On the Capacity Achieving Input of Amplitude Constrained Vector Gaussian Wiretap Channel

Antonino Favano*†, Luca Barletta*, Alex Dytso**
* Politecnico di Milano, Milano, 20133, Italy. Email: {antonino.favano, luca.barletta}@polimi.it
† Consiglio Nazionale delle Ricerche, Milano, 20133, Italy.
** New Jersey Institute of Technology, Newark, NJ 07102, USA. Email: alex.dytso@njit.edu

***Abstract*—This paper studies secrecy-capacity of an n-dimensional Gaussian wiretap channel under the peak-power constraint. This work determines the largest peak-power constraint $\bar{\mathsf{R}}_n$ such that an input distribution uniformly distributed on a single sphere is optimal; this regime is termed the low amplitude regime. The asymptotic of $\bar{\mathsf{R}}_n$ as n goes to infinity is completely characterized as a function of noise variance at both receivers. Moreover, the secrecy-capacity is also characterized in a form amenable for computation. Furthermore, several numerical examples are provided, such as the example of the secrecy-capacity achieving distribution outside of the low amplitude regime.**

I. Introduction

Consider the vector Gaussian wiretap channel with outputs

$$\mathbf{Y}_1 = \mathbf{X} + \mathbf{N}_1, \tag{1}$$

$$\mathbf{Y}_2 = \mathbf{X} + \mathbf{N}_2, \tag{2}$$

where $\mathbf{X} \in \mathbb{R}^n$ and where $\mathbf{N}_1 \sim \mathcal{N}(\mathbf{0}_n, \sigma_1^2\mathbf{I}_n)$ and $\mathbf{N}_2 \sim \mathcal{N}(\mathbf{0}_n, \sigma_2^2\mathbf{I}_n)$, and with $(\mathbf{X}, \mathbf{N}_1, \mathbf{N}_2)$ mutually independent. The output $\mathbf{Y}_1$ is observed by the legitimate receiver whereas the output $\mathbf{Y}_2$ is observed by the malicious receiver. We are interested in the scenario where the input $\mathbf{X}$ is limited by a peak-power constraint or amplitude constraint and assume that $\mathbf{X} \in \mathcal{B}_0(\mathsf{R}) = \{\mathbf{x} : \|\mathbf{x}\| \le \mathsf{R}\}$, i.e., $\mathcal{B}_0(\mathsf{R})$ is an n-ball centered at $\mathbf{0}$ of radius R. For this setting, the secrecy-capacity is

$$C_s(\sigma_1, \sigma_2, \mathsf{R}) = \max_{\mathbf{X}\in\mathcal{B}_0(\mathsf{R})} I(\mathbf{X};\mathbf{Y}_1) - I(\mathbf{X};\mathbf{Y}_2) \tag{3}$$

$$= \max_{\mathbf{X}\in\mathcal{B}_0(\mathsf{R})} I(\mathbf{X};\mathbf{Y}_1|\mathbf{Y}_2), \tag{4}$$

where the last step holds due to the degraded nature of the channel. It can be shown that $C_s(\sigma_1, \sigma_2, \mathsf{R}) = 0$ for $\sigma_1^2 \ge \sigma_2^2$. Therefore, in the remaining, we assume that $\sigma_1^2 < \sigma_2^2$.

We are interested in studying the input distribution $P_{\mathbf{X}^\star}$ that maximizes (4) in the low (but not vanishing) amplitude regime. Since closed-form expressions for secrecy-capacity are rare, we derive the secrecy-capacity in an integral form that is easy to evaluate. We also argue in Section II-C the solution to the secrecy-capacity can shed light on other problems unrelated to security.

A. Notation

The modified Bessel function of the first kind of order $v \ge 0$ will be denoted by $\mathsf{I}_v(x), x \in \mathbb{R}$. The following ratio of the Bessel functions will be commonly used in this work:

$$\mathsf{h}_v(x) = \frac{\mathsf{I}_v(x)}{\mathsf{I}_{v-1}(x)}, x \in \mathbb{R}, v \ge 0. \tag{5}$$

We denote the distribution of a random variable $\mathbf{X}$ by $P_{\mathbf{X}}$. The support set of $P_{\mathbf{X}}$ is denoted and defined as

$$\mathsf{supp}(P_{\mathbf{X}}) = \{\mathbf{x} : \text{ for every open set } \mathcal{D} \ni \mathbf{x} \text{ we have that } P_{\mathbf{X}}(\mathcal{D}) > 0\}. \tag{6}$$

The minimum mean squared error is denoted by

$$\mathrm{mmse}(\mathbf{X}|\mathbf{X}+\mathbf{N}) = \mathbb{E}\left[\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{X}+\mathbf{N}]\|^2\right]. \tag{7}$$

B. Literature Review

The wiretap channel was introduced by Wyner in [1], who also established the secrecy-capacity of the degraded wiretap channel. The wiretap channel plays a central role in network information theory; the interested reader is referred to [2]–[5] and reference therein for an in-detail treatment of the topic.

The secrecy-capacity of a scalar Gaussian wiretap channel with an average-power constraint was shown in [6] where the optimal input distribution was shown to be Gaussian. The secrecy-capacity of the MIMO wiretap channel was characterized in [7] and [8] where the Gaussian input was shown to be optimal. An elegant proof of optimality of Gaussian input, via the I-MMSE relationship [9], is given in [10].

The secrecy-capacity of the Gaussian wiretap channel under the peak-power constraint has received far less attention. The secrecy-capacity of the scalar Gaussian wiretap channel with an amplitude and power constraint was considered in [11] where the authors showed that the capacity-achieving input distribution $P_{X^\star}$ is discrete with finitely many support points. Recently, the result of [11] was sharpened in [12] by providing an explicit upper bound on the number of support points of $P_{X^\star}$ of the following from:

$$|\mathsf{supp}(P_{X^\star})| \le \rho\frac{\mathsf{R}^2}{\sigma_1^2} + O(\log(\mathsf{R})), \tag{8}$$

where $\rho = (2\mathrm{e}+1)^2\left(\frac{\sigma_2+\sigma_1}{\sigma_2-\sigma_1}\right)^2 + \left(\frac{\sigma_2+\sigma_1}{\sigma_2-\sigma_1}+1\right)^2$. The secrecy-capacity for the vector wiretap channel with a peak-power constraint was considered in [13] where it was shown that the optimal input distribution is concentrated on finitely many co-centric shells.

C. Contributions and Outline

The contributions and outlines of the paper are as follows. Section II discusses our assumptions and provides connections of the secrecy-capacity to other problems unrelated to security.

Section III presents our main results. Section IV is dedicated to numerical results and discusses structure of the optimal input beyond low amplitude regime. Section III and Section VI are dedicated to proofs. Section VII concludes paper. Due to space limitations, some of the proofs are omitted and can be found in the extended version of the paper [14].

II. Assumptions and Motivations

A. Assumptions

Consider the following function: for $y \in \mathbb{R}^+$

$$G_{\sigma_1,\sigma_2,\mathsf{R},n}(y) = \frac{\mathbb{E}\left[\frac{\mathsf{R}}{\|y+\mathbf{W}\|}\mathsf{h}_{\frac{n}{2}}\left(\frac{\mathsf{R}}{\sigma_2^2}\|y+\mathbf{W}\|\right)-1\right]}{\sigma_2^2} - \frac{\frac{\mathsf{R}}{y}\mathsf{h}_{\frac{n}{2}}\left(\frac{\mathsf{R}}{\sigma_1^2}y\right)-1}{\sigma_1^2}, \tag{9}$$

where $\mathbf{W} \sim \mathcal{N}(\mathbf{0}_{n+2}, (\sigma_2^2-\sigma_1^2)\mathbf{I}_{n+2})$. Notice that the function $G_{\sigma_1,\sigma_2,\mathsf{R},n}(y)$ is related to the derivative of the secrecy-density.

In this work, in order to make progress on the secrecy-capacity, we make the following *conjecture* about the ratio of the Bessel functions: for all $\mathsf{R} \geq 0, \sigma_2 > \sigma_1 \geq 0$ and $n \in \mathbb{N}$, the function $y \mapsto G_{\sigma_1,\sigma_2,\mathsf{R},n}(y)$ has *at most* one sign change.

In general, proving that $G_{\sigma_1,\sigma_2,\mathsf{R},n}(y)$ has at most one sign change is not easy. However, extensive numerical evaluations show that this property holds for any $n, \mathsf{R}, \sigma_1, \sigma_2$.

Therefore, the problem boils down to showing that there is at most one sign change for $y > 0$. Using this, we can give a sufficient condition for this conjecture to be true. Note that

$$G_{\sigma_1,\sigma_2,\mathsf{R},n}(y) \geq -\frac{1}{\sigma_2^2} + \frac{1}{\sigma_1^2} - \frac{\mathsf{R}}{\sigma_1^2 y}\mathsf{h}_{\frac{n}{2}}\left(\frac{\mathsf{R}}{\sigma_1^2}y\right) \tag{10}$$

$$\geq -\frac{1}{\sigma_2^2} + \frac{1}{\sigma_1^2} - \frac{\mathsf{R}^2}{\sigma_1^4 n}, \tag{11}$$

which is nonnegative, hence has no sign change for $y > 0$, if

$$\mathsf{R} < \sigma_1^2\sqrt{n\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}\right)}. \tag{12}$$

The inequality in (10) follows by $\mathsf{h}_{\frac{n}{2}}(x) \geq 0$ for $x \geq 0$; and (11) follows by $\mathsf{h}_{\frac{n}{2}}(x) \leq \frac{x}{n}$ for $x \geq 0$ and $n \in \mathbb{N}$.

B. Low Amplitude Regime

In this work a low amplitude regime is defined as follows.

Definition 1. *Let $\mathbf{X}_\mathsf{R} \sim P_{\mathbf{X}_\mathsf{R}}$ be uniform on $\mathcal{C}(\mathsf{R}) = \{\mathbf{x} : \|\mathbf{x}\| = \mathsf{R}\}$. The capacity in* (4) *is said to be in the low amplitude regime if $\mathsf{R} \leq \bar{\mathsf{R}}_n(\sigma_1^2, \sigma_2^2)$ where*

$$\bar{\mathsf{R}}_n(\sigma_1^2,\sigma_2^2) = \max\left\{\mathsf{R} : P_{\mathbf{X}_\mathsf{R}} = \arg\max_{\mathbf{X}\in\mathcal{B}_0(\mathsf{R})} I(\mathbf{X};\mathbf{Y}_1|\mathbf{Y}_2)\right\}. \tag{13}$$

If the set in (13) *is empty, we set $\bar{\mathsf{R}}_n(\sigma_1^2,\sigma_2^2) = 0$.*

The quantity $\bar{\mathsf{R}}_n(\sigma_1^2,\sigma_2^2)$ represents the largest radius R for which $P_{\mathbf{X}_\mathsf{R}}$ is secrecy-capacity-achieving.

One of the main goals of this work is to find $\bar{\mathsf{R}}_n(\sigma_1^2,\sigma_2^2)$.

C. Connections to Other Optimization Problems

The distribution $P_{\mathbf{X}_\mathsf{R}}$ occurs in a variety of statistical and information-theoretic applications. For example, consider the following two optimization problems:

$$\max_{\mathbf{X}\in\mathcal{B}_0(\mathsf{R})} I(\mathbf{X};\mathbf{X}+\mathbf{N}), \tag{14}$$

$$\max_{\mathbf{X}\in\mathcal{B}_0(\mathsf{R})} \mathrm{mmse}(\mathbf{X}|\mathbf{X}+\mathbf{N}), \tag{15}$$

where $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$. The first problem seeks to characterize the capacity of the point-to-point channel under the amplitude constraint, and the second problem seeks to find the largest minimum mean squared error under the assumption that the signal has bounded amplitude; the interested reader is referred to [15]–[17] for a detailed background on both problems.

Similarly to the wiretap channel, we can define the low amplitude regime for both problems as the largest R such that $P_{\mathbf{X}_\mathsf{R}}$ is optimal and denote these by $\bar{\mathsf{R}}_n^{\mathrm{ptp}}(\sigma^2)$ and $\bar{\mathsf{R}}_n^{\mathrm{MMSE}}(\sigma^2)$. We now argue that both $\bar{\mathsf{R}}_n^{\mathrm{ptp}}(\sigma^2)$ and $\bar{\mathsf{R}}_n^{\mathrm{MMSE}}(\sigma^2)$ can be seen as a special case of the wiretap solution. Hence, the wiretap channel provides and interesting unification and generalization of these two problems.

First, note that the point-to-point solution can be recovered from the wiretap by simply specializing the wiretap channel to the point-to-point channel, that is

$$\bar{\mathsf{R}}_n^{\mathrm{ptp}}(\sigma^2) = \lim_{\sigma_2\to\infty} \bar{\mathsf{R}}_n(\sigma^2,\sigma_2^2). \tag{16}$$

Second, to see that the MMSE solution can be recovered from the wiretap recall that by the I-MMSE relationship [9], we have that

$$\max_{\mathbf{X}\in\mathcal{B}_0(\mathsf{R})} I(\mathbf{X};\mathbf{Y}_1) - I(\mathbf{X};\mathbf{Y}_2) = \max_{\mathbf{X}\in\mathcal{B}_0(\mathsf{R})} \frac{1}{2}\int_{\sigma_1^2}^{\sigma_2^2} \frac{\mathrm{mmse}(\mathbf{X}|\mathbf{X}+\sqrt{s}\mathbf{Z})}{s^2}\mathrm{d}s \tag{17}$$

where $\mathbf{Z}$ is standard Gaussian. Now note that if we choose $\sigma_2^2 = \sigma_1^2 + \epsilon$ for some small enough $\epsilon > 0$, we arrive at

$$\max_{\mathbf{X}\in\mathcal{B}_0(\mathsf{R})} I(\mathbf{X};\mathbf{Y}_1) - I(\mathbf{X};\mathbf{Y}_2) \tag{18}$$

$$= \max_{\mathbf{X}\in\mathcal{B}_0(\mathsf{R})} \frac{\epsilon}{2}\frac{\mathrm{mmse}(\mathbf{X}|\mathbf{X}+\sqrt{\sigma_1^2}\mathbf{Z})}{\sigma_1^4}. \tag{19}$$

Consequently, for a small enough $\epsilon > 0$,

$$\bar{\mathsf{R}}_n^{\mathrm{MMSE}}(\sigma^2) = \bar{\mathsf{R}}_n(\sigma^2,\sigma^2+\epsilon). \tag{20}$$

III. Main Results

A. Characterizing the Low Amplitude Regime

Our first result characterizes the low amplitude regime.

Theorem 1. *Consider a function*

$$f(\mathsf{R}) = \int_{\sigma_1^2}^{\sigma_2^2} \frac{\mathbb{E}\left[\mathsf{h}_{\frac{n}{2}}^2\left(\frac{\|\sqrt{s}\mathbf{Z}\|\mathsf{R}}{s}\right) + \mathsf{h}_{\frac{n}{2}}^2\left(\frac{\|\mathsf{R}+\sqrt{s}\mathbf{Z}\|\mathsf{R}}{s}\right)\right]-1}{s^2}\mathrm{d}s \tag{21}$$

TABLE I: Values of $\bar{\mathsf{R}}_n^{\text{ptp}}(1)$, $\bar{\mathsf{R}}_n(1,\sigma_2^2)$, and $\bar{\mathsf{R}}_n^{\text{MMSE}}(1)$.

n	1	2	4	8	16	32
$\bar{\mathsf{R}}_n^{\text{ptp}}(1)$	1.666	2.454	3.580	5.158	7.367	10.472
$\bar{\mathsf{R}}_n(1,1000)$	1.664	2.450	3.575	5.151	7.357	10.458
$\bar{\mathsf{R}}_n(1,10)$	1.518	2.221	3.229	4.646	6.632	9.424
$\bar{\mathsf{R}}_n(1,1.5)$	1.161	1.687	2.444	3.513	5.013	7.124
$\bar{\mathsf{R}}_n(1,1.001)$	1.057	1.535	2.224	3.196	4.561	6.481
$\bar{\mathsf{R}}_n^{\text{MMSE}}(1)$	1.057	1.535	2.223	3.195	4.560	6.479

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$. *The input* $\mathbf{X}_{\mathsf{R}}$ *is secrecy-capacity achieving if and only if* $\mathsf{R} \leq \bar{\mathsf{R}}_n(\sigma_1^2,\sigma_2^2)$ *where* $\bar{\mathsf{R}}_n(\sigma_1^2,\sigma_2^2)$ *is given as the zero of*

$$f(\mathsf{R}) = 0. \tag{22}$$

Proof. See Section V. □

Remark 1. Note that (22) always has a solution. To see this observe that $f(0) = \frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} < 0$, and $f(\infty) = \frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} > 0$. Moreover, the solution is unique, because $f(\mathsf{R})$ is monotonically increasing for $\mathsf{R} \geq 0$.

The solution to (22) needs to be found numerically.[1] Since evaluating $f(\mathsf{R})$ is rather straightforward and not time-consuming, we opted for a binary search algorithm. In Table I, we show the values of $\bar{\mathsf{R}}_n(1,\sigma_2^2)$ for some values of σ_2^2 and n. Moreover, we report the values of $\bar{\mathsf{R}}_n^{\text{ptp}}(1)$ and $\bar{\mathsf{R}}_n^{\text{MMSE}}(1)$ from [15] in the first and the last row, respectively. As predicted by (16), we can appreciate the close match of the $\bar{\mathsf{R}}_n^{\text{ptp}}(1)$ row with the one of $\bar{\mathsf{R}}_n(1,1000)$. Similarly, the agreement between the $\bar{\mathsf{R}}_n^{\text{MMSE}}(1)$ row and the $\bar{\mathsf{R}}_n(1,1.001)$ row is justified by (20).

B. Large n Asymptotics

We now use the result in Theorem 1 to characterize the asymptotic behavior of $\bar{\mathsf{R}}_n(\sigma_1^2,\sigma_2^2)$.

Theorem 2.

$$\lim_{n\to\infty} \frac{\bar{\mathsf{R}}_n(\sigma_1^2,\sigma_2^2)}{\sqrt{n}} = c(\sigma_1^2,\sigma_2^2), \tag{23}$$

where $c(\sigma_1^2,\sigma_2^2)$ *is the solution of*

$$\int_{\sigma_1^2}^{\sigma_2^2} \frac{\frac{c^2}{\left(\frac{\sqrt{s}}{2}+\sqrt{\frac{s}{4}+c^2}\right)^2} + \frac{c^2(c^2+s)}{\left(\frac{s}{2}+\sqrt{\frac{s^2}{4}+c^2(c^2+s)}\right)^2} - 1}{s^2}\mathrm{d}s = 0. \tag{24}$$

In Fig. 1, for $\sigma_1^2 = 1$ and $\sigma_2^2 = 1.001, 1.5, 10$, we show the behavior of $\bar{\mathsf{R}}_n(1,\sigma_2^2)/\sqrt{n}$ and how its converges to $c(1,\sigma_2^2)$.

[1]To avoid any loss of accuracy in the numerical evaluation of $\mathsf{h}_v(x)$ for large values of x, we used the exponential scaling provided in the MATLAB implementation of $\mathsf{l}_v(x)$.

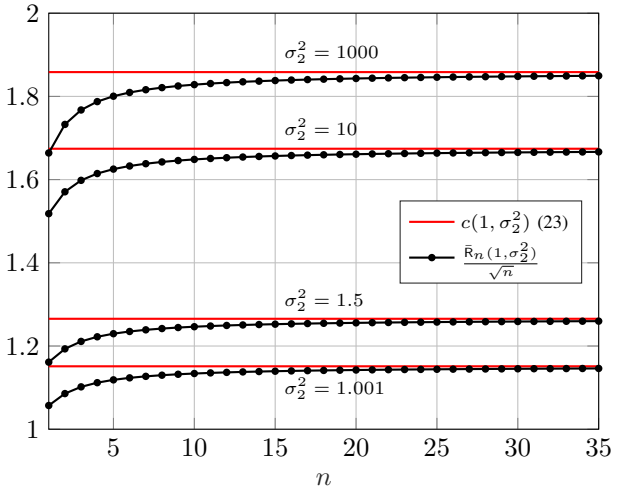

Fig. 1: Asymptotic behavior of $\bar{\mathsf{R}}_n(1,\sigma_2^2)/\sqrt{n}$ versus n for $\sigma_1^2 = 1$ and $\sigma_2^2 = 1.001, 1.5, 10, 1000$.

C. Capacity Expression in the Low Amplitude Regime

The result in Theorem 1 can also be used to establish the secrecy-capacity for all $\mathsf{R} \leq \bar{\mathsf{R}}_n(\sigma_1^2,\sigma_2^2)$ as is done next.

Theorem 3. *If* $\mathsf{R} \leq \bar{\mathsf{R}}_n(\sigma_1^2,\sigma_2^2)$, *then*

$$C_s(\sigma_1^2,\sigma_2^2,\mathsf{R}) = \frac{1}{2}\int_{\sigma_1^2}^{\sigma_2^2} \frac{\mathsf{R}^2 - \mathsf{R}^2\mathbb{E}\left[\mathsf{h}_{\frac{n}{2}}^2\left(\frac{\|\mathsf{R}+\sqrt{s}\mathbf{Z}\|\mathsf{R}}{s}\right)\right]}{s^2}\mathrm{d}s. \tag{25}$$

Proof: See Section VI. ■

IV. Beyond the Low Amplitude Regime

To evaluate the secrecy-capacity and find the optimal distribution $P_{\mathbf{X}^\star}$ beyond $\bar{\mathsf{R}}_n$ we rely on numerical estimations. We remark that, as pointed out in [13], the capacity-achieving distribution is isotropic and consists of finitely many co-centric shells. Keeping this in mind, we can find the optimal input distribution $P_{\mathbf{X}^\star}$ by just optimizing over $P_{\|\mathbf{X}\|}$ with $\|\mathbf{X}\| \leq \mathsf{R}$.

Let us denote by $\hat{C}_s(\sigma_1^2,\sigma_2^2,\mathsf{R})$ the numerical estimate of the secrecy-capacity and by $\hat{P}_{\|\mathbf{X}^\star\|}$ the optimal pmf of the input norm. To numerically evaluate $\hat{C}_s(\sigma_1^2,\sigma_2^2,\mathsf{R})$ and $\hat{P}_{\|\mathbf{X}^\star\|}$ we adapt the algorithmic procedure described in [18] by re-evaluating the Blahut-Arimoto recursion and the gradient of the secrecy-information.

In Fig. 2, we show with black circles the numerical estimate $\hat{C}_s(\sigma_1^2,\sigma_2^2,\mathsf{R})$ for $\sigma_1^2 = 1$, $\sigma_2^2 = 1.5, 10$, and $n = 2, 4$. For the same values of σ_1^2, σ_2^2, and n we also show, with the red lines, the analytical low amplitude regime capacity $C_s(\sigma_1^2,\sigma_2^2,\mathsf{R})$ from Theorem 3. Also, we show with blue dotted lines the secrecy-capacity under the average-power constraint $\mathbb{E}\left[\|\mathbf{X}\|^2\right] \leq \mathsf{R}^2$:

$$C_G(\sigma_1^2,\sigma_2^2,\mathsf{R}) = \frac{n}{2}\log\frac{1+\mathsf{R}^2/\sigma_1^2}{1+\mathsf{R}^2/\sigma_2^2} \geq C_s(\sigma_1^2,\sigma_2^2,\mathsf{R}), \tag{26}$$

where the inequality follows by noting that the average-power constraint $\mathbb{E}\left[\|\mathbf{X}\|^2\right] \leq \mathsf{R}^2$ is weaker than the amplitude

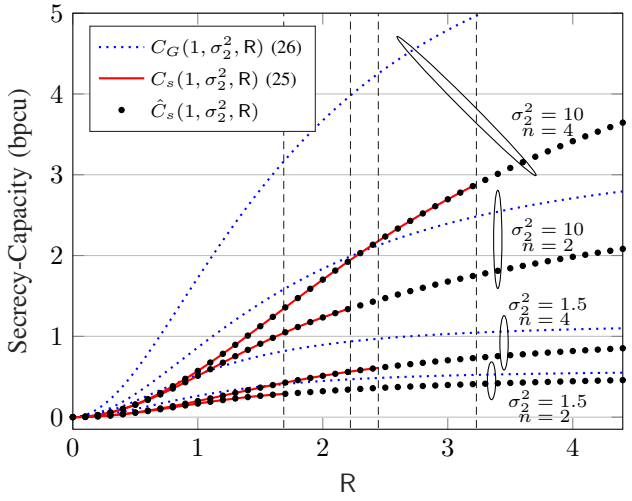

Fig. 2: Secrecy-capacity in bit per channel use (bpcu) versus R, for $\sigma_2^2 = 1.5, 10$ and $n = 2, 4$.

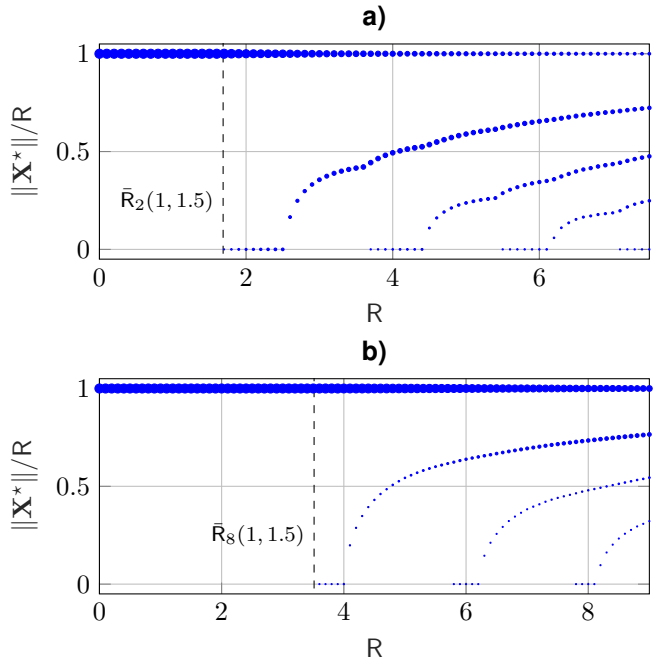

Fig. 3: Evolution of the numerically estimated $\hat{P}_{\|\mathbf{X}^\star\|}$ versus R for $\sigma_1^2 = 1$, $\sigma_2^2 = 1.5$, **a)** $n = 2$, and **b)** $n = 8$.

constraint $\|\mathbf{X}\| \le \mathsf{R}$. Finally, the dashed vertical lines show $\bar{\mathsf{R}}_n$ for the considered values of σ_1^2, σ_2^2, and n.

In Fig. 3, we show the evolution of the numerically estimated pmf $\hat{P}_{\|\mathbf{X}^\star\|}$ vs. R, for $\sigma_1^2 = 1$, $\sigma_2^2 = 1.5$, and $n = 2, 8$. The figure shows, at each R, the normalized amplitude mass points in the estimated pmf, while the size of the circles qualitatively shows the associated probability.

V. Proof of Theorem 1

A. KKT Conditions

Lemma 1. *$P_{\mathbf{X}^\star}$ maximizes* (4) *if and only if*

$$\Xi(\mathbf{x}; P_{\mathbf{X}^\star}) = C_s(\sigma_1^2, \sigma_2^2, \mathsf{R}),\ \mathbf{x} \in \mathsf{supp}(P_{\mathbf{X}^\star}), \tag{27}$$

$$\Xi(\mathbf{x}; P_{\mathbf{X}^\star}) \le C_s(\sigma_1^2, \sigma_2^2, \mathsf{R}),\ \mathbf{x} \in \mathcal{B}_0(\mathsf{R}), \tag{28}$$

where for $\mathbf{x} \in \mathbb{R}^n$

$$\Xi(\mathbf{x}; P_{\mathbf{X}^\star}) = \mathsf{D}(f_{\mathbf{Y}_1|\mathbf{X}}(\cdot|\mathbf{x})\|f_{\mathbf{Y}_1^\star}) - \mathsf{D}(f_{\mathbf{Y}_2|\mathbf{X}}(\cdot|\mathbf{x})\|f_{\mathbf{Y}_2^\star}) \tag{29}$$

$$= \mathbb{E}\left[g(\mathbf{Y}_1)|\mathbf{X} = \mathbf{x}\right], \tag{30}$$

and where

$$g(\mathbf{y}) = \mathbb{E}\left[\log \frac{f_{\mathbf{Y}_2^\star}(\mathbf{y}+\mathbf{N})}{f_{\mathbf{Y}_1^\star}(\mathbf{y})}\right] + n\log\left(\frac{\sigma_2}{\sigma_1}\right),\ \mathbf{y} \in \mathbb{R}^n, \tag{31}$$

with $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, (\sigma_2^2 - \sigma_1^2)\mathbf{I}_n)$.

Proof. This is a vector extension of [12, Lemma 1]. □

B. A New Necessary and Sufficient Condition

Theorem 4. *$P_{\mathbf{X}_\mathsf{R}}$ is optimal if and only if for all* $\|\mathbf{x}\| = \mathsf{R}$

$$\Xi(\mathbf{0}; P_{\mathbf{X}_\mathsf{R}}) \le \Xi(\mathbf{x}; P_{\mathbf{X}_\mathsf{R}}). \tag{32}$$

Moreover, if $\mathsf{R} < \sigma_1^2\sqrt{n\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)}$, *then $P_{\mathbf{X}_\mathsf{R}}$ is optimal.*

Proof. The secrecy-density $\Xi(\cdot; P_{\mathbf{X}_\mathsf{R}})$ is a function only of $\|\mathbf{x}\|$, thanks to the rotational symmetry of the Gaussian distribution and of $P_{\mathbf{X}_\mathsf{R}}$. In view of this, a way to prove condition (32) is to show that the maximum of $\|\mathbf{x}\| \mapsto \Xi(\|\mathbf{x}\|; P_{\mathbf{X}_\mathsf{R}})$ occurs at either $\|\mathbf{x}\| = 0$ or $\|\mathbf{x}\| = \mathsf{R}$. Next, we show that the derivative of $\Xi(\|\mathbf{x}\|; P_{\mathbf{X}_\mathsf{R}})$ makes at most one sign change, from negative to positive. This fact will prove the claim.

From Lemma 3 in [14], the derivative of Ξ is

$$\Xi'(\|\mathbf{x}\|; P_{\mathbf{X}_\mathsf{R}}) = \|\mathbf{x}\|\,\mathbb{E}\left[\widetilde{M}_2(\sigma_1 Q_{n+2}) - M_1(\sigma_1 Q_{n+2})\right] \tag{33}$$

where Q_{n+2}^2 is a noncentral chi-square random variable with $n+2$ degrees of freedom and noncentrality parameter $\frac{\|\mathbf{x}\|^2}{\sigma_1^2}$ and

$$M_i(y) = \frac{1}{\sigma_i^2}\left(\frac{\mathsf{R}}{y}\mathsf{h}_{\frac{n}{2}}\left(\frac{\mathsf{R}}{\sigma_i^2}y\right) - 1\right), \qquad i \in \{1,2\} \tag{34}$$

$$\widetilde{M}_2(y) = \mathbb{E}\left[M_2(\|y + \mathbf{W}\|)\right], \tag{35}$$

where $\mathbf{W} \sim \mathcal{N}(\mathbf{0}_{n+2}, (\sigma_2^2 - \sigma_1^2)\mathbf{I}_{n+2})$.

Note that $\Xi'(0; P_{\mathbf{X}_\mathsf{R}}) = 0$, and that $\Xi'(\|\mathbf{x}\|; P_{\mathbf{X}_\mathsf{R}}) > 0$ for sufficiently large $\|\mathbf{x}\|$; in fact, we have

$$\Xi'(\|\mathbf{x}\|; P_{\mathbf{X}_\mathsf{R}}) > \|\mathbf{x}\|\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) - \frac{\|\mathbf{x}\|}{\sigma_1^2}\mathbb{E}\left[\frac{\mathsf{R}}{\sigma_1 Q_{n+2}}\right] \tag{36}$$

$$= \|\mathbf{x}\|\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) - \frac{\|\mathbf{x}\|}{\sigma_1^2}\mathbb{E}\left[\frac{\mathsf{R}}{\|\mathbf{x}\|}\mathsf{h}_{\frac{n}{2}}\left(\frac{\|\mathbf{x}\|}{\sigma_1}Q_n\right)\right] \tag{37}$$

$$\ge \|\mathbf{x}\|\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) - \frac{\mathsf{R}}{\sigma_1^2}, \tag{38}$$

where (36) is by $0 \le \mathsf{h}_{\frac{n}{2}}(x) \le 1$, $x \ge 0$; (37) is by a change of measure in the expectation; and (38) is by $\mathsf{h}_{\frac{n}{2}}(x) \le 1$.

To conclude, we need to prove that $\Xi'(\|\mathbf{x}\|; P_{\mathbf{X}_\mathsf{R}})$ changes sign at most once. To that end, we will need the following lemma shown in [19, Theorem 3].

Lemma 2. *Let the pdf $f(x,\omega)$ be a positive-definite kernel that can be differentiated n times with respect to x for all ω, and let $\eta(\omega)$ be a function that changes sign n times. If*

$$M(x) = \int \eta(\omega) f(x,\omega) \mathrm{d}\omega, \tag{39}$$

can be differentiated n times, then $M(x)$ changes sign at most n times.

By using (33), the fact that the chi-square pdf is a positive defined kernel [19], and Lemma 2, the number of sign changes of $\Xi'(\|\mathbf{x}\|; P_{\mathbf{X}_\mathsf{R}})$ is upper-bounded by the number of sign changes of

$$\widetilde{M}_2(y) - M_1(y) = G_{\sigma_1,\sigma_2,\mathsf{R},n}(y),\ y > 0 \tag{40}$$

where $G_{\sigma_1,\sigma_2,\mathsf{R},n}(y)$ was defined and discussed in Section II and it was assumed that it has at most one sign change for $y > 0$. For example, a sufficient condition is given by

$$\mathsf{R} < \sigma_1^2 \sqrt{n\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)} \tag{41}$$

This concludes the proof. □

C. Estimation Theoretic Representation

To complete the proof we seek to re-write the condition in Theorem 4 in the estimation theoretic form. To that end, we need the following representation of the relative entropy [20]:

$$D(P_{\mathbf{X}_1+\sqrt{t}\mathbf{Z}} \| P_{\mathbf{X}_2+\sqrt{t}\mathbf{Z}}) = \frac{1}{2}\int_t^\infty \frac{g(s)}{s^2}\mathrm{d}s, \tag{42}$$

where

$$\begin{aligned} g(s) = &\mathbb{E}\left[\|\mathbf{X}_1 - \phi_2(\mathbf{X}_1 + \sqrt{s}\mathbf{Z})\|^2\right] \\ &- \mathbb{E}\left[\|\mathbf{X}_1 - \phi_1(\mathbf{X}_1 + \sqrt{s}\mathbf{Z})\|^2\right] \end{aligned} \tag{43}$$

and where $\phi_i(\mathbf{y}) = \mathbb{E}[\mathbf{X}_i | \mathbf{X}_i + \sqrt{s}\mathbf{Z} = \mathbf{y}]$, $i \in \{1,2\}$.

Another fact that will be important for our expression is

$$\mathbb{E}\left[\mathbf{X}_\mathsf{R} \mid \mathbf{X}_\mathsf{R} + \sqrt{s}\mathbf{Z} = \mathbf{y}\right] = \frac{\mathsf{R}\mathbf{y}}{\|\mathbf{y}\|}\mathsf{h}_{\frac{n}{2}}\left(\frac{\|\mathbf{y}\|\mathsf{R}}{s}\right), \tag{44}$$

see, for example, [15] for the proof.

Next, using (42) and (44) note that for any $\|\mathbf{x}\| = \mathsf{R}$ we have that for $i \in \{1,2\}$

$$D(P_{\mathbf{x}+\sqrt{\sigma_i^2}\mathbf{Z}} \| P_{\mathbf{X}_\mathsf{R}+\sqrt{\sigma_i^2}\mathbf{Z}}) \tag{45}$$

$$= \frac{1}{2}\int_{\sigma_i^2}^\infty \frac{\mathbb{E}\left[\|\mathbf{x} - \frac{\mathsf{R}(\mathbf{x}+\sqrt{s}\mathbf{Z})}{\|\mathbf{x}+\sqrt{s}\mathbf{Z}\|}\mathsf{h}_{\frac{n}{2}}\left(\frac{\|\mathbf{x}+\sqrt{s}\mathbf{Z}\|\mathsf{R}}{s}\right)\|^2\right]}{s^2}\mathrm{d}s \tag{46}$$

$$= \frac{1}{2}\int_{\sigma_i^2}^\infty \frac{\mathsf{R}^2 - \mathsf{R}^2\mathbb{E}\left[\mathsf{h}_{\frac{n}{2}}^2\left(\frac{\|\mathbf{x}+\sqrt{s}\mathbf{Z}\|\mathsf{R}}{s}\right)\right]}{s^2}\mathrm{d}s, \tag{47}$$

and

$$D(P_{\mathbf{0}+\sqrt{\sigma_i^2}\mathbf{Z}} \| P_{\mathbf{X}_\mathsf{R}+\sqrt{\sigma_i^2}\mathbf{Z}}) = \frac{1}{2}\int_{\sigma_i^2}^\infty \frac{\mathsf{R}^2\mathbb{E}\left[\mathsf{h}_{\frac{n}{2}}^2\left(\frac{\mathsf{R}\|\mathbf{Z}\|}{s}\right)\right]}{s^2}\mathrm{d}s. \tag{48}$$

Now, note that by using definition of $\Xi(\mathbf{x}; P_{\mathbf{X}_\mathsf{R}})$ in (30), and (47) and (48) we have that for $\|\mathbf{x}\| = \mathsf{R}$

$$\begin{aligned} &\Xi(\mathbf{x}; P_{\mathbf{X}_\mathsf{R}}) \\ &= D(P_{\mathbf{x}+\sqrt{\sigma_1^2}\mathbf{Z}} \| P_{\mathbf{X}_\mathsf{R}+\sqrt{\sigma_1^2}\mathbf{Z}}) - D(P_{\mathbf{x}+\sqrt{\sigma_2^2}\mathbf{Z}} \| P_{\mathbf{X}_\mathsf{R}+\sqrt{\sigma_2^2}\mathbf{Z}}) \\ &= \frac{1}{2}\int_{\sigma_1^2}^{\sigma_2^2} \frac{\mathsf{R}^2 - \mathsf{R}^2\mathbb{E}\left[\mathsf{h}_{\frac{n}{2}}^2\left(\frac{\|\mathbf{x}+\sqrt{s}\mathbf{Z}\|\mathsf{R}}{s}\right)\right]}{s^2}\mathrm{d}s, \end{aligned} \tag{49}$$

and

$$\begin{aligned} &\Xi(\mathbf{0}; P_{\mathbf{X}_\mathsf{R}}) \\ &= D(P_{\mathbf{0}+\sqrt{\sigma_1^2}\mathbf{Z}} \| P_{\mathbf{X}_\mathsf{R}+\sqrt{\sigma_1^2}\mathbf{Z}}) - D(P_{\mathbf{0}+\sqrt{\sigma_2^2}\mathbf{Z}} \| P_{\mathbf{X}_\mathsf{R}+\sqrt{\sigma_2^2}\mathbf{Z}}) \\ &= \frac{1}{2}\int_{\sigma_1^2}^{\sigma_2^2} \frac{\mathsf{R}^2\mathbb{E}\left[\mathsf{h}_{\frac{n}{2}}^2\left(\frac{\|\sqrt{s}\mathbf{Z}\|\mathsf{R}}{s}\right)\right]}{s^2}\mathrm{d}s \end{aligned} \tag{50}$$

Consequently, the necessary and sufficient condition in Theorem 4 can be equivalently written as

$$\int_{\sigma_1^2}^{\sigma_2^2} \frac{\mathbb{E}\left[\mathsf{h}_{\frac{n}{2}}^2\left(\frac{\|\sqrt{s}\mathbf{Z}\|\mathsf{R}}{s}\right) + \mathsf{h}_{\frac{n}{2}}^2\left(\frac{\|\mathbf{x}+\sqrt{s}\mathbf{Z}\|\mathsf{R}}{s}\right)\right] - 1}{s^2}\mathrm{d}s \le 0. \tag{51}$$

Now $\bar{\mathsf{R}}_n(\sigma_1^2, \sigma_2^2)$ will be the largest R that satisfies (51), which concludes the proof of Theorem 1.

VI. Proof of Theorem 3

Using the KKT conditions in (27), we have that for $\mathbf{x} = [\mathsf{R}, 0, \ldots, 0]$

$$\begin{aligned} C_s(\sigma_1^2, \sigma_2^2, \mathsf{R}) &= \mathsf{D}(f_{\mathbf{Y}_1|\mathbf{X}}(\cdot|\mathbf{x}) \| f_{\mathbf{Y}_1^\star}) - \mathsf{D}(f_{\mathbf{Y}_2|\mathbf{X}}(\cdot|\mathbf{x}) \| f_{\mathbf{Y}_2^\star}) \\ &= \frac{1}{2}\int_{\sigma_1^2}^{\sigma_2^2} \frac{\mathsf{R}^2 - \mathsf{R}^2\mathbb{E}\left[\mathsf{h}_{\frac{n}{2}}^2\left(\frac{\|\mathsf{R}+\sqrt{s}\mathbf{Z}\|\mathsf{R}}{s}\right)\right]}{s^2}\mathrm{d}s \end{aligned} \tag{52}$$

where the last expression was computed in (49).

VII. Conclusion

This paper focuses on the secrecy-capacity vector Gaussian wiretap channel under the peak-power (or amplitude constraint) in a so-called low (but not vanishing) amplitude regime. In this regime, the optimal input distribution $P_{\mathbf{X}_\mathsf{R}}$ is supported on a single sphere of radius R. The paper has identified the largest $\bar{\mathsf{R}}_n$ such that this distribution $P_{\mathbf{X}_\mathsf{R}}$ is optimal. In addition, the asymptotic of $\bar{\mathsf{R}}_n$ has been completely characterized as dimension n approaches infinity. As a by-product of the analysis, the capacity in the low amplitude regime has also been characterized in more or less closed-form. The paper has also provided a number of supporting numerical examples. As part of ongoing work, we are trying to resolve the conjecture that was made regarding the number of zeros of the function defined through the ratios of Bessel functions. An interesting but ambitious future direction would be to determine a regime in which a mixture of a mass point at zero and $P_{\mathbf{X}_\mathsf{R}}$ is optimal. We finally remark that the extension of the results of this paper to nondegraded wiretap channels is not trivial.

REFERENCES

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

[2] M. Bloch and J. Barros, *Physical-Layer Security:From Information Theory to Security Engineering*. Cambridge University Press, 2011.

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

[4] 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.

[5] 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.

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

[7] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas–Part II: The MIMOME wiretap channel," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5515–5532, 2010.

[8] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4961–4972, 2011.

[9] 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.

[10] R. Bustin, R. Liu, H. V. Poor, and S. Shamai, "An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, pp. 1–8, 2009.

[11] 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.

[12] L. Barletta and A. Dytso, "Scalar Gaussian wiretap channel: Bounds on the support size of the secrecy-capacity-achieving distribution," in *2021 IEEE Information Theory Workshop (ITW)*, 2021, pp. 1–6.

[13] A. Dytso, M. Egan, S. M. Perlaza, H. V. Poor, and S. S. Shitz, "Optimal inputs for some classes of degraded wiretap channels," in *2018 IEEE Information Theory Workshop (ITW)*, 2018, pp. 1–5.

[14] A. Favano, L. Barletta, and A. Dytso, "On the capacity achieving input of amplitude constrained vector Gaussian wiretap channel," *arXiv preprint arXiv:2202.00586*, 2022.

[15] A. Dytso, M. Al, H. V. Poor, and S. Shamai Shitz, "On the capacity of the peak power constrained vector gaussian channel: An estimation theoretic perspective," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3907–3921, 2019.

[16] A. Favano, M. Ferrari, M. Magarini, and L. Barletta, "The capacity of the amplitude-constrained vector Gaussian channel," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 426–431.

[17] J. C. Berry, "Minimax estimation of a bounded normal mean vector," *Journal of Multivariate Analysis*, vol. 35, no. 1, pp. 130–139, 1990.

[18] L. Barletta and A. Dytso, "Scalar Gaussian wiretap channel with peak amplitude constraint: Numerical computation of the optimal input distribution," *arXiv preprint arXiv:2111.11442*, 2021.

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

[20] S. Verdú, "Mismatched estimation and relative entropy," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3712–3720, 2010.