Amplitude Constrained Poisson Noise Channel: Properties of the Capacity-Achieving Input Distribution

Alex Dytso*, Luca Barletta**, Shlomo Shamai (Shitz)***
* New Jersey Institute of Technology, Newark, NJ 07102, USA, Email: alex.dytso@njit.edu
** Politecnico di Milano, Milano, 20133, Italy, Email: luca.barletta@polimi.it
*** Israel Institute of Technology, Technion City, Haifa 32000, Israel, Email: sshlomo@ee.technion.ac.il

***Abstract*—This work considers a Poisson noise channel with an amplitude constraint. It is well-known that the capacity-achieving input distribution for this channel is discrete with finitely many points. We sharpen this result by introducing upper and lower bounds on the number of mass points. In particular, the upper bound of order $\mathsf{A} \log^2(\mathsf{A})$ and lower bound of order $\sqrt{\mathsf{A}}$ are established where A is the constraint on the input amplitude. In addition, along the way, we show several other properties of the capacity and capacity-achieving distribution. For example, it is shown that the capacity is equal to $-\log P_{Y^\star}(0)$ where $P_{Y^\star}$ is the optimal output distribution. Moreover, an upper bound on the values of the probability masses of the capacity-achieving distribution and a lower bound on the probability of the largest mass point are established.**

I. Introduction

We consider a discrete-time memoryless Poisson channel. The output Y of this channel takes value on the set of nonnegative integers, and the input X takes value on the set of non-negative real numbers. The conditional probability mass function (pmf) of the output random variable Y given the input X that specifies the channel is given by

$$P_{Y|X}(y|x) = \frac{1}{y!} x^y \mathrm{e}^{-x},\, x \ge 0,\, y \in \mathbb{N} \cup \{0\}. \tag{1}$$

In (1), we use the standard convention that $0^0 = 1$ and $0! = 1$.

The capacity of this channel where the input X is subject to the amplitude constraint $0 \le X \le \mathsf{A}$ is given by

$$C(\mathsf{A}) = \max_{X:\, 0 \le X \le \mathsf{A}} I(X;Y), \qquad \mathsf{A} > 0. \tag{2}$$

Finding the capacity of this channel remains to be an elusive task. The goal of this work is to make progress on this problem by studying the properties of the capacity achieving distribution denoted by $P_{X^\star}$.

Prior Work: The now classical approach developed by Witsenhausen in [1] says that if the output alphabet has a cardinality n, then the support of the optimal input distribution cannot be more than n irrespective of the size of the input alphabet. However, since the output alphabet of the Poisson noise channel has a countably infinite alphabet, the Witsenhausen approach does not apply. Instead, the approach that has been applied to the Poisson noise channel largely follows the analyticity idea introduced by Smith in [2] in the context of amplitude constrained Gaussian noise channel. The interested reader is referred to [3] for a summary of these techniques. In this work, we also follow the latter technique. However, we considerably generalize and improve this approach. In what follows, we summarize the known results on the Poisson noise channel and highlight the elements of the new technique.

The discrete-time Poisson noise channel is suited to model low intensity, direct detection optical communication channels [4]; the interested reader is also referred to a survey on free space optical communications in [5]. The Poisson channel can be seen as a limiting case of the Binomial channel [6], which can be used to model the number of particles absorbed by a receiver unit in molecular communications [7]. A key difference in the mathematical formulation between the Poisson and the Binomial channel models is the infinite/finite nature of the output alphabet. In this work, we are concerned with the discrete-time channel; however, there also exists a large literature on continuous-time channels, and the interested reader is referred to a survey in [8].

The first major study of the capacity achieving distribution for the Poisson channel was undertaken in [9], where the authors consider the capacity of a Poisson channel in (1) with and without the additional power constraint on the input (i.e., $\mathbb{E}[X] \le \mathsf{P}$.)[1] The authors of [9] derived the Karush-Kuhn-Tucker (KKT) conditions that are necessary to study the structure of the capacity achieving distribution. These KKT conditions were then used to show that the support of an optimal input distribution for any A can contain at most one mass point in the open interval $(0,1)$. Moreover, for any $\mathsf{A} \le 1$, it was shown that the optimal input distribution consists of two mass points at 0 and A and the capacity is given by

$$P_{X^\star}(\mathsf{A}) = \frac{1}{\mathrm{e}^{\frac{\mathsf{A}}{\mathrm{e}^{\mathsf{A}}-1}} - \mathrm{e}^{-\mathsf{A}} + 1},\; P_{X^\star}(0) = 1 - P_{X^\star}(\mathsf{A}), \tag{3}$$

$$C(\mathsf{A}) = -\log\left(P_{X^\star}(0) + \mathrm{e}^{-\mathsf{A}} P_{X^\star}(\mathsf{A})\right). \tag{4}$$

The KKT conditions proposed in [9] were rigorously derived in [10, Corollary 1] and extended to a more general case that includes the possibility of a non-zero dark current parameter. Moreover, using the analyticity idea of [2], in [10] for $\mathsf{A} < \infty$ and any $\mathsf{P} > 0$ it was shown that the optimizing input distribution is unique and discrete with finitely many mass points. Moreover, for the case of $\mathsf{P} \ge \mathsf{A}$ (i.e., the power constraint is not active) and dark current is zero, it was shown that the distribution in (3) continues to be capacity achieving if and only if $\mathsf{A} \le \bar{\mathsf{A}}$ where $\bar{\mathsf{A}} \approx 3.3679$.

Further studies of the conditions under which the capacity achieving distribution is binary have been undertaken by the

[1] In the Poisson noise channel the power is measured in terms of the first moment of X.

authors of [11] and [12]. For example, in [11], it was shown that with both the amplitude and the power constraint, the optimal input distribution always contains a mass point at 0. Moreover, in the case of only an amplitude constraint, the optimal input distribution contains mass points at both 0 and A. In [12], it was shown that if $\mathsf{P} < \frac{\mathsf{A}}{2}$ and the dark current is large enough, the following binary distribution is optimal: $P_{X^\star}(0) = 1 - \frac{\mathsf{P}}{\mathsf{A}}$, $P_{X^\star}(\mathsf{A}) = \frac{\mathsf{P}}{\mathsf{A}}$. The capacity achieving distribution with only an average power constraint was considered in [13] and was shown to be discrete with infinitely many mass points.

The low power and the low amplitude asymptotics of the capacity have been studied in [14]–[18]. A number of papers have also focused on upper and lower bounds on the capacity. The first upper and lower bounds on the capacity have been derived in [9] for two situations: the case of the average power constraint only, and the case of both the average power and the amplitude constraint with $\mathsf{A} \le 1$. The authors of [19] derived upper and lower bounds, in the case of the average power constraint only, by focusing on the regime where both P and the dark current tend to infinity with a fixed ratio. Firm upper and lower bounds on the capacity in the case of only the average power constraint and no dark current have been derived in [14] and [15], and further improved in [13] and [20]. The most general bounds on the capacity that consider both the amplitude and the average power constraints on the input and hold for an arbitrary value of the dark current have been derived in [21]. The bounds in [21] have been shown to be tight in the regime where both the average power and the amplitude constraint approach infinity with a fixed ratio $\frac{\mathsf{P}}{\mathsf{A}}$.

In order to find an upper bound on the values of probability masses and a lower bound on the number of mass points, we will rely on the strong data-processing inequality for the relative entropy. The study of the strong data-processing inequalities has recently received some attention, and the interested reader is referred to [22]–[24].

To find an upper bound on the number of points in the support of $P_{X^\star}$, we will follow the proof technique developed in [25] for the Gaussian noise channel. The key step that we borrow from [25] is the use of the variation diminishing property, which is captured by the oscillation theorem of Karlin [26]. The oscillation theorem allows to upper bound the number of points in the support of $P_{X^\star}$ with the number of sign changes of a function that is related to the output distribution $P_{Y^\star}$. In the case of the Gaussian noise channel, in order to count the number of sign changes, one needs to resort to complex analytic techniques. In contrast, in the Poisson case, due to the discrete nature of the channel, we no longer need to rely on the complex analytic techniques, which simplifies this part of the analysis.

However, the analysis for the Poisson channel is not necessarily simpler than that in the Gaussian case. One crucial step in both proofs relies on finding a lower bound of the output pdf in the Gaussian case and the output pmf in the Poisson case. In the Gaussian case, this can be done by using Jensen's inequality, and the resulting bound is universal and is independent of $P_{X^\star}$. In the Poisson noise case, however, a distribution-independent bound cannot be obtained, and the lower bound on the tail depends on $P_{X^\star}$. Specifically, the lower bound depends on $P_{X^\star}(\mathsf{A})$. Therefore, to complete the proof, we need to also find a lower bound on $P_{X^\star}(\mathsf{A})$.

Some of the key intermediate steps in our proofs will rely on identities that connect information measures and estimation measures. In particular, we will rely on the expression for the conditional expectation derived in [27]. For further connections between estimation and information theoretic measure the interested reader is referred to [28] and [29] and references therein.

Notation: Throughout the paper, the deterministic scalar quantities are denoted by lower-case letters and random variables are denoted by uppercase letters.

We denote the distribution of a random variable X by P_X. The support set of P_X is denoted and defined as

$$\mathsf{supp}(P_X) = \{x : \text{ for every open set } \mathcal{D} \ni x \text{ we have that } P_X(\mathcal{D}) > 0\}. \quad (5)$$

Let g_1 and g_2 be non-negative functions, then

- $g_1(x) = O(g_2(x))$ means that there exists a constant $c > 0$ and x_0 such that $\frac{g_1(x)}{g_2(x)} \le c$ for all $x > x_0$;
- $g_1(x) = \Omega(g_2(x))$ means that $g_2(x) = O(g_1(x))$; and
- $g_1(x) = o(g_2(x))$ means $\lim_{x\to\infty} \frac{g_1(x)}{g_2(x)} = 0$.

II. Main Results

The main results of this paper are summarized in the following theorem.

Theorem 1. *The capacity and the capacity achieving distribution satisfy the following properties:*

- A New Capacity Expression: *For every* $\mathsf{A} \ge 0$*, the capacity is given by*
$$C(\mathsf{A}) = \log \frac{1}{P_{Y^\star}(0)}, \quad (6)$$
where $P_{Y^\star}$ *is the capacity achieving output distribution.*
- An Upper Bound on the Probabilities: *For every* $\mathsf{A} \ge 0$*,*

  Universal Bound:
$$P_{X^\star}(x) \le \mathrm{e}^{-\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}},\ x \in \mathsf{supp}(P_{X^\star}), \quad (7)$$
  Location Dependent Bound:
$$P_{X^\star}(x) \le \mathrm{e}^{-C(\mathsf{A}) - x\frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}}},\ x \in \mathsf{supp}(P_{X^\star}) \setminus \{0\}. \quad (8)$$
  In addition, if $|\mathsf{supp}(P_{X^\star})| = 2$*, then the bound in* (8) *becomes equality.*
- A Lower Bound on the Probability of the Largest Point: *For all* A *such that* $\mathrm{e}^{\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}} \ge 4$*, we have that*
$$P_{X^\star}(\mathsf{A}) \ge \frac{1 - 3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}}{2\mathsf{A}^{2\mathsf{A}\mathrm{e}\log(\mathsf{A})+2}\mathrm{e}^{-2\mathsf{A}+1}}. \quad (9)$$
- On the Location of Support Points: *Suppose that* $|\mathsf{supp}(P_{X^\star})| \ge 3$ *and let* $x^\star \in \mathsf{supp}(P_{X^\star}) \setminus \{0, \mathsf{A}\}$*. Then,*
$$\mathrm{e}^{-\sqrt{2(\log(\mathsf{A})-1)}} \le x^\star \le \mathsf{A} - 1. \quad (10)$$
- A Lower Bound on the Size of the Support: *For every* $\mathsf{A} \ge 0$*,*
$$|\mathsf{supp}(P_{X^\star})| \ge \mathrm{e}^{\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}. \quad (11)$$
- An Upper Bound on the Size of the Support: *For every* $\mathsf{A} \ge 0$*,*
$$|\mathsf{supp}(P_{X^\star})| \le \lceil \mathsf{A} - \log(P_{X^\star}(\mathsf{A})) - C(\mathsf{A}) \rceil + 2. \quad (12)$$

In addition, for all A *such that* $\mathrm{e}^{\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}} \geq 4$*, we have that*

$$|\mathsf{supp}(P_{X^\star})| \leq 2\mathrm{e}\mathsf{A}\log^2(\mathsf{A}) + 2\log(\mathsf{A}) - \mathsf{A} - \log\left(\frac{1-3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}}{2}\right) - C(\mathsf{A}) + 4. \quad (13)$$

A few comments are now in order.

A. Numerical Simulations

In order to aid our discussion, we have also numerically computed the optimal input distributions for values of A up to 15. Fig. 1 depicts the output of this simulation.

We note that there are several numerical recipes for generating an optimal input distribution [30]–[32]. However, most of these approaches ultimately optimize over the space of distributions, which is an infinite-dimensional space. As was already alluded to in [2] and [25], a firm upper bound on the number of mass points, such as the one in Theorem 1, allows us to move the optimization from the space of probability distributions to the space $\mathbb{R}^{2n}$ where n is the number of points. Working in $\mathbb{R}^{2n}$ allows us to employ methods such as the projected gradient ascent [33], which was used to generate the plots in Fig. 1.

B. On the Order of the Bounds

First, note that almost all of our bounds depend on the value of $C(\mathsf{A})$, which is currently unknown for $\mathsf{A} \geq \bar{\mathsf{A}} \approx 3.4$. However, this is not a limitation of our result, as we do have access to upper and lower bounds on $C(\mathsf{A})$ that are tight for large A such as those in [21], which suggest that

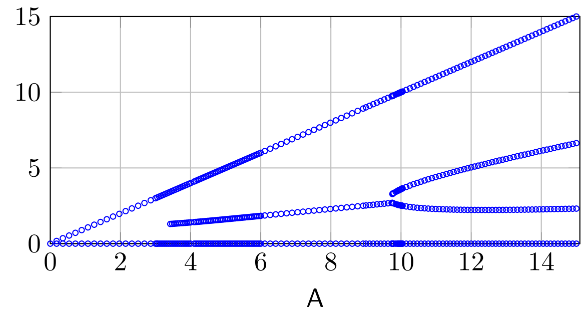
$$C(\mathsf{A}) = \frac{1}{2}\log(\mathsf{A}) - \frac{1}{2}\log\left(\frac{\pi\mathrm{e}}{2}\right) + o_{\mathsf{A}}(1). \quad (14)$$

Moreover, some of the bounds in Theorem 1, such as those in (9) and (13), while are firm, are meant to be used for large values of A. Therefore, combing the bounds in Theorem 1 with the bound in (14), we arrive at the following:

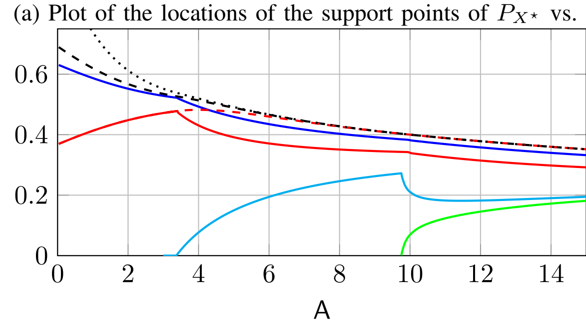
$$\Omega\left(\frac{\mathrm{e}^{2\mathsf{A}}}{\mathsf{A}^{2\mathsf{A}\mathrm{e}\log(\mathsf{A})+2}}\right) \leq P_{X^\star}(x) \leq O\left(\frac{1}{\sqrt{\mathsf{A}}}\right), \, x \in \mathsf{supp}(P_{X^\star}),$$
$$\Omega(\sqrt{\mathsf{A}}) \leq |\mathsf{supp}(P_{X^\star})| \leq O\left(\mathsf{A}\log^2(\mathsf{A})\right).$$

Thus, the order of the lower bound on the number of points is $\sqrt{\mathsf{A}}$, and the order of the upper bound on the number of points is $\mathsf{A}\log^2(\mathsf{A})$. It is interesting to speculate as to the reason why the bounds do not match and have different orders.

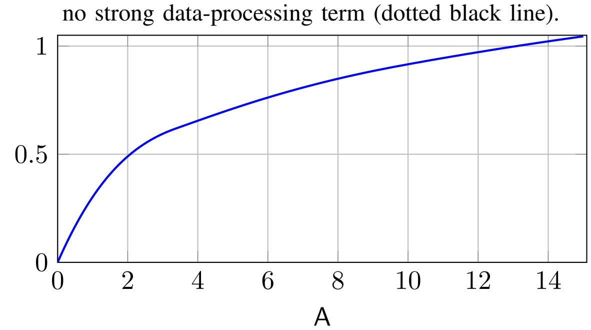
First, note that Theorem 1 presents two upper bounds on the number of points. The first and implicit bound in (12) depends on the value of $P_{X^\star}(\mathsf{A})$. The second bound in (13) is an explicit bound in terms of A and is derived by plugging in the lower bound on $P_{X^\star}(\mathsf{A})$ in (9) into the first bound in (12). We suspect that one of the reason why the bounds do not match is due to the lower bound on $P_{X^\star}(\mathsf{A})$ in (9), which we think is not tight. Hence, one interesting future direction is to improve the lower bound on the value of $P_{X^\star}(\mathsf{A})$ which, in view of (12), would lead to a better upper bound on $|\mathsf{supp}(P_{X^\star})|$. However, to the best of our knowledge, there are no other methods for finding lower or upper bounds on the probabilities of the optimal input distribution. Indeed, one of the contributions of this work is the introduction of two such methods, one for finding the upper bound on the values of probabilities and one for finding the lower bound.



(a) Plot of the locations of the support points of $P_{X^\star}$ vs. A.



(b) Plot of the probability values of the support points vs. A where points of the support satisfy $0 \leq x_1 \leq x_2 \leq \mathsf{A}$. The curves show: $P_{X^\star}(0)$ (solid blue line); $P_{X^\star}(\mathsf{A})$ (solid red line); $P_{X^\star}(x_1)$ (solid cyan line); $P_{X^\star}(x_2)$ (solid green line); upper bound in (7) (dashed black line); upper bound on $P_{X^\star}(\mathsf{A})$ in (8) (dashed red line); and upper bound in (7) with no strong data-processing term (dotted black line).



(c) Plot of $C(\mathsf{A})$ vs. A.

Fig. 1: Examples of the optimal input distributions and capacity vs. A. The probabilities for $\mathsf{A} > \bar{\mathsf{A}} \approx 3.4$ were computed numerically, and the probabilities for $\mathsf{A} \leq \bar{\mathsf{A}}$ are given in (3).

Finally, we would like to point out that numerical simulations are not useful for predicting the order of the number of points. For large A, the simulations become computationally demanding, and it is difficult to calculate the optimal input distribution and predict the order of the number of points.

C. On the Upper Bounds on the Probabilities in (7) and (8)

It is also interesting to ask how tight are the upper bounds on the probabilities in (7) and (8).

Note that the bound in (7) is universal and does not depend on the positions of the probability masses, while the bound in (8) depends on the position of the points. The advantage of the bound in (8) is that it can be tighter than the universal bound in (7). For example, in the regime where $\mathsf{A} \leq \bar{\mathsf{A}}$ where we only have two points in the support, the bound is achieved with equality. The clear disadvantage of the bound in (8) is that we do not know the location of the points (except 0 and A). However, such a bound might become useful once better estimates for the locations of the mass points are found. Some preliminary estimates of the locations are provided in (10).

Fig. 1b plots the upper bounds in (7) and (8) and compares them to the values of $P_{X^\star}$. To create the plot, in the regime $\mathsf{A} \le \bar{\mathsf{A}}$, we have used the exact expressions for $P_{X^\star}(\mathsf{A})$ and $P_{X^\star}(0)$ in (3). To create the plot in the regime $\mathsf{A} > \bar{\mathsf{A}}$, first note that the upper bounds in (7) can be loosened to

$$P_{X^\star}(x) \le \mathrm{e}^{-\frac{I(\widetilde{X};Y)}{1-\mathrm{e}^{-\mathsf{A}}}},\ x \in \mathsf{supp}(P_{X^\star}), \tag{15}$$

$$P_{X^\star}(x) \le \mathrm{e}^{-I(\widetilde{X};Y)-x\frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}}},\ x \in \mathsf{supp}(P_{X^\star}) \setminus \{0\}, \tag{16}$$

where we have used the fact that $C(\mathsf{A}) \ge I(\widetilde{X};Y)$ for any random variable $\widetilde{X} \in [0,\mathsf{A}]$. Therefore, since we can choose any $\widetilde{X} \in [0,\mathsf{A}]$, we selected it to be the one that is the output of the numerical simulation. The bound in (16) is only computed for $x = \mathsf{A}$ as we do not know the locations of other points and only have estimates for these. From the simulations in Fig. 1b, the bounds in (7) and (8) appear to be relatively tight.

The bound in (7) relies on the strong data-processing inequality. Specifically, the factor $\frac{1}{1-\mathrm{e}^{-\mathsf{A}}}$ in the exponent comes from using the strong data-processing inequality. The dotted black curve in Fig. 1b plots the loosened version of the bound in (7) that ignores the contribution of the strong data-processing inequality, that is we plot

$$P_{X^\star}(x) \le \mathrm{e}^{-I(\widetilde{X};Y)},\ x \in \mathsf{supp}(P_{X^\star}). \tag{17}$$

From the comparison in Fig. 1b, we see that the contribution of the strong data-processing inequality is non-trivial, especially for small and medium values of A.

D. On the Bound in (10)

In addition to finding bounds on the number of points and the values of the probabilities, we have also provided additional information about the location of the points. Specifically, (10) provides information about the location of support points other than 0 and A.

From the bound in (10), we see that the second-largest point can never be too close to A. Specifically, according to the bound in (10), the gap between A (i.e., the largest point) and the second-largest point is at least one. In fact, the numerical simulations shown in Fig. 1a suggest that this gap is much larger. In particular, the simulations suggest that the gap is not constant but is an increasing function of A. Therefore, one interesting future direction would be to verify this behavior and produce a better bound in (10) than $\mathsf{A} - 1$.

Similarly, from the lower bound in (10), we see that the second smallest point cannot be too close to the zero point. However, as A increases, the distance is allowed to get smaller. Note that our limited simulation results suggest a better lower bound, namely $x^\star \ge 1$. Therefore, one interesting future direction would be to either demonstrate the existence of a mass point in the range $(0, 1)$ or show that there is no such mass point. Note that the work of [9] already showed that there is at most one point in the range $(0, 1)$.

Beyond theoretical interest, the existence of the estimates for the mass points' location might also be of interest from the practical point of view. As the existence of such estimates can also impact the design of practical constellations for the Poisson noise channel.

E. On the Equivocation, Symbol Error Probability, and Entropy

It is well-known that the capacity-achieving distribution should be 'difficult' to detect or estimate on the per symbol basis. To make this statement explicit, we consider the equivocation $H(X^\star|Y^\star)$ and the probability of error under the maximum a posteriori (MAP) rule (i.e., $P_e = \mathbb{P}[X^\star \neq \hat{X}(Y^\star)]$ where $\hat{X}(Y^\star)$ is the MAP decoder). The plots of Fig. 2a and Fig. 2b show that the equivocation and P_e for the capacity achieving input have relatively high values. With Theorem 1 at our disposal, we can now show the following result regarding the asymptotic behavior of the error probability.

Proposition 1. *Let* $P_e = \mathbb{P}[X^\star \neq \hat{X}(Y^\star)]$ *where* $\hat{X}(y) = \arg\max_{x\in\mathsf{supp}(P_{X^\star})} P_{X^*|Y^*}(x|y)$*. Then,*

$$\liminf_{\mathsf{A}\to\infty} P_e \ge 1 - \sqrt{\frac{2}{\pi}}, \tag{18}$$

$$\liminf_{\mathsf{A}\to\infty} H(X^*|Y^*) \ge 2\left(1 - \sqrt{\frac{2}{\pi}}\right). \tag{19}$$

The entropy of the optimal input distribution vs. A is plotted in Fig. 2c. We observe a particular behavior of the entropy in the simulated range of A: the rate of increase has finite jumps approximately at the levels $\log(k)$ of entropy, where the cardinality of the optimal input distribution increases from k to $k{+}1$ points. These levels correspond to an approximate uniform input distribution on the k amplitude levels: this behavior is also confirmed by the mass probabilities plotted in Fig. 1b. When the rate of increase of the entropy is not compensated by a sufficiently large rate of decrease of the equivocation (see Fig. 2a), the rate of increase of capacity must be sustained by boosting the entropy: This is done by increasing the cardinality of the input distribution. It is interesting to understand how the rate of increase of capacity is split between entropy and equivocation: If one could prove that the equivocation is upper-bounded by a constant, then this would show that the equivocation does not provide degrees of freedom to channel capacity for large A, and thus the whole rate should be sustained by entropy by increasing the cardinality of $\mathsf{supp}(P_{X^\star})$. This hypothesis would imply $|\mathsf{supp}(P_{X^\star})| \approx \sqrt{\mathsf{A}}$ for large A.

III. Selected Proofs

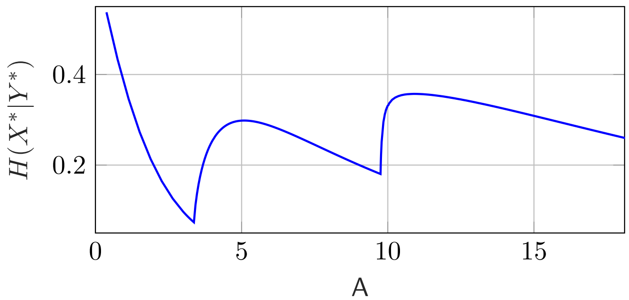
The starting point for most of our proofs are the following KKT conditions shown in [10].

Lemma 1. $P_{X^\star}$ *maximizes* (2) *if and only if*

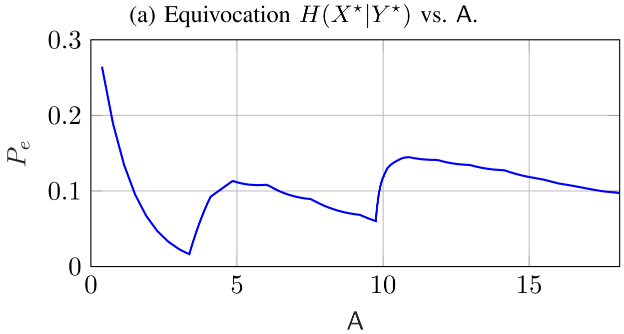
$$\mathsf{D}(P_{Y|X}(\cdot|x)\|P_{Y^\star}) \le C(\mathsf{A}),\ x \in [0,\mathsf{A}] \tag{20a}$$

$$\mathsf{D}(P_{Y|X}(\cdot|x)\|P_{Y^\star}) = C(\mathsf{A}),\ x \in \mathsf{supp}(P_{X^\star}). \tag{20b}$$

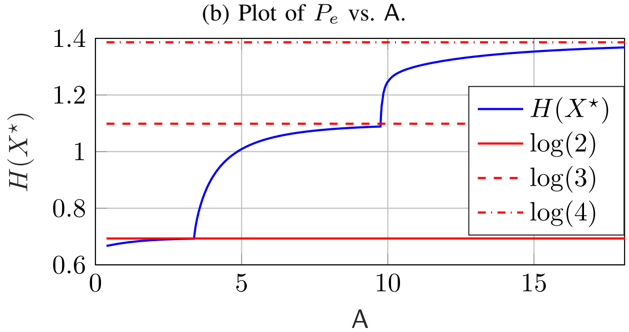
Due to space constraints, we only show a proof of the upper bound on the probabilities and the lower bound on the number of points, which rely on the data-processing argument.



(a) Equivocation $H(X^\star|Y^\star)$ vs. A.



(b) Plot of P_e vs. A.



(c) Entropy vs. A.

Fig. 2: Entropy (i.e., $H(X^\star)$), equivocation (i.e, $H(X^\star|Y^\star)$), and the probability of error of the optimal input distribution.

A. Proof of the Bound in (7) and of the Bound in (11)

We show two methods for finding bounds on the probabilities. The first method relies on the strong data-processing inequality and the second method relies on the exact expression for the values of the probability distribution. An interesting feature of both methods is that they work for all channels for which a capacity achieving distribution is discrete. Due to space constrains we only demonstrated the first technique.

Theorem 2. *For a channel $P_{Y|X}$ and the optimization problem*

$$C(\mathsf{A}) = \max_{0 \le X \le \mathsf{A}} I(X;Y), \tag{21}$$

suppose that a maximizing distribution $P_{X^\star}$ is discrete. Then,

$$P_{X^\star}(x) \le \mathrm{e}^{-\frac{1}{\eta_{\mathsf{KL}}(\mathsf{A};P_{Y|X})} C(\mathsf{A})}, \, x \in \mathsf{supp}(P_{X^\star}) \tag{22}$$

where $0 < \eta_{\mathsf{KL}}(\mathsf{A}; P_{Y|X}) \le 1$ is known as a contraction coefficient and is given by

$$\eta_{\mathsf{KL}}(\mathsf{A}; P_{Y|X}) = \sup_{Q_X:\, 0 \le X \le \mathsf{A},\, \mathsf{D}(Q_X\|P_{X^\star})<\infty,\, Q_X \to P_{Y|X} \to Q_Y} \frac{\mathsf{D}(Q_Y\|P_{Y^\star})}{\mathsf{D}(Q_X\|P_{X^\star})}. \tag{23}$$

Proof. Let Y_x be the output of the channel $P_{Y|X}$ when the input is $P_X = \delta_x$, where δ_x is the Dirac delta function centered in x. Next, suppose that $x \in \mathsf{supp}(P_{X^\star})$. Then, by using (20), we have that

$$\begin{aligned} C(\mathsf{A}) &= \mathsf{D}(P_{Y|X}(\cdot|x)\|P_{Y^\star}) && (24)\\ &= \mathsf{D}(P_{Y_x}\|P_{Y^\star}) && (25)\\ &\le \eta_{\mathsf{KL}}(\mathsf{A}; P_{Y|X})\mathsf{D}(\delta_x\|P_{X^\star}) && (26)\\ &= \eta_{\mathsf{KL}}(\mathsf{A}; P_{Y|X}) \log \frac{1}{P_{X^\star}(x)}, && (27) \end{aligned}$$

where in (26) we have used the strong data-processing inequality for the relative entropy [23] and where $\eta_{\mathsf{KL}}(\mathsf{A}; P_{Y|X})$ is defined in (23). This concludes the proof. □

The next result provides an upper bound on the contraction coefficient for the Poisson channel.

Lemma 2. *Let $P_{Y|X}$ be a Poisson channel as in* (1)*. Then, for all* $\mathsf{A} \ge 0$

$$\eta_{\mathsf{KL}}(\mathsf{A}; P_{Y|X}) \le 1 - \mathrm{e}^{-\mathsf{A}}. \tag{28}$$

This concludes the proof of the upper bounds on the values of the probabilities. The lower bound on the number of points in (13) is now a consequence of the upper bound on the values of $P_{X^\star}$:

$$1 = \sum_{x \in \mathsf{supp}(P_{X^\star})} P_{X^\star}(x) \le |\mathsf{supp}(P_{X^\star})| \mathrm{e}^{-\frac{1}{1-\mathrm{e}^{-\mathsf{A}}} C(\mathsf{A})}, \tag{29}$$

which simplifies to $\mathrm{e}^{\frac{1}{1-\mathrm{e}^{-\mathsf{A}}} C(\mathsf{A})} \le |\mathsf{supp}(P_{X^\star})|$.

IV. Conclusion

This work has focused on studying properties of the capacity-achieving distribution for the Poisson noise channel with an amplitude constraint. It was previously known that the capacity-achieving distribution for this channel is discrete with finitely many points. In this work, we sharpened this result in several ways.

First, by using a strong data-processing inequality, an upper bound on the values of the mass points has been shown. This upper bound on the probability values has been shown to lead to the lower bound on the number of support point of the optimal input distribution. Specifically, a lower bound of order $\sqrt{\mathsf{A}}$ has been established on the number of support points where A is the constraint on the amplitude.

Second, by using the variation-diminishing property of the Poisson kernel, the work has also established an upper bound on the number of the support points of the optimal input distribution. Specifically, an order $\mathsf{A} \log^2(\mathsf{A})$ bound has been established.

Finally, along the way, several other results have been shown. For example, a new compact expression for the capacity has been shown. In addition, a lower bound on the probability of the largest points of the optimal input distribution has been established. Furthermore, an estimate on the locations of the support other than 0 and A has been established.

References

[1] H. Witsenhausen, "Some aspects of convexity useful in information theory," *IEEE Transactions on Information Theory*, vol. 26, no. 3, pp. 265–271, 1980.

[2] J. G. Smith, "The information capacity of amplitude-and variance-constrained scalar Gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, 1971.

[3] 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. IEEE 52nd Annu.Conf. Info. Sci. Syst. (CISS)*. IEEE, 2018, pp. 1–6.

[4] J. P. Gordon, "Quantum effects in communications systems," *Proceedings of the IRE*, vol. 50, no. 9, pp. 1898–1908, 1962.

[5] M. A. Khalighi and M. Uysal, "Survey on free space optical communication: A communication theory perspective," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 2231–2258, 2014.

[6] C. Komninakis, L. Vandenberghe, and R. Wesel, "Capacity of the binomial channel, or minimax redundancy for memoryless sources," in *Proceedings. 2001 IEEE International Symposium on Information Theory (IEEE Cat. No. 01CH37252)*. IEEE, 2001, p. 127.

[7] N. Farsad, W. Chuang, A. Goldsmith, C. Komninakis, M. Médard, C. Rose, L. Vandenberghe, E. E. Wesel, and R. D. Wesel, "Capacities and optimal input distributions for particle-intensity channels," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 6, no. 3, pp. 220–232, 2020.

[8] S. Verdú, "Poisson communication theory," *International Technion Communication Day in Honor of Israel Bar-David*, vol. 66, 1999.

[9] R. McEliece, E. Rodemich, and A. Rubin, "The practical limits of photon communication," *Jet Propulsion Laboratory Deep Space Network Progress Reports*, vol. 42, pp. 63–67, 1979.

[10] 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.

[11] J. Cao, S. Hranilovic, and J. Chen, "Capacity-achieving distributions for the discrete-time Poisson channel Part i: General properties and numerical techniques," *IEEE Transactions on Communications*, vol. 62, no. 1, pp. 194–202, 2014.

[12] ——, "Capacity-achieving distributions for the discrete-time Poisson channel Part ii: Binary inputs," *IEEE Transactions on Communications*, vol. 62, no. 1, pp. 203–213, 2014.

[13] M. Cheraghchi and J. Ribeiro, "Improved capacity upper bounds for the discrete-time Poisson channel," *arXiv preprint arXiv:1801.02745*, 2018.

[14] A. Martinez, "Spectral efficiency of optical direct detection," *JOSA B*, vol. 24, no. 4, pp. 739–749, 2007.

[15] ——, "Achievability of the rate $\frac{1}{2}\log(1+\epsilon_s)$ in the discrete-time Poisson channel," *arXiv preprint arXiv:0809.3370*, 2008.

[16] A. Lapidoth, J. H. Shapiro, V. Venkatesan, and L. Wang, "The discrete-time Poisson channel at low input powers," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3260–3272, 2011.

[17] L. Wang and G. W. Wornell, "A refined analysis of the Poisson channel in the high-photon-efficiency regime," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4299–4311, 2014.

[18] ——, "The impact of dark current on the wideband Poisson channel," in *2014 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2014, pp. 2924–2928.

[19] D. Brady and S. Verdú, "The asymptotic capacity of the direct detection photon channel with a bandwidth constraint," in *28th Allerton Conf. Commun., Control and Comp.*, Oct. 1990, pp. 691–700.

[20] M. Cheraghchi and J. Ribeiro, "Non-asymptotic capacity upper bounds for the discrete-time Poisson channel with positive dark current," *arXiv preprint arXiv:2010.14858*, 2020.

[21] A. Lapidoth and S. M. Moser, "On the capacity of the discrete-time Poisson channel," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 303–322, 2009.

[22] M. Raginsky, "Strong data processing inequalities and ϕ -sobolev inequalities for discrete channels," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3355–3389, 2016.

[23] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and Bayesian networks," in *Convexity and Concentration*. Springer, 2017, pp. 211–249.

[24] F. du Pin Calmon, Y. Polyanskiy, and Y. Wu, "Strong data processing inequalities for input constrained additive noise channels," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1879–1892, 2017.

[25] A. Dytso, S. Yagli, H. V. Poor, and S. Shamai (Shitz), "The capacity achieving distribution for the amplitude constrained additive Gaussian channel: An upper bound on the number of mass points," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2006–2022, 2020.

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

[27] A. Dytso and H. V. Poor, "Estimation in Poisson noise: Properties of the conditional mean estimator," *IEEE Transactions on Information Theory*, vol. 66, no. 7, pp. 4304–4323, 2020.

[28] D. Guo, S. Shamai, and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1837–1849, 2008.

[29] R. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the Poisson channel," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1302–1318, 2012.

[30] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, 1972.

[31] C.-I. Chang and L. D. Davisson, "On calculating the capacity of an infinite-input finite (infinite)-output channel," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 1004–1010, 1988.

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

[33] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge university press, 2014.