A Detection-Theoretic and Computational Framework for Designing Geometrically Resilient Watermarking Systems

Pierre Moulin
University of Illinois at Urbana-Champaign

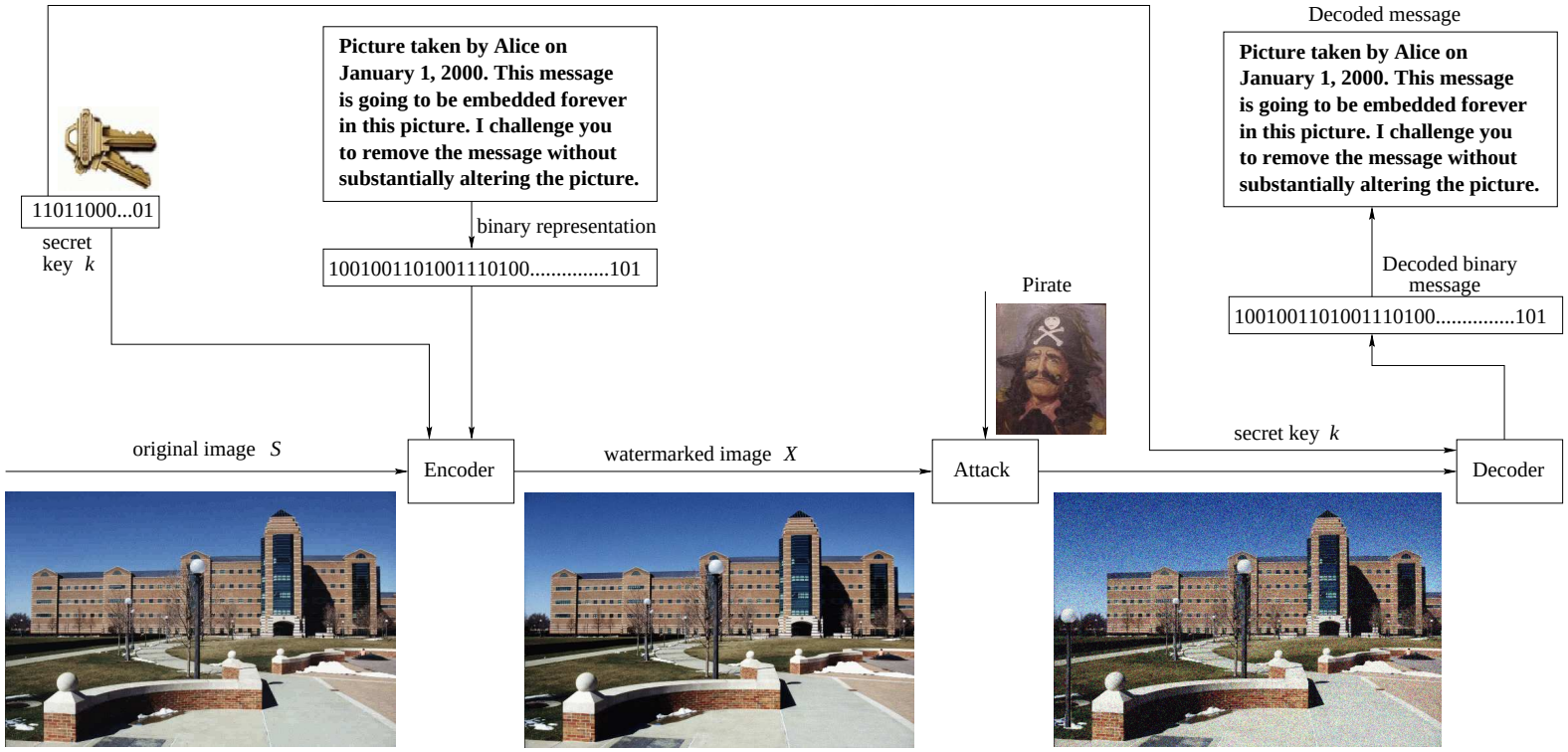
www.ifp.uiuc.edu/~moulin/talks/wacha05-slides.pdf

WaCha, Barcelona
June 8, 2005

Outline

- A communication model for geometric attacks
  - Role of Information Theory and Detection Theory
  - "Complexity" of geometric attacks
- Example: Unitary Geometric Attack Channels
- Invariant *vs* GLRT *vs* Pilot-based WM schemes

An Image Watermarking System



Attacks on Images

Original

JPEG, QF=10

4×4 median filtering

Gaussian filter ($\sigma = 3$)

Rotated by 10 degrees

Random bending

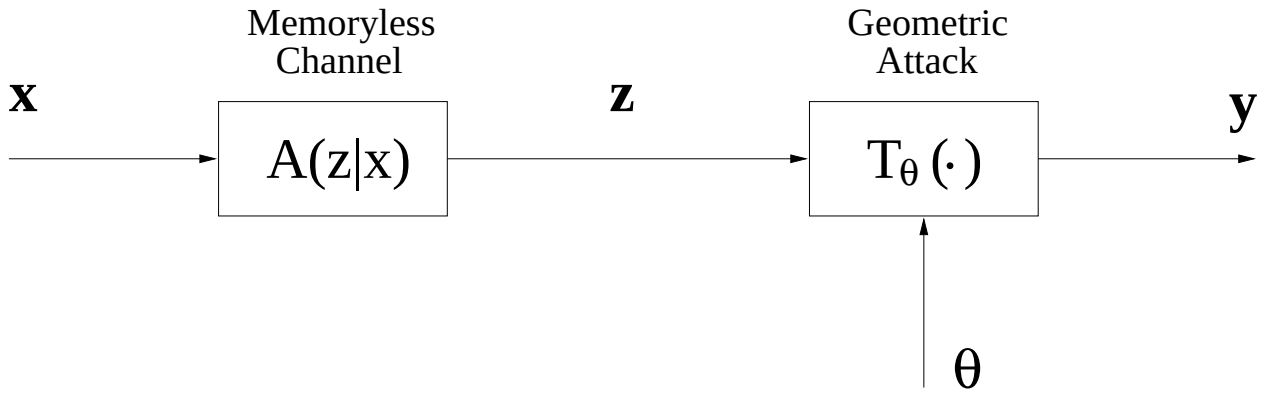
A Communication Model for Geometric Attacks

- Attacker maps watermarked $\mathbf{X} = (X_1, \cdots, X_n)$ into degraded $\mathbf{Y} = (Y_1, \cdots, Y_n)$ using stochastic mapping $p(\mathbf{y}|\mathbf{x})$.
- Distortion function $d(\mathbf{x}, \mathbf{y})$
- Feasible mappings satisfy a distortion constraint

  in average: $\mathbb{E}[d(\mathbf{X}, \mathbf{Y})] \leq D_2$

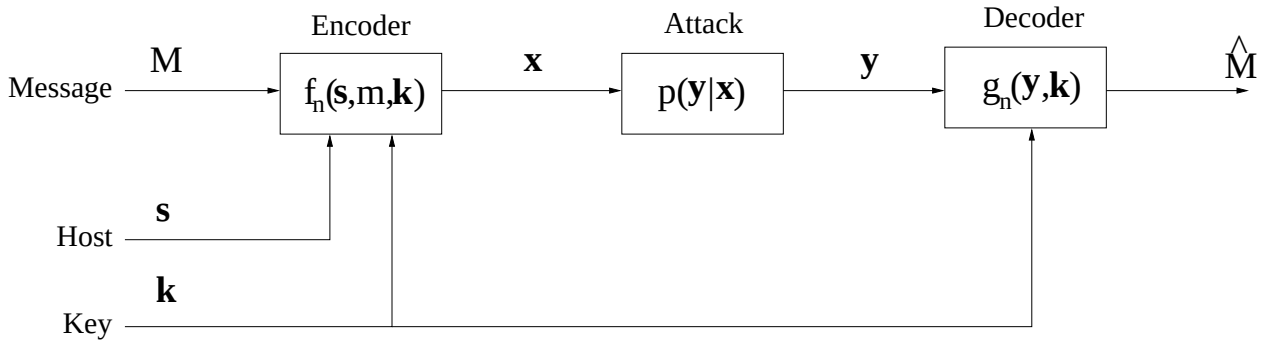
  or with probability one: $d(\mathbf{X}, \mathbf{Y}) \leq D_2$
- Would like "geometrically-inspired" $d(\mathbf{x}, \mathbf{y})$

Attack Model and Distortion Function [MM'02]

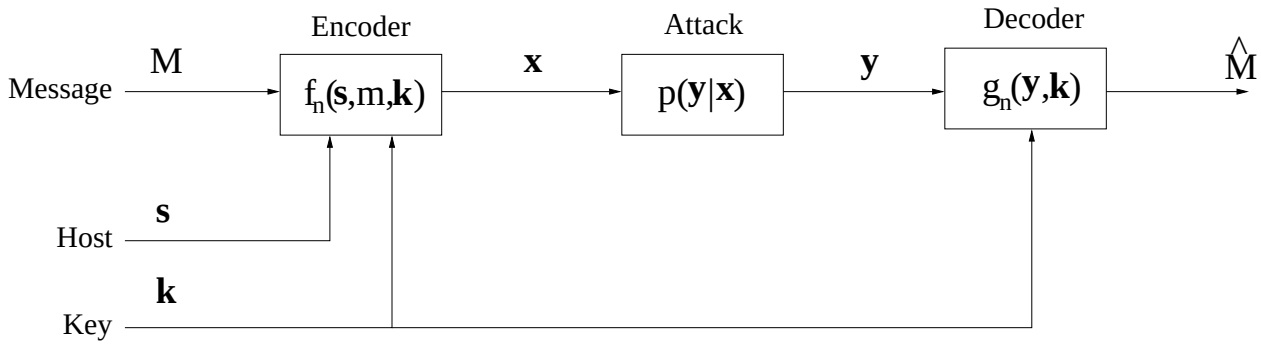


- Geometric (desynchronization) parameter $\theta \in \Theta$
- $T_\theta(\cdot)$ smooth, invertible mapping
- Additive distortion function $d_a(\mathbf{x}, \mathbf{z}) = \frac{1}{n} \sum_{i=1}^{n} d_a(x_i, z_i)$
- Distortion function $d(\mathbf{x}, \mathbf{y}) = \min_{\theta \in \Theta} d_a(\mathbf{x}, T_\theta^{-1}(\mathbf{y}))$ invariant to geometric attacks in class $\{T_\theta, \theta \in \Theta\}$
- Maximum distortion level D_2 for attacker

Information-Theoretic Setup



- Communications with side information (Gel'fand-Pinsker 1980)
- M uniformly distributed over message set $\mathcal{M}_n$
  - Coding problem: $R \triangleq \lim_{n\to\infty} \frac{1}{n} \log_2 |\mathcal{M}_n| > 0$
  - Detection problem: $\mathcal{M}_n$ independent of $n \quad \Rightarrow R = 0$
- Distortion levels D_1 and D_2
- Class of attacks: $\mathcal{P}_n \triangleq \{p_{\mathbf{Y}|\mathbf{X}}\}$
- Attacker knows f_n, g_n, selects $(A_{Z|X}, \theta) \sim p_{\mathbf{Y}|\mathbf{X}} \in \mathcal{P}_n$



- Minmax probability of error:

$$P_e^*(n, \mathcal{M}_n, \mathcal{P}_n) = \inf_{f_n, g_n} \sup_{p_{\mathbf{Y}|\mathbf{X}} \in \mathcal{P}_n} P_e(f_n, g_n, p_{\mathbf{Y}|\mathbf{X}})$$

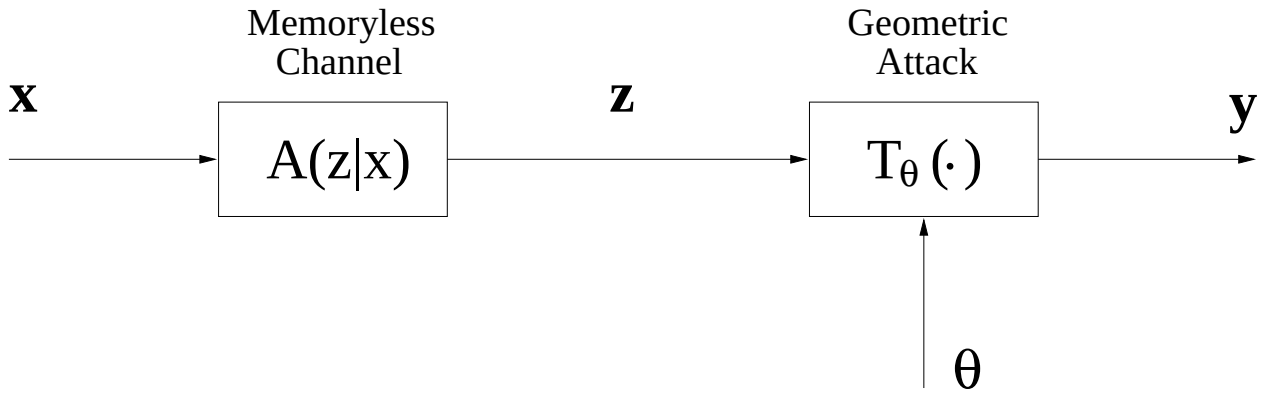
- Rate R is achievable if $\limsup_{n\to\infty} P_e^*(n, \mathcal{M}_n, \mathcal{P}_n) = 0$
- Supremum of achievable rates is **capacity** $C(D_1, D_2)$
- Error exponent

$$e^*(R, D_1, D_2) = \lim \inf_{n\to\infty} -\frac{1}{n} \log P_e^*(n, \mathcal{M}_n, \mathcal{P}_n), \quad 0 \le R \le C$$

- Write $P_e^*(n, \mathcal{M}_n, \mathcal{P}_n) \doteq 2^{-n\, e^*(R, D_1, D_2)}$

- Can derive expression for $C(D_1, D_2)$ for various classes of attacks involving *additive distortion functions*:
  - Memoryless attacks [MO'99]
  - Max-distortion attacks [CL'01, SM'03]
- Can also derive upper and lower bounds on $e^*(R, D_1, D_2)$ [SM'04] [MW'04]
- What happens under geometric attacks?

Complexity of Geometric Attacks



- Consider two cases: receiver knows θ or not
- If receiver knows θ, it can "undo" geometric attacks
- If receiver doesn't know θ but Θ is compact,
  - there is **no decrease in capacity**; $C(D_1, D_2)$ is achieved using traditional decoder, aided by pilot.
  - there is not even a decrease in $e_r^*(R, D_1, D_2)$, i.e., there exists a *universal decoder* against such geometric attacks

Standard WM Codes and Their Limitations

- Example: standard Quantization Index Modulation codes perform well against additive Gaussian attacks but are vulnerable to scaling attacks, delays, warping, etc.
- The main culprit is the minimum-Euclidean-distance decoder

Unitary Geometric Attack Channels

- Assume $\mathbf{s}, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$
- T_θ is a unitary matrix
  (geometric attack is linear and preserves signal energy)
- Example: *cyclic delay attack*
  - Attacker performs bandlimited interpolation of $\mathbf{x}$, applies cyclic delay $\theta \in [0, n]$, and resamples signal
- Assume $\mathbf{S} \sim \mathcal{N}(0, \Sigma)$ and $T_\theta \Sigma T_\theta^T$ is independent of θ
  $\Rightarrow$ statistics of $\mathbf{S}$ are invariant under T_θ

Example: M-ary Watermark Detection in iid Gaussian Noise

- Code rate $R = 0$
- Additive spread-spectrum embedding rule $\mathbf{x} = \mathbf{s} + \mathbf{w}_m$
- $M \leq n$ orthogonal watermarks $\mathbf{w}_m \in \mathbb{R}^n$,
  each with energy $\|\mathbf{w}_m\|^2 = nD_1$
- Watermark constellation $\mathcal{C} = \{\mathbf{w}_m\}$;
  transformed watermark constellation $\mathcal{C}_\theta = \{T_\theta \mathbf{w}_m\}$
- Total noise at receiver $\sim \mathcal{N}(0, \sigma^2 I_n)$
- Watermark to Noise ratio: $WNR = D_1/\sigma^2$
- Minimum distance of $\mathcal{C}_\theta$: $d_{\min} = \sqrt{2nWNR}$, same for all θ

Coherent Case: Detector knows θ

- Hypothesis test: $H_m \ : \ \mathbf{Y} \sim \mathcal{N}(T_\theta \mathbf{w}_m, \sigma^2 I_n), \quad m \in \mathcal{M}$
- Optimal likelihood ratio test (**LRT**) is a correlator-detector:

$$\hat{m} = \underset{m \in \mathcal{M}}{\operatorname{argmax}} \, \mathbf{y}^T T_\theta \mathbf{w}_m$$

- Error probability:

$$P_e \leq \frac{M-1}{2} Q(d_{\min}/2) \doteq e^{-n \frac{WNR}{4}}$$

- Computational complexity: no search, just $|\mathcal{M}|$ correlations $\Rightarrow |\mathcal{M}|$ ops/sample

Noncoherent Case: Detector doesn't know θ

- Hypothesis test:

$$H_m \ : \ \mathbf{Y} \sim \mathcal{N}(T_\theta \mathbf{w}_m, \sigma^2 I_n), \quad m \in \mathcal{M}, \quad \theta \in \Theta$$

- Worst-case error probability $\max_{\theta \in \Theta} P_e(f_n, g_n, \theta)$
- Can we do (nearly) as well as in the coherent case?
- What kind of detector g_n is (nearly) optimal?
- What kind of watermark code f_n should we use?

Taxonomy for Practical WM Schemes

- Invariant WM schemes
- Generalized Likelihood Ratio Test (GLRT) detectors
- Pilot-aided detection

Invariant Watermarks

- Invariant watermark: select embedding domain such that $p(\mathbf{y}|\theta, H_m)$ is independent of θ
  - θ is nonidentifiable
- Detector has same performance as in coherent case (against memoryless attacks in invariant domain)
- No increase in computational complexity
- Possible loss of robustness against memoryless attacks in original image domain
- And invariant domain does in general not exist!

Invariant Detection Tests

- Construct *good* detection statistics whose distribution is independent of θ
- Example: noncoherent detection of sinusoids (M-ary FSK) subject to cyclic delay attacks:

$$\mathbf{w}_m(i) = \sqrt{2D_1}\sin(2\pi f_m i), \quad 0 \le i < n, \quad f_m = (K+m)/n$$

- Detection statistics $z_m = \left|\sum_{i=0}^{n-1} y(n) e^{j2\pi f_m i}\right|^2, \quad m \in \mathcal{M}$
- Detection test: $\hat{m} = \text{argmax}_{m\in\mathcal{M}}\, z_m$
- Error probability $P_e \le (M-1)\, e^{-n\frac{WNR}{4}}$
- No loss in error exponent wrt coherent case

Generalized Likelihood Ratio Test (GLRT)

- **Step 1: Maximum-Likelihood Estimation:**

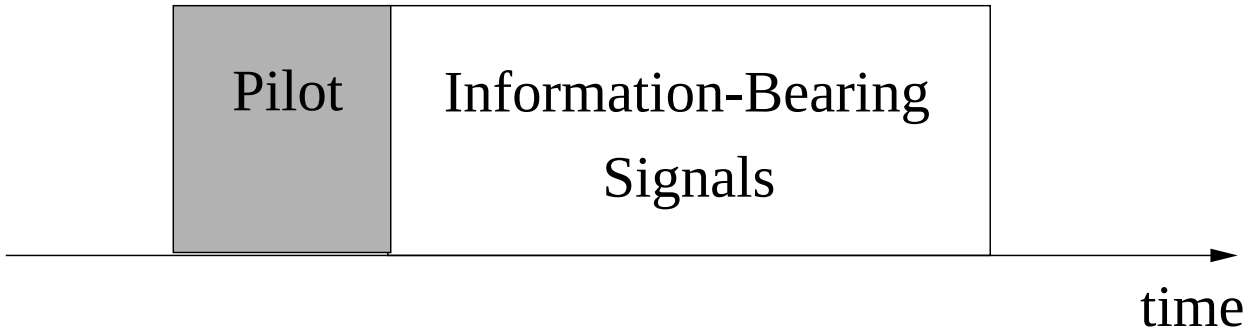
$$\begin{aligned}\hat{\theta}_m &\triangleq \operatorname*{argmax}_{\theta} p(\mathbf{y}|\theta, H_m) \\ &= \operatorname*{argmin}_{\theta} \|\mathbf{y} - T_\theta \mathbf{w}_m\|, \quad m \in \mathcal{M}\end{aligned}$$

- **Step 2: Correlator Detector:**

$$\hat{m} = \operatorname*{argmax}_{m \in \mathcal{M}} \mathbf{y}^T T_{\hat{\theta}_m} \mathbf{w}_m$$

- Asymptotic optimality of GLRT: $\theta \in \mathbb{R}$, still $P_e \doteq e^{-n\frac{WNR}{4}}$!
- Computational complexity: mostly $|\mathcal{M}|$ full searches

Pilot-Aided Detection



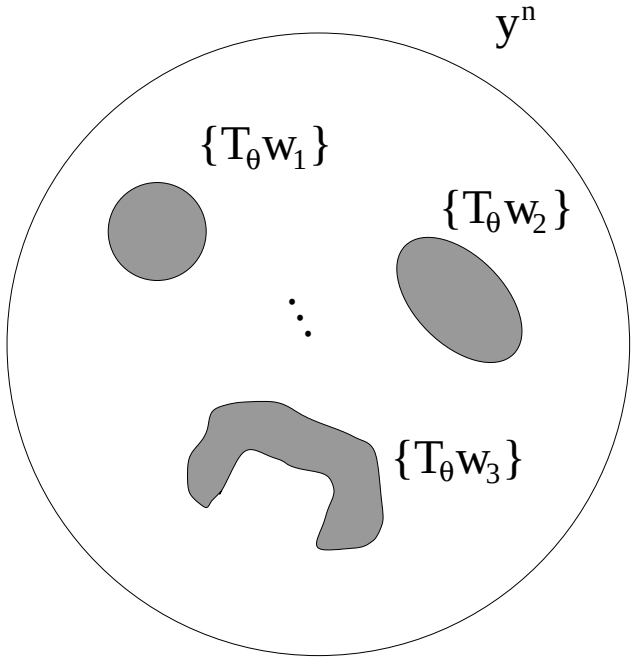
- Pilot known to receiver, conveys info about channel law $p_{\mathbf{Y}|\mathbf{X}}$
- Up to $n-1$ orthogonal WM's $\mathbf{w}_m$, each with energy nD_w
- Assume pilot $\mathbf{p} \in \mathbb{R}^n$ is orthogonal to $\{\mathbf{w}_m\}$, has energy nD_p.
- Transmit watermarked signal $\mathbf{x} = \mathbf{s} + \mathbf{w}_m + \mathbf{p}$
- Embedding distortion $= nD_1 = n(D_w + D_p) \quad \Rightarrow D_w < D_1$

- Computational complexity: mostly **one** full search (match $\mathbf{p}$ to $\mathbf{y}$)
- Reduces effective WNR by a factor of $1 - D_p/D_1$
  and therefore decreases error exponent
- Large estimation errors $\hat{\theta} - \theta$ also contribute to P_e
  $\Rightarrow$ optimal D_p results from large-deviations analysis

Detection *vs* computational-complexity tradeoff

More General Geometric Attacks

- Generally, T_θ is not unitary, not even linear



- How can we generalize the previous WM design/detection approaches?

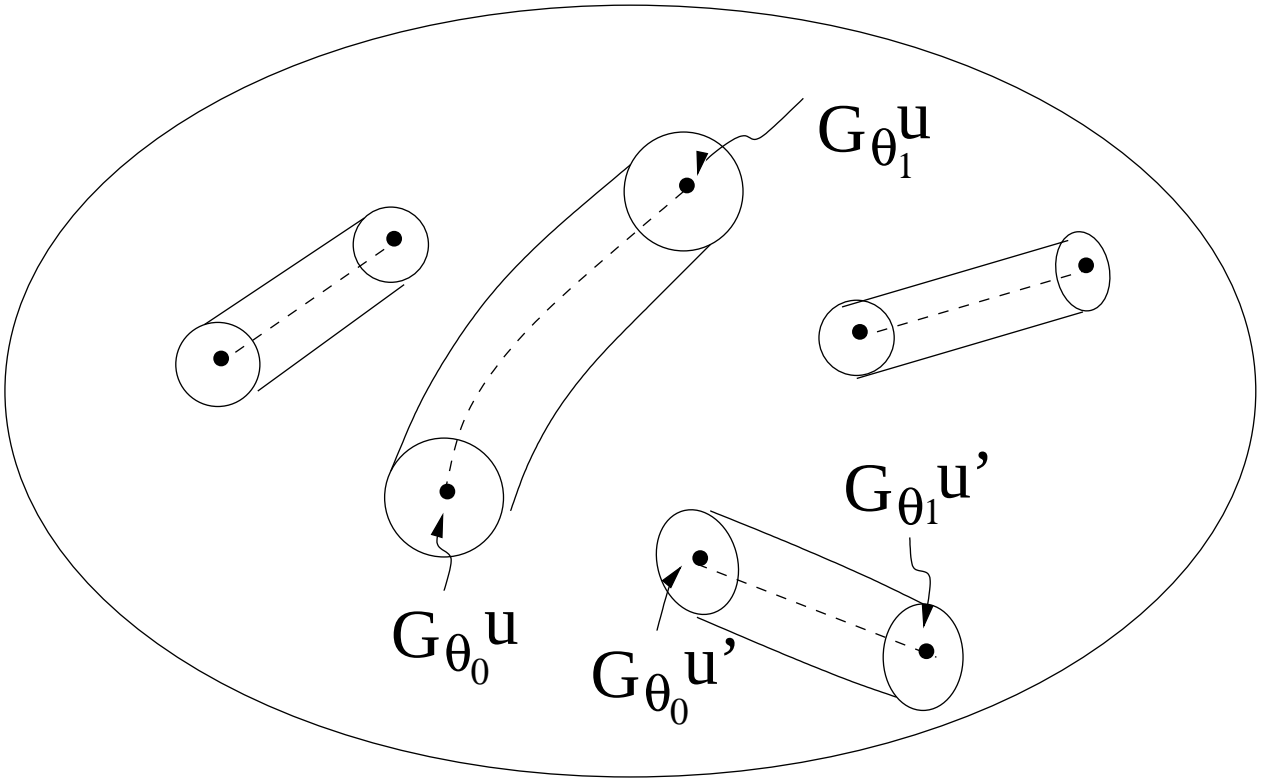
- *Invariant WM's*: **very hard** if not impossible to construct
- *GLRT approach*:
  - due to invertibility and smoothness of $T_\theta(\cdot)$,
    GLRT is **asymptotically optimal** as $n \to \infty$
    provided Θ is not "too complex" (e.g., $\Theta \in \mathbb{R}^d$ where $d \ll n$)
  - Proof is based on notion of competitive minimaxity [FM'02]
- *Pilot-based approach*:
  capacity-achieving, but **lower error exponents**

Fast Search

- Search for $\hat{\theta}_m = \text{argmax}_{\theta \in \Theta}\, p(\mathbf{y}|\theta, H_m), \quad m \in \mathcal{M}$
- Computational cost of full search (for discrete Θ) $\sim n\,|\mathcal{M}|\,|\Theta|$
- Replace full search by partial search
- Analogous to classical signal processing problems such as fast motion estimation in video, fast image registration, etc.

More General Watermarking Codes [M'03]

- Gelfand-Pinsker setup



- How to make it practical?

Conclusion

- Detection performance *vs* complexity tradeoffs
- Asymptotics ($n \to \infty$):
  - Assume θ is "low-dimensional" or belongs to compact set
  - Invariant watermarks may not exist...
  - Pilot-based schemes are capacity-achieving but cause loss in error exponents
  - GLRT is asymptotically optimal for our problem
- In practice:
  - GLRT-type detectors with fast search may be attractive
  - So are pilot-based schemes, if $|\mathcal{M}_n|$ is large

References

-[LN'98] Lapidoth and Narayan (IT 1998)

-[FL'98] Feder and Lapidoth (IT 1998)

-[MO'99] Moulin and O'Sullivan 1999 (IT 2003)

-[MM'02] Moulin and Mihcak (IP 2002)

-[CL'01] Cohen and Lapidoth 2001 (IT 2002)

-[FM'02] Feder and Merhav (IT 2002)

-[SM'03] Somekh-Baruch and Merhav (IT 2003)

-[SM'04] Somekh-Baruch and Merhav (IT 2004)

-[M'03] Moulin (SSP 2003)

-[MW'04] Moulin and Wang (ITW 2004)