On the Discreteness of the Capacity Achieving Measure of an Input-bounded Vector Gaussian Channel

Terence H. Chan[1], Steve Hranilovic, Frank R. Kschischang
Dept. of Electrical & Computer Engineering
University of Toronto
{terence, steve, frank}@comm.utoronto.ca

Abstract — In this work, we study the capacity-achieving input distribution of a discrete-time memoryless vector Gaussian noise channel subject to second-order average cost constraints, and input-bounded constraints. By generalizing the results of Smith and Shamai for the scalar additive Gaussian noise channel and the quadrature Gaussian noise channel, a necessary and sufficient condition for a distribution satisfying the average cost constraints, and the input-bounded constraints to be capacity achieving is derived. This capacity achieving distribution is shown to be "discrete" in nature.

I. Introduction

Shannon demonstrated that in the case of a scalar Gaussian noise channel subject to an average power constraint, the capacity achieving distribution is Gaussian. However, in practice, this source distribution is unrealizable due to its unbounded amplitude. A limit on the peak amplitude is necessary to more accurately reflect the physical limitations present in practical communication systems. In [1], Smith showed that the capacity achieving distribution of a discrete-time memoryless scalar Gaussian channel subject to both an average power constraint and a peak power constraint is a probability mass function. In [2], Shamai *et al.* extended Smith's result to quadrature Gaussian channel under an average and peak total power constraint. The capacity achieving input distribution is proved to be "discrete" in the sense that the phase is uniformly distributed and the distribution of the amplitude is discrete with a finite number of probability mass points.

Using a similar approach as [1] and [2], the capacity achieving input distributions of Rayleigh fading channels and non-coherent quadrature channels have been studied recently. In [3], Abou-Faycal *et al.* proved that the capacity achieving distribution of Rayleigh fading channels subject to an average power constraint is discrete with finite probability mass points, while in [4], Katz *et al.* proved that the capacity achieving distribution of noncoherent quadrature channels subject to an average power constraint is discrete with finite probability mass points in any bounded interval.

In this paper, we further generalize the results of Smith [1] and Shamai [2] by considering discrete-time memoryless vector Gaussian noise channels subject to two classes of constraints: (1) second-order average cost constraints, and (2) input-bounded constraints. Wireless optical communication systems, MIMO systems with deterministic channel gains are examples of practical systems that can be modeled in this way.

[1]This work is supported in part by The Croucher Foundation Fellowship.

II. Problem Formulation

Consider a vector Gaussian noise channel whose inputs are N-tuples, $\mathbf{X} = [X_1, X_2, \cdots, X_{\mathrm{N}}]^\top$, and outputs are M-tuple, $\mathbf{Y} = [Y_1, Y_2, \cdots, Y_{\mathrm{M}}]^\top$. The relation between $\mathbf{X}$ and $\mathbf{Y}$ is characterized by the equation: $\mathbf{Y} = \mathbf{AX} + \mathbf{Z}$ where $\mathbf{A}$ is a fixed $\mathrm{M} \times \mathrm{N}$ channel matrix and $\mathbf{Z} = [Z_1, Z_2, \cdots, Z_{\mathrm{M}}]^\top$ is Gaussian distributed with positive-definite covariance matrix Δ.

Assume that there are K second order average cost constraints of the form $\mathbf{G}_k(\mu) = \mathbf{E}_\mu\left[\mathbf{X}^\top \mathbf{Q}_k \mathbf{X} + \mathbf{L}_k \mathbf{X}\right] - \gamma_k \le 0$, where $\mathbf{Q}_k$ and $\mathbf{L}_k$ are given $\mathrm{N} \times \mathrm{N}$, and $1 \times \mathrm{N}$ matrices respectively, and the expectation is taken with respect to the input distribution μ. To better reflect the limitations in practical communication systems, an input-bounded constraint restricting the set of admissible channel inputs to be a closed, and bounded subset $\mathcal{S}$ in $\mathbb{R}^{\mathrm{N}}$ is imposed. Specifically, if the input distribution is μ, the input-bounded constraint is represented by $\mu(\mathcal{S}) = 1$, or simply $\mathbf{X} \in \mathcal{S}$, meaning that the set of admissible channel inputs is a subset of $\mathcal{S}$.

Let Λ_{phy} be the set of input distributions which satisfy the input-bounded constraint. An input distribution is called admissible if it satisfies both the average cost constraints and the input-bounded constraint. Let Λ be the set of admissible input distributions. The channel capacity of the proposed channel is thus given by $C = \sup_{\mu \in \Lambda} \mathbf{I}(\mu)$, where $\mathbf{I}(\mu)$ is the mutual information between $\mathbf{X}$ and $\mathbf{Y}$, given that the distribution of $\mathbf{X}$ is μ. Let $\mathbf{H}(\mathbf{Z})$ be the entropy of $\mathbf{Z}$, and $\mathbf{H}(\mu)$ be the entropy of $\mathbf{Y}$ given that the distribution of $\mathbf{X}$ is μ. Then the channel capacity problem is formulated as:

$$\begin{cases} \text{Maximize } \mathbf{I}(\mu) = \mathbf{H}(\mu) - \mathbf{H}(\mathbf{Z}) \\ \text{subject to } \mu \in \Lambda_{phy},\ \mathbf{G}_k(\mu) \le 0, \text{ for } k = 1, \cdots, \mathrm{K}. \end{cases}$$

III. Main results

In this section, we derive a necessary and sufficient condition for an input distribution to be capacity achieving, and show that the capacity achieving distribution is "discrete" in nature.

Let $\mathcal{B}_{\mathrm{N}}$ be the space of all bounded, and continuous real-valued functions defined on $\mathbb{R}^{\mathrm{N}}$ subject to the uniform norm, and $\mathcal{B}_{\mathrm{N}}^*$ be the dual space of $\mathcal{B}_{\mathrm{N}}$ subject to the weak* topology. It can be verified that the set of all input distributions is a subset of $\mathcal{B}_{\mathrm{N}}^*$, and the induced relative topology is metrizable. The corresponding metric is called the *Lévy* metric[6].

With respect to the *Lévy* metric, Λ is proved to be convex and (sequential) compact, while $\mathbf{I}(\mu)$ is a concave and continuous function of μ. Consequently, there exists a capacity achieving distribution. Furthermore, if the channel matrix $\mathbf{A}$ is left-invertible, $\mathbf{I}(\mu)$ is strictly concave, and hence the capacity achieving distribution is unique by the strict concavity of $\mathbf{I}(\mu)$.

Necessary and sufficient condition

Suppose μ_o is an input probability distribution. Let E_o be the set of points of increase of μ_o defined as

$$E_o = \left\{ \mathbf{x} \in \mathbb{R}^{\mathrm{N}} : \begin{array}{c} \mu_o(V) > 0 \text{ for any open} \\ \text{set } V \text{ containing } \mathbf{x} \end{array} \right\}.$$

Furthermore, let $\mathrm{P}_{\mathbf{Z}}(\mathbf{z})$ be the density function of $\mathbf{Z}$, $\mathrm{P}_{\mathbf{Y}}(\mathbf{y};\mu)$ be the density function of $\mathbf{Y}$ given that the distribution of $\mathbf{X}$ is μ, and $h(\mathbf{x};\mu_o) = -\int \mathrm{P}_{\mathbf{Z}}(\mathbf{A}\mathbf{x}-\mathbf{y}) \log \mathrm{P}_{\mathbf{Y}}(\mathbf{y};\mu_o) d\mathbf{y}$.

Theorem 1 *Let $\mu_o \in \Lambda$ and E_o be its set of points of increase. Then μ_o is capacity achieving if and only if there exists $\lambda_1 \cdots, \lambda_{\mathrm{K}} \geq 0$ such that for all $\mathbf{x} \in \mathcal{S}$,*

$$h(\mathbf{x};\mu_o) \leq \mathbf{H}(\mu_o) + \sum_{k=1}^{\mathrm{K}} \lambda_k(\mathbf{x}^\top \mathbf{Q}_k \mathbf{x} + \mathbf{L}_k\mathbf{x} - \gamma_k) \qquad (1)$$

and equality holds for all $\mathbf{x} \in E_o$.

Sketch of proof: Using the method of Lagrange multipliers, it can be proved that $\mu_o \in \Lambda$ is capacity achieving if and only if there exists $\lambda_1 \cdots, \lambda_{\mathrm{K}} \geq 0$ such that for $k = 1, \cdots, \mathrm{K}$, $\lambda_k \mathbf{G}_k(\mu_o) = 0$, and for $\mu \in \Lambda_{phy}$,

$$\int h(\mathbf{x};\mu_o)\, d\mu \leq \mathbf{H}(\mu_o) + \sum_{k=1}^{\mathrm{K}} \lambda_k \mathbf{G}_k(\mu). \qquad (2)$$

Now, suppose μ_o is capacity achieving. Then there exists $\lambda_1 \cdots, \lambda_{\mathrm{K}} \geq 0$ such that, $\lambda_k \mathbf{G}_k(\mu_o) = 0$ for $k = 1, \cdots, \mathrm{K}$, and for $\mu \in \Lambda_{phy}$, inequality (2) holds. For any $\mathbf{x} \in \mathcal{S}$, let μ be a distribution such that $\mu(\{\mathbf{x}\}) = 1$ and $\mu(\mathbb{R}^{\mathrm{N}} \backslash \{\mathbf{x}\}) = 0$. Substituting into (2),

$$h(\mathbf{x};\mu_o) \leq \mathbf{H}(\mu_o) + \sum_{k=1}^{\mathrm{K}} \lambda_k(\mathbf{x}^\top \mathbf{Q}_k \mathbf{x} + \mathbf{L}_k\mathbf{x} - \gamma_k) \qquad (3)$$

for $\mathbf{x} \in \mathcal{S}$. On the other hand, as $\int h(\mathbf{x};\mu_o)\, d\mu_o = \mathbf{H}(\mu_o)$, $\lambda_k \int (\mathbf{x}^\top \mathbf{Q}_k \mathbf{x} + \mathbf{L}_k\mathbf{x} - \gamma_k) d\mu_o = \lambda_k \mathbf{G}_k(\mu_o) = 0$, and $h(\mathbf{x};\mu_o)$ is a continuous function of $\mathbf{x}$, $h(\mathbf{x};\mu_o) = \mathbf{H}(\mu_o) + \sum_{k=1}^{\mathrm{K}} \lambda_k(\mathbf{x}^\top \mathbf{Q}_k \mathbf{x} + \mathbf{L}_k\mathbf{x} - \gamma_k)$ for $\mathbf{x} \in E_o$. Hence, the necessary condition is proved.

For the sufficient condition, notice that if inequality (1) is true, then $\int h(\mathbf{x};\mu_o) d\mu_o(\mathbf{x}) \leq \mathbf{H}(\mu_o) + \sum_{k=1}^{\mathrm{K}} \lambda_k \int (\mathbf{x}^\top \mathbf{Q}_k \mathbf{x} + \mathbf{L}_k\mathbf{x} - \gamma_k)\, d\mu_o$, and consequently $\sum_{k=1}^{\mathrm{K}} \lambda_k \mathbf{G}_k(\mu_o) \geq 0$. This implies that $\lambda_k \mathbf{G}_k(\mu_o) = 0$ for all $k = 1, \cdots, \mathrm{K}$. Now, if we integrate inequality (1) with respect to μ, we have

$$\int h(\mathbf{x};\mu_o)\, d\mu \leq \mathbf{H}(\mu_o) + \sum_{k=1}^{\mathrm{K}} \lambda_k \mathbf{G}_k(\mu),$$

and thus the sufficient condition is proved. $\square$

A complex-valued function $f(\mathbf{w})$ defined on $\mathbb{C}^{\mathrm{N}}$ is called holomorphic if it is analytic in each individual variable. A subset F in $\mathbb{R}^{\mathrm{N}}$ is called *sparse* if there exists a nonzero holomorphic function $f(\mathbf{w})$, such that $f(\mathbf{w}) = 0$ for all $\mathbf{w} \in F$. For examples, F is sparse if it is finite, whereas it is not sparse if it contains an open subset in $\mathbb{R}^{\mathrm{N}}$. In particular, when $\mathrm{N} = 1$ and F is bounded, F is sparse if and only if it is finite.

Theorem 2 *The capacity achieving distribution μ_o is "discrete" in the sense that its set of points of increase E_o is sparse.*

Sketch of proof: It can be proved that $h(\mathbf{x};\mu_o)$ and $\mathbf{H}(\mu_o) + \sum_{k=1}^{\mathrm{K}} \lambda_k(\mathbf{x}^\top \mathbf{Q}_k \mathbf{x} + \mathbf{L}_k\mathbf{x} - \gamma_k)$ can be extended holomorphically to $\mathbb{C}^{\mathrm{N}}$. If E_o is not sparse, then $h(\mathbf{x};\mu_o) - \mathbf{H}(\mu_o) - \sum_{k=1}^{\mathrm{K}} \lambda_k(\mathbf{x}^\top \mathbf{Q}_k \mathbf{x} + \mathbf{L}_k\mathbf{x} - \gamma_k) = 0$ for all $\mathbf{x} \in \mathbb{R}^{\mathrm{N}}$. On the other hand, it can be verified easily that $h(\mathbf{x};\mu_o) = -\mathrm{P}_{\mathbf{Z}}(\mathbf{r}) * \log \mathrm{P}_{\mathbf{Y}}(\mathbf{r};\mu_o)$, and $\mathbf{H}(\mu_o) + \sum_{k=1}^{\mathrm{K}} \lambda_k(\mathbf{x}^\top \mathbf{Q}_k \mathbf{x} + \mathbf{L}_k\mathbf{x} - \gamma_k)$ is equal to $\mathrm{P}_{\mathbf{Z}}(\mathbf{r}) * [\mathbf{H}(\mu_o) + \sum_{k=1}^{\mathrm{K}} \lambda_k(\mathbf{r}^\top \mathbf{Q}_k \mathbf{r} + \mathbf{L}_k\mathbf{r} - \gamma_k - \sum_{m,n} \mathbf{Q}_k^{(m,n)} \Delta^{(m,n)})]$ where $*$ is the convolution operator, $\mathbf{r} = \mathbf{A}\mathbf{x}$, $\mathbf{Q}_k^{(m,n)}$ and $\Delta^{(m,n)}$ are the $(m,n)^{th}$ entries of $\mathbf{Q}_k$ and Δ respectively. As the Fourier transform of $\mathrm{P}_{\mathbf{Z}}(\mathbf{z})$ is nonzero everywhere, $-\log \mathrm{P}_{\mathbf{Y}}(\mathbf{y};\mu_o) = \mathbf{H}(\mu_o) + \sum_{k=1}^{\mathrm{K}} \lambda_k \left(\mathbf{y}^\top \mathbf{Q}_k \mathbf{y} + \mathbf{L}_k \mathbf{y} - \gamma_k - \sum_{m,n} \mathbf{Q}_k^{(m,n)} \Delta^{(m,n)}\right)$. Consequently, $\mathbf{Y}$ is Gaussian, which further implies that $\mathbf{X}$ is Gaussian, contradicting to the assumption that $\mathbf{X}$ is bounded. Thus, E_o is sparse. $\square$

Example 1: Indoor wireless optical channel

Consider a wireless optical communication system in which the dominant noise source is the ambient light. Assuming that the symbol waveform is a linear combination of a basis of functions $\{\phi_n(t) : n = 1, \cdots, \mathrm{N}\}$, the channel is equivalent to a discrete time vector Gaussian noise channel $[Y_1, \cdots, Y_{\mathrm{N}}]^\top = [X_1, \cdots, X_{\mathrm{N}}]^\top + [Z_1, Z_2, \cdots, Z_{\mathrm{N}}]^\top$, such that the average power constraint is given by $\mathbf{E}_\mu \left[\sum_{n=1}^{\mathrm{N}} x_n (\int_0^T \phi_n(t)\, dt)\right] - \gamma \leq 0$. Moreover, the set of admissible channel inputs is the following closed and bounded subset [5]

$$\mathcal{S} = \left\{ (x_1, \cdots, x_{\mathrm{N}}) \in \mathbb{R}^{\mathrm{N}} : \forall 0 \leq t < T,\ 0 \leq \sum_{n=1}^{\mathrm{N}} x_n \phi_n(t) \leq s \right\}.$$

Consider the simplest case where there is only one function $\phi_1(t) = 1$ for $t \in [0,1]$ in the basis. In this case, the average power constraint is $\mathbf{E}_\mu[x] - \gamma \leq 0$ and the set of admissible channel inputs is equal to $\mathcal{S} = \{x \in \mathbb{R} : 0 \leq x \leq s\}$.

As the channel matrix $\mathbf{A}$ is the identity matrix, the capacity achieving distribution is unique. By Theorem 1, an input distribution μ_o is capacity achieving if and only if there exists $\lambda \geq 0$ such that

$$\left\{ \begin{array}{ll} h(\mathbf{x};\mu_o) \leq \mathbf{H}(\mu_o) + \lambda[x - \gamma] & \text{for all } \mathbf{x} \in \mathcal{S} \\ h(\mathbf{x};\mu_o) = \mathbf{H}(\mu_o) + \lambda[x - \gamma] & \text{for all } \mathbf{x} \in E_o, \end{array} \right. \qquad (4)$$

where E_o is the set of points of increase of μ_o. Note that, in this special case, it can be argued easily that $x = 0$ is a point of increase of the capacity achieving distribution. Hence, $\lambda = (\mathbf{H}(\mu_o) - h(0;\mu_o))/\gamma$.

Assume without loss of generality that that the noise power and the average optical power γ are equal to 1, and the peak optical intensity s is equal to 4.5. For each $b = 2, 3, \cdots$, let $\tau^{(b)}$ be the discrete admissible input distribution which maximizes the mutual information over those admissible input distributions with the number of probability mass points less than or equal to b.

Since the capacity achieving distribution is discrete, it must be equal to $\tau^{(b)}$ for some b. As $\tau^{(b)}$ can be found easily, the capacity achieving distribution can be found by verifying which $\tau^{(b)}$ satisfies condition (4). In Figure 1, we let $\mu_o = \tau^{(2)}$ and the two curves $y = h(\mathbf{x}; \mu_o)$ and $y = \mathbf{H}(\mu_o) + \lambda[x - \gamma]$ are plotted. From this figure, it is obvious that condition (4) is not satisfied, and thus $\tau^{(2)}$ is not capacity achieving. Repeating the same procedure for $\mu_o = \tau^{(3)}$, Figure 2 shows that condition (4) is satisfied, and hence, $\tau^{(3)}$ is capacity achieving.

Figure 3 plots the functions $\mathbf{I}(\tau^{(b)})$ for different peak optical intensity values s when $b = 2, 3, 4$ respectively. It is interesting to notice that when s is small, $\mathbf{I}(\tau^{(2)}) = \mathbf{I}(\tau^{(3)}) = \mathbf{I}(\tau^{(4)})$. This means that the capacity achieving distribution has only two probability mass points. As s increases, $\mathbf{I}(\tau^{(2)})$ becomes saturated. Hence, to achieve the channel capacity, an input distribution with more probability mass points is needed. Similar phenomenon happens as s continues to increase.

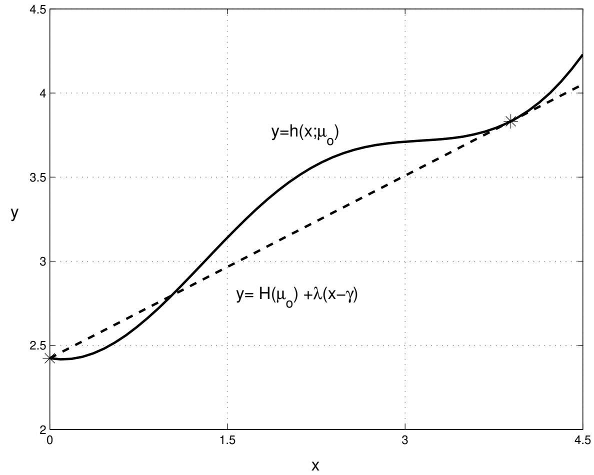

Figure 1: Check of condition (4) when $\mu_o = \tau^{(2)}$.

IV. Symmetry

In this section, we show that by exploiting the symmetry within a communication system, the problem of finding the capacity achieving distribution can be simplified.

Multiple transmitting antennae

Consider an additive vector Gaussian noise channel with N transmitting antennae and one receiving antenna. The input of the channel is a vector $\mathbf{X} = [X_1, \cdots, X_{\mathrm{N}}]^\top$ and the output of the channel is a scalar $\mathbf{Y} = Y$ which is equal to $\mathbf{AX} + Z$ where $\mathbf{A} = [A_1, \cdots, A_{\mathrm{N}}]$ is a row vector, and Z is a zero mean Gaussian noise. The average total power

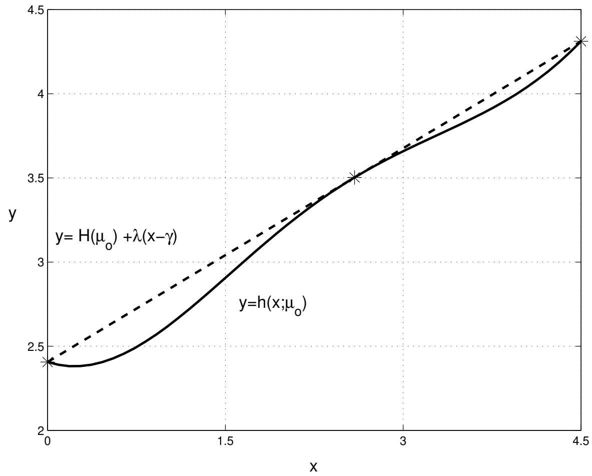

Figure 2: Check of condition (4) when $\mu_o = \tau^{(3)}$.

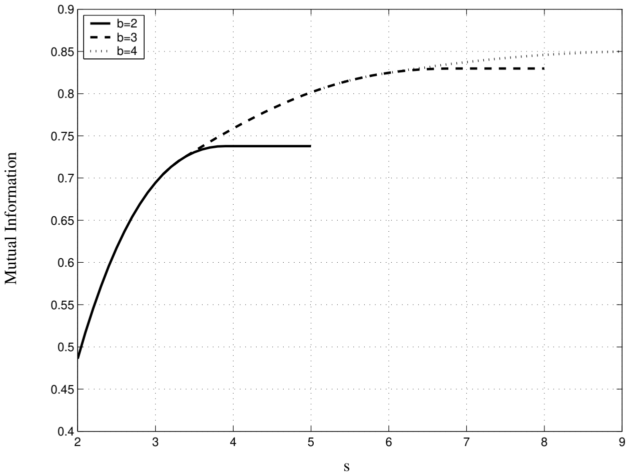

Figure 3: $\mathbf{I}(\tau^{(b)})$ versus peak optical power s.

constraint is given by $\mathbf{E}_\mu[\sum_{n=1}^{\mathrm{N}} X_n^2] \leq \gamma$, and the input-bounded constraint restricts the set of admissible channel inputs to be the set $\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^{\mathrm{N}} : \sum_{n=1}^{\mathrm{N}} x_n^2 \leq s\}$.

For any $\mathbf{x} \in \mathcal{S}$, let

$$\hat{\mathbf{x}} = \left(\frac{\sum_{n=1}^{\mathrm{N}} A_n x_n}{||\mathbf{A}||^2} \right) \mathbf{A}. \tag{5}$$

It is easy to show that $\sum_{n=1}^{\mathrm{N}} A_n \hat{x}_n = \sum_{n=1}^{\mathrm{N}} A_n x_n$, and $\sum_{n=1}^{\mathrm{N}} x_n^2 \geq \sum_{n=1}^{\mathrm{N}} \hat{x}_n^2$, with equality if and only if $\mathbf{x}$ is a scalar multiple of $\mathbf{A}^\top$. Based on the two observations, the capacity achieving distribution should "concentrate" all the probabilities on the "line", whose direction is characterized by the vector $\mathbf{A}^\top$. Therefore, the channel is reduced to a scalar channel $Y = \mathbf{AA}^\top X + Z$ subject to the average power constraint $\mathbf{E}_\mu[X^2||\mathbf{A}||^2] \leq \gamma$ and the input-bounded constraint corresponding to the set of admissible channel inputs $\mathcal{S} = \{x \in \mathbb{R} : x^2||\mathbf{A}||^2 \leq s\}$. And by Theorem 2 the capacity achieving distribution for this case is a probability mass function.

Orthogonal Multi-Pulse Modulation

Consider the following channel where orthogonal multi-pulse modulation is used. The output of the channel $\mathbf{Y} = [Y_1, \cdots, Y_{\mathrm{N}}]^\top$ is equal to the sum of the channel input $\mathbf{X} = [X_1, \cdots, X_{\mathrm{N}}]^\top$ and a zero-mean Gaussian noise $\mathbf{Z} = [Z_1, \cdots, Z_{\mathrm{N}}]^\top$. Assume that all the components of the noise are identical, and independently distributed with variance σ^2. The average total power constraint is thus given by $\mathbf{E}_\mu[\sum_{n=1}^{\mathrm{N}} X_n^2] \leq \gamma$. We are going to consider two different input-bounded constraints: (1) the rectangular constraint:

$$\mathcal{S} = \left\{\mathbf{x} \in \mathbb{R}^{\mathrm{N}} : |x_n| \leq s\right\} \tag{6}$$

and (2) the spherical constraint:

$$\mathcal{S} = \left\{\mathbf{x} \in \mathbb{R}^{\mathrm{N}} : \sum_{n=1}^{\mathrm{N}} x_n^2 \leq s\right\}. \tag{7}$$

Rectangular admissible region $\mathcal{S}$

Since the additive noise components of $[Z_1, \cdots, Z_{\mathrm{N}}]^\top$ are independent, for any input distribution μ of $[X_1, \cdots, X_{\mathrm{N}}]^\top$, it can be verified easily that

$$\mathbf{I}_\mu\left(\mathbf{X}; \mathbf{Y}\right) \quad \leq \quad \sum_{n=1}^{\mathrm{N}} \mathbf{I}_\mu\left(X_n; Y_n\right), \tag{8}$$

where the subscrpit μ is the input distribution.

Suppose μ_o is the unique capacity achieving input distribution, and $\mu_o^1, \cdots, \mu_o^{\mathrm{N}}$ are the corresponding marginal distributions of $X_1, \cdots, X_{\mathrm{N}}$ respectively. First, notice that by the uniqueness of the capacity achieving distribution, $\mathbf{E}_{\mu_o^n}[X_n^2] = \frac{\gamma}{\mathrm{N}}$. Second, suppose ν is the product probability measure of $\mu^1, \cdots, \mu^{\mathrm{N}}$. It can be verified easily that the input distribution ν satisfies the average cost constraint and the input-bounded constraint. Moreover, $\mathbf{I}_\nu\left(\mathbf{X}; \mathbf{Y}\right) = \sum_{n=1}^{\mathrm{N}} \mathbf{I}_\mu\left(X_n; Y_n\right) \geq \mathbf{I}_\mu\left(\mathbf{X}; \mathbf{Y}\right)$. By uniqueness of μ_o, $\mu_o = \nu$.

Consequently, the original channel can be decoupled into N independent and identical channels. Hence, to find the capacity achieving distribution of the vector channel is equivalent to finding the capacity achieving distribution of each individual component channel $Y_n = X_n + Z_n$ subject to average power constraint $\mathbf{E}_{\mu_o^n}[X_n^2] \leq \frac{\gamma}{\mathrm{N}}$ and input bounded constraint $|x_n| \leq s$.

As the capacity achieving distribution μ_o^n for each channel is a probability mass function, μ_o, which is equal to the product probability measure of $\mu_o^1, \cdots, \mu_o^{\mathrm{N}}$ is also a probability mass function, and the set of points of increase of μ_o is just the Cartesian product of the sets of points of increases of $\mu_o^1, \cdots, \mu_o^{\mathrm{N}}$.

Spherical admissible region $\mathcal{S}$

Now consider the case that the physical constraint is spherical. Suppose μ_o is capacity achieving. As the channel matrix is the identity matrix, the capacity achieving distribution μ_o is unique.

For any $\mathrm{N} \times \mathrm{N}$ orthonormal matrix $\mathbf{B}$, let $\tilde{\mathbf{X}} = \mathbf{B}\mathbf{X}$ and ρ be its underlying probability distribution. Then

$$\begin{aligned}
\mathbf{I}\left(\mathbf{X}; \mathbf{X}+\mathbf{Z}\right) &= \mathbf{H}\left(\mathbf{X}+\mathbf{Z}\right) - \mathbf{H}\left(\mathbf{X}+\mathbf{Z}|\mathbf{X}\right) && (9)\\
&= \mathbf{H}\left(\mathbf{B}\mathbf{X}+\mathbf{B}\mathbf{Z}\right) - \mathbf{H}\left(\mathbf{Z}\right) && (10)\\
&= \mathbf{H}\left(\tilde{\mathbf{X}}+\mathbf{Z}\right) - \mathbf{H}\left(\mathbf{Z}\right) && (11)\\
&= \mathbf{I}\left(\tilde{\mathbf{X}}; \tilde{\mathbf{X}}+\mathbf{Z}\right) && (12)
\end{aligned}$$

where the second equality follows from $\mathbf{B}$ is an orthonormal matrix, and the third equality follows from $\mathbf{BZ}$ and $\mathbf{Z}$ are identically distributed due to the components of $\mathbf{Z}$ is *i.i.d.* Gaussian distributed. It can be verified that μ_o satisfies the average cost constraint and the input-bounded constraint if and only if ρ also satisfies the two constraints. Since the capacity achieving distribution is unique, $\rho_o = \mu_o$.

For any $\mathbf{x} \in \mathbb{R}^{\mathrm{N}}$, the phase of $\mathbf{x}$ is defined as the unit vector $\frac{\mathbf{x}}{||\mathbf{x}||}$. Hence, every $\mathbf{x}$ is uniquely characterized by its phase and magnitude. The above result shows that the distribution of the magnitude is independent of the distribution of the phase, and the phase is uniformly distributed.

As a corollary, if $\mathbf{x}$ is a point of increase of μ_o, then $\mathbf{Bx}$ is also a point of increase of μ_o. Hence, the set of points of increases of μ_o is of the form $\{\mathbf{x} \in \mathbb{R}^{\mathrm{N}} : ||\mathbf{x}|| \in \mathcal{E}\}$. where $\mathcal{E}$ is the set of points of increase of the distribution of the magnitude. It can be checked easily that $\{\mathbf{x} \in \mathbb{R}^{\mathrm{N}} : ||\mathbf{x}|| \in \mathcal{E}\}$ is sparse if and only if $\mathcal{E}$ is finite. Hence, the distribution of the magnitude is discrete, and the capacity-achieving probability distribution concentrates its mass in concentric spheres centered at origin.

V. Conclusion

In this work, the capacity achieving distribution of a discrete-time memoryless vector Gaussian noise channel subject to second-order average cost constraints and input-bounded constraint is considered. A necessary and sufficient condition for an input distribution to be capacity achieving is derived, and the distribution is shown to be discrete in nature.

References

[1] J.G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inform. Contr.*, vol 18., pp. 203-219, 1971.

[2] S. Shamai(Shitz) and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inform. Theory*, vol 41, pp. 1060 -1071 July 1995.

[3] I.C. Abou-Faycal, M.D. Trott, S. Shamai(Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1290 - 1301, May 2001.

[4] M. Katz, S. Shamai(Shitz), "On the capacity-achieving distribution of the discrete-time non-coherent additive white noise channel," *IEEE Int. Symp. Inform. Theory* 2002, pp. 165.

[5] S. Hranilovic, F.R. Kschischang, "Optical Intensity-Modulated Direct Detection Channels: Signal Space and Lattice Codes," to appear in *IEEE Trans. Inform. Theory.*

[6] D.W. Stroock, *Probability theory, an analytic view.* Cambridge University Press 1993.