Coding Along Hermite Polynomials for Interference Channels

Emmanuel A. Abbe
Lizhong Zheng

***Abstract*— This paper analyzes the use of non-Gaussian input distributions over the Gaussian interference channel. It has been recently proved that the iid Gaussian code ensemble together with a decoder that treats interference as noise is sum-capacity achieving, if the interference is below a threshold. We show that, when the decoder treats interference as noise, and when the interference is above a threshold, the iid Gaussian ensemble can be strictly improved upon. In the block synchronous setting, the improvement is obtained by a Gaussian but non iid ensemble, whereas in the asynchronous setting, it is obtained from an iid but non Gaussian ensemble. The analysis of non-Gaussian ensembles is made possible by the use of the Hermite coordinate system.**

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), cf. [8], 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 [5], 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 input distributions more complex. First, the interference. The fact that there are other users in the system brings an interesting "frustration phenomena". Since the multiple users produce interference on each other, it is no longer clear that Gaussian inputs are the best input distributions for the overall system. Indeed, let us assume that there are two users and that the first one decides to use iid Gaussian inputs. Then the second user should use Gaussian inputs to maximize his own rate, but this would also cause the most harmful interference on the first user, if he treats interference as noise. How to treat the interference and how to optimally construct codes on such channels is still an open problem. Then comes the fading. Over a single-user AWGN channel with coherent fading, it is clear that maximizing $I(X;Y|H)$ is achieved by a Gaussian input, according to (1). However, even for the broadcast AWGN channel with coherent fading, it has been an open problem whether Gaussian inputs are still optimal or not. A reason for these open questions, is that Gaussian inputs are pretty much the only ones that can be analyzed over Gaussian noise channels, as non-Gaussian inputs leave most problems in an intractable form. In [1], a novel technique has been developed to analyze a class of non-Gaussian input distributions over Gaussian noise channels. It allowed in particular to find non-Gaussian input distributions that outperform Gaussian ones over the broadcast AWGN channel with coherent fading. In this paper, we focus on the other phenomena mentioned above, namely, the interference.

II. Problem Statement

We consider the symmetric memoryless interference channel (IC) with two users and white Gaussian noise. We denote the input of each user by X_1 and X_2, their power constraint by p, the interference/coupling coefficients a, and the respective noise Z_1 and Z_2 (independent standard Gaussian). We define the following expression

$$\begin{aligned} &S_{a,p}(X_1,X_2) \\ &= I(X_1;X_1+aX_2+Z_1) + I(X_2;X_2+aX_1+Z_2) \\ &= h(X_1+aX_2+Z_1) - h(aX_2+Z_1) \\ &\quad + h(X_2+aX_1+Z_2) - h(aX_1+Z_2), \end{aligned} \tag{4}$$

where X_1 and X_2 can be random vectors, with a covariance having a trace bounded by p. For any dimension of these random vectors, $S_{a,p}(X_1,X_2)$ is a lower bound to the sum-capacity. Moreover, it is tight by taking X_1 and X_2 of arbitrarily large block length. Now, to optimize this expression, a competitive situation takes place: Gaussian inputs maximize each entropy term, but we have different signs in front of these entropy terms. Of course, if $a = 0$, we have two parallel AWGN channels with no interference, and Gaussian inputs are optimal. We can then expect that this might still hold for small values of a. Indeed, it has been proved recently in [2], [6], [7], that the sum-capacity is achieved by treating interference as noise[1] and with iid Gaussian inputs as long as $pa^3 + a - 1/2 \leq 0$. On the other hand, if a gets larger and larger, there should be a moment for which treating interference as noise with iid Gaussian inputs becomes penalizing (we know this happens if $a \geq 1$). So what can we deduce from studying the maximizers of (4)? A random encoder can be drawn from a distribution which does not maximize (4) for any block length, but yet, a decoder may exist in order to have a capacity achieving code. For example, if $a \geq 1$, iid Gaussian inputs will achieve the sum-capacity if the receiver decodes both users (one can show that the problem is equivalent to having two MAC's). However, if $a \geq 1$, the iid Gaussian distribution does not maximize $S_{a,p}(X_1, X_2)$ (for the dimension 1, hence for arbitrary dimensions). However, if the Gaussian distribution does not maximize (4) for the dimension 1, it means that iid Gaussian inputs and treating interference as noise is not capacity achieving. The new maximizer which is not iid Gaussian may then achieve capacity by treating interference as noise. We know that this cannot happen if $pa^3 + a - 1/2 \leq 0$, but it is not known when and if this can happen for $pa^3 + a - 1/2 > 0$. Moreover, a code which treats interference as noise and whose encoder is drawn from a distribution can be capacity achieving only if the encoder is drawn from a distribution maximizing (4). Otherwise, we can encode from the distribution maximizing (4), treat interference as noise and achieve higher rates. Also, there might be a relationship between treating interference as noise and using Gaussian inputs. If the Gaussian inputs are no longer optimal, it may suggest that treating interference as noise is too harmful, in view of the game equilibrium of (1) and (3). Let us assume that a distribution defeats the iid Gaussian one in (4) when $pa^3 + a - 1/2 > 0$ and let a be sufficiently close to the root of $pa^3 + a - 1/2 = 0$ form the right. If by any arguments, the encoder drawn from the iid Gaussian distribution could be shown to still have a decoder which achieves the sum-capacity in this infinitesimal range, we would have that treating interference as noise is not optimal.

Hence, we are interested in finding for which regime of a, iid Gaussian inputs are no longer the maximizers of (4). We then distinguish the implication of such a threshold in both the synchronized and asynchronized users setting, as there will be an interesting distinction between these two cases. We recall how the synch and asynch settings are defined here. In the synch setting, each user of the IC sends their code word of a common block length n simultaneously, i.e., at time 1, they both send the first component of their code word, at time 2 the second component, etc. In the asynch setting, each user is still using code words of the same block length n, however, there might be a shift between the time at which the first and second users start sending their code words. We denote this shift by τ, and assume w.l.o.g. that $0 \leq \tau \leq n$. In the totally asynch setting, we assume that τ is drawn uniformly at random within $\{0, \ldots, n\}$. Note that if iid inputs are used and interference is treated as noise, whether the users are synch or asynch is not affecting the rate achievability, and (4) with an iid distribution for X_1 and X_2 can still be achieved for the totally asynch IC. However, if the users want to time-share over the channel uses, such as to fully avoid their interference, they will need synchronization.

Definition 1: Time sharing over a block length n (assumed to be even) with Gaussian inputs refers to using X_1 Gaussian with covariance $2P\hat{I}_{n/2}$ and X_2 Gaussian with covariance $2P\hat{I}^c_{n/2}$, where $\hat{I}_{n/2}$ is a diagonal matrix with $n/2$ 1's and 0's, and $\hat{I}^c_{n/2}$ flips the 1's and 0's on the diagonal.

Note that the time sharing requires a perfect synchronization of the users. We will see in section V that a blind time-sharing which allows to partially avoid interference can still be achieved for the totally asynch IC.

III. Local Geometry and Hermite coordinates

In this section, we present the results which will justify the use of non-Gaussian coding along Hermite polynomials. Roughly speaking, we consider distributions which are in the neighborhoods of Gaussian distributions. But the main result, consist in identifying an orthogonal coordinate system for these neighborhoods which is invariant under the addition of Gaussian noise. This is now explained in more details.

We first express (1) and (3) as follows,

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

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

where g_p denotes the Gaussian density with zero mean and variance p, and 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 geometry by looking at densities of the form

$$f_\varepsilon = g_p(1 + \varepsilon L), \tag{7}$$

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,

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

[1] treating interference as noise, means to use decoders that only require the knowledge of the other user's code book distribution but not the actual code book

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

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

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

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

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

where 0 denotes here the zero function. If (12) is obvious, (13) requires a non-trivial proof.

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

$$\Gamma^{(+)}: \quad L \in L_2(g_p; \mathbb{R}) \mapsto \frac{g_p L \star g_v}{g_{p+v}} \in L_2(g_{p+v}; \mathbb{R}). \quad (14)$$

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., $L_2(g_p; \mathbb{R})$ and $L_2(g_{p+v}; \mathbb{R})$, 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, 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

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

which gives $\|\Gamma^{(+)} L\|_{g_{p+v}} = \|\Lambda K\|_\lambda$. We have the following.

Proposition 1:

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

holds for each pair

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

where

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

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

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 $L_2(g_p; \mathbb{R})$, this can be found in [4]. Moreover, 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. However, it is formally only for even values of k that (10) is verified (this will be further discussed).

The following result is a version of Proposition 1 with the Gaussian measures.

Proposition 2:

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

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

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 (must) 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.

We now rewrite the results in the way we will mostly use them in our problems.

Theorem 1: H_k is an eigenfunction of the input/output perturbation operator $\Gamma^{(+)}$, in the sense that

$$\Gamma^{(+)} H_k^{[p]} = \left(\frac{p}{p+v}\right)^{k/2} H_k^{[p+v]}. \quad (18)$$

The corresponding eigenvalue is the input/output power ratio $\frac{p}{p+v}$ to the power $k/2$.

In words, we just saw that, 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 $k \geq 3$, but also for $k = 1$ as we will see in section IV).

IV. Hermite Coding: Formalities

With previous results, we will use the Hermite polynomials as our directions to perturb Gaussians. The Hermite polynomial corresponding to $k = 0$ is $H_0^{[p]} = 1$ and is clearly not a valid direction as it violates (9). But $H_0^{[p]} = 1$ also implies, from the orthogonality property of the Hermite basis, that $H_k^{[p]}$ satisfies (9) for any $k > 0$. However, it is only for k even that $H_k^{[p]}$ satisfies (8). On the other hand, for any $\delta > 0$, we have that $H_k^{[p]} + \delta H_{4k}^{[p]}$ satisfies (8), whether k is even or not. Now, if we consider the direction $-H_k^{[p]}$, (8) is not satisfied for both k even and odd. But again, for any $\delta > 0$, we have that $-H_k^{[p]} + \delta H_{4k}^{[p]}$ satisfies (8). Hence, in order to ensure (8), we will often work in the proofs with $\pm H_k^{[p]} + \delta H_{4k}^{[p]}$, although it will essentially allow us to reach the performance achieved by any $\pm H_k^{[p]}$ (odd or even), since we will then take δ arbitrarily small and use continuity arguments.

Convention: We will drop the variance upper script in the Hermite terms whenever a Gaussian density with specified variance is perturbed, i.e., the density $g_p(1 + \varepsilon H_k)$ always denotes $g_p(1+\varepsilon H_k^{[p]})$, no matter what p is. We can afford such a notation since Lemma 2 will always force the Hermite term to have the right variance in our problems. Same treatment is done for $\|\cdot\|$.

Now, in order to evaluate the entropy of a perturbation, i.e., $h(g_p(1+\varepsilon L))$, we can express it as the entropy of $h(g_p)$ minus the divergence gap, as in (11), and then use Lemma 1 for the approximation. But this is correct *if* $g_p(1+\varepsilon L)$ has the same first two moments as g_p. Hence, if L contains only H_k's with $k \geq 3$, the previous argument can be used. But if L contains H_1 and/or H_2 terms, the situation can be different. Next Lemma describes this.

Lemma 2: Let $\delta > 0$ and

$$b\widetilde{H}_k = \begin{cases} b(H_k + \delta H_{4k}), & \text{if } b \geq 0, \\ b(H_k - \delta H_{4k}), & \text{if } b < 0. \end{cases} \tag{19}$$

We have for any $\alpha_k \in \mathbb{R}$, $k \geq 1$, $\varepsilon > 0$

$$h(g_p(1+\varepsilon\sum_{k\geq 1}\alpha_k\widetilde{H}_k)) = h(g_p) - D(g_p(1+\varepsilon\sum_{k\geq 1}\alpha_k\widetilde{H}_k)||g_p) + \frac{\varepsilon\alpha_2}{\sqrt{2}}.$$

Finally, when we convolve two perturbed Gaussian distributions, we get $g_a(1+\varepsilon H_j)\star g_b(1+\varepsilon H_k) = g_{a+b} + \varepsilon[g_a H_j \star g_b + g_a \star g_b H_k] + \varepsilon^2 g_a H_j \star g_n H_k$. We already know from Theorem 1 how to describe the terms in ε, what we still need is to describe the term in ε^2. We have the following.

Lemma 3:

$$g_a H_1^{[a]} \star g_b H_1^{[b]} = \frac{\sqrt{2ab}}{a+b} g_{a+b} H_2^{[a+b]}$$

and

$$g_a H_k^{[a]} \star g_b H_l^{[b]} = C g_{a+b} H_{k+l}^{[a+b]}$$

where C is a constant depending only on a, b, k and l.

V. Results for the IC

Definition 2: Let

$$F_k(a,p) = \lim_{\delta\searrow 0}\lim_{\varepsilon\searrow 0}\frac{2}{\varepsilon^2}\left[S_{a,p}(X_1,X_2) - S_{a,p}(X_1^G,X_2^G)\right],$$

where $X_1 \sim g_p(1+\varepsilon\widetilde{H}_k)$, $X_2 \sim g_p(1-\varepsilon\widetilde{H}_k)$, with $\widetilde{H}_k$ defined in (19), and $X_1^G, X_2^G \overset{\text{iid}}{\sim} g_p$.

That is, $F_k(a,p)$ represents the gain (positive or negative) of using X_1 perturbed along H_k and X_2 perturbed along $-H_k$ with respect to using Gaussian distributions.

Theorem 2: We have for $k \geq 2$

$$F_k(a,p) = [a^2(1+p+pa^2)]^k - (1-a^k)^2(1+pa^2)^k.$$

For any fixed p, the function $F_k(\cdot,p)$ has a unique positive root, below which it is negative and above which it is positive.

Proposition 3: Treating interference as noise with iid Gaussian inputs does not achieve the sum-capacity of the symmetric IC (synch or asynch) and is outperformed by $X_1 \sim g_p(1+\varepsilon\widetilde{H}_3)$ and $X_2 \sim g_p(1-\varepsilon\widetilde{H}_3)$, if

$$F_3(a,p) > 0.$$

Proposition 4: For the symmetric synch IC, when treating interference as noise, time sharing improves on the iid Gaussian distribution if

$$F_2(a,p) > 0.$$

We now introduce the following definition.

Definition 3: Blind time sharing over a block length n (assumed to be even) with Gaussian inputs refers to using X_1 Gaussian with covariance $2P\text{diag}(10101010\ldots10)$ and X_2 Gaussian with covariance $2P\text{diag}(1,\ldots,1,0,\ldots,0)$. Note that this blind time sharing does not require any synchronization of the users, since no matter what τ is, it achieves the same rate in (4): the users are interfering in $n/4$ channel uses and have each a non-intefering channel in $n/4$ channel uses.

Proposition 5: For the symmetric asynch IC, when treating interference as noise, blind time sharing improves on the iid Gaussian distribution if

$$B_2(a,p) > 0,$$

where $B_2(a,p) = \frac{1}{4}(\log(1+2p) + \log(1+\frac{2p}{1+2a^2p})) - \log(1+\frac{p}{1+a^2p})$.

How to read these results: We have four thresholds to keep track of. When treating interference as noise we have:

- $T_1(p)$ is when $pa^3 + a - \frac{1}{2} = 0$. If $a \leq T_1(p)$, we know from [2], [6], [7] that iid Gaussian inputs are sum-capacity achieving[2].
- $T_2(p)$ is when $F_2(a,p) = 0$. If $a > T_2(p)$, we know from Prop. 4 that, if synchronization is permitted, time sharing improves on iid Gaussian inputs. This regime matches with the so-called moderate regime defined in [3].
- $T_3(p)$ is when $F_3(a,p) = 0$. If $a > T_3(p)$, we know from Prop. 3 that non-Gaussian distributions (opposites in H_3) improve on iid Gaussian inputs.
- $T_4(p)$ is when $B_2(a,p) = 0$. If $a > T_4(p)$, we know from Prop. 5 that, even if the users are totally asynchronized, blind time sharing improves on iid Gaussian inputs.

The question is now, how are these thresholds ranked. It turns out that

$$0 < T_1(p) < T_2(p) < T_3(p) < T_4(p).$$

And if $p = 1$, the above inequality reads as $0.424 < 0.605 < 0.680 < 1.031$. This implies the following for a decoder that treats interference as noise. Since $T_2(p) < T_3(p)$, it is first better to time share than using non-Gaussian distributions along H_3. But this is useful only if time-sharing is permitted, i.e., for the synch IC. However, for the asynchronized IC, since $T_3(p) < T_4(p)$, we are better off using the non-Gaussian distributions along H_3 before a Gaussian input scheme, even with blind time-sharing. Finally, we notice that there is still a gap between $T_1(p)$ and $T_2(p)$, and we cannot say if, in this range, iid Gaussian inputs are still optimal, or if another class of non-Gaussian inputs (far away from Gaussians) can outperform them. However, we have noticed something mysterious. In theorem 2, we asked for $k \geq 2$. Now, we can still wonder what the inequality achieved by requiring the expression of theorem 2 to be positive looks like for $k = 1$. And it turns out that this is precisely $pa^3 + a - \frac{1}{2} > 0$, i.e., the complement range delimited by $T_1(p)$. But the right hand side of theorem 2 for $k = 1$ is not equal to $F_1(a,p)$, and indeed, it would not make sense that moving along H_1, which changes the mean with a fixed second moment within Gaussians, would allow us to improve on the iid Gaussian scheme. Yet, getting to the exact same condition, when working on the problem of improving on the iid Gaussian scheme, seems to be a strange coincidence.

VI. Proofs

Proof of Proposition 3:
Let $\varepsilon, \delta > 0$ and let X_1 and X_2 be respectively distributed as $g_p(1 + \varepsilon[H_k + \delta H_{4k}])$ and $g_p(1 - \varepsilon[H_k - \delta H_{4k}])$, where $k \neq 1, 2$. We have

$$I(X_1, X_1 + aX_2 + Z_1) = h(X_1 + aX_2 + Z_1) - h(aX_2 + Z_1)$$

where X_k^G are independent Gaussian 0-mean and p-variance random variables. Let us first analyze $h(X_1 + aX_2 + Z_1)$. The density of $X_1 + aX_2 + Z_1$ is given by

$$g_p(1 + \varepsilon[H_k + \delta H_{4k}]) \star g_{a^2p}(1 - \varepsilon[H_k - \delta H_{4k}]) \star g_1, \quad (20)$$

which, from Theorem 1, is equal to

$$\begin{aligned} &g_{p+a^2p+1}(1+ \\ &\varepsilon\{\left[\left(\frac{p}{p+a^2p+1}\right)^{\frac{k}{2}} H_k + \delta\left(\frac{p}{p+a^2p+1}\right)^{2k} H_{4k}\right] \\ &-\left[\left(\frac{a^2p}{p+a^2p+1}\right)^{\frac{k}{2}} H_k - \delta\left(\frac{a^2p}{p+a^2p+1}\right)^{2k} H_{4k}\right] \\ &-\varepsilon L\}) \end{aligned}$$

where

$$L = \frac{g_p[H_k + \delta H_{4k}] \star g_{a^2p}[H_k - \delta H_{4k}] \star g_1}{g_{p+a^2p+1}}.$$

Note that each direction in each line of the $\{\cdot\}$ bracket above, including L, satisfy (8) and (9). Using Lemma 3, we have

$$\begin{aligned} L &= \frac{g_p[H_k + \delta H_{4k}] \star g_{a^2p+1}[\left(\frac{a^2p}{a^2p+1}\right)^{k/2} H_k - \left(\frac{a^2p}{a^2p+1}\right)^{2k} \delta H_{4k}]}{g_{p+a^2p+1}} \\ &= C_1 H_{2k}^{[p+a^2p+1]} + C_2 H_{5k}^{[p+a^2p+1]} + C_3 H_{8k}^{[p+a^2p+1]}, \end{aligned} \quad (21)$$

where C_1, C_2, C_3 are constants. Therefore, the density of $X_1 + aX_2 + Z_1$ is a Gaussian g_{p+a^2p+1} perturbed by the direction H_k in the order ε and several H_l with $l \geq 2k$ in the order ε^2 (and other directions but that have a δ order).
So we can use Lemma 2 and write

$$\begin{aligned} h(X_1 + aX_2 + Z_1) &= h(X_1^G + aX_2^G + Z_1) \\ &- D(X_1 + aX_2 + Z_1 || X_1^G + aX_2^G + Z_1) \end{aligned}$$

and using Lemma 1, we have

$$\begin{aligned} &D(X_1 + aX_2 + Z_1 || X_1^G + aX_2^G + Z_1) \\ &\doteq \frac{\varepsilon^2}{2} \| \left[\left(\frac{p}{p+a^2p+1}\right)^{\frac{k}{2}} H_k + \delta\left(\frac{p}{p+a^2p+1}\right)^{2k} H_{4k}\right] \\ &-\left[\left(\frac{a^2p}{p+a^2p+1}\right)^{\frac{k}{2}} H_k - \delta\left(\frac{a^2p}{p+a^2p+1}\right)^{2k} H_{4k}\right] \|^2 \\ &= \frac{\varepsilon^2}{2}(1-a^k)^2\left(\frac{p}{p+a^2p+1}\right)^k + \frac{\varepsilon^2}{2} o(\delta). \end{aligned}$$

Hence

$$\begin{aligned} h(X_1 + aX_2 + Z_1) &= h(X_1^G + aX_2^G + Z_1) \\ &- \frac{\varepsilon^2}{2}(1-a^k)^2\left(\frac{p}{p+a^2p+1}\right)^k + \frac{\varepsilon^2}{2} o(\delta). \end{aligned}$$

More directly, we get

$$D(aX_2 + Z_1 || aX_2^G + Z_1) \doteq \frac{\varepsilon^2}{2}\left(\frac{a^2p}{a^2p+1}\right)^k + \frac{\varepsilon^2}{2} o(\delta).$$

[2]this holds for both the synch and asynch settings, since this coding scheme does not requires synchronization

and

$$I(X_1, X_1 + aX_2 + Z_1) \doteq I(X_1^G, X_1^G + aX_2^G + Z_1) + \frac{\varepsilon^2}{2}\left[\left(\frac{a^2 p}{a^2 p + 1}\right)^k - (1 - a^k)^2\left(\frac{p}{p + a^2 p + 1}\right)^k\right] + \frac{\varepsilon^2}{2} o(\delta).$$

Finally, we have

$$I(X_2, X_2 + aX_1 + Z_2) = I(X_1, X_1 + aX_2 + Z_1)$$

and

$$I(X_1, X_1 + aX_2 + Z_1) + I(X_2, X_2 + aX_1 + Z_2) \doteq I(X_1^G, X_1^G + aX_2^G + Z_1) + I(X_2^G, X_2^G + aX_1^G + Z_2) + \varepsilon^2\left[\left(\frac{a^2 p}{a^2 p + 1}\right)^k - (1 - a^k)^2\left(\frac{p}{p + a^2 p + 1}\right)^k\right] + \varepsilon^2 o(\delta).$$

Hence, if for some $k \neq 3$ we have

$$\left(\frac{a^2}{a^2 p + 1}\right)^k - \frac{(a^k - 1)^2}{(p + a^2 p + 1)^k} > 0$$

we can improve on the iid Gaussian distributions g_p by using the respective Hermite perturbations.

Now, we could have started with X_1 and X_2 distributed as $g_p(1+\varepsilon b_k \widetilde{H}_k)$ and $g_p(1+\varepsilon c_k \widetilde{H}_k)$, where $b\widetilde{H}_k = b[H_k + H_{4k}]$ if $b \geq 0$ and $b\widetilde{H}_k = b[H_k - H_{4k}]$ otherwise. With similar expansions, we would then get that we can improve on the Gaussian distributions if for some some b_k, c_k and $k \neq 2$ we have

$$\left[\left(\frac{a^2 p}{a^2 p + 1}\right)^k - \frac{(a^2 p)^k + p^k}{(p + a^2 p + 1)^k}\right](b_k^2 + c_k^2) - 4\left(\frac{ap}{p + a^2 p + 1}\right)^k b_k c_k > 0.$$

But the quadratic function

$$(b, c) \in \mathbb{R}^2 \mapsto \gamma(b^2 + c^2) - 2\delta bc,$$

with $\delta > 0$, can be made positive if and only if $\gamma + \delta > 0$, and is made so by taking $b_k = -c_k$. Hence, the initial choice we made about X_1 and X_2 is optimal. Moreover, note that for this distribution of X_1 and X_2, we could have actually chosen $k = 2$ as well. Because, even if Lemma 2 tells us that we must use correction terms, these correction terms will cancel out when we consider the sum-rate, since $b_k = -c_k$ and since the correction is in ε. There is however another problem with using $k = 2$, which is that $g_p(1+\varepsilon H_2)$ has a larger second moment than p. However, if we use a scheme of block length 2, we can compensate this excess on the first channel use with the second channel use, and because of the symmetry, we can achieve the desired rate. But this is allowed only with synchronization. We could also have used perturbations that are mixtures of Hermite's, such as $g_p(1 + \varepsilon \sum_k b_k H_k)$. We would then get mixtures of previous equations as our condition. But in the current problem this will not be helpful. Finally, perturbing iid Gaussian inputs in a non iid way, i.e., to perturb different components in different Hermite directions, cannot improve on our scheme, from previous arguments. The only option which is not investigated here, is to perturb iid Gaussian inputs in a non independent manner.

Putting everything together, we have that if for some $k \geq 2$

$$\left(\frac{a^2}{a^2 p + 1}\right)^k - \frac{(a^k - 1)^2}{(p + a^2 p + 1)^k} > 0, \qquad (22)$$

using iid g_p Gaussian distributions is not optimal and can be defeated with the corresponding Hermite perturbations. Note that, if we work with $k = 1$, the proof sees the following modification. In (21), we now have a term in H_2. However, even if this term is in the order ε^2, we can no longer neglect it, since from Lemma 2, a $\varepsilon^2 H_2$ term in the direction comes out as a $\frac{\varepsilon^2}{\sqrt{2}}$ term in the entropy. Hence, we do not get the above condition for $k = 1$, but the one obtained by replacing $(a^k - 1)^2$ with $(a^2 + 1)$, and the condition for positivity can never be fulfilled.

VII. Discussion

We have used encoders drawn from non-Gaussian distributions along Hermite polynomials as introduced in [1]. 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 the means of the Hermite coordinates. We found in this paper that using non-Gaussian input distributions (along H_3) can strictly improve on the Gaussian distribution for the asynch IC, when treating interference as noise. We also recovered the threshold of the moderate regime by using H_2 perturbations in the synch setting, showing that this global threshold is reflected in our local setting. We also met mysteriously in our local setting the other global threshold found in [2], [6], [7], below which treating interference as noise with iid Gaussian inputs is optimal. We hope to understand this better with the work in progress.

The Hermite technique provides not only potential counter-examples to optimality of Gaussian inputs but it also gives insight on how the optimal distributions may look like. For example, in the IC problem, the perturbation in H_3 are better be done in an opposite manner for the two users, so has to make the mass of their distributions moving apart from each other. This may suggest a structure on the optimal distributions and would require more investigations. Finally, it would be interesting to compare the results obtained with the local analysis and the global ones. The fact that we have observed global results locally, as mentioned previously, gives hope for possible local to global extensions. Monotone paths, or variational calculus could then be helpful to go from local to global results. But another approach would be to consider all distributions with tails that are not heavier than the Gaussian ones. Then, we could described these distributions with Hermite polynomials and use similar results as presented here. Finally, an approach based on differential equations can also be considered. This is all work in progress.

Acknowledgment

References

[1] E. Abbe and L. Zheng, “Non-Gaussian Coding Along Hermite Polynomials for Fading Broadcast Gaussian Channel”, 2009.

[2] V.S. Annapureddy and V.V. Veeravalli, “Gaussian Interference Network: sum Capacity in the Low Interference Regime and New Outer Bounds on the Capacity Region”, *Submitted IEEE Trans. Inform. Theory*, 2008.

[3] M. Costa, ”On the Gaussian interference channel,” IEEE Trans. Inform. Theory, vol. IT-31, pp. 607-615, Sept. 1985.

[4] R. Courant and D. Hilbert, “Methods of Mathematical Physics”, Volume I, Wiley-Interscience, 1953.

[5] T. M. Cover and J. A. Thomas, “Elements of Information Theory”, John Wiley & Sons, New York, NY, 1991.

[6] A. S. Motahari, A. K. Khandani, “Capacity Bounds for the Gaussian Interference Channel”, CoRR abs/0801.1306: (2008).

[7] X. Shang, G. Kramer, B. Chen, “Outer bound and noisy-interference sum-rate capacity for symmetric Gaussian interference channels”, CISS 2008: 385-389.

[8] A.J. Stam, ”Some inequalities satisfied by the quantities of information of Fisher and Shannon”, *Information and Control*, 2: 101-112, 1959.