RESOURCE CONSTRAINED ADAPTIVE SENSING

by

Raghuram Rangarajan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2007

Doctoral Committee:

Professor Alfred O. Hero III, Chair
Professor Jeffrey A. Fessler
Professor Susan A. Murphy
Professor Demosthenis Teneketzis
Assistant Professor Clayton Scott

To my mom, my dad, and my brother

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks and deepest gratitude to Professor Alfred Hero for his invaluable guidance, encouragement, and patience during the course of my research. Through the years, I have come to admire Professor Hero's vast knowledge and deep insight on any scientific field, creativity in problem solving, and his exceptional time management skills. I consider myself extremely lucky to have found an advisor in Professor Alfred Hero and his attributes will definitely exert a great influence in all my future endeavors.

My sincere thanks and gratitude also goes to Raviv Raich, with whom I have collaborated on many of my research topics. His invaluable inputs on my research and his ability to breakdown problems have helped me find solutions much quicker and more efficiently. It has also been an absolute pleasure interacting with him on a day-to-day basis for the last 3 years on many other topics of research and life in general.

I am grateful to my committee members Professor Jeffrey Fessler, Professor Susan Murphy, Professor Demosthenis Teneketzis, and Professor Clayton Scott for their valuable input on my work and their helpful comments on my dissertation. I would like to extend my sincere thanks to all the faculty members at the EECS department for giving me a solid foundation on different topics in electrical engineering. Many thanks to my lab mates Michael Ting, Arvind Rao, Xing Zhou, Mark Kliger, Neal Patwari, Doron Blatt, and Dongxiao Zhu for creating a great work atmosphere and

the nice interactions over coffee.

This section would not be complete without acknowledging the support of my friends who have always be there for me and who have made my life in Ann Arbor so much more fun and memorable. In particular, my deep thanks to Ambalavanan Jayaraman, Somesh Srivastava, Deepak Sivaraman, Chandrasekaran Sethu, Arun Ramamurthy, Arun Kumar Panneerselvam, Satish Kumar, Nandakumar Vasudevan, and Harsha Chavali. Thanks also to the entire cricketing community and my friends at Indian Students Association for giving me exciting times.

Last but not the least, I would like to extend my love to my entire family - my mom, my dad, my brother, my sister-in-law, and my grandparents for their constant source of encouragement, unconditional support, sacrifices, and prayers throughout my life. They have been my pillars of strength and my source of energy. The credit to all my academic achievements and to every positive quality I possess goes to my parents.

Special thanks to my fiancèe Sujatha for her love, support, and encouragement during my last year of research.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER I

Introduction

1.1 Overview

Adaptive sensing has been an evolving field in signal processing over the past half century. The goal in adaptive sensing is to control the data acquisition process through adaptive design of system parameters based on acquired information to optimize performance. Adaptive sensing accounts for uncertainties, distortions, and changes in the sensing environment thereby leading to improved performance and better system efficiency over non-adaptive sensing methods. For example, in target detection, waveform adaptation can be exploited to mitigate the effect of a heavily cluttered environment and improve detection performance. As another example, agile radar systems can adaptively control their beam direction based on wide beam coarse resolution initial measurements to focus on the desired targets.

Many adaptive sensing tasks in radar imaging, communication systems, and wireless sensor networks are limited in resources such as energy, time, and bandwidth. Hence resource constraints form an important component in adaptive sensing, where decisions are made online using optimal resource allocation. For example, in radar imaging, an adaptive beamformer is constrained to detect a target in minimum time. In communication systems, there is a restriction on the number of transmitted wave-

forms at any time due to bandwidth constraints. In sensor networks, energy usage is constrained as replacing batteries in remote sensors may be infeasible. It is critical to consider resource limitations while designing data measurement strategies for adaptive sensing applications.

This dissertation develops new methods for resource constrained adaptive sensing in the context of parameter estimation and detection, sensor management, and target tracking. While the results in this thesis are applicable to many problems, we will illustrate our approach using examples from radar imaging and sensor networks.

1.2 Motivation

1.2.1 Radar imaging

Radar is a system that uses electromagnetic (EM) waves to identify the range, speed, altitude, and direction of moving and fixed objects such as aircrafts, tanks, and terrain. Radar was invented in 1944 during the second world war. Since then, radar has continued to be an important modality in surveillance applications such as target tracking and target detection (e.g., mines, trucks, and aircrafts) [143]. More recently, it has also been used for civilian applications such as auto and aircraft anti-collision systems, medical tomography, environmental monitoring [103,138], and autonomous navigation [33]. The main principle in radar is that EM waves scatter or reflect of objects that have significantly different dielectric properties. The strength of backscatter from a particular object depends on a variety of parameters and conditions such as frequency, polarization, observation angles, environment, weather, and shape of the target.

The objective in radar imaging is to form an accurate image of the distribution of the complex valued conductance of an unknown medium based on the radar backscatter or forward scatter. To image a random medium, the imaging area in space is

divided into voxels, volume elements in a three dimensional space, analogous to pixels in 2D space. Each voxel is often characterized by its reflection coefficient, which is the measure of the complex reflectivity of the medium and depends on the conductance of the target at a particular location. Radar imaging refers to the estimation of these reflection coefficients.

A mono-static radar imaging algorithm involves transmitting a waveform or a signal through the medium, measuring the backscattered signals reflected back from the objects or targets, and processing the received signals to form an image of the medium. Classical methods for focusing antenna sensitivity in a particular angular direction are performed using directional antennas [159] or by multiple antenna beamforming [38]. Recently, there has been a growing interest in cognitive radar (CR) [65] that can dynamically adapt to a changing environment. Such an adaptive sensing process can overcome several sources of active and passive interference such as noise, clutter, and jamming and improve the performance of radar imaging.

Moreover, the design of the transmitted waveforms is performed under an average energy constraint. Energy constraints are important in stealth applications, where accurate estimation is performed at low signal-to-noise ratios to avoid being detected by adversaries. A radar system must also detect and image targets in a limited time, which constrains the total transmit energy. Radar signal design subject to both average and peak power constraints is addressed in [8] and [142] using a control theoretic approach. There, the design is non-adaptive and the optimal continuous waveforms are shown to be on-off measurement patterns alternating between zero and peak power levels for a tracking example. Parameterized waveform selection for dynamic state estimation is explored in [78] and [79] where the shape of the waveforms are allowed to vary under constant transmit power. Closed-form solutions

to the parameter selection problem are found for a very restrictive set of cases such as one-dimensional target motions. More recently there has been a growing interest in dynamic waveform selection algorithms for tracking and detection [151]. However, there has been little effort in developing adaptive waveform design strategies that allocate different amounts of energy to the waveforms over time.

The main focus of this dissertation is to show that considerable performance improvements can be achieved using adaptive waveform design and optimal energy allocation for parameter estimation and detection problems. The theoretical results presented here can be applied to radar imaging, communications, sensor networks, and other applications where energy constraints are relevant.

The following section reviews resource allocation problems for target tracking and detection and strategies for solving them.

1.2.2 Target and sensor localization in sensor networks

A wireless sensor network system consists of spatially distributed sensors that collect data about the environment. With recent advances in radio frequency (RF) and micro-electro-mechanical systems (MEMS) integrated circuit (IC) design, the deployment of a dense network of wireless sensors has become feasible for a number of monitoring and control applications such as target tracking [94], environmental monitoring [95], manufacturing logistics [89], geographic routing, and precision agriculture [169].

Target tracking using sensor networks has been of significant interest in many military and civilian applications such as surveillance, vehicle tracking, robotics, biological research, and automotive collision warning systems. Depending on the models for the target trajectory and sensor measurements, tracking algorithms based on the Kalman Filter [120], extended Kalman filter [73], and Gaussian sum approxi-

mations [4] have been proposed. Particle filtering methods were then proposed for tracking, where the probability density of the state of the target (e.g., physical coordinates, velocity) is approximated on a set of discrete points [44]. While most target tracking algorithms assume the knowledge of sensor locations, this is often not the case. For dense networks, it is often impractical to manually position the sensors or too expensive to attach a GPS to every device in the network. To detect and track targets, the sensors must be able to automatically estimate their relative positions in the network. Several sensor self-localization algorithms [35, 39, 87, 88, 111, 124] have been proposed in the literature.

To extend the longevity of battery operated sensors, the tasks of sensor and target localization must be performed using minimal energy usage. The conservation of power in a sensor network can be performed in various ways, e.g., through optimal energy allocation for sensing, through minimal inter-sensor communications, and by optimal sensor scheduling. Since model-based tracking requires detailed sensed information and is computationally intensive, researchers have looked at the simpler problem of tracking in a binary sensing modality [6, 82]. The sensor outputs a high value, when the target is within a sensing range and outputs a low value, when the target falls outside its range. Based on the fusion of the sensor outputs, an approximate link level trajectory can be realized to track the target. This approach for a simple binary sensing measurement model is shown to require minimal power and is also analytically tractable [149].

Significant energy savings can also be realized by minimizing inter-sensor communications while implementing sensor localization algorithms. Classical methods for self-localization use multidimensional scaling (MDS) [124]; maximum likelihood estimation [108]; and convex optimization [45] algorithms. These are centralized

approaches, where all the sensor data measurements need to be communicated to a fusion center which computes the sensor location estimates. Such a process requires many inter-sensor measurements and high bandwidth. The number of inter-sensor communications can be reduced if the sensors can self-localize in a distributed fashion. This has resulted in many decentralized algorithms based on adaptive trilateration [111, 139] and successive refinement [35, 72] methods, which have shown to yield accurate sensor location estimates using much lower energy consumption than centralized methods.

Adaptive sensor management for target localization can further extend the service life of the operating sensors. The task in sensor scheduling is to activate the sensors optimally based on the current and past measurements to conserve power. The sensor scheduling strategy depends on the objective function to be optimized. Examples of objective functions that have been considered include minimizing track error, minimizing probability of false detection, and maximizing Fisher information [85]. For minimizing state estimation error, the optimal sensor to be chosen at the next time instant can be determined a priori and independent of measurements for the case of linear Gaussian systems [7,105]. For other general models, long term optimal scheduling of sensors is combinatorially hard and hence, implementable suboptimal solutions have been proposed [14, 86].

In this dissertation, we develop a novel sparsity penalized MDS algorithm for blind target tracking, i.e., a sensor network that can simultaneously track targets and localize sensors in the absence of prior knowledge of sensor locations. Furthermore, we also develop sensor and power management strategies that can be combined with our blind tracking algorithm to obtain further energy savings.

Next, we describe the advancements and contributions of this thesis in the context

Adaptive design	Parameter	Estimation		Detection
		Gau	Ryl	
two-step	scalar	✓	✓	✓
	multiple	✓		✓
multi-step	scalar	✓		
	multiple	✓		

Table 1.1: Research effort in the context of adaptive energy design for estimation and detection problems. Key: Gau: Gaussian, Ryl: Rayleigh.

of resource constrained adaptive sensing.

1.3 Contributions

The following lists contributions, where they appear in this dissertation, and corresponding publications.

- We develop the theory for adaptive waveform amplitude design (energy allocation) for estimating and detecting parameters in an unknown channel/medium under average energy constraints. The different modalities we considered are as follows (refer Table 1.1 for a summary):
  - Performance gain of at least 5dB over non-adaptive methods for estimating static parameters in linear Gaussian channel model using a 50 time step procedure (Chapter II and [126, 129, 130]).
  - Improvement of at least 1.6dB for estimation in a Rayleigh scattering channel model using a suboptimal adaptive two-step strategy (Chapter III and [127]).
  - Over 2dB improvement in performance for binary hypothesis testing in a linear Gaussian channel model using an optimal two-step strategy (Chapter IV).

- In the context of waveform design/sensor management, we develop a novel convex relaxation based adaptive waveform (sensor) subset selection algorithm for dynamic parameter estimation, which provides a linear time solution to the combinatorially hard problem (Chapter V and [128]).
- For target and sensor localization, we propose the sparsity penalized MDS algorithm with the following features (Chapter VI and [125, 131]):
  - Blind: the algorithm is capable of estimating sensor locations as well as target tracks in the absence of prior knowledge of sensor locations.
  - Decentralized: we perform the target and sensor location estimation locally through limited inter-sensor communications to minimize energy.
  - Fast convergence: The proposed iterative localization algorithm yields non-increasing cost functions and converges in a few iterations.
  - Robust: the MDS based link level tracking algorithm does not require specific target motion models and outperforms the likelihood ratio test based target localization. The nonparametric nature of our algorithm makes it attractive when RSS models are unavailable or inaccurate.
  - Multimodal: The algorithm has the capabilities to operate at link level when the target dynamics is poorly specified. However, it can incorporate a detailed target motion model into the framework, when available.

1.3.1 Personal perspective

The research results presented in this thesis for different problems vary in their depth of theoretical rigor and ease of practical implementation. In terms of theoretical contributions, I believe that my most significant work was the part on adaptive energy allocation strategies for the following reasons: I consider the formulation and

solution to the problem of optimal energy designs for a class of parameter estimation problems very challenging. The asymptotically optimal and suboptimal designs developed in this dissertation present significant performance improvements, which will interest the theorists from the adaptive control community. Moreover, these gains can also impact various research efforts where energy constraints are relevant, e.g., obtaining surveillance images in quickest time, limiting radiation on patients in medical imaging, or extending operating life of sensor networks. The energy designs for estimation in a Rayleigh scattering model and to detection and classification in linear models were natural extensions to the problem of estimation in linear models.

In terms of contributions towards direct applicability to the real world, the development of a sparsity penalized MDS algorithm for target and sensor localization presents a valuable contribution to a sensor network community on many counts. First, through our designs, we were able to show considerable theoretical gains over conventional methods in the context of target detection. More importantly, the proposed algorithm is designed to be implemented in real-time which increases its reach and impact on real world sensor network platforms. Finally, the performance improvements obtained on two real world data sets (ZebraNet and the UCSD wireless topology discovery) despite the many caveats present in them further reemphasizes the practical applicability of our algorithm.

1.4 Detailed outline of thesis

This thesis can be broadly divided as follows: the first four chapters deal with adaptive waveform and energy design in the context of radar imaging and channel estimation applications. Chapter V attempts to solve the waveform (sensor) subset selection problem. In Chapter VI, we focus on target and sensor localization in sensor

networks. Below, we provide an outline of each chapter in this thesis.

In Chapter II, we develop the theory of optimal energy allocation strategies for estimating deterministic parameters in a linear Gaussian model under an average energy constraint. We first derive the optimal energy design procedure for the estimation of a scalar parameter and the corresponding mean-squared error (MSE) for a two time step case. This adaptive two-step strategy yields an improvement of 1.7dB relative to the optimal non-adaptive strategy. We then propose a novel suboptimal solution to the N-step strategy and show an improvement of more than 5dB for $N = 50$. For the case of a vector parameter, we consider two different criteria for minimization of the MSE matrix: the minmax criterion and the trace criterion. For the minmax criterion, we prove that the performance improvements obtained for the scalar parameter case can be achieved for the vector parameter case. For the trace criterion, we show that energy control does not provide any advantages compared to the one-step strategy when the number of parameters to be estimated tends to infinity. However, we prove that using suitable waveform and energy control, it is possible to achieve an improvement of more than 5dB similar to the scalar parameter case. Applications of our results to MIMO and inverse scattering channel models are discussed.

Motivated by the energy design strategies for a linear model, Chapter III considers the problem of adaptive energy allocation for estimation in a Rayleigh fading model. In particular, we derive the optimal two-step energy allocation strategy for a Rayleigh inverse scattering problem. Since there is no closed-form solution to the adaptive two-step strategy, we present a suboptimal solution that yields a signal-to-noise ratio (SNR) improvement of 1.6dB over the non-adaptive design.

Chapter IV presents the problem of designing optimal energy allocation for binary

hypotheses testing in a linear Gaussian model under an average energy constraint. We consider a two time step problem, where the objective is to design the optimal energy allocation at the second step as a function of the first measurement. We develop optimal designs for the frequentist and Bayesian approaches to signal detection. We show that the optimal strategies yield a performance improvement of at least 2dB over the non-adaptive procedure.

In Chapter V, we divert our attention from optimal energy allocation to optimal waveform selection. The goal in this framework is to adaptively select a small subset of waveforms from a large waveform library that minimizes state prediction MSE given the past observations. This adaptive strategy differs from previous approaches to the subset selection problem since the optimal waveforms cannot be computed offline; it requires the previous observations. Since the optimal solution to the subset selection problem is combinatorially complex, we propose a convex relaxation to the problem and present a low complexity suboptimal solution. We consider a hidden Markov model (HMM) representation for the state and apply our algorithm to estimate the state in order to minimize the MSE. Through simulations, we show that the performance of this suboptimal procedure approaches that of the optimal waveform selector.

In Chapter VI of this dissertation, we consider the problem of target tracking in sensor networks. We assume no prior knowledge of the sensor locations and so we refer to such a problem as 'blind' tracking. Since any sensor localization algorithm can recover the sensor location estimates only up to a rotation and translation, we propose a novel sparsity penalized MDS algorithm to align the current time sensor location estimates to those of the previous time-frames. In the presence of a target, only location estimates of those sensors in the vicinity of a target change

relative to their previously estimated locations. Based on the differences in the sensor location estimates between two time-frames, we propose a novel perturbation based link level tracking algorithm, which localizes a target to within a small set of sensor links. Through a detailed numerical study, we show that our tracking algorithm outperforms the conventional likelihood ratio test (LRT) based tracking. In the absence of a target trajectory model, we also suggest methods for translating this link level estimate to actual target coordinates. As an application of our algorithm to real data, we simulate a sensor network collecting data on zebra tracks from the Sweetwater Game Reserve in Kenya available from the ZebraNet project undertaken by Princeton biologists. We also implement our localization algorithm for discovering the wireless network topology of the University of California, San Diego (UCSD) campus from real-time user trace data.

Conclusions and future research directions are presented in Chapter VII of this dissertation.

1.5 Connections

Chapter VI primarily focuses on algorithms for simultaneous target and sensor localization. Keeping in mind the issue of limited power and bandwidth, the algorithms can be implemented distributively to minimize communication and computational costs. However, further energy savings can be realized in this sensor network framework by making use of the resource constrained adaptive sensing designs developed in the earlier chapters. For example, consider the problem of sensor scheduling for target tracking, i.e., we need to activate only a small subset of sensors in the network to identify the targets. This problem can be cast as a waveform selection problem and we can borrow the sparsity constrained convex relaxation solution in Chapter V

to solve for the optimal sensor activation strategy. Given the activated sensors, the probability of the detection of a target can be improved from the energy allocation procedures described in Chapter IV. The sensor measurements need to be integrated over a period of time which can then be used for accurate target tracking. Depending on the integration time and the processing capabilities of the sensors, we can use the different temporal energy allocation designs from Chapters II and III to decrease the MSE in the estimation of the target locations.

1.6 Publications

The following publications are a product of the research covered by this thesis.

1. R. Rangarajan, R. Raich, and A. O. Hero III. Sparse multidimensional scaling for blind tracking in sensor networks'. To appear in *Advances in Sensor Networks*, Ed. V. Saligrama, Springer, NY, Dec. 2007.

2. R. Rangarajan, R. Raich, and A. O. Hero III. Blind tracking using sparsity penalized multidimensional scaling. To appear in *Proc. IEEE Workshop on Stat. Signal Processing*, Aug. 2007.

3. R. Rangarajan, R. Raich, and A. O. Hero III. Optimal sequential energy allocation for inverse problems. *IEEE J. Select. Topics in Signal Processing*, 1(1):67–78, Jun. 2007.

4. R. Rangarajan, R. Raich, and A. O. Hero III. Sequential energy allocation strategies for channel estimation. *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, 3:821–824, Apr. 2007.

5. R. Rangarajan, R. Raich, and A. O. Hero III. Single-stage waveform selection for adaptive resource constrained state estimation. *Proc. IEEE Intl. Conf.*

Acoust., Speech, Signal Processing, 3:672–675, May 2006.

6. R. Rangarajan, R. Raich, and A. O. Hero III. Sequential design for a Rayleigh inverse scattering problem. *Proc. IEEE Workshop on Stat. Signal Processing*, pages 625–630, Jul. 2005.

7. R. Rangarajan, R. Raich, and A. O. Hero III. Optimal experimental design for an inverse scattering problem. *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, 4:1117–1120, Mar. 2005.

8. R. Rangarajan and A. O. Hero III. Analysis of a multistatic adaptive target illumination and detection approach (MATILDA) to time reversal imaging. *IMA Hot topics Workshop on Adaptive Sensing and Multimode Data Inversion*, Jun. 2004.

CHAPTER II

Energy allocation for estimation in linear models

The context of this chapter is adaptive waveform design for estimating parameters of an unknown channel under average energy constraints. This chapter focuses on the simpler problem of adaptive waveform-amplitude design for which we obtain interesting analytical results. We treat an N-step design problem where a fixed waveform can be transmitted into the channel N times with amplitudes that can be chosen as a function of past channel outputs. For $N = 2$ and a linear Gaussian channel model, we derive the optimal amplitude to transmit at the second step as a function of the first measurement. This adaptive 2-step energy allocation strategy yields a mean-squared error (MSE) improvement of at least 1.7dB relative to the optimal non-adaptive strategy. Motivated by the optimal two-step strategy we propose a suboptimal adaptive N-step strategy that can achieve an MSE improvement of more than 5dB for $N = 50$. Applications of our results to MIMO and inverse scattering channel models are discussed.

2.1 Introduction

One of the important components in adaptive sensing problems such as channel estimation and radar imaging is energy management. Most applications are limited by peak power or average power. For example, in sensor network applications, sensors

have limited battery life and replacing them is expensive. Safety limits the peak transmit power in medical imaging problems. Energy is also a critical resource in communication systems where reliable communication is necessary at low signal-to-noise ratios. Hence it is important to consider energy limitations in waveform design problems. Most of the effort in previous research has focussed on waveform design under peak power constraints, e.g., sensor management. There has been little effort in developing adaptive waveform design strategies that allocate different amounts of energy to the waveforms over time.

Our goal in this chapter is to perform waveform amplitude design for adaptive sensing to estimate the set of unknown channel parameters or scattering coefficients under an average energy constraint. We formulate this problem as an experimental design problem in the context of sequential parameter estimation. We explain the methodology of experimental design, derive optimal designs, and show performance gains over non-adaptive design techniques. As a final step, we describe in detail how some applications of adaptive sensing such as channel estimation and radar imaging can be cast into this experimental design setting thereby leading to attractive performance gains compared to current literature. Next, we present a review of waveform design literature to provide a context for our work.

Note: The term 'sequential' is used in different contexts in the literature. In this chapter, 'sequential' means that at every time instant, the best signal to transmit is selected from a library that depends on past observations.

2.1.1 Related work - waveform design

Early work in waveform design focussed on selecting among a small number of measurement patterns [41]. Radar signal design using a control theoretic approach subject to both average and peak power constraints was addressed in [8] and [142].

The design was non-adaptive and the optimal continuous waveforms were shown to be on-off measurement patterns alternating between zero and peak power levels for a tracking example. In our design, the energy allocation to the waveforms over time are optimally chosen from a continuum of values. Parameterized waveform selection for dynamic state estimation was explored in [78] and [79] where the shape of the waveforms were allowed to vary under constant transmit power. Closed-form solutions to the parameter selection problem were found for a very restrictive set of cases such as one-dimensional target motions. More recently a dynamic waveform selection algorithm for tracking using a class of generalized chirp signals was presented in [151]. In contrast to these efforts, we focus our work in finding optimal waveform amplitudes under an average energy constraint for static parameter estimation. Sensor scheduling can be thought of as an adaptive waveform design problem under a peak power constraint [51] where the goal is to choose the best sensor at each time instant to provide the next measurement. The optimal sensor schedule can be determined a priori and independent of measurements for the case of linear Gaussian systems [7, 105]. The problem of optimal scheduling for the case of hidden Markov model systems was addressed in [86]. In table 2.1, we compare our work with existing literature via different categories.

2.1.2 Related work - sequential design for estimation

The concept of sequential design has been studied by statisticians for many decades [31, 56, 74, 80, 133, 144] and has found applications in statistics, engineering, biomedicine, and economics. Sequential analysis has been used to solve important problems in statistics such as change-point detection [63, 156], point and interval estimation [115], multi-armed bandit problems [135], quality control [147], sequential testing [167], and stochastic approximation [134]. Robbins pioneered the

Literature	Type of parameters				Type of design		Type of constraint			Type of control			
	D	R	LSD	NLSD	SQ	NSQ	EN	SN	NONE	EN	WV	SN	NONE
Waveform design [41]			✓	✓	✓				✓		✓		
Sensor scheduling [7, 86, 105]			✓	✓	✓			✓				✓	
Sequential estimation [23, 63]	✓				✓				✓				✓
Schweppe's design [8, 142]			✓			✓	✓			✓			
RLS [90]	✓				✓				✓				✓
Stein estimator [71, 158]	✓					✓			✓				✓
Kalman filter [120]			✓		✓				✓				✓
Our sequential approach	✓				✓		✓			✓			

Table 2.1: Key to the table: D-deterministic, R-random, LSD-Linear state dynamics, NLSD-non linear state dynamics, SQ-Sequential design, NSQ-Non sequential design, EN-energy, SN-sensors, WV-waveform parameters.

statistical theory of sequential allocation in his seminal paper [135]. Early research on the application of sequential design to problems of estimation was limited to finding asymptotically risk-efficient point estimates and fixed-width confidence intervals [23, 63, 83], i.e., sequential design was used to solve problems in which a conventional estimate, based on a sample whose size is determined by a suitably chosen stopping rule, achieves certain properties such as bounded risk. For the problem of estimating the mean under unknown variance, it was shown that a sequential two-step method guaranteed specified precision [5, 156, 157], which is not possible using a fixed sample. The statistical sequential design framework assumes a fixed measurement setup while acquiring the data and does not consider energy constraints. In this chapter, we adaptively design input parameters to alter the measurement patterns under an average energy constraint to obtain performance gains over non-adaptive strategies.

Another class of problems in sequential estimation is online estimation, where fast updating of parameter estimates are made in real time, called recursive identification in control theory, and adaptive estimation in signal processing. For example, consider the problem of estimating parameter $\boldsymbol{\theta}$ in the following model

$$y_i \quad = \quad \mathbf{x}_i^T \boldsymbol{\theta} + w_i, \quad i = 1, 2, \ldots, n,$$

where $\{\mathbf{x}_i\}$ are the sequence of inputs to the system, $\{w_i\}$ are independent identically distributed (i.i.d) Gaussian random variables with zero mean and $\{y_i\}$ are the set of received signals. The maximum likelihood estimate of $\boldsymbol{\theta}$ is given by the least squares (LS) solution, $\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T\right)^{-1} \left(\sum_{i=1}^n \mathbf{x}_i y_i\right)$. One way of computing the LS estimate is the recursive least squares approach (RLS) [90] which can be written

as

$$\begin{aligned}\hat{\boldsymbol{\theta}}_n &= \hat{\boldsymbol{\theta}}_{n-1} + \mathbf{P}_n \mathbf{x}_n (y_n - \mathbf{x}_n^T \hat{\boldsymbol{\theta}}_{n-1}) \\ \mathbf{P}_n &= \mathbf{P}_{n-1} - \frac{\mathbf{P}_{n-1}\mathbf{x}_n \mathbf{x}_n^T \mathbf{P}_{n-1}}{1 + \mathbf{x}_n^T \mathbf{P}_{n-1} \mathbf{x}_n},\end{aligned}$$

where $\mathbf{P}_n = \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T\right)^{-1}$. Using this recursive process, we avoid the complexity of computing the matrix inverse.

In the above formulation it was assumed that the input sequence $\{\mathbf{x}_i\}$ remains fixed. The problem of waveform design is relevant when inputs $\mathbf{x}_i$ can be adaptively chosen based on the past measurements $y_1, \ldots, y_{i-1}$. Measurement-adaptive estimation has application to a wide variety of areas such as communications and control, medical imaging, radar systems, system identification, and inverse scattering. By measurement-adaptive estimation we mean that one has control over the way measurements are made, e.g., through the selection of waveforms, projections, or transmitted energy. The standard solution for estimating parameters from adaptive measurements is the maximum likelihood (ML) estimator. For the case of classic linear Gaussian model, i.e., a Gaussian observation with unknown mean and known variance, it is well-known [120] that the ML estimator is unbiased and achieves the unbiased Cramér Rao lower bound (CRB). Many researchers have looked at improving the estimation of these parameters by adding a small bias to reduce the MSE. Stein showed that this leads to better estimators that achieve lower MSE than the ML estimator for estimating the mean in a multivariate Gaussian distribution with dimension greater than two [71, 158]. Other alternatives such as the shrinkage estimator [104], Tikhonov regularization [165] and covariance shaping least squares (CSLS) estimator [49] have also been proposed in the literature. While these pioneering efforts present interesting approaches to improve static parameter estimation

performance by introducing bias, none of them incorporate the notion of sequential design of input parameters. Our adaptive design of inputs effectively adds bias to achieve reduction in MSE.

In this chapter, we formulate a problem of sequentially selecting waveform amplitudes for estimating deterministic parameters of a linear Gaussian channel model under an average energy constraint over the waveforms and over the number of transmissions. In Section 2.2, the problem of experimental design [52, 171] for sequential parameter estimation is outlined and the analogy between this problem and the waveform design problem is explained. In Section 2.3, closed-form expressions for the optimal design parameters (e.g., energy allocation to the waveforms in the adaptive sensing context) and the corresponding minimum MSE in the single parameter (e.g., scatter coefficients in imaging, channel coefficients in channel estimation) case are derived for a two-step procedure (two time steps). In Section 2.4, we provide a suboptimal design for the two-step strategy, which takes into consideration a peak power constraint and achieves near optimal performance. Since the optimal solution requires the knowledge of parameters to be estimated, it is shown in Section 2.6 that the performance of this omniscient solution can be achieved with a parameter independent strategy. In Section 2.7, we describe an N-step sequential energy allocation procedure, which yields more than 5dB gain over non-adaptive methods. These results are extended to the vector parameter case in Section 2.8. Finally in Section 2.9, we show the applicability of this framework by recasting the problems of channel estimation and radar imaging to fit the statistical model of the sequential parameter estimation problem and applying the results from the previous sections to show the advantages of our approach over current literature for practical applications.

2.2 Problem statement

We denote vectors in $\mathbb{C}^M$ by boldface lower case letters and matrices in $\mathbb{C}^{M\times N}$ by boldface uppercase letters. The symbol $\|\cdot\|$ refers to the l_2-norm of a vector, i.e., $\|\mathbf{x}\| = \sqrt{\mathbf{x}^H\mathbf{x}}$, where $(\cdot)^H$ denotes the conjugate transpose. The terms MSE and SNR are abbreviations to mean-squared error and signal-to-noise ratio, respectively. Let $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_M]$ be the M-element vector of unknown parameters. The problem of estimating $\boldsymbol{\theta}$ in noise can then be written as

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta}) + \mathbf{n}_i, \quad i = 1, 2, \ldots, N, \tag{2.1}$$

where $\{\mathbf{n}_i\}$ is an i.i.d. random process corrupting the function of the parameters of interest $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta})$ and i denotes the time index. The T-element design parameter vectors, $\{\mathbf{x}_i\}_{i=1}^N$ can depend on the past measurements: $\mathbf{x}_i = \mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$, where $\mathbf{y}_i$ is the i^{th} K-element observation vector. In the context of adaptive sensing, $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta})$ represents the response of the medium, T and K denote the number of transmit and receive antennas respectively, $\{\mathbf{x}_i\}_{i=1}^N$ are the set of waveforms to be designed, $\boldsymbol{\theta}$ are the set of channel parameters or scattering coefficients to be estimated using the set of received signals $\{\mathbf{y}_i\}_{i=1}^N$. For the classic estimation problem in a linear Gaussian model, we have $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{H}(\mathbf{x}_i)\boldsymbol{\theta}$, $\mathbf{H}(\mathbf{x}_i) = [\mathbf{h}_1(\mathbf{x}_i), \mathbf{h}_2(\mathbf{x}_i), \ldots, \mathbf{h}_M(\mathbf{x}_i)]$ is a known $K \times M$ matrix and linear in $\mathbf{x}_i$ and $\mathbf{n}_i$ is a circularly symmetric complex Gaussian random variable with zero mean and covariance $\sigma^2\mathbf{I}$ denoted by $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. When $\mathbf{H}(\mathbf{x})$ is linear in $\mathbf{x}$, we can write $\mathbf{h}_l(\mathbf{x}) = \boldsymbol{H}_l\mathbf{x}, l = 1, 2, \ldots, M$. In this case $\mathbf{H}(\cdot)$ is uniquely determined by the matrices $\{\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_M\}$. The linear Gaussian model has been widely adopted in many studies [12,106] including channel estimation [15] and radar imaging [148] problems. The set of observations for parameter $\boldsymbol{\theta}$ can

then be written as

$$\mathbf{y}_i = \mathbf{H}(\mathbf{x}_i)\boldsymbol{\theta} + \mathbf{n}_i, \quad i = 1, 2, \ldots, N. \tag{2.2}$$

For the case of a scalar parameter θ_1, the observations are

$$\mathbf{y}_i = \mathbf{h}_1(\mathbf{x}_i)\theta_1 + \mathbf{n}_i, \quad i = 1, 2, \ldots, N. \tag{2.3}$$

An N-step design procedure specifies a sequence of functions $\{\mathbf{x}_i(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{i-1})\}_{i=1}^N$ corresponding to the N transmitted signal waveforms after receiving the previous measurements. An optimal N-step procedure selects the design vectors so that the MSE of the maximum likelihood (ML) estimator, $\hat{\theta}_1^{(N)}(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)$ is minimized subject to the average energy constraint, $\mathrm{E}\left[\sum_{i=1}^N \|\mathbf{x}_i\|^2\right] \leq E_0$, where E_0 is the total available energy and $\mathrm{E}[\cdot]$ denotes the statistical expectation over the space of received measurements. The ML estimator of θ_1 for the N-step procedure is given by

$$\hat{\theta}_1^{(N)} = \frac{\sum_{i=1}^N \mathbf{h}_1(\mathbf{x}_i)^H \mathbf{y}_i}{\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2} \tag{2.4}$$

and the corresponding $\mathrm{MSE}\left(\{\mathbf{x}_i\}_{i=1}^N\right) \triangleq \mathrm{E}\left[\left|\hat{\theta}_1^{(N)} - \theta_1\right|^2\right]$ is

$$\mathrm{MSE}^{(N)}\left(\{\mathbf{x}_i\}_{i=1}^N\right) = \mathrm{E}\left[\left|\frac{\sum_{i=1}^N \mathbf{h}_1(\mathbf{x}_i)^H \mathbf{n}_i}{\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2}\right|^2\right]. \tag{2.5}$$

Denote $E_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}) = \|\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})\|^2$, where $E_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$ represents the energy allocated to each time step i. Define $\mathcal{E}\left[\{\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})\}_{i=1}^N\right]$ as the average energy in the design parameters for the N-step procedure,

$$\mathcal{E}\left[\{\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})\}_{i=1}^N\right] = \mathrm{E}\left[\sum_{i=1}^N \|\mathbf{x}_i\|^2\right]. \tag{2.6}$$

The average energy constraint can be written as

$$\mathcal{E}\left[\{\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})\}_{i=1}^N\right] = \mathrm{E}\left[\sum_{i=1}^N E_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})\right] \leq E_0. \tag{2.7}$$

Our goal is to find the best sequence of the design vectors $\{\mathbf{x}_i\}_{i=1}^N$ to minimize the $\mathrm{MSE}^{(N)}\left(\{\mathbf{x}_i\}_{i=1}^N\right)$ in (2.5) under the average energy constraint in (2.7).

2.2.1 Non-adaptive strategy

As a benchmark for comparison, we consider the non-adaptive case where $\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$ is deterministic, independent of $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{i-1}$, $\|\mathbf{x}_i\|^2 = E_i$, and $\sum_{i=1}^N E_i \leq E_0$. Simplifying the expression for MSE in (2.5), we have

$$\mathrm{MSE}^{(N)} = \frac{\mathrm{E}\left[\left|\sum_{i=1}^N \mathbf{h}_1(\mathbf{x}_i)^H \mathbf{n}_i\right|^2\right]}{\left|\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2\right|^2}$$

Using the fact that $\{\mathbf{n}_i\|_{i=1}^N$ are i.i.d $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$, we obtain

$$\begin{aligned} \mathrm{MSE}^{(N)} &= \frac{\sigma^2}{\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^N E_i \frac{\|\mathbf{h}_1(\mathbf{x}_i)\|^2}{\|\mathbf{x}_i\|^2}} \\ &\geq \frac{\sigma^2}{E_0 \lambda_{\mathrm{m}}(\boldsymbol{H}_1)}, \end{aligned} \tag{2.8}$$

where equality is achieved iff $\forall i$ $\mathbf{x}_i \propto \mathbf{v}_{\mathrm{m}}$, the normalized eigenvector corresponding to $\lambda_{\mathrm{m}}(\boldsymbol{H}_1)$, the maximum eigenvalue of the matrix $\boldsymbol{H}_1^H \boldsymbol{H}_1$. Note

$$\lambda_{\mathrm{m}}(\boldsymbol{H}_1) = \max_{\mathbf{x}} (\mathbf{x}^H \boldsymbol{H}_1^H \boldsymbol{H}_1 \mathbf{x})/(\mathbf{x}^H \mathbf{x}) = \max_{\mathbf{x}} \|\mathbf{h}_1(\mathbf{x})\|^2 / \|\mathbf{x}\|^2. \tag{2.9}$$

Furthermore, the performance of the ML estimator does not depend on the energy allocation. Hence, without loss of generality we can assume all energy is allocated to the first transmission which implies that any N-step non-adaptive strategy is no better than the optimal one-step strategy. We define $\mathrm{SNR}\left(\{\mathbf{x}_i\}_{i=1}^N\right)$ as

$$\mathrm{SNR}^{(N)} = \frac{\lambda_{\mathrm{m}}(\boldsymbol{H}_1)\mathcal{E}\left[\{\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})\}_{i=1}^N\right]}{\sigma^2}. \tag{2.10}$$

Then the average energy constraint in (2.7) is equivalent to $\mathrm{SNR}^{(N)} \leq \mathrm{SNR}_0$, where $\mathrm{SNR}_0 = \lambda_{\mathrm{m}}(\boldsymbol{H}_1)E_0/\sigma^2$. We show in Appendix 2.11 that the problem of minimizing

$\mathrm{MSE}^{(N)}$ subject to $\mathrm{SNR}^{(N)} \leq \mathrm{SNR}_0$ is equivalent to minimizing $\mathrm{MSE}^{(N)} \times \mathrm{SNR}^{(N)}$. Thus we use the two minimization criteria interchangeably in the remainder of this chapter. The product of MSE and SNR is

$$\mathrm{MSE}^{(N)} \times \mathrm{SNR}^{(N)} = \mathrm{E}\left[\left|\frac{\sum_{i=1}^{N} \mathbf{h}_1(\mathbf{x}_i)^H \mathbf{n}_i}{\sum_{i=1}^{N} \|\mathbf{h}_1(\mathbf{x}_i)\|^2}\right|^2\right] \frac{\lambda_{\mathrm{m}}(\boldsymbol{H}_1)\mathrm{E}\left[\sum_{i=1}^{N} \|\mathbf{x}_i\|^2\right]}{\sigma^2} \tag{2.11}$$

and the minimum MSE for the one-step (or non-adaptive N-step) strategy satisfies

$$\mathrm{MSE}^{(1)}_{\mathrm{min}} \times \mathrm{SNR}_0 = 1. \tag{2.12}$$

While our goal is to find optimal input design parameters, $\{\mathbf{x}_j(\mathbf{y}_1, \ldots, \mathbf{y}_{j-1})\}_{j=1}^{N}$ which achieve minimum MSE, any suboptimal design that guarantees $\mathrm{MSE}^{(N)} \times \mathrm{SNR}_0 < 1$ is also of interest. We first look at a two-step sequential design procedure. A word of caution: in Sections 2.3 and 2.4 we develop optimal and suboptimal strategies where the solutions require the knowledge of the unknown parameter θ_1. However, in Section 2.6 we present a θ_1-independent design which asymptotically achieves the performance of the 'omniscient' strategies.

2.3 Omniscient optimal two-step sequential strategy

In the two-step sequential procedure, we have $N = 2$ time steps where in each time step $i = 1, 2$, we can control input design parameter $\mathbf{x}_i$ to obtain observation $\mathbf{y}_i$. For a two-step process, we have

$$\mathbf{y}_1 = \mathbf{h}_1(\mathbf{x}_1)\theta_1 + \mathbf{n}_1 \tag{2.13}$$

$$\mathbf{y}_2 = \mathbf{h}_1(\mathbf{x}_2(\mathbf{y}_1))\theta_1 + \mathbf{n}_2. \tag{2.14}$$

The ML estimator of θ_1 for a two-step procedure from (2.4) is

$$\hat{\theta}_1^{(2)} = \frac{\mathbf{h}_1(\mathbf{x}_1)^H \mathbf{y}_1 + \mathbf{h}_1(\mathbf{x}_2)^H \mathbf{y}_2}{\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2} \tag{2.15}$$

and its MSE from (2.5) is given by

$$\text{MSE}^{(2)}(\mathbf{x}_1, \mathbf{x}_2) = \text{E}\left[\frac{|\mathbf{h}_1(\mathbf{x}_1)^H\mathbf{n}_1 + \mathbf{h}_1(\mathbf{x}_2)^H\mathbf{n}_2|^2}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)^2}\right]. \tag{2.16}$$

We assume that the shape of the optimal designs, i.e., $\{\mathbf{x}_i/\|\mathbf{x}_i\|\}$ is the one-step optimum given by $\mathbf{v}_\text{m}$ defined below (2.8) and minimize the MSE over the energy of the design parameters. In other words, we search for the optimal energy design to the waveforms among those that have their shapes as $\mathbf{v}_\text{m}$. Denote $\|\mathbf{x}_1\| = \sqrt{E_0}\alpha_1$ and $\|\mathbf{x}_2(\mathbf{y}_1)\| = \sqrt{E_0}\alpha_2(\mathbf{y}_1)$. Under the sequential design framework, we select

$$\mathbf{x}_1 = \sqrt{E_0}\ \alpha_1\mathbf{v}_\text{m} \tag{2.17}$$

$$\mathbf{x}_2(\mathbf{y}_1) = \sqrt{E_0}\ \alpha_2(\mathbf{y}_1)\mathbf{v}_\text{m}, \tag{2.18}$$

where α_1 and $\alpha_2(\cdot)$ are real-valued scalars. The average energy constraint from (2.7) can then be written as

$$\text{E}\left[\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right] \leq 1. \tag{2.19}$$

We use Lagrangian multipliers to minimize the MSE in (2.16) with respect to α_1 and $\alpha_2(\cdot)$ under the energy constraint in (2.19). Substituting for $\mathbf{x}_1$ and $\mathbf{x}_2(\mathbf{y}_1)$ given by (2.17) and (2.18) respectively in (2.16) and adding the Lagrangian constraint we obtain the objective function to be minimized as

$$\begin{aligned} & \text{MSE}^{(2)}(\mathbf{x}_1, \mathbf{x}_2) + \gamma(\text{E}\left[\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right]) \\ = \ & \text{E}\left[\frac{|\mathbf{h}_1(\mathbf{x}_1)^H\mathbf{n}_1 + \mathbf{h}_1(\mathbf{x}_2)^H\mathbf{n}_2|^2}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)^2}\right] + \gamma\left(\alpha_1^2 + \text{E}\left[\alpha_2^2(\mathbf{y}_1)\right]\right) \end{aligned}$$

Using linearity of $\mathbf{h}_1(\cdot)$, the objective function can be written as

$$\frac{1}{E_0}\text{E}\left[\frac{|\alpha_1\mathbf{h}_1(\mathbf{v}_\text{m})^H\mathbf{n}_1 + \alpha_2(\mathbf{y}_1)\mathbf{h}_1(\mathbf{v}_\text{m})^H\mathbf{n}_2|^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2\|\mathbf{h}_1(\mathbf{v}_\text{m})\|^4} + \gamma E_0\left(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right)\right]$$

Taking the expectation over $\mathbf{n}_2$, the objective function becomes

$$\frac{1}{E_0}\text{E}\left[\frac{\alpha_1^2|\mathbf{h}_1(\mathbf{v}_\text{m})^H\mathbf{n}_1|^2 + \alpha_2^2(\mathbf{y}_1)\|\mathbf{h}_1(\mathbf{v}_\text{m})\|^2\sigma^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2\|\mathbf{h}_1(\mathbf{v}_\text{m})\|^4} + \gamma E_0\left(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right)\right]$$

Dividing numerator and denominator by $\|\mathbf{h}_1(\mathbf{v}_\mathrm{m})\|^2\sigma^2$, the minimization criterion simplifies to

$$= \frac{1}{\mathrm{SNR}_0}\mathrm{E}\left[\frac{\alpha_1^2\left|\tilde{n}_1(\mathbf{y}_1;\theta_1)\right|^2+\alpha_2^2(\mathbf{y}_1)}{(\alpha_1^2+\alpha_2^2(\mathbf{y}_1))^2}+\gamma\mathrm{SNR}_0\left(\alpha_1^2+\alpha_2^2(\mathbf{y}_1)\right)\right]$$

$$= \frac{1}{\mathrm{SNR}_0}\frac{1}{\alpha_1^2}\mathrm{E}\left[\frac{1}{\left(1+\frac{\alpha_2^2(\mathbf{y}_1)}{\alpha_1^2}\right)}-\frac{1-\left|\tilde{n}_1(\mathbf{y}_1;\theta_1)\right|^2}{\left(1+\frac{\alpha_2^2(\mathbf{y}_1)}{\alpha_1^2}\right)^2}+\gamma^{'}\left(1+\frac{\alpha_2^2(\mathbf{y}_1)}{\alpha_1^2}\right)\right], \tag{2.20}$$

where $\tilde{n}_1(\mathbf{y}_1;\theta_1)=\frac{\mathbf{h}_1(\mathbf{v}_\mathrm{m})^H}{\|\mathbf{h}_1(\mathbf{v}_\mathrm{m})\|}\left(\frac{\mathbf{y}_1-\mathbf{h}_1(\mathbf{x}_1)\theta_1}{\sigma}\right)=\frac{\mathbf{h}_1(\mathbf{v}_\mathrm{m})^H}{\|\mathbf{h}_1(\mathbf{v}_\mathrm{m})\|}\frac{\mathbf{n}_1}{\sigma}$ is a zero mean unit variance complex Gaussian random variable and $\gamma^{'}=\gamma\alpha_1^4\mathrm{SNR}_0$. Since the optimal solution to $\alpha_2(\mathbf{y}_1)$ depends on $\mathbf{y}_1$ only through the function $\tilde{n}_1(\mathbf{y}_1;\theta_1)$, we denote the solution as $\alpha_2\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)$. Let $g\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)=\left(1+\frac{\alpha_2^2(\tilde{n}_1(\mathbf{y}_1;\theta_1))}{\alpha_1^2}\right)$. Then the objective function can be written as

$$\frac{1}{\mathrm{SNR}_0}\frac{1}{\alpha_1^2}\mathrm{E}\left[\frac{1}{g\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)}-\frac{1-\left|\tilde{n}_1(\mathbf{y}_1;\theta_1)\right|^2}{g^2\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)}+\gamma^{'}g\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)\right]. \tag{2.21}$$

Differentiating and setting the objective function to zero, we have

$$g^3-\frac{1}{\gamma^{'}}g+2\frac{1-|\tilde{n}_1|^2}{\gamma^{'}}=0. \tag{2.22}$$

The function g that minimizes MSE is the root of the third-order polynomial in (2.22), real-valued and greater than or equal to 1. If more than one real-valued solution greater than 1 to the cubic equation exists, the optimal solution to g will be the root that achieves minimum MSE. The optimal g for every $\tilde{n}_1$ and $\gamma^{'}$ is denoted by $g_{\gamma^{'}}(\tilde{n}_1)$. Also $\mathrm{E}\left[g_{\gamma^{'}}(\tilde{n}_1)\right]=\frac{1}{\alpha_1^2}$. Therefore, finding α_1 that minimizes MSE is equivalent to finding $\gamma^{'}$ that minimizes MSE. We obtain $g_{\gamma^{'}}(\tilde{n}_1)$ for every $\gamma^{'}$ and use a brute force grid search to find the optimal $\gamma^{'}$ that minimizes the expression in (2.21). The MSE is minimized at $\gamma^{'*}\approx 0.22$, or $\alpha_1^*\approx 0.7421$. The optimal α_2 is given by the relation $\alpha_2^*\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)=\alpha_1^*\sqrt{\left(g_{\gamma^{'*}}\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)-1\right)}$. Since this solution

depends on the unknown parameter θ_1, we call this minimizer an "omniscient" energy allocation strategy. For the optimal solution, the product of MSE $\times$ SNR is

$$\text{MSE}^{(2)}_{\text{min}} \times \text{SNR}_0 \approx 0.68. \tag{2.23}$$

This corresponds to a 32% improvement in performance or a 1.67dB gain in terms of SNR for the two-step design when compared to the one-step procedure for which $\text{MSE}^{(1)}_{\text{min}} \times \text{SNR}_0 = 1$. $\text{MSE}^{(2)} \times \text{SNR}_0$ is plotted for various values of α_1 using both simulations (dotted) and theoretically (solid) in Fig. 2.1.

The theoretical performance curve (solid) was generated by evaluating the $\text{MSE}^{(2)}$ in (2.21) (without the constraint term) for various values of $\gamma^{'}$ (or α_1). Since the expectation in (2.21) depends only on random variable $\tilde{n}_1$, we construct a fine grid of $\tilde{n}_1$ and approximate the integral induced by the expectation as a Riemann sum. For the simulation curve (dotted), we generate 10000 samples of $\mathbf{n}_1$ and $\mathbf{n}_2$ distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. Using $\mathbf{n}_1$ and $\mathbf{x}_1$ from (2.17), we generate 10000 samples of $\mathbf{y}_1$ from (2.13). Using samples of $\mathbf{y}_1$, we generate samples of $\mathbf{x}_2(\mathbf{y}_1)$ from (2.18) and obtain 10000 samples of $\mathbf{y}_2$ from (2.14) using $\mathbf{x}_2(\mathbf{y}_1)$ and $\mathbf{n}_2$. We obtain an estimate of $\text{MSE}^{(2)}$ by evaluating the expectation in (2.16) through numerical averaging over the realizations.

The optimal energy allocation at the second step, $\alpha_2^{*2}\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)$ as shown in Fig. 2.2 (solid) is a thresholding function, i.e., α_2^* is zero for $|\tilde{n}_1|^2 \leq 0.59$. This solution implies that when the actual realization of the normalized noise along $\mathbf{h}_1(\mathbf{v}_{\text{m}})$ in the first step is small enough, then the second measurement becomes unnecessary. On the other hand, when the normalized noise along $\mathbf{h}_1(\mathbf{v}_{\text{m}})$ exceeds a threshold, then there is some merit in incorporating the information from the second measurement. The solution also suggests that the higher the noise magnitude at the first step, the more the energy that needs to be used. However, the probability of allocating

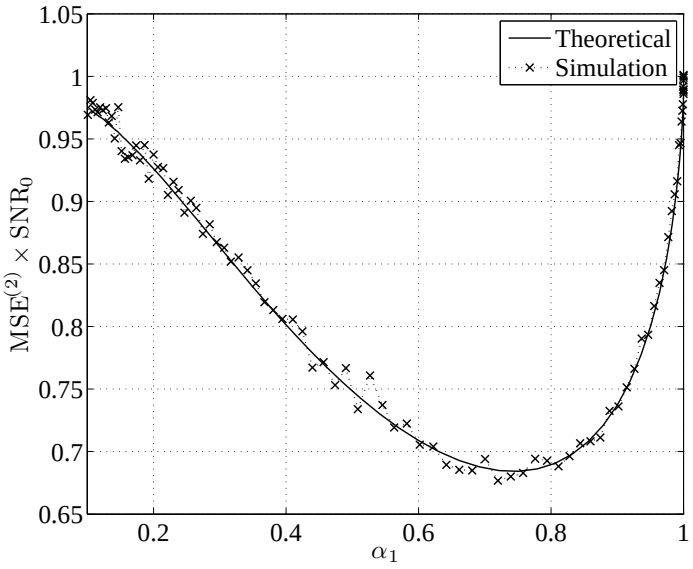


Figure 2.1: Reduction in MSE for varying values of α_1.

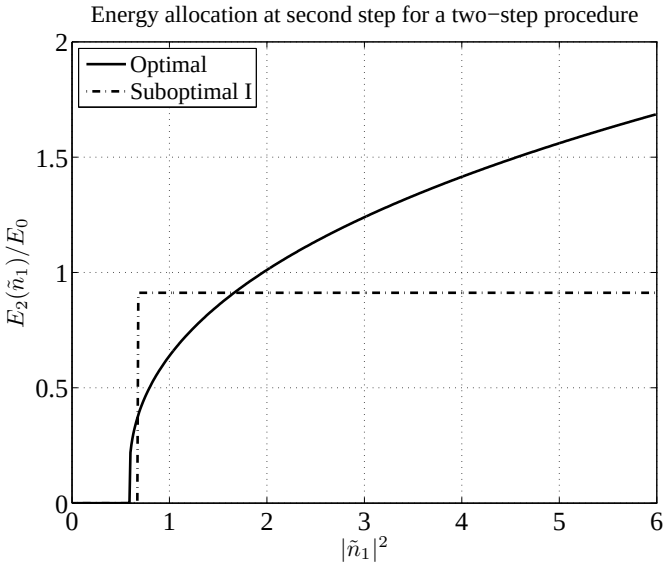


Figure 2.2: Plot of the optimal and suboptimal solution to the normalized energy transmitted at the second stage as functions of received signal at first stage.

energy greater than a particular value decreases exponentially with that energy value. Nevertheless in applications with a peak energy constraint, the transmission of the optimal energy at the second stage may not always be possible. Hence, in the following section we look at a suboptimal solution which takes into account this constraint and still achieves near optimal performance.

2.4 Omniscient suboptimal two-step strategy

The optimal solution in Section 2.3 is a thresholding function, where energy allocated to the second stage is zero if the noise magnitude at the first step is less than a threshold and increases with increasing noise magnitudes otherwise. For the suboptimal solution, we use a binary energy allocation strategy at the second stage based on the noise magnitude at the first step, i.e., we allocate a fixed nonzero energy if the noise magnitude is greater than a threshold else we allocate zero energy. The suboptimal solution to the design vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ is then of the form

$$\mathbf{x}_1 = \mathbf{v}_{\mathrm{m}} \sqrt{E_0}\, \alpha_1 \tag{2.24}$$

$$\mathbf{x}_2 = \mathbf{v}_{\mathrm{m}} \sqrt{E_0}\, \alpha_2 I\left(\left| \frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|} \frac{\mathbf{n}_1}{\sigma} \right|^2 > \rho \right) = \mathbf{v}_{\mathrm{m}} \sqrt{E_0}\, \alpha_2\, I\left(|\tilde{n}_1|^2 > \rho \right), \tag{2.25}$$

where $\tilde{n}_1$ is defined below (2.20), α_1, α_2 are design parameters independent of $\mathbf{y}_1$ and $I(\cdot)$ is the indicator function, i.e.,

$$I(A) = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false.} \end{cases}$$

The SNR of the suboptimal two-step procedure is

$$\mathrm{SNR}^{(2)} = \mathrm{SNR}_0 \left(\alpha_1^2 + \alpha_2^2 P\left(|\tilde{n}_1|^2 > \rho \right) \right). \tag{2.26}$$

The MSE of the ML estimator under this suboptimal solution using (2.16) is

$$\begin{aligned}\mathrm{MSE}^{(2)} &= \mathrm{E}\left[\frac{|\mathbf{h}_1(\mathbf{x}_1)^H\mathbf{n}_1+\mathbf{h}_1(\mathbf{x}_2)^H\mathbf{n}_2|^2}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2+\|\mathbf{h}_1(\mathbf{x}_2)\|^2)^2}\right] \\ &= \frac{1}{\mathrm{SNR}_0}\mathrm{E}\left[\frac{\alpha_1^2|\tilde{n}_1|^2+\alpha_2^2}{\left(\alpha_1^2+\alpha_2^2\right)^2}I\left(|\tilde{n}_1|^2\geq\rho\right)\right]+\frac{1}{\mathrm{SNR}_0}\mathrm{E}\left[\frac{|\tilde{n}_1|^2}{\alpha_1^2}I\left(|\tilde{n}_1|^2<\rho\right)\right].\end{aligned} \tag{2.27}$$

Denote $\beta=\frac{\alpha_1^2}{\alpha_1^2+\alpha_2^2}$, $0\leq\beta\leq 1$. Substituting for β in the expressions for $\mathrm{MSE}^{(2)}$ and $\mathrm{SNR}^{(2)}$ in (2.27) and (2.26), we obtain

$$\begin{aligned}\mathrm{MSE}^{(2)} &= \frac{1}{\mathrm{SNR}_0}\frac{1}{(\alpha_1^2+\alpha_2^2)}\left(\mathrm{E}\left[(\beta|\tilde{n}_1|^2+(1-\beta))I(|\tilde{n}_1|^2\geq\rho)+\frac{|\tilde{n}_1|^2}{\beta}I(|\tilde{n}_1|^2<\rho)\right]\right), \\ \mathrm{SNR}^{(2)} &= \mathrm{SNR}_0\left(\alpha_1^2+\alpha_2^2\right)\left(\beta+(1-\beta)P(|\tilde{n}_1|^2\geq\rho)\right).\end{aligned}$$

Using the fact that $\mathrm{E}\left[I(|x|^2\geq\rho)\right]=e^{-\rho}$ and $\mathrm{E}\left[|x|^2I(|x|^2\geq\rho)\right]=\rho e^{-\rho}$ when $x\sim\mathcal{CN}(0,1)$, the expressions for $\mathrm{MSE}^{(2)}$ and $\mathrm{SNR}^{(2)}$ simplify to

$$\begin{aligned}\mathrm{MSE}^{(2)} &= \frac{1}{\mathrm{SNR}_0}\frac{1}{(\alpha_1^2+\alpha_2^2)}\left(\beta\rho e^{-\rho}+e^{-\rho}+\frac{1}{\beta}(1-(1+\rho)e^{-\rho})\right), \\ \mathrm{SNR}^{(2)} &= \mathrm{SNR}_0\left(\alpha_1^2+\alpha_2^2\right)\left(\beta+(1-\beta)e^{-\rho}\right).\end{aligned} \tag{2.28}$$

Thus we have

$$\mathrm{MSE}^{(2)}\times\mathrm{SNR}^{(2)}=\left(\beta\rho e^{-\rho}+e^{-\rho}+\frac{1}{\beta}(1-(1+\rho)e^{-\rho})\right)\left(\beta+(1-\beta)e^{-\rho}\right). \tag{2.29}$$

Minimizing $\mathrm{MSE}^{(2)}\times\mathrm{SNR}^{(2)}$ with respect to β and ρ through a grid search for $\beta\in[0,1]$ and $\rho\in[0,\infty)$ yields $\beta^*\approx 0.37$, $\rho^*\approx 0.675$. It follows that $\alpha_1^*\approx 0.7319$, $\alpha_2^*\approx 0.9550$, and substituting for the optimal values of $\alpha_1^*,\alpha_2^*,\beta^*,\rho^*$ in (2.28) and multiplying by SNR_0, yields

$$\mathrm{MSE}^{(2)}_{\min}\times\mathrm{SNR}_0\approx 0.7143. \tag{2.30}$$

This translates to a 28.47% improvement in MSE performance or a 1.5dB savings in terms of SNR. Figure 2.3 plots the reduction in MSE for varying values of ρ at

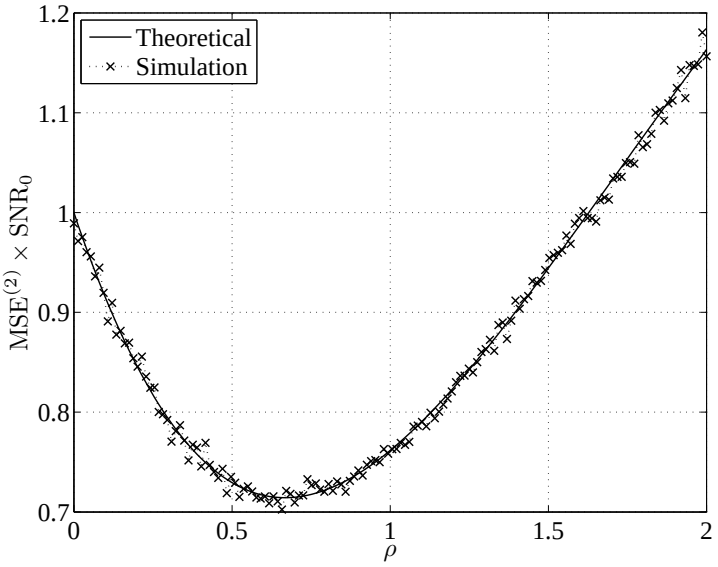


Figure 2.3: Theoretical versus simulation results for suboptimal strategy. Reduction in MSE for varying values of ρ at optimal $\alpha_1^* = 0.7319$.

optimal α_1^* and the reduction in MSE for varying values of α_1 at optimal ρ^* using simulation (dotted) and theoretically (solid) is shown in Fig. 2.4. The theoretical curves in both the figures are obtained by evaluating the expression for $\mathrm{MSE}^{(2)}$ in (2.28) for various values of α_1 and ρ, where α_2 is chosen to satisfy the SNR constraint with equality i.e., $\mathrm{SNR}^{(2)} = \mathrm{SNR}_0$. For the simulation curves, we generate 10000 samples of $\mathbf{y}_1$ in (2.13) using samples of $\mathbf{x}_1$ obtained from (2.24) and $\mathbf{n}_1$, where 10000 samples of $\mathbf{n}_1$ and $\mathbf{n}_2$ are generated from $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. Using samples of $\mathbf{y}_1$, we generate 10000 samples of $\mathbf{x}_2$ from (2.25) and then obtain 10000 samples of $\mathbf{y}_2$ in (2.14) using samples of $\mathbf{x}_2$ and $\mathbf{n}_2$. We then obtain an estimate of $\mathrm{MSE}^{(2)}$ by computing the expected value in (2.16) through numerical averaging over the 10000 realizations. The suboptimal solution to the energy design is shown in Fig. 2.2 by a dashed dotted line indicated as Suboptimal-I. Thus, while the suboptimal strategy limits the peak transmit power to $\max\left(\alpha_1^{*2}, \alpha_2^{*2}\right) E_0$, it is able to achieve near optimal performance.

In the previous two sections, we addressed the problem of minimizing MSE sub-

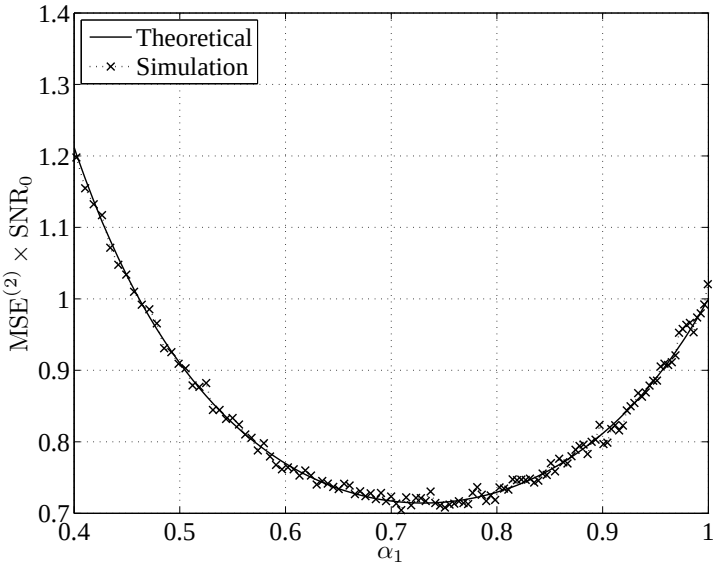


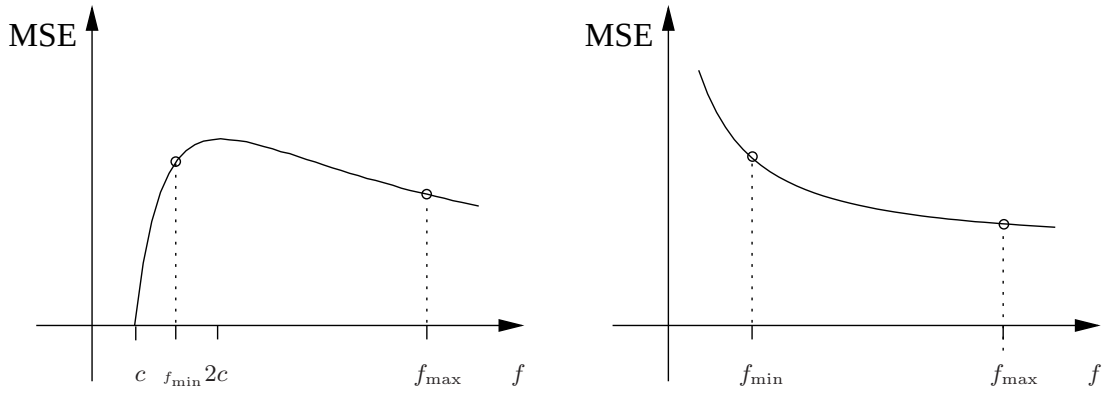
Figure 2.4: Theoretical versus simulation results for suboptimal strategy. Reduction in MSE for varying values of α_1 at optimal $\rho^* = 0.675$.

ject to an average energy constraint, $\mathrm{E}\left[\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2\right] \leq E_0$. An average energy constraint implies that the total allocated energy averaged over repeated trials of the two-step experiment is constrained to be less than or equal to E_0. This is less restrictive than the strict energy constraint $\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 \leq E_0$, as any solution satisfying this constraint satisfies the average energy constraint but not vice versa. The problem of minimizing the MSE in (2.16) under this strict energy constraint is presented in the following section.

2.5 Strict energy constraint solution

The strict energy constraint for a two-step procedure is $\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 \leq E_0$. The MSE for the two-step process given by (2.16) can be rewritten as

$$\begin{aligned} \mathrm{MSE}^{(2)} &= \sigma^2 \left\{ \frac{1}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)} - \frac{\|\mathbf{h}_1(\mathbf{x}_1)\|^2(1-|\tilde{n}_1|^2)}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)^2} \right\} \\ &= \sigma^2\left(\frac{1}{f} - \frac{c}{f^2}\right), \end{aligned} \tag{2.31}$$



(a) MSE(f) for $c > 0$ (b) MSE(f) for $c \leq 0$.

Figure 2.5: Typical plots of the MSE as a function of f.

where $f = (\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)$, $c = \|\mathbf{h}_1(\mathbf{x}_1)\|^2(1 - |\tilde{n}_1|^2)$, and $\tilde{n}_1$ defined below (2.20) is complex Gaussian with zero mean and unit variance. Let $\|\mathbf{x}_1\|^2 = E_1$ and $\|\mathbf{x}_2\|^2 = E_2$ such that $E_1 + E_2 \leq E_0$.

Minimizing the MSE with respect to $\mathbf{x}_2(\mathbf{y}_1)$ is equivalent to minimizing with respect to f. f is a function of $\mathbf{x}_1$ and $\mathbf{x}_2(\mathbf{y}_1)$. Since the squared norm of $\mathbf{x}_1$ and $\mathbf{x}_2$ is limited to E_1 and E_2 respectively, the support of f is restricted to $[f_{\min}, f_{\max}]$ obtained by minimizing and maximizing with respect to $\mathbf{x}_2$ respectively:

$$\begin{aligned} f_{\min} &= E_1' \lambda_m \quad \text{at} \quad \mathbf{x}_2 = \mathbf{v}_{m\perp}\sqrt{E_2} \\ f_{\max} &= E' \lambda_m \quad \text{at} \quad \mathbf{x}_2 = \mathbf{v}_m \sqrt{E_2}, \end{aligned}$$

where $E_1' = \frac{\|\mathbf{h}_1(\mathbf{x}_1)\|^2}{\lambda_m}$, $E' = E_1' + E_2$, and $\mathbf{v}_{m\perp}$ is an unit norm vector in the perpendicular space of $\boldsymbol{H}_1^H \boldsymbol{H}_1$, i.e., $\mathbf{h}_1(\mathbf{v}_{m\perp}) = 0$. When $\boldsymbol{H}_1^H \boldsymbol{H}_1$ is full rank, then there exists no vector in the orthogonal space, in which case an alternate solution, $E_2 = 0$ can be used to achieve $f_{\min}$. Since $E_2 = 0$ satisfies the energy constraint with inequality and consumes minimal energy, we take $E_2 = 0$ as the optimal solution to achieving $f_{\min}$.

Note that from Fig. 2.5, the MSE$^{(2)}(f)$ is either, monotonically increasing from

$f = 0$ to $2c$ and decreasing from $f = 2c$ to ∞ for $c > 0$, or strictly decreasing for $c \leq 0$. Since no local minimum exists, the minimum MSE will always occur at the end points of the support of f. Therefore, to minimize the MSE, we simply need to compare the $\text{MSE}^{(2)}$ values at $f_{\min}$ and $f_{\max}$. The optimal $\mathbf{x}_2$ is

$$\begin{aligned}\mathbf{x}_2(\mathbf{y}_1) &= \arg\max_{\mathbf{x}_2}(f)\ \mathbf{I}\left(\text{MSE}^{(2)}(f_{\max}) \leq \text{MSE}^{(2)}(f_{\min})\right) \\ &\quad + \arg\min_{\mathbf{x}_2}(f)\ \mathbf{I}\left(\text{MSE}^{(2)}(f_{\min}) < \text{MSE}^{(2)}(f_{\max})\right)\end{aligned}$$

Since $\mathbf{x}_2 = \mathbf{v}_m\sqrt{E_2}$ maximizes f, $\mathbf{x}_2 = \mathbf{v}_{m\perp}\sqrt{E_2}$ or $E_2 = 0$ minimizes f, and $\text{MSE}^{(2)}(f_{\max}) \leq \text{MSE}^{(2)}(f_{\min})$ is equivalent to $|\tilde{n}_1|^2 \geq \rho$, we conclude that the optimal $\mathbf{x}_2(\mathbf{y}_1)$ is

$$\mathbf{x}_2(\mathbf{y}_1) = \mathbf{v}_m\sqrt{E_2}\mathbf{I}\left(|\tilde{n}_1|^2 \geq \rho\right) + \mathbf{v}_{m\perp}\sqrt{E_2}\mathbf{I}\left(|\tilde{n}_1|^2 < \rho\right), \tag{2.32}$$

or equivalently,

$$\mathbf{x}_2(\mathbf{y}_1) = \mathbf{v}_m\sqrt{E_2}\mathbf{I}\left(|\tilde{n}_1|^2 \geq \rho\right), \tag{2.33}$$

where $\rho = \frac{E_1'}{2E_1'+E_2}$ and $\mathbf{I}(\cdot)$ is the indicator function. This solution implies that when the actual realization of the noise along $\mathbf{h}_1$ in the first transmission is small enough there is no advantage in using the measurement from the second step. Therefore, we transmit $\mathbf{x}_2 \propto \mathbf{v}_{m\perp}$, which makes the overall estimator only a function of the first measurement, or not transmit at the second step by having $E_2 = 0$. When the actual realization of the noise along $\mathbf{h}_1$ in the first transmission is not small enough, there is some merit in incorporating the information from the second measurement and therefore we select $\mathbf{x}_2 \propto \mathbf{v}_m$.

Substituting for $\mathbf{x}_2(\mathbf{y}_1)$ from (2.33) into (2.31), we obtain

$$\text{MSE}^{(2)} = \frac{\sigma^2}{E'\lambda_{\text{m}}}\left\{\mathbf{I}\left(|\tilde{n}_1|^2 \geq \rho\right)\left(\frac{E_2}{E'} + \frac{E_1'|\tilde{n}_1|^2}{E'}\right) + \mathbf{I}\left(|\tilde{n}_1|^2 < \rho\right)\left(\frac{E'|\tilde{n}_1|^2}{E_1'}\right)\right\}$$

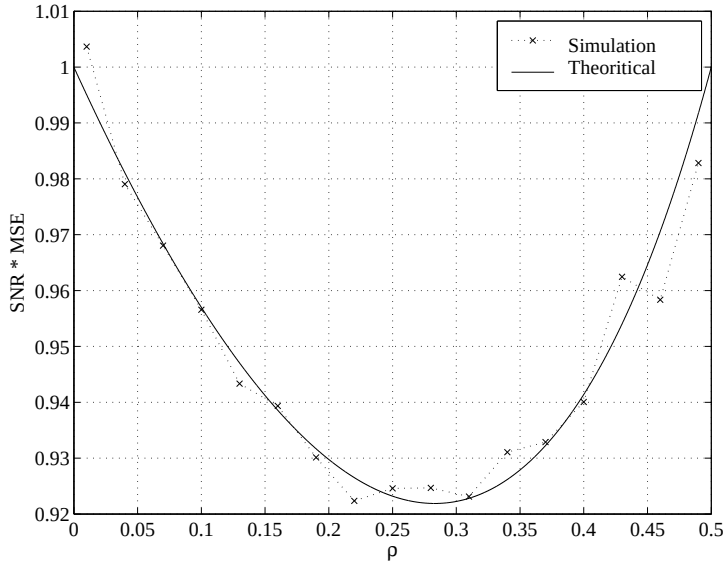


Figure 2.6: $\text{MSE}^{(2)} \times \text{SNR}^{(2)}$ vs. ρ.

where $\tilde{n}_1 \sim \mathcal{CN}(0,1)$. Taking the expectation over $\tilde{n}_1$, the $\text{MSE}^{(2)}$ simplifies to

$$\text{MSE}^{(2)} = \frac{\sigma^2}{E'\lambda_{\text{m}}}\left\{e^{-\rho}\frac{1-2\rho}{1-\rho} + \frac{\rho(1+\rho)}{1-\rho}e^{-\rho} + \frac{1-\rho}{\rho}(1-e^{-\rho}(1+\rho))\right\}. \tag{2.34}$$

We know that $E' = \frac{\|\mathbf{h}_1(\mathbf{x}_1)\|^2}{\lambda_m} + E_2$. $\text{MSE}^{(2)}$ is minimized when E' is maximized which happens when $\mathbf{x}_1 = \mathbf{v}_m\sqrt{E_1}$. The value of ρ that minimizes the $\text{MSE}^{(2)}$ is given by $\rho^* \approx 0.2831$. This implies that the optimal amount of energy allocated at the first stage is $E_1^* \approx 0.395E_0$ and the remaining energy, $E_2^* \approx 0.605E_0$ is used at the second stage. The minimum MSE will be given by,

$$\text{MSE}^{(2)} \approx \frac{1}{\text{SNR}^{(2)}}(0.9283). \tag{2.35}$$

We plot the numerical $\text{MSE}^{(2)}$ and the bias as a function of ρ and the exact $\text{MSE}(\rho)$ in Fig. 2.6 and 2.7 respectively. It is in fact easy to show that the bias of optimal estimator is zero. The simulation curve was generated by using the design of $\mathbf{x}_2$ given in by generating random 100000 samples of $\mathbf{y}_1$ and $\mathbf{y}_2$. The $\text{MSE}^{(2)}$ was then evaluated by numerically evaluating the expected value in (2.16). The theoretical curve was generated by evaluating the expression for $\text{MSE}^{(2)}$ in (2.34). We observe

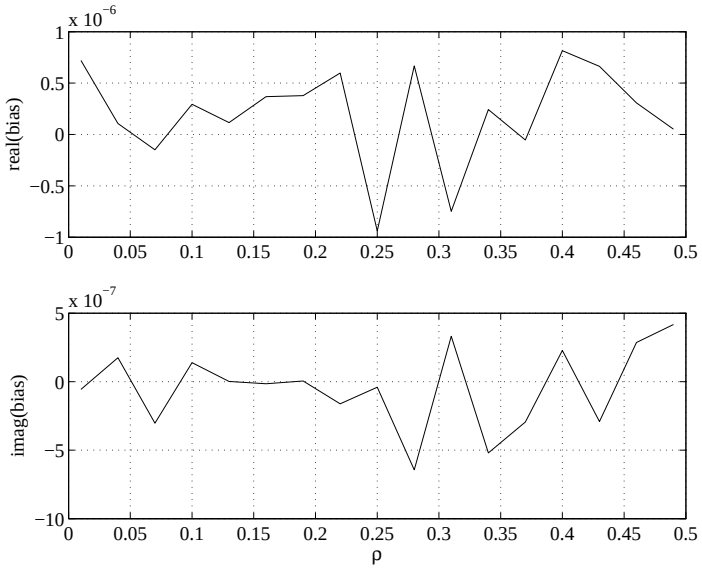


Figure 2.7: Re(Bias) vs. ρ.

that the simulation results for the $\text{MSE}^{(2)}$ and the bias agree with their analytical equivalents.

The optimal solution satisfies the the strict energy constraint with inequality but the average energy used is only $E_0(\alpha_1^{*2} + \alpha_2^{*2} e^{-\rho^*}) \approx 0.8550 E_0$. The solution to the two-step strategy under this strict energy constraint can also be derived by imposing an additional constraint, $\alpha_1^2 + \alpha_2^2 \leq 1$ to the suboptimal design problem described earlier in Section 2.4. In the following section, we design a θ_1-independent design strategy that achieves the optimal performance asymptotically and allows for any peak power constraint in the design.

2.6 Parameter independent two-step design strategy

2.6.1 Problem statement

Consider the optimal design for the two-step procedure

$$\begin{aligned}\mathbf{x}_1 &= \sqrt{E_0}\ \alpha_1^*\mathbf{v}_{\mathrm{m}} \\ \mathbf{x}_2 &= \sqrt{E_0}\ \alpha_2^*(\tilde{n}_1(\mathbf{y}_1;\theta_1))\mathbf{v}_{\mathrm{m}} = \sqrt{E_0}\ \alpha_2^*\left(\left|\frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|}\frac{(\mathbf{y}_1-\sqrt{E_0}\alpha_1\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\theta_1)}{\sigma}\right|\right)\mathbf{v}_{\mathrm{m}}.\end{aligned} \tag{2.36}$$

We showed that by designing α_1 and α_2 optimally we can gain up to 32% improvement in estimator performance. But the "omniscient" solution (2.36) depends on the parameter to be estimated. Here, we prove that we can approach the optimal two-step gain by implementing a θ_1-independent energy allocation strategy when θ_1 is bounded, i.e., $\theta_1 \in [\theta_{\min}, \theta_{\max}]$, $\theta_{\min}, \theta_{\max} \in \mathbb{R}$.

2.6.2 Solution

We describe the intuition behind the proposed solution in this subsection. The details of the proof are given in Appendix 2.12. Since we do not know the value of the actual parameter, we replace θ_1 by a 'guess' of θ_1, say θ_g, in the optimal solution to the design at the second step given in (2.36). The resulting suboptimal design is

$$\begin{aligned}\mathbf{x}_1 &= \sqrt{E_0}\ \alpha_1^*\mathbf{v}_{\mathrm{m}} \\ \mathbf{x}_2 &= \sqrt{E_0}\ \alpha_2^*\left(\left|\frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|}\frac{(\mathbf{y}_1-\sqrt{E_0}\alpha_1^*\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\theta_g)}{\sigma}\right|\right)\mathbf{v}_{\mathrm{m}} = \sqrt{E_0}\ \alpha_2^*\left(|\tilde{n}_1+z|\right)\ \mathbf{v}_{\mathrm{m}},\end{aligned} \tag{2.38}$$

where

$$z = \frac{\alpha_1^*\sqrt{E_0}\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|}{\sigma}(\theta_1-\theta_g) = \alpha_1^*\sqrt{\mathrm{SNR}_0}(\theta_1-\theta_g) \tag{2.39}$$

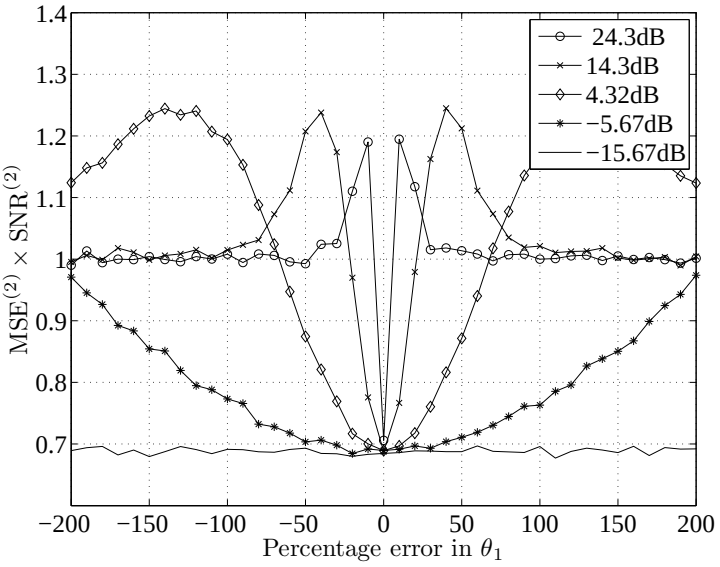


Figure 2.8: Plot of reduction in MSE versus percentage error in guess of parameter of θ_1 for various SNR.

and $\tilde{n}_1$, which is defined below (2.20) is $\mathcal{CN}(0,1)$. Substituting the above suboptimal solution in the expression for $\text{MSE}^{(N)} \times \text{SNR}^{(N)}$ in (2.11) and simplifying, we obtain

(2.40)

$$\text{MSE}^{(2)}(z)\times\text{SNR}^{(2)}(z) = \eta(z) = \text{E}\left[\frac{\alpha_1^{*2}|\tilde{n}_1|^2 + \alpha_2^{*2}(|\tilde{n}_1 + z|)}{(\alpha_1^{*2} + \alpha_2^{*2}(|\tilde{n}_1 + z|))^2}\right] \text{E}\left[\alpha_1^{*2} + \alpha_2^{*2}(|\tilde{n}_1 + z|)\right].$$

The optimal solution to $\text{MSE}^{(2)}(z) \times \text{SNR}^{(2)}(z)$ is achieved when $z = 0$. There are two ways that drive $z \to 0$. If $\theta_1 = \theta_g$, then $z = 0$ and we have $\eta(0) = \eta^* = \text{MSE}^{(2)}_{\text{min}} \times \text{SNR}_0 \approx 0.68$, the optimal two-step performance. Since θ_g is arbitrary, $|\theta_1 - \theta_g| > 0$; the two-step design is not optimal and therefore $\text{MSE}^{(2)} \times \text{SNR}_0 = \eta(z) > \eta^*$. The other way to achieve the optimal solution is to make SNR_0 as small as possible. Note that if SNR_0 is sufficiently small $\text{MSE}^{(2)} \times \text{SNR}^{(2)}$ approaches its minimal value. Since $\text{SNR}^{(2)} \leq \text{SNR}_0$, driving the SNR_0 to zero, drives the $\text{MSE}^{(2)}$ to infinity. To overcome this problem, we propose an $N \times 2$-*step procedure* to allow the SNR_0 to be fixed while driving $z \to 0$. The $N \times 2$-step algorithm is outlined in Fig. 2.9 and is shown through an illustration in Fig. 2.10. Any peak power constraint can also be satisfied using

the $N \times 2$-step strategy by choosing a sufficiently large N.

Figure 2.8 shows $\eta(z)$ in (2.40) as a function of the percentage error in the guess of θ_1, $100\left(\frac{\theta_1-\theta_g}{\theta_1}\right)$ for varying SNR_0. The plot indicates that when $\theta_g = \theta_1$, the optimal performance of the adaptive two-step strategy is achieved for all SNR. At high SNR, for certain values of $|\theta_1 - \theta_g|$, the two-step strategy defined by equations (2.37) and (2.38) performs worse than a single step strategy with signal-to-noise ratio SNR_0. This is because the solution presented in (2.37) and (2.38) in terms of scalar α_1^* and thresholding function $\alpha_2^*(\cdot)$ were optimized for $\tilde{n}_1 + z \sim \mathcal{CN}(0, 1)$, i.e., when $z = 0$. When $\theta_g \neq \theta_1$, the following happens: $z \neq 0$, $\tilde{n}_1 + z \sim \mathcal{CN}(z, 1)$, and the design parameters α_1^* and $\alpha_2^*(\cdot)$, which were found optimally for $\tilde{n}_1 + z \sim \mathcal{CN}(0, 1)$ $(z = 0)$ are no longer optimal. When $|\theta_1 - \theta_g|$ is large, z in (2.39) is a large constant and hence $\tilde{n}_1$ is a negligible term compared to z with high probability. In other words, $\alpha_2^*(\tilde{n}_1 + z)$ can be made arbitrarily close to $\alpha_2^*(z)$ with high probability as z tends to infinity. This implies that the strategy becomes equivalent to a two-step non-adaptive strategy with a specific non-adaptive energy distribution between the two steps whose performance is given by $\text{MSE}^{(2)} \times \text{SNR}_0 = 1$ from Section 2.2.1. Thus we observe that the performance of the two-step strategy tends to 1 for large $|\theta_1 - \theta_g|$. The most important information in the plot, however, is the performance of the two-step strategy under low SNR since each 2-step procedure in the $N \times 2$-step strategy works at $(1/N)^{\text{th}}$ of the total SNR. Hence as N becomes large, SNR in each experiment is very small and the lack of knowledge of θ_1 plays a negligible effect on the performance as z is made close to zero through the SNR factor.

- *Step 1: Perform N independent two-step suboptimal experiments with inputs $\frac{1}{\sqrt{N}}\mathbf{x}_1$ and $\frac{1}{\sqrt{N}}\mathbf{x}_2$ where $\mathbf{x}_1$ and $\mathbf{x}_2$ are given in* (2.37) *and* (2.38) *respectively, i.e., use energy E_0/N in each of the N experiments.*
  - The SNR of the $2N$-step procedure is $\mathrm{SNR}^{(2N)}(z) = N\mathrm{SNR}^{(2),1}(z) = \mathrm{SNR}^{(2)}(z/\sqrt{N})$ where $\mathrm{SNR}^{(2),k}$ is the SNR of the k^{th} two-step experiment. The first equality follows from the fact that $\{\mathrm{SNR}^{(2),k}\}_{k=1}^{N}$ are identical as the N experiments are independent while the second equality follows from the fact that each two-step experiment uses only $(1/N)^{\text{th}}$ of the total energy.
- *Step 2: Obtain ML estimate from each step as $\hat{\theta}_1^{(2),k}$ and average the N estimates to obtain the ML estimator of the $N \times 2$-step strategy as $\hat{\theta}_1^{(2N)} = \frac{1}{N}\sum_{k=1}^{N}\hat{\theta}_1^{(2),k}$.*
  - The MSE of $\hat{\theta}_1^{(2N)}$ is given by $\mathrm{MSE}^{(2N)}(z) = \frac{1}{N}\mathrm{MSE}^{(2),1}(z) = \mathrm{MSE}^{(2)}(z/\sqrt{N})$, where $\mathrm{MSE}^{(2),k}$ is the MSE of each two-step estimator $\hat{\theta}_1^{(2),k}$. The first equality follows from the fact that $\{\mathrm{MSE}^{(2),k}\}_{k=1}^{N}$ are identical as the N experiments are independent while the second equality follows from the fact that each two-step experiment uses only $(1/N)^{\text{th}}$ of the total energy.
- From Steps 1 and 2, we have $\mathrm{MSE}^{(2N)}(z) \times \mathrm{SNR}^{(2N)}(z) = \mathrm{MSE}^{(2)}(z/\sqrt{N})\mathrm{SNR}^{(2)}(z/\sqrt{N})$. As $N \to \infty$, $z/\sqrt{N} \to 0$ and $\mathrm{MSE}^{(2N)}(z) \times \mathrm{SNR}^{(2N)}(z) \to \eta^*$, i.e., minimal MSE is achieved. The details of the proof can be found in Appendix 2.12.

Figure 2.9: Description of the $N\times$ two-step procedure.

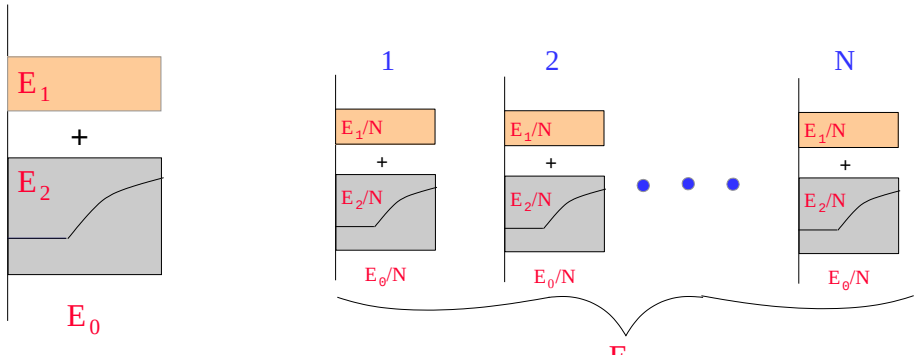


Figure 2.10: Illustration of the $N\times$ two-step procedure: the omniscient optimal two-step procedure, where energy E_1 is allocated to the first step and E_2 is chosen optimally at the second step based on the past measurements, is shown in Fig. (a). Figure (b) illustrates the $N \times 2$-step procedure, where N independent two-step experiments are performed with the energy design as the optimal two-step energy allocation strategy scaled through $1/N$ but with θ_g replacing θ_1. By averaging the estimates of the N two-step estimators, we asymptotically achieve optimal performance as $N \to \infty$.

2.7 Design of N-step procedure

In Sections 2.3 and 2.6, we derived the omniscient optimal two-step design to minimize the MSE and proved that the optimal performance can be achieved asymptotically using an $N \times 2$-step strategy. But the $N \times 2$-step strategy is a specific case of a $2N$-step design. In this section, we generalize the suboptimal solution from the 2-step case to the N-step case as follows: we assume that the shape of the design vector is fixed and look at the energy allocation among the various steps. The set of observations are as defined in (2.3). Let the shape of the design vector $\mathbf{x}_i$ be $\mathbf{v}_{\mathrm{m}}$ and the energy at step i, $E_i = \alpha_i^2(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$, i.e., $\mathbf{x}_i = \mathbf{v}_{\mathrm{m}}\alpha_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}),\ 1 \le i \le N$. Then

$$\alpha_1 = A_1 \tag{2.41}$$

$$\alpha_i = A_i I\left(\frac{|\sum_{j=1}^{i-1} \mathbf{h}_1(\mathbf{x}_j)^H \mathbf{n}_j|^2}{\sum_{j=1}^{i-1} \|\mathbf{h}_1(\mathbf{x}_j)\|^2 \sigma^2} \ge \rho_i\right), \quad i \ge 2, \tag{2.42}$$

where $\{A_i, \rho_i\}$ are design parameters. This approximate solution is motivated from the suboptimal thresholding solution to the two-step case derived in Section 2.4. Note that the definition of the amplitudes at each stage is recursive, i.e., the amplitude design α_i depends on past inputs $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$ which in turn depends on $\alpha_1, \ldots, \alpha_{i-1}$. To simplify our analysis, we make the assumption $\rho_1 \le \rho_2 \le \ldots \le \rho_N$. Then,

(2.43)

$$\alpha_2 = A_2 I\left(\left|\frac{\mathbf{h}_1(\mathbf{x}_1)^H}{\|\mathbf{h}_1(\mathbf{x}_1)\|}\frac{\mathbf{n}_1}{\sigma}\right|^2 \ge \rho_2\right) = A_2 I\left(\left|\frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|}\frac{\mathbf{n}_1}{\sigma}\right|^2 \ge \rho_2\right) = A_2 I\left(|\tilde{n}_1|^2 \ge \rho_2\right),$$

where $\tilde{n}_i = \frac{\mathbf{h}_1(\mathbf{v}_\mathrm{m})^H}{\|\mathbf{h}_1(\mathbf{v}_\mathrm{m})\|}\frac{\mathbf{n}_i}{\sigma}$ are i.i.d complex Gaussian random variables with zero mean and unit variance. The amplitude at the third stage simplifies to

$$\begin{aligned}
\alpha_3 &= A_3\left(\frac{|\mathbf{h}_1(\mathbf{x}_1)^H\mathbf{n}_1 + \mathbf{h}_1(\mathbf{x}_2)^H\mathbf{n}_2|^2}{\sigma^2\left(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2\right)} \geq \rho_3\right) \\
&= A_3 I\left(\frac{|A_1\tilde{n}_1 + A_2\tilde{n}_2|^2}{|A_1|^2+|A_2|^2} \geq \rho_3\right) I\left(|\tilde{n}_1|^2 \geq \rho_2\right) + A_3 I\left(|\tilde{n}_1|^2 \geq \rho_2\right) I\left(|\tilde{n}_1|^2 < \rho_2\right) \\
&= A_3 I\left(\frac{|A_1\tilde{n}_1 + A_2\tilde{n}_2|^2}{|A_1|^2+|A_2|^2} \geq \rho_3\right) I\left(|\tilde{n}_1|^2 \geq \rho_2\right). \qquad (2.44)
\end{aligned}$$

Following the same procedure, we simplify α_4 as

(2.45)

$$\alpha_4 = A_4 I\left(\frac{|A_1\tilde{n}_1 + A_2\tilde{n}_2 + A_3\tilde{n}_3|^2}{|A_1|^2+|A_2|^2+|A_3|^2} \geq \rho_4\right) I\left(\frac{|A_1\tilde{n}_1 + A_2\tilde{n}_2|^2}{|A_1|^2+|A_2|^2} \geq \rho_3\right) I\left(|\tilde{n}_1|^2 \geq \rho_2\right).$$

Thus, a general expression for α_i can be written as

$$\alpha_i = A_i \prod_{s=1}^{i-1} I\left(|w_s|^2 \geq \rho_{s+1}\right), \qquad (2.46)$$

where w_s is defined in (2.134). This form states that the stopping criteria at time step s is when the magnitude of the average noise, w_s drops below the threshold ρ_{s+1}. The goal is to minimize $\mathcal{G}_N = \mathrm{MSE}^{(N)} \times \mathrm{SNR}^{(N)}$ which from Appendix 2.13 is given by

$$\mathrm{MSE}^{(N)} \times \mathrm{SNR}^{(N)}(\mathbf{A}, \boldsymbol{\rho}) = \left(\sum_{i=1}^{N-1}\frac{T_i}{Q_i} + \frac{\tilde{T}_N}{Q_N}\right)\left(\sum_{i=1}^{N-1} Q_i P_i + Q_N\tilde{P}_N\right), \qquad (2.47)$$

where $\mathbf{A} = [A_1, \ldots, A_N], \boldsymbol{\rho} = [\rho_1, \ldots, \rho_N]$, Q_i, T_i, P_i are defined in (2.136), (2.137) and (2.139) respectively.

There is no closed-form solution to this $2N$ dimensional optimization. Instead we evaluate the performance of suboptimal solutions to the design vectors $\mathbf{A}$ and $\boldsymbol{\rho}$. For our simulations, we choose $\rho_i = (i-1)/(N-1)\ \rho_{\max},\ 1 \leq i \leq N$. Furthermore, we choose $\mathbf{A}$ as $\{A_i = d\ \alpha_1^*,\ \text{odd } i;\ A_i = d\ \alpha_2^*,\ \text{even } i\}$, where α_1^*, α_2^* are optimal values from the suboptimal solution presented in Section 2.4 and d is chosen to satisfy the

average energy constraint. We evaluate the performance of the N-step procedure with these parameters through theory and verify the theory using simulations.

Performance gains, $\mathcal{G}_N$ (in dB) are presented in Fig. 2.12. The theoretical performance curve was generated by evaluating the expressions for T_i and P_i in (2.137) and (2.139) using numerical integration and substituting them in (2.47) while the simulations are generated by sampling the distributions of $\{\mathbf{y}_i\}_{i=1}^N$ by first generating samples of $\{\alpha_i\}_{i=1}^N$ followed by an empirical estimate of the MSE. By designing this N-step procedure, we are essentially altering the Gaussian statistics of the measurement noise to obtain improvements in performance. In Fig. 2.11, we illustrate how the distribution of the estimation residuals changes with the number of the steps. We would like to point out that the simulation curve appears smoother than the theory curve in Fig. 2.12 as evaluating (2.137) and (2.139) involved high order integration. We see that in 50 steps, we are able to achieve gains of more than 5dB. In Section 2.6, we showed that the two-step gain can be achieved using an $N \times 2$-step strategy, i.e., in $2N$ steps. The basic motivating factor was to reduce the SNR in each experiment and achieve the diversity gain by increasing the number of steps. For the general N-step strategy, progressive reduction in SNR of each experiment implies that as the number of steps increases, the error of guessing θ_1 has a reduced effect on the overall performance. We demonstrate the achievability of performance for any N-step design in the following subsection.

Achievability of performance of any omniscient N-step design

For an N-step procedure, we need to design a sequence of input vectors $\{\mathbf{x}_i\}_{i=1}^N$ optimally under an average energy constraint to minimize the MSE in (2.5).

Theorem 2.7.1. *Let* $\mathcal{S} = \{\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}; \theta_1)\}_{i=1}^N$ *be any design of the input pa-*

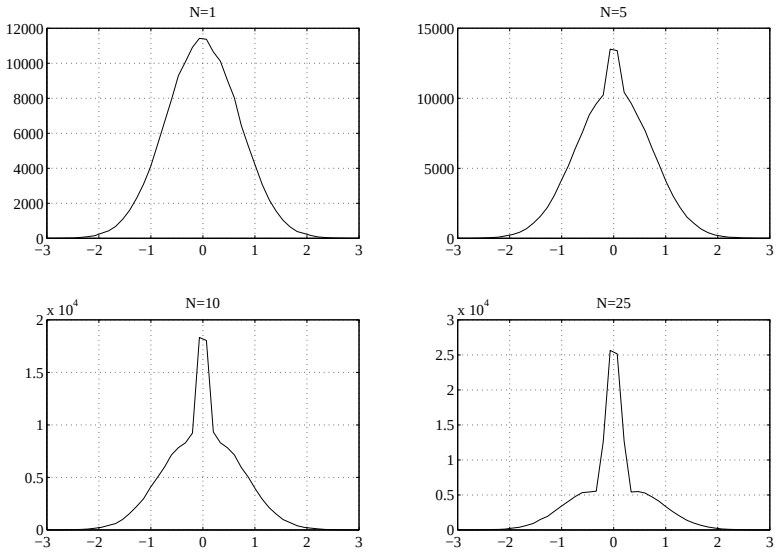


Figure 2.11: Distribution of residual noise versus number of steps.

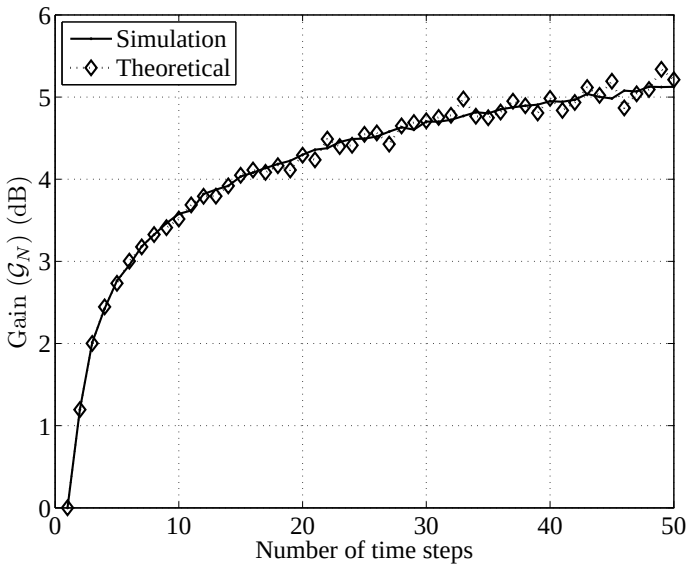


Figure 2.12: Plot of gains obtained through suboptimal N-step procedure as a function of N through theory and simulations.

rameters satisfying the following conditions:

- *Average energy constraint* - $\mathrm{E}\left[\sum_{i=1}^{N} \|\mathbf{x}_i\left(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1};\theta_1\right)\|^2\right] \leq E_0$.
- *Continuity - The design vector* $\mathbf{x}_i\left(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1};\theta_1\right)$ *is a continuous function of* $\{\mathbf{y}_j\}_{j=1}^{i-1}$ *or can assume the form of a thresholding function in (2.42).*

Then there exists a θ_1*-independent strategy whose performance can come arbitrarily close to* $MSE^{(N)}(\mathcal{S})$ *which assumes the knowledge of parameter* θ_1.

Proof. The proof is similar to the $N \times 2$-step strategy presented in Section 2.6, where the actual value of θ_1 in the optimal solution is replaced with a guess of θ_1. Refer to Appendix 2.16 for details. □

2.8 Sequential design for vector parameters

A general N-step procedure for the case of M unknown parameters can be written as

$$\mathbf{y}_i = \mathbf{H}(\mathbf{x}_i(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1}))\boldsymbol{\theta} + \mathbf{n}_i, \quad i = 1,2,\ldots,N, \tag{2.48}$$

where $\boldsymbol{\theta}$ is an M-element vector, $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0},\mathbf{R}_\mathbf{n})$, and $\mathbf{H}(\mathbf{x})$ is a $K \times M$ matrix. For the multiple parameter case, MSE is no longer a scalar. Various criteria such as trace, minmax, determinant of the MSE matrix can be considered as measures of performance under the multiple unknown setting.

2.8.1 Worst case error criterion

The component wise MSE for estimating specific parameters is given by the diagonal elements of the matrix $\mathrm{MSE} = \mathrm{E}\left[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^H\right]$. We seek to find the optimal energy allocation between the two design vectors, $\mathbf{x}_i(\{\mathbf{y}_j\}_{j=1}^{i-1}) = \mathbf{u}_\mathrm{m}\sqrt{E_0}\alpha_i(\{\mathbf{y}_j\}_{j=1}^{i-1})$,

$i = 1, 2$, that minimizes the worst case mean-squared error (*WC*-MSE) of the unknown parameters, where $\mathbf{u}_{\rm m}$ is any unit norm vector independent of past measurements, e.g., $\mathbf{u}_{\rm m}$ is chosen to minimize the one-step MSE. The ML estimate for a one-step process with energy E_0 is given by

$$\widehat{\boldsymbol{\theta}}^{(1)} = \frac{1}{\sqrt{E_0}} \mathbf{W}_{\mathbf{u}_{\rm m}} \mathbf{H}(\mathbf{u}_{\rm m})^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{y}_1 \tag{2.49}$$

and its corresponding MSE is

$$\text{MSE}^{(1)} = \frac{1}{E_0} \mathbf{W}_{\mathbf{u}_{\rm m}}, \tag{2.50}$$

where $\mathbf{W}_{\mathbf{u}_{\rm m}} = (\mathbf{H}(\mathbf{u}_{\rm m})^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{H}(\mathbf{u}_{\rm m}))^{-1}$. Define $\Phi(\mathbf{u}, \text{MSE}) = \mathbf{u}^H \text{MSE} \mathbf{u}$.

$$WC\text{-MSE} = \max_i \mathbf{e}_i^H \text{MSE} \mathbf{e}_i = \max_i \Phi(\mathbf{e}_i, \text{MSE}), \tag{2.51}$$

where $\mathbf{e}_i$ is an M-element vector with all zeros except for 1 in the i^{th} position. Then for a one-step process

$$WC\text{-MSE}^{(1)} = \max_i \Phi(\mathbf{e}_i, \text{MSE}^{(1)}) = \Phi(\mathbf{e}_{i^*}, \text{MSE}^{(1)}), \tag{2.52}$$

where i^* indicates the $\arg\max_i \Phi(\mathbf{e}_i, \text{MSE}^{(1)})$ and

$$\Phi(\mathbf{u}, \text{MSE}^{(1)}) = \frac{1}{E_0} \mathbf{u}^H \mathbf{W}_{\mathbf{u}_{\rm m}} \mathbf{u}. \tag{2.53}$$

The set of observations for the two-step process are

$$\mathbf{y}_1 = \sqrt{E_0} \alpha_1 \mathbf{H}(\mathbf{u}_{\rm m}) \boldsymbol{\theta} + \mathbf{n}_1 \tag{2.54}$$

$$\mathbf{y}_2 = \sqrt{E_0} \alpha_2(\mathbf{y}_1) \mathbf{H}(\mathbf{u}_{\rm m}) \boldsymbol{\theta} + \mathbf{n}_2. \tag{2.55}$$

For a two-step procedure, we need to design α_1 and $\alpha_2(\mathbf{y}_1)$ to minimize $WC\text{-MSE}^{(2)}$. From (2.141) in Appendix 2.14, we have

$$\Phi(\mathbf{u}, \text{MSE}^{(2)}) = \Phi(\mathbf{u}, \text{MSE}^{(1)}) \, \text{E} \left[\frac{\alpha_1^2 \left| \tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}) \right|^2 + \alpha_2^2(\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}))}{(\alpha_1^2 + \alpha_2^2(\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}))^2} \right], \tag{2.56}$$

where $\tilde{n}_1(\mathbf{y}_1;\boldsymbol{\theta}) = \frac{\mathbf{u}^H\mathbf{W}_{\mathbf{u}_\mathrm{m}}\mathbf{H}(\mathbf{u}_\mathrm{m})^H\mathbf{R}_\mathbf{n}^{-1}(\mathbf{y}_1-\mathbf{H}(\mathbf{x}_1)\boldsymbol{\theta})}{\sqrt{\mathbf{u}^H\mathbf{W}_{\mathbf{u}_\mathrm{m}}\mathbf{u}}}$ is a complex normal random variable with zero mean and unit variance. The error in (2.56) when minimized under the constraint $\alpha_1^2 + \mathrm{E}\left[\alpha_2^2(\tilde{n}_1)\right] \leq 1$ is exactly the same minimization derived for the single parameter case in Section 2.3. It follows that the optimal and suboptimal solutions to α_1 and $\alpha_2(\cdot)$ will hold for the multiple parameter case. In other words $\Phi(\mathbf{u}, \mathrm{MSE}^{(2)}) \approx 0.6821\ \Phi(\mathbf{u}, \mathrm{MSE}^{(1)})$. It follows that

$$(2.57)\quad \textit{WC}\text{-MSE}^{(2)} = \Phi(\mathbf{e}_{i^*}, \mathrm{MSE}^{(2)}) \approx 0.6821\ \Phi(\mathbf{e}_{i^*}, \mathrm{MSE}^{(1)}) = 0.6821\ \textit{WC}\text{-MSE}^{(1)}$$

and this performance can be achieved using a $\boldsymbol{\theta}$-independent strategy along similar lines to the derivation for the scalar parameter case in Section 2.6. The reduction in MSE in (2.57) holds for any M, the number of unknown parameters, as i^*, the index of the worst case error, can always be computed from (2.52) and (2.53) for any $M \in \mathbb{N}$. A similar result can be derived for the N-step procedure.

2.8.2 Trace criterion

For the multiple parameter case, the MSE is a matrix and we consider the trace as a measure of performance i.e., $\min_{\{\mathbf{x}_i\}_{i=1}^N} \mathrm{tr}(\mathrm{MSE}^{(N)}(\boldsymbol{\theta}))$, where $\mathrm{tr}(\cdot)$ denotes the trace. So far, we considered the problem of optimal and suboptimal strategies for energy allocation in an N-time step procedure. We assumed in our analysis that the waveform transmitted is the one-step optimal derived in Section 2.2.1. For the purposes of the trace criterion in the vector parameter case, we consider two possible strategies: first, performing N-step energy allocation under the constraint that the waveform transmit at every time instant is the one-step optimal for estimating the vector parameters. The second strategy is to provide a waveform and energy allocation simultaneously. We present the energy allocation procedure in the following section.

Energy design under fixed waveforms

The trace of the MSE matrix can be written in the following form,

$$\text{tr}(\text{MSE}) = \sum_{i=1}^{M} \mathbf{e}_i^H \text{MSE} \mathbf{e}_i = \sum_{i=1}^{M} \Phi(\mathbf{e}_i, \text{MSE}), \tag{2.58}$$

In this analysis, we will assume that all the elements of the matrix $\mathbf{H}$ defined in (2.48) can be controlled by the design vector $\mathbf{x}_i$ at every time instant. In other words, we assume the following statistical model

$$\mathbf{y}_i = \alpha_i \sqrt{E_0} \widetilde{\mathbf{X}}_i \boldsymbol{\theta} + \mathbf{n}_i, i = 1, 2, \ldots, N, \tag{2.59}$$

where $\widetilde{\mathbf{X}}_i$ is the $K \times M$ input design matrix and $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{R_n})$. The one-step estimator for $\boldsymbol{\theta}$ in this model is given by

$$\widehat{\boldsymbol{\theta}}^{(1)} = \frac{1}{\sqrt{E_0}} (\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1} \widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \mathbf{y}_1 \tag{2.60}$$

and the corresponding MSE can be derived using a similar derivation to (2.50) as

$$\text{MSE}^{(1)} = \frac{1}{E_0} (\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1}. \tag{2.61}$$

The trace of the $\text{MSE}^{(1)}$ matrix can then be written as

$$\text{tr}(\text{MSE}^{(1)}) = \frac{1}{E_0} \text{tr} \left\{ (\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1} \right\} = \frac{1}{E_0} \sum_{i=1}^{M} \mathbf{e}_i^H (\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1} \mathbf{e}_i. \tag{2.62}$$

For a two-step process, the set of observations can be written as

$$\mathbf{y}_1 = \alpha_1 \sqrt{E_0} \widetilde{\mathbf{X}}_1 \boldsymbol{\theta} + \mathbf{n}_1 \tag{2.63}$$

$$\mathbf{y}_2 = \alpha_2(\mathbf{y}_1) \sqrt{E_0} \widetilde{\mathbf{X}}_1 \boldsymbol{\theta} + \mathbf{n}_2, \tag{2.64}$$

where $\widetilde{\mathbf{X}}_1$ will be chosen as the optimal one-step design and α_1 and $\alpha_2(\mathbf{y}_1)$ are optimal energy allocation design parameters. The two-step ML estimator is given by

$$\widehat{\boldsymbol{\theta}}^{(2)} = \frac{1}{\sqrt{E_0}} (\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1} \widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \left(\frac{\alpha_1 \mathbf{y}_1 + \alpha_2(\mathbf{y}_1) \mathbf{y}_2}{\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)} \right) \tag{2.65}$$

and the corresponding MSE is

(2.66)
$$\mathrm{MSE}^{(2)} = \frac{1}{E_0}\mathrm{E}_{\mathbf{n}_1}\left[(\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}}_1)^{-1}\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1}\left(\frac{\alpha_1^2\mathbf{n}_1\mathbf{n}_1^H+\alpha_2^2(\mathbf{y}_1)\mathbf{R}_{\mathbf{n}}}{\left(\alpha_1^2+\alpha_2^2(\mathbf{y}_1)\right)^2}\right)\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}}_1(\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}}_1)^{-1}\right].$$

The trace of the $\mathrm{MSE}^{(2)}$ can be evaluated as

$$\begin{aligned}\mathrm{tr}(\mathrm{MSE}^{(2)}) &= \frac{1}{E_0}\left(\mathrm{E}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2(\mathbf{y}_1))^2}\mathbf{n}_1^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}}_1(\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}}_1)^{-2}\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{n}_1\right.\right.\\ &\qquad \left.\left.+\frac{\alpha_2^2}{(\alpha_1^2+\alpha_2^2(\mathbf{y}_1))^2}\mathrm{tr}\left\{(\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}}_1)^{-1}\right\}\right]\right) \qquad (2.67)\\ &= \frac{1}{E_0}\left(\mathrm{E}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2(\mathbf{y}_1))^2}\widetilde{\mathbf{n}}_1^H(\mathbf{y}_1;\boldsymbol{\theta})\mathbf{M}\widetilde{\mathbf{n}}_1(\mathbf{y}_1;\boldsymbol{\theta})\right.\right.\\ &\qquad \left.\left.+\frac{\alpha_2^2}{(\alpha_1^2+\alpha_2^2(\mathbf{y}_1))^2}\mathrm{tr}\left\{(\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}}_1)^{-1}\right\}\right]\right), \qquad (2.68)\end{aligned}$$

where

$$\mathbf{M} = \mathbf{R}_{\mathbf{n}}^{-1/2}\widetilde{\mathbf{X}}_1(\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}}_1)^{-2}\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1/2} \qquad (2.69)$$

and

$$\widetilde{\mathbf{n}}_1(\mathbf{y}_1;\boldsymbol{\theta}) = \mathbf{R}_{\mathbf{n}}^{-1/2}\mathbf{n}_1 = \mathbf{R}_{\mathbf{n}}^{-1/2}(\mathbf{y}_1-\alpha_1\sqrt{E_0}\widetilde{\mathbf{X}}_1\boldsymbol{\theta}). \qquad (2.70)$$

Since rotation or translation of the Gaussian random vector is still a Gaussian random vector, it follows that $\widetilde{\mathbf{n}}_1$ is i.i.d $\mathcal{CN}(\mathbf{0},\mathbf{I})$. Using circular invariance of trace of a matrix, i.e., $\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CAB})$, we have

$$\mathrm{tr}(\mathbf{M}) = \mathrm{tr}\left\{\mathbf{R}_{\mathbf{n}}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}})^{-2}\widetilde{\mathbf{X}}^H\mathbf{R}_{\mathbf{n}}^{-1/2}\right\} = \mathrm{tr}\left\{(\widetilde{\mathbf{X}}^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}})^{-1}\right\} \qquad (2.71)$$

Define $w_1(\mathbf{y}_1;\boldsymbol{\theta})$ as

$$w_1(\mathbf{y}_1;\boldsymbol{\theta}) = \frac{\widetilde{\mathbf{n}}_1^H(\mathbf{y}_1;\boldsymbol{\theta})\mathbf{M}\widetilde{\mathbf{n}}_1(\mathbf{y}_1;\boldsymbol{\theta})}{\mathrm{tr}(\mathbf{M})}. \qquad (2.72)$$

Substituting (2.72) and (2.71) in (2.68), we obtain

(2.73)
$$\mathrm{tr}(\mathrm{MSE}^{(2)}) = \mathrm{tr}(\mathrm{MSE}^{(1)})\left(\mathrm{E}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2(w_1(\mathbf{y}_1;\boldsymbol{\theta})))^2}w_1(\mathbf{y}_1;\boldsymbol{\theta})+\frac{\alpha_2^2(w_1)}{(\alpha_1^2+\alpha_2^2(w_1(\mathbf{y}_1;\boldsymbol{\theta})))^2}\right]\right),$$

where $\alpha_2(\mathbf{y}_1)$ is replaced with $\alpha_2(w_1(\mathbf{y}_1;\boldsymbol{\theta}))$ as the dependence of α_2 on $\mathbf{y}_1$ occurs through $w_1(\mathbf{y}_1;\boldsymbol{\theta})$. We will denote $w_1(\mathbf{y}_1;\boldsymbol{\theta})$ and $\widetilde{\mathbf{n}}_1^H(\mathbf{y}_1;\boldsymbol{\theta})$ as w_1 and $\widetilde{\mathbf{n}}_1$ respectively in the remainder of this discussion for convenience.

Let $\mathbf{M} = \mathbf{Q}\mathbf{D}\mathbf{Q}^H$ denote the eigenvalue decomposition of the matrix $\mathbf{M}$, where $\mathbf{D} = [d_1, \ldots, d_N]$ denotes the eigenvalues. Then

$$w_1 = \frac{\widetilde{\mathbf{n}}_1^H \mathbf{M} \widetilde{\mathbf{n}}_1}{\mathrm{tr}(\mathbf{M})} = \frac{\widehat{\mathbf{n}}_1^H \mathbf{D} \widehat{\mathbf{n}}_1}{\mathrm{tr}(\mathbf{D})} = \frac{\sum_{i=1}^N d_i |\widehat{\mathbf{n}}_{1,i}|^2}{\sum_{i=1}^N d_i} = \sum_{i=1}^N \left(\frac{d_i}{\sum_{i=1}^N d_i} \right) |\widehat{\mathbf{n}}_{1,i}|^2, \tag{2.74}$$

where $\widehat{\mathbf{n}}_1 = \mathbf{Q}^H \widetilde{\mathbf{n}}_1$ is again $\mathcal{CN}(0, \mathbf{I})$ since $\mathbf{Q}$ is unitary and $\widehat{\mathbf{n}}_{1,i}$ denotes the ith element of the vector $\widehat{\mathbf{n}}_1$. $\{|\widehat{\mathbf{n}}_{1,i}|^2\}_{i=1}^N$ are independent central χ_2^2 random variables, i.e., chi-square random variables with 2 degrees of freedom (exponentially distributed). Hence w_1 is central chi-square mixture with N degrees of freedom with

$$\mathrm{E}[w_1] = 1 \tag{2.75}$$

$$\mathrm{var}(w_1) = 2 \sum_{i=1}^N \left(\frac{d_i}{\sum_{i=1}^N d_i} \right)^2 \tag{2.76}$$

To find the optimal solution to the energy at the second stage, i.e., $\alpha_2(w_1)$ we need to minimize with respect to $\alpha_2(w_1)$ the function

$$\left(\mathrm{E}\left[\frac{\alpha_1^2}{(\alpha_1^2 + \alpha_2^2(w_1))^2} w_1 + \frac{\alpha_2^2(w_1)}{(\alpha_1^2 + \alpha_2^2(w_1))^2} \right] \right) \tag{2.77}$$

subject to the constraint on the average energy $\mathrm{E}[\alpha_1^2 + \alpha_2^2(w_1)] \leq 1$. The Lagrangian, the minimization condition plus the constraint, can be written as

$$\min_{\alpha_1, \alpha_2(w_1)} \mathrm{E}_{w_1} \left[\frac{\alpha_1^2 w_1 + \alpha_2^2(w_1)}{(\alpha_1^2 + \alpha_2^2(w_1))^2} \right] + \lambda \left(\alpha_1^2 + \mathrm{E}_{w_1} \left[\alpha_2^2(w_1) \right] \right) \tag{2.78}$$

$$= \min_{\alpha_1, \alpha_2(w_1)} \frac{1}{\alpha_1^2} \mathrm{E}_{\mathbf{n}_1} \left[\frac{w_1 + \alpha_2^2(w_1)/\alpha_1^2}{(1 + \alpha_2^2(w_1)/\alpha_1^2)^2} \right] + \lambda \alpha_1^2 \left(1 + \mathrm{E}_{w_1} \left[\alpha_2^2(w_1)/\alpha_1^2 \right] \right) \tag{2.79}$$

$$= \min_{\alpha_1, g(w_1)} \frac{1}{\alpha_1^2} \mathrm{E}_{w_1} \left[\frac{w_1 + g(w_1) - 1}{g^2(w_1)} \right] + \lambda \alpha_1^2 \mathrm{E}_{w_1} [g(w_1)] \tag{2.80}$$

$$= \min_{\alpha_1, g(w_1)} \frac{1}{\alpha_1^2} \mathrm{E}_{w_1} \left[\frac{1}{g(w_1)} - \frac{1 - w_1}{g^2(w_1)} + \lambda^{'} g(w_1) \right], \tag{2.81}$$

where $g(w_1) = 1+\alpha_2^2(w_1)/\alpha_1^2$ and $\lambda' = \lambda\alpha_1^2$. Differentiating and setting the derivative with respect to g to zero,

$$g^3 - \frac{1}{\lambda'}g + 2\frac{1-w_1}{\lambda'} = 0 \tag{2.82}$$

The structure of the optimal design for the multiple parameter case is the same as that of the optimal design for the single parameter case. The only difference between the solutions is that the optimal solution $\alpha_2(w_1)$ is a function of w_1, a χ^2-mixture with N degrees of freedom which reduces to being a exponentially distributed random variable when the number of parameters to be determined is reduced to 1. Furthermore to compute the exact reduction in the MSE, we need to know the general distribution of the χ^2-mixture. Various series expansions have been proposed in the literature for the distribution of a sum of χ^2-random variables: power series [145], χ^2 series [136], improved power series and Laguerre series expansions [84], Laguerre series for non central chi-square sum [101].

Before we proceed to find the optimal sequential design of energy for an N-step process, we solve for the optimal $\widetilde{\mathbf{X}}_1$ for an one-step process which yields the eigenvalues $d_1, d_2, \ldots, d_N$. To find the optimal one-step design vector, we need the following results.

Lemma 2.8.1. $N \geq M$

Proof. Consider the $M \times M$ matrix $(\widetilde{\mathbf{X}}^H\mathbf{R}_\mathbf{n}^{-1}\widetilde{\mathbf{X}})$ in the solution to the maximum likelihood estimator of $\boldsymbol{\theta}$ in (2.49).

$$(\widetilde{\mathbf{X}}^H\mathbf{R}_\mathbf{n}^{-1}\widetilde{\mathbf{X}}) = \left\{(\widetilde{\mathbf{X}}^H\mathbf{R}_\mathbf{n}^{-1/2})(\widetilde{\mathbf{X}}^H\mathbf{R}_\mathbf{n}^{-1/2})^H\right\}$$

Since for any matrix $\mathbf{A}$, $Rank(\mathbf{A}\mathbf{A}^H) = Rank(\mathbf{A})$, it follows that

$$Rank(\widetilde{\mathbf{X}}^H\mathbf{R}_\mathbf{n}^{-1}\widetilde{\mathbf{X}}) = Rank(\widetilde{\mathbf{X}}^H\mathbf{R}_\mathbf{n}^{-1/2}) = Rank(\widetilde{\mathbf{X}}) = \min(N, M). \tag{2.83}$$

Since we want the $M \times M$ matrix $(\widetilde{\mathbf{X}}^H \mathbf{R}_{\mathbf{n}}^{-1} \widetilde{\mathbf{X}})$ to be invertible, we need $Rank(\widetilde{\mathbf{X}}^H \mathbf{R}_{\mathbf{n}}^{-1} \widetilde{\mathbf{X}}) = M$. It follows that $N \geq M$ for any M. □

For the rest of the derivation, we consider the following assumptions:

- Without loss of generality we assume $N = M$ and furthermore $\widetilde{\mathbf{X}}$ is full rank.
- The receiver noises are independent and identically distributed.i.e., $\mathbf{R}_{\mathbf{n}} = \sigma^2 \mathbf{I}$.

Lemma 2.8.2. *For any positive $m \times m$ definite matrix $\mathbf{A}$, the following inequality holds,*

$$tr(\mathbf{A}^{-1}) \geq \sum_{i=1}^{M} (a_{i,i})^{-1}, \tag{2.84}$$

where $a_{i,i}$ is the ith diagonal element of A and equality iff $\mathbf{A}$ is diagonal.

Proof. The details of the proof can be found in [75]. □

Theorem 2.8.3. *$d_i = \sigma^2 N$, $i = 1, 2, \ldots, N$ minimizes $MSE^{(1)}$.*

Proof. The single-stage MSE given in (2.62) has the form

$$\mathrm{tr}(\mathrm{MSE}^{(1)}) = \frac{1}{E_0} (\widetilde{\mathbf{X}}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \widetilde{\mathbf{X}}_1)^{-1} = \frac{\sigma^2}{E_0} \mathrm{tr} \left\{ (\widetilde{\mathbf{X}}_1^H \widetilde{\mathbf{X}}_1)^{-1} \right\}, \tag{2.85}$$

when $\mathbf{R}_{\mathbf{n}} = \sigma^2 \mathbf{I}$. To obtain the minimum error we need to find $\widetilde{\mathbf{X}}_1$ optimally. Since the energy component was already extracted as a multiplying term of the form $\sqrt{E_0}\alpha_j(\mathbf{y}_1, \ldots, \mathbf{y}_{j-1})$, $\widetilde{\mathbf{X}}_1$ should satisfy $\|\widetilde{\mathbf{X}}_1\|_F = 1$ ($\|\cdot\|_F$ denotes the Frobenius norm). Using Lemma 2.8.2, it follows that the optimal $\widetilde{\mathbf{X}}_1$ satisfies

$$\widetilde{\mathbf{X}}_1^H \widetilde{\mathbf{X}}_1 = \mathrm{diag}(\|\mathbf{z}_1\|^2, \ldots, \|\mathbf{z}_N\|^2), \tag{2.86}$$

where $\widetilde{\mathbf{X}}_1 = [\mathbf{z}_1, \ldots, \mathbf{z}_N]$. We use Lagrange multipliers to solve the following optimization problem. The Lagrangian is given by

$$L(\widetilde{\mathbf{X}}_1, \lambda) = \min_{\widetilde{\mathbf{X}}_1} \operatorname{tr}\left\{\left(\widetilde{\mathbf{X}}_1^H \widetilde{\mathbf{X}}_1\right)^{-1}\right\} + \lambda(\operatorname{tr}(\widetilde{\mathbf{X}}_1^H \widetilde{\mathbf{X}}_1) - 1) \tag{2.87}$$

$$= \sum_{i=1}^{N} \left(\|\mathbf{z}_i\|^2\right)^{-1} + \lambda\left(\sum_{i=1}^{N} \|\mathbf{z}_i\|^2 - 1\right). \tag{2.88}$$

Setting $\frac{\partial L(\widetilde{\mathbf{X}}_1, \lambda)}{\partial \mathbf{z}_j} = 0$ for $j = 1, \ldots, N$, we obtain

$$\left(\|\mathbf{z}_j\|^2\right)^{-2} \mathbf{z}_j - \lambda \mathbf{z}_j = 0 \tag{2.89}$$

$$\left(1 - \lambda \|\mathbf{z}_j\|^4\right) \mathbf{z}_j = 0. \tag{2.90}$$

It follows that the solution to $\{\|\mathbf{z}_i\|^2\}_{i=1}^N$ is

$$\|\mathbf{z}_i\|^2 = \sqrt{\frac{1}{\lambda_0}}, \quad i = 1, \ldots, N \tag{2.91}$$

The optimal λ, $\lambda^* = N^2$ and

$$\|\mathbf{z}_i\|^2 = \frac{1}{N}, \quad i = 1, \ldots, N \tag{2.92}$$

Hence $\widetilde{\mathbf{X}}_1^H \widetilde{\mathbf{X}}_1 = \frac{1}{N}\mathbf{I}$ and

$$\mathbf{M} = \mathbf{R}_{\mathbf{n}}^{-1/2} \widetilde{\mathbf{X}}_1 (\widetilde{\mathbf{X}}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \widetilde{\mathbf{X}}_1)^{-2} \widetilde{\mathbf{X}}_1^H \mathbf{R}_{\mathbf{n}}^{-1/2} = \sigma^2 \, N^2 \widetilde{\mathbf{X}}_1 \widetilde{\mathbf{X}}_1^H$$

$$\left(\mathbf{M} - \sigma^2 N \mathbf{I}\right) \widetilde{\mathbf{X}}_1 = 0$$

$$\mathbf{M} = \sigma^2 N \mathbf{I} \tag{2.93}$$

It follows that all eigenvalues $d_1, \ldots, d_N$ of the matrix $\mathbf{M}$ are equal to $\sigma^2 \, N$. □

Since all eigenvalues are equal, the expression for w_1 in (2.74) can be simplified to

$$w_1 = \sum_{i=1}^{N} \left(\frac{d_i}{\sum_{i=1}^{N} d_i}\right) |\widehat{\mathbf{n}}_{1,i}|^2 = \frac{1}{N} \sum_{i=1}^{N} |\widehat{\mathbf{n}}_{1,i}|^2. \tag{2.94}$$

M	α_1^*	λ'^*	M	α_1^*	λ'^*	M	α_1^*	λ'^*
1.0000	0.7427	0.2077	11.0000	0.8781	0.5652	21.0000	0.9002	0.6373
2.0000	0.7910	0.3163	12.0000	0.8817	0.5766	22.0000	0.9003	0.6373
3.0000	0.8152	0.3787	13.0000	0.8817	0.5766	23.0000	0.9041	0.6502
4.0000	0.8295	0.4186	14.0000	0.8853	0.5883	24.0000	0.9041	0.6502
5.0000	0.8417	0.4535	15.0000	0.8890	0.6002	25.0000	0.9042	0.6502
6.0000	0.8512	0.4816	16.0000	0.8926	0.6123	26.0000	0.9080	0.6634
7.0000	0.8577	0.5012	17.0000	0.8927	0.6123	27.0000	0.9080	0.6634
8.0000	0.8643	0.5217	18.0000	0.8964	0.6247	28.0000	0.9080	0.6634
9.0000	0.8677	0.5323	19.0000	0.8965	0.6247	29.0000	0.9119	0.6768
10.0000	0.8746	0.5540	20.0000	0.9002	0.6373	30.0000	0.9119	0.6768

Table 2.2: Optimal values of α_1 and λ for various M, number of unknown parameters

From Appendix 2.15, we see that w_1 is a Gamma distribution whose probability density function and cumulative distribution function as

$$f_{w_1}(y) = \frac{N^N}{(N-1)!} y^{n-1} e^{-Ny},\ y \geq 0 \tag{2.95}$$

$$F_{w_1}(y) = 1 - e^{-Ny} \sum_{j=0}^{N} \frac{(Ny)^j}{j!}, \tag{2.96}$$

Using the above distribution and the optimal solution in (2.82) we solve for the optimal solution to $\alpha_2(w_1)$ at the second stage. We already know that the minimum occurs at $\alpha_1 \approx 0.7421$ for $M = 1$. We solve for the optimal solution and find the gain in MSE for various values of M. The optimal values of α_1 and λ' for varying values of M is shown in Table 2.2. In particular, the performance of the two-step sequential design are plotted versus varying values of α_1 theoretically (solid) and via simulations (dashed dotted) for $M = 2$ and $M = 3$ case in Fig. 2.13 and 2.14, respectively. Figure 2.15 plots the optimal reduction for the two-step design for increasing values of M theoretically (solid) and via simulations (dashed dotted). It is interesting to note that the reduction in MSE decreases as the number of parameters increases. In fact, as the number of unknown parameters M goes to infinity, the ratio tends to 1. We will prove this fact in the following sections.

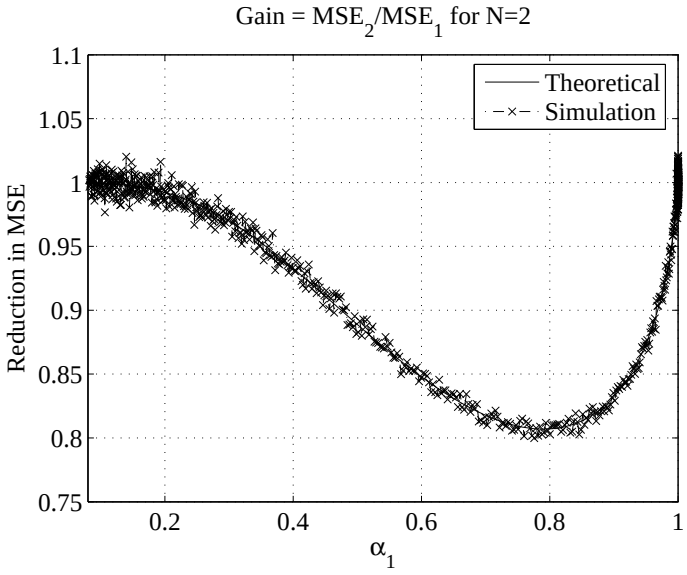


Figure 2.13: Plot of gain in two-step sequential design versus α_1 for $M = 2$ through theory and simulations

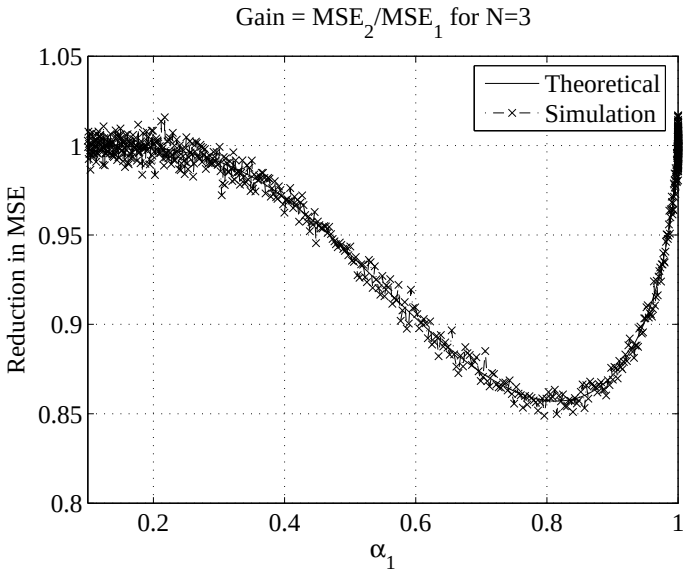


Figure 2.14: Plot of gain in two-step sequential design versus α_1 for $M = 3$ through theory and simulations

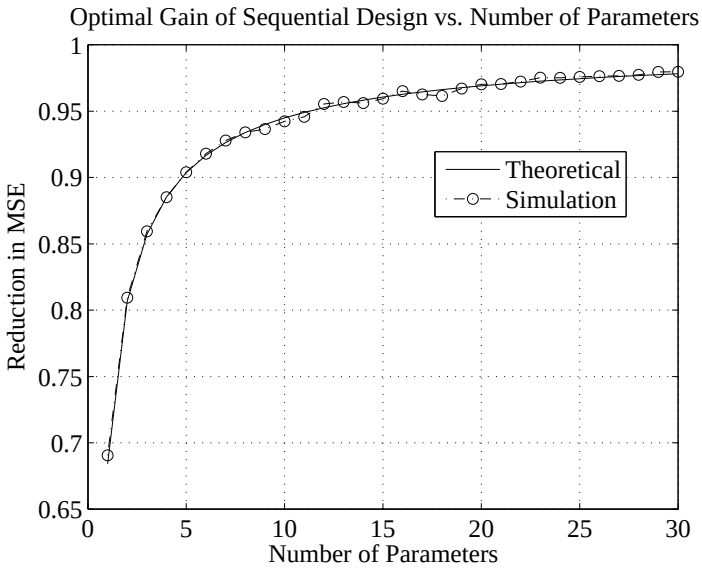


Figure 2.15: Plot of gain in two-step sequential design versus number of parameters M through theory and simulations

Suboptimal solution

As in the case of the single parameter case, we explore the performance gain for a suboptimal solution of the form,

$$\alpha_2(w_1) = \alpha_2 I(w_1 \geq \rho), \tag{2.97}$$

where α_2, ρ are chosen to satisfy the average energy constraint which can be written as

$$\alpha_1^2 + \alpha_2^2 \left(1 - F_\chi(w_1)\right) \leq 1. \tag{2.98}$$

This solution is motivated from the suboptimal solution presented in Section 2.4 for the scalar parameter case. Simplifying the energy constraint, we have

$$\alpha_2^2 = \frac{1 - \alpha_1^2}{1 - F_\chi(w_1)}, \tag{2.99}$$

where $F_\chi(w_1) = \mathrm{P}(w_1 \leq \rho)$ is the cumulative distribution function of the chi-square mixture. Substituting (2.97) in (2.73) yields,

$$\begin{aligned} \mathrm{tr}(\mathrm{MSE}^{(2)}) &= \mathrm{tr}(\mathrm{MSE}^{(1)})\mathrm{E}_{w_1}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2)^2} w_1 I(w_1 \geq \rho)\right. \\ &\left. + \frac{\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2} I(w_1 \geq \rho) + \frac{1}{\alpha_1^2} w_1 I(w_1 < \rho)\right] \qquad (2.100) \\ &= \mathrm{tr}(\mathrm{MSE}^{(1)})\left(\frac{1}{\alpha_1^2} + \left\{\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2)^2} - \frac{1}{\alpha_1^2}\right\}\mathrm{E}_{w_1}\left[w_1 I(w_1 \geq \rho)\right]\right. \\ &\left. + \frac{\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2}\left(1 - F_\chi(w_1)\right)\right), \qquad (2.101) \end{aligned}$$

where α_2 satisfies (2.99). We minimize the expression in (2.101) over $0 \leq \alpha_1 \leq 1$ and $\rho \in \mathcal{R}^+$. Figure 2.16 plots the gain in MSE as a function of the number of unknown parameters for the suboptimal solution through simulations along with the optimal gain. As is the case for a single parameter case, the best possible suboptimal design gives us an improvement of approximately 0.71. Also, it is worthwhile to note that the suboptimal solution provides near optimal performance. This is essentially due to the fact that the suboptimal solution is in accordance with the structure of optimal solution.

Note: The general two-step design procedure is given in terms of w_1 or in terms of $\{\widehat{\mathbf{n}}_{1,i}\}_{i=1}^N$ as in (2.74). Since

$$\widehat{\mathbf{n}}_1 = \mathbf{Q}^H\mathbf{R}_\mathbf{n}\mathbf{n}_1 = \mathbf{Q}^H\mathbf{R}_\mathbf{n}\left(\mathbf{y}_1 - \sqrt{E_0}\alpha_1\widetilde{\mathbf{X}}\boldsymbol{\theta}\right), \qquad (2.102)$$

it follows that the unknown $\boldsymbol{\theta}$ can be replaced by a guess $\boldsymbol{\theta}_g$ and a $K \times 2$-step procedure will yield the desired performance as in the single parameter case.

Asymptotic behavior of optimal design

Asymptotic distribution of w_1: When $d_1 = \ldots = d_N$, w_1 is a gamma distribution and it follows that the random variable w_1 asymptotically behaves as [66]

$$w_1 \sim \frac{1}{N}\chi_N \to \mathcal{N}(1, \frac{2}{N}) \quad as\ N \to \infty \qquad (2.103)$$

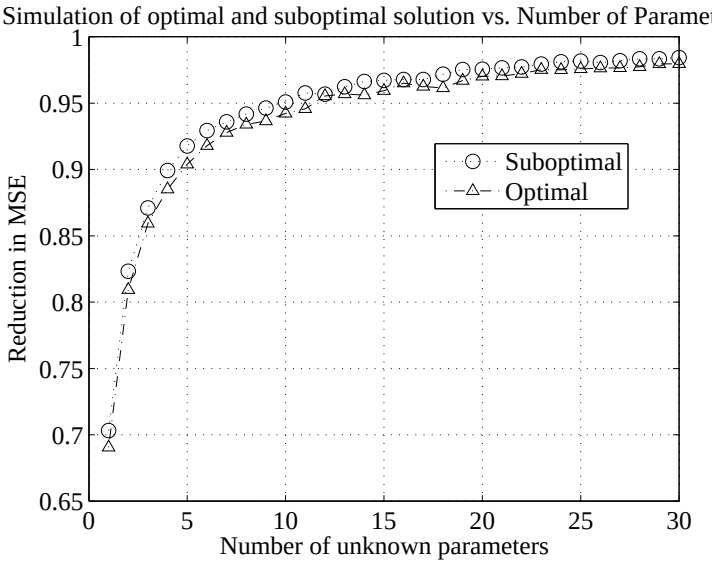


Figure 2.16: Plot of gains obtained in a suboptimal design versus number of unknown parameters to be estimated through simulations

Theorem 2.8.4. *tr(MSE$_2$) $\rightarrow$ tr(MSE$_1$) as $M \rightarrow \infty$, i.e., There is no gain in a sequential design procedure when the number of unknown parameters M goes to infinity.*

Proof. Since $N \geq M$, as $M \rightarrow \infty$, we have $w_1 \rightarrow \mathcal{N}(1, 2/N)$. It follows that asymptotically, the optimal design at the second step should concentrate all the energy at $w_1 = 1$, i.e., $\alpha_2(w_1) = \alpha_2\delta(w_1 - 1)$ and the corresponding energy constraint is $\alpha_1^2 + \alpha_2^2 \leq 1$. Hence the minimum MSE at the second step asymptotically is given by

$$\begin{aligned}
\mathrm{tr}(\mathrm{MSE}_2) &= \mathrm{tr}(\mathrm{MSE}_1)\mathrm{E}_{w_1}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2(w_1))^2}w_1 + \frac{\alpha_2^2(w_1)}{(\alpha_1^2+\alpha_2^2(w_1))^2}\right] \\
&= \lim_{M\rightarrow\infty} \mathrm{tr}(\mathrm{MSE}_1)\mathrm{E}_{w_1}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2\delta(w_1-1))^2}w_1 + \frac{\alpha_2^2\delta(w_1-1)}{(\alpha_1^2+\alpha_2^2\delta(w_1-1))^2}\right] \\
&= \lim_{M\rightarrow\infty} \mathrm{tr}(\mathrm{MSE}_1)\frac{1}{\alpha_1^2+\alpha_2^2} \\
&= \mathrm{tr}(\mathrm{MSE}_1)
\end{aligned}$$

□

Waveform and energy design for trace criterion

The problem of multiple parameter estimation is more complicated than estimation of a single parameter for the following reason. We showed in Section 2.2.1 that independent of the shape of $\mathbf{x}_i$, any non-adaptive energy allocation strategy is to assign all energy to the first step, i.e., a one-step strategy with energy E_0. But this is not true for the multiple parameter setting. Let us consider a simple example of estimating two parameters $\boldsymbol{\theta} = [\theta_1 \ \theta_2]^T$ in the model $\mathbf{y} = \mathbf{H}(\mathbf{x})\boldsymbol{\theta} + \mathbf{n}$, where

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 \\ 0 & x_2 \end{bmatrix}, \tag{2.104}$$

$\mathbf{x} = [x_1 \ x_2]^T$, $\mathbf{y} = [y_1 \ y_2]^T$, $\mathbf{n} = [n_1 \ n_2]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{R_n})$, and $\mathbf{R_n} = \sigma^2\mathbf{I}$. Then for a one-step process, we have $\mathrm{MSE}^{(1)}(\theta_1) = 2\sigma^2/x_1^2$ and $\mathrm{MSE}^{(1)}(\theta_2) = \sigma^2/x_2^2$. Minimizing $\mathrm{tr}(\mathrm{MSE}^{(1)}(\boldsymbol{\theta})) = \mathrm{MSE}^{(1)}(\theta_1) + \mathrm{MSE}^{(1)}(\theta_2)$ over the energy constraint $\|\mathbf{x}\|^2 \leq E_0 = 1$, we obtain $x_1 = x_2 = 1/\sqrt{2}$ and $\mathrm{tr}(\mathrm{MSE}^{(1)}_{\mathrm{min}}) = 6\sigma^2$. Now consider the following two-step non-adaptive strategy,

$$\begin{aligned} &\text{Step 1.} \qquad \mathbf{x} = [x_1 \ 0]^T, \qquad y_1 = x_1\theta_1 + n_1, \\ &\text{Step 2.} \qquad \mathbf{x} = [0 \ x_2]^T, \qquad [1 \ 1]\mathbf{y}_2 = 2x_2\theta_2 + [1 \ 1]\mathbf{n}_2. \end{aligned}$$

Minimizing the $\mathrm{tr}(\mathrm{MSE}^{(2)}(\boldsymbol{\theta})) = \mathrm{MSE}^{(2)}(\theta_1) + \mathrm{MSE}^{(2)}(\theta_2) = \sigma^2/x_1^2 + \sigma^2/2x_2^2$ over the energy constraint, we obtain $x_1 = x_2 = 1/\sqrt{2}$ and $\mathrm{tr}(\mathrm{MSE}^{(2)}_{\mathrm{min}}) = 3\sigma^2$. This translates to a 3dB gain in SNR for the two-step non-adaptive strategy over the one-step approach. We control the shape of the input $\mathbf{x} = [x_1 \ x_2]^T$ such that we have different energy allocation for each column of the matrix $\mathbf{H}$. By specifically designing the two-step non-adaptive strategy given in steps 1 and 2, we have reduced the estimation of the vector parameter $\boldsymbol{\theta} = [\theta_1, \theta_2]$ to two independent problems of estimating scalar parameters θ_1 and θ_2 respectively. For each of these scalar estimators, we design two

N-step sequential procedures ($2N$ steps in total) as in Section 2.7 for scalar controls x_1 and x_2 to obtain an improvement in performance of estimating $\boldsymbol{\theta}$. Applying the N-step design to both x_1 and x_2, we have $\mathrm{MSE}^{(N)}(\theta_1) = \mathcal{G}_N \mathrm{MSE}^{(2)}_{\min}(\theta_1)$ for the first N steps and $\mathrm{MSE}^{(N)}(\theta_2) = \mathcal{G}_N \mathrm{MSE}^{(2)}_{\min}(\theta_2)$ for the next N steps. Hence $\mathrm{tr}(\mathrm{MSE}^{(2N)}) = \mathcal{G}_N \mathrm{tr}(\mathrm{MSE}^{(2)}_{\min})$, where $\mathcal{G}_N$ is defined in (2.47). In other words, the MSE gains of the N-step procedure carry over to the vector parameter case as well.

2.9 Applications

2.9.1 MIMO channel estimation

It has been shown that multiple-input and multiple-output systems (MIMO) greatly increase the capacity of wireless systems [55, 100, 162] and hence MIMO has become an active area of research over the last decade [2, 132]. One important component in a MIMO system is the need to accurately estimate the channel state information (CSI) at the transmitter and receiver. This estimate has shown to play a crucial role in MIMO communications [25]. A recent and popular approach to channel estimation has been through the use of training sequences, i.e., known pilot signals are transmitted and channel is estimated using the received data and the pilot signals. A number of techniques for performing training based channel estimation have been proposed: maximum likelihood training method [102], least squares training [96], minimum mean squared estimation [140]. Recently, [15] proposed four different training methods for the flat block-fading MIMO system including the least squares and best linear unbiased estimator (BLUE) approach for the case of multiple LS channel estimates.

Problem formulation

In order to estimate the $r \times t$ channel matrix $\mathbf{\Theta}$ for a MIMO system with t transmit and r receive antennas, $N \geq t$ training vectors $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ are transmitted. The corresponding set of received signals can be expressed as [15, 93]

$$\mathbf{R} = \mathbf{\Theta X} + \mathbf{M}, \tag{2.105}$$

where $\mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_N]$ is a $r \times N$ matrix, $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_N]$ is the $r \times N$ matrix of sensor noise, $\mathbf{x}_i$ is the $t \times 1$ complex vector of transmitted signals, and $\mathbf{m}_i$ is the $r \times 1$ complex zero mean white noise vector. Let P_0 be the transmitted training power constraint, i.e., $\|\mathbf{X}\|_{\mathrm{F}}^2 = P_0$, $\| \cdot \|_F$ indicates Frobenius norm ($\|\mathbf{X}\|_{\mathrm{F}} = \sqrt{\mathrm{tr}(\mathbf{X}^H\mathbf{X})}$) and σ^2 denote the variance of receiver noise. Though $\mathbf{\Theta}$ is random, we estimate $\mathbf{\Theta}$ for a particular realization corresponding to the block of received data. The task of channel estimation is to recover the channel matrix $\mathbf{\Theta}$ based on the knowledge of $\mathbf{X}$ and $\mathbf{R}$ as accurately as possible under a transmit power constraint on $\mathbf{X}$. The standard LS solution and the corresponding estimation error can then be written as

$$\widehat{\mathbf{\Theta}}_{\mathrm{LS}} = \mathbf{R}\mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1} \tag{2.106}$$

$$\mathrm{MSE}_{\mathrm{LS}} = \frac{\sigma^2 t^2 r}{P_0}. \tag{2.107}$$

Assuming co-located transmitter and receiver arrays [123, 150] and multiple training periods available within the same coherency time (quasi-static) to estimate the channel, the set of received signals at the N time steps given by $\mathbf{R}_i = \mathbf{\Theta X}_i + \mathbf{M}_i$, $i = 1, 2, \ldots, N$, can be rewritten in the following form:

$$\mathbf{y}_i = \mathbf{H}(\mathbf{X}_i)\boldsymbol{\theta} + \mathbf{n}_i, \quad i = 1, 2, \ldots, N, \tag{2.108}$$

where $\mathbf{y}_i = \mathrm{vec}(\mathbf{R}_i), \boldsymbol{\theta} = \mathrm{vec}(\mathbf{\Theta}), \mathbf{n}_i = \mathrm{vec}(\mathbf{M}_i), \mathrm{vec}(\cdot)$ denotes the column-wise concatenation of the matrix, and $\mathbf{H}(\mathbf{X}_i) = (\mathbf{X}_i \otimes \mathbf{I})^T$ is a linear function of the input $\mathbf{X}_i$,

which is the same model described in (2.2). In [15], a method of linearly combining the estimates from each of the N stages was proposed and the MSE of the N stage estimator was shown to be $\mathrm{MSE}_{\mathrm{LS}}^{(N)} = \sigma^2 t^2 r / P_0$, where P_0 is the total power used in the N steps, i.e., $\sum_{i=1}^{N} \|\mathbf{X}_i\|_{\mathrm{F}}^2 \leq P_0$. If there are enough training samples, we could completely control the matrix $\mathbf{H}(\mathbf{X}_i)$ through the input $\mathbf{X}_i$ and make $\mathbf{H}(\mathbf{X}_i)$ orthogonal. In this case (2.108) along with the average power constraint $\mathrm{E}\left[\sum_i \|\mathbf{X}_i\|_{\mathrm{F}}^2\right] \leq P_0$ can benefit from adaptive energy allocation designs in Sections 2.7 and 2.8.2, where the problem is then separable into rt independent estimation problems of scalar parameters. Having N steps in the training sequence suggests an N-step energy allocation strategy. Hence it follows that using our strategy we are guaranteed to achieve the optimal error given by $\mathrm{MSE}^{(N)} \approx \mathcal{G}_N \sigma^2 t^2 r / P_0$, which we have shown to be at least 5dB (in 50 steps) better than any non-adaptive strategy.

2.9.2 Inverse scattering problem

The problem of imaging a medium using an array of transducers has been widely studied in many research areas such as mine detection, ultrasonic medical imaging [54], foliage penetrating radar, non-destructive testing [77], and active audio. The goal in an inverse scattering problem is to image and detect the present of scatterers (e.g., tanks, kidney stones, mines) in a medium using the backscattered signals.

Each scatterer is associated with a reflection coefficient, which is a measure of reflectivity of the scatterer. The higher the reflection coefficient, the higher the backscatter received. Accurate estimates of the reflection coefficients can be used to detect the reflective objects in the medium. When the number of scatterers (targets) of interest is sparsely distributed, the scattering coefficients are usually modeled as deterministic parameters and the phenomenon is termed as specular scattering. Multiple scattering arise in the presence of many scatterers creating a random en-

vironment. This random scattering is termed as diffusion scattering and in such a case, the scattering coefficients are modeled as Rayleigh distributed parameters. The multiple scattering tends to average out locally, so that the backscatter appears to be only a function of the intensity level in that region.

A recent approach [22] to the problem of imaging uses the concept of time reversal, which works by exploiting the reciprocity of a physical channel, e.g., acoustic, optical, or radio-frequency. One implication of reciprocity is that a receiver can reflect back a time reversed signal, thereby focusing the signal at the transmitter source [53]. Furthermore, with suitable prefiltering and aperture, the signal energy can also be focused on an arbitrary spatial location. This analysis assumes the noiseless scenario. For the noisy case, maximum likelihood estimation of deterministic point scatterers (specular component) was performed in [148].

We apply our concept of designing a sequence of measurements to image a medium of multiple scatterers using an array of transducers under a near-field approximation of the scatterers in the medium. Here, we discuss the application of our results to specular scattering, where the reflection coefficients are deterministic parameters. In Chapter III, we formulate a adaptive design of experiments for the Rayleigh scattering model.

Problem setting

We have N transducers located at positions $\{\mathbf{r}_k^a\}_{k=1}^N$, that transmit narrowband signals with center frequency ω rad/sec. The imaging area (or volume) is divided into V voxels at positions $\{\mathbf{r}_k^v\}_{k=1}^V$. The channel, denoted $\mathbf{a}_i$, between a candidate voxel i and the N transducers is given by the homogeneous Green's function as

$$\mathbf{a}_i = \left[\left(\frac{\exp(-j\omega/c\|\mathbf{r}_k^a - \mathbf{r}_i^v\|)}{\|\mathbf{r}_k^a - \mathbf{r}_i^v\|} \right)_{k=1\ldots N} \right]^T, \tag{2.109}$$

where c is the speed of light and $j = \sqrt{-1}$. This channel model is a narrowband near-field approximation, which ignores the effect of multiple scattering and has been widely adopted in other scattering studies, e.g., [18]. Each voxel can be characterized by its scatter coefficient, e.g., radar cross-section (RCS), $\{\theta_v\}_{v=1}^{V}$, which indicates the proportion of the received field that is re-radiated. Thus the channel between the transmitted field and the measured backscattered field at the transducer array is $\mathbf{A}\text{diag}(\boldsymbol{\theta})\mathbf{A}^T$, where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_V]$, $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_V]^T$, and $\text{diag}(\boldsymbol{\theta})$ denotes a $V \times V$ diagonal matrix with θ_i as its i^{th} diagonal element.

The probing mechanism for imaging of the scatter cross-section follows a sequential process, generating the following sequence of noise contaminated signals,

$$\mathbf{y}_i = \mathbf{A}\text{diag}(\boldsymbol{\theta})\mathbf{A}^T\mathbf{x}_i + \mathbf{n}_i = \mathbf{H}(\mathbf{x}_i)\boldsymbol{\theta} + \mathbf{n}_i,\ i = 1, 2, \ldots, N, \tag{2.110}$$

where $\mathbf{H}(\mathbf{x}_i) = \mathbf{A}\text{diag}(\mathbf{A}^T\mathbf{x}_i)$. The noises $\{\mathbf{n}_i\}$ are i.i.d complex normal random vectors with zero mean and a covariance matrix $\sigma^2\mathbf{I}$. The goal is to find estimates for the scattering coefficients $\boldsymbol{\theta}$ under the average energy constraint to minimize the MSE. If $\mathbf{A}$ is a square matrix, then we can condition $\text{diag}(\mathbf{A}^T\mathbf{x}_i)$ to have a single non zero component on any one of the diagonal elements, which translates to isolating the i^{th} column of $\mathbf{H}$ for any i. As in Section 2.8.2, we can perform V independent N-step experiments to guarantee the N-step gains of at least 5dB over the standard single step ML estimation for imaging [148]. If we are interested in optimally estimating any linear combination of the scattering coefficients, then the sequential strategy proposed in Section 2.8.1 can be used to achieve improvement in performance.

2.10 Conclusions

In this chapter, we considered the N-step adaptive waveform amplitude design problem for estimating parameters of an unknown channel under average energy

constraints. For a two-step problem, we found the optimal energy allocation at the second step as a function of the first measurement for a scalar parameter in the linear Gaussian model. We showed that this two-step adaptive strategy resulted in an improvement of at least 1.65dB over the optimal non-adaptive strategy. We then designed a suboptimal N-stage energy allocation procedure based on the two-step approach and demonstrated gains of more than 5dB in $N = 50$ steps. We extended our results to the case of vector parameters and provided applications of our design to MIMO and inverse scattering channel models.

2.11 Appendix: proof of equivalence

Denote the set of design parameters as $\mathbf{X} = \{\mathbf{x}_i(\mathbf{y}_1, \cdots, \mathbf{y}_{i-1})\}_{i=1}^N$. Let

$$\mathbf{X}^+ = \arg\min_{\mathbf{X}} \ \mathrm{MSE}^{(N)}(\mathbf{X})\mathrm{SNR}^{(N)}(\mathbf{X}) \tag{2.111}$$

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \ \mathrm{MSE}^{(N)}(\mathbf{X}) \text{ s.t } \ \mathrm{SNR}^{(N)}(\mathbf{X}) \leq \mathrm{SNR}_0. \tag{2.112}$$

Lemma 2.11.1. *For any $\beta \in \mathbb{R}$, $\beta\mathbf{X}^+$ is also a minimizer of the minimization criterion in* (2.111)*, where $\beta\mathbf{X} = \{\beta\mathbf{x}_i(\mathbf{y}_1, \cdots, \mathbf{y}_{i-1})\}_{i=1}^N$.*

Proof. From the energy definition in (2.6), the SNR definition in (2.10), and the property $\|\beta\mathbf{x}\| = \beta\|\mathbf{x}\|$, we obtain $\mathrm{SNR}^{(N)}(\beta\mathbf{X}^+) = \beta^2\mathrm{SNR}^{(N)}(\mathbf{X}^+)$. Using the scaling property of the linearity of $\mathbf{h}_1(\cdot)$, $\mathbf{h}_1(\beta\mathbf{x}) = \beta\,\mathbf{h}_1(\mathbf{x})$ in (2.5), we have $\mathrm{MSE}^{(N)}(\beta\mathbf{X}^+) = \frac{1}{\beta^2}\mathrm{MSE}^{(N)}(\mathbf{X}^+)$. Hence $\mathrm{MSE}^{(N)}(\beta\mathbf{X}^+)\mathrm{SNR}^{(N)}(\beta\mathbf{X}^+) = \mathrm{MSE}^{(N)}(\mathbf{X}^+)\mathrm{SNR}^{(N)}(\mathbf{X}^+)$, which is the minimum value of the criterion in (2.111). □

Since $\mathbf{X}^*$ minimizes the RHS of (2.112), we have

$$\mathrm{MSE}^{(N)}(\mathbf{X}^*) \leq \mathrm{MSE}^{(N)}(\beta\mathbf{X}^+), \tag{2.113}$$

where β satisfies $\mathrm{SNR}^{(N)}(\beta\mathbf{X}^+) \leq \mathrm{SNR}_0$. Similarly, from Lemma 2.11.1, we obtain

$$\mathrm{MSE}^{(N)}(\beta\mathbf{X}^+)\mathrm{SNR}^{(N)}(\beta\mathbf{X}^+) \leq \mathrm{MSE}^{(N)}(\mathbf{X}^*)\mathrm{SNR}^{(N)}(\mathbf{X}^*). \tag{2.114}$$

Multiplying (2.113) by $\mathrm{SNR}^{(N)}(\beta\mathbf{X}^+)$ and combining it with (2.114) yields

(2.115)

$$\mathrm{MSE}^{(N)}(\mathbf{X}^*)\mathrm{SNR}^{(N)}(\beta\mathbf{X}^+) \leq \mathrm{MSE}^{(N)}(\beta\mathbf{X}^+)\mathrm{SNR}^{(N)}(\beta\mathbf{X}^+) \leq \mathrm{MSE}^{(N)}(\mathbf{X}^*)\mathrm{SNR}^{(N)}(\mathbf{X}^*),$$

for $|\beta| \leq \sqrt{\frac{\mathrm{SNR}_0}{\mathrm{SNR}^{(N)}(\mathbf{X}^+)}}$. Choosing $\beta = \sqrt{\frac{\mathrm{SNR}^{(N)}(\mathbf{X}^*)}{\mathrm{SNR}^{(N)}(\mathbf{X}^+)}}$ satisfies this constraint and $\mathrm{SNR}^{(N)}(\beta\mathbf{X}^+) = \mathrm{SNR}^{(N)}(\mathbf{X}^*) \leq \mathrm{SNR}_0$. Replacing $\mathrm{SNR}^{(N)}(\beta\mathbf{X}^+)$ with $\mathrm{SNR}^{(N)}(\mathbf{X}^*)$ in (2.115), we obtain

(2.116)

$$\mathrm{MSE}^{(N)}(\mathbf{X}^*)\mathrm{SNR}^{(N)}(\mathbf{X}^*) \leq \mathrm{MSE}^{(N)}(\beta\mathbf{X}^+)\mathrm{SNR}^{(N)}(\mathbf{X}^*) \leq \mathrm{MSE}^{(N)}(\mathbf{X}^*)\mathrm{SNR}^{(N)}(\mathbf{X}^*),$$

It follows that $\mathrm{MSE}^{(N)}(\mathbf{X}^*) = \mathrm{MSE}^{(N)}(\beta\mathbf{X}^+)$, i.e., $\beta\mathbf{X}^+$ with $\beta = \sqrt{\frac{\mathrm{SNR}^{(N)}(\mathbf{X}^*)}{\mathrm{SNR}^{(N)}(\mathbf{X}^+)}}$ is the minimizer to the constrained minimization problem in (2.112). Furthermore,

Lemma 2.11.2. *$SNR^{(N)}(\mathbf{X}^*) = SNR_0$.*

Proof. By contradiction: If $\mathrm{SNR}^{(N)}(\mathbf{X}^*) < \mathrm{SNR}_0$, let $\mathrm{SNR}^{(N)}(\mathbf{X}^*) = \frac{1}{\beta}\mathrm{SNR}_0$ for some $\beta > 1$. Then by using the property $\|\alpha\mathbf{x}\| = \alpha\|\mathbf{x}\|$ in (2.6) and (2.10), we have $\mathrm{SNR}^{(N)}(\sqrt{\beta}\mathbf{X}^*) = \mathrm{SNR}_0$. Using linearity of $\mathbf{h}_1(\cdot)$ in (2.5), we obtain $\mathrm{MSE}^{(N)}(\sqrt{\beta}\mathbf{X}^*) = \frac{1}{\beta}\mathrm{MSE}^{(N)}(\mathbf{X}^*) < \mathrm{MSE}^{(N)}(\mathbf{X}^*)$. It follows that $\sqrt{\beta}\mathbf{X}^*$ satisfies the constraint and achieves a lower MSE than $\mathbf{X}^*$ which contradicts the fact that $\mathbf{X}^*$ is the minimum. □

2.12 Appendix: solution to problem 2.6.1

Properties of $\eta(z)$ in (2.40)

We list some of the properties of $\eta(z)$ which we will use to prove our results.

Proposition 2.12.1. *$\eta(z)$ achieves two-step minimum, i.e., $\eta(z)_{|z=0} = \eta^*$.*

Proposition 2.12.2. *$\eta(z)$ is an even function of z, i.e., $\eta(z) = \eta(|z|)$.*

Proof. $\eta(z)$ in (2.40) depends on z only through expected values of the form $\mathrm{E}_{\tilde{n}_1}\left[f(|\tilde{n}_1+z|)\right]$ for some continuous function f. Thus $\mathrm{E}_{\tilde{n}_1}\left[f(|\tilde{n}_1+z|)\right] = \mathrm{E}_{\tilde{n}_1}\left[f(\left|e^{-\angle z}\tilde{n}_1+|z|\right|)\right]$ $= \mathrm{E}_{\tilde{m}_1}\left[f(|\tilde{m}_1+|z||)\right]$, where $\tilde{m}_1 = e^{-\angle z}\tilde{n}_1$ is another complex Gaussian random variable with zero mean and unit variance. □

Proposition 2.12.3. *Optimal two-step minimum is achieved uniquely:* $\eta(z) > \eta(0)\ \forall z \in \mathbb{C} - \{0\}$. *Therefore,* $MSE^{(2)}(z) \times SNR^{(2)}(z) = \eta(z)$ *achieves a global minimum at* $z = 0$, *or* $\theta_g = \theta_1$.

Proof. By contradiction. If there exists z^+ such that $\eta(z^+) < \eta(0)$, then the design parameters α_1^* and $\alpha_2^*(\tilde{n}_1+z^+)$ will yield a $\mathrm{MSE}^{(2)} \times \mathrm{SNR}^{(2)} < \eta(0)$ which contradicts the fact that α_1^* and α_2^* achieves minimal $\eta(0)$. □

Proposition 2.12.4. *Continuity of* $\eta(z)$*: For any* $\epsilon > 0, \exists\ \delta > 0$ *such that for any* $0 \le z \le \delta$, $\eta^* \le \eta(z) \le (1+\epsilon)\eta^*$, *i.e., the optimal performance* $MSE^{(2)} \times SNR^{(2)} = \eta^*$ *can be approached within* ϵ *for any* θ_g *lying in the sphere* $0 \le |\theta_g - \theta_1| \le \frac{\delta}{\alpha_1\sqrt{SNR_0}}$. *Note that the sphere size increases as SNR decreases.*

Proof. It follows from the fact that the functions $\mathrm{E}\left[\cdot\right], \alpha_2^*(\cdot), f(\tilde{n}_1)$ are continuous. If the solution to $\alpha_2^*(\cdot)$ is the suboptimal thresholding function of the form (2.25) then the $\mathrm{MSE}^{(2)}(z)$ and $\mathrm{SNR}^{(2)}(z)$ are integrals of the probability density function of an independent Gaussian random variable over ellipsoids whose center is given through z and hence are still continuous functions in z. □

2.12.1 The $N\times$ two-step procedure

In this section, we take advantage of Proposition 2.12.4 to prove our result, i.e., the fact that the suboptimal solution presented in Section 2.6.2 can approach the optimal two-step solution in Section 2.3 when θ_g lies within a sphere centered at θ_1 with radius which increases as SNR decreases.

Consider N independent two-step experiments described below. The observations from the k^{th} experiment are

$$\mathbf{y}_1^k = \mathbf{h}_1(\mathbf{x}_1^k)\theta_1 + \mathbf{n}_1^k \tag{2.117}$$

$$\mathbf{y}_2^k = \mathbf{h}_1(\mathbf{x}_2^k)\theta_1 + \mathbf{n}_2^k,\ k = 1, 2, \ldots, N, \tag{2.118}$$

where $\mathbf{n}_1^k$ and $\mathbf{n}_2^k$ are i.i.d $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. The input design vectors for the N experiments are given by

$$\begin{aligned} \mathbf{x}_1^k &= \sqrt{\frac{E_0}{N(1+\epsilon)}}\alpha_1^*\mathbf{v}_{\text{m}} \\ \mathbf{x}_2^k &= \sqrt{\frac{E_0}{N(1+\epsilon)}}\alpha_2^*\left(\left|\frac{\mathbf{h}_1(\mathbf{v}_{\text{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\text{m}})\|}\frac{(\mathbf{y}_1^k - \mathbf{h}_1(\mathbf{x}_1^k)\theta_g)}{\sigma}\right|\right)\mathbf{v}_{\text{m}} \end{aligned} \tag{2.119}$$

$$= \sqrt{\frac{E_0}{N(1+\epsilon)}}\alpha_2^*\left(\left|\tilde{n}_1^k(\mathbf{y}_1^k;\theta_1) + z^{'}\right|\right)\mathbf{v}_{\text{m}},\ k = 1, 2, \ldots, N, \tag{2.120}$$

where

$$z^{'} = \frac{z}{\sqrt{N(1+\epsilon)}} = \sqrt{\frac{\text{SNR}_0}{N(1+\epsilon)}}\alpha_1^*(\theta_1 - \theta_g), \tag{2.121}$$

$$\begin{aligned} \tilde{n}_1^k(\mathbf{y}_1^k;\theta_1) &= \frac{\mathbf{h}_1(\mathbf{v}_{\text{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\text{m}})\|}\left(\frac{\mathbf{y}_1^k - \mathbf{h}_1(\mathbf{x}_1^k)\theta_1}{\sigma}\right) \\ &= \frac{\mathbf{h}_1(\mathbf{v}_{\text{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\text{m}})\|}\frac{\mathbf{n}_1^k}{\sigma},\ k = 1, 2, \ldots, N, \end{aligned} \tag{2.122}$$

and $\epsilon > 0$. The SNR in each experiment is

$$\begin{aligned} \text{SNR}^{(2),k}(z^{'}) &= \frac{\lambda_{\text{m}}(\boldsymbol{H}_1)}{\sigma^2}\text{E}\left[\|\mathbf{x}_1^k\|^2 + \|\mathbf{x}_2^k\|^2\right] \\ &= \frac{\text{SNR}_0}{N(1+\epsilon)}\left(\alpha_1^{*2} + \text{E}\left[\alpha_2^{*2}\left(\left|\tilde{n}_1^k + z^{'}\right|\right)\right]\right). \end{aligned} \tag{2.123}$$

Then the overall SNR in the $2N$ experiments (N two-step procedures) is given by

$$\begin{aligned} \text{SNR}^{(2N)}(z^{'}) &= \sum_{k=1}^{N}\text{SNR}^{(2),k}(z^{'}) \\ &= N\ \text{SNR}^{(2),1}(z^{'}) \\ &= \frac{\text{SNR}_0}{(1+\epsilon)}\left(\alpha_1^{*2} + \text{E}\left[\alpha_2^{*2}\left(\left|\tilde{n}_1^1 + z^{'}\right|\right)\right]\right), \end{aligned} \tag{2.124}$$

since $\{\tilde{n}_1^k\}_{k=1}^N$ are i.i.d $\mathcal{CN}(0,1)$ and hence the expected value of $\alpha_2^{*2}(\cdot)$ is identical and independent of k. The ML estimator for the k^{th} two-step process, $\hat{\theta}_1^{(2),k}$ is given by

$$\hat{\theta}_1^{(2),k} = \frac{\{\mathbf{h}_1(\mathbf{x}_1^k)\}^H \mathbf{y}_1^k + \{\mathbf{h}_1(\mathbf{x}_2^k)\}^H \mathbf{y}_2^k}{\|\mathbf{h}_1(\mathbf{x}_1^k)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2^k)\|^2}. \tag{2.125}$$

Since the N experiments are independent, the estimators $\{\hat{\theta}_1^{(2),k}\}_{k=1}^N$ are i.i.d random variables. The ML estimator for the $N \times 2$-step procedure is

$$\hat{\theta}_1^{(2N)} = \frac{1}{N} \sum_{k=1}^{N} \hat{\theta}_1^{(2),k} \tag{2.126}$$

and the corresponding MSE is

$$\text{MSE}^{(2N)}(z') = \text{E}\left[\left|\theta_1 - \hat{\theta}_1^{(2N)}\right|^2\right] = \frac{1}{N^2} \sum_{k=1}^{N} \text{MSE}^{(2),k}(z') = \frac{1}{N} \text{MSE}^{(2),1}(z'), \tag{2.127}$$

where $\text{MSE}^{(2),k}(z') = \text{E}\left[\left|\theta_1 - \hat{\theta}_1^{(2),k}\right|^2\right]$ are identical as the N two-step experiments are independent. Further, the input designs for each 2-step experiment given by (2.119) and (2.120) are the same as the suboptimal input designs in (2.37) and (2.38) with energy $E_0/(N(1+\epsilon))$, and using a similar derivation to (2.40), we obtain

$$\text{MSE}^{(2),k}(z') = \frac{N(1+\epsilon)}{\text{SNR}_0} \text{E}\left[\frac{\alpha_1^{*2}\left|\tilde{n}_1^k + z'\right|^2 + \alpha_2^{*2}\left(\left|\tilde{n}_1^k + z'\right|\right)}{\left(\alpha_1^{*2} + \alpha_2^{*2}\left(\left|\tilde{n}_1^k + z'\right|\right)\right)^2}\right]. \tag{2.128}$$

Lemma 2.12.5. *Given $\epsilon > 0$, $\theta_1, \theta_g \in [\theta_{\min}, \theta_{\max}]$, $\theta_{\min}, \theta_{\max} \in \mathbb{R}$, $\exists N_0$ such that $\forall N \geq N_0$, $\frac{SNR_0}{(1+\epsilon)} \leq SNR^{(2N)}(z') \leq SNR_0$.*

Proof. $\frac{\text{SNR}_0}{(1+\epsilon)} \leq \text{SNR}^{(2N)}(z'), \forall N \in \mathbb{N}$ follows from the fact $\text{SNR}^{(2N)}(z')$ is an even function of z and achieves minimum at $z' = 0$. (2.12.5.1)

Since $\text{E}[\cdot]$ and $\alpha_2^*(\cdot)$ are continuous functions, $\text{SNR}^{(2N)}(z')$ is continuous everywhere and hence at $z' = 0$. Thus for every $\zeta > 0$, $\exists \delta > 0$ such that $|z' - 0| < \delta$ implies $|\text{SNR}^{(2N)}(z') - \frac{\text{SNR}_0}{(1+\epsilon)}| < \zeta$. Choose $\zeta = \frac{\epsilon\ \text{SNR}_0}{(1+\epsilon)}$, we have $\text{SNR}^{(2N)}(z') \leq \text{SNR}_0$. (2.12.5.2)

Choose $N_0 = \left\lceil \frac{\alpha_1^{*2}\lambda_{\mathrm{m}}(\boldsymbol{H}_1)E_0|\theta_{\max}-\theta_{\min}|^2}{\sigma^2\delta^2(1+\epsilon)} \right\rceil$, from (2.12.5.1) and (2.12.5.2), the result follows. Q.E.D. □

Theorem 2.12.6. *Given $\Delta > 0$, $\theta_1, \theta_g \in [\theta_{\min}, \theta_{\max}]$, $\exists N_0$ such that $\forall N \geq N_0$ we have $\frac{\eta^*}{SNR_0} \leq MSE^{(2N)}(z^{'}) \leq \frac{(1+\Delta)\eta^*}{SNR_0}$.*

In other words, this theorem states that we can asymptotically achieve the performance of the optimal two-step estimator η^*/SNR_0 using the $N\times$ two-step procedure when θ_1 is bounded.

Proof. Comparing the product of expressions in (2.128) and (2.124) to the expression for $\eta(\cdot)$ in (2.40) and using Proposition 2.12.2, we have for a single two-step procedure that

$$\begin{aligned} \mathrm{MSE}^{(2),k}(z^{'}) \times \mathrm{SNR}^{(2),k}(z^{'}) &= \eta\left(z^{'}\right) \\ &= \eta\left(\alpha_1^*|\theta_1-\theta_g|\sqrt{\frac{\mathrm{SNR}_0}{N(1+\epsilon)}}\right). \end{aligned} \tag{2.129}$$

Since $\mathrm{MSE}^{(2N)} = \frac{1}{N}\mathrm{MSE}^{(2),k}$ and $\mathrm{SNR}^{(2N)} = N\mathrm{SNR}^{(2),k}$, the total MSE satisfies

$$\begin{aligned} \mathrm{MSE}^{(2N)}(z^{'}) \times \mathrm{SNR}^{(2N)}(z^{'}) &= \frac{1}{N}\mathrm{MSE}^{(2),k}(z^{'}) \times N\mathrm{SNR}^{(2),k}(z^{'}) \\ &= \eta\left(\alpha_1^*|\theta_1-\theta_g|\sqrt{\frac{\mathrm{SNR}_0}{N(1+\epsilon)}}\right). \end{aligned} \tag{2.130}$$

Using Proposition 2.12.3, $\mathrm{MSE}^{(2N)}(z^{'}) \times \mathrm{SNR}^{(2N)}(z^{'}) \geq \eta^*$. From RHS of *Lemma* 2.12.5, it follows that

$$\mathrm{MSE}^{(2N)}(z^{'}) \geq \frac{\eta^*}{\mathrm{SNR}^{(2N)}(z^{'})} \geq \frac{\eta^*}{\mathrm{SNR}_0}. \tag{2.131}$$

Since $\eta(\cdot)$ is continuous everywhere and at $z^{'} = 0$, it follows that for every $\mu^{'} > 0$, $\exists \delta^{'} > 0$ such that $|z^{'} - 0| \leq \delta^{'}$ implies $|\eta(z^{'}) - \eta^*| \leq \mu^{'}$. Choose $\mu^{'} = \eta^*\epsilon^{'}$, we obtain

$\eta(z^{'}) \leq (1+\epsilon^{'})\eta^{*}$. Thus for $N_0 = \left\lceil \frac{\alpha_1^{*2}\lambda_{\rm m}(\boldsymbol{H}_1)E_0|\theta_{\max}-\theta_{\min}|^2}{\sigma^2(1+\epsilon)} \right\rceil \max\left(\frac{1}{\delta}, \frac{1}{\delta'}\right)^2$, we have

$$\begin{aligned} \mathrm{MSE}^{(2N)}(z^{'}) &\leq \frac{(1+\epsilon^{'})\eta^{*}}{\mathrm{SNR}^{(2N)}(z')} \\ &\leq \frac{(1+\epsilon^{'})(1+\epsilon)\eta^{*}}{\mathrm{SNR}_0} \quad \text{from } \textit{Lemma } 2.12.5 \\ &\leq \frac{(1+\Delta)\eta^{*}}{\mathrm{SNR}_0}, \end{aligned} \tag{2.132}$$

where $\epsilon^{'} = \left(\frac{\Delta-\epsilon}{1+\epsilon}\right)$ and $0 < \epsilon < \Delta$. From (2.131) and (2.132), we have the result. □

2.13 Appendix: derivation of the N-step procedure

The design of the N-step procedure given by (2.42) can be written as

$$\alpha_i = A_i \prod_{s=1}^{i-1} I\left(|w_s|^2 \geq \rho_{s+1}\right),\ 1 \leq i \leq N, \tag{2.133}$$

where

$$w_s = \frac{\sum_{j=1}^{s} A_j \tilde{n}_j}{\sqrt{\sum_{j=1}^{s} |A_j|^2}}. \tag{2.134}$$

Then $\mathbf{w} = [w_1, \ldots, w_N]^T$ is a zero mean complex normal vector. Define the sets $U_i = \{|w_i|^2 \geq \rho_{i+1}\}, 1 \leq i \leq N$. For the set of events $\{U_i\}_{i=1}^{N-1}$, the set of events $\{D_i = \cap_{k=1}^{i-1} U_k \cap U_i^c,\ 1 \leq i \leq N-1, \cap_{k=1}^{N-1} U_k,\ i = N\}$ are disjoint and satisfy $I(\mathbf{w} \in \{D_i\}_{i=1}^{N}) = 1,\ \forall \mathbf{w} \in \mathbb{C}^N$. Hence

$$I\left(I(|w_1|^2 < \rho_2) + \sum_{i=2}^{N-1} I(|w_i|^2 < \rho_{i+1}) \prod_{s=1}^{i-1} I(|w_s|^2 \geq \rho_{s+1}) + \prod_{s=1}^{N-1} I(|w_s|^2 > \rho_{s+1})\right) = 1.$$

Substituting this expression inside the expectation for the $\mathrm{MSE}^{(N)}$ in (2.5), we obtain

$$\begin{aligned} \mathrm{MSE}^{(N)} = \mathrm{E}\Bigg[& \frac{|\sum_{i=1}^{N} \mathbf{h}_1(\mathbf{x}_i)^H \mathbf{n}_i|^2}{|\sum_{i=1}^{N} \|\mathbf{h}_1(\mathbf{x}_i)\|^2|^2} \Bigg(I(|w_1|^2 < \rho_2) + \\ & \sum_{i=2}^{N-1} I(|w_i|^2 < \rho_{i+1}) \prod_{s=1}^{i-1} I(|w_s|^2 \geq \rho_{s+1}) + \prod_{s=1}^{N-1} I(|w_s|^2 > \rho_{s+1}) \Bigg) \Bigg] \end{aligned}$$

$$
\begin{aligned}
&= \frac{\sigma^2}{\lambda_{\rm m}(\boldsymbol{H}_1)}\Bigg(\frac{1}{|A_1|^2}\mathrm{E}\left[|w_1|^2 I(|w_1|^2<\rho_2)\right] \\
&\quad + \sum_{i=2}^{N-1}\frac{1}{\sum_{j=1}^{i}|A_j|^2}\mathrm{E}\left[|w_i|^2 I(|w_i|^2<\rho_{i+1})\prod_{s=1}^{i-1} I(|w_s|^2>\rho_{s+1})\right] \\
&\quad + \frac{}{\sum_{j=1}^{N}|A_j|^2}\mathrm{E}\left[|w_N|^2\prod_{s=1}^{N-1} I(|w_s|^2>\rho_{s+1})\right]\Bigg)
\end{aligned}
$$

$$
\mathrm{MSE}^{(N)} = \frac{\sigma^2}{\lambda_{\rm m}(\boldsymbol{H}_1)}\left\{\sum_{i=1}^{N-1}\frac{T_i}{Q_i}+\frac{\tilde{T}_N}{Q_N}\right\}, \tag{2.135}
$$

where

$$
Q_i = \sum_{s=1}^{i}|A_s|^2, \tag{2.136}
$$

$$
\begin{aligned}
T_1 &= \mathrm{E}\left[|w_1|^2 I(|w_1|^2<\rho_2)\right] = 1-(1+\rho_2)e^{-\rho_2}, \\
T_i &= \mathrm{E}\left[|w_i|^2 I(|w_i|^2<\rho_{i+1})\prod_{s=1}^{i-1} I(|w_s|^2>\rho_{s+1})\right], \qquad (2.137) \\
\tilde{T}_i &= \mathrm{E}\left[|w_i|^2\prod_{s=1}^{i-1} I(|w_s|^2>\rho_{s+1})\right].
\end{aligned}
$$

The SNR of this N-step process is given by

$$
\begin{aligned}
\mathrm{SNR}^{(N)} &= \frac{\lambda_{\rm m}(\boldsymbol{H}_1)}{\sigma^2}\mathrm{E}\left[\sum_{i=1}^{N}\|\mathbf{x}_i\|^2\right] \\
&= \frac{\lambda_{\rm m}(\boldsymbol{H}_1)}{\sigma^2}\Big\{A_1^2\mathrm{E}\left[I(|w_1|^2<\rho_2)\right] \\
&\quad + \sum_{i=2}^{N-1}\left(\sum_{j=1}^{i-1}|A_j|^2\right)\mathrm{E}\left[I(|w_i|^2<\rho_{i+1})\prod_{s=1}^{i-1} I(|w_s|^2\geq\rho_{s+1})\right] \\
&\quad + \left(\sum_{j=1}^{N}|A_j|^2\mathrm{E}\left[\prod_{s=1}^{N-1} I(|w_s|^2\geq\rho_{s+1})\right]\right)\Big\} \\
&= \frac{\lambda_{\rm m}(\boldsymbol{H}_1)}{\sigma^2}\left\{\sum_{i=1}^{N-1}Q_iP_i+Q_N\tilde{P}_N\right\}, \qquad (2.138)
\end{aligned}
$$

where

$$
\begin{aligned}
P_1 &= \mathrm{E}\left[I(|w_1|^2 < \rho_2)\right] = |A_1|^2(1-e^{-\rho_2}), \\
P_i &= \mathrm{E}\left[I(|w_i|^2 < \rho_{i+1}) \prod_{s=1}^{i-1} I(|w_s|^2 \geq \rho_{s+1})\right], \\
\tilde{P}_i &= \mathrm{E}\left[\prod_{s=1}^{i-1} I(|w_s|^2 \geq \rho_{s+1})\right] = 1 - \sum_{k=1}^{i-1} P_k.
\end{aligned} \tag{2.139}
$$

From (2.135) and (2.138), we have

$$
\mathrm{MSE}^{(N)} \times \mathrm{SNR}^{(N)} = \left(\sum_{i=1}^{N-1} \frac{T_i}{Q_i} + \frac{\tilde{T}_N}{Q_N}\right)\left(\sum_{i=1}^{N-1} Q_i P_i + Q_N \tilde{P}_N\right). \tag{2.140}
$$

2.14 Appendix: derivation of two-step minmax criteria

The ML estimate and the MSE for the two-step process described by (2.54) and (2.55) are given by

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}^{(2)} &= \frac{1}{\sqrt{E_0}} \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{H}(\mathbf{u}_\mathrm{m})^H \mathbf{R}_\mathbf{n}^{-1} \left(\frac{\alpha_1 \mathbf{y}_1 + \alpha_2(\mathbf{y}_1)\mathbf{y}_2}{\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)}\right), \\
\mathrm{MSE}^{(2)} &= \frac{1}{E_0} \mathrm{E}\left[(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(2)})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(2)})^H\right] \\
&= \frac{1}{E_0} \mathrm{E}\left[\mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{H}(\mathbf{u}_\mathrm{m})^H \mathbf{R}_\mathbf{n}^{-1} \left(\frac{\alpha_1^2 \mathbf{n}_1 \mathbf{n}_1^H + \alpha_2^2(\mathbf{y}_1)\mathbf{R}_\mathbf{n}}{\left(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right)^2}\right) \mathbf{R}_\mathbf{n}^{-1} \mathbf{H}(\mathbf{u}_\mathrm{m}) \mathbf{W}_{\mathbf{u}_\mathrm{m}}\right].
\end{aligned}
$$

Then,

$$
\begin{aligned}
\Phi(\mathbf{u}, \mathrm{MSE}^{(2)}) &= \mathbf{u}^H \mathrm{MSE}^{(2)} \mathbf{u} \\
&= \frac{1}{E_0} \mathrm{E}\left[\frac{\alpha_1^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \left|\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{H}(\mathbf{u}_\mathrm{m})^H \mathbf{R}_\mathbf{n}^{-1} \mathbf{n}_1\right|^2 \right. \\
&\quad \left. + \frac{\alpha_2^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{u}\right] \\
&= \frac{1}{E_0} \mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{u}\ \mathrm{E}\left[\frac{\alpha_1^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \left|\frac{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{H}(\mathbf{u}_\mathrm{m})^H \mathbf{R}_\mathbf{n}^{-1} \mathbf{n}_1}{\sqrt{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{u}}}\right|^2 \right. \\
&\quad \left. + \frac{\alpha_2^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2}\right] \\
&= \Phi(\mathbf{u}, \mathrm{MSE}^{(1)})\ \mathrm{E}\left[\frac{\alpha_1^2 |\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta})|^2 + \alpha_2^2(\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}))}{(\alpha_1^2 + \alpha_2^2(\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta})))^2}\right],
\end{aligned} \tag{2.141}
$$

where

$$\tilde{n}_1(\mathbf{y}_1;\boldsymbol{\theta}) = \frac{\mathbf{u}^H\mathbf{W}_{\mathbf{u}_\mathrm{m}}\mathbf{H}(\mathbf{u}_\mathrm{m})^H\mathbf{R}_\mathbf{n}^{-1}(\mathbf{y}_1-\mathbf{H}(\mathbf{x}_1)\boldsymbol{\theta})}{\sqrt{\mathbf{u}^H\mathbf{W}_{\mathbf{u}_\mathrm{m}}\mathbf{u}}} = \frac{\mathbf{u}^H\mathbf{W}_{\mathbf{u}_\mathrm{m}}\mathbf{H}(\mathbf{u}_\mathrm{m})^H\mathbf{R}_\mathbf{n}^{-1}\mathbf{n}_1}{\sqrt{\mathbf{u}^H\mathbf{W}_{\mathbf{u}_\mathrm{m}}\mathbf{u}}} \tag{2.142}$$

is distributed $\mathcal{CN}(0,1)$.

2.15 Appendix: distribution of Gaussian mixture

Let $Y=\sum_{i=1}^N \alpha_i X_i^2$ where $\{X_i\}_{i=1}^N$ are independent unit normal random variables. Denote $\boldsymbol{\alpha}=[\alpha_1,\ldots,\alpha_N]$. Then the density ($g_N(\boldsymbol{\alpha},y)$) and distribution ($G_N(\boldsymbol{\alpha},y)$) function of Y is given as [84]

$$g_N(\boldsymbol{\alpha},y) = \left(\frac{y}{2}\right)^{\frac{n}{2}-1}\sum_{k=0}^{\infty}\frac{c_k(-1)^k\left(\frac{y}{2}\right)^k}{2\Gamma(\frac{n}{2}+k)} \tag{2.143}$$

$$G_N(\boldsymbol{\alpha},y) = \left(\frac{y}{2}\right)^{\frac{n}{2}}\sum_{k=0}^{\infty}\frac{c_k(-1)^k\left(\frac{y}{2}\right)^k}{\Gamma(\frac{n}{2}+k+1)}, \tag{2.144}$$

where c_k are determined by

$$c_0 = \Pi_{j=1}^n \alpha_j^{-\frac{1}{2}} \tag{2.145}$$

$$d_k = \frac{1}{2}\sum_{j=1}^n \alpha_j^{-k},\ k\geq 1 \tag{2.146}$$

$$c_k = \frac{1}{k}\sum_{r=0}^{k-1} d_{k-r}c_r \tag{2.147}$$

Since $w_1=\frac{1}{N}\sum_{i=1}^N|\widehat{\mathbf{n}}_{1,i}|^2=\frac{1}{2N}\sum_{i=1}^{2N}X_i^2$ where $\{X_i=\sqrt{2}\mathrm{Re}(\widehat{\mathbf{n}}_{1,i})\}_{i=1}^N$ and $\{X_i=\sqrt{2}\mathrm{Im}(\widehat{\mathbf{n}}_{1,i})\}_{i=N+1}^{2N}$ are independent unit normal random variables. $\boldsymbol{\alpha}=\frac{1}{2N}[1,\ldots,1]$ yields

$$c_k = \binom{N+k-1}{k}(2N)^{N+k},\ k\geq 0 \tag{2.148}$$

and the distribution of w_1 can then be written as

$$f_{w_1}(y) = \frac{N^N}{(N-1)!}y^{n-1}e^{-Ny},\ y\geq 0 \tag{2.149}$$

$$F_{w_1}(y) = 1-e^{-Ny}\sum_{j=0}^{N}\frac{(Ny)^j}{j!}, \tag{2.150}$$

Note that when all α_i's are equal, we have a mean of $2N$ independent unit normal random variables equivalent to a sum of N independent exponential random variables which is indeed a gamma distribution.

2.16 Appendix: proof of theorem 2.7.1

Proof. Since the shape of the design vectors $\mathbf{x}_i$ is $\mathbf{v}_{\mathrm{m}}$, we can write the set of transmitted signals as

(2.151)

$$\mathbf{x}_i\left(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1}\right)=\mathbf{v}_{\mathrm{m}}\sqrt{E_0}\alpha_i\left(\mathbf{y}_1-\mathbf{h}_1(\mathbf{x}_1)\theta_1,\ldots,\mathbf{y}_{i-1}-\mathbf{h}_1(\mathbf{x}_{i-1})\theta_1\right),\quad i=1,2,\ldots,K.$$

The inherent problem with any design $\mathcal{S}$ is the fact that the transmitted signal depend on the past through the noise magnitudes in the previous stages, i.e., we need to know the value of θ_1 to achieve the optimal performance. We overcome the dependence on the parameter of interest using a strategy similar to the one in Section 2.12.1. We replace θ_1 with a guess of θ_1 namely θ_g in the solution in (2.151). Then we have,

$$\begin{aligned}
\mathbf{x}_i &= \mathbf{v}_{\mathrm{m}}\sqrt{E_0}\alpha_i\left(\{\mathbf{y}_k-\mathbf{h}_1(\mathbf{x}_k)\theta_g\}_{k=1}^{i-1}\right)\\
&= \mathbf{v}_{\mathrm{m}}\sqrt{E_0}\alpha_i\left(\left\{\mathbf{n}_k+\sqrt{E_0}\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\alpha_k(\theta_1-\theta_g)\right\}_{k=1}^{i-1}\right)\\
&= \mathbf{v}_{\mathrm{m}}\sqrt{E_0}\alpha_i\left(\{\mathbf{n}_k+\alpha_k\mathbf{z}\}_{k=1}^{i-1}\right), \qquad (2.152)
\end{aligned}$$

where $\mathbf{z}=\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\sqrt{E_0}(\theta_1-\theta_g)$. Then the MSE of this N-step procedure can be written as

$$\mathrm{MSE}^{(K)}(\mathbf{z})=\frac{\sigma^2}{\lambda_{\mathrm{m}}E_0}\mathrm{E}\left[\frac{\left|\sum_{i=1}^{K}\alpha_i\left(\{\mathbf{n}_k+\alpha_k\mathbf{z}\}_{k=1}^{i-1}\right)\widetilde{n}_i\right|^2}{\left(\sum_{i=1}^{K}\alpha_i^2\left(\{\mathbf{n}_k+\alpha_k\mathbf{z}\}_{k=1}^{i-1}\right)\right)^2}\right] \qquad (2.153)$$

under the average constraint given by

$$\mathrm{SNR}^{(K)}(\mathbf{z})=\mathrm{SNR}_0\mathrm{E}\left[\sum_{i=1}^{K}\alpha_i^2\left(\{\mathbf{n}_k+\alpha_k\mathbf{z}\}_{k=1}^{i-1}\right)\right]\leq\mathrm{SNR}_0. \qquad (2.154)$$

Denote $\eta^{(K)}(\mathbf{z})$ as

$$\eta^{(K)}(\mathbf{z}) \;=\; \mathrm{MSE}^{(K)}(\mathbf{z}) \times \mathrm{SNR}^{(K)}(\mathbf{z}) \tag{2.155}$$

Proposition 2.16.1. *$MSE^{(K)}(z)$ and $SNR^{(K)}(z)$ are continuous functions of* $\mathbf{z}$

Proof. If the functions $\{\alpha_i\left(\{\mathbf{n}_k+\mathbf{z}\}_{k=1}^{i-1}\right)_{i=1}^{K}$ are continuous, then it follows that $\mathrm{MSE}^{(K)}(\mathbf{z})$ and $\mathrm{SNR}^{(K)}(\mathbf{z})$ are continuous since $\mathrm{E}[\cdot]$ and pdf of $\{\mathbf{n}_k\}_{k=1}^{K}$ are continuous functions. Also if the solution to $\{\alpha_i\}_{i=1}^{K}$ are thresholding functions of the form (2.42) then the $\mathrm{MSE}^{(K)}(\mathbf{z})$ and $\mathrm{SNR}^{(K)}(\mathbf{z})$ are integrals of the probability density function of independent gaussian random variable over ellipsoids whose center is given through $\mathbf{z}$ and hence are still continuous functions in $\mathbf{z}$. It then implies that $\eta^{(K)}(\mathbf{z})$ is also continuous in $\mathbf{z}$. □

$N \times K$-step procedure

Similar to the $N \times 2$-step procedure, we now construct an $N \times K$-step process, where we assume that the average energy in each of the N steps equals E_0/N. Then the ML estimate of θ_1 for the k^{th} K-step procedure is given by

$$\widehat{\theta}_1^{(K),k} = \frac{\sum_{j=1}^{N}\{\mathbf{h}_1(\mathbf{x}_j^k\}^H\mathbf{y}_j^k}{\sum_{j=1}^{N}\|\mathbf{h}_1(\mathbf{x}_j^k)\|^2}, \tag{2.156}$$

where the input design vector is

$$\mathbf{x}_j^k = \mathbf{v}_{\mathrm{m}}\sqrt{\frac{E_0}{N(1+\epsilon)}}\alpha_j\left(\left\{\mathbf{n}_i^k + \frac{\alpha_i\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right\}_{i=1}^{j-1}\right),\ 1\le i\le N,\ 1\le j\le K \tag{2.157}$$

and $\{\mathbf{n}_i^k\}_{i=1,k=1}^{K,N}$ are independent complex normal noises generated at the k^{th} step on the i^{th} stage. Then the overall ML estimate of θ_1 for the $N \times K$-step procedure is given by

$$\widehat{\theta}_1^{(K),N} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\theta}_1^{(K),i} \tag{2.158}$$

and the corresponding MSE is

(2.159)

$$\mathrm{MSE}^{(K),N}(\mathbf{z}) = \mathrm{E}\left[\|\theta_1 - \widehat{\theta}_1^{(K),N}\|^2\right] = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}\left[\|\theta_1 - \widehat{\theta}_1^{(K),i}\|^2\right] = \frac{1}{N}\mathrm{MSE}^{(K),1}(\mathbf{z}),$$

where $\mathrm{MSE}^{(K),1}(\mathbf{z})$ indicates the MSE of the first K^{th} estimator from the N stages and is given by

$$\mathrm{MSE}^{(K),1}(\mathbf{z}) = \mathrm{E}\left[\left\|\theta_1 - \widehat{\theta}_1^{(K),1}\left(\alpha_j\left(\left\{\mathbf{n}_k^1 + \frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right\}_{k=1}^{j-1}\right)\right)\right\|^2\right] \tag{2.160}$$

$$= N(1+\epsilon)\mathrm{MSE}^{(K)}\left(\frac{\alpha_k \mathbf{z}}{\sqrt{N(1+\epsilon)}}\right) \tag{2.161}$$

Substituting the expression for $\mathrm{MSE}^{(K),1}(\mathbf{z})$ in (2.161), we obtain

$$\mathrm{MSE}^{(K),N}(\mathbf{z}) = (1+\epsilon)\mathrm{MSE}^{(K)}\left(\frac{\mathbf{z}}{N(1+\epsilon)}\right) \tag{2.162}$$

The SNR of the $N \times K$-step procedure is

$$\begin{aligned}
\mathrm{SNR}^{(K),N}(\mathbf{z}) &= \sum_{j=1}^{N}\mathrm{SNR}^{(K),j} \\
&= \sum_{j=1}^{N}\sum_{i=1}^{K}\mathrm{E}\left[\left\|\mathbf{x}_i^j\left(\left\{\mathbf{n}_k^j + \frac{\alpha_k \mathbf{z}}{\sqrt{N(1+\epsilon)}}\right\}_{k=1}^{i-1}\right)\right\|^2\right] \\
&= \sum_{j=1}^{N}\frac{\mathrm{SNR}_0}{N(1+\epsilon)}\mathrm{E}\left[\sum_{i=1}^{K}\alpha_i^2\left(\left\{\mathbf{n}_k^j + \frac{\alpha_k \mathbf{z}}{\sqrt{N(1+\epsilon)}}\right\}_{k=1}^{i-1}\right)\right] \\
&= \frac{\mathrm{SNR}_0}{(1+\epsilon)}\mathrm{E}\left[\sum_{i=1}^{K}\alpha_i^2\left(\left\{\mathbf{n}_k^1 + \frac{\alpha_k \mathbf{z}}{\sqrt{N(1+\epsilon)}}\right\}_{k=1}^{i-1}\right)\right] \\
&= \frac{1}{(1+\epsilon)}\mathrm{SNR}^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right). \qquad (2.163)
\end{aligned}$$

From (2.162),(2.163), and (2.155), it follows that

$$\begin{aligned}
\mathrm{MSE}^{(K),N}(\mathbf{z})\mathrm{SNR}^{(K),N}(\mathbf{z}) &= \mathrm{MSE}^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right)\mathrm{SNR}^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right) \\
&= \eta^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right) \qquad (2.164)
\end{aligned}$$

Using continuity of $\mathrm{SNR}^{(K)}(\mathbf{z})$, for $\theta_1 \in [\theta_{\min}, \theta_{\max}]$, we have for any $\zeta_1 > 0$, $\exists \delta_1 > 0$ such that for $\left|\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right| \leq \delta_1$ we have $|\mathrm{SNR}^{(K),N}(\mathbf{z}) - \frac{\mathrm{SNR}_0}{(1+\epsilon)}| \leq \epsilon_1$. Choosing $\zeta_1 = \frac{\epsilon}{(1+\epsilon)}\mathrm{SNR}_0$, we obtain

$$\frac{1-\epsilon}{(1+\epsilon)}\mathrm{SNR}_0 \leq \mathrm{SNR}^{(K),N}(\mathbf{z}) \leq \mathrm{SNR}_0. \tag{2.165}$$

The condition $\left|\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right| \leq \delta_1$ is equivalent to $N \geq N_0$, where $N_0 = \left\lceil \frac{\lambda_{\mathrm{m}} E_0 |\theta_{\max} - \theta_{\min}|^2}{\delta_1(1+\epsilon)} \right\rceil$. Similarly continuity of $\eta^{(K)}(\mathbf{z})$ yields, for any $\zeta_2 > 0$, $\exists \delta_2 > 0$ such that for $\left|\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right| \leq \delta_2$, or equivalently $N \geq N_1$, where $N_1 = \left\lceil \frac{\lambda_{\mathrm{m}} E_0 |\theta_{\max} - \theta_{\min}|^2}{\delta_2(1+\epsilon)} \right\rceil$, we have

$$|\eta^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right) - \eta^*| \leq \zeta_2. \tag{2.166}$$

Choose $\zeta_2 = \epsilon_2 \eta^*$, then $\exists N \geq \max(N_0, N_1)$, such that

$$\begin{aligned} \mathrm{MSE}^{(K),N}(\mathbf{z}) &= \frac{\eta^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right)}{\mathrm{SNR}^{(K),N}(\mathbf{z})} \qquad (2.167) \\ &\leq \frac{(1+\epsilon_2)\eta^*}{\mathrm{SNR}^{(K),N}(\mathbf{z})} \\ &\leq \frac{(1+\epsilon_2)(1+\epsilon)}{1-\epsilon}\frac{\eta^*}{\mathrm{SNR}_0} \\ &\leq (1+\delta)\frac{\eta^*}{\mathrm{SNR}_0}, \qquad (2.168) \end{aligned}$$

where $\delta = (1+\epsilon_2)\frac{(1+\epsilon)}{(1-\epsilon)} - 1$. Further

$$\begin{aligned} \mathrm{MSE}^{(K),N}(\mathbf{z}) &= \frac{\eta^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right)}{\mathrm{SNR}^{(K),N}(\mathbf{z})} \\ &\geq (1-\epsilon_2)\frac{\eta^*}{\mathrm{SNR}_0}. \qquad (2.169) \end{aligned}$$

Since $1+\delta > 1+\epsilon_2$ implies $\delta > \epsilon_2$ implies $1-\epsilon_2 > 1-\delta$. Therefore from (2.168) and (2.169),

$$\left|\mathrm{MSE}^{(K),N}(\mathbf{z}) - \frac{\eta^*}{\mathrm{SNR}_0}\right| \leq \delta \tag{2.170}$$

Hence we can achieve the performance of a K-step design asymptotically using an $N \times K$-step design strategy. However in practice, an $N \times K$-step procedure might be an overkill to achieve the optimal performance ϵ close. This is because for the K-step strategy, the energy at each time step is already scaled by a factor proportional to $1/K$ which implies the effect of the unknown parameter θ_1 has a reduced effect on the overall performance even without the $N \times K$-step approach. $\square$

CHAPTER III

Energy allocation for estimation in a nonlinear model

In Chapter II we designed optimal energy allocation strategies for estimation in a linear channel model and discussed their applications to a linear inverse scattering problem. In this chapter, we generalize the solution of the optimal energy allocation for a nonlinear Rayleigh scattering medium. We study the effect of noise on the estimation of Rayleigh scatterer reflection magnitudes by setting up a sequential design of experiments. We derive an expression for the mean squared error for estimating the scattering magnitudes. Employing some of the design strategies for the case of the linear channel presented in the previous chapter, we derive suboptimal two-step energy allocation solutions that yields a 1.7dB performance improvement over the single step experiment for a Rayleigh fading channel. Closed-form expressions for the optimal transmission scheme and the minimum mean squared error are provided.

3.1 Introduction

Over the past decade, the problem of imaging has been widely studied in areas such as non-destructive testing [77], land mine detection, active audio, underwater acoustics [70] and ultrasonic medical imaging [163]. One recent approach to imaging uses the concept of time reversal [22,76]. Though time reversal imaging methods have been studied in detail for both deterministic and random scattering environments,

the performance of these methods in the presence of receiver noise has not been thoroughly studied. A detailed review of the time reversal imaging literature and its performance analysis is given in Appendix 3.5.

Design of experiments is another area which has found wide range of applications in statistical decision making [52, 171]. Sequential design [31, 133, 144] uses the knowledge of the past measurements to improve upon the performance of an estimator. Applied to the problem of imaging a scattered medium, a carefully designed sequence of measurements sounding the channel could alter the statistics of the next measurement to yield an overall reduction in mean squared error (MSE).

The results of Chapter II were shown to be applicable for imaging a specular scattering medium with a simple additive Gaussian measurement noise. In this chapter, we consider the problem of imaging a Rayleigh scattering medium and systematically study the effect of receiver noise on the imaging performance. We evaluate the imaging performance through the MSE of the least squared (LS) estimates of the scatterer reflection powers. For the case of a single dense scatterer, we obtain a closed-form expression for the MSE under the optimal transmission strategy. We assume that the spatial properties of the transmitted signals are fixed and find the optimal energy allocation scheme between the two transmissions involved in steps (i) and (iii) under the constraint that the total transmitted energy is fixed. We then show that we achieve a better performance than a single step strategy using this two-step design. We then extend the results to the case of constrained optimization.

This chapter is organized as follows: in Section 3.2, we present the concept of imaging a Rayleigh scattering medium using an iterative process of array measurements. In Section 3.3, we formulate the MSE criterion for LS estimation of the scatterer reflection powers. We offer an optimal two-step design that minimizes the MSE and

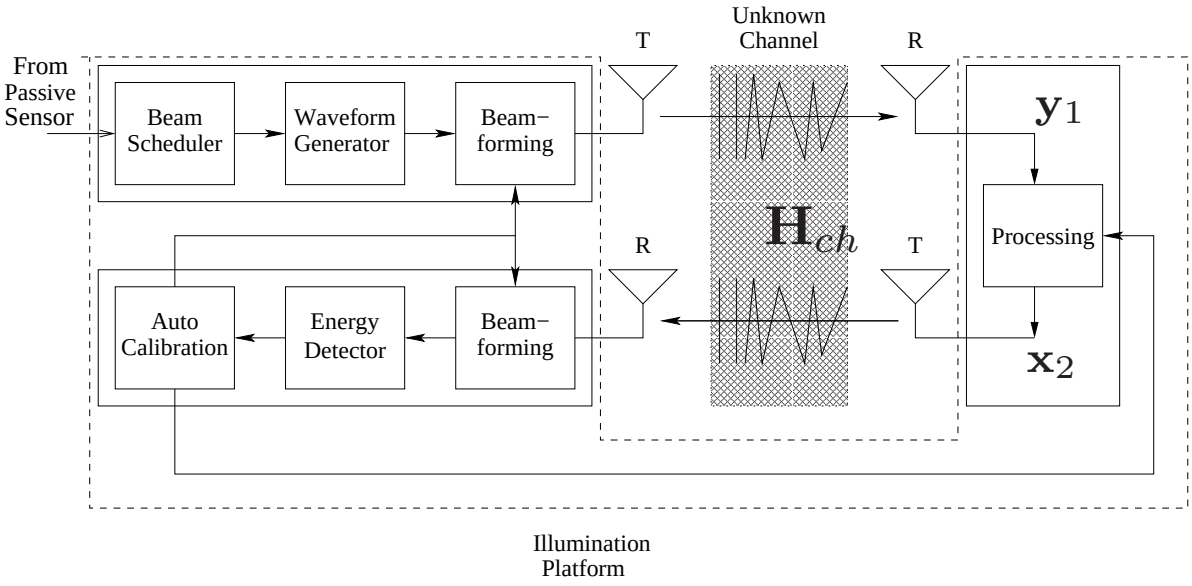


Figure 3.1: Measurement setup

show that this strategy outperforms the conventional beamformer. Furthermore, we provide simulation results to verify the optimal solution obtained analytically. We conclude this chapter in Section 3.4.

3.2 Model and mathematical description

The block diagram in Fig. 3.1 provides a high level description of the system. The signal flow in the block diagram is read clockwise from the upper left corner of the diagram. The three blocks surrounded by the box on the upper left of the diagram incorporate voxel selection (beam scheduling), spatio-temporal waveform selection and beamsteering followed by transmission into the medium, denoted as a dispersive spatio-temporal channel function $\mathbf{H}_{ch}$. The block on the right of the diagram processes the received backscattered signal and reinserts it into the medium $\mathbf{H}_{ch}$. Let us consider W transducers at locations $\{\mathbf{r}_k^a\}_{k=1}^W$ and M scatterers at unknown locations $\{\mathbf{r}_k^v\}_{k=1}^M$ with scattering strengths $\mathbf{d} = [d_1, \ldots, d_M]^T$ in the medium. The

most general channel model can be written as,

$$\mathbf{H}_{ch} = \mathbf{H}\mathbf{D}\mathcal{H}^T \tag{3.1}$$

where $\mathbf{D} = \text{diag}(\mathbf{d})$ are the scattering strengths along the diagonal, $\mathbf{H}$ denotes the random Green's function and $\mathcal{H}$ represents the full Green function taking into account the effective of both the background and multiple scattering [62, 148, 152]. The Green's function $\mathbf{H}$ and $\mathcal{H}$ between antenna i and scatterer j satisfy

$$\mathcal{H}(\mathbf{r}_j^v, \mathbf{r}_i^a) = \mathbf{H}(\mathbf{r}_j^v, \mathbf{r}_i^a) + \sum_{k \neq j} \mathbf{H}(\mathbf{r}_j^v, \mathbf{r}_k^v) d_k \mathcal{H}(\mathbf{r}_k^v, \mathbf{r}_i^a)$$

For e.g., if we assume second order scattering, then

$$\mathcal{H}(\mathbf{r}_j^v, \mathbf{r}_i^a) = \mathbf{H}(\mathbf{r}_j^v, \mathbf{r}_i^a) + \sum_{k \neq j} \mathbf{H}(\mathbf{r}_j^v, \mathbf{r}_k^v) d_k \mathbf{H}(\mathbf{r}_k^v, \mathbf{r}_i^a) \tag{3.2}$$

Most studies consider a homogeneous medium [148] where the standard Green's function between any candidate voxel i at location $\mathbf{r}$ and the W transducers is given by

$$\mathbf{h}_i(\mathbf{r}) = \left[\left(\frac{\exp(-j\omega/c \|\mathbf{r}_k^a - \mathbf{r}\|)}{\|\mathbf{r}_k^a - \mathbf{r}\|} \right)_{k=1 \ldots W} \right]^T. \tag{3.3}$$

For fixed point scatterers in the medium, $\mathbf{d}$ are the scattering strengths that need to be estimated, a problem that was studied in Chapter II of this dissertation. When the number of scatterers in the medium is large, there is random scattering. For random scattering, the scattering coefficients are random variables and are typically modeled as Rayleigh (or Rician) distributed in magnitude, i.e., the coefficients are independent circularly symmetric complex normal random variables with $\text{E}[d_i] = 0$, $\text{E}[|d_i|^2] = r_d(i)$, $\text{E}[d_i d_{i_1}] = 0$, $\text{E}[d_i (d_{i_1})^*] = 0$; $i_1 \neq i$. This implies that the elements of the channel matrix $\mathbf{H}_{ch}$ are jointly complex normal random variables and hence Rayleigh in magnitude. Note that matrix $\mathbf{H}$ is $W \times M$, $\mathbf{D}$ is $M \times M$, and $\mathbf{H}_{ch}$ is $W \times W$.

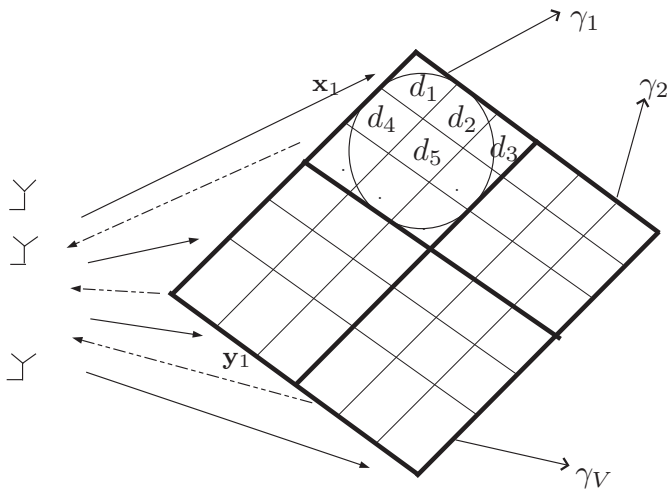


Figure 3.2: Scattering medium

The two-step probing mechanism for estimating the power distribution of the scatter coefficients $\mathbf{d}$ involves four signal processing steps and generates the following sequence of noise contaminated signals.

Step 1(a): The transducer array transmits a complex amplitude vector, $\mathbf{x}_1$.

Step 1(b): The transducer array receives the backscattered signal $\mathbf{H}_{ch}\mathbf{x}_1$ plus noise $\mathbf{n}_1$,

$$\mathbf{y}_1 = \mathbf{H}_{ch}\ \mathbf{x}_1 + \mathbf{n}_1. \tag{3.4}$$

Step 2(a): The transducer array transmits $\mathbf{x}_2 = \mathbf{x}_2(\mathbf{y}_1)$ which, in general, is a function of $\mathbf{y}_1$ allowing the system to exploit information about $\mathbf{H}_{ch}$ in the signal $\mathbf{y}_1$.

Step 2(b): The transducer array receives the second backscattered signal,

$$\mathbf{y}_2 = \mathbf{H}_{ch}\ \mathbf{x}_2 + \mathbf{n}_2. \tag{3.5}$$

The noises $\mathbf{n}_1, \mathbf{n}_2$ are i.i.d complex normal random vectors with zero mean and covariance matrix $\sigma^2\mathbf{I}$. For a Rayleigh scattering medium, $\mathbf{y}_1$ is complex normal with

a mean of $\mathrm{E}[\mathbf{y}_1] = 0$ and a covariance matrix given by

$$\begin{aligned} \mathbf{R}_{\mathbf{y}_1} &= \mathrm{E}\left[\mathbf{y}_1\mathbf{y}_1^H\right] \\ &= \sum_{l=1}^{V} \gamma_l \mathbf{R}_l(\mathbf{x}_1) + \sigma^2\mathbf{I}, \end{aligned} \tag{3.6}$$

where

$$\mathbf{R}_l(\mathbf{x}_1) = \frac{\int_{\mathbf{r}} \hat{r}_l(\mathbf{r})\mathbf{h}_l(\mathbf{r})\mathbf{h}_l(\mathbf{r})^H |\mathbf{h}_l^T(\mathbf{r})\mathbf{x}_1|^2 d\mathbf{r}}{\int_{\mathbf{r}} \hat{r}_l(\mathbf{r}) d\mathbf{r}} \tag{3.7}$$

$$\gamma_l = \int_{\mathbf{r}} \hat{r}_l(\mathbf{r}) d\mathbf{r}, \quad l = 1, \ldots, V, \tag{3.8}$$

and $\hat{r}_l(\mathbf{r})$ are the autocorrelation coefficients of the scatterer distribution. Our goal is to estimate $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_V]^T$, the fixed non-negative, Rayleigh scattering reflection powers (intensities) that are determined through the statistics of the scatter coefficients $\{d_i\}$. The interpretation of the scattering magnitudes can be stated as follows: when the number of random scatterers is large, a neighborhood of these scatterers undergo multiple scattering which averages out asymptotically to yield a constant scatterer reflection power for the particular region of interest. Figure 3.2 illustrates this procedure. The imaging area is divided into V cells and $\{\gamma_i\}_{i=1}^V$ denotes the average scatterer reflection power in the cells.

Step 2 (i.e., 2(a) and 2(b)) can be repeated to generate a sequential n-step procedure where the n transmitted signals would be $\mathbf{x}_j = \mathbf{x}_j(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{j-1})$, $j = 1, 2, \ldots, n$. For the n-step procedure, the sequence of received signals are distributed as

$$\begin{aligned} \mathbf{y}_j|\mathbf{x}_j &\sim \mathcal{CN}(0, \mathbf{R}_{\mathbf{y}_j}), \quad j = 1, 2, \ldots, n, \\ \mathbf{R}_{\mathbf{y}_j} &= \sum_{i=1}^{V} \gamma_i \mathbf{R}_i\left(\mathbf{x}_j(\{\mathbf{y}_k\}_{k=1}^{j-1})\right) + \sigma^2\mathbf{I}. \end{aligned} \tag{3.9}$$

Given the observations $\mathbf{y}_1, \ldots, \mathbf{y}_{j-1}$ at any step j, the objective is to design the next transmitted signal $\mathbf{x}_j(\mathbf{y}_1, \ldots, \mathbf{y}_{j-1})$ in order to improve the estimator performance.

3.3 Mean squared error calculation

We divide the analysis of the MSE into three parts: The one-stage estimator, an optimal two-step non sequential design for estimation and a two-step sequential design estimator. In the general setup, our goal is *to design a sequence of experiments to improve upon the performance of a one-step estimator under the constraint that the total transmitted energy, E_0 is fixed.*

3.3.1 One-step estimator

Given N sample observations of $\mathbf{y}_1$ $(W \times 1)$, an estimator $\widehat{\boldsymbol{\gamma}}_1(\{\mathbf{y}_{1k}\}_{k=1}^N)$ can be obtained by least squares fitting of $\gamma_1, \ldots, \gamma_V$ to the set of W^2 equations,

$$\widehat{\mathbf{R}}_{\mathbf{y}_1} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{y}_{1k}\mathbf{y}_{1k}^H = \sum_{i=1}^{V} \gamma_i \mathbf{R}_i(\mathbf{x}_1) + \sigma^2\mathbf{I} \tag{3.10}$$

Equation (3.10) can be rewritten as

$$\mathbf{M}\widehat{\boldsymbol{\gamma}}_1 \;=\; \mathbf{vec}\left(\widehat{\mathbf{R}}_{\mathbf{y}_1} - \sigma^2\mathbf{I}\right),$$

where $\mathbf{M} = [\mathbf{vec}\left(\mathbf{R}_1(\mathbf{x}_1)\right), \ldots, \mathbf{vec}\left(\mathbf{R}_V(\mathbf{x}_1)\right)]$ and $\mathbf{vec}(\mathrm{X})$ returns a vector obtained by stacking the columns of the matrix X. Given (3.10), the LS estimate of $\boldsymbol{\gamma}$ is given by,

$$\widehat{\boldsymbol{\gamma}}_1 \;=\; \left(\mathbf{M}^H\mathbf{M}\right)^{-1}\mathbf{M}^H\mathbf{vec}\left(\widehat{\mathbf{R}}_{\mathbf{y}_1} - \sigma^2\mathbf{I}\right). \tag{3.11}$$

The MSE for the one-step estimator can be written as

$$\begin{aligned} \mathrm{MSE}_1 &= \mathrm{E}\left[\left(\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}\right)\left(\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}\right)^H\right] \\ &= \frac{1}{N}\left(\mathbf{M}^H\mathbf{M}\right)^{-1}\boldsymbol{\Pi}\left(\mathbf{M}^H\mathbf{M}\right)^{-1}, \end{aligned} \tag{3.12}$$

where

$$\boldsymbol{\Pi} = N\ \mathbf{M}^H\mathrm{E}\left[\ \mathbf{vec}(\widehat{\mathbf{R}}_{\mathbf{y}_1} - \mathbf{R}_{\mathbf{y}_1})\mathbf{vec}(\widehat{\mathbf{R}}_{\mathbf{y}_1} - \mathbf{R}_{\mathbf{y}_1})^H\right]\mathbf{M}.$$

Using the fourth-order moment property for complex Gaussian vectors described in Appendix 3.7,

$$\mathbf{\Pi}_{i,j} = \mathrm{tr}\left[\mathbf{R}_i(\mathbf{x}_1)\mathbf{R}_{\mathbf{y}_1}\mathbf{R}_j(\mathbf{x}_1)\mathbf{R}_{\mathbf{y}_1}\right]. \tag{3.13}$$

Furthermore,

$$\begin{aligned}\left(\mathbf{M}^H\mathbf{M}\right)_{i,j} &= \mathbf{vec}\left(\mathbf{R}_i(\mathbf{x}_1)\right)^H \mathbf{vec}\left(\mathbf{R}_j(\mathbf{x}_1)\right) \\ &= \mathrm{tr}\left[\mathbf{R}_i(\mathbf{x}_1)\mathbf{R}_j(\mathbf{x}_1)\right]. \end{aligned} \tag{3.14}$$

Specialization to the single dense scatterer case

For a single unknown dense scatterer (γ_1), the LS one-step estimator of its reflection power from (3.11) and the corresponding MSE_1 from (3.12) become

$$\widehat{\gamma}_1^{(1)}(\mathbf{y}_1) = \frac{\mathrm{tr}\left(\mathbf{R}_1(\mathbf{x}_1)(\widehat{\mathbf{R}}_{\mathbf{y}_1} - \sigma^2\mathbf{I})\right)}{\mathrm{tr}\left(\mathbf{R}_1^2(\mathbf{x}_1)\right)}, \tag{3.15}$$

$$\begin{aligned}\mathrm{MSE}_1(\mathbf{x}_1) &= \frac{1}{N}\left(\frac{\gamma^2\mathrm{tr}\left(\mathbf{R}_1^4(\mathbf{x}_1)\right)}{\mathrm{tr}^2\left(\mathbf{R}_1^2(\mathbf{x}_1)\right)} + \frac{\sigma^4}{\mathrm{tr}\left(\mathbf{R}_1^2(\mathbf{x}_1)\right)}\right. \\ &\quad \left. + \frac{2\sigma^2\gamma\mathrm{tr}\left(\mathbf{R}_1^3(\mathbf{x}_1)\right)}{\mathrm{tr}^2\left(\mathbf{R}_1^2(\mathbf{x}_1)\right)}\right). \end{aligned} \tag{3.16}$$

Note that though the received signals are corrupted by complex normal noise, the estimate $\widehat{\gamma}_1^{(1)}$ is real as both matrices $\mathbf{R}_1(\mathbf{x}_1)$ and $\mathbf{R}_{\mathbf{y}_1}$ are Hermitian symmetric.

3.3.2 Two-step non sequential design

We find the optimal two-step non sequential energy design for estimating a scatterer reflector power γ_1 of a dense scatterer. For an N-step non sequential procedure, the energy allocation between the various steps is determined a priori and independent of the measurements at each time step. We assume we transmit a fixed unit norm waveform $\mathbf{v}$ and design only the energy allocation between the various steps. In other words, the transmit signal at time step i is given by $\mathbf{x}_i = \sqrt{E_i}\mathbf{v}$. For a

N-step non sequential process, the set of estimates of γ_1 is given by

$$\widehat{\gamma}_1^{(1),i}(\mathbf{y}_i) = \frac{\mathrm{tr}\left(\mathbf{R}_1(\mathbf{x}_i)(\widehat{\mathbf{R}}_{\mathbf{y}_i} - \sigma^2\mathbf{I})\right)}{\mathrm{tr}\left(\mathbf{R}_1^2(\mathbf{x}_i)\right)}, i = 1, 2, \ldots, N \tag{3.17}$$

and the corresponding MSE is

$$\begin{aligned} \mathrm{MSE}_{1i}(\mathbf{x}_i) &= \frac{1}{N}\left(\frac{\gamma^2 \mathrm{tr}\left(\mathbf{R}_1^4(\mathbf{x}_i)\right)}{\mathrm{tr}^2\left(\mathbf{R}_1^2(\mathbf{x}_i)\right)} + \frac{\sigma^4}{\mathrm{tr}\left(\mathbf{R}_1^2(\mathbf{x}_i)\right)}\right. \\ &\quad \left. + \frac{2\sigma^2\gamma \mathrm{tr}\left(\mathbf{R}_1^3(\mathbf{x}_i)\right)}{\mathrm{tr}^2\left(\mathbf{R}_1^2(\mathbf{x}_i)\right)}\right), i = 1, 2, \ldots, N. \end{aligned} \tag{3.18}$$

From (3.7), we have $\mathbf{R}_1(\mathbf{x}_i) = E_i\mathbf{R}_1(\mathbf{v})$. Substituting this expression in the MSE term, we obtain

$$\mathrm{MSE}_{1i}(\mathbf{x}_i) = \frac{1}{N}\left(\frac{\gamma^2 \mathrm{tr}\left(\mathbf{R}_1^4(\mathbf{v})\right)}{\mathrm{tr}^2\left(\mathbf{R}_1^2(\mathbf{v})\right)} + \frac{1}{E_i^2}\frac{\sigma^4}{\mathrm{tr}\left(\mathbf{R}_1^2(\mathbf{v})\right)} + \frac{1}{E_i}\frac{2\sigma^2\gamma \mathrm{tr}\left(\mathbf{R}_1^3(\mathbf{v})\right)}{\mathrm{tr}^2\left(\mathbf{R}_1^2(\mathbf{v})\right)}\right), \tag{3.19}$$

$$= t_0 + \frac{t_1}{\mathrm{SNR}_i} + \frac{t_2}{\mathrm{SNR}_i^2}, \quad i = 1, 2, \ldots, N, \tag{3.20}$$

where

$$t_0 = \frac{1}{N}\frac{\gamma^2 \mathrm{tr}\left(\mathbf{R}_1^4(\mathbf{v})\right)}{\mathrm{tr}^2\left(\mathbf{R}_1^2(\mathbf{v})\right)}, \tag{3.21}$$

$$t_1 = \frac{1}{N}\frac{2\gamma \mathrm{tr}\left(\mathbf{R}_1^3(\mathbf{v})\right)}{\mathrm{tr}^2\left(\mathbf{R}_1^2(\mathbf{v})\right)}, \tag{3.22}$$

$$t_2 = \frac{1}{N}\frac{1}{\mathrm{tr}\left(\mathbf{R}_1^2(\mathbf{v})\right)}, \tag{3.23}$$

and $\mathrm{SNR}_i = E_i/\sigma^2, i = 1, 2, \ldots, N$.

The N-step non sequential estimate of γ_1 can be written as

$$\widehat{\gamma}_1^{(N)} = \frac{\sum_{i=1}^N w_i \widehat{\gamma}_1^{(1),i}}{\sum_{i=1}^N w_i}, \tag{3.24}$$

where $\{w_i\}_{i=1}^N$ are the weights. These weights are optimally chosen to minimize the MSE of the overall estimator. The MSE of the N-step estimator is given by

$$\begin{aligned} \mathrm{MSE}_1^{(N)} &= \mathrm{E}\left[(\widehat{\gamma}_1^{(N)} - \gamma_1)^2\right] \\ &= \frac{\sum_{i=1}^N w_i \mathrm{MSE}_{1i}}{\left(\sum_{i=1}^N w_i\right)^2}, \end{aligned} \tag{3.25}$$

where the last equality follows from the fact that the N one-step estimators are independent. Minimizing $\mathrm{MSE}_1^{(N)}$ with respect to the weights w_i subject to a constraint on their sum $\sum_{i=1}^{N} w_i$, we obtain

$$w_i = \frac{1}{\mathrm{MSE}_{1i}}, \quad i = 1, 2, \ldots N. \tag{3.26}$$

Substituting the expression for weights from (3.26) in (3.25) yields,

$$\mathrm{MSE}_1^{(N)} = \frac{1}{\sum_{i=1}^{N} \frac{1}{\mathrm{MSE}_{1i}}} \tag{3.27}$$

The objective is to minimize the MSE expression in (3.27) subject to an energy constraint $\sum_{i=1}^{N} E_i \leq E_0$, where E_0 is the total available energy. The energy constraint can be written as

$$\sum_{i=1}^{N} \mathrm{SNR}_i \leq \mathrm{SNR}_0, \tag{3.28}$$

where $\mathrm{SNR}_0 = E_0/\sigma^2$. For a two-step process, the goal is to find the optimal SNR_1 and SNR_2 that minimizes $\mathrm{MSE}_1^{(2)}$. Since $\mathrm{MSE}_{1i} \propto 1/\mathrm{SNR}_i$, an optimal solution satisfies the energy constraint with equality. If not, we can always scale the energies to satisfy the constraint with equality and obtain a lower MSE. Substituting for the expressions of MSE_{1i} from (3.18) in (3.27), the derivative of (3.27) with respect to SNR_1 is given by

$$\frac{\partial \mathrm{MSE}_1^{(2)}}{\partial \mathrm{SNR}_1} = \frac{\mathrm{MSE}_{11}\frac{\partial \mathrm{MSE}_{12}}{\partial \mathrm{SNR}_1} + \mathrm{MSE}_{12}\frac{\partial \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1}}{\mathrm{MSE}_{11} + \mathrm{MSE}_{12}} - \frac{\mathrm{MSE}_{11}\mathrm{MSE}_{12}}{(\mathrm{MSE}_{11} + \mathrm{MSE}_{12})^2}\left(\frac{\partial \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1} + \frac{\partial \mathrm{MSE}_{12}}{\partial \mathrm{SNR}_1}\right) \tag{3.29}$$

Setting the derivative to zero yields

$$\frac{1}{(\mathrm{MSE}_{11})^2}\frac{\partial \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1} + \frac{1}{(\mathrm{MSE}_{12})^2}\frac{\partial \mathrm{MSE}_{12}}{\partial \mathrm{SNR}_1} = 0, \tag{3.30}$$

where

$$\frac{\partial \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1} = \frac{-t_1}{\mathrm{SNR}_1^2} - \frac{2t_2}{\mathrm{SNR}_1^3}, \tag{3.31}$$

$$\frac{\partial \mathrm{MSE}_{12}}{\partial \mathrm{SNR}_1} = \frac{t_1}{(\mathrm{SNR}_0 - \mathrm{SNR}_1)^2} + \frac{2t_2}{(\mathrm{SNR}_0 - \mathrm{SNR}_1)^3},$$

$$= \frac{t_1}{\mathrm{SNR}_2^2} + \frac{2t_2}{\mathrm{SNR}_2^3}, \tag{3.32}$$

where the last equality uses the fact that $\mathrm{SNR}_2 = \mathrm{SNR}_0 - \mathrm{SNR}_1$. Substituting the above expressions and the expressions for $\mathrm{MSE}_{1i}, i = 1, 2$ from (3.20) in (3.30), we obtain

(3.33)

$$\frac{1}{(t_0 + \frac{t_1}{\mathrm{SNR}_1} + \frac{t_2}{\mathrm{SNR}_1^2})^2} \left(\frac{-t_1}{\mathrm{SNR}_1^2} - \frac{2t_2}{\mathrm{SNR}_1^3} \right) = \frac{1}{(t_0 + \frac{t_1}{\mathrm{SNR}_2} + \frac{t_2}{\mathrm{SNR}_2^2})^2} \left(\frac{t_1}{\mathrm{SNR}_2^2} + \frac{2t_2}{\mathrm{SNR}_2^3} \right).$$

The optimal solution requires a solution to a sixth order equation. By symmetry, the value of SNR_1 that minimizes $\mathrm{MSE}_1^{(2)}$ will also be the value of SNR_2 that minimizes $\mathrm{MSE}_1^{(2)}$. Then $\mathrm{SNR}_1 = \mathrm{SNR}_2 = \mathrm{SNR}_0/2$ will be one solution satisfying (3.33) but it remains to be verified whether this solution is a local minima or a local maxima. Further the minima could also occur at the end points. In order to evaluate whether the solution minimizes or maximizes the expression, we need to compute the second order derivative of the $\mathrm{MSE}_1^{(2)}$ with respect to SNR_1. Computing the second derivative using the expression of the first derivative in (3.29) we obtain

$$\begin{aligned} \frac{\partial^2 \mathrm{MSE}_1^{(2)}}{\partial {\mathrm{SNR}_1}^2} &= \frac{2}{\mathrm{MSE}_{11} + \mathrm{MSE}_{12}} \left(\frac{\mathrm{MSE}_{12}}{\mathrm{MSE}_{11}} \frac{\partial \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1} + \frac{\mathrm{MSE}_{11}}{\mathrm{MSE}_{12}} \frac{\partial \mathrm{MSE}_{12}}{\partial \mathrm{SNR}_1} \right) \\ &\quad + \left(\frac{1}{\mathrm{MSE}_{11}^2} \left\{ \frac{\partial^2 \mathrm{MSE}_{11}}{\partial {\mathrm{SNR}_1}^2} - \frac{2}{\mathrm{MSE}_{11}} \left(\frac{\partial \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1} \right)^2 \right\} \right) \\ &\quad + \left(\frac{1}{\mathrm{MSE}_{12}^2} \left\{ \frac{\partial^2 \mathrm{MSE}_{12}}{\partial {\mathrm{SNR}_1}^2} - \frac{2}{\mathrm{MSE}_{12}} \left(\frac{\partial \mathrm{MSE}_2}{\partial \mathrm{SNR}_1} \right)^2 \right\} \right), \end{aligned} \tag{3.34}$$

where

$$\frac{\partial^2 \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1{}^2} = \frac{2t_1}{\mathrm{SNR}_1^3} + \frac{6t_2}{\mathrm{SNR}_1^4} \tag{3.35}$$

$$\frac{\partial^2 \mathrm{MSE}_{12}}{\partial \mathrm{SNR}_1{}^2} = -\frac{2t_1}{\mathrm{SNR}_2^3} - \frac{6t_2}{\mathrm{SNR}_2^4} \tag{3.36}$$

Evaluating the expressions for MSE_{11} and MSE_{12}, $\frac{\partial \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1}$, $\frac{\partial \mathrm{MSE}_{12}}{\partial \mathrm{SNR}_1}$, $\frac{\partial^2 \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1{}^2}$, and $\frac{\partial^2 \mathrm{MSE}_{12}}{\partial \mathrm{SNR}_1{}^2}$ in (3.20), (3.31), (3.32), (3.35), and (3.36) respectively at $\mathrm{SNR}_1 = \mathrm{SNR}_2 = \mathrm{SNR}_0/2$, we obtain

$$\begin{aligned}
\mathrm{MSE}_{11}(E_0/2) &= \mathrm{MSE}_{12}(E_0/2) = t_0 + \frac{2t_1}{\mathrm{SNR}_0} + \frac{4t_2}{\mathrm{SNR}_0^2} \\
\frac{\partial \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1}_{|E_1=E_0/2} &= -\frac{\partial \mathrm{MSE}_{12}}{\partial \mathrm{SNR}_1}_{|E_1=E_0/2} = \frac{-4t_1}{\mathrm{SNR}_0^2} - \frac{16t_2}{\mathrm{SNR}_0^3} \\
\frac{\partial^2 \mathrm{MSE}_{11}}{\partial \mathrm{SNR}_1{}^2}_{|E_1=E_0/2} &= \frac{\partial^2 \mathrm{MSE}_{12}}{\partial \mathrm{SNR}_1{}^2}_{|E_1=E_0/2} = \frac{16t_1}{\mathrm{SNR}_0^3} + \frac{96t_2}{\mathrm{SNR}_0^4}.
\end{aligned}$$

Substituting the above values in the second derivative expression in (3.34), we have

$$\begin{aligned}
\frac{\partial^2 \mathrm{MSE}_1^{(2)}}{\partial \mathrm{SNR}_1{}^2}_{|E_1=E_0/2} &= \frac{2}{\mathrm{MSE}_0^2}\left\{\frac{16t_1}{\mathrm{SNR}_0^3} + \frac{96t_2}{\mathrm{SNR}_0^4} - \frac{2}{\mathrm{MSE}_0}\left(\frac{-4t_1}{\mathrm{SNR}_0^2} - \frac{16t_2}{\mathrm{SNR}_0^3}\right)^2\right\} \\
&= \frac{32}{\mathrm{MSE}_0^3 \mathrm{SNR}_0^6}\left(t_0 t_1 \mathrm{SNR}_0^3 + 6t_0 t_2 \mathrm{SNR}_0^2 - 8t_2^2\right). \qquad (3.37)
\end{aligned}$$

Hence for the solution $E_1 = E_2 = E_0/2$ to be a minimum, the following condition must be satisfied,

$$t_0 t_1 \mathrm{SNR}_0^3 + 6t_0 t_2 \mathrm{SNR}_0^2 - 8t_2^2 \geq 0. \tag{3.38}$$

For a given set of parameters $\mathbf{R}_1(\mathbf{x}_1)$ and a fixed waveform $\mathbf{v}$, the SNR_0 should satisfy the condition in (3.38) for the equal energy allocation to be a solution. Figure 3.3 plots the $\mathrm{MSE}_1^{(2)}$ using the expression in (3.27) for varying values of E_1/E_0 between 0 and 1 at different SNR_0's for a channel defined by $\mathbf{R}_1(\mathbf{x}_1)$ and a waveform $\mathbf{v}$ chosen at random. The second derivative of the MSE is shown along each curve corresponding to a particular SNR_0. We observe that for all SNR_0 whose second derivative is non

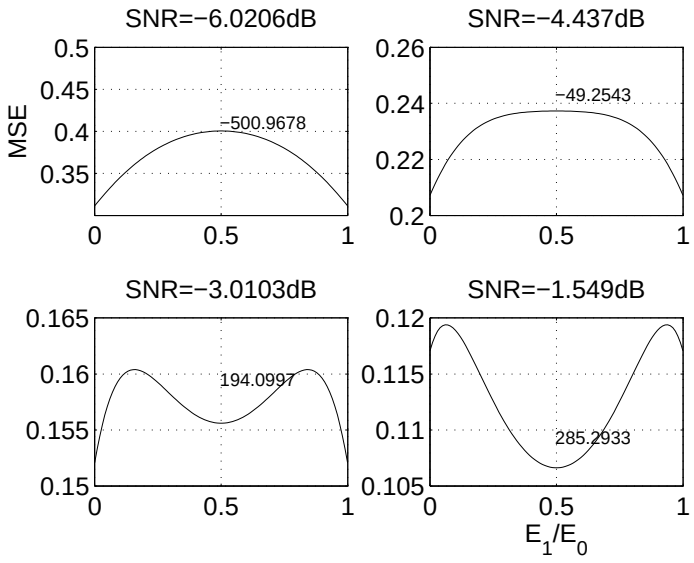


Figure 3.3: Plot of $\text{MSE}_1^{(2)}$ versus E_1/E_0 for varying SNR_0.

negative, the MSE attains a minimum at $E_1 = E_2 = E_0/2$. On the other hand, if the second derivative is positive then we have local maxima at $E_1 = E_2 = E_0/2$ while the minima occurs at the end points, where all energy is allocated to either the first or the second step.

Comparing design with linear Gaussian model

At low SNR, the MSE_{1i} can be approximated as

$$\text{MSE}_{1i} \approx \frac{\sigma^4 t_2}{E_i^2} \tag{3.39}$$

and the corresponding two-step estimator is given by

$$\text{MSE}_1^{(2)} \approx \frac{\sigma^4 t_2}{E_1^2 + E_2^2} \tag{3.40}$$

The MSE was optimized under the energy constraint $E_1 + E_2 \leq E_0$. For the two-step non sequential design in a linear Gaussian model discussed in Chapter II, we minimized the two-step MSE give by

$$\text{MSE}^{(2)} = \frac{\sigma^2 \lambda_{\text{m}}}{E_1 + E_2} \tag{3.41}$$

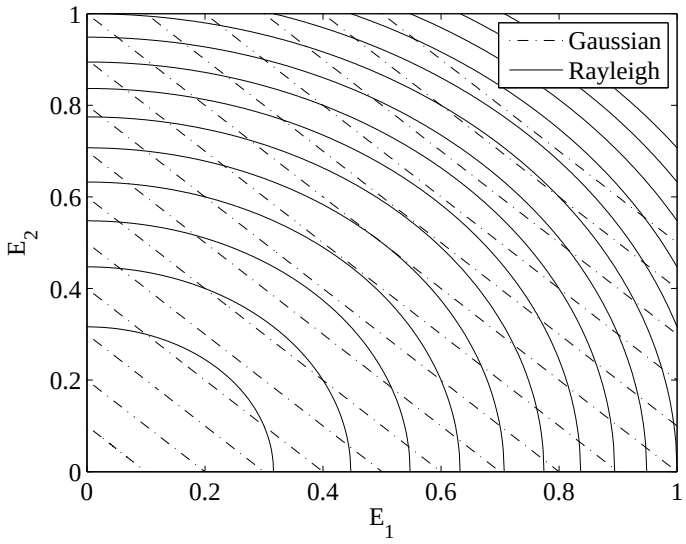


Figure 3.4: Contours of constant MSE for the Rayleigh and Gaussian case.

subject to the same constraint $E_1 + E_2 \leq E_0$. This suggested that any possible solution to E_1 and E_2 satisfying the constraint with equality is an optimal solution. In Fig. 3.4, we plot the contours of constant MSE for the Rayleigh and the Gaussian channel in the single dense scatterer case. For the Rayleigh channel, the optimal solution to the problem will lie at the end points of the support of E_1 and E_2, since a larger circle indicates a smaller SNR and the largest circle that can be drawn to intersect the constraint line $E_1 + E_2 = E_0$ is that circle which touches the two end points of the straight line at $E_1 = E_0$ and $E_2 = E_0$. From our derivation in the previous section, the second derivative becomes proportional to $-t_2^2 < 0$ at low SNR indicating that $E_1 = E_2 = E_0/2$ is a maxima in this low SNR case and not a minima.

3.3.3 Two-step sequential design

For a two-step sequential design, we search for a waveform $\mathbf{x}_2(\mathbf{y}_1)$ which yields a lower MSE than that achievable using $\mathbf{x}_1$ under the constraint that $\mathrm{E}\left[E_1 + E_2\right] \leq E_0$ where E_1 and E_2 are the average energies used in the first and second transmissions

respectively. We assume here that the spatial properties of $\mathbf{x}_1$ and $\mathbf{x}_2$ are fixed, and go after the energy allocation between the two steps that minimizes the MSE. The transmitted signal at the second step $\mathbf{x}_2(\mathbf{y}_1)$ can be written as

$$\mathbf{x}_2(\mathbf{y}_1) = \sqrt{E_2(\mathbf{y}_1)}\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} = \sqrt{E_2(\mathbf{y}_1)}\widetilde{\mathbf{x}}_1, \tag{3.42}$$

where $E_2(\mathbf{y}_1) = \|\mathbf{x}_2(\mathbf{y}_1)\|^2$, $E_2 = \mathrm{E}\left[\|\mathbf{x}_2(\mathbf{y}_1)\|^2\right]$, $E_1 = \|\mathbf{x}_1\|^2$, and $\widetilde{\mathbf{x}}_1$ is the normalized version of $\mathbf{x}_1$. We first look at the two-step design for a single scatterer case. Let $\widehat{\gamma}_1^{(1)}(\mathbf{y}_1)$ and $\widehat{\gamma}_1^{(1)}(\mathbf{y}_2)$ be the LS estimates of γ_1 obtained from the two steps by transmitting signals $\mathbf{x}_1$ and $\mathbf{x}_2(\mathbf{y}_1)$ respectively. The overall two-step estimate of γ is

$$\widehat{\gamma}_1^{(2)} = \frac{w_1\widehat{\gamma}_1^{(1)}(\mathbf{y}_1) + w_2\widehat{\gamma}_1^{(1)}(\mathbf{y}_2)}{w_1 + w_2}, \tag{3.43}$$

where the weights w_1 and w_2 are chosen to minimize the MSE. The MSE of the two-step estimate is

$$\begin{aligned} \mathrm{MSE}_2 &= \mathrm{E}\left[\left(\widehat{\gamma}_1^{(2)} - \gamma\right)^2\right] = \mathrm{E}_{\mathbf{y}_1}\left[\mathrm{E}_{\mathbf{y}_2|\mathbf{y}_1}\left[\left(\widehat{\gamma}_1^{(2)} - \gamma\right)^2\right]\right] \\ &= \mathrm{E}_{\mathbf{y}_1}\left[\mathrm{MSE}_2|\mathbf{y}_1\right], \end{aligned} \tag{3.44}$$

where

$$\mathrm{MSE}_2|\mathbf{y}_1 = \frac{w_1^2(\widehat{\gamma}_1^{(1)}(\mathbf{y}_1) - \gamma)^2 + w_2^2(\mathrm{MSE}_1(\mathbf{x}_2(\mathbf{y}_1)))}{(w_1 + w_2)^2}.$$

Minimizing $\mathrm{MSE}_2|\mathbf{y}_1$ with respect to w_1 and w_2, we get $w_1\left(\widehat{\gamma}_1^{(1)}(\mathbf{y}_1) - \gamma\right)^2 = w_2\mathrm{MSE}_1(\mathbf{x}_2(\mathbf{y}_1))$. Substituting for the optimal weights in equation (3.44), the two-stage MSE is

$$\mathrm{MSE}_2 = \mathrm{E}_{\mathbf{y}_1}\left[\frac{1}{\left(\frac{1}{(\widehat{\gamma}_1^{(1)}(\mathbf{y}_1) - \gamma)^2} + \frac{1}{\mathrm{MSE}_1(\mathbf{x}_2(\mathbf{y}_1))}\right)}\right]. \tag{3.45}$$

In Section 2.3 of the previous chapter, the optimal solution to the two-step sequential energy allocation procedure for the additive Gaussian channel model was

found to be a thresholding strategy. A similar thresholding solution to energy at the second stage can be written as,

$$E_2(\mathbf{y}_1) = A\ I\left(\left[\frac{\widehat{\gamma}_1^{(1)}(\mathbf{y}_1) - \gamma}{\sqrt{\mathrm{MSE}_1(\mathbf{x}_1)}}\right]^2 > \rho\right), \tag{3.46}$$

where $I(\cdot)$ is the indicator function and A is chosen to satisfy the energy constraint. This solution implies that if the particular realization of $\widehat{\gamma}_1$ was closer than average to the true value, then it is fairly accurate and thus there is no need to retransmit energy.

We first look at the solution to this problem at low SNR (SNR $= \frac{E_0}{\sigma^2}$). At low SNR, the MSE for the two stages can be approximated as

$$\mathrm{MSE}_1(\mathbf{x}_1) \approx \frac{\mathrm{H}}{NE_1^2}, \tag{3.47}$$

$$\mathrm{MSE}_1(\mathbf{x}_2(\mathbf{y}_1)) \approx \frac{\mathrm{H}}{NE_2^2(\mathbf{y}_1)}, \tag{3.48}$$

where

$$\mathrm{H} = \frac{\sigma^4}{\mathrm{tr}\left(\mathbf{R}_1^2(\widetilde{\mathbf{x}}_1)\right)}.$$

When N is large, the averaging associated with first estimate of γ drives the standardized MSE $\left(\frac{\widehat{\gamma}_1^{(1)}(\mathbf{y}_1)-\gamma}{\sqrt{\mathrm{MSE}_1(\mathbf{x}_1)}}\right)$ to asymptotically zero mean unit variance normal random variable, n_1. Substituting for n_1 and $\mathrm{MSE}_1(\mathbf{x}_1)$, $\mathrm{MSE}_1(\mathbf{x}_2(\mathbf{y}_1))$ from equations (3.47), (3.48) into equation (3.45), the MSE for the two-step design is

$$\begin{aligned} \mathrm{MSE}_2 &= \frac{\mathrm{H}}{N}\,\mathrm{E}_{\mathrm{n}_1}\left[\frac{\mathrm{n}_1^2 I(\mathrm{n}_1^2 > \rho)}{E_1^2 + \mathrm{n}_1^2 E_2^2(\mathbf{y}_1)}\right] \\ &= \frac{\mathrm{H}}{N}\mathrm{E}_{\mathrm{n}_1}\left[\frac{\mathrm{n}_1^2 I(\mathrm{n}_1^2 > \rho)}{E_1^2 + \mathrm{n}_1^2 A^2} + \frac{\mathrm{n}_1^2 I(\mathrm{n}_1^2 \le \rho)}{E_1^2}\right] \end{aligned} \tag{3.49}$$

So our goal now is to minimize this two-step MSE for the optimal energy allocation between the two steps subject to the energy constraint which can be written as

$$\mathrm{E}_{\mathbf{y}_1}\left[E_1 + E_2(\mathbf{x}_2(\mathbf{y}_1))\right] \le E_0 \tag{3.50}$$

Substituting the suboptimal energy solution from equation (3.46) into (3.50), we obtain

$$\begin{aligned} E_1 + A\ \mathrm{E}\left[I(|\mathrm{n}_1|^2 > \rho)\right] &\leq E_0 \\ A &\leq E_0 \frac{(1-\alpha)}{2\mathrm{Q}(\sqrt{\rho})}, \end{aligned} \tag{3.51}$$

where $\alpha = \frac{E_1}{E_0}$ is the fraction of energy allocated to the first step. Putting back the constraint into the MSE_2 expression in equation (3.49) we get

$$\begin{aligned} \mathrm{MSE}_2 &= \frac{\mathrm{H}}{NE_0^2}\mathrm{E}_{\mathrm{n}_1}\left[\frac{\mathrm{n}_1^2 I(\mathrm{n}_1^2 > \rho)}{\alpha^2 + \left(\frac{\mathrm{n}_1(1-\alpha)}{2\mathrm{Q}(\sqrt{\rho})}\right)^2} + \frac{\mathrm{n}_1^2 I(\mathrm{n}_1^2 \leq \rho)}{\alpha^2}\right] \\ &= \frac{H}{NE_0^2}\left(2\int_{\sqrt{\rho}}^{\infty} \frac{\mathrm{n}_1^2}{\alpha^2 + \mathrm{n}_1^2\left(\frac{(1-\alpha)}{2\mathrm{Q}(\sqrt{\rho})}\right)^2}\mathrm{f}(\mathrm{n}_1)d\mathrm{n}_1 \right. \\ &\quad \left. + \frac{1}{\alpha^2}\left[-\sqrt{\frac{2\rho}{\pi}}e^{\frac{\rho}{2}} + 1 - 2\mathrm{Q}(\sqrt{\rho})\right]\right), \end{aligned} \tag{3.52}$$

where the integral is evaluated numerically. Minimizing MSE_2 in the above expression, the optimal solution to ρ and α and the corresponding MSE at low SNR is found to be

$$\rho_{\mathrm{opt}} \approx 0.8885, \quad \alpha_{\mathrm{opt}} = \frac{E_{1\mathrm{opt}}}{E_0} \approx 0.66, \tag{3.53}$$

$$\mathrm{MSE}_2(\gamma_1) \approx 0.6821\ \frac{H}{NE_0^2} = 0.6821\ \mathrm{MSE}_1(\gamma_1) \tag{3.54}$$

corresponding to a reduction in MSE by 68%. Figures 3.6 and 3.7 show the analytical (solid line) and simulation (dashed line) plots of the gain $= \frac{\mathrm{MSE}_2}{\mathrm{MSE}_1}$ as a function of ρ for α_{opt} and as a function of α for ρ_{opt} at SNR $= -10$dB. Since $\frac{\mathrm{MSE}_2}{\mathrm{MSE}_1} < 1$, we obtain a reduction in MSE using our sequential design approach. The plot of gain in MSE vs. SNR corresponding to α_{opt} and ρ_{opt} at low SNR is shown in Fig. 3.8 through simulation (dashed line) and analytically (solid line).

Our design procedure is more critical at a lower SNR for the following reason. It is important to note that the solution to $\mathbf{x}_2(\mathbf{y}_1)$ and the weights depend on the value of γ which is unknown. However, if we are given information of the form $\gamma_1 \in [\gamma_a, \gamma_b]$ for any $-\infty < \gamma_a, \gamma_b < \infty$, then it is possible to incorporate this knowledge in making the optimal decision for $\mathbf{x}_2$ by replacing γ_1 with γ_g in (3.46) :

$$E_2(\mathbf{y}_1) = A\ \mathbf{I}\left(\left[\frac{\widehat{\gamma}_1^{(1)}(\mathbf{y}_1) - \gamma}{\sqrt{\mathrm{MSE}_1(\mathbf{x}_1)}} + \frac{\sqrt{N}E_1(\gamma_1 - \gamma_g)}{\sqrt{\mathrm{H}}}\right]^2 > \rho\right), \tag{3.55}$$

where γ_g is a guess of γ_1. Since γ_1 is bounded, the guess term $\left|\frac{\sqrt{N}E_1(\gamma-\gamma_g)}{\sqrt{\mathrm{H}}}\right|$ is also bounded. For a typical low SNR scenario, the energy transmitted tends to zero thereby making this term negligibly small. As a result, there is no loss of optimality due to the guess factor γ_g in the solution in (3.55). To demonstrate this concept, we plot the gain in MSE versus the error in the guess of γ for varying SNR in Fig. 3.9. The figure validates the fact that as SNR decreases, the error in the guess of γ plays a negligible role in the gain in MSE.

Performance under non negativity constraint

The performance of the one-step LS estimator of $\boldsymbol{\theta}$ was described in Section 3.3.1 without imposing the restriction that $\boldsymbol{\gamma}$ is a non-negative quantity. The LS solution $(\widehat{\boldsymbol{\gamma}}_1)$ allows for negative estimates of $\boldsymbol{\gamma}$. In practice, quadratic programming should be used to solve for $\boldsymbol{\gamma}$ when $\boldsymbol{\gamma} \geq 0$. For a single dense scatterer, the constrained solution $(\gamma_1 \geq 0)$ can be written as

$$\widehat{\gamma}_1^{(1),\mathrm{c}} = \widehat{\gamma}_1^{(1)} I\left(\widehat{\gamma}_1^{(1)} \geq 0\right). \tag{3.56}$$

Denote $\mathrm{MSE}_1^{\mathrm{c}}$ as the one-step MSE of the constrained LS estimator. $\mathrm{MSE}_1^{\mathrm{c}}$ can be evaluated as follows:

$$\begin{aligned}
\mathrm{MSE}_1^{\mathrm{c}} &= \mathrm{E}\left[\left|\widehat{\gamma}_1^{(1),\mathrm{c}} - \gamma_1\right|^2\right] \\
&= \mathrm{E}\left[\left|\left(\widehat{\gamma}_1^{(1)} - \gamma_1\right) I\left(\widehat{\gamma}_1^{(1)} \geq 0\right) - \gamma_1 I\left(\widehat{\gamma}_1^{(1)} < 0\right)\right|^2\right] \\
&= \mathrm{E}\left[\left|\left(\widehat{\gamma}_1^{(1)} - \gamma_1\right)\right|^2 I\left(\widehat{\gamma}_1^{(1)} \geq 0\right) + \gamma_1^2 I\left(\widehat{\gamma}_1^{(1)} < 0\right)\right] \\
&= \mathrm{MSE}_1\left(\mathrm{E}\left[\left(\frac{\widehat{\gamma}_1^{(1)} - \gamma_1}{\sqrt{\mathrm{MSE}_1}}\right)^2 I\left(\left(\frac{\widehat{\gamma}_1^{(1)} - \gamma_1}{\sqrt{\mathrm{MSE}_1}}\right) \geq \frac{-\gamma_1}{\sqrt{\mathrm{MSE}_1}}\right)\right]\right) \\
&\quad + \frac{\gamma_1^2}{\mathrm{MSE}_1} I\left(\left(\frac{\widehat{\gamma}_1^{(1)} - \gamma_1}{\sqrt{\mathrm{MSE}_1}}\right) < \frac{-\gamma_1}{\sqrt{\mathrm{MSE}_1}}\right) \qquad (3.57)
\end{aligned}$$

When N is large, the averaging associated with first estimate of γ drives the standardized MSE $\left(\frac{\widehat{\gamma}_1^{(1)}(\mathbf{y}_1) - \gamma_1}{\sqrt{\mathrm{MSE}_1(\mathbf{x}_1)}}\right)$ to asymptotically zero mean unit variance normal random variable, m_1.

$$\mathrm{MSE}_1^{\mathrm{c}} = \mathrm{MSE}_1 \mathrm{E}\left[m_1^2 I(m_1 \geq -s_1) + s_1^2 I(m_1 < -s_1)\right], \qquad (3.58)$$

where $s_1 = \frac{-\gamma_1}{\sqrt{\mathrm{MSE}_1}}$. Simplifying (3.58), we obtain

$$\mathrm{MSE}_1^{\mathrm{c}} = \mathrm{MSE}_1\left(1 - \frac{1}{\sqrt{2\pi}} s_1 \exp(-s_1^2/2) - Q(s_1) + s_1^2 Q(s_1)\right), \qquad (3.59)$$

where $Q(x)$ is the Q-function given by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt. \qquad (3.60)$$

At high SNR, $\mathrm{MSE}_1^{\mathrm{c}} \approx \mathrm{MSE}_1$, This follows from the fact that for N large, and high SNR, $s_1 \to \infty$, and hence

$$\begin{aligned}
\mathrm{MSE}_1^{\mathrm{c}} &= \lim_{s_1 \to \infty} \mathrm{MSE}_1\left(1 - \frac{1}{\sqrt{2\pi}} s_1 \exp(-s_1^2/2) - Q(s_1) + s_1^2 Q(s_1)\right) \\
&= \mathrm{MSE}_1. \qquad (3.61)
\end{aligned}$$

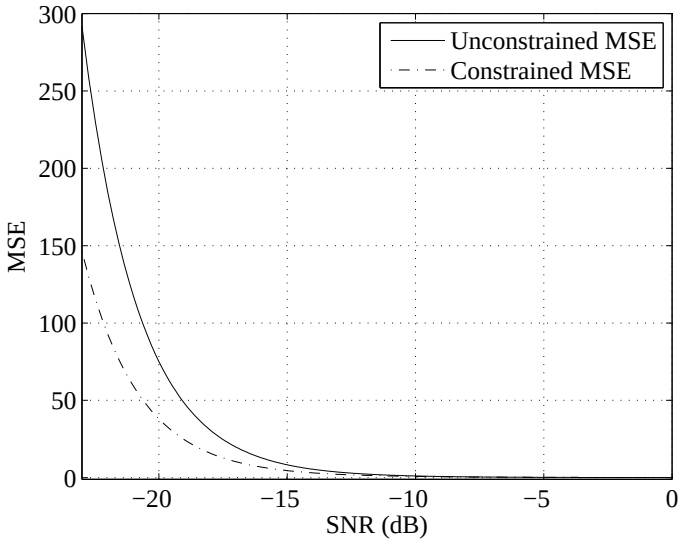


Figure 3.5: Plot of the one-step MSE as a function of SNR for the constrained and the unconstrained solutions.

Similarly, at low SNR, $s_1 \to 0$, and therefore

$$\begin{aligned} \mathrm{MSE}_1^{\mathrm{c}} &= \lim_{s_1 \to 0} \mathrm{MSE}_1 \left(1 - \frac{1}{\sqrt{2\pi}} s_1 \exp(-s_1^2/2) - Q(s_1) + s_1^2 Q(s_1) \right) \\ &= \frac{\mathrm{MSE}_1}{2} \end{aligned} \tag{3.62}$$

Thus the performance of the new estimator under the nonnegativity constraint is the same as the performance without the constraint at high SNR and there is a 3dB gain in performance at low SNR. The performance of the unconstrained and the constrained solutions is plotted as a function of SNR for the one-step strategy in Fig. 3.5. Using the same type of two-step design applied for the unconstrained case, we derive an expression for the two-step constrained MSE in Appendix 3.6. Further we show that the gain in this constrained case for low SNR is

$$\frac{\mathrm{MSE}_2^{\mathrm{c}}(\gamma_1)}{\mathrm{MSE}_1^{\mathrm{c}}(\gamma_1)} \approx 0.1263 \tag{3.63}$$

and all the above discussions regarding the optimal solution and the guess of γ approach can be directly extended to this constrained optimization.

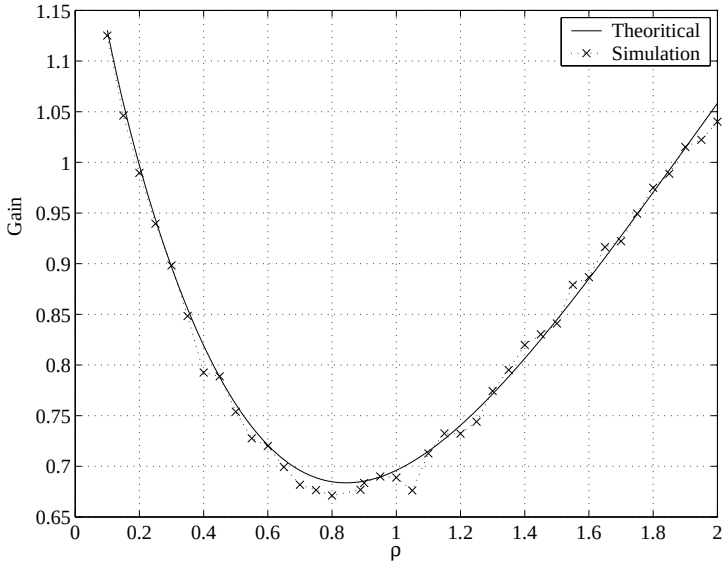


Figure 3.6: Gain(α_{opt}) vs. ρ. at SNR $= -10$dB.

3.4 Conclusions

The problem of imaging a Rayleigh scattering medium using an array of sensors rises in many applications. We obtained the MSE for the LS solution to the scatterer reflection powers of the dense scattering regions. For a two-step sequential design, we found the optimal transmission scheme that minimizes the MSE and proved that we can gain over conventional one-step strategies. The gains in MSE obtained analytically are verified through simulations. We also extended the results to the constrained optimization case. One strategy for extending this result to the case of estimating a vector of Rayleigh scattering magnitudes is to sequentially perform two-step procedures on each of the dense regions in space. We also intend to solve the problem of optimizing the transmitted spatial waveform rather than just looking at the energy allocation, a topic that is addressed in Chapter V of this dissertation. In addition, we need to generalize this approach from a two-step method to an iterative sequence of measurements.

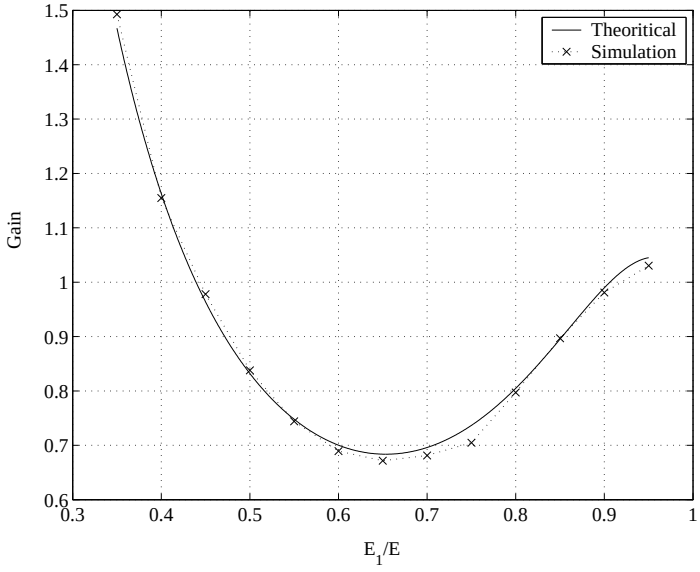


Figure 3.7: Gain(ρ_{opt}) vs. α. at SNR $= -10$dB.

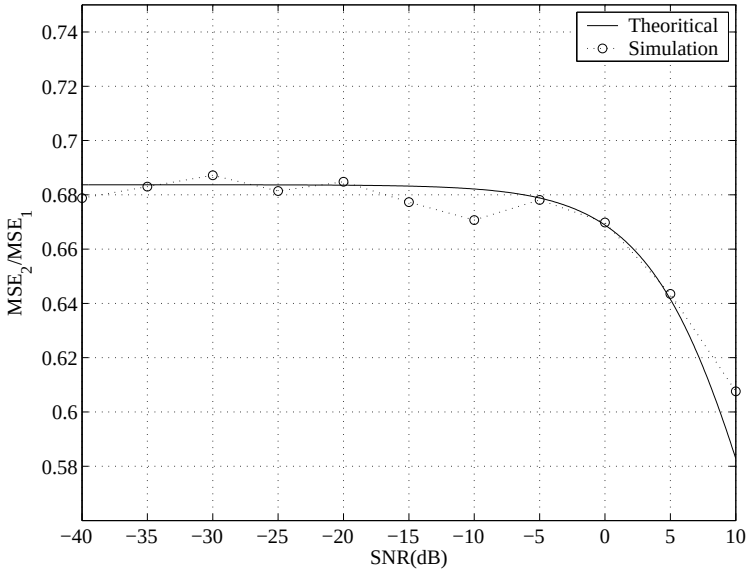


Figure 3.8: Gain vs. SNR for low SNR optimal $\alpha_{\text{opt}} \approx 0.66$ and $\rho_{\text{opt}} \approx 0.8885$.

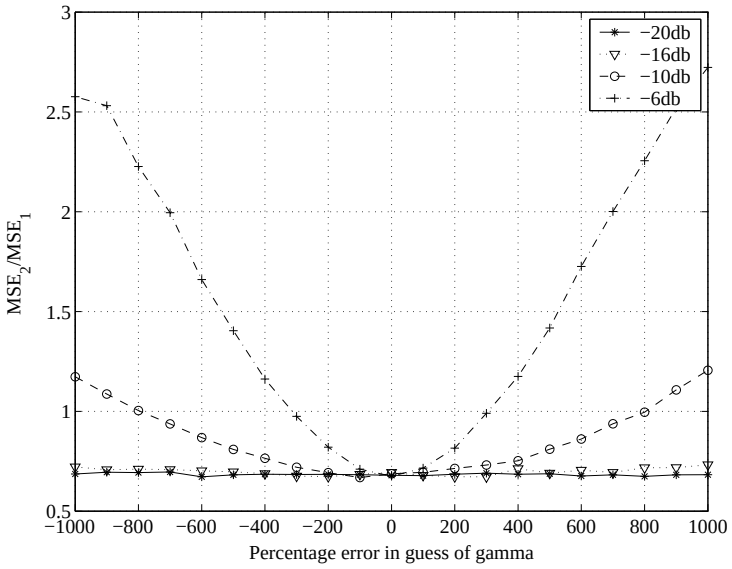


Figure 3.9: Gain vs. Percentage error in guess of γ_1 for varying SNR.

3.5 Appendix: time reversal imaging

In this appendix, we analyze the performance of time reversal imaging for a homogeneous medium with single scattering in the presence of noise under energy constraints. We use the Cramer Rao bound as a measure of performance and show that the optimal waveform design at both the steps of a two time step process is the optimal beamformer, i.e., time reversal does not provide any gains under these conditions. Future work would be to provide a similar analysis for multiple scattering in random media, as it is under these conditions that time reversal has its advantages.

Introduction

In the last decade, there have been many experimental and theoretical developments involving the concept of time reversal. In time reversal, an array of transducers receives signals from a localized source, time reverses it and retransmits into the medium to focus near the source [53]. It has been shown that the cross-range resolution is better in a random medium than in a homogeneous medium, both exper-

imentally [54,91,153] and in theory [10,21,47,116]. This is because the rich scattering in a noisy random medium tends to carry more information about the source and the phenomenon is termed as super-resolution of the time reversal process in random media [47].

The applications of time reversal can be broadly classified into two categories - focusing or the ability to localize energy on particular points (scatterers) of a target and time reversal imaging, i.e., in the context of recovering the target scattering characteristics. For the purposes of focusing energy using time reversal [76, 121, 122], the measured fields are time reversed and transmitted back into the medium to enable focusing, a term often referred to as physical time reversal. Iterative time reversal techniques [107] have also been used to achieve selective focusing. In [81], the concept of adaptive beamforming [37] and interference cancellation is used to focus energy at a specified location. This has led to extensive applications in ultrasonic nondestructive testing (finding cracks and defects) [77], medical therapy (detection of tumors and kidney stones) [163] and military applications (submarine detection, mines or objects buried under sediments) [70, 97, 160].

In computational time reversal imaging, the received signals are back propagated into a fictitious medium to form the image of a random scattering medium [20]. The fictitious medium usually tends to be a homogeneous medium with constant sound speed where the channel characteristics are completely known. This computational time reversal process is also termed as migration, or back propagation in geophysics [17, 32], and in x-ray crystallography respectively [113].

Since back propagation into a homogeneous medium in general is not the optimal solution, researchers have focused attention on imaging a random media. It is shown that the random Green's function can be replaced by its expectation because of the

self averaging property of time reversal [19, 116]. Using the above property and the moment formula [69, 137], the random Green's function is approximated through the Green's function of the homogeneous medium and the effective aperture, and this result is used to perform interferometric imaging in [20]. In [27], wideband time reversal imaging of an elastic target is performed in an acoustic waveguide by modeling the unknown random Green's function using a set of parameters and determining them via a genetic algorithm by employing a cost function obtained through the quality of the time reversal image.

While there has been much study of the performance of time reversal retrofocusing methods for both deterministic and random scattering environments, the performance of time reversal imaging methods in the presence of transducer noise and unknown transducer position has not been thoroughly studied. Uncalibrated arrays are especially a problem for radar applications of TRM due to small illumination wavelengths, amplifying positional uncertainty and distorting the estimated scatter field. Additive noise introduced at the receiver before and after time reversal will also degrade TRM performance by increasing estimator variance. The time reversal probing method considered here takes a double noise hit due to the combination of receiver noise seeping into the transmitted time reversed signal and also into the received backscatter signal. In this report we systematically study the effect of such noise and sensor calibration factors on the imaging performance of TRM and compare the same with the conventional beam forming methods under the constraint that the energy transmitted in the two cases is the same. The imaging performance is measured by the Cramèr-Rao Lower Bounds (CRB) on variance of any unbiased scatter-coefficient estimator. The CRB permits quantitative comparisons of TRM to conventional imaging methods independently of any specific estimation or image

reconstruction algorithm.

Mathematical description of time reversal probing

We assume the Born approximation channel model which ignores the effect of multiple scattering. Furthermore, we assume that the scattering coefficients are a set of deterministic parameters to be estimated. Hence the channel can be written as $\mathbf{H}_{ch} = \mathbf{H}\mathbf{D}\mathbf{H}^T$, where $\mathbf{H}$ is the homogeneous Green's function presented in Section 3.2. The set of received signals for time reversal probing is similar to the set of equations presented in Section 3.2 except that the transmit signal at the two time steps $\mathbf{x}_1$ and $\mathbf{x}_2(\mathbf{y}_1)$ have a specified structure as a result of time reversal probing. The time reversal probing mechanism for imaging of the scatter cross section in each voxel follows a 4-step process.

Step 1: The transducers first send $\mathbf{x}_1$, the projection of the signal $\mathbf{z}$ onto the beam-steering matrix $\mathbf{B}_1$, i.e., $\mathbf{x}_1 = \mathbf{B}_1\mathbf{z}$.

Step 2: Next, the transducers receive the backscattered signal, after it has traveled through the channel $\mathbf{H}_{ch}$,

$$\mathbf{y}_1 = \mathbf{H}_{ch}\mathbf{x}_1 + \mathbf{n}_1 = \mathbf{H}\mathbf{D}\mathbf{H}^T\mathbf{B}_1\mathbf{z} + \mathbf{n}_1. \tag{3.64}$$

Step 3: Next, the transducers time-reverse the signal, amplify it by a factor a, project it onto the beam-steering vector, and re-transmit, i.e.,

$$\mathbf{x}_2 = \mathbf{B}_2\mathbf{y}_1^* = \mathbf{B}_2\mathbf{H}^*\mathbf{D}^*\mathbf{H}^H\mathbf{B}_1^*\mathbf{z}^* + \mathbf{B}_2\mathbf{n}_1^*. \tag{3.65}$$

Step 4: Finally, the transducers receive the backscattered, time-reversed signal $\mathbf{y}_2$, where

$$\mathbf{y}_2 = \mathbf{H}_{ch}\mathbf{x}_2 + \mathbf{n}_2 = \mathbf{H}\mathbf{D}\mathbf{H}^T\mathbf{B}_2\mathbf{H}^*\mathbf{D}^*\mathbf{H}^H\mathbf{B}_1^*\mathbf{z}^* + \mathbf{H}\mathbf{D}\mathbf{H}^T\mathbf{B}_2\mathbf{n}_1^* + \mathbf{n}_2. \tag{3.66}$$

The noises $\mathbf{n}_1, \mathbf{n}_2$ are i.i.d complex normal random vectors with zero mean and a covariance matrix $\sigma^2\mathbf{I}$.

3.5.1 CRB for scattering coefficients

How well does probing the medium with the time reversal mechanism described above permit the estimation of $\mathbf{d}$? We use the CRB to generate lower bounds on the variance of any unbiased estimators which use the measurements described above. First, we consider the bounds on estimation of $\mathbf{d}$ via the measurement of $\mathbf{y}_1$ to provide a baseline of the possible performance of a non-time-reversal measurement method. Then, we calculate the bound on estimation of $\mathbf{d}$ via the measurement of $\mathbf{y}_2$, ie., the performance using time reversal. We consider the case of calibrated arrays, i.e., known transducer positions, in this section.

The FIM (CRB^{-1}) for a set of parameters $\boldsymbol{\theta}$ given a vector of complex normal observations with mean $\boldsymbol{\mu}(\boldsymbol{\theta})$ and covariance matrix $\mathbf{R}(\boldsymbol{\theta})$ is given by [75],

(3.67)
$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = 2\mathrm{Re}\left\{\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta})^H}{\partial\theta_i}\mathbf{R}^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta})}{\partial\theta_j}\right\} + \mathrm{tr}\left[\mathbf{R}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{R}(\boldsymbol{\theta})}{\partial\theta_i}\mathbf{R}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{R}(\boldsymbol{\theta})}{\partial\theta_j}\right],$$

where $\mathrm{Re}\{\cdot\}$ indicates the real operator and $\mathbf{I}(\boldsymbol{\theta})$ is the $V \times V$ Fisher Information Matrix. The FIM above is defined for real parameters while the parameters $\mathbf{d}$ are in general complex. If the FIM for the set of real and imaginary parameters, $\boldsymbol{\theta} = [\mathbf{d}_{\mathrm{Re}}, \mathbf{d}_{\mathrm{Im}}]^T$, is denoted by $2V \times 2V$ matrix $\begin{bmatrix} \mathrm{Re}(\mathbf{M}) & \mathrm{Im}(\mathbf{M}) \\ -\mathrm{Im}(\mathbf{M}) & \mathrm{Re}(\mathbf{M}) \end{bmatrix}$, then the FIM for the set of complex parameters is given by $2\mathbf{M}$ [75]. In the following section, we compute the CR lower bounds of scatter coefficients and attempt to find the best spatial beam forming matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ which minimizes the CRB for a general beam forming and TR methods.

Optimal beam steering for calibrated probing without time reversal.

Since both $\mathbf{B}_1$ and $\mathbf{z}$ are in our control, it is enough to optimize over a beam steering vector $\mathbf{x}_1 = \mathbf{B}_1\mathbf{z}$.

Considering the measurement $\mathbf{y}_1$ as given in (3.64), the distribution of $\mathbf{y}_1$ is

$$\mathbf{y}_1 \sim \mathcal{CN}\left(\mathbf{HDH}^T\mathbf{x}_1, \sigma^2\mathbf{I}\right). \tag{3.68}$$

Then the derivative of the mean $\boldsymbol{\mu}_1$ is given by

$$\frac{\partial \boldsymbol{\mu}_1(\boldsymbol{\theta})}{\partial d_{i\mathrm{Re}}} = \mathbf{H}\mathbf{e}_i\mathbf{e}_i^T\mathbf{H}^T\mathbf{x}_1 \tag{3.69}$$

where $\mathbf{e}_i$ is vector with all zeros except for a one in the i^{th} place. The FIM_1 for the real part of the scattering coefficients would be

$$\begin{aligned}\mathrm{Re}(\mathbf{M})_{i,j} &= \frac{2}{\sigma^2}\mathrm{Re}\left\{\frac{\partial \boldsymbol{\mu}_1(\boldsymbol{\theta})}{\partial d_{i\mathrm{Re}}}^H \frac{\partial \boldsymbol{\mu}_1(\boldsymbol{\theta})}{\partial d_{j\mathrm{Re}}}\right\} \\ &= \frac{2}{\sigma^2}\mathrm{Re}\left\{\mathbf{e}_i^T\mathbf{H}^H\mathbf{x}_1^*(\mathbf{e}_i^T\mathbf{H}^H\mathbf{H}\mathbf{e}_j)\mathbf{e}_j^T\mathbf{H}^T\mathbf{x}_1\right\}.\end{aligned}$$

The FIM and CRB for the complex parameters is then given by

$$\mathrm{FIM}_1 = \frac{4}{\sigma^2}\left\{\mathrm{diag}^*(\mathbf{H}^T\mathbf{x}_1)(\mathbf{H}^H\mathbf{H})\mathrm{diag}(\mathbf{H}^T\mathbf{x}_1)\right\}, \tag{3.70}$$

$$\mathrm{CRB}_1 = \frac{\sigma^2}{4}\left\{\mathrm{diag}^{-1}(\mathbf{H}^T\mathbf{x}_1)(\mathbf{H}^H\mathbf{H})^{-1}\mathrm{diag}^{-1}(\mathbf{H}^T\mathbf{x}_1)^*\right\} \tag{3.71}$$

Our goal is to find the best $\mathbf{x}_1 = \mathbf{v}$ that will minimize the CRB for the scatter coefficients. Since the minimization of the CRB over $\mathbf{v}$ has no closed form solution, we find a suboptimal solution by maximizing the trace of the fisher information matrix FIM_1. The trace of the FIM is

$$\mathrm{tr}(\mathrm{FIM}_1) = \sum_{i=1}^{V}\mathrm{FIM}_{1\,i,i} = \frac{4}{\sigma^2}\sum_{i=1}^{V}|\mathbf{e}_i^T\mathbf{H}^T\mathbf{v}|^2|\mathbf{H}^H\mathbf{H}|_{i,i}. \tag{3.72}$$

We maximize the above expression of the $\mathrm{tr}(\mathrm{FIM}_1)$ subject to the constraint on the energy transmitted, i.e., $\|\mathbf{B}_1\mathbf{z}\|^2 = \|\mathbf{v}\|^2 \leq E$ where E is the total energy.

Solving using Lagrange multipliers, the normalized optimal $\mathbf{v}$, $\mathbf{v}_{\text{opt}}$ is the eigen vector corresponding to the maximum eigen value of the matrix $\widehat{\mathbf{H}} = \sum_{i=1}^{V} \mathbf{h}_{yi}^{*}\mathbf{h}_{yi}^{T}\|\mathbf{h}_{yi}\|^2$. The optimal FIM_1 is then given by

$$\text{FIM}_1 = \frac{4E}{\sigma^2}\text{diag}^{*}(\mathbf{H}^T\mathbf{v}_{\text{opt}})(\mathbf{H}^H\mathbf{H})\text{diag}(\mathbf{H}^T\mathbf{v}_{\text{opt}}). \tag{3.73}$$

Optimal FIM for calibrated probing using time reversal : bi-static case

In a bi-static case, we observe both the received signals $\mathbf{y}_1$ and $\mathbf{y}_2$. We need to obtain the best possible $\mathbf{v}$ and $\mathbf{B}_2$ for the new Fisher Information Matrix given by,

$$\begin{aligned}
\text{FIM}_2 &= \text{FIM}_{\mathbf{y}_1,\mathbf{y}_2}(\boldsymbol{\theta}) \\
&= -\text{E}\left[\nabla^2 \log\, f(\mathbf{y}_1,\mathbf{y}_2|\boldsymbol{\theta})\right] \\
&= -\text{E}\left[\nabla^2 \left(\log\, f(\mathbf{y}_1|\boldsymbol{\theta}) + \log\, f(\mathbf{y}_2|\mathbf{y}_1,\boldsymbol{\theta})\right)\right] \\
&= -\text{E}\left[\nabla^2 \log\, f(\mathbf{y}_1|\boldsymbol{\theta})\right] - \text{E}_{\mathbf{y}_2,\mathbf{y}_1}\left[\nabla^2 \log\, f(\mathbf{y}_2|\mathbf{y}_1,\boldsymbol{\theta})\right] \\
&= \text{FIM}_{1_{\mathbf{y}_1}}(\boldsymbol{\theta}) - \text{E}_{\mathbf{y}_1}\left[\text{E}_{\mathbf{y}_2|\mathbf{y}_1}\left[\nabla^2 \log\, f(\mathbf{y}_2|\mathbf{y}_1,\boldsymbol{\theta})\right]\right] \\
&= \text{FIM}_{1_{\mathbf{y}_1}}(\boldsymbol{\theta}) + \text{E}_{\mathbf{y}_1}\left[\text{FIM}_{\mathbf{y}_2|\mathbf{y}_1}(\boldsymbol{\theta})\right],
\end{aligned}$$

where $\mathbf{y}_1 \sim \mathcal{CN}(\mathbf{HDH}^T\mathbf{x}_1, \sigma^2\mathbf{I})$ and $\mathbf{y}_2|\mathbf{y}_1 \sim \mathcal{CN}(\mathbf{HDH}^T\mathbf{B}_2\mathbf{y}_1^{*}, \sigma^2\mathbf{I})$.

We only need to optimize the second term for the best possible $\mathbf{B}_2$ subject to an energy constraint as the first term is independent of $\mathbf{B}_2$. We once again maximize the trace of the Fisher Information Matrix over all possible $\mathbf{B}_2$, i.e., maximize the trace,

$$\text{tr}\left(\text{E}_{\mathbf{y}_1}\left[\text{FIM}_{1_{\mathbf{y}_2|\mathbf{y}_1}}\right]\right) = \frac{4}{\sigma^2}\sum_{i=1}^{V}\text{E}_{\mathbf{y}_1}\left[|\mathbf{H}^T\mathbf{B}_2\mathbf{y}_1^{*}|_i^2(\mathbf{H}^H\mathbf{H})_{i,i}\right] \tag{3.74}$$

subject to the energy constraints $\text{E}\left[\|\mathbf{B}_2\mathbf{y}_1^{*}\|^2\right] \leq E_2$, $\|\mathbf{v}\|^2 \leq E_1$ and $E_1 + E_2 = E$ where E_1 and E_2 are the energy used in the first and second transmissions. Using

Lagrange multipliers and setting the derivative with respect to $\mathbf{B}_2$ to zero, we get

$$\left(\frac{4}{\sigma^2}\sum_{i=1}^{V}\mathbf{H}^*\mathbf{e}_i\mathbf{e}_i^T\mathbf{H}^T(\mathbf{H}^T\mathbf{H}^*)_{i,i}-\lambda_2^*\mathbf{I}\right)\mathbf{B}_2 \;=\; 0, \tag{3.75}$$

where the energy constraint is $E_2 = \mathrm{tr}\left(\mathbf{B}_2^H\mathbf{B}_2(\mathrm{E}_{\mathbf{y}_1}\left[\mathbf{y}_1^*\mathbf{y}_1^T\right])\right)$ and λ_2 is the Lagrange multiplier variable.

Using the above equations, we can generate a rank one solution to $\mathbf{B}_2$ given by $\mathbf{B}_2 = c_2\mathbf{v}_2\mathbf{v}_2^H$ where $\|\mathbf{v}_2\|^2 = 1$. Substituting for $\mathbf{B}_2$ back in equation (3.75) and in the energy constraint, we get the optimal $\mathbf{v}_2 = \mathbf{v}_{\mathrm{opt}}$ and $c_2 = \frac{E_2}{\sigma^2}$. Thus the Fisher Information Matrix using time reversal is

$$\begin{aligned}
\mathrm{FIM}_2 &= \mathrm{FIM}_{1_{\mathbf{y}_1}}(\boldsymbol{\theta}) + \mathrm{E}_{\mathbf{y}_1}\left[\mathrm{FIM}_{\mathbf{y}_2|\mathbf{y}_1}(\boldsymbol{\theta})\right] \\
&= \frac{4}{\sigma^2}\left\{\mathrm{diag}^*(\mathbf{H}^T\mathbf{x}_1)(\mathbf{H}^H\mathbf{H})\mathrm{diag}(\mathbf{H}^T\mathbf{x}_1) + \right. \\
&\qquad \left. |c_2|^2\mathrm{E}_{\mathbf{y}_1}\left[|\mathbf{v}_2^H\mathbf{y}_1^*|^2\mathrm{diag}^*(\mathbf{H}^T\mathbf{v}_2)(\mathbf{H}^H\mathbf{H})\mathrm{diag}(\mathbf{H}^T\mathbf{v}_2)\right]\right\} \\
&= \frac{4}{\sigma^2}\left\{\mathrm{diag}^*(\mathbf{H}^T\mathbf{x}_1)(\mathbf{H}^H\mathbf{H})\mathrm{diag}(\mathbf{H}^T\mathbf{x}_1) + E_2\ \mathrm{diag}^*(\mathbf{H}^T\mathbf{v}_{\mathrm{opt}})(\mathbf{H}^H\mathbf{H})\mathrm{diag}(\mathbf{H}^T\mathbf{v}_{\mathrm{opt}})\right\}.
\end{aligned}$$

FIM_2 depends on $\mathbf{x}_1$ only through the first term and hence the optimal solution to $\mathbf{x}_1$ is the same as before, i.e.,$\mathbf{x}_1 = \sqrt{E_1}\mathbf{v}_{\mathrm{opt}}$. Hence

$$\begin{aligned}
\mathrm{FIM}_2 &= \frac{4(E_1+E_2)}{\sigma^2}\mathrm{diag}^*(\mathbf{H}^T\mathbf{v}_{\mathrm{opt}})(\mathbf{H}^H\mathbf{H})\mathrm{diag}(\mathbf{H}^T\mathbf{v}_{\mathrm{opt}}) \\
&= \frac{4E}{\sigma^2}\mathrm{diag}^*(\mathbf{H}^T\mathbf{v}_{\mathrm{opt}})(\mathbf{H}^H\mathbf{H})\mathrm{diag}(\mathbf{H}^T\mathbf{v}_{\mathrm{opt}}) \qquad (3.76) \\
&= \mathrm{FIM}_1 \qquad (3.77)
\end{aligned}$$

This analysis proves that beam forming alone is optimal with regards to maximizing the trace of the FIM for the single scattering model. Nevertheless it would be worthwhile to study the optimality while minimizing the CRB for multiple scattering models.

3.6 Appendix: two-step sequential strategy, constrained case

The constrained solution for a two-step procedure is given by

$$\widehat{\gamma}_1^{(2),\mathrm{c}} = \frac{w_1\widehat{\gamma}_1^{(1),c}(\mathbf{y}_1) + w_2\widehat{\gamma}_1^{(1),c}(\mathbf{y}_2)}{w_1 + w_2}, \tag{3.78}$$

where $\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_1)$ and $\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_2)$ are the constrained LS estimators of γ_1 obtained from each of the two-steps, i.e.,

$$\begin{aligned}
\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_1) &= \widehat{\gamma}_1^{(1)}(\mathbf{y}_1) I\left(\widehat{\gamma}_1^{(1)}(\mathbf{y}_1) \geq 0\right) \\
\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_2) &= \widehat{\gamma}_1^{(1)}(\mathbf{y}_2) I\left(\widehat{\gamma}_1^{(1)}(\mathbf{y}_2) \geq 0\right).
\end{aligned}$$

The corresponding MSE of the constrained two-step estimator can be evaluated as

$$\begin{aligned}
\mathrm{MSE}_2^{\mathrm{c}} &= \mathrm{E}\left[\left(\widehat{\gamma}_1^{(2),\mathrm{c}} - \gamma_1\right)^2\right] \\
&= \mathrm{E}_{\mathbf{y}_1}\left[\mathrm{E}_{\mathbf{y}_2|\mathbf{y}_1}\left[\left(\widehat{\gamma}_1^{(2),\mathrm{c}} - \gamma_1\right)^2\right]\right] \\
&= \mathrm{E}_{\mathbf{y}_1}\left[\mathrm{MSE}_2^{\mathrm{c}}|\mathbf{y}_1\right], \qquad (3.79)
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{MSE}_2^{\mathrm{c}}|\mathbf{y}_1 &= \frac{w_1^2(\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_1) - \gamma_1)^2 + w_2^2(\mathrm{MSE}_1^{\mathrm{c}}(\mathbf{x}_2(\mathbf{y}_1)))}{(w_1 + w_2)^2} + \\
&\qquad \frac{2w_1w_2(\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_1) - \gamma_1)\mathrm{E}_{\mathbf{y}_2|\mathbf{y}_1}\left[(\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_2) - \gamma_1)\right]}{(w_1 + w_2)^2}. \qquad (3.80)
\end{aligned}$$

The bias term in the expression for $\mathrm{MSE}_2^{\mathrm{c}}|\mathbf{y}_1$ in (3.80) is

$$\begin{aligned}
b_2 &= \mathrm{E}_{\mathbf{y}_2|\mathbf{y}_1}\left[(\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_2) - \gamma_1)\right] \\
&= \mathrm{E}_{y_2|y_1}\left[(\widehat{\gamma}_1^{(1)}(\mathbf{y}_2) - \gamma_1) I\left(\widehat{\gamma}_1^{(1)}(\mathbf{y}_2) \geq 0\right) - \gamma_1 I\left(\widehat{\gamma}_1^{(1)}(\mathbf{y}_2) < 0\right)\right] \\
&= \mathrm{E}_{y_2|y_1}\left[\sqrt{\mathrm{MSE}_1(E_2)}\frac{\widehat{\gamma}_1^{(1)}(\mathbf{y}_2) - \gamma_1}{\sqrt{\mathrm{MSE}_1(E_2)}} I\left(\frac{\widehat{\gamma}_1^{(1)}(\mathbf{y}_2) - \gamma_1}{\sqrt{\mathrm{MSE}_1(E_2)}} \geq -\frac{\gamma}{\sqrt{\mathrm{MSE}_1(E_2)}}\right)\right. \\
&\qquad \left. -\gamma_1 I\left(\frac{(\widehat{\gamma}_1^{(1)}(\mathbf{y}_2) - \gamma_1)}{\sqrt{\mathrm{MSE}_1(E_2)}} < -\frac{\gamma_1}{\sqrt{\mathrm{MSE}_1(E_2)}}\right)\right], \qquad (3.81)
\end{aligned}$$

where $\mathrm{MSE}_1(E_2)$ is the MSE of the second-step LS estimator which can be evaluated using the expression in (3.20) for a fixed waveform $\mathbf{v}$ at energy E_2. The expression $\frac{\widehat{\gamma}_1^{(1)}(\mathbf{y}_2)-\gamma_1}{\sqrt{\mathrm{MSE}_1(E_2)}}$ can be approximated by a zero mean unit variance random variable for sufficiently large number of observations denoted by m_2. Then the bias term can be computed as

$$\begin{aligned} b_2 &= \sqrt{\mathrm{MSE}_1(E_2)}\,(m_2 I(m_2 \geq s_2) - s_2 I(m_2 < s_2)) \\ &= \sqrt{\mathrm{MSE}_1(E_2)}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-s_2\right) - s_2 Q(s_2)\right), \end{aligned} \tag{3.82}$$

where $s_2 = \gamma_1/\sqrt{\mathrm{MSE}_1(E_2)}$. As in the sequential design case, we choose weights w_1 and w_2 proportional to $1/\left(\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_1) - \gamma_1\right)^2$ and $1/\mathrm{MSE}_1^{\mathrm{c}}(\mathbf{x}_2(\mathbf{y}_1))$ respectively. Substituting the expression for the weights in (3.80) and evaluating $\mathrm{MSE}_2^{\mathrm{c}}$, we obtain

$$\begin{aligned} \mathrm{MSE}_2^{\mathrm{c}} &= \mathrm{E}\left[\frac{\left(\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_1) - \gamma_1\right)^2 \mathrm{MSE}_1^{\mathrm{c}}(\mathbf{x}_2(\mathbf{y}_1))}{\left(\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_1) - \gamma_1\right)^2 + \mathrm{MSE}_1^{\mathrm{c}}(\mathbf{x}_2(\mathbf{y}_1))}\right] \\ &+ \mathrm{E}\left[\frac{2b_2\left(\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_1) - \gamma_1\right)^2 \mathrm{MSE}_1^{\mathrm{c}}(\mathbf{x}_2(\mathbf{y}_1))}{\left(\left(\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_1) - \gamma_1\right)^2 + \mathrm{MSE}_1^{\mathrm{c}}(\mathbf{x}_2(\mathbf{y}_1))\right)^2}\right]. \end{aligned} \tag{3.83}$$

Let $m_{1\mathrm{c}} = \frac{\widehat{\gamma}_1^{(1),\mathrm{c}}(\mathbf{y}_1)-\gamma_1}{\sqrt{\mathrm{MSE}_1(E_1)}}$. Then $\mathrm{MSE}_2^{\mathrm{c}}$ can be rewritten as

$$\begin{aligned} \mathrm{MSE}_2^{\mathrm{c}} &= \mathrm{E}\left[\frac{m_{1\mathrm{c}}^2 \mathrm{MSE}_1(E_1)\mathrm{MSE}_1^{\mathrm{c}}(E_2)}{m_{1\mathrm{c}}^2\mathrm{MSE}_1(E_1) + \mathrm{MSE}_1^{\mathrm{c}}(E_2)}\right] \\ &+ \mathrm{E}\left[\frac{2b_2 m_{1\mathrm{c}}^2 \mathrm{MSE}_1(E_1)\mathrm{MSE}_1^{\mathrm{c}}(E_2)}{\left(m_{1\mathrm{c}}^2\mathrm{MSE}_1(E_1) + \mathrm{MSE}_1^{\mathrm{c}}(E_2)\right)^2}\right], \end{aligned} \tag{3.84}$$

where $\mathrm{MSE}_1^{\mathrm{c}}(\mathbf{x}_2(\mathbf{y}_1))$ is replaced by $\mathrm{MSE}_1^{\mathrm{c}}(E_2)$ since the energy of transmit signal $\mathbf{x}_2(\mathbf{y}_1)$ is $E_2(\mathbf{y}_1)$. At low SNR, we proved in Section 3.3.3 that $\mathrm{MSE}_1(E_1) \approx \frac{\mathrm{H}}{NE_1^2}$. Further we showed earlier in this section that the MSE of the constrained estimator at low SNR is $\mathrm{MSE}_1^{\mathrm{c}}(E_2) \approx \mathrm{MSE}_1(E_2)/2 \approx \frac{\mathrm{H}}{2NE_2^2}$. Similar to the two-step sequential design for unconstrained estimation, we look for a suboptimal solution to the

energy distribution of the form $E_2 = AI(m_{1c}^2 > \rho)$. Substituting the expressions for $\mathrm{MSE}_1(E_1)$, $\mathrm{MSE}_1^c(E_2)$, and E_2 in the expression for MSE_2^c in (3.84), we obtain

$$\begin{aligned}
\mathrm{MSE}_2^c(E_1,\rho) &= \frac{\mathrm{H}}{2NE_0^2}\mathrm{E}\left[\left\{\frac{2E_0^2m_{1c}^2}{E_1^2+2A^2m_{1c}^2}+\frac{4E_0^2E_1Ab_2m_{1c}^3}{\left(E_1^2+2A^2m_{1c}^2\right)^2}\right\}I(m_{1c}^2>\rho)\right. \\
&\qquad \left.+\frac{2E_0^2m_{1c}^2}{E_1^2}I(m_{1c}^2<\rho)\right] \\
&= \mathrm{MSE}_1^c(E_0)\left(2E_0^2\mathrm{E}\left[\frac{m_{1c}^2}{E_1^2+2A^2m_{1c}^2}I(m_{1c}^2>\rho)\right]\right. \\
&\qquad +4E_0^2E_1Ab_2\mathrm{E}\left[\frac{m_{1c}^3}{\left(E_1^2+2A^2m_{1c}^2\right)^2}I(m_{1c}^2>\rho)\right] \\
&\qquad \left.+\frac{2E_0^2}{E_1^2}\mathrm{E}\left[m_{1c}^2I(m_{1c}^2<\rho)\right]\right). \qquad (3.85)
\end{aligned}$$

The distribution of m_{1c} is given by

$$P(m_{1c}<x)=\begin{cases}1-Q(x) & x<s_1\\ Q(x) & x=-s_1\\ 0 & \text{o.w}\end{cases}$$

The expression for $\mathrm{MSE}_2^c(E_1,\rho)$ in (3.85) can be further simplified using the results from integral calculus provided in Appendix 3.8 and minimum error is obtained by numerically evaluating the expression on a grid of E_1 and ρ. Then the optimal solution is given by

$$\mathrm{MSE}_{2,\ \mathrm{opt}}^c \approx 0.1215\mathrm{MSE}_1^c. \qquad (3.86)$$

3.7 Appendix: results from matrix theory

Let $\Delta\mathbf{R}_x = \frac{1}{Y}\sum_{i=1}^{Y}\mathbf{x}_i\mathbf{x}_i^H - \mathbf{R}_x$ and $\Delta\mathbf{R}_y = \frac{1}{Y}\sum_{i=1}^{Y}\mathbf{y}_i\mathbf{y}_i^H - \mathbf{R}_y$ where $\mathbf{x}_i$ and $\mathbf{y}_i$ are zero mean complex normal i.i.d with covariance matrices $\mathbf{R}_x$ and $\mathbf{R}_y$ respectively. Furthermore, $\mathrm{E}[\mathbf{x}_i\mathbf{y}_i] = \mathbf{R}_{xy}$ and $\mathrm{E}[\mathbf{x}_i\mathbf{y}_j] = 0$. Then the following results hold:

$$\mathrm{E}\left[\Delta\mathbf{R}_x\mathbf{M}\Delta\mathbf{R}_y\right] = \frac{1}{Y}\mathbf{R}_{xy}\mathrm{tr}\left\{\mathbf{M}\mathbf{R}_{xy}\right\}, \qquad (3.87)$$

$$\mathrm{E}\left[\mathrm{tr}\{\mathbf{A}\Delta\mathbf{R}_x\}\mathrm{tr}\{\mathbf{B}\Delta\mathbf{R}_y\}\right] = \frac{1}{Y}\mathrm{tr}\{\mathbf{A}\mathbf{R}_{xy}\mathbf{B}\mathbf{R}_{xy}\}. \qquad (3.88)$$

3.8 Appendix: result from integral calculus

An expression for the following indefinite integral is given below:

$$\int \frac{x^3 \exp(-x^2)dx}{(1+a^2x^2)^2} = \frac{1}{2}\frac{\exp(-x^2)}{a^4(a^2x^2+1)} - \frac{1}{2}\frac{(a^2+1)\exp(\frac{1}{a^2})E_i(1, x^2+\frac{1}{a^2})}{a^6},$$

where $E_i(1,x)$ is the exponential integral given by

$$E_i(1,x) = \int_x^\infty \frac{\exp(-t)}{t}dt \quad , \; x > 0. \tag{3.89}$$

CHAPTER IV

Energy allocation for detection in linear models

In Chapters II and III, we derived optimal energy allocation strategies for estimation under two different channel models and showed more than 5dB improvement in performance when compared to non adaptive procedures. In this chapter, we pose a similar problem of optimal energy allocation for detection. Like in Chapter II, we consider a linear Gaussian channel model and solve the two-step energy allocation problem under the frequentist and Bayesian approaches. We show that such a two-step approach yields a performance improvement of atleast 2dB over non adaptive methods.

4.1 Introduction

The problem of detection of signals in noise has been studied in depth for many years now. Signal detection is formulated as a hypothesis testing problem in the theory of statistical inference [120]. In binary hypothesis testing, a decision is to be made on the presence (H_1) or absence of a target (H_0) based on a set of observations. In a parametric setting, the observation X has a probability density function $f(x;\theta)$ based on the set of unknown parameters $\theta \in \Theta$. Under each hypothesis $\mathrm{H}_i, i = 1, 2$, the parameter θ belongs to the set Θ_i, where $\Theta_1 \cap \Theta_2 = \phi$ and $\Theta_1 \cup \Theta_2 = \Theta$. The H_0 hypothesis is commonly referred to as the null hypothesis, since it refers to the

absence of target and H_1 is the alternative hypothesis. We can express the problem as testing the two hypotheses

$$\begin{aligned} \mathrm{H}_0 : \theta &\in \Theta_0, X \sim f(x;\theta) \\ \mathrm{H}_1 : \theta &\in \Theta_1, X \sim f(x;\theta). \end{aligned}$$

The hypotheses are said to be simple when the sets Θ_0 and Θ_1 are singleton, i.e., $\Theta_0 = \theta_0$ and $\Theta_1 = \theta_1$. If either of the sets have more than one element, then the hypotheses are said to be composite. Associated with these hypotheses are three main cost functions: probability of false alarm (P_{FA}) (deciding H_1 when H_0 is the truth), probability of miss (P_M) (deciding H_0 when H_1 is the truth), and probability of detection (P_D) (deciding H_1 when H_1 is the truth).

There are two main approaches to hypothesis testing: the frequentist approach and the Bayesian approach. In the Bayesian strategy, θ is assumed to be a random variable and a prior $f(\theta)$ is assigned for θ. The prior probabilities on H_0 and H_1 is given by

$$P(\mathrm{H}_i) = P(\theta \in \Theta_i) = \int_{\Theta_i} f(\theta) d\theta, \quad i = 1, 2. \tag{4.1}$$

There is a cost assigned for making wrong decisions and the objective in the Bayesian approach is to find a strategy to minimize the average cost. For a special case of minimizing average probability of error, the cost function P_e can be written as

$$P_e = P_{FA} P(\mathrm{H}_0) + P_M P(\mathrm{H}_1). \tag{4.2}$$

The optimal decision rule in this case is the Bayes Likelihood Ratio test (BLR) given by

$$\Lambda_{\mathrm{BLT}} : \frac{f(x|\mathrm{H}_1)}{f(x|\mathrm{H}_0)} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \eta, \tag{4.3}$$

where η is the optimal Bayesian threshold

$$\eta = \frac{P(\mathrm{H}_0)}{P(\mathrm{H}_1)}. \tag{4.4}$$

Rearranging terms in (4.3), we obtain

$$\frac{P(\mathrm{H}_1|x)}{P(\mathrm{H}_0|x)} \gtrless 1, \tag{4.5}$$

which is the maximum a posteriori test or the MAP rule. The Bayesian approach only ensures best average performance with respect to the selected prior and provides no guaranteed protection against false alarm and miss.

A frequentist approach to detection is to find a strategy to maximize probability of detection (P_D) subject to a fixed false alarm level. A test is said to be of false alarm level $\alpha \in [0, 1]$ if

$$\max_{\theta \in \Theta_0} P_{FA}(\theta) \leq \alpha. \tag{4.6}$$

which reduces to

$$P_{FA}(\theta_0) \leq \alpha. \tag{4.7}$$

for the simple hypotheses case. The optimal decision rule for a simple hypotheses in the frequentist approach is given by the Neyman Pearson (NP) strategy. Let $\phi(x)$ be the decision rule defined by

$$\phi(x) = \begin{cases} 1, & \text{decide } \mathrm{H}_1 \\ 0, & \text{decide } \mathrm{H}_0 \end{cases}$$

Then the NP theorem states that the optimal test $\phi^*(x)$ is of the form

$$\phi^*(x) = \begin{cases} 1, & f(x;\theta_1) > \eta f(x;\theta_0) \\ q, & f(x;\theta_1) = \eta f(x;\theta_0) \\ 0 & f(x;\theta_1) < \eta f(x;\theta_0), \end{cases}$$

where η and q are chosen to satisfy $\mathrm{E}_{\theta_0}[\phi^*] = \alpha$. This is also referred to as the most powerful (MP) test of level α since the power function of a test ϕ is $\beta(\theta) = P_D(\theta)$. The randomization parameter q is necessary only if the cumulative density function of the likelihood ratio, $f(x;\theta_1)/f(x;\theta_0)$, has a discontinuity at $1-\alpha$.

If the likelihood is continuous everywhere, then the NP likelihood ratio test (LRT) can be rewritten as

$$\Lambda_{\mathrm{LRT}} : \frac{f(x|\mathrm{H}_1)}{f(x|\mathrm{H}_0)} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \eta, \tag{4.8}$$

where η is chosen to satisfy $P_{FA} = P(\Lambda_{\mathrm{LRT}} > \eta|\mathrm{H}_0)$. Thus the LRT is identical to BLT for the simple hypotheses case, except for the value of η. Though most practical applications are composite in nature, the simple hypotheses case has continued to be a focus of interest for many applications as they are easier to solve and provides a tool for obtaining closed form solutions to basic detection problems. For example, many communication channel need an optimal decision rule for classifying M symbols based on received observations which can be formulated as a multiple hypotheses testing problem with each hypotheses being a singleton set, i.e., $\mathrm{H}_i : \theta = \theta_i,\ i = 1, 2, \ldots, M$.

Keeping in mind the energy constraint in many such problems, the focus of this chapter is to find optimal energy allocation strategies to a detection problem with observations gathered from a sequence of steps. We consider a simple hypotheses testing problem in a classic linear Gaussian model under an average energy constraint, where an optimal decision rule needs to be designed based on observations from a sequence of steps. The goal is to find the optimal energy allocation among the various steps satisfying the energy constraint that minimizes a particular error criterion.

The organization of the chapter is as follows: We describe the hypotheses testing problem formulation in Section 4.2. We then find the optimal two-step energy allocation strategy while using the frequentist approach in Section 4.3. We outline

a drawback of the frequentist strategy for this particular problem and provide the optimal two-step energy allocation for the alternative Bayesian approach in Section 4.4. We provide scope for future work in the conclusion in Section 4.5.

4.2 Problem formulation

Consider the following binary hypothesis testing problem

$$\begin{aligned} \mathrm{H}_0 : \theta &= \theta_0 \\ \mathrm{H}_1 : \theta &= \theta_1, \end{aligned}$$

where θ_0 and θ_1 are known parameters. The set of observations for a linear Gaussian model under each hypothesis is given by

$$\begin{aligned} \mathrm{H}_0 : y_1 &= x_1\theta_0 + n_1, \\ \mathrm{H}_1 : y_1 &= x_1\theta_1 + n_1. \end{aligned} \tag{4.9}$$

Signal x_1 can be intepreted as transmit power and n_1 is additive white Gaussian noise (AWGN) with zero mean and variance σ^2, i.e., $n_1 \sim \mathcal{N}(0, \sigma^2)$.

4.3 Frequentist approach

In the frequentist framework, the decision rule maximizes the probability of correct detection subject to a specified false alarm level.

4.3.1 Single step solution

Given x_1, the Likelihood Ratio Test (LRT) for the observations in (4.9) can be written as

$$\begin{aligned}\Lambda^1_{\text{LRT}} &= \frac{f(y_1;\theta_1)}{f(y_1;\theta_0)} \\ &= \frac{1}{\sigma^2}\left(x_1 y_1(\theta_1-\theta_0) - \frac{x_1^2(\theta_1^2-\theta_0^2)}{2}\right) \\ &= \frac{\theta_1}{\sigma^2}\left(x_1 y_1 - \frac{x_1^2\theta_1}{2}\right) \underset{\text{H}_0}{\overset{\text{H}_1}{\lessgtr}} \eta.\end{aligned}$$

We assume $\theta_1 - \theta_0 > 0$. Then θ_0 can be set to zero without loss of generality. Furthermore, $x_1 > 0$ as any performance achieved by a negative x_1 can be achieved with $|x_1|$. Since θ_1, and x_1 are known positive values, the LRT is equivalent to a test of the form $y_1 \gtrless \eta'$. The optimal region of interest does not depend on the parameters and hence the test is a *Uniform Most Powerful* (UMP) test. η' satisfies

$$P_{FA} = \alpha = P(y_1 > \eta' | \text{H}_0) = 1 - \mathcal{N}(\eta'), \tag{4.10}$$

where α is the false alarm level. Denote β as the probability of correct detection. Then for the one-step case,

$$\beta = 1 - \mathcal{N}(\mathcal{N}^{-1}(1-\alpha) - d), \tag{4.11}$$

where $d = \frac{x_1\theta_1}{\sigma}$. Under energy constraint $|x_1|^2 \leq E_0$, the performance is given by (4.11) with $d = d_0 = \frac{\sqrt{E_0}\theta_1}{\sigma} = \sqrt{\text{SNR}_0}\theta_1$, where $\text{SNR}_0 = E_0/\sigma^2$ is the signal-to-noise ratio.

4.3.2 Two-step design

For a one-step procedure, the transmitted power x_1 modifies the density of the observation y_1 by a simple shift in mean. In a two-step procedure the received measurement y_1 is used to determine the transmitted power at the second step.

Specifically, we control the input $x_2(y_1)$ at the second step based on y_1 to obtain observation y_2. Based on observations y_1 and y_2, an optimal decision is made on the hypotheses $\mathrm{H}_0 vs. \mathrm{H}_1$. The set of observations at the two time steps under each hypothesis can be written as

$$\mathrm{H}_0 : \begin{cases} y_1 = x_1\theta_0 + n_1 \\ y_2 = x_2(y_1)\theta_0 + n_2 \end{cases} \tag{4.12}$$

$$\mathrm{H}_1 : \begin{cases} y_1 = x_1\theta_1 + n_1 \\ y_2 = x_2(y_1)\theta_1 + n_2, \end{cases} \tag{4.13}$$

where n_1, n_2 are i.i.d zero mean and variance σ^2 Gaussian random variables, i.e., $n_i \sim \mathcal{N}(0, \sigma^2)$. The two-step log likelihood ratio test is given by

$$\begin{aligned} \Lambda^2_{\mathrm{LRT}} &= \log \frac{f(y_1, y_2|\mathrm{H}_1)}{f(y_1, y_2|\mathrm{H}_0)} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\lessgtr}} \eta, \\ &= z_1 + z_2 \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\lessgtr}} \eta, \end{aligned} \tag{4.14}$$

where

$$z_1 = \frac{\theta_1}{\sigma^2}\left(x_1 y_1 - \frac{x_1^2\theta_1}{2}\right), \tag{4.15}$$

$$z_2 = \frac{\theta_1}{\sigma^2}\left(x_2(y_1) y_2 - \frac{x_2^2(y_1)\theta_1}{2}\right). \tag{4.16}$$

are the log likelihood ratios evaluated at the two steps separately.

We attempt to satisfy the average energy constraint under the null hypothesis H_0 while maintaining a specified false alarm. Such an energy constraint makes sense for a sensor network in a remote location, where targets are rare. In other words, most of the time, the null hypothesis is valid and threshold exceedances are false alarms. An analogous formulation could be achieved for the miss probability under H_1. Thus, the goal is to find the optimal design parameters x_1, $x_2(y_1)$ and η to maximize the

probability of detection (P_D) subject to a fixed false alarm level α and subject to an average energy constraint $\mathrm{E}_{\mathrm{H}_0}\left[|x_1|^2 + |x_2(y_1)|^2\right] \leq E_0$. Without loss of generality, the energy constraint can be rewritten as $\mathrm{E}_{\mathrm{H}_0}\left[x_1^2 + x_2^2(y_1)\right] \leq E_0$. The objective function is given by

$$H = P_D - \mu(P_{FA} - \alpha) - \lambda(\mathrm{E}_{\mathrm{H}_0}\left[|x_1|^2 + |x_2(y_1)|^2\right] - E_0), \tag{4.17}$$

where μ and λ are Lagrangian multipliers chosen to satisfy the constraints.

Before we optimize for x_1 and $x_2(y_1)$, we derive expressions for P_D and P_{FA}. The cumulative distribution function (CDF) of z_1 and z_2 under H_0 is given by

$$F_{z_1}(z|\mathrm{H}_0) = 1 - \mathrm{Q}\left(\frac{z + t_1/2}{\sqrt{t_1}}\right) \tag{4.18}$$

$$F_{z_2}(z|y_1, \mathrm{H}_0) = 1 - \mathrm{Q}\left(\frac{z + t_2(y_1)/2}{\sqrt{t_2(y_1)}}\right). \tag{4.19}$$

where $t_1 = \frac{x_1^2\theta_1^2}{\sigma^2}$ and $t_2(y_1) = \frac{x_2^2(y_1)\theta_1^2}{\sigma^2}$. Similarly the distribution under H_1 can be written as,

$$F_{z_1}(z|\mathrm{H}_1) = 1 - \mathrm{Q}\left(\frac{z - t_1/2}{\sqrt{t_1}}\right) \tag{4.20}$$

$$F_{z_2}(z|y_1, \mathrm{H}_1) = 1 - \mathrm{Q}\left(\frac{z - t_2(y_1)/2}{\sqrt{t_2(y_1)}}\right). \tag{4.21}$$

Further the density of z_1 under H_0 and H_1 satisfies the following relation,

$$f_{z_1}(z|\mathrm{H}_1) = f_{z_1}(z|\mathrm{H}_0)e^z. \tag{4.22}$$

Since there is a one-to-one mapping between y_1 and z_1 through (4.15), we refer to the design parameter $x_2(y_1)$ as $x_2(z_1)$. Similarly $t_2(y_1) = t_2(z_1)$. The probability of

correct detection (P_D) is given by

$$\begin{aligned} P_D &= P(z_1 + z_2 \geq \eta | \mathrm{H}_1) \\ &= \mathrm{E}_{\mathrm{H}_1}\left[P(z_2 \geq \eta - z_1 | \mathrm{H}_1, z_1)\right] \\ &= \mathrm{E}_{\mathrm{H}_1}\left[\mathrm{Q}\left(\frac{\eta - z_1 - t_2(z_1)/2}{\sqrt{t_2(z_1)}}\right)\right] \\ &= \mathrm{E}_{\mathrm{H}_0}\left[\mathrm{Q}\left(\frac{\eta - z_1 - t_2(z_1)/2}{\sqrt{t_2(z_1)}}\right) e^{z_1}\right], \end{aligned} \tag{4.23}$$

where the last equality follows from the relation in (4.22). The probability of false alarm (P_{FA}) is

$$\begin{aligned} P_{FA} &= P(z_1 + z_2 \geq \eta | \mathrm{H}_0) \\ &= \mathrm{E}_{\mathrm{H}_0}\left[P(z_2 \geq \eta - z_1 | \mathrm{H}_0, z_1)\right] \\ &= \mathrm{E}_{\mathrm{H}_0}\left[\mathrm{Q}\left(\frac{\eta - z_1 + t_2(z_1)/2}{\sqrt{t_2(z_1)}}\right)\right]. \end{aligned} \tag{4.24}$$

The average energy constraint under H_0 can be equivalently expressed as

$$E_{\mathrm{avg}} = \mathrm{E}_{\mathrm{H}_0}\left[x_1^2 + x_2^2(z_1)\right]. \tag{4.25}$$

Substituting (4.23), (4.24), and (4.25) in the expression for the objective function in (4.17), we obtain

$$\begin{aligned} H &= \mathrm{E}_{\mathrm{H}_0}\left[\mathrm{Q}\left(\frac{\eta - z_1 - t_2(z_1)/2}{\sqrt{t_2(z_1)}}\right) e^{z_1}\right] - \mu \mathrm{E}_{\mathrm{H}_0}\left[\mathrm{Q}\left(\frac{\eta - z_1 + t_2(z_1)/2}{\sqrt{t_2(z_1)}}\right)\right] \\ &\quad -\lambda \mathrm{E}_{\mathrm{H}_0}\left[x_2^2(z_1)\right] + \mu\alpha + \lambda E_0 + \lambda x_1^2 \\ &= \mathrm{E}_{\mathrm{H}_0}\left[\mathrm{Q}\left(\frac{\eta - z_1 - t_2(z_1)/2}{\sqrt{t_2(z_1)}}\right) e^{z_1} - \mu \mathrm{Q}\left(\frac{\eta - z_1 + t_2(z_1)/2}{\sqrt{t_2(z_1)}}\right) - \lambda' t_2(z_1)\right] \\ &\quad +\mu\alpha + \lambda E_0 + \lambda x_1^2, \end{aligned}$$

where $\lambda' = \sigma^2\lambda/\theta_1^2$. We now find the optimal $t_2(z_1)$ and η that maximize P_D under these constraints. Setting $\frac{\partial H}{\partial \eta} = 0$ for each z_1, we obtain

$$\exp\left(-\frac{1}{2t_2(z_1)}(\eta - z_1 - t_2(z_1)/2)^2\right) \exp z_1 = \mu \exp\left(-\frac{1}{2t_2(z_1)}(\eta - z_1 + t_2(z_1)/2)^2\right).$$

Expanding the terms inside the exponentials and canceling terms common to both sides yields

$$\eta = \log_e(\mu). \tag{4.26}$$

The derivative of H with respect to $t_2(z_1)$ is given by

$$\begin{aligned} \frac{\partial H}{\partial t_2} &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\eta - z_1 - t_2/2)^2}{2t_2}\right) \exp(z_1) \left(\frac{\eta - z_1 + t_2/2}{2t_2\sqrt{t_2}}\right) \\ &\quad -\mu \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\eta - z_1 + t_2/2)^2}{2t_2}\right) \left(\frac{\eta - z_1 - t_2/2}{2t_2\sqrt{t_2}}\right) - \lambda'. \end{aligned} \tag{4.27}$$

Substituting for μ from (4.26) and multiplying by $\exp(-\eta)$ on both sides, we obtain

$$\begin{aligned} \frac{\partial H}{\partial t_2} &= \frac{1}{\sqrt{2\pi}} \exp(z_1 - \eta) \exp\left(-\frac{(\eta - z_1 - t_2/2)^2}{2t_2}\right) \left(\frac{\eta - z_1 + t_2/2}{2t_2\sqrt{t_2}}\right) \\ &\quad -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\eta - z_1 + t_2/2)^2}{2t_2}\right) \left(\frac{\eta - z_1 - t_2/2}{2t_2\sqrt{t_2}}\right) - \lambda'', \end{aligned} \tag{4.28}$$

where $\lambda'' = \exp(-\eta)\lambda'$.

To solve for the optimal t_2 using the above expression, consider the following parametric substitution:

$$\eta - z_1 = u\sqrt{t_2} - \frac{t_2}{2}. \tag{4.29}$$

Substituting (4.29) in (4.28) and setting the expression for the derivative in (4.28) to zero, we obtain

(4.30)

$$\exp\left(-u\sqrt{t_2} + \frac{t_2}{2}\right) \exp\left(-\frac{(u - \sqrt{t_2})^2}{2}\right) \left(\frac{u}{2t_2}\right) - \exp\left(-\frac{u^2}{2}\right) \left(\frac{u - \sqrt{t_2}}{2t_2}\right) = \sqrt{2\pi}\lambda'',$$

Simplifying the above expression yields

$$t_2(z_1) = \frac{1}{8\pi(\lambda'')^2} \exp(-u^2). \tag{4.31}$$

Substituting $t_2(z_1)$ from (4.31) in (4.29), we have

$$z_1 = \log_e(\mu) - \frac{u}{2\sqrt{2\pi}\lambda''} \exp(-u^2/2) - \frac{1}{16\pi(\lambda'')^2} \exp(-u^2). \tag{4.32}$$

The expressions in (4.31) and (4.32) yield the optimal solution to the transmit power at the second stage as a function of the observation at the first step in parametric form. Constants $\lambda^{''}$ and μ are chosen to satisfy the energy constraint and false alarm constraint respectively.

Figure 4.1 illustrates the structure of the optimal solution, $t_2^2(z_1)$ generated by the parametric solution in (4.31) and (4.32) for various values of $\lambda^{''}$ at fixed $\eta \approx -2.3$ and $x_1 = 1$. This can be thought of as a plot of the energy function $x_2^2(z_1)$ with $\theta_1 = 1$ and $\sigma = 1$. We observe that the optimal solution has mean greater than zero. The intuition behind such a solution is the following: The distribution of z_1 under H_0 is $\mathcal{N}(-t_1/2, t_1)$ while the distribution of z_1 under H_1 is $\mathcal{N}(t_1/2, t_1)$. Thus a realization of z_1 which is close $-t_1/2$ implies that with reasonably high probability the realization came from the H_0 distribution. On the other hand if the realization of z_1 is $t_1/2$ or greater, then with high probability, the realization came from H_1. Since we constrain the energy under the null hypothesis H_0, the solution assigns very low or no energy at the second step for those realizations of z_1 which are $-t_1/2$ or lower. However if the realization of z_1 is high, then the algorithm is able to assign large amounts of energy at the second step to gain on the performance of the algorithm as the realizations were more likely from H_1. However for very large z_1 values, the algorithm resets the energy level to zero for two reasons. Firstly, increase of energy levels for large z_1 are tightly controlled by the energy constraint. Secondly, for large z_1, the accuracy of the decision at the first step improves, reducing the need for energy allocation to the second step. The test will clearly choose H_1 as the solution in such cases.

Figure 4.2 shows the ROC curve for the optimal two-step solution (solid) with the following parameters: $\sigma = 1$, $\theta_0 = 0$, $\theta_1 = 1$, $E_0 = 1$. We implement the optimal two-

step solution as follows: We fix a false alarm level of α by choosing μ appropriately. We then take a grid of values of x_1 between 0 and 1; for each value of x_1, we find the optimal $x_2(z_1)$ which satisfies the energy constraint by a numerical search on the values of λ''. We then find the optimal x_1 and the corresponding $x_2(z_1)$ that maximizes P_D. This process is repeated for various values of α. We also plot the corresponding optimal one-step strategy (dashed dotted) with $E_0 = 1$. We notice a significant improvement in performance especially at low false alarm rates while at other places the two-step ROC is at least as good as the one-step solution.

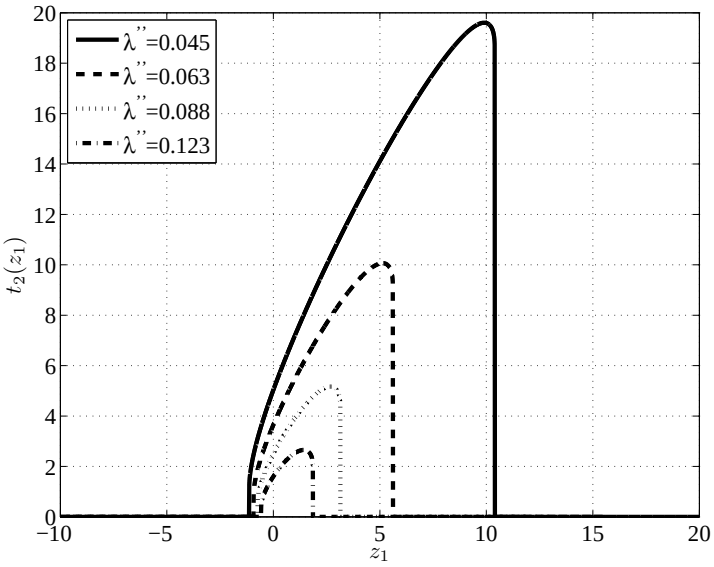


Figure 4.1: Plot showing the optimal solution to energy at second stage for various values of λ'' with $x_1 = 1$ and $\mu = -2.3$.

In table 4.1, we present a selected pair of points P_D and P_{FA} that lead to the optimal two-step ROC in Fig. 4.2. For each pair, we also present the average energy under the H_1 hypothesis. We know already that the average energy under H_0 hypothesis is exactly E_0. However we observe from the table that the average energy consumed under H_1 for low P_{FA} is more than E_0. This is because of the bias in the optimal solution $x_2(z_1)$ towards the mean of the H_1 hypothesis discussed earlier.

P_{FA}	P_D	$\mathrm{E}_{\mathrm{H}_1}\left[x_1^2 + x_2^2\right]$
0.0081	0.2281	1.9151
0.0093	0.2372	1.8903
0.0106	0.2466	1.8868
0.0122	0.2565	1.8819
0.0183	0.2843	1.7939
0.0209	0.2950	1.7894
0.0241	0.3078	1.7767
0.0366	0.3445	1.6755
0.0422	0.3603	1.6535

Table 4.1: Probability of detection, false alarm values for optimal two-step solution along with energy consumed under H_1 hypothesis.

Thus in reality, if the observations come from H_1, the solution proposes to use more energy than E_0 for most false alarm rates which is counter intuitive in light of the restriction of the total energy. However, the average energy constraint is well defined in a Bayesian setting as we will see in the next section.

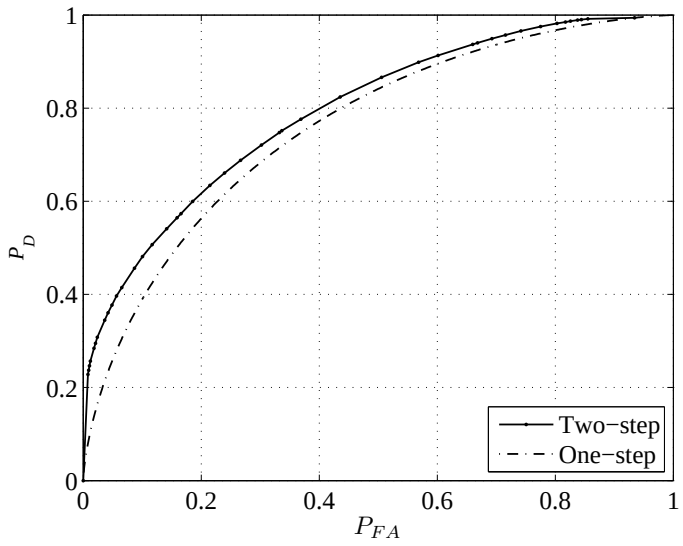


Figure 4.2: Optimal ROC curves for the optimal one-step and two-step strategies in the frequentist setting.

4.4 Bayesian approach

In the frequentist approach, we considered the energy allocation problem for simple hypotheses testing ($\theta = \theta_0, \theta = \theta_1$). Since the optimal LRT was UMP, the results

also hold for a composite hypotheses under suitable regions of Θ_0 and Θ_1. For example, a composite hypotheses test of the form: $\mathrm{H}_0 : \theta = \theta_0$, $\mathrm{H}_1 : \theta > \theta_0$ or $\theta < \theta_0$. In this section, we consider the energy allocation for hypotheses testing using the following Bayesian framework:

$$\mathrm{H}_0 : \theta \in \Theta_0 \tag{4.33}$$

$$\mathrm{H}_1 : \theta \in \Theta_1, \tag{4.34}$$

where θ has a prior density $f(\theta)$. The problem of finding optimal energy allocation for an arbitrary $f(\theta)$ is not feasible. Hence, we restrict our treatment to the following simple hypotheses case:

$$\mathrm{H}_0 : \theta = \theta_0 \quad \text{with prob. } p \tag{4.35}$$

$$\mathrm{H}_1 : \theta = \theta_1 \quad \text{with prob. } 1-p. \tag{4.36}$$

The Bayesian approach is to maximize the average probability of error, which from the laws of total probability is given by

$$P_{\text{avg}} = P(\mathrm{H}_0|\mathrm{H}_1)P(\mathrm{H}_1) + P(\mathrm{H}_1|\mathrm{H}_0)P(\mathrm{H}_0) \tag{4.37}$$

$$= (1 - P_D)p + P_{FA}(1-p). \tag{4.38}$$

4.4.1 One-step solution

Using the optimal Bayesian likelihood ratio test illustrated by (4.3), the decision rule can be written as

$$\log\left(\frac{f(y|H_1)}{f(y|H_0)}\right) \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \eta^*, \tag{4.39}$$

where the optimal Bayes threshold is given by the relation

$$\eta^* = \log\left(\frac{p}{1-p}\right). \tag{4.40}$$

The test simplifies to

$$\frac{\theta_1}{\sigma^2}\left(x_1 y_1 - \frac{x_1^2\theta_1}{2}\right) \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\lessgtr}} \eta^*, \tag{4.41}$$

when without loss of generality we assume $\theta_0 = 0$. Then the probability of correct detection, P_D and the probability of false alarm, P_{FA} are given by

$$P_D = \mathrm{Q}\left(\frac{\eta^*\sigma^2}{\theta_1 x_1} - \frac{x_1\theta_1}{2\sigma}\right) \tag{4.42}$$

$$P_{FA} = \mathrm{Q}\left(\frac{\eta^*\sigma^2}{\theta_1 x_1} + \frac{x_1\theta_1}{2\sigma}\right), \tag{4.43}$$

The average energy constraint for the one-step case is given by

$$E_{\text{avg}} = \mathrm{E}\left[x_1^2|\mathrm{H}_0\right] P(\mathrm{H}_0) + \mathrm{E}\left[|x_1|^2|\mathrm{H}_1\right] P(\mathrm{H}_1) = x_1^2 \leq E_0. \tag{4.44}$$

Substituting the expressions for P_D and P_{FA} in (4.42) and (4.43) in (4.38), the minimum average probability of error for the one-step strategy is

(4.45)

$$P_{\text{avg}} = (1-p)\left(1 - \mathrm{Q}\left(\frac{\eta^*}{\theta_1 \mathrm{SNR}_0} - \frac{\sqrt{\mathrm{SNR}_0}\theta_1}{2}\right)\right) + p\left(\mathrm{Q}\left(\frac{\eta^*}{\theta_1 \mathrm{SNR}_0} + \frac{\sqrt{\mathrm{SNR}_0}\theta_1}{2}\right)\right)$$

4.4.2 Two-step solution

The set of equations governing each hypothesis are the same set of equations from the frequentist approach given in (4.12) and (4.13). The optimal two-step Bayes test is the same as the LRT presented in (4.14), where η is chosen optimally to minimize the criterion in the Bayesian setting. Rather than maximize P_D subject to a false alarm constraint, the objective here is to design the optimal energy allocation strategy to minimize average probability of error. The objective function for the two-step strategy can be written as

$$H_b = (1-p)(1-P_D) + pP_{FA} + \lambda\left(\mathrm{E}\left[t_1^2 + t_2^2(z_1)\right] - \frac{E_0\theta_1^2}{\sigma^2}\right), \tag{4.46}$$

where the probabilities of false alarm and correct detection are the same expressions given in (4.24) and (4.23), respectively, since the test remains the same. Further we have reparameterized the energy allocation parameters x_1 and $x_2(z_1)$ through t_1 and $t_2(z_1)$ defined below (4.19). The average energy constraint can be simplified as

$$\begin{aligned} E_{\text{avg}} &= \mathrm{E}\left[t_1^2 + t_2^2(z_1)\right] \\ &= t_1^2 + p\mathrm{E}_{\mathrm{H}_0}\left[t_2^2(z_1)\right] + (1-p)\mathrm{E}_{\mathrm{H}_1}\left[t_2^2(z_1)\right] \\ &= x_1^2 + \left(p + (1-p)e^{z_1}\right)\mathrm{E}_{\mathrm{H}_0}\left[t_2^2(z_1)\right]. \end{aligned} \tag{4.47}$$

Substituting the expressions for P_{FA}, P_D, and E_{avg} from (4.24), (4.23), and (4.47), respectively in (4.46), we obtain

$$\begin{aligned} H_b &= (1-p)\left(1 - \mathrm{E}_{\mathrm{H}_0}\left[\mathrm{Q}\left(\frac{\eta - z_1 - t_2(z_1)/2}{\sqrt{t_2(z_1)}}\right)e^{z_1}\right]\right) \\ &\quad + p\mathrm{E}_{\mathrm{H}_0}\left[\mathrm{Q}\left(\frac{\eta - z_1 + t_2(z_1)/2}{\sqrt{t_2(z_1)}}\right)\right] \\ &\quad - \lambda\left(x_1^2 + \left(p + (1-p)e^{z_1}\right)\mathrm{E}_{\mathrm{H}_0}\left[t_2(z_1)\right] - \frac{E_0\theta_1^2}{\sigma^2}\right). \end{aligned} \tag{4.48}$$

Setting the partial derivative of H_b with respect to η to zero yields,

$$\begin{aligned} &(1-p)\exp\left(-\frac{1}{2t_2(z_1)}(\eta - z_1 - t_2(z_1)/2)^2\right)\exp z_1 \\ &\quad - p\exp\left(-\frac{1}{2t_2(z_1)}(\eta - z_1 + t_2(z_1)/2)^2\right) = 0. \end{aligned} \tag{4.49}$$

The optimal solution to η is given by

$$\eta^* = \log\left(\frac{p}{1-p}\right), \tag{4.50}$$

which is the same optimal solution as in the one-step strategy. Setting $\frac{\partial H_b}{\partial t_2} = 0$, we obtain

$$\begin{aligned} &\frac{1-p}{\sqrt{2\pi}}\exp\left(-\frac{(\eta - z_1 - t_2/2)^2}{2t_2}\right)\exp(z_1)\left(\frac{\eta - z_1 + t_2/2}{2t_2\sqrt{t_2}}\right) \\ &\quad - \frac{p}{\sqrt{2\pi}}\exp\left(-\frac{(\eta - z_1 + t_2/2)^2}{2t_2}\right)\left(\frac{\eta - z_1 - t_2/2}{2t_2\sqrt{t_2}}\right) = \lambda\left(p + (1-p)\exp(z_1)\right). \end{aligned}$$

Dividing by $1-p$ on both sides and multiplying by $\exp(-\eta^*)$, we have

$$\exp\left(-\frac{(\eta^*-z_1-t_2/2)^2}{2t_2}\right)\exp(z_1-\eta^*)\left(\frac{\eta^*-z_1+t_2/2}{2t_2\sqrt{t_2}}\right) - \exp\left(-\frac{(\eta-z_1+t_2/2)^2}{2t_2}\right)\left(\frac{\eta-z_1-t_2/2}{2t_2\sqrt{t_2}}\right) = \sqrt{2\pi}\lambda\left(1+\exp(z_1-\eta^*)\right).$$

Using the same parametric substitution as (4.29) and simplifying, we obtain

$$\frac{1}{2\sqrt{t_2}}\exp(-\frac{u^2}{2}) = \sqrt{2\pi}\lambda(1+\exp\left(\frac{t_2}{2}-u\sqrt{t_2}\right). \tag{4.51}$$

The optimal solution to t_2 in parametric form is given by

$$t_2 = \frac{1}{8\pi\lambda^2}\frac{1}{\left(\exp(-u^2/2)+\exp-(u-\sqrt{t_2})^2/2\right)^2}, \tag{4.52}$$

where u and z_1 are given by the relation

$$\eta^*-z_1 = u\sqrt{t_2}-\frac{t_2}{2}. \tag{4.53}$$

The solution to this equation is not obtainable in closed form. Hence we solve it numerically and if there is more than one solution, we pick the lower valued solution as the optimal t_2 to minimize energy. Furthermore, $t_2(z_1)$ is symmetric about η^*, i.e., $t_2'(z_1) = t_2'(-z_1)$, where $t_2'(z_1) = t_2(z_1+\eta^*)$. This result is easy to verify. Substituting for u from (4.53) in (4.52) and simplifying, the optimal $t_2(z_1)$ satisfies

$$\sqrt{t_2(z_1)} = \frac{1}{2\sqrt{2\pi}\lambda}\frac{\exp\left(-\frac{1}{2t_2(z_1)}(z_1-\eta^*-t_2(z_1)/2)^2\right)}{1+\exp(z_1-\eta^*)}. \tag{4.54}$$

Then $t_2'(z_1) = t_2(z_1+\eta^*)$ is given by

$$\sqrt{t_2'(z_1)} = \frac{1}{2\sqrt{2\pi}\lambda}\frac{\exp-\left(\frac{1}{2t_2'(z_1)}(z_1-t_2'(z_1)/2)^2\right)}{1+\exp(z_1)}. \tag{4.55}$$

Replacing z_1 with $-z_1$, we obtain

$$\begin{aligned}
\sqrt{t_2'(-z_1)} &= \frac{1}{2\sqrt{2\pi}\lambda}\frac{\exp\left(-\frac{1}{2t_2'(-z_1)}(-z_1 - t_2'(-z_1)/2)^2\right)}{1+\exp(-z_1)} \\
&= \frac{1}{2\sqrt{2\pi}\lambda}\frac{\exp\left(\frac{1}{2t_2'(-z_1)}(z_1 + t_2'(-z_1)/2)^2\right)\exp(z_1)}{1+\exp(z_1)} \\
&= \frac{1}{2\sqrt{2\pi}\lambda}\frac{\exp\left(-\frac{1}{2t_2'(-z_1)}(z_1 - t_2'(-z_1)/2)^2\right)}{1+\exp(z_1)}. \qquad (4.56)
\end{aligned}$$

Comparing (4.55) and (4.56), $t_2'(z_1)$ and $t_2'(-z_1)$ satisfy the same equation for optimality. It follows that the solution to $t_2'(z_1)$ and $t_2'(-z_1)$ is identical.

The optimal solution given by the parametric form in (4.52) and (4.53) is plotted in Fig. 4.3 for various values of λ at $p = 0.5$. For $p = 0.5$, $\eta^* = 0$ and the optimal curves are symmetric about zero. Figure 4.4 shows the optimal solution to the energy at the second step for different values of p with fixed $\lambda = 0.03$. We observe that the energy curves are symmetric about the lines $z_1 = \eta^*$, where $\eta^* = \log(p/(1-p))$.

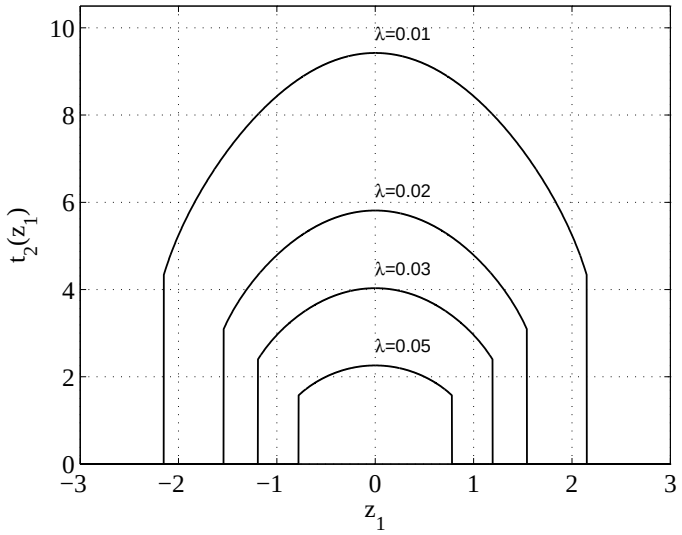


Figure 4.3: Plot showing the optimal solution to energy at second stage for various values of λ. We set $p = 0.5, \theta_0 = 0, \theta_1 = 1$, and $\sigma = 1$.

Figure 4.5 shows the performance of the optimal one-step (dotted) and two-step strategies (solid), i.e., minimum average error versus p, the probability of H_0 for

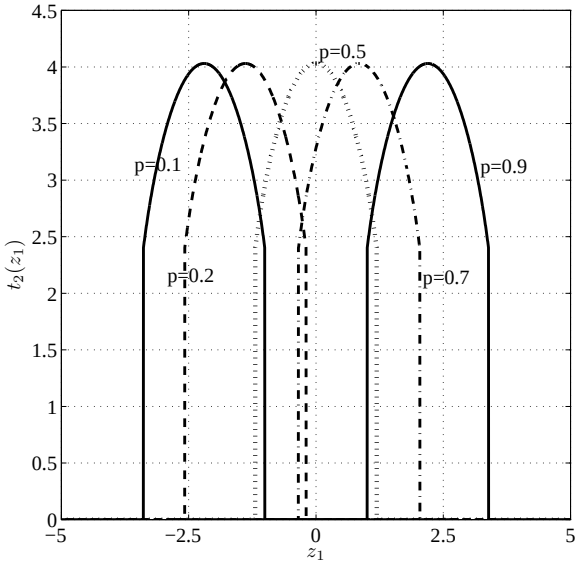


Figure 4.4: Plot of the optimal solution to energy at second step for different prior probabilities at $\lambda = 0.03$.

3 different SNRs, i.e., energy levels. The following values were assumed for the parameters to generate the plot: $\theta_0 = 0, \theta_1 = 1, \sigma = 1$, and the energy allocated to the first stage is 1 ($x_1 = 1$). Rather than optimizing the energy allocation between steps 1 and 2, we assume that $x_1 = 1$, and find the optimal λ that yields $\mathrm{E}\left[x_2^2(z_1)\right] = E_0 - 1$. Using the optimal design, we compute the average probability of error by evaluating the values of P_D and P_{FA} in (4.42) and (4.43), respectively using numerical integration. The optimal one-step performance was generated using (4.45).

Figure 4.6 plots the optimal performance of both the one-step and two-step theoretically and via simulations for a SNR of 3dB, i.e., $E_0 = 2$. The theoretical curve for the one-step procedure was generated using (4.45). For the simulation curve, for each value of p, we generated $N = 100000$-element random vector which had zeros with probability p and one otherwise corresponding to the H_0 and H_1 hypothesis. Let N_0 represent the number of zeros and N_1 represent the number of ones. Based on this random vector, N samples of y_1 were generated using (4.9) and the optimal Bayes threshold test in (4.41) was performed. We compute N_{10} and N_{01} as

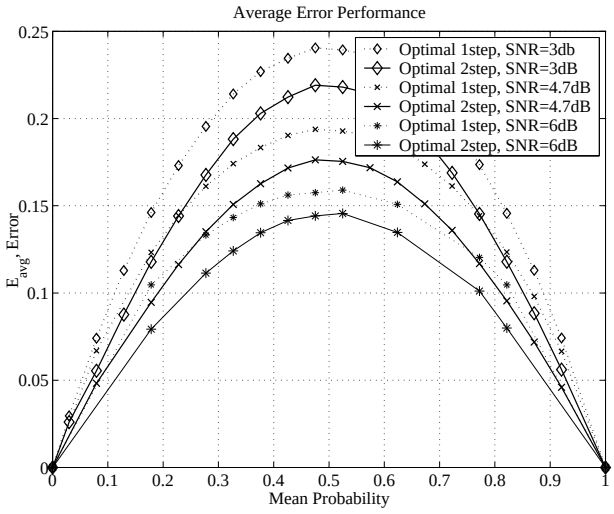


Figure 4.5: Minimum average error versus prior probability for varying SNR for the optimal one-step and two-step cases.

the number of times a decision was classified as H_1 when it was H_0 and viceversa, respectively. The probability of miss and false alarm are then computed as N_{10}/N_0 and N_{01}/N_1, respectively. The average probability of error was evaluated empirically as $N_{10}/N + N_{01}/N$. For the two-step procedure, we first find the optimal second step energy allocation and generated the analytical curve by computing the expressions for P_D and P_{FA} in (4.42) and (4.43), respectively using numerical integration. For the simulation curve, we generate the samples in a similar fashion to the one-step procedure. We first create the N-element random vector of ones and zeros based on prior probability p and generate samples of both y_1 and y_2 using the expressions in (4.12) and (4.13). After performing the optimal two-step Bayes thresholding test in (4.14) with $\eta = \eta^*$, we compute empirical estimates of P_D and P_{FA} by counting the average number of times a wrong classification occurs. We use the estimates of P_D and P_{FA} to compute the average probability of error. We observe from Fig. 4.6 that the analysis and the simulation results match accurately.

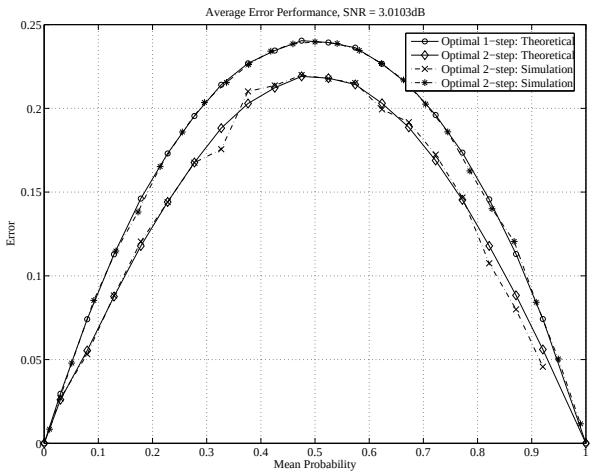


Figure 4.6: Minimum average error versus prior probability at SNR=3dB theoretically and by simulation.

4.5 Conclusions

We considered the binary hypotheses testing problem and found the optimal energy allocation strategies for the two-step procedure under an average energy constraint for the frequentist and the Bayesian approaches. In the frequentist framework, we constrained the energy under the null hypothesis H_0 under the assumption that target presence is rare. The constraint is satisfied under H_0 but not under H_1. To obtain a strategy that satisfies the energy constraint under H_0 or H_1, we turn to a Bayesian formulation. For the Bayesian formulation, the average energy constraint is defined as the average energy expended under both the hypothesis. The Bayesian optimal detector framework yielded a similar solution as the frequentist optimal detector for the parameter values investigated and the optimal solution yielded at least a 2dB improvement in performance. Future work is to extend this principle to an M-ary hypotheses testing procedure.

CHAPTER V

Adaptive waveform selection for state estimation

Chapters II-IV focused on adaptive waveform amplitude design for estimation and detection. In this chapter, we shift our attention towards the problem of optimal adaptive waveform selection. We would like to choose a small subset from a given set of waveforms that minimizes state prediction mean squared error (MSE) given the past observations. This differs from previous approaches to this problem since the minimization problem cannot be generally solved by offline solutions (e.g., matching pursuit (MP), basis pursuit (BP)); the optimal waveforms must be selected after observations are received. Since the optimal solution to the subset selection problem is combinatorially complex, we propose a convex relaxation of the problem and provide a low complexity suboptimal solution. We present a hidden Markov model for the state and show that the performance of this suboptimal procedure for state estimation approaches that of the globally optimal procedure.

5.1 Introduction

Over the past decade, the problem of optimal waveform design has found important applications in synthetic aperture radar (SAR), automatic target recognition and radar astronomy [13]. Based on the application, waveform design may depend

on various optimality criteria, e.g., target classification [155], accurate reconstruction of a high resolution radar image [154], or estimating a set of target parameters. One implication of choosing the set of transmitted waveforms optimally is that the backscattered signals will contain maximum target information.

Most of the work in the area of waveform design involves finding the best functional form of the waveforms suited to a particular task, e.g., design of waveforms from the radar ambiguity function for narrowband signals [16] or design of wideband waveforms to resolve targets in dense target environments [110]. However, closed-form solutions to waveform design problems are only possible under restrictive model assumptions, e.g., one-dimensional target motions [79].

In this chapter, we focus on the optimal waveform selection problem rather than the design of waveforms, i.e., we would like to choose only a small subset from a given library of waveforms. One application of the waveform subset selection problem is in hyperspectral imaging [141]. A hyperspectral radar surveillance system mounted on an aircraft or spacecraft obtains an spectral image using the energy reflected or emitted by each voxel on the ground for many bands of the optical spectrum. Different objects (e.g., minerals, metal, and soil) possess unique spectral signatures in a particular spectral band. These signatures are then used to identify the objects. However, transmission of the data cube (layers of spectral images) from the radar system to a processing unit is difficult and expensive due to the large volume of data. As an example, a data cube from the NASA imaging spectrometer is 500Mb in size. An adaptive waveform selector can be used to find the optimal combination of spectral bands that yields a minimal data set containing complete information about the objects. Furthermore, for problems where exact solutions to waveform designs are infeasible, a waveform selection procedure can provide an approximation

to the optimal design solution. To assess the performance of a particular subset of waveforms, we need to define an optimization criterion such as expected reward or risk.

The problem of choosing p out of M possible waveforms becomes a high complexity combinatorial optimization problem. E.g., if there are $M = 128$ waveforms and we need to select $p = 32$ element subset, there are more than 10^{30} combinations of indices that need to be checked. Significant work has been focused on approximation methods based on convex relaxation which lead to sparse solutions [98]. One type of convex penalty is the lasso, a shrinkage method which imposes an l_1-norm constraint on the optimization problem [164]. By nature of the constraint, making the weighting of the constraint larger causes some of the coefficients to be zero thus giving rise to a suboptimal sparse solution to the subset selection problem. The matching pursuit (MP) [99] and basis pursuit (BP) [30] algorithms use interior point methods and greedy stepwise regression, respectively to solve the lasso but involve considerable complexity. The more recently proposed least angle regression (LARS) is an efficient stepwise algorithm, which yields the exact solution to the lasso in linear time.

Recently, conditions under which the l_1-norm constrained solution is equivalent to the optimal subset selection is derived in [46, 60]. As a result, the use of l_1-norm constrained convex optimization to obtain spare representations has become increasingly popular [98]. However, most of these problems that deal with sparse regression are offline strategies, where the library of waveforms are chosen a priori based on accumulated data. In contrast, we propose to solve the online subset selection problem by construction of adaptive waveform libraries and using the LARS to obtain the subset selection in linear time.

In this chapter, we consider the expected state prediction MSE as a measure of

performance and find the optimal subset that minimizes this expected reward given the past measurements (online strategy). We relax this combinatorially complex problem into an optimization problem under l_1-norm constraint and propose a low complexity suboptimal solution whose performance approaches that of the optimal subset selection. We then consider a numerical example to illustrate this approach and provide simulation results to compare the various solutions.

The organization of the chapter is as follows: In Section 5.2, we present the waveform selection problem. Section 5.3 proposes a suboptimal solution. We contrast our solution with previous offline approaches in Section 5.4. In Section 5.5, we solve the problem for a specific model. Section 5.6 addresses the computational complexity of the proposed solution and Section 5.7 provides simulation results. We conclude this chapter in Section 5.8.

5.2 Problem formulation

We consider the waveform selection problem for a hyperspectral radar system, where the radar can transmit and receive energy over multiple channels simultaneously. We restrict the number of waveforms transmitted at any time to be a small subset of p out of M available waveforms. Denote the state at time t as $\mathbf{s}_t$ and let the received signals corresponding to a single transmit waveform $\boldsymbol{\phi}_i$ be denoted as $\mathbf{y}_t^i,\ i = 1, \ldots, M$. We restrict our attention to single stage policies, i.e., myopic policies that seek to maximize an expected reward conditioned on the immediate past.

Let $\{i_1, \ldots, i_p\} \in \{1, \ldots, M\}$ denote the indices of the p different waveforms taken from a set of M $(M \geq p)$ waveforms. Denote the set of past observations as $\mathbf{Y}^{t-1} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1}]$. We solve the optimal subset selection problem by maximizing the

expected reduction in the variance of the optimal state estimator after an action (choosing p out of M waveforms) is taken:

$$\max_{i_1,\ldots,i_p} \Big\{ \mathrm{E}\left[\|\mathbf{s}_t - \mathrm{E}\left[\mathbf{s}_t \middle| \mathbf{Y}^{t-1}\right]\|^2 \middle| \mathbf{Y}^{t-1}\right] - \mathrm{E}\left[\left\| \mathbf{s}_t - \mathrm{E}\left[\mathbf{s}_t \middle| \mathbf{y}_t^{i_1}, \ldots, \mathbf{y}_t^{i_p}, \mathbf{Y}^{t-1}\right]\right\|^2 \middle| \mathbf{Y}^{t-1}\right] \Big\}. \tag{5.1}$$

Since the first term is independent of $\{i_1, \ldots, i_p\}$, the maximization in (5.1) can be equivalently expressed as

$$\min_{i_1,\ldots,i_p} \mathrm{E}\left[\|\mathbf{s}_t - \widehat{\mathbf{s}}_t(i_1, \ldots, i_p)\|^2 \middle| \mathbf{Y}^{t-1}\right], \tag{5.2}$$

where

$$\widehat{\mathbf{s}}_t(i_1, \ldots, i_p) \;=\; \mathrm{E}\left[\mathbf{s}_t \middle| \mathbf{y}_t^{i_1}, \ldots, \mathbf{y}_t^{i_p}, \mathbf{Y}^{t-1}\right]. \tag{5.3}$$

The minimization in (5.2) requires one to evaluate (5.3) for all $\binom{M}{p}$ possibilities of $i_1, \ldots, i_p$. Two fundamental difficulties are encountered in solving (5.2): computation of the conditional expectation (5.3); and combinatorial minimization of (5.2). In the tracking examples considered here the computation of (5.3) is not difficult. Since the complexity of problem is exponential in M (for fixed p/M), we propose a low complexity suboptimal solution for (5.2) whose performance approaches that of the optimal one.

5.3 Proposed solution

As an alternative to exhaustively searching over $\binom{M}{p}$ possible subsets we pose the following sparsity constrained prediction surrogate:

$$\min_{\boldsymbol{\gamma}} \mathrm{E}\left[\left\| \mathbf{s}_t - \sum_i \gamma_i g_i(\mathbf{y}_t^1, \ldots, \mathbf{y}_t^M, \mathbf{Y}^{t-1}) \right\|^2 \middle| \mathbf{Y}^{t-1}\right] + \beta \|\boldsymbol{\gamma}\|_l, \tag{5.4}$$

where $\beta \geq 0$, $\|\boldsymbol{\gamma}\|_l$, $0 \leq l \leq 1$ is a sparseness inducing penalty and $\{g_i\}$ is a set of base predictors of $\mathbf{s}_t$ and the linear combination of these predictors approximates the exact solution in (5.3). When $\|\boldsymbol{\gamma}\|_l = \|\boldsymbol{\gamma}\|_0$ is the l_0-norm, $g_i(\mathbf{y}_t^1, \ldots, \mathbf{y}_t^M, \mathbf{y}_{t-1}) = g_i(\mathbf{y}_t^{i_1}, \ldots, \mathbf{y}_t^{i_p}, \mathbf{y}_{t-1}) = \widehat{\mathbf{s}}_t$ in (5.3), i indexes over the $\binom{M}{p}$ combinations of indices $i_1, \ldots, i_p$, the solution of (5.4) yields the optimal solution (5.2) for sufficiently large β. A surrogate investigated by many [40,164] for the l_0-norm penalty is the l_1-norm penalty $\|\boldsymbol{\gamma}\|_1$ which will be adopted here. In the special case that g_i depends only on a single variable $\mathbf{y}_t^i$ the regression in (5.4) is equivalent to using a simple generalized additive model (GAM) [64]. We further assume that $g_i(\mathbf{y}_t^i) = \mathrm{E}\left[\mathbf{s}_t | \, \mathbf{y}_t^i, \mathbf{y}_{t-1}\right]$. Thus the constrained prediction problem can be formulated as

$$\min_{\boldsymbol{\gamma}} \mathrm{E}\left[\left\|\mathbf{s}_t - \sum_{i=1}^{M} \gamma_i \mathrm{E}\left[\mathbf{s}_t | \mathbf{y}_t^i, \mathbf{Y}^{t-1}\right]\right\|^2 \middle| \mathbf{Y}^{t-1}\right] + \beta\|\boldsymbol{\gamma}\|_1, \tag{5.5}$$

and β is chosen such that exactly p out of the M γ_i's are nonzero. This quadratic optimization in $\boldsymbol{\gamma}$ under l_1-norm constraint is a convex problem and can be evaluated in a straightforward fashion using standard techniques, e.g., [40,48,114,164]. We first find the range of β that gives rise to a sparse solution with exactly p nonzero elements and fix it to that value in the range which gives the minimum unconstrained error. We take the indices of the p nonzero components of $\boldsymbol{\gamma}$ corresponding to this β as the solution to the waveform subset selection problem in (5.2).

5.4 Comparison to previous strategies

We contrast our approach to the previous offline strategies in Fig. 5.1. In algorithms such as MP and BP, the library of waveforms are chosen to minimize the worst case error in the approximation of the function f chosen from the functional space $\mathcal{P}$. In contrast, we design an online waveform library based on the observed measurements that can be best used to predict the state at the next time instant.

Matching Pursuit (MP)

- Function to be approximated: $f \in \mathcal{Q}$, where $\mathcal{Q}$ denotes the functional space.
- Basis functions: $\{g_i\} \in \mathcal{Q}$.
- For choosing the optimal waveform library:

$$\min_{\boldsymbol{\alpha}} \max_{f \in \mathcal{P}} \|f - \sum_i \alpha_i g_i\|^2 + \lambda \|\boldsymbol{\alpha}\|_0.$$

- For finding the subset selector for a fixed function f_0:

$$\min_{\boldsymbol{\alpha}} \|f_0 - \sum_i \alpha_i g_i\|^2 + \lambda \|\boldsymbol{\alpha}\|_0.$$

Our approach

- State variable to be predicted: $\mathbf{s}_t$ belongs to the probability space $(\Sigma, \mathcal{F}, \mathcal{P})$.
- Basis functions: $\{g_i(\mathbf{Y}^{t-1})\} \in (\Sigma, \mathcal{F}, \mathcal{P})$ depend on the past observations $\mathbf{Y}^{t-1}$.
- The optimal subset selector chooses $\boldsymbol{\alpha}$ to

$$\min_{\boldsymbol{\alpha}} \mathrm{E}\left[\|\mathbf{s}_t - \sum_i \alpha_i g_i(\mathbf{Y}^{t-1})\|^2 | \mathbf{Y}^{t-1}\right] + \lambda \|\boldsymbol{\alpha}\|_0.$$

Figure 5.1: A comparison of our approach with previous offline methods.

The optimal subset selector is chosen to minimize the average error, which implies that the optimal choice of waveforms depends on the higher order statistics between the state and the base predictors.

5.5 Numerical study

To illustrate our approach, we consider the following problem: At time $t = 1$, we assume without loss of generality that an arbitrary waveform index η from $\{1, \ldots, M\}$ is chosen and waveform $\boldsymbol{\phi}_\eta$ is transmitted into the medium. The received signal at the first stage can then be written as

$$\mathbf{y}_1 = \mathbf{L}(\boldsymbol{\phi}_\eta)\mathbf{s}_1 + \mathbf{n}_1 = \mathbf{L}_\eta \mathbf{s}_1 + \mathbf{n}_1, \tag{5.6}$$

where $\mathbf{L}(\cdot)$ is based on the channel model, $\mathbf{n}_1$ is receiver noise, and $\mathbf{s}_1$ is the initial state. We consider the state update equation as a hidden Markov model (HMM), equivalent to a Gaussian mixture model, defined as

$$\mathbf{s}_t \;=\; \mathbf{A}\;\mathbf{s}_{t-1} + \mathbf{I}_t\;\mathbf{w}_{1,t} + (1-\mathbf{I}_t)\;\mathbf{w}_{0,t},\;\; t = 2,3,\ldots, \tag{5.7}$$

where $\{\mathbf{w}_{i,t}, i=0,1\}_t$ are independent Normal random vectors with mean $\boldsymbol{\mu}_i$ and covariance matrix $\mathbf{R}_{\mathbf{w}_i}$, $\mathbf{A}$ is a fixed matrix and $\mathbf{I}_t$ are i.i.d Bernoulli random variables with success probability q.

Assume the initial state $\mathbf{s}_1$ is a Normal random vector with zero mean and covariance matrix $\mathbf{R}_{\mathbf{s}}$. Receiver noises $\{\mathbf{n}_t\}$ are i.i.d Normal with zero mean and covariance matrix $\mathbf{R}_n$ and $\{\mathbf{n}_t, \{\mathbf{w}_{i,t}, i=0,1\}, \mathbf{I}_t, \mathbf{s}_1\}$ are all independent. The model (5.7) captures the non-Gaussian nature of the tracking problem where the state dynamics switch at random between the hidden states $\mathbf{I}_t = 1$ and $\mathbf{I}_t = 0$. The received signal at time $t=2$ corresponding to transmission of waveform $\boldsymbol{\phi}_i$ can be written as

$$\mathbf{y}_2^i \;=\; \mathbf{L}_i\mathbf{s}_2 + \mathbf{n}_2^i,\;\; i = 1,\ldots,M. \tag{5.8}$$

Our goal is to maximize expected reduction in the variance of the state estimator after sending the waveforms $\{\boldsymbol{\phi}_{i_k}\}_{k=1}^p$ and receiving the backscatter $\mathbf{y}_2^{i_1},\ldots,\mathbf{y}_2^{i_p}$, i.e.,

$$\min_{i_1,\ldots,i_p} \mathrm{E}\left[\left\|\mathbf{s}_2 - \mathrm{E}\left[\mathbf{s}_2 \middle|\, \mathbf{y}_2^{i_1},\ldots,\mathbf{y}_2^{i_p},\mathbf{y}_1\right]\right\|^2 \middle|\, \mathbf{y}_1\right]. \tag{5.9}$$

For the proposed GAM prediction problem under l_1-norm constraint, we need to minimize

$$\mathrm{E}\left[\left\|\mathbf{s}_2 - \sum_{i=1}^{M}\gamma_i \mathrm{E}\left[\mathbf{s}_2 \middle|\, \mathbf{y}_2^i,\mathbf{y}_1\right]\right\|^2 \middle|\, \mathbf{y}_1\right] + \beta\|\boldsymbol{\gamma}\|_1 \tag{5.10}$$

with respect to $\boldsymbol{\gamma}$ and use the nonzero indices obtained through this method as our solution to the subset selection problem.

Given $\mathbf{I}_2 = k \in \{0, 1\}$, the random vectors $\mathbf{x}_2$, $\mathbf{y}_2^i$ and $\mathbf{y}_1$ are jointly Gaussian. Let $\mathbf{y} = [{\mathbf{y}_2^{i_1}}^T, \ldots, {\mathbf{y}_2^{i_p}}^T, {\mathbf{y}_1}^T]^T$. Then the joint distribution can be written as

$$\left(\begin{array}{c} \mathbf{s}_2 \\ \mathbf{y} \end{array} \right)_{\mathbf{I}_2=k} = \mathcal{N}\left[\left(\begin{array}{c} \boldsymbol{\mu}_k \\ \boldsymbol{\mu}_{\mathbf{y},k} \end{array} \right), \left(\begin{array}{cc} \mathbf{R}_{\mathbf{s}_2,k} & \mathbf{R}_{\mathbf{s}_2,k,\mathbf{y}} \\ \mathbf{R}^H_{\mathbf{s}_2,k,\mathbf{y}} & \mathbf{R}_{\mathbf{y},k} \end{array} \right) \right], \tag{5.11}$$

where

$$\boldsymbol{\mu}_{\mathbf{y},k} = \left[\begin{array}{c} \left[\mathbf{L}_{i_1}^H, \ldots, \mathbf{L}_{i_p}^H \right]^H \boldsymbol{\mu}_k \\ 0 \end{array} \right], \tag{5.12}$$

$$\mathbf{R}_{\mathbf{s}_2,k,\mathbf{y}} = \left[(\mathbf{R}_{\mathbf{s}_2,k}) \left[\mathbf{L}_{i_1}^H, \ldots, \mathbf{L}_{i_p}^H \right], \mathbf{A}\mathbf{R}_s \mathbf{L}_\eta^H \right], \tag{5.13}$$

$$\mathbf{R}_{\mathbf{s}_2,k} = \mathbf{R}_{\mathbf{w}_k} + \mathbf{A}\mathbf{R}_s\mathbf{A}^H. \tag{5.14}$$

If $\mathbf{y}_1$ is a $N \times 1$ vector, then $\mathbf{R}_y$ is a $N(p+1) \times N(p+1)$ matrix whose mn-th block is given by

$$\begin{aligned} \mathbf{R}_{\mathbf{y},k_{m,n}} &= \mathbf{L}_{i_m}\mathbf{R}_{\mathbf{s}_2,k}\mathbf{L}_{i_n}^H + \mathbf{R}_n\delta(m-n), 1 \le m, n \le p \\ \mathbf{R}_{\mathbf{y},k_{m,p+1}} &= \mathbf{R}^H_{\mathbf{y},k_{p+1,m}} = \mathbf{L}_{i_m}\mathbf{A}\mathbf{R}_\mathbf{s}\mathbf{L}_\eta^H,\ 1 \le m \le p. \\ \mathbf{R}_{\mathbf{y},k_{p+1,p+1}} &= \mathbf{L}_\eta\mathbf{R}_\mathbf{s}\mathbf{L}_\eta^H + \mathbf{R}_n. \end{aligned}$$

Since the random vectors $\mathbf{s}_2, \mathbf{y}_2^{i_1}, \ldots, \mathbf{y}_2^{i_p}, \mathbf{y}_1$ are jointly Gaussian, the conditional mean of $\mathbf{s}_2$ given $\mathbf{y}$ and $\mathbf{I}_2 = k$ can be evaluated as

$$\begin{aligned} &\mathrm{E}\left[\mathbf{s}_2 | \, \mathbf{y}_2^{i_1}, \ldots, \mathbf{y}_2^{i_p}, \mathbf{y}_1, \mathbf{I}_2 = k \right] \\ &\qquad = \boldsymbol{\mu}_k + \mathbf{R}_{\mathbf{s}_2,k,\mathbf{y}}\mathbf{R}_{\mathbf{y},k}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y},k} \right), \end{aligned}$$

and the conditional mean estimator is

$$\begin{aligned} \mathrm{E}\left[\mathbf{s}_2 | \, \mathbf{y}_2^{i_1}, \ldots, \mathbf{y}_2^{i_p}, \mathbf{y}_1 \right] &= \sum_{k=0}^{1} \mathrm{E}\left[\mathbf{s}_2 | \, \mathbf{y}_2^{i_1}, \ldots, \mathbf{y}_2^{i_p}, \mathbf{y}_1, \mathbf{I}_2 = k \right] \\ &\qquad \mathrm{P}\left(\mathbf{I}_2 = k | \, \mathbf{y}_2^{i_1}, \ldots, \mathbf{y}_2^{i_p}, \mathbf{y}_1 \right), \end{aligned} \tag{5.15}$$

where the conditional probability of $\mathbf{I}_2$ can be found using Bayes formula:

$$\begin{aligned} \Pi_k(\mathbf{y}) &= \mathrm{P}\left(\mathbf{I}_2 = k|\,\mathbf{y}\right) \\ &= \mathrm{P}\left(\mathbf{I}_2 = k|\,\mathbf{y}_2^{i_1}, \ldots, \mathbf{y}_2^{i_p}, \mathbf{y}_1\right) \\ &= \frac{f(\mathbf{y}|\,\mathbf{I}_2 = k)\mathrm{P}(\mathbf{I}_2 = k)}{\sum_i f(\mathbf{y}|\,\mathbf{I}_2 = i)\mathrm{P}(\mathbf{I}_2 = i)}, \end{aligned} \tag{5.16}$$

where $f(\mathbf{y}|\,\mathbf{I}_2 = k)$

$$= \frac{|\mathbf{R}_{\mathbf{y},k}|^{-0.5}}{(\sqrt{2\pi})^{N/2}} \exp\left(-0.5(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y},k})^H \mathbf{R}_{\mathbf{y},k}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y},k})\right)$$

and $\mathrm{P}(\mathbf{I}_2 = 1) = q$. Thus equation (5.15) can be rewritten as

$$\mathrm{E}\left[\mathbf{s}_2|\,\mathbf{y}\right] = \sum_{k=0}^{1} \Pi_k(\mathbf{y}) \left(\boldsymbol{\mu}_k + \mathbf{R}_{\mathbf{s}_2,k,\mathbf{y}} \mathbf{R}_{\mathbf{y},k}^{-1} \left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y},k}\right)\right). \tag{5.17}$$

The MSE criterion in (5.9) can now be evaluated by substituting for the conditional expectation from (5.17). For the suboptimal criterion in (5.10), we need to find $\mathrm{E}\left[\mathbf{s}_2|\,\mathbf{y}_2^i, \mathbf{y}_1\right]$ which is a specific case of (5.17) with $p = 1$, i.e., $\mathbf{y} = [\mathbf{y}_2^i, \mathbf{y}_1]$. It is worthwhile to note that even in the case of $q = 0$ or 1 for which the target dynamics are linear Gaussian, the solution to (5.10) is suboptimal, i.e., it is not equivalent to the conditional expectation (5.3). This is because the predictor does not take into account the spatial correlation between the received signals $\mathbf{y}_2^1, \ldots, \mathbf{y}_2^M$. However, if the received signals are scalars, then the l_1-norm constrained solution to (5.10) can be shown to be the optimal solution for the Gaussian case.

5.6 Computational complexity

The estimator given in (5.17) is in closed-form and hence the major complexity in finding the optimal solution is in its evaluation for all $\binom{M}{p}$ possible combinations of waveforms. Instead we use the suboptimal solution given by (5.10) to find the best p waveforms to be transmitted at the second stage. We use the recently proposed

LARS algorithm (Least Angle Regression) [48] to solve for (5.10) which requires only the same order of magnitude of computational effort as the ordinary least squares solution. The algorithm uses the fact that the solution to (5.10) is piecewise linear in β and hence one can obtain the exact solution in $\min(p, M-p)$ steps either by doing a forward selection or backward elimination procedure.

5.7 Simulation results

Based on the formulation in Section 5.5, we perform a simulation for the simple case of $M = 5$ different waveforms. This will allow us to quantify the gap between the optimal solution (5.3) and the solution to the approximation (5.10). We assume a radar receiver array with $N = 25$ antenna elements so that the received signals $\mathbf{y}_1, \mathbf{y}_2$ are 25×1 vectors. The state vector is assumed to be a $N_s \times 1$ vector with $N_s = 10$. The correlation matrices $\mathbf{R}_n, \mathbf{R}_{\mathbf{w}_0}, \mathbf{R}_{\mathbf{w}_1}, \mathbf{R}_{\mathbf{s}}$ are identity matrices. The mean vectors $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ are 10×1 vectors consisting of all zeros and all 0.1 respectively. The Bernoulli random variables $\mathbf{I}_t$ takes the value 1 with probability $q = 0.4$. We assumed the channel model to be linear and selected the waveforms $\{\boldsymbol{\phi}_i\}_{i=1}^{M}$ at random over 25 dimensional unit sphere. These waveforms are unit norm and have cross correlation less than 0.1. We simulated the performance of the optimal subset selector along with the l_1-norm constrained convex problem under this setting. The performance criteria considered in the simulations is shown in Table 2.2.

We first present the MSE of the l_1-norm penalized solution found from (5.5) (solid line, GAM with l_1) as a function of the sparseness regularization parameter β in Fig. 5.2. For each value of β, we also show the corresponding l_0-norm of optimal $\boldsymbol{\gamma}$ (on top of the solid line) in the figure. The MSE is a increasing function of β and as explained earlier, we notice that increasing β induces more sparseness in the

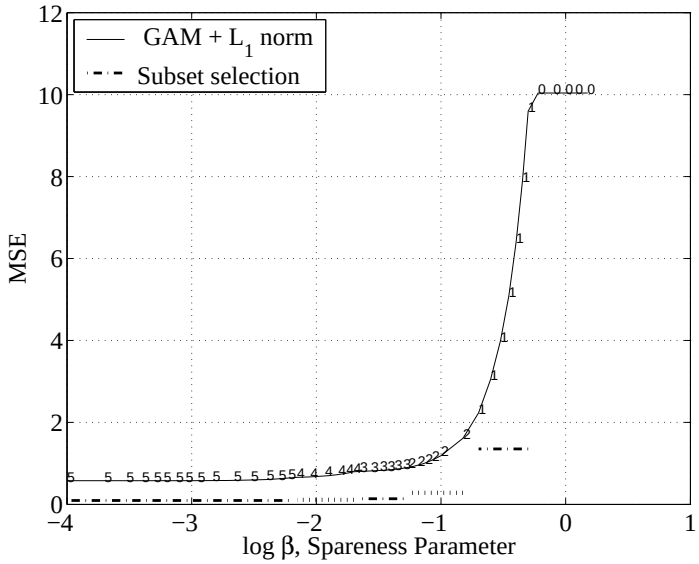


Figure 5.2: Minimum MSE for optimal subset selection (dotted and dashed-dotted line) and l_1-norm constrained solution (solid line)with respect to β. $\|\boldsymbol{\gamma}\|_0$, corresponding to the number of nonzero components in the optimal solution of $\boldsymbol{\gamma}$ for constrained optimization is shown adjacent to the solid line as a function of β.

solution. When β is large, the MSE converges to the variance of the state parameter. We also plot the MSE of the optimal subset selection solution (dashed line) corresponding to the l_0-norm obtained through the l_1-norm constrained solution. We see a clear difference in performance between the two techniques. This is because of two main reasons: The primary reason is the fact that we find a suboptimal solution by assuming the GAM estimator of the form in (5.5) rather than the optimal estimator given in (5.3). The other reason is due to the fact that we solve the minimization problem subject to an l_1-norm constraint rather than an l_0-norm constraint.

In Fig. 5.3, we plot the performance of state estimators mentioned in Table 5.1. We observe that the performance of GAM under l_0-norm constraint is indeed found to be optimal for $\|\boldsymbol{\gamma}\|_0 = 1$ case and clearly suboptimal for other cases due to the restrictive additive model. Finally we see that our proposed solution has a significant performance gain as compared to the simple l_1-norm constrained minimization and approaches the optimal subset selection performance. This suggests that we can

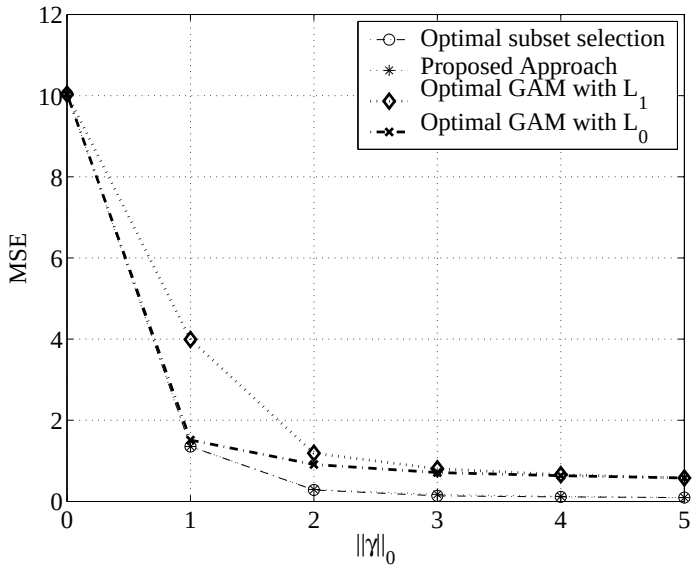


Figure 5.3: Minimum MSE for the optimal subset selection problem (circle), optimal GAM with l_1 constraint (diamond), optimal GAM with l_0 constraint (cross) and the proposed approach (star) as a function of $\|\boldsymbol{\gamma}\|_0$.

considerably reduce the computational complexity of the problem and at the same time achieve nearly optimal performance using such a design approach.

5.8 Conclusions

We considered the problem of optimal waveform selection. We optimally choose a small subset of waveforms that minimizes the state prediction MSE given the past

Approach	Form of predictor	Constr.
Subset Selection	$\mathrm{E}\left[\mathbf{s}_2 \middle\vert \mathbf{y}_2^{i_1}, \ldots, \mathbf{y}_2^{i_p}, \mathbf{y}_1\right]$	-
GAM + l_0	$\sum_i \gamma_i \mathrm{E}\left[\mathbf{s}_2 \middle\vert \mathbf{y}_2^{i}, \mathbf{y}_1\right]$	$\|\boldsymbol{\gamma}\|_0$
GAM + l_1	$\sum_i \gamma_i \mathrm{E}\left[\mathbf{s}_2 \middle\vert \mathbf{y}_2^{i}, \mathbf{y}_1\right]$	$\|\boldsymbol{\gamma}\|_1$
Optimal predictor	$\mathrm{E}\left[\mathbf{s}_2 \middle\vert \mathbf{y}_2^{1}, \ldots, \mathbf{y}_2^{M}, \mathbf{y}_1\right]$	-
Proposed Solution	Use optimal from GAM + l_1 in subset selection	$\|\boldsymbol{\gamma}\|_1$

Table 5.1: Form of predictors

observations. We observe that the optimal subset selection is a combinatorially complex optimization problem and hence infeasible. We proposed a suboptimal solution through convex relaxation which achieves near optimal performance. We considered a particular model and compared the performance of the various strategies through simulation. This problem is a natural extension to the problem of optimal energy allocation under energy constraints using sequential design strategies presented in Chapters II and III. It would be worthwhile to explore the possibilities for designing a waveform selector together with optimal energy allocation.

CHAPTER VI

Sparsity penalized MDS for blind tracking in sensor networks

In this chapter, we consider the problem of tracking a moving target using sensor network measurements. We assume no prior knowledge of the sensor locations and so we refer to this tracking as 'blind'. We use the distributed weighted multidimensional scaling (dwMDS) algorithm to obtain estimates of the sensor positions. Since dwMDS can only recover sensor position estimates up to rotation and translation, there is a need for alignment of sensor positions from one time frame to another. We introduce a sparsity constraint to dwMDS to align current time sensor positions estimates with those of the previous time frame. We formulate a local implementation of the sparsity penalized MDS algorithm using optimization transfer to minimize communication costs. In the presence of a target, location estimates of sensors in the vicinity of the target will vary from their initial values. Based on the differences in the sensor location estimates between two time-frames, we propose a novel perturbation based link level tracking algorithm which identifies a set of senors localizing a target. Using a detailed numerical study, we show that perturbation based link level tracking outperforms LRT based target localization. Applications of the algorithm to two real world data sets are discussed in detail.

6.1 Introduction

Wireless sensor networks have been deployed for a number of monitoring and control applications such as target tracking [94], environmental monitoring [95], manufacturing logistics [89], geographic routing, and precision agriculture [169]. For many target tracking applications such as anomaly detection [67, 170], species distribution and taxonomy [59], and surveillance [24], the main purpose of the sensor network is to locate and track changes in remote environments. For example, species distribution and classification are currently documented using sightings, captures, and trap locations, which involve considerable manpower, time, and effort. Deploying mobile sensors with cameras can improve remote counts of the species as they move around in the environment. For surveillance applications, the sensors must be able to locate where the intruders or the vehicles are moving in the network. Another example is the problem of locating equipment in a warehouse. The sensors that tag the equipment must register their physical locations and activate an alarm if they are about to exit the building. As another example, in secure protocol and network routing it is critical to track anomalies such as worm activity, flash crowds, outages, and denial-of-service (DOS) attacks in the network.

Automatic self-configuration and self-monitoring of sensor networks is the key enabling technology for these tracking applications. To respond to changes in the sensor network, it is critical to know where the changes are occurring. Data measurements from the sensors must be registered to their physical locations in the network in order to make optimal decisions. For dense sensor networks, the large size makes it impractical for humans to manually enter the physical location of the sensors and it is too expensive to attach the GPS to every device in the network. The sensors must

have the capabilities to automatically estimate their relative positions and detect changes in the network at low cost, e.g., with minimum battery power.

Self-localization algorithms can be broadly classified into two categories, centralized strategies and decentralized strategies. In a centralized approach, all the data collected by the sensors must be communicated to the fusion center which then makes a decision based on this information. Algorithms that use multidimensional scaling (MDS) [146], maximum likelihood estimation [108], and convex optimization [45] have been proposed for centralized estimation and have shown to perform well. However, this may be impractical when the sensors operate with limited power and bandwidth. For networks with thousands of sensors, transmission of sensor data to a fusion center overwhelm the low-bandwidth capacity of sensor networks. Furthermore, remote sensors are frequently battery operated and battery replacement may be infeasible or expensive.

The need to conserve power and bandwidth has set the stage for more efficient decentralized strategies for localization. Among the popular approaches are adaptive trilateration [111,139] and successive refinement [35,72] algorithms. In trilateration, each sensor gathers information about its location with respect to anchor nodes, also referred to as seeds [109], through a shortest path. Using the range estimates from the seeds, a sensor uses trilateration to estimate its location in the network. In successive refinement algorithms, each sensor localizes its position in its own coordinate system based on the information communicated from only its neighbors. Sensors refine their location estimates iteratively using updates from neighboring sensors and finally merge their local coordinates systems, effectively finding the solution to the localization problem.

The dwMDS algorithm proposed in [35] is one such successive refinement algo-

rithm, where a global cost function is divided into multiple local cost functions at each sensor location and the computational load involved in finding the sensor location estimates is divided among the sensors in a distributed fashion. The allocation of non-negative continuous weights to the measured data overcomes the problem of combining local maps to one global map, a problem that is common to other decentralized methods [72].

Recently, there has been research emphasis on localization based on a moving target, called a 'mobile' in [28, 117, 161]. The mobile moves randomly in the network while transmitting signals thereby allowing the sensors to estimate their range to the mobile. This provides a large number of measurements with greater diversity which helps overcome environmental obstacles and enables improved estimation of the sensor node locations.

Sensor localization is frequently viewed as an essential prelude to the monitoring and tracking of active phenomena. Target tracking and detection has been one such motivating application of sensor networks [4, 73, 85, 168]. Most target tracking applications assume known sensor locations or estimate the location of sensor nodes separately before employing the tracking algorithm. The standard model used for describing the state dynamics of a moving target is the linear Gaussian model [120]. When the measurement model is also Gaussian, the optimal tracker is given by the Kalman filter. For nonlinear state space and measurement models, other techniques such as extended Kalman filter (EKF) [73], unscented Kalman filter (UKF) [168], and Gaussian sum approximation [4] have been proposed. Particle filtering algorithms were then formulated for target tracking, where the probability density of the state is approximated by a point mass function on a set of discrete points [44]. The discrete points are chosen through importance sampling. The advantage of particle

filtering is its applicability to a large range of densities, noise processes, and measurement models. Most prior work on tracking consider a model-based approach, which requires a detailed probabilistic model of the unknown target dynamics, more sensed information, and is computationally intensive.

More recently, researchers have looked at the simpler problem of tracking in a binary sensing modality [6, 82]. The sensor outputs a high value, when the target is within a sensing range and outputs a low value, when the target falls outside its range. Based on the fusion of the sensor outputs, an approximate link level trajectory can be realized to track the target. Link level tracking has many attractive features, the most important of which is that it does not require a dynamical model for the target, which is fundamental to most tracking algorithms in the literature [11]. This approach for a simple binary sensing measurement model is shown to require minimal power and is also analytically tractable [149]. Moreover, the goal of certain sensor networks is to obtain an estimate of the location of the targets, or detect changes in the network. For example, in military applications, the sensors can locate a target relative to the network and the network can activate the appropriate sensors to identify the target. For animal tracking in biological research, it may be sufficient to have a low resolution tracking algorithm to monitor animal behavior and interactions with their own clan and with other species. This procedure can also be interpreted as a target detection problem implemented for multiple time steps.

Distributed target detection methods have been proposed in the literature [112] in the context of designing an optimal decision statistics at the sensor fusion center. The detection problem has also been addressed for under communication constraints, where the sensor transmitting the information needs to send an optimal summary of the gathered information to the fusion center [29]. In the context of anomaly

detection in internet data, approximate density of incoming traffic is constructed for each location. Distance between densities is then used as a similarity measure in the MDS algorithm to form a map of the internet network. By performing MDS over time, it is shown that anomalies such as network scans, worm attacks, and denial of service attacks can be identified and classified [50, 119]. For wormhole detection in ad hoc sensor networks, most research efforts require mobile nodes equipped with special hardware or GPS devices [26, 68] to localize the wormhole.

In contrast to previous approaches, we propose a sparsity penalized MDS algorithm which localizes the sensor nodes and performs link level tracking when no prior information about the sensor locations are available. The principle behind our proposed algorithm is the following: in the 'acquisition phase' or initialization, an initial estimate of sensor locations is acquired using the dwMDS algorithm [35]. Any localization algorithm recovers sensor location estimates only up to a rotation, reflection, and translation in the absence of anchor nodes, i.e., certain sensors which have knowledge of their positions. This is because the distance information used for finding sensor location estimates depends only on the differences in the sensor locations so that the positions of the sensors in the network can be rotated and translated without changing these distances. However, once the sensors have been initially localized, it is only the relative sensor locations that are critical to the problem of target detection. Hence, during the tracking phase, we introduce a sparsity constraint to the dwMDS problem formulation, which attempts to fix the alignment of the sensor network with respect to the alignment of the localized network at the previous time instance. By doing so, we keep monitoring the network with respect to a fixed geometry obtained by the localization algorithm at the first time instance ($t = 1$). We rely on the fact that accurate estimates of the sensor locations are

obtained during the acquisition phase, i.e., in the absence of targets. The sparsity constraint only reassigns a small fraction of the sensor locations, while maintaining the locations of remaining sensors close to their previous estimates.

When the sensor network is then used for tracking, only the sensors affected by the presence of a target have their location estimates perturbed, while the rest of the location estimates remain unchanged. Based on the differences in the sensor location estimates between two time-frames, we propose a novel perturbation based link level tracking algorithm, which localizes a target to within a small set of sensor links. Figure 6.1 shows the localization process in the absence of targets. The actual sensor locations are marked as circles and the anchor nodes are highlighted using squares. The sensors communicate among themselves and the anchor nodes to obtain location estimates indicated as crossed circles. Figure 6.2 shows the localization process in the presence of a target. The measurements of the sensor nodes closest to the target are affected and the sensor location estimates appear further apart than they are in reality. This change in the sensor location estimates can be used to perform link level tracking.

In the absence of a target trajectory model, we give a flavor of how this algorithm can be extended to estimate actual target coordinates using standard state space tracking algorithms like Kalman trackers and particle filters. Furthermore, the algorithm we present here can be used to design optimal sensor scheduling strategies for tracking to limit power consumption in sensor networks.

This chapter is organized as follows: Section 6.2 formally introduces the problem of sensor localization. Section 6.3 introduces the classical MDS algorithm and its variations. We then present our sparsity penalized dwMDS algorithm in Section 6.4. In Section 6.5, we explain how this algorithm can be applied for link level tracking

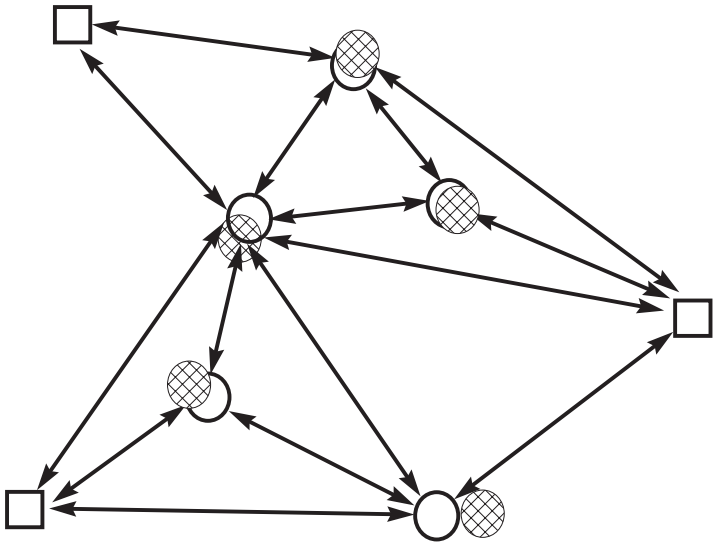


Figure 6.1: Localization in the absence of target. Anchor nodes (square), true sensor locations (circle), estimated sensor coordinates (crossed circle).

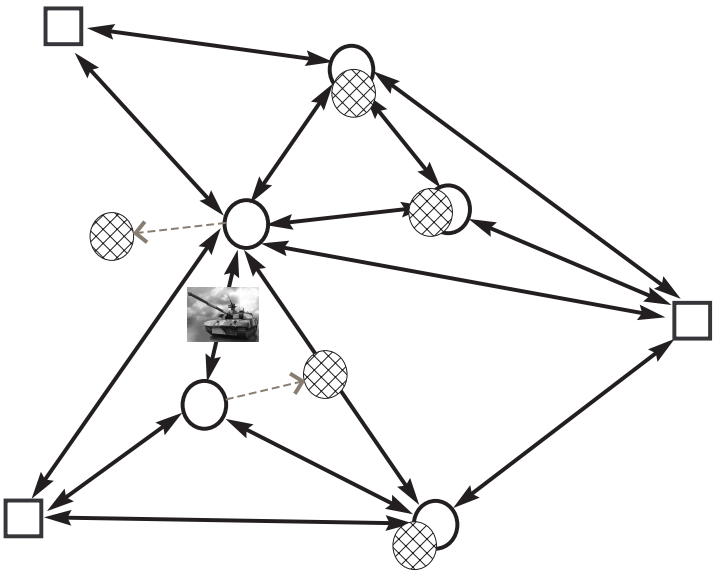


Figure 6.2: Link level tracking based on localization in the presence of target.

and compare its performance with LRT based tracking. In Section 6.6, we apply our algorithm on two real world data sets: the ZebraNet project and the UCSD wireless topology discovery project. Finally, we conclude this chapter in Section 6.7 by discussing the extensions of this formulation for model-based multiple target tracking and sensor management strategies.

6.2 Problem formulation

We begin by introducing the nomenclature used in this chapter. We denote vectors in $\mathbb{R}^M$ by boldface lowercase letters and matrices in $\mathbb{R}^{M\times N}$ by boldface uppercase letters. The identity matrix is denoted by $\mathbf{I}$. We use $(\cdot)^T$ to denote the transpose operator. We denote the l_2-norm of a vector by $\|\cdot\|$, i.e., $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}}$. A Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ is denoted as $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$.

The purpose of the sparsity constrained MDS algorithm is to simultaneously localize and track targets. We first formally state the sensor localization problem. Consider a network of N nodes in d dimensional space. The localization algorithms can be applied to arbitrary d ($d < N$) dimensional spaces. Since applications for localization typically occur in physical space, we will restrict our attention to $d = 2, 3$ dimensions. Let $\{\mathbf{x}_i\}_{i=1}^N, \mathbf{x}_i \in \mathbb{R}^d$ be the true location of the N sensors. The locations of the first $n (n < N)$ sensor nodes are unknown. The remaining $m = N - n$ sensor nodes $\{\mathbf{x}_i\}_{i=n+1}^N$ are anchor nodes, i.e., whose locations are known. We introduce the anchor nodes to keep the formulation as general as possible. Later, we set $m = 0$ for anchor free localization. Denote $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ as the $d \times N$ matrix of actual sensor locations. Let $\mathbf{D} = (d_{i,j})_{i,j=1}^N$ be the matrix of the true inter-sensor distances, where $d_{i,j}$ denotes the distance between sensor i and sensor j. It is common that some wireless sensor networks may have imperfect a priori knowledge about the lo-

cations of certain sensor nodes. This information is encoded through parameters r_i and $\bar{\mathbf{x}}_i$, where $\bar{\mathbf{x}}_i$ is the sensor location and r_i is the corresponding confidence weight. If $\bar{\mathbf{x}}_i$ is unavailable, then we set $r_i = 0$. The problem setting is explained through an illustration of a sensor network in Fig. 6.3. In this sensor network, each sensor communicates to its three nearest neighbors and hence, the weights corresponding to links between non neighboring sensors are zero.

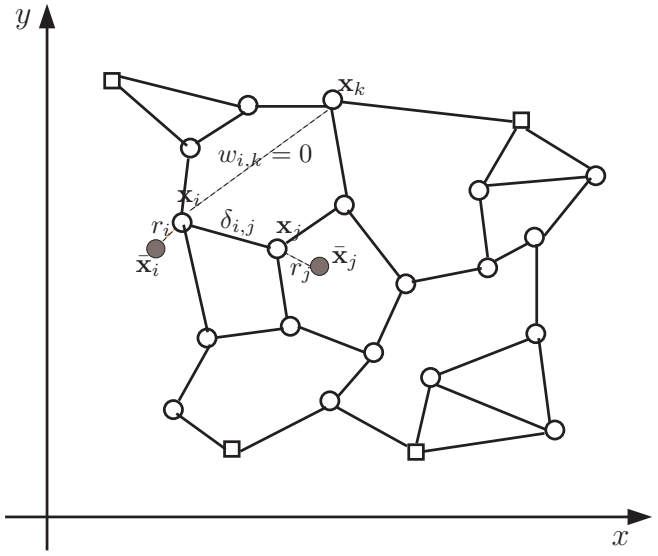


Figure 6.3: Sensor localization setup: Anchor nodes (square), sensor nodes (circle), a priori sensor locations (blocked circle). The communicating sensors are connected using solid lines. The non neighboring sensor links have zero weight.

Sensor localization is the process of estimating the location of the n sensor nodes $\{\mathbf{x}_i\}_{i=1}^{n}$ given $\{\mathbf{x}_i\}_{i=n+1}^{n+m}$, $\{r_i\}$, $\{\bar{\mathbf{x}}_i\}$ and pairwise range measurements $\{\delta_{i,j}^t\}$ taken over time $t = 1, 2, \ldots, K$. The indices (i, j) run over a subset of $\{1, 2, \ldots, N\} \times \{1, 2, \ldots, N\}$. The range measurements can be obtained by sensing modalities such as time-of-arrival (TOA), received signal strength (RSS), or proximity.

6.3 Classical MDS and variations

Multidimensional scaling (MDS) is a methodology for recovering underlying low dimensional structure in high dimensional data. The measured data can come from confusion matrices, group data, or any other (dis)similarity measures. MDS has found numerous applications in cognitive science, marketing, ecology, information science, and manifold learning [39, 42]. In the context of sensor localization, the goal in MDS is to discover the sensor locations (lower dimensional embedding) from the inter-sensor distances obtained by a given sensing method (high dimensional data) [35, 72, 146].

Classical MDS [58] provides a closed-form solution to the sensor locations when the inter-sensor measurements are the inter-sensor Euclidean distances, i.e., in the absence of noise or nonlinear effects. When all pairwise range measurements are available, we can compute the complete matrix of distances:

$$d_{i,j} = \|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)}. \tag{6.1}$$

Denote by $\mathbf{D}^{(2)}$ the matrix of squared distances, i.e., $\mathbf{D}^{(2)} = (d_{i,j}^2)_{i,j=1}^N$. Then $\mathbf{D}^{(2)}$ can be rewritten as

$$\mathbf{D}^{(2)} = \boldsymbol{\psi}\mathbf{1}^T - 2\mathbf{X}^T\mathbf{X} + \mathbf{1}\boldsymbol{\psi}^T, \tag{6.2}$$

where $\mathbf{1}$ is an N-element vector of ones and $\boldsymbol{\psi} = [\mathbf{x}_1^T\mathbf{x}_1, \mathbf{x}_2^T\mathbf{x}_2, \ldots, \mathbf{x}_N^T\mathbf{x}_N]^T$. Let $\mathbf{H} = \mathbf{I} - (1/N)\mathbf{1}\mathbf{1}^T$. Multiplying on the left of $\mathbf{D}^{(2)}$ by $-1/2\mathbf{H}$ and the right by $\mathbf{H}$, we obtain

$$\mathbf{A} = -\frac{1}{2}\mathbf{H}\mathbf{D}^{(2)}\mathbf{H} = \mathbf{H}\mathbf{X}^T\mathbf{X}\mathbf{H}. \tag{6.3}$$

Let $\mathbf{Y} = \mathbf{H}\mathbf{X}$. Given $\mathbf{A}$, one can discover the matrix $\mathbf{X}$ to within a rotation and

translation by solving the following variational problem

$$\min_{\mathbf{Y}} \|\mathbf{A} - \mathbf{Y}^T\mathbf{Y}\|_F^2, \tag{6.4}$$

where $\|\cdot\|_F$ indicates the Frobenius norm and the search space is over all full rank $d \times N$ matrices. The solution to $\mathbf{X}$ is then given by

$$\mathbf{X} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_d^{1/2})\mathbf{V}_1^T, \tag{6.5}$$

where the singular value decomposition (SVD) of $\mathbf{A}$ is given by

$$\mathbf{A} = [\mathbf{V}_1 \ \mathbf{V}_2] \ \text{diag}(\lambda_1, \ldots, \lambda_d, \lambda_{d+1}, \ldots, \lambda_N) \ [\mathbf{V}_1 \ \mathbf{V}_2]^T. \tag{6.6}$$

The matrix $\mathbf{V}_1$ consists of the eigenvectors of the first d eigenvalues $\lambda_1, \ldots, \lambda_d$, while the rest of the $N-d$ eigenvectors are represented as $\mathbf{V}_2$. The term $\text{diag}(\lambda_1, \ldots, \lambda_N)$ refers to a $N \times N$ diagonal matrix with λ_i as its ith diagonal element. Though the solution to the classical MDS is obtained in closed-form, the algorithm has the following deficiencies:

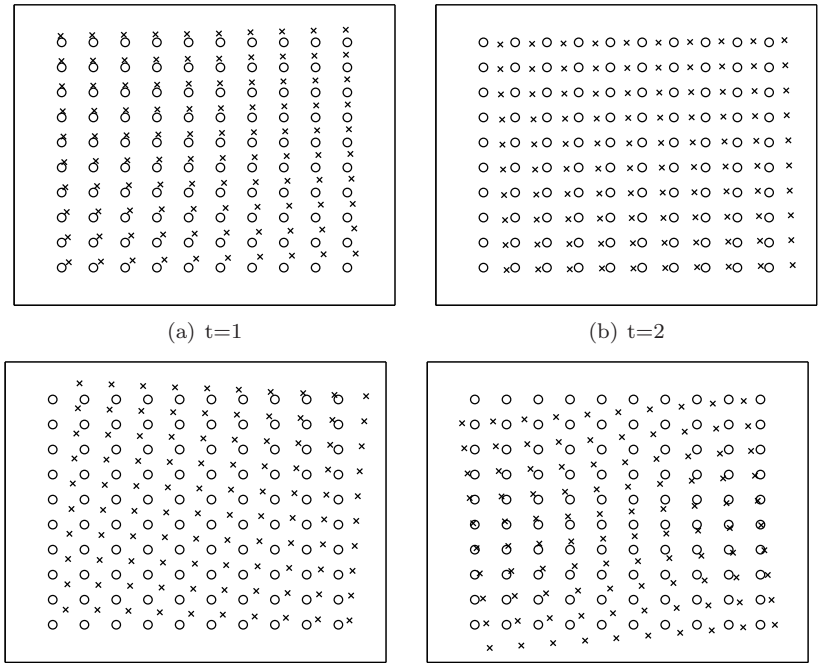
1. MDS requires knowledge of all inter-sensor distances. Obtaining all pairwise range measurements is prohibitive due to the size of the sensor network and the limited power of the sensors. Furthermore, the SVD solution requires the transmission of the range information to the fusion center which then performs the MDS algorithm. Due to power and bandwidth limitations in the sensor network, this process is infeasible.

2. The inter-sensor range measurements $\delta_{i,j}$ are corrupted by environment and receiver noise which further degrades the quality of the measurements, i.e., $\delta_{i,j}$ is only an estimate of the inter-sensor distance $d_{i,j}$.

3. MDS uses the squared distance matrix which tends to amplify the measurement noise, resulting in poor performance.

As mentioned in Section 6.1, there has been significant effort directed towards designing decentralized strategies for sensor localization. However, consistent reconstruction (without rotation or translation) of the sensor locations is attainable only in the presence of anchor nodes. If the current localization algorithms are implemented for anchor free localization, the sensor location estimates can assume different alignments when localization is performed at different time instants. The idea of using perturbations in sensor location estimates between two time frames to identify the presence of a target is infeasible in the absence of an alignment.

To illustrate this phenomenon, we implement the dwMDS algorithm for sensor localization in the absence of anchor nodes and in the absence of target. By absence of target, we mean that the inter-sensor measurements yield a reasonably accurate estimate of the inter-sensor distances. We provide snapshots of the sensor location estimates (cross) along with their actual locations (circle) in Fig. 6.4 as a function of time. Observe that the geometry of the network is maintained, while the true locations are subject to rotation and translation. Now consider a target moving through this network. In this scenario, the localization process is affected by inaccurate inter-sensor measurements in the vicinity of the target. When anchor nodes are present, only those sensor locations in the vicinity of the targets are localized to a position different from previous time. So a time varying dwMDS algorithm can potentially yield accurate target localization to within a small set of sensor links. However, with anchor free localization, such a process becomes infeasible due to the alignment problem. In the following section, we propose a methodology that overcomes this problem.



(a) t=1 (b) t=2

(c) t=3 (d) t=4

Figure 6.4: Anchor free sensor localization by dwMDS. True sensor locations (circle), estimated sensor locations (cross).

6.4 Sparsity penalized MDS

Consider using the MDS algorithm independently to obtain the sensor location estimates at time $t-1$ and at time t. Alignment between these two sets of points can be performed in various ways. For example, in Procrustes analysis [57] alignment is performed by finding the optimal affine transformation of one set of nodes that yields the set closest to the second set of points in the least squares sense. Let us consider this alignment procedure for aligning two sets of sensor location estimates from two time instants. Let at time $t-1$, there was no target and at time t, a target appears in the network. The sensor measurements are altered due to the reflection/attenuation from the target and the sensor localization algorithm yields sensor location estimates that are perturbed from their previous time frame estimates. However, while performing the alignment, this perturbation becomes less pronounced as the smoothing

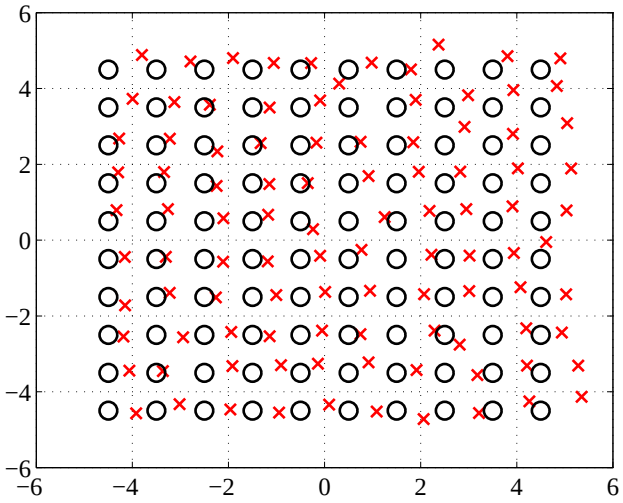


Figure 6.5: Original sensor locations (circle) and location estimates obtained from dwMDS algorithm (cross) are the input data to Procrustes alignment.

penalty (least squares) will distribute the error equally among all the sensor location estimates. Moreover, in the presence of noise, this procedure is only degraded further and the alignment cannot provide unchanged sensor locations estimates from their previously estimated values. The errors in the sensor location estimates between two time steps may also accumulate over time resulting in more alignment errors. As an example, a set of estimated coordinates obtained from the dwMDS algorithm (cross) and the actual coordinates (circle) shown in Fig. 6.5 are aligned using Procrustes analysis. The aligned coordinates as a result of least squares smoothing are shown in Fig. 6.6. Using such an alignment, it is not possible to identify a perturbation of a sensor location as that of the one caused by the presence of the target. Furthermore, such a procedure would require a fusion center to gather all the information to perform this analysis.

In contrast, we introduce a sparseness penalty on the distances between the sensor location estimates $\mathbf{x}_i$ at time t and location estimates $\mathbf{x}_i^{(t-1)}$ at time $t-1$ directly to

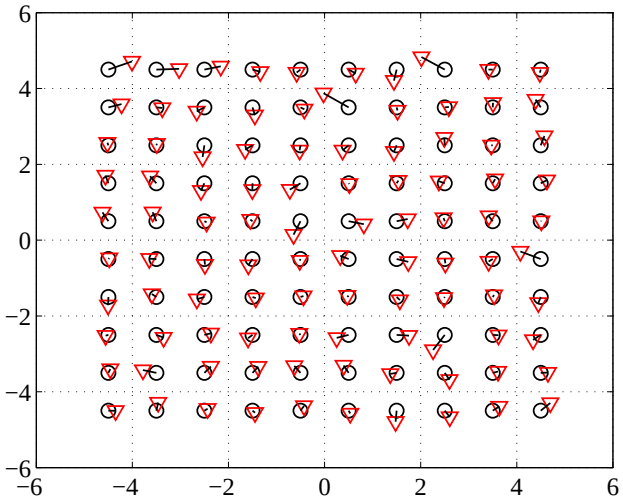


Figure 6.6: Original sensor locations (circle) and aligned sensor location estimates (triangle) obtained from Procrustes analysis.

the sensor localization algorithm. Construct a vector of Euclidean distances between the location estimates at time t and at time $t-1$

$$\mathbf{g}^{(t)} = \left[\|\mathbf{x}_1 - \mathbf{x}_1^{(t-1)}\|, \ldots, \|\mathbf{x}_n - \mathbf{x}_n^{(t-1)}\|\right]^T. \tag{6.7}$$

Define the l_0-measure of a vector $\mathbf{v} = [v_1, v_2, \ldots, v_n]$ as the number of nonzero elements given by

$$\|\mathbf{v}\|_0 \triangleq \sum_{i=1}^{n} I(v_i \neq 0), \tag{6.8}$$

where $I(\cdot)$ is the indicator function. Using an l_0-constraint on the distance vector $\mathbf{g}^{(t)}$ of the form $\|\mathbf{g}^{(t)}\|_0 \leq q$, we guarantee that no more than q of the location estimates will vary from their previous time frame values. Minimizing a cost function under the l_0-constraint requires a combinatorial search which is computationally infeasible. Define the l_p-measure of a vector $\mathbf{v}$ as

$$\|\mathbf{v}\|_p \triangleq \Big(\sum_{i=1}^{n} |v_i|^p\Big)^{1/p}. \tag{6.9}$$

For a quadratic cost function, an l_p-constraint ($0 < p \leq 1$) induces a sparse solution. Among all l_p sparsifying constraints, only $p = 1$ offers a convex relaxation to the l_0-constraint [46]. To promote sparsity, we next advocate the use of the l_p-constraint as a penalty term via the Lagrange multiplier in the dwMDS algorithm to solve for the sensor location estimates. Hence the term *sparsity penalized MDS*.

The cost function of the dwMDS algorithm [35] is motivated by the variational formulation of the classical MDS, which attempts to find sensor location estimates that minimize the inter-sensor distance errors. Keeping in mind that it is the geometry of the sensor network which is crucial for tracking, we present a novel extension of the dwMDS algorithm through the addition of the sparseness inducing l_p-constraint. At any time t, we seek to minimize the overall cost function $C^{(t)}$ given by

$$C^{(t)} = \sum_{1\leq i\leq n} \sum_{i\leq j\leq n+m} \sum_{1\leq l\leq M} w_{i,j}^{(t),l} \left(\delta_{i,j}^{(t),l} - d_{i,j}(\mathbf{X})\right)^2 + \sum_{i=1}^{n} r_i \|\bar{\mathbf{x}}_i - \mathbf{x}_i\|^2 + \lambda \|\mathbf{g}^{(t)}\|_p^p. \tag{6.10}$$

The Euclidean distance $d_{i,j}(\mathbf{X})$ is defined in (6.1). For each time t, there are M range measurements $\delta_{i,j}^{(t),l}$ for each sensor link i, j. As in [35], the weights $w_{i,j}^{(t),l}$ can be chosen to quantify the accuracy of the predicted distances. When no measurement is made between sensor i and sensor j, $w_{i,j}^{(t),l} = 0$. Furthermore, the weights are symmetric, i.e., $w_{i,j}^{(t),l} = w_{j,i}^{(t),l}$, and $w_{i,i}^{(t),l} = 0$. If available, the a priori information of sensor locations is encoded through the penalty terms $\{r_i \|\bar{\mathbf{x}}_i - \mathbf{x}_i\|^2\}$. Finally, we introduce an l_p-constraint ($0 \leq p \leq 1$) on the distances between the sensor locations at time t and the estimated sensor locations at time $t-1$. The Lagrange multiplier of the sparseness penalty is denoted as λ. We can tune the value of λ to yield the desired sparsity level in $\mathbf{g}^{(t)}$. Later, when we apply the algorithm for tracking, the sparseness will be advantageous as only those sensors which are highly affected by

the target will vary from their initial positions, thereby allowing for a detection of the target through the process of relative sensor localization. To solve this optimization problem, we propose to use the successive refinement technique, where each sensor node i updates its location estimate by minimizing the global cost function $C^{(t)}$, after observing range measurements at node i and receiving position estimates from its neighboring nodes.

6.4.1 Minimizing cost function by optimization transfer

Unlike classical MDS for which we could obtain a closed-form expression for the sensor location estimates, there is no closed-form solution to minimizing $C^{(t)}$. The original algorithm for minimizing the least squares MDS cost function in (6.10) with $\lambda = 0$ and $\{r_i = 0\}$ used gradient methods with elaborate step-size procedures [87]. A majorization method for solving the nonlinear least squares problem was introduced in [43]. This procedure can be viewed as a special case of optimization transfer through surrogate objective functions [92] (e.g., the popular EM algorithm) and has shown to work well for sensor localization [35]. In this work, we generalize the SMACOF (scaling by majorizing a complicated function [61]) majorization algorithm by including an alignment penalty to address the problem of sensor localization and tracking.

A majorizing function $T(\mathbf{x}, \mathbf{y})$ of $C(\mathbf{x})$ is a function $T : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, which satisfies the following properties: $T(\mathbf{x}, \mathbf{y}) \geq C(\mathbf{x}),\ \forall \mathbf{y}$ and $T(\mathbf{x}, \mathbf{x}) = C(\mathbf{x})$. In other words, the majorizing function upper bounds the original cost function. Using this property, we can formulate an iterative minimization procedure as follows: denote the initial condition as $\mathbf{x}_0$. Starting from $n = 1$, obtain $\mathbf{x}_n$ by solving

$$\mathbf{x}_n = \arg\min_{\mathbf{x}} T(\mathbf{x}, \mathbf{x}_{n-1}),$$

until a convergence criterion for $C(\mathbf{x})$ is met. We can easily observe that this iterative scheme always produces a non decreasing sequence of cost functions, i.e.,

$$C(\mathbf{x}_{n+1}) \leq T(\mathbf{x}_{n+1}, \mathbf{x}_n) \leq T(\mathbf{x}_n, \mathbf{x}_n) = C(\mathbf{x}_n).$$

The first and last relations follows from the properties of majorizing functions while the middle inequality follows from the fact that $\mathbf{x}_{n+1}$ minimizes $T(\mathbf{x}, \mathbf{x}_n)$. Now the trick is to choose a majorizing function that can be minimized analytically, e.g., a linear or quadratic function. We propose a quadratic majorizing function $T^{(t)}(\mathbf{X}, \mathbf{Y})$ for the global cost $C^{(t)}(\mathbf{X})$. Minimizing $C^{(t)}(\mathbf{X})$ through the majorization algorithm is the simple task of minimizing the quadratic function $T^{(t)}(\mathbf{X}, \mathbf{Y})$, i.e.,

$$\frac{\partial T^{(t)}(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{x}_i} = 0, \quad i = 1, 2, \ldots, n. \tag{6.11}$$

If we denote the estimates of the sensor nodes at iteration k as $\mathbf{X}^k$, the recursion for the update of location estimates for node i from (6.11) is given by

$$\mathbf{x}_i^k = \frac{1}{a_i} \left(c_i + \mathbf{X}^{k-1} \mathbf{b}_i^{k-1} \right), \tag{6.12}$$

where $\mathbf{b}_i^{k-1}$, a_i, and c_i are defined in (6.46)-(6.49) respectively. The details of the derivation of the sparsity penalized MDS algorithm can be found in Section 6.8. For each sensor i, the j^{th} element of the vector $\mathbf{b}_i^{k-1}$ depends on the weight $w_{i,j}$. Since the weights of the nodes not in the neighborhood of the sensor are zero, the corresponding elements in the vector $\mathbf{b}_i^{k-1}$ are also zero; therefore the update rule for node i in (6.12) will depend only the location of its nearest neighbors and not on the entire matrix $\mathbf{X}^{k-1}$. This facilitates the distributed implementation of the algorithm. The proposed algorithm is summarized in Fig. 6.7. We illustrate the majorization procedure in Fig. 6.8. The original cost function (solid) and the corresponding surrogate (dotted) is presented for every iteration, along with the track of the estimates at iteration k (circle).

Inputs: $\{\bar{w}_{i,j}^{(t)}\}$, $\{\bar{\delta}_{i,j}^{(t)}\}$, $\{r_i\}$, $\{\bar{\mathbf{x}}_i\}$, $\{\mathbf{x}_i^{(t-1)}\}$, ϵ, $\mathbf{X}^0$ (initial condition for iterations).

Set $k = 0$, compute cost function $C^{(t),0}$ and a_i from equations (6.10) and (6.48) respectively

repeat

- k=k+1
- for $i = 1$ to n
  - compute $\mathbf{b}_i^{k-1}$ from equation (6.46)
  - $\mathbf{x}_i^k = \frac{1}{a_i}\left(c_i + \mathbf{X}^{k-1}\mathbf{b}_i^{k-1}\right)$
  - compute $C_i^{(t),k}$
  - update $C^{(t),k}$ to $C^{(t),k} - C_i^{(t),k-1} + C_i^{(t),k}$
  - communicate $\mathbf{x}_i^k$ to neighbors of sensor i (nodes for which $w_{i,j} > 0$)
  - communicate $C^{(t),k}$ to next node $((i+1) \mod n)$
- end for

until $C^{(t),k} - C^{(t),k-1} < \epsilon$

Figure 6.7: Description of the sparsity constrained MDS algorithm.

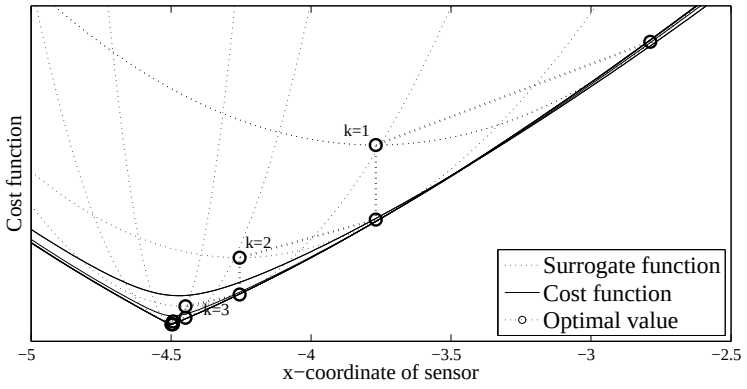


Figure 6.8: Majorization procedure: cost function (solid curve), surrogate function (dotted curve), optimal location estimate at each iteration (circle). Only a single coordinate is updated in this picture.

Our proposed algorithm introduces a sparseness penalty on the distance between estimate at time $t-1$ and the current estimate. If the sparsity regularization parameter λ is not chosen properly, many sensor positions estimates might slowly vary with time, thereby creating cumulative error in the sensor localization. An interesting way to counteract this problem would be to penalize the distance between the current estimate and the initial estimate at $t = 1$. Using such a constraint would mean that the sensors are always compared to the fixed initial frame and errors do not accumulate over time. The implementation of this algorithm would be straightforward as it would simply involve changing the index $t-1$ to 1 in the original algorithm presented in Fig. 6.7. However, using the estimate from time $t-1$ has the property that it is easily adapted to the case of mobile sensors.

6.4.2 Implementation

Weights: When RSS measurements are used to compute distance estimates, the weights are set using the locally weighted regression methods (LOESS) scheme [34] similar to one used in the dwMDS algorithm [35]. The weight assignment is given by

$$w_{i,j} = \begin{cases} \exp\left(-\delta_{i,j}^2/h_{i,j}^2\right), & \text{if i and j are neighbors} \\ 0, & \text{otherwise,} \end{cases}$$

where $h_{i,j}$ is the maximum distance measured by either sensor i or j. A naive equal weight assignment to all measurements is also shown to work well with our algorithm.

Initialization: For the successive refinement procedure, the sensor locations estimates must be initialized for every time frame. Though several initialization algorithms have been proposed in the literature, we use a naive random initialization. We would like to point out that the initialization is not a critical component to our

algorithm, as we are solely interested in the alignment of sensors in the network and not on the exact locations. Irrespective of the initial estimates, the sparseness penalty will ensure that the estimated sensor locations are relatively close to those of previous time frames. In our simulations, we observe that our algorithm is fairly robust with respect to the initial estimates.

Neighborhood selection: Traditionally, the neighbors are chosen based on the distance measure obtained from the RSS measurements, i.e., select all sensors within a distance R as your neighbors. When the RSS measurements are noisy, there is a significant bias in the neighborhood selection rule. This method has a tendency to select sensors which are, on average, less than the actual distances $\|\mathbf{x}_i - \mathbf{x}_j\|$. We use a simple two-stage adaptive neighborhood selection rule proposed in [35] to overcome the effect of this bias:

In the first stage, the dwMDS algorithm is run with a neighborhood structure based on the available range measurements, i.e., set $w_{i,j} = 0$ if $\delta_{i,j} > d_R$ for some reference distance d_R. After convergence, this step provides an interim estimate $\{\hat{\mathbf{x}}_i\}$ of the sensors locations.

In the second step, these predicted distances from the estimated sensor locations are used to compute a new neighborhood structure, by assigning $w_{i,j} = 0$ if $\|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j\| > d_R$. Some neighbors with larger range measurements will be added while some others with low range measurements will be removed. Then, using $\{\hat{\mathbf{x}}_i\}$ as an initial condition and the new neighborhood structure, the dwMDS algorithm is rerun, resulting in the final location estimates. We remark that this 2-step algorithm does not imply twice the computation. The dwMDS algorithm is based on majorization, and each iteration brings the location estimates one step closer to the optimal

estimate. Since the first step only needs to provide coarse localization information, it does not need to be very accurate, and so the dwMDS algorithm can be stopped quickly with a large ϵ. The second step will likely require fewer iterations to converge as we begin with a good set of initial estimates.

Range measurement models: The inter-sensor measurements can be obtained by RSS, TOA, or proximity. Any one of these approaches can be used in our algorithm. Our sparsity constrained dwMDS algorithm is fairly robust to either of these measurement models. For the simulations in this chapter, we use the RSS to obtain a range measurement between two sensors. It can be shown through the central limit theorem (CLT) that the RSS is log-normal in its distribution [36], i.e., if $P_{i,j}$ is the measured power by sensor i transmitted by sensor j in milliWatts, then $10\log_{10}(P_{i,j})$ is Gaussian. Thus $P_{i,j}$ in dBm is typically modeled as

$$\begin{aligned} P_{i,j} &\sim \mathcal{N}(\bar{P}_{i,j}, \sigma_0^2) \\ \bar{P}_{i,j} &= P_0 - 10n_p \log\left(\frac{d_{i,j}}{d_0}\right) \end{aligned} \tag{6.13}$$

where $\bar{P}_{i,j}$ is the mean received power at distance $d_{i,j}$, σ_0 is the standard deviation of the received power in dBm, and P_0 is received power in dBm at a reference distance d_0. n_p is referred to as the path-loss exponent that depends on the multipath in the environment. Given the received power, we use maximum likelihood estimation to compute the range, i.e., distance between the sensor nodes i and j. The maximum likelihood estimator of $d_{i,j}$ is given by

$$\delta_{i,j} = d_0 10^{((P_0 - P_{i,j})/10n_p)}. \tag{6.14}$$

Simulation of tracker without a target

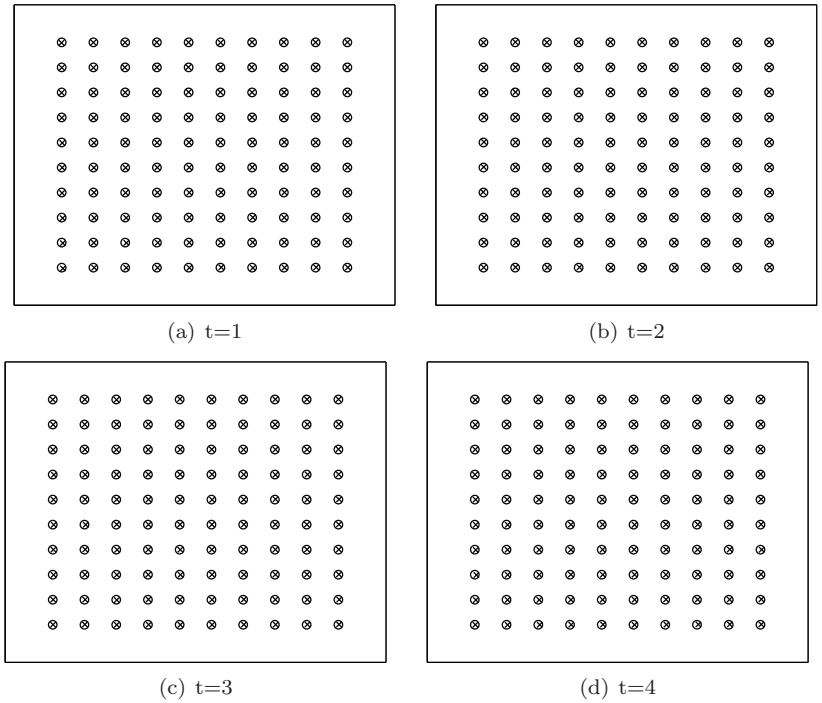
The simulation parameters are chosen as follows: we deploy a 10×10 uniform grid of sensors in a network. We consider anchor free localization, i.e., $m = 0$ and we

assume we make a single inter-sensor measurement ($M = 1$). We set the sparseness parameter λ to produce a change in the location estimates for only a small portion of the sensors. The value of λ will depend on the size of the network and the noise in the measurements. If the RSS measurements are very noisy, then range estimates become inaccurate which tend to vary the sensor location estimates. Hence λ is selected to ensure that sensor location estimates remain aligned with the previous time frame estimates. In this simulation, we set $\lambda = 0.1$ and the noise variance $\sigma_0 = 0.15$. Each sensor communicates with its 15 nearest neighbors. The weights of the RSS measurements were chosen based on the LOESS scheme described earlier. The weights of links for non communicating sensors were set to zero.

We demonstrate the result of the sparsity constrained MDS algorithm on this sensor network as a function of time in Fig. 6.9. The true locations are denoted as circles and the estimated locations as crosses. In contrast to the least squares alignment produced by Procrustes analysis in Fig. 6.6, we observe that we are able to achieve near perfect alignment between two sets of sensor location estimates. The presence of a target will perturb this alignment and in the following section, we discuss the scope for target localization based on this phenomenon.

6.5 Tracking using sparse MDS

Here we present an algorithm for performing link level tracking using the sparsity constrained MDS algorithm. By link level tracking, we refer to localization of targets to within a set of inter-sensor links. Though link level tracking does not need any assumptions on the dynamical target motion model, it is important to know the effect of the target on the inter-sensor measurements. Researchers have proposed various models for the signal strength measurements ranging from the traditional



(a) t=1 (b) t=2

(c) t=3 (d) t=4

Figure 6.9: Anchor free sensor localization by sparsity penalized MDS. True sensor locations (circle), sensor position estimates (cross).

linear Gaussian model to the binary sensing model. These are approximate statistical models and the distribution of the measurements in the presence of a target remains an open question.

To model the statistics under the setting of vehicle tracking, we conducted experiments using RF sensors hardware in the presence of a target. We constructed a fine grid of locations, where the target was placed and RSS measurements were recorded between two static sensors for positions on the grid. A detailed description of the experimental setup is given in Appendix 6.9. Upon gathering the data, we fit the following statistical model in the presence of target. The RSS measurements under this H_1 hypothesis at sensor link i, j are distributed as

$$\begin{aligned} P_{i,j}^k|\hat{P}_{i,j} &\sim \mathcal{N}(\hat{P}_{i,j}, \sigma_0^2), \text{ i.i.d, } k = 1, 2, \ldots, M \\ \hat{P}_{i,j} &\sim \mathcal{N}(\bar{P}_{i,j}, \sigma_1^2), \end{aligned} \tag{6.15}$$

where $P_{i,j}^k$ is the k^{th} inter-sensor measurement when the target is in the neighborhood of the sensors. The M sensor link measurements are correlated through the random variable $\hat{P}_{i,j}$. The values obtained from our actual experiments were $\sigma_0 \approx 0.1463$dBm and $\sigma_1 \approx 1.5$dBm. The noise variance in the measurements σ_1 was roughly an order of 10 times larger than σ_0. In other words, RSS measurements tend to have a larger variance due to scattering and attenuation of the signals in the presence of a target. A confidence measure for such a log-normal distribution of the RSS data is obtained using the Kolmogorov-Smirnov (KS) test in [118] and the model is shown to work well for sensor localization. We assume this statistical model for the RSS measurements, when the target is within a specified distance R of the sensor link i, j. The distance R depends on the reflectivity of the object. If the object is highly reflective, then the variation in the RSS measurements is detected by more links.

Based on the H_0 and H_1 hypothesis given in (6.13) and (6.15) respectively, we formulate the optimal decision statistic to detect a presence of a target in a particular sensor link using the LRT. The LRT for each link i, j is given by

$$\left| \frac{1}{M} \sum_{l=1}^{M} P_{i,j}^{(t),l} - P_{i,j}^{'} \right| \underset{\text{H}_0}{\overset{\text{H}_1}{\gtrless}} \gamma, \tag{6.16}$$

where γ is chosen to satisfy a false alarm level and $P_{i,j}^{'}$ is the mean received power in the sensor link estimated using an initial set of range measurements. We assume that the sensor network is in its steady state operation mode. We do not consider the transient effects in the measured data when it is obtained in the absence of any target. This most powerful test of level α yields the probability of correct detection

$$\beta = 2Q\left(Q^{-1}(\alpha/2) \sqrt{\frac{\sigma_0^2}{\sigma_0^2 + M\sigma_1^2}} \right). \tag{6.17}$$

A derivation of the decision rule and its performance can be found in Appendix 6.10. We show that the performance of the optimal detector is dependent on the

number of samples M available for the inter-sensor measurements. As M becomes very large, the probability of correct detection β tends to 1. However, if only few samples are available, β may not approach 1 and misdetect type errors may become non negligible. In such a case, instead of using the LRT, we can use a test on the variation of the sensor location estimates at time t from their estimates at a previous time τ ($\tau < t$). In other words, we can perform a simple hypothesis test for each link of the form

$$\left\| d_{i,j}^{(t)} - d_{i,j}^{(\tau)} \right\| \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \gamma_{i,j}, \tag{6.18}$$

where $d_{i,j}^{(t)} = \|\mathbf{x}_i^{(t)} - \mathbf{x}_j^{(t)}\|$ and $\{\mathbf{x}_i^{(t)}\}$ are the sensor location estimates obtained from the sparsity penalized MDS algorithm.

Simulation of tracker in the presence of target

We present our results by simulating moving targets in a uniform 10×10 grid of sensors. We set $m = 0$, i.e., no anchor nodes. We assume no a priori knowledge of the sensor coordinates, i.e., $r_i = 0$. Each sensor communicates only to its 15 nearest neighbors and the weights for those links were chosen by the LOESS strategy. The rest of the weights were set to zero. We obtain $M = 50$ data measurements for each communicating sensor link in the network. We set the sparseness parameter λ to produce a change in the location estimates for only a small portion of the sensors. We allow any number of targets to appear in a sensor network with probability 0.4. Though our algorithm is robust to randomly moving targets in the network, we consider a state-space model for the purposes of this simulation to produce a visually pleasing target trajectory. We apply the sparsity constrained MDS algorithm as multiple targets move through the sensor network.

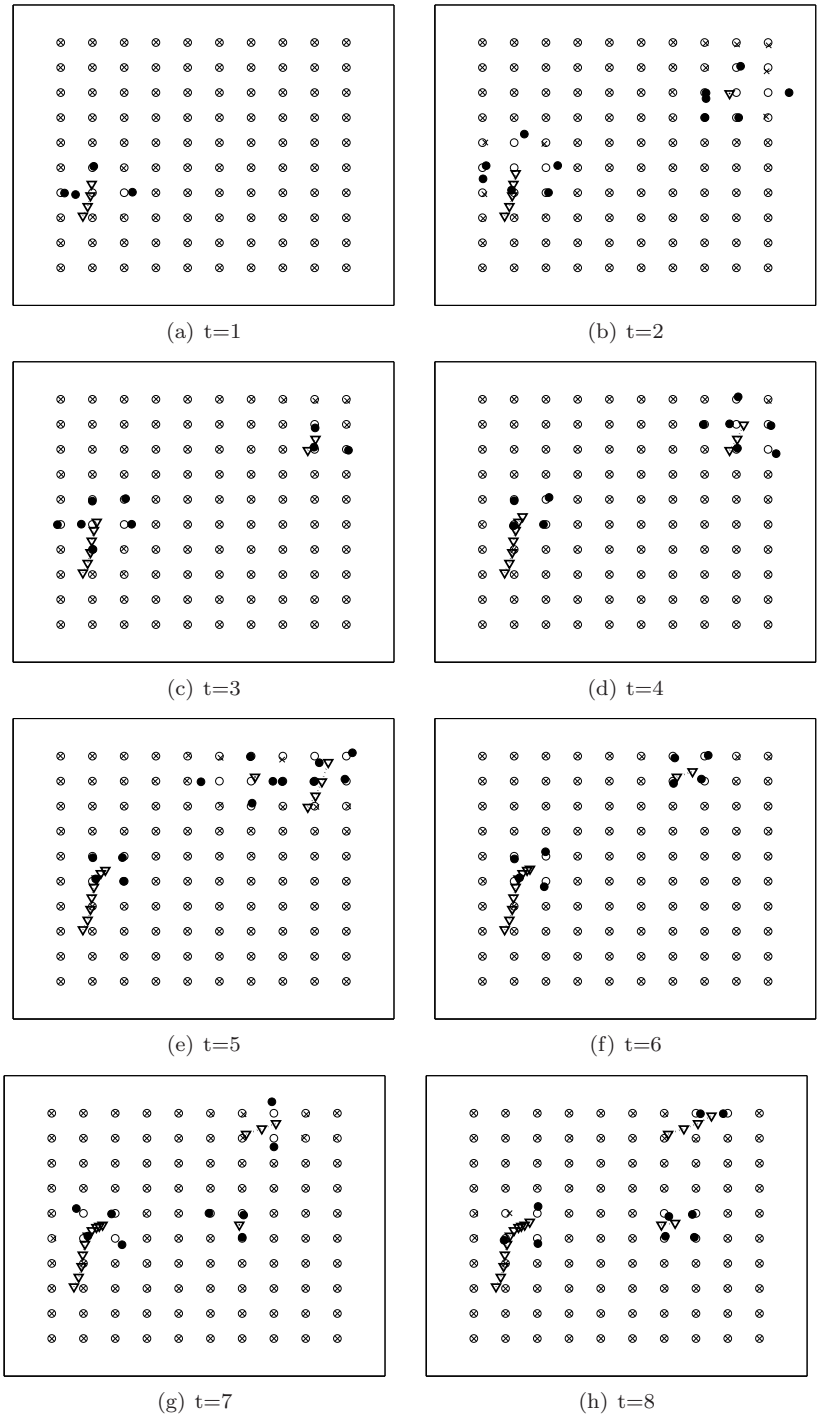
The results are shown in Fig. 6.10. The true sensor locations are shown as circles

and the estimated sensor locations are indicated using crosses. The sensors corresponding to those sensor links that declared a target present using the DBT are shown in filled circles. The target trajectories are shown as inverted triangles. We observe that as the targets move, the sparsity constrained MDS algorithm reconstructs changes in sensor positions while the majority of the estimates sensor locations are unchanged from time step to time step. Thus, in conjunction with sparse dwMDS, the DBT is able to localize the targets to within a small set of sensor links.

6.5.1 Numerical Study

We analyze the performance of the localization algorithms using ROC curves. We consider the following setup: we deploy a 10×10 uniform grid of sensors in a network (see Fig. 6.12). We consider anchor free localization, i.e., $m = 0$ and make a single inter-sensor measurement ($M = 1$) at each time frame. We assume no a priori knowledge of the sensor coordinates, i.e., $r_i = 0$. Each sensor communicates only to its $NN = 8$ nearest neighbors and the weights for those links were chosen by the LOESS strategy [35]. The rest of the weights were set to zero. Furthermore, we set noise variances σ_0 and σ_1 defined in (6.13) and (6.15), respectively as $\sigma_0 = 1$ and $\sigma_1 = 5\sigma_0 = 5$. Sensor links within a radius $R = 1.5$ indicate the presence of a target, i.e., follow the H_1 hypothesis. We set the reference distance d_0 defined below (6.13) to be $d_0 = 1$ and the path loss exponent $\eta = 2$. We set the sparseness parameters $\lambda = 2.5$ and $p = 1$ to produce a change in the location estimates for only a small portion ($< 10\%$) of the sensors.

We begin by considering the case of random appearance of targets in the sensor network, i.e., targets appear at different locations every time instant. For the distance based target localization algorithm (DBT), we set $\tau = 0$ in (6.18), i.e., we compare our distance estimates to a fixed initial frame. For every time instant, the DBT



(a) t=1 (b) t=2

(c) t=3 (d) t=4

(e) t=5 (f) t=6

(g) t=7 (h) t=8

Figure 6.10: Anchor free sensor localization by sparsity constrained MDS in the presence of targets. True sensor locations (circle), estimated sensor locations (cross), sensors localizing the target (blocked circle), target trajectory (inverted triangle).

and the LRT are performed on each active sensor link and the process is repeated for 5000 target locations. The resulting ROC curve is presented in Fig. 6.11. The ROC for the LRT using simulations is indicated using circles and the corresponding theoretical curve obtained from (6.17) is shown as a solid line. We observe that the simulation and the theoretical curves match for the LRT. The ROC for the DBT is shown using a dashed line. The DBT algorithm yields higher probability of correct detection than the LRT for most false alarm levels. For example, at false alarm level $\alpha = 0.3$, β for the DBT is approximately 0.89 which is 5% more than that of the LRT, which yields $\beta \approx 0.84$.

The intuition for this result is as follows: in the presence of a target, the RSS measurements of the sensor links are spatially-correlated. The presence of a target in a given link implies that with high probability the target is present in neighboring sensor links. However, the RSS model in (6.15) specifies only the distribution of the measurements independently on each link. The LRT makes complete use of the RSS measurements but is limited in its performance as the optimal decision statistic for each sensor link i, j is independent of other sensor link measurements. On the other hand, the DBT finds the active sensor links only based on the estimated distances through sparsity penalized MDS. However, since the inter-sensor distances are computed at each sensor using information from its nearest neighbors, this method makes an implicit use of the spatial correlation of the measurements in its decision statistic, which results in an improvement in performance.

Next, we consider the case of a moving target, where we assumed a standard state-space target motion model (for the purpose of a visually pleasing trajectory). We repeated the same algorithms for 5000 such trajectories. The LRT based algorithm yields the same performance curve as the test is independent of whether the target

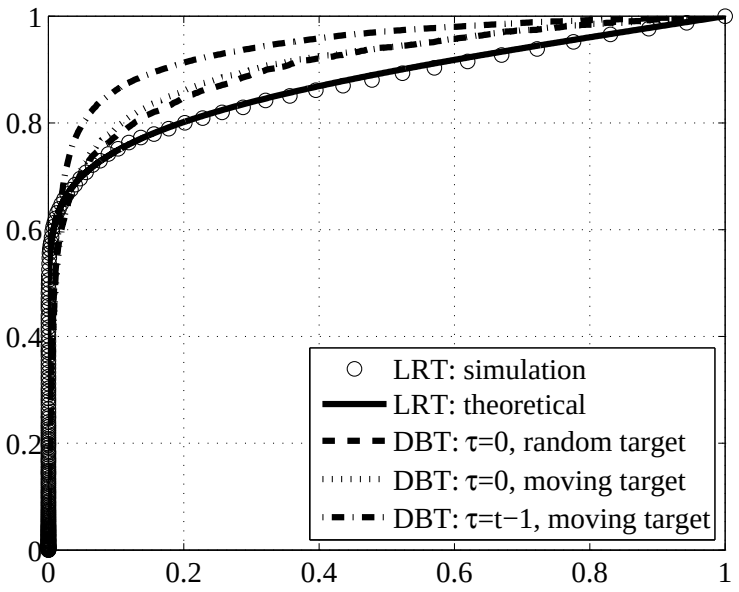


Figure 6.11: ROC curve for the LRT and the DBT link level tracking algorithm. LRT (solid line), DBT for a random target with $\tau = 0$ (dashed), DBT for a moving target with $\tau = 0$ (dotted), and DBT for a moving target with $\tau = t - 1$ (dashed dotted).

is moving or not. The resulting ROC curve for the DBT is presented as a dotted line in Fig. 6.11. Since we continue to base our decision rule on the fixed initial frame ($\tau = 0$), we observe that the performance of the DBT is also similar to the case of random target appearances.

In the case of a moving target, the RSS measurements are also temporally-correlated. Given a set of sensors indicating a presence of a target at a particular time, there is a high probability that the target is in the vicinity of these sensors at the next time frame. To make use of the temporal correlation, we can compare the current estimated distances to the estimated distances from the previous time-frame rather than the initial frame, i.e., set $\tau = t - 1$ instead of $\tau = 0$. The temporal correlation of the RSS measurements is captured in the DBT through the sparsity constraint used for aligning the sensors locations estimates. In other words, with high probability the sensor location estimates that are perturbed in the previous

time-frame will also be perturbed in the current time-frame, thereby increasing the probability of detection.

The results for $\tau = t - 1$ are presented in Fig. 6.11 using a dashed dotted line. We observe that the performance gains are higher than the DBT performed only with spatial smoothing ($\tau = 0$) as such a decision rule incorporates both spatial and temporal correlations of the target dynamics. For example, for $\alpha = 0.1$, β for the LRT is 0.75. The result of spatial smoothing alone yields $\beta \approx 0.79$. By performing both spatial and temporal smoothing, we can obtain $\beta \approx 0.86$ through our algorithm, which corresponds to a 15% increase in performance.

We make the following observations for the two proposed tests:

- The DBT for link level tracking outperforms the LRT as it can account for the spatial and the temporal correlations in the target motion.

- The LRT outperforms DBT for low false alarm levels ($\alpha < 0.01$) for the following reasons: first, the DBT we considered is suboptimal as we did not optimize the performance over the choice of sparsity (p, λ). Furthermore, the LRT uses an optimal decision statistic and the exact measurements to perform the test.

- The issue of space-time sampling is key to the performance of the DBT. Any scenario that exhibits high spatial correlations (e.g., highly reflective targets or more sensors/unit area) can yield further improvement in performance of the DBT. If the sampling time for the sensors and the computation time of the DBT algorithm is much faster than the target motion, the DBT can yield better performance by taking advantage of more temporal correlations.

- The disadvantage of LRT in this setting is that the test is performed independently on each sensor link. Further improvements in the probability of detection

can be achieved when the LRT is derived for the full spatio-temporal model.

- In the performance analysis, we assumed steady state operation, i.e., perfect knowledge of the inter-sensor distances are obtained a priori in the absence of target. If such knowledge is unavailable and distances need to be estimated, the LRT tracker must be modified to a generalized likelihood ratio test (GLRT). The DBT can estimate the initial set of distances more accurately from the RSS measurements by taking advantage of spatial correlations and hence can yield a higher probability of detection than the GLRT.

Spatial localization from link level localization

Our objective is to approximately locate the target relative to the location of the sensors. There are a number of ways in which this link level estimate can be translated into estimated target coordinates in space. For example, one could use as an estimate the midpoint of the convex hull generated by the positions of those sensors that detect the target according to the LRT or the DBT. An example of the midpoint tracking algorithm is shown in Fig. 6.12. Another estimate can be found by the intersection of convex regions corresponding to the sensor links that show the presence of the target through the optimal decision rule. These estimates do not require a physical model of the target trajectory. However, given a target motion model, standard filtering techniques such as the Kalman filter or particle filter (PF) can be used to obtain refined target position estimates from the link level data.

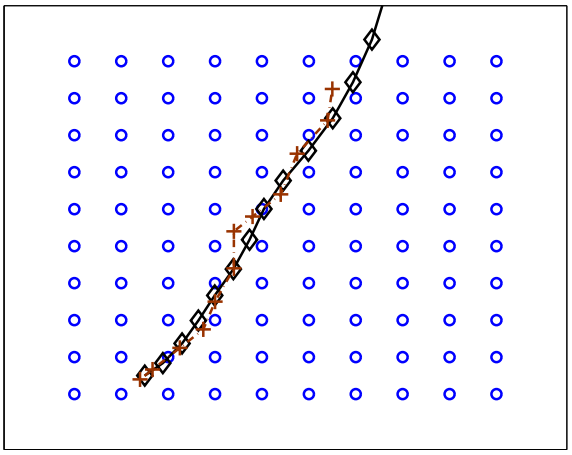


Figure 6.12: A simple tracking algorithm based on link level tracking. True sensor locations (circle), true trajectory of the target (diamond), estimated trajectory (plus).

6.6 Real world applications

6.6.1 ZebraNet database

ZebraNet data set [1] is a collection of zebra movement traces obtained from real-world deployments at the Sweetwaters Game Reserve near Nanyuki, Kenya during the summer of 2005 for a period of 10 days. Four zebras were fitted with sensor nodes consisting of a global positioning system (GPS), simple microcontroller CPU, wireless transceiver, and non-volatile storage to hold logged data and their locations were recorded for a period of 7 days. The raw zebra tracks are shown in Fig. 6.13. The four zebras are marked using a circle, plus, triangle, and a cross respectively.

The ZebraNet sensor nodes used node-to-node communication to propagate measured data towards the base station in a store-and-forward manner. Data was collected on the zebra locations once every 8 minutes to enable the nodes to remain functional for a longer period of time. The zebras on which the hardware set was placed were chosen by biologists to monitor behavior patterns among the zebras. For example, one of the nodes was on a male, which had a tendency to move around

[1] http://www.princeton.edu/ mrm/zebranet.html

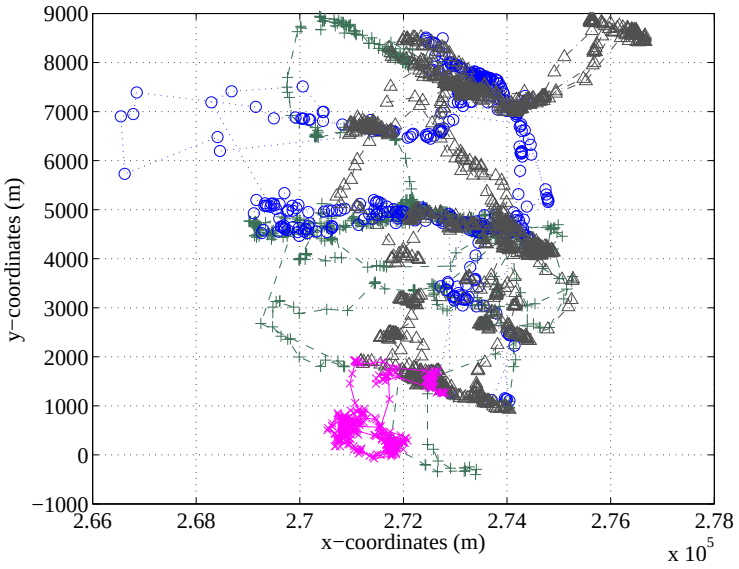


Figure 6.13: Raw tracks for 4 zebras denoted by a circle, plus, triangle, and cross.

looking for a mate. Another sensor was placed on a female, which seemed to be a leader of the large herd. The following information was downloaded to the base station for each of the sensor nodes: the node number; the location in Universal Transverse Mercator (UTM) format; corresponding time rounded-off to the closest minute; and energy status of the node when the measurements were taken.

Our objective is use to these animal traces to test our target localization algorithms. To simulate the performance of the sparse dwMDS algorithm, we superimpose a sensor network consisting of 300 randomly placed sensors on the region of interest in the Sweetwaters Game Reserve. We assume the same hypothesis model presented in (6.50) for the measurements in the presence and absence of target. We assume that each sensor communicates with the 25 nearest neighbors. Based on the noise level, we set a suitable sparsity level λ proportional to the noise variance which ensures that approximately only 5% of the sensor location estimates can vary from those of the previous time frames. We apply the sparsity penalized dwMDS algorithm to this data as the zebras move around in the reserve. The results of the

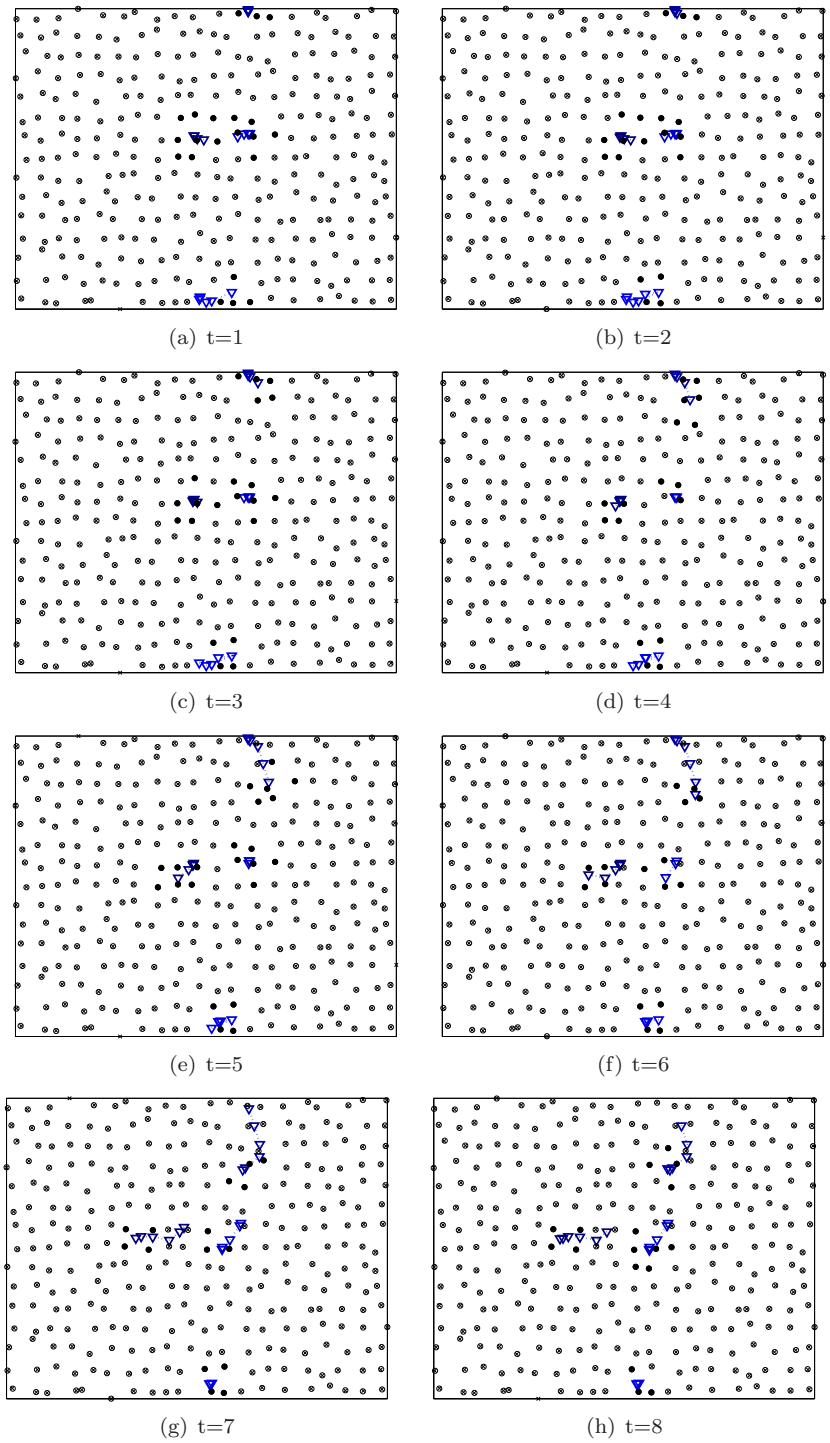
algorithm are presented in Fig. 6.14, shown at various time frames.

Numerical Study

We study the performance of the DBT using the zebra tracks rather than tracks generated by a state-space motion model. We consider the same set of parameters as in the numerical study in Section 6.5.1, i.e., $m = 0$, $M = 1$, $r_i = 0$, $d_0 = 1$, $NN = 8$, $\eta = 2$, $\sigma_0 = 1$, and $\sigma_1 = 5$. We consider 300 randomly deployed sensors in the network (see Fig. 6.14). The denser spatial sampling is useful for tracking multiple targets (5 zebras) and also provides us a setting for addressing the effect of spatial sampling on the DBT performance.

In order to compare the performance of the zebra tracks to the previous results, we run the DBT algorithm on the 5 zebra tracks sequentially, i.e., we assume that there is only one zebra roaming in the network at any time. In the first step, we set $\tau = 0$ in (6.18) and compare the distance estimates to the fixed initial frame. The performances of the DBT and the LRT tests are averaged over the 5 zebra tracks. The resulting ROC curves are presented in Fig. 6.15. The ROC of the LRT shown as a solid line is the same curve as the ones presented earlier. The ROC curve for the DBT is shown using a dashed dotted line. A similarly obtained ROC curve of the DBT from the previous study (see dotted line in Fig. 6.11) is shown as a dotted line. By comparing the perturbation to the fixed initial frame, we only perform spatial smoothing of the sensor location estimates. We observe that the denser sampling of sensors have resulted in better spatial smoothing, which eliminates more false alarms resulting in an improved performance. For example, at a false alarm level $\alpha = 0.01$, the DBT with 100 sensors yields $\beta \approx 0.48$, while the DBT with 300 sensors yields $\beta \approx 0.66$.

Next, we perform the DBT on the same set of tracks using $\tau = t - 1$ in (6.18).



(a) t=1 (b) t=2

(c) t=3 (d) t=4

(e) t=5 (f) t=6

(g) t=7 (h) t=8

Figure 6.14: Sparsity constrained MDS on the ZebraNet data. True sensor locations (circle), estimated sensor locations (cross), sensors localizing the target (blocked circle), target trajectory (inverted triangle).

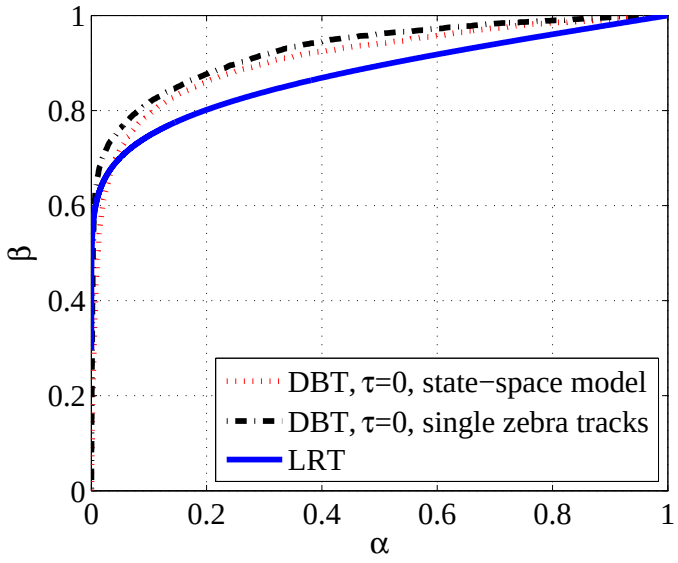


Figure 6.15: ROC curve for the LRT and the DBT link level tracking algorithm using individual zebra tracks with $\tau = 0$. LRT (solid line), DBT for a moving target from state-space model using 10×10 grid of sensors (dotted), and DBT for the zebra tracks (dashed dotted) using 300 randomly located sensors.

The resulting ROC curve is presented in Fig. 6.16. The performance of the DBT with $\tau = t - 1$ is shown as a dashed line. The dashed dotted line represents the performance of DBT with $\tau = 0$, while the solid line denotes the ROC of the DBT with $\tau = t - 1$ from the previous study with 10×10 uniform grid of sensors. We observe that the probability of detection of the DBT with $\tau = t - 1$ for a specified false alarm is mostly higher than that of $\tau = 0$ DBT ROC curve. However, the performance improvement between the two methods (with $\tau = 0$ and $\tau = t - 1$) is lower than performance improvement between the same two methods from the previous study. A possible reason for this behavior can be explained as follows: in the previous study, we assumed a high sampling rate, so that the sensors can make use of the temporal correlations in the measurements to improve performance. On the other hand, the zebra tracks were sampled only once in 8 minutes, which eliminates most temporal correlations in the data.

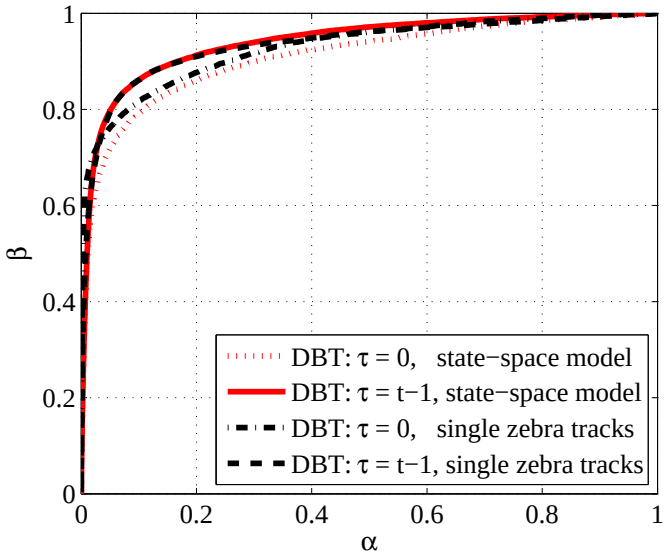


Figure 6.16: ROC curve for DBT link level tracking algorithm using individual zebra tracks with $\tau = t - 1$. DBT for a moving target from state-space model using 10×10 grid of sensors (dotted), and DBT for the zebra tracks (dashed dotted) using 300 randomly located sensors.

Finally, we analyze the performance of the DBT for simultaneously tracking the 5 zebras. The resulting ROC curves for the case of $\tau = 0$ and $\tau = t-1$ are presented in Fig. 6.17. The case of $\tau = 0$ is shown as a dotted curve and the dashed dotted curve represents the DBT performance with $\tau = t - 1$. The performance gains for the multiple targets case are similar to the case of tracking the individual zebras sequentially. This suggests the DBT algorithm is well equipped to handle the case of multi-target tracking.

6.6.2 UCSD wireless trace data

The wireless topology discovery (WTD) project[2] was undertaken by researchers at University of California San Diego (UCSD). The project collects data on dynamic characteristics and user behavior in a real world wireless network. The primary objective of the WTD project was to test and develop reliable routing protocols in

[2]http://nile.usc.edu/MobiLib

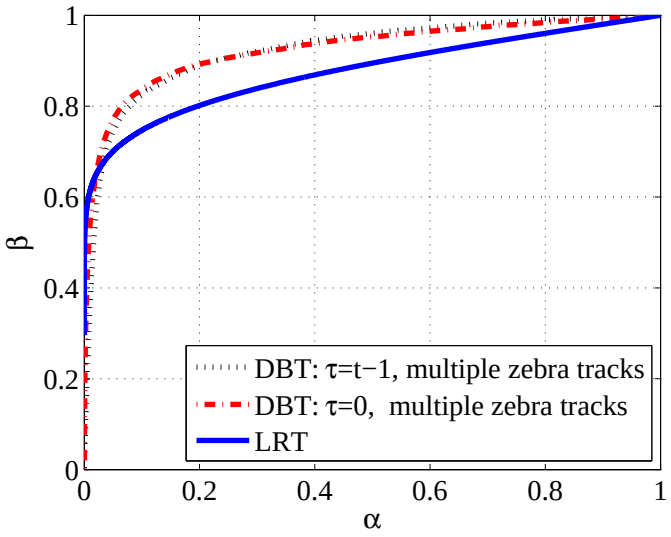


Figure 6.17: ROC curve for DBT link level tracking algorithm using multiple zebra tracks. LRT (solid line), DBT for zebra tracks with $\tau = 0$ (dotted), and DBT for zebra tracks with $\tau = t - 1$ (dashed dotted).

a geographically constrained wireless network. To collect data, 275 UCSD freshman were handed HP Jordana PDAs which were equipped with symbol 802.11 compact flash cards and the WTD data collection software. The software recorded all access points (AP) sensed by the user every 20 seconds. The trace data were collected over a 11 week period and then transferred to a centralized database for analysis. The data indicated that around 300 APs were sensed over this time among which only 200 of them had knowledge about their locations.

The trace data collected consisted of the following information: user identity, sample time, AP identity, RSS, and AC/battery power indicator. The coordinates of the known APs were also provided in the database. Our objective was to recover user trajectories over time and to reconstruct the network topology (locations of the unknown APs) using available data. The map of the known AP locations on the UCSD campus is shown in Fig. 6.18. The data samples at a particular time instant for a single user are shown in Fig. 6.19. The APs sensed by the user are shown using

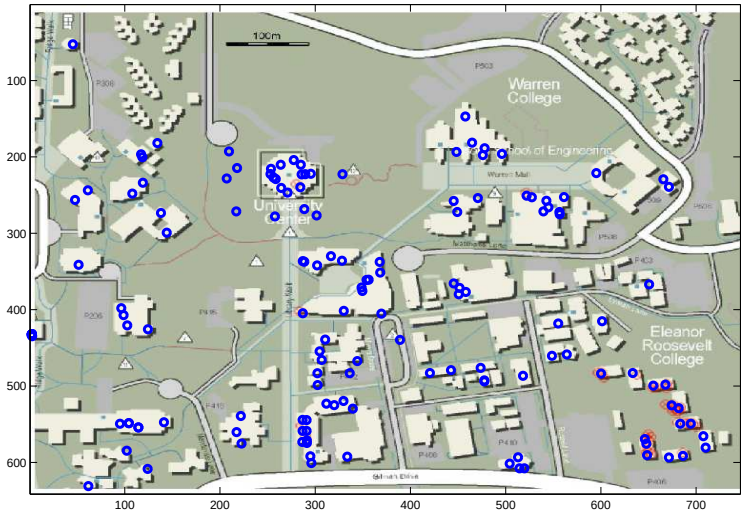


Figure 6.18: Campus map showing the 200 known AP locations. Only the horizontal (x,y) part of the 3D coordinates (x,y,z) are shown. In addition to these locations, there were 100 APs at unknown locations.

blocked circles. The corresponding RSS values are shown next to these APs.

To reconstruct the user trajectories, we need to estimate user locations over time based on RSS observed to the various access points. The RSS values provided in the database were 5-bit quantized values between 0 and 31. Since the mapping from these quantized values to the actual signal strength in dB is unknown, we needed to calibrate the quantized numbers to signal strength values.

Calibration

Consider a group of U users in the wireless network. Denote the location of user i at time t as $\mathbf{x}_i^t$. Denote the set of N known AP locations as $\{\mathbf{y}_k\}_{k=1}^N$. Let the unknown locations of the remaining M APs be $\{\mathbf{z}_k\}_{k=1}^M$. Given RSS measurement $P_{i,j}$ between user i and AP j, the ML estimate of the distance is given by

$$\delta_{i,j} = d_0 10^{((P_0 - P_{i,j})/10n_p)}, \tag{6.19}$$

where we set reference distance $d_0 = 1$. Since there are no reference points given through the data, P_0 and n_p are unknown and need to calibrated. Furthermore,

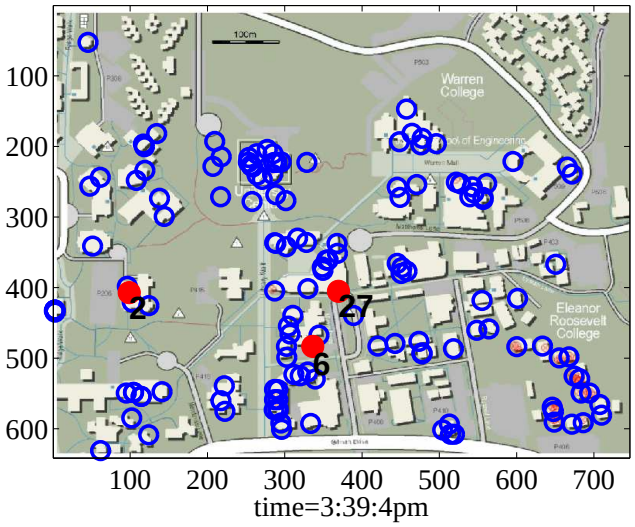


Figure 6.19: Sample RSS data from a single user to APs. The sensed APs by the user are indicated in blocked red along with their RSS measurements.

values for $P_{i,j}$ are given through the quantized numbers and the actual mapping to RSS is not known. Here we assume that 5-bit quantized values is scaled and translated version of the actual RSS, i.e., if the quantized values are given by $P^q_{i,j}$, then the RSS value is given by $P_{i,j} = aP^q_{i,j} + b$. Hence the distance estimate in terms of the quantized RSS values can be written as

$$\delta_{i,j} = 10^{\left((P^q_0 - P^q_{i,j})/10n^q_p\right)}, \tag{6.20}$$

where $n^q_p = n_p/a$ and $P^q_0 = (P_0 - b)/a$. We provide an iterative least squares solution for estimating P^q_0 and n^q_p that yields this mapping between the RSS integer values and the distances. We start with random initial conditions for P^q_0 and n^q_p and compute the user location estimates through the dwMDS algorithm based on the RSS values between the users and the APs. We construct a scatter plot between the estimated distances and the observed quantized RSS values. Using this map, we construct a least squares fit to obtain a new set of values for P^q_0 and n^q_p. This process is repeated until the values of P^q_0 and n^q_p converge. A block diagram of the algorithm is presented

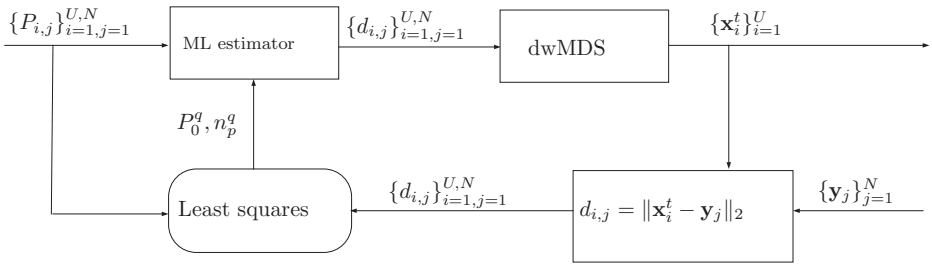


Figure 6.20: A block diagram of the iterative least squares procedure for estimating calibration parameters.

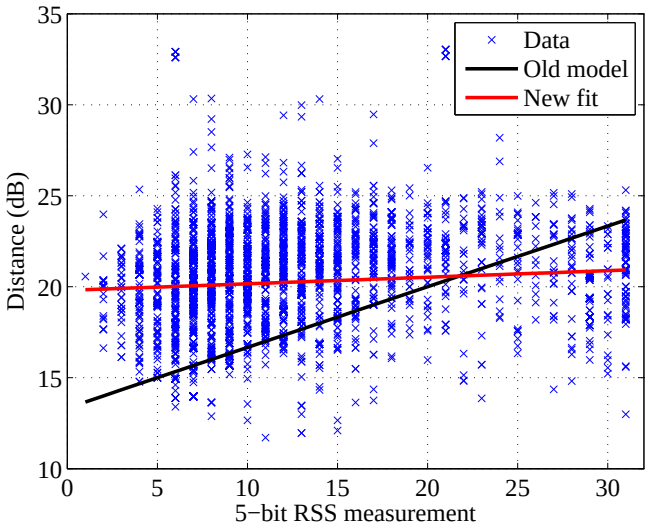


Figure 6.21: Estimated distances along with the least square fit to the data at the first iteration.

in Fig. 6.20. The histogram after the first step of this iteration is shown in Fig. 6.21. The black line is the initially assumed map obtained by taking the logarithm of (6.20), where P_0^q and n_p^q are set to initial values used for generating the map. The blue crosses indicate the estimated distances (dB) for various quantized values of the RSS based on these initial conditions. The red line is the least squares fit to the generated data. The two lines have different slopes which suggests that the actual fit is not the same as the assumed fit. A similar graph generated after 15 iterations is shown in Fig. 6.22. We observe that the two lines nearly coincide suggesting that the values of P_0^q and n_p^q are a stable fit to the data.

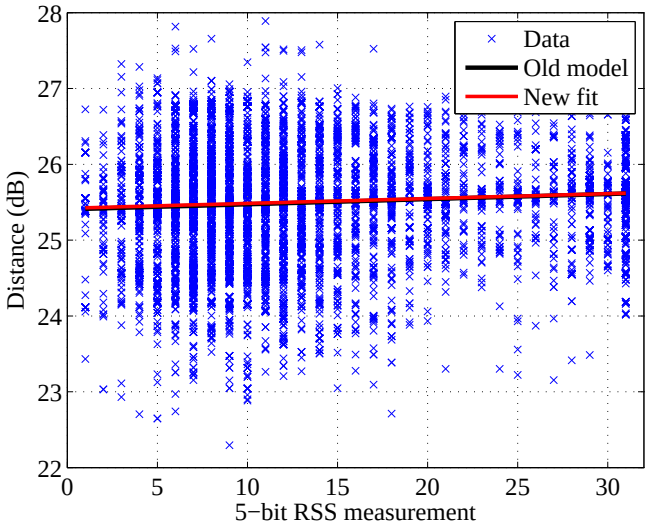


Figure 6.22: Estimated distances along with the least square fit to the data after 15 iterations.

We observe from Fig. 6.22 that there is a large variance in the data and the mean distance estimates of the RSS integer values are similar. One reason for this large variance is due to the fact that we fit a global shadowing model to the APs. However, in reality, some of the APs might be more accessible to the user (e.g., on top of a tower) than others (e.g., within a building). Furthermore, APs might sense users using different power levels, which needs normalization. Below, we list a number of limitations of this UCSD data set that made its analysis difficult.

Limitations of the data

1. Noisy and inconsistent data.

2. The RSS measurements recorded by the user to the various APs are not normalized. A user can be sensing two different APs operating at two different power levels. Furthermore, the user could be operating his/her PDA using AC or battery power which may affect the RSS value.

3. 5-bit quantized RSS values. A 5-bit RSS value yields only 32 distinct distance

estimates from the user to the AP which is insufficient to accurately locate the user.

4. No reference RSS values. The map between the quantized numbers and the RSS values is an important missing piece in this problem. We alleviate the need for the map to a small extent by the calibration process described earlier.

5. The RSS values are recorded every 20 seconds, which is an insufficient sampling rate for recovering smooth user trajectories.

Simultaneous localization of targets and APs

The information about the unknown AP locations are given by the RSS values measured by the users at various time instants to these APs. One would expect that since many of the AP locations are known, and the target user locations can be estimated from these known-location AP data, we could recover the positions of the unknown APs (they are after-all like any other fixed target at an unknown location once target positions are approximate known). The problem of optimally localizing the unknown APs and the targets can be formulated as an Euclidean distance matrix completion problem (EDMCP) [3, 9]. This is a classical problem in geometry and can be stated as the problem of recovering the set of all pairwise distances between n points given only a subset of these distances. The inter-point distance matrix $\mathbf{D}$ can be written as

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_{yy} & \mathbf{D}_{yx} & \mathbf{D}_{yz} \\ \mathbf{D}_{xy} & \mathbf{D}_{xx} & \mathbf{D}_{xz} \\ \mathbf{D}_{zy} & \mathbf{D}_{zx} & \mathbf{D}_{zz} \end{pmatrix}, \tag{6.21}$$

where $\mathbf{D}_{yy}$ is the distance matrix between the known AP locations, $\mathbf{D}_{yx}$ is the distance matrix between the known AP locations and the users, and the rest of the

sub-distance matrices are similarly defined. Among these sub-matrices, only $\mathbf{D}_{yy}$ is completely known. The distance matrices $\mathbf{D}_{yx}$ and $\mathbf{D}_{zx}$ are partially known since different users are within range of only a fraction of the known and unknown APs. Given the partial matrix,

$$\mathbf{D}^* = \begin{pmatrix} \mathbf{D}_{yy} & \mathbf{D}_{yx} & * \\ \mathbf{D}_{xy} & * & \mathbf{D}_{xz} \\ * & \mathbf{D}_{zx} & * \end{pmatrix}, \tag{6.22}$$

our objective is to reconstruct the complete distance matrix $\mathbf{D}$ in (6.21). Once $\mathbf{D}$ is recovered we can recover the user tracks and unknown AP locations.

We consider two cases: (a) the exact completion problem, i.e., when the partial matrix is exact. (b) the approximate completion problem, i.e., when the known distances are estimated from noisy RSS measurements.

Exact completion problem

A set of necessary and sufficient conditions for existence of a unique solution to the exact completion problem can be summarized by the following theorem [9].

Theorem 6.6.1. *Let* $\mathbf{A}$ *be a* $N \times N$ *partial distance matrix in* $\mathbb{R}^k$*. Let* $G = (V, E)$ *be an undirected graph with* $V = 1, 2, \ldots, N$*,* $E = \{(i, j) \mid a_{i,j} \textit{is specified}\}$*, and whose specified entries are chordal (see below for definition of a chordal graph). Let* $\mathcal{S}$ *be the collection of all minimal vertex separators of* G*. Then* $\mathbf{A}$ *admits a unique completion to a distance matrix in* $\mathbb{R}^k$ *if and only if*

$$\mathbf{B} = \begin{pmatrix} 0 & \mathbf{e}^T \\ \mathbf{e} & \mathbf{A}(S) \end{pmatrix} \textit{ has rank } k+2 \textit{ for any } S \in \mathcal{S}, \tag{6.23}$$

where $\mathbf{e}$ *is a column vector of ones and* $\mathbf{A}(S)$ *is a matrix formed by using the set of vertices in* S*.*

Definition: A graph is *chordal* if each of its cycles of four or more vertices has a chord, which is an edge joining two nodes that are not adjacent in the cycle.

However, the solution to the exact completion problem for an arbitrary known set of partial matrix entries cannot be obtained in closed-form [166]. But we show that for a specific structure of the partial distance matrix, a closed-form solution to the exact completion problem can be obtained. The key idea behind finding a solution to the EDMCP is the rank deficiency of the distance matrix.

Theorem 6.6.2. *Let* $\mathbf{A}$ *be a* $(N+M)\times(N+M)$ *partial distance matrix with rank* $k+2$ *and the following structure:*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \tag{6.24}$$

where $\mathbf{A}_{11}$ *is* $N\times N$*,* $\mathbf{A}_{12}$ *is* $N\times M$*, and* $\mathbf{A}_{22}$ *is an* $M\times M$ *matrix. Given,* $\mathbf{A}_{11}, \mathbf{A}_{12}$*, there exists an unique Euclidean matrix completion to* $\mathbf{A}$ *given by* $\mathbf{A}_{22} = \mathbf{A}_{21}\mathbf{A}_{11}^{+}\mathbf{A}_{12}$ *if rank*$(\mathbf{A}_{11}) = k+2$*.*

Proof. The set of nonadjacent vertices for the partially complete distance matrix $\mathbf{A}$ defined in (6.24) is given by $N_v = \{(i,j) \mid N+1 \leq i,j \leq N+M\}$. The corresponding collection of minimal vertex separators of the graph G is a singleton set $S = \{1, 2, \ldots, N\}$. When $N > k+2$, $\mathbf{B}$ in (6.23) has rank $k+2$ if $\mathbf{A}_{11}$ has rank $k+2$. Let $r = k+2$. From Theorem 6.6.1, there exists a unique solution to the exact completion problem. The eigendecomposition of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T,$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r]$ are the corresponding set of orthonormal eigenvectors. Let $\mathbf{V} = [\mathbf{V}_1^T \mathbf{V}_2^T]^T$, where $\mathbf{V}_1$ is $N\times r$ and $\mathbf{V}_2$ is $M\times r$.

Then the sub-matrices can be written as

$$\mathbf{A}_{11} = \mathbf{V}_1 \mathbf{\Lambda} \mathbf{V}_1^T \tag{6.25}$$

$$\mathbf{A}_{12} = \mathbf{V}_1 \mathbf{\Lambda} \mathbf{V}_2^T \tag{6.26}$$

$$\mathbf{A}_{22} = \mathbf{V}_2 \mathbf{\Lambda} \mathbf{V}_2^T \tag{6.27}$$

Since $\mathbf{A}_{11}$ is full rank, the pseudo inverse of $\mathbf{A}_{11}$ can be written as

$$\mathbf{A}_{11}^+ = \mathbf{V}_1 (\mathbf{V}_1^T \mathbf{V}_1)^{-1} \mathbf{\Lambda}^{-1} (\mathbf{V}_1^T \mathbf{V}_1)^{-1} \mathbf{V}_1^T. \tag{6.28}$$

Then

(6.29)

$$\mathbf{A}_{21} \mathbf{A}_{11}^+ \mathbf{A}_{12} = \mathbf{V}_2 \mathbf{\Lambda} \mathbf{V}_1^T \left(\mathbf{V}_1 (\mathbf{V}_1^T \mathbf{V}_1)^{-1} \mathbf{\Lambda}^{-1} (\mathbf{V}_1^T \mathbf{V}_1)^{-1} \mathbf{V}_1^T \right) \mathbf{V}_1 \mathbf{\Lambda} \mathbf{V}_2^T = \mathbf{V}_2 \mathbf{\Lambda} \mathbf{V}_2^T = \mathbf{A}_{22}.$$

□

This result does not hold for the following degenerate case: when all the points yielding the partial matrix $\mathbf{A}_{11}$ lie on a k-dimensional sphere such that the rank($\mathbf{A}_{11}$) is $k+1$. It is easy to verify that a unique solution exists for this case using Theorem 6.6.1 but Theorem 6.6.2 does not yield the optimal completion. However, in a real world setting, the probability that all the randomly deployed sensors lie in a 3-dimensional sphere is nearly zero.

Approximate completion problem

For the UCSD trace data, $\mathbf{D}_{yx}$ and $\mathbf{D}_{zx}$ are only partially known and noisy. Hence the theorem cannot be applied to this data set. For the case of approximate completions, the optimal Euclidean distance matrix can be found by solving the following minimization problem analogous to MDS. Let $\mathbf{A} = \{a_{i,j}\}$ be a pre-distance matrix, where only some elements $a_{i,j}$ are given. Let $\mathbf{W}$ be a symmetric weight matrix

with nonnegative elements. For e.g., $w_{i,j} = 1$ if $a_{i,j}$ is given and zero otherwise. Then the closest distance matrix to $\mathbf{A}$ can be found by minimizing the objective function

$$\min_{\mathbf{D}} \|\mathbf{W} \circ (\mathbf{A} - \mathbf{D})\|_F^2 \qquad \text{s.t} \qquad \mathbf{D} \in \mathcal{D}, \tag{6.30}$$

where $\circ$ denotes the Hadamard product and $\mathcal{D}$ is the convex cone of Euclidean distance matrices. The objective function can be rewritten as

$$\min_{\mathbf{D}} \sum_{i,j} w_{i,j} \left(a_{i,j} - d_{i,j}\right)^2, \qquad \text{s.t} \qquad \mathbf{D} \in \mathcal{D}. \tag{6.31}$$

A semi-definite programming solution to this problem is provided in [1]. In the context of recovering the locations, the distributed weighted MDS algorithm minimizes precisely the same cost function, i.e.,

$$\min_{\mathbf{X}} \sum_{i,j} w_{i,j} \left(a_{i,j} - \|\mathbf{x}_i - \mathbf{x}_j\|\right)^2, \tag{6.32}$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ are the set of locations which yields the distance matrix $\mathbf{D}$. In other words, the dwMDS yields another approach to the approximate completion problem. We adopt a two-step dwMDS procedure to localize the unknown users and APs.

First, we consider all users that sense at least 4 known APs and one unknown AP. Using only the knowledge of the RSS values between the users and the known APs, we estimate the locations of the users in the network. We then use the set of user locations with the corresponding RSS values to the unknown APs to estimate the location of the unknown APs. This two stage process is illustrated through Fig. 6.23. Since we do know the ground truth of the actual location of the unknown APs, we randomly choose a small set of known APs and add them to the set of unknown APs. We reconstruct the location of unknown APs (and those APs assumed unknown) using different user trajectories. For the set of APs with known coordinates,

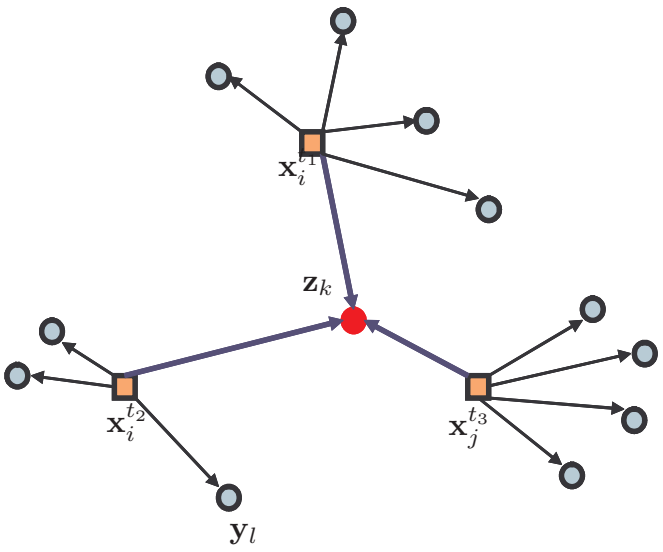


Figure 6.23: Two stage procedure for constructing network topology. First, through the known APs, user locations are estimated. Based on these locations, the unknown APs are localized.

we obtain a 1σ confidence region on the network topology which serves as a measure of performance of the localization algorithm. We illustrate the performance of the method in Fig. 6.24. The known AP locations which were originally assumed unknown are shown using filled circles. The mean estimates of the APs are shown as triangles. The black ellipse is the uncertainty in the estimation of the unknown AP. The error in the mean estimate of the AP location is roughly 30m. The actual estimated locations of the unknown APs are shown in Fig. 6.25. The already known AP locations are indicated as circles while the location estimates of the unknown APs are shown as triangles.

Scope for improvement

Despite a number of limitations presented in this UCSD data set, we were able to attain reasonable accuracy on the localization of the APs. We offer some suggestions to improve upon the localization performance.

1. Multiple local fits to the RSS model rather than a single global fit: To accurately

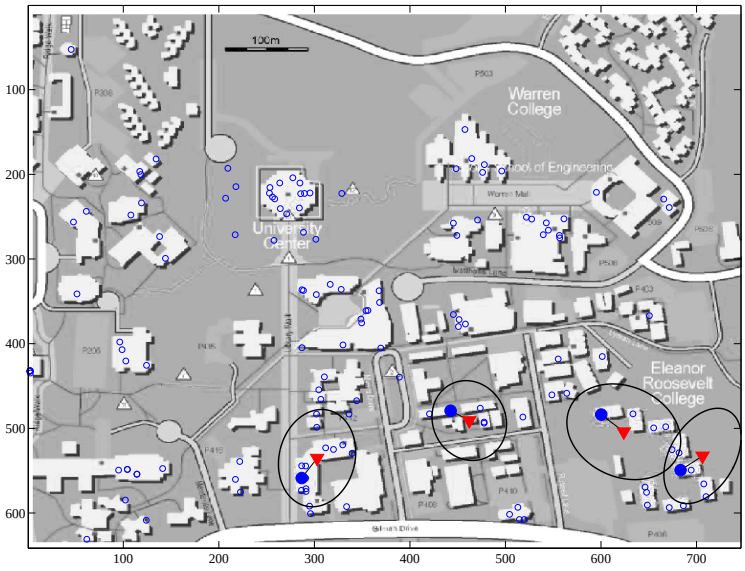


Figure 6.24: Location estimates of APs assumed to be unknown and the corresponding uncertainty ellipses. Only horizontal coordinates of the 3D coordinate estimates are shown.

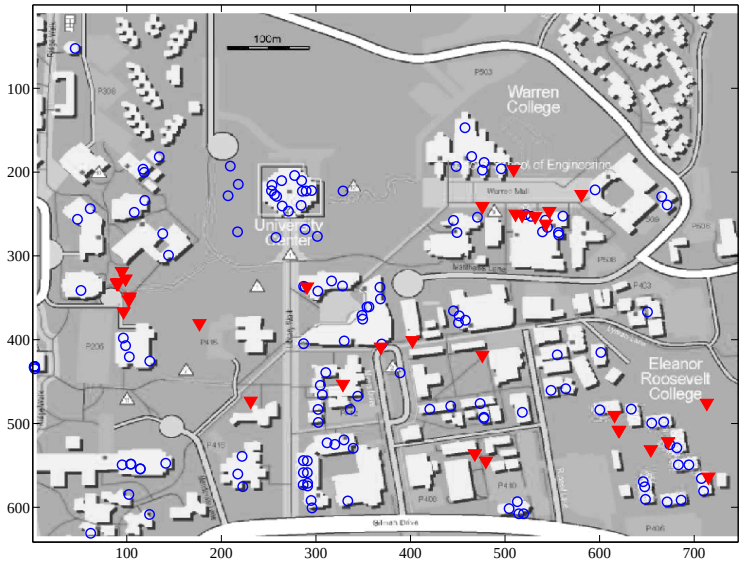


Figure 6.25: Location estimates of the unknown APs are shown in red. The known AP locations are shown in blue.

estimate distances from RSS measurements, we fit a single log normal model to the RSS data. However, the attenuation in the RSS sensed by a user might be different for different APs. For example, users might receive a higher RSS value from an AP on a roof top rather than an AP on the same building one floor below. This suggests that the log normal model can be adaptively chosen for each AP based on its relative location in the network.

2. Inclusion of side information: We estimated the user locations with no assumptions on the user trajectories. One can make smoothness assumptions on the user tracks by restricting the user location estimate at the current time to be in a neighborhood of the user location estimate from the previous time instant. Moreover, we can impose further constraints on the user and AP locations by making use of the topology of the UCSD campus. For example, an user cannot be located on a building or floating in air, and an AP cannot be in the middle of a road.

6.7 Conclusions

In this chapter, we presented the sparsity penalized MDS algorithm for simultaneous localization and tracking. In this problem one is interested in tracking a targets position relative to the sensor coordinates. The subset selection capability of our proposed sparsity constraint allows the algorithm to find only those sensors which have significantly changed their location estimate due to the presence of a target. We use these sensors to perform link level tracking. We formulate a model for the inter-sensor RSS measurements in the presence and absence of a target by conducting actual experiments in free space. Using this model, we propose a perturbation based algorithm for link level target tracking. Through a numerical study, we showed

that for a large range of false alarm levels, the DBT outperforms the LRT as it is able to perform spatial and temporal smoothing without the need for target motion models. The nonparametric nature of our algorithm makes it attractive when RSS models are unavailable or inaccurate. We then tested our localization algorithms on two real world applications: localizing zebras from the ZebraNet data and recovering the wireless network topology of the UCSD campus. Currently, we are analyzing the effect of space-time sampling (via the increase of sensors, faster sampling of measurements) on the performance of the DBT. We are also in pursuit of optimal sensor scheduling strategies for physical level tracking.

6.8 Appendix: derivation of sparsity penalized dwMDS

To simplify our derivation, we divide the global cost function into multiple local cost functions as follows:

$$C^{(t)} = \sum_{i=1}^{n} C_i^{(t)} + c^{(t)}, \tag{6.33}$$

where $c^{(t)}$ is a constant independent of the sensor locations $\mathbf{X}$ and the local cost function at each sensor node i is

$$\begin{aligned} C_i^{(t)} = \sum_{j=1, j\neq i}^{n} \bar{w}_{i,j}^{(t)} (\bar{\delta}_{i,j}^{(t)} - d_{i,j}(\mathbf{X}))^2 + 2 \sum_{j=n+1}^{n+m} \bar{w}_{i,j}^{(t)} (\bar{\delta}_{i,j}^{(t)} - d_{i,j}(\mathbf{X}))^2 \\ + r_i \|\bar{\mathbf{x}}_i - \mathbf{x}_i\|^2 + \lambda \|\mathbf{x}_i - \mathbf{x}_i^{(t-1)}\|^p, \end{aligned} \tag{6.34}$$

where $\bar{w}_{i,j}^{(t)} = \sum_{l=1}^{M} w_{i,j}^{(t),l}$ and $\bar{\delta}_{i,j}^{(t)} = \sum_{l=1}^{M} w_{i,j}^{(t),l} \delta_{i,j}^{(t),l} / \bar{w}_{i,j}^{(t)}$. The cost function $C_i^{(t)}$ depends only the measurements made by sensor node i and the positions of the neighboring nodes, i.e., nodes for which $w_{i,j}^{(t),l} > 0$; $C_i^{(t)}$ is local to node i [35]. The local cost function in (6.34) can be rewritten as

$$C_i^{(t)}(\mathbf{X}) = c_1^{(t)} + c_2^{(t)}(\mathbf{X}) - c_3^{(t)}(\mathbf{X}) + c_4^{(t)}(\mathbf{X}), \tag{6.35}$$

where

$$
\begin{aligned}
c_1^{(t)} &= \sum_{j=1,j\neq i}^{n} \bar{w}_{i,j}^{(t)} (\bar{\delta}_{i,j}^{(t)})^2 + 2 \sum_{j=n+1}^{n+m} \bar{w}_{i,j}^{(t)} (\bar{\delta}_{i,j}^{(t)})^2 \\
c_2^{(t)}(\mathbf{X}) &= \sum_{j=1,j\neq i}^{n} \bar{w}_{i,j}^{(t)} d_{i,j}^2(\mathbf{X}) + 2 \sum_{j=n+1}^{n+m} \bar{w}_{i,j}^{(t)} d_{i,j}^2(\mathbf{X}) + r_i \|\bar{\mathbf{x}}_i - \mathbf{x}_i\|^2 \\
c_3^{(t)}(\mathbf{X}) &= 2 \sum_{j=1,j\neq i}^{n} \bar{w}_{i,j}^{(t)} \bar{\delta}_{i,j}^{(t)} d_{i,j}(\mathbf{X}) + 4 \sum_{j=n+1}^{n+m} \bar{w}_{i,j}^{(t)} \bar{\delta}_{i,j}^{(t)} d_{i,j}(\mathbf{X}) \\
c_4^{(t)}(\mathbf{X}) &= \lambda \|\mathbf{x}_i - \mathbf{x}_i^{(t-1)}\|^p. \qquad (6.36)
\end{aligned}
$$

The term $c_1^{(t)}$ is independent of $\mathbf{x}_i$. The term $c_2^{(t)}$ is quadratic in $\mathbf{x}_i$. Terms $c_3^{(t)}$ and $c_4^{(t)}$ are neither affine nor quadratic functions of $\mathbf{x}_i$. A majorizing function for the term $c_3^{(t)}$ is motivated by the following Cauchy-Schwarz inequality,

$$
d_{i,j}(\mathbf{X}) = \|\mathbf{x}_i - \mathbf{x}_j\| \geq \frac{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{y}_i - \mathbf{y}_j)}{d_{i,j}(\mathbf{Y})}, \qquad \forall \mathbf{Y}, \qquad (6.37)
$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$. For $c_4^{(t)}$, we present a quadratic majorizing function, which can be obtained from the following relation

$$
\alpha^{p/2} \leq \alpha_0^{p/2} + \frac{p}{2}(\alpha - \alpha_0)(\alpha_0)^{(\frac{p}{2}-1)}, \qquad \forall \alpha, \alpha_0 > 0. \qquad (6.38)
$$

The above inequality follows from a linear approximation to the concave function $f(\alpha) = \alpha^{p/2}$ via Taylor series expansion. Choosing $\alpha = \|\mathbf{x}_i - \mathbf{x}_i^{t-1}\|^2$ and $\alpha_0 = \|\mathbf{y}_i - \mathbf{x}_i^{t-1}\|^2$ yields

$$
\|\mathbf{x}_i - \mathbf{x}_i^{t-1}\|^p \leq \|\mathbf{y}_i - \mathbf{x}_i^{t-1}\|^p + \frac{p}{2} \frac{\|\mathbf{x}_i - \mathbf{x}_i^{t-1}\|^2 - \|\mathbf{y}_i - \mathbf{x}_i^{t-1}\|^2}{\|\mathbf{y}_i - \mathbf{x}_i^{t-1}\|^{2-p}}, \qquad (6.39)
$$

the majorizing function for the $c_4^{(t)}$ term. Substituting the inequalities from (6.37)

and (6.39) in (6.35), we obtain the majorizing function for the local cost function as

$$\begin{aligned} T_i^{(t)}(\mathbf{X},\mathbf{Y}) &= c_1^{(t)} + \sum_{j=1,j\neq i}^{n} \bar{w}_{i,j}^{(t)} d_{i,j}^2(\mathbf{X}) + 2\sum_{j=n+1}^{n+m} \bar{w}_{i,j}^{(t)} d_{i,j}^2(\mathbf{X}) + r_i\|\bar{\mathbf{x}}_i - \mathbf{x}_i\|^2 \\ &\quad +2\sum_{j=1,j\neq i}^{n} \bar{w}_{i,j}^{(t)}\bar{\delta}_{i,j}^{(t)} \frac{(\mathbf{x}_i-\mathbf{x}_j)^T(\mathbf{y}_i-\mathbf{y}_j)}{d_{i,j}(\mathbf{Y})} \\ &\quad +4\sum_{j=n+1}^{n+m} \bar{w}_{i,j}^{(t)}\bar{\delta}_{i,j}^{(t)} \frac{(\mathbf{x}_i-\mathbf{x}_j)^T(\mathbf{y}_i-\mathbf{y}_j)}{d_{i,j}(\mathbf{Y})} \\ &\quad + \lambda\|\mathbf{y}_i - \mathbf{x}_i^{(t-1)}\|^p + \frac{\lambda p}{2}\frac{\|\mathbf{x}_i-\mathbf{x}_i^{(t-1)}\|^2 - \|\mathbf{y}_i-\mathbf{x}_i^{(t-1)}\|^2}{\|\mathbf{y}_i-\mathbf{x}_i^{(t-1)}\|^{2-p}}. \end{aligned} \tag{6.40}$$

Since $T_i^{(t)}(\mathbf{X},\mathbf{Y})$ is a majorizing function to $C_i^{(t)}(\mathbf{X})$, it is easy to verify that the function $T^{(t)}(\mathbf{X},\mathbf{Y}) = \sum_{i=1}^n T_i^{(t)}(\mathbf{X},\mathbf{Y})$ is a majorizing function to the global cost function $C^{(t)}(\mathbf{X})$. The partial derivative of $T^{(t)}(\mathbf{X},\mathbf{Y})$ with respect to $\mathbf{x}_i$ is straightforward as all the expressions in (6.40) are linear or quadratic in $\mathbf{x}_i$. The partial derivative of $T^{(t)}(\mathbf{X},\mathbf{Y})$ with respect to $\mathbf{x}_i$ is given by

$$\frac{\partial T^{(t)}(\mathbf{X},\mathbf{Y})}{\partial \mathbf{x}_i} = \frac{\partial T_i^{(t)}(\mathbf{X},\mathbf{Y})}{\partial \mathbf{x}_i} + \sum_{k\neq i}\frac{\partial T_k^{(t)}(\mathbf{X},\mathbf{Y})}{\partial \mathbf{x}_i}, \tag{6.41}$$

where

$$\begin{aligned} \frac{\partial T_i^{(t)}(\mathbf{X},\mathbf{Y})}{\partial \mathbf{x}_i} &= 2\sum_{j=1,j\neq i}^{n}\left(\bar{w}_{i,j}^{(t)}(\mathbf{x}_i-\mathbf{x}_j) - \bar{w}_{i,j}^{(t)}\bar{\delta}_{i,j}^{(t)}\frac{(\mathbf{y}_i-\mathbf{y}_j)}{\|\mathbf{y}_i-\mathbf{y}_j\|}\right) \\ &\quad +4\left(\sum_{j=n+1}^{n+m}\bar{w}_{i,j}^{(t)}(\mathbf{x}_i-\mathbf{x}_j) - \bar{w}_{i,j}^{(t)}\bar{\delta}_{i,j}^{(t)}\frac{(\mathbf{y}_i-\mathbf{y}_j)}{\|\mathbf{y}_i-\mathbf{y}_j\|}\right) \\ &\quad +2r_i(\mathbf{x}_i-\bar{\mathbf{x}}_i) + \lambda p\frac{(\mathbf{x}_i-\mathbf{x}_i^{(t-1)})}{\|\mathbf{y}_i-\mathbf{x}_i^{(t-1)}\|^{2-p}} \end{aligned} \tag{6.42}$$

and

$$\frac{\partial T_k^{(t)}(\mathbf{X},\mathbf{Y})}{\partial \mathbf{x}_i} = 2\left(\bar{w}_{i,k}^{(t)}(\mathbf{x}_i-\mathbf{x}_k) - \bar{w}_{i,k}^{(t)}\bar{\delta}_{i,k}^{(t)}\frac{(\mathbf{y}_i-\mathbf{y}_k)}{\|\mathbf{y}_i-\mathbf{y}_k\|}\right). \tag{6.43}$$

Substituting (6.42) and (6.43) in (6.41) yields,

$$\begin{aligned}\frac{\partial T^{(t)}(\mathbf{X},\mathbf{Y})}{\partial \mathbf{x}_i} &= 4\left(\sum_{j=1,j\neq i}^{n+m} \bar{w}_{i,j}^{(t)}(\mathbf{x}_i-\mathbf{x}_j) - \bar{w}_{i,j}^{(t)}\bar{\delta}_{i,j}^{(t)}\frac{(\mathbf{y}_i-\mathbf{y}_j)}{\|\mathbf{y}_i-\mathbf{y}_j\|}\right) \\ &\quad +2r_i(\mathbf{x}_i-\bar{\mathbf{x}}_i) + \lambda p\frac{(\mathbf{x}_i-\mathbf{x}_i^{(t-1)})}{\|\mathbf{y}_i-\mathbf{x}_i^{(t-1)}\|^{2-p}}. \qquad (6.44)\end{aligned}$$

Setting the derivatives to zero yields the following recursive update rule

$$\mathbf{x}_i^k = \frac{1}{a_i}\left(c_i + \left[\mathbf{x}_1^{(k-1)},\ldots,\mathbf{x}_N^{(k-1)}\right]\mathbf{b}_i^{(k-1)}\right), \qquad (6.45)$$

where $\mathbf{x}_i^k$ denotes the location of node i at iteration k. Furthermore, $\mathbf{b}_i^k = [b_1^k, b_2^k, \ldots, b_N^k]$ and

$$b_i^k = 4\left(\sum_{j=1,j\neq i}^{n+m}\frac{\bar{w}_{i,j}^{(t)}\bar{\delta}_{i,j}^{(t)}}{\|\mathbf{x}_i^k-\mathbf{x}_j^k\|}\right), \qquad (6.46)$$

$$b_j^k = 4\left(\bar{w}_{i,j}^{(t)} - \frac{\bar{w}_{i,j}^{(t)}\bar{\delta}_{i,j}^{(t)}}{\|\mathbf{x}_i^k-\mathbf{x}_j^k\|}\right), \quad j\neq i, \qquad (6.47)$$

$$a_i = 4\sum_{j=1,j\neq i}^{n+m}\bar{w}_{i,j}^{(t)} + 2r_i + \frac{\lambda p}{\|\mathbf{x}_i^k-\mathbf{x}_i^{t-1}\|^{2-p}}, \qquad (6.48)$$

$$c_i = 2r_i\bar{\mathbf{x}}_i + \frac{\lambda p\mathbf{x}_i^{(t-1)}}{\|\mathbf{x}_i^{k-1}-\mathbf{x}_i^{(t-1)}\|}. \qquad (6.49)$$

The dwMDS algorithm in [35] obtains a recursive update for location $\mathbf{x}_i$ by setting the derivatives of the surrogate to the i^{th} local cost function ($T_i^{(t)}(\mathbf{X},\mathbf{Y})$) to zero. This is equivalent to minimizing the global cost function only under anchor free localization ($m=0$) and no a priori information ($r_i=0$). However, in our algorithm, we use the local cost functions only to derive a majorizing function for the global cost function and not in the minimization. Moreover, the algorithm is still decentralized in its implementation though we minimize the global cost function with respect to the sensor locations $\mathbf{X}$.

6.9 Appendix: experimental setup

To model the inter-sensor measurements in the presence of a target, we conducted experiments using following setup: we constructed a fine grid of locations in a free-space environment, where the target was placed and RSS measurements were recorded between two static sensors for positions on the grid. Crossbow Technology Inc MICA2 motes (MPR400 915 MHz models) running the TinyOS operating system were used for the experiment. Packets containing several bytes of useful information, e.g., broadcast source, packet number, local battery voltage can be sent through the Chipcon CC1000 radio, which operated at 910MHz. NesC code was tailored for two of the motes, the transmitter and the receiver, while the third mote relayed packets to a laptop using readily-available code. The radio of the transmitting motes was programmed to broadcast at various power settings. The radio of the receiving motes measured the RSS of each received packet, and then broadcasted another packet containing this information intended for the base node. Packets were labeled with their source to avoid confusion. The motes used in the experiment are shown in Fig. 6.26(a). An aluminium wrapped basketball shown in Fig. 6.26(b) was used as a reflective target. The environment under which the experiments were performed is fairly clutter-free and is shown in Fig. 6.26(c). The experimental setup is illustrated through Fig. 6.26(d). The statistical likelihood model described in (6.15) was formulated based on the measurements collected from this experiment.

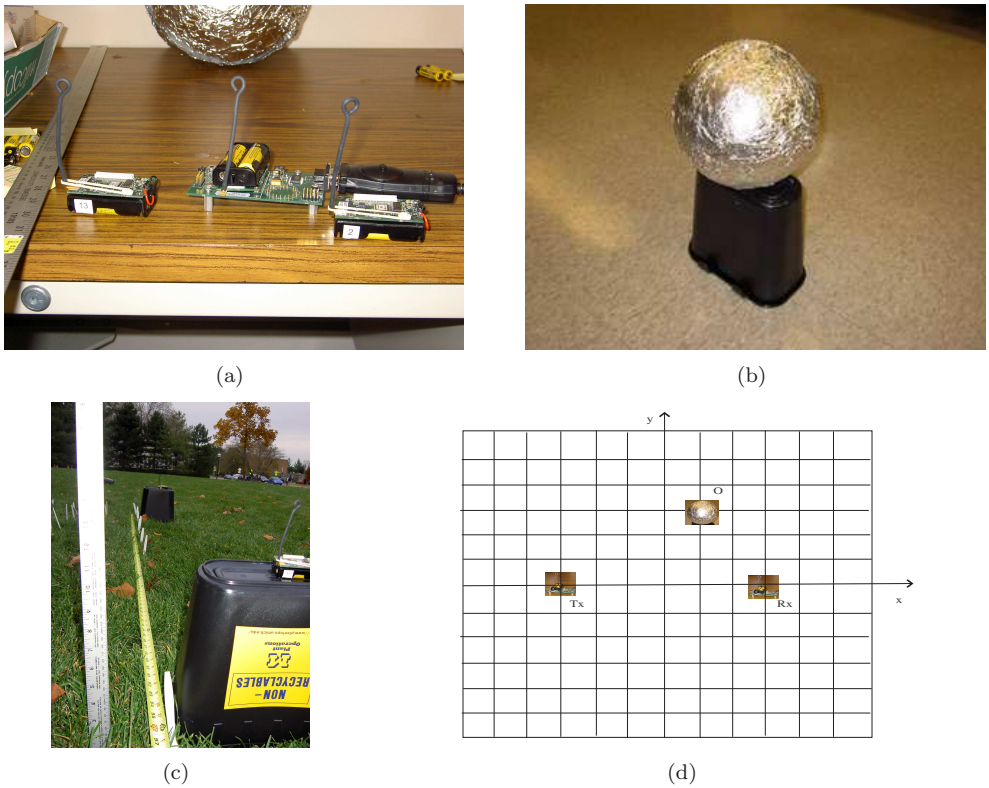


Figure 6.26: Experimental setup for obtaining statistics of RSS measurements. (a) Crossbow Technology Inc MICA2 sensors, (b) reflective target, (c) environment, (d) experimental setup.

6.10 Appendix: optimal likelihood ratio test

To test the presence of a target on a sensor link i, j, we pose the following hypotheses testing problem

$$\begin{aligned} \mathrm{H}_0 &: P_1, \ldots, P_M \sim \mathcal{N}(\bar{P}, \sigma_0^2) \\ \mathrm{H}_1 &: P_1, \ldots, P_M | \hat{P} \sim \mathcal{N}(\hat{P}, \sigma_0^2), \text{ i.i.d}, \quad \hat{P} \sim \mathcal{N}(\bar{P}, \sigma_1^2), \end{aligned}$$

where $P_1, \ldots, P_M$ are the measurements made by a particular link i, j. We leave out the indices i, j in the measurements for brevity. $\bar{P}$ is the mean received power in the sensor link i, j. We assume it can be obtained during the system setup in the absence of targets. Denote the measurements by the M-element vector $\mathbf{p} = [P_1, P_2, \ldots, P_M]^T$. Then the hypotheses can be written as

$$\begin{aligned} \mathrm{H}_0 &: \mathbf{p} \sim \mathcal{N}(\bar{P}\mathbf{1}, \sigma_0^2 \mathbf{I}) \\ \mathrm{H}_1 &: \mathbf{p} \sim \mathcal{N}(\bar{P}\mathbf{1}, \sigma_1^2 \mathbf{1}\mathbf{1}^T + \sigma_0^2 \mathbf{I}). \end{aligned} \tag{6.50}$$

To construct the LRT, we first compute the log likelihood ratio as

$$\begin{aligned} \Lambda &= \log \left(\frac{f(\mathbf{p}|\mathrm{H}_1)}{f(\mathbf{p}|\mathrm{H}_0)} \right) \\ &= \frac{1}{2}(\mathbf{p} - \bar{P}\mathbf{1})^T (\mathbf{C}_0^{-1} - \mathbf{C}_1^{-1})(\mathbf{p} - \bar{P}\mathbf{1}) + \frac{1}{2} \log \left(\frac{|\mathbf{C}_0|}{|\mathbf{C}_1|} \right), \end{aligned} \tag{6.51}$$

where $\mathbf{C}_0 = \sigma_0^2 \mathbf{I}$, $\mathbf{C}_1 = \sigma_1^2 \mathbf{1}\mathbf{1}^T + \sigma_0^2 \mathbf{I}$ and $|\mathbf{C}|$ denotes the determinant of a matrix $\mathbf{C}$. The eigendecompositions of the covariance matrices $\mathbf{C}_0$ and $\mathbf{C}_1$ can be written as

$$\begin{aligned} \mathbf{C}_0 &= \mathbf{V}_0 \mathbf{D}_0 \mathbf{V}_0^T, \\ \mathbf{C}_1 &= \mathbf{V}_1 \mathbf{D}_1 \mathbf{V}_1^T, \end{aligned}$$

where $\mathbf{D}_i$ is a diagonal matrix composed of the eigenvalues $\lambda_1^i, \ldots, \lambda_M^i$ and $\mathbf{V}_i$ is the matrix of corresponding eigenvectors. The eigenvalues of the covariance matrix

$\mathbf{C}_1$ are given by $\lambda_1^1 = \sigma_1^2 M + \sigma_0^2$ and $\lambda_i^1 = \sigma_0^2,\ i = 2, \ldots, M$. The corresponding eigenvectors are $\mathbf{v}_1 = \mathbf{1}/\sqrt{M}, \mathbf{v}_2, \ldots, \mathbf{v}_M$, where $\{\mathbf{v}_i\}_{i=1}^M$ are a set of orthogonal unit norm vectors. The eigenvalues of $\mathbf{C}_0$ are all σ_0^2 and it is easy to verify that $\mathbf{v}_1, \ldots, \mathbf{v}_M$ are eigenvectors to $\mathbf{C}_0$, i.e., $\mathbf{V}_0 = \mathbf{V}_1$. Thus

$$\mathbf{C}_0^{-1} - \mathbf{C}_1^{-1} = \mathbf{V}_0 \mathrm{diag}\left(\frac{M\sigma_1^2}{M\sigma_1^2 + \sigma_0^2}, 0, \ldots, 0\right) \mathbf{V}_0^T = \frac{\sigma_1^2 M}{\sigma_1^2 M + \sigma_0^2} \frac{\mathbf{1}\mathbf{1}^T}{M}. \tag{6.52}$$

Substituting (6.52) in (6.51) and collecting constant terms at the right hand side yields the optimal LRT as

$$|\bar{p} - \bar{P}| \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \gamma, \tag{6.53}$$

where $\bar{p} = \sum_{i=1}^M P_i/M$ is the minimal sufficient statistics of this test. Under H_0, $\bar{p}$ is distributed as $\mathcal{N}(\bar{P}, \sigma_0^2/M)$ and under H_1, $\bar{p}$ is $\mathcal{N}(\bar{P}, \sigma_0^2/M + \sigma_1^2)$. We find γ to satisfy a false alarm of level α, i.e.,

$$P\left(|\bar{p} - \bar{P}| > \gamma | \mathrm{H}_0\right) = 2Q\left(\frac{\sqrt{M}\gamma}{\sigma_0}\right) = \alpha, \tag{6.54}$$

which implies $\gamma = (\sigma_0/\sqrt{M})Q^{-1}(\alpha/2)$. The probability of correct decision, β is then given by

$$\begin{aligned} \beta &= P\left(|\bar{p} - \bar{P}| > \gamma | \mathrm{H}_1\right) \\ &= 2Q\left(\frac{\gamma}{\sqrt{\sigma_0^2/M + \sigma_1^2}}\right) \\ &= 2Q\left(Q^{-1}(\alpha/2)\sqrt{\frac{\sigma_0^2}{\sigma_0^2 + M\sigma_1^2}}\right). \end{aligned} \tag{6.55}$$

CHAPTER VII

Conclusion and future directions

That's one giant leap for me, one small step for mankind.

The need for optimal allocation of resources (e.g., energy, sensors, and bandwidth) in many adaptive sensing applications has been of much interest to different research communities in recent years. The research presented in this dissertation attempts to take one small step forward in the direction of providing optimal resource management strategies to adaptive sensing. While the work was motivated by practical problems in radar and sensor networks, the results of this dissertation can be widely applied. The theoretical results of Chapters II-IV and asymptotic analysis in Chapter II can be applied to any application where energy constraints make sense, e.g., X-ray CT, radiation oncology, and communications. We were able to show that considerable performance improvements can be achieved if adaptive designs use energy optimally.

The latter part of the thesis aimed towards developing optimal algorithms for target and sensor localization in sensor networks that could be implemented real-time and useful for analyzing real data. Along the way, we also suggested methods of applying the experimental designs for adaptive resource allocation to radar imaging,

channel estimation, and sensor network applications.

We conclude this dissertation by providing some ideas for future research.

Energy allocation

The first part of this dissertation considered the problem of waveform design and energy allocation strategies for a general class of estimation and detection problems. Applications of these results to inverse scattering, channel estimation, and target tracking problems were discussed. It would be satisfying to see these designs implemented in hardware for such applications to confirm the theory and simulations presented here.

Another potential application of the energy allocation procedure is to medical imaging and radiation therapy. Computed Tomography (CT) is a widely used method for generating three dimensional images of the internal organs of a body using multiple two dimensional X-ray images. There has been a growing need for higher resolution images which can improve the detection of abnormalities (tumors, cysts, infections) at an earlier stage. However, higher resolution comes at a cost of higher radiation dose which could produce harmful reactions on the human body. An energy allocation procedure similar to the one proposed in Chapters II-IV could prove extremely useful to generate high quality images without having to expose the body to high amounts of radiation. By giving lower levels of energy for a slightly longer time period of time, one could minimize high levels of radiation to the patient. It would be worthwhile to study whether such a strategy could be applied in this context.

Sensor scheduling

In the previous chapter, we implemented the sparsity penalized MDS algorithm that could simultaneously localize sensors as well as targets. However, further savings could be realized if we could design a sensor management strategy that enables only those sensors which are necessary to identify the target. In the absence of any target motion model, one could devise a simple strategy based on the sparsity penalized MDS as follows: given the set of tagged sensors, i.e., sensor links with high output through the distance based or LRT methods, we could activate all sensors within a specified neighborhood and perform MDS locally to obtain the new position of the target. This approach will enable the sensors further away from the target to go into sleep mode and could be activated at a later stage.

If we are given a target motion model, then we could borrow the concepts of sparse waveform selection method proposed in Chapter V of this dissertation for sparse sensor selection. Consider the problem of choosing s sensors to activate at each time step in a sensor network of n wireless sensors ($s \ll n$) to minimize a target state prediction error. This problem falls into the framework of sparse selection described in Chapter V. We could then implement a convex relaxation solution and show near optimal performance to this sensor selection problem.

Completion problems

The study of the sparsity penalized MDS as a real-time application tool for the analysis of the UCSD wireless trace data has opened many interesting problems for wireless topology discovery. The UCSD wireless traces capture signal strength measurements of users within range of the various APs on campus. However, more than 60% of the AP locations were unknown, thereby raising the question of whether

the rest of the network topology could be discovered. In particular, given a set of distances from the user to the known and unknown AP points, is it possible for us to discover the entire pairwise distance matrix which then can be translated into location estimates of the user and the AP points? We presented simple conditions under which an accurate reconstruction is feasible in the absence of noise and with the knowledge of all distances between the users and the APs. The presence of noise and missing data (only certain user-AP distances are known) brings in the topic of least squares reconstruction. In this context, there is extensive research on Euclidean distance matrix completion problems (EDMCP) which finds a solution to this problem via semi definite programming. It would be worthwhile to study the theory of EDMCP and implement this approach for topology discovery. Furthermore, inclusion of side information would add an additional dimension to the problem. For example, incorporating constraints such as building layouts, attenuation map for each AP, and information about transmit signal strengths should provide significant performance improvement to the localization algorithm.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] S. Al-Homidan and H. Wolkowicz. Approximate and exact completion problems for Euclidean distance matrices using semidefinite programming. *Linear Algebra and its Applications*, 406:109–141, 2005.

[2] S. M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE J. Sel. Areas Commun.*, 16:1451–1458, 1998.

[3] A. Y. Alfakih. On the uniqueness of Euclidean distance matrix completions. *Linear Algebra and its Applications*, 370:1–14, 2003.

[4] D. L. Alspach and H. W. Sorensen. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Trans. Automat. Contr.*, 82:1032–1063, 1987.

[5] F. J. Anscombe. Sequential estimation. *J. Roy. Statist. Soc. Ser. B*, 15:1–29, 1953.

[6] A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Demirbas, M. Gouda, Y. Choi, T. Herman, and S. Kulkarni. A line in the sand: a wireless sensor network for detection, classification, and tracking. *Computer Networks*, 46(5):605–634, 2004.

[7] M. Athans. On the determination of optimal costly measurement strategies for linear stochastic systems. *Automatica*, 8:397–412, 1972.

[8] M. Athans and F. C. Schweppe. Optimal waveform design via control theoretic principles. *Inform. Control*, 10:335–377, 1967.

[9] M. Bakonyi and C. R. Johnson. The Euclidean distance matrix completion problem. *SIAM J. Matrix. Anal. Appl.*, 16(2):646–654, April 1995.

[10] G. Bal, G. C. Papanicolaou, and L. Ryzhik. Self averaging in time reversal for the parabolic wave equation. *Stochastics and Dynamics*, 2:507–531, 2002.

[11] Y. Bar-Shalom. *Multitarget Multisensor Tracking: Advanced Applications*. Artech House, 1990.

[12] B. Baygun and A. O. Hero III. Optimal simultaneous detection and estimation under a false alarm constraint. *IEEE Trans. on Info. Thy.*, 41(3):688–703, 1995.

[13] M. R. Bell. Information theory and radar waveform design. *IEEE Trans. on Inform. Theory*, 39(5):1578–1597, Sep. 1993.

[14] D. P. Bertsekas and D. Castanon. Rollout algorithms for stochastic scheduling problems. *J. Heuristics*, 5(1):89–108, 1999.

[15] M. Biguesh and A. B. Gershman. Training based MIMO channel estimation: A study of estimator tradeoffs and optimal trainign signals. *IEEE Trans. Signal Processing*, 54(3):884–892, 2006.

[16] R. E. Blahut, W. M. Miller, and C. H. Wilcox. *Radar and Sonar, Part I.* Springer-Verlag, New York, 1991.

[17] N. Bleistein, J. K. Cohen, and J. W. Stockwell Jr. *Mathematics of Multidimensional Seismic Imaging, Migration and Inversion*. Springer, New York, 2001.

[18] D. Bliss, K. W. Forysthe, A. O. Hero III, and A. F. Yegulalp. Environmental issues for MIMO capacity. *Signal Processing*, 50(9):2128–2142, 2002.

[19] P. Blomgren, G. C. Papanicolaou, and H. Zhao. Super-resolution in time reversal acoustics. *J. Acoust. Soc. Am.*, 111:238–248, 2002.

[20] L. Borcea, G. Papanicolaou, and C. Tsogka. Theory and applications of time reversal and interferometric imaging. *J. Acoust. Soc. Am*, 19:5139–4164, 2003.

[21] L. Borcea, G. C. Papanicolaou, and C. Tsogka. Estimation of the refocusing resolution for time reversal in scattering media. *SIAM J. Multiscale Modeling and Simulation*, 10:10–20, 2003.

[22] L. Borcea, G. C. Papanicolaou, C. Tsogka, and J. Berryman. Imaging and time reversal in random media. *Inverse Problems*, 18:1247–1279, 2002.

[23] V. Z. Borisov and V. V. Konev. On sequential parameter estimation in discrete time process. *Automat. Remote Contr.*, 38:58–64, 1977.

[24] R. R. Brooks, P. Ramanathan, and A. M. Sayeed. Distributed target classification and tracking in sensor networks. *Proc. IEEE*, 91(8):1163–1171, 2003.

[25] C. Buianu and L. Tong. Channel estimation for space-time orthogonal block codes. *IEEE Trans. Signal Processing*, 50:2515–2528, 2002.

[26] S. Capkun, L. Buttyan, and J. Hubaux. SECTOR: Secure tracking of node encounters in multi-hop wireless networks. In *Proc. 1st ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN)*, pages 21–32, 2003.

[27] L. Carin, H. Liu, T. Yoder, L. Couchman, B. Houston, and J. Bucaro. Wideband time-reversal imaging of an elastic target in an acoustic waveguide. *J. Acoust. Soc. Am.*, 115(1):259–268, 2004.

[28] V. Cevher and J. H. McClellan. Sensor array calibration via tracking with the extended Kalman filter. In *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, volume 5, pages 2817–2820, 2001.

[29] J. Chamberland and V. V. Veeravalli. Decentralized detection in sensor networks. *IEEE Trans. Signal Processing*, 51(2):407–416, 2003.

[30] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Scientific Computing*, 20(1):33–61, 1998.

[31] H. Chernoff. Sequential design of experiments. *Ann. Math. Statist.*, 30:755–770, 1959.

[32] J. F. Claerbout. *Fundamentals of Geophysical Data Processing: with Applications to Petroleum Prospecting*. Blackwell, Palo Alto, CA, 1985.

[33] S. Clark and H. Durrant-Whyte. Autonomous land vehicle navigation using millimeter wave radar. In *Proc. Intl. Conf. Robotics and Automation*, volume 4, pages 3697–3702, 1998.

[34] W. Cleveland. Robust locally weighted regression and smoothing scatterplots. *J. Am. Statist. Assoc.*, 74(368):829–836, 1979.

[35] J. Costa, N. Patwari, and A. O. Hero III. Distributed multidimensional scaling with adaptive weighting for node localization in sensor networks. *ACM J. Sensor Networking*, 2(1):39–64, 2006.

[36] A. J. Coulson, A. G. Williamson, and R. G. Vaughan. A statistical basis for lognormal shadowing effects in multipath fading channels. *IEEE Trans. on Veh. Tech.*, 46(4):494–502, 1998.

[37] H. Cox. Robust adaptive beamforming. *IEEE Trans. Acoust., Speech, Signal Processing*, 35:1365–1376, 1987.

[38] H. Cox, R. Zeskind, and M. Owen. Robust adaptive beamforming. *IEEE Trans. Acoust., Speech, Signal Processing*, 35(10):1365–1376, Oct. 1987.

[39] T. Cox and M. Cox. *Multidimensional Scaling.* Chapman & Hall, London, 1994.

[40] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure App. Math.*, 57(11):1413–1457, Nov. 2004.

[41] F. Daum. A system approach to multiple target tracking. *Multitarget-Multisensor Tracking:Application and Advances*, 2:149–181, 1992.

[42] M. L. Davidson. *Multidimensional scaling.* Wiley, New York, NY, 1983.

[43] J. de Leeuw. Applications of convex analysis to multidimensional scaling. In J. Barra, F. Brodeau, G. Romier, and B. van Custem, editors, *Recent developments in statistics*, pages 133–145. North Holland Publishing Company, Amsterdam, The Netherlands, 1977.

[44] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez. Particle filtering. *IEEE Signal Processing Magazine*, 20(5):19–38, 2003.

[45] L. Doherty, K. S. Pister, and L. E. Ghaoui. Convex position estimation in wireless sensor networks. In *Proc. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, volume 3, pages 1655–1663, 2001.

[46] D. L. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. on Inform. Theory*, 52(1):6–18, 2006.

[47] D. R. Dowling and D. R. Jackson. Narrow band performance of phase conjugate arrays in dynamic random media. *J. Acoust. Soc. Am.*, 91:3257–3277, 1992.

[48] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Ann. Statist.*, 32(2):407–499, 2004.

[49] Y. C. Eldar and A. V. Oppenheim. Covariance shaping least-squares estimation. *IEEE Trans. Signal Processing*, 51(3):686–697, 2003.

[50] C. Estan, S. Savage, and G. Varghese. Automatically inferring patterns of resource consumption in network traffic. In *ACM SIGCOMM*, pages 137–148, 2003.

[51] J. S. Evans and V. Krishanmurthy. Optimal sensor scheduling for hidden markov model state estimation. *Int. J. Contr.*, 74(18):1737–1742, 2001.

[52] V. V. Fedorov. *Theory of Optimal Experiments.* Academic Press, New York, 1972.

[53] M. Fink. Time reversal acoustics. *Physics Today*, pages 34–40, 1997.

[54] M. Fink and C. Pradha. Acoustic time-reversal mirrors. *Inverse Problems*, 17:1–38, 2001.

[55] G. J. Foschini and M. J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Commun.*, 6:311–335, 1998.

[56] S. G. Ghurye and H. Robbins. Two-stage experiments for estimating the difference between means. *Biometrika*, 41:146–152, 1954.

[57] J. C. Gower and G .B. Dijksterhuis. *Procrustes Problems*. Oxford University Press, 2004.

[58] M. J. Greenacre. *Theory and Applications of Correspondence Analysis*. Academic Press Inc., London, UK, 1984.

[59] C. J. Gregory, R. R. Carthy, and L. G. Pearlstine. Survey and monitoring of species at risk at camp blanding training site, northeastern florida. *Southeastern Naturalist*, 5(3):473–498, 2006.

[60] R. Gribonval and M. Nielsen. Sparse decompositions in unions of bases. *IEEE Trans. on Inform. Theory*, 49(12):3320–3325, 2003.

[61] P. Groenen. *The majorization approach to multidimensional scaling: some problems and extensions*. DSWO Press, 1993.

[62] F. K. Gruber, E. A. Marengo, and A. J. Devaney. Time reversal imaging with multiple signal classification considering multilpe scattering between the targets. *J. Acoust. Soc. Am.*, 115:3042–3047, 2004.

[63] J. B. S. Haldane. On a method of estimating frequencies. *Biometrika*, 33:222–225, 1945.

[64] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer Series in Statistics. Springer Verlag, New York, 2000.

[65] S. Haykin. Cognitive radar: a way of the future. *IEEE Signal Processing Magazine*, 23(1):30–40, Jan. 2006.

[66] A. O. Hero III. *Statistical Methods for Signal Processing*. Course Notes, Ann Arbor, MI, 2005.

[67] A. O. Hero III. Geometric entropy minimization (GEM) for anomaly detection and localization. In *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2006.

[68] Y. Hu, A. Perrig, and D. Johnson. Packet leashes: A defense against wormhole attacks in wireless ad hoc networks. In *Proc. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, volume 3, pages 1976–1986, 2003.

[69] A. Ishimaru. *Wave Propagation and Scattering in Random Media*. Piscataway, NJ:IEEE, 1997.

[70] D. R. Jackson and D. R. Dowling. Phase conjugation in under water acoustics. *J. Acoust. Soc. Am.*, 89:171–181, 1991.

[71] W. James and C. M. Stein. Estimation with quadratic loss. In *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, volume 1, pages 361–379, 1961.

[72] X. Ji and H. Zha. Sensor positioning in wirless ad-hoc sensor networks with multidimensional scaling. In *Proc. IEEE Infocom*, pages 2652–2661, 2004.

[73] S. J. Julier and J. K. Uhlman. A new extension of the kalman filter to nonlinear systems. In *Proc. AeroSense: Eleventh Intl. Symp. on Aerospace/Defense Sensing, Simulations and Control Multi Sensor Fusion, Tracking and Resource Management II*, volume 3068, pages 182–193, 1997.

[74] K. B. Gray, Jr. Sequential selection of experiments. *Ann. Math. Statist.*, 39:1953–1977, 1968.

[75] S. M. Kay. *Statistical Estimation*. Prentice-Hall, Englewood-Cliffs, N.J., 1991.

[76] E. Kerbat, C. Prada, D. Cassereau, and M. Fink. Imaging in the presence of grain noise using the decomposition of the time reversal operator. *J. Acuost. Soc. Am.*, 113:1230–1240, 2003.

[77] E. Kerbrat, C. Prada, and M. Fink. Ultrasonic nondestructive testing of scattering media using the decomposition of the time-reversal operator. *IEEE Trans. on Ultra., Ferro., and Freq., Control*, 49:1103–1112, 2002.

[78] D. J. Kershaw and R. J. Evans. Optimal waveform selection for tracking systems. *IEEE Trans. Inform. Theory*, 40(5):1536–1550, 1994.

[79] D. J. Kershaw and R. J. Evans. Waveform selective probabilistic data association. *IEEE Trans. Aerosp. Electron. Syst.*, 33(4):1180–1188, 1997.

[80] J. Kiefer. Optimal experimental designs. *J. Roy. Statist. Soc. Ser. B(Methodological)*, 21:272–319, 1959.

[81] J. Kim, H.C. Song, and W.A. Kuperman. Adaptive time reversal mirror. *J. Acoust. Soc. Am*, 109:1817–1825, 2001.

[82] W. Kim, K. Mechitov, J. Y. Choi, and S. Ham. On target tracking with binary proximity sensors. In *Proc. Fourth Intl Symposium Information Processing in Sensor Networks*, pages 301–308, 2005.

[83] V. V. Konev and T. L. Lai. Estimators with prescribed precision in stochastic regression models. *Sequential Anal.*, 14:179–192, 1995.

[84] S. Kotz, N. L. Johnson, and D. W. Boyd. Series representations of distributions of quadratic forms in normal variables. i. central case. *Ann. Math. Statist.*, 38(3):823–837, 1967.

[85] C. M. Kreucher. *An Information-based Approach to Sensor Resource Allocation*. PhD thesis, University of Michigan, Ann Arbor, 2005.

[86] V. Krishanmurthy. Algorithms for optimal scheduling and management of hidden markov model sensors. *IEEE Trans. Signal Processing*, 50(6):1382–1396, 2002.

[87] J. Kruskal. Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hyptothesis. *Psychometrika*, 29:1–27, 1964.

[88] J. Kruskal. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29:115–129, 1964.

[89] J. Kumagai and S. Cherry. Sensors and sensibility. *IEEE Spectrum*, 41(7):22–28, 2004.

[90] P. R. Kumar and P. Varaiya. *Stochastic Systems - Estimation, Identificiation and Adaptive Control*. Prentice-Hall, Englewood Cliffs, N.J, 1986.

[91] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson. Phase conjugation in the ocean: Experimental demonstration of an acoustic time reversal mirror. *J. Acoust. Soc. Am.*, 103:25–40, 1998.

[92] K. Lange, D. R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, 9(1):1–20, 2000.

[93] E. G. Larsson and P. Stoica. *Space-Time Block Coding for Wireles Communications*. Cambridge Univ. Press, Cambridge, U.K, 2003.

[94] M. S. Lee and Y. H. Kim. An efficient multitarget tracking for car applications. *IEEE Trans. Industrial Electronics*, 50(2):397–399, 2003.

[95] M. Leoncini, G. Resta, and P. Santi. Analysis of a wirless sensor dropping problem for wide-area environmental monitoring. In *Fourth Intl. Symp. on Inform. Processing Sensor Networks (IPSN)*, pages 239–245, 2005.

[96] Y. Li. Optimal training sequences for ofdm systems with multiple transmite antennas. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, volume 3, pages 1478–1482, 2000.

[97] H. Liu and L. Carin N. Dasgupta. Time reversal imaging for wideband underwater target classification. In *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, volume 5, pages 6–10, 2003.

[98] D. Malioutov, M. Cetin, and A. S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Processing*, 53(8):3010–3022, Aug. 2005.

[99] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Processing*, 41(12):3397–3415, Dec. 1993.

[100] C. C. Martin, J. H. Winters, and N. R. Sollenberger. Multiple-input multiple-output (mimo) radio channel measurements. *Proc. VTC'00*, 2:774–779, 2000.

[101] A. C. Martinez and F. L . Blazquez. Distribution of a sum of weighted noncentral chi-square variables. *Sociedad de Estadistica e Investigacion Operativa Test*, 14(2):397–415, 2005.

[102] T. L. Marzetta. BLAST training: Estimating channel characteristics for high capacity space-time wireless. In *Proc. 37th Annual Allerton Conf. Communications, Control, Computing*, pages 958–966, Sep. 1999.

[103] D. Massonnet and K. L. Feigl. Radar interferometry and its application to changes in earth's surface. *Review of Geophysics*, 36(4):441–500, 1998.

[104] L. S. Mayer and T. A. Willke. On biased estimation in linear models. *Technometr.*, 15:497–508, 1973.

[105] L. Meier, J. Perschon, and R. M. Dressier. Optimal control of measurement subsystems. *IEEE Trans. Automat. Contr.*, AC-12:528–536, 1967.

[106] M. Mobed and W. L. Root. Estimation of constrained parameters in a linear model with multiplicative and additive noise. *IEEE Trans. on Info. Thy.*, 40(1):56–66, 1994.

[107] G. Montaldo, M. Tanter, and M. Fink. Revisiting iterative time reversal processing: Application to detection of multiple targets. *J. Acoust. Soc. Am*, 115:776–784, 2004.

[108] R. L. Moses, D. Krishanmurthy, and R. Patterson. A self-localization method for wireless sensor networks. *EURASIP J. Applied Signal Processing*, 4:348–358, 2003.

[109] R. Nagpal, H. Shrobe, and J. Bachrach. Organizing a global coordinate system from local information on an ad hoc sensor network. In *Proc. 2nd Intl Workshop on Inform. Processing Sensor Networks (IPSN), Lecture Notes in Computer Science*, volume 2634, pages 333–348, 2003.

[110] H. Naparst. Dense target signal processing. *IEEE Trans. on Inform. Theory*, IT-37(2):317–327, Mar. 1991.

[111] D. Niculescu and B. Nath. Ad hoc positioning systems. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, volume 5, pages 2926–2931, 2001.

[112] R. Niu, P. K. Varshney, and Q. Cheng. Distributed detection in a large wireless sensor network. *Information Fusion*, 7:380–394, 2006.

[113] C. J. Nolan and M. Cheney. Synthetic aperture inversion. *Inverse Problems*, 18:221–235, 2002.

[114] M. Osborne, B. Presnell, and B. Turlach. A new approach to variable selection in least squares problems. *IMA J. Numeric. Anal.*, 20(3):389–403, 2000.

[115] E. S. Page. Continuous inspection schemes. *Biometrika*, 41:100–114, 1954.

[116] G. Papanicolaou, L. Ryzhik, and K. Solna. Statistical stability in time reversal. *SIAM J. Appl. Math.*, 64:1133–1155, 2004.

[117] P. Pathirana, N. Bulusu, S. Jha, and A. Savkin. Node localization using mobile robots in delay-tolerant sensor networks. *IEEE Trans. Mobile Computing*, 4(3):285–296, 2005.

[118] N. Patwari, A. O. Hero III, M. Perkins, N. S. Correal, and R. J. O'Dea. Relative location estimation in wireless sensor networks. *IEEE Trans. Signal Processing*, 51(8):2137–2148, 2003.

[119] N. Patwari, A. O. Hero III, and A. Pacholski. Manifold learning visualization of network traffic data. In *Proc. Workshop on Mining Network Data, Philadelphia, PA*, pages 191–196, 2005.

[120] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, N.Y, 1988.

[121] C. Prada and M. Fink. Eigenmodes of the time reversal operator: A solution to selective focusing in multiple-target media. *Wave Motion*, 20:151–163, 1994.

[122] C. Prada, S. Manneville, D. Spoliansky, and M. Fink. Decomposition of the time reversal operator: Detection and selective focusing on two scatterers. *J. Acoust. Soc. Am*, 99:2067–2076, 1996.

[123] G. G. Raleigh and J. M. Cioffi. Spatio-temporal coding for wireless communication. *IEEE Trans. Commun.*, 46(3):357–366, 1998.

[124] J. Ramsay. Some statistical approaches to multidimensional scaling data. *J. R. Statist. Soc.*, 145(3):285–312, 1982.

[125] R. Rangarajan, R. Raich, and A. O. Hero III. Blind tracking using sparsity penalized multidimensional scaling. to appear in *Proc. IEEE Workshop on Stat. Signal Processing*, Aug. 2007.

[126] R. Rangarajan, R. Raich, and A. O. Hero III. Optimal experimental design for an inverse scattering problem. In *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, volume 4, pages 1117–1120, 2005.

[127] R. Rangarajan, R. Raich, and A. O. Hero III. Sequential design for a Rayleigh inverse scattering problem. In *Proc. IEEE Workshop on Stat. Signal Processing*, pages 625–630, 2005.

[128] R. Rangarajan, R. Raich, and A. O. Hero III. Single-stage waveform selection for adaptive resource constrained state estimation. In *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, volume 3, pages 672–675, 2006.

[129] R. Rangarajan, R. Raich, and A. O. Hero III. Optimal sequential energy allocation for inverse problems. *IEEE J. Select. Topics in Signal Processing*, 1(1):67–78, Jun. 2007.

[130] R. Rangarajan, R. Raich, and A. O. Hero III. Sequential energy allocation strategies for channel estimation. In *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, volume 3, pages 821–824, 2007.

[131] R. Rangarajan, R. Raich, and A. O. Hero III. Sparsity penalized mds for blind tracking in sensor networks. In V. Saligrama, editor, *Advances in Sensor Networks*. Springer, NY, 2007.

[132] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas. Transmit beam-forming and power control for cellular wireless systems. *IEEE J. Sel. Areas Commun.*, 16(8):1437–1450, 1998.

[133] D. Richter. Two-stage experiments for estimating a common mean. *Ann. Math. Statist.*, 31:1164–1173, 1960.

[134] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statist.*, 22:400–407, 1951.

[135] H. E. Robbins. Some aspects of sequential design of experiments. *Bull. Amer. Math. Soc.*, 58:527–535, 1952.

[136] H. Rubens. Probability content of regions under spherical normal distributions. iv. the distribution of homogeneous and non homogeneous quadratic functions of normal variables. *Ann. Math. Statist.*, 33(2):542–570, 1962.

[137] S. M. Rytov, A. Yu. Kravtsov, and V. I. Tatarskii. *Principles of Statistical Radiophysics*. Springer-Verlag, Berlin, 1989.

[138] S. Saatchi, J. V. Soares, and M. Alves. Mapping deforestation and land use in amazon rainforest by using sir-c imagery. *Remote Sensing of Environment*, 59(2):191–202, 1997.

[139] A. Savvides, H. Park, and M. B. Srivastava. The bits and flops of the n-hop multilateration primitive for node localization problems. In *Proc. Intl Workshop on Sensor Nets & Apps.*, pages 112–121, 2002.

[140] A. Scaglione and A. Vosoughi. Turbo estimation of channel and symbols in precoded MIMO systems. In *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, volume 4, pages 413–416, 2004.

[141] S. M. Schweizer and J. M. F. Moura. Efficient detection in hyperspectral imagery. *IEEE Trans. Image Processing*, 10(4):584–597, Apr. 2001.

[142] F. C. Schweppe and D. L. Gray. Radar signal design subject to simultaneous peak and average constraints. *IEEE Trans. Inform. Theory*, IT-12:13–26, 1966.

[143] W. R. Scott, J. S. Martin, and G. D. Larson. Experimental model for a seismic landmine detection systems. *IEEE Trans. Geoscience and Remote Sensing*, 39(6):1155–1164, 2001.

[144] D. Seigmund. Herbert Robbins and sequential analysis. *Ann. Statist.*, 31:349–365, 2003.

[145] B. K. Shah and C. G. Khatri. Distribution of a definite quadratic form for non-central normal variables. *Ann. Math. Statist.*, 32(3):883–887, 1961.

[146] Y. Shang, W. Rumi, Y. Zhang, and M. P. Fromherz. Localization from mere connectivity. In *Proc. 4th ACM Intl. Symp. on Mobile ad hoc networking and computing*, pages 201–212, 2003.

[147] W. A. Shewhart. *Economic Control of Manufactured Products*. Van Nostrand Reinhold, Newyork, 1931.

[148] G. Shi and A. Nehorai. Maximum likelihood estimation of point scatterers for computational time-reversal imaging. *Communications in Information and Systems*, 5(2):227–256, 2005.

[149] N. Shrivastava, R. Mudumbai, U. Madhow, and S. Suri. Target tracking with binary proximity sensors: fundamental limits, minimal descriptions, and algorithms. In *Proc. ACM 4th Intl. Conf. Embedded networked sensor systems (SenSys)*, pages 251–264, 2006.

[150] O. Simeone and U. Spagnolini. Channel estimation for block-fading frequency selective Rayleigh MIMO channels. *IEEE Trans. Signal Processing*, 52(11):3265–3277, 2004.

[151] S. P. Sira, A. P. Suppappola, and D. Morrell. Waveform scheduling in wideband environments. In *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, volume 5, pages 1121–1124, 2006.

[152] R. K. Snieder and J. A. Scales. Time reversed imaging as a diagnostic of wave and particle chaos. *Physical Review E*, 58:5668–5675, 1998.

[153] H.C. Song, W.A. Kuperman, and W.S.Hodgkiss. Iterative time reversal in the ocean. *J. Acoust. Soc. Am*, 105:3176–3184, 1999.

[154] S. M. Sowelam and A. H. Tewfik. Optimal waveforms for wideband radar imaging. *J. Franklin Institute - Engg. App. Math.*, 335B(8):1341–1366, Nov. 1998.

[155] S. M. Sowelam and A. H. Tewfik. Waveform selection in radar target classification. *IEEE Trans. on Inform. Theory*, 46(3):1014–1029, May 2000.

[156] C. M. Stein. A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.*, 16:243–258, 1945.

[157] C. M. Stein. Some problems in sequential estimation. *Econometrica*, 17:77–78, 1949.

[158] C. M. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proc. 3rd Berkeley Symp. Math. Stat. Prob.*, volume 1, pages 197–206, 1956.

[159] R. J. Sullivan. *Radar Foundations for Imaging and Advanced Concepts*. Scitech Publishing, Raleigh, NC, 2003.

[160] A. Sutin, A. Sarvazyan, P. Johnson, and J. TenCate. Land mine detection by time reversal acousto-seismic method. *J. Acoust. Soc. Am.*, 115:2384, 2004.

[161] C. Taylor, A. Rahimi, J. Bachrach, H. Shrobe, and A. Grue. Simultaneous localization, calibration, and tracking in an ad hoc sensor network. In *Proc. 5th Intl. Conf. Proc. Information Processing in Sensor Networks (IPSN)*, pages 27–33, 2006.

[162] I. Telatar. Capacity of multiple antenna Gaussian channels. *Eur. Trans. Telecommun.*, 10(6):585–95, 1999.

[163] J. L. Thomas, F. Wu, and M. Fink. Time reversal focusing applied to lithotrypsy. *Ultrason. Imaging*, 18:106–121, 1996.

[164] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Statist. Soc.*, 58:267–288, Nov. 1996.

[165] A. N. Tikhonov and V. Y. Arsenin. *Solution to Ill-Posed Problems*. V. H. Winston, Washington D.C, 1977.

[166] M. W. Trosset. Distance matrix completion by numerical optimization. Technical report, Rice University, Houston, Texas, 1997.

[167] A. Wald. Sequential tests of statistical hypotheses. *Ann. Math. Statist.*, 16:117–186, 1945.

[168] E. A. Wan and R. Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *IEEE Symp. Adaptive Systems for Signal Processing, Communications, and Control*, pages 153–158, 2000.

[169] N. Wang, M. H. Wang, and N. Q. Zhang. Wireless sensors in agriculture and food industry: Recent developments and future perspective. *Computers and electronics in agriculture*, 50(1):1–14, 2006.

[170] W. Wang and B. Bhargava. Visualization of wormholes in sensor networks. In *Proc. ACM Workshop on wireless security*, pages 51–60, 2004.

[171] M. B. Zarrop. *Optimal Experiment Design for Dynamic System Identification.* Springer-Verlag, New York, 1979.