Multi-Modal Inverse Scattering for Detection and Classification Of General Concealed Targets:

Landmines, Targets Under Trees, Underground Facilities

Research Team:

Lawrence Carin, Leslie Collins, Qing Liu
Duke University

Waymond Scott and James McClellan
Georgia Institute of Technology

George Papanicolaou
Stanford University

Sponsors: DARPA (Dr. Doug Cochran), ARO (Dr. Russell Harmon)
• Sensing of landmines, targets under trees, underground structures and targets in a water channel are very distinct missions, although they fall under the general problem of sensing concealed targets in the presence of a complex, stochastic environment

• Rather than focusing on one of these areas, we exploit their inter-relationships to investigate the general concealed-target problem

• Particular examples will be investigated by connecting members of the MURI to appropriate members of the user community (e.g. landmines: Army Countermine Office, Ft. Belvoir, VA)

• Undertake a multi-modal (multi-sensor) approach to effect inversion, with insights from the evolving inversion used to refine/optimize the inter- and intra-sensor parameters

• Will exploit the fact that future (and current) DoD systems will rely increasingly on autonomous vehicles (e.g. multiple robots, UAVs and UUVs that can be positioned to optimize the sensor platform as the inversion is undertaken)
Multi-Modal Inversion of General Targets Embedded In Arbitrary Stochastic Layered Media

General Adaptive Algorithms and Phenomenological Insights

Landmines
Underground Structures
Concealed Ground Structures

Application-Specific Questions/Issues
Multiple Sensors

Multi-Sensor Physics-Based Constrained RDOF

Phenomenology from Forward Models

Direct Inversion

Reverse-Time Migration

General Nonlinear Inversion

Statistical Inversion

Adaptive Coarse-to-Fine Pruning

Adaptive HMM
Multi-sensor mutual information
Multi-sensor, adaptive Bayesian

Optimize Intra-Sensor Parameters
Optimize Inter-Sensor Parameters

Inversion Confidence

Inversion complete

Phenomenology from Forward Models
Key to Program Success: Close Coupling to User Community

- MURI team members have a track record of working together and with the responsible parties within the DoD (and contractors)

- To assure that the research addresses relevant issues, and is transitioned to the appropriate organizations, it is essential that the MURI team members work closely with the user community

- Close relationships exist with the DoD landmine community (NVESD), and similar relationships are planned for the TUT and underground-structures problems

- Research relevance assured through goal of processing data generated by the appropriate DoD community
Duke Team Presentations

8:50- 9:20 Lawrence Carin
“Optimal Experiments for Adaptive Sensing: Regression, Detection and Classification”

9:20-9:50 Leslie Collins
“Bayesian Adaptive Multi-Modality Processing for the Discrimination of Landmines”

9:50-10:20 Qing Liu
“Fast Forward Solvers for Electromagnetic and Elastic Scattering”

10:20-10:45 Break

10:45-11:15 George Papanicolaou
“Imaging in Clutter”

11:15-12:15 Waymond Scott and James McClellan
“Detection of Obscured Objects: Signal Processing & Modeling”

www.ee.duke.edu/~lcarin/MURI.html
Optimal Experiments for Adaptive Sensing: Regression, Detection and Classification

Lawrence Carin, Xuejun Liao, Shihao Yu and Yan Zhang
Department of Electrical & Computer Engineering
Duke University
Durham, NC 27708-0291

www.ee.duke.edu/~lcarin
Outline

• Optimal adaptive regression with application to EMI sensing of buried targets

• Optimal adaptive detection and classification with application to multi-aspect sensing (SAR, UUV, UAV)

• Optimal adaptive training for buried targets with application to UXO

• Future directions
Magnetization Tensor for EMI Sensing

• Magnetic fields of a rotated EMI dipole

\[
H_z(\Theta \mid r_s, \omega_s) = \frac{z - z_s}{[(r - r_s)^T(r - r_s)]^3}(r - r_s)^T U^T M(\omega_s) U(r - r_s)
\]

\[
M(\omega) = \text{diag}[m_{p0} + \sum_k \frac{\omega m_{pk}}{\omega - j\omega_{pk}}, m_{p0} + \sum_k \frac{\omega m_{pk}}{\omega - j\omega_{pk}}, m_{z0} + \sum_k \frac{\omega m_{zk}}{\omega - j\omega_{zk}}]
\]

• The target EMI dipole parameters to be inverted are

\[
\Theta = [m_{p0}, m_{pk}, \omega_{pk}, m_{z0}, m_{zk}, \omega_{zk}, x, y, z, \theta, \phi]^T
\]

• The sensor parameters that may be varied are

\[
p = [x_s, y_s, z_s, \omega_s]^T
\]
Mobile induction sensor
target dipole

x-y-z: sensor coordinate system
Measurement Model

• Set of $N$ measurement parameters and observations represented $(p_n, O_n)_{n=1,N}$

• The observation is modeled as the magnetic field plus additive WGN

$$O = H_z(\Theta | r_s, \omega_s) + G$$

• Given $N$ measurements $(p_n, O_n)_{n=1,N}$ the target model parameters are estimated as

$$\hat{\Theta} = \arg \min_{\Theta} g(\Theta) = \arg \min_{\Theta} \sum_{n=1}^{N} |H_z(\Theta | p_n) - O_n(p_n)|^2$$

subject to : $m_{p0}, m_{pk}, \omega_{pk}, m_{z0}, m_{zk}, \omega_{zk} \geq 0$
Statistical Characterization

- The error between the observation and the model is modeled as WGN, where we are assuming that the model is correct

\[ O - f(p; \Theta) = e(p, \Theta) = e_R(p, \Theta) + je_I(p, \Theta) \]

- The associated density function is

\[ p(e; \Theta) = \sqrt{\frac{\beta}{2\pi}} \exp\left(-\frac{1}{2} \beta e_R^2\right) \sqrt{\frac{\beta}{2\pi}} \exp\left(-\frac{1}{2} \beta e_I^2\right) = p(e_R; \Theta)p(e_I; \Theta) \]

- Associated Fisher information matrix (iid)

\[ J = \beta \sum_{n=1}^{N} \text{Re}\{[\nabla_\Theta f(p_n; \Theta)][\nabla_\Theta f(p_n; \Theta)]^H\} \]

- Assumption of additive WGN is analogous to linearization with arbitrary noise

\[ f_L(p; \Theta) \equiv f(p; \hat{\Theta}_N) + (\Theta - \hat{\Theta}_N)^T \nabla_\Theta f(p; \Theta) \bigg|_{\Theta = \hat{\Theta}} \]
Information Gain of New Measurement

- Measure of information in $N$ iid measurements

\[
q(\{p_1,\ldots, p_N\}) = |J(p_1,\ldots, p_N)| = \sum_{n=1}^{N} J^n(p_n)
\]

where

\[
J^n(p_n) = \beta \Re \left\{ [\nabla \phi f(p_n; \Theta)][\nabla \phi f(p_n; \Theta)]^H \right\}
\]

- Information gain by adding measurement $N+1$

\[
\delta_q(p) = \ln q(\{p_1,\ldots, p_N, p_{N+1}\}) - \ln q(\{p_1,\ldots, p_N\}) = \ln \left| I + \beta F^T B_N^{-1} F \right|
\]

\[
F = [F_R, F_I] = [\Re \{\nabla \phi f(p_{N+1}; \Theta)\}, \Im \{\nabla \phi f(p_{N+1}; \Theta)\}]
\]

\[
B_N = \sum_{n=1}^{N} J^n
\]

- Sensor parameters for measurement $N+1$

\[
p_{N+1} = \arg \max_p \delta_q(p) = \arg \max_p \ln \left| I + \beta F^H B_N^{-1} F \right|
\]
Accounting for Measurement Cost

- Previous theory chose new sensor parameters based on information gain alone

- Desirable to account as well for measurement “cost”

- Simple augmentation of cost function

\[ \psi(p) = \delta_q(p) - \mu c(p) \]

\[ = ln \left| I + \beta F^H B_N^{-1} F \right| - \mu \sqrt{(p_{N+1} - p_N)^T W^T W (p_{N+1} - p_N)} \]
Sensor-frequency used in the fixed grid

Sensor Frequency (Hz)

Iteration Number
source dipole
numbers: sensor dipole
optimal sensor-frequency from the search strategy

Sensor Frequency (Hz)

Iteration Number

Sensor Frequency (Hz)
Mean Squared Error (MSE) vs. Number of Data Points Used

- MSE with optimal sampling points
- MSE with fixed sampling points

The graph shows the mean squared error for different numbers of data points used. The error is significantly lower with optimal sampling points compared to fixed sampling points, especially as the number of data points increases.
No Constraints On Sensor Motion

digits = sensor positions, star = sources, angles = [87.9, 4.9] Deg.
optimal sensor-frequency as a function of iteration number
Constrained Sensor Motion

digits = sensor positions, star = sources, angles = [128.2, 7.1] Deg.

source dipole numbers: sensor dipole

$x$ position (meters) $y$ position (meters)
optimal sensor-frequency as a function of iteration number
Information-Gain Surface, Measurement One

Fisher information gain

y position (meters) 0 50 -50 -50 x position (meters)
Information-Gain Surface, Measurement Two
Information-Gain Surface, Measurement Three
Information-Gain Surface, Measurement Four
Information-Gain Surface, Measurement Five
true dipole, dipole fitted from all selected data, and dipole fitted from the initial single datum

<table>
<thead>
<tr>
<th></th>
<th>( m_{p0} )</th>
<th>( m_{p1} )</th>
<th>( \omega_{p1} )</th>
<th>( m_{z0} )</th>
<th>( m_{z1} )</th>
<th>( \omega_{z1} )</th>
<th>( x_n )</th>
<th>( y_n )</th>
<th>( z_n )</th>
<th>( \theta )</th>
<th>( \phi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>true dipole 1</td>
<td>20.29</td>
<td>33.77</td>
<td>498.05</td>
<td>73.60</td>
<td>23.51</td>
<td>679.38</td>
<td>-5.00</td>
<td>-5.00</td>
<td>10.00</td>
<td>1.53</td>
<td>0.09</td>
</tr>
<tr>
<td>true dipole 2</td>
<td>28.36</td>
<td>44.61</td>
<td>440.11</td>
<td>41.62</td>
<td>82.48</td>
<td>767.61</td>
<td>-3.00</td>
<td>-5.00</td>
<td>10.00</td>
<td>1.53</td>
<td>0.09</td>
</tr>
<tr>
<td>last fit. dipole 1</td>
<td>28.36</td>
<td>44.61</td>
<td>440.11</td>
<td>41.62</td>
<td>82.48</td>
<td>767.61</td>
<td>-3.00</td>
<td>-5.00</td>
<td>10.00</td>
<td>1.53</td>
<td>0.09</td>
</tr>
<tr>
<td>last fit. dipole 2</td>
<td>20.29</td>
<td>33.77</td>
<td>498.05</td>
<td>73.60</td>
<td>23.51</td>
<td>679.38</td>
<td>-5.00</td>
<td>-5.00</td>
<td>10.00</td>
<td>1.53</td>
<td>0.09</td>
</tr>
<tr>
<td>last fit. dipole 3</td>
<td><strong>0.00</strong></td>
<td><strong>21.91</strong></td>
<td><strong>543.54</strong></td>
<td><strong>10.72</strong></td>
<td><strong>1268.05</strong></td>
<td><strong>-8.18</strong></td>
<td><strong>0.00</strong></td>
<td><strong>1.53</strong></td>
<td><strong>0.09</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ini fit. dipole 1</td>
<td>54.92</td>
<td>58.49</td>
<td>772.05</td>
<td>48.61</td>
<td>3.41</td>
<td>498.53</td>
<td>-5.25</td>
<td>-8.16</td>
<td>4.66</td>
<td>2.14</td>
<td>2.37</td>
</tr>
<tr>
<td>ini fit. dipole 2</td>
<td>0.90</td>
<td>46.89</td>
<td>360.33</td>
<td>72.39</td>
<td>51.27</td>
<td>536.37</td>
<td>-2.81</td>
<td>8.87</td>
<td>2.14</td>
<td>2.37</td>
<td></td>
</tr>
<tr>
<td>ini fit. dipole 3</td>
<td>38.34</td>
<td>13.77</td>
<td>417.47</td>
<td>37.52</td>
<td>9.54</td>
<td>434.88</td>
<td>-3.37</td>
<td>7.80</td>
<td>2.14</td>
<td>2.37</td>
<td></td>
</tr>
</tbody>
</table>
Outline

• Optimal adaptive regression with application to EMI sensing of buried targets

• Optimal adaptive detection and classification with application to multi-aspect sensing (SAR, UUV, UAV)

• Optimal adaptive training for buried targets with application to UXO

• Future directions
• Scattering Data Can be Segmented into Angular Bins Characterized by particular physics
• Each such angular range termed a state (S1, S2, ..., SN)
Hidden Markov Models

\[ \pi_1 \quad \pi_2 \quad \pi_3 \]

\[
\begin{align*}
\pi_1 \quad &\quad \pi_2 \quad &\quad \pi_3 \\
S_1 \quad &\quad S_2 \quad &\quad S_3
\end{align*}
\]

\[
\begin{align*}
a_{11} \quad &a_{13} \quad &a_{33} \\
S_1 \quad &S_2 \quad &S_3
\end{align*}
\]
Theory of Optimal Experiments

• In previous HMM classification studies the path of the sensor was fixed, with uniform sampling spatially.

• For a UUV/UAV, one may be able to adaptively control sensor path.

• Question: Can we optimize the UUV/UAV path to most efficiently classify a submerged target (mine)?

• Employ the theory of optimal experiments.

• Assume that we wish to distinguish between \( K \) targets, and we perform \( t \) measurements, or observations \( O_t = \{O_1, \ldots, O_t\} \).

• ML classification: Choose target \( i \) if \( p(O_i | T_i) > p(O_i | T_k), \forall T_k \neq T_i \).
Theory of Optimal Experiments - 2

• Assume that we have performed $t$ measurements, at which we have measured, $O_t = \{O_1, \ldots, O_t\}$. The measurements are performed at the relative angles

$$\Delta \theta_2, \ldots, \Delta \theta_t$$

• Note: Since we don’t know the target orientation, we can only control the relative sensor angles with respect to the target

• Assume we have the target-dependent statistical models

$$p(O_1, \ldots, O_t \mid \Delta \theta_2, \ldots, \Delta \theta_t, T_k)$$

for example with an HMM

• At what relative angle $\Delta \theta_{t+1}$ should we perform measurement $t+1$, to optimize target classification?
Theory of Optimal Experiments - 3

• Consider the probabilities \( p(T_k \mid O_1, \ldots, O_t, \Delta \theta_2, \ldots, \Delta \theta_t) \) for all targets

• Idea: Choose the \( t+1 \) measurement to minimize the entropy characteristic of these density functions

\[
H_t(\Delta \theta_2, \ldots, \Delta \theta_t) = -\sum_{k=1}^{K} p(T_k \mid \Delta \theta_2, \ldots, \Delta \theta_t, O_1, \ldots, O_t) \log p(T_k \mid \Delta \theta_2, \ldots, \Delta \theta_t, O_1, \ldots, O_t)
\]

• Issue: When considering different \( \Delta \theta_{t+1} \) we do not actually have access to \( O_{t+1} \)

• Solution: We compute the expected entropy, integrating over our density function for \( O_{t+1} \)

• Key: With HMM we need not assume independent measurements + HMM efficient
We can perform this efficiently utilizing our previous work with HMMs.
## Confusion Matrix (5 points)

### Uniform Sampling: 5°

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>0.7667</td>
<td>0.0139</td>
<td>0.0333</td>
<td>0.0528</td>
<td>0.0133</td>
</tr>
<tr>
<td>$t_2$</td>
<td>0.0694</td>
<td>0.7278</td>
<td>0.1111</td>
<td>0</td>
<td>0.0917</td>
</tr>
<tr>
<td>$t_3$</td>
<td>0.0194</td>
<td>0.1861</td>
<td>0.7083</td>
<td>0.0111</td>
<td>0.0750</td>
</tr>
<tr>
<td>$t_4$</td>
<td>0.0417</td>
<td>0</td>
<td>0.0222</td>
<td>0.9083</td>
<td>0.0278</td>
</tr>
<tr>
<td>$t_5$</td>
<td>0.0389</td>
<td>0.0972</td>
<td>0.0500</td>
<td>0.0389</td>
<td>0.7750</td>
</tr>
</tbody>
</table>

**Ave = 77.72%**

### Uniform Sampling: 45°

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$t_4$</th>
<th>$t_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>0.9417</td>
<td>0.0278</td>
<td>0.0222</td>
<td>0.0056</td>
<td>0.0028</td>
</tr>
<tr>
<td>$t_2$</td>
<td>0.0722</td>
<td>0.7917</td>
<td>0.1194</td>
<td>0</td>
<td>0.0167</td>
</tr>
<tr>
<td>$t_3$</td>
<td>0.0333</td>
<td>0.2333</td>
<td>0.7167</td>
<td>0</td>
<td>0.0167</td>
</tr>
<tr>
<td>$t_4$</td>
<td>0.0028</td>
<td>0</td>
<td>0</td>
<td>0.9556</td>
<td>0.0417</td>
</tr>
<tr>
<td>$t_5$</td>
<td>0.0056</td>
<td>0.0083</td>
<td>0.0111</td>
<td>0.0361</td>
<td>0.9389</td>
</tr>
</tbody>
</table>

**Ave= 86.89%**
Optimal Sampling:

<table>
<thead>
<tr>
<th></th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_3$</th>
<th>$T_4$</th>
<th>$T_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$</td>
<td>0.9750</td>
<td>0.0139</td>
<td>0.0028</td>
<td>0</td>
<td>0.0083</td>
</tr>
<tr>
<td>$T_2$</td>
<td>0.0333</td>
<td>0.8806</td>
<td>0.0833</td>
<td>0</td>
<td>0.0028</td>
</tr>
<tr>
<td>$T_3$</td>
<td>0.0083</td>
<td>0.1333</td>
<td>0.8444</td>
<td>0</td>
<td>0.0139</td>
</tr>
<tr>
<td>$T_4$</td>
<td>0.0056</td>
<td>0</td>
<td>0</td>
<td>0.9639</td>
<td>0.0306</td>
</tr>
<tr>
<td>$T_5$</td>
<td>0.0028</td>
<td>0.0028</td>
<td>0.0056</td>
<td>0.0083</td>
<td>0.9806</td>
</tr>
</tbody>
</table>

Ave= 92.89%
Objective Function

![Graph showing the relationship between sample rate and delta entropy with a peak at Δθ = 89. The graph has a logarithmic scale on the y-axis, ranging from $10^{-29}$ to $1.2$. The x-axis represents the sample rate, ranging from 0 to 90.]
Sample positions=93 141 189 237 285, delta=0 48 48 48 48
Number of Observations

Sample positions=93 183 247 252 273, delta=0 90 64 5 21

- Targ 1
- Targ 2
- Targ 3
- Targ 4
- Targ 5
Optimal Multi-Aspect Target Detection

- Previous results considered optimal sensing for *classification* problem

- We assumed knowledge of HMMs for each of the targets

- Often we have HMMs for known targets but may come across objects we haven’t seen previously

- Can we extend theory of optimal experiments to case for which we only have HMMs for a subset of objects we may encounter?

- Relevant for target detection
Optimal Multi-Aspect Target Detection - 2

• Assume that we have performed $t$ measurements, at which we have measured, 
  $O_t = \{O_1, \ldots, O_t\}$. The measurements are performed at the relative angles 
  $\Delta \theta_2, \ldots, \Delta \theta_t$

• The log likelihood of these measurements, for the known target $T$ is

  $\log p(O_1, O_2, \ldots, O_t | \Delta \theta_1, \Delta \theta_2, \ldots, \Delta \theta_t, T)$

• We perform the next measurement at the change of angle $\Delta \theta_{t+1}$ that maximizes the expected increase in log likelihood

  $\Delta \theta_{t+1} = \arg \max_{\Delta \theta_{t+1}} E_{O_{t+1}|O_1, \ldots, O_t, \Delta \theta_1, \ldots, \Delta \theta_t, \Delta \theta_{t+1}} [\log p(O_1, \ldots, O_t, O_{t+1} | \Delta \theta_1, \ldots, \Delta \theta_t, \Delta \theta_{t+1}, T)]$

• This corresponds to minimizing the entropy (uncertainty) of whether the target under interrogation corresponds to target $T$
NSWC Data: 6 Targets 40-80 kHz

Bullet

Plastic Cone

55 Gal. Drum

Rock 1

Rock 2

Log
ROC curve between Target 5 and NSWC Targets

- Adaptive
- 5 degree
- 22 degree
Outline

• Optimal adaptive regression with application to EMI sensing of buried targets

• Optimal adaptive detection and classification with application to multi-aspect sensing (SAR, UUV, UAV)

• Optimal adaptive training for buried targets with application to UXO

• Future directions
Background

- UXO versus clutter: a classification problem
- Labeled data and unlabeled data
- Data is easy to observe, but excavation is very costly
- Reducing cost by the smart learning, no prior training data required
Step 1: Collect Data
**Step 2**: Choose Representative Signatures (Basis Functions), No Digging
Step 3: Dig First Hole to Learn Label
Step 4: Refine Classifier and then Dig Hole Two
Step 5: Refine Classifier and then Dig Hole Three, etc
Step 6: Obtain Set of Labeled Signatures and Complete Classifier Design after Information Gain Accrued By New Labels is Minimal
Three parts for building the classifier:

1. Select basis functions sequentially to build the *structure* of the classifier using all available data (unlabeled). The structure is designed to occupy the maximum information from the data set.

2. Sequentially select the most informative data to the classifier structure determined from Part 1, discover its label, and refine the information of the structure.

3. Estimate the weights of the classifier with labeled data selected from Part 2.
Classifier design: Kernel Machine

Kernel machine is a linear classifier in high dimensional space, it can produce complex decision boundaries while keep the generalization ability high.

\[ f_n(x) = \sum_{i=1}^{n} w_{n,i} K(c_i, x) + w_{n,0} = w_n^T \Phi_n(x) \]

Basis functions: \( \Phi_n(\cdot) = [1, K(c_1, \cdot), K(c_2, \cdot), \ldots, K(c_n, \cdot)]^T \)

Weights: \( w_n = [w_{n,0}, w_{n,1}, w_{n,2}, \ldots, w_{n,n}]^T \)

Kernel: \( K(c_i, x) \) Comparing the similarity between \( c_i \) and \( x \) in high dimensional space.

Relevant vectors: \( c_i, i = 1, \ldots, n \)
With the whole unlabeled data set \( X = \{x_1, x_2, \ldots, x_N\} \), basis functions are selected sequentially by Fisher Information Criteria (FIC) to capture the data characteristics.

Assume a linear kernel machine has \( n \) basis functions, then the Fisher information matrix for the weights is

\[
M_n = \sum_{i=1}^{N} \Phi_n(x_i)\Phi_n^T(x_i)
\]

\( M_n \) indicates the information from data set \( X \) for the current structure of the classifier. Its inverse is the Cramer-Rao Bound.
When a new kernel $\Phi_{n+1}(\bullet) = K(c_{n+1}, \bullet)$, $c_{n+1} \in X$ is added to basis functions, the new Fisher information matrix is

$$M_{n+1} = \sum_{i=1}^{N} \begin{bmatrix} \Phi_n(x_i) \\ \Phi_{n+1}(x_i) \end{bmatrix} \begin{bmatrix} \Phi_n^T(x_i) & \Phi_{n+1}(x_i) \end{bmatrix}$$

The information gain by adding a new kernel is defined as follows. It can also be viewed as the drop of CRB.

$$\Delta I_{\text{kernel}} = \ln \text{Det}(M_{n+1}) - \ln \text{Det}(M_n)$$
Classifier design: Selection of Basis Functions (3)

For the new kernel $K(c_{n+1}, \bullet)$, the information gain is equal to

$$\ln \Delta I_{\text{kernel}}(c_{n+1}) = \sum_{i=1}^{N} \Phi_{n+1}^{2}(x_i) - \left( \sum_{i=1}^{N} \Phi_{n+1}(x_i) \Phi_{n}^{T}(x_i) \right) M_{n}^{-1} \left( \sum_{i=1}^{N} \Phi_{n}(x_i) \Phi_{n+1}(x_i) \right)$$

Then the optimal new kernel is selected by a greedy search to maximize the information gain

$$c_{n+1} = \arg \max_{c \in X \setminus \{c_1, \ldots, c_n\}} \ln \Delta I_{\text{ker nel}}(c)$$
After a set of most informative basis functions $\Phi_{n_0}(\bullet)$ have been selected, in order to determine the classifier, the labeled data are required to estimate the weights.

Another FIC is utilized to assign a priority to each unlabeled data, and those more informative to the estimation are labeled first by digging them out. Therefore, the classifier is determined with substantial less excavating effort, which is the most costly part for UXO problem.
Labeled data selection is also a sequential process. Assuming we have a subset of labeled data $X_L \subset X$, $X_L = \{x_1, x_2, ..., x_L\}$, the Fisher information matrix with the data set $X_L$ is

$$M_{n_0}(X_L) = \sum_{l=1}^{L} \Phi_{n_0}(x_l)\Phi_{n_0}^T(x_l)$$

When a new datum $x_{L+1}$ is added, the new Fisher information matrix with the data set $X_{L+1}$ is

$$M_{n_0}(X_{L+1}) = M_{n_0}(X_L) + \Phi_{n_0}(x_{L+1})\Phi_{n_0}^T(x_{L+1})$$
The information gain by adding a new datum is defined by

$$\Delta I_{data} = \ln \text{Det}(M_{n_0}(X_{L+1})) - \ln \text{Det}(M_{n_0}(X_L))$$

For the datum $x_{L+1}$, the information gain is

$$\ln \Delta I_{data}(x_{L+1}) = 1 + \Phi^T_{n_0}(x_{L+1})M^{-1}_{n_0}(X_L)\Phi_{n_0}(x_{L+1})$$

Then the next most informative datum is selected by a “greedy” search, and its label is recovered for the estimation of weights.

$$x_{L+1} = \arg \max_{x \in X \setminus X_L} \ln \Delta I_{data}(x)$$
Classifier design: Weights Estimation

With $n_0$ basis functions $\Phi_{n_0}(\bullet)$ and $L_0$ labeled data $\{X_{L_0}, Y_{L_0}\}$ selected by FIC, the weights $w_{n_0}$ of the classifier are obtained via a Best Linear Unbiased Estimation (BLUE).

$$w_{n_0} = M_{n_0}^{-1}(X_{L_0}) \sum_{l=1}^{L_0} \Phi_{n_0}(x_l)y_l$$

$$M_{n_0}(X_{L_0}) = \sum_{l=1}^{L_0} \Phi_{n_0}(x_l)\Phi_{n_0}^T(x_l)$$
Overall Procedure

• **Collect** magnetometer and induction data at a given site (NO training data)

• **Select** a set of example signatures for basis function (NO digging required)

• **Sequentially** dig up UXO/clutter, one at a time, to learn labels of particular signatures

• Using the signatures and labels, refine classifier

• **Using** new classifier, choose next example to dig and learn label

• Refine classifier

• **Continue** digging and learning labels until the information gain to classifier is minimal

• Classifier design complete

• Using final classifier, discriminate UXO from clutter for remaining un-excavated objects
Discussion

• The proposed method is tested with the data from the JPG V demonstration site, which is deployed to assess different technologies for UXO detection and discrimination under realistic scenarios.

• Both GEM-3 data and magnetometer data are considered. Features are extracted via model fitting. Magnetometer feature extraction is similar to that of the induction sensor, but much simpler. A typical set of features are utilized for the classification.

• The data is from two adjoining areas, and includes totally 40 UXO, which are 60 mm and 81 mm mortars; 57 mm, 76 mm, 105 mm, 152 mm, and 155 mm projectiles; and 2.75 inch rockets. After the model fitting based prescreening operation, 260 clutters are left.
Results

• There are 16 UXO and 112 clutters in area 3, which is originally assigned as the training area. We observe that most UXO items presented in one area are also appeared in the other area, which seems to be a manual effort to keep the training data matched to the detecting data. We feel this may not be feasible in practice.

• The ROC curve of the proposed active training method is presented in comparison with results of a standard Kernel Matching Pursuits (KMP) from 100 random training selections. KMP is the most related classifier to the method proposed here, and it has achieved superior results comparing with some state of the art methods.
Number of labeled data: 128
Number of labeled data:  90
Number of labeled data: 60
Number of labeled data: 40
Future Directions

• Thus far have considered optimization of parameters (e.g. position, frequency, etc.) of single sensors
  
  - EMI
  - Acoustic
  - Magnetometer

• Now must place optimal sensing into context of multiple sensors

• Multiple sensors operating collectively (e.g. “team” of robots and/or UAVs)