Sensor Networks: Studies on the Variance of Estimation, Improving Event/Anomaly Detection, and Sensor Reduction Techniques Using Probabilistic Models

Philip Allen Chin

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Mechanical Engineering

Michael J. Roan
Alexander Leonessa
Cory M. Papenfuss

(June 15, 2012)
Blacksburg, VA

Keywords: direction of arrival, source localization, Cramer-Rao bounds, distributed sensor network, sensor reduction, event monitoring, event detection, anomaly detection, direction of travel, joint probability, geospatial intelligence, correlation, classifier

Sensor Networks: Studies on the Variance of Estimation, Improving Event/Anomaly Detection, and Sensor Reduction Techniques Using Probabilistic Models

Philip Allen Chin

ABSTRACT

Sensor network performance is governed by the physical placement of sensors and their geometric relationship to the events they measure. To illustrate this, the entirety of this thesis covers the following interconnected subjects: 1) graphical analysis of the variance of the estimation error caused by physical characteristics of an acoustic target source and its geometric location relative to sensor arrays, 2) event/anomaly detection method for time aggregated point sensor data using a parametric Poisson distribution data model, 3) a sensor reduction or placement technique using Bellman optimal estimates of target agent dynamics and probabilistic training data (Goode, Chin, & Roan, 2011), and 4) transforming event monitoring point sensor data into event detection and classification of the direction of travel using a contextual, joint probability, causal relationship, sliding window, and geospatial intelligence (GEOINT) method.

DEDICATION

I would like to dedicate this thesis first and foremost to my loving wife, and secondly to my two sons and my parents. All of this is for you.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank Dr. Roan for his firm lessons in improving my research skills. Without this, I would undoubtedly have not been prepared for the future. I also thank him for his great ability to maintain research funding for students like myself. Thanks to Dr. Leonessa for creating a healthy research environment, and thanks to Dr. Papenfuss for heavily facilitating empirical research skills through leading by example.

This research was supported by U.S. Army Engr. R&D Center contract number W913E5-09-C-0006 and U.S. Department of Defense Contract No. N00014-09-1-0609.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Research Motivation

Monitoring and surveillance applications are increasingly being researched and pursued. Distributed sensor networks can be implemented to improve knowledge about events at remote locations. The sensor data depends on the types of sensors, physical placement, noise of the surrounding environment, and the number of sources. If sensors are implemented and placed correctly, the data from a sensor network can be used for detection, classification, and alerting of unwanted events. The goal of this thesis is to use signal processing and statistical methods to model data to increase sensing capabilities.

The placement of sensors to obtain precise measurements is a key issue when designing and planning a sensor network. The goal is to place the sensor network in a manner such that the variance of the error of the measurements will be at minimum. It is required that the parameter estimation of the variance of the error possess two qualities: 1) minimum variance, 2) unbiased. The Cramer-Rao lower bound is the minimum variance unbiased (MVU) estimator for a signal in white Gaussian noise and guarantees, if it exists, that no unbiased estimator can exist whose variance is lower. Two case studies of acoustic sensing and the Cramer-Rao lower bound of the measurement error are presented. The first study derives the Cramer-Rao bounds for the direction of arrival (bearing only) of two narrowband acoustic sources using one uniform linear array. The second study derives the Cramer-Rao Bounds for source localization (bearing and range) of one narrowband acoustic source using two uniform linear arrays.

The improvement of existing distributed sensor networks are also topics of great interest. Current topics in the research area of sensor placement and reduction include maximizing coverage, cost efficiency, and hierarchical communication schemes. In this thesis, however, a method is presented to reduce sensors based off of 1)guarding a target destination and 2)known typical environmental dynamics. Using the above prior known information, sensors in a new or existing network can be placed to achieve: 1)more efficient data processing, 2)higher performance of the signal processing application, and 3)a decrease in the number of material resources used.

The methods chosen to place and reduce the number of sensors can directly affect the method chosen to conduct event monitoring, event detection, and anomaly detection. Event monitoring networks such as precipitation and temperature stations have been implemented successfully, gaining general trend knowledge about a specific process. For example, Gaussian processes and correlation coefficient (normalized covariance) techniques are used to characterize the relationships between multiple sensor locations. Gaussian processes model the data as mixtures of multiple Gaussian distributions, whereas correlation coefficient techniques involve learning the covariance matrix. Instead of learning general trends, the focus has increasingly turned to the detection of specific events. This area of research is called event detection. Current event detection methods rely strongly upon time correlation, convolution, correlation coefficient, and pattern matching techniques. Time correlation is a good example of a pattern matching technique where a sensor's data is matched with another to find the best matching lead/lag time. The main goal is to learn a model of typical sensor data so that specific events and deviation away from normal activity can be detected. The subject of finding

deviation from normalcy is commonly referred to as anomaly detection, which is a subgroup research area of event detection. Chapter 8 is an example of applying a statistical distribution (Poisson) model to data to improve anomaly detection performance.

Although pattern matching and correlation coefficient methods are commonly applied in the area of event detection, their weakness can be found in their problem formulation. These methods assume that there is only one underlying process that is generating the data which is measured by the sensors. For example, one person walking in one direction is an example of one underlying process. This can be contrasted with multiple people walking in multiple directions, which is an example of multiple underlying processes. Chapter 10 shows that human movement and the traveling of paths in two or multiple directions is a scenario where pattern matching techniques are ineffective. This thesis presents a novel method for the event detection of the direction of travel for bidirectional temporally dynamic travelled paths. The premise of this work is that prior knowledge about the placement of distributed sensors can be incorporated into a probabilistic model which results in an increase in performance of event detection.

1.2. Problem Statement and Scope of Thesis

Due to the nature of different applications, the topics considered in this thesis have separate problem statements and scopes. Chapter 3, Chapter 4, and Chapter 5 present pre-established precision limits for the estimation of the direction of arrival and source localization using acoustic arrays. The direction of arrival (DOA) problem occurs when a single array estimates the angle at which a source is located. The source

localization problem occurs when multiple arrays simultaneously estimate the position of a source. The scope is specifically limited to non-moving narrowband sources. The goal of these chapters is to provide extensive graphical analysis of the Cramer-Rao lower bound (CRB) limits to precision measurements. The CRB functions can then be used to find an optimal sensor and target geometric relationship where the precision is maximized, but this is beyond the scope of this thesis.

Chapter 9 presents a method to reduce the number sensors needed to conduct surveillance in a closed area of surveillance, one direction of travel scenario. The sensor reduction problem can be stated as such: given the current placement of sensors, how can sensors be removed to improve efficiency of the system and enhance detection of anomalous trajectories? The goal is to provide an estimate of a pattern characterizing typical environmental dynamics using a one-step transition probability statistical model. This probabilistic model is used as a measurement to show that an anomalous evader is more likely to be detected by the trajectory they travel when compared to typical trajectories.

Chapter 6, Chapter 8, and Chapter 10 present methods to statistically model data and conduct event monitoring, event detection, and anomaly detection. The scope of these chapters is specifically limited to the application of time stamped, Boolean, aggregated, point sensor data. Chapter 8 specifically addresses the design and use of a Poisson data model for anomaly detection purposes, whereas Chapter 10 presents a new non-pattern matching event detection method.

1.3. Original Contributions

The original contributions of this thesis are:

1. Extensive graphical analysis of existing Cramer-Rao lower bound expressions for acoustic array sensing for the cases of: 1)direction of arrival (DOA) of two sources using one linear array, and 2)source localization of one source using two linear arrays. The CRB and its graphical analysis can be used as an objective function to determine the sensor placement to obtain the most precise measurements.
2. The improvement of detecting anomalies in point sensor data using a Poisson data model and numerical scores based on the difference from learned activity patterns. This is an example where if a statistical data model matches the physical generation mechanism of the data (people arriving and leaving), then statistical distribution models are applicable. Learning the parameters and patterns of the underlying statistical distribution is shown to result in an increased ability to detect anomalies.
3. A one-step transition probabilistic method to learn the typical pattern of environmental agent dynamics. This method is also used as a measurement to show how an anomalous evader is more probable to be detected.
4. A contextual, joint probability, causal relationship, sliding window, and geospatial intelligence (GEOINT) event detection and classification method for the direction of travel of agents on a bidirectional path under dynamic temporal traffic conditions. This contribution is a new approach to event detection as opposed to conventional time correlation, convolution, and pattern matching approaches. It shows that event detection can be enhanced by relating surrounding sensors and activity.

1.4. Organization of Thesis

The remainder of this thesis is organized in the following manner. Chapter 2 presents a literature review. Chapter 3 discusses the background concepts of the Cramer-Rao lower bound precision limits for both cases of the direction of arrival and the source localization of an acoustic source. Chapter 4 presents models and graphical analysis of the Cramer-Rao lower bound for the directions of arrival (Stoica & Nehorai, 1990), (Kay, 1993) for a uniform linear array sensing two acoustic non-moving narrowband sources. Chapter 5 presents the model from (Kozick & Sadler, 2004) for the Cramer-Rao lower bound for the source localization of a single acoustic source using two uniform linear arrays.

Chapter 6 and Chapter 7 present background concepts and a literature review for event monitoring, event detection, and anomaly detection topics. Chapter 8 presents a Poisson distribution data model for time stamped, Boolean, aggregated, point sensor data for the application of anomaly detection. This chapter is concluded with receiver operation characteristic (ROC) curves showing an increase in performance in anomaly detection over comparable summation methods. Chapter 9 is a self-supporting reprinted and individually copyrighted copy of a conference publication that presents a method to reduce the number of sensors and increase the measure of an anomalous trajectory. Chapter 10 is a self-supporting conference paper ready for publication which proposes a new non-pattern matching technique for the detection and classification of the direction of travel in crowded surveillance scenes.

Chapter 2. Literature Review

2.1. Literature Review for Direction of Arrival (DOA) Estimation

An actively pursued topic in research is improving the precision estimation of the direction of arrival (DOA) of narrow-band signals using sensor arrays (Stoica & Nehorai, 1990), (Song & Ritcey, 1996), (Wong, Reilly, Wu, & Qiao, 1992), (Zeira & Friedlander, 1996), (Jaffer, 1988), (Stoica & Sharman, 1990), (Wilson, 1998), (Stoica, Moses, Friedlander, & Soderstrom, 1989), (Satish & Kashyap, 1996), (Nguyen & Van Trees, 1994), (Stoica & Arye, 1989), (Kundu, 1996), (Swingler, 2002), (Swingler, 1993).

The estimation of the DOA is affected by the number of sources and signals, the channel environment that signals travel through, and how signals arrive at the geometric array(s) configuration. The main approaches to estimating the DOA are operating on the sample covariance matrix of the observed sensor outputs or modeling the data (Stoica & Nehorai, 1990). Most data models assume that the noise is temporally and spatially uncorrelated. Although the aforementioned uncorrelated noise model is in general agreed upon, a main difference in literature focuses on the source signal model. Two main source signal models are used in literature (Stoica & Nehorai, 1990): 1)unconditional model, 2)conditional model. The unconditional model assumes the signals to me random, whereas the conditional model assumes the signals to be a hybrid of both deterministic and stochastic properties. Both source signal models are presented in Chapter 4 and the CRB using the conditional model will be evaluated. A summary of main DOA literature and their contributions can be cited as follows.

(Stoica & Nehorai, 1990) distinguishes between the conditional and unconditional source signal models. They derive the explicit expressions for the covariance matrix and CRB for the unconditional model for use in the maximum likelihood estimation of the DOA. They show that the unconditional model provides tighter bounds than the conditional model, and that when the signal to noise ratio (SNR) and/or the number of sensors increases, then the conditional and unconditional signal models result in the same performance.

(Wong, Reilly, Wu, & Qiao, 1992), propose a Bayesian maximum a posterior (MAP) estimate of the DOA's of multiple sources. Their contribution is that they do not assume a spatially uncorrelated noise covariance matrix and use a Jeffrey's non-informative prior instead. In their second paper, (Reilly & Wong, 1992), they show results that show that the MAP DOA estimation method outperforms the MUSIC methods for unknown correlated white noise.

Another proposed approach to improve DOA estimates for narrowband sources with known center frequencies is (Zeira & Friedlander, 1996). By assuming this prior known characteristic of the signal, the CRB for multiple non-coherent sources is shown to achieve the single source CRB with small spatial separation relative to the beamwidth of the array. This paper proposes the use of a deterministic parametric source signal model, specifically polynomial phase signals. Examples of evaluated signal models are linear and quadrature frequency modulated (FM) signals as well as pure sinusoidal phase. Like (Zeira & Friedlander, 1996), this thesis presents a case study of the CRB when the center frequencies are known. Unlike (Zeira & Friedlander, 1996), however, Chapter 4

presents analysis of the scenario where a second source worsens the CRB of the DOA estimate of a first source due to varying the center frequency.

Other proposed methods, (Song & Ritcey, 1996), focus on how a random environment, describing the propagation of waves through an oceanic medium can be incorporated into the unconditional signal model. Their contribution is that they show the difference in the relationship between the covariance matrices of the received signals for the cases of a homogeneous (isotropic) and random Gaussian environment. The key parameter used to link the two signal models is the fluctuation strength, which is used to measure multiple scattering effects. The numerical results presented show that the mismatching of the fluctuation strength parameter result in the degradation of the DOA estimation performance away from the CRB. A key comparison made is between a parametric maximum likelihood estimator (MLE) method and the conventional beamforming method. Numerical results indicate that the beamformer method MSE does not vary with the fluctuation strength and that the MLE outperforms the beamforming method. Extensions to (Song & Ritcey, 1996) include (Wilson, 1998) which considers the effects of atmospheric turbulence on the DOA estimation. Their results show that the coherence of an acoustic wave-front is significantly affected by constant or intermittent atmospheric turbulence. They also show the degradation of the CRB for the DOA when the distance to the source or the frequency of the source is increased.

Much literature has focused on improving the MLE method for the DOA problem. (Jaffer, 1988) proposed a method to improve existing MLE of the DOA by showing that the MLE could be expressed as a function of only the angle parameters, thus reducing the parameter set needed for numerical optimization. (Stoica & Sharman, 1990)

investigated the multiple signal classification (MUSIC) algorithm (Stoica & Arye, 1989) and their previously proposed DOA estimation algorithms called the method of direction estimation (MODE). This comparison of algorithms was motivated by the observation that the MLE methods were not statistically efficient for a practical and finite number of sensors. Unlike the above literature, this thesis focuses on the physical interpretation of the CRB for sensor placement purposes and not the algorithms to determine the most statistically efficient (achieving more accuracy with a lesser amount of data samples) estimator.

A main problem plaguing the practical application of the DOA theory is calculating the CRB (covariance matrix) without inversion of the FIM. (Satish & Kashyap, 1996) proposed a Taylor series expansion of the inverse of the data covariance matrix instead of using a first order approximation of the likelihood function for a large number of observed data points as in (Stoica & Nehorai, 1990). (Satish & Kashyap, 1996) suggested that the application of the maximum likelihood estimator for the DOA problem requires the use of a larger number of sensors on an array instead of a larger number of observations recorded, which is unrealistic to assume. Like (Satish & Kashyap, 1996), this thesis shows results that the CRB on the variance of the estimate of the DOA worsens as the number of sensors on an array is decreased. Furthermore, this thesis also makes note of one example of literature (Brennan, 1961) that is not consistent with other literature because of its CRB expression derived from increasing number of sensors for a single snapshot.

The CRB is not the only optimal bound for the precision of estimates. (Sadler & Kozick, 2006) surveys the time delay estimation (TDE) problem which is strongly related

to the DOA problem through array geometry. The key parameters of the CRB, Ziv-Zakai, and Weiss-Weinstein bounds are the signal time-bandwidth product and the SNR. When the SNR drops below a certain threshold, the performance degrades for both the DOA and TDE problems, and the CRB is no longer the limit to precision. The Ziv-Zakai and Weiss-Weinstein bounds are used to predict the limit to precision below the SNR threshold, and also to practically determine where the SNR threshold occurs (Nguyen & Van Trees, 1994).

Interestingly, an unusual addition to the DOA literature includes simple algebraic approximations of the CRB (Swingler, 1993). Similar to the purpose of the case studies presented in this thesis, (Swingler, 1993) proposed a simplified CRB expression for a uniform linear array measuring two closely spaced sources (approximately less than a beamwidth) with unequal SNR, with prior information that the sources are uncorrelated. Their results show intuitively that when the SNR of one of the sources is high, then the CRB on the DOA of that source is a strong function of the other interfering source's SNR. This means that when one source is strong, its variance of the DOA estimate is primarily due to the weaker source's strength.

The above literature review of the DOA and associated problems is complete (with the caveat that many more papers and books could have been cited) in the sense that all of the above literature are related directly to obtaining, comprehending, and analyzing the CRB for the DOA problem. Furthermore, the references above were chosen because they were the most straightforward with their notation for the application purposes of our two case studies. Next, a literature review of a very similar and related problem of localizing an acoustic source is presented.

2.2. Literature Review for Source Localization

Acoustic source localization is the problem of locating a source in the environment given the measurements of the acoustic field. The source localization problem encompasses the DOA problem and the unknown parameters include both the range and bearing, as opposed to just the bearing for the DOA problem. Assuming the acoustic far-field plane wave model, the sound intensity (Kinsler, Frey, Coppens, & Sanders, 1999) of the source decays at the rate of the inverse of the distance squared (not to be confused with the sound pressure). Although it is true that a passive array could determine relative distance (closer or farther away) assuming a constant source amplitude, a more reliable method using multiple arrays is chosen. The standard source localization method triangulates a source using multiple arrays. This method relies on a distributed processing scheme where each array makes a DOA estimate which is transmitted to a data fusion center. This distributed processing scheme requires low communication bandwidth and therefore less power to transmit and receive, but the source localization will perform worse than an optimal central processing data fusion method (Kozick & Sadler, 2004). Using multiple arrays and a central processing scheme that enables the calculation of the coherence of signals between different arrays, source localization can be achieved with the limits to precision given by the CRB. The CRB for simultaneous range and bearing can be mapped into a two dimensional plane, indicating that a source is, at best, within an area.

A main piece of literature for the acoustic source localization problem using distributed sensor arrays is (Kozick & Sadler, 2004). Although distributed processing schemes had already been proposed, (Wax & Kailath, 1985), the main contribution of

(Kozick & Sadler, 2004) was modeling the frequency selective coherence for signals that arrive at distributed (separate) array locations. Consistent with the plane wave assumption, perfect spatial coherence across any individual array is assumed. Two distributed processing methods are discussed: 1)each array transmits only its DOA bearing estimate to the data fusion center (bearings only triangulation), 2)each array transmits its bearing estimate and the data from one sensor. The distributed processing methods are then compared to the centralized joint processing scheme. Through numerical CRB examples, it is shown that: 1)the first distributed processing method (triangulation) listed above is equivalent to the central processing scheme with zero coherence, 2)the second distributed processing method listed above is equivalent to the centralized processing scheme for all coherences. The conclusion is that any knowledge about the coherence of the signals between different array locations reduces the CRB when the coherence is nonzero. The centralized processing scheme is shown to have tighter bounds and is characterized in terms of the power spectra, signals, coherence between the signals, noise, and the number of observations. In this thesis, only the central processing scheme is presented because it provides the tightest bounds for the source localization area. The following discussed literature outlines the motivation that drove the formulation of (Kozick & Sadler, 2004).

One main literature that discusses decentralized processing is (Weinstein, 1981). They show that the maximum likelihood procedure can be decentralized. The main motivation for this approach is that the maximization procedure over multiple parameters with limited data becomes increasingly complex. Their proposed decentralized maximum likelihood (DML) method is applied by: 1) partitioning the data and unknown

parameter vectors, and 2)simultaneously forming the MLE of each individual unknown parameter. Their main assumption is that the signals arrive at in a fully coherent manner at the different sensor sites. Given the above assumption, they show that their DML pair-wise processing of received signals is near optimal for distributed sensing sites for source localization.

Another key paper that proposes a decentralized processing source localization technique with zero coherence is (Wax & Kailath, 1985). The motivation for this paper stemmed from the fact that the standard triangulation source localization method provides sub-optimal performance. They label the key trade off as between performance and communication load. To exploit the communication issue, they propose a decentralized processing scheme where each sub-array sends its sample covariance matrix to the fusion center. Since the sample covariance matrix contains all the information about the signals relative to each other at a sub-array, it is not necessary to send the data itself. The sample covariance matrices were also chosen because of their smallest eigenvalue and its corresponding eigenvector is orthogonal to the location vector of the source. Once all the covariance matrices have been received for a given time period, the source localization problem is expressed as a search for the smallest distance from orthogonality. The problem of multiple sources is briefly discussed and the conclusion is that spurious minima due to the intersections of multiple sources requires the data association step of triangulation. This method therefore yields no practical advantage for the localization of multiple sources.

Fully coherent and decentralized estimates are discussed in the context of the DOA problem in (Stoica, Nehorai, & Söderström, 1995). They consider a model where a

very large array is split into sub-arrays due to computational and physical constraints. They show that the two proposed MODE algorithms are statistically optimal for decentralized estimates because they give the minimum variance DOA estimates. The computational advantage of the MODE algorithm comes from the eigen-decomposition of the sample covariance matrix. The conclusions state that their work could be used to help derive optimal splitting of a large array into sub-arrays by minimizing the accuracy degradation. Although this paper compares the MODE method for both centralized and decentralized processing schemes, it does not address degradation of the source localization precision due to the loss of coherence.

Although (Kozick & Sadler, 2004) paved the path to obtain more realistic precision limits by proposing a source localization model with coherence relationships explicitly included, they only considered a multi-array, single source case. (Erling, Roan, & Gramann, 2007) extended their model to the multi-array, multiple source case. The main contribution that this paper makes besides extending previously published models is noting that multiple sources can have unforeseen effects on the uncertainty of the source localization due to the signals, array geometry and noise. They essentially suggest that one must explore the model with multiple sources and different parameters to verify and comprehend the CRB physically. The case studies presented in Chapter 4 and Chapter 5 are in line with this approach and are intended to offer insight to solve the problem of sensor placement to maximize precision.

Chapter 3. Background: Limits to Precision, Cramer-Rao Lower Bounds (CRB), Direction(s) of Arrival (DOA), and Source Localization

3.1. Introduction

A common problem in all science disciplines involves characterizing data. Characterization is a necessary process because the data observations themselves are typically not the quantities to be measured (Bos, 2007). In this chapter, it is assumed that the mathematical relationship between the data and a model of the data is known. When the model of the data is known, characterization of the data becomes the process of parameter estimation. Different estimators of the same parameters, applied to the same data observations, typically have different precision. Although many types of signal models and estimators for their parameters exist, the problem scope is narrowed down by only considering general Gaussian random variable (RV) signal models and the case of unbiased estimators, those which converge to the true expected value of the parameters estimated. In this section, the motivation to study the limits to precision of estimators instead of specific estimators is discussed, and the general steps (Van Trees, 2002) to finding the CRB are outlined. These two steps are: 1) Obtain the expression of the Fisher information matrix (FIM) for the given signal model, 2) Obtain the CRB expression by taking the inverse of the FIM. The background information is as follows.

3.2. Background Definitions

In this section, useful and needed definitions for operating on RV's are presented. The expected value of a RV is the average value that a RV will take when considering all

possible values that the RV could be. Let $f(x)$ be the PDF of an RV X. Then the expression for the expected value of a RV X is

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx \tag{3.1}$$

The concept of the expected value can be expanded to arbitrary functions of RV's by

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \tag{3.2}$$

Note that the expected value is taken with respect to the underlying PDF of the RV.

Moments are quantities that are of great interest when studying RV's. They measure the shape of a set of points. (Papoulis & Pillai, 2002) distinguishes between central, absolute, and generalized moments. The n^{th} central moment about a real variable c can be expressed as

$$\mu_n = E[(x-c)^n] = \int_{-\infty}^{\infty} (x-c)^n f(X)dx \tag{3.3}$$

For example, for a normal distribution, the first central moment (n=1) about the real variable c=0 is the mean,μ and the second central moment (n=2) about the mean c=μ is the variance. The statistical moments are applied to the log-likelihood functions to obtain the CRB expressions.

When estimating unknown parameters, a key property to aim to obtain is the sufficiency of a statistic (calculation or function) with respect to the statistical model being applied to the data. A statistic is sufficient if the data sample it was calculated with contains no extra information about the unknown parameter than the statistic calculation itself (Kay, 1993), (Papoulis & Pillai, 2002). The principle of a sufficient statistic can be most easily illustrated by considering the Neyman-Fisher factorization (Kay, 1993). Let

$p(x;\theta)$ be the PDF of RV X with unknown parametersθ. A function $T(x)$ is a sufficient statistic if the PDF can be factorized into arbitrary functions h and g such that

$$p(x;\theta) = h(x)g(T(x),\theta) \tag{3.4}$$

Equation (3.4) states that the function g is dependent only on the data *x* through $T(x)$ and that the function h is dependent only on the data *x*. Since the statistic $T(x)$ contains all knowledge about the data, it is a sufficient statistic for the unknown parameter(s) θ. Sufficiency can be interpreted using the likelihood principle which is described below. If two separate sets of data have the same value for the sufficient statistic calculation $T(x)$, then the inferences from the likelihood about the unknown parameters θ will be the same. Sufficiency also helps to determine which data samples are important to our estimation.

In signal processing, coherence is used to measure the extent that one signal can be predicted from another signal using an optimum linear least squares relationship. The coherence of two signals x and y, is defined as (Bendat & Piersol, 1986).

$$\gamma_{xy}^2 = \frac{\left|G_{xy}(f)\right|^2}{G_{xx}(f)G_{yy}(f)} \quad where \tag{3.5}$$

with limits $0 \leq \gamma_{xy}^2 \leq 1$ and where $G(f)$ are the one sided auto/cross power spectrums.

3.3. Unbiased Estimators

For an estimator to be unbiased, the estimator must, on the average, return the true value of the unknown parameter. Let the unknown parameter be symbolized by θ, the estimator be $\hat{\theta}$,and the interval that the unknown parameter occurs on be $a < \theta < b$. Then an estimator is unbiased if

$$E(\hat{\theta}) = \theta \qquad a < \theta < b \tag{3.6}$$

To illustrate this principle, consider an example of unbiased and biased estimators for a direct current (DC) value in white Gaussian noise (WGN). Let the observed data with unknown parameter limits $-\infty < \theta < \infty$ be

$$x(n) = \theta + w(n) \qquad \text{for } n = 0,1, \dots, N-1 \tag{3.7}$$

If the goal is to find the average value of the signal $x(n)$, then a good estimator is the sample mean.

$$\begin{aligned} E(\hat{\theta}) &= E\left[\frac{1}{N}\sum_{n=0}^{N-1} x(n)\right] \\ &= \frac{1}{N}\sum_{n=0}^{N-1} E[x(n)] \\ &= \frac{1}{N}\sum_{n=0}^{N-1} \theta \\ &= \theta \end{aligned} \tag{3.8}$$

Since the sample mean estimate of the unknown parameter is equal to the true value, the sample mean is in this example an unbiased estimator. Now let us consider the signal in (3.8) again but this time using a modified sample mean as the estimator.

$$\hat{\theta} = \frac{1}{2N}\sum_{n=0}^{N-1} x(n) \tag{3.9}$$

Applying the expected value to the modified sample mean estimator in (3.9), results in (3.10).

$$\begin{aligned} E(\hat{\theta}) &= \frac{1}{2}\theta \\ &\neq \theta \ \ if \ \ \theta \neq 0 \end{aligned} \tag{3.10}$$

Since the expected value of the modified sample mean estimator is not the true value of the parameter, the modified sample mean in (3.9) is a biased estimator. To restate (3.6)

in a more formal and continuous manner, let $p(x;\theta)$ be the probability density function (PDF) for $x = [x(0), x(1), \ldots, x(N-1)]$. Then the estimator $\hat{\theta} = g(x)$ is unbiased if

$$E(\hat{\theta}) = \int g(x)p(x;\theta)dx = \theta \quad \forall \ \theta \tag{3.11}$$

If an estimator $\hat{\theta}$ is biased, then the bias is defined as

$$b(\theta) = E(\hat{\theta}) - \theta \tag{3.12}$$

Next, the attention turns to outlining how some optimality criterion is established for estimators.

3.4. Justification of Measurement for Optimal Estimators

As it was seen in the previous section, not all estimators are equivalent in performance. A method is required to quantify the difference between the estimate and the true value of the unknown parameter. To gauge the difference in the performance of estimators, a well known and reasonable measurement is the mean square error (MSE).

$$\text{mse}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \tag{3.13}$$

The MSE measures the average, or expected value of the squares of the errors. The MSE of the estimator, however, cannot be found unless the true value is known. To further illustrate that the MSE is an impractical performance measure for estimators, consider the following mathematical explanation. By expanding the quadratic term in (3.13) and distributing the expected value operator, (3.14) is obtained.

$$\text{mse}(\hat{\theta}) = E\left[\hat{\theta}^2\right] - 2\theta E[\hat{\theta}] + \theta^2 \tag{3.14}$$

Solving (3.12) for θ and substituting it into (3.14) results in (3.15).

$$\begin{aligned} \text{mse}(\hat{\theta}) &= E\left[\hat{\theta}^2\right] - 2\big(E[\hat{\theta}] - b(\theta)\big)E[\hat{\theta}] + \big(E[\hat{\theta}] - b(\theta)\big)^2 \\ &= E\big[\hat{\theta}^2\big] - \big(E\big[\hat{\theta}\big]\big)^2 + b^2(\theta) \end{aligned} \tag{3.15}$$

$$= Var[\hat{\theta}] + b^2(\theta)$$

Equation (3.15) states that the MSE of the estimator is a function of the variance and the bias of the estimator. No matter how the MSE is expressed, for example (3.13)-(3.15), it is still a function of the bias and unknown parameter and is therefore not practical or calculable. Since the bias is unobtainable, (3.15) shows that the MSE is characterized by the variance of the estimate of the unknown parameter θ. This motivates the main criterion discussed in this section, the minimum variance criterion.

3.5. Minimum Variance Criterion

In the previous section (3.15) showed that the MSE is a function of the variance of the estimator and the bias of the unknown parameter. If the scope of our problem is narrowed down by considering only unbiased estimators, the task of minimizing the MSE now turns into one of minimizing the variance of an unbiased estimator (MVU). The definition of an MVU can be stated as follows. Consider an exhaustive set of estimators $\{\hat{\theta}_i\}\ \forall\ i = 1, \dots, I$, for the unknown parameter θ. The i^{th} estimator is said to be the MVU of θ if

$$\text{mvu}(\theta) = \begin{cases} \hat{\theta}_i \text{ if } \exists\ Var[\hat{\theta}_i] = \min\limits_{\theta}[\{Var[\hat{\theta}_1], Var[\hat{\theta}_2], \dots, Var[\hat{\theta}_I]\}] \\ D.N.E. \text{ otherwise} \end{cases} \tag{3.16}$$

Equation (3.16) states that the MVU estimator is the minimum of the set of variances of each estimator for all values of the unknown parameter θ. If the variance of an estimator is not minimum for all values of θ, then it is not the MVU. Since there are an infinite number of estimators, the MVU might not be findable, even if it exists. Since the conclusion that the MVU might or might not exist has been introduced, the discussion on

specific estimators is ended, and a related discussion on the statistical principle underlying the bounds of the variance of estimators is started.

3.6. Likelihood Principle

The FIM, and therefore the CRB, is based on the likelihood principle which can be distinguished from probability in the following manner using the normal distribution as an example. Let the data set that has been obtained be $x = [x(0), x(1), \dots, x(N-1)]$ and follow a model such that the normal random variable (RV) $X = x$ is defined by $X \sim N(\mu, \sigma^2)$. When the PDF of the RV is viewed as a function of the unknown parameter(s) θ, it is called the likelihood function. The unknown parameters for a normal distribution are μ, the mean, and/or σ^2, the variance such that $\theta = [\mu, \sigma^2]$. The likelihood function for the RV *X* described above can be expressed as (3.17).

$$L(\theta = [\mu, \sigma^2] | x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{3.17}$$

Although the notation for the left hand side of (3.17) is interchangeable in some texts, it is preferred in this thesis to write the likelihood function as a function that is conditioned on the fixed data. This notation is preferred also because it is consistent with Bayesian methods' notations (Lee, 2004). The left hand side of (3.17) reads: "The likelihood of the parameter set θ, given the data set *x*". The likelihood function essentially represents the probability of a parameter being specific values, given that the data that has already occurred. Thus the likelihood function in (3.17) represents a "family" of normal distributions which can be contrasted to the concept of a single distribution that a PDF represents.

Regardless of the choice for notation, the main idea behind the likelihood function is that all of the information about the unknown parameter(s) θ is contained within the 1)observed data and 2)the PDF underlying the observed data. The likelihood is most often used in the natural logarithm form called the log-likelihood for several reasons: 1)the logarithm is a monotonically increasing function that has the same maximum value at the same point as the likelihood, 2)the mathematics simplify when taking the derivative of the function, especially when working with any statistical distribution with terms belonging to the exponential family.

3.7. Fisher Score and Information

The Fisher score is the gradient with respect to the unknown parameters of the (natural) log-likelihood function shown

$$\frac{\partial \ln\big(L(\theta|\mathrm{x})\big)}{\partial \theta} \text{ or } \frac{\partial \ln\big(\mathrm{p}(\mathrm{x}|\theta)\big)}{\partial \theta} \tag{3.18}$$

The physical interpretation of the Fisher score is the change in the log-likelihood function with respect to the unknown parameters. The higher the change with respect to the unknown parameters, the higher the Fisher score is, which implies that a log-likelihood function that is more sharp/pointy contains more information about the unknown parameter than a non-sharp function. The Fisher score is used in the "regularity" condition (Kay, 1993) that needs to be satisfied in order to be able to apply the CRB.

$$E\left[\frac{\partial \ln\big(L(\theta|\mathrm{x})\big)}{\partial \theta}\right] = 0 \quad \forall\ \theta \tag{3.19}$$

The expectation is taken with respect to the PDF $p(x;\theta)$. Equation (3.19) states that the change of the log-likelihood function with respect to the unknown parameters on average, and over the entire data interval of *x* where the PDF is defined, is zero. The concept of

the regularity condition is abstract, but can be restated in terms of its application as follows. The regularity condition ensures that the PDF underlying the data is continuously differentiable and that the derivative of the log-likelihood function with respect to the unknown parameters (Fisher score) exists and is a finite value (contains a finite amount of information). The regularity condition also states that the first moment (the mean) of the Fisher score is zero.

The second moment of the Fisher score is the variance of the Fisher score, and is defined as the Fisher information. The Fisher information is a method of measuring the amount of information that an RV contains about the unknown parameters and is shown in (3.20).

$$I(\theta) = E\left[\left(\frac{\partial \ln\left(L(\theta|\mathrm{x})\right)}{\partial \theta}\right)^2\right] \tag{3.20}$$

It can be shown by identity (Kay, 1993) that the Fisher information can be equivalently be expressed as (3.21)

$$I(\theta) = E\left[\left(\frac{\partial \ln\left(L(\theta|\mathrm{x})\right)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln\left(L(\theta|\mathrm{x})\right)}{\partial \theta^2}\right] \tag{3.21}$$

Equation (3.21) shows that the physical interpretation of the Fisher information is a measure of the curvature (second derivative) of the log-likelihood function. Intuitively, the more Fisher information that is contained about the unknown parameters, the less the variance of an estimate of the unknown parameters will be. Two main points about the Fisher information and CRB relationship are: 1)the Fisher information is inversely proportional to the variance of an unbiased estimator (the CRB), 2)the Fisher information is the measure of all the information about the unknown parameters, therefore the variance of the estimator will always be greater than or, at best, equal to the inverse of it.

The two points about the Fisher information and CRB relationship are summarized in (3.22).

3.8. Cramer-Rao Lower Bound

$$\text{var}\left(\hat{\theta}\right) \geq \frac{1}{-E\left[\frac{\partial^2 \ln\left(L(\theta|\text{x})\right)}{\partial \theta^2}\right]} \tag{3.22}$$

For unbiased estimators, the Cramer-Rao Lower Bound (CRB) is the lower bound on the variance. If the CRB is shown to exist, then no other unbiased estimator of the parameters exists that has less variance than it. The CRB is thus interpreted as a limit to the precision of the measurements. The reason that the CRB is used is because of the following two information theoretic properties (Kay, 1993): 1)it is a nonnegative value, 2)it is additive for independent data observations. The second property results from the property of the natural log function.

The $n\,x\,n$ matrix form of the CRB occurs when there are multiple unknown parameters such that $\theta = [\theta_1, \dots, \theta_n]$. The scalar unknown parameter case for the Fisher information (3.21) can be written in its matrix form as (3.23), where the indices are $i = (1, \dots, n)$ and $j = (1, \dots, n)$.

$$[I(\theta)]_{ij} = -E\left[\frac{\partial^2 \ln\left(L(\theta|\text{x})\right)}{\partial \theta_i \partial \theta_j}\right] \tag{3.23}$$

The CRB for the estimate of the i^{th} unknown parameter can then be expressed as

$$var\left(\hat{\theta}_i\right) \geq [I^{-1}(\theta)]_{ii} \tag{3.24}$$

3.9. Conclusions

This chapter discussed needed statistical definitions, unbiased estimators, the justification for optimal estimators (MSE and MVU), the likelihood principle, Fisher score and information, and the expressions for the Cramer-Rao lower bound on the variance of the estimator. Although all of the above topics are statistical, they still can be physically interpreted through the applied calculus. The main points of this chapter can be summarized in the following statements. Unbiased estimators are sought after because the goal is to obtain estimates of parameters that on average will converge to the true value of the parameter being estimated. The motivation to use a minimum variance criterion comes from the impracticalness of the mean squared error. In most practical situations, the true value of the unknown parameter will be unknown so the next best approach is to determine how much the estimates vary. The likelihood principle expresses the chosen distribution's PDF in terms of the unknown parameters and conditions them on the observed data which becomes "fixed". The concept of Fisher information measures how much information is contained about the unknown parameters in the log-likelihood function. The CRB is the bound for the variance of an unbiased estimator and is therefore the limit to precision that can be obtained. With the background discussed in this chapter, the CRB theory is readily applicable to the acoustic direction of arrival (DOA) and source localization case studies.

Chapter 4. Cramer-Rao Performance Bounds for the Directions of Arrival for a Uniform Linear Array Sensing Two Acoustic Sinusoidal Sources

4.1. Abstract

Consider the problem of estimating the directions of arrival (DOA) of two narrowband plane wave acoustic sources impinging on a single uniform linear array of sensors. The variance of the estimates of the DOAs can be quantified using Cramer Rao lower bounds (CRB). The precision of the measurement of the DOA is dependent on the number of sensors, spacing of the sensors, frequency and phase of the acoustic sources, signal to noise ratio, number of samples gathered in time, and the variance of the environmental noise. Given that no range information is given and that only one sensing array is available, the parameter estimation of the DOA is the applicable model available in literature. A stochastic/deterministic "conditional" hybrid model (Stoica & Nehorai, 1990) of the directions of arrival is derived and CRB theory is applied to the cases of: 1) unknown frequency, 2) unknown frequency and phase. It is shown that the CRB of the measurement error is a function of the true bearings to the sources, worsens with more uncertainty due to unknown phase, and that the bounds correspond directly to the physical characteristics of two acoustic sources interfering with each other.

4.2. Introduction

Consider the case where there are two narrowband plane wave sinusoidal signals/sources received at one uniform linear array (ULA) of sensors.

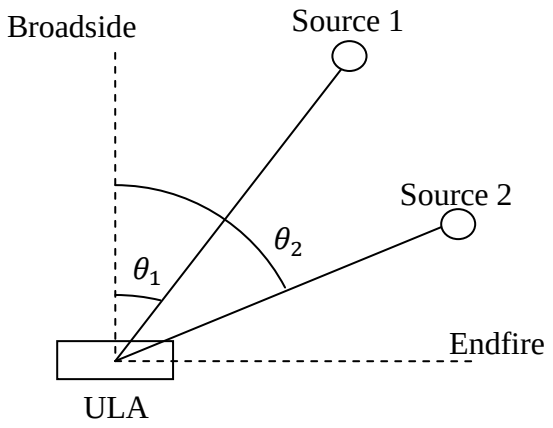


Figure 4-1: Bearings from ULA to two acoustic sources

The measured output of the array and unbiased estimate $\hat{\theta}$ for a direction of arrival (DOA) of an acoustic source can be written as:

$$\hat{\theta} = \theta + v \tag{4.1}$$

where θ is the true bearing to the source and v is the additive zero mean Gaussian noise defined by (4.2).

$$v \sim N(0, {\sigma_v}^2) \tag{4.2}$$

The noise variance ${\sigma_v}^2$ is caused by characteristics of the environment, array geometry, and signal properties. Chapter 3 discussed the motivation for Cramer-Rao lower bound (CRB) analysis and in this chapter, the CRB of the directions of arrival (DOA) are graphically analyzed.

The goal is to estimate the DOA's of the two acoustic sources upon the ULA. The differences between the two main theoretical models of the signals used in DOA estimation will be discussed and the conditional stochastic/deterministic hybrid model will be used to derive the CRB expressions. The major assumptions are that the following parameters are known about our system:

1. Center frequencies of sources

2. Variance of noise
3. Signal to noise ratio (SNR)
4. Time average of the samples of our received signals converges to a mean value given that enough samples are processed at one time to yield an accurate mean

The first assumption is specific to our application and the justification of the other assumptions will be discussed at pertinent times. Examples of the CRB derived using the conditional model will be presented and analyzed, yielding a map of the variance of the measurement of the DOA parameters.

4.3. Signal Model and Assumptions

The source signals can be modeled in two different manners (Stoica & Nehorai, 1990) as an Unconditional Model (UM) or Conditional Model (CM).

4.3.1. Unconditional Signal Model

The unconditional model (UM) represents the received source signals as a zero mean Gaussian random variable (RV):

$$Y(t) \sim N(0, R) \tag{4.3}$$

where the covariance matrix R is defined by (4.4).

$$R = APA^* + \sigma^2 I \tag{4.4}$$

The complex conjugate transpose is denoted by "*", and P is the covariance matrix of the original signals $x(t)$:

$$P = \left(\frac{1}{N}\right) \sum_{t=1}^{N} x(t) x^*(t) \tag{4.5}$$

Thus if the received signals were observed, then the sample data covariance matrix is represented by:

$$\hat{R} = \left(\frac{1}{N}\right)\sum_{t=1}^{N} y(t)y^*(t) \tag{4.6}$$

The UM also assumes that the original signal $x(t)$ is a stochastic random variable itself, hence the model is fully stochastic. The UM is named so because it does not depend on, or is not conditional to the explicit expressions for the original signals.

4.3.2. Conditional Signal Model

Now the focus turns to the conditional model (CM). Using the CM, the received source signals are represented as non-zero mean Gaussian random variables (RV):

$$Y(t) \sim N(\mu, \sigma^2) = N(Ax(t), \sigma_{noise}^2 I) \tag{4.7}$$

where the mean vector is the array manifold *A* multiplied by the sensed signals, and the variance is due to uncorrelated complex Gaussian noise both temporally and spatially. Thus the properties of the noise *e(t)* are:

$$\begin{aligned} &E[e(t)e^*(s)] = \sigma_{noise}^2 \delta_{t,s} \\ &E[e(t)e^T(s)] = 0 \quad \text{(for all t and s)} \end{aligned} \tag{4.8}$$

The signals and noise are also assumed to be uncorrelated:

$$E[x(t)e^T(s)] = 0 \quad \text{(for all t and s)} \tag{4.9}$$

The CM also assumes that the original signal $x(t)$ is nonrandom, in the sense that all characteristics of $x(t)$ are specifically realized in $y(t)$ for all time. The CM is named because it depends on, or is conditional to the explicit expressions for the original signals. The CM was chosen as the model to work with, because:

1. The characteristics of our received signals were known (testing purposes)

2. The Cramer-Rao (lower) Bound (CRB) of the CM is known to provide similar performance to the CRB of the UM for weakly correlated sources (Stoica & Nehorai, 1990)
3. The CRB of the CM is also similar to the UM for highly correlated sources, given that the sample amount is large (Stoica & Nehorai, 1990)

4.3.3. Conditional Model Signal Definitions

The CM can be fully defined in the following manner:

$$Y(t) \sim N(\mu, \sigma^2) = N(Ax(t), \sigma_{noise}^2 I) \tag{4.10}$$

Let the two acoustic sinusoidal sources be represented by:

$$\begin{gathered} x(t) = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \\ s_1 = \mathrm{a}_1 \sin(\omega_1 t + \varphi_1) \\ s_2 = \mathrm{a}_2 \sin(\omega_2 t + \varphi_2) \end{gathered} \tag{4.11}$$

The direction of arrival (DOA) for each source can be modeled as a complex delay as:

$$A_{mn}(\omega_n, \theta_n) = e^{-jm\alpha_n} \tag{4.12}$$

where:

$$\begin{gathered} \theta_n = \text{unknown DOA angle to estimate} \\ \alpha_n = \frac{\omega_n d}{c} \sin(\theta_n) \\ m = 0 \ldots M - 1 \ \text{(Sensor Index)} \\ n = 1 \ldots 2 \ \text{(Signal Index)} \\ c = \text{speed of sound} \\ d = \text{distance between sensors on ULA} \end{gathered} \tag{4.13}$$

Thus the means of the signals sensed by an ULA will have dimensions of (m x 1) and has the form:

$$\mu_{mn} = A_{mn}x = \begin{bmatrix} s_1 + s_2 \\ \vdots \\ e^{-jm\alpha_1}s_1 + e^{-jm\alpha_2}s_2 \end{bmatrix} \tag{4.14}$$

The covariance matrix is diagonal with dimensions of (m x m) and has the form:

$$C(\theta) = \begin{bmatrix} \sigma_{noise}^2 & 0 \cdots & 0 \\ 0 & \ddots & \vdots \\ \vdots & & 0 \\ 0 & \cdots 0 & \sigma_{noise}^2 \end{bmatrix} \tag{4.15}$$

With the mean and variance of the CM defined, the Fisher information matrix and the Cramer-Rao Lower Bound are ready to be explored.

4.4. Fisher Information Matrix and Cramer-Rao Lower Bound

4.4.1. Fisher Information Matrix and Cramer-Rao Lower Bound Relationships

Recalling the unbiased estimator in (4.1), the expected value of the mean squared error of the estimate is:

$$E\left[\left(\hat{\theta} - \theta\right)^2\right] \geq J^{-1}(\theta) \tag{4.16}$$

The precision to which θ can be estimated is limited by the Fisher information matrix (FIM) which is denoted here by J. Extending (4.16) to the multiple unknown parameter case (from now on let θ =unknown parameter vector) and by simplifying the left hand side the Cramer-Rao lower bound (CRB) inequality relationship is found to be:

$$var\left(\hat{\theta}_i\right) \geq [J^{-1}(\theta)]_{ii} \tag{4.17}$$

The CRB states that the variance of the parameter estimated, or the error of the parameter estimate will always be greater or equal to its corresponding diagonal component of the inverse FIM. Hence the CRB is the best estimation precision obtainable. Block forms for computing (4.17) can be found in (Papoulis & Pillai, 2002).

4.4.2. Fisher Information Matrix for a General Case Gaussian Random Variable

Consider a general Gaussian random variable:

$$G = N(\mu(\theta), C(\theta)) \tag{4.18}$$

Then the FIM for the general Gaussian random variable for one sample in time can be written as (Stoica & Nehorai, Performance study of conditional and unconditional direction of arrival estimation, 1990), (Kay, 1993):

$$J_{ij}(\theta) = Re\left\{\left[\frac{\partial\mu(\theta)}{\partial\theta_i}\right]^* C(\theta)^{-1}\left[\frac{\partial\mu(\theta)}{\partial\theta_j}\right]\right\} + \frac{1}{2}tr\left\{C(\theta)^{-1}\frac{\partial C(\theta)}{\partial\theta_i}C(\theta)^{-1}\frac{\partial C(\theta)}{\partial\theta_j}\right\} \tag{4.19}$$

where "Re" and "tr" denote the real part and the matrix trace respectively. For the conditional signal model (4.10)-(4.15), the covariance matrix is not a function of θ, thus the FIM reduces to:

$$J_{ij}(\theta) = Re\left\{\left[\frac{\partial\mu(\theta)}{\partial\theta_i}\right]^* C(\theta)^{-1}\left[\frac{\partial\mu(\theta)}{\partial\theta_j}\right]\right\} \tag{4.20}$$

Considering the joint probability of multiple independent and identically distributed samples of a Gaussian RV in time along with a diagonal covariance matrix, the FIM further simplifies into (Stoica & Nehorai, 1990), (Kay, 1993):

$$J_{ij}(\theta) = \frac{N}{\sigma^2_{noise}} Re\left\{\frac{1}{N}\sum_{t=1}^{N}\left[\frac{\partial\mu(\theta)}{\partial\theta_i}\right]^*\left[\frac{\partial\mu(\theta)}{\partial\theta_j}\right]\right\} \tag{4.21}$$

where *N* is the number of samples. Thus the averaging summation across time appears in the FIM expression when considering multiple samples. In essence, a linearly increasing amount of information is gained as the number of time samples increases. There are three major points to note about (4.17) and (4.21):

1. The CRB is inversely proportional to N, the number of samples, hence the CRB will improve as the number of time samples is increased.

2. The CRB is directly proportional to the variance of the signal noise.
3. The summation and averaging of samples over the observation time will yield the amplitude of the received signals. Let the $SNR = \frac{a^2}{\sigma_{noise}^2}$. Then the CRB will be inversely proportional to the SNR.

The key points mentioned above imply that the unknown variables are scaling factors of the CRB, and that they won't change the partial derivatives given that a sufficient amount of time samples are processed.

4.4.3. Discrepancies in Other Direction of Arrival Cramer-Rao Bound Literature

Other nonmainstream literature, (Brennan, 1961) and (Gadre, Roan, & Stilwell, 2008), suggest that the angular accuracy of the array can be determined by studying the CRB found with respect to summing and averaging all signals received at the sensors for one snapshot of time. Pursuing the CRB for the DOA parameter estimation using their methods with either complex or real signal models ends up in solutions not obtainable in closed solution form when trying to solve the multiple sources problem. This is due to the small amount of sensors, which does not allow the signal components of the Fisher information to be averaged out. All literature reviewed except for those mentioned in this section were found to follow the UM or CM theory to obtain the CRB for the DOA. The main difference between models can be described in the following manner. The non-mainstream way to increase Fisher information and the precision of DOA measurements is to increase the number of sensors. This can be contrasted with the more practical main method which increases Fisher information due to the number of observations obtained. Both models are technically correct, but by using the mainstream model, closed form

solutions can be found. Our graphical analysis will also include the case of increasing number of sensors on an array. In the next section, examples and graphical analysis of the CRB's for the two source, one array, DOA parameter estimation problem are presented.

4.5. Example 1: Two Sources, Same Frequency, Unknown DOA Angle

Consider the simplest case where only the two DOA's are unknown and are sensed by one ULA. The unknown parameter vector denoted in bold is:

$$\boldsymbol{\theta} = [\theta_1 \theta_2] \tag{4.22}$$

To find the CRB or the variance of the first source's DOA estimation (4.17), it is found that:

$$var(\hat{\theta}_1) \geq [J^{-1}(\boldsymbol{\theta})]_{11} = f(\theta_1, \theta_2, \omega_1, \omega_2, SNR) \tag{4.23}$$

To evaluate (4.17), (4.21), and (4.23), the following constants are assigned:

$$\begin{aligned} &c = 1497 \quad (m/s \quad \text{speed of sound freshwater @ temp=25 C}) \\ &d = 0.22 \quad (m) \\ &\omega_1 = (2\pi)(11{,}800) \quad (rad/s \quad \text{frequency of source 1}) \\ &\omega_2 = \omega_1 \quad (rad/s \quad \text{frequency of source 2}) \\ &SNR = 5 \quad (\text{unitless}) \end{aligned} \tag{4.24}$$

Both of the unknown DOA's are varied from front endfire to back endfire positions avoiding side ambiguity such that:

$$\begin{aligned} -90 \leq \theta_1 \leq 90 \\ -90 \leq \theta_2 \leq 90 \end{aligned} \tag{4.25}$$

To study the effects of a second source on the first source (Papoulis & Pillai, 2002), the expression for the upper left diagonal component of a two dimensional CRB is derived as (4.26) and vary the number of sensors on the ULA.

$$[J^{-1}(\boldsymbol{\theta})]_{11} = \left(\frac{1}{J_{11}}\right)\left(\frac{1}{1-(J_{12}J_{21})/(J_{11}J_{22})}\right) \tag{4.26}$$

The leftmost component of the right hand side of (4.26) is the CRB for the single source case which is shown in Figure 4-2. The rightmost component of the right hand side of (4.26) is the effects that a second unknown source has on the first source which is shown in Figure 4-3. The entirety of (4.26) is the CRB for the estimate of the first source and is shown in Figure 4-4.

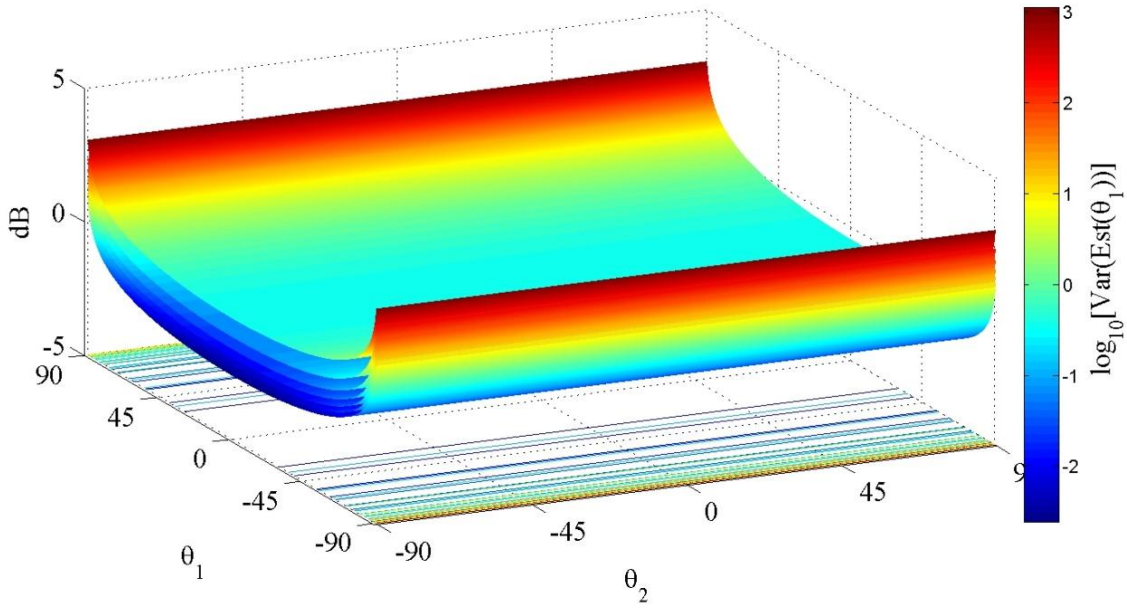


Figure 4-2: CRB same frequency case, single source

Figure 4-2 illustrates that the CRB for the single source case is not a function of any parameters related to the second source and is constant when varying θ_2. The variance of the estimate of θ_1 increases when the DOA moves away from broadside (0 degrees). The stacked multiple surface plots represent the CRB decreasing due to the

number of sensors on the ULA being increased from two to eight sensors. This characteristic remains the same for the remainder of our analysis.

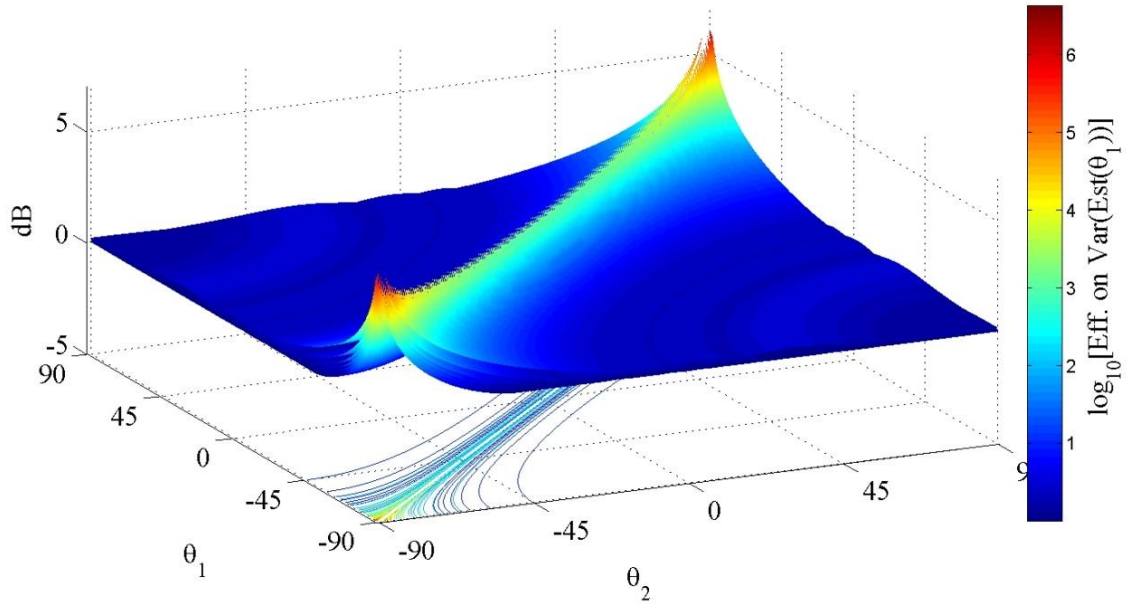


Figure 4-3: CRB same frequency case: Effects of 2nd source on CRB of 1st source

Figure 4-3 illustrates that a second source will increase the variance of the estimate of the first source, especially when the second source is near or at the same DOA as the first source. This is represented graphically by the high variance ridge and the symmetry of the effects about the $\theta_1 = \theta_2$ line. The second source increases the variance the most when it is at the same angle as the first source and at endfire position.

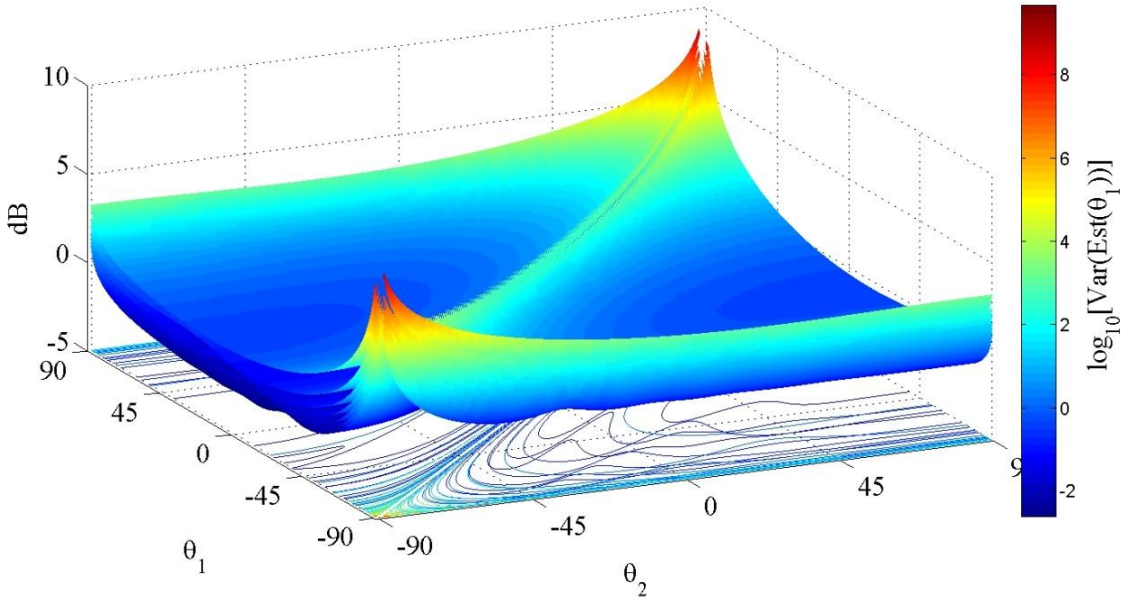


Figure 4-4: CRB same frequency case: Total CRB of 1st source

Figure 4-4 illustrates that the total CRB of the DOA for the estimate of the first source (4.26) is a combination of the single source CRB and the effects of the second source. The total CRB is characterized by increasingly higher variance when moving towards endfire positions and the highest variance when at endfire and an interfering second source at the same DOA.

4.6. Example 2: Two Sources, Different Frequency, Unknown DOA Angle

The two source different frequency case adjusts the frequencies of (4.23) to (4.27).

$$\omega_1 = (2\pi)(11{,}800) \quad (rad/s \quad \text{frequency of source 1})$$
$$\omega_2 = (2\pi)(11{,}200) \quad (rad/s \quad \text{frequency of source 2}) \tag{4.27}$$

The single source expression is independent of the second sources frequency and remains the same as in Figure 4-2. The effects a second unknown on the first source is shown in Figure 4-5 and the entirety of (4.26) is shown in Figure 4-6.

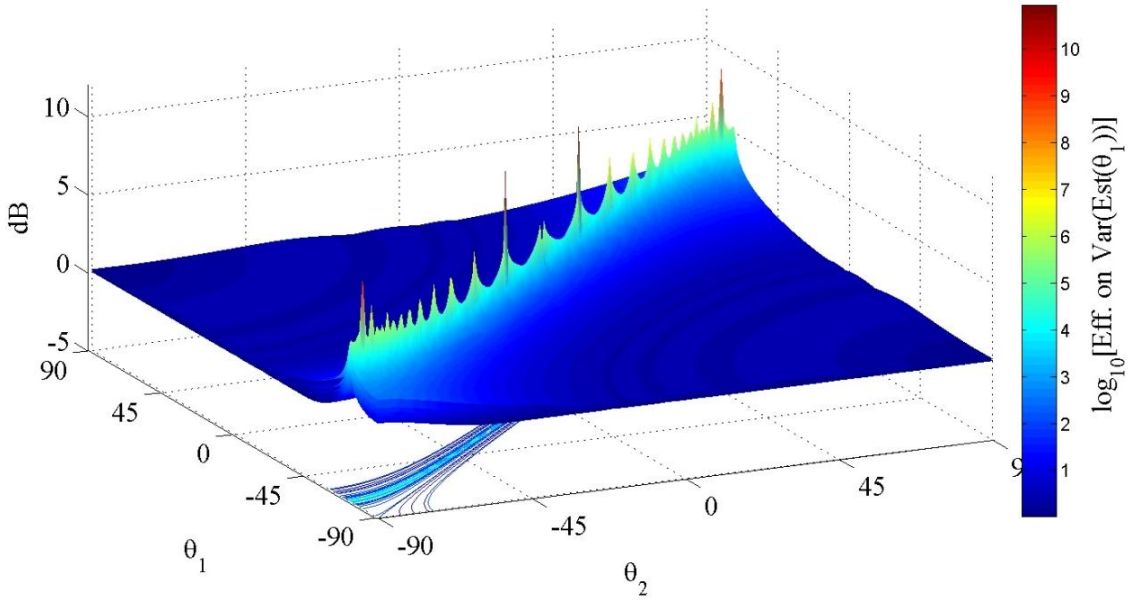


Figure 4-5: CRB different frequency case: Effects of 2nd source on CRB of 1st source

Figure 4-5 illustrates that when the two sources have different frequencies, the symmetry of the variance is offset off of the $\theta_1 = \theta_2$ line. High variance points are still characterized by a ridge, representing the two sources having near or the same DOA. The ridge is no longer smooth as in Figure 4-3 but contains distributed spikes. This is due to the mismatch of signal frequencies interfering at periodic angle combinations.

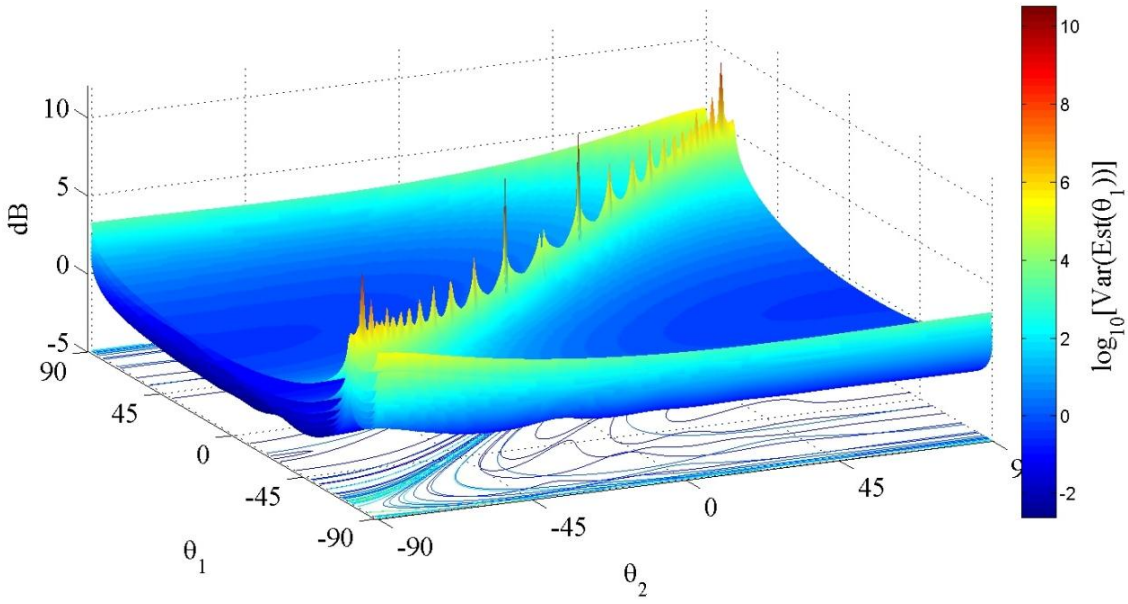


Figure 4-6: CRB different frequency case: Total CRB of 1st source

Figure 4-6 illustrates again that the total CRB of the DOA on the first source is a combination of the effects of a second source and the single source case.

4.7. Example 3: Comparison of Same Frequency and Different Frequency Cases, Unknown DOA

A direct comparison of same and different frequency CRB's is shown as follows.

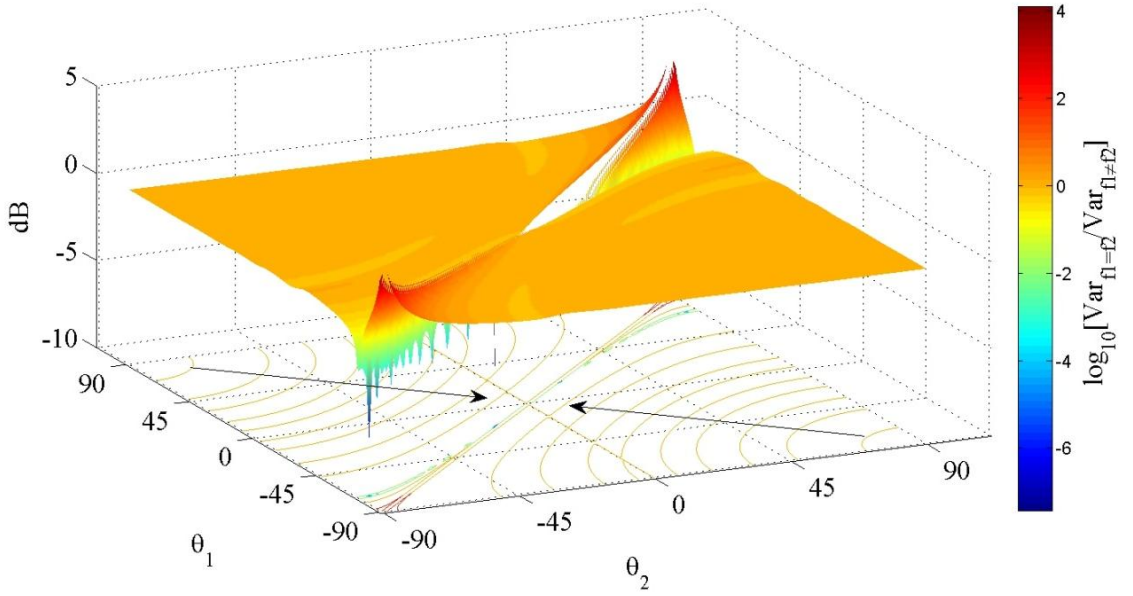


Figure 4-7: Comparison of total CRB for same and different frequencies cases

Figure 4-7 illustrates the comparison of same and different source frequencies for the total CRB while increasing the number of sensors on the ULA. The offset of the symmetry due to the frequency mismatch is more visible. The arrows indicate the movement of small oscillating ridges as the number of sensors increases. Several conclusions from the above examples can be drawn:

1. The variance is the highest when the sources are either at the same angles or at endfire positions
2. Sources at different frequencies create more complex CRB's and could possibly create unforeseen sensing variance effects (Erling, Roan, & Gramann, 2007) when also considering array geometry
3. The difference between $var_{f1\neq f2} and\ var_{f1=f2}$ for the $\theta_1 \neq \theta_2$ and $\theta_1 \approx \theta_2$ (not equal to and not approximately equal to) region are spatial ridges
4. The inclusion of more sensors and the gain in information results in a decrease in the variance of the estimate and increases the precision of the estimate

4.8. Example 4: Two Sources, Same Frequency, Unknown DOA Angle and Unknown Signal Phase

Consider the case where both the DOA's and the source phases of each source are unknown. The unknown parameter vector is:

$$\boldsymbol{\theta} = [\theta_1 \ \varphi_1 \ \theta_2 \ \varphi_2] \tag{4.28}$$

The change in the number of unknown parameters changes the dimensions of (4.21) from two by two, as in examples 1-3, to dimensions of four by four. The CRB of the estimate of the first source's DOA (4.26) can be rewritten as (4.29) (Papoulis & Pillai, 2002)

$$[J^{-1}(\boldsymbol{\theta})]_{11} = \left(\frac{1}{J_{11}}\right)\left(\frac{1}{1 - B^T G^{-1} B / J_{11}}\right) \tag{4.29}$$

where the matrices *B* and *G* are block matrices within the FIM defined by (4.30).

$$J = \begin{pmatrix} J_{11} & B^T \\ B & G \end{pmatrix} \tag{4.30}$$

The constants assigned, the evaluation process, and the single source case are the same as in examples 1-3. The effects of the three unknown parameters on the CRB of the DOA of the first source are graphed while varying the number of sensors as shown below.

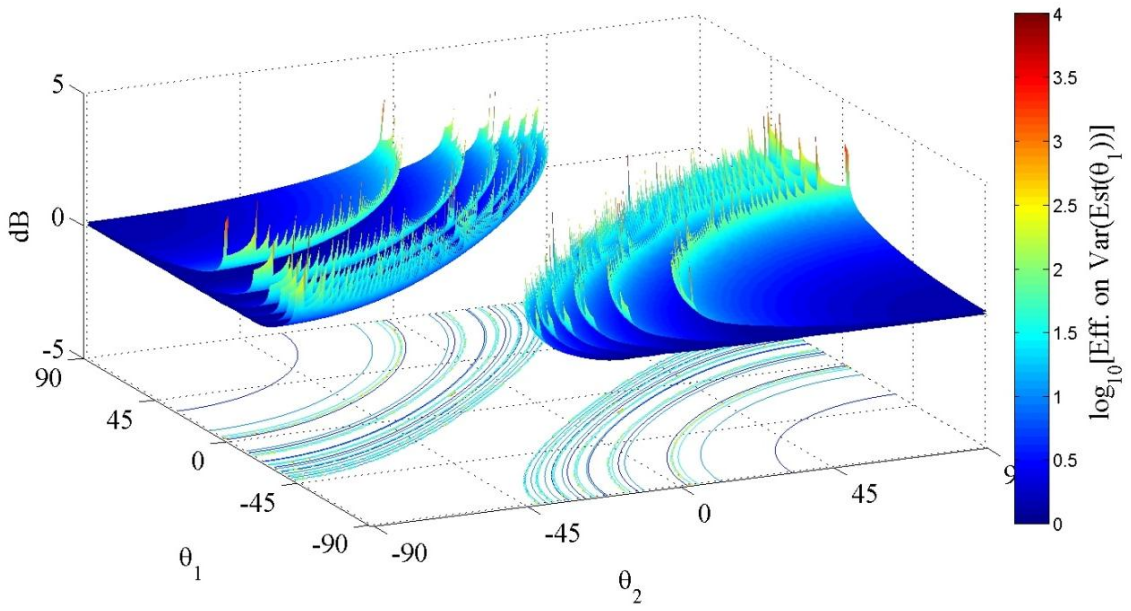


Figure 4-8: CRB same frequency case: Effects of 2nd source, unknown DOA and phase, unknown 1st source phase on CRB of 1st source

Figure 4-8 shows that the remaining unknown parameter set of $[\varphi_1\ \theta_2\ \varphi_2]$ increases the variance of the estimate towards infinity at specific lines of contour. The effects in Figure 4-8 also indicate that an increase from three to eight sensors on an array decreases the variance for regions away from the $\theta_1 = \theta_2$ line. Figure 4-9 below shows that the total CRB when both angle and phase are unknown, is higher than the case where the DOA's are unknown as shown in Figure 4-4.

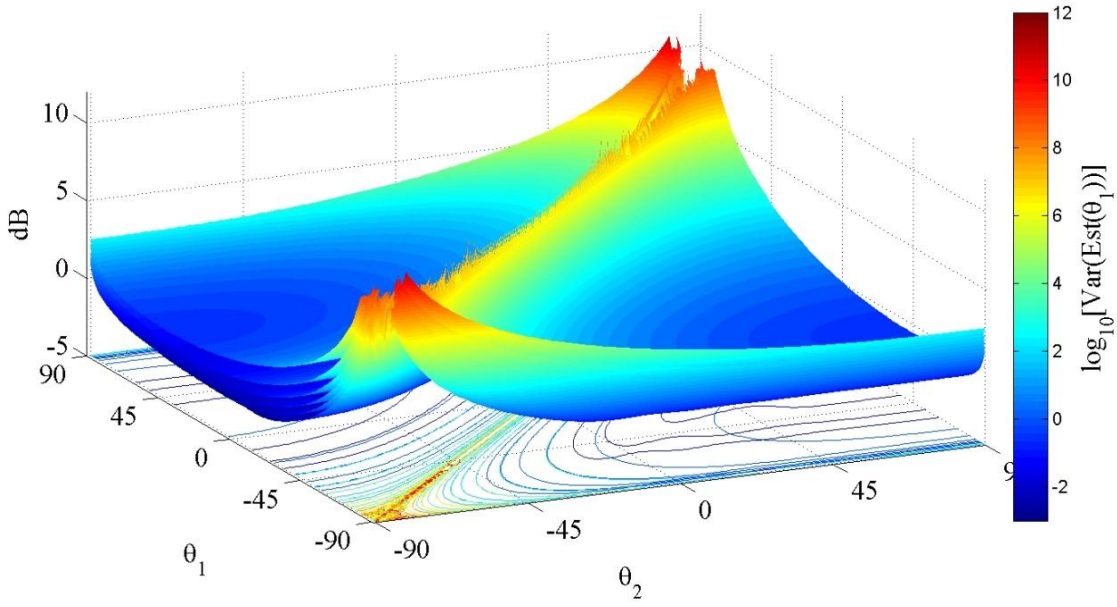


Figure 4-9: CRB same frequency case: Total CRB of 1st source, unknown phase and angle

4.9. Example 5: Two Sources, Different Frequency, Unknown DOA Angle and Unknown Signal Phase

The frequencies of the two sources were then changed to the same values as example 2.

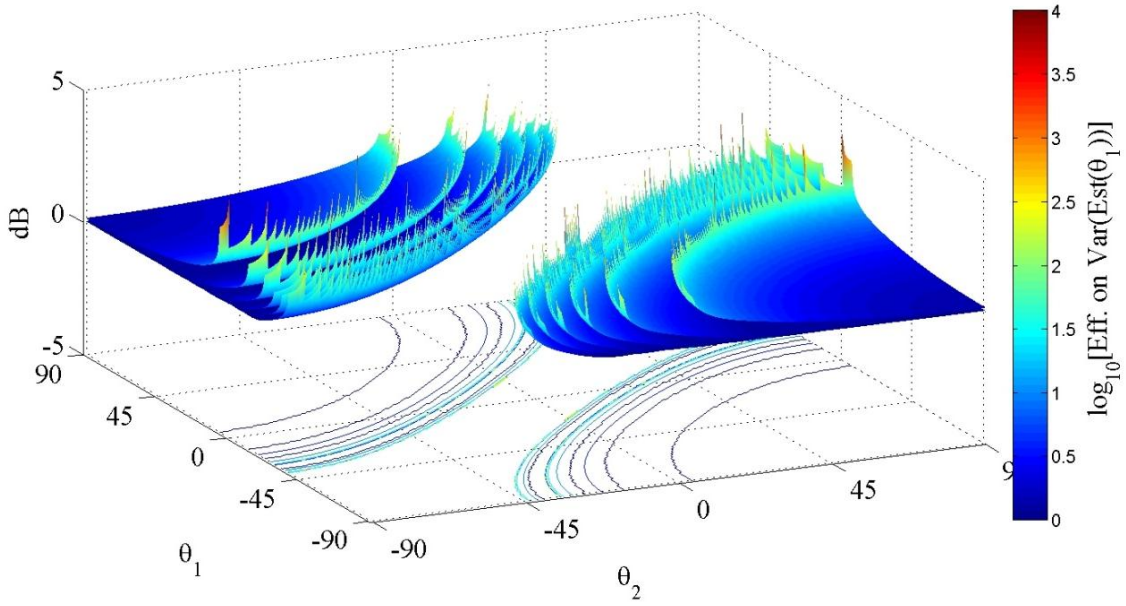


Figure 4-10: CRB different frequency case: Effects of 2nd source on CRB of 1st source, unknown DOA and phase

Figure 4-10 illustrates that the effects for the different frequency, unknown DOA and phase case have the same pattern as the same frequency case Figure 4-8, except the spatial ridges occur at different DOA combinations. below, Figure 4-11 shows that the total CRB in this case is the highest at double broadside DOA's out of all the above total CRB examples.

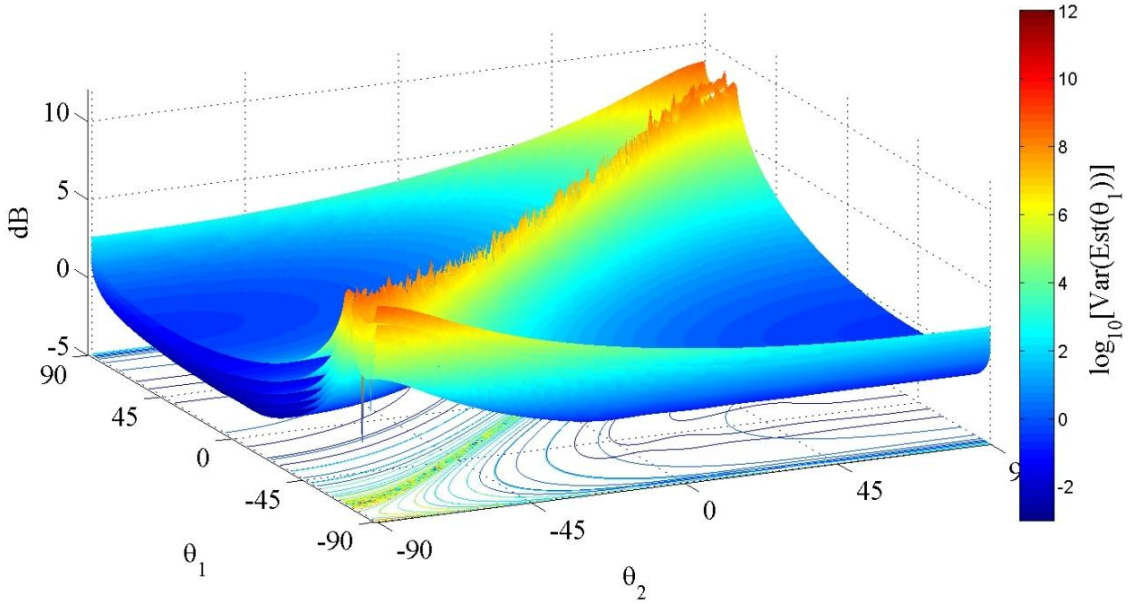


Figure 4-11: CRB different frequency case: Total CRB of 1st source, unknown phase and angle

4.10. Example 6: Comparison of Same Frequency and Different Frequency Cases, Unknown DOA Angle and Unknown Signal Phase

A direct comparison of same and different frequency CRB's is shown as follows.

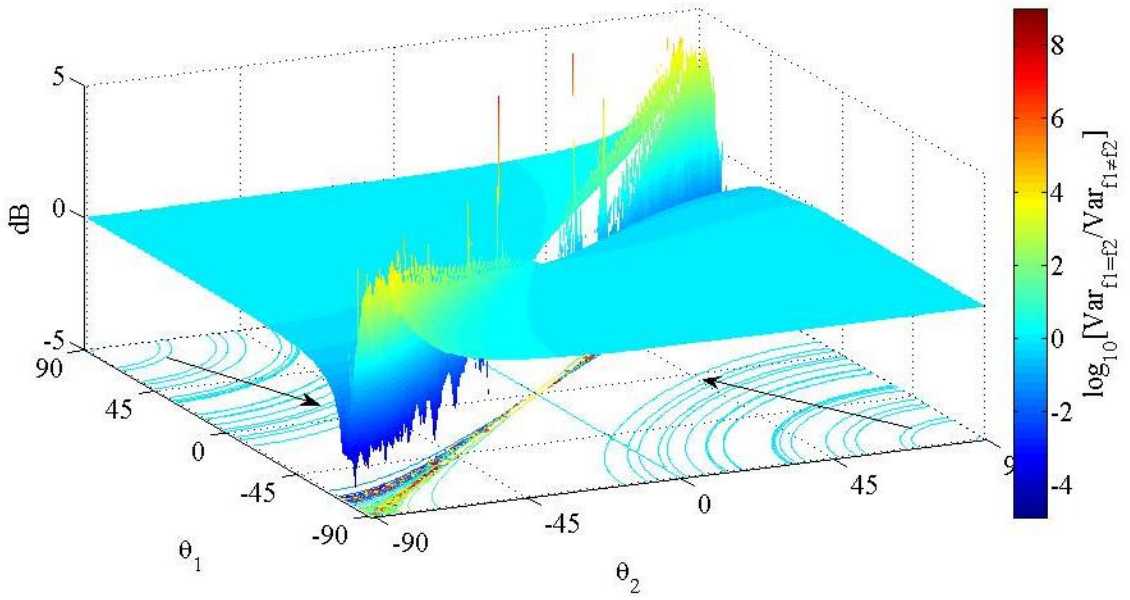


Figure 4-12: Comparison of total CRB for same and different frequencies cases

Figure 4-12 illustrates the same offset of symmetry due to the frequency mismatch as Figure 4-7.

For this example, it is shown that all the conclusions for examples 1-3 are also true for examples 4-6. By considering the signal phases as unknown parameters, the results show that the CRB's are greater than those in examples 1-3. This is due to the increase in the number of unknown parameters. As the number of unknown parameters increases, the CRB worsens. The effects shown in Figure 4-8 and Figure 4-10 indicate that the closed form CRB effects expressions for the unknown DOA and phase cases are numerically unstable due to missing trigonometric function cancelations. The results shown in Figure 4-9 and Figure 4-11, however, indicate that numerical expressions for the total variance of the first source's estimate (which include the effects) are numerically stable.

4.11. Conclusions: Two Sources, One Array DOA Problem

The parameter estimation DOA problem considered in this chapter was for one array, and two sources. It was shown that the physical characteristics of two acoustic sources interfering with each other were represented accurately by the CRB's generated in simulation. By examining examples 1-6, it is found that the general shape of the variance maps do not change for the two source DOA problem, but the scale and range of the variance do. Similar to the line of thinking discussed in (Erling, Roan, & Gramann, 2007), the analysis of the performance limitations of the DOA problem might have unforeseen effects. This unpredictability in the variance of estimates can be attributed to signal, environmental, array geometry, and a combination of all three contributors. Despite the informality and subjectivity of the likelihood principle on which the CRB theory is based, the results were graphically validated with the physicality of the two sources, one array case. The DOA CRB maps and/or FIM's can serve as cost functions to minimize the variance and error of the measurement of the DOA's. The best measurements of the DOA's will occur at the combinations of θ_1 and θ_2 that give minimum values of the CRB variance maps.

With the addition of more arrays and/or range estimates available, the DOA problem discussed here becomes the source localization problem discussed specifically in (Kozick & Sadler, 2004) and (Erling, Roan, & Gramann, 2007). Source localization applies the same Cramer-Rao theory to the unknown x and y coordinates instead of the directions of arrival, yielding performance bounds in terms of areas surrounding the target.

Chapter 5. Cramer-Rao Performance Bounds for the Source Localization of a Single Source Using Two Uniform Linear Arrays

5.1. Abstract

Consider the problem of using a network of two uniform linear arrays to localize the position of one nonmoving narrowband plane wave acoustic source on a plane. It is assumed that the individual arrays have full and instantaneous communication of data between each other and the focus is solely on the Cramer-Rao performance bounds representing geometric error of the source localization. The error of the source localization is dependent on geometry (bearing and range to the source), the received signals (frequency, cross power spectral density, coherence, signal to noise ratio, and number of samples gathered), and array (orientation and configuration) properties. In this chapter, the geometric model for the narrowband signal case described by (Kozick & Sadler, 2004) is outlined and the dependency of the Cramer-Rao bounds on the aforementioned properties is shown for multiple cases.

5.2. Introduction

Source localization of an acoustic source is a continually studied problem. Both central and decentralized signal processing schemes have been proposed for source localization as presented in Chapter 2. Typical decentralized methods rely on the triangulation of a source by processing multiple DOA estimates from individual arrays that are transmitted to a central processing unit. The distributed processing scheme

requires less bandwidth, but the localization accuracy is sub-optimal. In this chapter, a centralized processing scheme discussed specifically in (Kozick & Sadler, 2004) is considered. The motivation for this case study is to determine how the geometry of two uniform linear arrays and their position relative to an acoustic source affect the precision limits to source localization as depicted below in Figure 5-1.

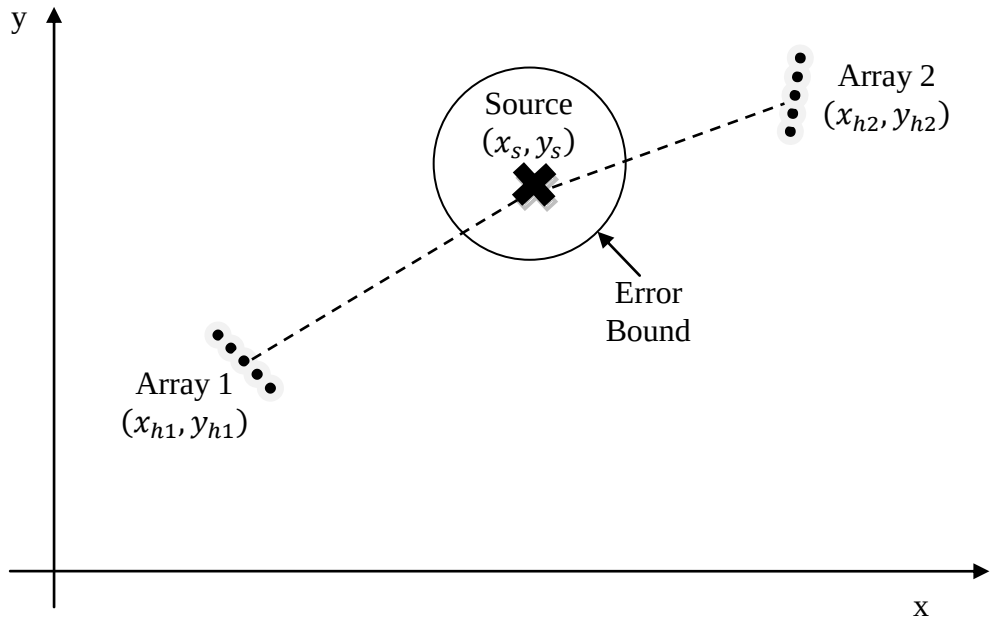


Figure 5-1: Two linear arrays sensing one source and error bound of source localization

The problem can be stated as: Given that the arrays have both bearing and range information available about the source, and that both arrays share this information with each other unlimitedly (not communication bandwidth limited), what is the size, shape, and direction of the error bound on the source localization when the coherence is varied? By modeling the received signals as zero-mean, jointly wide-sense stationary, Gaussian random processes, the Cramer-Rao theory discussed in Chapter 3 is used to bound the error of the source's position in the (x, y) plane.

5.3. Kozick, Sadler Signal Model Equation Summary

The following equations from (Kozick & Sadler, 2004) and notes were used to obtain the Cramer-Rao bounds (CRB). Let a single nonmoving source be located at (x_s, y_s) in the (x, y) plane and let $h \in \{1 \dots H\}$ arrays sense the source with their reference sensor be located at (x_h, y_h) as shown in Figure 5-1. The location of the n^{th} sensor $n = \{1 \dots N_h\}$ on an array is $(x_h + \Delta x_{hn}, y_h + \Delta y_{hn})$ and the relative locations Δx_{hn} and Δy_{hn} are defined below in (5.1).

$$\begin{aligned} \Delta x_{hn} &= r \sin \rho \\ \Delta y_{hn} &= r \cos \rho \end{aligned} \tag{5.1}$$

Equation (5.1) defines the relative locations in terms of polar coordinate distance r and angle ρ as shown below in Figure 5-2.

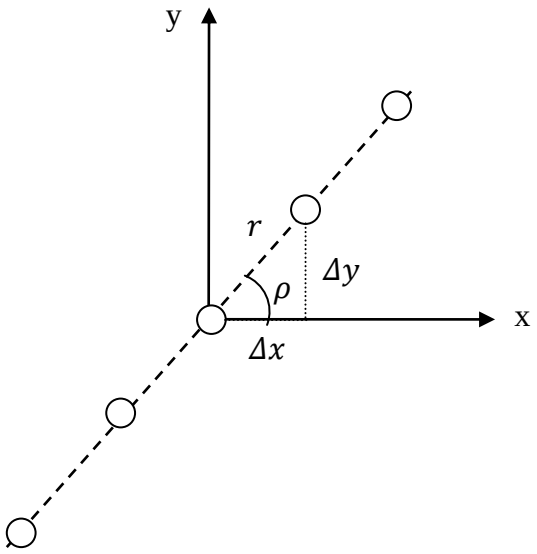


Figure 5-2: Inter-sensor Spacing and Array Orientation

Using the above definitions, the propagation time from the source to the reference sensor of any array can be expressed as (5.2)

$$\tau_h = \frac{d_h}{c} = \frac{1}{c}[(x_s - x_h)^2 + (y_s - y_h)^2]^{\frac{1}{2}} \tag{5.2}$$

where d_h is the distance from array *h* to the source, and *c* is the speed of propagation. Equation (5.2) explicitly shows dependency of the model on range between the source and arrays.

The propagation time model can be broken down into two parts: source to reference sensor, and reference sensor to other sensors on the same array. Using the far-field plane wave approximation, the propagation time from the reference sensor to each of the other sensors on an array can then be expressed as (5.3)

$$\begin{aligned} \tau_{hn} &= -\frac{1}{c}\left[\frac{x_s - x_h}{d_h}\Delta x_{hn} + \frac{y_s - y_h}{d_h}\Delta y_{hn}\right] \\ &= -\frac{1}{c}\left[\cos(\varphi_h)\Delta x_{hn} + \sin(\varphi_h)\Delta y_{hn}\right] \end{aligned} \tag{5.3}$$

where φ_h is the bearing of the source relative to array *h*. Equation (5.3) also states that the propagation time depends explicitly on the bearing of the source with respect to an array and also the orientation of the array. The total propagation time from the source to sensor *n* on array *h* is $\tau_h + \tau_{hn}$.

Let the vector of reference sensors be $s(t) = [s_1(t), \dots, s_H(t)]^T$ be modeled as zero mean jointly wide-sense stationary Gaussian processes and let the received signal at sensor *n* on array *h* be denoted as $s_h\big(t - (\tau_h + \tau_{hn})\big)$. Then all the properties of the received reference sensor signals on each array are described by the *H x H* dimensional cross-correlation matrix (5.4).

$$R_s(\tau) = E[s(t+\tau)s(t)^T] \tag{5.4}$$

The (g, h) individual entry of (5.4) is defined by the cross correlation function of the received signals at the g^{th} and h^{th} arrays as (5.5).

$$r_{s,gh}(\tau) = E\left[s_g(t+\tau)s_h(t)\right] \tag{5.5}$$

All of the properties of the cross correlation function can also be expressed in the frequency domain by taking the Fourier transform of the cross correlation, yielding the cross power spectral density (CPSD) functions and matrix.

$$\begin{aligned} G_{s,gh}(\omega) &= \mathcal{F}\left[r_{s,gh}(\tau)\right] = \int_{-\infty}^{\infty} r_{s,gh}(\tau)e^{-j\omega\tau}d\tau \\ G_s(\omega) &= \mathcal{F}[R_s(\tau)] \end{aligned} \tag{5.6}$$

The diagonal components of the CPSD matrix are the power spectral density (PSD) functions which measure the average signal power per frequency. The off-diagonal components of the CPSD matrix measure the likeness of the spectrums of the signals of two different arrays ($g \neq h$) and can be expressed as

$$G_{s,gh}(\omega) = \gamma_{s,gh}(\omega)\left[G_{s,gg}(\omega)G_{s,hh}(\omega)\right]^{1/2} \tag{5.7}$$

where $\gamma_{s,gh}(\omega)$ is the coherence function defined in Chapter 3 with magnitude lower limit of zero (no correlation) and higher limit of one (perfect correlation). The coherence function measures the frequency domain correlation of two signals received at array g and h. . Partial spatial coherence is formally introduced into the model in (5.7) and its effects are one of the main focuses of (Kozick & Sadler, 2004). Another main assumption of (Kozick & Sadler, 2004) is that there is perfect spatial coherence across individual arrays. This implies that the random propagation effects are assumed to have negligible impact on relative intra-array propagation delays (5.3) and the bearing of the source relative to individual arrays.

Next, additive noise signals are introduced to the received signal model.

$$z_{hn}(t) = s_h\big(t - (\tau_h + \tau_{hn})\big) + w_{hn}(t) \tag{5.8}$$

The noise signals $w_{hn}(t)$ are modeled as zero-mean, jointly wide-sense stationary, Gaussian random processes, which are uncorrelated at each sensor and are uncorrelated with the original source signals $s_h(t)$ such that

$$\begin{aligned} E\big[w_{gm}(t+\tau)w_{hn}(t)\big] &= r_w(\tau)\delta_{gh}\delta_{mn} \\ E\big[w_{gm}(t+\tau)s_h(t)\big] &= 0 \end{aligned} \tag{5.9}$$

where $r_w(\tau)$ is the auto correlation function of the noise and δ is the Kronecker delta. The noise PSD is the Fourier transform of the noise auto correlation function (5.10).

$$G_w(\omega) = \mathcal{F}[r_w(\tau)] \tag{5.10}$$

The array geometry becomes part of the model by considering the array manifold. The array manifold is characterized by frequency and the delay between the reference sensor and the other remaining sensors on an array. Assuming the far-field plane wave assumption and omnidirectional sensor response, the array manifold for an array with N_h sensors, can be expressed as

$$\begin{aligned} a_h(\omega) &= \begin{bmatrix} e^{-j\omega\tau_{h1}} \\ \vdots \\ e^{-j\omega\tau_{hN_h}} \end{bmatrix} \\ &= \begin{bmatrix} \exp\left[j\frac{\omega}{c}\big((\cos\varphi_h)\Delta x_{h1} + (\sin\varphi_h)\Delta y_{h1}\big)\right] \\ \vdots \\ \exp\left[j\frac{\omega}{c}\Big((\cos\varphi_h)\Delta x_{hN_h} + (\sin\varphi_h)\Delta y_{hN_h}\Big)\right] \end{bmatrix} \end{aligned} \tag{5.11}$$

The relative time delay between array g and array h can be expressed as (5.12).

$$\begin{aligned} D_{gh} &= \tau_g - \tau_h \;\; \forall\, g,h \;\text{ where }\; g \neq h \\ &= \frac{1}{c}\big[(x_s - x_g)^2 + (y_s - y_g)^2\big]^{1/2} - \frac{1}{c}\big[(x_s - x_h)^2 + (y_s - y_h)^2\big]^{1/2} \end{aligned} \tag{5.12}$$

Since the received signals were modeled with zero-mean, jointly wide-sense stationary, uncorrelated (with other sensors and with the signals) Gaussian random processes, the CPSD can be expressed as (5.13) where *I* is the identity matrix.

$$G_Z(\omega) = G_s(\omega) + G_w(\omega)I \tag{5.13}$$

Using the source signal CPSD definitions in (5.6)-(5.7) and the noise PSD definition (5.10), the sensed signals CPSD (5.13) can be rewritten as

$$G_Z(\omega) = \begin{bmatrix} a_1(\omega)a_1(\omega)^{\dagger}G_{s,11}(\omega) & \cdots & a_1(\omega)a_H(\omega)^{\dagger}\exp(-j\omega D_{1H})G_{s,1H}(\omega) \\ \vdots & \ddots & \vdots \\ a_H(\omega)a_1(\omega)^{\dagger}\exp(+j\omega D_{1H})G_{s,1H}(\omega)^{*} & \cdots & a_H(\omega)a_H(\omega)^{\dagger}G_{s,HH}(\omega) \end{bmatrix} + G_w(\omega)I \tag{5.14}$$

where the superscripts * denotes the complex conjugate and † denotes the complex conjugate transpose. Equation (5.14) shows that the model includes delays between source and reference sensor, and between the reference sensor and remaining sensors on each array.

Before applying Fisher information and Cramer-Rao lower bound (CRB) theory, the trigonometric functions of the bearings of the source with respect to an array are rewritten in terms of range coordinates.

$$\cos(\varphi_h) = \frac{x_s - x_h}{\left[(x_s - x_h)^2 + (y_s - y_h)^2\right]^{1/2}} \quad \forall \; h$$

$$\sin(\varphi_h) = \frac{y_s - y_h}{\left[(x_s - x_h)^2 + (y_s - y_h)^2\right]^{1/2}} \quad \forall \; h \tag{5.15}$$

Equations (5.12) and (5.15) essentially eliminate the bearing variables from the model and allow the Cramer-Rao bounds to be written only in terms of the (x, y) coordinates of the source.

5.4. Fisher Information Matrix and Cramer-Rao Bounds for Source Localization

Let the unknown parameters for the source localization be the (x, y) coordinates of the source

$$\boldsymbol{\theta} = [x_s, y_s] \tag{5.16}$$

The Cramer-Rao Bounds for the wideband source case (Kozick & Sadler, 2004), (Friedlander, 1984) when sampling at an angular frequency of ω_s, and observing T samples, is (5.17)

$$J_{ij} = \frac{T}{4\pi}\int_0^{\omega_s} \mathrm{tr}\left\{\frac{\partial G_Z(\omega)}{\partial \theta_i} G_Z(\omega)^{-1} \frac{\partial G_Z(\omega)}{\partial \theta_j} G_Z(\omega)^{-1}\right\} d\omega \tag{5.17}$$

where "tr" denotes the trace. To simplify (5.17) into narrowband form, consider a source that has a narrowband PSD which is nonzero in a small band of frequencies limited by the bandwidth $\Delta\omega$ as in (5.18).

$$\omega_0 - \frac{\Delta\omega}{2} \leq \omega \leq \omega_0 + \frac{\Delta\omega}{2} \tag{5.18}$$

If the bandwidth of the narrowband is chosen $\Delta\omega$ small enough that all ω frequency dependent components of (5.17) are approximated well by their evaluation at the value of ω_0, then the narrowband FIM can be approximated by (5.19).

$$J_{ij} = \frac{T\Delta\omega}{\omega_s} \mathrm{tr}\left\{\frac{\partial G_Z(\omega_0)}{\partial \theta_i} G_Z(\omega_0)^{-1} \frac{\partial G_Z(\omega_0)}{\partial \theta_j} G_Z(\omega_0)^{-1}\right\} \tag{5.19}$$

With the FIM expression found, the CRB theory in Chapter 3 can now be applied. The narrowband expression shows that the Cramer-Rao bound (the inverse of the FIM) will be directly proportional to the sampling frequency, and will be inversely proportional to the number of samples and the bandwidth. As noted before, these CRB expressions

represent jointly processed data from all the sensors on all the arrays, thus sidestepping the issue of communication delays between arrays and a data fusion center.

5.5. Methods to Graph Source Localization Cramer-Rao Bounds

Once the FIM is found the source localization performance can be evaluated by solving the quadratic ellipse equation (5.20) for (x, y) coordinates.

$$[x \quad y]\, J \begin{bmatrix} x \\ y \end{bmatrix} = 1 \tag{5.20}$$

The ellipse equation is solved by finding the eigenvalues and eigenvectors of the FIM and applying the following relationships discussed in (Kozick & Sadler, 2004), (Strang, 1988).

$$\Lambda = \{\lambda_1, \lambda_2\} = \text{set of eigenvalues of } J$$

$$\begin{bmatrix} v_{1_x} \\ v_{1_y} \end{bmatrix} and \begin{bmatrix} v_{2x} \\ v_{2_y} \end{bmatrix} = \text{corresponding eigenvectors}$$

$$\lambda_1 < \lambda_2 \quad \text{(eigenvalues relationship)}$$

$$CRB_{major\ radius} = \frac{1}{\sqrt{\lambda_1}} \tag{5.21}$$

$$CRB_{minor\ radius} = \frac{1}{\sqrt{\lambda_2}}$$

$$\phi_{rot.\ ellipse} = \text{atan2}\left(v_{1_y}, v_{1_x}\right) = \text{four quadrant inverse tangent}$$

The eigenvalues of the FIM measure the amount of Fisher information about the unknown parameters and the eigenvectors measure the relative angle of rotation. Equation (5.21) shows that the major and minor radii of the CRB are inversely proportional to the square root of the eigenvalues of the FIM, which is consistent with the CRB FIM inverse relationship. With the CRB defined for source localization and a

method to graph the performance limitations, two examples were evaluated to show the accuracy of the CRB.

5.6. Example 1: Parallel Array Formation, Cramer-Rao Bound Evaluation

The following parameters were used for all the examples: a bandwidth of 1 Hz centered at $\omega = 2\pi(11.8e3)$, $SNR = 10$ dB, T=1000 samples, $f_s = 24\text{e}3\ samples/s$, and $\gamma_s(\omega) = [0, 0.5, 1]$. Three CRB ellipses were graphed per source location. The largest ellipse represents the zero coherence case, and the smallest ellipse represents the perfect coherence case $\gamma_s(\omega) = 1$. The array parameters were set to have a sensor spacing of 0.022m, using 7 sensors on each array. The source locations were moved away from the center position in increments of 50m while the array locations were kept constant. The sizes of the arrays were magnified to show their orientation.

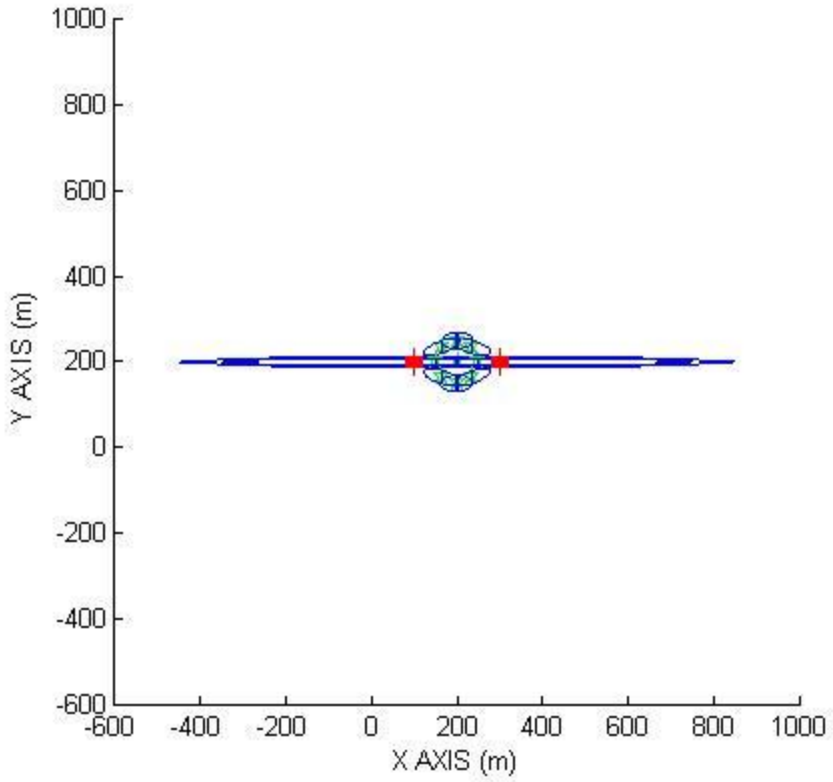


Figure 5-3: CRB: parallel formation, source at center and 50m away

Figure 5-3 illustrates that a source that lies on the broadside angle for both arrays produces the largest CRB bound. The double-broadside position is a blind spot which can sense almost perfectly on one axis, but with no precision on the other remaining axis. This special source/array combination is consistent with the physical limitedness of the capability of both arrays being able to accurately measure angles, but not range.

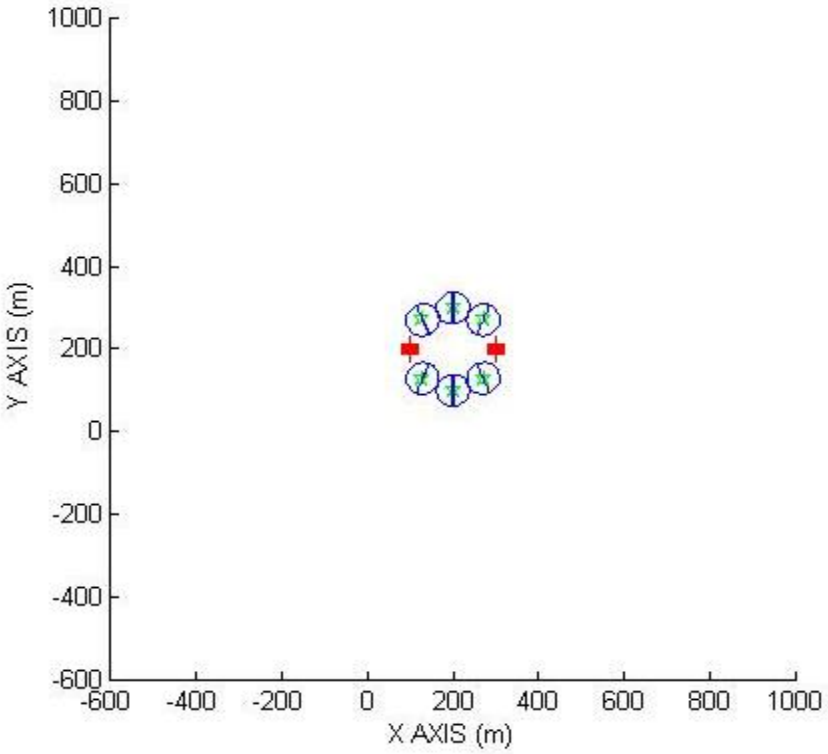


Figure 5-4: CRB: parallel formation, source at 100m away

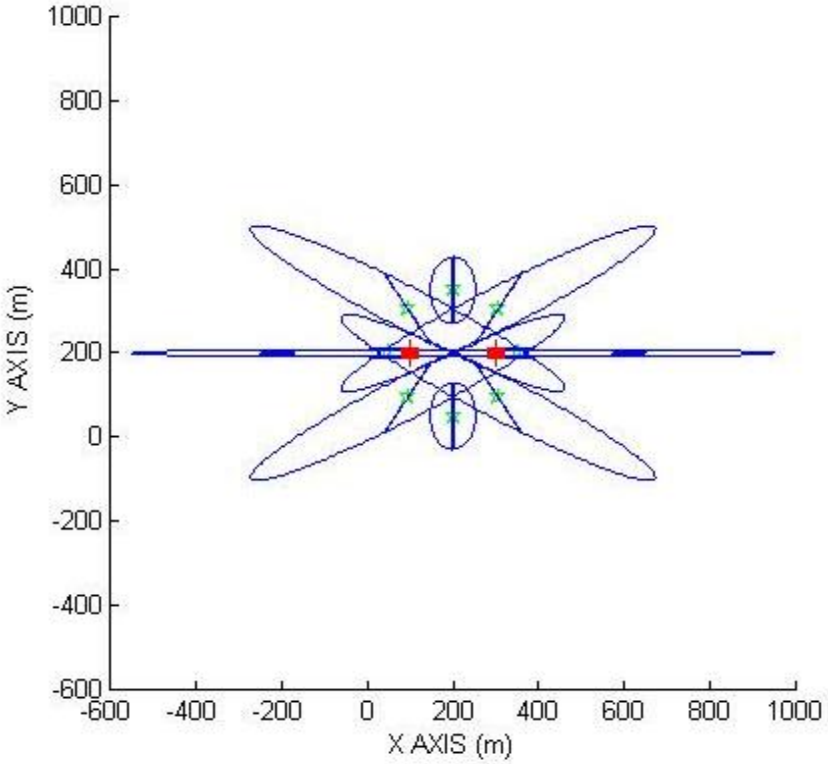


Figure 5-5: CRB: parallel formation, source at 150m away

Figure 5-4 and Figure 5-5 illustrate that the source moving further away from the center of the parallel formation is characterized by an increase in the size of the CRB's. It can also be observed in Figure 5-5 that when the source is at an end-fire sensing angle for one array, the direction of the CRB relies heavier upon the other array that is in a better position and angle to sense it.

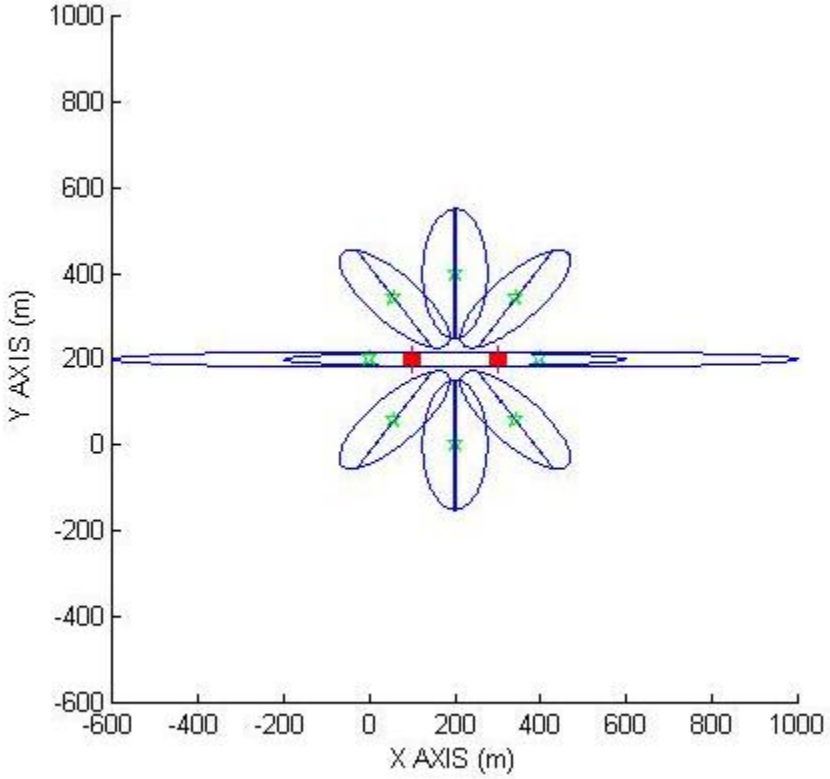


Figure 5-6: CRB: parallel formation, source at 200m away

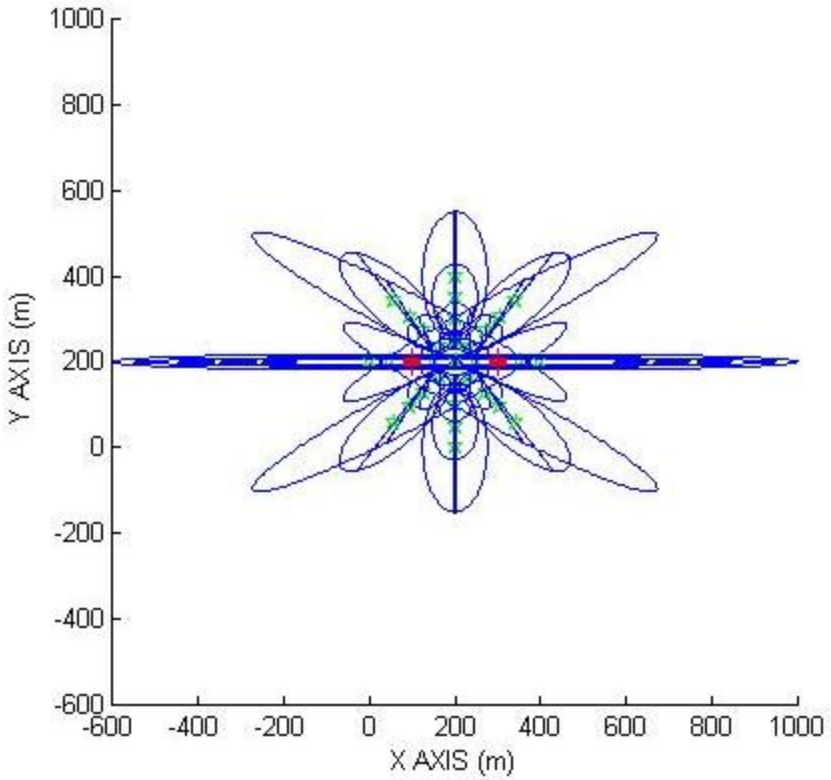


Figure 5-7: CRB: parallel formation, Source at 0, 50, 100, 150, 200m away

Graphing all the above cases at once, it is shown that the parallel formation generates symmetric CRB's about the vertical and horizontal center source location.

5.7. Example 2: Cross Array Formation, Cramer-Rao Bound Evaluation

For the next example, an array/source combination is presented where the two arrays are at 90 degrees from each other, and the source center starts at broadside for both arrays.

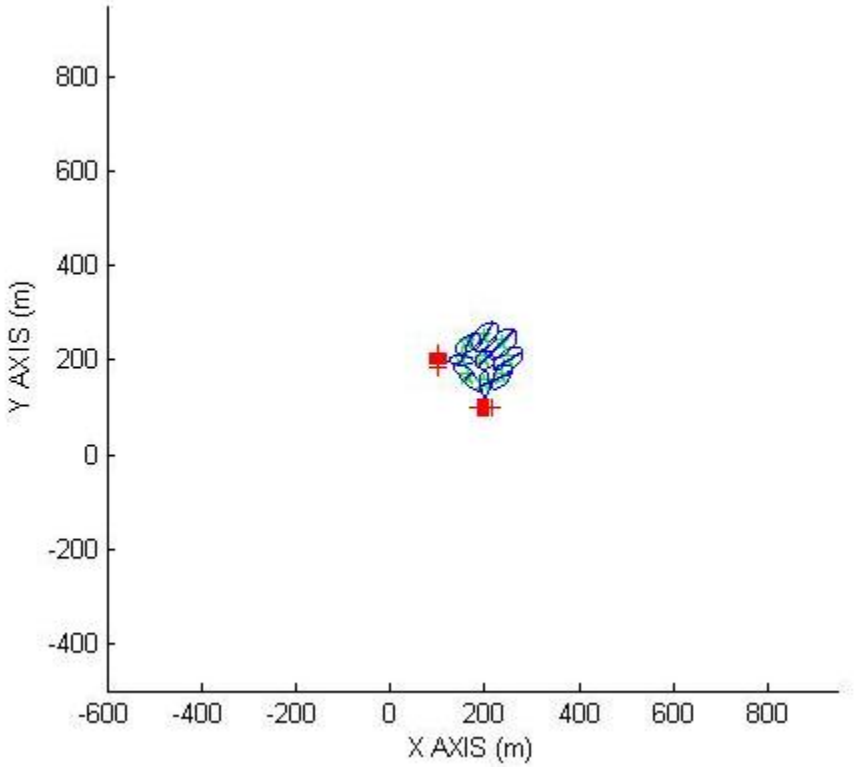


Figure 5-8: CRB: Cross Formation, Source at Center and 50m Away

Figure 5-8 shows that when the arrays are in cross formation, the center source location produces the smallest CRB ellipse. As the source moves further away from broadside for both arrays, the size of the CRB increases.

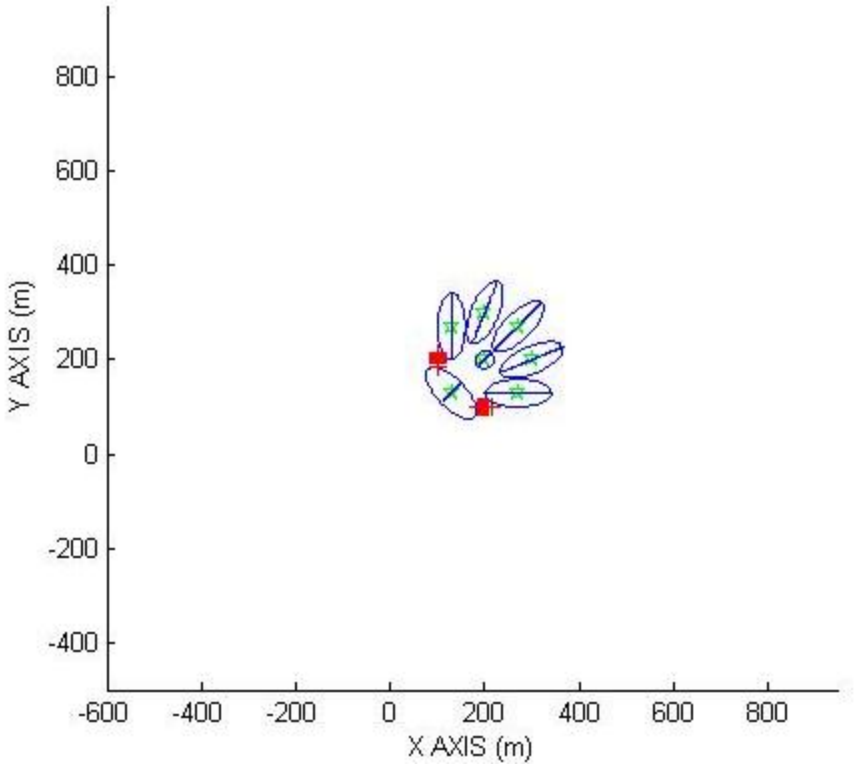


Figure 5-9: CRB: Cross Formation, Source at 100m Away

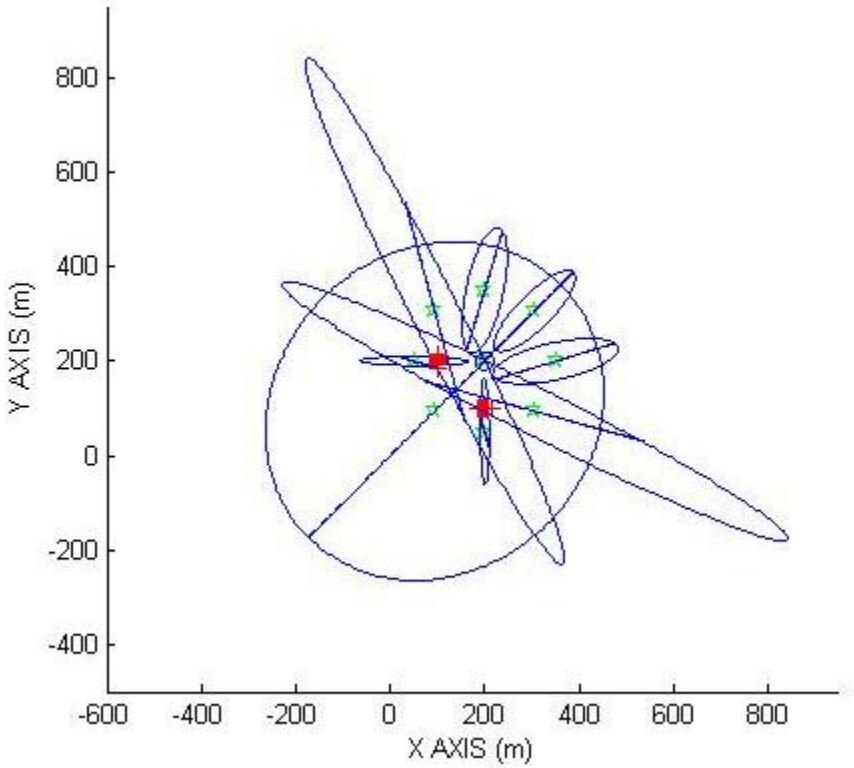


Figure 5-10: CRB: Cross Formation, Source at 150m Away

Figure 5-10 shows that if the source is located close to end-fire for both arrays and receiving signals with zero coherence, then the CRB will have an extremely large size in both the vertical and horizontal directions.

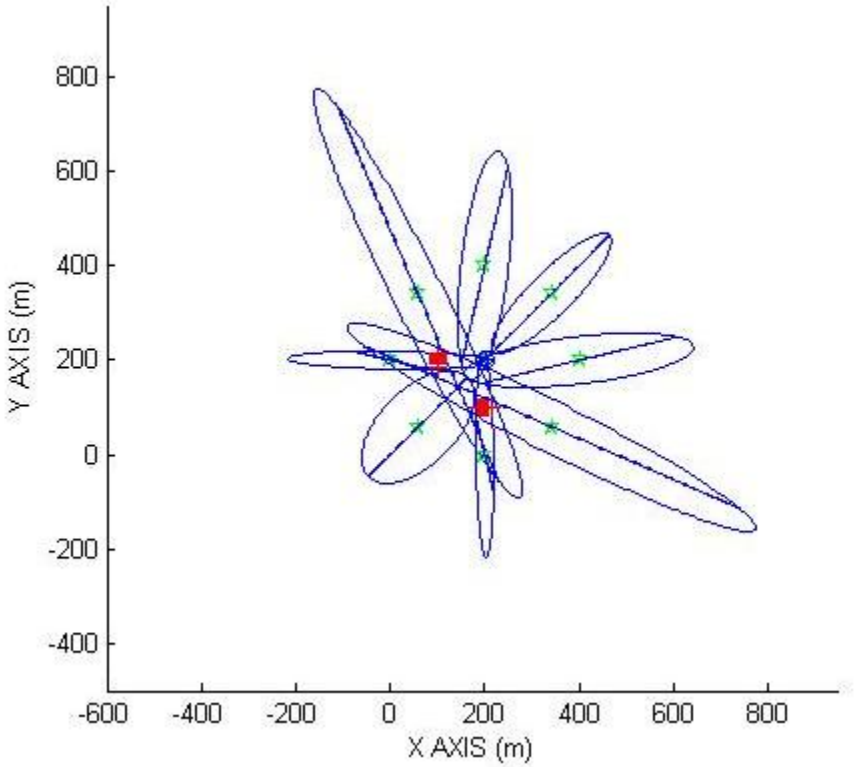


Figure 5-11: CRB: Cross Formation, Source at 200m Away

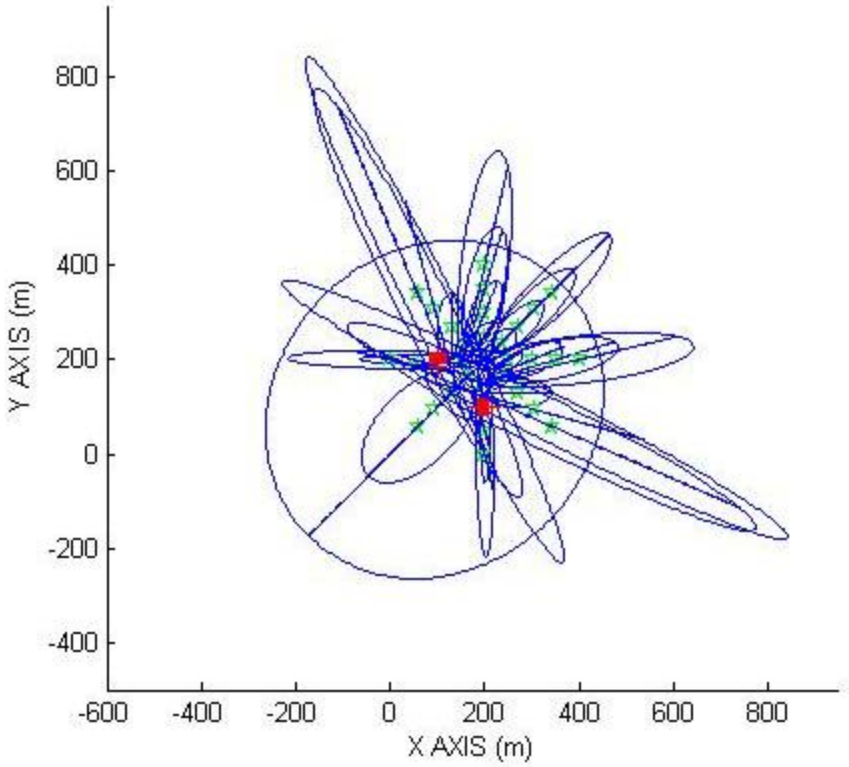


Figure 5-12: CRB: Cross Formation, Source at 0, 50, 100, 150, 200m away

By graphing all the above source locations, it is shown that the source localization will have less error in the direction where the sensing capabilities are the best due to array orientation.

5.8. Conclusions and Extensions

The Cramer-Rao bounds were graphed and verified for the case of two uniform linear arrays, one non moving source. After review of the equations and examples, it can be concluded that the CRB's size and direction are defined by the geometry (bearing and range to the source), the received signals (frequency, cross power spectral density, coherence, signal to noise ratio, and number of samples gathered), and array (orientation and configuration) properties. The physical limitations of uniform linear arrays were verified by special cases that yield unique CRB's. An extension to multisource

localization using a network of arrays using a very similar model can be found in (Erling, Roan, & Gramann, 2007).

Chapter 6. Literature Review for Event monitoring, Event Detection, and Anomaly Detection

6.1. Introduction

Event monitoring, event detection, and anomaly detection are areas of active current research. These topics, although closely named, are distinguished from each other in the following manner. In this thesis, event monitoring refers to the process by which sensed data is analyzed for general trends and relationships are extracted. Event detection is referred to as the process by which data is actively searched for a pattern. This pattern is usually a defined signature unique to the event that is sought after. Anomaly detection, also sometimes referred to as outlier detection, is referred to as the process of finding abnormal events by a comparison to some normal events. Although the above definitions of the research areas overlap and have seemingly close names, the conceptual methods applied are quite different from each other. The following is a literature review of these three topics that are not explicitly discussed in Chapter 8 and Chapter 10.

6.2. Event Monitoring

Event monitoring and the placement strategy of sensors is an ongoing and current research topic in literate. A good example of a monitoring network of sensors is (Bodik, Hong, Guestrin, Madden, Paskin, & Thibaux, 2004). The goal of this study was to provide a free access data set of time-stamped topology information, along with humidity, temperature, light and voltage values so that researchers could benchmark

scientific tests on the data. Another good example of a monitoring network is (Widmann & Bretherton, 1999). In this study, 522 stations measured precipitation was measured daily for the Pacific Northwest from years 1949 to 1994. Although the above event monitoring networks and data sets measure specific physical events, they are alike because the recorded measurements were taken by sparse sensor placements and at sparse interval times. The sparseness described in the above experiment produces data with similar sparse properties as those that are studied in Chapter 8 and Chapter 10.

(Krause, Singh, & Carlos, 2008) operate on the data from (Bodik, Hong, Guestrin, Madden, Paskin, & Thibaux, 2004) and (Widmann & Bretherton, 1999) using a Gaussian process model based technique to determine the relationship between sensors. One of the key contributions made is the use of a method to estimate the nonstationary spatial covariance structure (Nott & Dunsmuir, 2002), which defines the Gaussian process kernel. The model based technique is a typical event monitoring method to learn general trends. Chapter 8 uses a technique similar to the Gaussian process, but uses the Poisson distribution and histograms to characterize the data based on average number of occurrences during a time period. Unlike the event monitoring Gaussian process techniques discussed above, the goal of Chapter 10 is not to learn general trends, but to relate and detect individual agents moving through a dynamic area of surveillance.

6.3. Event Detection

The research area of event detection is widespread and applies to many different applications. Some literature focuses on the actual detection of an event while others try to enhance event detection performance by strategically sensing according to some objective function (Vaisenberg, Mehrotra, & Ramanan, 2009). Regardless of the

application, the main approach to event detection is matching data with a predefined pattern (Xue, Luo, Chen, & Liu, 2006), (Li, Liu, & Chen, 2008), (Guralnik & Srivastava, 1999), (Xue, Luo, & Pung, 2011). These patterns can be spatial, temporal, or spatiotemporal. Chapter 8 presents a Poisson distribution data model, non-pattern matching, and threshold based anomaly detection method, whereas Chapter 10 presents a non-pattern matching, non-threshold based event detection method.

(Xue, Luo, & Pung, 2011) state the main problems of event detection in a very concise manner. They note two key problems and issues which are:

1. How to model an event based on the sensor deployment and user inputs
2. How to develop an algorithm to detect an event by data matching

They also make the typical key observation that an event generates a specific spatiotemporal signature which can be thought of as a distribution itself. They call these event data distributions. One key component that (Xue, Luo, & Pung, 2011) use is the application of regression analysis to learn the event data distributions as opposed to the typical Gaussian process method. This process of learning the event data distributions is the process of model selection where the goal is to find the best fitting model. The event data distributions learned from regression techniques can be formulated into contour maps (Xue, Luo, Chen, & Liu, 2006). The main contribution of (Xue, Luo, & Pung, 2011) to the event detection area is a concept of relative region map matching. The concept is called relative because the matching pattern can float within a grid and its purpose is to detect dynamic events.

Another main event detection technique involves assigning a set threshold value to detect an event. (Li, Liu, & Chen, 2008) state that set threshold based approaches

work for the detection of simple events, but fail in the detection of more complex spatiotemporal events. Their contribution is describing complex physical phenomena in three dimensions with spatiotemporal data patterns for the application of gas leakage sensing in coal mines. Particular to this application, they develop an event feature matching pattern that shows the spreading or evolution of an event pattern. Their method is based off of an event being detected only if the orthogonal polyhedral models, which are polyhedrons with three orthogonal faces, of the data map and the defined pattern match. Although the title of (Li, Liu, & Chen, 2008) suggests that it is a novel approach by being a non-threshold based method, it is still an exact pattern matching application.

Depending on the application, event detection sometimes requires different approaches to finding an event and different definitions of the event to detect. (Guralnik & Srivastava, 1999) consider the event detection problem from the statistical viewpoint of change-point detection. Change-points are defined as time points at which the behavior of the data changes. Their main contribution consists of two algorithms that detect change-points either from a non-real-time viewpoint, where an entire batch of data is available, or from a real-time perspective, where data is gathered incrementally. A maximum likelihood criterion is then used to determine whether further segmentation and the detection of more change-points in the segment of data should be applied or not.

6.4. Anomaly Detection

Anomaly detection, sometimes referred to as outlier or novelty detection has been studied in many forms throughout literature. According to (Barnett & Lewis, 1994), informal approaches to outlier detection can be traced back to the middle of the eighteenth century. The problem was posed as whether or not to draw a line between

some observations dividing them into those that should be retained for use, and those that should be rejected. Since then, (Barnett & Lewis, 1994) concisely stated two major goals consistent with the work in this thesis as:

1. The multiple ways in which outliers arise
2. The different goals for which the outliers are examined for

The first point above refers to the physical event that the data represents. (Barnett & Lewis, 1994) note that two major causes for outliers are inherent variability and imperfect measurement error. The second point above refers to the intention of the user after data has been examined for outliers. The outliers found could be retained, rejected, or robustly accommodated for (Barnett & Lewis, 1994). Accommodation methods essentially keep outliers, but move or replace them according to strategy, while keeping the inference of the entirety of the data set mostly undisturbed.

(Hawkins, 1980) defines an outlier as "an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism". The mechanisms referred to in (Hawkins, 1980) are divided into two main categories:

1. The data obtained comes from a heavy tailed distribution (this is equivalent to the inherent variability mentioned in (Barnett & Lewis, 1994))
2. The data comes from two or multiple distributions.

Interestingly, the first point above is the same observation as (Barnett & Lewis, 1994), but the second point is unique. The proposed event detection method presented in Chapter 10 is in line with the second point above. (Hawkins, 1980) also notes that there are several main ways to treat outliers: 1) estimation (rejection of outliers is a form of

obtaining a robust estimator), 2) smoothly vary the weights of the outliers, 3) move outliers to a prefixed position (Winsorization), 4) sound an alarm or warning, and 5) classification of the observations which is consistent with the second point above from (Barnett & Lewis, 1994). Having considered the above discussion from (Barnett & Lewis, 1994) and (Hawkins, 1980), it can be concluded that outlier detection depends strictly on the physical application that data is generated from and the intent or purpose of the user.

Much work in the area of anomaly detection has been summarized by review and survey papers (Hodge & Austin, 2004), (Patcha & Park, 2007), (Chandola, Banerjee, & Kumar, 2009), (Markou & Singh, 2003), (Ben-Gal, 2005), (Osborne & Overbay, 2004). (Hodge & Austin, 2004) state three fundamental approaches to outlier detection as: 1) find the outliers without prior knowledge about the data, 2) model what is normal and abnormal, and 3) model only what is normal or what is abnormal. (Patcha & Park, 2007) addresses anomaly detection from a cyber intrusion detection perspective. (Chandola, Banerjee, & Kumar, 2009) distinguish between types of anomalies as point, contextual, and collective. The Poisson data model anomaly detection methods in Chapter 8 are designed for point and collective anomalies, whereas the event detection method proposed in Chapter 10 is contextual. (Chandola, Banerjee, & Kumar, 2009) concludes that contextual and distributed anomaly detection is increasingly applicable. (Markou & Singh, 2003) address novelty detection which they define as the identification of new unknown data signatures or signals that machine learning systems don't consider when in a training phase. (Ben-Gal, 2005) provides a short taxonomy of outlier detection methods mostly consistent with all the above literature. (Osborne & Overbay, 2004) give

many and more specific reasons of possible causes of outliers which is consistent with point one from (Barnett & Lewis, 1994) above. Regardless of the source, a general taxonomy of contributing research areas, criterion for method choice, and applied methods in the area of anomaly detection is depicted below in Figure 6-1.

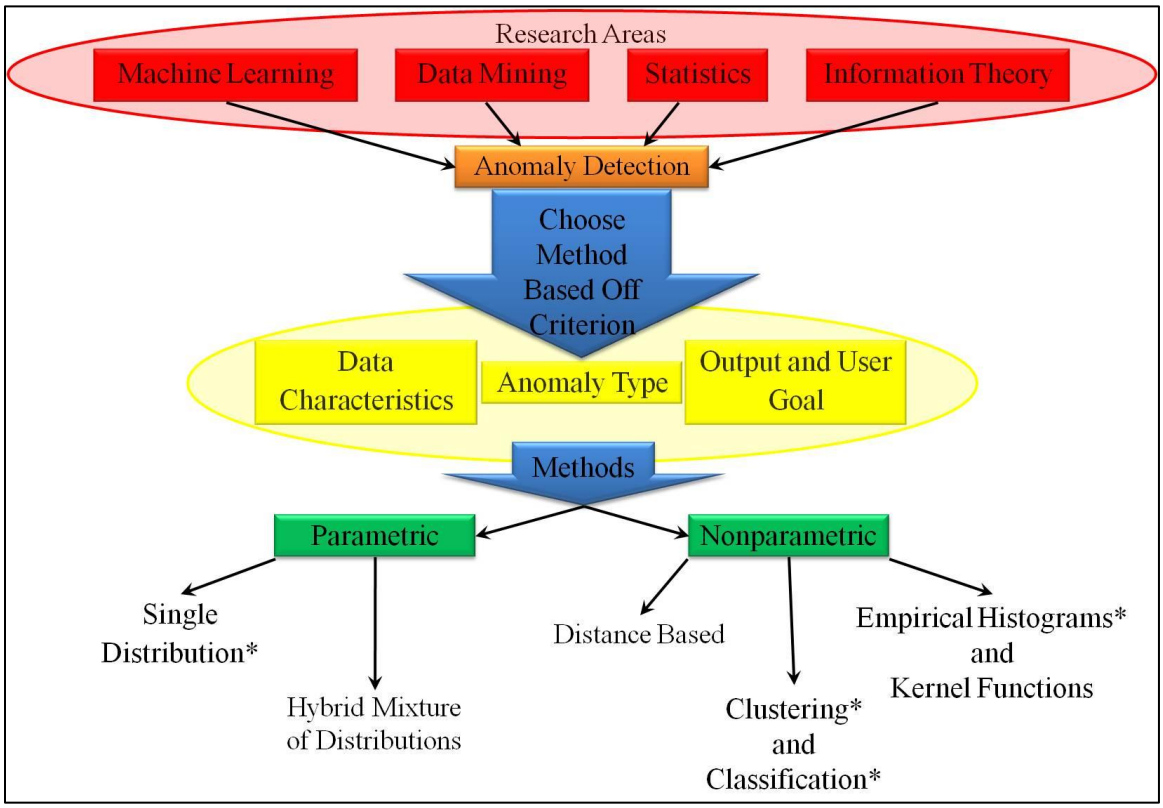


Figure 6-1: Taxonomy of anomaly detection: similar research areas, criterion, and methods where * denotes a method applied directly in this thesis

The Poisson distribution has been used in the area of anomaly detection to: model time series of network traffic flow (Chen, 2008), monitor the occurrences of clustered interactions between humans (Panangadan, 2004), and investigate anomalies driven by changes in weather, such as El-Niño (Rajagopalan, 1997) . Both (Chen, 2008) and (Panangadan, 2004) apply the Poisson model and flag anomalous events after some comparison to a threshold value, whereas (Rajagopalan, 1997) uses a Poisson rate parameter to characterize the probability of trends in the anomaly rate for yearly weather.

It is unclear from the previous work, however, what dimension to model as a Poisson process when dealing with multidimensional datasets such as that in this study (year, month, day, minute, etc.). For the scenario of monitoring trajectory-based activity of a facility discussed in Chapter 8, consistent daily trends were found via inspection when using a time bin width of one minute. The Poisson distribution is known to model arrival rate processes when (Brooks, 2007): 1)the events occurring can be counted in whole numbers, 2)the occurrences are independent from each other, 3)the average frequency of occurrence (Poisson parameter) for a given time period is known or can be found, and 4)all relevant information about an event is contained in an event occurring, but no information is contained in an event not occurring. The last point above states the fundamental difference between the binomial (7.5) and Poisson (7.8) PMF's. In summary, the Poisson distribution was chosen for anomaly detection because of its direct applicability to arrival count and arrival rate problems, and because they model the occurrence of rare events within a given time interval (Devore, 1995).

Much work has been done in the area of ambient intelligence using passive point sensor networks (Singla, Cook, & Edgecombe, 2010), (Wren & al, 2006) in closed smart environments, but little work has been done in the area of sensor processing in open environments where many different agents are present (Nakashima & al., 2010). Two main obstacles in processing this data are handling sensor noise (false alarms) and relating sensor alarms to individual agents moving through the environment. Additionally, these alarms must be able to describe activity from a macro and microscopic viewpoint while relating activities to normal or abnormal events. To accomplish this, Chapter 10 presents an event detection method that uses geospatial

intelligence data to relate data from multiple sensors to multiple agents' physical trajectories.

Chapter 7. Background: Event Monitoring, Event Detection, and Anomaly Detection

7.1. Introduction

The search for general trends in data is called event monitoring. Event detection, however, is the result of a process where data is monitored and searched for the occurrence of a specific event. If the event that is searched for is deemed abnormal, the problem is called anomaly detection.

The background discussed in this chapter will cover the topics needed to understand the methods to statistically model a point sensor network. Point sensors, such as beam-break (BB) and passive infrared (PIR) sensors, are triggered when agents move across a sensor's sensing line, area, or path. Statistical distribution parameter estimation is a straightforward method to conduct event monitoring. The background information discussed in this chapter will discuss pertinent continuous (normal and Student T) and discrete (Bernoulli, binomial, multinomial, and Poisson) distributions. Nonparametric estimation will also be introduced via histogram and smoothing techniques.

After the statistical distributions are discussed, methods to apply parameter estimation are introduced. Classification and clustering techniques such as the K-means and expectation maximization (EM) algorithms are a proven method to learn the parameters of a statistical distribution. Once the data is fit to the framework of a statistical distribution, and patterns are learned, event and anomaly detection methods can be applied. This chapter is concluded with a literature review of event monitoring, event detection, and anomaly detection.

7.2. Axioms of Probability

To more formally define probability, it is appropriate to first review the three main axioms of probability stated as (Taylor & Karlin, 1998)

1. The sample space is defined as a set Ω, and the elements within Ω correspond to every possible outcome of an experiment and the probability of the sample space is $P[\Omega] = 1$
2. The probability of any event A is nonnegative, $P[A] \geq 0$
3. Let $A_1, \dots, A_n$ be events, then if events $A_1, \dots, A_n$ are disjoint, then $P[\bigcup_{n=1}^{\infty} A_n] = \sum_{n=1}^{\infty} P[A_n]$

By combining the first and second axioms, a probability measure summary is created that can be expressed as

$$0 = P[\emptyset] \leq P[A] \leq P[\Omega] = 1 \quad \text{for} \quad A \in \Omega \tag{7.1}$$

where $\emptyset$ is the null or empty set.

7.3. Statistical Distributions

Although there are numerous statistical distributions to study the properties of, the following statistical distributions are directly pertinent to our applications. The distributions are separated into two major categories, continuous and discrete.

7.3.1. Continuous Distributions

The normal, referred to synonymously as a Gaussian, distribution is the most important statistical distribution in probability and statistics (Devore, 1995). A continuous probability density function (PDF) gives the probability that a random

variable (RV) is equal to a specific value, for example $X = x$. The PDF of a normal distribution given the mean μ and variance σ^2 is defined as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \text{ for } -\infty < x < \infty \tag{7.2}$$

Recall that the definitions for the mean and variance are defined in terms of central moments in Chapter 3.

Unlike the normal distribution governed by two parameters, the T distribution (also called student t distribution) is governed by only one parameter, ν, called the degrees of freedom. The PDF of the T distribution is defined as (7.3)

$$p(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)} \tag{7.3}$$

where Γ is the Gamma function defined by $\Gamma(n) = (n-1)!$. The overall shape of the T distribution is the same as the normal distribution except that it is always centered around a zero mean value, the peak probability is lower, and the probabilities are spread wider than the normal distribution (Devore, 1995) as shown below in Figure 7-1.

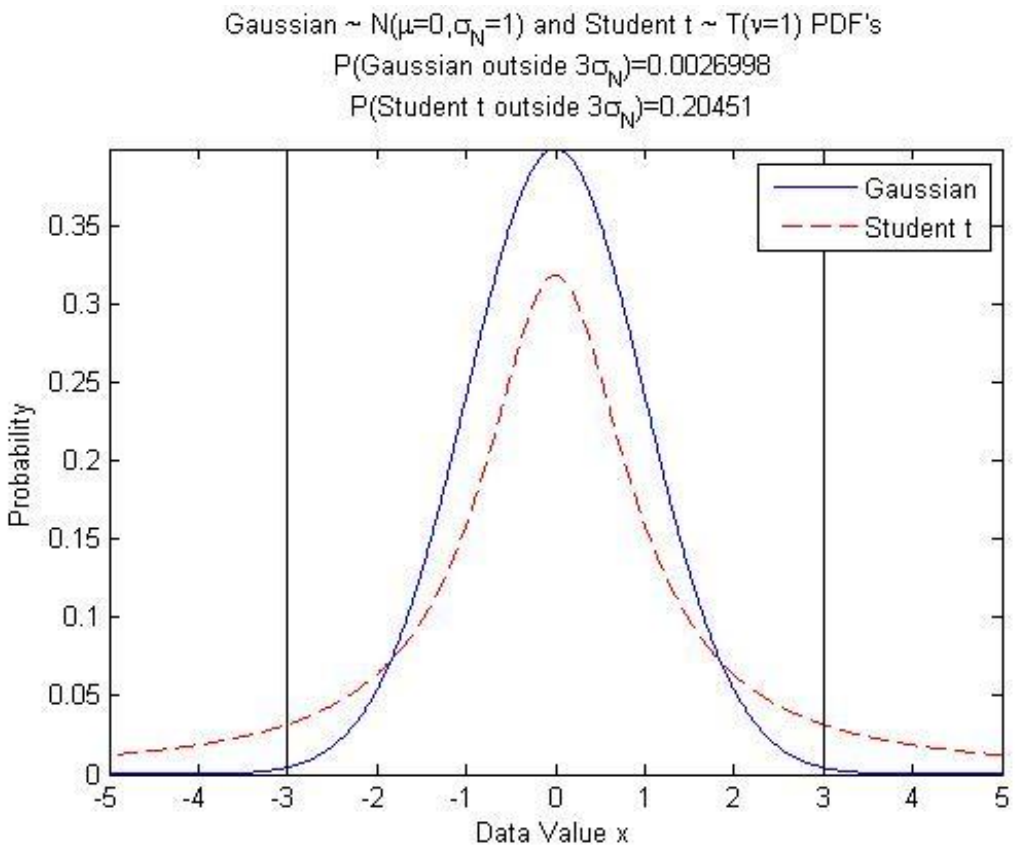


Figure 7-1: Comparison of normal and T distributions

Figure 7-1 illustrates that even though the two distributions are very alike in shape, that fitting the data to the wrong distribution could have strong negative effects on the outcome of the calculation. The above example illustrates that when searching for outliers lying outside the plus/minus 3σ data value, the amount of probability of the outlier values is completely different for the two distributions.

7.3.2. Discrete Distributions

The Bernoulli distribution is the simplest discrete probability distribution that symbolizes a single experiment where the outcome is either true or false. The probability mass function (PMF) of the Bernoulli distribution is defined as

$$f(k;p) = \begin{cases} p & \text{if } k = 1 \\ 1-p & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases} \tag{7.4}$$

The binomial distribution is a simple discrete probability distribution that is an extension of the Bernoulli distribution. The popular use of this statistical distribution is due to the fact that many scientific procedures conform to its definition requirements which can be stated (Devore, 1995) in the following manner.

1. An experiment consists of a sequence of *n* trials. The number of trials *n* is fixed before the experiment is begun
2. Each of the *n* trials is identical to the other
3. Each of the *n* trials is independent of the other trials (Ex. the outcome of the first trial does not affect the outcome of the second trial)
4. The probability of success/failure, denoted by *p* is constant for all *n* trials

The above requirements for an experiment to be modeled by a binomial distribution show that the distribution is essentially a chain or series of Bernoulli true/false experiments. The PMF of the binomial distribution is defined as (7.5)

$$b(k;n,p) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases} \tag{7.5}$$

where *k* is the number of successful trials out of *n* total trials and $\binom{n}{k}$ denotes the number of combinations of size *k* that can be made from *n* independent trials and is defined as

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \tag{7.6}$$

Recall that combinations are unordered subsets (Devore, 1995) and therefore measure the number of unordered possibilities of outcomes for an experiment.

The multinomial distribution is a generalization of the binomial distribution for multiple possible outcomes. The joint PMF for the multinomial distribution of *r* possible outcome variables is (7.7) (Taylor & Karlin, 1998)

$$f(k_1, \ldots, k_r; n, p_1, \ldots, p_r) = \begin{cases} \frac{n!}{k_1! \cdots k_r!} p_1^{k_1} \ldots p_r^{k_r} & \text{if } k_1 + \cdots + k_r = 1 \\ 0 & \text{otherwise} \end{cases} \quad (7.7)$$

where $p_i > 0$ for $i = 1, \ldots, r$ and $\sum_i p_i = 1$.

The Poisson distribution gives the probability that *k* events occur given the expected or average number of events that occur, λ. The PDF of the Poisson distribution is defined as

$$P(k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!} \text{ for } k = 0, 1, 2, \ldots \quad (7.8)$$

Like the T distribution, the Poisson distribution only has a single parameter, λ, which is the expected number of occurrences.

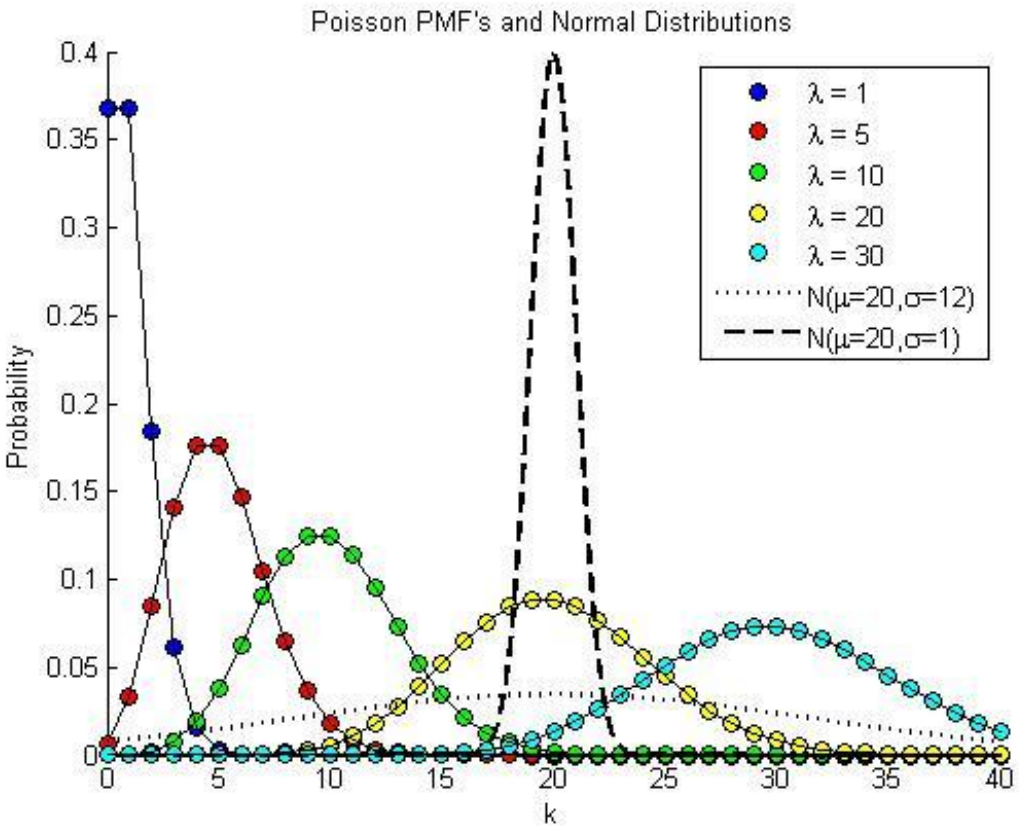


Figure 7-2: Comparison of Poisson probability mass functions when varying λ and normal distributions

Figure 7-2 shows a comparison of the Poisson and normal distributions. For the Poisson distribution, as the expected number, λ, is increased, both the mean and the variance are increased simultaneously. This is contrasted to the normal distribution where the mean and variance can be adjusted individually.

7.4. Histogram Nonparametric Method

Histograms are the oldest and most commonly used nonparametric density estimator (Wand & Jones, 1995). In contrast to the parameterized statistical distributions discussed above, histograms are characterized by the size or width of the bins, the number of bins, and the positioning of the bin edges. The bins are not required to be of uniform size, but

this is the simplest case. The number of observations or occurrences in each bin is counted. To use histograms in the context of probability, each bin is normalized by the total amount of observations in all bins considered at once. This can be expressed in the following manner. Let *N* be the total number of bins such that $n = \{1, \dots, N\}$ and let k_n be the number of occurrences of an event in bin *n*. Then the probability that a random variable will occur in the n^{th} bin can be expressed as

$$p(X = n) = \frac{k_n}{\sum_{n=1}^{N} k_n} \tag{7.9}$$

7.5. Methods to Learn Parameters of Distributions

7.5.1. K-means Clustering Algorithm

The K-means clustering algorithm is a well known algorithm used to group data (Kadous, 2002). The algorithm is named as such because "K" represents the number of clusters or groups in the data that the mean values will be calculated for. Given that the number of groups is known, the algorithm works by soft classifying a data point to a cluster by minimizing the distance to a centroid. The centroid locations are updated and the process is started again. The classification is called soft because each iteration allows the possibility of a data point being classified in the same or a new cluster. The algorithm is outlined by (Alpaydin, 2004), (Kadous, 2002), and (Redmond & Heneghan, 2007). Although Euclidean distance is primarily used in the algorithm, any measurement definition may be substituted for the distance depending on the application.

7.5.2. Expectation Maximization (EM) Algorithm

The Expectation Maximization (EM) algorithm is a more general algorithm for parameter estimation. The EM algorithm was defined by (Dempster, Laird, & Rubin,

1977) which combined previously proposed special circumstance work by many earlier authors. The EM algorithm works by alternating between two major steps until some user specified precision tolerance is reached (McLachlan & Krishnan, 2008), (Dellaert, 2002), (Dinov, 2008). The first step of the EM algorithm operates by finding the expectation of the data log-likelihood function with respect to current parameter estimates. The second step maximizes the expectation from the first step and provides new parameter estimates. If the tolerance is not reached, the two steps are repeated iteratively until it is, or an iteration limit is reached.

7.5.3. Example of K-means Clustering and EM Algorithm

The following are graphical examples of K-means and EM algorithm clustering on the same mixture of two Gaussians.

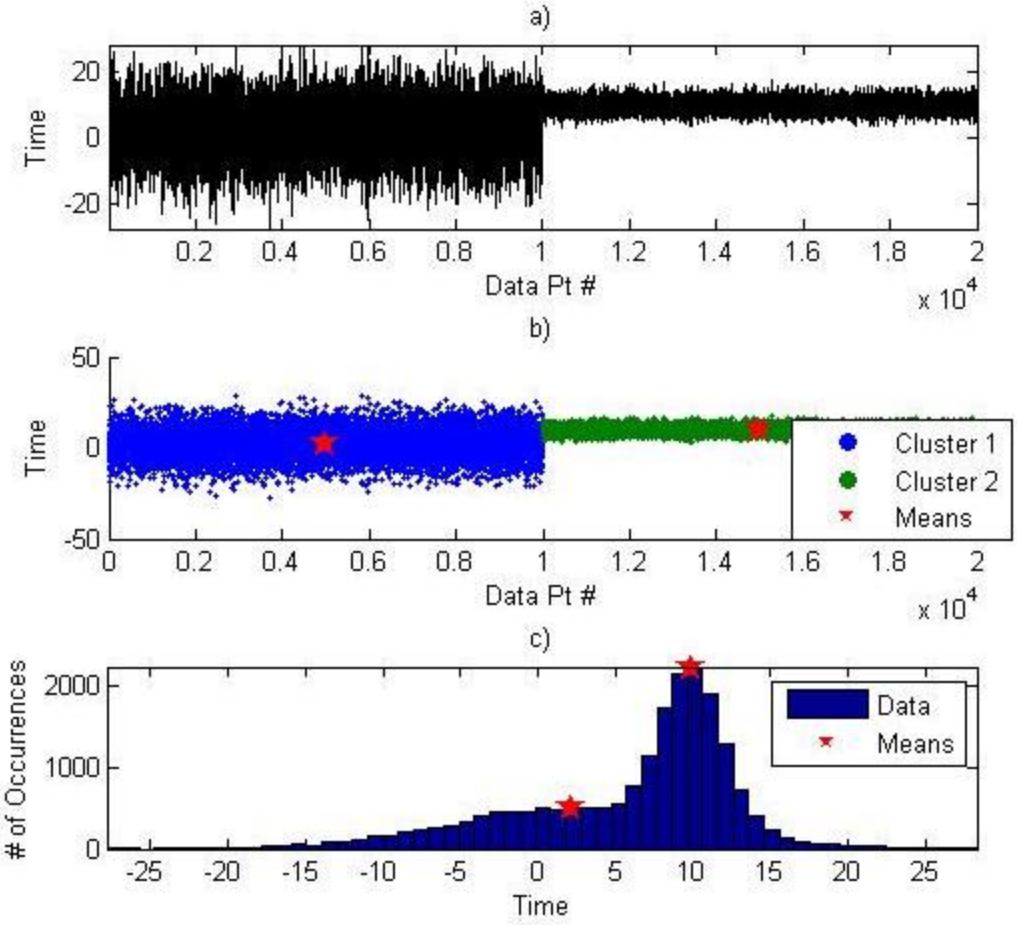


Figure 7-3: K-Means algorithm example: a)Mixture of two Gaussian distributions in time, b)K-Means applied to time, c)K-Means output plotted over histogram

Next the EM algorithm is applied to the same data set of a mixture of two Gaussian distributions.

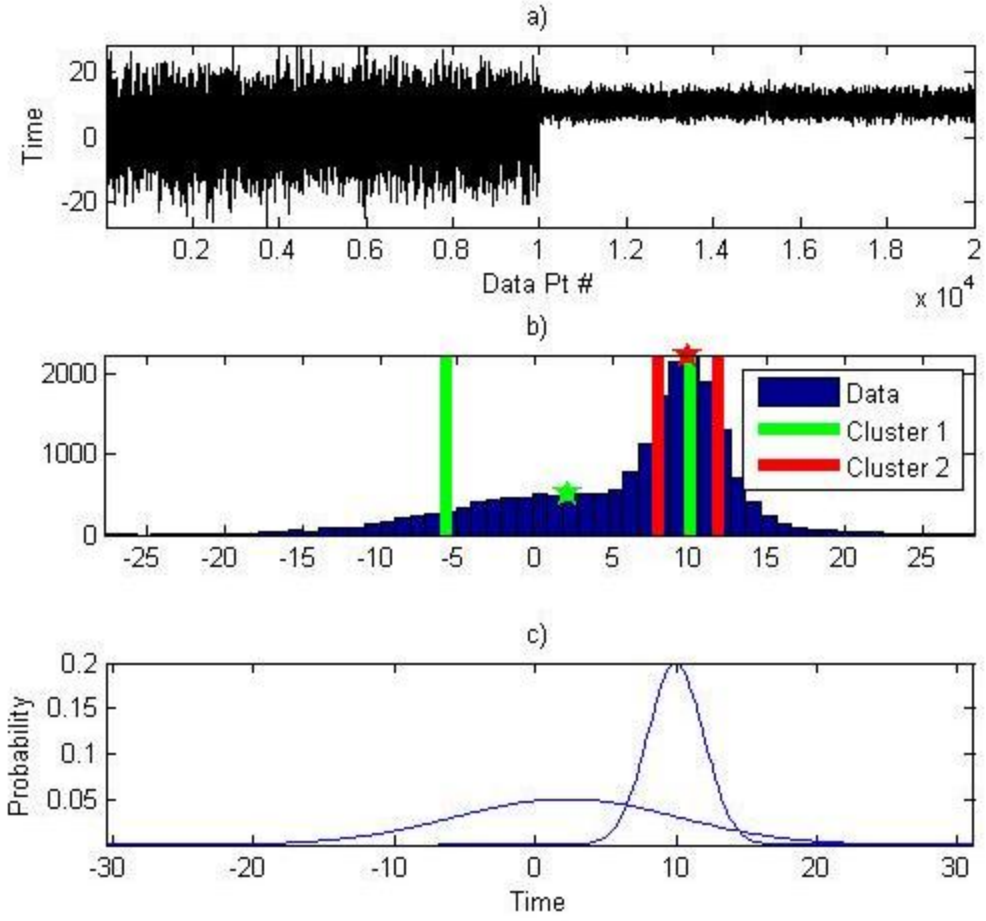


Figure 7-4: EM algorithm example: a)Mixture of two Gaussian distributions in time, b)EM algorithm applied to time to find mean and standard deviation estimates, c)probability density function of output of EM algorithm

The resulting graphical examples show how the K-Means and EM algorithm are useful to estimate the parameters of a Gaussian mixture model (GMM). Both algorithms can be applied in higher dimensions as well.

Chapter 8. Poisson Data Model Anomaly Detection

8.1. Introduction

This chapter considers a Poisson data model for the problem of anomaly detection for a network of point sensors. The anomaly detection method is tested on simulated data. The design and testing process is as follows. The simulation data is generated and the time-stamped binary data from a distributed sensor network is aggregated into time bins. The simulation data is then separated into training and validation data sets. The training data is modeled as a Poisson distribution and the Poisson parameter for each time bin is learned. Next, four Poisson data model numerical scores were designed and tested for their ability to detect anomalies. Each numerical score is compared to a specified threshold value. If the numerical score exceeds the threshold, then an anomaly is flagged and reported. All of the Poisson numerical scores considered were based off of the difference from learned regular activity. The Poisson data model numerical scores were then compared to simple summation numerical scores. The receiver operating characteristic (ROC) curve results indicate that the Poisson data model anomaly detection method has higher performance for detecting single sensor, single occurrence anomalies. The chapter is organized in the following manner. First, the simulation of a distributed point sensor network monitoring a facility is presented. Second, the Poisson data model is presented. Next, the numerical scores are defined. Finally, the chapter is concluded with ROC curve analysis.

8.2. Simulation Data

In an effort to generate a controlled set of data, a simulation of an eight point sensor network shown below in Figure 8-1 was constructed based on pedestrian and vehicular agents' trajectories.

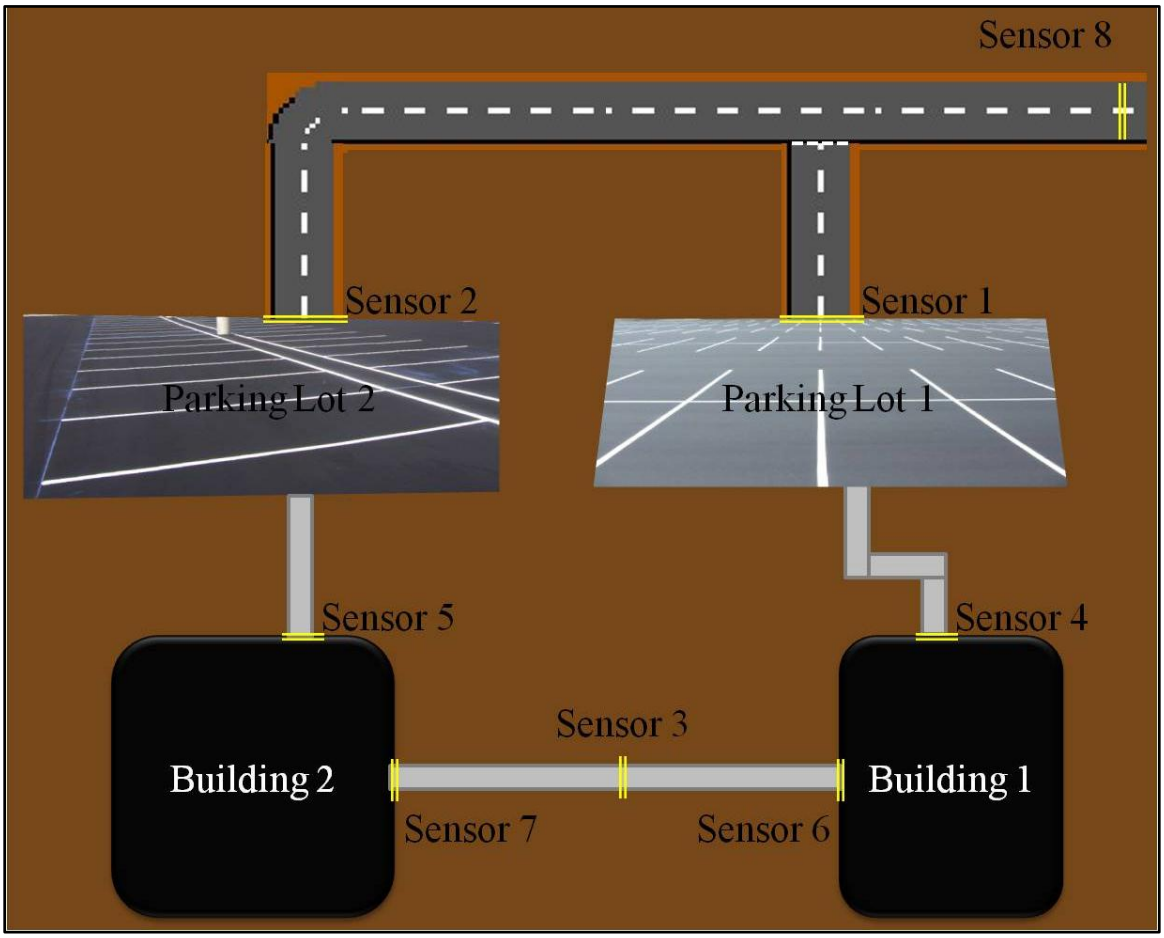


Figure 8-1: Facility simulation sensor placement

The simulation was designed to control and vary the:

1. multiple types of trajectories due to agents moving
2. number of trajectories starting at a specific time, characterized by the time of day
3. origin of trajectories
4. distance between sensors
5. travel speed
6. probability that a trajectory occurs or not (binomial and multinomial distributions)

7. path that an agent chose to traverse (binomial and multinomial distributions)

Points one through three above imply that the simulation is characterized by multiple alarms caused by multiple agents which follow some general trend activity pattern for the time span of one day. Points four and five indicate that the timing between alarms at different sensors, due to the trajectories, are governed by travel parameters such as the distance between sensors and travel speed. Point four above is directly integrating geospatial intelligence (GEOINT) into the simulation. Point five above is indirectly integrating GEOINT into the model because the types of paths traversed have different travel speeds. For example roads, sidewalks, or beaten dirt paths all have different travel speeds due to the purpose of what they are used for. These travel parameters (points four and five), as well as the number of trajectories generated at a specific time of the day (point two above), were treated as Gaussian random variables. Points five and six above indicate that the simulation was further randomized by treating the occurrence of a trajectory and the choice of path taken by agents as binomial or multinomial random variables. In summary, the above simulation approach created a dataset of statistically distributed time-stamped sensor alerts and provided ground-truth data enabling the analysis of meta-tagged events. Exemplar simulation data is shown below in Figure 8-2.

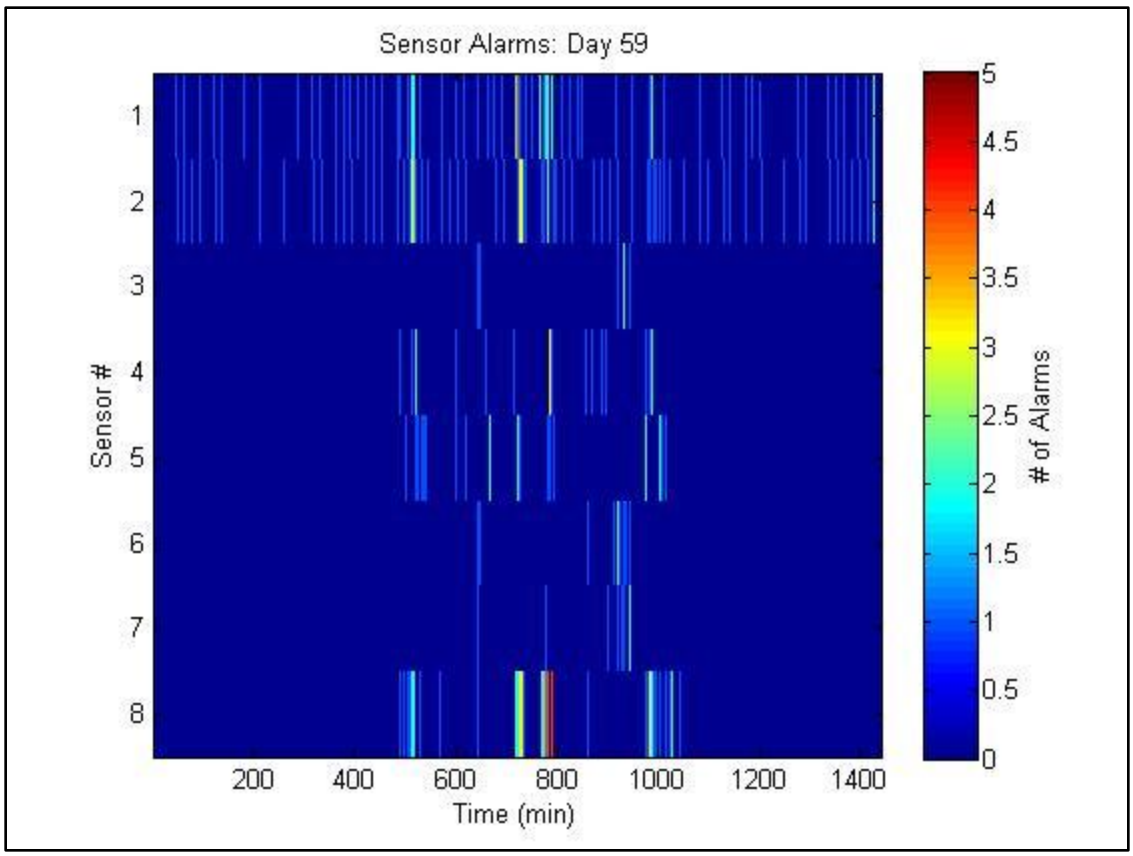


Figure 8-2: Exemplar data for eight sensor simulation

The simulated sensor network data consisted of 260 days of training data, and 260 days of validation data with 5 types of anomalies embedded in the validation data. The five anomaly types were sensor saturation, sensor drop-out, population increase, population decrease, and inter-building movement change. The validation set contained 12 randomly selected and randomly embedded anomalies.

8.3. Poisson Data Model

The Poisson distribution gives the probability, $P(k|\lambda_{ij})$, of k alarms at sensor i, at time j, given λ_{ij} expected alarms. The data for each sensor and for each time bin in minutes was assumed to be a Poisson random variable. The distribution parameter λ_{ij} was calculated by summing and averaging the training data set across days. The Poisson

probability that an alarm occurs for sensor *i*, at time *j*, on day *d* is given by (8.1), while the average Poisson probability over a training period of days *D* is shown in (8.2). This model is equivalent to learning the non-homogeneous Poisson parameter $\lambda_i(j=t)$ for sensor *i*.

$$Pr(k_{ijd}|\lambda_{ij}) = \frac{(\lambda_{ij})^{k_{ijd}} e^{-\lambda_{ij}}}{k_{ijd}!} \tag{8.1}$$

$$Pr(k_{ij}|\lambda_{ij}) = \frac{\sum_{d=1}^{D} \frac{(\lambda_{ij})^{k_{ijd}} e^{-\lambda_{ij}}}{k_{ijd}!}}{D} \tag{8.2}$$

Four numerical scores (*NS*), which vary from 0 to 1, were developed to detect anomalies from the expected values of the Poisson parameter λ_{ij} and comparison to thresholds. Equation (8.3) detects single sensor outages and saturation using the absolute value of the difference between the training mean and the observed Poisson probabilities. After comparing to a threshold value, the remaining alarms are then scanned for consecutive anomalies. NS_2, NS_3, and NS_4 were designed to detect multiple (simultaneous) sensor anomalies (i.e., increases and decreases in employee population and inter-building movement). Equation (8.4) is the absolute value of the ratio of change from the average for each sensor. Equation (8.5) is the average of NS_2 for all sensors. Equation (8.6) is the absolute value of the ratio of change from the average for all sensors.

$$NS_1 = \Delta Pr = \left|Pr(k_{ijd}|\lambda_{ij}) - Pr(k_{ij}|\lambda_{ij})\right| > \text{thresh}_1 \tag{8.3}$$

$$NS_2 = \left|1 - \frac{\sum_{j=1}^{T} Pr(k_{ijd}|\lambda_{ij})}{\sum_{j=1}^{T} Pr(k_{ij}|\lambda_{ij})}\right| > \text{thresh}_2 \tag{8.4}$$

$$NS_3 = \frac{\sum_{i=1}^{N}(NS_2)}{N} > \text{thresh}_3 \tag{8.5}$$

$$NS_4 = \left| 1 - \frac{\sum_{i=1}^{N} \sum_{j=1}^{T} Pr(k_{ijd}|\lambda_{ij})}{\sum_{i=1}^{N} \sum_{j=1}^{T} Pr(k_{ij}|\lambda_{ij})} \right| > \text{thresh}_4 \quad (8.6)$$

The general process for detecting anomalies using the Poisson data model numerical scores is depicted below in Figure 8-3.

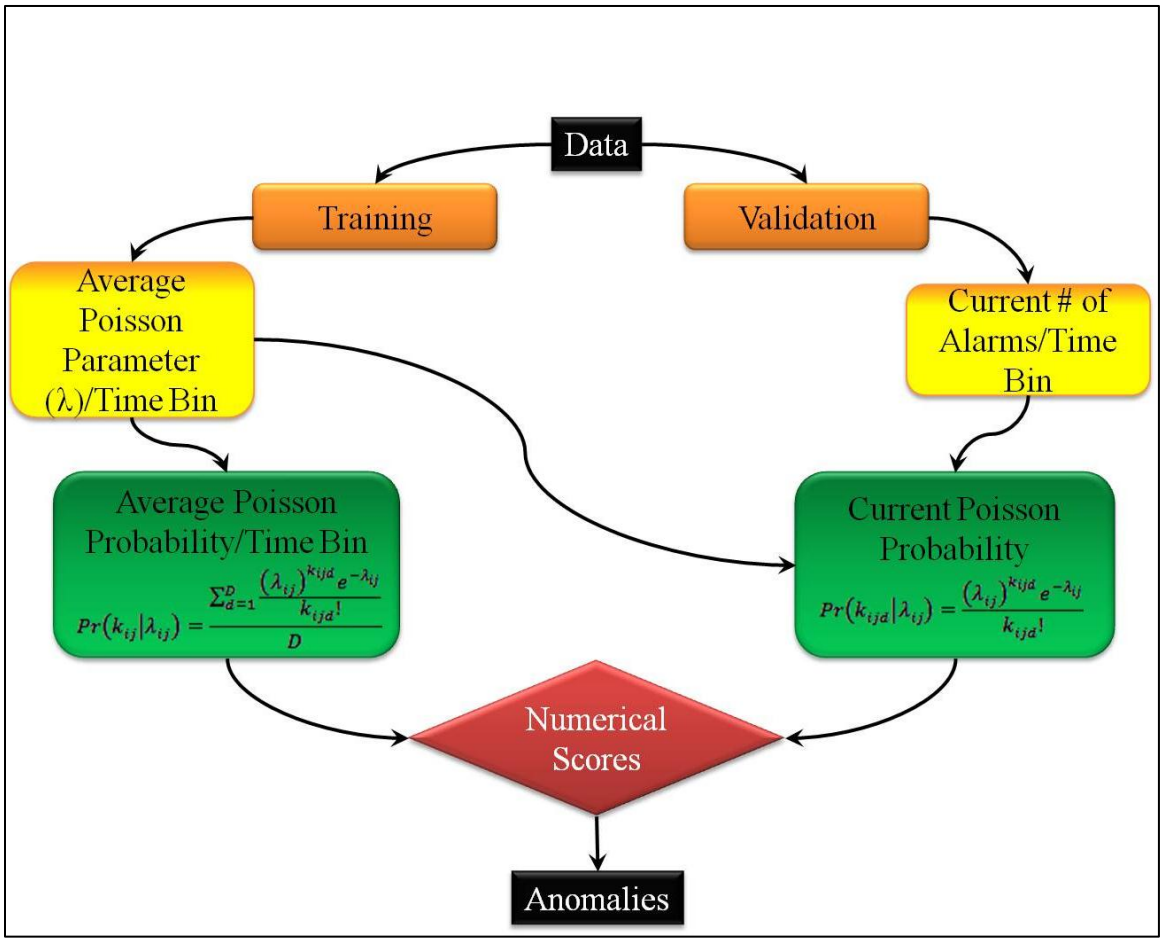


Figure 8-3: Poisson data model anomaly detection process flow chart

Figure 8-3 illustrates that the data is first split into training and validation data. Using the training data, the average Poisson parameter is learned for each time bin. Once the average Poisson parameter is calculated, it can be used to calculate the current Poisson probability of the validation data. The average and current Poisson probabilities are then compared using the numerical scores, which results in the detection of anomalies if they exist.

Benchmarks to compare the Poisson numerical scores to were developed from summations and averages of sensor alarms instead of the average Poisson parameter. The following substitutions described below in (8.7)-(8.8) were made in (8.3)-(8.6) were made to obtain the comparable summation benchmarks.

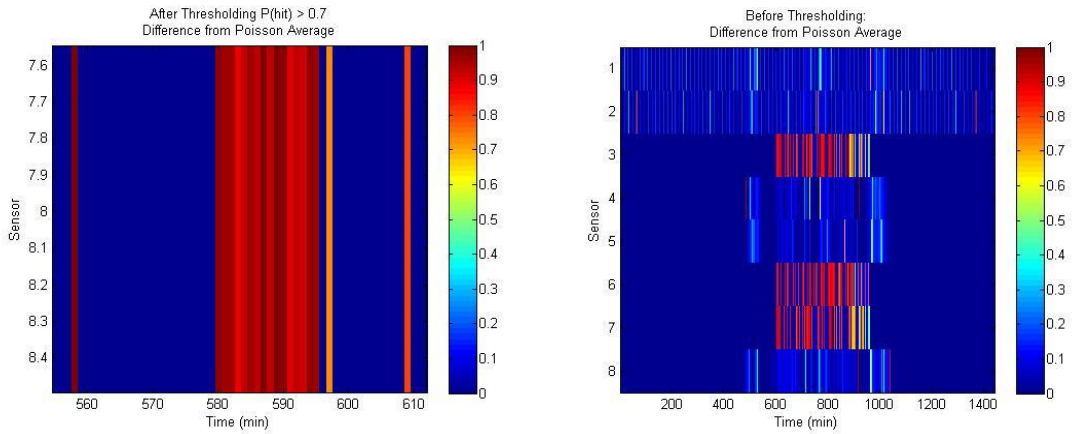
$$Pr(k_{ijd}|\lambda_{ij}) \rightarrow k_{ijd} \tag{8.7}$$

$$Pr(k_{ij}|\lambda_{ij}) \rightarrow k_{ij} \tag{8.8}$$

The comparable simple summation numerical scores to NS_1, NS_2, NS_3, and NS_4 are labeled as S_1, S_2, S_1, and S_4 respectively.

8.4. Simulation Poisson Data Model Anomaly Detection and Receiver Operating Characteristic (ROC) Curve Results

Three of the 12 anomalies were sensor outage occurring at times where no alarms were regularly present, which resulted in a 25% type II error rate. Figure 8-4(a) shows an example of single sensor saturation detection using NS_1, whereas Figure 8-4(b) shows the detection of increased inter-building movement using NS_2, NS_3, and NS_4.



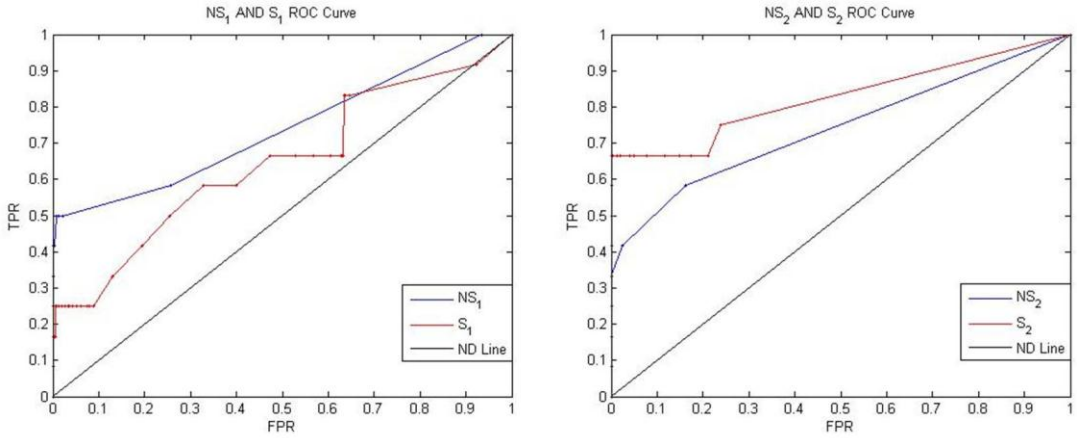
(a) (b)

Figure 8-4: a)NS_1 Single sensor saturation detected, b)NS_2, NS_3, and NS_4 Inter-building movement increase detected

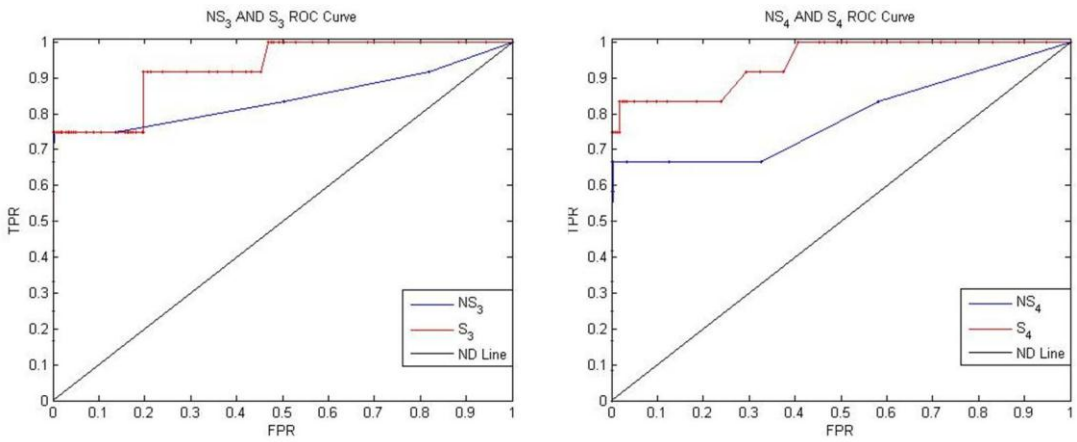
The threshold values in (8.3)-(8.6) were varied to produce ROC curves for the Poisson and summation numerical scores. Each numerical score was analyzed individually. The true positive rate (TPR) was calculated by dividing the number of true-positives by the number of positives (12 anomalies). The false positive rate (FPR) was calculated by dividing the number of false-positives by the number of negatives. The number of negatives was calculated based off of the following rule shown below in (8.9).

$$\#\,Negatives = \begin{cases} (\#\,sensors)(\#\,days) - (12\ anomalies)\ if\ \text{single sensor anomaly} \\ (\#\,days) - (12\ anomalies)\ if\ \text{multiple sensor anomaly} \end{cases} \tag{8.9}$$

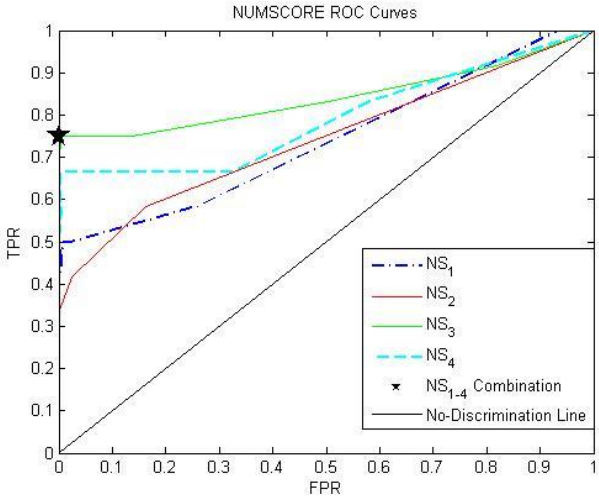
The ROC curve results for the Poisson and simple summation benchmark numerical scores are shown below in Figure 8-5(a)-(e).



(a) (b)



(c) (d)



(e)

Figure 8-5: (a)NS_1 and S_1, (b)NS_2 and S_2, (c)NS_3 and S_3, (d)NS_4 and S_4, and (e) NS_1, NS_2, NS_3, NS_4 ROC curves

The Poisson and summation ROC curves were compared to the no-discrimination line and to each other. Figure 8-5(a)-(e) shows that all methods developed have positive classification characteristics. Figure 8-5(a) shows that the Poisson method is a better classifier of single sensor anomalies, whereas Figure 8-5(b)-(d) shows that simple summations and averages outperform the Poisson method for multiple (simultaneous) sensor anomaly detections. These results are due to the characteristics of the Poisson distribution, which heavily weight large deviations from the expected value λ_{ij}. The Poisson distribution, however, did not heavily weigh slight deviations over multiple

sensors, which is consistent with the results above. Figure 8-5(e) shows an operating point when all Poisson numerical scores are combined, suggesting that a combination of numerical scores can be optimized over regions of an ROC curve.

8.5. Conclusions

The Poisson data model and the numerical scores were shown to have positive anomaly detection and classification characteristics on simulated. The lack of ground truth in real data motivated the use of a trajectory based simulation. The Poison data model anomaly detection method was found to be more effective on single sensor, highly anomalous alarms than simple summation techniques. Given that the real data is more sparse and random than the simulation, it is unclear how much the ROC performance would degrade compared to simple summation and average methods. The results, however, show that NS_1 is best suited to detect anomalies from realistic data of this type. It should also be noted that once a training period has ended, the Poisson data model anomaly detection method does not require any prior knowledge of the anomaly type or form.

One possible extension of this work would be to introduce the concept of relating neighboring time bins (minutes) for a sensor. This approach allows for the variability of an individual's (or agent's) tendency to perform an action at regular time intervals; e.g., variation in the time of arrival to work. Instead of applying the Poisson distribution to a specific time bin across a given dimension, it could be applied to overlapping windows of time.

Chapter 9. A Sensor Reduction Technique Using Bellman Optimal Estimates of Target Agent Dynamics

Brian J. Goode, Philip A. Chin and Michael J. Roan, "A sensor reduction technique using Bellman optimal estimates of target agent dynamics", Proc. SPIE 8050, 805018 (2011);

DOI abstract link:
http://dx.doi.org/10.1117/12.884108

9.1. Abstract

Reducing the number of sensors in a sensor network is of great interest for a variety of surveillance and target tracking scenarios. The time and resources needed to process the data from additional sensors can delay reaction time to immediate threats and consume extra financial resources. There are many methods to reduce the number of sensors by considering hardware capabilities alone. However, by incorporating an estimate of environment and agent dynamics, sensor reduction for a given scenario may be achieved using Bellman optimality principles. We propose a method that determines the capture regions where sensors can be eliminated. A capture region is defined as a section of the surveillance field, where using a causal relationship to the other sensors, an event may be determined using fast marching semi-Lagrangian (FMSL) solution techniques. This method is applied to a crowded hallway scenario with two possible exits, one primary, and one alternate. It is desired to determine if a target deviates from the crowd and moves toward the alternate exit. A proximity sensor grid is placed above the crowd to record the number of people that pass through the hallway. Our result shows

that the Bellman optimal approximation of the capture set for the alternate exit identifies the region of the surveillance field where sensors are needed, allowing the others to be removed.

9.2. Introduction

One topic of interest in distributed sensing is sensor placement to monitor targets in a surveillance field[5, 7, 12]. By strategically placing sensors according to single or multiple objectives [14], the total number of sensors in a distributed network can be reduced for more efficient data processing and decreased number of material resources. There are multiple approaches to solving the geometric coverage[13] sensor placement problem such as combinatorial optimization[4], Voronoi diagrams[12], and genetic algorithms[1]. Others consider a data-driven combinatorial optimization approach to maximize mutual information [11] between sensor locations. However, in this paper, the authors propose a sensor reduction technique using a dynamic trajectory data-driven model to reduce the coverage area of the surveillance field, while still maintaining the ability to determine outcomes in the uncovered area. By approximating the dynamics of agents in the surveillance field, certain activities can be identified without having to cover the entire area.

A classic example of this is the Zermelo navigation problem in optimal control theory[15]. In this problem, a swimmer (agent) is trying to reach the shore in optimal time. However, the current (environment) in the river (surveillance field) provides a constant sweep downstream. Using viability theory, Cardaliaguet, Quincampoix, and Saint-Pierre[3] formed a new version of the problem where there is a waterfall downstream of the river. The problem becomes determining whether or not the swimmer

will be able to make it to shore before reaching the waterfall. In their analysis, they use the dynamics, and not the trajectories of the water current and agent to find where in the river it is possible for the swimmer to reach shore to safety. This forms a barrier in the river where, if located upstream of the barrier, the agent will be able to make it to the shore to safety. However, if located downstream, the swimmer reaches the waterfall first. In terms of surveillance, if one wanted to know using sensors that the swimmer is making progress toward the shore, the sensor coverage would only need to span the river upstream of the barrier. Otherwise, it is known where the swimmer's trajectory will terminate.

Finding these barriers is known as the ``Game of Kind" in the classic optimal control literature[10]. Equivalently, one could use a time optimal value function to find barriers in the state space[9]. If the time optimal value for a state is infinite, then the agent will not be able to traverse to the target. Contrarily, if the value is finite at a state, then it is possible for the agent to reach the target. There are several methods for calculating the value function[2] using Bellman optimal dynamic programming techniques. These infinite horizon solution techniques begin calculating the value at the target and expand out into the state space (surveillance field). The fast marching (FMSL) method[6] uses this to its advantage by ceasing the calculation as soon as there are no more states that will reach the target. The fast marching algorithm is based on the semi-Lagrangian solution method that converges over the entire state space[8] and is used in this work.

This paper proceeds by giving an overview of target capture sets [9]. We show how the formulation can be applied to a surveillance field with environment and agent

dynamics to reduce the coverage area needed in a surveillance field to determine target reachability. An algorithm adapted from Falcone[6] is given to determine the capture sets and eliminate sensors from a sensor grid using the Bellman optimal dynamic programming principle. The formulation and algorithm are then applied to a crowded hallway where it is desired to know if an agent will exit from the main exit following the crowd or divert to an alternate exit as soon as possible. The result of the optimal control calculations are applied to remove sensors in the surveillance field where it is not possible for an agent to exit through the alternate exit due to the motion of the crowd. The joint probabilities of the trajectories are then shown to become more distinct as the sensors are eliminated, leading to an increase in the ability to detect the diverting agent.

9.3. Preliminaries

The problem consists of three distinct entities: (1) a surveillance field, (2) an environment, and (3) agents. The surveillance field is a bounded region where the sensors are to monitor activity. The environment is a set of dynamics that are either known or estimated in the surveillance field. These are assumed to be associated with ``common" movement in the surveillance field. Lastly, agents are the entities under surveillance. In terms of application, these can include people, cars, etc. Agents have a set of dynamics that act as a disturbance on the environment dynamics. Agents that are not detectable will have a minimal disturbance on the environment, and agents that are the target of surveillance by engaging in anomalous behavior will have a larger disturbance impact on the environment dynamics.

The surveillance field consists of a planar state space to which the trajectories of the agents are projected. A state, x, is a member of the compact state space $X \subset \mathbb{R}^2$. The

agents' dynamics are a function of the control, d, and consists of a finite set $D \subset \mathbb{N}$. It is assumed that d is a mapping to a compact set in $\mathbb{R}$. A solution of the system is a trajectory, $\psi = (x, d)$, where $\psi \in \Psi$ and each trajectory in Ψ must satisfy

$$\dot{x}^t = \mathbf{f}(x^t, d^t) = \mathbf{f}_E(x^t) + \mathbf{f}_A(x^t, d^t) \tag{9.1}$$

for all time, $t \in \mathbb{R}^+$, from some initial state, x^{t_0}, where $\mathbf{f}_E(x^t)$ and $\mathbf{f}_A(x^t, d^t)$ are the dynamic contributions of the environment and agent respectively. The form of the evolution equations implies the dynamics of the two entities are exogenous and separable. State evolutions given the disturbance function, $d(\cdot)$, are found using

$$x^{t_f} = S(t_f, t_0; d(\cdot))x^{t_0} \tag{9.2}$$

where S is the state transition operator that finds the value of x^{t_f} from an initial state x^{t_0} given the control function $d(\cdot)$. An agent seeks to generate a trajectory in Ψ such that $x^{t>t^*} \in \mathcal{T}$ for some $t_0 < t^* < \infty$ where $\mathcal{T} \subset X$. $\mathcal{T}$ is termed the target set and is positively invariant. In summary, these equations represent the physical limitations that govern agents' motion through the surveillance field given both the agents' and environmental dynamics.

9.3.1. Capture Set

The kinematic capture set, $\mathcal{C}(\mathcal{T})$, is the set of initial states, x^{t_0}, from which there exists a valid trajectory to the target set, $\mathcal{T}$ with some choice of open-loop control, $d(\cdot)$. Namely,

$$\mathcal{C}(\mathcal{T}) := \left\{x^{t_0} \in X | \exists d(\cdot)\ s.t.\ (x, d) \in \Psi, x^{t>t^*} \in \mathcal{T}, t_0 < t^* < \infty\right\}, \tag{9.3}$$

where x^t is found using the state transition operator of (9.2). This set designates those states from where it is possible for the agent to traverse to the target set given the governing dynamics of (9.2) and the set of agent controls, D.

Calculating the capture set, $\mathcal{C}(\mathcal{T})$ for a given target, $\mathcal{T}$, is not trivial, because all the possible controls, $d(\cdot)$ for the agent must be considered. In order to calculate this set, we use Bellman optimality to determine from which initial states it is possible for an agent to traverse to the target set. This method is advantageous for this purpose because it considers the trajectory over the infinite time horizon. Therefore, we now present an optimal control problem using the Hamilton-Jacobi-Bellman equation (HJB) that allows us to accomplish this task.

9.3.2. Finding the Capture Set

Here, properties of the infinite horizon solution to the HJB equations are used to establish the initial states where the agent is capable of reaching a target set, $\mathcal{T}$. In solving the HJB equation, the result is a value function, $V(x)$ that represents the ``cost to go" of a state in the state space. By choosing an appropriate running cost, $g(x, d)$, then the value function will reveal which states actually reach the target. In the following, the cost to go is strictly finite throughout the state space, and 0 in the target set. Therefore, the value function will have finite value for all states that reach $\mathcal{T}$.

Proposition 1 *Assume an invariant set $\mathcal{T}$, a control law $d(\cdot)$ and the value, $V(x)$, is the solution to the partial differential equation (Dynamic Programming Principle)*

$$0 = \min_{d \in D}[g(x, d) + \nabla V(x)\mathbf{f}(x, d)] \tag{9.4}$$

subject to the boundary condition

$$V(x) = 0, \qquad \forall x \in \mathcal{T} \tag{9.5}$$

and the constraints

$$g(x, d) = \begin{cases} 0, & \forall x \in \mathcal{T} \\ (0, \infty), & otherwise \end{cases} \tag{9.6}$$

The agent is able to guide the state to the set, $\mathcal{T}$, from an initial state $x_0 \in X$ when $V(x) < \infty$.

Proof. The proof follows from the running cost, $g(x, d)$, being finite and accruing at every point on the trajectory that is not in $\mathcal{T}$. The function $g(x, d)$ is defined to have 0 value on $\mathcal{T}$. Since $\mathcal{T}$ is assumed invariant, if the trajectory never reaches, then $V(x)$ will tend toward infinity.

From this result, the definition of the capture set may be restated as,

$$\mathcal{C}(\mathcal{T}) := \{x^{t_0} \in X | V(x) < \infty\} \tag{9.7}$$

where $V(x)$ is the solution to Proposition 1. In this work, a fast marching semi-Lagrangian (FMSL) method is used to find the solution to the optimal control problem in Proposition 1. The FMSL method presented in the subsequent section is based on of the results of [6], which will be summarized with some modest changes to the algorithm where noted.

9.3.3. Algorithm for Calculating the Capture Set

The advantage of the Fast Marching algorithm is that it terminates when the level set of $V(x)$ ceases to propagate, and minimal time is spent calculating the function at states where it is not possible to traverse into $\mathcal{T}$. In this work, we seek to solve (9.4). In order to implement the FMSL method, both the state space and time have to be discretized. The discretized equation becomes

$$V(x_i) = \min_{d \in D}\{V(x_i + hf(x_i, d))\} + hg(x_i, d) \tag{9.8}$$

This is the discrete dynamic programming problem. In choosing a value for $g(x_i, d)$ it is convenient to use time optimality as a running cost.

$$g(x_i, d) = \begin{cases} 0, if\ \ x_i \in \mathcal{T} \\ 1, otherwise \end{cases} \tag{9.9}$$

Substituting in for $g(x_i, d)$, and utilizing the Kružkov transformation $w(x_i) = 1 - e^{-\lambda h}$, the dynamic programming problem becomes

$$w(x_i) = \min_{d \in D}\{e^{-\lambda h} w(x_i + h\mathbf{f}(x_i, d))\} + 1 - e^{-\lambda h} \tag{9.10}$$

where $h = \Delta t$, and $\lambda \in (0,1)$. The original value (time) may be found by reversing the transformation with, $V(x) = \frac{-ln(1-w(x))}{\lambda}$. The Kružkov transformation and the scaling factor, λ, are used to bound $w(x) \in (0,1]$ because $V(x)$ can have an infinite value at some states. By using this change of variables $w(x) = 1$ corresponds to $V(x) = \infty$.

The term $w(x_i + h\mathbf{f}(x_i, d))$, is approximated by interpolating the value of $w(x_i)$ at the neighboring nodes. This is shown in Figure 9-1, where the arrows represent the vectogram initiating from the node labeled ``A". Therefore, in this case, the nodes labeled with the numerals I, II, III will be used in the interpolation. The choice of time step h determines the magnitude of the 1^{st} order approximation, $x_i + h\mathbf{f}(x_i, d)$. For convergence, h should be chosen such that the vectogram of possible trajectories is contained within the neighboring nodes. The following algorithm presents the methods by which the solution to (9.10) is found.

Algorithm 1 *Solution method for (9.10):*

1. Nodes of the set $\mathcal{T}$ are placed in the set, A and $w(x_i) = 0, \forall x_i \in \mathcal{T}$. For all other nodes, $x_i \notin \mathcal{T}$, $w(x_i) = 1$.
2. All neighbors of the nodes in A are placed in N.
3. $w(x_i)$ is calculated for all $x_i \in N$ using (9.10).

4. The node in N with the lowest value, $w(x_i)$, is removed from N and placed in A. Neighbors of this node are then added to N if they are reachable (have dynamics that can reach A).
5. Repeat until N is empty.

Figure 9-1 provides an illustration of the different components of this algorithm. The white circles represent the accepted set, A. The black circles are the set of neighboring nodes, N and the gray circles are the set of nodes that have not been considered.

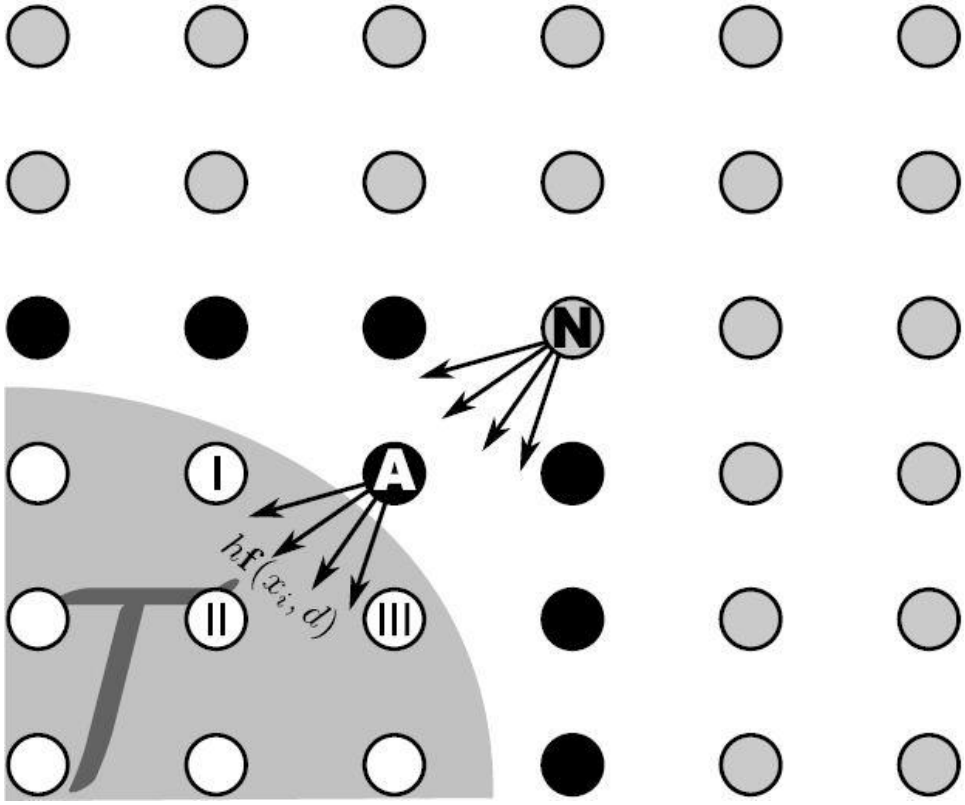


Figure 9-1: Illustration of Alg. 1 on the plane

As an example of a transition, consider the node labeled ``A". It is closest to $\mathcal{T}$, which means that it will have the minimal value. Therefore, it is moved from the set $N$ to $A$. Consequently, the node labeled ``N" will be added to the set N, because it is the only neighbor of the newly accepted node that is not already included. Following the steps in

the algorithm, the process of interpolating the value at each of the nodes in N, repeats until there are no more nodes that are in the neighboring set.

By solving (9.10), it is possible to find the set of states, $\{x_i \in X | V(x_i) < \infty\}$ from which trajectories will traverse into $\mathcal{T}$ given an agent's control set, D. The set may be equivalently written as $\{x_i \in X | w(x_i) < 1\}$. Using this formulation, the capture set can be approximated by

$$\tilde{\mathcal{C}}(\mathcal{T}) = \{x_i \in X | w(x_i) < 1\} \tag{9.11}$$

The quality of the approximation is dependent on the quality of the resolution of the state space.

9.4. Dynamics of Hallway Surveillance

In the previous section, the mathematical formulation given can be used to determine from which states in the surveillance field it is possible for an agent to reach a specified target. In this section, we apply this mathematical formulation to a surveillance scenario where the surveillance field consists of a hallway filled with moving people. The dynamics of the environment model have the crowd move from one end of the hallway to a main exit, E_1. We are interested in detecting if an agent moves toward an alternate exit, E_2. An agent in this scenario is capable of moving in any direction, but is also subject to the environmental dynamics of the crowd that keeps the agents moving in the horizontal direction. We assume that there are proximity sensors that indicate whether or not an agent has entered the square region surrounding the sensor or not. In the following, each of the equations describing the entities are presented.

9.4.1. Surveillance Field

The surveillance field for this work is a planar state space, $X \subset \mathbb{R}^2$. Agent trajectories evolve in this space. To simulate a rectangular hallway, the horizontal state is bounded such that $x_1 \in [x_{1,min}, x_{1,max}]$, and the vertical state is bounded by $x_2 \in [x_{2,min}, x_{2,max}]$. The main exit is defined to be

$$E_1 := \{x \in X | x_1 = x_{1,max}\} \tag{9.12}$$

and the second exit is defined in the state space as

$$E_2 := \{x \in X | x_1 \in [e_{2,min}, e_{2,max}], x_2 = x_{2,min}\}. \tag{9.13}$$

The set of all initial states, where it is assumed that all trajectories start, is the entrance of the hallway such that

$$X_0 = \{x \in X | x_1 = x_{1,min}\}. \tag{9.14}$$

To determine the exact initial condition, $x_0 \in X_0$, of an individual agent trajectory, the initial vertical state, $x_{0,2}$ must be specified. This is determined following

$$x_{0,2} : N\left(0, \sqrt{\frac{1}{2} x_{2,max}}\right), \tag{9.15}$$

and all agents are assumed to enter into the surveillance field at random initial times. All agents in the surveillance field have dynamics subject to (9.1), which has both an environmental and individual agent component. The next sections outline the specific form of this evolution equation in terms of the environment and agent dynamics.

9.4.2. Environmental Dynamics Model

The environmental dynamics approximate the flow of a crowd of people through the hallway (surveillance field). As a simple model, it is assumed that people move horizontally through the hallway with an average speed, $v_p > 0$, which is a constant

parameter. To account for local randomness in the trajectories, the vertical state is subject to a random input. The following model was used to approximate the crowd motion.

$$\mathbf{f}_E(x) = \dot{x}_1\dot{x}_2 = \begin{Bmatrix} v_p \\ A_1\sin(R_1) + A_2\sin(\omega x + 2\pi R_2) \end{Bmatrix} \tag{9.16}$$

where R_1 and R_2: $N(0,1)$. Parameters R_1 and R_2 approximate a Gaussian random noise and phase input, respectively, to the vertical direction movement.

9.4.3. Agent Model

Agents in the hallway have two choices. They can either follow the crowd dynamics and exit from E_1, or attempt to exit the hallway from E_2. To capture this, the agent makes a decision using its control, $d \in D$, where $D = \{0,1\}$. For the agent attempting to exit from E_2, two types of models were used to test the surveillance system. A difficult trajectory to detect is approximated by the agent Type I dynamics:

$$\mathbf{f}_A(x,d) = \dot{x}_1\dot{x}_2 = d\begin{Bmatrix} 0 \\ -c \end{Bmatrix} \tag{9.17}$$

where $c > 0$ and is a constant. The agent Type I dynamics show that the agent attempting to exit from E_2 does so with a constant downward velocity. Agent Type II has the following set of dynamics:

$$\mathbf{f}_A(x,d) = \dot{x}_1\dot{x}_2 = d\begin{Bmatrix} 0 \\ -\gamma v_p e^{\gamma x_1} \end{Bmatrix} \tag{9.18}$$

where γ is some positive constant. The Type II dynamics model the evader as becoming more and more aggressive with attempting to flee as E_1 draws near. In this formulation the control, d, acts to turn the agent dynamics, $\mathbf{f}_A(x,d)$, on or off. Therefore, the agent has the choice to either follow the crowd or attempt to exit. In this work, we assume that this choice is made by the agent prior to entering the hallway, and that d remains constant once it is chosen.

9.5. Sensor Model

The surveillance field is monitored by a rectangular grid of sensors, S, with i sensors on the horizontal direction and j sensors in the vertical direction. s_{ij} denotes the sensor at location (i, j), and there exists a mapping from (i, j) to $\mathbb{R}^2$. The proximity sensor data is generated assuming no intensity or measure of closeness to the center of proximity sensor. If an agent trajectory moves into the proximity region of a sensor, then a hit is recorded at the earliest time this occurs. Each sensor records the total number of hits recorded over K time intervals,

$$H_{s_{ij}} = \sum_{k=1}^{K} h_{s_{ij}}(k) \tag{9.19}$$

where $h_{s_{ij}}(k)$ is the number of hits at sensor s_{ij} in the k^{th} time interval. The time intervals occur at regular time steps, dT. For this work, we take the time step to be $dT = \frac{\Delta x_{1,sensor}}{v_p}$ which corresponds to the time it takes an agent subscribing to the crowd dynamics to move from one column of sensors in the grid to the next.

9.5.1. One-Step Probability of Traversing Two Sensors

The goal is to calculate the probability of a joint path being traveled. A comparison of the joint probability of the trajectories of agents choosing to be a part of the crowd using the environment dynamics to the joint probability of the alternate trajectories will show that the alternate paths are less probable. Trajectories from each proximity sensor s_{ij} can travel to any sensor above, below, and to the right. The set that identifies these neighboring sensors is S'_{ij}. This arrangement is consistent with the one-way direction of travel from the environment dynamics in the hallway. Two different

methods were used to calculate the probability of a transition from a sensor, s_{ij} to its neighboring sensors, S'_{ij}. The first is the classic Conditional Probability method, and the second is a Conditional Probability Ratio method. In both methods, all possible paths are considered by including all of the neighboring sensors in the conditional probability calculation.

Conditional Probability: The conditional probability that $H_{s'_{ij}}$ alarms occur at a neighboring sensor, $s'_{ij} \in S'_{ij}$ given $H_{s_{ij}}$ alarms at sensor s_{ij} is

$$P\left(H_{s'_{ij}} | H_{s_{ij}}\right) = \frac{\#\ \text{of outcomes in } \mathrm{H_{s_{ij}}} \text{and } \mathrm{H_{s'_{ij}}}}{\#\ \text{ofoutcomes in } \mathrm{H_{s_{ij}}}}$$

$$= \frac{\min[\{H_{s_{ij}}, H_{s'_{ij}}\}]}{\sum_{s'_{ij}} \min[\{H_{s_{ij}}, H_{s'_{ij}}\}]}$$

Conditional Probability Ratio Method: This is an alternative to the conditional probability method. First, a ratio between the number of alarms is defined for a sensor and its neighbors using,

$$J = \begin{cases} \frac{H_{s_{ij}}}{H_{s'_{ij}}}, & \text{if } H_{s_{ij}} \leq H_{s'_{ij}} \\ \\ \frac{H_{s'_{ij}}}{H_{s_{ij}}}, & \text{if } H_{s_{ij}} > H_{s'_{ij}} \end{cases} \tag{9.20}$$

Then, the probability that $H_{s'_{ij}}$ alarms occur at a neighboring sensor, $s'_{ij} \in S'_{ij}$ given $H_{s_{ij}}$ alarms at sensor s_{ij} is

$$P\left(H_{s'_{ij}} | H_{s_{ij}}\right) = \frac{J}{\sum_{s'_{ij}} J} \tag{9.21}$$

Both methods shown above are practical to calculate the probability of a sensor transition. We use the ratio method in this work, because the latter ratio method more accurately weights the transition probabilities from the sensor data.

9.5.2. Joint Probability of Traversing Paths

Since we are monitoring trajectories in the surveillance field, the joint probability of traversing multiple sensor proximity regions is needed. To accomplish this, we define a path in the sensor grid as

$$\mathcal{S} := \{s_{ij}^1, s_{ij}^2, \dots, s_{ij}^N\} \tag{9.22}$$

where $\mathcal{S}$ is a specific sequence of N sensors, s_{ij}, that are traversed by an agent trajectory, ψ. We note the position in the sequence using $\mathcal{S}_1 = s_{ij}^1, \mathcal{S}_2 = s_{ij}^2$, and so on. To be a valid sequence the following criterion must hold

$$\mathcal{S}_n \in \mathcal{S}'_{\mathcal{S}_{n-1}}, \qquad \forall n \in \{2, \dots, N\}. \tag{9.23}$$

The rest of this formulation assumes that $\mathcal{S}$ is a valid sequence. Otherwise, the transition probability is 0. We begin by considering the probability of a specific one-step transition when $N = 2$, with the path beginning at an initial sensor, $\mathcal{S}_1$. The joint probability of transitioning between two adjacent sensors is

$$P(H_{\mathcal{S}_1}, H_{\mathcal{S}_2}) = P(H_{\mathcal{S}_2} | H_{\mathcal{S}_1}) P(H_{\mathcal{S}_1}). \tag{9.24}$$

It follows that the joint probability of traversing three sensors ($N = 3$) is

$$P(H_{\mathcal{S}_1}, H_{\mathcal{S}_2}, H_{\mathcal{S}_3}) = P(H_{\mathcal{S}_3} | H_{\mathcal{S}_2}, H_{\mathcal{S}_1}) P(H_{\mathcal{S}_2}, H_{\mathcal{S}_1}). \tag{9.25}$$

After applying the Markov property on the RHS, this simplifies to

$$P(H_{\mathcal{S}_1}, H_{\mathcal{S}_2}, H_{\mathcal{S}_3}) = P(H_{\mathcal{S}_3} | H_{\mathcal{S}_2}) P(H_{\mathcal{S}_2} | H_{\mathcal{S}_1}) P(H_{\mathcal{S}_1}). \tag{9.26}$$

The joint probability of traversing a sequence, $\mathcal{S}$, of N sensors is given by

$$P(\mathcal{S}) = P\left(H_{\mathcal{S}_1}, \ldots, H_{\mathcal{S}_N}\right) = P\left(H_{\mathcal{S}_N}\middle|H_{\mathcal{S}_{N-1}}\right)P\left(H_{\mathcal{S}_{N-1}}\right)\middle|H_{\mathcal{S}_{N-2}}\right) \ldots P\left(H_{\mathcal{S}_2}\middle|H_{\mathcal{S}_1}\right)P\left(H_{\mathcal{S}_1}\right) \quad (9.27)$$

$$= \prod_{n=1}^{N-1} P(H_{\mathcal{S}_{n+1}}|H_{\mathcal{S}_n})P(H_{\mathcal{S}_1}) \quad (9.28)$$

9.6. Sensor Reduction Using Bellman Optimality

Having developed the models for the environment, agents, and sensors, we show how the dynamics of the environment and agents can be integrated to eliminate sensors. To do this we use the optimal control problem outlined in Sec. 2. Specifically, we will use Proposition 1 to estimate the capture set $\tilde{\mathcal{C}}(\mathcal{T})$. **Table 9-1** shows all the parameters used in the subsequent example calculations.

Table 9-1: Model parameters used for sensor reduction calculations.

Surveillance Field	**Environment Dynamics**	**Agent Dynamics**	**Sensor Model**
$x_{1,min} = 0$ $x_{1,max} = 50$ $x_{2,min} = -10$ $x_{2,max} = 10$ $e_{2,min} = 25$ $e_{2,max} = 30$	$v_p = 2$ $A_1 = 0.5$ $A_2 = 0.1571$ $\omega = 0.1571$	$\gamma = 0.122$ $c = 0.8$	$\Delta x_{1,sensor} = 2$ $\Delta x_{2,sensor} = 2$

The parameters in **Table 9-1** are applied to the surveillance and sensor model that is illustrated in Figure 9-2. The sensor grid, S, presently covers the entire surveillance field, X. In the following calculations, we will determine the sensors the set of sensors, $\{S \cap \tilde{\mathcal{C}}(\mathcal{T})\}$, needed to monitor E_2.

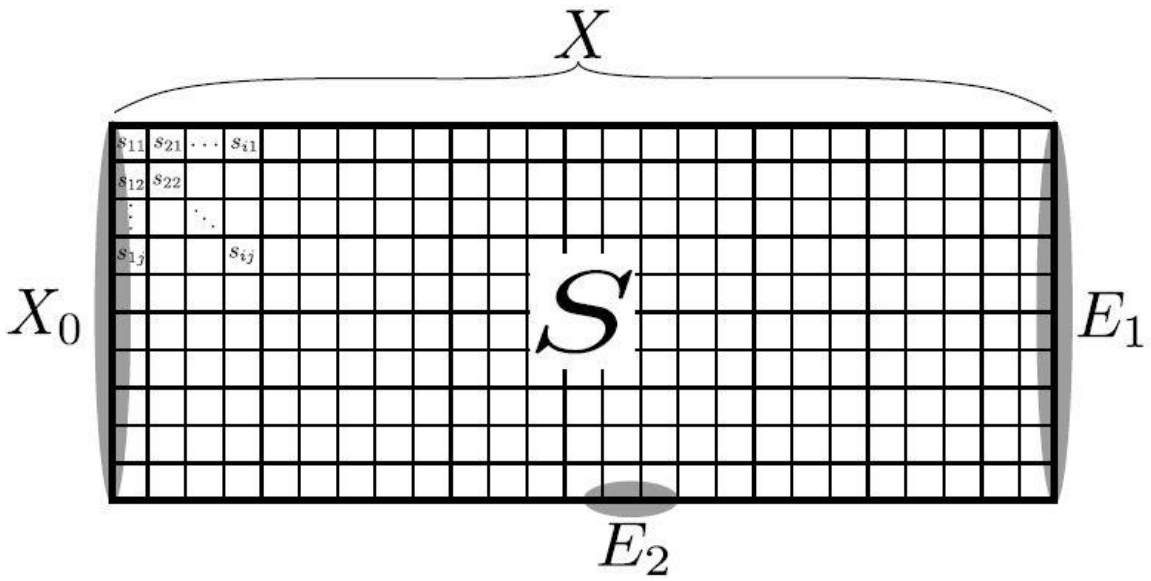


Figure 9-2: This is the layout of the surveillance field. The entrance is X_0, and the two exits are E_1 and E_2. Agents entering the surveillance field will be monitored by the sensor grid, S.

In particular, we want to determine from which states in the state space it is possible for an agent to succeed in exiting through exit, E_2. Therefore, we set $\mathcal{T} = E_2$, and proceed with calculating $V(x)$ in Proposition 1 using Alg. 1. However, in order to use Alg. 1, the state space must be discretized into a grid lattice. We have chosen for this work to discretize the states with a 0.5 node spacing in both directions of the lattice. Lastly, $\lambda = 0.1$ was chosen as the bounding rate for the change of variables from $V(x)$ to $w(x_i)$.

9.6.1. Determining the Capture Set

We first consider the dynamics associated with agent Type I. To approximate the dynamics of the agent, we ignore the noise term of the crowd. So at any point in time, the agent has the option of following the crowd or trying to move down the crowd toward the exit. The capture set, $\tilde{\mathcal{C}}(\mathcal{T})$ for this scenario is shown in Figure 9-3.

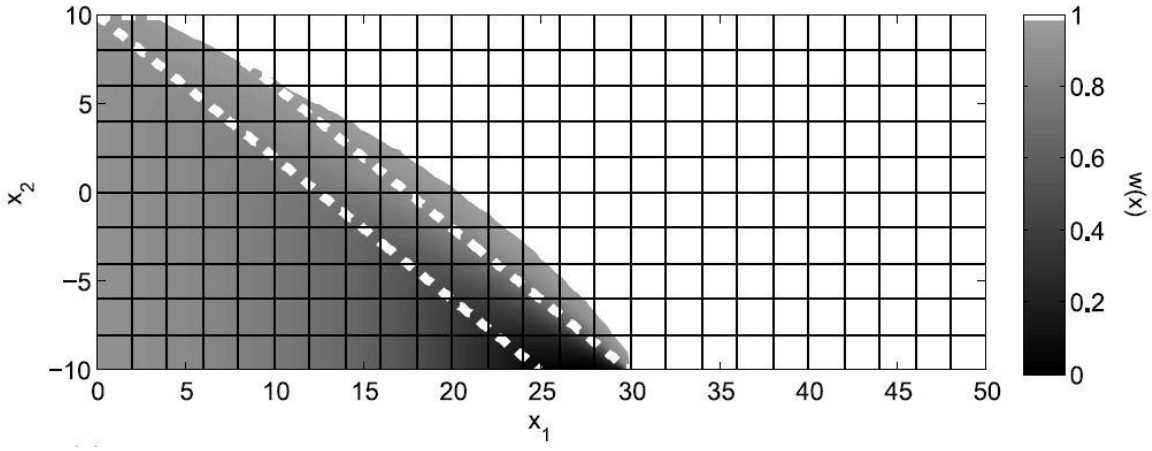


Figure 9-3: $w(x)$ is plotted over the state space, assuming a Type I agent. The trajectories show two of the extreme routes that could be taken by an agent from $x_0 = (0,10)$. The gray region depicts the capture set, $\tilde{\mathcal{C}}(\mathcal{T})$

The gray region in the figure shows all the initial states where it is possible for the agent to reach E_2 in finite time. This is the capture set. The darker regions of the capture set show the states where it takes a smaller amount of time to reach E_2, and the lighter states show longer times to reach the alternate exit. Also shown are two extreme trajectories emanating from the initial point $x_0 = (0,10)$. The trajectory terminating on the left most region of E_2 is the case where the agent enters the surveillance field and immediately heads for the exit. Here, $d = 1$ beginning at $t = t_0$. The other trajectory shows the case where the agent can follow the crowd dynamics until $x_1 = 5$, must try to head for the exit. If the agent were to follow the crowd for any longer, it would be forced to exit from E_1 due to the crowd input of the dynamics. This trajectory can be interpreted as the right barrier of the true capture set, $\mathcal{C}(\mathcal{T})$, where the surveillance field boundaries serve as the other boundaries. As long as the agent does not pass this path, then it is possible to reach the exit by playing the proper control value, d. We see that the estimated capture set, $\tilde{\mathcal{C}}(\mathcal{T})$, for this example follows this trajectory. The error seen by the curvature of the set, is a result of numerical error in the calculation and the discretization of the grid.

Therefore, to adequately monitor the surveillance field for a Type I agent seeking to exit from E_2, the only sensors needed are those whose proximities are covered, even if just partially, by the capture set. In this scenario, over half of the sensors can be eliminated, because it is known that if the agent is not within the capture set, it must exit from E_1.

In Figure 9-4, the same calculations were performed except the Type II agent dynamics are shown. In comparing the two cases, we see that the gray capture set is much larger. This is because the speed of the agent toward exiting the crowd is assumed to increase exponentially with x_1. Therefore, there are more states as x_1 increases from which the agent can reach the target.

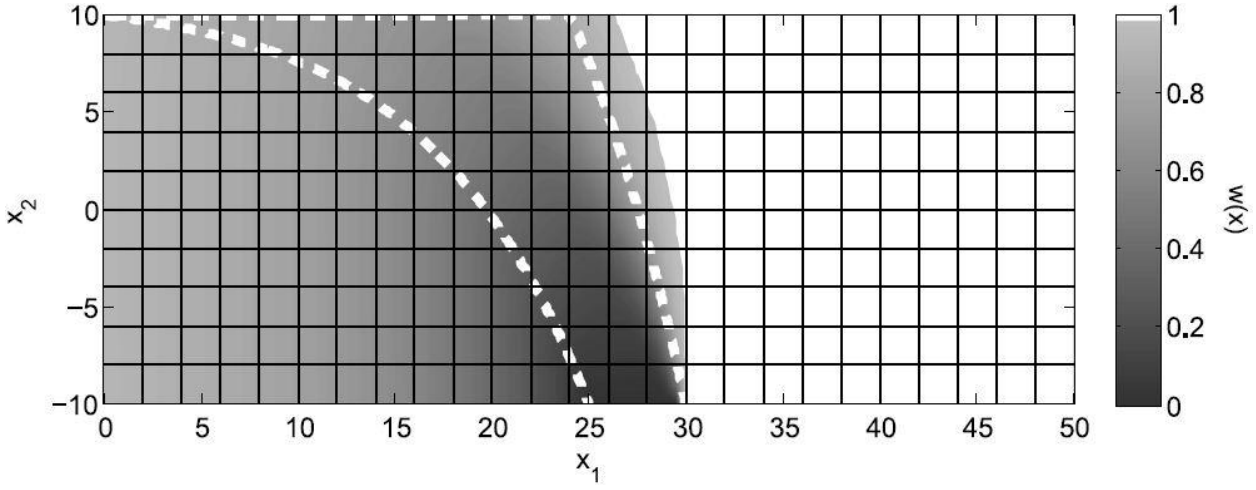


Figure 9-4: $w(x)$ is plotted over the state space, assuming a Type II agent. The trajectories show two of the extreme routes that could be taken by an agent from $x_0 = (0,10)$. The gray region depicts the capture set, $\tilde{\mathcal{C}}(\mathcal{T})$

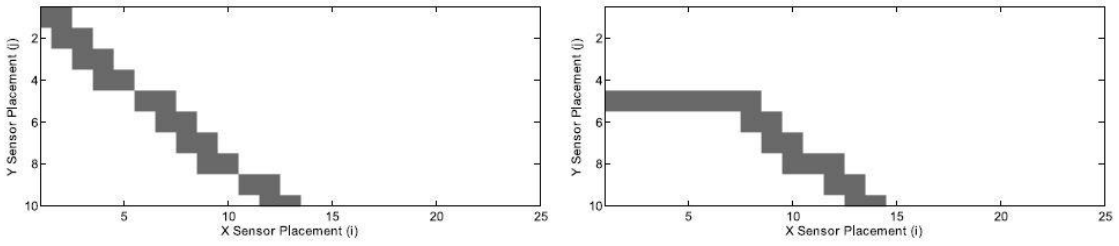
Once again two extreme trajectories are shown that initiate at $x_0 = (0,10)$. The first trajectory which terminates on the left side of the exit, E_2, occurs when $d = 1$ for the entire time the agent enters the surveillance field. Likewise, the other trajectory that terminates on the far side of the exit, occurs when the agent plays $d = 0$ for as long as

possible. This trajectory also acts as a barrier where the agent is guaranteed to reach the target if located on the left side of the path it produces. We also note how the estimated capture set approximates this trajectory reasonably well in that every sensor the trajectory crosses through is also a member of the capture set.

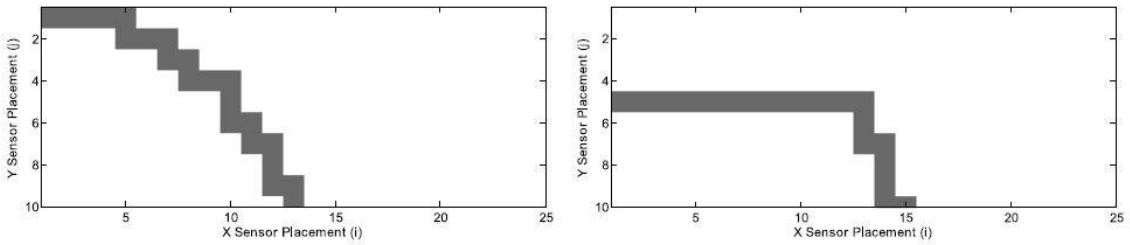
Although these examples are simple in nature, they show that the calculation of the estimated capture set reasonably approximates the barrier trajectory. As more complex dynamics are considered, the barrier trajectory becomes more difficult to determine, and the estimated capture set can give a good approximation of where to focus in on the surveillance field. Furthermore, this calculations do not take a long time to complete. In both of the examples above, the times to calculate $w(x_i)$ were on the order of 16-17s on a 1.7 GHz Pentium M Processor with 1 GB RAM. However, this number is subject to fluctuate with different approximation parameters [9].

9.6.2. Agent Tracking with the Reduced Sensor Field

In this section we apply the sensor model from Sec. 4 to the surveillance field and compare the probability that a sensor sequence, $\mathcal{S}$, occurs with both the full and reduced sensor networks. We consider four agent trajectories (different from the previous section), two from each agent type shown as sensor sequences in Figure 9-5 below.



(a)$\mathcal{S}$, Case 1 (b)$\mathcal{S}$, Case 2



(c)$\mathcal{S}$, Case 3 (d)$\mathcal{S}$, Case 4

Figure 9-5: Four trajectories are studied for the probability of a particular path occurring in the surveillance field

Agent Type I is represented in cases 1 and 2, and agent Type II is represented in cases 3 and 4. To turn the trajectories into a sensor sequence, $x \mapsto \mathcal{S}$, the trajectory was overlaid onto the sensor field. Then the order of the sensors was recorded as the trajectory entered into the sensor's proximity. In general, the steeper the downward trend of the trajectory, the easier it is to detect. Each agent type is given an ``easy'' trajectory (Figure 9-5 (a) and (c)) and a ``harder'' trajectory to detect (Figure 9-5 (b) and (d)). In both cases 1 and 3, the initial condition is $x_0 = (0,10)$ and $d = 1, \forall x_1$. The initial condition for cases 2 and 4 is $x_0 = (0,0)$. In case 2,

$$d = \begin{cases} 0, & x_1 < 25 \\ 1, & otherwise \end{cases}$$

and for case IV,

$$d = \begin{cases} 0, & x_1 < 15 \\ 1, & otherwise \end{cases}$$

to generate the initial straight path before the agent diverts toward the alternate exit. Once the sensor sequence for the diverting agents were determined, the joint probability is calculated for an agent traversing sensor sequence, $\mathcal{S}$ for each of the four cases in the presence of a crowd in the hallway. To simulate the crowd, 500 trajectories were created using the crowd dynamics, by setting $d = 0$ and giving the agents a random initial state

$x_0 \in X_0$. One hundred of the five hundred trajectories are shown in Figure 9-6, where we see that the the normal distribution of the initial states places most of the crowd agents in the center of the hallway. Furthermore, the random motion in the x_2 direction is seen clearly.

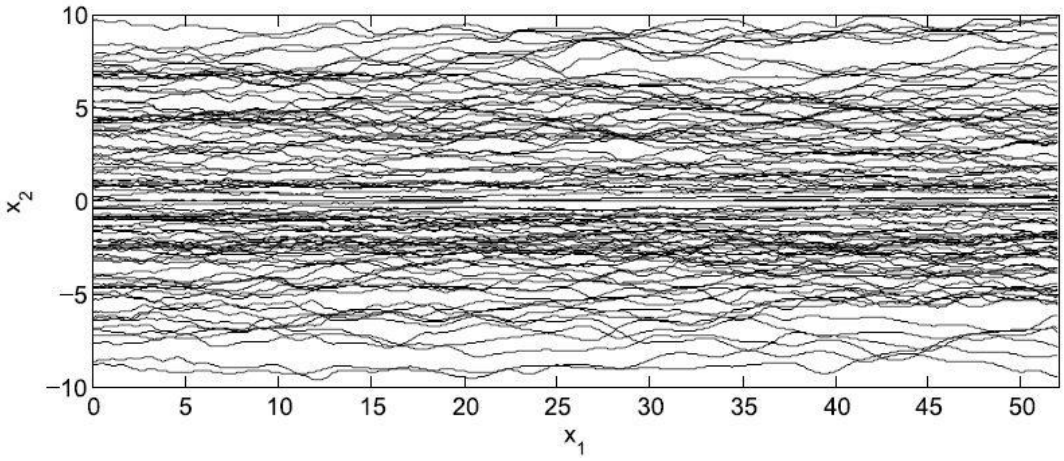


Figure 9-6: This figure shows a sample of 100 trajectories from the total of 500 used to form the crowd hit count

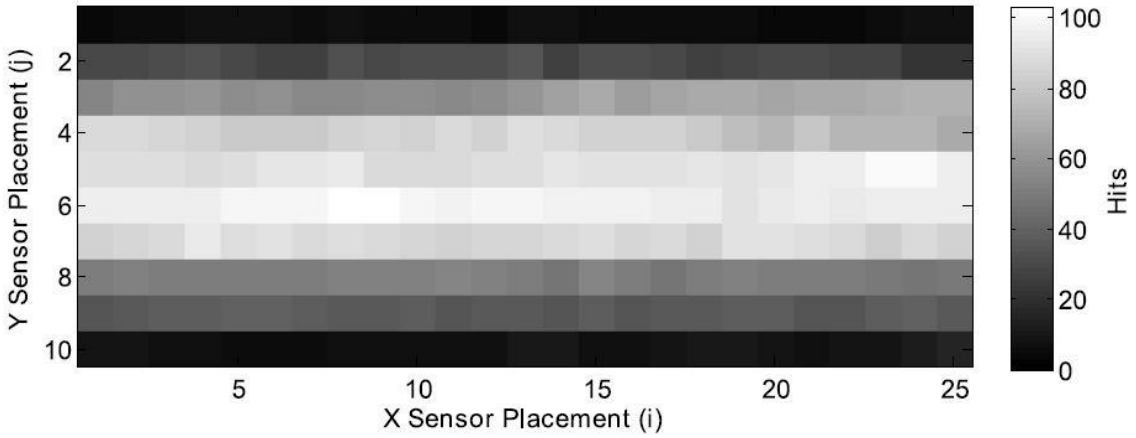
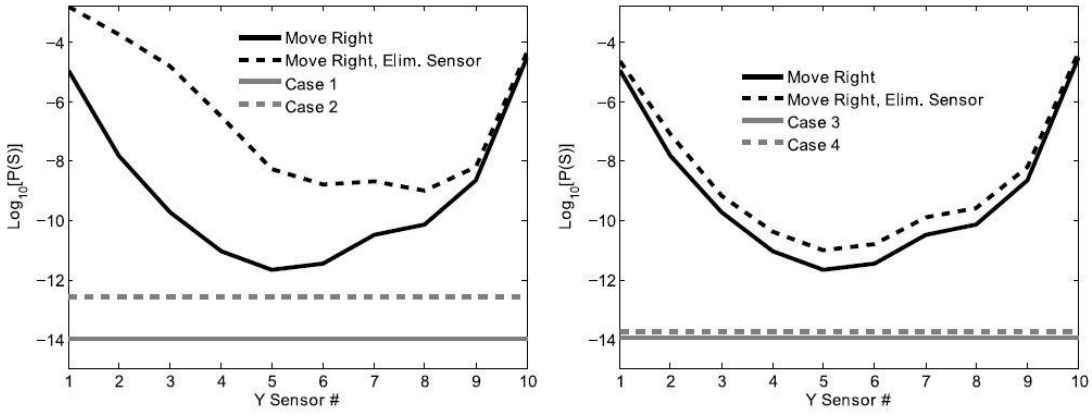


Figure 9-7: This shows the hit count for the random trajectories of 500 agents using the crowd dynamics, $d = 0$

Once the trajectories were recorded, a sensor hit count shown in Figure 9-7 was created by first assigning each of the trajectories a random entrance time. Then, the hits were recorded as the trajectories entered the proximity region of the sensor. Because the

initial condition of the trajectories is a normal distribution, most of the hits appear in the center of the plot. These hit counts are now used to assess the probability of specific sensor sequences occurring.

Equation (9.27) was used to find the joint probability of a sensor sequence occurring with the hit counts given in Figure 9-7. The results of these calculations are shown in Figure 9-8. Here, the probability of each of the case trajectories from Figure 9-5 associated with an agent type are plotted against a case of the probability of an agent moving right following a crowd dynamic from each of the possible initial sensors. The black solid line represents the probability of an agent moving right and exiting from E_1 when the full sensor network is utilized. The black dashed line represents the same trajectory, but calculated using the partial sensor network. The gray lines represent probabilities of sensor sequences for their respective cases. We note that the specific case probabilities only occur at the initial sensor of the trajectory, but have been represented by a line across all initial x_2 sensors for ease of comparison.



(a)Agent Type I (b)Agent Type II

Figure 9-8: Probability of sensor sequences occurring in the surveillance field

In both Figure 9-8 (a) and (b), the probabilities for the move right case are higher than cases 1-4. The convexity of the probability distribution of the move right scenarios is due to the normal distribution of the trajectories. Since more agents tend to be located in the center of the surveillance field, the random walk motion in the vertical direction makes it unlikely for an agent to travel a completely straight line. Anytime the move right scenario is higher than the agent diverging case is beneficial, because the trajectory of an agent heading for E_2 is much smaller than the agents heading for E_1. Therefore, the sensor network would identify this type of behavior as very anomalous. When comparing the ``easy" and ``hard" cases of each type, the dashed line representing the ``hard" cases are always higher than the solid line because these trajectories appear more likely. This is due to the period of time that the agent travels straight before diverting. However, the advantage of the eliminated sensors is seen when comparing the agent types. When comparing the two figures, agent Type I in Figure 9-8 (a) shows a much larger gap between the dashed and solid lines for the move right scenarios than the Type II agent in Figure 9-8 (b). The reason for this increased gap in the Type I agent movement is because more sensors can be eliminated when x_1 is small due to the agent's inability to reach E_2. On the contrary, agent Type II dynamics offer little warning before tending toward the alternate exit. Therefore, there are very few clues early on that can help to determine the ultimate path of the agent.

From these results, scenarios that will benefit most from the sensor reduction techniques are those that can eliminate sensors early along a trajectory. When these types of cases arrive, the dynamics of the agent are such that their intentions are governed more by the environment than cases where the sensors cannot be removed sooner. Therefore,

the environment acts as a causal inferencing tool with which we can make assumptions of the agent's dynamics to estimate the outcome of an agent's trajectory. The result is a sensor network with a reduced number of sensors.

9.7. Conclusions

In this work, we exploit properties of the value function produced by solving the Hamilton-Jacobi-Bellman equation for infinite horizon optimal control problems. With an estimate of the dynamics of the agents and environment of the surveillance field, it is possible to determine where sensors should be placed to determine if an agent will reach a specified target. This information allows for sensor reduction of a sensor network, leading to more efficient detection capabilities. This theory is applied to a hallway detection scenario where it is desired to detect if an agent intends to leave the hallway through an alternate exit. With this scenario we have shown how the capture set calculation approximates the true capture set for different agent dynamics. Probability calculations are given and we show how the reduced sensor network makes the probability of the alternate trajectory more distinct. Future work on this topic will involve more complex environments and anomaly detection with a realistic data set.

9.8. Acknowledgments

Authors 1, 3 supported by ONR Contract No. N00014-06-1-0356. Authors 2, 3 supported by U.S. Army Eng. R&D Center Contract No. W913E5-09-C-0006.

9.9. References

[1] Barrett, S.R. Optimizing Sensor Placement for Intruder Detection with Genetic Algorithms. *Intelligence and Security Informatics, 2007 IEEE*, pages 185 -188, 2007.

[2] Bertsekas, Dimitri P. *Dynamic Programming and Optimal Control*, volume 2. Athena Scientific, 2 edition, 2001.

[3] Cardaliaguet, P. and Quincampoix, M. and Saint-Pierre, P. Optimal times for constrained nonlinear control problems without local controllability. *Applied Mathematics & Optimization*, 36:21-42, 1997.

[4] Chakrabarty, K. and Iyengar, S.S. and Hairong Qi and Eungchun Cho. Grid coverage for surveillance and target location in distributed sensor networks. *Computers, IEEE Transactions on*, 51(12):1448 - 1453, 2002.

[5] Clouqueur, Thomas and Phipatanasuphorn, Veradej and Ramanathan, Parameswaran and Saluja, Kewal K. Sensor deployment strategy for target detection. *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications* in WSNA '02, pages 42--48, New York, NY, USA, 2002. ACM.

[6] Emiliano Cristiani and Maurizio Falcone. A Fast Marching Method for Pursuit-Evasion Games. *Communications to SIMAI Congress*, 1, 2006.

[7] Dhillon, S.S. and Chakrabarty, K. Sensor placement for effective coverage and surveillance in distributed sensor networks. *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, pages 1609 -1614, 2003.

[8] M. Falcone. Numerical Methods for Differential Games Based on Partial Differential Equations. *International Game Theory Review*, 8(2):231-272, 2006.

[9] Brian Goode and Andrew Kurdila and Mike Roan. A Graph Theoretical Approach Toward a Switched Feedback Controller for Pursuit Evasion Scenarios. *Proc. American Control Conference, 2011*. to be published.

[10] Rufus Isaacs. *Differential Games*. John Wiley and Sons, Inc., New York, NY, 1965.

[11] Krause, Andreas and Singh, Ajit and Guestrin, Carlos. Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies. *J. Mach. Learn. Res.*, 9:235-284, 2008.

[12] Jing Li and Ruchuan Wang and Haiping Huang and Lijuan Sun. Voronoi-Based Coverage Optimization for Directional Sensor Networks. *Wireless Sensor Network*, 1:417-424, 2009.

[13] Joseph O'Rourke. *Art Gallery Theorems and Algorithms* of The International Series of Monographs on Computer Science. Oxford University Press, New York, NY, 1987.

[14] Wettergren, Thomas A. and Costa, Russell. Optimal placement of distributed sensors against moving targets. *ACM Trans. Sen. Netw.*, 5:1-25, 2009.

[15] E. Zermelo. Über das Navigationproble bei Ruhender oder Veranderlicher Windverteilung. *Z. Angrew. Math. und. Mech.*, (11), 1931.

Chapter 10. Transforming Ground Surveillance Event Monitoring Into Event Detection Using Geospatial Intelligence

10.1. Abstract

Distributed and unattended sensor networks can be used to monitor activity possibly leading to the detection or prediction of events of interest. Most event detection methods have relied heavily on the time correlation of data with predefined user known events. These methods passively operate on data by matching patterns, giving little thought to the multiple physical phenomenon that generate patterns in the sensed data and the relationship between sensors. Correlation techniques do not necessarily imply causality and detection, but rather the potential existence of a relationship. However, by considering the physical mechanism that generates patterns in data at multiple sensor sites and prior knowledge about how sensors are related, the true detection of events can be increased. The authors propose a contextual joint probability (JP), causal relationship, and sliding window event detection method to improve the detection and classification of trajectories using prior geospatial intelligence (GEOINT) relationships between sensors. This method is applied to a ground surveillance scenario where the goal is to detect and classify the direction of travel from data gathered by an activity monitoring network of point source beam break sensors on a bidirectional path under dynamic temporal traffic conditions. Experimental data, analysis, and results are presented, showing a significant improvement over time correlation methods for detecting and classifying the direction of

travel, thus transforming an event monitoring sensor network into an event detection system.

Keywords - Unattended ground sensors; ground surveillance; point sensors; event detection; event monitoring; classification; direction of travel; joint probability; geospatial intelligence; sliding window; correlation

10.2. Introduction

Distributed unmanned sensor networks can provide key intelligence information by conducting surveillance. The area of surveillance (AOS) that is being monitored for activity governs the type of physical phenomena that causes the sensor measurements. Point sensors, whether they be beam-break, proximity, or passive infrared (PIR), are useful tools for sensing entering, exiting, and roaming that occurs in the AOS. In this paper we refer to point sensing as any type of sensor with non-overlapping sensing areas. One major problem with point sensor networks is that they must be placed to simultaneously provide coverage [1] and detect unique target characteristic events. Even if an event is detected, it must then be classified correctly according to its specific characteristics. Another major problem with ground point sensor networks is that signals can be caused by multiple phenomena, namely localized and non-localized activity caused by trajectories traversing single and/or multiple sensors in multiple directions which mask each other from detection. The main goal in this paper is to take event monitoring point sensor data and separate, extract, detect, and classify events caused by trajectories.

The research area of event monitoring is extensive in literature. Good examples of monitoring networks have included studies on acid rain fall [2] and indoor laboratory

temperature [3]. Ground surveillance of human agents is much more complicated due to the combination of localized and non-localized movement, the intent and motive of humans underlying their movement, and movement due to different types of travel (ex. vehicle or walking) which is depicted in Figure 10-1. Instead of simply monitoring activity, and classifying activity as high or low, the authors' goal is to link or unlink the activity to nearby and surrounding sensors to detect events.

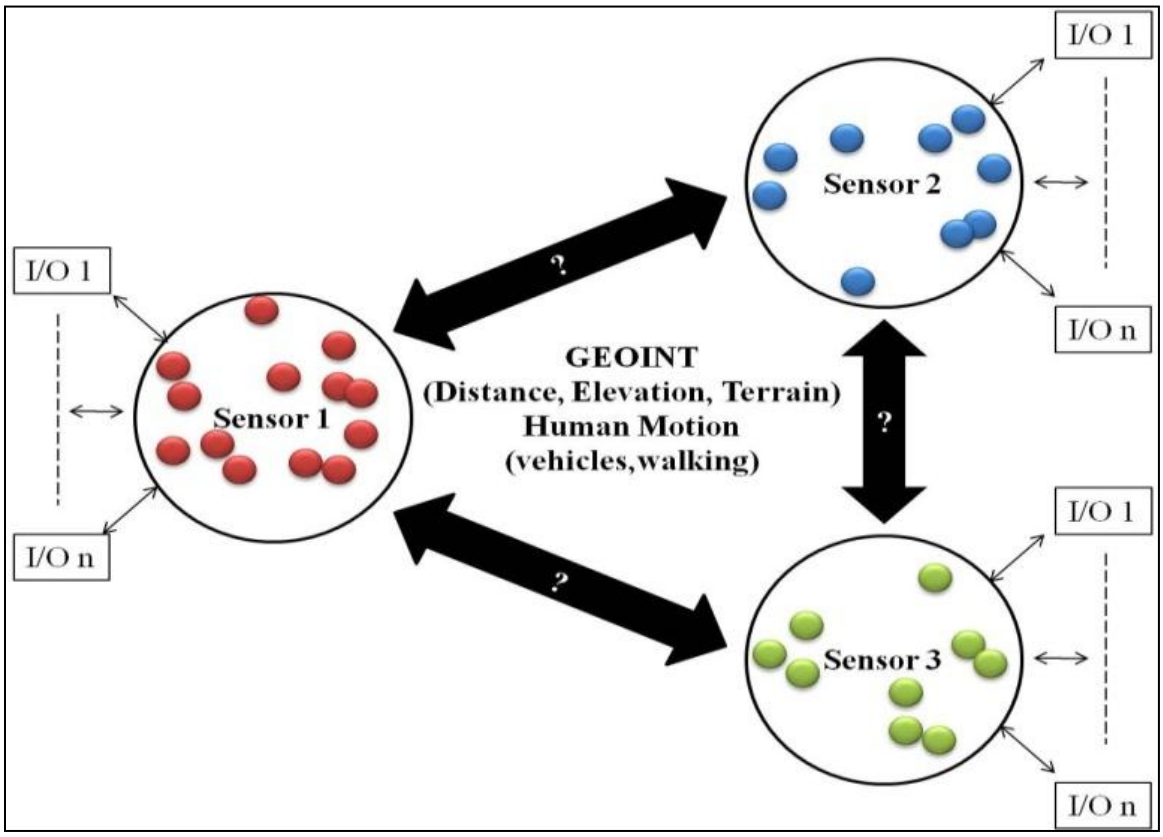


Figure 10-1: Ground sensor network related by GEOINT, travel type parameters, and simultaneous input/output (I/O) sources

Currently, the main approach to event detection is time cross-correlating new or existing data with a predefined pattern [4] and flagging an event for detection if certain likeness measure conditions or thresholds are met. Although cross-correlation (CC) is a

very intuitive approach to finding a pattern, we find that it fails when applied to data characterized by: 1)generation due to multiple physical phenomena, 2)missing data due to occlusion of targets in a crowded AOS, and 3)sensor noise. In this paper, the authors propose a contextual joint probability (JP), causal relationship, and sliding window event detection method to improve the classification of trajectories using prior geospatial intelligence (GEOINT). The method is considered "contextual" because the JP takes all activity of related sensors into account simultaneously. The relationship between sensors is limited to causal/non-causal to establish possible points of origin for trajectories, and the sliding window solves the question of "what data is relevant?" when working with a real-time streaming data scenario. Prior known distances, elevation, terrain, and typical travel methods are used as GEOINT to link the data from multiple sensors. The authors show through results that the combination of GEOINT and the above contextual methods leads to an increase in the correct classification of trajectories over conventional cross-correlation techniques.

The remainder of this paper can be outlined as follows. Section II discusses causal/non-causal and sliding window relationships between multiple sensors. Section III describes the contextual joint probabilistic model of the sensor data. Section IV shows how GEOINT based probability filters are generated and how they dictate the suitable lengths for the sliding window. Section V outlines comparable cross-correlation techniques for event detection and Section VI presents results of analyzing experimental data which show a significant increase in performance of the classification of direction of travel over the cross-correlation methods. Section VII discusses conclusions and how the proposed methods are immediately applicable.

10.3. Causality and Sliding Window Technique

10.3.1. Causal and Non-causal Sensor Relationships

The classification of multiple directions of travel across point sensors requires that we take causality into consideration. By labeling sensors as causal or non-causal, we establish possible points of origin for trajectories and activity. Figure 10-1 illustrates a three point sensor network where each sensor is measuring the level of some activity. The sensors are related to each other due to GEOINT factors such as distance, elevation, and terrain type. The travel of agents between sensors is governed by these GEOINT factors as well as the mode of transportation. Whenever causal activity occurs at any of the sensor locations, each and every related sensor is simultaneously considered as a non-causal sensor and searched for delayed and similar activity. For example, in Figure 10-1, if sensor 1 and sensor 2 have simultaneous activity, then both sensor 1 and 2 will be searched for causal and non-causal activity whereas sensor 3 will only be scanned for non-causal activity. The causality notation for sensor alarms is defined as follows. The set that identifies the causal sensor set is S, and the set that identifies the non-causal sensor set of the i^{th} causal sensor is S_i' such that $s_i \in S$ and $s_{ij}' \in S_i' \ \forall \ j$,where the prime symbol is used to denote non-causality. The number of alarms sensed at causal and non-causal sensors and times respectively are $Hs_i(t_c)$ and $Hs_{ij}'(t_{nc})$, where t_c and t_{nc} are causal and non-causal times described below in (10.1).

10.3.2. Sliding Window Technique

A sliding window method is used because we are only interested in GEOINT linked data and not the entire history of each sensor. In other work, [5,6], a sliding window technique is proposed to limit the amount of data considered, yet they give no reasoning

on how to choose the size of the window. In this paper, we link the size of the sliding window to both causality and GEOINT. The window is broken down into causal and non-causal sections shown in Figure 10-2.

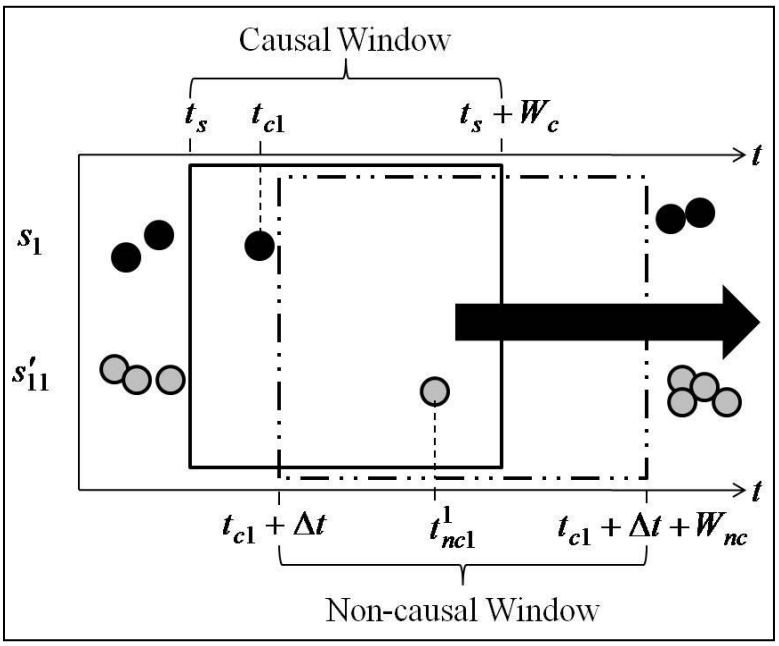


Figure 10-2: Sliding window example

The sliding window time limits for causal and non-causal windows respectively are

$$\begin{aligned} t_s < t_{ck} \le t_s + W_c \\ t_{ck} + \Delta t < t^k_{ncl} \le t_{ck} + \Delta t + W_{nc} \end{aligned} \tag{10.1}$$

where $k=\{1\ldots K\}$ is the index to individual causal time occurrences of t_c, $l=\{1\ldots L\}$ is the index to individual non-causal time occurrences of t^k_{nc} belonging to the k^{th} causal time occurrence (this will be referred to as a "proper" combination of k and l later), and K and L are the total number of causal and non-causal alarms within the sliding window respectively. The adjustable parameters are, t_s, the starting time of the window, Δt, the time bin size, W_c, the length of the causal window, and W_{nc}, the length of the non-causal window. Methods to choose appropriate values for W_c and W_{nc} according to prior

GEOINT are discussed in Section IV. The number of time bins that the causal sliding window moves is Δt_s with lower limit (10.2).

$$\Delta t_s \geq \Delta t \tag{10.2}$$

10.4. Joint Probability Data Model

Joint probability (JP) is typically used to characterize the relationship between multiple related random variables. In this paper, the discrete case of JP is used to further the contextual event detection concept because of its ability to relate amounts of localized activity at one sensor to the total amount of activity at all related sensors. The interpretation and calculation of JP depends on the number of directions of travel we consider simultaneously. We first consider the case where only a single direction of travel is used to make a JP calculation, and then we extend this to consider two directions of travel simultaneously.

10.4.1. Single Direction of Travel

The joint probability mass function (JPMF) for a simple two random variable (RV) case where $x \in X$ and $y \in \bar{Y}'$ has the following property [7].

$$\sum_x \sum_y P(X = x, \bar{Y}' = y) = 1 \tag{10.3}$$

Equation (10.3) states that we obtain a total probability of 1 by marginalizing the JPMF over both RV's *X* and $\bar{Y}'$. To build on this notation, now let *X* be the RV for the i^{th} causal sensor such that $X = X_i$, and let $\bar{Y}_i'$ be a set of non-causal sensor RV's belonging to the i^{th} causal sensor such that $\bar{Y}_i' = \{Y_j\} \ \forall \ j = \{1 \ldots J\}$ where *i* and *j* are the same causal and non-causal sensor indices as in Section II A. Substituting these new definitions into (10.3) we

obtain the marginalizing property for the i^{th} causal sensor and its corresponding J related non-causal sensors.

$$\sum_{j}\sum_{x_i}\sum_{y_j} P(X_i = x_i, Y_j = y_j) = 1 \tag{10.4}$$

Let the number of alarms $Hs_i(t_c)$ and $Hs'_{ij}(t_{nc})$ be discrete random variables with time sample spaces shown in (10.5).

$$\begin{aligned} T_c &= \{\varnothing, t_{ck}\} \ \forall \ k = \{1 \ldots K\} \\ T_{nc} &= \{\varnothing, t^k_{ncl}\} \ \forall \ l = \{1 \ldots L\} \end{aligned} \tag{10.5}$$

where $\varnothing$ is the empty set, K and L are the same as in Section II B. Then the JPMF of the number of alarms that occur at the i^{th} causal sensor and j^{th} non-causal sensor can be expressed as

$$\begin{aligned} &num_{ikjl} = \min(\{Hs_i(t_{ck}), Hs'_{ij}(t^k_{ncl})\}) \\ &P\left(Hs_i(t_c), Hs'_{ij}(t_{nc})\right) = \\ &\begin{cases} \dfrac{num_{ikjl}}{\sum_{k}\sum_{j}\sum_{l} num_{ikjl}} \ \forall \ j \ \ iff \ \text{proper} k, l \\ 0 \ \text{otherwise} \end{cases} \end{aligned} \tag{10.6}$$

which has dimensions of (W_c x W_{nc}). The numerator of (10.6) is the number of outcomes in $Hs_i(t_{ck})$ and $Hs'_{ij}(t^k_{ncl})$ whereas the denominator of (10.6) is the total number of outcomes that the i^{th} causal sensor can produce in its J non-causal sensors. Note that the minimum of the set is taken because the number of outcomes in the intersection of any causal and non-causal activity pair is equal to the minimum amount of activity at either sensing location. For example, five alarms at one sensor location is allowed to cause up to five more alarms at a related sensor location, but not greater than five. We can now formalize the marginalizing property of (10.4) for JP (10.6) as (10.7).

$$\sum_{j}\sum_{t_c}\sum_{t_{nc}} P(Hs_i(t_c), Hs'_{ij}(t_{nc})) = 1 \tag{10.7}$$

Algorithm DT1, shown in Figure 10-3, explicitly presents how (10.6) is calculated and states that each causal alarm occurrence becomes a reference point for finding non-causal sensor alarms. Step six is counting the number of outcomes between a specific pair of causal and non-causal sensor alarms, where as step ten is the total number of outcomes which is used as a normalizing factor for the JP calculation. Figure 10-4 illustrates how the JP is calculated for a single direction of travel.

Algorithm DT1:

Inputs:
- a. Sliding Window of data (limits found in (1))
- b. Sensor Sets $S = \{s_i\}$ and $s'_{ij} \in S' \ \forall \ j$

1 Find times of causal alarms in S (Find t_{ck})
2 For each causal alarm time t_{ck} where k=[1…K]
3 For each non-causal sensor in S' j=[1…J]
4 Find times of non-causal alarms (Find t^k_{ncl})
5 For each non-causal alarm time t^k_{ncl} where l=[1…L]
6 $C_{ikjl} = \min(\{Hs_i(t_{ck}), Hs'_{ij}(t^k_{ncl})\})$
7 End l
8 End j
9 End k
10 $C_i = \sum_k \sum_j \sum_l C_{ikjl}$
11 $P(Hs_i(t_c), Hs'_{ij}(t_{nc})) = \frac{C_{ikjl}}{C_i}$

Figure 10-3: Algorithm DT1, joint probability, one direction of travel

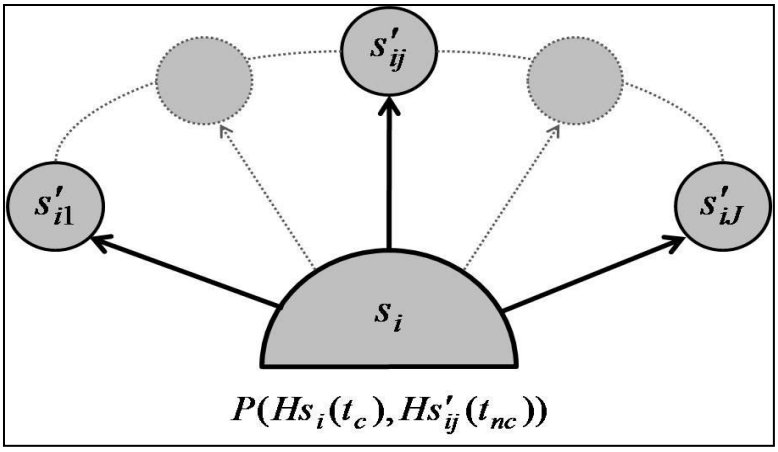


Figure 10-4: Illustration of Algorithm DT1

10.4.2. Two Directions of Travel

To further advance the framework of contextualizing all activity occurring at surrounding sensors, we now consider two directions of travel simultaneously between sensors in the JP calculation. The single direction of travel only used activity moving away from the causal sensor, whereas two directions of travel uses outward and inward activity. The sensor sets for causal, non-causal, outward, and inward travel are summarized in Table 10-1.

Table 10-1: Sensor sets for bidirectional travel

	Outward	**Inward**
Causal	$S = \{s_i\}$	$S = \{s'_{ij}\} \ \forall \ j$
Non-causal	$S'_i = \{s'_{ij}\} \ \forall \ j$	$S'_j = \{s'_{jn}\} \ \forall \ n$

The notation for causal and non-causal sensor sets remains the same as the single direction of travel case, but the sets change. The inward travel causal set is the outwards non-causal set, and the inward non-causal set is $S'_j = \{s'_{jn}\} \ \forall \ n$, where n=$\{1 \ldots N\}$ is the

index to the n^{th} non-causal sensor belonging to the j^{th} inward causal sensor. Algorithm DT2, shown in Figure 10-5 and illustrated in Figure 10-6, explicitly shows the JP calculation for bidirectional travel and incorporates steps from Algorithm DT1.

Algorithm DT2:

Inputs:

a. Sliding Window of data (limits found in (1))
b. Outward sensor Sets $S = \{s_i\}$ and $S' = \{s'_{ij}\}\ \forall\ j$
c. Inward sensor Sets $S = \{s'_{ij}\}\ \forall\ j$ and $S' = \{s'_{jn}\}\ \forall\ n$

2 For outward and inward direction of travel, let [IO]={O or I}
3 If [IO]={O}, assign input b
4 Else (if [IO]={I}), assign input c, End If
5 For each causal sensor in S, m=[1…M]
6 Alg. DT1 (steps 1 through 5)
7 Alg. DT1 step 6: $C_{i\,[IO]\,mkjl} = \min(\{Hs_{i\,[IO]\,m}(t_{ck}), Hs'_{i\,[IO]\,mj}(t^{k}_{ncl})\})$
8 Alg. DT1 (steps 7 through 9)
9 End m
10 End [I/O]
11 $C_{i\,[IO]\,kjl} = \sum_{m} C_{i\,[IO]\,mkjl}$
12 $C_i = \sum_{[IO]} \sum_{m} \sum_{k} \sum_{j} \sum_{l} C_{i\,[IO]\,mkjl}$
13 $P\left(Hs_{i\,[IO]}(t_c), Hs'_{i\,[IO]\,j}(t_{nc})\right) = \frac{C_{i\,[IO]\,kjl}}{C_i}$

Figure 10-5: Algorithm DT2, joint probability, two directions of travel

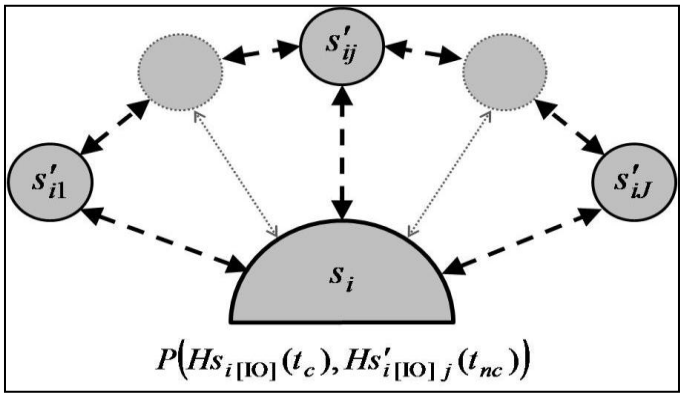


Figure 10-6: Illustration of Algorithm DT2

10.5. GEOINT Based Probability Filters

10.5.1. Travel Time GEOINT Prior Model

Distributed networks of ground sensors measuring ground activity are related by GEOINT parameters such as distance, elevation, and terrain as shown in Figure 10-1. By incorporating this prior intelligence, we can determine how likely activity at separate locations is related to each other by the time that it takes to travel between two points. Much work has been done on travel time estimates between two points [8]. Although travel times can be estimated and modeled through a host of methods including empirical nonparametric smoothed histograms [9], we consider a simple unimodal, symmetric (no skewness), parametric, normal model based on known GEOINT prior (GP) parameters in (10.8).

$$\begin{aligned} T_{trav} &= \mathrm{N}\left(\mu = \frac{d_{ij}}{\theta_s(terr)}, \sigma^2\left(d_{ij}, \theta_s(terr)\right)\right) \\ f(t_{trav}) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t_{trav}-\mu)^2/(2\sigma^2)} \end{aligned} \tag{10.8}$$

The mean can be calculated using d_{ij}, the level distance between the i^{th} and j^{th} sensors and, $\theta_s(terr)$, the travel speed which is a function of the terrain. The variance can be expressed as a function of the same independent parameters as the mean and can be learned directly from observed training data.

A GP probability filter is created by discretizing (10.8) into bins of size Δt as described in (10.1). The lengths of the sliding windows are linked directly to the parameters of (10.8) can be chosen according to the following rules.

$$\begin{aligned} &\Delta t_s \le W_c/2 \\ &W_c \ge \mu \\ &\mu \le W_{nc} \le 3\mu \end{aligned} \tag{10.9}$$

The first rule states that the sliding window should advance in time at an interval less than half of the causal window length. The second rule states that the length of the causal sliding window should be, at minimum, the length of travel time mean and the third rule states that the length of the non-causal sliding window should be centered around twice the travel time mean to equally consider events that occur both before and after the mean travel time. The filter is applied by multiplying times the JPMF in (10.6) or step 13 Figure 10-5. Let G_{ij} be the GEOINT based filter between the i^{th} and j^{th} sensor with dimensions of (1 x W_{nc}). The filter output with dimensions of (W_c x W_{nc}), is then (10.10).

$$P_{filt} = P\left(Hs_i(t_c), Hs'_{ij}(t_{nc})\right) G_{ij}(t_{nc}) \;\; \forall \;\; t_c, t_{nc} \tag{10.10}$$

10.5.2. Classification of Direction of Travel

The authors propose a method to detect direction of travel by a two stage soft and hard classification scheme. First we transform the causal and non-causal JP into units of total elapsed time $t_{elap} = t_c + t_{nc}$. Let A_{ijq} denote the type of event that occurs from travel between the i^{th} and j^{th} sensors, during the q^{th} sliding window t_s. Then the probability that an event occurs is (10.11).

$$P(A_{ijq}(t_{elap})) = \sum_{t_c} \sum_{t_{nc}} P\left(Hs_i(t_c), Hs'_{ij}(t_{nc})\right) \tag{10.11}$$

For every section of sliding window data with limits (10.1), a soft classification is made by taking a maximum of the set of (10.11).

$$\underset{JPGPsoft}{Classifier}(t_{elap}, q) = \max_{i,j}[\{A_{ijq}(t_{elap})\}] \tag{10.12}$$

Once the sliding window has completed all soft classifications for a time ($t_{elap} < t_s$), then a hard classification is made by taking the mode of (10.12) for each t_{elap}.

$$\underset{JPGPhard}{Classifier}(t_{elap}) = \underset{q}{\text{mode}}\left[\left\{\underset{JPGPsoft}{Classifier}(t_{elap}, q)\right\}\right] \tag{10.13}$$

A similar two stage soft and hard classification scheme will also be applied to the cross-correlation techniques described in Section V.

10.6. Time Correlation Techniques

Time correlation is a standard method used to find leading and lagging patterns between signals. The travel time between any two sensors is assumed wide sense stationary (WSS), in the sense that the mean and variance do not vary with respect to time. The Euclidean normalized time cross-correlation (CC) between the i^{th} and j^{th} sensors signals of length T at time $\tau = [-(T-1)...(T-1)]$ is defined as

$$R_{Eij}(\tau) = \frac{\sum_{t=0}^{T-1} s_i(t) s_j(t+\tau)}{\left|s_i(t)\right| \left|s_j(t)\right|} \tag{10.14}$$

where the $|\cdot|$ symbol denotes the Euclidean norm. To contextualize (10.14) at each τ into the possibility of the sensors synchronizing across all other times τ, we normalize again with respect to the summation of (10.14).

$$R_{ij}(\tau) = \frac{R_{Eij}(\tau)}{\sum_{\tau} R_{Eij}(\tau)} \tag{10.15}$$

To use time correlation as means to classify the direction of travel, we first separate the CC into lag and lead components according to the delay.

$$R_{ij\,\mathrm{Lag}}(\tau) = R_{ij}(\tau) \quad \forall \quad \tau \le 0$$
$$R_{ij\,\mathrm{Lead}}(\tau) = R_{ij}(\tau) \quad \forall \quad \tau \ge 0 \tag{10.16}$$

The lag CC component, with negative values of τ, represent the j^{th} sensor lagging behind the i^{th} sensor whereas the lead CC components, with positive values of τ, represent the j^{th} sensor leading the i^{th} sensor. Let B_{ijq} denote the type of event that occurs from travel between the i^{th} and j^{th} sensors, during the q^{th} sliding window t_s. Then the probability that an event occurs is (10.17).

$$P(B_{ijq}(t_{elap})) = \sum_{\tau \le 0} R_{ijq\,\mathrm{Lag}}(\tau)$$
$$P(B_{jiq}(t_{elap})) = \sum_{\tau \ge 0} R_{ijq\,\mathrm{Lead}}(\tau) \tag{10.17}$$

Since it is unlikely that any two sensors match exactly, we propose the summation of the lag and lead components in (10.17), and then comparison to each other to form the CC classifier rule for direction of travel in (10.18).

$$\underset{CCsoft}{Classifier}(t_{elap}, q) = \max_{i,j}[\{P(B_{ijq}), P(B_{jiq})\}] \tag{10.18}$$

Equation (10.18) states that alarms will be classified as the same direction of travel as the maximum of the set of lag and lead CC component summations. The same hard classifying stage as in (10.13) is then applied.

10.7. Experiment and Results

In order to compare the joint probability GEOINT prior (JPGP) based filter and the CC methods, a time synchronized test-bed of three beam break sensors, consisting of two paths in series, was set up along with video for validation purposes. The locations of the sensors in Figure 10-7 were chosen to be a set of stairs which exhibit 1) bidirectional traffic, 2) dynamic temporal traffic conditions (uncrowded to crowded), and 3)

differential GEOINT (elevation change). Three data sets were recorded and their characteristics are shown in Table 10-2.

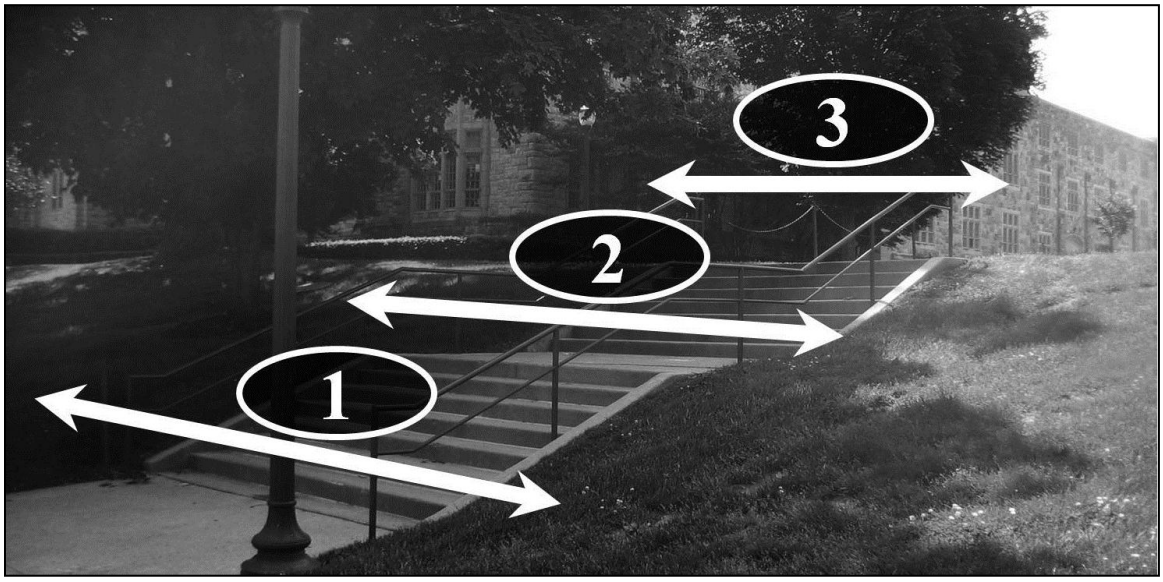


Figure 10-7: Sensor locations

Table 10-2: Characteristics of data sets

	Characteristic			
	Length (min)	***# of people up/down***	***# crowded incidents (≥ 3 people)***	***# Nonzero alarm time bins before/after aggregation***
Set 1	27	30/64	10	323/259
Set 2	19	130/100	48	832/590
Set 3	26	22/66	6	302/242
Set 4	24	31/121	24	491/377

Column two and three indicate that number of people traversing the path for a given time period was dynamic. Column four indicates that the data sets recorded a range of crowdedness, even when put into perspective with the total number of people traversing the paths in column three. Although the time data was gathered at a sampling frequency

of 10KHz, the fifth column shows that aggregation of the data into one second time bins resulted in a 19% to 29% decrease in the number of nonzero time bins. Set 1's video was used as training data to verify and learn the mean and variance parameters (10.8) for the individual paths shown below in Table 10-3.

Table 10-3: Learned GEOINT parameters

	Parameters d=13.75 (ft)		
Path	θ_s ***(ft/s)***	μ ***(s)***	σ ***(s)***
$\mathbf{1 \rightarrow 2}$	3.23	4.25	1.05
$\mathbf{2 \rightarrow 3}$	3.44	4.00	1.09
$\mathbf{3 \rightarrow 2}$	3.75	3.67	1.03
$\mathbf{2 \rightarrow 1}$	3.72	3.70	1.01

The one and two direction of travel JPGP and CC methods were tested on datasets two through four. The parameters used and the results are summarized in Table 10-4.

Table 10-4: Parameters and true/false positive rates

PARAMETERS $W_c = 10, W_{nc} = 8, \Delta t_s = (W_c / 2), \Delta t = 1, T = W_c + W_{nc}$			
	TPR/FPR		
	1 Dir.	***2Dir.***	***CC***
Set 2	0.68/0.32	0.71/0.29	0.55/0.45
Set 3	0.87/0.13	0.88/0.12	0.52/0.48
Set 4	0.84/0.16	0.83/0.17	0.56/0/44

All methods were found to have a 0% type II error rate for detecting activity (failing to report activity when it occurred) for all data sets. Table 10-4 indicates that the joint JPGP method outperforms the CC method for the classification of the direction of travel at a minimum of 13% during overly crowded situations. We find that the two direction method outperforms the one direction method for two datasets, indicating that the JPGP

methods are robust to the number of directions considered. Next, all methods were tested for robustness by varying the sliding window parameters and the results are graphed in receiver operating characteristic (ROC) plots in Figure 10-8 along with the line of no discrimination (ND).

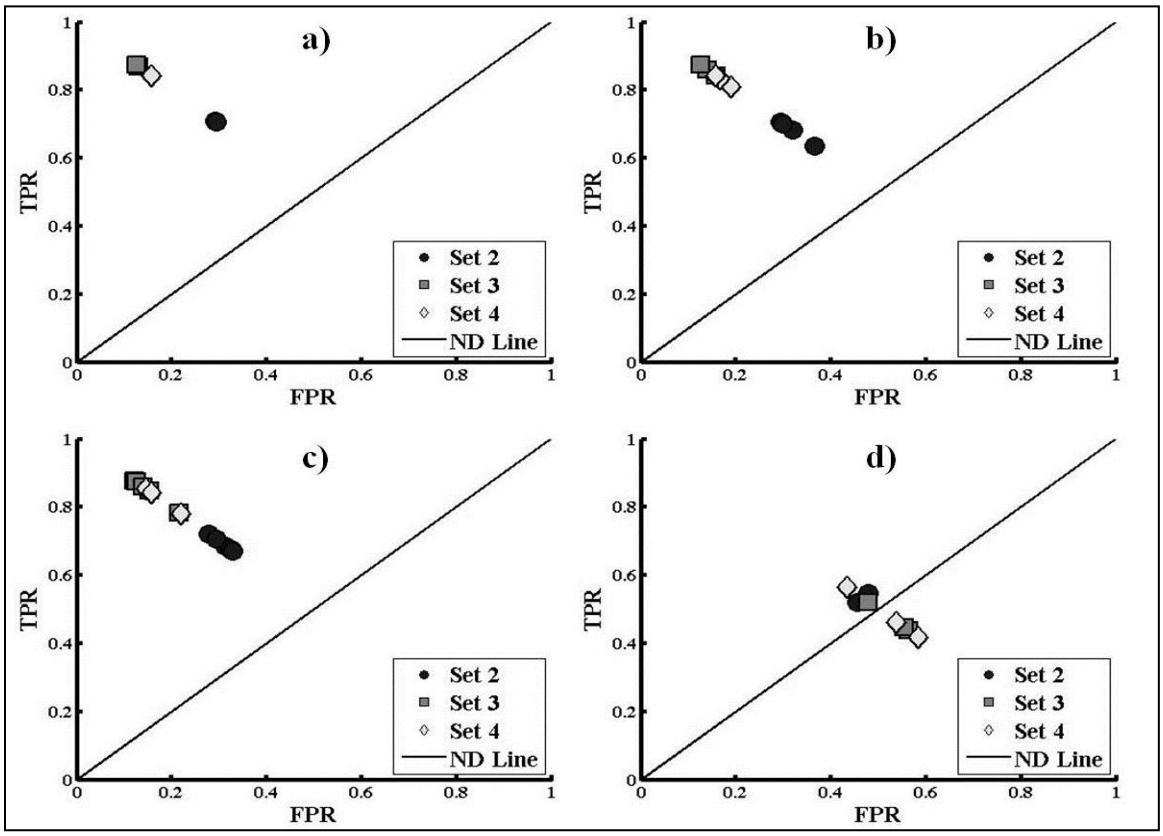


Figure 10-8: ROC plots when using parameters in Table 10-4 and varying a) $1 \le \Delta t_s \le 5$, b) $1 \le W_c \le 20$, c) $4 \le W_{nc} \le 12$, d) $10 \le T \le 26$

Figure 10-8 a) shows that Δt_s does not affect the operating point as long as the rule (10.9) is followed. Figure 10-8 b)-c) indicate that deviation away from central values of W_c and W_{nc} suggested in (10.9) results in a decrease in the effectiveness of JPGP methods. Figure 10-8d) shows that variation of the sliding window length T for the CC method results in negative discrimination, which is undesirable. Figure 10-8 a)-d)

indicate that the decrease in the effectiveness of all methods due to varying individual sliding window parameters is characterized by linear decay towards the ND line.

10.8. Conclusions

The proposed joint probability GEOINT prior and similar CC event detection methods were compared. The empirical results show that the application of simple GEOINT prior filters to JPMF's, which relate local activity to surrounding activity, significantly outperform conventional CC methods for the detection and classification of the direction of travel. The main, but not necessarily valid, objection to the proposed JP GEOINT prior event detection methods is that it could possibly incorrectly detect coincidental unlinked activity at separate sensor locations. The proposed joint probability GEOINT prior methods counter this by 1) contextualizing localized activity into all surrounding activity, 2) relying on redundant sensor placement in the AOS. The results show that the joint probability GEOINT prior methods are feasible and immediately applicable for event detection and classification of the direction of travel in dynamic, multi-path, and crowded AOS. Immediate extension topics of this research include: scalability to wide AOS, heterogeneous types of sensors, travel time estimation, and uncontrollable random sensor placement.

10.9. References

[1] K. Chakrabarty, S. Iyengar, H. Qi, and E. Cho, "Grid coverage for surveillance and target location in distributed sensor networks," IEEE Trans. Computers, vol. 51, no. 12, December 2002.

[2] W.F. Caselton and J.V. Zidek, Statistics in the Environmental and Earth Sciences, Chapter 2, Quality data networks that minimize entropy, Halsted Press, 1992, pp. 10-38.

[3] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: theory, efficient algorithms and empirical studies," Journal of Machine Learning Research, 2008.

[4] M. Li, Y.Liu, and L. Chen, "Nonthreshold-based event detection for 3D environment monitoring in sensor networks," IEEE Trans. Knowledge and Data Engineering, vol. 20, no. 12, December 2008.

[5] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos, "Online outlier detection in sensor data using non-parametric models," Proceedings of the 32nd international conference on Very large data bases (VLDB '06), VLDB Endowment, pp. 187-198.

[6] V. Guralnik and J. Srivastava, "Event detection from time series data," Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '99), ACM, New York, NY, USA, pp. 33-42.

[7] J.L. Devore, Probability and Statistics for Engineering and the Sciences, 4th ed., Brooks/Cole Pub. Co., Pacific Grove, CA, 1995.

[8] D.J. Dailey, "Travel-time estimation using cross-correlation techniques," Transportation Research Part B., vol. 27, April 1993, pp. 97-107.

[9] M.P. Wand, and M.C. Jones, Kernel Smoothing. 1st ed., Chapman & Hall, New York, 1995.

Bibliography

Alpaydin, E. (2004). *Introduction to machine learning.* Cambridge, Mass.: MIT Press.

Barnett, V., & Lewis, T. (1994). *Outliers in Statistical Data.* West Sussex: Wiley.

Bendat, J. S., & Piersol, A. G. (1986). *Random Data.* New York: Wiley.

Ben-Gal, I. (2005). Outlier Detection. In *Data Mining and Knowledge Discovery Handbook* (pp. 131-146). Springer.

Bodik, P., Hong, W., Guestrin, C., Madden, S., Paskin, M., & Thibaux, R. (2004). Intel Lab Data, . *http://db.csail.mit.edu/labdata/labdata.html*.

Bos, A. V. (2007). *Parameter Estimation for Scientists and Engineers.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Brennan, L. (1961). Angular accuracy of a phased array radar. *Antennas and Propagation, IRE Transactions on* , 268-275.

Brooks, E. B. (2007, August 24). Retrieved from Statistics The Poisson Distribution: http://www.umass.edu/wsp/statistics/lessons/poisson/index.html

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 15:1--15:58.

Chen, D. (2008). A novel method for network anomaly detection using superstatistics. *International Conference on Complex, Intelligent and Software Intensive Systems* (pp. 595-598). IEEE.

Chin, P. A., & Roan, M. (2011). Transforming ground surveillance event monitoring into event detection using geospatial intelligence. *Technologies for Homeland Security (HST), 2011 IEEE International Conference on.* Waltham, Massachusetts: IEEE.

Dellaert, F. (2002). *The Expectation Maximization Algorithm, Technical Report number GIT-GVU-02-20.* Atlanta, Georgia: Georgia Institute of Technology.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1-38.

Devore, J. L. (1995). *Probability and Statistics for Engineering and the Sciences.* Pacific Grove: Duxbury Press.

Dinov, I. D. (2008, 12 9). *Statistics Online Computational Resource, UC Los Angeles*. Retrieved from Generalized Expectation Maximization: http://escholarship.org/uc/item/1rb70972

Erling, J., Roan, M., & Gramann, M. (2007). Performance Bounds for Multisource Parameter Estimation Using a Multiarray Network. *Signal Processing, IEEE Transactions on*, 4791-4799.

Friedlander, B. (1984). On the Cramer-Rao bound for time delay and doppler estimation. *IEEE Trans. Inform. Theory*, 575–580.

Gadre, A., Roan, M., & Stilwell, D. (2008). *Sensor error model for a uniform linear array Technical Report No. VaCAS-2008-01*. Blacksburg: Virginia Center for Autonomous Systems (VaCAS).

Goode, B., Chin, P., & Roan, M. (2011). A sensor reduction technique using Bellman optimal estimates of target agent dynamics. *SPIE Conference on Defense, Security, and Sensing* (pp. Proc. SPIE 8050, 805018). Orlando, FL: SPIE.

Guralnik, V., & Srivastava, J. (1999). Event detection from time series data. *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '99)* (pp. 33-42). New York: ACM.

Hawkins, D. M. (1980). *Identification of Outliers.* New York: Chapman and Hall.

Hodge, V., & Austin, J. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 85-126.

Jaffer, A. (1988). Maximum likelihood direction finding of stochastic sources: a separable solution. *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, (pp. 2893-2896).

Kadous, M. W. (2002, 12 10). *Doing the Search*. Retrieved from http://www.cse.unsw.edu.au/~waleed/phd/html/node58.html#fig:kmeans.alg

Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing Estimation Theory*. Upper Saddle River, New Jersey: Prentice-Hall Inc.

Kinsler, L., Frey, A., Coppens, A., & Sanders, J. (1999). *Fundamentals of Acoustics.* New York: Wiley.

Kozick, R., & Sadler, B. (2004). Source localization with distributed sensor arrays and partial spatial coherence. *Signal Processing, IEEE Transactions on*, 601- 616.

Krause, A., Singh, A., & Carlos, G. (2008). Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies. *Journal of Machine Learning Research*, 235-284.

Kundu, D. (1996). Modified MUSIC algorithm for estimating DOA of signals. *Signal Processing, 48*(1), 85-90.

Lee, P. (2004). *Bayesian Statistics an Introduction* (3rd ed.). New York: Oxford University Press Inc.

Li, M., Liu, Y., & Chen, L. (2008). Nonthreshold-based event detection for 3D environment monitoring in sensor networks. *IEEE Trans. Knowledge and Data Engineering*, 1699-1711.

Markou, M., & Singh, S. (2003). Novelty detection: a review—part 1: statistical approaches. *Signal Processing*, 2481 - 2497.

McLachlan, G. J., & Krishnan, T. (2008). *The EM algorithm and extensions.* Hoboken, N.J.: Wiley-Interscience.

Nakashima, H., & al., e. (2010). *Handbook of Ambient Intelligence and Smart Environments.* New York: Springer Publishing.

Nguyen, H., & Van Trees, H. L. (1994). Comparison of performance bounds for DOA estimation. *Proc. 7th SP Workshop Statistical Signal Array Processing*, (pp. 313-316).

Nott, D. J., & Dunsmuir, W. T. (2002). Estimation of nonstationary spatial covariance structure. *Biometrika*, 819–829.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation*.

Panangadan, A. (2004). Detecting anomalous human interactions using laser range-finders. *International Conference on Intelligent Robots and Systems* (pp. 2136-2141). IEEE.

Papoulis, A., & Pillai, S. (2002). *Probability, random variables, and stochastic processes* (6 ed.). Boston: McGraw-Hill.

Patcha, A., & Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, 3448 - 3470.

Rajagopalan, B. (1997). Anomalous ENSO Occurrences: An Alternate View. *Journal of Climate*, 2351-2357.

Redmond, S. J., & Heneghan, C. (2007). A method for initialising the K-means clustering algorithm using kd-trees. *Pattern Recognition Letters*, 965 - 973.

Reilly, J., & Wong, K. (1992). Estimation of the directions of arrival of signals in unknown correlated noise. II. Asymptotic behavior and performance of the MAP approach. *Signal Processing, IEEE Transactions on*, 2018-2028.

Sadler, B. M., & Kozick, R. J. (2006). A survey of time delay estimation performance bounds. *Sensor Array and Multichannel Processing, 2006. Fourth IEEE Workshop on*, 282-288.

Satish, A., & Kashyap, R. (1996). Maximum likelihood estimation and Cramer-Rao bounds for direction of arrival parameters of a large sensor array. *Antennas and Propagation, IEEE Transactions on*, 478-491.

Singla, G., Cook, D., & Edgecombe, M. (2010). Recognizing Independent and Joint Activities Among Multiple Residents in Smart Environments. *J. Ambient Intelligent Human Computing, 1*(57-63).

Song, B.-G., & Ritcey, J. (1996). Angle of arrival estimation of plane waves propagating in random media. *Journal of the Acoustical Society of America,* 1370-1379.

Stoica, P., & Arye, N. (1989). MUSIC, maximum likelihood, and Cramer-Rao bound. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 720-741.

Stoica, P., & Nehorai, A. (1990, October). Performance study of conditional and unconditional direction of arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 38*, 1783-1795.

Stoica, P., & Sharman, K. (1990). Maximum likelihood methods for direction-of-arrival estimation. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 1132-1143.

Stoica, P., Moses, R., Friedlander, B., & Soderstrom, T. (1989). Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 378-392.

Stoica, P., Nehorai, A., & Söderström, T. (1995). Decentralized array processing using the MODE algorithm. *Circuits, Systems, and Signal Processing*, 17-38.

Strang, G. (1988). *Linear Algebra and its Applications* (3rd ed.). New York: Harcourt Brace & Company.

Swingler, D. (1993). Simple approximations to the Cramer-Rao lower bound on direction of arrival for closely spaced sources. *Signal Processing, IEEE Transactions on*, 1668-1672.

Swingler, D. (2002). CRB approximations for a horizontal array observing a narrow-band target with partial coherence. *Oceanic Engineering, IEEE Journal of*, 305-309.

Taylor, H. M., & Karlin, S. (1998). *An Introduction to Stochastic Modeling.* San Diego: Academic Press.

Vaisenberg, R., Mehrotra, S., & Ramanan, D. (2009). Exploiting semantics for scheduling real-time data collection from sensors to maximize event detection.

Multimedia Computing and Networking 2009. Proceedings of SPIE - The International Society for Optical Engineering.

Van Trees, H. L. (2002). *Optimum Array Processing.* New York: John Wiley & Sons Inc.

Wand, M., & Jones, M. (1995). *Kernel Smoothing.* London: Chapman & Hall.

Wax, M., & Kailath, T. (1985). Decentralized processing in sensor arrays. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 1123- 1129.

Weinstein, E. (1981). Decentralization of the Gaussian maximum likelihood estimator and its applications to passive array processing. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 945- 951.

Widmann, M., & Bretherton, C. (1999, May). 50 km resolution daily precipitation for the Pacific Northwest, 1949-94, http://www.jisao.washington.edu/data_sets/widmann/.

Wilson, D. K. (1998). Performance bounds for acoustic direction-of-arrival arrays operating in atmospheric turbulence . *Journal of Acoustical Society of America*, 1306-1319.

Wong, K., Reilly, J., Wu, Q., & Qiao, S. (1992). Estimation of the directions of arrival of signals in unknown correlated noise. I. The MAP approach and its implementation. *Signal Processing, IEEE Transactions on*, 2007-2017.

Wren, C., & al, e. (2006). Similarity-based Analysis for Large Networks of Ultra-Low Resolution Sensors. *Pattern Recognition, 39*, 1918 - 1931.

Xue, W., Luo, Q., & Pung, H. K. (2011). Modeling and detecting events for sensor networks. *Information Fusion, Special Issue on Information Fusion in Future Generation Communication Environments*, 176-186.

Xue, W., Luo, Q., Chen, L., & Liu, Y. (2006). Contour map matching for event detection in sensor networks. *Proceedings of the 2006 ACM SIGMOD international conference on Management of data* (pp. 145--156). ACM.

Zeira, A., & Friedlander, B. (1996). Direction of arrival estimation using parametric signal models. *Signal Processing, IEEE Transactions on*, 339-350.