Asymptotically Optimum Universal Watermark Embedding and Detection in the High SNR Regime

Pedro Comesaña, Neri Merhav, and Mauro Barni

Abstract

The problem of optimum watermark embedding and detection was addressed in a recent paper by Merhav and Sabbag, where the optimality criterion was the maximum false–negative error exponent subject to a guaranteed false–positive error exponent. In particular, Merhav and Sabbag derived universal asymptotically optimum embedding and detection rules under the assumption that the detector relies solely on second order joint empirical statistics of the received signal and the watermark. In the case of a Gaussian host signal and a Gaussian attack, however, closed–form expressions for the optimum embedding strategy and the false–negative error exponent were not obtained in that work. In this paper, we derive the false negative error exponent for any given embedding strategy and use such a result to show that in general the optimum embedding rule depends on the variance of the host sequence and the variance of the attack noise. We then focus on high SNR regime, deriving the optimum embedding strategy for such a set-up. In this case a universally optimum embedding rule turns out to exist and to be very simple with an intuitively–appealing geometrical interpretation. The effectiveness of the newly proposed embedding strategy is evaluated numerically.

Index Terms

Hypothesis testing, Neyman–Pearson, watermark detection, watermark embedding, watermarking.

P. Comesaña is with the Signal Theory and Communications Department, University of Vigo, Campus Lagoas-Marcosende, Vigo 36310, Spain, (phone: +34 986 818655, fax: +34 986 812116, e-mail: pcomesan@gts.tsc.uvigo.es), N. Merhav is with the Department of Electrical Engineering, Technion – I.I.T., Haifa 32000, Israel, (phone/fax: +972-4-8294737, e-mail: merhav@ee.technion.ac.il), M. Barni is with the Department of Information Engineering, University of Siena, Via Roma 56, Siena 53100, Italy, (phone: phone: +39 0577 234850 int. 1005, fax: +39 0577 233630, e-mail: barni@dii.unisi.it).

This work was partially supported by the Italian Ministry of Research and Education under FIRB project no. RBIN04AC9W, by Xunta de Galicia under Projects 07TIC012322PR (FACTICA), 2006/150 (Consolidation of Research Units), and by the Spanish Ministry of Science and Innovation under projects COMONSENS (ref. CSD2008-00010) and SPROACTIVE (ref. TEC2007-68094-C02-01/TCM).

I. INTRODUCTION

About a decade ago, the community of researchers in the field of watermarking and data hiding has learned about the importance and relevance of the problem of channel coding with non–causal side information at the transmitter [1], and in particular, its Gaussian version – *writing on dirty paper*, due to Costa [2], along with its direct applicability to watermarking, cf. [3], [4]. Costa's main result is that the capacity of the additive white Gaussian noise (AWGN) channel with an additional independent interfering signal, known non–causally to the transmitter only, is the same as if this interference was available at the decoder as well (or altogether non–existent). When applied in the realm of watermarking and data hiding, this means that the host signal (playing the role of the interfering signal), should not be actually considered as additional noise, since the embedder (the transmitter) can incorporate its knowledge upon generating the watermarked signal (the codeword). The methods based on this paradigm, usually known as *side-informed* methods, can even asymptotically eliminate (under some particular conditions) the interference of the host signal, that was previously believed to be inherent to any watermarking system.

Ever since the relevance of Costa's result to watermarking has been observed, numerous works have been published about the practical implementation of the side–informed paradigm for the so-called *multi–bit watermarking* [4], [5], [6], [7] case, where the decoder estimates the transmitted message among many possible messages. Far less attention has been devoted, however, to the problem of deciding on the presence or absence of a given watermark in the observed signal. In fact, in most of the works that deal with this binary hypothesis testing problem, usually known as zero–bit (a.k.a. one–bit) watermarking, the watermarking displacement signal does not depend on the host[1] [8], [9], [10], [11], [12] that then interferes with the watermark, thus contributing to augment the error probability. To the best of our knowledge, exceptions to this statement are the works by Cox *et al.* [3], [13], Liu and Moulin [14], Merhav and Sabbag [15] and Furon *et al.* [16], [17]. In the next few paragraphs, we briefly describe the main results contained in these works.

Cox et al. [3], [13]: In [3], Cox *et al.* introduce the paradigm of watermarking as a coded communication system with side information at the embedder. Based on this paradigm, and by considering a statistical model for attacks, the authors propose a detection rule based on the Neyman–Pearson criterion. The resulting detection region is replaced by the union of two hypercones; mathematically, this detection rule is given by $\frac{|\mathbf{s}^t \cdot \mathbf{u}|}{\|\mathbf{s}\| \cdot \|\mathbf{u}\|} \geq \tau(\alpha)$, where $\mathbf{s}$ is the received signal, $\mathbf{u}$ is the watermark, $\mathbf{s}^t$ is the transpose of $\mathbf{s}$, $\mathbf{s}^t \cdot \mathbf{u}$ is the inner product of $\mathbf{s}$ and $\mathbf{u}$, α is the maximum allowed false–positive probability, and $\tau(\alpha)$ is the decision threshold, which is a function of α. In a successive paper [13], Miller *et al.* also compare the performance of the strategy of [3] to other typical embedding strategies. No attempt is made to jointly design the optimum embedding and detection rules.

In [18] Furon and Bas used a set of (sligthly modified) double hypercones for zero–bit watermarking applications, and proposed to design the embedding strategy in such a way to maximize the minimum distance to the detection boundary.

[1]This is not really the case in practical scenarios, where the watermarking displacement signal must be perceptually shaped; nevertheless, when performing theoretical analysis the Euclidean norm is extensively used for the sake of analysis simplicity, therefore neglecting perceptual considerations. In any case, the dependency produced by perceptual considerations is not intended to reduce the host–interference effect.

Liu and Moulin [14]: In [14], both false–positive and false–negative error exponents are studied for the zero–bit watermarking problem, both for additive spread spectrum (Add-SS) and a quantization index modulation (QIM) technique [4]. The constraint on the embedding distortion is expressed in terms of the mean Euclidean norm of the watermarking displacement signal, and the non–watermarked signal is also assumed to be attacked (with attacks that impact the false–positive error probability). For Add-SS, exact expressions of the error exponents of both false–positive and false–negative probabilities are derived. For QIM, the authors provide bounds only. These results show that although the error exponents of QIM are indeed larger than those obtained by public Add-SS (where the host signal is not available at the detector), they are still smaller than those computed for private Add-SS (where the host signal is also available at the detector). This seems to indicate that the interference due to the host is not completely removed.

A practical scheme where quantization–based methods are used for zero–bit watermarking purposes was proposed by Pérez-Freire *et al.* in [19]. In that work several detection regions are proposed, based on the geometry of the quantization noise at the detector; the corresponding false–positive and false–negative error probabilities are calculated.

Merhav and Sabbag [15]: In [15], the problem of zero–bit watermarking is approached from an information–theoretic point of view. Optimum embedders and detectors are sought, in the sense of minimum false–negative probability subject to the constraint that the false–positive exponent is guaranteed to be at least as large as a given prescribed constant $\lambda > 0$, under a certain limitation on the kind of empirical statistics gathered by the detector. Another feature of the analysis in [15] is that the statistics of the host signal are assumed unknown. The proposed asymptotically optimum detection rule compares the empirical mutual information between the watermark $\mathbf{u}$ and the received signal $\mathbf{y}$ to a threshold depending on λ. In the Gaussian case, this boils down to thresholding the absolute value of the empirical correlation coefficient between these two signals. Merhav and Sabbag also derive the optimal embedding strategy for the attack–free case and derive a lower bound on the false–negative error exponent. Furthermore, the optimization problem associated with optimum embedding is reduced to an easily implementable 2D problem yielding a very simple embedding rule. In the same paper, Merhav and Sabbag also study the scenario where the watermarked signal is attacked. In this case, however, closed–form expressions for the error exponents and the optimum embedding rule are not available due to the complexity of the involved optimizations.

Furon et al. [16], [17]: In [16] Furon *et al.* propose to use the discrimination (i.e., the Kullback-Leibler Divergence) between the probability density function (pdf) of the original host signal and the pdf of its watermarked and attacked version in order to quantify the goodness of zero–bit embedding strategies. The considered attack is based on adding AWGN to the watermarked content, and scaling the resulting signal in order to have the same variance of the original host. The argument put forward [16] is that a high discrimination is a necessary condition to have good detection performances, so the watermark detection problem is equivalent to finding the embedding function that maximizes the discrimination; be aware that this analysis requires a perfect knowledge of the statistics of all the involved signals. By using this measure, the authors analyze the effect of considering quantization–based approaches,

as well as the Improved Spread Spectrum [20] technique, showing that the later achieves optimal performance for asymptotically long sequences. In the second part of [16], and in [17], Furon uses the Pitman–Noether theorem [21] to derive the form of the best detector for a given embedding function, and the best embedding function for a given detection function. By combining these results, a differential equation is obtained, that the author refers to as the *fundamental equation of zero-bit watermarking*. Furon shows that many of the most popular watermarking methods in the literature can be seen as special cases of the fundamental equation, ranging from Add-SS, multiplicative spread spectrum, or JANIS [22] (a zero–bit watermarking technique previously proposed by Furon *et al.*, where the detector statistic is heuristically computed as an n-order function, and the watermarking displacement signal is a scaled version of its gradient), to a two-sheet hyperboloid, or even combinations of the previous techniques with watermarking on a projected domain [23], or watermarking based on lattice quantization. Compared with the framework introduced in [15], two important differences must be highlighted:

- In [17], the watermarking displacement signal is constrained to be a function of the host signal which is scaled to yield a given embedding distortion. This means that in this set–up the direction of the watermarking displacement signal can not be changed as a function of the allowed embedding distortion.
- One of the conditions that must be verified to apply the Pitman–Noether theorem is that the power of the watermarking displacement signal goes to zero when the dimensionality increases without bound. In fact, Furon hypothesizes that this is the reason why neither the absolute normalized correlation nor the normalized correlation are solutions of the fundamental equation.

In this paper, we extend the results of [15] by deriving the false negative error exponent for any given embedding strategy in the Gaussian set–up, that is, for a Gaussian host signal and a Gaussian attack channel. As in [15], we assume that the detector is of limited resources, specifically, that it relies only on the Euclidean norm of the received signal and the empirical correlation between the received signal and the watermark. We then use the optimal (under the mentioned constraints) detector obtained in [15] to derive the optimum embedding strategy in the Neyman–Pearson sense of maximizing the false–negative error exponent for a given guaranteed false–positive error exponent. In particular, we show that the optimum embedding rule depends on the variance of both the host sequence and the attacking noise. In the second part of the paper, we turn our attention to the high-SNR regime, where the variance of the attacking noise is much smaller than the variance of the host signal and the embedding distortion. For this set-up a class of universal (asymptotically) optimum embedding strategies is derived, in the sense that they do not depend on the variances of the host sequence and the attacking noise. Closed-form expressions for asymptotically optimum embedding rules are also derived. We then consider one particular embedding strategy in the class derived before fitting the case of a vanishingly small (yet strictly positive) false negative error exponent. The performance of the new scheme is evaluated numerically, showing that in addition to be asymptotically optimum in the considered set-ups, the proposed scheme provides good performance in a wide range of settings, including realistic situations.

The remaining part of the paper is organized as follows: In Section II, we introduce notation conventions and formalize the problem. In Section III, the asymptotically optimum detection region is derived. In Section IV, we

use it to derive the false–negative error exponent for a generic embedding rule. The optimization of the false–negative error exponent resulting in the derivation of the high-dimensionality asymptotically optimum embedding is addressed in Section V. Section VI is devoted to the evaluation of the performance of the embedding rules derived in Section V for various settings. Finally, the main results of this work are summarized in Section VII where some suggestions for future research are also outlined.

II. Notation and Problem Formulation

Throughout the sequel, we denote scalar random variables by capital letters (e.g., V), their realizations with corresponding lower case letters (e.g., v), and their alphabets, with the respective script font (e.g., $\mathcal{V}$). The same convention applies to n–dimensional random vectors and their realizations, using bold face fonts (e.g., $\mathbf{V}$, $\mathbf{v}$). The alphabet of each corresponding n–vector will be taken to be the n–th Cartesian power of the alphabet of a single component, which will be denoted by the alphabet of a single component with a superscript n (e.g.,$\mathcal{V}^n$). The i-th component of a vector $\mathbf{V}$ is denoted V_i. The probability law of a random vector $\mathbf{V}$ is described by its pdf $f_{\mathbf{V}}(\mathbf{v})$. The equality in the exponential scale as a function of n will be denoted by $\doteq$; more precisely, if $\{a_n\}$ and $\{b_n\}$ are two positive sequences, $a_n \doteq b_n$ means that $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$.

Let $\mathbf{u}$ and $\mathbf{x}$, both $n-$dimensional vectors, be the *watermark sequence* and the *host sequence*, respectively. While u_i, $i = 1, \ldots, n$, the components of $\mathbf{u}$, take on binary values in $\mathcal{U} = \{-1, +1\}$, the components of $\mathbf{x}$, namely, x_i, $i = 1, \ldots, n$, take values in $\mathcal{X} = \mathbb{R}$. The embedder receives $\mathbf{x}$ and $\mathbf{u}$, and produces the *watermarked sequence* $\mathbf{y}$, yet another n–dimensional vector with components in $\mathcal{Y} = \mathbb{R}$. We refer to the difference signal $\mathbf{w} = \mathbf{y} - \mathbf{x}$ as the *watermarking displacement signal*. The embedder must keep the embedding distortion $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{y} - \mathbf{x}\|^2 = ||\mathbf{w}||^2$ within a prescribed limit, i.e., $d(\mathbf{x}, \mathbf{y}) \leq nD$, where $D > 0$ is the maximum allowed distortion per dimension, uniformly for every $\mathbf{x}$ and $\mathbf{u}$.

The output signal of the transmitter may either be the unaltered original host $\mathbf{x}$, in the non–watermarked case, or the vector $\mathbf{y}$, in the watermarked case. In both cases, the output signal is subjected to an attack, which yields a *forgery* signal, denoted by $\mathbf{s}$. The action of the attacker is modeled by a channel, which is given in terms of a conditional probability density of the forgery given the input it receives, $W(\mathbf{s}|\mathbf{x})$ – in the non–watermarked case, or $W(\mathbf{s}|\mathbf{y})$ – in the watermarked case. For the sake of convenience, we define $\mathbf{z}$ as the noise vector added by the attacker, i.e., the difference between the forgery signal $\mathbf{s}$ and the channel input signal, which is the transmitter output ($\mathbf{x}$ or $\mathbf{y}$, depending on whether the signal is watermarked or not). We assume that $\mathbf{z}$ is a Gaussian vector with zero–mean, i.i.d. components, all having variance σ_Z^2.[2]

The detector partitions $\mathbb{R}^n$ into two complementary regions, Λ (a.k.a. the detection region) and Λ^c. If $\mathbf{s} \in \Lambda$, the detector decides that the watermark is present (hypothesis H_1), otherwise it decides that the watermark is absent (hypothesis H_0). We assume that the detector knows the watermark $\mathbf{u}$, but does not know the host signal $\mathbf{x}$ (blind

[2]Although different additive noise variances could be considered depending on the fact of the transmitted signal being watermarked or not, we will not distinguish the case where those variances are different, as due to the circular symmetry of the Gaussian noise, it is irrelevant for the subsequent derivation.

or public watermarking). The design of the optimum detection region for the attack–free case was studied in [15], and it is generalized to the case of Gaussian attacks in Section III.

The performance of a zero–bit watermarking system is usually measured in terms of the tradeoff between the *false positive* probability of deciding that the watermark is present when it is actually absent, i.e.,

$$P_{fp} = \int_{\Lambda} \mathrm{d}\mathbf{s} \cdot [2\pi(\sigma_X^2 + \sigma_Z^2)]^{-n/2} \cdot \exp\left\{-\frac{\|\mathbf{s}\|^2}{2(\sigma_X^2 + \sigma_Z^2)}\right\} \tag{1}$$

and the *false negative* probability, of deciding that the watermark is absent when it is actually present, i.e.,

$$P_{fn} = \int_{\Lambda^c} \mathrm{d}\mathbf{s} \int_{\mathbb{R}^n} \mathrm{d}\mathbf{x} \cdot (2\pi\sigma_X^2)^{-n/2} \cdot \exp\left\{-\frac{\|\mathbf{x}\|^2}{2\sigma_X^2}\right\} \cdot (2\pi\sigma_Z^2)^{-n/2} \cdot \exp\left\{-\frac{\|\mathbf{s} - f(\mathbf{x},\mathbf{u})\|^2}{2\sigma_Z^2}\right\}, \tag{2}$$

where f is the embedding function, that is, $\mathbf{y} = f(\mathbf{x},\mathbf{u})$. As n grows without bound, these probabilities normally decay exponentially. The corresponding exponential decay rates, i.e., the *error exponents*, are defined as

$$E_{fp} \triangleq \lim_{n\to\infty} -\frac{1}{n} \ln P_{fp}, \tag{3}$$

$$E_{fn} \triangleq \lim_{n\to\infty} -\frac{1}{n} \ln P_{fn}. \tag{4}$$

The aim of this paper is to devise a detector as well as an embedding rule for a zero–mean, i.i.d. Gaussian host with variance σ_X^2 and a zero–mean memoryless Gaussian attack channel with noise power σ_Z^2, where the detector is limited to base its decision on the empirical energy of the received signal and its empirical correlation with $\mathbf{u}$. Both σ_X^2 and σ_Z^2 are assumed unknown to the detector, while the embedder knows them[3]. We seek high-dimensionality asymptotically optimum embedding and detection rules in the sense of maximizing the false–negative error exponent, E_{fn}, subject to the constraint that $E_{fp} \geq \lambda$, where λ is a prescribed positive real.

III. Optimum Detection rule

In [15], an asymptotically optimum detector is derived for the discrete case and for the continuous Gaussian case. In the latter case, it is shown that if the detector is limited to base its decision on the empirical energy of the received signal, $\frac{1}{n}\sum_{i=1}^n s_i^2$, and its empirical correlation with the watermark, $\frac{1}{n}\sum_{i=1}^n u_i s_i$, then an asymptotically optimum decision strategy, in the above defined sense, is to compare the (Gaussian) empirical mutual information, given by:

$$\hat{I}_{\mathbf{us}}(U;S) = -\frac{1}{2}\ln\left[1 - \frac{\left(\frac{1}{n}\sum_{i=1}^n u_i s_i\right)^2}{\left(\frac{1}{n}\sum_{i=1}^n u_i^2\right)\left(\frac{1}{n}\sum_{i=1}^n s_i^2\right)}\right] = -\frac{1}{2}\ln\left[1 - \frac{\left(\frac{1}{n}\sum_{i=1}^n u_i s_i\right)^2}{\frac{1}{n}\sum_{i=1}^n s_i^2}\right] \tag{5}$$

to λ, or equivalently, to compare the absolute normalized correlation

$$|\hat{\rho}_{\mathbf{us}}| = \frac{\left|\frac{1}{n}\sum_{i=1}^n u_i s_i\right|}{\sqrt{\frac{1}{n}\sum_{i=1}^n s_i^2}}, \tag{6}$$

[3]We will remove this assumption in the second part of the paper where we focus on the high SNR regime.

to $\sqrt{1-e^{-2\lambda}}$, i.e., the detection region is the union of two hypercones, around the vectors $\mathbf{u}$ and $-\mathbf{u}$, with a spread depending on λ. This decision rule of thresholding the empirical mutual information, or empirical correlation, is intuitively appealing since the empirical mutual information is an estimate of the degree of statistical dependence between two data vectors.[4]

For the present setting, we have to extend the analysis to incorporate the Gaussian attack channel. This turns out to be a straightforward task, since in the non–watermarked case (pertaining to the false–positive constraint), $\mathbf{s}$ continues to be Gaussian – the only effect of the channel is to change its variance, which is assumed unknown to the detector anyhow. Thus, the detection rule outlined above continues to be asymptotically optimum also in our setting.

Before we proceed with the derivation of the optimum embedder, it is instructive to look more closely at the dependence of the detection region on the false–positive exponent λ. As mentioned earlier, the choice of λ imposes a threshold that must be compared with (6) in order to provide the detector output. This is equivalent to establishing the limit angle of the detection region, that we will denote by $\beta = \arccos(\sqrt{1-e^{-2\lambda}}) = \arcsin(e^{-\lambda}) \in [0, \pi/2]$. Letting $\theta = \arccos(\hat{\rho}_{\mathbf{us}})$, we then have:

$$\begin{aligned} P_{fp} &= \Pr\{\hat{\rho}^2_{\mathbf{us}} > 1 - e^{-2\lambda} | H_0\} \\ &= \Pr\{0 \le \theta < \beta | H_0\} + \Pr\{\pi - \beta < \theta \le \pi | H_0\} \\ &= 2\Pr\{0 \le \theta < \beta | H_0\} = \frac{2A_n(\beta)}{A_n(\pi)} = 1 - I_{\cos(\beta)^2}(1/2, (n-1)/2) \doteq e^{n \ln(\sin \beta)}, \end{aligned} \tag{7}$$

where $A_n(\theta)$ is the surface area of the n–dimensional spherical cap cut from a unit sphere centered in the origin, by a right circular cone of half angle θ, and $I_{(\cdot)}(\cdot,\cdot)$ is the Regularized Incomplete Beta Function. In (7), we used the fact that in the non–watermarked case, where $\mathbf{s}$ is a zero–mean Gaussian vector with i.i.d. components, independent of $\mathbf{u}$, the normalized vector $\mathbf{s}/\|\mathbf{s}\|$ is uniformly distributed over the surface of the n–dimensional unit sphere, as there are no preferred directions.

IV. The False–Negative Exponent

In this section, we derive the false–negative error exponent as a function of the watermarking displacement signal $\mathbf{w}$. In order to do that, and without loss of generality, we apply the Gram–Schmidt orthogonalization procedure to the vectors $\mathbf{u}$, $\mathbf{x}$ and $\mathbf{w}$,[5] and then select the remaining $n-3$ orthonormal basis vectors for $\mathbb{R}^n$ in an arbitrary manner; the nth basis vector will be denoted by $\mathbf{e}_n$. After transforming to the resulting coordinate system, the above vectors have the forms $\mathbf{u} = (\sqrt{n}, 0, 0, \ldots, 0)$, $\mathbf{x} = (x_1, x_2, 0, \ldots, 0)$, $x_2 \ge 0$, $\mathbf{w} = (w_1, w_2, w_3, 0, \ldots, 0)$ and

[4]It is also known from the literature on universal decoding that the maximum mutual information (MMI) decoder, which selects the codeword having the highest empirical mutual information with the channel output vector, is universally optimum (in the random coding error exponent sense) for memoryless channels.

[5]In case $\mathbf{x}$ lies in the subspace spanned by $\mathbf{u}$ (i.e., $\mathbf{x}$ is proportional to $\mathbf{u}$), an arbitrary unit vector, orthogonal to $\mathbf{u}$ can be chosen as a second basis vector as part of the Gram–Schmit procedure. Similarly, if $\mathbf{w}$ lies in the subspace spanned by the two previous vectors, then an arbitrary unit vector orthogonal to both can be selected as the third basis vector.

$\mathbf{y} = (x_1+w_1, x_2+w_2, w_3, 0, \ldots, 0)$, while all the components of the noise sequence $\mathbf{z}$ will remain, in general, non–null. For the sake of convenience, in the remainder of the paper we will consider the normalized vector $\bar{\mathbf{w}} = \frac{1}{\sqrt{n}}\mathbf{w}$ instead of $\mathbf{w}$. The false–negative error exponent derived in this section, will be used later to derive asymptotically optimal embedding rules subject to the distortion constraint, $\|\mathbf{w}\|^2 \leq nD$, which corresponds to the constraint $\|\bar{\mathbf{w}}\|^2 \leq D$. For convenicence, we also define the function

$$S(x) = \frac{1}{2}(x - \ln x - 1).$$

Our first main result is the following.

Theorem 1: Let P_{fp}, P_{fn} and their corresponding error exponents E_{fp} and E_{fn}, be defined as in eqs. (1), (2), (3) and (4), respectively. Let $\bar{\mathbf{w}} = (\bar{w}_1, \bar{w}_2, \bar{w}_3) \in \mathbb{R}^3$ be given, and let $\Lambda = \{\mathbf{s} : \ \hat{\rho}^2_{\mathbf{us}} \geq 1 - e^{-2\lambda}\}$. Then,

$$E_{fn} = \min_{r\in\mathbb{R}^+} \min_{\bar{x}_1\in\mathbb{R}} \min_{(\bar{z}_1,\bar{z}_2,\bar{z}_3)\in\mathbb{R}^3} \min_{q\geq[T_1]_+} \left\{ S\left(\frac{q}{\sigma_Z^2}\right) + \frac{\bar{x}_1^2}{2\sigma_X^2} + S\left(\frac{r}{\sigma_X^2}\right) + \frac{\bar{z}_1^2+\bar{z}_2^2+\bar{z}_3^2}{2\sigma_Z^2} \right\}, \tag{8}$$

where $[u]_+ = \max\{0, u\}$, and

$$T_1 = T_1(r, \bar{\mathbf{w}}, \mathbf{t}) \quad = \quad (\bar{x}_1 + \bar{z}_1 + \bar{w}_1)^2 \tan^2\beta - (\sqrt{r} + \bar{w}_2 + \bar{z}_2)^2 - (\bar{w}_3 + \bar{z}_3)^2.$$

Proof. From (6), a false–negative event occurs whenever

$$\frac{(x_1 + w_1 + z_1)^2}{(x_1 + w_1 + z_1)^2 + (x_2 + w_2 + z_2)^2 + (w_3 + z_3)^2 + \sum_{j=4}^n z_j^2} < \cos^2\beta,$$

where $w_1^2 + w_2^2 + w_3^2 \leq nD$. This is equivalently to:

$$\begin{aligned} & (x_1 + \sqrt{n}\bar{w}_1 + z_1)^2 \tan^2\beta - (x_2 + \sqrt{n}\bar{w}_2 + z_2)^2 - (\sqrt{n}\bar{w}_3 + z_3)^2 \\ = \quad & (x_1 + \sqrt{n}\bar{w}_1 + z_1)^2 \tan^2\beta - \\ & \left[\sqrt{nr} + \sqrt{n}\bar{w}_2 + z_2\right]^2 - (\sqrt{n}\bar{w}_3 + z_3)^2 < \sum_{j=4}^n z_j^2 = nq, \end{aligned}$$

where

$$\begin{aligned} r \quad &\triangleq \quad \frac{1}{n}\sum_{j=2}^n x_j^2 = \frac{x_2^2}{n}, \\ q \quad &\triangleq \quad \frac{1}{n}\sum_{j=4}^n z_j^2. \end{aligned}$$

By defining

$$\begin{aligned} \bar{\mathbf{x}} \quad &\triangleq \quad \frac{\mathbf{x}}{\sqrt{n}}, \\ \bar{\mathbf{z}} \quad &\triangleq \quad \frac{\mathbf{z}}{\sqrt{n}}, \\ T \quad &\triangleq \quad (\bar{x}_1 + \bar{w}_1 + \bar{z}_1)^2 \tan^2\beta - (\sqrt{r} + \bar{w}_2 + \bar{z}_2)^2 - (\bar{w}_3 + \bar{z}_3)^2, \end{aligned} \tag{9}$$

a false negative event is now defined by the condition $q > T$. Next, observe that $\frac{nQ}{\sigma_Z^2}$, where Q designates the random variable associated with q, is a χ^2 random variable with $n-3$ degrees of freedom, i.e., its density is given by

$$f_Q(q) = \begin{cases} \frac{n}{\sigma_Z^2}\left(\frac{1}{2}\right)^{(n-3)/2} \frac{1}{\Gamma\left(\frac{n-3}{2}\right)} \left(\frac{nq}{\sigma_Z^2}\right)^{\left(\frac{n-3}{2}-1\right)} \exp\left\{-\frac{nq}{2\sigma_Z^2}\right\}, & \text{if } q \geq 0 \\ 0, & \text{elsewhere} \end{cases}. \tag{10}$$

By the same token, $\frac{nR}{\sigma_X^2}$, where R is the random variable associated with r, is a χ^2 distribution with $n-1$ degrees of freedom, and so

$$f_R(r) = \begin{cases} \frac{n}{\sigma_X^2}\left(\frac{1}{2}\right)^{(n-1)/2} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{nr}{\sigma_X^2}\right)^{\left(\frac{n-1}{2}-1\right)} \exp\left\{-\frac{nr}{2\sigma_X^2}\right\}, & \text{if } r \geq 0 \\ 0, & \text{elsewhere} \end{cases}. \tag{11}$$

On the other hand,

$$f_{\bar{X}_1}(\bar{x}_1) = \frac{\sqrt{n}\exp\left\{-n\bar{x}_1^2/(2\sigma_X^2)\right\}}{\sqrt{2\pi\sigma_X^2}},$$

and, equivalently,

$$f_{\bar{Z}_i}(\bar{z}_i) = \frac{\sqrt{n}\exp\left\{-n\bar{z}_i^2/(2\sigma_Z^2)\right\}}{\sqrt{2\pi\sigma_Z^2}},$$

where $1 \leq i \leq 3$. Therefore, the probability of false negative is given by:

$$\begin{aligned} P_{fn} &= \int\limits_{\mathbb{R}^+} dr \int\limits_{\mathbb{R}} d\bar{x}_1 \int\limits_{\mathbb{R}^3} d\bar{z}_1 d\bar{z}_2 d\bar{z}_3 \int\limits_{[T]_+}^{\infty} dq \frac{n}{\sigma_Z^2}\left(\frac{1}{2}\right)^{(n-3)/2} \frac{1}{\Gamma\left(\frac{n-3}{2}\right)} \left(\frac{nq}{\sigma_Z^2}\right)^{\left(\frac{n-3}{2}-1\right)} \exp\left\{-\frac{nq}{2\sigma_Z^2}\right\} \\ &\quad \frac{n^{3/2}\exp\left\{-\frac{n(\bar{z}_1^2+\bar{z}_2^2+\bar{z}_3^2)}{2\sigma_Z^2}\right\}}{(2\pi\sigma_Z^2)^{3/2}} \frac{n}{\sigma_X^2}\left(\frac{1}{2}\right)^{(n-1)/2} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{nr}{\sigma_X^2}\right)^{\left(\frac{n-1}{2}-1\right)} \exp\left\{-\frac{nr}{2\sigma_X^2}\right\} \frac{\sqrt{n}\exp\left\{-\frac{n\bar{x}_1^2}{2\sigma_X^2}\right\}}{\sqrt{2\pi\sigma_X^2}}. \end{aligned}$$

The last integral becomes

$$\begin{aligned} \lim_{n\to\infty} -\frac{1}{n}\ln P_{fn} &= -\frac{1}{2} - \frac{1}{2} - \lim_{n\to\infty}\frac{1}{n}\ln \int\limits_{\mathbb{R}^+} dr \int\limits_{\mathbb{R}} d\bar{x}_1 \int\limits_{\mathbb{R}^3} d\bar{z}_1 d\bar{z}_2 d\bar{z}_3 \int\limits_{[T_1]_+}^{\infty} dq \ \exp\left\{-\frac{n(\bar{z}_1^2+\bar{z}_2^2+\bar{z}_3^2)}{2\sigma_Z^2}\right\} \\ &\quad \exp\left\{\left(\frac{n-3}{2}-1\right)\ln\left(\frac{q}{\sigma_Z^2}\right) - \frac{nq}{2\sigma_Z^2} + \left(\frac{n-1}{2}-1\right)\ln\frac{r}{\sigma_X^2} - \frac{nr}{2\sigma_X^2} - \frac{n\bar{x}_1^2}{2\sigma_X^2}\right\}, \end{aligned}$$

where we used the fact that

$$\lim_{n\to\infty}\frac{1}{n}\ln\left[\frac{(1/2)^{\frac{n}{2}} n^{\frac{n}{2}}}{\Gamma(n/2)}\right] = \frac{1}{2}.$$

Finally, by using Laplace method of integration (see, e.g., [24]), we observe that the exponential rate of this multi–dimensional integral is dominated by the point at which the integrand is maximum, thus obtaining the result asserted in the theorem and completing the proof.

V. Optimum Watermark embedding

Having calculated E_{fn} as a function of $\hat{\mathbf{w}}$, we can now characterize a class of asymptotically optimum embedding functions, i.e., those that maximize E_{fn}. To this end, we must take into account that the embedder has access to the host signal, but not to the attacking signal (the noise). Formally, we can write the false–negative error exponent when the embedder designs the watermarking displacement signal trying to maximize E_{fn} as

$$E_{fn} = \min_{r\in\mathbb{R}^+}\min_{\bar{x}_1\in\mathbb{R}}\max_{\hat{\mathbf{w}}:||\hat{\mathbf{w}}||^2\le D}\min_{\bar{\mathbf{z}}\in\mathbb{R}^3}\min_{q\ge[T]_+}\left\{S\left(\frac{q}{\sigma_Z^2}\right)+S\left(\frac{r}{\sigma_X^2}\right)+\frac{\bar{x}_1^2}{2\sigma_X^2}+\frac{\|\mathbf{z}\|^2}{2\sigma_Z^2}\right\}. \tag{12}$$

Note that the dependence of E_{fn} on $(\bar{w}_1,\bar{w}_2)$ is through T only.

From this formula, we can derive the following conclusions about the optimal values of $\bar{x}_1,\bar{w}_3$ and $\bar{z}_3$, henceforth denoted $\bar{x}_1^*,\bar{w}_3^*,\bar{z}_3^*$, respectively:

- $\bar{x}_1^*=0$: given the definition of T, and the fact that the embedder knows the host signal when computing the watermarking displacement signal, $\bar{w}_1$ could be chosen so to have the same sign of $\bar{x}_1$, and the embedder would take advantage of any value of $\bar{x}_1\neq 0$ to maximize T, and therefore maximize E_{fn}. Formally,

$$\begin{aligned}
&\max_{\hat{\mathbf{w}}:||\hat{\mathbf{w}}||^2\le D}\min_{\bar{\mathbf{z}}\in\mathbb{R}^3}\min_{q\ge[T]_+}\left\{S\left(\frac{q}{\sigma_Z^2}\right)+S\left(\frac{r}{\sigma_X^2}\right)+\frac{\bar{x}_1^2}{2\sigma_X^2}+\frac{\|\bar{\mathbf{z}}\|^2}{2\sigma_Z^2}\right\}\\
\ge\ &\max_{\hat{\mathbf{w}}:||\hat{\mathbf{w}}||^2\le D}\min_{\bar{\mathbf{z}}\in\mathbb{R}^3}\min_{q\ge[T]_+}\left\{S\left(\frac{q}{\sigma_Z^2}\right)+S\left(\frac{r}{\sigma_X^2}\right)+\frac{\|\bar{\mathbf{z}}\|^2}{2\sigma_Z^2}\right\}\\
\ge\ &\max_{\hat{\mathbf{w}}:||\hat{\mathbf{w}}||^2\le D}\min_{\bar{\mathbf{z}}\in\mathbb{R}^3}\min_{q\ge[T_2]_+}\left\{S\left(\frac{q}{\sigma_Z^2}\right)+S\left(\frac{r}{\sigma_X^2}\right)+\frac{\|\bar{\mathbf{z}}\|^2}{2\sigma_Z^2}\right\},
\end{aligned}$$

  where

$$T_2 \ \triangleq\ (\bar{w}_1+\bar{z}_1)^2\tan^2\beta-(\sqrt{r}+\bar{w}_2+\bar{z}_2)^2-(\bar{w}_3+\bar{z}_3)^2,$$

  and the second inequality is based on the fact that $T\ge T_2$, as the embedder would select $\bar{w}_1$ to be of the same sign as $\bar{x}_1$, so that for any $\bar{z}_1$ there exists $\bar{z}_1'$ for which $\bar{z}_1^2\ge(\bar{z}_1')^2$ and $(\bar{x}_1+\bar{w}_1+\bar{z}_1)^2\ge(\bar{w}_1+\bar{z}_1')^2$. Note that equality between the first and the last expressions is only achieved when $\bar{x}_1=0$.
- $\bar{w}_3^*=0$: According to the definition of T and the fact that the determination of the worst noise takes into account the choice of the watermark, $\bar{z}_3$ could be chosen to have the same sign as $\bar{w}_3$. Therefore, any value of $\bar{w}_3\neq 0$ would yield a smaller minimum T, which is not desired by the embedder.
  Specifically, let T_3 be the value of T when $\bar{w}_3=0$. We have:

$$T_3 \ \triangleq\ (\bar{x}_1+\bar{w}_1+\bar{z}_1)^2\tan^2\beta-(\sqrt{r}+\bar{w}_2+\bar{z}_2)^2-(\bar{z}_3)^2.$$

  Given that in the optimization of (12), one has the freedom to choose the sign of $\bar{z}_3$, it is clear that the selected value will satisfy $\bar{z}_3\cdot\bar{w}_3\ge 0$, so $(\bar{w}_3+\bar{z}_3)^2\ge(\bar{z}_3)^2$, and consequently $T\le T_3$, achieving equality only when $\bar{w}_3^*=0$.
- $\bar{z}_3^*=0$. We calculate the $\bar{z}_3$ that minimizes T_3, subject to the constraint $\bar{z}_2^2+\bar{z}_3^2=K$, for any arbitrary

budget (i.e., values providing a fixed value of the last term of (12)) available for $\bar{z}_2$ and $\bar{z}_3$, K. To this end, we consider the fact that $\mathbf{z}$ is chosen based on the knowledge of $\mathbf{w}$ and $\mathbf{x}$, so $\bar{z}_2$ can be chosen to have the same sign as $\sqrt{r}+\bar{w}_2$, as this is the sign that minimizes T. Therefore, we can write T_3 as

$$\begin{aligned} T_3 &= (\bar{x}_1+\bar{w}_1+\bar{z}_1)^2\tan^2\beta - \left(\sqrt{r}+\bar{w}_2+\sqrt{K-\bar{z}_3^2}\right)^2 - (\bar{z}_3)^2 \\ &= (\bar{x}_1+\bar{w}_1+\bar{z}_1)^2\tan^2\beta - (\sqrt{r}+\bar{w}_2)^2 - K + \bar{z}_3^2 - 2(\sqrt{r}+\bar{w}_2)\sqrt{K-\bar{z}_3^2} - \bar{z}_3^2, \end{aligned}$$

which is obviously minimized when $\bar{z}_3 = 0$.

Incorporating these facts, eq. (12) can be rewritten as

$$E_{fn} = \min_{r\in\mathbb{R}^+} \max_{\bar{w}_1,\bar{w}_2:\bar{w}_1^2+\bar{w}_2^2\le D} \min_{(\bar{z}_1,\bar{z}_2)\in\mathbb{R}^2} \min_{q\ge[T]_+} \left\{ S\left(\frac{q}{\sigma_Z^2}\right) + S\left(\frac{r}{\sigma_X^2}\right) + \frac{\bar{z}_1^2+\bar{z}_2^2}{2\sigma_Z^2} \right\}, \tag{13}$$

where now

$$T = (\bar{w}_1+\bar{z}_1)^2\tan^2\beta - (\sqrt{r}+\bar{w}_2+\bar{z}_2)^2. \tag{14}$$

The most important conclusion from (13) is that, in general, the asymptotically optimum watermarking displacement signal depends on σ_X^2 and σ_Z^2. This implies that the watermark embedding strategy that solves (13) is not universal.

A. Optimum watermark embedding in high SNR regime

An interesting situation takes place when the variance of the attacking noise is much smaller than the variance of the host sequence, i.e., $\sigma_Z^2 << \sigma_X^2$, which we refer to as the *high SNR regime*. The high SNR regime is motivated by situations of non–malicious attacks, where the modification of the watermarked signal is very small compared with the host signal. For fixed (but arbitrary) σ_X^2, the high SNR regime is, of course, equivalent to a vanishing σ_Z^2. It should be noted that the high SNR regime poses limitations neither on the value of σ_X^2 nor on its ratio to D. As it will be shown below, our proposed class of optimum embedding strategies does not depend on these quantities, so it is universal in that sense (similarly as in [15]).

Since the target function in (13) is monotonically decreasing with σ_Z^2 for a given $(r,\bar{w}_1,\bar{w}_2,\bar{z}_1,\bar{z}_2)$, E_{fn} itself is monotonically decreasing with σ_Z^2. Therefore, the limit of E_{fn} as $\sigma_Z^2\to 0$, whether finite or infinite, must exist. The following theorem asserts that E_{fn} converges to a finite limit and a universally optimum embedding rule exists in the large n limit.

Theorem 2: In the high SNR regime, i.e. $\frac{\sigma_X^2}{\sigma_Z^2}\to\infty$, the maximum false–negative exponent, subject to the constraint $\bar{w}_1^2+\bar{w}_2^2\le D$, is given by $S\left(\max\left\{1,\frac{D}{\sigma_X^2\cos^2\beta}\right\}\right)$, and it is attained by the set of equally optimal embedding strategies defined by:

$$M(r) = \left\{ (\bar{w}_1,\bar{w}_2) : \bar{w}_1^2+\bar{w}_2^2\le D, \text{and if } r < \frac{D}{\cos^2\beta} \text{ then } \bar{w}_1^2 > \frac{1}{\tan^2\beta}\left[\sqrt{r}+\bar{w}_2\right]^2 \right\}. \tag{15}$$

Proof.

For the sake of notational simplicity, we define $\xi \triangleq \frac{\sigma_X^2}{\sigma_Z^2}$. We are interested in $\lim_{\xi\to\infty} E_{fn}$, which we denote as $E_{fn}^{\text{high-SNR}}$. We also define

$$f(r,\bar{z}_1,\bar{z}_2,q) = S\left(\frac{\xi q}{\sigma_X^2}\right) + S\left(\frac{r}{\sigma_X^2}\right) + \frac{\xi(\bar{z}_1^2+\bar{z}_2^2)}{2\sigma_X^2}. \tag{16}$$

The proof is based on the following chain of inequalities:

$$\begin{aligned}
& \min_{r\in\mathbb{R}^+} \min_{(\bar{z}_1,\bar{z}_2)\in\mathbb{R}^2} \min_{q\geq[T(\bar{w}_1^*(r),\bar{w}_2^*(r))]_+} f(r,\bar{z}_1,\bar{z}_2,q) \\
\leq\ & \min_{r\in\mathbb{R}^+} \max_{\bar{w}_1,\bar{w}_2:\bar{w}_1^2+\bar{w}_2^2\leq D} \min_{(\bar{z}_1,\bar{z}_2)\in\mathbb{R}^2} \min_{q\geq[T(\bar{w}_1,\bar{w}_2)]_+} f(r,\bar{z}_1,\bar{z}_2,q) \\
=\ & E_{fn} \\
\leq\ & \min_{r\in\mathbb{R}^+} \max_{\bar{w}_1,\bar{w}_2:\bar{w}_1^2+\bar{w}_2^2\leq D} \min_{q\geq[T(\bar{w}_1,\bar{w}_2)]_+} f(r,0,0,q) \\
\leq\ & \max_{\bar{w}_1,\bar{w}_2:\bar{w}_1^2+\bar{w}_2^2\leq D} \min_{q\geq[T(\bar{w}_1,\bar{w}_2)]_+} f(r^*,0,0,q),
\end{aligned} \tag{17}$$

where we have made explicit the dependency of T upon $(\bar{w}_1,\bar{w}_2)$, and where $r^* = \max\left\{\sigma_X^2, \frac{D}{\cos^2\beta}\right\}$. Here, $(\bar{w}_1^*(r),\bar{w}_2^*(r))$ stands for an arbitrary embedding rule in $M(r)$. When $r < \frac{D}{\cos^2\beta}$, the set $M(r)$ is nonempty. As an example, for the embedding strategy $(\bar{w}_1,\bar{w}_2) = (\text{sign}(x_1)|\sqrt{D-r\cos^4\beta}|, -\sqrt{r}\cos^2\beta)$,[6] the constraint $\bar{w}_1^2 > \frac{1}{\tan^2\beta}\left[\sqrt{r}+\bar{w}_2\right]^2$ can be rewritten as

$$(D - r\cos^4\beta)\tan^2\beta > r\sin^4\beta, \tag{18}$$

which is equivalent to $D\tan^2\beta > r\sin^2\beta$, that always holds whenever $r < \frac{D}{\cos^2\beta}$. On the other hand, when $r \geq \frac{D}{\cos^2\beta}$, any embedding strategy that meets the distortion constraint belongs to $M(r)$. We first prove that all these embedding rules satisfy $\bar{w}_1^2 \leq \frac{1}{\tan^2\beta}\left[\sqrt{r}+\bar{w}_2\right]^2$. To this end, we use a *reductio ad absurdum* argument: assume, conversely, that there is at least one embedding rule $(\bar{w}_1,\bar{w}_2)$ such that $\bar{w}_1^2+\bar{w}_2^2 \leq D$ and

$$\bar{w}_1^2 > \frac{1}{\tan^2\beta}\left[\sqrt{r}+\bar{w}_2\right]^2. \tag{19}$$

Since $|\bar{w}_2| \leq \frac{\sqrt{D}}{\cos\beta} \leq \sqrt{r}$, (19) is equivalent to $|\bar{w}_1|\tan\beta - \bar{w}_2 > \sqrt{r}$. Note that due to the embedding power constraint and its monotonicity in $\bar{w}_1$, $\max_{(\bar{w}_1,\bar{w}_2):\bar{w}_1^2+\bar{w}_2^2\leq D}|\bar{w}_1|\tan\beta - \bar{w}_2$ is equivalent to $\max_{\bar{w}_1}|\bar{w}_1|\tan\beta + |\sqrt{D-\bar{w}_1^2}|$. The solution to this optimization problem is $\bar{w}_1 = \pm\sqrt{D}\sin\beta$, being the value of the target function $\frac{\sqrt{D}}{\cos\beta}$. Therefore, on the one hand, we have that for any $(\bar{w}_1,\bar{w}_2)$ satisfying the distortion constraint, $|\bar{w}_1|\tan\beta - \bar{w}_2 \leq \frac{\sqrt{D}}{\cos\beta}$, whereas on the other hand, from (19) we can say that $|\bar{w}_1|\tan\beta - \bar{w}_2 > \sqrt{r} \geq \frac{\sqrt{D}}{\cos\beta}$, proving that whenever $r \geq \frac{D}{\cos^2\beta}$ and $\bar{w}_1^2+\bar{w}_2^2 \leq D$, then $\bar{w}_1^2 \leq \frac{1}{\tan^2\beta}\left[\sqrt{r}+\bar{w}_2\right]^2$, regardless of $(\bar{w}_1,\bar{w}_2)$.

From an intuitive point of view, the previous derivation means that if one fixes in the optimization described in (13) $\bar{z}_1$ and $\bar{z}_2$ to be null (and consequently obtains an upper bound of (13)), then the detection region is the

[6]The sign$(\cdot)$ function value is $+1$ or -1, depending on the sign of its argument; if its argument were 0, then $+1$ or -1 can be arbitrarily returned. This choice of the sign of $\bar{w}_1$ is related to the first bullet of Sect. V.

hypercone

$$(\bar{w}_1)^2 \tan^2 \beta - (\sqrt{r} + \bar{w}_2)^2 \geq 0,$$

or equivalently

$$|\bar{w}_1| \tan \beta - \bar{w}_2 \geq \sqrt{r},$$

where we have assumed that $\sqrt{r} \geq |\bar{w}_2|$. In that case, whenever $r \geq \frac{D}{\cos^2 \beta}$ the host signal is too large, in the sense that the embedder will not have power enough to produce a watermarked signal in the detection region.

a) Upper bound: First, we study the behavior of $\max_{\bar{w}_1, \bar{w}_2 : \bar{w}_1^2 + \bar{w}_2^2 \leq D} \min_{q \geq [T(\bar{w}_1, \bar{w}_2)]_+} f(r^*, 0, 0, q)$. Here, the optimization problem can be written as

$$\max_{\bar{w}_1, \bar{w}_2 : \bar{w}_1^2 + \bar{w}_2^2 \leq D} \min_{q \geq [T]_+} S\left(\frac{\xi q}{\sigma_X^2}\right) + S\left(\frac{r^*}{\sigma_X^2}\right), \tag{20}$$

or equivalently,

$$\max_{\bar{w}_1, \bar{w}_2 : \bar{w}_1^2 + \bar{w}_2^2 \leq D} S\left(\max\left[\frac{\xi T}{\sigma_X^2}, 1\right]\right) + S\left(\frac{r^*}{\sigma_X^2}\right), \tag{21}$$

where

$$T \quad = \quad \bar{w}_1^2 \tan^2 \beta - (\sqrt{r^*} + \bar{w}_2)^2. \tag{22}$$

First, we prove that (21) vanishes whenever $\frac{D}{\cos^2 \beta} < \sigma_X^2$. To this end, note that $r^* = \sigma_X^2$, and as was shown above, $|\bar{w}_1| \tan \beta - \bar{w}_2 \leq \frac{\sqrt{D}}{\cos \beta}$ for any embedding strategy satisfying the distortion constraint, which yields $T < 0$ for any $(\bar{w}_1, \bar{w}_2)$. Considering both results together, $r^* = \sigma_X^2$ and $T < 0$, provide a null value for (21). For this reason, in the following we assume that $\frac{D}{\cos^2 \beta} \geq \sigma_X^2$.

Maximization of (21) is equivalent to maximize (22). Therefore, when $\frac{D}{\cos^2 \beta} \geq \sigma_X^2$, the embedding strategies $(\bar{w}_1(r^*), \bar{w}_2(r^*))$ solving (21) must satisfy

$$(\bar{w}_1^{opt}, \bar{w}_2^{opt}) = \arg \max_{(\bar{w}_1, \bar{w}_2) : \bar{w}_1^2 + \bar{w}_2^2 \leq D} \bar{w}_1^2 \tan^2 \beta - (\sqrt{r^*} + \bar{w}_2)^2. \tag{23}$$

Since the target function is monotonically increasing with $\bar{w}_1^2$, the maximum is achieved for $\bar{w}_1^2 + \bar{w}_2^2 = D$, which allows to represent the optimization problem as

$$(\bar{w}_1^{opt}, \bar{w}_2^{opt}) = \arg \max_{(\bar{w}_1, \bar{w}_2) : \bar{w}_1^2 + \bar{w}_2^2 \leq D} \bar{w}_1^2 \tan^2 \beta - \left[\sqrt{r^*} - \sqrt{D - \bar{w}_1^2}\right]^2. \tag{24}$$

Equating the partial derivative of the target function with respect to $\bar{w}_1$ to zero, and solving for $\bar{w}_1$, we obtain three

solutions:

$$\begin{cases} \bar{w}_1 = 0 \\ \bar{w}_1 = -\sqrt{D - r^* \cos^4 \beta} \\ \bar{w}_1 = \sqrt{D - r^* \cos^4 \beta} \end{cases} . \tag{25}$$

Considering the second partial derivative, we see that for $\bar{w}_1^{opt} = \pm\sqrt{D - r^* \cos^4 \beta} = \pm\sqrt{D} \sin \beta$ one obtains maxima of the target function, yielding $\bar{w}_2^{opt} = -\sqrt{r^*} \cos^2 \beta = -\sqrt{D} \cos \beta$, and a corresponding value of $T = D \tan^2 \beta - r^* \sin^2 \beta = 0$, for any value of ξ.

Summarizing, in this part we have proven that for any embedding strategy satisfying the distortion constraint, the false–negative is upper bounded by $S\left(\frac{r^*}{\sigma_X^2}\right)$.

b) Lower bound: Consider now the problem

$$\min_{r \in \mathbb{R}^+} \min_{(\bar{z}_1, \bar{z}_2) \in \mathbb{R}^2} \min_{q \geq [T(\bar{w}_1^*(r), \bar{w}_2^*(r))]_+} f(r, \bar{z}_1, \bar{z}_2, q). \tag{26}$$

Defining $K_1 \triangleq \xi \bar{z}_1^2$, $K_2 \triangleq \xi \bar{z}_2^2$, $\eta_1 \triangleq \text{sgn}(\bar{z}_1)$, $\eta_2 \triangleq \text{sgn}(\bar{z}_2)$, this optimization problem is equivalent to

$$\begin{aligned} &\min_{r \in \mathbb{R}^+} \min_{(\eta_1, \eta_2) \in \{-1,+1\}^2} \min_{(K_1, K_2) \in (\mathbb{R}^+)^2} \frac{K_1 + K_2}{2\sigma_X^2} + S\left(\frac{r}{\sigma_X^2}\right) + \\ &S\left(\max\left\{1, \frac{\xi}{\sigma_X^2}\left[\left(\bar{w}_1^*(r) + \eta_1\sqrt{\frac{K_1}{\xi}}\right)^2 \tan^2\beta - \left(\sqrt{r} + \bar{w}_2^*(r) + \eta_2\sqrt{\frac{K_2}{\xi}}\right)^2\right]\right\}\right). \end{aligned} \tag{27}$$

where we have used the fact that

$$\min_{q \geq T} S\left(\frac{\xi q}{\sigma_X^2}\right) = S\left(\max\left\{1, \frac{\xi T}{\sigma_X^2}\right\}\right). \tag{28}$$

As was proven in the derivation of the upper bound, the false–negative error exponent is bounded, independently of ξ, by a finite constant, which we shall denote by E_{fn}^u. Since the lower bound on E_{fn}, in eq. (27), is the sum of three non-negative terms, the first of which increases without bound as K_1 and/or K_2 go to ∞ the existence of a uniform upper bound, E_{fn}^u, implies that a necessary condition for a point $(r, \eta_1, \eta_2, K_1, K_2)$ to solve the minimization problem (27) is that each term of (27) is smaller than or equal to E_{fn}^u. Applying this consideration to the term $K_1/(2\sigma_X^2)$, we have $\frac{K_1}{2\sigma_X^2} \leq E_{fn}^u$, hence enabling us to confine the search over K_1 to the interval $[0, K_u]$, where $K_u = 2E_{fn}^u \sigma_X^2$. The same comment applies, of course, to K_2. Consequently, the lower bound in (27) is equivalent to

$$\begin{aligned} &\min_{r \in \mathbb{R}^+} \min_{(\eta_1, \eta_2) \in \{-1,+1\}^2} \min_{(K_1, K_2) \in [0, K_u]^2} \frac{K_1 + K_2}{2\sigma_X^2} + S\left(\frac{r}{\sigma_X^2}\right) + \\ &S\left(\max\left\{1, \frac{\xi}{\sigma_X^2}\left[\left(\bar{w}_1^*(r) + \eta_1\sqrt{\frac{K_1}{\xi}}\right)^2 \tan^2\beta - \left(\sqrt{r} + \bar{w}_2^*(r) + \eta_2\sqrt{\frac{K_2}{\xi}}\right)^2\right]\right\}\right). \end{aligned} \tag{29}$$

Now, the second argument of the max operator is quadratic in $\sqrt{\xi}$, i.e., it is of the form as $a_2\xi + a_1\sqrt{\xi} + a_0$, where a_0, a_1 and a_2 are independent of ξ. Therefore, there exists a value of ξ, which we will denote by ξ_0, such that

$a_2\xi + a_1\sqrt{\xi} + a_0$ is either monotonically increasing for all $\xi \geq \xi_0$, or monotonically decreasing and less than unity for all $\xi \geq \xi_0$, depending on the signs of a_1 and a_2.[7] Accordingly, for any $\xi \geq \xi_0$, $\max\{1, a_2\xi + a_1\sqrt{\xi} + a_0\}$ is either strictly larger than one and monotonically increasing (in the former case), or it is equal 1 (in the latter case). In either case, it is monotonically non–decreasing. Considering the fact that the function $S(x)$ is monotonically increasing for $x \geq 1$, the target function in (29) is monotonically non–decreasing for $\xi \geq \xi_0$. Thus, as $\xi \to \infty$, this function has a limit (finite or infinite) for any fixed $(r, \eta_1, \eta_2, K_1, K_2)$. The same applies to the behavior of (29) as ξ goes to infinity. As (29) is known to be upper bounded by E_{fn}^u for any ξ, its limit must be finite.

Let us first assume that there are arbitrarily large values of ξ for which the solution to (29) satisfies $r < \frac{D}{\cos^2\beta}$. Then, by the definition of $M(r)$, $\bar{w}_1^2 \tan^2\beta - [\sqrt{r} + \bar{w}_2]^2 > 0$. On the other hand,

$$\begin{aligned}\lim_{\xi\to\infty} &\left(\bar{w}_1^*(r) + \eta_1\sqrt{\tfrac{K_1}{\xi}}\right)^2 \tan^2\beta - \left(\sqrt{r} + \bar{w}_2^*(r) + \eta_2\sqrt{\tfrac{K_2}{\xi}}\right)^2 \\ &= \left(\bar{w}_1^*(r)\right)^2 \tan^2\beta - \left(\sqrt{r} + \bar{w}_2^*(r)\right)^2, \end{aligned} \tag{30}$$

where we have taken into account that both K_1 and K_2 are bounded by $K_u < \infty$. Therefore, the right argument of the $\max$ operator in (29) would grow without bound as $\xi \to \infty$, yielding an unbounded value of (29) when ξ goes to infinity. However, (29) is upper bounded by E_{fn}^u irrespectively of ξ, which is a contradiction. Thus, for all sufficiently large ξ, the solution to (29) must satisfy $r \geq \frac{D}{\cos^2\beta}$. We can then rewrite (29) as

$$\begin{aligned}&\lim_{\xi\to\infty} \min_{r\in\mathbb{R}^+} \min_{(\eta_1,\eta_2)\in\{-1,+1\}^2} \min_{(K_1,K_2)\in[0,K^u]^2} \frac{K_1+K_2}{2\sigma_X^2} + S\left(\frac{r}{\sigma_X^2}\right) + \\ &S\left(\max\left\{1, \frac{\xi}{\sigma_X^2}\left[\left(\bar{w}_1^*(r) + \eta_1\sqrt{\frac{K_1}{\xi}}\right)^2 \tan^2\beta - \left(\sqrt{r} + \bar{w}_2^*(r) + \eta_2\sqrt{\frac{K_2}{\xi}}\right)^2\right]\right\}\right) \end{aligned} \tag{31}$$

$$\begin{aligned}&= \lim_{\xi\to\infty} \min_{r\geq\frac{D}{\cos^2\beta}} \min_{(\eta_1,\eta_2)\in\{-1,+1\}^2} \min_{(K_1,K_2)\in[0,K^u]^2} \frac{K_1+K_2}{2\sigma_X^2} + S\left(\frac{r}{\sigma_X^2}\right) + \\ &S\left(\max\left\{1, \frac{\xi}{\sigma_X^2}\left[\left(\bar{w}_1^*(r) + \eta_1\sqrt{\frac{K_1}{\xi}}\right)^2 \tan^2\beta - \left(\sqrt{r} + \bar{w}_2^*(r) + \eta_2\sqrt{\frac{K_2}{\xi}}\right)^2\right]\right\}\right) \end{aligned} \tag{32}$$

$$\geq \min_{r\geq\frac{D}{\cos^2\beta}} \min_{(\eta_1,\eta_2)\in\{-1,+1\}^2} \min_{(K_1,K_2)\in[0,K^u]^2} \frac{K_1+K_2}{2\sigma_X^2} + S\left(\frac{r}{\sigma_X^2}\right), \tag{33}$$

whose solution has $K_1 = K_2 = 0$, independently of η_1 and η_2, and $r = \sigma_X^2$ whenever $\sigma_X^2 \geq \frac{D}{\cos^2\beta}$, or $r = \frac{D}{\cos^2\beta}$ whenever $\sigma_X^2 < \frac{D}{\cos^2\beta}$. From an intuitive point of view, this result shows that having $\bar{z}_1 \neq 0$ or $\bar{z}_2 \neq 0$ in (13) is too expensive, in the sense of producing a large increase in the cost function (as σ_Z^2 is arbitrarily small), but not significantly modifying (14).

Summarizing, in this part we have proven that for any embedding strategy belonging to $M(r)$, the false–negative error exponent is lower bounded by $S\left(\frac{r^*}{\sigma_X^2}\right)$.

[7] If $a_2 < 0$, or if $a_2 = 0$ and $a_1 < 0$: $a_2\xi + a_1\sqrt{\xi} + a_0 \leq 1$, for any $\xi \geq \xi_0$. If $a_2 > 0$, or if $a_2 = 0$ and $a_1 \geq 0$: $a_2\xi + a_1\sqrt{\xi} + a_0$ is monotonically increasing for any $\xi \geq \xi_0$.

This asymptotic lower bound of the error exponent coincides with the upper bound previously derived, thus proving that this is the false–negative error exponent in the high–SNR scenario, and showing also the optimality of the embedding strategies described by $M(r)$ in the high–SNR scenario. This completes the proof of Theorem 2.

B. Optimum watermark embedding for small false negative error exponents

In the previous subsection we have characterized a family of embedding strategies that yields the optimum false–negative error exponent in the high-SNR scenario. A natural question that may arise is whether there is a particular embedding strategy in this family, which exhibits good performance not only in the high-SNR regime, but in more general situations. In this short subsection, we focus on the case where the false negative error exponent E_{fn} is very small. From eq. (13), we see that a necessary condition for E_{fn} to vanish is that $\bar{z}_1 \to 0$, $\bar{z}_2 \to 0$. This brings us back to the problem

$$(\bar{w}_1^*, \bar{w}_2^*) = \arg \max_{(\bar{w}_1, \bar{w}_2): \bar{w}_1^2 + \bar{w}_2^2 \leq D} (\bar{w}_1)^2 \tan^2 \beta - (\sqrt{r} + \bar{w}_2)^2, \tag{34}$$

that was studied in the derivation of the upper bound in the proof of Theorem 2, for $r = r^*$. For a general r, the solution is $\bar{w}_1^* = \pm\sqrt{D - r\cos^4\beta}$, and $\bar{w}_2^* = -\sqrt{r}\cos^2\beta$. In the next sections, we will see that this embedding rule has a nice geometrical interpretation, and most of all, it guarantees fairly good performance even when the high-SNR assumption does not hold.

C. Discussion

First of all, we will look at the false–negative error exponent in the high–SNR regime of the embedding strategies in $M(r)$ as a function of the false–positive error exponent λ. For the embedding strategies in $M(r)$, one can see that $S\left(\max\left\{1, \frac{D}{\sigma_X^2 \cos^2\beta}\right\}\right)$ is equivalent to

$$\lim_{\sigma_Z^2 \to 0} E_{fn}^* \quad = \quad \begin{cases} 0, & \text{if } \frac{D}{1-e^{-2\lambda}} \leq \sigma_X^2 \\ \frac{1}{2}\left[\frac{D}{\sigma_X^2(1-e^{-2\lambda})} - \ln\left(\frac{D}{\sigma_X^2(1-e^{-2\lambda})}\right) - 1\right] & \text{elsewhere} \end{cases} . \tag{35}$$

In view of (35), it is interesting to note that as long as $D > \sigma_X^2$, $E_{fn}^* > 0$ for any λ. In fact, under these conditions, the asymptotic value of E_{fn} when $\lambda \to \infty$ is

$$\frac{1}{2}\left[\frac{D}{\sigma_X^2} - \ln\left(\frac{D}{\sigma_X^2}\right) - 1\right], \tag{36}$$

coinciding with the result of [15] (Corollary 1).

On the other hand, when $D \leq \sigma_X^2$ another interesting point which reflects the goodness of the class of optimum strategies for the high-SNR regime is the computation of the range of values of λ for which $E_{fn} > 0$ can be achieved. In this case, the condition to be verified is

$$\frac{D}{1 - e^{-2\lambda}} > \sigma_X^2, \tag{37}$$

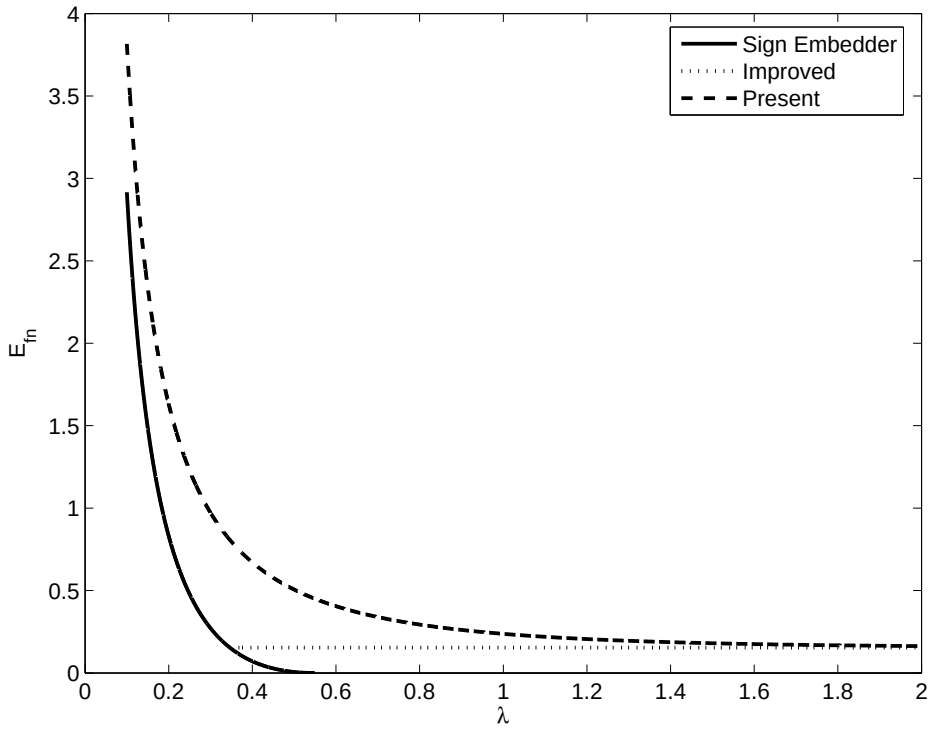

Fig. 1. Comparison of the error exponents obtained by the sign embedder described by Merhav and Sabbag [15], its improved version, and the technique presented in this work. $\sigma_X^2 = 1$ and $D = 2$.

implying that

$$\lambda < -\frac{1}{2}\ln\left(1 - \frac{D}{\sigma_X^2}\right) = \lambda_1, \text{ for } D \leq \sigma_X^2, \tag{38}$$

whereas for the sign embedder [15], the values of λ for which $E_{fn} > 0$ are those such that

$$\frac{D}{\sigma_X^2} > \frac{1 - e^{-2\lambda}}{e^{-2\lambda}}, \tag{39}$$

or, equivalently,

$$\lambda < -\frac{1}{2}\ln\left(\frac{\sigma_X^2}{D + \sigma_X^2}\right) = \lambda_2, \text{ for all } D. \tag{40}$$

Given that $\lambda_1 > \lambda_2$, larger values of false positive error exponents are allowed (while still keeping $E_{fn} > 0$) by the embedding rules in $M(r)$. In Figure 1 we compare the bounds on the false–negative exponent for the attack–free case found in [15], with the real value derived here. As it can be seen, the improvement brought by the optimum embedding strategies is significant, especially for small λ.

As we already saw in the general case, even in the high SNR and the small false negative error exponent regimes the optimum watermarking displacement signal $\mathbf{w}$, and therefore the watermarked sequence $\mathbf{y}$, lies in the plane spanned by the watermark $\mathbf{u}$ and the host signal $\mathbf{x}$. This allows us to express the optimum watermarking

displacement signal, as well as the watermarked sequence as a combination of the host signal and the watermark, leading to the following result:

Corollary 1: Whenever $D \geq r\cos^2\beta$, the optimum watermarked signal resulting from the embedding rule derived in section V-B is given by $\mathbf{y} = a\mathbf{x} + b\mathbf{u}$, with:

$$\begin{aligned} a &= 1 - \frac{\cos^2\beta}{\cos\alpha}, \\ b &= \sqrt{r}\cdot\tan\alpha\cos^2\beta + \text{sign}(x_1)\left|\sqrt{D - r\cos^4\beta}\right|, \\ \alpha &= \arcsin\left(\frac{<\mathbf{x},\mathbf{u}>}{||\mathbf{x}||\cdot||\mathbf{u}||}\right). \end{aligned}$$

Proof. From Theorem 2 and the result in Sect. V-B, we have:

$$\begin{aligned} y_1 &= \sqrt{nr}\sin\alpha + \text{sign}(x_1)\left|\sqrt{n(D - r\cos^4\beta)}\right|, \\ y_2 &= \sqrt{nr}[\cos\alpha - \cos^2\beta]. \end{aligned} \tag{41}$$

On the other hand, $y_2 = a\sqrt{nr}\cos\alpha$, and so, we can conclude that $a = 1 - \frac{\cos^2\beta}{\cos\alpha}$. To find b, we use $y_1 = a\sqrt{nr}\sin\alpha + b\sqrt{n}$, which when combined with (41), gives the value of b which is asserted in Corollary 1. This completes the proof of Corollary 1.

More importantly the optimum embedding strategy derived in Sect. V-B depends neither on σ_X^2 nor on σ_Z^2, that is the optimum embedding rule for the high SNR regime (and the low E_{fn} scenario) defines a universally optimum embedding rule. Furthermore, in the two asymptotic cases analyzed in Sect. V-A and V-B both $\bar{z}_1$ and $\bar{z}_2$ goes to zero, so T in (14) is just reduced to Miller's *et al.* [13] measure of robustness, geometrically interpreted by Furon and Bas in [18].

The geometrical interpretation of the embedding strategy derived in Sect. V-B is the following: the embedder devotes part of the allowed distortion budget to scale down the host signal, thus reducing its interference, and then injects the remaining energy in the direction of the watermark. Concretely, the watermarking displacement signal is orthogonal to the detection boundary until the watermarked signal is in the detection region, and then it is parallel to the detection region hypercone axis; due to this geometrical interpretation, we will denote the embedding strategy derived in Sect. V-B as *Orthogonal to the Boundary, and then Parallel to the Axis* (OBPA). This geometrical interpretation explains why, whenever the watermarked signal is within the detection region, only its component in the direction of the watermark (i.e., b) depends on D. For illustration, we compare OBPA strategy derived in this work, and the sign-embedder introduced in [15]. For the sign embedder, the watermarked signal is given by $\mathbf{y}_{se} = \mathbf{x} + \text{sign}(\mathbf{x}^t\cdot\mathbf{u})\sqrt{D}\mathbf{u}$, so the watermarking displacement signal can be written as $\mathbf{w}_{se} = \text{sign}(\mathbf{x}^t\cdot\mathbf{u})\sqrt{D}\mathbf{u}$. The two strategies are compared in Fig. 2, where it is easy to see that the OBPA strategy is that of minimizing the embedding distortion necessary for obtaining a watermarked signal. It is also interesting to observe that the new embedding technique we have introduced could not be described by [17], as in that case the watermarking displacement signal direction is just a function of the host signal, and it is scaled for obtaining the desired distortion.

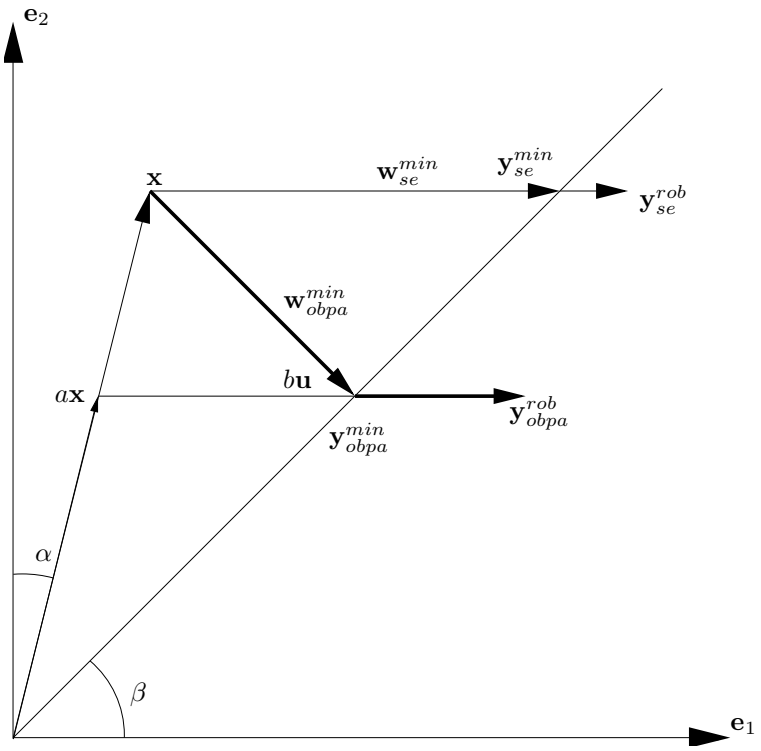

Fig. 2. Geometrical interpretation of the optimum embedding problem, and comparison between the sign-embedder and the OBPA embedder. $\mathbf{w}_{obpa}^{min}$ and $\mathbf{w}_{se}^{min}$ denote the minimum norm watermarking displacement signals that produce signals in the detection region, for both the OBPA embedder and the sign embedder, respectively. The corresponding watermarked signals are $\mathbf{y}_{obpa}^{min}$ and $\mathbf{y}_{se}^{min}$. $\mathbf{e}_1$ and $\mathbf{e}_2$ denote the first two basis vectors obtained by the Gram-Schmidt procedure described in Sect. IV, with $\mathbf{e}_1$ being proportional to $\mathbf{u}$. Furthermore, one can see the watermarked signals for the OBPA embedder and the sign embedder when part of the embedding distortion can be used to gain some robustness to noise (denoted by $\mathbf{y}_{obpa}^{rob}$ and $\mathbf{y}_{se}^{rob}$), and the composition of $\mathbf{y}_{obpa}^{min}$ as $a\mathbf{x} + b\mathbf{u}$.

Another way to look at Sect. V-B is by evaluating a joint condition on the embedding distortion and the false–positive exponent (or equivalently on β) that allows to obtain a positive false–negative error exponent: if $T \leq 0$, then the optimization on (r, q) in (13) is performed on the region $[0, \infty) \times [0, \infty)$, so any pair (σ_Z^2, σ_X^2), even with $\sigma_Z^2 = 0$, will be in the allowed region, yielding a vanishing error exponent. The condition that permits to avoid this situation is $r \leq \frac{D}{\cos^2 \beta}$. Indeed, when $D = r \cos^2 \beta$ the watermarked signal is the intersection of the boundary of the detection region and the perpendicular vector to that boundary that goes through $\mathbf{x}$. On the other hand, when $D < r \cos^2 \beta$, even in the high-SNR regime case, one cannot ensure that the embedding distortion constraint allows to produce a signal in the detection region, so the embedding function in that case will not be so important. In fact, regardless of the embedding function we choose, the false negative error exponent would vanish.

This last consideration also establishes a connection with the high-SNR analysis. Due to the absence of noise, the only source of false negative errors is that the embedding distortion is not enough for moving the host signal into the detection region, i.e. $D < r \cos^2 \beta$. Nevertheless, whenever $D > r \cos^2 \beta$ a set of equally optimal embedding strategies exists; indeed, all the embedding strategies able to move the host signal into the detection region with a minimum-normed distorting vector, i.e. moving the host signal to the detection boundary with distortion $r \cos^2 \beta$,

yield the same false–negative error exponent, regardless of the exact point where the watermarked signal lies inside the detection region. This explains why Theorem 2 describes a set of equally optimal embedding strategies. It is worth remarking that the OBPA embedding strategy belongs to the class of optimum embedding functions defined by Theorem 2; nevertheless, it is not the only example in the literature belonging to such class. For example, both the embedding strategy proposed by Merhav and Sabbag in [15], and that proposed by Furon and Bas in [18] when just one double hypercone is considered, satisfy the condition set by Theorem 2.

VI. Performance evaluation

Given a particular embedding strategy, equation (8) allows to numerically evaluate the corresponding false negative error exponent. In fact, the optimization problem expressed in equation (8) is rather easy to solve numerically given that it implies an optimization over three parameters only, namely r, $\bar{z}_1$ and $\bar{z}_2$, as the minimization over q is equivalent to compute $\max(\sigma_Z^2, T)$. Similarly, the computation of the false–negative error exponent for the optimum embedder in the general case, that, as it was mentioned above, will not yield a universal embedder (as it requires the knowledge of both σ_X^2 and σ_Z^2), is obtained as the solution of the optimization problem described in (13). In that case the number of involved parameters is four, namely r, $\bar{w}_1$, $\bar{z}_1$ and $\bar{z}_2$, since the maximum over $\bar{w}_2$ is achieved for $\bar{w}_2 = \sqrt{D - \bar{w}_1^2}$.

In the following, we show the results that we obtained by computing numerically the optimum (non–universal) false–negative error exponent, and compare them against the false–negative error exponent obtained with the OBPA embedding rule, and against those of two popular embedding rules, namely the sign embedder rule introduced in [15] and the Broken Arrows strategy introduced in [18]. For the latter method, and through the rest of the paper, we will focus on the particular case where just one double hypercone is considered.

In order to be able to clearly see the differences among the various embedding strategies, the values of λ should be large enough, or equivalently the values of β small, as for small values of λ all the considered strategies are asymptotically equivalent. Therefore, trying to analyze the behavior of the various schemes for large values of λ, Fig. 3 shows the false–negative error exponents when the host variance takes a very small value, concretely $\sigma_X^2 = 1$, for $D = 2$ and $\sigma_Z^2 = 1$. In this plot one can see that although the Broken Arrows strategy is slightly better than the OBPA embedding strategy for small λ, the situation completely changes for large values of λ. In effect, when λ is increased, and consequently E_{fn} is decreased, the optimal performance of the OBPA embedding strategy in that scenario is clearly observed. In fact, one can see that the OBPA strategy is asymptotically optimal for large values of λ (in the sense of those values yielding E_{fn} close to zero). It is also remarkable the good behavior of the new embedding strategy in the full range of considered values of λ, not only for the large values. Finally, as expected, the values of E_{fn} obtained for the optimal (non-universal) embedding strategy are always the largest ones.

The scenario considered in Fig. 3 and described in the previous paragraph is not a realistic one. Typically, $\sigma_X^2 >> D$ and $\sigma_X^2 >> \sigma_Z^2$. In order to assess the performance of OBPA in more practical setups, in Fig. 4 the false–negative error exponent is plotted as a function of λ when $\sigma_X^2 = 1$, $D = 0.1$, and $\sigma_Z^2 = 1$. As mentioned earlier and as intuition suggests, the maximum value of λ providing positive false–negative error exponent is much

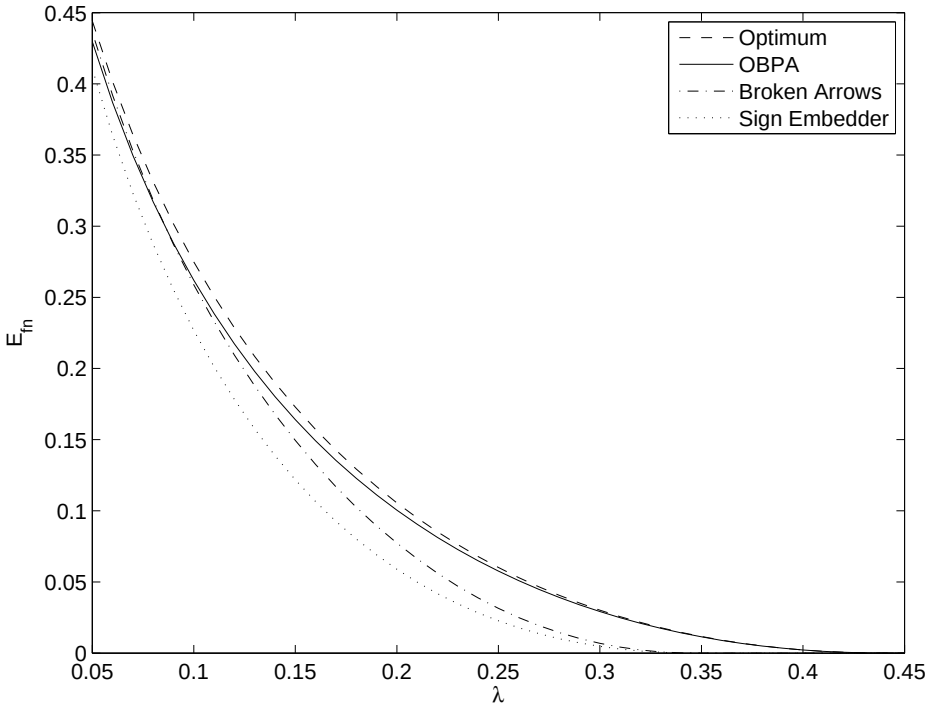

Fig. 3. Comparison of the errors exponents obtained by the sign embedder described by Merhav and Sabbag [15], the Broken Arrows strategy proposed by Furon and Bas [18] when just a secret direction is considered, the solution of (13), and the OBPA embedding technique. $\sigma_X^2 = 1$, $D = 2$, $\sigma_Z^2 = 1$.

smaller in this case, implying that the angle of the double hypercone defining the detection region is much larger. Therefore, the differences among the embedding strategies are minimal, and as a consequence, the obtained error exponents are virtually the same for Broken Arrows, OBPA, and the optimal embedder described by (13).

VII. Conclusions

In this paper we considered the derivation of a Neyman–Pearson asymptotically optimum zero–bit watermarking scheme in a Gaussian setting, when the detector is limited to base its decisions on second order empirical statistics only. In particular we extended previous works in this direction by considering the presence of noise. The main contributions of the paper can be summarized as follows: i) we derived the false negative error exponent for any embedding strategy; ii) we derived a min-max-min expression for the the optimal embedding strategy in a general context; iii) we derived a class of universally optimum embedding strategies in the high-SNR; iv) we proposed a new embedding rule, chosen among the optimal embedding rules for the high-SNRr regime, that is particularly suited to the case of low E_{fn} values; v) we derived the false negative error exponent of the new embedding rule and that of some previously proposed methods; vi) finally, we have shown the good (though not optimal) behavior of the new scheme in a wide range of set-ups including those most relevant from a practical point of view. Interestingly, the new embedding strategy we introduced is very simple thus opening the door to practical

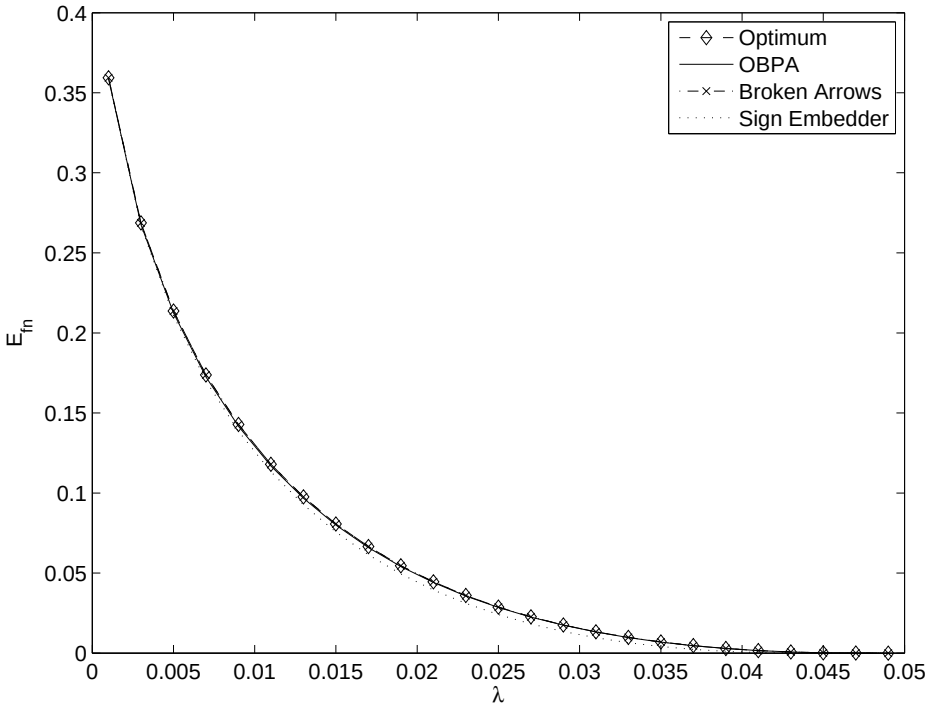

Fig. 4. Comparison of the errors exponents obtained by the sign embedder described by Merhav and Sabbag [15], the Broken Arrows strategy proposed by Furon and Bas [18] when just a secret direction is considered, the solution of (13), and the OBPA embedding technique. $\sigma_X^2 = 1$, $D = 0.1$, $\sigma_Z^2 = 0.1$.

implementations. This work can be extended in many interesting directions, including non-Gaussian settings, more complicated attacks, like de-synchronization attacks [25], [26], more detailed empirical statistics gathered by the detector, and the introduction of security considerations in the picture [27].

REFERENCES

[1] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Information and Control*, vol. 9, no. 1, pp. 19–31, 1980.

[2] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[3] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1127–1141, July 1999.

[4] B. Chen and G. W. 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.

[5] M. Ramkumar and A. N. Akansu, "Signaling methods for multimedia steganography," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1100–1111, April 2004.

[6] A. Abrardo and M. Barni, "Informed watermarking by means of orthogonal and quasi-orthogonal dirty paper coding," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 824–833, February 2005.

[7] F. Pérez-González, C. Mosquera, M. Barni, and A. Abrardo, "Rational dither modulation: a high-rate data-hiding method invariant to gain attack," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3960–3975, October 2005.

[8] J. R. Hernández, M. Amado, and F. Pérez-González, "DCT-domain watermarking techniques for still images: detector performance analysis and a new structure," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 55–68, January 2000.

[9] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A new decoder for the optimum recovery of non-additive watermarks," *IEEE Transactions on Image Processing*, vol. 10, no. 5, pp. 755–766, May 2001.

[10] X. Huang and B. Zhang, "Statistically robust detection of multiplicative spread-spectrum watermarks," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 1–13, March 2007.

[11] M. Noorkami and R. M. Mersereau, "A framework for robust watermarking of H.264-encoded video with controllable detection performance," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 14–23, March 2007.

[12] W. Liu, L. Dong, and W. Zeng, "Optimum detection for spread-spectrum watermarking that employs self-masking," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 645–654, December 2007.

[13] M. L. Miller, I. J. Cox, and J. A. Bloom, "Informed embedding: Exploiting image and detector information during watermark insertion," in *IEEE International Conference on Image Processing (ICIP)*, vol. 3, Vancouver, BC, Canada, September 2000, pp. 1–4.

[14] T. Liu and P. Moulin, "Error exponents for one-bit watermarking," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, Hong Kong, April 2003, pp. 65–68.

[15] N. Merhav and E. Sabbag, "Optimal watermark embedding and detection strategies under limited detection resources," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 255–274, January 2008.

[16] T. Furon, J. Josse, and S. L. Squin, "Some theoreical aspects of watermarking detection," in *Proceedings of SPIE, Security, Steganography and Watermarking of Multimedia contents VIII*, E. J. Delp III and P. W. Wong, Eds., vol. 6072. San Jose, CA, USA: SPIE, January 2006.

[17] T. Furon, "A constructive and unifying framework for zero-bit watermarking," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 2, pp. 149–163, June 2007.

[18] T. Furon and P. Bas, "Broken arrows," *EURASIP Journal on Information Security*, vol. 2008, pp. 1–13, 2008, doi:10.1155/2008/597040.

[19] L. Pérez-Freire, P. Comesaña, and F. Pérez-González, "Detection in quantization-based watermarking: Performance and security issues," in *Proceedings of SPIE, Security, Steganography, and Watermarking of Multimedia Contents VII*, E. J. Delp III and P. W. Wong, Eds., vol. 5681. San Jose, CA, USA: SPIE, January 2005, pp. 721–733.

[20] 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.

[21] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer Texts in Electrical Engineering, 1994.

[22] T. Furon, B. Macq, N. Hurley, and G. Silvestre, "JANIS: Just Another N-order side-Informed watermarking Scheme," in *IEEE International Conference on Image Processing (ICIP)*, vol. 2, Rochester, NY, USA, September 2002, pp. 153–156.

[23] 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.

[24] R. Wong, *Asymptotic Approximations of Integrals*. SIAM, 2001.

[25] M. Barni, "Effectiveness of exhaustive search and template matching against watermark desynchronization," *IEEE Signal Processing Letters*, vol. 12, no. 2, pp. 158–161, February 2005.

[26] A. D'Angelo, M. Barni, and N. Merhav, "Expanding the class of watermark desynchronization attacks," in *Proceedings of 9-th ACM Multimedia Security Workshop*, Dallas, Texas, 20-21 September 2007.

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

PLACE
PHOTO
HERE

Pedro Comesaña (M'09) received the telecommunications engineer degree and the Ph. D. degree in telecommunications engineering from the University of Vigo, Vigo, Spain, in 2002 and 2006, respectively.

Currently, he is an Assistant Professor with the University of Vigo. In 2004, he spent six months with the Technische Universiteit Eindhoven, Eindhoven, The Netherlands. In 2006, he was with the Information Hiding Laboratory at the National University of Ireland (University College Dublin), Dublin, Ireland for six months. In 2007 and 2008, he was with the Università degli Studi di Siena, Siena, Italy, for two periods spanning ten months. His research interests are data hiding, multimedia signal processing, and digital communications.

Dr. Comesaña was recipient of the Spanish Official Institute of Telecommunications Engineers Award to the the Best Ph. D. Thesis in Security and Defense, 2006.

PLACE
PHOTO
HERE

Neri Merhav (S'86–M'87–SM'93–F'99) was born in Haifa, Israel, on March 16, 1957. He received the B.Sc., M.Sc., and D.Sc. degrees from the Technion, Israel Institute of Technology, in 1982, 1985, and 1988, respectively, all in electrical engineering.

From 1988 to 1990 he was with AT&T Bell Laboratories, Murray Hill, NJ, USA. Since 1990 he has been with the Electrical Engineering Department of the Technion, where he is now the Irving Shepard Professor. During 1994–2000 he was also serving as a consultant to the Hewlett–Packard Laboratories – Israel (HPL-I). His research interests include information theory, statistical communications, and statistical signal processing. He is especially interested in the areas of lossless/lossy source coding and prediction/filtering, relationships between information theory and statistics, detection, estimation, as well as in the area of Shannon Theory, including topics in joint source–channel coding, source/channel simulation, and coding with side information with applications to information hiding and watermarking systems. Another recent research interest concerns the relationships between Information Theory and statistical physics.

Dr. Merhav was a co-recipient of the 1993 Paper Award of the IEEE Information Theory Society and he is a Fellow of the IEEE since 1999. He also received the 1994 American Technion Society Award for Academic Excellence and the 2002 Technion Henry Taub Prize for Excellence in Research. From 1996 until 1999 he served as an Associate Editor for Source Coding to the IEEE TRANSACTIONS ON INFORMATION THEORY. He also served as a co–chairman of the Program Committee of the 2001 IEEE International Symposium on Information Theory. He is currently on the Editorial Board of FOUNDATIONS AND TRENDS IN COMMUNICATIONS AND INFORMATION THEORY.

PLACE
PHOTO
HERE

Mauro Barni graduated in electronic engineering at the University of Florence in 1991. He received the PhD in informatics and telecommunications in October 1995. He has carried out his research activity for over 20 years first at the Department of Electronics and Telecommunication of the University of Florence, then at the Department of Information Engineering of the University of Siena where he works as associate Professor. During the last decade he has been studying the application of image processing techniques to copyright protection and authentication of multimedia (digital watermarking). He is author/co-author of about 250 papers published in international journals and conference proceedings, and holds three patents in the field of digital watermarking. He is co-author of the book ”Watermarking Systems Engineering: Enabling Digital Assets Security and other Applications”.

He participated to several National and European research projects on diverse topics, including computer vision, multimedia signal processing, remote sensing, digital watermarking, IPR protection.

He serves as associate editor of the IEEE Trans. Information Forensics and Security and the IEEE Trans. on Circuits and system for Video Technology. He was the founding editor of the EURASIP Journal on Information Security. He was the general chairman of the 2004 edition of IEEE workshop on Multimedia Signal Processing (MMSP’04) and the 2005 edition of the International Workshop on Digital Watermarking (IWDW’05). Prof. Barni is the chairman of the IEEE Information Forensic and Security technical Committee (IFS-TC) of the IEEE Signal Processing Society. He is a senior member of the IEEE and EURASIP.