When is a maximal invariant hypothesis test better than the GLRT?

Hyung Soo Kim and Alfred Hero

Dept. of EECS,
The University of Michigan

Oct. 2000

Outline

1. Detection Criteria
2. Motivating Example: Poisson image classification
3. Sufficency vs Invariance
4. Example: ATR for deep hide scenario

Detection Criteria

Generalized Likelihood Ratio (GLR) Test

$$\frac{\max_{\theta\in\Theta_1} f(\mathbf{X};\theta)}{\max_{\theta\in\Theta_0} f(\mathbf{X};\theta)} = \frac{f(\mathbf{X};\hat{\theta}_1)}{f(\mathbf{X};\hat{\theta}_0)} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

- GLRT is a function of MLEs $(\hat{\theta}_0, \hat{\theta}_1)$.
- GLRT is not asymptotically UMP unless MLEs are consistent.
- GLRT performance can be very poor (not even unbiased) in finite sample regime.
- GLRT optimization (maximization) may be intractable.

Maximal Invariant (MI) Test

: a function of *Maximal Invariant* – Systematic method for data reduction which accounts for parameter uncertainty

Poisson imaging example

Observe: Planar Poisson process $\{\underline{x}_i,\ i = 1, \ldots, N\}$

Intensity: $\lambda(\underline{x})$, $\underline{x} = (x_1, x_2)$.

$$\begin{aligned} H_0 &: \lambda(\underline{x}) = \lambda_0(\underline{x}) \\ H_1 &: \lambda(\underline{x}) = \lambda_\rho(\underline{x}), \quad \rho \in \mathcal{R}. \end{aligned}$$

where

$$\tilde{\lambda}_0(\underline{x}) = \left\{ \begin{array}{ll} \frac{\alpha}{a_3}, & \|\underline{x}\| \in (r_2, r_3] \\ \frac{1-2\alpha}{a_2}, & \|\underline{x}\| \in (r_1, r_2] \\ \frac{\alpha}{a_1}, & \|\underline{x}\| \in [0, r_1] \end{array} \right., \quad \tilde{\lambda}_\rho(\underline{x}) = \left\{ \begin{array}{ll} \frac{\rho(\underline{x})}{a_3}, & \|\underline{x}\| \in (r_2, r_3] \\ 0, & \|\underline{x}\| \in (r_1, r_2] \\ \frac{1-1/a_3}{a_1}, & \|\underline{x}\| \in [0, r_1] \end{array} \right. .$$

and $\rho = \rho(\underline{x}) \in \mathcal{R}$ has known radial marginals

$$\int_0^{2\pi} \rho(r\cos(\psi), r\sin(\psi))d\psi = \frac{2\pi}{a_3}, \quad r_2 < r \leq r_3.$$

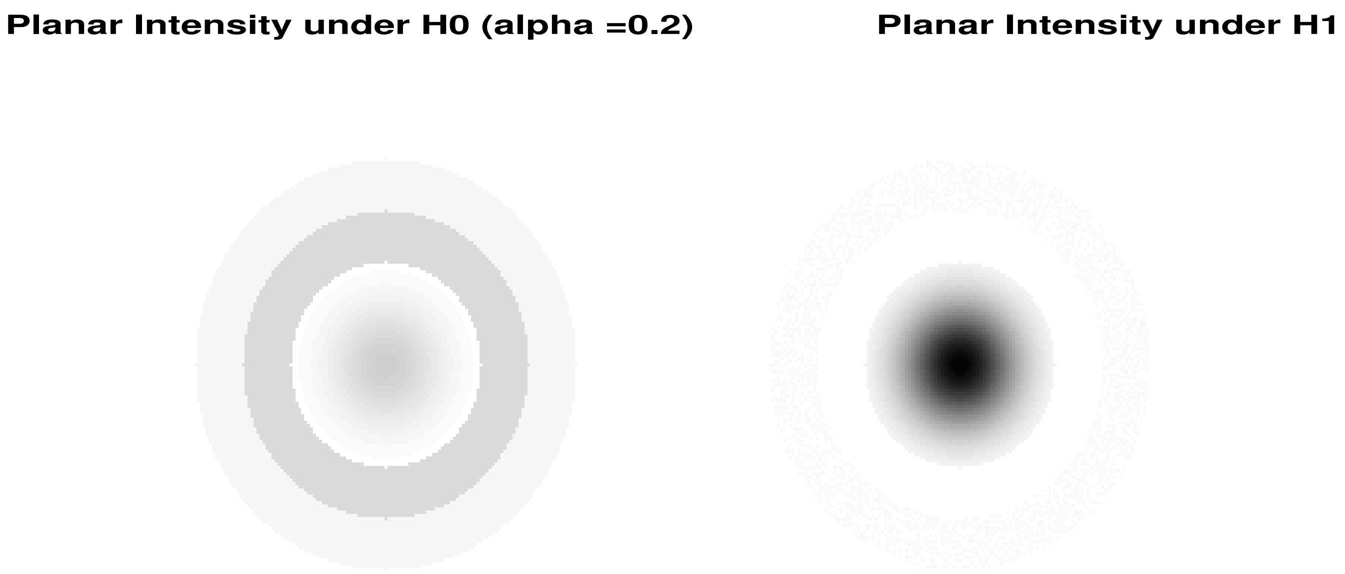


Figure 1: Planar intensities $\lambda(\underline{x})$ supported on nested rings at known radii. Under H_0 the intensity, shown for parameter $\alpha = 0.2$, is completely known. Under H_1 the outer ring has an unknown intensity whose integral over any fixed radius contour is equal to a known constant.

GLRT:

$$\max_{\rho\in\mathcal{R}} \frac{f_\rho(\underline{x})}{f_0(\underline{x})} = \left[\frac{1-1/a_3}{\alpha}\right]^{N_1} \max_{\rho\in\mathcal{R}} \prod_{i=1}^{N_3} \frac{\rho(\underline{x}_i)}{\alpha} \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \gamma$$

GLR critical region (large Λ): decide H_1 if $N_3 > 0$.

$$P_{FA} = 1 - P(N_3 = 0|H_0) = 1 - \exp\left(-\alpha\Lambda\right)$$

$$P_D = 1 - P(N_3 = 0|H_1) = 1 - \exp\left(-\frac{\Lambda}{a_3}\right).$$

Alternative: notice that model is *invariant* with respect to rotation of coordinate system about O:

$$\boxed{\underline{x} = re^{j\psi}}$$

$\Rightarrow$ Data reduction: throw away phase angle ψ of $\underline{x}$

$$H_0 \ : \ \lambda(r) = \lambda_0(r)$$

$$H_1 \ : \ \lambda(r) = \lambda_1(r)$$

where

$$\hat{\lambda}_0(r) = \begin{cases} \alpha \, \frac{2\pi r}{a_3}, & r \in (r_2, r_3] \\ 1 - 2\alpha \, \frac{2\pi r}{a_2}, & r \in (r_1, r_2] \\ \alpha \, \frac{2\pi r}{a_1} & r \in [0, r_1] \end{cases},$$

$$\hat{\lambda}_1(r) = \begin{cases} \frac{1}{a_3} \, \frac{2\pi r}{a_3}, & r \in (r_2, r_3] \\ 0, & r \in (r_1, r_2] \\ 1 - 1/a_3 \, \frac{2\pi r}{a_1}, & r \in [0, r_1] \end{cases}.$$

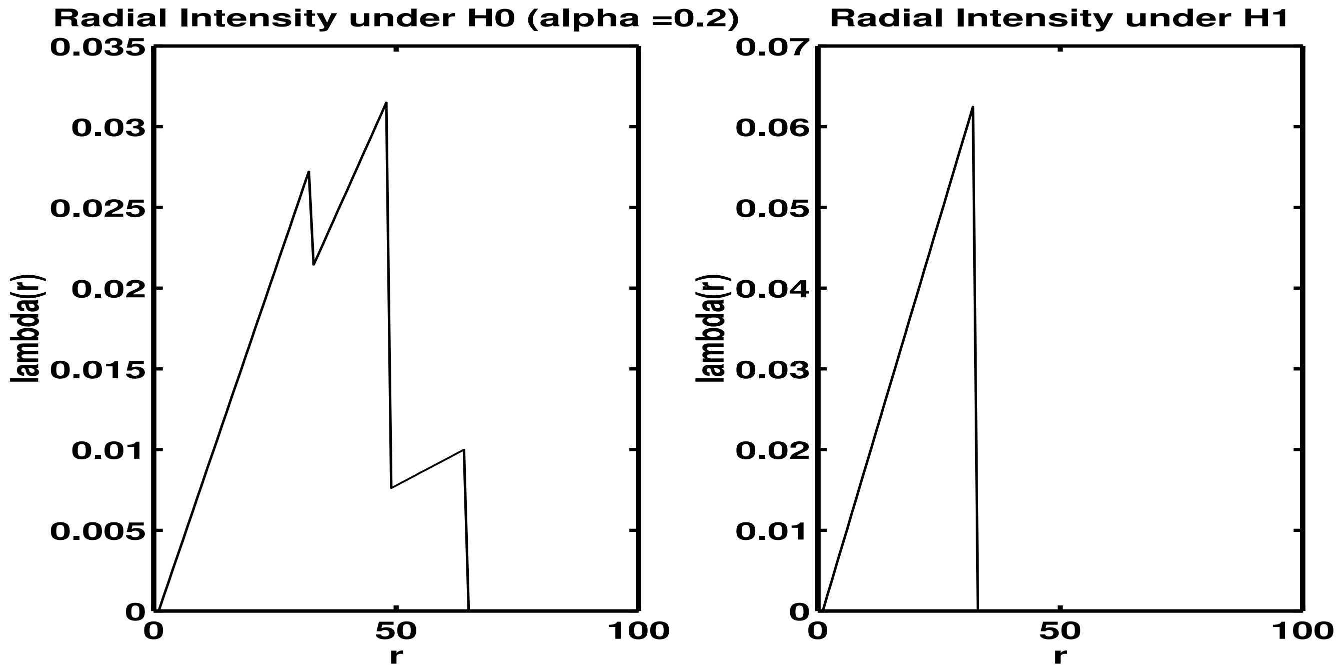


Figure 2: Marginal radial intensities $\lambda(r)$ under H_0 and H_1 derived from the planar intensities in Fig. 1.

Invariant LRT ($\Lambda \gg 0$):

$$\frac{f_1(r)}{f_0(r)} = \left[\frac{1 - 1/a_3}{\alpha}\right]^{N_1} \left[\frac{1/a_3}{\alpha}\right]^{N_3} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma,$$

or equivalently (assuming $a_3 \geq 1/\alpha$ and $\alpha \in [0, \frac{1}{2}]$)

$$N_1 + N_3\, c \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

where $c = \log\left[\frac{1}{a_3}\,\frac{1}{\alpha}\right] / \log\left[(1 - 1/a_3)\,\frac{1}{\alpha}\right]$.

$\boxed{\text{No dependence on } \rho \Rightarrow \text{UMPI test}}$

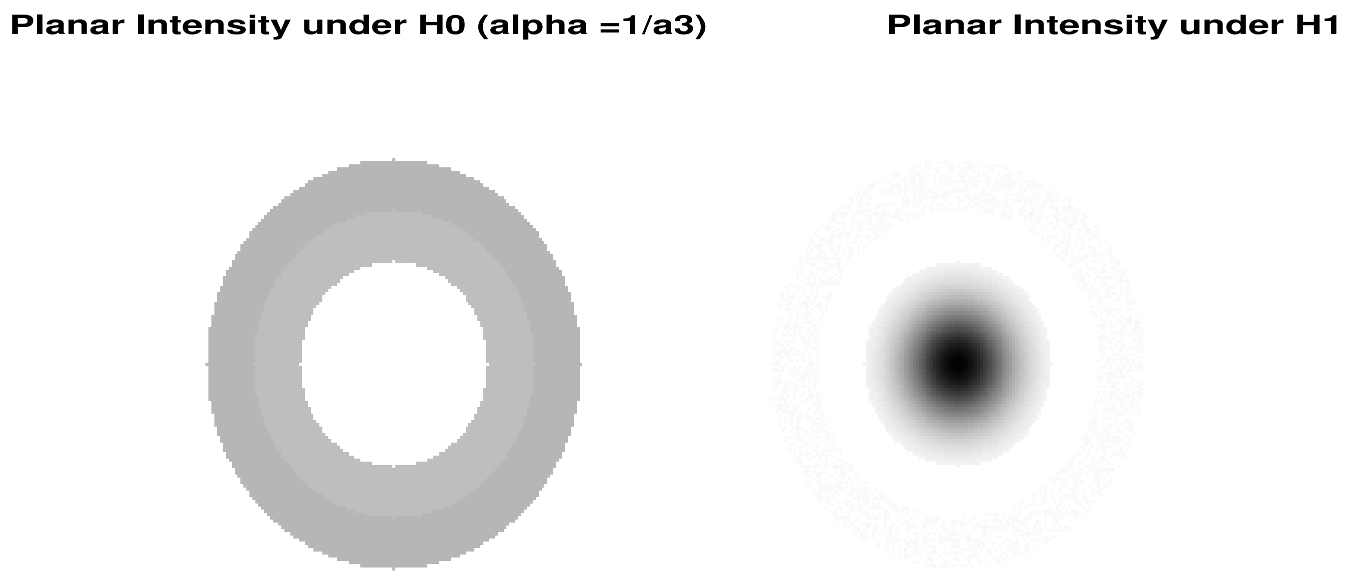


Figure 3: Same H_0 and H_1 planar intensities as in Fig. 1 except that $\alpha = 1/a_3$ where a_3 is the area of the outer ring.

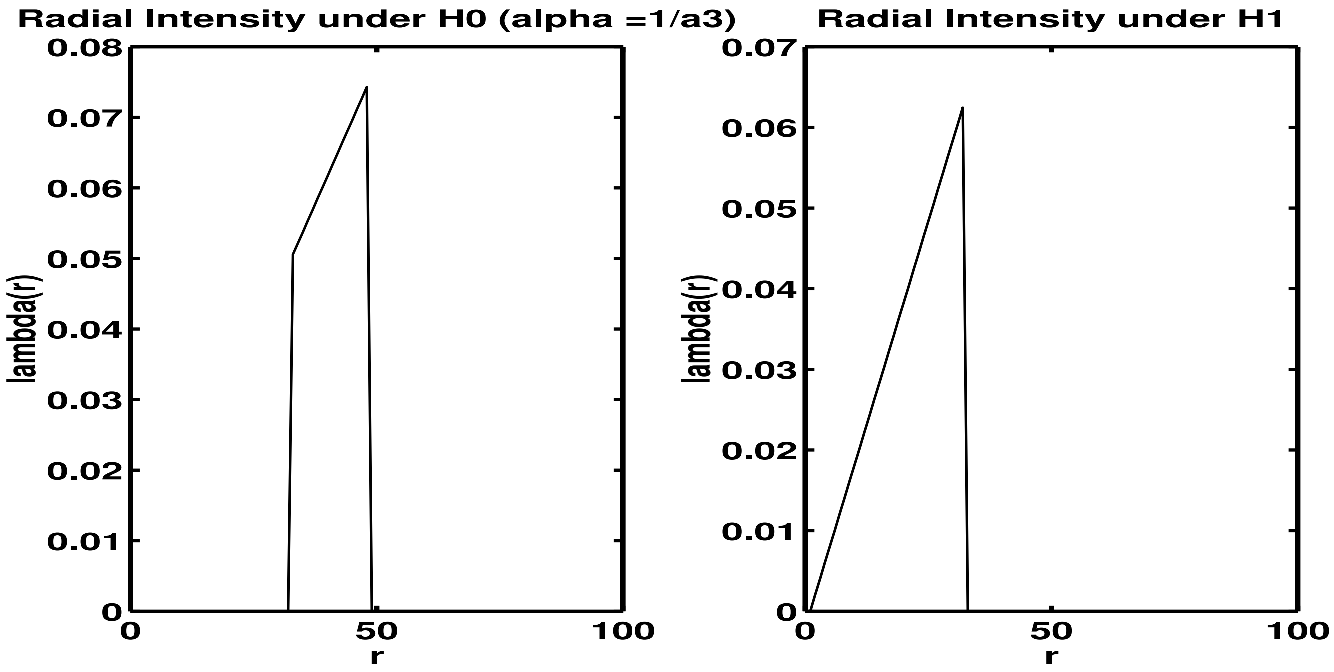


Figure 4: Marginal radial intensities $\lambda(r)$ under H_0 and H_1 derived from the planar intensities in Fig. 2.

Special Case: $\alpha = 1/a_3$

MPI critical region: decide H_1 if $N_1 > 0$

The (large λ) false alarm probability and probability of detection of the MPRI test are

$$P_{FA} \; = \; 1 - P(N_1 = 0|H_0) = 1 - \exp\left(-\alpha\Lambda\right)$$

$$P_D \; = \; 1 - P(N_1 = 0|H_1) = 1 - \exp\left(-(1 - 1/a_3)\Lambda\right).$$

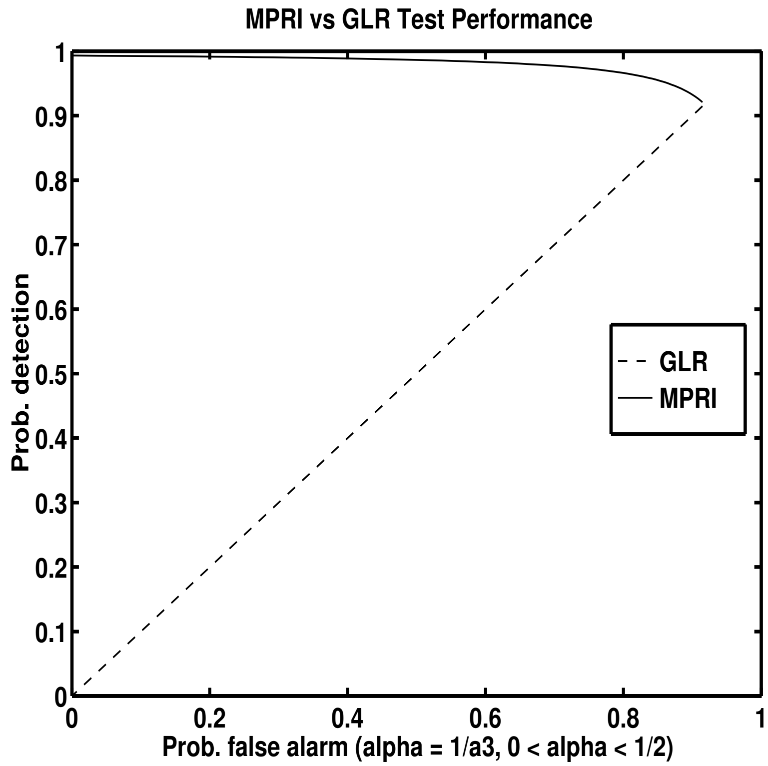


Figure 5: Probability of detection as a function of $P_F = 1 - e^{-\alpha\Lambda}$ associated with GLR and MPRI tests of H_0 vs. H_1 for the case of $a_3 = 1/\alpha$ shown Fig. 2.

Linear Multivariate Image Model

$$\mathbf{X} = [\underline{x}_1, \dots, \underline{x}_n]$$

$\underline{x}_k$ $(m \times 1)$: lexicographic ordered k-th sub-image

$$\mathbf{X} = \mathbf{S}\,\mathbf{A}\,\mathbf{B} + \mathbf{N}$$

GMANOVA model :

- $\mathbf{S} = [\underline{s}_1, \dots, \underline{s}_p]$ $(m \times p)$: target signatures (*known*)
- $\mathbf{A} = diag([a_1, \dots, a_p])$ $(p \times p)$: target amplitudes (*unknown*)
- $\mathbf{B} = \begin{bmatrix} \underline{b}_1^H \\ \vdots \\ \underline{b}_p^H \end{bmatrix}$ $(p \times n)$: target locations (*known*)

Multivariate Gaussian Noise Model

$\mathbf{N}$ $(m \times n)$: multivariate Gaussian matrix
with i.i.d. columns $(m \times 1)$ each having covariance matrix $\mathbf{R}$ $(m \times m)$

$$vec\{\mathbf{N}\} \sim \mathcal{N}(\underline{0}, \mathbf{R} \bigotimes \mathbf{I}_n)$$

$$cov[vec\{\mathbf{N}\}] = \mathbf{R} \bigotimes \mathbf{I}_n = \begin{bmatrix} \begin{matrix} \mathbf{R} & \mathbf{O} \\ \mathbf{O} & \mathbf{R} \end{matrix} & \mathbf{O} \\ \mathbf{O} & \begin{matrix} \mathbf{R} & \mathbf{O} \\ \mathbf{O} & \mathbf{R} \end{matrix} \end{bmatrix}_{(mn \times mn)}$$

Note : $vec\{F\mathbf{N}H\} \sim \mathcal{N}(\underline{0}, F\mathbf{R}F^H \bigotimes HH^H)$

Statistical Reduction via Sufficiency

Minimal Sufficient Statistic $T(\mathbf{X})$

$$\Lambda(\mathbf{X};\theta_0,\theta_1) \quad \triangleq \quad \frac{f(\mathbf{X};\theta_1)}{f(\mathbf{X};\theta_0)} = h(T(\mathbf{X});\theta_0,\theta_1)$$

$$\{\mathbf{X} : \Lambda(\mathbf{X};\theta_0,\theta_1) = \lambda\}_{\lambda>0} \quad = \quad \{\mathbf{X} : T(\mathbf{X}) = t\}_t$$

- $T(\mathbf{X})$ indexes the constant contours (orbits) of Λ for any fixed $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$.
- $T(\mathbf{X})$ achieves maximal data reduction while preserving information necessary to estimate θ_0 and θ_1, and discriminate between H_0 and H_1.
- Distribution of $T(\mathbf{X})$ depends on particular values of $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$.

Statistical Reduction via Invariance

Maximal Invariant Statistic $Z(\mathbf{X})$

$$\begin{aligned}\tilde{\Lambda}(\mathbf{X}) &\triangleq \{\frac{f(\mathbf{X};\theta_1)}{f(\mathbf{X};\theta_0)} : \theta_0 \in \Theta_0, \theta_1 \in \Theta_1\} \\ &= \{\tilde{h}(Z(\mathbf{X});\theta_0,\theta_1) : \theta_0 \in \Theta_0, \theta_1 \in \Theta_1\}\end{aligned}$$

- $Z(\mathbf{X})$ indexes the orbits of the set function $\tilde{\Lambda}$ for $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$.
- Maximal Invariant can be defined utilizing simple topological group structure of Θ_0 and Θ_1.

Group Characterization

Let $\mathcal{G}$ be a group of transformations $g : \chi \to \chi$ acting on $\mathbf{X}$.

Suppose that for each θ,

$$\exists\,! \quad \bar{\theta} = \bar{g}(\theta) \quad \text{such that} \quad f_\theta(g(\mathbf{X})) = f_{\bar{\theta}}(\mathbf{X}).$$

If $\bar{g}(\Theta_0) = \Theta_0,\ \bar{g}(\Theta_1) = \Theta_1$, then the decision problem is invariant to $\mathcal{G}$.

$\mathbf{Z} = Z(\mathbf{X})$ is *maximal invariant* iff

1. (invariant property) $Z(g(\mathbf{X})) = Z(\mathbf{X}) \quad$ for all $\quad g \in \mathcal{G}$,
2. (maximal property) $Z(\mathbf{X}) = Z(\mathbf{Y}) \ \Rightarrow\ \mathbf{Y} = g(\mathbf{X})$ for some $g \in \mathcal{G}$.

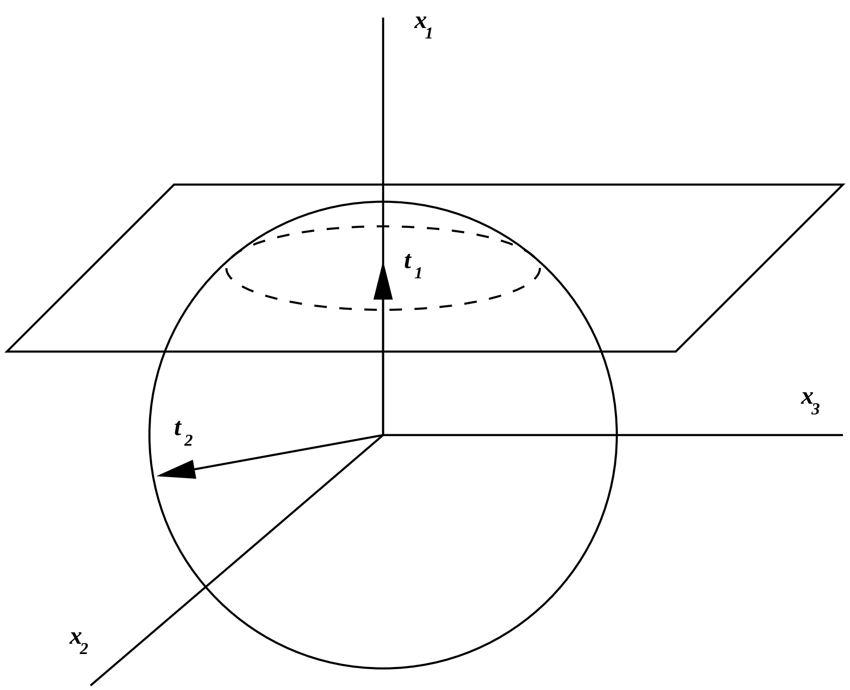
<u>Example</u> : Gaussian vector ($m = 1$) with unknown a, σ^2

$$\underline{\mathbf{x}}^T = a\,[1, 0, \ldots\;, 0] + \underline{N}, \qquad \underline{N} \sim \mathcal{N}(\underline{0}^T, \sigma^2\mathbf{I})$$

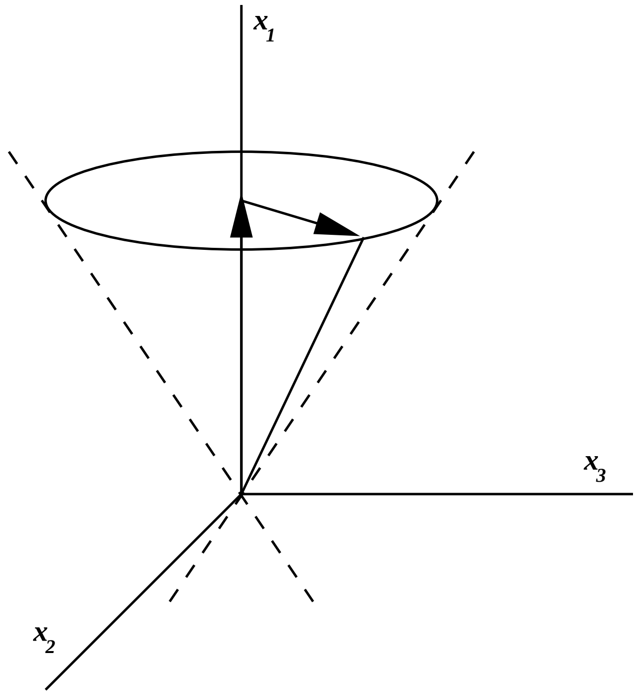
$$\begin{cases} \Theta_0 = \{a = 0, \sigma^2 > 0\} \\ \Theta_1 = \{a \neq 0, \sigma^2 > 0\} \end{cases}$$

$$\longrightarrow \qquad \ln \Lambda_{\theta_0,\theta_1}(\underline{\mathbf{x}}) = \frac{a}{\sigma_1^2} \cdot \underline{\mathbf{e}}_1^T \underline{\mathbf{x}} + \frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2\sigma_1^2} \cdot \underline{\mathbf{x}}^T \underline{\mathbf{x}} - \frac{a^2}{2\sigma_1^2} + n \ln \frac{\sigma_0}{\sigma_1}$$

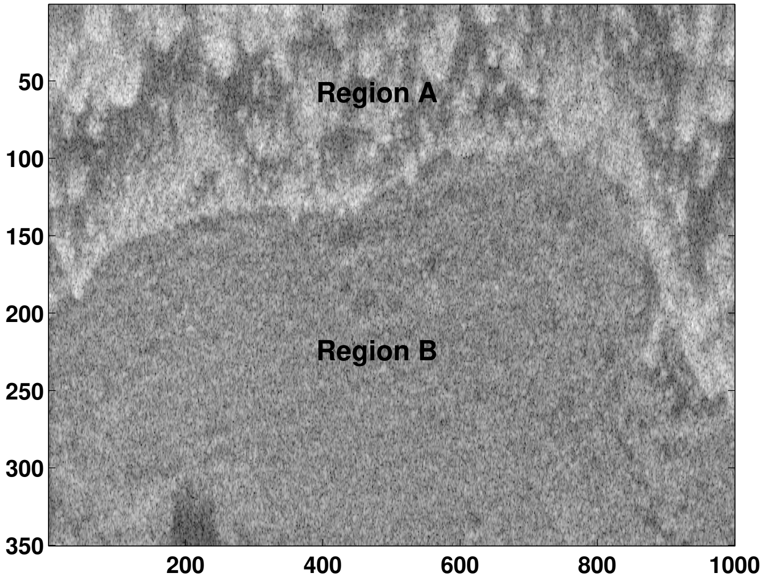
- **Sufficiency Orbit** : $[\underline{\mathbf{e}}_1^T \underline{\mathbf{x}}, \underline{\mathbf{x}}^T \underline{\mathbf{x}}] = [t_1, t_2]$ (circle in $\mathbf{R}^3$)
- **Invariance Orbit** : $|\underline{\mathbf{e}}_1^T \underline{\mathbf{x}}|^2 / \underline{\mathbf{x}}^T \underline{\mathbf{x}} = z$ (cone in $\mathbf{R}^3$)



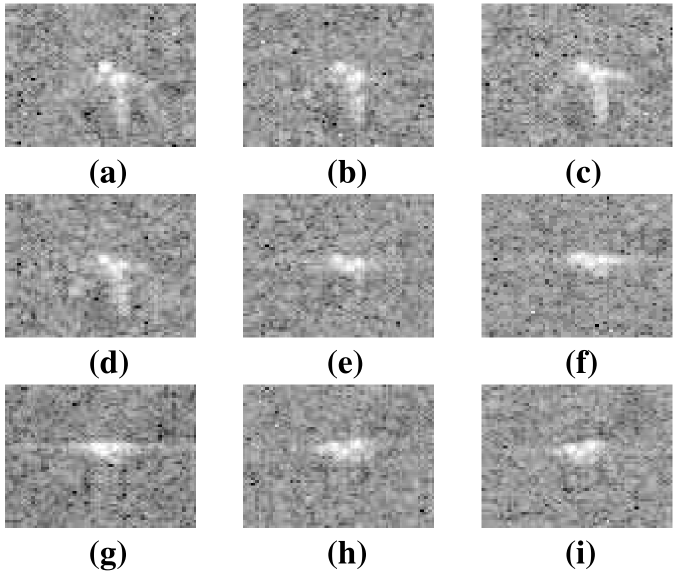
Sufficiency orbit



Invariance orbit



SAR Clutter image with a target in the boundary



Target images

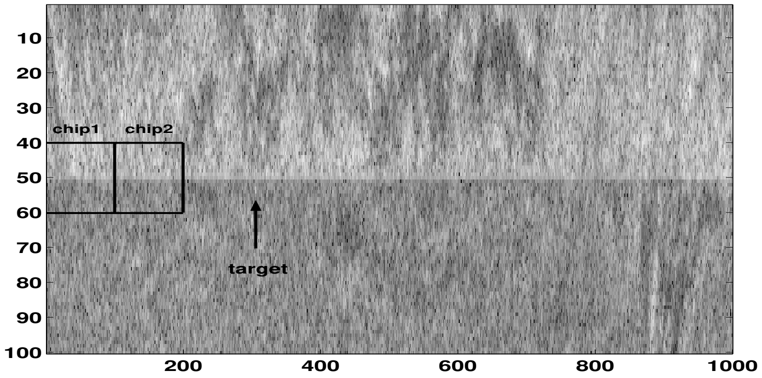


Image realigned along the extracted boundary

Partially Known (*Structured*) Clutter

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_A & \mathbf{O} \\ \mathbf{O} & \mathbf{R}_B \end{bmatrix}$$

- Case 1. : $\mathbf{R}_A > 0$, $\mathbf{R}_B > 0$
- Case 2. : $\mathbf{R}_A > 0$, $\mathbf{R}_B = \sigma^2\mathbf{I}$, $\sigma^2 > 0$
- Case 3. : $\mathbf{R}_A > 0$, $\mathbf{R}_B = \mathbf{I}$

$$\begin{aligned} \mathbf{X} &= a \begin{bmatrix} \underline{s}_A \\ \underline{s}_B \end{bmatrix} \underline{e}_1^T + \begin{bmatrix} \mathbf{N}_A \\ \mathbf{N}_B \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix} = \begin{bmatrix} \underline{x}_{A1} & \mathbf{X}_{A2} \\ \underline{x}_{B1} & \mathbf{X}_{B2} \end{bmatrix} = \begin{bmatrix} x_{A11} & \underline{x}_{A12} \\ \underline{x}_{A21} & \mathbf{X}_{A22} \\ x_{B11} & \underline{x}_{B12} \\ \underline{x}_{B21} & \mathbf{X}_{B22} \end{bmatrix} \end{aligned}$$

GLR Test Statistics

$\mathbf{R}_A$	$\mathbf{R}_B$	Log GLR : $\frac{1}{n}\ln\Lambda = \max_a\{\cdot\}$
?	?	$\ln\left[\frac{1+p(0,\underline{s}_A,\mathbf{X}_A)}{1+p(a,\underline{s}_A,\mathbf{X}_A)}\right] + \ln\left[\frac{1+p(0,\underline{s}_B,\mathbf{X}_B)}{1+p(a,\underline{s}_B,\mathbf{X}_B)}\right]$
?	$\sigma^2\mathbf{I}$	$\ln\left[\frac{1+p(0,\underline{s}_A,\mathbf{X}_A)}{1+p(a,\underline{s}_A,\mathbf{X}_A)}\right] + m_B \cdot \ln\left[\frac{q(0,\underline{s}_B,\mathbf{X}_B)}{q(a,\underline{s}_B,\mathbf{X}_B)}\right]$
?	$\mathbf{I}$	$\ln\left[\frac{1+p(0,\underline{s}_A,\mathbf{X}_A)}{1+p(a,\underline{s}_A,\mathbf{X}_A)}\right] + \frac{1}{n}[q(0,\underline{s}_B,\mathbf{X}_B) - q(a,\underline{s}_B,\mathbf{X}_B)]$

where m_B = number of rows in $\mathbf{X}_B$, n = number of columns in $\mathbf{X}_B$, and

$$p(a,\underline{s}_A,\mathbf{X}_A) = (\underline{x}_{A1} - a\underline{s}_A)^H(\mathbf{X}_{A2}\mathbf{X}_{A2}^H)^{-1}(\underline{x}_{A1} - a\underline{s}_A)$$

$$q(a,\underline{s}_B,\mathbf{X}_B) = tr\left\{(\mathbf{X}_B - a\underline{s}_B\underline{e}_1^T)^H(\mathbf{X}_B - a\underline{s}_B\underline{e}_1^T)\right\}$$

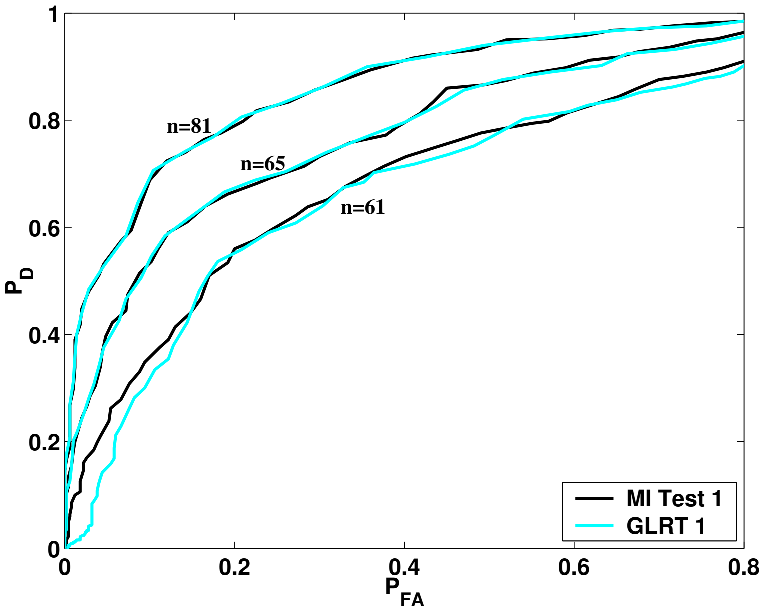
MI Test Statistics

$\mathbf{R}_A$	$\mathbf{R}_B$	MI Test : $T = \dfrac{\left| \begin{bmatrix} \underline{s}_A^H & \underline{s}_B^H \end{bmatrix} \begin{bmatrix} \mathbf{K}_A & \mathbf{O} \\ \mathbf{O} & \mathbf{K}_B \end{bmatrix}^{-1} \begin{bmatrix} \underline{x}_{A1} \\ \underline{x}_{B1} \end{bmatrix} \right|^2}{\begin{bmatrix} \underline{s}_A^H & \underline{s}_B^H \end{bmatrix} \begin{bmatrix} \mathbf{K}_A & \mathbf{O} \\ \mathbf{O} & \mathbf{K}_B \end{bmatrix}^{-1} \begin{bmatrix} \underline{s}_A \\ \underline{s}_B \end{bmatrix}}$
?	?	$\mathbf{K}_A = q_A \mathbf{X}_{A2} \mathbf{X}_{A2}^H, \quad \mathbf{K}_B = q_B \mathbf{X}_{B2} \mathbf{X}_{B2}^H$
?	$\sigma^2 \mathbf{I}$	$\mathbf{K}_A = q_A \mathbf{X}_{A2} \mathbf{X}_{A2}^H, \quad \mathbf{K}_B = v_2 \mathbf{I}$
?	$\mathbf{I}$	$\mathbf{K}_A = q_A \mathbf{X}_{A2} \mathbf{X}_{A2}^H, \quad \mathbf{K}_B = v_3 \mathbf{I}$

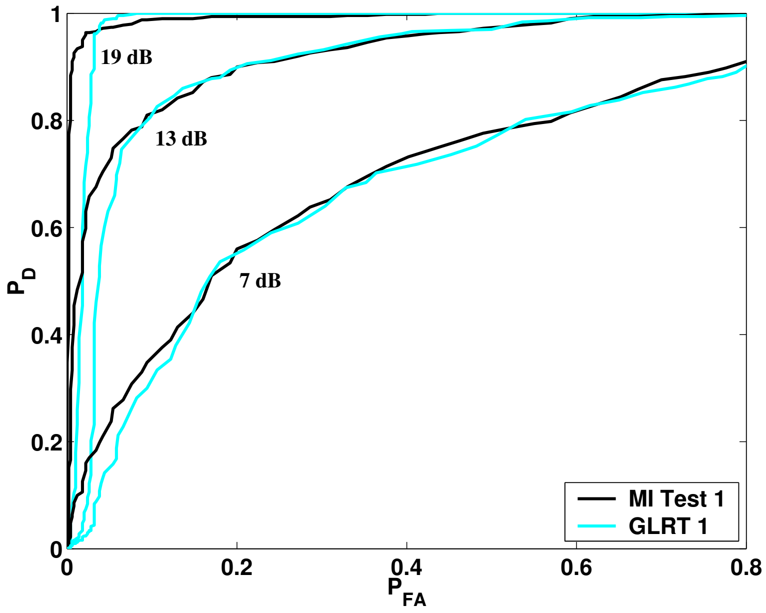
where

$$q_A = 1 + \underline{x}_{A1}^H (\mathbf{X}_{A2} \mathbf{X}_{A2}^H)^{-1} \underline{x}_{A1} \ , \ v_2 = \frac{1}{m_B} \mathrm{tr}\{\mathbf{X}_B^H \mathbf{X}_B\}$$

$$q_B = 1 + \underline{x}_{B1}^H (\mathbf{X}_{B2} \mathbf{X}_{B2}^H)^{-1} \underline{x}_{B1} \ , \ v_3 = n$$



Case 1. (SNR=7dB)



Case 1. ($n = 61$)

Conclusion

- Detection algorithms for inhomogeneous clutter
  - GLR extended to case of block structured covariance.
  - MI test proposed as an alternative.
- GLR/MI tests significantly outperform existing tests (Kelly's test, Bose and Steinhardt's test).

- GLR and MI tests are complementary.
  1. GLRT is asymptotically optimal (UMP) for large n.
  2. For small n and low SNR, MI test outperforms GLR for low P_{FA}.
- Suggests hybrid GLR/MI for optimal performance.