Efficient Resource Allocation Schemes for Search

by

Eran Bashan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2008

Doctoral Committee:

Professor Alfred O. Hero III, Co-Chair
Professor Jeffrey A. Fessler, Co-Chair
Professor Robert W. Keener
Professor Demosthenis Teneketzis
Assistant Professor Michael Wakin

© Eran Bashan 2008
All Rights Reserved

To my late grandfather, Yehuda Borenstein, a meister who taught me to believe

ACKNOWLEDGEMENTS

This work has seen support from many individuals, and I would like to use this opportunity to name a few. To begin with, I would like to thank Professor Jeff Fessler. He convinced me to join the University of Michigan and start this journey under his wings. I have enjoyed Professor Fessler's knowledge, inspiration, guidance, mentorship, and above all friendship and kind words in times of need. Also, I am in great debt to Professor Hero for his support and help throughout the majority of this research. Working with Professor Hero is both a challenging and a rewarding experience. I admire his patience with his students, vast knowledge, and imagination.

This work has benefited from numerous discussions with the rest of my committee members. I would like to thank Professor Teneketzis, Professor Keener, and Professor Wakin for their support, input, and helpful comments. At the same time I wish to thank Raviv Raich with whom I worked closely on a major section of my thesis.

One of the best things that happened to me at Michigan is the exposure to a group of exceptional individuals whom I was lucky to have as colleagues in both Professors Fessler and Hero research groups. Specifically, I would like to thank Dan Ruan who was always there when a tough problem was waiting to be solved. Also, I'd like to mention Matt Jacobson, Somesh Srivastava, Kim Khalsa, Seyoung Chun, Mark Kliger, Ami Weisel, Kevin Carter, and Raghuram Rangarajan. I would like to thank my friends at Michigan: Ioannis Poulakakis, David Shuman, and Petar Momcilovic and elsewhere Evyatar, Harel, Tani, Oren, Rotem, Amir, Israel, Dikla, Ayelet, and

many others for their support and encouragement throughout the past years.

Finally, and most of all, I have to thank my wife Iris who was there for me to support and help through the harshest of times. I also wish to thank my parents, my brothers, and my sister for their distant support of my endeavors. Last but not least I'm grateful to Herbert and Barbara Goldberg for taking us into their extended family and creating a place we can call home away from home.

PREFACE

Every journey starts somewhere, and this one started back in March of 1996. After a few good years with the military and several months in South America, I landed in Florida and started a long motorcycle journey throughout the United States. Some 25,000 miles and seven months later, I was on my way back to Israel to start my undergrad education. At that point I knew it is only a matter of time until I will be back in this magnificent country. Of course, I had no way of knowing that I would be riding my motorcycle throughout four Michigan winters.

My military education taught me the principles of war, which to a large extent are suitable for more peaceful situations as well. Two of these principles led me to this current research: 1) Mass - Bring decisive force to bear at critical times and places. 2) Economy of Force - Allocate minimum essential combat power to secondary efforts. This research deals with efficient resource allocation schemes and captures my quest to create efficient systems.

On the morning of October 6, 2006, I was rushed to the ER at UM Hospital after an unpleasant encounter I had with a deer on the steering bar of my motorcycle at approximately 55 mph. Twelve hours and some 400 x-rays later (mainly due to head and chest CT) it turned out there was nothing wrong with me, I was shaken not stirred. Recovered from that unorthodox experience it hit me (along with the deer): there has got to be a better way to do these things. In the time since the accident I have developed methodology to improve the way we search for needles-in-a-haystack.

These methods focus search efforts onto unknown regions of interest, limiting the efforts wasted scanning benign locations. There are numerous applications for which this may apply, ranging from: air traffic control and missile launch detection to medical imaging and early detection of cancer tumors. It is my sincere hope that this work will find its way to the real world and help us reduce wasted efforts. Whether it is in shortening the time we wait through an airport security screening line or by reducing radiation exposure of healthy tissue, I hope that I made something good with my time at Michigan.

May 2008

Ann Arbor, MI.

Figure P.1: March 22^{nd}, 2008, I've almost made it home. Image taken on Michigan Ave. (US-12).

TABLE OF CONTENTS

LIST OF FIGURES

Figure

ABSTRACT

Efficient Resource Allocation Schemes for Search

by
Eran Bashan

Co-Chairs: Alfred O. Hero III and Jeffrey A. Fessler

This thesis concerns the problem of efficient resource allocation under constraints. In many applications a finite budget is used and allocating it efficiently can improve performance. In the context of medical imaging the constraint is exposure to ionizing radiation, e.g., computed tomography (CT). In radar and target tracking time spent searching a particular region before pointing the radar to another location or transmitted energy level may be limited. In airport security screening the constraint is screeners' time. This work addresses both static and dynamic resource allocation policies where the question is: How a budget should be allocated to maximize a certain performance criterion.

In addition, many of the above examples correspond to a needle-in-a-haystack scenario. The goal is to find a small number of details, namely 'targets', spread out in a far greater domain. The set of 'targets' is named a region of interest (ROI). For example, in airport security screening perhaps one in a hundred travelers carry prohibited item and maybe one in several millions is a terrorist or a real threat. Nevertheless, in most aforementioned applications the common resource allocation

policy is exhaustive: all possible locations are searched with equal effort allocation to spread sensitivity.

A novel framework to deal with the problem of efficient resource allocation is introduced. The framework consists of a cost function trading the proportion of efforts allocated to the ROI and to its complement. Optimal resource allocation policies minimizing the cost are derived. These policies result in superior estimation and detection performance compared to an exhaustive resource allocation policy. Moreover, minimizing the cost has a strong connection to minimizing both probability of error and the *Cramér-Rao* bound on estimation mean square error. Furthermore, it is shown that the allocation policies asymptotically converge to the omniscient allocation policy that knows the location of the ROI in advance. Finally, a multi-scale allocation policy suitable for scenarios where targets tend to cluster is introduced. For a sparse scenario exhibiting good contrast between targets and background this method achieves significant performance gain yet tremendously reduces the number of samples required compared to an exhaustive search.

CHAPTER I

Introduction

Consider the problem of efficiently using a given budget to perform a certain task. By a 'task' we generally mean the process of acquiring data and we refer to the acquisition 'cost' as budget. This setting is common to many daily applications where the question is how to best utilize the budget at hand. In the context of medical imaging the budget or constraint is the total amount of ionizing radiation a patient may safely endure, e.g. computed tomography (CT). In radar and target tracking we are limited by the duration of time we can spend searching a particular region before moving the radar beam to another location or by the energy level of the transmitted signal. In airport security screening we are limited by the total amount of screeners time at any given day. Time is also an issue in medical imaging modalities such as single photon electron CT (SPECT) or magnetic resonance imaging (MRI) where a long scan duration may result in either motion blur or a need to keep the patient inside the scanner for too long.

In addition, in many of the examples discussed above the scanned medium bears little interest for the actual application. What we are looking for is a small number of details, namely targets, spread out in a far greater domain. We call the collection of all targets in a given domain a region of interest (ROI). In medical imaging we

may look for a small tumor, maybe less than one cubic centimeter, placed somewhere inside the torso. For target tracking/detection applications the connection may be more obvious as we are looking for vehicles on the ground, airplanes in the air, or vessels in the sea. In a general screening process it is also more often than not the case where the proportion of 'targets' within the screened population is relatively small. For example, percentage of women exhibiting breast cancer is roughly 12% of the screened population. In airport security screening perhaps one in a hundred travelers forget a prohibited item in his belonging and maybe one in several million is a terrorist or a real threat. Nevertheless, in most of the applications mentioned above the common search/resource allocation scheme is exhaustive where all possible locations are searched with equal effort allocation to spread sensitivity. For example, the same energy level is used in each CT projection when we image the torso to detect lung cancer tumor.

A resource allocation policy may be either static or dynamic. A static allocation is one where we predetermine how to allocate efforts before any action is taken. In active radar this is equivalent to predefining the beam pattern and the (angular) scan speed and trajectory. This way, regardless of finding, the radar beam would keep scanning the domain in a fix pattern and with a fix resolution. Exhaustive resource allocation scheme is a special case among static schemes. Dynamic resource allocation scheme allows the allocation policy to change over time. If this change is data dependent, then we call such schemes *adaptive sensing*. In a sense, *adaptive sensing* schemes utilize past observation to modify the way some reserved budget is being deployed to acquire future observations. This thesis considers both static and dynamic resource allocation schemes. We start by designing a non uniform SPECT scanner system and move on to derive optimal adaptive allocation schemes.

The main contributions of this dissertation are:

i. We design a nonuniform SPECT system to minimize estimation error of a linear estimator subject to bandwidth constraints. Analytic expression of the expected estimation variance is given and reduction of up to 70% of estimation variance is achieved.

ii. We provide a novel framework for the problem of dynamic resource allocation. This framework utilizes a new cost function that accounts for effort distribution inside and outside the ROI.

iii. We show that minimizing our cost function is strongly connected to minimizing both error probability and estimation error.

iv. An optimal Adaptive Resource Allocation Policy, which we call $ARAP$, is derived for a two stage allocation procedure.

v. We show that $ARAP$ is asymptotically optimal in terms of our cost function.

vi. We derive a multi-scale version of $ARAP$, namely M-ARAP, utilizing a coarse scale for the first stage, then refining the data over a small part of the scanned domain at the second stage.

vii. We show that M-ARAP maintains most of the properties of $ARAP$, yet significantly reducing the total number of measurements used.

viii. A bound on the expected 'waste' due to multiscaling is provided.

We suggest that future work would connect our adaptive sensing methods to *compressive sensing* to yield a new sampling paradigm for sparse signals.

The rest of this thesis is organized as follows: We continue with a literature review of related work for the rest of this chapter. Chapter II propose a method for designing a nonuniform two-pinhole SPECT system. In Chapter III we formulate a novel framework for the problem of adaptive resource allocation and derive two stage optimal effort allocation schemes, namely *ARAP*. Chapter IV discusses several benefits in multiscaling and introduces M-ARAP. Finally, in Chapter V we conclude and point out to possible future work.

1.1 SPECT system design

The first application of the tracer principle was made in 1911 by Hevesy. However, the bridge to modern nuclear medicine and its emphasis on imaging awaited the development of imaging devices, which first appeared in the late 1940's. Cassen et al. developed the first planar gamma-ray scanning device. Their rectilinear scanner produced a pattern of dots representing the distribution of radiotraces within a patient body. In the early 1950's, Anger was the first to use pinhole collimation to increase resolution in small regions. The image was projected, through the pinhole, onto a scintillating screen with photographic film behind it. The overall system was highly inefficient and it required extremely long exposure times (principally due to losses in the film). These inefficiencies resulted in extremely high radiation doses to patients. By the end of the 1950's, Anger had replaced the photographic film with an array of photomultiplier tubes (PMT). This design becomes the basis for today's Anger camera. Kuhl and Edwards were the first to present tomographic images produced using the Anger Camera in 1963. By 1970's a series of innovations in rotating cameras led researchers to look for improved resolution in the reconstructed images. In 1978 Vogel et al. reported experiments with Anger cameras using a seven-pinhole

collimator, but distracting artifacts were present in the images due to limitations in the reconstruction methods available at that time. It was not until the 1990's that researchers were able to use multi-pinhole collimators in Emission Tomography (ET) scanners. Stationary PET scanners were pioneered by Robertson and Bozzo et al. by the 1960's and by Ter-Pogossian et al. in the early 1970's. Today ET scanners (both PET and SPECT) are of the most important medical imaging modalities, providing images that reveal subtle information about physiological processes in humans and animals.

Since the beginning, the reconstructed image resolution was a constant challenge for all system designers. In [50] Rogulski et al. showed that improvements in detector resolution can lead to both improved spatial resolution in the image and improved counting efficiency, through the design of multiple-pinhole coded-aperture system. Their group had since been dealing with feasibility issues and had build several such SPECT systems (see [43, 44, 59]), where optimizing parameters such as number of pinholes, pinholes geometry and their diameter is essential during the design process. In [59] they also report on first attempts of using a multi-resolution system, i.e., having pinholes with different diameters, where the image is reconstructed using OS-EM algorithm (Ordered Subsets Expectation Maximization). Unfortunately, no conclusive conclusions regarding this setup are given. Ivanovic et al. had also reported [27] experiments of a multiple pinhole scanner where they had optimize, among other system parameters, the number of pinholes their aperture size and their geometry. In [51] Schramm et al. describe another multiple pinhole SPECT scanner and reports improvements in the system resolution and sensitivity when compared to a single pinhole system. Meng et al., in [40], describes a study of a Compton scattering enhanced multiple pinhole imager. They show that scattered detected photons

information may be used in addition to photons detected after going through the pinhole collimator to improve the overall system performance. In [39] Meng and Clinthorne modified the *Uniform Cramér-Rao Bound* (UCRB), suggested in [26], and presented tradeoffs between some resolution constraint and the reconstructed image variance. They compare the bounds for various multiple pinhole systems and show that increasing the number of pinholes yield improved system performance. However, none of the aforementioned papers attempt to analyze a system with different pinhole diameters, and [59] is the only one that even considers such a design. Intuition leads to consider a system where a wide pinhole, yielding many counts and improves Signal to Noise Ratio (SNR), is combined with a narrow pinhole that provides good resolution but a noisy image, to result in a low-noise high-resolution reconstructed image. In Chapter II we seek to minimize the estimation variance subject to some constraint on the overall system resolution.

1.1.1 Resolution measures

Among researchers in the imaging science there is a convention that system performance is a task dependent measure. For example, the human vision system deals very well with "white noises". Human's vision automatically averages out zero mean "salt and pepper" type noises. Therefore, noisy images do not necessarily imply that the system producing them is not useful. On the other hand, image blur is very conspicuous for the human eye. Thus, we wish to look at sharp images with as little spatial bias as possible. As a result, quantifying imaging system performance is not a trivial task. In [15] den Dekker and van den Bos survey some common resolution measures. Starting from classical resolution criteria such as Rayleigh's just resolved criteria and Houston's Full Width Half Maximum (FWHM) criteria both considering

a noise free system to measure precision effects on system resolution, and compare the different measures. Evidentially they claim that, since noise free systems are not of practical interest, all resolution measures should consider SNR. Moreover, they show similarities between different resolution measures ranging from the image estimated variance to information and decision theory resolution measures, since they are all deeply affected by SNR. A different approach is taken in [57] where Wang and Li state 5 axioms that a good resolution measure should obey, then show that any such measure should be proportional to the standard deviation of the point spread function of the imaging system. There are several other resolution criteria in the literature that we are not going to mention here.

In [20] Fessler suggests another approach by minimizing the estimated image variance subject to a certain constraint on the FWHM of the system. Fessler analyzes a single pinhole system considering a specific estimation scheme, namely kernel-based indirect density estimator. For the suggested criteria Fessler shows that the optimal pinhole diameter should be proportional to the desired system FWHM. Moreover, for a specific case Fessler provides a close form solution for the problem showing that the ratio of the optimal pinhole diameter to the system FWHM is constant. In Chapter II we use that criteria to find the optimal pinhole diameters in a two pinhole system subject to some constraint on the system frequency response of the point spread function.

1.2 Search and dynamic resource allocation

The problem of dynamic resource allocation is connected to many different research fields. Although often called by different names and hidden under different frameworks, the concept of dynamic resource allocation is apparent in: adaptive

sensing, Markov decision problems, multi-armed bandit problems, sensor management, search, and multi-scale hypothesis testing. Previous work on adaptive sensing considered the problem of how to spatially distribute samples to recover an underlying signal. In Markov decision problems one is looking to find an action policy that maximizes a certain reward function depending on the different states and actions. Multi-Armed Bandit (MAB) problems are a class of sequential resource allocation problems concerned with allocating one or more resources among several alternative projects. In search theory the objective is often to find a search sequence (policy) that maximizes the probability of detecting a target hidden in one of many cells. Multi-scale hypothesis testing problem concerns with making a multi-hypothesis decision process more efficient by lumping parts of the hypothesis space together.

1.2.1 Adaptive sampling

Most of the previous work on adaptive sampling (AS), which sometimes appears in the literature as active learning or active sampling, has concentrated on estimating functions in noise. Castro et al. [13] present asymptotical analysis and shows that for piecewise constant functions adaptive sampling methods can capitalize on spatial properties of the function. By focusing samples to the estimated vicinity of the boundaries, adaptive sampling methods yields nearly optimal convergence rate, in terms of estimation mean square error (MSE). It is also shown that for spatially homogeneous functions adaptive sampling has no advantages over passive sampling. Nowak et al. [42], Castro et al. [12], and Willett et al. [58] consider different applications characterized by spatially inhomogeneous functions, for which adaptive sampling methods can be efficiently used. In [11], Castro et al. show that for certain classes of piecewise constant signals *compressed sensing* is as efficient as adaptive

sampling, in terms of the estimation error convergence rate. Work in the field of *Compressed Sensing* (CS) challenges the traditional signal processing sampling requirements. Recent results show that a relatively small number of random projections of a signal can contain most of its salient information. It follows that if a signal is compressible in some orthonormal basis, it can be accurately recovered from random projections even when they are contaminated with noise [9, 24]. Candes and Tao [9] introduce the Dantzig selector (DS) algorithm which solves an l_1-norm minimization problem to reconstruct a sparse signal (defined below) in $\mathbb{R}^Q$ from a very limited set of $N < Q$ noisy observations. Their algorithm converges to the true solution as long as the measurement operator obeys the uniform uncertainty principle (UUP). They provide an upper bound on the mean squared error (MSE) which, remarkably, is proportional up to a $C \log Q$ factor of the noise level σ^2. Haupt and Nowak present a similar result but their measurement operator is randomized rather than following the UUP [24]. Most of the previous work in both AS and CS is limited to inhomogeneous signals and is of limited usage for the problem considered in this thesis.

Sparse signals: A signal is considered sparse if its value is zero, or almost zero, in most places. Strong sparsity is defined when most of the signal elements must be exactly zero and is quantified by the fraction of nonzero elements. Weak sparsity is defined when most of the signal elements are very small and is quantified by the rate at which the sorted nonzero amplitudes decay. Although we consider homogenous signals, we assume that the support of the ROI is small as compared to the entire support of the signal. Therefore, we refer to such signals as sparse. Sparsity is used in a variety of applications: signal compression, reconstruction, approximation, source

separation and localization, and target tracking or detection [4,18,23,37,38,41,54,61]. Most of the related research considers post processing tasks. Matching pursuit [38] use a greedy algorithm to select base elements from the dictionary. Algorithms like FOCUSS [23] use sparsity to reconstruct a signal from limited samples. Nafie et al. [41] address the problem of subset selection. Wohlberg [61] provides reconstruction error bounds for several sparse signal bases. Sparse solutions using l_1 penalty are used in [37] to improve performance in direction-of-arrival estimation. Tropp lays theoretical foundations for convex relaxation techniques for sparse optimization problems [54]. Escoda et al. incorporate *a priori* knowledge of the signal structure to compensate for a potentially coherent dictionary [18]. An algorithm that adapts a dictionary to a given training set is given in [4]. In our work, we would like to utilize the sparsity during the data acquisition phase as a pre-processing task.

1.2.2 Sensor management

Sensor management is a very wide topic with many different applications and we refer the interested reader to [25]. However, in the context of this work we mention two questions discussed in the literature that are where to point a sensor and in what mode to operate a sensor for the next observation. Assume an agile array of sensors is used to scan a certain domain. At each time step, one chooses which grid point (cell) to search next and in what mode. Generally, a number of existing targets need to be tracked while new targets are being looked for. Kastella looks at such porblems under low SNR [31]. He introduces the discrimination gain based on the Kullback-Leibler information to quantify the usefulness of the next measurement. Using a myopic strategy, Kastella shows that pointing the sensor to the cell maximizing the discrimination gain decreases the probability of incorrectly detecting where a target

is. Kreucher et al. [33, 34] show that integrating the sensor management algorithm with the target tracking algorithm via the posterior joint multi-target probability density (JMPD), helps to predict which measurement would prove most informative in terms of increasing the information gain. Our approach differs as we consider selecting the sensor operation mode from a continuous rather than discrete set of modes. Krishnamurthy [35, 36] is interested in the problem of how to manage the sensor to keep track of multiple targets already acquired. He uses a multi-arm bandit formulation involving hidden Markov models to derive solutions to that problem. In [36], an optimal algorithm is formulated to track multiple targets. Since the optimal approach has prohibitive computational complexity, suboptimal methods are given and numerical examples are presented. In [35], the problem is reversed and a single target is observed from a collection of sensors. Again, approximate methods for the optimal solution are formulated due to its intractability.

Adaptive energy allocation is addressed in [47–49]. Rangarajan et al. consider the problem of adaptive amplitude design for estimating parameters of an unknown medium under average energy constraints (fix energy constraints in [47]). They treat an N time-steps design problem and provide an optimal solution for the case of $N = 2$ in terms of minimizing estimation MSE. However, they do not consider the parameter vector of interest to be sparse and as a result only minor gains are possible. Using our method we show asymptotic gains in MSE inversely proportional to the sparsity of the scanned domain.

Multi-Armed Bandit problems: In the classical MAB problem (see[1] [25] chapter 6) at each instant of time a single resource is allocated to one of many competing

[1]Most of the following paragraph is taken from the referred book and was written by Aditya Mahajan and Demosthenis Teneketzis.

projects. The project to which the resource is allocated can change its state; the remaining projects do not change state. In a variants of the MAB problem one or more resources are dynamically allocated among several projects; new projects may arrive; all projects may change state; delays may be incurred by the reallocation of resources, etc. In general, sequential resource allocation problems can be solved by dynamic programming. Dynamic programming, which is based on backwards induction, provides a powerful method for the solution of dynamic optimization problems, but suffers from the curse of dimensionality. The special structure of the classical MAB problem has led to the discovery of optimal index-type allocation policies that can be computed by forward induction (see [22] for more details), which is computationally less intensive than backward induction. Researchers have also discovered conditions under which forward induction leads to the discovery of optimal allocation policies for variants of the classical MAB. In the approach taken here we assume a different degree of freedom. The action we take affects the way we collect information regarding certain states, rather than causing a change in the states.

1.2.3 Search theory

The field of *Search Theory* considers the following problem: a single target is hidden in one of Q boxes. Each box is equipped with prior, detection, and false alarm probabilities. A desirable search policy maximizes the probability of correctly detecting the location of the target. For review of the problem and reference therein see [6]. From the earlier work of Kadane [30] on "whereabouts search" to a more recent work of Castanon [10] on "dynamic hypothesis testing", the question remains which cell to sample next in order to maximize the probability of detecting the location of the target. Castanon shows that a myopic strategy is optimal for certain

noise characteristics. Although search theory has generated much research for more than six decades, most of the work has concentrated on searching one box at a time. In our case, we relax this stringent restriction. Song and Teneketzis [52] generalize the framework to a search more than one cell at each step. They derive two condition under which a search policy is optimal for either a fixed horizon or for any horizon respectively. Most search theory literature considers independent measurements between neighboring cells and over time. This model enables an offline calculation of a compact static table listing the probability of detecting the target while searching a cell at a given time. In turns, the table is used to find an optimal search policy. Castanon extends this model and consider the case of dependent cells where the probability table is dynamically updated [10]. This is also the case in our work, although we introduce dependency between cells in a dual manner: over time and via a sparsity constraint.

To the best of our knowledge, Posner [46] was the first to consider searching more than one box at a time. He considers the problem of using a radar to locate a satellite lost in a region of the sky containing Q cells. His goal is to minimize the expected total search duration, and the idea is to search the cells where the satellite is most likely to be first. Assuming a uniform prior, the competing strategy exhaustively searches each cell for time t_1 with an expected search time of $t_1(Q+1)/2$. Posner suggests a preliminary search yielding a non-uniform likelihood function, followed by a search of all cells for a time t_1 in a descending likelihood order. For the preliminary search he allows to widen the radar beam and measure k cells for a time t in each measurement. Moreover, Posner allows to take as many preliminary searches as necessary. In his model, the detection probability increases in t and decreases in k. Posner's model assumes that the search is stopped as soon as the satellite has been

found. Therefore, sequentially searching the cells with the highest likelihood reduces the overall expected search time. Posner shows that the optimal solution minimizing the expected search time takes a single preliminary search, in which $k = 1$ and t is small (i.e., take a sneak peek at each cell), then uses the returns to sort the cells in a decreasing likelihood order and finally measure each cell again in the new order. By minimizing expected search time Posner imposes a 'soft' resource constraint on the total time used in the search process. In the approach taken in this work, we use the posterior distribution in a similar manner as the likelihood is used in [46], although we consider a different cost function. The Bayesian framework we use suffice to show optimality of our search policy.

1.2.4 Multi-scale hypothesis testing

A search problem can be interpreted as a multiple hypothesis testing problem where we know only one hypothesis out of many is true. A natural extension is a multiple hypothesis testing were more than one hypothesis is true. Dorfman [16] considers the problem of detection a defective members of large population. A simple way of approaching that problem is sampling all members of the population then testing each sample individually, which is equivalent to an exhaustive search. In large populations such approach is tedious, e.g. airport security, and may lead to additional detection errors due to mechanical or human imperfections. Dorfman considers the problem of weeding out all syphilitic men among a large population (inductees to the armed forces). Since the test used to detect the presence or absence of "syphilitic antigen" is very sensitive, Dorfman suggests to pool blood samples of different individuals together and test the pool rather than testing each sample individually. If the pool is tested positive (defective) then each one of the pool

constituents is tested separately. If the prevalence rate of the syphilitic antigen is low great savings can be achieved. Dorfman continue by finding the optimal pool size per a given prevalence rate. Sterrett [53] improves on Dorfman method by suggesting that individual samples from a defective pool would be retested only until one of them was found defective. Next the samples yet to be tested are pooled again and retested. If the prevalence rate is low there is a good chance that the new subgroup would be cleared and the testing can be stopped. Both Dorfman and Sterrett use a binomial model (B-model) for the underlying population distribution. Pfeifer and Enis [45] modify this model (M-model) by considering sampling from two distribution: one is composed of only zeros while the other contain only positive values. Therefore, each blood pool sampled results in a number representing the 'defectiveness' of that pool (if zero then the pool is cleared). Under the M-model it is possible to test a subgroup of the original defective group and still learn about the remaining untested members of the original group. Thus additional savings are possible compared to the original Dorfman procedure.

Frakt et al. [21] consider the problem of anomaly detection and localization from noisy tomographic data. In effort to reduce the problem of testing hypothesis over a space extremely large in cardinality, they propose a hierarchical framework that minimize computation and zooms in on the right hypothesis. The main difference between Frakt work and the previously mentioned papers is that Dorfman procedures are merely a sum of individual samples while Frakt et al. consider a general affine statistic of the samples. However, the latter requires first sampling the entire 'population' on a fine scale. Our goal is to finely sample the population only where it is needed and therefore is along the line of Dorfman's procedures.

Abdel-Samad and Tewfik [1–3] consider the problem of maximizing the probabil-

ity of correctly detecting a target hidden in M discrete cells given a total of $L < M$ observations. They suggest using a hierarchical search scheme, i.e. recursively dividing the M cells into m groups until each group contains a single cell, then use $l < L$ measurements at each step to decide on which group to focus next. Their concern is how to allocate the L measurements between the different levels of the hierarchical tree, where SNR is decreasing as the number of cells in each group increases. In [1] they present one offline solution and two online solution to allocate repeated measurements at each level of the tree, i.e. l is fixed. They conclude that the dynamic method they call binary look-ahead search performs best at high SNR. That method use a binary tree, $m = 2$, but at each stage consider all previous measurement to decide where to go next. In [2] they continue, analyzing an offline scheme, by establishing a lower bound on L, for a given error probability, as a function of SNR. In addition, they now let l_i, the number of measurement at each step, vary. Again they conclude that $m = 2$ is the best choice. Finally, in [3] they resort to a sequential multi hypothesis testing to provide online enhancement of measurement allocation. By using sequential hypothesis testing less measurement are needed, on average, to achieve the same probability of error that an offline batch processing algorithm requires. Hence, the remaining measurements are used to reduce the probability of error.

1.3 Applications

Different researchers considered different applications for which the above mentioned methods have been applied. We are primarily interested in static search problems. In static search the target location remains unchanged during the search. Slow or small changes compared to the search duration or the signal support size,

respectively, will be considered in future work. Relevant applications are medical imaging and early tumor detection, static target detection, and screening.

In medical imaging, we are specifically interested in early detection of breast cancer tumors. About one out of every eight women will experience breast cancer over a 90-year life span. If detected at an early stage, the patient stands an excellent recovery chance. However, detecting early stage tumors is a hard task, especially among younger women. Microwave imaging technology provides high contrast between normal breast fatty tissue and tumors and is a promising imaging modality for this application [7, 14, 19, 63, 64]. Bond et al. [7] suggest an exhaustive search policy for early detection of breast cancer. Although microwave energy is a non-ionizing radiation, it generates heat within the scanned tissue, which limits the energy level that can be safely used for a scan. Additionally, since this is an active radar system, the SNR depends on the amplitude of the transmitted signal. Hence, a search policy that would concentrate energy around region of interests should outperform an exhaustive search for a given total energy budget.

In Section 3.5 we provide an illustrative example of our method when applied to a synthetic aperture radar (SAR) imaging system. We show we can save time or improve performance in acquiring the content of an unknown ROI. Another possible extension of this work is to apply it to the airport security problem.

CHAPTER II

The Two-pinholes problem

2.1 Introduction

A common issue in medical imaging system design is how to optimize certain parameters to achieve a desired system performance. In [20] Fessler analyzes the tradeoff between spatial resolution and noise for a simple, single-pinhole, imaging system with a position sensitive photon-counting detector. This chapter explores the following problem: in a two-pinholes imaging system, should the pinhole sizes differ? We follow the work started in [20], and extended it to a two (independent) pinholes imaging system. We consider image recovery algorithms based on indirect density estimation methods using kernels that are based on apodized inverse filters. In [20] Fessler used this method to show that for a single pinhole system the optimal pinhole diameter ω, in terms of minimum estimation variance, is proportional to the Kernel function parameter β, which is also the system Full Width Half Maximum. Moreover, for a Gaussian profile pinhole, a closed form expressions for both the estimation variance and the proportionality constant was provided. We extend the expressions given in [20] to hold for an imaging system in which the estimate is formed by a convex sum of two images recovered from each pinhole independently. In addition we consider three types of constraints on the system frequency response

over which we seek the minimum variance. For a Gaussian shape pinhole with a Gaussian apodizing filter we provide a closed form solution for the problem. We show that under the first two constraints it is beneficial to design the two pinholes to have a different diameter. We further show that for the above setting the BW-RMS (bandwidth root mean square) constraint is not sufficient since it is possible to satisfy any such constraint with a system that yields zero variance, which basically means a blank recovered image. Finally, we perform a non-parametric variance minimization for a single pinhole system considering the same constraints and compare our results to the one derived for a two pinhole system. Ultimately, when designing medical imaging systems one would like the system images to be useful (have high-resolution) for physicians. Unfortunately, minimizing the estimated image variance subject to the mentioned constraints does not assure that desired property.

Since the beginning, the reconstructed image resolution was a constant challenge for all system designers. In [50] Rogulski et al. showed that improvements in detector resolution can lead to both improved spatial resolution in the image and improved counting efficiency, through the design of multiple-pinhole coded-aperture system. Their group had since been dealing with feasibility issues and had build several such SPECT systems (see [43, 44, 59]), where optimizing parameters such as number of pinholes, pinholes geometry and their diameter is essential during the design process. In [59] they also report on first attempts of using a multi-resolution system, i.e., having pinholes with different diameters, where the image is reconstructed using OS-EM algorithm (Ordered Subsets Expectation Maximization). Unfortunately, no conclusive conclusions regarding this setup are given. Ivanovic et al. had also reported [27] experiments of a multiple pinhole scanner where they had optimize, among other system parameters, the number of pinholes their aperture size and their geometry.

In [51] Schramm et al. describe another multiple pinhole SPECT scanner and reports improvements in the system resolution and sensitivity when compared to a single pinhole system. Meng et al., in [40], describes a study of a Compton scattering enhanced multiple pinhole imager. They show that scattered detected photons information may be used in addition to photons detected after going through the pinhole collimator to improve the overall system performance. In [39] Meng and Clinthorne modified the *Uniform Cramér-Rao Bound* (UCRB), suggested in [26], and presented tradeoffs between some resolution constraint and the reconstructed image variance. They compare the bounds for various multiple pinhole systems and show that increasing the number of pinholes yield improved system performance. However, none of the aforementioned papers attempt to analyze a system with different pinhole diameters, and [59] is the only one that even considers such a design. In this chapter we analyze a system combining data from two independent pinholes of different diameters. Intuition leads to consider a system where a wide pinhole, yielding many counts and improves Signal to Noise Ratio (SNR), is combined with a narrow pinhole that provides good resolution but a noisy image, to result in a low-noise high-resolution reconstructed image. Our objective is to minimize the estimation variance subject to some constraint on the overall system resolution. We show that it is not always beneficial to use identical set of pinholes, which is what current systems use.

Among researchers in the imaging science there is a convention that system performance is a task dependent measure. For example, the human vision system deals very well with "white noises". Humans vision automatically average out zero mean "salt and pepper" type noises. Therefore, noisy images do not necessarily imply that the system producing them is not useful. On the other hand, image blur is very

conspicuous for the human eye. Thus, we wish to look at sharp images with as little spatial bias as possible. As a result, quantifying imaging system performance is not a trivial task. In [15] den Dekker and van den Bos survey some common resolution measures. Starting from classical resolution criteria such as Rayleigh's just resolved criteria and Houston's Full Width Half Maximum (FWHM) criteria both considering a noise free system to measure precision effects on system resolution, and compare the different measures. Evidentially they claim that, since noise free systems are not of practical interest, all resolution measures should consider SNR. Moreover, they show similarities between different resolution measures ranging from the image estimated variance to information and decision theory resolution measures, since they are all deeply affected by SNR. A different approach is taken in [57] where Wang and Li state 5 axioms that a good resolution measure should obey, then show that any such measure should be proportional to the standard deviation of the point spread function of the imaging system. There are several other resolution criteria in the literature that we are not going to mention here.

In [20] Fessler suggests another approach by minimizing the estimated image variance subject to a certain constraint on the system's FWHM. Fessler analyzes a single pinhole system considering a specific estimation scheme, namely kernel-based indirect density estimator. For the suggested criteria Fessler shows that the optimal pinhole diameter should be proportional to the desired system FWHM. Moreover, for a specific case Fessler provides a close form solution for the problem showing that the ratio of the optimal pinhole diameter to the system FWHM is $1/\sqrt{2}$. In the rest of this chapter we use the criteria suggested in [20] to find the optimal pinhole diameters in a two pinhole system subject to some constraint on the system frequency response of the point spread function.

2.2 Summary of main results and notation borrowed from [20]

We first review a one pinhole system. Consider an emitting object with emission-rate density $\lambda(\underline{x})$ having unit emissions per unit time per unit volume. The emission rate density $\lambda(\underline{x})$ is defined over a subset Ω of $\mathbb{R}^d$, and we concentrate on $d = 2$ (planar imaging). We assume that the time-ordered sequence of emissions originated from statistically independent random spatial locations $\{\underline{X}_1, \underline{X}_2, \ldots\}$ drawn from a Poisson spatial point process having rate $\lambda(\underline{x})$. Let $s(\underline{x})$ denote the *sensitivity function* of the emission system, i.e., $s(\underline{x})$ is the probability that a photon emitted from a location $\underline{x}$ is detected somewhere by the system. When the system detects an emission, the probability density that the emission originated from a spatial location $\underline{x}$ is given by

$$f(\underline{x}) = \frac{\lambda(\underline{x})s(\underline{x})}{\int \lambda(\underline{x}')s(\underline{x}')d\underline{x}'} = \frac{\lambda(\underline{x})s(\underline{x})}{r}, \tag{2.1}$$

where $r \stackrel{\triangle}{=} \int \lambda(\underline{x})s(\underline{x})d\underline{x}$ is the total rate of detected events, with units of detected counts per unit time. Let $\{\underline{V}_n^{(i)}\}_{n=1}^N$ be the recorded position of some photon measured by a position sensitive device, where $i = 1, 2, ...Q$ represents the number of independent pinholes. We use kernel-based indirect density estimation to estimate the density, $f(\underline{x})$, of the unknown source $\underline{x}$. This method can be described, for a single pinhole system ($Q = 1$), as

$$\hat{f}(\underline{x}) = \frac{1}{N}\sum_{n=1}^{N} g_\beta(\underline{x}, \underline{V}_n^{(1)}), \tag{2.2}$$

where we suppose that the imaging system records a total of N events during a prespecified time t_0. By assumption, N is a Poisson random variable with mean

$$E\{N\} = t_0 \int \lambda(\underline{x}) s(\underline{x}) d\underline{x} = t_0 r. \tag{2.3}$$

In addition, we must have $\int g_\beta(\underline{x}, \cdot) d\underline{x} = 1$ to assure that $\hat{f}(\underline{x})$ integrate to one. Next we use (2.1) to estimate $\hat{\lambda}(\underline{x})$ as

$$\hat{\lambda}(\underline{x}) = \frac{\hat{f}(\underline{x})}{s(\underline{x})} r = \frac{\hat{f}(\underline{x})}{s(\underline{x})} \frac{E\{N\}}{t_0} = \frac{\hat{f}(\underline{x})}{s(\underline{x})} \frac{N}{t_0}, \tag{2.4}$$

where N is used as an estimate of $E\{N\}$. In general the recorded measurements $\{\underline{V}_n^{(i)}\}$ are indirectly related to the emitted photons $\{\underline{X}_n\}$ through some conditional pdf $f(\underline{v}|\underline{x})$. This pdf includes both the pinhole collimator response function as well as the detector response function. We consider a shift-invariant system[1], i.e., $f(\underline{v}|\underline{x}) = h(\underline{v} - \underline{x})$. Since $f(\underline{v}|\underline{x})$ is a conditional pdf in $\underline{v}$, it has to integrate to one. In addition we assume that the kernel function is also shift-invariant, i.e., $g_\beta(\underline{x} - \underline{v})$. The design problem is to choose the pinhole diameter ω, where the pinhole response function is defined by $h_0(\underline{x}) = \frac{1}{\omega^2} t(\underline{x}/\omega)$, and $t(\underline{x})$ is a transmissivity function, normalized in such a way that $\int t(\underline{x}) d\underline{x} = 1$. The Fourier transform of the pinhole response function is $H_0(\underline{\nu}) = T(\omega \underline{\nu})$. Define the apodized inverse filter

$$G_\beta(\underline{\nu}) \triangleq \frac{A(\beta \underline{\nu})}{H_0(\underline{\nu})}, \tag{2.5}$$

where $A(\beta \underline{\nu})$ is a user-chosen apodizing function, which is also the overall PSF$(\underline{\nu})$ for a single pinhole. We further simplify the problem by assuming that the sensitivity function is space invariant, i.e. $s(\underline{x}) = s_0$, where s_0 depends on the pinhole diameter

[1]For example a scanner system where the emitting body is being scanned with pinhole detectors.

ω. Therefore, in the spatial domain we have the following results, the systems overall point spread function is given by

$$\text{psf}(\underline{x},\underline{x}') = \int g_\beta(\underline{x}-\underline{v})h(\underline{v}-\underline{x}')d\underline{v} = \tag{2.6}$$

$$\int g_\beta(\underline{x}-\underline{x}'-\underline{x}'')h(\underline{x}'')d\underline{x}'' = (g_\beta * h)(\underline{x}-\underline{x}'),$$

where $\underline{x}'' = \underline{v} - \underline{x}'$. The estimator mean is given by

$$E\{\hat{\lambda}(\underline{x})\} = \mu(\underline{x}) = \int \text{psf}(\underline{x},\underline{x}')\lambda(\underline{x}')d\underline{x}' = (g_\beta * h * \lambda)(\underline{x}), \tag{2.7}$$

and the estimator variance is

$$\sigma^2(\underline{x}) = \frac{1}{t_0 s_0}(g_\beta^2 * h * \lambda)(\underline{x}). \tag{2.8}$$

Finally, Fessler shows (see [20] pp-249) that for a single pinhole with a Gaussian profile, assuming $s_0 = \left(\frac{\omega}{\kappa}\right)^2$, and if we choose a Gaussian apodizing function $A(\beta\underline{\nu}) = e^{-\pi(\rho/\kappa)^2}$, where $\kappa = 2\sqrt{\frac{\ln 2}{\pi}}$ is a constant that depends on the pinhole profile, the estimation variance is approximately

$$\sigma^2(\underline{x}) \cong \tilde{\lambda}(\underline{x})\frac{\kappa^{2d}}{2^{d/2}t_0}(\beta^2\omega^2 - \omega^4)^{-d/2} = \frac{c(\underline{x})}{\beta^2\omega^2 - \omega^4}, \tag{2.9}$$

where $\tilde{\lambda}(\underline{x}) = h * \lambda(\underline{x})$, $d = 2$, and $c(\underline{x}) = \frac{\tilde{\lambda}(\underline{x})\kappa^4}{2t_0}$ is a function of all the nuisance parameters and the underlying source density. Equation (2.9) holds as long as both the pinhole width ω and the kernel width β are relatively small compared with the variations in $\lambda(\underline{x})$. Hence, we must have $\omega \leqslant \omega_{\max}$ and $\beta \leqslant \beta_{\max}$, for some constants $(\omega_{\max}, \beta_{\max})$.

2.3 A two pinhole system

Consider a two pinhole system, where two independent pinhole systems scan the emitting object and their measurements are jointly used to estimate $\lambda(\underline{x})$. Specifically, we use $\left\{\underline{V}_n^{(1)}\right\}_{n=1}^{N_1}$ and $\left\{\underline{V}_n^{(2)}\right\}_{n=1}^{N_1}$ to estimate $\hat{\lambda}(\underline{x})$ through $\hat{f}(\underline{x})$, where

$$\hat{f}(\underline{x}) = \frac{\alpha}{N_1}\sum_{n=1}^{N_1} g_{\beta_1}(\underline{x}, \underline{V}_n^{(1)}) + \frac{(1-\alpha)}{N_2}\sum_{n=1}^{N_2} g_{\beta_2}(\underline{x}, \underline{V}_n^{(2)}). \tag{2.10}$$

N_1, N_2 are each system total number of detected events during the same measurement period t_0, g_{β_1}, g_{β_2} are two estimation kernels, and $\alpha \in (0,1)$ is a convex sum parameter. It can be easily shown that for the shift invariant case, the two pinhole equivalent of (2.7) and (2.8) are

$$E\{\hat{\lambda}(\underline{x})\} = \mu(\underline{x}) = \left[(\alpha g_{\beta_1} * h_1 + (1-\alpha) g_{\beta_2} * h_2) * \lambda\right](\underline{x}), \tag{2.11}$$

$$\sigma^2(\underline{x}) = \left[\left(\frac{\alpha^2}{t_0 s_1} g_{\beta_1}^2 * h_1 + \frac{(1-\alpha)^2}{t_0 s_2} g_{\beta_2}^2 * h_2\right) * \lambda\right](\underline{x}). \tag{2.12}$$

If we consider a Gaussian pinhole profile, and a Gaussian apodizing functions we may use (2.9) to formulate the following problem: find $\alpha, \beta_1, \beta_2, \omega_1, \omega_2$, with $\omega_i \leqslant \omega_{\max}$, $\beta_i \leqslant \beta_{\max}$, minimizing

$$\sigma^2(\underline{x}) \cong \frac{c_1(\underline{x})\alpha^2}{\beta_1^2\omega_1^2 - \omega_1^4} + \frac{c_2(\underline{x})(1-\alpha)^2}{\beta_2^2\omega_2^2 - \omega_2^4}, \tag{2.13}$$

where, $c_i(\underline{x}) = \frac{\tilde{\lambda}_i(\underline{x})\kappa^4}{2t_0}$, and $\tilde{\lambda}_i(\underline{x}) = h_i * \lambda(\underline{x})$, $i = 1, 2$, subject to some constraint on the PSF $H(\underline{\nu})$, given by

$$H(\underline{\nu}) = H(||\underline{\nu}||) = \alpha e^{-\pi\left(\frac{\beta_1}{\kappa}||\underline{\nu}||\right)^2} + (1-\alpha)e^{-\pi\left(\frac{\beta_2}{\kappa}||\underline{\nu}||\right)^2}. \tag{2.14}$$

Note that in [20], β represents the Full Width Half Maximum (FWHM) of the system. Therefore, Fessler looks for the optimal pinhole diameter that minimizes variance subject to a given β. In our case, the FWHM is a function of β_i and α, which are part of the optimization space. Hence we need different constraints on the system PSF.

Next we note that by taking partial derivatives of (2.13) w.r.t. ω_1 and ω_2 and setting them to zero we find, as in the single pinhole case, that the optimal diameters are proportional to the kernel parameters β_1, β_2 respectively with the same ratio

$$\omega_{i_{\min}} = \frac{\beta_i}{\sqrt{2}}, \qquad i = 1, 2. \tag{2.15}$$

Hence, the optimal pinhole diameters have a one-to-one mapping to the optimal kernel parameters. Therefore, (2.13) simplifies to

$$\sigma^2(\underline{x}) \cong \frac{4c_1(\underline{x})\alpha^2}{\beta_1^4} + \frac{4c_2(\underline{x})(1-\alpha)^2}{\beta_2^4}, \tag{2.16}$$

and the optimization space is reduced. Furthermore, if we assume that we examine the problem over some small region of the image where the underlying source is fairly uniform, then since both h_i's integrate to one we have $c_1(\underline{x}) \cong c_2(\underline{x}) \equiv c(\underline{x})$, which yields

$$\sigma^2(\underline{x}) \approx 4c(\underline{x}) \left[\frac{\alpha^2}{\beta_1^4} + \frac{(1-\alpha)^2}{\beta_2^4} \right] \triangleq \sigma_0^2(\underline{x}). \tag{2.17}$$

2.3.1 Resolution constraints

Note that (2.17) is inversely proportional to β_1^4, β_2^4 and therefore inversely proportional to the pinhole diameters to the power of four. Hence infinite diameter

pinholes[2] would result in zero estimation variance. However, this trivial solution is in essence a 'DC' like filter where only the mean of the underlying density is being estimated. To avoid the trivial solution we consider constrained optimization of (2.17), imposing some additional "resolution" constraints on (2.14). By constraining $H(\underline{\nu})$ we force a certain bandwidth (BW) on the system. Assuming that blur is sometime caused by insufficient BW limiting the system ability to preserve edges of the underlying image, we hope that the BW constraint would translate to good resolution performances. Note that $H(\underline{0}) = 1$ and $H(\underline{\nu})$ is a decreasing function of $||\underline{\nu}||$. Because imaging science field lacks a canonical definition for "resolution", we consider three different types of constraints on the system frequency response (2.14). The first one is a hard frequency constraint

$$H(\underline{\nu}_0) \geqslant \epsilon, \tag{2.18}$$

for some $\epsilon \in [0, 1]$ and some $\underline{\nu}_0$. Since (2.14) is monotonically decreasing in $\underline{\nu}$, (2.18) prevents the optimal solution from converging to a 'DC' filter. The second constraint is the Bandwidth Root Mean Square (BW-RMS) measure suggested in [57], defined as

$$\sqrt{\frac{\int_{\mathbb{R}^2} ||\underline{\nu}||^2 |H(\underline{\nu})|^2 d\underline{\nu}}{\int_{\mathbb{R}^2} |H(\underline{\nu})|^2 d\underline{\nu}}} \geqslant \eta. \tag{2.19}$$

Under statistical interpretation, (2.19) can be thought off as the standard deviation of the frequency component $\underline{\nu}$ having the following distribution

$$\frac{|H(\underline{\nu})|^2}{\int_{\mathbb{R}^2} |H(\underline{\nu})|^2 d\underline{\nu}}. \tag{2.20}$$

By forcing the "frequency variance" to be larger than some η, we want to guarantee that the optimal frequency response would not degenerate into the trivial solution.

[2] This is hypothetically speaking only, as something with an infinite diameter can hardly be called a pinhole.

The third constraint we use is a variation of the second one. We force the area beneath the square of the magnitude response to be greater than some constant, namely

$$\int_{\mathbb{R}^2} |H(\underline{\nu})|^2 d\underline{\nu} \geqslant \gamma. \tag{2.21}$$

Since in the Gaussian case $H_0(0) = 1$ the latter is a reasonable bandwidth constraint. Note that in (2.18) the constraint has no units, in (2.19) the constraint have units of inverse distance squared, and in (2.21) the units are inverse distance to the power of four.

2.3.2 Single pinhole

We first apply all three constraints to the single pinhole case, to serve as a reference for the results that follows in the two pinhole case. Our objective is to minimize (2.9) subject to the constraints (2.18),(2.19) and (2.21). First we note that in the aforementioned Gaussian case

$$\int_{\mathbb{R}^2} |H(\underline{\nu})|^2 d\underline{\nu} = \int_0^{2\pi} \int_0^{\infty} \rho H^2(\rho) d\phi d\rho \tag{2.22}$$

and that

$$\int_0^{2\pi} \int_0^{\infty} \rho e^{-\gamma\rho^2} d\phi d\rho = \frac{\pi}{\gamma}. \tag{2.23}$$

In addition

$$\int_{\mathbb{R}^2} ||\underline{\nu}||^2 |H(\underline{\nu})|^2 d\underline{\nu} = \int_0^{2\pi} \int_0^{\infty} \rho^3 H^2(\rho) d\phi d\rho, \tag{2.24}$$

and

$$\int_0^{2\pi} \int_0^{\infty} \rho^3 e^{-\gamma\rho^2} d\phi d\rho = \frac{\pi}{\gamma^2}. \tag{2.25}$$

Hence, our problem can be formulated as find $\beta_{\min}$

$$\beta_{\min} = \arg\min_{\beta>0} \frac{4c(\underline{x})}{\beta^4}, \tag{2.26}$$

such that (s.t.) either one of (2.27)-(2.29) holds

$$e^{-\pi\left(\frac{\beta}{\kappa}\rho_o\right)^2} = \epsilon, \tag{2.27}$$

$$\frac{\kappa}{\beta\sqrt{2\pi}} \geqslant \eta, \tag{2.28}$$

$$\frac{\kappa^2}{2\beta^2} \geqslant \gamma. \tag{2.29}$$

The solution, for each of the three constraints, can be found easily by solving (2.27) and by taking the maximal β allowed by constraints (2.28) and (2.29). These yields

$$\sigma_0^2(\underline{x}) = \frac{4c(\underline{x})}{\kappa^4}\frac{\pi^2\rho_0^4}{\ln^2\epsilon}, \tag{2.30}$$

$$\sigma_0^2(\underline{x}) = \frac{4c(\underline{x})}{\kappa^4}4\pi^2\eta^4, \tag{2.31}$$

and

$$\sigma_0^2(\underline{x}) = \frac{4c(\underline{x})}{\kappa^4}4\gamma^2 \tag{2.32}$$

corresponding to (2.27)-(2.29) respectively. From the last three equations, one can see that by increasing each constraint, i.e. let $\epsilon \to 1$ or $\eta, \gamma \to \infty$ the estimation variance increases. This is a desired behavior when performing constraint optimization, as it shows the conflict between the cost function and the different constraints. Moreover, it shows that an infinite bandwidth system would have infinite estimation variance.

2.3.3 Two independent pinholes

Hard Frequency constraint

Starting with the first constraint we want to minimize (2.17) s.t.

$$H(\rho_0) = \alpha e^{-\pi\left(\frac{\beta_1}{\kappa}\rho_0\right)^2} + (1-\alpha)e^{-\pi\left(\frac{\beta_2}{\kappa}\rho_0\right)^2} = \epsilon. \tag{2.33}$$

Taking derivative of (2.17) with respect to (w.r.t.) α and setting it equal to zero yields

$$\alpha_0 = \frac{\beta_1^4}{\beta_1^4 + \beta_2^4}. \tag{2.34}$$

Naturally, if $\beta_1 = \beta_2 \equiv \beta$, we have $\alpha_0 = 1/2$, in which case (2.33) is only a function of $\beta_2 = \beta$, and we have $\beta_0 = \frac{\kappa}{\rho_0}\sqrt{\frac{\ln \epsilon^{-1}}{\pi}}$. Plugging everything back into (2.17) yields

$$\sigma_0^2(\underline{x}) = \frac{2c(\underline{x})}{\kappa^4}\frac{\pi^2\rho_0^4}{\ln^2 \epsilon}, \tag{2.35}$$

which is, as expected, half of the variance expression in (2.30). However, if $\beta_1 \neq \beta_2$ we may solve (2.33) for α and get

$$\alpha_0 = \frac{\epsilon - e^{-\pi\left(\frac{\beta_2}{\kappa}\rho_0\right)^2}}{e^{-\pi\left(\frac{\beta_1}{\kappa}\rho_0\right)^2} - e^{-\pi\left(\frac{\beta_2}{\kappa}\rho_0\right)^2}}. \tag{2.36}$$

Due to the symmetry of the problem in β_1 and β_2 we may assume without loss of generality (wlog) that $\beta_2 > \beta_1$, then, since $\alpha \in (0,1)$ we have

$$0 \leqslant \epsilon - e^{-\pi\left(\frac{\beta_2}{\kappa}\rho_0\right)^2} \leqslant e^{-\pi\left(\frac{\beta_1}{\kappa}\rho_0\right)^2} - e^{-\pi\left(\frac{\beta_2}{\kappa}\rho_0\right)^2}, \tag{2.37}$$

After some simple manipulations these yield

$$\beta_{\max} \geqslant \beta_2 > \frac{\kappa}{\rho_0}\sqrt{\frac{\ln \epsilon^{-1}}{\pi}} > \beta_1. \tag{2.38}$$

Plugging (2.36) into (2.17) we get

(2.39)

$$\sigma_0^2(\underline{x}) = \frac{4c(\underline{x})}{\left(e^{-\pi\left(\frac{\beta_1}{\kappa}\rho_o\right)^2} - e^{-\pi\left(\frac{\beta_2}{\kappa}\rho_o\right)^2}\right)^2}\left[\frac{\left(\epsilon - e^{-\pi\left(\frac{\beta_2}{\kappa}\rho_o\right)^2}\right)^2}{\beta_1^4} + \frac{\left(e^{-\pi\left(\frac{\beta_1}{\kappa}\rho_o\right)^2} - \epsilon\right)^2}{\beta_2^4}\right].$$

To further analyze (2.39) fix arbitrary ρ_o, then since $\beta_2 > \beta_1$, we note that if $\beta_2 \gg \beta_1$ (2.39) is a function dominated by β_1, given by

$$\sigma_{0,\beta_1}^2(\underline{x}) = \frac{4c(\underline{x})}{e^{-2\pi\left(\frac{\beta_1}{\kappa}\rho_o\right)^2}}\frac{\epsilon^2}{\beta_1^4}, \tag{2.40}$$

where for the simplicity of analysis we assumed $\beta_2 \to \infty$. To minimize (2.40) we maximize its denominator, which can be found by taking derivative w.r.t. β_1. The solution

$$\beta_{1_{\min}} = \frac{\kappa}{\rho_o\sqrt{\pi}}, \tag{2.41}$$

is surprisingly independent of ϵ. Nevertheless, for a feasible solution we must have $\ln\epsilon^{-1} > 1$. Substituting (2.41) into (2.40) yields

$$\sigma_{0,\beta_{1_{\min}}}^2(\underline{x}) = \frac{2c(\underline{x})}{\kappa^4}\frac{\pi^2\rho_0^4 2\epsilon^2}{e^{-2}}. \tag{2.42}$$

To know if (2.42) is better than (2.35) we need to find when $2e^2\epsilon^2\ln^2\epsilon < 1$, which we may solve numerically. In figure 2.1 we numerically evaluate (2.42) and (2.35) versus ϵ to identify areas where one design is preferable to the other. For completion, a single pinhole system variance was added to the plot. One can see that as long as $\epsilon \in (0, 1/8)$, (2.42) is less than (2.35). In which case it is better to design a system with two pinholes of different diameter, where the larger pinhole diameter approaches

infinity and the smaller can be derived from (2.41) and (2.15). Substituting these results into (2.36) yields

$$\alpha_0 = e\epsilon. \tag{2.43}$$

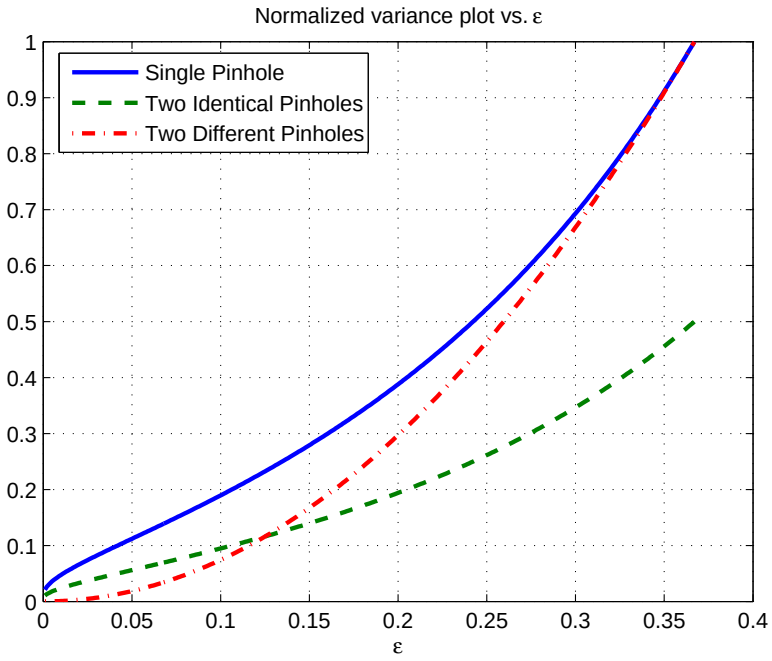

Figure 2.1: Variance comparison between a single pinhole system (blue, solid) to a two identical pinholes system (green, dashed) and a two different pinholes system (red, dashed-dot).

BW-RMS constraint

Thus far we have established a first case where it is beneficial to design two independent pinholes of different width. Moreover, we note that our result suggested that we let one of the two pinholes to become infinitely wide, making it a 'DC' spatial filter. Our next objective is to minimize (2.17) s.t. the BW-RMS constraint, given in (2.19), holds. Using (2.22)-(2.25), we note that (2.19) can be solved analytically

and has the following form

$$\frac{\int_{\mathbb{R}^2} \underline{\nu}^2 |H(\underline{\nu})|^2 d\underline{\nu}}{\int_{\mathbb{R}^2} |H(\underline{\nu})|^2 d\underline{\nu}} = \tag{2.44}$$

$$\frac{\int_0^{2\pi}\int_0^{\infty} \rho^3 [\alpha^2 e^{-\pi \frac{\beta_1^2}{\kappa^2}\rho^2} + 2\alpha(1-\alpha) e^{-\frac{\pi}{\kappa^2}(\beta_1^2+\beta_2^2)\rho^2} + (1-\alpha)^2 e^{-\pi \frac{\beta_2^2}{\kappa^2}\rho^2}] d\phi d\rho}{\int_0^{2\pi}\int_0^{\infty} \rho [\alpha^2 e^{-\pi \frac{\beta_1^2}{\kappa^2}\rho^2} + 2\alpha(1-\alpha) e^{-\frac{\pi}{\kappa^2}(\beta_1^2+\beta_2^2)\rho^2} + (1-\alpha)^2 e^{-\pi \frac{\beta_2^2}{\kappa^2}\rho^2}] d\phi d\rho} =$$

$$\frac{\kappa^2}{\pi} \left[\frac{\frac{\alpha^2}{4\beta_1^4} + \frac{2\alpha(1-\alpha)}{\left(\beta_1^2+\beta_2^2\right)^2} + \frac{(1-\alpha)^2}{4\beta_2^4}}{\frac{\alpha^2}{2\beta_1^2} + \frac{2\alpha(1-\alpha)}{\beta_1^2+\beta_2^2} + \frac{(1-\alpha)^2}{2\beta_2^2}} \right] \geqslant \eta^2 \tag{2.45}$$

Numerical minimization of (2.17) s.t. (2.45) holds, suggest that it is possible to choose α, β_1 and β_2 that satisfy the constraint yet yield variance approaching zero. To demonstrate a limiting case behavior we set $\alpha = \beta_1^{2.5}$, $\beta_2 = \beta_1^{-2}$ and let β_1 approach zero. This setting simplifies (2.45) as follows

$$\xi(\beta_1) = \frac{\kappa^2}{\pi} \left[\frac{\frac{\beta_1}{4} + 2\beta_1^{10.5} + \frac{\beta_1^8}{4}}{\frac{\beta_1^3}{2} + 2\beta_1^{6.5} + \frac{\beta_1^4}{2}} \right] \geqslant \eta^2, \tag{2.46}$$

and in the limit

$$\lim_{\beta_1 \to 0} \xi(\beta_1) = \frac{\kappa^2}{2\pi\beta_1^2} \to \infty. \tag{2.47}$$

On the other hand, along that path (2.17) has the following form

$$\sigma^2_{0,\beta_1}(\underline{x}) = 4c(\underline{x}) \left[\beta_1 + \beta_1^8\right], \tag{2.48}$$

and

$$\beta_1 \to 0 \qquad \Rightarrow \qquad \sigma^2_{0,\beta_1}(\underline{x}) \to 0. \tag{2.49}$$

This suggests that for our second constraint (BW-RMS), it is always possible to achieve minimal variance yet satisfying the constraint if we pick one very small pinhole and one very large pinhole. Note that, along that trajectory, the system frequency response (2.14) has the following form (where $\beta_1 \to 0$)

$$H(\rho) = \beta_1^{2.5} e^{-\pi\left(\frac{\beta_1 \rho}{\kappa}\right)^2} + e^{-\pi\left(\frac{\rho}{\kappa\beta_1^2}\right)^2} \cong \begin{cases} 1, & \rho = 0 \\ \beta_1^{2.5}, & \rho > 0 \end{cases}, \tag{2.50}$$

so once again, most of our estimation comes from the infinitely wide pinhole.

Area constraint

The third constraint can be extracted from (2.45) and is

$$\kappa^2 \left[\frac{\alpha^2}{2\beta_1^2} + \frac{2\alpha(1-\alpha)}{\beta_1^2 + \beta_2^2} + \frac{(1-\alpha)^2}{2\beta_2^2} \right] \geqslant \gamma. \tag{2.51}$$

Rearranging (2.51) as a polynomial in α we get

$$\alpha^2 \frac{\left(\beta_1^2 - \beta_2^2\right)^2}{2\beta_1^2\beta_2^2(\beta_1^2 + \beta_2^2)} + \alpha \frac{\beta_2^2 - \beta_1^2}{\beta_2^2(\beta_1^2 + \beta_2^2)} + \frac{1}{2\beta_2^2} \geqslant \frac{\gamma}{\kappa^2}. \tag{2.52}$$

Although we cannot solve analytically this constraint optimization problem, we can show that it has a minimum at $\alpha = \frac{1}{2}$ and $\beta_1 = \beta_2$. Consider the following Lagrangian

(2.53)

$$\psi(\underline{x}) = 4c(\underline{x}) \left[\frac{\alpha^2}{\beta_1^4} + \frac{(1-\alpha)^2}{\beta_2^4} \right] + \phi \left[\alpha^2 \frac{\left(\beta_1^2 - \beta_2^2\right)^2}{2\beta_1^2\beta_2^2(\beta_1^2 + \beta_2^2)} + \alpha \frac{\beta_2^2 - \beta_1^2}{\beta_2^2(\beta_1^2 + \beta_2^2)} + \frac{1}{2\beta_2^2} - \frac{\gamma}{\kappa^2} \right],$$

where ϕ is a Lagrange multiplier. Taking its derivative w.r.t. α we get,

$$\frac{\partial \psi(\underline{x})}{\partial \alpha} = 4c(\underline{x}) \left[\frac{2\alpha}{\beta_1^4} - \frac{2(1-\alpha)}{\beta_2^4} \right] + \phi \left[\alpha \frac{\left(\beta_1^2 - \beta_2^2\right)^2}{\beta_1^2\beta_2^2(\beta_1^2 + \beta_2^2)} + \frac{\beta_2^2 - \beta_1^2}{\beta_2^2(\beta_1^2 + \beta_2^2)} \right]. \tag{2.54}$$

Recall from (2.34), that $\beta_1 = \beta_2 \Rightarrow \alpha = \frac{1}{2}$. Hence, we note that $\beta_1 = \beta_2$ is a stationary point of (2.54). Taking the second derivative we verify that it is a minimum of $\psi(\underline{x})$, since $\beta_1 = \beta_2$ yields,

$$\frac{\partial^2 \psi(\underline{x})}{\partial \alpha^2} = 8c(\underline{x}) \left[\frac{1}{\beta_1^4} + \frac{1}{\beta_2^4} \right] + \phi \frac{\left(\beta_1^2 - \beta_2^2\right)^2}{\beta_1^2 \beta_2^2 (\beta_1^2 + \beta_2^2)} = 8c(\underline{x}) \left[\frac{1}{\beta_1^4} + \frac{1}{\beta_2^4} \right] > 0. \tag{2.55}$$

Note that we can get an analytical solution for this specific case, since under those assumptions, (2.51) simplifies to

$$\frac{1}{2\beta_2} \geqslant \frac{\gamma}{\kappa^2}, \tag{2.56}$$

which results in

$$\beta_{\min} = \frac{\kappa}{\sqrt{2\gamma}}. \tag{2.57}$$

Therefore, (2.17) has the following form

$$\sigma_0^2(\underline{x}) = \frac{4c(\underline{x})}{\kappa^4} 2\gamma^2, \tag{2.58}$$

which, as expected is half of (2.32). Numerical evaluation of that constraint optimization problem shows that (2.57) is a global minimum.

2.4 Simulation results

Although we had allowed, for the sake of analysis, one of the pinholes diameter (as well as the reconstruction kernel width) to approach infinity, we recall that our variance approximation is only valid when both parameters are bounded. To verify that our previous results holds for real systems, we consider a system where one pinhole is larger than the other, yet reasonably bounded. We choose a target resolution

of 2[mm] and therefore designed the system with two identical pinholes to have a FWHM of 2[mm]. We then designed a second system with different pinholes diameters that match the same frequency response constraint as the first system but has lower variance. Treating each pinhole independently, we note that β_i is the FWHM of each subsystem, and if $\beta_1 = \beta_2 \equiv \beta$ than it is also the FWHM of the overall system. Since our reference system has a FWHM of 2[mm], we restrict $\beta_{\max}$ to be not much greater than 2[mm], and keep $\beta_1 < \beta_2 \leqslant \beta_{\max}$. We used the Derenzo phantom shown in figure 2.2 to test our system, where the two smallest discs diameters were chosen to "challenge" the systems. For brevity, from hereon we name the equivalent diameter pinholes the "standard" system, and the different diameter pinholes the "new" system. Results are presented and summarized below. We used $N = 10000$ as the mean of the Poisson random variables N_i's to compute the total number of counts in the standard system. The mean number of counts of the new system was scaled appropriately. Each system imaged the Derenzo phantom 500 times and simulation statistics were caculated. Figure 2.3 shows four images reconstruction of the Derenzo phantom using each of the systems described in details below. The upper left image was acquired by a system with two identical diameter pinhole, set to achieve a FWHM of 2[mm]. The lower left image was reconstructed by a system that satisfy the BW-RMS constraint. Both images on the right column were derived from systems satisfying the hard frequency constraint. In the upper one of the two a standard reconstruction kernel was used, while in the lower one a jinc masked kernel with sign alternating tail was used. See the appendix for a full size images of the different reconstructions.

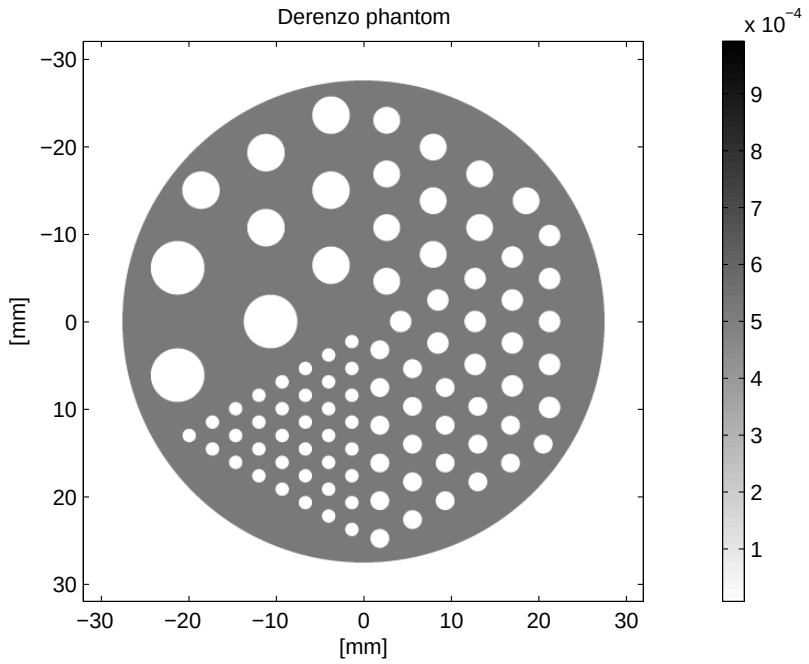

Figure 2.2: A cold spot Derenzo phantom used as an input to our systems. The discs radii are [1.53 2.15 2.45 3.06 4.28 6.12] [mm].

2.4.1 Hard frequency constraint

As previously shown, the new system outperforms (i.e., had a lower estimation variance) the standard system when $\epsilon \in (0, \frac{1}{8})$. We arbitrarily set $\epsilon = 0.05$, where our analysis predict a 67% improvement of the variance. We let $\beta_{\max} = 3$, which results in $\beta_1 = 1.257$, $\beta_2 = 3$, $\alpha = 0.16$ and $\rho_o = 0.4587$. Numerical evaluation of (2.16) suggests a 35% reduction in the variance versus 67% gain if we let $\beta_2 \to \infty$. Figure 2.4 shows the two systems frequency response functions. We see that although both functions satisfy the constraint they do so in a different manner. The new system frequency response decays slower than the standard, as it is a combination of two Gaussians. Figures 2.5 and 2.6 shows simulation results of this setup, where we concentrate on the estimators mean and variance. We clearly see that new system outperform the standard system by achieving the predicted variance gain as seen in figures 2.5 and 2.6. Moreover, figure 2.6 also validates the accuracy

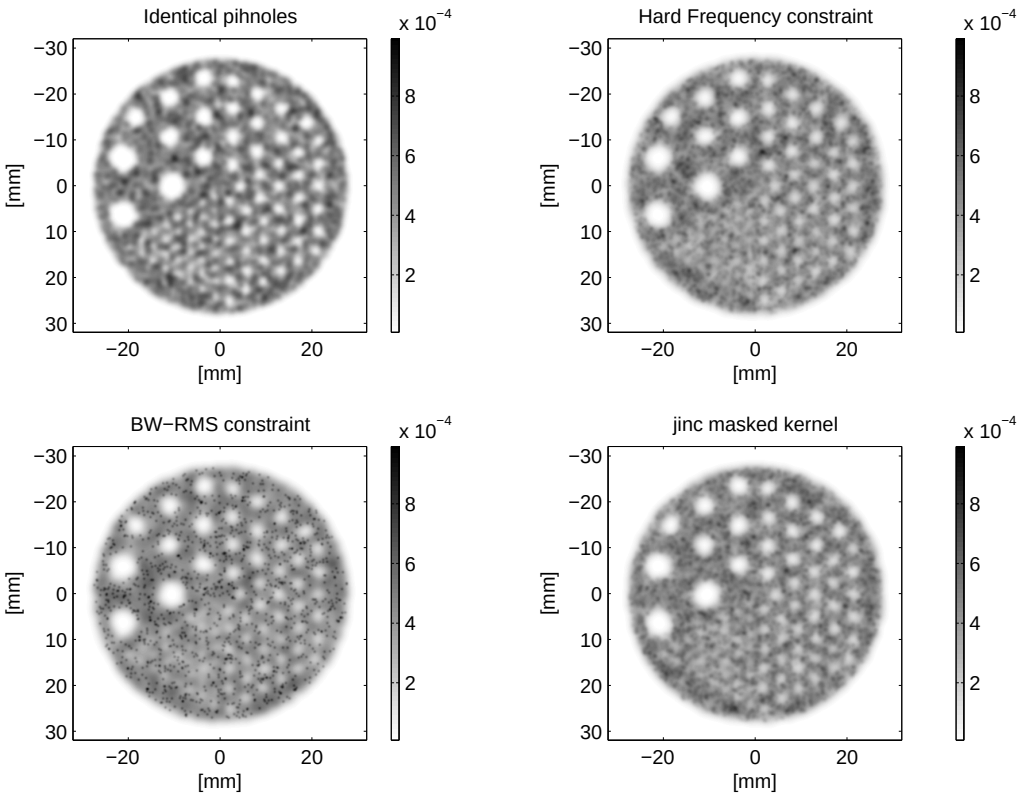

Figure 2.3: We show samples of reconstructed images of all systems used in the mentioned simulations. Top left, a standard system. Top right, a new system satisfying the hard frequency constraint. Bottom left, a new system satisfying the BW-RMS constraint. Bottom right, a new system using a jinc masked reconstruction kernel satisfying hard frequency constraint.

of our variance approximation, as the predicted variance curve agrees well with the calculated variance of each system. The actual variance gain that we see is about 39%. As in any other equivalent estimation problem (see [26], [17]), we note the bias-variance tradeoff effect. The new system, with a lower estimated image variance also produce blurrier images (see the upper row in figure 2.5). Thus, despite the fact that both system satisfy the same hard frequency constraint, it is not clear if the new system visually "outperforms" the standard system application wise.

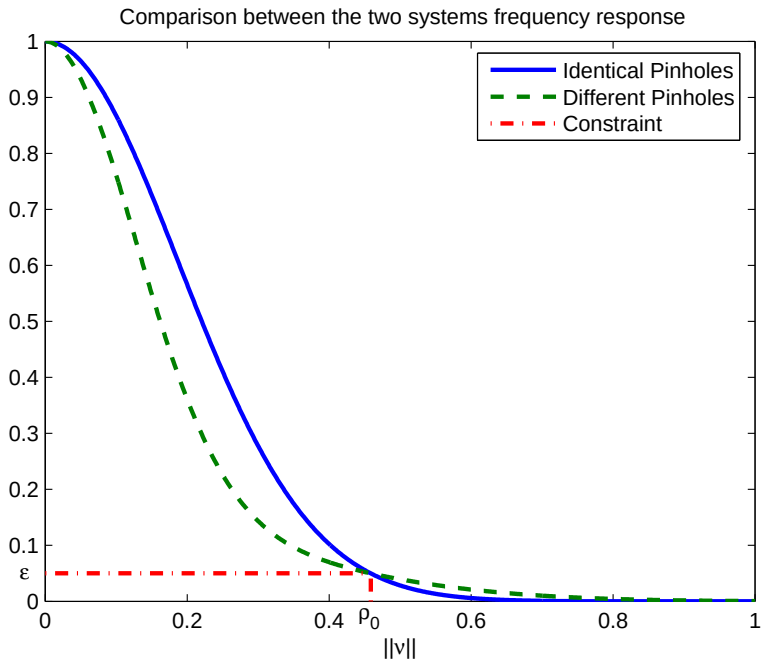

Figure 2.4: Hard Frequency constraint: Comparison of the two systems frequency response functions showing that both response satisfy the constraint. The solid-blue line represents the standard system, while the green-dashed line represents the new system.

2.4.2 BW-RMS constraint

Once again we select η that match $\beta = 2$ in the standard system. To compare our result to the previous system simulated, we choose to simulate the case $\beta_{\max} = 3$. From figure 2.7 we derive the following system: $\beta_1 = 0.7348, \beta_2 = 3$ and $\alpha = 0.07$. The expected improvement in the variance is about 12%, and it is also the variance gain that we see in the simulation results. Note that the setting of the new system simulated in section 2.4.1 resulted in a better improvement of the variance. Figures 2.8 and 2.9 shows the simulation results. We note that the new system with the lower estimation variance also produce a blurrier images (figure 2.8 top row), and that, as expected, the variance reduction is about 12%. It is interesting to compare this system's performance to the one we consider for the previous constraint. Lower variance reduction in the current system should yield better images and this may

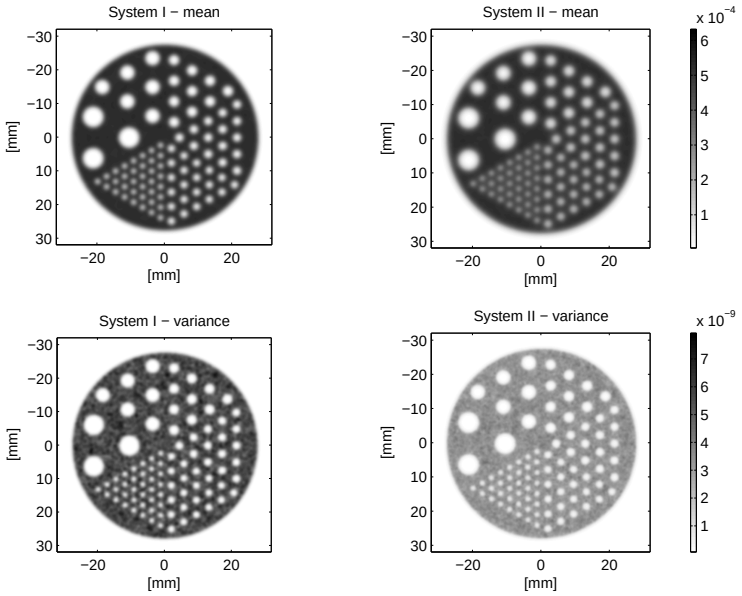

Figure 2.5: Hard Frequency constraint: Comparison of the mean (top row) and variance (bottom row) of the two systems. The new system results are presented on the right column and the standard system on the left.

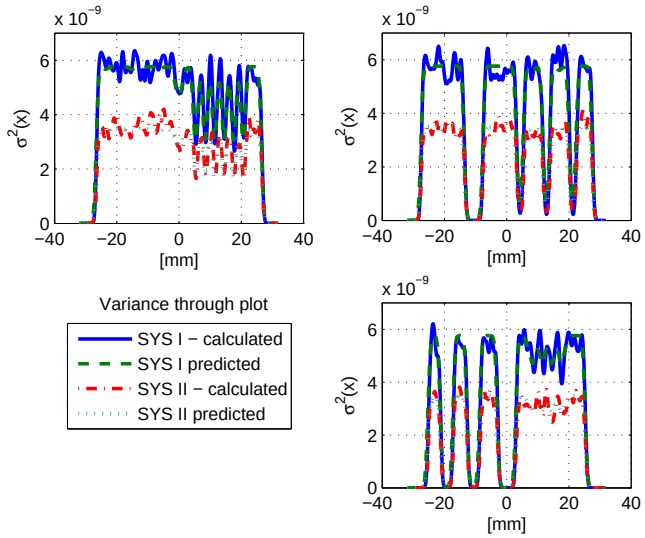

Figure 2.6: Hard Frequency constraint: Variance plots through slices of the Derenzo phantom. We compare the two system performance as well as validating the accuracy of our variance approximations.

be the case if we compare the estimators means (top right image in figures 2.5 and 2.8). However, comparing single realizations, e.g. figures 2.12 and 2.14, we note that the previous system had a more appealing results. This is due to the different in parameterizations. Although we now have $\beta_1 = 0.7348$ which should give sharper results, $\alpha = 0.07$ prevents the estimator from giving enough weight to the "better" reconstruction.

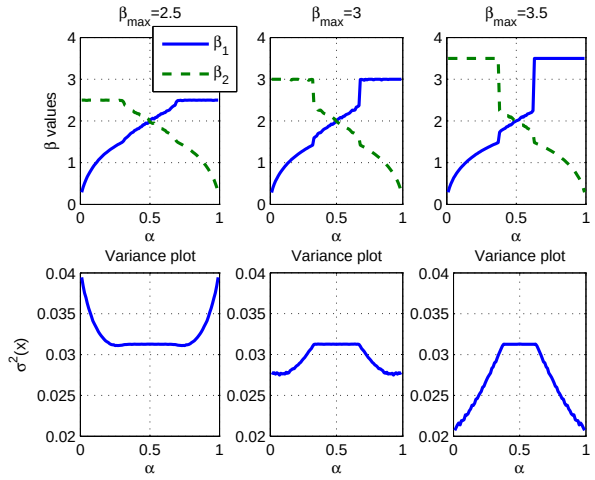

Figure 2.7: BW-RMS constraint: In the upper row we plot the search results for β_i vs. α, and in the lower row we plot the estimation variance vs. α for the β_i's given in each upper figure.

2.4.3 Reconstruction Kernel comparison

So far we have seen some discrepancy between minimizing the variance of the estimated image and its actual visual appearance. It seems that our new systems, though minimizing the estimation variance compared to the standard system, produce less appealing images. In an effort to improve the new system performance I tried using a different reconstruction kernel. In all the simulation results displayed so far we used the same Gaussian reconstruction kernel $g_\beta(\cdot)$. This kernel is strictly positive, and note that we only require $\int g_\beta = 1$. To reduce the blur in the reconstructed images I also explored using a non-strictly positive kernel. It was thought

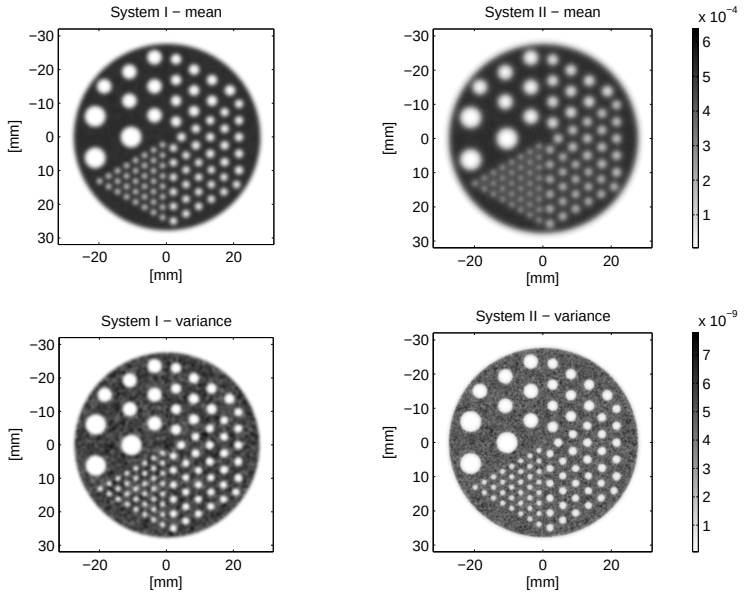

Figure 2.8: BW-RMS constraint: Comparison of the mean (top row) and variance (bottom row) of the two systems. The new system results are presented on the right column and the standard system on the left.

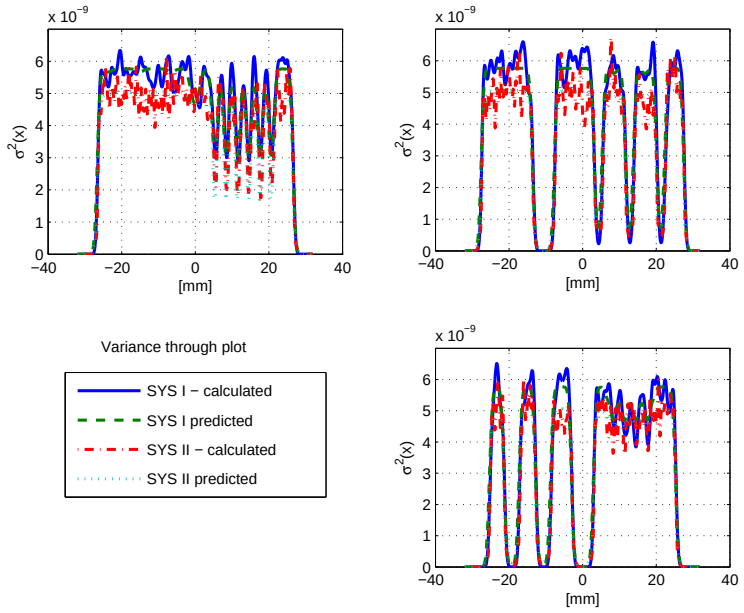

Figure 2.9: BW-RMS constraint: Variance plots through slices of the Derenzo phantom. We compare the two system performance as well as validating the accuracy of our variance approximations.

that a reconstruction kernel with negative (or partially negative) tail would reduce the blur in the reconstructed image. We used the parameters of the hard frequency constraint system, namely $\beta_1 = 1.257,\ \beta_2 = 3$ and $\alpha = 0.16$, and compared two other reconstruction kernels. The first one had a strictly negative tail, and the second had a sinusoidal tail. Both kernels resulted in negligible differences both in the estimation variance and the reconstructed image blur. As it turns out, the dominant feature of a reconstruction kernel is its FWHM. Since our new kernels only differ from the Gaussian kernel in their tail sections, the difference in the final result was insignificant.

2.5 Single pinhole with a general reconstruction kernel

To complete our analysis we wish to explore the possibilities of other useful reconstruction kernels. Based on the general results in [20] (equation (17) in page 246), we note that the estimation variance is proportional to the integral of $|G_\beta(\underline{\nu})|^2$. Assuming that the Fourier transform of the reconstruction kernel is real, we pose the following minimization problem in a Calculus of Variations format. Find $f : \mathbb{R}^2 \to \mathbb{R}_+,\ f \in L^2$ that minimizes

$$J(f) = \int f^2(\underline{\nu}) d\underline{\nu}, \tag{2.59}$$

subject to constraints (2.19) and (2.21), where the system frequency response is

$$H(\underline{\nu}) = f(\underline{\nu}) H_0(\underline{\nu}), \tag{2.60}$$

and $H_0(\underline{\nu})$ is the fourier transform of $h(\underline{v} - \underline{x})$. Starting with condition (2.21), we form the following Lagrangian

$$J(f, \phi) = \int f^2(\underline{\nu}) d\underline{\nu} + \phi\left(\int |f(\underline{\nu}) H_0(\underline{\nu})|^2 d\underline{\nu} - \gamma\right), \tag{2.61}$$

where ϕ is a Lagrange multiplier. If we further assume that $H_0(\underline{\nu})$ is real, e.g., a Gaussian profile pinhole, then our problem becomes

$$\min_{f,\phi} \left\{ J(f,\phi) = \frac{1}{2} \left[\int f^2(\underline{\nu})(1+\phi H_0^2(\underline{\nu}))d\underline{\nu} - \phi\gamma \right] \right\}, \tag{2.62}$$

where the factor $\frac{1}{2}$ was added as a matter of convenience. Taking derivative of $J(f,\phi)$ w.r.t. f and equating it to zero, yields

$$\frac{\partial}{\partial f} J(f,\phi) = J'_f(f,\phi) = \int f(\underline{\nu})(1+\phi H_0^2(\underline{\nu}))d\underline{\nu} = 0. \tag{2.63}$$

First, we obviously ignore the trivial solution $f = 0$ almost everywhere (a.e.). Also, since $f \in L^2$, we may rule out any solution of the form of a Dirac delta function. Another valid solution for (2.63) can be found if $f \perp (1+\phi H_0^2) \Rightarrow < f, H_0^2 >= c$, for some constant $c \in \mathbb{R}$. However $< f, H_0^2 >= c$, implies that we may decompose $f = f' + c'$, where c' is a constant and $< f', H_0^2 >= 0$. Since f is the Fourier transform of the reconstruction kernel it is in general a function of non-compact support. Hence it cannot contain an additive constant since a constant function defined on $\mathbb{R}^2$ is not integrable. Finally, consider the following function

$$f(\underline{\nu}) = \begin{cases} f_0, & ||\underline{\nu}|| < \alpha \\ 0, & \text{otherwise} \end{cases}. \tag{2.64}$$

For any such function, we have

$$J'_f(f,\phi) = f_0 \int_{||\underline{\nu}||<\alpha} (1+\phi H_0^2(\underline{\nu}))d\underline{\nu} = 0, \tag{2.65}$$

as long as

$$\phi = -\frac{-\pi\alpha^2}{\int_{||\underline{\nu}||<\alpha} H_0^2(\underline{\nu})d\underline{\nu}}. \tag{2.66}$$

To satisfy the constraint we must have

$$\int_{||\underline{\nu}||<\alpha} |f(\underline{\nu})H_0(\underline{\nu})|^2 d\underline{\nu} = \gamma. \tag{2.67}$$

Hence,

$$f_0 = \sqrt{\frac{\gamma}{\int_{||\underline{\nu}||<\alpha} H_0^2(\underline{\nu})d\underline{\nu}}}, \tag{2.68}$$

and, therefore $\alpha \to 0 \Rightarrow f_0 \to \infty$. Since $J(f,\phi)$ is convex in f, (2.64) is a minimum, and the corresponding minimal variance is

$$J(f) = \int f^2(\underline{\nu})d\underline{\nu} = f_0^2 \pi\alpha^2 = \frac{\gamma\pi\alpha^2}{\int_{||\underline{\nu}||<\alpha} H_0^2(\underline{\nu})d\underline{\nu}}. \tag{2.69}$$

If we further assume that $H_0(\underline{0}) = c \neq 0$, we have

$$\lim_{\alpha\to 0} J(f) = \frac{\gamma}{c^2}. \tag{2.70}$$

We note that (2.70) is a lower bound of the variance in this case, since $H_0(\underline{0}) = 1$ and H_0 was assumed to be a non-increasing function. Hence, the lower bound of the estimation variance is proportional to the constraint. This means that minimizing the variance, subject to (2.21) yields a very low bandwidth for the reconstruction kernel, which in turns implies poor resolution of the overall system.

Next we seek to minimize (2.59) subject to (2.19). Rearranging the constraint we derive the following functional

$$J(f,\phi) = \frac{1}{2}\left[\int f^2(\underline{\nu})d\underline{\nu} + \phi\int f^2(\underline{\nu})H_0^2(\underline{\nu})(||\underline{\nu}||^2 - \eta)d\underline{\nu}\right]. \tag{2.71}$$

Taking derivative w.r.t. f, we get

$$J'_f(f,\phi) = \int f(\underline{\nu})\left[1 + \phi H_0^2(\underline{\nu})(||\underline{\nu}||^2 - \eta)\right] d\underline{\nu} = 0. \tag{2.72}$$

Consider a solution of the form (2.64), then[3]

$$\phi = \frac{-\pi\alpha^2}{\int_{||\underline{\nu}||<\alpha} H_0^2(\underline{\nu})(||\underline{\nu}||^2 - \eta)d\underline{\nu}} \tag{2.73}$$

is enough to guaranty that (2.72) holds. However, using constraint (2.19) to derive f_0 we note that f_0 cancels out

$$\frac{f_0^2 \int_{||\underline{\nu}||<\alpha} ||\underline{\nu}||^2 H_0^2(\underline{\nu})d\underline{\nu}}{f_0^2 \int_{||\underline{\nu}||<\alpha} H_0^2(\underline{\nu})d\underline{\nu}} \geqslant \eta^2, \tag{2.74}$$

and, thus we may set $f_0 = 1$ and it remains to figure out α. Rearranging (2.74) we get

$$\int_{||\underline{\nu}||<\alpha} (||\underline{\nu}||^2 - \eta^2) H_0^2(\underline{\nu})d\underline{\nu} \geqslant 0. \tag{2.75}$$

Therefore, we must have $\alpha > \eta$. However the exact lower limit of α depends on the specific H_0. Minimizing (2.59) suggests the selection of the minimal α allowed and a minimal $f_0 > 0$. This yields the behavior that we had witnessed in the two pinhole case, i.e., we may find a trajectory, in (α, f_0) plane, that satisfy the constraint yet yielding $\sigma^2(\underline{x}) \to 0$. To conclude this section we note that the first constraint, which was forcing the frequency response to exceed a certain value at a specific frequency is not suitable for the Calculus of variations framework.

2.6 Conclusions and future work

We have shown examples where selecting different diameter pinholes results in a lower estimation variance compared to a standard system with equal width pinholes. Moreover, we showed limiting behavior of the considered systems. Under the hard

[3]Since $H_0 \neq 0$ a.e. the denominator of (2.73) is not zero.

frequency constraint the system degenerates to a single pinhole of constant width and constant weight α. For the BW-RMS constraint the system degenerates to a composition of a very narrow pinhole with very little weighting coefficient and a very wide pinhole that yield most of the estimated image. We had also shown roots for that behavior in a single pinhole system. All of the above holds for an estimator reconstructing the underlying images as a linear combination of the two independent system measurements.

Notwithstanding the reduction in estimation variance, we had come to question the validity of our cost function as a performance measure for an actual imaging system. It is either that minimizing variance has little merit for capturing image quality, considering the bias-variance tradeoff, or that we failed to choose proper constraints for the minimization problem. As mentioned, quantifying resolution is a daunting task. Thus, we leave for future work the task of finding a more suitable cost function or constraints or both. A possible such constraint would be the FWHM of the overall system, although we could not come up with a good way of utilizing it. We believe that future work should consider multiplexing between the different pinholes, as this is often the case for real systems.

There are obvious advantages for using a multi-resolution design for SPECT scanners. The constant struggle over SNR and resolution encourage us to find a way, perhaps non-linear, to combine a high-resolution noisy image with a low resolution low noise background image. Such an estimator, although hard to analyze, may produce better images for the actual tasks they are being used for by physicians. Another viable future direction is to consider using a (two) multi-diameter pinhole scanner in an adaptive manner. Such a system may start by producing a coarse image, then refine the image by re-scanning parts of the object where details are

needed using a finer pinhole diameter.

2-A Appendix: Full size reconstructed images of the different systems

For comparison purposes we present the four images shown in figure 2.3 in full size. Figure 2.11 was generated with the standard system while 2.12 and 2.13 were generated by a system satisfying the hard frequency constraint where the latter used the jinc masked reconstruction kernel. Finally, figure 2.14 was generated by a system using the BW-RMS constraint. We can see that in figures 2.12-2.14 it is hard to see the smallest radii discs due to the extra blur. Also, in figure 2.14 we see the nature of the BW-RMS constraint driving the image to appear flat with some details (black spots) contributed by the smaller pinhole. In addition, in figure 2.10 (which is the same as figure 2.7) we show, for the BW-RMS constraint, the behavior of β_1 β_2 and the expected variance as a function of α plotted for different $\beta_{\max}$. We observe the following: (i) As expected the results are symmetric in β, i.e. for every triplet $(\beta_1, \beta_2, \alpha)$ there is a matching triplet $(\beta_2, \beta_1, 1-\alpha)$. (ii) The reference standard system ($\beta_1 = \beta_2 = 2$) is a specific case of this grid search with $\alpha = \frac{1}{2}$. (iii) As we increase $\beta_{\max}$ the optimal solution, in terms of minimal variance, converges to $\beta_1 \to 0$ and $\sigma_0^2(\underline{x}) \to 0$. We also note that the gain in the expected variance in the new system is far less than the one expected by the new system satisfying the hard frequency constraint, yet the reconstructed image look much worse.

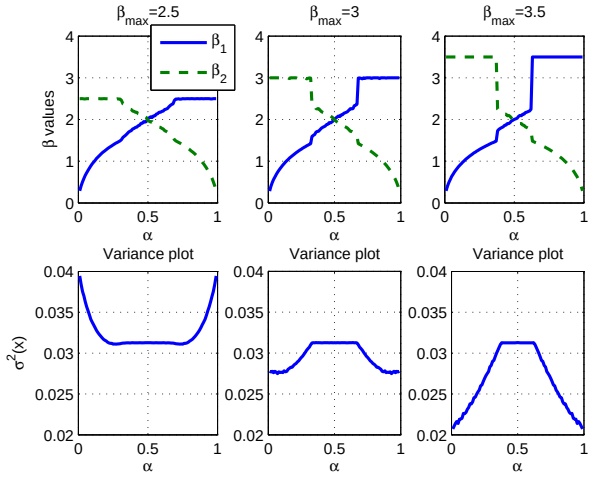

Figure 2.10: BW-RMS constraint: In the upper row we plot the search results for β_i vs. α, and in the lower row we plot the estimation variance vs. α for the β_i's given in each upper figure.

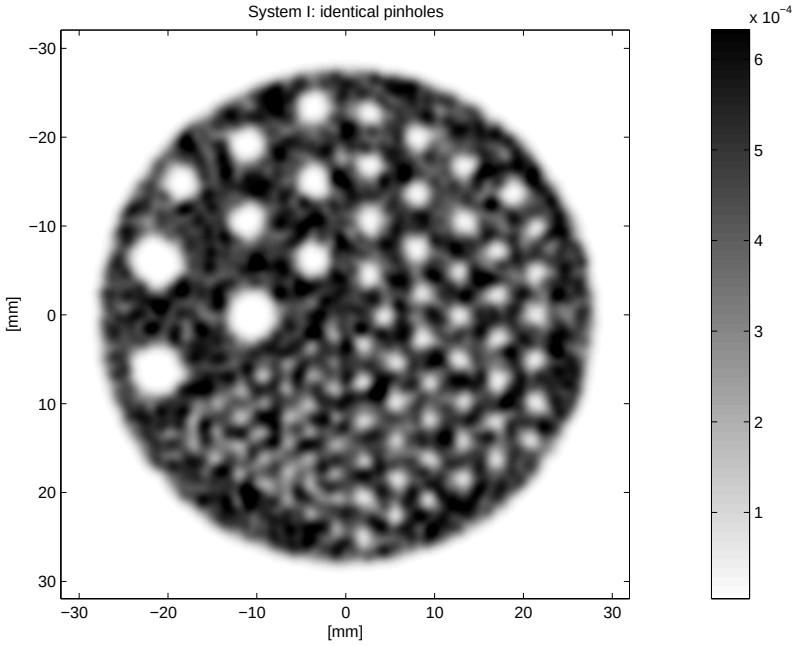

Figure 2.11: Standard system, image reconstruction of the Derenzo phantom.

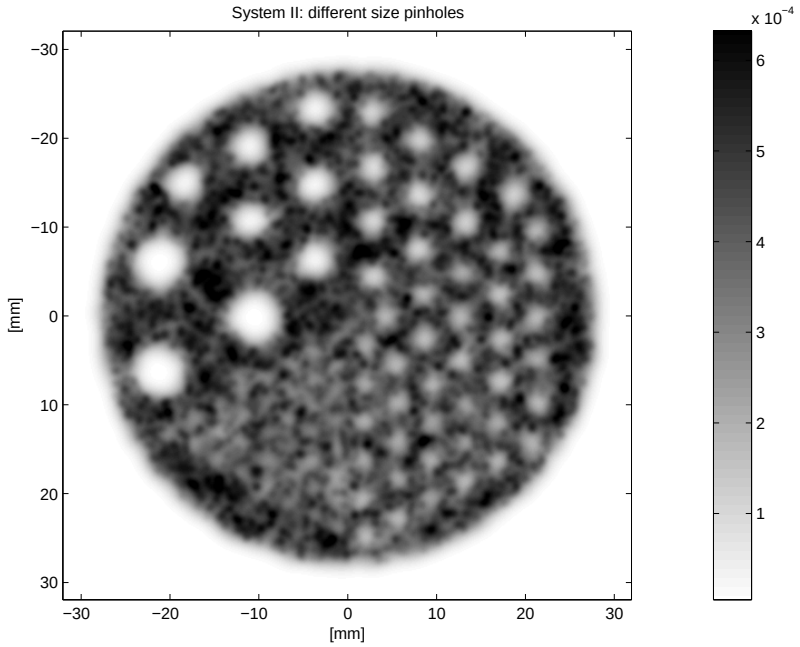

Figure 2.12: Hard frequency constraint: New system Gaussian reconstruction kernel, image reconstruction of the Derenzo phantom.

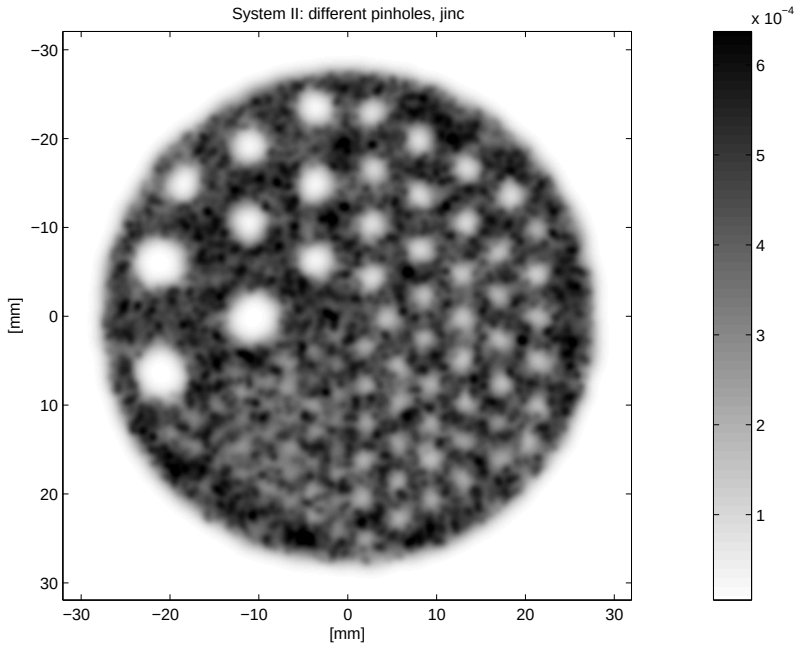

Figure 2.13: Hard frequency constraint: New system, image reconstructed using jinc masked kernel.

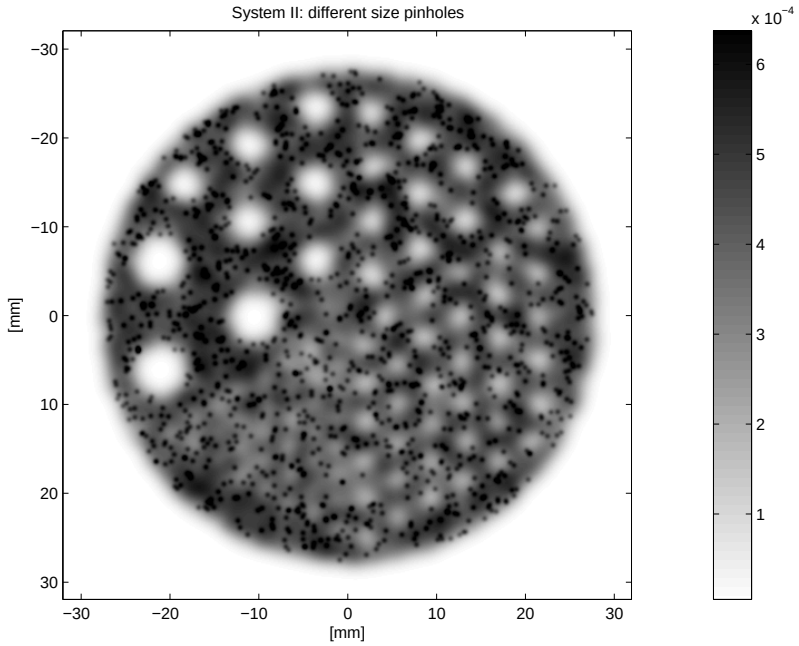

Figure 2.14: BW-RMS constraint: New system, image reconstruction of the Derenzo phantom.

CHAPTER III

Optimal Two-stage Search for Sparse Targets using Convex Criteria

3.1 Introduction

This chapter considers the problem of estimating and detecting sparse signals over a large area of an image or other medium. We introduce a novel cost function that captures the tradeoff between allocating energy to signal regions, called regions of interest (ROI), versus exploration of other regions. We show that minimizing our cost guarantees reduction of both the error probability over the unknown ROI and the mean square error (MSE) in estimating the ROI content. Two solutions to the resource allocation problem, subject to a total resource constraint, are derived. Asymptotic analysis shows that the estimated ROI converges to the true ROI. We show that our adaptive sampling method outperforms exhaustive search and are nearly optimal in terms of MSE performance. An illustrative example of our method in SAR imaging is given.

This problem arises in many applications including: target detection and classification, computer aided diagnosis, and screening. For example, in a radar reconnaissance mission, active radar may be used to image a given scene. A typical system is designed to detect targets exceeding a minimal profile. This minimal target profile

dictates the scan parameters such as the energy level the radar transmits and the scan time duration. Moreover, targets usually occupy a small section of the scanned area, called the ROI. Most systems consider exhaustive search with equal energy allocation, defined as a search policy where all cells are searched with equal effort allocated to each cell, to spread sensitivity over all locations. As a result, a relatively large portion of the energy is allocated outside the ROI. This excess energy could be used to better illuminate the ROI. Furthermore in surveillance applications, by deploying energy over an unknown area the searcher risks exposure. Reducing the scan energy outside the ROI reduces exposure risks.

As another application, consider the task of early detection of tumors using medical imaging, e.g., X-ray computed tomography (CT). Early detection is concerned with detecting small tumors. In many cases, little *a priori* knowledge about the tumor location exists. Consider the area containing the tumor as an unknown ROI. Lifetime radiation exposure constraints limit the total energy that can be used in a specific scan. There are two questions: *a*) where are tumors located? *b*) what kind of tumors does the ROI contain? This combined detection/estimation suggests using adaptive sampling over the image to improve both detection and estimation performance.

The search problem considered in this chapter bears some similarity to Posner's work on minimizing expected search time for finding a satellite lost in the sky [46]. Posner suggests a two step procedure. First, briefly search the entire domain to generate a likelihood function of possible satellite locations. Second, sequentially search the domain again in order of decreasing likelihood of satellite position. Posner's model assumes that the search is stopped as soon as the satellite has been found and that detection probability is increasing with search time. Therefore, se-

quentially searching the cells with the highest likelihood reduces the overall expected search time. By minimizing expected search time Posner imposes a 'soft' resource constraint on the total time used to search each cell. In this chapter, we adopt a Bayesian framework for sequential search for multiple objects obscured by noise. Thus the posterior distribution of objects presence replaces the likelihood of object presence used in [46]. The use of the posterior in place of the likelihood guarantees minimization of an average cost and a stronger sense of optimality.

Although we search for multiple targets within a signal, we focus on applications where the total support of the signal part containing targets is small compared to the entire signal support. Such signals can be viewed as sparse and we define the sparsity parameter p as the proportion of the signal support containing targets. Johnstone and Silverman consider a Bayesian framework for estimating sparse signals from a single measurements vector in [29]. They consider thresholding the vector entries and setting the estimated signal equal zero for all measurements below the threshold. Thus, significant gains in estimation mean square error (MSE) are achieved for small p. They utilize a Bayesian framework to find an optimal threshold minimizing the average squared error. We also use the Bayesian framework to find an optimal threshold. However, we use this threshold to generate additional measurements of all signal elements exceeding it. This adaptive sampling approach and added measurements enables higher gains in reducing estimation MSE. Wipf and Rao use sparse Bayesian learning in the problem of basis selection from overcomplete dictionaries [60]. They use a parameterized prior on the basis weight coefficients to encourage sparse solution to an otherwise l_2-norm minimization problem. The parameters of the prior are estimated from the data along with the basis weights using an expectation-maximization (EM) algorithm. However, the EM algorithm uses a

single measurement of the underlying signal. In this paper, the posterior distribution is used to partially remeasure the underlying signal. By spatially focusing the measurement process onto the ROI we can better estimate the correct signal values.

Adaptive sampling or active learning for the purpose of estimating functions in noise has been considered in [13]. Castro et al. show that for piecewise constant functions, active learning methods can capitalize on the highly localized nature of the boundary by focusing the sampling process in the estimated vicinity of the boundary. Thus, estimation error converges at higher rates than with passive sampling. A two-step approach is proposed where first a rough estimate of the underlying signal is generated and second the signal is resampled in the vicinity of the estimated boundaries. We use a similar two-stage approach where previous measurements are used to determine where to sample next. However, our work differs in two aspects as we do not limit the discussion to a class of spatially inhomogeneous signals and we consider the additional aspect of resource allocation. While [13] assume identical sampling procedure for all samples, e.g. similar measurement noise variance, we consider different sampling procedures between stages and among different spatial locations within a particular stage.

Resource allocation in the context of adaptive waveform amplitude design for estimating parameters of an unknown medium under average energy constraints is discussed in [49]. Rangarajan et al. derive an optimal amplitude allocation for the second stage in a two-stage problem as a function of the first stage measurements. In this work we consider the more general problem of waveform design under a total energy constraint. Therefore, measurements at the first stage are used to select the optimal waveform among all possible waveforms with some bounded total energy. Thus we are able to focus the sampling process onto the ROI.

The cost function considered here is similar in nature to a terminal/total reward utility function. This formulation is used in some multi arm bandit (MAB) problems (see [25], pp. 123). The difference between MAB and our formulation of the search problem is that each action we take affects the posterior probability distribution of target locations, and our method is greedy.

In this chapter, we focus on adaptively determining the ROI that contains targets of interest. Two main contributions in this chapter are: 1) we introduce a novel convex cost function for optimizing the ROI search; 2) we provide two-stage optimal and suboptimal adaptive search policies with respect to (w.r.t.) our cost function. Remarkably, this leads to solutions that minimize both the Chernoff bound on error probability and the *Cramér-Rao* bound on estimating the parameter values within the ROI. The optimal and suboptimal policies are greedy search algorithms with complexity order proportional to the discrete signal support, Q. The optimal policy rank orders the posterior distribution values then finds an optimal threshold and assigns additional effort to all cells with posterior values exceeding the threshold. On the other hand, the suboptimal policy uses the posterior distribution values to assign additional effort to all cells, thus requiring an order $\mathcal{O}(\log Q)$ fewer computations. Both policies outperform an exhaustive search scheme in terms of post-processing tasks such as target detection and ROI estimation. Moreover, an asymptotic performance analysis at high signal to noise ratio (SNR) is given and shows that the estimated ROI converges to the true ROI and the performance gain approaches a theoretical limit, which is inversely proportional to the sparsity of the signal.

Our optimal resource allocation policies are derived for a two-stage resource allocation problem, based on a Gaussian observation model, and assume statistically independent targets. Our methods result in 6 [dB] performance gain estimating the

parameters value within the ROI, at SNR of 13 [dB] and $p = 1\%$ (see Fig. 3.1). In addition, the radar imaging example in Section 3.5 uses a non-Gaussian model but still results in significant performance gain.

The rest of this chapter is organized as follows: Section 3.2 formally states the problem and introduces our cost function. In Section 3.3, we present optimal and suboptimal solutions for the resource allocation problem. Section 3.4 includes thorough performance evaluation of the two policies as compared to an exhaustive policy for both detection and estimation. An illustrative imaging example of our methods using synthetic aperture radar data is given in Section 3.5. Finally, we conclude and point to future work in Section 3.6.

3.2 Problem formulation

Consider a discrete space $\mathcal{X} = \{1, 2, \ldots, Q\}$ containing Q cells and equipped with a probability measure P. We use Ψ to denote an ROI in $\mathcal{X}$, i.e., $\Psi \subseteq \mathcal{X}$. In the sequel Ψ will be a randomly selected sparse subset of $\mathcal{X}$. We assume $|\Psi| \ll |\Psi^c|$, where $|\Psi|$ equals the number of elements in Ψ and Ψ^c is the relative complement $\mathcal{X} \setminus \Psi$ of Ψ. Exact definition of the ROI is application dependent. In radar target localization the ROI is the collection of all cells containing targets and target related phenomena, e.g., target shadows. In a medical imaging application, such as an early detection of breast cancer, the tumor has to be placed in some context in order to classify the cancer stage. For example, If a tumor is diagnosed as DCIS (Ductal Carcinoma In Situ) the patient is considered at stage 0. Therefore, in case of the latter, the ROI may be defined as the collection of all cells containing targets plus some neighboring cells.

Let I_i be an indicator function of the ROI such that

$$I_i = \begin{cases} 1, & i \in \Psi \\ 0, & \text{Otherwise} \end{cases} \tag{3.1}$$

and $\{p_i = \Pr(I_i = 1)\}_{i=1}^{Q}$ is an associated set of prior probabilities. We assume that the I_i's are independent. Let $\boldsymbol{I}_\Psi = [I_1, \ldots, I_Q]'$ be a vector corresponding to the set of all indicators and $(\cdot)'$ denotes the transpose operator. We say that the presence of a target affects cell i if $i \in \Psi$. Define the random vector $\boldsymbol{Y} : \mathcal{X} \to \mathbb{R}^Q$ and consider the conditional probability $p(\boldsymbol{Y}|\boldsymbol{I}_\Psi; \lambda)$, where λ is defined below.

Consider a sequential experiment where cell i may be sampled T times. By sampling we mean that $\boldsymbol{y}(t)$ a realization of $\boldsymbol{Y}$ is observed at time t. Let the distribution $\lambda(i,t) \geqslant 0$ denote the search effort allocated to cell i at time t, with

$$\sum_{t=1}^{T} \sum_{i=1}^{Q} \lambda(i,t) = 1, \quad 0 \leqslant \lambda(i,t) \leqslant 1 \tag{3.2}$$

and $\{\lambda(i,t)\}$ is a mapping from past observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$ to the probability simplex and is called an effort allocation policy, or equivalently, a search policy. We focus here on a deterministic mappings λ, although a more general random mapping can also be incorporated into our framework but is not presented here. We assume that the 'quality' of the samples is an increasing function of the allocated effort to the associated cell, e.g. measured in terms of Fisher information or inverse variance. In general, effort might be: time, computing power, complexity, cost, or energy, that is allocated to acquiring a particular cell location. Define the cumulative search effort allotted to cell i, as

$$\Lambda(i) = \sum_{t=1}^{T} \lambda(i,t). \tag{3.3}$$

Consider the following cost function

$$J(\Lambda) = \sum_{i=1}^{Q} I_i A(\Lambda(i)) + \sum_{i=1}^{Q} B(\Lambda(i)), \tag{3.4}$$

where $A(\cdot)$ and $B(\cdot)$ are decreasing functions that may depend on additional parameters. This restriction ensures that allocating more effort to cells reduces the over-all cost. Note that our cost function (3.4) depends directly on the ROI via the summand of the first sum on the right hand side (r.h.s.) of (3.4). Choosing $B(\cdot) = 0$ emphasizes focusing efforts on the ROI. Selecting $A(x) = \frac{2\nu-1}{x}$ and $B(x) = \frac{1-\nu}{x}$, with $\nu \in [\frac{1}{2}, 1]$, simplifies $J(\Lambda)$ to

$$J(\Lambda) = \sum_{i=1}^{Q} \frac{\nu I_i + (1-\nu)(1-I_i)}{\Lambda(i)}, \tag{3.5}$$

which has some intuitive and appealing properties. Minimizing this cost function concentrates ν of the total effort over the ROI Ψ and $(1-\nu)$ to its complement Ψ^c with ν being the dividing factor. Setting $\nu = 1$ focuses all the effort at the ROI, while $\nu = \frac{1}{2}$ results in an exhaustive effort allocation policy, i.e., equal energy allocated to all cells inside and outside the ROI. The choice $\nu = 1$ has some very intuitive and appealing properties. For example, in the context of estimating a deterministic signal $\boldsymbol{\mu}$ in additive Gaussian noise, minimizing (3.5) is equivalent to minimizing the *Cramér-Rao* lower bound on $\mathrm{E}\{\sum_i I_i(\hat{\mu}_i - \mu_i)^2\}$, see Appendix 3-A for details. In a sense, ν controls the tradeoff between exploitation of signal in the ROI ($\nu = 1$) and exploration of the entire signal ($\nu = \frac{1}{2}$).

In addition, consider a binary Gaussian hypothesis testing problem. Define the null hypothesis, $\mathcal{H}_0(i)$, as $\mu_i = 0$, and the alternative, $\mathcal{H}_1(i)$, as $\mu_i > 0$ with a prior probability $\pi = \Pr(\mathcal{H}_1)$. Consider the task of deciding between the two hypothe-

sis, $\forall i \in \Psi$. The probability of error, i.e., making the wrong decision, defined as $P_e = \Pr(\text{decide } \mathcal{H}_0|\mathcal{H}_1)\pi + \Pr(\text{decide } \mathcal{H}_1|\mathcal{H}_0)(1-\pi)$, can be broken into two parts: misdetect probability P_m over Ψ, and false alarm probability P_{fa} over Ψ^c. With $\nu = 1$, we show in Appendix 3-B that minimizing (3.5) is equivalent to uniformly minimizing the Chernoff bound on the probability of error P_m over the ROI. Setting $\nu = \frac{1}{2}$, most of the energy is spread over Ψ^c due to the assumed sparsity $|\Psi| \ll |\Psi^c|$. This leads to a lower Chernoff bound on P_{fa}, or correspondingly, fewer false alarms. If $\nu \in (\frac{1}{2}, 1)$ we trade the two cases, either relaxing the upper bound on P_m or on P_{fa}. In Section 3.4 we corroborate this intuition with simulation results, where adaptive measurement policies derived from minimization of (3.5) are used to generate data that is used for both estimation and detection tasks.

Next, we provide an achievable lower bound on our cost function (3.5).

Lemma 3.1. *The cost function (3.5) is lower bounded by*

$$J(\Lambda) \geqslant \left[\sqrt{\nu}|\Psi| + \sqrt{1-\nu}(Q - |\Psi|)\right]^2. \tag{3.6}$$

This lower bound is achievable with a two-level effort allocation scheme Λ_o, equal over the ROI and equal over its complement, defined as

$$\Lambda_o(i) = \begin{cases} \frac{\sqrt{\nu}}{\sqrt{\nu}|\Psi| + \sqrt{1-\nu}(Q-|\Psi|)}, & i \in \Psi \\ \frac{\sqrt{1-\nu}}{\sqrt{\nu}|\Psi| + \sqrt{1-\nu}(Q-|\Psi|)}, & i \in \Psi^c \end{cases}. \tag{3.7}$$

Proof. For a nonnegative series $\{a_i\}$, Cauchy-Schwarz inequality provides

$$\left(\sum_{i=1}^{Q} \frac{a_i}{\Lambda(i)}\right)\left(\sum_{i=1}^{Q} \Lambda(i)\right) \geqslant \left(\sum_{i=1}^{Q} \sqrt{\frac{a_i}{\Lambda(i)}}\sqrt{\Lambda(i)}\right)^2. \tag{3.8}$$

Since $\sum_{i=1}^{Q} \Lambda(i) = 1$, (3.8) simplifies to

$$\sum_{i=1}^{Q} \frac{a_i}{\Lambda(i)} \geqslant \left(\sum_{i=1}^{Q} \sqrt{a_i} \right)^2 . \tag{3.9}$$

Substituting $a_i = \nu I_i + (1-\nu)(1-I_i)$, i.e., $\forall i \in \Psi : a_i = \nu$ and $\forall i \in \Psi^c : a_i = 1-\nu$, into the r.h.s. of (3.9) yields

$$\left(\sum_{i=1}^{Q} \sqrt{a_i} \right)^2 = \left[\sqrt{\nu}|\Psi| + \sqrt{1-\nu}(Q - |\Psi|) \right]^2 . \tag{3.10}$$

Noting that on the l.h.s. of (3.9) we have $J(\Lambda) = \sum_{i=1}^{Q} \frac{a_i}{\Lambda(i)}$ proves the validity of (3.6). To prove the second part of the *lemma* note that

$$J(\Lambda_o) = \sum_{i \in \Psi} \frac{\nu}{\Lambda_o(i)} + \sum_{i \in \Psi^c} \frac{1-\nu}{\Lambda_o(i)}, \tag{3.11}$$

and algebra yields

$$J(\Lambda_o) = \left[\sqrt{\nu}|\Psi| + \sqrt{1-\nu}(Q - |\Psi|) \right]^2 , \tag{3.12}$$

which is exactly the r.h.s. of (3.6). This completes the proof. □

Discussion

We would like to point out the potential performance gains using our cost function (3.5). Let Λ_U denote an exhaustive search policy with $\Lambda_U(i) = \frac{1}{Q}$. As previously mentioned, Λ_U is the optimal search policy for $\nu = \frac{1}{2}$ in (3.7). For a general ν, the cost (3.5) associated with Λ_U is

$$J(\Lambda_U) = Q\left[\nu|\Psi| + (1-\nu)(Q - |\Psi|)\right], \tag{3.13}$$

with the two special cases of

$$J(\Lambda_U) = \begin{cases} \frac{Q^2}{2}, & \nu = 0.5 \\ \\ Q|\Psi|, & \nu = 1 \end{cases}. \tag{3.14}$$

Since Λ_U does not offer any adaptivity, any good adaptive effort allocation scheme Λ should result in $J(\Lambda) \leqslant J(\Lambda_U)$. Therefore, we define the performance gain in [dB] as

$$G(\Lambda) = -10 \log \frac{J(\Lambda)}{J(\Lambda_U)}. \tag{3.15}$$

For $\nu = 1$, $J(\Lambda_o) = |\Psi|^2$ and thus the optimal gain is

$$G(\Lambda_o) = -10 \log \frac{|\Psi|}{Q}, \tag{3.16}$$

which is achieved by the omniscient energy allocation strategy that knows the ROI and consequently concentrates all energy in the ROI. Define $p^* = \frac{|\Psi|}{Q}$, then $p^* \to 0$ forces $G(\Lambda_o) \to \infty$. Consequently, a good sampling method should yield large gains in a sparse setting, i.e., when p^* is small the ROI is small. In the following, we will develop a sampling method that exploits these gains. Taking the derivative of the r.h.s. in (3.6) w.r.t. ν, it can be shown that $J(\cdot) \geqslant |\Psi|^2$ for all $\nu \in [\frac{1}{2}, 1]$ and $\frac{|\Psi|}{Q} \leqslant \frac{1}{2}$ (see Appendix 3-C). In other words, if $\frac{|\Psi|}{Q} \leqslant \frac{1}{2}$ the optimal gain is achieved by Λ_o for $\nu = 1$. Unfortunately, Λ_o represents the omniscient search policy and is not a feasible policy since we do not know the location of the ROI in advance. Our goal is to derive a feasible policy that comes close to achieving the performance of Λ_o for unknown ROI.

In this chapter, we restrict our attention to minimizing the expected value of (3.5). This probabilistic setting utilizes the conditional distribution $p(\boldsymbol{Y}|\boldsymbol{I}_\Psi;\lambda)$ in our model. Assuming we observe realizations of $\boldsymbol{Y}$, our goal is to find a search policy

$$\hat{\lambda}(i,t) = \arg\min_{\lambda(i,t)} \mathrm{E}\left\{\sum_{i=1}^{Q} \frac{\nu I_i + (1-\nu)(1-I_i)}{\Lambda(i)}\right\}, \tag{3.17}$$

where $\Lambda(i)$ is given by (3.3). Next, we introduce a total energy constraint.

3.2.1 Energy allocation with energy constraint

Consider a fixed-target radar measurements in the presence of noise. We assume that a radar transmits energy in some known beam pattern to probe a collection of cells. We further assume that the radar is subject to an energy constraint, and that observations obey the following model

$$y_j(t) = \sum_{i=1}^{Q} h_{ij}(t)\sqrt{\lambda(i,t)}\theta_i(t)I_i + n_j(t), \quad t = 1, 2, \ldots, T, \tag{3.18}$$

where $h_{ij}(t)$ are known weights corresponding to the beam pattern, $\lambda(i,t)$ is the energy allocated for measuring cell i, $\theta_i(t)$ is a random return from cell i, and $n_j(t)$ is an additive observation noise, all at time t. Note that since the indicator of the ROI I_i is independent of t, this model corresponds to a static target scenario. We assume that the additive noise $n_j(t)$ is independent for different j and t. Also, assume that the positive $\theta_i(t)$ follow a prior distribution $p_\Theta(\theta)$ and are independent for different i but may be dependent for different t. The model in (3.18) can be written in vector form

$$\boldsymbol{y}(t) = H(t)'\mathrm{diag}\{\sqrt{\boldsymbol{\lambda}(t)}\}\mathrm{diag}\{\boldsymbol{\theta}(t)\}\boldsymbol{I}_\Psi + \boldsymbol{n}(t), \tag{3.19}$$

where

$$\begin{aligned}
\boldsymbol{y}(t) &= [y_1(t), y_2(t), \ldots, y_{M_t}(t)]', \\
\boldsymbol{\lambda}(t) &= [\lambda(1,t), \lambda(2,t), \ldots, \lambda(Q,t)]', \\
\boldsymbol{\theta}(t) &= [\theta_1(t), \theta_2(t), \ldots, \theta_Q(t)]', \\
\boldsymbol{I}_\Psi &= [I_1, I_2, \ldots, I_Q]', \\
\boldsymbol{n}(t) &= [n_1(t), n_2(t), \ldots, n_{M_t}(t)]',
\end{aligned}$$

and $[H(t)]_{ij} = h_{ij}(t)$. The notation $\sqrt{\boldsymbol{\lambda}(t)}$ denotes a $Q \times 1$ vector with $[\sqrt{\boldsymbol{\lambda}(t)}]_i = \sqrt{\lambda(i,t)}$, while the operator $\mathrm{diag}\{\boldsymbol{x}\}$ corresponds to a $Q \times Q$ (square) diagonal matrix with $x(i)$ as its ii^{th} element. The objective is to specify a resource allocation policy $\lambda(i,t)$ that improves the 'quality' of the measurements over the ROI as measured by (3.17). In addition we assume an energy constraint

$$\frac{1}{\lambda_T} \sum_{i=1}^{Q} \sum_{t=1}^{T} \lambda(i,t) = 1. \tag{3.20}$$

At time t, the energy allocated to cell i may depend on past observations, i.e., $\lambda(i,t)$ is a function of $\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots, \boldsymbol{y}(t-1)$. For brevity, we use the notation $\lambda(i,t)$ to denote $\lambda(i,t;\boldsymbol{y}(1),\boldsymbol{y}(2),\ldots,\boldsymbol{y}(t-1))$. Following (3.3), define Λ the cumulative energy distributed to cell i as $\Lambda(i) = \sum_{t=1}^{T} \lambda(i,t)$. Our cost function is $J(\Lambda)$, as defined in (3.5), and our goal is to minimize the expected cost in the r.h.s. of (3.17) over all possible energy allocations $\lambda(i,t)$, subject to (3.20). Consider $\nu = 1$ and let $p = \Pr(I_i = 1)$ be a uniform prior distribution on the location of targets, where p represents the sparsity of the vector $\boldsymbol{I}_\Psi$, i.e. $|\Psi|$ is a Binomial random variable (r.v.) with $\mathrm{E}|\Psi| = pQ$. Define $p^* = \frac{|\Psi|}{Q}$ then $\mathrm{E}\{p^*\} = p$ and $\mathrm{var}(p^*) = \frac{p(1-p)}{Q}$. To find an upper bound on possible performance gains we use $J(\Lambda_o) \leqslant J(\Lambda)$ for any Λ and,

thus

$$G(\Lambda) \leqslant G(\Lambda_o) = -10 \log p^*. \tag{3.21}$$

In Appendix 3-D we use (3.21) and Bernstein's inequality to prove, for $\delta \geqslant \exp\{-Q\frac{3p}{8(1-p)}\}$, that

$$\Pr(G(\Lambda) \leqslant -10 \log p + \epsilon(\delta, p, Q)) \geqslant 1 - \delta. \tag{3.22}$$

In the SAR imaging example from Section 3.5 this yields

$$\Pr(G(\Lambda) \leqslant 20 + 0.333) \geqslant 0.999.$$

3.3 Search policy

In the following section we solve the optimization problem for $T = 2$. The idea is to expend at $t = 1$ a portion of the total energy to learn about the data domain and obtain a posterior distribution. Then use it at stage $t = 2$ to refine the posterior and estimate the ROI. We will solve

$$\hat{\lambda}(i,t) = \arg\min_{\lambda(i,t)} \mathrm{E}\left\{ \sum_{i=1}^{Q} \frac{\nu I_i + (1-\nu)(1-I_i)}{\sum_{t=1}^{2} \lambda(i,t)} \right\}, \tag{3.23}$$

subject to the energy constraint $\frac{1}{\lambda_T} \sum_i \sum_t \lambda(i,t) = 1$. Initially, the prior distribution on targets location is uniform, $p_i = p$ for all i. This could be modified if there were prior knowledge on targets location. Let $\lambda^{(t)}$ be the total energy spent at search step t with $\frac{1}{\lambda_T} \sum_t \lambda^{(t)} = 1$, $t = 1, 2$. Our goal is to optimize the energy distribution between the two steps and among cells in each step. See [5] for a discussion of the case of $T > 2$.

3.3.1 Optimal two stage search policy

With $T = 1$ and a uniform prior, we show in Appendix 3-E.1 that the minimizer of the cost (3.23) is an equal energy allocation

$$\lambda(i,1) = \frac{\lambda^{(1)}}{Q} \triangleq \lambda_1. \tag{3.24}$$

Let Γ be the set of all search policies with $\lambda(i,1) = \lambda_1$. For $T = 2$ we find the optimal search policy $\lambda \in \Gamma$ minimizing (3.23). Since, $\lambda_T = Q\lambda_1 + \lambda^{(2)}$ optimizing the total effort allocated for each step is equivalent to finding an optimal pair $(Q\lambda_1, \lambda_T - Q\lambda_1)$, which involves minimizing over a single variable. Hence the cost function (3.23) simplifies to

$$J(\Lambda) = \sum_{i=1}^{Q} \mathrm{E}\left\{ \frac{\nu I_i + (1-\nu)(1-I_i)}{\lambda_1 + \lambda(i,2)} \right\}, \tag{3.25}$$

where expectation is taken w.r.t. $\boldsymbol{y}(1)$ and $\boldsymbol{I}_\Psi$. Note that $\lambda(i,2)$ depends only on $\boldsymbol{y}(1)$, thus $\boldsymbol{y}(2)$ does not affect the cost function and can be omitted from the expectation. In addition, (3.25) is constant in $\boldsymbol{\theta}$ and therefore we omit it from the expectation as well. Rewriting (3.25) using iterated expectation yields

$$J(\Lambda) = \sum_{i=1}^{Q} \mathrm{E}\left\{ \mathrm{E}\left\{ \frac{\nu I_i + (1-\nu)(1-I_i)}{\lambda_1 + \lambda(i,2)} \middle| \boldsymbol{y}(1) \right\} \right\}. \tag{3.26}$$

Note that I_i is a binary r.v.. In addition, given $\boldsymbol{y}(1)$, $\lambda(i,2)$ is deterministic. Hence (3.26) becomes

$$\sum_{i=1}^{Q} \mathrm{E}\left\{ \frac{\mathrm{E}\left\{\nu I_i + (1-\nu)(1-I_i) | \boldsymbol{y}(1)\right\}}{\lambda_1 + \lambda(i,2)} \right\} = \tag{3.27}$$

$$\sum_{i=1}^{Q} \mathrm{E}\left\{ \frac{\nu \mathrm{Pr}\left(I_i = 1 | \boldsymbol{y}(1)\right) + (1-\nu)(1 - \mathrm{Pr}\left(I_i = 1 | \boldsymbol{y}(1)\right))}{\lambda_1 + \lambda(i,2)} \right\}. \tag{3.28}$$

Using Bayes rule we obtain

$$\mathrm{Pr}(I_i = 1 | \boldsymbol{y}(1)) = \frac{P(\boldsymbol{y}(1)|I_i = 1)p}{P(\boldsymbol{y}(1)|I_i = 1)p + P(\boldsymbol{y}(1)|I_i = 0)(1-p)} \stackrel{\triangle}{=} p_{I_i|\boldsymbol{y}(1)}, \tag{3.29}$$

where $P(\boldsymbol{y}|I_i) = \int P(\boldsymbol{y}|I_i, \theta_i) p_\Theta(\theta) d\theta$ is the given conditional probability model describing the measurement dependency on the target. Finally, we rewrite our cost function on the r.h.s. of (3.25) and solve

$$\hat{\lambda}(i,t) = \arg \min_{\lambda_1, \lambda(i,2)} \mathrm{E}\left\{ \sum_{i=1}^{Q} \frac{\nu p_{I_i|\boldsymbol{y}(1)} + (1-\nu)(1 - p_{I_i|\boldsymbol{y}(1)})}{\lambda_1 + \lambda(i,2)} \right\}. \tag{3.30}$$

Let λ_1 be the energy allocated to each cell at the first step, with $Q\lambda_1 \leqslant \lambda_T$. Define the r.v.

$$W_j = \nu p_{I_j|\boldsymbol{Y}} + (1-\nu)(1 - p_{I_j|\boldsymbol{Y}}),\ j = 1, 2, \ldots, Q, \tag{3.31}$$

with vector of corresponding realizations $\boldsymbol{w} = [w_1, w_2, \ldots, w_Q]'$. Let $\tau : \mathcal{X} \to \mathcal{X}$ be a permutation operator that corresponds to the rank ordering of the w_i's:

$$\tau(j) = \arg \min_{i=1,\ldots,Q} \{w_i : w_i \geqslant w_{\tau(j-1)}\}, \qquad j \in \{1, 2, \ldots, Q\}, \tag{3.32}$$

with $w_{\tau(0)} \stackrel{\triangle}{=} 0$. Whenever the r.h.s. of (3.32) is not unique we select an arbitrary i satisfying $w_{\tau(1)} \leqslant w_{\tau(2)} \leqslant \ldots \leqslant w_{\tau(Q)}$. The solution λ_1, $\lambda(i,2)$ of (3.30) will depend on an integer threshold k_0 that is equal to the number of cells that are not searched at $t = 2$. Assuming[1] $w_{\tau(1)} > 0$, define k_0, the threshold parameter, as $k_0 = 0$ if

$$\frac{\lambda_T}{\lambda_1} > \frac{\sum_{i=1}^{Q} \sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(1)}}}, \tag{3.33}$$

[1]The special case of $w_{\tau(1)} = 0$ is treated in Appendix 3-E.2.

otherwise $k_0 \in \{1, \ldots, Q-1\}$ is the integer satisfying

$$\frac{\sum_{i=k_0+1}^{Q} \sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(k_0+1)}}} < \frac{\lambda_T}{\lambda_1} - k_0 \leqslant \frac{\sum_{i=k_0+1}^{Q} \sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(k_0)}}}. \tag{3.34}$$

An optimal solution for the second step $\lambda(i,2)$ is given in Appendix 3-E and the special case where $w_{\tau(i)} = 0$, for all i, is treated there (Appendix 3-E.2). Its properties are summarized below: Given λ_1 and k_0, it is shown that the optimal energy allocation $\lambda(\tau(i),2)$ minimizing the cost (3.30) is of the form

$$\lambda(\tau(i),2) = \left(\frac{\lambda_T - k_0\lambda_1}{\sum_{j=k_0+1}^{Q} \sqrt{w_{\tau(j)}}} \sqrt{w_{\tau(i)}} - \lambda_1 \right) I(i > k_0). \tag{3.35}$$

Substituting (3.35) into (3.30), the optimization problem is equivalent to finding λ_1^* minimizing

$$\lambda_1^* = \arg\min_{\lambda_1} \mathrm{E} \left\{ \sum_{i=1}^{Q} \frac{W_{\tau(i)}}{\lambda_1 + \left(\frac{\lambda_T - k_0\lambda_1}{\sum_{j=k_0+1}^{Q} \sqrt{W_{\tau(j)}}} \sqrt{W_{\tau(i)}} - \lambda_1 \right) I(i > k_0)} \right\} \tag{3.36}$$

$$= \arg\min_{\lambda_1} \mathrm{E} \left\{ \frac{1}{\lambda_1} \sum_{i=1}^{k_0} W_{\tau(i)} + \frac{1}{\lambda_T - k_0\lambda_1} \sum_{i=k_0+1}^{Q} \sum_{j=k_0+1}^{Q} \sqrt{W_{\tau(i)} W_{\tau(j)}} \right\}, \tag{3.37}$$

where if $k_0 = 0$ then the first summation in (3.37) equals zero. We can find λ_1^* via a one dimensional search. In summary, the optimal policy minimizing (3.30) is

Algorithm 3.1. Two stage Adaptive Resource Allocation Policy (ARAP) λ_A

Step 1: Allocate $\lambda_A(i,1) = \lambda_1^*$ to each cell and measure $\boldsymbol{y}(1)$.

Step 2: Given $\boldsymbol{y}(1)$ compute posteriors $P_{I_i|y_i(1)}$ defined in (3.29), then sort the W_i's defined in (3.31).

Step 3: Use λ_1^* and the ordered statistic $w_{\tau(i)}$ to find k_0 using (3.33) and (3.34).

Step 4: Given k_0, apply $\lambda(i,2)$ the energy allocation to cell i as

(3.38)

$$\lambda_A(\tau(i),2) = \lambda(\tau(i),2) = \left(\frac{\lambda_T - k_0\lambda_1^*}{\sum_{j=k_0+1}^{Q} \sqrt{w_{\tau(j)}}} \sqrt{w_{\tau(i)}} - \lambda_1^* \right) I(i > k_0),$$

and measure $\boldsymbol{y}(2)$.

Note that *ARAP* is a water-filling algorithm. This is a direct consequences of the facts that the cost function minimized in (3.30) is convex, $\lambda(i,t)$ is positive, and our constraint is linear of the form $\sum_{i,t} \lambda(i,t) = C$ (see [8] pp. 245).

3.3.2 Properties of the optimal energy allocation

Theorem 3.2. *For some $\nu \in [\frac{1}{2}, 1]$ and λ_1^*, let $\Lambda_A(i)$ be the search policy obtained using ARAP. Then, for a uniform prior distribution, we have*

$$J(\Lambda_A) \leqslant J(\Lambda_U), \tag{3.39}$$

with equality achieved if $P_{I_i|\boldsymbol{y}(1)} = c,\ \forall i$.

Proof. Note that an equal effort allocation policy can be broken into any arbitrarily

number T of consecutive search steps, as long as for all i

$$\Lambda_U(i) = \frac{1}{\lambda_T} \sum_{t=1}^{T} \lambda(i,t) = \frac{1}{Q}. \tag{3.40}$$

Without loss of generality let $T = 2$ and let Γ denote the family of all effort allocation policies with $\lambda(i,1) = \lambda_1$ (equal effort allocation) at the first stage

$$\Gamma = \left\{ \lambda(i,t) : \sum_{i=1}^{Q} \sum_{t=1}^{2} \lambda(i,t) = \lambda_T,\ \lambda(i,1) = \lambda_1 \geqslant 0,\ \lambda(i,2) \geqslant 0 \right\}. \tag{3.41}$$

Since $ARAP$ yields the optimal effort allocation for any set of posterior distribution $\{P_{I_i|\boldsymbol{y}(1)}\}_{i=1}^{Q}$, we have $J(\Lambda_A) \leqslant J(\Lambda_U)$. If $P_{I_i|\boldsymbol{y}(1)} = c$ for all i, then $w_i = c',\ \forall i$, for which $ARAP$ yields $k_0 = 0$. Thus, from (3.38) we obtain $\lambda_A(i,2) = \frac{\lambda_T}{Q} - \lambda_1$, or, equivalently, $\Lambda_A(i) = \frac{\lambda_T}{Q} = \Lambda_U(i)$. This completes the proof. □

$ARAP$ is optimal over all policies that allocate energy uniformly at the first step. In [5] we give the optimal ($ARAP$) strategy for a general case of a non-uniform prior. However, in this general case $\lambda^*(i,1)$ depends on the specific prior and is a function of i. Therefore, the optimization problem involves searching a $Q+1$ dimensional space and is computationally exhaustive.

Asymptotic (high SNR) properties of $ARAP$

Next we prove some asymptotic properties of $ARAP$, when $\nu = 1$. By asymptotic conditions we mean high SNR. We define SNR as the signal to noise ratio per cell for an equal energy allocation, i.e., SNR$=\frac{\lambda_T/Q}{\sigma^2}$. For $ARAP$, we show:

$$\mathrm{E}\{k_0\} \to (1-p)Q, \tag{3.42}$$

$$\lambda_1^* \to 0, \tag{3.43}$$

$$G(\Lambda_A) \to -10 \log p^*. \tag{3.44}$$

These asymptotic properties are proved under the condition that

$$p_{I_i|\boldsymbol{y}(1)} \to I_i, \tag{3.45}$$

in probability. In Appendix 3-F we prove the validity of (3.45) for the Gaussian case, where $y_i(1) \sim \mathcal{N}(\sqrt{\lambda_1}\theta_i I_i, \sigma^2)$, and we speculate that (3.45) holds for other cases as well. Hence, for high SNR, the set $\{p_{I_i|y_i}\}_{i=1}^{Q}$ can be approximated as, i.i.d., Bernoulli r.v.'s, with $\Pr(p_{I_i|y_i} = 1) = p$. In addition, $\nu = 1$ provides $w_i = p_{I_i|\boldsymbol{y}(1)}$ and therefore, for any realization, $\exists \tilde{k}$ such that the ordered posterior probabilities obey

$$p_{I_{\tau(i)}|y_{\tau(i)}} = \begin{cases} 0, & i \leqslant \tilde{k} \\ 1, & i > \tilde{k} \end{cases}. \tag{3.46}$$

Evaluating inequalities (3.34) for $k_0 = \tilde{k}$, yields

$$Q < \frac{\lambda_T}{\lambda_1} \leqslant \infty. \tag{3.47}$$

Hence, $\forall \lambda_1 > 0$, the desired solution of (3.34) is $k_0 = \tilde{k}$. Moreover, $\tilde{k} = Q - \sum_i I(p_{I_i|y_i} = 1)$ equals the total number of zeros in the sequence, and is a Binomial random variable distributed $\tilde{k} \sim B(Q, 1-p)$, thus

$$\mathrm{E}\{k_0\} = (1-p)Q. \tag{3.48}$$

Let $\widehat{\Psi} = \{i \in \mathcal{X} : \lambda(i,2) > 0\}$ denote the second stage allocation set. Note that together (3.45) and (3.42) suggests that $\widehat{\Psi}$ converges to Ψ or $|\Psi \triangle \widehat{\Psi}| \to 0$, where

'$\triangle$' is the symmetric set difference operator. To see that, note

$$i \in \Psi \Rightarrow p_{I_i|\boldsymbol{y}(1)} \to 1 \Rightarrow i \in \widehat{\Psi}, \tag{3.49}$$

thus $\Psi \subseteq \widehat{\Psi}$. At the same time

$$i \in \widehat{\Psi} \Rightarrow p_{I_i|\boldsymbol{y}(1)} = 1 \Rightarrow i \in \Psi, \tag{3.50}$$

and therefore $\widehat{\Psi} \subseteq \Psi$.

To prove (3.43), we substitute $\tilde{k}$ for k_0 and use the fact that $w_{\tau(i)} = p_{I_{\tau(i)}|y_{\tau(i)}} = 0$ for $\tau(i) \leqslant \tilde{k}$, thus (3.37) yields

$$\lambda_1^* = \arg\min_{\lambda_1} \mathrm{E}\left\{\frac{(Q-\tilde{k})^2}{\lambda_T - \tilde{k}\lambda_1}\right\}, \tag{3.51}$$

and since $\tilde{k} \in \{1, 2, \ldots, Q-1\}$, $\frac{(Q-\tilde{k})^2}{\lambda_T - \tilde{k}\lambda_1}$ is monotonically increasing in λ_1, regardless of the expectation operator. Thus, the minimizer w.r.t. λ_1 is achieved at the lowest possible value for λ_1. Since $\lambda_1 \in (0, \frac{\lambda_T}{Q})$, the best one can do is to allocate the minimum feasible positive energy value to all cells at the first step. Thus, most resources are 'saved' for the second step, which is depict in (3.44) that claims that Λ_A converges to Λ_o and is asymptotically optimal.

For large Q we have $p^* \to p$ in probability. This leads our optimal search policy to approach the lower bound for the cost function. Recall from (3.7) that the best possible search policy uniformly allocate the total among all cells in the ROI. Hence the expected minimal cost is $\mathrm{E}\{\frac{|\Psi|^2}{\lambda_T}\} = \frac{p^2Q^2 + p(1-p)Q}{\lambda_T}$, letting $\lambda_1 \to 0$ in our cost function yields

$$\lim_{SNR\to\infty} \mathrm{E}\left\{\frac{(Q-\tilde{k})^2}{\lambda_T - \tilde{k}\lambda_1}\right\} \cong \frac{\mathrm{E}\{(Q-\tilde{k})^2\}}{\lambda_T} = \frac{p^2Q^2 + p(1-p)Q}{\lambda_T}, \tag{3.52}$$

therefore, $G(\Lambda_A) \to -10 \log p$. Hence, in this Q the asymptotic gain is proportional to $\log \frac{1}{p}$ and we conclude that higher gains are achieved for sparser signals.

3.3.3 Suboptimal two stage search policy

Note that $ARAP$ requires sorting the w_i's and solving Q inequalities to find k_o. This requires an order $\mathcal{O}(\log Q)$ computations. As a simple alternative to $ARAP$, we consider a search policy where $\lambda(i,1) = \lambda_1$ and

$$\lambda(i,2) = \frac{\lambda_T - Q\lambda_1}{\sum_{j=1}^{Q} \sqrt{w_j}} \sqrt{w_i},$$

leading to a corresponding cumulative energy allocation Λ. Substituting Λ in (3.30) yields a single variable optimization problem and we grid search λ_1 to find λ_1^* minimizing the expected cost. Finally we define $\lambda_{so}(i,1) = \lambda_1^*$ and

$$\lambda_{so}(i,2) = \frac{\lambda_T - Q\lambda_1^*}{\sum_{j=1}^{Q} \sqrt{w_j}} \sqrt{w_i}, \tag{3.53}$$

with its equivalent cumulative energy allocation Λ_{so}. The simple allocation policy (3.53) is optimal, i.e. minimizing (3.30) when $\nu = 1$, for two extreme cases: (i) uniform posterior distribution, (ii) posterior distribution vector $[p_{I_1|\boldsymbol{y}(1)}, \ldots, p_{I_Q|\boldsymbol{y}(1)}]$ with L elements for which $p_{I_i|\boldsymbol{y}(1)} = 1$ and $(Q-L)$ elements for which $p_{I_i|\boldsymbol{y}(1)} = 0$. For (i) we get an equal energy allocation, while for (ii), (3.53) results in

$$\lambda(i,2) = \begin{cases} \frac{\lambda_T - Q\lambda_1}{L}, & p_{I_i|\boldsymbol{y}(1)} = 1 \\ \\ 0, & \text{Otherwise} \end{cases}, \tag{3.54}$$

both equivalent to the optimal mapping (3.38). Although (3.53) does not make the analytical evaluation of the expectation in (3.25) tractable, it is less computationally

demanding than the optimal solution. In fact, since λ_1^* is a function of SNR and the sparsity p it can be computed offline. Thus, λ_{so} is a direct mapping from the observation space to the search space. In Section 3.4 we compare the two policies and show that λ_{so} nearly optimizes (3.25).

3.4 Comparison of search algorithms

Assume either λ_A or λ_{so} were used to generate data vectors $(\boldsymbol{y}(1), \boldsymbol{y}(2))$. A natural question is whether or not this data is better in some sense than the measurement vector $\boldsymbol{y}$ obtained using the standard exhaustive search policy. In this section, we compare performance of both the optimal and suboptimal effort allocation policies to those achieved by exhaustive search. We start by showing performance gains in both estimation and detection due to our adaptive measuring schemes. Next, we compare the performance (3.25) achieved by λ_A and λ_{so} to show that λ_{so} is nearly optimal. In the following section we assume $\theta_i(t) = \theta_i \geqslant 0$ are independent and identically distributed (i.i.d.) truncated Gaussian random variables with mean $\mu_\theta = 1$ and standard deviation $\sigma_\theta = 0.25$, for all i.

3.4.1 Estimation post-processing

Consider the problem of estimating the true value of a target return θ_i from $y(t)$ given by

$$y_i(t) = \sqrt{\lambda(i,t)}\theta_i I_i + n_i(t), \tag{3.55}$$

where $n_i(t) \sim \mathcal{N}(0, \sigma^2)$, $t = 0$ represent an exhaustive search with $\lambda(i,0) = \frac{\lambda_T}{Q} \stackrel{\triangle}{=} \lambda_0$, and $t = 1, 2$ are due to either the optimal or suboptimal measurement policies λ_A or λ_{so}. For estimation the choice of $\nu = 1$ seems natural. Recall that in Section 3.2

we claimed that minimizing our cost is entirely equivalent, in the Gaussian case, to minimizing the estimation mean squared error (MSE). Assuming $\theta_i \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$ we use a Bayesian framework for estimating θ_i based on its prior distribution. The optimal estimator minimizing the MSE is the conditional mean estimator (CME). We compare the performance of the CME $\mathrm{E}\{\theta_i|\boldsymbol{y}(0)\}$ for an exhaustive search policy to the CME $\mathrm{E}\{\theta_i|\boldsymbol{y}(1), y_i(2)\}$ for either $ARAP$ or the suboptimal search policy. The MSE of the CME for the exhaustive search policy is given by

$$\mathrm{var}\{\theta_i|y_i(0)\} = \sigma_\theta^2 - \frac{\lambda_0\sigma_\theta^4}{\sigma^2 + \lambda_0\sigma_\theta^2} = \frac{\sigma_\theta^2}{1 + \lambda_0\frac{\sigma_\theta^2}{\sigma^2}}. \tag{3.56}$$

The competing estimator is a Naive Bayes estimator [62] of $\mathrm{E}\{\theta|\boldsymbol{y}(1), y_i(2)\}$, which is derived under the assumption that $(\boldsymbol{y}(1), y_i(2))$ are independent. The Naive Bayes estimator is

$$\hat{\theta}_i = \mu_\theta + \frac{y_i(1)\sqrt{\lambda_1} + y_i(2)\sqrt{\lambda(i,2)} - \mu_\theta(\lambda_1 + \lambda(i,2))}{\lambda_1 + \lambda(i,2) + \sigma^2/\sigma_\theta^2}. \tag{3.57}$$

Monte-Carlo simulations were used to estimate the MSE of (3.57). In Figure 3.1, we plot the MSE performance gain $g(\lambda)$, defined as

$$g(\lambda) = 10\log\frac{\mathrm{var}(\theta_i|y_i(0))}{\mathrm{MSE}(\hat{\theta}_i)} \tag{3.58}$$

as a function of SNR, where Monte-Carlo simulations were used to estimate the MSE of (3.57). We chose $Q = 4096$, $p \in \{\frac{1}{10}, \frac{1}{100}\}$, and each point on the figure represents an average over θ based on 2000 realizations. Curves with crosses and circles represents the optimal policy ($ARAP$) λ_A and the suboptimal policy λ_{so} respectively. Note that $ARAP$ yield better estimation performance compared to the suboptimal policy. The high gains in Figure 3.1 reinforce the connection between our cost function (3.5) and estimation MSE.

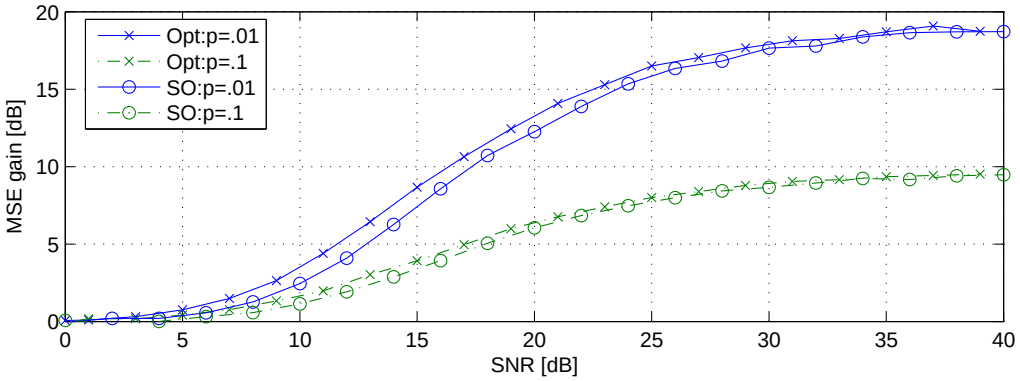

Figure 3.1: Gain in MSE for the CME in (3.57) based on an adaptive search compared to the MSE of the CME for an exhaustive search policy (3.56). Curves with crosses correspond to $ARAP$, for p values of $\frac{1}{100}$ and $\frac{1}{10}$, while curves with circles represent the suboptimal adaptive policy. The MSE gain for $ARAP$ is slightly higher than that of the suboptimal mapping. Note that using our methods results in about 6 [dB] gain in MSE at SNR value of 12 [dB] for sparsity level of 1%. In addition MSE gain is inversely proportional to the sparsity, hence higher gains can be expected for application where $|\Psi| \ll Q$.

3.4.2 Detection post-processing

Non-adaptive detection: Consider the problem of correctly detecting whether cell i contains a target based on a sample $y_i(0)$. As before, we assume the samples follow the model (3.55) and that $\theta_i \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$ are i.i.d.. Thus, for an exhaustive search policy $y_i(0) \sim \mathcal{N}(\sqrt{\lambda_0}\mu_\theta I_i, \sigma_{y_r}^2)$, where $\sigma_y^2 = \sigma^2 + \lambda_0\sigma_\theta^2$. Given $y_i(0)$ the measurement of pixel i, our goal is to decide between

$$\begin{aligned} \mathcal{H}_0 &: y_i \sim \mathcal{N}(0, \sigma^2), \\ \mathcal{H}_1 &: y_i \sim \mathcal{N}(\sqrt{\lambda_0}\mu_\theta, \sigma^2 + \lambda_0\sigma_\theta^2). \end{aligned} \tag{3.59}$$

For a known σ_y^2, the uniformly most powerful test for this simple binary hypothesis testing problem is a likelihood ratio test (LRT). The performance of this non-adaptive LRT in terms of its receiver operating characteristic (ROC) curve is easily calculated.

The power β of this level α LRT is [32]

$$\beta = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \sqrt{\frac{\lambda_0 \mu_\theta^2}{\sigma^2}}\right), \tag{3.60}$$

where $\Phi(\cdot)$ is the normal cumulative distribution function.

Adaptive detection: First we derive the likelihood function $f(\boldsymbol{y}(2), \boldsymbol{y}(1); \boldsymbol{I}_\Psi)$. Bayes rule provides

$$f(\boldsymbol{y}(2), \boldsymbol{y}(1)) = f(\boldsymbol{y}(2)|\boldsymbol{y}(1)) f(\boldsymbol{y}(1)), \tag{3.61}$$

but, given $\boldsymbol{y}(1)$, the measurements at the second step are independent for different cells and thus

$$f(\boldsymbol{y}(2), \boldsymbol{y}(1)) = \prod_{i=1}^{Q} f(y_i(2)|\boldsymbol{y}(1)) f(y_i(1)). \tag{3.62}$$

Therefore, the LRT statistic T_j is

$$T_j = \frac{\prod_{i=1}^{Q} f(y_i(2)|\boldsymbol{y}(1), I_j = 1) f(y_i(1)|I_j = 1)}{\prod_{i=1}^{Q} f(y_i(2)|\boldsymbol{y}(1), I_j = 0) f(y_i(1)|I_j = 0)} = \tag{3.63}$$

$$\frac{f(y_j(2)|\boldsymbol{y}(1), I_j = 1) f(y_j(1)|I_j = 1)}{f(y_j(2)|\boldsymbol{y}(1), I_j = 0) f(y_j(1)|I_j = 0)}. \tag{3.64}$$

For our model, we have $y_i(1) \sim \mathcal{N}(\sqrt{\lambda_1}\mu_\theta I_i, \sigma^2 + \lambda_1 \sigma_\theta^2 I_i)$ and, given $\boldsymbol{y}(1)$, the second step measurements are also Gaussian $y_i(2) \sim \mathcal{N}(\sqrt{\lambda(i,2)}\mu_\theta I_i, \sigma^2 + \lambda(i,2)\sigma_\theta^2 I_i)$. Substituting these distributions into (3.64) provides the following LRT

(3.65)

$$T_i = \frac{\sigma^2}{\sqrt{(\sigma^2 + \lambda_1 \sigma_\theta^2)(\sigma^2 + \lambda(i,2)\sigma_\theta^2)}} \frac{\exp\left\{-\frac{\left(y_i(2) - \mu_\theta\sqrt{\lambda(i,2)}\right)^2}{2(\sigma^2 + \lambda(i,2)\sigma_\theta^2)} - \frac{\left(y_i(1) - \mu_\theta\sqrt{\lambda_1}\right)^2}{2(\sigma^2 + \lambda_1\sigma_\theta^2)}\right\}}{\exp\left\{-\frac{y_i^2(1)}{2\sigma^2} - \frac{y_i^2(2)}{2\sigma^2}\right\}} \gtrless \gamma.$$

Note that for all cells where $\lambda(i,2) = 0$ this test is a function of $y_i(1)$ alone.

Next, we compare the theoretical ROC curve (3.60) to the empirical ROC curve calculated for the adaptive LRT (3.65) performed on the data pair $(\boldsymbol{y}(1), \boldsymbol{y}(2))$ using the optimal λ_A and the suboptimal λ_{so}. We conducted multiple runs for varying SNR levels and observed that with $\nu = 1$ (3.65) provides higher detection probability than the non-adaptive LRT having performance (3.60) for false alarm levels lower than 30%. At SNR values close to 0 [dB] the difference between the two tests is negligible, but increases with SNR. Note that, for very low false alarm levels λ_A performs better than λ_{so} in terms of detection probability. However, for higher test levels the suboptimal search policy λ_{so} yield better detection performances.

Results are presented in Figure 3.2 (a) and (b) for SNR=10 [dB], $Q = 1024$, and either $p = 0.1$ or $p = 0.01$. Monte-Carlo simulations were conducted with $\nu = 1$ and each point on the figures represents an average over 2000 realizations. Detection probability was averaged over the entire ensemble and over all pixels inside and outside the ROI. At 10 [dB] the ROC curves are very sharp and are plotted on a logarithmic scale. The solid curve represents (3.60) the non-adaptive LRT with equal energy allocation. Curves with crosses represents λ_A, while curves with circles represents λ_{so}. It is evident that the ROC curves of different tests have different slopes for low false alarm values. Moreover, for high false alarm values no adaptive policy outperforms the exhaustive search policy. Figure 3.2(b) zooms in to better illustrate the differences for $P_{fa} \in [0.005, 0.5]$. One can see that the optimal search policy has the best performances up to $\alpha = 3\%$, at which point the suboptimal policy yields higher detection probability. The exhaustive search policy outperforms both the adaptive methods for $\alpha > 30\%$.

Finally we compare detection probability values, for a fixed false alarm rate, and

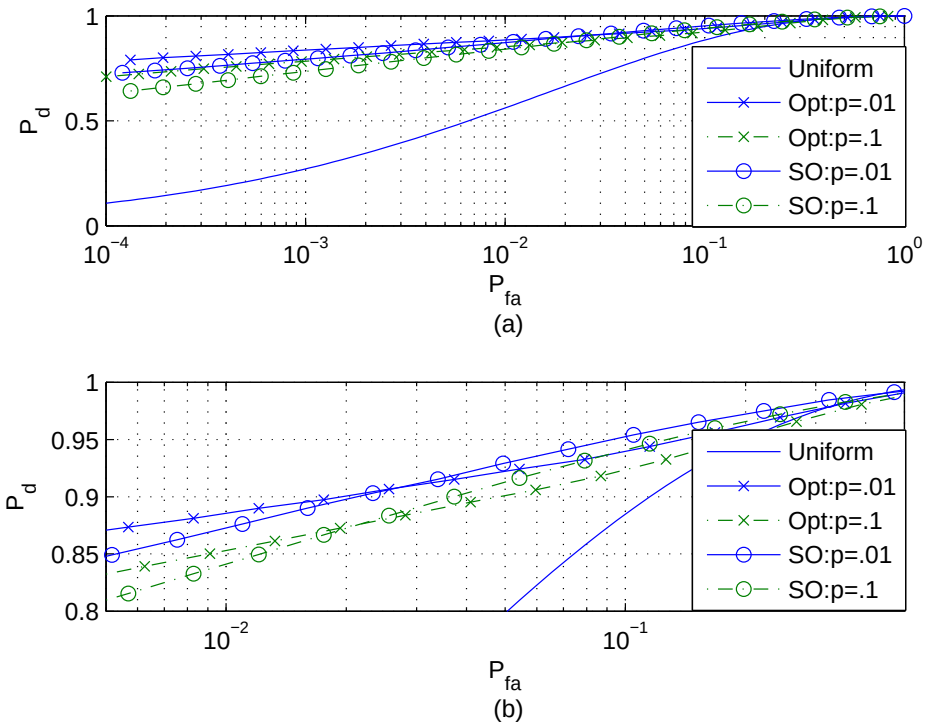

Figure 3.2: ROC curves for the LRT tests based on an exhaustive search scheme and the two adaptive policies measurements scheme, for $p = 0.1$ and $p = 0.01$ and SNR of 10 [dB]. (a) shows the entire ROC curve while (b) zooms in on false alarm probability values less than 0.5. The simulation suggests that our adaptive search policies outperforms an exhaustive search policy in terms of detection probability for false alarm values lower than 30%.

estimation MSE gain $g(\cdot)$ in (3.58) as a function of ν. Results are shown in Figure 3.3. The curve with triangle markers represents estimation MSE gain and its corresponding values are indexed on the right axis of the figure. The other curves represent detection probability for a given test level, with the detection values indexed at the left axis. All curves are a function of ν. For the selected operating point it is clear that it is best to choose $\nu = 1$, since it maximizes both detection and estimation performance. In [5] we compare (3.65) to a test using a detection optimized measurement and show that λ_A with $\nu < 1$ is nearly optimal for detection.

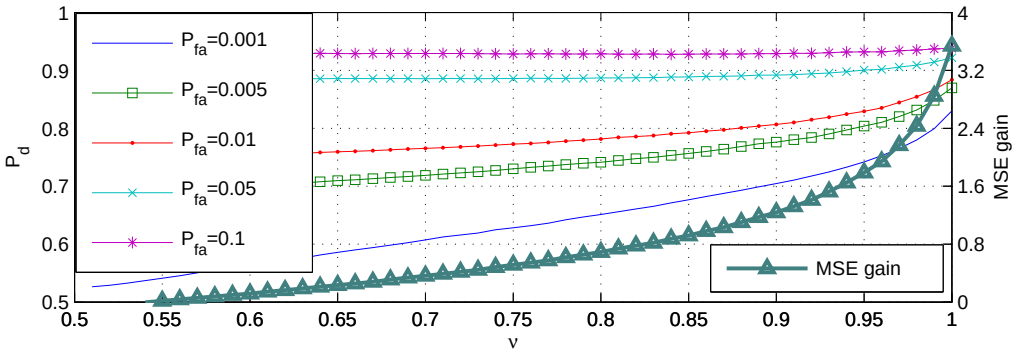

Figure 3.3: Detection probability, for a fixed test level, and estimation MSE gain $g(\cdot)$ in (3.58) as a function of ν when SNR is 10 [dB] and $p = 0.01$. Note that the MSE gain values (curve with triangular markers) are given on the r.h.s. of the figure. Since MSE gain is defined over the true ROI it increases with ν.

3.4.3 Achievable cost

As a final comparison between our two policies, we compare the average performance gain $G(\cdot)$, defined as the expected value of (3.15), achieved by the two search policies for $\nu = 1$. We chose $Q = 8192$ as the total number of cells and the sparsity values of $p \in \{\frac{1}{10}, \frac{1}{100}, \frac{1}{1000}\}$, i.e., a mean of roughly 800, 80, and 8 targets per realization, respectively. Results are shown in Figures 3.4(a) and 3.4(b) where the curves with crosses and circles describe the expected gain $\mathrm{E}\{G(\Lambda_A)\}$ with optimal and with suboptimal $\mathrm{E}\{G(\Lambda_{so})\}$ allocation, respectively. Figure 3.4(a) shows the behavior of the gain for the two policies for SNR values of 0 [dB] to 40 [dB]. Figure 3.4(b) zooms in on SNR values of 0 [dB] to 13 [dB]. Each point on a graph represents 500 runs in a Monte Carlo simulation. As can be seen from Figure 3.4(a), at extreme high or low SNR values the performance gain of the two policies coincide. Figure 3.4(a) indicates that the gain converges to its theoretical limit given by (3.44). The largest gap in gain between the two algorithms is near the transition zone: SNR between 5-15 [dB], and the gap is less than 2 [dB]. Evidently, the simpler suboptimal mapping rule does not significantly degrade performance gain.

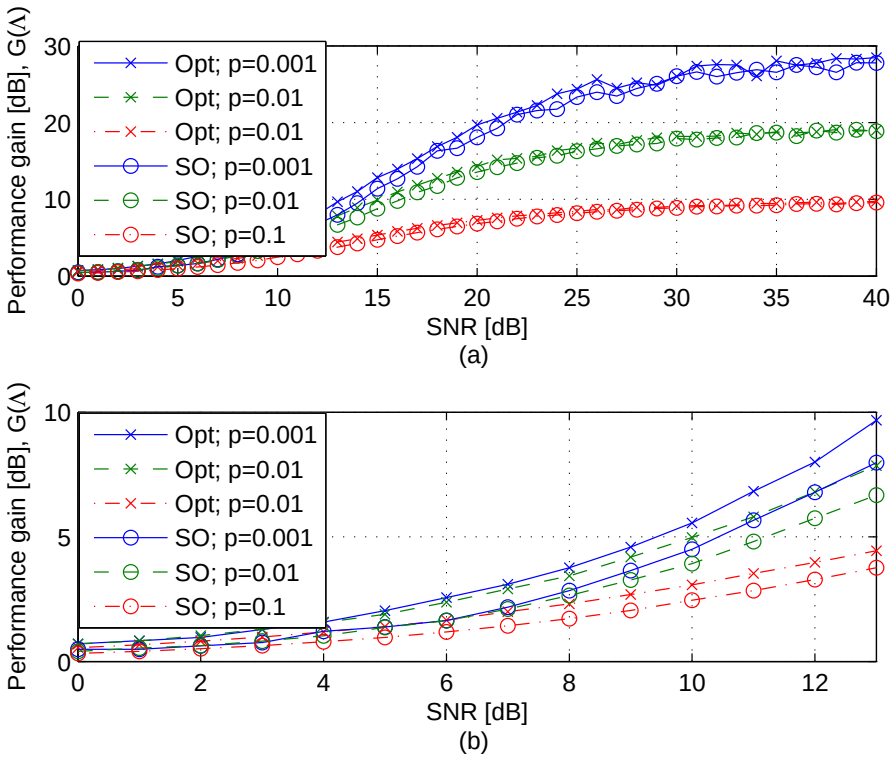

Figure 3.4: The cost gain compared to an exhaustive search for both our optimal and suboptimal energy allocation schemes. (a) shows that both algorithms converges to the asymptotic predicted gain, at $-10\log p$. (b) enhances the difference between our two policies for SNR values in the range of $0-13$ [dB].

Figure 3.5 compares the percentage of effort $\frac{\lambda_1 Q}{\lambda_T}$ allocated in the first step for both policies. While for SNR values greater than 25 [dB] the curves overlap, this is not the case for low SNR values. As measurement quality decreases, $ARAP$ invests more energy at the first step. Considering the difference between the two policies, this result makes sense: after the first step $ARAP$ ignores all cells with posterior probability values lower than some threshold. Decision errors at that stage can no longer be compensated, i.e., the cumulative effort distribution $\Lambda_A(i)$ for those cells will remain unchanged. Hence, more effort has to be allocated to the first step to reduce decision errors, i.e., improve the agreement between $\widehat{\Psi}$ and Ψ. On the other hand, the suboptimal mapping invests energy in all Q cells at the second step. Thus,

it has a chance to compensate for poor estimated posterior probability values. As a general remark we note that the incentive to use our methods increases with SNR (e.g. Figure 3.1). Therefore, differences between the two policies in the low SNR regime are of lesser importance.

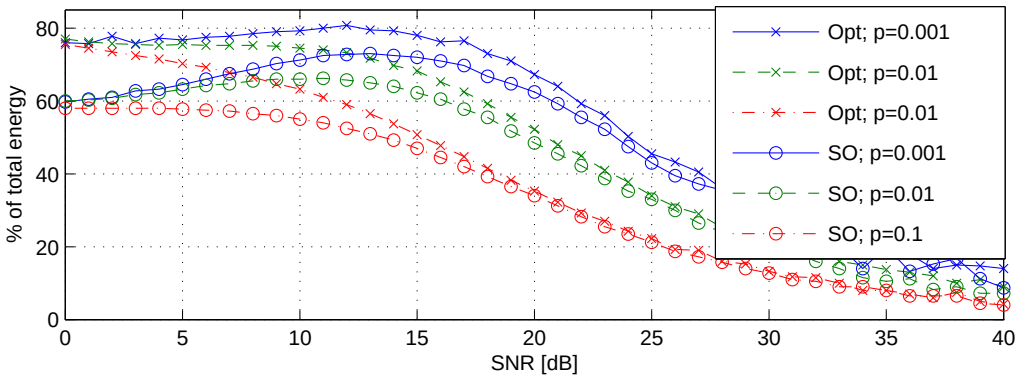

Figure 3.5: The proportion of energy invested at the first step for the two algorithms λ_A and λ_{so}. Curves correspond to prior probability values of $0.001, 0.01$ and 0.1. As seen, the optimal search policy invest more energy at the first step. However, for SNR > 25 [dB] the two are essentially equivalent.

3.5 Application - detecting and estimating a ROI in a SAR image

Consider the task of imaging a large area using a satellite equipped with a synthetic aperture radar (SAR) system. Assume we have satellite access at two different occasions for a limited time. In SAR imaging the measurement quality improves as the dwell time increases. Therefore, with limited access to satellite time we face the question of how to best utilize the resource available to us, i.e., where to dwell and for how long to dwell in the SAR system. Problems concerning where to point a sensor to are discussed in [31,33,34]. However, by allowing an additional continuous degree of freedom for the varying dwell time we consider a different set of feasible policies. Assume the satellite has enough time to acquire all cells of the image at each stage, if so desired. Assuming our goal is to detect and identify targets spread out in a

large area we propose the following two stage search policy based on *ARAP*: The first stage performs an exhaustive search with equal dwell time over all cells yielding a preliminary image $\boldsymbol{y}(1)$ and an estimate $\widehat{\Psi}$ of the ROI Ψ. In the second stage, let *ARAP* allocate the dwell time in a non uniform manner to the different cells via (3.38), i.e., focusing the search on $\widehat{\Psi}$. Finally, we combine both measurements to form an image on which we detect the ROI and estimate its content. The competing strategy performs an exhaustive search with equal dwell time at each cell twice, with no adaptation between scans, than uses the arithmetic mean of the two independent scans as the estimated image.

We used a SAR image, taken from the Sandia National Laboratories website[2] as an example of a "sparse" image. The image displays two columns of tanks in a field and its sparsity ratio is $p < 0.01$. Let X denote the original image and let $\boldsymbol{x}$ be a lexicographic ordering of X. We emulated the effect of the SAR varying dwell time as modulated speckle noise variance in the post-processed SAR image. In particular, the SAR image after the first stage (equal dwell time at all pixels) is modeled as:

$$\boldsymbol{x}_1 = (1 + z_1)\boldsymbol{x}, \tag{3.66}$$

where z_1 is a zero mean uniform random variable with $\mathrm{var}(z_1) = \frac{1}{\lambda_1}$ noted as $z_1 \sim \mathrm{U}[-\sqrt{\frac{3}{\lambda_1}}, \sqrt{\frac{3}{\lambda_1}}]$ and $\lambda_1 = \lambda_0$. A tank template shown in Figure 3.6, was applied as a matched filter to the noisy image X_1 yielding $\widetilde{X}_1$. The input to the *ARAP* algorithm $\boldsymbol{y}(1) = \frac{\widetilde{\boldsymbol{x}}_1 - \overline{x}_1}{\sqrt{\frac{1}{Q}(\widetilde{\boldsymbol{x}}_1 - \overline{x}_1)'(\widetilde{\boldsymbol{x}}_1 - \overline{x}_1)}}$ was the variance normalized version of $\widetilde{X}_1$, where $\overline{x}_1 = \frac{1}{Q}\sum_{i=1}^{Q} \widetilde{\boldsymbol{x}}_1(i)$. *ARAP* was used to obtain a search policy for the second step $\lambda(i, 2)$ via (3.38). All indices i with $\lambda(i, 2) < \lambda_1$ were set to zero and their cumulative search effort was redistributed among the rest of the cells in a proportional manner.

[2] `http://www.sandia.gov/RADAR/images/rtv_tanks_9in.jpg`

The SAR return $\boldsymbol{x}_2$ was modeled as

$$\boldsymbol{x}_2 = (\boldsymbol{x} + \boldsymbol{x} \odot \boldsymbol{z}_2) \odot \boldsymbol{I}_{\widehat{\Psi}}, \tag{3.67}$$

where $z_2(i) \sim \mathrm{U}[0, \sqrt{\frac{12}{\lambda(i,2)}}]$ for all $i \in \widehat{\Psi}$ and zero otherwise, and $\odot$ denotes an element by element multiplication operator. This resulted in a non-uniform variance for different cells in $\widehat{\Psi}$ with $\mathrm{var}(z_2(i)) = \frac{1}{\lambda(i,2)}$. Note that $\boldsymbol{x}_2 = 0$ for all $i \in \widehat{\Psi}^c$.

We considered several suboptimal linear estimators of $\boldsymbol{x}$ based on $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ that performed comparably. The estimator presented here $\widehat{X}$ was defined via $\widehat{\boldsymbol{x}}$ as

$$\widehat{\boldsymbol{x}} = \left(\frac{\boldsymbol{x}_1 \odot I\{i \in \widehat{\Psi}^c\}}{\sqrt{1 + 1/\lambda_1}} + \boldsymbol{x}_2 \odot \mathrm{vec}\left\{ \frac{1}{\sqrt{1 + 1/\lambda(i,2)}} \right\} \right) \odot \boldsymbol{s}, \tag{3.68}$$

where $\mathrm{vec}\{r_i\} = [r_1, r_2, \ldots, r_Q]'$ and $\boldsymbol{s}(i) = \left(\sqrt{1 + \frac{1}{\lambda_1}} + \sqrt{1 + \frac{I\{i \in \widehat{\Psi}\}}{\lambda(i,2)}} \right)^{-1}$. The

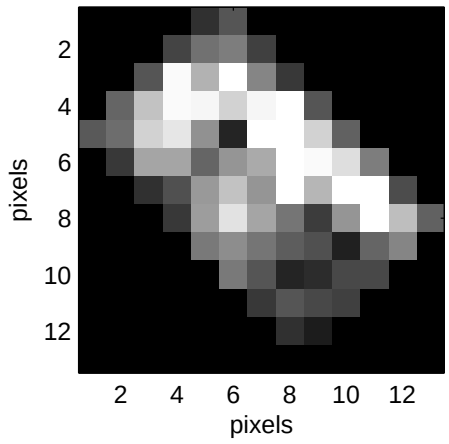

Figure 3.6: The above (13×13) tank template was used as a matched filter to filter the noisy data X_1 and generate $\boldsymbol{y}(1)$.

estimator (3.68) is compared to an image reconstructed from two exhaustive searches with equal effort allocation $(\boldsymbol{x}_{U1}, \boldsymbol{x}_{U2})$ given in (3.66), defined as $\boldsymbol{x}_U = (\boldsymbol{x}_{U1} + \boldsymbol{x}_{U2})/2$. SNR was defined as $10 \log 2\lambda_0$.

Results are presented in Figures 3.7 and 3.8 for SNR values of 4 and 0 [dB], respectively. Figure 3.7(a) show the original image, Figure 3.7(b) and (d) show a single realization of images reconstructed using exhaustive search and $ARAP$ via (3.68), respectively. Figure 3.7(c) show the effort allocated by $ARAP$ at the second stage for that specific realization. Although all targets are identifiable in Figure 3.7(b), they appear clearly in Figure 3.7(d). Figure 3.8 focuses on the ROI to demonstrate the superiority of $ARAP$ compared to the exhaustive search policy. Figures 3.8(a) and (b) show a single realization of the two search methods exhaustive and $ARAP$ at 0 [dB] respectively, while Figs. 3.8(c) and (d) display a 1D profile, going through the left columns of tanks, of 100 different realizations of each policy respectively. It is evident that variations in profiles in images reconstructed using $ARAP$ are much smaller than those in images reconstructed from exhaustive search policy.

This illustrative example demonstrates the potential utility of our method in SAR imaging. Note that energy allocation is equivalent to dwell time allocation in this example.

3.6 Conclusions and future work

We introduced a novel convex cost function and showed that its minimization relates to minimizing error probability or estimation MSE over an unknown ROI. A closed form solution for the second stage in a two-stage optimal search policy was provided, and numeric search for the first step was presented. A closed form low complexity approximation for the two step minimization problem was also presented and it was shown to perform comparably to the optimal solution. In a high SNR the performance of the optimal and approximated algorithms was shown to converge to the ideal omniscient limit. For the detection task, the two search policies introduced

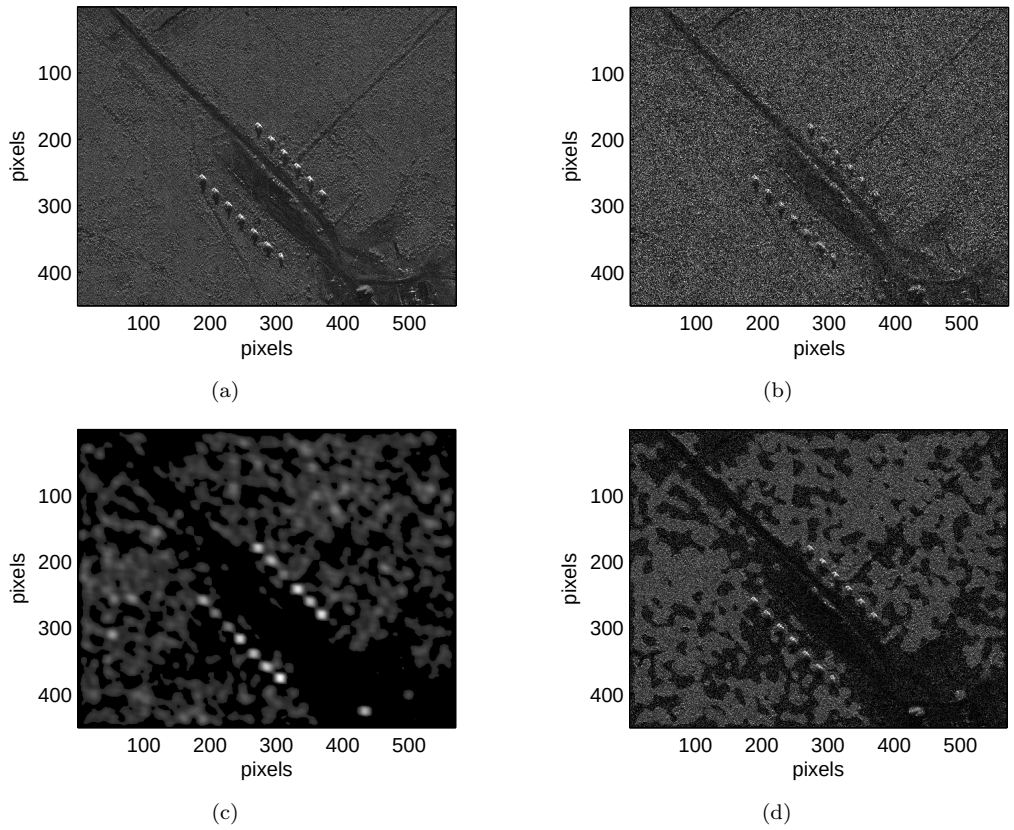

Figure 3.7: SAR imaging example, SNR=4 [dB]. (a) Original image. (b) Image reconstructed using two exhaustive searches. (c) Effort allocation using $ARAP$ at the second stage. (d) Image resulted from (3.68) using $ARAP$.

outperformed the one step exhaustive measurement scheme when the false alarm is less than 30%. For estimation, comparing the MSE of estimated values within the ROI, our adaptive search policies dominate the exhaustive search policy. The search policy is parameterized by ν which varies from $\frac{1}{2}$ to 1 and controls the energy allocated within the ROI. An offline lookup table can be generated for the optimal ν in terms of the sparseness p and SNR of the target in the data. Finally, an illustrative example of our method for SAR imaging was presented.

This approach is applicable to tumor detection where a cluster of calcification may appear around the lesion. In this case multiscale hypothesis testing methods presented in [21] may be relevant. Frakt et al. deals with anomaly detection once

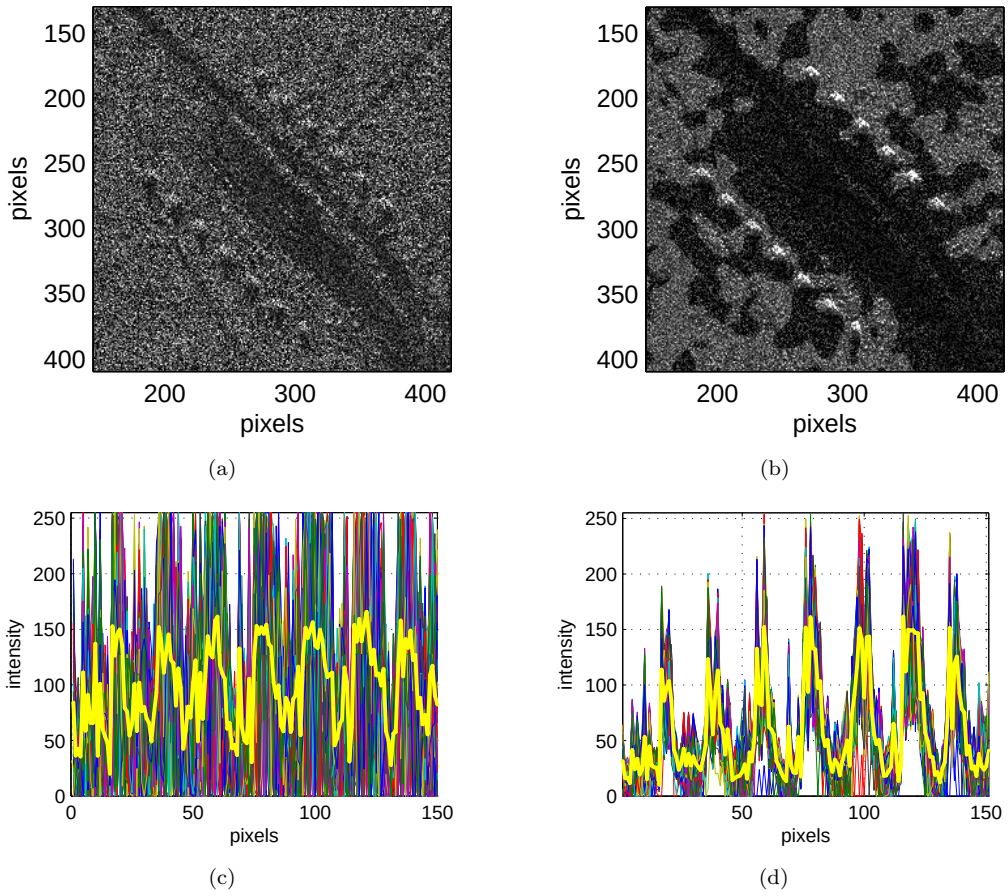

Figure 3.8: SAR imaging example, SNR=0 [dB]. (a) Image reconstructed using two exhaustive searches, targets are not easily identifiable. (b) Image resulted from (3.68) using *ARAP*. Figures (c) and (d) compare a 1D profile going through the targets on the lower left column for 100 different realizations. (c) Profiles of images reconstructed from an exhaustive search. (d) Profiles of images reconstructed using *ARAP*. The bold bright line on both figures represent the mean profile of the different realizations. Evidently, variations of profiles of images due to *ARAP* are much smaller compared to variations of profiles of images resulted from an exhaustive scan.

measurements, at a fine resolution, have been acquired. Our goal is to generate fine resolution measurements only where the are most informative. Another interesting area of application is to compressive sensing. Work such as [56] and [55] consider the problem of sampling a sparse medium via an arbitrary affine transformation. In cases of sparse signals complete reconstruction of the underlying signal can be accomplished with only a few samples. This exploitation of sparsity is analogous to the methods proposed in this chapter.

3-A Appendix: Equivalency of minimizing our cost to minimizing the *Cramér-Rao* bound

Consider the problem of estimating an unknown signal in the presence of a Gaussian noise. Let $\boldsymbol{y}_{m\times 1}$ be a measurements vector given by

$$\boldsymbol{y} = \text{diag}\{\sqrt{\boldsymbol{\lambda}}\}\boldsymbol{\mu} + \boldsymbol{n}, \tag{3-A.1}$$

where $\boldsymbol{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\text{I})$. Let Ψ be a collection of indices, namely the ROI, with $|\Psi| \ll m$. Assume that the effort to be distributed is the measurement energy λ_i associated with measuring an element of $\boldsymbol{y}$, with $\sum_i \lambda_i = 1$. Define $z_i = \frac{y_i}{\sqrt{\lambda_i}}$, hence

$$z_i \sim \mathcal{N}\left(\mu_i, \frac{\sigma^2}{\lambda_i}\right). \tag{3-A.2}$$

Recall that the *Cramér-Rao* bound on the variance of any unbiased estimator $\hat{\mu}_i(z_i)$ of μ_i is given as

$$\text{VAR}(\hat{\mu}_i) \geqslant \frac{\sigma^2}{\lambda_i}, \tag{3-A.3}$$

which is also the measurement variance for this case. From (3.7) we know that the minimizer of our cost (3.5) when $\nu = 1$ is a uniform energy allocation within the ROI. Hence the optimal solution minimizing our cost $\lambda_i = \frac{1}{|\Psi|} I(i \in \Psi)$ also uniformly minimizes the *Cramér-Rao* bound on all $\hat{\mu}_i$ where $i \in \Psi$.

3-B Appendix: Chernoff bound on the probability of error and our cost function

Let us consider simple binary, hypothesis testing problem of classifying an observation X as coming from one of two possible classes (hypotheses) $\mathcal{H}_0, \mathcal{H}_1$. Let π_0, π_1 denote the a priori probabilities on the hypotheses, and let $p_0(x), p_1(x)$ denote

the conditional probability density functions given the true hypothesis. Then, Bayes decision rule minimizes the probability of error P_e by choosing the hypothesis with the largest a posteriori probability. The Chernoff bound on P_e states that for any $\alpha \in [0,1]$

$$P_e \leqslant \pi_0^{\alpha}\pi_1^{1-\alpha}\int_X [p_0(x)]^{\alpha}\,[p_1(x)]^{1-\alpha}\,dx. \tag{3-B.1}$$

Further let X be a Gaussian r.v. with known variance σ^2 and unknown mean with the two hypothesis corresponds to $\{\mathcal{H}_0 : \mu_0 = 0\}$ and $\{\mathcal{H}_1 : \mu_1 > 0\}$. In this case (3-B.1) yields

$$P_e \leqslant \pi_0^{\alpha}\pi_1^{1-\alpha}e^{-\frac{\alpha(1-\alpha)\mu_1^2}{2\sigma^2}}, \tag{3-B.2}$$

so the Chernoff bound on the probability of error P_e is a decreasing function of the unknown mean μ and an increasing function of the variance σ^2. Let $\boldsymbol{w}_{Q\times 1} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathrm{I})$ and consider the problem of detecting between $\mu_i = 0$ and $\mu_i > 0$. Let $\Psi = \{i : \mu_i > 0\}$ and assume $|\Psi| \ll Q$. Further assume that by allocating energy $\lambda(i)$ to the i^{th} measurement we increase the *detectability index* $\frac{\mu_i}{\sigma}$, e.g. as in (3.18). For such a scenario, minimizing our cost function (3.5) with $\nu = 1$ results in uniformly allocating energy to all cells in Ψ. Utilizing (3-B.2) it implies uniformly minimizing P_e over Ψ.

3-C Appendix: Showing a global lower bound on our cost

In Section 3.2 we claimed that the r.h.s. of (3.6) is lower bounded by $|\Psi|^2$ for all $\nu \in [\frac{1}{2}, 1]$ and $\frac{|\Psi|}{Q} \leqslant \frac{1}{2}$. To prove this claim let $f(\nu) = [\sqrt{\nu}|\Psi| + \sqrt{1-\nu}(Q - |\Psi|)]^2$, then we need to show

$$f(\nu) \geqslant |\Psi|^2. \tag{3-C.1}$$

First note that

$$f(\frac{1}{2}) = \frac{Q^2}{2} > |\Psi|^2 = f(1), \tag{3-C.2}$$

since $\frac{|\Psi|}{Q} \leqslant \frac{1}{2}$. Next we show that $\dot{f}(\nu) \leqslant 0$ for all $\nu \in [\frac{1}{2}, 1]$, i.e., $f(\nu)$ is a decreasing function over the interval of interest. When combined with (3-C.2) this proves the claim. Note that

$$\dot{f}(\nu) = 2\left[\sqrt{\nu}|\Psi| + \sqrt{1-\nu}(Q - |\Psi|)\right]\left[\frac{|\Psi|}{\sqrt{\nu}} - \frac{Q-|\Psi|}{\sqrt{1-\nu}}\right], \tag{3-C.3}$$

hence it suffice to show

$$\left[\frac{|\Psi|}{\sqrt{\nu}} - \frac{Q-|\Psi|}{\sqrt{1-\nu}}\right] \leqslant 0.. \tag{3-C.4}$$

Rearranging (3-C.4) we need to show

$$\sqrt{\frac{1-\nu}{\nu}} \leqslant \frac{Q-|\Psi|}{|\Psi|} = \frac{Q}{|\Psi|} - 1. \tag{3-C.5}$$

But $\nu \in [\frac{1}{2}, 1]$ provides $\frac{1-\nu}{\nu} \leqslant 1$, while $\frac{|\Psi|}{Q} \leqslant \frac{1}{2}$ results in $\frac{Q}{|\Psi|} - 1 \geqslant 1$ and hence the inequality in (3-C.5) holds. This completes the proof.

3-D Appendix: Showing an upper bound on the gain

Let I_i be a Bernoulli r.v. with $\Pr(I_i = 1) = p$. Define $|\Psi| = \sum_{i=1}^{Q} I_i$, then $|\Psi|$ is a Binomial r.v. and $|\Psi| \sim \mathrm{B}(Q, p)$. Further let $p^* = \frac{|\Psi|}{Q}$ then $\mathrm{E}\{p^*\} = p$ and $\mathrm{var}(p^*) = \frac{p(1-p)}{Q}$.

Claim 1. *For an arbitrary effort allocation policy Λ and $\delta \geqslant \exp\{-Q\frac{3p}{8(1-p)}\}$ we have*

$$\Pr(G(\Lambda) \leqslant -10\log p + \epsilon) \geqslant 1 - \delta, \tag{3-D.1}$$

where $\epsilon = -10 \log \epsilon'$ *and* $\epsilon' < 1$ *is the solution of*

$$(1-\epsilon')^2 + (4-\epsilon')\frac{2}{3}\frac{1-p}{pQ}\ln\delta = 0. \tag{3-D.2}$$

Proof. For all Λ

$$J(\Lambda_o) \leqslant J(\Lambda), \tag{3-D.3}$$

and therefore

$$G(\Lambda) \leqslant G(\Lambda_o) = -10 \log p^*, \tag{3-D.4}$$

thus

$$\Pr(G(\Lambda) \leqslant -10\log p + \epsilon) \geqslant \Pr(-10\log p^* \leqslant -10\log p + \epsilon), \tag{3-D.5}$$

for some $\epsilon > 0$. Next, we evaluate the expression on the r.h.s. of (3-D.5). Note that

$$\Pr(-10\log p^* \leqslant -10\log p + \epsilon) = \tag{3-D.6}$$

$$\Pr\left(\log p^* \geqslant \log p - \frac{\epsilon}{10}\right) = \tag{3-D.7}$$

$$\Pr\left(p^* \geqslant \frac{p}{10^{\frac{\epsilon}{10}}}\right) = \Pr(p^* \geqslant \epsilon' p) = 1 - \Pr(p^* \leqslant \epsilon' p), \tag{3-D.8}$$

where $\epsilon' = \frac{1}{10^{\frac{\epsilon}{10}}} < 1$. Note that

$$\Pr(p^* \leqslant \epsilon' p) = \Pr(p^* - p \leqslant p(\epsilon' - 1)) = \tag{3-D.9}$$

$$\Pr(p - p^* \geqslant p(1-\epsilon')). \tag{3-D.10}$$

For a sequence of i.i.d. r.v.'s $\{X_i\}_{i=1}^n$ with $\mathrm{E}\{X_i\} = 0$, Bernstein's inequality provides

$$\Pr\left(\sum_{i=1}^{n} X_i \geqslant t\right) \leqslant \exp\left\{-\frac{t^2/2}{\sum_i \mathrm{E}X_i^2 + M\frac{t}{3}}\right\}, \tag{3-D.11}$$

for $t > 0$, and where $|X_i| \leqslant M$ with probability 1. Let $X_i = p - I_i$, then assuming $p \leqslant \frac{1}{2}$ yield $M = 1 - p$ and

$$\Pr(p - p^* \geqslant t) \leqslant \exp\left\{-\frac{Q\frac{t^2}{2}}{(1-p)\left(p + \frac{t}{3}\right)}\right\}. \tag{3-D.12}$$

Utilizing (3-D.12) we obtain

$$\Pr(p - p^* \geqslant p(1-\epsilon')) \leqslant \exp\left\{-Q\frac{3}{2}\frac{p}{1-p}\frac{(1-\epsilon')^2}{4-\epsilon'}\right\} \triangleq \delta. \tag{3-D.13}$$

When combined, (3-D.5)-(3-D.13) provide

$$\Pr(G(\Lambda) \leqslant -10\log p + \epsilon) \geqslant 1 - \delta. \tag{3-D.14}$$

This completes the proof. □

Discussion

Bernstein's inequality requires $t > 0$ which hold provided that $\epsilon' \in (0, 1)$. Unfortuntaely, this means that we do not have $\delta \to 0$ since substituting $\epsilon' = 0$ in (3-D.13) yields

$$\delta \geqslant \exp\left\{-Q\frac{3}{8}\frac{p}{1-p}\right\}, \tag{3-D.15}$$

and Q and p are given. Nevertheless, δ may still be sufficiently small to make *Claim 1* attractive. In Section 3.5 we provide a SAR imaging example of our methods. A

(450[pix] $\times$ 570[pix]) image containing 13 targets each roughly (13[pix] $\times$ 13[pix]) is examined. This translates to $p < 0.01$ and $Q = 256,500$. Therefore, setting $\delta = 10^{-3}$ yields $\epsilon = 0.333$ or

$$\Pr(G(\Lambda) \leqslant -10 \log p + 0.333) \geqslant 1 - 10^{-3},$$

i.e., the gain of any effort allocation policy Λ is smaller than 20.333[dB] with probability grater than 0.999.

3-E Appendix: Minimizing the cost function

Let $\boldsymbol{Y}(t) \in \mathcal{Y}$ be a $(Q \times 1)$ random vector (r.v.) with a probability density function (pdf) $p_{\boldsymbol{Y}(t)}(\boldsymbol{y}) > 0$, for all $\boldsymbol{y} \in \mathcal{Y}$ and $t \in \{1, 2, \ldots, T\}$, representing random observations. Let $\boldsymbol{I}_\Psi = [I_1, I_2, \ldots, I_Q]'$ be a r.v. with $\Pr(I_i = 1) = p_i$. Let $Y_1^t = \{\boldsymbol{Y}(1), \boldsymbol{Y}(2), \ldots, \boldsymbol{Y}(t)\}$ be a collection of all observation up to time t. Define $x(i,t) : Y_1^{t-1} \to \mathbb{R}_+$, then, for some $\nu \in (\frac{1}{2}, 1]$, our goal is to find $\{\hat{x}(i,t)\}$ with $i = 1, \ldots, Q$ and $t = 1, \ldots, T$, such that

$$\hat{x}(i,t) = \arg\min_{x(i,t)} \mathrm{E}\left\{ \sum_{i=1}^{Q} \frac{\nu I_i + (1-\nu)(1-I_i)}{\sum_{t=1}^{T} x(i,t)} \right\},$$

where expectation is taken w.r.t. $\boldsymbol{I}_\Psi$ and Y_1^{T-1}, subject to $\sum_{i=1}^{Q} \sum_{t=1}^{T} x(i,t) = X$.

E.1 The case of $T = 1$

For $T = 1$, our cost function has the following form

$$\hat{x}(i) = \arg\min_{x(i)} \mathrm{E}\left\{ \sum_{i=1}^{Q} \frac{\nu I_i + (1-\nu)(1-I_i)}{x(i)} \right\},$$

subject to $\sum_{i=1}^{Q} x(i) = X$, with the expectation taken w.r.t. $\boldsymbol{I}_\Psi$. Note that $\mathrm{E}\{I_i\} = p_i$, so $\hat{x}_i$ can be derived using Lagrange multiplier, i.e., finding the minimizer of

$$L(\boldsymbol{x}, \lambda) = \sum_{i=1}^{Q} \frac{\nu p_i + (1-\nu)(1-p_i)}{x(i)} + \lambda \left(\sum_{i=1}^{Q} x(i) - X \right).$$

Taking derivatives and setting them equal to zero, yields

$$\hat{x}(i) = \frac{X\sqrt{\nu p_i + (1-\nu)(1-p_i)}}{\sum_{j=1}^{Q} \sqrt{\nu p_j + (1-\nu)(1-p_j)}}.$$

E.2 The case of $T = 2$

Consider the following problem, let $x(i,1) = x_1 > 0$ for all i, then $\sum_{i=1}^{Q} x(i,1) = Qx_1$. Our goal is minimize the cost, i.e., find

$$\hat{x}(i,2) = \arg\min_{x(i,2)} \mathrm{E} \left\{ \sum_{i=1}^{Q} \frac{\nu I_i + (1-\nu)(1-I_i)}{x_1 + x(i,2)} \right\}, \tag{3-E.1}$$

subject to

$$\sum_{i=1}^{Q} x(i,2) = X - Qx_1 \geqslant 0, \tag{3-E.2}$$

and $x(i,2) : \boldsymbol{Y}(1) \to \mathbb{R}_+$. For brevity, let $x_i(\boldsymbol{Y}) = x(i,2)$, and note that in (3-E.1) expectation is taken w.r.t. $\boldsymbol{I}_\Psi$ and $\boldsymbol{Y}(1)$. Using iterated expectation, we obtain

(3-E.3)

$$\mathrm{E} \left\{ \sum_{i=1}^{Q} \frac{\nu I_i + (1-\nu)(1-I_i)}{x_1 + x(i,2)} \right\} = \mathrm{E} \left\{ \sum_{i=1}^{Q} \mathrm{E} \left\{ \frac{\nu I_i + (1-\nu)(1-I_i)}{x_1 + x_i(\boldsymbol{Y})} \middle| \boldsymbol{Y}(1) \right\} \right\}.$$

Given $\boldsymbol{Y}(1)$ the denominator is deterministic and expectation can be applied to the numerator, therefore (3-E.3) becomes

$$\mathrm{E}\left\{\sum_{i=1}^{Q}\frac{\nu\mathrm{Pr}(I_i=1|\boldsymbol{Y}(1))+(1-\nu)(1-\mathrm{Pr}(I_i=1|\boldsymbol{Y}(1)))}{x_1+x_i(\boldsymbol{Y})}\right\}.$$

Define $p_{I_i|\boldsymbol{Y}}=\mathrm{Pr}(I_i=1|\boldsymbol{Y}(1))$ and $W_i=\nu p_{I_i|\boldsymbol{Y}}+(1-\nu)(1-p_{I_i|\boldsymbol{Y}})$, we use Lagrange multiplier to minimize

$$\mathrm{E}\left\{\sum_{i=1}^{Q}\frac{W_i}{x_1+x_i(\boldsymbol{Y})}\right\}+\int_{\mathcal{Y}}\lambda'(\boldsymbol{y})\left(\sum_{i=1}^{Q}x_i(\boldsymbol{y})-(X-Qx_1)\right)d\boldsymbol{y}=$$

$$\sum_{i=1}^{Q}\int_{\mathcal{Y}}\left[\frac{w_i}{x_1+x_i(\boldsymbol{y})}f_{\boldsymbol{Y}}(\boldsymbol{y})+\lambda'(\boldsymbol{y})\left(x_i(\boldsymbol{y})-(X-Qx_1)\right)\right]d\boldsymbol{y}. \tag{3-E.4}$$

where $w_i=W_i(\boldsymbol{y})$ is a realization of the r.v. W_i. Since $f_{\boldsymbol{Y}}(\boldsymbol{y})$ is strictly positive, define $\lambda(\boldsymbol{y})=\frac{\lambda'(\boldsymbol{y})}{f_{\boldsymbol{Y}}(\boldsymbol{y})}$, then, $\hat{x}_i$, the minimizer of (3-E.4) is also the minimizer of

$$\sum_{i=1}^{Q}\int_{\mathcal{Y}}\left[\frac{w_i}{x_1+x_i(\boldsymbol{y})}+\lambda(\boldsymbol{y})x_i(\boldsymbol{y})\right]d\boldsymbol{y},$$

Note that our problem has translated to minimizing a separable sum of integrals of a positive integrands. Hence, finding

$$\hat{x}_i(\boldsymbol{Y})=\arg\min_{x_i(\boldsymbol{Y})}\frac{w_i}{x_1+x_i(\boldsymbol{Y})}+\lambda(\boldsymbol{Y})x_i(\boldsymbol{Y}),$$

suffice. Solving for $x_i(\boldsymbol{y})$ given the Lagrange multiplier $\lambda(\boldsymbol{Y})$ yields

$$x_i(\boldsymbol{Y})=\begin{cases}\sqrt{\frac{w_i}{\lambda(\boldsymbol{Y})}}-x_1, & \sqrt{\frac{w_i}{\lambda(\boldsymbol{Y})}}>x_1\\ 0, & \text{Otherwise}\end{cases}, \tag{3-E.5}$$

which can be also written as $x_i(\boldsymbol{Y}) = (\sqrt{\frac{w_i}{\lambda(\boldsymbol{Y})}} - x_1) I(\lambda(\boldsymbol{Y}) < \frac{w_i}{x_1^2})$, where $I(\cdot)$ is an indicator function. Utilizing the constraint $\sum_i x_i(\boldsymbol{Y}) = X - Qc_1$, we obtain

$$\frac{1}{\sqrt{\lambda(\boldsymbol{Y})}} \sum_{i=1}^{Q} \sqrt{w_i} I\left(\lambda(\boldsymbol{Y}) < \frac{w_i}{x_1^2}\right) = X - x_1 \left(Q - \sum_{i=1}^{Q} I\left(\lambda(\boldsymbol{Y}) < \frac{w_i}{x_1^2}\right)\right). \tag{3-E.6}$$

Note that

$$Q = \sum_{i=1}^{Q} I\left(\lambda(\boldsymbol{Y}) < \frac{w_i}{x_1^2}\right) + \sum_{i=1}^{Q} I\left(\lambda(\boldsymbol{Y}) \geqslant \frac{w_i}{x_1^2}\right), \tag{3-E.7}$$

substituting (3-E.7) into (3-E.6) and rearranging, yields

$$\sqrt{\lambda(\boldsymbol{Y})} = \frac{\sum_{i=1}^{Q} \sqrt{w_i} I\left(\lambda(\boldsymbol{Y}) < \frac{w_i}{x_1^2}\right)}{X - x_1 \sum_{i=1}^{Q} I\left(\lambda(\boldsymbol{Y}) \geqslant \frac{w_i}{x_1^2}\right)}. \tag{3-E.8}$$

Next, use $\tau : \mathcal{X} \to \mathcal{X}$ defined in (3.32) such that $w_{\tau(1)} \leqslant w_{\tau(2)} \leqslant \ldots \leqslant w_{\tau(Q)}$.

Case $\lambda(\boldsymbol{Y})x_1^2 < w_{\tau(1)}$

If $\lambda(\boldsymbol{Y})x_1^2 < w_{\tau(1)}$, for all $\boldsymbol{Y}$, then $\lambda(\boldsymbol{Y})x_1^2 < w_{\tau(i)}$ for all i and $\sqrt{\lambda(\boldsymbol{Y})x_1^2} = \frac{\sum_{i=1}^{Q} \sqrt{w_{\tau(i)}}}{X}$. For which case, the cost minimizer is achieved at

$$x_{\tau(i)}(\boldsymbol{Y}) = \frac{X\sqrt{w_{\tau(i)}}}{\sum_{j=1}^{Q} \sqrt{w_j}} - x_1,$$

and, for all i

$$\frac{X}{x_1} \geqslant \frac{\sum_{j=1}^{Q} \sqrt{w_j}}{\sqrt{w_{\tau(i)}}}. \tag{3-E.9}$$

Case $\lambda(\boldsymbol{Y})x_1^2 \geqslant w_{\tau(Q)}$

For the other extreme, we have $\lambda(\boldsymbol{Y})x_1^2 \geqslant w_{\tau(i)}$ for all i, which in turn provides $\lambda(\boldsymbol{Y}) = 0$, due to (3-E.8). This can only hold if both $\nu = 1$ and $p_{I_i|\boldsymbol{Y}} = 0$ for all i. If this is the case, then for any choice of $x_i(\boldsymbol{Y})$ the overall cost is zero.

Case $w_{\tau(k)} \leqslant \lambda(\boldsymbol{Y})x_1^2 < w_{\tau(k+1)}$**, for some** $k \in \{1, 2, \ldots, Q-1\}$

The interesting case is when $\lambda(\boldsymbol{Y})$ has some intermittent value, i.e.,

$$\frac{\sqrt{w_{\tau(k)}}}{x_1} \leqslant \frac{\sum_{i=k+1}^{Q}\sqrt{w_{\tau(i)}}}{X - kx_1} < \frac{\sqrt{w_{\tau(k+1)}}}{x_1}. \tag{3-E.10}$$

Since all the terms in (3-E.10) are positive, we rewrite the inequality as

$$\frac{\sum_{i=k+1}^{Q}\sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(k+1)}}} + k < \frac{X}{x_1} \leqslant \frac{\sum_{i=k+1}^{Q}\sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(k)}}} + k. \tag{3-E.11}$$

Define

$$\begin{aligned} \mathrm{LB}(k) &= \frac{\sum_{i=k+1}^{Q}\sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(k+1)}}} + k, \\ \mathrm{UB}(k) &= \frac{\sum_{i=k+1}^{Q}\sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(k)}}} + k. \end{aligned} \tag{3-E.12}$$

To show that (3-E.11) makes sense, we need to show that for any X of interest there exists a unique $k = k_0 \in \{1, 2, \ldots, Q-1\}$ such that $\frac{X}{x_1} \in (\mathrm{LB}(k_0), \mathrm{UB}(k_0)]$. To do so we show that following: (i) $\mathrm{LB}(k) \leqslant \mathrm{UB}(k)$, (ii) $\mathrm{LB}(k-1) = \mathrm{UB}(k)$, and (iii) $\frac{X}{x_1} \in (\mathrm{LB}(Q-1), \mathrm{UB}(1)]$. Start with (i)

$$\frac{\sum_{i=k+1}^{Q}\sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(k+1)}}} + k - \left(\frac{\sum_{i=k+1}^{Q}\sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(k)}}} + k\right) = \tag{3-E.13}$$

$$\left(\frac{\sqrt{w_{\tau(k)}}-\sqrt{w_{\tau(k+1)}}}{\sqrt{w_{\tau(k)}w_{\tau(k+1)}}}\right)\sum_{i=k+1}^{Q}\sqrt{w_{\tau(i)}}\leqslant 0,$$

since $w_{\tau(k)}\leqslant w_{\tau(k+1)}$. Note that $w_{\tau(k)}=w_{\tau(k+1)}$ implies that the interval $(\mathrm{LB}(k),\mathrm{UB}(k)]$ is empty. To prove (ii) we substitute $(k-1)$ for k in the expression for $\mathrm{LB}(k)$ in (3-E.12). This yields

$$\mathrm{LB}(k-1)-\mathrm{UB}(k)=\frac{\sqrt{w_{\tau(k)}}}{\sqrt{w_{\tau(k)}}}-1=0. \tag{3-E.14}$$

Combining (i) and (ii) provides $\mathrm{LB}(k)\leqslant\mathrm{UB}(k)=\mathrm{LB}(k-1)$, i.e., $\mathrm{LB}(k)$ is decreasing in k. Finally, we need to show $\frac{X}{x_1}\in(\mathrm{LB}(Q-1),\mathrm{UB}(1)]$. First, note from (3-E.9) that $\frac{X}{x_1}>\mathrm{UB}(1)$ belongs to *Case 1* for which a separate solution was provided. Second, from (3-E.12) we obtain $\mathrm{LB}(Q-1)=Q$. If $Qc_1=X$ then $x_i(\boldsymbol{Y})=0$, for all i, due to the constraint (3-E.2). Thus, for a two-step problem we must have $\frac{X}{x_1}>Q=\mathrm{LB}(Q-1)$. Hence, $\frac{X}{x_1}\in(\mathrm{LB}(Q-1),\mathrm{UB}(1)]$ as required. Together (i), (ii), and (iii) proves existence and uniqueness of the solution, since

1. $\frac{X}{x_1}\in(\mathrm{LB}(Q-1),\mathrm{UB}(1)]$,

2. $\bigcup_{k=1}^{Q}(\mathrm{LB}(k),\mathrm{UB}(k)]=(\mathrm{LB}(Q-1),\mathrm{UB}(1)]$,

3. $(\mathrm{LB}(i),\mathrm{UB}(i)]\cap(\mathrm{LB}(j),\mathrm{UB}(j)]=\emptyset,\ \forall i\neq j$.

To conclude this section we note that $\hat{x}_i(\boldsymbol{Y})$ the minimizer of (3-E.1) is given by (3.38) replacing x_1 for λ_1^* and $\hat{x}_{\tau(i)}(\boldsymbol{Y})$ for $\lambda(\tau(i),2)$. A general solution for the case of a non-uniform prior distribution is given in [5].

3-F Convergence of $p_{I_i|\boldsymbol{y}(1),\boldsymbol{\theta}}$ for the Gaussian case

For brevity, we prove the convergence given $\boldsymbol{\theta}$. This still holds for a random target return, e.g. $\theta_i \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, but the derivation is a little messier. Consider a simplified version of the measurement model (3.18) given as

$$y_i(1) = \sqrt{\lambda_1} I_i \theta_i(1) + n_i(1), \tag{3-F.1}$$

where we substitute $t = 1$, $\lambda(i,1) = \lambda_1$, and $h_{ij}(1) = \delta_{ij}$ with δ_{ij} being the Kronecker delta. We further assume that the r.v. $\theta_i > 0$. For brevity we suppress the time dependency from here on. In addition we assumed $\Pr(I_i = 1) = p$ for all i and the noise samples are i.i.d. Gaussian with $n_i \sim \mathcal{N}(0, \sigma^2)$. This setting results in i.i.d. measurements distributed according to

$$y_i \sim \mathcal{N}(\sqrt{\lambda_1} I_i \theta_i, \sigma^2), \tag{3-F.2}$$

Therefore, the posterior probabilities $p_{I_i|\boldsymbol{y}}$ depends only on the i^{th} sample, i.e., $p_{I_i|\boldsymbol{y}} = p_{I_i|y_i}$. For the Gaussian case, $p_{I_i|y_i}$ has an explicit form given as

$$p_{I_i|y_i} = \frac{p \exp\{-\frac{1}{2\sigma^2}(y_i - \theta_i\sqrt{\lambda_1})^2\}}{p \exp\{-\frac{1}{2\sigma^2}(y_i - \theta_i\sqrt{\lambda_1})^2\} + (1-p)\exp\{-\frac{1}{2\sigma^2} y_i^2\}}, \tag{3-F.3}$$

which can be rearranged as

$$p_{I_i|y_i} = \frac{1}{1 + \frac{1-p}{p} \exp\left\{-\frac{\theta_i \lambda_i}{\sigma^2}\left(\frac{y_i}{\sqrt{\lambda_1}} - \frac{\theta_i}{2}\right)\right\}}. \tag{3-F.4}$$

Let $z_i = \frac{y_i}{\sqrt{\lambda_1}}$, then $z_i \sim \mathcal{N}(\theta_i I_i, \eta^2)$ where $\eta^2 = \frac{\sigma^2}{\lambda_1}$ and at high SNR $\eta^2 \to 0$. Substituting y_i with z_i in (3-F.4) provides

$$p_{I_i|y_i} = p_{I_i|z_i} = \frac{1}{1 + \frac{1-p}{p} \exp\left\{-\frac{\theta_i}{\eta^2}\left(z_i - \frac{\theta_i}{2}\right)\right\}}. \tag{3-F.5}$$

Claim 2. *The conditional probability $p_{I_i|y_i}$ defined in (3-F.5) converges in probability to I_i as $\eta \to 0$ for any $\theta_i > 0$, i.e., $\forall \epsilon > 0$ and some $\delta > 0$ there exists $\eta_0(\epsilon)$, such that $\forall \eta \leqslant \eta_0(\epsilon)$*

$$\Pr(|p_{I_i|y_i}(\eta) - I_i| > \delta) \leqslant \epsilon, \tag{3-F.6}$$

The intuition behind *Claim 2* is derived from the following limit

$$\lim_{\eta^2 \to 0} p_{I_i|y_i} = \begin{cases} 0, & z_i < \frac{\theta_i}{2} \\ \\ 1, & z_i > \frac{\theta_i}{2} \end{cases}. \tag{3-F.7}$$

Proof. Using Chebyshev's inequality we prove *Claim 2* for the case of $I_i = 0$. A symmetry argument suggests that the same line of proof can be applied for the case of $I_i = 1$. Define $A = \{z_i : z_i > \frac{\theta_i}{2}\}$ then

$$\Pr(|p_{I_i|y_i} - I_i| > \delta) = \Pr(A)\Pr(|p_{I_i|y_i} - I_i| > \delta|A) +$$

$$\Pr(A^c)\Pr(|p_{I_i|y_i} - I_i| > \delta|A^c). \tag{3-F.8}$$

For $I_i = 0$ we have[3]

$$\Pr(p_{I_i|y_i} > \delta) \leqslant \Pr(A) + \Pr\left(p_{I_i|y_i} > \delta \,\middle|\, z_i \leqslant \frac{\theta_i}{2}\right). \tag{3-F.9}$$

We show that both elements on the r.h.s. (3-F.9) approach zero as $\eta \to 0$. For some $\alpha > 0$, Chebyshev's inequality provides

$$\Pr\left(|z_i - \theta_i I_i| \geqslant \alpha\right) \leqslant \frac{\eta^2}{\alpha^2}. \tag{3-F.10}$$

[3]Since $p_{I_i|y_i} \geqslant 0$ we can replace $p_{I_i|y_i} = |p_{I_i|y_i}|$.

Note that

$$\Pr(z_i > \tfrac{\theta_i}{2}) \leqslant \Pr(|z_i - \theta_i I_i| \geqslant \tfrac{\theta_i}{2}), \quad I_i = 0,$$

(3-F.11)

$$\Pr(z_i < \tfrac{\theta_i}{2}) \leqslant \Pr(|z_i - \theta_i I_i| \geqslant \tfrac{\theta_i}{2}), \quad I_i = 1,$$

and therefore

$$\Pr(A) \leqslant \Pr\left(|z_i| \geqslant \frac{\theta_i}{2}\right) \leqslant \frac{4\eta^2}{\theta_i^2}. \tag{3-F.12}$$

Next, we examine the term on r.h.s. of (3-F.9). Note that

$$\Pr\left(p_{I_i|y_i} > \delta \,\middle|\, z_i \leqslant \frac{\theta_i}{2}\right) = \tag{3-F.13}$$

$$\Pr\left(\frac{1}{1 + \frac{1-p}{p}\exp\left\{\frac{\theta_i}{\eta^2}\left(\frac{\theta_i}{2} - z_i\right)\right\}} > \delta \,\middle|\, z_i \leqslant \frac{\theta_i}{2}\right) = \tag{3-F.14}$$

$$\Pr\left(\frac{\theta_i}{2} - z_i < \frac{\eta^2}{\theta_i}\ln\frac{p}{1-p}\frac{1-\delta}{\delta} \,\middle|\, z_i \leqslant \frac{\theta_i}{2}\right). \tag{3-F.15}$$

Let

$$\epsilon' = \frac{\eta^2}{\theta_i}\ln\frac{p}{1-p}\frac{1-\delta}{\delta}, \tag{3-F.16}$$

then (3-F.13) is equivalent to

$$\Pr\left(z_i \in \left[\frac{\theta_i}{2} - \epsilon', \frac{\theta_i}{2}\right]\right) \tag{3-F.17}$$

First, note that $\delta > p$ results in $\epsilon' < 0$ and hence (3-F.17) equals zero. Second, since $f_Z(z) \leqslant f_Z(0) = \frac{1}{\sqrt{2\pi\eta^2}}$ we have the following simple bound for any $\epsilon' > 0$

$$\Pr\left(z_i \in \left[\frac{\theta_i}{2} - \epsilon', \frac{\theta_i}{2}\right]\right) \leqslant \epsilon' \frac{1}{\sqrt{2\pi\eta^2}} = \eta \frac{1}{\theta_i\sqrt{2\pi}} \ln \frac{p}{1-p} \frac{1-\delta}{\delta}, \tag{3-F.18}$$

thus

$$\Pr\left(p_{I_i|y_i} > \delta \,\middle|\, z_i \leqslant \frac{\theta_i}{2}\right) \leqslant \eta \frac{1}{\theta_i\sqrt{2\pi}} \ln \frac{p}{1-p} \frac{1-\delta}{\delta}. \tag{3-F.19}$$

Finally, substituting (3-F.19) and (3-F.12) in (3-F.9) we obtain

$$\Pr(p_{I_i|y_i} > \delta) \leqslant \frac{4\eta^2}{\theta_i^2} + \eta \frac{1}{\theta_i\sqrt{2\pi}} \ln \frac{p}{1-p} \frac{1-\delta}{\delta}. \tag{3-F.20}$$

Define $\eta_0(\epsilon)$ as

$$\eta_0(\epsilon) = \sup_{\eta>0} \left\{ \frac{4\eta^2}{\theta_i^2} + \eta \frac{1}{\theta_i\sqrt{2\pi}} \ln \frac{p}{1-p} \frac{1-\delta}{\delta} \leqslant \epsilon \right\}, \quad \forall \delta > 0, \tag{3-F.21}$$

then using (3-F.20) we have

$$p_{I_i|y_i} \to 0 \tag{3-F.22}$$

in probability when $I_i = 0$. In a similar manner, it can be shown that

$$p_{I_i|y_i} \to 1 \tag{3-F.23}$$

in probability for $I_i = 1$. □

CHAPTER IV

The multi-scale search problem

4.1 Introduction

This chapter considers the problem of estimating and detecting sparse signals over a large area of an image or other medium from noisy observations. We further assume that the nonzero signal elements, i.e., targets, tend to cluster together in areas which we call regions of interest (ROI). In Chapter III we introduced a novel cost function and solved a related minimization problem, which yielded an asymptotically optimal adaptive resource allocation policy, namely *ARAP*. *ARAP* uses a two-stage approach: first the entire domain is scanned, second a subset of it, suspected as part of the ROI, is re-scanned. For signals in $\mathbb{R}^Q$, *ARAP* requires $N > Q$ measurements to complete. In this work we explore tradeoffs between the total number of expected measurements and performance gain. We use the knowledge that targets tend to cluster together and modify *ARAP* (M-ARAP) in a multi-scale fashion. In the first stage M-ARAP takes $M_1 < Q$ measurements that cover the entire signal domain on a coarse grid, then re-sample part of the signal domain on a fine grid at the second stage. While we lose optimality, we show that the estimated ROI yielded by M-ARAP still converges to the true ROI while the overall number of samples is greatly reduced and significant estimation gains are maintained compared to an exhaustive

search policy. We show that for sparse signals with sufficient ROI to background contrast we have 7 [dB] reduction in estimation error at signal to noise ratio of 5 [dB] with $N < \frac{Q}{5}$ samples.

This scenario is common in radar target detection where large scattering objects may occupy consecutive pixels on the Radar screen and appear as a cluster of targets. Nevertheless, the overall area occupied by targets is small compared to the area a scanning radar system covers. In early detection of cancer tumors such as in breast cancer, the diameter of a typical tumor is a few millimeters to 1.5 centimeters. Hence, on a fine grid a tumor may appear as a cluster of targets, yet, its overall volume is very small compared to the volume of the entire breast.

The notion of saving measurements when sampling sparse signals has received significant amount of attention in recent years. Work in the field of *Compressed Sensing* (CS) challenges the traditional signal processing sampling requirements. Recent results show that a relatively small number of random projections of a signal can contain most of its salient information. It follows that if a signal is compressible in some orthonormal basis, it can be accurately recovered from random projections even when they are contaminated with noise [9, 24]. Candes and Tao [9] introduce the Dantzig selector (DS) algorithm which solves an l_1-norm minimization problem to reconstruct a sparse signal in $\mathbb{R}^Q$ from a very limited set of $N < Q$ noisy observations. Their algorithm converges to the true solution as long as the measurement operator obeys the uniform uncertainty principle (UUP). They provide an upper bound on the mean squared error (MSE) which, remarkably, is proportional up to a $C \log Q$ factor of the noise level σ^2. Haupt and Nowak present a similar result but their measurement operator is randomized rather than following the UUP [24]. Although both methods work extremely well for (almost) truly sparse signals, they

do not focus efforts to specifically acquire the ROI. The measurement efforts are uniformly spread at random locations in a single stage. While we share the goal of reducing the overall number of measurements we chose an orthogonal approach. Rather than randomly selecting the views used for the reconstruction, we learn from past measurements and sequentially focus search efforts onto spatial locations of the signal corresponding to an estimated ROI.

Abdel-Samad and Tewfik [2] study the problem of how to best allocate N measurements to find a single target hidden in Q cells for radar target localization when $N < Q$. They propose a hierarchical approach recursively grouping the Q cells into $q < Q$ groups in a tree like structure. They assume that the radar beam pattern can be shaped accordingly and that signal to noise ratio (SNR) decreases as the group size increases. Their multiple hypothesis testing approach is computationally intense and does not account for large N and Q ($Q = 64$ is used in their example). The solution complexity limits the number of measurements they can allocate at each stage of the sequential search. In a similar manner we also note degradation in performance as the scale increases. The main difference from our work is that they account for a single target and the multi-hypothesis test cannot accept more than one correct hypothesis at a single stage. Therefore, their method is not applicable for large scale search problems of the type discussed here.

To the best of our knowledge, multiscale search approach was introduced in the early 1940's. Dorfman considered the problem of detecting defective members of a large population in the context of weeding out all syphilitic men called up for induction [16]. The test was so sensitive and accurate that Dorfman suggests the following procedure: 1) draw blood from each candidate, 2) use half of each sample to create a pool containing samples from n individual subjects, 3) test the pool. If a pool

was tested positive the other half sample of each pool member was individually tested to detect the defective member. In the case of low disease prevalence rates a great amount of time savings can be achieved. Dorfman type procedures use a binary model (B-model) and do not account for false alarms or missed detections, which in our setting is equivalent to an infinite SNR. Therefore they do not require the additional degree of freedom of resource allocation and are only concerned with minimizing the total number of samples required. An optimal group size n can be analytically evaluated for each disease prevalence rate. Dorfman procedures enjoyed great success due to their simplicity and effectiveness. Pfeifer modifies the binomial model (M-model) and considers test values that were either zero (for negative) or greater than zero (for the degree of contamination) [45]. This way, when a pool is tested positive, each sample of a subgroup from the pool reveals information regarding the other pool members, thus even greater savings are achieved. However, the modified model still does not account for false alarms or missed detections. Although we use a structure similar to the Dorfman procedure, the presence of noise creates a significant difference between our work and the previous work done on blood pooling.

Finally, in Chapter III we examined utilities in a two-stage adaptive measuring scheme for detection and estimation of sparse signals containing targets in noise. We assumed independent and identical distribution (i.i.d.) of targets among pixels and presented a two stage approach for resource allocation under a fixed energy constraint, namely $ARAP$. In this chapter we look at detecting and estimating clusters of targets under additional sampling or time constraints. Since targets are assumed to be clustered, a logical approach is to use a coarse measurement scale to reduce the number of samples used at the first stage. A modified version of $ARAP$ is introduced to select locations and energy levels with which areas in the scanned domain

are re-sampled on a fine grid. The two measurements can be later combined to both detect and estimate the region of interest (ROI) and its content. In Chapter III we showed $ARAP$ to be asymptotically optimal in terms of estimation gains being inversely proportional to the signal sparsity. Here performance would depend on sparsity, number of samples allowed, and the inherent contrast of the signal represented by the '*detectability index*'. We show that with sufficient contrast we improve estimation accuracy by more than 7 [dB] while using less than 20% of the number of measurements used by a traditional sampling scheme at SNR of 5 [dB] and sparsity of 0.1%. Moreover, estimation gains increase with SNR leading to surprisingly large gains.

The rest of this chapter is organized as follows. In Section 4.2 we formally introduce the search problem and define a cost function. Section 4.3 discusses our search methods and analyzes their properties. In Section 4.4 we compare the performance of our adaptive multiscale approach to an exhaustive search. We conclude and point out future work in Section 4.5.

4.2 Problem formulation

Consider a discrete space $\mathcal{X} = \{1, 2, \ldots, Q\}$ containing Q cells and equipped with a probability measure P. We use Ψ to denote a region of interest (ROI) in $\mathcal{X}$, i.e., $\Psi \subseteq \mathcal{X}$. In the sequel Ψ will be a randomly selected small subset of $\mathcal{X}$. Thus, we assume $|\Psi| \ll |\Psi^c|$, where $|\Psi|$ equals the number of elements in Ψ and Ψ^c is the relative complement, $\mathcal{X} \setminus \Psi$, of Ψ. Exact definition of the ROI is application dependent. In radar target localization the ROI is the collection of all cells containing targets and target related phenomena, e.g., target shadows. In a medical imaging application, such as early detection of breast cancer, where tumor boundaries are

poorly defined, the ROI may be defined as the collection of all cells containing targets (a tumor) plus some neighboring cells.

Let I_i be an indicator function of the ROI such that

$$I_i = \begin{cases} 1, & i \in \Psi \\ 0, & \text{Otherwise} \end{cases} \tag{4.1}$$

and $\{p_i = \Pr(I_i = 1)\}_{i=1}^{Q}$ is an associated set of prior probabilities. In Chapter III we considered the case where the I_i's are i.i.d. and $p_i = p = \mathrm{E}\{\frac{|\Psi|}{Q}\}$ for all i. In this chapter we consider the case where it is highly likely that targets tend to cluster together and occupy neighboring pixels. In a one dimensional case we let the random variable Ξ denote a cluster length and assume $\Xi \leqslant \xi_0$ and $p(\Xi)$ is known, e.g. $\Pr(\Xi = i) = \frac{1}{\xi_0}$ for $i = 1, 2, \ldots, \xi_0$. Let $\boldsymbol{I}_\Psi = [I_1, \ldots, I_Q]'$ be a vector corresponding to the set of all indicators and $(\cdot)'$ denote the transpose operator. We say that the presence of a target affects cell i if $i \in \Psi$. We define the random vector $\boldsymbol{Y} : \mathcal{X} \to \mathbb{R}^M$ and consider the conditional probability $p(\boldsymbol{Y}|\boldsymbol{I}_\Psi)$.

Consider a sequential experiment where in the first $T-1$ stages we observe $\boldsymbol{Y}(t) : \mathcal{X} \to \mathbb{R}^{M_t}$ with $M_t < Q$ for $t = 1, 2, \ldots, T-1$ and in stage T we observe a selected subset of $\boldsymbol{Y}(T) : \mathcal{X} \to \mathbb{R}^Q$, i.e., we observe $y_i(T)$ for $i \in \widehat{\Psi} \subseteq \{1, 2, \ldots, Q\}$. This formulation allows us to limit the total number of observations to $N = \sum_{t=1}^{T-1} M_t + |\widehat{\Psi}|$. Let the distribution $\lambda(i,t) \geqslant 0$ denote the search effort allocated to cell i at time t, with

$$\sum_{t=1}^{T} \sum_{i=1}^{Q} \lambda(i,t) = 1, \quad 0 \leqslant \lambda(i,t) \leqslant 1 \tag{4.2}$$

and $\{\lambda(i,t)\}$ is a mapping from past observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$ to the probability

simplex and is called an effort allocation policy. The set $\widehat{\Psi}$ is formally defined as

$$\widehat{\Psi} = \{i \in \mathcal{X} : \lambda(i,T) > 0\}. \tag{4.3}$$

The combination of $\{\lambda(i,t)\}$, M_t, and $\widehat{\Psi}$ is termed a search policy. We focus here on a deterministic mappings λ, although a more general random mapping can also be incorporated into our framework but is not presented here. We assume that a sample's 'quality' is an increasing function of the allocated effort to the associated cell, e.g. measured in terms of information or inverse variance. In general, effort might be: computing power, complexity, cost, or energy that is allocated to acquiring a particular cell location. Define the cumulative search effort allotted to cell i as

$$\Lambda(i) = \sum_{t=1}^{T} \lambda(i,t). \tag{4.4}$$

In Chapter III we introduced the following cost function

$$J(\Lambda) = \sum_{i=1}^{Q} \frac{\nu I_i + (1-\nu)(1-I_i)}{\Lambda(i)}, \tag{4.5}$$

with $\nu \in [\frac{1}{2}, 1]$. Minimizing (4.5) subject to a total energy constraint λ_T yielded *ARAP*, for the case of $T = 2$ and $M_1 = Q$, which is summarized below.

Algorithm 4.1. Two stage Adaptive Resource Allocation Policy (ARAP) λ_A

Step 1: Allocate $\lambda_A(i,1) = \lambda_1^*$ to each cell and measure $\boldsymbol{y}(1)$.

Step 2: Given $\boldsymbol{y}(1)$ compute posterior probabilities, $p_{I_i|\boldsymbol{y}(1)} \stackrel{\triangle}{=} \Pr(I_i = 1|\boldsymbol{y}(1))$ and $w_i = \nu p_{I_i|\boldsymbol{y}(1)} + (1-\nu)(1-p_{I_i|\boldsymbol{y}(1)})$, then rank order the w_i's.

Step 3: Use λ_1^* and the ordered statistic $w_{\tau(i)}$ to find an optimal threshold k_0.

Step 4: Given k_0, apply $\lambda(i,2)$, the energy allocation, to cell i as

(4.6)

$$\lambda_A(\tau(i),2) = \lambda(\tau(i),2) = \left(\frac{\lambda_T - k_0\lambda_1^*}{\sum_{j=k_0+1}^{Q} \sqrt{w_{\tau(j)}}} \sqrt{w_{\tau(i)}} - \lambda_1^* \right) I(i > k_0),$$

and measure $\boldsymbol{y}(2)$.

For $ARAP$ the normalized number of observations, $N^* = \frac{Q+(Q-k_0)}{Q} = 1 + (1 - \frac{k_0}{Q})$, is a random variable that was shown to converge in probability to $1+p$ at high SNR, where $p = \mathrm{E}\{\frac{|\Psi|}{Q}\}$. Our goal in this work is to add a sampling constraint on top of the fixed energy constraint to limit the total search time or the number of measurements. We use a modified version of $ARAP$ to achieve that goal. We modify $ARAP$ in two ways: *i*) Keep $M_1 < Q$ and let N^* be a random variable with bounded $\mathrm{E}\{N^*\}$, *ii*) Let $M_1 < Q$ and k_0 be fixed, thus $N^* = \frac{M_1}{Q} + (1 - \frac{k_0}{Q})$ is also fixed. Note that we do not intend to solve a new optimization problem but rather to modify an existing optimal solution, namely $ARAP$, in a manner that suits the additional constraints.

4.3 Search policy for $T = 2$ under total effort and time constraints

Let $\boldsymbol{\Theta} = [\Theta_1, \ldots, \Theta_Q]'$ be a random vector where $\Theta_i \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$ are i.i.d. random variables (r.v.) corresponding to targets' profiles. We assume known background mean μ_b that is smaller than μ_θ and let $(\mu_\theta - \mu_b)$ represent the inherent contrast of the search problem. Without loss of generality we further assume $\mu_b = 0$. Let $\boldsymbol{I}_\Psi$ be a vector of indicators marking whether or not cell i contains a target. Note that clustering or spatial correlation between targets location would govern the distribution of $\boldsymbol{I}_\Psi$. Consider the following measurement model:

$$\tilde{\boldsymbol{y}}(t) = H(t)\mathrm{diag}\{\sqrt{\boldsymbol{\lambda}(t)}\}\mathrm{diag}\{\boldsymbol{\theta}\}\boldsymbol{I}_\Psi + \boldsymbol{n}(t), \tag{4.7}$$

where $H(t)$ is an $(M_t \times Q)$ matrix describing the measurement operator (e.g. beam forming), $\boldsymbol{\lambda}(t)$ is a vector describing resource allocation at time t, $[\sqrt{\boldsymbol{x}}]_i$ equals $\sqrt{x_i}$, $\mathrm{diag}\{\boldsymbol{x}\}$ is a square matrix with $[\mathrm{diag}\{\boldsymbol{x}\}]_{ii} = x_i$ and $[\mathrm{diag}\{\boldsymbol{x}\}]_{ij} = 0$ for $i \neq j$, and $\boldsymbol{n}(t) \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \mathrm{I}_{\mathrm{M_t}})$ where $\mathrm{I}_{\mathrm{M_t}}$ is an $(M_t \times M_t)$ identity matrix. In our model (4.7), both $H(t)$ and $\boldsymbol{\lambda}(t)$ are design parameters subject to the mentioned constraints. They are kept separate to ease the design procedure. We focus our attention here on the following simple design: let $\boldsymbol{h}_j' \in \mathbb{R}^Q$, the rows of $H(1)$, be a shifted version of $\boldsymbol{h}_1'$ specified by its entries:

$$H_{1j} = \begin{cases} \frac{1}{L}, & j \leqslant L \\ 0, & j > L \end{cases} \tag{4.8}$$

and $\lambda_i(1) = \lambda_1$ for all i. Let $< \boldsymbol{x}, \boldsymbol{y} >= \boldsymbol{y}'\boldsymbol{x}$ denote the standard inner product in $\mathbb{R}^2$ and assume $< \boldsymbol{h}_i, \boldsymbol{h}_j >= 0$ for all $i \neq j$, and that $L\sum_{i=1}^{M_t} \boldsymbol{h}_i' = \boldsymbol{1}'$, where $\boldsymbol{1} \in \mathbb{R}^Q$ is the vector of all ones. Hence, without overlapping between the rows of $H(1)$, a

scalar version of the measurement model (4.7) is given by

$$\tilde{y}_j(1) = \frac{\sqrt{\lambda_1}}{L} \sum_{i=(j-1)L+1}^{jL} \theta_i I_i + n_j(1), \qquad j = 1, 2, \ldots, \frac{Q}{L}, \tag{4.9}$$

where $L \geqslant \xi_0$ is a factor of Q. Let $\mathcal{X}_j = \{(j-1)L+1, \ldots, jL-1, jL\}$ denote the support being averaged by the sample $\tilde{y}_j(1)$ and $|\mathcal{X}_j| = L$ for all j. With small p, large Q, and $L \ll Q$ the probability that $\mathcal{X}_j$ contains more than one cluster of targets is negligible. Therefore, we assume that $\mathcal{X}_j$ contains at most one cluster of targets of length ξ_j. Let $\mathcal{H}_0$ denote the event $\mathrm{E}\{\tilde{y}_j(1)\} = 0$, i.e., the support $\mathcal{X}_j$ contains no targets or $I_i = 0$ for all $i \in \mathcal{X}_j$. Then $\tilde{y}_j(1)|\mathcal{H}_0$ is a zero mean Gaussian r.v. with variance σ^2. Let $\mathcal{H}_1$ denote the complement event where $\mathcal{X}_j$ contains ξ_j targets. The observation $\tilde{y}_j(1)$ given $\mathcal{H}_1$ is a Gaussian mixture and its mean depends on ξ_j, L, and the position of the cluster within $\mathcal{X}_j$. Further assume $\Pr(\xi_j = k) = \frac{1}{\xi_0}$ for $k = 1, \ldots, \xi_0$ and zero otherwise, therefore we have

$$\mathrm{E}\{\tilde{y}_j(1)|\mathcal{H}_1\} = \frac{\mu_\theta}{\xi_0} \frac{\sqrt{\lambda_1}}{L} \sum_{\xi=1}^{\xi_0} \xi \frac{\xi + L - 1}{L + 3\xi - 3} \triangleq \mu_1, \tag{4.10}$$

and

$$\mathrm{var}\{\tilde{y}_j(1)|\mathcal{H}_1\} = \sum_{i=1}^{\xi_0} q_i \left(\mathrm{E}\{y_i^2\} - 2m_i\mu_1 + \mu_1^2\right) \triangleq \sigma_1^2, \tag{4.11}$$

where the probability of each Gaussian q_i is given by

$$q_i = \frac{L + 4\xi_0 + 1 - 5i}{\frac{3}{2}\xi_0^2 + (L - \frac{3}{2})\xi_0}, \tag{4.12}$$

$m_i = i\frac{\sqrt{\lambda_1}}{L}\mu_\theta$, $s_i^2 = \sigma^2 + i\frac{\lambda_1}{L^2}\sigma_\theta^2$, and $\mathrm{E}\{y_i^2\} = s_i^2 + m_i^2$ (see Appendix 4-A).

4.3.1 First search policy

We define the following Multiscale Adaptive Resource Allocation Policy (M-ARAP) based on *ARAP*:

Algorithm 4.2. M-ARAP: Two stage Multiscale *ARAP* λ_M

Step 1: Allocate $\lambda_M(j,1) = \lambda_1^*$ to each support $\mathcal{X}_j$ and measure $\tilde{\boldsymbol{y}}(1)$ in (4.7).

Step 2: Define $\boldsymbol{y}(1) : \mathbb{R}^{M_1} \rightarrow \mathbb{R}^Q$ as an interpolated version of $\tilde{\boldsymbol{y}}(1)$, then compute posterior probabilities $p_{I_i|\boldsymbol{y}(1)} \stackrel{\triangle}{=} \Pr(I_i = 1|\boldsymbol{y}(1))$ and $w_i = \nu p_{I_i|\boldsymbol{y}(1)} + (1-\nu)(1 - p_{I_i|\boldsymbol{y}(1)})$.

Step 3: Rank order the w_i's using (4.14), then use λ_1^* and the ordered statistic $w_{\tau(i)}$ to find a threshold k_0 via (4.15) and (4.16).

Step 4: Given k_0, apply $\lambda(i,2)$, the energy allocation, to cell i as (4.13)

$$\lambda_M(\tau(i),2) = \lambda(\tau(i),2) = \left(\frac{\lambda_T - k_0 \frac{\lambda_1^*}{|\mathcal{X}_j|}}{\sum_{j=k_0+1}^{Q} \sqrt{w_{\tau(j)}}} \sqrt{w_{\tau(i)}} - \frac{\lambda_1^*}{|\mathcal{X}_j|} \right) I(i > k_0),$$

and measure $\boldsymbol{y}(2)$ using $H(2) = \mathrm{I}_Q$ and $[\boldsymbol{\lambda}(2)]_i = \lambda(i,2)$.

To complete the definition of M-ARAP, define the permutation operator $\tau : \mathcal{X} \rightarrow \mathcal{X}$ corresponding to the rank ordering of the w_i's as

$$\tau(j) = \arg \min_{i=1,\ldots,Q} \{w_i : w_i \geqslant w_{\tau(j-1)}\}, \qquad j \in \{1,2,\ldots,Q\}, \tag{4.14}$$

with $w_{\tau(0)} \stackrel{\triangle}{=} 0$. Whenever the r.h.s. of (4.14) is not unique we select an arbitrary i satisfying $w_{\tau(1)} \leqslant w_{\tau(2)} \leqslant \ldots \leqslant w_{\tau(Q)}$. Then, assuming $w_{\tau(1)} > 0$, define k_0, the

threshold parameter, as $k_0 = 0$ if

$$\frac{\lambda_T}{\lambda_1^*/L} > \frac{\sum_{i=1}^{Q}\sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(1)}}}, \tag{4.15}$$

otherwise $k_0 \in \{1, \ldots, Q-1\}$ is the integer satisfying

$$\frac{\sum_{i=k_0+1}^{Q}\sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(k_0+1)}}} < \frac{\lambda_T}{\lambda_1^*/L} - k_0 \leqslant \frac{\sum_{i=k_0+1}^{Q}\sqrt{w_{\tau(i)}}}{\sqrt{w_{\tau(k_0)}}}. \tag{4.16}$$

A proof that there exists a unique solution for (4.16) and its properties are given in Chapter III.

Note that if $M_1 = Q$ then M-ARAP is equivalent to *ARAP*, provided that λ_1^* is correctly defined. To define λ_1^* accordingly let

$$\lambda_1^* = \arg \min_{\lambda_1 \in \left(0, \frac{\lambda_T}{M_1}\right)} \mathrm{E}\left\{ \sum_{i=1}^{Q} \frac{\nu I_i + (1-\nu)(1-I_i)}{\Lambda(i)} \right\}, \tag{4.17}$$

where $\Lambda(i) = \frac{\lambda_1}{L} + \lambda(i,2)$ and $\lambda(i,2)$ is defined in (4.13) substituting λ_1 for λ_1^*. Note that $\Lambda_M(i) = \sum_{t=1}^{2} \lambda_M(i,t)$ depends on $p_{I_i|\boldsymbol{y}(1)}$, which, in turns depends on the statistical characteristics of $\boldsymbol{\Theta}$. To analyze expected performance of M-ARAP we explore the properties of $p_{I_i|\boldsymbol{y}(1)}$.

4.3.2 Detectability index and asymptotic properties of $p_{I_i|\boldsymbol{Y}(1)}$ when $\nu = 1$

Let $|\mathcal{X}_j| = L$ denote an observed support size for the first stage in M-ARAP. Let the true expected sparsity of the observed signal be $p = \frac{\mathrm{E}\{|\Psi|\}}{Q}$, and let Ξ_j be a random variable corresponding to the total number of targets at support $\mathcal{X}_j$. We assume that the distribution $p(\Xi)$, or at least $\mathrm{E}\{\Xi\}$, is known. In the current section we analyze asymptotic properties of M-ARAP, where by asymptotic we mean high SNR and large Q. We show:

1. Given that a target is present in support $\mathcal{X}_j$, the posterior probabilities $p_{I_i|\boldsymbol{Y}(1)} \rightarrow 1$ in probability for all $i \in \mathcal{X}_j$. For the complement case where $I_i = 0$ for all $i \in \mathcal{X}_j$ we have $p_{I_i|\boldsymbol{Y}(1)} \rightarrow 0$.
2. The expected normalized number of samples N^* is bounded, i.e., $\lim \Pr\left(\frac{1}{L} \leqslant N^* \leqslant \frac{1}{L} + pL\right) = 1$.
3. The normalized symmetric set difference $\frac{|\Psi \triangle \widehat{\Psi}|}{Q}$ is bounded by $p(L - \mathrm{E}\{\Xi\})$ with probability.

Recall that the *detectability index*, d, is defined as

$$d = \frac{|\mathrm{E}\{y_j|\mathcal{H}_1\} - \mathrm{E}\{y_j|\mathcal{H}_0\}|}{\sqrt{\mathrm{var}(y_j|\mathcal{H}_0)}}, \tag{4.18}$$

and under $\mathcal{H}_0$ we have $y_j \sim \mathcal{N}(0, \sigma^2)$. Substituting (4.10) into (4.18) yields

$$d = \frac{\mu_\theta}{\xi_0 L}\sqrt{\frac{\lambda_1}{\sigma^2}}\sum_{\xi=1}^{\xi_0} \xi \frac{\xi + L - 1}{L + 3\xi - 3}, \tag{4.19}$$

which is proportional to μ_θ and to the square root of effective SNR $\frac{\lambda_1}{\sigma^2}$, where$\lambda_1 = \frac{\lambda_1^*}{L}$. Therefore, we expect improved power for a likelihood ratio test (LRT) performed on $\boldsymbol{y}(1)$ as either the inherent contrast μ_θ or effective SNR increase. Furthermore, the posterior probabilities $\{p_{I_i|\boldsymbol{Y}(1)}\}$ are defined as

$$p_{I_i|\boldsymbol{y}(1)} = \frac{p_i f_1(\boldsymbol{y}(1))}{p_i f_1(\boldsymbol{y}(1)) + (1 - p_i) f_0(\boldsymbol{y}(1))}, \tag{4.20}$$

where $f_j(\boldsymbol{y}(1))$ denotes the probability density function (pdf) of a specific observation conditioned on $\mathcal{H}_j$ for $j = 0, 1$. From (4.20) we have $p_{I_i|\boldsymbol{y}(1)} = 0$ when $p_i = 0$. If $p_i > 0$ we rearrange $p_{I_i|\boldsymbol{y}(1)}$ as

$$p_{I_i|\boldsymbol{y}(1)} = \frac{1}{1 + \frac{1-p_i}{p_i}\frac{f_0(\boldsymbol{y}(1))}{f_1(\boldsymbol{y}(1))}}. \tag{4.21}$$

Note that $\frac{1-p_i}{p_i}\frac{f_0(\boldsymbol{y}(1))}{f_1(\boldsymbol{y}(1))} \propto \frac{1}{\text{LRT}}$, hence large d implies $\frac{1}{\text{LRT}} \to 0$ with high probability under $\mathcal{H}_1$, or, equivalently $p_{I_i|\boldsymbol{y}(1)} \to 1$. Moreover, using a Gaussian approximation for $f_1(\boldsymbol{y}(1))$ we estimate $\frac{1-p_i}{p_i}\frac{f_0(\boldsymbol{y}(1))}{f_1(\boldsymbol{y}(1))}$ as follows

$$\frac{1-p_i}{p_i}\frac{f_0(\boldsymbol{y}(1))}{f_1(\boldsymbol{y}(1))} \cong \frac{1-p_i}{p_i}\frac{\sigma_1}{\sigma}e^{\frac{\mu_1^2}{2\sigma_1^2}}\exp\left\{-\frac{y_i^2(1)}{2}\frac{\sigma_1^2-\sigma^2}{\sigma_1^2\sigma^2} - y_i(1)\frac{\mu_1}{\sigma_1^2}\right\} \tag{4.22}$$

$$= C_1 \exp\left\{-\frac{y_i^2(1)}{2}C_2 - y_i(1)\frac{\mu_1}{\sigma_1^2}\right\}, \tag{4.23}$$

where $C_1 = \frac{1-p_i}{p_i}\frac{\sigma_1}{\sigma}e^{\frac{\mu_1^2}{2\sigma_1^2}}$ and $C_2 = \frac{\sigma_1^2-\sigma^2}{\sigma_1^2\sigma^2} > 0$ since $\sigma_1^2 > \sigma^2$.

Claim 3. *For every small p_i we can bound C_1 and use the Gaussian approximation of $f_1(\boldsymbol{y}(1))$ to show that $p_{I_i|\boldsymbol{y}(1)} \to 1$ with probability under $\mathcal{H}_1$. Similarly, $p_{I_i|\boldsymbol{y}(1)} \to 0$ with probability under $\mathcal{H}_0$.*

Proof. See Appendix 4-B for the proof. □

Note that the Gaussian approximation is useful for low SNR, low contrast, or small $\frac{\xi_0}{L}$ when the different modes of the Gaussian mixture are both centered relatively close and/or have a high variance. At high SNR we can use a simpler bound to show convergence. Assume that if a target exists it has length $\xi_j = \xi_0 = 1$. Hence $\boldsymbol{y}_j(1)|\mathcal{H}_1 \sim \mathcal{N}(\frac{\sqrt{\lambda_1}}{L}\mu_\theta, \sigma^2 + \lambda_1\sigma_\theta^2)$ and the posterior probability $p_{I_i|\boldsymbol{y}(1)}|\mathcal{H}_1 \to 1$ since it follows the performance of an LRT for a simple binary hypothesis testing problem with a Gaussian distribution. Recall that an LRT is the uniformly most powerful test for distinguishing between two Gaussians with $\mu_1 > \mu_0$, where μ_r is the mean conditioned on $\mathcal{H}_r$ and $r = 0, 1$.

In Chapter III we used the asymptotic consistency property of $p_{I_i|\boldsymbol{y}(1)}$ to show that the threshold parameter k_0 converges to the true sparsity of the scanned domain

$(1-p)Q$. For λ_M we can provide an asymptotic bound on k_0 with probability. The logic of the bound is that p can be used to bound K, the number of supports $\mathcal{X}_j$'s for which $\mathcal{H}_1$ is true, then $k_0 \geqslant Q - KL$. Therefore, we have the following:

Claim 4. *The normalized number of samples N^* used by M-ARAP is upper bounded by $\frac{1}{L} + pL$ with probability for*[1] *$(\mu_1, \sigma_1^2) \in \mathcal{R}$, where p is the true sparsity of the underlying domain, i.e.,*

$$\lim_{\mathrm{SNR}\to\infty} \Pr\left(\frac{1}{L} \leqslant N^* \leqslant \frac{1}{L} + pL\right) = 1, \tag{4.24}$$

for some large Q.

Proof. To prove (4.24) note first that

$$N^* = \frac{M_1 + |\widehat{\Psi}|}{Q} = \frac{1}{L} + \frac{|\widehat{\Psi}|}{Q} \geqslant \frac{1}{L}, \tag{4.25}$$

hence the left side of the inequality is trivial and it suffices to show

$$\lim_{\mathrm{SNR}\to\infty} \Pr\left(\frac{|\widehat{\Psi}|}{Q} \leqslant pL\right) = 1. \tag{4.26}$$

In other words, we have to show that there exists Q_0 and SNR_0 such that for all $\mathrm{SNR} > \mathrm{SNR}_0$, $Q > Q_0$, and for an arbitrary small $\epsilon > 0$

$$\Pr\left(\frac{|\widehat{\Psi}|}{Q} \leqslant pL + \epsilon\right) = 1. \tag{4.27}$$

Note that $|\widehat{\Psi}| \sim \mathrm{B}(Q, p_0)$ is a binomial random variable with $\mathrm{E}\{|\widehat{\Psi}|\} = p_0 Q$, where p_0 is the expected proportion of supports $\mathcal{X}_j$'s for which $\mathcal{H}_1$ is true. Since we assumed

[1] the convergence region of $p_{I_i|\boldsymbol{y}(1)}$

a sparse domain with $\mathrm{E}\{\frac{|\Psi|}{Q}\} = p \ll 1$, we have $p_0 \leqslant pL$, where $pL < 1$ represents a worst case scenario when all targets are of length one and each resides in a different support. Therefore: (i) from Appendix 4-B we have $p_{I_i|\boldsymbol{y}(1)}|\mathcal{H}_r \to r$ with probability, for high SNR, where $r = 0, 1$. (ii) In Chapter III we showed that k_0 converges in probability to the transition value in the asymptotically Bernoulli sequence $p_{I_i|\boldsymbol{y}(1)}$. (iii) Therefore $\frac{|\widehat{\Psi}|}{Q} \to p_0 \leqslant pL$ in the mean square sense, for large Q, since $\mathrm{var}(\frac{|\widehat{\Psi}|}{Q}) = \frac{p_0(1-p_0)}{Q}$. This completes the proof. □

Claim 5. *The expected proportion of the area that is scanned by M-ARAP at the second stage but does not contain targets is bounded with probability, i.e.,*

$$\Pr\left(\frac{|\Psi \vartriangle \widehat{\Psi}|}{Q} \leqslant p(L - \mathrm{E}\{\Xi\})\right) = 1. \tag{4.28}$$

Proof. From *Claim 3* we know that for high SNR and large Q the posterior probabilities can be approximated as a Bernoulli sequence with $\Pr(p_{I_i|\boldsymbol{y}(1)} = 1) = p_0$. In Chapter III we showed that $k_0 \to \tilde{k}$, where $\tilde{k}$ marks the transition point in the ordered posterior probabilities sequence. Let $H(j) \equiv \mathcal{H}_r$ denote the fact that $\mathcal{H}_r$ is true for support $\mathcal{X}_j$ and $r = 0, 1$. Hence, we have

$$\forall i \in \mathcal{X}_j \quad , i \in \Psi \Rightarrow H(j) \equiv \mathcal{H}_1 \Rightarrow p_{I_i|\boldsymbol{y}(1)} \to 1, \tag{4.29}$$

which, in turn provides $i \in \widehat{\Psi}$, or $\Psi \subseteq \widehat{\Psi}$. Therefore,

$$\mathrm{E}\left\{\frac{|\Psi \vartriangle \widehat{\Psi}|}{Q}\right\} = \frac{\mathrm{E}\{|\Gamma|\}}{Q}, \tag{4.30}$$

where $\Gamma = \widehat{\Psi} \setminus \Psi = \{i : i \in \widehat{\Psi}, i \notin \Psi\}$. Note that,

$$\mathrm{E}\{|\Gamma|\} = \mathrm{E}\left\{\sum_{j=1}^{Q/L} I(H(j) \equiv \mathcal{H}_1)(L - \xi_j)\right\}, \tag{4.31}$$

and ξ_j denotes the cluster length in support $\mathcal{X}_j$. Since the indicator $I(H(j) \equiv \mathcal{H}_1)$ and the cluster length ξ_j are independent r.v., and $\Pr(I(H(j) \equiv \mathcal{H}_1)) = p_0$ we obtain

$$\mathrm{E}\{|\Gamma|\} = p_0 \left(L\frac{Q}{L} - \sum_{j=1}^{Q/L} \mathrm{E}\{\xi_j\}\right). \tag{4.32}$$

We further assume that the ξ_j's are i.i.d. and hence

$$\mathrm{E}\{|\Gamma|\} = p_0 \frac{Q}{L}(L - \mathrm{E}\{\Xi\}). \tag{4.33}$$

Let Z be a binomial r.v. with $Z \sim \mathrm{B}(Q, p_0)$, then, by construction $\frac{|\Gamma|}{Q} = \frac{L - E\{\Xi\}}{LQ} Z$ and hence

$$\mathrm{var}\left(\frac{|\Gamma|}{Q}\right) = \frac{(L - E\{\Xi\})^2}{L^2 Q} p_0(1 - p_0). \tag{4.34}$$

Therefore, $\frac{|\Gamma|}{Q} \to \frac{p_0}{L}(L - \mathrm{E}\{\Xi\})$ in the mean square sense for large Q. Since $p_0 < pL$ we obtain the desired result. This completes the proof. □

Corollary 4.1. *Claim 5 provides a bound on the cardinality of the redundant support which M-ARAP covers at the second stage. Not surprisingly, this bound is a function of both L and Ξ. In a sense this is the penalty we pay for using large L compared to the expected cluster length $\mathrm{E}\{\Xi\}$. This bound can be used to evaluate the tradeoff between reducing the number of measurements and expected estimation gains. To do*

so, we use our cost function (4.5) and assume that asymptotically we learn $\widehat{\Psi}$ *at almost no cost. Thus, an optimal allocation policy is given by*

$$\lambda_{M_o}(i,2) = \begin{cases} \frac{\lambda_T}{|\widehat{\Psi}|}, & i \in \widehat{\Psi} \\ 0, & \text{otherwise} \end{cases}. \tag{4.35}$$

Hence, $\Lambda_{M_o}(i) = \frac{\lambda_T}{|\widehat{\Psi}|} I(i \in \widehat{\Psi})$ and with $\nu = 1$ the expected cost is

$$J(\Lambda_{M_o}) = \sum_{i \in \Psi} \frac{1}{\lambda_T / |\widehat{\Psi}|} = \frac{|\widehat{\Psi}||\Psi|}{\lambda_T}. \tag{4.36}$$

At the same time, the cost associated with an exhaustive search policy $\Lambda_U(i) = \frac{\lambda_T}{Q}$ equals $J(\Lambda_U) = \frac{Q|\Psi|}{\lambda_T}$. Define the gain function

$$G(\Lambda) = 10 \log \frac{J(\Lambda_U)}{J(\Lambda)}, \tag{4.37}$$

then

$$G(\Lambda_{M_o}) = 10 \log \frac{Q}{|\widehat{\Psi}|} = 10 \log \frac{Q}{|\Psi| + |\Psi \triangle \widehat{\Psi}|}. \tag{4.38}$$

Using *Claim 5* we obtain

$$G(\Lambda_{M_o}) \geqslant 10 \log \frac{Q}{pQ(1 + L - \mathrm{E}\{\Xi\})} = 10 \log \frac{1}{p} - 10 \log(1 + L - \mathrm{E}\{\Xi\}), \tag{4.39}$$

where $10 \log(1 + L - \mathrm{E}\{\Xi\})$ is the gain penalty that we pay due to multiscaling.

4.3.3 Second search policy

Our second search policy is based on the sub-optimal search method λ_{so} given in Chapter III. However, we keep k_0 fixed and hence name this policy λ_{M_f}. This means

that we start by rank ordering $\lambda_{M_f}(\tau(i),2)$ according to

$$\lambda_{M_f}(\tau(i),2) = \frac{\lambda_T - Q\frac{\lambda_1^*}{|\mathcal{X}_j|}}{\sum_{j=1}^{Q}\sqrt{w_{\tau(j)}}}\sqrt{w_{\tau(i)}}, \tag{4.40}$$

then set

$$\lambda_{M_f}(\tau(i),2) = 0, \quad \forall i \leqslant k_0. \tag{4.41}$$

To complete the definition of λ_{M_f}, let

$$\lambda_1^* = \arg \min_{\lambda_1 \in \left(0, \frac{\lambda_T}{M_1}\right)} \mathrm{E}\left\{\sum_{i=1}^{Q} \frac{\nu I_i + (1-\nu)(1-I_i)}{\Lambda(i)}\right\}, \tag{4.42}$$

where $\Lambda(i) = \frac{\lambda_1}{L} + \lambda(i,2)$ and $\lambda(i,2)$ is defined via (4.40) and (4.41) substituting λ_1 for λ_1^* in the former.

4.4 Performance and comparison of search algorithms

4.4.1 Estimation

Assume that cell l belongs to the ROI or equivalently $I_l = 1$. Our next goal is to estimate θ_l using the measurement pair $(\boldsymbol{y}(1), \boldsymbol{y}(2))$. Let $\theta_l \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, we use a Bayesian framework to estimate θ_l based on its prior distribution. The optimal estimator minimizing the MSE is the conditional mean estimator (CME). We compare the performance of the CME for M-ARAP, $\mathrm{E}\{\theta_l|\boldsymbol{y}(1), y_l(2)\}$, to the CME $\mathrm{E}\{\theta_l|\boldsymbol{y}(0)\}$ for an exhaustive search policy, with

$$y_i(0) = \sqrt{\lambda_0}\theta_i I_i + n_i(0), \quad n_i(0) \sim \mathcal{N}(0, \sigma^2) \tag{4.43}$$

and $\lambda_0 = \frac{\lambda_T}{Q}$. The MSE of the CME for an exhaustive search policy is given by

$$\mathrm{var}\{\theta_l|y_l(0)\} = \sigma_\theta^2 - \frac{\lambda_0\sigma_\theta^4}{\sigma^2 + \lambda_0\sigma_\theta^2} = \frac{\sigma_\theta^2}{1 + \lambda_0\frac{\sigma_\theta^2}{\sigma^2}}. \tag{4.44}$$

Let $l \in \mathcal{X}_j$, then $y_l(1)$ is given as an interpolated version of $\tilde{y}_j(1)$. Consider a zero order hold interpolation scheme, i.e., $y_i(1) = \tilde{y}_j(1)$ for all $i \in \mathcal{X}_j$, then

$$y_l(1) = \frac{\sqrt{\lambda_1}}{L} \sum_{i=(j-1)L+1}^{jL} \theta_i I_i + n_l(1), \quad (j-i)L < l \leqslant jL. \tag{4.45}$$

Furthermore, $I_l = 1$ implies that $\mathcal{H}_1$ is true for $\mathcal{X}_j$ and hence $y_l(1)$ is a Gaussian mixture with mean and variance defined in (4.10) and (4.11) respectively. We approximate the distribution of $y_l(1)$ using Gaussian distribution with

$$y_l(1) \sim \mathcal{N}(\mu_1, \sigma_1^2). \tag{4.46}$$

The conditional distribution of $y_l(2)$ given $\boldsymbol{y}(1)$ is also Gaussian and defined as

$$y_l(2) \sim \mathcal{N}\left(\sqrt{\lambda_M(i,2)}\mu_\theta, \sigma^2 + \lambda_M(i,2)\sigma_\theta^2\right). \tag{4.47}$$

The competing estimator is a Naive Bayes estimator [62] of $\mathrm{E}\{\theta_l|\boldsymbol{y}(1), y_l(2)\}$, which is derived under the assumption that $(\boldsymbol{y}(1), y_l(2))$ defined in (4.46) and (4.47), respectively, are independent. This is not optimal but serves as a good comparison benchmark. Let $\boldsymbol{v}_l = [y_l(1)\ y_l(2)]'$, then the Naive Bayes estimator is given by

$$\hat{\theta}_l \stackrel{\triangle}{=} \mathrm{E}\{\theta_l|\boldsymbol{v}_l\} = \mu_\theta + \mathrm{cov}(\theta_l, \boldsymbol{v}_l)\mathrm{cov}^{-1}(\boldsymbol{v}_l)(\boldsymbol{v}_l - \mathrm{E}\{\boldsymbol{v}_l\}). \tag{4.48}$$

Note that

$$\mathrm{cov}(\theta_l, \boldsymbol{v}_l) = \sigma_\theta^2 \left[\frac{\sqrt{\lambda_1}}{L} \quad \sqrt{\lambda_M(l,2)}\right], \tag{4.49}$$

$$\mathrm{cov}(\boldsymbol{v}_l) = \begin{bmatrix} \sigma_1^2 & \frac{\sqrt{\lambda_1\lambda_M(l,2)}}{L}\sigma_\theta^2 \\ \frac{\sqrt{\lambda_1\lambda_M(l,2)}}{L}\sigma_\theta^2 & \sigma^2 + \lambda_M(l,2)\sigma_\theta^2 \end{bmatrix}, \tag{4.50}$$

and therefore

(4.51)

$$\hat{\theta}_l = \mu_\theta + \frac{\sigma^2 \frac{\sqrt{\lambda_1}}{L} \left(y_l(1) - \mu_1\right) + \sqrt{\lambda_M(l,2)} \left(\sigma_1^2 - \frac{\lambda_1 \sigma_\theta^2}{L^2}\right) \left(y_l(2) - \sqrt{\lambda_M(l,2)}\mu_\theta\right)}{\sigma_1^2(\sigma^2 + \lambda_M(l,2)\sigma_\theta^2) - \frac{\lambda_1 \lambda_M(l,2)}{L^2}\sigma_\theta^4} \sigma_\theta^2.$$

We conducted multiple simulations and representative results are seen in Figs. 4.1-4.3. In both figures 4.1 and 4.3 we plot the MSE performance gain $g(\lambda)$, defined as

(4.52)
$$g(\lambda) = 10 \log \frac{\mathrm{var}(\theta_l | y_l(0))}{\mathrm{MSE}(\hat{\theta}_l)}$$

as a function of SNR (Fig. 4.1) and the detectability index given in (4.19) (Fig. 4.3). Monte-Carlo simulations were used to estimate the MSE of (4.51). We chose $Q = 12,000$, $p = \frac{1}{1000}$, and each point on the figure represents an average over θ based on 500 realizations. We let signal to noise ratio, defined as $10 \log \frac{\lambda_T/Q}{\sigma^2}$, vary from 0 to 40 [dB], used contrast level $\mu_\theta \in \{0.5, 1, 2, 4, 8\}$, and set $\sigma_\theta^2 = \frac{1}{16}$. Different lengths L were simulated for the first stage, but we present here the cases of $L = 8$ and $L = 32$ since it is enough to understand the general trends. Maximal cluster length $\xi_0 = 6$ was chosen and $\xi \sim U\{1, 2, 3, 4, 5, 6\}$ was simulated. Curves with different markers represent different contrast level μ_θ. Figure 4.2 show curves of the normalized number of samples N^* for the same settings.

Note that as opposed to $ARAP$ we do not claim optimality of M-ARAP and indeed the optimal gain of 30 [dB] is not realized. Moreover, asymptotic gains decreases as L increases. This is natural since the posterior probabilities $p_{I_i|\boldsymbol{y}(1)}$ are identical within each support. Hence, if the resource allocation scheme λ_M suspects that a target exists in $\mathcal{X}_j$, all cells within this support receive the same effort allocation for

the second stage. As the difference $(L - \mathrm{E}\{\Xi\})$ increases, this translates to wasted resources according to *Claim 5*. Fig. 4.1 shows asymptotic gain of 25 [dB] for $L = 8$ and 20 [dB] for $L = 32$ both higher than (4.39) which yielded 22.6 and 15.3 [dB] respectively. Therefore, we conclude that our bound (4.28) is not tight.

Similar phenomenon was observed by Posner in [46] when he considered a multi-scale framework to minimize the expected search time. He concluded that there are no benefits for multiscaling when the goal is to minimize search time. However, his work had accounted for the case of a single target, a restriction that we have relaxed here. Moreover, the normalized number of samples N^* is lower bounded by $\frac{1}{L}$, hence there is a tradeoff between possible estimation gain and reducing the overall number of measurements as suggested in *Corollary 4.1*. We show here that if the contrast is high enough ($\mu_\theta > 2$ in our case) we can both save measurements (according to *claim 4*) and enjoy significant estimation gain within the ROI. Figure 4.2 shows the expected saving in measurements or N^* for the scenario depict in Fig. 4.1. Solid curves represent $L = 8$ while dash-dot curves represent $L = 32$, with different contrast levels identified using the same markers as in Fig. 4.1. Combining the information on both figures shows measurement saving per estimation gain. For example for SNR of 15 [dB] and relatively low contrast level of $\mu_\theta = 2$, M-ARAP with $L = 8$ yields about 10 [dB] performance gain in estimation while using only 14% of the number of samples used by an exhaustive search. Similar performance gain is achieved by M-ARAP for $L = 32$ with $\mu_\theta = 4$ and about 22% of the samples. We used the fixed amount of samples version of our adaptive method, λ_{M_f}, given in (4.40) and (4.41) with values of $N^* \in (0.15, 0.25)$ but did not see significant differences compared to the results of λ_M displayed here.

In Fig. 4.3 we plot estimation gains vs. the detectability index since it incorpo-

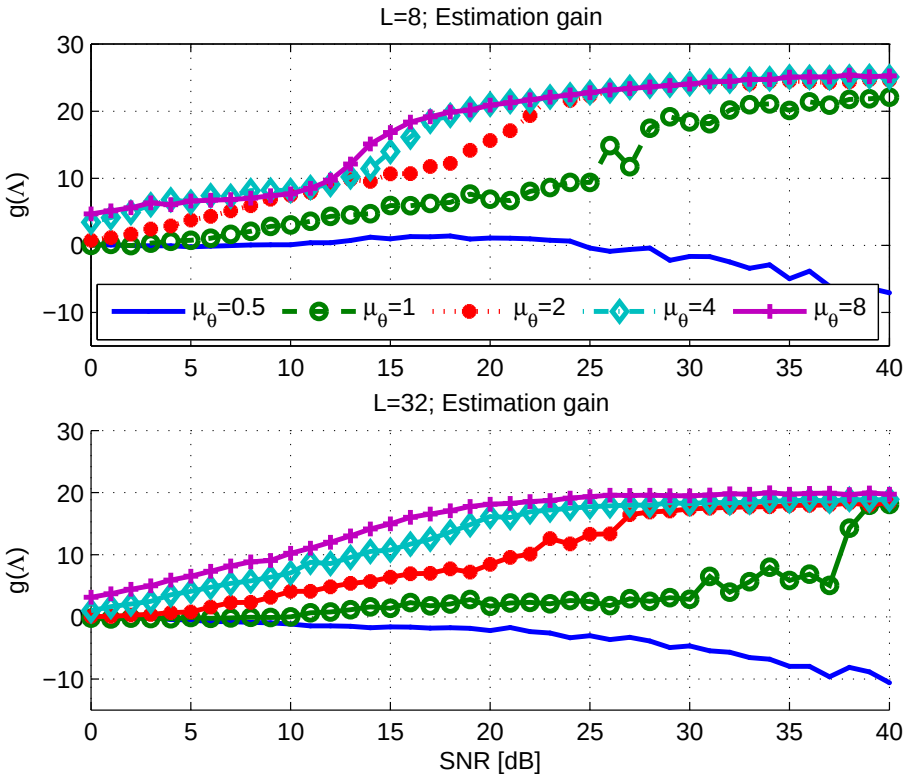

Figure 4.1: We plot estimation gains as a function of SNR for different contrast levels. The upper plot show gains for $L = 8$ while the lower plot show gains for $L = 32$. Note that without sufficient contrast ($\mu_\theta = \frac{1}{2}$), λ_M results in performance loss. However, for high contrast significant gains of 10 [dB] are achieved at SNR values less than 15 [dB]. At the same time we only use about 15% of the samples compared to an exhaustive search. Note that the asymptotic lower bound on the gain (4.39) yields 22.6 [dB] and 15.3 [dB] for $L = 8$ and $L = 32$ respectively. Since in both plots the gains exceed the bound we conclude that the bound is not tight.

rates both the contrast level and the SNR in a single parameter. One would expect similar gains for similar detectability index values regardless of the actual contrast or SNR. However, this is not necessarily the case here. Fig. 4.3 displays estimation gains vs. detectability index for $L = 8$ and $L = 32$ in the upper and lower plots respectively. While $L = 32$ results in the expected behavior, this is clearly not the case for $L = 8$. The reason is that the detectability index was derived using a single Gaussian approximation for the Gaussian mixture pdf of $\boldsymbol{y}(1)$. This approximation is valid when the different modes of the mixture are relatively close by, which is the case for $L = 32$. As conditions improves, i.e., either SNR or contrast increase or

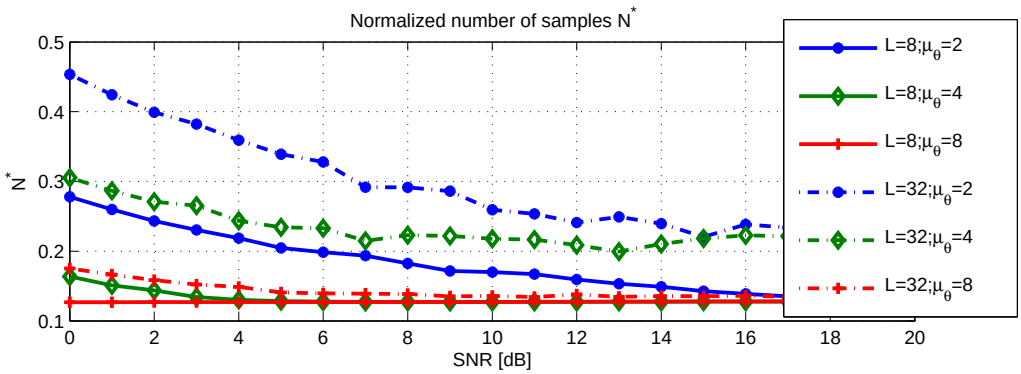

Figure 4.2: We plot the normalized number of samples N^* as a function of SNR for $L = 8$, $L = 32$, and different contrast levels $\mu_\theta \in \{2, 4, 8\}$. These N^* values are associated with estimation gains seen in Fig. 4.1 for SNR values ranging from 0 to 20 [dB] (the left half of the SNR axis in Fig 4.1). For example for a relatively low contrast of $\mu_\theta = 2$, SNR of 15 [dB], and $L = 8$, estimation performance gain of 10 [dB] is achieved with only 14% of the sampling used by exhaustive search.

L get closer to ξ_0, the approximation does not hold anymore. What we see in the upper plot is that performance is actually better than what the detectability index predicts. Additional discussion regarding the statistical model used here is given in Section 4.4.3.

4.4.2 Detection

Non-adaptive detection Consider the problem of correctly detecting whether cell l contains a target based on a sample $y_l(0)$. As before, we assume the samples follow the model (4.43) and that $\theta_l \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$ are i.i.d.. Thus, for an exhaustive search policy $y_l(0) \sim \mathcal{N}(\sqrt{\lambda_0}\mu_\theta I_l, \sigma_{y_r}^2)$. Given $y_l(0)$, the measurement of pixel l, our goal is to decide between

$$\begin{aligned} H_0 &: y_l(0) \sim \mathcal{N}(0, \sigma^2), \\ H_1 &: y_l(0) \sim \mathcal{N}(\sqrt{\lambda_0}\mu_\theta, \sigma^2 + \lambda_0\sigma_\theta^2). \end{aligned} \tag{4.53}$$

For a known $\sigma_{y_r}^2$, the uniformly most powerful test for this binary hypothesis testing problem is a likelihood ratio test (LRT). The performance of this non-adaptive LRT

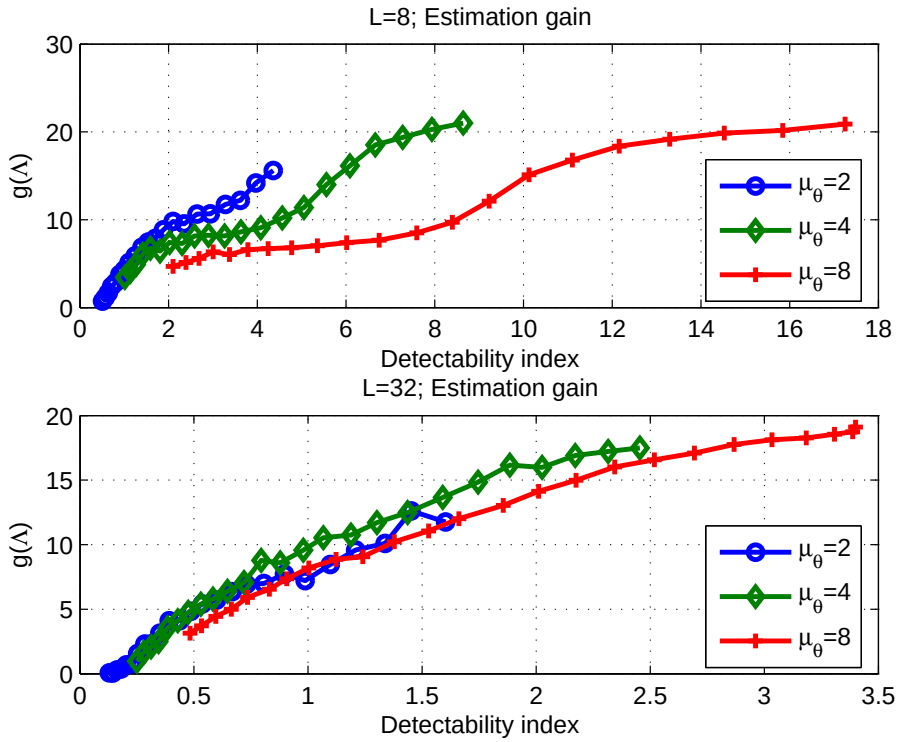

Figure 4.3: We plot estimation gains vs. detectability index. Note that for large L (lower plot) the detectability index can be used to predict performance gain regardless of the actual contrast or SNR. This is not the case when L is close to the expected maximum target length ξ_0 as seen in the upper plot for $L = 8$.

in terms of its receiver operating characteristic (ROC) curve is easily calculated. The power β of this level α LRT is

$$\beta = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \sqrt{\frac{\lambda_0 \mu_\theta^2}{\sigma^2}}\right), \tag{4.54}$$

where $\Phi(\cdot)$ is the normal cumulative distribution function [32].

Adaptive detection Using Bayes rule, the likelihood function $f(\boldsymbol{y}(2), \boldsymbol{y}(1); \boldsymbol{I}_\Psi)$ equals

$$f(\boldsymbol{y}(2), \boldsymbol{y}(1)) = f(\boldsymbol{y}(2)|\boldsymbol{y}(1)) f(\boldsymbol{y}(1)), \tag{4.55}$$

but, given $\boldsymbol{y}(1)$, the measurements at the second step are independent for different cells and thus

$$f(\boldsymbol{y}(2), \boldsymbol{y}(1)) = \prod_{i=1}^{Q} f(y_i(2)|\boldsymbol{y}(1)) f(y_i(1)). \tag{4.56}$$

Therefore, the LRT statistic, T_l, is

$$T_l = \frac{\prod_{i=1}^{Q} f(y_i(2)|\boldsymbol{y}(1), I_l = 1) f(y_i(1)|I_l = 1)}{\prod_{i=1}^{Q} f(y_i(2)|\boldsymbol{y}(1), I_l = 0) f(y_i(1)|I_l = 0)} =$$

$$\frac{f(y_l(2)|\boldsymbol{y}(1), I_l = 1) f(y_l(1)|I_l = 1)}{f(y_l(2)|\boldsymbol{y}(1), I_l = 0) f(y_l(1)|I_l = 0)}. \tag{4.57}$$

Let $l \in \mathcal{X}_j$ and $y_l(1)$ is given in (4.45). Note that given $I_l = 1$, $y_l(1) \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $y_l(2) \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are defined in (4.46) and (4.47), respectively. Under the null hypothesis, $I_l = 0$, we have $y_l(2) \sim \mathcal{N}(0, \sigma^2)$ for the second measurement. However, for the first measurement we have to account for two scenarios: (a) $\mathcal{X}_j$ does not contain any targets. (b) $\mathcal{X}_j$ contains at most $\min(L-1, \xi_0)$ targets. Note that the probability of (a) is

$$\Pr\left(\prod_{i \in \mathcal{X}_j} (1 - I_i) = 1\right) = q_0 \cong (1-p)^L,$$

while the probability of (b) is $(1 - q_0)$. Since we are mainly interested in sparse scenarios where $p \ll 1$ we have q_0 close to one and therefore approximate the pdf of $y_l(1)|(I_l = 0)$ as a Gaussian mixture with

$$y_l(1)|(I_l = 0) \sim q_0 \mathcal{N}(0, \sigma^2) + (1 - q_0)\mathcal{N}(\mu_b, \sigma_b^2), \tag{4.58}$$

where (μ_b, σ_b^2) represent case (b) and are evaluated using (4.10) and (4.11), respectively, replacing ξ_0 with $\min(L-1, \xi_0)$. Furthermore, we approximate the r.h.s. of (4.58) as $\mathcal{N}(\mu_0, \sigma_0^2)$, where

$$\begin{aligned}\mu_0 &= (1-q_0)\mu_b, \\ \sigma_0^2 &= q\sigma^2 + (1-q)\sigma_b^2 + q(1-q)\mu_b^2.\end{aligned}$$

Collecting all the previous results we obtain:

$$H_0 : \begin{cases} y_l(1) \sim \mathcal{N}(\mu_0, \sigma_0^2) \\ y_l(2)|\boldsymbol{y}(1) \sim \mathcal{N}(0, \sigma^2) \end{cases}$$

(4.59)

$$H_1 : \begin{cases} y_l(1) \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ y_l(2)|\boldsymbol{y}(1) \sim \mathcal{N}(\mu_2, \sigma_2^2) \end{cases}.$$

Substituting these distributions into (4.57) provides the following LRT

$$T_l = \frac{\sigma\sigma_0}{\sigma_1\sigma_2} \frac{\exp\left\{-\frac{(y_l(2)-\mu_2)^2}{2\sigma_2^2} - \frac{(y_l(1)-\mu_1)^2}{2\sigma_1^2}\right\}}{\exp\left\{-\frac{y_l(2)^2}{2\sigma^2} - \frac{(y_l(1)-\mu_0)^2}{2\sigma_0^2}\right\}} \gtrless \gamma,$$

which can also be expressed as

$$T_l = \frac{y_l^2(2)}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\sigma_2^2}\right) + y_l(2)\frac{\mu_2}{\sigma_2^2} + \frac{y_l^2(1)}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) + y_l(1)\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2}\right) + \tag{4.60}$$

$$\frac{\mu_0^2}{2\sigma_0^2} - \frac{\mu_1^2}{2\sigma_1^2} - \frac{\mu_2^2}{2\sigma_2^2} + \log\frac{\sigma\sigma_0}{\sigma_1\sigma_2} \gtrless \gamma',$$

where $\gamma' = \log\gamma$. Note that $\lambda_M(l,2) = 0$ yields $\mu_2 = 0$ and $\sigma_2^2 = \sigma^2$, hence, for all cells in $\widehat{\Psi}^c$ the test (4.60) is a function of $y_l(1)$ alone. Next, we compare the

theoretical ROC curve (4.54) to the empirical ROC curve calculated for the adaptive LRT (4.60) performed on the data pair $(\boldsymbol{y}(1), \boldsymbol{y}(2))$ when λ_M was used as a search policy.

For simulations we used $Q = 12,000$, $p = \frac{1}{1000}$, and each point on the figures represents 500 Monte-Carlo runs. Detection probability was averaged over the entire ensemble and over all pixels inside and outside the ROI. We conducted multiple runs for varying SNR levels and observed that with $\nu = 1$ and high contrast the non-adaptive LRT (4.54) does much better than (4.60). Interestingly enough, for low contrast, low SNR, and $L \simeq \xi_0$ (support size roughly equals the maximal expected target length) the converse is true. Figs. 4.4 and 4.5 show ROC curves for SNR of 0 [dB], $\mu_\theta \in \{0.5, 1, 2, 4, 8\}$, and $\sigma_\theta^2 = \frac{1}{16}$ for $L = 8$ and $L = 32$, respectively. In all figures dash-dot curves represent a LRT performed on exhaustive search data (4.54) and different markers represent varying contrast levels μ_θ. The upper plot in Fig. 4.4, where $L = 8$ and SNR is 0 [dB], shows that for low contrast levels $\mu_\theta \leqslant 2$ (4.60) does better than the non-adaptive LRT. However, as soon as the contrast improves the situation is reversed (lower plot). When L is increased to 32 in Fig. 4.5 the non-adaptive LRT outperforms λ_M for all contrast levels we tried.

In Fig. 4.6 we set SNR of 10 [dB] and compare ROC curves for $L = 8$ and $L = 32$ in the upper and lower plots respectively. Both plots show contrast levels of $\mu_\theta \in \{0.5, 1, 2\}$. For $\mu_\theta \leqslant 1$ and $L = 8$ we get improved detection performance for λ_M for very low false alarm probability levels. Other than that, the LRT test (4.54) does much better than (4.60). This shows that detection performance on data acquired using our adaptive policy does not outperform the non-adaptive measuring scheme as SNR improves.

We conclude by pointing out that our cost function (4.5) with $\nu = 1$ is not optimal

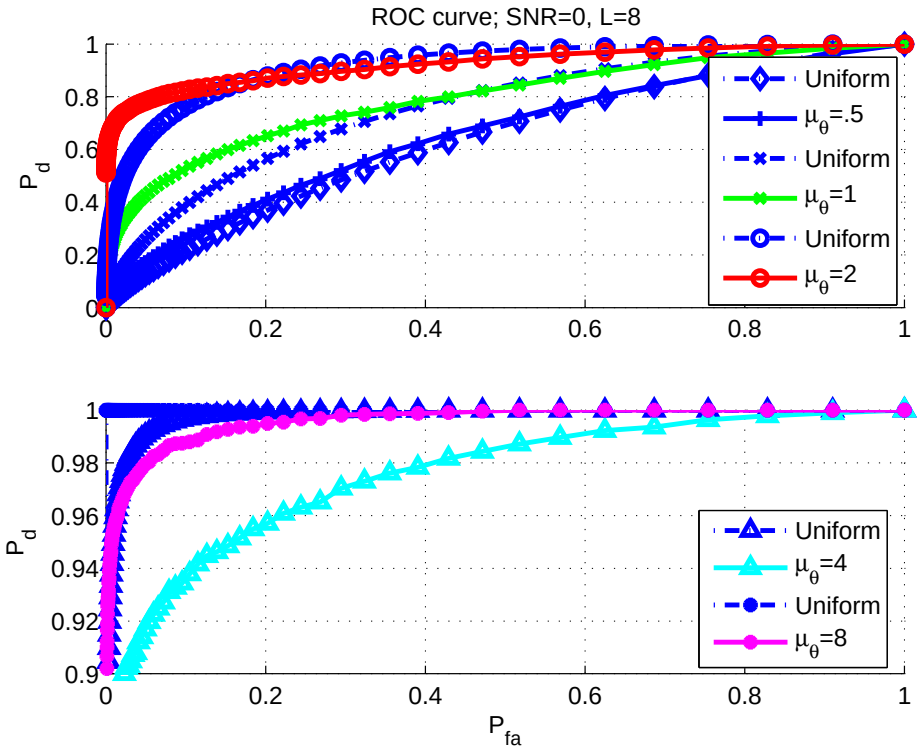

Figure 4.4: Receiver operating characteristic (ROC) curves for SNR of 0 [dB] and $L = 8$ of data acquired using M-ARAP vs. a non-adaptive exhaustive data acquisition. The upper plot shows low contrast levels of $\mu_\theta \leqslant 2$, while the lower plot focuses on high contrast levels. Dash-dot curves represent the ROC of the non-adaptive LRT and different markers represent contrast. In the upper plot we see improved detection performance of M-ARAP compared to the non-adaptive scheme. However, this is reversed for the lower plot.

for detection as was pointed out in Chapter III. It is possible to get improved detection performances when optimizing over ν. However, we leave this experiment to be done for a specific application.

4.4.3 Statistical models approximation

Most of the results in Section 4.3 rely on a Gaussian mixture (GM) statistical model and a Gaussian approximation for the GM given in (4.10) and (4.11) and developed in Appendix 4-A. In the following section we provide a feel for the statistical nature of the data and discuss the validity of the single Gaussian approximation of the GM. As seen in Fig. 4.3 this approximation, which supported the detectability

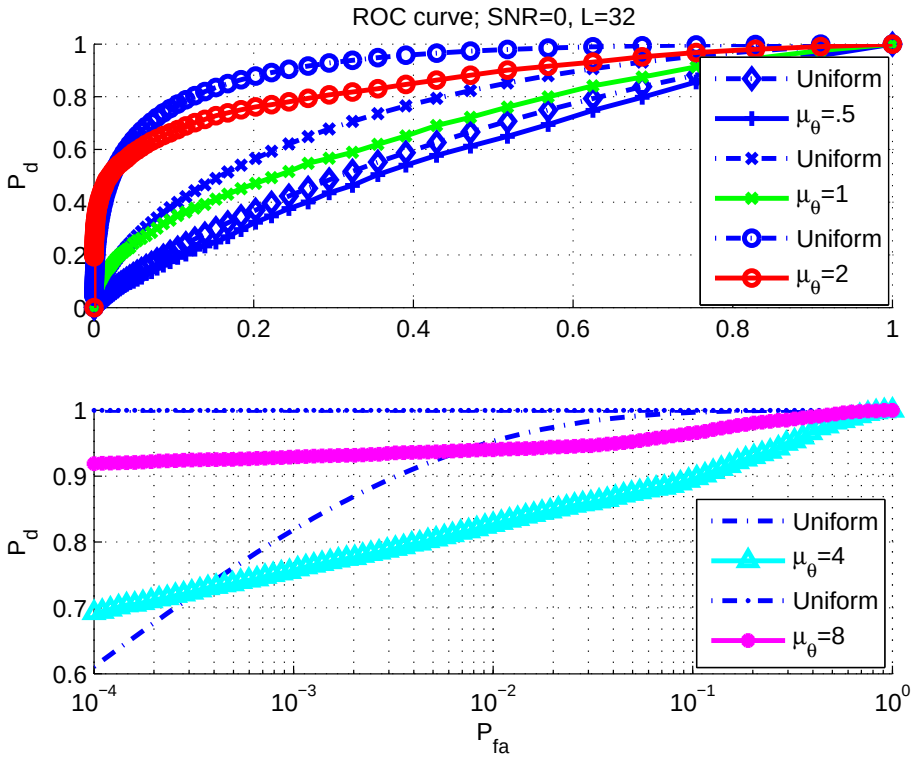

Figure 4.5: Receiver operating characteristic (ROC) curves for SNR of 0 [dB] and $L = 32$ of data acquired using M-ARAP vs. a non-adaptive exhaustive data acquisition. The upper plot shows low contrast levels of $\mu_\theta \leqslant 2$, while the lower plot focuses on high contrast levels. Dash-dot curves represent the ROC of the non-adaptive LRT and different markers represent contrast. The ROC curves due to the non-adaptive LRT dominate the ROC curves resulting from M-ARAP.

index analysis, is only accurate for low values of d or low SNR, small $\frac{\xi_0}{L}$, and low contrast μ_θ. We examine the pdf of $\boldsymbol{y}(1)|\mathcal{H}_1$ under different settings to better understand this phenomenon. To generate enough data under $\mathcal{H}_1$, we used simulations with $Q = 16,384$, $p = \frac{1}{100}$ and 500 Monte-Carlo runs.

We start by showing a scenario where the single Gaussian approximation is a reasonable approximation to the Gaussian mixture. With either $(L = 8, \mu_\theta = 2)$ or $(L = 32, \mu_\theta = 8)$ and SNR of 10 [dB] the detectability index (4.19) is roughly one. Figs. 4.7 and 4.8 show the probability distribution functions (pdf) and cumulative distribution functions (cdf) of $\boldsymbol{y}(1)|\mathcal{H}_1$ for these scenarios, respectively. The upper plot corresponds to $(L = 8, \mu_\theta = 2)$ while the lower plot corresponds to $(L = 32, \mu_\theta =$

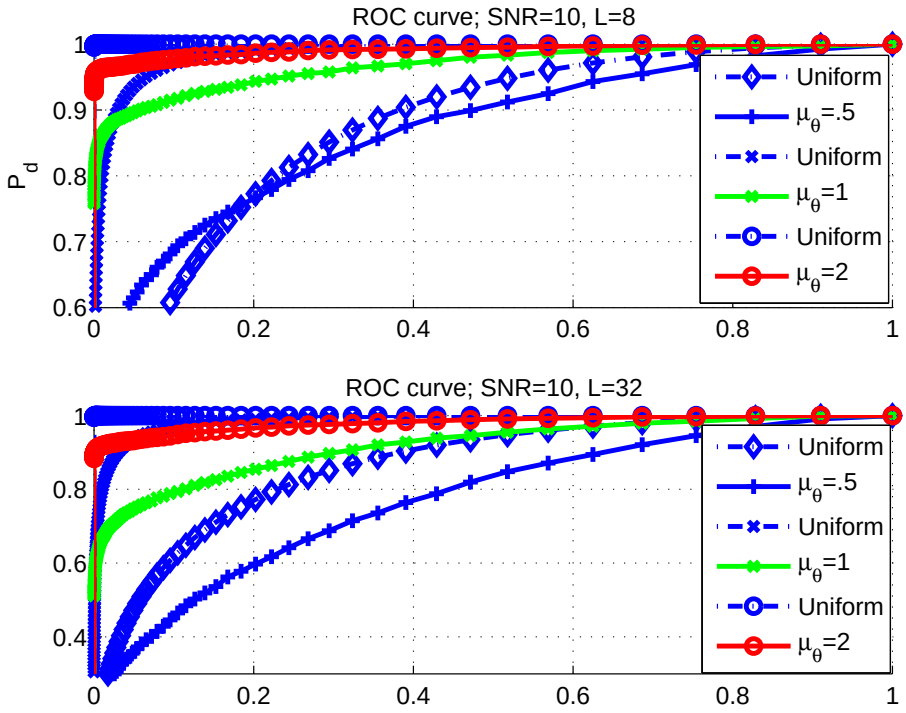

Figure 4.6: ROC curves for the two tests at SNR of 10 [dB] and low contrast levels of $\mu_\theta \leqslant 2$. In the upper plot we set $L = 8$, while $L = 32$ was chosen for the lower plot. Note that increasing SNR improves test performance of the non-adaptive scheme far better than for M-ARAP. This is mainly attributed to the fact that with $\nu = 1$, M-ARAP does not spend enough energy characterizing the alternative and focuses the sampling energy onto the ROI.

8) on both figures. While we are not interested in goodness-of-fit type results, it is clearly seen that the single Gaussian approximation is not far from the actual GM distribution.

On the other hand, Fig. 4.9 represents a case where the single Gaussian approximation begins to breakdown. In both plots SNR is 20 [dB] and the contrast level $\mu_\theta = 4$, while $L = 8$ and $L = 32$ for the upper and lower plots respectively. Both plots shows distinctive modes for the true GM and it is clear that a single Gaussian is no longer a reasonable representation of the true pdf. Nevertheless, this does not suggest that the posterior probabilities $p_{I_i|\boldsymbol{y}(1)}$ will no longer converge to either one or zero under the corresponding hypothesis $\mathcal{H}_r$. As mentioned in Section 4.3.2 and

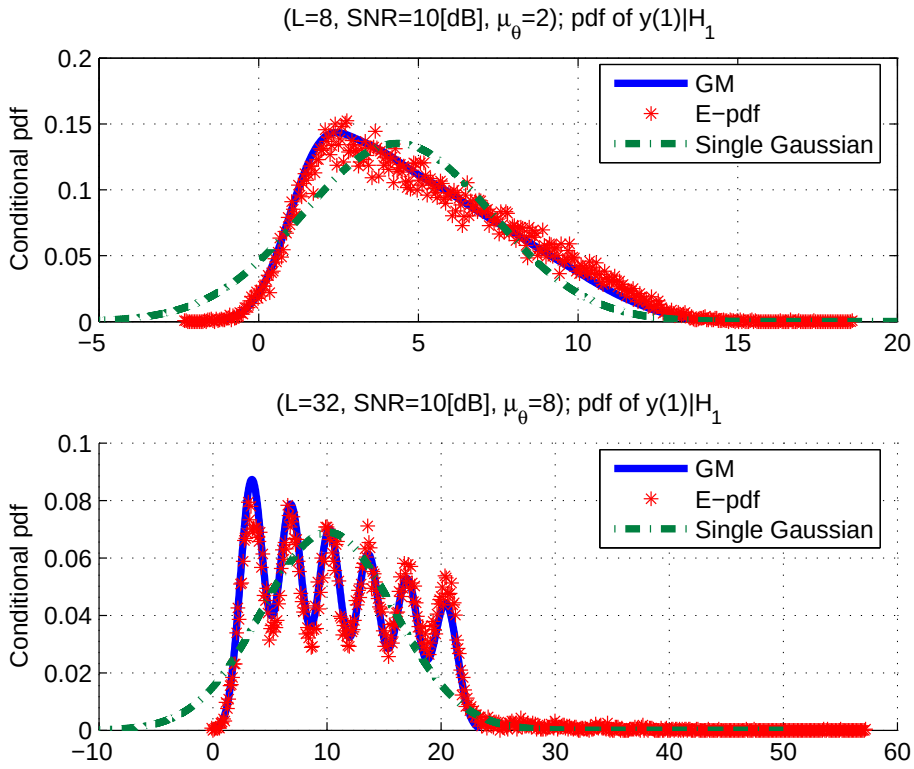

Figure 4.7: We plot the pdf of the predicted Gaussian mixture (solid blue), simulations histogram (asterisk red), and the single Gaussian approximation (dash-dot green) with mean and variance given in (4.10) and (4.11). For the upper plot we have $L = 8$ and $\mu_\theta = 2$, while for the lower plot $L = 32$ and $\mu_\theta = 8$. For both plots SNR is 10 [dB] and the detectability index (4.19) is roughly one.

Appendix 4-B, for this case we can use a worst case scenario to serve as a lower bound for the convergence rate of the $p_{I_i|\boldsymbol{y}(1)}$'s to the true value. This can be easily visualized from the left mode of the GM on the upper plot. This mode represents the appearance of a single target within a given support $\mathcal{X}_j$, i.e., a cluster of targets with length one. Still, as SNR and contrast improves the consistency property of the plug in estimator of the posterior probability, (4.20) holds.

4.5 Conclusions and future work

We hypothesized that for sparse signals with good inherent contrast we can improve on our previous search policy $ARAP$ by reducing the number of measurements

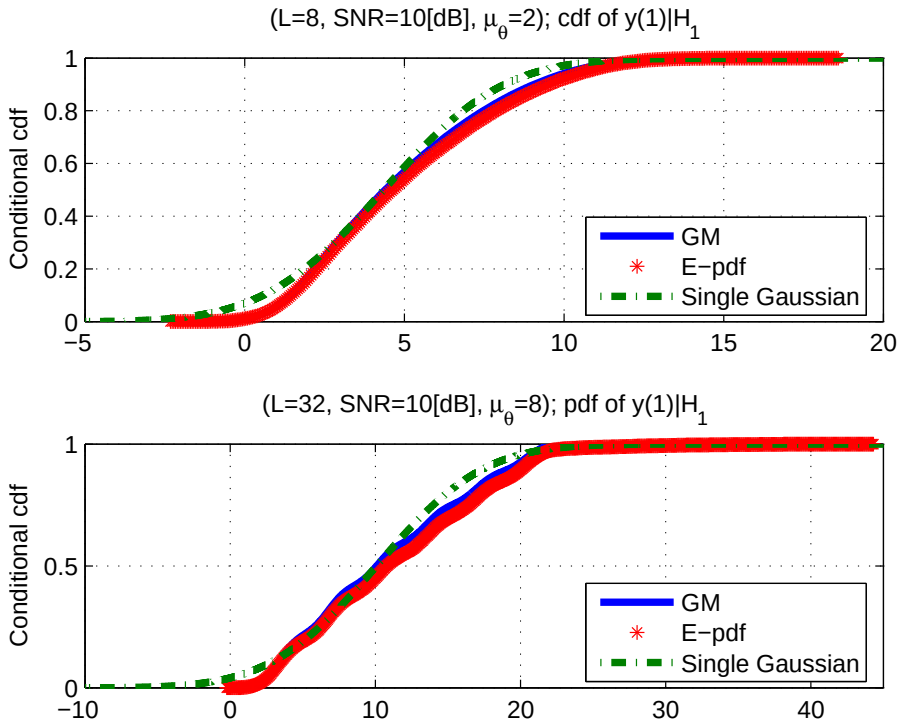

Figure 4.8: We plot the cdf of the predicted Gaussian mixture (solid blue), simulations histogram (asterisk red), and the single Gaussian approximation (dash-dot green) with mean and variance given in (4.10) and (4.11). For the upper plot we have $L = 8$ and $\mu_\theta = 2$, while for the lower plot $L = 32$ and $\mu_\theta = 8$. For both plots SNR is 10 [dB] and the detectability index (4.19) is roughly one. As one can see, the Gaussian approximation is not far from the true GM distribution.

yet preserving estimation gains. M-ARAP was derived using the same cost function introduced in Chapter III under additional total number of measurements or time constraints. We used multi-scaling where collections of cells were measured in the first step to save measurements and decide where we need to focus the measurement scheme and produce fine resolution at the second stage. We showed that significant saving in measurements can be achieved while maintaining high estimation gains. Specific examples showed 10-17 [dB] gain in estimation performance, for different contrast levels, using less than 20% of the samples needed to perform an exhaustive search, at SNR of 15 [dB]. Other methods capable of such saving in the sampling process are active sampling (AS) and compressive sensing (CS). However,

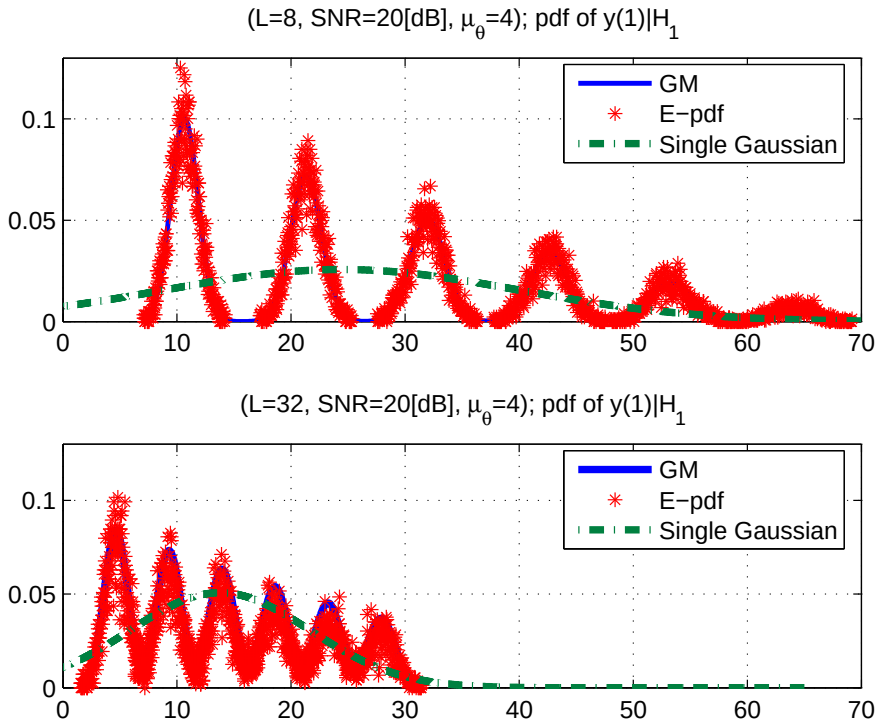

Figure 4.9: We plot the pdf of the predicted Gaussian mixture (solid blue), simulations histogram (asterisk red), and the single Gaussian approximation (dash-dot green) with mean and variance given in (4.10) and (4.11). For both plots SNR is 20 [dB] and $\mu_\theta = 4$, while $L = 8$ in the upper plot and $L = 32$ for the lower plot. Note that while the single Gaussian approximation in the lower plot is still a somewhat reasonable approximation of the GM this is not the case for the upper plot. This explains the different behavior exhibited by the curves in Fig. 4.3.

both methods lack the additional degree of freedom of 'resource allocation' that was suggested here. Since AS and CS consider the problem of "where to sample" without the "how to sample" part, there is limited, if any, estimation gain associated with either method. Moreover, both AS and CS where suggested for sampling spatially homogeneous signals. We do not impose such a constraint here and only require reasonable contrast between the ROI and the signal background. Previous work on blood pooling [16, 45, 53] considered detecting defective blood samples in a large population. The main concern is to shorten the search time and there is an inherent assumption of an infinite contrast [16]. In addition, as opposed to the clustering

scenario discussed here, blood pooling assumes that it is most likely that each pool contains a single target [53]. By deciding "how to sample" and considering clusters we improve estimation gains to all cells within the ROI.

Our methods displayed here have not been optimized for detection. Since we allow many degrees of freedom, we leave this optimization for future work. This optimization part may also be application dependent. We have yet to find a good application for our methods and this is another part of our intended future work. For some potential applications, like airport security screening or early detection of breast cancer tumors, tools that can use a multi-scale measurement scheme have to be developed. For other applications, like radar imaging or CT, sensor sensitivity or resolution may still need to be improved to achieve the potential benefits our method suggests.

4-A Appendix: Mean and variance of $Y_j(1)$ given $\mathcal{H}_1$

We derive the mean and variance of our measurement model (4.9) under the assumption that $\mathcal{X}_j$ contains a cluster of targets of length ξ_j. We further assume that $\xi_j \leqslant \xi_0 \leqslant |\mathcal{X}_j| = L$ and that $\Pr(\xi_j = k) = \frac{1}{\xi_0}$ for $k = 1, 2, \ldots, \xi_0$. In this case, $y_j(1)$ is a Gaussian mixture with ξ_0 components, each have a mean $m_i = \mu_\theta \frac{\sqrt{\lambda_1}}{L} i$ and variance $s_i^2 = \sigma^2 + i\frac{\lambda_1}{L^2}\sigma_\theta^2$. To evaluate the conditional mean of $Y_j(1)$ we enumerate all possible intersection of a targets with length ξ and supports of length L. Then, using the total probability theorem we obtain

$$\mathrm{E}\{y_j(1)|\mathcal{H}_1\} = \frac{\sqrt{\lambda_1}}{L}\mathrm{E}\left\{\mathrm{E}\left\{\sum_{i=(j-1)L+1}^{jL}\theta_i I_i \,\middle|\, \xi_j\right\}\right\} \tag{4-A.1}$$

$$= \mu_\theta \frac{1}{\xi_0} \frac{\sqrt{\lambda_1}}{L} \sum_{\xi=1}^{\xi_0} \frac{1}{L+\xi-1} \left(4 \sum_{i=1}^{\xi-1} i + (L-\xi+1)\xi \right) \tag{4-A.2}$$

$$= \frac{\mu_\theta}{\xi_0} \frac{\sqrt{\lambda_1}}{L} \sum_{\xi=1}^{\xi_0} \xi \frac{\xi+L-1}{L+3\xi-3} = \mu_{y_j(1)|\mathcal{H}_1} \triangleq \mu_1. \tag{4-A.3}$$

Note that the probability of each Gaussian q_i equals

$$q_i = \frac{L + 4\xi_0 + 1 - 5i}{\frac{3}{2}\xi_0^2 + (L - \frac{3}{2})\xi_0}. \tag{4-A.4}$$

Therefore, the conditional pdf is

$$f_{Y(1)|\mathcal{H}_1}(y) = \sum_{i=1}^{\xi_0} q_i \mathcal{N}(m_i, s_i^2), \tag{4-A.5}$$

where $\mathcal{N}(m_i, s_i^2)$ is the i'th Gaussian pdf with mean m_i and variance s_i^2. The conditional variance of $y_j(1)$ is given by

$$\text{var}\{y_j(1)|\mathcal{H}_1\} = \sum_{i=1}^{\xi_0} q_i \int (y-\mu_1)^2 \frac{1}{\sqrt{2\pi s_i^2}} \exp\left\{ -\frac{1}{2s_i^2}(y-m_i)^2 \right\} dy \tag{4-A.6}$$

$$= \sum_{i=1}^{\xi_0} q_i \left(\text{E}\{y_i^2\} - 2m_i\mu_1 + \mu_1^2 \right), \tag{4-A.7}$$

where $\text{E}\{y_i^2\} = s_i^2 + m_i^2$.

4-B Appendix: Showing $p_{I_i|y_i(1)} \to r$ under $\mathcal{H}_r$

We show $p_{I_i|y_i(1)}|\mathcal{H}_1 \to 1$ and the complementary claim can be proved in a symmetric manner. Recall the definition of $p_{I_i|\boldsymbol{y}(1)}$

$$p_{I_i|\boldsymbol{y}} = \frac{p_i f_1(\boldsymbol{y})}{p_i f_1(\boldsymbol{y}) + (1-p_i) f_0(\boldsymbol{y})} = \frac{1}{1 + \frac{1-p_i}{p_i} \frac{f_0(\boldsymbol{y})}{f_1(\boldsymbol{y})}}, \tag{4-B.1}$$

where for brevity the time dependency '(1)' was omitted. From (4.9) we note that the posterior probability depends on the specific y_j for all $i \in \mathcal{X}_j$ yielding

$$p_{I_i|\boldsymbol{y}} = \frac{1}{1 + \frac{1-p_i}{p_i}\frac{f_0(y_j)}{f_1(y_j)}} \stackrel{\triangle}{=} p_{I_i|y_j}. \tag{4-B.2}$$

Using the Gaussian approximation (4.22) we show that for an arbitrary small $\delta > 0$ and $\epsilon' > 0$ there exists $\tilde{\mu}_1$ and $\tilde{\sigma_1}^2$ such that for $(\mu_1, \sigma_1^2) \in \mathcal{R}$

$$\Pr(p_{I_i|y_j} > 1 - \epsilon') > 1 - \delta, \tag{4-B.3}$$

where $\mathcal{R} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geqslant \tilde{\mu}_1, x_2 \leqslant \tilde{\sigma_1}^2\}$. To prove (4-B.3) we note that $p_{I_i|y_j} > 1 - \epsilon'$ if and only if

$$\frac{1-p_i}{p_i}\frac{f_0(y_j)}{f_1(y_j)} < \epsilon'' = \frac{1}{1-\epsilon'} - 1. \tag{4-B.4}$$

Following (4.22) and (4.23) let

$$z = \sigma_1 e^{\frac{\mu_1^2}{2\sigma_1^2}} \exp\left\{-\frac{y_j^2(1)}{2}\frac{\sigma_1^2 - \sigma^2}{\sigma_1^2\sigma^2} - y_j(1)\frac{\mu_1}{\sigma_1^2}\right\},$$

then it suffices to show that

$$\Pr(z < \epsilon) > 1 - \delta, \tag{4-B.5}$$

and $\epsilon = \sigma\epsilon''\frac{p_i}{1-p_i}$. Using total probability we obtain

(4-B.6)

$$\Pr(z < \epsilon) = \Pr\left(z < \epsilon \,\middle|\, y_j > \frac{\mu_1}{2}\right)\Pr\left(y_j > \frac{\mu_1}{2}\right) + \Pr\left(z < \epsilon \,\middle|\, y_j \leqslant \frac{\mu_1}{2}\right)\Pr\left(y_j \leqslant \frac{\mu_1}{2}\right),$$

and claim that under $\mathcal{H}_1$ there exists $\mu_1(\epsilon)$ and $\sigma_1^2(\epsilon)$ for which $\Pr(y_j \leqslant \frac{\mu_1}{2}) < \delta$ for all $\mu_1 \geqslant \mu_1(\epsilon)$ and $\sigma_1^2 \leqslant \sigma_1^2(\epsilon)$. Therefore, it is sufficient to assume $y_j > \frac{\mu_1}{2}$ and show

that $\exists \mathcal{R} \neq \emptyset$ for which

$$\sigma_1 e^{\frac{\mu_1^2}{2\sigma_1^2}} \exp\left\{-\frac{y_j^2}{2}\frac{\sigma_1^2-\sigma^2}{\sigma_1^2\sigma^2} - y_j\frac{\mu_1}{\sigma_1^2}\right\} < \epsilon,\ \forall(\mu_1,\sigma_1^2) \in \mathcal{R}. \tag{4-B.7}$$

Note that (4-B.7) holds if and only if

$$\phi(y_j) = \frac{\sigma_1^2-\sigma^2}{2\sigma_1^2\sigma^2}y_j^2 + \frac{\mu_1}{\sigma_1^2}y_j + \log\epsilon - \log\sigma_1 - \frac{\mu_1^2}{2\sigma_1^2} > 0. \tag{4-B.8}$$

Since $\sigma_1 > \sigma$, $\phi(y_j)$ is a convex function, which is greater than zero for all $y_j > \max(\text{roots}[\phi(y_j)])$ and roots$[\phi(\cdot)]$ are the two solutions of $\phi(\cdot) = 0$. Note that for $y_j = 0$ it is easy to find an ϵ for which $\phi(0) < 0$, hence the equation $\phi(y_j) = 0$ has two solutions. The set $\mathcal{R}$ is defined by all pairs (μ_1, σ_1^2) satisfying

$$\phi\left(\frac{\mu_1}{2}\right) = \frac{\sigma_1^2-\sigma^2}{2\sigma_1^2\sigma^2}\frac{\mu_1^2}{4} + \log\epsilon - \log\sigma_1 > 0, \tag{4-B.9}$$

or, equivalently

$$\sigma_1 \exp\left\{-\frac{\mu_1^2}{8}\frac{\sigma_1^2-\sigma^2}{\sigma_1^2\sigma^2}\right\} \leqslant \epsilon. \tag{4-B.10}$$

Therefore, for $(\mu_1, \sigma_1^2) \in \mathcal{R}$ we have

$$\Pr(z < \epsilon) \geqslant 1 - \delta. \tag{4-B.11}$$

Substituting (4.11) and (4.10) for σ_1 and μ_1 in (4-B.10) respectively, we note that the support $\mathcal{R}$ is a function of the inherent contrast μ_θ, the window length L, the effort allocation at the first stage λ_1, and the cluster statistics $p(\Xi)$.

CHAPTER V

Conclusions and Future Work

5.1 Conclusions

We had started this journey believing that for sparse scenarios there has to be a more efficient way to allocate resource compared to an exhaustive scheme. In Chapter II we investigated design of nuclear imaging scanners. We established that considerable gains in image reconstruction variance are possible using a non-uniform system design. While this measure of performance may not be the best for evaluation of system performance it indicates that further investigation is warranted. Our later work on dynamic resource allocation schemes in Chapters III and IV shows great promise. The main contribution of this dissertation is in forming a novel way to think about dynamic resource allocation problems. We introduced a new cost function capturing the tradeoffs between allocating efforts inside and outside the region of interest. Then, we showed that minimizing this cost is equivalent to reducing both error probability for detecting the ROI correctly and estimation MSE in characterizing the ROI content.

We derived adaptive resource allocation policy, namely $ARAP$, and showed that it is capable of concentrating resource onto the true unknown ROI. $ARAP$ was shown to converge to the oracle allocation policy that knows the location of the ROI in

advance and allocates resources only where it matters. We showed that using $ARAP$ as a pre-processing step, i.e., generating measurements, yields data of a better quality compared to an exhaustive allocation policy. The data quality translates to improved detection and estimation performance. Estimation gains converges to the optimal gains achieved by the omniscient or oracle policy. For detection it is necessary to allocate some efforts to characterize the 'alternative' or the ROI complement. Therefore, the tradeoff parameter ν has to be optimize to provide best combined (detection and estimation) performance. This can be done generically, as seen in Chapter III, but yield better results if done in an application specific context. The price we pay for the significantly improved data yielded by $ARAP$ is an increase in the total number of measurements. We have devised M-ARAP specifically to address this aspect.

Since in many of the application we had considered targets tend to cluster it seems natural that we should try and take advantage of that knowledge. By clustering we do not mean that independent targets appear in groups. In our understanding, a cluster is a result of search steps smaller than targets size. Therefore, a single target is highly likely to occupy several consecutive 'pixels' in the measurements grid. For example, air traffic control (ATC) radar is designed to detect targets of minimal size at a certain range. However, targets that are close by appear much larger than a resolution cell on the radar screen. With that in mind, in Chapter IV we designed M-ARAP to follow the principle of $ARAP$ but use a multi-scale framework. In essence, we believe that targets location can be resolved from a coarse 'image' where spatial resolution is 'sacrificed' to save measurements. The coarse image is then used to decide which parts of the scanned domain should be re-scanned on a fine grid and how much resource should we allocate to each point. For sparse scenarios with

sufficient ROI to background contrast, M-ARAP is capable of producing significant gains while using less than 20% of the measurements required by an exhaustive search. M-ARAP maintained most of the properties of *ARAP* such as convergence to the true ROI yet with much fewer measurements. We explored the tradeoff between estimation gains and measurement saving and conclude that it is best to keep the coarse grid resolution at about 2-3 times of the expected cluster size.

5.2 Future work

For future work, we suggest to connect our work to work done in compressive sampling (CS). For sparse scenarios one can use CS as a replacement for the coarse grid used by M-ARAP. By reducing the number of measurement additional effort can be allocated to each random projection collected by CS measuring operator. This data can be processed to yield posterior probabilities which are then used to define the second allocation step. Moreover, we deal with scenario that are not necessarily sparse in the strict sense, i.e., the complement of the ROI is not limited to be zero. However, we believe that there exist a transformation from the signal space to another space where the signals we are looking at are truly sparse, i.e., the transformed signal values outside the ROI are almost or exactly zero. If such a transformation exist, then compressive sampling methods can be directly applied to the scenarios we consider. Furthermore, a most recent paper by Ji et al. introduces Bayesian CS [28]. Among other things they suggest to adaptively optimize the CS measurement operator in a non random manner using relevance vector machine. This points out to another possible connection between *ARAP* and CS that needs to be explored. A practical issue to address is how to realize CS measurement operator for some of the applications mentioned here. It seems that applications such as radar,

CT, or screening are defined on some discrete grid and it is hard to visualize a radar taking measurements as a random affine transformation of its entire search domain. Nevertheless, the common assumption of sparseness make it worthwhile to further explore such connections.

Another direction worth exploring is the connection to adaptive sampling (AS) such as the work described in [12, 13, 58] and others. The literature on AS address the question of where to sample next or how to focus the sampling process, which was also addressed in this work. Using coarse to fine grid measurement scheme, AS sequentially refines the sampling pattern around suspected edges in an acquired image. The motivation to do so is justified by the assumption that the measured signal has some continuity in a lower dimensional space. Also, since AS does not consider the resource allocation part of the problem it is possible to extend the number of measurement stages beyond two. We believe that techniques used in AS can be combined with policy like M-ARAP to improve performance, but it is yet to be determined how exactly the two should be connected.

Another interesting extension of our work is to extend the horizon beyond a two-stage allocation policy. A straight forward extension suffers from the curse of dimensionality since the optimization problem becomes combinatorically complex. A possible solution is to use a rolling horizon approach. Using a rolling horizon one may estimate potential advantages by going to $T = 3, 4, \ldots$. However, at the end of the first stage the posterior distribution (or prior distribution for the following stages) on the location of targets is no longer uniform. While in [5] we introduce a general version of $ARAP$ that account for a non-uniform prior, the optimization becomes dependent on the specific prior. More specifically, the optimal allocation

lies in a $Q+1$ dimensional space[1]. Therefore, due to the fact that the cardinality of the set of all possible priors is uncountably infinite, the optimization cannot be done offline. Moreover, resource allocated to refine/improve the posterior distribution will no longer be available to be distributed over the estimated ROI.

Finally, it would be interesting to do a sensitivity analysis for the performance of the allocation policies suggested here. Checking how sensitive $ARAP$ is for uncertainties in the sparsity parameter p may yield new insights. In addition, it is worthwhile to analyze what do one lose by assuming a uniform prior as oppose to incorporating prior knowledge. While the general version of $ARAP$ accounts for a non-uniform prior, a uniform prior leads to a robust system design. If $ARAP$ is not very sensitive to such model mismatches it increases the motivation to use it in real life applications.

[1] accounting for an optimal waveform in $\mathbb{R}^Q$ plus its overall energy.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] A. A. Abdel-Samad and A. H. Tewfik. Search strategies for radar target localization. In *Proceeding 1999 International Conference on Image Processing*, volume 3, pages 862–866, October 1999.

[2] A. A. Abdel-Samad and A. H. Tewfik. Hierarchical radar target localization. In *Proceeding 2000 International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 3033–3036, June 2000.

[3] A. A. Abdel-Samad and A. H. Tewfik. Sequential techniques in hierarchical radar target localization. In *Proceeding 2000 International Conference on Image Processing*, volume 1, pages 697–700, September 2000.

[4] M. Aharon, M. Elad, and A. Bruckstein. k-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transaction Signal Processing*, 54(11):4311–4322, November 2006.

[5] E. Bashan, R. Raich, and A. O. Hero III. Adaptive sampling: Efficient search schemes under resource constraints. Technical Report 385, University of Michigan, Communications and Signal Processing Lab., October 2007.

[6] S. J. Benkoski, M. G. Monticino, and J. R. Weisinger. A survey of the search theory literature. *Naval Research Logistics*, 38(4):469–494, 1991.

[7] E. J. Bond, L. Xu, S. C. Hagness, and B. D. van Veen. Microwave imaging via space-time beamforming for early detection of breast cancer. *IEEE Transactions on Antennas and Propagation*, 51(8):1690–1705, August 2003.

[8] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004.

[9] E. Candes and T. Tao. The dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351, 2007.

[10] D. A. Castanon. Optimal search strategies in dynamic hypothesis testing. *IEEE Transaction on Systems, Man and Cybernetics*, 25(7):1130–1138, July 1995.

[11] R. Castro, J. Haupt, and R. Nowak. Compressed sensing vs. active learning. In *Proceedings 2006 International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages III–III, May 2006.

[12] R. Castro, R. Willet, and R. Nowak. Coarse-to-fine manifold learning. In *Proceedings 2004 International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages iii–992–5, May 2004.

[13] R. Castro, R. Willett, and R. Nowak. Faster rates in regression via active learning. In *Proceedings of the Neural Information Processing Systems Conference (NIPS) 2005*, Vancouver, Canada, December 2005.

[14] S. K. Davis, H. Tandradinata, S. C. Hagness, and B. D. van Veen. Ultrawideband microwave breast cancer detection: a detection-theoretic approach using the generalized likelihood ratio test. *IEEE Transaction on Biomedical Engineering*, 52(7):1237–1250, July 2005.

[15] A. den Dekker and A. van den Bos. Resolution: a survey. *Journal Optical Society of America*, 14(3):547–557, March 1997.

[16] R. Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440, December 1943.

[17] Y. C. Eldar. Minimum variance in biased estimation: Bounds and asymptotically optimal estimators. *IEEE Transactions on Signal Processing*, 52(7):1915–1930, July 2004.

[18] O. D. Escoda, L. Granai, and P. Vandergheynst. On the use of a priori information for sparse signal approximations. *IEEE Transactions on Signal Processing*, 54(9):3468–3482, September 2006.

[19] E. C. Fear, P. M. Meaney, and M. A. Stuchly. Microwaves for breast cancer detection? *IEEE Potentials*, 22(1):12–18, Febuary-March 2003.

[20] J. A. Fessler. Spatial resolution and noise tradeoffs in pinhole imaging system design: a density estimation approach. *OPTICS EXPRESS*, 2(6):237–253, March 1998.

[21] A. B. Frakt, W. C. Karl, and A. S. Willsky. A multiscale hypothesis testing approach to anomaly detection and localization from noisy tomographic data. *IEEE Transactions on Image Processing*, 7(6):825–837, June 1998.

[22] J. C. Gittins. *Multi-Armed Bandit Allocation Indices*. John Wiley and Sons, 1989.

[23] I.F. Gorodnitsky and B.D. Rao. Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45(3):600–616, March 1997.

[24] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. *Information Theory, IEEE Transactions on*, 52(9):4036–4048, September 2006.

[25] A. O. Hero III, D. A. Castañón, D. Cochran, and K. Kastella, editors. *Foundations and Applications of Sensor Management*. Springer, Boston, MA, 2007.

[26] A. O. Hero III, J. A. Fessler, and M. Usman. Exploring estimator bias-variance tradeoffs using the uniform cr bound. *IEEE Transactions on Signal Processing*, 44(8):2026–2041, August 1996.

[27] M. Ivanovic, D. A. Webber, and S. Loncaric. Multi-pinhole collimator optimization for high resolution spect imaging. In *IEEE Nuclear Science Symposium*, volume 2, pages 1097–1101, November 1997.

[28] S. Ji, Y. Xue, and L. Carin. Bayesian compressive sensing. *Signal Processing, IEEE Transactions on*, 56(6):2346–2356, June 2008.

[29] I. M. Johnstone and B. W. Silverman. Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Annals of Statistics*, 32:1594–1649, 2004.

[30] J. B. Kadane. Optimal whereabouts search. *Operations Research*, 19:894–904, July-August 1971.

[31] K. Kastella. Discrimination gain to optimize detection and classification. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 27(1):112–116, January 1997.

[32] S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Prentice-Hall PTR, January 1998.

[33] C. Kreucher, K. Kastella, and A. O. Hero III. Multitarget tracking using the joint multitarget probability density. *IEEE Transactions on Aerospace and Electronic Systems*, 41(4):1396–1414, October 2005.

[34] C. Kreucher, K. Kastella, and A. O. Hero III. Sensor management using an active sensing approach. *Signal Processing*, 85(3):607–624, March 2005.

[35] V. Krishnamurthy. Algorithms for optimal scheduling and management of hidden markov model sensors. *IEEE Transactions on Signal Processing*, 50(6):1382–1397, June 2002.

[36] V. Krishnamurthy and R. J. Evans. Hidden markov model multiarm bandits: a methodology for beam scheduling in multitarget tracking. *IEEE Transactions on Signal Processing*, 49(12):2893–2908, December 2001.

[37] D. Malioutov, M. Cetin, and A. S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Transactions on Signal Processing*, 53(8):3010–3022, August 2005.

[38] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, December 1995.

[39] L. J. Meng and N. H. Clinthorne. A modified uniform cramer-rao bound for multiple pinhole aperture design. *IEEE Transactions on Medical Imaging*, 23(7):896–902, July 2004.

[40] L. J. Meng, W. L. Rogers, N. H. Clinthorne, and J. A. Fessler. Feasibility study of compton scattering enchanced multiple pinhole imager for nuclear medicine. *IEEE Transactions on Nuclear Science*, 50(5):1609–1671, October 2003.

[41] M. Nafie, A. H. Tewfik, and M. Ali. Deterministic and iterative solutions to subset selection problems. *IEEE Transactions on Signal Processing*, 50(7):1591–1601, July 2002.

[42] R. Nowak, U. Mitra, and R. Willett. Estimating inhomogeneous fields using wireless sensor networks. *IEEE Journal on Selected Areas in Communication*, 22(6):999–1006, August 2004.

[43] T. E. Peterson, D. W. Wilson, and H. H. Barrett. A small-animal imaging system based on silicon strip detectors. In *Proceedings 2002, IEEE International Symposium on Biommedical Imaging*, pages 533–536, July 2002.

[44] T. E. Peterson, D. W. Wilson, and H. H. Barrett. Ultrahigh-resolution small-animal imaging using a silicon detector. In *Nuclear Science Symposium Conference Record*, volume 3, pages 1984–1987, October 2003.

[45] C. G. Pfeifer and P. Enis. Dorfman-type group testing for a modified binomial model. *Journal of the American Statistical Association*, 73(363):588–592, September 1978.

[46] E. Posner. Optimal search procedures. *IEEE Transactions on Information Theory*, 9(3):157–160, July 1963.

[47] R. Rangarajan, R. Raich, and A. O. Hero III. Optimal experimental design for an inverse scattering problem. In *Proceedings. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 1117–1120, March 2005.

[48] R. Rangarajan, R. Raich, and A. O. Hero III. Sequential design of experiments for a rayleigh inverse scattering problem. In *2005 IEEE/SP 13th Workshop on Statistical Signal Processing*, pages 625–630, July 2005.

[49] R. Rangarajan, R. Raich, and A. O. Hero III. Optimal sequential energy allocation for inverse problems. *IEEE Journal on Selected Topics in Signal Processing*, 1(1):67–78, June 2007.

[50] M. M. Rogulski, H. B. Barber, H. H. Barrett, R. L. Shoemaker, and J. M. Woolfenden. Ultra-high-resolution brain spect imaging: simulation results. *IEEE Transactions on Nuclear Science*, 40(4):1123–1129, August 1993.

[51] N. U. Schramm, G. Ebel, U. Engeland, T. Schurrat, M. Béhé, and T. M. Behr. High-resolution spect using multipinhole collimation. *IEEE Transactions on Nuclear Science*, 50(3):315–320, June 2003.

[52] Nah-Oak Song and D. Teneketzis. Discrete search with multiple sensors. *Mathematical Methods of Operations Research*, 60(1):1–13, September 2004.

[53] A. Sterrett. On the detection of defective members of large populations. *The Annals of Mathematical Statistics*, 28(4):1033–1036, December 1957.

[54] J. A. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, March 2006.

[55] M. B. Wakin and R. G. Baraniuk. Random projections of signal manifolds. In *Proceedings IEEE Conference on Acoustics, Speech and Signal Processing*, volume 5, pages V–V, May 2006.

[56] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk. An architecture for compressive imaging. In *IEEE International Conference on Image Processing*, pages 1273–1276, October 2006.

[57] G. Wang and Y. Li. Axiomatic approach for quantification of image resolution. *IEEE Signal Processing Letters*, 6(10):257–258, October 1999.

[58] R. Willett, A. Martin, and R. Nowak. Backcasting: adaptive sampling for sensor networks. In *Third International Symposium on Information Processing in Sensor Networks*, pages 124–133, April 2004.

[59] D. W. Wilson, H. H. Barrett, and L. R. Furenlid. A new design for a spect small-animal imager. In *Nuclear Science Symposium Conference Record*, volume 3, pages 1826–1829, November 2001.

[60] D. P. Wipf and B. D. Rao. Sparse Bayesian learning for basis selection. *IEEE Transactions on Signal Processing*, 52(8):2153–2164, August 2004.

[61] B. Wohlberg. Noise sensitivity of sparse signal representations: reconstruction error bounds for the inverse problem. *IEEE Transactions on Signal Processing*, 51(12):3053–3060, December 2003.

[62] B. Wray. Learning classification rules using Bayes. In *Proceedings of the sixth international workshop on Machine learning*, pages 94–98, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc.

[63] L. Xu, E. J. Bond, B. D. van Veen, and S. C. Hagness. An overview of ultra-wideband microwave imaging via space-time beamforming for early-stage breast-cancer detection. *IEEE Antennas and Propagation Magazine*, 47(1):19–34, Febuary 2005.

[64] L. Xu, S. K. Davis, S. C. Hagness, D. W. van der Weide, and B. D. van Veen. Microwave imaging via space-time beamforming: experimental investigation of tumor detection in multilayer breast phantoms. *IEEE Transactions on Microwave Theory and Techniques*, 52(8):1856–1865, August 2004.