Coding Along Hermite Polynomials for Gaussian Noise Channels

Emmanuel A. Abbe
IPG, EPFL
Lausanne, 1015 CH
Email: emmanuel.abbe@epfl.ch

Lizhong Zheng
LIDS, MIT
Cambridge, MA 02139
Email: lizhong@mit.edu

***Abstract*— This paper shows that the capacity achieving input distribution for a fading Gaussian broadcast channel is not Gaussian in general. The construction of non-Gaussian distributions that strictly outperform Gaussian ones, for certain characterized fading distributions, is provided. The ability of analyzing non-Gaussian input distributions with closed form expressions is made possible in a local setting. It is shown that there exists a specific coordinate system, based on Hermite polynomials, which parametrizes Gaussian neighborhoods and which is particularly suitable to study the entropic operators encountered with Gaussian noise.**

I. Introduction

Let a memoryless additive white Gaussian noise (AWGN) channel be described by $Y = X + Z$, where $Z \sim \mathcal{N}(0, v)$ is independent of X. If the input is subject to an average power constraint given by $\mathbb{E}X^2 \leq p$, the input distribution maximizing the mutual information is Gaussian. This is due to the fact that under second moment constraint, the Gaussian distribution maximizes the entropy, hence

$$\arg \max_{X:\mathbb{E}X^2=p} h(X+Z) \sim \mathcal{N}(0,p). \tag{1}$$

On the other hand, for an additive noise channel, if we use a Gaussian input distribution, i.e., $X \sim \mathcal{N}(0,p)$, among noises with bounded second moment, the noise minimizing the mutual information is again Gaussian. This can be shown by using the *entropy power inequality* (EPI), which reduces in this setting to

$$\arg \min_{Z:\, h(Z)=\frac{1}{2}\log 2\pi e v} h(X+Z) \sim \mathcal{N}(0,v) \tag{2}$$

and implies

$$\arg \min_{Z:\mathbb{E}Z^2=v} h(X+Z) - h(Z) \sim \mathcal{N}(0,v). \tag{3}$$

Hence, in the single-user setting, Gaussian inputs are the best inputs to fight Gaussian noise and Gaussian noise is the worst noise to face with Gaussian inputs. This provides a game equilibrium between user and nature, as defined in [3], p. 263.

With these results, many problems in information theory dealing with Gaussian noise can be solved. However, in the network information theory setting, two interesting new phenomena make the search for the optimal codes more complex. First, the interference. With the interference, a frustration phenomena appears. Assuming Gaussian additive noise for each user, if the users treat interference as noise, the game equilibrium described previously provides now a conflicting situation: for inputs that are Gaussian distributed, we also have the worst kind of interference (Gaussian distributed). Should interference not be treated as noise, or should non-Gaussian ensembles be considered? These are still open questions. Another interesting difference between the single user and network setting concerns the fading. In a single user setting, for an AWGN channel with a coherent fading, (1) allows to show that the Gaussian input distribution achieves the capacity (no matter what the fading distribution is). Just like for the case with no fading. In a Gaussian broadcast channel (BC), (1) and (3) allow to show that Gaussian inputs are again capacity achieving. But if we consider a fading Gaussian BC, with a coherent fading (preserving the degradedness property), there are no known results using (1) and (3) which allow to prove that the Gaussian inputs are capacity achieving. Either we are missing theorems to prove this assertion or it is simply wrong. In this paper, we show that this assertion is wrong.

Even for the following simple structure of coherent Fading Gaussian BC, where

$$Y_i = HX + Z_i, \quad i = 1,2$$

with $Z_1 \sim \mathcal{N}(0, v_1)$, $Z_2 \sim \mathcal{N}(0, v_2)$, $v_1 < v_2$ and H is arbitrarily distributed but is the same for both users; the input distributions achieving the capacity region boundaries are unknown. Since this is a degraded BC, the capacity region is given by all rate pairs $(I(X;Y_1|U,H), I(U;Y_2|H))$ with $U - X - (Y_1, Y_2)$. The optimal input distributions, i.e., the distributions of (U, X) achieving the capacity region boundary, are given by the following optimization, where $\mu \in \mathbb{R}$,

$$\arg \max_{\substack{(U,X):\, U-X-(Y_1,Y_2) \\ \mathbb{E}X^2 \leq p}} I(X;Y_1|U,H) + \mu I(U;Y_2|H). \tag{4}$$

Note that the objective function in the above maximization is

$$h(Y_1|U,H) - h(Z_1) + \mu h(Y_2|H) - \mu h(Y_2|U,H).$$

Now, each term in the above is individually maximized by a Gaussian distribution, but these terms are combined with different signs, so there is a competitive situation and the maximizer is not obvious. When $\mu \leq 1$, one can show that Gaussian distributions are optimal. Also, if H is compactly supported, and if v is small enough as to make the support of

978-1-4244-4313-0/09/$25.00

H and $1/vH$ non overlapping, the optimal input distribution is Gaussian (cf. [5]). However, in general the optimal distributions is unknown.

Note that a similar competitive situation occurs for the interference channel. Let the inputs be X_1 and X_2, the interference coefficients a and b, and the noise Z_1 and Z_2 (independent standard Gaussian). The following expression

$$\begin{aligned} &I(X_1; X_1 + aX_2 + Z_1) + I(X_2; X_2 + bX_1 + Z_2) \\ &= h(X_1 + aX_2 + Z_1) - h(aX_2 + Z_1) \\ &+ h(X_2 + bX_1 + Z_2) - h(bX_1 + Z_2), \end{aligned}$$

is a lower bound to the sum-capacity, which is tight by considering X_1 and X_2 of arbitrary block length n. Now, Gaussian distributions are maximizing each entropy term in the above, hence, since they appear with different signs, there is a competitive situation. Would we then prefer to take X_1 and X_2 Gaussian, "very non-Gaussian" or "slightly non-Gaussian"? Although these questions are not formally defined, the dilemma posed by them can still be understood.

II. Problem Statement

We will use the fading Gaussian BC problem as a motivation for our more general goal. For this specific problem, we want to know if/when the distribution of (U, X) maximizing (4) is Gaussian or not.

Our more general goal is to understand better the problem posed by the competitive situations described in the Introduction. For this purpose, we formulate a mathematical problem in the next section.

III. Local Geometry and Hermite Coordinates

Let g_p denote the Gaussian density with zero mean and variance p. We start by changing the notation and rewrite (1) and (3) as optimizations over the input distributions, i.e.

$$\arg \max_{f:\, m_2(f)=p} h(f \star g_v) = g_p \tag{5}$$

$$\arg \min_{f:\, m_2(f)=p} h(f \star g_v) - h(f) = g_p \tag{6}$$

where the functions f are density functions on $\mathbb{R}$, i.e., positive integrable functions integrating to 1, and having a well-defined entropy and second moment $m_2(f) = \int_{\mathbb{R}} x^2 f(x)dx$.

We now consider the local analysis. We define densities of the form

$$f_\varepsilon(x) = g_p(x)(1 + \varepsilon L(x)), \quad x \in \mathbb{R}, \tag{7}$$

where $L : \mathbb{R} \to \mathbb{R}$ satisfies

$$\inf_{x\in\mathbb{R}} L(x) > -\infty \tag{8}$$

$$\int_{\mathbb{R}} L(x) g_p(x) dx = 0. \tag{9}$$

Hence, with these two constraints, f_ε is a valid density for ε sufficiently small. It is a perturbed Gaussian density, in a "direction" L. Observe that,

$$m_2(f_\varepsilon) = p \quad \text{iff} \quad M_2(L) = \int_{\mathbb{R}} x^2 L(x) g_p(x) dx = 0. \tag{10}$$

We are now interested in analyzing how these perturbations affects the output through an AWGN channel. Note that, if the input is a Gaussian g_p perturbed in the direction L, the output is a Gaussian g_{p+v} perturbed in the direction $\frac{(g_p L)\star g_v}{g_{p+v}}$, since $f_\varepsilon \star g_v = g_{p+v}(1 + \varepsilon \frac{(g_p L)\star g_v}{g_{p+v}})$.

Convention: $g_p L \star g_v$ refers to $(g_p L) \star g_v$, i.e., the multiplicative operator precedes the convolution one.

For simplicity, let us assume in the following that the function L is a polynomial satisfying (8), (9).

Lemma 1: We have

$$\begin{aligned} D(f_\varepsilon || g_p) &= \frac{1}{2}\varepsilon^2 \|L\|^2_{g_p} + o(\varepsilon^2) \\ D(f_\varepsilon \star g_v || g_p \star g_v) &= \frac{1}{2}\varepsilon^2 \|\frac{g_p L \star g_v}{g_{p+v}}\|^2_{g_{p+v}} + o(\varepsilon^2), \end{aligned}$$

where $\|L\|^2_{g_p} = \int_{\mathbb{R}} L^2(x) g_p(x) dx$.

Moreover, note that for any density f, if its first and second moments are $m_1(f) = a$ and $m_2(f) = p + a^2$, we have

$$h(f) = h(g_{a,p}) - D(f||g_{a,p}). \tag{11}$$

Hence, the extremal entropic results of (5) and (6) are locally expressed as

$$\arg \min_{L:\, M_2(L)=0} \|\frac{g_p L \star g_v}{g_{p+v}}\|^2_{g_{p+v}} = 0 \tag{12}$$

$$\arg \max_{L:\, M_2(L)=0} \|\frac{g_p L \star g_v}{g_{p+v}}\|^2_{g_{p+v}} - \|L\|^2_{g_p} = 0, \tag{13}$$

where 0 denotes here the zero function. If (12) is obvious, (13) requires a check which will be done in section V.

Now, if we want to make headway on the competitive situations presented in the introduction, we need more refined results than the ones above. Let us define the following mapping,

$$\Gamma^{(+)} : \ L \in L_2(g_p) \mapsto \frac{g_p L \star g_v}{g_{p+v}} \in L_2(g_{p+v}), \tag{14}$$

where $L_2(g_p)$ denotes the space of real functions having a finite $\|\cdot\|_{g_p}$ norm. This linear mapping gives, for a given perturbed direction L of a Gaussian input g_p, the resulting perturbed direction of the output through additive Gaussian noise g_v. The norm of each direction in their respective spaces, i.e., $L_2(g_p)$ and $L_2(g_{p+v})$, gives how far from the Gaussian distribution these perturbations are. Note that if L satisfies (8)-(9), so does $\Gamma^{(+)} L$ for the measure g_{p+v}. The result in (13) (worst noise case) tells us that this mapping is a contraction, but for our goal, what would be helpful is a spectral analysis of this operator, to allow more quantitative results than the extreme-case results of (12) and (13). In order to do so, one can express $\Gamma^{(+)}$ as an operator defined and valued in the same space, namely L_2 with the Lebesgue measure λ, which is done by inserting the Gaussian measure in the operator argument. We then proceed to a singular function/value analysis. Formally, let $K = L\sqrt{g_p}$, which gives $\|K\|_\lambda = \|L\|_{g_p}$, and let

$$\Lambda : \ K \in L_2(\lambda; \mathbb{R}) \mapsto \frac{\sqrt{g_p} K \star g_v}{\sqrt{g_{p+v}}} \in L_2(\lambda; \mathbb{R}) \tag{15}$$

which gives $\|\Gamma^{(+)}L\|_{g_{p+v}} = \|\Lambda K\|_\lambda$. We want to find the singular functions of Λ, i.e., denoting by Λ^t the adjoint operator of Λ, we want to find the eigenfunctions K of $\Lambda^t\Lambda$.

IV. Results

A. General Result

The following proposition gives the singular functions and values of the operator Λ defined in previous section.

Proposition 1:

$$\Lambda^t \Lambda K = \gamma K, \quad K \neq 0$$

holds for each pairs

$$(K,\gamma) \in \{(\sqrt{g_p}H_k^{[p]}, (\frac{p}{p+v})^k)\}_{k\geq 0},$$

where $H_k^{[p]}(x) = \frac{1}{\sqrt{k!}}H_k(x/\sqrt{p})$ and

$$H_k(x) = (-1)^k e^{x^2/2}\frac{d^k}{dx^k}e^{-x^2/2}, \quad k \geq 0, x \in \mathbb{R}.$$

The polynomials $H_k^{[p]}$ are the normalized Hermite polynomials (for a Gaussian distribution having variance p) and $\sqrt{g_p}H_k^{[p]}$ are called the Hermite functions. For any $p > 0$, $\{H_k^{[p]}\}_{k\geq 0}$ is an orthonormal basis of $L_2(g_p)$, this can be found in [4]. One can check that H_1, respectively H_2 perturb a Gaussian distribution into another Gaussian distribution, with a different first moment, respectively second moment. For $k \geq 3$, the H_k perturbations are not modifying the first two moments and are moving away from Gaussian distributions. Since $H_0^{[p]} = 1$, the orthogonality property implies that $H_k^{[p]}$ satisfies (9) for any $k > 0$. However, it is formally only for even values of k that (10) is verified (although we will see in section V that essentially any k can be considered in our problems). The following result contains the property of Hermite polynomials mostly used in our problems, and expresses Proposition 1 with the Gaussian measures.

Proposition 2:

$$\Gamma^{(+)}H_k^{[p]} = \frac{g_p H_k^{[p]} \star g_v}{g_{p+v}} = (\frac{p}{p+v})^{k/2}H_k^{[p+v]}, \qquad (16)$$

$$\Gamma^{(-)}H_k^{[p+v]} = H_k^{[p+v]} \star g_v = (\frac{p}{p+v})^{k/2}H_k^{[p]}. \qquad (17)$$

Last proposition implies proposition 1, since

$$\Gamma^{(-)}\Gamma^{(+)}L = \gamma L \quad \Longleftrightarrow \quad \Lambda^t\Lambda K = \gamma K$$

for $K = L\sqrt{g_p}$.

Comment: these properties of Hermite polynomials and Gaussian measures are likely to be already known, in a different context or with different notations. However, what is particularly interesting here, are not only these properties by themselves, but mainly the fact that they are precisely emerging from our information theoretic setting and are helpful to solve our problems.

In words, we just saw that H_k is an eigenfunction of the input/output perturbation operator $\Gamma^{(+)}$, in the sense that $\Gamma^{(+)}H_k^{[p]} = \left(\frac{p}{p+v}\right)^{k/2} H_k^{[p+v]}$. Hence, over an additive Gaussian noise channel g_v, if we perturb the input g_p in the direction $H_k^{[p]}$ by an amount ε, we will perturb the output g_{p+v} in the direction $H_k^{[p+v]}$ by an amount $\left(\frac{p}{p+v}\right)^{k/2}\varepsilon$. Such a perturbation in H_k implies that the output entropy is reduced (compared to not perturbing) by $\left(\frac{p}{p+v}\right)^{k}\frac{\varepsilon^2}{2}$ (if $k \geq 3$).

B. Fading Gaussian Broadcast Result

Proposition 3: Let $Y_i = HX + Z_i$, $i = 1,2$, with X such that $\mathbb{E}X^2 \leq p$, $Z_1 \sim \mathcal{N}(0,v)$, $0 < v < 1$ and $Z_2 \sim \mathcal{N}(0,1)$. There exists fading distributions and values of v for which the capacity achieving input distribution is non-Gaussian. More precisely, let U be any auxiliary random variable, with $U - X - (Y_1,Y_2)$. Then, there exists p, v, a distribution of H and μ such that

$$(U,X) \mapsto I(X;Y_1|U,H) + \mu I(U;Y_2|H) \qquad (18)$$

is maximized by a non jointly Gaussian distribution.

In the proof, we find a counter-example to Gaussian being optimal for H binary (and of course other counter-examples can be found). In order to defeat Gaussian codes, we construct codes using the Hermite coordinates. The proof also gives conditions on the fading distribution and the noise variance v for which these codes are strictly improving on Gaussian ones.

V. Proofs

We start by reviewing the proof of (13), as it brings interesting facts. We then prove the main result.

Proof of (13)*:*
We assume first that we insist on constraining f_ε to have zero mean and variance exactly p. Using the Hermite basis, we express L as $L = \sum_{k\geq 3}\alpha_k H_k^{[p]}$ (L must have such an expansion, since it must have a finite $L_2(g_p)$ norm, to make sense of the original expressions). Using (16), we can then express (13) as

$$\sum_{k\geq 3}\alpha_k^2\left(\frac{p}{p+v}\right)^k - \sum_{k\geq 3}\alpha_k^2 \qquad (19)$$

which is clearly negative. Hence, we have proved that

$$\|\frac{g_p L \star g_v}{g_{p+v}}\|^2_{g_{p+v}} \leq \|L\|^2_{g_p} \qquad (20)$$

and (13) is maximized by taking $L = 0$. Note that we can get tighter bounds than the one in previous inequality, indeed the tightest, holding for H_3, is given by

$$\|\frac{g_p L \star g_v}{g_{p+v}}\|^2_{g_{p+v}} \leq \left(\frac{p}{p+v}\right)^3 \|L\|^2_{g_p} \qquad (21)$$

(this clearly holds if written as a series like in (19)). Hence, locally the contraction property can be tightened, and locally, we have stronger EPI's, or worst noise case. Namely, if $\nu \geq \left(\frac{p}{p+v}\right)^3$, we have

$$\arg\min_{f:\, m_1(f)=0, m_2(f)=p} h(f\star g_v) - \nu h(f) = g_p \qquad (22)$$

and if $\nu < \left(\frac{p}{p+v}\right)^3$, g_p is outperformed by non-Gaussian distributions. Now, if we consider the constraint $m_2(f) \leq p$, which in particular, allows now to have $m_1(f) > 0$ and $m_2(f) = p$, we get that if $\nu \geq \frac{p}{p+v}$,

$$\arg \min_{f:\, m_2(f) \leq p} h(f \star g_v) - \nu h(f) = g_p \tag{23}$$

and if $\nu < \frac{p}{p+v}$, g_p is outperformed by $g_{p-\delta}$ for some $\delta > 0$. It would then be interesting to study if these tighter results hold in a greater generality than for the local setting.

Proof of Proposition 2:
We refer to (18) as the mu-rate. Let us first consider Gaussian codes, i.e., when (U, X) is jointly Gaussian, and see what mu-rate they can achieve. Without loss of generality, we can assume that $X = U + V$, with U and V independent and Gaussian, with respective variance Q and R satisfying $P = Q + R$. Then, (18) becomes

$$\frac{1}{2}\mathbb{E}\log(1 + \frac{RH^2}{v}) + \mu\frac{1}{2}\mathbb{E}\log\frac{1 + PH^2}{1 + RH^2}. \tag{24}$$

Now, we pick a μ and look for the optimal power R that should be allocated to V in order to maximize the above expression. We are interested in cases for which the optimal R is not at the boundary but at an extremum of (24), and if the maxima is unique, the optimal R is found by the first derivative check, which gives $\mathbb{E}\frac{H^2}{v+RH^2} = \mu\mathbb{E}\frac{H^2}{1+RH^2}$. Since we will look for μ, v, with $R > 0$, previous condition can be written as

$$\mathbb{E}\frac{RH^2}{v + RH^2} = \mu\mathbb{E}\frac{RH^2}{1 + RH^2}. \tag{25}$$

We now check if we can improve on (24) by moving away from the optimal jointly Gaussian (U, X). There are several ways to perturb (U, X), we consider here a first example. We keep U and V independent, but perturb them away from Gaussian's in the following way:

$$p_{U_\varepsilon}(u) = g_Q(u)(1 + \varepsilon(H_3^{[Q]}(u) + \delta H_4)) \tag{26}$$

$$p_{V_\varepsilon}(v) = g_R(v)(1 - \varepsilon(H_3^{[R]}(v) - \delta H_4)) \tag{27}$$

with $\varepsilon, \delta > 0$ small enough. Note that these are valid density functions and that they preserve the first two moments of U and V. The reason why we add δH_4, is to ensure that (10) is satisfied, but we will see that for our purpose, this can essentially be neglected. Then, using Lemma 2, the new distribution of X is given by

$$p_X(x) = g_P(x)(1 + \varepsilon\left(\frac{Q}{P}\right)^{\frac{3}{2}} H_3^{[P]} - \varepsilon\left(\frac{R}{P}\right)^{\frac{3}{2}} H_3^{[P]}) + f(\delta)$$

where $f(\delta) = \delta g_P(x)\varepsilon(\left(\frac{Q}{P}\right)^{\frac{4}{2}} H_4^{[P]} + \left(\frac{R}{P}\right)^{\frac{4}{2}} H_4^{[P]})$, which tends to zero when δ tends to zero. Now, by picking $P = 2R$, we have

$$p_X(x) \quad = \quad g_P(x) + f(\delta). \tag{28}$$

Hence, by taking δ arbitrarily small, the distribution of X is arbitrarily close to the Gaussian distribution with variance P.

We now want to evaluate how these Hermite perturbations perform, given that we want to maximize (18), i.e.,

$$h(Y_1|U, H) - h(Z_1) + \mu h(Y_2|H) - \mu h(Y_2|U, H). \tag{29}$$

We wonder if, by moving away from Gaussians, the gain achieved for the term $-h(Y_2|U, H)$ is higher than the loss suffered from the other terms. With the Hermite structure described in previous section, we are able to precisely measure this and we get

$$\begin{aligned} &h(Y_1|U = u, H = h) \\ &= h(g_{hu,v+Rh^2}(1 - \varepsilon\left(\frac{Rh^2}{v + Rh^2}\right)^{\frac{3}{2}} H_3^{[hu,v+Rh^2]})) \\ &= \frac{1}{2}\log 2\pi e(v + Rh^2) - \frac{\varepsilon^2}{2}\left(\frac{Rh^2}{v + Rh^2}\right)^3 + o(\varepsilon^2) + o(\delta) \end{aligned}$$

$$\begin{aligned} &h(Y_2|U = u, H = h) \\ &= \frac{1}{2}\log 2\pi e(1 + Rh^2) - \frac{\varepsilon^2}{2}\left(\frac{Rh^2}{1 + Rh^2}\right)^3 + o(\varepsilon^2) + o(\delta) \end{aligned}$$

and because of (28)

$$h(Y_2|H = h) = \frac{1}{2}\log 2\pi e(1 + Ph^2) + o(\varepsilon^2) + o(\delta).$$

Therefore, collecting all terms, we find that for U_ε and V_ε defined in (26) and (27), expression (29) reduces to

$$\begin{aligned} I_G - \frac{\varepsilon^2}{2}\mathbb{E}\left(\frac{RH^2}{v + RH^2}\right)^3 + \mu\frac{\varepsilon^2}{2}\mathbb{E}\left(\frac{RH^2}{1 + RH^2}\right)^3 \\ + o(\varepsilon^2) + o(\delta) \end{aligned} \tag{30}$$

where I_G is equal to (24) (which is the mu-rate obtained with Gaussian inputs). Hence, if for some distribution of H and some v, we have that

$$\mu\mathbb{E}\left(\frac{RH^2}{1 + RH^2}\right)^k - \mathbb{E}\left(\frac{RH^2}{v + RH^2}\right)^k > 0, \tag{31}$$

when $k = 3$ and R is optimal for μ, we can take ε and δ small enough in order to make (30) strictly larger than I_G. We have shown how, if verified, inequality (31) leads to counter-examples of the Gaussian optimality, but with similar expansions, we would also get counter-examples if the following inequality holds for any power k instead of 3, as long as $k \geq 3$. Let us summarize what we obtained: Let R be optimal for μ, which means that (25) holds if there is only one maxima (not at the boarder). Then, non-Gaussian codes along Hermite's strictly outperforms Gaussian codes, if, for some $k \geq 3$, (31) holds. If the maxima is unique, this becomes

$$\frac{\mathbb{E}T(v)^k}{\mathbb{E}T(1)^k} < \frac{\mathbb{E}T(v)}{\mathbb{E}T(1)}$$

where

$$T(v) = \frac{RH^2}{v + RH^2}.$$

So we want the Jensen gap of $T(v)$ for the power k to be small enough compared to the Jensen gap of $T(1)$.

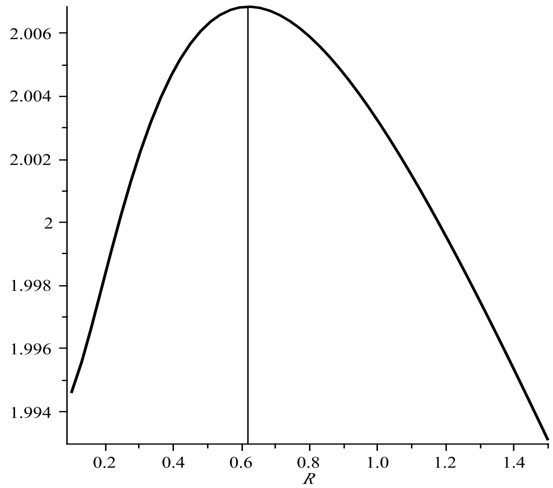

Fig. 1. Gaussian mu-rate, i.e., expression (24), plotted as a function of R for $\mu = 5/4$, $v = 1/4$, $P = 1.24086308$ and H binary $\{1; 10\}$. Maxima at $R = 0.62043154$.

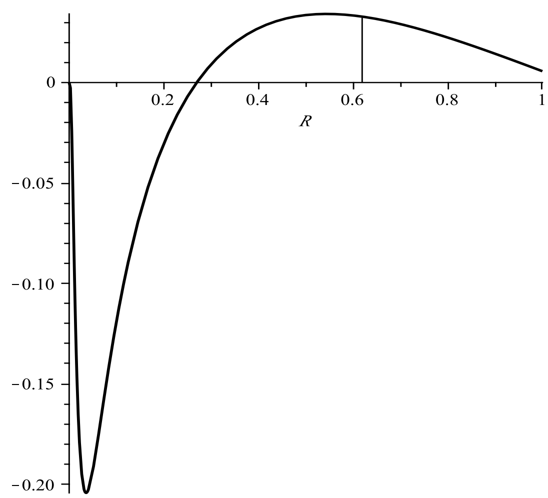

Fig. 2. LHS of (31) as a function of R, for $\mu = 5/4$, $v = 1/4$, $k = 8$ and H binary $\{1; 10\}$, positive at $R = 0.62043154$.

We now give an example of fading distribution for which the above conditions can be verified. Let H be binary, taking values 1 and 10 with probability half and let $v = 1/4$. Let $\mu = 5/4$, then for any values of P, the maximizer of (24) is at $R = 0.62043154$, cf. Figure 1, which corresponds in this case to the unique value of R for which (25) is satisfied. Hence if P is larger than this value of R, there is a corresponding fading BC for which the best Gaussian code splits the power on U and V with $R = 0.62043154$ to achieve the best mu-rate with $\mu = 5/4$. To fit the counter-examples with the choice of Hermite perturbations made previously, we pick $P = 2R$. Finally, for these values of μ and R, (31) can be verified for $k = 8$, cf. Figure 2, and the corresponding Hermite code (along H_8) strictly outperforms any Gaussian codes.

Note that we can consider other non-Gaussian encoders, such as when U and V are independent with U Gaussian and V non-Gaussian along Hermite's. In this case, we get a different condition than (31), which is stronger in general for fixed values of the parameters, but which can still be verified, making V non-Gaussian strictly better than Gaussian.

VI. Discussion

We have introduced the use of encoders drawn from non-Gaussian distributions along Hermite polynomials. If the performance of non-Gaussian inputs is usually hard to analyze, we showed how the neighborhoods of Gaussian inputs can be analytically analyzed by use of the Hermite coordinates. This allowed us to use nuanced version of the usual extremal entropic results, and in particular, to show that Gaussian inputs are in general not optimal for degraded fading Gaussian BC, although they might still be optimal for many fading distributions. The Hermite technique provides not only potential counter-examples to optimality of Gaussian inputs but it also gives insight on problems for which a competitive situations does not imply the obvious optimality of Gaussian inputs. For example, in the considered problem, the Hermite technique gives a condition on what kind of fading distribution and degradedness (values of v) non-Gaussian inputs must be used. It also tells us how, locally, the optimal encoder is defined. In this paper, we considered fading BC's, but many other competitive situations can be tackled with this tool, particularly, since a multi-letter generalization of the current technique can be carried out (to appear). Finally, in a different context, local results could be "lifted" to corresponding global results in [1]. There, the localization is made with respect to the channels and not the input distribution, yet, it would be interesting to compare the local with the global behavior for the current problem too. A work in progress aims to replace the local neighborhood with a global one, consisting of all sub-Gaussian distributions.

Acknowledgment

The authors would like to thank Tie Liu and Shlomo Shamai for pointing out problems relevant to the application of the proposed technic, as well as Daniel Stroock for discussions on Hermite polynomials. For the first author, this work has partially been done under a postdoctoral fellowship at LIDS, MIT.

References

[1] E. Abbe and L. Zheng, "Linear universal decoding: a local to global geometric approach", *Submitted IEEE Trans. Inform. Theory*, 2008.
[2] T. M. Cover, "Comments on broadcast channel", *IEEE Trans. on Inform. Theory* vol. 44, n. 6, pp. 2524-2530, Oct. 1998.
[3] T. M. Cover and J. A. Thomas, "Elements of Information Theory", John Wiley & Sons, New York, NY, 1991.
[4] D. W. Stroock, "A Concise Introduction to the Theory of Integration", Third Edition, Birkhäuser, 1999.
[5] D. Tuninetti and S. Shamai, "On two-user fading Gaussian broadcast channels with perfect channel state information at the receivers", *Int. Symp. on Inform. Theory (ISIT)*, 2003.