Stability of Bernstein's Characterization of Gaussian Vectors and a Soft Doubling Argument

Mohammad Mahdi Mahvari and Gerhard Kramer
Technical University of Munich, Germany
Emails: {mahdi.mahvari, gerhard.kramer}@tum.de

***Abstract*—Stability properties of Bernstein's characterization of Gaussian vectors are derived. Stability leads to a soft doubling argument through which one can prove capacity theorems without requiring the existence of capacity-achieving distributions.**

I. Introduction

The following characterization of vector Gaussian distributions builds on work of Kac [1] and Bernstein [2] and is a special case of the main results in [3]–[5].

Theorem 1. *Consider the independent[1] d-dimensional random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$. If $\mathbf{X}_1 + \mathbf{X}_2$ and $\mathbf{X}_1 - \mathbf{X}_2$ are also independent then $\mathbf{X}_1$ and $\mathbf{X}_2$ are Gaussian with the same covariance matrix.*

We refer to Theorem 1 is as Bernstein's theorem. The result has been used to establish the optimality of Gaussian functions and Gaussian random vectors for several inequalities, including inequalities with applications to reliable communications over channels with additive Gaussian noise (AGN). In the following, we review applications of Bernstein's theorem and motivate stability theorems.

A. Applications of Bernstein's Theorem

Lieb [6] used the "$O(2)$ rotation invariance of products of centered Gaussians" to show that Gaussian functions achieve equality in the generalized Brascamp-Lieb inequality [7]. Lieb's method is closely related to Bernstein's theorem as it considers products of a function with the vector arguments $(\mathbf{x}_1 + \mathbf{x}_2)/\sqrt{2}$ and $(\mathbf{x}_1 - \mathbf{x}_2)/\sqrt{2}$. A special case of the generalized Brascamp-Lieb inequality is Young's inequality which is met with equality by Gaussian functions. Carlen [8] used Lieb's technique to prove that Gaussian functions achieve equality in the logarithmic Sobolev inequalities; he refers to the technique as a "doubling trick".

More recently, Theorem 1 was applied to communications problems by establishing that Gaussian random vectors are optimal for certain functional and extremal inequalities [9]–[16]. For instance, the doubling trick was used to characterize the capacity region, or capacity points, of vector Gaussian broadcast channels [17]–[20], multiaccess channels with feedback [21], relay channels [22], Z-interference channels [23], [24], Gray-Wyner networks [25], source coding with a causal helper [26], and two-way wiretap channels [27].

B. Motivation

A key step in applying Bernstein's theorem to communications problems is establishing the existence of distributions achieving rate tuples on the boundaries of capacity regions. One motivation for this paper is to investigate the necessity of this step. After all, in practice one can only approach rather than achieve the capacity of noisy channels, so requiring a capacity supremum to be a maximum is more of mathematical than engineering relevance. Moreover, if a maximizing distribution exists and is continuous then stability is important because practical modulation alphabets are finite.

A second motivation for this paper is that the existence proof in [17] requires several technical theorems on the convergence of sequences of distributions, including Prokhorov's theorem, the converse of the Scheffé-Riesz theorem, Lévy's continuity theorem, and a theorem of Godavarti-Hero [28]. We instead wish to have a proof that requires only basic theory, and that considers individual distributions rather than sequences of distributions.

A third motivation is a mathematical one, namely to extend the stability of Bernstein's theorem from scalars to vectors. For scalars, this stability is based on the stability of the Cauchy functional equation for bi-infinite and finite intervals; treating vectors requires extensions to multivariate functions. We also use stability results for vector differential entropies based on individual distributions [29].

C. Stability of Bernstein's Theorem

Several stability results are described in [30, Sec. 4] with different assumptions on the sums and differences of independent X_1 and X_2. For example, in [30, Thm. 4.4] the random variables X_1 and X_2 are said to be ϵ-dependent[2] if

$$\sup_{x_1, x_2 \in \mathbb{R}} |F_{X_1,X_2}(x_1,x_2) - F_{X_1}(x_1)F_{X_2}(x_2)| < \epsilon. \quad (1)$$

Define the Gaussian c.d.f.s $\mathcal{F}_i$ with means m_i and variances σ_i^2, $i = 1, 2$. Let $\mathrm{E}[X]$ and $\mathrm{Var}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right]$ denote the expectation and variance of X.

[1]By "independent" we mean statistical independence as opposed to, e.g., linear independence.

[2]This property is called ϵ-independent in [30, Sec. 2] but it seems more natural to identify independence with 0-dependence rather than 0-independence.

Now suppose $X_1 + X_2$ and $X_1 - X_2$ are ϵ-dependent for some ϵ satisfying $0 < \epsilon < 1$, and that $\mathrm{E}\left[|X_i|^{2(1+\delta)}\right] < \infty$ for $i = 1, 2$ and some δ satisfying $0 < \delta \leq 1$. Then [30, Thm. 4.4] states that

$$\sup_{x \in \mathbb{R}} |F_{X_i}(x) - \mathcal{F}_i(x)| < c\,(-\ln \epsilon)^{-1/2} \tag{2}$$

for $i = 1, 2$, where c is independent of ϵ, $m_i = \mathrm{E}[X_i]$, and $\sigma_i^2 = (\mathrm{Var}[X_1] + \mathrm{Var}[X_2])/2$.

The discussion in [30] describes several other stability metrics, including the Lévy metric [31]. We instead follow [32], [33] (see also [34]) and consider a metric in the characteristic function (c.f.) domain. Let $j = \sqrt{-1}$ and let

$$f_X(t) = \mathrm{E}\left[e^{jtX}\right],\; f_{X_1,X_2}(t_1,t_2) = \mathrm{E}\left[e^{jt_1X_1 + jt_2X_2}\right] \tag{3}$$

be the c.f.s of X and (X_1, X_2), respectively. For example, the c.f. of a Gaussian X with mean m and variance σ^2 is

$$\Phi(t) = e^{jmt - \frac{1}{2}\sigma^2 t^2}, \quad t \in \mathbb{R}. \tag{4}$$

X_1 and X_2 are said to be ϵ-dependent in the c.f. domain if

$$\sup_{t_1,t_2 \in \mathbb{R}} |f_{X_1,X_2}(t_1,t_2) - f_{X_1}(t_1) f_{X_2}(t_2)| \leq \epsilon. \tag{5}$$

The paper [33] develops the following stability theorem. Let $\mathcal{P}_\epsilon$ be the class of (X_1, X_2) for which X_1 and X_2 are independent and $X_1 + X_2$ and $X_1 - X_2$ are ϵ-dependent in the c.f. domain. Then we have (see [33, Thm. 1])

$$c_1\epsilon \leq \sup_{(X_1,X_2)\in\mathcal{P}_\epsilon} \max_{i=1,2} \left(\sup_{t\in\mathbb{R}} |f_{X_i}(t) - \Phi_i(t)| \right) \leq c_2\epsilon \tag{6}$$

for Gaussian c.f.s Φ_i, $i = 1, 2$, where c_1 and c_2 are positive constants independent of ϵ. The bounds (6) imply that the scaling proportional to ϵ is the best possible in general.

D. Multivariate Stability

Gabovič [35] established stability for a vector form of the Darmois-Skitovič theorem [3], [4] that generalizes Bernstein's theorem. However, there are several differences to our result. Perhaps the most important is that the random vectors must satisfy a special condition "to prevent the 'leakage' of a significant probabilistic mass to infinity" [35, p. 5]; this restriction seems to prevent Gabovič's stability from truly generalizing Bernstein's theorem. Also, there is no claim of an identical covariance matrix for $\mathbf{X}_1$ and $\mathbf{X}_2$.

Thus, there are important advantages of studying stability in the c.f. domain. First, one need not *a-priori* exclude certain random vectors. Second, one can prove a common covariance matrix for sufficiently small ϵ, i.e., the stability theory generalizes Theorem 1. Third, convergence is proportional to ϵ which is the best possible scaling, see (6) and Theorems 3 and 4 below. Finally, one can relate ϵ-dependence to mutual information by using Pinsker's inequality, see Lemma 2 below.

E. Organization

This paper is organized as follows. Sec. II develops notation and reviews properties of multivariate c.f.s. Sec. III states and proves our main stability theorems (Theorems 3-5). Sec. IV develops "soft" versions of the "hard" doubling arguments in [17] for point-to-point channels, and outlines extensions to product channels and two-receiver broadcast channels. Sec. V concludes the paper. Due to space constraints, most proofs are given in a separate version of this paper [36].

II. Preliminaries

A. Basic Notation

The p-norm for d-dimensional vectors is written as

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{d} |x_i|^p \right)^{1/p} \tag{7}$$

and we write $\|\mathbf{x}\|_\infty = \max_{1\leq i\leq d} |x_i|$. We usually consider the 1-norm that we write as $\|\mathbf{x}\| = \|\mathbf{x}\|_1$. For complex-valued functions on $\mathbb{R}^d$ we write

$$\|f\|_p = \left(\int_{\mathbb{R}^d} |f(\mathbf{t})|^p \, d\mathbf{t} \right)^{1/p}. \tag{8}$$

For a square matrix $\mathbf{Q}$, we write $\det \mathbf{Q}$ for the determinant of $\mathbf{Q}$, and $\mathbf{Q}' \preceq \mathbf{Q}$ if $\mathbf{Q} - \mathbf{Q}'$ is positive semi-definite. The $d \times d$ identity matrix is written as $\mathbf{I}_d$. The vector with zero entries except for a 1 in entry i is written as $\mathbf{e}_i$.

The distribution, c.d.f., mean, and covariance matrix of $\mathbf{X}$ are written as $P_\mathbf{X}$, $F_\mathbf{X}$, $\mathbf{m}_\mathbf{X} = \mathrm{E}[\mathbf{X}]$, and

$$\mathbf{Q}_\mathbf{X} = \mathrm{E}\left[(\mathbf{X} - \mathbf{m}_\mathbf{X})(\mathbf{X} - \mathbf{m}_\mathbf{X})^T\right] \tag{9}$$

respectively, where $\mathbf{t}^T$ is the transpose of $\mathbf{t}$. The distribution $P_\mathbf{X}$ is absolutely continuous (a.c.) with respect to the Lebesgue measure if and only if a probability density function (p.d.f.) exists that we write as $p_\mathbf{X}$.

The notation $h(p)$, $h(\mathbf{X})$, $I(\mathbf{X}; \mathbf{Y})$, and $D(p\|q)$ refers to the differential entropy of the p.d.f. p, the differential entropy of $\mathbf{X}$, the mutual information of $\mathbf{X}$ and $\mathbf{Y}$, and the informational divergence of the p.d.f.s p and q, respectively. We often discard subscripts on probability distributions and other functions for notational convenience.

B. Multivariate Characteristic Functions

The characteristic function (c.f.) of the d-dimensional real-valued $\mathbf{X}$ evaluated at $\mathbf{t} \in \mathbb{R}^d$ is

$$f_\mathbf{X}(\mathbf{t}) = \mathrm{E}\left[e^{j\mathbf{t}^T\mathbf{X}}\right]. \tag{10}$$

If the p.d.f. $p_\mathbf{X}$ exists then $p_\mathbf{X}$ and $f_\mathbf{X}$ can be interpreted as a Fourier transform pair. The c.f. of the pair $\mathbf{X}_1, \mathbf{X}_2$ evaluated at $\mathbf{t}_1, \mathbf{t}_2$ is

$$f_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{t}_1,\mathbf{t}_2) = \mathrm{E}\left[e^{j\mathbf{t}_1^T\mathbf{X}_1 + j\mathbf{t}_2^T\mathbf{X}_2}\right] \tag{11}$$

Note that $\mathbf{X}_1$ and $\mathbf{X}_2$ need not have the same dimension.

Four properties of c.f.s are as follows, see [37, p. 55]: $f_\mathbf{X}(\mathbf{0}) = 1$; $|f_\mathbf{X}(\mathbf{t})| \leq 1$; $f_\mathbf{X}(-\mathbf{t}) = f_\mathbf{X}(\mathbf{t})^*$ where x^* is the complex conjugate of x; $f_{\mathbf{X}}$ is uniformly continuous and therefore non-vanishing in a region around $\mathbf{t}=\mathbf{0}$.

We state two properties of pairs of random vectors. First, $\mathbf{X}_1$ and $\mathbf{X}_2$ are statistically independent if and only if $f_{\mathbf{X}_1,\mathbf{X}_2}$ factors as $f_{\mathbf{X}_1}f_{\mathbf{X}_2}$. Second, we state a lemma that relates the c.f.s of pairs of random vectors and their mutual information. We also define two versions of ϵ-dependence.

Lemma 2. *For random vectors $\mathbf{X}_1, \mathbf{X}_2$ with joint p.d.f. $p_{\mathbf{X}_1,\mathbf{X}_2}$ and for all $\mathbf{t}_1 \in \mathbb{R}^{d_1}$ and $\mathbf{t}_2 \in \mathbb{R}^{d_2}$ we have*

$$\begin{aligned} &|f_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{t}_1,\mathbf{t}_2) - f_{\mathbf{X}_1}(\mathbf{t}_1)f_{\mathbf{X}_2}(\mathbf{t}_2)| \\ &\quad \le \|p_{\mathbf{X}_1,\mathbf{X}_2} - p_{\mathbf{X}_1}p_{\mathbf{X}_2}\| \\ &\quad \le \sqrt{2I(\mathbf{X}_1;\mathbf{X}_2)} \end{aligned} \tag{12}$$

where the mutual information is measured in nats.

Definition 1. *Let ϵ and T be non-negative constants. The random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ are* (ϵ, T)-dependent *if*

$$\sup_{\|\mathbf{t}_1\|\le T, \|\mathbf{t}_2\|\le T} |f_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{t}_1,\mathbf{t}_2) - f_{\mathbf{X}_1}(\mathbf{t}_1)f_{\mathbf{X}_2}(\mathbf{t}_2)| \le \epsilon. \tag{13}$$

Similarly, $\mathbf{X}_1$ and $\mathbf{X}_2$ are ϵ-dependent *if they are (ϵ, T)-dependent for all non-negative T.*

The ϵ-dependence of Definition 1 can be interpreted as (ϵ,∞)-dependence. Also, $\mathbf{X}_1$ and $\mathbf{X}_2$ are 0-dependent (or $(0,\infty)$-dependent) if and only if they are independent.

C. Gaussian Vectors

We write $\mathbf{X} \sim \mathcal{N}(\mathbf{m}_{\mathbf{X}}, \mathbf{Q}_{\mathbf{X}})$ if $\mathbf{X}$ is Gaussian with mean $\mathbf{m}_{\mathbf{X}}$ and covariance matrix $\mathbf{Q}_{\mathbf{X}}$, i.e., the p.d.f of $\mathbf{X}$ is

$$\phi_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\det(2\pi\mathbf{Q}_{\mathbf{X}})^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_{\mathbf{X}})^T\mathbf{Q}_{\mathbf{X}}^{-1}(\mathbf{x}-\mathbf{m}_{\mathbf{X}})} \tag{14}$$

where we assumed that $\mathbf{Q}_{\mathbf{X}}$ is invertible. More generally, the Gaussian c.f. is

$$\Phi_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^T\left(j\mathbf{m}_{\mathbf{X}} - \frac{1}{2}\mathbf{Q}_{\mathbf{X}}\,\mathbf{t}\right)} \tag{15}$$

and we have $|\Phi_{\mathbf{X}}(\mathbf{t})| = 1$ if and only if $\mathbf{t}$ lies in the null space of $\mathbf{Q}_{\mathbf{X}}$. Otherwise, $|\Phi_{\mathbf{X}}(c\cdot\mathbf{t})|$ strictly decreases from 1 to 0 as c increases from $c=0$ to $c=\infty$.

III. Stability Theorems

This section states three stability theorems for d-dimensional random vectors. Theorem 3 considers the local stability for a finite interval around $\mathbf{t}=\mathbf{0}$. Theorem 4 extends Bernstein's Theorem to include local stability, and extends the scalar theory in [33] to vectors. We emphasize that both theorems have a common covariance matrix $\widehat{\mathbf{Q}}$ which is not the case in [33] but is important to generalize Theorem 1 and to develop further results for product channels and for broadcast channels. Theorem 5 gives two stability results: one for differential entropy and one for correlation matrices. Again, due to space constraints, proofs are provided in the separate paper [36].

Theorem 3. *Suppose $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent random vectors, and $\mathbf{X}_1+\mathbf{X}_2$ and $\mathbf{X}_1-\mathbf{X}_2$ are (ϵ,T)-dependent. Also, suppose there is a constant $p>0$ such that*

$$|f_i(\mathbf{t})| \ge p \quad \text{for } \|\mathbf{t}\| \le T \text{ and } i=1,2. \tag{16}$$

Then for $0<\epsilon\le p^4/[360d^2(d+1)]$ and $\|\mathbf{t}\|\le T/2$ we have

$$|f_i(\mathbf{t}) - \Phi_i(\mathbf{t})| \le C(\epsilon)\cdot|\Phi_i(\mathbf{t})|, \quad i=1,2 \tag{17}$$

where for some mean vectors $\widehat{\mathbf{m}}_i$, $i=1,2$, and for some common covariance matrix $\widehat{\mathbf{Q}}$ we have the Gaussian c.f.s

$$\Phi_i(\mathbf{t}) = e^{\mathbf{t}^T\left(j\widehat{\mathbf{m}}_i - \frac{1}{2}\widehat{\mathbf{Q}}\,\mathbf{t}\right)}, \quad i=1,2 \tag{18}$$

and the error term is

$$C(\epsilon) = \frac{720d^2(d+1)}{p^4}\,\epsilon. \tag{19}$$

Theorem 4. *Suppose $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent random vectors, and $\mathbf{X}_1+\mathbf{X}_2$ and $\mathbf{X}_1-\mathbf{X}_2$ are ϵ-dependent. Then for all ϵ below some positive threshold, for all $\mathbf{t}\in\mathbb{R}^d$, and for $i=1,2$ we have*

$$|f_i(\mathbf{t}) - \Phi_i(\mathbf{t})| \le \tilde{C}\epsilon \tag{20}$$

for the Gaussian c.f.s (18), and for a constant $\tilde{C}$ independent of ϵ and $\mathbf{t}$. In particular, if $\epsilon=0$ then $\mathbf{X}_1$ and $\mathbf{X}_2$ are Gaussian with the same covariance matrix.

Theorem 5. *Consider the random vectors $\mathbf{Y}_1=\mathbf{X}_1+\mathbf{Z}_1$ and $\mathbf{Y}_2=\mathbf{X}_2+\mathbf{Z}_2$ where $\mathbf{X}_1,\mathbf{X}_2,\mathbf{Z}_1,\mathbf{Z}_2$ are mutually independent, $\mathbf{Y}_1,\mathbf{Y}_2$ have finite second moments, and the noise vectors $\mathbf{Z}_i\sim\mathcal{N}(\mathbf{0},\mathbf{Q}_{\mathbf{Z}_i})$, $i=1,2$, are non-degenerate. Suppose $\mathbf{Y}_1+\mathbf{Y}_2$ and $\mathbf{Y}_1-\mathbf{Y}_2$ are ϵ-dependent. Then for all ϵ below some positive threshold and for $i=1,2$ we have*

$$|h(\mathbf{Y}_i) - h(\mathbf{Y}_{g,i})| \le B(\epsilon) \tag{21}$$

where the $\mathbf{Y}_{g,i}$ are Gaussian with the same covariance matrix, and thus $h(\mathbf{Y}_{g,1}) = h(\mathbf{Y}_{g,2})$, and

$$\mathrm{E}\left[\mathbf{Y}_{g,i}\mathbf{Y}_{g,i}^T\right] \preceq \mathrm{E}\left[\mathbf{Y}_i\mathbf{Y}_i^T\right] + B(\epsilon)\,\mathbf{I}_d. \tag{22}$$

where $B(\epsilon)\to 0$ as $\epsilon\to 0$ and $B(\epsilon)=0$ if $\epsilon=0$.

IV. Soft Doubling for AGN Channels

This section shows how to combine the stability of Bernstein's theorem with the doubling argument in [17] to obtain a soft doubling argument that does not require the existence of maximizers. We consider point-to-point channels, product channels, and broadcast channels with AGN that have applications to cellular wireless networks [38].

A. Point-to-Point Channels

An AGN channel has output

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z} \tag{23}$$

where $\mathbf{X},\mathbf{Y},\mathbf{Z}\in\mathbb{R}^d$ and $\mathbf{Z}\sim\mathcal{N}(\mathbf{0},\mathbf{Q}_{\mathbf{Z}})$ is independent of $\mathbf{X}$. Consider the optimization problem:

$$V(\mathbf{Q}) := \sup_{\mathbf{X}:\,\mathrm{E}[\mathbf{X}\mathbf{X}^T]\preceq\mathbf{Q}} I(\mathbf{X};\mathbf{Y}). \tag{24}$$

We use Theorem 5 and a soft doubling argument to prove the following known result.

Proposition 6. *For the AGN channel (23) we have*

$$I(\mathbf{X};\mathbf{Y}) \le \frac{1}{2}\log\frac{\det(\mathbf{Q_X}+\mathbf{Q_Z})}{\det\mathbf{Q_Z}} \tag{25}$$

with equality if $\mathbf{X}$ *is Gaussian.*

Proof. Equality holds in (25) if $\mathbf{X}$ is Gaussian so it remains to prove the inequality. Note that we may assume $\mathrm{E}[\mathbf{X}]=\mathbf{0}$ because $I(\mathbf{X};\mathbf{Y})$ does not depend on translation of $\mathbf{X}$.

Now consider $\mathbf{Y}_1=\mathbf{X}_1+\mathbf{Z}_1$ and $\mathbf{Y}_2=\mathbf{X}_2+\mathbf{Z}_2$, where $\mathbf{X}_1,\mathbf{X}_2\sim P_\mathbf{X}$ and $\mathbf{Z}_1,\mathbf{Z}_2\sim P_\mathbf{Z}$ are mutually independent. Further define

$$\mathbf{X}_+ := \frac{1}{\sqrt{2}}(\mathbf{X}_1+\mathbf{X}_2), \quad \mathbf{X}_- := \frac{1}{\sqrt{2}}(\mathbf{X}_1-\mathbf{X}_2). \tag{26}$$

Note that $\mathrm{E}\left[\mathbf{X}_+\mathbf{X}_+^T\right]\preceq\mathbf{Q}$ and $\mathrm{E}\left[\mathbf{X}_-\mathbf{X}_-^T\right]\preceq\mathbf{Q}$. Also, define

$$\mathbf{Y}_+ := \frac{1}{\sqrt{2}}(\mathbf{Y}_1+\mathbf{Y}_2), \quad \mathbf{Y}_- := \frac{1}{\sqrt{2}}(\mathbf{Y}_1-\mathbf{Y}_2) \tag{27}$$

$$\mathbf{Z}_+ := \frac{1}{\sqrt{2}}(\mathbf{Z}_1+\mathbf{Z}_2), \quad \mathbf{Z}_- := \frac{1}{\sqrt{2}}(\mathbf{Z}_1-\mathbf{Z}_2) \tag{28}$$

so that $\mathbf{Y}_+=\mathbf{X}_++\mathbf{Z}_+$ and $\mathbf{Y}_-=\mathbf{X}_-+\mathbf{Z}_-$ where the noise vectors $\mathbf{Z}_+,\mathbf{Z}_-\sim P_\mathbf{Z}$ are independent.

Now suppose $\epsilon>0$ and

$$I(\mathbf{X};\mathbf{Y}) = V(\mathbf{Q_X})-\epsilon. \tag{29}$$

We then have

$$\begin{aligned}
2V(\mathbf{Q_X}) &= I(\mathbf{X}_1;\mathbf{Y}_1)+I(\mathbf{X}_2;\mathbf{Y}_2)+2\epsilon \\
&= I(\mathbf{X}_1\mathbf{X}_2;\mathbf{Y}_1\mathbf{Y}_2)+2\epsilon \\
&= I(\mathbf{X}_+\mathbf{X}_-;\mathbf{Y}_+\mathbf{Y}_-)+2\epsilon \\
&\overset{(a)}{=} \underbrace{I(\mathbf{X}_+;\mathbf{Y}_+)}_{\le V(\mathbf{Q_X})}+\underbrace{I(\mathbf{X}_-;\mathbf{Y}_-)}_{\le V(\mathbf{Q_X})}-I(\mathbf{Y}_+;\mathbf{Y}_-)+2\epsilon \\
&\le 2V(\mathbf{Q_X})-I(\mathbf{Y}_+;\mathbf{Y}_-)+2\epsilon
\end{aligned} \tag{30}$$

where step (a) follows by

$$p(\mathbf{y}_+,\mathbf{y}_-|\mathbf{x}_+,\mathbf{x}_-)=p(\mathbf{y}_+|\mathbf{x}_+)\,p(\mathbf{y}_-|\mathbf{x}_-)$$

for all $\mathbf{x}_+,\mathbf{x}_-,\mathbf{y}_+,\mathbf{y}_-$. Lemma 2 and (30) give

$$\begin{aligned}
&\left|f_{\mathbf{Y}_+,\mathbf{Y}_-}(\mathbf{t}_1,\mathbf{t}_2)-f_{\mathbf{Y}_+}(\mathbf{t}_1)f_{\mathbf{Y}_-}(\mathbf{t}_2)\right| \\
&\quad\le\sqrt{2I(\mathbf{Y}_+;\mathbf{Y}_-)}\le 2\sqrt{\epsilon}
\end{aligned} \tag{31}$$

so $\mathbf{Y}_+$ and $\mathbf{Y}_-$ are $(2\sqrt{\epsilon})$-dependent.

For sufficiently small ϵ, Theorem 5 gives

$$\begin{aligned}
I(\mathbf{X};\mathbf{Y}) &= h(\mathbf{Y}_1)-h(\mathbf{Z}) \\
&\le h(\mathbf{Y}_{g,1})-\frac{1}{2}\log\det(2\pi\,\mathbf{Q_Z})+B\left(2\sqrt{\epsilon}\right) \\
&= \frac{1}{2}\log\frac{\det\mathbf{Q}_{\mathbf{Y}_{g,1}}}{\det\mathbf{Q_Z}}+B\left(2\sqrt{\epsilon}\right)
\end{aligned} \tag{32}$$

where for some mean vector $\mathbf{m}_{g,1}$ we have

$$\begin{aligned}
\mathbf{Q}_{\mathbf{Y}_{g,1}} &= \mathrm{E}\left[\mathbf{Y}_{g,1}\mathbf{Y}_{g,1}^T\right]-\mathbf{m}_{g,1}\mathbf{m}_{g,1}^T \\
&\preceq \mathrm{E}\left[\mathbf{Y}_1\mathbf{Y}_1^T\right]+B\left(2\sqrt{\epsilon}\right)\mathbf{I}_d \\
&\overset{(a)}{=} \mathbf{Q_X}+\mathbf{Q_Z}+B\left(2\sqrt{\epsilon}\right)\mathbf{I}_d
\end{aligned} \tag{33}$$

and step (a) follows by $\mathrm{E}[\mathbf{X}]=\mathbf{0}$. Moreover, the ϵ in (29) can be chosen as close to zero as desired because $V(\mathbf{Q_X})$ is a supremum. Finally, observe that if $I(\mathbf{X};\mathbf{Y})\le J+\epsilon$ for all $\epsilon>0$ then $I(\mathbf{X};\mathbf{Y})\le J$. □

B. Product Channels and Broadcast Channels

The proof of Proposition 6 uses an AGN product channel with two outputs $\mathbf{Y}_1$ and $\mathbf{Y}_2$ where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ have the same covariance matrix. More generally, suppose the covariance matrices are different. Since the noise is non-degenerate, this channel is equivalent to the AGN product channel considered in [17, Sec. I.A], namely[3]

$$\mathbf{Y}_{11} = \mathbf{G}_1\mathbf{X}_1+\mathbf{Z}_{11} \tag{34}$$
$$\mathbf{Y}_{22} = \mathbf{G}_2\mathbf{X}_2+\mathbf{Z}_{22} \tag{35}$$

where $\mathbf{G}_1=\mathbf{Q}_{\mathbf{Z}_1}^{-1/2}$, $\mathbf{G}_2=\mathbf{Q}_{\mathbf{Z}_2}^{-1/2}$, the noise vectors $\mathbf{Z}_{11},\mathbf{Z}_{22}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)$ are independent, and $(\mathbf{X}_1,\mathbf{X}_2)$ is independent of $(\mathbf{Z}_{11},\mathbf{Z}_{22})$.

For the channel (34)-(35) we prove several results that let us treat the two-receiver AGN broadcast channel with

$$\mathbf{Y}_1 = \mathbf{G}_1\mathbf{X}+\mathbf{Z}_1 \tag{36}$$
$$\mathbf{Y}_2 = \mathbf{G}_2\mathbf{X}+\mathbf{Z}_2 \tag{37}$$

where $\mathbf{G}_1,\mathbf{G}_2$ are invertible, and $\mathbf{Z}_1,\mathbf{Z}_2\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)$ are independent. Note that the input $\mathbf{X}$ is common to both sub-channels, and that $\mathbf{Z}_1,\mathbf{Z}_2$ have the same covariance matrix.

Define the expressions (see [17])

$$s_\lambda(\mathbf{X}) := I(\mathbf{X};\mathbf{Y}_1)-\lambda I(\mathbf{X};\mathbf{Y}_2) \tag{38}$$
$$s_\lambda(\mathbf{X}|\mathbf{V}) := I(\mathbf{X};\mathbf{Y}_1|\mathbf{V})-\lambda I(\mathbf{X};\mathbf{Y}_2|\mathbf{V}) \tag{39}$$
$$S_\lambda(\mathbf{X}) := \sup_{p(\mathbf{v}|\mathbf{x}):\,\mathbf{V}-\mathbf{X}-\mathbf{Y}_1\mathbf{Y}_2} s_\lambda(\mathbf{X}|\mathbf{V}) \tag{40}$$
$$V_\lambda(\mathbf{Q}) := \sup_{\mathbf{X}:\,\mathrm{E}[\mathbf{X}\mathbf{X}^T]\preceq\mathbf{Q}} S_\lambda(\mathbf{X}) \tag{41}$$

where $S_\lambda(\mathbf{X})$ is the upper concave envelope of $s_\lambda(\mathbf{X})$ as a function of $p(\mathbf{x})$.

The longer version [36] of this paper now re-proves a key result from [39] which states that a Gaussian $\mathbf{X}$ is optimal for the problem (41) and one does not require $\mathbf{V}$. This theorem was also re-proved in [17, Thm. 1] through a series of propositions. Our proof follows similar steps but we do not require the existence of a maximizer, and we must replace the independence result [17, Prop. 2] with other steps.

Theorem 7 (See [39, Thm. 8])**.** *If* $\lambda>1$ *then we have* $V_\lambda(\mathbf{Q})=s_\lambda(\mathbf{X}_g)$ *for some* $\mathbf{X}_g\sim\mathcal{N}(\mathbf{0},\widehat{\mathbf{Q}})$ *with* $\widehat{\mathbf{Q}}\preceq\mathbf{Q}$.

Theorem 7 can be used to show that Gaussian signaling is optimal for two-receiver broadcast channels with dedicated

[3] We replace the $\mathbf{G}_{11},\mathbf{G}_{22}$ in [17, Sec. I.A] with $\mathbf{G}_1,\mathbf{G}_2$.

(also called private) messages for each receiver, see [39, Sec. IV.A] and [17, Sec. III.A]. Moreover, the method described above extends to two-receiver broadcast channels with a common message since the proof of Theorem 2 in [17] applies similar steps.

We remark that [17, Sec. II.B] and Theorem 7 treat the case $\lambda > 1$ while [39, Thm. 8] includes $\lambda = 1$. However, as pointed out in [17, Remark 9], the case $\lambda = 1$ can be treated by showing that a capacity function is convex and bounded while the case $\lambda < 1$ can be treated by reversing the roles of $\mathbf{Y}_1$ and $\mathbf{Y}_2$.

V. Conclusions

The stability of Bernstein's characterization of Gaussian distributions was extended to vectors. The theory led to a soft doubling argument that establishes the optimality of Gaussian vectors for point-to-point channels with AGN.

It seems that soft doubling can replace hard doubling in general. However, whether soft doubling can provide new inequalities and capacity theorems that hard doubling cannot remains to be seen. For example, if a communications model has a strict cost constraint such as $\mathrm{E}\left[\|\mathbf{X}\|^2\right] < P$ then one can turn to stability rather than, e.g., relaxing the constraint to $\mathrm{E}\left[\|\mathbf{X}\|^2\right] \leq P$, proving the existence of a maximizer (if possible) and applying Theorem 1. In this sense, stability seems to be more flexible than requiring the existence of maximizing distributions, just as suprema are more flexible than maxima.

VI. Acknowledgments

This work was supported by the 6G Future Lab Bavaria and by the German Research Foundation (DFG).

References

[1] M. Kac, "On a characterization of the normal distribution," *American J. Math.*, vol. 61, no. 3, pp. 726–728, 1939.

[2] S. N. Bernshtein, "On a property characterizing Gauss' law," *Trudy Leningr. Bolitekhn.*, vol. 217, no. 3, pp. 21–22, 1941.

[3] G. Darmois, "Analyse générale des liaisons stochastiques," *Rev. Inst. Intern. Stat.*, vol. 21, pp. 2–8, 1953.

[4] V. P. Skitovic, "On a property of the normal distribution," *Dokl. Akad. Nauk SSSR*, vol. 89, pp. 217–219, 1953.

[5] S. G. Ghurye and I. Olkin, "A characterization of the multivariate normal distribution," *Ann. Math. Stat.*, vol. 33, no. 2, pp. 533–541, 1962.

[6] E. H Lieb, "Gaussian kernels have only Gaussian maximizers," *Invent. Math.*, vol. 102, pp. 179–208, Dec. 1990.

[7] H. J. Brascamp and E. H Lieb, "Best constants in Young's inequality, its converse, and its generalization to more than three functions," *Advances Mathem.*, vol. 20, no. 2, pp. 151–173, 1976.

[8] E. A. Carlen, "Superadditivity of Fisher's information and logarithmic Sobolev inequalities," *J. Functional Analysis*, vol. 101, no. 1, pp. 194–211, 1991.

[9] C. Nair, "An extremal inequality related to hypercontractivity of Gaussian random variables," in *Inf. Theory Applic. Workshop*, 2014.

[10] T. A. Courtade and J. Jiao, "An extremal inequality for long Markov chains," in *Allerton Conf. Commun., Control, Computing*, 2014, pp. 763–770.

[11] T. A. Courtade, "A strong entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2173–2192, 2017.

[12] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Information-theoretic perspectives on Brascamp-Lieb inequality and its reverse," *arXiv preprint arXiv:1702.06260*, 2017.

[13] J. Liu, *Information Theory from a Functional Viewpoint*, Ph.D. thesis, Princeton University, 2018.

[14] J. Liu, T. A. Courtade, P. W. Cuff, and S. Verdú, "A forward-reverse Brascamp-Lieb inequality: Entropic duality and Gaussian optimality," *Entropy*, vol. 20, no. 6, pp. 418, 2018.

[15] V. Anantharam, V. Jog, and C. Nair, "Unifying the Brascamp-Lieb inequality and the entropy power inequality," *IEEE Trans. Inf. Theory*, pp. 1–1, 2022.

[16] Efe A. and T. A. Courtade, "Entropy inequalities and Gaussian comparisons," *arXiv preprint arXiv:2206.14182*, 2022.

[17] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel With private and common messages," *IEEE Tran. Inf. Theory*, vol. 60, no. 4, pp. 2087–2104, 2014.

[18] H.-F. Chong and Y.-C. Liang, "On the capacity region of the parallel degraded broadcast channel with three receivers and three-degraded message sets," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5017–5041, 2016.

[19] V. Ramachandran and S. R. B. Pillai, "Feedback-capacity of degraded Gaussian vector BC using directed information and concave envelopes," in *Nat. Conf. Commun.*, 2017, pp. 1–6.

[20] Z. Goldfeld and H. H. Permuter, "MIMO Gaussian broadcast channels with common, private, and confidential messages," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2525–2544, 2019.

[21] E. Sula, M. Gastpar, and G. Kramer, "Sum-rate capacity for symmetric Gaussian multiple access channels with feedback," *IEEE Trans. Inf. Theory*, vol. 66, no. 5, pp. 2860–2871, May 2020.

[22] A. El Gamal, A. Gohari, and C. Nair, "A strengthened cutset upper bound on the capacity of the relay channel and applications," *IEEE Trans. Inf. Theory*, 2022.

[23] A. Gohari, C. Nair, and D. Ng, "An information inequality motivated by the Gaussian Z-interference channel," in *IEEE Int. Symp. Inf. Theory*, 2021, pp. 2744–2749.

[24] M. Costa, C. Nair, D. Ng, and Y. N. Wang, "On the structure of certain non-convex functionals and the Gaussian Z-interference channel," in *IEEE Int. Symp. Inf. Theory*, 2020, pp. 1522–1527.

[25] M. Gastpar and E. Sula, "Relaxed Wyner's common information," in *IEEE Inf. Theory Workshop*, 2019, pp. 1–5.

[26] Shraga I. Bross, "Source coding with a causal helper," *Information*, vol. 11, no. 12, pp. 553, 2020.

[27] Z. Wang, *Secret Key Agreement over Gaussian Two-way Wiretap Channel*, Ph.D. thesis, Univ. Toronto, 2021.

[28] M. Godavarti and A. Hero, "Convergence of differential entropies," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 171–176, 2004.

[29] H. Ghourchian, A. Gohari, and A. Amini, "Existence and continuity of differential entropy for a class of distributions," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1469–1472, 2017.

[30] E. Lukacs, "Stability theorems," *Advances in Appl. Probab.*, vol. 9, no. 2, pp. 336–361, 1977.

[31] P. Lévy, *Théorie de l'Addition des Variables Aléatoires*, Gauthier-Villars, Paris, 1937.

[32] L. Klebanov and R. Yanushkyavichyus, "ε-dependence of $X_1 + X_2$ and $X_1 - X_2$," *Litovskii Matemat. Sbornik (Lietuvos Matemat. Rinkinys)*, vol. 25, no. 3, pp. 83–92, 1985, English translation: Inst. Mathem. Cybernet., Academy Sci. Lithuanian SSR, pp. 236-242, 1985.

[33] L. B. Klebanov and R. V. Yanushkyavichyus, "Estimation of stability in S. N. Bernshtein's Theorem," *Theory Probab. & Its Appl.*, vol. 30, no. 2, pp. 383–386, 1986.

[34] R. Yanushkevichius, "On stability of the B. Gnedenko characterization," *J. Math. Sci.*, vol. 92, no. 3, pp. 3960–3971, 1998.

[35] Y. R. Gabovic, "The stability of the characterization of the normal distribution by the Skitovic-Darmois theorem," *Zap. Naucn. Sem. Leningrad Otdel. Mat. Inst. Steklov.*, vol. 61, 1976.

[36] M. M. Mahvari and G. Kramer, "Stability of Bernstein's theorem and soft doubling for vector Gaussian channels," *arXiv preprint arXiv:2212.04484*, 2022.

[37] N. G. Ushakov, *Selected Topics in Characteristic Functions*, VSP, Utrecht, The Netherlands, 1999.

[38] Y.-H. Kim and G. Kramer, "Information theory for cellular wireless networks," in *Information Theoretic Perspectives on 5G Systems and Beyond*, pp. 10–92. Cambridge Univ. Press, 2022.

[39] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839–1851, May 2007.