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

Yanlin Geng, *Member, IEEE,* and Chandra Nair, *Member, IEEE*

***Abstract*—A novel method for establishing the optimality of Gaussian auxiliary random variables in multiterminal information theory problems is developed. This method is then employed to show that Marton's inner bound achieves the capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages.**

***Index Terms*—multiuser channels, channel capacity, Gaussian distribution optimality**

I. Introduction

Channels with additive Gaussian noise are a commonly used model in wireless communications. Computing the capacity regions or bounds on the capacity regions for these classes of channels is of wide interest. Bounds on capacity regions or capacity regions themselves are oftentimes represented using auxiliary random variables and evaluations of these bounds (or regions) reduce to optimization problems and computation of the "extremal" auxiliary random variables. In several instances involving Gaussian noise, it turns out that the extremal (optimal) auxiliaries are Gaussian random variables. However proving the optimality of Gaussian distributions is usually cumbersome; almost always invoking the entropy-power-inequality (EPI), or some recent variations using only some elements from its proof.

In the following sections we develop a novel way of proving the optimality of Gaussian input distributions 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. As an illustration of this technique we will recover some existing results and then demonstrate its effectiveness by obtaining the capacity region of a well-studied problem that had resisted solution using traditional techniques: the capacity region of the vector Gaussian channel with both private and common messages.

For the two-receiver Gaussian vector broadcast channel with private messages, the capacity region was established [1] by showing that a certain pair of inner and outer bounds yield identical regions. This argument, though effective, could not be generalized to compute the capacity region of the vector Gaussian channel with both private and common messages.

Y. Geng and C. Nair are with the Department of Information Engineering, The Chinese University of Hong Kong.

The paper was in part presented at the IEEE International Symposium on Information Theory, 2012.

A. Definitions

Broadcast channel [2] 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 the messages (M_0, M_1) and receiver Y_2 to recover the 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 [3].

A *broadcast channel* is characterized by a probability transition matrix $\mathrm{q}(y_1, y_2|x)$. The following broadcast channel is referred to as the Gaussian vector 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 $\mathbf{X}, \mathbf{Z}_1, \mathbf{Z}_2 \in \mathbb{R}^t$ are mutually independent[1] random vectors, $G_1, G_2 \in \mathbb{R}^{t\times t}$ are channel gain matrices, and noises $\mathbf{Z}_1, \mathbf{Z}_2$ are Gaussian distributed random vectors.

Remark 1. We make the following assumptions regarding channel gain matrices and noise covariances.

1) We will assume that all our channel gain matrices are invertible. The reason for this assumption is the following: we are working with inner and outer bounds to capacity regions represented in terms of mutual information between the channel inputs (or auxiliary random variables) and the channel outputs. The mutual information terms, and hence the inner and outer bounds, are continuous functions of the channel gain matrices. We establish capacity regions by showing that the inner and outer bounds coincide. Since the set of all invertible matrices form a dense set, by continuity of the bounds, the inner and outer bounds will coincide for all channel gain matrices.
2) We also assume that all the Gaussian noise vectors are $\mathcal{N}(0, I)$ for the following reason. The mean of the Gaussian noise does not affect the capacity region. If the covariance matrix is invertible, then one can multiply by another invertible matrix to transform the covariance matrix to the identity matrix. On the other hand, if the

[1]One can relax the mutual independence to the following assumption: $\mathbf{X}$ is independent of $\mathbf{Z}_1$ and $\mathbf{X}$ is independent of $\mathbf{Z}_2$. Since the capacity region of a broadcast channel depends only on the marginal distributions $\mathrm{q}_1(\mathbf{y}_1|\mathbf{x})$ and $\mathrm{q}_2(\mathbf{y}_2|\mathbf{x})$ (see [3]) the assumption of mutual independence is not restrictive for the purpose of characterizing the capacity region.

covariance matrix is non-invertible, then by a suitable linear transformation, one can get a noise-less channel which has infinite capacity; an uninteresting scenario.

A *product broadcast channel*, consisting of a sender $(\mathbf{X}_1, \mathbf{X}_2)$ and two receivers $(\mathbf{Y}_{11}, \mathbf{Y}_{12})$ and $(\mathbf{Y}_{21}, \mathbf{Y}_{22})$, is a broadcast channel whose transition probability has the form $\mathrm{q}_1(\mathbf{y}_{11}, \mathbf{y}_{21}|\mathbf{x}_1) \times \mathrm{q}_2(\mathbf{y}_{12}, \mathbf{y}_{22}|\mathbf{x}_2)$. A Gaussian vector 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 $\mathbf{Z}_{11}, \mathbf{Z}_{12}, \mathbf{Z}_{21}, \mathbf{Z}_{22} \sim \mathcal{N}(0, I)$ are i.i.d.; and they are also independent of $(\mathbf{X}_1, \mathbf{X}_2)$.

A *two-letter* version of a (broadcast) channel is a product (broadcast) channel where the components are identical, i.e. $\mathrm{q}_1(\cdot|\cdot) = \mathrm{q}_2(\cdot|\cdot)$.

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 ideas employed in this paper: (i) the use of a two-letter channel to identify that the optimizing distributions are Gaussians, (ii) the factorization of concave envelopes that relates a function on a product channel to its counterparts in the component channels, and (iii) the use of a min-max interchange to deal with a linearized expression. The second and third ideas had been partly developed in the context of discrete memoryless broadcast channels [4].

We present our first idea using the well studied problem of maximizing mutual information in Section II-A. We then present the second idea in Section II-B, the results of which will be used to give an alternate proof of the capacity region for the private messages case. The arguments are then generalized in Section II-C, and the results there will be used to determine the capacity region of the case with private and common messages. The capacity regions will be established in Section III and here we will also incorporate the min-max idea that was alluded to earlier.

Since we are working with continuous alphabets, our approach involves some mathematical technicalities that need to be taken care of; we defer these arguments to the Appendices. These arguments in the Appendices are stated in a general manner so as to enable future applications of these ideas by invoking the results directly. We also illustrate an adaptation of our technique to the vector Gaussian wiretap setting in Appendix III.

B. A couple of mathematical preliminaries

We present some elementary results regarding additive Gaussian channels which will be useful later.

Proposition 1. *Consider the following two-letter Gaussian product channel*

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

where $\mathbf{Z}_1$ *and* $\mathbf{Z}_2$ *are independent and distributed as* $\mathcal{N}(0, I)$. *Define*

$$\mathbf{X}_{\theta_1} = \frac{1}{\sqrt{2}}(\mathbf{X}_1 + \mathbf{X}_2), \quad \mathbf{X}_{\theta_2} = \frac{1}{\sqrt{2}}(\mathbf{X}_1 - \mathbf{X}_2),$$
$$\mathbf{Y}_{\theta_1} = \frac{1}{\sqrt{2}}(\mathbf{Y}_1 + \mathbf{Y}_2), \quad \mathbf{Y}_{\theta_2} = \frac{1}{\sqrt{2}}(\mathbf{Y}_1 - \mathbf{Y}_2).$$

Then $I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2) = I(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}; \mathbf{Y}_{\theta_1}, \mathbf{Y}_{\theta_2})$.

Proof. Since the linear transformations involved here amount to multiplication by a unitary matrix, we have $h(\mathbf{Y}_{\theta_1}, \mathbf{Y}_{\theta_2}) = h(\mathbf{Y}_1, \mathbf{Y}_2)$ and $h(\mathbf{Y}_{\theta_1}, \mathbf{Y}_{\theta_2}|\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}) = h(\mathbf{Z}_{\theta_1}, \mathbf{Z}_{\theta_2}) = h(\mathbf{Z}_1, \mathbf{Z}_2) = h(\mathbf{Y}_1, \mathbf{Y}_2|\mathbf{X}_1, \mathbf{X}_2)$ where $\mathbf{Z}_{\theta_1} = \frac{1}{\sqrt{2}}(\mathbf{Z}_1 + \mathbf{Z}_2)$, $\mathbf{Z}_{\theta_2} = \frac{1}{\sqrt{2}}(\mathbf{Z}_1 - \mathbf{Z}_2)$. An alternate proof is to observe that mutual information is preserved under bijective transformations. □

Remark 2. An interesting consequence of additive noise having a Gaussian distribution is that $\mathbf{Z}_{\theta_1}$ and $\mathbf{Z}_{\theta_2}$ are again independent and distributed according to $\mathcal{N}(0, I)$. Hence $(\mathbf{Y}_{\theta_1}, \mathbf{Y}_{\theta_2})$ can be regarded as the output of the *same product channel* when the input is distributed according to $(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2})$. This observation is peculiar to additive Gaussian noise channels.

Proposition 2. *In Gaussian vector 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

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

The first equality uses the independence between $\mathbf{Y}_{11}$ and $\mathbf{Y}_{22}$; the second equality uses the independence between $\mathbf{Z}_{11}$ and $\mathbf{X}_1$, and the independence between $\mathbf{Z}_{22}$ and $\mathbf{X}_2$.

On the other hand, since $\mathbf{Z}_{11}, \mathbf{Z}_{22}$ are mutually independent and independent of the pair $\mathbf{X}_1, \mathbf{X}_2$ we have

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

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

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

Hence, by the uniqueness of the characteristic functions, $G_{11}\mathbf{X}_1$ and $G_{22}\mathbf{X}_2$ are independent; and since G_{11} and G_{22} are invertible, $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent. □

II. Optimality of Gaussian via factorization of concave envelopes

We devise a new technique to show that Gaussian distributions achieve the maximum value of an optimization problem, subject to a covariance constraint. The technique developed here allows us to obtain new results as well as greatly simplify the proofs of existing results. Loosely speaking, we develop a machinery that can map the traditional single letterization arguments into proofs of optimality of Gaussian distributions.

The main idea behind the approach is to show that if a certain $\mathbf{X}$ (centered to have zero mean) achieves the maximum value of an optimization problem, then so does $\frac{1}{\sqrt{2}}(\mathbf{X}_1+\mathbf{X}_2)$ and $\frac{1}{\sqrt{2}}(\mathbf{X}_1-\mathbf{X}_2)$; where $\mathbf{X}_1, \mathbf{X}_2$ are two i.i.d. copies of $\mathbf{X}$. To show this, we go to the two-letter version of the channel, use a *factorization property* of the function involved (this is related to the traditional single letterization arguments), and then use Proposition 1 to move from the pair $\mathbf{X}_1, \mathbf{X}_2$ to $\frac{1}{\sqrt{2}}(\mathbf{X}_1 + \mathbf{X}_2)$. Further we will show that $\frac{1}{\sqrt{2}}(\mathbf{X}_1+\mathbf{X}_2)$ and $\frac{1}{\sqrt{2}}(\mathbf{X}_1 - \mathbf{X}_2)$ have to be independent as well, which forces the initial distribution to be Gaussian (see Theorem 3 and Corollary 3 in Appendix I-A). Alternately, one can repeat the averaging procedure inductively and use the Central Limit Theorem to conclude that a Gaussian distribution achieves the maximum.

Remark 3. In all the optimization problems considered in this paper, we assume that the maximizers are centered to have zero-mean. This zero-mean assumption is a consequence of mutual information being unchanged by centering. Since we employ an upper bound on the input covariance matrix note that centering only decreases $\mathrm{E}(\mathbf{X}\mathbf{X}^T)$ and thus the centered variables remain feasible and do not change the objective function value.

Remark 4. The remarkable similarity of the structure of the arguments that follow for the three optimization problems considered in this section for which we show the optimality of Gaussian distributions is worth noting. In particular the first example, though trivial, contains some of the key elements.

A. Example 1: Mutual information

Consider an additive Gaussian noise channel $\mathrm{q}(\mathbf{y}|\mathbf{x})$ given by $\mathbf{Y} = G\mathbf{X} + \mathbf{Z}$, where $G \in \mathbb{R}^{t\times t}$ is invertible and $\mathbf{Z} \sim \mathcal{N}(0, I)$ is independent of $\mathbf{X}$. Given a positive semi-definite matrix $K \succeq 0$, consider the following optimization problem:

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

Consider a product channel $\mathrm{q}_1(\mathbf{y}_1|\mathbf{x}_1) \times \mathrm{q}_2(\mathbf{y}_2|\mathbf{x}_2)$. The inequality in the proposition below may be called the *factorization property* of mutual information.

Proposition 3. *The following inequality holds for product channels*

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

Further, for a product Gaussian noise channel, if a pair of random variables $(\mathbf{X}_{1*}, \mathbf{X}_{2*})$ achieves equality above then $\mathbf{X}_{1*}$ and $\mathbf{X}_{2*}$ must be independent.

Proof. Observe that

$$\begin{aligned} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2) &= h(\mathbf{Y}_1, \mathbf{Y}_2) - h(\mathbf{Y}_1, \mathbf{Y}_2|\mathbf{X}_1, \mathbf{X}_2) \\ &\overset{(a)}{=} h(\mathbf{Y}_1, \mathbf{Y}_2) - h(\mathbf{Y}_1|\mathbf{X}_1) - h(\mathbf{Y}_2|\mathbf{X}_2) \\ &= I(\mathbf{X}_1;\mathbf{Y}_1) + I(\mathbf{X}_2;\mathbf{Y}_2) - I(\mathbf{Y}_1;\mathbf{Y}_2) \end{aligned}$$

where equality (a) is true since the channel has a product form. Further, if equality holds then $\mathbf{Y}_{1*}$ and $\mathbf{Y}_{2*}$ must be independent, which from Proposition 2 implies that $\mathbf{X}_{1*}$ and $\mathbf{X}_{2*}$ are independent. □

Proposition 4. *Let $p_*(\mathbf{x})$ be a zero mean distribution that attains*[2] *$\mathsf{V}^{\mathrm{q}}(K)$ and let $(\mathbf{X}_1, \mathbf{X}_2) \sim p_*(\mathbf{x}_1)p_*(\mathbf{x}_2)$. Then the following random variables $\mathbf{X}_{\theta_1} = \frac{1}{\sqrt{2}}(\mathbf{X}_1 + \mathbf{X}_2)$, $\mathbf{X}_{\theta_2} = \frac{1}{\sqrt{2}}(\mathbf{X}_1 - \mathbf{X}_2)$ are independent and also attain $\mathsf{V}^{\mathrm{q}}(K)$.*

Proof. Let $\mathbf{Y}_{\theta_1} = \frac{1}{\sqrt{2}}(\mathbf{Y}_1 + \mathbf{Y}_2)$, $\mathbf{Y}_{\theta_2} = \frac{1}{\sqrt{2}}(\mathbf{Y}_1 - \mathbf{Y}_2)$. Consider the two-letter product channel $\mathrm{q}(\mathbf{y}_1|\mathbf{x}_1) \times \mathrm{q}(\mathbf{y}_2|\mathbf{x}_2)$ and observe that

$$\begin{aligned} 2\mathsf{V}^{\mathrm{q}}(K) &\overset{(a)}{=} I(\mathbf{X}_1;\mathbf{Y}_1) + I(\mathbf{X}_2;\mathbf{Y}_2) \\ &\overset{(b)}{=} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2) \\ &\overset{(c)}{=} I(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}; \mathbf{Y}_{\theta_1}, \mathbf{Y}_{\theta_2}) \\ &\overset{(d)}{\le} I(\mathbf{X}_{\theta_1};\mathbf{Y}_{\theta_1}) + I(\mathbf{X}_{\theta_2};\mathbf{Y}_{\theta_2}) \\ &\overset{(e)}{\le} \mathsf{V}^{\mathrm{q}}(K) + \mathsf{V}^{\mathrm{q}}(K) = 2\mathsf{V}^{\mathrm{q}}(K). \end{aligned}$$

Here (a) holds since $p_*(\mathbf{x})$ achieves $\mathsf{V}^{\mathrm{q}}(K)$; (b) holds since $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent and the channel has a product form; (c) is a consequence of Proposition 1; and (d) follows from Proposition 3. Finally (e) follows from the definition of $\mathsf{V}^{\mathrm{q}}(K)$ since the channels $p_{\mathbf{Y}_{\theta_1}|\mathbf{X}_{\theta_1}}$ and $p_{\mathbf{Y}_{\theta_2}|\mathbf{X}_{\theta_2}}$ are the same as $p_{\mathbf{Y}|\mathbf{X}}$ (Remark (2)) and

$$\mathrm{E}(\mathbf{X}_{\theta_1}\mathbf{X}_{\theta_1}^T) = \mathrm{E}(\mathbf{X}_{\theta_2}\mathbf{X}_{\theta_2}^T) = \frac{1}{2}\left(\mathrm{E}(\mathbf{X}_1\mathbf{X}_1^T) + \mathrm{E}(\mathbf{X}_2\mathbf{X}_2^T)\right) \preceq K.$$

Since the extremes match, all inequalities must be equalities. Hence (d) must be an equality, which implies from Proposition 3 that $\mathbf{X}_{\theta_1}$ and $\mathbf{X}_{\theta_2}$ are independent. The equality (e) implies $I(\mathbf{X}_{\theta_1};\mathbf{Y}_{\theta_1}) = I(\mathbf{X}_{\theta_2};\mathbf{Y}_{\theta_2}) = \mathsf{V}^{\mathrm{q}}(K)$ as desired. □

From the above propositions we get an alternate proof of this well known result:

Proposition 5. *$\mathsf{V}^{\mathrm{q}}(K)$ is attained when (and only when) the input $\mathbf{X}$ is distributed as $\mathcal{N}(0, K)$.*

Proof. Using Proposition 4 we have shown that any zero mean maximizer $\mathbf{X} \sim p_*(\mathbf{x})$ that attains $\mathsf{V}^{\mathrm{q}}(K)$ has the following property: If $\mathbf{X}_1$ and $\mathbf{X}_2$ 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 I-A) we have that $\mathbf{X} \sim \mathcal{N}(0, K_*)$ for some $K_* \preceq K$. Using the monotonicity of the $\log|\cdot|$ function we deduce that $K_* = K$, thus establishing the uniqueness of the maximizer.

Alternately, one could also use the following approach: For any zero mean maximizer $\mathbf{X}$, Proposition 4 implies that the

[2]The proof of the existence of a maximizer can be inferred from Proposition 17, Theorem 4, and Proposition 18 in Appendix II-A.

corresponding $\frac{1}{\sqrt{2}}(\mathbf{X}_1+\mathbf{X}_2)$ also achieves the maximum. Proceeding by induction, we can use the Central Limit Theorem to deduce that a Gaussian distribution is also a maximizer. (In this regard see the arguments in Appendix IV.) There is a subtle difference between the arguments however; the former one ensures the uniqueness of the maximizer to be Gaussian while the latter one only yields that Gaussian is a maximizer. □

Remark 5. In the examples that follow we do not have any such monotonicity. Hence, the techniques we introduce will only establish that the optimizing distributions are Gaussian, which is sufficient for establishing a computable characterization of the capacity region. Additional properties of the maximizer may be inferred using standard optimization techniques.

B. Example 2: Difference of mutual informations

Consider a broadcast channel $\mathrm{q}(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x})$. For $\lambda > 1$ let the following function of $p(\mathbf{x})$ be defined by

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

For $(V,\mathbf{X})$ such that $V \to \mathbf{X} \to (\mathbf{Y}_1,\mathbf{Y}_2)$ is a Markov chain, let $\mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X}|V) := I(\mathbf{X};\mathbf{Y}_1|V) - \lambda I(\mathbf{X};\mathbf{Y}_2|V)$.

Further define the *upper concave envelope*[3] of $\mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X})$ as

$$\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}) := \mathfrak{C}(\mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X})).$$

It is a straightforward exercise to see that

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

We define $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}|V) := \sum_v p(v)\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}|V=v)$ for V (with a finite alphabet) and its natural extension for an arbitrary V.

Remark 6. Since $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X})$ is concave in $p(\mathbf{x})$ we have $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}|V) \le S^{\mathrm{q}}_\lambda(\mathbf{X})$ by Jensen's inequality. One may also note that if $W \to V \to \mathbf{X}$ is Markov, then $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}|W,V) = \mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}|V)$ because $p(\mathbf{x}|w,v) = p(\mathbf{x}|v)$.

For a product broadcast channel $\mathrm{q}_1(\mathbf{y}_{11},\mathbf{y}_{21}|\mathbf{x}_1) \times \mathrm{q}_2(\mathbf{y}_{12},\mathbf{y}_{22}|\mathbf{x}_2)$ we define, in a similar fashion as above,

$$\begin{aligned}\mathsf{s}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_1,\mathbf{X}_2) &:= I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{11},\mathbf{Y}_{12})\\ &\quad - \lambda I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{21},\mathbf{Y}_{22}).\end{aligned}$$

We also define the quantities $\mathsf{s}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_1,\mathbf{X}_2|V)$, $\mathsf{S}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_1,\mathbf{X}_2)$ and $\mathsf{S}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_1,\mathbf{X}_2|V)$ similarly. The inequality in the following proposition is referred to as the *factorization* of $\mathsf{S}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_1,\mathbf{X}_2)$.

[3]The upper concave envelope of a function $f(x)$ is the smallest concave function $g(x)$ such that $g(x) \ge f(x), \forall x$. In particular $g(x)$ can be expressed as $g(x) = \sup_{p(x):\mathrm{E}(X)=x} \mathrm{E}(f(X))$.

Proposition 6. *The following inequalities holds for product broadcast channels*

$$\begin{aligned}\mathsf{S}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_1,\mathbf{X}_2) &\le \mathsf{S}^{\mathrm{q}_1}_\lambda(\mathbf{X}_1|\mathbf{Y}_{22}) + \mathsf{S}^{\mathrm{q}_2}_\lambda(\mathbf{X}_2|\mathbf{Y}_{11})\\ &\le \mathsf{S}^{\mathrm{q}_1}_\lambda(\mathbf{X}_1) + \mathsf{S}^{\mathrm{q}_2}_\lambda(\mathbf{X}_2).\end{aligned}$$

Further, for a Gaussian product broadcast channel, if a particular triple $(V_*,\mathbf{X}_{1*},\mathbf{X}_{2*})$ satisfies

$$\begin{aligned}\mathsf{s}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_{1*},\mathbf{X}_{2*}|V_*) &= \mathsf{S}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_{1*},\mathbf{X}_{2*})\\ &= \mathsf{S}^{\mathrm{q}_1}_\lambda(\mathbf{X}_{1*}) + \mathsf{S}^{\mathrm{q}_2}_\lambda(\mathbf{X}_{2*}),\end{aligned}$$

then all of the following must be true:

1) $\mathbf{X}_{1*}$ *and* $\mathbf{X}_{2*}$ *are conditionally independent given* V_*,
2) $\mathsf{S}^{\mathrm{q}_1}_\lambda(\mathbf{X}_{1*}) = \mathsf{s}^{\mathrm{q}_1}_\lambda(\mathbf{X}_{1*}|V_*)$,
3) $\mathsf{S}^{\mathrm{q}_2}_\lambda(\mathbf{X}_{2*}) = \mathsf{s}^{\mathrm{q}_2}_\lambda(\mathbf{X}_{2*}|V_*)$.

Proof. For any $(V,\mathbf{X}_1,\mathbf{X}_2)$ such that $V \to (\mathbf{X}_1,\mathbf{X}_2) \to (\mathbf{Y}_{11},\mathbf{Y}_{12},\mathbf{Y}_{21},\mathbf{Y}_{22})$ is Markov, observe

$$\begin{aligned}&\mathsf{s}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_1,\mathbf{X}_2|V)\\
&\quad= I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{11},\mathbf{Y}_{12}|V)\\
&\qquad - \lambda I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{21},\mathbf{Y}_{22}|V)\\
&\quad= I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{11}|V) + I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{12}|V,\mathbf{Y}_{11})\\
&\qquad - \lambda I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{22}|V) - \lambda I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{21}|V,\mathbf{Y}_{22})\\
&\quad\overset{(a)}{=} I(\mathbf{X}_1;\mathbf{Y}_{11}|V) + I(\mathbf{X}_2;\mathbf{Y}_{12}|V,\mathbf{Y}_{11})\\
&\qquad - \lambda I(\mathbf{X}_2;\mathbf{Y}_{22}|V) - \lambda I(\mathbf{X}_1;\mathbf{Y}_{21}|V,\mathbf{Y}_{22})\\
&\quad\overset{(b)}{=} I(\mathbf{X}_1;\mathbf{Y}_{11}|V,\mathbf{Y}_{22}) + I(\mathbf{X}_2;\mathbf{Y}_{12}|V,\mathbf{Y}_{11})\\
&\qquad - \lambda I(\mathbf{X}_2;\mathbf{Y}_{22}|V,\mathbf{Y}_{11}) - \lambda I(\mathbf{X}_1;\mathbf{Y}_{21}|V,\mathbf{Y}_{22})\\
&\qquad - (\lambda-1) I(\mathbf{Y}_{11};\mathbf{Y}_{22}|V)\\
&\quad\overset{(c)}{\le} \mathsf{S}^{\mathrm{q}_1}_\lambda(\mathbf{X}_1|\mathbf{Y}_{22}) + \mathsf{S}^{\mathrm{q}_2}_\lambda(\mathbf{X}_2|\mathbf{Y}_{11}) - (\lambda-1) I(\mathbf{Y}_{11};\mathbf{Y}_{22}|V)\\
&\quad\overset{(d)}{\le} \mathsf{S}^{\mathrm{q}_1}_\lambda(\mathbf{X}_1) + \mathsf{S}^{\mathrm{q}_2}_\lambda(\mathbf{X}_2) - (\lambda-1) I(\mathbf{Y}_{11};\mathbf{Y}_{22}|V)\\
&\quad\overset{(e)}{\le} \mathsf{S}^{\mathrm{q}_1}_\lambda(\mathbf{X}_1) + \mathsf{S}^{\mathrm{q}_2}_\lambda(\mathbf{X}_2).\end{aligned}$$

Here (a) and (b) hold since given V we have the Markov chain $(\mathbf{Y}_{11},\mathbf{Y}_{21}) \to \mathbf{X}_1 \to \mathbf{X}_2 \to (\mathbf{Y}_{12},\mathbf{Y}_{22})$ for the product broadcast channel; (c) follows from the definition of $\mathsf{S}^{\mathrm{q}}_\lambda(\cdot|\cdot)$ and the Markov chains $(V,\mathbf{Y}_{22}) \to \mathbf{X}_1 \to (\mathbf{Y}_{11},\mathbf{Y}_{21})$ and $(V,\mathbf{Y}_{11}) \to \mathbf{X}_2 \to (\mathbf{Y}_{12},\mathbf{Y}_{22})$; (d) holds since $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X})$ is concave in $p(\mathbf{x})$; finally $\lambda > 1$ implies (e). Thus, by noticing (c) and (e) above, we have

$$\begin{aligned}&\sup_{\substack{p(v|\mathbf{x}_1,\mathbf{x}_2):\\ V\to(\mathbf{X}_1,\mathbf{X}_2)\to(\mathbf{Y}_1,\mathbf{Y}_2)}} \mathsf{s}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_1,\mathbf{X}_2|V)\\
&\qquad= \mathsf{S}^{\mathrm{q}_1\times\mathrm{q}_2}_\lambda(\mathbf{X}_1,\mathbf{X}_2)\\
&\qquad\le \mathsf{S}^{\mathrm{q}_1}_\lambda(\mathbf{X}_1|\mathbf{Y}_{22}) + \mathsf{S}^{\mathrm{q}_2}_\lambda(\mathbf{X}_2|\mathbf{Y}_{11})\\
&\qquad\le \mathsf{S}^{\mathrm{q}_1}_\lambda(\mathbf{X}_1) + \mathsf{S}^{\mathrm{q}_2}_\lambda(\mathbf{X}_2).\end{aligned}$$

In a Gaussian product broadcast channel if $(V_*,\mathbf{X}_{1*},\mathbf{X}_{2*})$ satisfies the given equality condition, then inequalities $(c),(d)$ and (e) are tight. Since $\lambda > 1$ we must have $I(\mathbf{Y}_{11*};\mathbf{Y}_{22*}|V_*) = 0$, i.e. $\mathbf{Y}_{11*}$ and $\mathbf{Y}_{22*}$ are conditionally independent given V_*. By Proposition 2, $\mathbf{X}_{1*}$ and $\mathbf{X}_{2*}$ are conditionally independent given V_*. Hence, using (c), (d) and (e) we obtain

$$\begin{aligned}\mathsf{S}^{\mathrm{q}_1}_\lambda(\mathbf{X}_{1*}) &= I(\mathbf{X}_{1*};\mathbf{Y}_{11*}|V_*,\mathbf{Y}_{22*}) - \lambda I(\mathbf{X}_{1*};\mathbf{Y}_{21*}|V_*,\mathbf{Y}_{22*})\\ &= I(\mathbf{X}_{1*};\mathbf{Y}_{11*}|V_*) - \lambda I(\mathbf{X}_{1*};\mathbf{Y}_{21*}|V_*),\\
\mathsf{S}^{\mathrm{q}_2}_\lambda(\mathbf{X}_{2*}) &= I(\mathbf{X}_{2*};\mathbf{Y}_{12*}|V_*,\mathbf{Y}_{11*}) - \lambda I(\mathbf{X}_{2*};\mathbf{Y}_{22*}|V_*,\mathbf{Y}_{11*})\\ &= I(\mathbf{X}_{2*};\mathbf{Y}_{12*}|V_*) - \lambda I(\mathbf{X}_{2*};\mathbf{Y}_{22*}|V_*).\end{aligned}$$

This completes the proof. □

1) Maximizing the concave envelope subject to a covariance constraint: Consider a Gaussian vector broadcast channel $\mathrm{q}(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x})$. For $K \succeq 0$, define

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

From the definition of $\mathsf{S}^{\mathrm{q}}_{\lambda}(\mathbf{X})$ it is clear that

$$\mathsf{V}^{\mathrm{q}}_{\lambda}(K) = \sup_{\substack{(V,\mathbf{X}):\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K \\ V\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)}} \mathsf{s}^{\mathrm{q}}_{\lambda}(\mathbf{X}|V).$$

Proposition 7. *There is a pair of random variables $(V_*, \mathbf{X}_*)$ with $|\mathcal{V}_*| \le \frac{t(t+1)}{2} + 1$ and $\mathrm{E}(\mathbf{X}_*\mathbf{X}_*^T) \preceq K$ such that*

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

Further, we can assume that the conditional law of $\mathbf{X}_*|(V_* = v_*)$ has zero mean for every v_*.

Proof. The existence of a maximizer and the cardinality bound on V_* is established in Appendix II-A. The centering conditioned on each $V_* = v_*$ does not change the mutual information terms and hence $\mathsf{s}^{\mathrm{q}}_{\lambda}(\mathbf{X}_*|V_*)$ remains unchanged. Note that the centering continues to satisfy the covariance constraint. □

The goal of this section is to show that a single Gaussian distribution achieves $\mathsf{V}^{\mathrm{q}}_{\lambda}(K)$, i.e. we can take V_* to be trivial and $\mathbf{X}_* \sim \mathcal{N}(0, K_*)$, $K_* \preceq K$.

Remark 7. This result is known and was first shown by Liu and Viswanath [5] using perturbation based techniques. We re-derive the result here to illustrate our technique and then our final result, a more involved example, in the next section is new.

Proposition 8. *Let $(V_*, \mathbf{X}_*) \sim p_*(v, \mathbf{x})$ attain $\mathsf{V}^{\mathrm{q}}_{\lambda}(K)$, with $|\mathcal{V}| = m \le \frac{t(t+1)}{2} + 1$; and let $\mathbf{X}_v$ denote a centered random variable (zero mean) distributed according to the conditional distribution $p_*(\mathbf{x}|V = v)$. 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 $p_*(v, \mathbf{x})$. Define*

$$\tilde{V} = (V_1, V_2), \quad \mathbf{X}_{\theta_1}|\big(\tilde{V} = (v_1, v_2)\big) \sim \frac{1}{\sqrt{2}}\left(\mathbf{X}_{v_1} + \mathbf{X}_{v_2}\right),$$
$$\mathbf{X}_{\theta_2}|\big(\tilde{V} = (v_1, v_2)\big) \sim \frac{1}{\sqrt{2}}\left(\mathbf{X}_{v_1} - \mathbf{X}_{v_2}\right).$$

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

1) *$\mathbf{X}_{\theta_1}$ and $\mathbf{X}_{\theta_2}$ are conditionally independent given $\tilde{V}$,*
2) *$\mathsf{V}^{\mathrm{q}}_{\lambda}(K) = \mathsf{s}^{\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_1}|\tilde{V})$,*
3) *$\mathsf{V}^{\mathrm{q}}_{\lambda}(K) = \mathsf{s}^{\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_2}|\tilde{V})$.*

Proof. Let $K_v := \mathrm{E}(\mathbf{X}_v\mathbf{X}_v^T)$. Consider the two-letter broadcast channel $\mathrm{q}(\mathbf{y}_{11}, \mathbf{y}_{21}|\mathbf{x}_1) \times \mathrm{q}(\mathbf{y}_{12}, \mathbf{y}_{22}|\mathbf{x}_2)$. We have

$$\begin{aligned}
2\mathsf{V}^{\mathrm{q}}_{\lambda}(K) &\overset{(a)}{=} \mathsf{s}^{\mathrm{q}}_{\lambda}(\mathbf{X}_1|V_1) + \mathsf{s}^{\mathrm{q}}_{\lambda}(\mathbf{X}_2|V_2) \\
&\overset{(b)}{=} \mathsf{s}^{\mathrm{q}\times\mathrm{q}}_{\lambda}(\mathbf{X}_1, \mathbf{X}_2|V_1, V_2) \\
&\overset{(c)}{=} \mathsf{s}^{\mathrm{q}\times\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}|\tilde{V}) \\
&\overset{(d)}{\le} \mathsf{S}^{\mathrm{q}\times\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}) \\
&\overset{(e)}{\le} \mathsf{S}^{\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_1}) + \mathsf{S}^{\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_2}) \\
&\overset{(f)}{\le} \mathsf{V}^{\mathrm{q}}_{\lambda}(K) + \mathsf{V}^{\mathrm{q}}_{\lambda}(K) = 2\mathsf{V}^{\mathrm{q}}_{\lambda}(K).
\end{aligned}$$

Here (a) holds because $p_*(v, \mathbf{x})$ achieves $\mathsf{V}_{\lambda}(K)$; (b) holds because $(V_1, \mathbf{X}_1)$ and $(V_2, \mathbf{X}_2)$ are independent; (c) is a consequence of Proposition 1; (d) follows from the definition; (e) is a consequence of Proposition 6; finally (f) follows from the definition of $\mathsf{V}^{\mathrm{q}}_{\lambda}(K)$ by noticing

$$\begin{aligned}
\mathrm{E}(\mathbf{X}_{\theta_1}\mathbf{X}_{\theta_1}^T) &= \mathrm{E}(\mathbf{X}_{\theta_2}\mathbf{X}_{\theta_2}^T) \\
&= \sum_{v_1,v_2} p_*(v_1)p_*(v_2) \cdot \frac{1}{2}(K_{v_1} + K_{v_2}) \\
&= \sum_{v=1}^{m} p_*(v)K_v \preceq K.
\end{aligned}$$

Since the extremes match, all inequalities must be equalities. Notice (d) being an equality means that $p(\tilde{v}|\mathbf{x}_{\theta_1}, \mathbf{x}_{\theta_2})$ achieves $\mathsf{S}^{\mathrm{q}\times\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2})$. From Proposition 6, (d) and (e) being equalities implies that $\mathbf{X}_{\theta_1}$ and $\mathbf{X}_{\theta_2}$ are conditionally independent given $\tilde{V}$, $p(\tilde{v}|\mathbf{x}_{\theta_1})$ achieves $\mathsf{S}^{\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_1})$, and $p(\tilde{v}|\mathbf{x}_{\theta_2})$ achieves $\mathsf{S}^{\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_2})$. Finally from (f) we know $\mathsf{S}^{\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_1}) = \mathsf{V}^{\mathrm{q}}_{\lambda}(K) = \mathsf{S}^{\mathrm{q}}_{\lambda}(\mathbf{X}_{\theta_2})$. □

As a consequence, for any pair (v_1, v_2), $\mathbf{X}_{v_1} + \mathbf{X}_{v_2}$ and $\mathbf{X}_{v_1} - \mathbf{X}_{v_2}$ are independent. Combined with the fact that $\mathbf{X}_{v_1}$ and $\mathbf{X}_{v_2}$ are independent zero mean random variables, Corollary 3 in Appendix I-A implies that $\mathbf{X}_{v_1}$ and $\mathbf{X}_{v_2}$ are Gaussians with the same covariance matrix. Since (v_1, v_2) is arbitrary, all Gaussians $\mathbf{X}_{v_i}$ have the same covariance matrix, say K_*. Clearly $K_* \preceq K$. Let $\mathbf{X}_* \sim \mathcal{N}(0, K_*)$. Then

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

Hence we obtain the following theorem (originally established in [5]).

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

Proof. The existence is clear from the preceding argument and here, we only comment on the uniqueness. First we show that if a zero mean random variable $\mathbf{X}$ is a maximizer, that is $\mathsf{V}^{\mathrm{q}}_{\lambda}(K) = \mathsf{s}^{\mathrm{q}}_{\lambda}(\mathbf{X})$, it must be Gaussian. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be two i.i.d. copies of $\mathbf{X}$. Applying Proposition 8 (take V to be the trivial random variable), we obtain that $\mathbf{X}_1 + \mathbf{X}_2$ and $\mathbf{X}_1 - \mathbf{X}_2$ are also independent. Hence, from Corollary 3, $\mathbf{X}$ must be a Gaussian. Suppose $\mathsf{V}^{\mathrm{q}}_{\lambda}(K)$ has two Gaussian maximizers, say $\mathbf{G}_1 \sim \mathcal{N}(0, K_1)$ and $\mathbf{G}_2 \sim \mathcal{N}(0, K_2)$ such that $K_1, K_2 \preceq K$

and $K_1 \neq K_2$. Consider random variables $(V, \mathbf{X})$ such that V is binary (say uniformly distributed), $\mathbf{X}|(V=1) \sim \mathcal{N}(0, K_1)$ and $\mathbf{X}|(V=2) \sim \mathcal{N}(0, K_2)$. Then note that $(V, \mathbf{X})$ also attains $\mathsf{V}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K)$ and satisfies the covariance constraint. Now from Proposition 8, taking $v_1 = 1$ and $v_2 = 2$, we obtain that $\mathbf{G}_1 + \mathbf{G_2}$ is independent of $\mathbf{G}_1 - \mathbf{G_2}$, clearly impossible as $K_1 \neq K_2$ (see Corollary 3). □

Remark 8. Notice that we never used the precise form of $\mathsf{S}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X})$ but just the implications of Proposition 6. In the next section we will define a new concave envelope that also satisfies a condition similar to Proposition 6, and then establish the optimality of Gaussian distributions. In general, the Gaussian optimality can be established if one shows the factorization property (as mentioned earlier, this is related to the single-letterization arguments), the existence of the maximizer (in this regard see the arguments in Appendix II), and the invariance of the expressions with respect to the rotation operations (usually a consequence of the additive Gaussian noise model).

The following corollary will be useful later.

Corollary 1. *If $\mathbf{X} \sim \mathcal{N}(0, K)$ then there exists a decomposition of $\mathbf{X}$ into $\mathbf{X}_* \sim \mathcal{N}(0, K_*)$ and an independent random variable $\mathbf{X}' \sim \mathcal{N}(0, K - K_*)$, $K_* \preceq K$ such that $\mathsf{S}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}) = \mathsf{s}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_*) = \mathsf{V}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K)$. Further, this decomposition (i.e. the corresponding covariance matrix K_*) is unique.*

Proof. From Theorem 1, there exists $\mathbf{X}_* \sim \mathcal{N}(0, K_*)$, $K_* \preceq K$ such that $\mathsf{s}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_*) = \mathsf{V}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K)$. Let $\mathbf{X}' \sim \mathcal{N}(0, K - K_*)$ be independent of $\mathbf{X}_*$, and $\mathbf{X} = \mathbf{X}' + \mathbf{X}_*$. By definition, $\mathsf{S}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}) \leq \mathsf{V}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K)$. On the other hand since $\mathbf{X}|(\mathbf{X}' = \mathbf{x}') \sim \mathbf{X}_* + \mathbf{x}'$ we have $\mathsf{s}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|\mathbf{X}') = \mathsf{s}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_*)$. From the Markov chain $\mathbf{X}' \to \mathbf{X} \to (\mathbf{Y}_1, \mathbf{Y}_2)$ we obtain

$$\begin{aligned}\mathsf{S}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}) &= \sup_{V: V \to \mathbf{X} \to (\mathbf{Y}_1, \mathbf{Y}_2)} \mathsf{s}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|V) \\ &\geq \mathsf{s}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|\mathbf{X}') = \mathsf{s}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_*) = \mathsf{V}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K).\end{aligned}$$

The uniqueness is a direct consequence of Theorem 1. This finishes the proof. □

C. Example 3: A new extremal inequality

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

Consider a broadcast channel $\mathrm{q}(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x})$. For $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)$, where $\lambda_i > 0, i = 0, 1, 2$, $\lambda_2 > \lambda_1$, $\alpha \in [0, 1]$ and $\bar{\alpha} := 1 - \alpha$, consider the following function of $p(\mathbf{x})$ defined by

$$\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}) := -\lambda_0 \alpha I(\mathbf{X}; \mathbf{Y}_1) + (\lambda_2 - \lambda_0 \bar{\alpha}) I(\mathbf{X}; \mathbf{Y}_2) + \lambda_1 \mathsf{S}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}),$$

where $\mathsf{S}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X})$ is defined in Section II-B. As before, we define some terms based on $\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X})$. For $(W, \mathbf{X})$ such that $W \to \mathbf{X} \to (\mathbf{Y}_1, \mathbf{Y}_2)$ is Markov, let

$$\begin{aligned}\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|W) := &-\lambda_0 \alpha I(\mathbf{X}; \mathbf{Y}_1|W) + (\lambda_2 - \lambda_0 \bar{\alpha}) I(\mathbf{X}; \mathbf{Y}_2|W) \\ &+ \lambda_1 \mathsf{S}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}|W).\end{aligned}$$

Further define the upper concave envelope of $\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X})$ as

$$\mathsf{T}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}) := \mathfrak{C}(\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X})).$$

It is easy to see that

$$\begin{aligned}&\mathfrak{C}(\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X})) \\ &= \sup_{W: W \to \mathbf{X} \to (\mathbf{Y}_1, \mathbf{Y}_2)} \begin{aligned}[t]&\lambda_0 \alpha I(\mathbf{X}; \mathbf{Y}_1|W) + \lambda_1 \mathsf{S}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}|W) \\ &+ (\lambda_2 - \lambda_0 \bar{\alpha}) I(\mathbf{X}; \mathbf{Y}_2|W)\end{aligned} \\ &= \sup_{W: W \to \mathbf{X} \to (\mathbf{Y}_1, \mathbf{Y}_2)} \mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|W).\end{aligned}$$

We also define $\mathsf{T}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|U) := \sum_u p(u) \mathsf{T}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|U = u)$ for finite U and its natural extension for arbitrary U.

For a product broadcast channel $\mathrm{q}_1(\mathbf{y}_{11}, \mathbf{y}_{21}|\mathbf{x}_1) \times \mathrm{q}_2(\mathbf{y}_{12}, \mathbf{y}_{22}|\mathbf{x}_2)$ we define

$$\begin{aligned}&\mathsf{t}^{\mathrm{q}_1 \times \mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_1, \mathbf{X}_2) \\ &:= -\lambda_0 \alpha I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_{11}, \mathbf{Y}_{12}) \\ &\quad + (\lambda_2 - \lambda_0 \bar{\alpha}) I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_{21}, \mathbf{Y}_{22}) + \lambda_1 \mathsf{S}^{\mathrm{q}_1 \times \mathrm{q}_2}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}_1, \mathbf{X}_2),\end{aligned}$$

and also the terms $\mathsf{t}^{\mathrm{q}_1 \times \mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_1, \mathbf{X}_2|W)$, $\mathsf{T}^{\mathrm{q}_1 \times \mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_1, \mathbf{X}_2)$ and $\mathsf{T}^{\mathrm{q}_1 \times \mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_1, \mathbf{X}_2|W)$. The inequality in the following proposition is referred to as the *factorization* of $\mathsf{T}^{\mathrm{q}_1 \times \mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_1, \mathbf{X}_2)$.

Proposition 9. *When $\lambda_0 > \lambda_2$ the following inequality holds for product broadcast channels*

$$\begin{aligned}\mathsf{T}^{\mathrm{q}_1 \times \mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_1, \mathbf{X}_2) &\leq \mathsf{T}^{\mathrm{q}_1}_{\boldsymbol{\lambda}}(\mathbf{X}_1|\mathbf{Y}_{22}) + \mathsf{T}^{\mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_2|\mathbf{Y}_{11}) \\ &\leq \mathsf{T}^{\mathrm{q}_1}_{\boldsymbol{\lambda}}(\mathbf{X}_1) + \mathsf{T}^{\mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_2).\end{aligned}$$

Further, for a Gaussian product broadcast channel, if a particular triple $(W_*, \mathbf{X}_{1*}, \mathbf{X}_{2*})$ satisfies

$$\begin{aligned}\mathsf{t}^{\mathrm{q}_1 \times \mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_{1*}, \mathbf{X}_{2*}|W_*) &= \mathsf{T}^{\mathrm{q}_1 \times \mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_{1*}, \mathbf{X}_{2*}) \\ &= \mathsf{T}^{\mathrm{q}_1}_{\boldsymbol{\lambda}}(\mathbf{X}_{1*}) + \mathsf{T}^{\mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_{2*}),\end{aligned}$$

then all of the following must be true:

1) *$\mathbf{X}_{1*}$ and $\mathbf{X}_{2*}$ are conditionally independent given W_*,*
2) *$\mathsf{T}^{\mathrm{q}_1}_{\boldsymbol{\lambda}}(\mathbf{X}_{1*}) = \mathsf{t}^{\mathrm{q}_1}_{\boldsymbol{\lambda}}(\mathbf{X}_{1*}|W_*)$,*
3) *$\mathsf{T}^{\mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_{2*}) = \mathsf{t}^{\mathrm{q}_2}_{\boldsymbol{\lambda}}(\mathbf{X}_{2*}|W_*)$.*

Proof. For any Markov chain $W \to (\mathbf{X}_1, \mathbf{X}_2) \to$

$(\mathbf{Y}_{11},\mathbf{Y}_{12},\mathbf{Y}_{21},\mathbf{Y}_{22})$, observe

$$\begin{aligned}
&\mathsf{t}_{\boldsymbol{\lambda}}^{\mathrm{q}_1\times\mathrm{q}_2}(\mathbf{X}_1,\mathbf{X}_2|W)\\
&=-\lambda_0\alpha I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{11},\mathbf{Y}_{12}|W)\\
&\quad+(\lambda_2-\lambda_0\bar{\alpha})I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_{21},\mathbf{Y}_{22}|W)+\lambda_1\mathsf{S}_{\frac{\lambda_2}{\lambda_1}}^{\mathrm{q}_1\times\mathrm{q}_2}(\mathbf{X}_1,\mathbf{X}_2|W)\\
&\overset{(a)}{=}-\lambda_0\alpha\big(I(\mathbf{X}_1;\mathbf{Y}_{11}|W,\mathbf{Y}_{22})+I(\mathbf{X}_2;\mathbf{Y}_{12}|W,\mathbf{Y}_{11})\big)\\
&\quad+(\lambda_2-\lambda_0\bar{\alpha})\big(I(\mathbf{X}_2;\mathbf{Y}_{22}|W,\mathbf{Y}_{11})+I(\mathbf{X}_1;\mathbf{Y}_{21}|W,\mathbf{Y}_{22})\big)\\
&\quad-(\lambda_0-\lambda_2)I(\mathbf{Y}_{11};\mathbf{Y}_{22}|W)+\lambda_1\mathsf{S}_{\frac{\lambda_2}{\lambda_1}}^{\mathrm{q}_1\times\mathrm{q}_2}(\mathbf{X}_1,\mathbf{X}_2|W)\\
&\overset{(b)}{\le}-\lambda_0\alpha\big(I(\mathbf{X}_1;\mathbf{Y}_{11}|W,\mathbf{Y}_{22})+I(\mathbf{X}_2;\mathbf{Y}_{12}|W,\mathbf{Y}_{11})\big)\\
&\quad+(\lambda_2-\lambda_0\bar{\alpha})\big(I(\mathbf{X}_2;\mathbf{Y}_{22}|W,\mathbf{Y}_{11})+I(\mathbf{X}_1;\mathbf{Y}_{21}|W,\mathbf{Y}_{22})\big)\\
&\quad-(\lambda_0-\lambda_2)I(\mathbf{Y}_{11};\mathbf{Y}_{22}|W)\\
&\quad+\lambda_1\mathsf{S}_{\frac{\lambda_2}{\lambda_1}}^{\mathrm{q}_1}(\mathbf{X}_1|W,\mathbf{Y}_{22})+\lambda_1\mathsf{S}_{\frac{\lambda_2}{\lambda_1}}^{\mathrm{q}_2}(\mathbf{X}_2|W,\mathbf{Y}_{11})\\
&\overset{(c)}{\le}\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}_1}(\mathbf{X}_1|\mathbf{Y}_{22})+\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}_2}(\mathbf{X}_2|\mathbf{Y}_{11})-(\lambda_0-\lambda_2)I(\mathbf{Y}_{11};\mathbf{Y}_{22}|W)\\
&\overset{(d)}{\le}\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}_1}(\mathbf{X}_1)+\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}_2}(\mathbf{X}_2)-(\lambda_0-\lambda_2)I(\mathbf{Y}_{11};\mathbf{Y}_{22}|W)\\
&\overset{(e)}{\le}\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}_1}(\mathbf{X}_1)+\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}_2}(\mathbf{X}_2).
\end{aligned}$$

Here (a) is similar to step (b) in the proof of Proposition 6 since $W\to(\mathbf{X}_1,\mathbf{X}_2)\to(\mathbf{Y}_{11},\mathbf{Y}_{12},\mathbf{Y}_{21},\mathbf{Y}_{22})$ is Markov; (b) is from Proposition 6 by the definition of $\mathsf{S}_{\lambda}^{\mathrm{q}}(\cdot|\cdot)$ and that $W\to(\mathbf{X}_1,\mathbf{X}_2)\to(\mathbf{Y}_{11},\mathbf{Y}_{12},\mathbf{Y}_{21},\mathbf{Y}_{22})$ is Markov; (c) is due to the Markov chains $(W,\mathbf{Y}_{22})\to\mathbf{X}_1\to(\mathbf{Y}_{11},\mathbf{Y}_{21})$ and $(W,\mathbf{Y}_{11})\to\mathbf{X}_2\to(\mathbf{Y}_{12},\mathbf{Y}_{22})$; (d) is due to the concavity of $\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\cdot)$.

Now for $(W_*,\mathbf{X}_{1*},\mathbf{X}_{2*})$, since the end-to-end equality holds, from $\lambda_0>\lambda_2$ and (e) we have $I(\mathbf{Y}_{11*};\mathbf{Y}_{22*}|W_*)=0$. By Proposition 2 we have that $\mathbf{X}_{1*}$ and $\mathbf{X}_{2*}$ are conditionally independent given W_*, which implies the Markov chain $(\mathbf{Y}_{11*},\mathbf{Y}_{21*})\to\mathbf{X}_{1*}\to W_*\to\mathbf{X}_{2*}\to(\mathbf{Y}_{12*},\mathbf{Y}_{22*})$. Now using the equality observe that

$$\begin{aligned}
&\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}_1}(\mathbf{X}_{1*})\\
&=-\lambda_0\alpha I(\mathbf{X}_{1*};\mathbf{Y}_{11*}|W_*,\mathbf{Y}_{22*})\\
&\quad+(\lambda_2-\lambda_0\bar{\alpha})I(\mathbf{X}_{1*};\mathbf{Y}_{21*}|W_*,\mathbf{Y}_{22*})+\lambda_1\mathsf{S}_{\frac{\lambda_2}{\lambda_1}}^{\mathrm{q}_1}(\mathbf{X}_{1*}|W_*,\mathbf{Y}_{22*})\\
&\overset{(e)}{=}-\lambda_0\alpha I(\mathbf{X}_{1*};\mathbf{Y}_{11*}|W_*)+(\lambda_2-\lambda_0\bar{\alpha})I(\mathbf{X}_{1*};\mathbf{Y}_{21*}|W_*)\\
&\quad+\lambda_1\mathsf{S}_{\frac{\lambda_2}{\lambda_1}}^{\mathrm{q}_1}(\mathbf{X}_{1*}|W_*,\mathbf{Y}_{22*})\\
&\overset{(f)}{=}-\lambda_0\alpha I(\mathbf{X}_{1*};\mathbf{Y}_{11*}|W_*)+(\lambda_2-\lambda_0\bar{\alpha})I(\mathbf{X}_{1*};\mathbf{Y}_{21*}|W_*)\\
&\quad+\lambda_1\mathsf{S}_{\frac{\lambda_2}{\lambda_1}}^{\mathrm{q}_1}(\mathbf{X}_{1*}|W_*),
\end{aligned}$$

where (e) is from the Markov chain $\mathbf{Y}_{22*}\to W_*\to\mathbf{X}_{1*}\to(\mathbf{Y}_{11*},\mathbf{Y}_{21*})$ and (f) is from the Markov chain $\mathbf{Y}_{22*}\to W_*\to\mathbf{X}_{1*}$. Similar result holds for $\mathbf{X}_{2*}$. This completes the proof. □

For $K\succeq 0$, define

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

Proposition 10. *There exists a pair $(W_*,\mathbf{X}_*)$ with $|\mathcal{W}_*|\le\frac{t(t+1)}{2}+1$ and $\mathrm{E}(\mathbf{X}_*\mathbf{X}_*^T)\preceq K$ such that*

$$\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K)=\mathsf{t}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_*|W_*).$$

Further, we can assume that the conditional law of $\mathbf{X}_*|(W_*=w_*)$ has zero mean for every w_*.

Proof. The existence of a maximizer and the cardinality bound on W_* are established in Appendix II-B. The centering argument works as before. □

Proposition 11. *Let $(W_*,\mathbf{X}_*)\sim p_*(w,\mathbf{x})$ attain $\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K)$, with $|\mathcal{W}|=m\le\frac{t(t+1)}{2}+1$; and let $\mathbf{X}_w$ denote a zero mean random variable distributed according to the conditional distribution $p_*(\mathbf{x}|W=w)$. 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 $p_*(w,x)$. Define*

$$\begin{aligned}
\tilde{W}=(W_1,W_2),\quad &\mathbf{X}_{\theta_1}|\big(\tilde{W}=(w_1,w_2)\big)\sim\frac{1}{\sqrt{2}}(\mathbf{X}_{w_1}+\mathbf{X}_{w_2}),\\
&\mathbf{X}_{\theta_2}|\big(\tilde{W}=(w_1,w_2)\big)\sim\frac{1}{\sqrt{2}}(\mathbf{X}_{w_1}-\mathbf{X}_{w_2}).
\end{aligned}$$

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

1) *$\mathbf{X}_{\theta_1}$ and $\mathbf{X}_{\theta_2}$ are conditionally independent given $\tilde{W}$,*
2) *$\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K)=\mathsf{t}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_{\theta_1}|\tilde{W})$,*
3) *$\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K)=\mathsf{t}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_{\theta_2}|\tilde{W})$.*

Proof. The proof mirrors that of Proposition 8. Let $K_w=\mathrm{E}(\mathbf{X}_w\mathbf{X}_w^T)$. Consider a two-letter broadcast channel $\mathrm{q}(\mathbf{y}_{11},\mathbf{y}_{21}|\mathbf{x}_1)\times\mathrm{q}(\mathbf{y}_{12},\mathbf{y}_{22}|\mathbf{x}_2)$. We have

$$\begin{aligned}
2\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K)&\overset{(a)}{=}\mathsf{t}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_1|W_1)+\mathsf{t}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_2|W_2)\\
&\overset{(b)}{=}\mathsf{t}_{\boldsymbol{\lambda}}^{\mathrm{q}\times\mathrm{q}}(\mathbf{X}_1,\mathbf{X}_2|W_1,W_2)\\
&\overset{(c)}{=}\mathsf{t}_{\boldsymbol{\lambda}}^{\mathrm{q}\times\mathrm{q}}(\mathbf{X}_{\theta_1},\mathbf{X}_{\theta_2}|\tilde{W})\\
&\overset{(d)}{\le}\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}\times\mathrm{q}}(\mathbf{X}_{\theta_1},\mathbf{X}_{\theta_2})\\
&\overset{(e)}{\le}\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_{\theta_1})+\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_{\theta_2})\\
&\overset{(f)}{\le}\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K)+\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K)=2\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K).
\end{aligned}$$

Here (a) comes because $p_*(w,\mathbf{x})$ achieves $\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K)$; (b) because $(W_1,\mathbf{X}_1)$ and $(W_2,\mathbf{X}_2)$ are independent; (c) is a consequence of Proposition 1; (e) is a consequence of Proposition 9; and (f) follows from the definition of $\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K)$ by noticing that

$$\begin{aligned}
\mathrm{E}(\mathbf{X}_{\theta_1}\mathbf{X}_{\theta_1}^T)&=\mathrm{E}(\mathbf{X}_{\theta_2}\mathbf{X}_{\theta_2}^T)\\
&=\sum_{w_1,w_2}p_*(w_1)p_*(w_2)\cdot\frac{1}{2}(K_{w_1}+K_{w_2})\\
&=\sum_{w=1}^{m}p_*(w)K_w\preceq K.
\end{aligned}$$

Since extremes of the chain of inequalities match, all inequalities must be equalities. Notice (d) being an equality means that $p(\tilde{w}|\mathbf{x}_{\theta_1},\mathbf{x}_{\theta_2})$ achieves $\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}\times\mathrm{q}}(\mathbf{X}_{\theta_1},\mathbf{X}_{\theta_2})$. Now from Proposition 9, (d) and (e) being equalities implies that $\mathbf{X}_{\theta_1}$ and $\mathbf{X}_{\theta_2}$ are conditionally independent given $\tilde{W}$, $p(\tilde{w}|\mathbf{x}_{\theta_1})$ achieves $\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_{\theta_1})$, and $p(\tilde{w}|\mathbf{x}_{\theta_2})$ achieves $\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_{\theta_2})$. Finally from (f) we know $\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_{\theta_1})=\hat{\mathsf{V}}_{\boldsymbol{\lambda}}^{\mathrm{q}}(K)=\mathsf{T}_{\boldsymbol{\lambda}}^{\mathrm{q}}(\mathbf{X}_{\theta_2})$. □

As a consequence, for any fixed (w_1,w_2), $\mathbf{X}_{w_1}+\mathbf{X}_{w_2}$ and $\mathbf{X}_{w_1}-\mathbf{X}_{w_2}$ are independent. Combined with the fact that $\mathbf{X}_{w_1}$ and $\mathbf{X}_{w_2}$ are independent zero mean random variables, from Corollary 3 in Appendix I-A, we obtain that $\mathbf{X}_{w_1}$ and $\mathbf{X}_{w_2}$ are Gaussians with the same covariance matrix. Since (w_1,w_2) is

arbitrary, all Gaussians $\mathbf{X}_{w_i}$ have the same covariance matrix, say K_*. Clearly $K_* \preceq K$. Let $\mathbf{X}_* \sim \mathcal{N}(0, K_*)$. Then

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

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

Theorem 2. *There exists $\mathbf{X}_* \sim \mathcal{N}(0, K_*)$, $K_* \preceq K$ such that $\hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K) = \mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_*)$. Further the zero mean maximizer is unique.*

Corollary 2. *If $\mathbf{X} \sim \mathcal{N}(0, K)$ then there exists a decomposition into $\mathbf{X}_{1*} \sim \mathcal{N}(0, K_1)$ and an independent random variable $\mathbf{X}_{2*} \sim \mathcal{N}(0, K_2)$, $K_1 + K_2 = K_* \preceq K$ such that $\mathsf{T}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}) = \mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_{1*} + \mathbf{X}_{2*}) = \hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K)$ and $\mathsf{S}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}_{1*} + \mathbf{X}_{2*}) = \mathsf{s}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}_{1*}) = \mathsf{V}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(K_1 + K_2)$. Further, this decomposition (i.e. the corresponding covariance matrices) is unique.*

Proof. From Theorem 2, there exists $\mathbf{X}_* \sim \mathcal{N}(0, K_*)$, $K_* \preceq K$ such that $\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_*) = \hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K)$. Now let $\mathbf{X}' \sim \mathcal{N}(0, K - K_*)$ be independent of $\mathbf{X}_*$, and let $\mathbf{X} = \mathbf{X}' + \mathbf{X}_*$. Thus $\mathbf{X} \sim \mathcal{N}(0, K)$. By definition $\mathsf{T}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}) \leq \hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K)$. On the other hand since $\mathbf{X}|(\mathbf{X}' = \mathbf{x}') \sim \mathbf{X}_* + \mathbf{x}'$ we have $\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|\mathbf{X}') = \mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_*)$. From Markov chain $\mathbf{X}' \to \mathbf{X} \to (\mathbf{Y}_1, \mathbf{Y}_2)$ we obtain

$$\begin{aligned}\mathsf{T}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}) &= \sup_{W: W \to \mathbf{X} \to (\mathbf{Y}_1, \mathbf{Y}_2)} \mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|W) \\ &\geq \mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|\mathbf{X}') = \mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_*) = \hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K).\end{aligned}$$

Now by Corollary 1, it is possible to split $\mathbf{X}_*$ into independent $\mathbf{X}_{1*}$ and $\mathbf{X}_{2*}$, such that $\mathsf{S}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}_{1*} + \mathbf{X}_{2*}) = \mathsf{s}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}_{1*}) = \mathsf{V}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(K_1 + K_2)$, is possible by Corollary 1. Further uniqueness of the covariance matrices is a consequence of Theorem 2 and Corollary 1. □

III. Two capacity regions

A. Gaussian vector broadcast channel with private messages

Consider a Gaussian vector broadcast channel $\mathrm{q}(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x})$ with only private messages. We will show, using the results from Section II-B, that an inner bound (Bound 2) to the capacity region matches an outer bound (Bound 1) to the capacity region under this setting. The proof will also show that Gaussian random variables suffice to characterize the capacity region, thus making the capacity region computable.

Consider the Körner-Marton outer bound [6] and Marton's inner bound [6] to the capacity region, $\mathcal{C}$, of the broadcast channel.

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

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

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 convex closure of the union of rate pairs (R_1, R_2) satisfying*

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

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

Denote this region as $\mathcal{I}$. This inner bound is obtained by taking the convex closure of the achievable region obtained using Theorem 2 in [6] by taking $W = \emptyset$. One is allowed to take the convex closure since any point in the interior of the convex closure can be obtained via time-sharing.

One can adapt these inner and outer bounds to the additive Gaussian setting by introducing a power constraint, i.e. $\mathrm{tr}(\mathrm{E}(\mathbf{X}\mathbf{X}^T)) \leq P$. Instead here we impose a covariance constraint $\mathrm{E}(\mathbf{X}\mathbf{X}^T) \preceq K$ and denote $\mathcal{I}_K, \mathcal{C}_K, \mathcal{O}_K$ to be the corresponding inner bound, capacity region, and outer bound. If one determines the capacity region under a covariance constraint, then the capacity region under the trace constraint can be obtained by taking the union over all the covariance matrices satisfying the trace constraint. By definition, we have $\mathcal{I}_K \subseteq \mathcal{C}_K \subseteq \mathcal{O}_K$. We now wish to show that $\mathcal{O}_K \subseteq \mathcal{I}_K$. We will show this inclusion using the supporting hyperplanes characterization of closed convex sets.

A closed and bounded convex set can be characterized by the intersection of its supporting hyperplanes. The regions $\mathcal{I}_K, \mathcal{O}_K$ are closed and bounded subsets in the first quadrant. Clearly the following hyperplanes

$$\begin{aligned}R_1 \geq 0, \quad R_1 &\leq \max_{\mathbf{X}: \mathrm{E}(\mathbf{X}\mathbf{X}^T) \preceq K} I(\mathbf{X}; \mathbf{Y}_1) =: C_1^K, \\ R_2 \geq 0, \quad R_2 &\leq \max_{\mathbf{X}: \mathrm{E}(\mathbf{X}\mathbf{X}^T) \preceq K} I(\mathbf{X}; \mathbf{Y}_2) =: C_2^K\end{aligned}$$

are supporting hyperplanes to $\mathcal{I}_K$ and $\mathcal{O}_K$. Further the points $(C_1^K, 0)$ and $(0, C_2^K)$ lie on the boundary of $\mathcal{I}_K$ as well as that of $\mathcal{O}_K$. Therefore to show that the regions coincide, it suffices to show, for $\lambda_1, \lambda_2 > 0$, that

$$\max_{(R_1, R_2) \in \mathcal{O}_K} \lambda_1 R_1 + \lambda_2 R_2 \leq \max_{(R_1, R_2) \in \mathcal{I}_K} \lambda_1 R_1 + \lambda_2 R_2.$$

In the rest of this section we will show that, for $\lambda > 1$, (equivalently taking $\lambda_2 > \lambda_1$ above)

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

The case for $\lambda < 1$ is dealt with similarly by interchanging roles of $\mathbf{Y}_1$ and $\mathbf{Y}_2$.

Remark 9. To show the $\lambda = 1$ case, observe that the function

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

is convex and bounded in λ when $\lambda \in (0, 2)$ (more generally, any bounded open interval containing 1 would suffice) which implies that $C(\lambda)$ is continuous in λ at $\lambda = 1$ (see Proposition 17, Chapter 5 [7]). Thus characterizing $C(\lambda)$ for all $\lambda > 1$ would also characterize $C(\lambda)$ at $\lambda = 1$.

Thus Marton's inner bound and Körner-Marton's outer bound will match under a covariance constraint.

Observe that

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

Here (a) is obtained using two constraints in the outer bound, namely $R_2 \leq I(V;\mathbf{Y}_2)$ and $R_1+R_2 \leq I(V;\mathbf{Y}_2)+I(\mathbf{X};\mathbf{Y}_1|V)$ and (b) is from the definition of $\mathsf{V}^{\mathrm{q}}_{\lambda}(K)$.

From Corollary 1, we know that there exists $\mathbf{X}_* \sim \mathcal{N}(0,K_*)$, $K_* \preceq K$ such that $\mathsf{V}^{\mathrm{q}}_{\lambda}(K)=\mathsf{s}^{\mathrm{q}}_{\lambda}(\mathbf{X}_*)$. Now let $V_* \sim \mathcal{N}(0,K-K_*)$ be independent of $\mathbf{X}_*$ and let $\mathbf{X}=V_*+\mathbf{X}_*$. Thus $\mathbf{X}\sim\mathcal{N}(0,K)$ maximizes $\lambda I(\mathbf{X};\mathbf{Y}_2)$ (subject to the covariance constraint) and

$$I(\mathbf{X};\mathbf{Y}_1|V_*)-\lambda I(\mathbf{X};\mathbf{Y}_2|V_*)=\mathsf{s}^{\mathrm{q}}_{\lambda}(\mathbf{X}|V_*)=\mathsf{s}^{\mathrm{q}}_{\lambda}(\mathbf{X}_*)=\mathsf{V}^{\mathrm{q}}_{\lambda}(K).$$

Hence

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

Proposition 12 (Dirty paper coding)**.** *Let $\mathbf{X}=V_*+\mathbf{X}_*$ and V_*, $\mathbf{X}_*$ be independent Gaussians with covariances $K-K_*$, K_* respectively for some $0\preceq K_*\preceq K$. Let $\mathbf{Y}_1=G_1\mathbf{X}+\mathbf{Z}_1$, where $\mathbf{Z}_1\sim\mathcal{N}(0,I)$ is independent of $(V_*,\mathbf{X}_*)$. Set $U_*=\mathbf{X}_*+AV_*$ where $A=K_*G_1^T(G_1K_*G_1^T+I)^{-1}$; then*

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

Proof. This well-known identification (see Chapter 9.5 of [3]) stems from the celebrated paper [8]. □

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

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

According to Marton's inner bound, since $(U_*,V_*)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)$ is Markov and $\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K$, the rate pair $R_2=I(V_*;\mathbf{Y}_2)$, $R_1=I(U_*;\mathbf{Y}_1)-I(U_*;V_*)$ belongs to $\mathcal{I}_K$. Hence

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

Thus the inner and outer bounds match for vector Gaussian product channels establishing its capacity region. Further observe that equality (hence the extreme points of the capacity region) can be obtained using Gaussian distributions, thus making the region computable.

B. Gaussian vector broadcast channel with common message

Consider a Gaussian vector broadcast channel $\mathrm{q}(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x})$ with common and private message requirements. Let $\hat{\mathcal{C}}$ denote the capacity region. As in Section III-A we establish the capacity region by showing that a certain outer bound and a certain inner bound to the capacity region match. In particular, we consider the UVW outer bound [9] and Marton's inner bound ([6], see the guided exercise 10.(c) in page 391 of [10]) to the capacity region of the broadcast channel with private and common messages. Further we show that extreme points are achievable using auxiliaries that are jointly Gaussian and hence the regions are computable.

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

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

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

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

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

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

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

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

Given a covariance constraint $\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K$, let $\hat{\mathcal{C}}_K$, $\hat{\mathcal{O}}_K$, and $\hat{\mathcal{I}}_K$ denote the capacity region, outer bound, and the inner bound for a Gaussian broadcast channel computed under this input constraint. Given the supporting hyperplanes characterization of bounded and closed convex sets, using a similar reasoning as in Section III-A, it suffices to characterize $\max_{(R_0,R_1,R_2)\in\hat{\mathcal{C}}_K}\lambda_0R_0+\lambda_1R_1+\lambda_2R_2$ for $\lambda_0,\lambda_1,\lambda_2>0$. Without loss of generality, we can assume $\lambda_2>\lambda_1$ since the case $\lambda_2<\lambda_1$ can be dealt with similarly by interchanging $\mathbf{Y}_1$ and $\mathbf{Y}_2$, and the case $\lambda_2=\lambda_1$ follows via a continuity argument as discussed in Remark (9). We further Proposition that it suffices to restrict ourselves to the case $\lambda_0>\lambda_2$. This is due to the following elementary observation: If a rate triple (R_0,R_1,R_2) belongs to $\hat{\mathcal{C}}$ then so does the triple $(0,R_1,R_2+R_0)$. This inference is immediate by treating the common message to be part of the private message to receiver $\mathbf{Y}_2$. Now since $\lambda_0R_0+\lambda_1R_1+\lambda_2R_2\leq 0\cdot R_0+\lambda_1R_1+\lambda_2(R_2+R_0)$

we have

$$\begin{aligned}&\max_{(R_0,R_1,R_2)\in\hat{\mathcal{C}}_K}\lambda_0R_0+\lambda_1R_1+\lambda_2R_2\\&=\max_{(0,R_1,R_2)\in\hat{\mathcal{C}}_K}\lambda_1R_1+\lambda_2R_2\\&=\max_{(R_1,R_2)\in\mathcal{C}_K}\lambda_1R_1+\lambda_2R_2,\end{aligned}$$

where $\mathcal{C}_K$ is the private messages capacity region that was already characterized in Section III-A.

Hence to characterize $\hat{\mathcal{C}}_K$, it suffices to show that for all $\lambda_0>\lambda_2>\lambda_1>0$ we have

$$\begin{aligned}&\max_{(R_0,R_1,R_2)\in\hat{\mathcal{O}}_K}\lambda_0R_0+\lambda_1R_1+\lambda_2R_2\\&\le\max_{(R_0,R_1,R_2)\in\hat{\mathcal{I}}_K}\lambda_0R_0+\lambda_1R_1+\lambda_2R_2.\end{aligned}$$

Remark 10. The setting $\lambda_0\ge\lambda_1+\lambda_2$ can be deduced from the degraded message sets capacity region and an earlier result [11]; however this setting of parameters will be subsumed in our treatment.

For any $\alpha\in[0,1]$ observe that

$$\begin{aligned}&\max_{(R_0,R_1,R_2)\in\hat{\mathcal{O}}_K}\lambda_0R_0+\lambda_1R_1+\lambda_2R_2\\&\overset{(a)}{\le}\sup_{\substack{(V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}}\begin{array}{c}\lambda_0I(W;\mathbf{Y}_1)+\bar\alpha\lambda_0I(W;\mathbf{Y}_2)\\+\lambda_2I(V;\mathbf{Y}_2|W)+\lambda_1I(\mathbf{X};\mathbf{Y}_1|V,W)\end{array}\\&\overset{(b)}{=}\sup_{\substack{(V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}}\begin{array}{c}\alpha\lambda_0(I(X;\mathbf{Y}_1)-I(\mathbf{X};\mathbf{Y}_1|W))\\+\bar\alpha\lambda_0(I(X;\mathbf{Y}_2)-I(\mathbf{X};\mathbf{Y}_2|W))\\+\lambda_2(I(\mathbf{X};\mathbf{Y}_2|W)-I(\mathbf{X};\mathbf{Y}_2|V,W))\\+\lambda_1I(\mathbf{X};\mathbf{Y}_1|V,W)\end{array}\\&\overset{(c)}{\le}\sup_{\substack{W\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}}\begin{array}{c}\alpha\lambda_0I(X;\mathbf{Y}_1)+\bar\alpha\lambda_0I(X;\mathbf{Y}_2)\\-\alpha\lambda_0I(\mathbf{X};\mathbf{Y}_1|W)-\bar\alpha\lambda_0I(\mathbf{X};\mathbf{Y}_2|W)\\+\lambda_2I(\mathbf{X};\mathbf{Y}_2|W)+\lambda_1\mathsf{S}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}|W)\end{array}\\&\le\sup_{\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}(\alpha\lambda_0I(X;\mathbf{Y}_1)+\bar\alpha\lambda_0I(X;\mathbf{Y}_2))\\&\quad+\sup_{\substack{W\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}}\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|W)\\&=\sup_{\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}(\alpha\lambda_0I(X;\mathbf{Y}_1)+\bar\alpha\lambda_0I(X;\mathbf{Y}_2))+\hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K).\end{aligned}$$

Here (a) follows from the first, third, and fourth constraints of the UVW outer bound; (b) is due to the Markov chain $(V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)$; and (c) follows from the definition of $\mathsf{S}^{\mathrm{q}}_{\lambda}(\cdot|\cdot)$ by noticing that conditioned on W, $V\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)$ is Markov.

From Corollary 2, we know that there exist independent random variables $\mathbf{X}_{1*}\sim\mathcal{N}(0,K_1)$, $\mathbf{X}_{2*}\sim\mathcal{N}(0,K_2)$, $K_1+K_2\preceq K$, such that $\hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K)=\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_{1*}+\mathbf{X}_{2*})$ and $\mathsf{S}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}_{1*}+\mathbf{X}_{2*})=\mathsf{s}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}_{1*})$. 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 maximizes $\alpha\lambda_0I(\mathbf{X};\mathbf{Y}_1)+\bar\alpha\lambda_0I(\mathbf{X};\mathbf{Y}_2)$ and attains $\hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K)$ simultaneously. Indeed, from Corollary 2, the covariance matrices K_1 and K_2 are unique, i.e. there is a unique such decomposition.

In order to conform to notation in the bounds, let $V_*=\mathbf{X}_{2*}$, implying $\mathbf{X}=W_*+\mathbf{X}_{1*}+V_*$.

$$\begin{aligned}&\max_{(R_0,R_1,R_2)\in\mathcal{O}_K}\lambda_0R_0+\lambda_1R_1+\lambda_2R_2\\&\le\alpha\lambda_0I(\mathbf{X};\mathbf{Y}_1)+\bar\alpha\lambda_0I(\mathbf{X};\mathbf{Y}_2)-\alpha\lambda_0I(\mathbf{X};\mathbf{Y}_1|W_*)\\&\quad+(\lambda_2-\bar\alpha\lambda_0)I(\mathbf{X};\mathbf{Y}_2|W_*)+\lambda_1I(\mathbf{X};\mathbf{Y}_1|V_*,W_*)\\&\quad-\lambda_2I(\mathbf{X};\mathbf{Y}_2|V_*,W_*)\\&=\alpha\lambda_0I(W_*;\mathbf{Y}_1)+\bar\alpha\lambda_0I(W_*;\mathbf{Y}_2)+\lambda_2I(V_*;\mathbf{Y}_2|W_*)\\&\quad+\lambda_1I(\mathbf{X};\mathbf{Y}_1|V_*,W_*).\end{aligned}$$

Now using Proposition 12 choose $U_*=\mathbf{X}_{1*}+\tilde AV_*$ as before to have

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

Hence

$$\begin{aligned}&\max_{(R_0,R_1,R_2)\in\hat{\mathcal{O}}_K}\lambda_0R_0+\lambda_1R_1+\lambda_2R_2\\&\le\alpha\lambda_0I(W_*;\mathbf{Y}_1)+\bar\alpha\lambda_0I(W_*;\mathbf{Y}_2)+\lambda_2I(V_*;\mathbf{Y}_2|W_*)\\&\quad+\lambda_1\big(I(U_*;\mathbf{Y}_1|W_*)-I(U_*;V_*|W_*)\big)\\&\le\sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}}\begin{array}{c}\alpha\lambda_0I(W;\mathbf{Y}_1)\\+\bar\alpha\lambda_0I(W;\mathbf{Y}_2)+\lambda_2I(V;\mathbf{Y}_2|W)\\+\lambda_1\big(I(U;\mathbf{Y}_1|W)-I(U;V|W)\big).\end{array}\end{aligned}$$

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

$$\begin{aligned}&\max_{(R_0,R_1,R_2)\in\mathcal{O}_K}\lambda_0R_0+\lambda_1R_1+\lambda_2R_2\\&\le\min_{\alpha\in[0,1]}\sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}}\begin{array}{c}\alpha\lambda_0I(W;\mathbf{Y}_1)+\bar\alpha\lambda_0I(W;\mathbf{Y}_2)\\+\lambda_2I(V;\mathbf{Y}_2|W)\\+\lambda_1I(U;\mathbf{Y}_1|W)\\-\lambda_1I(U;V|W).\end{array}\end{aligned}$$

To complete the proof that the inner and outer bounds match we present the following Proposition 13 (essentially established in [4]). We will defer the proof of this proposition to Appendix I-B.

Proposition 13. *The following min-max interchange holds:*

$$\begin{aligned}&\min_{\alpha\in[0,1]}\sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}}\begin{array}{c}\alpha\lambda_0I(W;\mathbf{Y}_1)+\bar\alpha\lambda_0I(W;\mathbf{Y}_2)\\+\lambda_2I(V;\mathbf{Y}_2|W)+\lambda_1I(U;\mathbf{Y}_1|W)\\-\lambda_1I(U;V|W)\end{array}\\&=\sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}}\min_{\alpha\in[0,1]}\begin{array}{c}\alpha\lambda_0I(W;\mathbf{Y}_1)+\bar\alpha\lambda_0I(W;\mathbf{Y}_2)\\+\lambda_2I(V;\mathbf{Y}_2|W)\\+\lambda_1I(U;\mathbf{Y}_1|W)\\-\lambda_1I(U;V|W)\end{array}\\&=\sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}}\begin{array}{c}\lambda_0\min\{I(W;\mathbf{Y}_1),I(W;\mathbf{Y}_2)\}\\+\lambda_2I(V;\mathbf{Y}_2|W)+\lambda_1I(U;\mathbf{Y}_1|W)\\-\lambda_1I(U;V|W).\end{array}\end{aligned}$$

Now using Marton's inner bound we can always achieve the 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

$$\begin{aligned}&\max_{(R_0,R_1,R_2)\in\hat{\mathcal{O}}_K}\lambda_0R_0+\lambda_1R_1+\lambda_2R_2\\&\le\sup_{\substack{(U,V,W)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)\\ \mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K}}\begin{array}{c}\lambda_0\min\{I(W;\mathbf{Y}_1),I(W;\mathbf{Y}_2)\}\\+\lambda_2I(V;\mathbf{Y}_2|W)+\lambda_1I(U;\mathbf{Y}_1|W)\\-\lambda_1I(U;V|W)\end{array}\\&\le\max_{(R_0,R_1,R_2)\in\hat{\mathcal{I}}_K}\lambda_0R_0+\lambda_1R_1+\lambda_2R_2.\end{aligned}$$

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

C. 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 [11]. What we have established in the above arguments can be phrased as

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

where W, V are independent zero mean Gaussians with covariances K_w, K_v respectively, and $\mathbf{X}=U+V+W$, 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 [11] can be cast as

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

where (as before) W, V are independent zero mean Gaussians with covariances K_w, K_v respectively, and $\mathbf{X}=U+V+W$, 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 $\mathcal{R}_K=\hat{\mathcal{C}}_K$, i.e. in particular the following proposition.

Proposition 14. *The following min-max interchange holds:*

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

where the random variables $W\sim\mathcal{N}(0,K_w)$, $V\sim\mathcal{N}(0,K_v)$, and $U\sim\mathcal{N}(0,K-K_w-K_v)$ are mutually independent, and $\mathbf{X}=U+V+W$. Further the Markov relationship $(W,U,V)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)$ holds.

Proof. The non-trivial direction is to establish that

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

Let us define

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

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 and the uniqueness is a consequence of Corollary 2.

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

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

Comparing the first and last term, we obtain that

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

Case 1: If $\alpha^*\in(0,1)$ we have that $I(W_\beta;\mathbf{Y}_1)\leq I(W_\beta;\mathbf{Y}_2)$ whenever $\beta\leq\alpha^*$ and $I(W_\beta;\mathbf{Y}_1)\geq I(W_\beta;\mathbf{Y}_2)$ whenever $\beta\geq\alpha^*$. Consider the sequence of unique Gaussian maximizers $W_{\beta_n}, V_{\beta_n}, \mathbf{X}_{\beta_n}$ that attain $SR_{\boldsymbol{\lambda}}(\beta_n)$ as $\beta_n\uparrow\alpha^*$. By continuity of $SR_{\boldsymbol{\lambda}}(\alpha)$ in α (indeed it is Lipschitz continuous), and by the compactness of the set $\{(K_v,K_w): K_v,K_w\succeq 0, K_v+K_w\preceq K\}$, there is a convergent subsequence of the associated covariance matrices. Thus there is a maximizer $W^{(b)}_{\alpha^*}, V^{(b)}_{\alpha^*}, \mathbf{X}^{(b)}_{\alpha^*}$ that attains $SR_{\boldsymbol{\lambda}}(\alpha^*)$ such that $I(W^{(b)}_{\alpha^*};\mathbf{Y}_1)\leq I(W^{(b)}_{\alpha^*};\mathbf{Y}_2)$.

Similarly approaching α^* from above, we obtain another maximizer $W^{(a)}_{\alpha^*}, V^{(a)}_{\alpha^*}, \mathbf{X}^{(a)}_{\alpha^*}$ that attains $SR_{\boldsymbol{\lambda}}(\alpha^*)$ such that $I(W^{(a)}_{\alpha^*};\mathbf{Y}_1)\geq I(W^{(a)}_{\alpha^*};\mathbf{Y}_2)$. However by the uniqueness of the Gaussian maximizer at any $\alpha\in[0,1]$ we must have $I(W^{(b)}_{\alpha^*};\mathbf{Y}_1)=I(W^{(a)}_{\alpha^*};\mathbf{Y}_1)$ and $I(W^{(b)}_{\alpha^*};\mathbf{Y}_2)=I(W^{(a)}_{\alpha^*};\mathbf{Y}_2)$. This implies that $I(W^{(b)}_{\alpha^*};\mathbf{Y}_1)=I(W^{(a)}_{\alpha^*};\mathbf{Y}_1)=(W^{(b)}_{\alpha^*};\mathbf{Y}_2)=I(W^{(a)}_{\alpha^*};\mathbf{Y}_2)$.

Let us denote the unique maximizer as $W_{\alpha^*}, V_{\alpha^*}, \mathbf{X}_{\alpha^*}$. We now have $I(W_{\alpha^*};\mathbf{Y}_1)=I(W_{\alpha^*};\mathbf{Y}_2)$ and hence, as desired, we obtain

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

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

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

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

Thus we have established the required min-max interchange. □

IV. 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 Gaussian vector 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.

ACKNOWLEDGEMENTS

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 authors would like to thank Hon-Fah Chong and Yeow-Khiang Chia for alerting that an earlier version did not contain a proof for Proposition 14. The authors also wish to thank the anonymous referees for various suggestion that vastly improved the readability of the paper and for suggesting to include a proof of the wiretap setting as an illustration of our technique.

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

REFERENCES

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

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

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

[4] Y. Geng, A. Gohari, C. Nair, and Y. Yu, "On marton's inner bound and its optimality for classes of product broadcast channels," *Information Theory, IEEE Transactions on*, vol. 60, no. 1, pp. 22–41, 2014.

[5] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *Information Theory, IEEE Transactions on*, vol. 53, pp. 1839 –1851, may 2007.

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

[7] H. Royden, *Real analysis*. Macmillan, 1988.

[8] M. Costa, "Writing on dirty paper (corresp.)," *Information Theory, IEEE Transactions on*, vol. 29, pp. 439 – 441, may 1983.

[9] C. Nair, "A note on outer bounds for broadcast channel," *Presented at International Zurich Seminar*, 2010, http://arXiv.org/abs/1101.0640.

[10] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL, USA: Academic Press, Inc., 1982.

[11] H. Weingarten, Y. Steinberg, and S. Shamai, "On the capacity region of the multi-antenna broadcast channel with common messages," in *Information Theory, 2006 IEEE International Symposium on*, pp. 2195 –2199, july 2006.

[12] S. G. Ghurye and I. Olkin, "A characterization of the multivariate normal distribution," *The Annals of Mathematical Statistics*, vol. 33, no. 2, pp. pp. 533–541, 1962.

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

[14] D. D. Boos, "A converse to scheffe's theorem," *Annals of Statistics*, vol. 13, no. 1, pp. 423–427, 1985.

[15] M. Godavarti and A. O. Hero, "Convergence of differential entropies," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 171–176, 2004.

[16] L. Bunt, *Bijdrage tot de theorie der convexe puntverzamelingen*. PhD thesis, Univ. Groningne, Amsterdam, 1934.

[17] V. Y. Protasov and M. E. Shirokov, "Generalized compactness in linear spaces and its applications," *MATHEMATICS*, vol. 200, no. 5, pp. 697–722, 2009.

[18] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas - part II: The mimome wiretap channel," *Information Theory, IEEE Transactions on*, vol. 56, no. 11, pp. 5515–5532, 2010.

[19] F. Oggier and B. Hassibi, "The secrecy capacity of the mimo wiretap channel," *Information Theory, IEEE Transactions on*, vol. 57, no. 8, pp. 4961–4972, 2011.

[20] T. Liu and S. Shamai, "A note on the secrecy capacity of the multiple-antenna wiretap channel," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2547–2553, 2009.

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

Yanlin Geng (M'12) Yanlin Geng received his B.Sc. (mathematics) and M.Eng. (signal and information processing) from Peking University, and Ph.D. (information engineering) from The Chinese University of Hong Kong in 2006, 2009, and 2012, respectively. He is currently a postdoctoral researcher in the Information Engineering department at The Chinese University of Hong Kong.

Chandra Nair (M'02) Chandra Nair is an Associate Professor in the Information Engineering department of the Chinese University of Hong Kong. Dr. Nair received his Bachelor of Technology (B.Tech) degree in Electrical Engineering from the Indian Institute of Technology (IIT), Madras in 1999. Concurrently, he also completed a four year nurture program in Mathematics at the Institute of Mathematical Sciences (IMSc). He received a Masters (2002) and PhD (2005) in electrical engineering from Stanford University. Subsequently he was a postdoctoral fellow at the theory group in Microsoft Research (Redmond) for two years. Following this he joined the IE department, CUHK, as an assistant professor in Fall 2007. His research interests are on fundamental problems in various interdisciplinary pursuits involving information theory, combinatorial optimization, statistical physics, and algorithms.

APPENDIX I
SOME KNOWN RESULTS

A. A characterization of Gaussian distributions

Theorem 3 (Theorem 1 in [12])**.** *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 [13]. There were scalar versions of this known since the 30s, including Bernstein's theorem. The proof relies on solving the functional equation satisfied by the characteristic functions.

Corollary 3. *If $\mathbf{X}_1$ and $\mathbf{X}_2$ are 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. Let $\hat{\mathbf{X}}_i := \mathbf{X}_i - \mathrm{E}(\mathbf{X}_i)$, $i = 1, 2$. Notice that shifting random variables by constants doesn't affect the independence, we have

$$\mathrm{E}((\hat{\mathbf{X}}_1+\hat{\mathbf{X}}_2)(\hat{\mathbf{X}}_1-\hat{\mathbf{X}}_2)^T) = \mathrm{E}(\hat{\mathbf{X}}_1+\hat{\mathbf{X}}_2)\,\mathrm{E}(\hat{\mathbf{X}}_1-\hat{\mathbf{X}}_2)^T = 0.$$

On the other hand

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

Thus they have the same covariance matrix. □

B. Min-max theorem

We reproduce the following Corollary from the Appendix of [4].

Corollary 4 (Corollary 2 in [4])**.** *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 Proposition 13.

Proof of Proposition 13:

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

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

The following set is a convex set:

$$\mathcal{A} = \left\{ (a_1, a_2) : \begin{array}{c} a_i \leq T_i(p(u,v,w,\mathbf{x})), i = 1, 2 \\ \text{for some } p \in \mathcal{P} \end{array} \right\}$$

To show this, suppose we have (a_1, a_2), $(b_1, b_2) \in \mathcal{A}$, and $p, q \in \mathcal{P}$ such that $a_i \leq T_i(p)$, $b_i \leq T_i(q)$, $i = 1, 2$. Consider a new distribution $r(u, v, \tilde{w}, \mathbf{x})$ where $\tilde{W} = (Q, W)$, Q is Bernoulli(α), and $r(u, v, (0, w), \mathbf{x}) = \bar{\alpha}\, p(u, v, w, \mathbf{x})$, $r(u, v, (1, w), \mathbf{x}) = \alpha\, q(u, v, w, \mathbf{x})$. Clearly $\mathrm{E}_r(\mathbf{X}\mathbf{X}^T) = \bar{\alpha}\,\mathrm{E}_p(\mathbf{X}\mathbf{X}^T) + \alpha\,\mathrm{E}_q(\mathbf{X}\mathbf{X}^T) \preceq K$ thus $r \in \mathcal{P}$. Now $T_i(r) = \lambda_0 I_r(Q;\mathbf{Y}_i) + \bar{\alpha}T_i(p) + \alpha T_i(q) \geq \bar{\alpha}a_i + \alpha b_i$, $i = 1, 2$, which means $\bar{\alpha}(a_1, a_2) + \alpha(b_1, b_2) \in \mathcal{A}$. Thus $\mathcal{A}$ is convex.

Hence from Corollary 4, we have

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

□

APPENDIX II
EXISTENCE OF MAXIMIZING DISTRIBUTIONS

The aim of this section is to give formal proofs of Propositions 7 and 10 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 yielding 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[4]. 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.

A. Properties of additive Gaussian noise

In this section we will establish the validity of Proposition 7. For this we need some tools and results from analysis.

We first establish certain smoothness properties of distributions obtained according to $\mathbf{Y} = \mathbf{X}+\mathbf{Z}$, where $\mathbf{Z} \sim \mathcal{N}(0, I)$ is independent of $\mathbf{X}$. 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.

[4]A normal topological space is one where every two disjoint closed sets have disjoint open neighbourhoods.

For simplicity of notation, we consider the scalar case. Let $\tilde{F}(x) = \mathrm{P}(X \leq x)$. Note that $0 \leq \tilde{F}(x) \leq 1$. Then we see that since $f_z(z)$ has a density, we have

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

Thus we have

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

By the Dominated Convergence Theorem (to justify interchange of derivative and integration) Y has a density given by

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

Hence

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

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

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

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 $\nabla f_{\mathbf{Y}}(\mathbf{y})$. For the setting $\mathbf{Y} = G\mathbf{X} + \mathbf{Z}$, we have the same result.

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

Proposition 15 (Lemma 1 in [14])**.** *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*

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

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 $|f_n(\mathbf{y}) - f_n(\mathbf{y}_1)| < \epsilon\ \forall n \geq n(\mathbf{y},\epsilon)$, 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

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

Proposition 16. *Let $\{\mathbf{X}_n\}$ be any sequence of random variables and let $\mathbf{Y}_n = \mathbf{X}_n + \mathbf{Z}$ where $\mathbf{Z} \sim \mathcal{N}(0, I)$ is independent of $\{\mathbf{X}_n\}$. 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 the density is clear from Remark (13). To see the uniform equicontinuity observe that by the mean value theorem

$$\begin{aligned}|f_n(\mathbf{y}+\boldsymbol{\Delta}) - f_n(\mathbf{y})| &= |\nabla f_n(\mathbf{y}') \cdot \boldsymbol{\Delta}| \\ &\overset{(a)}{\leq} \|\nabla f_n(\mathbf{y}')\|_1 \|\boldsymbol{\Delta}\|_\infty \\ &\leq \|\nabla f_n(\mathbf{y}')\|_1 \|\boldsymbol{\Delta}\|_2\end{aligned}$$

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: 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 $\mathrm{P}(\mathbf{X}_n \notin C_\epsilon) \leq \epsilon,\ \forall n$.

Proposition 17. *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 $C_\epsilon = \{\mathbf{x} : \|\mathbf{x}\|_2^2 \leq \frac{1}{\epsilon}\mathrm{tr}(K)\}$. By Markov's inequality $\mathrm{P}(\|\mathbf{X}_n\|^2 > \frac{1}{\epsilon}\mathrm{tr}(K)) \leq \frac{\epsilon\, \mathrm{E}(\|\mathbf{X}_n\|^2)}{\mathrm{tr}(K)} \leq \epsilon,\ \forall n.$ □

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 $\mathbf{X}_{n_i} \overset{w}{\Rightarrow} \mathbf{X}_*$.*

Proposition 18. *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 by the Levy's continuity theorem) since $\Phi_{\mathbf{Y}_n}(\mathbf{t}) = \Phi_{\mathbf{X}_n}(\mathbf{t}) e^{-\|\mathbf{t}\|^2/2}$. The second part (a stronger proposition than weak convergence) comes from Proposition 15. 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 $\|\mathbf{y}_n\| \to \infty$ (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 $|\mathbf{y}_n| \to \infty$ since the total integral is one). 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 $\mathrm{tr}(K)$, and the pointwise convergence from the second part. □

Theorem 5 (Theorem 1 in [15])**.** *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*

$$\max\{\sup_{\mathbf{y}} f_i(\mathbf{y}), \sup_{\mathbf{y}} f_*(\mathbf{y})\} \le F,$$
$$\max\{\int \|\mathbf{y}\|^\kappa f_i(\mathbf{y})d\mathbf{y}, \int \|\mathbf{y}\|^\kappa f_*(\mathbf{y})d\mathbf{y}\} \le L$$

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 $\liminf_i h(\mathbf{Y}_i) \ge h(\mathbf{Y}_*)$ coming due to the upper bound on the densities and $\limsup_i h(\mathbf{Y}_i) \le h(\mathbf{Y}_*)$ due to the moment constraints. A similar kind of result can be found in Appendix 3A of [3].

We now have the tools to prove Proposition 7.

Proof of Proposition 7:

Proof. Define

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

Let $\mathbf{X}_n$ be a sequence of random variables such that $\mathrm{E}(\mathbf{X}_n\mathbf{X}_n^T) = \hat{K}$ and $\mathsf{s}^{\mathsf{q}}_\lambda(\mathbf{X}_n) \uparrow \mathsf{v}^{\mathsf{q}}_\lambda(\hat{K})$. By the covariance constraint (Proposition 17) we know that the sequence of random variables $\mathbf{X}_n$ forms a tight sequence and by Theorem 4 there exists $X^*_{\hat{K}}$ and a convergent subsequence such that $\mathbf{X}_{n_i} \overset{w}{\Rightarrow} \mathbf{X}^*_{\hat{K}}$. From Proposition 18 we have that $h(\mathbf{Y}_{1n_i}), h(\mathbf{Y}_{2n_i}) \to h(\mathbf{Y}^*_{1\hat{K}}), h(\mathbf{Y}^*_{2\hat{K}})$ and hence $\mathsf{s}^{\mathsf{q}}_\lambda(\mathbf{X}^*_{\hat{K}}) = \mathsf{v}^{\mathsf{q}}_\lambda(\hat{K})$. We have the following trivial bound

$$\mathsf{v}^{\mathsf{q}}_\lambda(\hat{K}) = \mathsf{s}^{\mathsf{q}}_\lambda(\mathbf{X}^*_{\hat{K}}) \le I(\mathbf{X}^*_{\hat{K}}; \mathbf{Y}_1) \le \frac{1}{2}\log|I + G_1\hat{K}G_1^T|.$$

Recall that $\mathsf{V}^{\mathsf{q}}_\lambda(K)$ is defined using a convex combination as follows

$$\mathsf{V}^{\mathsf{q}}_\lambda(K) = \sup_{\substack{(V,\mathbf{X}):\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K \\ V\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)}} \mathsf{s}^{\mathsf{q}}_\lambda(\mathbf{X}|V).$$

Hence to obtain the best convex combination subject to the covariance constraint it suffices to restrict ourselves to the family of maximizers $\mathbf{X}^*_{\hat{K}}$ for $\hat{K} \succeq 0$. Thus, we can see that

$$\mathsf{V}^{\mathsf{q}}_\lambda(K) = \sup_{\substack{\alpha_i, \hat{K}_i : \alpha_i \ge 0, \sum_i \alpha_i = 1 \\ \sum_i \alpha_i \hat{K}_i \preceq K}} \sum_i \alpha_i \mathsf{v}^{\mathsf{q}}_\lambda(\hat{K}_i),$$

where $\{\alpha_i\}$ denotes a finite convex combination. It takes $\frac{t(t+1)}{2}$ constraints to preserve the covariance matrix and one constraint to preserve $\sum_i \alpha_i \mathsf{v}^{\mathsf{q}}_\lambda(\hat{K}_i)$. Hence, by using the Bunt-Carathedory theorem[5], we can restrict ourselves to convex combinations of at most $m := \frac{t(t+1)}{2} + 1$ points, i.e.

$$\mathsf{V}^{\mathsf{q}}_\lambda(K) = \sup_{\substack{\alpha_i, \hat{K}_i : \alpha_i \ge 0, \sum_{i=1}^m \alpha_i = 1 \\ \sum_{i=1}^m \alpha_i \hat{K}_i \preceq K}} \sum_{i=1}^m \alpha_i \mathsf{v}^{\mathsf{q}}_\lambda(\hat{K}_i).$$

[5]We need to use Bunt's extension [16] 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 the vanilla Caratheordory at the expense of one extra cardinality.

Consider any sequence of convex combinations $(\{\alpha_i^n\}, \{K_i^n\})$ that approaches the supremum as $n \to \infty$. Using compactness of the m-dimensional simplex, we can assume w.l.o.g. that $\alpha_i^n \overset{n\to\infty}{\to} \alpha_i^*, i = 1, .., m$. If any $\alpha_i^* = 0$, since $\alpha_i^n K_i^n \preceq K$ and $\mathsf{v}^{\mathsf{q}}_\lambda(K_i^n) \le \frac{1}{2}\log|I + G_1 K_i^n G_1^T|$ it is easy to see that $\alpha_i^n \mathsf{v}^{\mathsf{q}}_\lambda(K_i^n) \overset{n\to\infty}{\to} 0$. Thus we can assume that $\min_{i=1,..m} \alpha_i^* = \alpha^* > 0$. This implies that $K_i^n \preceq \frac{2}{\alpha^*}K$ for large enough n uniformly in i. Hence we can find a convergent subsequence for each $i, 1 \le i \le m$, so that $K_i^{n_k} \overset{k\to\infty}{\to} K_i^*$. Putting these together, we have

$$\mathsf{V}^{\mathsf{q}}_\lambda(K) = \sum_{i=1}^m \alpha_i^* \mathsf{v}^{\mathsf{q}}_\lambda(\hat{K}_i^*),$$

or in other words, we can find a pair of random variables $(V_*, \mathbf{X}_*)$ with $|V_*| \le \frac{t(t+1)}{2} + 1$ such that $\mathsf{V}^{\mathsf{q}}_\lambda(K) = \mathsf{s}^{\mathsf{q}}_\lambda(\mathbf{X}_*|V_*)$. □

B. Continuity in a pathwise sense on concave envelopes

In this section we will establish the validity of Proposition 10. For this we need more tools and results from analysis.

Proposition 19. *For $\lambda > 1$, there exists C_λ such that $\mathsf{s}^{\mathsf{q}}_\lambda(\mathbf{X}) \le C_\lambda$ for all $\mathbf{X}$.*

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

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

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

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

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 the 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 $\mathbb{R}^t$. Notice that t-dimensional subspace of $\mathbb{R}^t$ is unique, that is $L_t = \mathbb{R}^t$, we have

$$\mu_1(A) = \max_{L_t} \min_{0\ne u\in L_t} \frac{u^T A u}{u^T u} = \min_{0\ne u\in\mathbb{R}^t} \frac{u^T A u}{u^T u},$$
$$\mu_t(A) = \min_{L_t} \max_{0\ne u\in L_t} \frac{u^T A u}{u^T u} = \max_{0\ne u\in\mathbb{R}^t} \frac{u^T A u}{u^T u}.$$

Thus for any non-zero $u \in \mathbb{R}^t$ we have $\mu_1(A) \leq \frac{u^T A u}{u^T u} \leq \mu_t(A)$. Now

$$\begin{aligned}
&\mu_j(K+\Sigma) \\
&= \min_{L_j} \max_{0\neq u\in L_j} \left(\frac{u^T K u}{u^T u} + \frac{u^T \Sigma u}{u^T u}\right) \\
&\begin{cases} \geq \min_{L_j} \max_{0\neq u\in L_j} \left(\frac{u^T K u}{u^T u} + \mu_1(\Sigma)\right) = \mu_j(K) + \mu_1(\Sigma), \\ \leq \min_{L_j} \max_{0\neq u\in L_j} \left(\frac{u^T K u}{u^T u} + \mu_t(\Sigma)\right) = \mu_j(K) + \mu_t(\Sigma). \end{cases}
\end{aligned}$$

Hence we have $\mu_j(K) + \mu_1(\Sigma) \leq \mu_j(K+\Sigma) \leq \mu_j(K) + \mu_t(\Sigma)$, $j = 1, 2, \ldots, t$. Now

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

where $\mu_* = \max\{0, \frac{1}{\lambda-1}(\mu_1(\Sigma_2) - \lambda\mu_t(\Sigma_1))\}$, and (a) holds since $\mu_j(K) \geq 0$, the derivative of the function $f(x) := \log(x + \mu_t(\Sigma_1)) - \lambda \log(x + \mu_1(\Sigma_2))$ is zero at $x_* = \frac{1}{\lambda-1}(\mu_1(\Sigma_2) - \lambda\mu_t(\Sigma_1))$, negative when $x > x_*$ and positive when $x < x_*$. Finally the upper bound is finite by noticing that $\mu_t(\Sigma_1), \mu_1(\Sigma_2) > 0$ since the positive semi-definite matrices Σ_1 and Σ_2 are invertible. □

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

Recall our definition

$$\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}) = \mathfrak{C}(\mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X})) = \sup_{p(v|\mathbf{x}):V\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)} \mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X}|V).$$

Taking $V = \mathbf{X}$ we observe that $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}) \geq 0$. Define $\bar{\mathsf{s}}^{\mathrm{q}}_\lambda(\mathbf{X}) = \max\{\mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X}), 0\}$. Now note that $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}) = \mathfrak{C}(\bar{\mathsf{s}}^{\mathrm{q}}_\lambda(\mathbf{X}))$, since $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}) \geq 0$.

Let $\bar{\mathsf{s}}^{\mathrm{q},m}_\lambda(\mathbf{X})$ be $\bar{\mathsf{s}}^{\mathrm{q}}_\lambda(\mathbf{X})$ restricted to $\mathcal{A}_m$. Consider a sequence $\mathbf{X}_n \in \mathcal{A}_m$ such that $\mathbf{X}_n \overset{w}{\Rightarrow} \mathbf{X}_*$. Since the second moments are uniformly bounded, similar arguments as in Proposition 7 will imply that $\bar{\mathsf{s}}^{\mathrm{q},m}_\lambda(\mathbf{X}_n) \to \bar{\mathsf{s}}^{\mathrm{q},m}_\lambda(\mathbf{X}_*)$. Let $\mathsf{s}^{\mathrm{q},m}_\lambda(\mathbf{X})$ be the continuous extension of $\bar{\mathsf{s}}^{\mathrm{q},m}_\lambda(\mathbf{X})$ from $\mathcal{A}_m$ on to $\mathcal{P}$. This exists due to the Tietze Extension Theorem (stated below).

Theorem 6 (The 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.*

Further observe that the function $\mathsf{s}^{\mathrm{q},m}_\lambda(\mathbf{X})$ is bounded and non-negative since $\bar{\mathsf{s}}^{\mathrm{q},m}_\lambda(\mathbf{X})$ is bounded (from above by C_λ) and non-negative.

The following result follows from a recent result in [17]. The convex hull of a function $f(\mathbf{X})$ is the lower convex envelope, or equivalently $-\mathfrak{C}(-f(\mathbf{X}))$.

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 [17]. □

An immediate corollary, which follows from the fact that convex hull of $f(\mathbf{X})$ is $-\mathfrak{C}(-f(\mathbf{X}))$, 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 $\mathsf{S}^{\mathrm{q},m}_\lambda(\mathbf{X})$ to be concave envelope of $\mathsf{s}^{\mathrm{q},m}_\lambda(\mathbf{X})$. From Corollary 5 we have that $\mathsf{S}^{\mathrm{q},m}_\lambda(\mathbf{X})$ is continuous; Further since $\mathsf{s}^{\mathrm{q},m}_\lambda(\mathbf{X})$ is bounded and non-negative, so is $\mathsf{S}^{\mathrm{q},m}_\lambda(\mathbf{X})$. Continuity in particular implies that

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

Proposition 20 (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 $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_n) \to \mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_*)$.*

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

$$\begin{aligned}
&\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_n) \leq \mathsf{S}^{\mathrm{q},m}_\lambda(\mathbf{X}_n) + \epsilon, \\
\overset{n\to\infty}{\Longrightarrow} \quad &\limsup_n \mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_n) \leq \mathsf{S}^{\mathrm{q},m}_\lambda(\mathbf{X}_*) + \epsilon, \\
\overset{m\to\infty}{\Longrightarrow} \quad &\limsup_n \mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_n) \leq \mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_*) + \epsilon.
\end{aligned}$$

Similarly $\forall m > M_\epsilon$

$$\begin{aligned}
&\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_n) \geq \mathsf{S}^{\mathrm{q},m}_\lambda(\mathbf{X}_n) - \epsilon, \\
\overset{n\to\infty}{\Longrightarrow} \quad &\liminf_n \mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_n) \geq \mathsf{S}^{\mathrm{q},m}_\lambda(\mathbf{X}_*) - \epsilon, \\
\overset{m\to\infty}{\Longrightarrow} \quad &\liminf_n \mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_n) \geq \mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_*) - \epsilon.
\end{aligned}$$

Hence $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_n) \to \mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_*)$ provided we show the uniform convergence (in m) of $\mathsf{S}^{\mathrm{q},m}_\lambda(\mathbf{X}_n) \to \mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_n)$. Given $\epsilon > 0$ consider a V such that $\mathsf{S}^{\mathrm{q}}_\lambda(\mathbf{X}_n) \leq \mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X}_n|V) + \frac{\epsilon}{4}$. Observe that V induces a probability measure on the space of all probability measures. We now bound the induced probability measure assigned to distributions such that $\mathrm{E}(\|\mathbf{X}\|^2) \geq m$. Since $\mathrm{E}(\|\mathbf{X}_n\|^2) \leq \mathrm{tr}(K)$, 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 $\frac{\mathrm{tr}(K)}{m}$. Hence their contribution to $\mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X}_n|V)$ is at most $\frac{C_\lambda \mathrm{tr}(K)}{m}$, where C_λ is the global upper

bound on $\mathsf{s}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X})$. Thus by taking m large enough we can make this smaller than $\frac{\epsilon}{4}$. Hence

$$\mathsf{S}^{\mathrm{q},m}_{\boldsymbol{\lambda}}(\mathbf{X}_n) \geq \mathsf{s}^{\mathrm{q},m}_{\boldsymbol{\lambda}}(\mathbf{X}_n|V) \geq \mathsf{s}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_n|V) - \frac{\epsilon}{4} \geq \mathsf{S}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_n) - \frac{\epsilon}{2}.$$

Similar argument (taking V' such that $\mathsf{S}^{\mathrm{q},m}_{\boldsymbol{\lambda}}(\mathbf{X}_n) \leq \mathsf{s}^{\mathrm{q},m}_{\boldsymbol{\lambda}}(\mathbf{X}_n|V') + \frac{\epsilon}{4}$) also shows that $\mathsf{S}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_n) \geq \mathsf{S}^{\mathrm{q},m}_{\boldsymbol{\lambda}}(\mathbf{X}_n) - \frac{\epsilon}{2}$. Hence for all $m > \frac{4C_{\boldsymbol{\lambda}}\mathrm{tr}(K)}{\epsilon}$ we have that $|\mathsf{S}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_n) - \mathsf{S}^{\mathrm{q},m}_{\boldsymbol{\lambda}}(\mathbf{X}_n)| \leq \epsilon$ uniformly in n as desired. □

We now have the tools to prove Proposition 10.

Proof of Proposition 10:

Proof. The proof is similar to Proposition 7. Define

$$\begin{aligned}\hat{\mathsf{v}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\hat{K}) &:= \sup_{\mathbf{X}:\mathrm{E}(\mathbf{X}\mathbf{X}^T)=\hat{K}} \mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}) \\ &= \sup_{\mathbf{X}:\mathrm{E}(\mathbf{X}\mathbf{X}^T)=\hat{K}} \begin{matrix} -\lambda_0\alpha I(\mathbf{X};\mathbf{Y}_1) \\ +(\lambda_2-\lambda_0\bar{\alpha})I(\mathbf{X};\mathbf{Y}_2) \\ +\lambda_1\mathsf{S}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}) \end{matrix}.\end{aligned}$$

Let $\mathbf{X}_n$ be a sequence of random variables such that $\mathrm{E}(\mathbf{X}_n\mathbf{X}_n^T) = \hat{K}$ and $\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_n) \uparrow \hat{\mathsf{v}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\hat{K})$. By the covariance constraint (Proposition 17) we know that the sequence of random variables $\mathbf{X}_n$ forms a tight sequence and by Theorem 4 there exists $\mathbf{X}^*_{\hat{K}}$ and a convergent subsequence such that $\mathbf{X}_{n_i} \overset{w}{\Rightarrow} \mathbf{X}^*_{\hat{K}}$. Proposition 18 yields that $h(\mathbf{Y}_{1n_i}), h(\mathbf{Y}_{2n_i}) \to h(\mathbf{Y}^*_{1\hat{K}}), h(\mathbf{Y}^*_{2\hat{K}})$ and Proposition 20 yields $\mathsf{S}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}_{n_i}) \to \mathsf{S}^{\mathrm{q}}_{\frac{\lambda_2}{\lambda_1}}(\mathbf{X}^*_{\hat{K}})$. Hence $\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}^*_{\hat{K}}) = \hat{\mathsf{v}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\hat{K})$. Since $\hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K)$ is defined using a convex combination as follows

$$\hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K) = \sup_{\substack{(W,\mathbf{X}):\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K \\ W\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)}} \mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|W),$$

to obtain the maximizer subject to the covariance constraint it suffices to restrict ourselves to the family of maximizers $\mathbf{X}^*_{\hat{K}}$ for $\hat{K} \succeq 0$. Thus, we can see that

$$\hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K) = \sup_{\alpha_i,\hat{K}_i:\sum_i \alpha_i\hat{K}_i\preceq K} \sum_i \alpha_i \hat{\mathsf{v}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\hat{K}_i),$$

where $\{\alpha_i\}$ denotes a finite convex combination. It takes $\frac{t(t+1)}{2}$ constraints to preserve the covariance matrix and one constraint to preserve $\mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}|W)$. Hence by Bunt-Caratheodory's theorem and a similar argument as in the proof of Proposition 7, we can find a pair of random variables $(W_*, \mathbf{X}_*)$ with $|W_*| \leq \frac{t(t+1)}{2} + 1$ such that $\hat{\mathsf{V}}^{\mathrm{q}}_{\boldsymbol{\lambda}}(K) = \mathsf{t}^{\mathrm{q}}_{\boldsymbol{\lambda}}(\mathbf{X}_*|W_*)$. □

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

Proposition 21. *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) $\exists\, \kappa > 1$ such that $\mathrm{E}(\|\mathbf{X}_n\|^\kappa) \leq B$ $\forall n$ (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 in the proof of Proposition 20, 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 proposition 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.

APPENDIX III
GAUSSIAN VECTOR WIRETAP CHANNEL

In this section we will show how the techniques we introduced in this paper can be adapted to establish the optimality of Gaussian auxiliary random variables in the vector Gaussian wiretap channel setting. We only provide a brief outline since the details mimic the arguments in Section II-C.

Consider a vector Gaussian wiretap channel

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

where $\mathbf{Z}_1 \sim \mathcal{N}(0, I)$, $\mathbf{Z}_2 \sim \mathcal{N}(0, I)$, and the matrices G_1 and G_2 are invertible (see Remark (1)). Further we impose a covariance constraint on the input, i.e. $\mathrm{E}(\mathbf{X}\mathbf{X}^T) \preceq K$. The goal of the wiretapper setting is to communicate a message M to receiver $\mathbf{Y}_1$ while keeping the eavesdropper $\mathbf{Y}_2$ ignorant of the message. For formal description and known results in this setting the readers are urged to refer to Chapter 22 in [3].

The secrecy capacity, C_{S}, for vector Gaussian wiretap channel under a covariance constraint is given by

$$\begin{aligned}&\sup_{\substack{(Q,U,\mathbf{X}):\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K \\ (Q,U)\to\mathbf{X}\to(\mathbf{Y}_1,\mathbf{Y}_2)}} I(U;\mathbf{Y}_1|Q) - I(U;\mathbf{Y}_2|Q) \\ &= \sup_{\mathbf{X}:\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K} \mathfrak{C}\begin{pmatrix} I(\mathbf{X};\mathbf{Y}_1) - I(\mathbf{X};\mathbf{Y}_2) \\ +\mathfrak{C}(I(\mathbf{X};\mathbf{Y}_2) - I(\mathbf{X};\mathbf{Y}_1)) \end{pmatrix},\end{aligned}$$

where the equality is an immediate consequence of the definition of an upper concave envelope of a function. The conditioning on Q in the capacity expression is due to the covariance constraint; note that in the discrete memoryless setting one does not need this conditioning in the capacity formula. The achievability and the converse for this formula

follows from arguments in the discrete memoryless setting, see [3]. (Note that due to the covariance constraint one must stop the converse argument at equality labeled (d) on Page 555 of [3].)

Here we just outline the key steps (it could be a useful template for other settings as well), which are identical to those in Sections II-B and II-C.

- Consider the objective function $\mathfrak{C}(I(\mathbf{X};\mathbf{Y}_1) - \lambda I(\mathbf{X};\mathbf{Y}_2) + \mathfrak{C}(I(\mathbf{X};\mathbf{Y}_2) - \lambda I(\mathbf{X};\mathbf{Y}_1)))$, where $\lambda > 1$.
- Observe that $\mathfrak{C}(I(\mathbf{X};\mathbf{Y}_2) - \lambda I(\mathbf{X};\mathbf{Y}_1))$ satisfies the factorization property (see Proposition 6), and there is a maximizer for $\sup_{\mathbf{X}:\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K} \mathfrak{C}(I(\mathbf{X};\mathbf{Y}_2) - \lambda I(\mathbf{X};\mathbf{Y}_1))$ (see Proposition 7).
- Observe that $\mathfrak{C}(I(\mathbf{X};\mathbf{Y}_1) - \lambda I(\mathbf{X};\mathbf{Y}_2) + \mathfrak{C}(I(\mathbf{X};\mathbf{Y}_2) - \lambda I(\mathbf{X};\mathbf{Y}_1)))$ satisfies the factorization property (see Proposition 9), and there is a maximizer for $\sup_{\mathbf{X}:\mathrm{E}(\mathbf{X}\mathbf{X}^T)\preceq K} \mathfrak{C}(I(\mathbf{X};\mathbf{Y}_1) - \lambda I(\mathbf{X};\mathbf{Y}_2) + \mathfrak{C}(I(\mathbf{X};\mathbf{Y}_2) - \lambda I(\mathbf{X};\mathbf{Y}_1)))$ (see Proposition 10).
- By going to the two-letter version of the channel and using the invariance of mutual information with respect to rotations, in a fashion identical to the proof of Proposition 11, one obtains that the optimizer is a Gaussian.
- Observe the continuity of the maximum value at $\lambda = 1$ and let $\lambda \downarrow 1$. (see Remark (9)).

Using the above arguments we obtain that

$$C_{\mathrm{S}} = \frac{1}{2} \max_{0\preceq K_2 \preceq K_1 \preceq K} \log \frac{|I + G_1 K_1 G_1^T|}{|I + G_2 K_1 G_2^T|} + \log \frac{|I + G_2 K_2 G_2^T|}{|I + G_1 K_2 G_1^T|}.$$

The results in [18]–[20] seems to indicate that the maximum value is attained by setting $K_2 = 0$. As mentioned earlier (see Remark (5)), the techniques introduced in this paper are aimed at showing that the maximizers are Gaussian and further properties of the maximizers can be attained using standard optimization techniques.

Appendix IV
Alternate path to Theorem 1

Below, we will give an elementary proof of Theorem 1 without invoking Corollary 3. However this approach only shows that Gaussian is a maximizer and not necessarily the unique maximizer[6].

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 $(V^n, \tilde{\mathbf{X}}_n)$ achieves $\mathsf{V}_\lambda^{\mathrm{q}}(K)$ where $\tilde{\mathbf{X}}_n | \big(V^n = (v_1, v_2, .., v_n)\big) \sim \frac{1}{\sqrt{n}} \left(\mathbf{X}_{v_1} + \mathbf{X}_{v_2} + \cdots + \mathbf{X}_{v_n}\right)$. 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 Proposition 8. □

Consider $(V^n, \mathbf{X}^n) \sim \prod_{i=1}^n p_*(v_i, \mathbf{x}_i)$, where $p_*(v, \mathbf{x})$ achieves $\mathsf{V}_\lambda^{\mathrm{q}}(K)$. 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, $\mathcal{T}^{(n)}(V) := \{v^n : \big|\big|\{i : v_i = v\}\big| - np_*(v)\big| \leq n\omega_n p_*(v),\ \forall v \in [1:m]\}$. 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 Chebyshev's inequality)

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

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 $\hat{\mathbf{X}}_n \sim \tilde{\mathbf{X}}_n | v^n$, where by $\tilde{\mathbf{X}}_n | v^n$ we mean $\tilde{\mathbf{X}}_n | (V^n = v^n)$ for ease of notation.

Proposition 22. $\hat{\mathbf{X}}_n \overset{w}{\Rightarrow} \mathcal{N}(0, \sum_{v=1}^m p_*(v) K_v)$

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 $\hat{\mathbf{X}}_{n,i}^{\mathbf{c}} \sim \frac{1}{\sqrt{n}} \mathbf{c}^T \cdot \mathbf{X}_{v_i}$ and $\hat{\mathbf{X}}_{n,i}^{\mathbf{c}}$ be independent random variables over i. Note that $\sum_{i=1}^n \hat{\mathbf{X}}_{n,i}^{\mathbf{c}} \sim \mathbf{c}^T \hat{\mathbf{X}}_n$.

Note that

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

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

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[7] we have $\sum_{i=1}^n \hat{\mathbf{X}}_{n,i}^{\mathbf{c}} \overset{w}{\Rightarrow} \mathcal{N}(0, \mathbf{c}^T \sum_v p_*(v) K_v \mathbf{c})$. Hence $\hat{\mathbf{X}}_n \overset{w}{\Rightarrow} \mathcal{N}(0, \sum_v p_*(v) K_v)$ (Cramer-Wold device). □

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

Proposition 23. *Given any $\delta > 0$, there exists N_0 such that $\forall n > N_0$ we have for all $v^n \in \mathcal{T}^{(n)}(V)$*

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

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

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

[6] It is possible that this approach can be extended to provide a proof of uniqueness as well, but we do not pursue it here.

[7] We adopt the notation in Theorem (4.5), Chapter 2 in [21].

However from Proposition 22 we know that $\tilde{\mathbf{X}}_{n_k}|v^{n_k} \stackrel{w}{\Rightarrow} \mathbf{X}^*$ and from Proposition 18 we have $\mathsf{s}^{\mathrm{q}}_\lambda(\tilde{\mathbf{X}}_{n_k}|v^{n_k}) \to \mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X}^*)$, a contradiction. □

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

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 $\mathsf{V}^{\mathrm{q}}_\lambda(K)$. Hence

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

For a given v^n, let $\hat{\mathbf{X}} \sim \tilde{\mathbf{X}}_n|v^n$. Then note that $\mathrm{E}(\hat{\mathbf{X}}\hat{\mathbf{X}}^T) \preceq \sum_{v=1}^m K_v$. Thus $\mathsf{s}^{\mathrm{q}}_\lambda(\hat{\mathbf{X}}) \le I(\hat{\mathbf{X}};\mathbf{Y}_1) \le C$ for some fixed constant C that is independent of v^n. Thus using Proposition 23 we can upper bound $\mathsf{V}^{\mathrm{q}}_\lambda(K)$ for large n by

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

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 $\mathsf{V}^{\mathrm{q}}_\lambda(K) \le \mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X}^*) + \delta$; but $\delta > 0$ is arbitrary, hence $\mathsf{V}^{\mathrm{q}}_\lambda(K) \le \mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X}^*)$. The other direction $\mathsf{V}^{\mathrm{q}}_\lambda(K) \le \mathsf{s}^{\mathrm{q}}_\lambda(\mathbf{X}^*)$ is trivial from the definition of $\mathsf{V}^{\mathrm{q}}_\lambda(K)$ and the fact that $\sum_v p_*(v)K_v \preceq K$. □