LOOKING FOR THE SAME NEEDLE IN MULTIPLE HAYSTACKS: PERFORMANCE BOUNDS

Raviv Raich and Zeyu You

School of EECS, Oregon State University, Corvallis, OR, 97331-5501, USA

raich@eecs.oregonstate.edu, youz@onid.orst.edu

ABSTRACT

We consider the problem of finding the same pattern in multiple sets. This problem can be applied in a variety of signal processing and machine learning problems including DNA sequencing and detection of electrical signatures. In our problem setting, each set contains only a single unknown pattern of interest among many other patterns. To understand the performance limitations associated with this setting, we focus on the evaluation of the Cramér-Rao lower bound (CRLB). We introduce a probabilistic model for the problem. The random position of a pattern in a given set gives rise to a mixture model and consequently a non trivial CRLB analysis. We present the derivation of the CRLB for the problem and provide a numerical evaluation of the CRLB. We verify our expression for the CRLB against the mean-squared-error of an iterative implementation of the maximum likelihood estimator.

***Index Terms*—** Cramér-Rao lower bound, Mean Squared Error, Pattern Matching

1. INTRODUCTION

We 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 [1], 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 [2], sinusoidal model with unknown frequencies [3], or a steering vector with unknown angles or delays [4]. In computer science, fast pattern matching [5] 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 [6] enables the process of pattern discovery.

The problem formulation presents two important challenges. The first challenge pertains to the development of algorithms that would blindly identify a repeated pattern among multiple sets. The second challenge involves modelling and performance limitations study. In this paper, we focus on the latter. Our contributions in this paper are as follows: (i) we formulate the problem of finding the same needle in multiple haystacks as an inference problem and present a novel probabilistic model for the problem; and (ii) we obtain performance limitations for this problem using the derivation and analysis of Cramér-Rao lower bound (CRLB).

2. PROBLEM FORMULATION

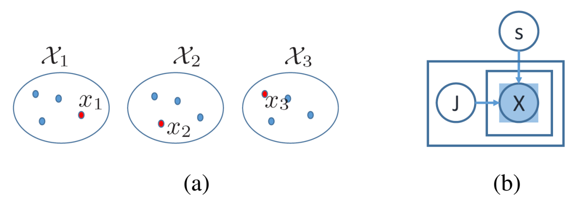


Fig. 1: (a) Our setting: each set $\mathcal{X}_i$ is assumed to contain one instance of a desired element s. Our goal is to identify the desired element s along with the most similar element in each set, i.e., $x_i \in \mathcal{X}_i$. (b) A graphical model for the alignment problem

2.1. Problem Setup

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. 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.

2.2. A Probabilistic 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

$$\mathbf{x}_{ij} = \begin{cases} s + \nu_{ij} & j = j_i \\ \nu_{ij} & j \neq j_i \end{cases} \tag{1}$$

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:

$$f(X|s) \quad = \quad \prod_{i=1}^{N} f(X_i|s). \tag{2}$$

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:

$$f(X_i|s) \quad = \quad \frac{1}{n_i} \sum_{j=1}^{n_i} f(X_i|J_i = j, s). \tag{3}$$

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) yields

$$f(X_i|s) = \prod_{j'=1}^{n_i} f_0(\mathbf{x}_{ij'}) \cdot \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{f_1(\mathbf{x}_{ij}|s)}{f_0(\mathbf{x}_{ij})}. \tag{4}$$

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 (4), we find $f(X_i|s)$, substitute it into (2), and obtain

$$f(X|s) = \prod_{i=1}^{N} \Big(e^{-\frac{\|s\|^2}{2\sigma^2}} \prod_{j'=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}^{d}} e^{-\frac{\|\mathbf{x}_{ij'}\|^2}{2\sigma^2}} \frac{1}{n_i} \sum_{j=1}^{n_i} e^{\frac{s^T \mathbf{x}_{ij}}{\sigma^2}}\Big). \tag{5}$$

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

$$\log f(X|s) = K - \frac{N\|s\|^2}{2\sigma^2} + \sum_{i=1}^{N} \log\Big(\sum_{j=1}^{n_i} e^{\frac{s^T \mathbf{x}_{ij}}{\sigma^2}}\Big). \tag{6}$$

The log-likelihood can be used to facilitate the derivation of the ML estimator as well as the derivation of the CRLB. In this paper, we set our goal to gain understanding of the performance limitations associated with estimation of unknown template s. We proceed with the derivation of the CRLB.

3. PERFORMANCE ANALYSIS

The Cramér-Rao lower bound (CRLB) on the MSE of an unbiased estimator of s is given by the inverse of the Fisher information matrix (FIM) $\text{FIM} = E[\frac{\log f(X|s)}{ds} \frac{\log f(X|s)}{ds}^T]$ [7]. Since the X_is are generated in an independent fashion, we have $\text{FIM} = \sum_i \text{FIM}_i$ where $\text{FIM}_i = E[\frac{\log f(X_i|s)}{ds} \frac{\log f(X_i|s)}{ds}^T]$ is the FIM for a single set X_i [7]. Following the derivation in the Appendix, we obtain the expression for FIM_i:

$$\text{FIM}_i = \frac{b(\rho, n_i)}{\sigma^2}\Big(I + \frac{a(\rho, n_i) - b(\rho, n_i)}{b(\rho, n_i)} \frac{ss^T}{\|s\|^2}\Big) \tag{7}$$

where

$$a(\rho, n) \quad = \quad E_Z[(\sqrt{\rho}(1 - W_1) - \textstyle\sum_{j=1}^{n} W_j Z_j)^2] \tag{8}$$

$$b(\rho, n) \quad = \quad \textstyle\sum_{j=1}^{n} E_Z[W_j^2] \tag{9}$$

$$Z_j \quad \sim \quad \mathcal{N}(0, 1), \qquad j = 1, 2, \ldots, n \tag{10}$$

$$W_j \quad = \quad \frac{e^{\rho\delta_{j1} + \sqrt{\rho}Z_j}}{\sum_{l=1}^{n} e^{\rho\delta_{l1} + \sqrt{\rho}Z_l}}, \qquad j = 1, 2, \ldots, n \tag{11}$$

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) $a(\rho, n), b(\rho, n) \rightarrow 1$ as $\rho \rightarrow \infty$ and (ii) $a(\rho, n), b(\rho, n) \rightarrow \frac{1}{n}$ as $\rho \rightarrow 0$ (see Fig. 2). For the special

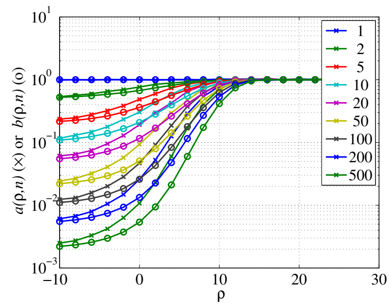


Fig. 2: Plot of the function $a(\rho, n)$ ($\times$) and $b(\rho, n)$ ($\circ$) as a function of ρ for $n \in \{1, 2, 5, 10, 20, 50, 100, 200, 500\}$.

case in which all sets have 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 $N \cdot \text{FIM}_1$ or explicitly as

$$\text{FIM} = \frac{Nb(\rho, n)}{\sigma^2}\Big(I + \frac{a(\rho, n) - b(\rho, n)}{b(\rho, n)} \frac{ss^T}{\|s\|^2}\Big). \tag{12}$$

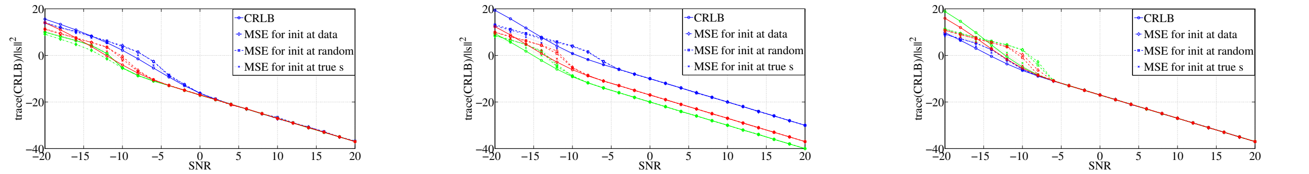
The CRLB is computed by inverting the FIM using the Sherman-Morrison formula [8]:

$$\text{CRLB} = \frac{\sigma^2}{Nb(\rho, n)}\Big(I - \frac{a(\rho, n) - b(\rho, n)}{a(\rho, n)} \frac{ss^T}{\|s\|^2}\Big). \tag{13}$$

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

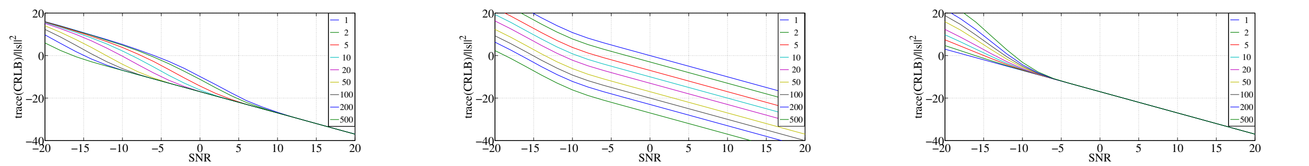
$$\frac{E[\|\hat{s} - s\|^2]}{\|s\|^2} \geq \frac{1}{N\text{SNR}}\Big(\frac{d - 1}{d} \frac{1}{b(d\text{SNR}, n)} + \frac{1}{d} \frac{1}{a(d\text{SNR}, n)}\Big), \tag{14}$$

where $\text{SNR} = \rho/d$ is the ratio between the energy of the signal $\|s\|^2$ and the total energy for a d-dimensional noise vector $\sigma^2 d$.



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

Fig. 3: 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.



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

Fig. 4: Relative CRLB as a function of SNR.

4. NUMERICAL EVALUATION

In this section, 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.

4.1. Verifying the CRLB

To verify the CRLB, we compare the CRLB to the MSE obtained by applying the ML estimator. To that end we derived an iterative implementation of the ML estimator. Due to space limitations, we omit the lengthy derivation. The derivation follows the approach of [9] for minimization of sum of convex and concave functions as we have in (6). The ML update equations are given by

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

Note that upon convergence the resulting equation is identical to the equation obtained by differentiating the log-likelihood in (6) with respect to s and setting to zero.

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. 3, 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.

4.2. CRLB analysis

Next, we focus on the evaluation of the relative CRLB for the problem (14). 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: (i) $N = 50$, $n = 20$, and $d \in \{1, 2, 5, 10, 20, 50, 100, 200, 500\}$; (ii) $N = 50$, $d = 100$, and $n \in \{1, 2, 5, 10, 20, 50, 100, 200, 500\}$; and (iii) $n = 20$, $d = 100$, and $N \in \{1, 2, 5, 10, 20, 50, 100, 200, 500\}$. We present the relative CRLB for settings (i), (ii), and (iii) in Fig. 4. 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.

5. CONCLUSION

We presented a problem setting in which an unknown element present in multiple sets is sought after. We presented a statistical model in which the element of interest is corrupted by Gaussian noise and is placed among noisy elements. The study of CRLB for the problem revealed that the relative mean squared error associated with the estimation of the unknown element of interest depends on the signal-to-noise ratio, element dimension, number of elements per set, and the number of sets. When the SNR is large, the CRLB approaches the single element per set case CRLB. For a medium SNR, a trade-off can be obtained: to achieve the same relative CRLB, the difficulty of determining the position of the element in a set can be offset by including more sets. Our current work aims to further the application of the model derived in this paper towards efficient estimation algorithms in terms of both computational complexity and accuracy.

A. APPENDIX: SINGLE SET 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 (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

$$\log f(X_1|s) = K - \frac{\|s\|^2}{2\sigma^2} + \log(\sum_{j=1}^{n} e^{\frac{s^T \mathbf{x}_j}{\sigma^2}}). \tag{15}$$

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

$$\frac{\log f(X_1|s)}{ds} = \frac{1}{\sigma^2} \sum_{j=1}^{n} w_j(\mathbf{x}_j - s), \quad \text{where} \tag{16}$$

$$w_j = e^{\frac{s^T \mathbf{x}_j}{\sigma^2}} / (\sum_{j=1}^{n} e^{\frac{s^T \mathbf{x}_j}{\sigma^2}}) \tag{17}$$

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

$$\text{FIM}_1 = E[\frac{\log f(X_1|s)}{ds} \frac{\log f(X_1|s)}{ds}^T |J = 1]. \tag{18}$$

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 (17)), 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:

$$\mathbf{x}_j = s\delta_{j1} + \sigma U \mathbf{z}_j \tag{19}$$

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 (17) and express w_j in terms of Z as

$$w_j = \frac{e^{\rho\delta_{j1} + \sqrt{\rho} Z_{1j}}}{\sum_{l=1}^{n} e^{\rho\delta_{l1} + \sqrt{\rho} Z_{1l}}}. \tag{20}$$

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 (19):

$$U^T(\mathbf{x}_j - s) = \|s\| \mathbf{e}_1 \delta_{j1} + \sigma \mathbf{z}_j - \|s\| \mathbf{e}_1 \tag{21}$$

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 (21), we can re-write FIM_1 as

$$\text{FIM}_1 = \frac{1}{\sigma^2} U M U^T, \qquad \text{where} \tag{22}$$

$$M = \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} E\big[w_{j_1} w_{j_2} (\sqrt{\rho} \mathbf{e}_1 (\delta_{j_1 1} - 1) + \mathbf{z}_{j_1}) \cdot (\sqrt{\rho} \mathbf{e}_1 (\delta_{j_2 1} - 1) + \mathbf{z}_{j_2})^T\big]. \tag{23}$$

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

$$M_{kl} = \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} E[w_{j_1} w_{j_2} (\sqrt{\rho} \delta_{k1} (\delta_{j_1 1} - 1) + Z_{k j_1}) \cdot (\sqrt{\rho} \delta_{l1} (\delta_{j_2 1} - 1) + Z_{l j_2})]. \tag{24}$$

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

$$M_{kl} = E\Big[\big(\sum_{j=1}^{n} w_j (\sqrt{\rho} \delta_{k1} (\delta_{j1} - 1) + Z_{kj})\big)^2\Big]. \tag{25}$$

For the case of $k = l = 1$, we have $\delta_{l1} = \delta_{k1} = 1$. Hence the argument of the square in (25) is $\sum_{j=1}^{n} w_j (\sqrt{\rho} (\delta_{j1} - 1) + Z_{1j}) = -\sqrt{\rho} \sum_{j=2}^{n} w_j + \sum_{j=1}^{n} w_j Z_{1j} = -\sqrt{\rho}(1 - w_1) + \sum_{j=1}^{n} w_j Z_{1j}$. Substituting $\sum_{j=1}^{n} w_j (\sqrt{\rho}(\delta_{j1} - 1) + Z_{1j}) = -\sqrt{\rho}(1 - w_1) + \sum_{j=1}^{n} w_j Z_{1j}$ into (25) yields

$$M_{11} = E[(\sqrt{\rho}(1 - w_1) - \sum_{j=1}^{n} w_j Z_{1j})^2]. \tag{26}$$

We continue with the evaluation of M_{kl} terms for which $k = 2, \ldots, n$ and $l = 1$. Substituting $\delta_{k1} = 0$ into (24), we simplify M_{kl} as

$$\begin{aligned} M_{k1} &= \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} E[w_{j_1} w_{j_2} Z_{k j_1} (\sqrt{\rho}(\delta_{j_2 1} - 1) + Z_{1 j_2})] \\ &= \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} E[w_{j_1} w_{j_2} (\sqrt{\rho}(\delta_{j_2 1} - 1) + Z_{1 j_2})] E[Z_{k j_1}] \\ &= 0, \end{aligned} \tag{27}$$

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

$$M_{kl} = \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} E[w_{j_1} w_{j_2} Z_{k j_1} Z_{l j_2}] = \sum_{j=1}^{n} E[w_j^2] \delta_{kl} \tag{28}$$

since $E[w_{j_1} w_{j_2} Z_{k j_1} Z_{l j_2}] = E[w_{j_1} w_{j_2}] E[Z_{k j_1} Z_{l j_2}] = E[w_{j_1} w_{j_2}] \cdot \delta_{kl} \delta_{j_1 j_2}$. 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 $M = \text{diag}([M_{11}, M_{22}, \ldots, M_{22}])$. 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

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

A. REFERENCES

[1] 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*, vol. 262, no. 5131, pp. 208–214, 1993.

[2] 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*, vol. 33, no. 4, pp. 823–831, 1985.

[3] Richard Roy and Thomas Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 984–995, 1989.

[4] 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*, vol. 46, no. 2, pp. 405–418, 1998.

[5] Donald E Knuth, James H Morris, Jr, and Vaughan R Pratt, "Fast pattern matching in strings," *SIAM journal on computing*, vol. 6, no. 2, pp. 323–350, 1977.

[6] Jason S Papadopoulos and Richa Agarwala, "Cobalt: constraint-based alignment tool for multiple protein sequences," *Bioinformatics*, vol. 23, no. 9, pp. 1073–1079, 2007.

[7] Steven M Kay, *Fundamentals of statistical signal processing, volume I: Estimation theory (v. 1)*, Prentice Hall, 1993.

[8] 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*, vol. 21, no. 1, pp. 124–127, 1950.

[9] Alan L Yuille and Anand Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, 2003.