AN ABSTRACT OF THE THESIS OF

Zeyu You for the degree of Master of Science in Electrical and Computer Engineering presented on June 10, 2014.

Title: A Statistical Inference Framework for Finding Recurring Patterns in Large Data with Applications to Energy Management

Abstract approved: ____

Raviv Raich

We consider the problem of finding unknown patterns that are recurring across multiple sets. For example, finding multiple objects that are present in multiple images or a short DNA code that is repeated across multiple DNA sequences. We first consider a simple problem of finding a single unknown pattern in multiple data sets. For time series data, the problem can also be formulated as a blind joint delay estimation. The non-convex nature of the problem presents a few challenges. Here, we introduce a novel algorithm to estimate the unknown pattern, which is guaranteed to yield an error within a factor of two of that of the optimal solution. Using mixture modeling, we propose a natural extension to the approach that allows the detection of multiple templates placed across multiple sets. Applications to home energy management are considered.

A Statistical Inference Framework for Finding Recurring Patterns in Large Data with Applications to Energy Management

by

Zeyu You

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented June 10, 2014
Commencement June 2014

Master of Science thesis of Zeyu You presented on June 10, 2014.

APPROVED:

Major Professor, representing Electrical and Computer Engineering

Director of the School of Electrical Engineering and Computer Science

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Zeyu You, Author

ACKNOWLEDGEMENTS

I would like to acknowledge my major advisor Prof. Raviv Raich, without his patience and kindness guidance, I would not make so much progress so far. His motivation, enthusiasm and immerse knowledge helped me to work through this thesis.

Besides my advisor, I would like to thank Intel corporation for partially funding this work and extend a special thanks to Tomm Aldridge and James Song of Intel Labs. I would also like to thank the Pecan Street Research Institute and Intel corporation for providing the dataset used in this thesis.

Moreover, I would like to thank the rest of my thesis committee: Prof. Thinh Nguyen, Prof. Xiaoli Fern and Prof. for their support and advice.

I would also like to thank our group members, Evgenia Chunikhina, Anh Pham, Sumanth Avadhani and Xin Li, for their support on my thesis defense and all the fun we have had in the past year.

Last but not the least, I would like to thank my family, especially my mum for always supporting me throughout my life and help me go through all these difficult periods.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

Chapter 1: Introduction

We are interested in the problem of finding recurring patterns in multiple sets. Finding repeated DNA sequences and motif discovery are among the multiple application areas. In this thesis, the problem is formulated as a statistical inference problem, a few statistical models are proposed and analyzed. We develop both a single pattern model and a multiple pattern model for inference and derive a maximum likelihood estimation solution. Due to the non-convex nature of the problem, we introduce novel algorithms to solve the non-convex optimization tasks and provide theoretical guarantees on the solution approach. A key motivating application is the problem of recognizing voltage envelope signatures of home appliances.

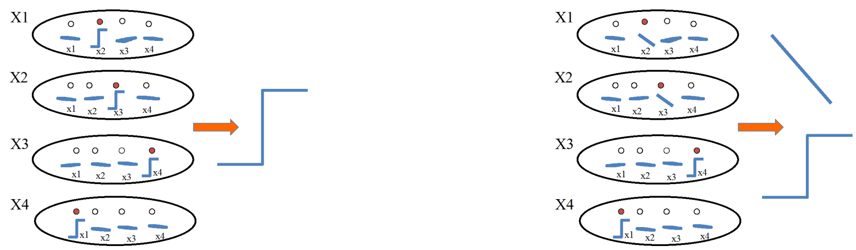
1.1 A key motivating application

The problem of home appliances recognition based on voltage measurement is the key motivating application for this work. Several approaches have been proposed to the problem for disaggregated end-use energy [6]. Most approaches are concentrated on active and reactive power or current signatures. In this work, we focus on voltage envelope transient responses [7]. To the best of our knowledge, the use of activation signatures in appliance recognition has only been lightly explored as opposed to complex power analysis [15], spectral signatures [28], or harmonics of current [9]. Appliance recognition from voltage signatures is challenging because: (i) Voltage envelope changes are subtle. In order to keep the home power grid stable, voltage envelope increase or decrease are under a small percent of the overall voltage envelope value. With small changes in voltage, the quantization error becomes significant. (ii) Voltage envelope transient response may not capture the variations among different appliances (many responses may look like step decrease). However, there are many advantages in using voltage features: (i) Voltage can be monitored at any power outlet while current and power can only be measured at the appliance. (ii) Voltage envelopes are rich in information especially in different work modes and various models.

To explore the possible advantages in the voltage approach, we aim at developing a system which can learn the voltage activation patterns for home appliances. The challenges are: (i) How to learn the voltage signatures? (ii) How to develop a detection or classification system to detect the signatures and recognizing their labels?

1.2 Statement of the problem

Many methods and algorithms have been developed and studied for pattern recognition. While we focus on the problem of estimating multiple different templates from N multi-instance bags containing only one of the multiple templates (see Fig. 1.1(b)), we start by introducing the simpler problem of estimating a single template from N multi-instance bags each containing only one occurrence of the desired template (see Fig. 1.1(a)). In Fig. 1.1(a) and (b), the dot over the template indicates the position of template in the bag.



(a) Single template (b) Multiple templates

Figure 1.1: Recognition of templates in multiple sets.

1.3 Objectives

The aim of this thesis is to provide an inference framework on the problem of recognizing unknown recurring patterns in multiple sets. To investigate the problem of identifying voltage signatures of home appliances, we focus on the following tasks:

1. Develop a statistical model for finding a single pattern recurring in multiple sets.

2. Solve the ML estimation problem of the unknown pattern analytically.

3. Extend the model into recognizing multiple patterns that are recurring in multiple sets.

4. Develop algorithms that solve ML estimation of multiple unknown patterns.

5. Evaluate the performance of the proposed approaches analytically and quantitatively for both cases.

6. Apply the algorithms for recognition of home appliances.

1.4 Related Work

We first consider the problem of finding the same unknown element in multiple sets. This problem may arise in different application areas including but not limited to: pattern matching, sequence alignment in DNA sequencing, and dictionary learning. The problem presents multiple challenges. First, no a-priori information is provided for the element of interest. The search for the element of interest must be performed blindly. This is different than matched filtering in which an element in a set is matched with multiple known templates. The second challenge is computational. When comparing two sets, one can compare every element in the first set to every element in the second set. The complexity associated with comparisons of elements from multiple sets grows exponentially in the number of sets.

Template or pattern matching has been explored in several areas. In [16], a Gibbs Sampling framework for estimating and identifying multiple patterns in the DNA sequences is proposed. In communications and signal processing, matched filtering and correlation analysis have been used in the context of joint delay or angle of arrival estimation. A pre-specified signal structure is a common assumption, e.g., a predefined transmitted signal [27], sinusoidal model with unknown frequencies [23], or a steering vector with unknown angles or delays [26]. In computer science, fast pattern matching [13] for text strings is preformed given a pre-specified template. The formulation in our paper differs from the aforementioned frameworks in that we are interested in an unknown pattern. A closer setup in bioinformatics involves alignment of multiple sequences. While the reference sequence is not defined, scoring different alignments using the COBALT tool [19] enables the process of pattern discovery.

When the object changes to multiple patterns, the main focus may be changed. Finding recurring patterns in data can be applied to various areas, such as finding regulatory sequences in DNA [16], pattern matching in strings [13], and audio motif discovery for bioacoustic applications [18].

Different approaches have been proposed for a pre-specified pattern matching. A Gibbs sampling framework for estimating and identifying multiple patterns in the DNA sequences is proposed in [16], while a graph based WINNOWER algorithm for finding a signal in sampled DNA sequence is proposed in [20]. In computer science, fast pattern matching [13] for text strings has been widely used. Dynamic time warping (DTW) is also a well-known algorithm for a matching problem that allows variations in time [2]. If the pattern of interest is unknown, the problem becomes a blind pattern recognition problem. In [8], a parameter-free CK distance approach with probabilistic early abandoning is proposed for audio motif discovering on large data archives. Finding the most similar pair in long sequence is their focus.

A natural extension to the single pattern matching involves the recognition of multiple recurring patterns. For multiple motif identification and alignment of protein sequences, [1] proposes a combination of search and refinement algorithm. For speaker identification [22], a robust text-independent Gaussian mixture model is proposed.

1.5 Structure of the thesis

The rest of this thesis is organized as follows. Section 2 introduces the blind joint delay estimation model and a maximum likelihood estimation of the model parameters. Section 3 proposes an approach for a single pattern recognition framework. A statistical model is introduced and ML estimator with performance bounds are provided. Section 4 generalizes the single pattern recognition model into K-pattern model. An EM-based algorithm with robust initialization and Majorization-minimization refinement is introduced for solving the model. Section 5 evaluates the performance of proposed algorithm on synthetic and real-world datasets. In the last section, summary of this work, list of publications and future directions are presented.

Chapter 2: A blind joint delay estimation for single pattern

For continuous data, we develop the corresponding bind joint delay model originated from home appliance signatures recognition task.

2.1 Model

In order to identify the activation pattern from voltage envelope measurements, we need to estimate the offset parameter b_i and the delay τ_i for each observed noisy template y_i. Following the Gaussian iid assumption with $n_i(t) \sim \mathcal{N}(0, \sigma^2)$, the negative log-likelihood [25] of the observation can be written as $\frac{1}{2\sigma^2}\sum_{i,t} \|y_i(t) - s(t-\tau_i) - b_i\|^2 + const$. Hence, the optimization associated with ML is equivalent to the following minimization problem:

$$\min_{\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \|y_i(t) - (s(t-\tau_i) + b_i)\|^2, \tag{2.1}$$

where $\theta = [\tau_1, \ldots, \tau_n, b_1, \ldots, b_n, s(1), \ldots, s(T_0)]^T$ is the vector of unknown parameters.

2.2 Solution approach

To perform the minimization, we propose to eliminate the b_i's, then the $s(t)$ and finally the τ_i's. By [25], the resulting ML estimate of the b_i's is given by $\hat{b}_i^{ML} = \bar{y}_i - \overline{s(t-\tau_i)}$. Substituting $\hat{b}_i^{ML}$ for $i = 1, 2 \ldots, n$ into (2.1) yields

$$\min_{\tau,\tilde{s}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\tilde{y}_i(t) - \tilde{s}(t-\tau_i))^2, \tag{2.2}$$

where $\tau = [\tau_1, \ldots, \tau_n]^T$, $\tilde{y}_i(t) = y_i(t) - \bar{y}_i$, $\tilde{s}(t) = s(t) - \overline{s(t)}$ and $\tilde{s} = [\tilde{s}(1), \ldots, \tilde{s}(T)]^T$. Note that $\sum_t \tilde{s}(t) = 0$. Next, we exploit the fact that $s(t) = 0$ for $t \notin \{1, \ldots, T_0\}$. Consequently $\tilde{s}(t) = -\bar{s}$ for $t \notin \{1, \ldots, T_0\}$ and (2.2) can be rewritten as

$$\min_{\tau,\tilde{s}} \sum_{i=1}^{N} \sum_{t=1}^{T_0} (\tilde{y}_i(t+\tau_i) - \tilde{s}(t))^2 + \sum_{i=1}^{N} \sum_{t\in T(\tau_i)} (\tilde{y}_i(t) + \bar{s})^2, \tag{2.3}$$

where $T(\tau_i) = [1, \tau_i] \cup [\tau_i + T_0 + 1, T]$. Next, if we expand $\sum_{t\in T(\tau_i)} (\tilde{y}_i(t) + \bar{s})^2$ as $\sum_{t\in T(\tau_i)} (\tilde{y}_i(t) - \bar{y}_{iT(\tau_i)})^2 + \sqrt{T-T_0}(\bar{y}_{iT(\tau_i)} + \bar{s})^2$ then we can rewrite (2.3) as

$$\begin{aligned}\min_{\tau,\tilde{s}} (\sum_{i=1}^{N} \sum_{t=1}^{T_0} (\tilde{y}_i(t+\tau_i) - \tilde{s}(t))^2 + \sqrt{T-T_0}(\bar{y}_{iT(\tau_i)} + \bar{s})^2) +& \\ \sum_{i=1}^{N} \sum_{t\in T(\tau_i)} (\tilde{y}_i(t) - \bar{y}_{iT(\tau_i)})^2.&\end{aligned} \tag{2.4}$$

2.3 Optimization reformulation

We construct the $(T_0+1) \times (T - T_0 + 1)$ matrix Y_i such that its kth column given by $[y_i(k), \ldots, y_i(k+T_0-1), \sqrt{T-T_0}\bar{y}_{iT(k)}]^T$ and vector $\tilde{\mathbf{s}} = [\tilde{s}(1), \ldots, \tilde{s}(T_0), -\sqrt{T-T_0}\bar{s}]^T$ and rewrite (2.4) as

$$\min_{\tilde{\mathbf{s}},\tau} \sum_{i=1}^{N} \|Y_i e_{\tau_i} - \tilde{\mathbf{s}}\|^2 + \sum_{i=1}^{N} \phi_i(\tau_i), \tag{2.5}$$

where $\phi_i(\tau_i) = \sum_{t\in T(\tau_i)} (\tilde{y}_i(t) - \bar{y}_{iT(\tau_i)})^2$ and e_k is the canonical vector with 1 at the k^{th} place and 0 otherwise. Next, we obtain the ML estimate of $\tilde{\mathbf{s}}$ by differentiating

(2.5) with respect to $\tilde{\mathbf{s}}$ and setting to zero. The resulting ML estimate for $\tilde{\mathbf{s}}$ is given by $\tilde{\mathbf{s}} = \frac{1}{N}\sum_{i=1}^{N} Y_i e_{\tau_i}$. After substituting the ML estimate of $\tilde{\mathbf{s}}$ in (2.5), we obtain a minimization only with respect to τ

$$\min_{\tau} \sum_{i=1}^{N} \|Y_i e_{\tau_i} - \frac{1}{N}\sum_{j=1}^{N} Y_j e_{\tau_j}\|^2 + \sum_{i=1}^{N} \phi_i(\tau_i). \tag{2.6}$$

While the resulting minimization involves only τ, it is still non-trivial. The τ_i's are integers and hence the domain of the problem is non-convex leading to a non-convex optimization problem. Note that (2.6) can also be written as

$$\min_{\tau} \frac{1}{2N} \sum_{i=1}^{N}\sum_{j=1}^{N} \|Y_i e_{\tau_i} - Y_j e_{\tau_j}\|^2 + \sum_{i=1}^{N} \phi_i(\tau_i). \tag{2.7}$$

In the reformulation of (2.7), each term in the summation involves only two delay terms τ_i and τ_j. The equivalence between (2.6) and (2.7) is due to the following result. For vectors $u_1, u_2, \ldots, u_n$ we have $\frac{1}{2N}\sum_{ij}\|u_i - u_j\|^2 = \sum_i \|u_i - \bar{u}\|^2$ where $\bar{u} = \frac{1}{N}\sum_{i=1}^{N} u_i$. This can be proven by expanding both LHS and RHS into the term $\sum_i \|u_i\|^2 - \|\bar{u}\|^2$. The LHS can be expanded as $\frac{1}{2N}\sum_{ij}\|u_i - u_j\|^2 = \frac{1}{2N}\sum_{ij}\|u_i\|^2 + \|u_j\|^2 - 2u_i^T u_j = \frac{1}{2N}(2N\sum_i \|u_i\|^2 - 2(\sum_i u_i)^T(\sum_j u_j) = \frac{1}{2N}(2N\sum_i \|u_i\|^2 - 2N^2\|\bar{u}\|^2 = \sum_i \|u_i\|^2 - N\|\bar{u}\|^2$. Similarly, the RHS can be expanded as $\sum_i \|u_i - \bar{u}\|^2 = \sum_i \|u_i\|^2 - 2\bar{u}^T u_i + \|\bar{u}\|^2 = \sum_i \|u_i\|^2 - \|\bar{u}\|^2$.

Denote the number of delays for each τ_i by $M = T - T_0$. The computational complexity of minimizing (2.7) with respect to the delays τ in a brute-force manner is $\mathcal{O}(M^N)$ [10]. For example, if $N = 30$ and the number of delays is $M = 100$, then $M^N = 10^{60}$. This prompts us to propose an approximate solution with significantly lower computational complexity. The proposed solution guarantees no more than twice of the global minimum achieved by the objective in (2.7).

2.4 Approximate solution for non-convex minimization

We could apply the graph-based approximation described for robust initialization in single pattern recognition section to solve the problem. Since the objective in (2.7) can be viewed as a sum of edge weight in a graph given by $D_{ij} = \|Y_i e_{\tau_i} - Y_j e_{\tau_j}\|^2$ and a sum of node penalties $\phi_i(\tau_i)$. Since the sum runs over all pairs of (i, j), the graph is a complete graph. We propose to replace the single complete graph by N bipartite graphs [3] (see Fig. 3.2). The ith bipartite graph contains only $N-1$ edges placed between the ith node and all other nodes.

To obtain the approximate ML solution $\hat{\tau}_{AML}$, we begin by solving N minimizations. The ith minimization is given by

$$\tau^i = \arg\min_{\tau} f_i(\tau), \text{ where} \tag{2.8}$$

$$f_i(\tau) = \sum_{j=1 \neq i}^{N} \big(\|Y_i e_{\tau_i} - Y_j e_{\tau_j}\|^2 + \phi_j(\tau_j)\big). \tag{2.9}$$

Then, $\tau^{AML} = \tau^{i^*}$, where

$$i^* = \arg\min_{i} f_i(\tau^i). \tag{2.10}$$

Although the objectives $f_i(\tau)$ differ from our original objective in (2.7) they are tightly connected.

2.5 Theoretical guarantees

For both estimators, we establish a lower and upper bounds:

$$\frac{1}{2N}\sum_i f_i(\tau^i) \leq f(\tau^*) \leq \min_i f_i(\tau^i).$$

The bound holds for both $\tau^* = \tau^{ML}$ and $\tau^* = \tau^{AML}$.

Starting with τ^{ML}. For the lower bound, it is easy to see that $f(\tau) = \frac{1}{2N}\sum_i f_i(\tau)$ and hence $f(\tau^{ML}) = \min_\tau f(\tau) = \min_\tau \frac{1}{2N}\sum_i f_i(\tau) \geq \frac{1}{2N}\sum_i \min_\tau f_i(\tau) = \frac{1}{2N}\sum_i f_i(\tau^i)$. For the upper bound, we have $f(\tau) = \min_s \sum_i \|Y_i e_{\tau_i} - s\|^2 \leq \sum_i \|Y_i e_{\tau_i} - Y_j e_{\tau_j}\|^2 = f_j(\tau)$ for all j and hence $f(\tau^{ML}) = \min f(\tau) \leq \min_\tau f_i(\tau) = f_i(\tau_i)$. Next, we show the same bound for τ^{AML}. For the lower bound, we have $f(\tau^{ML}) \leq f(\tau^{AML})$ and hence $\frac{1}{2N}\sum_i f_i(\tau^i) \leq f(\tau^{AML})$. For the upper bound we have $f(\tau) \leq f_j(\tau)$ hold for any τ and j. Hence setting $j = i^*$ and $\tau = \tau^{AML} = \tau^{i^*}$, yields $f(\tau^{AML}) \leq f_{i^*}(\tau^{i^*}) = \min_i f_i(\tau_i)$.

Since $N \min_i f_i(\tau^i) \leq \sum_i f_i(\tau^i)$, we can further bound the lower bound by $\frac{1}{2}\min\limits_i f_i(\tau^i)$. Therefore,

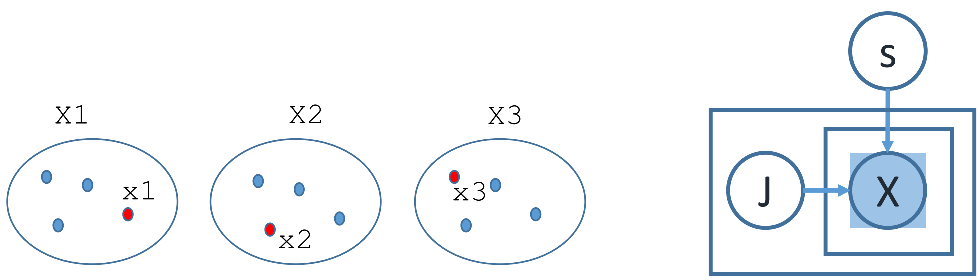
$$\frac{1}{2}\min_i f_i(\tau^i) \leq f(\tau^{ML}) \leq f(\tau^{AML}) \leq \min_i f_i(\tau^i).$$

This sandwich inequality guarantees $f(\tau^{ML}) \leq f(\tau^{AML}) \leq 2f(\tau^{ML})$. This bound suggests that the proposed approach yields a solution objective within a factor of 2 from the optimal solution objective.

The main advantage of the proposed algorithms is the relatively low computational complexity. The minimization in (4.8) can be implemented as follows. For each of the M values of τ_i, $N-1$ separate minimizations over M values of τ_j can be performed yielding a computational complexity of the order $\mathcal{O}(M^2N)$. Since this minimization is applied for every i, the overall computational complexity is $\mathcal{O}((MN)^2)$. T his is the computational complexity obtained by comparing every one of M delay windows in every one of N observed sequences with every one of the M delay windows in all other $N-1$ observed sequences.

Chapter 3: Single Pattern Recognition

Consider the problem of finding the same unknown pattern across multiple sets. To formulate this problem, consider N subsets $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_N$ of the d-dimensional Euclidean space $\mathbb{R}^d$, i.e., $\mathcal{X}_i \subseteq \mathbb{R}^d$ for $i = 1, 2, \ldots, N$. Each set is assumed to contain only one instance of the unknown pattern of interest (see Fig. 3.1(a)) among other patterns. Our goal is to obtain the pattern of interest. In general, no distinguishing characteristics are provided for the unknown pattern and hence it cannot be found when only one set is available. The fact that the pattern of interest is repeated in each set is key to its estimation. We proceed with a detailed probabilistic model for the problem.



(a) (b)

Figure 3.1: (a) Our setting: each set $\mathcal{X}_i$ is assumed to contain one instance of a desired element s. Our goal is to identify the desired element s along with the most similar element in each set, i.e., $x_i \in \mathcal{X}_i$. (b) A graphical model for the alignment problem

3.1 Model

To model the problem of finding the same unknown element in multiple sets in a noisy setting, we start with a generative model for the collection of sets. We begin by generating N sets, each containing one instance of the pattern of interest in an independent fashion. For the ith set, we assume the following generative process. Sample the ith set position RV J_i uniformly in $\{1, 2, \ldots, n_i\}$. Then, generate the n_i elements in $\mathcal{X}_i$ according to

$$\mathbf{x}_{ij} = \begin{cases} s + \nu_{ij} & j = j_i \\ \nu_{ij} & j \neq j_i \end{cases} \tag{3.1}$$

for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n_i$ where s is a deterministic unknown signal, the noise terms ν_{ij}s are iid $\mathcal{N}(0, \sigma^2 I)$.

We determine the joint distribution of $\mathcal{X}_1, \ldots, \mathcal{X}_N$ based on the aforementioned generative process. For each i we organize the elements of $\mathcal{X}_i$ in a $d \times n_i$ matrix $X_i = [\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in_i}]$ and consider joint distribution of the observations represented by the observation matrix $X = [X_1, \ldots, X_N]$ given the unknown vector s. Since we assume that sets are generated in an independent fashion, we express the joint distribution of sets as a product of their marginal PDFs:

$$f(X|s) \quad = \quad \prod_{i=1}^{N} f(X_i|s). \tag{3.2}$$

Since the position of the vector s, J_i, is a latent random variable uniform over the set of positions $\{1, 2, \ldots, n_i\}$, we use the following marginalization of J_i to obtain $f(X_i|s) = \sum_{j=1}^{n_i} f(X_i|J_i = j, s)P(J_i = j)$, where $f(X_i|J_i = j, s)$ denotes the PDF of X_i with s

positioned in the jth element of X_i. As a result, we express $f(X_i|s)$ as a mixture:

$$f(X_i|s) \quad = \quad \frac{1}{n_i}\sum_{j=1}^{n_i} f(X_i|J_i = j, s). \tag{3.3}$$

We denote the PDF of a single element $\mathbf{x}_{ij}$ which does not contain s as $f_0(\cdot)$ and the PDF of a single element which contains s as $f_1(\cdot|s)$. Assuming that the elements in each set are drawn independently conditioned on $J_i = j$, we can express $f(X_i|j_i = j, s)$ as a product of $n - 1$ iid RVs which follow f_0 and one RV which follows f_1: $f(X_i|J_i = j, s) = f_1(\mathbf{x}_{ij}|s)\prod_{j'=1\neq j}^{n_i} f_0(\mathbf{x}_{ij'})$. An alternative version of $f(X_i|J_i = j, s)$ is given by $f(X_i|J_i = j, s) = \frac{f_1(\mathbf{x}_{ij}|s)}{f_0(\mathbf{x}_{ij})}\prod_{j'=1}^{n_i} f_0(\mathbf{x}_{ij'})$. Substituting this expression for $f(X_i|J_i = j, s)$ into (3.3) yields

$$f(X_i|s) = \prod_{j'=1}^{n_i} f_0(\mathbf{x}_{ij'}) \cdot \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{f_1(\mathbf{x}_{ij}|s)}{f_0(\mathbf{x}_{ij})}. \tag{3.4}$$

Under the f_0 model, $\mathbf{x}_{ij}$ is distributed $\mathcal{N}(0, \sigma^2 I)$ and under the f_1 model, $\mathbf{x}_{ij}$ is distributed $\mathcal{N}(s, \sigma^2 I)$. Therefore the ratio $\frac{f_1(\mathbf{x}_{ij}|s)}{f_0(\mathbf{x}_{ij})} = \exp(-\|s\|^2/(2\sigma^2))\exp(s^T\mathbf{x}_{ij}/\sigma^2)$. Substituting this ratio and f_0 into (3.4), we find $f(X_i|s)$, substitute it into (3.2), and obtain

$$f(X|s) = \prod_{i=1}^{N}(e^{-\frac{\|s\|^2}{2\sigma^2}}\prod_{j'=1}^{n_i}\frac{1}{\sqrt{2\pi\sigma^2}^d}e^{-\frac{\|\mathbf{x}_{ij'}\|^2}{2\sigma^2}}\frac{1}{n_i}\sum_{j=1}^{n_i} e^{\frac{s^T\mathbf{x}_{ij}}{\sigma^2}}). \tag{3.5}$$

Note that $f(X|s)$ can be expressed as $f(X|s) = A(X)B(s)\cdot \prod_{i=1}^{N}\sum_{j=1}^{n_i}\exp(s^T\mathbf{x}_{ij}/\sigma^2)$, where $A(X) = \prod_{i=1}^{N}\prod_{j'=1}^{n_i}\sqrt{2\pi\sigma^2}^{-d}\exp(-\|\mathbf{x}_{ij'}\|^2/(2\sigma^2))\frac{1}{n_i}$ is only a function of the observations $X_1, \ldots, X_n$ and $B(s) = \exp(-N\|s\|^2/(2\sigma^2))$ is only a function of the parameter vector s. Note that in general the PDF $f(X|s)$ is not a member of the exponential family. However, the aforementioned modeling approach yields a fairly simple log-likelihood

$$\log f(X|s) = K - \frac{N\|s\|^2}{2\sigma^2} + \sum_{i=1}^{N} \log\left(\sum_{j=1}^{n_i} e^{\frac{s^T \mathbf{x}_{ij}}{\sigma^2}}\right). \tag{3.6}$$

The log-likelihood can be used to facilitate the derivation of the ML estimator as well as the derivation of the CRLB.

3.2 Maximum Likelihood Estimation

In order to obtain the ML estimator of $\mathbf{s}$, we consider a minimization problem of the negative objective:

$$\begin{aligned} \min_{\mathbf{s}} \quad f(\mathbf{s}) &= u(\mathbf{s}) - v(\mathbf{s}), \text{where,} \\ u(\mathbf{s}) &= \frac{\|\mathbf{s}\|^2}{2\sigma^2}; \\ v(\mathbf{s}) &= \frac{1}{N} \sum_{i=1}^{N} \log\left(\sum_{j=1}^{n_i} e^{\frac{\mathbf{s}^T \mathbf{x}_{ij}}{\sigma^2}}\right). \end{aligned}$$

Since $u(\mathbf{s})$ and $v(\mathbf{s})$ are both real-valued convex functions, $f(\mathbf{s})$ is a convex-concave function and may contain multiple local solutions. We propose majorization-minimization (MM) approach [14]. The general idea is to construct a majorizing function $g(\mathbf{s}, \mathbf{s}^{(t)})$ such that (i) $g(\mathbf{s}, \mathbf{s}^{(t)}) \geq f(\mathbf{s})$ for any $\mathbf{s}, \mathbf{s}^{(t)}$; and (ii) $g(\mathbf{s}, \mathbf{s}^{(t)}) = f(\mathbf{s})$ for any $\mathbf{s}$. Minimizing $g(\mathbf{s}, \mathbf{s}^{(t)})$ function instead of $f(\mathbf{s})$ results in the following update rule $\mathbf{s}^{(t+1)} = \arg\min_{\mathbf{s}} g(\mathbf{s}, \mathbf{s}^{(t)})$, which yields non increasing sequence of the objective, i.e., $f(\mathbf{s}^{(t+1)}) \leq f(\mathbf{s}^{(t)})$.

A simple upper bound function $g(\mathbf{s}, \mathbf{s}^{(t)})$ can be obtained by linearizing the convex function $v(\mathbf{s})$. Since $v(\mathbf{s}) \geq v(\mathbf{s}^{(t)}) + (\mathbf{s}-\mathbf{s}^{(t)})^T \Delta v(\mathbf{s}^{(t)})$, then $f(\mathbf{s}) \leq u(\mathbf{s}) - v(\mathbf{s}^{(t)}) - (\mathbf{s} - \mathbf{s}^{(t)})^T \Delta v(\mathbf{s}^{(t)}) := g(\mathbf{s}, \mathbf{s}^{(t)})$ [14]. Therefore, the upper bound $g(\mathbf{s}, \mathbf{s}^{(t)})$ is:

$$
\begin{aligned}
g(\mathbf{s}, \mathbf{s}^{(t)}) \quad &= \tfrac{\|\mathbf{s}\|^2}{2\sigma^2} - \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \cdot \frac{\sum_{j=1}^{n_i} e^{\frac{\mathbf{s}^{(t)T}\mathbf{x}_{ij}}{\sigma^2}} \cdot \frac{x_{ij}}{\sigma^2}}{\sum_{j=1}^{n_i} e^{\frac{\mathbf{s}^{(t)T}\mathbf{x}_{ij}}{\sigma^2}}} \\
&\cdot \big(\mathbf{s} - \mathbf{s}^{(t)}\big) - v(\mathbf{s}^{(t)}).
\end{aligned}
$$

By minimizing $g(\mathbf{s}, \mathbf{s}^{(t)})$ with respect to $\mathbf{s}$, we obtain the update rule:

$$
\mathbf{s}^{(t+1)} \quad = \quad \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{e^{\frac{\mathbf{s}^{(t)T}\mathbf{x}_{ij}}{\sigma^2}}}{\sum_{k=1}^{n_i} e^{\frac{\mathbf{s}^{(t)T}\mathbf{x}_{ik}}{\sigma^2}}} \mathbf{x}_{ij}. \tag{3.7}
$$

Due to the non-convexity of the objective, the ML solution depend on the initialization. However, we introduce a core idea which suggests that despite the non-convex nature of the problem, a close to optimal solution can be obtained. We rely on the observation that the log-likelihood can be approximated using the soft-max approximation of the max function: $\log(\sum_i e^{\alpha_i}) \approx \max_i \alpha_i$, yielding,

$$
\begin{aligned}
\frac{1}{N} \sum_{i=1}^{N} \log G_i(X_i|\mathbf{s}) \quad &\approx C - \tfrac{\|\mathbf{s}\|^2}{2\sigma^2} + \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \max_j \frac{\mathbf{s}^T \mathbf{x}_{ij}}{\sigma^2} \\
&= \max_{j_1,\ldots,j_N} C - \tfrac{\|\mathbf{s}\|^2}{2\sigma^2} + \frac{\mathbf{s}^T \frac{1}{N} \sum_i \mathbf{x}_{ij_i}}{\sigma^2}.
\end{aligned} \tag{3.8}
$$

Consequently, ML can be approximated by

$$
\max_{\mathbf{s},\mathbf{j}} -\frac{\|\mathbf{s}\|^2}{2} + \mathbf{s}^T \frac{1}{N} \sum_i \mathbf{x}_{\mathbf{ij_i}}, \tag{3.9}
$$

or as a minimization problem

$$\min_{\mathbf{s},\mathbf{j}} \sum_{i=1}^{N} \|\mathbf{x}_{ij_i} - \mathbf{s}\|^2 - \sum_{i=1}^{N} \|\mathbf{x}_{ij_i}\|^2, \tag{3.10}$$

where $\mathbf{j} = [j_1, j_2, \ldots, j_N]^T$. This problem is a non-trivial integer programming. A solution to a more general form is proposed in [29]:

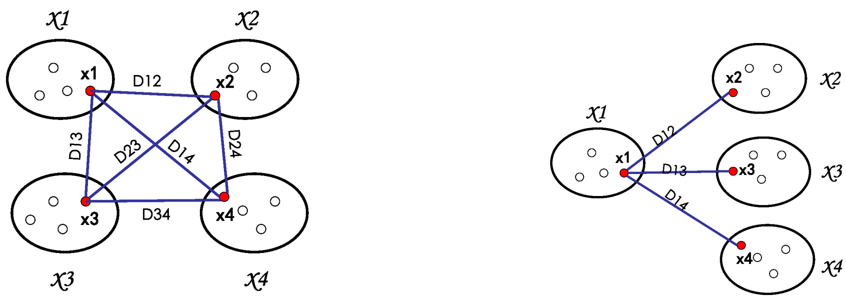
$$\min_{\mathbf{s},\mathbf{j}} \sum_{i=1}^{N} \|\mathbf{x}_{ij_i} - \mathbf{s}\|^2 + \sum_{i=1}^{N} \phi_i(\mathbf{x}_{ij_i}), \tag{3.11}$$

where $\phi_i(\mathbf{x}_{ij_i}) \geq 0$. Minimizing the objective in (3.11) with respect to $\mathbf{s}$ results in $\mathbf{s} = \frac{1}{N}\sum_{i=1}^{N} x_{ij_i}$. After substituting $\mathbf{s}$ back into (3.11), a minimization problem only with respect to $\mathbf{j}$ is obtained:

$$\hat{\mathbf{j}} = \arg\min_{\mathbf{j}} \ f(\mathbf{j}), \qquad \text{where,}$$

$$f(\mathbf{j}) = \tfrac{1}{2N} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \|\mathbf{x}_{i_1 j_{i_1}} - \mathbf{x}_{i_2 j_{i_2}}\|^2 + \sum_{i=1}^{N} \phi_i(\mathbf{x}_{ij_i}). \tag{3.12}$$

The objective in (3.12) can be viewed as a sum of edge weight in a graph given by $D_{i_1 i_2} = \|\mathbf{x}_{i_1 j_{i_1}} - \mathbf{x}_{i_2 j_{i_2}}\|^2$ and a sum of node penalties $\phi_i(\mathbf{x}_{ij_i})$. The graph is a complete graph since the sum runs over all pairs of (i_1, i_2). The solution for the complete graph requires a brute-force search which results in computational complexity $\mathcal{O}(M^N)$, where M is the number of instances per bag. To reduce the computational complexity, the proposed algorithm in [29] replaces the single complete graph by N bipartite graphs (see Fig. 3.2), reducing the computational complexity to $\mathcal{O}(M^2N^2)$ [3]. For each bipartite graph, we set aside the ith bag and calculate the sum of the squared distances from one instance in bag i to the other instance in all other bags as a function of $f_i(\mathbf{j}_i)$. Instead



complete graph bi-partite graph

Figure 3.2: Graphical representation of two approach: (3.12) and (3.13).

of minimizing the objective in (3.12), a sub-optimal solution $\tilde{\mathbf{j}}$ is obtained by solving N independent minimizations. For each i, we solve

$$\mathbf{j}^i = \arg\min_{\mathbf{j}} f_i(\mathbf{j}), \text{ where} \tag{3.13}$$

$$f_i(\mathbf{j}) = \sum_{i_2=1\neq i}^{N} \left(\|\mathbf{x}_{ij_i} - \mathbf{x}_{i_2 j_{i_2}}\|^2 + \phi_{i_2}(\mathbf{x}_{i_2 j_{i_2}})\right). \tag{3.14}$$

Then, the vector of position estimate is determined by $\tilde{\mathbf{j}} = \mathbf{j}^{i^*}$, where

$$i^* = \arg\min_i f_i(\mathbf{j}^i). \tag{3.15}$$

In [29], it is shown that the minimum of the objective $f(\mathbf{j})$ can be bounded using the $f_i(\mathbf{j})$'s as follows:

$$\frac{1}{2}\min_i f_i(\mathbf{j}^i) \leq f(\hat{\mathbf{j}}) \leq f(\tilde{\mathbf{j}}) \leq \min_i f_i(\mathbf{j}^i).$$

This sandwich inequality guarantees $f(\hat{\mathbf{j}}) \leq f(\tilde{\mathbf{j}}) \leq 2f(\hat{\mathbf{j}})$. Consequently, the bound suggests that the bi-partite approach yields a solution which guarantees that $f(\tilde{\mathbf{j}})$ the objective value in (3.12) evaluated at the sub-optimal solution is no more than the twice of its global minimum $f(\hat{\mathbf{j}})$.

Naturally this approach can be applied to the minimization in (3.10) by setting $\phi_i(\mathbf{x}_{ij_i}) = \max_t \|\mathbf{x}_{it}\|^2 - \|\mathbf{x}_{ij_i}\|^2$ in (3.11). Consequently the minimum of the objective $\sum_{i=1}^N \|\mathbf{x}_{ij_i} - \mathbf{s}\|^2 + \sum_{i=1}^N (\max_j \|\mathbf{x}_{ij}\|^2 - \|\mathbf{x}_{ij_i}\|^2)$ can be approached within a factor of 2. Moreover, this result suggests that the approximate solution $\mathbf{s}^*$

$$\mathbf{s}^* = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_{i\tilde{\mathbf{j}}} \tag{3.16}$$

can offer a feasible robust initialization to iterative methods for solving the ML in (3.6).

In effort to obtain the global solution, we propose the combination of the initialization in (3.16) and the iterations in (3.7). Inspired by this approach for solving the ML problem for the single template case, we proceed with a mixture model generalization for the multiple template case.

3.3 Performance Analysis: Cramér-Rao lower bound (CRLB) Analysis

The CRLB on the MSE of an unbiased estimator of s is given by the inverse of the Fisher information matrix (FIM) FIM $= E[\frac{\log f(X|s)}{ds} \frac{\log f(X|s)}{ds}^T]$ [11]. Since the X_is are generated in an independent fashion, we have FIM $= \sum_i \mathrm{FIM}_i$ where $\mathrm{FIM}_i = E[\frac{\log f(X_i|s)}{ds} \frac{\log f(X_i|s)}{ds}^T]$ is the FIM for a single set X_i [11]. Following the derivation in the Appendix, we obtain the expression for FIM_i:

$$\mathrm{FIM}_i = \frac{b(\rho, n_i)}{\sigma^2}(I + \frac{a(\rho, n_i) - b(\rho, n_i)}{b(\rho, n_i)} \frac{ss^T}{\|s\|^2}) \tag{3.17}$$

where

$$
\begin{aligned}
a(\rho, n) &= E_Z[(\sqrt{\rho}(1 - W_1) - \textstyle\sum_{j=1}^{n} W_j Z_j)^2] && (3.18) \\
b(\rho, n) &= \textstyle\sum_{j=1}^{n} E_Z[W_j^2] && (3.19) \\
Z_j &\sim \mathcal{N}(0, 1), \qquad j = 1, 2, \ldots, n && (3.20) \\
W_j &= \frac{e^{\rho\delta_{j1} + \sqrt{\rho} Z_j}}{\sum_{l=1}^{n} e^{\rho\delta_{l1} + \sqrt{\rho} Z_l}}, \qquad j = 1, 2, \ldots, n && (3.21)
\end{aligned}
$$

and $\rho = \frac{\|s\|^2}{\sigma^2}$. Here $a(\rho, n)$ and $b(\rho, n)$ are defined as expectations of functions of (W, Z, ρ, n) wrt RVs Z_js keeping in mind that the RV W_js are dependent on $(Z_1, \ldots, Z_n, \rho, n)$. Both $a(\rho, n)$ and $b(\rho, n)$ have the same limits: (i) $a(\rho, n), b(\rho, n) \to 1$ as $\rho \to \infty$ and (ii) $a(\rho, n), b(\rho, n) \to \frac{1}{n}$ as $\rho \to 0$ (see Fig. 3.3). For the special case in which all sets have

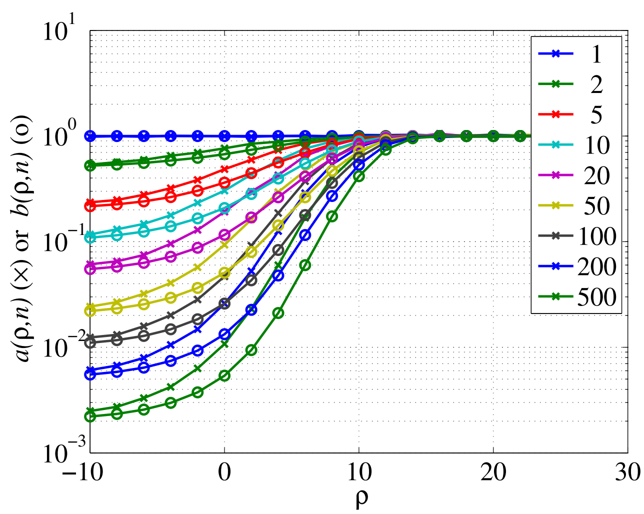


Figure 3.3: Plot of the function $a(\rho, n)$ ($\times$) and $b(\rho, n)$ ($\circ$) as a function of ρ for $n \in \{1, 2, 5, 10, 20, 50, 100, 200, 500\}$.

the same number of elements $n_i = n$, further simplification is possible. In this case, $\text{FIM}_i = \text{FIM}_1$ for $i = 1, 2, \ldots, N$. The FIM for s given $X_1, \ldots, X_N$ can be obtained as

$N \cdot \mathrm{FIM}_1$ or explicitly as

$$\mathrm{FIM} = \frac{Nb(\rho, n)}{\sigma^2}(I + \frac{a(\rho, n) - b(\rho, n)}{b(\rho, n)} \frac{ss^T}{\|s\|^2}). \tag{3.22}$$

The CRLB is computed by inverting the FIM using the Sherman-Morrison formula [24]:

$$\mathrm{CRLB} = \frac{\sigma^2}{Nb(\rho, n)}(I - \frac{a(\rho, n) - b(\rho, n)}{a(\rho, n)} \frac{ss^T}{\|s\|^2}). \tag{3.23}$$

To determine the relative error given by $\frac{E[\|\hat{s}-s\|^2]}{\|s\|^2}$, we apply the trace to $E[(\hat{s} - s)(\hat{s} - s)^T] \geq \mathrm{CRLB}$ and obtain

$$\frac{E[\|\hat{s} - s\|^2]}{\|s\|^2} \geq \frac{1}{N\mathrm{SNR}}(\frac{d-1}{d} \frac{1}{b(d\mathrm{SNR}, n)} + \frac{1}{d} \frac{1}{a(d\mathrm{SNR}, n)}), \tag{3.24}$$

where $\mathrm{SNR} = \rho/d$ is the ratio between the energy of the signal $\|s\|^2$ and the total energy for a d-dimensional noise vector $\sigma^2 d$. Note that when $\rho \to \infty$, $CRLB \to \frac{d\sigma^2}{N}$, when $\rho \to 0$, $CRLB \to \frac{nd\sigma^2}{N}$.

Chapter 4: Multiple Pattern Recognition

We introduce a novel non-Gaussian mixture model based on the single pattern model in [21]. Due to the non-convex nature of the problem, multiple local solutions may arise. To address this problem, we propose novel robust initialization and iterative updates. Based on mixture modeling approach, we first show estimation performance on synthetic data. Then, we present detection performance on real world dataset and show a significant increase in performance compared to the approaches of [21] and [29].

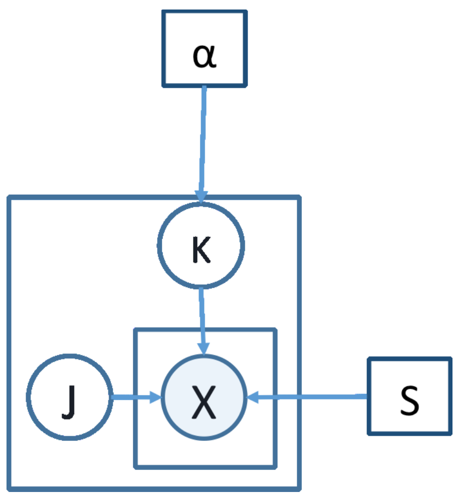


Figure 4.1: A graphical model for the K-Pattern alignment problem

To formulate this problem, consider N subsets $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_N$ of the d-dimensional Euclidean space $\mathbb{R}^d$, i.e., $\mathcal{X}_i \subseteq \mathbb{R}^d$ for $i = 1, 2, \ldots, N$. Each set is assumed to contain only one of K possible patterns $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K\}$ among other instances (see Fig. 1.1(a)). Our goal is to obtain the K patterns of interest.

4.1 Statistical K-pattern Model

To model the problem of finding the K-unknown elements in multiple sets in a noisy setting, we extend the single pattern model in [21] as shown in Fig. 4.1. We introduce hidden template id RV K in addition to the position of the template J in a given bag.

For each bag i, we organize the elements of $\mathcal{X}_i$ in a $d \times n_i$ matrix $X_i = [\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}]$ and consider joint distribution of the observations represented by the observation matrix $X = [X_1, \ldots, X_N]$ given the unknown vectors $\mathbf{s}_1, \ldots, \mathbf{s}_K$. We introduce the class prior probability α_k that satisfies $0 < \alpha_k < 1, \sum_{k=1}^{K} \alpha_k = 1$ for each probability density function $G(X_i|\mathbf{s}_k) = f(X_i|\mathbf{s}_k)$ in (3.6). Since we assume that sets are generated in an independent fashion, we express the joint distribution of sets as a product of their marginal PDFs:

$$\Lambda(X;\theta) = \prod_{i=1}^{N} f_i(X_i;\theta) \tag{4.1}$$

$$f_i(X_i;\theta) = \sum_{k=1}^{K} \alpha_k G(X_i|\mathbf{s}_k), \tag{4.2}$$

where $G(X_i|\mathbf{s}_k)$ is a the ith bag probability density function conditioned on template pattern $\mathbf{s}_k$, and $\theta = \{\alpha_1, \alpha_2, \ldots, \alpha_K, \mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K\}$. Then, the log-likelihood function is:

$$\log \Lambda(X;\theta) = \sum_{i=1}^{N} \log(\sum_{k=1}^{K} \alpha_k G(X_i|\mathbf{s}_k)). \tag{4.3}$$

Although the expectation maximization algorithm has been well-developed to solve the parameter estimation problem in mixture models, the optimization of a non-convex objective is non-trivial.

4.2 Expectation Maximization

Expectation Maximization (EM) is an iterative solution to maximum likelihood [17]. Specifically, the iterations offer a non-decreasing sequence of the likelihood function. In general, the auxiliary function $Q(\theta, \theta^{(t)})$ is:

$$\begin{aligned} Q(\theta, \theta^{(t)}) = \quad & E[\log P(X_1, X_2, \ldots, X_N, k_1, k_2, \ldots, k_N; \theta) \\ & |X_1, X_2, \ldots, X_N, \theta^{(t)}] \end{aligned}$$

The iterations are performed in two steps. In the E-step, the auxiliary function is computed as:

$$Q(\theta, \theta^{(t)}) = \sum_{i=1}^{N} \sum_{k=1}^{K} p_i^{(t)}(k|\theta^{(t)}) \log(\alpha_k G(X_i|\mathbf{s}_k)).$$

Here, $p_i^{(t)}(k|\theta^{(t)}) = \frac{\alpha_k^{(t)} G(X_i|\mathbf{s}_k^{(t)})}{\sum_{l=1}^{K} \alpha_l^{(t)} G(X_i|\mathbf{s}_l^{(t)})}$ represents the probability that the ith bag was generated by component K.

In the M-step, we maximize the auxiliary function $\max_\theta \; Q(\theta, \theta^{(t)})$ to obtain the update rule:

$$\alpha_k^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} p_i^{(t)}(k|\theta^{(t)}), \tag{4.4}$$

$$\begin{aligned} \mathbf{s}_k^{(t+1)} = \; & \arg\max_{\mathbf{s}_k} \sum_{i=1}^{N} p_i^{(t)}(k|\theta^{(t)}) \cdot \\ & \Big(C - \frac{\|\mathbf{s}_k\|^2}{2\sigma^2} + \log\big(\sum_{j=1}^{n_i} e^{\frac{\mathbf{s}_k^T \mathbf{x}_{ij}}{\sigma^2}}\big)\Big). \end{aligned} \tag{4.5}$$

The optimization in (4.5) involves the sum of convex-concave functions that cannot be solved in closed-form. We propose to solve (4.5) and obtain $\mathbf{s}_k^{(t+1)}$ by using a method described in Section 3.2. First, we find a robust initialization for $\mathbf{s}_k^{(t+1)}$ (i.e., $\mathbf{s}_k^{(t+1,0)}$). Then we use MM approach to refine the solution.

Algorithm 1 Expectation Maximization for the mixture model

1: Initialize $\theta^0 = \{\alpha_1^0, \alpha_2^0, \ldots, \alpha_K^0, s_1^0, s_2^0, \ldots, s_K^0\}$.
2: **procedure** EMforMGF(θ^0, X)
3: **while** Likelihood $\Lambda(X;\theta)$ not converged **do**
4: E-step: compute membership probability $p_{ik}^{(t+1)} = \frac{\alpha_k^{(t)} G(x_i|s_k^{(t)})}{\sum_{l=1}^K \alpha_l^{(t)} G(x_i|s_l^{(t)})}$
5: M-step: $\max_\theta \quad Q(\theta, \theta^{(t)})$ to obtain s_k
6: Running Procedure: $\hat{s_k}$ =MMforS(s_k^0, X)
7: Return θ

4.3 Robust Initialization

There are two sets of initialization parameters $\alpha_k^0 = \{\alpha_1^0, \alpha_2^0, \ldots, \alpha_K^0\}$ and $\mathbf{s}_k^0 = \{\mathbf{s}_1^0, \mathbf{s}_2^0, \ldots, \mathbf{s}_K^0\}$. The initialization of Gaussian mixture model is a well-known problem (e.g., see [5]). We can directly apply initialization techniques for the α_k^0 and $\mathbf{s}_k^0$, while initializing $\mathbf{s}_k^{(t+1,0)}$ is our focus.

By approximating the log of sum of exponential functions with the largest term in the sum $\log(\sum_{j=1}^{n_i} e^{\mathbf{s}_k^T \mathbf{x}_{ij}}) \approx \max_j \ \mathbf{s}_k^T \mathbf{x}_{ij}$ and $p_{ik} = p_i(k|\theta),\ w_{ik} = \frac{p_{ik}}{\sum_{i=1}^N p_{ik}}$, the approximated maximization problem in (4.5) becomes:

$$
\begin{aligned}
&\max_{\mathbf{s}_k} \qquad \sum_{i=1}^N w_{ik} \cdot \big(-\frac{\|\mathbf{s}_k\|^2}{2} + \max_{j_i} \quad \mathbf{s}_k^T \mathbf{x}_{ij_i}\big), \quad \text{or,} \\
&\max_{\mathbf{s}_k, \mathbf{j}} \qquad \sum_{i=1}^N w_{ik} \cdot \big(-\frac{\|\mathbf{s}_k\|^2}{2} + \quad \mathbf{s}_k^T \mathbf{x}_{ij_i}\big).
\end{aligned} \tag{4.6}
$$

We first solve for $\mathbf{s}_k$ by taking the derivative of the objective function with respect to $\mathbf{s}_k$ and setting to zero. We obtain the solution for $\mathbf{s}_k$ as $\mathbf{s}_k = \sum_{i=1}^{N} w_{ik}\mathbf{x}_{ij}$. Substituting $\mathbf{s}_k$ back into (4.6), yields:

$$\begin{aligned} \max_{\mathbf{j}} \quad & \frac{1}{2}\Big(\sum_{i=1}^{N} w_{ik}\mathbf{x}_{ij_i}\Big)^2 \quad \text{or,} \\ \max_{\mathbf{j}} \quad & \frac{1}{2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N} w_{i_1k}w_{i_2k}\mathbf{x}_{i_1j_{i_1}}^T\mathbf{x}_{i_2j_{i_2}}, \end{aligned}$$

which can be written as,

$$\begin{aligned} & \min_{\mathbf{j}} \quad f^{(k)}(\mathbf{j}), \qquad \text{where} \\ f^{(k)}(\mathbf{j}) \quad & = \tfrac{1}{2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N} w_{i_1k}w_{i_2k}||\mathbf{x}_{i_1j_{i_1}} - \mathbf{x}_{i_2j_{i_2}}||^2 \\ & + \sum_{i_1=1}^{N} w_{i_1k}\big(\max_{t} \quad ||\mathbf{x}_{i_1t}||^2 - ||\mathbf{x}_{i_1j_{i_1}}||^2\big). \end{aligned} \tag{4.7}$$

The objective in (4.7) can be viewed as a weighted sum of edge weight in a graph given by $D_{i_1i_2} = \|\mathbf{x}_{i_1j_{i_1}} - \mathbf{x}_{i_2j_{i_2}}\|^2$ and a weighted sum of node penalties $\phi_i(j) = \max_t \; ||\mathbf{x}_{it}||^2 - ||\mathbf{x}_{ij}||^2$.

This problem is similar to the single pattern matching problem. We apply the bipartite graph approach for each pattern to robustly initialize $\mathbf{s}_k^{(t)}$ for each iteration with estimated $\hat{\mathbf{s}}_k$. Since (4.7) is similar to (3.12), we can use the same procedure to obtain the ML solution $\hat{\mathbf{j}}_k$. Using $f_i^{(k)}(\mathbf{j})$ functions and solving N minimizations for each pattern individually, we obtain the approximate solution $\tilde{\mathbf{j}}_k$:

$$\begin{aligned} \mathbf{j}_k^{i_1} &= \arg\min_{\mathbf{j}} f_{i_1}^{(k)}(\mathbf{j}_{i_2}), \text{ where} \\ f_{i_1}^{(k)}(\mathbf{j}_{i_2}) &= \sum_{i_2=1\neq i_i}^{N} w_{i_2 k}\big(\|\mathbf{x}_{i_1 j_{i_1}} - \mathbf{x}_{i_2 j_{i_2}}\|^2 + \phi_{i_2}(j_{i_2})\big). \end{aligned}$$

Then, $\tilde{\mathbf{j}}_k = \mathbf{j}_k^{i_k^*}$, where

$$i_k^* = \arg\min_{i} f_{i_1}^{(k)}(\mathbf{j}_{i_2}).$$

Based on the approximate solution $\tilde{\mathbf{j}}_k$, we directly obtain the approximate estimation for $\mathbf{s}_k^*$:

$$\mathbf{s}_k^* = \sum_{i=1}^{N} w_{ik}\mathbf{x}_{i\tilde{\mathbf{j}}_k}. \tag{4.8}$$

Moreover, we can still establish a lower and upper bound for each pattern k:

$$\frac{1}{2}\sum_{i_1} w_{i_1 k} f_{i_2}^{(k)}(\mathbf{j}_k^{i_1}) \leq f^{(k)}(\tilde{\mathbf{j}}_{(k)}) \leq \min_{i_1} f_{i_1}^{(k)}(\mathbf{j}_k^{i_1}).$$

Since $\sum_{i_1} w_{i_1 k} \min_{i_1} f_{i_1}^{(k)}(\mathbf{j}_k^{i_1}) \leq \sum_{i_1} w_{i_1 k} f_{i_1}^{(k)}(\mathbf{j}_k^{i_1})$, we can further bound the lower bounded by $\frac{1}{2}\min_{i_1} f_{i_1}^{(k)}(\mathbf{j}_k^{i_1})$. Therefore,

$$\frac{1}{2}\min_{i_1} f_{i_1}^{(k)}(\mathbf{j}_k^{i_1}) \leq f^{(k)}(\hat{\mathbf{j}}_k) \leq f^{(k)}(\tilde{\mathbf{j}}_k) \leq \min_{i_1} f_{i_1}^{(k)}(\mathbf{j}_k^{i_1}).$$

This bound shows that the robust initialization finds out an approximated template such that the corresponding objective is within a factor of 2 from the optimal solution objective.

Algorithm 2 Robust Initialization

Input p_{ik} from previous E-step in EM algorithm
Compute $w_{ik} = \frac{p_{ik}}{\sum_{i=1}^{N} p_{ik}}$
procedure SEARCHGOODINSTANCES(w_{ik}, X)
  for bagid i_1 in 1,...,N **do**
    for bagid i_2 in 1,...,N$\neq i_1$ **do**
      Compute weighted distance matrix $D_{j_{i_1} j_{i_2}} = w_{i_2 k}\big(\|\mathbf{x}_{i_1 j_{i_1}} - \mathbf{x}_{i_2 j_{i_2}}\|^2 + \phi_{i_2}(j_{i_2})\big)$
      Find smallest instance position for each i_1: $[j_1^*, j_2^*, \ldots, j_N^*]$=minindex($D^T_{j_{i_1} j_{i_2}}$)
      Compute $v = v + D^T_{j_{i_1} j_{i_2}}$
    Find overall smallest distance value for each i_1:
    MinVal(i_1)=minimum value(v)
    MinIdx(i_1)=minimum index(v)
  $[i_1^*]$=min(MinVal(i_1))
  Get $[j_1*, j_2*, \ldots$ MinIdx(i_1^*)$, \ldots, j_N*]$ from optimal position collection in bag i_1^*.
  Return $\mathbf{s}_k = \sum_{i=1}^{N} w_{ik}\mathbf{x}_{ij}$

4.4 Majorization-mimimization for ML Refinement

In the M-step of the EM algorithm, a separate update rule is used for each $\mathbf{s}_k$ (see (4.6)). We can directly apply MM algorithm for each individual minimization:

$$\min_{\mathbf{s}_k} \quad \tilde{f}_k(\mathbf{s}_k), \qquad \text{where}$$
$$\tilde{f}_k(\mathbf{s}_k) = \frac{\|\mathbf{s}_k\|^2}{2\sigma^2} - \sum_{i=1}^{N} w_{ik} \log\Big(\sum_{j=1}^{n_i} e^{\frac{\mathbf{s}_k^T \mathbf{x}_{ij}}{\sigma^2}}\Big),$$

where $p_{ik} = p_i(k|\theta)$ and $w_{ik} = \frac{p_{ik}}{\sum_{i=1}^{N} p_{ik}}$. The upper bound of the objective $g_k(\mathbf{s}_k, \mathbf{s}_k^{(t')})$ is a majorizing function which satisfies $\tilde{f}_k(\mathbf{s}_k) \leq g_k(\mathbf{s}_k, \mathbf{s}_k^{(t')})$. By minimizing g_k function, a solution of $\mathbf{s}_k*$ is obtained in the t'th iteration and it provides an input to the $(t'+1)$th iteration:

$$\begin{aligned} g_k(\mathbf{s}_k, \mathbf{s}_k^{(t')}) \quad &= \frac{\|\mathbf{s}_k\|^2}{2\sigma^2} - \sum_{i=1}^N w_{ik} \frac{\sum_{j=1}^{n_i} e^{\frac{\mathbf{s}_k^{(t')T}\mathbf{x}_{ij}}{\sigma^2}} \cdot \frac{\mathbf{x}_{ij}}{\sigma^2}}{\sum_{j=1}^{n_i} e^{\frac{\mathbf{s}_k^{(t')T}\mathbf{x}_{ij}}{\sigma^2}}} \\ &\cdot (\mathbf{s}_k - \mathbf{s}_k^{(t')}) - \sum_{i=1}^N w_{ik} \log\left(\sum_{j=1}^{n_i} e^{\frac{\mathbf{s}_k^{(t')T}\mathbf{x}_{ij}}{\sigma^2}}\right). \end{aligned}$$

Then by setting $\frac{\delta g_k(\mathbf{s}_k, \mathbf{s}_k^{(t')})}{\delta \mathbf{s}} = 0$, we have the update rule:

$$\begin{aligned} \mathbf{s}_k^{(t'+1)} &= \sum_{i=1}^N w_{ik}^{(t'+1)} \sum_{j=1}^{n_i} W_{ijk}^{(t'+1)} \cdot \mathbf{x}_{ij}, \quad \text{where,} \\ w_{ik}^{(t'+1)} &= \frac{p_{ik}^{(t')}}{\sum_{i=1}^N p_{ik}^{(t')}}, \; W_{ijk}^{(t'+1)} = \frac{e^{\frac{\mathbf{s}_k^{(t')T}\mathbf{x}_{ij}}{\sigma^2}}}{\sum_{j=1}^{n_i} e^{\frac{\mathbf{s}_k^{(t')T}\mathbf{x}_{ij}}{\sigma^2}}}. \end{aligned} \tag{4.9}$$

Using a combination of robust initialization and iterative implementation of the ML estimator of $\mathbf{s}_k$, we can obtain the solution of $\mathbf{s}_k^{(t+1)}$ in (4.5).

Algorithm 3 Majorization-minimization for template $\mathbf{s}_k$

1: RobustInitialize $s\mathbf{s}_k^0 = \{\mathbf{s}_1^0, \mathbf{s}_2^0, \dots, \mathbf{s}_K^0\}$.
2: **procedure** MMforS($\mathbf{s}_k^0, X$)
3: **while** Likelihood $f(\mathbf{s}; \mathbf{s}^{(t')})$ not converged **do**
4: Recalculate $Q_{ik}^{(t'+1)} = \frac{p_{ik}^{(t')}}{\sum_{i=1}^N p_{ik}^{(t')}}$ from E-step of EM
5: Recalculate $W_{ijk}^{(t'+1)} = \frac{e^{\frac{\mathbf{s}_k^{(t')T}\mathbf{x}_{ij}}{\sigma^2}}}{\sum_{j=1}^{n_i} e^{\frac{\mathbf{s}_k^{(t')T}\mathbf{x}_{ij}}{\sigma^2}}}$
6: Update $\mathbf{s}_k^{(t'+1)} = \sum_{i=1}^N Q_{ik}^{(t'+1)} \sum_{j=1}^{n_i} W_{ijk}^{(t'+1)} \cdot \mathbf{x}_{ij}$
7: Return $\mathbf{s}_k^{final}$

Chapter 5: Experiments and Results

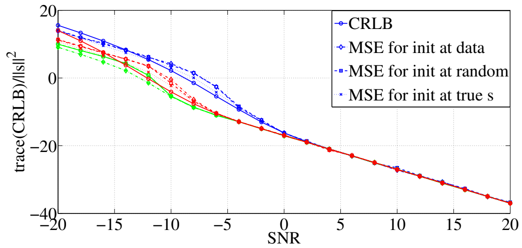
In this section, we evaluate our proposed method on both synthetic data set and on a real world data set of electric appliance activations (*Source: Pecan Street Research Institute*). We perform numerical experiments to verify the CRLB against the MSE of an iterative implementation of the ML estimator and to gain further insight into the expression for the CRLB. For single pattern recognition task, we evaluate our proposed method and compare it with Woody's method [4]. We first deploy our signature estimation procedure and compute sum of squared errors for both methods. Then we use the estimated signature to detect activation events of multiple devices from voltage measurements taken from multiple homes. For multiple pattern recognition task, we evaluate our methods in terms of Receiver Operating Characteristic curve (ROC) and Area Under the ROC curve (AUC) and also compare the results to the results presented in [29]. We also show the improved performance on ROC and AUC based on the mixture model.

5.1 Synthetic Data on Single Pattern

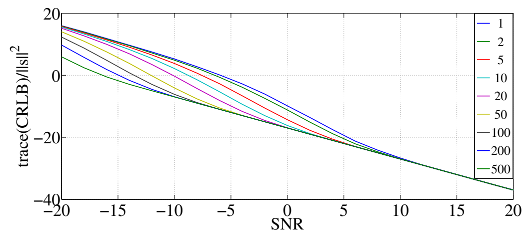
Consider the nominal setting of $N = 50$ sets with $n = 20$ $d = 100$-dimensional elements in each set for SNR $\in \{-20\text{dB}, -18\text{dB}, \ldots, 20\text{dB}\}$. We vary one parameter (d, n, and N) at a time in $\{10, 50, 100\}$ to evaluate the MSE of the ML estimator and the CRLB as a function of SNR. For each combination of parameters ($\{N, n, d, \text{SNR}\}$), we generate 100 independent Monte-Carlo (MC) realizations based on the model. For each realization, we apply the iterative implementation of the ML estimator initialized (i) at random with multiple restarts, (ii) by averaging over the largest norm element from each set and (iii) at the true value of s. Using the 100 MC runs, we estimate the MSE by averaging the squared estimation error. In Fig. 5.2 ((a), (c), (e)), we present the CRLB as a function of the SNR along with the MSE of the iterative implementation of the ML estimator. We observe that for SNR $\geq$ 0dB the MSE of the ML estimator agrees with the formula of the CRLB while for SNR $<$ 0dB, the MSE of the ML deviates from the CRLB. The random initialization and average max energy template methods are outperformed by initializing at the true s. This is expected since for low SNR the ML estimator is no longer unbiased, however, the method of initializing with the true s biases the ML estimator favorably.

Next, we focus on the evaluation of the relative CRLB for the problem (3.24). We evaluate the performance bound as a function of SNR $\in \{-20\text{dB}, -18\text{dB}, \ldots, 20\text{dB}\}$ for three different settings of the parameters: (b) $N = 50$, $n = 20$, and $d \in \{1, 2, 5, 10, 20, 50, 100, 200, 500\}$; (d) $N = 50$, $d = 100$, and $n \in \{1, 2, 5, 10, 20, 50, 100, 200, 500\}$; and (f) $n = 20$, $d = 100$, and $N \in \{1, 2, 5, 10, 20, 50, 100, 200, 500\}$. We present the relative CRLB for settings (b), (d), and (f) in Fig. 5.2. we observe that an increase in SNR, element dimension d, number of sets observed N, or a decrease the number of elements in each set n yields a decrease in the relative CRLB. We also notice that it is possible to achieve an under $-10dB$ relative CRLB, for fairly low values of SNR by either increasing the dimension d or the number of sets N. This suggests that while an increase in the number of elements in each set (i.e., larger haystacks) degrades the performance, using more sets (i.e., increasing N) allows us to compensate for this performance degradation.

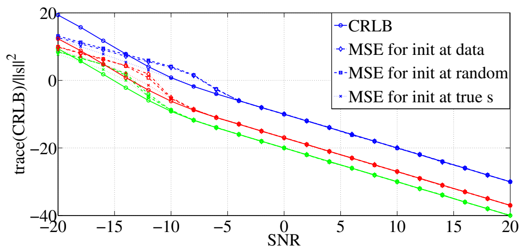
The derivation of CRLB is in the Appendix.



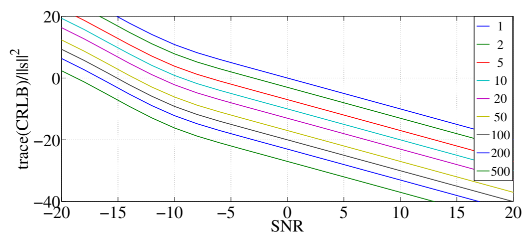
(a) $d \in \{10, 50, 100\}$, $n = 20$, $N = 50$



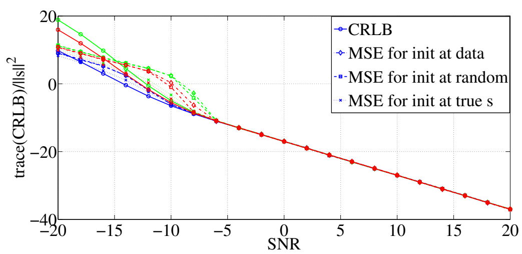
(b) $d \in \{1, 2, 5, 10, 50, 100, 200, 500\}$



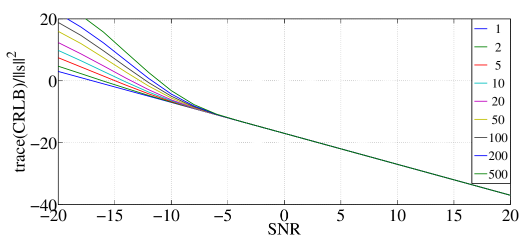
(c) $d = 100$, $n = 20$, $N \in \{10, 50, 100\}$



(d) $N \in \{1, 2, 5, 10, 50, 100, 200, 500\}$



(e) $d = 100$, $n \in \{10, 50, 100\}$, $N = 50$



(f) $n \in \{1, 2, 5, 10, 50, 100, 200, 500\}$

Figure 5.1: (a), (c), (e): Relative CRLB and MSE of the ML estimator initialized using three methods as a function of SNR. Parameter values 10, 50, 100 are shown in blue, red, and green, respectively. (b), (d), (f): Relative CRLB as a function of SNR.

5.2 Synthetic Data on Multiple Patterns

An extent synthetic data test on multiple pattern tasks are given as follows: The X_i's are generated in an independent fashion based on the K-pattern Model, where the template id for bag i is uniformly sampled in $\{1, 2, \ldots, K\}$ and the template position in the ith bag J_i is uniformly sampled in $\{1, 2, \ldots, M\}$. We choose $K = 3$ and ground-truth templates $\mathbf{s}_1(t), \mathbf{s}_2(t), \mathbf{s}_3(t)$ are designed as:

$$\begin{aligned}
\mathbf{s}_1(t) &= u(t + D/2) - u(t - D/2), \quad \text{for} \quad t = 0, 1, 2, \ldots, D; \\
\mathbf{s}_2(t) &= t, \quad \text{for} \quad 0 \leq t \leq D; \\
\mathbf{s}_3(t) &= -t, \quad \text{for} \quad 0 \leq t \leq D.
\end{aligned}$$

Note that $u(t)$ is a step function. We normalize each vector $\mathbf{s}_i = [\mathbf{s_i(1), s_i(2), \ldots, s_i(t)}]^{\mathbf{T}}$ using $\mathbf{s}_i/||\mathbf{s}_i||$ and set it as our new $\mathbf{s}_i$ for all $i = 1, 2, 3$.

Since the estimated accuracy is affected by the set of parameters $\{D, M, N, \text{SNR}\}$ (D-dimension of the template, M-number of instances per bag, N-number of bags and SNR-signal to noise ratio), we perform numerical experiments to analyze the mean squared error (MSE) of the iterative implementation of the ML estimator against different setups of parameters. Then, we also perform a detection task based on a maximum a-posterior probability (MAP) detector using the estimated patterns.

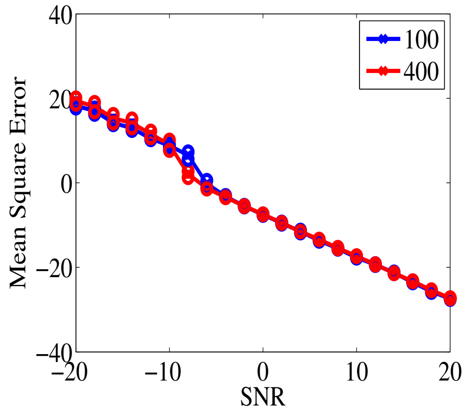
To analyze the estimation performance with respect to different parameters, we start with the same nominal setting as the single pattern recognition task. Then we vary one parameter (D, M, and N) at a time as $D \in \{100, 400\}$ and $M, N \in \{10, 50\}$ to evaluate the MSE of the ML estimator as a function of SNR. For each combination of parameters $\{N, M, D, \text{SNR}\}$, we generate 50 independent Monte-Carlo (MC) realizations based on our mixture model. Since EM is sensitive to the initialization, we use 10 iterations of different random values of α_k^0 and $\mathbf{s}_k^0$ and choose the estimate yielding the largest likelihood value. Using the 50 MC runs, we compute the sum of each k empirical MSE with the mean and its confidence interval. In Fig. 5.2, we present the MSE as a function of the SNR of the iterative implementation of the ML estimator. Increasing SNR and the number of bags N yields a decrease in the relative MSE, while increasing template dimension D and the number of instances in each bag M yields a small increase in the relative MSE when SNR is less than $-10dB$. We also notice that it is possible to achieve an under $-10dB$ relative MSE, for fairly low values of SNR by either increasing the dimension D or the number of sets N. This suggests that using more sets compensate for the performance degradation when choosing larger number of elements in each set.

In order to verify that recognizing more patterns will increase the performance significantly in learning task, we designed a GLRT framework for detecting the position $\mathbf{J}$ of unknown patterns $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K\}$ given a new dataset $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$. We denote $\mathbf{x}_j*$ as an instance in the set that contains one of the true templates $\mathbf{s}_k \in \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K\}$. Our goal is to detect the position j in each bag and analyze the performance of our detectors as a function of k.

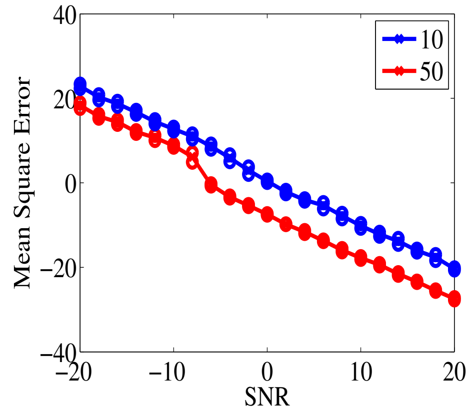
By maximizing the posterior probability of $\mathbf{J}$ and $\mathbf{K}$, which can be written as $P(J=j, K=k|X) = \frac{f(X|J=j,K=k)\cdot P(J=j)P(K=k)}{\sum_{j=1}^{M}\sum_{k=1}^{K} f(X|J=j,K=k)\cdot P(J=j)P(K=k)} \propto f(X|J=j,K=k)P(J=j)P(K=k)$, we can directly obtain the detector as

$$\max_{J,K} \; P(J=j, K=k|X).$$

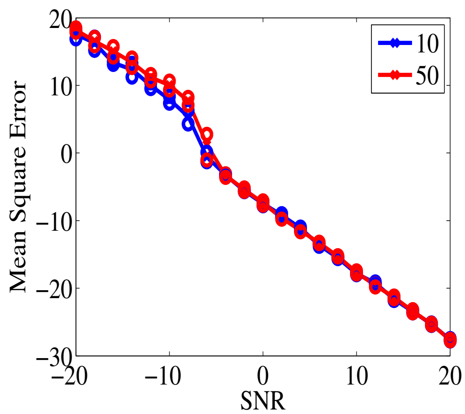
To simplify the notation, we omit the dependence of $P(J=j, K=k|X;\alpha,\mathbf{s})$ on α and $\mathbf{s}$ and write it as $P(J=j, K=k|X)$. Since for each bag, $f(X_i|J=j, K=k) = \prod_{j=1}^{M} f_0(x_j) \cdot \frac{f_1(x_j|s_k)}{f_0(x_j)}$, and based on the Gaussian model for f_1 and f_0, the log of the



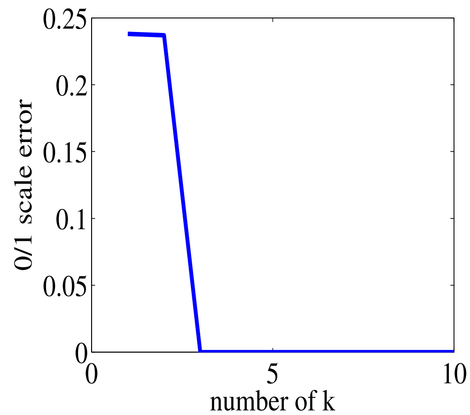
(a) $D \in \{100, 400\}$, $M = 20$, $N = 50$



(b) $N \in \{10, 50\}$, $D = 100$, $M = 20$



(c) $M \in \{10, 50\}$, $D = 100$, $N = 50$



(d) (d) Detection error vs. K

Figure 5.2: MSE of the ML estimator as a function of SNR in (i)-(iii) and detection error vs. K in (iv) .

posterior probability can be rewritten as:

$$\begin{aligned}
\log(f(X|J = j, K = k)) &= -\frac{||x_j - s_k||^2}{2\sigma^2} + \frac{||x_j||^2}{2\sigma^2} + C; \\
\log(P(J = j)) &= -\log(M); \\
\log(P(K = k)) &= \log(\alpha_k).
\end{aligned}$$

By taking the negative $\log P(J = j, K = k|X)$, we obtain the detector as:

$$\min_{j,k} \quad \frac{2s_k^T x_j - ||s_k||^2}{2\sigma^2} + \log(\alpha_k). \tag{5.1}$$

In this experiment, we apply this detector to the synthetic data set with 50 bags and we detect the position of the pattern based on the K-pattern estimation results of $\hat{\mathbf{s}}_k, \hat{\alpha}_k$. If the position of a pattern is true, we count it as a hit, otherwise, we count it as a miss. The error given by $P(J \neq j|X)$ is presented in Fig. 5.2(d) as a function of the number of the templates.

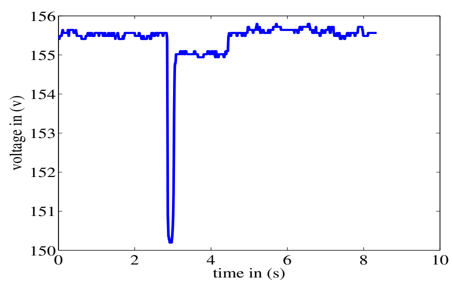
5.3 Real-world Data for Single Pattern Recognition

In our experiments, we use the Pecan Street dataset (*Source: Pecan Street Research Institute*). The dataset contains four homes of disaggregated, time-sampled electricity usage data with 120 sampling frequency. The data set includes voltage and apparent power readings for both the whole home and disaggregated household appliances in a period of 25 days. Since the voltage peak to peak (V_{pp}) waveform is corrupted by spike noise, we apply a five-tap median filter to de-spike the voltage waveforms.

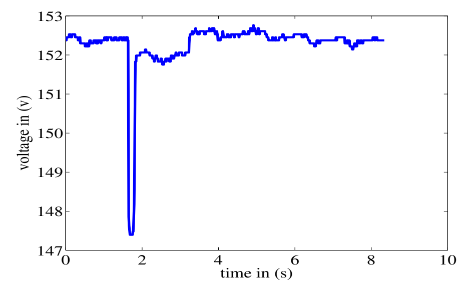
Home appliance recognition from voltage envelope measurements relies on the unique signatures associated with each appliance. To extract voltage envelope waveforms containing the appliance activation transient response, a power meter measurement of the appliance of interest provides a rough interval in which the activation response is present. Since the precise start time of the activation is unavailable, blind joint delay estimation is key to this problem. In our problem formulation, a set of N signals containing the appliance activation signature are extracted from training data for each appliance,

$$y_i(t), \qquad i = 1, 2, \ldots, N, \quad 1 \le t \le T.$$

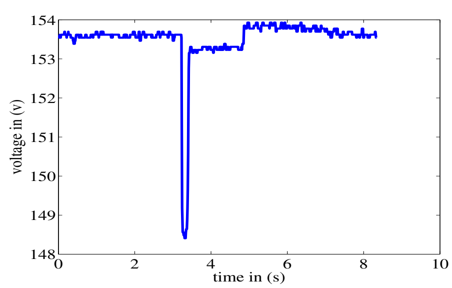
Figure 5.3 (a)-(c) shows three different templates $y_1(t)$ to $y_3(t)$ containing the activation signature from the same appliance.



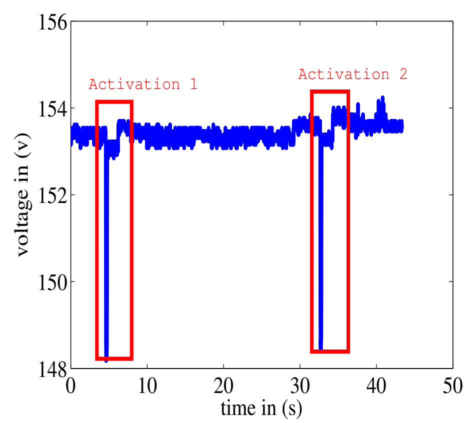
(a) $y_1(t)$



(b) $y_2(t)$



(c) $y_3(t)$



(d) Detection

Figure 5.3: Three air-conditioning activation events (a)-(c) and template detection illustration (d)

Our goal is to detect the presence of an activation signature in a new (test) signal (see Fig. 5.3 (d)) using the information from the training data $y_i(t)$ for $i = 1, 2, \ldots, n$

and $1 \leq t \leq T$.

We identify two tasks: (i) detect the presence of a signature in a new signal, and (ii) obtain an accurate estimate of the signature present in the multiple training templates.

5.3.1 Generalized Likelihood Ratio Test Dector

The GLRT framework is a common and powerful statistical test method to determine between multiple hypothesis models which involve unknown parameters. The GLRT [12] for observation vector x is given by:

$$\frac{\max_{\theta_1} \; p(x|H_1, \theta_1)}{\max_{\theta_0} \; p(x|H_0, \theta_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \rho, \tag{5.2}$$

where θ_0 and θ_1 are the unknown parameters associated with the statistical model under hypothesis H_0 and H_1, respectively, and ρ is the non-negative test threshold. The test can be rephrased in terms of the negative log-likelihood as $\min_{\theta_1}(-\log p(x|H_1, \theta_1)) - \min_{\theta_0}(-\log p(x|H_0, \theta_0)) \underset{H_1}{\overset{H_0}{\gtrless}} \rho'$, where $\rho' = -\log \rho$ is a real-valued threshold [12].

$$\min_{\theta_1}(-\log p(x|H_1, \theta_1)) - \min_{\theta_0}(-\log p(x|H_0, \theta_0)) \underset{H_1}{\overset{H_0}{\gtrless}} \rho', \tag{5.3}$$

Based on the Gaussian model for H_0, we have$-\log f(\mathbf{y}_{test}|H_0, \theta_0) = \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_{test}(t) - B)^2 + c$ where $c = \frac{T}{2}\log(2\pi\sigma^2)$ and similarly, based on H_1 we have$-\log f(\mathbf{y}_{test}|H_1, \theta_1) = \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_{test}(t) - s(t-\tau) - B)^2 + c$. Minimizing the negative log-likelihood for H_0 and H_1 respectively yields $B_{H_0} = \bar{y}_{test} \triangleq \frac{1}{T}\sum_{t=1}^{T} y_{test}(t)$ and $B_{H_1} = \bar{y}_{test} - \frac{1}{T}\sum_{t=1}^{T} s(t-\tau)$. Substituting the optimal values of B into the respective negative log-likelihood yields the following form for the simplified GLRT:

$$\min_{\tau} \sum_{t=1}^{T} (y_{test}(t) - \bar{y}_{test} - (s(t-\tau) - \overline{s(t-\tau)}))^2 \\ - \sum_{t=1}^{T} (y_{test}(t) - \bar{y}_{test})^2 \underset{H_1}{\overset{H_0}{\gtrless}} \rho'', \tag{5.4}$$

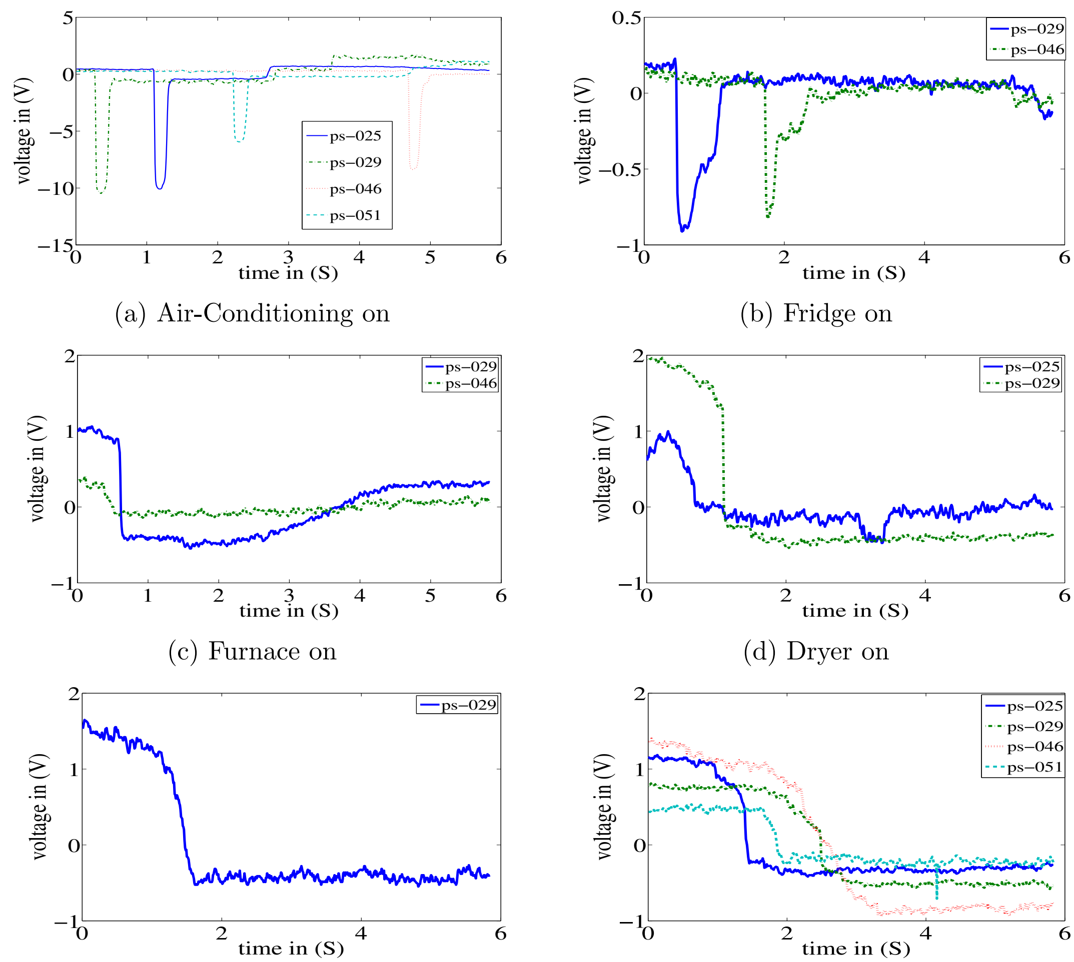
where $\overline{s(t-\tau)} \triangleq \frac{1}{T}\sum_{t=1}^{T} s(t-\tau)$ and $\rho'' = 2\rho'\sigma^2$. Expanding the quadratic form of the first term on the LHS of (5.4) yields the following simplification of the test to a simple correlation test as our detector [12]:

$$\max_{\tau} \sum_{t=1}^{T} (y_{test}(t) - \bar{y}_{test}(t))(s(t-\tau) - \overline{s(t-\tau)}) \underset{H_0}{\overset{H_1}{\gtrless}} \rho'' \tag{5.5}$$

where $\rho'' = \rho'\sigma^2 - \frac{1}{2}\sum_t (s(t-\tau) - \overline{s(t-\tau)})^2)$. The resulting detector compares the maximum sample cross-covariance function to a threshold to determine the presence or absence of the template s. It is closely related to the well-known matched filter [12, p. 95] in which a test signal is correlated with a given template.

5.3.2 Signature Maximum Likelihood Estimation

Our goal is to learn the activation signature for each appliance using the training data and to test the detection performance obtained using a detector which uses the estimated signature. In our experiment, we split four home data into training data (in the period 11/17/2012-11/25/2012 with around 50 activations per appliance) and test data (in the period 11/26/2012-12/11/2012 with around 80 activations per appliance). The ground truth (based on the independent measurement from a commercial power meter) regarding the activation events is obtained by identifying a power increase from 0 to 80 watt or more.



(a) Air-Conditioning on (b) Fridge on

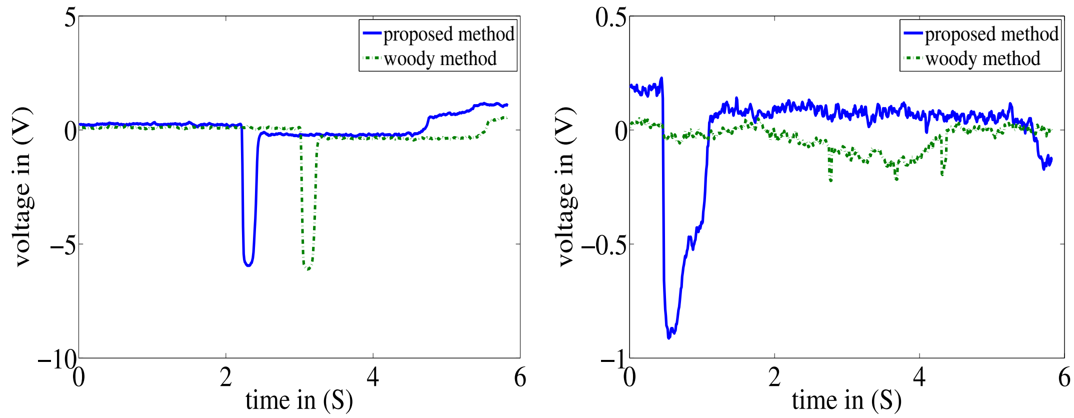
(c) Furnace on (d) Dryer on

(e) Microwave on (f) Oven on

Figure 5.4: Activation patterns of six household appliances from four homes.

For the training phase, we obtain activation events from the training data by extracting a segment $y_i(t)$ of $T = 1000$ samples around the reported activation time for each event i in the training dataset. We consider an activation signature window size of $T_0 = 700$. We use (3.13)-(3.15) to find the delay of the activation signature within each segment. For each segment, we extract the portion associated with the activation signature and average following (2.5). Similarly, we apply the Woody's method [4] to obtain a signature for each device. The mean square error (MSE) $\frac{1}{N}\sum_{i=1}^{N}\|Y_i e_{\tau_i} - \tilde{\mathbf{s}}\|^2$ is presented in the Table 5.1.

After the training process, we generate distinct activation patterns of each appliance in each home. In Fig. 5.6, we present activation patterns of six appliances in four homes (PS-025, PS-029, PS-046, and PS-051). Based on the activation patterns estimated during the training phase, we apply the detector in (5.5) to the test data. We apply the detection scheme to each hourly file in a period of more than ten days and acquire the receiver operating characteristic (ROC) curve for each appliance in all homes. We present the area under the ROC curve (AUC) for each of the appliances available in each of the homes in Table 5.1. We observe that for most of the appliances the AUC is over 80%. Additionally, we observe that for devices which have a distinct single consistent activation pattern such as air-conditioning, both the proposed method and the Woody's method achieve AUC of over 0.9 (e.g., see air-conditioning signature in Fig. 5.5(a)). However, we notice that for some of the other appliances, Woody's method fails to find the activation pattern yielding a low AUC of 0.5 (e.g., see fridge signature in Fig. 5.5(b)). Moreover, when a given appliance has more than one activation pattern, the detection performance degrades for all algorithms tested. The template obtained by averaging over the multiple activation patterns may not resemble either of the patterns. Additionally, when one of the activation signatures is prominent, the average follows it closely. However, during the test phase, the less prominent activation signatures of a given appliance may not be detected.



(a) Air-conditioning (b) Fridge

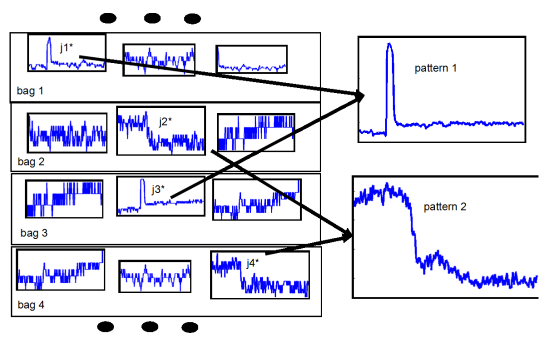
Figure 5.5: Template comparison for the proposed method and the Woody's method [4].

House ID	App. Name	MSE. Our Method	MSE. Woody's Method	AUC Our Method	AUC Woody's Method
PS-025	Air-Cond.	2517.93	3201.17	0.95066	0.90309
PS-025	Oven	1812.61	3243.28	0.52177	0.38571
PS-029	Air-Cond.	5356.00	3723.36	0.91496	0.88241
PS-029	Fridge	1573.47	4605.86	0.71906	0.30876
PS-029	Furnace	1582.34	2201.17	0.86338	0.39473
PS-029	Dryer	3812.68	7316.96	0.99142	0.55087
PS-029	Microwave	2168.59	5440.45	0.87869	0.47560
PS-029	Oven	1953.54	2323.37	0.91030	0.53450
PS-046	Air-Cond.	1548.87	2366.34	0.84892	0.85404
PS-046	Fridge	1303.00	2142.41	0.49252	0.49213
PS-046	Furnace	623.93	690.28	0.53887	0.55045
PS-046	Oven	4193.05	5024.09	0.91824	0.49346
PS-051	Air-Cond.	2730.66	2569.54	0.91311	0.92936
PS-051	Oven	2115.58	2599.95	0.78501	0.47497

Table 5.1: MSE for the estimated template and AUC for the proposed method and for Woody's method [4]

5.4 Real-world Data for Multiple Pattern Recognition

it is shown that the blind joint delay estimation for single activation pattern yields mean AUC around 75%. However, for some appliances such as oven, the AUC is as low as 50%. Our goal is jointly identify multiple activation patterns together for the same appliance in a multiple bag setting. We show an example of oven (in Fig. 5.6) with two activation patterns repeated multiple times.



(e) Oven Activations

Figure 5.6: Examples of Identifying Two Activation Patterns of Oven.

5.4.1 Comparison with Single Pattern Model

To make the comparison fair, we use the same amount of training examples and choose a window of size 700 (i.e., $D = 700$) during the training phase. We compare the ROC and AUC in [29] with those of the K-pattern model for both single activation appliances and multi-activation appliances.

Since not all appliances have multiple activation patterns, we test the performance of our proposed algorithm by increasing K (the number of patterns). Based on the activation patterns estimated during the training phase, we apply the same detector

$\max_\tau \sum_{t=1}^{T}(y_{test}(t) - \bar{y}_{test}(t))(s(t-\tau) - \overline{s(t-\tau)}) \underset{H_0}{\overset{H_1}{\gtrless}} \rho''$ as in [29] to each hourly file in the test data with a period of more than ten days and acquire the ROC curve for each appliance in each of the four homes. In [29], because the model is not robust to outliers, the training data has been filtered. To make the comparison fair, we also apply the filtering process to the training data such that the training examples are free from outliers. The corresponding AUCs for all appliances available in each home on both single pattern model and K-pattern model is present in the TABLE 5.2.

House ID	App. Name	AUC (Single)	AUC (Mixture$K=1$)
PS025	Air-Cond.	0.95066	0.95228
PS025	Oven	0.52177	0.52076
PS029	Air-Cond.	0.91496	0.91496
PS029	Fridge	0.71906	0.68795
PS029	Furnace	0.86338	0.86519
PS029	Dryer	0.99142	0.98460
PS029	Microwave	0.87869	0.87926
PS029	Oven	0.91030	0.91602
PS046	Air-Cond.	0.84892	0.85882
PS046	Fridge	0.49252	0.49680
PS046	Furnace	0.53887	0.57652
PS046	Oven	0.91824	0.95471
PS051	Air-Cond.	0.91311	0.91418
PS051	Oven	0.78501	0.77505

Table 5.2: AUCs of single pattern model vs. mixture model with $K = 1$ on test dataset

We observe that mixture model with $K = 1$ attains a reasonable ROC curves and the similar AUCs compared to the single pattern model in [29]. We also notice that AUCs for most appliances in the mixture model is slightly higher than the single pattern model (10/14 entries are higher in the mixture model in Table 5.2). This suggests that the mixture model, built upon the single pattern model, is not decreasing the performance of the single pattern model. Moreover, we concentrate on increasing the AUCs for the appliances revealing more than one activation pattern.

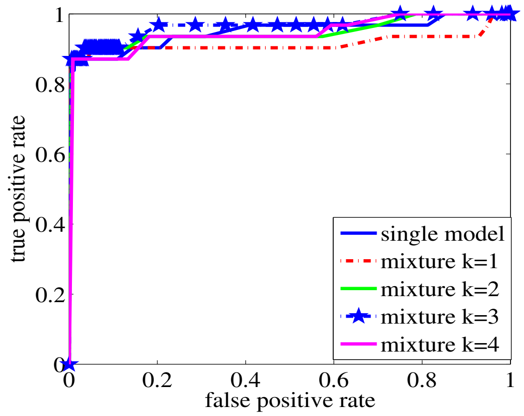
5.4.2 K-pattern Model Results

By increasing the number of activation patterns K in the model, more than one activation pattern can be identified, but each pattern would be more coarse. This is due to the effect of averaging with less bags for each pattern. To capture the variation among patterns and maintain the completeness of the training data, we train the mixture model on the unfiltered training data. Then, we apply the same detector to test for different K.

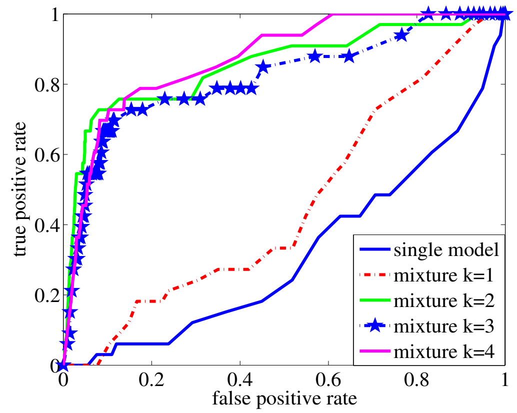
For appliances with only one activation pattern, such as air-conditioning, furnace, and microwave, considering a larger K model would not affect the performance significantly (see Fig. 5.7(a) and (c)). For those appliances with multiple activation patterns, such as oven, dryer and fridge, mixture model captures more variations of the activation patterns yielding a significant improvement in detection accuracy (see Fig. 5.7(b) and (d)).

In the case of $K = 1$, the performance of single pattern model and mixture model is similar (see TABLE 5.2). To test the performance of mixture model by the effect of varying K, we present the AUCs for $K = 1$, $K = 2$, $K = 3$ and $K = 4$ in four homes which is shown in TABLE 5.3. Even though increasing the number K in the training phase is more time consuming, the detection accuracy has increased in the testing phase. The computation complexity of training the mixture model with K components is K times more than the single pattern model. However, the AUCs improved significantly for $K = 3$ than the single pattern model, especially for those appliances containing multiple activation patterns (see Table 5.2 and Table 5.3). Moreover, without manually filtering the training data, we can save time and avoid human intervention.

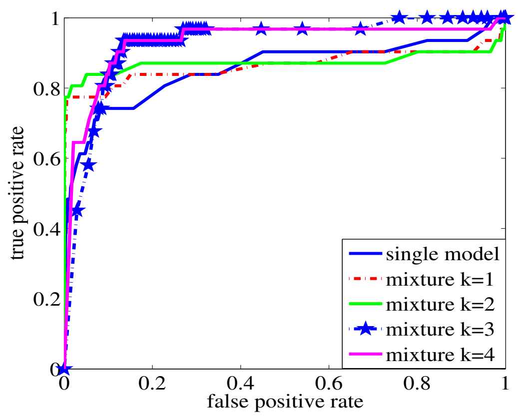
We observe that the AUCs for most appliances change slightly when varying K from



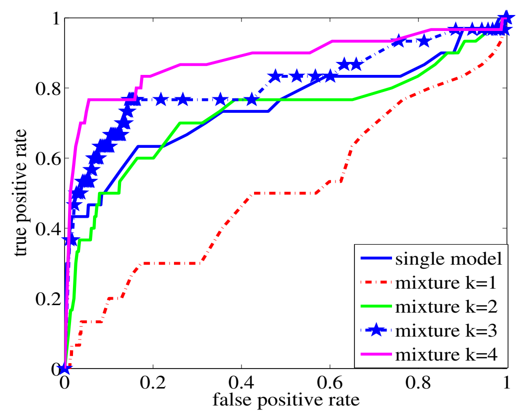
(a) PS025 Air-cond. ROC



(b) PS046 Oven ROC



(c) PS029 Furnace ROC



(d) PS046 Fridge ROC

Figure 5.7: ROC plots for Single pattern detection and for multiple pattern detection.

1 to 4, while some appliances change significantly, such as oven in home PS025, fridge, dryer and microwave in home PS029 and air-conditioning, fridge in home PS046. We notice that the AUC may not increase by increasing K because outliers can be recognized as a pattern introducing more 'false alarms'. Even if some appliances have only one activation pattern, the mixture model approach increases the AUC by capturing the variations in patterns. In practice, K can be selected using cross-validation to prevent

House ID	App. Name	AUC ($K=1$)	AUC ($K=2$)	AUC ($K=3$)	AUC ($K=4$)
PS025	Air-Cond.	0.91536	0.94718	0.96430	0.94521
PS025	Oven	0.62589	0.77490	0.77117	0.78919
PS029	Air-Cond.	0.89135	0.93373	0.93373	0.93373
PS029	Fridge	0.69454	0.82033	0.82255	0.77734
PS029	Furnace	0.92872	0.87298	0.92531	0.92872
PS029	Dryer	0.14849	0.98840	0.96926	0.97028
PS029	Microwave	0.70571	0.94661	0.93492	0.92364
PS029	Oven	0.92116	0.95478	0.95478	0.91151
PS046	Air-Cond.	0.27775	0.85115	0.94300	0.95887
PS046	Fridge	0.50963	0.72579	0.81084	0.87933
PS046	Furnace	0.50576	0.55790	0.57826	0.52459
PS046	Oven	0.45611	0.85403	0.81768	0.87384
PS051	Air-Cond.	0.91362	0.93661	0.96314	0.96314
PS051	Oven	0.77862	0.78036	0.75660	0.79800

Table 5.3: AUCs of mixture model by varying K

overfitting.

Chapter 6: Conclusion and Future Work

6.1 Summary

We presented a problem setting in which an unknown pattern present in multiple sets is sought after. We first presented a single pattern statistical model in which the element of interest is corrupted by Gaussian noise and is placed among noisy elements. We extended the model to statistical mixture model for finding multiple patterns across multiple sets. We proposed an iterative algorithm for solving ML estimation of the unknown pattern with theoretical guarantees on the optimal solution. Then, we extended the solution by using an EM-based inference framework with robust initialization approach. We tested the performance of our proposed algorithms on both synthetic dataset and real world dataset. The results on synthetic data showed that for high SNR, MSE for multiple patterns would achieve a similar performance as that of the single pattern.

In real world dataset, we first provided a formulation of the problem of automatic detection of electric appliance activation from voltage measurements as a blind joint delay estimation. We then presented a GLRT detection scheme based on activation signatures estimated in the maximum likelihood framework. In the experiment, we achieved an improved detection performance relative to Woody's method. For most appliances, the AUC achieved by our method was over 80%.

A disadvantage of the single pattern model is whenever there are multiple activation patterns of the same appliance, the model may not resemble any of the patterns. However, this can be solved by the K-pattern model. For some appliances, we observed a significant performance increase when using the K pattern model instead of the single pattern model. Moreover, if a home appliance has only one activation pattern, using the mixture model maintained the performance of the single pattern model. The mixture model introduces significant performance improvement relative to the single pattern model when an appliance exhibits multiple activation patterns.

6.2 Publications

The following is a list of publications I have worked on during my master period from 2012 September to 2014 June.

6.2.1 Journal papers

1. Zeyu, You and Raich, Raviv. Learning Recurring Unknown Patterns with Robustness to Outliers. *Signal Processing, IEEE Transactions on*, IEEE, 2015, In preparation.

6.2.2 Conference papers

2. Zeyu, You, Raich, Raviv and Yonghong, Huang. An Inference Framework for Detection of Home Appliance Activation from Voltage measurements, *International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2014 IEEE International Conference on*, IEEE, 2014.

3. Raich, Raviv and Zeyu, You. Looking for the Same Needle in Multiple Haystacks: Performance Bounds, *International Conference on Acoustics, Speech, and Signal*

Processing (ICASSP), 2014 IEEE International Conference on, IEEE, 2014.

4. Zeyu, You, Raich, Raviv and Yonghong, Huang. Mixture modeling and inference for recognition of multiple recurring unknown patterns, *International Joint Conference on Neural Networks (IJCNN), 2014 IEEE International Conference on*, IEEE-WCCI, 2014.

6.3 Future Work

The future directions of this work are divided into three different categories:

1. Making the algorithm robust to outliers: Training data may contain pure noise examples among more pronounced transient responses. We would like to develop algorithms that are robust to pure noise examples.

2. Reducing the computational complexity from quadratic to linear: The original approach of solving the non-convex objective for pattern recognition is quadratic in the number of total instances from all bags. We want to develop an algorithm that reduces computational complexity.

3. Classification: The previous detection tasks are mainly to detect the presence of the activation template in the data. We want to develop a classifier that can determine the label of the patterns (activation templates). We are also interested in classification of appliances across multiple homes.

Bibliography

[1] Timothy L Bailey and Charles Elkan. Fitting a mixture model by expectation maximization to discover motifs in bipolymers, 1994.

[2] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.

[3] John Adrian Bondy and U. S. R. Murty. *Graph Theory with Applications.* North-Holland, 1976.

[4] Aline Cabasson and Olivier Meste. Time delay estimation: A new insight into the Woody's method. *Signal Processing Letters, IEEE*, 15:573–576, 2008.

[5] Usama M Fayyad, Cory Reina, and Paul S Bradley. Initialization of iterative refinement clustering algorithms. In *KDD*, pages 194–198, 1998.

[6] Jon Froehlich, Eric Larson, Sidhant Gupta, Gabe Cohn, Matthew Reynolds, and Shwetak Patel. Disaggregated end-use energy sensing for the smart grid. *Pervasive Computing, IEEE*, 10(1):28–39, 2011.

[7] Sidhant Gupta, Matthew S Reynolds, and Shwetak N Patel. ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139–148. ACM, 2010.

[8] Yuan Hao, Mohammad Shokoohi-Yekta, George Papageorgiou, and Eamonn Keogh. Parameter-free audio motif discovery in large data archives. In *IEEE International Conference on Data Mining (ICDM)*. IEEE, 2013.

[9] George William Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.

[10] Ronald A Iltis and Laurence Mailaender. An adaptive multiuser detector with joint amplitude and delay estimation. *Selected Areas in Communications, IEEE Journal on*, 12(5):774–785, 1994.

[11] Steven M Kay. Fundamentals of statistical signal processing, volume i: Estimation theory (v. 1). 1993.

[12] Steven M Kay. *Fundamentals of statistical signal processing: detection theory*. Prentice-hall, 1998.

[13] Donald E Knuth, James H Morris, Jr, and Vaughan R Pratt. Fast pattern matching in strings. *SIAM journal on computing*, 6(2):323–350, 1977.

[14] Gert R Lanckriet and Bharath K Sriperumbudur. On the convergence of the concave-convex procedure. In *Advances in neural information processing systems*, pages 1759–1767, 2009.

[15] Christopher Laughman, Kwangduk Lee, Robert Cox, Steven Shaw, Steven Leeb, Les Norford, and Peter Armstrong. Power signature analysis. *Power and Energy Magazine, IEEE*, 1(2):56–63, 2003.

[16] Charles E Lawrence, Stephen F Altschul, Mark S Boguski, Jun S Liu, Andrew F Neuwald, and John C Wootton. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *science*, 262(5131):208–214, 1993.

[17] Todd K Moon. The expectation-maximization algorithm. *Signal processing magazine, IEEE*, 13(6):47–60, 1996.

[18] Abdullah Mueen, Eamonn J Keogh, Qiang Zhu, Sydney Cash, and M Brandon Westover. Exact discovery of time series motifs. In *SIAM International Conference on Data Mining (SDM)*, pages 473–484. SIAM, 2009.

[19] Jason S Papadopoulos and Richa Agarwala. Cobalt: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, 23(9):1073–1079, 2007.

[20] Pavel A Pevzner, Sing-Hoi Sze, et al. Combinatorial approaches to finding subtle signals in dna sequences. In *ISMB*, pages 269–278, 2000.

[21] Raviv Raich and You Zeyu. Looking for the same needle in multiple haystacks: Performance bounds. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014.

[22] Douglas A Reynolds and Richard C Rose. Robust text-independent speaker identification using gaussian mixture speaker models. *Speech and Audio Processing, IEEE Transactions on*, 3(1):72–83, 1995.

[23] Richard Roy and Thomas Kailath. Esprit-estimation of signal parameters via rotational invariance techniques. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(7):984–995, 1989.

[24] Jack Sherman and Winifred J Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, 1950.

[25] Jan Sijbers and AJ Den Dekker. Maximum likelihood estimation of signal amplitude and noise variance from MR data. *Magnetic Resonance in Medicine*, 51(3):586–594, 2004.

[26] A-J Van der Veen, Michaela C Vanderveen, and Arogyaswami Paulraj. Joint angle and delay estimation using shift-invariance techniques. *Signal Processing, IEEE Transactions on*, 46(2):405–418, 1998.

[27] H Wang and M Kaveh. Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(4):823–831, 1985.

[28] Peter Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *Audio and Electroacoustics, IEEE Transactions on*, 15(2):70–73, 1967.

[29] You Zeyu, Raviv Raich, and Huang Yonghong. An inference framework for detection of home appliance activation from voltage measurements. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014.

APPENDICES

Appendix A: CRLB derivation

A.1 Single set Fisher Information Matrix (FIM)

We derive the expression for $\text{FIM}_1 = E[\frac{\log f(X_1|s)}{ds}\frac{\log f(X_1|s)}{ds}^T]$. The log-likelihood of s given $X_1 = [\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n_1}]$ is obtained by setting $N = 1$ in (3.6). To simplify the derivation of of FIM_1, we omit the dependence on i and write $\mathbf{x}_{1j}$ simply as $\mathbf{x}_j$ and n_1 as n. Consequently, we rewrite $\log f(X_1|s)$ as

$$\log f(X_1|s) = K - \frac{\|s\|^2}{2\sigma^2} + \log(\sum_{j=1}^{n} e^{\frac{s^T\mathbf{x}_j}{\sigma^2}}). \tag{A.1}$$

The derivative of the log-likelihood $\log f(X_1|s)$ wrt to s is given by:

$$\frac{\log f(X_1|s)}{ds} = \frac{1}{\sigma^2}\sum_{j=1}^{n} w_j(\mathbf{x}_j - s), \quad \text{where} \tag{A.2}$$

$$w_j = e^{\frac{s^T\mathbf{x}_j}{\sigma^2}} / (\sum_{j=1}^{n} e^{\frac{s^T\mathbf{x}_j}{\sigma^2}}) \tag{A.3}$$

are sum-to-one non-negative weights that depend on (X_1, s). Due to symmetry in the position of s, the distribution of $\frac{\log f(X_1|s)}{ds}$ is invariant of J and hence

$$\text{FIM}_1 = E[\frac{\log f(X_1|s)}{ds}\frac{\log f(X_1|s)}{ds}^T|J = 1]. \tag{A.4}$$

We can further simplify this as

$$\text{FIM}_1 = \frac{1}{\sigma^4} E_X[(\sum_k w_k(\mathbf{x}_k - s))(\sum_k w_k(\mathbf{x}_k - s))^T | J = 1] \tag{A.5}$$

Since we proceed with the calculation of FIM_1 under the assumption that $J = 1$, we assume $\mathbf{x}_1 \sim \mathcal{N}(s, \sigma^2 I)$ and $\mathbf{x}_j \sim \mathcal{N}(0, \sigma^2 I)$ for $j = 2, \ldots, n$.

Due to the dependencies between the w_j's and $\mathbf{x}_j$'s (see (A.3)), the computation of the FIM is non-trivial. To simplify the dependencies, we consider a change of coordinates. First, we introduce the $d \times n$ matrix Z whose entries are iid zero mean unit variance Gaussian random variables, $Z_{lk} \sim \mathcal{N}(0, 1)$. The jth column of Z is given by $\mathbf{z}_j = [Z_{1j}, Z_{2j}, \ldots, Z_{dj}]^T$. Then, we express $\mathbf{x}_j$ in terms of Z as:

$$\mathbf{x}_j = s\delta_{j1} + \sigma U \mathbf{z}_j \tag{A.6}$$

where $U = [\frac{s}{\|s\|}, u_2, \ldots, u_d]$ is a unitary matrix and δ_{ab} is the delta function, which satisfies $\delta_{ab} = 1$ if $a = b$ and 0 otherwise. Note that with the exception of the first column of matrix U all other columns can be chosen arbitrarily while maintaining the orthonormality. To express the w_i's in terms of Z, we substitute $\frac{s^T \mathbf{x}_j}{\sigma^2} = \rho\delta_{j1} + \sqrt{\rho} Z_{1j}$ into (A.3) and express w_j in terms of Z as

$$w_j = \frac{e^{\rho\delta_{j1} + \sqrt{\rho} Z_{1j}}}{\sum_{l=1}^n e^{\rho\delta_{l1} + \sqrt{\rho} Z_{1l}}}. \tag{A.7}$$

Note that for all $j = 1, 2, \ldots, n$, w_j depends only on $Z_{11}, \ldots, Z_{1n}$ and is independent of $Z_{l1}, \ldots, Z_{ln}$ for all $l = 2, \ldots, n$. Next, we express the score, $\frac{d \log f(X_1|s)}{ds}$, in the new coordinates. Since the score depends on $(\mathbf{x}_j - s)$, we compute its new coordinates using

(A.6):

$$U^T(\mathbf{x}_j - s) = \|s\|\mathbf{e}_1\delta_{j1} + \sigma\mathbf{z}_j - \|s\|\mathbf{e}_1 \tag{A.8}$$

where $\mathbf{e}_t$ denotes the canonical vector in which the tth element is one and all other elements are zero. Using the variable substitution in (A.8), we can re-write FIM_1 as

$$\text{FIM}_1 = \frac{1}{\sigma^2}UMU^T, \qquad \text{where} \tag{A.9}$$

$$M = \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E\big[w_{j_1}w_{j_2}(\sqrt{\rho}\mathbf{e}_1(\delta_{j_11}-1)+\mathbf{z}_{j_1})\cdot (\sqrt{\rho}\mathbf{e}_1(\delta_{j_21}-1)+\mathbf{z}_{j_2})^T\big]. \tag{A.10}$$

We proceed with the calculation of the entries of matrix M. The kl term of the matrix M is given by

$$M_{kl} = \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1}w_{j_2}(\sqrt{\rho}\delta_{k1}(\delta_{j_11}-1)+Z_{kj_1})\cdot (\sqrt{\rho}\delta_{l1}(\delta_{j_21}-1)+Z_{lj_2})]. \tag{A.11}$$

If $k = l$, we can simplify M_{kl} as

$$M_{kl} = E\Big[\big(\sum_{j=1}^{n} w_j(\sqrt{\rho}\delta_{k1}(\delta_{j1}-1)+Z_{kj})\big)^2\Big]. \tag{A.12}$$

For the case of $k = l = 1$, we have $\delta_{l1} = \delta_{k1} = 1$. Hence the argument of the square in (A.12) is $\sum_{j=1}^{n} w_j(\sqrt{\rho}(\delta_{j1}-1)+Z_{1j}) = -\sqrt{\rho}\sum_{j=2}^{n} w_j + \sum_{j=1}^{n} w_jZ_{1j} = -\sqrt{\rho}(1-w_1) + \sum_{j=1}^{n} w_jZ_{1j}$. Substituting $\sum_{j=1}^{n} w_j(\sqrt{\rho}(\delta_{j1}-1)+Z_{1j}) = -\sqrt{\rho}(1-w_1) + \sum_{j=1}^{n} w_jZ_{1j}$

into (A.12) yields

$$
\begin{aligned}
M_{11} &= \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1}w_{j_2}(\sqrt{\rho}(\delta_{j_1 1}-1)+Z_{1j_1})(\sqrt{\rho}(\delta_{j_2 1}-1)+Z_{1j_2})] && \text{(A.13)}\\
&= \rho\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1}w_{j_2}(\delta_{j_2 1}-1)(\delta_{j_2 1}-1)] && \text{(A.14)}\\
&+ \sqrt{\rho}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1}w_{j_2}(\delta_{j_1 1}-1)Z_{1j_2}] && \text{(A.15)}\\
&+ \sqrt{\rho}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1}w_{j_2}(Z_{1j_1}(\delta_{j_2 1}-1)] && \text{(A.16)}\\
&+ \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1}w_{j_2}Z_{1j_1}Z_{1j_2}] && \text{(A.17)}\\
&= \rho E[(\sum_{j_1=2}^{n} w_j)^2] - 2\sqrt{\rho}E[\sum_{j_1=2}^{n} w_{j_1}\sum_{j_2=1}^{n} w_{j_2}Z_{1j_2}] + E[(\sum_{j=1}^{n} w_jZ_{1j})^2] && \text{(A.18)}\\
&= \rho E[(1-w_1)^2] - 2\sqrt{\rho}E[(1-w_1)\sum_{j=1}^{n} w_jZ_{1j}] + E[(\sum_{j=1}^{n} w_jZ_{1j})^2] && \text{(A.19)}\\
&= E[(\sqrt{\rho}(1-w_1)-\sum_{j=1}^{n} w_jZ_{1j})^2]. && \text{(A.20)}
\end{aligned}
$$

We continue with the evaluation of M_{kl} terms for which $k = 2, \ldots, n$ and $l = 1$. Substituting $\delta_{k1} = 0$ into (A.11), we simplify M_{kl} as

$$
\begin{aligned}
M_{k1} &= \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1}w_{j_2}Z_{kj_1}(\sqrt{\rho}(\delta_{j_2 1}-1)+Z_{1j_2})]\\
&= \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1}w_{j_2}(\sqrt{\rho}(\delta_{j_2 1}-1)+Z_{1j_2})]E[Z_{kj_1}]\\
&= 0, && \text{(A.21)}
\end{aligned}
$$

where the second equality holds due to the independence between Z_{kj} for $k = 2, \ldots, n$

and $j = 1, \ldots, n$ and (Z_{1j}, w_j) for $j = 1, \ldots, n$ and the third equality hold since all Z_{kj} are zero mean. By symmetry $M_{1k} = M_{k1} = 0$. Continue with $k, l = 2, \ldots, n$. Recognizing that $\delta_{k1} = \delta_{l1} = 0$, we simplify M_{kl} as

$$M_{kl} = \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1} w_{j_2} Z_{kj_1} Z_{lj_2}] \tag{A.22}$$

$$= \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1} w_{j_2}] E[Z_{kj_1} Z_{lj_2}] \tag{A.23}$$

$$= \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} E[w_{j_1} w_{j_2}] \delta_{kl}\delta_{j_1 j_2} \tag{A.24}$$

$$= \sum_{j=1}^{n} E[w_j^2]\delta_{kl} \tag{A.25}$$

since $E[w_{j_1} w_{j_2} Z_{kj_1} Z_{lj_2}] = E[w_{j_1} w_{j_2}] E[Z_{kj_1} Z_{lj_2}] = E[w_{j_1} w_{j_2}] \cdot \delta_{kl}\delta_{j_1 j_2}$. Note that $M_{kk} = \sum_{j=1}^{n} E[w_j^2]$ for $k = 2, \ldots, n$ and $M_{kl} = 0$ for $k \neq l$. The matrix M is a diagonal matrix and is given by $M = \mathrm{diag}([M_{11}, M_{22}, \ldots, M_{22}])$. We can write M as

$$M = (M_{11} - M_{22})\mathbf{e}_1\mathbf{e}_1^T + M_{22} I.$$

Multiplying on the left with U and on the right with U^T and dividing by σ^2, we obtain FIM_1 as

$$\frac{1}{\sigma^2} U M U^T = \frac{1}{\sigma^2}((M_{11} - M_{22})\frac{ss^T}{\|s\|^2} + M_{22} I).$$

For the special case of $n_i = n$ for all i, we have FIM$= \frac{N}{\sigma^2}((M_{11} - M_{22})\frac{ss^T}{\|s\|^2} + M_{22} I) = \frac{N M_{22}}{\sigma^2}(\frac{M_{11} - M_{22}}{M_{22}}\frac{ss^T}{\|s\|^2} + I)$ and consequently CRLB$= \frac{\sigma^2}{N M_{22}}(-\frac{M_{11} - M_{22}}{M_{11}}\frac{ss^T}{\|s\|^2} + I)$ (since $(I + auu^T)(I - buu^T) = I + (a - b - ab)uu^T = I$, if $b = \frac{a}{a+1}$). We notice that the CRLB in the direction of the unit vector $s/\|s\|$ is given by $\frac{\sigma^2}{nM_{11}}$ and $\frac{\sigma^2}{nM_{22}}$ in any other direction orthogonal to s.

A.2 Extra equations

Note that since the expected value of the score is 0, we have

$$
\begin{aligned}
0 = \sum_{j_1=1}^{n} E[w_j(\sqrt{\rho}\mathbf{e}_1(\delta_{j1}-1)+z_j)] &= \sqrt{\rho}\mathbf{e}_1 E[\sum_{j_1=1}^{n} w_j(\delta_{j1}-1)] + \sum_{j_1=1}^{n} E[w_j z_j] &\quad (A.26)\\
&= -\sqrt{\rho}\mathbf{e}_1 E[\sum_{j=2}^{n} w_j] + \sum_{j=1}^{n} E[w_j z_j] &\quad (A.27)\\
&= \sqrt{\rho}\mathbf{e}_1 E[w_1 - 1] + \sum_{j=1}^{n} E[w_j z_j] &\quad (A.28)
\end{aligned}
$$

$$
\sum_{j=1}^{n} E[w_j Z_{1j}] = \sqrt{\rho} E[1 - w_1] \quad (A.29)
$$