

From almost Gaussian to Gaussian

Max H. M. Costa, Olivier Rioul

▶ **To cite this version:**

Max H. M. Costa, Olivier Rioul. From almost Gaussian to Gaussian. 34th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2014), Sep 2014, Amboise, France. pp.67-73, 10.1063/1.4905964 . hal-02286942

HAL Id: hal-02286942

https://telecom-paris.hal.science/hal-02286942

Submitted on 10 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

From Almost Gaussian to Gaussian

Max H. M. Costa* and Olivier Rioul†

*University of Campinas - Unicamp, Brazil, max@fee.unicamp.br
†Télécom ParisTech - Institut Mines-Télécom - 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 approximately Gaussian random vector after they are "smoothed" by an arbitrarily distributed random vector of finite power. These bounds are important to establish the optimality of the corner points in the capacity region of Gaussian interference channels. A problematic issue in a previous attempt to establish these bounds was detected in 2004 and the mentioned corner points have since been dubbed "the missing corner points". The importance of the given bounds comes from the fact that they induce Fano-type inequalities for the Gaussian interference channel. Usual Fano inequalities are based on a communication requirement. In this case, the new inequalities are derived from a non-disturbance constraint. The upper bound on the difference of differential entropies is established by the data processing inequality (DPI). For the lower bound, we do not have a complete proof, but we present an argument based on continuity and the DPI.

Keywords: Gaussian interference channels, the missing corner points, Gaussian approximation, Kullback-Leibler distance, data processing inequality.

1. INTRODUCTION

The bounding problem we consider is motivated by the Gaussian interference channel. To simplify the discussion we investigate a simpler version, the Z Gaussian interference channel, where interference is restricted to only one direction. This model is depicted in Fig. 1. The channel outputs are given by $Y_1' = X_1' + Z_1'$ and $Y_2' = X_2' + aX_1' + Z_2'$, where X_1' and X_2' are input (transmitter) signals constrained to have average powers Q_1 and Q_2, respectively, a is an interference gain in the interval $(0,1)$, and Z_1' and Z_2' are Gaussian noise variables of unit variance. The receivers are interested in messages sent by their respective, same indexed, transmitters. Thus X_1' encodes a message addressed to receiver 1 and X_2' conveys a message to receiver 2.

This model is a particular case of the general Gaussian interference channel, which exhibits interference in both directions. Like in the general model, finding the channel capacity region for arbitrary parameters has been an open problem for over 35 years. In the case of strong interference, when $a \geq 1$, the capacity region is known [1, 2, 3]. In this situation, the unintended receiver is sure to have the means to decode the interference. Also, when $a = 0$, the problem has a trivial solution. The problem is open in the case of weak interference, where the interference gain a satisfies $0 < a < 1$.

This work has been partially supported by FAPESP and CAPES.

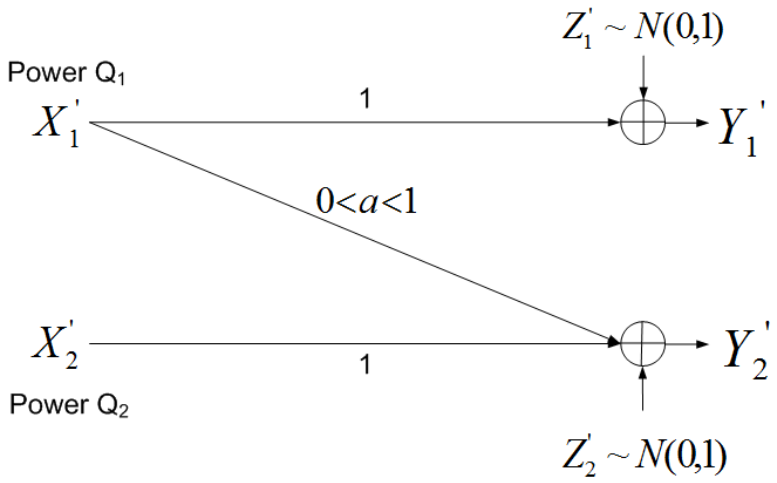


FIGURE 1. Z-Gaussian interference channel.

It can be shown that the capacity region of the Z Gaussian interference channel with weak interference is identical to the capacity region of an equivalent degraded Gaussian interference channel [4], a model shown in Fig. 2. For simplicity, we chose to use the more common notation, without the primes, in this degraded channel, as this will be our working model.

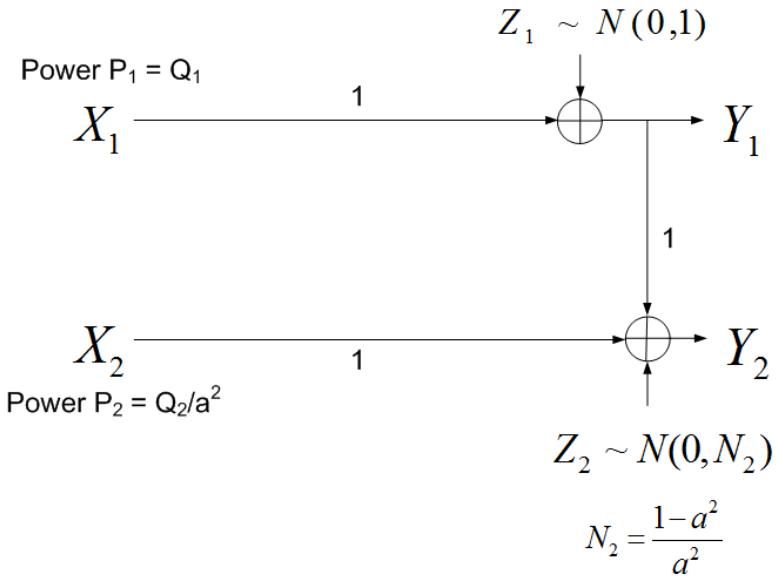


FIGURE 2. Degraded Gaussian interference channel.

Like the Z-Gaussian interference channel, the degraded Gaussian interference channel is characterized by three parameters, namely the two powers P_1 and P_2, and the additional independent noise in the second receiver, of power N_2. These parameters are related to the parameters of the original Z-channel by $P_1 = Q_1$, $P_2 = Q_2/a^2$, and $N_2 = (1-a^2)/a^2$. Moreover, since $0 < a < 1$, the additional noise power N_2 is positive. This degraded version of the interference channel is more amenable to evidentiate the effects of the interference caused by X_1 over Y_2.

2. 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. 2 has outputs given by

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

We use the usual notation of denoting vectors by boldface types and scalar random variables by capital Roman types. $H(.)$ is used to denote Shannon's entropy and $h(.)$ denotes differential entropy. $I(.;.)$ denotes mutual information.

One of the corner points of the capacity region of this channel is known as the Sato point. Its coordinates are $(R_1,R_2) = (\frac{1}{2}\log(1+P_1), \frac{1}{2}\log(1+\frac{P_2}{1+P_1+N_2}))$, 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. Sato obtained this corner point by showing that it is also in the border of the capacity region of the broadcast channel with power $P = P_1 + P_2$ [5].

The two corner points in the achievable region of the degraded Gaussian interference channel are shown in Fig. 3. The Sato point is the star in the lower right.

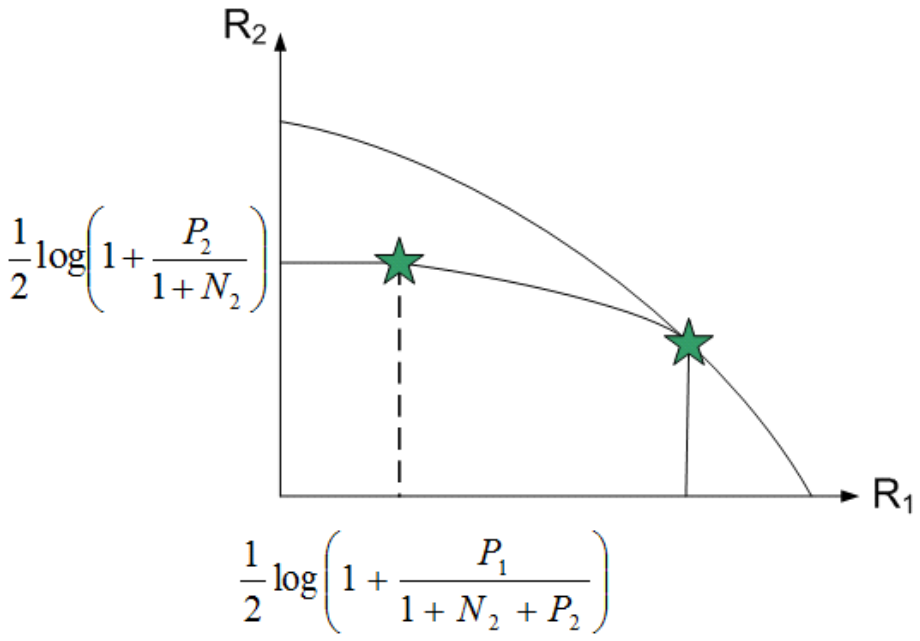


FIGURE 3. Achievable rate region for degraded Gaussian interference channel.

The corner point in the other side of the capacity region (the upper left star), known as the missing corner point [6], 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 drop. The point coordinates are $(R_1,R_2) = (\frac{1}{2}\log(1+\frac{P_1}{1+P_2+N_2}), \frac{1}{2}\log(1+\frac{P_2}{1+N_2}))$. Note that 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 indeed 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 have to 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 X_1 at Y_2 goes to zero, where n is 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 is the motivation for 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 seen at the output of a channel with $\mathbf{X}_1$ in 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

$$H(X_1|Y_2) = H(X_1)+h(Y_2|X_1)-h(Y_2) = H(X_1)+h(Z_1+Z_2+X_2)-h(Y_2), \quad (2)$$

and

$$H(X_1|Y_3) = H(X_1)+h(Y_3|X_1)-h(Y_3) = H(X_1)+h(Z_1+Z_2+Z_3)-h(Y_3). \quad (3)$$

Subtracting gives

$$H(X_1|Y_3)-H(X_1|Y_2) = h(Z_1+Z_2+Z_3)-h(Z_1+Z_2+X_2)+h(Y_2)-h(Y_3) \quad (4)$$

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 [4] (the almost Gaussian assumption).

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

$$h(\mathbf{Y}_2)-h(\mathbf{Y}_3) \le H(\mathbf{X}_1|\mathbf{Y}_3)-H(\mathbf{X}_1|\mathbf{Y}_2) \le n\varepsilon_1+h(\mathbf{Y}_2)-h(\mathbf{Y}_3) \quad (5)$$

Therefore to bound the difference of equivocations we need to bound the difference between the differential entropies of $\mathbf{Y}_2$ and $\mathbf{Y}_3$. The first attempt to bound this difference [4, Appendix B] used Pinsker inequality [7], but the achieved bound had a dependence on n that grew faster than linearly, as pointed out in [8]. In the sequel we consider lower and upper bounds for this difference of differential entropies.

3. UPPER BOUND

In this Section 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 $H(\mathbf{X}_2|\mathbf{Y}_2) \le n\varepsilon_2$, where ε_2 goes to

zero as the probability of error disappears. Since $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, we observe that

$$\begin{aligned} I(\mathbf{X}_1;\mathbf{Y}_2) &= I(\mathbf{X}_1;\mathbf{Y}_2|\mathbf{X}_2) - I(\mathbf{X}_1;\mathbf{X}_2|\mathbf{Y}_2) \\ &\geq I(\mathbf{X}_1;\mathbf{Y}_2|\mathbf{X}_2) - n\varepsilon_2 \\ &= H(\mathbf{X}_1) - H(\mathbf{X}_1|\mathbf{X}_1+\mathbf{Z}_1+\mathbf{Z}_2) - n\varepsilon_2 \\ &= I(\mathbf{X}_1;\mathbf{X}_1+\mathbf{Z}_1+\mathbf{Z}_2) - n\varepsilon_2 \\ &\geq I(\mathbf{X}_1;\mathbf{Y}_3) - n\varepsilon_2, \end{aligned} \tag{6}$$

by the data processing inequality (DPI). Equivalently,

$$\begin{aligned} h(\mathbf{Y}_3) - h(\mathbf{Y}_2) &\leq h(\mathbf{Y}_3|\mathbf{X}_1) - h(\mathbf{Y}_2|\mathbf{X}_1) + n\varepsilon_2 \\ &= h(\mathbf{Z}_1+\mathbf{Z}_2+\mathbf{Z}_3) - h(\mathbf{Z}_1+\mathbf{Z}_2+\mathbf{X}_2) + n\varepsilon_2 \\ &\leq n(\varepsilon_1+\varepsilon_2), \end{aligned} \tag{7}$$

by the almost Gaussian assumption.

4. 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 $0 \leq h(g) - h(f) = n\varepsilon_1$. Then we can write the Kullback-Leibler (KL) divergence of f from g as $D(f\|g) = h(g) - h(f) + \int (g-f)\log g = h(g) - h(f) \leq n\varepsilon_1$, 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. We assume the existence of p for ease of notation since Gaussian smoothing guarantees the existence of all considered densities. We denote the convolution of g and p by $g*p$. By the DPI we have that

$$0 \leq D(g*p\|f*p) \leq D(g\|f) \leq \int (f-g)\log f. \tag{8}$$

Expanding as before we get

$$0 \leq h(f*p) - h(g*p) + \int (f*p - g*p)\log f*p \leq \int (f-g)\log f, \tag{9}$$

or

$$h(f*p) - h(g*p) \leq \int (g*p - f*p)\log f*p + \int (f-g)\log f. \tag{10}$$

Equivalently

$$h(\mathbf{Y}_3) - h(\mathbf{Y}_2) \geq \int (f*p - g*p)\log f*p + \int (g-f)\log f. \tag{11}$$

This inequality is key to getting the desired lower bound. Since the second integral above is strictly negative (cf. (8)) and by continuity of smoothing with increasing values of variance, we argue that the first integral is positive. For the first integral to change sign as a result of the "smoothing" by $\mathbf{X}_1$, it would have to equal zero for some density p. At this point our argument is conjectural in nature as we do not have a proof of the impossibility of the first integral being zero. We believe this would only happen if the smoothed distributions, $f * p$ and $g * p$, were identical, a clear impossibility.

Fig. 4 shows extreme possible paths for the divergences between f and g before and after smoothing by p. Note that $\int (g-f)\log g = 0$, by hypothesis, in one extreme (before smoothing), and $\int (f*p - g*p)\log(f*p)$ is reduced to some arbitrarily small value in the other extreme.

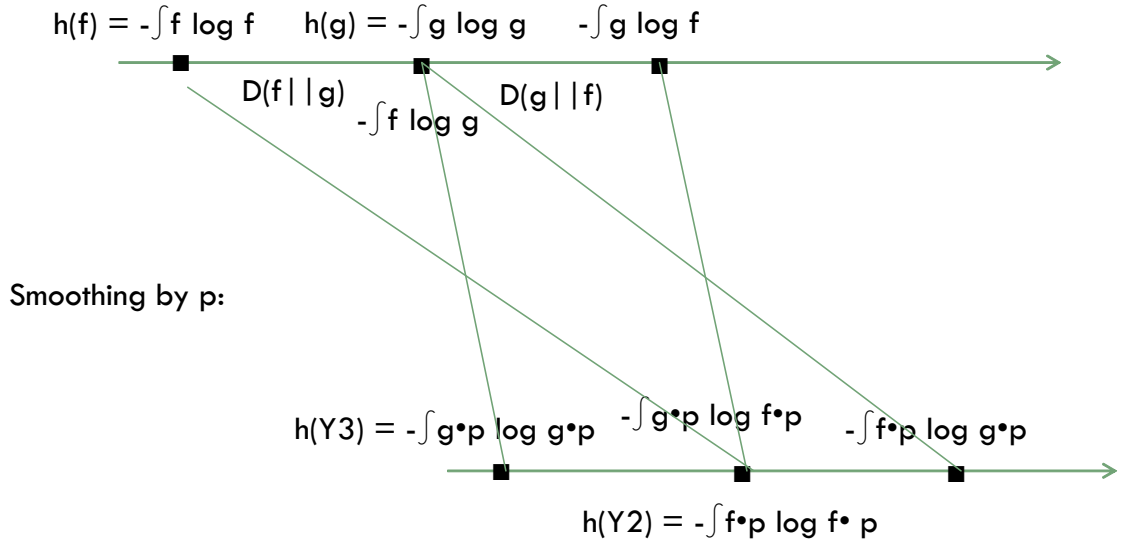


FIGURE 4. Extreme relative positions of divergences between f and g, before and after smoothing by p.

If we can circumvent this difficulty, we would obtain

$$h(\mathbf{Y}_3) - h(\mathbf{Y}_2) \geq \int (g-f)\log f, \tag{12}$$

a lower bound that does not depend on p. Now note that

$$0 \leq D(g\|f) + D(f\|g) = \int (f-g)\log f. \tag{13}$$

We argue that these two KL divergences approach zero at the same rate, based on their variations with respect to f, as f approaches g, both of the same order. This is verified simply by considering the variations of the two divergences with respect to f. Therefore $\int (f-g)\log f \leq 2n\varepsilon_1$, and our lower bound for $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$ is equal to $-2n\varepsilon_1$.

It is interesting to note, somewhat surprisingly, that the difference $h(\mathbf{Y}_3) - h(\mathbf{Y}_2)$ can be negative. This may happen if the arbitrary smoothing by $\mathbf{X}_1$ can somehow partially compensate for the non-Gaussianity of f and make $f * p$ look more Gaussian than $g * p$. This observation is in line with the interesting findings of Abbe and Zheng in [9].

5. DISCUSSION

From (5), the argued lower and upper bounds on the difference of differential entropies lead to lower and upper 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 [4, Appendix C] to establish the legitimacy 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.

In [10] Sason presents a detailed study of the corner points of the capacity region of the two-user Gaussian interference channel. The present considerations are consistent with his findings.

To whisper or to shout. One interesting observation of this result is that there seems to be no way for X_1 to communicate with Y_1 without significantly disturbing the rate of the second link other than by making its own signal understood at the unintended, degraded output Y_2. So, one may produce less inconvenience by shouting than by whispering.

ACKNOWLEDGMENTS

We would like thank Chandra Nair for his invaluable critique and assistance. His dedicated and friendly help were instrumental for our arguments. We also thank Igal Sason, Gerhard Kramer and Lizhong Zheng for their inspiring discussions.

We have recently become aware of [11], where results to establish the corner point are achieved with a minimum mean square error (MMSE) approach, for a class of p distributions. We thank Ronit Bustin for having sent us their manuscript.

REFERENCES

1. A. Carleial, *IEEE Trans. Inform. Theory* **21**, 569–570 (1975).
2. T. S. Han, and K. Kobayashi, *IEEE Trans. Inform. Theory* **27**, 49–60 (1981).
3. H. Sato, *IEEE Trans. Inform. Theory* **27**, 786–788 (1981).
4. M. H. M. Costa, *IEEE Trans. Inform. Theory* **31**, 607–615 (1985).
5. H. Sato, *IEEE Trans. Inform. Theory* **24**, 637–640 (1978).
6. G. Kramer, *Int. Zurich Seminar on Communications (IZS)* pp. 162–165 (2006).
7. I. Csiszár, and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd Ed.*, Cambridge, 2011.
8. I. Sason, *IEEE Trans. Inform. Theory* **50**, 1345–1356 (2004).
9. E. A. Abbe, and L. Zheng, *IEEE Trans. Inform. Theory* **58**, 721–733 (2012).
10. I. Sason, *arXiv:1306. 4934[cs.IT]* (2013).
11. R. Bustin, H. V. Poor, and S. Shamai, *arXiv:1404. 6690v1 [cs.IT]* (2014).