Hermite Polynomials as Provably Good Functions to Watermark White Gaussian Hosts

Teddy Furon*
INRIA - Temics
Campus de Beaulieu
35042 Rennes, France
teddy.furon@irisa.fr

ABSTRACT

In the watermark detection scenario, also known as zero-bit watermarking, a watermark, carrying no hidden message, is inserted in content. The watermark detector checks for the presence of this particular weak signal in content. The article looks at this problem from a classical detection theory point of view, but with side information enabled at the embedding side: the watermark signal is a function of the host content. Our study is twofold. The first issue is to design the best embedding function for a given detection function (a Neyman-Pearson detector structure is assumed). The second issue is to find the best detection function for a given embedding function. This yields two conditions, which are mixed into one 'fundamental' partial differential equation.

Solutions of this fundamental equation are heavily dependent on the probability distribution function of the host signals. This conference paper is an extract of [7], where we only look at white gaussian hosts. This gives birth to polynomials solutions known as Hermite polynomial, whose extension is the JANIS watermarking scheme, invented heuristically some years ago.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Theory

Keywords

Watermarking, Detection theory, Pitman Noether theorem

1. INTRODUCTION

In the past six years, side-informed embedding strategies have been shown to greatly improve watermark *decoding*. They exploit knowledge of the host signal during the construction of the watermark signal. The theory underlying these side-informed schemes was presented in the famous paper "Writing on Dirty paper" by M. Costa in 1983. Our work gives some theoretical aspects of the achievable performances when using side-information at the embedding side but for the watermark *detection* problem (a.k.a. zero-bit watermarking [4, Sect. 2.2.3]). This surprisingly received almost no study compared to the issue of watermark decoding, although it is perceived as a non trivial problem [11].

The trade-off between payload of the hidden message and robustness is a well known fact in watermarking. The main rationale for zero-bit watermarking is that the maximum robustness that a watermarking primitive can inherently offer, is expected as the payload is reduced to the minimum.

Some copy protection platforms [1] use watermarks as flags whose presence warns compliant devices that the piece of content they are dealing with, is a copyrighted material. Content access and copy protection are tackled by cryptographic primitives. Watermarking just prevents the 'analog hole' [16, 6, 9]. Therefore, zero-bit watermarking is sufficient in this application.

Copyright protection is the most famous application of watermarking. However, hiding the name of the author in his Work is just a fact having no legal value. In Europe, the author must be first a member of an author society, then he registers his Work. The only legal proof is to give evidence that the suspicious image is indeed a version of a Work duly registered in an author society's database. Consequently, this is a yes/no question, which can be solved by detecting the presence or absence of a watermark previously embedded by an author society. Here again, zero-bit watermarking is sufficient for this application.

The attacker obviously knows which content is watermarked. In the copy protection application, for instance, there is no point in attacking a personal video which is a free content.

2. STRATEGY AND NOTATION

Our goal is not to derive an accurate statistical model of the host signal as done in prior works. On contrary, a very basic assumption, ie. white gaussian distribution, is considered in order to stress the major role of side information at the embedding side.

2.1 Embedding side

The embedder transforms an original host signal $\mathbf{s}$ into a watermarked content $\mathbf{y} = \mathbf{f}(\mathbf{s}) = \mathbf{s} + \mathbf{x}$. The host signal or

*supported by the ACI-SI Fabriano French National Project.

MM&Sec'06, September 26–27, 2006, Geneva, Switzerland.

channel state $\mathbf{s}$ is a vector of n components of the original content, modeled as random variables. The notational key of the article is to decompose the watermark signal $\mathbf{x}$ as a unit power vector $\mathbf{w}$ and an amplitude θ.

$$\mathbf{f}(\mathbf{s}) = \mathbf{s} + \mathbf{x} = \mathbf{s} + \theta\mathbf{w}(\mathbf{s}). \quad (1)$$

$\mathbf{w}(.)$ is a smooth function from $\mathbb{R}^n$ to $\mathbb{R}^n$, with the constraint $\mathrm{E}\{\|\mathbf{w}(\mathbf{s})\|^2\} = n$ ($\mathrm{E}\{.\}$ denotes the expectation). It is a kind of direction pointing to an acceptance region of $\mathbb{R}^n$, towards which the host signal should be pushed. The scalar θ controls the gain or amplitude of the watermark signal. In practice, host contents might bear different watermark power depending on their individual masking property. This change might even occur within a content, such that we should resort to a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ gathering positive and small gains affecting each sample. In this case, the developments given in the sequel are still possible with a vectorial notation. Yet, for simplicity's sake, we assume that θ is taken as the average gain of $\boldsymbol{\theta}$.

Both parts of the watermark signal depends on the host content, either through side information, either for some perceptual reasons. Unfortunately, in blind schemes, side information is not made available at the detection side. Moreover, we wish to maintain a low detector's complexity; hence, no human visual or auditive system can recreate an estimate of θ. Another point is that strong attacks certainly spoil this estimation, and no study have been done on the robustness of the detector with respect to this parameter. In our strategy, the detector just knows that the watermark amplitude is positive and small. We believe this model allows a great flexibility which eases practical implementations of highly robust watermarking schemes.

2.2 Detection side

Upon receipt of signal $\mathbf{r}$, the detector makes a binary decision: $d = 1$ ($d = 0$) means that, according to the detector, the piece of content under scrutiny is watermarked (resp. it has not been watermarked). There are two hypotheses: Under hypothesis $\mathcal{H}_0$, the detector receives an original content $\mathbf{r} = \mathbf{r}_0 = \mathbf{s}$ (see end of the introduction for justifications), whereas under hypothesis $\mathcal{H}_1$, the detector receives a watermarked and possibly attacked content $\mathbf{r} = \mathbf{r}_1$. Probability of false alarm P_{fa} and power of the test P_p are given by

$$P_{fa} = \Pr\{d = 1|\mathcal{H}_0\} \quad ; \quad P_p = \Pr\{d = 1|\mathcal{H}_1\}. \quad (2)$$

In zero-bit watermarking, no symbol is transmitted. Our problem is then fundamentally different from the communication of one bit because, under hypothesis $\mathcal{H}_0$, no processing is applied and $\mathbf{s}$, given by Nature, is directly sent to the detector.

We assume the detector has the structure of a Neyman-Pearson test. First, it applies a detection function $t(\mathbf{r})$ mapping from $\mathbb{R}^n$ to $\mathbb{R}$. Then, this scalar is compared to a threshold τ: $d = 1$ if $t(\mathbf{r}) > \tau$, 0 else. The threshold is given by the constraint of a significance level α such that $P_{fa} = \mathrm{E}\{(t(\mathbf{r}) > \tau)|\mathcal{H}_0\} \leq \alpha$. Moreover, we assume without loss of generality, that, under hypothesis $\mathcal{H}_0$, $t(\mathbf{r})$ is a centered random variable with unit variance:

$$\mathrm{E}\{t(\mathbf{r})|\mathcal{H}_0\} = 0, \qquad \mathrm{Var}\{t(\mathbf{r})|\mathcal{H}_0\} = 1 \quad (3)$$

If not the case, it is easy to built the test $\tilde{t}(\mathbf{r}) = (t(\mathbf{r}) - \mathrm{E}\{t(\mathbf{r})|\mathcal{H}_0\})/\sqrt{\mathrm{Var}\{t(\mathbf{r})|\mathcal{H}_0\}}$.

2.3 Pitman Noether efficacy

In this article, the tests are compared asymptotically for $n \to +\infty$. The Pitman-Noether theorem indicates that the best test has the higher efficacy η (or efficiency per sample), whose general definition is given, for instance, in [15, Sect. III.C.3]. In our case it simply reads:

$$\eta = n^{-1} \left(\frac{\partial}{\partial\theta} \mathrm{E}\{t(\mathbf{r})|\mathcal{H}_1\}\Big|_{\theta=0} \right)^2. \quad (4)$$

The proof of this theorem is based on an asymptotic study where the alternative hypothesis $\mathcal{H}_1$ has a vanishing parameter $\theta_n = kn^{-2}$, with k a positive constant and $t(\mathbf{r}) - \mathrm{E}\{t(\mathbf{r})\}$ is assumed, as $n \to \infty$, to be normal distributed with mean 0 and variance 1, both under $\mathcal{H}_1$ and under $\mathcal{H}_0$. These assumptions brings important restriction to our study.

The original theorem compares two detection functions, whereas we use it to compare two watermarking schemes composed each of two functions: embedding and detection. The methodology remains however the same.

3. DETECTION OF WEAK SIGNAL DEPENDENT ON SIDE INFORMATION

The goal of this section is to give the expressions for best detector and best embedding function. We mean 'best' in the sense of the Pitman Noether theorem, ie. such as they maximized the efficacy. This section doesn't consider any attack.

3.1 Best detector for a given embedding

In this subsection, embedding function $\mathbf{w}$ is fixed. A well known corollary of the Pitman Noether theorem [15, Sect. III.C.3] states that the Locally Most Powerful test in $\theta = 0$ is asymptotically the best. A Cauchy-Schwarz inequality gives:

$$\begin{aligned} \frac{\partial}{\partial\theta} \mathrm{E}\{t(\mathbf{r})|\mathcal{H}_1\}\Big|_{\theta=0} &= \int_{\mathbb{R}^n} t(\mathbf{r}) \left.\frac{\partial}{\partial\theta} p(\mathbf{r}|\mathcal{H}_1)\right|_{\theta=0} dv \\ &\leq \sqrt{\mathrm{Var}\{t(\mathbf{r})|\mathcal{H}_0\}} \sqrt{\mathrm{E}\{t_0^2(\mathbf{r})|\mathcal{H}_0\}} \\ &= \sqrt{\mathrm{E}\{t_0^2(\mathbf{r})|\mathcal{H}_0\}}, \end{aligned}$$

with equality for the LMP test:

$$t(\mathbf{r}) = k_t t_0(\mathbf{r}) = k_t \frac{1}{p(\mathbf{r}|\mathcal{H}_0)} \left.\frac{\partial p(\mathbf{r}|\mathcal{H}_1)}{\partial\theta}\right|_{\theta=0}, \quad (5)$$

where k_t a positive constant whose role is explained below. The use of the LMP with $\theta = 0$ is reinforced in practice by the fact the watermark power is very weak compared to the host power.

When there is no attack, $p(\mathbf{r}|\mathcal{H}_0) = p_{\mathbf{S}}(\mathbf{r})$ and $p(\mathbf{r}|\mathcal{H}_1) = p_{\mathbf{Y}}(\mathbf{r})$. We assume that $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \sigma_S^2\mathbf{I})$ and that there exists $\bar{\theta} > 0$, such that function $f(\mathbf{s})$ is invertible at least when $0 \leq \theta \leq \bar{\theta}$: $\mathbf{s} = f^{-1}(\mathbf{y})$. This allows to write $p_{\mathbf{Y}}(\mathbf{r}) = p_{\mathbf{S}}(f^{-1}(\mathbf{r}))|J_{f^{-1}}(\mathbf{r})|$, with the last term being the determinant of the Jacobian matrix of f^{-1} taken at $(\mathbf{r}, \theta)$. Developing this last equation, we finally get this expression:

$$t(\mathbf{r}) = k_t \sigma_S^{-2} \mathbf{r}^T \mathbf{w}(\mathbf{r}) - k_t \mathrm{div}(\mathbf{w}(\mathbf{r})). \quad (6)$$

The first term corresponds to the classical linear correlation based test (except that in our case the watermark signal is not fixed), whereas the second term is not null whenever side information is enabled at the embedding side.

One can show [7] that this detection function is centered under hypothesis $\mathcal{H}_0$, as required in subsection 2.2, provided $\mathrm{E}\{\|\mathbf{w}(\mathbf{s})\|^2|\mathcal{H}_0\} < +\infty$. The constant k_t enforces that $\mathrm{Var}\{t(\mathbf{r})|\mathcal{H}_0\} = 1$:

$$k_t = \left(\int_{\mathbb{R}^n} \frac{1}{p(\mathbf{r}|\mathcal{H}_0)} \left[\frac{\partial p(\mathbf{r}|\mathcal{H}_1)}{\partial \theta} \right]_{\theta=0}^2 dv \right)^{-1/2}. \quad (7)$$

However, the asymptotic gaussianity is not granted in general. It has to be checked for each particular case.

Finally, the efficiency per element for such tests reads:

$$\eta = n^{-1} k_t^{-2}. \quad (8)$$

3.2 Best embedding for a given detection

The detection function t being fixed ($t(\mathbf{s})$ is assumed to be centered, unit variance random variable under $\mathcal{H}_0$ and asymptotically gaussian distributed), we write:

$$\frac{\partial}{\partial \theta} \mathrm{E}\{t(\mathbf{r})|\mathcal{H}_1\}\Big|_{\theta=0} = \mathrm{E}\left\{ \frac{\partial}{\partial \theta} t(\mathbf{s} + \theta \mathbf{w}(\mathbf{s}))\Big|_{\theta=0} \right\} \quad (9)$$

$$= \mathrm{E}\{\mathbf{w}(\mathbf{s})^T \nabla t(\mathbf{s})\}. \quad (10)$$

It appears that, for a given t, it is important to let $\mathbf{w}(\mathbf{s}) \propto \nabla t(\mathbf{s})$, $\forall \mathbf{s} \in \mathbb{R}^n$. The efficacy is then upper bounded by the following Cauchy-Schwarz inequality:

$$\eta \leq n^{-1} \mathrm{E}\{\|\mathbf{w}(\mathbf{s})\|^2\} \mathrm{E}\{\|\nabla t(\mathbf{s})\|^2\} \quad (11)$$

with equality when:

$$\mathbf{w}(\mathbf{s}) = k_w \nabla t(\mathbf{s}) \quad \forall \mathbf{s} \in \mathbb{R}^n. \quad (12)$$

where k_w is a normalizing constant to achieve $\mathrm{E}\{\|\mathbf{w}(\mathbf{s})\|^2\} = n$:

$$k_w = \sqrt{n / \mathrm{E}\{\|\nabla t(\mathbf{s})\|^2\}}. \quad (13)$$

(11) and (13) give the efficacy for such tests:

$$\eta = n k_w^{-2} = \mathrm{E}\{\|\nabla t(\mathbf{s})\|^2\}. \quad (14)$$

3.3 Synthesis

For the moment, we know how to design the best embedding function for a given detection function, and how to design the best detection function for a given embedding function. Insert (12) in (6) yields a partial differential equation, that we loosely name 'fundamental equation of zero-bit watermarking':

$$(k_t k_w)^{-1} t(\mathbf{r}) - \sigma_S^{-2} \mathbf{r}^T \nabla t(\mathbf{r}) + \nabla^2 t(\mathbf{r}) = 0, \quad (15)$$

$\nabla^2 t(\mathbf{r})$ being the Laplacian of $t(\mathbf{r})$. Hence, the best couple of detection/embedding functions $\{t, \mathbf{w}\}$ is $\{t^\star, k_w \nabla t^\star\}$, with $t^\star$ a fundamental solution, ie. a solution of (15). Note that (8) and (14) are still valid. Therefore, it is possible to build a scheme of a given η (virtually, as high as possible), provided (15) admits a solution with $(k_w k_t)^{-1} = \eta$, which is also asymptotically gaussian distributed.

4. FUNDAMENTAL SOLUTIONS

We are not able to find a general solution of the fundamental equation, even with the restriction to the white gaussian case. The following gives some examples.

Table 1: Polynomial solutions of the scalar Gaussian case $s \sim \mathcal{N}(0,1)$.

η	$w(s)$	$t(r)$
1	1	r
2	s	$\frac{-1+r^2}{\sqrt{2}}$
3	$\frac{-1+s^2}{\sqrt{2}}$	$\frac{-3r+r^3}{\sqrt{6}}$
4	$\frac{-3s+s^3}{\sqrt{6}}$	$\frac{3-6r^2+r^4}{2\sqrt{6}}$
5	$\frac{3-6s^2+s^4}{2\sqrt{6}}$	$\frac{15r-10r^3+r^5}{2\sqrt{30}}$
6	$\frac{15s-10s^3+s^5}{2\sqrt{30}}$	$\frac{-15+45r^2-15r^4+r^6}{12\sqrt{5}}$
7	$\frac{-15+45s^2-15s^4+s^6}{12\sqrt{5}}$	$\frac{-105r+105r^3-21r^5+r^7}{12\sqrt{35}}$

4.1 The scalar case

The host samples are i.i.d. such that $p_{\mathbf{S}}(\mathbf{s}) = \prod_{i=1}^n p_S(s_i)$. Moreover, our strategy here is to maintain this statistical independence while embedding the watermark: $\mathbf{w}(\mathbf{s}) = (\epsilon_1 w(s_1), \ldots, \epsilon_n w(s_n))^T$, where $\boldsymbol{\epsilon}$ is a secret vector, with for instance, $\epsilon_i = \pm 1 \, \forall i \in \{1, \ldots, n\}$. (6) shows that the detection function has the following form $t(\mathbf{r}) = \sum_{i=1}^n \epsilon_i t(r_i)$ which is asymptotically gaussian thanks to the central limit theorem. (15) boils down to a scalar second-order ordinary differential equation with non constant coefficients:

$$\eta t(r) - \sigma_S^{-2} r t'(r) + t''(r) = 0. \quad (16)$$

The solution is a linear combination of two 'independent' (*ie.* their Wronskian is not null) confluent hypergeometric functions of the first kind taken in $r^2/2$:

$$t^{(a)}(r) = k_{t_1} \cdot {}_1F_1\left(-\frac{\sigma_S^2 \eta}{2}, \frac{1}{2}, \frac{r^2}{2\sigma_S^2} \right), \quad (17)$$

$$t^{(b)}(r) = k_{t_2} . r \cdot {}_1F_1\left(\frac{1 - \sigma_S^2 \eta}{2}, \frac{3}{2}, \frac{r^2}{2\sigma_S^2} \right). \quad (18)$$

If $\sigma_S^2 \eta$ is an even integer, $t^{(a)}$ is a polynomial function. If $\sigma_S^2 \eta$ is an odd integer, $t^{(b)}$ is a polynomial function. A much more elegant way to see this is to recognize this later differential equation as the Hermite equation when η is a positive integer and $\sigma_S^2 = 1$. Therefore, if $\eta \sigma_S^2 = k \in \mathbb{N}$, $t_k(r) = \kappa_k H_k(r/\sigma_S)$, H_k being the Hermite polynomial of order k. Another definition is given by the Rodrigues formula:

$$H_k(x) = (-1)^k e^{x^2/2} \frac{\partial^k}{\partial x^k} e^{-x^2/2}. \quad (19)$$

This family of polynomials is known to be orthogonal with a weighting function[1] $\exp(-r^2/2)$. This implies that elements of this polynomial family satisfies:

$$\mathrm{E}\{t_k(r) t_\ell(r)|\mathcal{H}_0\} = \delta(k - \ell), \quad (20)$$

δ being the Kronecker delta function.

Table 4.1 gives the expressions of the first elements of this family and their associated embedding function. The first line of this table is the well known direct spread spectrum scheme with a linear correlator, optimal detector in the

[1]This is the 'probabilist' definition of Hermite polynomials. However, these polynomials take different form according to the chosen standardization. For instance, $\kappa_k = 1/\sqrt{k!}$ when the coefficient of highest order of H_k is set to 1.

Gaussian i.i.d. case. The second line is known as the proportional or multiplicative embedding, first proposed in [3, Sect. 4.2] for perceptual reason. A higher efficacy is another inherent advantage of proportional embedding. The remaining lines of this table generalize this idea to new schemes (as far as the author knows).

4.2 The vector case

4.1 uses the cartesian system where the embedding processes in a sample wise manner. We generalize this idea to block based watermarking scheme assuming there exists an integer p dividing n so that $\mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^p \ldots \times \mathbb{R}^p$ and that $p_{\mathbf{S}}(\mathbf{s}) = \prod_{i=1}^{n/p} p(s_{(i-1)p+1}, \ldots, s_{(i-1)p+p})$. If $t^{(p)}$ is a solution of the fundamental equation in $\mathbb{R}^p$ with a given efficacy, then $t^{(n)}(\mathbf{r}) = \sqrt{p/n} \sum_{i=1}^{n/p} t^{(p)}(r_{(i-1)p+1}, \ldots, r_{(i-1)p+p})$ is a solution in $\mathbb{R}^n$ yielding the same efficacy. This realizes a statistically independent embedding in the sense that the block of p watermark samples only depends on the same block of p host samples. Parameter p must be fixed for all n to ensure the asymptotic gaussian distribution of the detection function. The issue is now on finding solutions $t^{(p)}$. A usual technique is the separation of variables method in a specific orthogonal coordinate system [13].

4.3 Separation of variables

In the cartesian coordinate system, the separation of variables method considers a solution $t^{(p)}(\mathbf{r}) = \prod_{i=1}^{p} t_{\eta_i}(r_i)$, where each t_{η_i} has to satisfy (16) with their own efficacy η_i. The resulting efficacy of $t^{(p)}$ is then $\eta = \sum_{i=1}^{p} \eta_i$. This gives birth to an extension of the polynomial family which is indeed based on the multivariate Hermite polynomials, indexed by the p-uple $\mathbf{k} \in \mathbb{N}^p$: $H_{\mathbf{k}}(\mathbf{r}) = \prod_{i=1}^{p} H_{k_i}(r_i)$. Two different elements of this family are orthogonal in the sense that $\mathrm{E}\{t_{\mathbf{k}}(\mathbf{r}) t_{\boldsymbol{\ell}}(\mathbf{r})\} = \delta(\mathbf{k}-\boldsymbol{\ell})$, even having the same efficacy: $|\mathbf{k}|_1 = |\boldsymbol{\ell}|_1$ (L^1-norm).

This extension of the polynomial family is illustrated in the following example. JANIS, a zero-bit watermarking scheme invented heuristically some years ago [8, 5], is a block based fundamental solution for white gaussian host. Its detection function is the following one:

$$t(\mathbf{r}) = \sqrt{\frac{p}{n}} \sum_{i=1}^{n/p} \prod_{j=1}^{p} \frac{r_{(i-1)p+j}}{\sigma_S}. \tag{21}$$

Note that r_k appears only once in the detection function, $\forall k \in \{1, \ldots, n\}$. It is easy to see that $\mathbf{r}^T \nabla t(\mathbf{r}) = p t(\mathbf{r})$ and $\nabla^2 t(\mathbf{r}) = 0$. Thus, JANIS with order p is a solution to (15) provided that $\eta \sigma_S^2 = p$. This can be interpreted as follows: this is a block based watermarking scheme built on the p multivariate Hermite polynomial $H_{(1,\ldots,1)}$. This theoretical framework proves the optimality of the heuristic JANIS scheme. Note that this result is also given by the generalized Rodrigues formula:

$$t(\mathbf{r}) = (-1)^{|\mathbf{k}|_1} \kappa_{\mathbf{k}} \frac{1}{p_{\mathbf{S}}(\mathbf{r})} \frac{\partial^{k_1} \ldots \partial^{k_p}}{\partial r_1^{k_1} \ldots \partial r_p^{k_p}} p_{\mathbf{S}}(\mathbf{r}) \tag{22}$$

Separation of variables can be done on another coordinate system. The spherical coordinate system $(\rho, \theta_1, \ldots, \theta_{p-1})$ is in general adapted to isotropic host distributions, ie. $p_{\mathbf{S}}(\mathbf{s}) = f(\rho)$ with $\rho = \|\mathbf{s}\|$. In our case, we have: $f(\rho) \propto e^{-\rho^2/\sigma_S^2}$. For instance, we seek a function $t(\mathbf{r}) = t(\rho, \theta_{p-1})$ whose expression in the spherical coordinate system is $U(\rho)V(\theta_{p-1})$. Separating variables in (15) yields two equations:

$$KV + (n-2)\cot\theta V' + V'' = 0,$$
$$(\eta\rho^2 - K)U + \left((n-1)\rho - \rho^3\sigma_S^2\right)U' + \rho^2 U'' = 0,$$

with $K \in \mathbb{R}$. The choice $U(\rho) = \rho^2$ and $V(\theta) = p\cos^2\theta - 1$ is a solution provided $K = 2p$ and $\eta\sigma_S^2 = 2$: $t(\rho, \theta_{p-1}) = k_t\rho^2(p\cos^2\theta_{p-1} - 1)$. This solution is interpreted as

$$t(\mathbf{r}) = k_t \left((\sqrt{p}\mathbf{r}^T\mathbf{e}_p)^2 - \|\mathbf{r}\|^2 \right), \tag{23}$$

i.e. the measure of robustness given in Cox *et al.* book [4, Eq.(5.13)]: $t(\mathbf{r}) = cst$ defines a p-dimensional two-sheet hyperboloid. This acceptance region is closed to a one-sheet hypercone, acceptance region of the well-known normalized correlation [4], or a double-sheet hypercone, acceptance region of the squared normalized correlation [12]. Yet, neither the famous normalized correlation $t(\mathbf{r}) = \mathbf{r}^T.\mathbf{e}_p/\|\mathbf{r}\| = \cos\theta_{p-1}$, nor its squared version are fundamental solutions for gaussian hosts. As a concluding remark, although the spherical coordinate system seems to provide solutions quite different than the polynomial family, it appears that the later solution can be rewritten as a mixture of second order Hermite polynomials: $t(\mathbf{r}) = k_t\sqrt{2}(pH_2(r_p) - \sum_{i=1}^{p} H_2(r_i))$.

4.4 Sparsity

Many possible coordinate systems allow a separation of variables, but their investigation is out of scope in this paper. However, we would like here to rediscover a famous principle in watermarking. Suppose we know a solution $t^\star$ to the scalar equation : $\eta^\star t^\star(x) + L(x)t^{\star\prime}(x) + t^{\star\prime\prime}(x) = 0$. We would like to extend this scalar solution to higher space dimension considering a solution in the form: $t = t^\star \circ g$, with $g : \mathbb{R}^n \to \mathbb{R}$ a derivable function. Gradient and laplacian have the following expressions:

$$\nabla t(\mathbf{r}) = t^{\star\prime}(g(\mathbf{r}))\nabla g(\mathbf{r}), \tag{24}$$
$$\nabla^2 t(\mathbf{r}) = t^{\star\prime\prime}(g(\mathbf{r}))\|\nabla g(\mathbf{r})\|^2 + t^{\star\prime}(g(\mathbf{r}))\nabla^2 g(\mathbf{r}), \tag{25}$$

and the fundamental equation becomes:

$$\eta t^{\star\prime}(g(\mathbf{r}))\left(-\frac{\eta}{\eta^\star}L(g(\mathbf{r})) + \sigma_S^{-2}\mathbf{r}^T\nabla g(\mathbf{r}) + \nabla^2 g(\mathbf{r})\right)$$
$$+ \quad t^{\star\prime\prime}(g(\mathbf{r}))\left(\|\nabla g(\mathbf{r})\|^2 - \frac{\eta}{\eta^\star}\right) = 0 \tag{26}$$

We restrict our analysis to a linear form, ie. a projection $g(\mathbf{r}) = \mathbf{r}^T\boldsymbol{\lambda}$. Then, t is a fundamental solution with efficacy $\eta = \eta^\star\|\boldsymbol{\lambda}\|^2$, provided we have:

$$L(\mathbf{r}^T\boldsymbol{\lambda}) = \frac{1}{\|\boldsymbol{\lambda}\|^2\sigma_S^2}\mathbf{r}^T\boldsymbol{\lambda}. \tag{27}$$

$L(x)$ is thus the score (ie. the ratio $p'(x)/p(x)$) associated to $\mathcal{N}(0, \|\boldsymbol{\lambda}\|^2\sigma_S^2)$, the distribution of $\mathbf{r}^T\boldsymbol{\lambda}$. Consequently, we discover here a possible extension of the polynomial family to the vector case with fundamental solutions of the form

$$t_k(\mathbf{r}) = \kappa_k H_k(\mathbf{r}^T\boldsymbol{\lambda}/\|\boldsymbol{\lambda}\|\sigma_S), \tag{28}$$

whose efficacy is $\eta = k/\sigma_S^2$. However, asymptotic gaussianity is not granted and a similar strategy as the block watermarking above-mentioned should be use to regularise this issue.

This kind of solutions illustrates the principle known as sparsity or time sharing [14, Sect. 5.2 and 8.2], where the watermark embedding is processed on the projection $\mathbf{r}^T\boldsymbol{\lambda}$.

5. ATTACK NOISE

When there is an attack, the received signal under $\mathcal{H}_1$ is $\mathbf{r}_1 = \mathbf{a}(\mathbf{y})$. The attack channel $\mathbf{a}$ is defined through a conditional probability distribution $p_a(\mathbf{r}_1|\mathbf{y})$, whose associated attack power is $\sigma_a^2 = \int\int \|\mathbf{r}_1 - \mathbf{y}\|^2 p_a(\mathbf{r}_1|\mathbf{y}) p_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} d\mathbf{r}_1/n$. The parameters of the attack channel are unknown at the detection side. We would like to keep the detection as simple as possible so that the estimation of these parameters is not tractable in this strategy. The performance of the detector should slowly degrade with the strength of the attack for a robust watermarking scheme.

The Pitman Noether might then become useless because there is a disruption between the two hypothesis: $\mathcal{H}_1$ doesn't asymptotically converge to $\mathcal{H}_0$, due to the presence of the attack channel only under $\mathcal{H}_1$.

We present here two ways to tackle this problem, changing our framework in order to enforce the Pitman Noether theorem. A first idea is to restrict our analysis to a fixed WNR (watermark to noise power ratio): $\theta_n^2/\sigma_a^2 = g$. The received signal can be written as: $\mathbf{r}_1 = \mathbf{s} + \theta_n \mathbf{w}(\mathbf{s}) + \theta_n \sqrt{g}\tilde{\mathbf{z}}$, with $\mathrm{E}\{\|\tilde{\mathbf{z}}\|^2\} = n$. Therefore, the power of the difference signal $\mathbf{r}_1 - \mathbf{r}_0$ asymptotically vanishes with θ_n^2. The second idea considers attacks with fixed DNR (document -ie. host- to noise power ratio) where signals are corrupted by the same attack under both hypothesis as T. Liu and P. Moulin did [10]. Yet, the targeted applications as described in our introduction do not a priori motivate this possibility because the attack of unprotected content are clearly unlikely. We argue that a 'soft' attack lead on original pieces of content yields regular content, in the sense that it still statistically looks like a content. In other words, we restrict our study to attack channels changing the value of the feature vectors, but not modifying their inherent statistical structure.

Under both attack models, the fundamental equation appears to be statistically robust in the sense that it is not modified by the presence of noise. However, this is only true for very particular conditions as described in the sequel.

5.1 Fixed WNR attacks

This subsection only shows that the fundamental equation remains unchanged when watermarked signals go through a fixed WNR AWGN attack channel.

5.1.1 Best embedding for a given detection

As usual, we write:

$$\begin{aligned}\left.\frac{\partial}{\partial\theta}\mathrm{E}\{t(\mathbf{r})|\mathcal{H}_1\}\right|_{\theta=0} &= \left.\frac{\partial}{\partial\theta}\mathrm{E}_{\mathbf{S}}\mathrm{E}_{\tilde{\mathbf{Z}}} t(\mathbf{s}+\theta\mathbf{w}(\mathbf{s})+\theta\sqrt{g}\tilde{\mathbf{z}})\right|_{\theta=0} \\ &= \mathrm{E}_{\mathbf{S}}\{\mathbf{w}(\mathbf{s})^T \nabla t(\mathbf{s}) + \mathrm{E}_{\tilde{\mathbf{Z}}}\sqrt{g}\tilde{\mathbf{z}}^T \nabla t(\mathbf{s})\}\end{aligned}$$

We assume $\tilde{\mathbf{z}}$ is independent of $\mathbf{s}$ and centered, so that the second term is null. We find back the same best embedder as (12).

5.1.2 Best detection for a given embedding

The pdf of $\mathbf{r}_1 = \mathbf{y} + \sqrt{g}\theta\tilde{\mathbf{z}}$ is given by the following convolution:

$$p_{\mathbf{R}_1}(\mathbf{r}) = \int p_{\mathbf{Y}}(\mathbf{u}) p_{\sqrt{g}\theta\tilde{\mathbf{Z}}}(\mathbf{r}-\mathbf{u}) d\mathbf{u}, \tag{29}$$

whose derivative is composed of two terms:

$$\begin{aligned}\left.\frac{\partial}{\partial\theta}p_{\mathbf{R}_1}(\mathbf{r})\right|_{\theta=0} &= \int \left.\frac{\partial}{\partial\theta}p_{\mathbf{Y}}(\mathbf{u})\right|_{\theta=0} \lim_{\theta\to 0} p_{\sqrt{g}\theta\tilde{\mathbf{Z}}}(\mathbf{r}-\mathbf{u}) d\mathbf{u} \\ &+ \int p_{\mathbf{S}}(\mathbf{u}) \left.\frac{\partial}{\partial\theta}p_{\sqrt{g}\theta\tilde{\mathbf{Z}}}(\mathbf{r}-\mathbf{u})\right|_{\theta=0} d\mathbf{u} \quad (30)\end{aligned}$$

We assume that $\tilde{\mathbf{z}}$ is normal distributed, so that its limit $\lim_{\theta\to 0} p_{\sqrt{g}\theta\tilde{\mathbf{Z}}}(\mathbf{r}-\mathbf{u})$ is the Dirac distribution. Hence, the first term is, as detailed in Sect. 3.1, the derivative of the pdf without attack.

The second term is calculated being inspired by some proofs of the De Bruijn's identity (see [2, Th. 16.6.2]). It corresponds to the derivative of the pdf of $\mathbf{a}(\mathbf{s}) = \mathbf{s} + \sqrt{g}\theta\tilde{z}$ with respect to θ. In one hand, we have:

$$\frac{\partial}{\partial\theta}p_{\mathbf{a}(\mathbf{S})}(\mathbf{r}) = \int p_{\mathbf{S}}(\mathbf{u}) \left(\frac{\|\mathbf{r}-\mathbf{u}\|^2}{g\theta^3} - \frac{n}{\theta}\right) p_{\sqrt{g}\theta\tilde{\mathbf{Z}}}(\mathbf{r}-\mathbf{u}) d\mathbf{u} \tag{31}$$

On the other hand, it appears that:

$$\begin{aligned}\nabla^2 p_{\mathbf{a}(\mathbf{S})}(\mathbf{r}) &= \int p_{\mathbf{S}}(\mathbf{u}) \left(\frac{\|\mathbf{r}-\mathbf{u}\|^2}{g^2\theta^4} - \frac{n}{g\theta^2}\right) \\ &\qquad p_{\sqrt{g}\theta\tilde{\mathbf{Z}}}(\mathbf{r}-\mathbf{u}) d\mathbf{u} \\ &= \frac{1}{g\theta}\frac{\partial}{\partial\theta}p_{\mathbf{a}(\mathbf{S})}(\mathbf{r}).\end{aligned}$$

Finally, the second term is null because:

$$\left.\frac{\partial}{\partial\theta}p_{\mathbf{a}(\mathbf{S})}(\mathbf{r})\right|_{\theta=0} = \lim_{\theta\to 0} g\theta\nabla^2 p_{\mathbf{a}(\mathbf{S})}(\mathbf{r}) = 0, \tag{32}$$

and we find back the same best detection function as (6). To conclude, neither the best embedding nor the best detection have changed under the fixed WNR AWGN attack channel. Hence, the fundamental equation still holds in this framework.

5.2 Fixed DNR attacks

We consider a different framework where the hypothesis are now: $\mathcal{H}_0 : \mathbf{r}_0 = \mathbf{a}(\mathbf{s})$ against $\mathcal{H}_1 : \mathbf{r}_1 = \mathbf{a}(\mathbf{s} + \theta\mathbf{w}(\mathbf{s}))$. What is the impact on the detection and embedding functions?

As already said, our analysis only holds for channel attacks conserving the statistical structure of the host signal. The restrictions are as follows. For gaussian host $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, the attack is the SAWGN channel: $\mathbf{a}(\mathbf{s}) = \gamma(\mathbf{s}+\mathbf{z})$, with $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma_Z^2\mathbf{I}_n)$ independent of $\mathbf{s}$ and $\gamma = 1/\sqrt{1+\sigma_Z^2}$. The attack is a Wiener filtering for this very simple case, which maintains $p(\mathbf{r}|\mathcal{H}_0)$ as a normal distribution. The expression (6) of the best detection function given the embedding function is thus not modified.

This is not the case for the best embedding function given the detection function. For the class of attack channel considered in this paper, $\mathbf{a}(\mathbf{s}) = \gamma(\mathbf{s}+\mathbf{z})$, (9) is modified as follows:

$$\begin{aligned}\left.\frac{\partial}{\partial\theta}\mathrm{E}\{t(\mathbf{r})|\mathcal{H}_1\}\right|_{\theta=0} &= \mathrm{E}_{\mathbf{S}}\mathrm{E}_{\mathbf{Z}} \left.\frac{\partial}{\partial\theta} t(\gamma(\mathbf{s}+\theta\mathbf{w}(\mathbf{s})+\mathbf{z}))\right|_{\theta=0} \\ &= \gamma\mathrm{E}_{\mathbf{S}}\{\mathbf{w}(\mathbf{s})^T \mathrm{E}_{\mathbf{Z}}\{\nabla t(\gamma(\mathbf{s}+\mathbf{z}))\}\} \quad (33)\end{aligned}$$

This last equation shows that the best strategy at the embedding side should set

$$\mathbf{w}(\mathbf{s}) \propto \mathrm{E}_Z\{\nabla t(\gamma(\mathbf{s}+\mathbf{z}))\}. \tag{34}$$

This implies that the embedder knows the attack channel parameters. This counter attack may not be realistic in general. However, there are some cases where the counter attack (34) is surprisingly simple because it is indeed identical to the regular embedding strategy (12) whatever the parameters of the attack channel. This occurs when t is such that $\mathrm{E}_Z\{\nabla t(\gamma(\mathbf{s}+\mathbf{z}))\} = h(\gamma,\sigma_Z)\nabla t(\mathbf{s})$. As a consequence, the fundamental equation (15) derived in the no attack case, remains valid under these particular attack cases. The efficiency per element is then equal to $\eta(\gamma,\sigma_Z) = \gamma^2 h^2(\gamma,\sigma_Z)\eta(1,0)$.

For the Hermite polynomial family, we rewrite the Wiener filtering denoting $\tilde{z} = \sigma_Z^{-1} z$ distributed as $\mathcal{N}(0,1)$ and $\alpha = \arccos(\gamma)$. A not so familiar identity of the Hermite polynomials is the following:

$$H_{\ell-1}\quad(\cos(\alpha)s + \sin(\alpha)\tilde{z}) =$$
$$\textstyle\sum_{k=0}^{\ell-1}\quad \binom{\ell-1}{k}\cos^k(\alpha)\sin^{\ell-1-k}(\alpha)H_k(s)H_{\ell-1-k}(\tilde{z}).$$

$\mathrm{E}_Z\{t'_\ell(\gamma(s+z))\}$ reduces to $\kappa_\ell \ell \gamma^{\ell-1} H_{\ell-1}(s) = \gamma^{\ell-1} t'_\ell(s)$ because $\mathrm{E}_{\tilde{Z}}\{H_k(\tilde{z})\} = \delta(k)$. Consequently, the polynomial family is a set of fundamental solutions for i.i.d. gaussian hosts and SAWGN attacks with Wiener filtering, whose efficiency per element is given by $\eta(\gamma,\sigma_Z) = \ell\gamma^{2\ell}$. Wiener filtering means that $\gamma = (1+\sigma_Z^2)^{-1/2}$. Two noticeable exemptions are t_1 and t_2, whose efficiency follows the same rule whatever the value of γ of the SAWGN channel. Last but not least: the higher the 'original' efficiency $\eta(1,0) = \ell$, the less robust is the scheme in the sense that $\eta(\gamma,\sigma_Z)/\eta(1,0) = (1+\sigma_Z^2)^{-\eta(1,0)}$ decreases faster with the strength of the attack. We have found this feature in other watermarking schemes [7].

The same analysis also holds for the extension of the polynomial family to the vector case. For instance, JANIS is a solution of the fundamental equation for i.i.d. gaussian hosts and SAWGN attack, such that $\mathrm{E}_Z\{\nabla t(\gamma(\mathbf{s}+\mathbf{z}))\} = \gamma^{p-1}\nabla t(\mathbf{s})$. The Wiener filtering restriction is not necessary as JANIS is based on first order Hermite polynomials. This gives the following efficiency $\eta(\gamma,\sigma_Z) = p\gamma^{2p}$ which follows the same decreasing rule as the polynomial family.

6. CONCLUSION

Zero-bit watermarking pertains to test hypothesis. The efficacy is a core notion in this field because of the Pitman Noether theorem asymptotically comparing test performances. Working on the expression of the efficacy when side information is enabled at the embedding side gives a fundamental equation. This equation still holds when an attack channel is considered with some very restrictive assumptions. When focusing on white gaussian hosts, Hermite polynomials form a class of solutions of the fundamental solution. JANIS indeed appears to be a mixture of multivariate low order Hermite polynomials. This shows that JANIS is provably good in the sense that it is a solution of the fundamental equation under SAWGN attacks.

Our future work is to see whether the generalized Rodrigues formula (22) can work with different probability distribution functions such as gaussian with covariant matrix $\mathbf{R}_x \neq \sigma_X^2 \mathbf{I}_n$ or exponential family.

7. REFERENCES

[1] J. Andreaux, A. Durand, T. Furon, and E. Diehl. Copy protection system for digital home networks. *IEEE Signal Processing Magazine*, 21(2):100–108, March 2004. Special Issue on Digital Right Management.

[2] T. Cover and J. Thomas. *Elements of information theory*. Number ISBN-0-471-06259-6 in Wiley series in telecommunications. Wiley, 1991.

[3] I. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1687, December 1997.

[4] I. Cox, M. Miller, and J. Bloom. *Digital Watermarking*. Morgan Kaufmann Publisher, 2001.

[5] J. Delhumeau, T. Furon, N. Hurley, and G. Silvestre. Improved polynomial detectors for side-informed watermarking. In *Security and Watermarking of Multimedia Contents IV*, pages 311–321, Santa Clara, Cal., USA, January 2003. SPIE Electronic Imaging.

[6] E. Diehl and T. Furon. Closing the analog hole. In *Proc. IEEE Int. Conf. Consumer Electronics*, pages 52–53, 2003.

[7] T. Furon. A constructive and unifying framework for zero-bit watermarking. *to be submitted to IEEE Trans. Information Forensics and Security*, 2006.

[8] T. Furon, G. Silvestre, and N. Hurley. JANIS: Just Another N-order side-Informed Scheme. In *Proc. of Int. Conf. on Image Processing ICIP'02*, volume 2, pages 153–156, Rochester, NY, USA, September 2002.

[9] E. Lin, A. Eskicioğlu, R. Lagendijk, and E. Delp. Advances in digital video content protection. *Proc. of IEEE*, 93(1):171–183, jan 2005.

[10] T. Liu and P. Moulin. Error exponents for one-bit watermarking. In *Proc. of ICASSP*, Hong-Kong, apr 2003.

[11] N. Merhav. An information-theoretic view of watermarking embedding-detection and geometric attacks. presented at WaCha05, available at `www.ee.technion.ac.il/people/merhav/`, jun 2005.

[12] N. Merhav and E. Sabbag. Optimal watermarking embedding and detection strategies under limited detection resources. submitted to IEEE Trans. on Inf. Theory, 2006.

[13] P. Moon and D. E. Spencer. Theorems on separability in riemannian n-space. *Proc. Amer. Math. Soc.*, 3:635–642, 1952.

[14] P. Moulin and R. Koetter. Data hiding codes. *Proceedings of the IEEE*, dec 2005.

[15] H. V. Poor. *An introduction to signal detection and estimation*, volume 2nd edition. Springer, 1994.

[16] Wikipedia. Analog hole. *Wikipedia, The Free Encyclopedia*, `http://en.wikipedia.org/w/index.php?title=Analog_hole&oldid=38835021`, 2006.