# An Information–Theoretic View of Watermark –Detection and Geometric Attacks

Neri Merhav

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 32000, Israel

WaCha ‘05, Barcelona, Spain, June 2005

# Outline

- An IT approach to WM embedding and detection (no attacks).
- Extending the framework to geometric attacks (GA's).
- Discussing a few variations and extensions (in time permits).

Bottom line: The ES approach is as good as if there was no GA.

# Background

The most popular approach — additive embedding:

Given a host vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ and a WM, $\boldsymbol{w} = (w_1, \ldots, w_n)$ ($w_i \in \{\pm 1\}$), the stegovector $\boldsymbol{y} = (y_1, \ldots, y_n)$ is created by

$$\boldsymbol{y} = \boldsymbol{x} + \alpha \boldsymbol{w},$$

where the choice of $\alpha$ trades off quality (distortion) and detectability.

Given this embedding function, $\boldsymbol{x}$ is like "noise", and in the Gaussian case, classical detection theory tells that the best detector is based on correlation.

# Background (Cont'd)

In classical detection theory, the additive structure (of the channel) is given, but here we have the freedom to select any function

$$\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{w}).$$

For an aribtrary $f$, the correlation detector is no longer necessarily optimal.

How should we design $f$ together with the detector?

# The Approach

Joint optimization of the embedder and detector is not easy.

Instead, let us confine ourselves to a certain class of detectors – the class of all detectors that base their decisions on a given set of statistics extracted from $\boldsymbol{y}$ and $\boldsymbol{w}$.

**Examples:**

- The joint empirical distribution $\hat{P}(w, y) = n(w, y)/n$.
- Correlation $\frac{1}{n}\sum_{i=1}^{n} w_i y_i$, energy $\frac{1}{n}\sum_{i=1}^{n} y_i^2$.
- Similar joint statistics of $\boldsymbol{y}$ and shifted versions of $\boldsymbol{w}$.

# Problem Definition

Let $\boldsymbol{w}$ be given, and $\boldsymbol{x} \sim P$ – a finite alphabet memoryless source.

A decision rule partitions $\mathcal{Y}^n$ two complementary regions $\Lambda$ and $\Lambda^c$:

$\boldsymbol{y} \in \Lambda \rightarrow$ decide for $H_1: \ \boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{w})$.

$\boldsymbol{y} \in \Lambda^c \rightarrow$ decide for $H_0: \ \boldsymbol{y} = \boldsymbol{x}$.

Neyman–Pearson criterion:

Minimize $P_{e_1} =$ Pr$\{f(\boldsymbol{x}, \boldsymbol{w}) \in \Lambda^c\}$ (FN)

s.t. $P_{e_2} =$ Pr$\{\boldsymbol{x} \in \Lambda\} \leq e^{-\lambda n}$ (FP constraint)

and $d(\boldsymbol{x}, \boldsymbol{y}) \leq nD$.

# Asymptotically Optimal Detector and Embedder

Among all detectors that are based on $\hat{P}(w, y)$, the following one is asymptotically optimum in the error–exponent sense:

$$\Lambda_* = \{\boldsymbol{y} : \; \ln P(\boldsymbol{y}) + n\hat{H}(Y|W) + \lambda n \leq 0\}$$

regardless of the choice of $f$!

The optimum $f$ is now the following:

$$f^*(\boldsymbol{x}, \boldsymbol{w}) = \text{argmin}_{\boldsymbol{y}:\; d(\boldsymbol{x},\boldsymbol{y}) \leq nD}[\ln P(\boldsymbol{y}) + n\hat{H}(Y|W)].$$

# Computational Complexity of $f^*$

Although it involves minimization over a sphere $\{\boldsymbol{y}:\ d(\boldsymbol{x},\boldsymbol{y}) \leq nD\}$, it does not really take an exponentially large exhaustive search.

Note that for an additive distortion measure

$$d(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{n} d(x_i, y_i),$$

both $d(\boldsymbol{x},\boldsymbol{y})$ and $[\ln P(\boldsymbol{y}) + n\hat{H}(Y|W)]$ depend on $(\boldsymbol{x},\boldsymbol{w})$ only via the joint empirical distribution of $(\boldsymbol{x},\boldsymbol{w},\boldsymbol{y})$, and the search over empirical distributions is polynomial in $n$.

# Universality in the Covertext Statistics

If $P$ is unknown (but still known to be memoryless), it makes sense to require that $P\{\boldsymbol{x} \in \Lambda\} \leq e^{-\lambda n}$ would hold for every memoryless $P$.

The resulting version of $\Lambda_*$ is a mutual information detector, i.e., it accepts $H_1$ iff

$$\hat{I}(W;Y) \geq \lambda,$$

and the corresponding embedder would be

$$f^*(\boldsymbol{x}, \boldsymbol{w}) = \text{argmax}_{\boldsymbol{y}:\ d(\boldsymbol{x},\boldsymbol{y}) \leq nD} \hat{I}(W;Y).$$

# Continuous Alphabets

One of the most customary models is $\boldsymbol{x} \sim \mathcal{N}(0, \sigma^2 I)$, where $\sigma^2$ is unknown.

Suppose that our class of detectors depend on the correlation $\sum_{i=1}^n w_i y_i$ and the energy $\sum_{i=1}^n y_i^2$, then the corresponding mutual information detector compares

$$\hat{\rho}^2 = \frac{(\frac{1}{n}\sum_{i=1}^n w_i y_i)^2}{\frac{1}{n}\sum_{i=1}^n y_i^2}$$

to a threshold.

The embedder, in the case of quadratic distortion $d(x,y) = (x-y)^2$, maximizes $\hat{\rho}^2$ s.t. $\sum_{i=1}^n (x_i - y_i)^2 \leq nD$.

# Continuous Alphabets (Cont'd)

Consider the optimization problem

$$\max \frac{\left(\sum_{i=1}^{n} w_i y_i\right)^2}{\sum_{i=1}^{n} y_i^2} \text{ s.t. } \sum_{i=1}^{n} (x_i - y_i)^2 \le nD.$$

Every solution $\boldsymbol{y}$ can be represented as

$$\boldsymbol{y} = a\boldsymbol{x} + b\boldsymbol{w} + \boldsymbol{z}$$

where $\boldsymbol{z}$ is orthogonal to $\boldsymbol{x}$ and $\boldsymbol{w}$.

Note, however, that WLOO, $\boldsymbol{z} = 0$ as every non–zero $\boldsymbol{z}$ increases the denominator without increasing the numerator and without improving the distance to $\boldsymbol{x}$. Thus,

$$\boldsymbol{y} = a\boldsymbol{x} + b\boldsymbol{w}$$

# Continuous Alphabets (Cont'd)

It remains to optimize only over two parameters, $a$ and $b$. Some simple manipulations reduce this further to a one–dimensional line search.

Note that

$$\boldsymbol{y} = a^*\boldsymbol{x} + b^*\boldsymbol{w}$$

is not a linear embedder, as $a^*$ and $b^*$ depend on $\boldsymbol{x}$ and $\boldsymbol{w}$.

# Attacks

In the case of attack, the detector sees a “forgery” $\boldsymbol{z}$, instead of $\boldsymbol{y}$, where $\boldsymbol{z}$ is created from $\boldsymbol{y}$ via an attack channel $W(\boldsymbol{z}|\boldsymbol{y})$.

In the case of a memoryless channel, the optimal detector is the same as before except that $\boldsymbol{y}$ is replaced by $\boldsymbol{z}$.

The optimal embedder would then be

$$f^*(\boldsymbol{x}, \boldsymbol{w}) = \text{argmin}_{\boldsymbol{y}:\ d(\boldsymbol{x},\boldsymbol{y}) \leq nD} \sum_{\boldsymbol{z} \in \Lambda_*^c} W(\boldsymbol{z}|\boldsymbol{y}).$$

# Geometric Attacks

A geometric attack can be thought of as an (unknown) transformation of the coordinates of the signal/image (e.g., a cyclic shift).

We will model it as a randomly chosen permutation: Given a set of $M$ permutations $\{\pi_1, \ldots, \pi_M\}$, let:

$$W(\boldsymbol{z}|\boldsymbol{y}) = \frac{1}{M}\sum_{i=1}^{M} 1\{\boldsymbol{z} = \pi_i(\boldsymbol{y})\}.$$

# Geometric Attacks (cont'd)

It can be shown that if $M$ is sub–exponential in $n$, an exhaustive search applied to $\Lambda_*$:

$$\Lambda_{ES} = \{\boldsymbol{z} : \ \ln P(\boldsymbol{z}) + n \cdot \min_i \hat{H}(Y|W^i) + \lambda n \leq 0\}$$

is not only optimum in the error exponent sense, but it also as good as if there was no attack.

An asymptotically optimum embedder:

$$f^*(\boldsymbol{x}, \boldsymbol{w}) = \text{argmin}_{\boldsymbol{y}:\ d(\boldsymbol{x},\boldsymbol{y}) \leq nD} [\ln P(\boldsymbol{y}) + n \cdot \max_j \min_i \hat{H}(Y^j|W^i)].$$

This seems to require a full search over the sphere because of the maxmin.

# Geometric Attacks (cont'd)

If, however, the set of permutations $\{\pi_1, \ldots, \pi_M\}$ forms a group, with the operations:

$$\pi_i(\pi_j(\cdot)) = \pi_{i\star j}(\cdot), \quad \pi_i^{-1}(\cdot) = \pi_{i^{-1}}(\cdot),$$

then the above can be simplified to

$$f^*(\boldsymbol{x}, \boldsymbol{w}) = \text{argmin}_{\boldsymbol{y}:\ d(\boldsymbol{x},\boldsymbol{y}) \le nD}[\ln P(\boldsymbol{y}) + n \cdot \min_i \hat{H}(Y|W^i)],$$

where the order of the minimizations can be interchanged, and so, the computational complexity is proportional to $M$, which is sub–exponential.

# Closing Remarks

- We have developed on IT framework for WM embedding and detection.
- The above leads to the principle of maximum mutual information.
- Good embedders are not linear in general.
- The ES approach is asymptotically optimum.
- Easy to extend to the case of private WM – better exponents.
- For multi–bit WM, the capacity is $C = \max\{H(Y|X) : \ Ed(X,Y) \leq D\}$, in both private and public settings, with and without GA's.