An Upper Bound on the Number of Mass Points in the Capacity Achieving Distribution for the Amplitude Constrained Additive Gaussian Channel

Semih Yagli[†], Alex Dytso[†], H. Vincent Poor[†], Shlomo Shamai (Shitz)[*],
[†] Princeton University, Princeton, NJ 08544, USA, Email: {syagli, adytso, poor}@princeton.edu
[*] Israel Institute of Technology, Technion City, Haifa 32000, Israel, Email: sshlomo@ee.technion.ac.il

Abstract—This paper studies an n-dimensional additive Gaussian noise channel with a peak-power-constrained input. It is well known that, in this case, the capacity-achieving input distribution is supported on finitely many concentric shells. However, due to the previous proof technique, neither the exact number of shells of the optimal input distribution nor a bound on it was available.

This paper provides an alternative proof of the finiteness of the number shells of the capacity-achieving input distribution and produces the first firm upper bound on the number of shells, paving an alternative way for approaching many such problems. In particular, for every dimension n, it is shown that the number of shells is given by $O(\mathsf{A}^2)$ where A is the constraint on the input amplitude. Moreover, this paper also provides bounds on the number of points for the case of $n = 1$ with an additional power constraint.

I. INTRODUCTION

We consider an additive noise channel where the input-output relationship is given by

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}, \tag{1}$$

where the input $\boldsymbol{X} \in \mathbb{R}^n$ is independent of the standard Gaussian noise $\boldsymbol{Z} \in \mathbb{R}^n$. We are interested in finding the capacity of the channel in (1) subject to the constraint that $\boldsymbol{X} \in \mathcal{B}_0(\mathsf{A})$ where $\mathcal{B}_0(\mathsf{A})$ is an n-ball centered at zero with radius A (i.e., peak-power constrained input), that is

$$C_n(\mathsf{A}) = \max_{\boldsymbol{X} \,:\, \boldsymbol{X} \in \mathcal{B}_0(\mathsf{A})} I(\boldsymbol{X}; \boldsymbol{Y}). \tag{2}$$

In his seminal paper [1], for the case of $n = 1$, Smith has shown that an optimizing distribution in (2) is unique, symmetric, and, perhaps surprisingly, discrete with finitely many mass points. Using tools such as the Identity Theorem from complex analysis, Smith has proven that the cardinality of the support set of the optimal input distribution cannot be infinite, and, thus, must be finite. Employing this proof by contradiction, Shamai and Bar-David [2] have extended the method of Smith to $n = 2$, and showed that, in this setting, the maximizing input random variable is given by $\boldsymbol{X}^\star = R^\star \cdot \boldsymbol{U}^\star$ where the magnitude $R^\star$ is discrete with finitely many points and the random unit vector $\boldsymbol{U}^\star$, which is independent of $R^\star$, has a uniform phase on $[0, 2\pi)$. In other words, the support of $\boldsymbol{X}^\star$ consists of finitely many concentric shells. As a matter of fact, this phenomena that

the optimal input distribution lies on finitely many concentric spheres remains true for any $n \geq 2$, cf. [3], [4] and [5].

Regrettably, the method of proof by contradiction does not lead to a characterization of the number of spheres (number of mass points when $n = 1$) in the capacity-achieving input distribution. In fact, as of the writing of this paper, very little is known about the structure of that distribution, and a very simple question remains open about 50 years after Smith's contribution: *When $n = 1$, what is the cardinality of the support of the optimal input distribution as a function of A?*

In this work, we provide the first firm upper bound on the number of points for $n = 1$ and the number of shells for every $n \geq 2$, partially answering the above question. Furthermore, for the case of $n = 1$, using similar methods, we also provide an upper bound on the cardinality of the support of the distribution achieving

$$C(\mathsf{A}, \mathsf{P}) = \max_{\substack{X \,:\, |X| \leq \mathsf{A} \\ \mathbb{E}[X^2] \leq \mathsf{P}}} I(X; Y). \tag{3}$$

a) Prior Work: The history of the problem begins with [1], where Smith proves the discreteness of the capacity-achieving input distribution and also shows the optimality of the equiprobable binary input on $\{\pm\mathsf{A}\}$ so long as $\mathsf{A} \leq 0.1$. Sharma and Shamai [6] extend the result of Smith, and argue that an equiprobable input on $\{\pm\mathsf{A}\}$ is optimal if and only if $\mathsf{A} \leq \bar{\mathsf{A}} \approx 1.665$. The proof of the result in [6], which generalizes to vector channels, is shown in [7].

A progress on the algorithmic aspect of computing the optimal input distribution is made in [8] which proposes an iterative procedure that converges to the a capacity achieving distribution based on the cutting-plane method. The bound on the number of mass points found in this work is relevant for numerical methods as it reduces the optimization space.

A number of papers have also focused on upper and lower bounds on the capacity in (2). Broadly speaking, there are three types of capacity upper bounding approaches. The first approach uses the maximum entropy principle [9, Chapter 12] and upper bounds the output differential entropy, $h(Y)$, subject to some moment constraint [10]. The second approach uses a dual capacity characterization where the maximization of the mutual information over the input distribution is replaced by minimization of the relative entropy over the output distribution. A suboptimal choice of an output distribution in the dual capacity expression results in an upper bound on the capacity [11], [12], [13]. The third approach uses a characterization of the mutual information as an integral of the minimum mean square error (MMSE) [14], and leads to

ISIT 2019

an upper bound by replacing the optimal estimator in the MMSE term by a suboptimal one [7].

There is also a substantial literature that extends the proof recipe of Smith to the other channels. For example, the approach of Smith for showing discreteness of an optimal input distribution has been extended to complex Gaussian channels [2], Rayleigh fading channels [15], and Poisson channels [16]. For an overview of the literature on various optimization methods that show discreteness of a capacity-achieving distribution the interested reader is referred to [17].

 b) Contributions and Paper Outline: In what follows:
1) Section II presents our main results;
2) Section III provides an upper bound on the number of extreme points, for the scalar case, of an arbitrary output probability density function (pdf) of the Gaussian channel described in (1). The proof of this result exploits the analyticity of the Gaussian density together with Tijdeman's Number of Zeros Lemma [18, Lemma 1];
3) Section IV provides the proof of our main result for the case of $n = 1$. The idea behind this proof is to show that the maximum number of extreme points of the output pdf provides an upper bound on the number of mass points of the optimal input distribution. The main element of the proof relies on Karlin's Oscillation Theorem and the bounds on the number of extreme points developed in Section III. The proof for the vector case ($n \geq 2$) follows along the same lines as the proof for the scalar case ($n = 1$), albeit with a more involved algebra, therefore it is omitted to abide the space constraints; and
4) Section V concludes the paper with some final remarks.

 c) Notation: Throughout the paper, the deterministic scalar quantities are denoted by lower-case letters, deterministic vectors are denoted by bold lowercase letters, random variables are denoted by uppercase letters, and random vectors are denoted by bold uppercase letters (e.g., x, $\boldsymbol{x}$, X, $\boldsymbol{X}$). We denote the distribution of a random vector $\boldsymbol{X}$ by $P_{\boldsymbol{X}}$. Moreover, we say that a point $\boldsymbol{x}$ is in the support, denoted by supp($P_{\boldsymbol{X}}$), of the distribution $P_{\boldsymbol{X}}$ if for every open set $\mathcal{O} \ni \boldsymbol{x}$ we have that $P_{\boldsymbol{X}}(\mathcal{O}) > 0$. The number of zeros of a function $f \colon \mathbb{R} \to \mathbb{R}$ on the interval $\mathcal{I}$ is denoted by $\mathrm{N}(\mathcal{I}, f)$. Similarly, if $f \colon \mathbb{C} \to \mathbb{C}$ is a function on the complex domain, $\mathrm{N}(\mathcal{D}, f)$ denotes the number of its zeros in the region $\mathcal{D}$.

II. Main Result

Theorem 1, stated below, gives the first firm upper bound on the support size of the capacity-achieving input of the additive Gaussian channel with an amplitude constraint.

Theorem 1. *Consider the amplitude constrained scalar additive Gaussian channel $Y = X + Z$ where the input X, satifying $|X| \leq \mathsf{A}$, is assumed to be independent from the noise $Z \sim \mathcal{N}(0, 1)$. Assuming $\mathsf{A} \geq 1$, let $P_{X^\star}$ be the optimizing input distribution for this channel. Then, $P_{X^\star}$ is a symmetric discrete distribution with*

$$\sqrt{1 + \frac{2\mathsf{A}^2}{\pi \mathrm{e}}} \leq |\mathsf{supp}(P_{X^\star})| \leq a_2 \mathsf{A}^2 + a_1 \mathsf{A} + a_0, \quad (4)$$

where

$$a_2 = 9\mathrm{e} + 6\sqrt{\mathrm{e}} + 5,$$
$$a_1 = 6\mathrm{e} + 2\sqrt{\mathrm{e}},$$
$$a_0 = \mathrm{e} + 2\log\left(4\sqrt{\mathrm{e}} + 2\right) + 1.$$

Shown in Section IV, the proof of Theorem 1 uses an upper bound on the number of extreme points of Gaussian convolution that can be found in Section III.

Remark 1. Observe that the upper and lower bounds in (4) are not of equal order. We conjecture that the order of the lower bound is tight. See the extended version of this paper in [19] for a more detailed discussion.

Theorem 2. *Consider the amplitude constrained vector additive Gaussian channel $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}$ where the input $\boldsymbol{X}$, satisfying $\|\boldsymbol{X}\| \leq \mathsf{A}$, is assumed to be independent from the white Gaussian noise $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. Let $\boldsymbol{X}^\star \sim P_{\boldsymbol{X}^\star}$ be the optimizing input for this channel. Then, $P_{\boldsymbol{X}^\star}$ is unique, radially symmetric, and the distribution of its amplitude, namely $P_{\|\boldsymbol{X}^\star\|}$, is a discrete distribution with*

$$\left|\mathsf{supp}\left(P_{\|\boldsymbol{X}^\star\|}\right)\right| \leq \mathfrak{a}_{n_2} \mathsf{A}^2 + \mathfrak{a}_{n_1} \mathsf{A} + \mathfrak{a}_{n_0}, \quad (5)$$

where

$$\mathfrak{a}_{n_2} = 4 + 4\mathrm{e} + \sqrt{8\mathrm{e} + 4},$$
$$\mathfrak{a}_{n_1} = \left(3 + 4\mathrm{e} + \sqrt{2\mathrm{e} + 1}\right) n + \sqrt{\frac{32}{n - 1}},$$
$$\mathfrak{a}_{n_0} = \log \frac{\mathrm{e}^2 \sqrt{\pi}\, \Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$
$$+ \left(3 + 4\mathrm{e} + \sqrt{2\mathrm{e} + 1}\right) \log \frac{\mathrm{e}^{\frac{n}{2}} \sqrt{\pi}\, \Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

Omitted because of the space constraint, the proof of Theorem 2 benefits from the same technique that is used in the proof of Theorem 1. For details, see [19].

Remark 2. Note that when the vector channel is of dimension 2, Theorem 2 gives an upper bound on the number of shells of the optimal input distribution for the additive complex Gaussian channel with an amplitude constraint.

For the sake of demonstrating the versatility of our novel method, shown next[1] is an upper bound on the support size of the optimal input distribution for the scalar additive Gaussian channel with both a peak- and an average-power constraints.

Theorem 3. *Consider the amplitude and power constrained scalar additive Gaussian channel $Y = X + Z$ where the input X, satisfying $|X| \leq \mathsf{A}$ and $\mathbb{E}[|X|^2] \leq \mathsf{P}$, is assumed to be independent from the noise $Z \sim \mathcal{N}(0, 1)$. Assuming $\mathsf{A} \geq 1$, let $P_{X^\star}$ be the optimizing input distribution for this channel. Then, $P_{X^\star}$ is a symmetric discrete distribution with*

$$\sqrt{1 + \frac{2\min\{\mathsf{A}^2, 3\mathsf{P}\}}{\pi \mathrm{e}}} \leq |\mathsf{supp}(P_{X^\star})| \quad (6)$$
$$\leq a_{\mathsf{P}_2} \mathsf{A}_{\mathsf{P}}^2 + a_{\mathsf{P}_1} \mathsf{A}_{\mathsf{P}} + a_{\mathsf{P}_0}, \quad (7)$$

where

$$\mathsf{A}_{\mathsf{P}} = \frac{\mathsf{A}\mathsf{P}}{\mathsf{P} - \log(1 + \mathsf{P})\mathbf{1}\{\mathsf{P} < \mathsf{A}^2\}},$$
$$a_{\mathsf{P}_2} = (1 + 2\lambda_{\mathsf{P}})(9\mathrm{e} + 6\sqrt{\mathrm{e}} + 1) + 2(2 - \lambda_{\mathsf{P}})(1 - 2\lambda_{\mathsf{P}}),$$
$$a_{\mathsf{P}_1} = (1 + 2\lambda_{\mathsf{P}})(6\mathrm{e} + 2\sqrt{\mathrm{e}}),$$
$$a_{\mathsf{P}_0} = (1 + 2\lambda_{\mathsf{P}})\mathrm{e} + 2\log\left(\frac{2 + 4\sqrt{\mathrm{e}}(1 + 2\lambda_{\mathsf{P}})}{1 - 2\lambda_{\mathsf{P}}}\right) + 1,$$
$$\lambda_{\mathsf{P}} = \frac{\log(1 + \mathsf{P})}{2\mathsf{P}} \mathbf{1}\{\mathsf{P} < \mathsf{A}^2\}.$$

---

[1]Proof is omitted because of space constraints. See [19] for the details.

Remark 3. In the case of $P \geq A^2$, the power constraint is inactive and Theorem 3 recovers the result of Theorem 1.

III. BOUNDS ON THE NUMBER OF EXTREME POINTS OF A GAUSSIAN CONVOLUTION

This section presents some of the main tools required in our analysis. Specifically, given an *unknown constant* $0 \leq \kappa_1 \leq \max_b f_Y(b)$, our aim is to find a worst case upper bound on the number of zeros of the shifted output pdf

$$f_Y - \kappa_1.$$

Here f_Y denotes the pdf of the random variable $Y = X + Z$, where X is an *arbitrary* zero mean[2] random variable at the input of the channel satisfying the amplitude constraint: $|X| \leq A$; Z is the standard Gaussian random variable independent from X; and Y is the random variable induced by the input X at the output of this additive Gaussian channel.

As a starting point, before chasing after the number of zeros of $f_Y - \kappa_1$, the following lemma shows that the zeros of $f_Y - \kappa_1$ are always contained on an interval that is only "slightly" larger than $[-A, A]$.

Remark 4. Since the capacity-achieving input distribution for the channel in question is unknown, throughout this paper, the bounds are uniform over all possible inputs X satisfying $|X| \leq A$. Equivalently, the bounds involving the output pdf f_Y are uniform over all possible output distributions.

Lemma 1 (On the Location and Finiteness of Zeros). *For a fixed $\kappa_1 \in \left(0, \frac{1}{\sqrt{2\pi}}\right]$ there exists some $B_{\kappa_1} = B_{\kappa_1}(A) < \infty$ such that*

$$N\left(\mathbb{R}, f_Y - \kappa_1\right) = N\left([-B_{\kappa_1}, B_{\kappa_1}], f_Y - \kappa_1\right) < \infty. \quad (8)$$

In other words, there are finitely many zeros of $f_Y(y) - \kappa_1$ all of which are contained within the interval $[-B_{\kappa_1}, B_{\kappa_1}]$. Moreover, B_{κ_1} can be upper bounded as follows:

$$B_{\kappa_1} \leq A + \log^{\frac{1}{2}}\left(\frac{1}{2\pi\kappa_1^2}\right). \quad (9)$$

Since the exact value of the constant κ_1 is unknown, in counting the number of zeros of $f_Y - \kappa_1$, a worst-case approach needs to be taken. In an attempt at doing so, the following elementary result from calculus provides a bound on the number of zeros of a function in terms of the number of its extreme points. As simple as it is, Lemma 2 is one of the key steps in this paper. It states that, to find a bound on the number of zeros of $f_Y - \kappa_1$, it suffices to find a bound on that of f_Y', eliminating the dependence on the nuisance constant κ_1.

Lemma 2. *Suppose that f is continuous on $[-R, R]$ and differentiable on $(-R, R)$. If $N([-R, R], f) < \infty$, then*

$$N([-R, R], f) \leq N([-R, R], f') + 1, \quad (10)$$

where f' denotes the derivative of f.

Thanks to Lemma 2, to upper bound the number of zeros of $f_Y - \kappa_1$, all that is needed is to find an upper bound on the number of zeros of the derivative of f_Y, namely

$$f_Y'(y) = \frac{1}{\sqrt{2\pi}}\mathbb{E}\left[(X - y)\exp\left(-\frac{(y - X)^2}{2}\right)\right]. \quad (11)$$

At this point, there are several trajectories that one could follow to produce an upper bound on the number of zeros, see the full version of this paper [19] for a detailed discussion. The method used in this paper is based on Tijdeman's Number of Zeros Lemma, which is presented next.

Lemma 3 (Tijdeman's Number of Zeros Lemma). *Let R, s, t be positive numbers such that $s > 1$. For the complex valued function $f \neq 0$ which is analytic on $|z| < (st + s + t)R$, its number of zeros $N(\mathcal{D}_R, f)$ within the disk $\mathcal{D}_R = \{z : |z| \leq R\}$ satisfies*

$$N(\mathcal{D}_R, f) \leq \log_s \frac{\max_{|z| \leq (st+s+t)R} |f(z)|}{\max_{|z| \leq tR} |f(z)|}. \quad (12)$$

The following two lemmas, whose proofs (see [19]) are omitted because of space constraints, find upper and lower bounds on absolute value of the complex analytic extension[3] of f_Y' over a disc of finite radius centered at the origin.

Lemma 4. *Let $f_Y' : \mathbb{R} \to \mathbb{R}$ as in (11) and let $\breve{f}_Y' : \mathbb{C} \to \mathbb{C}$ denote its complex extension. Then,*

$$\max_{|z| \leq B} \left|\breve{f}_Y'(z)\right| \leq \frac{1}{\sqrt{2\pi}}(A + B)\exp\left(\frac{B^2}{2}\right). \quad (13)$$

Lemma 5. *Let $f_Y' : \mathbb{R} \to \mathbb{R}$ as in (11) and let $\breve{f}_Y' : \mathbb{C} \to \mathbb{C}$ denote its complex extension. For any $|X| \leq A \leq B$,*

$$\max_{|z| \leq B} \left|\breve{f}_Y'(z)\right| \geq \frac{A}{\sqrt{2\pi}}\exp\left(-2A^2\right). \quad (14)$$

By assembling the results of Lemmas 3, 4 and 5, Theorem 4 below provides an upper bound on the number of oscillations of a Gaussian convolution.

Theorem 4 (Bound on the Number of Oscillations of f_Y). *Suppose X is a symmetric, real-valued random variable and Z is standard Gaussian. Assuming $|X| \leq A < R$, let $Y = X + Z$ and denote by f_Y the probability density function of Y. Then, the number of extreme points of f_Y, namely the number of zeros of f_Y', within the interval $[-R, R]$ satisfies*

$$N([-R, R], f_Y')$$
$$\leq \min_{s > 1}\left\{\frac{\frac{((A+R)s+A)^2}{2} + 2A^2 + \log\left(2 + \frac{(A+R)s}{A}\right)}{\log s}\right\}. \quad (15)$$

Proof. Let $\mathcal{D}_R \subset \mathbb{C}$ be a disk of radius R centered at $z_0 = 0$, and note that

$$N([-R, R], f_Y')$$
$$\leq N(\mathcal{D}_R, \breve{f}_Y') \quad (16)$$
$$\leq \min_{s > 1, t \geq \frac{A}{R}}\left\{\log_s \frac{\max_{|z| \leq (st+s+t)R} |\breve{f}_Y'(z)|}{\max_{|z| \leq tR} |\breve{f}_Y'(z)|}\right\} \quad (17)$$
$$\leq \min_{s > 1, t \geq \frac{A}{R}}\left\{\frac{\frac{(st+s+t)^2 R^2 + 4A^2}{2} + \log\left(1 + \frac{(st+s+t)R}{A}\right)}{\log s}\right\} \quad (18)$$
$$= \min_{s > 1}\left\{\frac{\frac{((A+R)s+A)^2 + 4A^2}{2} + \log\left(2 + \frac{(A+R)s}{A}\right)}{\log s}\right\}, \quad (19)$$

---

where (16) follows because zeros of f'_Y are also zeros of $\breve{f}'_Y$; (17) is a consequence of Lemma 3; (18) follows from Lemma 4 with $\mathsf{B} \leftarrow (st+s+t)R$ and Lemma 5 with $\mathsf{B} \leftarrow tR$; and in (19) we use the fact that $t = \frac{\mathsf{A}}{R}$ is the minimizer. $\square$

Finally, combining the results of Lemmas 1 and 2, and Theorem 4, the following corollary presents the desired result of this section.

Corollary 1. *Given an arbitrary constant $\kappa_1 \in \left(0, \frac{1}{2\pi}\right)$, suppose $R > \mathsf{A} + \log^{\frac{1}{2}}\left(\frac{1}{2\pi\kappa_1^2}\right)$. Then, the number of zeros of $f_Y - \kappa_1$ satisfies*

$$
\begin{aligned}
&\mathrm{N}(\mathbb{R}, f_Y - \kappa_1) \\
&= \mathbb{N}([-R, R], f_Y - \kappa_1) \quad (20) \\
&\leq 1 + \min_{s>1}\left\{\frac{\frac{((\mathsf{A}+R)s+\mathsf{A})^2}{2} + 2\mathsf{A}^2 + \log\left(2 + \frac{(\mathsf{A}+R)s}{\mathsf{A}}\right)}{\log s}\right\}. \quad (21)
\end{aligned}
$$

IV. PROOF OF THE MAIN RESULT

This section proves the result presented in Theorem 1. The first ingredient of the proof is the following characterization of the optimal input distribution shown in [1, Corollary 1].

Lemma 6. *Consider the amplitude constrained scalar additive Gaussian channel $Y = X + Z$ where the input X, satisfying $|X| \leq \mathsf{A}$, is independent from the noise $Z \sim \mathcal{N}(0,1)$. Then, $P_{X^\star}$ is the capacity-achieving input distribution if and only if*

$$
\begin{aligned}
i(x; P_{X^\star}) &= C(\mathsf{A}), \quad x \in \mathsf{supp}(P_{X^\star}), \quad (22) \\
i(x; P_{X^\star}) &\leq C(\mathsf{A}), \quad x \in [-\mathsf{A}, \mathsf{A}], \quad (23)
\end{aligned}
$$

where

$$
i(x; P_{X^\star}) = \int_{\mathbb{R}} \frac{\mathrm{e}^{-\frac{(y-x)^2}{2}}}{\sqrt{2\pi}} \log \frac{1}{f_{Y^\star}(y)}\mathrm{d}y - h(Z), \quad (24)
$$

with $h(Z) = \log\sqrt{2\pi\mathrm{e}}$ denoting the differential entropy of the standard Gaussian distribution, and $f_{Y^\star}(y)$ denoting the output pdf induced by the input $P_{X^\star}$, that is, for $X \sim P_{X^\star}$,

$$
f_{Y^\star}(y) = \frac{1}{\sqrt{2\pi}}\mathbb{E}\left[\mathrm{e}^{-\frac{(y-X)^2}{2}}\right]. \quad (25)
$$

Remark 5. An immediate consequence of Lemma 6 is the fact that $x \in \mathsf{supp}(P_{X^\star}) \implies i(x; P_{X^\star}) - C(\mathsf{A}) = 0$. In other words,

$$
\begin{aligned}
|\mathsf{supp}(P_{X^\star})| &\leq \mathrm{N}([-\mathsf{A}, \mathsf{A}], \Xi_\mathsf{A}(\cdot; P_{X^\star})) \quad (26) \\
&\leq \mathrm{N}(\mathbb{R}, \Xi_\mathsf{A}(\cdot; P_{X^\star})), \quad (27)
\end{aligned}
$$

where the function $\Xi(\cdot; P_{X^\star})\colon \mathbb{R} \to \mathbb{R}$ is defined as

$$
\Xi_\mathsf{A}(x; P_{X^\star}) = i(x; P_{X^\star}) - C(\mathsf{A}). \quad (28)
$$

a) Connecting the Number of Oscillation of f_Y to the Number of Masses in P_X : This section gives an alternative proof that $P_{X^\star}$ is discrete by relating the cardinality of $\mathsf{supp}(P_{X^\star})$ to the number of zeros of the shifted output pdf $f_{Y^\star} - \mathrm{e}^{-C(\mathsf{A})-h(Z)}$. The following definition sets the stage.

Definition 1 (Sign Change of a Function). *The number of sign changes of a function ξ is given by*

$$
\mathscr{S}(\xi) = \sup_{m\in\mathbb{N}}\left\{\sup_{y_1<\cdots<y_m} \mathscr{N}\{\xi(y_i)\}_{i=1}^m\right\}, \quad (29)
$$

where $\mathscr{N}\{\xi(y_i)\}_{i=1}^m$ is the number of changes of sign of the sequence $\{\xi(y_i)\}_{i=1}^m$.

Proven in [21], the following theorem is the main tool in connecting the number of zeros of an output pdf $f_{Y^\star}$ to the number of mass points of a capacity-achieving input distribution $P_{X^\star}$.

Theorem 5 (Oscillation Theorem). *Given open intervals $\mathbb{I}_1$ and $\mathbb{I}_2$, let $p\colon \mathbb{I}_1 \times \mathbb{I}_2 \to \mathbb{R}$ be a strictly Polyá type-∞ function.[4] For an arbitrary y, suppose $p(\cdot, y)\colon \mathbb{I}_1 \to \mathbb{R}$ is an n-times differentiable function. Assume that μ is a measure on $\mathbb{I}_2$, and let $\xi\colon \mathbb{I}_2 \to \mathbb{R}$ be a function with $\mathscr{S}(\xi) = n$. For $x \in \mathbb{I}_1$, define*

$$
\Xi(x) = \int \xi(y)p(x,y)\mathrm{d}\mu(y). \quad (30)
$$

If $\Xi\colon \mathbb{I}_1 \to \mathbb{R}$ is an n-times differentiable function, then either $\mathscr{S}(\Xi) \leq \mathrm{N}(\mathbb{I}_1, \Xi) \leq n$, or $\Xi \equiv 0$.

Note that Theorem 5 is applicable in our setting as the Gaussian distribution is a member of the set of Polyá type-∞ functions [21]. The following result shows the connection between the support size of $P_{X^\star}$ and the number of zeros of the optimal output pdf $f_{Y^\star}$.

Lemma 7. *The support set of the capacity-achieving input distribution $P_{X^\star}$ satisfies*

$$
|\mathsf{supp}(P_{X^\star})| \leq \mathrm{N}\left([-R, R], f_{Y^\star} - \kappa_1\right) < \infty. \quad (31)
$$

where $\kappa_1 = \mathrm{e}^{-C(\mathsf{A})-h(Z)}$ and $R > \mathsf{A} + \log^{\frac{1}{2}}\left(\frac{1}{2\pi\kappa_1^2}\right)$.

Proof. First, observe that $\Xi_\mathsf{A}(x; P_{X^\star})$, defined in (28), can be written as follows:

$$
\Xi_\mathsf{A}(x; P_{X^\star}) = \int_{\mathbb{R}} \frac{\xi_\mathsf{A}(y)}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(y-x)^2}{2}}\mathrm{d}y, \quad (32)
$$

$$
\text{where } \xi_\mathsf{A}(y) = \log\frac{1}{f_{Y^\star}(y)} - C(\mathsf{A}) - h(Z). \quad (33)
$$

Next, using the fact that the Gaussian distribution is a member of Polyá type-∞ functions,

$$
\begin{aligned}
|\mathsf{supp}(P_{X^\star})| &\leq \mathrm{N}(\mathbb{R}, \Xi_\mathsf{A}(\cdot; P_{X^\star})) \quad (34) \\
&\leq \mathscr{S}(\xi_\mathsf{A}) \quad (35) \\
&\leq \mathrm{N}(\mathbb{R}, \xi_\mathsf{A}) \quad (36) \\
&= \mathrm{N}\left(\mathbb{R}, f_{Y^\star} - \kappa_1\right) \quad (37) \\
&= \mathrm{N}\left([-R, R], f_{Y^\star} - \kappa_1\right) < \infty, \quad (38)
\end{aligned}
$$

where (34) is a consequence of Lemma 6, see Remark 5; (35) follows from Theorem 5; (36) follows because the number of zeros is an upper bound on the number of sign changes; (37) follows by observing that $\xi_\mathsf{A}(y) = 0$ if and only if $f_{Y^\star}(y) - \kappa_1 = 0$; and finally (38) follows from Lemma 1. $\square$

b) Proof of the Upper Bound: We begin by simplifying the previously provided upper bound on B_{κ_1}. Note that an amplitude constraint $|X| \leq \mathsf{A}$ induces a second moment constraint $\mathbb{E}[X^2] \leq \mathsf{A}^2$, and therefore

$$
C(\mathsf{A}) = \max_{\substack{|X|\leq\mathsf{A} \\ \mathbb{E}[X^2]\leq\mathsf{A}^2}} I(X; Y) \leq \frac{1}{2}\log\left(1 + \mathsf{A}^2\right). \quad (39)
$$

---

[4]A function $f : \mathbb{I}_1 \times \mathbb{I}_2 \to \mathbb{R}$ is said to be strictly Polyá type-n if $\det\left([p(x_i, y_j)]_{i,j=1}^m\right) > 0$ for all $1 \leq m \leq n$, and for all $x_1 < \cdots < x_m \in \mathbb{I}_1$, and $y_1 < \cdots < y_m \in \mathbb{I}_2$. If f is Polyá type-n for all $n \in \mathbb{N}$, then f is Polyá type-∞.

Since the differential entropy of a standard normal distribution is $h(Z) = \frac{1}{2}\log(2\pi e)$, (39) implies that

$$\frac{1}{\kappa_1} = \exp(C(\mathsf{A}) + h(Z)) \leq \sqrt{2\pi e\,(1+\mathsf{A}^2)}. \tag{40}$$

Capitalizing on the bound in (9),

$$\mathsf{B}_{\kappa_1} \leq \mathsf{A} + \sqrt{1 + \log(1+\mathsf{A}^2)} \leq 2\mathsf{A} + 1, \tag{41}$$

where the last inequality follows because $\log(1+x) \leq x$ and $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$.

As the finalizing step, letting $R \leftarrow (2\mathsf{A}+1)$ in Lemma 7 above, an application of Corollary 1 in Section III yields

$$\mathsf{N}([-\mathsf{B}_{\kappa_1}, \mathsf{B}_{\kappa_1}], f_{Y^\star} - \kappa_1)$$
$$= \mathsf{N}([-2\mathsf{A}-1, 2\mathsf{A}+1], f_{Y^\star} - \kappa_1) \tag{42}$$
$$\leq 1 + \min_{s>1}\left\{\frac{\frac{((3\mathsf{A}+1)s+\mathsf{A})^2}{2} + 2\mathsf{A}^2 + \log\left(2 + \frac{(3\mathsf{A}+1)s}{\mathsf{A}}\right)}{\log s}\right\} \tag{43}$$
$$\leq 1 + \min_{s>1}\left\{\frac{\frac{((3s+1)\mathsf{A}+s)^2}{2} + 2\mathsf{A}^2 + \log\left(2 + 4s\right)}{\log s}\right\} \tag{44}$$
$$\leq a_2\mathsf{A}^2 + a_1\mathsf{A} + a_0, \tag{45}$$

where (44) follows because $3\mathsf{A} + 1 \leq 4\mathsf{A}$ for[5] $\mathsf{A} \geq 1$; and (45) follows by choosing a suboptimal value $s = \sqrt{e}$ in the minimization and letting $a_2 = 9e + 6\sqrt{e} + 5$, $a_1 = 6e + 2\sqrt{e}$ and $a_0 = e + 2\log(4\sqrt{e} + 2) + 1$. $\square$

Remark 6. A more careful optimization of (45) over the parameter s would lead to better absolute constants a_0, a_1 and a_2. However, the order A^2 in (45) would not change.

c) Proof of the Lower Bound: Using the fact that the optimizing input distribution is discrete with finitely many points and denoting by $H(P_{X^\star})$ the entropy of the optimizing input distribution $P_{X^\star}$, it follows that

$$\frac{1}{2}\log\left(1 + \frac{2\mathsf{A}^2}{\pi e}\right) \leq \max_{X:\,|X|\leq\mathsf{A}} I(X;Y) \tag{46}$$
$$\leq H(P_{X^\star}) \tag{47}$$
$$\leq \log\left(|\mathsf{supp}(P_{X^\star})|\right), \tag{48}$$

where (46) is thanks to Shannon [22, Section 25]. $\square$

V. CONCLUDING REMARKS

This paper has introduced several new tools to study the capacity of amplitude constrained additive Gaussian channels. Not only are the introduced tools strong enough to show that the optimal input distribution is discrete with finite support, but they are also able to provide concrete upper bounds on the number of elements in that support. Moreover, the method has been demonstrated to be easily generalizable to other settings such as a scalar additive Gaussian channel with both peak and average power constraints. In addition to the scalar cases, the method is shown to work for a vector Gaussian channel with an amplitude constraint A. In particular, for an optimal input $\boldsymbol{X}^\star$ it has been shown that its magnitude $\|\boldsymbol{X}^\star\|$, is a discrete random variable with at most $O(\mathsf{A}^2)$ number of mass points for any fixed dimension n. Finally, it is highly likely

that the presented approach generalizes to other (possibly non-additive) channels where channel transition probability is given by a Polyá type-∞ function, e.g., the Poisson channel.

REFERENCES

[1] J. G. Smith, "The information capacity of amplitude-and variance-constrained scalar Gaussian channels," *Inform. and Contr.*, vol. 18, no. 3, 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. Inf. Theory*, vol. 41, no. 4, pp. 1060–1071, 1995.

[3] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, June 2005.

[4] B. Rassouli and B. Clerckx, "On the capacity of vector Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6884–6903, December 2016.

[5] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), "When are discrete channel inputs optimal? - Optimization techniques and some new results," in *Proc. Conf. on Inf. Sci. and Sys.*, Princeton, NJ, USA, March 2018, pp. 1–6.

[6] N. Sharma and S. Shamai (Shitz), "Transition points in the capacity-achieving distribution for the peak-power limited AWGN and free-space optical intensity channels," *Probl. Inf. Transm.*, vol. 46, no. 4, pp. 283–299, 2010.

[7] A. Dytso, H. V. Poor, and S. Shamai (Shitz), "On the capacity of the peak power constrained vector Gaussian channel: An estimation theoretic perspective," 2018. [Online]. Available: https://arxiv.org/pdf/1901.03264.pdf

[8] J. Huang and S. P. Meyn, "Characterization and computation of optimal distributions for channel coding," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2336–2351, 2005.

[9] T. Cover and J. Thomas, *Elements of Information Theory: Second Edition*. Wiley, 2006.

[10] A. Dytso, M. Goldenbaum, S. Shamai (Shitz), and H. V. Poor, "Upper and lower bounds on the capacity of amplitude-constrained MIMO channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, 2017, pp. 1–6.

[11] A. L. McKellips, "Simple tight bounds on capacity for the peak-limited discrete-time channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Chicago, IL, 2004, pp. 348–348.

[12] A. Thangaraj, G. Kramer, and G. Böcherer, "Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4172–4182, 2017.

[13] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, 2003.

[14] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[15] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, 2001.

[16] S. S. Shamai, "Capacity of a pulse amplitude modulated direct detection photon channel," *IEE Proceedings I (Communications, Speech and Vision)*, vol. 137, no. 6, pp. 424–430, 1990.

[17] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), "When are discrete channel inputs optimal?: Optimization techniques and some new results," in *Proc. 52nd Annu. Conf. Info. Scien. and Syst. (CISS)*, Princeton, USA, 2018, pp. 1–6.

[18] R. Tijdeman, "On the number of zeros of general exponential polynomials," in *Indagationes Mathematicae (Proceedings)*, vol. 74. North-Holland, 1971, pp. 1–7.

[19] A. Dytso, S. Yagli, H. V. Poor, and S. Shamai, "Capacity achieving distribution for the amplitude constrained additive Gaussian channel: An upper bound on the number of mass points," 2019. [Online]. Available: https://arxiv.org/pdf/1901.03264.pdf

[20] J. G. Smith, "On the Information Capacity of Peak and Average Power Constrained Gaussian Channels," PhD dissertation, University of California, 1969.

[21] S. Karlin, "Pólya type distributions, ii," *The Ann. Math. Stat.*, vol. 28, no. 2, pp. 281–308, 1957.

[22] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 379-423, 623-656, 1948.

[5]The unessential assumption that $\mathsf{A} \geq 1$ is just for simplifying the presentation. In the case when $\mathsf{A} \leq 1$, the optimality of $P_{X^\star}$ that is equiprobable on $\mathcal{X} = \{-\mathsf{A}, \mathsf{A}\}$ is known [6].