_*(v^n)\mathsf{s}_\lambda(\tilde{\mathbf{X}}_n|v^n) = \sum_{v^n\in\mathcal{T}^{(n)}(V)} p_*(v^n)\mathsf{s}_\lambda(\tilde{\mathbf{X}}_n|v^n) + \sum_{v^n\notin\mathcal{T}^{(n)}(V)} p_*(v^n)\mathsf{s}_\lambda(\tilde{\mathbf{X}}_n|v^n).$$

For a given v^n, let $\hat{\mathbf{X}} \sim \mathbf{X}_n|v^n$. Then note that $\mathrm{E}(\hat{\mathbf{X}}\hat{\mathbf{X}}^T) \preceq \sum_{v=1}^m K_v$. Thus $\mathsf{s}_\lambda(\hat{\mathbf{X}}) \leq I(\hat{\mathbf{X}}; \mathbf{Y}_1) \leq C$ for some fixed constant C that is independent of v^n. Thus using Claim 17 we can upper bound $\mathrm{V}_\lambda(K)$ for large n by

$$\begin{aligned}\mathrm{V}_\lambda(K) &= \sum_{v^n\in\mathcal{T}^{(n)}(V)} p_*(v^n)\mathsf{s}_\lambda(\tilde{\mathbf{X}}_n|v^n) + \sum_{v^n\notin\mathcal{T}^{(n)}(V)} p_*(v^n)\mathsf{s}_\lambda(\tilde{\mathbf{X}}_n|v^n) \\ &\leq \sum_{v^n\in\mathcal{T}^{(n)}(V)} p_*(v^n)(\mathsf{s}_\lambda(\mathbf{X}^*)+\delta) + C\sum_{v^n\notin\mathcal{T}^{(n)}(V)} p_*(v^n) \\ &= \mathrm{P}(v^n \in \mathcal{T}^{(n)})(\mathsf{s}_\lambda(\mathbf{X}^*)+\delta) + C\,\mathrm{P}(v^n \notin \mathcal{T}^{(n)}).\end{aligned}$$

Here $\mathbf{X}^* \sim \mathcal{N}(0, \sum_v p_*(v)K_v)$. Since $\mathrm{P}(v^n \in \mathcal{T}^{(n)}) \to 1$ as $n \to \infty$ we get $\mathrm{V}_\lambda(K) \leq \mathsf{s}_\lambda(\mathbf{X}^*) + \delta$; but $\delta > 0$ is arbitrary, hence $\mathrm{V}_\lambda(K) \leq \mathsf{s}_\lambda(\mathbf{X}^*)$. The other direction $\mathrm{V}_\lambda(K) \leq \mathsf{s}_\lambda(\mathbf{X}^*)$ is trivial from the definition of $\mathrm{V}_\lambda(K)$ and the fact that $\sum_v p_*(v)K_v \preceq K$. □

[5] We adopt the notation in Theorem (4.5), Chapter 2 in [5].