AN EXTREMAL INEQUALITY RELATED TO HYPERCONTRACTIVITY OF GAUSSIAN RANDOM VARIABLES

CHANDRA NAIR[1 †]

[1]*Dept. of Information Engineering,*
The Chinese University of Hong Kong

Abstract

We establish that Gaussian distributions are the optimizers for a particular optimization problem related to determining the hypercontractivity parameters for a pair of jointly Gaussian random variables.

1. Introduction

Hypercontractive inequalities have played an important role in physics, mathematics, and theoretical computer science. In this work we will use a recently established [3] formulations of hypercontractivity parameters using information measures to prove an extremal inequality and use this to give an alternate proof of Nelson's result for scalar Gaussian random variables.

For a random variable Z, let $\|Z\|_p := \mathrm{E}(|Z|^p)^{1/p}$ for $p \geq 1$. A pair of random variables (X, Y) are said to be (p, q)-*hypercontractive*, for $1 \leq q \leq p$, if

$$\|E(g(Y)|X)\|_p \leq \|g(Y)\|_q$$

holds for every measurable function $g(Y)$.

Given $p \geq 1$, define the following related quantities:

$$\begin{aligned} q_p^*(X; Y) &:= \inf\{q : (X, Y) \text{ is } (p, q)\text{-}hypercontractive\}, \\ r_p(X; Y) &:= \frac{q_p^*(X; Y)}{p}, \\ s_p(X; Y) &:= \frac{q_p^*(X; Y) - 1}{p - 1}, p > 1. \end{aligned}$$

For finite valued random variables $(X, Y) \sim \mu_{X,Y}$ the following theorem yields an alternate expression for hypercontractive parameters in terms of information measures.

Theorem 1 (Theorem 1 in [3]). *The hypercontractive ratio $r_p(X; Y)$ is also given by any of the following expressions*

(a)

$$\sup_{\nu_{X,Y} \ll \mu_{X,Y}} \frac{D_{KL}(\nu_Y \| \mu_Y)}{p D_{KL}(\nu_{X,Y} \| \mu_{X,Y}) - (p - 1) D_{KL}(\nu_X \| \mu_X)}$$

† Corresponding author: Chandra Nair, email: chandra.nair@gmail.com

(b)

$$\sup_U \frac{I(U;Y)}{pI(U;X,Y) - (p-1)I(U;X)}$$

(c)

$$\inf\{\lambda : H(Y) - \lambda pH(X,Y) + \lambda(p-1)H(X) = \mathfrak{K}[H(Y) - \lambda pH(X,Y) + \lambda(p-1)H(X)]_\mu\},$$

where $\mathfrak{K}[f(\cdot)]_\mu$ denotes the lower convex envelope of the function $f(\cdot)$ (over joint distributions) evaluated at the joint distribution $\mu_{X,Y}$.

It is clear that the first two expressions carry over to arbitrary random variables as well using standard techniques in probability theory. Suppose $\mu_{X,Y}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$ and $[X\ Y]$ has a well-defined covariance matrix K, then it is reasonably straightforward to deduce from Theorem 1 that

$$r_p(X;Y) = \inf\{\lambda : \mathrm{h}(Y) - \lambda p\mathrm{h}(X,Y) + \lambda(p-1)\mathrm{h}(X) = \mathfrak{K}[\mathrm{h}(Y) - \lambda p\mathrm{h}(X,Y) + \lambda(p-1)\mathrm{h}(X)]_\mu\}.$$

The main result of this paper is the following theorem.

Theorem 2. *For any $p > 1$, $K \succeq 0$, $\frac{1}{p} < \lambda < 1$, there exists a $0 \preceq K' \preceq K$ and $(X',Y') \sim \mathcal{N}(0,K')$ such that for any $(X,Y) \sim \mu_{X,Y}$ with covariance matrix $K_{X,Y} \preceq K$ the following inequality holds:*

$$\mathfrak{K}[\mathrm{h}(Y) - \lambda p\mathrm{h}(X,Y) + \lambda(p-1)\mathrm{h}(X)]_\mu \geq \mathrm{h}(Y') - \lambda p\mathrm{h}(X',Y') + \lambda(p-1)\mathrm{h}(X').$$

Further if $(\hat{X},\hat{Y}) \sim \mathcal{N}(0,K)$ then equality is achieved.

Theorem 3 (with *V. Anantharam*)**.** *Let $(X,Y) \sim \mathcal{N}(0, \left[\begin{smallmatrix} 1 & \alpha \\ \alpha & 1 \end{smallmatrix}\right])$ and $\lambda = \frac{(p-1)\alpha^2+1}{p}$, then for any $0 \preceq K' \preceq \left[\begin{smallmatrix} 1 & \alpha \\ \alpha & 1 \end{smallmatrix}\right]$ and $(X',Y') \sim \mathcal{N}(0,K')$ we have*

$$\mathrm{h}(Y') - \lambda p\mathrm{h}(X',Y') + \lambda(p-1)\mathrm{h}(X') \geq \mathrm{h}(Y) - \lambda p\mathrm{h}(X,Y) + \lambda(p-1)\mathrm{h}(X).$$

Conversely, when $(X,Y) \sim \mathcal{N}(0, \left[\begin{smallmatrix} 1 & \alpha \\ \alpha & 1 \end{smallmatrix}\right])$ and $\frac{1}{p} < \lambda < \frac{(p-1)\alpha^2+1}{p}$, then there exists $K' \preceq \left[\begin{smallmatrix} 1 & \alpha \\ \alpha & 1 \end{smallmatrix}\right], K' \succeq 0$ and $(X',Y') \sim \mathcal{N}(0,K')$ such that

$$\mathrm{h}(Y') - \lambda p\mathrm{h}(X',Y') + \lambda(p-1)\mathrm{h}(X') < \mathrm{h}(Y) - \lambda p\mathrm{h}(X,Y) + \lambda(p-1)\mathrm{h}(X).$$

Thus combining Theorems 2 and 3 immediately implies that for $(X,Y) \sim \mathcal{N}(0, \left[\begin{smallmatrix} 1 & \alpha \\ \alpha & 1 \end{smallmatrix}\right])$ we jave

$$r_p(X;Y) = \frac{(p-1)\alpha^2+1}{p}$$

establishing the Gaussian hypercontractivity result.

2. Main

2.1. Proof of Theorem 2: The proof uses the techniques developed in [2] to establish the optimality of Gaussian distributions in multi-terminal information theory settings. It is immediate that

$$\mathfrak{K}[\mathrm{h}(Y) - \lambda p\mathrm{h}(X,Y) + \lambda(p-1)\mathrm{h}(X)]_\mu = \inf_U \mathrm{h}(Y|U) - \lambda p\mathrm{h}(X,Y|U) + \lambda(p-1)\mathrm{h}(X|U).$$

Consider the following minimization problem over pairs of random variables (X,Y) satisfying a covariance constraint K.

$$V(K) := \inf_{\substack{(X,Y)\sim\mu_{XY} \\ K_{XY}\preceq K}} \mathfrak{K}[\mathrm{h}(Y) - \lambda p\mathrm{h}(X,Y) + \lambda(p-1)\mathrm{h}(X)]_\mu.$$

Using techniques similar to those in [2] one can establish that the infimum is achieved; and in particular that there is a triple (U,X,Y) with $|\mathcal{U}| \leq 4$ and $K_{XY} \preceq K$ such that

$$V(K) = \mathrm{h}(Y|U) - \lambda p\mathrm{h}(X,Y|U) + \lambda(p-1)\mathrm{h}(X|U).$$

Let (U_1, X_1, Y_1) and (U_2, X_2, Y_2) be two i.i.d. copies of the minimizer (U, X, Y). Hence

$$2V(K) = \mathrm{h}(Y_1, Y_2|U_1, U_2) - \lambda p\mathrm{h}(X_1, X_2, Y_1, Y_2|U_1, U_2) + \lambda(p-1)\mathrm{h}(X_1, X_2|U_1, U_2).$$

Let $X_+ = \frac{X_1+X_2}{\sqrt{2}}$, $X_- = \frac{X_1-X_2}{\sqrt{2}}$, $Y_+ = \frac{Y_1+Y_2}{\sqrt{2}}$, and $Y_- = \frac{Y_1-Y_2}{\sqrt{2}}$. Thus we have

$$2V(K) = \mathrm{h}(Y_+, Y_-|U_1, U_2) - \lambda p\mathrm{h}(X_+, X_-, Y_+, Y_-|U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+, X_-|U_1, U_2).$$

Observe the following similar sets of manipulations.

i) Note

$$\begin{aligned}
2V(K) &= \mathrm{h}(Y_+, Y_-|U_1, U_2) - \lambda p\mathrm{h}(X_+, X_-, Y_+, Y_-|U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+, X_-|U_1, U_2)\\
&= \mathrm{h}(Y_+|U_1, U_2) - \lambda p\mathrm{h}(X_+, Y_+|U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+|U_1, U_2)\\
&\quad + \mathrm{h}(Y_-|X_+, Y_+, U_1, U_2) - \lambda p\mathrm{h}(X_-, Y_-|X_+, Y_+, U_1, U_2) + \lambda(p-1)\mathrm{h}(X_-|X_+, Y_+, U_1, U_2)\\
&\quad + I(X_+; Y_-|Y_+, U_1, U_2) + \lambda(p-1)I(Y_+; X_-|X_+, U_1, U_2)\\
&\geq V(K) + V(K) + I(X_+; Y_-|Y_+, U_1, U_2) + \lambda(p-1)I(Y_+; X_-|X_+, U_1, U_2).
\end{aligned}$$

The last inequality holds by the definition (minimality) of $V(K)$ combined with the observation that $K_{X_+Y_+} \preceq K, K_{X_-Y_-} \preceq K$. This implies

$$I(X_+; Y_-|Y_+, U_1, U_2) = I(Y_+; X_-|X_+, U_1, U_2) = 0. \tag{1}$$

Similarly interchanging the roles of (X_+, Y_+) with (X_-, Y_-) we also obtain

$$I(X_-; Y_+|Y_-, U_1, U_2) = I(Y_-; X_+|X_-, U_1, U_2) = 0. \tag{2}$$

ii) Alternately,

$$\begin{aligned}
2V(K) &= \mathrm{h}(Y_+, Y_-|U_1, U_2) - \lambda p\mathrm{h}(X_+, X_-, Y_+, Y_-|U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+, X_-|U_1, U_2)\\
&= \mathrm{h}(Y_+|U_1, U_2) - \lambda p\mathrm{h}(X_+, Y_+|U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+|U_1, U_2)\\
&\quad + \mathrm{h}(Y_-|U_1, U_2) - \lambda p\mathrm{h}(X_-, Y_-|U_1, U_2) + \lambda(p-1)\mathrm{h}(X_-|U_1, U_2)\\
&\quad - I(Y_+; Y_-|U_1, Y_2) + \lambda pI(X_+, Y_+; X_-, Y_-|U_1, U_2) - \lambda(p-1)I(X_+; X_-|U_1, U_2)\\
&\geq V(K) + V(K) - I(Y_+; Y_-|U_1, Y_2) + \lambda pI(X_+, Y_+; X_-, Y_-|U_1, U_2) - \lambda(p-1)I(X_+; X_-|U_1, U_2),
\end{aligned}$$

where the last inequality holds for the similar reason as above. This implies

$$-I(Y_+; Y_-|U_1, Y_2) + \lambda pI(X_+, Y_+; X_-, Y_-|U_1, U_2) - \lambda(p-1)I(X_+; X_-|U_1, U_2) \leq 0. \tag{3}$$

iii) A third way of decomposing yields

$$\begin{aligned}
2V(K) &= \mathrm{h}(Y_+, Y_-|U_1, U_2) - \lambda p\mathrm{h}(X_+, X_-, Y_+, Y_-|U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+, X_-|U_1, U_2)\\
&= \mathrm{h}(Y_+|X_-, Y_-, U_1, U_2) - \lambda p\mathrm{h}(X_+, Y_+|X_-, Y_-, U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+|X_-, Y_-, U_1, U_2)\\
&\quad + \mathrm{h}(Y_-|X_+, Y_+, U_1, U_2) - \lambda p\mathrm{h}(X_-, Y_-|X_+, Y_+, U_1, U_2) + \lambda(p-1)\mathrm{h}(X_-|X_+, Y_+, U_1, U_2)\\
&\quad + I(X_-, Y_-; Y_+|U_1, U_2) + I(X_+; Y_-|Y_+, U_1, U_2) - \lambda pI(X_-, Y_-; X_+, Y_+|U_1, U_2)\\
&\quad + \lambda(p-1)I(Y_+; X_-|X_+, U_1, U_2) + \lambda(p-1)I(X_-, Y_-; X_+|U_1, U_2)\\
&\geq V(K) + V(K) + I(X_-, Y_-; Y_+|U_1, U_2) + I(X_+; Y_-|Y_+, U_1, U_2) - \lambda pI(X_-, Y_-; X_+, Y_+|U_1, U_2)\\
&\quad + \lambda(p-1)I(Y_+; X_-|X_+, U_1, U_2) + \lambda(p-1)I(X_-, Y_-; X_+|U_1, U_2).
\end{aligned}$$

This implies

$$\begin{aligned}
&I(X_-, Y_-; Y_+|U_1, U_2) + I(X_+; Y_-|Y_+, U_1, U_2) - \lambda pI(X_-, Y_-; X_+, Y_+|U_1, U_2)\\
&\quad + \lambda(p-1)I(Y_+; X_-|X_+, U_1, U_2) + \lambda(p-1)I(X_-, Y_-; X_+|U_1, U_2) \leq 0.
\end{aligned}$$

Using equation (1) and (2) the above constraint reduces to

$$I(Y_+; Y_-|U_1, Y_2) - \lambda pI(X_+, Y_+; X_-, Y_-|U_1, U_2) + \lambda(p-1)I(X_+; X_-|U_1, U_2) \leq 0. \tag{4}$$

Combining (3) and (4) yields

$$I(Y_+; Y_-|U_1, Y_2) - \lambda pI(X_+, Y_+; X_-, Y_-|U_1, U_2) + \lambda(p-1)I(X_+; X_-|U_1, U_2) = 0. \tag{5}$$

iv) A fourth way of manipulation yields

$$\begin{aligned} 2V(K) &= \mathrm{h}(Y_+, Y_-|U_1, U_2) - \lambda p \mathrm{h}(X_+, X_-, Y_+, Y_-|U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+, X_-|U_1, U_2) \\ &= \mathrm{h}(Y_+|X_-, Y_-, U_1, U_2) - \lambda p \mathrm{h}(X_+, Y_+|X_-, Y_-, U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+|X_-, Y_-, U_1, U_2) \\ &\quad + \mathrm{h}(Y_-|X_+, U_1, U_2) - \lambda p \mathrm{h}(X_-, Y_-|X_+, U_1, U_2) + \lambda(p-1)\mathrm{h}(X_-|X_+, U_1, U_2) \\ &\quad + I(X_+; Y_-|U_1, U_2) + I(Y_+; X_-|Y_-, U_1, U_2) - \lambda p I(X_-, Y_-; X_+|U_1, U_2) \\ &\quad + \lambda(p-1) I(X_-, Y_-; X_+|U_1, U_2) \\ &\geq V(K) + V(K) + I(X_+; Y_-|U_1, U_2) + I(Y_+; X_-|Y_-, U_1, U_2) - \lambda I(X_-, Y_-; X_+|U_1, U_2). \end{aligned}$$

This implies

$$I(X_+; Y_-|U_1, U_2) + I(Y_+; X_-|Y_-, U_1, U_2) - \lambda I(X_-, Y_-; X_+|U_1, U_2) \leq 0.$$

Using equation (2) this reduces to

$$I(X_+; Y_-|U_1, U_2) - \lambda I(X_-; X_+|U_1, U_2) \leq 0. \tag{6}$$

Similarly interchanging the roles of (X_+, Y_+) with (X_-, Y_-) we also obtain

$$I(X_-; Y_+|U_1, U_2) - \lambda I(X_-; X_+|U_1, U_2) \leq 0. \tag{7}$$

v) Finally a fifth way of manipulation yields

$$\begin{aligned} 2V(K) &= \mathrm{h}(Y_+, Y_-|U_1, U_2) - \lambda p \mathrm{h}(X_+, X_-, Y_+, Y_-|U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+, X_-|U_1, U_2) \\ &= \mathrm{h}(Y_+|X_-, Y_-, U_1, U_2) - \lambda p \mathrm{h}(X_+, Y_+|X_-, Y_-, U_1, U_2) + \lambda(p-1)\mathrm{h}(X_+|X_-, Y_-, U_1, U_2) \\ &\quad + \mathrm{h}(Y_-|Y_+, U_1, U_2) - \lambda p \mathrm{h}(X_-, Y_-|Y_+, U_1, U_2) + \lambda(p-1)\mathrm{h}(X_-|Y_+, U_1, U_2) \\ &\quad + I(Y_+; X_-, Y_-|U_1, U_2) - \lambda p I(X_-, Y_-; Y_+|U_1, U_2) \\ &\quad + \lambda(p-1) I(Y_-; X_+|U_1, U_2) + \lambda(p-1) I(Y_-; X_+|X_-, U_1, U_2) \\ &\geq V(K) + V(K) - (\lambda p - 1) I(X_-, Y_-; Y_+|U_1, U_2) + \lambda(p-1) I(Y_-; X_+|U_1, U_2) \\ &\quad + \lambda(p-1) I(Y_-; X_+|X_-, U_1, U_2). \end{aligned}$$

This implies

$$-(\lambda p - 1) I(X_-, Y_-; Y_+|U_1, U_2) + \lambda(p-1) I(Y_-; X_+|U_1, U_2) + \lambda(p-1) I(Y_-; X_+|X_-, U_1, U_2) \leq 0.$$

Using equation (2) this reduces to

$$-(\lambda p - 1) I(Y_-; Y_+|U_1, U_2) + \lambda(p-1) I(Y_-; X_+|U_1, U_2) \leq 0. \tag{8}$$

Similarly interchanging the roles of (X_+, Y_+) with (X_-, Y_-) we also obtain

$$-(\lambda p - 1) I(Y_-; Y_+|U_1, U_2) + \lambda(p-1) I(Y_+; X_-|U_1, U_2) \leq 0. \tag{9}$$

We will now combine our observations to deduce that

$$I(X_+, Y_+; X_-, Y_-|U_1, U_2) = 0.$$

Towards this, observe that

$$\begin{aligned} 0 &\leq I(X_+; Y_-|X_-, Y_+, U_1, U_2) \\ &= I(X_+; Y_-, X_-|Y_+, U_1, U_2) - I(X_+; X_-|Y_+, U_1, U_2) \\ &= I(X_+, Y_+; Y_-, X_-|U_1, U_2) - I(Y_+; Y_-, X_-|U_1, U_2) - I(X_+, Y_+; X_-|U_1, U_2) + I(Y_+; X_-|U_1, U_2) \\ &= I(X_+, Y_+; Y_-, X_-|U_1, U_2) - I(Y_+; Y_-|U_1, U_2) - I(X_+; X_-|U_1, U_2) + I(Y_+; X_-|U_1, U_2) && \text{(Using (1), (2))} \\ &= \frac{1}{\lambda p} I(Y_+; Y_-|U_1, U_2) + \frac{(p-1)}{p} I(X_+; X_-|U_1, U_2) - I(Y_+; Y_-|U_1, U_2) && \text{(Using (5))} \\ &\quad - I(X_+; X_-|U_1, U_2) + I(Y_+; X_-|U_1, U_2) \\ &= I(Y_+; X_-|U_1, U_2) - \frac{\lambda p - 1}{\lambda p} I(Y_+; Y_-|U_1, U_2) - \frac{1}{p} I(X_+; X_-|U_1, U_2). \end{aligned}$$

We thus obtain

$$I(Y_+;X_-|U_1,U_2) \geq \frac{\lambda p-1}{\lambda p} I(Y_+;Y_-|U_1,U_2) + \frac{1}{p} I(X_+;X_-|U_1,U_2). \tag{10}$$

Combining (7) with (10) we obtain

$$\lambda I(X_+;X_-|U_1,U_2) \geq \frac{\lambda p-1}{\lambda p} I(Y_+;Y_-|U_1,U_2) + \frac{1}{p} I(X_+;X_-|U_1,U_2) \iff \lambda I(X_+;X_-|U_1,U_2) \geq I(Y_+;Y_-|U_1,U_2). \tag{11}$$

In the above simplification, we used $\lambda p > 1$.

Combining (9) with (10) we obtain

$$\begin{aligned} \frac{\lambda p-1}{\lambda(p-1)} I(Y_+;Y_-|U_1,U_2) \geq \frac{\lambda p-1}{\lambda p} I(Y_+;Y_-|U_1,U_2) + \frac{1}{p} I(X_+;X_-|U_1,U_2) \quad \iff \\ \frac{\lambda p-1}{\lambda(p-1)} I(Y_+;Y_-|U_1,U_2) \geq I(X_+;X_-|U_1,U_2). \end{aligned} \tag{12}$$

Multiplying (11) by $\frac{(\lambda p-1)}{\lambda(p-1)}$ and using (12) we obtain

$$\frac{\lambda(\lambda p-1)}{\lambda(p-1)} I(X_+;X_-|U_1,U_2) \geq I(X_+;X_-|U_1,U_2) \iff \frac{\lambda p(\lambda-1)}{\lambda(p-1)} I(X_+;X_-|U_1,U_2) \geq 0,$$

which is only possible when $I(X_+;X_-|U_1,U_2) = 0$ as $\lambda < 1$.

Substituting $I(X_+;X_-|U_1,U_2) = 0$ into (11) we obtain that $I(Y_+;Y_-|U_1,U_2) = 0$. Now using (5) we obtain that $I(X_+,Y_+;X_-,Y_-|U_1,U_2) = 0$.

Let $(X^{(u)},Y^{(u)})$ denote a random variable distributed according to the conditional law $\mu_{X,Y|U}(X,Y|U=u)$. For every $u_1,u_2 \in \mathcal{U}$ we know the following two conditions hold:

(a) (X_1,Y_1) and (X_2,Y_2) are conditionally independent given $U_1 = u_1, U_2 = u_2$ (by construction).

(b) $\left(\frac{X_1+X_2}{\sqrt{2}}, \frac{Y_1+Y_2}{\sqrt{2}}\right)$ and $\left(\frac{X_1-X_2}{\sqrt{2}}, \frac{Y_1-Y_2}{\sqrt{2}}\right)$ are conditionally independent given $U_1 = u_1, U_2 = u_2$ (This holds since $I(X_+,Y_+;X_-,Y_-|U_1,U_2) = 0$).

From the Skitovic-Darmois characterization of Gaussian distributions we obtain that $(X^{(u)},Y^{(u)}) \sim \mathcal{N}(0,K')$. There is no dependence of u for K'. (See [2] for a detailed reasoning.). Let $(X',Y') \sim \mathcal{N}(0,K')$.

This yields the first (main) part of Theorem 2, i.e.

$$V(K) = \mathrm{h}(Y') - \lambda p \mathrm{h}(X',Y') + \lambda(p-1)\mathrm{h}(X') \leq \mathfrak{K}[\mathrm{h}(Y) - \lambda p \mathrm{h}(X,Y) + \lambda(p-1)\mathrm{h}(X)]_\mu.$$

The second part is immediate by choosing $U \sim \mathcal{N}(0,K-K')$ independent of (X',Y') and observing that by setting $(X,Y) = U + (X',Y') \sim \mathcal{N}(0,K)$

$$\mathrm{h}(Y|U) - \lambda p \mathrm{h}(X,Y|U) + \lambda(p-1)\mathrm{h}(X|U) = \mathrm{h}(Y') - \lambda p \mathrm{h}(X',Y') + \lambda(p-1)\mathrm{h}(X').$$

This establishes Theorem 2. □

2.2. Proof of Theorem 3 Reparameterize K' as

$$K' = \begin{bmatrix} \sqrt{\frac{(1+\alpha)}{2}} & -\sqrt{\frac{(1-\alpha)}{2}} \\ \sqrt{\frac{(1+\alpha)}{2}} & \sqrt{\frac{(1-\alpha)}{2}} \end{bmatrix} C \begin{bmatrix} \sqrt{\frac{(1+\alpha)}{2}} & \sqrt{\frac{(1+\alpha)}{2}} \\ -\sqrt{\frac{(1-\alpha)}{2}} & \sqrt{\frac{(1-\alpha)}{2}} \end{bmatrix},$$

and note that

$$\begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{(1+\alpha)}{2}} & -\sqrt{\frac{(1-\alpha)}{2}} \\ \sqrt{\frac{(1+\alpha)}{2}} & \sqrt{\frac{(1-\alpha)}{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{(1+\alpha)}{2}} & \sqrt{\frac{(1+\alpha)}{2}} \\ -\sqrt{\frac{(1-\alpha)}{2}} & \sqrt{\frac{(1-\alpha)}{2}} \end{bmatrix}.$$

Thus $0 \preceq K' \preceq \left[\begin{smallmatrix} 1 & \alpha \\ \alpha & 1 \end{smallmatrix}\right]$ is equivalent to $0 \preceq C \preceq \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$.

Let $\beta \in (0, -\pi)$ be defined according to $\cos\left(\frac{\beta}{2}\right) = \sqrt{\frac{(1+\alpha)}{2}}$. Without loss of generality, one may express C as

$$C = \mu \begin{bmatrix} \cos\left(\theta - \frac{\beta}{2}\right) \\ -\sin\left(\theta - \frac{\beta}{2}\right) \end{bmatrix} \left[\cos\left(\theta - \frac{\beta}{2}\right) - \sin\left(\theta - \frac{\beta}{2}\right)\right] + \mu\gamma \begin{bmatrix} \sin\left(\theta - \frac{\beta}{2}\right) \\ \cos\left(\theta - \frac{\beta}{2}\right) \end{bmatrix} \left[\sin\left(\theta - \frac{\beta}{2}\right) \cos\left(\theta - \frac{\beta}{2}\right)\right]$$

for some $\theta \in [0, 2\pi), 0 \le \mu, \gamma \le 1$.

Thus we express K' as

$$K' = \mu \begin{bmatrix} \cos^2(\theta - \beta) + \gamma \sin^2(\theta - \beta) & \cos(\theta - \beta)\cos(\theta) + \gamma \sin(\theta - \beta)\sin(\theta) \\ \cos(\theta - \beta)\cos(\theta) + \gamma \sin(\theta - \beta)\sin(\theta) & \cos^2(\theta) + \gamma \sin^2(\theta) \end{bmatrix}$$

When $(X, Y) \sim \mathcal{N}(0, [\begin{smallmatrix} 1 & \alpha \\ \alpha & 1 \end{smallmatrix}])$ then

$$\begin{aligned} \mathrm{h}(Y) - \lambda p \mathrm{h}(X, Y) + \lambda(p-1)\mathrm{h}(X) &= \frac{1}{2}\log(2\pi e) - \frac{\lambda p}{2}\log(2\pi e)^2(1-\alpha^2) + \frac{\lambda(p-1)}{2}\log 2\pi e \\ &= -\frac{\lambda p}{2}\log(\sin^2\beta) - \frac{\lambda p + \lambda - 1}{2}\log(2\pi e). \end{aligned} \tag{13}$$

When $(X', Y') \sim \mathcal{N}(0, K')$ then

$$\begin{aligned} &\mathrm{h}(Y') - \lambda p \mathrm{h}(X', Y') + \lambda(p-1)\mathrm{h}(X') \\ &= \frac{1}{2}\log\left(2\pi e\mu\left(\cos^2(\theta-\beta) + \gamma\sin^2(\theta-\beta)\right)\right) - \frac{\lambda p}{2}\log\left((2\pi e)^2\mu^2\gamma\sin^2\beta\right) \\ &\quad + \frac{\lambda(p-1)}{2}\log\left(2\pi e\mu\left(\cos^2(\theta) + \gamma\sin^2(\theta)\right)\right) \\ &= -\frac{\lambda p}{2}\log(\gamma\sin^2\beta) - \frac{\lambda p + \lambda - 1}{2}\log(2\pi e\mu) + \frac{1}{2}\log\left(\cos^2(\theta-\beta) + \gamma\sin^2(\theta-\beta)\right) \\ &\quad + \frac{\lambda(p-1)}{2}\log\left(\cos^2(\theta) + \gamma\sin^2(\theta)\right) \end{aligned} \tag{14}$$

Using the representations in (13) and (14) Theorem 3 can be reformulated[1] as the following Lemma.

Lemma 1. *Let* $\lambda = \frac{p\cos^2\beta + \sin^2\beta}{p}$, *then for any* $\theta \in (0, 2\pi], 0 \le \gamma \le 1$ *we have*

$$\gamma^{\lambda p} \le \left(\cos^2(\theta-\beta) + \gamma\sin^2(\theta-\beta)\right)\left(\cos^2(\theta) + \gamma\sin^2(\theta)\right)^{\lambda(p-1)}$$

Conversely, when $\frac{1}{p} < \lambda < \frac{p\cos^2\beta + \sin^2\beta}{p}$, *then there exists* $\theta \in (0, 2\pi], 0 \le \gamma \le 1$ *such that*

$$\gamma^{\lambda p} > \left(\cos^2(\theta-\beta) + \gamma\sin^2(\theta-\beta)\right)\left(\cos^2(\theta) + \gamma\sin^2(\theta)\right)^{\lambda(p-1)}.$$

Note that we have set $\mu = 1$, which is the worst case since $-(\lambda p + \lambda - 1)\log\mu \ge 0$ and any other choice $\mu < 1$ would make the inequality (comparison between (13) and (14)) weaker.

Observe that

$$\begin{aligned} \gamma^{\lambda p} &\le \left(\cos^2(\theta-\beta) + \gamma\sin^2(\theta-\beta)\right)\left(\cos^2(\theta) + \gamma\sin^2(\theta)\right)^{\lambda(p-1)} \quad \Longleftrightarrow \\ \left(\frac{1}{\gamma}\right)^{1-\lambda} &\le \left(\frac{1}{\gamma}\cos^2(\theta-\beta) + \sin^2(\theta-\beta)\right)\left(\frac{1}{\gamma}\cos^2(\theta) + \sin^2(\theta)\right)^{\lambda(p-1)}. \end{aligned} \tag{15}$$

Concavity of the logarithm function and the Jensen's inequality implies

$$\begin{aligned} \frac{1}{\gamma}\cos^2(\theta-\beta) + \sin^2(\theta-\beta) &\ge \left(\frac{1}{\gamma}\right)^{\cos^2(\theta-\beta)}, \\ \frac{1}{\gamma}\cos^2(\theta) + \sin^2(\theta) &\ge \left(\frac{1}{\gamma}\right)^{\cos^2(\theta)} \end{aligned}$$

[1] The reparamterization and subsequent reforumulation is due to V. Anantharam.

Thus (15) holds whenever

$$1 - \lambda \leq \cos^2(\theta - \beta) + \lambda(p-1)\cos^2\theta. \tag{16}$$

Substituting $\lambda = \frac{p\cos^2\beta + \sin^2\beta}{p}$ we would like to show the validity of (16), i.e. show equivalently that

$$\sin^2(\theta - \beta) \leq \left(\frac{p\cos^2\beta + \sin^2\beta}{p}\right)(p\cos^2\theta + \sin^2\theta).$$

The above inequality can be rearranged as

$$\left(\sqrt{p}\cos\theta\cos\beta - \frac{1}{\sqrt{p}}\sin\theta\sin\beta\right)^2 \geq 0, \tag{17}$$

which holds. This proves the first part of Lemma 1.

To show the second part, take θ to satisfy $\tan\theta = p\cot\beta$, and take $\gamma = 1 - \epsilon$. We expand both sides of (15) in ϵ for $\epsilon > 0$ and observe that if

$$(1 - \lambda) > \cos^2(\theta - \beta) + \lambda(p-1)\cos^2\theta \tag{18}$$

then for $\epsilon > 0$ small enough and $\gamma = 1 - \epsilon$ we would have

$$\left(\frac{1}{\gamma}\right)^{1-\lambda} > \left(\frac{1}{\gamma}\cos^2(\theta - \beta) + \sin^2(\theta - \beta)\right)\left(\frac{1}{\gamma}\cos^2(\theta) + \sin^2(\theta)\right)^{\lambda(p-1)}$$

as desired. (In other words equality holds in (15) at $\gamma = 1$ and we are comparing the derivates as $\gamma \uparrow 1$).

At $\lambda = \frac{p\cos^2\beta + \sin^2\beta}{p}$ we know that equality holds in (18) using (16), (17), and our choice $\tan\theta = p\cot\beta$. Hence when $\lambda < \frac{p\cos^2\beta + \sin^2\beta}{p}$ the inequality is strict in (18) as desired. This establishes Lemma 1. □

3. Conclusion

In this manuscript we provide an alternate proof of the characterization of $r_p(X; Y)$ for Gaussian random variables. This work uses the technique developed in [2] for showing the optimality of Gaussian distributions. The novel element in this work is the method used to deduce the conditional independence of (X_+, Y_+) and $X_-, Y_-)$ by combining various ways of single-letterizing a two-letter expression. This could be potentially very useful in several multi-terminal information theory situations.

Historical background and Acknowledgements

This work is directly motivated by the work of the author with Venkat Anantharam, Amin Gohari, Sudeep Kamath in [1] which deals with the equivalent version of Theorem 1 when $p \to \infty$. Motivated by this work, the author then generalized the equivalence for finite p and obtained the result in Theorem 1. The author wishes to deeply acknowledge numerical experiments conducted by Sudeep Kamath which prompted the proof of Theorem 1. This work is an attempt to gauge the usefulness of the alternate characterization by trying to rederive existing results in the new framework. The author was struggling with the proof of Theorem 3 and Venkat Anantharam greatly simplified the statement by his reformulation to Lemma 1. The author also had sought help from Amin Gohari in the proof of Theorem 3, though their efforts did not pan out. Thus this work owes a deep rooted acknowledgement to Venkat Anantharam, Amin Gohari, and Sudeep Kamath.

This work was supported by an AoE grant (Project No. AoE/E-02/08) and two GRF grants (Project Nos. 415810 and 415612) from the UGC of the Hong Kong Special Administrative Region, China.

References

[1] Venkat Anantharam, Amin Aminzadeh Gohari, Sudeep Kamath, and Chandra Nair, On maximal correlation, hypercontractivity, and the data processing inequality studied by erkip and cover, CoRR **abs/1304.6133** (2013).

[2] Yanlin Geng and Chandra Nair, The capacity region of the two-receiver vector gaussian broadcast channel with private and common messages, IEEE Trans. Inf. Theory (accepted) (2014).

[3] C Nair, Equivalent formulations of hypercontractivity using information measures, (Extended Abstract) to be presented at International Zurich Seminar (2014).