RICE UNIVERSITY

Antenna Arrays for Wireless CDMA Communication Systems

by

Raghavendra K. Madyastha

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Approved, Thesis Committee:

Behnaam Aazhang, Chair
Professor in Electrical and Computer Engineering

Don H. Johnson
Professor in Electrical and Computer Engineering

Joseph R. Cavallaro
Associate Professor in Electrical and Computer Engineering

Danny C. Sorensen
Professor of Computational and Applied Mathematics

Houston, Texas

May, 1997

Antenna Arrays for Wireless CDMA Communication Systems

Raghavendra K. Madyastha

Abstract

The estimation of code delays along with amplitudes and phases of different users constitutes the first stage in the demodulation process in a CDMA communication system. The delay estimation stage is termed the *acquisition* stage and forms the bottleneck for the detection of users' bitstreams; accurate detection necessitates accurate acquisition. Most existing schemes incorporate a single sensor at the receiver, which leads to an inherent limit in the *acquisition based capacity*, which is the number of users that can be simultaneously acquired. In this thesis we combine the benefits of spatial processing in the form of an antenna array at the receiver along with code diversity to gain an increase in the capacity of the system. An additional parameter to be estimated now is the direction of arrival (DOA) of each user. We demonstrate the gains in parameter estimation with the incorporation of spatial diversity. We propose two classes of delay-DOA estimation algorithms — a maximum likelihood algorithm and a subspace based algorithm (MUSIC). With reasonable assumptions on the system we are able to derive computationally efficient estimation algorithms and demonstrate the gains achieved in exploiting multiple sensors at the receiver. In addition, we also investigate the benefits of spatial diversity in linear multiuser detection. We consider two linear multiuser detectors, the decorrelating detector and the linear MMSE detector (chosen for their near-far properties) and characterize the performance increase in the multisensor case. We observe that in many cases, the gain can be directly captured in terms of the number of sensors in the array.

Acknowledgments

"Do not go gentle into that good night,
Rage, rage against the dying of the light."
— Dylan Thomas

The journey has not been easy, it has not been smooth. A large part of it has been uphill and the legs are weary. Yet now it draws to a close; the chapter is finally complete. And reflecting on the path taken, the conclusion is inescapable: it has been worth it! If I am but a goulash of experiences, then Rice is definitely the strongest flavor. This university's hallowed grounds have contributed more to my character than anything or anybody else. Rice will remain forever etched in my memory.

I would like to express my sincere and heartfelt gratitude to my mentor and friend, Dr. Behnaam Aazhang. He has been a constant source of inspiration, encouragement, humor and money, none of which I could have done without. When I found my confidence floundering and my faith in my abilities waning, he was the spark that ignited my hope and fueled my creativity. He was the gentle wind in my sail. He was the strength that kept me going and going and going .... He gave me two of the greatest gifts any teacher can give a pupil, the gift of desire and the gift of trust; the unquenchable desire for knowledge along with the trust in myself to pursue it. Dr. Aazhang — Behnie, I will simply say "Thank you. You're the greatest."

I would like to thank my other committee members, Dr. Johnson, Dr. Cavallaro and Dr. Sorensen for so readily consenting to be on my committee and for all their constructive comments and suggestions. I would also like to thank Ari Hottinen and Jorma Lilleberg of Nokia, Inc. for their advice, support and patience.

Nora Quiocho, efficiency and warmth personified, did much to make my life in the

department as smooth and enjoyable as it could be and I owe her my sincere thanks. Without her egging on and encouragement my graduation might never have occurred.

I could not have achieved what I did without the incredibly fruitful discussions with my officemates Andrew Sendonaris and Steve Bensley. I am indebted to my friend Kishore Kota for not letting me go "nonlinear" in the final days of my struggle. I would also like to thank Ram, Brinda, Partha, Mahesh, Kang and all my other friends for being there whenever I needed them and making my life at Rice so fascinating and full.

Finally, this acknowledgement would most certainly be incomplete without an expression of my sincere gratitude to my parents and brothers for enabling me to be what I am today. They never doubted that I would make it even though I gave them enough cause for doubt and concern. This degree is as much theirs as it is mine.

Contents

Illustrations

I would like to dedicate this thesis to my grandfather who was instrumental in imparting a sense of curiosity in me, without which I might not be in this fortunate position.

Thank you *Tata*.

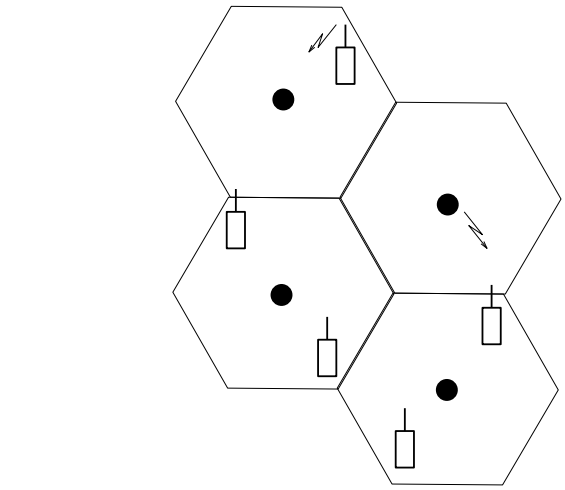
Chapter 1

Introduction

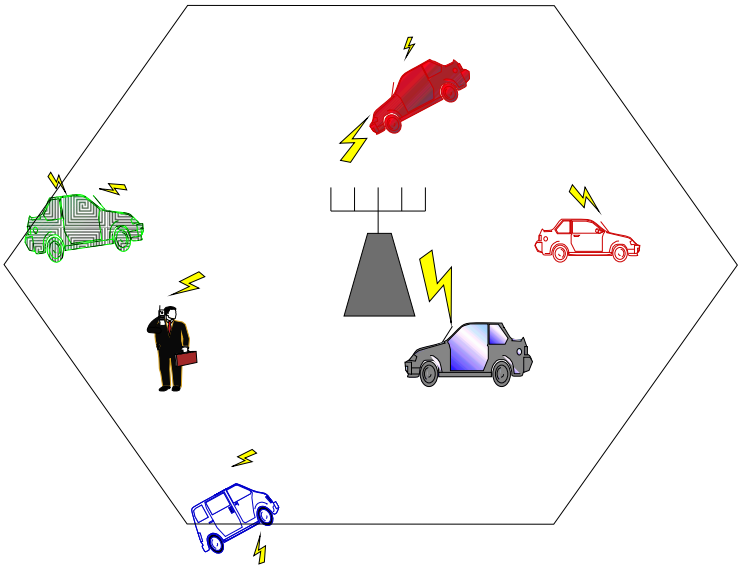
The concept of conveying information by successive choices from a finite alphabet is the essence of data communications. The rapid increase in the demand for information transfer in recent years necessitates increasingly reliable data transmission at ever increasing rates. An example of an emerging communications network is the mobile cellular phone or wireless communications network. Such a system is called a multiple access system because typically more than one user or subscriber is accessing the network at a given time. An immediate issue is the efficient distribution of available resources, typically bandwidth, amongst the different users [1,2].

In a typical cellular communication system a geographical region is divided into a number of cells, the coverage of each cell depending on the average traffic in the cell (based on experimental measurements). Figure 1.1(a) depicts a schematic of one such "honeycomb" of idealized hexagonal cells, with Figure 1.1(b) representing an individual cell. The mobiles in each cell (whether vehicle or human) communicate with a central processing unit called the base station. The base station in turn processes all incoming calls from the cell and redirects them to a *mobile trunk swtiching unit* (MTSN) which redirects the calls to other base stations or to the terrestrial wired network. The base station is also responsible for directing incoming calls from the MTSN to the individual mobiles as well as tracking the signal of a mobile in hand-off mode. In the realm of voice communications, a typical cell might be required to support 40 users at 9600 baud, each incurring a bit error rate of 10^{-3}.

There are three extant multiple access schemes: frequency division multiple access (FDMA), time division multiple access (TDMA) and code division multiple access



(a)



(b)

Figure 1.1 : (a) Idealized schematic of a cellular system depicting hexagonal cells with base stations. (b) Close-up of a particular cell in the above system. Each user communicates with the base station in its cell.

(CDMA). FDMA is the oldest method to date and uses the intuitively obvious approach of dividing the available bandwidth into frequency bins, one for each user. TDMA is a more recent scheme wherein each user has the entire bandwidth at its disposal, but is restricted to transmitting in fixed time slots. The number of users that can be supported is determined by the number of slots available, either in frequency or time. Both the above systems are inherently inflexible in that the number of frequency or time slots (or channels) allocated is fixed *a priori*; this leads to wastage of resources when the number of users is less than the number of channels.

CDMA is the most recent multiple access strategy and falls under a class of techniques called *spread spectrum* techniques [3, 4]. The term spread spectrum is used to signify the fact that the data stream is *modulated* or coded so that the overall transmission rate is much higher than what is needed to transmit the actual data. The basic idea is to replace a single data bit by a coded waveform that can only be detected when this code is known to the receiver. By utilizing intelligent encoding of the users' signals, each user is allowed the entire use of the time-frequency plane. We would ideally like to ensure that all the codes assigned are mutually orthogonal in order to be able to maximally differentiate amongst them. In practice however, this is not realistic due to the asynchronous manner of operation that CDMA supports, i.e., the users are not constrained to transmit at any particular times. This also allows for a larger number of possible codes with the overall system performance degrading gracefully as the number of users increases.

The two common modulation techniques for spread spectrum are direct sequence (DS) and frequency hopping (FH). Frequency hopping is an extension of the idea that a good way to prevent interference from unwanted signals is to move the desired signal around as much as possible in the frequency domain so that the operating frequency cannot be estimated. The desired user's signal is subjected to a predetermined pattern of "hopping" among certain frequencies. Direct sequence CDMA (DS-CDMA) signals are generated by multiplying the desired signal by a larger-bandwidth *spread-*

ing sequence. These spreading sequences, also called spreading codes, are also called pseudonoise sequences due to their resemblance to stochastic uncorrelated noise sequences. DS-CDMA schemes allow all the users in a system to transmit over the entire available bandwidth, achieving inter-user distinction through distinct spreading codes assigned to individual users. A feature of direct sequence techniques is the fine resolution in time that they offer, thereby allowing differentiation of multiple copies of signals that may arrive at different times at a receiver due to reflections from large objects nearby, also called *multipath* [5,6]. In addition, these sequences are designed to have low cross correlations amongst themselves so that the interference due to the other users, also termed *multiple access interference* (MAI), is kept as low as possible. The above reasons, along with reasonable costs of implementation, have led to direct sequence code division multiple access gaining considerable ground in the recent past as a viable protocol for commercial digital cellular communications. The emergence of the IS-95 standard for commercial cellular communications, has focused interest in deploying CDMA systems both in the cellular as well as personal communications arenas. Henceforth we will restrict our discussion of CDMA to direct sequence CDMA.

1.1 Motivation: The acquisition problem

The first stage in the demodulation of a CDMA signal is the so-called *acquisition* stage wherein the receiver attempts to lock onto the phase of the desired user's code. The communication link originating at the mobile and terminating at the base station, also called the uplink or reverse link, in a general wireless CDMA system is inherently asynchronous in nature, i.e., different signals arrive at the receiver with different relative time-offsets (for example, see Figure 1.2). The detection of a particular user's transmitted bits involves the correlation of the received waveform with a copy of the corresponding code at the receiver. Accurate correlation necessitates an accurate estimate of the user's timing offset. In the presence of multiple access

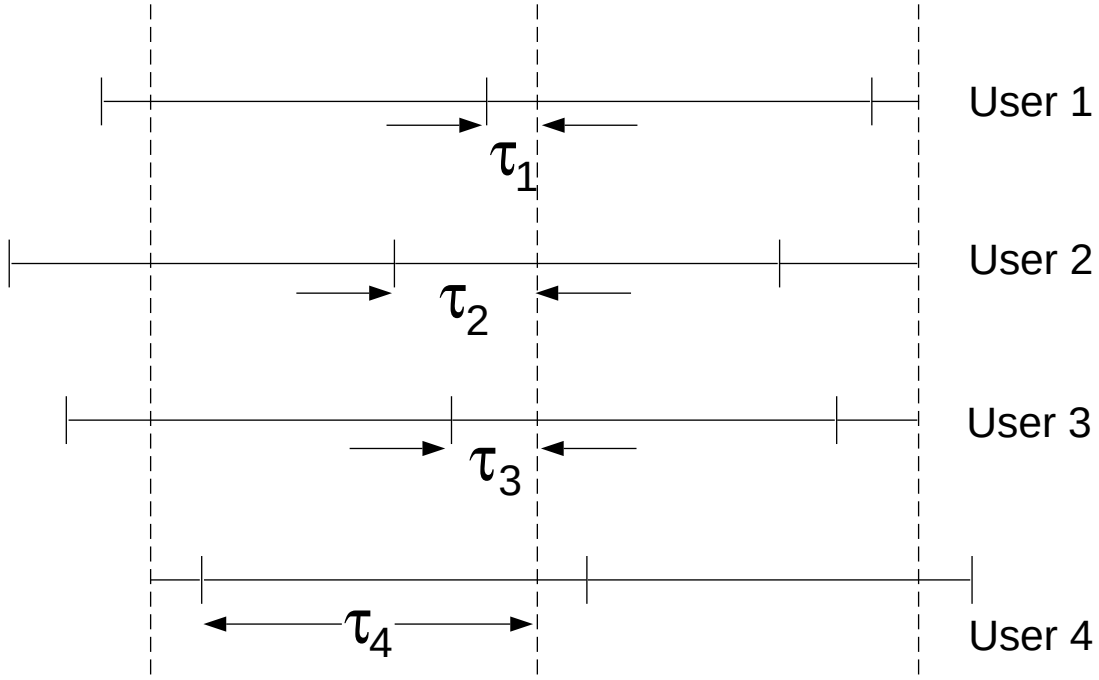


Figure 1.2 : A schematic of 2 successive bits each of four asynchronous users, with delays τ_1, τ_2, τ_3 and τ_4, arriving at a base station. The vertical dashed lines represent the base station's initial estimate of the bit boundaries.

interference (MAI) acquiring this timing or delay information is not a trivial task, especially since preferred codes have a very narrow and peaky autocorrelation structure (see Figure 1.3). Once an initial estimate of each user's delay is calculated, the next task is to track these delays over the subsequent incoming bits. The two stage process of acquiring the users' timings and tracking them as they change is called synchronization. In this paper we focus on the first stage, i.e, the acquisition phase. We must emphasize at this point that the tracking phase of the overall synchronization is an important problem in its own right and merits a separate study. This is because the channel impulse from the mobile to the base station is in constant flux due to the relative motion of the mobile and its surroundings — this contributes to constantly changing delays as well as occasionally varying *numbers* of delays.

Conventional CDMA systems treat the MAI as noise and therefore use single-

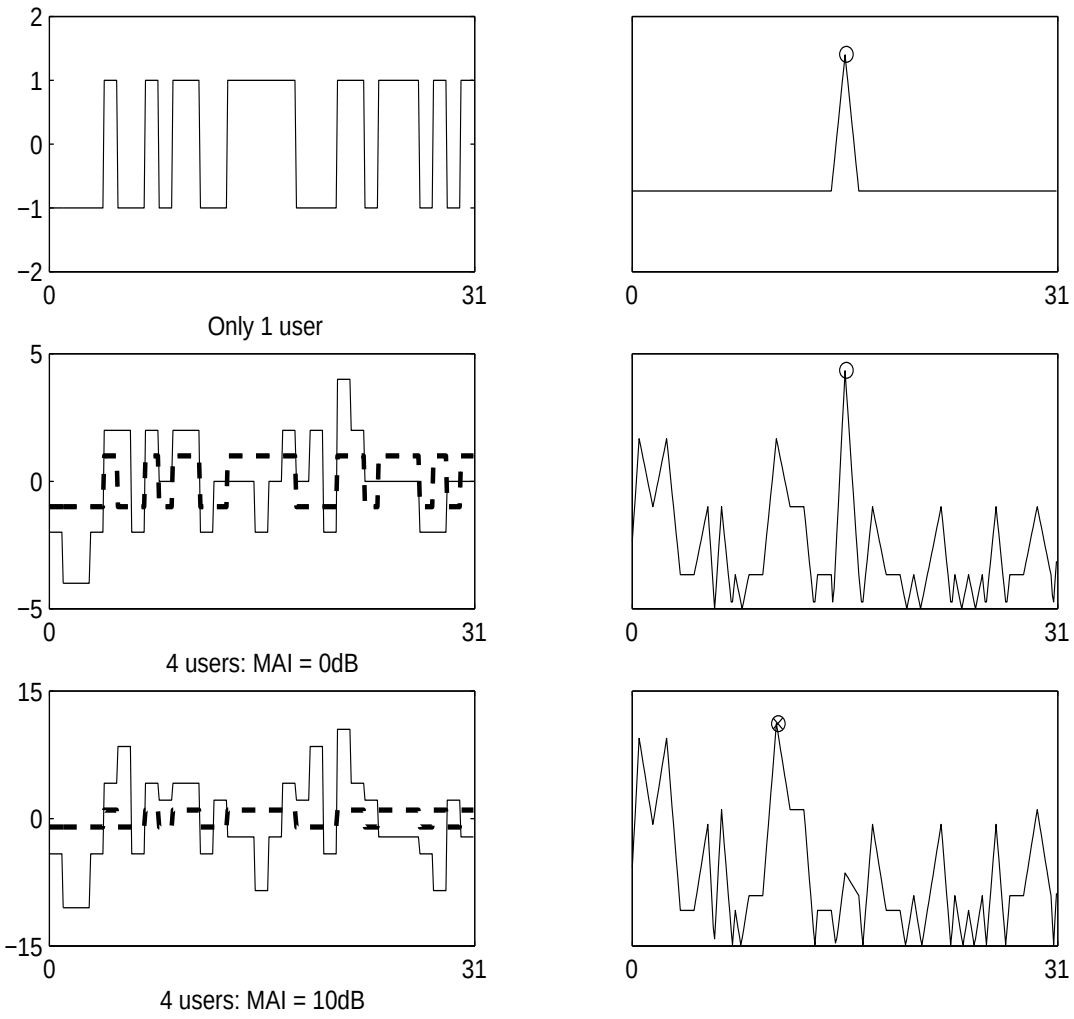


Figure 1.3 : The plot depicts the effects of MAI on the output of the correlator corresponding to the desired user's code. The left hand plots depict the received signal; the right hand plots depict the output of the sliding correlator corresponding to user 1, with the maximum denoted by a 'o'. The delay of user 1 corresponds to the argument of the maximum value. In the top plate we see only the desired user present and hence the timing is easy to estimate. In the middle plate we see that with mild interference it might still be possible to estimate the user's delay. The plates at the bottom represent 4 users with 10dB MAI and we see that the timing estimate is wrong. The dashed curve in the left-hand middle and bottom plots represents the desired user's code.

user schemes such as sliding correlators with sequential searching for acquisition purposes [7–10]. Therefore, they are subject to degradations incurred due to widely different power levels of different users — this is termed the *near-far problem*. Stringent power control is therefore required to avoid a deterioration in acquisition and tracking capabilities. Several joint parameter maximum likelihood techniques have been proposed to overcome the deficiencies of conventional schemes, but while these yield good results, they are computationally intensive [11–14]. An effective compromise was achieved in [15–17] by taking into account the structure of the signals arriving at the receiver. The common thread to these different approaches is the exploiting of the fact that the transmitted signals lie in a structured subspace. Estimating this subspace is crucial to the estimation of the desired parameters. A key drawback to all the above mentioned techniques which employ a single sensor at the base station receiver is that they are limited in the number of users' whose delays can be simultaneously estimated. We would like to be able to design receivers that alleviate this limitation while retaining good near-far propertes and involving a low computational complexity.

1.2 Focus of the Thesis

A main focus of this thesis is to increase the *acquisition-based capacity* [18], which is defined as the number of users whose delays can be simultaneously estimated. This is accomplished by incorporating spatial diversity into the acquisition algorithm through the means of an *antenna array* at the receiver. An antenna array [19] is a collection of antennas or sensors in a particular geometry, each receiving correlated versions of the same signal. This reception of correlated versions of the received signal at different sensors across the array increases the effective signal to noise ratio and gives rise to the spatial diversity that increases the acquisition-based capacity. The use of antenna arrays also enables the estimation of the angles of arrival, also referred to with some abuse of terminology as *directions of arrival* (DOAs), of the different users.

The estimation of the user's DOAs affords many immediate advantages [4,20,21]. In the context of cellular communications, direction of arrival information can be used for emergency call location and to aid soft handoff, where multiple base stations keep track of a particular mobile as it makes transitions across cell boundaries. The ability to "direction find" also leads to potential reduction in transmit power as the transmitters can direct their signals more accurately toward the intended receiver. This information can also be exploited through beamforming techniques to limit the MAI. In this thesis we develop a unified framework for simultaneously estimating the delays and DOAs of all transmitting users under radically different algorithms. This information is then incorporated in detector structures to limit the MAI, as in Chapter 5 (see also [22]), where we see that increasing the number of sensors dramatically improves detection performance.

1.3 Overview of the Thesis

Techniques exploiting the multidimensional benefits of antenna arrays for acquisition of spread spectrum signals have not received much attention in the literature; [21,23,24] are three papers that discuss such techniques. The studies in [21,23] introduce a reference based (involving training sequences) least squares technique for antenna array based synchronization. The multiple sensors are employed merely to increase the effective SNR and no attempt is made to estimate the DOA simultaneously. In addition, the spreading codes are assumed orthogonal and the arriving signals are assumed chip synchronous. This assumption, in the face of typical chip asynchronicity, can lead to a maximum loss of 3dB and an average loss of about 1.3dB. The structure of the interference is not exploited either. The technique proposed in [24] is an interesting maximum likelihood technique involving the transmission of training sequences, with no *a priori* assumption on the array structure or noise covariance matrix. As in the above mentioned techniques, only code delay estimation is dealt with (no DOA estimation), the interference structure is not exploited and the delays

are assumed chip synchronous (which is not generally the case).

In this dissertation we in turn first focus on a maximum likelihood approach that decouples the multiuser problem to a set of single user ones. The algorithm draws upon certain computationally elegant features of a maximum likelihood algorithm for DOA estimation presented in [25]. Our presentation generalizes the algorithm to the simultaneous estimation of DOA as well as code delay. The algorithm described here also assumes the transmission of training sequences by all the users that are being acquired. We are not constrained by any particular array structure and we make no assumption of chip synchronicity. The additive noise is assumed to be a circularly complex zero mean Gaussian random vector, but no *a priori* assumption is made on its covariance. We represent the additive noise component at each sensor as a sum of two components: one that is uncorrelated across the array (this could model receiver noise plus uncorrelated contributions due to background/other-cell propagation), and the other that is completely correlated across the array (this could model the contributions from the other-cell interference that are essentially seen as the same, sensor to sensor). With this noise model, we can express the theoretical noise covariance matrix as a Kronecker product of two smaller covariance matrices. In fact, this particular noise structure falls under a broader class of spatiotemporal correlations known as *separable* correlations [19]. This disjunction of space and time leads to a further decomposition of the joint DOA-delay estimation; we are now able to estimate all the users' DOAs and delays individually and separately. If we consider a widely used model for the array geometry, the uniform linear array (ULA), it can be shown [25–27] that DOA estimation can be reduced to a 2×2 eigenvalue problem. Estimating the code delay, as we will demonstrate, can be reduced to rooting N quadratic equations and determining the maximum of $3N$ values, where N is the length of each spreading code. It can be shown that the estimates of delays, DOAs and amplitudes are asymptotically unbiased [28]. We also derive Cramér-Rao bounds on the various mean-squared estimation errors. In the development of the algorithm

we will impose an uncorrelated assumption on the received signals, thereby limiting our discussion to single path channels. Multipath channels need to be handled via a modification in the algorithm which is not discussed here. The maximum likelihood algorithm is presented in Chapter 3.

The maximum likelihood algorithm described above requires the transmission of preambles or training sequences to facilitate the estimation process. In situations of high traffic the transmission of training sequences limits the data rate and hence there is a need to design *blind* estimation algorithms. These algorithms perform parameter estimation with no prior knowledge of the transmitted bits. In Chapter 4 we present a blind algorithm that only requires knowledge of the users' spreading codes. The algorithm is based on a subspace decomposition technique first introduced in [16, 17] for a single sensor receiver; we generalize this to a multisensor case. The received signal at the antenna-array is modeled as a linear combination of certain elementary signal vectors. These vectors that span a *signal subspace* are parameterized by a set of unknown parameters that we are trying to estimate, in this case, the timing and DOA. Generic subspace based methods estimate these parameters in either of two ways: by maximizing the projection of the signal vectors into the signal subspace, or by minimizing the projection of the signal vectors into the orthogonal *noise subspace*. These subspaces are in turn estimated using the eigenvalue decomposition (EVD) of the correlation matrix of the received signal (or the singular value decomposition (SVD) of the data matrix). In Chapter 4 of this thesis we use the MUSIC algorithm [19, 26, 27, 29, 30] to estimate the timing and DOA. We again assume a slowly fading single path model for each user; an understanding of this case is essential to the generalization to the multipath case, which is not treated here.

Once the transmitting users in a cell are synchronized, various detection strategies can be employed to estimate their transmitted bits. Receiver structures that take into account the interference structure are termed *multiuser receivers*. Many multiuser receivers have been designed, additionally, to combat the egregious effects of the

near-far problem [31–36]. In this context we investigate a class of multiuser detectors called linear multiuser detectors and examine the effects of multiple sensors on their performance. In particular, we focus on two members of the class of linear multiuser detectors, the decorrelating detector [33, 34] and the MMSE detector [37, 35]. Both of these suboptimal detectors have been shown to possess good near-far resistant properties, which make them attractive choices for cellular CDMA systems. It can be shown in these cases that the performance in terms of near-far resistance and probability of error is enhanced with the use of multiple sensors over a single sensor and this enhancement is proportional to the number of sensors in the antenna array. This is the focus of Chapter 5.

Chapter 2

Direct Sequence CDMA System Model

We assume a K-user narrow band direct sequence CDMA system with BPSK (Binary Phase Shift Keying) modulation with each transmitted signal selected from a binary alphabet and limited to $[0, T]$, where T is the symbol period. In general, each user transmits a zero mean stationary bit sequence with i.i.d components, and different users are independent of each other. In this development we will assume a single path channel and hence, all bitstreams are uncorrelated with each other.

The baseband representation of the k^{th} user's transmitted signal is given by

$$s_k(t) = \sqrt{2P_k} \sum_i b_{k,i} c_k(t - iT) \, , \tag{2.1}$$

where P_k is the transmitted power, $b_{k,i} \in \{+1, \ -1\}$ is the i^{th} transmitted bit and $c_k(t)$ is the spreading waveform. The spreading or code waveform is composed of N chips and if we assume BPSK for the spreading modulation we have

$$c_k(t) = \sum_{n=0}^{N-1} c_{k,n} \Pi(t - nT_c) \, , \tag{2.2}$$

where $c_{k,n} \in \{+1, \ -1\}$ and the chip pulse waveform $\Pi(t)$ is a rectangular pulse of duration T_c. We will assume that the extent of the spreading code is one bit period and hence we have $T = NT_c$. An example of a length-7 spreading code is given in Figure 2.1 with the chip pulse waveform $\Pi(t)$ shown also. Practical communication systems use more bandwidth efficient pulse shapes, such as raised cosines [1], for the code generation.

The front end of the receiver consists of an antenna array consisting of M sensors arranged in a specific geometry. We assume that all the transmitting sources are in

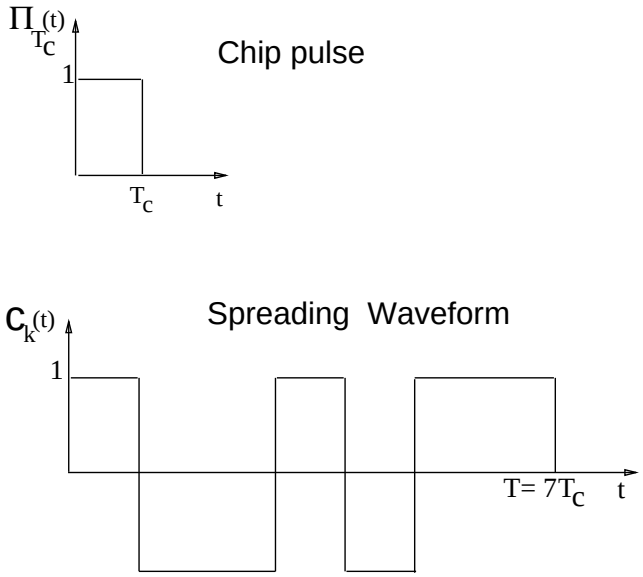
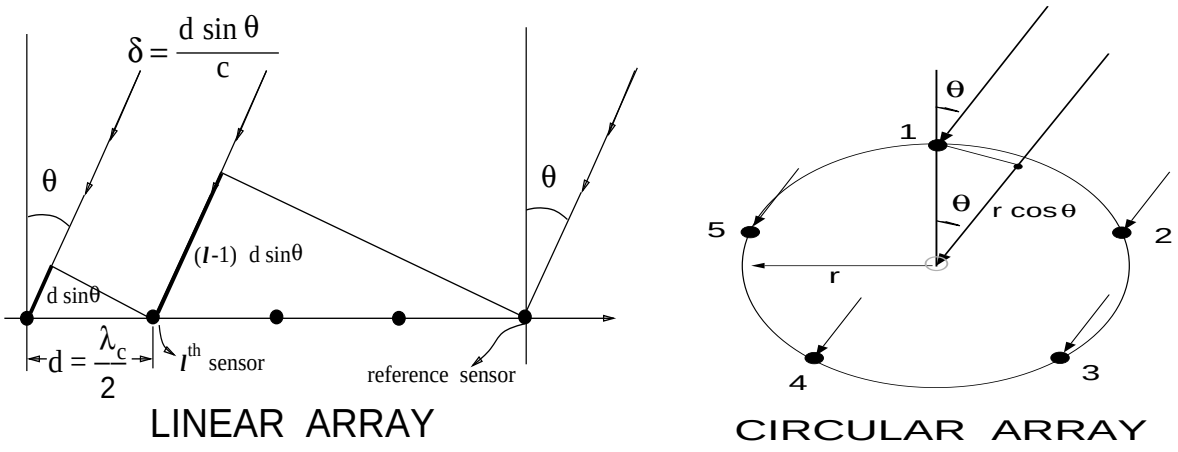


Figure 2.1 : A typical length-7 spreading code.

the array's *far field* [19], i.e., the wavefront emitted by a source has little curvature with respect to the arrays dimensions. In other words, the direction of propagation is approximately the same at each sensor. The array response vector, which is the response to a plane wave propagating from such a source and impinging on the array at an angle θ_k with the vertical, is

$$\mathbf{p}(\theta_k) = [p^{(1)}(\theta_k), \ldots, p^{(m)}(\theta_k), \ldots, p^{(M)}(\theta_k)]^\top \tag{2.3}$$

and is determined solely by the geometry of the array. For instance, in Figure 2.2 we see two array geometries depicted. The schematic on the left is a 5-element uniform linear array with intersensor spacing d. The array has one signal impinging on it from an angle θ. Since the transmitted signals are assumed narrowband, the signals at the different sensors are just phase-shifted versions of each other, with the phase shift depending on the differences in times of arrival of the signal at the different sensors. We designate the rightmost sensor to be the reference sensor. Then we see that the difference in times of arrival of the signal at the m^{th} and the reference sensors is given by $\tau^{(m)} = (m-1)d\sin\theta/c$, where c is the speed of light in free space. Now the array



(a) (b)

Figure 2.2 : Plot (a) depicts a 5-element uniform linear array (ULA) with intersensor spacing d. The reference sensor for the ULA is taken to be the extreme right one. Plot (b) depicts a 5-element uniform circular array (UCA) where the radius of the circle is r. The reference sensor in this case is taken to be a fictitious sensor located at the center of the array.

response vector from (2.3) can be written as

$$\mathbf{p}(\theta_k) = [1, \ \ldots, \ e^{j2\pi f_c(m-1)\delta \sin\theta_k}, \ \ldots, \ e^{j2\pi f_c(M-1)\delta \sin\theta_k}]^\top, \tag{2.4}$$

where $\delta = d/c$ and f_c represents the carrier frequency of the propagating wave. We typically assume that $d = \lambda_c/2$, with λ_c being the wavelength of the propagating wave, to avoid "spatial aliasing" [19], whence the m^{th} element of $\mathbf{p}(\theta_k)$ now becomes $\exp(j\pi \sin\theta_k)$ (in a broadband application we would require $d \leq \lambda_{min}/2$). If $d > \lambda_c/2$ we suffer from undersampling and if $d < \lambda_c/2$ we get oversampling. We note from (2.4) that $\mathbf{p}(\theta_k) = \mathbf{p}(\pi - \theta_k)$ leading to the so called *left-right ambiguity*. Schematic (b) in Figure 2.2 depicts a uniform circular array of radius r with 5 sensors. The reference is an imaginary sensor at the center of the array. It can be shown that the array response vector in this case is given by

$$\mathbf{p}(\theta_k) = [e^{j2\pi f_c\delta \cos\theta_k}, \ \ldots, \ e^{j2\pi f_c\delta \cos(2\pi(m-1)/M-\theta_k)}, \ \ldots, \ e^{j2\pi f_c\delta \cos(2\pi(M-1)/M-\theta_k)}]^\top, \tag{2.5}$$

where $\delta = r/c$. We see that there are no two angles $-\pi/2 \leq \theta_i, \theta_j \leq \pi/2$ such that $\mathbf{p}(\theta_i) = \mathbf{p}(\theta_j)$; however, we note that $\mathbf{p}(\theta_k) = \mathbf{p}^*(\pi + \theta_k)$ leading to indistinguishability between two opposing wavefronts. In addition, [38] demonstrate that the intersensor spacing can be several multiples of $\lambda_{min}/2$ without spatial aliasing. In the context of broadband array processing the use of a circular array over a ULA affords many advantages [38].

An *a priori* assumption on array geometry is not crucial to the development of the algorithms in this paper; however, given the narrowband realm of our system, we will consider a uniform linear array (ULA) to derive an elegant DOA estimation algorithm in Chapter 3 [26]. The narrow band assumption imposed on the transmitted signals seems to contradict their spread spectrum nature. However, we stress that it is only in the context of the antenna array that they are considered narrow band. This implies that the carrier frequency is much larger than the message bandwidth and for most reasonably sized antenna arrays, this also implies that the time taken to traverse physical array is much smaller than the inverse of the message bandwidth. Hence, the envelope characteristics of the signal do not vary across the array. For example, in a typical cellular CDMA communication system, the carrier frequency is 1GHz and the chip rate (message bandwidth) is 1MHz and we see that treating the system as narrowband is valid.

Each of the users transmits through a different time varying channel whose parameters we will assume are constant in the time taken to estimate them. Accordingly, the received signal at the m^{th} sensor is a superposition of attenuated and delayed signals transmitted by all the K users and is given by

$$r^{(m)}(t) = \sum_{k=1}^{K} w_k s(t - \tau_k)\, p^{(m)}(\theta_k) + \nu^{(m)}(t). \tag{2.6}$$

In the above equation the symbols have the following significance :

- w_k: Complex amplitude with which the k^{th} signal is received and includes contributions from the channel attenuation and the local oscillator phase offset

with respect to the transmitter.

- τ_k: Relative delay with respect to the transmitter, modulo T the bit period.
- θ_k: Direction of arrival (DOA) at the array with respect to an axis in the plane of the array (see Fig. 2.2 for an example).
- $\nu(t)$: Additive white Gaussian noise assumed to have some correlation across the array. We ascribe the following model to the noise $\nu(t) = \beta\eta(t) + (1-\beta)\eta^{(m)}(t)$. The component $\eta(t)$ is completely correlated across the array and can arise from other-cell interference — this component is taken to be the same at each sensor; $\eta^{(m)}(t)$ is the noise element independent from sensor to sensor, arising from a combination of receiver noise and the uncorrelated contributions from other-cell transmissions ($\beta \in [0, 1)$).

The noise components $\eta(t)$ and $\eta^{(m)}(t)$ are uncorrelated random processes and each is assumed to be white and Gaussian with zero-mean and double-sided spectral density of $\mathcal{N}_0/2$. In a typical cellular environment, the number of other-cell interferers at a base station is much larger than the number of users within the cell. Since these interferers are independent and bounded, we can invoke the central limit theorem [39] to model the other-cell interference as Gaussian.

The continuous time signal at the m^{th} sensor is discretized by sampling the output of a chip-matched filter, which is a simple integrate and dump operation due to the rectangular chip pulse, at the chip rate

$$r^{(m)}[n] = \frac{1}{T_c} \int_{nT_c}^{(n+1)T_c} r^{(m)}(t)dt \tag{2.7}$$

Since each spreading vector is periodic with period N, $r^{(m)}[n]$ is wide sense cyclo-stationary and observation vectors $\mathbf{r}_i^{(m)}$ formed by blocking together N successive outputs $r^{(m)}[n]$ are wide sense stationary. The observation vector, $\mathbf{r}_i^{(m)} \in \mathbb{C}^N$, at time

i is formed as

$$\mathbf{r}_i^{(m)} = \left[r^{(m)}[iN],\ r^{(m)}[iN+1],\ \ldots,\ r^{(m)}[iN+N-1]\right]^\top . \tag{2.8}$$

Since the system is asynchronous, each observation vector can be viewed as a linear combination of $2K$ *signal vectors* — 2 components from each user due to the past and current bits. We can now write $\mathbf{r}_i^{(m)}$ as

$$\begin{aligned} \mathbf{r}_i^{(m)} &= \sum_{k=1}^{K} w_k \left(b_{k,i-1}\mathbf{c}_k^R \ + \ b_{k,i}\mathbf{c}_k^L\right) p_k^{(m)} \ + \ \boldsymbol{\nu}_i^{(m)} \\ &= \mathbf{A}(\boldsymbol{\tau})\,\mathbf{W}\mathbf{b}_i \cdot p_k^{(m)} \ + \ \beta\boldsymbol{\eta}_i \ + \ (1-\beta)\boldsymbol{\eta}_i^{(m)} \ , \end{aligned} \tag{2.9}$$

where

- The diagonal matrix $\mathbf{W} \in \mathbb{C}^{2K\times 2K}$ of the form $\mathbf{W} = diag([w_1, w_1, \ldots, w_K, w_K])$, using Matlab notation.
- The N-dimensional complex circular gaussian noise vector $\boldsymbol{\nu}_i^{(m)} = \beta\boldsymbol{\eta}_i + (1-\beta)\boldsymbol{\eta}_i^{(m)}$, where $\boldsymbol{\eta}_i\ , \boldsymbol{\eta}_i^{(m)} \in \mathbb{C}^N$ and are distributed as $\mathsf{N}(\mathbf{0}, \sigma^2\mathbf{I})$ where $\sigma^2 = \frac{\mathcal{N}_0}{2T_c}$.
- We use the notation $p_k^{(m)}$ to denote $p^{(m)}(\theta_k)$*.
- The vectors $\mathbf{c}_k^R$ and $\mathbf{c}_k^L$ represent the *right* and *left* signal vectors for the k^{th} user and depend only on the delay τ_k and the spreading code.
- The $N \times 2K$ matrix $\mathbf{A}(\boldsymbol{\tau}) = \left[\mathbf{c}_1^R\ \mathbf{c}_1^L\ \ldots\ \mathbf{c}_K^R\ \mathbf{c}_K^L\right]$ where the parameter vector $\boldsymbol{\tau} = (\tau_1,\ \tau_2, \ldots\ \tau_K)$.
- The bit vector $\mathbf{b}_i = [b_{1,i-1},\ b_{1,i}\ , \ldots\ b_{K,i-1},\ b_{K,i}]^\top$.

Given that the k^{th} user's delay $\tau_k \in [0, NT_c)$, we can decompose it as follows: $\tau_k/T_c = q + \gamma$ where $q \in \{0, 1, \ldots, N-1\}$ and $\gamma \in [0, 1)$. The signal vectors

*Henceforth we will abbreviate $\mathbf{p}_k \equiv \mathbf{p}(\theta_k)$.

$\mathbf{c}_k^R$ and $\mathbf{c}_k^L$ can then be written as [15, 16]

$$\begin{aligned} \mathbf{c}_k^R(\tau_k) &= (1-\gamma)\,\mathbf{c}_k^R[q] \quad + \quad \gamma\,\mathbf{c}_k^R[q+1] \\ \mathbf{c}_k^L(\tau_k) &= (1-\gamma)\,\mathbf{c}_k^L[q] \quad + \quad \gamma\,\mathbf{c}_k^L[q+1]\,, \end{aligned} \tag{2.10}$$

where

$$\begin{aligned} \mathbf{c}_k^R[q] &= [c_{k,N-q} \ldots \; c_{k,N-1}\; 0 \ldots 0]^\top \\ \mathbf{c}_k^L[q] &= [0 \ldots 0 \; c_{k,0} \ldots \; c_{k,N-q-1}]^\top . \end{aligned} \tag{2.11}$$

An example of the left and right signal vectors at a given sensor for a code length

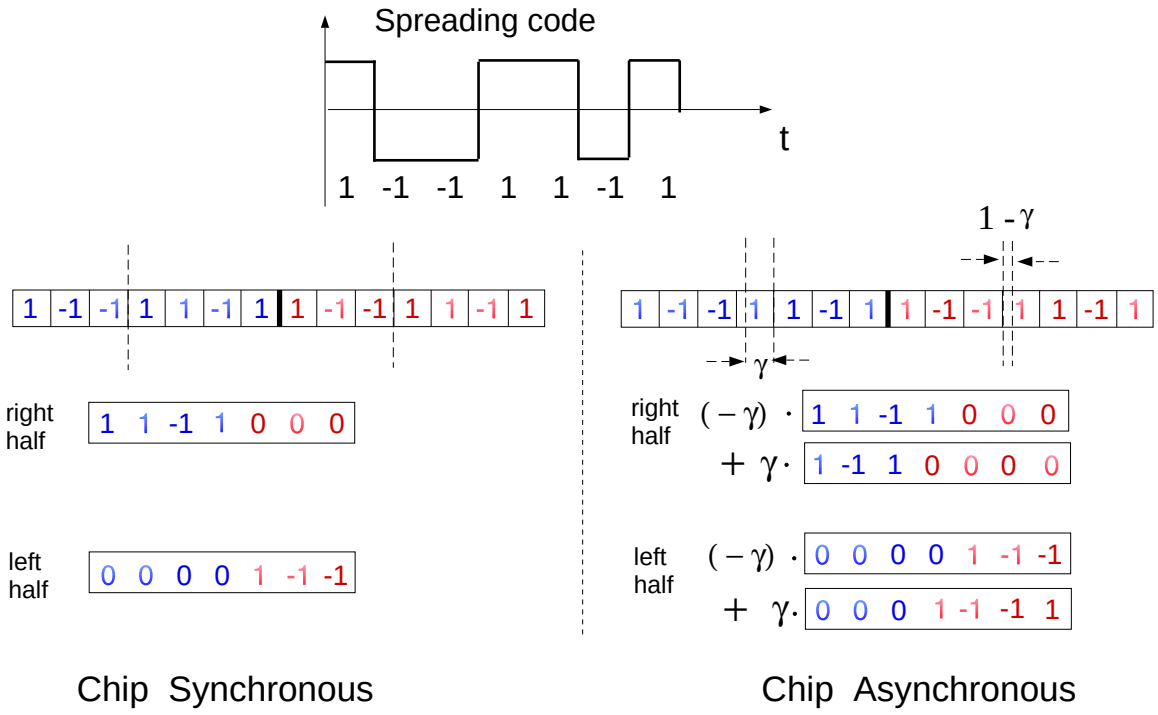


Figure 2.3 : Typical left and right signal vectors at one sensor for a length-7 spreading code. The left plot depicts the chip-synchronous case and the right plot depicts the chip-asynchronous case where the variable γ is used to denote the chip-asynchronism.

of $N = 7$ is shown in Figure 2.3. We use the notation $\mathbf{c}_k^R(\cdot), \mathbf{c}_k^L(\cdot)$ to denote continuous arguments and $\mathbf{c}_k^R[\cdot], \mathbf{c}_k^L[\cdot]$ to denote discrete arguments. Where an argument is omitted, it will be understood to be continuous.

The observation vector across the array, $\mathbf{r}_i \in \mathbb{C}^{MN}$, is now formed as a concatenation of individual sensor observation vectors as

$$\mathbf{r}_i = \left[\mathbf{r}_i^{(1)^\top}, \mathbf{r}_i^{(2)^\top}, \ldots, \mathbf{r}_i^{(M)^\top} \right]^\top . \tag{2.12}$$

The signal model for $\mathbf{r}_i$ in (2.12) can now be written as

$$\mathbf{r}_i = \mathcal{A}(\boldsymbol{\tau},\, \boldsymbol{\theta})\mathbf{W}\, \mathbf{b}_i \;+\; \boldsymbol{\nu}_i \,, \tag{2.13}$$

where $\boldsymbol{\theta} = (\theta_1,\, \theta_2, \ldots\, \theta_K)$,

$$\mathcal{A} = \left[\mathbf{a}_1^R \;\; \mathbf{a}_1^L, \ldots, \mathbf{a}_K^R \;\; \mathbf{a}_K^L \right] = \begin{pmatrix} \mathbf{A}\boldsymbol{\Phi}^{(1)} \\ \mathbf{A}\boldsymbol{\Phi}^{(2)} \\ \vdots \\ \mathbf{A}\boldsymbol{\Phi}^{(M)} \end{pmatrix}, \quad \boldsymbol{\Phi}^{(m)} = \begin{pmatrix} p_1^{(m)}\, \mathbf{I}_2 & & & \mathbf{0} \\ & p_2^{(m)}\, \mathbf{I}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & p_K^{(m)}\, \mathbf{I}_2 \end{pmatrix} \tag{2.14}$$

and

$$\boldsymbol{\nu}_i = \begin{pmatrix} \beta\boldsymbol{\eta}_i + (1-\beta)\boldsymbol{\eta}_i^{(1)} \\ \vdots \\ \beta\boldsymbol{\eta}_i + (1-\beta)\boldsymbol{\eta}_i^{(M)} \end{pmatrix} . \tag{2.15}$$

The $(2k-1)^{\text{th}}$ and $2k^{\text{th}}$ columns of $\mathcal{A}$, given by $\mathbf{a}_k^R$ and $\mathbf{a}_k^L$, are termed the k^{th} user's multisensor signal vectors and from (2.14) it is easy to see that

$$\mathbf{a}_k^R(\tau, \theta) = \mathbf{p}_k(\theta) \otimes \mathbf{c}_k^R(\tau) \,, \qquad \mathbf{a}_k^L(\tau, \theta) = \mathbf{p}_k(\theta) \otimes \mathbf{c}_k^L(\tau) \,, \tag{2.16}$$

where the symbol "$\otimes$" represents the Kronecker product defined as

$$\mathbf{C}^{mp \times nq} = \mathbf{A}^{m \times n} \otimes \mathbf{B}^{p \times q} = \begin{pmatrix} \mathbf{A}_{1,1}\mathbf{B} & \mathbf{A}_{1,2}\mathbf{B} & \cdots & \mathbf{A}_{1,n}\mathbf{B} \\ \mathbf{A}_{2,1}\mathbf{B} & \mathbf{A}_{2,2}\mathbf{B} & \cdots & \mathbf{A}_{2,n}\mathbf{B} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \mathbf{A}_{m,1}\mathbf{B} & \mathbf{A}_{m,2}\mathbf{B} & \cdots & \mathbf{A}_{m,n}\mathbf{B} \end{pmatrix} . \tag{2.17}$$

Thus, $\mathbf{a}_k^R$ and $\mathbf{a}_k^L$ depend on the unknown parameters τ_k and θ_k as well as on the known code and array structures.

The covariance matrix of the circular complex Gaussian noise vector, $\boldsymbol{\nu}_i \in \mathbb{C}^{MN}$, can be written as

$$\mathbf{K} = E[\boldsymbol{\nu}_i \, \boldsymbol{\nu}_i^H] = \mathbf{K}_1 \otimes \mathbf{K}_2 \, ,$$

where

$$\mathbf{K}_1^{M\times M} = \begin{pmatrix} \beta^2 + (1-\beta)^2 & \beta^2 & \cdots & \beta^2 \\ \beta^2 & \beta^2 + (1-\beta)^2 & \cdots & \beta^2 \\ \vdots & \vdots & \vdots & \vdots \\ \beta^2 & \beta^2 & \cdots & \beta^2 + (1-\beta)^2 \end{pmatrix} \text{ and } \mathbf{K}_2^{N\times N} = \sigma^2 \mathbf{I} \, . \tag{2.18}$$

We make particular note of the fact that the closer β is to 1, the more ill-conditioned $\mathbf{K}_1$ is, with $\beta = 1$ resulting in non-invertibility.

The estimation process involves obtaining estimates for the k^{th} user's delay (τ_k), direction of arrival (θ_k) as well as amplitude (w_k). This requires a knowledge of the codes of the different users as well as the geometry of the array, both of which are manifest in the signal vectors. In Chapter 3 we will discuss a maximum likelihood technique that exploits the Kronecker product structure of the signal vectors and the noise covariance matrix to obtain the required estimates for each user.

In Chapter 4 we adopt a subspace based approach to parameter estimation. The incoming observations are decomposed into two mutually orthogonal subspaces vis-a-vis the SVD of a data matrix. The structure of these subspaces is utilized to arrive at estimates for τ_k and θ_k.

Chapter 3

Maximum Likelihood Estimation of Delay and DOA

3.1 Maximum Likelihood Algorithm

In this chapter we will develop a maximum likelihood (ML) based technique for delay (τ_k), DOA (θ_k) and amplitude (w_k) estimation. The algorithm can be seen to generalize the ML technique for DOA estimation presented in [25] to the simultaneous estimation of delays and DOAs for the desired users. In a situation where the observation vector $\mathbf{r}_i$ depends on a parameter vector $\boldsymbol{\xi}$ that is either deterministic but unknown or whose a priori statistics are unknown, an estimate that is widely used is the maximum likelihood estimate of $\boldsymbol{\xi}$, given by [19]

$$\hat{\boldsymbol{\xi}}_{ML} = \arg\max_{\boldsymbol{\xi}} p(\mathbf{r}_i|\boldsymbol{\xi}). \tag{3.1}$$

In our setup, $\mathbf{r}_i$ is a function of the delays $\boldsymbol{\tau}$, the DOAs $\boldsymbol{\theta}$, the complex amplitudes $\mathbf{w} = (w_1, \ldots, w_K)$, the noise covariance matrix $\mathbf{K}$, which is assumed unknown and the transmitted bits $\mathbf{b}_i$. We assume that the bits $\mathbf{b}_i$ are known since, otherwise, the maximization of the likelihood function with respect to all the above unknowns is not a well posed problem. This is accomplished either by requiring that all the users being acquired transmit training sequences (acquisition phase) or by the receiver operating in a decision-directed mode (tracking phase).

Given L observations $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_L$ we see from (2.9) that they are not independent — $\mathbf{r}_i$ is correlated with $\mathbf{r}_{i+1}$, $\forall i$. This might raise doubts about the tractability of the joint density of the observations; however, we can see that $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_L$ are conditionally independent given $\mathbf{b} = [\mathbf{b}_0^\top, \mathbf{b}_1^\top, \ldots, \mathbf{b}_L^\top]^\top$, the transmitted bits. We

can now form the joint conditional probability density function of $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_L$ as

$$p(\mathbf{r}_1, \ldots, \mathbf{r}_L \mid \mathcal{A}, \mathbf{W}, \mathbf{b}) = \frac{1}{\pi^{MNL}|\mathbf{K}|^L} \exp\left\{ -\sum_{i=1}^{L} (\mathbf{r}_i - \mathcal{A}\mathbf{W}\mathbf{b}_i)^H \mathbf{K}^{-1} (\mathbf{r}_i - \mathcal{A}\mathbf{W}\mathbf{b}_i) \right\} \tag{3.2}$$

where $|\cdot|$ represents the determinant operator. The log-likelihood function can therefore be expressed as

$$\begin{aligned} \mathcal{L} &= -MN \ln \pi - \ln |\mathbf{K}| - \operatorname{tr}\left\{ \frac{1}{L} \sum_{i=1}^{L} (\mathbf{r}_i - \mathcal{A}\mathbf{W}\mathbf{b}_i)^H \mathbf{K}^{-1} (\mathbf{r}_i - \mathcal{A}\mathbf{W}\mathbf{b}_i) \right\} \\ &= -MN \ln \pi - \ln |\mathbf{K}| - \operatorname{tr}\left\{ \mathbf{K}^{-1} \cdot \frac{1}{L} \sum_{i=1}^{L} (\mathbf{r}_i - \mathcal{A}\mathbf{W}\mathbf{b}_i)(\mathbf{r}_i - \mathcal{A}\mathbf{W}\mathbf{b}_i)^H \right\} \end{aligned} \tag{3.3}$$

where $\operatorname{tr}(\cdot)$ represents the trace operator. The last equality follows from the following property of the trace: $\operatorname{tr}\{\mathbf{ABC}\} = \operatorname{tr}\{\mathbf{CAB}\} = \operatorname{tr}\{\mathbf{BCA}\}$.

We first maximize (3.3) with respect to $\mathbf{K}$ by using an easily verifiable fact ([19]) :

Lemma 3.1

If $\mathbf{B}$ *is a positive definite Hermitian matrix, then* $f(\mathbf{K}) = -\ln|\mathbf{K}| - \operatorname{tr}\{\mathbf{K}^{-1}\mathbf{B}\}$ *is maximized by* $\widehat{\mathbf{K}} = \mathbf{B}$.

Proof : The proof of the above Lemma follows from the expressions for the derivative of $\ln|\mathbf{K}|$ and $\operatorname{tr}\{\mathbf{K}^{-1}\mathbf{B}\}$ with respect to the matrix $\mathbf{K}$

$$\nabla_{\mathbf{K}} \ln|\mathbf{K}| = \mathbf{K}^{-\top}, \qquad \nabla_{\mathbf{K}} \operatorname{tr}\{\mathbf{K}^{-1}\mathbf{B}\} = \mathbf{K}^{-1}\mathbf{B}\mathbf{K}^{-\top}.$$

■

Thus, the maximum likelihood estimate of $\mathbf{K}$ is

$$\widehat{\mathbf{K}}(\mathcal{A}, \mathbf{W}) = \frac{1}{L} \sum_{i=1}^{L} (\mathbf{r}_i - \mathcal{A}\mathbf{W}\mathbf{b}_i)(\mathbf{r}_i - \mathcal{A}\mathbf{W}\mathbf{b}_i)^H, \tag{3.4}$$

Substituting this into $\mathcal{L}$, we obtain the effective consolidated log likelihood to be $\mathcal{L} = -\ln|\widehat{\mathbf{K}}|$, the maximization of which is equivalent to the minimization of

$$\mathcal{L}_1 = |\widehat{\mathbf{K}}(\mathcal{A}, \mathbf{W})| = \left| \frac{1}{L} \sum_{i=1}^{L} (\mathbf{r}_i - \mathcal{A}\mathbf{W}\mathbf{b}_i)(\mathbf{r}_i - \mathcal{A}\mathbf{W}\mathbf{b}_i)^H \right|. \tag{3.5}$$

Direct maximization of $\mathcal{L}_1$ in (3.5) with respect to $\{\boldsymbol{\tau}, \boldsymbol{\theta}, \mathbf{w}\}$ is rather intractable and hence we adopt an indirect approach. We notice from (3.5) that $\mathcal{L}_1$ is a function of $\{\boldsymbol{\tau}, \boldsymbol{\theta}, \mathbf{w}\}$ only through the product $\mathcal{A}(\boldsymbol{\tau}, \boldsymbol{\theta})\mathbf{W}(\mathbf{w})$. We can therefore very conveniently capture the dependence of $\mathcal{L}_1$ on all the unknown parameters through a single matrix $\mathcal{Y} \in \mathbb{C}^{MN \times 2K}$, where $\mathcal{Y} = \mathcal{A}\mathbf{W}$. The received observations can now be written as

$$\mathbf{r}_i = \mathcal{Y}\mathbf{b}_i + \boldsymbol{\nu}_i. \tag{3.6}$$

The ML estimation procedure is now carried out in two steps:

(i) We first form the unconstrained ML estimate of $\mathcal{Y}$, given by $\widehat{\mathcal{Y}}$. This is akin to a minimum mean square estimate of $\mathcal{Y}$.

(ii) Having obtained $\widehat{\mathcal{Y}}$, we obtain the estimates $\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{w}}$ by minimizing the weighted least squares fit between $\mathcal{Y}$ and its unstructured estimate $\widehat{\mathcal{Y}}$.

En route to obtaining $\widehat{\mathcal{Y}}$ we define the following sample correlation matrices

$$\widehat{\mathbf{R}}_{rr} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{r}_i \mathbf{r}_i^H\,; \quad \widehat{\mathbf{R}}_{rb} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{r}_i \mathbf{b}_i^H\,; \quad \widehat{\mathbf{R}}_{bb} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{b}_i \mathbf{b}_i^H\,. \tag{3.7}$$

The problem now reduces to

$$\widehat{\mathcal{Y}} = \arg\min_{\mathcal{Y}} \mathcal{L}_2 = \arg\min_{\mathcal{Y}} \left| \widehat{\mathbf{R}}_{rr} - \mathcal{Y}\widehat{\mathbf{R}}_{rb}^H - \widehat{\mathbf{R}}_{rb}\mathcal{Y}^H + \mathcal{Y}\widehat{\mathbf{R}}_{bb}\mathcal{Y}^H \right|. \tag{3.8}$$

Motivated by [35] we establish the following lemma :

Lemma 3.2

If the matrices $\mathbf{P}, \mathbf{C} \in \mathbb{C}^{MN \times 2K}$ *and* $\mathbf{R} \in \mathbb{C}^{MN \times MN}$ *is a positive definite matrix, then for any positive definite Hermitian matrix* $\mathbf{Q} \in \mathbb{C}^{2K \times 2K}$,

$$\arg\min_{\mathbf{C}} |F(\mathbf{C})| = \arg\min_{\mathbf{C}} \left| \mathbf{R} - \mathbf{P}\mathbf{C}^H - \mathbf{C}\mathbf{P}^H + \mathbf{C}\mathbf{Q}\mathbf{C}^H \right| = \mathbf{P}\mathbf{Q}^{-1}. \tag{3.9}$$

Proof: For any matrix $\mathbf{Z} \in \mathbb{C}^{MN \times 2K}$ we can see that

$$\begin{aligned} F(\mathbf{PQ^{-1}} + \mathbf{Z}) &= \mathbf{R} - \mathbf{PQ^{-1}P}^H + \mathbf{ZQZ}^H \\ &= F(\mathbf{PQ^{-1}}) + \mathbf{ZQZ}^H. \end{aligned}$$

Without loss of generality we can assume that $F(\mathbf{C})$ is a positive definite matrix function of $\mathbf{C}$ (this is seen from the fact that $F(\cdot)$ can be thought of as originating from a form similar to $\widehat{\mathbf{K}}(\mathcal{Y})$ in (3.4)), $F(\mathbf{PQ^{-1}}) = \mathbf{R} - \mathbf{PQ^{-1}P}^H$ is a positive definite matrix. Therefore $F(\mathbf{PQ^{-1}})$ can be written as $\mathbf{CC}^H$, with $\mathbf{C}$ invertible and we have

$$\begin{aligned} \left|F(\mathbf{PQ^{-1}} + \mathbf{Z})\right| &= \left|\mathbf{CC}^H + \mathbf{ZQZ}^H\right| \\ &= \left|\left|\mathbf{C}\right|\right|^2 \cdot \left|\mathbf{I} + \mathbf{C^{-1}ZQZ}^H\mathbf{C}^{-H}\right| \\ &= \left|F(\mathbf{PQ^{-1}})\right| \cdot \left|\mathbf{I} + \mathbf{C^{-1}ZQZ}^H\mathbf{C}^{-H}\right|. \end{aligned} \tag{3.10}$$

It then follows from the positive definiteness of $\mathbf{C^{-1}ZQZ}^H\mathbf{C}^{-H}$ that

$$\arg\min_{\mathbf{C}} \left|F(\mathbf{PQ^{-1}} + \mathbf{Z})\right| = \mathbf{PQ^{-1}},$$

because any nonzero $\mathbf{Z}$ causes an increase in the product of the eigenvalues of $F(\mathbf{PQ^{-1}} + \mathbf{Z})$. Hence the Lemma. ■

From (3.9) we see that

$$\widehat{\mathcal{Y}} = \widehat{\mathbf{R}}_{rb}\widehat{\mathbf{R}}_{bb}^{-1}, \tag{3.11}$$

which yields

$$\widehat{\mathbf{K}} = \widehat{\mathbf{R}}_{rr} - \widehat{\mathbf{R}}_{rb}\widehat{\mathbf{R}}_{bb}^{-1}\widehat{\mathbf{R}}_{rb}^H. \tag{3.12}$$

From (3.11) and (3.12), we see that the unconstrained estimates are derived completely from the received observations and a knowledge of the training sequences. From (3.8) and (3.11) we see that calculating $\widehat{\mathcal{Y}}$ amounts to forming a minimum mean squared estimate of $\mathcal{Y}$ with $\widehat{\mathcal{Y}}$ being that estimate and $\widehat{\mathbf{K}}$ being the residual minimum mean squared error.

By making use of the expressions for $\widehat{\mathbf{K}}$ and $\widehat{\mathcal{Y}}$, (3.5) can be rewritten as

$$\begin{aligned}\mathcal{L}_1 &= \left|\widehat{\mathbf{R}}_{rr} - \mathcal{Y}\widehat{\mathbf{R}}_{rb}^H - \widehat{\mathbf{R}}_{rb}\mathcal{Y}^H + \mathcal{Y}\widehat{\mathbf{R}}_{bb}\mathcal{Y}^H\right| \\ &= \left|\widehat{\mathbf{R}}_{rr} - \widehat{\mathbf{R}}_{rb}\widehat{\mathbf{R}}_{bb}^{-1}\widehat{\mathbf{R}}_{rb}^H + \widehat{\mathcal{Y}}\widehat{\mathbf{R}}_{bb}\widehat{\mathcal{Y}}^H - \mathcal{Y}\widehat{\mathbf{R}}_{rb}^H - \widehat{\mathbf{R}}_{rb}\mathcal{Y}^H + \mathcal{Y}\widehat{\mathbf{R}}_{bb}\mathcal{Y}^H\right| \qquad (3.13) \\ &= \left|\widehat{\mathbf{K}}\right| \cdot \left|\mathbf{I} + \widehat{\mathbf{K}}^{-1}(\mathcal{Y} - \widehat{\mathcal{Y}})\widehat{\mathbf{R}}_{bb}(\mathcal{Y} - \widehat{\mathcal{Y}})^H\right| .\end{aligned}$$

We note that minimizing $\left|\mathbf{I} + \widehat{\mathbf{K}}^{-1}(\mathcal{Y} - \widehat{\mathcal{Y}})\widehat{\mathbf{R}}_{bb}(\mathcal{Y} - \widehat{\mathcal{Y}})^H\right|$ is equivalent to minimizing $\ln\left|\mathbf{I} + \widehat{\mathbf{K}}^{-1}(\mathcal{Y} - \widehat{\mathcal{Y}})\widehat{\mathbf{R}}_{bb}(\mathcal{Y} - \widehat{\mathcal{Y}})^H\right|$. Using properties of eigenvalues and norms of matrices, we show in Appendix A that

$$\min_{\boldsymbol{\tau},\boldsymbol{\theta},\mathbf{W}} \ln\left|\mathbf{I} + \widehat{\mathbf{K}}^{-1}(\mathcal{Y} - \widehat{\mathcal{Y}})\widehat{\mathbf{R}}_{bb}(\mathcal{Y} - \widehat{\mathcal{Y}})^H\right| \stackrel{\equiv}{\longrightarrow} \min_{\boldsymbol{\tau},\boldsymbol{\theta},\mathbf{W}} \mathrm{tr}\{\widehat{\mathbf{R}}_{bb}(\mathcal{Y} - \widehat{\mathcal{Y}})^H\widehat{\mathbf{K}}^{-1}(\mathcal{Y} - \widehat{\mathcal{Y}})\}, \qquad (3.14)$$

where the symbol $\stackrel{\equiv}{\longrightarrow}$ denotes "is asymptotically equivalent to". Intuitively, this is interpreted in the following manner : Since $\widehat{\mathcal{Y}}$ is a consistent estimate of $\mathcal{Y}$, the difference $(\mathcal{Y} - \widehat{\mathcal{Y}}) \longrightarrow \mathbf{0}$ in the number of observations L. Invoking a familiar example from the scalar case, $\ln(1 + x) \approx x$ for small x, we see that the above approximation is not unreasonable (and this is proved in Appendix A). This greatly simplifies the problem since it replaces the nonlinear functional $|\cdot|$ by a linear one $\mathrm{tr}\{\cdot\}$. The problem now reduces to

$$\{\hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{w}}\} = \arg\min_{\boldsymbol{\tau},\boldsymbol{\theta},\mathbf{w}} \mathrm{tr}\left\{\widehat{\mathbf{R}}_{bb}(\mathcal{Y} - \widehat{\mathcal{Y}})^H\widehat{\mathbf{K}}^{-1}(\mathcal{Y} - \widehat{\mathcal{Y}})\right\} . \qquad (3.15)$$

Since the different users' training sequences are chosen independent of one another and since we are investigating a single path channel, for large L the sample correlation matrix $\widehat{\mathbf{R}}_{bb}$ is diagonal in structure. The multidimensional optimization problem in (3.15) now dramatically reduces to a 3-dimensional problem for each user k,

$$\{\hat{\tau}_k, \hat{\theta}_k, \hat{w}_k\} = \arg\min_{\tau_k,\theta_k,w_k} \mathcal{L}_3 = \arg\min_{\tau_k,\theta_k,w_k} \Big[(\mathbf{y}_{2k-1} - \hat{\mathbf{y}}_{2k-1})^H\widehat{\mathbf{K}}^{-1}(\mathbf{y}_{2k-1} - \hat{\mathbf{y}}_{2k-1}) + (\mathbf{y}_{2k} - \hat{\mathbf{y}}_{2k})^H\widehat{\mathbf{K}}^{-1}(\mathbf{y}_{2k} - \hat{\mathbf{y}}_{2k})\Big] , \qquad (3.16)$$

where $\mathbf{y}_m = (\mathcal{Y})_m$, $\hat{\mathbf{y}}_m = (\widehat{\mathcal{Y}})_m$ and $(\mathbf{H})_m$ represents the m^{th} column of the matrix $\mathbf{H}$. In (3.16) the two terms account for the right and left portions of the k^{th} user's spreading code. It should be mentioned that if the channel introduced multipath, $\widehat{\mathbf{R}}_{bb}$ would be block diagonal and so, while the users would still be decoupled, the function to be minimized in (3.16) would involve a greater number of terms.

From the definition of $\mathcal{Y}$ we see that

$$\mathbf{y}_{2k-1} = w_k \mathbf{a}_k^R(\tau_k, \theta_k), \quad \mathbf{y}_{2k} = w_k \mathbf{a}_k^L(\tau_k, \theta_k)\,. \tag{3.17}$$

Since $\mathbf{y}_{2k-1}$ and $\mathbf{y}_{2k}$ are linear functions of w_k, we can obtain calculate $\hat{w}_k$, by simply setting $\partial \mathcal{L}_3 / \partial \mathrm{w}_k^* = 0$, where $*$ represents complex conjugation. With $\mathbf{y}_{2k-1}, \mathbf{y}_{2k}$ defined in (3.17), $\hat{w}_k$ is obtained as

$$\hat{w}_k = \frac{\hat{\mathbf{y}}_{2k-1}^H \widehat{\mathbf{K}}^{-1} \hat{\mathbf{a}}_k^R + \hat{\mathbf{y}}_{2k}^H \widehat{\mathbf{K}}^{-1} \hat{\mathbf{a}}_k^L}{\hat{\mathbf{a}}_k^{R^H} \widehat{\mathbf{K}}^{-1} \hat{\mathbf{a}}_k^R + \hat{\mathbf{a}}_k^{L^H} \widehat{\mathbf{K}}^{-1} \hat{\mathbf{a}}_k^L}\,, \tag{3.18}$$

where $\hat{\mathbf{a}}_k^{\{R/L\}} \triangleq \mathbf{a}_k^{\{R/L\}}(\hat{\tau}_k, \hat{\theta}_k)$.

If we substitute for $\hat{w}_k$ from (3.18) in (3.16), it can be shown that $\hat{\tau}_k, \hat{\theta}_k$ can be calculated as

$$\begin{aligned} \{\hat{\tau}_k, \hat{\theta}_k\} &= \arg \max_{\tau_k, \theta_k} \ \mathcal{L}_4 \\ &= \arg \max_{\tau_k, \theta_k} \frac{\left| \hat{\mathbf{y}}_{2k-1}^H \widehat{\mathbf{K}}^{-1} \mathbf{a}_k^R + \hat{\mathbf{y}}_{2k}^H \widehat{\mathbf{K}}^{-1} \mathbf{a}_k^L \right|^2}{\mathbf{a}_k^{R^H} \widehat{\mathbf{K}}^{-1} \mathbf{a}_k^R + \mathbf{a}_k^{L^H} \widehat{\mathbf{K}}^{-1} \mathbf{a}_k^L}\,. \end{aligned} \tag{3.19}$$

The problem in (3.19) can be interpreted geometrically as a subspace fitting problem. To see this, let us define

$$\begin{aligned} \mathbf{d}_k^R &\triangleq \widehat{\mathbf{K}}^{-1/2} \mathbf{a}_k^R\,, \qquad & \hat{\mathbf{z}}_{2k-1} &\triangleq \widehat{\mathbf{K}}^{-1/2} \hat{\mathbf{y}}_{2k-1} \\ \mathbf{d}_k^L &\triangleq \widehat{\mathbf{K}}^{-1/2} \mathbf{a}_k^L\,, \qquad & \hat{\mathbf{z}}_{2k} &\triangleq \widehat{\mathbf{K}}^{-1/2} \hat{\mathbf{y}}_{2k}\,. \end{aligned}$$

If we further define the $(2MN \times 1)$ matrices

$$\mathbf{D}_k = \left[\mathbf{d}_k^{R^H} \ \ \mathbf{d}_k^{L^H} \right]^H \text{ and } \mathbf{Z}_k = \left[\hat{\mathbf{z}}_{2k-1}^H \ \ \hat{\mathbf{z}}_{2k}^H \right]^H\,, \tag{3.20}$$

the maximization in (3.19) can now be recast as

$$\begin{aligned} \hat{\tau}_k, \hat{\theta}_k &= \arg \max_{\tau_k, \theta_k} \ \mathrm{tr} \left\{ \Pi_{\mathbf{D}_k}(\tau_k, \theta_k) \ \mathbf{Z}_k \mathbf{Z}_k^H \right\} \\ &\equiv \arg \min_{\tau_k, \theta_k, c} \ \| \mathbf{Z}_k - c \, \mathbf{D}_k(\tau_k, \theta_k) \|_F^2 , \end{aligned} \tag{3.21}$$

where $\Pi_{\mathbf{D}_k} = \mathbf{D}_k (\mathbf{D}_k^H \mathbf{D}_k)^{-1} \mathbf{D}_k^H$ is the projection matrix onto the column space of $\mathbf{D}_k$ and the complex number $c = (\mathbf{D}_k^H \mathbf{D}_k)^{-1} \mathbf{D}_k^H \mathbf{Z}_k$ [29]. We see that (3.21) highlights the one-dimensional subspace fit between the transformed data, embodied in $\mathbf{Z}_k$, and the transformed signal vector, embodied in $\mathbf{D}_k$ (see Figure 3.1 for a geometrical schematic of the 1-d subspace fit). Specifically, we choose those parameters $\{\tau_k, \theta_k\}$ so as to maximize the fit between the two one-dimensional subspaces as shown in (3.21).

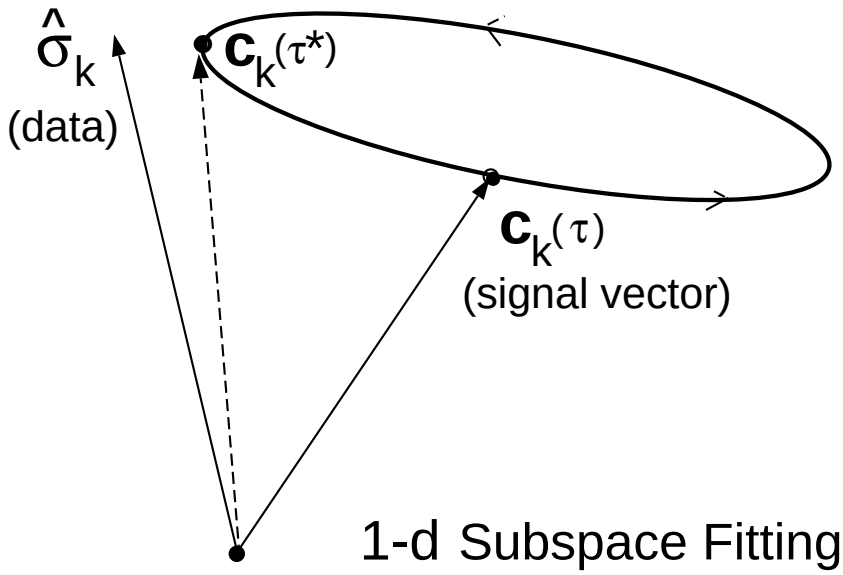


Figure 3.1 : The figure depicts a schematic representation of the 1-dimensional subspace fit, as a function of $\tau \in [0, T)$ and $\theta \in [-\pi/3, \pi/3)$, that the ML estimation reduces to. The vector $\mathbf{Z}_k$ represents the transformed data vector and the vector $\mathbf{D}_k(\tau, \theta)$ represents the transformed signal vector as given in (3.20). For different values of τ and θ, $\mathbf{D}_k(\tau, \theta)$ defines a two dimensional manifold, simply represented by the shaded ellipse. The parameters $\{\tau^*, \theta^*\}$ represent the optimum pair, i.e., $\mathbf{D}_k(\tau^*, \theta^*)$ represents the point in the ellipse that is closest to $\mathbf{Z}_k$.

It is clear that (3.19) describes a 2-dimensional optimization problem which could

still be computationally burdensome. We would ideally like to be able to decouple the delay and DOA estimation problems. We recall from (2.18) and (3.17) that $\mathbf{K}$, the true covariance matrix, and $\mathbf{Y}_k \stackrel{\triangle}{=} [\mathbf{y}_{2k-1} \;\; \mathbf{y}_{2k}]$ are endowed with a Kronecker product structure. It can also be shown [40] that $\widehat{\mathbf{K}}$ and $\widehat{\mathcal{Y}}$ are consistent estimates (because they are maximum likelihood estimates), i.e., $\widehat{\mathbf{K}} \stackrel{L}{\longrightarrow} \mathbf{K}$, $\widehat{\mathcal{Y}} \stackrel{L}{\longrightarrow} \mathcal{Y}$ where $\stackrel{L}{\longrightarrow}$ denotes the asymptotic limit in L, the number of observations. Therefore if we can approximate $\widehat{\mathbf{K}}$ and $\widehat{\mathbf{Y}}_k = [\hat{\mathbf{y}}_{2k-1} \;\; \hat{\mathbf{y}}_{2k}]$ as

$$\widehat{\mathbf{K}} \approx \widehat{\mathbf{K}}_1 \otimes \widehat{\mathbf{K}}_2 \qquad \text{and} \qquad \widehat{\mathbf{Y}}_k \approx \boldsymbol{\rho}_k \otimes \boldsymbol{\sigma}_k \,, \tag{3.22}$$

where $\widehat{\mathbf{K}}_1 \in \mathbb{C}^{M\times M}$, $\widehat{\mathbf{K}}_2 \in \mathbb{C}^{N\times N}$, $\boldsymbol{\rho}_k \in \mathbb{C}^{M\times 1}$ and $\boldsymbol{\sigma}_k \in \mathbb{C}^{N\times 2}$ with $\boldsymbol{\sigma}_k = [\boldsymbol{\sigma}_{k1} \;\; \boldsymbol{\sigma}_{k2}]$, then we expect that

$$\widehat{\mathbf{K}}_1 \stackrel{L}{\longrightarrow} \alpha_1 \mathbf{K}_1 \qquad\qquad \boldsymbol{\rho}_k \stackrel{L}{\longrightarrow} \alpha_3 \mathbf{p}_k$$
$$\widehat{\mathbf{K}}_2 \stackrel{L}{\longrightarrow} \alpha_2 \mathbf{K}_2 \qquad\qquad [\boldsymbol{\sigma}_{k1} \;\; \boldsymbol{\sigma}_{k2}] \stackrel{L}{\longrightarrow} \alpha_4 \left[\mathbf{c}_k^R \;\; \mathbf{c}_k^L\right]$$

where $\alpha_1, \ldots, \alpha_4$ are scalars and don't affect the maximization. The approximation ("$\approx$" in (3.22)) is in the Frobenius norm sense, i.e., for example

$$\{\widehat{\mathbf{K}}_1, \widehat{\mathbf{K}}_2\} = \arg \min_{\substack{\mathbf{Q}_1 \in \mathbb{C}^{M\times M} \\ \mathbf{Q}_2 \in \mathbb{C}^{N\times N}}} \left\| \widehat{\mathbf{K}} - \mathbf{Q}_1 \otimes \mathbf{Q}_2 \right\|_F^2 . \tag{3.23}$$

It is shown in Appendix B that the above approximation amounts to calculating the largest left singular vector and right singular vector of an $M^2 \times N^2$ matrix (the reader is also referred to [56]).

The function $\mathcal{L}_4$ defined in (3.19) can now be reduced to

$$\mathcal{L}_4 \approx \frac{\left| (\boldsymbol{\rho}_k \otimes \boldsymbol{\sigma}_{k1})^H \left(\widehat{\mathbf{K}}_1^{-1} \otimes \widehat{\mathbf{K}}_2^{-1}\right) (\mathbf{p}_k \otimes \mathbf{c}_k^R) + (\boldsymbol{\rho}_k \otimes \boldsymbol{\sigma}_{k2})^H \left(\widehat{\mathbf{K}}_1^{-1} \otimes \widehat{\mathbf{K}}_2^{-1}\right) (\mathbf{p}_k \otimes \mathbf{c}_k^L) \right|^2}{(\mathbf{p}_k \otimes \mathbf{c}_k^R)^H \left(\widehat{\mathbf{K}}_1^{-1} \otimes \widehat{\mathbf{K}}_2^{-1}\right) (\mathbf{p}_k \otimes \mathbf{c}_k^R) + (\mathbf{p}_k \otimes \mathbf{c}_k^L)^H \left(\widehat{\mathbf{K}}_1^{-1} \otimes \widehat{\mathbf{K}}_2^{-1}\right) (\mathbf{p}_k \otimes \mathbf{c}_k^L)}$$

$$= \underbrace{\frac{\left| \boldsymbol{\rho}_k^H \widehat{\mathbf{K}}_1^{-1} \mathbf{p}_k \right|^2}{\mathbf{p}_k^H \widehat{\mathbf{K}}_1^{-1} \mathbf{p}_k} \cdot \frac{\left| \boldsymbol{\sigma}_{k1}^H \widehat{\mathbf{K}}_2^{-1} \mathbf{c}_k^R + \boldsymbol{\sigma}_{k2}^H \widehat{\mathbf{K}}_2^{-1} \mathbf{c}_k^L \right|^2}{\mathbf{c}_k^{RH} \widehat{\mathbf{K}}_2^{-1} \mathbf{c}_k^R + \mathbf{c}_k^{LH} \widehat{\mathbf{K}}_2^{-1} \mathbf{c}_k^L}}_{\mathcal{L}_5} \tag{3.24}$$

where we have made use of the following properties of the Kronecker product [41] :

Fact 3.1

1. *If* $\mathbf{A}\,, \mathbf{B}$ *are invertible and* $\mathbf{Y} = \mathbf{A} \otimes \mathbf{B}$ *then* $\mathbf{Y}^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$.

2. $(\mathbf{C} \otimes \mathbf{D}) \cdot (\mathbf{G} \otimes \mathbf{H}) = \mathbf{CG} \otimes \mathbf{DH}$ *if the matrix dimensions are appropriate.*

3. *If* $\alpha_1\,, \alpha_2$ *are scalars,* $\alpha_1 \otimes \alpha_2 = \alpha_1 \cdot \alpha_2$.

The delay and DOA subparts thus decouple and we can now deal with the one dimensional optimizations separately. If we relabel the two parts of $\mathcal{L}_5$ as

$$\begin{aligned} F_1(\theta) &= \frac{\left|\mathbf{p}_k^H(\theta)\,\widehat{\mathbf{K}}_1^{-1}\boldsymbol{\rho}_k\right|^2}{\mathbf{p}_k^H(\theta)\,\widehat{\mathbf{K}}_1^{-1}\mathbf{p}_k(\theta)} \\ &= \boldsymbol{\rho}_k^H\widehat{\mathbf{K}}_1^{-1/2}\left\{(\widehat{\mathbf{K}}_1^{-1/2}\mathbf{p}_k)\left(\mathbf{p}_k^H\widehat{\mathbf{K}}_1^{-1}\mathbf{p}_k\right)^{-1}(\widehat{\mathbf{K}}_1^{-1/2}\mathbf{p}_k)^H\right\}\widehat{\mathbf{K}}_1^{-1/2}\boldsymbol{\rho}_k \\ F_2(\tau) &= \frac{\left|\boldsymbol{\sigma}_{k1}^H\widehat{\mathbf{K}}_2^{-1}\mathbf{c}_k^R(\tau) + \boldsymbol{\sigma}_{k2}^H\widehat{\mathbf{K}}_2^{-1}\mathbf{c}_k^L(\tau)\right|^2}{\mathbf{c}_k^{R\,H}(\tau)\widehat{\mathbf{K}}_2^{-1}\mathbf{c}_k^R(\tau) + \mathbf{c}_k^{L\,H}(\tau)\widehat{\mathbf{K}}_2^{-1}\mathbf{c}_k^L(\tau)}\,, \end{aligned} \tag{3.25}$$

we are faced with the two maximization problems : $\max_\theta F_1$, $\max_\tau F_2$. We see that these individual problems have a form similar to that in (3.21) and hence, can be posed as subspace fitting problems. The maximization of $F_1(\theta)$ in (3.25) in case of a ULA can be carried out through a non-iterative process as discussed in [25, 26, 42] and we briefly present it here for completeness without delving into great detail.

Estimation of θ_k

Recall that the array response vector for a M-sensor ULA with half wavelength spacing is given by

$$\mathbf{p}(\theta_k) = \left[1, e^{j\phi_k}, e^{j\,2\phi_k}, \ldots, e^{j\,(M-1)\phi_k}\right]^\top,$$

where $\phi_k = \pi \sin\theta_k$. Our main aim is to reparametrize the function $F_1(\theta)$ so as to avoid a search in the θ parameter. This is achieved through the coefficients of a first

order complex polynomial of the form

$$b_{k,0} + b_{k,1} z = b_{k,1}(z - e^{j\phi_k}) . \tag{3.26}$$

The coefficients $b_{k,0}$ and $b_{k,1}$ are configured such that $e^{j\phi_k}$ is a root of the polynomial. Therefore, given $b_{k,0}$ and $b_{k,1}$, θ_k is easily found as

$$\theta_k = \sin^{-1}\left(\frac{1}{\pi}\arg\left(-\frac{b_{k,0}}{b_{k,1}}\right)\right). \tag{3.27}$$

If we now define an $(M \times M-1)$ Toeplitz matrix by

$$\mathbf{B}^H = \begin{pmatrix} b_{k,0} & b_{k,1} & \dots & 0 \\ 0 & b_{k,0} & b_{k,1} & \dots \\ & \ddots & \ddots & \\ 0 & & b_{k,0} & b_{k,1} \end{pmatrix}, \tag{3.28}$$

we see that $\mathbf{B}$ is of rank $M-1$ and $\mathbf{B}^H \mathbf{p}_k = \mathbf{0}$ (because of the way $b_{k,0}$ and $b_{k,1}$ are defined). Therefore $\widehat{\mathbf{K}}_1^{1/2}\mathbf{B}$ is also of rank $M-1$ and

$$\left(\widehat{\mathbf{K}}_1^{1/2}\mathbf{B}\right)^H \left(\widehat{\mathbf{K}}_1^{-1/2}\mathbf{p}_k\right) = \mathbf{0},$$

which together imply that the columns of $\widehat{\mathbf{K}}_1^{1/2}\mathbf{B}$ span the null space of $\widehat{\mathbf{K}}_1^{-1/2}\mathbf{p}_k$, i.e.,

$$\left(\widehat{\mathbf{K}}_1^{-1/2}\mathbf{p}_k\right)\left(\mathbf{p}_k^H\widehat{\mathbf{K}}_1^{-1}\mathbf{p}_k\right)^{-1}\left(\widehat{\mathbf{K}}_1^{-1/2}\mathbf{p}_k\right)^H = \widehat{\mathbf{K}}_1^{1/2}\mathbf{B}(\mathbf{B}^H\widehat{\mathbf{K}}_1\mathbf{B})^{-1}\mathbf{B}^H\widehat{\mathbf{K}}_1^{1/2}.$$

Therefore the maximization of $F_1(\theta)$ can be equivalently posed as

$$\begin{aligned} &\max_{\theta}\ \boldsymbol{\rho}_k^H\,\widehat{\mathbf{K}}_1^{-1/2}\left(\widehat{\mathbf{K}}_1^{-1/2}\mathbf{p}_k\right)\left(\mathbf{p}_k^H\widehat{\mathbf{K}}_1^{-1}\mathbf{p}_k\right)^{-1}\left(\widehat{\mathbf{K}}_1^{-1/2}\mathbf{p}_k\right)^H\widehat{\mathbf{K}}_1^{-1/2}\boldsymbol{\rho}_k \\ \equiv &\min_{\mathbf{b}}\ \boldsymbol{\rho}_k^H\mathbf{B}(\mathbf{B}^H\widehat{\mathbf{K}}_1\mathbf{B})^{-1}\mathbf{B}^H\boldsymbol{\rho}_k\,, \end{aligned} \tag{3.29}$$

where $\mathbf{b} = [b_{k,0}\ \ b_{k,1}]^\top$. It can be shown [25, 42] that the above minimization can be carried out in two steps:

1. First perform the minimization

$$\min_{\mathbf{b}}\ \boldsymbol{\rho}_k^H\mathbf{B}\mathbf{B}^H\boldsymbol{\rho}_k. \tag{3.30}$$

We can ignore the contribution due to $(\mathbf{B}^H\widehat{\mathbf{K}}_1\mathbf{B})^{-1}$ to obtain a first estimate of $b_{k,0}, b_{k,1}$ because $\boldsymbol{\rho}_k$ is a consistent estimate of $\mathbf{p}_k$ and hence $\mathbf{B}^H\boldsymbol{\rho}_k \xrightarrow{L} \mathbf{0}$. This implies that the estimate of $b_{k,0}, b_{k,1}$ obtained in this step is also consistent.

2. Replace the entries of $\mathbf{b}$ in $(\mathbf{B}^H\widehat{\mathbf{K}}_1\mathbf{B})^{-1}$ by the consistent estimates obtained in (3.30). Carry out the minimization in (3.29) with $(\mathbf{B}^H\widehat{\mathbf{K}}_1\mathbf{B})^{-1}$ replaced by it's consistent estimate.

The minimization in the above two steps can be further simplified by noting that

$$\mathbf{B}^H\boldsymbol{\rho}_k = \tilde{\boldsymbol{\rho}}^H\mathbf{b}\,,$$

where the matrix $\tilde{\boldsymbol{\rho}}$ defined as

$$\tilde{\boldsymbol{\rho}}^H = \begin{pmatrix} \rho_{k,2} & \rho_{k,1} \\ \vdots & \vdots \\ \rho_{k,M} & \rho_{k,M-1} \end{pmatrix} \tag{3.31}$$

can be thought of as a "data" matrix. In the above formulation $\rho_{k,i}$ is the i^{th} element of the $(M \times 1)$ vector $\boldsymbol{\rho}_k$. Now the problem in (3.29) can be reduced to the following equivalent problem

$$\min_{\mathbf{b}} \; \mathbf{b}^H\tilde{\boldsymbol{\rho}}\;\widetilde{\mathbf{B}}\;\tilde{\boldsymbol{\rho}}^H\mathbf{b}, \tag{3.32}$$

where $\widetilde{\mathbf{B}} = \mathbf{I}$ or $\widetilde{\mathbf{B}} = (\mathbf{B}^H\widehat{\mathbf{K}}_1\mathbf{B})^{-1}$ (depending on which of the above two steps are being carried out) with $\mathbf{B}$ obtained from the first step in the two-stage minimization process described above. We impose further constraints on $\mathbf{b}$ by taking into account the form of the solution in (3.26) [42]. These conditions are captured by

$$b_{k,1} = b_{k,0}^* \quad \text{and} \quad (\text{Re}\{b_{k,0}\})^2 + (\text{Im}\{b_{k,0}\})^2 = 1 \tag{3.33}$$

The two conditions in (3.33) imposed on $\mathbf{b}$ reduce the optimization problem in (3.32) to a 2×2 eigenvalue problem in $[\text{Re}\{b_{k,0}\},\ \text{Im}\{b_{k,0}\}]^\top$. Thus we see that $\hat{\theta}_k$ can be obtained through a non-iterative process and involves the solution to a 2×2 eigenvalue problem.

Estimation of τ_k

We will now inspect the maximization of F_2 more carefully. The objective function F_2 is not a continuously differentiable function of τ [15]; rather, it is *piecewise* differentiable over each chip period. If we write $\tau/T_c = q + \gamma$ where $q \in \{0, 1, \ldots, N-1\}$ and $\gamma \in [0, 1)$, we remember from (2.10) that

$$\begin{aligned} \mathbf{c}_k^R(\tau) &= (1-\gamma)\,\mathbf{c}_k^R[q] \;+\; \gamma\,\mathbf{c}_k^R[q+1] \\ \mathbf{c}_k^L(\tau) &= (1-\gamma)\,\mathbf{c}_k^L[q] \;+\; \gamma\,\mathbf{c}_k^L[q+1]\,. \end{aligned}$$

Therefore, over the chip period $[qT_c, (q+1)T_c)$, we see that

$$\frac{d}{d\tau}\mathbf{c}_k^{R/L} = \frac{d}{d\gamma}\mathbf{c}_k^{R/L} \;\triangleq\; \boldsymbol{\zeta}_k^{R/L}[q] = \;\mathbf{c}_k^{R/L}[q+1] \;-\; \mathbf{c}_k^{R/L}[q]\,. \tag{3.34}$$

The function we are trying to maximize is

$$\begin{aligned} F_2(\tau) &= \frac{\left|\boldsymbol{\sigma}_{k1}^H \widehat{\mathbf{K}}_2^{-1}\mathbf{c}_k^R + \boldsymbol{\sigma}_{k2}^H \widehat{\mathbf{K}}_2^{-1}\mathbf{c}_k^L\right|^2}{\mathbf{c}_k^{R^H}\widehat{\mathbf{K}}_2^{-1}\mathbf{c}_k^R + \mathbf{c}_k^{L^H}\widehat{\mathbf{K}}_2^{-1}\mathbf{c}_k^L} \\ &= \frac{|\alpha_1(q)|^2\,\gamma^2 + \alpha_{12}(q)\,\gamma + |\alpha_2(q)|^2}{\beta_1(q)\,\gamma^2 + \beta_2(q)\,\gamma + \beta_3(q)}\,, \end{aligned} \tag{3.35}$$

where $\alpha_1\,, \alpha_{12}\,, \alpha_2\,, \beta_1\,, \beta_2\,, \beta_3$ are defined as

$$\begin{aligned} \alpha_1(q) &\triangleq \;\boldsymbol{\sigma}_{k1}^H\,\widehat{\mathbf{K}}_2^{-1}\,\boldsymbol{\zeta}_k^R[q] \;+\; \boldsymbol{\sigma}_{k2}^H\,\widehat{\mathbf{K}}_2^{-1}\,\boldsymbol{\zeta}_k^L[q] \\ \alpha_2(q) &\triangleq \;\boldsymbol{\sigma}_{k1}^H\,\widehat{\mathbf{K}}_2^{-1}\,\mathbf{c}_k^R\,[q] \;+\; \boldsymbol{\sigma}_{k2}^H\,\widehat{\mathbf{K}}_2^{-1}\,\mathbf{c}_k^L\,[q] \\ \alpha_{12}(q) &\triangleq \;2\mathrm{Re}\,(\alpha_1\alpha_2^*) \end{aligned}$$

and

$$\begin{aligned} \beta_1(q) &\triangleq \;\boldsymbol{\zeta}_k^{R^H}[q]\,\widehat{\mathbf{K}}_2^{-1}\,\boldsymbol{\zeta}_k^R[q] \;+\; \boldsymbol{\zeta}_k^{L^H}[q]\,\widehat{\mathbf{K}}_2^{-1}\,\boldsymbol{\zeta}_k^L[q] \\ \beta_2(q) &\triangleq \;2\mathrm{Re}\left(\boldsymbol{\zeta}_k^{R^H}[q]\,\widehat{\mathbf{K}}_2^{-1}\,\mathbf{c}_k^R\,[q] \;+\; \boldsymbol{\zeta}_k^{L^H}[q]\,\widehat{\mathbf{K}}_2^{-1}\,\mathbf{c}_k^L\,[q]\right) \\ \beta_3(q) &\triangleq \;\mathbf{c}_k^{R^H}[q]\,\widehat{\mathbf{K}}_2^{-1}\,\mathbf{c}_k^R\,[q] \;+\; \mathbf{c}_k^{L^H}[q]\,\widehat{\mathbf{K}}_2^{-1}\,\mathbf{c}_k^L\,[q] \end{aligned} \tag{3.36}$$

Differentiation of (3.35) with respect to γ yields the optimum argument, $\gamma^* \in [0,1)$ as one of the roots of the following quadratic equation

$$\left(|\alpha_1|^2\beta_2 - \alpha_{12}\beta_1\right)\gamma^2 + 2\left(|\alpha_1|^2\beta_3 - |\alpha_2|^2\beta_1\right)\gamma + \left(\alpha_{12}\beta_3 - |\alpha_2|^2\beta_2\right) = 0\,. \qquad (3.37)$$

We now differentiate $F_2(\tau)$ over N chip periods and calculate the $2N$ roots. Since each interval is half-open, the maximum in each $[q,(q+1))$ can either occur at one of the two roots in the interval, or at the left end-point. We now select that $\hat{\tau} = (\hat{q}+\hat{\gamma})T_c$ of these $3N$ candidates that maximizes $F_2(\hat{\tau})$.

Consistency of estimates

It is straightforward to show that the estimates of τ_k, θ_k and w_k are consistent. Let us denote the estimates of the true parameters $\{\tau_k, \theta_k, w_k\}$ by $\{\hat{\tau}_k, \hat{\theta}_k, \hat{w}_k\}$. Since $\hat{\mathbf{y}}_{2k-1}$ and $\hat{\mathbf{y}}_{2k}$ are consistent estimates, it follows that

$$\hat{\mathbf{y}}_{2k-1} \xrightarrow{L} w_k\,\mathbf{a}_k^R(\tau_k,\theta_k) \;\text{ and }\; \hat{\mathbf{y}}_{2k} \xrightarrow{L} w_k\,\mathbf{a}_k^L(\tau_k,\theta_k)\,. \qquad (3.38)$$

The consistency in the estimate of w_k now follows directly from (3.38) and (3.18). Examining (3.19) and using (3.38), we see that for large L we can express $\mathcal{L}_4$ as

$$\mathcal{L}_4 = |w_k|^2\,\frac{\left|\tilde{\mathbf{a}}_k^H(\tau_k,\theta_k)\,\widetilde{\widehat{\mathbf{K}}}^{-1}\,\tilde{\mathbf{a}}_k(\hat{\tau}_k,\hat{\theta}_k)\right|^2}{\tilde{\mathbf{a}}_k^H(\hat{\tau}_k,\hat{\theta}_k)\,\widetilde{\widehat{\mathbf{K}}}^{-1}\,\tilde{\mathbf{a}}_k(\hat{\tau}_k,\hat{\theta}_k)}\,, \qquad (3.39)$$

where

$$\tilde{\mathbf{a}}_k = \begin{pmatrix}\mathbf{a}_k^R\\ \mathbf{a}_k^L\end{pmatrix} \quad\text{and}\quad \widetilde{\widehat{\mathbf{K}}}^{-1} = \begin{pmatrix}\widehat{\mathbf{K}}^{-1} & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{K}}^{-1}\end{pmatrix}.$$

Applying the Cauchy-Schwartz Lemma to (3.39) we obtain

$$\frac{\left|\tilde{\mathbf{a}}_k^H(\tau_k,\theta_k)\,\widetilde{\widehat{\mathbf{K}}}^{-1}\,\tilde{\mathbf{a}}_k(\hat{\tau}_k,\hat{\theta}_k)\right|^2}{\tilde{\mathbf{a}}_k^H(\hat{\tau}_k,\hat{\theta}_k)\,\widetilde{\widehat{\mathbf{K}}}^{-1}\,\tilde{\mathbf{a}}_k(\hat{\tau}_k,\hat{\theta}_k)} \;\le\; \tilde{\mathbf{a}}_k^H(\tau_k,\theta_k)\,\widetilde{\widehat{\mathbf{K}}}^{-1}\,\tilde{\mathbf{a}}_k(\tau_k,\theta_k), \qquad (3.40)$$

with equality iff $\tilde{\mathbf{a}}_k(\hat{\tau}_k, \hat{\theta}_k) = \tilde{\mathbf{a}}_k(\tau_k, \theta_k)$. We can easily choose array configurations and spreading codes such that two signal vectors $\tilde{\mathbf{a}}_k(\tau_1, \theta_1) = \tilde{\mathbf{a}}_k(\tau_2, \theta_2)$ iff $\tau_1 = \tau_2$ and $\theta_1 = \theta_2$ (most reasonable array configurations $\mathbf{p}(\theta)$ and spreading codes $\mathbf{c}(\tau)$ satisfy this property [28]; the reader is also referred to Chapter 4 for a further clarification of the above idea). Since we are interested in maximizing $\mathcal{L}_4$, by invoking the above fact we see from (3.40) that $\mathcal{L}_4$ is maximized only when $\{\hat{\tau}_k, \hat{\theta}_k\} = \{\tau_k, \theta_k\}$. This yields consistency of the delay and DOA estimates as well.

3.2 Analysis of Performance — Cramér-Rao Bound

In this section we calculate lower bounds on the mean squared errors of the parameters estimated in Section 3.1 via the Cramér-Rao bound. Since the estimates are unbiased [25,30], this amounts to calculating the inverse of the Fisher information matrix.

We can calculate the Cramér-Rao bound in terms of the following quantities :

$$\Delta_1 \stackrel{\Delta}{=} \left(\mathcal{A}^{R^H}\mathbf{K}^{-1}\mathcal{A}^R \;+\; \mathcal{A}^{L^H}\mathbf{K}^{-1}\mathcal{A}^L\right) \odot \widehat{\mathbf{R}}_{bb}^{\top} \tag{3.41}$$

$$\Delta_2 \stackrel{\Delta}{=} \left(\mathbf{\Gamma}^{R^H}\mathbf{K}^{-1}\mathbf{\Gamma}^R \;+\; \mathbf{\Gamma}^{L^H}\mathbf{K}^{-1}\mathbf{\Gamma}^L\right) \odot \left(\mathbf{W}\widehat{\mathbf{R}}_{bb}\mathbf{W}^H\right)^{\top} \tag{3.42}$$

$$\Delta_3 \stackrel{\Delta}{=} \left(\Theta^{R^H}\mathbf{K}^{-1}\Theta^R \;+\; \Theta^{L^H}\mathbf{K}^{-1}\Theta^L\right) \odot \left(\mathbf{W}\widehat{\mathbf{R}}_{bb}\mathbf{W}^H\right)^{\top} \tag{3.43}$$

$$\mathbf{\Upsilon}_1 \stackrel{\Delta}{=} \left(\mathbf{\Gamma}^{R^H}\mathbf{K}^{-1}\mathcal{A}^R \;+\; \mathbf{\Gamma}^{L^H}\mathbf{K}^{-1}\mathcal{A}^L\right) \odot \left(\widehat{\mathbf{R}}_{bb}\mathbf{W}^H\right)^{\top} \tag{3.44}$$

$$\mathbf{\Upsilon}_2 \stackrel{\Delta}{=} \left(\Theta^{R^H}\mathbf{K}^{-1}\mathcal{A}^R \;+\; \Theta^{L^H}\mathbf{K}^{-1}\mathcal{A}^L\right) \odot \left(\widehat{\mathbf{R}}_{bb}\mathbf{W}^H\right)^{\top} \tag{3.45}$$

$$\mathbf{\Upsilon}_3 \stackrel{\Delta}{=} \left(\mathbf{\Gamma}^{R^H}\mathbf{K}^{-1}\Theta^R \;+\; \mathbf{\Gamma}^{L^H}\mathbf{K}^{-1}\Theta^L\right) \odot \left(\mathbf{W}\widehat{\mathbf{R}}_{bb}\mathbf{W}^H\right)^{\top} \tag{3.46}$$

where

$$\begin{aligned}
\mathcal{A}^R &\stackrel{\Delta}{=} \left[\mathbf{a}_1^R,\ \ldots,\mathbf{a}_K^R\right] = \left[\mathbf{p}_1\otimes\mathbf{c}_1^R,\ \ldots,\mathbf{p}_K\otimes\mathbf{c}_K^R\right] \\
\mathbf{\Gamma}^R &\stackrel{\Delta}{=} \left[\frac{\partial\mathbf{a}_1^R}{\partial\gamma_1},\ \ldots,\frac{\partial\mathbf{a}_K^R}{\partial\gamma_K}\right] = \left[\mathbf{p}_1\otimes\frac{\partial\mathbf{c}_1^R}{\partial\gamma_1},\ \ldots,\mathbf{p}_K\otimes\frac{\partial\mathbf{c}_K^R}{\partial\gamma_K}\right] \\
\Theta^R &\stackrel{\Delta}{=} \left[\frac{\partial\mathbf{a}_1^R}{\partial\theta_1},\ \ldots,\frac{\partial\mathbf{a}_K^R}{\partial\theta_K}\right] = \left[\frac{\partial\mathbf{p}_1}{\partial\theta_1}\otimes\mathbf{c}_1^R,\ \ldots,\frac{\partial\mathbf{p}_K}{\partial\theta_K}\otimes\mathbf{c}_K^R\right]
\end{aligned} \tag{3.47}$$

The definitions for $\mathcal{A}^L, \mathbf{\Gamma}^L, \Theta^L$ follow along similar lines and $\tau_m = (q_m + \gamma_m) \cdot T_c$. The matrix $\mathbf{W}$ now represents a $K \times K$ diagonal matrix of users' amplitudes. The symbol "$\odot$" represents the Hadamard product of two matrices defined by $[\mathbf{A} \odot \mathbf{B}]_{ij} = \mathbf{A}_{ij} \cdot \mathbf{B}_{ij}$.

Theorem 3.1

The Cramér-Rao bound $\mathrm{CRB}(\boldsymbol{\tau}, \boldsymbol{\theta})$ *is given by*

$$\mathrm{CRB}^{-1}(\boldsymbol{\tau}, \boldsymbol{\theta}) = 2L \begin{pmatrix} \mathrm{Re}\left[\Delta_2 - \mathbf{\Upsilon}_1^H \Delta_1^{-1}\, \mathbf{\Upsilon}_1\right] & \mathrm{Re}\left[\mathbf{\Upsilon}_3 - \mathbf{\Upsilon}_1^H \Delta_1^{-1}\, \mathbf{\Upsilon}_2\right] \\ \mathrm{Re}^{\top}\left[\mathbf{\Upsilon}_3 - \mathbf{\Upsilon}_1^H \Delta_1^{-1}\, \mathbf{\Upsilon}_2\right] & \mathrm{Re}\left[\Delta_3 - \mathbf{\Upsilon}_2^H \Delta_1^{-1}\, \mathbf{\Upsilon}_2\right] \end{pmatrix} \tag{3.48}$$

Proof : The derivation is based on a similar exposition in [25,28] and is included in Appendix C. ■

The first K diagonal terms of the inverse of the matrix specified in (3.48) yield the CRBs for the K delay estimates and the next K diagonal terms in the above matrix yield the CRBs for the DOA estimates. The individual calculation of $\mathrm{CRB}(\boldsymbol{\tau})$ or $\mathrm{CRB}(\boldsymbol{\theta})$ is analytically difficult for a general array and code set. For large number of observations, L, and uncorrelated signals the CRB for both parameter estimates reduces to a diagonal matrix each.

Theorem 3.2

For large L *and uncorrelated signals and for* $\mathbf{K}$ *of the form* $\mathbf{K} = \mathbf{K}_1 \otimes \sigma^2 \mathbf{I}$, $\mathrm{CRB}(\tau)$ *and* $\mathrm{CRB}(\theta)$ *can be written as*

$$\mathrm{CRB}(\boldsymbol{\tau}) = \frac{1}{2L} \left(\frac{|\mathbf{W}|^2}{\sigma^2} \right)^{-1} \mathcal{P}^{-1} \cdot \mathrm{Re}^{-1} \left[\ddot{\mathcal{C}} - \dot{\mathcal{C}}^H \mathcal{C}^{-1}\, \dot{\mathcal{C}} \right] \tag{3.49}$$

$$\mathrm{CRB}(\boldsymbol{\theta}) = \frac{1}{2L} \left(\frac{|\mathbf{W}|^2}{\sigma^2} \right)^{-1} \mathcal{C}^{-1} \cdot \mathrm{Re}^{-1} \left[\ddot{\mathcal{P}} - \dot{\mathcal{P}}^H \mathcal{P}^{-1}\, \dot{\mathcal{P}} \right] \tag{3.50}$$

where

$$\mathcal{P} = \mathrm{diag}\left[\mathbf{p}_k^H \mathbf{K}_1^{-1} \mathbf{p}_k\right] \qquad \mathcal{C} = \mathrm{diag}\left[\mathbf{c}_k^{R^H} \mathbf{c}_k^R + \mathbf{c}_k^{L^H} \mathbf{c}_k^L\right]$$

$$\dot{\mathcal{P}} = \mathrm{diag}\left[\frac{\partial \mathbf{p}_k^H}{\partial \theta_k}\, \mathbf{K}_1^{-1} \mathbf{p}_k\right] \qquad \dot{\mathcal{C}} = \mathrm{diag}\left[\boldsymbol{\zeta}_k^{R^H} \mathbf{c}_k^R + \boldsymbol{\zeta}_k^{L^H} \mathbf{c}_k^L\right]$$

$$\ddot{\mathcal{P}} = \mathrm{diag}\left[\frac{\partial \mathbf{p}_k^H}{\partial \theta_k}\, \mathbf{K}_1^{-1} \frac{\partial \mathbf{p}_k}{\partial \theta_k}\right] \qquad \ddot{\mathcal{C}} = \mathrm{diag}\left[\boldsymbol{\zeta}_k^{R^H} \boldsymbol{\zeta}_k^R + \boldsymbol{\zeta}_k^{L^H} \boldsymbol{\zeta}_k^L\right]$$

Proof : Since $\widehat{\mathbf{R}}_{bb}$ is a consistent estimate, $\widehat{\mathbf{R}}_{bb} \xrightarrow{L} \mathbf{R}_{bb}$ and hence for large L, we can replace $\widehat{\mathbf{R}}_{bb}$ with $\mathbf{R}_{bb} = \mathbf{I}$ (since the signals are uncorrelated). We now only have to concern ourselves with the diagonal entries of the matrices $\Delta_1, \Delta_2, \Delta_3, \boldsymbol{\Upsilon}_1, \boldsymbol{\Upsilon}_2$ and $\boldsymbol{\Upsilon}_3$ defined in (3.41)–(3.46). These quantities can therefore be condensed to diagonal matrices

$$\begin{aligned}
\Delta_1 &= \frac{1}{\sigma^2}\,\mathcal{P}\,\mathcal{C}\,, & \boldsymbol{\Upsilon}_1 &= \frac{1}{\sigma^2}\,\mathcal{P}\,\dot{\mathcal{C}}\mathbf{W}^H \\
\Delta_2 &= \frac{1}{\sigma^2}\,\mathcal{P}\,\ddot{\mathcal{C}}\,|\mathbf{W}|^2\,, & \boldsymbol{\Upsilon}_2 &= \frac{1}{\sigma^2}\,\dot{\mathcal{P}}\,\mathcal{C}\mathbf{W}^H \\
\Delta_3 &= \frac{1}{\sigma^2}\,\ddot{\mathcal{P}}\,\mathcal{C}\,|\mathbf{W}|^2\,, & \boldsymbol{\Upsilon}_3 &= \frac{1}{\sigma^2}\,\dot{\mathcal{P}}^H\dot{\mathcal{C}}\,|\mathbf{W}|^2
\end{aligned}$$

Substituting the above expressions in (3.48) we obtain

$$\mathrm{CRB}^{-1}(\tau\,,\theta) = 2L\begin{pmatrix} \frac{|\mathbf{W}|^2}{\sigma^2}\mathrm{Re}\left[\mathcal{P}\left(\ddot{\mathcal{C}} - \dot{\mathcal{C}}^H\mathcal{C}^{-1}\dot{\mathcal{C}}\right)\right] & \mathbf{0} \\ \mathbf{0} & \frac{|\mathbf{W}|^2}{\sigma^2}\mathrm{Re}\left[\mathcal{C}\left(\ddot{\mathcal{P}} - \dot{\mathcal{P}}^H\mathcal{P}^{-1}\dot{\mathcal{P}}\right)\right] \end{pmatrix}, \tag{3.51}$$

and hence the theorem. ■

To understand the behavior of the CRBs with M, N, L and SNR, we make a few simplifying assumptions :

1. A ULA at the receiver with the intersensor spacing being half the carrier wavelength (to avoid spatial aliasing), where the array response vector is of the form

$$\mathbf{p}_k = [1 \;\; e^{j\pi\sin\theta_k} \;\; \cdots \;\; e^{j(M-1)\pi\sin\theta_k}]^\top, \tag{3.52}$$

2. Random codes — $\mathbf{c}_{k,n} \in \{+1, -1\}$ with equal probability.

With the above assumptions, and defining $\phi_k \stackrel{\triangle}{=} \pi\sin(\theta_k)$, it can be shown, through

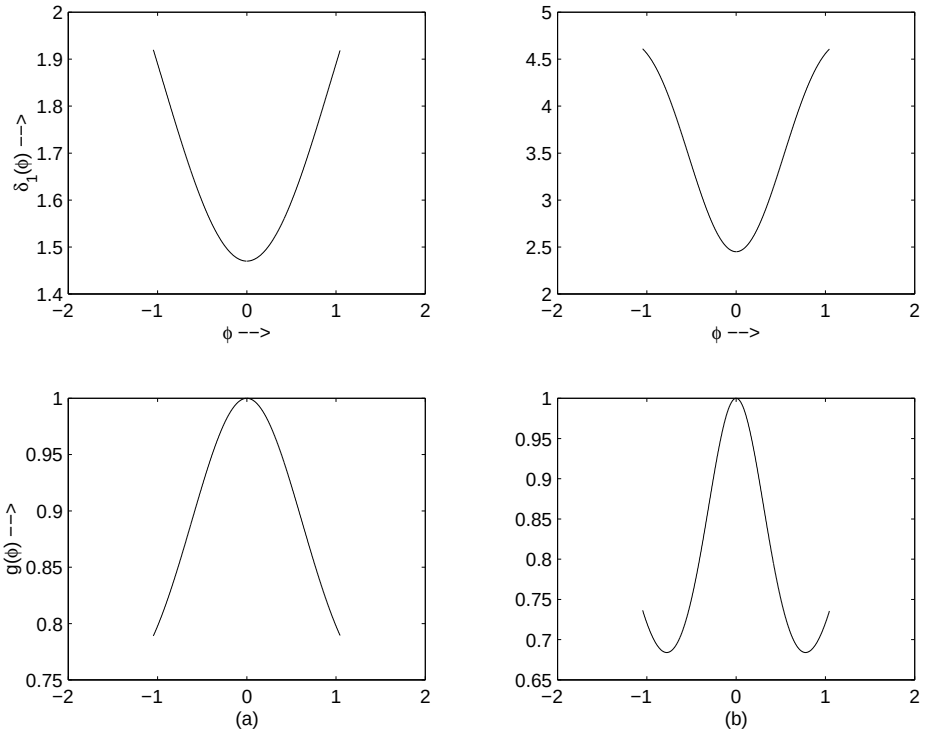


Figure 3.2 : Plots of $g(\phi)$ and $\delta_1(\phi)$ versus ϕ, depicting the trade off in the simultaneous estimation of $\hat{\tau}_k$ and $\hat{\theta}_k$. The top plots depict $\delta_1(\phi)$ and the bottom plots represent $g(\phi)$. Plots (a) were made for $M = 3$ and plots (b) for $M = 5$.

a fair amount of tedious algebra, that

$$\mathrm{CRB}(\gamma_k) = \frac{(1-\beta)^2}{2LMN\ \mathrm{SNR}_k} \cdot \frac{(M\beta^2 + (1-\beta)^2)}{\delta_1(\phi_k)} \cdot \left(\gamma_k^2 + (1-\gamma_k)^2\right) \tag{3.53}$$

$$\mathrm{CRB}(\theta_k) = \frac{6(1-\beta)^2}{LM(M^2-1)N\ \mathrm{SNR}_k} \cdot \frac{1}{(\gamma_k^2 + (1-\gamma_k)^2)} \cdot \left(\frac{1}{1 - \dfrac{12\ \beta^2\ \delta_2^2(\phi_k)}{(M^2-1)\delta_1(\phi_k)}} \right), \tag{3.54}$$

where the variables $\delta_1(\phi_k)$ and $\delta_2(\phi_k)$ are defined by

$$\begin{aligned} \delta_1(\phi_k) &= \left(M^2 - \frac{\sin^2 M\phi_k/2}{\sin^2 \phi_k/2} \right) \beta^2 + M(1-\beta)^2 \\ \delta_2(\phi_k) &= \frac{M \sin(\phi_k/2)\cos(M\phi_k/2) - \cos(\phi_k/2)\sin(M\phi_k/2)}{2\sin^2 \phi_k/2}. \end{aligned} \tag{3.55}$$

We observe that the CRBs decrease with increasing number of observations L, code length N, array size M, and SNR. In light of the fact that the estimate of τ_k is decoupled from the estimate of θ_k, the dependence of CRB(γ_k) on M might seem disconcerting initially. However, it must be noted that the decoupling is exact only asymptotically in L, or equivalently in SNR. For finite L, increasing the number of sensors M effectively increases the SNR and hence we expect this to influence the estimates (especially in low SNR regimes, as we will observe later). Another interesting observation can be made on closer inspection of (3.53), (3.54) and (3.55) and Figure 3.2, which graphs $g(\phi)$ and $\delta_1(\phi)$, with $g(\phi)$ defined by

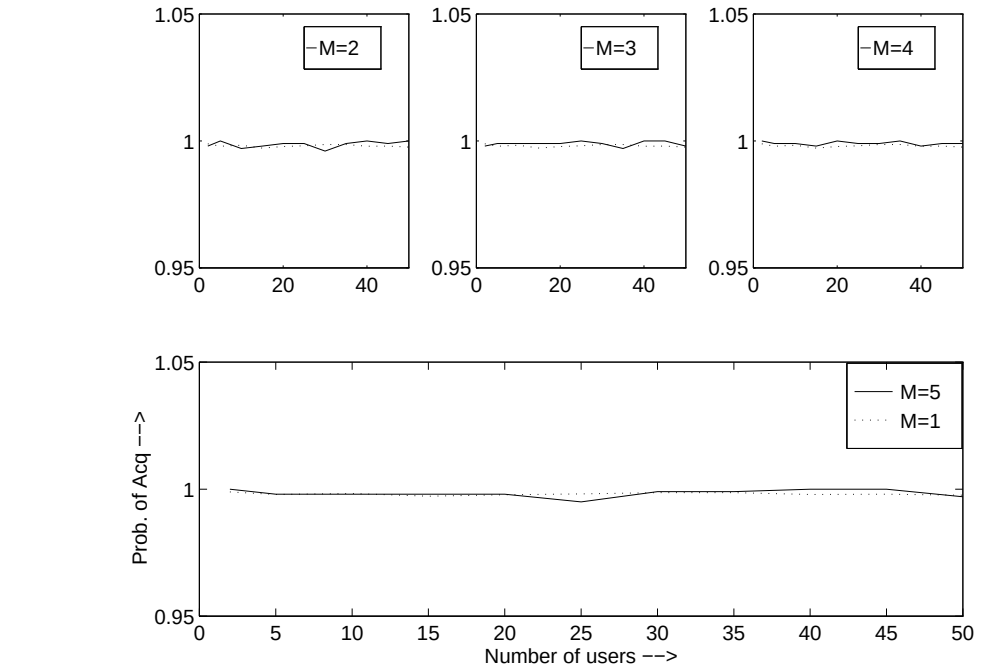
$$g(\phi) = 1 - \frac{12\ \beta^2\ \delta_2^2(\phi_k)}{(M^2-1)\delta_1(\phi_k)}.$$

We see that the γ_k that minimizes CRB(γ_k), i.e., $\gamma_k = 0.5$, actually maximizes CRB(θ_k). Furthermore, from Figure 3.2 we note that for nonzero β, the θ_k that minimizes CRB(θ_k), i.e., $\theta_k = 0$, in turn maximizes CRB(γ_k). Thus there appears to be a performance trade-off between estimating τ_k and θ_k when both are jointly estimated.

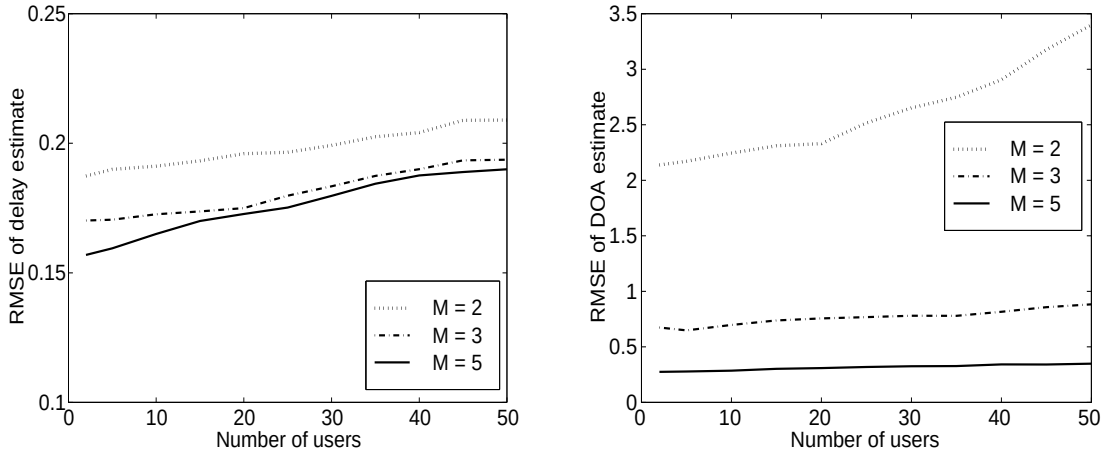
3.3 Numerical Results

We proceed to describe the numerical simulations that we conducted to evaluate the performance of the afore-mentioned estimators. A code length of $N = 31$ was used in all the simulations. Furthermore, the array structure assumed was a uniform linear array (ULA). The delays of all the users were assumed uniformly distributed in $[0.1, 31)$ chips and the DOAs were assumed to be uniformly distributed in $[-60, 60]^\circ$, corresponding to one sector of 120° in a cell. The noise covariance structure postulated was as in (2.18) with σ^2 being determined by the desired user's SNR and β being fixed at 0.3. Each point in the subsequent plots was generated using 1000 Monte-Carlo trials.

Performance .vs. K: In Fig. 3.3(a)–(c) we depict the performance of the esti-



(a)



(b) (c)

Figure 3.3 : Performance of ML estimators for delay and DOA as a function of the number of users. Plot(a) portrays the probability of acquisition; plots(b) and (c) graph the RMS errors of the delay and DOA estimates in chips and degrees respectively. The training sequence length is 200 bits. The SNR of the desired user is 6dB per sensor. Length of each spreading code is 31.

mators in terms of the probability of acquisition and root mean squared errors of the delay and DOA estimates with increasing number of users (K) in the system. Since our main goal behind using antenna arrays is to increase the acquisition based capacity of the system, we compare the performance of the algorithm with 1, 3 and 5 sensors in the array. The performance is measured in terms of the probability of acquisition and the root mean squared errors of the various estimates. The probability of acquisition is defined to be the probability that $|\hat{\tau}_k - \tau_k| < T_c$. The SNR of the desired user at each sensor was fixed at a rather high value of 6dB to highlight the effect of increasing K. The multiple access interference (MAI) for each interferer, which is the ratio of the interferer's and desired user's received energies, was taken to be uniformly distributed in $[0, 20]$dB to simulate users at different distances from the base station, and the window length used was $L = 200$ bits. The probability of acquisition curves in Fig. 3.3(a) show no variation across K. A rather surprising observation is that there is no gain in increasing the number of array elements (at this SNR) — as we can see, the probability of acquisition and mean squared error of the delay estimate are relatively invariant to increasing the number of sensors, while the DOA estimate is significantly affected. Careful deliberation leads us to realize, however, that since we succeeded in decoupling the delay and DOA estimation problems, the parameter that directly governs the delay estimate is the code length N (if we were to change N we would notice a concomitant change in the quality of the delay estimates). We remark here that the ML estimators are not dimension limited like subspace-based estimators [27, 29]. This is reflected in the fact that even with a single sensor, we are able to acquire in excess of 31 users, this being the dimension of the covariance matrix.

Performance .vs. L: In Figs. 3.4(a)–(c) we investigate the behavior of the estimates as a function of increasing window size for SNRs of 0 and 6 dB. Once again, we examine the performance with $M = 1, 3$ and 5 sensors in the array with $K = 25$ users in the system. The MAI was uniformly distributed in $[0, 20]$dB. It is desirable to have

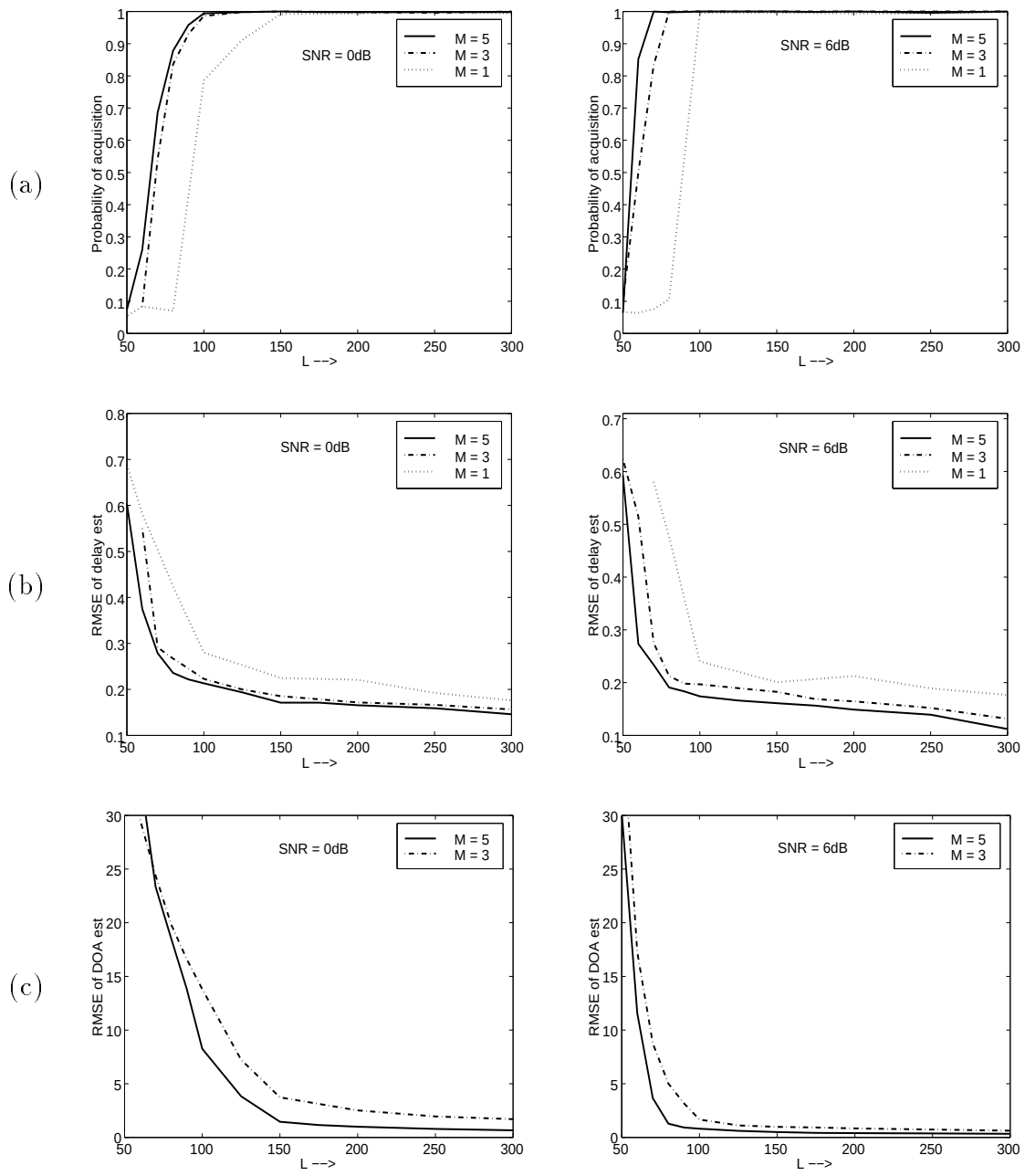
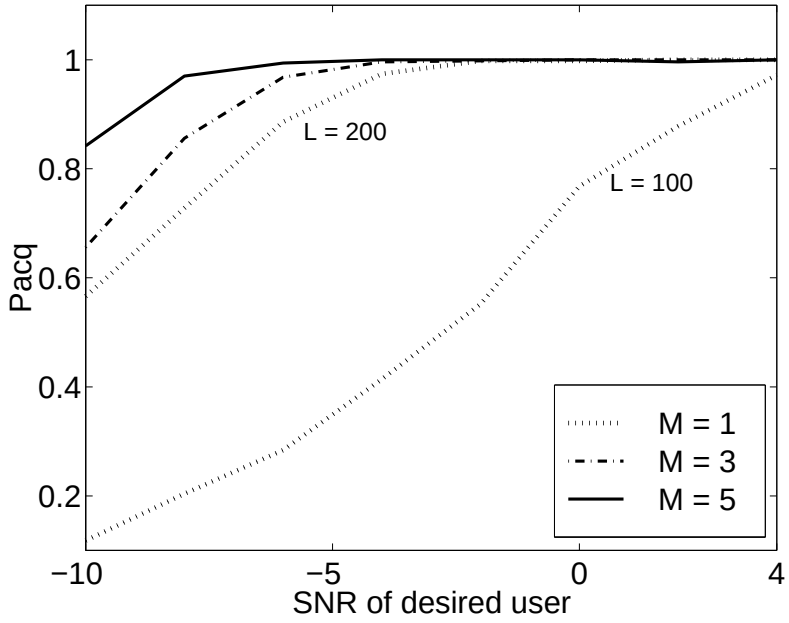


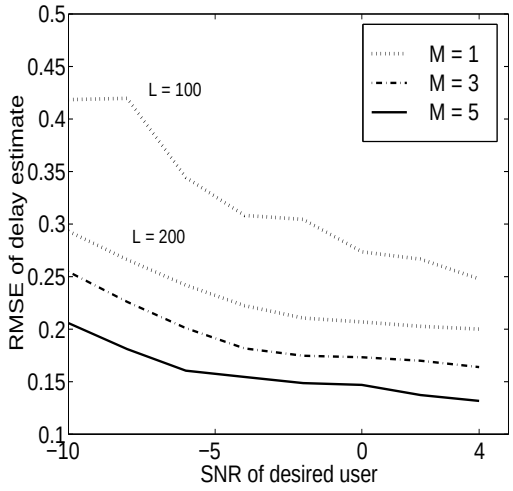
Figure 3.4 : Performance of ML estimators for delay and DOA as a function of window length. Plots (a) portrays the probability of acquisition; plots (b) and (c) graph the RMS errors of the delay and DOA estimates in chips and degrees respectively. The number of users in the sector is 25. The performance is depicted for two SNR levels for the desired user, 0dB on the left and 6dB on the right. Length of each spreading code is 31.

acquisition occur in the shortest time possible to minimize the impact on data rates. Fig. 3.4(a) graphs the probability of acquisition while the root mean squared errors of the delay and DOA estimates are graphed in Figs. 3.4(b) and (c) respectively. Since the dimension of the covariance matrix is $MN \times MN$, we observe that its estimate for $L < MN$ is rank deficient; hence, in place of the inverse we used the pseudo-inverse with little performance degradation as can be seen in Fig. 3.4(a). The limiting factor on the number of observations, it would appear, is actually governed by the invertibility of the sample correlation matrix of the training sequences, $\widehat{\mathbf{R}}_{bb}$, and not that of the covariance matrix estimate, $\widehat{\mathbf{K}}$. Thus, for preamble lengths larger than $2K$ the quality of the estimates begin to improve. This imparts a subspace flavor to the algorithm; however, as pointed out earlier, the algorithm is not inherently dimension limited. We see that a preamble length of 150 bits is sufficient for acquisition in most SNR regimes. In fact, we observe that acquisition occurs more rapidly with 3 or 5 sensors as opposed to just one sensor.

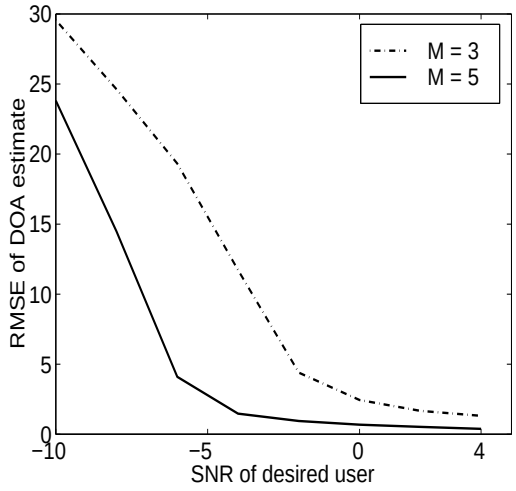
Performance .vs. SNR: We now depict the behavior of the algorithm as a function of the SNR of the desired user at each sensor. The number of interfering users in the system was assumed to be 24, each of whose power level relative to the desired user was uniformly distributed in $[0, 20]$dB. The preamble length was chosen to be 200 bits for $M = 1, 3$ and 5 sensors. Fig. 3.5(a) graphs the probability of acquisition with increasing SNRs while Figs. 3.5(b) and (c) portray plots of the root mean squared errors of the delay and DOA estimates. We see that the performance for SNRs as low as 0dB is uniformly good for all the three array sizes. However, as the SNR is decreased further, we see a definite delineation in performance across the number of sensors in the array. Thus, there is a definite advantage to employing multiple sensors at the receiver. The inherent trade-off between SNR and preamble length is shown in the same figure where we have graphed the probability of acquisition for a preamble length of 100 bits for one sensor (the behavior for larger number of sensors can be inferred from this behavior). The degradation in performance is easily explained:



(a)



(b)



(c)

Figure 3.5 : Performance of ML algorithm with increasing SNR for different number of sensors. Plot(a) portrays the probability of acquisition; plots(b) and (c) graph the RMS errors of the delay and DOA estimates in chips and degrees respectively. The number of users is 25 and the MAI is uniformly distributed in $[0, 20]$dB. The preamble length is 200 bits. Code length = 31.

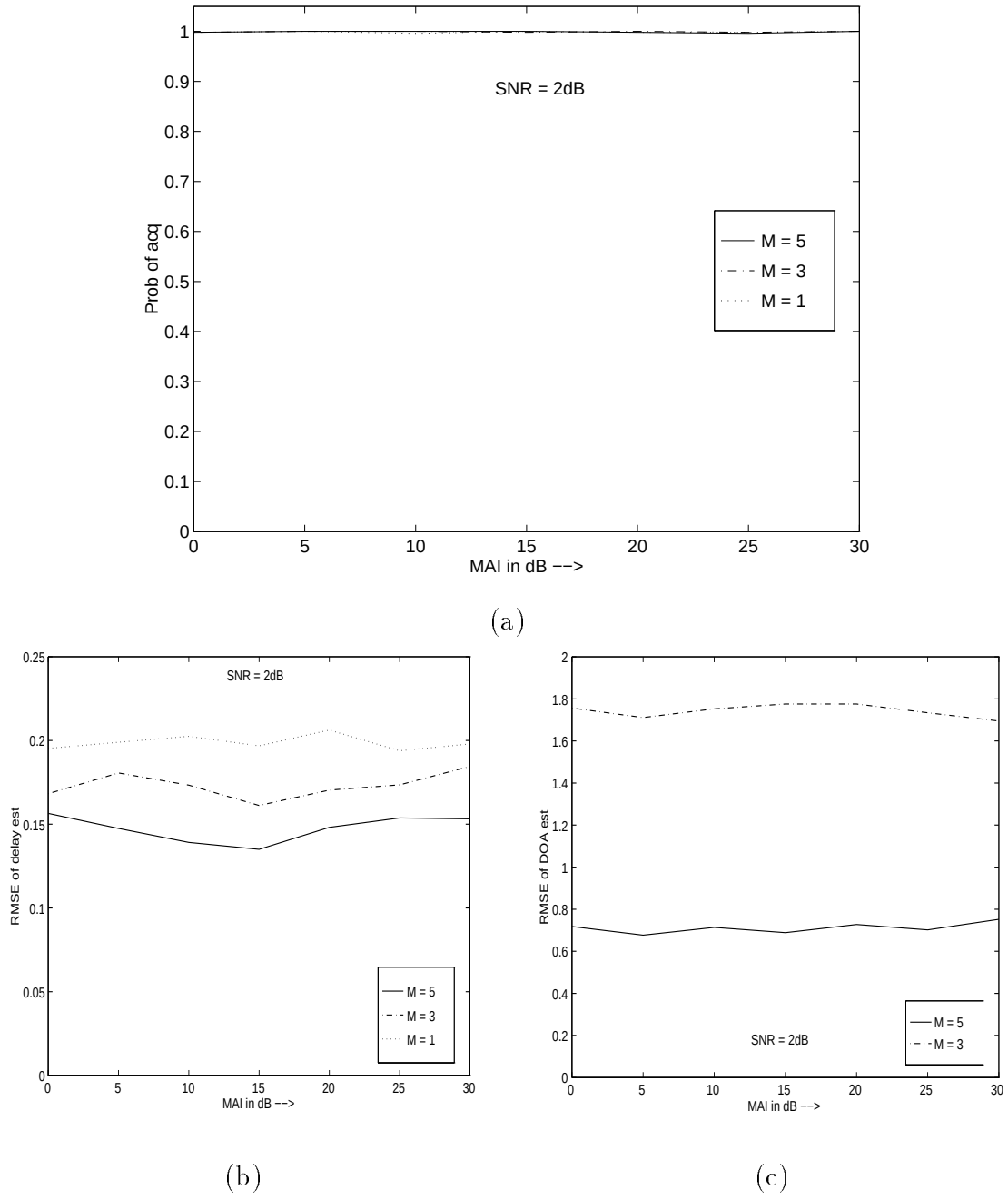


Figure 3.6 : Performance of ML algorithm with increasing MAI for different number of sensors. Plot(a) portrays the probability of acquisition; plots(b) and (c) graph the RMS errors of the delay and DOA estimates in chips and degrees respectively. The number of users is 25 and the preamble length is 200 bits. The SNR of the desired user is 2dB at each sensor. Code length = 31.

since the estimates are maximum likelihood estimates, consistency decrees that the larger the preamble size, the more accurate the estimates of the covariance matrix $\widehat{\mathbf{K}}$ and the matrix $\widehat{\mathcal{Y}}$, which results in improved performance.

Performance .vs. MAI: From the expression in (3.19) we see that the estimates for τ_k and θ_k are independent of the received amplitude and hence, we expect the algorithm to be near-far resistant. This property is captured in the plots in Figs. 3.6(a)–(c). Fig. 3.6(a) graphs the probability of acquisition while (b) and (c) depict root mean squared error estimates in delay and DOA. The graphs were generated for a system with 24 interfering users. An MAI level of xdB signifies that the interferer powers are all uniformly distributed in $[0, x]$dB. The SNR of the desired user at each sensor was 2dB and the preamble length was fixed at 200 bits. Fig. 3.6(a) validates our claim of near-far resistance; furthermore, we notice that the probability of acquisition is independent of array size. However, the mean squared errors decrease with increased array size, leading to gains in the performance of any detector that makes use of these estimates.

Chapter 4

Subspace Based Techniques for Delay and DOA Estimation

Subspace based parameter estimation algorithms, also called geometric algorithms, such as MUSIC [43], ESPRIT [44], Weighted Subspace Fitting [45], et al., exploit the knowledge of the structure of the transmitted signals and the noise. The aim is to identify the signal contribution in the received observations that are corrupted by additive Gaussian noise.

We first introduce the concept of an *array manifold*. The signal vectors $\mathbf{a}_k^R(\tau,\theta), \mathbf{a}_k^L(\tau,\theta) \in \mathbb{C}^{MN}$ each describe a two dimensional manifold for τ and θ ranging over the entire parameter space, $\tau \in [0, NT_c)$, $\theta \in [-\Theta, \Theta)$, where Θ can be conveniently viewed as an angular limit for one sector of a cell in a cellular communication system. These individual manifolds are called the array manifolds and are distinct for each user due to the distinct codes assigned (see (2.16) for the definition of $\mathbf{a}_k^R(\tau,\theta)$ and $\mathbf{a}_k^L(\tau,\theta)$). In conventional array processing terms, this is the equivalent of different array patterns for different users. Thus we see that the individual spreading codes provide levels of diversity that would not be possible in conventional DOA estimation — the individual users are now distinguished by their different spreading codes and delays *as well as* their DOAs. We will assume that the mapping, f, from the parameter set $\{\boldsymbol{\tau}, \boldsymbol{\theta}\}$ to $\mathcal{A}(\boldsymbol{\tau}, \boldsymbol{\theta})$ is invertible (which is not too stringent an imposition). In other words, given $\mathbf{a}_k^R(\tau,\theta)$ and $\mathbf{a}_k^L(\tau,\theta)$ the parameters τ and θ are uniquely determined. Thus, we have

$$\mathbf{a}_k^{R/L}(\tau_1,\theta_1) = \mathbf{a}_k^{R/L}(\tau_2,\theta_2) \Longrightarrow \tau_1 = \tau_2 \text{ and } \theta_1 = \theta_2. \tag{4.1}$$

In the absence of noise, the K received signals across the array lie in a $2K$-

dimensional subspace of $\mathbb{C}^{MN}$, the *signal subspace* $\mathcal{S}$, spanned by the $2K$ columns of $\mathcal{A}$. The array manifolds for user k each intersect the signal subspace at one point given by $\mathbf{a}_k^R(\tau_k, \theta_k)$ and $\mathbf{a}_k^L(\tau_k, \theta_k)$ respectively, where τ_k and θ_k are the true values of the respective parameters [29, 44]. Thus $\{\tau_k, \theta_k\}$ for each user are found as those parameters that render the span of the columns of $\mathcal{A}(\boldsymbol{\tau}, \boldsymbol{\theta})$, which we denote by $\mathcal{R}\{\mathcal{A}(\boldsymbol{\tau}, \boldsymbol{\theta})\}$, equal to $\mathcal{S}$. In the presence of additive noise the received observations now span all of $\mathbb{C}^{MN}$, which can be decomposed into a signal subspace and an orthogonal subspace, the *noise subspace* $\mathcal{N}$. If we had perfect knowledge of the signal subspace we would find that the $2K$ array manifolds still intersect this subspace at exactly one point each. However, in the presence of noise, it is unreasonable to expect perfect knowledge of the signal subspace; this must be estimated from the received observations. Now, the array manifolds no longer intersect this subspace, implying that $\mathcal{R}\{\mathcal{A}(\boldsymbol{\tau}, \boldsymbol{\theta})\} \neq \hat{\mathcal{S}}$, the estimated signal subspace. Instead, the parameters τ_k and θ_k for each user are estimated by maximizing the normalized projection of the signal vectors onto the observed (or estimated) signal subspace or equivalently, by minimizing their projection onto the estimated noise subspace $\hat{\mathcal{N}}$. This latter approach, termed MUSIC, is adopted in this study. The above statements are quantified in the subsequent paragraphs.

4.1 Subspace Estimation

To decompose the whole space into the constituent signal and noise subspaces we make use of the correlation matrix of the received observations. An eigenvalue decomposition of this correlation matrix serves as the vehicle for the above mentioned decomposition. We can rewrite the received vector in (2.13) as

$$\mathbf{r}_i = \mathcal{A}\mathbf{x}_i + \boldsymbol{\nu} \ , \tag{4.2}$$

where $\mathbf{x}_i = \mathbf{W}\mathbf{b}_i$. With the assumptions made on the noise statistics (zero mean, identically distributed at each sensor and uncorrelated from sensor to sensor), the

correlation matrix of the received observations is

$$\mathbf{R}_{rr} = E[\mathbf{r}_i \mathbf{r}_i^H] = \mathcal{A}\mathbf{R}_{xx}\mathcal{A}^H + \sigma^2\mathbf{I} \tag{4.3}$$

where the diagonal matrix $\mathbf{R}_{xx} = E[\mathbf{x}_i \mathbf{x}_i^H]$, the power in the received bits. In the single path case, this matrix is full rank; in the multipath case it is rank deficient. In this chapter we will assume a slow fading single path environment. The eigenvector decomposition of $\mathbf{R}_{rr}$ can be written as $\mathbf{R}_{rr} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ where $\mathbf{V}$ is a unitary matrix. If we partition the eigenvector matrix $\mathbf{V}$ into two matrices, $\mathbf{V}_\mathcal{S} \in \mathbb{C}^{MN\times 2K}$ and $\mathbf{V}_\mathcal{N} \in \mathbb{C}^{MN\times(MN-2K)}$ as $\mathbf{V} = [\mathbf{V}_\mathcal{S}\ \mathbf{V}_\mathcal{N}]$, we can then express $\mathbf{R}_{rr}$ as

$$\mathbf{R}_{rr} = \begin{bmatrix}\mathbf{V}_\mathcal{S} & \mathbf{V}_\mathcal{N}\end{bmatrix} \begin{bmatrix}\mathbf{\Lambda}_\mathcal{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_\mathcal{N}\end{bmatrix} \begin{bmatrix}\mathbf{V}_\mathcal{S}^H \\ \mathbf{V}_\mathcal{N}^H\end{bmatrix} = \mathbf{V}_\mathcal{S}\mathbf{\Lambda}_\mathcal{S}\mathbf{V}_\mathcal{S}^H + \mathbf{V}_\mathcal{N}\mathbf{\Lambda}_\mathcal{N}\mathbf{V}_\mathcal{N}^H. \tag{4.4}$$

In the above representation the columns of $\mathbf{V}_\mathcal{S}$ and $\mathbf{V}_\mathcal{N}$ span the signal and noise subspaces respectively and the submatrices $\mathbf{\Lambda}_\mathcal{S}$ and $\mathbf{\Lambda}_\mathcal{N}$ represent the associated eigenvalues. If the matrix product $\mathcal{A}\,\mathbf{R}_{xx}\mathcal{A}^H$ in (4.3) is of full rank $2K$, the $2K$ largest eigenvalues are just the nonzero eigenvalues of $\mathcal{A}\,\mathbf{R}_{xx}\mathcal{A}^H$ incremented by σ^2 and the $MN - 2K$ smallest eigenvalues of $\mathbf{R}_{rr}$ are all equal to σ^2, i.e., $\lambda_{2K+1} = \cdots = \lambda_{MN} = \sigma^2$. Therefore we have $\mathbf{\Lambda}_\mathcal{N} = \sigma^2\mathbf{I}$ and (4.4) can then be written as

$$\mathbf{R}_{rr} = \mathbf{V}_\mathcal{S}\mathbf{\Lambda}_\mathcal{S}\mathbf{V}_\mathcal{S}^H + \sigma^2\mathbf{V}_\mathcal{N}\mathbf{V}_\mathcal{N}^H = \mathbf{V}_\mathcal{S}\widetilde{\mathbf{\Lambda}}_\mathcal{S}\mathbf{V}_\mathcal{S}^H + \sigma^2\mathbf{I} \tag{4.5}$$

where $\widetilde{\mathbf{\Lambda}}_\mathcal{S} = \mathbf{\Lambda}_\mathcal{S} - \sigma^2\mathbf{I}$.

In a practical situation the correlation matrix $\mathbf{R}_{rr}$ has to be estimated from the observations and an estimate over a window of L data vectors at the i^{th} time instant is formulated as

$$\widehat{\mathbf{R}}_{rr}(i) = \frac{1}{L}\sum_{j=i-L+1}^{i} \mathbf{r}_j \mathbf{r}_j^H. \tag{4.6}$$

Hence we only have estimates for the signal and noise subspaces, $\widehat{\mathbf{V}}_\mathcal{S}$ and $\widehat{\mathbf{V}}_\mathcal{N}$ respectively.

Instead of an eigenvalue decomposition of $\widehat{\mathbf{R}}_{rr}$, a singular value decomposition (SVD) of the accumulated $(MN \times L)$ data matrix, $\mathbf{Y} = [\mathbf{r}_{i-L+1}, \ldots, \mathbf{r}_i]$ can be performed. The SVD of $\mathbf{Y}$ is preferred over the eigenvalue decomposition because of its superior numerical stability and ease of pipelined implementations [46, 47]. As before, we partition $\widehat{\mathbf{V}}$ as $\widehat{\mathbf{V}} = [\widehat{\mathbf{V}}_{\mathcal{S}} \quad \widehat{\mathbf{V}}_{\mathcal{N}}]$ where $\widehat{\mathbf{V}}_{\mathcal{S}}$ and $\widehat{\mathbf{V}}_{\mathcal{N}}$ form bases for the *estimated* signal and noise subspaces respectively. The estimate $\widehat{\mathbf{R}}_{rr}(i)$ can be shown to be the maximum likelihood estimate of $\mathbf{R}_{rr}$ [48] and accordingly, $\widehat{\mathbf{V}}_{\mathcal{S}}$ and $\widehat{\mathbf{V}}_{\mathcal{N}}$ are the ML (and hence consistent) estimates of $\mathbf{V}_{\mathcal{S}}$ and $\mathbf{V}_{\mathcal{N}}$.

4.2 Parameter Estimation

A comparison of (4.3) and (4.5) reveals that the columns of $\mathbf{V}_{\mathcal{S}}$ and $\mathcal{A}(\boldsymbol{\tau}, \boldsymbol{\theta})$ span the same subspace, the signal subspace, for the *true* values of parameter vectors $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$. This suggests a natural method for estimating these parameters: for user k, choose τ_k and θ_k such that the corresponding pair of signal vectors $\mathbf{a}_k^R(\tau_k, \theta_k)$ and $\mathbf{a}_k^L(\tau_k, \theta_k)$ are orthogonal to $\mathbf{V}_{\mathcal{N}}$, i.e.,

$$\left\{\hat{\tau}_k, \hat{\theta}_k\right\} = \arg_{\tau,\theta}\left\{ \left[\mathbf{a}_k^R(\tau, \theta) \; \mathbf{a}_k^L(\tau, \theta)\right] \perp \mathbf{V}_{\mathcal{N}}\right\},$$

With only finite sample estimates of the noise subspace at our disposal, signal vectors $\mathbf{a}_k^R$ and $\mathbf{a}_k^L$ are not completely orthogonal to $\widehat{\mathbf{V}}_{\mathcal{N}}$; we now choose $\{\hat{\tau}_k, \hat{\theta}_k\}$ to minimize the projection of the sum $\mathbf{a}_k(\tau, \theta) = \mathbf{a}_k^R + \mathbf{a}_k^L$ into $\widehat{\mathbf{V}}_{\mathcal{N}}$ (if $\mathbf{a}_k^R$ and $\mathbf{a}_k^L$ are orthogonal to $\mathbf{V}_{\mathcal{N}}$ then so is their sum, but not the vice-versa; hence we are making an approximation here but, as we will see, the performance loss is minimal). Accordingly we have the MUSIC algorithm first proposed in [43]

$$\left\{\hat{\tau}_k, \hat{\theta}_k\right\} = \arg\min_{\tau,\theta} \; e(\tau, \theta) = \arg\min_{\tau,\theta} \; \left\|\mathbf{a}_k(\tau, \theta)^H \; \widehat{\mathbf{V}}_{\mathcal{N}}\right\|^2 \tag{4.7}$$

From (2.10), (2.16) and the definition of $\mathbf{a}_k(\tau, \theta)$, we see that it can also be written

as

$$\begin{aligned}\mathbf{a}_k(\tau,\theta) &= \mathbf{p}_k(\theta) \otimes \Big((1-\gamma)\mathbf{c}_k[q] + \gamma\mathbf{c}_k[q+1]\Big) \\ &= \mathbf{p}_k(\theta) \otimes \mathbf{c}_k(\tau),\end{aligned} \tag{4.8}$$

where τ is decomposed as $\tau/T_c = q + \gamma$, $q \in \{0, \ldots, N-1\}$, $\gamma \in [0,1)$ and $\mathbf{c}_k[\cdot] = \mathbf{c}_k^R[\cdot] + \mathbf{c}_k^L[\cdot]$. We therefore see, from the structure of the array response vector $\mathbf{p}_k(\theta)$ and the spreading code vector $\mathbf{c}_k(\tau)$, that the function $e(\tau,\theta)$ is a continuously differentiable function of θ but only piecewise (over a chip) continuously differentiable in τ. From (4.7) and (4.8) we can write $e(\tau,\theta)$ as

$$e(\tau,\theta) = \Big[\mathbf{p}_k(\theta) \otimes \Big(\mathbf{c}_k[q] + \gamma\boldsymbol{\zeta}_k[q]\Big)\Big]^H \cdot \mathbf{Q} \cdot \Big[\mathbf{p}_k(\theta) \otimes \Big(\mathbf{c}_k[q] + \gamma\boldsymbol{\zeta}_k[q]\Big)\Big], \tag{4.9}$$

with $\boldsymbol{\zeta}_k[q] = (\mathbf{c}_k[q+1] - \mathbf{c}_k[q])$ and $\mathbf{Q} = \widehat{\mathbf{V}}_{\mathcal{N}}\widehat{\mathbf{V}}_{\mathcal{N}}^H$. For a particular DOA, $\theta = \theta_m$, we see from (4.9) that, for $\tau \in [q, q+1)$, the MUSIC norm $e(\tau,\theta_m)$ can be written as

$$e\big(\tau,\theta_m\big) = \Big(\mathbf{u}_k + \gamma\mathbf{v}_k\Big)^H \mathbf{Q}\ \Big(\mathbf{u}_k + \gamma\mathbf{v}_k\Big), \tag{4.10}$$

where the vectors $\mathbf{u}_k, \mathbf{v}_k \in \mathbb{C}^{MN\times 1}$ and $\mathbf{u}_k = \mathbf{p}_k(\theta_m) \otimes \mathbf{c}_k[q]$, $\mathbf{v}_k = \mathbf{p}_k(\theta_m) \otimes \boldsymbol{\zeta}_k[q]$. Thus, for a fixed θ, the function can be minimized over the τ variable in closed form. The optimum γ as a function of θ_m in the chip interval $[q, q+1)$ is obtained as

$$\hat{\gamma}(\theta_m) = \frac{-\mathrm{Re}\left[\mathbf{u}_k^H\mathbf{Q}\mathbf{v}_k\right]}{\mathbf{v}_k^H\mathbf{Q}\mathbf{v}_k}, \quad \hat{\tau}(q,\theta_m) = (q + \hat{\gamma}(\theta_m))\,T_c. \tag{4.11}$$

Since the function e is only piecewise differentiable over chip intervals, we obtain the minimum over $\gamma \in [0,1)$ in each of the N chip intervals and compare these N function values to obtain the global minimum for the particular $\theta = \theta_m$ as

$$q^* = \arg\min_q \left\{e\left(\hat{\tau}(q,\theta_m),\theta_m\right)\right\}, \quad q = 0, \ldots, N-1,$$

and subsequently $\hat{\tau}_k(q^*,\theta_m) = (q^* + \hat{\gamma})T_c$. We then minimize over θ by selecting the $\hat{\tau}_k(q^*,\theta_m), m = 1, \ldots, M_\theta$ that yields the minimum corresponding function value, thus obtaining the optimum $\{\hat{\tau}_k, \hat{\theta}_k\}$ pair for that choice of grid, where M_θ is the number of points in the grid.

The above procedure still requires a one dimensional search over the θ parameter. The computational complexity of the algorithm is directly linked to the accuracy in the solution via the grid spacing in the θ domain; the finer the grid the more accurate but more computationally complex the solution. We therefore attempt to approximate the array response vector, $\mathbf{p}_k(\theta)$, in such a way as to lead to a simpler solution for the θ parameter also. The approximation is achieved by expanding $\mathbf{p}_k(\theta)$ in a Taylor series expansion and limiting the expansion to the first order term. For $s \in [-\theta_h, \theta_h)$, the array response vector $\mathbf{p}_k(\theta_m + s)$ can be expanded about a point θ_m as

$$\mathbf{p}_k(\theta_m + s) = \mathbf{p}_k(\theta_m) + s\,\dot{\mathbf{p}}_k(\theta_m), \tag{4.12}$$

where $\dot{\mathbf{p}}_k(\theta_m) = \partial \mathbf{p}_k / \partial\theta \Big|_{\theta=\theta_m}$. We are interested in the local properties of the approximate and hence we focus our attention on the approximate MUSIC norm $\widehat{e}$ in the region $\{(\tau,\theta) : q \le \tau < q+1, \;\; \theta_m - \theta_h \le \theta < \theta_m + \theta_h\}$ where $q = 0, \ldots, N-1$ and $m = 1, \ldots, M_I$. For clarity of exposition we ignore the indices q and m of chip and angle respectively and write $\widehat{e}$ as

$$\widehat{e}(\gamma, s) = \left[\left(\mathbf{p}_k + s\dot{\mathbf{p}}_k\right) \otimes \left(\mathbf{c}_k + \gamma\boldsymbol{\zeta}_k\right)\right]^H \mathbf{Q} \left[\left(\mathbf{p}_k + s\dot{\mathbf{p}}_k\right) \otimes \left(\mathbf{c}_k + \gamma\boldsymbol{\zeta}_k\right)\right]. \tag{4.13}$$

Direct differentiation of $\widehat{e}$ with respect to γ and s yields a set of two coupled nonlinear equations in γ and s which proves intractable to solve. Inspecting (4.13), it can be seen that, due to the non-negative definite nature of $\mathbf{Q}$, $\widehat{e}(\gamma, s)$ is individually convex in either variable (though not necessarily jointly so). This suggests an iterative approach whereby the function is held fixed along one coordinate and minimized along the other; this process is then repeated interchanging the coordinates until a suitable neighborhood of the minimum is reached. For instance, if $\gamma^{(0)}$ is chosen as the starting point, it is easily seen from (4.13) that the optimal s is given by

$$s^{(1)} = \frac{-\mathrm{Re}\left[\tilde{\mathbf{u}}_k^H \mathbf{Q} \tilde{\mathbf{v}}_k\right]}{\tilde{\mathbf{v}}_k^H \mathbf{Q} \tilde{\mathbf{v}}_k},$$

where $\tilde{\mathbf{u}}_k = \mathbf{p}_k \otimes \mathbf{c}_k(\tau^{(0)})$, $\tilde{\mathbf{v}} = \dot{\mathbf{p}}_k \otimes \mathbf{c}_k(\tau^{(0)})$ and $\tau^{(0)} = (q + \gamma^{(0)})T_c$. Given $s^{(1)}$ and hence $\theta^{(1)} = \theta_m + s^{(1)}$, we update value of γ to $\gamma^{(1)}$ from (4.11), where $\mathbf{u}_k$ and $\mathbf{v}_k$ are now functions of $\theta^{(1)}$. Obviously, if the number of iterations, j, required for $\{\gamma^{(j)}, s^{(j)}\}$ to fall within a "suitable" neighborhood of the true minimum $\{\gamma^*, s^*\}$ in the region is large, the savings in computation over the previous method are negligible. As described in the next section, we see that a surprisingly small number of iterations (two) suffice to fall within a reasonable neighborhood of $\{\gamma^*, s^*\}$.

Having obtained a suitable pair $\{\gamma^{(j)}, s^{(j)}\}$ in the region specified by $\{(\tau, \theta) : q \leq \tau < q+1, \;\; \theta_m - \theta_h \leq \theta < \theta_m + \theta_h\}$, the global minimum over all the regions, $q = 0, \ldots, N-1, \;\; m = 1, \ldots, M_I$ is calculated by comparing the approximate MUSIC norm at these local optima and selecting the pair corresponding to the minimum.

4.3 Performance of Estimators

In this section we examine the estimates of τ_k and θ_k obtained via the MUSIC algorithm described above. The main objective is to demonstrate the superior performance of multisensor estimates over their single sensor counterparts. Subspace based algorithms in general, and MUSIC in particular, are dimension limited, i.e., the number of signals' parameters that can be estimated is limited by the dimension of the observation correlation matrix. In the current context, this implies that for an M-sensor antenna array, the number of delay-DOA pairs that can be estimated are limited by the relation $MN - 2K \geq 0$ and therefore, $K \leq MN/2$. We see, therefore, that there is an immediate gain of M for the multisensor receiver over the single sensor one. This gain is highlighted in the subsequent plots. All the simulations described in this section were done for code lengths of $N = 31$; Gold codes were assigned to different users. The delay of each user was chosen uniformly distributed in $[0, NT_c)$ and the DOA of each user was taken to be uniformly distributed in a sector of width $-\pi/3 \leq \theta \leq \pi/3$ radians, corresponding to one sector in a typical 3-sector cell. Furthermore, the antenna array structure was taken to be uniform linear, with

the intersensor distance equal to half the carrier wavelength ($d = \lambda_c/2$) to prevent spatial aliasing. The array response vector is then given by (2.4)

$$\mathbf{p}(\theta_k) = [1,\ e^{j\pi \sin\theta_k},\ \ldots,\ e^{j(M-1)\pi \sin\theta_k}]^\top.$$

It is important to note that the algorithm is not constrained by any specific underlying array geometry; specifying it to be a uniform linear array just simplifies some analysis. One drawback with the ULA, as mentioned in Chapter 2, is the absolute indistinguishability of two plane waves arriving from directions θ_0 and $\pi - \theta_0$, which could lead to interference from adjacent sectors. One way to rectify this problem would be to use a more complicated array geometry. However, it should be noted that since different users are distinguished by different codes, the signal vectors of two such sources would still be distinct, allowing their resolution. Thus, although we "lose" the angle of arrival dimension for resolution, the code diversity remains.

Performance .vs. K: We first highlight the increased acquisition-based capacity with increasing number of sensors M in Figure 4.1 for $M = 1$ and 3 sensors. This is most efficaciously done by examining the probability of acquisition (P_{acq}) as a function of increasing number of users in the system. Acquisition was deemed to have occurred if $|\hat{\tau_k} - \tau_k| \leq T_c$. The SNR of the desired user at each sensor was fixed at 8dB and the window length used to form subspace estimates was $L = 200$ bits. The interferers and the desired user were chosen to have the same powers (i.e., the multiple access interference, MAI = 0dB) since we were mainly interested in highlighting the effect of reduced noise-subspace dimension on the estimation. As expected, the performance with $M = 1$ drops much sooner than for $M = 3$; in fact, for $K > 10$ the single sensor performance degrades appreciably (since $N = 31$, the maximum number of user delays that can be theoretically estimated with 1 sensor is $K = \lfloor N/2 \rfloor$, which is 15 in this case; but in actuality we see that K has to be considerably less than $\lfloor N/2 \rfloor$ for acceptable acquisition performance). With just 3 sensors we see that we are able to simultaneously acquire a substantially increased number of users — the maximum number of users that can now be theoretically acquired is $K = \lfloor MN/2 \rfloor$, which is 46.

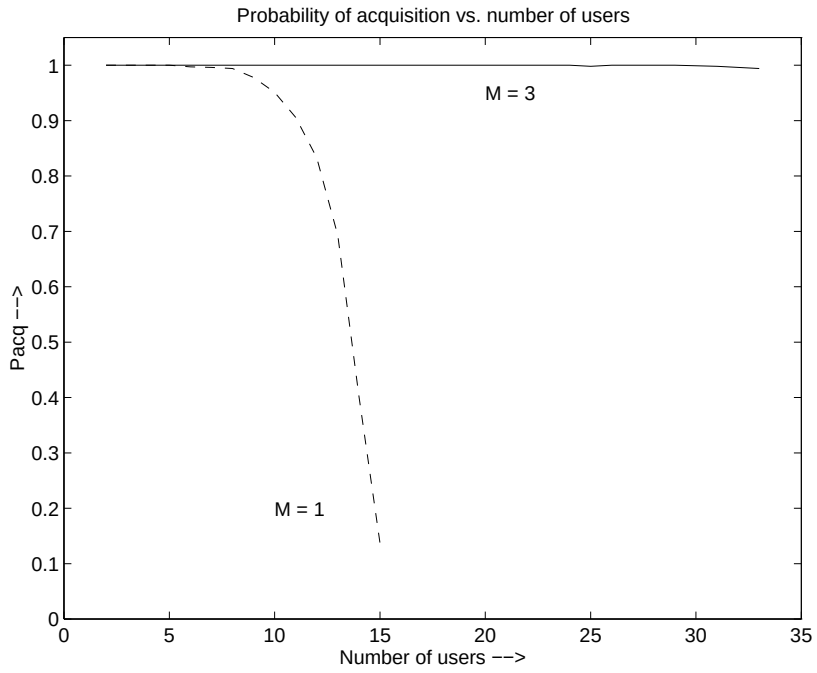
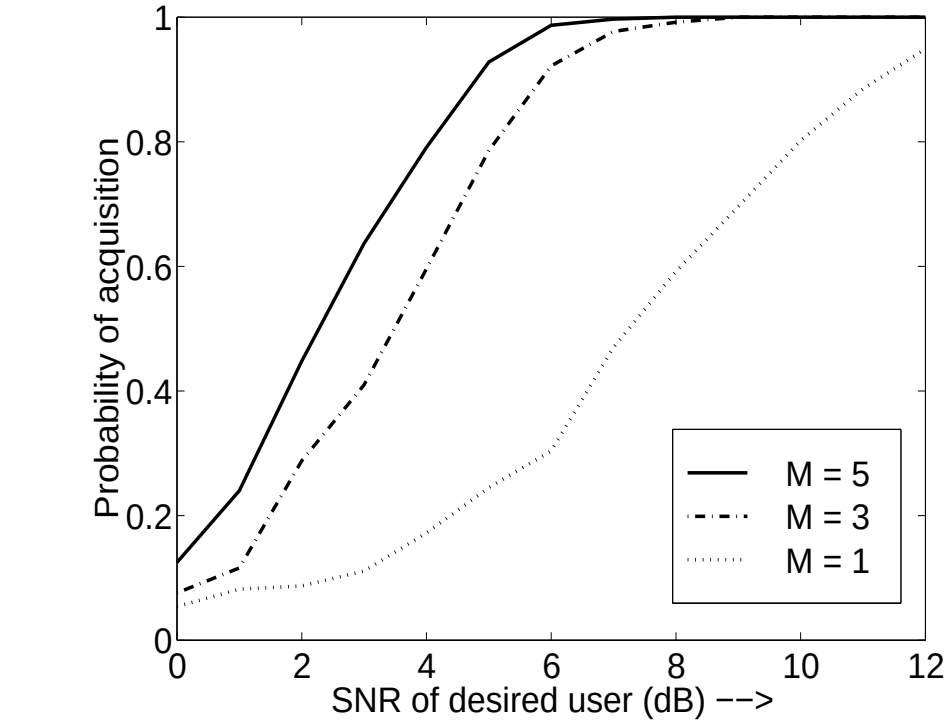
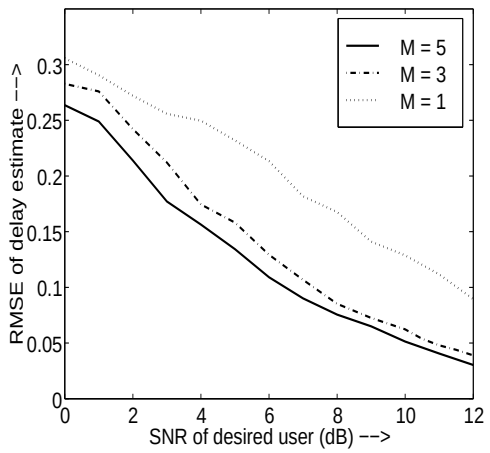


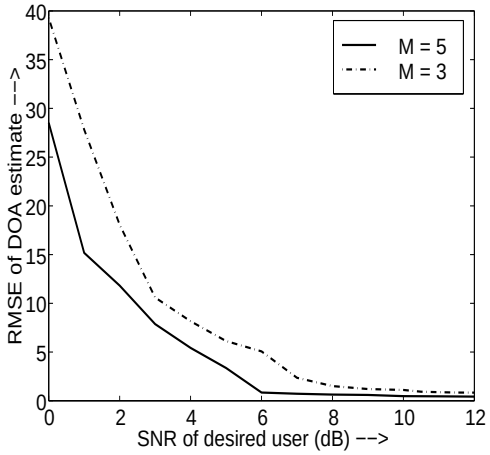
Figure 4.1 : Performance of the subspace-based acquisition algorithm (MUSIC) as a function of increasing number of users (K). The plot graphs the probability of acquisition for $M = 1, 3$ sensors. The window length, $L = 200$ bits; SNR of the desired user at each sensor = 8dB; code length = 31.



(a)

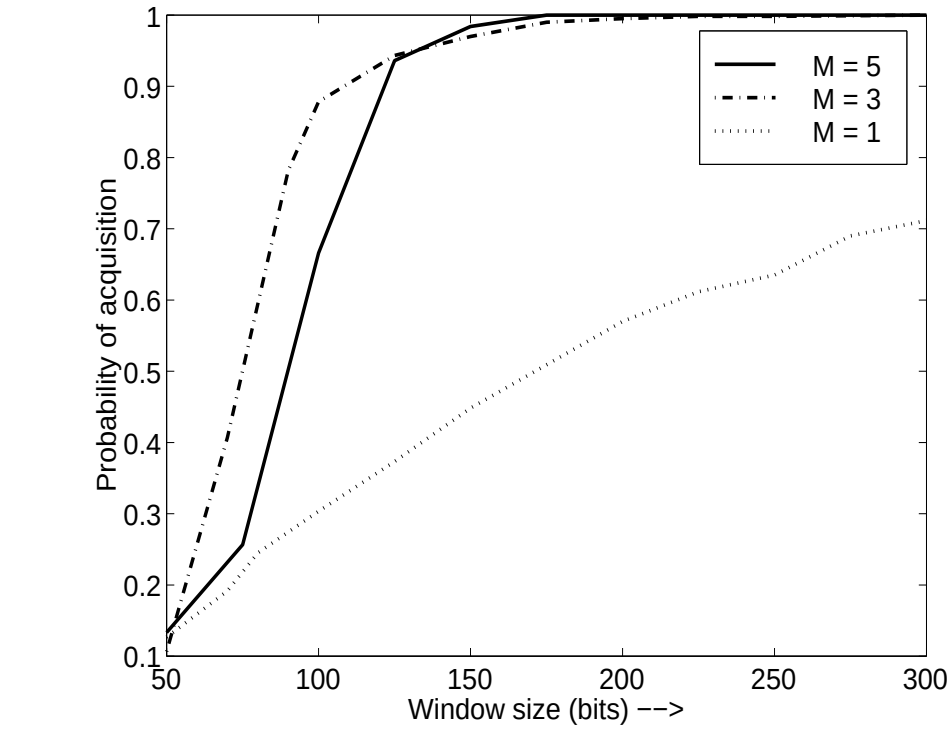


(b)

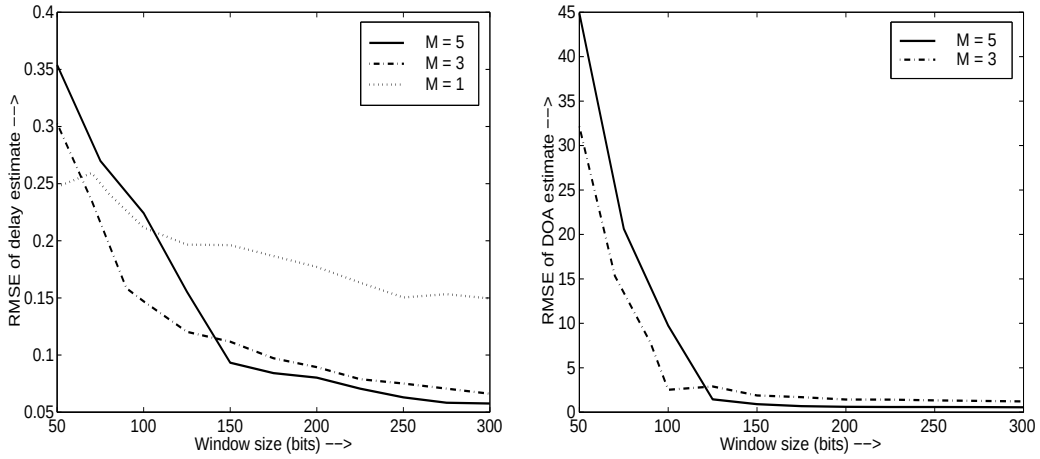


(c)

Figure 4.2 : Performance of the subspace-based acquisition algorithm (MUSIC) as a function of increasing SNR per sensor of the desired user. Plot (a) graphs the probability of acquisition for $M = 1, 3$ and 5 sensors. Plots (b) and (c) graph the RMSE of the delay and DOA for $M = 1, 3$ and 5 sensors. The window length, $L = 200$ bits; number of users in the system, $K = 10$; MAI = 20dB; code length = 31.

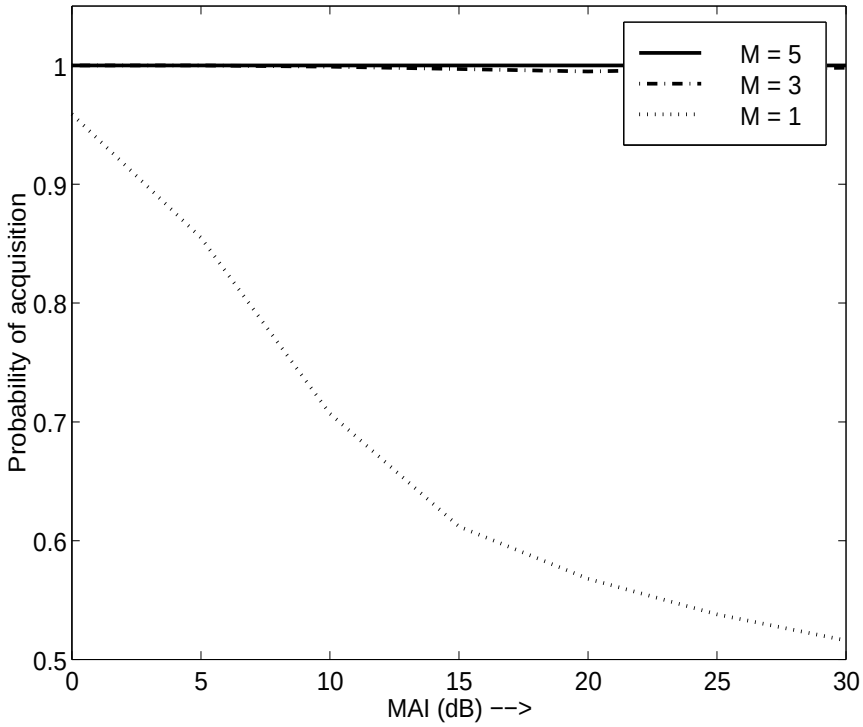


(a)

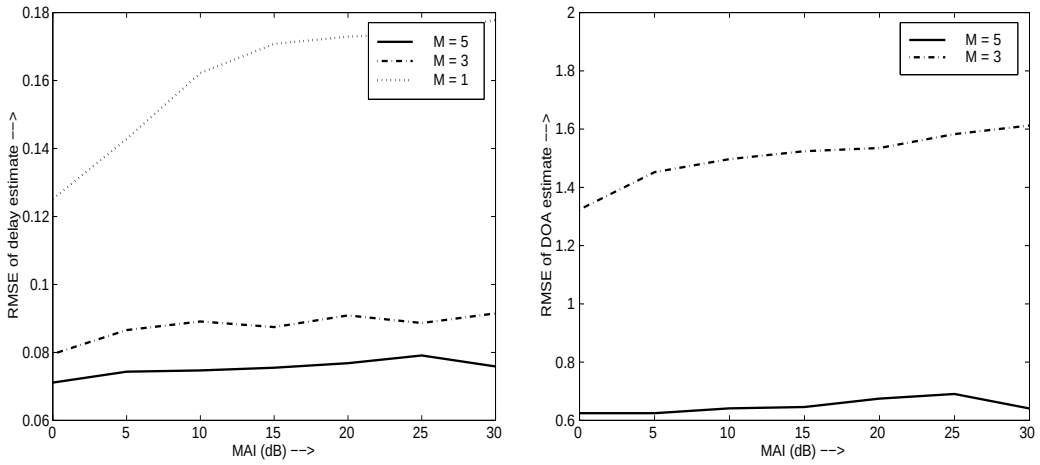


(b) (c)

Figure 4.3 : Performance of the subspace-based acquisition algorithm (MUSIC) as a function of increasing window size, L, in bits. Plot (a) graphs the probability of acquisition for $M = 1, 3$ and 5 sensors. Plots (b) and (c) graph the RMSE of the delay and DOA for $M = 1, 3$ and 5 sensors. The number of users in the system $K = 10$; MAI = 20dB; SNR of the desired user = 8dB; code length = 31.



(a)



(b) (c)

Figure 4.4 : Performance of the MUSIC algorithm as a function of increasing MAI of the interfering users. Plot (a) graphs the probability of acquisition for $M = 1, 3$ and 5 sensors. Plots (b) and (c) graph the RMSE of the delay and DOA for $M = 1, 3$ and 5 sensors. The window length $L = 200$ bits; number of users in the system $K = 10$; SNR of the desired user = 8dB; code length = 31.

Performance .vs. SNR: Figure 4.2(a) depicts the probability of acquisition as a function of the desired user's SNR at the receiver, for $M = 1, 3, 5$. The number of users was taken to be 10 so as to facilitate comparison across different number of sensors. The window size L was taken to be 200. Figures 4.2(b) and (c) depict the associated root mean squared errors (RMSE) of the delay and DOA estimates, given that acquisition has occurred. The performance is seen to increase with increased M; in fact, there is a significant increase in P_{acq} in going from 1 to 3 sensors. This is because as M increases, the dimension of the noise subspace increases, and with it the associated decrease in RMSE of the estimates [49]. However, we notice that at high SNRs there is little performance difference between $M = 3$ and $M = 5$ sensors as long as the number of users satisfies $K \leq \lfloor MN/2 \rfloor$; lower values of SNR, however, entail an appreciable performance difference.

Performance .vs. L: We next depict the performance of the delay and DOA estimates with increasing window length (number of observations) L. Since subspace based techniques depend heavily on the structure of the observation correlation matrix, we expect the accuracy of the estimates to be poor when the correlation matrix is rank deficient. This is confirmed by the plots in Figure 4.3(a), (b) and (c) which portray P_{acq} and the RMSE of the delay and DOA estimates respectively. As a result of the dependence on the dimensionality of the correlation matrix, the more the number of sensors the larger the required window for forming the correlation matrix estimate. As concluded earlier, for higher values of SNR we observe that a ULA with just 3 sensors performs comparably with its 5 sensor sibling.

Performance .vs. MAI: Our next goal is to examine the behavior of the MUSIC estimates with increasing multiple access interference (MAI). We say an estimation algorithm is near-far resistant if the accuracy of the estimates is invariant to increasing interferers' powers. The number of users in this simulation was again fixed at 10 and the window size L was taken to be 200. Figures 4.4(a), (b) and (c) depict P_{acq} and the root mean squared errors of the delay and DOA estimates as a function of increasing

MAI for $M = 1, 3, 5$. With just one sensor we observe that the performance degrades alarmingly with increased interference. Thus, it is only if the number of users is *well within* the dimensional limitations of the subspace based algorithms that we observe the algorithms to be near-far resistant.

Having established the increase in acquisition capacity as well as estimate accuracy with increasing number of sensors in the receiver, we now compare the performance of the exact and approximate MUSIC algorithms described in Section 4.2. The primary difference between the two methods is in the number of points in the θ coordinate at which the function is evaluated; the exact method involves minimizing the ideal MUSIC norm at *each* point on a fine grid in the θ variable, while the second, more approximate, method minimizes a bi-quadratic approximation to the MUSIC norm over a small number of disjoint intervals in the θ variable.

The two methods are compared by evaluating their performance in terms of P_{acq} and the RMSEs as a function of increasing SNR of the desired user. The number of observations used to form subspace estimates is fixed at $L = 200$. We evaluate the algorithms for different numbers of users, $K = 5, 15$ and 25 and for a fixed number of sensors, $M = 5$. This is summarized in Figure 4.5(a), (b) and (c) which depict P_{acq} and the root mean squared errors of the delay and DOA estimates respectively. Each plot is generated for $K = 5, 15$ and 25, thereby showing the behavior of the approximate estimate as a function of increasing number of users. As the number of users increases, we observe the expected decrease in estimate accuracies. We also observed that as K was increased, the number of intervals, M_I, required to get acceptable performance from the approximate Taylor series technique increased. Specifically, for $K = 5, 15$ and 25 users, we used $M_I = 5, 6$ and 7 intervals respectively. This can be explained by noting that as the number of users increases, the cross-correlations with the desired user increase, causing the MUSIC norm to vary more across the delay and DOA parameters. The number of intervals required to locally capture the function behavior therefore increase. As seen from the plots, the performance of the Taylor

series method is slightly inferior to the exact search method but the degradation is acceptable in light of the computational savings. It was also observed that the number of iterations required to fall within a suitably small neighborhood of the minimum in each region was only *two*. Therefore, if M_θ is the number of points in the search grid for the θ variable in the exact MUSIC method, the computational savings obtained for comparable accuracy in the estimates is on the order of $M_\theta/(2M_I)$. If the grid refinement is 0.5 degree with the sector width being 120°, and the number of Taylor series intervals chosen is $M_I = 6$, the computational savings is on the order of a factor of 20.

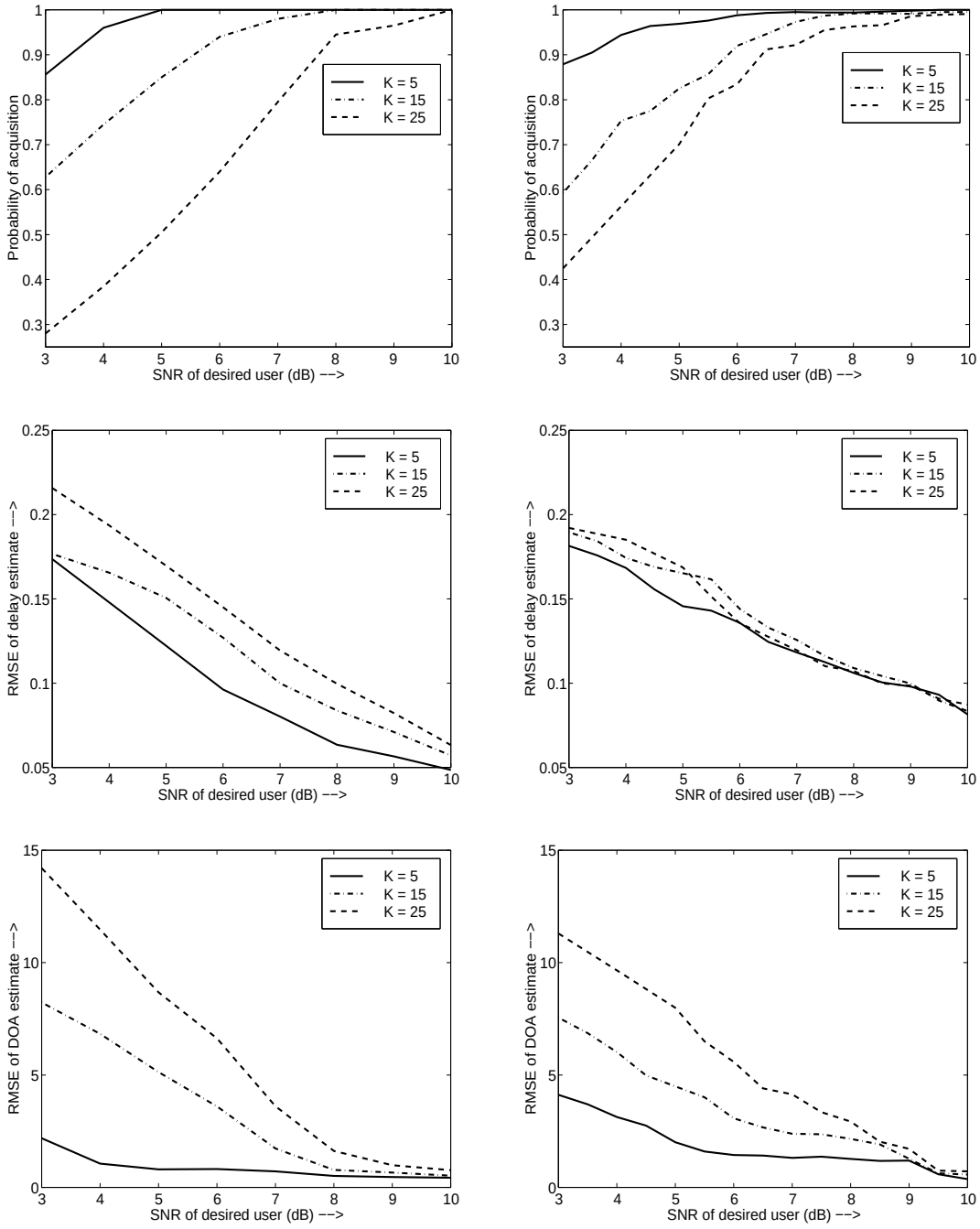


Figure 4.5 : Comparing the two estimation algorithms (MUSIC and approximate MUSIC) versus SNR for $M = 5$ sensors and $K = 5, 15$ and 25 users. The plots in column (a) represent the exact MUSIC algorithm while the plots on the right, in column (b), depict the Taylor series approximation MUSIC algorithm.

Chapter 5

Linear Detectors with Antenna Arrays

In the previous section we examined the aquisition based capacity of an antenna array receiver and established that the capacity increases as a function of the number of sensors. The parameters estimated were the delay τ_k and direction of arrival (DOA) θ_k of each user. In this section we examine the benefits of incorporating multiple sensors in linear multiuser detection. Specifically we analyze two detector structures, the decorrelating detector and the minimum mean squared error (MMSE) detector, and demonstrate the advantages of added spatial diversity via multiple antennas. The results obtained here are consistent with those obtained in [22]; we extend their study to the asynchronous combining of decision statistics. We consider linear multiuser detectors due to their simplicity and the relative ease of implementation and analysis; furthermore, we restrict ourselves to the realm of *single-shot* detectors (the gains derived from multiple sensors in this case can be extrapolated to sequence detection).

The observation vector at a single sensor (say, at sensor 1) can be given by $\mathbf{r}_i^{(1)} = \mathbf{AWb}_i + \boldsymbol{\nu}_i^{(1)}$. Since the matrix $\mathbf{A}$ contains the users' codes, we can now define the single sensor bank of matched filter outputs by

$$\mathbf{y}_i^{(1)} = \mathbf{A}^\top \mathbf{r}_i^{(1)} = \mathbf{CWb}_i + \mathbf{A}^\top \boldsymbol{\nu}_i^{(1)}, \tag{5.1}$$

where $\mathbf{A}$ is the code matrix, $\mathbf{C} = \mathbf{A}^\top\mathbf{A}$ is the code correlation matrix, $\mathbf{W}$ is a $2K \times 2K$ diagonal matrix of received complex amplitudes, the received bit vector $\mathbf{b}_i = [b_{1,i-1}\ b_{1,i} \ldots b_{K,i-1}\ b_{K,i}]^\top$ and $\boldsymbol{\nu}_i^{(1)} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Now, recall from (2.13) that the received observation vector across the array can be written as

$$\mathbf{r}_i = \mathcal{A}(\boldsymbol{\tau},\ \boldsymbol{\theta})\ \mathbf{Wb}_i + \boldsymbol{\nu}_i, \quad \boldsymbol{\nu}_i \sim N(\mathbf{0}, \sigma^2\mathbf{I}),$$

For simplicity we assume that the noise is temporally and spatially white (incorporating a known covariance structure in this case poses no other problem that complicating the algebra). Because the system is asynchronous, each of the two "halves" of each user's code, $\mathbf{c}_k^R, \mathbf{c}_k^L$, can be considered as a virtual user's code. Given the estimates of the delay and DOA of each user (obtained in the previous sections), we can then obtain the outputs of a bank of matched filters matched to the $2K$ signal vectors $\{\mathbf{a}_k^R, \mathbf{a}_k^L\}_{k=1}^K$, which are defined in (2.16). If we denote by $\mathbf{y}_i^{(\mathsf{M})}$ (not to be confused with the output of the M^{th} sensor) the bank of matched filter outputs of the M-sensor array at time iT

$$\begin{aligned} \mathbf{y}_i^{(\mathsf{M})} = \mathcal{A}^H \mathbf{r}_i &= \mathcal{A}^H \mathcal{A}\, \mathbf{W}\, \mathbf{b}_i \;+\; \mathcal{A}^H \boldsymbol{\nu}_i \\ &= \mathcal{C}\mathbf{W}\mathbf{b}_i \;+\; \mathcal{A}^H \boldsymbol{\nu}_i, \end{aligned} \tag{5.2}$$

where $\mathcal{C} = \mathcal{A}^H\mathcal{A}$ represents the multi-sensor code correlation matrix. We can then rewrite $\mathbf{y}_i^{(\mathsf{M})}$ in (5.2), by recalling the definition of $\mathcal{A}$ from (2.14), as

$$\mathbf{y}_i^{(\mathsf{M})} = \boldsymbol{\Phi}^{(1)^H}\mathbf{A}^\top \mathbf{r}_i^{(1)} + \boldsymbol{\Phi}^{(2)^H}\mathbf{A}^\top \mathbf{r}_i^{(2)} + \cdots + \boldsymbol{\Phi}^{(\mathsf{M})^H}\mathbf{A}^\top \mathbf{r}_i^{(\mathsf{M})}. \tag{5.3}$$

Since all the delay information is contained in $\mathbf{A}$ and all the DOA information is captured in $\boldsymbol{\Phi}$, this can be interpreted as a code-matched filter operation at each sensor, followed by an "angle-matched" filter or conventional beamformer (refer to Figure 5.1). To see this, we recall the structure of $\boldsymbol{\Phi}^{(\ell)}$ from (2.14) and realize that successive pairs of entries of the $(2K \times 1)$ vector $\mathbf{y}_i^{(\mathsf{M})}$ can be written as (ignoring the time index i)

$$\begin{pmatrix} \left(\mathbf{y}^{(\mathsf{M})}\right)_{2k-1} \\ \left(\mathbf{y}^{(\mathsf{M})}\right)_{2k} \end{pmatrix} = \begin{pmatrix} \mathbf{p}_k^H \mathbf{v}_k^R \\ \mathbf{p}_k^H \mathbf{v}_k^L \end{pmatrix}, \tag{5.4}$$

where the notation $(\mathbf{v})_k$ is used to denote the k^{th} element in the vector $\mathbf{v}$. The vectors $\mathbf{v}_k^R$, $\mathbf{v}_k^L \in \mathbb{C}(M \times 1)$ and contain the right and left code-matched filter outputs from the M sensors,

$$\begin{aligned} \mathbf{v}_k^R &= [\mathbf{c}_k^{R^H}\mathbf{r}^{(1)}, \ldots, \mathbf{c}_k^{R^H}\mathbf{r}^{(\mathsf{M})}]^\top \\ \mathbf{v}_k^L &= [\mathbf{c}_k^{L^H}\mathbf{r}^{(1)}, \ldots, \mathbf{c}_k^{L^H}\mathbf{r}^{(\mathsf{M})}]^\top. \end{aligned}$$

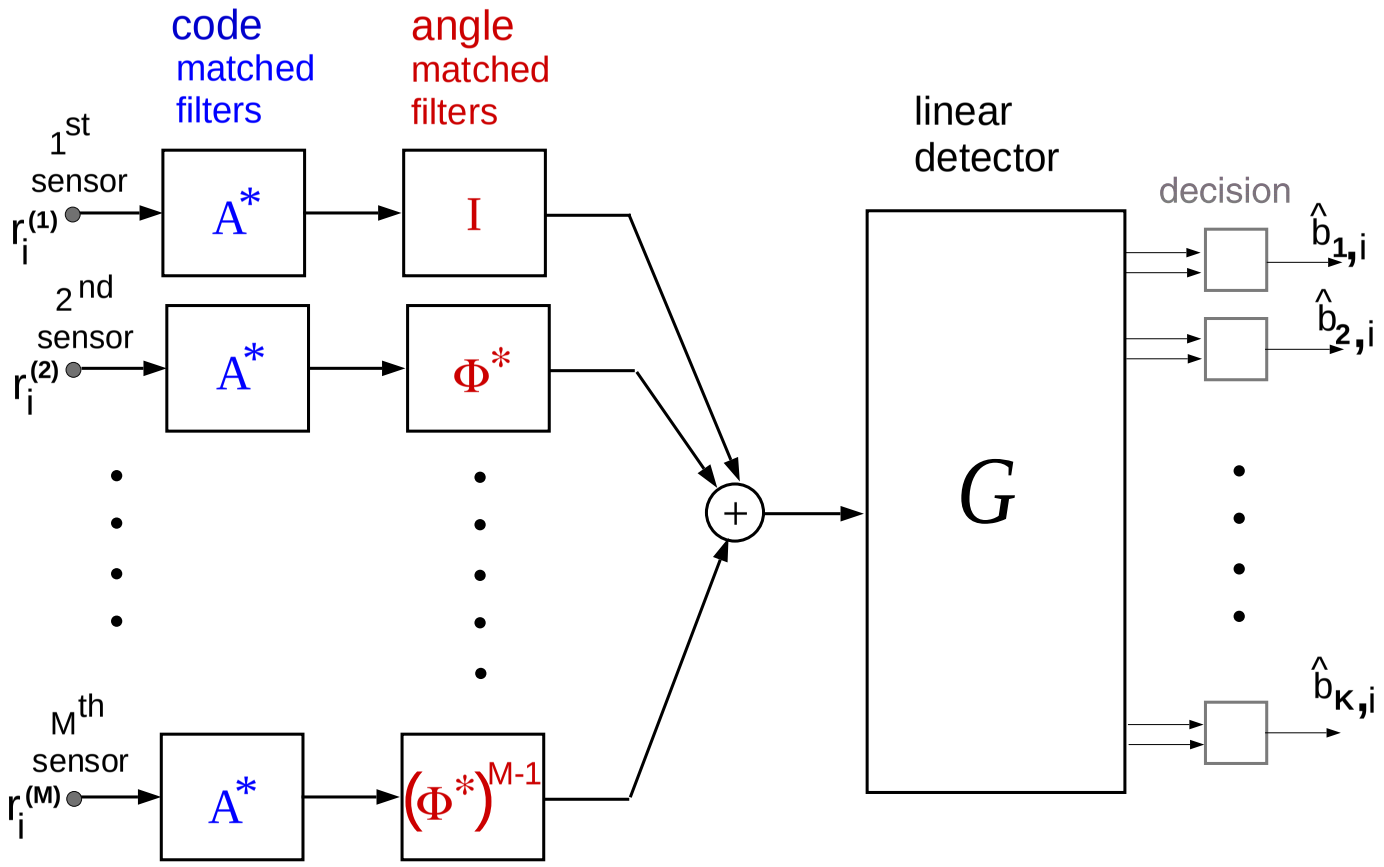


Figure 5.1 : Schematic of a multisensor linear detector, represented by $\mathcal{G}$. The blocks $\mathbf{A}$ and $\mathbf{\Phi}^{m-1}$ represent the code and angle matched filters. The linear detector $\mathcal{G}$ transforms a length $2K$ vector into a length $2K$ vector. Successive pairs of outputs are linearly combined to produce bit estimates.

Thus, the compound matched filters at each sensor are matched to the users' delays as well as directions of arrival, which would be optimum for the multisensor case with one user in the presence of AWGN only.

5.1 Linear Detectors

Our aim in this section is to quantify the gains obtained from multiple sensors at the receiver. A single-sensor linear detector, $\mathbf{G} \in \mathbb{C}^{K\times K}$, in a synchronous system operates on the vector of matched filter outputs at one sensor, say $\mathbf{y}_i^{(1)}$, to yield a vector of decision statistics for all the users

$$\mathbf{z}_i^{(1)} = \mathbf{G}\mathbf{y}_i^{(1)} = \mathbf{GCW}\mathbf{b}_i + \mathbf{GA}^\top \boldsymbol{\nu}_i^{(1)}. \tag{5.5}$$

Under the assumption of BPSK modulation and restricting ourselves to the first user's bit without loss of generality, we then obtain the required bit estimate as $\hat{\mathbf{b}}_i = \text{sign}(\mathbf{z}_i^{(1)})_1$. The output signal-to-interference ratio (SIR), which is defined as the ratio of the signal power to the interference power at the detector output, where the interference includes the additive noise component, can serve as a measure of detector performance. For large code lengths and number of users, the probability of error for user 1 at the output of the detector can be closely approximated as $Pe_1 = Q(\sqrt{SIR})$, where $Q(\cdot)$ is defined by $Q(x) = 1/\sqrt{2\pi}\ \int_x^\infty \exp(-t^2/2)dt$. From (5.5) we can write $(\mathbf{z}_i^{(1)})_1$ as

$$(\mathbf{z}_i^{(1)})_1 = (\mathbf{GC})_{1,1} w_1 b_{1,i} + \sum_{k=2}^{K} (\mathbf{GC})_{1,k} w_k b_{k,i} + (\tilde{\nu}_i)_1\,, \tag{5.6}$$

where $(\tilde{\nu}_i)_1$ is the first entry of the vector $\mathbf{GA}^\top \boldsymbol{\nu}_i^{(1)}$ with variance $\sigma^2(\mathbf{GCG}^\top)_{1,1}$. The SIR can now be written as

$$SIR = \frac{(\mathbf{GC})_{1,1}^2 w_1^2}{\sigma^2(\mathbf{GCG}^\top)_{1,1} + \sum_{k=2}^{K} (\mathbf{GC})_{1,k}^2 w_k^2}. \tag{5.7}$$

In the asynchronous case, the current bit $b_{1,i}$ influences two successive observation vectors $\mathbf{r}_i$ and $\mathbf{r}_{i+1}$ (remember that $\mathbf{b}_i = [b_{1,i-1}, \underline{b_{1,i}}, \ldots, b_{K,i-1}, b_{K,i}]^\top$ and $\mathbf{b}_{i+1} =$

$[\underline{b_{1,i}}, b_{1,i+1}, \ldots, b_{K,i}, b_{K,i+1}]^\top$). The decision statistic of the first user should reflect this dependence and hence is given by

$$(\mathbf{z}_i^{(1)})_1 = (1-\alpha)\big(\mathbf{G}\mathbf{y}_i^{(1)}\big)_2 + \alpha\big(\mathbf{G}\mathbf{y}_{i+1}^{(1)}\big)_1. \tag{5.8}$$

In the above equation $\mathbf{G} \in \mathbb{C}^{2K\times 2K}$ which reflects the additional virtual user for each of the K actual transmitting users. The real number $\alpha \in [0,1]$ and can be chosen to optimize an appropriate objective function. From the dependence of (5.8) on α we might be tempted to choose $\alpha = \tau_1/T$. However, this not entirely unreasonable choice of α does not optimize any apparent criterion. We choose α to minimize the SIR at the output of the detector. In the asynchronous case, the SIR at the output of $\mathbf{G}$, albeit more algebraically cumbersome, can be calculated as in the synchronous case and is given by

$$SIR(\alpha) = \frac{\big((1-\alpha)(\mathbf{GC})_{2,2} + \alpha(\mathbf{GC})_{1,1}\big)^2 w_1^2}{\sigma^2\Big((1-\alpha)^2(\mathbf{GCG}^\top)_{2,2} + \alpha^2(\mathbf{GCG}^\top)_{1,1}\Big) + \mathcal{I}}. \tag{5.9}$$

The multiple access interference (MAI) term $\mathcal{I}$ is given by

$$\mathcal{I} = \alpha^2\mathcal{I}_1 + (1-\alpha)^2\mathcal{I}_2 + 2\alpha(1-\alpha)\mathcal{I}_3, \tag{5.10}$$

with the individual components defined by

$$\mathcal{I}_1 = \sum_{j=2}^{2K} (\mathbf{GC})_{1,j}^2 \mathbf{W}_{j,j}^2 \tag{5.11}$$

$$\mathcal{I}_2 = \sum_{\substack{j=1\\ j\neq 2}}^{2K} (\mathbf{GC})_{2,j}^2 \mathbf{W}_{j,j}^2 \tag{5.12}$$

$$\mathcal{I}_3 = (\mathbf{GC})_{1,(3:2K)} \left(\widetilde{\mathbf{W}}^2 \otimes \mathbf{E}\right) (\mathbf{GC})_{2,(3:2K)}^\top, \tag{5.13}$$

where $\widetilde{\mathbf{W}}$ is a $(K-1\times K-1)$ diagonal matrix of all the users' amplitudes except user 1, $\mathbf{E} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and the notation $(\mathbf{A})_{i,(j:k)}$ represents a row vector containing

the elements $(\mathbf{A}_{i,j}, \ldots, \mathbf{A}_{i,k})$. We can now minimize the SIR in (5.9) as a function of α as follows. Define the following terms

$$n_1 = (\mathbf{GC})_{1,1}^2 \; ; \qquad n_2 = (\mathbf{GC})_{2,2}^2 \; ; \qquad n_3 = (\mathbf{GC})_{1,1}(\mathbf{GC})_{2,2}$$
$$d_1 = \sigma^2(\mathbf{GCG}^\top)_{1,1} + \mathcal{I}_1 \; ; \qquad d_2 = \sigma^2(\mathbf{GCG}^\top)_{2,2} + \mathcal{I}_2 \; ; \qquad d_3 = \mathcal{I}_3$$

Then, the optimum α is given by $\alpha_* = 1/(1+t)$ where t is the root of the quadratic equation

$$(d_2 n_3 - d_3 n_2)t^2 + (d_2 n_1 - d_1 n_2)t + (d_3 n_1 - d_1 n_3) = 0, \tag{5.14}$$

such that $\alpha_* \in [0, 1]$. It can be shown that the above quadratic equation has both real roots and in fact, only one root is positive. Thus there is only one choice for α_*.

The multisensor extension of the single sensor linear detector, $\mathbf{G}$, follows quite naturally (the reader is referred to Figure 5.1 for a schematic of multisensor linear detector). Analogous to (5.8), the decision statistic for the first user can be given by

$$(\mathbf{z}_i^{(\mathsf{M})})_1 = (1-\alpha)\big(\mathcal{G}\mathbf{y}_i^{(\mathsf{M})}\big)_2 + \alpha\big(\mathcal{G}\mathbf{y}_{i+1}^{(\mathsf{M})}\big)_1, \tag{5.15}$$

where $\mathcal{G} \in \mathbb{C}^{2K \times 2K}$ is the M-sensor linear detector and $\mathbf{y}_i^{(\mathsf{M})}$ is the vector of multisensor matched filter outputs from (5.2). The corresponding SIR, as in the single sensor asynchronous case, can be written as

$$SIR(\alpha) = \frac{\left|(1-\alpha)(\mathcal{GC})_{2,2} + \alpha(\mathcal{GC})_{1,1}\right|^2 w_1^2}{\sigma^2\Big((1-\alpha)^2(\mathcal{GCG}^H)_{2,2} + \alpha^2(\mathcal{GCG}^H)_{1,1}\Big) + \mathcal{I}}, \tag{5.16}$$

where the MAI term $\mathcal{I} = \alpha^2\mathcal{I}_1 + (1-\alpha)^2\mathcal{I}_2 + 2\alpha(1-\alpha)\mathrm{Re}\{\mathcal{I}_3\}$, with each component defined by

$$\mathcal{I}_1 = \sum_{j=2}^{2K} |\mathcal{GC}|_{1,j}^2 \mathbf{W}_{j,j}^2$$
$$\mathcal{I}_2 = \sum_{\substack{j=1 \\ j\neq 2}}^{2K} |\mathcal{GC}|_{2,j}^2 \mathbf{W}_{j,j}^2$$
$$\mathcal{I}_3 = (\mathcal{GC})_{1,(3:2K)} \left(\widetilde{\mathbf{W}}^2 \otimes \mathbf{E}\right) (\mathcal{GC})_{2,(3:2K)}^H.$$

The optimum value for α is found exactly as in the single sensor case as the solution to a quadratic equation of the form (5.14).

In the sequel we will investigate two linear multiuser receivers and demonstrate the performance enhancement achieved by employing multiple sensors at the base station receiver. To maintain the near-far resistant properties of the subspace based estimators, we only consider linear detectors that are also near-far resistant; specifically, we restrict ourselves to the decorrelating detector [34,50] and the linear MMSE detector [37,35].

5.1.1 Decorrelating Detector

The decorrelating detector $\mathcal{G}_{\text{dec}}$ can be viewed as the unconstrained maximum likelihood (or least squares) estimate of $\mathcal{G}$ in additive white gaussian noise. From (5.2) we see that $\mathcal{G}_{\text{dec}} = \mathcal{C}^{-1}$, where the multisensor code correlation matrix $\mathcal{C} = \mathcal{A}^H\mathcal{A}$. Following (5.2) and (5.15), we can write the decision statistic for the i^{th} bit of the first user as

$$(\mathbf{z}_i^{(\mathsf{M})})_1 = (1-\alpha)\big(\mathbf{W}\mathbf{b}_i\big)_2 + \alpha\big(\mathbf{W}\mathbf{b}_{i+1}\big)_1 + \tilde{\nu}, \tag{5.17}$$

where the additive gaussian noise term is

$$\tilde{\nu} = (1-\alpha)(\mathcal{C}^{-1}\mathcal{A}^H\boldsymbol{\nu}_i)_{2,2} + \alpha(\mathcal{C}^{-1}\mathcal{A}^H\boldsymbol{\nu}_{i+1})_{1,1}. \tag{5.18}$$

The MAI is eliminated at the output of the detector ($\mathcal{I} = 0$) and the individual users' bits are decorrelated from one another at the cost of noise enhancement [34]. We also note that the detector does not require amplitude information making it a natural choice when the amplitude is rapidly varying. From (5.9) the SIR can be given by

$$SIR^{(\mathsf{M})}(\alpha) = \frac{w_1^2}{\sigma^2\left((1-\alpha)^2\mathcal{C}_{2,2}^{-1} + \alpha^2\mathcal{C}_{1,1}^{-1}\right)}.$$

From (5.14), the optimum α can be calculated simply as

$$\alpha_* = \frac{\mathcal{C}_{2,2}^{-1}}{\mathcal{C}_{1,1}^{-1} + \mathcal{C}_{2,2}^{-1}} \tag{5.19}$$

and the maximum SIR is now given by

$$SIR_*^{(\mathsf{M})} = \frac{w_1^2}{\sigma^2}\left(\frac{1}{\mathcal{C}_{1,1}^{-1}} + \frac{1}{\mathcal{C}_{2,2}^{-1}}\right) \tag{5.20}$$

The probability of error in this case is completely specified by the SIR since $\mathcal{I} = 0$. Another well known measure of detector performance is the *asymptotic efficiency* (η_1) for a particular user. If we define the effective energy of user 1, $e_1(\sigma)$, as the energy that user 1 would have to transmit, in the *absence* of interfering users, to achieve the same bit-error rate as in the multiple access channel, then the asymptotic efficiency can be defined as [50]

$$\eta_1 = \lim_{\sigma \to 0} e_1(\sigma)/w_1\,. \tag{5.21}$$

Thus, it is a measure of the loss in efficiency that user 1 experiences due to the presence of MAI and very little noise (to emphasize the MAI). From (5.17), (5.18) and [34] the probability of error in detecting the first user's bits, $Pe_1^{(\mathsf{M})}$, and the asymptotic efficiency, $\eta_1^{(\mathsf{M})}$, using M sensors are given by

$$Pe_1^{(\mathsf{M})} = Q\left(SIR_*^{(\mathsf{M})}\right), \qquad \eta_1^{(\mathsf{M})} = \frac{1}{\mathcal{C}_{1,1}^{-1}} + \frac{1}{\mathcal{C}_{2,2}^{-1}}, \tag{5.22}$$

while the corresponding quantities for a single sensor receiver are

$$Pe_1^{(1)} = Q\left(SIR^{(1)}\right), \qquad \eta_1^{(1)} = \frac{1}{\mathbf{C}_{1,1}^{-1}} + \frac{1}{\mathbf{C}_{2,2}^{-1}}, \tag{5.23}$$

where $SIR^{(1)} = \frac{w_1^2}{\sigma^2}\left(1/\mathbf{C}_{1,1}^{-1} + 1/\mathbf{C}_{2,2}^{-1}\right)$.

In order to obtain more insight into the gain obtained from multiple sensors, let us examine the structure of $\mathcal{C}$ more closely. From (2.14) we can write $\mathcal{C}$ as

$$\begin{aligned}\mathcal{C} &= \left[\mathbf{\Phi}^{(1)^H}\mathbf{A}^\top \;\cdots\; \mathbf{\Phi}^{(\mathsf{M})^H}\mathbf{A}^\top\right]\cdot\left[\mathbf{\Phi}^{(1)^H}\mathbf{A}^\top \;\cdots\; \mathbf{\Phi}^{(\mathsf{M})^H}\mathbf{A}^\top\right]^H \\ &= \mathbf{\Phi}^{(1)^H}\mathbf{C}\mathbf{\Phi}^{(1)} + \cdots + \mathbf{\Phi}^{(\mathsf{M})^H}\mathbf{C}\mathbf{\Phi}^{(\mathsf{M})},\end{aligned} \tag{5.24}$$

where $\mathbf{C}$ represents the single sensor code correlation matrix, $\mathbf{C} = \mathbf{A}^\top\mathbf{A}$. We now establish a few properties of $\mathcal{C}$ via the following theorem.

Theorem 5.1

Given a correlation matrix $\mathcal{C}$ defined by (5.24), the following properties hold :

(a) $\mathcal{C}$ is symmetric positive semi-definite (s.p.d), denoted henceforth, with a little abuse of notation, by $\mathcal{C} \succeq \mathbf{0}$.

(b) For a uniform linear array (ULA), the diagonal elements of the multisensor and single sensor correlation matrices are related by $(\mathcal{C}^{-1})_{n,n} \leq (\mathbf{C}^{-1})_{n,n}$.

(c) In fact, we can tighten the inequality: $(\mathcal{C}^{-1})_{n,n} \leq \frac{1}{M} \cdot (\mathbf{C}^{-1})_{n,n}$.

Proof :

1. Proposition (a) follows immediately from the definition of $\mathcal{C}$ as $\mathcal{A}^H\mathcal{A}$.

2. To prove Proposition (b), we make note of a few properties of s.p.d matrices [41]. If $\mathbf{X}, \mathbf{Y} \succeq \mathbf{0}$, we have $\mathbf{X^{-1}}, (\mathbf{X}+\mathbf{Y}) \succeq \mathbf{0}$ and hence $(\mathbf{X}+\mathbf{Y})^{-1}, (\mathbf{X^{-1}} + \mathbf{Y^{-1}})^{-1} \succeq \mathbf{0}$. Also, we have $\mathbf{Z} \stackrel{\triangle}{=} \mathbf{X^{-1}}(\mathbf{X^{-1}} + \mathbf{Y^{-1}})^{-1}\mathbf{X^{-1}} \succeq \mathbf{0}$ (this follows because $\mathbf{A}^H\mathbf{X}\mathbf{A} \succeq \mathbf{0}$ for any matrix $\mathbf{A}$, and $\mathbf{X^{-1}}$ is symmetric). Hence $\mathbf{Z}_{n,n} \geq 0$ (diagonal elements of s.p.d matrices are ≥ 0). Therefore, expanding $(\mathbf{X}+\mathbf{Y})^{-1}$ as

$$(\mathbf{X}+\mathbf{Y})^{-1} = \mathbf{X^{-1}} - \underbrace{\mathbf{X^{-1}}(\mathbf{X^{-1}} + \mathbf{Y^{-1}})^{-1}\mathbf{X^{-1}}}_{\mathbf{Z} \succeq 0}, \tag{5.25}$$

we see that $(\mathbf{X}+\mathbf{Y})^{-1}_{n,n} \leq \mathbf{X}^{-1}_{n,n}$. From (5.24), we have

$$\mathcal{C}^{-1} = \left(\mathbf{\Phi}^{(1)^H}\mathbf{C}\mathbf{\Phi}^{(1)} + \cdots + \mathbf{\Phi}^{(\mathsf{M})^H}\mathbf{C}\mathbf{\Phi}^{(\mathsf{M})}\right)^{-1}.$$

From (2.4) we see that the m^{th} element of the array response vector $\mathbf{p}_k$ for a ULA is given by, $\mathbf{p}_{k,m} = \exp(j(m-1)\pi \sin\theta_k) = \exp(j(m-1)\phi_k)$, where $\phi_k \stackrel{\triangle}{=} \pi \sin\theta_k$. This implies that $\mathbf{\Phi}^{(m)} = \mathbf{\Phi}^{m-1}$, where

$$(\mathbf{\Phi}^{(m)})_{2k-1,2k-1} = (\mathbf{\Phi}^{(m)})_{2k,2k} = \exp(j(m-1)\phi_k).$$

Therefore $\mathbf{\Phi}^{(1)} = \mathbf{I}$ and therefore ${\mathbf{\Phi}^{(1)}}^H \mathbf{C} \mathbf{\Phi}^{(1)} = \mathbf{C}$. If we now define $\mathbf{X} \stackrel{\triangle}{=} \mathbf{C}$ and $\mathbf{Y} \stackrel{\triangle}{=} ({\mathbf{\Phi}}^H \mathbf{C} \mathbf{\Phi} + \cdots + {\mathbf{\Phi}^{M-1}}^H \mathbf{C} \mathbf{\Phi}^{M-1})$, we see that $\mathbf{X}, \mathbf{Y} \succeq \mathbf{0}$. This follows from the definition of $\mathbf{C}$ as $\mathbf{A}^\top \mathbf{A}$ and the properties of s.p.d matrices discussed above. Hence, directly applying the result obtained from (5.25), we have that

$$\begin{aligned} \left({\mathbf{\Phi}^{(1)}}^H \mathbf{C} \mathbf{\Phi}^{(1)} + \cdots + {\mathbf{\Phi}^{(M)}}^H \mathbf{C} \mathbf{\Phi}^{(M)} \right)^{-1}_{n,n} &= \left(\mathbf{C} + \cdots + {\mathbf{\Phi}^{M-1}}^H \mathbf{C} \mathbf{\Phi}^{M-1} \right)^{-1}_{n,n} \\ &\leq \mathbf{C}^{-1}_{n,n}, \end{aligned} \tag{5.26}$$

proving (b).

3. To prove (c), we make use of an interesting property of s.p.d matrices [51]. Given an integer q, we denote by $\mathcal{M}_q$ the space of all $q \times q$ matrices with complex entries and we denote by $\mathcal{H}_q(a,b)$ the space of all hermitian matrices with spectra in the interval $(a,b) \subset \mathbb{R}$. It can be shown that $\mathcal{H}_q(a,b)$ is a convex set [51].

   Definition 5.1 A twice continuously differentiable function $f(\cdot) : (a,b) \to \mathbb{R}$ is said to be a convex matrix function on $\mathcal{H}_q(a,b)$ if $f : \mathcal{H}_q(a,b) \to \mathcal{H}_q(-\infty, \infty)$ satisfies

$$(1-\alpha) f(\mathbf{A}) + \alpha f(\mathbf{B}) \succeq f\big((1-\alpha)\mathbf{A} + \alpha \mathbf{B}\big)$$

   for all $\mathbf{A}, \mathbf{B} \in \mathcal{H}_q(a,b)$ and all $\alpha \in [0,1]$. The notation $\mathbf{A} \succeq \mathbf{B}$ signifies that $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$.

   It can be shown that the function $f(t) = 1/t$ is a convex matrix function on $\mathcal{H}_q(0,\infty), q = 1, 2, \ldots,$ i.e., on the space of Hermitian positive definite matrices of all dimensions [51, pp.547–549]. Therefore, since each of the matrices ${\mathbf{\Phi}^{(m)}}^H \mathbf{C} \mathbf{\Phi}^{(m)} \in \mathcal{H}_{2K}(0,\infty),\ m = 1, \ldots, M$, we can use the convexity of the

matrix function $1/t$ on $\mathcal{H}_{2K}(0,\infty)$ to obtain

$$\frac{1}{M}\left(\mathbf{\Phi}^{(1)^H}\mathbf{C}\mathbf{\Phi}^{(1)}\right)^{-1}+\cdots+\frac{1}{M}\left(\mathbf{\Phi}^{(M-1)^H}\mathbf{C}\mathbf{\Phi}^{(M-1)}\right)^{-1}\succeq \left(\frac{1}{M}\,\mathbf{\Phi}^{(1)^H}\mathbf{C}\mathbf{\Phi}^{(1)}+\cdots+\frac{1}{M}\,\mathbf{\Phi}^{(\mathsf{M})^H}\mathbf{C}\mathbf{\Phi}^{(\mathsf{M})}\right)^{-1} \tag{5.27}$$

$$\Longrightarrow \frac{1}{M^2}\Big(\left(\mathbf{\Phi}^{(1)^H}\mathbf{C}\mathbf{\Phi}^{(1)}\right)^{-1}+\cdots+\left(\mathbf{\Phi}^{(\mathsf{M})^H}\mathbf{C}\mathbf{\Phi}^{(\mathsf{M})}\right)^{-1}\Big)\succeq \left(\mathbf{\Phi}^{(1)^H}\mathbf{C}\mathbf{\Phi}^{(1)}+\cdots+\mathbf{\Phi}^{(\mathsf{M})^H}\mathbf{C}\mathbf{\Phi}^{(\mathsf{M})}\right)^{-1}=\mathcal{C}^{-1}. \tag{5.28}$$

Because the array is a ULA, the complex exponential terms in $\mathbf{\Phi}^{(m)}$ yield

$$\left(\mathbf{\Phi}^{(m)^H}\mathbf{C}\mathbf{\Phi}^{(m)}\right)^{-1}_{n,n}=\mathbf{C}^{-1}_{n,n}. \tag{5.29}$$

The relations expressed in (5.29) hold for each diagonal term (n,n). Now comparing diagonal terms of the matrices on both sides of the $\succeq$ sign in (5.28), we obtain

$$\mathcal{C}^{-1}_{n,n}\le\frac{1}{M^2}\left(\left(\mathbf{\Phi}^{(1)^H}\mathbf{C}\mathbf{\Phi}^{(1)}\right)^{-1}+\cdots\left(\mathbf{\Phi}^{(\mathsf{M})^H}\mathbf{C}\mathbf{\Phi}^{(\mathsf{M})}\right)^{-1}\right)_{n,n}=\frac{1}{M}\mathbf{C}^{-1}_{n,n}. \tag{5.30}$$

The first inequality above follows from the fact that if $\mathbf{A}\succeq\mathbf{B}$, then $\mathbf{B}_{n,n}\le\mathbf{A}_{n,n},\ \forall n$. This proves Proposition (c) and completes the proof. ■

From Part (b) of the above theorem and Eqs. (5.22), (5.23) we see that

$$Pe_1^{(\mathsf{M})}\le Pe_1^{(1)},\qquad \eta_1^{(\mathsf{M})}\ge\eta_1^{(1)}.$$

Therefore, the performance enhancement of an M-sensor antenna array receiver over a single sensor counterpart is evident in the decreased probability of error and the superior near-far resistant properties (since the near-far resistance of a detector, $\overline{\eta}=\min_k\eta_k,\ k=1,\ldots,K$). Part (c) quantifies the gain more accurately; we now see that $\eta_1^{(\mathsf{M})}\ge M\cdot\eta_1^{(1)}$ with the associated decrease in the probability of error (since $Pe_1^{(\cdot)}=Q\left(\sqrt{w_1^2\eta_1^{(\cdot)}/\sigma^2}\right)$), yielding a gain that is directly dependent on the number of sensors, M.

Numerical Results

The above analytic results are corroborated by simulations of performance below. In all the following figures we assume perfect knowledge of the delays and DOAs of all the users since the focus is on the gain obtained in using multiple sensors at the receiver. The delays of the users were generated uniformly in $[0, NT_c)$ and the DOAs were uniformly distributed in $[-\pi/3, \pi/3)$ radians. Each user was assigned a Gold code of length $N = 31$ and the interfering users' powers were distributed uniformly in $[0, 20]$ dB relative to the power of the desired user.

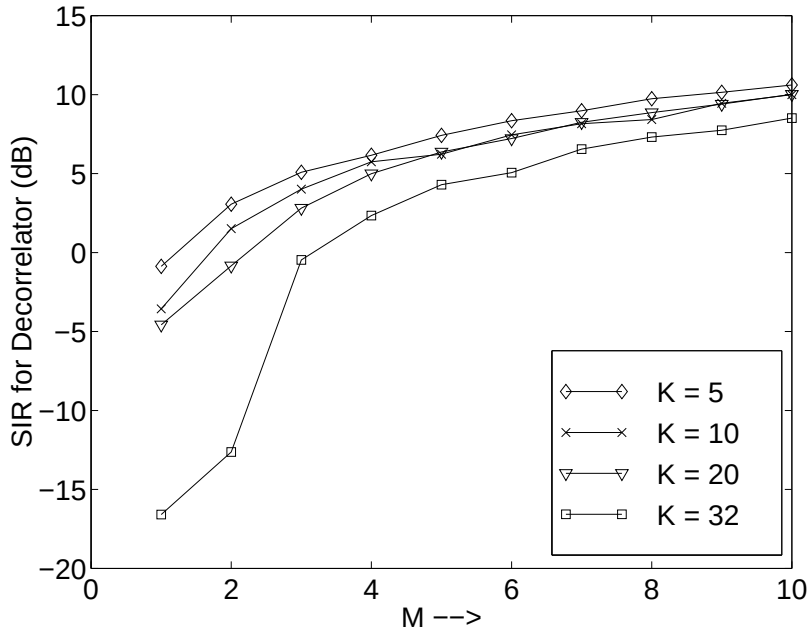


Figure 5.2 : Plot of SIR_1 at the output of the decorrelating detector as a function of the number of sensors in the array for different numbers of total users in the system. The desired user's SNR was 8dB and the MAI was uniformly distributed in $[0, 20]$dB.

In Figure 5.2 we graph the maximum SIR (from simulations) of user 1 at the output of the decorrelating detector as a function of M. As expected we observe an increase in SIR with an increase in array size (in fact, the SIR increases approximately linearly in M, as predicted by (5.20) and Theorem 5.1.

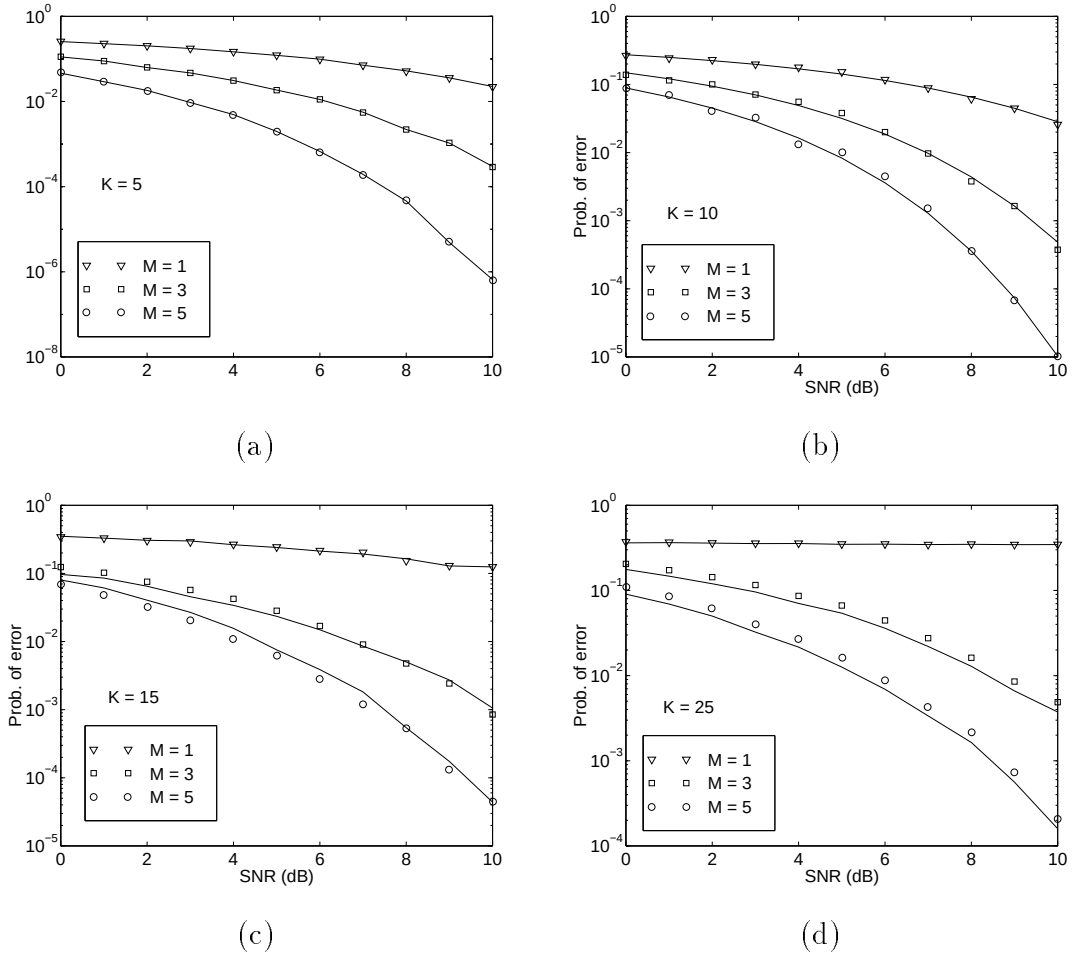


Figure 5.3 : Performance of the multisensor decorrelating detector versus SNR of the desired user for $M = 1, 3$ and 5 sensors. Plots (a), (b), (c) and (d) depict the performance for $K = 5, 10, 15$ and 25 users in the system respectively. The MAI is 20dB. The solid lines represent the theoretical curves obtained from. The '∇', '$\Box$' and '$\circ$' symbols represent $M = 1, 3$ and 5 sensors respectively.

Since the probability of error in the case of the decorrelating detector is a direct function of SIR, as captured in (5.22) we see that the performance of a multisensor decorrelating detector improves significantly over the single-sensor case (because the SIR appears as the argument of a Q-function). This is indeed captured in Figures 5.3(a), (b), (c) and (d), where we observe the performance of the decorrelating receiver by graphing the probability of error against increasing SNR for different number of users in the system. We compare the performance for different number of sensors, 1, 3 and 5, at the receiver. Plots 5.3(a), (b), (c) and (d) represent $K = 5, 10, 15$ and 25 users in the system respectively. As expected, the performance increases (and we note, rather dramatically) as a function of the number of sensors. This can be attributed to the increased dimensionality of the system and hence the corresponding increase in effective SNR; with M sensors, we observe the effects of M-diversity. We also note that the performance decreases as the number of users increases. Since the single sensor decorrelating detector is given by $\mathbf{C^{-1}} = (\mathbf{A}^\top \mathbf{A})^{-1}$, where $\mathbf{A} \in \mathbb{R}^{N \times 2K}$, its rank is determined by the rank of $\mathbf{A}$ which is $\min\{2K, N\}$. Therefore if the number of users K is such that $2K > N$, the detector becomes rank deficient. In practice, as $2K$ approaches N, the matrix $\mathbf{C^{-1}}$ becomes ill conditioned and performance drops. Thus, in Figures 5.3 (c) and (d) where $K = 15$ and 25 respectively, we notice that the single sensor receiver performs quite poorly. Geometrically speaking, since the decorrelating detector attempts to orthogonalize the users, with more than $\lfloor N/2 \rfloor$ users in the system, they cannot all be orthogonalized and hence the performance degradation. In the multisensor case, $\mathcal{G}_{\text{dec}} = (\mathcal{A}^H \mathcal{A})^{-1}$, where $\mathcal{A} \in \mathbb{C}^{MN \times 2K}$ and hence $2K$ can get much larger before performance degrades — with a greater number of dimensions at our disposal we can orthogonalize more users. The gain is a factor of M, the number of sensors.

It should be remarked that the above results were derived for the case where the noise is both temporally *and* spatially white. If the noise has a non-white but known correlation structure, $\boldsymbol{\nu}_i \sim N(\mathbf{0}, \mathbf{K})$, then this must be factored in the calculation

of $Pe_1^{(\mathsf{M})}, \eta_1^{(\mathsf{M})}, Pe_1^{(1)}$ and $\eta_1^{(1)}$. Quantitative comparisons in this case become more difficult to make.

5.1.2 MMSE Detector

In an M-sensor synchronous system, the linear MMSE detector can be cast as the solution to the problem [37]

$$\min_{\mathcal{G}} E[\|\mathbf{W}\mathbf{b}_i - \mathcal{G}\mathbf{y}_i\|^2],$$

and is given by

$$\mathcal{G}_{\text{mmse}} = (\mathcal{C} + \sigma^2\mathbf{W}^{-2})^{-1}. \tag{5.31}$$

The linear MMSE detector can also be shown to maximize the SIR in both synchronous as well as asynchronous cases over the class of linear detectors (thus, it is the analogue of the minimum variance beamformer in conventional array signal processing terms). Unlike the decorrelating detector, the probability of error for user 1 at the output of the linear MMSE detector, which is the true measure of performance, is not strictly a function of the SIR [37]. However, for large code lengths and number of users, the approximation $Pe_1 \approx Q(\sqrt{SIR_1})$ is quite accurate. We now proceed to prove that in the synchronous case, the SIR (defined in (5.7)) increases with the number of sensors.

From (5.31) and (5.7) we can rewrite the SIR of user 1 in the following manner

$$1 - \frac{1}{SIR_1^{(\mathsf{M})} + 1} = \Big\{ \Big[\mathbf{I} \odot \sigma^{-1}\mathbf{W}\mathcal{C}\left(\mathcal{C} + \sigma^2\mathbf{W}^{-2}\right)^{-1}\Big] \Big[\mathbf{I} \odot \left(\mathcal{C} + \sigma^2\mathbf{W}^{-2}\right)^{-1} \left(\mathcal{C} + \sigma^{-2}\mathcal{C}\mathbf{W}^2\mathcal{C}\right)\left(\mathcal{C} + \sigma^2\mathbf{W}^{-2}\right)^{-1}\Big]^{-1} \Big[\mathbf{I} \odot \left(\mathcal{C} + \sigma^2\mathbf{W}^{-2}\right)^{-1}\sigma^{-1}\mathcal{C}\mathbf{W}\Big] \Big\}_{1,1}, \tag{5.32}$$

where "$\odot$" represents the Hadamard product defined in Chapter 3. We now establish the following theorem.

Theorem 5.2

For an M-sensor ULA we have the following inequality

$$SIR_1^{(\mathsf{M})} \geq SIR_1^{(1)}. \tag{5.33}$$

Proof : To prove the theorem, it is sufficient to show that

$$1 - \frac{1}{SIR_1^{(\mathsf{M})} + 1} \geq 1 - \frac{1}{SIR_1^{(1)} + 1}.$$

The proof now follows from a few simple manipulations of the matrices in (5.32). We first note that

$$\left(\mathcal{C} + \sigma^2 \mathbf{W}^{-2}\right)^{-1} \left(\mathcal{C} + \sigma^{-2} \mathcal{C} \mathbf{W}^2 \mathcal{C}\right) = \sigma^{-2} \mathbf{W}^2 \mathcal{C}$$

and therefore (5.32) reduces to

$$\begin{aligned} 1 - \frac{1}{SIR_1^{(\mathsf{M})} + 1} &= \left(\mathbf{I} \odot \mathcal{C} \left(\mathcal{C} + \sigma^2 \mathbf{W}^{-2}\right)^{-1}\right)_{1,1} \\ &= \left(\mathbf{I} \odot \left(\mathbf{I} + \sigma^2 \mathbf{W}^{-2} \mathcal{C}^{-1}\right)^{-1}\right)_{1,1}. \end{aligned}$$

Therefore, to prove that $SIR_1^{(\mathsf{M})} \geq SIR_1^{(1)}$, we need to show that

$$\begin{aligned} \left(\mathbf{I} \odot \left(\mathbf{I} + \sigma^2 \mathbf{W}^{-2} \mathcal{C}^{-1}\right)^{-1}\right)_{1,1} &\geq \left(\mathbf{I} \odot \left(\mathbf{I} + \sigma^2 \mathbf{W}^{-2} \mathbf{C}^{-1}\right)^{-1}\right)_{1,1} \\ \text{i.e., } \left(\mathbf{I} + \sigma^2 \mathbf{W}^{-2} \mathcal{C}^{-1}\right)^{-1}_{1,1} &\geq \left(\mathbf{I} + \sigma^2 \mathbf{W}^{-2} \mathbf{C}^{-1}\right)^{-1}_{1,1}. \end{aligned} \tag{5.34}$$

But (5.34) follows immediately because $\mathcal{C}^{-1}_{1,1} \leq \mathbf{C}^{-1}_{1,1}$ (from Theorem 5.1). Hence we have the theorem. ■

In this case, however, we notice that the performance improvement cannot be directly quantified in terms of the number of sensors M.

Turning our attention to the asynchronous case now, we have the decision statistic given by

$$(\mathbf{z}_i^{(\mathsf{M})})_1 = (1 - \alpha)\left(\mathcal{G}\mathcal{C}\mathbf{W}\mathbf{b}_i\right)_2 + \alpha\left(\mathcal{G}\mathcal{C}\mathbf{W}\mathbf{b}_{i+1}\right)_1 + \tilde{\nu},$$

where the additive gaussian noise term is

$$\tilde{\nu} = (1-\alpha)(\mathcal{G}\mathcal{A}^H \boldsymbol{\nu}_i)_{2,2} + \alpha(\mathcal{G}\mathcal{A}^H \boldsymbol{\nu}_{i+1})_{1,1}\,.$$

The optimum α and the resultant SIR can now be given analogous to (5.14) and (5.16). In this case, the analytical demonstration of the increase in SIR with number of sensors is a daunting task. However, it is reasonable to expect that the above property still holds (the asynchronous nature of the system cannot negate basic intuition!) and we resort to simulations to substantiate our claim.

Numerical Results

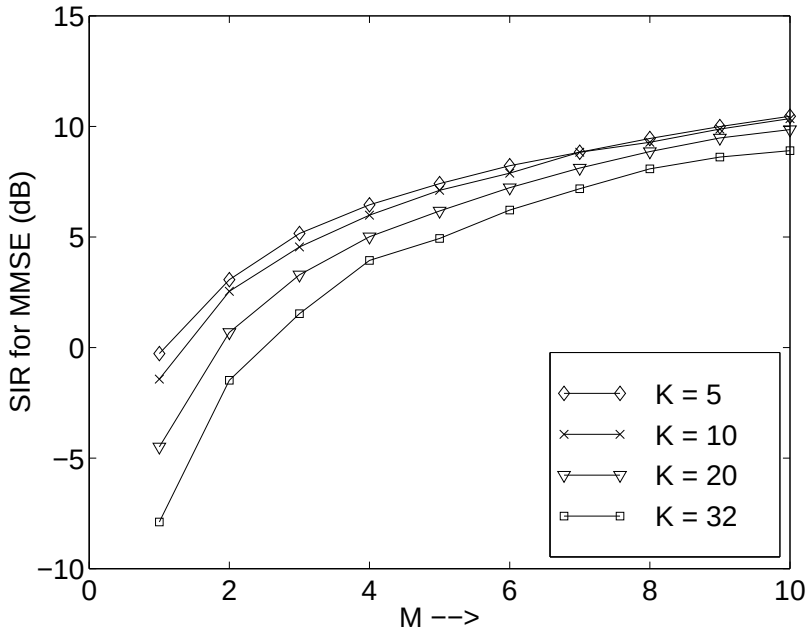


Figure 5.4 : Plot of SIR_1 for a linear MMSE detector in an asynchronous system versus the number of sensors in the array. The SIR is graphed for different number of users in the system. The desired user's SNR was 8dB and the MAI was uniformly distributed in $[0, 20]$dB.

Figure 5.4 demonstrates the benefits of increased number of sensors in an asynchronous system in terms of the SIR of user 1 at the output of the MMSE detector.

We observe that the rate of increase of the SIR with M is also approximately linear, as in the case of the decorrelator.

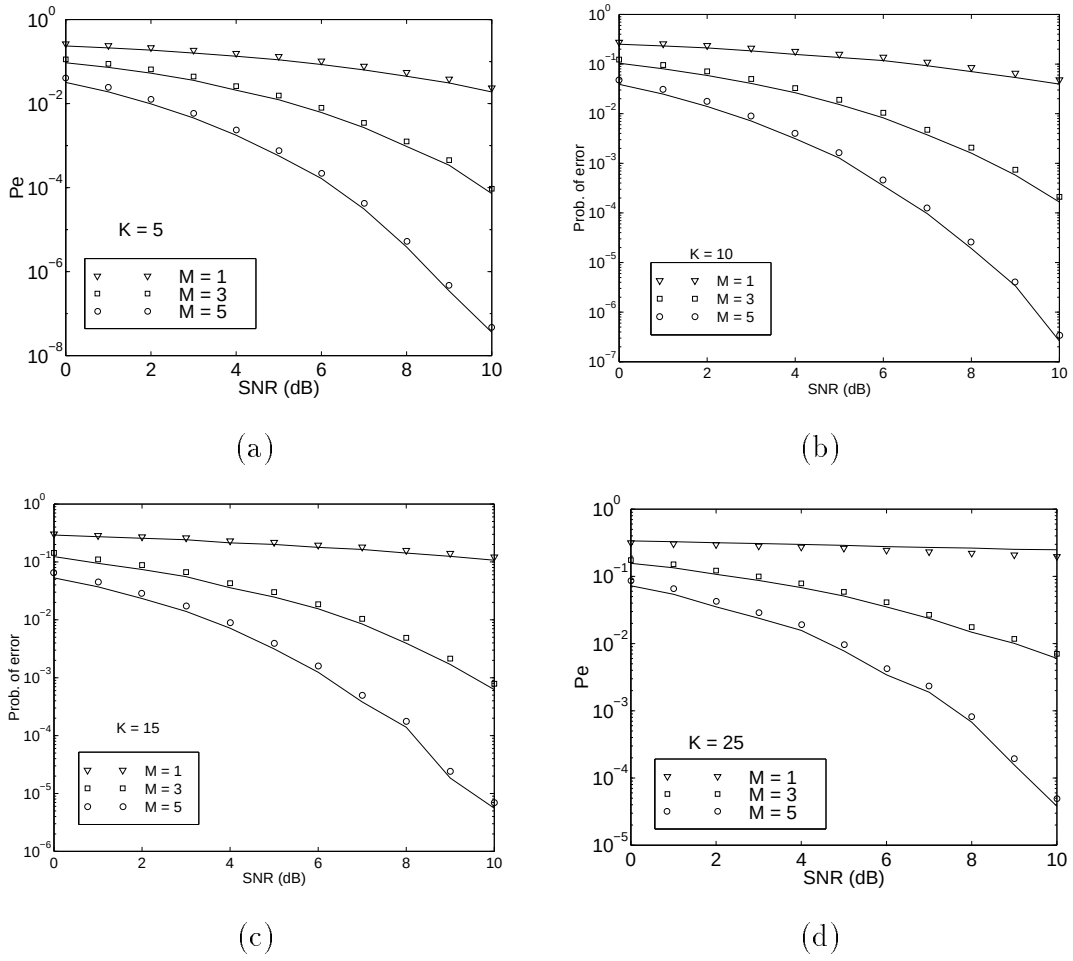


Figure 5.5 : Performance of the multisensor linear MMSE detector versus SNR of the desired user for $M = 1, 3$ and 5 sensors. Plots (a), (b), (c) and (d) depict the performance for $K = 5, 10, 15$ and 25 users in the system respectively. The MAI is 20dB. The solid lines represent the "theoretical" curves obtained from $Q(\sqrt{SIR_1})$. The ∇, $\square$ and $\circ$ symbols represent $M = 1, 3$ and 5 sensors respectively.

The near-far resistance properties of the MMSE detector are identical to those of the decorrelating detector [50] and hence the gain in near-far resistant properties from employing M sensors is identical to the decorrelating detector case. However, since the probability of error is not a direct function of SIR, analytical demonstration

of the superiority of multiple sensors is not straightforward. However, since we know from Figure 5.4 that SIR_1 increases with M, we expect the same to hold for Pe_1, especially since Pe_1 is closely approximated by $Q(\sqrt{SIR_1})$ for large code lengths and number of users. We illustrate the gains via simulations of the probability of error versus increasing SNR in Figure 5.5 for array sizes of $M = 1, 3$ and 5 sensors. Figures 5.5(a), (b), (c) and (d) depict the probabilities of error for user 1 with 4, 9, 14 and 24 interfering users in the system respectively. As we increase the number of users we notice a degradation in performance across all array sizes, especially for one sensor. Even though the MMSE detector does not become singular as $2K \rightarrow N$, it does become ill-conditioned. The analogy to the minimum variance beamformer carries over and we see that we cannot "null out" more than $\lfloor N/2 \rfloor$ interferers. With increasing array sizes, we have an increased number of dimensions in the system and hence we can null out an increased number of interferers as seen from the plots of $M = 3$and 5 in Figure 5.5. Also plotted in Figure 5.5 are the "theoretical Pe" curves, i.e., we plot the approximation for Pe_1 given by $Pe_1 \approx Q(\sqrt{SIR_1})$ for the various combinations of K and M. We notice a very close agreement between simulation and this approximation even for small number of users, $K = 5$. The code length used in the simulations was $N = 31$.

Chapter 6

Conclusions

This thesis investigates the gain achieved in incorporating multiple sensors in the form of an antenna array at the base station in a CDMA cellular communication system. Increasing the number of sensors in the array increases the effective SNR of each user and hence, we expect a performance enhancement over single-sensor receivers; this enhancement is quantified. Multisensor algorithms for acquisition in CDMA systems have not received much attention in the literature, [21,23,24] featuring the prominent approaches.

The main contributions of this thesis are twofold: we develop a framework for simultaneous estimation of delay and DOA for transmitting users in a cellular CDMA system; we also demonstrate the advantages of increased number of sensors in both parameter estimation as well as bit detection scenarios. We first develop a maximum likelihood algorithm to estimate the delays, DOAs and amplitudes of a set of transmitting users in the reverse link of a wireless CDMA communication network. The algorithm is seen to be a generalization of an algorithm proposed by [25] for DOA estimation. Traditional acquisition schemes employ a single sensor at the receiver and are therefore limited in the number of delays they can simultaneously estimate. We utilize multiple sensors via antenna arrays at the receiver to surmount this handicap. The presence of multiple sensors not only contributes to increased dimensionality through spatial diversity, but also facilitates important procedures such as source location and soft handoff. In addition to their delays, we are now able to estimate the direction of arrival (DOA) of individual signals. The additive complex noise in the system is assumed to be zero-mean, Gaussian but no assumption is made on its covariance

structure. We see that by incorporating simplifying assumptions on the signal and noise structures we obtain an elegant decoupling of the multiuser estimation problem into the constituent single-user ones and a further decoupling of the joint delay-DOA-amplitude estimation for each user. The proposed algorithm is seen to possess many useful properties. Since the delay and DOA estimates are independent of the amplitude, these estimators are near-far resistant. The maximum likelihood algorithm can be cast as a subspace fitting problem; however, we notice that the estimators are not dimensionally limited as traditional subspace-based estimators that decompose the correlation matrix of the observations are. This coupled with the relative computational simplicity makes them ideal for acquisition in scenarios where preambles can be accomodated. We observe the advantage of multiple sensors through the uniformly enhanced DOA estimates as well as the improved delay estimates in low SNR situations. Furthermore, the preamble size is not prohibitively large either — for the number of antennas considered, i.e., 1, 3 and 5, satisfactory acquisition occurred with 150 bits. An interesting outcome of the joint delay-DOA estimation, as evinced by the Cràmer-Rao bound in (3.53) (3.54), is the trade-off between the accuracy in estimating both parameters simultaneously. At present, the only limitation of the method is the imposition of the single path channel model. Multipath introduces correlation among the received training signals and the algorithm has to be modified to account for this effect, making it necessarily more complex. It is in this direction that the authors are currently navigating.

In the next chapter we adopt a blind subspace-based acquisition scheme based on the well known MUSIC algorithm for the joint delay-DOA estimation problem. The transmitted signals lie in a structured subspace called the signal subspace. This subspace along with its orthogonal complement, the noise subspace, is estimated via an SVD of a data matrix. We notice that the number of data points required for reasonable estimates is around $M \cdot N$, the dimension of the observation correlation matrix. The presence of more than one sensor significantly increases the number of

users' delays that can be estimated through an increase in the dimension of the noise subspace. The gain is proportional to the number of sensors in the array, M; we can now estimate the parameters of $\lfloor MN/2 \rfloor$ users. We observe that the estimation algorithm is near-far resistant since the signal and noise subspaces are independent of the received amplitude; however, it should be remarked that this property holds only when the number of impinging signals is well within the dimensionality of the system. In the context of joint delay-DOA estimation we have proposed a computationally simpler alternative to the full MUSIC algorithm via the Taylor series expansion of the array response vector, and we notice that the performance loss, in terms of the probability of acquisition as well as mean squared estimate errors, is quite small.

The joint delay-DOA information can be exploited to yield lower probabilities of error in the detection of the transmitted bits of the users. This gain is highlighted in the thesis through a study of two single-shot linear multiuser detectors, the decorrelating detector and the linear MMSE detector. Due to the asynchronous reception of users' bitstreams, we make use of the detector outputs at two successive time instants to form a decision on each bit. We derive the optimal linear combination of these successive outputs for a general linear detector, based on maximizing the output signal to interference ratio (SIR). The gain in performance due to multiple sensors, quantified by the gain in output SIR, is found to depend directly on the number of sensors in the array.

6.1 Future Work and Issues

The acquisition algorithms described above were derived for the case of a single path channel. An issue that warrants addressing is the generalization to multipath fading channels. It is well known that fading dramatically decreases performance [2] and diversity techniques need to be implemented to mitigate the effects of fading. The presence of multipath introduces correlation among the received signals and hence the estimation algorithms need to be modified to incorporate this knowledge. A viable

technique for handling intersignal coherence is through sub-array averaging [19] where spatial averaging is enhanced at the expense of resolution. More complex subspace techniques such as weighted subspace fitting (WSF) [52] or probabilistic subspace techniques [26] can be employed to resolve signal coherence; while these methods are indeed effective, they involve added computational complexity.

In this study we treated the transmitted signals as narrowband. While current systems can certainly be classified as narrowband from the point of view of the antenna array, it is doubtful that future generation wireless CDMA systems with chip rates in excess of 10MHz truly fall under the same umbrella. Therefore the investigation of broadband techniques assumes importance. In this connection we must mention that uniform linear arrays are fraught with many disadvantages [38] and it behooves one to investigate more sophisticated array geometries; one such example is the uniform circular array of Figure 2.2.

In conclusion we would like to briefly touch on an issue not addressed in this thesis: *tracking* the estimated parameters and subspaces. The study and incorporation of algorithms that adaptively track the noise or signal subspaces is of paramount importance since the SVD in this problem is of complexity $O(M^2N^2)$. Current techniques perform K-dimensional subspace tracking in $O(NK)$ complexity, where N is the overall dimension [47, 53, 54]. Therefore, subspace tracking offers a computationally attractive alternative to the complete SVD with little performance loss. In addition, with subspace tracking, typical nonstationarities such as slowly varying delays, DOAs as well as amplitudes can be tracked. Implementing such tracking algorithms on real time processors is therefore an issue that is rapidly gaining importance in the current world of dedicated parallel architectures for signal processing applications [55].

Bibliography

[1] R. D. Gitlin, J. F. Hayes, and S. B. Weinstein, *Data Communications Principles*. Plenum Press, 1992.

[2] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 2nd ed., 1989.

[3] M. Simon, J. Omura, R. Scholtz, and B. Levitt, *Spread spectrum communications*, vol. I,II and III. Rockville MD: Computer Science Press, 1985.

[4] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Addison-Wesley Wireless Communications Series, Reading, Massachusetts: Addison-Wesley, 1995.

[5] J. G. Proakis and M. Salehi, *Communication Systems Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[6] G. L. Turin, "Introduction to spread-spectrum antimultipath techniques and their application to urban digital radio," *Proceedings of the IEEE*, vol. 68, pp. 328–353, Mar. 1980.

[7] R. L. Pickholtz, D. L. Schilling, and L. B. Milstein, "Theory of spread-spectrum communications — a tutorial," *IEEE Trans. Commun.*, vol. COM-30, pp. 904–912, May 1982.

[8] R. B. Ward and K. Yiu, "Acquisition of pseudonoise signals by recursion-aided sequential estimation," *IEEE Trans. Commun.*, pp. 784–794, Aug. 1977.

[9] A. Polydoros and M. K. Simon, "Generalized serial search code acquisition: the equivalent circular state diagram approach," *IEEE Trans. on Comm.*, vol. 32, pp. 550–560, May 1984.

[10] V. M. Jovanovic, "Analysis of strategies for serial-search spread-spectrum code acquisition – direct approach," *IEEE Trans. on Comm.*, vol. 36, pp. 1208–1220, Nov. 1988.

[11] Z. Xie, C. K. Rushforth, R. Short, and T. K. Moon, "Joint signal detection and parameter estimation in multiuser communications," *IEEE Trans. on Commun.*, vol. 41, no. 7, pp. 1208–1216, 1993.

[12] A. Radović and B. Aazhang, "Iterative algorithms for joint data detection and delay estimation for code division multiple access communication systems," in *Proc. of the 31st Annual Allerton Conference on C ommunication, Control, and Computing*, (Allerton House, Monticello, IL), 1993.

[13] S. Y. Miller and S. C. Schwartz, "Parameter estimation for asynchronous multiuser communication," in *Proceedings of the 1989 Conference on Information Sciences and Systems*, pp. 294–299, 1989.

[14] Y. Steinberg and H. V. Poor, "Sequential amplitude estimation in multiuser communications," *IEEE Trans. Inform. Theory*, vol. IT-40, pp. 11–20, Jan. 1994.

[15] S. E. Bensley and B. Aazhang, "Maximum likelihood synchronization of a single user for code division multiple access communication systems," *Accepted for publication in IEEE Trans. Commun.*, 1995.

[16] S. E. Bensley and B. Aazhang, "Subspace-based channel estimation for code division multiple access communication systems," *IEEE Trans. on Commun.*, vol. 44, pp. 1009–1020, Aug. 1996.

[17] E. G. Ström, S. Parkvall, S. L. Miller, and B. E. Ottersen, "Propagation delay estimation in asynchronous direct-sequence code-division multiple access systems," *Submitted to IEEE Trans. on Commun.*, 94.

[18] U. Madhow and M. Pursley, "Acquisition in direct-sequence spread-spectrum communication networks: An asymptotic analysis," *IEEE Trans. on Inform. Theory*, vol. 39, pp. 903–912, May 1993.

[19] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[20] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 1996.

[21] R. T. Compton, "An adaptive array in a spread-spectrum communication system," *Proceedings of the IEEE*, vol. 66, pp. 289–298, Mar. 1978.

[22] S. Y. Miller and S. C. Schwartz, "Integrated spatial-temporal detectors for asynchronous gaussian multiple-access channels," *IEEE Trans. Commun.*, vol. COM-43, no. 2/3/4, pp. 396–411, 1995.

[23] L. E. Brennan and I. S. Reed, "An adaptive signal processing algorithm for communications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-18, pp. 124–130, 1982.

[24] D. Dlugos and R. Scholtz, "Aquisition of spread spectrum signals by an adaptive array," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 37, pp. 1253–1270, August 1989.

[25] J. Li, B. Halder, P. Stoica, and M. Viberg, "Computationally efficient angle estimation for signals with known waveforms," *IEEE Trans. on Signal Processing*, vol. 43, pp. 2154–2163, Sept. 1995.

[26] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. ASSP*, vol. 38, pp. 1132–1143, 1990.

[27] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in *Radar Array Processing* (S. Haykin, J. Litva, and T. J. Shepherd, eds.), vol. 25 of *Springer Series in Information Sciences*, Springer-Verlag, 1993.

[28] P. Stoica and A. Nehorai, "Music, maximum likelihood and cramer-rao bound," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 37, pp. 720–741, May 1989.

[29] A. Van der Veen, E. Deprettere, and A. Swindlehurst, "Subspace-based signal analysis using singular value decomposition," *Proceedings of the IEEE*, vol. 81, pp. 1277–1308, Sept. 1993.

[30] P. Stoica and A. Nehorai, "Performance comparison of subspace rotation and music methods for direction estimation," *IEEE Trans. Signal Processing*, vol. 39, pp. 446–453, Feb. 1991.

[31] M. K. Varanasi and B. Aazhang, "Multistage detection for asynchronous code-division multiple-access communications," *IEEE Trans. Commun.*, vol. COM-38, April 1990.

[32] N. B. Mandayam and B. Aazhang, "Gradient estimation for sensitivity analysis and adaptive multiuser interference rejection in code division multiple access systems," *IEEE Trans. Commun.*, 1994. Submitted for publication.

[33] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 123–136, Jan. 1989.

[34] R. Lupas and S. Verdu, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. COM-38, no. 4, pp. 496–508, April, 1990.

[35] M. L. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Th.*, vol. IT-41, no. 3, pp. 944–960, 1995.

[36] Z. Xie, R. Short, and C. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.*, vol. 8, May 1990.

[37] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3178–3188, 1994.

[38] C. U. Padmini and P. S. Naidu, "Circular array and estimation of direction of arrival of a broadband source," *Signal Processing*, vol. 37, pp. 243–254, 1994.

[39] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 3rd ed., 1991.

[40] L. Ljung, *System Identification Theory for the User*. Prentice-Hall, Inc., 1987.

[41] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[42] Y. Bresler and A. Macovski, "On the number of signals resolvable by a uniform linear arrau," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 6, December 1986.

[43] R. O. Schmidt, *A signal subspace approach to multiple emitter location and spectral estimation*. PhD thesis, Stanford University, Stanford, CA, 1981.

[44] R. Roy and T. Kailath, "ESPRIT — estimation of signal parameters via rotational invariant techniques," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[45] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. on Signal Processing*, vol. 39, pp. 1110–1120, May 1991.

[46] K. Kota and J. R. Cavallaro, "Pipelining multiple SVDs on a single processor array," *Advanced Signal Processing Algorithms, Architectures and Implementations V*, vol. 2296, pp. 612–623, 1994.

[47] M. Moonen, P. V. Dooren, and J. Vandewalle, "A systolic array for SVD updating," *SIAM J. Matrix Anal. Appl.*, vol. 14, pp. 353–371, 1993.

[48] H. L. van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 1968.

[49] P. Stoica and A. Nehorai, "Statistical efficiency study of direction estimation methods part I: Analysis of MUSIC and preliminary study of MLM," in *Advances in Spectrum Analysis and Array Processing — Volume II* (S. Haykin, ed.), pp. 263–305, Englewood Cliffs, New Jersey: Prentice Hall, 1991.

[50] S. Verdu, "Multiuser detection," in *Advances in Statistical Signal Processing: Detection and Estimation*, vol. 2, pp. 369–409, 1993.

[51] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[52] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. ASSP*, vol. 39, pp. 2436–2449, 1991.

[53] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, vol. 43, pp. 95–107, Jan. 1995.

[54] G. W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Process.*, vol. 40, pp. 1535–1541, June 1992.

[55] C. Sengupta, J. R. Cavallaro, and B. Aazhang, "Solving the SVD updating problem for subspace tracking on a fixed sized linear array of processors," in *Proceedings of ICASSP*, 1997.

[56] C. F. Van Loan and N. P. Pitsianis, "Approximation with Kronecker products," in *Linear Algebra for Large Scale and Real Time Applications* (M. S. Moonen and G. H. Golub, eds.), pp. 293–314, Kluwer Publications, 1993.

[57] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore: The Johns Hopkins University Press, 2nd ed., 1989.

Appendix A

Proof of (3.14)

In this appendix we will show that for large L,

$$\min_{\boldsymbol{\xi}} \ln \left| \mathbf{I} + \widehat{\mathbf{K}}^{-1}(\mathcal{Y} - \widehat{\mathcal{Y}})\widehat{\mathbf{R}}_{bb}(\mathcal{Y} - \widehat{\mathcal{Y}})^H \right| \approx \min_{\boldsymbol{\xi}} \mathrm{tr}\{\widehat{\mathbf{R}}_{bb}(\mathcal{Y} - \widehat{\mathcal{Y}})^H \widehat{\mathbf{K}}^{-1}(\mathcal{Y} - \widehat{\mathcal{Y}})\}, \tag{A.1}$$

where $\boldsymbol{\xi} = \{\boldsymbol{\tau}, \boldsymbol{\theta}, \mathbf{w}\}$. In other words, minimizing the two functions in A.1 is asymptotically equivalent. Since the sample correlation matrix $\widehat{\mathbf{R}}_{bb}$ is positive definite we can write it as $\widehat{\mathbf{R}}_{bb} = \mathbf{R}\mathbf{R}^H$ and therefore we can define

$$\mathbf{Z}\mathbf{Z}^H \stackrel{\triangle}{=} (\mathcal{Y} - \widehat{\mathcal{Y}})\widehat{\mathbf{R}}_{bb}(\mathcal{Y} - \widehat{\mathcal{Y}})^H.$$

Since $\widehat{\mathcal{Y}}$ is a consistent estimate [40], $(\mathcal{Y} - \widehat{\mathcal{Y}}) \xrightarrow{L} \mathbf{0}$ and therefore

$$\mathbf{Z} \stackrel{\triangle}{=} (\mathcal{Y} - \widehat{\mathcal{Y}})\,\mathbf{R} \xrightarrow{L} \mathbf{0}.$$

Now, $\min \ln \left| \mathbf{I} + \widehat{\mathbf{K}}^{-1}\mathbf{Z}\mathbf{Z}^H \right|$ is equivalent to $\min \ln \left| \mathbf{I} + \mathbf{Z}^H\widehat{\mathbf{K}}^{-1}\mathbf{Z} \right|$, which follows from the fact that, for any two matrices $\mathbf{A} \in \mathbb{C}^{r\times s}, \mathbf{B} \in \mathbb{C}^{s\times r}$, we have

$$|\mathbf{I}_r + \mathbf{A}\mathbf{B}| = |\mathbf{I}_s + \mathbf{B}\mathbf{A}|.$$

Let us denote the eigenvalues of $\mathbf{Z}^H\widehat{\mathbf{K}}^{-1}\mathbf{Z}$ as $\lambda_1, \lambda_2, \ldots, \lambda_{MN}$. Then

$$\ln \left| \mathbf{I} + \mathbf{Z}^H\widehat{\mathbf{K}}^{-1}\mathbf{Z} \right| = \sum_{i=1}^{MN} \ln(1 + \lambda_i). \tag{A.2}$$

We now define an appropriate matrix norm $|||\cdot|||$ on $\mathbb{C}^{MN\times MN}$ and the corresponding vector norm $\|\cdot\|$ on $\mathbb{C}^{MN}$. The norms we choose to operate under are the

matrix 2-norm and vector 2-norm, henceforth denoted by $|||\cdot|||$ and $\|\cdot\|$ respectively, and defined as

$$|||\mathbf{A}||| = \left(\sum_{i}^{MN} \sum_{j}^{MN} \mathbf{A}_{i,j}^2 \right)^{1/2}$$

$$\|\mathbf{a}\| = \left(\sum_{i}^{MN} \mathbf{a}_i^2 \right)^{1/2}.$$

Then from norm inequalities [41] we have

$$\begin{aligned} 0 \leq \lambda_j \leq \rho\left(\mathbf{Z}^H \widehat{\mathbf{K}}^{-1} \mathbf{Z}\right) &\leq \left|\left|\left| \mathbf{Z}^H \widehat{\mathbf{K}}^{-1} \mathbf{Z} \right|\right|\right| \\ &\leq \left|\left|\left| \widehat{\mathbf{K}}^{-1} \right|\right|\right| \cdot |||\mathbf{Z}\mathbf{Z}^H||| = \left|\left|\left| \widehat{\mathbf{K}}^{-1} \right|\right|\right| \cdot \left|\left|\left| \sum_i \mathbf{z}_i \mathbf{z}_i^H \right|\right|\right| \leq \left|\left|\left| \widehat{\mathbf{K}}^{-1} \right|\right|\right| \cdot \sum_i \|\mathbf{z}_i\|^2, \end{aligned} \quad \text{(A.3)}$$

for all $j = 1, 2, \ldots, MN$, where $\mathbf{z}_i$ is the i^{th} column of $\mathbf{Z}$ and the spectral radius, $\rho(\cdot)$ is defined as

$$\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

The first inequality in (A.3) follows because $\widehat{\mathbf{K}}^{-1}$ is non-negative definite and hence, so is $\mathbf{Z}^H \widehat{\mathbf{K}}^{-1} \mathbf{Z}$. The second and third inequalities follow from the definition of $\rho(\cdot)$ and the remaining inequalities are standard matrix norm inequalities.

Since $\mathbf{Z} \xrightarrow{L} \mathbf{0}$, it follows that $\|\mathbf{z}_i\| \xrightarrow{L} 0$ and therefore, from (A.3), $\lambda_j \xrightarrow{L} 0 \ \forall j$. We can now rewrite (A.2) as

$$\sum_{i=1}^{MN} \ln(1 + \lambda_i) \approx \sum_{i=1}^{MN} \lambda_i = \text{tr}\left\{ \mathbf{Z}^H \widehat{\mathbf{K}}^{-1} \mathbf{Z} \right\} = \text{tr}\left\{ \widehat{\mathbf{R}}_{bb} (\mathcal{Y} - \widehat{\mathcal{Y}})^H \widehat{\mathbf{K}}^{-1} (\mathcal{Y} - \widehat{\mathcal{Y}}) \right\} \quad \text{(A.4)}$$

and this completes the proof.

Appendix B

Kronecker Product Approximation of a Matrix

Given a matrix $\widehat{\mathbf{K}} \in \mathbb{C}^{MN \times MN}$, we describe an algorithm to approximate $\widehat{\mathbf{K}}$ as $\widehat{\mathbf{K}}_1 \otimes \widehat{\mathbf{K}}_2$ where $\widehat{\mathbf{K}}_1 \in \mathbb{C}^{M \times M}$ and $\widehat{\mathbf{K}}_2 \in \mathbb{C}^{N \times N}$; the approximation is carried out in the Frobenius norm sense, i.e.,

$$\left\{\widehat{\mathbf{K}}_1, \widehat{\mathbf{K}}_2\right\} = \arg \min_{\substack{\mathbf{K}_1 \in \mathbb{C}^{M \times M} \\ \mathbf{K}_2 \in \mathbb{C}^{N \times N}}} \left\| \widehat{\mathbf{K}} - \mathbf{K}_1 \otimes \mathbf{K}_2 \right\|_F^2 . \tag{B.1}$$

For a more complete treatise on kronecker product approximations of a given matrix the reader is referred to [56]. Now define the matrix operation $vec(\cdot)$ as follows :

$$\mathrm{vec}(\mathbf{A}) = \begin{pmatrix} (\mathbf{A})_1 \\ (\mathbf{A})_2 \\ \vdots \\ (\mathbf{A})_n \end{pmatrix} , \tag{B.2}$$

where $\{(\mathbf{A})_i\}_{i=1}^{n}$ are columns of $\mathbf{A}$. If $\mathbf{k}_1 = \mathrm{vec}(\mathbf{K}_1)$ and $\mathbf{k}_2 = \mathrm{vec}(\mathbf{K}_2)$, it is easy to see that the two matrices $\mathbf{k}_1 \mathbf{k}_2^\top \in \mathbb{C}^{M^2 \times N^2}$ and $\mathbf{K}_1 \otimes \mathbf{K}_2 \in \mathbb{C}^{MN \times MN}$ contain exactly the same elements, but in different permutations. We define $\widetilde{\mathbf{K}} \in \mathbb{C}^{M^2 \times N^2}$ to be simply a rearrangement of the elements of $\widehat{\mathbf{K}}$ performed to mirror the rearrangement involved in transforming $\mathbf{K}_1 \otimes \mathbf{K}_2$ to $\mathbf{k}_1 \mathbf{k}_2^\top$. Now we have converted the problem in (B.1) to

$$\left\{\hat{\mathbf{k}}_1 , \hat{\mathbf{k}}_2\right\} = \arg \min_{\substack{\mathbf{k}_1 \in \mathbb{C}^{M^2 \times 1} \\ \mathbf{k}_2 \in \mathbb{C}^{N^2 \times 1}}} \left\| \widetilde{\mathbf{K}} - \mathbf{k}_1 \mathbf{k}_2^\top \right\|_F^2 . \tag{B.3}$$

This now reduces to a familiar SVD problem [57] — if the SVD of $\widetilde{\mathbf{K}}$ can be written as $\widetilde{\mathbf{K}} = \mathbf{U} \Sigma \mathbf{V}^H$, the vectors $\hat{\mathbf{k}}_1$ and $\hat{\mathbf{k}}_2$ are calculated as

$$\hat{\mathbf{k}}_1 = \sqrt{\sigma_1}\, \mathbf{u}_1$$

$$\hat{\mathbf{k}}_2 = \sqrt{\sigma_1}\, \mathbf{v}_1$$

The matrices $\widehat{\mathbf{K}}_1$ and $\widehat{\mathbf{K}}_2$ are now obtained as

$$\widehat{\mathbf{K}}_1 = \mathrm{vec}^{-1}(\hat{\mathbf{k}}_1)\,, \qquad \widehat{\mathbf{K}}_2 = \mathrm{vec}^{-1}(\hat{\mathbf{k}}_2)\,, \tag{B.4}$$

where the operation vec^{-1} has the obvious meaning.

Appendix C

Derivation of CRB($\boldsymbol{\tau}, \boldsymbol{\theta}$)

From (3.3) we have the log-likelihood function of the L observations to be

$$\mathcal{L} = -MNL \cdot \ln \pi - L \cdot \ln |\mathbf{K}| - \left\{ \sum_{i=1}^{L} (\mathbf{r}_i - \mathcal{A}\mathbf{x}_i)^H \mathbf{K}^{-1} (\mathbf{r}_i - \mathcal{A}\mathbf{x}_i) \right\} \tag{C.1}$$

Denote the unknown parameter vector by

$$\boldsymbol{\xi} = [\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_{M^2N^2}, w_1, , \ldots, w_K, \tau_1, \ldots, \tau_K, \theta_1, \ldots, \theta_K],$$

where $\{\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_{M^2N^2}\}$ are the M^2N^2 entries of the unknown covariance matrix $\mathbf{K}$. Then, the CRB for $\boldsymbol{\xi}$ is given by

$$\text{CRB}(\boldsymbol{\xi}) = \left(E[\Psi\, \Psi^\top]\right)^{-1}, \quad \Psi = \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}_1}, \ldots, \frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}_i}, \ldots, \frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}_P}\right]^\top, \tag{C.2}$$

where $P = M^2N^2 + 3K$, the number of unknown parameters. Differentiating with respect to $\mathbf{k}_m$, $m = 1, 2, \ldots, M^2N^2$ we obtain

$$\frac{\partial \mathcal{L}}{\partial \mathbf{k}_m} = -L \cdot \text{tr}\left\{ \mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial \mathbf{k}_m} \right\} + \sum_{i=1}^{L} \boldsymbol{\nu}_i^H \mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial \mathbf{k}_m} \mathbf{K}^{-1} \boldsymbol{\nu}_i \tag{C.3}$$

Differentiating with respect to $\tau_m = q_m + \gamma_m$ (piecewise over successive chip periods), θ_m, $\overline{w}_m$, $\widetilde{w}_m, m = 1, \ldots, K$, where $\overline{w}_m \triangleq \text{Re}(w_m)$ and $\widetilde{w}_m \triangleq \text{Im}(w_m)$ yield

$$\frac{\partial \mathcal{L}}{\partial \tau_m} = 2 \sum_{i=1}^{L} \text{Re}\left(\mathbf{X}_m^H(i-1, i) \boldsymbol{\Gamma}_m^H \mathbf{K}^{-1} \boldsymbol{\nu}_i\right) \tag{C.4}$$

$$\frac{\partial \mathcal{L}}{\partial \theta_m} = 2 \sum_{i=1}^{L} \text{Re}\left(\mathbf{X}_m^H(i-1, i) \Theta_m^H \mathbf{K}^{-1} \boldsymbol{\nu}_i\right) \tag{C.5}$$

$$\frac{\partial \mathcal{L}}{\partial \overline{w}_m} = 2 \sum_{i=1}^{L} \text{Re}\left(\mathbf{B}_m^H(i-1, i) \mathcal{A}_m^H \mathbf{K}^{-1} \boldsymbol{\nu}_i\right) \tag{C.6}$$

$$\frac{\partial \mathcal{L}}{\partial \widetilde{w}_m} = 2 \sum_{i=1}^{L} \mathrm{Im}\left(\mathbf{B}_m^H(i-1,i)\mathcal{A}_m^H \mathbf{K^{-1}} \boldsymbol{\nu}_i\right) \tag{C.7}$$

wherein we have

$$\mathbf{X}_m(i-1,i) = [x_{m,i-1}\ ,\ x_{m,i}]^\top$$

$$\mathbf{B}_m(i-1,i) = [b_{m,i-1}\ ,\ b_{m,i}]^\top$$

$$\mathcal{A}_m = [\mathbf{a}_m^R\ \ \mathbf{a}_m^L] = [\mathbf{p}_m \otimes \mathbf{c}_m^R \quad \mathbf{p}_m \otimes \mathbf{c}_m^L]$$

$$\boldsymbol{\Gamma}_m = \left[\frac{\partial \mathbf{a}_m^R}{\partial \gamma_m} \quad \frac{\partial \mathbf{a}_m^L}{\partial \gamma_m}\right] = \left[\mathbf{p}_m \otimes \frac{\partial \mathbf{c}_m^R}{\partial \gamma_m} \quad \mathbf{p}_m \otimes \frac{\partial \mathbf{c}_m^L}{\partial \gamma_m}\right]$$

$$\Theta_m = \left[\frac{\partial \mathbf{a}_m^R}{\partial \theta_m} \quad \frac{\partial \mathbf{a}_m^L}{\partial \theta_m}\right] = \left[\frac{\partial \mathbf{p}_m}{\partial \theta_m} \otimes \mathbf{c}_m^R \quad \frac{\partial \mathbf{p}_m}{\partial \theta_m} \otimes \mathbf{c}_m^L\right] \tag{C.8}$$

From the equations (C.4)–(C.7) we can write

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\tau}} = 2 \sum_{i=1}^{L} \mathrm{Re}\left(\mathbf{X}^H(i-1,i)\boldsymbol{\Gamma}^H \mathbf{K^{-1}} \boldsymbol{\nu}_i\right) \tag{C.9}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} = 2 \sum_{i=1}^{L} \mathrm{Re}\left(\mathbf{X}^H(i-1,i)\Theta^H \mathbf{K^{-1}} \boldsymbol{\nu}_i\right) \tag{C.10}$$

$$\frac{\partial \mathcal{L}}{\partial \overline{\mathbf{w}}} = 2 \sum_{i=1}^{L} \mathrm{Re}\left(\mathbf{B}^H(i-1,i)\mathcal{A}^H \mathbf{K^{-1}} \boldsymbol{\nu}_i\right) \tag{C.11}$$

$$\frac{\partial \mathcal{L}}{\partial \widetilde{\mathbf{w}}} = 2 \sum_{i=1}^{L} \mathrm{Im}\left(\mathbf{B}^H(i-1,i)\mathcal{A}^H \mathbf{K^{-1}} \boldsymbol{\nu}_i\right) \tag{C.12}$$

where $\boldsymbol{\Gamma}$ and Θ have their obvious interpretation and $\mathbf{X}, \mathbf{B}$ are defined as

$$\mathbf{X}(i-1,i) = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_K \end{pmatrix} \quad , \qquad \mathbf{B}(i-1,i) = \begin{pmatrix} \mathbf{B}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_K \end{pmatrix}$$

Since $\boldsymbol{\nu}_i$ is a zero mean circular complex Gaussian random vector all moments of the form $E[\boldsymbol{\nu}_j^\top]$ and $E[\boldsymbol{\nu}_i^H \boldsymbol{\nu}_j \boldsymbol{\nu}_j^\top] = 0\ \forall i,j$ [28]. Hence we see from (C.3) that

$\partial\mathcal{L}/\partial\mathbf{k}_m$ is uncorrelated with all the other derivatives in (C.9)–(C.12) as well as with $\partial\mathcal{L}/\partial\mathbf{k}_n$, $m \neq n$. This implies that the entries in CRB($\boldsymbol{\xi}$) due to $\mathbf{k}_m$, $m = 1, \ldots, M^2N^2$ do not affect CRB(τ) or CRB(θ). Exploiting the independence and correlation properties of $\boldsymbol{\nu}_i$ mentioned above, we now obtain after some calculations

$$E\left[\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\tau}}\right)\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\tau}}\right)^{\top}\right] = 2\,\mathrm{Re}\sum_{i=1}^{L}\mathbf{X}^H(i-1,i)\boldsymbol{\Gamma}^H\mathbf{K}^{-1}\boldsymbol{\Gamma}\mathbf{X}(i-1,i) \tag{C.13}$$

$$E\left[\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\right)^{\top}\right] = 2\,\mathrm{Re}\sum_{i=1}^{L}\mathbf{X}^H(i-1,i)\Theta^H\mathbf{K}^{-1}\Theta\mathbf{X}(i-1,i) \tag{C.14}$$

$$E\left[\left(\frac{\partial\mathcal{L}}{\partial\overline{\mathbf{w}}}\right)\left(\frac{\partial\mathcal{L}}{\partial\overline{\mathbf{w}}}\right)^{\top}\right] = 2\,\mathrm{Re}\sum_{i=1}^{L}\mathbf{B}^H(i-1,i)\mathcal{A}^H\mathbf{K}^{-1}\mathcal{A}\mathbf{B}(i-1,i) \tag{C.15}$$

$$E\left[\left(\frac{\partial\mathcal{L}}{\partial\widetilde{\mathbf{w}}}\right)\left(\frac{\partial\mathcal{L}}{\partial\widetilde{\mathbf{w}}}\right)^{\top}\right] = 2\,\mathrm{Re}\sum_{i=1}^{L}\mathbf{B}^H(i-1,i)\mathcal{A}^H\mathbf{K}^{-1}\mathcal{A}\mathbf{B}(i-1,i) \tag{C.16}$$

$$E\left[\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\tau}}\right)\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\right)^{\top}\right] = 2\,\mathrm{Re}\sum_{i=1}^{L}\mathbf{X}^H(i-1,i)\boldsymbol{\Gamma}^H\mathbf{K}^{-1}\Theta\mathbf{X}(i-1,i) \tag{C.17}$$

$$E\left[\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\tau}}\right)\left(\frac{\partial\mathcal{L}}{\partial\overline{\mathbf{w}}}\right)^{\top}\right] = 2\,\mathrm{Re}\sum_{i=1}^{L}\mathbf{X}^H(i-1,i)\boldsymbol{\Gamma}^H\mathbf{K}^{-1}\mathcal{A}\mathbf{B}(i-1,i) \tag{C.18}$$

$$E\left[\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\tau}}\right)\left(\frac{\partial\mathcal{L}}{\partial\widetilde{\mathbf{w}}}\right)^{\top}\right] = -2\,\mathrm{Im}\sum_{i=1}^{L}\mathbf{X}^H(i-1,i)\boldsymbol{\Gamma}^H\mathbf{K}^{-1}\mathcal{A}\mathbf{B}(i-1,i) \tag{C.19}$$

$$E\left[\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\mathcal{L}}{\partial\overline{\mathbf{w}}}\right)^{\top}\right] = 2\,\mathrm{Re}\sum_{i=1}^{L}\mathbf{X}^H(i-1,i)\Theta^H\mathbf{K}^{-1}\mathcal{A}\mathbf{B}(i-1,i) \tag{C.20}$$

$$E\left[\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\mathcal{L}}{\partial\widetilde{\mathbf{w}}}\right)^{\top}\right] = -2\,\mathrm{Im}\sum_{i=1}^{L}\mathbf{X}^H(i-1,i)\Theta^H\mathbf{K}^{-1}\mathcal{A}\mathbf{B}(i-1,i) \tag{C.21}$$

$$E\left[\left(\frac{\partial\mathcal{L}}{\partial\overline{\mathbf{w}}}\right)\left(\frac{\partial\mathcal{L}}{\partial\widetilde{\mathbf{w}}}\right)^{\top}\right] = -2\,\mathrm{Im}\sum_{i=1}^{L}\mathbf{B}^H(i-1,i)\mathcal{A}^H\mathbf{K}^{-1}\mathcal{A}\mathbf{B}(i-1,i) \tag{C.22}$$

Because of the "almost-diagonal" structure of $\mathbf{X}$ and $\mathbf{B}$ the expressions in (C.13)

– (C.22) can be rewritten. For example,

$$2\,\mathrm{Re}\sum_{i=1}^{L}\mathbf{X}^H(i-1,i)\mathbf{\Gamma}^H\mathbf{K^{-1}}\mathbf{\Gamma}\,\mathbf{X}(i-1,i) = 2\,\mathrm{Re}\left(\mathbf{\Gamma}^{R^H}\mathbf{K^{-1}}\mathbf{\Gamma}^R\odot\left[\sum_{i=1}^{L}\mathbf{x}_{i-1}\mathbf{x}_{i-1}^H\right]^\top + \mathbf{\Gamma}^{L^H}\mathbf{K^{-1}}\mathbf{\Gamma}^L\odot\left[\sum_{i=1}^{L}\mathbf{x}_i\mathbf{x}_i^H\right]^\top + 2\,\mathbf{\Gamma}^{R^H}\mathbf{K^{-1}}\mathbf{\Gamma}^L\odot\left[\sum_{i=1}^{L}\mathbf{x}_{i-1}\mathbf{x}_i^H\right]^\top\right), \quad \text{(C.23)}$$

$$2\,\mathrm{Im}\sum_{i=1}^{L}\mathbf{X}^H(i-1,i)\mathbf{\Gamma}^H\mathbf{K^{-1}}\mathcal{A}\,\mathbf{B}(i-1,i)$$
$$= 2\,\mathrm{Im}\left(\mathbf{\Gamma}^{R^H}\mathbf{K^{-1}}\mathcal{A}^R\odot\left[\sum_{i=1}^{L}\mathbf{x}_{i-1}\mathbf{b}_{i-1}^H\right]^\top + \mathbf{\Gamma}^{L^H}\mathbf{K^{-1}}\mathcal{A}^L\odot\left[\sum_{i=1}^{L}\mathbf{x}_i\mathbf{b}_i^H\right]^\top\right), \quad \text{(C.24)}$$

where $\mathbf{\Gamma}^R, \mathbf{\Gamma}^L, \mathcal{A}^R, \mathcal{A}^L$ are defined in (3.47) in Chapter 3, the vector of users' bits $\mathbf{b}_i = [b_{1,i},\ b_{2,i},\ \ldots,\ b_{K,i}]^\top$ and $\mathbf{x}_i = \mathbf{W}\mathbf{b}_i$. Since $\mathbf{b}_{i-1}$ is uncorrelated with $\mathbf{b}_i$, for large L we have

$$\sum_i \mathbf{b}_{i-1}\mathbf{b}_{i-1}^H = L\cdot\widehat{\mathbf{R}}_{bb}, \qquad \sum_i \mathbf{b}_i\mathbf{b}_i^H = L\cdot\widehat{\mathbf{R}}_{bb}$$
$$\sum_i \mathbf{x}_{i-1}\mathbf{x}_{i-1}^H = L\cdot\mathbf{W}\widehat{\mathbf{R}}_{bb}\mathbf{W}^H, \qquad \sum_i \mathbf{x}_i\mathbf{x}_i^H = L\cdot\mathbf{W}\widehat{\mathbf{R}}_{bb}\mathbf{W}^H$$
$$\sum_i \mathbf{x}_{i-1}\mathbf{b}_{i-1}^H = L\cdot\widehat{\mathbf{R}}_{bb}\mathbf{W}^H, \qquad \sum_i \mathbf{x}_i\mathbf{b}_i^H = L\cdot\widehat{\mathbf{R}}_{bb}\mathbf{W}^H$$
$$\sum_i \mathbf{b}_{i-1}\mathbf{b}_i^H = \mathbf{0}, \qquad \sum_i \mathbf{x}_{i-1}\mathbf{b}_i^H = \mathbf{0}.$$

Incorporating the above simplifications, the right hand sides of the ten equations (C.13) – (C.22) can be rewritten as $2L\ \mathrm{Re}\Delta_2$, $2L\ \mathrm{Re}\Delta_3$, $2L\ \mathrm{Re}\Delta_1$, $2L\ \mathrm{Re}\Delta_1$, $2L\ \mathrm{Re}\mathbf{\Upsilon}_3$, $2L\ \mathrm{Re}\mathbf{\Upsilon}_1$, $-2L\ \mathrm{Im}\mathbf{\Upsilon}_1$, $2L\ \mathrm{Re}\mathbf{\Upsilon}_2$, $-2L\ \mathrm{Im}\mathbf{\Upsilon}_2$ and $-2L\ \mathrm{Im}\Delta_1$ respectively, where the matrices $\Delta_1, \Delta_2, \Delta_3, \mathbf{\Upsilon}_1, \mathbf{\Upsilon}_2, \mathbf{\Upsilon}_3$ are defined in (3.41)–(3.46).

We are now ready to calculate $\mathrm{CRB}(\boldsymbol{\tau}, \boldsymbol{\theta})$. We have established that the entries in $\mathrm{CRB}(\boldsymbol{\xi})$ due to $\mathbf{k}_m\ , m = 1, \ldots, M^2N^2$ do not affect $\mathrm{CRB}(\mathbf{w})$, $\mathrm{CRB}(\boldsymbol{\tau})$ or $\mathrm{CRB}(\boldsymbol{\theta})$.

From (C.2) we can now write $\mathrm{CRB}(\overline{\mathbf{w}}, \widetilde{\mathbf{w}}, \boldsymbol{\tau}, \boldsymbol{\theta})$ as

$$\mathrm{CRB}^{-1}(\overline{\mathbf{w}}, \widetilde{\mathbf{w}}, \boldsymbol{\tau}, \boldsymbol{\theta}) = 2L \begin{pmatrix} \mathrm{Re}\Delta_1 & \mathrm{Im}\Delta_1 & \mathrm{Re}\boldsymbol{\Upsilon}_1^H & \mathrm{Re}\boldsymbol{\Upsilon}_2^H \\ -\mathrm{Im}\Delta_1 & \mathrm{Re}\Delta_1 & \mathrm{Im}\boldsymbol{\Upsilon}_1^H & \mathrm{Im}\boldsymbol{\Upsilon}_2^H \\ \mathrm{Re}\boldsymbol{\Upsilon}_1 & -\mathrm{Im}\boldsymbol{\Upsilon}_1 & \mathrm{Re}\Delta_2 & \mathrm{Re}\boldsymbol{\Upsilon}_3 \\ \mathrm{Re}\boldsymbol{\Upsilon}_2 & -\mathrm{Im}\boldsymbol{\Upsilon}_2 & \mathrm{Re}\boldsymbol{\Upsilon}_3^H & \mathrm{Re}\Delta_3 \end{pmatrix}. \tag{C.25}$$

We define the matrices $\mathbf{H}_1, \mathbf{H}_2$ and $\mathbf{G}$ as follows:

$$\mathbf{H}_1 = \begin{pmatrix} \mathrm{Re}\Delta_1 & \mathrm{Im}\Delta_1 \\ -\mathrm{Im}\Delta_1 & \mathrm{Re}\Delta_1 \end{pmatrix}, \quad \mathbf{H}_2 = \begin{pmatrix} \mathrm{Re}\Delta_2 & \mathrm{Re}\boldsymbol{\Upsilon}_3 \\ \mathrm{Re}\boldsymbol{\Upsilon}_3^H & \mathrm{Re}\Delta_3 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathrm{Re}\boldsymbol{\Upsilon}_1 & -\mathrm{Im}\boldsymbol{\Upsilon}_1 \\ \mathrm{Re}\boldsymbol{\Upsilon}_2 & -\mathrm{Im}\boldsymbol{\Upsilon}_2 \end{pmatrix}.$$

With the above definitions $\mathrm{CRB}(\overline{\mathbf{w}}, \widetilde{\mathbf{w}}, \boldsymbol{\tau}, \boldsymbol{\theta})$ now simplifies to

$$\mathrm{CRB}^{-1}(\overline{\mathbf{w}}, \widetilde{\mathbf{w}}, \boldsymbol{\tau}, \boldsymbol{\theta}) = 2L \begin{pmatrix} \mathbf{H}_1 & \mathbf{G}^\top \\ \mathbf{G} & \mathbf{H}_2 \end{pmatrix}. \tag{C.26}$$

It can be shown [28] with the aid of a standard matrix inversion result that (C.26) results in an expression for $\mathrm{CRB}(\boldsymbol{\tau}, \boldsymbol{\theta})$ of the form

$$\mathrm{CRB}^{-1}(\boldsymbol{\tau}, \boldsymbol{\theta}) = 2L \left(\mathbf{H}_2 - \mathbf{G}^\top \mathbf{H}_1^{-1} \mathbf{G}\right). \tag{C.27}$$

Since Δ_1 is hermitian, it is easily seen that

$$\mathbf{H}_1^{-1} = \begin{pmatrix} \mathrm{Re}\Delta_1^{-1} & \mathrm{Im}\Delta_1^{-1} \\ -\mathrm{Im}\Delta_1^{-1} & \mathrm{Re}\Delta_1^{-1} \end{pmatrix} \tag{C.28}$$

and subsequently, incorporating (C.28) in (C.27) we obtain the result in Theorem 3.1.