At Every Corner:
Determining Corner Points of Two-User Gaussian Interference Channels

Olivier Rioul
LTCI CNRS, Télécom ParisTech, Université Paris-Saclay, 75013 Paris, France
Email: olivier.rioul@telecom-paristech.fr

Abstract—The corner points of the capacity region of the two-user Gaussian interference channel under strong or weak interference are determined using the notions of almost Gaussian random vectors, almost lossless addition of random vectors, and almost linearly dependent random vectors. In particular, the “missing” corner point problem is solved in a manner that differs from previous works in that it avoids the use of integration over a continuum of SNR values or of Monge-Kantorovich transportation problems.

I. INTRODUCTION

This work is about the complete determination of corner points of the capacity region of the two-user Gaussian interference channel. Some classical ingredients are Fano’s inequality, the data processing inequality (DPI), the maximum entropy (MaxEnt) property under a power constraint, the entropy power inequality (EPI), and the concavity of the entropy power. Interestingly, only weak forms of the latter two are required. To these ingredients we add the notions of almost Gaussian random vectors, almost lossless addition of random vectors, and almost linearly dependent random vectors.

In particular, the determination of the second corner point under weak interference is known as the Costa’s corner point conjecture. This conjecture has been settled recently and independently by Polyanskiy and Wu [1] (using optimal transport theory) and Bustin et al. [2, 3] (using the I-MMSE relation). The approach described here is a natural continuation from previous work [4–7] that is very close in spirit to the solution of Polyanskiy and Wu but can be thought as more direct as it avoids the use of integration over a continuum of SNR values or of Monge-Kantorovich transportation problems.

II. DEFINITIONS AND NOTATIONS

Throughout the paper we consider zero-mean random vectors taking values in \( \mathbb{R}^n \) and let \( \| \cdot \| \) denote the Euclidean norm in \( \mathbb{R}^n \). Consider the two-user Gaussian interference channel in standard form:

\[
\begin{align*}
Y_1 &= X_1 + \sqrt{b} X_2 + Z_1, \\
Y_2 &= \sqrt{a} X_1 + X_2 + Z_2,
\end{align*}
\]

where the joint distribution of the Gaussian noises \( (Z_1, Z_2) \) at the decoder sides is not relevant as there is no cooperation between the receivers. We find it notionally convenient to set \( Z_1 = Z_2 = Z \). The corresponding noise powers are \( N_1 = b N_2 = N \). Sender \( i \) for \( i = 1, 2 \) produces a uniformly distributed \( M_i \)-ary message \( W_i \), where \( W_1 \) and \( W_2 \) are independent. Encoder \( i \) maps \( W_i \) to a random vector \( X_i \in \mathbb{R}^n \) of dimension \( n \) which satisfies the power constraint \( \|X_i\|^2 \leq n P_i \). Decoder \( i \) maps the output \( Y_i \) to an \( M_i \)-ary decoded message \( \hat{W}_i \).

The capacity region of the channel may be defined as the set of all limit points of all sequences \( (R_1, R_2) \) for which the corresponding sequence of encoding and decoding functions with \( M_i = e^{n R_i} \) are such that \( \mathbb{P}\{\hat{W}_i \neq W_i\} \) tend to 0 as \( n \to +\infty \). Note that \( R_1, R_2, W_1, W_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2 \) all depend on the dimension \( n \). However, \( P_1, P_2 \) are constants, independent of \( n \). Because \( n \) is taken arbitrarily large, it is convenient to use the following notation.

Definition 1 (Almost Inequalities \( \lesssim \) and \( \gtrsim \)). Let \( \epsilon(n) \) denote any positive function of \( n \) which tends to 0 as \( n \to +\infty \) (thus we can write, for example, \( \epsilon(n) + \epsilon(n) = \epsilon(n) \)). Given real number sequences \( A_n, B_n \), we write \( A_n \lesssim B_n \) (\( A_n \) is almost less than \( B_n \)) if

\[
A_n \leq B_n + n \epsilon(n) \iff B_n \geq A_n - n \epsilon(n).
\]

We also write \( B_n \gtrsim A_n \) (\( B_n \) is almost greater than \( A_n \)).

The capacity region is a subset of the rectangle \( R_1 \leq C_1, R_2 \leq C_2 \), where \( C_i = (1/2) \log(1 + P_i/N) \) with two corner points \( (C_1, C_2) \) and \( (C_1', C_2) \). That \( (C_1', C_2) \) is a corner point is established by showing that it is achievable and that for any \( (R_1, R_2) \) for which the associated probability of error tends to 0 as \( n \to +\infty \),

\[
n R_1 \gtrsim n C_1 \implies n R_2 \lesssim n C_2'.
\]

That \( (C_1', C_2) \) is a corner point is similarly characterized by:

\[
n R_2 \gtrsim n C_2 \implies n R_1 \lesssim n C_1'.
\]

Achievability is generally not a problem and is done using classical ingredients such as random coding, onion peeling and rate splitting. Therefore, in this paper, we focus exclusively on the derivation of the converse [3].

III. PRELIMINARIES

Throughout the paper \( X \sim \mathcal{N} \) denotes a white Gaussian vector of the same variance as \( X \).

Lemma 1. The condition \( n R_1 \gtrsim n C_1 \) in (3a) implies

(a) \( h(X_1 + Z) \gtrsim h(X_1^2 + Z) \);
(b) \( I(X_1; Y_1) \gtrsim I(X_1; X_1 + Z) \).

The symmetrical lemma holds for (3b).
Proof: By the classical derivation of the converse:
\[
nR_1 = H(W_1) \lesssim I(W_1; Y_1) \quad \text{(Fano) (4a)}
\]
\[
\lesssim I(X_1; Y_1) \quad \text{(DPI) (4b)}
\]
\[
\lesssim I(X_1; X_1 + Z) \quad \text{(DPI again) (4c)}
\]
\[
= h(X_1 + Z) - h(Z) \quad \text{(4d)}
\]
\[
\leq nC_1 \quad \text{(MaxEnt) (4e)}
\]
Thus \(nR_1 \gtrsim nC_1\) amounts to saying that all quantities in (4) are at distance \(\leq n\epsilon\). This implies, in particular, (a) from (4c) and (b) from (4e).

Remark 1. Condition \(nR_1 \gtrsim nC_1\) also implies \(I(W_1; Y_1) \gtrsim I(X_1; Y_1)\) which holds (with equality) if the encoder mapping is invertible. In that case \(nR_1 \gtrsim nC_1 \iff (a),(b)\).

Lemma 2 (AG and AL properties). Let \(X\) have power constraint \(\frac{1}{2}E[\|X\|^2] \leq P\). We say that \(X\) is almost (white) Gaussian (AG) if
\[
h(X) \gtrsim h(X^G).\] (5)
Let \(Z\) and \(Z'\) be mutually independent (not necessarily Gaussian) vectors, independent of \(X\). We say that \(X + Z + Z'\) is almost lossless (AL) compared to \(X + Z\) (with respect to \(X\)) if
\[
I(X; X + Z + Z') \gtrsim I(X; X + Z).\] (6)
Thus (a), (b) in Lemma 1 are equivalent to:
(a) \(X_1 + Z\) is AG;
(b) \(X_1 + \sqrt{b}X_2 + Z\) is AL compared to \(X_1 + Z\) w.r.t. \(X_1\).
The latter condition means that adding interference \(bX_2\) in \(Y_1\) almost does not decrease information. This becomes vacuous in the case of no interference (\(b = 0\)). If \(b \neq 0\), condition (b) is equivalent to:
(b') \(X_1 + \sqrt{b}X_2 + Z\) is AL compared to \(\sqrt{b}X_2 + Z\) w.r.t. \(X_2\).
This is a direct consequence of the following lemma, which is particularly important as it allows one to pass from one transmission to the other.

Lemma 2 (Fork Lemma). Let \(X_1, X_2\) and \(Z\) be independent. If \(X_1 + X_2 + Z\) is AL compared to \(X_1 + Z\) w.r.t. \(X_1\), then it is also AL compared to \(X_2 + Z\) w.r.t. \(X_2\).
Proof: \(I(X_2; X_1 + X_2 + Z) - I(X_2; X_2 + Z) = h(X_1 + X_2 + Z) - h(X_1 + Z) - h(X_2 + Z) + h(Z) = I(X_1; X_1 + X_2 + Z) - I(X_1; X_1 + Z)\).
To simplify the derivations in the remainder of the paper, we restrict ourselves the case of a Gaussian \(Z\)-interference channel with one of the interference parameters (e.g., \(b\)) equal to zero:
\[
Y_1 = X_1 + Z
\]
\[
Y_2 = X_2 + \sqrt{b}X_1 + Z.\] (7)
The general determination of corner points will follow in the general case of two-sided interference by noting that removing an interference link can only enlarge the capacity region, as explained in [8] Table I.

IV. Corner Points Under Strong Interference

The very strong interference case \((a \geq 1 + P_2/N)\) is well-known [9]. One has \((C'_1 = C_1, C'_2 = C_2)\) and in this case there is no need to prove (3). For strong interference \((1 \leq a \leq 1 + P_2/N)\) the corner points are known and given by (3) below. The usual derivation follows from that of the capacity region of the multiple access channel and from the result of Han and Kobayashi [10] and Sato [11], who showed that both receivers should be able to decode both messages \(W_1\) and \(W_2\).

Lemma 3. Let \(X_i = \sqrt{t}X\) and \(Z\) be Gaussian independent of \(X\). Then \(I(X; X_i + Z), or h(X_i + Z), is nondecreasing in t\).
Proof: Let \(u = \frac{1}{\sqrt{t}}, Z_u = \sqrt{t}Z\) so that \(I(X; X_u + Z) = I(X; X + Z_u)\) and let \(Z'\) be an independent copy of \(Z\). By the DPI and the divisibility property of the Gaussian, \(\forall \delta > 0, I(X; X + Z_u) \geq I(X; X + Z_u + Z'_\delta) = I(X; X + Z_{u+\delta})\).

Proposition 1. For the strong \(Z\)-interference Gaussian channel,
\[
C'_1 = \frac{1}{2} \log \left(1 + \frac{aP_1 + P_2}{N} \right) - C_2 \quad \text{(8a)}
\]
\[
= \frac{1}{2} \log \left(1 + \frac{aP_1}{P_1 + P_2} \right)\] (8b)
\[
C'_2 = \frac{1}{2} \log \left(1 + \frac{aP_1 + P_2}{N} \right) - C_1
\]
\[
= \frac{1}{2} \log \left(1 + \frac{(a-1)P_1 + P_2}{P_1 + N} \right).\] (8c)

Proof of Proposition 1 First suppose that \(nR_1 \gtrsim nC_1\). From Lemma 1 \(X_1 + Z\) is AG. Therefore, from (4a)–(4b) where index 1 is replaced by 2,
\[
nR_2 \lesssim I(X_2; Y_2) \quad \text{(9a)}
\]
\[
= h(Y_2) - h(\sqrt{a}X_1 + Z) \quad \text{(9b)}
\]
\[
\leq h(Y_2) - h(X_1 + Z) \quad \text{(Lemma 3) (9c)}
\]
\[
\lesssim h(Y_2) - h(Z) - nC_1 \quad \text{(AG) (9d)}
\]
\[
\leq nC'_2 \quad \text{(MaxEnt) (9e)}
\]
which proves that \(nR_2 \lesssim nC'_2\) (cf. (3c)).

Next suppose that \(nR_2 \gtrsim nC_2\). From Lemma 1 written for transmission 2, \(X_2 + Z\) is AG and \(aX_1 + X_2 + Z\) is AL compared to \(X_2 + Z\) w.r.t. \(X_2\). Since \(a \neq 0\), by Lemma 2 \(aX_1 + X_2 + Z\) is AL compared to \(aX_1 + Z\) w.r.t. \(X_1\). Therefore, from (4a)–(4b),
\[
nR_1 \lesssim I(X_1; Y_1) = I(X_1; X_1 + Z) \quad \text{(10a)}
\]
\[
\lesssim I(X_1; \sqrt{a}X_1 + Z) \quad \text{(Lemma 3) (10b)}
\]
\[
\lesssim I(X_1; \sqrt{a}X_1 + X_2 + Z) \quad \text{(AL) (10c)}
\]
\[
= h(Y_2) - h(X_2 + Z) \quad \text{(10d)}
\]
\[
\lesssim h(Y_2) - h(Z) - nC_2 \quad \text{(AG) (10e)}
\]
\[
\leq nC'_1 \quad \text{(MaxEnt) (10f)}
\]
which proves that \(nR_1 \leq nC'_1\) (cf. (3b)).
V. SATO’S CORNER POINT

For weak interference \(a < 1\), Sato [12] has found that the first corner point is given by (11) below. The usual derivation follows from the equivalence between Gaussian \(Z\)-interference channel and a “fully” degraded version proved in [8], the fact that it can be considered as a broadcast channel with input power given by \(P_1 + P_2\) [12], and the derivation of the capacity region of the Gaussian (degraded) broadcast channel by Bergmans [13]. We give a simple proof based on the following lemma which is an direct consequence of the EPI.

Lemma 4. Let \(X_1 = \sqrt{t}X\) and \(Z\) be Gaussian independent of \(X\). If \(X + Z\) is AG then so is \(X_1 + Z\) for any \(0 < t < 1\).

Proof: Let \(u = 1/t > 1\), \(Z_u = \sqrt{t}Z\) and let \(Z'\) be an independent copy of \(Z\). By the DFI for divergence and the divisibility property of the Gaussian, \(h(X_1^2 + Z) - h \left(X + Z \right) = h(X^2 + Z) - h \left(X + Z \right) = (X^2 + Z) - h \left(X + Z \right) \leq D \left( X + Z \right)\) \(\leq h(X^2 + Z) - h(X + Z)\).

Remark 2. By noting that \(X\) is AG if and only if its entropy power \(N(X)\) satisfies \(N(X) \geq N(X) - \epsilon(n)\), it is readily seen that the general EPI \(N(X + Y) \geq N(X) + N(Y)\) for independent \(X, Y\) implies that if \(X\) and \(Y\) are AG, then so is \(X + Y\). Thus the conclusion of Lemma 4 is also obtained using the EPI where one of the variables is Gaussian: \(N(X + Z) \geq N(X) + N(Z)\).

It is interesting to note, however, that the EPI is not even required: only the DFI applied to divergence was necessary in the above proof, which is strictly weaker than the EPI. In fact, \(D \left( X + Z \right) \leq D \left( X \right)\) is equivalent to \(N(X + Z) \geq N(X) + N(Z)\) \(\left(N(X)/N(X^2)\right)\) where \(N(X)/N(X^2) \leq 1\).

Proposition 2. For the weak \(Z\)-interference Gaussian channel,

\[
C_2' = \frac{1}{2} \log \left(1 + \frac{P_2}{aP_1 + N}\right).
\]  

Proof: Suppose that \(nR_1 \geq nC_1\). From Proposition 1 \(X_1 + Z\) is AG. By Lemma 4 \(\sqrt{t}X_1 + Z\) is also AG. Therefore, from (4a)–(4b) written for \(i = 2\),

\[
\begin{align*}
nR_2 &\leq I(2; X_2) = h(Y_2) - h(\sqrt{t}X_1 + Z) \quad \text{(AG)} \quad \text{(12a)} \ \text{AG}\ \text{(12b)} \\
&\leq nC_2' \quad \text{(MaxEnt)} \quad \text{(12c)}
\end{align*}
\]

which proves that \(nR_2 \leq nC_2'\) (cf. (3a)).

VI. ALMOST LINEAR DEPENDENCE

For any two (zero-mean) \(n\)-dimensional random vectors \(U, V\) with finite average powers we define their correlation coefficient by

\[
\rho(U, V) = \frac{\mathbb{E}\{U \cdot V\}}{\sqrt{\mathbb{E}\{U^2\}} \sqrt{\mathbb{E}\{V^2\}}}
\]  

where \(\cdot\) denotes the scalar product. By Cauchy-Schwarz inequality one has \(\rho(U, V) \leq 1\) with equality if and only if \(U\) and \(V\) are linearly dependent in the sense that \(U = \lambda V\) a.e. for some \(\lambda \in \mathbb{R}\).

Definition 3 (ALD property). We say that \(U\) and \(V\) are almost linearly dependent (ALD) if

\[
1 - |\rho(U, V)| \leq \epsilon(n).
\]  

(Recall that \(\epsilon(n)\) denotes any positive function of \(n\) which tends to \(0\) as \(n \to +\infty\).)

We now consider \(Y = X_2 + Z\) of variance \(Q \leq P_2 + N\) and the interference term \(X = \sqrt{n}X_1\).

Remark 3. Since \(Z\) is Gaussian, it is proven in [12, App. II.A] that \(Y = X_2 + Z\) has a continuous density (see also [13] Lemma 1). Similarly \(X + Y = (\sqrt{n}X_1 + X_2) + Z\) also has a continuous density. In contrast, \(X\) is proportional to a code distribution that is typically discrete.

Clearly \(Y^G = X_2 + Z\) satisfies the inequality \(h(Y) \leq h(Y^G)\). However, the interference term \(X\) might very well be such that the opposite inequality \(h(X + Y) \geq h(X + Y^G)\) holds after addition. We now aim at bounding the difference \(h(X + Y) - h(X + Y^G)\).

Lemma 5. One has

\[
\log \left( h(X + Y) - h(X + Y^G) \right) < c \cdot n \cdot \sqrt{1 - \rho(Y, Y^G)}
\]  

where \(c\) is a constant (independent of \(n\)).

Proof: The continuous p.d.f. \(q\) of \(X + Y\) takes the form

\[
q(u) = \mathbb{E}\{q(u|X)\} = \frac{\mathbb{E}\exp \left( -\sqrt{2}Q \right)}{2Q^{n/2}}.
\]  

Since \(D(X + Y \| X + Y^G) \geq 0\), we have

\[
\log \frac{q(\tilde{u})}{q(u)} = \log \frac{\mathbb{E}\exp \left( -\sqrt{2}Q \right)}{2Q^{n/2}}
\]  

Now for any \(u \in \mathbb{R}^n\), \(\|u - X\|^2 - \|\tilde{u} - X\|^2 = (\|u\|^2 - \|\tilde{u}\|^2 + 2X \cdot (\tilde{u} - u) \leq |\|u\|^2 - \|\tilde{u}\|^2 + 2\sqrt{anP_1} \|\tilde{u} - u\|\). It follows that

\[
\log \frac{q(\tilde{u})}{q(u)} \leq \frac{|\|u\|^2 - \|\tilde{u}\|^2 + 2\sqrt{anP_1} \|\tilde{u} - u\|}{2Q}
\]  

1This particular instance of Cauchy-Schwarz inequality can be proved by considering the discriminant of the nonnegative quadratic form \(\lambda \rightarrow \mathbb{E}\{U^2 + \lambda V^2\}\). Alternatively, one has \(\mathbb{E}(U^2 Y) \leq \sum_{i=1}^n \mathbb{E}(U^2 V_i) \leq \sqrt{\mathbb{E}\{U^2\}} \sqrt{\mathbb{E}\{V_i^2\}} \leq \sqrt{\mathbb{E}\{U^2\}} \mathbb{E}\{V_i^2\}\) where the Cauchy-Schwarz inequality is applied twice (for random variables and for vectors).

2In fact, the density of \(Y\) is indefinitely differentiable, bounded, positive, tends to zero at infinity and all its derivatives are also bounded and tend to zero at infinity [14, App. B]; but we shall not need this result here.
where the identical terms $\mathbb{E}\exp(-\|\bar{u} - X\|^2/2Q)$ in the numerator and denominator were cancelled. Plugging this inequality into (17) and noting that $X + Y^G = (X + Y) = Y^G - Y$ we obtain

$$
\begin{align*}
  h(X + Y) - h(X + Y^G) & \leq \mathbb{E}[\|X + Y\|^2]/2Q - \mathbb{E}[\|X + Y^G\|^2]/2Q \\
 & + \frac{\sqrt{\det P_1}}{Q} \sqrt{\mathbb{E}[\|Y\|^2] + \mathbb{E}[\|Y^G\|^2] - 2 \mathbb{E}[Y \cdot Y^G]} \\
  &= n \sqrt{2a_1 P_1/Q} \cdot \sqrt{1 - \rho(Y, Y^G)}
\end{align*}
$$

(20)

where the first two terms in $\mathbb{E}[Y \cdot Y^G]$ were cancelled.

The result of Lemma 5 shows that if $Y$ and $Y^G$ are ALD such that $1 - \rho(Y, Y^G) \leq \epsilon(n)$, then $h(Y^G) - h(X + Y)$ is almost positive: it can be negative, but not by much. In order to obtain a value $\rho(Y, Y^G)$ close to one, the next lemma shows that it is sufficient to assume a dependence of the form $Y = F(Y^G)$ where $F$ is “almost linear”.

**Lemma 6.** One can always assume that $Y = F(Y^G)$ where the change of variable $F$ has a triangular Jacobian matrix $J$ with positive diagonal elements such that

$$
\rho(Y, Y^G) = \frac{1}{n} \mathbb{E}[\text{Tr}(J)] \geq 0.
$$

(22)

Of course, a truly linear dependence of the form $Y = \lambda Y^G$ implies $\lambda = 1$ (since $Y$ and $Y^G$ have the same variance), hence $J = I$ (identity matrix), in keeping with the fact that $\rho(Y, Y^G) = 1$ in this case.

**Proof:** The change of variable of this lemma is well known as Knüdhe map in the theory of convex bodies [17, p. 126], [18, p. 312], [19, Thm. 3.4], [20, Thm. 1.3.1]. For completeness we give Knüdhe’s proof [21]. By Remark 3, $Y$ has a continuous density. For each $y^G \in \mathbb{R}$, define $F_1(y^G)$ such that

$$
\int_{-\infty}^{F_1(y^G)} p_Y(y) \, dy = \int_{-\infty}^{y^G} p_{Y^G}(y) \, dy.
$$

(23)

Clearly $F_1$ is increasing and differentiating gives

$$
p_{Y^G}(F_1(y^G)) \frac{\partial F_1}{\partial y^G}(y^G) = p_{Y^G}(y^G)
$$

(24)

which proves the result in one dimension: $Y$ has the same distribution as $F_1(Y^G)$ where the Jacobian is positive. Next for each $y^G, y^G_2$, in $\mathbb{R}$, define $F_2(y^G, y_2^G)$ such that

$$
\int_{-\infty}^{F_2(y^G, y_2^G)} p_{Y_1, Y_2}(F_1(y_1^G), y_2^G) \, dy_1 = \int_{-\infty}^{y_2^G} p_{Y^G_1, Y^G_2}(y_1^G, y_2^G) \, dy_1.
$$

(25)

Again $F_2$ is increasing in $y_2^G$ and differentiating gives

$$
p_{Y_1, Y_2}(F_1(y_1^G), F_2(y_1^G, y_2^G)) \frac{\partial F_1}{\partial y_1^G}(y_1^G) \frac{\partial F_2}{\partial y_2^G}(y_1^G, y_2^G) = p_{Y_1^G, Y_2^G}(y_1^G, y_2^G).
$$

(26)

Continuing in this manner we arrive at

$$
p_{Y_1, Y_2, \ldots, Y_n}(F_1(y_1^G), F_2(y_1^G, y_2^G), \ldots, F_n(y_1^G, y_2^G, \ldots, y_n^G)) \times \frac{\partial F_1}{\partial y_1^G}(y_1^G) \frac{\partial F_2}{\partial y_2^G}(y_1^G, y_2^G) \cdots \frac{\partial F_n}{\partial y_n^G}(y_1^G, y_2^G, \ldots, y_n^G)
$$

$$
= p_{Y_1^G, Y_2^G, \ldots, Y_n^G}(y_1^G, y_2^G, \ldots, y_n^G)
$$

(27)

which shows that $Y$ has the same distribution as $F(Y^G) = (F_1(Y^G), F_2(Y^G, Y_2^G), \ldots, F_n(Y^G, Y_2^G, \ldots, Y_n^G))$. The Jacobian matrix $J$ of $F$ is triangular with positive diagonal elements are positive since by construction each $F_k$ is increasing in $y_k^G$. For convenience we choose to define $(Y, Y^G)$ such that $Y = F(Y^G)$. By Stein’s lemma,

$$
\rho(Y, Y^G) = \frac{1}{nQ} \sum_{i=1}^{n} \mathbb{E}(Y_i^G, F_i(Y^G))
$$

(28)

$$
= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(\frac{\partial F_i}{\partial y_i^G}(Y)) = \frac{1}{n} \mathbb{E}(\text{Tr}(J)).
$$

(29)

**Proposition 3.** If $Y$ is AG, then $Y$ and $Y^G$ are ALD and

$$
I(X; X + Y^G) \geq I(X; X + Y).
$$

(30)

The latter equation also reads, with our previous notations,

$$
I(X; \sqrt{a}X_1 + X_2^G + Z) \geq I(X_1; \sqrt{a}X_1 + X_2 + Z).
$$

(31)

**Proof:** By making the change of variable in the expression of $Y = F(Y^G)$ one obtains

$$
\rho(Y) = h(F(Y^G)) = h(Y^G) + \mathbb{E}\log \text{det } J
$$

(32)

Thus, since $Y$ is AG, $\mathbb{E}\log \text{det } J \geq 0$. On the other hand from (22) by Hadamard’s inequality,

$$
\rho(Y, Y^G) = \frac{1}{n} \mathbb{E}(\text{Tr}(J)) \geq \mathbb{E}\sqrt{\text{det } J}
$$

(33)

$$
\geq \frac{1}{n} \mathbb{E}\log \text{det } J
$$

(34)

which shows that $Y$ and $Y^G$ are ALD, such that $1 - \rho(Y, Y^G) \leq \epsilon(n)$. From Lemma 5 it follows that $h(X + Y) - h(X + Y^G) \leq 0$, hence $I(X; X + Y) = h(X + Y) - h(Y) \leq h(X + Y^G) - h(Y^G) = I(X; X + Y^G)$.}

VII. THE “MISSING” CORNER POINT

For weak interference $a < 1$, Costa [8] has stated that the second corner point is given by (20) below. A problematic issue in the proof was detected by Sason [22] and the corner point has been later dubbed “missing” [23]. Recently, Polyanskiy and Wu [11] solved the missing corner point problem using optimal transport theory by showing Lipschitz continuity of differential entropy with respect to the Wasserstein distance and Talagrand’s transportation-information inequality. An independent solution using the I-MMSE approach was given by Bustin et al. [2, 3] for a restricted subset of inputs—and later more generally—by integration of the MMSE over a continuum of SNR values. We provide yet another solution to the problem in continuation of previous investigations [4–7].
that is close to Polyanisky and Wu’s but sidesteps the use of the Wasserstein distance. Our proof is based on Prop. 4 and the following lemma.

Lemma 7. Let $Z$ be Gaussian independent of $X$ and write $Z_u = \sqrt{u}Z$. For any positive $u < u' < u''$, there exists a constant independent of $u$ such that

$$I(X; X + Z_u) - I(X; X + Z_u') \geq \mu \cdot (I(X; X + Z_u') - I(X; X + Z_{u''}))$$  \hspace{1cm} (35)

Consequently, $I(X; X + Z_{u''}) \geq I(X; X + Z_u)$ implies $I(X; X + Z_u) \geq I(X; X + Z_{u''})$.

Proof: Letting $t = 1/u > t' = 1/u' > t'' = 1/u''$, it is equivalent to show that $I(X; X_u + Z) - I(X; X_t + Z) \geq \mu \cdot (I(X; X_t + Z) - I(X; X_{t''} + Z))$. But this holds with $\mu = \frac{t}{t''}$ by concavity of $t \mapsto I(X; X_t + Z)$.

Remark 4. The concavity of $I(X; X_t + Z)$ or $h(X_t + Z)$ is a consequence of the concavity of the entropy power $N(X_t + Z)$, but is strictly weaker as remarked in [1], since a concave function is not always exponentially concave. In fact, it can be shown that $I(X; X_t + Z)$ is equivalent to the concavity of $N(X + Z)$. By taking the logarithm, this implies concavity of both $h(X_t + Z)$ and $h(X + Z)$. While the latter can be shown directly using the DPI [25], the former requires de Bruijn’s identity or the I-MMSE relation [26].

Proposition 4. For the weak $Z$-interference Gaussian channel,

$$C'_f = \frac{1}{2} \log \left(1 + \frac{aP_1}{P_2 + N}\right).$$  \hspace{1cm} (36)

Proof: Suppose that $nR_2 \gtrsim nC_2$. From Proposition 1 written for transmission 2, $X_2 + Z$ is AG and adding interference $\alpha X_1$ in $Y_2 = \sqrt{\alpha}X_1 + X_2 + Z$ is AL w.r.t. $X_2$. Since $\alpha \neq 0$, by the Fork Lemma (Lemma 3), this implies that adding $X_2$ in $Y_2 = \sqrt{\alpha}X_1 + X_2 + Z$ is AL compared to $\sqrt{\alpha}X_1 + Z$ w.r.t. $X_1$.

Therefore,

$$nC'_f = h(\sqrt{\alpha}X_1 + X_2 + Z) - h(X_2 + Z) \geq h(\sqrt{\alpha}X_1 + X_2^2 + Z) - h(X_2^2 + Z) \geq I(X_1; \sqrt{\alpha}X_1 + X_2^2 + Z)$$

$$\geq I(X_1; \sqrt{\alpha}X_1 + X_2 + Z) \geq I(X_1; \sqrt{\alpha}X_1 + \sqrt{\alpha}Z) \geq I(X_1; X_1 + Z) = I(X_1; Y_1) \geq nR_1$$

(see (35)–(37))

which proves that $nR_1 \gtrsim nC'_f$ (cf. (3b)). Notice that we have used Lemma 2 for $u = a, u' = N$ and $u'' = P_2 + N$ in the form: $I(X_1; \sqrt{\alpha}X_1 + X_2^2 + Z) \geq I(X_1; \sqrt{\alpha}X_1 + Z)$ implies $I(X_1; \sqrt{\alpha}X_1 + Z) \geq I(X_1; \sqrt{\alpha}X_1 + \sqrt{\alpha}Z)$.

VIII. CONCLUSION

In this work, a complete determination of corner points of the capacity region of the two-user Gaussian interference channel is carried out, using the notions of almost Gaussian random vectors, almost lossless addition of random vectors, and almost linearly dependent random vectors. The resulting proofs use some basic properties of Shannon’s information theory. Interestingly, only weak forms the entropy power inequality and the concavity of the entropy power are required. This approach does not aim at finding best possible constants but yields a rigorous proof for the determination of Costa’s “missing” corner point which can be thought of as a variation of the solution of Polyanisky and Wu which does not recours to optimal transport theory nor to estimation theory.

ACKNOWLEDGMENTS

The author would like to thank Flavio Calmon, Max Costa, Michèle Wigger and Yihong Wu for their discussions.

REFERENCES

[1] Y. Polyanisky and Y. Wu, “Wasserstein continuity of entropy and outer bounds for interference channels,” 2015, draft at http://arxiv.org/abs/1504.04419.
[2] R. Bustin, H. V. Poor, and S. Shamai, “The effect of maximal rate codes on the interfering message rate,” in Proc. ISIT’14, Honolulu, Hawaii, USA, July 2014, pp. 91–95, longer draft available at http://arxiv.org/abs/1404.6690.
[3] ——, “Optimal point-to-point codes in interference channels: An incremental approach,” 2015, draft at http://arxiv.org/abs/1510.08213.
[4] M. H. M. Costa and O. Rioul, “From almost Gaussian to Gaussian,” in AIP Proc. Int. Workshop on Bayesian Inference and Maximum Entropy Methods (MaxEnt), Amboise, France, Sept. 21–26, 2014.
[5] ——, “From almost Gaussian to Gaussian: Bounding differences of differential entropies,” in Information Theory and Applications Workshop (ITA 2015), San Diego, Feb. 2–6 2015.
[6] O. Rioul and M. H. M. Costa, “Almost there: Corner points of Gaussian interference channels,” in Information Theory and Applications Workshop (ITA 2015), San Diego, Feb. 2–6 2015.
[7] ——, “On some almost properties,” in Information Theory and Applications Workshop (ITA 2016), San Diego, Jan. 31–Feb. 5 2016.
[8] M. H. M. Costa, “On the Gaussian interference channel,” IEEE Trans. Inf. Theory, vol. 31, no. 5, pp. 607–615, Sept. 1985.
[9] A. B. Carleial, “A case where interference does not reduce capacity,” IEEE Trans. Inf. Theory, vol. 21, no. 5, pp. 569–570, Sept. 1975.
[10] T. S. Han and K. Kobayashi, “A new achievable rate region for the interference channel,” IEEE Trans. Inf. Theory, vol. 27, no. 1, pp. 49–60, Jan. 1981.
[11] H. Sato, “The capacity of the Gaussian interference channel under strong interference,” IEEE Trans. Inf. Theory, vol. 27, no. 6, pp. 786–788, Nov. 1981.
[12] ——, “On degraded Gaussian two-user channels,” IEEE Trans. Inf. Theory, vol. 24, no. 5, pp. 637–640, Sept. 1978.
[13] P. P. Bergmans, “A simple converse for broadcast channels with additive white Gaussian noise,” IEEE Trans. Inf. Theory, vol. 20, no. 2, pp. 279–280, March 1974.
[14] Y. Geng and C. Nair, “The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages,” IEEE Trans. Inf. Theory, vol. 60, no. 4, pp. 2087–2104, Apr. 2014.
[15] O. Rioul, “Information theoretic proofs of entropy power inequalities,” IEEE Transactions on Information Theory, vol. 57, no. 1, pp. 33–55, Jan. 2011.
[16] ——, “Yet another proof of the entropy power inequality,” 2016, available at http://arxiv.org/abs/1606.05969.
[17] V. D. Milman and G. Schechtman, Asymptotic Theory of Finite Dimensional Normed Spaces, ser. Lecture Notes in Mathematics. Springer, 1986, vol. 1200.
[18] R. Schneider, Convex Bodies: The Brunn-Minkowski Theory. Cambridge University Press, 1993.
[19] A. A. Giannopoulos and V. D. Milman, Asymptotic convex geometry: A short overview,” in Different Faces of Geometry, S. Donaldson, Y. Eliashberg, and M. Gromov, Eds. Springer, 2004, vol. 3, pp. 87–162.
[20] S. Artstein-Avidan, A. Giannopoulos, and V. D. Milman, *Asymptotic Geometric Analysis I*. Amer. Math. Soc., 2015.

[21] H. Knothe, “Contributions to the theory of convex bodies,” *Michigan Math. J.*, vol. 4, pp. 39–52, 1957.

[22] I. Sason, “On achievable rate regions for the Gaussian interference channel,” *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1345–1356, June 2004.

[23] G. Kramer, “Review of rate regions for interference channels,” in *Proc. IEEE Int. Zurich Seminar on Communications (IZS)*, Feb. 22–24, 2006, pp. 162–165.

[24] M. H. M. Costa, “A new entropy power inequality,” *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 751–760, Nov. 1985.

[25] J. Fahs and I. Abou-Fayçal, “A new tight upper bound on the entropy of sums,” *Entropy*, vol. 17, pp. 8312–8324, Dec. 2015.

[26] D. Guo, S. S. (Shitz), and S. Verdú, “Mutual information and minimum mean-square error in gaussian channels,” *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.