Generalized Improved Spread Spectrum Watermarking Robust Against Translation Attacks

by

Neslihan Gerek

B. S., in Electrical and Electronics Engineering, Boğaziçi University, 2005

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Master of Science in Electrical and Electronics Engineering

Boğaziçi University

2008

Generalized Improved Spread Spectrum Watermarking Robust Against Translation Attacks

APPROVED BY:

Assist. Prof. Kıvanç Mıhçak (Thesis Supervisor)	.....
Prof. Bülent Sankur	.....
Assist. Prof. Serdar Kozat	.....

DATE OF APPROVAL: 03.09.2008

ACKNOWLEDGEMENTS

I would like to thank Kıvanç Mıhçak, my supervisor, for his many suggestions and constant support during this research. I am also thankful to Onur Özyeşil for sharing his knowledge of Detection and Estimation Theory and providing his comments on analytical derivations, to Bülent Sankur and Serdar Kozat for their valuable comments which improve the quality of this thesis.

And I am grateful to my parents for their patience and love.

ABSTRACT

Generalized Improved Spread Spectrum Watermarking Robust Against Translation Attacks

In this work, we consider the ISS (improved spread spectrum) watermarking [1] framework, and propose a generalized version of it, termed "generalized improved spread spectrum" (GISS), where we achieve both host-interference cancelation and robustness to "translation" attacks up to some tolerance. In particular, we reduce the correlation between the watermark and the host, not only at the embedding location, but also within an a-priori-defined neighborhood around it. We show that the resulting framework leads to a constrained quadratic optimization problem, where the cost function and the constraint represent the amount of host interference on the watermark and the norm of the resulting "host interference cancelation sequence" (HICS), respectively. We provide a closed-form analytical solution to this optimization problem and experimentally demonstrate its effectiveness for 1D signals.

Also, we propose three different methods (search/correlation, joint MAP and focused MAP methods) for the decoding of embedded binary information and analyze the performances of these methods in terms of the probability of decoding error. Regarding these performances, we provide closed-form solutions for focused MAP and joint MAP detectors and experimental results of their performances. Moreover, for the search/correlation decoding method that we introduce, we provide analytical and experimental results regarding the performance.

ÖZET

Öteleme Saldırılarına Karşı Gürbüz Genelleştirilmiş Geliştirilmiş Yayılı Spektrum Damgalama

Bu çalışmada GYS (geliştirilmiş yayılı spektrum) damgalama [1] yöntemini göz önünde tutarak onun genelleştirilmiş bir biçimini ileri sürdük. "Genelleştirilmiş geliştirilmiş yayılı spektrum" (GGYS) yönteminde hem konakçı girişiminin iptal edilmesini, hem de belli bir toleransa kadar öteleme saldırılarına karşı gürbüzlüğü başardık. Özellikle, damga ile konakçı sinyal arasındaki ilinti sadece gömme yapılan konumda değil, o noktanın önsel-tanımlanmış komşuluğunda da azaltıldı. Elde ettiğimiz çatı, maliyet fonksiyonu damga üzerindeki konakçı girişimi, kısıtı da sonuçta oluşan "konakçı girişimini iptal eden dizi"nin (KGID) normu olan bir kısıtlı eniyileme problemine ulaştı. Bu eniyileme problemi için kapalı yapıda analitik sonucu sağladık ve 1 boyutlu sinyaller için etkinliğini deneysel sonuçlarla gösterdik.

Ayrıca, gömülmüş ikili bilgiyi çözebilmek için üç farklı kodçözme yöntemi (arama/ilintileme, birleşik MAP ve odaklanmış MAP yöntemleri) ileri sürdük ve bu yöntemlerin kodçözme hata olasılığı bakımından başarımını çözümledik. Bu başarımlarla ilgili olarak, birleşik MAP ve odaklanmış MAP kodçözücüleri için kapalı yapıda çözümler ve başarımları ile ilgili deneysel sonuçlar sağladık. Bundan başka, sunduğumuz arama/ilintileme yöntemi için analitik ve deneysel sonuçlar sağladık.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

$\mathcal{A}(.)$	Sub-area function
b	Binary information
$\mathbf{c}$	Host interference cancelation sequence
$f_\theta(.)$	Translation attack
$\mathbf{H}$	Linear correlation transform matrix
K	Core watermark signal size
L	Maximum shift amount
M	Tolerance region size
$\bar{N}$	Host signal size
N	Inner portion of the host signal size
$\bar{\mathbf{n}}$	Additive white gaussian noise
$\bar{\mathbf{r}}$	Received signal at the decoder end
$\mathbf{r}$	Inner portion of the received signal at the decoder end
$\bar{\mathbf{s}}$	Watermark embedded signal
$\mathbf{s}$	Inner portion of the watermark embedded signal
$\mathbf{u}$	Watermark signal
$\bar{\mathbf{x}}$	Host signal
$\mathbf{x}$	Inner portion of the host signal
α	Watermark strength
θ	Translation amount
i.i.d.	Independent and identically distributed
SVD	Singular value decomposition

1. INTRODUCTION

Digital processing and distribution of multimedia content offers many advantages such as high signal quality, software installation instead of hardware, ease of editing, copying and transmitting. On the other hand, progressive development in the digital processing area makes secure data transmission difficult.

Information hiding refers to a class of problems where secure data transmission is carried out by hiding a message in the host data. Since different scenarios for information hiding problem yield different problems, they have to be treated differently.

Data embedding is one of the classes of information hiding problem. Here, the hidden message is unknown at the receiver end; however, the receiver knows that there is a hidden message in the received data. Thus, the goal is to decode the message with high reliability and this problem is a *decoding* problem. Naturally this problem is closely related to classical communication problems. An information-theoretic approach towards finding the capacity in a game-theoretic framework has been presented in [7, 8].

The aim of data embedding is to embed a signal (namely "watermark") into the host data, while the embedded signal is not detectable without the secret key. Novel applications of this problem appeared in the 1990's. The first of a series of annual international workshops on information hiding was held in 1996. Special issues of of the *IEEE Journal on Selected Areas in Communications* and of the *Proceedings of the IEEE* were devoted to copyright and privacy protection in 1998 and 1999, respectively [2, 3]. A non-technical survey of the problem of information hiding can be found in [4]. State-of-the-art watermarking techniques are discussed in a recent book by Cox *et al.* [5]. Another closely related subject is given by Eggers [6].

A general scheme for digital watermarking is given in Figure 1.1. Watermark signal (WM) is embedded to the host signal by using a secret key at the embedder

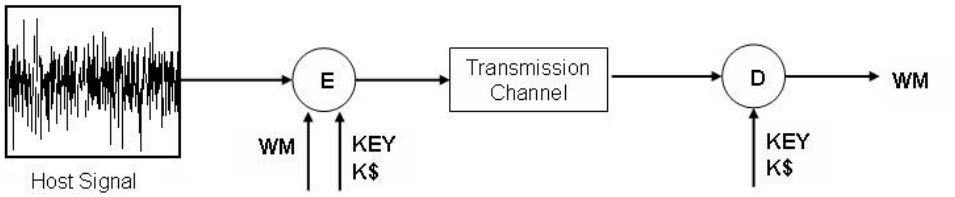


Figure 1.1. General Scheme for Digital Watermarking.

(E). Without the knowledge of the secret key, it is not possible to extract the information embedded in the host signal. The watermark embedded signal passes through a transmission channel, and may subject to attacks such as lossy compression, geometric distortion, AWGN attack, or any signal processing operation. At the receiver side, the watermark is extracted using the secret key.

In this thesis, we consider the spread spectrum watermarking problem and present a generalized version of it, which is robust against translation attacks. The main contributions of this thesis can be stated as follows:

- We generalized the Improved Spread Spectrum (ISS) method proposed by Malvar *et. al* [1] such that decorrelation of the watermark signal is performed not only on the host signal, but also on several shifted versions of it. Therefore, host interference is reduced in the presence of translation attacks.
- We analyze three decoding methods in the presence of Additive White Gaussian Noise (AWGN) and translation attacks.

In the rest of the introduction, we present some possible applications for watermarking schemes in Sec. 1.1. We give the properties of the watermark in Sec. 1.2, and in Sec. 1.3, we outline the rest of the thesis. This chapter is intended to form a brief introduction for readers who are not familiar with Watermarking. Other readers can skip the remaining sections and proceed with the rest of the thesis with little or no loss of information.

1.1. Watermarking Applications

Digital watermarking has a wide range of applications including owner identification, transactional watermarks, copy control, broadcast monitoring and covert communication. We refer the reader to [9] for an extensive description of the various applications. The broad classification of the applications is as follows:

i. Owner Identification : Such applications are usually targeted by robust watermarking algorithms and intended for commercial purposes. In such a situation, a company that produces and sells digital audio clips or a movie company that sells its products over the Internet is concerned with copyright issues. In particular, it is very profitable for hackers to crack these products and sell them at a cheaper price. In such situations, original producers would like to have legally valid proof that they are the real owners. Robust signature casting is a possible solution in such cases. A standardized robust watermarking scheme, accepted by the courts, would require the attackers to remove the watermark, while preserving the quality of the data, before selling the modified data. By design this should be very difficult to achieve.

ii. Transactional Watermarks (Fingerprinting) : Fingerprinting applications allow a content owner or content distributor to identify the source of an illegal copy, because all individual copy is embedded with a unique watermark. In such a situation, a movie company inserts user IDs in each product before selling it. Whenever an unauthorized user is caught playing the movie or selling it, that user would be identified.

iii. Copy Control : It is possible for recording and playback devices to react embedded signals. For example, a recording device might inhibit recording of a signal if it detects a watermark that indicates recording is prohibited.

iv. Broadcast Monitoring : We can use watermarks for broadcast monitoring by putting a unique watermark in each media signal (video or sound clip) before the broadcast. Automated broadcast monitoring stations can then receive broadcasts and look for these watermarks identifying when and where each clips appears.

v. Covert Communication : These are mainly applications of steganography. In

many situations, such as military and intelligence applications, people would like to send messages to each other without being detected. In such cases, the adversary is usually the enemy. It is often vital that the enemy does not detect the presence of a secret message transmitted.

1.2. Constraints, Goals and Practical Difficulties

In designing watermarking algorithms, we come across several design constraints. These constraints could be somewhat different for each intended application. However it is still possible to classify them broadly as follows:

i. Fidelity : Watermarking methods need to maintain a certain level of perceptual quality in the data that are produced after embedding. Otherwise the watermarked data would be useless.
ii. Attack Strength : The watermarked data will possibly undergo attacks. An attack can be intentional or unintentional. In both cases, the watermarked data should be usable after the attack. This provides a bound on the strength of the attacks that are considered.
iii. Computational Complexity : In some applications, watermarking schemes have to operate in real time. Thus, it is desired that they are fast. Hence, in such scenarios, we have a constraint on the computational complexity of the proposed methods.

The goal of a watermarking scheme should be discussed separately for verification and decoding problems. In a decoding scheme, the goal is to minimize the probability of error in the decoded message at the receiver, where the probability of error is given by the probability that the decoded message is not equal to the embedded message. In a verification scheme, the probability of error takes a different form; the receiver makes a binary decision: a watermark is present or not. In this case, it is desired to minimize the probability of error, which is given by a linear combination of probability of miss (the probability of declaring a hidden message is not present, even though it is) and probability of false alarm (the probability of declaring a hidden message is present,

even though it is not).

1.3. Thesis Organization

The organization of the thesis is as follows. In Sec. 2, we provide the notation that is used throughout the thesis and give brief explanations of spread spectrum (SS) (Sec. 2.2.1) and Improved Spread Spectrum (ISS) (Sec. 2.2.2) watermarking methods. In Sec. 3, we specify the formal problem statement. In Sec. 4, we introduce the resulting optimization problem of GISS method and provide the analytical solution. In Sec. 5, we model the attack that we utilize for experiments and analytical inference. In Sec. 6, we present three different decoding methods, namely Search/Correlation Decoding Method (Sec. 6.1), Joint MAP Decoding Method (Sec. 6.2), and Focused MAP Decoding Method (Sec. 6.3). In Sec. 7, we show experimental results comparing different encoding and decoding methods that we define in the previous sections. We conclude with final discussions in Sec. 8.

2. NOTATION AND PRIOR ART

In this section we provide the notation we use and state the theoretical foundations of the proposed method by explaining Spread Spectrum (SS) and Improved Spread Spectrum (ISS) watermarking methods briefly.

2.1. Notation

Bold upper- and lower-case letters represent matrices and vectors, respectively. Corresponding regular letters with subscripts represent individual elements. For example, $\mathbf{a} \in \mathbb{R}^N$ is a vector and $a_i \in \mathbb{R}$ is its i^{th} element; given the matrix $\mathbf{A}$, A_{ij} is its $(i,j)^{th}$ element, $\mathbf{A}^T$ and $r(\mathbf{A})$ denote its transpose and rank, respectively. $\langle\,.\,,\,.\,\rangle$ represents the inner product which induces the Euclidean (L_2) norm; $\mathbf{I}_K$ denotes the identity matrix of size $K \times K$. Also, $\mathbf{a} \sim \mathcal{N}(\mu, \mathbf{\Sigma})$ represents that the random variable $\mathbf{a}$ is Gaussian distributed with mean μ and covariance $\mathbf{\Sigma}$. Similarly, $\mathbf{a} \sim \mathcal{U}(S)$ represents that the random variable $\mathbf{a}$ is uniformly distributed over the set S.

Definition 2.1.1 *Given* $\mathbf{A} \in \mathbb{R}_{M \times N}$*, such that* $r(\mathbf{A}) = K$*,* Singular Value Decomposition (SVD) *of* $\mathbf{A}$ *is unique (up to ordering) and defined as*

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T, \tag{2.1}$$

where $\mathbf{U} \in \mathbb{R}_{M \times K}$*,* $\mathbf{V} \in \mathbb{R}_{N \times K}$*,* $\mathbf{\Lambda} \in \mathbb{R}_{K \times K}$ *are called the left-singular vector matrix, the right-singular vector matrix, and the singular value matrix of* $\mathbf{A}$ *respectively. The matrix* $\mathbf{\Lambda}$ *is diagonal and these diagonal elements are termed as the singular values of* $\mathbf{A}$*.*

2.2. Prior Art

2.2.1. Basic Spread Spectrum (SS) Watermarking Method

In its most basic form, SS method assumes that one bit of information is embedded in a vector of N coefficients, achieving a bit rate of $1/N$ bits per sample. In this case, the watermarked signal is given by

$$\mathbf{s} = \mathbf{x} + b\mathbf{u},$$

where $\mathbf{x}$ and $\mathbf{u} \in \mathbb{R}^N$ are the host and the watermark signals, respectively; $b \in \{\pm 1\}$ represents the embedded bit. If the attack channel can be modeled as additive noise, then the received signal at the decoder is given by

$$\mathbf{y} = \mathbf{s} + \mathbf{n} = \mathbf{x} + b\mathbf{u} + \mathbf{n}.$$

Decoding is performed by checking the sign of the normalized statistics produced by the correlation detector:

$$\gamma \triangleq \frac{\langle \mathbf{y}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} = \frac{\langle b\mathbf{u} + \mathbf{x} + \mathbf{n}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2}; \quad \hat{b} = \text{sign}(\gamma), \tag{2.2}$$

where γ and $\hat{b}$ denote the detection statistics and the decoded bit, respectively. See [10] for further details.

2.2.2. Improved Spread Spectrum (ISS) Watermarking Method

The main idea in this method is to reduce the correlation between the host signal and the watermark by modulating the energy of the watermark at the embedding process using the projection of the host signal on the watermark. The resulting embedding method, namely ISS [1], is a slightly modified version of the conventional SS

embedding method. In this case the watermarked signal is represented as

$$\mathbf{s} = \mathbf{x} + \lambda(x, b)\mathbf{u}.$$

Here, the value x represents the projection coefficient of the host $\mathbf{x}$ on the watermark $\mathbf{u}$, i.e. $x \triangleq \langle \mathbf{x}, \mathbf{u} \rangle / \|\mathbf{u}\|^2$. Amplitude of the embedded watermark is controlled by the function $\lambda(x, b)$, which can be defined in various ways. In case of linearity we get

$$\mathbf{s} = \mathbf{x} + (\alpha b - \lambda_{ISS} x)\, \mathbf{u};$$

where α controls the distortion level and λ_{ISS} controls the removal of the host interference on the detection statistics. We refer the interested reader to [1] for further details.

3. PROBLEM STATEMENT

We study an additive watermarking method that we name "generalized improved spread spectrum"(GISS)[1] . In this scheme, the host signal is first decorrelated from the watermark signal in order to achieve robustness to translation attacks to some extent, and then embedding of the watermark and binary information is performed. Subsequently, the watermarked signal is subject to an attack, and model of this attack is explained in Sec. 5. The aim of the detector is to detect the embedded binary information in the received signal. Decorrelation of host signal and the watermark signal is first proposed by Malvar *et. al* [1], and this scheme is briefly explained in Sec. 2.2.2.

In this setup, we assume that the host signal is a long stream of data and we embed the watermark in a relatively small portion of it. Also, we assume that this long stream of data is subject to a *cyclic shift attack* (see eq. (5.1)), which yields a *translation attack* for the watermark embedded part. This setup represents the problem of robust watermarking for long data streams where extraction of watermark is potentially subject to synchronization-like imperfections, i.e., the exact location of the embedded bit is unknown at the receiver end. Embedding of the watermark to a portion of the long data stream is performed using the "sub-area functions" provided in the following definition.

Definition 3.0.1 *Sub-area functions* $\mathcal{A}^{pre} : \mathbb{R}^{\bar{N}} \longmapsto \mathbb{R}^{N_1}$, $\mathcal{A} : \mathbb{R}^{\bar{N}} \longmapsto \mathbb{R}^{N}$, $\mathcal{A}^{post} : \mathbb{R}^{\bar{N}} \longmapsto \mathbb{R}^{N_2}$ *satisfy*

$$\begin{aligned} [\boldsymbol{a}^{pre} = \mathcal{A}^{pre}(\bar{\boldsymbol{a}})] &\iff [\boldsymbol{a}_i^{pre} = \bar{\boldsymbol{a}}_i \ ; 1 \leq i \leq N_1] \\ [\boldsymbol{a} = \mathcal{A}(\bar{\boldsymbol{a}})] &\iff [\boldsymbol{a}_i = \bar{\boldsymbol{a}}_{N_1+i} \ ; 1 \leq i \leq N] \\ \left[\boldsymbol{a}^{post} = \mathcal{A}^{post}(\bar{\boldsymbol{a}})\right] &\iff \left[\boldsymbol{a}_i^{post} = \bar{\boldsymbol{a}}_{N_1+N+i} \ ; 1 \leq i \leq N_2\right] \end{aligned}$$

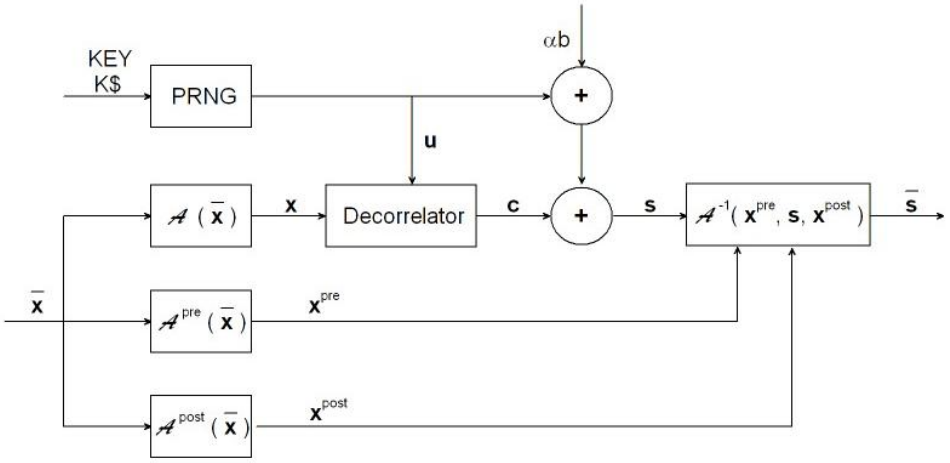
[1]This approach can potentially be extended to include multiplicative watermarking scheme as well. This constitutes part of our future research.

where $\bar{N} = N_1 + N_2 + N$ *and* $N \ll \bar{N}$ *and* $\bar{\boldsymbol{a}} = \begin{bmatrix} \boldsymbol{a}^{pre} \\ \boldsymbol{a} \\ \boldsymbol{a}^{post} \end{bmatrix}$.

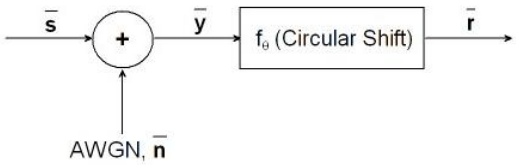
Using the sub-area functions $\mathcal{A}^{pre}, \mathcal{A}$ and $\mathcal{A}^{post}$ defined above, the signal flow in our problem is given as follows:

- For a sufficiently long stream $\bar{\mathbf{x}} \in \mathbb{R}^{\bar{N}}$, the watermark embedded portion of the original stream $\mathbf{x} \in \mathbb{R}^{N}$ is given by $\mathbf{x} = \mathcal{A}(\bar{\mathbf{x}})$, i.e., we subdivide the original stream into three portions where $\bar{\mathbf{x}} = \begin{bmatrix} \mathbf{x}^{pre} \\ \mathbf{x} \\ \mathbf{x}^{post} \end{bmatrix}$ and $\mathbf{x}^{pre} = \mathcal{A}^{pre}(\bar{\mathbf{x}})$, $\mathbf{x}^{post} = \mathcal{A}^{post}(\bar{\mathbf{x}})$.
- Using the inner portion of the host signal, i.e., $\mathbf{x}$, host interference cancellation sequence (HICS), $\mathbf{c}$, is designed to minimize the correlation between the watermark signal and several shifted versions of the inner portion of the signal, subject to a norm constraint (see Sec. 4).
- The binary information $b \in \{-1\,,\,+1\}$, the watermark strength α and the watermark signal $\mathbf{u}$ constitute the embedded watermark via a multiplicative relation, i.e., the embedded watermark is given by $\alpha b\mathbf{u}$.
- The embedded watermark, together with the HICS, is added to the inner portion $\mathbf{x}$ to form the inner portion of the transmitted stream $\bar{\mathbf{s}} \in \mathbb{R}^{\bar{N}}$. We thus have, $\bar{\mathbf{s}} = \begin{bmatrix} \mathbf{x}^{pre} \\ \mathbf{s} \\ \mathbf{x}^{post} \end{bmatrix}$ where $\mathbf{s} = \mathcal{A}(\bar{\mathbf{s}}) = \mathbf{x} + \mathbf{c} + \alpha b\mathbf{u}$ and $\mathbf{x}^{pre} = \mathcal{A}^{pre}(\bar{\mathbf{s}})$, $\mathbf{x}^{post} = \mathcal{A}^{post}(\bar{\mathbf{s}})$.
- The channel between the transmitter and the receiver is composed of an additive noise and a cyclic shift attack. Thus, in terms of the transmitted stream $\bar{\mathbf{s}}$ the received signal $\bar{\mathbf{r}}$ is given by $\bar{\mathbf{r}} = t\,(\bar{\mathbf{s}}) = f_\theta\,(\bar{\mathbf{y}}) = f_\theta\,(\bar{\mathbf{s}} + \bar{\mathbf{n}})$, where the overall attack imposed by the channel, the additive noise and cyclic shift attack are represented by $t\,(.)$, $\bar{\mathbf{n}}$ and $f_\theta\,(.)$, respectively (see Sec. 5), where θ denotes cyclic shift amount.

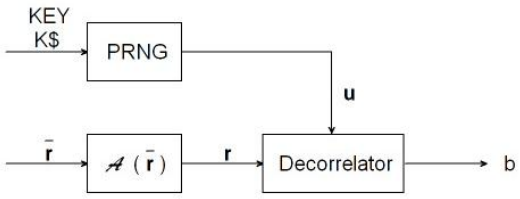
The signal flow in our setup that we explained above is summarized by Figure 3.1. Here, K$ is the secret key shared by the embedder and the decoder. PRNG stands for Psuedo-Random number generator, b is the embedded bit and AWGN stands for Additive White Gaussian Noise.



(a)



(b)



(c)

Figure 3.1. GISS watermarking: (a) the embedder, (b) the attack channel, (c) the decoder

The host signal $\bar{\mathbf{x}}$ and the additive noise $\bar{\mathbf{n}}$ are assumed to be *zero mean* random vectors having *Gaussian* distributions with covariance matrices $\sigma^2_{\mathbf{x}}\mathbf{I}_{\bar{N}}$ and $\sigma^2_{\mathbf{n}}\mathbf{I}_{\bar{N}}$, respectively. Also, the embedded information, b, is equiprobable binary: $b \in \{1, -1\}$, where $p(b = 1) = p(b = -1) = \frac{1}{2}$.

For the watermarking operation described above, we solve the following problems:

- HICS Design : HICS is used to reduce the interference resulting from the host signal, which increases the probability of detection error. In Sec. 4, we solve the problem of designing HICS as a function of host signal (subject to a power constraint on HICS) to minimize the correlation between the watermark signal and several shifted variants of the host signal.
- Decoding Analysis : We analyze the error performance of three decoding methods; namely "search/correlation decoding" (Sec. 6.1), "joint MAP decoding" (Sec. 6.2) and "focused MAP decoding" (Sec. 6.3). Also, we provide experimental results to study the performances of these three methods.

4. EMBEDDING VIA GISS APPROACH

In this section, we provide our watermark embedding method using GISS approach. The ISS method explained in Sec. 2.2.2 aims to decrease the correlation only between the watermark signal and the host signal. This approach relies on the assumption that in the extraction of the watermark, the location of the watermark signal is *exactly known*. However, in our setup we assume that the channel between the embedder and the decoder sides introduces a cyclic shift on the long stream resulting in a linear shift on the watermark embedded portion. We also assume that the amount of this shift is uniformly distributed on the set $\{-L, -L+1, ..., L\}$, where L denotes the maximum shift amount satisfying $L < N_1$ and $L < N_2$ in order to assure that the cyclic shifts on the long stream reduce to linear shifts on the watermark embedded portion. Hence, the exact location of the watermark signal is unknown to the decoder side, instead, only the *probabilistic characteristics* of the location is known at the decoder side. As a result, the GISS approach aims to minimize the correlation between the watermark signal and several shifted versions of the host signal in order to decrease the probability of detection error at the decoder side for our setup. In this section, we construct the "host interference cancellation sequence" $\mathbf{c} \in \mathbb{R}^N$ in order to realize the minimization of the correlation mentioned above.

Let $\mathbf{w} \in \mathbb{R}^K$ be a zero mean random vector having Gaussian distribution with variance $\sigma_{\mathbf{u}}\mathbf{I}_K$ denote the "core watermark" signal, where we assume $K < N$. This signal is generated by a pseudo-random number generator (PRNG) using the secret key as a seed. To represent the embedding of $\mathbf{w}$ into the watermark embedded portion of the host signal $\mathbf{x} = \mathcal{A}(\bar{\mathbf{x}})$, we define $\mathbf{u} \in \mathbb{R}^N$ satisfying

$$\mathbf{u}_i = \begin{cases} \mathbf{w}_i - \frac{N-K}{2} & \frac{N-K}{2} < i \leq \frac{N+K}{2} \\ 0 & \text{else} \end{cases} .$$

In GISS approach, the modified version of the inner portion of the host signal $\mathbf{x}$ is given by $\mathbf{x} + \mathbf{c}$. Now, in order to represent the correlation between a shifted version of

$\mathbf{x}+\mathbf{c}$ and the watermark $\mathbf{u}$, we consider the original version of $\mathbf{x}+\mathbf{c}$ and a version of the watermark that is shifted in reverse direction with the same amount. To find the correlation values between all possible shifted versions of $\mathbf{x}+\mathbf{c}$ and $\mathbf{u}$ we define the following "linear correlation transform matrix".

Definition 4.0.2 *The matrix* $\mathbf{H}$*, the rows of which are the cyclic shifted versions of the watermark signal* $\mathbf{u}$*, is called "linear correlation transform matrix", i.e.,* $\mathbf{H} \in \mathbb{R}^{2L+1\times N}$ *satisfies*

$$\mathbf{H}_i = \mathbf{u}^i \quad \textit{where} \quad \mathbf{u}^i_j = \mathbf{u}_{(L+1+j-i)\, mod(N)}; \qquad 1 \leq i \leq 2L+1, \qquad 1 \leq j \leq N \quad .$$

Remark 4.0.1 *Note that we currently confine ourselves to the usage of i.i.d. Gaussian distributed watermark and all the subsequent derivations are carried out accordingly. The reason of choosing an i.i.d. process is intuitively clear: since our final goal is to minimize interferences under translation attacks, usage of a "self-correlated" watermark would presumably hinder achieving this goal. However, the aforementioned argument is still qualitative. As such, it seems to be an interesting and open problem to find out the "optimal" distribution of the watermark in the sense of the decoder probability of error, which is beyond*

Remark 4.0.2 *In the following sections, we call* $M \triangleq 2L+1$ *the "tolerance region size". M refers to the size of the domain outside of which the probability of the existence of the location of the watermark is equal to zero. Throughout the thesis, we assume that* $M+K<N$*, i.e., in case of a translation attack, core of the watermark will remain in the inner portion of the received signal.*

Remark 4.0.3 *Note that as the system designer, we have the freedom to choose* N*. Therefore, as long as* $M+K<\bar{N}$*, we can find out some* $N<\bar{N}$ *which satisfies the constraint of* $M+K<N$*, and this is needed for analytical convenience.*

Remark 4.0.4 *If we consider* $\mathbf{H}$ *defined above, the correlation values between the shifted versions of the modified inner portion of the host signal, i.e.,* $\mathbf{x}+\mathbf{c}$*, and the watermark* $\mathbf{u}$ *are given by* $\mathbf{H}(\mathbf{x}+\mathbf{c})$*. Therefore* $\mathbf{H}$ *allows us to represent the correlation values in a compact vector form.*

Remark 4.0.5 *We assume that* $\mathbf{H}$ *is full-rank.*

Remark 4.0.6 *Since the location of the watermark is uniformly distributed in the tolerance region, the individual correlation values do not differ in terms of their priorities in the minimization problem, i.e., the correlation values are* not weighted *for the minimization.*

Next we define our main problem in this section, i.e., construction of the optimal HICS that minimizes the correlation between shifted versions of $\mathbf{x}+\mathbf{c}$ and $\mathbf{u}$.

Definition 4.0.3 *"The optimal HICS" is given by*

$$\mathbf{c}_{opt} \triangleq \underset{\substack{\mathbf{c}\in\mathbb{R}^N \\ \|\mathbf{c}\|^2 \leq A}}{\operatorname{argmin}} \| \mathbf{H}\left(\mathbf{x}+\mathbf{c}\right) \|^2 \tag{4.1}$$

Remark 4.0.7 *Parameter* A *controls the distortion, induced by adding* $\mathbf{c}$ *to* $\mathbf{x}$*. This practical situation decreases the detection performance of the optimal HICS in cases where the variances of the host signal elements* $\mathbf{x}_i$ *are high compared to parameter* A*.*

In Proposition 4.0.1, we provide the construction of $\mathbf{c}_{opt}$ defined by (4.1).

Proposition 4.0.1 *The optimal HICS in (4.1) is given by*

$$\mathbf{c}_{opt} = -\mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T\mathbf{x}, \tag{4.2}$$

where $\mathbf{V}$ *is the right-singular vector matrix of the linear correlation transform matrix* $\mathbf{H}$ *and* $\widetilde{\mathbf{S}} \in \mathbb{R}^{M \times M}$ *is a diagonal matrix, satisfying*

$$\widetilde{\mathbf{S}}_{ii} = \frac{\sigma_i^2}{\lambda_{opt} + \sigma_i^2}, \quad 1 \leq i \leq M.$$

for which, σ_i *is the* i^{th} *singular value of* $\mathbf{H}$ *and* λ_{opt} *is the optimal Lagrange multiplier for the constraint optimization problem given by (4.1).*

Proof:

See Appendix A. ∎

Corollary 4.0.1 *For the optimal HICS, the inner portion of the transmitted sequence,* $\mathbf{s}$ *is given by*

$$\begin{aligned} \mathbf{s} &= \mathbf{x} + \mathbf{c}_{opt} + \alpha b \mathbf{u} \\ &= (\mathbf{I} - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T)\mathbf{x} + \alpha b \mathbf{u}. \end{aligned}$$

Remark 4.0.8 *At optimality, the constraint in (4.1) is active, i.e.,* λ_{opt} *is non-zero and chosen such that* $\|\mathbf{c}\|^2 \leq A$ *is satisfied with equality. Then* λ_{opt} *determines the tradeoff between the power of the HICS,* $\mathbf{c}$*, and the norm of the correlation vector between the HICS-embedded signal,* $\mathbf{x} + \mathbf{c}$*, and the linear correlation transform matrix,* $\mathbf{H}$*. Increasing* λ_{opt} *jointly reduces the effect of the cost function in the Lagrangian, and increases the effect of the constraint function. Thus, it can be shown that* λ_{opt} *is non-negative and monotonic decreasing in* A*.*

In Fig. 4.1, we present the experimental values for the expected value of the parameter A, when the Lagrangian multiplier λ is active and signal to watermark ratio (SWR), watermark to noise ratio (WNR) and tolerance size, M, are set to 5, 5 and 21, respectively. Details of calculating the SWR and WNR will be explained in Sec. 7. Using different SWR, WNR and tolerance region size values, we can obtain different values

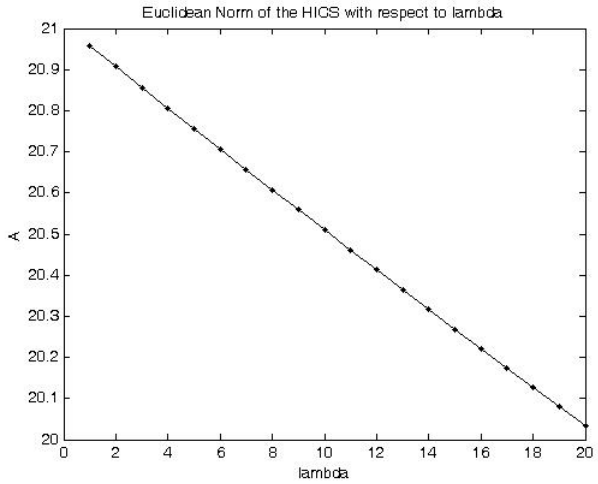


Figure 4.1. Eucliean norm of the HICS, with respect to lambda. SWR, WNR, and tolerance region size is set to 5, 5 and 21, respectively.

for the parameter A for the same λ. Nevertheless, we aim to visualize the behavior of the Euclidean norm of the HICS with respect to λ. λ can be chosen in an iterative fashion so as to satisfy a desired distortion. Note that an analogous approach is carried out in some state of the art lossy image compression algorithms; see for example [11].

5. ATTACK MODELING

In this section, we define the attack model which we assume to be inserted in the channel between the embedder and the decoder. The attack induced by the channel is modeled as a combination of AWGN attack and translation attack (Figure 3.1(b)).

Definition 5.0.4 *Overall attack is defined by the mapping* $t = \mathbb{R}^{\bar{N}} \longmapsto \mathbb{R}^{\bar{N}}$ *such that;*

$$t(\bar{\mathbf{s}}) = f_\theta(\bar{\mathbf{y}}) = f_\theta(\bar{\mathbf{s}} + \bar{\mathbf{n}})$$

where $\bar{\mathbf{n}} \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\mathbf{n}} \mathbf{I}_{\bar{N}})$ *is AWGN with zero-mean and variance* $\sigma_{\mathbf{n}}$ *and* $f_\theta(.) : \mathbb{R}^{\bar{N}} \longmapsto \mathbb{R}^{\bar{N}}$ *is cyclic shifting operator such that for any* $\mathbf{a} \in \mathbb{R}^{\bar{N}}$

$$[\mathbf{b} = f_\theta(\mathbf{a})] \Longleftrightarrow \left[\mathbf{b}_i = \mathbf{a}_{[(i-1+\theta) mod \bar{N}+1]} \ : \ 1 \leq i \leq \bar{N}\right] \tag{5.1}$$

Here, θ *is the translation amount and it is uniformly distributed on the set* $\{-L, -L+1, ..., L\}$*, where* L *denotes the maximum shift amount as defined in Section 3.*

Thus, the final attacked signal $\bar{\mathbf{r}} \in \mathbb{R}^{\bar{N}}$ is given by $\bar{\mathbf{r}} = f_\theta(\bar{\mathbf{y}})$, and this signal is the input for the decoder end (Figure 1.1).

Definition 5.0.5 *According to the sub-area definition of 3.0.1, we have:*

$$\mathbf{r}_1 \triangleq \mathcal{A}_1(\bar{\mathbf{r}}) \ ; \quad \mathbf{r}_2 \triangleq \mathcal{A}_2(\bar{\mathbf{r}}) \quad ; \mathbf{r} \triangleq \mathcal{A}(\bar{\mathbf{r}})$$
$$\mathbf{y}_1 \triangleq \mathcal{A}_1(\bar{\mathbf{y}}) \ ; \quad \mathbf{y}_2 \triangleq \mathcal{A}_2(\bar{\mathbf{y}}) \quad ; \mathbf{y} \triangleq \mathcal{A}(\bar{\mathbf{y}})$$

6. GISS DECODING

In this section, we propose three decoding methods, search/correlation decoding, joint MAP decoding and focused MAP decoding. These methods are used to extract binary information from the received signal, $\bar{\mathbf{r}}$, at the decoder end (See Figure 1.1).

Decoding is performed with the knowledge of the watermark signal, $\mathbf{u}$, at the decoder side for all decoding methods. The watermark signal is generated by seeding the shared key $K\$$ to a PRNG. Details of the watermark generating process is explained in Sec. 4.

In our setup, we assume that the received signal is watermarked, therefore, detection of the watermark signal is not our concern; our aim is to decode the embedded bit. Search/correlation and joint MAP decoding methods provides embedding location of the bit along with the sign information.

6.1. Search/Correlation Decoding - Method I

Search/correlation method first finds correlation values for all points in the tolerance region, producing the correlation vector, γ, and assigns the location where the absolute maximum correlation value is attained as the embedded location, $\hat{\theta}_1$. The sign of the correlation value determines the sign of the embedded bit, $\hat{b}_1$. These steps can be shown as follows:

1. Find $\hat{\theta}_1 \triangleq \text{argmax}_{-L\leq\theta\leq L} \mid < \mathcal{A}\left(f_\theta^{-1}(\bar{\mathbf{r}})\right), \alpha\mathbf{u} > \mid$
2. Then, $\hat{b}_1 = \text{sign}\left[\mathcal{A}(f_{\hat{\theta}_1}^{-1}(\bar{\mathbf{r}})), \alpha\mathbf{u} >\right]$

Here, $\mathcal{A}(.)$ is the sub-area function defined in Definition 3.0.1. $f_\theta^{-1}(.)$ is the inverse operation of the cyclic shift attack, $f_\theta(.)$, which is induced by the channel. Remember that, this shift is uniformly distributed on the set $\{-L, -L+1, ..., L\}$, so the inverse shifting is also performed for all values on the same set, and correlation values with

the inverse shifted received signal and the watermark signal is calculated in order to find the maximum correlation value.

The performance of this method can be evaluated by analyzing the probability of error. Assuming that we know the translation amount, the probability of error can be shown as:

$$\begin{aligned} P_e &= Pr\left[b \neq \hat{b}_1 \mid \theta\right] \\ P_e &= Pr\left[b \neq \hat{b}_1 \mid \theta, \hat{\theta}_1; \text{ s.t. } \hat{\theta}_1 \neq \theta\right] .Pr\left[\hat{\theta}_1 \neq \theta \mid \theta\right] \\ &\quad + Pr\left[b \neq \hat{b}_1 \mid \theta, \hat{\theta}_1; \text{ s.t. } \hat{\theta}_1 = \theta\right] .Pr\left[\hat{\theta}_1 = \theta \mid \theta\right] \end{aligned} \tag{6.1}$$

where θ is the actual translation and $\hat{\theta}_1$ is the detected translation amount found by Method I.

Here, we assume that there is a constraint on the magnitude of correlation of the shifted versions of the watermark;

$$\delta_{ij} = | < \mathbf{u}^i, \mathbf{u}^j > | < \kappa \quad \text{for all} \quad i \neq j \qquad 0 \leq i, j \leq 2L+1$$

where, $\kappa < 1$, $\mathbf{u}^i$ is the i^{th} circular rotated version of $\mathbf{u}$, or, equivalently, the i^{th} row of the linear correlation transform matrix, $\mathbf{H}$, as defined in Definition 4.0.2.

Remark 6.1.1 *Per our assumption and set up, we know that* $\delta_{ii} = 1$, *and* $\forall\ j \neq i$, $| \delta_{ij} | < \kappa < 1$.

Definition 6.1.1 *In the following derivations, we will use use* $\hat{\mathbf{y}} = \mathcal{A}(f_\theta^{-1}(\bar{\mathbf{r}}))$ *as the inner portion of the inverse shifted received signal. Using this signal and the linear correlation transform matrix,* $\mathbf{H}$, *defined in Definition 4.0.2, we will obtain correlation vector,* γ, *as follows:*

$$\gamma = \mathbf{H} \cdot \hat{\mathbf{y}}$$

If we know the translation amount, then the inner portion of the inverse shifted received signal, $\hat{\mathbf{y}}$ will be the inner portion of the AWGN attacked watermark embedded signal, $\mathcal{A}(\bar{\mathbf{y}})$, and the correlation vector will be:

$$\gamma = \mathbf{H} \cdot \mathcal{A}(\bar{\mathbf{y}}) = \underbrace{\mathbf{H} \cdot (\mathbf{I}_N - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T) \cdot \mathcal{A}(\bar{\mathbf{x}}) + \mathbf{H} \cdot \mathcal{A}(\bar{\mathbf{n}})}_{\triangleq \mathbf{Z}} + b\alpha \mathbf{H} \cdot \mathbf{u} \tag{6.2}$$

Note that, $\mathbf{Z}$ is also gaussian with $\mathbf{0}$-mean, since we assume that the host and the noise signals are gaussian with distributions $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\mathbf{x}} \mathbf{I}_{\bar{N}})$ and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\mathbf{n}} \mathbf{I}_{\bar{N}})$, respectively.

Proposition 6.1.1 *Auto-covariance of* $\mathbf{Z}$ *is:*

$$\Sigma_{\mathbf{Z}} = \mathbf{U}[\Sigma^2 - 2\Sigma\widetilde{\mathbf{S}}\Sigma + \Sigma\widetilde{\mathbf{S}}^2\Sigma]\mathbf{U}^T\sigma^2_{\mathbf{x}} + \mathbf{U}\Sigma\mathbf{U}^T\sigma^2_{\mathbf{n}} \tag{6.3}$$

Proof:
See Appendix B. ∎

Remark 6.1.2 *Since the watermark is assumed to be deterministic for the receiver who knows the key,* γ *is gaussian with mean* $b\alpha\mathbf{\Delta}$*, where* $\mathbf{\Delta_j} = \delta_{ij}$*, and variance* $\Sigma_{\mathbf{Z}}(j,j)$*.*

The probability of error expression given in the eq. (6.1) is analyzed in three steps, as follows:

1. Find $Pr\left[b \neq \hat{b}_1 \mid \hat{\theta}_1 \neq \theta\right]$.
   We know that $\gamma_j \sim \mathcal{N}(\alpha b \delta_{ij}, \Sigma_{\mathbf{Z}}(j,j))$ where $\delta_{ij} =< \mathbf{u}^i, \mathbf{u}^j >$. For this case,

$$P_e = Pr\left[\gamma_j > 0 \mid b = -1\right].\frac{1}{2} + Pr\left[\gamma_j < 0 \mid b = +1\right].\frac{1}{2}$$

   If $b = -1 \quad \Rightarrow \quad \gamma_j \sim \mathcal{N}(-\alpha\delta_{ij}, \Sigma_{\mathbf{Z}(j,j)})$.

If $b = +1 \quad \Rightarrow \quad \gamma_j \sim \mathcal{N}(\alpha\delta_{ij}, \Sigma_{\mathbf{Z}(j,j)})$.
Then, $P_e = Pr\left[b \neq \hat{b}_1 \mid \hat{\theta}_1 \neq \theta\right] = Q\left(\frac{\alpha\delta_{ij}}{\sqrt{\Sigma_{\mathbf{Z}(j,j)}}}\right)$.
Overall expression for $\pm b$ is shown as:

$$Q\left(\frac{\alpha\kappa}{\sqrt{\Sigma_{\mathbf{Z}}(j,j)}}\right) \leq Pr\left[b \neq \hat{b}_1 \mid \theta \neq \hat{\theta}_1\right] \leq Q\left(\frac{-\alpha\kappa}{\sqrt{\Sigma_{\mathbf{Z}}(j,j)}}\right)$$

and $\frac{1}{2} - \epsilon \leq Pr\left[b \neq \hat{b}_1 \mid \theta \neq \hat{\theta}_1\right] \leq \frac{1}{2} + \epsilon$.

2. Find $Pr[b \neq \hat{b}_1 \mid \theta = \hat{\theta}_1]$.
   If $\theta = \hat{\theta}_1$, then $E[\gamma_i] = E[\gamma_j] = \alpha b\|\mathbf{u}\|^2$ and $\mathrm{Var}(\gamma_i) = \mathrm{Var}(\gamma_j) = \Sigma_{\mathbf{Z}}(i,i)$

$$Pr\left[b \neq \hat{b}_1 \mid \theta = \hat{\theta}_1\right] = Q\left(\frac{\alpha\|\mathbf{u}\|}{\sqrt{\Sigma_{\mathbf{Z}}(i,i)}}\right)$$

3. Find $Pr(\theta = \hat{\theta}_1)$.
   It is not possible to find logical bounds for the $Pr(\theta = \hat{\theta}_1)$. So we decided to find numerical values by running tests on synthetic images. The results of the experiments are presented in Sec. 7.

6.2. Joint MAP Decoding - Method II

The MAP algorithm is a symbol-by-symbol estimator which accepts observations together with *a priori* symbol probabilities and produces *a posteriori* symbol probabilities. The decoded symbols are declared to be the ones with the maximum a posteriori probability.

We deploy the MAP algorithm for decoding in two different ways. In our first MAP decoding method, we decode the embedded bit and the embedding location jointly. In the second MAP decoding method, we find the marginal distribution of the received signal conditioned on the embedded bit, and apply the MAP decoding algorithm on this distribution. Details of this method will be explained in Sec. 6.3.

In the joint MAP method, we find out joint MAP estimates of (b, θ) such that:

$$(\hat{b}_2, \hat{\theta}_2) = \begin{array}{c} \text{argmax} \\ b \in \{\pm 1\} \\ -L \leq \theta \leq L \end{array} \quad p(\bar{\mathbf{r}}|b, \theta) \tag{6.4}$$

First note that since $f_\theta(.)$ is an invertible mapping, we have

$$p(\bar{\mathbf{r}}|b, \theta) = p(\hat{\mathbf{y}}|b)|_{\hat{\mathbf{y}}=f_\theta^{-1}(\bar{\mathbf{r}})} \tag{6.5}$$

Assume decoder knows $\mathbf{u}$ and assume that $\theta = 0$ (i.e., in the absence of a shifting attack). Then;

$$\begin{aligned} \bar{\mathbf{r}} = \bar{\mathbf{y}} &= \bar{\mathbf{s}} + \bar{\mathbf{n}} \\ \mathbf{r} = \mathbf{y} &= \mathbf{x} + \mathbf{c} + \mathcal{A}(\bar{\mathbf{n}}) + \alpha b \mathbf{u} \\ &= (\mathbf{I}_N - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T)\mathbf{x} + \mathcal{A}(\bar{\mathbf{n}}) + \alpha b \mathbf{u} \end{aligned}$$

Remember that, $\mathbf{y}$ and $\mathbf{x}$ are the inner portions of the AWGN added watermark embedded signal, $\bar{\mathbf{y}}$, and the host signal, $\bar{\mathbf{x}}$, respectively. Moreover, we will show the inner portion of the AWGN signal, $\bar{\mathbf{n}}$ as $\mathbf{n}$, in the rest of the thesis.

Definition 6.2.1 *Conditioned on b, we can define the statistical characteristics of* $\mathbf{y}$ *as follows:*

$$\begin{aligned} \mathbf{y} &\sim \mathcal{N}(\alpha b \mathbf{u}, \Sigma_{\mathbf{y}}) \\ \textit{where } \Sigma_{\mathbf{y}} &\triangleq E\left[(\mathbf{y} - \alpha b \mathbf{u})(\mathbf{y} - \alpha b \mathbf{u})^T\right] \\ &= E\left\{\left[(\mathbf{I} - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T)\mathbf{x} + \mathbf{n}\right]\left[\mathbf{x}^T(\mathbf{I} - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T) + \mathbf{n}^T\right]\right\} \\ &= \sigma_{\mathbf{x}}^2(\mathbf{I} - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T)(\mathbf{I} - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T) + \sigma_{\mathbf{n}}^2\mathbf{I}_N \\ &= \sigma_{\mathbf{x}}^2(\mathbf{I} - 2\mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T + \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T\mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T) + \sigma_{\mathbf{n}}^2\mathbf{I}_N \\ &= \mathbf{V}(\sigma_{\mathbf{n}}^2\mathbf{I}_N + \sigma_{\mathbf{x}}^2\Lambda)\mathbf{V}^T + (\sigma_{\mathbf{n}}^2 + \sigma_{\mathbf{x}}^2)\mathbf{W}\mathbf{W}^T \end{aligned} \tag{6.6}$$

Here $\Lambda \triangleq \mathbf{I} - 2\widetilde{\mathbf{S}} + \widetilde{\mathbf{S}}^2$ *(a positive definite diagonal matrix) and* $\mathbf{W}(N \times (N-M))$ *defined such that* $\mathbf{V}\mathbf{V}^T$ *and* $\mathbf{W}\mathbf{W}^T$ *are algebraic complements of each other.*

Definition 6.2.2 Λ *in the eq. (6.6) is a function of the singular values of the linear correlation transform matrix,* $\mathbf{H}$ *and lagrange multiplier* λ*. We have, for all* $i \in \{1, 2, ..., N\}$*:*

$$\begin{aligned} \Lambda_{ii} &= 1 - 2\frac{\sigma_i^2}{\lambda + \sigma_i^2} + \frac{\sigma_i^4}{(\lambda + \sigma_i^2)^2} \\ &= \frac{(\lambda + \sigma_i^2)^2 - 2\sigma_i^2(\lambda + \sigma_i^2) + \sigma_i^4}{(\lambda + \sigma_i^2)^2} \\ &= \left(\frac{\lambda}{\lambda + \sigma_i^2}\right)^2 \\ &= \left(1 + \frac{\sigma_i^2}{\lambda}\right)^{-2} \\ \textit{i.e., } \Lambda &= \left(\mathbf{I}_N + \frac{1}{\lambda}\Sigma_{\mathbf{H}}^2\right)^{-2} \quad . \end{aligned} \tag{6.7}$$

Taking the distributions of the host, watermark and noise signals into consideration, the AWGN added watermark embedded signal, $\bar{\mathbf{y}}$, is composed of three parts, the first and last parts have i.i.d. gaussian distributions with zero mean and their variance are related to the variances of the host and the noise signals. The watermark embedded part (the inner portion) is also a gaussian distribution, with mean $b\alpha\mathbf{u}$ and covariance matrix $\Sigma_{\mathbf{y}}$, which is defined in eq. (6.6). Then the joint distribution of $\bar{\mathbf{y}}$ is as follows:

$$p_{\bar{\mathbf{Y}}|b}(\bar{\mathbf{y}}|b) = \begin{cases} p_{\mathbf{Y}_1}(\mathbf{y}_1) & 1 \leq i < N_1 \\ p_{\mathbf{Y}|b}(\mathbf{y}|b) & N_1 \leq i < N_1 + N \\ p_{\mathbf{Y}_2}(\mathbf{y}_2) & N_1 + N \leq i < \bar{N} \end{cases}$$

We know that subareas of $\bar{\mathbf{y}}$ are independent from each other. Then we have:

$$p(\bar{\mathbf{y}}|b) = p(\mathbf{y}_1) \cdot p(\mathbf{y}_2) \cdot p(\mathbf{y}|b) \tag{6.8}$$

Proposition 6.2.1 *The joint MAP estimate given in eq. (6.4) can be rewritten using the probability distributions as follows:*

$$(\hat{b}_2, \hat{\theta}_2) = \underset{\substack{b \in \{\pm 1\} \\ -L \leq \theta \leq L}}{\operatorname{argmin}} \left[\frac{1}{\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2} \left[\left\| \mathcal{A}_1(f_\theta^{-1}(\bar{\mathbf{r}})) \right\|^2 + \left\| \mathcal{A}_2(f_\theta^{-1}(\bar{\mathbf{r}})) \right\|^2 \right] + \frac{1}{2} \left\| \Sigma_{\mathbf{y}}^{-1/2} \left[\mathcal{A}(f_\theta^{-1}(\bar{\mathbf{r}})) - \alpha b \mathbf{u} \right] \right\|^2 \right] \quad (6.9)$$

Proof:

See Appendix C. ∎

Remark 6.2.1 *The role of inclusion "pre" and "post" regions in eq. (6.9) may not be clear at first sight (since they do not posses any embedded information). However, not that in order to solve eq. (I-2), we carry out a brute force search of complexity* $2(2L+1)$*, which includes "trying" all possible values of* $\theta \in \{-L, -L+1, ..., L\}$*. When the true value of* θ*is tried, the incorporation of the first term of eq. (I-2) is "close" to* $\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2$*. However, for the wrong values of* θ*, they will clearly affect the outcome, which justifies incorporating both "pre" and "post" regions in the MAP decoding process.*

6.3. Focused MAP Decoding - Method III

In this method, we focus on finding out the MAP estimate of the embedded bit only. At the decoder side, we have the distribution information of the received signal conditioned on the embedded bit and the embedding location. Therefore, the decoded bit can be found as follows:

$$\hat{b}_3 = \underset{b \in \{\pm 1\}}{\operatorname{argmax}} p(\bar{\mathbf{r}}|b).$$

Proposition 6.3.1 *The focused MAP estimate of the embedded bit can be found as follows:*

$$\hat{b}_3 = \underset{b\in\{\pm 1\}}{\operatorname{argmax}}\, p(\bar{\mathbf{r}}|b) = \sum_{\theta=-L}^{L} p(\bar{\mathbf{r}}|b,\theta)$$

Proof:

See Appendix D ∎

Therefore, using the eq. (6.9) in the proposition 6.3.1:

$$\hat{b}_3 = \underset{b\in\{\pm 1\}}{\operatorname{argmax}} \sum_{\theta=-L}^{L} \exp\left\{ -\frac{\left\|\mathcal{A}_1(f_\theta^{-1}(\bar{\mathbf{r}}))\right\|^2 + \left\|\mathcal{A}_2(f_\theta^{-1}(\bar{\mathbf{r}}))\right\|^2}{2(\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{n}}^2)} - \frac{1}{2}\left\|\Sigma_{\mathbf{y}}^{-1/2}(\hat{\mathbf{y}} - \alpha b\mathbf{u})\right\|^2 \right\}$$

Remark 6.3.1 *In Method II, we jointly find estimates of the embedded bit and the translation amount (which contrast with finding an estimate of the embedded bit only, covered in Method III). Method II is more useful in situations where we aim to estimate the attack as well.*

Remark 6.3.2 *Defining the error events:*

$$\begin{aligned} E_2 &= Pr\left(\hat{b}_2, \hat{\theta}_2 \neq (b,\theta)\right) = Pr\left((\hat{b}_2 \neq b)\ OR\ (\hat{\theta}_2 \neq \theta)\right) \\ E_3 &= Pr(\hat{b}_3 \neq b) \end{aligned}$$

Note that Method II (resp. Method III) has $Pr(E_2)$ (resp. $Pr(E_3)$). Thus, we expect that $Pr\left(\hat{b}_2, \hat{\theta}_2 \neq (b,\theta)\right) \geq Pr(\hat{b}_3 \neq b)$. Further, we also expect $Pr(\hat{\theta}_1 \neq \theta) \geq Pr(\hat{\theta}_2 \neq \theta)$ since Method I is empirical and Method II relies on MAP; this expectation is confirmed by our experiments.

7. EXPERIMENTAL RESULTS

We evaluate the proposed scheme on synthetic data. The experimental setup is designed to compare the performances of three spread spectrum watermarking methods defined in the Sec. 4 using three different decoding algorithms presented in Sec. 6, under the attack model defined in Sec. 5. We set the host signal unit-variance zero-mean Gaussian distributed and modify the energy of the AWGN via the input parameter $SNR = SWR + WNR$, where $SNR = \frac{\|\mathcal{A}(\bar{\mathbf{x}})\|^2}{\|\mathcal{A}(\bar{\mathbf{n}})\|^2}$. "Signal to Watermark Ratio (SWR)" is defined as the "total embedding distortion" (distortion induced by the watermark signal and decorrelator sequence), used in order to modify the energy of the watermark signal. For SS method, this value is set as watermark strength, directly, i.e., $SWR = \frac{\|\mathcal{A}(\bar{\mathbf{x}})\|^2}{\|\alpha_{SS}\mathbf{w}\|^2}$. For ISS and GISS methods, since the decorrelation sequence adds distortion to host signal, watermark strength is adjusted to make the total distortion equal to the SWR, i.e., $SWR = \frac{\|\mathcal{A}(\bar{\mathbf{x}})\|^2}{\|(\alpha_{ISS}b + \lambda_{ISS}x)\mathbf{w}\|^2}$ for the ISS method and $SWR = \frac{\|\mathcal{A}(\bar{\mathbf{x}})\|^2}{\|\mathbf{c} + b\alpha_{GISS}\mathbf{w}\|^2}$ for the GISS method. The inner portion of the host signal, $\mathbf{x}$, the watermark signal, $\mathbf{u}$, and HICS, $\mathbf{c}$, are all vectors of size $N = 1000$. The core watermark signal, $\mathbf{w}$ is a vector of size $K = 100$. The translation amount, L, is controlled by tolerance region size, M, as defined in Sec. 5. In all experiments, the "total embedding distortion" (distortion induced by the watermark signal and decorrelator sequence) on the host image is fixed for all embedding methods, via modifying the energy of the watermark signal.

Probability of error calculation is realized by producing 50000 i.i.d. host signals and i.i.d. watermark signals, each. For all single realizations, the watermark signal is embedded to host signal using three embedding methods and the embedding bit is extracted using the three decoding methods. For search/correlation and joint MAP decoding methods, embedding location is also found.

The performances of embedding and decoding algorithms are evaluated by means of probability of error calculations in terms of bit error, location error and symbol error as follows.

- Probability of Bit Error $Pr(b \neq \hat{b})$: While calculating the probability of bit error with respect to SNR, tolerance region size is set to 21 and λ used in the calculation of HICS is set to 1. Figures 7.1, 7.2, 7.3 and 7.4 show the probability of error performances of the embedding methods with respect to SNR, under different decoding methods and different SWR values.

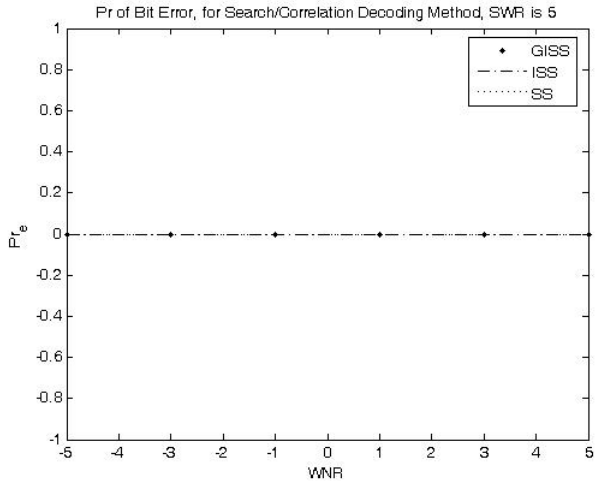
  When the SWR is low, the embedded bit can be decoded correctly using search/ correlation and joint MAP decoding methods (Figure 7.1). Increasing the SWR effects the decoding performance for GISS embedding of the search/correlation method more than the other methods (Figure 7.3 and 7.4). For all SWR values, joint MAP decoding method has the lowest probability of error values for the GISS embedding method.
- Probability of Location Error $Pr(\theta \neq \hat{\theta})$: Location information is not provided by the focused MAP decoder method. Therefore, the experiments are run on the other two decoding algorithms. Here, we set the SWR to 15 and vary the magnitude of the WNR from -5 to $+5$ with a step length of size 2; and the size of the tolerance region from 1 to 51 with a step length of size 10. Figures 7 and 7.6 show the probability of location error performances of the embedding methods with respect to tolerance region size, under different decoding algorithms.

  The experimental results show that for both decoding methods and for all WNR values that are used in the experiment, GISS method outperforms the other two embedding methods in terms of localization of the watermark signal.
- Probability of Symbol Error $Pr(b \neq \hat{b} \vee \theta \neq \hat{\theta})$: The experimental setup for calculation of probability of symbol error is similar to the setup explained above for the probability of location error. Figure 7.7 and 7.8 shows the probability of symbol error performances of the embedding methods with respect to tolerance region size, under different decoding algorithms.

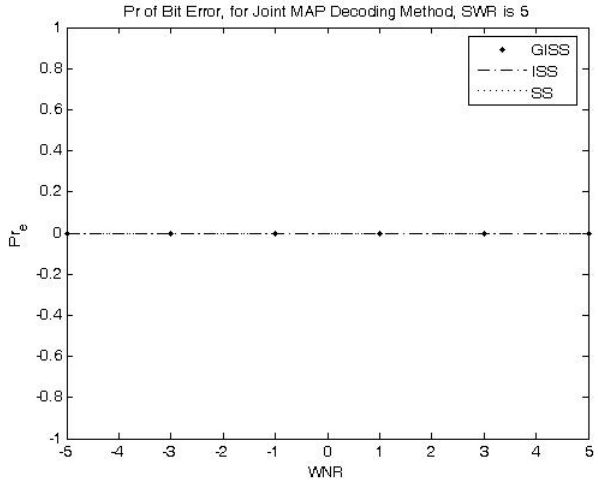
  Symbol error performance of both methods are approximately similar to their location error performances, since once we correctly localize the watermark signal in the received signal, probability of bit error reduces.

Considering probability of bit, location and symbol errors together, if our aim is to decode the bit and locate the watermark signal in the received signal correctly, then the best course of action will be to use the GISS embedding method at the embedding side and joint MAP method at the decoding side. Search/correlation method has a better performance in terms probability of bit error, when the embedding method is either SS or ISS, especially for high SWR values. However, at almost all SWR values, decoding GISS embedded signal with joint MAP has the best performance in terms of both probability of bit and location errors.

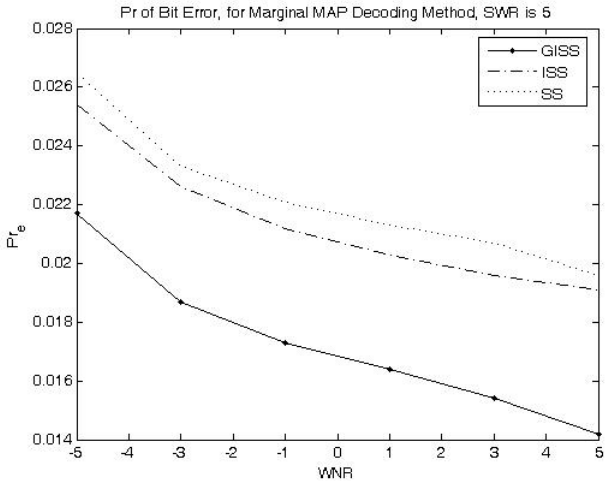
Note that, there is a bunch parameters in the evaluation of the performances, which are SWR, WNR, size of the tolerance region, lagrange multiplier, i.e., λ, of the optimization problem, signal and core watermark lengths. Thus we can produce different experimental results by varying these parameters. Before preparing the experimental setup, we run tests on smaller sets of host and watermark signals to determine the intervals of SWR, WNR and λ parameters, in order to obtain reasonable results. Sizes of the tolerance region, host and watermark signals are chosen in the limits of the equipments and programs used to realize the experiment.



(a)

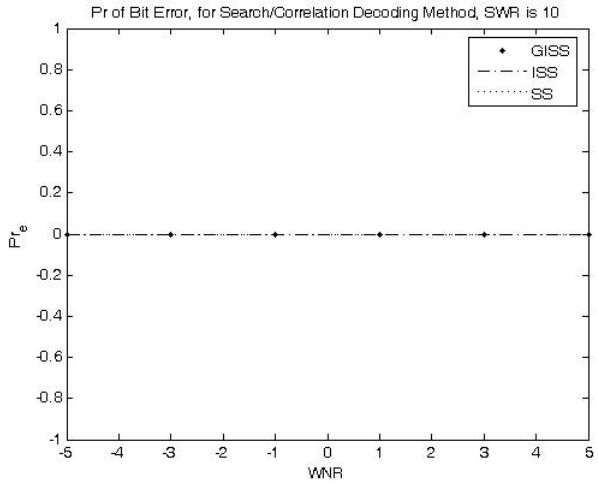


(b)

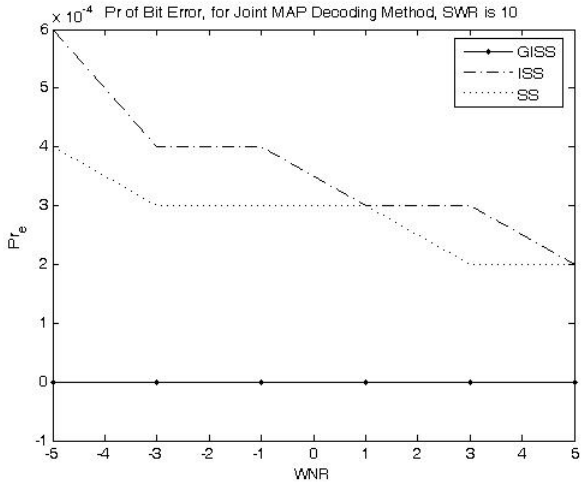


(c)

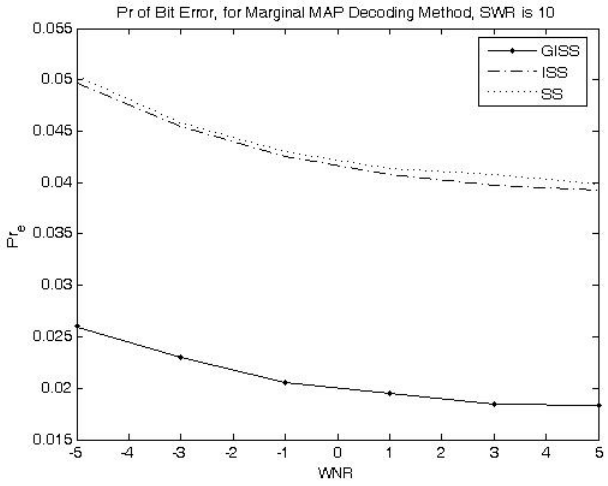
Figure 7.1. Probability of bit error with respect to WNR, SWR is 5: (a) searh/correlation method, (b) joint MAP method, (c) focused MAP method



(a)

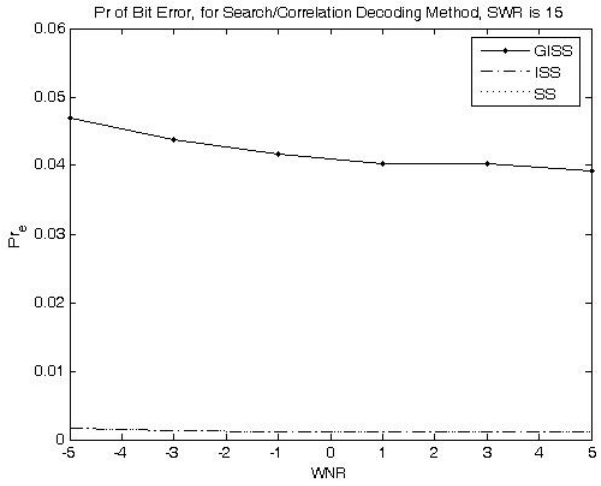


(b)

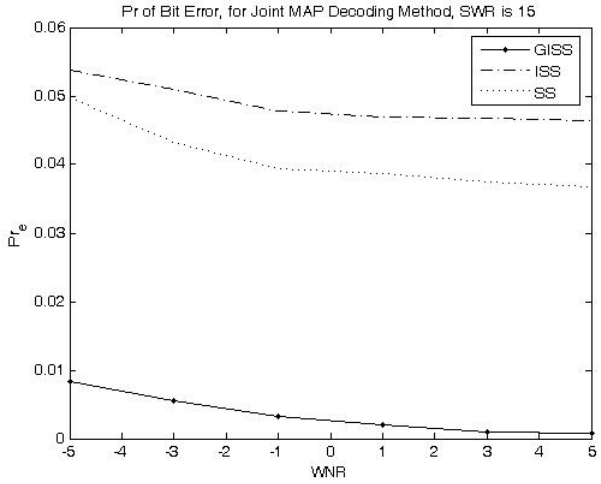


(c)

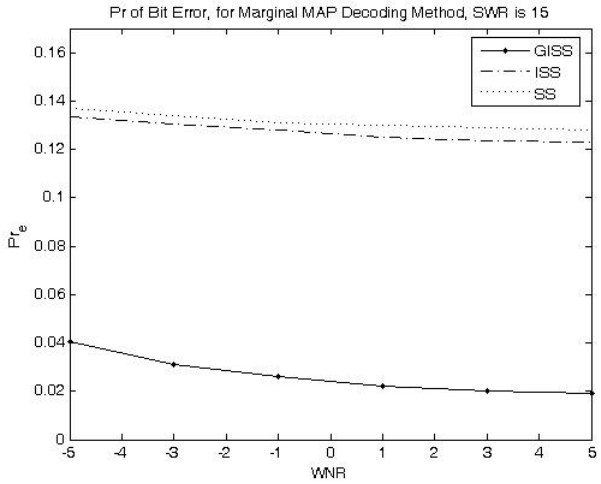
Figure 7.2. Probability of bit error with respect to WNR, SWR is 10: (a) searh/correlation method, (b) joint MAP method, (c) focused MAP method



(a)

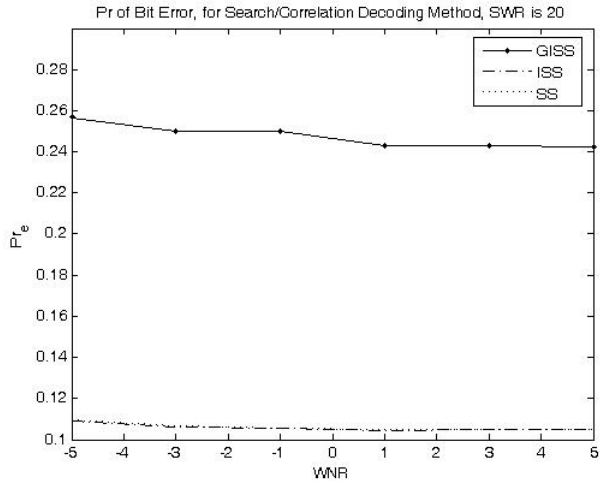


(b)

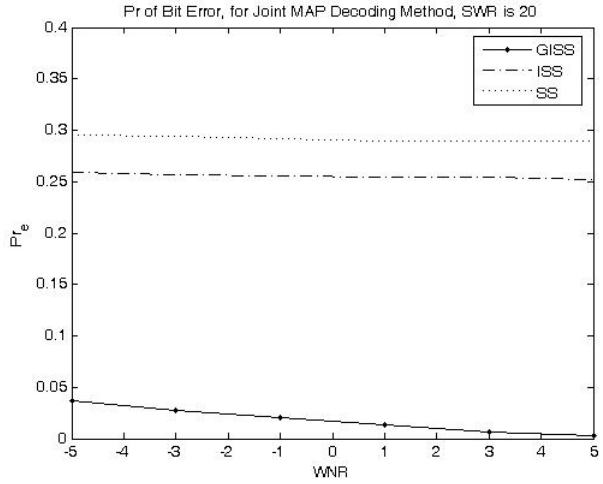


(c)

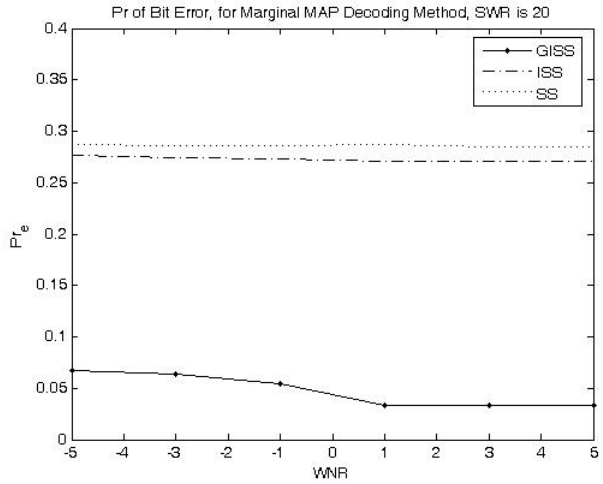
Figure 7.3. Probability of bit error with respect to WNR, SWR is 15: (a) searh/correlation method, (b) joint MAP method, (c) focused MAP method



(a)

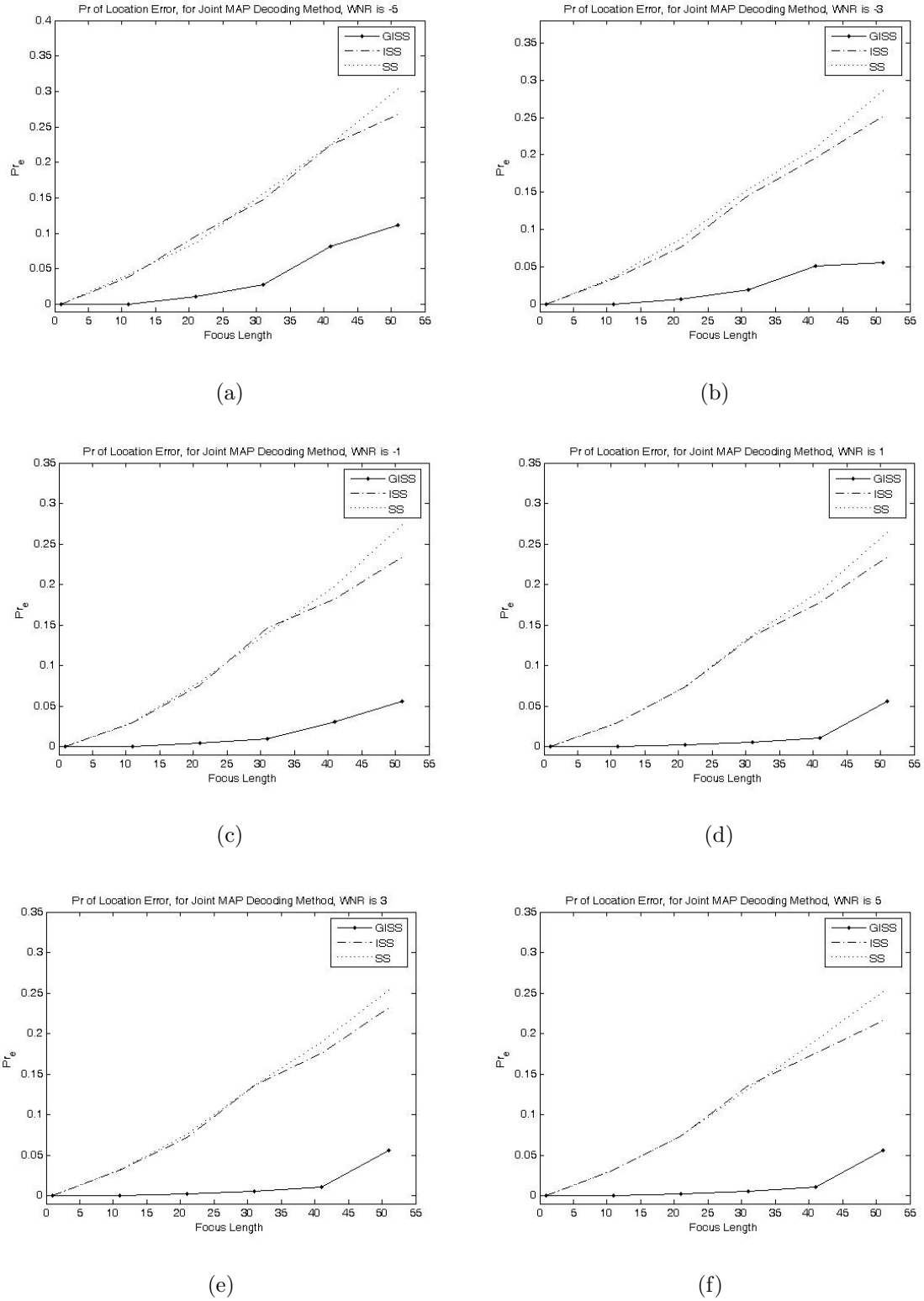


(b)



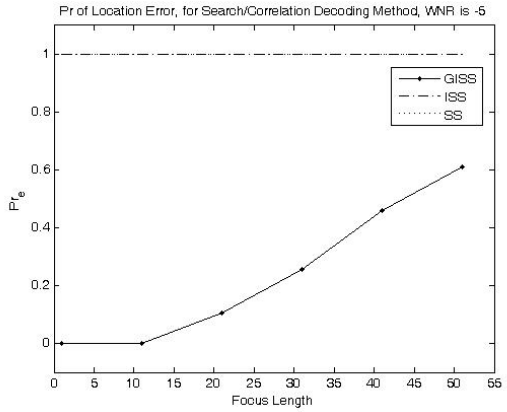
(c)

Figure 7.4. Probability of bit error with respect to WNR, SWR is 20: (a) searh/correlation method, (b) joint MAP method, (c) focused MAP method

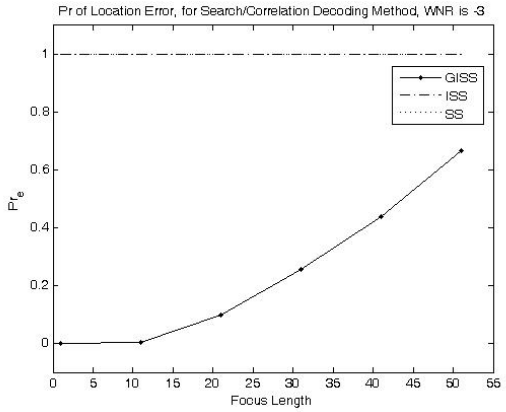


(a) (b) (c) (d) (e) (f)

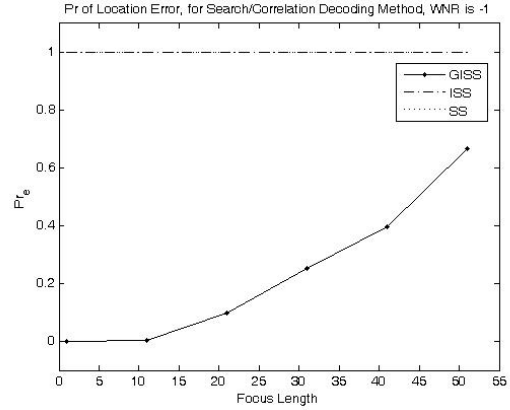
Figure 7.5. Probability of location error for joint MAP method using different WNR values, SNR is 15: (a) WNR is -5, (b) WNR is -3, (c) WNR is -1, (d) WNR is 1, (e) WNR is 3, (f) WNR is 5



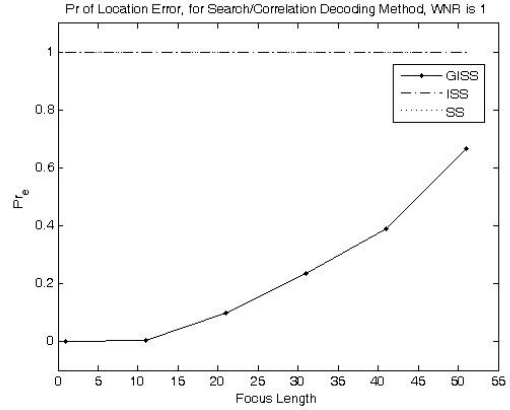
(a)



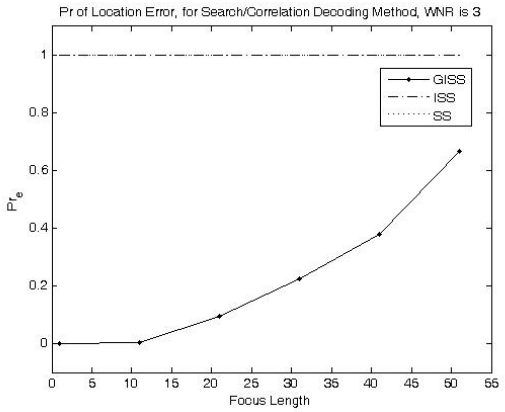
(b)



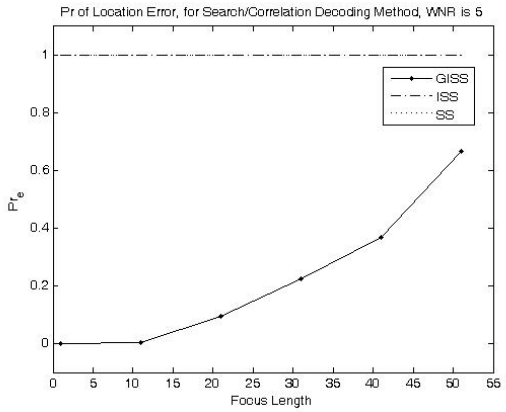
(c)



(d)

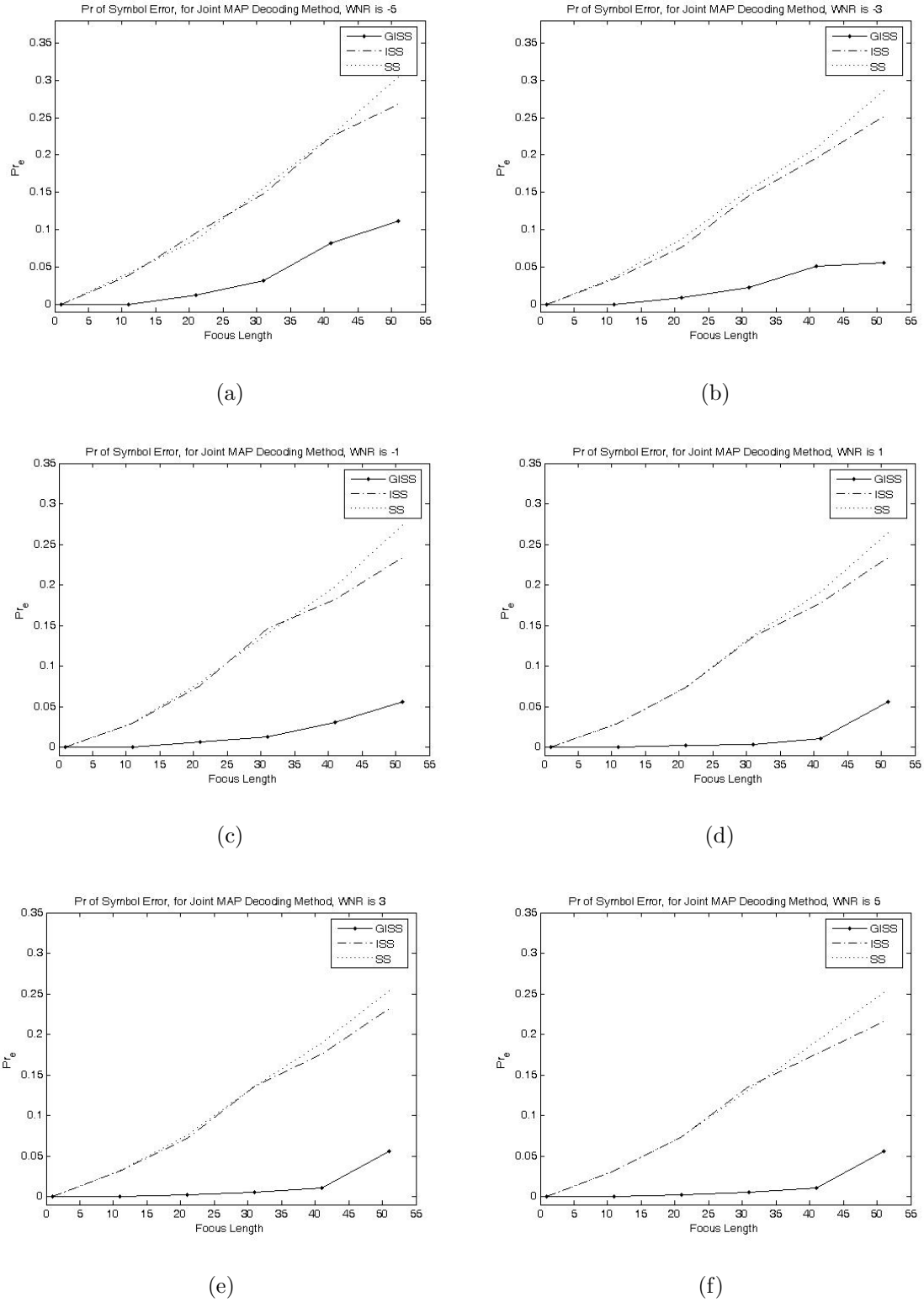


(e)



(f)

Figure 7.6. Probability of location error for search/correlation method using different WNR values, SNR is 15: (a) WNR is -5, (b) WNR is -3, (c) WNR is -1, (d) WNR is 1, (e) WNR is 3, (f) WNR is 5

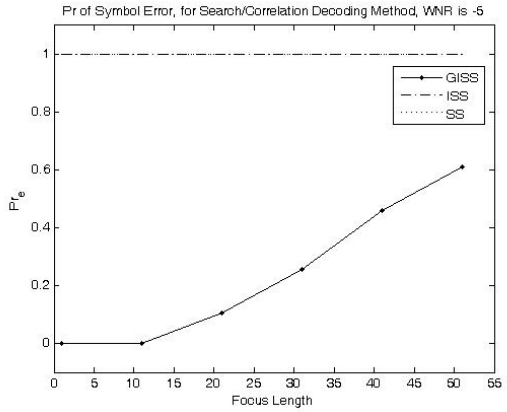


(a) (b)

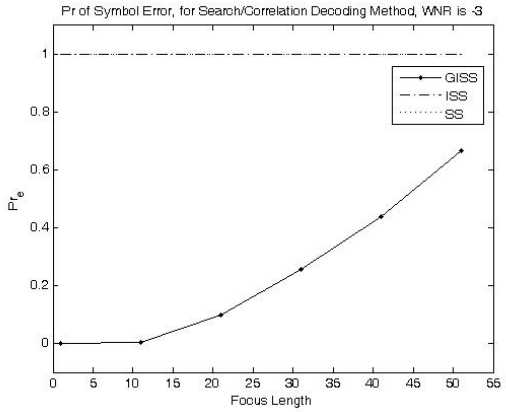
(c) (d)

(e) (f)

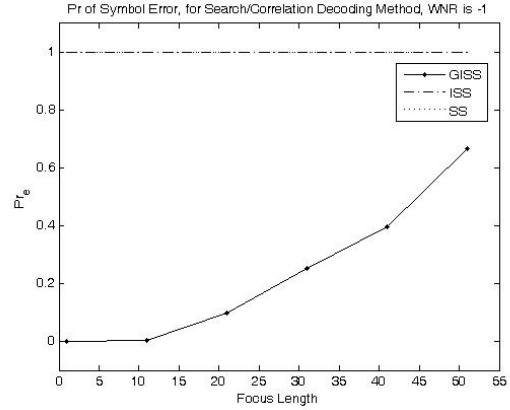
Figure 7.7. Probability of symbol error for joint MAP method using different WNR values, SNR is 15: (a) WNR is -5, (b) WNR is -3, (c) WNR is -1, (d) WNR is 1, (e) WNR is 3, (f) WNR is 5



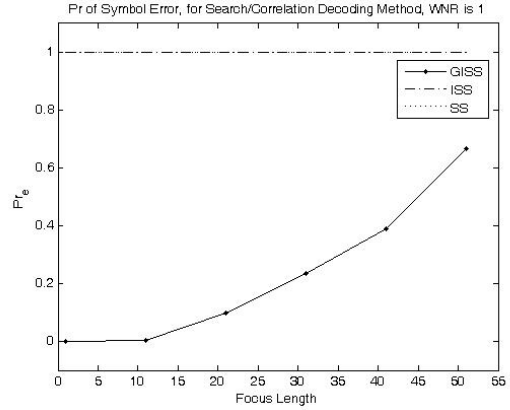
(a)



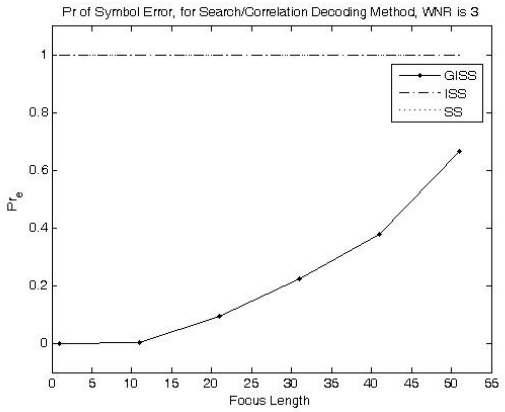
(b)



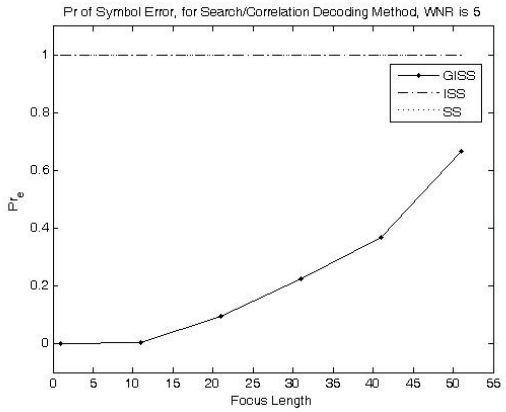
(c)



(d)



(e)



(f)

Figure 7.8. Probability of symbol error for search/correlation method using different WNR values, SNR is 15: (a) WNR is -5, (b) WNR is -3, (c) WNR is -1, (d) WNR is 1, (e) WNR is 3, (f) WNR is 5

8. CONCLUSIONS

In spread-spectrum watermarking systems, it is well-known that the correlation between the host signal and the watermark degrades the performance of the system. Improved Spread-Spectrum [1] aims to provide a solution to this problem via modifying the host prior to mark embedding, such that the correlation between the host signal and the watermark is reduced as much as possible. In the improved spread spectrum method, this is achieved via "decorrelating" the watermark and the host at the region of watermark embedding. In this case, geometric attacks are still problematic.

In this thesis, we propose a new watermarking strategy, termed Generalized Improved Spread Spectrum, where the strategy we follow is similar to, but an extended version of the one of Improved Spread Spectrum, so as to include robustness against translation-type geometric attacks. As a result of our method, we introduce local host interference cancelation sequence in order to reduce the correlation between the host signal and the watermark, not only at the region where the watermark is inserted, but also within a certain neighborhood. Therefore, robustness against translation-type attacks is attained.

Next, we define three decoding methods, namely search/correlation decoding, joint MAP decoding and focused MAP decoding in order to extract the binary information embedded in the received signal.

Finally, we apply the embedding schemes mentioned in the thesis to signals and demonstrate their effectiveness experimentally, using the proposed decoding algorithms. The experimental results shows that GISS method has a better decoding performance against translation attacks than other spread spectrum methods when the decoding algorithm is one of the MAP methods. Moreover, GISS method outperforms ISS and SS methods in terms of localization of the embedded watermark in the received signal, using any of the decoding methods.

Another inference from the experimental results can be the performance evaluation of the decoding methods. The search/correlation method has a better bit error performance in terms of ISS and SS methods compared to other decoding methods. However, this method has a very poor localization performance. Joint MAP decoding method is expected to perform better in terms of localizing the watermark, because this method tries to reduce the bit and location probability of error jointly.

In this thesis, we assume that the translation attack inserted by the channel has uniform distribution and it is confined with the maximum shift amount. In our future research, we plan to embed more than one watermark into the long stream of host data, and by detecting the embedding location, we will try to model the translation attack induced by the channel. Then using this model, we are planning to derive a more realistic distribution for the translation attack. In the experimental setup, we use synthetic data to realize the host signal. As a part of our future research, we plan to use practical multimedia data in the experiments and compare the results with the synthetic data results.

APPENDIX A: PROOF OF PROPOSITION 4.0.1

Consider the nonlinear constraint optimization problem given by (4.1). Here, we can write the Lagrangian for this optimization problem as

$$\begin{aligned}\mathfrak{L} &= \|\mathbf{H}(\mathbf{x}+\mathbf{c})\|^2 + \lambda\|\mathbf{c}\|^2 \\ &= \mathbf{c}^T\left(\mathbf{H}^T\mathbf{H}+\lambda\mathbf{I}_N\right)\mathbf{c} + 2\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{c} + \mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}.\end{aligned}$$

First, note that $\mathbf{H}^T\mathbf{H}+\lambda\mathbf{I}_N$ is positive-definite since $\lambda > 0$ and $\mathbf{H}^T\mathbf{H}$ is positive-semidefinite. As a result, $\mathfrak{L}$, which is quadratic in $\mathbf{c}$, is also convex in $\mathbf{c}$. Then, the necessary and sufficient condition for optimality is given by

$$\left[\nabla_{\mathbf{c}}\mathfrak{L}|_{\mathbf{c}=\mathbf{c}_{opt}} = 0\right] \quad \Longleftrightarrow \quad \left[\mathbf{c}_{opt} = -\left(\mathbf{H}^T\mathbf{H}+\lambda_{opt}\mathbf{I}_N\right)^{-1}\mathbf{H}^T\mathbf{H}\mathbf{x}\right], \tag{I-1}$$

where λ_{opt} is the value of the Lagrange multiplier λ at optimality.

Now, let SVD of $\mathbf{H}$ be given by $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Then, we can rewrite $\mathbf{c}_{opt}$ given by (I-1) as

$$\begin{aligned}\mathbf{c}_{opt} &= -\left(\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T+\lambda_{opt}\mathbf{I}_N\right)^{-1}\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{x} \\ &= -\left(\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T+\lambda_{opt}\mathbf{V}\mathbf{V}^T+\lambda_{opt}\mathbf{W}\mathbf{W}^T\right)^{-1}\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{x} \\ &= -\left(\mathbf{V}\left(\mathbf{\Sigma}^2+\lambda_{opt}\mathbf{I}_N\right)\mathbf{V}^T+\lambda_{opt}\mathbf{W}\mathbf{W}^T\right)^{-1}\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{x} \\ &= -\left(\mathbf{V}\left(\mathbf{\Sigma}^2+\lambda_{opt}\mathbf{I}_N\right)^{-1}\mathbf{V}^T+\frac{\mathbf{W}\mathbf{W}^T}{\lambda_{opt}}\right)\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{x} \qquad \text{(I-2)} \\ &= -\mathbf{V}\left(\mathbf{\Sigma}^2+\lambda_{opt}\mathbf{I}_N\right)^{-1}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{x} \qquad \text{(I-3)} \\ &= -\mathbf{V}\left(\mathbf{I}_N+\lambda_{opt}\mathbf{\Sigma}^{-2}\right)^{-1}\mathbf{V}^T\mathbf{x} \\ &= -\mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T\mathbf{x} \qquad \text{(I-4)}\end{aligned}$$

where $\mathbf{W}\mathbf{W}^T$ is the algebraic complement of $\mathbf{V}\mathbf{V}^T$. Here, (I-2) can be verified by direct computation, (I-3) follows from the fact that $\mathbf{W}^T\mathbf{V} = 0$, $\widetilde{\mathbf{S}}$ in (I-4) is diagonal and satisfies $\widetilde{\mathbf{S}}_{ii} = \frac{\sigma_i^2}{\lambda_{opt}+\sigma_i^2}$ for $1 \leq i \leq M$, where $\sigma_i = \mathbf{\Sigma}_{ii}$. So we are done. $\square$

APPENDIX B: PROOF OF PROPOSITION 6.1.1

Consider the correlation vector at the decoder end defined by the eq. (6.2). Here, we assume that the watermark signal is known by the decoder side, and define the indeterministic part as $\mathbf{Z}$:

$$\mathbf{Z} = \mathbf{H} \cdot (\mathbf{I}_N - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T) \cdot \mathcal{A}(\bar{\mathbf{x}}) + \mathbf{H} \cdot \mathcal{A}(\bar{\mathbf{n}})$$

In order to analyze the statistical behaviors of the received signal under given assumptions, we evaluate the covariance matrix of $\mathbf{Z}$:

$$\begin{aligned}
\Sigma_{\mathbf{Z}} &= \mathrm{E}(\mathbf{Z}\mathbf{Z}^T) \\
&= \mathbf{H}(\mathbf{I}_N - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T) \cdot (\mathbf{I}_N - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T)\mathbf{H}^T \cdot \sigma_{\mathbf{x}} + \mathbf{H}\mathbf{H}^T\sigma_{\mathbf{n}} \\
&= (\mathbf{H}\mathbf{H}^T - 2\mathbf{H}\mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T + \mathbf{H}\mathbf{V}\widetilde{\mathbf{S}}\underbrace{\mathbf{V}^T\mathbf{V}}_{\mathbf{I}_M}\widetilde{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T)\sigma_{\mathbf{x}}^2 + \mathbf{H}\mathbf{H}^T\sigma_{\mathbf{n}} \\
&= (\mathbf{H}\mathbf{H}^T - 2\mathbf{H}\mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T + \mathbf{H}\mathbf{V}\widetilde{\mathbf{S}}^2\mathbf{V}^T\mathbf{H}^T)\sigma_{\mathbf{x}}^2 + \mathbf{H}\mathbf{H}^T\sigma_{\mathbf{n}}
\end{aligned}$$

Let the SVD of $\mathbf{H}$ be given by $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^T$. Then, since $\widetilde{\mathbf{S}}(i,i) = \frac{\sigma_i}{\lambda+\sigma_i}$; $\widetilde{\mathbf{S}}$ can be represented as: $\widetilde{\mathbf{S}} = \Sigma \cdot (\lambda\mathbf{I}_M + \Sigma)^{-1}$. Moreover, $\mathbf{H}\mathbf{H}^T = \mathbf{U}\Sigma\mathbf{V}^T \cdot \mathbf{V}\Sigma\mathbf{U}^T = \mathbf{U}\Sigma^2\mathbf{U}^T$.

Analyzing all the terms one-by-one, we can obtain the following results;

- $\mathbf{H}\mathbf{H}^T = \mathbf{U}\Sigma^2\mathbf{U}^T$
- $\mathbf{H}\mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T\mathbf{H}^T = (\mathbf{U}\Sigma\mathbf{V}^T)(\mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T)(\mathbf{V}\Sigma\mathbf{U}^T) = \mathbf{U}\Sigma\widetilde{\mathbf{S}}\Sigma\mathbf{U}^T$
- $\mathbf{H}\mathbf{V}\widetilde{\mathbf{S}}^2\mathbf{V}^T\mathbf{H}^T = (\mathbf{U}\Sigma\underbrace{\mathbf{V}^T)(\mathbf{V}}\widetilde{\mathbf{S}}^2\underbrace{\mathbf{V}^T)(\mathbf{V}}\Sigma\mathbf{U}^T) = \mathbf{U}\Sigma\widetilde{\mathbf{S}}^2\Sigma\mathbf{U}^T$

Then;

$$\Sigma_{\mathbf{Z}} = \mathbf{U}[\Sigma^2 - 2\Sigma\widetilde{\mathbf{S}}\Sigma + \Sigma\widetilde{\mathbf{S}}^2\Sigma]\mathbf{U}^T\sigma_{\mathbf{x}}^2 + \mathbf{U}\Sigma\mathbf{U}^T\sigma_{\mathbf{n}}^2$$

Remember that $\gamma = \mathbf{H}(\mathbf{I}_N - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T)\mathbf{x} + \mathbf{H}\mathbf{n} + b\alpha\mathbf{H}\mathbf{u}$ and which can be simplified for ease in calculations as $\gamma = \mathbf{Z} + b\alpha\mathbf{H}\mathbf{u}$ where $\mathbf{Z} = \mathbf{H}(\mathbf{I}_N - \mathbf{V}\widetilde{\mathbf{S}}\mathbf{V}^T)\mathbf{x} + \mathbf{H}\mathbf{n}$. Now, since both noise and host signals are assumed to be gaussian signals, we can conclude that $\mathbf{Z}$ is also a gaussian signal with mean $\mathbf{0}$ and variance $\Sigma_{\mathbf{Z}} = \mathbf{U}[\Sigma^2 - 2\Sigma\widetilde{\mathbf{S}}\Sigma + \Sigma\widetilde{\mathbf{S}}^2\Sigma]\mathbf{U}^T\sigma_{\mathbf{x}}^2 + \mathbf{U}\Sigma\mathbf{U}^T\sigma_{\mathbf{n}}^2$. □

APPENDIX C: PROOF OF PROPOSITION 6.2.1

We know that the distributions of outer portions of the AWGN added watermark embedded signal, $\bar{\mathbf{y}}$, are $\mathbf{y}_1 \sim \mathcal{N}(\mathbf{0}, (\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)\mathbf{I}_{N_1})$ and $\mathbf{y}_2 \sim \mathcal{N}(\mathbf{0}, (\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)\mathbf{I}_{N_2})$. Then,

$$\begin{aligned} p(\mathbf{y}_1) &= \prod_{i=1}^{N_1} \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2}} \exp\left[-\frac{\mathbf{y}_{1,i}^2}{2(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)}\right] \\ p(\mathbf{y}_2) &= \prod_{i=1}^{N_2} \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2}} \exp\left[-\frac{\mathbf{y}_{2,i}^2}{2(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)}\right] \end{aligned}$$

Moreover, the distribution of inner portions of $\bar{\mathbf{y}}$ is also gaussian; i.e., $\mathbf{y} \sim \mathcal{N}(b\alpha\mathbf{u}, \Sigma_{\mathbf{y}})$, where $\Sigma_{\mathbf{y}}$ is defined in eq. (6.6). Thus we have:

$$p(\mathbf{y}|b) = \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma_{\mathbf{y}}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y} - \alpha b\mathbf{u})^T\Sigma^{-1}(\mathbf{y} - \alpha b\mathbf{u})\right]$$

where $\Sigma_{\mathbf{y}}$ is determined via (6.6) and (6.7). Thus, we have:

$$\log p(\mathbf{y}_1) = -\frac{N_1}{2}\log[2\pi(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)] - \frac{1}{2(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)}\sum_{i=1}^{N_1}\mathbf{y}_{1,i}^2 \quad \text{(I-1)}$$

$$\log p(\mathbf{y}_2) = -\frac{N_2}{2}\log[2\pi(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)] - \frac{1}{2(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)}\sum_{i=1}^{N_2}\mathbf{y}_{2,i}^2 \quad \text{(I-2)}$$

$$\log p(\mathbf{y}|b) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_{\mathbf{y}}| - \frac{1}{2}(\mathbf{y} - \alpha b\mathbf{u})^T\Sigma_{\mathbf{y}}^{-1}(\mathbf{y} - \alpha b\mathbf{u}) \quad \text{(I-3)}$$

Using (I-1), (I-2), (I-3) in (6.8);

$$\begin{aligned} \log p(\bar{\mathbf{y}}|b) &= \log p(\mathbf{y}_1) + \log p(\mathbf{y}_2) + \log p(\mathbf{y}|b) \\ &= -\frac{\bar{N}}{2}\log(2\pi) - \frac{N_1 + N_2}{2}\log(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2) - \frac{1}{2}\log|\Sigma_{\mathbf{y}}| \\ &\quad -\frac{1}{2(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)}\left(\sum_{i=1}^{N_1}\mathbf{y}_{1,i}^2 + \sum_{i=1}^{N_2}\mathbf{y}_{2,i}^2\right) \\ &\quad -\frac{1}{2}(\mathbf{y} - \alpha b\mathbf{u})^T\Sigma_{\mathbf{y}}^{-1}(\mathbf{y} - \alpha b\mathbf{u}) \end{aligned} \quad \text{(I-4)}$$

Using (I-4) in (6.5), we have:

$$\begin{aligned}\log p(\bar{\mathbf{r}}|b,\theta) &= K - \frac{1}{2(\sigma_{\mathbf{x}}^2)+\sigma_{\mathbf{n}}^2}\left[\left\|\mathcal{A}_1(f_\theta^{-1}(\bar{\mathbf{r}}))\right\|^2 + \left\|\mathcal{A}_2(f_\theta^{-1}(\bar{\mathbf{r}}))\right\|^2\right] \\ &\quad -\frac{1}{2}\left\|\Sigma_{\mathbf{y}}^{-1/2}\left[\mathcal{A}(f_\theta^{-1}(\bar{\mathbf{r}})) - \alpha b\mathbf{u}\right]\right\|^2 \end{aligned} \tag{I-5}$$

where $K \triangleq -\frac{\bar{N}}{2}\log(2\pi) - \frac{N_1+N_2}{2}\log(\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{n}}^2) - \frac{1}{2}\log|\Sigma_{\mathbf{y}}|$ is constant in (b,θ).

Remember that the Joint MAP estimate is as follows:

$$\begin{aligned}(\hat{b}_2,\hat{\theta}_2) = \quad & \underset{\substack{b\in\{\pm 1\}\\ -L\le\theta\le L}}{\text{argmax}} \quad p(\bar{\mathbf{r}}|,b,\theta),\end{aligned}$$

which can be rewritten as:

$$\begin{aligned}(\hat{b}_2,\hat{\theta}_2) = \quad & \underset{\substack{b\in\{\pm 1\}\\ -L\le\theta\le L}}{\text{argmax}} \quad \log p(\bar{\mathbf{r}}|,b,\theta)\end{aligned}$$

Then, taking the eq. (I-5) into consideration, we have

$$\begin{aligned}(\hat{b}_2,\hat{\theta}_2) &= \underset{\substack{b\in\{\pm 1\}\\ -L\le\theta\le L}}{\text{argmin}} \quad \left[\frac{1}{\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{n}}^2}\left[\left\|\mathcal{A}_1(f_\theta^{-1}(\bar{\mathbf{r}}))\right\|^2 + \left\|\mathcal{A}_2(f_\theta^{-1}(\bar{\mathbf{r}}))\right\|^2\right]\right. \\ &\qquad \left. +\frac{1}{2}\left\|\Sigma_{\mathbf{y}}^{-1/2}\left[\mathcal{A}(f_\theta^{-1}(\bar{\mathbf{r}})) - \alpha b\mathbf{u}\right]\right\|^2\right]\end{aligned}$$

□

APPENDIX D: PROOF OF PROPOSITION 6.3.1

The focused MAP estimate is based on the Bayes rule, which relates the conditional and marginal probabilities of two random events. We have:

$$\begin{aligned} p(\bar{\mathbf{r}}|b) &= \sum_{\theta=-L}^{L} p(\bar{\mathbf{r}}, \theta|b) \\ &= \sum_{\theta=-L}^{L} p(\bar{\mathbf{r}}|b, \theta).p(\theta) \end{aligned}$$

(since θ is independent of b). Then,

$$p(\bar{\mathbf{r}}|b) = \sum_{\theta=-L}^{L} p(\bar{\mathbf{r}}|b, \theta)\frac{1}{2L+1} \qquad (\text{since } \theta \sim \mathcal{U}(\{-L, ..., L\}))$$

Thus,

$$\hat{b}_3 = \underset{b\in\{\pm 1\}}{\operatorname{argmax}}\, p(\bar{\mathbf{r}}|b) = \underset{b\in\{\pm 1\}}{\operatorname{argmax}} \sum_{\theta=-L}^{L} p(\bar{\mathbf{r}}|b, \theta)$$

where:

$$\begin{aligned} p(\bar{\mathbf{r}}|b,\theta) &= p(\mathbf{y}_1)|_{\mathbf{y}_1=\mathcal{A}_1(f_\theta^{-1}(\bar{\mathbf{r}}))} \cdot p(\mathbf{y}_2)|_{\mathbf{y}_2=\mathcal{A}_2(f_\theta^{-1}(\bar{\mathbf{r}}))} \cdot p(\mathbf{y})|_{\mathbf{y}=\mathcal{A}(f_\theta^{-1}(\bar{\mathbf{r}}))} \qquad \text{(I-1)} \\ &= (2\pi)^{-\bar{N}/2} \cdot (\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)^{-(N_1+N_2)/2} \cdot |\Sigma_{\mathbf{y}}|^{-1/2} \\ &\quad \cdot \exp\left[-\frac{1}{(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2)}\left(\|\mathbf{y}_1\|^2 + \|\mathbf{y}_2\|^2\right)\right] \\ &\quad \cdot \exp\left[-\frac{1}{2}\left\|\Sigma_{\mathbf{y}}^{-1/2}(\mathbf{y} - \alpha b\mathbf{u})\right\|^2\right] | \qquad \text{(I-2)} \\ &\qquad \mathbf{y}_1 = \mathcal{A}_1(f_\theta^{-1}(\bar{\mathbf{r}})) \\ &\qquad \mathbf{y}_2 = \mathcal{A}_2(f_\theta^{-1}(\bar{\mathbf{r}})) \\ &\qquad \mathbf{y} = \mathcal{A}(f_\theta^{-1}(\bar{\mathbf{r}})) \end{aligned}$$

Note that, (I-1) follows using the eq. (6.8), and (I-2) follows the distribution information given in the Appendix C.

REFERENCES

1. Malvar H. S. and D. A. F. Florencio, "Improved Spread Spectrum: A New Modulation Technique for Robust Watermarking", *IEEE Trans. Signal Proc.*, Vol. 51, No. 4, 2003.

2. *IEEE Journal on Selected Areas in Communications*. Special Issue on Copyright and Privacy Protection, vol. 16, no. 4, May 1998.

3. *Proceedings of the IEEE*. Special Issue on Identification and Protection of Multimedia Information, vol. 87, no.7, July 1999.

4. Petitcolas F. A. P., R. J. Anderson, and M. G. Kuhn, "Information Hiding - a Survey," *Proceedings of the IEEE*. Special Issue on Protection of Multimedia Content, vol. 87, no. 7, pp. 1062–1078, july 1999.

5. Cox I. J., M. L. Miller, and J. A. Bloom, *Digital Watermarking*. San Fransisco: Morgan Kaufmann Publishers, 2001.

6. Egger J., "Information Embedding and Digital Watermarking as Communication with Side Information," Ph.D. dissertation, University of Erlangen, Erlangen, Germany, 2002.

7. Moulin P., J. A. O'sullivan, "Information Theoretic Analysis of Information Hiding," *IEEE Trans. Inform. Theory*, vol. 49, pp. 563–593, 1999.

8. Moulin P., M. K. Mıhçak, *IEEE Trans. Inform. Theory*, vol. 50, pp. 272–289, 2004.

9. Cox I. J., M. L. Miller, and J. A. Bloom, "Watermarking Applications and Their Properties," presented at international Conference on Information Technology, Las Vegas, 2000.

10. Cox I. J., J. Kilian, F. T. Leighton, and T. Shamoon, "Secure Spread Spectrum

Watermarking for Multimedia", *IEEE Trans. Image Proc.*, Vol. 6, No. 12, pp. 1637–1687, 1997.

11. Lopresto S. M., K. Ramch, M. T. Orchard, "Image Coding Based on Mixture Modeling of Wavelet Coefficients and a Fast Estimation-Quantization Framework," presented at Data Compression Conference, Snowbird, Utah, 1997.

12. Horn R. A. and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

13. Lichtenauer J., I. Setyawan, T. Kalker, and R. Lagendijk, "Exhaustive Geometrical Search and the False Positive Watermark Detection Probability," *SPIE EI, Security and Watermarking of Multimedia Contents V*, San Jose, CA, Jan. 2003.

14. Cox I. J., M. L. Miller, "A review of watermarking and the importance of perceptual modeling," *Proc. of the Electronic Imaging*, pp. 92–97, Feb. 1997.

15. Cox I. J., J. Kilian, F. T. Leighton, and T. Shamoon, "A secure robust watermark for multimedia," *presented at the First International Information Hiding Workshop*, Cambridge, UK, Apr. 1996.

16. Zhao J., "Applying digital watermarking techniques to online multimedia commerce," *In Proc. of the International Conference on Imaging Science, Systems, and Applications*, Las Vegas, 1997.

17. Tirkel A. Z., C. F. Osborne, "Image watermarking - a spread spectrum application," *In Proc. IEEE Int. Symposium on Spread Spectrum Techniques and Applications*, pp. 785–789, 1996.

18. Coumou D. J., G. Sharma, "Watermark synchronization for featurebased embedding: Application to speech," *IEEE ICME*, pp. 849–852, 2006.

19. Swanson M. D., M. Koyabashi, A. H. Tewfik, "Multimedia data-embedding and watermarking strategies," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1064–1087, June 1998.

20. Borwein J. M., *Convex Analysis and Nonlinear Optimization : Theory and Examples*. New York : Springer, c2006

21. Chen B., G. W. Wornell, "An information-theoretic approach to the design of robust digital watermarking systems," *in Proceedings of IEEE International Conferenceon Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2061–2064, 1999.

22. Chen B., 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, pp. 1423–1443, May 2001.