ROBUST DETECTION, CLASSIFICATION AND CLUSTERING

Progress over period 11/95 - 8/98

Alfred Hero
Dept. of EECS
University of Michigan

Principal Objective: To develop weak-target detection/classification/clustering algorithms which are robust to target and clutter variability yet maintain highest possible discrimination capability.

Methods: CFAR target detectors, iterative estimation, graph theoretic clustering algorithms, subspace processed moment matrices.

RESEARCH ACCOMPLISHMENTS

- **Detection algorithms for inhomogeneous clutter**: Max invariant and GLR tests (Hero&Guillouet [1]).
- **Accelerated EM and SAGE ML/PML algorithms**: proximal-point with Kullback penalty (Chretien&Hero [2],[3], [4]).
- **Theory and application of k-shortest graphs**: entropy estimation, robust clustering, image registration, resource allocation (Hero&Michel [5],[6],[7],[8],)
- PCA for higher order moment matrices: non-parametric target detection and classification (Hero, O'Neill&Williams [9],[10]).

Detection of Deep Hide Targets

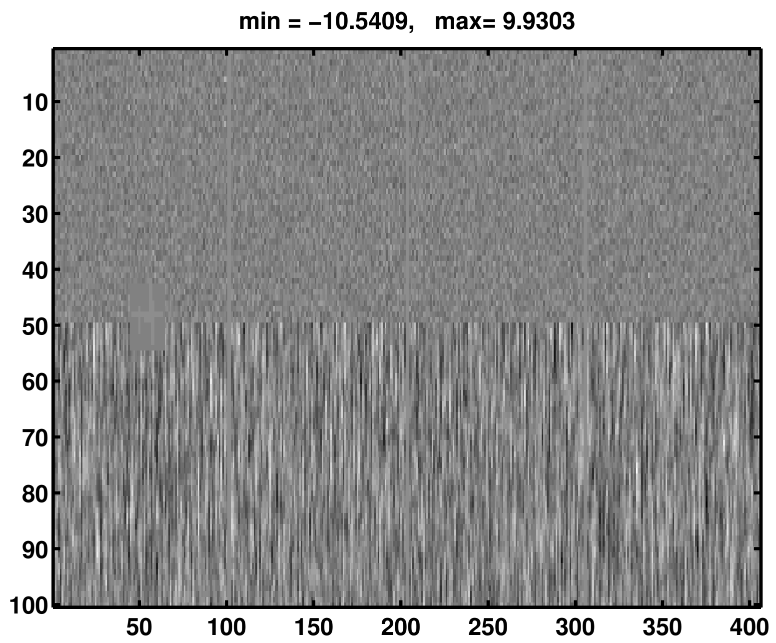


Figure 1: Known deep hide target on clutter boundary

Objective: Adaptive target detection in structured unknown clutter.

Methods: GLR and maximal invariant tests for structured MANOVA

Challenges:

- previous GLR work applies only to non-structured homogeneous clutter (Kelly:AES86, Kelly&Forsythe:TR848, Chen&Reed:AES87, ...)
- extension of group invariance approaches to SAR/IR imaging

Accomplishments:

- GLR extended to case of block structured clutter covariance.
- Maximal invariant tests extended to multiple targets (Hero&Guillouet:ICIP97 [1])
- quantitative improvement observed in P_D for $P_F < 0.1$.

Hypothesis testing problem:

$$H_0 \;:\; \mathbf{Y} = \mathbf{N}$$

$$H_1 \;:\; \mathbf{Y} = \mathbf{S}\,\mathbf{a}\,\mathbf{B} + \mathbf{N}$$

- $\mathbf{S} = [\underline{s}_1, \ldots, \underline{s}_p]$ = known $m \times p$ matrix of spatial target signatures
- $\mathbf{a}$ = unknown $p \times p$ matrix
- $\mathbf{B}$ known $p \times n$ indicator matrix of target-containing chips
- $\text{vec}\{\mathbf{N}\} \sim \mathcal{N}_{mn}\left(\underline{0}, \mathbf{R} \otimes \mathbf{I}_n\right)$
- $\mathbf{R}$ unknown **block diagonal** clutter covariance

GLR Test:

$$\Lambda(\mathbf{Y}) = \frac{\max_{\mathbf{a} \neq 0, \mathbf{R} > 0} f(\mathbf{Y};\; \mathbf{a},\; \mathbf{R})}{\max_{\mathbf{a} = 0, \mathbf{R} > 0} f(\mathbf{Y};\; \mathbf{a},\; \mathbf{R})} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

$\mathbf{R}_{AB}$	$\mathbf{R}_A$	$\mathbf{R}_B$	Log GLR
?	?	?	$2/n\ \ln(\Lambda) = \ln\left(\frac{1+q(0,\hat{R})}{1+q(\hat{a}_0,\hat{R})}\right)$
0	?	?	$2/n\ \ln(\Lambda) = \ln\left(\frac{1+q_A(0,\hat{R}_A)}{1+q_A(\hat{a}_1,\hat{R}_A)}\ \frac{1+q_B(0,\hat{R}_B)}{1+q_B(\hat{a}_1,\hat{R}_B)}\right)$
0	$\sigma^2\mathbf{C}_A$	?	$2/n\ \ln(\Lambda) = \ln\left(\frac{1+q_B(0,\hat{R}_B)}{1+q_B(\hat{a}_2,\hat{R}_B)}\frac{q_A(0,C_A)}{q_A(\hat{a}_2,C_A)}\right)$
0	$\mathbf{C}_A$	?	$2/n\ \ln(\Lambda) = \ln\left(\frac{1+q_B(0,\hat{R}_B)}{1+q_B(\hat{a}_3,\hat{R}_B)}\right) + q_A(0,C_A) - q_A(\hat{a}_3,C_A)$

Table 1: GLR Test statistics for various block structured clutter covariance

Special case: Single known target $\mathbf{S} = \underline{s}\ (m \times 1)$

- $q_A(a, R)$ is a quadratic form:

$$q_A(a, R_A) = (\underline{\bar{x}}_i - a\underline{s}_A)^H \mathbf{R}_A^{-1} (\underline{\bar{x}}_i - a\underline{s}_A)$$

- $\hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{a}_3$ are solutions to cubic equation

Group Invariance approach:

- Find group of transformations $\mathcal{G}$ leaving model invariant
- Find maximal invariant $\mathbf{Z}$ under group actions
- Compute MPI detector using $\mathbf{Z}$

$$\Lambda(\mathbf{Z}) = \frac{f(\mathbf{Z};\ \mathrm{a}, \mathbf{R})}{f(\mathbf{Z};\ 0, \mathbf{R})} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

OR

- Implement deflection criterion constructed on $\mathbf{Z}$

$$Q(\mathbf{Z}) \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

Numerical Example

Assumptions

- presegmented clutter regions
- 51 subimages acquired from 20×40 sliding data windows
- crosshair targets

Target detection algorithms implemented:

- Maximal invariant test
- Mismatched GLR using global clutter covariance matrix R
- Block covariance constrained GLR's implemented with cube rooting

Legends for graphs

- Kelly(uns) is GLRT which ignores the boundary information and uses the full covariance matrix estimate.
- Kelly(str) is the Kelly(uns) test implemented with a block diagonal covariance estimator.
- B-S (Bose-Steinhardt) is maximal invariant test for known target which accounts for block diagonal clutter covariance.
- GLRT1 is GLR test which assumes unknown clutter covariance on each side of boundary
- GLRT2 is GLR test which assumes clutter on one side of boundary is known white ($\mathbf{R}_A$ = identity matrix).
- GLRT3 is GLR test which assumes clutter on one side of boundary is white with unknown power ($\mathbf{R}_A = \sigma^2\mathbf{I}$)

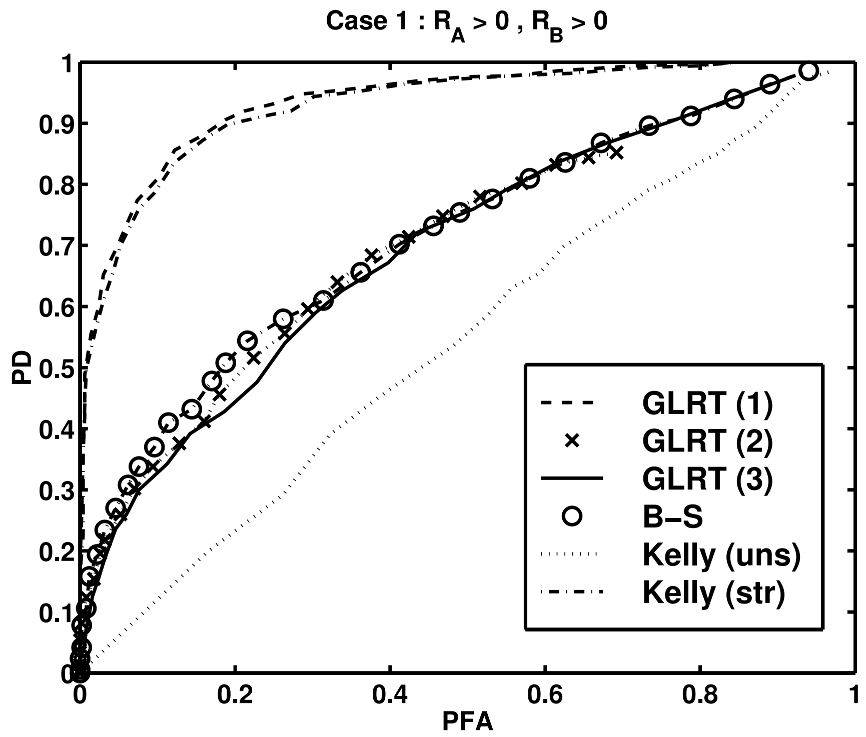


Figure 2: ROC comparisons for the case of a known target at known location on boundary between lowpass and highpass clutter regions. Note GLRT1 performs slightly better than Kelly(uns).

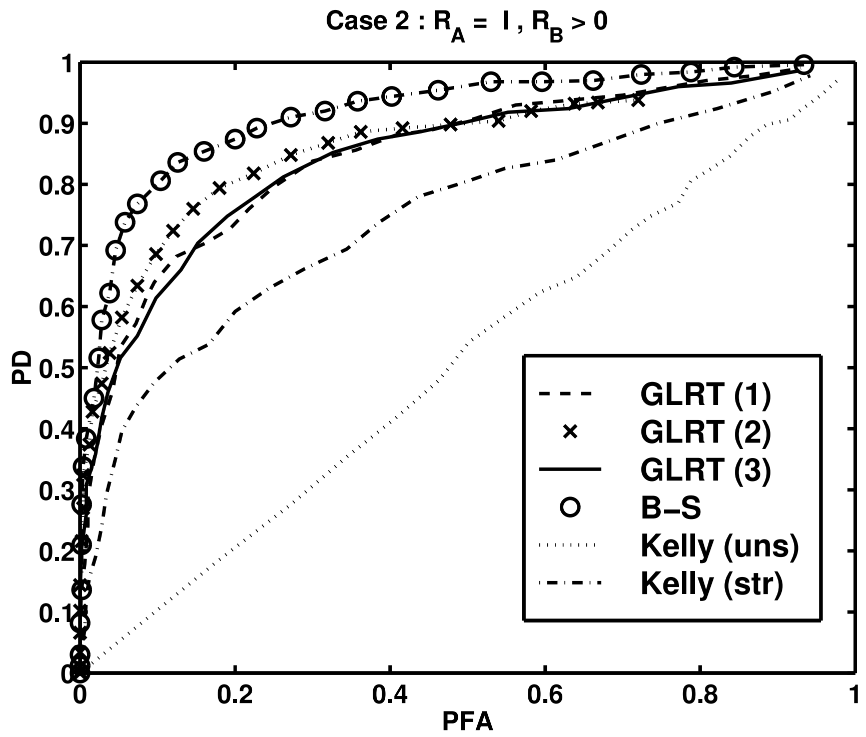


Figure 3: ROC comparisons for the case of a known target at known location on boundary between lowpass clutter region and a white clutter regionwith known power. Note B-S performs better than GLRT2.

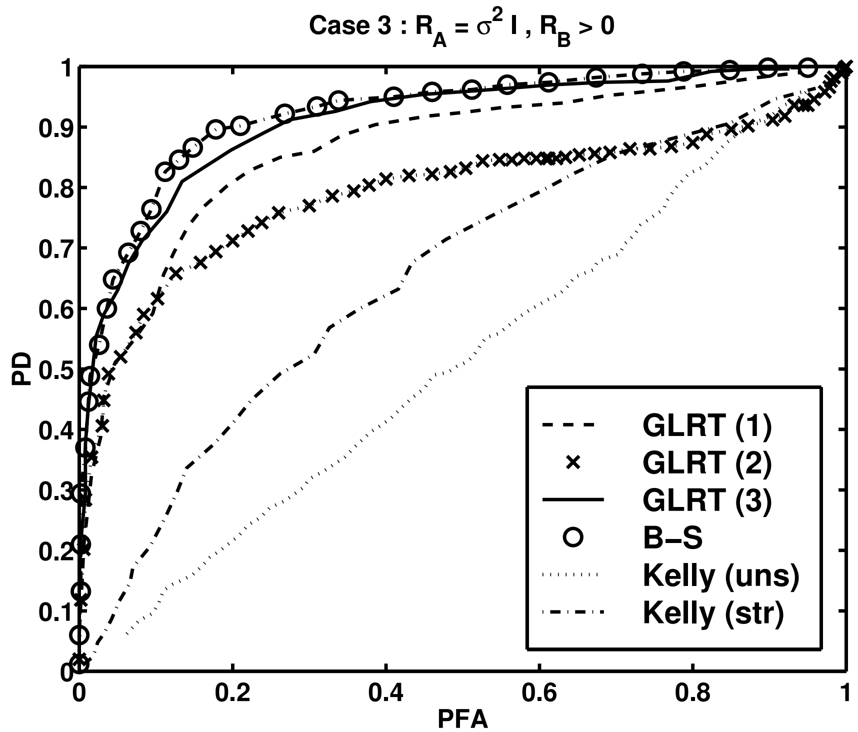
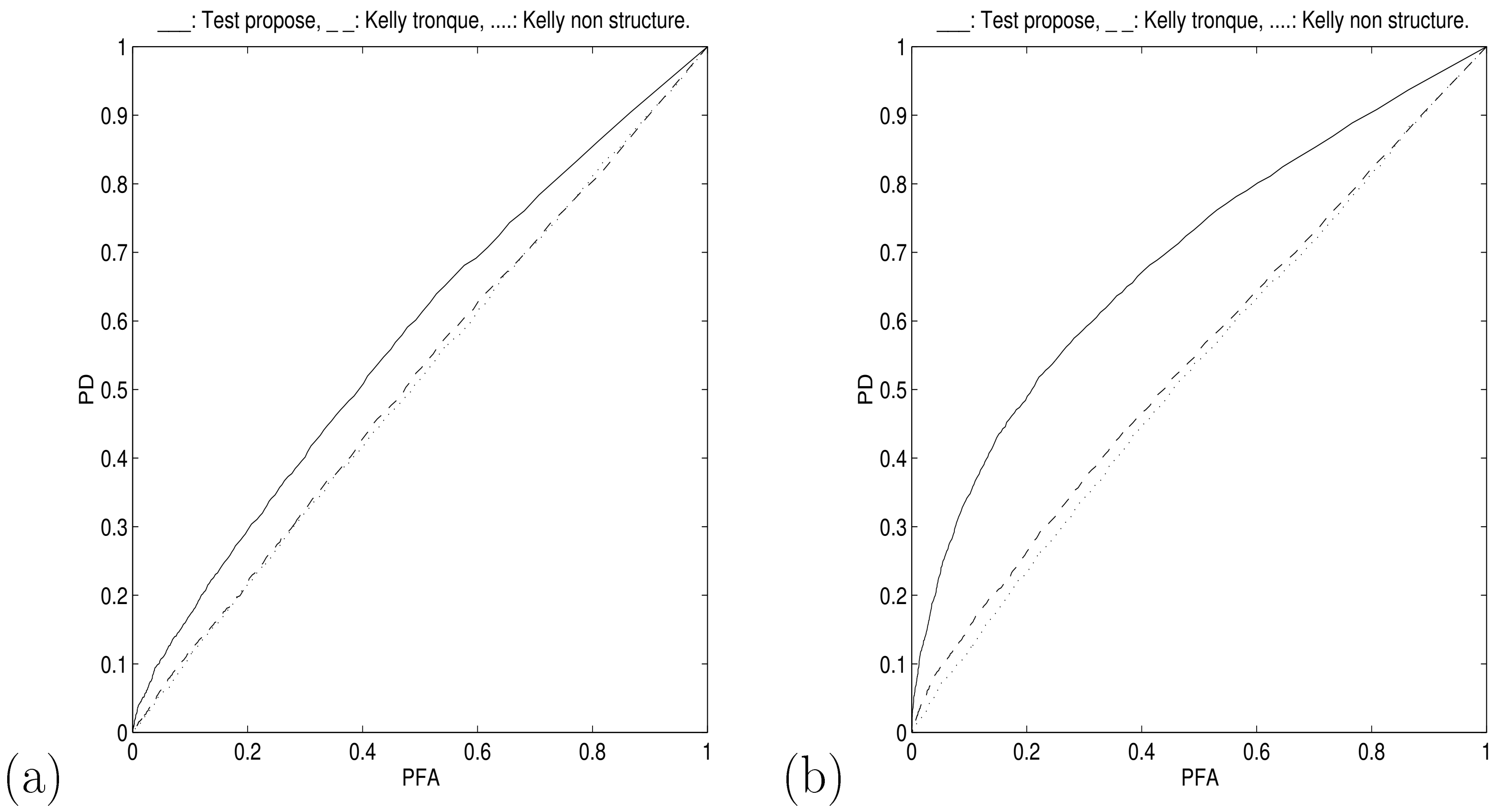


Figure 4: ROC comparisons for the case of a known target at known location on boundary between lowpass clutter region and a white clutter region with unknown power. Note B-S performs slightly better than GLRT3.



Single target at **unknown** location. ROC curves for Kelly's GLRT (structured and unstructured) and new max invariant test: (a) SNR = 0dB, (b) SNR = 20dB (From Hero&Guillouet:ICIP97).

Extensions

- Sensitivity to boundary segmentation errors
- Pose uncertainty: target subspace detection
- Polarimetric SAR
- Multiple regions: b-block structured maximal invariant
- Non-Gaussian clutter: elliptical family of densities, heavy-tailed mixture models
- **GLR via iterative optimization**: Proximal point-EM

Proximal-Point/EM Algorithms

Objective: maximize the convex (penalized) log likelihood

$$\ln f(Y;\ \underline{\theta}) - \beta P(\underline{\theta})$$

Challenges: ensuring fast convergence rate, low computation per iteration, stability, monotonicity.

<u>Our Solution</u>: (Chretien&Hero:ISIT98, Chretien&Hero:SIAM98)

Develop new class of stable, monotonic and rapidly convergent hybrid EM algorithms

Expectation Maximization Algorithm (Dempster, Laird&Rubin:JRSS78)

$$\underline{\theta}^{k+1} = \mathrm{argmax}_{\underline{\theta}} Q(\underline{\theta}|\underline{\theta}^k), \qquad k = 1, 2, \ldots$$

- $Q(\underline{\theta}|\underline{\theta}^k) = E[\ln f(X;\ \underline{\theta})|Y,\ \underline{\theta}^k]$
- X is more informative "complete data"

Advantages of EM:

- Stability: likelihood is guaranteed to increase in k
- Globally convergent for convex likelihood:
- Likelihood typically increases to $> 95\%$ of maximum within few iterations

Disadvantage of EM: Slow asymptotic convergence $\Rightarrow$ SAGE

Proximal Point Algorithm (PPA) (Rockafellar:SIAM76)

$$\underline{\theta}^{k+1} = \text{argmax}_{\underline{\theta}}\{\ln f(Y;\underline{\theta}) - \lambda_k \|\underline{\theta} - \underline{\theta}^k\|^2\}, \qquad k = 1, 2, \ldots$$

PPA with Kullback Penalty (Chretien&Hero:SIAM98)

$$\underline{\theta}^{k+1} = \text{argmax}_{\underline{\theta}}\left\{\ln f(Y;\underline{\theta}) - \lambda_k K(\underline{\theta},\ \underline{\theta}^k|Y)\right\}$$

- $K(\underline{\theta},\ \underline{\theta}^k|Y) = E\left[\ln \frac{f(X|Y;\underline{\theta})}{f(X|Y;\underline{\theta}^k)}\right]$
- $\{\lambda_k\} > 0$ sequence of relaxation parameters

$$\lambda_k > 0, \qquad \text{and} \qquad \lambda_k \to 0$$

Computational form of Kullback-PPA:

$$\underline{\theta}^{k+1} = \text{argmax}_{\underline{\theta}} \left\{ (1-\lambda_k) \ln f(Y;\underline{\theta}) + \lambda_k Q(\underline{\theta} \mid \underline{\theta}^k) \right\}, \qquad k = 1, 2, \ldots$$

Properties: (Chretien&Hero:SIAM98)

- Kullback-PPA has monotone likelihood property for any $\lambda_k > 0$
- Bundle mechanism (conjugate subgradient) (LeMarechal:75) can be applied for non-differentiable $f(Y;\underline{\theta})$, $Q(\underline{\theta};\underline{\theta}^k)$
- Obtain superlinear convergence rate for differentiable case
- Under local quadratic approximation to $\ln f(Y;\underline{\theta})$ Kullback-PPA becomes hybrid EM/Newton algorithm
- Kullback-PPA generalizes to coordinatewise optimization: hybrid SAGE/Newton

Example: Maximum likelihood deblurring

$$\underline{Y} = \mathbf{A}\underline{\theta} + \underline{n}_y \quad (n \text{ data points})$$

* $\mathbf{A}$: known blurring matrix

* $\underline{n}_y$: Poisson noise with covariance

$$\mathrm{cov}_{\underline{\theta}}\,(\mathrm{diag}[\mathbf{A}\underline{\theta} + \underline{r}])$$

* $\underline{\theta}$: unknown vector (image)

$\Rightarrow$ Compare EM, SAGE, NCA, Prox-SAGE

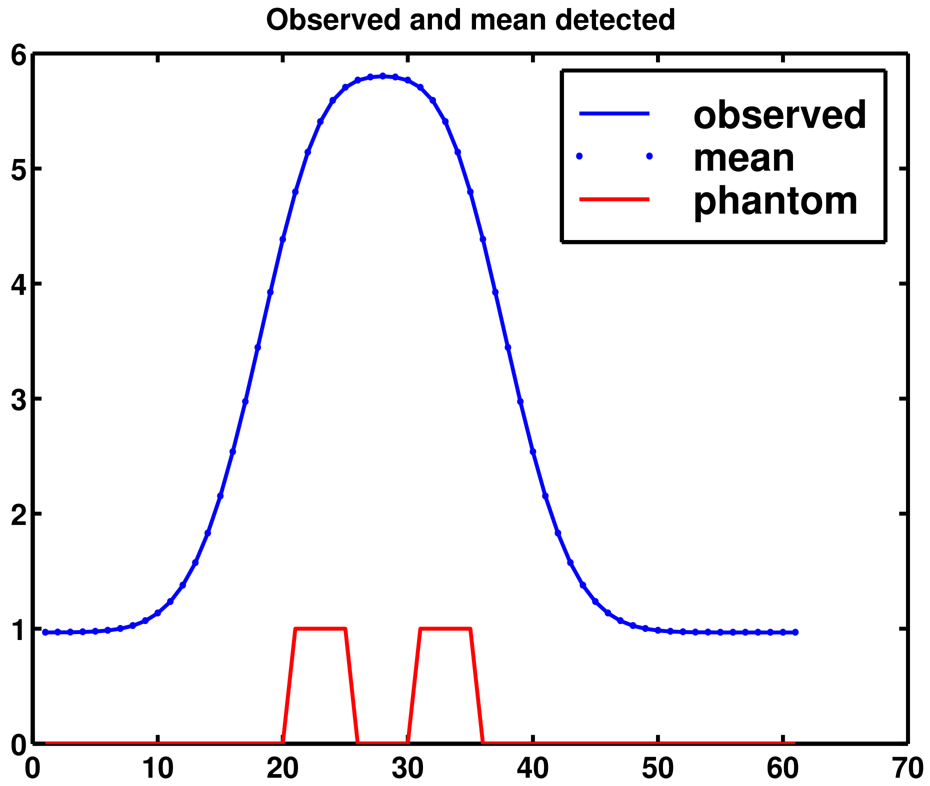


Figure 6: True bilevel signal and blurred signal

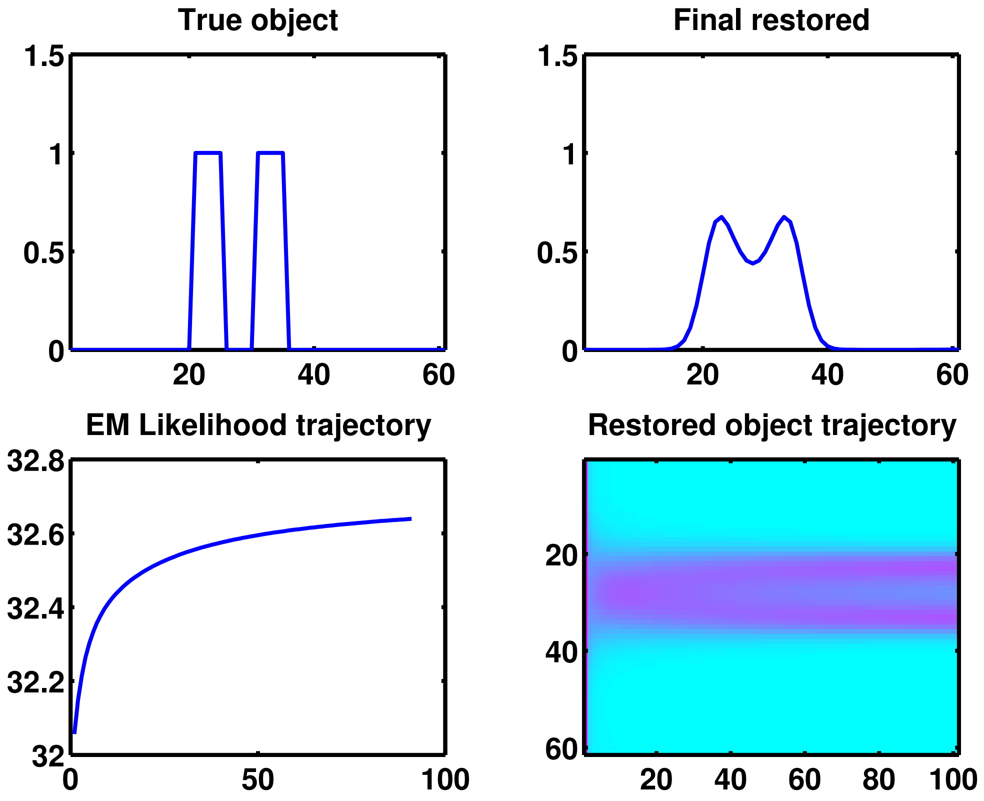


Figure 7: EM restoration

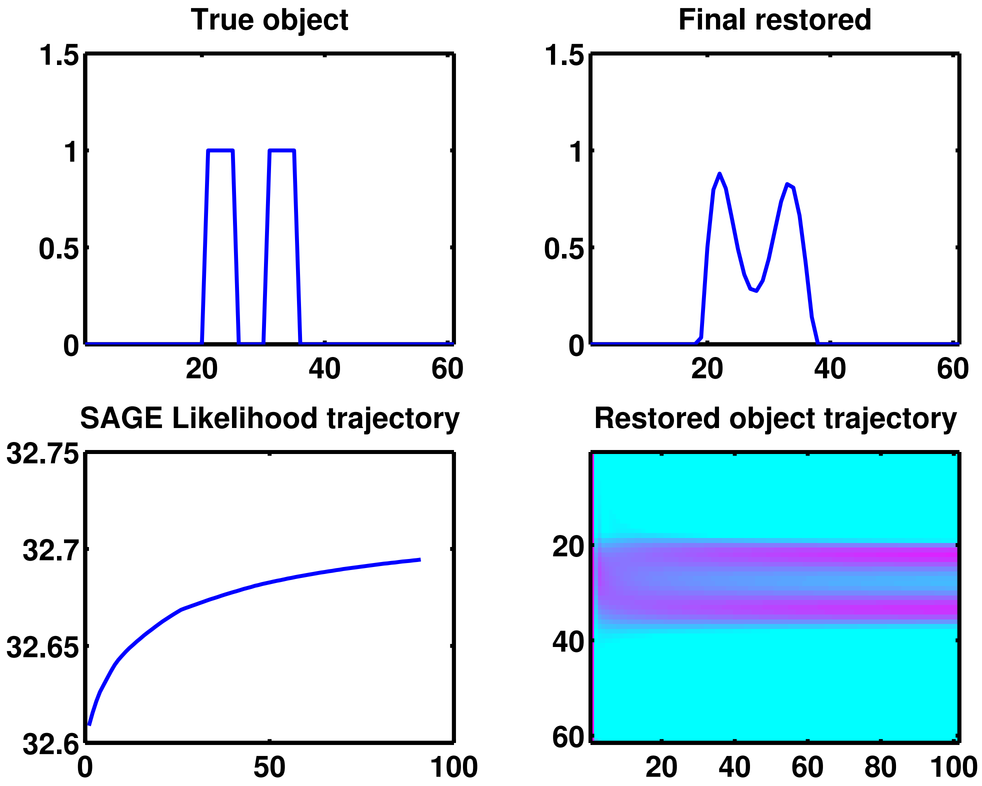


Figure 8: SAGE restoration

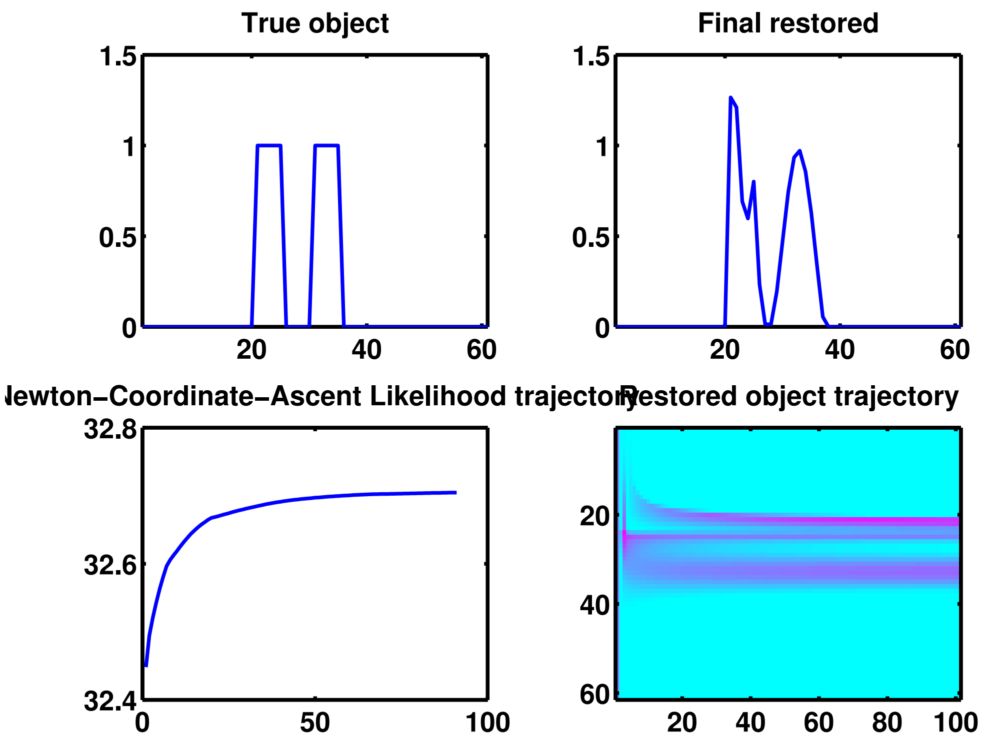


Figure 9: Netwon Coordinate Ascent reconstruction

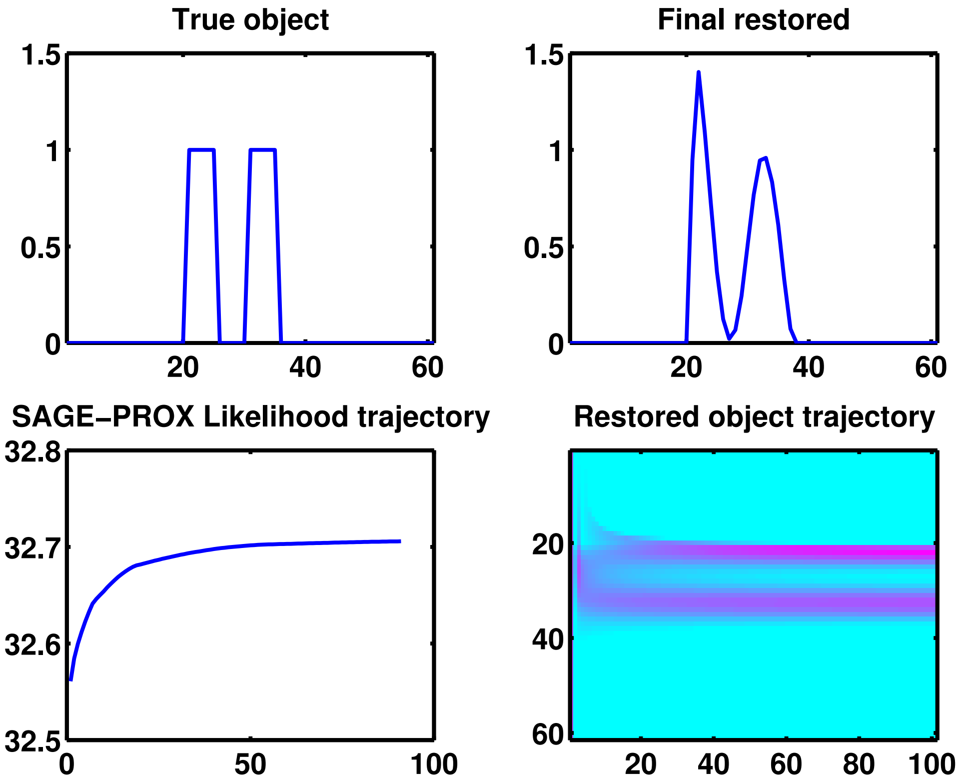


Figure 10: Prox-SAGE reconstruction

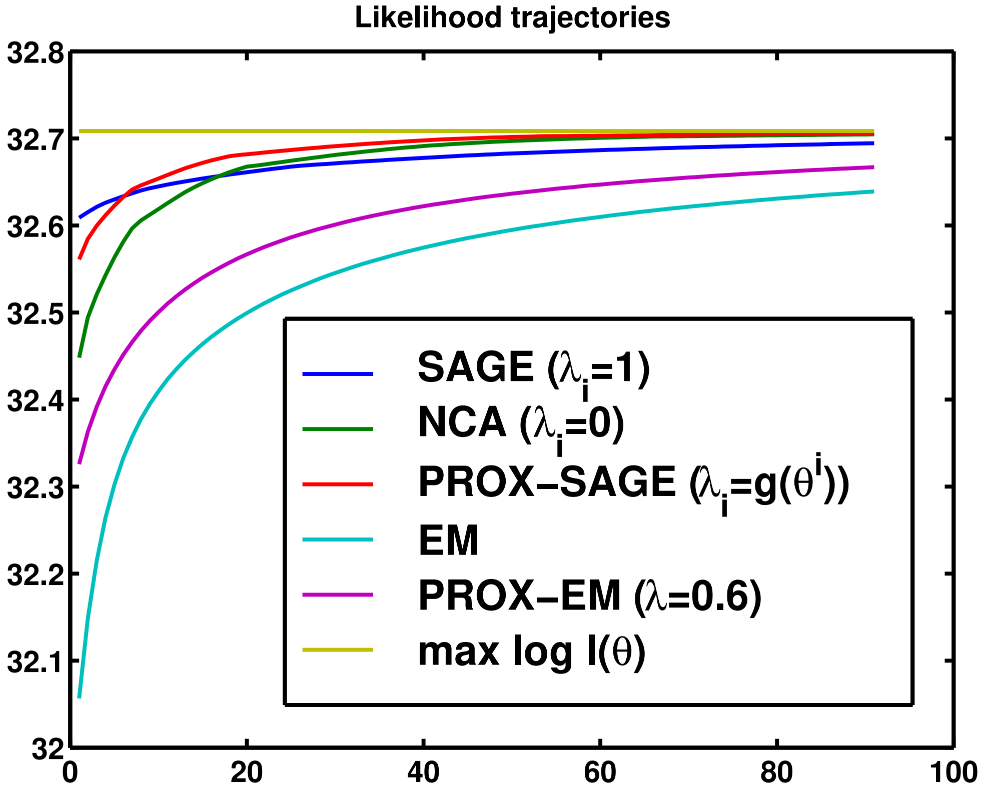
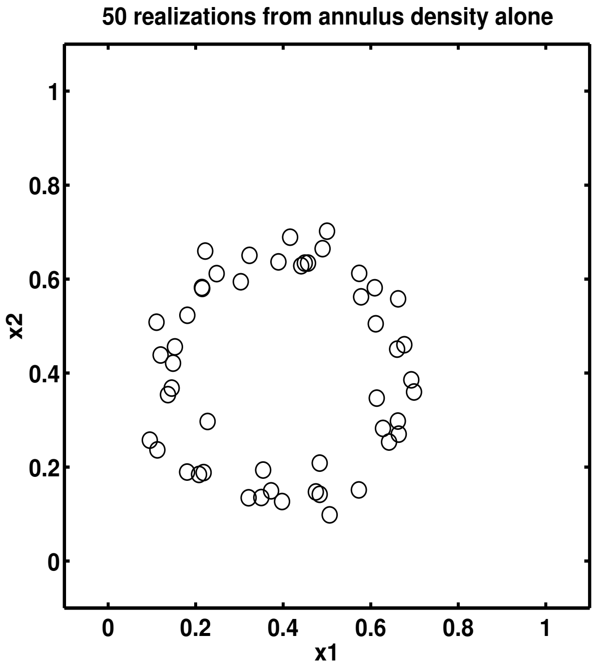
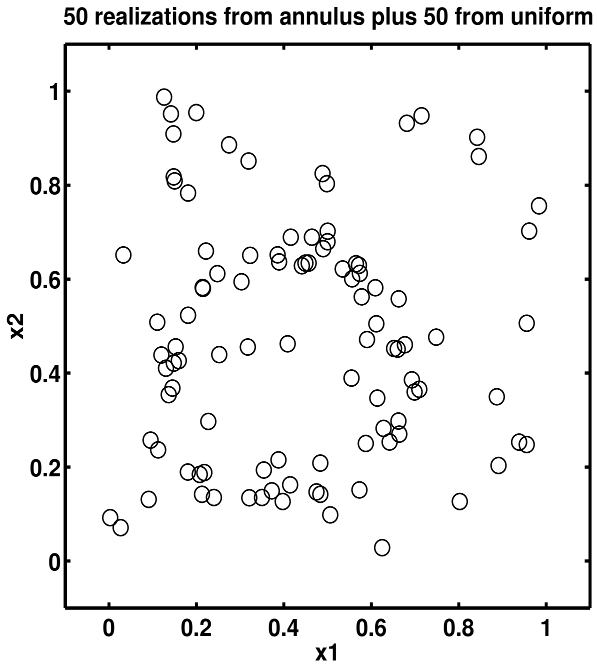


Figure 11: Likelihood trajectory comparisons

Theory and Application of k-shortest Graphs

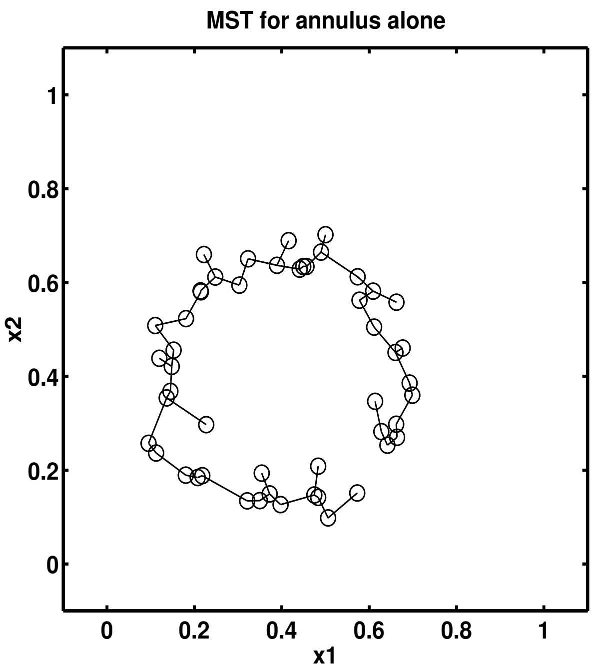


(a)

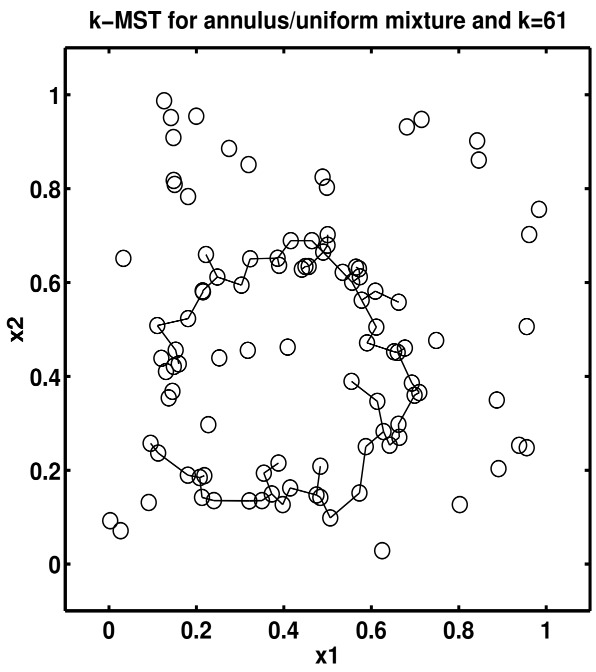


(b)

Definition: MST over $\{x_i\}_{i=1}^n$ in $\mathbb{R}^n$ is **shortest tree spanning the** n **points**



(a)



(b)

Definition: k-shortest MST over $\{x_i\}_{i=1}^n$ in $\mathbb{R}^n$ is **shortest MST spanning any k points**

Accomplishments (Hero&Michel:IT98, Hero&Michel:SPIE98):

- Developed polynomial-time greedy approximations to k-MST.
- Established asymptotic properties of k-MST length function (Renyi entropy).
- Established quantitative robustness of k-MST via influence curve analysis (multi-dimensional rank order statistics).

Applications:

- Robust entropy estimation - extension of Ahmad&Lin:IT76, Vasicek:JRSS76, Hall&Morton:AnnInstStatMath93
- Clustering/classification with false alarms - extension of Toussaint:PR80, Zahn:TC71, Banks:Interface92
- Robust image registration and two-sample matching - extension of Viola&Wells:MIT95, Friedman&Rafsky:AnnStat79)

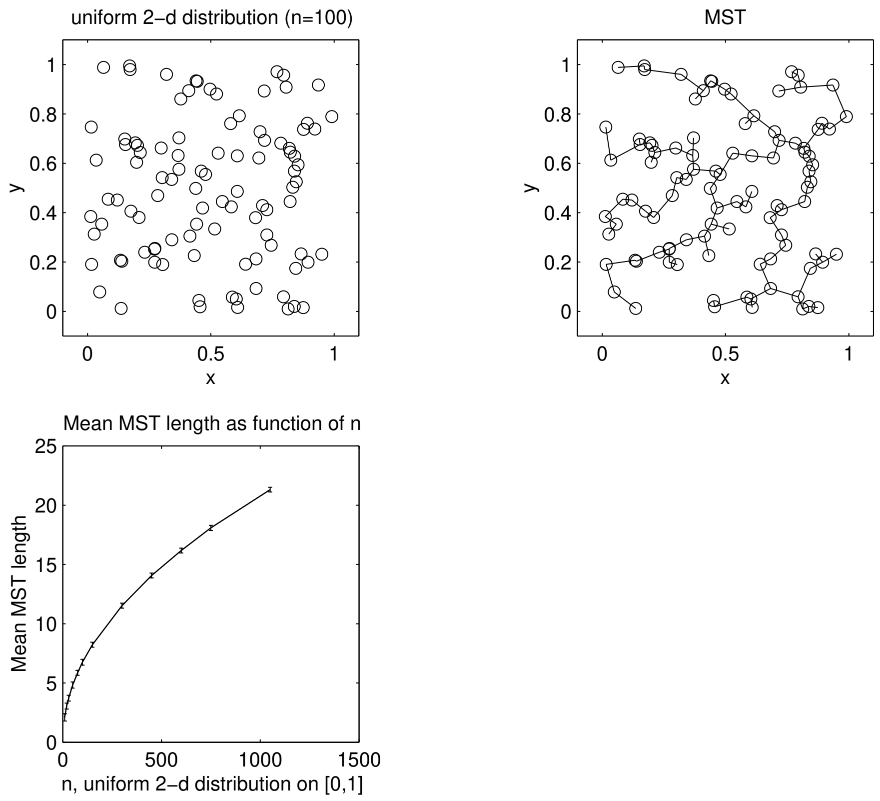


Figure 14: 2D Uniform sample study.

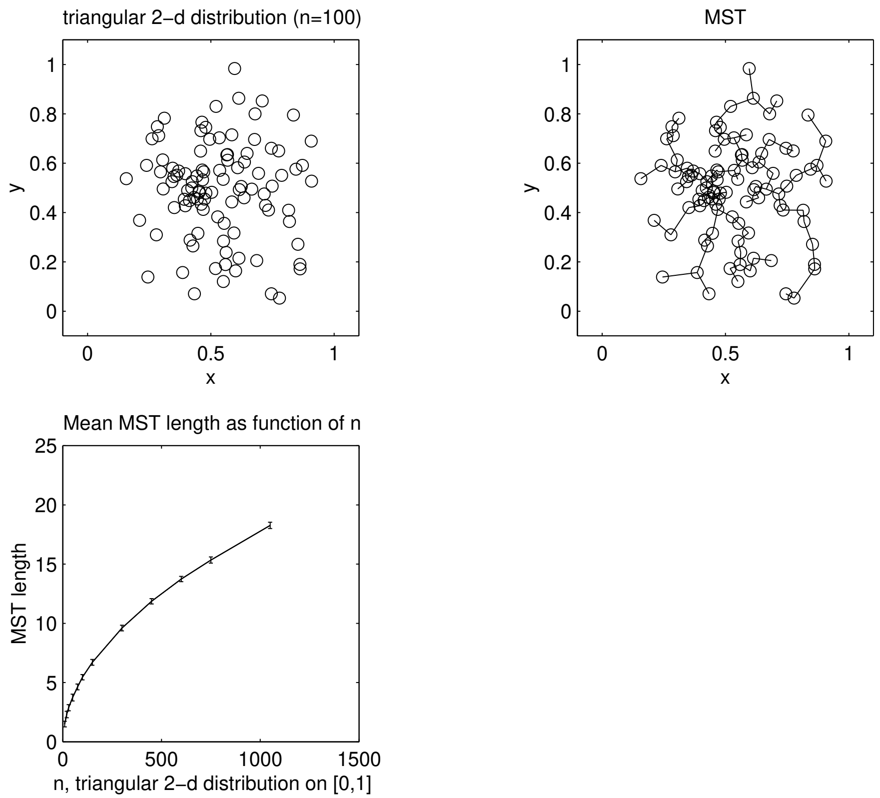


Figure 15: 2D Triangular sample study.

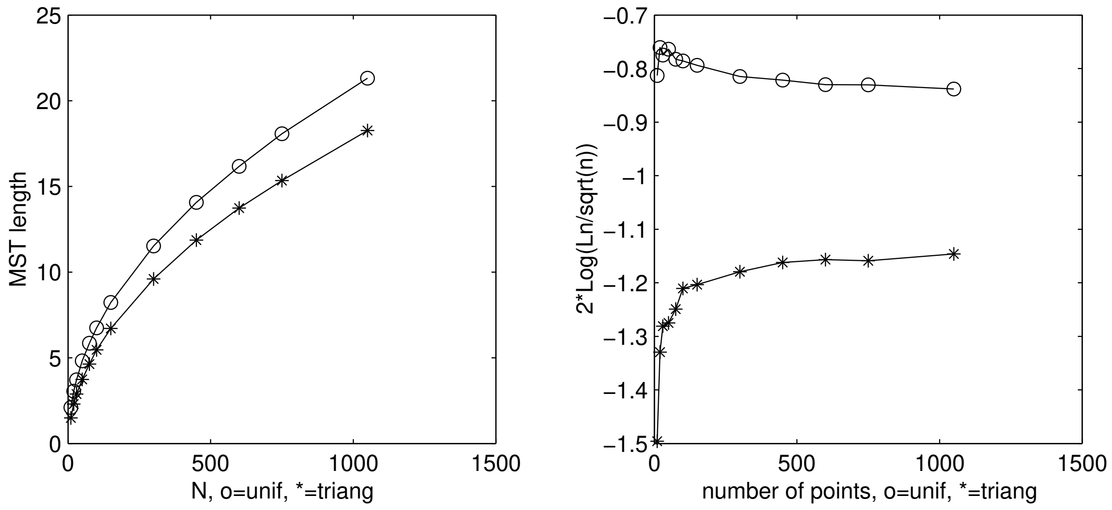


Figure 16: MST Length and log length comparisons.

Principal asymptotic result: extension of Steele:AnnAppProb1988, Redmond&Yukich:AnnAppProb94:

Theorem (Hero&Michel:IT98): For any $\alpha \in (0,1]$ the sum $L_{n,k} = \Sigma\,|e|^\gamma$ of the power-weighted edges of the k-MST satisfies

$$\frac{1}{1-\nu} \ln\left(L_{n,\lfloor \alpha n \rfloor}/(\lfloor \alpha n \rfloor)^\nu\right) \;\rightarrow\; \min_{A:P(A)\geq\alpha} R_\nu(X|A) + \ln \beta_{d,\gamma} \quad (a.s.)$$

where $R_\nu(X|A)$ is the Renyi entropy of the conditional distribution $f(x|X \in A)$ of order $\nu = (d-\gamma)/d$

$$R_\nu(X|A) = \frac{1}{1-\nu}\ \ln \int f^\nu(x|X \in A)dx.$$

Application to Image Registration

Minimum entropy property:

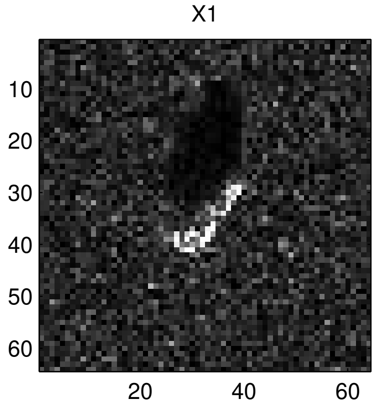
$$R_\nu([X,Y]) \geq R_\nu(X)$$

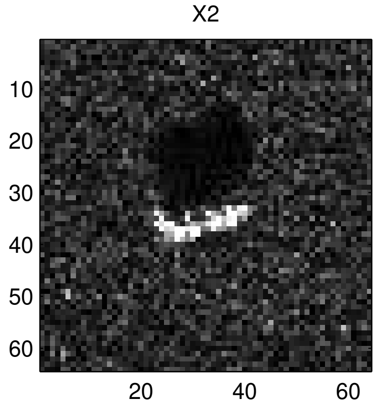
Two step registration process:

1. Select α, $\nu \in [0,1]$
2. Implement outlier rejection on X, Y via k-MST
3. Entropy minimization via MST over coordinate transformations on Y

Applications:

1. **MSTAR targets in clutter**
2. MRI brain image registration



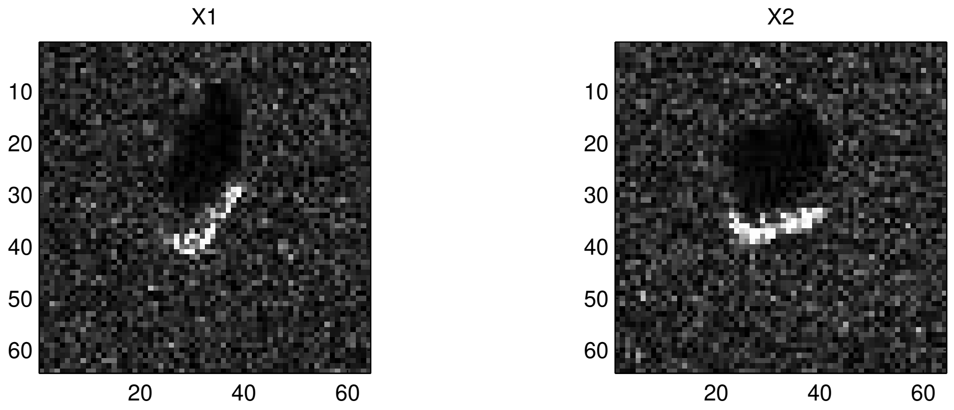


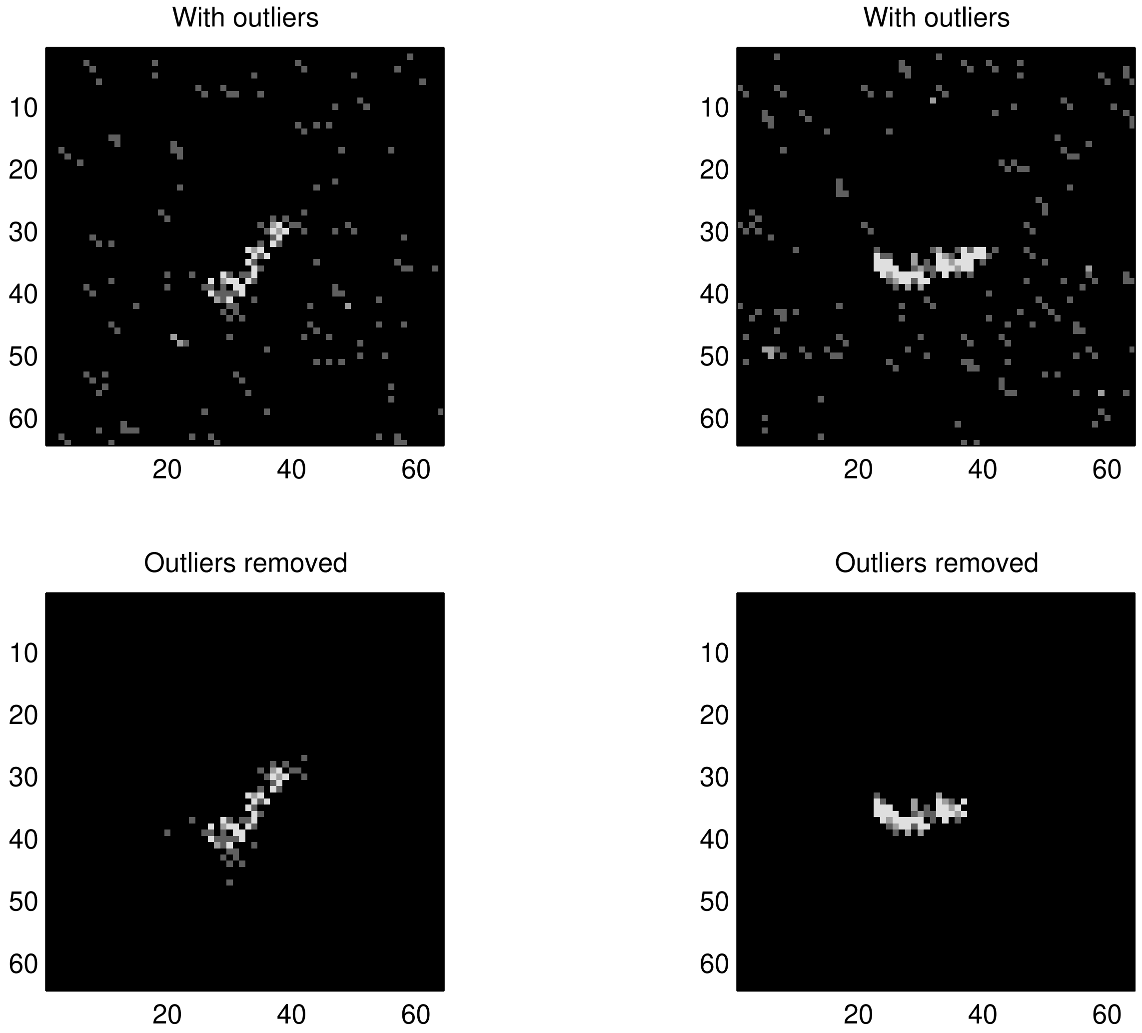
BTR-70 Personnel Carrier - Quadrant 3 ($180° - 270°$)

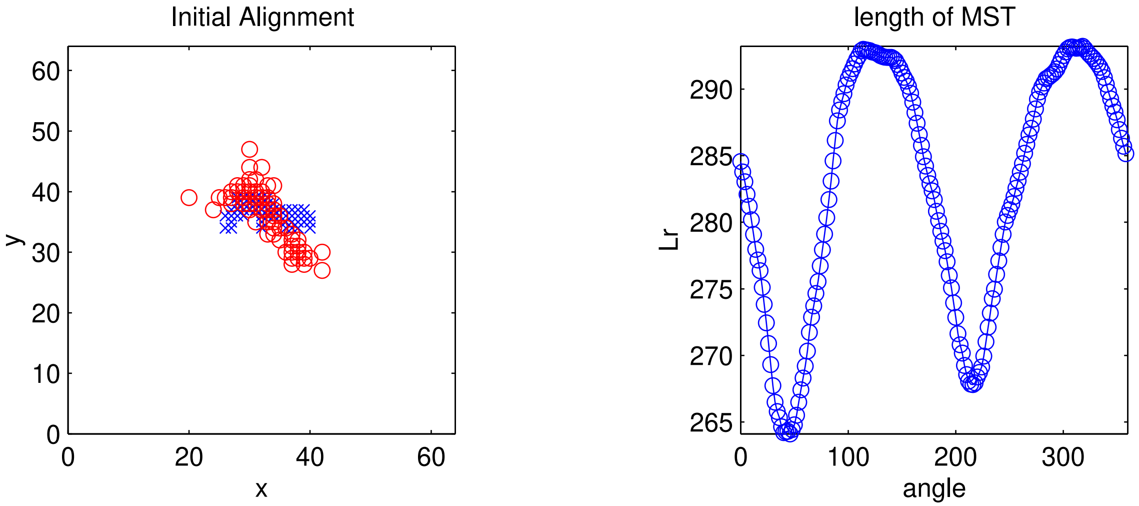
Targets at 214° and 255° (spotlight mode)

Depression angle 14.95° and 15.12°

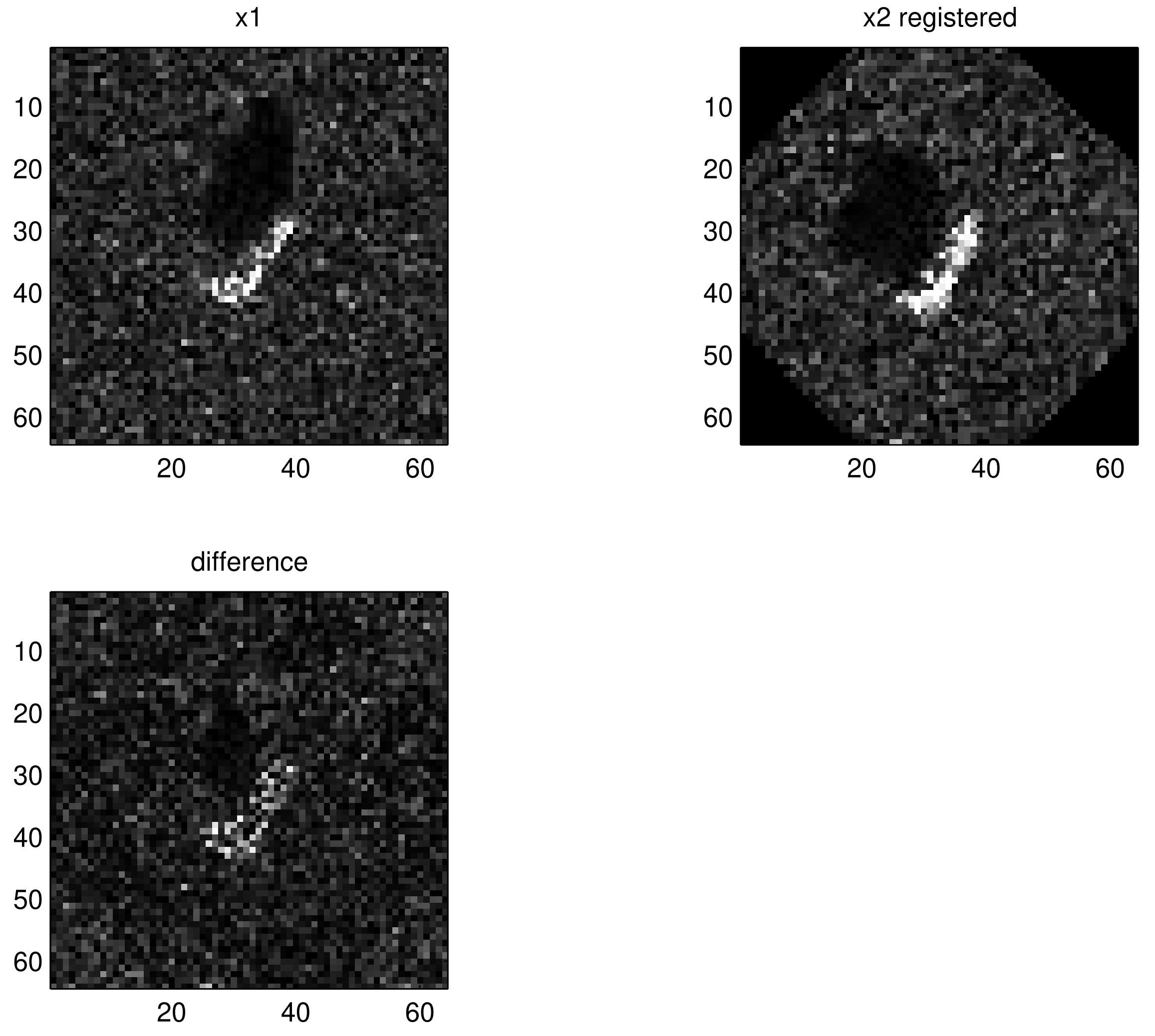
Images from MBVLab's MSTAR Radar Image Gallery
(http://www.mbvlab.wpafb.af.mil/public/MBVDATA/img_intro.htm)

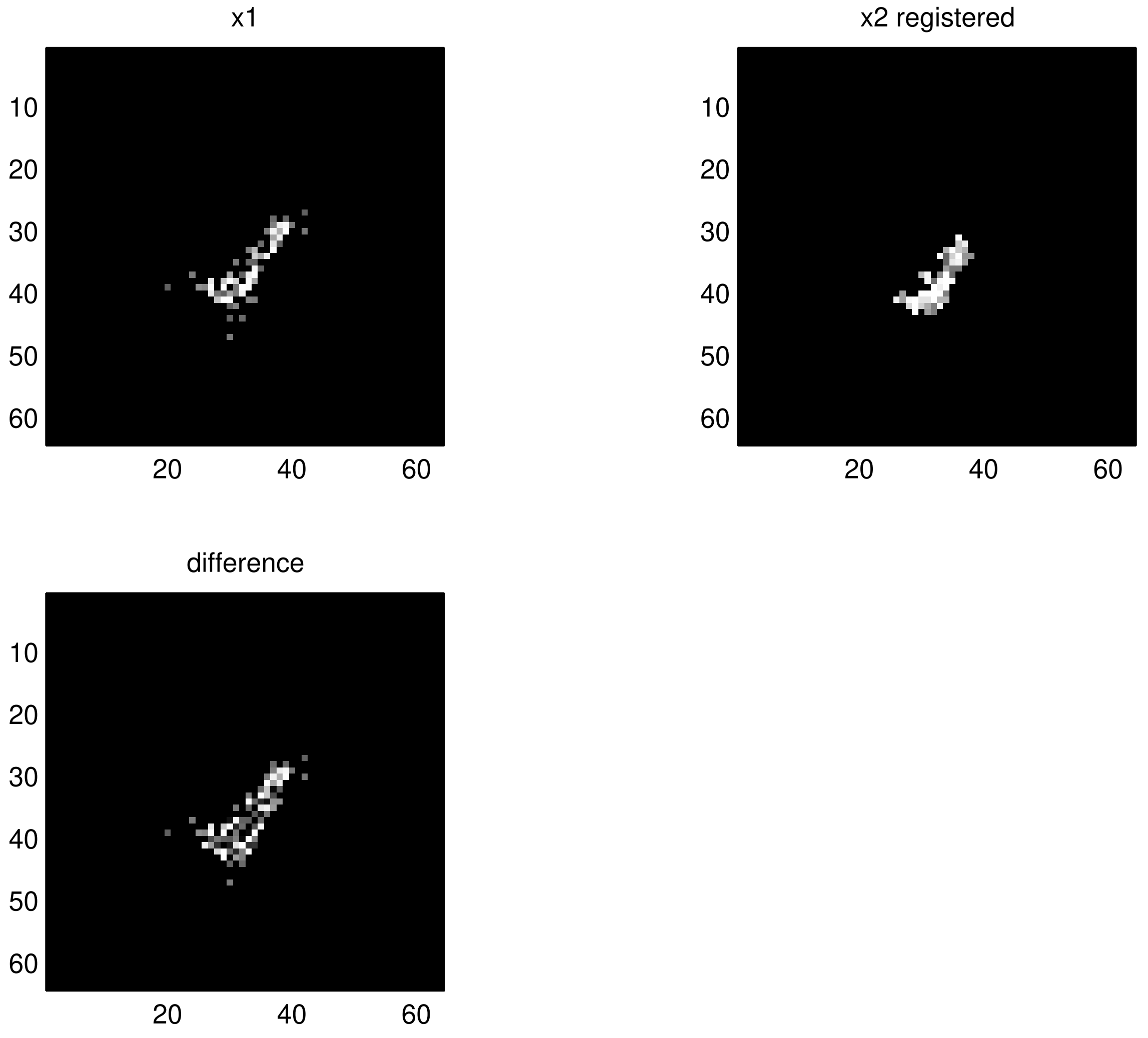

X1
X2
10
20
30
40
50
60
20
40
60


With outliers
With outliers
Outliers removed
Outliers removed
10
20
30
40
50
60
20
40
60



Initial Alignment
60
50
40
30
20
10
0
y
0
20
40
60
x
length of MST
290
285
280
275
270
265
Lr
0
100
200
300
angle


x1
10
20
30
40
50
60
20
40
60
x2 registered
10
20
30
40
50
60
20
40
60
difference
10
20
30
40
50
60
20
40
60


x1
10
20
30
40
50
60
20
40
60
x2 registered
10
20
30
40
50
60
20
40
60
difference
10
20
30
40
50
60
20
40
60

Future Work:

- Selection of α for outlier rejection, e.g. linear prediction criteria.
- Quantitative comparison to histogram, RBF, kernel estimators of entropy.
- Convergence analysis of k-MST for spatially correlated feature vectors.

*

References

[1] A. Hero and C. Guillouet, "Robust detection of SAR/IR targets via invariance," in *IEEE Int. Conf. on Image Processing*, pp. 472–475, 1997.

[2] S. Chretien and A. Hero, "Acceleration of the EM algorithm via proximal point iterations," in *Proc. of IEEE Symposium on Information Theory*, MIT, Cambridge, August 1998.

[3] S. Chretien and A. Hero, "Generalized proximal point algorithms and bundle implementations," *SIAM Journ on Control and Optimization*, vol. Submitted, , Sept., 1998.

[4] S. Chretien and A. Hero, "Generalized proximal point algorithms and bundle implementations," Technical Report CSPL-313, Comm. and Sig. Proc. Lab. (CSPL), Dept. EECS, University of Michigan, Ann Arbor, Aug. 1998.

[5] A. Hero and O. Michel, "Robust estimation of point process intensity features using K-minimal spanning trees," in *Proc. of IEEE Symposium on Information Theory*, Ulm, Germany, July 1997.

[6] A. Hero and O. Michel, "Asymptotic theory of greedy approximations to minimal k-point random graphs," *IEEE Trans. on Inform. Theory*, Submitted for publication, Mar. 1998.

[7] A. Hero and O. Michel, "Robust entropy estimation strategies based on edge weighted random graphs," in *Meeting of Intl. Soc. for Optical Engin. (SPIE)*, San Diego, CA, July 1998.

[8] A. Hero and O. Michel, "Asymptotic theory of greedy approximations to minimal K-point random graphs," Technical Report CSPL-314, Communications and Signal Processing Laboratory (CSPL), Dept. EECS, The Univ. of Michigan, Ann Arbor MI 48109-2122, 1998.

[9] A. Hero, J. O'Neill, and W. Williams, "Moment matrices for recognition of spatial pattern in noisy images," in *IEEE Int. Conf. on Image Processing*, pp. 378–381, 1997.

[10] A. O. Hero, J. O'Neill, H. Hadinejad-Mehram, and W. J. Williams, "Moments and moment matrices for invariant classification of noise contaminated spatial patterns," Technical Report 307, Comm. and Sig. Proc. Lab. (CSPL), Dept. EECS, University of Michigan, Ann Arbor, Feb., 1997.