Optimal Inputs for Some Classes of Degraded Wiretap Channels

Alex Dytso, Malcolm Egan, Samir Perlaza, H Vincent Poor, Shlomo Shamai

▶ **To cite this version:**

Alex Dytso, Malcolm Egan, Samir Perlaza, H Vincent Poor, Shlomo Shamai. Optimal Inputs for Some Classes of Degraded Wiretap Channels. 2018 IEEE Information Theory Workshop (ITW), Nov 2018, Guangzhou, China. hal-01884159

HAL Id: hal-01884159
https://inria.hal.science/hal-01884159

Submitted on 29 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal Inputs for Some Classes of Degraded Wiretap Channels

Alex Dytso, Malcolm Egan, Samir M. Perlaza, H. Vincent Poor and Shlomo Shamai (Shitz)

***Abstract*—In this paper, an analysis of an input distribution that achieves the secrecy capacity of a general degraded additive noise wiretap channel is presented. In particular, using convex optimization methods, an input distribution that achieves the secrecy capacity is characterized by conditions expressed in terms of integral equations. The new conditions are used to study the structure of the optimal input distribution for three different additive noise cases: vector Gaussian; scalar Cauchy; and scalar exponential.**

I. Introduction

The degraded wiretap channel consisting of a transmitter, a legitimate receiver and an eavesdropper is a fundamental information theoretic model. A key feature of the degraded wiretap channel is that capacity is achieved via a binning strategy. Different variants of this channel including discrete memoryless [1], scalar Gaussian [2] and Poisson [3] wiretap channels have been extensively studied, including the impact of peak power constraints [4].

In parallel, memoryless point-to-point channels have been studied with general noise models and constraints. This includes characterizations of the capacity and also the structure of the optimal input distributions. Although initial work focused on particular noise models and constraint sets [5], there has been recent success in establishing general conditions for an optimal input to be compactly supported and discrete or have unbounded support [6]. These results have all exploited convex optimization methods over sets of probability measures in a crucial way.

A fact that has not been widely exploited is that the secrecy rate for the degraded wiretap channel is concave. This ensures that many of the same methods developed to study point-to-point channels are also applicable to wiretap channels. As such, a natural question is the structure of the optimal input distributions and the interactions between these optimal distributions, the noise model and the constraint set.

In this paper, an optimal input distribution for degraded additive wiretap channels is studied under three noise models: Gaussian; Cauchy; and exponential. By exploiting the convex optimization methods originating in the study of point-to-point channels, we establish necessary and sufficient conditions for an input distribution to be optimal. These techniques also provide a means for deriving constraints on the input that guarantee that a given input distribution is optimal under general noise models.

Although Gaussian wiretap channels have been widely studied in [7], [8] and [9], the structure of the optimal input distribution is not known for degraded vector Gaussian wiretap channels under general linear constraints. Two examples of this class of constraints beyond power are absolute moment and logarithmic constraints. To this end, we provide a characterization of an optimal input distribution and, in particular, establish conditions when this distribution is discrete and compactly supported. As a byproduct of our analysis in the Gaussian noise case, we establish a new extremal inequality relying on conditions when an input uniformly distributed on a spherical shell is optimal.

We show that the same methodology also applies to scalar Cauchy and exponential noise channels. In the Cauchy noise case, we establish a constraint on the input such that a Cauchy input is the unique optimal input. For the exponential noise model, we study an optimal input distribution corresponding to a first moment constraint. In particular, we establish that an optimal input for the point-to-point exponential noise channels studied in [10] does not correspond to an optimal input for the degraded wiretap channel. This is, perhaps, surprising since for several known additive models (e.g., Gaussian and Cauchy) the input distribution that achieves the capacity of a point-to-point also achieves the capacity of the degraded wiretap channel.

A. Notation

Vectors are denoted by bold lowercase letters, random vectors by bold uppercase letters, and matrices by bold uppercase sans serif letters (e.g., $\mathbf{x}$, $\mathbf{X}$, $\mathbf{X}$). We denote the distribution of a random vector $\mathbf{X}$ by $P_{\mathbf{X}}$. Moreover, we say that a point x is in the support of the distribution $P_{\mathbf{X}}$ if for every open set $\mathcal{O}$ such that $x \in \mathcal{O}$ we have that $P_{\mathbf{X}}(\mathcal{O}) > 0$ and denote the collection of the support points of $P_{\mathbf{X}}$ as $\mathcal{E}(P_{\mathbf{X}})$. The set of all the distributions over a set $\mathcal{X}$ is denoted by $\triangle(\mathcal{X})$. The Gaussian distribution with mean μ and variance σ^2 is denote by $\mathcal{N}(\mu, \sigma^2)$. The Cauchy distribution with the location parameter μ and scaling parameter k is denoted by $\mathcal{C}(\mu, k)$. The Dirac delta measure at $\mathbf{x}_1$ is denote by $\delta_{\mathbf{x}_1}(\mathbf{x})$.

A. Dytso, S. M. Perlaza, and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ, USA (e-mail: {adytso, poor}@princeton.edu).

M. Egan and S. M. Perlaza are with the Laboratoire CITI (a joint laboratory between the Université de Lyon, INRIA, and INSA de Lyon), 6 Avenue des Arts, F-69621, Villeurbanne, France (e-mail: {malcom.egan, samir.perlaza}@inria.fr).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa, Israel (e-mail: sshlomo@ee.technion.ac.il).

This work was supported in part by the U. S. National Science Foundation under Grant CNS-1702808, and by the European Union's Horizon 2020 Research And Innovation Programme, grant agreement no. 694630.

We use the following parametrization of the mutual information in terms of the input distribution $P_{\mathbf{X}}$ and the random transformation $P_{\mathbf{Y}|\mathbf{X}}$:

$$I(P_{\mathbf{X}}, P_{\mathbf{Y}|\mathbf{X}}) \triangleq I(\mathbf{X};\mathbf{Y}).$$

We also define the following quantity that is akin to the information density:

$$i(\mathbf{x}, P_{\mathbf{X}}, P_{\mathbf{Y}|\mathbf{X}}) \triangleq \mathbb{E}\left[\log \frac{\mathrm{d}P_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X})}{\mathrm{d}P_{\mathbf{Y}}(\mathbf{Y})} \mid \mathbf{X}=\mathbf{x}\right],$$

where $P_{\mathbf{Y}}$ is the distribution of the channel output $\mathbf{Y}$. The differential entropy of a continuous random vector $\mathbf{X}$ is denoted by $h(\mathbf{X})$.

Let $n \in \mathbb{N}$ be fixed. An n-ball and an $(n-1)$-sphere of radius r centered at the origin are respectively denoted by $\mathcal{B}_0(r) \triangleq \{x : \|x\|_2 \leq r\}$ and $\mathcal{C}(r) \triangleq \{x : \|x\|_2 = r\}$, where $\|\cdot\|_2$ denotes the Euclidian norm.

Let $f(x)$ and $g(x)$ be two real-valued functions. We use the Landau notation $f(x) = o(g(x))$ to mean that for *every* $c > 0$ there exists an x_0 such that $f(x) < cg(x)$ for all $x \geq x_0$. Moreover, we say that $f(x) = \omega(g(x))$ if $\mathrm{g}(x) = o(f(x))$.

Due to space limitations, some of the proofs are omitted and can be found in an extended version of this paper [11].

II. Wiretap Channel

Consider an n-dimensional memoryless wiretap channel (WC) described by the Markov kernel $P_{\mathbf{Y}_1\,\mathbf{Y}_2|\mathbf{X}}$ with input $\mathbf{X}$ in $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big)$ and outputs $(\mathbf{Y}_1, \mathbf{Y}_2)$ in $\big(\mathbb{R}^{2n}, \mathcal{B}(\mathbb{R}^{2n})\big)$. The output $\mathbf{Y}_1$ is observed by the legitimate receiver whereas the output $\mathbf{Y}_2$ is observed by the malicious receiver. The input distribution $P_{\mathbf{X}}$ is such that given a set $\mathcal{X} \subset \mathbb{R}^n$ and a function $f: \mathcal{X} \to \mathbb{R}$, it satisfies:

$$P_{\mathbf{X}} \in \mathcal{F}(\mathcal{X}, f), \tag{1}$$

where,

$$\mathcal{F}(\mathcal{X}, f) \triangleq \Big\{ Q_{\mathbf{X}} \in \triangle\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big) : \mathcal{E}(Q_{\mathbf{X}}) = \mathcal{X} \text{ and } \mathbb{E}_{Q_{\mathbf{X}}}[f(\mathbf{X})] \leq 0 \Big\}. \tag{2}$$

With a slight abuse of notation, the set $\mathcal{F}(\mathcal{X}, f)$ in (2) is denoted by

$$\mathcal{F}(\mathcal{X}) \triangleq \{Q_{\mathbf{X}} \in \triangle\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big) : \mathcal{E}(Q_{\mathbf{X}}) = \mathcal{X}\}, \tag{3}$$

whenever the constraint is only over the support of the input distribution; or

$$\mathcal{F}(f) \triangleq \{Q_{\mathbf{X}} \in \triangle\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big) : \mathbb{E}_{Q_{\mathbf{X}}}[f(\mathbf{X})] \leq 0\}, \tag{4}$$

whenever the constraint is only on the expectation of f. In the following, the constraint set $\mathcal{F}(\mathcal{X}, f)$ is assumed to be equipped with the topology of weak convergence, which is known to be metrized by the Lévy-Prokhorov metric [12].

This analysis is restricted to the case of physically degraded WCs, i.e., WCs for which the Markov kernel $P_{\mathbf{Y}_1\,\mathbf{Y}_2|\mathbf{X}}$ factorizes as

$$P_{\mathbf{Y}_1\,\mathbf{Y}_2|\mathbf{X}} = P_{\mathbf{Y}_1|\mathbf{X}} P_{\mathbf{Y}_2|\mathbf{Y}_1}. \tag{5}$$

The maximum secrecy rate of a WC, denoted by $C_{\mathrm{s}} \in \mathbb{R}_+$, is referred to as the *secrecy capacity*. The following lemma fully characterizes the secrecy capacity of a degraded WC.

Lemma 1 (Secrecy Capacity of a WC)**.** *The secrecy capacity of an n-dimensional memoryless WC described by the Markov kernel $P_{\mathbf{Y}_1\,\mathbf{Y}_2|\mathbf{X}}$ with input $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n\big)$ and output $\big(\mathbb{R}^{2n}, \mathcal{B}(\mathbb{R}^{2n})\big)$ of the form in* (5) *subject to the input constraint in* (1) *is:*

$$C_{\mathrm{s}} = \sup_{P_{\mathbf{X}} \in \mathcal{F}(\mathcal{X}, f)} I(P_{\mathbf{X}}, P_{\mathbf{Y}_1|\mathbf{X}}) - I(P_{\mathbf{X}}, P_{\mathbf{Y}_2|\mathbf{X}}). \tag{6}$$

Throughout the rest of the paper the following assumption is made.

Assumption 1. *There exists at least one solution to the optimization problem in* (6)*, denoted by $P^\star_{\mathbf{X}}$, and it satisfies $\mathbb{E}_{P^\star_{\mathbf{X}}}[f(\mathbf{X})] = 0$.*

The next lemma provides sufficient conditions for Assumption 1 to hold true using the extreme value theorem [13].

Lemma 2. *Assumption 1 holds true if the constraint set $\mathcal{F}(\mathcal{X}, f)$ is compact in the topology of weak convergence and the objective function in* (6)*, i.e., $I(P_{\mathbf{X}}, P_{\mathbf{Y}_1|\mathbf{X}}) - I(P_{\mathbf{X}}, P_{\mathbf{Y}_2|\mathbf{X}})$, is weakly continuous on $\mathcal{F}(\mathcal{X}, f)$.*

III. Preliminaries

The following notion of weak or directional derivative will be useful in our analysis.

Definition 1. (The Gâteaux Derivative.) *Let $\mathcal{G}$ be a locally convex topological space. For any two elements $P \in \mathcal{G}$ and $Q \in \mathcal{G}$, the Gâteaux derivative of a functional $G : \mathcal{G} \to \mathbb{R}$ at P in the direction of Q is*

$$\Delta_Q G(P) \triangleq \lim_{\lambda \to 0} \frac{G\big((1-\lambda)P + \lambda Q\big) - G(P)}{\lambda}. \tag{7}$$

The functional G is said to be Gâteaux differentiable at P if its Gâteaux derivative exists at P for all $Q \in \mathcal{G}$.

The following theorem provides a characterization of the Gâteaux derivative of the solution to the optimization problem in (6).

Theorem 1. *Let $G : \mathcal{F}(\mathcal{X}, f) \to \mathbb{R}$ be the functional given by*

$$G(P_{\mathbf{X}}) = I(P_{\mathbf{X}}, P_{\mathbf{Y}_1|\mathbf{X}}) - I(P_{\mathbf{X}}, P_{\mathbf{Y}_2|\mathbf{X}}). \tag{8a}$$

Then, if Assumption 1 holds, the functional G is Gâteaux differentiable at $P^\star_{\mathbf{X}}$ and the derivative is given by

$$\begin{aligned}
&\Delta_{Q_{\mathbf{X}}} G(P^\star_{\mathbf{X}}) \\
&= \mathbb{E}_{Q_{\mathbf{X}}}\left[\log\left(\frac{\mathrm{d}P_{\mathbf{Y}_1|\mathbf{X}}(\mathbf{Y}_1|\mathbf{X})}{\mathrm{d}P_{\mathbf{Y}_1}(\mathbf{Y}_1; P^\star_{\mathbf{X}})}\right)\right] - I(P^\star_{\mathbf{X}}, P_{\mathbf{Y}_1|\mathbf{X}}) \\
&\quad - \mathbb{E}_{Q_{\mathbf{X}}}\left[\log\left(\frac{\mathrm{d}P_{\mathbf{Y}_2|\mathbf{X}}(\mathbf{Y}_2|\mathbf{X})}{\mathrm{d}P_{\mathbf{Y}_2}(\mathbf{Y}_2; P^\star_{\mathbf{X}})}\right)\right] + I(P^\star_{\mathbf{X}}, P_{\mathbf{Y}_2|\mathbf{X}})
\end{aligned} \tag{8b}$$

for all $Q_{\mathbf{X}} \in \mathcal{F}(\mathcal{X}, f)$, where for all $k \in \{1, 2\}$, given a joint distribution $P_{\mathbf{Y}_k|\mathbf{X}} P^\star_{\mathbf{X}}$, the term $P_{\mathbf{Y}_k}(\mathbf{Y}_k; P^\star_{\mathbf{X}})$ denotes the marginal distribution of $\mathbf{Y}_k$.

Proof: The proof follows by generalizing the argument used in [6]. In particular, the Gâteaux differentiability of $G(P_{\mathbf{X}})$ follows from the differentiability of $I(P_{\mathbf{X}}, P_{\mathbf{Y}_1|\mathbf{X}})$ and $I(P_{\mathbf{X}}, P_{\mathbf{Y}_2|\mathbf{X}})$, and the linearity of the Gâteaux derivative. ■

Using Theorem 1, a sufficient and necessary condition for an input distribution to be the solution to (6) can be stated.

Theorem 2. *Let Assumption 1 hold. Then, $P_{\mathbf{X}^\star}$ is a solution to* (6) *if and only if there exists a strictly positive real λ such that the following hold:*

(a) *for all $\mathbf{x} \in \mathcal{E}(P_{\mathbf{X}}^\star)$*

$$\begin{aligned} &i(\mathbf{x}, P_{\mathbf{X}}^\star, P_{\mathbf{Y}_1|\mathbf{X}}) - i(\mathbf{x}, P_{\mathbf{X}}^\star, P_{\mathbf{Y}_2|\mathbf{X}}) \\ &\quad - \lambda(f(\mathbf{x}) - \mathbb{E}_{P_{\mathbf{X}}^\star}[f(\mathbf{X})]) \\ &= I(P_{\mathbf{X}}^\star, P_{\mathbf{Y}_1|\mathbf{X}}) - I(P_{\mathbf{X}}^\star, P_{\mathbf{Y}_2|\mathbf{X}}); \text{ and} \end{aligned} \tag{9a}$$

(b) *for all $\mathbf{x} \in \mathcal{X} \setminus \mathcal{E}(P_{\mathbf{X}}^\star)$*

$$\begin{aligned} &i(\mathbf{x}, P_{\mathbf{X}}^\star, P_{\mathbf{Y}_1|\mathbf{X}}) - i(\mathbf{x}, P_{\mathbf{X}}^\star, P_{\mathbf{Y}_2|\mathbf{X}}) \\ &\quad - \lambda(f(\mathbf{x}) - \mathbb{E}_{P_{\mathbf{X}}^\star}[f(\mathbf{X})]) \\ &< I(P_{\mathbf{X}}^\star, P_{\mathbf{Y}_1|\mathbf{X}}) - I(P_{\mathbf{X}}^\star, P_{\mathbf{Y}_2|\mathbf{X}}). \end{aligned} \tag{9b}$$

Theorem 2 provides a basis for characterizing the structure of an optimal input in a range of linear channels with different types of noise, e.g., Gaussian noise, Cauchy noise, and exponential noise.

IV. Gaussian Noise

This section focuses on the case in which the Markov kernel $P_{\mathbf{Y}_1\,\mathbf{Y}_2|\mathbf{X}}$ factorizes as $P_{\mathbf{Y}_1|\mathbf{X}} P_{\mathbf{Y}_2|\mathbf{X}}$ and for all $\mathbf{x} \in \mathbb{R}^n$, the probability measure $P_{\mathbf{Y}_1|\mathbf{X}=\mathbf{x}}$ is $\mathcal{N}(\sqrt{\mathsf{snr}}_1\mathbf{x}, \mathbf{I}_n)$ and the probability measure $P_{\mathbf{Y}_2|\mathbf{X}=\mathbf{x}}$ is $\mathcal{N}(\sqrt{\mathsf{snr}}_2\mathbf{x}, \mathbf{I}_n)$, with $\mathsf{snr}_1 \geq \mathsf{snr}_2$ and $\mathbf{I}_n$ the n-dimensional identity matrix. More specifically, let $\mathbf{Z}_1$ and $\mathbf{Z}_2$ be n-dimensional vectors whose entries are independent and identically distributed Gaussian random variables with zero means and unit variances. Hence, for all $k \in \{1, 2\}$,

$$\mathbf{Y}_k = \sqrt{\mathsf{snr}_k}\mathbf{X} + \mathbf{Z}_k, \tag{10}$$

where the input $\mathbf{X}$ follows a distribution $P_{\mathbf{X}}$ that satisfies (1), for some specific set $\mathcal{X}$ and function f. This particular case is referred to as *the degraded Gaussian WC*.

The following theorem is a consequence of Theorem 1.

Theorem 3. *Let Assumption 1 hold in the degraded Gaussian WC. Then, $P_{\mathbf{X}}^\star$ possesses the following properties:*

- (i) *if $\mathsf{snr}_1 > \mathsf{snr}_2$, then $P_{\mathbf{X}}^\star$ is unique;*
- (ii) *if $f(\mathbf{x}) = \omega(\|\mathbf{x}\|^2)$, then there exists some $R > 0$ such that $\mathcal{E}(P_{\mathbf{X}}^\star) \subset \mathcal{B}_0(R)$ (i.e., $P_{\mathbf{X}}^\star$ has bounded support);*
- (iii) *if $f(\mathbf{x}) = a\|\mathbf{x}\|^2 - b$, for some strictly positive reals a and b, then $P_{\mathbf{X}}^\star$ is $\mathcal{N}(0, \frac{b}{a}\mathbf{I}_n)$. Moreover, this is the only choice of f under which a Gaussian distribution is optimal;*
- (iv) *if $f(\mathbf{x}) = o(\|\mathbf{x}\|^2)$, then for all $R > 0$ we have that $\mathcal{E}(P_{\mathbf{X}}^\star) \cap \mathcal{B}_0(R)^c \neq \emptyset$; and*
- (v) *if the function $f(\cdot)$ satisfies the following: (a) $f(\mathbf{x}) \neq c\|\mathbf{x}\|^2$ for any constant c; (b) $f(\cdot)$ is a radial function, i.e., it depends on $\mathbf{x}$ only through $\|\mathbf{x}\|$; and (c) $f(\cdot)$ is an analytic function of $\|\mathbf{x}\|$. Then,*

$$\mathcal{E}(P_{\mathbf{X}}^\star) = \bigcup_{i=1}^{N} \mathcal{C}(r_i), \tag{11}$$

where $N \leq \infty$ (possibly infinite) and where the sequence $\{r_i\}_{i=1}^N$ does not have an accumulation point.

A. Extremal Inequalities for Gaussian Noise Case

Let $\mathbf{X}_G$ be an n-dimensional vector whose entries are Gaussian random variables with zero means and finite variances. Using this notation, the simplified version of the extremal inequality in [14] can be written as

$$\begin{aligned} &\max_{P_{\mathbf{X}} \in \mathcal{F}(\mathbb{R}^n, \|\mathbf{x}\|^2 - c)} h(\sqrt{\mathsf{snr}_1}\mathbf{X} + \mathbf{Z}_1) - h(\sqrt{\mathsf{snr}_2}\mathbf{X} + \mathbf{Z}_2) \\ &= h(\sqrt{\mathsf{snr}_1}\mathbf{X}_G + \mathbf{Z}_1) - h(\sqrt{\mathsf{snr}_2}\mathbf{X}_G + \mathbf{Z}_2), \end{aligned} \tag{12}$$

for a given positive real c. Several interesting inequalities that are reminiscent of this extremal inequality can be obtained from Theorem 3. The following theorem presents one of these results.

Theorem 4. *Consider an n-dimensional degraded Gaussian WC under the assumption that $P_{\mathbf{X}} \in \mathcal{F}(\mathcal{B}_0(R))$ for some $R > 0$ such that $\sqrt{\mathsf{snr}_1}R \leq \sqrt{n}$. Let also $\mathbf{X}^\circ$ be uniformly distributed in $\mathcal{C}(R)$. Then, the following holds*

$$\begin{aligned} &\max_{P_{\mathbf{X}} \in \mathcal{F}(\mathcal{B}_0(R))} h(\sqrt{\mathsf{snr}_1}\mathbf{X} + \mathbf{Z}_1) - h(\sqrt{\mathsf{snr}_2}\mathbf{X} + \mathbf{Z}_2) \\ &= h(\sqrt{\mathsf{snr}_1}\mathbf{X}^\circ + \mathbf{Z}_1) - h(\sqrt{\mathsf{snr}_2}\mathbf{X}^\circ + \mathbf{Z}_2). \end{aligned} \tag{13}$$

Proof: The proof follows by setting $P_{\mathbf{X}}^\star = P_{\mathbf{X}^\circ}$ in (9) and characterizing conditions on R such that the sufficient and necessary conditions in (9) still hold. ■

The conditional version of the extremal inequality in (12) can be used to prove a converse for the Gaussian noise broadcast channel (BC) with the power constraint on $\mathbf{X}$. An interesting extension would be to use the inequality in (13) to establish the converse for the Gaussian noise BC with an amplitude constraint.

V. Cauchy Noise

This section focuses on the one-dimensional case in which the Markov kernel $P_{Y_1\,Y_2|X}$ factorizes as $P_{Y_1|X} P_{Y_2|X}$ and for all $x \in \mathbb{R}$, the probability measure $P_{Y_1|X=x}$ is $\mathcal{C}(x, \gamma_1)$ and the probability measure $P_{Y_2|X=x}$ is $\mathcal{C}(x, \gamma_2)$, with $\gamma_2 \geq \gamma_1 > 0$. More specifically, let N_1 and N_2 be independent and distributed following a Cauchy distribution with location parameters equal to zero and scale parameters equal to γ_1 and γ_2, respectively. Hence, for all $k \in \{1, 2\}$, the probability density function (pdf) of N_k is

$$f_{N_k}(x) = \frac{1}{\pi\gamma_k\left(1 + \left(\frac{x}{\gamma_k}\right)^2\right)}, \tag{14}$$

and

$$Y_k = X + N_k, \tag{15}$$

where the input X follows a distribution P_X that satisfies (1), for some specific set $\mathcal{X}$ and function f. This particular case is referred to as *the degraded Cauchy WC*.

The following lemma is instrumental to state the main results in this section.

Lemma 3. *Let P_U be $\mathcal{C}(0,P)$ and for all $u \in \mathbb{R}$, let $P_{V|U=u}$ be $\mathcal{C}(u,\gamma)$, with $P>0$ and $\gamma>0$. Then,*

$$\begin{aligned} i(u;P_U,P_{V|U}) = \log\left(\left(\frac{P+2\gamma}{P+\gamma}\right)^2 + \left(\frac{u}{P+\gamma}\right)^2\right) \\ + \log\left(\pi(P+\gamma)\right) - \log(4\pi\gamma), \end{aligned} \tag{16}$$

and

$$I(P_U,P_{V|U}) = \log\left(\frac{P+\gamma}{\gamma}\right). \tag{17}$$

The following theorem presents a necessary and sufficient condition for the Cauchy distribution with null location parameter and scale parameter $P>0$, to be the optimal input distribution of a degraded Cauchy WC.

Theorem 5. *Consider the degraded Cauchy WC subject to $P_X \in \{Q_X \in \triangle\big(\mathbb{R},\mathcal{B}(\mathbb{R})\big) : \mathbb{E}_{Q_X}[f(X)]=0\}$, for some particular function f. Let also $P>0$ be fixed and X_C be $\mathcal{C}(0,P)$. Then,*

$$\begin{aligned} &\max_X I(X;X+N_1) - I(X;X+N_2) \\ &= I(X;X_C+N_1) - I(X_C;X_C+N_2), \end{aligned} \tag{18}$$

if and only if

$$f(x) = \lambda \log\left(\frac{\left(\frac{P+2\gamma_1}{P+\gamma_1}\right)^2 + \left(\frac{x}{P+\gamma_1}\right)^2}{\left(\frac{P+2\gamma_2}{P+\gamma_2}\right)^2 + \left(\frac{x}{P+\gamma_2}\right)^2}\right), \tag{19}$$

where f is unique up to the multiplicative constant λ.

Proof: Denote by $P_X^\star$ the solution to the optimization problem in (6) and assume that it is $\mathcal{C}(0,P)$. Then, using Lemma 3, the lefthand side of the necessary and sufficient condition in (9a) becomes

$$\begin{aligned} &i(x;P_X^\star,P_{Y_1|X}) - i(x;P_X^\star,P_{Y_2|X}) \\ &= \log\left(\left(\frac{P+2\gamma_1}{P+\gamma_1}\right)^2 + \left(\frac{x}{P+\gamma_1}\right)^2\right) \\ &+ \log\left(\pi(P+\gamma_1)\right) - \log(4\pi\gamma_1) \\ &- \log\left(\left(\frac{P+2\gamma_2}{P+\gamma_2}\right)^2 + \left(\frac{x}{P+\gamma_2}\right)^2\right) \\ &- \log\left(\pi(P+\gamma_2)\right) + \log(4\pi\gamma_2), \end{aligned} \tag{20}$$

and the mutual information on the right side of (9a) becomes

$$I(P_X,P_{Y_1|X}) - I(P_X,P_{Y_2|X}) = \log\left(\frac{\gamma_2}{\gamma_1}\frac{P+\gamma_1}{P+\gamma_2}\right). \tag{21}$$

By combining (20) and (21) the sufficient and necessary condition in (9a) can be written as follows:

$$\log\left(\frac{\left(\frac{P+2\gamma_1}{P+\gamma_1}\right)^2 + \left(\frac{x}{P+\gamma_1}\right)^2}{\left(\frac{P+2\gamma_2}{P+\gamma_2}\right)^2 + \left(\frac{x}{P+\gamma_2}\right)^2}\right) = \lambda f(x). \tag{22}$$

This concludes the proof. ∎

Theorem 5 shows that the input constraint set for which the Cauchy input is optimal in the Cauchy WC depends on the channel parameters γ_1 and γ_2. Note that this is not the case for the Gaussian wiretap channel with the second moment constraint in which the optimal Gaussian input depends only on the input power constraint. A similar situation also arises in the Cauchy point-to-point channel. More specifically, by letting $\gamma_2 \to \infty$, it follows from Theorem 5 that the secrecy capacity of the degraded Cauchy WC is the capacity of a point-to-point Cauchy channel, which was already addressed in [15]. The following corollary describes this observation.

Corollary 1. *Consider the Cauchy point-to-point channel subject to $P_X \in \{Q_X \in \triangle\big(\mathbb{R},\mathcal{B}(\mathbb{R})\big) : \mathbb{E}_{Q_X}[f(X)]=0\}$, for some particular function f. Let also $P>0$ be fixed and X_C be $\mathcal{C}(0,P)$. Then,*

$$\max_X I(X;X+N) = I(X_C;X_C+N), \tag{23}$$

if and only if

$$f(x) = \log\left(\left(\frac{P+2\gamma}{P+\gamma}\right)^2 + \left(\frac{x}{P+\gamma}\right)^2\right) - \log(4). \tag{24}$$

Note that the input constraint set for which the Cauchy input is optimal in the point-to-point channel in Corollary 1 depends indeed on the channel parameter γ in (24). One interpretation of this result is that the constraint in (24) is not an input constraint but rather an output constraint. That is

$$\mathbb{E}[f(X)] = \mathbb{E}\left[\log\left(1+\frac{Y^2}{d^2}\right) - \log(4)\right] = 0, \tag{25}$$

for some constant d.

VI. Exponential Noise

This section focuses on the one-dimensional case in which the Markov kernel $P_{Y_1\,Y_2|X}$ factorizes as $P_{Y_1|X}P_{Y_2|X}$ and for all $k\in\{1,2\}$, the output Y_k is

$$Y_k = X + N_k, \tag{26}$$

where N_k follows an exponential distribution with parameter λ_k. That is, for all $x\geq 0$, the pdf of N_k is $f_{N_k}(x) = \lambda_k \mathrm{e}^{\lambda_k x}$, with $\lambda_1 > \lambda_2$. The input X is assumed to follow a distribution P_X that satisfies (1), with $\mathcal{X} = \mathbb{R}^+ \triangleq \{x : x\geq 0\}$ and a specific function f. This particular case is referred to as *the exponential WC*.

The following lemma shows that the exponential WC is degraded and thus, the result of Theorem 2 holds.

Lemma 4. *Let U be a random variable with a pdf such that for all $u \geq 0$,*

$$f_U(u) = \alpha\delta_0(u) + (1-\alpha)\lambda \mathrm{e}^{-\lambda u}, \tag{27}$$

for some $\alpha \in [0,1]$. Let also V be an exponential random variable with a parameter β and independent of U. Then, the random variable $Z = U + V$ has the following pdf such that for all $z \geq 0$:

$$f_Z(z) = \alpha\beta \mathrm{e}^{-\beta z} + (1-\alpha)\,\lambda\beta \frac{\mathrm{e}^{-\lambda z} - \mathrm{e}^{-\beta z}}{\beta - \lambda}. \tag{28}$$

Note than when $\alpha = \frac{\lambda}{\beta}$ and $\lambda \leq \beta$, the random variable Z in Lemma 4 is an exponential random variable. This implies the following corollary.

Corollary 2. *An exponentially distributed random variable is self-decomposable.*

In the exponential WC, in contrast to the Gaussian WC and the Cauchy WC, the noise is not closed under convolutions for all choices of parameters. Nevertheless, according to Lemma 4, exponential noise is self-decomposable; that is, for each exponential random variable Z, there exists two independent random variables Z_1 and Z_2 such that $Z \stackrel{d}{=} Z_1 + Z_2$ where Z_1 has a pdf according to (27) and Z_2 is an exponential random variable. This fact can be readily used to show that the pdf in (27) can achieve the capacity of a point-to-point channel under the first moment constraint on the input [10].

Theorem 6. *Let N be an exponential random variable with parameter $\beta > 0$ and $X^\star$ be distributed according to* (27) *with $\alpha = \frac{c}{c+\beta}$ and $\frac{1}{\lambda} = \frac{1}{c} + \frac{1}{\beta}$. Then, for all $P_X \in \{Q_X \in \triangle\big(\mathbb{R}, \mathcal{B}(\mathbb{R}^+)\big) : \mathbb{E}_{Q_X}[X - \frac{1}{c}] = 0\}$, it holds that*

$$I(X; X+N) \leqslant I(X^\star; X^\star + N). \tag{29}$$

The optimality of the distribution in (27) for a point-to-point channel with a first moment constraint in (29) comes from the fact that it induces an exponential distribution on the output of the channel which is an entropy maximizing distribution under the first moment constraint. As was shown in Section IV and Section V, for Gaussian and Cauchy noise, this entropy maximization paradigm can be extend from a point-to-point channel to a corresponding wiretap channel. In other words, for Gaussian and Cauchy noises the distribution that achieves the capacity of a point-to-point channel and induces a maximum entropy distribution on the output of the channel also achieves the capacity of a corresponding wiretap channel and induces a maximizing entropy distribution on both outputs of the wiretap channel. Interestingly, however, unlike for the Gaussian and the Cauchy cases, the distribution that achieves the capacity of a point-to-point exponential noise channel no longer achieves the capacity of an exponential noise wiretap channel.

Theorem 7. *Suppose that the constraint function is given by $f(x) = x - \frac{1}{P}$. Then, for $\lambda_i > 0$, the input distribution in* (27) does not *achieve the capacity of a wiretap channel in* (26).

The result of Theorem 7 is, perhaps, surprising as it shows that the distribution that achieves the capacity of a point-to-point channel does not achieve the capacity of a degraded wiretap channel. We suspect that the reason the input distribution in (6) does not achieve capacity of a wire-tap channel has to do with the fact that the exponential distribution is not a stable distribution. In fact, we conjecture that the only additive channels for which the maximizing input random variable is the same (up to a linear transformation) for a point-to-point and wiretap channels are those with stable noise.

VII. Conclusion

The capacity achieving input distributions for non-Gaussian degraded wiretap channels have been considered. Using convex optimization methods, a new characterization for the optimal inputs has been derived. By using this characterization, optimal inputs for vector Gaussian, scalar Cauchy and scalar exponential noise channels have been studied. Moreover, the optimal input for degraded vector Gaussian wiretap channels has been obtained in general settings. As a byproduct, a new extremal inequality has been demonstrated and an avenue of future work is to establish a proof via I-MMSE methods [16].

References

[1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[2] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, 1978.

[3] A. Laourine and A. B. Wagner, "The degraded Poisson wiretap channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 12, pp. 7073–7085, 2012.

[4] O. Ozel, E. Ekrem, and S. Ulukus, "Gaussian wiretap channel with amplitude and variance constraints," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5553–5563, 2015.

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

[6] 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. Inf. Sci. Syst. (CISS)*, Princeton, NJ, USA, 2018, to appear.

[7] F. Oggier and B. Hassibi, "A perspective on the MIMO wiretap channel," *Proc. of IEEE*, vol. 103, no. 10, pp. 1874–1882, 2015.

[8] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Information theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4–5, pp. 355–580, 2009.

[9] H. V. Poor and R. F. Schaefer, "Wireless physical layer security," *Proc. the Natl. Acad. Sci. U.S.A.*, vol. 114, no. 1, pp. 19–26, 2017.

[10] S. Verdú, "The exponential distribution in information theory," *Problemy Peredachi Informatsii*, vol. 32, no. 1, pp. 100–111, 1996.

[11] A. Dytso, M. Egan, M. S. Perlaza, H. V. Poor, and S. Shamai (Shitz), "On the capacity of non-Gaussian wiretap channels and extremal inequalities," 2018. [Online]. Available: http://www.princeton.edu/%7Eadytso/papers/ITW2018Extended.pdf

[12] R. M. Dudley, *Real Analysis and Probability*. Cambridge University Press, 2002, vol. 74.

[13] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, 1997.

[14] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839 –1851, 2007.

[15] J. Fahs and I. Abou-Faycal, "A Cauchy input achieves the capacity of a Cauchy channel under a logarithmic constraint," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, 2014, pp. 3077–3081.

[16] R. Bustin, R. Liu, H. V. Poor, and S. Shamai (Shitz), "An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel," *EURASIP J. Wireless Commun. and Netw.*, vol. 2009, no. 1, p. 370970, 2009.