FAST PHYSICS–BASED METHODS FOR WIDEBAND ELECTROMAGNETIC
INDUCTION DATA ANALYSIS

By
GANESAN RAMACHANDRAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

1
© 2010 Ganesan Ramachandran
ACKNOWLEDGMENTS

I thank my mother, JothiSree Venkataraman; father, Ramachandran Sethunarayanan; and wife, Swetha Seetharaman, for their relentless love and support.

I thank my advisor, Paul Gader, for his guidance and encouragement throughout my tenure at the University of Florida. I thank my committee Paul Gader, Joseph Wilson, Arunava Banerjee, and John Harris for their insight and guidance which has steered my research and bettered resulting contributions.

I thank my lab mates for their support and am thankful for their ability to endure my shenanigans. I thank Andres Mendez-Vasquez, Seniha Esen Yuksel, Xuping Zhang and Gyeongyong Heo for their encouragement, suggestions and aid in my research.

I thank colleagues, Jim Keller, Hichem Frigui, Peter Torrione and Dominic Ho, for their collaboration on a variety of research projects.

I thank Russell Harmon of Army Research Office (ARO), Richard Weaver and Mark Locke of Night Vision and Electronic Sensors Directorate (NVESD), and Waymond Scott of Georgia Tech, for their support of my research.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>3</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>6</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>7</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>9</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>Wideband Electromagnetic Induction Data and Analysis</td>
<td>11</td>
</tr>
<tr>
<td>Statement of Problem</td>
<td>12</td>
</tr>
<tr>
<td>Overview of Research</td>
<td>12</td>
</tr>
<tr>
<td>2 LITERATURE REVIEW</td>
<td>14</td>
</tr>
<tr>
<td>Induced Field of a Homogeneous Sphere</td>
<td>14</td>
</tr>
<tr>
<td>Spectral Models</td>
<td>17</td>
</tr>
<tr>
<td>Bell-Miller Models</td>
<td>17</td>
</tr>
<tr>
<td>First order Cole-Cole model</td>
<td>19</td>
</tr>
<tr>
<td>Nonlinear least squares optimization</td>
<td>21</td>
</tr>
<tr>
<td>Bishay’s circle fitting method</td>
<td>21</td>
</tr>
<tr>
<td>Xiang’s direct inversion method</td>
<td>23</td>
</tr>
<tr>
<td>Bayesian inversion</td>
<td>24</td>
</tr>
<tr>
<td>Discrete Spectrum of Relaxation Frequencies</td>
<td>24</td>
</tr>
<tr>
<td>Kth Order Cole-Cole</td>
<td>26</td>
</tr>
<tr>
<td>3 TECHNICAL APPROACH</td>
<td>27</td>
</tr>
<tr>
<td>Argand Diagrams of WEMI Responses</td>
<td>28</td>
</tr>
<tr>
<td>Prototype Angle Matching</td>
<td>29</td>
</tr>
<tr>
<td>Gradient Angle Model Algorithm</td>
<td>30</td>
</tr>
<tr>
<td>Stability of Lookup Table Step in GRANMA</td>
<td>32</td>
</tr>
<tr>
<td>Classifier Design</td>
<td>34</td>
</tr>
<tr>
<td>Gradient Angles in Parts Algorithm</td>
<td>34</td>
</tr>
<tr>
<td>Filter Design</td>
<td>36</td>
</tr>
<tr>
<td>Classifier Design</td>
<td>37</td>
</tr>
<tr>
<td>Dielectric Relaxation estimation using Sparse Models</td>
<td>37</td>
</tr>
<tr>
<td>Non-negative Least Squares Optimization</td>
<td>38</td>
</tr>
<tr>
<td>Convex Optimization with Sparsity Constraint</td>
<td>38</td>
</tr>
<tr>
<td>Quadratic programming</td>
<td>40</td>
</tr>
<tr>
<td>Linear programming</td>
<td>41</td>
</tr>
<tr>
<td>Joint Sparse Estimation of Dielectric Relaxations</td>
<td>42</td>
</tr>
<tr>
<td>Gradient Newton Methods</td>
<td>44</td>
</tr>
</tbody>
</table>
4 RESULTS ........................................................................................................................................53

Comparison on Synthetic Data ...............................................................................................53
  Experiment 1: Data from first order model .....................................................................53
  Observations ....................................................................................................................54
  Experiment 2: Comparison of Joint Sparse Methods on Data from second order model ..................................................................................................................56
  Experiment 3: Comparison of Cole-Cole and DSRF Dictionaries ..................................58
  Experiment 4: Dictionary update .....................................................................................60
  Experiment 5: Double Dictionary search – Proof of concept ..........................................61
  Experiment 6: Dictionary Correlation Analysis ..............................................................63

Comparison on Landmine Data ..............................................................................................64
  Data Sets ..........................................................................................................................64
  Data Filtering ...................................................................................................................65
  Experiment 6: Classification of objects using first order model .....................................65
    Experiment setup ......................................................................................................65
    Experiments and observations ..................................................................................66
    PRAM analysis .........................................................................................................66
    GRANMA analysis ..................................................................................................66
    Bishay method and Xiang inversion analysis ..........................................................68
  Feature Selection .............................................................................................................68
  Observations ....................................................................................................................68
  Experiment 7: Analysis of GRANPA on landmine data .................................................69
  Experiment 8: Analysis of Dictionary methods on landmine data ..................................72

5 CONCLUSIONS ....................................................................................................................76

LIST OF REFERENCES ...............................................................................................................77

BIOGRAPHICAL SKETCH .........................................................................................................79
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-1</td>
<td>Comparison of convergence statistics: Part (1)</td>
</tr>
<tr>
<td>4-2</td>
<td>Comparison of convergence statistics: Part (2)</td>
</tr>
<tr>
<td>4-3</td>
<td>Nomenclature and Proportion</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>2-1</td>
<td>Pictorial representation of the data collection setup</td>
</tr>
<tr>
<td>2-2</td>
<td>Geometry terms $g_n(r, d_{T, d_T}, R_T, R_S)$ for n from 1 to 4</td>
</tr>
<tr>
<td>2-3</td>
<td>Shape terms $\chi_n(kr)$ for n from 1 to 4</td>
</tr>
<tr>
<td>2-4</td>
<td>Effect of varying $\tau$ and $\alpha$</td>
</tr>
<tr>
<td>2-5</td>
<td>Real vs. imaginary parts of data from a first order Cole-Cole model</td>
</tr>
<tr>
<td>2-6</td>
<td>Effect of pole spread on DSRF coefficients</td>
</tr>
<tr>
<td>3-1</td>
<td>Argand plots of WEMI response</td>
</tr>
<tr>
<td>3-2</td>
<td>Angle plots of WEMI response</td>
</tr>
<tr>
<td>3-3</td>
<td>Plot of gradient angle $m$ in degrees</td>
</tr>
<tr>
<td>3-4</td>
<td>A) $\partial m/\partial \tau$ and B) $\partial m/\partial \alpha$ of the proposed model</td>
</tr>
<tr>
<td>3-5</td>
<td>Argand diagram of a low metal Anti-Personnel mine in the near field</td>
</tr>
<tr>
<td>3-7</td>
<td>Estimated relaxations A) $\tau=10^{-5}$, B) $\tau=10^{-4}$ and C) $\tau=10^{-3.4}$</td>
</tr>
<tr>
<td>4-1</td>
<td>Real vs Imaginary part of the induced response of a sample</td>
</tr>
<tr>
<td>4-2</td>
<td>Fitting Error vs. Signal Energy for different parameter estimation methods</td>
</tr>
<tr>
<td>4-3</td>
<td>Actual vs. estimated parameter values for noise free case</td>
</tr>
<tr>
<td>4-4</td>
<td>Actual vs. estimated parameter values for $\rho = 10^{-4}$</td>
</tr>
<tr>
<td>4-5</td>
<td>Histogram image of Number of terms vs. Fitting error</td>
</tr>
<tr>
<td>4-6</td>
<td>Convergence diagnostics for dictionary search and update</td>
</tr>
<tr>
<td>4-8</td>
<td>Two nonzero rows of the weight matrix $W^T$</td>
</tr>
<tr>
<td>4-9</td>
<td>Correlation regions of a Cole-Cole dictionary region</td>
</tr>
<tr>
<td>4-10</td>
<td>Down-track filter template</td>
</tr>
<tr>
<td>4-11</td>
<td>Plot showing the relationship between $\tau$, $\alpha$ and $E_L$ for different mine types</td>
</tr>
<tr>
<td>4-12</td>
<td>$\tau$ and $\alpha$ for different object types</td>
</tr>
</tbody>
</table>
4-13 ROC curves of different landmine detection algorithms. ..................................................69
4-14 Change in Argand diagram with distance.................................................................69
4-15 Change in amplitude with distance..............................................................................70
4-18 Parameter values estimated using joint sparse DSRF along down-track....................73
4-19 Relaxations estimated using a Cole-Cole dictionary ..................................................74
4-20 Usage of dictionary elements by the landmine data. ................................................74
4-21 Entropy of estimated relaxations for 28 different mine types....................................75
Three methods of object recognition using wideband electromagnetic induction data are described. A fourth method, which extends an existing algorithm to extract features using a dictionary based search, was developed to analyze objects. Emphasis was given to speed of execution as our interest is in their real-time performance.

Wideband electromagnetic induction data may consist of a wide range of frequencies starting from a few Hz to a few hundred thousand Hz. In addition to the object, the data usually has information about the sensor geometry, orientation of data collection setup and the medium in which the object lies.

The first method is called the Prototype Angle Matching (PRAM) algorithm that takes a non-parametric approach using the gradient angle between the real and imaginary components of the data. It classifies using distance from prototypes whose gradient angles have been measured. It is fast and does not make any assumptions about the data.

The second method, the Gradient Angle Model Algorithm (GRANMA), is based on a novel analytical derivation of the gradient angle using a first order Cole-Cole model. The analytical derivation reduces the number of parameters from four to two, enabling a fast look-up approach to nearest-neighbor classification schemes. Furthermore, the other two parameters can
easily be estimated from the first two. The method is demonstrated to be much faster and more robust than existing methods.

The third method, Gradient Angles in Parts Algorithm (GRANPA) uses a piecewise Cole-Cole modeling approach to attempt to estimate parameters of higher order models. It estimates the frequency segments in the data that follow the Cole-Cole model in an automated way and then uses the same setup as the GRANMA to extract the parameters.

The fourth set of methods, collectively referred to as SPARse Dielectric Relaxation estimation (SPARDIR), use a model that generalizes the Discrete Spectrum of Relaxation Frequencies (DSRF) model. The SPARDIR algorithms assume that the data are formed by a weighted combination of the Cole-Cole models and use a gradient Newton framework to search for parameters. A variety of combinations of $L_1$ and $L_2$ norm-based objective functions and constraints are investigated to seek sparse, physically meaningful parameter estimates. Furthermore, SPARDIR algorithms are devised that perform joint sparse estimation of parameters over a set of measurements and compared to SPARDIR and DSRF algorithms that perform point-wise sparse estimation.

Classification and parameter estimation results are given on sets of real, measured data as well as synthetic data sets for which the true parameter values are known. In these experiments, GRANMA performed better, more robust and with higher classification rates, than several existing algorithms and is faster than all but one. The Joint SPARDIR algorithms more accurately estimated the true underlying model parameters for more general models than previous work. In addition, the Joint SPARDIR algorithms are general; they are not specific to sensor type.
CHAPTER 1
INTRODUCTION

Wideband Electromagnetic Induction Data and Analysis

Wideband electromagnetic induction (WEMI) sensors have been used in a wide variety of applications ranging from finding mineral ore deposits [1] to monitoring blood glucose levels [2]. They are mainly divided into time domain and frequency domain sensors based on the domain of interpretation. The time domain sensors operate on the principle that when a time varying electrical current from a primary transmitter is injected through a dielectric medium, part of the electrical power is stored in the medium. When the current is stopped, the stored energy dissipates. This energy dissipation creates a secondary electric field which can be measured by a receiver. The rate at which the energy dissipates depends on the dielectric properties of the medium. This phenomenon is known as induced polarization (IP) [3].

Frequency domain sensors use a discrete number of sinusoidal signals. The secondary electric field or the induced field differs from the primary in amplitude and phase. This can be represented as the result of change in the complex impedance of the sensor [4]. By having the primary field consist a wide range of frequencies, we can get a complex spectral signature of the object under investigation. The main appeal of using electromagnetic induction is that different materials have different spectral signatures that can be used to detect and identify them. Also, different materials have different bandwidths, i.e., frequencies at which they produce maximum induction response.

There exists a multitude of electromagnetic induction sensors or metal detectors as they are more commonly known. Some commercial handheld metal detectors use a small range of frequencies and have applications such as treasure hunting or airport security. Their designs vary in the bandwidth of interest for the minerals they are tuned for. Most of them use the energy in
the secondary field for detection and seldom have discrimination capabilities between different materials within same bandwidth.

Wideband EMI sensors, as their name indicates, use a wide range of frequencies to identify different liquids or solids. Their frequencies of operation can range from a few hundreds of Hz to a few KHz. They are used in a variety of applications such as mineral prospecting, biological tissue analysis and landmine detection.

**Statement of Problem**

Parametric models offer a framework for characterization and classification of different objects in a physically interpretable way. Most parametric models for wideband EMI data are based on simplifications of Maxwell’s equations to characterize the secondary field. But they often suffer from one or more of the following assumptions

- The object to be identified is suspended in vacuum
- The object comprises of a single homogeneous material
- The object is of a simple geometric shape (sphere, cylinder etc)
- The surrounding medium is magnetically transparent
- Non-targets or clutter can be modeled

Even though most parametric models make such assumptions, they are still highly nonlinear. Parameter estimation is a nontrivial problem in most cases. Also most algorithms use iterative methods that require good knowledge of search ranges and starting values. We need models that are reasonably accurate to fit the observed data and estimated parameters to be consistent for identical objects to enable classification. Also, we need parameter estimation methods that are fast enough to be employed in practical applications.

**Overview of Research**

This research involved the development and analysis of both parametric and nonparametric models to detect, analyze and identify different metallic objects using frequency domain
wideband EMI sensors. The primary objective was to develop fast, physics-based algorithms for
landmine detection and extend them to other applications. We first developed models to
characterize spectral shapes at a given location and then modeled the variation in the parameters
with respect to distance and orientation of the sensor from the object.

The first method uses only the shape information in the spectral signature and designs a
prototype matching framework. It is nonparametric and hence does not make any assumptions
about the data distribution. It creates a fast anomaly detection framework for landmine detection.

The second method uses a novel gradient angle approach to solving the first order Cole-
Cole equation. The Cole-Cole equation has been used over half a century in the characterization
of dielectric properties of different minerals. The Gradient Angle Model Algorithm (GRANMA)
uses a fast lookup table method to avoid local optima and is able to model most signatures with
acceptable accuracy.

The third method Gradient Angles in Parts Algorithm (GRANPA) models more
complicated spectral shapes by extending the GRANMA algorithm to attempt piecewise
modeling approach. It uses the same framework of the GRANMA algorithm and hence is much
faster than existing methods.

The fourth method SPARse DIelectric Relaxation estimation (SPARDIR) uses a dictionary
of dielectric relaxations approach to analyze the underlying dielectric properties of the object. It
separates the properties of the object from the sensor setup.
CHAPTER 2
LITERATURE REVIEW

The following is a review of current literature pertinent to modeling and processing WEMI data. Most methods attempt to solve either one or both of two problems. Some try to characterize the spectral shape of the observed data at a given location with respect to the object, and others try to characterize the spatial pattern of energy around the object at a given frequency. Unified methods that can characterize shape and spatial energy pattern are rare except for a few basic shapes like sphere, cylinder etc. First, therefore, the analytical solution to a homogeneous sphere field in the field of a coaxial coil is presented to discuss the different factors that influence induced response. Next, the cases of more complicated models are reviewed. Finally, a brief review of WEMI models and parameter estimation methods in the context of unexploded ordnance (UXO) detection is given.

**Induced Field of a Homogeneous Sphere**

![Diagram of Induced Field of a Homogeneous Sphere](image)

Figure 2-1. Pictorial representation of the data collection setup
The simplest and most used analytical model in the WEMI literature is of a homogeneous sphere in the field of a coaxial coil [4]. Let $r$, $\sigma$ and $\mu$ denote radius, conductivity and permeability of the sphere respectively. Let $R_T$ and $R_S$ denote the radii of the transmit and receive coils located at $d_T$ and $d_S$ respectively from the sphere’s center.

Let $P_n^I(x)$ denote Legendre polynomials and $J_n(x)$ denote modified Bessel functions of the first kind. Then, the response $V^{(s)}$ of the sphere is given by

$$V^{(s)} = 2\pi \mu_0 I_\omega \frac{R_S R_T}{(d_T + R_T)^2} \sum_{n=1}^{\infty} g_n(r, d_T, d_S, R_T, R_S) \chi_n (kr)$$

where $g_n(r, d_T, d_S, R_T, R_S) = \frac{r^{2n+1}}{2n(n+1)} \frac{P_n^I(d_T/[d_T^2 + R_T^2]^{1/2})}{(d_T^2 + R_T^2)^{n+1/2}} \frac{P_n^I(d_S/[d_S^2 + R_S^2]^{1/2})}{(d_S^2 + R_S^2)^{n+1/2}}$, known as the geometry term and, $\chi_n (kr) = I_n (kr) + jQ_n (kr) = \left[ \frac{(n+1)\mu/\mu_0 + n}{n(\mu/\mu_0 - 1)} \right] J_{n+1/2} (kr) - kr J_{n-1/2} (kr)$, known as the shape term with $k^2 = j\sigma\mu\omega = \frac{2j}{\delta^2}$, and $\mu_0 = 4\pi 10^{-7}$ is the permittivity of free space and $\delta$ is the skin depth of the material defined as $\delta = \sqrt{\frac{2}{\omega\mu\sigma}}$.

The geometry term $g_n(r, d_T, d_S, R_T, R_S)$ controls the relative importance of different order terms and also the magnitude of the response. The shape term $\chi_n (kr)$ defines the shape of the spectral response. The first order term ($n=1$) is known as the dipole response. Most often only the first order approximation is used as the higher order terms fall off quickly for large distances and small spheres. In the example shown in Figure 2-2 and Figure 2-3 for the case with $R_S = R_T = 0.3m$ and $a = 0.05m$ that are typical in landmine detection sensors, only $n = 1$ needs to be accounted for.
Figure 2-2. Geometry terms $g_n(r, d_1, d_2, R_T, R_S)$ for $n$ from 1 to 4. The terms are shown for $d_1=d_2$ in the range of 0 to 0.3m, $R_1=R_2=0.2$ m, $r=0.05$ m. $n>1$ terms are negligible compared to $n=1$ term.

Figure 2-3. Shape terms $\chi_n(kr)$ for $n$ from 1 to 4. The terms are shown for $\omega=2\pi 10^{-6}$ to $2\pi 10^5$ Hz $\sigma=5.8x10^7$ with A) $\mu=\mu_0$ and B) $\mu=100\mu_0$. The x and y axes represent the real and imaginary parts respectively.
Since most of the objects of interest are neither spheres nor homogeneous, complete mathematical analysis is not possible. Therefore there have been a number of attempts to model the induced response using parametric models, some of which are discussed in the following section.

**Spectral Models**

One of the most common approaches in the study of wideband EMI response of objects is to model the spectral shape. Most of the shape modeling approaches either make some assumptions on the shape and use a parametric model, or use a set of basis functions to represent it non-parametrically.

**Bell-Miller Models**

Miller et al. [5] showed that for most objects, the analytical solution for the homogeneous sphere can be approximated by the first order term. For \( n = 1 \), the Bessel function terms simplify to hyperbolic sines and cosines. The shape term in Equation 2-1 can be simplified as,

\[
\chi_i(kr) = \frac{[\mu_0 + \mu_i k^2 r^2 + \mu] \sinh(kr) - kr[2\mu + \mu_0] \cosh(kr)}{[\mu_0 + \mu_i k^2 r^2 - \mu] \sinh(kr) - kr[\mu - \mu_0] \cosh(kr)}
\]

(2-2)

They show that in the case of highly permeable objects (\( \mu >> \mu_0 \)), this formula can be simplified into a 3-parameter model to characterize more general but compact shapes as,

\[
\chi_i(kr) = a \left( s + \left( \frac{j\omega\tau}{2} \right)^{1/2} \right) \left( \frac{j\omega\tau}{1 + 1/2} \right)
\]

(2-3)

with \( a, s, \) and \( \tau \) known as the amplitude, shift and relaxation time respectively. For a compact shaped object, the EMI response can be approximated by the dipole moment \( \mathbf{m} \) (3x1 vector) induced in the target by the primary field \( \mathbf{h}_0 \) (3x1 vector) created by the transmitter coil [6]. The approximation neglects higher order multi-pole contributions to the response, and is valid if the
distance from the sensor to the object is large compared to the dimensions of the object. For a
harmonic field oscillating at frequency $\omega$,

$$m e^{j\omega t} = V P h_0 e^{j\omega t}$$

(2-4)

where $V$ is the volume of the object (scalar) and $P$ is the magnetic polarizability tensor [7] (3x3
matrix for each frequency $\omega$). $P$ fully characterizes the EMI dipole response. The elements $P_{ij}(\omega)$
of $P$ are a function of $\omega$ and depend on the object’s electrical properties, shape and on its
orientation in the primary field. They are complex numbers corresponding to the frequency-
dependent phase shift between primary and induced field.

Since the induced field is causal, $P$ has the property of being symmetric. This means that it
can be diagonalized and be represented by its Eigen values $\lambda_i$ and Eigen vectors $u_i$ as

$$P = U \Lambda U^T,$$

where $U = [u_1, u_2, u_3]$ and $\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$. At any given location and angle of the measurement
setup with respect to the object, what we observe is a linear combination of the Eigen values. If
$a_i(d)$ denotes the strength of the induced field at distance $d$ along the Eigen direction $i$, then the
response of a compact object according to the 3-parameter Bell-Miller model is given by,

$$z_i(\omega, d) = (I_i(\omega, d) + jQ_i(\omega, d))_{p_i} = a_i(d) \left[ s_i(d) + \frac{(j \omega \tau_i)^{\frac{1}{2}} - 2}{(j \omega \tau_i)^{\frac{1}{2}} + 1} \right]$$

(2-5)

The model assumes that the value $\tau_i$ remains the same along a given Eigen direction around the
influence of the object. For a sphere, all the Eigen values are equal and the response is identical
in any direction.

The rest of this document deals with the influence of orientation and distance separately.

The initial work assumes that the primary and secondary coils are equidistant from the object and
the object’s response can be approximated by its dipole approximation unless mentioned otherwise.

By replacing the exponent with $\alpha$ instead of $\frac{1}{2}$, the above model can be used to characterize non-compact shapes. The 4-parameter model is given by,

$$z(\omega) = (I(\omega) + jQ(\omega))_{p_i} = a \left( s + \frac{(j\omega\tau)^\alpha - 2}{(j\omega\tau)^\alpha + 1} \right)$$  \hspace{1cm} (2-6)

The factor $\alpha$ controls the width of the quadrature part and hence controls the bandwidth of the material. The bandwidth in this context represents the range of frequencies where the object under study has a significant imaginary part.

**First order Cole-Cole model**

Cole et al.[8], proposed a four parameter model in the context of mineral prospecting. Denoting the WEMI response at infinite and zero frequencies respectively as $z_\infty$ and $z_0$, it is given as,

$$z(\omega) = z_\infty + \frac{z_0 - z_\infty}{1 + (j\omega\tau)^\alpha}$$  \hspace{1cm} (2-7)

Since both the first order Cole-Cole model and the four parameter Bell-Miller model have the same number of parameters and are of similar form, they warrant further analysis. From Equation 2-6, we can write the zero and infinite frequency responses of the four parameter Bell-Miller model as,

$$z(0) = a(s - 2)$$
$$z(\infty) = a(s + 1)$$  \hspace{1cm} (2-8)

With simple re-arranging of terms, the amplitude and shift can be shown to be, $a = \frac{z_\infty - z_0}{3}$, and
\[ s = \frac{2z_\infty + z_0}{z_\infty - z_0} \]. Substituting the notations for \( a \) and \( s \), the four-parameter model becomes,

\[ z(\omega) = z_\infty + \frac{z_0 - z_\infty}{1 + (j\omega \tau)^\alpha} = \text{first order Cole-Cole}. \]

Hence, the 4-parameter model proposed by Miller et al. defaults to 1\(^{\text{st}}\) order Cole-Cole model [8].

![Figure 2-4. Effect of varying \( \tau \) and \( \alpha \).](image)

Figure 2-4 shows the effect of varying \( \tau \) and \( \alpha \) in the shape of the curve in the In-phase vs. Quadrature-phase plot, also known as the Argand diagram. As \( \tau \) goes outside a boundary of values, only part of the shape is visible in the Argand diagram due to the finite bandwidth of operation. Therefore the range of \( \tau \) is limited by the bandwidth of the sensor.

In most unexploded ordnance (UXO) objects, there are soft metal rings near the tail of the projectile, known as driving bands to facilitate being fired by a gun. This can be modeled by extending the 4-parameter model into a 5-parameter model as,
\[
\mathbf{z}(\omega) = (\mathbf{I}(\omega) + j\mathbf{Q}(\omega))_{ps} = \left(a + \frac{(j\omega\tau)^{\alpha} - 2}{(j\omega\tau)^{\alpha} + 1} + b \left(\frac{j\omega\tau_{\text{Loop}}}{(j\omega\tau_{\text{Loop}})^{\alpha} + 1}\right)^{\alpha} - 2\right)
\]

where \(\tau_{\text{Loop}} = 10^{1.295}\tau\) and \(c_{\text{Loop}} = 0.943\) are fixed empirically. This can be shown as a special case of 2\textsuperscript{nd} order Cole-Cole model.

While Bell-Miller models provide a framework for analyzing the electromagnetic induction spectra, they don’t provide a fast way of estimating the parameters. There have been many attempts at finding exact or approximate solutions to the first order Cole-Cole model, some of which are described in the following sections.

**Nonlinear least squares optimization**

The most straightforward albeit computationally intensive method to estimate the parameters is to use iterative least squares curve fitting approach. If \(\omega_1, \omega_2, \ldots, \omega_N\) are the frequencies at which measurements are taken, then the parameters are given by,

\[
\{\hat{a}, \hat{s}, \hat{\tau}, \hat{c}\} = \arg\min_{a,s,\tau,c} \sum_{k=1}^{N} \left| \mathbf{z}(\omega_k) - a \left( s + \frac{(j\omega_k\tau)^{\alpha} - 2}{(j\omega_k\tau)^{\alpha} + 1}\right) \right|^2
\]

with the constraint that the parameters are real. This is usually achieved by using an iterative least squares curve fitting algorithm. However, it requires a good initial estimate for converging to the global or at least a good optimum. In practice, this requires the algorithm to be run multiple times with different random initial conditions to find the optimal solution. The ways of selecting the initial conditions and the number of epochs required are usually heuristic. It is also a common practice to first divide the data by its total norm and then finally rescale the estimated amplitude.

**Bishay’s circle fitting method**

Claim 2-1: If the spectral signature follows the Cole-Cole model, then the plot of \(\mathbf{I}(\omega)\) vs. \(\mathbf{Q}(\omega)\) will be a segment of a circle.
**Proof:** Let us assume that the points from a first order Cole-Cole model indeed form a segment of a circle. Let the center of the circle be at \((a_c, -b_c)\), then the points from the model intersect the x-axis at \((z_0, 0)\) and \((z_\infty, 0)\).

From Figure 2-5, the \(I(\omega)\) coordinate of the center lies where \(Q(\omega)\) reaches its maximum. Taking the derivative \(Q(\omega)\) of with respect to \(\omega\),

![Figure 2-5. Real vs. imaginary parts of data from a first order Cole-Cole model.](image)

At the maximum of \(Q(\omega)\), the derivative is zero. Ignoring cases where \(\omega = 0\), \(\omega = \infty\) and \(\alpha = 0\), the maximum happens at \(\omega = 1/\tau\). Denoting the maximum by \(Q_{\text{max}}\) we get

\[
Q_{\text{max}} = Q(\omega = 1/\tau) = \frac{3a}{2} \tan\left(\frac{\alpha \pi}{4}\right).
\]

The x coordinate of the center is given by \(a_c = I(\omega=1/\tau) = a(s-1/2)\). From the figure, radius is given by \(r = b_c + Q_{\text{max}}\). By using \((I(\omega)-a_c)^2 + (Q(\omega)+b_c)^2 = r^2\) constraint, we get

\[
(I(\omega)-a(s-0.5))^2 + (Q(\omega)+r-Q_{\text{max}})^2 = r^2
\]

Substituting for \(I(\omega), Q(\omega)\) and \(Q_{\text{max}}\) and solving for \(r\), we get
Therefore, the first order Cole-Cole equation can be written as
\[
\left( I(\omega) - a \left( s - \frac{1}{2} \right) \right)^2 + \left( Q(\omega) - \frac{3a}{2} \left( \csc \left( \frac{\alpha \pi}{2} \right) - \tan \left( \frac{\alpha \pi}{2} \right) \right) \right)^2 = \left( \frac{3a}{2} \csc \left( \frac{\alpha \pi}{2} \right) \right)^2 \tag{2-11}
\]

The cases where \( \alpha > 1 \) correspond to the segment being greater than a semicircle and vice versa. If \( N_f \) denotes the number of frequencies where measurements are made, then in cases where \( z(0) \) and \( z(\infty) \) can be measured, the values of \( \tau \) and \( \alpha \) can be estimated as [9]:
\[
\alpha = \frac{1}{N_f} \sum_{k=1}^{N_f} 2 \left[ \tan^{-1} \left( \frac{Q(\omega_k)}{I(\omega_k) - z(0)} \right) + \tan^{-1} \left( \frac{Q(\omega_k)}{z(\infty) - I(\omega_k)} \right) \right] \tag{2-12}
\]
\[
\tau = \frac{1}{N_f} \sum_{k=1}^{N_f} \frac{1}{\omega_k} \left[ \frac{(I(\omega_k) - z(0))^2 + Q(\omega_k)^2}{(z(\infty) - I(\omega_k))^2 + Q(\omega_k)^2} \right]^{1/2} \tag{2-13}
\]

Since in most practical applications, \( z(0) \) and \( z(\infty) \) cannot be directly measured, they have to be estimated. If the measured data includes the maximum of \( Q(\omega) \) at frequency \( \omega_m \), then
\[
2I(\omega_m) \approx z(0) + z(\infty). \]
Since \( z(0) \) and \( z(\infty) \) are the \( I(\omega) \)-axis intercepts, they are given by, \( z(0) = a_c - (r^2 - b_c^2)^{1/2} \) and \( z(\infty) = a_c + (r^2 - b_c^2)^{1/2} \). The value of \( b_c \) is positive if \( \alpha > 1 \) and is negative if \( \alpha < 1 \). The values of \( a_c, b_c \) and \( r \) are found by using least squares search with the constraint \( (I(\omega) - a_c)^2 + (Q(\omega) - b_c)^2 = r^2 \). This approach works well for spectral shapes that fall on a circle but is not robust against noise and deviations from the assumed shape. Also, in cases where \( z(\omega) \) has flat regions, \( r \approx \infty \) which makes the estimate unstable.

**Xiang’s direct inversion method**

Pelton et al. [10] formulated the Cole-Cole model in terms of induced polarization (IP) parameters in the form,
\[
\mathbf{z}(\omega) = \mathbf{z}(0) \left[ 1 - m \left( 1 - \frac{1}{1 + (j \omega \pi)^2} \right) \right], \text{ with } m = 1 - \frac{z(\infty)}{z(0)}. \tag{2-14}
\]
The Xiang inversion method [11] converts the parameter estimation into a 1-D problem using elimination of variable to estimate the parameters of the Pelton model, and hence is more computationally efficient, but restricts the values of $m$ and $\alpha$ to be between 0 and 1. This proves to be a handicap in modeling certain low metal mines as shown in the experiments.

**Bayesian inversion**

Since direct inversion of the first Cole-Cole may lead to sub-optimal solutions, a Bayesian approach [1] can be used to find the global optimum. $z(0)$ and $\tau$ are characterized using Jeffreys priors or alternately, $\log_{10}(z(0))$ and $\log_{10}(\tau)$ by uniform probability density. Also, by making a change of variable from $m$ to $m'$ using $m' = \log_{10}(m/(1-m))$, it can be represented by a uniform pdf. And finally, by characterizing $\alpha$ by a uniform pdf, the Bayesian *a posteriori* probability distribution is given by:

$$p(\theta) = a \tau \cdot z(0)m(1-m)\pi(\theta)\exp[-0.5*(z - z_r(\theta))^T C_{dd}^{-1}(z - z_r(\theta))]$$

(2-15)

where $p(\theta)$ represents the *a priori* probability density of the model parameters, $z_r(\theta)$ the reconstructed data and $C_{dd}$ is the NxN data covariance matrix. The individual parameters are found by finding the respective marginal pdfs. Though this method can produce accurate estimates for the parameters with the right choice of priors, it is computationally intensive as the posteriors do not have closed form solutions and hence need to be estimated by n-dimensional grid search.

**Discrete Spectrum of Relaxation Frequencies**

The EMI response of a target can be modeled as a sum of real exponentials which, in frequency domain can be represented as poles in the spectrum [12]. They are represented as:

$$z(\omega) = c_0 + \sum_{k=1}^{K} \frac{c_k}{1 + j \omega \tau_k}, \text{ with } \tau_k \geq 0 \text{ and } c_k \geq 0.$$

(2-16)
The values $c_k$ are estimated by searching for the $\zeta = 1/\tau_k$ values in the region where $\omega \tau_k > 10^{-2}$ to $10^2$. Since $c_k$ are restricted to be real and positive, the matrix setup shown in Equation 2-16 is solved using non-negative least-squares for real and imaginary components in parallel.

$$
\begin{bmatrix}
\mathbf{z}(\omega_1) \\
\mathbf{z}(\omega_2) \\
\vdots \\
\mathbf{z}(\omega_N)
\end{bmatrix} =
\begin{bmatrix}
1 & \frac{1}{1+ja_1/\zeta^1} & \frac{1}{1+ja_2/\zeta^2} & \cdots & \frac{1}{1+ja_M/\zeta^M} \\
1 & \frac{1}{1+ja_1/\zeta^1} & \frac{1}{1+ja_2/\zeta^2} & \cdots & \frac{1}{1+ja_M/\zeta^M} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \frac{1}{1+ja_1/\zeta^1} & \frac{1}{1+ja_2/\zeta^2} & \cdots & \frac{1}{1+ja_M/\zeta^M}
\end{bmatrix}
\begin{bmatrix}
c_0 \\
c_1 \\
\vdots \\
c_M
\end{bmatrix}
$$

$\mathbf{z} = \Theta \vec{c}$

(2-17)

Figure 2-6. Effect of pole spread on DSRF coefficients. A) Argand diagram of the original and reconstructed signals. B) DSRF pole locations and their strengths. Since DSRF assumes the pole spread to be unity, it splits a single pole into many for low $\alpha$ values. In other words, DSRF tries to model the curve in the Argand diagram by a weighted sum of semi-circles.
If the data is in fact generated from a Cole-Cole model with $\alpha<1$, the DSRF splits a single pole into multiple poles to make an accurate fit. In other words, DSRF tries to model the curve in the Argand diagram which is a segment of a circle smaller than a semi-circle by a weighted sum of semi-circles. This makes the characterization and classification of objects difficult.

**K\textsuperscript{th} Order Cole-Cole**

In the analysis of dielectric properties of biological tissues, the spectral shapes are more complicated than that of the first order Cole-Cole model [13]. By generalizing the DSRF, we may be able to model complicated shapes with lesser number of parameters. This leads to a generalized version of the Cole-Cole equation as:

$$z(\omega) = c_0 + \sum_{k=1}^{K} \frac{c_k}{1 + (j\omega \tau_k)^\alpha_k}, \text{ with } c_k \geq 0.$$  \hspace{1cm} (2-18)

This form of K\textsuperscript{th} order Cole-Cole model with $K=4$ is widely used in breast tumor detection and other wideband EMI applications in medicine and biology. Direct inversion of a K\textsuperscript{th} order Cole-Cole model is mathematically intractable. There however, exist a few methods that attempt to find approximate solutions.

If the order $K$ is known, then an iterative least squares method can be used to find the $c_k$’s and $\tau_k$’s. The finite difference time domain method is also widely employed by the biological tissue analysis community. Both these methods however require a good estimate of the initial conditions and are computationally inefficient.
CHAPTER 3
TECHNICAL APPROACH

The focus of this research was on developing tools for detecting, analyzing and identifying objects from sets of spectral measurements obtained from a wideband EMI sensor. Three specific approaches, namely Prototype Angle Matching (PRAM), Gradient Angle Model Algorithm (GRANMA) and Gradient Angle in Parts Algorithm (GRANPA), using the gradient angle between the real and imaginary components of the data were developed each building on the previous one and in the order of increasing complexity on simple spectral signatures. GRANMA and GRANPA express the data in terms of their dielectric properties by extracting features related to the underlying physical phenomena. PRAM and GRANMA identify objects based on distance to one or more nearest prototypes. PRAM is a fast prototype matching method based on gradient angles calculated numerically on the spectral signatures. GRANMA uses the gradient of an analytical model, and reduces the number of parameters to be estimated simultaneously from four to two. It significantly reduces the complexity of the search using a two stage approach. GRANPA uses piecewise curve fitting to model more complicated spectra. It provides unique solutions to model near field effects and is faster than existing methods.

A fourth method, SPARse DIelectric Relaxation estimation (SPARDIR), was introduced that looks for a jointly sparse solution to provide the most compact representation of the data. SPARDIR also separates the underlying physical phenomena from the sensor setup. Two solutions to this algorithm, SPARDIR2 and SPARDIR1 were developed based on minimizing quadratic and absolute error respectively. Although both methods provide sparse solutions and approximate the data well, the first one is faster whereas the second one provides solutions closer to the true underlying physics.
In the following section, we briefly review the Argand diagram and explain the four methods, PRAM, GRANMA, GRANPA and SPARDIR.

**Argand Diagrams of WEMI Responses**

Argand diagrams have been one of the widely used tools to study complex data. They are also known as Nyquist plots in the context of control theory and Cole-Cole plots in electrochemistry. It is a plot of In-phase component $I(\omega)$ against the Quadrature-phase component $Q(\omega)$. Figure 3-1 shows that it provides a consistent framework to identify specific mine types. Mines with similar metal content have similar signatures (which may not be true in the case of Ground Penetrating Radar). This property can be exploited in landmine detection and classification.

![Argand plots of WEMI response for a A) boxed AP mine and a B) circular AT mine at different locations and depths. Different points in the curve represent the WEMI response at different frequencies.](image)

Figure 3-1 shows that though the shapes are similar, there is variability in the amplitude and in the shift in In-phase component between different candidates of same mine type. The
The slope between the points in the Argand diagram is a powerful tool for discrimination. Figure 3-2 shows that, the variability between different candidates is reduced as the gradient angle is independent of amplitude and shift, two of the most difficult and unreliable (Fails, et al., 2007) values to estimate. Here the angles are calculated numerically by using a two-sided gradient.

\[
\begin{align*}
\partial Q(\omega) &= Q(\omega + \Delta \omega) - Q(\omega - \Delta \omega) \\
\partial I(\omega) &= I(\omega + \Delta \omega) - I(\omega - \Delta \omega)
\end{align*}
\]

The gradient angle is defined as

\[
m(\omega) = atan2(\partial Q(\omega), \partial I(\omega))
\]

Using the MATLAB® definition of \( atan2 \), the equation for the gradient angle is given by:

\[
m(\omega) = 2 \tan^{-1} \left( \frac{\partial Q(\omega)}{\partial I(\omega) + \sqrt{(\partial I(\omega))^2 + (\partial Q(\omega))^2}} \right).
\]

### Prototype Angle Matching

The Prototype Angle Matching method is a nearest prototype classifier. The prototype of each object class is defined to be point-wise median. More precisely, if \( T_i \) represents the set of indices of all candidates of object type \( i \), then the corresponding set of gradient angles is given by \( M'(\omega) = \{ m_k(\omega) : \text{for all } k \in T_i \} \) and the prototype by:

\[
P_i(\omega) = \text{median}[m_k(\omega)]
\]

If \( z \) is a test vector, then the confidence that \( z \) is of target class is defined by

\[
\text{conf}(z) = \frac{1}{1 + e^{-\gamma(d(z,P_i)-\delta)}}
\]

where \( d_i(z) = \min_i d(z,P_i) \) is defined to be the distance between \( z \) and the target class \( t \) and \( \gamma \) and \( \delta \) are the rate and bias parameters of the logistic function to be estimated by cross validation [15].
Figure 3-2. Angle plots of WEMI response for a A) circular AT mine and B) boxed AP mine and at different locations and depths.

**Gradient Angle Model Algorithm**

Parameters of the Cole-Cole model have been used successfully [14] to model and classify metallic objects. The most widely used parameter estimation methods use nonlinear optimization methods, which are slow due to the presence of local minima. The simpler methods are faster but not robust against noise and deviations from the model. In this section, it is shown that by analytically differentiating the Cole-Cole model, a fast lookup algorithm can be formulated. This is one of the contributions of the present research.

Separating the $I(\omega)$ and $Q(\omega)$ components using MATHEMATICA®, (verified numerically and by MATLAB® symbolic math toolbox) the four-parameter model becomes:
\[ I(\omega) = a \left( s + 1 - \frac{3 \left( 1 + (\omega \tau)^{2} \cos \left( \frac{\alpha \pi}{2} \right) \right)}{1 + (\omega \tau)^{2} + 2(\omega \tau)^{\alpha} \cos \left( \frac{\alpha \pi}{2} \right)} \right), \ \text{and} \]  

\[ Q(\omega) = a \left( \frac{3(\omega \tau)^{\alpha} \sin \left( \frac{\alpha \pi}{2} \right)}{1 + (\omega \tau)^{2} + 2(\omega \tau)^{\alpha} \cos \left( \frac{\alpha \pi}{2} \right)} \right). \]  

Using the same framework as Equation 3-2, the equation for the gradient angle for four-parameter model (assuming \( a > 0 \)) is

\[ m(\omega) = 2 \tan^{-1} \left( \frac{-1 + (\omega \tau)^{\alpha}}{1 + (\omega \tau)^{\alpha} \tan \left( \frac{\alpha \pi}{4} \right)} \right). \]  

The range of \( m(\omega) \) is given by evaluating it at the extrema of \( \omega \tau : \)

\[ m(\omega) \big|_{\omega \tau = 0} = \frac{\alpha \pi}{2} \quad \text{and} \quad m(\omega) \big|_{\omega \tau = \infty} = -\frac{\alpha \pi}{2}. \]

These equations show that the angle varies between \( \alpha \pi/2 \) and \( -\alpha \pi/2 \) which is adequate to model all angles when \( \alpha \in [0, 2] \). They also show that the three-parameter Bell-Miller model (Miller, et al., 2001), where \( \alpha \) is restricted to be equal to 0.5, can only characterize angles between \(-\pi/4\) and \(\pi/4\). Analysis of empirical data shows the latter range to be inadequate.

Since the equation for the gradient angle involves only two variables, a lookup table was created for \( m(\omega) \) for a range of values of \( \tau \) and \( \alpha \). The elements of the table are denoted by \( \hat{m}(\omega, \alpha, \tau) \). Values of \( \tau \) and \( \alpha \) were found by calculating the best match between \( m(\omega) \) and table values using least mean absolute error:

\[ \hat{E}(\alpha, \tau) = \sum_{\omega} \left| m(\omega) - \hat{m}(\omega, \alpha, \tau) \right| \]  

The lookup table error \( E_L \) is defined as follows:
\[ E_L = \hat{E}(\hat{\alpha}, \hat{\tau}) \quad \text{where} \quad \hat{\alpha}, \hat{\tau} = \arg \min_{\alpha, \tau} \hat{E}(\alpha, \tau) \]  
\hspace{1cm} (3-9)

The amplitude \( a \) was found by substituting the estimates \( \tau \) and \( \alpha \) in Equation 3-6 and finding the maximum likelihood estimate:

\[ \hat{a} = \frac{Q(\omega)}{\text{Im} \left( \frac{(j \omega \hat{\tau})^\alpha - 2}{(j \omega \hat{\tau})^\alpha + 1} \right)} \]  
\hspace{1cm} (3-10)

where \( \hat{\tau} \) and \( \hat{\alpha} \) are lookup table estimates, and \( N_f \) is the number of measured frequencies.

Finally, the shift \( s \) was found by substituting the estimates of \( a, \tau \) and \( \alpha \) in Equation 3-5:

\[ \hat{s} = \frac{I(\omega)}{\hat{a}} - \text{Re} \left[ \frac{(j \omega \hat{\tau})^\alpha - 2}{(j \omega \hat{\tau})^\alpha + 1} \right] \]  
\hspace{1cm} (3-11)

The goodness of fit for this method was measured as the normalized error between the actual data and the fit:

\[ E_F = \frac{\sum_{k=1}^{N_f} |z(\omega_k) - \hat{a} \left( \hat{s} + \frac{(j \omega_k \hat{\tau})^\alpha - 2}{(j \omega_k \hat{\tau})^\alpha + 1} \right)^2}{\sum_{k=1}^{N_f} |z(\omega_k)|^2}. \]  
\hspace{1cm} (3-12)

**Stability of Lookup Table Step in GRANMA**

All the parameter estimates depend on the accuracy of the gradient and on the resolution of the table. Therefore, first it was necessary to see how a small error in the parameter estimate in the \((\tau, \alpha)\) space due to the finite size of the lookup table affects the angle estimate. Figure 3-3 shows the dependence of \( m \) on \( \tau \) and \( \alpha \) for a fixed \( \omega \). It shows two regions of interest, one being \( \alpha = 0 \), and another \( \alpha = 2 \). Figure 3-4 shows that near regions where \( \omega \tau = 1 \), there is a large change in \( m \) for small changes in \( \tau \) and \( \alpha \) due to the fact that the denominator in Cole-Cole equation becomes zero. Therefore, regions too close to \( \alpha = 2 \) were avoided.
Figure 3-3. Plot of gradient angle $m$ in degrees vs. $\omega \tau$ and $\alpha$. It shows a discontinuity in $m$ when $\alpha = 2$ and $\omega \tau = 1$.

Figure 3-4. A) $\partial m / \partial \tau$ and B) $\partial m / \partial \alpha$ of the proposed model, computed numerically, shown at 330Hz. It shows the pattern that, near to regions where $\omega \tau \approx 1$ and $\alpha \approx 2$, there is a big difference in the angles for a small change in $\tau$ or $\alpha$. 
Next, it was necessary to analyze the robustness of the estimated parameters with respect to noise in the data and in the gradient estimates. Since the noise in the parameters is non-linearly linked to the noise in the data, it is difficult to derive expressions linking the noise in them. Therefore, the stability of the method with respect to noise in the data was analyzed in the context of classifier design.

Classifier Design

The classification was done on six features using a soft $K$ nearest neighbor method. The features are $a$, $\tan^{-1}(s)$, $\log_{10}(\tau)$, $\alpha$, $E_L$ & $E_F$ defined in Equations 3-9 to 3-12. The value of $K$ was found as the one that gave minimum variance without sacrificing performance. The confidence value for object $i$ is given by

$$\text{conf}(i) = \frac{\sum_{j=1}^{K} n'_j}{\sum_{j=1}^{K} m'_j}$$

(3-13)

where $n'_j$ is the distance to $j^{th}$ nearest non-mine neighbor, and $m'_j$ is the distance to $j^{th}$ mine neighbor. The contribution here is the estimation of the Cole-Cole parameters, not in the classifier design. Any suitable classifier could be used here.

Gradient Angles in Parts Algorithm

The first order Cole-Cole model together with GRANMA is useful for modeling objects that show a single relaxation in their spectral signatures. This, however, may not be enough to model objects that have more complicated spectral shapes. The $K^{th}$ order Cole-Cole has found many applications especially in electromagnetic studies in biology [13]. However, there are no direct inversion methods available to estimate the parameters and existing methods are too slow to be used in a real-time environment like landmine detection [13].
If the order $K$ is known and the relaxations express themselves in frequency bands with little overlap, then a piecewise curve fitting approach can be used to estimate $K^{th}$ order Cole-Cole parameters.

Figure 3-5. Argand diagram of a low metal Anti-Personnel mine in the near field. It shows that a single pole model is not enough to capture the information contained. The red and magenta circles represent output of two first order Cole-Cole models and the black asterisks represent their sum.

This figure shows an example a spectrum with two relaxations that dominate in different frequency bands. Therefore, if the frequency bands are known, then we can characterize the frequency bands individually using a single Cole-Cole model and estimate the parameters using GRANMA.

$$z(\omega) = h_1(\omega)z_1(\omega) + (1 - h_1(\omega))z_2(\omega) \quad (3-14)$$

where $h_1(\omega)$ is a low pass filter with frequency response:
\[ h_i(\omega) = \begin{cases} 1, & \omega \in [0, \omega_b) \\ 0, & \omega \geq \omega_b \end{cases} \], with \( \omega_b \) as the cutoff frequency.

This approach can be extended to a general \( K \)th order case as a bank of band-pass filters:

\[ z(\omega) = \sum_{k=1}^{K} h_k(\omega)z_k(\omega), \text{ with } \sum_{k=1}^{K} h_k(\omega) = 1 \text{ for all } \omega. \]  \hspace{1cm} (3-15)

Now, the problem of estimating \( K \)th order Cole-Cole equations simplifies into finding the band-pass filters, so that individual bands can be modeled using a first order Cole-Cole model.

**Filter Design**

Figure 3-6. A) Argand diagram of a low metal Anti-Personnel mine in the near field. B) gradient angle at different frequencies. C) slope of gradient angle
There are many ways of finding the optimal cutoff frequency in two filter case. One of the easiest is to search exhaustively in the frequencies of interest (usually around the middle frequency) to optimize for the least lookup table error. Another method would be to find the location of the gradient angle minimum around the center frequency as shown in Figure 3-6 B. Another option would be to choose the frequency where slope of gradient angle crosses from negative to positive around the center frequency as shown in Figure 3-6 C. We found the zero crossing to be a reliable estimate of the cutoff frequency for landmine data. Experimental results demonstrated that a less heuristic method should be pursued. There is one merit that should be mentioned here. GRANPA provides a simple and fast way of estimating the near field effects.

**Classifier Design**

A soft kNN classifier similar to the one used in GRANMA case could be used. Relevance Vector Machines (RVM) also show promise in case of GRANMA features and could be used here. This method is applicable only when the order \( K \) is known and the relaxations are bandwidth limited.

**Dielectric Relaxation estimation using Sparse Models**

We can use similar to the framework used by DSRF to model such curves. By extending the matrix \( T \) to include different values of \( \tau \) and \( \alpha \), we get:

\[
\begin{bmatrix}
\mathbf{z}(\omega_1) \\
\mathbf{z}(\omega_2) \\
\vdots \\
\mathbf{z}(\omega_N)
\end{bmatrix}
= 
\begin{bmatrix}
1 & 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & 1 \\
1 & 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & 1 \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
1 & 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & 1
\end{bmatrix}
\begin{bmatrix}
c_0 \\
c_1 \\
\vdots \\
c_M\text{ML}
\end{bmatrix}
\]
or,

\[ z = \Theta c \]

\[
\theta_{i,j} = \frac{1}{n^{(j-1)\bmod L}} \quad n = 1 + ((j-1)\bmod L), \quad m = \text{Integer}\left[\frac{j}{L}\right]
\]

(3-16)

The values of \( c_k \) can be found using many different approaches, some of which are described in the following section. Since the \( c_k \)'s are restricted to be real whereas, \( \Theta \) and \( z \) are complex, the optimization is done as

\[
\begin{bmatrix}
\Re(z) \\
\Im(z)
\end{bmatrix} = 
\begin{bmatrix}
\Re(\Theta) \\
\Im(\Theta)
\end{bmatrix} c.
\]

(3-17)

**Non-negative Least Squares Optimization**

This uses the same approach as in DSRF. This approach has the advantage of being able to estimate \( c_k \)'s and the order \( K \) simultaneously. However, this almost always over-estimates \( K \), and needs to be modified with a sparsity constraint on the \( c_k \)'s as described in the next method.

**Convex Optimization with Sparsity Constraint**

Both DSRF and SPARDIR can be written in matrix form as \( z = \Theta c \) with \( \Theta \) being the dictionary of dielectric relaxations of interest. Each column in \( \Theta \) corresponds to a realization of a first order Cole-Cole or Debye model normalized to have a unit \( L_2 \) norm. In general, the dimension of \( z \) is lesser than the number of columns in \( \Theta \). Also the columns themselves are correlated. This leads to an over-complete problem with multiple solutions. However, if we assume that only a few \( c_k \)'s are nonzero, then we are essentially looking for the smallest order Cole-Cole or Debye model to fit our data. This transforms the problem of estimating the dielectric relaxations into a constrained optimization problem given as

\[
(P_0) \text{ Minimize } \| \tilde{c} \|_0 \text{ subject to } z = \Theta c
\]

(3-18)
with \( \|c\|_0 \) being the L_0 norm which is the number of nonzero elements in \( c \). Even though L_0 norm does not satisfy the requirements for a norm, it is still referred to as a norm.

The constrained optimization problem as given in Equation 3-18 is NP hard \([16]\). There is no direct method of optimizing with the L_0 norm constraint. However, there are three alternate approaches to solving such problems. These approaches are desirable because they can converge to the original solution for problem P_0 under certain conditions. If we let \( u \) be the parameter controlling the trade-off between fitting error and the sparsity of the final solution, then

- (P_1) L_1 constrained optimization: minimize \( u \|z - \Theta c\|_2^2 + (1 - u) \sum_{k=1}^{L} |c_k| \).  
  \( \text{(3-19)} \)

- (P_2) L_p constrained optimization:
  
  minimize \( u \|z - \Theta c\|_2^2 + (1 - u) \sum_{k=1}^{L} |c_k|^p \), with \( 0 < p < 1 \)  
  \( \text{(3-20)} \)

- (P_3) Iterative reweighted optimization:
  
  minimize \( u \|z - \Theta c\|_2^2 + (1 - u) \sum_{k=1}^{L} |\gamma_k| |c_k|^{(n+1)} \), with \( 0 < p < 1 \), and \( \epsilon \approx 0 \).  
  \( \text{(3-21)} \)

There are multiple ways of solving the above mentioned three optimization problems. Some of the famous ones are quadratic programming, linear programming, projected gradient and gradient descent.

The uniqueness of the solutions to problems P_1-P_3 depends on the correlation between the columns of \( \Theta \). If \( \mu = \max_{i,j \in [1,L], i \neq j} \Theta_i^T \Theta_j \) denotes the maximum correlation between any two columns in the dictionary, then if the number of nonzero coefficients or model order \( K \) for the true L_0 solution satisfies
\[ K = \| \tilde{c} \|_0 < 0.5 (1 + 1/\mu) \] (3-22)

[16] then the L_1 norm based solution of P_1 is the same as the L_0 norm based solution in Equation 3-18 [16]. The L_p norm based solution of P_2 can equal to the L_0 norm based solution of P_0 under more relaxed conditions. In general, the conditions depend on the Restricted Isometry Property (RIP) of the matrix \( \Theta \) [18]. It states that to produce K-sparse solutions, the matrix \( \Theta \) has to satisfy

\[
(1 - \delta_K) \| z \|_2^2 \leq \| \Theta z \|_2^2 \leq (1 + \delta_K) \| z \|_2^2 \forall z \in \mathbb{R}^L, \| z \|_0 \leq K
\] (3-23)

with \( \delta_K < 1 \). For L_1 norm solution to reach the L_0 norm solution, the condition is \( \delta_K < \sqrt{2} - 1 \). For an L_p norm solution to give a unique L_0 solution, the condition is

\[
\delta_K < 1 - \frac{1}{1 + 2(\sqrt{2} - 1)\left(1 + \frac{1}{K}\right)^{1 - \frac{1}{p}}} \quad [18].
\]

When \( p \) is close to zero, \( \delta_K \) is close to 1. So the interval \([1 - \delta_K, 1 + \delta_K]\) is larger, yielding more relaxed conditions. This makes L_p optimization to give better solutions than with \( p = 1 \). In other words, as \( p \) decreases, the condition for reaching the L_0 solution becomes easier to satisfy. However, when \( p < 1 \), the problem is no longer convex and hence achieving a global optimum depends on the algorithm and its initial conditions.

**Quadratic programming**

The optimization problems shown in Equation 3-19 and Equation 3-21 can be solved by quadratic programming which tries to minimize the function \( 0.5x^TAx + b^Tx \) with respect to \( x \) for \( x \geq 0 \). For a simple \( x \geq 0 \) constrained optimization, \( A = \Theta^T \Theta \) and \( b = -z^T \Theta \). When we add the L_1 constraint, the inputs become \( A = \Theta^T \Theta \) and \( b = -z^T \Theta + 1 \), where \( 1 \) is a vector of ones of appropriate length. For iterative reweighted optimization with a given \( p \), the input \( A \) remains the same while \( b = -z^T \Theta + \gamma \), with \( \gamma \) being a vector containing the \( \gamma_k \)'s. We used MATLAB®'s...
quadprog.m function to implement this method. This iterative schedule is summarized in the following pseudo-code

Q_IR

1. \( A \leftarrow \Theta^T \Theta \)
2. \( \gamma_k \leftarrow 1 \) for all \( k \)
3. Scale \( z \) for unit norm and positive sign
4. \( b \leftarrow -z^T \Theta + \gamma \)
5. Iteration \( \leftarrow 1 \)
6. \textbf{while} \ |\text{ObjectiveFunctionValue} - \text{PreviousObjectiveFunctionValue}| > \text{ChangeThreshold} \textbf{do}
   a. \( \text{PreviousObjectiveFunctionValue} \leftarrow \text{ObjectiveFunctionValue} \)
   b. Compute \( c \) to minimize \( 0.5 c^T A c + b^T c \)
   c. Update \( \gamma_k = 1/|c_k| \)
   d. \textbf{if} \( \gamma_k > \text{SparsityParameterThreshold} \) \textbf{then}
      i. Remove \( k^{th} \) column of \( \Theta \)
      ii. Remove \( \gamma_k \) from \( \gamma \)
      iii. Update \( A \leftarrow \Theta^T \Theta \)
   \textbf{end if}
   e. Update \( b \leftarrow -z^T \Theta + \gamma \)
   f. Update \( \text{ObjectiveFunctionValue} \leftarrow 0.5 c^T A c + b^T c \)
   g. Iteration \( \leftarrow \) Iteration + 1
7. \textbf{end while}
8. Re-scale \( c \) with original scale and sign of \( z \).

Linear programming

The linear programming approaches focus on minimizing functions of the form \( f^T x \) subject to \( z = \Theta c \). For \( L_1 \) constrained optimization, \( f \) is a vector of ones. For iterative reweighted optimization \( f = \gamma \), with \( \gamma_k \)'s being updated iteratively. This makes the approach a sequence of linear programming solutions all with the same region of feasibility. We used MATLAB®'s linprog.m function to implement this method. This iterative schedule is summarized in the following pseudo-code

L_IR

1. \( \gamma_k \leftarrow 1 \) for all \( k \)
2. Scale \( z \) for unit norm and positive sign
3. \( f \leftarrow \gamma \)
4. Iteration \( \leftarrow 1 \)
5. while |ObjectiveFunctionValue - PreviousObjectiveFunctionValue| > ChangeThreshold
   a. PreviousObjectiveFunctionValue ← ObjectiveFunctionValue
   b. Compute $c$ to minimize $f^T c$
   c. Update $\gamma_k = 1/|c_k|$
   d. if $\gamma_k > \text{SparsityParameterThreshold}$ then
      i. Remove $k^{th}$ column of $\Theta$
   end if
   e. Update $f ← \gamma$
   f. Update ObjectiveFunctionValue ← $f^T c$
   g. Iteration ← Iteration + 1
6. end while
7. Re-scale $c$ with the original scale and sign of $z$.

There are two common ways to implement linear programming namely, the simplex
method and the interior point method. The explanation of those methods is beyond the scope of
this thesis.

**Joint Sparse Estimation of Dielectric Relaxations**

All the methods mentioned so far use a single observation to identify objects. In practical
use, most often multiple measurements are taken around an object with varying orientations and
distances. If we can assume the observed data is a weighted sum of parts with each part due to a
different component, then multiple measurements around an object should contain the same set
of parts albeit with different set of weights. For example if $\theta_1$ and $\theta_2$ are responses due to two
sources, then observations are given by $z_1 = w_1 \theta_1 + w_2 \theta_2$ and $z_2 = w_3 \theta_1 + w_4 \theta_2$, with $w_1$ to $w_4$ being
the appropriate weights. Therefore it is preferable to use all the responses related to an object
together to get a more stable estimate.

The optimization now becomes a joint sparse estimation with the group of observations
coming from the same small subset of the dictionary. If $H$ denotes the Heaviside step function,
then the optimization function for $N_o$ observations is given by,
(JP0) Minimize $K = \sum_{k=1}^{L} H \left( \max_{j \in [1,N_0]} |c_{k,j}| \right)$ subject to

$$Z = [z^{(1)} \ z^{(2)} \ldots \ z^{(L)}] = \Theta [c^{(1)} \ c^{(2)} \ldots \ c^{(N_0)}] = \Theta C.$$  

(3-24)

Joint sparsity implies that the matrix $C$ is row sparse i.e. only a few rows have nonzero elements. The problem given in Equation 3-22 is NP hard, therefore approximations similar to Equations 3-18 to 3-20 can be extended in the joint case as

- (JP1) $L_{p,q}$ constrained optimization: minimize $u \| Z - \Theta C \|_p^2 + \left( 1 - u \right) \sum_{k=1}^{L} \left( \sum_{j=1}^{N_0} |c_{k,j}|^q \right)^p$,  

  with $0 < p < 1$ and $q > 1$.

- (JP2) Iterative reweighted optimization:

  minimize $u \| Z - \Theta C \|_p^2 + \left( 1 - u \right) \sum_{k=1}^{L} \gamma_k \left( \sum_{j=1}^{N_0} |c_{k,j}|^q \right)^p$,  

  $\gamma_k^{(n+1)} = \frac{1}{\left( \sum_{j=1}^{O} |c_{j,k}| \right)^{1-p}} \varepsilon$,  

  with $0 < p < 1$, $q \geq 1$ and $\varepsilon \approx 0$.

- (JP3) $L_1$ optimization: minimize $u \| Z - \Theta C \|_1 + \left( 1 - u \right) \sum_{k=1}^{L} \sum_{j=1}^{N_0} |c_{k,j}|$,  

  (3-27)

The advantage of such joint sparse optimization as opposed to optimizing each observation individually is that we can uniquely estimate a $K$ sparse solution with $K < 0.5(1/\mu + \text{Rank}(Z))$ as opposed to $0.5(1+1/\mu)$ [17]. This enables us to estimate higher order models for a given dictionary, or design dictionaries with more elements which in turn produce more accurate answers.

As in the single observation case, the quadratic programming or linear programming approaches can be used to solve Equation 3-24. The $A$ matrix remains the same, while $b^{(i)} = z^{(i)T} \Theta + \gamma^T$. The entries in the $\gamma$ vector now correspond to the sums along the rows of $C$ rather than
the individual elements in each of the columns. This definition of $\gamma$ enforces joint sparsity by using the same sparsity parameter $\gamma_k$ along all columns of a given row of $C$.

**Gradient Newton Methods**

The quadratic and linear programming methods work well when all the weights are non-negative. However, this may not be the case with all WEMI data. For example, in landmine detection using a Dipole transmitter antenna and a Quadrupole receiver antenna, each of the relaxations can have positive or negative weights. In case of multiple observations the magnitude and sign of the weights vary with respect to the relative distance and orientation between the object and the antenna setup. In such cases, a gradient descent approach can be used to solve Equations 3-24 to 3-26.

For the objective functions defined in Equations 3-24 to 3-26 of the form $uE + (1-u)R$, Newton methods are given by the update equation for time step $t+1$ as $C^{t+1} = C^t - \eta \Delta C^t$, where $\eta$ denotes the step size $\Delta C^t$ is the update which is given by

$$vec(\Delta C^t) = \left( u \frac{\partial^2 E^t}{\partial N^2} + (1-u) \frac{\partial^3 R^t}{\partial N^3} \right)^{-1} \left( u \frac{\partial E^t}{\partial N} + (1-u) \frac{\partial R^t}{\partial N} \right),$$

with $E$ denoting the quadratic or linear fitting error term, $R$ the regularization term and $v$ the vectorized version of $C$. $vec(\cdot)$ denotes the vectorization operator.

**$L_{p,q}$ Regularized Optimization**

To solve Equation 3-24, let us denote the quadratic fitting error and the regularization terms as $E = \|\Theta C - Z\|_2^2$ & $R = \left( \sum_{l=1}^L \left( \sum_{j=1}^{N_l} c_{lj} \right)^p \right)^\frac{1}{p} = \sum_{l=1}^L (R_{l,p})^\frac{1}{p} \right)$.

Taking partial derivatives, and using $\odot$ to denote element-wise multiplication,
\[
\frac{\partial E}{\partial \Theta} = -2\Theta^T (\Theta C - Z)
\]

For \( q = 2 \)

\[
\frac{\partial R}{\partial \Theta} = C \circ p \begin{bmatrix}
R_1^{p-1} \\
\vdots \\
R_L^{p-1}
\end{bmatrix}
\]

To derive the Hessian matrix, first we need to vectorize the weight matrix.

Let \( v = \text{vec}(C) \)

\[
\frac{\partial^2 E}{\partial v^2} = \begin{bmatrix}
\Theta^T \Theta & 0 & 0 & 0 & \cdots \\
0 & \Theta^T \Theta & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]

If \( k_1 = k \mod L \) and \( m_1 = m \mod L \), with \( L \) being the number of columns of \( \Theta \), \( v_k = c_{i_1,j_1}, k = (i_1-1)L + j_1 \) and \( m = (i_2-1)L + j_2 \), then the Hessian for the regularization term is given by,

\[
\frac{\partial^2 R}{\partial v_k \partial v_m} = \begin{cases}
4p(p-1)R_{ki}^{p-2}v_k^2 + 2pR_{ki}^{p-1}, & \text{if } k = m \\
4p(p-1)R_{ki}^{p-2}v_kv_m, & \text{if } k_1 = m_1 \\
0, & \text{otherwise}
\end{cases}
\]

The Hessian matrix of the gradient descent controls the speed of convergence along each of the weights. The matrix needs to be positive definite for the solution to converge. The problem in using a Newton approach directly on \( L_{p,q} \) regularized optimization is that the second order gradient matrix has \( p(p-1) \) term in the diagonal. Since \( p < 1 \), this may lead to negative diagonal entries which can result in the solution to diverge.
Iterative Reweighted Optimization

We can extend the same treatment to JP2 as we did for JP1.

\[
\begin{bmatrix}
\gamma_1 & 0 & 0 & \ldots & 0 \\
0 & \gamma_2 & 0 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \gamma_L
\end{bmatrix}
\]

Let \( G = \ldots \) \( (3-30) \)

For \( q = 2 \), using a similar approach as in \( L_{p,q} \) optimization, the equations simplify to

\[
C = u \left( u \Theta^T \Theta + (1 - u)G \right)^{-1} \Theta^T Z
\]

\( (3-31) \)

This setup provides us with a simple equation that improves the solution by iteratively updating \( G \) and \( C \). This iterative schedule is summarized in the following pseudo-code

\( \text{L2}_\text{IR} \)

1. \( \gamma_k \leftarrow 1 \) for all \( k \)
2. Scale all columns of \( Z \) for positive sign
3. Compute \( G \) according to Equation 3-30
4. Compute \( C \) according to Equation 3-31
5. Iteration \( \leftarrow 1 \)
6. while \( |\text{ObjectiveFunctionValue} - \text{PreviousObjectiveFunctionValue}| > \text{ChangeThreshold} \)
   a. PreviousObjectiveFunctionValue \( \leftarrow \) ObjectiveFunctionValue
   b. Update \( \gamma_k \leftarrow \frac{1}{\sum_{j=1}^{N_0} c_{k,j}^2} \)
   c. if \( \gamma_k > \text{SparsityParameterThreshold} \) then
      i. Remove \( k \)th column of \( \Theta \)
      ii. Remove \( \gamma_k \) from \( \gamma \)
      iii. Update \( G \) according to Equation 3-30
      iv. Update \( C \) according to Equation 3-31
   end if
   d. Update ObjectiveFunctionValue according to Equation 3-26
   e. Iteration \( \leftarrow \) Iteration + 1
7. end while
8. Re-scale \( C \) with original sign of \( Z \).
\textbf{L}_1 \text{ Optimization}

To solve Equation 3-26, let \( E = \| \Theta C - Z \| \) & \( R = \sum_{k=1}^{L} \gamma_{k} \left( \sum_{j=1}^{N_{e}} | c_{k,j} | \right) \). Directly applying second derivative approach is not possible as the derivative of \( L_1 \) norm is the sign or signum function which is discontinuous at 0 and has derivative equal to 0 elsewhere. However, the hyperbolic tangent function can be used as a smooth approximation to the signum function with a scaling constant \( \beta \). As \( \beta \) increases, the approximation becomes better. Using the approximation and taking partial derivatives, we obtain

\[
\frac{\partial E}{\partial C} = 2 \Theta^T \text{sign}(\Theta C - Z) \approx 2 \Theta^T \tanh(\beta(\Theta C - Z))
\]

(3-32)

\[
\frac{\partial R}{\partial C} = 2 \text{sign}(C) \cdot \begin{bmatrix} \gamma_1 \\ . \\ \gamma_L \end{bmatrix} \cdot \begin{bmatrix} \gamma_1 \\ . \\ \gamma_L \end{bmatrix} \approx 2 \text{tanh}(\beta C) \cdot \begin{bmatrix} 1 & \ldots & 1 \\ . & \ldots & . \\ . & \ldots & . \end{bmatrix}
\]

(3-33)

Using \( \downarrow \) to denote entries along a column, the update equation can be written as

\[
\Delta C_{\downarrow} = \left( u H_{\downarrow}^E + (1-u) H_{\downarrow} \right) \left( u \frac{\partial E_{\downarrow}}{\partial C} + (1-u) \frac{\partial R_{\downarrow}}{\partial C} \right),
\]

(3-34)

with \( H_{\downarrow}^E = G \cdot R_{\downarrow} \), where \( G \) is the diagonal matrix with \( \gamma_k \)'s as its entries as defined in the previous method and \( \cdot \) denotes the matrix product.

Let \( R_{\downarrow} = \beta \cdot \begin{bmatrix} \text{sech}^2(\beta c_{1,j}) & 0 & \ldots & 0 \\ 0 & \text{sec h}^2(\beta c_{2,j}) & 0 & \ldots \\ . & . & . & . \\ 0 & 0 & 0 & \ldots \text{sec h}^2(\beta c_{L,j}) \end{bmatrix} \)

(3-35)
Using $e_j = \Theta C - Z_{i,j}$ to denote the $j^{th}$ column of the fitting error matrix, we can write the Hessian matrix for the $L_1$ fitting error as

$$H_j^E = \beta \Theta^T \cdot \Theta$$

(3-36)

This approximation has the property that it bridges the gap between $L_1$ and $L_2$ methods. When $\beta$ is near one, the approximation is close to the $L_2$ method because $\tanh(x) \approx x$ for small values of $x$. As $\beta$ increases the results become closer and closer to the $L_1$ approach. This iterative schedule is summarized in the following pseudo-code

**L1_IR**

1. $\gamma_k \leftarrow 1$ for all $k$
2. $\beta \leftarrow 1$
3. Scale all columns of $Z$ for positive sign
4. Compute $G$ according to Equation 3-30
5. Compute $C$ according to Equation 3-31
6. Iteration $\leftarrow 1$
7. while $|\text{ObjectiveFunctionValue} - \text{PreviousObjectiveFunctionValue}| > \text{ChangeThreshold}$
   a. PreviousObjectiveFunctionValue $\leftarrow$ ObjectiveFunctionValue
   b. Compute $\partial E / \partial C$ according to Equation 3-32
   c. Compute $\partial R / \partial C$ according to Equation 3-33
   d. Compute $R_j$ according to Equation 3-35
   e. Compute Hessian of regularization term $H_j^R \leftarrow G.R_j$
   f. Compute Hessian of absolute error term according to Equation 3-36
   g. Update $C$ according to Equation 3-34
   h. Update $\gamma_k \leftarrow \frac{1}{N_{\gamma}} \sum_{j=1}^{N_{\gamma}} |C_{k,j}|$
   i. if $\gamma_k > \text{SparsityParameterThreshold}$ then
      i. Remove $k^{th}$ column of $\Theta$
      ii. Remove $\gamma_k$ from $\gamma$
      iii. Update $G$ according to Equation 3-30
iv. Update \( \mathbf{C} \) according to Equation 3-34

\[ \text{end if} \]

j. Update \( \text{ObjectiveFunctionValue} \) according to Equation 3-27

k. \( \text{Iteration} \leftarrow \text{Iteration} + 1 \)

8. \( \text{end while} \)

9. Re-scale \( \mathbf{C} \) with original sign of \( \mathbf{Z} \).

**Dictionary Learning using Adaptive Kernels**

The set of methods discussed so far use a fixed dictionary. In most parametric modeling approaches, the dictionary is populated using a specific model whose parameters are varied to create dictionary elements. In the present case the model is a first order Cole-Cole model. The task of creating a dense enough dictionary to represent data of interest while keeping the correlation between the elements small is not trivial. For example, the Cole-Cole dictionary for the frequencies between 300 Hz to 90000 Hz equispaced in log scale can have only a maximum of 11 elements for a correlation threshold of 0.9. This is highly restrictive and still the correlation is too high. One solution to alleviate this problem would be to make the dictionary tuned to the data. Let \( \theta_k(\omega) = \frac{1}{1 + (i\omega\tau_k)^{\alpha_k}} \) denote \( k \)th dielectric relaxation. By extending the same gradient descent approach to the regularized error function defined in Equation 3-26, we get for \( L_2 \) error,

\[
\frac{\partial E}{\partial \log \tau_k} = 2 \left( \frac{\partial \theta_k}{\partial \log \tau_k} \right)^T (\Theta \mathbf{C} - \mathbf{Z}) \quad \& \quad \frac{\partial R}{\partial \tau_k} = 0
\]

(3-37)

with

\[
\frac{\partial \theta_k(\omega)}{\partial \log \tau_k} = -\frac{\alpha_k(i\omega\tau_k)}{\left(1 + (i\omega\tau_k)^{\alpha_k}\right)}
\]

(3-38)

\[
\frac{\partial E}{\partial \alpha_k} = 2 \left( \frac{\partial \theta_k}{\partial \alpha_k} \right)^T (\Theta \mathbf{C} - \mathbf{Z}) \quad \& \quad \frac{\partial R}{\partial \alpha_k} = 0
\]

(3-39)

with

\[
\frac{\partial \theta_k(\omega)}{\partial \alpha_k} = -\frac{(i\omega\tau_k)^{\alpha_k} \log(i\omega\tau_k)}{\left(1 + (i\omega\tau_k)^{\alpha_k}\right)^2}
\]

(3-40)
Note that since \( \tau \) varies in exponential scale, the update is done using \( \log_e(\tau) \). The Hessians are given by

\[
\frac{\partial^2 \theta_k(\omega)}{\partial \log \tau_k \partial \log \tau_m} = \begin{cases} 
\alpha_k^2 (i \omega \tau_k)^{\alpha_k} \left( -1 + (i \omega \tau_k)^{\alpha_k} \right) \left( 1 + (i \omega \tau_k)^{\alpha_k} \right), & k = m \\
0, & \text{otherwise}
\end{cases},
\]

\[
\frac{\partial^2 \theta_k(\omega)}{\partial \alpha_k \partial \alpha_m} = \begin{cases} 
(i \omega \tau_k)^{\alpha_k} \left( -1 + (i \omega \tau_k)^{\alpha_k} \right) \left( \log(i \omega \tau_k) \right)^2, & k = m \\
0, & \text{otherwise}
\end{cases}
\]

Since the diagonal elements of the Hessian cannot be guaranteed to be positive, only a first order update is used. Also, since the parameters are restricted to be real, only the real part of the update is used. Using the same approach as Equations 3-37 to 3-40, the update equations for L1 error are given by

\[
\frac{\partial E}{\partial \log \tau_k} = 2 \left( \frac{\partial \theta_k}{\partial \log \tau_k} \right)^T \tanh(\beta(\Theta C - Z)) \quad \& \quad \frac{\partial R}{\partial \tau_k} = 0
\]

(3-41)

and

\[
\frac{\partial E}{\partial \alpha_k} = 2 \left( \frac{\partial \theta_k}{\partial \alpha_k} \right)^T \tanh(\beta(\Theta C - Z)) \quad \& \quad \frac{\partial R}{\partial \alpha_k} = 0
\]

(3-42)

The dictionary elements are adapted along with their corresponding weights as described below

AKL1_IR

1. \( \gamma_k \leftarrow 1 \) for all \( k \)
2. \( \beta \leftarrow 1 \)
3. Scale all columns of \( Z \) for positive sign
4. Compute \( G \) according to Equation 3-30
5. Compute \( C \) according to Equation 3-31
6. Iteration \( \leftarrow 1 \)
7. while \(|ObjectiveFunctionValue - PreviousObjectiveFunctionValue| > ChangeThreshold\)
   a. \( \beta \leftarrow 1.005 \beta \)
   b. PreviousObjectiveFunctionValue \( \leftarrow \) ObjectiveFunctionValue
   c. Compute \( \partial E / \partial C \) according to Equation 3-32
   d. Compute \( \partial R / \partial C \) according to Equation 3-33
e. Compute $R_j$ according to Equation 3-35
f. Compute Hessian of regularization term $H_j^R \leftarrow G.R_j$
g. Compute Hessian of absolute error term according to Equation 3-36
h. Update $C$ according to Equation 3-34
i. Update $\gamma_k \leftarrow \frac{1}{N_{\tilde{u}} \sum_{j=1}^{N_{\tilde{u}}} |c_{k,j}|}$

j. if $\gamma_k > \text{SparsityParameterThreshold}$ then
   i. Remove $k^{th}$ column of $\Theta$
   ii. Remove $\gamma_k$ from $\gamma$
   iii. Update $G$ according to Equation 3-30
   iv. Update $C$ according to Equation 3-34
   v. $\beta \leftarrow 1$
end if
k. Update $\tau_k$ according to Equation 3-41
l. Update $\alpha_k$ according to Equation 3-42
m. Update ObjectiveFunctionValue according to Equation 3-27
n.Iteration $\leftarrow$ Iteration + 1

8. end while
9. Re-scale $C$ with original sign of $Z$.

Double Dictionary Search

In most practical cases the joint sparse estimation is done over observations that are generated by collecting data by varying the distance and orientation of the sensor setup with respect to the object. The observations are then related to one another by the response function of the transmitter and the receiver.

For example, Figure 3-7 shows the weights of individual relaxations for a high metal anti-tank mine along a path known as down-track. We can see that the weights themselves follow a pattern that is dependent on the dipole transmitter and quadrupole receiver configuration. We can create another dictionary of such shapes and constrain the weights to a linear combination of them as $C = (HW)^T$, with $H$ being the dictionary of down-track shapes. The matrix $W$ now serves as row sparse weight matrix for the dielectric relaxations and picks the down-track shapes by the column values along such rows. By designing an appropriate dictionary, we can get
accurate estimates by a small number of such shape elements. Each column of $\mathbf{W}$ is sparse but not necessarily jointly sparse along multiple rows. The optimization is done using

$$E = \left\| \mathbf{W}^T \mathbf{H}^T - \mathbf{Z} \right\|_2^2 \quad \text{and regularization} \quad R = \sum_{j=1}^{L} \lambda_j \sum_{i=1}^{N_{\text{shapes}}} W_{ij}^2$$

(3-43)

Figure 3-7. Estimated relaxations A) $\tau = 10^{-5}$, B) $\tau = 10^{-4}$ and C) $\tau = 10^{-3.4}$ and their weights for a high metal anti-tank mine along a given direction across the object.

The update is done as a two step process, first with optimizing $\mathbf{C}$ with respect to the data to be a row sparse matrix and then next optimizing $\mathbf{W}$ to be generally sparse with respect to the estimate of $\mathbf{C}$. The weights are initialized using least squares pseudo-inverse solution

$$\mathbf{W} = \left( \mathbf{H}^T \mathbf{H} + \lambda \mathbf{I} \right)^{-1} \mathbf{H}^T \mathbf{Z} \Theta (\Theta^T \Theta + \lambda \mathbf{I})^{-1}$$

with $\lambda$ controlling the ill-posedness of the matrix inverse operation.
CHAPTER 4
RESULTS

We implemented and tested the four algorithms and their variants discussed in the previous chapter. In the following sections we discuss the test results of our proposed methods. First, to illustrate the effectiveness of one method against another, synthetic data sets are created with varying levels of complexity in the spectral shapes. Then, the methods are compared on a set of landmine data collected at two different locations.

Comparison on Synthetic Data

Experiment 1: Data from first order model

Figure 4-1. Real vs Imaginary part of the induced response of a sample generated using Cole-Cole model with increasing levels of additive noise.

To compare different algorithms that estimate the parameters of the 1st order Cole-Cole model, 100 instances of data were generated using the 1st order Cole-Cole model with the four parameters $a$, $s$, $\tau$ and $\alpha$ randomly sampled from uniform distributions over the intervals $(10^{-8},1)$, $(-10,10)$, $(10^{-7},10^{-1})$ and $[0.1,1.4]$ respectively at 21 frequencies ranging from 330 Hz to 90300
Hz in log scale. These were typical values seen in the landmine detection framework. The experiments were repeated with zero mean Gaussian noise of variance \( \sigma^2(\omega) \) added at each frequency. More precisely, the noisy versions of the signal are given by,

\[
z'(\omega) = z(\omega) + N(0, \sigma^2) \text{ where } \sigma = \rho \sqrt{ \sum_{k=1}^{N/2} [\Re[z(\omega_k)]^2 + \Im[z(\omega_k)]^2] }, \text{ with } \rho = [10^{-4}, 10^{-3}, 10^{-2}].
\]

The effect of additive noise on the signal is shown by Argand diagrams in Figure 4-1. We see that \( \rho = 10^{-3} \) was the point where noise started dominating the signal.

**Observations**

Figure 4-2 shows the fitting error as a function of the signal energy for different parameter estimation methods. The circle fitting method by Bishay [9] worked well and was the fastest method in the noise free case. However its performance degraded quickly in the presence of noise. Figure 4-3 and Figure 4-4 compare the estimated values in the y axis to the true values in the x axis. Therefore, a good estimate should lie on the \( y = x \) line. Figure 4-3 shows that even for a low fitting error, the estimated parameters might be far from the true underlying parameters. It also shows that in noise free case, the estimation of the parameter \( \alpha \) was to be the least reliable one. GRANMA produced the most robust results for noise levels up to \( \rho = 10^{-3} \). When \( \rho = 10^{-2} \), all methods fail most of the time.

The experiments were carried out in MATLAB on a Linux based server with 10GB memory and quad-core processor with each core operating at 2Ghz. Bishay’s circle fitting method was the fastest with speeds about 10x faster than GRANMA which was in turn faster than the iterative least squares method by 10x - 20x. The Xiang inversion method varied in speed from 2x – 10x slower than GRANMA , depending on the number of iterations it required to converge.
Figure 4-2. Fitting Error vs. Signal Energy for different parameter estimation methods of Cole-Cole model at different noise levels.

Figure 4-3. Actual vs. estimated parameter values for different Cole-Cole parameter estimation methods in the noise free case.
Figure 4-4. Actual vs. estimated parameter values for different Cole-Cole parameter estimation methods when $\rho = 10^{-4}$. GRANMA produced the most robust estimates and Bishay and Xiang methods started to degrade in performance.

**Experiment 2: Comparison of Joint Sparse Methods on Data from second order model**

The methods discussed in the previous experiment assume that the data come from a 1<sup>st</sup> order Cole-Cole model. In most practical applications, the spectral shape of the data follows higher order models. For example, we observed most low metal anti-personnel mines in our data set exhibit behavior that is best described by a 2<sup>nd</sup> or 3<sup>rd</sup> order Cole-Cole model. In this experiment, we compared the different dictionary search methods to simultaneously estimate the order of the model along with the parameters of the model. The objective functions for the different algorithms seek to produce models that accurately represent a set of observations with a small number of dictionary elements.

To test this property, 100 different synthetic data sets were created using the same procedure. A dictionary was created by varying $\tau$ and $\alpha$ in the range $(10^{-9}, 10)$ and $(0, 1)$. The dictionary was pruned so that no two elements had correlation greater than 0.96. This dictionary was used for all 100 trials. In each trial, a two-step process was used to generate synthetic data.
First, two random dictionary elements were selected from an (almost) uniform distribution over the pruned dictionary. They were restricted to have correlation less than 0.3. Then, a data set of 25 spatially sequential data vectors was then generated by varying the weights of the two relaxations using two differentiated Gaussian shapes in the down-track direction. The parameter estimation algorithms were then applied to the data for 50,000 iterations with high weight given to the sparsity terms of the objective functions.

![Histogram image of Number of terms vs. Fitting error for different dictionary search methods.](image)

Figure 4-5. Histogram image of Number of terms vs. Fitting error for different dictionary search methods. From the top left to the bottom right, A) $L_2$ error with iterative reweighted $L_2$ regularization, B) approximate $L_1$ error with iterative reweighted approximate $L_1$ regularization, C) quadratic programming with iterative reweighted $L_1$ regularization, D) linear programming with iterative reweighted $L_1$ regularization, E) quadratic programming with fixed $L_1$ regularization, F) linear programming with fixed $L_1$ regularization.

To compare the effectiveness of the algorithms, the histograms of fitting errors for different numbers of nonzero terms is plotted in Figure 4-5. The plots are created by first plotting
parametric curves for each trial. Each curve plots number of nonzero rows of the weight matrix vs. fitting error and the parameter is the iteration index. Histograms are then created by overlaying all the parametric curves in one plot and counting how many times each cell is covered. Here we observed that our L₁ and L₂ based methods outperformed off the shelf quadratic and linear programming methods.

**Experiment 3: Comparison of Cole-Cole and DSRF Dictionaries**

In this experiment, the effect of choosing the right model to form the dictionary was studied. Two dictionaries were designed based on the DSRF (α = 1) and Cole-Cole (α not necessarily 1) models respectively. We refer to these two dictionaries as the DSRF and Cole-Cole dictionaries. The maximum correlation between elements was restricted to 0.96 and both had 11 elements each. This experiment had two parts: (1) synthetic data were generated from the Cole-Cole model and α not close to 1 and (2) synthetic data were generated from the DSRF model. In each part, we compare the two dictionaries by analyzing their convergence (using the L₁_IR method) to the correct number of parameters and the closeness to the true parameters. As before, each part of the experiment consisted of 100 trials.

In part (1), data were generated for each trial by sampling the number of relaxations from a uniform distribution over {1, 2, 3}. Once the number of relaxations was selected, then that number of relaxations was sampled by sampling τ and α uniformly over the intervals (10⁻⁷, 1) and (0.2, 0.8) respectively. The relaxations were combined using a preset combination of weights to generate 25 data vectors as in the previous experiment. Gaussian noise with 10% of the energy in the signal was added. Since both DSRF and Cole-Cole dictionaries used the same number of elements, this experiment compares the effect of offsets in the τ and α spaces respectively. The convergence statistics are given in Table 4-1.
Table 4-1. Comparison of convergence statistics: Part (1).

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>Convergence to true number of relaxations</th>
<th>Converged to the closest relaxations in the parameter space</th>
<th>Converged to relatively closer distance in the parameter space</th>
<th>Performed the same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cole-Cole</td>
<td>47%</td>
<td>0%</td>
<td>38%</td>
<td>53%</td>
</tr>
<tr>
<td>DSRF</td>
<td>9%</td>
<td>0%</td>
<td>9%</td>
<td></td>
</tr>
</tbody>
</table>

The meanings of columns two and three seem clear. The fourth and fifth columns of Table 4-1 require some discussion. The spacing between dictionary elements in the parameter space varies non-uniformly because the parameter quantization depends on the correlation between the elements. The spacing must be taken into account when calculating distance between the true and estimated parameters since no algorithm can overcome the inherent quantization error. More discussion of this will be given later. The fourth column measures the percentage of trials for which each algorithm was closer to the true parameter set. The fifth column is the percentage of cases for which they both the same distance from the true parameters. The DSRF dictionary based method converged to the true number of relaxations only 9% of the time compared to the Cole-Cole dictionary based method which converged 47% of the time.

In the second part, it was required that $\alpha = 1$. Data were then generated in each trial with one, two or three relaxations with $\tau$ uniformly sampled over the interval $(10^{-7}, 1)$ as described in part (1). As can be seen in Table 4-2, both dictionaries produce the correct number of relaxations but the DSRF errors are smaller more often. This should be expected because the quantization in the line $\alpha = 1$ is coarser for Cole-Cole than DSRF.
Table 4-2. Comparison of convergence statistics: Part (2)

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>Convergence to true number of relaxations</th>
<th>Converged to the closest relaxations in the parameter space</th>
<th>Converged to relatively closer distance in the parameter space</th>
<th>Performed the same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cole-Cole</td>
<td>100%</td>
<td>5%</td>
<td>15%</td>
<td>10%</td>
</tr>
<tr>
<td>DSRF</td>
<td>100%</td>
<td>21%</td>
<td>69%</td>
<td></td>
</tr>
</tbody>
</table>

Experiment 4: Dictionary update

The previous experiment assumed that the sample $z$ is truly a linear combination of one or more of the dictionary elements. This however is not practical as the correlation between elements dictates the size of the dictionary and hence its resolution in the parameter space. Therefore, in this experiment the dictionary is adapted to the data using the adaptive kernel matching pursuit approach described in the technical approach. In this experiment, two relaxations were used for every trial. They were selected by sampling $\tau$ and $\alpha$ uniformly over the intervals $(10^{-7}, 1)$ and $(0, 1)$ respectively. Figure 4-7 (f) is interesting. Each different colored curve represents the distance of a particular dictionary element from the nearest true relaxation in the parameter space. At the iteration for which a dictionary element is pruned, the curve ceases since that element is no longer part of the solution. By the end of the iterations, there are only two curves remaining, each of which have distance value almost zero. Thus, the algorithm converges to the true relaxations and only the true relaxations.
Experiment 5: Double Dictionary search – Proof of concept

To prove the concept of using double dictionary search we generated data with 2 relaxations taken from a dictionary of 23 elements. A down-track shape dictionary of 100 Gaussian pulses with different means and variances is designed with each pulse of duration 49 samples. This led to a total of 49 observations with their relaxations changing according to two opposing Gaussian pulse shaped signals with means at locations 22 and 32 respectively. Zero mean Gaussian random noise with 10% relative energy compared to the signal is added to the data to test the robustness of the estimates.
Figure 4-7. Weights of individual relaxations using double dictionary search. The estimated values are noisy versions of the original shape.

Figure 4-8. Two nonzero rows of the weight matrix $W^T$ showing row sparsity and nonzero values in columns corresponding to the down-track shape element.

Figure 4-7 and Figure 4-8 show that the results are accurate. This shows that a dictionary of down-track shapes for objects of interest can be designed for a given sensor configuration. Since this method is highly specific to data sets with particular sensor configurations, it is reserved for future work.
**Experiment 6: Dictionary Correlation Analysis**

The effectiveness of any dictionary based search depends on the correlation between its elements. This is because most dictionary search methods use the L₂ norm of the fitting error. Since the data and the elements are normalized to have a unit L₂ norm, the fitting error becomes dependant only on the correlation. Therefore, in an L₂ norm based solution, two elements with high correlation between them are almost indistinguishable.

Figure 4-9. Correlation regions of a Cole-Cole dictionary region. The black asterisks represent the DSRF dictionary elements chosen. Each of the different colors represent the regions that have 0.96 correlation or more with at least one or more of the dictionary elements. None of the dictionary elements can be used.

Figure 4-9 shows the regions in the Cole-Cole parameter space that are correlated with the DSRF dictionary elements. This plot demonstrates that the Cole-Cole dictionary must be coarser along the line $\alpha = 1$ or else no dictionary elements with $\alpha \neq 1$ can be used.
Comparison on Landmine Data

Data Sets

The experiments were performed on two data sets. Each data set contains measurements over buried mines, buried clutter objects and blanks. The first set contains 62 different types of objects including 26 different types of mines collected over 11 adjoining lanes divided into 220 grid cells. The second set contains 24 different types of objects including 12 different types of mines collected over 12 adjoining lanes divided into 225 cells. Table 4-1 shows the distribution of objects in the data. The data were collected in two opposite directions over the same set of grid locations using a Wideband EMI sensor developed by W. Scott [14]. The object is assumed to be at the center of the grid cell and position errors were fixed manually.

Table 4-3. Nomenclature and Proportion

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Set 1</td>
</tr>
<tr>
<td>HMAP</td>
<td>High Metal Anti-Personnel mine</td>
<td>10%</td>
</tr>
<tr>
<td>LMAP</td>
<td>Low Metal Anti-Personnel mine</td>
<td>21.4%</td>
</tr>
<tr>
<td>HMAT</td>
<td>High Metal Anti-Tank mine</td>
<td>3.6%</td>
</tr>
<tr>
<td>LMAT</td>
<td>Low Metal Anti-Tank mine</td>
<td>15.9%</td>
</tr>
<tr>
<td>HMC</td>
<td>High Metal Clutter</td>
<td>15.5%</td>
</tr>
<tr>
<td>MMC</td>
<td>Medium Metal Clutter</td>
<td>0%</td>
</tr>
<tr>
<td>LMC</td>
<td>Low Metal Clutter</td>
<td>0%</td>
</tr>
<tr>
<td>NMC</td>
<td>Non-Metallic Clutter or Blank</td>
<td>33.6%</td>
</tr>
</tbody>
</table>

In the first data set, each grid cell is of dimension 1.5m x 1.5m and in the second one; they are 1m x 1m. The WEMI sensor collects complex responses in 21 frequencies at 1cm intervals in 3 channels. This results in a 21 x 3 x 150 or 21 x 3 x 100 complex data matrix for each cell.
Data Filtering

Figure 4-10. Down-track filter template

The measured data are filtered in the down-track direction by convolving with a zero mean template shown in Figure 4-10. This is used to remove ground response and sensor drift and also to improve signal to noise ratio. Filtered data is used in our experiments for classification of objects modeled with a 1st order Cole-Cole model. Only the data from the center channel and at the center of the cell is used. This makes the first data set of size 220 grid cells x 2 directions and the second data set of size 225 x 2. Unfiltered data is used in dictionary based feature extraction methods as it represents the underlying physical phenomenon better.

Experiment 6: Classification of objects using first order model

Experiment setup

A 10 fold cross-validation was performed. Care was taken to avoid having the same cell represented in both training and test sets simultaneously. To get fair representation of all objects in the training data, object stratification was used to create the training and testing folds. This method groups objects by type and splits them equally in all folds.
Experiments and observations

The experiments were carried out for different algorithms in their increasing order of complexity.

PRAM analysis

In this approach, first the gradient angles were computed for each grid cell. Since there are 28 different mines are in the data, one prototype per mine type was created from the training data as per Equation 3-2. The non-mines were considered as outliers and not given prototypes. This led to a one class classification approach with a test sample belonging to the mine class with confidence given by Equation 3-3.

GRANMA analysis

Figure 4-11 plot showing the relationship between $\tau$, $\alpha$ and $E_L$ for different mine types.
In this approach, the landmine data were assumed to follow first order Cole-Cole models and the parameters were extracted using GRANMA. The lookup table estimates of $\tau$ and $\alpha$ for these data sets are shown in Figure 4-11. We have already seen from Figure 3-3 and the corresponding section that, the error surface and the update gradients are highly nonlinear especially near regions where $\omega\tau = 1$ and $\alpha \approx 2$. But, Figure 4-12 shows that searching for $\alpha$ between 0.1 and 1.4 (in linear scale) and for $\tau$ in the range of $10^{-6}$ to 1 (in log scale) was sufficient for the data in the present experiments. This justifies the choice of using a lookup table method. Future research may be targeted towards using alternative methods to a lookup table such as regression and direct inversion.

![Figure 4-12. $\tau$ and $\alpha$ for different object types.](image)

The features are $\log_{10}(A)$, $s$, $\log_{10}(\tau)$, $\alpha$, $\log_{10}(E_L)$ and $E_F$ making a feature set of size 220 x 2 for the first data set and of size 225 x 2 for the second data set.
Feature vector

\[ F = \begin{bmatrix} \frac{\log_{10}(\tilde{r})}{\tau_{\text{max}}} & \frac{\hat{\alpha}}{\alpha_{\text{max}}} & \frac{\log_{10}(E_{\text{max}})}{A_{\text{max}}} & \frac{\log_{10}(\hat{A})}{s_{\text{max}}} & \hat{s} & E_F \end{bmatrix} \] (4-1)

where \( A_{\text{max}}, s_{\text{max}}, \tau_{\text{max}}, \alpha_{\text{max}} \) and \( E_{\text{max}} \) are scaling parameters.

Figure 4-12 shows that for different object types form their own clusters in the \((\alpha, \tau)\) space whereas non-metallic clutter (NMC) is spread all over. This is because there is no specific region in the lookup table that matches with NMC.

**Bishay method and Xiang inversion analysis**

In this approach, the landmine data was assumed to follow first order Cole-Cole model and the parameters were extracted using Xiang inversion and Bishay method and the classification is done using the same soft kNN method described for GRANMA.

**Feature Selection**

Since the Xiang inversion and Bishay method produced 5 features and GRANMA produced 6 features, each algorithm is evaluated for all its feature subsets. This makes \(2^6-1 = 63\) classification experiments for GRANMA and \(2^5-1 = 31\) experiments each for the Bishay method and Xiang inversion method. The results are compared using best ROC curves.

**Observations**

GRANMA outperformed the other Cole-Cole feature extraction methods in terms of classification accuracy while Bishay method was the fastest. Even though all methods estimated the same features and used same type of classifier, the consistency in parameter estimates contributed to the difference in performance.
Figure 4-13. ROC curves of different landmine detection algorithms with their corresponding best features. ErrLookup and ErrFit denote the lookup table error and the fitting error respectively as defined in Equation 3-8 and Equations 3-11.

Experiment 7: Analysis of GRANPA on landmine data

Figure 4-14. Change in Argand diagram with distance for a Low Metal Anti-Personnel Mine buried at 1". 
Figure 4-15. Change in amplitude with distance for a Low Metal Anti-Personnel Mine buried at 1”. The location nearest to the object is noted by the blue ellipse. The x axis denotes observation number along down track.

Figure 4-14 shows the variation in the spectral shape of unfiltered data as measurements are taken at increasing distance from the object. GRANMA and other first order approximation methods model the far field well and produce consistent parameter estimates for different instances of the same type of object. However, in the near field, data looks more piecewise first order. Figure 4-15 shows the variation in amplitude with distance. It shows that the amplitude crosses zero closest to the object. Also, the observation closest to the zero crossing shows the maximum near field effect. Therefore GRANPA was used to extract parameters from the observation closest to the zero crossing and used to initialize an iterative search procedure to connect the parameters of near and far field. This iterative schedule is summarized in the following pseudo-code

**GRANPA_ILS**

1. Find zero crossing location \( d_0 \)
2. Choose either positive or negative amplitude side
3. Estimate order and parameters using GRANPA \( S_0 \)
4. \( \text{CurrentObservation} \leftarrow \text{Start of Positive Amplitude Side} \)
5. \( S_{\text{start}} \leftarrow S_0 \)
6. \( \text{while CurrentObservation} < \text{EndOfPositiveAmplitudeSide} \)
   a. \( \text{CurrentObservation} \leftarrow \text{Move to next observation} \)
   b. Use iterative least squares with initial conditions \( S_{\text{start}} \)
Figure 4-16. Change in parameter values for a second order approximation estimated using GRANPA. The x axis denotes observation number along down track. The amplitude zero crossing happens at observation number 90.

Figure 4-16 shows that as the distance from the object increases, the parameter values change. While it is expected that the amplitude and the shift can change with distance, the values of \( \tau \) and \( \alpha \) should not as they are dielectric properties of the object under study.
Figure 4-17. Parameter values estimated using DSRF along down-track. The x axis denotes distance along down track. The amplitude zero crossing happens at location 0.8m from the start. The horizontal lines correspond to parameters estimated using joint sparse optimization.

Figure 4-17 shows the difference between individual and joint sparse optimization using DSRF. Without the joint sparsity constraint, the dictionary elements selected at different locations are free to change.
Figure 4-18. Parameter values estimated using joint sparse DSRF along down-track. The x axis denotes observation number along down track. The y axes denote the weight of corresponding dielectric relaxations. The amplitude zero crossing happens at observation number 35 from the start of the segment.

Figure 4-18 shows DSRF parameters estimated on a segment of the same data using a joint sparsity constraint. This constrains the $\tau$ values to be constant along the down-track and only the amplitudes of the relaxations change.
Figure 4-19. Relaxations estimated using a Cole-Cole dictionary on the left vs. those estimated using a DSRF dictionary. The x axis corresponds to different cell locations for the same mine type.

Figure 4-20. Usage of dictionary elements by the landmine data. We want similar mines to use the same dictionary elements and hence lesser number of them.
Figure 4-19 shows parameters estimated using joint sparse dictionary search method with a Cole-Cole model dictionary on the left and a Debye model based dictionary used by DSRF on the right. The number of elements in each dictionary is fixed so that the comparison is done on the model used rather than resolution of the dictionary. The x axis shows different ground locations of the same landmine type. Figure 4-20 shows the number of times each dictionary element was selected for the entire landmine data set. Since it is desirable to have not only sparse estimates but also the same estimates at different conditions for the same object, the entropy of the dictionary element usage is shown in Figure 4-21. The Cole-Cole model is not just able to produce a sparser representation than the DSRF, but also a more consistent one.

![Figure 4-21. Entropy of estimated relaxations for 28 different mine types.](image-url)
CHAPTER 5
CONCLUSIONS

We have developed four algorithms to detect, analyze and identify objects using wideband electromagnetic induction data. The first and simplest PRAM method is quick and doesn’t make any assumptions about the data. It removes the variability in the amplitude and drifts in the data and has been found to be useful in designing a pre-screener to mark interesting sections of ground for further analysis. The second method GRANMA gives a robust way of estimating the Cole-Cole parameters and provides consistent features that can be used for object classification. The third method GRANPA uses the same framework as GRANMA and is a quick way of modeling near field effects. The fourth method SPARDIR separates the sensor setup from the dielectric properties of objects and provides a framework to study them.

Future work may be targeted towards speeding up GRANMA by searching only a section of the lookup table using the range of the gradient angles. More statistical analysis on the convergence of the developed L₁ version of SPARDIR needs to be done. Another area of promise is using SPARDIR to extract features identify objects.
LIST OF REFERENCES


BIOGRAPHICAL SKETCH

Ganesan Ramachandran received the Bachelor of Engineering degree in Electronics and Communication Engineering from the Regional Engineering College, Tiruchirappalli, Tamil Nadu, India in May 1999. He received his Master of Science and Doctor of Philosophy degrees in Electrical and Computer Engineering from the University of Florida in August 2002 and May 2010 respectively.

Currently, he is a research assistant in the Computational Science and Intelligence Lab in the Computer and Information Sciences and Engineering Department at the University of Florida. Research includes the development of algorithms, methodologies, and models with a solid mathematical and/or statistical base with applications to landmine detection. Current and previous research applies these methods to a variety of data including electromagnetic induction, radar, weather and biomedical.