Spread Spectrum Watermarking Security

Luis Pérez-Freire[†] and Fernando Pérez-González, *Member, IEEE*

Abstract

This paper presents both theoretical and practical analyses of the security offered by watermarking and data hiding methods based on spread spectrum. In this context, security is understood as the difficulty of estimating the secret parameters of the embedding function based on the observation of watermarked signals. On the theoretical side, the security is quantified from an information-theoretic point of view by means of the equivocation about the secret parameters. The main results reveal fundamental limits and bounds on security and provide insight into other properties, such as the impact of the embedding parameters, and the tradeoff between robustness and security. On the practical side, workable estimators of the secret parameters are proposed and theoretically analyzed for a variety of scenarios, providing a comparison with previous approaches, and showing that the security of many schemes used in practice can be fairly low.

Index Terms

Spread spectrum, watermarking security, information leakage, equivocation, parameter estimation, PCA, ICA, Constant Modulus

EDICS Category: WAT-SSPM, WAT-THEO, WAT-OTHA

I. Introduction

The notion of security in watermarking has started to being considered several years ago [1], [2], although the first attempt at providing a mathematical framework for assessing watermarking security was [3], which gave rise to other related works as [4] and [5]. The present paper considers the security of spread spectrum methods for watermarking and data hiding following the same guidelines as in the aforementioned works. The fundamentals of this approach to security assessment can be found in [4], where the attacks to the security of watermarking and data hiding methods are defined as those aimed at gaining knowledge about the secret parameters of the system. The key assumptions are that the watermarker owns a secret key that he/she repeatedly uses to watermark contents, and an attacker is able to gather several signals (observations) that were watermarked with the same secret key; if the attacker manages

The authors are with the Signal Theory and Communications Dept., ETSI Telecommunications, University of Vigo, 36310 Vigo, Spain (e-mail: {lpfreire,fperez}@gts.tsc.uvigo.es)

[†] Luis Pérez-Freire holds a temporary postdoctoral research position in the Security Labs of Thomson R&D, France.

This work was supported in part by *Xunta de Galicia* under Projects 07TIC012322PR (FACTICA), 2007/149 (REGACOM), 2006/150 ("Consolidation of Research Units"), and by the Spanish Ministry of Science and Innovation under projects COMONSENS (ref. CSD2008-00010) of the CONSOLIDER-INGENIO 2010 Program and SPROACTIVE (ref. TEC2007-68094-C02-01/TCM).

to estimate this secret key from the observations at hand, then he/she has completely "broken" the watermarking system [3],[6]. An additional assumption is that all parameters of the watermarking scheme are known (except the secret key), according to Kerckhoffs' principle; hence, the attacker is only interested in disclosing the secret key. The information about the key provided by the observations is quantified by means of the Shannon's mutual information, and the remaining uncertainty or "equivocation" about the key is measured by the differential entropy of the key conditioned on the observations, which can be related to the lowest attainable error in the estimation of the secret key. The number of observations needed to achieve a certain estimation accuracy can be regarded to as the "security level" of the watermarking scheme. This approach has been utilized in [6] and [7] for analyzing the security of lattice DC-DM methods, and it is used here for developing a complete security analysis of spread spectrum methods, i.e., those methods that perform watermark embedding in a secret subspace by modulating a "secret carrier" with the symbols to be embedded. Specifically, three well-known data hiding methods are considered: additive Spread Spectrum [8], attenuated Spread Spectrum [9], and Improved Spread Spectrum [10]. Two different scenarios for security assessment are considered, according to the classification given in [3]:

1) Known Message Attack (KMA): the attacker is assumed to have access to watermarked signals and the messages embedded in each of those signals. This scenario constitutes the basis for the study of more involved scenarios and provides the main insight into the security problem (influence of the embedding parameters). It is also useful for the study of security in some watermark detection scenarios.
2) Watermarked Only Attack (WOA): the only information available to the attacker are the watermarked signals, without any knowledge of the embedded messages. As such, WOA models most of the data hiding scenarios of practical interest.

The reader must be aware that the security of spread spectrum watermarking has been first addressed in [3]. An obvious difference between the present paper and [3] is that we resort to the Shannon's equivocation instead of the Fisher information for performing the theoretical analysis. Further contributions of the present paper, which have not been previously addressed by other authors, are the following: the evaluation of the security in asymptotic conditions, the evaluation of the tradeoff robustness-security in the considered embedding functions, the derivation of bounds on the estimation performance, the theoretical analysis of the estimators proposed in [3], the proposal of new estimators, and their application to practical scenarios. Some of these contributions have been partially presented in [11].

Spread spectrum methods continue to be widely used, as many embedding functions existing nowadays are based on spreading. Thus, the analysis presented in this paper is expected to provide useful insights in the identification of security weaknesses of current spread spectrum schemes and the design of improved ones. In this regard, we want to remark that spread-spectrum-based embedding functions with improved security features have already been proposed by other authors in [12]. One of these embedding functions (Natural Watermarking) achieves perfect secrecy for Gaussian hosts. The other one (Circular Watermarking) is strongly related to the ISS embedding function analyzed in this paper, and it is briefly addressed in Section VII-A2 from a practical point of view. Another interesting contribution from [12] is the proposal of a general formulation for clarifying the different degrees of security existent in data hiding. Based on Kerckhoffs' principle, the following classification of data hiding schemes (in increasing order of security) is proposed: insecure, key secure, subspace secure, and stegosecure. These security classes are related to the degree of concealment of the secret keys. All the methods analyzed in the present paper pertain to the category of insecure, with the exception of Circular Watermarking, which is key secure. The practical estimators used in [12] are the same as those used in [3].

The remaining of this paper is organized as follows. In Section II, the problem is formalized, introducing the working assumptions and the notation. Section III studies the security of the classical spread spectrum embedding function, whereas Section IV addresses the security when host rejection is considered. In Section V, we provide bounds for the estimation of the spreading vector based on the information-theoretic analysis of the previous sections. In Section VI, practical methods for estimating the secret key are proposed and analyzed, and Section VII shows experimental results obtained on real images, also briefly discussing the extension of the proposed estimators to other scenarios. In Section VIII we consider the links, in terms of security, between the methods studied here and other methods that are strongly related to the spread spectrum formulation followed in this paper. Finally, the conclusions are summarized in Section IX.

II. FORMAL PROBLEM STATEMENT AND NOTATION

In this section, the problem of watermarking security is formalized. Hereinafter, boldface letters denote column vectors, whereas italicized letters denote scalar variables. The considered scenario is the following: the secret key of a certain user is reutilized N_o times for watermarking a set of host signals $\{\mathbf{x}_i \in \mathbb{R}^{n\times 1},\ i = 1,\ldots,N_o\}$, where $x_{i,j}$ is the jth component of the ith host vector. These vectors $\mathbf{x}_i$ are usually obtained through some partition of the digital item (image, audio signal, etc.) to be watermarked. The secret key is a n-dimensional vector $\mathbf{s}$ that parameterizes the embedding function, and it is usually termed "secret carrier" or "spreading vector" (we will use both terms without distinction). For the family of methods considered in this paper, the embedding function can be written as

$$\mathbf{y}_i = \mathbf{x}_i + (-1)^{m_i}\mathbf{s} + \Psi(\mathbf{x}_i, \mathbf{s}) = \mathbf{x}_i + \mathbf{w}_i,\ i = 1,\ldots,N_o, \tag{1}$$

where $m_i \in \mathcal{M} = \{0,1\}$ denotes the embedded message that modulates $\mathbf{s}$, and the function $\Psi : \mathbb{R}^{n\times 1}\times\mathbb{R}^{n\times 1} \rightarrow \mathbb{R}^{n\times 1}$ is used for host-rejection purposes. The resulting embedding rate of the scheme is $R \triangleq \log(2)/n$. We will consider that the objective of the attacker is to obtain an estimate of $\mathbf{s}$ using the information contained in the sequence of observations $\{\mathbf{o}_i,\ i = 1,\ldots,N_o\}$, where $\mathbf{o}_i = [\mathbf{y}_i^T, m_i]^T$ in the KMA case,[1] and $\mathbf{o}_i = \mathbf{y}_i$ in the WOA case.

The signals involved in the theoretical security analysis will be modeled as random variables, which will be denoted by capital letters. Instantiations of these random variables will be denoted by lowercase letters as in (1). For the theoretical analysis, the host signals $\{\mathbf{X}_i, i = 1,\ldots,N_o\}$ will be assumed to be i.i.d. and Gaussian-distributed with zero mean: $\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \sigma_X^2\mathbf{I}_n)$, where $\mathbf{I}_n$ denotes the identity matrix of size $n \times n$. Furthermore, the $\mathbf{X}_i$ are assumed to be mutually independent. The messages $\{M_i, i = 1,\ldots,N_o\}$ embedded in different observations are assumed to be mutually independent and equiprobable in $\mathcal{M} = \{0,1\}$. As for the spreading vector, it is assumed to be Gaussian-distributed and i.i.d, $\mathbf{S} \sim \mathcal{N}(0, \sigma_S^2\mathbf{I}_n)$. The motivation for this choice is twofold: 1) on one hand, the

[1] The notation $[a, b]$ indicates concatenation of the row vectors a and b.

Gaussian distribution for $\mathbf{S}$ is a long-standing assumption [8] that has been frequently recalled in later theoretical analysis (also for mathematical simplicity); 2) on the other hand, for a given embedding distortion, the Gaussian distribution maximizes the a priori entropy of $\mathbf{S}$, which is desirable from the security standpoint.

The information about $\mathbf{S}$ provided by the observations is frequently referred to as "information leakage". This information will be quantified in an information-theoretic manner by means of the mutual information between $\mathbf{S}$ and the observations [13], which is denoted by $I(\mathbf{O}_1,\ldots,\mathbf{O}_{N_o};\mathbf{S})$. The remaining uncertainty about $\mathbf{S}$, given N_o observations, will be represented by means of the "equivocation" or "residual entropy", defined as $h(\mathbf{S}|\mathbf{O}_1,\ldots,\mathbf{O}_{N_o})$, where $h(A)$ denotes the differential entropy of the continuous random variable A [13]. For a discrete random variable B, the entropy will be denoted by $H(B)$. Throughout this paper, all the entropies and mutual informations will be expressed in natural units.

Notice that the a priori uncertainty about $\mathbf{S}$ is given by $h(\mathbf{S})$, and that the equivocation is a decreasing function of N_o. The number of observations (N_o) needed to achieve a certain value of the equivocation is regarded to as the security level of the watemarking method. For providing a fair comparison between the security level of the different methods, they will be studied under the same conditions of embedding distortion. The embedding distortion per dimension is defined as $D_w \triangleq \frac{1}{n}E[||\mathbf{W}_i||^2]$. Thus, D_w quantifies the power of the watermark. Furthermore, we define the Document to Watermark Ratio (DWR) for quantifying the relative powers between the host and the watermark:

$$\text{DWR} \triangleq 10\log_{10}\frac{\frac{1}{n}E[||\mathbf{X}_i||^2]}{D_w} = 10\log_{10}\xi,$$

where $\xi \triangleq \sigma_X^2/D_w$, the operator $E[\cdot]$ denotes mathematical expectation, and $||\mathbf{x}||^2 \triangleq \sum_i x_i^2$ denotes the squared Euclidean norm of $\mathbf{x}$.

Other notational conventions are the following: $\Gamma(z)$ denotes the complete Gamma function. If $\mathbf{X}\sim\mathcal{N}(\mathbf{0},\sigma_X^2\mathbf{I}_n)$, then $T \triangleq ||\mathbf{X}||^2$ follows a Chi-squared distribution with n degrees of freedom, which is denoted as $\chi^2(n,\sigma_X)$. If $\mathbf{X}\sim\mathcal{N}(\mathbf{v},\sigma_X^2\mathbf{I}_n)$, then $T' \triangleq ||\mathbf{X}||^2$ follows a noncentral Chi-squared, denoted by $\chi'^2(n,\mathbf{v},\sigma_X)$. The probability density function (pdf) of a continuous random variable A is denoted by $f(a)$. The transpose of a vector $\mathbf{x}$ is denoted by $\mathbf{x}^T$. The estimate of the vector $\mathbf{x}$ is denoted by $\hat{\mathbf{x}}$.

III. Additive Spread Spectrum (add-SS)

Binary add-SS, as proposed by Cox et al. [8], is the most popular and widely studied watermarking method. No host rejection is performed, so the embedding function is simply given by

$$\mathbf{Y}_i = \mathbf{X}_i + (-1)^{M_i}\mathbf{S},\ \text{for } i=1,\ldots,N_o. \tag{2}$$

The embedding distortion results in $D_w=\sigma_S^2$. The security level for KMA and WOA scenarios is addressed below.

A. KMA scenario

Under the assumptions made in Section II, the KMA scenario can be seen as a simple additive Gaussian channel. Although this scenario was already considered in [4], a simple, alternative derivation is given in this paper for the sake of completeness. Furthermore, this derivation introduces some results that will be used later on. Let us denote

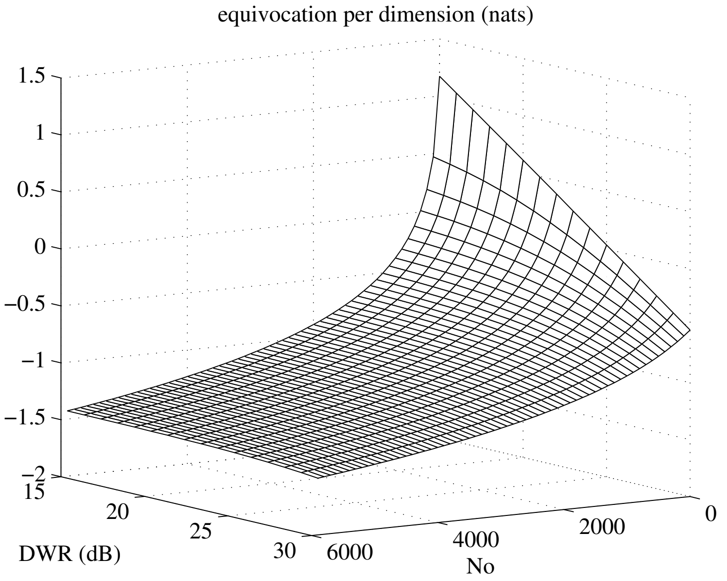


Fig. 1. Equivocation per dimension for add-SS in the KMA scenario.

by $\bar{\mathbf{S}}_{N_o}$ the random variable $\mathbf{S}$ conditioned on N_o observations. From the embedding function of add-SS, it follows that the components $\{\bar{S}_i,\ i=1,\ldots,n\}$, of $\bar{\mathbf{S}}_{N_o}$ are all mutually independent and Gaussian-distributed [14]. Given a particular realization $\{\mathbf{Y}_1=\mathbf{y}_1,\ldots,\mathbf{Y}_{N_o}=\mathbf{y}_{N_o},M_1=m_1,\ldots,M_{N_o}=m_{N_o}\}$, we have $\bar{\mathbf{S}}_{N_o}\sim\mathcal{N}(\mathbf{v},\sigma^2_{\bar{S}_{N_o}}\mathbf{I}_n)$, with

$$v_i = \frac{\sigma_S^2}{N_o\sigma_S^2+\sigma_X^2}\boldsymbol{\mu}^T\mathbf{y}^{(i)}, \quad i=1,\ldots,n, \tag{3}$$

$$\sigma^2_{\bar{S}_{N_o}} = \frac{\sigma_X^2\sigma_S^2}{N_o\sigma_S^2+\sigma_X^2}, \tag{4}$$

where $\boldsymbol{\mu}\triangleq[(-1)^{m_1},\ldots,(-1)^{m_{N_o}}]^T$, and $\mathbf{y}^{(i)}\triangleq[y_{1,i},\ldots,y_{N_o,i}]^T$. Since $\bar{\mathbf{S}}_{N_o}$ is i.i.d. Gaussian, its entropy is given by $h(\bar{\mathbf{S}}_{N_o})=\frac{n}{2}\log(2\pi e\sigma^2_{\bar{S}_{N_o}})$, which does not depend on the particular realization of the observations. Hence, we can conclude that the equivocation per dimension is

$$\frac{1}{n}h(\mathbf{S}|\mathbf{Y}_1,\ldots,\mathbf{Y}_{N_o},M_1,\ldots,M_{N_o})_{\text{add-SS}}=\frac{1}{2}\log\left(2\pi e\frac{\sigma_S^2}{1+N_o\cdot\xi^{-1}}\right). \tag{5}$$

Now, it is straightforward to see that the information leakage per dimension reads as

$$\frac{1}{n}I(\mathbf{Y}_1,\ldots,\mathbf{Y}_{N_o};\mathbf{S}|M_1,\ldots,M_{N_o})_{\text{add-SS}}=\frac{1}{2}\log\left(1+N_o\cdot\xi^{-1}\right). \tag{6}$$

The information leakage (equivocation) is concave (convex) and strictly increasing (decreasing) with N_o, and its increasing (decreasing) rate is dependent on the DWR. Although (5) depends on the value of σ_S^2, for large N_o we have $\frac{1}{n}h(\mathbf{S}|\mathbf{Y}_1,\ldots,\mathbf{Y}_{N_o},M_1,\ldots,M_{N_o})_{\text{add-SS}}\approx\frac{1}{2}\log\left(2\pi e\sigma_X^2/N_o\right)$, as can be seen in Fig. 1. Notice that both (5) and (6) are independent of n, meaning that the information leakage about each dimension is independent of the total number of dimensions. In other words, the difficulty of estimating each component of $\mathbf{S}$ does not depend on its total length when the embedded messages are a priori known.

B. WOA scenario

The WOA scenario can be seen as an additive Gaussian channel with an unknown scaling factor which, according to the binary transmission scheme given in Eq. (2) and the assumptions stated in Section II, takes values ±1 equiprobably in each channel use. In this case, an exact expression for the information leakage cannot be obtained, so we derive upper and lower bounds. Following an approach similar to [7], by resorting to the chain rule for mutual informations [13], the information leakage is rewritten as

$$\begin{aligned} I(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};\mathbf{S})_{\text{add-SS}} &= I(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};\mathbf{S},M_1,\dots,M_{N_o})_{\text{add-SS}} - I(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};M_1,\dots,M_{N_o}|\mathbf{S})_{\text{add-SS}} \\ &= I(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};\mathbf{S}|M_1,\dots,M_{N_o})_{\text{add-SS}} + I(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};M_1,\dots,M_{N_o})_{\text{add-SS}} \\ &- I(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};M_1,\dots,M_{N_o}|\mathbf{S})_{\text{add-SS}}. \end{aligned} \tag{7}$$

The first term of Eq. (7) has been already calculated in (6). The second and third terms of (7) represent the amount of information that can be learned by an attacker and by a fair user, respectively, about the sequence of embedded messages. These quantities are studied in Appendix A, resulting in the following upper and lower bounds to the information leakage per dimension:

$$\begin{aligned} \frac{1}{n}I(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};\mathbf{S})_{\text{add-SS}} &\leq \frac{1}{2}\log\left(1+N_o\cdot\xi^{-1}\right)+\frac{N_o}{n}I(\bar{\mathbf{X}}_{N_o}+(-1)^{M_{N_o}}\mathbf{V}_{N_o};M_{N_o}|\mathbf{V}_{N_o}) \\ &- \frac{N_o}{n}I(\mathbf{X}_1+(-1)^{M_1}\mathbf{S};M_1|\mathbf{S}),\ \text{for } N_o\geq 2, \end{aligned} \tag{8}$$

and

$$\begin{aligned} \frac{1}{n}I(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};\mathbf{S})_{\text{add-SS}} &\geq \frac{1}{2}\log\left(1+N_o\cdot\xi^{-1}\right)+\frac{1}{n}\sum_{i=2}^{N_o}I(\bar{\mathbf{X}}_i+(-1)^{M_i}\mathbf{V}_i;M_i|\mathbf{V}_i) \\ &- \frac{N_o}{n}I(\mathbf{X}_1+(-1)^{M_1}\mathbf{S};M_1|\mathbf{S}),\ \text{for } N_o\geq 2. \end{aligned} \tag{9}$$

In the expressions (8) and (9), $\bar{\mathbf{X}}_i\sim\mathcal{N}(\mathbf{0},(\sigma_X^2+\sigma_{\bar{S}_{i-1}}^2)\mathbf{I}_n)$, with $\sigma_{\bar{S}_{i-1}}^2$ given by (4), and $\mathbf{V}_i\sim\mathcal{N}(\mathbf{0},\frac{(i-1)\sigma_S^4}{(i-1)\sigma_S^2+\sigma_X^2}\mathbf{I}_n)$. The second and third terms of (8) and (9) must be computed by taking into account that

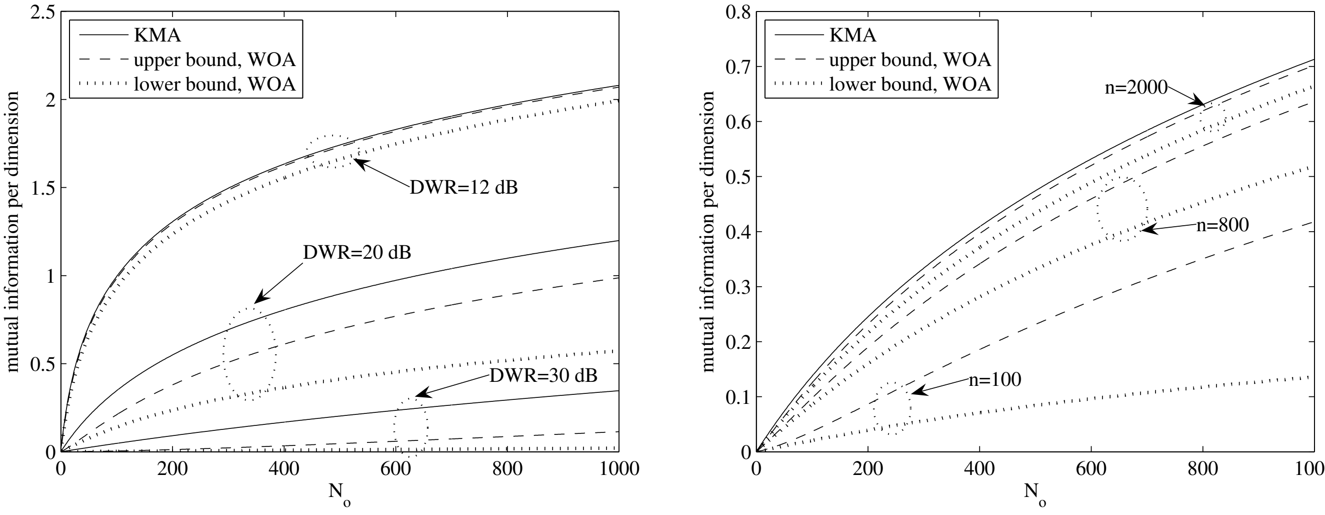
$$I(\mathbf{X}_1+(-1)^{M_1}\mathbf{S};M_1|\mathbf{S})_{\text{add-SS}} = E\left[h((-1)^{M_1}||\mathbf{S}||^2+\mathbf{X}_1^T\mathbf{S}|\mathbf{S}=\mathbf{s})\right]-\frac{1}{2}E\left[\log\left(2\pi e\sigma_X^2\|\mathbf{S}\|^2\right)\right],$$

where the expectation is taken over $\mathbf{S}$. Notice that the above upper and lower bounds differ only in their second term. Nevertheless, they cannot be given in closed-form, so numerical integration (on a scalar domain) is needed. This integration is straightforward because, under the i.i.d. Gaussian assumption for $\mathbf{S}$, $||\mathbf{S}||^2$ follows a Chi-square distribution $\chi^2(n,\sigma_S)$. A comparison between the information leakage (per dimension) in KMA and WOA scenarios is shown in Fig. 2:

- Fig. 2(a) shows that, when the parameter n is fixed, decreasing the DWR increases the information leakage in the KMA scenario (recall Eq. (6)), and simultaneously reduces the gap between KMA and WOA.
- Fig. 2(b) shows the effect of varying the length of the spreading vector, n, when the DWR is fixed. In this case, we can see that the information leakage of the KMA scenario is approached as n is increased. Thus, the security level of the WOA scenario is strongly dependent on the value of n, contrarily to the KMA scenario.

The asymptotic behavior of the information leakage is formalized below. First, let us define the "loss function" as

$$\begin{aligned} \delta(N_o)_{\text{add-SS}} &\triangleq I(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};\mathbf{S}|M_1,\dots,M_{N_o})_{\text{add-SS}} - I(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};\mathbf{S})_{\text{add-SS}} \\ &= h(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};M_1,\dots,M_{N_o}|\mathbf{S})_{\text{add-SS}} - h(\mathbf{Y}_1,\dots,\mathbf{Y}_{N_o};M_1,\dots,M_{N_o})_{\text{add-SS}}. \end{aligned} \tag{10}$$



(a) $n = 100$

(b) DWR = 25 dB

Fig. 2. Comparison between the information leakage in KMA and WOA scenarios for add-SS. Figures 2(a) and 2(b) show the effect of varying the DWR and n, respectively.

The loss function represents the information about $\mathbf{S}$ that is lost due to the a priori ignorance of the embedded messages, and it is non-negative for all N_o.

Theorem 1: The loss function for add-SS in the WOA scenario can be upper bounded as

$$\delta(N_o)_{\text{add-SS}} \leq \log(2) + \sum_{i=2}^{N_o} H\left(\frac{\tau_i^{\frac{n}{2}}}{2}\right), \tag{11}$$

where

$$\tau_i = \frac{i\sigma_S^2\sigma_X^2 + \sigma_X^4}{(i-1)\sigma_S^4 + i\sigma_S^2\sigma_X^2 + \sigma_X^4},$$

and $H(\cdot)$ denotes the binary entropy function. The right hand side of (11) is decreasing with n and DWR^{-1}, and the following asymptotic properties hold:

1) For fixed n, $\lim_{\text{DWR}\to-\infty} \delta(N_o)_{\text{add-SS}} \leq \log(2)$.
2) For fixed DWR, $\lim_{n\to\infty} \delta(N_o)_{\text{add-SS}} \leq \log(2)$.

Proof: See Appendix B. ■

Theorem 1 basically states that the ignorance of the embedded messages does not affect the difficulty of estimating $\mathbf{S}$ if either the DWR or the embedding rate R (recall that $R = \log(2)/n$) are small enough. Although the first case is of virtually null relevance in practice, the second case is of major importance for practical applications. When high robustness is sought, the watermark is usually embedded at very low rates that allow to recover the message with low complexity. In such case a few observations suffice to obtain an estimate of $\mathbf{S}$ that in turn allows accurate recovery of the embedded message. Nevertheless, it is important to point out, as realized before in [3] (by means of blind source separation theoretic arguments), that the penalty to pay for not knowing the messages M_i comes in the form of an ambiguity in the sign of $\mathbf{S}$, independently of n and of the DWR. An alternative way for proving this ambiguity is to show that the a posteriori probability of the spreading vector in the WOA scenario is independent of its sign. One can easily check that

$$f(\mathbf{s}_0|\mathbf{Y}_i=\mathbf{y}_i)=\frac{f(\mathbf{y}_i|\mathbf{S}=\mathbf{s}_0)f(\mathbf{s}_0)}{f(\mathbf{y}_i)}=\frac{f(\mathbf{y}_i|\mathbf{S}=-\mathbf{s}_0)f(-\mathbf{s}_0)}{f(\mathbf{y}_i)}=f(-\mathbf{s}_0|\mathbf{Y}_i=\mathbf{y}_i)\ \forall\ \mathbf{s}_0\in\mathbb{R}^n, \tag{12}$$

so the sign ambiguity becomes patent. Notice that (12) holds regardless of the statistical distribution of the host, but it is needed that $\mathbf{S}$ is circularly symmetric. In the information-theoretic analysis, the sign ambiguity is reflected in the factor $\log(2)$ shown in Theorem 1. The sign ambiguity is irreducible without a priori knowledge of the embedded messages. For instance, in the particular case where all $M_i = M$ in the N_o observations, the sign ambiguity still exists, and it cannot be undone unless one of the M_i is known by the attacker. In any case, the sign ambiguity does not prevent from estimating the one-dimensional subspace spanned by $\mathbf{s}$, a feature that will be exploited by practical estimators in Section VI.

IV. Spread Spectrum with host rejection

After studying the security properties of add-SS, the influence of the host rejection mechanisms in the security level is addressed in this section. Two particular methods are studied: “attenuated spread spectrum” [9] and “improved spread spectrum” [10].

A. *Attenuated Spread Spectrum (γ-SS)*

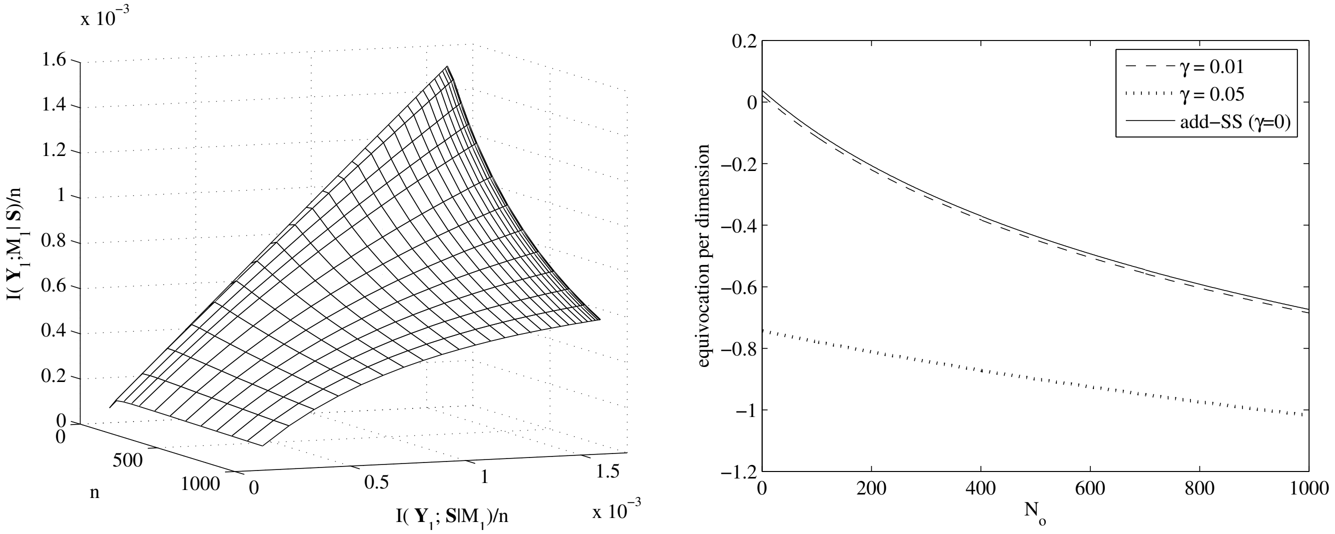
The attenuated spread spectrum technique proposed in [9] consists in attenuating the host prior to embedding, in order to optimize the power transmission subject to an MSE distortion constraint (the embedding distortion). The embedding function is as follows:

$$\mathbf{Y}_i=(1-\gamma)\cdot\mathbf{X}_i+(-1)^{M_i}\cdot\mathbf{S},\ \text{for } i=1,\ldots,N_o, \tag{13}$$

where $0\leq\gamma\leq 1$ is a host-rejection parameter. The embedding distortion of this scheme is given by $D_w=\gamma^2\sigma_X^2+\sigma_S^2$. We will refer to this technique in the following as γ-SS. The parameter γ can be adjusted so as to optimize some performance measure, usually the error probability. Thus, the performance (in terms of robustness) of γ-SS is at least as good as that of add-SS, as the latter is just a particular case of γ-SS for $\gamma = 0$. However, its security level can be shown to be always worse than that of add-SS for the same values of DWR and n. In order to provide a fair comparison between both schemes, we impose that $\sigma_S^2=\sigma_X^2\left(\xi^{-1}-\gamma^2\right)$ so as to get $D_w=\sigma_S^2$ as in add-SS. Note that this condition restricts the value of γ to the interval $0\leq\gamma\leq\xi^{-\frac{1}{2}}$, i.e. in practical scenarios (with low embedding distortion) γ takes values close to 0. It is easy to see that the results obtained for add-SS can be straightforwardly adapted to γ-SS replacing σ_S^2 by $\sigma_X^2\left(\xi^{-1}-\gamma^2\right)$ and σ_X^2 by $(1-\gamma)^2\sigma_X^2$ in the corresponding expressions. The equivocation for the KMA scenario, for instance, results in

$$\frac{1}{n}h(\mathbf{S}|\mathbf{Y}_1,\ldots,\mathbf{Y}_{N_o},M_1,\ldots,M_{N_o})_{\gamma\text{-SS}}=\frac{1}{2}\log\left(2\pi e\frac{\sigma_X^2(\xi^{-1}-\gamma^2)(1-\gamma)^2}{(1-\gamma)^2+N_o(\xi^{-1}-\gamma^2)}\right). \tag{14}$$

The expression above can be shown to be monotonically decreasing with γ. The results for add-SS WOA can be easily generalized as well to γ-SS. The most interesting consequence of introducing the parameter γ is the existence of a tradeoff between robustness and security. This tradeoff is illustrated in Fig. 3(a), which shows the plot of the



(a) (b)

Fig. 3. γ-SS for KMA scenario and DWR = 25 dB. Tradeoff robustness-security as a result of varying γ in the interval $[0, \xi^{-\frac{1}{2}}]$, for $N_o = 1$ (a) and equivocation per dimension (b).

information leakage for $N_o = 1$ (in the KMA scenario) vs. the achievable rate for a fair user. The latter can be computed by numerical integration by taking into account that

$$I((1-\gamma)\mathbf{X}_1 + (-1)^{M_1}\mathbf{S}; M_1|\mathbf{S})_{\gamma\text{-SS}} = E\left[h((-1)^{M_1}||\mathbf{S}||^2 + (1-\gamma)\mathbf{X}_1^T\mathbf{S}|\mathbf{S})\right] - \frac{1}{2}E\left[\log\left(2\pi e(1-\gamma)^2\sigma_X^2\|\mathbf{S}\|^2\right)\right].$$

Fig. 3(a) basically shows that increasing the achievable rate for fair users will provide more information for attackers interested in estimating $\mathbf{S}$. As can be seen, the tradeoff is also dependent of n, since this parameter affects the achievable rate. Fig. 3(b) shows the equivocation per dimension in the KMA scenario for several values of the parameter γ, evidencing the degradation of the security level as the host rejection is increased.

B. Improved Spread Spectrum (ISS)

ISS [10] is the result of introducing a host-interference-rejection mechanism in add-SS, fundamentally different from γ-SS in that ISS attenuates the host only in the direction of embedding, thus saving in embedding distortion and improving the performance of the latter in terms of robustness. We will consider the linear version of ISS, whose embedding function is as follows:

$$\mathbf{Y}_i = \mathbf{X}_i + (-1)^{M_i}\nu\mathbf{S} - \lambda\frac{\mathbf{X}_i^T\mathbf{S}}{||\mathbf{S}||^2}\mathbf{S}, \text{ for } i = 1, \ldots, N_o, \tag{15}$$

where $0 \leq \lambda \leq 1$ is the host-rejection parameter, and ν is a parameter for fixing the embedding distortion. The embedding distortion in ISS can be computed as follows. For the i-th observation and a particular $\mathbf{s}$ we have

$$E[||\mathbf{W}_i||^2|\mathbf{S} = \mathbf{s}] = E\left[\left\|\left((-1)^{M_i}\nu - \lambda\frac{\mathbf{X}_i^T\mathbf{s}}{||\mathbf{s}||^2}\right)\mathbf{s}\right\|^2\right] = \nu^2||\mathbf{s}||^2 + \lambda^2\sigma_X^2. \tag{16}$$

Finally, for a zero-mean Gaussian spreading vector, $D_w = \frac{1}{n}E[||\mathbf{W}_i||^2] = \nu^2\sigma_S^2 + \frac{\lambda^2}{n}\sigma_X^2$, which is the same result as for a spreading vector with constant norm equal to $n\sigma_S^2$ (the case originally considered in [10]). For a fair comparison with add-SS, the embedding distortion is fixed to $D_w = \sigma_S^2$, as proposed by Malvar and Florêncio [10]. This imposes

$$\nu = \left(1 - \frac{\lambda^2\xi}{n}\right)^{\frac{1}{2}}. \tag{17}$$

Since ν must be real, the maximum allowable value for λ is determined by $\min\{1, \sqrt{n\xi^{-1}}\}$. Clearly, λ can be made arbitrarily close to 1 by increasing n, thus achieving complete rejection of the host interference. In general, the parameter λ is tuned so as to optimize the performance of ISS in terms of error probability. Clearly, ISS with $\lambda = 0$ is equivalent to add-SS as described in Section III. We are concerned in this section with determining the effect on the security level of using $\lambda > 0$. The study will be carried out for the KMA scenario, where

$$I(\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}; \mathbf{S}|M_1, \ldots, M_{N_o})_{\text{ISS}} = h(\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}|M_1, \ldots, M_{N_o}) - h(\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}|\mathbf{S}, M_1, \ldots, M_{N_o}). \quad (18)$$

The second term is easy to compute: given the messages and $\mathbf{S} = \mathbf{s}$, the observations are mutually independent, following a Gaussian distribution

$$\mathbf{Y}_i|\mathbf{S} = \mathbf{s}, M_i = m_i \sim \mathcal{N}((-1)^{m_i}\nu\mathbf{s}, \mathbf{\Sigma}_{\mathbf{S}}), \quad (19)$$

with $\mathbf{\Sigma}_{\mathbf{S}} = E\left[(\mathbf{Y}_i - (-1)^{m_i}\nu\mathbf{s})^T \cdot (\mathbf{Y}_i - (-1)^{m_i}\nu\mathbf{s})\right]$ the covariance matrix of the $\mathbf{Y}_i$ conditioned on the realization of $\mathbf{S}$ and M_i. The eigenvalue decomposition of $\mathbf{\Sigma}_{\mathbf{S}}$ is given by $= \mathbf{U}_{\mathbf{S}}\mathbf{\Lambda}\mathbf{U}_{\mathbf{S}}^T$, where

$$\mathbf{\Lambda} = \begin{bmatrix} (1-\lambda)^2\sigma_X^2 & 0 \\ 0 & \sigma_X^2\mathbf{I}_{n-1} \end{bmatrix}, \quad (20)$$

and $\mathbf{U}_{\mathbf{S}}$ is a unitary matrix whose first column is colinear to $\mathbf{s}$. That is, the watermarked signal $\mathbf{Y}_i$ can be seen as a signal $(-1)^{M_i}\nu\mathbf{s}$ transmitted in a Gaussian channel with noise correlated with the signal. It turns out that in ISS not only the mean of the observations provides information about $\mathbf{S}$, but also the covariance matrix of the noise (here, the attenuated host). Using (19) and (20) we can write

$$h(\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}|\mathbf{S}, M_1, \ldots, M_{N_o}) = \sum_{i=1}^{N_o} h(\mathbf{Y}_i|M_i, \mathbf{S}) = \frac{N_o}{2}\log\left((2\pi e)^n \cdot (\sigma_X^2)^n \cdot (1-\lambda)^2\right). \quad (21)$$

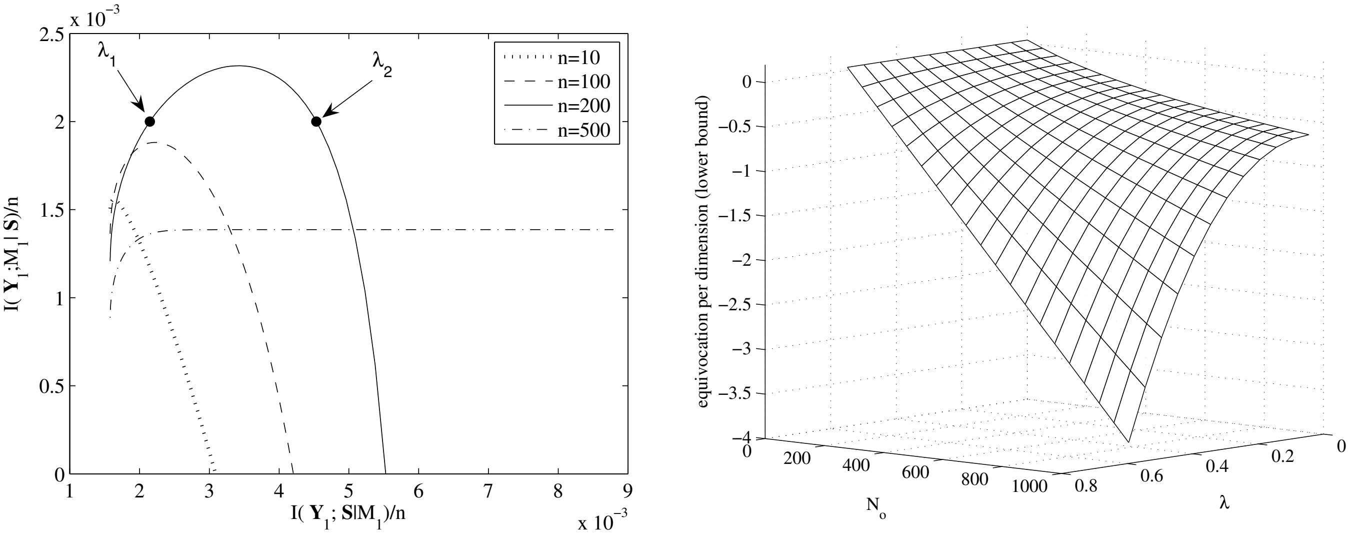
Since the first term of (18) is hard to compute analytically, we first formalize the general behavior of the information leakage in the following theorem for $N_o = 1$.

Theorem 2: The information leakage in ISS is a convex and increasing function of the host-rejection parameter λ, and for $N_o = 1$ it is given by

$$\frac{1}{n}I(\mathbf{Y}_1; \mathbf{S}|M_1)_{\text{ISS}} = \frac{1}{2}\log\left(1 + \frac{\lambda(\lambda-2)}{n} + \nu^2\xi^{-1}\right) - \frac{1}{n}\log(1-\lambda). \quad (22)$$

Proof: In Appendix C, the exact value of the information leakage for $N_o = 1$ is shown to be given by (22). If we compute the first and second derivatives of the information leakage in terms of λ, we find out that the function is convex and increasing in the interval $\lambda \in [0, \min\{1, \sqrt{n\xi^{-1}}\}]$. ■

In ISS, the achievable rate depends on the value of λ. The optimum λ that maximizes this rate depends on the DWR and the power of the attacking noise (see [10] for further discussion). This behavior in conjunction with Theorem 2 shows that, similarly to the γ-SS scheme, the host-rejection mechanism of ISS induces a tradeoff between information leakage and achievable rate. This tradeoff is illustrated in Fig. 4(a) by plotting $I(\mathbf{Y}_1; \mathbf{S}|M_1)/n$ vs $I(\mathbf{Y}_1; M_1|\mathbf{S})/n$ in terms of λ, for $\lambda \in [0, \min\{1, \sqrt{n\xi^{-1}}\}]$. It can be noticed the concavity of the curves, which present a global maximum of the achievable rate for a certain λ_{max} (dependent of n). Increasing λ beyond this value has the double



(a) (b)

Fig. 4. ISS for KMA scenario and DWR = 25 dB. Tradeoff robustness-security as a result of varying λ in the interval $[0, \min\{1, \sqrt{n\xi^{-1}}\}]$ (a), and lower bound on the equivocation per dimension for $n = 100$ (b).

(negative) effect of not increasing further the achievable rate but increasing the information leakage. Notice that for $\lambda = \sqrt{n\xi^{-1}} < 1$, from (17) we have $\nu = 0$, so $I(\mathbf{Y}_1; M_1|\mathbf{S}) = 0$. Even in this case, we can see in Fig. 4(a) that the information leakage is not null, due to the dependence of the covariance matrix ($\mathbf{\Sigma_S}$) on the spreading vector $\mathbf{S}$. For a watermarker, the values of λ for which the tradeoff curve is at the left of the maximum are preferable. For example, the values λ_1 and λ_2 depicted in Fig. 4(a) yield the same achievable rate for $n = 200$, but λ_1 produces a smaller information leakage than λ_2.

For the case $N_o \geq 1$, an upper bound to the information leakage is derived in Appendix D, obtaining

$$\frac{1}{n} I(\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}; \mathbf{S}|M_1, \ldots, M_{N_o})_{\text{ISS}} \leq \frac{1}{2}\log\left(\left(1 + \frac{\lambda(\lambda-2)}{n}\right)^{N_o}\left(1 + \frac{N_o \nu^2 \sigma_S^2}{\sigma_X^2\left(1 + \frac{\lambda(\lambda-2)}{n}\right)}\right)\right) - \frac{N_o}{n}\log(1-\lambda). \quad (23)$$

The bound (23) on the information leakage produces a lower bound on the equivocation, which is plotted in Fig. 4(b) for different values of λ and compared to add-SS. We can see that the equivocation decreases as λ increases, in accordance with Theorem 2. This bound can be used to derive a conservative security level. Notice that (23) coincides with (22) for $N_o = 1$.

REMARK 1: As expected, for $\lambda = 0$ (which implies $\nu = 1$) the bound (23) coincides with (6), the information leakage for add-SS. This is because the hypothesis of independence used in the first bounding of (78) of Appendix D is fulfilled for $\lambda = 0$.

REMARK 2: For $n \to \infty$, (23) tends to (6). This means that asymptotically there is no penalty in security level for using host rejection in one dimension, constituting a major advantage over γ-SS. Remember that the latter performs host rejection in all dimensions, and as such it cannot benefit from increasing n for concealing the information about $\mathbf{S}$. Nevertheless, we want to remark that increasing n in ISS has the same effect as for add-SS stated in Theorem 1, namely, that the information leakage in the WOA scenario approaches that of the KMA scenario when $n \to \infty$. This can be easily proved for ISS following similar guidelines as those of Appendix B.

V. Bounds on the estimation error

In this section we provide fundamental performance bounds for practical estimators of the spreading vector. These bounds are based on the information-theoretic results derived in the previous sections. The aim is to translate the equivocation into other measures that result useful for the evaluation of the security from a practical point of view. The first bound is concerned with the mean-squared error between the spreading vector ($\mathbf{s}$) and its estimate ($\hat{\mathbf{s}}$), and the second one with the normalized correlation between $\mathbf{s}$ and $\hat{\mathbf{s}}$. The achievability of each bound is also discussed.

A. Bound on the mean-squared error (MSE)

Let us define the estimation error as $\mathbf{e} \triangleq \mathbf{s} - \hat{\mathbf{s}}$ and its variance per dimension as $\sigma_E^2 \triangleq \frac{\text{tr}(\mathbf{\Sigma}_E)}{n}$, where $\mathbf{\Sigma}_E$ is the covariance matrix of $\mathbf{e}$, and $\text{tr}(\mathbf{\Sigma}_E)$ denotes the trace of $\mathbf{\Sigma}_E$. As shown in [6], we have the following lower bound:

$$\sigma_E^2 \geq \frac{1}{2\pi e} \exp\left(\frac{2}{n} h(\mathbf{S}|\mathbf{O}_1, \ldots, \mathbf{O}_{N_o})\right). \tag{24}$$

Hence, the equivocation can be regarded as the exponent of the estimation error lower bound. Inserting (5) into (24), for add-SS in the KMA scenario we have $\sigma_{E_{\text{KMA}}}^2 \geq \sigma_S^2/(1+N_o \cdot \xi^{-1}) \approx \sigma_X^2/N_o$, which is achievable when the estimation error is zero-mean and Gaussian-distributed, and the approximation holds for large N_o. This bound coincides with the Cramér-Rao lower bound, which was calculated in [4], and with the variance of the MMSE estimator [14]. This is not surprising, as the MMSE estimator is unbiased and its estimation error is Gaussian-distributed, thus fulfilling the conditions for achieving the bound.

A similar bound for the WOA scenario could be obtained by inserting the corresponding equivocation into (24). Taking into account Theorem 1 we find out that $\lim_{n\to\infty} \sigma_{E_{\text{WOA}}}^2 \geq \sigma_S^2/(1+N_o \cdot \xi^{-1})$, exactly as for KMA. However, we must bear in mind that, contrarily to KMA, the latter bound is obviously not achievable due to the sign ambiguity in the estimate of $\mathbf{S}$ (recall Sect. III-B). Hence, the best estimate possible (for $N_o \to \infty$) is $\hat{\mathbf{S}} = \pm\mathbf{S}$ with probability $1/2$ each, which leads to an error variance $2\sigma_S^2$, the minimum achievable without knowledge of the embedded messages.

B. Bound on the normalized correlation

In spread spectrum methods using binary antipodal constellations, exact knowledge of $\mathbf{s}$ is not necessary for performing correct decoding. Usually, decoding is implemented by means of a cross-correlation operation, estimating the message embedded in $\mathbf{y}_i$ as $\hat{m}_i = \text{sign}\{\mathbf{y}_i^T \cdot \mathbf{s}\}$. Also, in watermark detection applications based on spread spectrum, the detector decides on the presence of the watermark upon the angle between $\mathbf{y}_i$ and $\mathbf{s}$. In other words, the norm of $\mathbf{s}$ is important for the embedding operation (e.g. for controlling the embedding distortion), but not for detection/decoding. This implies that the attacker is mainly interested in disclosing the direction of $\mathbf{s}$, which spans the subspace where the watermark is contained. Thus, it is useful to quantify the difficulty in estimating the direction of $\mathbf{s}$. The natural performance measure is the normalized correlation, defined as

$$\rho \triangleq \frac{\hat{\mathbf{s}}^T \mathbf{s}}{||\hat{\mathbf{s}}|| \cdot ||\mathbf{s}||} = \cos(\phi) \in [-1, 1], \tag{25}$$

where ϕ denotes the angle between $\mathbf{s}$ and $\hat{\mathbf{s}}$. The closer to 1 is the value of ρ, the more accurate is the estimate of $\mathbf{s}$. The relation between ρ and the Cramér-Rao bound on the estimation error of $\mathbf{S}$ has been pointed out in [3, Sect. V], although not in deep detail. Here we pursue a bound on ρ using the equivocation.

Notice that the vector $\mathbf{s} \in \mathbb{R}^n$ can be expressed by means of its norm and an n-dimensional unit vector colinear to $\mathbf{s}$, i.e. we consider the transformation $\mathbf{s} \rightarrow (q, \mathbf{r})$, with $q = ||\mathbf{s}||$ and $\mathbf{r}$ a unit vector in the direction of $\mathbf{s}$. Thus, we have a coordinate change $\mathbb{R}^n \rightarrow \mathbb{R}^+ \times \mathbb{R}^n$.

Lemma 1: For an unbiased estimator, the mean value of the normalized correlation can be bounded from above as

$$E\left[\cos(\Phi)\right] \leq 1 - \frac{n}{4\pi e}\exp\left(\frac{2}{n}h(\mathbf{R}|\mathbf{O}_1, \ldots, \mathbf{O}_{N_o})\right), \tag{26}$$

where $h(\mathbf{R}|\mathbf{O}_1, \ldots, \mathbf{O}_{N_o})$ represents the equivocation about $\mathbf{R}$, given N_o observations.

Proof: Let us define the estimation error as $\mathbf{d} \triangleq \mathbf{r} - \hat{\mathbf{r}}$. From Eq. (24), for an unbiased estimator we have

$$\frac{E\left[||\mathbf{D}||^2\right]}{n} = \frac{\text{tr}(\mathbf{\Sigma}_D)}{n} \geq \frac{1}{2\pi e}\exp\left(\frac{2}{n}h(\mathbf{R}|\mathbf{O}_1, \ldots, \mathbf{O}_{N_o})\right). \tag{27}$$

By the cosine theorem, we have that $||\mathbf{d}||^2 = 2(1 - \cos(\phi))$, with ϕ the angle between $\mathbf{r}$ and $\hat{\mathbf{r}}$. Combining this with (27), we arrive at (26). ■

Lemma 1 relates the normalized correlation with the equivocation about $\mathbf{R}$. The a priori equivocation, $h(\mathbf{R})$, achieves its maximum when $\mathbf{R}$ is uniformly distributed over the surface of the unit-radius hypersphere. Note that this is the case when $\mathbf{S}$ is i.i.d. Gaussian, as we are assuming in this paper. We are concerned now with the equivocation about $\mathbf{R}$. Using the coordinate change $\mathbf{s} \rightarrow (q, \mathbf{r})$ introduced above, the differential entropy of $\mathbf{S}$ can be rewritten as

$$h(\mathbf{S}) = h(Q, \mathbf{R}) + E[\log(J)], \tag{28}$$

where J denotes the Jacobian of the coordinate change and the expectation is taken over $\mathbf{S}$. This change of coordinates can be seen as a QR factorization [15], $\mathbf{s} = q \cdot \mathbf{r}$, with $q \in \mathbb{R}$, $\mathbf{r} \in \mathbb{R}^n$, and $||\mathbf{r}|| = 1$. The Jacobian of this QR factorization is given by $J = Q^{n-1}$ [16]. Hence, using this result and (28), we can lowerbound the equivocation on the embedding direction as

$$h(\mathbf{R}|\mathbf{O}_1, \ldots, \mathbf{O}_{N_o}) \geq h(\mathbf{S}|\mathbf{O}_1, \ldots, \mathbf{O}_{N_o}) - h(Q|\mathbf{O}_1, \ldots, \mathbf{O}_{N_o}) - (n-1)E\left[E[\log(Q)|\mathbf{O}_1 = \mathbf{o}_1, \ldots, \mathbf{O}_{N_o} = \mathbf{o}_{N_o}]\right], \tag{29}$$

where the inner expectation is taken over Q, and the outer expectation is over the observations. Equality in (29) is achieved when the norm and direction of $\mathbf{S}$ are mutually independent. Using (29), we will specialize the bound of Lemma 1 to add-SS in the KMA scenario.

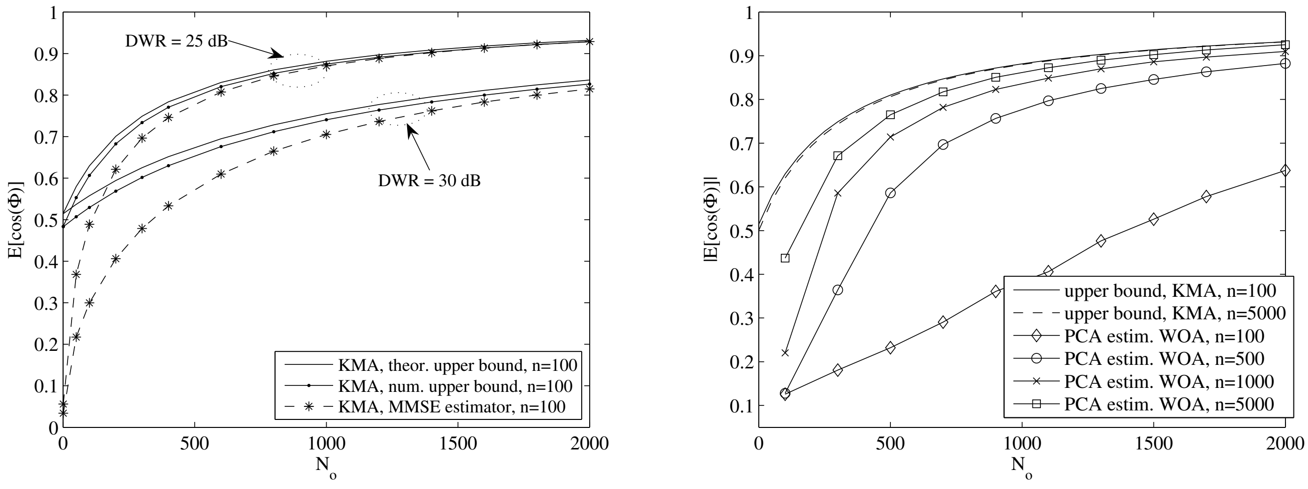
Lemma 2: For add-SS in the KMA scenario, the equivocation about $\mathbf{R}$ can be bounded from below as

$$\begin{aligned} h(\mathbf{R}|\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}, M_1, \ldots, M_{N_o}) \quad &\geq \quad \frac{n-1}{2}\log\left(\frac{2\pi e}{n\sigma_S^2}\right) + \frac{n}{2}\log\left(\frac{\sigma_S^2}{1 + N_o\xi^{-1}}\right) \\ -\frac{1}{2}\log\left(n\sigma_S^2 - \frac{2\sigma_S^2}{1 + N_o\xi^{-1}}\left(\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}\right)^2 {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -\frac{nN_o}{2}\xi^{-1}\right)^2\right)&, \end{aligned} \tag{30}$$

where ${}_1F_1$ and Γ denote the confluent hypergeometric function of the first kind and the complete Gamma function, respectively [17].

Proof: The first term in the right hand side of (29) is given by (5). The remaining terms, related to the norm of the spreading vector, are upper bounded in Appendix E. The combination of these results yields (30). ■

We have represented in Fig. 5(a) the upper bound on the normalized correlation for add-SS resulting from the insertion of (30) in (26). In order to support the tightness of the bounds derived in Appendix E, the theoretical bound



(a)

(b)

Fig. 5. Upper bound to the normalized correlation for add-SS. Comparison with practical estimators for KMA with $n = 100$ (a) and WOA with DWR=25 dB (b).

is compared to the result using numerical integration in the two rightmost terms of (29). Moreover, for checking the tightness of the bound on $E[\cos(\Phi)]$, we have computed numerically (by Monte Carlo) the average normalized correlation resulting from applying the MMSE estimator of $\mathbf{S}$. As can be seen, the bound is loose for small N_o, but it becomes tight as N_o is increased. The reason is that the distributions of $\mathbf{R}$ and Q conditioned on the observations become approximately independent when N_o is increased and because the boundings based on Jensen's inequality are also asymptotically tight when the variance of the considered random variable approaches 0.

As seen in Section III-B, when the embedding rate is small enough, the information that the WOA scenario provides about $\mathbf{S}$ approaches that of KMA except for the sign ambiguity. This ambiguity also arises when evaluating a practical estimator, but we can get rid of it simply by taking the absolute value of ρ as performance measure. Notice that, in this case, we would be evaluating the accuracy in the estimation of the subspace spanned by $\mathbf{s}$. This performance evaluation is illustrated in Fig. 5(b), showing that the accuracy of the subspace estimation in the WOA case tends (as expected) to that of the KMA as n is increased. The curves for WOA were obtained empirically with the "PCA estimator", that will be introduced in Section VI.

REMARKS: Notice that the theoretical bound in Fig. 5(b) remains approximately invariant with n. Moreover, it can be checked that it is approximately independent of the specific values of σ_S or σ_X, depending only on ξ. Thus, it looks more appealing than the MSE bound for evaluating the security level.

VI. PRACTICAL ESTIMATORS OF THE SPREADING VECTOR

After the theoretical analysis carried out in the previous sections, we are interested now in evaluating the security from a practical point of view. We will focus on the ISS embedding function, since the attacks devised for it are applicable to add-SS and γ-SS as well. The purpose of this section is twofold: on one hand, we will analyze the approaches previously proposed for tackling the spreading vector estimation problem, highlighting their limitations; on the other hand, new estimators for the WOA scenario will be proposed and analyzed. For the analysis, the host will be assumed to be i.i.d. Gaussian again. This analysis is performed under asymptotic conditions ($N_o \rightarrow \infty$), in order to show the fundamental limitations of each method. The comparison between the different estimators on a practical scenario will be deferred until Section VII.

Hereinafter, for the sake of clarity, the spreading vector used by the watermarker and which the attacker wants to estimate will be denoted by $\mathbf{s}_0$. We will assume, for simplicity, that $\mathbf{s}_0$ is of unit norm. Hence, the embedding function that we consider is

$$\mathbf{y}_i = \mathbf{x}_i + (-1)^{m_i}\nu\mathbf{s}_0 - \lambda(\mathbf{x}_i^T\mathbf{s}_0)\mathbf{s}_0, \tag{31}$$

which is equivalent to (15) if we use $\nu = \left(n\sigma_S^2 - \lambda^2\sigma_X^2\right)^{\frac{1}{2}}$. Furthermore, by assuming $\mathbf{s}_0$ to be of unit norm, the comparison of the different estimators considered is straightforward, since all of them have been devised for estimating unit norm vectors.

We will focus on the WOA scenario, which is the most interesting in practice (for work on estimators for the KMA scenario, see [3] for add-SS, and [11],[18] for ISS). As explained in Section III-B, in this scenario there is an inherent sign ambiguity that cannot be removed when estimating $\mathbf{s}_0$. In addition, estimation of the embedding direction is usually enough for the attacker's purposes, as discussed in Section V-B. For these reasons, we will consider that the objective of the attacker is to estimate the one-dimensional subspace spanned by $\mathbf{s}_0$, and the estimator's performance will be measured in terms of the absolute value of the normalized correlation, $|\rho| = \frac{|\hat{\mathbf{s}}_0^T\mathbf{s}_0|}{||\hat{\mathbf{s}}_0||\cdot||\mathbf{s}_0||}$.

A. Previous approaches for the WOA scenario

We analyze the estimation setup proposed in [3], based on Independent Component Analysis (ICA) and Principal Component Analysis (PCA). ICA and PCA are well known statistical tools for performing blind source separation (BSS) [19]. PCA was applied for the first time to the watermarking security problem in [20] for estimating the embedding subspace for spread spectrum modulations in the WOA scenario. This approach was later refined in [3] by means of a two-step procedure: first, PCA is applied for identifying the embedding subspace, and later ICA is applied in that subspace in order to completely recover the secret carriers. For the application of ICA, they resort to the FastICA algorithm [21]. The main difference between the practical setup of [3] and ours is that the authors of [3] consider a multibit embedding function, where several secret carriers can be embedded at once. However, the estimation of N_b carriers by means of the FastICA method is performed through the application of FastICA to N_b one-bit problems and proper orthogonalization after each iteration. Hence, the performance of FastICA is in last instance determined by that of the one-bit estimator. For these reasons, we focus our analysis below on the one-bit setup (the multibit setup with several secret carriers is briefly considered in Section VII-A). For the technical details of the analysis that are omitted here, the reader is referred to [18, Chap. IV].

1) Principal Component Analysis (PCA) : The use of PCA in the context of spread spectrum security is based on the following rationale: for large spreading sequences, the variance in the direction of the watermark dominates over the remaining directions, so for carrier estimation (assuming multiple carriers) it suffices to keep for the ICA only the subspace defined by the eigenvectors with the largest associated eigenvalues. Moreover, if there is only one secret carrier, PCA by itself can be successful, so further application of ICA is not necessary. This reasoning holds in many practical scenarios, especially in watermark detection applications, where n uses to be in the order of thousands.[2] However, when considering data hiding applications, the situation may be different. Let us denote by $\mathbf{Q}$ the covariance matrix of the observations. For $N_o \to \infty$, an eigenvalue decomposition yields $\mathbf{Q} = \mathbf{V}\mathbf{D}\mathbf{V}^T$, with

$$\begin{aligned} \mathbf{V} &= [\mathbf{s}_0, \mathbf{V}_{\mathbf{s}_0}] \in \mathbb{R}^{n\times n}, \\ \mathbf{D} &= \begin{bmatrix} \nu^2 + (1-\lambda)^2\sigma_X^2 & 0 \\ 0 & \sigma_X^2 \cdot \mathbf{I}_{n-1} \end{bmatrix}, \end{aligned} \tag{32}$$

where $\mathbf{V}_{\mathbf{s}_0} \in \mathbb{R}^{n\times n-1}$ is a unitary matrix whose columns span the orthogonal complement of the subspace spanned by $\mathbf{s}_0$. When a single carrier is being used, the estimator of $\mathbf{s}_0$ by PCA is simply given by [3]

$$\hat{\mathbf{s}}_0 = \mathbf{V}[\arg\max_i D_{i,i}], \tag{33}$$

where $\mathbf{V}[k]$ denotes the kth column of the matrix $\mathbf{V}$, and $D_{i,i}$ is the ith element in the diagonal of the matrix $\mathbf{D}$, both defined in (32). The performance of this estimator, which will be referred to as the “PCA estimator”, was already plotted for add-SS (i.e. with $\lambda = 0$) in Fig. 5(b), where we can see that it works remarkably well, especially as n is increased. In order to perform a correct estimate, the variance in the direction of $\mathbf{s}_0$ must be larger than in the remaining directions. For the i.i.d. Gaussian host, this is equivalent to

$$\nu^2 + (1-\lambda)^2\sigma_X^2 > \sigma_X^2 \Leftrightarrow \text{DWR} < 10\log_{10}\left(\frac{n}{2\lambda}\right) \tag{34}$$

as can be seen from (32). If the pdf of the host signal is not circular (i.e. with non-diagonal covariance matrix), then the condition for ensuring correct estimation is more restrictive, as one must take into account the directions with the highest variance. In any case, the watermarker could easily fool the PCA estimator by properly tuning the parameters n and λ so as to reduce the variance in the embedding direction (possibly loosing some robustness).

2) Independent Component Analysis (ICA): In BSS, the idea behind ICA methods is to optimize a cost function that measures the mutual independence between the separated sources. Maximization of independence is equivalent to maximization of the squared “negentropy”, which can be intuitively interpreted as the divergence from gaussianity [22]. ICA is not restricted to the BSS paradigm, but it is often used as a tool for extracting “interesting” components of high-dimensional data, which is precisely the target pursued here. Indeed, under this point of view, ICA can be seen as a way of performing Projection Pursuit [23] rather than BSS. Among the variety of ICA tools existing in the literature, we will focus on the approach used in [3] and [24], where the ICA cost function is defined as [21]

$$J_{\text{ICA}}(\mathbf{s}) = \left(E\left[g(\mathbf{Y}^T\mathbf{s})\right] - E\left[g(U)\right]\right)^2, \tag{35}$$

with $U \sim \mathcal{N}(0, \text{var}(\mathbf{Y}^T\mathbf{s}))$, and $g(\cdot)$ is the so-called “contrast function”. The ICA estimator results in

$$\hat{\mathbf{s}}_0 = \arg\max_{\mathbf{s}} J_{\text{ICA}}(\mathbf{s}). \tag{36}$$

Intuitively, (36) looks for the unit-norm vector $\mathbf{s}$ that maximizes the divergence between the distribution of $\mathbf{Y}^T\mathbf{s}$ and a Gaussian distribution with the same variance. The optimal choice of the contrast function is $g(z) = \log(f(z))$, where $f(z)$ is the pdf of the independent component to be estimated (the optimality is in the sense of asymptotic variance of

[2] This is, for instance, the practical scenario considered by Cayre et al. [3].

the estimation error [25]). For an i.i.d. Gaussian host, the pdf of this component is $f(z) = K \cdot \exp\left(-\frac{z^2}{2\sigma_Z^2}\right)\cosh\left(\frac{z\nu}{\sigma_Z^2}\right)$, where K is a constant and $\sigma_Z^2 = (1-\lambda)^2\sigma_X^2$. Hence, making the change of variable $a = \nu/\sigma_Z^2$, the optimal ICA cost function for our problem results in [18, Chap. IV]

$$J_{\text{ICA}}^{blind}(\mathbf{s}, a) = \left(E\left[\log\cosh(a \cdot \mathbf{Y}^T\mathbf{s})\right] - E\left[\log\cosh(a \cdot U)\right]\right)^2, \tag{37}$$

where the parameter a can be fixed by the attacker. The notation "*blind*" is adopted for emphasizing the fact that, as in [3] and [24], the cost function does not depend on any parameter to be estimated from the observed data. It is interesting to note that the choice of the $\log\cosh$ contrast function coincides with the recommendation in [21] for general purpose ICA, although the reasoning for arriving there is essentially different. Assuming an i.i.d. Gaussian host, the term $\mathbf{Y}^T\mathbf{s}$ in (37) is a binary Gaussian mixture, according to [18, Chap. IV]

$$\mathbf{Y}^T\mathbf{s} \sim \frac{1}{2}\left(\mathcal{N}(\nu\rho, \sigma_X^2||\mathbf{t}||^2) + \mathcal{N}(-\nu\rho, \sigma_X^2||\mathbf{t}||^2)\right), \text{ with } ||\mathbf{t}||^2 = ||\mathbf{s} - \lambda\rho\mathbf{s}_0||^2 = 1 + \rho^2(\lambda^2 - 2\lambda). \tag{38}$$

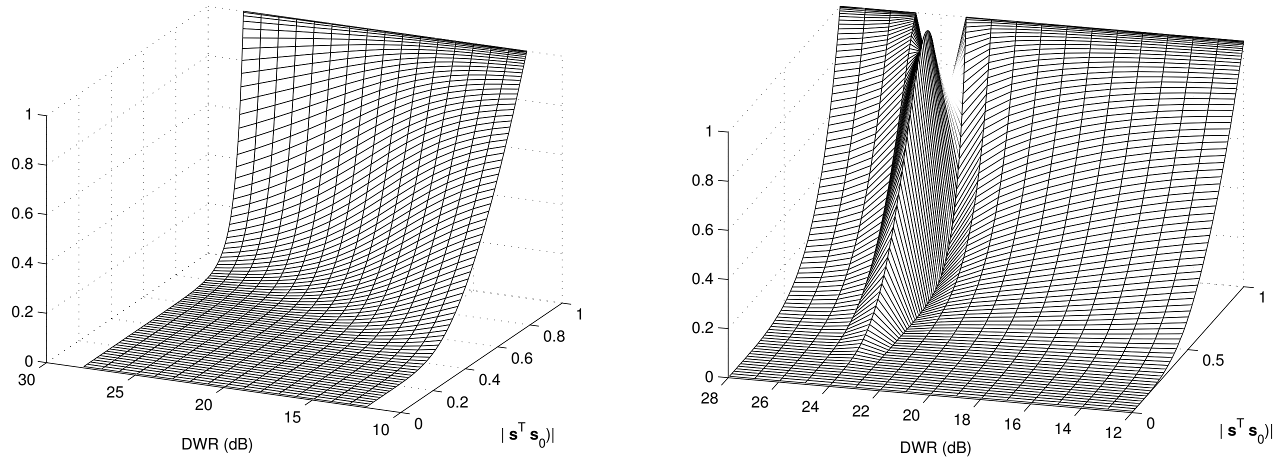
The expectations in (37) have no closed-form expression, so their evaluation requires numerical integration. Notice that the cost function depends solely on the value of ρ, not on the particular realization of $\mathbf{s}_0$. Fig. 6(a) shows the cost function versus $|\rho|$ and the DWR for $a = 1$ and $\lambda = 0.5$. For each DWR, the plots have been normalized by the maximum value of the cost function for ease of comparison. This normalized cost function has shown to be virtually insensitive to the chosen value of a. Although $J_{\text{ICA}}^{blind}(\mathbf{s}, a)$ is convex and its maximum is clearly located at $|\rho| = 1$ as desired, this cost function is not well suited to practical applications where the DWR is moderately high, because of its remarkable flatness for small ρ. Due to this flatness, the initial vector must be very close to $\mathbf{s}_0$ for assuring convergence when the cost function is to be optimized iteratively. Indeed, in real experiments the ICA estimator has shown to get stuck most of the times at $\rho \approx 0$, a fact that is probably due to the noise in the estimation of the cost function with a finite number of samples.

As can be guessed from Fig. 6(a), $J_{\text{ICA}}^{blind}(\mathbf{s}, a)$ becomes a suitable cost function when the DWR is small. Hence, a possible strategy is to apply PCA in a first step for reducing the dimensionality by keeping only the subspace of largest variance. This would reduce the effective DWR, helping the blind ICA estimator. Nevertheless, an important drawback of this approach (specially if the host is not white) is that the attacker does not know a priori the dimensionality of the subspace that contains the spreading vector. Furthermore, if PCA is fooled, the subsequent ICA step is automatically fooled, because it would perform the search for the spreading vector in the wrong subspace.

The reader must notice that most ICA estimators (including FastICA [21]) work with whitened observations. However, the whitening operation is not inherent to ICA nor strictly necessary. In fact, it is not considered in the ICA cost function originally proposed by Hyvärynen [25]. The interested reader is referred to [18, Sect. VI-B], where the ICA cost function with whitening has been evaluated. The results therein point out that whitening does not add any improvement to ICA; instead, it is even harmful because it leads to further flattening of the cost function.

B. New estimators for the WOA scenario

The main objective of this section is to develop new estimators that work in scenarios where blind ICA and PCA fail, constituting this way a wider battery of methods for performing practical security tests. The interested reader



(a) $J_{\text{ICA}}^{blind}(\mathbf{s}, a)$ with $a = 1$

(b) $J_{\text{ICA}}^{inf}(\mathbf{s}, a)$ with $a = 1$

Fig. 6. Cost surface of blind and informed ICA, for $n = 200$ and $\lambda = 0.5$.

can find in [18, Chap. IV] some technical details that are missing here due to the lack of space, along with another estimator termed "Approximate Maximum Likelihood" (AML).

1) Informed ICA : We have empirically observed that the drawbacks mentioned in Sect. VI-A2 for the blind ICA estimator can be partially overcome if the variance of the random variable U in $J_{\text{ICA}}^{blind}(\mathbf{s}, a)$ is fixed to a proper constant value, instead of varying it according to var$(\mathbf{Y}^T\mathbf{s})$. We term the new cost function $J_{\text{ICA}}^{inf}(\mathbf{s}, a)$, where "$inf$" stands for "informed". After some experiments, it was found that fixing that value to σ_X^2 yields good results. In order to keep the estimator blind, we compute an estimate of σ_X from the observations. Our estimate of σ_X is

$$\hat{\sigma}_X = \left(\frac{1}{n}\text{tr}\left(\mathbf{Q}\right)\right)^{\frac{1}{2}} = \left(\frac{1}{n}\sum_{i=1}^{n} D_{i,i}\right)^{\frac{1}{2}}, \tag{39}$$

where tr$(\mathbf{Q})$ is the covariance matrix of the observations, and $D_{i,i}$ are the diagonal elements of the matrix of eigenvalues defined in (32). The expression of $J_{\text{ICA}}^{inf}(\mathbf{s}, a)$ is still given by (35), with the only difference that $U \sim \mathcal{N}(0, \hat{\sigma}_X^2)$. Fig. 6(b) shows the cost surface of $J_{\text{ICA}}^{inf}(\mathbf{s}, a)$ (obtained again by means of numerical integration) under the same conditions as $J_{\text{ICA}}^{blind}(\mathbf{s}, a)$. As can be seen, $J_{\text{ICA}}^{inf}(\mathbf{s}, a)$ is convex for all DWRs except for a small range, and no problems of flatness for small ρ appear, so an iterative optimization algorithm can easily reach the maximum. Although $J_{\text{ICA}}^{inf}(\mathbf{s}, a)$ has shown to be quite sensitive to value of the variance we fix for U, it performs reasonably well with real images (see next section). Similarly to $J_{\text{ICA}}^{blind}(\mathbf{s}, a)$, the value of a has no noticeable influence on the normalized cost function.

2) Constant Modulus (CM) criterion : According to (38), when $\mathbf{Y}$ is correlated with a vector $\mathbf{s}$ orthogonal to $\mathbf{s}_0$ (i.e. with $\rho = 0$), the resulting random variable is zero-mean Gaussian with variance σ_X^2. However, when $\mathbf{Y}$ is correlated with $\mathbf{s}_0$, we have $\mathbf{Y}^T\mathbf{s}_0 \sim \frac{1}{2}\left(\mathcal{N}(\nu, (1-\lambda)^2\sigma_X^2) + \mathcal{N}(-\nu, (1-\lambda)^2\sigma_X^2)\right)$. Hence, if we take the modulus of $\mathbf{Y}^T\mathbf{s}$, it should lie (in average) closer to ν as $\mathbf{s}$ approximates $\mathbf{s}_0$. The main idea behind the CM method is to define a cost function that penalizes the deviations from ν. A possible cost function is then

$$J_{\text{CM}}(\mathbf{s}) = E\left[\left((\mathbf{Y}^T\mathbf{s})^2 - \nu^2\right)^2\right] = E\left[(\mathbf{Y}^T\mathbf{s})^4\right] - 2\nu^2 \cdot E\left[(\mathbf{Y}^T\mathbf{s})^2\right] + \nu^4. \tag{40}$$

The validity of this approach for performing estimation of $\mathbf{s}_0$ is now discussed. In general, the attacker has no a priori information about the embedding parameter ν, so he needs an estimate before applying the CM method. We assume now that the attacker manages to obtain an estimate $\hat{\nu} = \nu + \tilde{\nu}$ from the observations at hand. The new CM cost function, now termed "blind", is

$$J_{\text{CM}}^{blind}(\mathbf{s}, \hat{\nu}) = E\left[(\mathbf{Y}^T\mathbf{s})^4\right] - 2\hat{\nu}^2 \cdot E\left[(\mathbf{Y}^T\mathbf{s})^2\right] + \hat{\nu}^4 = J_{\text{CM}}(\mathbf{s}) - 2E\left[(\mathbf{Y}^T\mathbf{s})^2\right]\left(\tilde{\nu}^2 + 2\nu\tilde{\nu}\right) + \hat{\nu}^4 - \nu^4, \qquad (41)$$

and the blind CM estimator is defined as

$$\hat{\mathbf{s}}_0 = \arg\min_{\mathbf{s}} J_{\text{CM}}^{blind}(\mathbf{s}, \hat{\nu}). \qquad (42)$$

By definition, the CM estimator is essentially equivalent to the methods for blind equalization based on the constant modulus criterion which are well known in the literature of Digital Communications [26] and have been extensively studied there. Nevertheless, the problem setup in our case is different, so a new analysis of the CM cost function for our problem is justified. The fourth and second order moments involved in (41) have been computed in [18, Chap. IV], finding that the blind CM cost function results in a fourth degree polynomial in the normalized correlation,

$$\begin{aligned} J_{\text{CM}}^{blind}(\mathbf{s}, \hat{\nu}) = J_{\text{CM}}^{blind}(\rho, \hat{\nu}) \;&=\; \rho^4\left(\nu^4 - 12\nu^2\lambda\sigma_X^2 + 6\nu^2\lambda^2\sigma_X^2 + 12\lambda^2\sigma_X^4 - 12\lambda^3\sigma_X^4 + 3\lambda^4\sigma_X^4\right) \\ &\quad -2\rho^2\left(\nu^2 - 3\sigma_X^2 + 2\nu\tilde{\nu} + \tilde{\nu}^2\right)\left(\nu^2 - 2\lambda\sigma_X^2 + \lambda^2\sigma_X^2\right) \\ &\quad +3\sigma_X^4 - 2\nu^2\sigma_X^2 + \nu^4 + 4\nu^3\tilde{\nu} + 6\nu^2\tilde{\nu}^2 + 4\nu\tilde{\nu}^3 + \tilde{\nu}^4 - 4\nu\tilde{\nu}\sigma_X^2 - 2\tilde{\nu}^2\sigma_X^2. \qquad (43) \end{aligned}$$

In the particular case of $\lambda = 0$ (add-SS) and $\tilde{\nu} = 0$ (perfect estimate of ν), Eq. (43) admits a clearer expression:

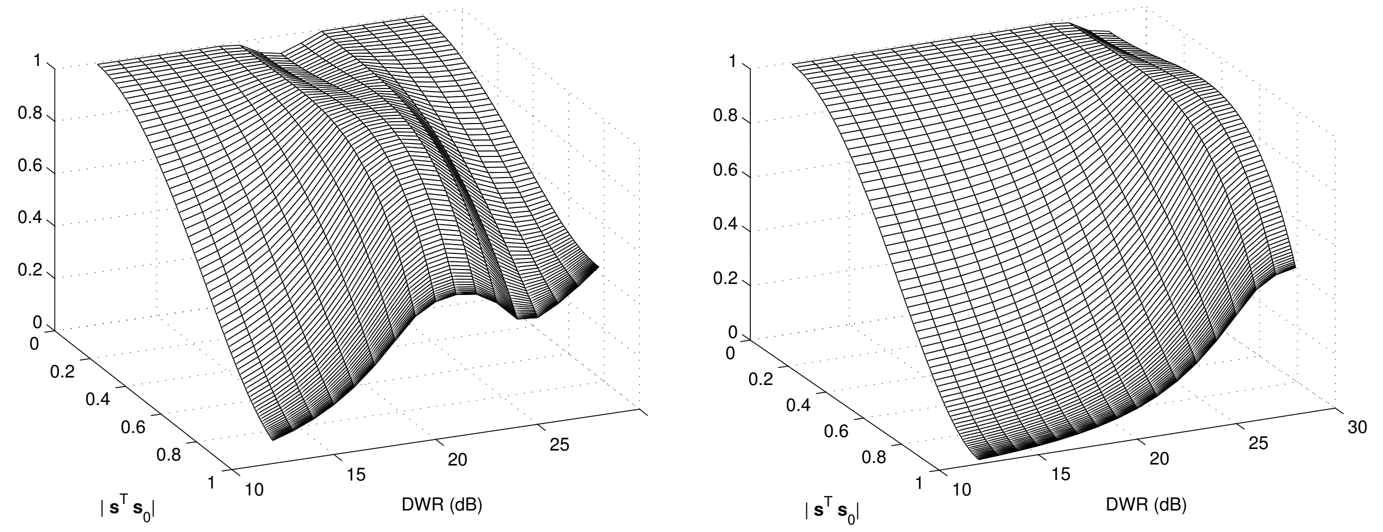
$$J_{\text{CM}}^{blind}(\mathbf{s}, \hat{\nu})\,|_{\lambda=0,\ \hat{\nu}=\nu} = \rho^4\nu^4 - 2\rho^2\nu^2(\nu^2 - 3\sigma_X^2) + 3\sigma_X^4 - 2\nu^2\sigma_X^2 + \nu^4. \qquad (44)$$

Now let us denote by ρ_{min} the value of ρ for which the global minimum of (44) is achieved. From the attacker's point of view, it is desirable that $|\rho_{min}|$ is as close to 1 as possible. For $\nu \leq \sqrt{3}\sigma_X$, the minimum of (44) is always achieved for $\rho_{min} = 0$. In turn, for $\nu > \sqrt{3}\sigma_X$, if the attacker wants to achieve $|\rho_{min}| \geq \tau \in [0, 1]$, then the condition $\text{DWR} < 10\log_{10}\left(\frac{n(1-\tau^2)}{3}\right)$ must hold. This means that for achieving $|\rho_{min}| \geq 0.9$ the DWR must be below 1 dB and 18 dB for $n = 20$ and $n = 1000$, respectively. Bear in mind that, when optimizing iteratively the CM cost function, the condition derived above is necessary but not sufficient, in general, for arriving at the optimum that guarantees $|\rho_{min}| > \tau$, since (44) is not necessarily monotonically decreasing for $|\rho| \in [0, 1]$.

For $\lambda > 0$ it is hard to infer from (43) the conditions for the successful estimate of $\mathbf{s}_0$, so we will plot the cost function for different embedding parameters. First of all, we propose to estimate $\hat{\nu}$ as

$$\hat{\nu}_{blind} \triangleq \max_i \left\{(D_{i,i})^{\frac{1}{2}}\right\}, \qquad (45)$$

where $D_{i,i}$ are the diagonal elements of $\mathbf{D}$, defined in (32). This choice is justified by the fact, observed in [18, Chap. IV], that underestimating the true value of ν may have a harmful effect in the cost function, but using $\hat{\nu} > \nu$ can even be beneficial for the attacker (notice that $\hat{\nu}_{blind} \geq \nu$). The resulting cost surface of the blind CM estimator is shown in Fig. 7 for $\lambda = 0.5$. Notice that for $n = 200$ there exists a small range of DWRs (around 20 dB) with a local minimum close to $|\rho| = 0$. For those DWRs, an iterative optimization of the cost function would get stuck at $|\rho| \approx 0$ with high probability. However, for the remaining DWRs, the shape of the cost function is very appealing. Increasing the parameter n (Fig. 7(b)) has the effect of increasing the range of DWRs where the function is unimodal (i.e. with no local minima) and achieves its minimum for $|\rho| = 1$. As can be seen, the range of DWRs with undesired local minima is moved towards higher DWRs.



(a) $n = 200$

(b) $n = 1000$

Fig. 7. Cost surface of the blind CM method for ISS with $\lambda = 0.5$.

C. Final remarks

Although the analysis carried out here for the ICA and blind CM cost functions was made considering Gaussian hosts, the conclusions can be extended to more general host distributions under some mild assumptions. All the cost functions are based on functionals of the type $\mathbf{Y}^T\mathbf{s}$. Hence, if the components of the observations are approximately independent, the resulting random variable can be well approximated for a wide variety of host distributions by a Gaussian whenever n is sufficiently large, by virtue of the Central Limit Theorem. This is true, for instance, when the embedding is made on the DCT or DWT domains, where the coefficients are distributed according to a Generalized Gaussian. This latter case is explicitly analyzed in [18, Chap. IV] for the blind CM cost function.

Further remarks come from the links between the theoretical security analysis and the behavior of the estimators:

1) From Section III-B, it is known that large spreading sequences provide more information to the attacker than short ones. Thus, it should be easier for the attacker to perform estimation of $\mathbf{s}_0$ as n is increased. This is the case for the PCA estimator, as discussed in Section VI-A1, and also for the blind CM estimator, under the light of Fig. 7.

2) From Section IV-B, it is known that increasing λ, i.e. increasing the host rejection, provides more information about $\mathbf{s}_0$. This additional information is effectively translated into an advantage for the blind CM estimator, which is reflected in its cost function [18, Chap. IV]. However, the PCA estimator works better for $\lambda = 0$, according to (34).

VII. Experimental results

In practice, the expectations and covariance matrices needed for the estimators proposed in Section VI are replaced by the corresponding sample estimates. Furthermore, the ICA and CM estimators do not admit closed form expressions, so one has to resort to iterative optimization algorithms. From the wide variety of algorithms available in the literature, we chose a family of optimization methods that permits us to easily extend the proposed approach to other related watermarking scenarios, as we will see in Section VII-A. In any case, the rigorous analysis of these optimization methods is out of the scope of the present paper.

Given the geometrical structure of the problem we are considering, it is natural to think of tools for optimization on curved spaces or “manifolds” that exploit unitary constraints, such as Grassman and Stiefel manifolds. The advantage of optimizing directly on these manifolds is that one naturally gets rid of the unit-norm constraint (applied on the norm of $\mathbf{s}_0$) in our optimization problems. In the case of a single vector estimation, the Stiefeld manifold becomes simply the surface of the unit sphere in $\mathbb{R}^n$, and the Grassman manifold is the set of one-dimensional spaces that can be spanned by a n-dimensional vector. The interested reader is referred to [27] for an introductory explanation on these manifolds, or to [18, Chap. IV] for the most basic definitions, relevant to our problem. For our practical implementations we chose a conjugate gradient method over the Grassman manifold, whose complete description can be found in [27].

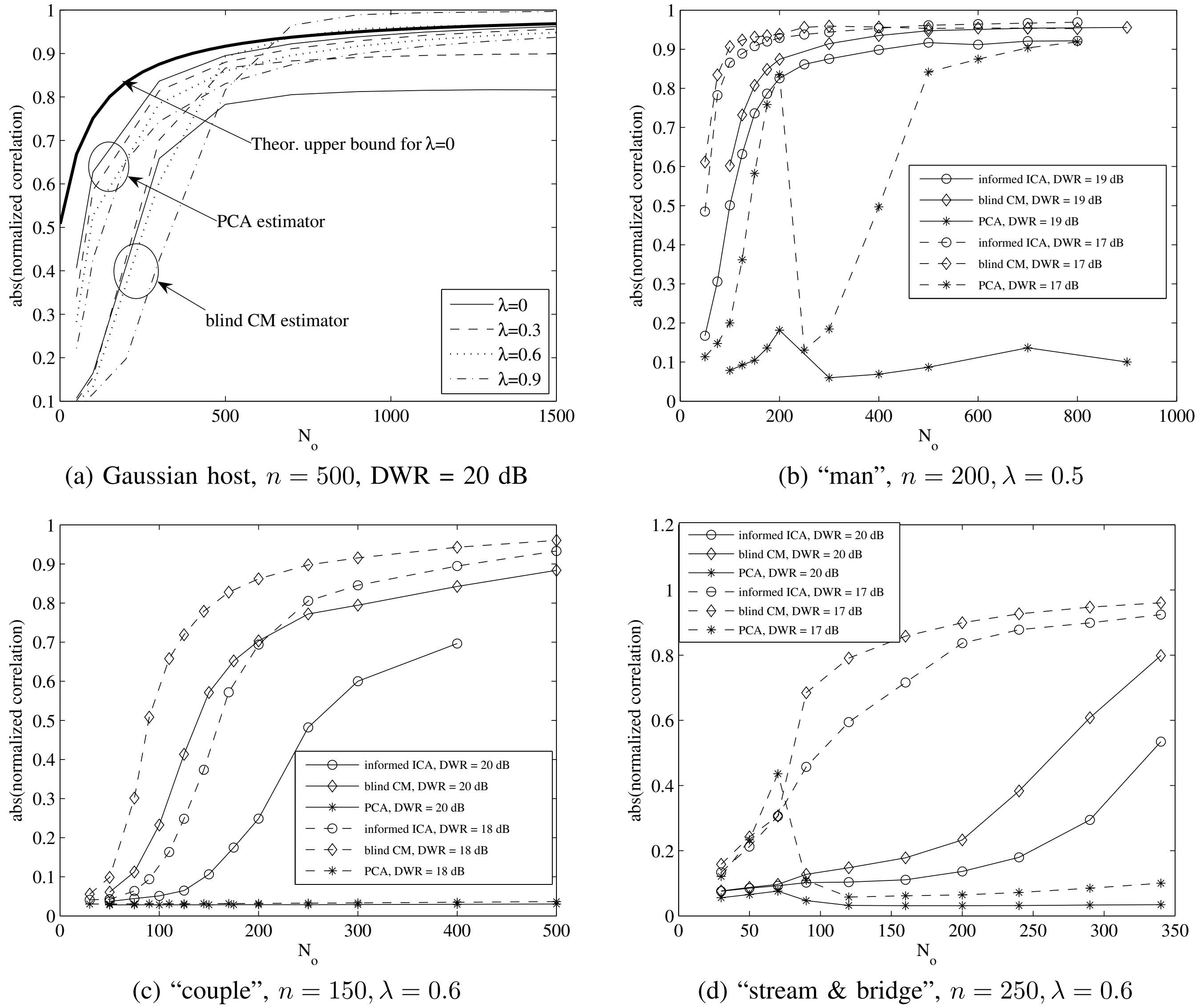
As explained in Section VI-A2, the blind ICA approach fails when directly applied to our estimation problem. This is why we focus here on the “blind CM” and “informed ICA” estimators, that will be compared with the PCA estimator. Fig. 8(a) shows the comparison between the PCA and blind CM estimators for i.i.d. Gaussian host and different values of λ. As can be seen, the PCA estimator achieves its best performance for $\lambda = 0$, as explained in the previous section, and it is below the theoretical upper bound (derived in Section V-B) in all cases. As for the blind CM estimator, it can be seen that it clearly benefits from increasing λ. In general, PCA presents better performance than blind CM for small N_o, but this is not necessarily true as N_o is increased. Fig. 8(b)-(c)-(d) show results obtained for real images: “man”, “couple”, and “stream & bridge”, available for download in [28]. All of them have been watermarked in the DCT domain, using a subset of coefficients that is assumed to be known by the attacker, and which is the input to the estimators (thus, depending on the size of the image and the value of n, the maximum allowed N_o is different in each case). This scenario is less favorable for PCA than with the i.i.d. Gaussian host, because the embedding direction is not guaranteed to have the highest variance. As can be seen, the blind CM estimator offers in all cases better performance than the informed ICA. It is also interesting to see that the PCA estimator gets trapped in the directions of maximum variance that, in general, are far from the direction of the spreading vector. Furthermore, it can be observed that for PCA the estimation accuracy may decrease (see Fig. 8(b), 8(d)) when the number of observations is increased. This is due to the fact that the host vectors are not i.i.d., and consequently the directions of maximum variance depend on the considered region of the image.

A. Extension to other scenarios

1) Estimation of several carriers (multibit data hiding): We consider here the case where several bits of information are embedded in the same host signal by means of several carriers, a setup that could correspond to scenarios of multiuser information embedding, for instance. The embedding function for a multibit ISS system with embedding rate $R = \frac{1}{n}\log(N_b)$ can be written as

$$\mathbf{Y}_i = \mathbf{X}_i + \nu \sum_{k=1}^{N_b} (-1)^{M_i} \mathbf{s}_k - \lambda \sum_{k=1}^{N_b} \frac{\mathbf{X}_i^T \mathbf{s}_k}{||\mathbf{s}_k||^2} \mathbf{s}_k, \text{ for } i = 1, \ldots, N_o, \tag{46}$$

where for simplicity we have considered that the parameters ν and λ are the same for all carriers. Hence, the secrecy of the system relies on the matrix of secret carriers $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_{N_b}] \in \mathbb{R}^{n \times N_b}$.



(a) Gaussian host, $n = 500$, DWR = 20 dB

(b) “man”, $n = 200, \lambda = 0.5$

(c) “couple”, $n = 150, \lambda = 0.6$

(d) “stream & bridge”, $n = 250, \lambda = 0.6$

Fig. 8. Performance comparison of blind CM, informed ICA and PCA for Gaussian hosts and real images. Results averaged over 100 realizations of the spreading vector in each case.

For this kind of multibit setups, the authors of [3] have already proposed an estimator based on PCA and ICA. Note that the sole application of PCA can disclose, at most, the subspace spanned by the secret carriers, so the use of ICA in a subsequent step is necessary in order to disclose the carriers. However, the application of PCA to this scenario presents the drawback noted in Sect. VI-A1 for single carrier schemes: if the variance in the embedding subspace is not the largest, then PCA will be fooled. If PCA is not successful, ICA will not be either because it will work on the wrong subspace.

The methods for optimization on manifolds introduced above can be effectively adapted to this scenario. In most practical scenarios, the carriers are independently generated, thus being approximately orthogonal if n is large enough. If this is the case, then the matrix of carriers $\mathbf{S}$ is (approximately) unitary. For performing estimation of the carriers in the multibit scenario we resort to a "deflation" approach which, although clearly suboptimal,[3] yields good performance at reasonably low complexity, and basically consists in estimating the carriers one by one, properly orthogonalizing the observations by means of the Gram-Schmidt procedure (see [18, Chap. IV] for more details). Of course there is an ambiguity in the order of the estimated carriers, which is inherent to the multibit scenario and cannot be undone, as noted in [3]. Fig. 9(a) shows the estimation results when "man" is watermarked with 3 different carriers, after removing the ambiguity in the order of the estimated carriers. As can be seen, the blind CM estimator outperforms the PCA-ICA estimator in this scenario. For the latter, we have performed reduction to 3 dimensions with PCA, as in [3], followed by the application of ICA with the $\log\cosh$ contrast function.

2) Circular modulations: Some authors have proposed several variations of the spread spectrum embedding function aimed at improving its security. In [24], the "circular watermarking" method was proposed as a variation of the multibit ISS. In order to make more difficult the estimation of the carriers, an additional random sequence is used for getting the watermarked signal sphered in the embedding subspace. Similarly to the multibit ISS scheme, more than one carrier is needed for implementing this additional randomization. According to [24], the embedding function of the "Circular ISS" reads as

$$\mathbf{Y}_i = \mathbf{X}_i + \nu \sum_{k=1}^{N_b} (-1)^{M_i} \mathbf{s}_k d_k - \lambda \sum_{k=1}^{N_b} \frac{\mathbf{X}_i^T \mathbf{s}_k}{||\mathbf{s}_k||^2} \mathbf{s}_k, \text{ for } i = 1, \ldots, N_o, \tag{47}$$

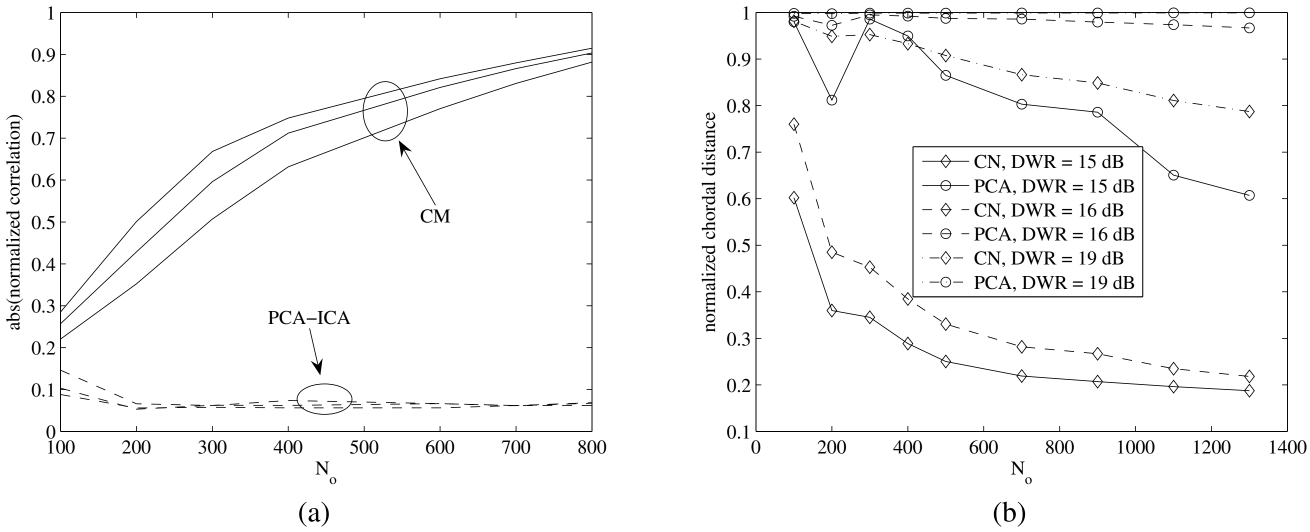
where $\mathbf{d} = [d_1, \ldots, d_{N_b}]$ is a vector uniformly distributed in the N_b-dimensional unit sphere. Hence, the attacker must estimate again a matrix of secret carriers $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_{N_b}] \in \mathbb{R}^{n \times N_b}$.

The Circular ISS scheme is another example of the tradeoff between robustness and security: it has the drawback of impairing the communication between embedder and decoder, since the latter ignores the randomization sequence $\mathbf{d}$, which is changed for each watermarked vector. However, the security over plain ISS is increased because the Circular ISS scheme effectively conceals the secret carriers, in such a way that it its impossible for the attacker to find their direction. Nevertheless, this modulation does not properly conceal the embedding subspace, since the watermarked signal in that subspace is strongly symmetric, with a squared norm close to $N_b \nu^2$ with high probability. This structure can be exploited by an attacker for estimating the subspace spanned by $\mathbf{S}$. If the attacker manages to disclose such subspace, then he can use this knowledge, e.g. for jamming the communication with low distortion, by applying an interfering signal in this particular subspace. Albeit this is truly a harmful attack, the Circular ISS modulation prevents from more optimal attacks, such as those based on bit flipping, which are only possible when the direction of the carriers has been accurately estimated. Using the nomenclature introduced in [12], Circular ISS is said to be key secure but not subspace secure.

For the purpose of subspace estimation one can think of a generalized CM cost function, that in fact can be interpreted as a Constant Norm criterion (CN) [29]:

$$J_{CN}^{blind}(\mathbf{S}, \nu) = E\left[\left(||\mathbf{Y}^T \mathbf{S}||^2 - \hat{\nu}^2 N_b\right)^2\right], \tag{48}$$

[3] A presumably better approach would be to perform optimization directly in the Stiefeld manifold for estimating the whole set of carriers at once, although this path will not be pursued in this paper. Of course, a suitable cost function must be defined for such problem.



(a) (b)

Fig. 9. In (a) we have the performance of "blind CM" and the joint PCA-ICA estimator for "man" watermarked with multibit ISS, $N_b = 3$, $n = 200$, $\lambda = 0.8$ and DWR = 17 dBs. In (b) we have the performance of the "blind CN" and PCA estimators for "man" watermarked with "circular ISS", $N_b = 2$, $n = 200$, and $\lambda = 0.7$. Results averaged over 100 realizations of the spreading vector for each image.

where for $\hat{\nu}$ we use the same estimate proposed for the blind CM estimator. For measuring the performance of the CN estimator we resort to the "chordal distance" [30], which is the natural measure of distance between two subspaces spanned by unitary matrices $\mathbf{P}$ and $\mathbf{Q}$, and is defined as $d_c(\mathbf{P}, \mathbf{Q}) = \frac{1}{\sqrt{2}}||\mathbf{P}\mathbf{P}^T - \mathbf{Q}\mathbf{Q}^T||_F$, where $||\cdot||_F$ denotes the Frobenius norm for matrices.[4] The chordal distance achieves its maximum, $\sqrt{N_b}$, when the two subspaces are perfectly orthogonal, and it equals 0 when both matrices generate the same subspace. For optimizing (48) we resort again to the Grassman manifold. Fig. 9(b) shows the performance comparison (in terms of the chordal distance normalized by $\sqrt{N_b}$) between the CN and PCA estimators applied to "man" with $N_b = 2$. As can be seen, in this scenario CN performs significantly better than PCA.

VIII. Links with other methods

In this section we briefly consider other watermarking methods which are strongly related to the spread spectrum formulation studied in this paper. Though this is by no means intended to be a rigorous analysis, the considerations provided here may help to link the security of the considered methods to the analysis carried out in this paper.

A. New spread spectrum methods

We consider two spread spectrum methods that have been recently proposed for watermark detection: Broken-Arrows (BA) [32], and the method by Comesaña et al. [33]. Watermark detection is modeled as a binary hypothesis testing problem where the watermark detector has to decide whether the received signal has been watermarked or not. If this received signal belongs to a certain region of the space, called "acceptance region" (which is dependent

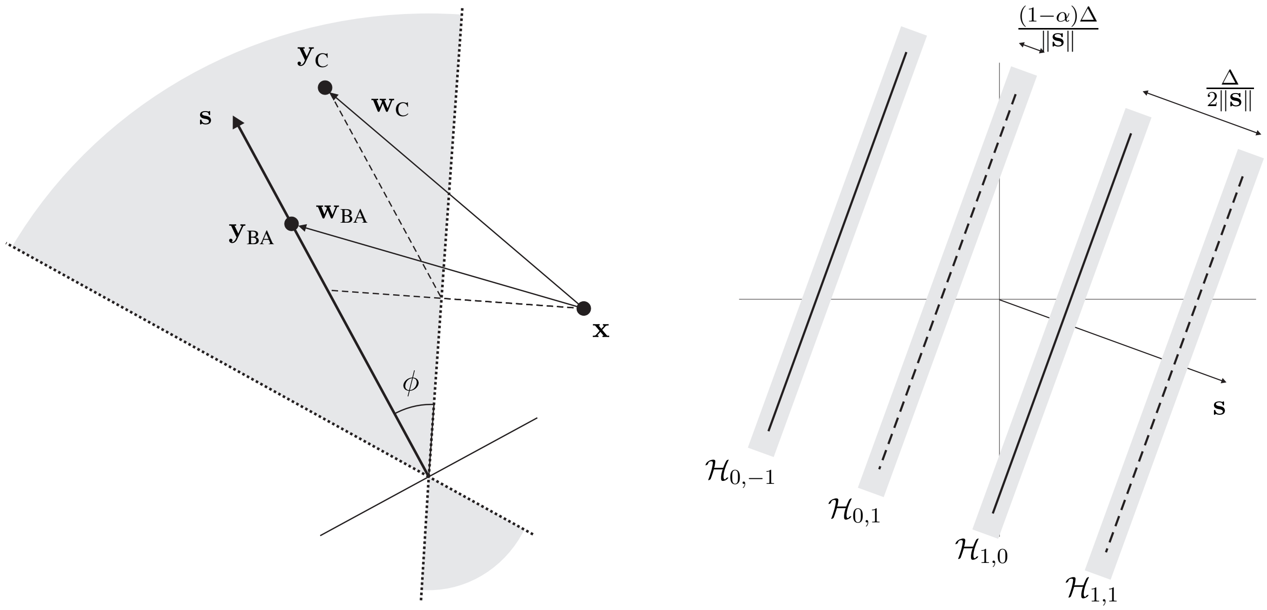
[4] As defined, the computation of the chordal distance involves matrices of size $n \times n$. Thus, when n is very large it is advisable to resort to other definitions which relate the notions of Principal Angle and Singular Value Decomposition, allowing to compute the chordal distance recursively [31, Chapt. II].

on the watermarking method and a secret key), then it is spotted as watermarked; otherwise it is considered to not be watermarked. For this reason, watermark detection is usually known as "one-bit" or "zero-bit" watermarking, depending on the author.

In general, the one-bit watermarking problem involves the joint design of an embedding function and an acceptance region. Comesaña's method has been proved to be optimal in the presence of additive Gaussian attacks if the host is Gaussian as well. The embedding function of both methods [32], [33] corresponds to the generic embedding rule given by Eq. (1), and the acceptance region in both cases consists in a double hypercone, with half-angle ϕ, symmetric around the coordinate origin. In both methods, the embedding distortion is constrained by imposing a maximum value p to the norm of the watermark vector (i.e. $||\mathbf{w}|| \leq p$). The considered embedding functions are illustrated in Fig. 10(a) for $n = 2$, and explained in the following. The vertex of the double cone is in the coordinate origin. The host to be watermarked is represented by the dot labeled as $\mathbf{x}$. The secret parameter of the embedding function is the direction of the axis of the cone, labeled as $\mathbf{s}$ (the spreading vector).

1) For the BA method, embedding is performed by moving the host perpendicularly to the boundary of the cone. The objective is to introduce the watermarked signal as deep as possible into the cone. Once reached the axis of the cone, if the norm of the watermark is below the distortion constraint, the remaining distortion is spent in introducing further the watermarked signal into the cone by moving it over the axis. The final watermarked signal is $\mathbf{y}_{\text{BA}}$, and the corresponding watermark vector $\mathbf{w}_{\text{BA}}$.
2) In Comesaña's method, the host is moved perpendicularly to the boundary of the cone until the latter is reached. Once there, the watermarked signal is moved in the direction of the axis $\mathbf{s}$ into the cone as much as allowed by the embedding distortion constraint. $\mathbf{w}_{\text{C}}$ and $\mathbf{y}_{\text{C}}$ denote the watermark and the watermarked signal, respectively.

When considering the n-dimensional case, interesting links with add-SS can be found. The false alarm probability, which is the probability of taking a non-watermarked signal as watermarked, depends on the angle of the hypercone. As n is increased, if the false alarm probability wants to be kept constant, the angle of the cone must be increased as well. Asymptotically ($n \to \infty$), the angle ϕ approaches $\pi/2$, and both embedding functions become approximately equivalent. Moreover, in the asymptotic case they become similar to add-SS, since the watermark would be approximately parallel to $\mathbf{s}$. This shows that the security of BA and Comesaña's method will approach that of add-SS as n is increased and the false alarm probability is kept constant. For a small ϕ, however, these embedding functions are obviously very different from add-SS, and there are also differences between both strategies. Compared to Comesaña's method, in BA the watermarked signal is always closer to the axis of the cone, $\mathbf{s}$. From a security point of view, this can be a drawback since the attacker could take advantage of the "clusterization" of the watermarked signals. On the other hand, the variance in the direction of the secret vector $\mathbf{s}$ is larger for Comesaña's method than for BA. This may represent a drawback, as studied in this paper, since it may lead to easier identification of the embedding subspace.


$\mathbf{y}_{\mathrm{C}}$
$\mathbf{w}_{\mathrm{C}}$
$\mathbf{s}$
$\mathbf{y}_{\mathrm{BA}}$
$\mathbf{w}_{\mathrm{BA}}$
$\mathbf{x}$
ϕ
$\frac{(1-\alpha)\Delta}{\|\mathbf{s}\|}$
$\frac{\Delta}{2\|\mathbf{s}\|}$
$\mathbf{s}$
$\mathcal{H}_{0,-1}$
$\mathcal{H}_{0,1}$
$\mathcal{H}_{1,0}$
$\mathcal{H}_{1,1}$

(a) Broken Arrows and Comesaña et al.

(b) ST-DM

Fig. 10. (a) Geometrical interpretation of the embedding function in the spread spectrum schemes proposed in [32] and [33]. (b) Geometrical interpretation of ST-DM data hiding for $n = 2$. In both cases, the shaded regions depict the possible values for the watermarked signal.

B. Spread Transform - Dither Modulation (ST-DM)

ST-DM is a well-known data hiding method. It can be considered as a hybrid scheme, since it combines a quantization strategy with spread spectrum. Its embedding function can be written as [34],[35]

$$\mathbf{y}_i = \mathbf{x}_i + \alpha \frac{\left(Q_{m_i}(\mathbf{x}_i^T \mathbf{s}) - \mathbf{x}_i^T \mathbf{s}\right)}{\|\mathbf{s}\|^2} \mathbf{s}, \tag{49}$$

where $0 \leq \alpha \leq 1$ is the distortion compensation parameter, and $Q_{m_i}(\cdot)$ is a scalar quantizer that depends on the embedded bit m_i (we consider only binary transmission schemes) with its centroids distributed according to

$$\Delta\mathbb{Z} + (-1)^{m_i} \frac{\Delta}{4}, \text{ for } m_i = 0, 1. \tag{50}$$

We assume that the secrecy relies only on the spreading vector, $\mathbf{s}$. The embedding function (49) imposes a very specific structure to the watermarked signal, as depicted in Fig. 10(b). The embedding operation in ST-DM is equivalent to quantizing the signal $\mathbf{Y}_i$ to the nearest hyperplane $\mathcal{H}_{m,j}$, $m = \{0, 1\}$; $j = \{-\infty, \ldots, \infty\}$, which is defined as

$$\mathcal{H}_{m,j} = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{s}^T \mathbf{y} = (-1)^m \frac{\Delta}{4} + j\Delta\}. \tag{51}$$

As can be seen, these hyperplanes are perpendicular to $\mathbf{s}$ and they are equidistant to each other. In Fig. 10(b), $\mathcal{H}_{m,j}$, $m = \{0, 1\}$; $j = \{-\infty, \ldots, \infty\}$, denotes the hyperplane corresponding to the jth centroid of the quantizer corresponding to bit m. The distance between adjacent hyperplanes is $\frac{\Delta}{2\|\mathbf{s}\|}$.

It is a known result [36] that, if the DWR is kept constant and n is increased, then the ratio $\frac{\Delta}{\|\mathbf{s}\|}$ is increased as well. Therefore, assuming a zero-mean host signal with finite variance, when n is sufficiently large, the probability that the host signal is quantized to other hyperplane different from $\mathcal{H}_{0,0}$ or $\mathcal{H}_{1,0}$ (i.e. the hyperplanes which are closest to the origin) is negligible. The consideration of the asymptotic case ($n \rightarrow \infty$) shows an interesting connection between

ST-DM and ISS. Notice that when all the hyperplanes (equivalently, the centroids) but the two closest to the origin are neglected, the embedding function of ST-DM can be written as

$$\mathbf{Y}_i = \mathbf{X}_i + (-1)^{M_i}\frac{\alpha\Delta}{4\|\mathbf{S}\|^2}\mathbf{S} - \frac{\alpha\mathbf{X}_i^T\mathbf{S}}{\|\mathbf{S}\|^2}\mathbf{S}, \tag{52}$$

where it is easy to identify the resemblance with Eq. (15). In fact, it can be shown that (52) and (15) are equivalent in terms of robustness and security. Thus, the security of ST-DM can be well approximated by that of ISS as n is increased (and the DWR is kept constant); however, this assertion does not necessarily hold for smaller values of n.

IX. Conclusions

The security of spread-spectrum-based data hiding methods has been investigated from theoretical and practical points of view. Among the theoretical results obtained in this paper, we would like to remark the following:

1) Under the same conditions of embedding distortion (i.e. keeping constant the power of the watermark), the decrease of the embedding rate (equivalently, the increase of n) has a harmful impact in the security level of WOA scenarios. In limiting cases of zero-rate watermarking, known for being robust to blind attacks, the penalty for ignoring the embedded messages becomes negligible, representing a serious threat to the security of the system.

2) A tradeoff between security and robustness has been shown to exist in the methods that perform host rejection. For the schemes studied in this paper, which cover a wide range of the spread spectrum schemes considered in the literature, host rejection can significantly decrease the security level of plain spread spectrum (add-SS). Nevertheless, different host rejection strategies can yield very different results: whereas the penalty for the ISS scheme vanishes as n is increased, the security level of γ-SS cannot be improved by increasing n.

In the practical side of the security problem, we have analyzed the previously proposed ICA and PCA estimators, whose good performance in certain practical scenarios had been demonstrated (see e.g. [3],[24]). In particular, PCA is very effective for circular hosts or when n is in the order of thousands, since in this case the variance in the embedding subspace is very large. On the other hand, ICA works also very well when applied on the appropriate subspace. In this paper we have tried to highlight the limitations of ICA and PCA. The point was to show that these estimators may fail in situations where, a priori, the estimation of the secret carrier seems easier for the attacker (e.g. using smaller n). In the end, we have shown that PCA and ICA can be fooled by appropriately choosing the embedding parameters. As for the new estimators proposed in this paper (informed ICA and blind CM), the statistical analysis performed in Section VI-B showed that they also present some drawbacks when facing certain combinations of embedding parameters. However, they have proved to work in a number of scenarios were the previous approaches have failed, as seen in Section VII. The combination of new and previous methods provides a wider battery of estimators for performing practical security tests that work for most practical situations. Finally, we want to note that with the chosen optimization algorithm, computational problems may appear if the size of the spreading vector (n) is very large. In this case, one should look for other optimization methods more suitable for "large scale" problems. In general, a complete and rigorous analysis of a practical estimator requires the consideration of three elements: the theoretical cost function, the finite sample effects in its estimation, and the optimization method itself.

Finally, we have considered in a preliminary manner the links between the security of the methods analyzed in this paper and the security of other methods strongly related to the spread spectrum formulation. Our preliminary considerations point out that, in certain asymptotic setups, the security of these methods approaches that of add-SS or ISS. In any case, they deserve a rigorous security analysis in the future.

Acknowledgements

The authors want to thank Dr. Pierre Moulin for his advice in the preliminary versions of this paper, and to Dr. Roberto López-Valcarce for his valuable comments on the Constant Modulus algorithm.

Appendix A
Bounds on the information leakage in the WOA scenario for add-SS

For obtaining the bounds to the information leakage, we will resort to the expression given by (7). Since its first term has been already computed in (6), we will focus on the remaining terms. For a fair user with perfect knowledge of $\mathbf{S}$, the observations are all mutually independent. Hence, the third term can be rewritten as

$$I(\mathbf{Y}_1, \dots, \mathbf{Y}_{N_o}; M_1, \dots, M_{N_o}|\mathbf{S})_{\text{add-SS}} = N_o \cdot I(\mathbf{Y}_1; M_1|\mathbf{S}). \tag{53}$$

For computing (53) we take advantage of the fact that the statistic $\mathbf{Y}_i^T\mathbf{S}$ is a sufficient statistic for decoding M_i when $\mathbf{S}$ is known by the decoder [14]. Thus, we have

$$\begin{aligned} I(\mathbf{Y}_1; M_1|\mathbf{S}) &= I(\mathbf{Y}_1^T\mathbf{S}; M_1|\mathbf{S}) = h(\mathbf{Y}_1^T\mathbf{S}|\mathbf{S}) - h(\mathbf{Y}_1^T\mathbf{S}|M_1, \mathbf{S}) \\ &= h(\mathbf{Y}_1^T\mathbf{S}|\mathbf{S}) - \frac{1}{2}\left(\log(2\pi e\sigma_X^2) + E[\log(\|\mathbf{S}\|^2)]\right), \end{aligned} \tag{54}$$

where we have used that $\mathbf{Y}_1^T\mathbf{S}|M_1 = m_1, \mathbf{S} = \mathbf{s} \sim \mathcal{N}(\|\mathbf{s}\|^2(-1)^{m_1}, \|\mathbf{s}\|^2\sigma_X^2)$. For the term $h(\mathbf{Y}^T\mathbf{S}|\mathbf{S})$ we must take into account that $\mathbf{Y}_1^T\mathbf{S}|\mathbf{S} = \mathbf{s} \sim \frac{1}{2}\left(\mathcal{N}(\|\mathbf{s}\|^2, \|\mathbf{s}\|^2\sigma_X^2) + \mathcal{N}(-\|\mathbf{s}\|^2, \|\mathbf{s}\|^2\sigma_X^2)\right)$, and that $h(\mathbf{Y}_1^T\mathbf{S}|\mathbf{S}) = E\left[h(\mathbf{Y}_1^T\mathbf{S}|\mathbf{S} = \mathbf{s})\right]$, where the expectation is taken over $\mathbf{S}$. The computation of the expectations in (54) requires numerical integration, taking into account that $||\mathbf{S}||^2 \sim \chi^2(n, \sigma_S)$. For the second term of (7) we can write

$$\begin{aligned} I(\mathbf{Y}_1, \dots, \mathbf{Y}_{N_o}; M_1, \dots, M_{N_o})_{\text{add-SS}} &= N_o \cdot H(M_1) - H(M_1, \dots, M_{N_o}|\mathbf{Y}_1, \dots, \mathbf{Y}_{N_o}) \\ &= N_o \cdot H(M_1) - \sum_{i=1}^{N_o} H(M_i|\mathbf{Y}_1, \dots, \mathbf{Y}_{N_o}, M_1, \dots, M_{i-1}) \\ &\leq N_o \cdot H(M_1) - N_o \cdot H(M_{N_o}|\mathbf{Y}_1, \dots, \mathbf{Y}_{N_o}, M_1, \dots, M_{N_o-1}) \qquad (55) \\ &= N_o \cdot I(\mathbf{Y}_{N_o}; M_{N_o}|\mathbf{Y}_1, \dots, \mathbf{Y}_{N_o-1}, M_1, \dots, M_{N_o-1}), \qquad (56) \end{aligned}$$

where the inequality (55) follows from the fact that conditioning reduces entropy [13]. Conditioned on a particular realization of the observations, we have that $\mathbf{Y}_{N_o} = \mathbf{X}_{N_o} + (-1)^{M_{N_o}} \cdot \bar{\mathbf{S}}_{N_o-1}$, with $\bar{\mathbf{S}}_{N_o-1}$ a random variable that follows the distribution of $\mathbf{S}$ conditioned on the $N_o - 1$ past observations, i.e. $\bar{\mathbf{S}}_{N_o} \sim \mathcal{N}(\mathbf{v}, \sigma^2_{\bar{S}_{N_o-1}}\mathbf{I}_n)$, with $\mathbf{v}$ and $\sigma^2_{\bar{S}_{N_o-1}}$ given by (3) and (4), respectively. Hence, considering a random variable $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\bar{S}_{N_o-1}}\mathbf{I}_n)$, and noticing that $(-1)^{M_{N_o}}\mathbf{N}$ is identically distributed to $\mathbf{N}$, then we can write

$$\mathbf{Y}_{N_o} = \mathbf{X}_{N_o} + (-1)^{M_{N_o}}(\mathbf{N} + \mathbf{v}) = \bar{\mathbf{X}}_{N_o} + (-1)^{M_{N_o}}\mathbf{v}, \tag{57}$$

where $\bar{\mathbf{X}}_{N_o} \triangleq \mathbf{X}_{N_o} + \mathbf{N} \sim \mathcal{N}(\mathbf{0}, (\sigma_X^2 + \sigma_{\bar{S}_{N_o-1}}^2)\mathbf{I}_n)$. Clearly, this implies that

$$I(\mathbf{Y}_{N_o}; M_{N_o}|\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}, M_1, \ldots, M_{N_o-1})_{\text{add-SS}} = I(\mathbf{Y}_{N_o}; M_{N_o}|\mathbf{V}_{N_o}), \tag{58}$$

where the components of $\mathbf{V}_{N_o}$ are the realizations of (3). Since $\boldsymbol{\mu}^T\mathbf{y}^{(i)}$ is zero-mean Gaussian with variance $(N_o - 1)\cdot(\sigma_X^2 + (N_o-1)\sigma_S^2)$, then $\mathbf{V}_{N_o} \sim \mathcal{N}\left(\mathbf{0}, \frac{(N_o-1)\sigma_S^4}{(N_o-1)\sigma_S^2+\sigma_X^2}\mathbf{I}_n\right)$. Hence, (58) becomes

$$\begin{aligned} I(\mathbf{Y}_{N_o}; M_{N_o}|\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o-1}, M_1, \ldots, M_{N_o-1})_{\text{add-SS}} &= I(\mathbf{Y}_{N_o}^T\mathbf{V}_{N_o}; M_{N_o}|\mathbf{V}_{N_o}) \\ = h(\mathbf{Y}_{N_o}^T\mathbf{V}_{N_o}|\mathbf{V}_{N_o}) - h(\mathbf{Y}_{N_o}^T\mathbf{V}_{N_o}|M_{N_o}, \mathbf{V}_{N_o}) = h(\mathbf{Y}_{N_o}^T\mathbf{V}|\mathbf{V}_{N_o}) &- E\left[\frac{1}{2}\log\left(2\pi e(\sigma_X^2 + \sigma_{\bar{S}_{N_o}}^2)\|\mathbf{V}_{N_o}\|^2\right)\right], \end{aligned} \tag{59}$$

where we have used again the fact that $\mathbf{Y}_{N_o}^T\mathbf{V}_{N_o}$ is a sufficient statistic for decoding (see [14]). Notice that the first term in the right hand side of (54) and (59) must be computed by means of numerical integration. Finally, combining (7) with (6), (53), (54), (56) and (59) we arrive at (8), the final expression of the upper bound.

A lower bound on the mutual information can be obtained by taking into account that

$$\begin{aligned} I(\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}; M_1, \ldots, M_{N_o})_{\text{add-SS}} &= \sum_{i=1}^{N_o}\sum_{j=1}^{N_o} I(\mathbf{Y}_i; M_j|\mathbf{Y}_1, \ldots, \mathbf{Y}_{i-1}, M_1, \ldots, M_{j-1}) \\ &\geq \sum_{i=2}^{N_o} I(\mathbf{Y}_i; M_i|\mathbf{Y}_1, \ldots, \mathbf{Y}_{i-1}, M_1, \ldots, M_{i-1}). \end{aligned} \tag{60}$$

Note that each term of (60) was already computed in (59). The final expression of the lower bound is given in (9).

Appendix B
Proof of Theorem 1

Consider the lower bound to the information leakage (9) where we take now $\log(2)$ as an upper bound of the third term. This gives an upper bound to the loss function (10). Now, recall that the loss function is always non-negative. Thus, it can be bounded as

$$N_o\log(2) - \sum_{i=2}^{N_o} I(\mathbf{Y}_i; M_i|\mathbf{V}_i) \geq \delta(N_o)_{\text{add-SS}} \geq 0, \text{ for } N_o \geq 2, \tag{61}$$

where $\mathbf{Y}_i = \bar{\mathbf{X}}_i + (-1)^{M_i}\mathbf{V}_i$, $\mathbf{V}_i \sim \mathcal{N}(\mathbf{0}, \frac{(i-1)\sigma_S^4}{(i-1)\sigma_S^2+\sigma_X^2}\mathbf{I}_n)$, and $\bar{\mathbf{X}}_i \sim \mathcal{N}(\mathbf{0}, (\sigma_X^2 + \sigma_{\bar{S}_{i-1}}^2)\mathbf{I}_n)$ with $\sigma_{\bar{S}_{i-1}}^2$ given by (4). The first term in the right hand side of (61) can be rewritten as $I(\mathbf{Y}_i; M_i|\mathbf{V}_i) = H(M_i) - H(M_i|\mathbf{Y}_i, \mathbf{V}_i)$. By inserting this expression into (61), we obtain

$$\log(2) + \sum_{i=2}^{N_o} H(M_i|\mathbf{Y}_i, \mathbf{V}_i) \geq \delta(N_o)_{\text{add-SS}} \geq 0, \tag{62}$$

Now we focus on the second term in the right hand side of (62). Using Fano's inequality [13] and taking into account that $|\mathcal{M}|$=2, the conditional entropy of M_i can be bounded from above as

$$H(M_i|\mathbf{Y}_i, \mathbf{V}_i) \leq H(P_e|\mathbf{V}_i), \tag{63}$$

where $P_e|\mathbf{v}_i$ denotes the probability of decoding error conditioned on $\mathbf{V}_i$, and $H(P_e)$ is the binary entropy function [13]. Hence, (61) is rewritten as

$$\log(2) + \sum_{i=2}^{N_o} H(P_e|\mathbf{V}_i) \geq \delta(N_o)_{\text{add-SS}} \geq 0, \tag{64}$$

For computing the error probability we define the scalar random variable $Z_i \triangleq \mathbf{Y}_i^T \mathbf{V}_i$, which is a sufficient statistic for the decoding of M_i. The statistic of Z_i conditioned on $\mathbf{v}_i$ is

$$Z_i|\mathbf{V}_i = \mathbf{v}_i \sim \frac{1}{2}\left(\mathcal{N}\left(-||\mathbf{v}_i||^2, (\sigma_X^2 + \sigma_{\bar{S}_{i-1}}^2)||\mathbf{v}_i||^2\right) + \mathcal{N}\left(||\mathbf{v}_i||^2, (\sigma_X^2 + \sigma_{\bar{S}_{i-1}}^2)||\mathbf{v}_i||^2\right)\right). \tag{65}$$

Eq. (63) holds for any decoder, and in particular for a decoder based on sign-decision. For the latter, $P_e|\mathbf{v}_i = \Pr\{Z_i > 0|M_i = 1, \mathbf{V}_i = \mathbf{v}_i\} = \Pr\{Z_i \leq 0|M_i = -1, \mathbf{V}_i = \mathbf{v}_i\}$. Obviously, from (65), we have

$$\Pr\{Z_i > 0|M_i = 1, \mathbf{V}_i = \mathbf{v}_i\} = Q\left(\frac{t_i^{\frac{1}{2}}}{(\sigma_X^2 + \sigma_{\bar{S}_{i-1}}^2)^{\frac{1}{2}}}\right), \tag{66}$$

where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt$ denotes the Gaussian Q-function, and $T_i \triangleq ||\mathbf{V}_i||^2 \sim \chi^2(n, \sigma_{V_i})$, with $\sigma_{V_i}^2 = \frac{(i-1)\sigma_S^4}{(i-1)\sigma_S^2 + \sigma_X^2}$. Using the well known Chernoff bound [37], we can upper bound the error probability (66) as

$$\Pr\{Z_i > 0|M_i = 1, \mathbf{V}_i = \mathbf{v}_i\} \leq \frac{1}{2}\exp\left(\frac{-t_i}{2(\sigma_X^2 + \sigma_{\bar{S}_{i-1}}^2)}\right). \tag{67}$$

Since $H(P_e)$ is increasing in $P_e \in [0, 0.5]$, (67) can be used to upper bound $H(P_e|\mathbf{V}_i = \mathbf{v}_i)$. Hence,

$$\begin{aligned} H(P_e|\mathbf{V}_i) &\leq E\left[H\left(\frac{1}{2}\exp\left(\frac{-T_i}{2(\sigma_X^2 + \sigma_{\bar{S}_{i-1}}^2)}\right)\right)\right] \\ &\leq H\left(\frac{1}{2}E\left[\exp\left(\frac{-T_i}{2(\sigma_X^2 + \sigma_{\bar{S}_{i-1}}^2)}\right)\right]\right) = H\left(\frac{1}{2}\left(\frac{\sigma_X^2 + \sigma_{\bar{S}_{i-1}}^2}{\sigma_X^2 + \sigma_{\bar{S}_{i-1}}^2 + \sigma_{V_i}^2}\right)^{\frac{n}{2}}\right) \triangleq H\left(\frac{\tau_i^{\frac{n}{2}}}{2}\right), \end{aligned} \tag{68}$$

where the second inequality follows from Jensen's inequality [13]. By combining (68) with (64), (11) is obtained. Now, the asymptotic results of Theorem 1 follow easily:

1) For proving the first result, we rewrite the term τ_i in (68), after some algebraic manipulations, as

$$\tau_i = \frac{i\sigma_S^2\sigma_X^2 + \sigma_X^4}{(i-1)\sigma_S^4 + i\sigma_S^2\sigma_X^2 + \sigma_X^4}. \tag{69}$$

In the limit, we have $\lim_{\sigma_S \to \infty} H\left(\frac{1}{2}\tau_i^{\frac{n}{2}}\right) = \lim_{\sigma_X \to 0} H\left(\frac{1}{2}\tau_i^{\frac{n}{2}}\right) = 0$ for $i \geq 2$. Hence, $\lim_{\text{DWR} \to -\infty} \delta(N_o)_{\text{add-SS}} \leq \log(2)$. This proves the first asymptotic result of Theorem 1.

2) The term τ_i in (68) is strictly smaller than 1; hence, (68) decreases with n, and $\lim_{n \to \infty} H\left(\frac{1}{2}\tau_i^{\frac{n}{2}}\right) = 0$. Thus, $\lim_{n \to \infty} \delta(N_o)_{\text{add-SS}} \leq 0$, proving the second asymptotic result of Theorem 1.

APPENDIX C
PROOF OF EQ. (22) IN THEOREM 2

For computing the information leakage for ISS, we first need to prove the following lemma.

Lemma 3: For the ISS embedding function, $\mathbf{Y}_i$ conditioned on the embedded message M_i is i.i.d. Gaussian.

Proof: For $N_o = 1$, $\mathbf{Y}_i$ is the sum of an i.i.d. Gaussian ($\mathbf{S}$) and another Gaussian whose covariance matrix depends on $\mathbf{S}$. Noticing that the sum of Gaussian random variables is Gaussian, we have that $\mathbf{Y}_i$ is Gaussian. If a Gaussian random variable is circularly symmetric, then it is i.i.d. Thus, it remains to be proved that the pdf of $\mathbf{Y}$ is circularly symmetric, i.e. that its pdf is invariant under rotations, or equivalently, $f(\mathbf{y}|M = 0) = f(\mathbf{H}\mathbf{y}|M = 0)$, for $\mathbf{H} \in \mathbb{R}^{n \times n}$ any unitary matrix.

In the following we drop the subindex i from the notation for simplicity, and we assume that the embedded message is $m = 0$ without loss of generality. The pdf $f(\mathbf{Hy}|M = 0)$ is calculated as

$$f(\mathbf{Hy}|M = 0) = \int_{\mathbb{R}^n} f(\mathbf{Hy}|\mathbf{S} = \mathbf{s}, M = 0) \cdot f(\mathbf{s})d\mathbf{s} = K \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\left((\mathbf{Hy} - \nu\mathbf{s})^T \mathbf{\Sigma}_{\mathbf{S}}^{-1}(\mathbf{Hy} - \nu\mathbf{s}) + \frac{||\mathbf{s}||^2}{\sigma_S^2}\right)\right) d\mathbf{s}, \quad (70)$$

where K is a constant (notice, from Eq. (20), that $|\mathbf{\Sigma}_{\mathbf{S}}|$ does not depend on $\mathbf{s}$). Taking into account that $\mathbf{HH}^T = \mathbf{I}_n$, the first term of the exponent in (70) can be rewritten as $(\mathbf{y} - \nu\mathbf{H}^T\mathbf{s})^T(\mathbf{H}^T\mathbf{\Sigma}_{\mathbf{S}}\mathbf{H})^{-1}(\mathbf{y} - \nu\mathbf{H}^T\mathbf{s})$. Hence, by realizing that $||\mathbf{s}||^2 = ||\mathbf{H}^T\mathbf{s}||^2$ (or equivalently, that $f(\mathbf{s})$ is circularly symmetric), we can write

$$f(\mathbf{Hy}|M = 0) = \int_{\mathbb{R}^n} f(\mathbf{y}|\mathbf{H}^T\mathbf{S} = \mathbf{H}^T\mathbf{s}, M = 0) \cdot f(\mathbf{H}^T\mathbf{s})d\mathbf{s} = f(\mathbf{y}|M = 0), \quad (71)$$

where for the last equality we have used the change of variable $\mathbf{s}' = \mathbf{H}^T\mathbf{s}$ (notice that the Jacobian of this change of variable is 1). This concludes the proof of the lemma. ■

By Lemma 3, the entropy of $\mathbf{Y}_j$ conditioned on M_j is

$$h(\mathbf{Y}_j|M_j) = \frac{1}{2}\left(\frac{n}{2}\log(2\pi e\sigma_{Y_0}^2) + \frac{n}{2}\log(2\pi e\sigma_{Y_1}^2)\right), \quad (72)$$

with $\sigma_{Y_0}^2$ and $\sigma_{Y_1}^2$ the variance of the components of $\mathbf{Y}_j$ conditioned on $M_j = 0$ and $M_j = 1$, respectively. Let $Y_{j,i}$ be the ith component of the jth observation, $Y_{j,i} = X_{j,i} + (-1)^{M_j}\nu S_i - \lambda\frac{\mathbf{X}_j^T\mathbf{S}}{||\mathbf{S}||^2}S_i$, for $i = 1, \ldots, n$. Since the components of $\mathbf{X}_j$ are zero-mean, it is easy to see that $E[Y_{j,i}|M_j = m_j] = 0$. Thus, the variance of $Y_{j,i}$ conditioned on M_j is given by

$$E\left[Y_{j,i}^2|M_j = m_j\right] = \sigma_X^2 + \nu^2\sigma_S^2 - 2\lambda\sigma_X^2 \cdot E\left[\left(\frac{S_i}{||\mathbf{S}||}\right)^2\right] + \lambda^2 \cdot E\left[\frac{(\mathbf{X}_j^T\mathbf{S})^2 S_i^2}{||\mathbf{S}||^4}\right]$$

$$= \sigma_X^2 + \nu^2\sigma_S^2 - 2\lambda\sigma_X^2 \cdot E\left[\left(\frac{S_i}{||\mathbf{S}||}\right)^2\right] + \lambda^2\sigma_X^2\left(E\left[\left(\frac{S_i}{||\mathbf{S}||}\right)^4\right] + E\left[\sum_{l=1,l\neq i}^{n}\left(\frac{S_l}{||\mathbf{S}||}\right)^2\left(\frac{S_i}{||\mathbf{S}||}\right)^2\right]\right). \quad (73)$$

We need to compute the second and fourth order statistics of $S_i/||\mathbf{S}||$. For a vector $\mathbf{S}$ isotropically distributed (i.e., with its probability density function invariant under rotations), the random variable defined as $\mathbf{S}' \triangleq \frac{\mathbf{S}}{||\mathbf{S}||}$ is uniformly distributed on the surface of the n-dimensional sphere of unit radius. Notice that the Gaussian vector with i.i.d. components, which is the case of interest for us, is indeed isotropically distributed. The marginal probability density function of one component S_i' can be computed by integrating the probability density function of $\mathbf{S}'$ over the surface of the $(n-1)$-dimensional sphere with radius $\sqrt{1-(s_i')^2}$, yielding

$$f(s_i') = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi} \cdot \Gamma\left(\frac{n-1}{2}\right)} \cdot \left(1 - s_i^2\right)^{\frac{n-3}{2}} \;\; \forall\, i = 1, \ldots, n;\; s_i' \in [-1, 1], \quad (74)$$

where $\Gamma(\cdot)$ denotes the complete Gamma function. Due to the symmetry of the pdf, it is easy to see that $E[S_i'] = 0$, and the moment generating function $\int_{-1}^{1}(s_i')^p f(s_i')ds_i'$ yields $E[(S_i')^2] = 1/n$ and $E[(S_i')^4] = 3/(2n+n^2)$. For computing the other statistic involved in (73), the joint pdf $f(s_i', s_j'),\; i \neq j$ must be calculated. Conditioned on $S_i' = s_i'$, the remaining components of $\mathbf{S}'$ are uniformly distributed over the surface of a $(n-1)$-dimensional sphere of radius $r_i = \sqrt{1-(s_i')^2}$. Hence, the conditional marginal pdf of S_j' is given by

$$f(s_j'|S_i' = s_i') = \frac{\Gamma\left(\frac{n-1}{2}\right)}{r_i \cdot \sqrt{\pi} \cdot \Gamma\left(\frac{n-2}{2}\right)} \cdot \left(1 - \left(\frac{s_j'}{r_i}\right)^2\right)^{\frac{n-4}{2}} \quad \forall\, j = 1, \ldots, n;\; j \neq i;\; s_j' \in [-r_i, r_i]. \quad (75)$$

After some algebraic simplifications, we arrive at the following expression for the joint pdf:

$$f(s_i', s_j') = \frac{n-2}{2\pi} \cdot \frac{r^2 \left(1 - (s_i')^2\right)^{\frac{n}{2}} \left(1 - \left(\frac{s_j'}{r_i}\right)^2\right)^{\frac{n}{2}}}{(s_i')^2 + (s_j')^2 - 1}. \tag{76}$$

Now we can calculate $E[(S_i')^k \cdot (S_j')^k]$, finding out that S_i' and S_j' are uncorrelated, but $E[(S_i')^2 \cdot (S_j')^2] = 1/(2n+n^2)$. Substituting in (73) we obtain

$$E\left[Y_{j,i}^2 | M_j = m_j\right] = \sigma_X^2 + \nu^2\sigma_S^2 - 2\lambda\frac{\sigma_X^2}{n} + \lambda^2\sigma_X^2\left(\frac{3}{2n+n^2} + \frac{n-1}{2n+n^2}\right) = \sigma_X^2 + \nu^2\sigma_S^2 + \sigma_X^2\frac{\lambda(\lambda-2)}{n}. \tag{77}$$

Since (77) is independent of the actual value of M_j, it follows that (72) is given by

$$h(\mathbf{Y}_j | M_j) = \frac{n}{2}\log\left(2\pi e\left(\sigma_X^2 + \nu^2\sigma_S^2 + \sigma_X^2\frac{\lambda(\lambda-2)}{n}\right)\right).$$

Finally, combining this result with (21) for $N_o = 1$ and rearranging terms, we arrive at (22).

Appendix D

Upper bound to the entropy of the observations in the KMA scenario for ISS

An upper bound to the entropy of the observations can be derived as follows:

$$h(\mathbf{Y}_1, \dots, \mathbf{Y}_{N_o} | M_1, \dots, M_{N_o}) \leq \sum_{i=1}^{n} h(Y_{1,i}, \dots, Y_{N_o,i} | M_1, \dots, M_{N_o}) \leq \sum_{i=1}^{n} \frac{1}{2} E\left[\log\left((2\pi e)^{N_o} |\mathbf{\Sigma}_{\mathbf{Y}_i}|\right)\right], \tag{78}$$

where $\mathbf{\Sigma}_{\mathbf{Y}_i}$ denotes the covariance matrix of $Y_{1,i}, \dots, Y_{N_o,i} | M_1 = m_1, \dots, M_{N_o} = m_{N_o}$, and the expectation is taken over all possible realizations of the messages sequences. It can be seen that, for a particular realization, the off-diagonal terms of the covariance matrix are given by

$$\mathbf{\Sigma}_{\mathbf{Y}_i}(j,k) = E\left[Y_{j,i} \cdot Y_{k,i} | M_1 = m_1, \dots, M_{N_o} = m_{N_o}\right] = (-1)^{m_j + m_k}\nu^2 \cdot E\left[S_i^2\right] = (-1)^{m_j+m_k}\nu^2\sigma_S^2,\ j \neq k. \tag{79}$$

The diagonal terms have been already calculated in (77). As can be seen, the covariance matrix $\mathbf{\Sigma}_{\mathbf{Y}_i}$ is of the form

$$\mathbf{\Sigma}_{\mathbf{Y}_i} = \begin{bmatrix} P + C & (-1)^{m_1+m_2}C & \cdots & (-1)^{m_1+m_{N_o}}C \\ (-1)^{m_2+m_1}C & P + C & \cdots & (-1)^{m_2+m_{N_o}}C \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{m_{N_o}+m_1}C & (-1)^{m_{N_o}+m_2}C & \cdots & P + C \end{bmatrix}, \tag{80}$$

with $C = \nu^2\sigma_S^2$, and $P = \sigma_X^2(1 + \frac{\lambda(\lambda-2)}{n})$. Taking into account that multiplying a row or a column of a matrix by a scalar multiplies the determinant by that scalar, the determinant of the above matrix results in $P^{N_o}\left(1 + \frac{N_o C}{P}\right)$, independently of the actual values of m_i. We can insert this result in Eq. (78), arriving at

$$h(\mathbf{Y}_1, \dots, \mathbf{Y}_{N_o} | M_1, \dots, M_{N_o}) \leq \frac{n}{2}\log\left((2\pi e)^{N_o}(\sigma_X^2)^{N_o}\left(1 + \frac{\lambda(\lambda-2)}{n}\right)^{N_o}\left(1 + \frac{N_o\nu^2\sigma_S^2}{\sigma_X^2\left(1 + \frac{\lambda(\lambda-2)}{n}\right)}\right)\right). \tag{81}$$

This result is finally combined with (21) for obtaining (23).

Appendix E

Upper bounds to the conditional entropy and the log-expectation of the norm

In the KMA scenario, the conditional pdf of the spreading vector follows a nonzero-mean Gaussian $\mathcal{N}(\mathbf{v}, \sigma^2_{\bar{S}_{N_o}})$, where $\mathbf{v}$ and $\sigma^2_{\bar{S}_{N_o}}$ are given by (3) and (4), respectively. The bounds derived in this appendix are based on the fact that the squared norm of a nonzero-mean Gaussian follows a noncentral Chi-square distribution. For the sake of notational clarity, let us define the random variable $T \sim \chi'^2(n, \mathbf{v}, \sigma_{\bar{S}_{N_o}})$. We will first derive the bound for the log-expectation of the norm. We can write

$$E\left[\log\left(T^{\frac{1}{2}}\right)\right] = \frac{1}{2}E\left[\log(T)\right] \leq \frac{1}{2}\log\left(E[T]\right) = \frac{1}{2}\log\left(n\sigma^2_{\bar{S}_{N_o}} + ||\mathbf{v}||^2\right), \tag{82}$$

where the upper bound follows from Jensen's inequality [13]. Since $\mathbf{V} \sim \mathcal{N}\left(\mathbf{0}, \frac{N_o\sigma_S^4}{N_o\sigma_S^2+\sigma_X^2}\mathbf{I}_n\right)$, the third term of (29) can be bounded from above as follows:

$$\begin{aligned}(n-1)E\left[E[\log(Q)|\mathbf{O}_1 = \mathbf{o}_1, \ldots, \mathbf{O}_{N_o} = \mathbf{o}_{N_o}]\right] &\leq \frac{(n-1)}{2}E\left[\log(n\sigma^2_{\bar{S}_{N_o}} + ||\mathbf{V}||^2)\right]\\ &\leq \frac{(n-1)}{2}\log\left(n\sigma^2_{\bar{S}_{N_o}} + n\frac{N_o\sigma_S^4}{N_o\sigma_S^2+\sigma_X^2}\right) = \frac{1}{2}\log(n\sigma_S^2),\end{aligned} \tag{83}$$

where we have applied again Jensen's inequality. An upper bound to the second term of (29) is now derived. First, note that $h(T^{\frac{1}{2}}) \leq \frac{1}{2}\log(2\pi e \cdot \text{var}(T^{\frac{1}{2}}))$. For the variance, we have [37, Chapter 2]

$$\text{var}(T^{\frac{1}{2}}) = n\sigma^2_{\bar{S}_{N_o}} + ||\mathbf{v}||^2 - \left(\sqrt{2}\sigma_{\bar{S}_{N_o}}\exp\left(-\frac{||\mathbf{v}||^2}{2\sigma^2_{\bar{S}_{N_o}}}\right)\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}\ {}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{||\mathbf{v}||^2}{2\sigma^2_{\bar{S}_{N_o}}}\right)\right)^2, \tag{84}$$

where ${}_1F_1(\alpha;\beta;x)$ denotes the confluent hypergeometric function of the first kind [17]. Hence, the second term of (29) can be upper bounded as

$$\begin{aligned}h(Q|\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}, M_1, \ldots, M_{N_o}) \leq \frac{1}{2}\log(2\pi e) + \frac{1}{2}\log&\left(n\sigma^2_{\bar{S}_{N_o}} + n\frac{N_o\sigma_S^4}{N_o\sigma_S^2+\sigma_X^2}\right.\\ &\left.- 2\sigma^2_{\bar{S}_{N_o}}\left(\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}\right)^2 \cdot E\left[\left(\exp\left(-\frac{||\mathbf{V}||^2}{2\sigma^2_{\bar{S}_{N_o}}}\right)\ {}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{||\mathbf{V}||^2}{2\sigma^2_{\bar{S}_{N_o}}}\right)\right)^2\right]\right),\end{aligned} \tag{85}$$

Using the Kummer transformation [17, Chapter 13], the expectation in (85) can be written as

$$E\left[\left(\exp\left(-\frac{||\mathbf{V}||^2}{2\sigma^2_{\bar{S}_{N_o}}}\right)\ {}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{||\mathbf{V}||^2}{2\sigma^2_{\bar{S}_{N_o}}}\right)\right)^2\right] = E\left[{}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -\frac{||\mathbf{V}||^2}{2\sigma^2_{\bar{S}_{N_o}}}\right)^2\right]. \tag{86}$$

By considering the integral representation of the hypergeometric function [17] and using Leibniz's rule [17, Chapter 3], it can be shown that ${}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -z\right)^2$ is convex in z. Thus, (86) can be lower bounded using Jensen's inequality:

$$E\left[{}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -\frac{||\mathbf{V}||^2}{2\sigma^2_{\bar{S}_{N_o}}}\right)^2\right] \geq {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -\frac{n\frac{N_o\sigma_S^4}{N_o\sigma_S^2+\sigma_X^2}}{2\sigma^2_{\bar{S}_{N_o}}}\right)^2 = {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -\frac{nN_o\sigma_S^2}{2\sigma_X^2}\right)^2. \tag{87}$$

Combining (85), (86), and (87), we obtain

$$\begin{aligned}&h(Q|\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_o}, M_1, \ldots, M_{N_o})\\ &\quad\leq \frac{1}{2}\log(2\pi e) + \frac{1}{2}\log\left(n\sigma_S^2 - \frac{2\sigma_S^2}{1+N_o\xi^{-1}}\left(\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}\right)^2\ {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -\frac{nN_o}{2}\xi^{-1}\right)^2\right).\end{aligned} \tag{88}$$

Finally, combining (29) with (5), (83), (88), and simplifying terms, the lower bound (30) follows.

REFERENCES

[1] T. Kalker, "Considerations on watermarking security," in *IEEE International Workshop on Multimedia Signal Processing*, Cannes, France, October 2001, pp. 201–206.

[2] M. Barni, F. Bartolini, and T. Furon, "A general framework for robust watermarking security," *Signal Processing*, vol. 83, no. 10, pp. 2069–2084, February 2003.

[3] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: theory and practice," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, October 2005.

[4] P. Comesaña, L. Pérez-Freire, and F. Pérez-González, "Fundamentals of data hiding security and their application to spread-spectrum analysis," in *7th Information Hiding Workshop, IH05*, ser. Lecture Notes in Computer Science. Barcelona, Spain: Springer Verlag, June 2005.

[5] P. Bas and J. Hurri, "Security of DM quantization watermarking schemes: a practical study for digital images," in *Fourth International Workshop on Digital Watermarking*, M. Barni, I. Cox, T. Kalker, and H. J. Kim, Eds., vol. 3710. Siena, Italy: Springer, September 2005, pp. 186–200.

[6] L. Pérez-Freire, F. Pérez-González, T. Furon, and P. Comesaña, "Security of lattice-based data hiding against the known message attack," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 4, pp. 421–439, December 2006.

[7] L. Pérez-Freire and F. Pérez-González, "Exploiting security holes in lattice data hiding," in *9th Information Hiding Workshop, IH07*, ser. Lecture Notes in Computer Science. Saint Malo, France: Springer Verlag, June 2007.

[8] I. J. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for images, audio and video," *IEEE Transactions on Image Processing*, vol. 6, pp. 1673–1687, December 1997.

[9] P. Moulin and A. Ivanović, "The zero-rate spread-spectrum watermarking game," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1098–1117, April 2003.

[10] H. S. Malvar and D. A. F. Florêncio, "Improved Spread Spectrum: a new modulation technique for robust watermarking," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 898–905, April 2003.

[11] L. Pérez-Freire, P. Moulin, and F. Pérez-González, "Security of spread-spectrum-based data hiding," in *Security, Steganography, and Watermarking of Multimedia Contents IX*, Edward J. Delp III and P. W. Wong, Eds., vol. 6505. San Jose, California, USA: SPIE, January 2007.

[12] F. Cayre and P. Bas, "Kerckhoffs-based embedding security classes for woa data hiding," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 1–15, March 2008.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley series in Telecommunications, 1991.

[14] H. V. Poor, *An introduction to signal detection and estimation*, 2nd ed. New York: Springer, 1998.

[15] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed., ser. Johns Hopkins Studies in Mathematical Sciences. Johns Hopkins University Press, 1996.

[16] J. A. Díaz-García and G. González-Farías, "Singular random matrix decompositions: Jacobians," *Journal of Multivariate Analysis*, vol. 93, no. 2, pp. 296–312, 2005.

[17] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1964.

[18] L. Pérez-Freire, "Digital watermarking security," Ph.D. dissertation, University of Vigo, Spain, 2008.

[19] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, ser. Adaptive and learning systems for signal processing, communications and control. John Wiley & Sons, 2001.

[20] G. Döerr and J.-L. Dugelay, "Danger of low-dimensional watermarking subspaces," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Montreal, Canada, 17-21 May 2004, pp. 93–96.

[21] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999.

[22] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.

[23] P. J. Huber, "Projection pursuit," *The Annals of Statistics*, vol. 13, no. 2, pp. 435–475, June 1985.

[24] P. Bas and F. Cayre, "Achieving subspace or key security for WOA using natural or circular watermarking," in *ACM Multimedia and Security Workshop*, Geneva, Switzerland, September 2006.

[25] A. Hyvärinen, "One-unit contrast functions for Independent Component Analysis: a statistical analysis," in *Proceedings of the VII IEEE Workshop on Neural Networks for Signal Processing*, Amelia Island, FL, USA, 24-26 September 1997, pp. 388–397.

[26] C. R. Jonhson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: a review," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1927–1950, 1998.

[27] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.

[28] "The USC-SIPI Image Database," available at http://sipi.usc.edu/database/.

[29] A. Goupil and J. Palicot, "New algorithms for blind equalization: the Constant Norm Algorithm family," *IEEE Transaction on Signal Processing*, vol. 55, no. 4, pp. 1436–1444, April 2007.

[30] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in grassmanian spaces," *Experimental Mathematics*, vol. 5, no. 2, pp. 139–159, 1996.

[31] M. E. Argentati, "Principal angles between subspaces as related to Rayleigh quotient and Rayleigh Ritz inequalities with applications to eigenvalue accuracy and an eigenvalue solver," Ph.D. dissertation, University of Colorado, Denver, 2003.

[32] T. Furon and P. Bas, "Broken Arrows," *Eurasip Journal on Information Security*, 2008, to appear.

[33] P. Comesaña, M. Barni, and N. Merhav, "Asymptotically optimum embedding strategy for one-bit watermarking under Gaussian attacks," in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, Edward J. Delp III, P. W. Wong, J. Dittmann, and N. Memon, Eds., vol. 6819. San Jose, California, USA: SPIE, January 2008.

[34] B. Chen and G. Wornell, "Quantization Index Modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[35] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar Costa Scheme for information embedding," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, April 2003, special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery.

[36] F. Pérez-González, F. Balado, and J. R. Hernández, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 960–980, April 2003, Special Issue on Signal Processing for Data Hiding in Digital Media & Secure Content Delivery.

[37] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.