From Almost Gaussian to Gaussian
Bounding Differences of Differential Entropies

Max H. M. Costa

School of Electrical and Computer Engineering
University of Campinas – Unicamp, Brazil
max@fee.unicamp.br

Olivier Rioul

Télécom ParisTech
CNRS LTCI, Paris, France
olivier.rioul@telecom-paristech.fr

Abstract—**We consider lower and upper bounds on the difference of differential entropies of a Gaussian random vector and an almost Gaussian random vector after both are "smoothed" by an arbitrarily distributed random vector of finite power. These bounds are important to prove the optimality of corner points in the capacity region of Gaussian interference channels. The upper bound, presented in MaxEnt-2014, follows from the data processing inequality (DPI). For the lower bound we consider a class of almost Gaussian distributions and use the DPI and a symmetry argument. We also show a counterexample that disproves a conjecture we proposed in MaxEnt-2014 regarding a certain difference of integrals.**

I. Introduction

The bounding problem we consider is motivated by the Gaussian interference channel. For simplification, we consider the degraded Gaussian interference channel. This model is shown in Fig. 1.
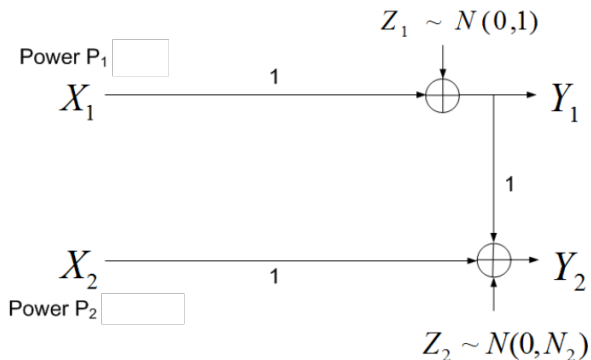

Fig. 1. Degraded Gaussian interference channel.

The channel outputs are given by $Y_1 = X_1 + Z_1$ and $Y_2 = X_1 + X_2 + Z_1 + Z_2$, where X_1 and X_2 are independent input signals constrained to have average powers P_1 and P_2, respectively, and Z_1 and Z_2 are Gaussian noise variables with variances equal to 1 and N_2, respectively. The receivers are interested in messages sent by their respective transmitters. Thus X_1 encodes a message addressed to receiver 1 and X_2 sends a message to receiver 2.

II. Preliminaries

The capacity of an additive white Gaussian noise channel $Y = X + Z$, with X constrained to power P and Z distributed according to $\mathcal{N}(0, N)$, is well known to be

$C = \frac{1}{2} \log_2(1 + P/N)$ bits per transmission, or bits per dimension.

The degraded interference channel of Fig. 1 has outputs given by

$$\begin{cases} Y_1 &= X_1 + Z_1 \\ Y_2 &= X_1 + X_2 + Z_1 + Z_2. \end{cases} \tag{1}$$

We denote vectors by boldface letters and scalar random variables by capital Roman letters. $H(\cdot)$ is used to denote Shannon's entropy and $h(\cdot)$ denotes differential entropy. $I(\cdot; \cdot)$ denotes mutual information.

One of the corner points of the capacity region of this channel is known as the Sato point. Its coordinates are

$$(R_1, R_2) = (\tfrac{1}{2} \log(1 + P_1), \tfrac{1}{2} \log(1 + \tfrac{P_2}{1 + P_1 + N_2})),$$

the rate pair produced when transmitter 1 occupies the channel at its maximum possible rate and transmitter 2 takes up whatever is leftover, considering X_1's signal as noise [1]. Fig. 2 shows the two corner points in the achievable rate region
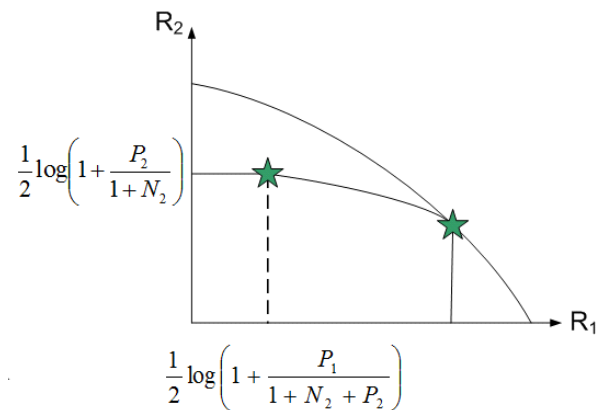

Fig. 2. Achievable rate region for the degraded Gaussian interference channel.

of the degraded Gaussian interference channel. The Sato point is the star in the lower right. It is a point in the curved boundary of the capacity region of the associated broadcast channel with power $P = P_1 + P_2$, also graphed in Fig. 2.

The corner point in the other side of the achievable region (the upper left star), known as the missing corner point [2], is produced when transmitter 2 takes up the channel at its maximum possible rate and requires that transmitter 1 do not

inflict a reduction in its rate (R_2) by more than an infinitesimal amount. The point coordinates are

$$(R_1, R_2) = (\tfrac{1}{2}\log(1 + \tfrac{P_1}{1+P_2+N_2}), \tfrac{1}{2}\log(1 + \tfrac{P_2}{1+N_2})).$$

Thus R_1 is the capacity of a Gaussian channel with signal power P_1 and noise power $1 + P_2 + N_2$, and R_2 is the capacity of a channel with signal power P_2 and noise power $1 + N_2$. To establish that this is a corner point in the capacity region of the degraded channel we need to demonstrate that all communication between X_1 and Y_1 will be fully decoded by Y_2 as well, even if this is not a requirement of the interference channel model.

The difficulty in demonstrating this bound arises from a lack of a Fano type inequality that requires the conditional equivocation of $\mathbf{X}_1$ given $\mathbf{Y}_2$, $H(\mathbf{X}_1|\mathbf{Y}_2)$, to be n times a small number ϵ that goes to zero as the probability of error of decoding $\mathbf{X}_1$ at $\mathbf{Y}_2$ goes to zero, with n being the code length. This Fano type inequality does not exist because Y_2 is not required to decode X_1 in the interference channel model.

To produce a Fano type inequality that leads to the desired demonstration, we consider the difference between $H(\mathbf{X}_1|\mathbf{Y}_2)$ and $H(\mathbf{X}_1|\mathbf{Y}_3)$, where $\mathbf{Y}_3$ is an auxiliary random vector produced by the sum of $\mathbf{X}_1$ and an independent Gaussian noise vector of covariance $(1+N_2+P_2)I$, I being the identity matrix. This motivates the title of this paper. We need to bound the difference in the equivocation of $\mathbf{X}_1$ when it is seen at $\mathbf{Y}_2$, a signal that is produced by $\mathbf{X}_1$ added to the almost Gaussian signal $\mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{X}_2$, and when it is seen at the output of a channel with $\mathbf{X}_1$ as the input and additive white Gaussian noise of power $(1+N_2+P_2)$. To simplify the model we let Y_3 be given by $Y_3 = X_1 + Z_1 + Z_2 + Z_3$, where Z_3 is Gaussian with variance P_2.

Consider scalar variables for a moment. By the chain rule we have

$$
\begin{aligned}
H(X_1|Y_2) &= H(X_1) + h(Y_2|X_1) - h(Y_2) \\
&= H(X_1) + h(Z_1 + Z_2 + X_2) - h(Y_2)
\end{aligned}
\tag{2}
$$

and

$$
\begin{aligned}
H(X_1|Y_3) &= H(X_1) + h(Y_3|X_1) - h(Y_3) \\
&= H(X_1) + h(Z_1 + Z_2 + Z_3) - h(Y_3).
\end{aligned}
\tag{3}
$$

Subtracting gives

$$
\begin{aligned}
H(X_1|Y_3) - H(X_1|Y_2) = {}& h(Z_1 + Z_2 + Z_3) \\
& - h(Z_1 + Z_2 + X_2) + h(Y_2) - h(Y_3).
\end{aligned}
\tag{4}
$$

Since the Gaussian distribution maximizes entropy given a power constraint, and from the hypothesis related to the extreme point where R_2 is maximal, we have that $h(Z_1+Z_2+Z_3) - h(Z_1+Z_2+X_2)$ is bounded below by zero and above by a diminishing ϵ_1 [3] (the almost Gaussian assumption).

Now we consider the vector version of (4). We have that

$$
\begin{aligned}
h(\mathbf{Y}_2) - h(\mathbf{Y}_3) &\le H(\mathbf{X}_1|\mathbf{Y}_3) - H(\mathbf{X}_1|\mathbf{Y}_2) \\
&\le n\epsilon_1 + h(\mathbf{Y}_2) - h(\mathbf{Y}_3)
\end{aligned}
\tag{5}
$$

It is clear that, to bound the difference of equivocations $H(\mathbf{X}_1|\mathbf{Y}_3) - H(\mathbf{X}_1|\mathbf{Y}_2)$, we need to obtain upper and lower bounds for the difference of differential entropies $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$. An earlier attempt to bound this difference [3, Appendix B] resulted in a bound with a faster than linear dependence on n, as noted in [4], which impaired the usefulness of the bound.

In [5], we presented an upper bound based on the data processing inequality (DPI) for $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$, which is repeated below for convenience.

III.   UPPER BOUND

We obtain an upper bound on $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$. Since X_2 needs to communicate with Y_2, we have from Fano's inequality that $H(\mathbf{X}_2|\mathbf{Y}_2) \le n\epsilon_2$, where ϵ_2 goes to zero as the probability of error disappears. As $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, we observe that

$$
\begin{aligned}
I(\mathbf{X}_1; \mathbf{Y}_2) &= I(\mathbf{X}_1; \mathbf{Y}_2|\mathbf{X}_2) - I(\mathbf{X}_1; \mathbf{X}_2|\mathbf{Y}_2) \\
&\ge I(\mathbf{X}_1; \mathbf{Y}_2|\mathbf{X}_2) - n\epsilon_2 \\
&= H(\mathbf{X}_1) - H(\mathbf{X}_1|\mathbf{X}_1 + \mathbf{Z}_1 + \mathbf{Z}_2) - n\epsilon_2 \\
&= I(\mathbf{X}_1; \mathbf{X}_1 + \mathbf{Z}_1 + \mathbf{Z}_2) - n\epsilon_2 \\
&\ge I(\mathbf{X}_1; \mathbf{Y}_3) - n\epsilon_2,
\end{aligned}
\tag{6}
$$

by the DPI. Equivalently,

$$
\begin{aligned}
h(\mathbf{Y}_3) - h(\mathbf{Y}_2) &\le h(\mathbf{Y}_3|\mathbf{X}_1) - h(\mathbf{Y}_2|\mathbf{X}_1) + n\epsilon_2 \\
&= h(\mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{Z}_3) \\
&\quad - h(\mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{X}_2) + n\epsilon_2 \\
&\le n(\epsilon_1 + \epsilon_2),
\end{aligned}
\tag{7}
$$

by the almost Gaussian assumption.

IV.   TOWARDS A LOWER BOUND

We now consider a lower bound for the difference $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$. Let f denote the probability density function of $\mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{X}_2$, the almost Gaussian density, and let g denote the Gaussian density of $\mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{Z}_3$. We assume that the covariance matrices of f and g are identical and that $0 \le h(g) - h(f) \le n\epsilon_1$. We also write $f = g + \Delta f$, at this point making no particular assumptions about Δf. Then we can expand the Kullback-Leibler (KL) divergence of f from g as $D(f\|g) = h(g) - h(f) + \int (g-f)\log g = h(g) - h(f) \le n\epsilon_1$, using the assumption of equal covariance matrices.

Now we assume that $\mathbf{X}_1$ has density p, arbitrary except for the finite power, and consider the effect of smoothing f and g by the addition of $\mathbf{X}_1$. Strictly, $\mathbf{X}_1$ may not have a density as it is typically drawn from a finite set of codewords (a "porcupine" distribution). We assume the existence of p for ease of notation since Gaussian smoothing guarantees the existence of all considered probability densities. We denote the convolution of g and p by $g * p$. By the DPI we have that

$$0 \le D(f * p \| g * p) \le D(f\|g) \le n\epsilon_1. \tag{8}$$

Expanding as before we get

$$0 \le h(g * p) - h(f * p) + \int (g*p - (g + \Delta f)*p)\log(g*p)$$
$$\le n\epsilon_1, \tag{9}$$

or

$$0 \le h(g * p) - h(f * p) - \int (\Delta f * p)\log(g * p) \le n\epsilon_1. \tag{10}$$

Equivalently,

$$0 \le h(\mathbf{Y}_3) - h(\mathbf{Y}_2) - \int (\Delta f * p) \log(g * p) \le n\epsilon_1. \quad (11)$$

This inequality is key to getting the desired lower bound. In [5], in order to establish the lower bound, we conjectured that the positive integral given by $D(f\|g) + D(g\|f) = \int (f - g) \log f$ would maintain its sign after smoothing by p. This conjecture turns out to be false and we present a counter-example in the Appendix.

V. LOWER BOUND FOR A CLASS OF DISTRIBUTIONS f

To establish the required lower bound, we can map the loci of authorized points for the quantities $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$ and $-\int (\Delta f * p) \log(g * p)$ in a graph, as shown in Fig. 3. The valid points are laid on a diagonal strip of width $n\epsilon_1$ and the upper bound on $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$ further restricts the allowed values to the left of the $n(\epsilon_1 + \epsilon_2)$ mark in the horizontal axis.
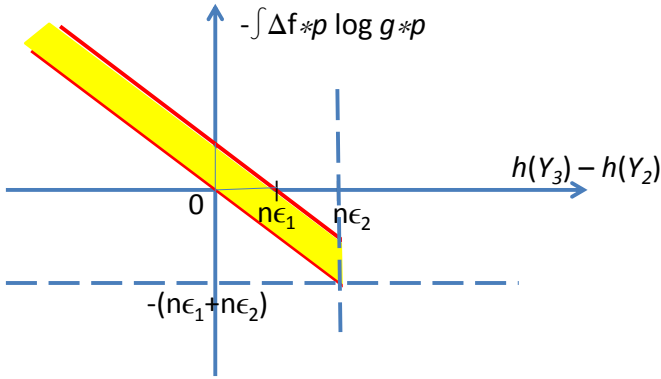

Fig. 3. Region of possible values for $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$ and $-\int (\Delta f * p) \log(g * p)$.

We now look at the characteristics of the deviation function Δf. Notice that $\int \Delta f = \int x \Delta f = \int \|x\|^2 \Delta f = 0$, as the first order moments of f and g are identical[1]. It is known that $D(f\|g) = 0$ if and only if $f = g$ almost everywhere. The smoothing produced by the additive Gaussian noise takes care of excluding problematic convergence behavior. Since $D(f\|g) = \int (f - g) \le n\epsilon_1$, we assume that Δf will converge pointwisely to 0 as $n \to \infty$ and $\epsilon_1 \to 0$. Thus we consider the class of densities $f = g + \Delta f$ for which $|\Delta f| \ll g$. Because of the assumption that Δf converges to 0, we argue that this condition will be prevalent among all qualifying distributions f. Considering this class of distributions, we note that the capacities of the channel from X_2 to Y_2 (with $X_1 = 0$) induced by distributions $f + \Delta f$ and $f - \Delta f$ are the same, up to second order terms. This can be seen by expanding

$$0 \le h(g) - h(g + \Delta f)$$
$$= -\int g \log g + \int (g + \Delta f) \log(g + \Delta f) \quad (12)$$
$$= \int g \log \frac{g + \Delta f}{g} + \int \Delta f \log \frac{g + \Delta f}{g},$$

[1] We assume, without loss of generality, that $EX_2 = EZ_3 = 0$.

where we used the fact that $\int \Delta f \log g = \int \|x\|^2 \Delta f = 0$. For $\Delta f \ll g$ we can use Taylor series expansion to obtain

$$h(g) - h(g + \Delta f) = \int \Delta f - \frac{(\Delta f)^2}{2g} + \frac{(\Delta f)^2}{g} + o(\frac{(\Delta f)^2}{g})$$
$$= \int \frac{(\Delta f)^2}{2g} + o(\frac{(\Delta f)^2}{g}) \le n\epsilon_1. \quad (13)$$

Similarly we have

$$0 \le h(g) - h(g - \Delta f) = \int \frac{(\Delta f)^2}{2g} + o(\frac{(\Delta f)^2}{g}) \le n\epsilon_1. \quad (14)$$

To visualize the fact that $f + \Delta f$ and $f - \Delta f$ induce the same achievable rates, we can imagine a hypersphere of radius $\sqrt{n(1 + N_2 + P_2)}$. The typical set of the X_2 codewords plus the Gaussian noise will be in the shell of this hypersphere with positive deviations from g at positions that are related to codeword locations in the $\sqrt{nP_2}$-radius codebook hypersphere. So changing $f + \Delta f$ to $f - \Delta f$ is similar to replacing crests by troughs in the bumpy hyperspherical shell, which amounts to a simple rotation in the X_2 codebook.

We now note that any valid bound for functions of Δf will also imply a symmetric bound corresponding to $-\Delta f$ as these two deviation functions from g are equally able to produce achievable rates. Considering that the vertical ordinate $-\int (\Delta f * p) \log(g * p)$ depends of f only through Δf, we can, by symmetry, further limit the authorized points in the plane to those lying below the $n(\epsilon_1 + \epsilon_2)$ mark in the vertical axis. Finally we observe that the authorized region for the points is situated in the square box of sides $2n(\epsilon_1 + \epsilon_2)$ centered at the origin, as depicted in Fig. 4. Thus we have the desired lower bound on $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$.
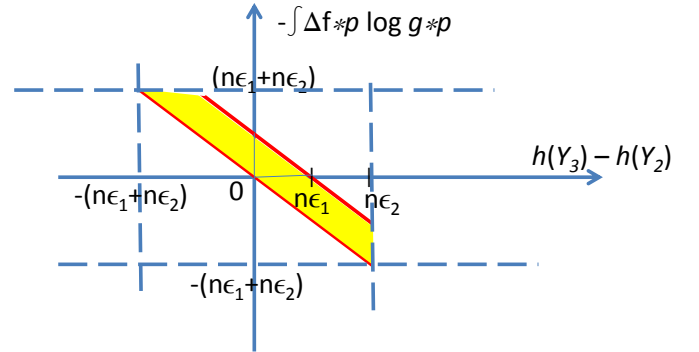

Fig. 4. Bounded supports for horizontal and vertical ordinates.

These arguments are good indication that we can indeed limit the support of $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$ to a vanishing interval around the origin. This is valid for the class of distributions $f = g + \Delta f$ for which $\Delta f \ll g$, that is, close to the limit as $f \to g$, but with no restrictions on the interfering distribution p of $\mathbf{X}_1$.

VI. CONCLUSION

In this paper we investigate upper and lower bounds for a difference of differential entropies that arise in the study of Gaussian interference channels. The upper bound follows from the DPI and was presented earlier in [5]. The lower bound

is established for a class of distributions f such that, for high values of n, the deviation Δf from g is negligible. The argued bounds on the difference of differential entropies lead to upper and lower bounds on $H(\mathbf{X}_1|\mathbf{Y}_3) - H(\mathbf{X}_1|\mathbf{Y}_2)$. This provides a Fano type inequality that allows us to apply the lower bound on equivocation given in [3, Appendix C] to establish the validity of the second corner point of the Z-Gaussian interference channel. This lower bound on equivocation is a consequence of the concavity of the entropy power with increasing Gaussian smoothing. As in [3], this result can be extended to the corner points in the standard Gaussian interference channel with less than strong interference, i.e., with interference gains in the interval $(0,1)$.

In conclusion, we make reference to two recent papers [6], [7] that also address this intriguing limiting problem.

APPENDIX
A. THE SIGN OF $\int (f * p - g * p) \log(f * p)$

In our MaxEnt paper [5], we conjectured that the positive integral given by $D(f\|g) + D(g\|f) = \int (f - g) \log f \geq 0$ would maintain its sign after smoothing by p. This conjecture turns out to be false and we present below a counter-example in one dimension ($n = 1$) in which $\int (f * p - g * p) \log(f * p)$ assumes negative values.

We use the Hermite coordinate system described in [8]. Let $f = (1 + \epsilon H)g_P$ and $p = (1 + \epsilon H')g_Q$, where g_P and g_Q denote zero-mean Gaussian densities of variances P and Q, respectively. Here H and H' are polynomials of even degree, expressed as sums of Hermite polynomials of order ≥ 3 in the corresponding Hilbert space $\mathcal{L}^2(g_P)$ and $\mathcal{L}^2(g_Q)$, respectively. Because of the degrees are even, the quantity ϵ can be set to some small positive value such that both f and p are non negative. Also, since H and H' are orthogonal to polynomials of degree ≤ 2 in their corresponding Hilbert spaces, f and p share the same moments up to second order as g_P and g_Q, respectively. Thus $\int f = \int p = 1$, $\int x f(x)\,\mathrm{d}x = \int x p(x)\,\mathrm{d}x = 0$, and $\int x^2 f(x)\,\mathrm{d}x = P$, $\int x^2 p(x)\,\mathrm{d}x = Q$.

We now compute $f * p$ and $g * p$ using an extended version of [8, Theorem 2] which is stated in the last section of [8]. For completeness we provide a short proof in the next section of this Appendix. One obtains

$$
\begin{aligned}
f * p &= (1 + \epsilon\widetilde{H} + \epsilon\widetilde{H}' + \epsilon^2 H'')g_{P+Q}, \\
g * p &= (1 + \epsilon\widetilde{H}')g_{P+Q},
\end{aligned}
\tag{15}
$$

for appropriately computed polynomial terms $\widetilde{H}$, $\widetilde{H}'$ and H''. Expanding to second-order terms, this leads to

$$
\int (f * p - g * p) \log(f * p) = \epsilon^2 \left(\|\widetilde{H}\|^2 + \langle \widetilde{H} \mid \widetilde{H}' \rangle\right) + o(\epsilon^2),
\tag{16}
$$

with norm and inner product defined in $\mathcal{L}^2(g_{P+Q})$. This quantity can be negative because the inner product can be sufficiently negative to more than cancel the squared norm of $\widetilde{H}$. Consider, for example, $P = Q = 1$, and $H = H_3 + \delta H_4$ and $H' = -2H_3 + \delta H_4$, where δ equals some arbitrarily small positive value (intended to yield even degree polynomials [8]) and H_3 and H_4 are Hermite polynomials of degrees 3 and 4, respectively. Then a simple calculation yields $\|\widetilde{H}\|^2 +$

$\langle \widetilde{H} \mid \widetilde{H}' \rangle = 1^2 2^{-3} + \delta^2 2^{-4} - 2 \cdot 2^{-3} + \delta^2 2^{-4} = (-1 + \delta^2)/8$ which is negative. This disproves our conjecture regarding the sign of $\int (f * p - g * p) \log(f * p)$.

B. A SHORT PROOF OF THE EXTENDED THEOREM 2 IN [8]

Theorem 1. *Let* $H_k^P(x) = \frac{1}{\sqrt{k!}} H_k(x/\sqrt{P})$, $k \in \mathbb{N}$, *be the Hermite orthonormal basis of* $\mathcal{L}^2(g_P)$ *as defined in [8]. Then*

$$
(H_k^P \cdot g_P) * (H_l^Q \cdot g_Q) = \sqrt{\binom{k+l}{k} \frac{P^k Q^l}{(P+Q)^{k+l}}} \cdot H_{k+l}^{P+Q} \cdot g_{P+Q}.
\tag{17}
$$

Proof: By the definition of the Hermite polynomial, the kth order derivative of g_P is given by $g_P^{(k)} = (-1)^k \left(\frac{\sqrt{k!}}{P^{k/2}}\right) H_k^P \cdot g_P$ where the constant term is due to the normalization in the definition of H_k^P. Using this definition and the well-known relation $g_P * g_Q = g_{P+Q}$, we obtain

$$
\begin{aligned}
H_k^P \cdot g_P * H_l^Q \cdot g_Q &= (-1)^{k+l} \frac{P^{k/2} Q^{l/2}}{\sqrt{k!l!}} g_P^{(k)} * g_Q^{(l)} \\
&= (-1)^{k+l} \frac{P^{k/2} Q^{l/2}}{\sqrt{k!l!}} g_{P+Q}^{(k+l)} \\
&= \frac{P^{k/2} Q^{l/2}}{\sqrt{k!l!}} \frac{\sqrt{(k+l)!}}{(P+Q)^{(k+l)/2}} H_{k+l}^{P+Q} g_{P+Q}
\end{aligned}
$$

∎

REFERENCES

[1] H. Sato, "On degraded Gaussian two-user channels," *IEEE Trans. Inform. Theory*, vol. 24, pp. 637–640, Sept. 1978.

[2] G. Kramer, "Review of rate regions for Interference Channels," *Int. Zurich Seminar on Communications (IZS)*, pp. 162–165, Feb. 2006.

[3] M. H. M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inform. Theory*, vol. 31, pp. 607–615, Sept. 1985.

[4] I. Sason, "On achievable rate regions for the Gaussian interference channel," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1345–1356, June 2004.

[5] M. H. M. Costa and O. Rioul, "From almost Gaussian to Gaussian," *in AIP Proc. Int. Workshop on Baysean Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt)*, Amboise, France, Sept. 2014.

[6] I. Sason, "On the corner points of the capacity region of a two-user Gaussian interference channel," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 2739–2743, Honolulu, Hawaii, USA, July 2014. Longer draft at http://arxiv.org/abs/1306.4934.

[7] R. Bustin, H. V. Poor and S. Shamai, "The effect of maximal rate codes on the interfering message rate," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 91–95, Honolulu, Hawaii, USA, July 2014. Longer draft at http://arxiv.org/abs/1404.6690.

[8] E. A. Abbe and L. Zheng, "A coordinate system for Gaussian networks," *IEEE Trans. Inform. Theory*, vol. 58, pp. 721–733, Feb. 2012.