Compressive Sensor Networks: Fundamental Limits and Algorithms

by

John Zheng Sun

B.S. Electrical and Computer Engineering, Cornell University (2007)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

Author .....
Department of Electrical Engineering and Computer Science
September 2, 2009

Certified by .....
Vivek K Goyal
Esther and Harold E. Edgerton Associate Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Terry P. Orlando
Chairman, Department Committee on Graduate Theses

Compressive Sensor Networks: Fundamental Limits and Algorithms

by

John Zheng Sun

Submitted to the Department of Electrical Engineering and Computer Science
on September 2, 2009, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

Compressed sensing is a non-adaptive compression method that takes advantage of natural sparsity at the input and is fast gaining relevance to both researchers and engineers for its universality and applicability. First developed by Candès *et al.*, the subject has seen a surge of high-quality results both in its theory and applications. This thesis extends compressed sensing ideas to sensor networks and other bandwidth-constrained communication systems. In particular, we explore the limits of performance of compressive sensor networks in relation to fundamental operations such as quantization and parameter estimation.

Since compressed sensing is originally formulated as a real-valued problem, quantization of the measurements is a very natural extension. Although several researchers have proposed modified reconstruction methods that mitigate quantization noise for a fixed quantizer, the optimal design of such quantizers is still unknown. We propose to find the optimal quantizer in terms of minimizing quantization error by using recent results in functional scalar quantization. The best quantizer in this case is not the optimal design for the measurements themselves but rather is reweighted by a factor we call the sensitivity. Numerical results demonstrate a constant-factor improvement in the fixed-rate case.

Parameter estimation is an important goal of many sensing systems since users often care about some function of the data rather than the data itself. Thus, it is of interest to see how efficiently nodes using compressed sensing can estimate a parameter, and if the measurements scalings can be less restrictive than the bounds in the literature. We explore this problem for time difference and angle of arrival, two common methods for source geolocation. We first derive Cramér-Rao lower bounds for both parameters and show that a practical block-OMP estimator can be relatively efficient for signal reconstruction. However, there is a large gap between theory and practice for time difference or angle of arrival estimation, which demonstrates the CRB to be an optimistic lower bound for nonlinear estimation. We also find scaling laws for time difference estimation in the discrete case. This is strongly related to partial support recovery, and we derive some new sufficient conditions that show a

very simple reconstruction algorithm can achieve substantially better scalings than full support recovery suggests is possible.

Thesis Supervisor: Vivek K Goyal
Title: Esther and Harold E. Edgerton Associate Professor of Electrical Engineering

Acknowledgments

Marvin Weisbord once said. "Teamwork is the quintessential contradiction of a society grounded in individual achievement," and these words resonate most strongly in the academic community. In preparing this thesis, I have benefited from the support of many talented and genuinely wonderful people that I wish to acknowledge.

First and foremost, I thank my advisor Vivek Goyal for his advice and mentorship. Our numerous discussions have been invaluable and his suggestions are always insightful and revealing.

I also acknowledge the members the STIR group for their helpful comments, friendship and encouragement. Specifically, I thank Lav Varshney for the wealth of his knowledge and important pointers to papers on universal delay estimation, Dan Weller for advice on estimation problems, Vinith Misra for discussions on functional quantization, and Adam Zelinski for providing code for sparsity-enforcing reconstruction algorithms. I also appreciate the help of Eric Strattman on the administrative side.

There are several external collaborators that helped shape aspects of this thesis and I thank them as well. Joel Goodman, Keith Forsythe, Ben Miller, and Andrew Bolstad at Lincoln Laboratory worked with me on the Compressive Sensor Networks project. Sundeep Rangan provided excellent suggestions that led to some proofs on partial support recovery. Ramin Samadani, Mitchell Trott, and John Apostolopoulos of HP Labs were very gracious in hosting me as a Researcher in Residence and helped me better understand estimation performance bounds.

On a more personal level, I am indebted to old and new friends for making the last two years particularly memorable. The most important thanks goes to my parents for teaching me that the quest for knowledge is lifelong and always being supportive. I thank my sister Jenny for her infectious boundless energy, and to Grace for possessing the uncanny ability to make me smile. I am grateful to Jon, Yi Zheng, Mitch and Peter for friendships that defy distance, and to all my new friends that has made MIT as fun as it is enlightening.

Contents

List of Figures

Chapter 1

Introduction

Sensor networks are prevalent in today's technological landscape and have inspired important engineering innovations in a variety of military, industrial, and environmental applications. However, the nodes of these sensor networks must oftentimes survey wide bands, requiring high-rate analog-to-digital conversion, expensive processors, and large data flows to propagate information through the network. In the case where the data of interest is *sparse* in some basis, meaning it has few degrees of freedom, there is an opportunity to filter the measured signal intelligently and use a sub-Nyquist sampling rate while still maintaining the fidelity of the data. Another way of saying this is that it is possible to sample closer to the *information rate* of the signal rather than being restricted above the Nyquist rate. Moreover, utilizing the underlying sparsity allows for data compression, which eases the communication load between nodes. Efficient techniques for exploiting sparsity in signal representation and estimation have caused an enthusiastic reexamination of data acquisition in sensor networks.

Compressive sensor networks (CSN) exploit a new compression paradigm called compressed sensing (CS) to non-adaptively filter and compress sparse signals. As described in Section 2.1, CS refers to the estimation of a signal at a resolution higher than the number of data samples by taking advantage of sparsity or compressibility of the signal and randomization in the measurement process. CSN nodes contain analog-to-digital converters (ADCs) that use CS principles such as signal spreading

and random sampling. This allows the ADCs to sample significantly slower than the Nyquist rate, making the hardware simpler to design and cheaper to manufacture. Moreover, we can transmit the compressed version of the input signal and ease communication loads.

This thesis addresses extensions of compressed sensing to quantization and estimation, both important operations in sensor networks. We focus on fundamental limits and practical algorithms, thereby abstracting out the actual data-collecting machinery. Thus, although the thesis is related to CSN in theme, the questions we study are of independent interest in the compressed sensing literature.

1.1 A Peek at Compressed Sensing

We present a quick summary of compressed sensing (CS) and introduce the notation that will be used in the thesis. We will provide a more detailed look at CS in Section 2.1.

Consider a length-N input vector x that is K-sparse in some orthonormal basis Ψ, such that a length-N vector $u = \Psi^{-1}x$ will have only K nonzero elements. Define the *sparsity pattern* J to be the set of indices of the nonzero elements in u. Also define the *sparsity ratio* to be $\alpha \triangleq K/N$.

Now let a length-M *measurement vector* be $y = \Phi x$, where $\Phi \in \mathbb{R}^{M \times N}$ is the *sensing matrix*. We define the *downsampling rate* to be $d \triangleq N/M$. In general, since $d > 1$, we cannot recover x from the measurements y since Φ is underdetermined. The major innovation in compressed sensing for the case of sparse u is that the recovery of x from y via some computationally-tractable reconstruction method can be guaranteed asymptotically almost surely for *random* sensing matrices Φ. This means the CS encoder is simply a matrix multiplication, making it linear and non-adaptive. Moreover, the decoder is well-studied and implementable in practice.

Intuitively, compressed sensing works because the information rate is much lower than the Nyquist rate and hence sampling via Shannon's famed theorem is redundant. For certain choices of Φ, specifically when the sparsity and sampling (Ψ and

Φ respectively) are *incoherent*, CS provides theoretical bounds on the performance of signal or sparsity pattern recovery.

Many practical reconstruction methods, or decoders, have been proposed including convex programs like basis pursuit and greedy algorithms like orthogonal matching pursuit (OMP). The algorithms pertinent to the thesis are discussed in Section 2.1.3.

1.2 Thesis Outline

The thesis will explore optimality criteria and practical algorithms for certain aspects of CSN. We tackle three problems that bound the performance of sampling, transmission and inference of sparse signals using a CS setup. These problems are addressed in separate chapters in this thesis. Before that, we begin with some background on compressed sensing, quantization and inference bounds in Chapter 2.

Chapter 3 addresses the design of optimal quantizers at the sensor's ADC to minimize distortion due to quantization noise. Although quantization is a necessary step in any practical ADC, it is oftentimes neglected in theoretical explorations. We find an approximation to the optimal quantizer and quantify its performance.

Chapter 4 looks at the performance of estimators for time difference of arrival (TDOA), angle of arrival (AoA), and signal reconstruction using the CSN framework. We present heuristic algorithms for estimating such parameters at a fusion center given the random measurement data from CSN nodes. We also derive Cramér-Rao lower bounds to quantify the error variance of optimal estimators and compare them to practical ones. TDOA and AOA are useful in tracking and surveillance, and this chapter aims to see if CSN can be applied to these situations.

Finally, we look at scaling laws for discrete time difference estimation (dTDOA) in Chapter 5. This specific case of the previous is concerned with sparse discrete-time signals that are periodic and delayed (circularly shifted) by some integer amount. We aim to find how the compression factor scales with signal length and sparsity so that the time difference can be recovered with high probability.

Chapter 2

Background

This thesis extends compressed sensing to two fundamental areas in communications: quantization and inference. This chapter introduces compressed sensing in the context of its history, theory and applications. Also, relevant concepts in quantization and inference are discussed.

We first introduce the notation for the rest of the thesis. Scalars and vectors are in lowercase, while matrices are in uppercase. Subscripts are used to indicate entries of a vector or matrix. A random variable is always bolded while its realizations are unbolded. This is used carefully to distinguish between when we care about the variable being random or just the realizations of it for computational manipulations.

2.1 Compressed Sensing

Almost exclusively, we consider the setup described in Section 1.1. Except when mentioned, we assume without loss of generality that Ψ is the identity matrix I_N and hence the input vector x is sparse. The sensing matrix Φ is chosen to satisfy certify certain conditions to be discussed and the measurement vector is $y = \Phi x + \eta$, where η is measurement noise. As a reminder, x has length N with K nonzero entries and y has length M.

In this section, we present some pointers to various works on compressed sensing. For those interested in delving deeper into the literature, we recommend the March

2008 issue of the IEEE Signal Processing Magazine, especially [1], as a starting point.

2.1.1 History

Compressed sensing was developed in three landmark 2006 papers by Candès *et al.* [2], Donoho [3], and Candès and Tao [4]. However, the idea of exploiting sparsity in undersampled mixed measurements has a long and rich history. The substitution of the natural ℓ_0 pseudonorm ($||x||_0$ = number of nonzeros in x) with an ℓ_1 norm ($\|x\|_1 = \sum |x_i|$) in the sparsity constraint to create a convex optimization problem is well-known in many communities, including geophysics [5] and medical imaging [6]. This trick was later formalized in the harmonic analysis community, and the proofs in these works relate strongly to the uncertainty principle [7, 8] and the idea of mutual incoherence between the sparse and measurement bases [9, 10, 11].

In [2], Candès *et al.* contributed several key ideas to distinguish compressed sensing from previous work. The most important of these is using *random* sensing as both a practical way to reduce observed measurements and a tool for proving sufficiency in the number of measurements needed for signal recovery. Also essential is exploiting practical algorithms developed previously for sparse expansions of signals (called atomic decomposition) to sparse signal recovery. In particular, the authors considered a convex optimization (which simplified to a linear program) to find the best sparse time signal given a random subset of frequencies, and determined that the number of measurements for perfect signal recovery with high probability scales as $\mathcal{O}(K \log(N/K))$.

Later papers by Donoho [3] and Candès and Tao [4], generalized compressed sensing to signals sparse in arbitrary bases and a broader class of *compressible* signals that satisfy

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{1/p} \leq R \tag{2.1}$$

for some constant R and $0 < p \leq 1$. For compressible signals, perfect recovery is impossible but the minimax error of the K-sparse solution found from using $\mathcal{O}(K \log N)$ random measurements is bounded by the error of the best K coefficients. Succinctly

stated, $K \log N$ random measurements is approximately as good as the K most informative ones.

Later extensions to measurements with additive noise were proposed by Candès *et al.* [12], Haupt and Nowak [13], and Candès and Tao [14].

2.1.2 Incoherence and the Restricted Isometry Property

In the first CS papers, sensing is almost always assumed to be random and most of the derivations hinge on properties of random matrices. However, compressed sensing can be applied much more generally and measurement scaling laws can be derived as long as both the sparse basis Ψ and sampling matrix Φ obey either incoherence or the restricted isometry property (RIP). We will now briefly describe both methods and contrast them.

Coherence, introduced earlier in [9] for atomic decomposition, is an intuitive measure of similarity between two bases and corresponds to the largest correlation between any atom pair. Mathematically, given matrices Φ and Ψ representing orthonormal bases in $\mathbb{R}^N$, the coherence $\mu(\Phi, \Psi)$ is

$$\mu(\Phi, \Psi) = \sqrt{N} \max_{1 \leq k,j \leq N} |\langle \phi_k, \psi_j \rangle| , \tag{2.2}$$

where ϕ_k and ψ_j are columns (or atoms) of Φ and Ψ respectively. The two bases are considered *incoherent* if $\mu(\Phi, \Psi)$ is small. Two applicable examples include the time-frequency basis pair, with $\mu(\Phi, \Psi) = 1$, and random bases, which are incoherent with most natural bases.

In CS, it is known that signal recovery via ℓ_1 minimization is only possible for $M \asymp \mu^2(\Phi, \Psi) K \log N$ [15]. The number of measurements is minimized when the sensing and sparsity are incoherent. This validates the scenarios presented in [2] and [4] since in both situations the coherence term is close to $\mathcal{O}(1)$. Moreover, coherence allows one to determine what types of sensing approach the CS bounds and gives intuition on how sparsity and sensing must be "orthogonal" to each other.

The restricted isometry property, described in [16], is satisfied by matrices Ψ and

Φ if the smallest possible δ for

$$(1-\delta)\|u\|_2^2 \leq \|\Phi\Psi u\|_2^2 \leq (1+\delta)\|u\|_2^2 \tag{2.3}$$

is not close to 1. This must hold for all K-sparse vectors u. An interpretation of these inequalities is that the compression $\Phi\Psi$ somewhat preserves the norm of all K-sparse signals, allowing them to be recovered later.

For sampling matrices where RIP holds, the ℓ_1 minimization will have bounded error [12], meaning for some constant c,

$$\|\hat{u} - u\|_2 \leq c\|u_K - u\|_1, \tag{2.4}$$

where u_K is the best K-sparse estimate and $\hat{u}$ is the ℓ_1-minimization solution. This means that, for choices of Φ that satisfy RIP, exact reconstruction is possible for K-sparse inputs. However, it also bounds compressible and noisy signals as well. It has been shown that random matrices satisfy RIP for $M \asymp \mathcal{O}(K\log(N/K))$ and hence be suitable for ℓ_1 minimization.

Both incoherence and the RIP provides conditions on Φ and Ψ for the CS model to perform successful signal reconstruction. The RIP provides stronger statements and can be extended easily to compressible signals. However, it is less intuitive and much more difficult to validate. Incoherence is a weaker condition but can be useful for choices of Φ where RIP will not hold.

2.1.3 Reconstruction Algorithms

The reconstruction of the sparse input x from the measurement vector y and sensing matrix Φ is a well-studied problem in the last few years. Reconstruction algorithms usually fall into three categories: combinatorial searches, convex programs and greedy algorithms.

The combinatorial methods are the most intuitive but unfortunately not computationally tractable. If the signal is known to be exactly K-sparse, then it must lie

on one of the K-sparse planes in $\mathbb{R}^N$. There are exactly $\binom{N}{K}$ such subspaces and one can do an exhaustive search to find the best solution constrained on them. Another way of formulating this problem is to solve the combinatorial optimization problem

$$\hat{x} = \arg\min_x \|x\|_0, \quad \text{subject to} \quad y = \Phi x, \tag{2.5}$$

where $\|x\|_0$ is the ℓ_0 pseudonorm, or the number of nonzero terms [2]. This corresponds to the ML estimator studied in [17, 18]. For this class, $M \asymp \mathcal{O}(K)$ measurements are needed to perfectly reconstruct the original sparse signal in the noiseless setting. In the noisy setting, [19] shows that $M \asymp \mathcal{O}(K \log(N-K))$ is a sufficient condition for perfectly recovering the sparsity pattern.

Convex relaxations of the sparsity constraint reduce the computational costs dramatically and were discussed in the original CS papers. Specifically, a linear program

$$\hat{x} = \arg\min_x \|x\|_1, \quad \text{subject to} \quad y = \Phi x, \tag{2.6}$$

gives accurate signal recovery with overwhelming probability for Φ chosen randomly provided M is large enough. This is known in the literature under the name basis pursuit [20]. As shown in [2], the ℓ_1 minimization is successful almost surely (compared to the ℓ_0 minimization being successful always) if the number of measurements is $\mathcal{O}(K \log(N/K))$. Later work sharpens this sufficient condition to $M \geq 2K \log(N/M)$ [21].

With additive Gaussian noise, perfect reconstruction of a K-sparse signal is impossible. A modified quadratic program called lasso [22] is often used to find a solution with bounded error for $M \asymp \mathcal{O}(K \log(N-K))$. Lasso takes the form

$$\hat{x} = \arg\min_x \left(\|y - \Phi x\|_2^2 + \mu \|x\|_1 \right), \tag{2.7}$$

with the regularization parameter μ dependent on the Gaussian noise variance. As a sample result, lasso leads to proper detection of the nonzero indices, called sparsity pattern recovery, with high probability if $M \sim 2K \log(N-K) + K$ under certain

conditions on Φ, μ, and the scaling of the smallest entry of x [17]. Several algorithmic methods for determining the reconstruction from lasso for a given μ have been studied [22, 23], but the proper choice for μ is an interesting design problem.

One method to visualize the set of solutions formed by lasso is homotopy continuation [24]. HC considers the regularization parameter μ at an extreme point (e.g. very large μ so the reconstruction is all zero) and sweeps μ so that all sparsities and the resulting reconstructions are obtained. It is shown that there are N values of μ where the lasso solution changes sparsity, or equivalently $N+1$ intervals over which the sparsity does not change. For μ in the interior of one of these intervals, the reconstruction is determined uniquely by the solution of an affine system of equations involving a submatrix of Φ. In particular, for a specific choice μ and observed random measurements y,

$$2\Phi_{J_\mu}^T \Phi_{J_\mu} \hat{x} + \mu \operatorname{sgn}(\hat{x}) = 2\Phi_{J_\mu}^T y, \tag{2.8}$$

where Φ_{J_μ} is the submatrix of Φ with columns corresponding to the nonzero entries $J_\mu \subset \{1, 2, \ldots, N\}$ of $\hat{x}$.

A final class of reconstruction algorithms is made up of greedy heuristics that are known to be good for sparse signal approximations for overcomplete dictionaries [25, 26]. In particular, an algorithm called orthogonal matching pursuit (OMP) is shown to be successful in signal recovery for the scaling $M \asymp \mathcal{O}(K \log N)$ [27].

2.1.4 Extensions and Applications of CS

Compressed sensing has reinvigorated the study of sampling in applied mathematics, statistics, and computational science. We will briefly mention some current work in extending CS theory and applications.

Many researchers are trying to generalize the rather rigid constraints of the landmark papers on CS, which restricts the sparse signal to be discrete time and continuous valued. Lu and Do [28], and Eldar [29] have extended the CS framework to analog signals. Goyal *et al.* [30] and others consider practical communication constraints in terms of quantization of the measurements, which we further extend in Chapter 3.

Other authors like Fletcher *et al.* [31] and Saligrama *et al.* [32] explore the asymptotic bounds of sparsity pattern recovery rather than signal recovery. Extensions to simultaneous sparsity for multi-sensor compressed sensing has also be widely studied, most commonly with an "ℓ_1 of the ℓ_2" sparsity cost in the measurement matrix of sensor readings [33, 34].

Other current work focuses on improving existing reconstruction algorithms, both in computational complexity and in the number of measurements needed. Some interesting papers include thresholded basis pursuit [35], CoSaMP [36], and subspace pursuit [37].

Finally, numerous applications have also been developed using the CS paradigm in a variety of areas. Many researchers are applying CS to magnetic resonance imaging [38] since fast sampling is essential. Other relevant EE-style applications include finding users in a wireless network [39] and single-pixel imaging [40]. Moreover, compressed sensing has found applications in fields as diverse as astronomy [41], integrated circuits [42], and neuroscience [43].

2.2 Quantization

The quantization of real-world measurements is an essential consideration for digital systems. When an analog signal $x(t)$ is processed by an ADC, a digital (discrete-value, discrete-time) signal $x[n]$ is produced. In most cases, the sampling in time is uniform but there is flexibility is choosing the values (or levels) of the discrete-valued output. The best choice for the number and values of the levels is a developed field of research that is surveyed in [44].

We define a quantizer Q as a mapping from the real line to a countable set of points $\mathcal{C} = \{c_i\}$. In particular, if we partition the real line into a set of intervals $\mathcal{P} = \{P_i\}$, then $Q(x) = c_i$ if $x \in P_i$. A more communications-flavored definition of quantization is as a source-coding problem. Each point c_i is associated with a string of bits b_i, and hence an input x is mapped to the string b_i through a lossy encoder S. A corresponding decoder $\hat{S}$ then maps b_i to c_i. The quantizer is therefore

$Q(\cdot) = \hat{S}(S(\cdot))$. This view is inspired by the canonical works of Shannon [45] and allows us understand the cost of quantization in terms of the expected *rate*, or length of each bitstring.

The design of the *codebook* $\mathcal{C}$ and *partition* $\mathcal{P}$ has been studied extensively. The case when each observation is considered independently is called *scalar* quantization. Alternatively, sets of observations quantized together is called *vector* quantization. Another category of variation is whether the bitstrings are of a single length, called *fixed-rate* quantization, or can vary, called *variable-rate* or *entropy-coded* quantization. In most real-world applications, fixed-rate scalar quantization is used. For simplicity, the levels are usually chosen to be equidistant from one another, called *uniform* quantization. However, other types of quantization can lead to significant gains.

2.2.1 Optimal Quantization

Usually, one wishes to design the quantizer Q to minimize some cost. In the literature, the cost is usually the mean-squared error (MSE). Hence, for a probabilistic input $\mathbf{y}$, the quantizer is found by solving the optimization

$$\min_{Q} E\left[\|\mathbf{y} - Q(\mathbf{y})\|^2\right]. \tag{2.9}$$

For the fixed-rate case and a set of rates $\{R_i\}$, the constraint is the maximum number of quantization levels for each $\mathbf{y_i}$ being less than 2^{R_i}. For the entropy-coded case, the entropy of the codebook for each $\mathbf{y_i}$ is less than R_i.

Finding analytical results for quantization as formulated is difficult because the function Q is not continuous. Therefore, we use the high-resolution approximation with R_i large to form continuous representations of quantizers [46, 47]. We define the *(normalized) quantizer point density function* to be $\lambda_i(t)$, such that $\lambda_i(t)\delta$ is the approximate fraction of quantizer reproduction points for $\mathbf{y_i}$ in an interval centered at t with width δ. In the fixed-rate case, using a surprisingly simple application of Hölder's inequality [48], the optimal point density for a given source distribution

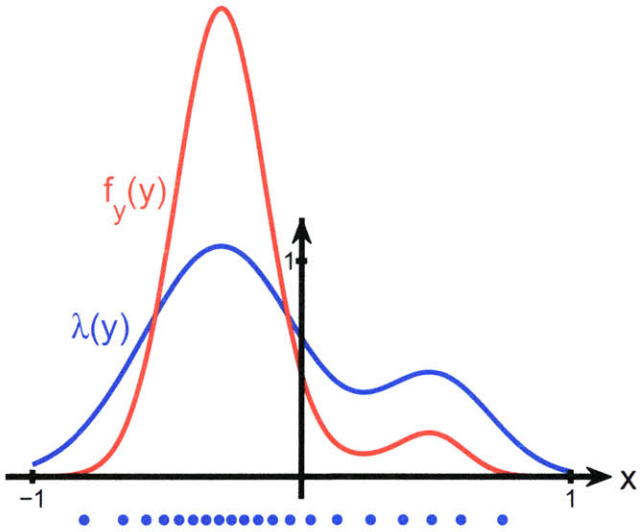


Figure 2-1: A source $\mathbf{y}$ and its point density function, defined as $\lambda_i(t) \propto f_{\mathbf{y_i}}^{1/3}(t)$. A finite-rate quantizer is represented by the dots corresponding to the placement of codewords.

$f_{\mathbf{y_i}}(\cdot)$ is

$$\lambda_i(t) = \frac{f_{\mathbf{y_i}}^{1/3}(t)}{\int f_{\mathbf{y_i}}^{1/3}(t')dt'}. \tag{2.10}$$

The distortion corresponding to this point density is

$$D(\{R_i\}) \approx \sum_{i=1}^{M} 2^{-2R_i} E\left[\frac{\mathbf{y_i}}{12\lambda_i^2(\mathbf{y_i})}\right]. \tag{2.11}$$

To design a quantizer with a certain rate, we can simply partition the cumulative point density into equidistant intervals and find the corresponding codeword. Figure 2-1 shows the point density and a sample quantizer for a source $\mathbf{y}$.

A similar derivation applies for entropy-coded quantization. We will refer to [44] for the details but point out the key result that the optimal point density $\lambda(t)$ is constant on the support of $\mathbf{y_i}$.

2.2.2 Distributed Functional Scalar Quantization

In many applications, one might desire to minimize the quantization error of some function $g(\mathbf{y})$ rather than the source $\mathbf{y}$ itself. *Distributed functional scalar quantiza-*

tion (DFSQ) [49, 50] addresses how to design a set of quantizers to discretize each entry of $\mathbf{y}$. Unlike (2.9), the optimality criterion is now

$$\min_{Q} E\left[|g(\mathbf{y}) - g(Q(\mathbf{y}))|^2\right], \tag{2.12}$$

subject to similar codebook size or entropy constraints.

A canonical example illuminating the gains of DFSQ is separately scalar quantizing a set of variables $\{\mathbf{y_i}\}$ when we wish to minimize distortion on $\max(\{\mathbf{y_i}\})$. Intuitively, we should have a higher concentration of codewords for larger values of $\mathbf{y_i}$ because there is a higher probability that it will be relevant. DFSQ quantifies this optimal quantizer and we note a significant operational distortion-rate improvement.

To apply the theory discussed in [50], we need $g(\cdot)$ and $f_{\mathbf{y}}(\cdot)$ to satisfy certain conditions:

C1. $g(\mathbf{y})$ is (piecewise) smooth and monotonic for each $\mathbf{y_i}$.

C2. The partial derivative $g_i(y) = \partial g(y)/\partial y_i$ is (piecewise) defined and bounded for each i.

C3. The joint pdf of the source variables $f_{\mathbf{y}}(y)$ is smooth and supported in a compact subset of $\mathbb{R}^M$.

For a valid $g(\cdot)$ and $f_{\mathbf{y}}(\cdot)$ pair, we define a set of functions

$$\gamma_i(t) = \left(E\left[|g_i(\mathbf{y})|^2 \mid \mathbf{y_i} = t\right]\right)^{1/2}. \tag{2.13}$$

We call $\gamma_i(t)$ the *sensitivity* of $g(\mathbf{y})$ with respect to the source variable $\mathbf{y_i}$.

In the fixed-rate case, the optimal point density is

$$\lambda_i(t) = C\left(\gamma_i^2(t) f_{\mathbf{y_i}}(t)\right)^{1/3}, \tag{2.14}$$

for some normalization constant C. This leads to a total operational distortion-rate

$$D(\{R_i\}) \approx \sum_{i=1}^{M} 2^{-2R_i} E\left[\frac{\gamma_i^2(\mathbf{y_i})}{12\lambda_i^2(\mathbf{y_i})}\right]. \tag{2.15}$$

The sensitivity $\gamma_i(t)$ serves to reshape the quantizer, giving better resolution to regions of $\mathbf{y_i}$ that have more impact on $g(\mathbf{y})$, thereby reducing MSE. One way of looking at these results then is that this is simply an ordinary quantization design problem, but with the pdf weighted by the sensitivity. We caution that this reweighting need not integrate to unity so it is not a valid pdf.

In the entropy-coded case, the optimal point density is proportional to the sensitivity ($\lambda_i(t) \propto \gamma_i(t)$). This leads to a distortion

$$D(\{R_i\}) \approx \sum_{i=1}^{M} \frac{1}{12} \|\gamma_i^2\|_1^2 2^{2h(\mathbf{y_i})+2E[\log_2 \gamma_i(\mathbf{y_i})]} 2^{-2R_i}. \tag{2.16}$$

The theory of DFSQ can be extended to a vector of functions, where $\hat{\mathbf{x}}_\mathbf{j} = g^{(j)}(\mathbf{y})$ for $1 \leq j \leq N$. Since the cost function is additive in its components, we can show that the overall sensitivity for each component $\mathbf{y_i}$ is

$$\gamma_i(t) = \frac{1}{N} \sum_{j=1}^{N} \gamma_i^{(j)}(t), \tag{2.17}$$

where $\gamma_i^{(j)}(t)$ is the sensitivity of the function $g^{(j)}(\mathbf{y})$ with respect to $\mathbf{y_i}$.

2.3 Inference Performance Bounds

In the inference literature, there are several performance bound families that quantify the expected error variance of optimal parameter estimators. The simplest and most popular of these is the Cramér-Rao lower bound (CRB), which is a measure of the average curvature of the log-likelihood function for the observed data with respect to the parameter in question. We will briefly define both Fisher information and the CRB, using notation from [51].

Definition 2.1. Given a set of observations $\mathbf{x}$ parametrized by a non-random variable

θ, the *Fisher information* $I(\theta)$ is defined as

$$I(\theta) = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right] = E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right)^2\right].$$

Definition 2.2. In the multivariate case with a vector of parameters θ, the *Fisher information matrix* $I(\theta)$ is defined as

$$I(\theta)_{ij} = E\left[\frac{\partial}{\partial\theta_i}\ln p(\mathbf{x};\theta)\frac{\partial}{\partial\theta_j}\ln p(\mathbf{x};\theta)\right].$$

Definition 2.3. The *Cramér-Rao lower bound* is defined as the inverse of the Fisher information. The error variance for any unbiased estimator $\hat{\theta}$ is bounded below by the CRB, such that

$$\mathrm{Var}\left(\hat{\theta}\right) \geq \frac{1}{I(\theta)}.$$

In the multivariate case, the variance of an unbiased estimator $\hat{\theta}_i$ is bounded by

$$\mathrm{Var}\left(\hat{\theta}_i\right) \geq \left[I(\theta)^{-1}\right]_{ii}.$$

Due to its simplicity, the CRB tends to be too optimistic of a bound. A variety of papers show that the CRB is never tight in nonlinear parameter estimation except in the asymptotic-SNR regime [52, 53]. Of relevance to this thesis, it is not tight for TDOA and AOA estimation from measured data for realistic SNR scenarios. Even more troubling, there is a well-known threshold effect at low SNR, when the actual error variance deviates dramatically from the bound. There are several other classes of bounds, including the Barankin [54], Ziv-Zakai [52] and Weiss-Weinstein [53], that are tighter. However, they are more complex to analyze and implement.

Chapter 3

Quantization of Random Measurements

One of the major limitations of the original formulation of compressed sensing is that all quantities are purely continuous-valued, making the model unrealistic in practical systems like sensor networks. One emerging topic in CS research is applying quantization to CS measurements for transmission or storage while maintaining reconstruction fidelity. Most current research focuses on the design of reconstruction algorithms to reduce quantization error while keeping the quantizer design fixed. This chapter considers the reverse case, when the reconstruction algorithm is known and the quantizer is designed to minimize distortion. We utilize recent results in functional quantization, described in Section 2.2.2, to approximate the best quantizer for a CS system.

3.1 Related Work

Quantized compressed sensing (QCS) is fast gaining interest as researchers begin to apply compressed sensing to practical systems. Current work can be separated into two categories: ones that consider asymptotic reconstruction performance assuming a mean-squared error (MSE) distortion metric, and ones providing algorithmic modifications to existing reconstruction methods for mitigating quantization error.

The first work for asymptotic performance of QCS is by Candès and Romberg [55] and considers uniform scalar quantization on random measurements for compressible signals. The authors find the worst-case distortion (using Kolmogorov entropy) for uniform quantization is within a $(\log R)^2$ factor of the optimal encoding. Later work show that, in exactly sparse signals, the penalty of using scalar quantization is much more severe [56, 30]. Bounds for reconstruction distortion in the presence of quantization are presented in [57].

Algorithmically, several modifications to existing reconstruction methods have been used to reduce quantization error. In [12], quantization is treated as iid bounded noise, and reconstruction is found via a relaxed convex optimization

$$\hat{x} = \arg\min_{x} \|x\|_1, \quad \text{subject to} \quad \|y - \Phi x\|_2^2 < \epsilon, \tag{3.1}$$

where ϵ is determined by the noise, or quantization rate. Extensions to this optimization include adding a sign consistency constraint in the low-rate case [58], and applying a different ℓ_p norm on the fidelity constraint [59]. Other modifications include quantized subspace pursuit [57] and vector quantization through binning of quantizer output indexes [60].

3.2 Contribution

As mentioned before, previous works take a reconstruction-centric view of quantization. Our contribution is to reduce distortion by designing the quantizer intelligently based on knowledge of the processing that will occur later to the values being quantized. The key observation is that QCS measurements are used as arguments in a *nonlinear* reconstruction function. Thus, designing a quantizer for the reconstruction is not equivalent to designing a quantizer for the measurements, as demonstrated in an example in Appendix 3.B.

To tackle this problem, we model the reconstruction as a vector-valued function $\hat{\mathbf{x}} = G(\mathbf{y})$ dependent on the observed measurements $\mathbf{y}$, and we wish to minimize

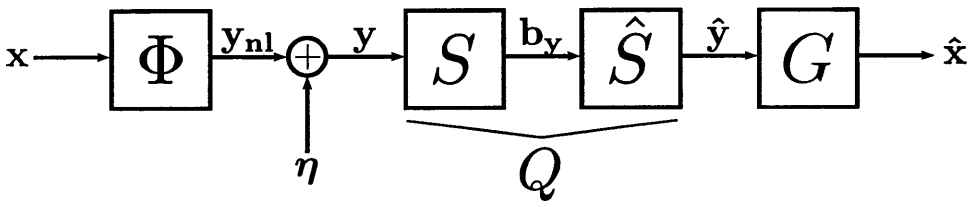


Figure 3-1: A compressed sensing model with quantization of noisy measurements $\mathbf{y}$. The vector $\mathbf{y_{nl}}$ denotes the noiseless random measurements.

distortion resulting from quantization of $\mathbf{y}$. Hence, this is exactly a functional quantization problem (c.f. Section 2.2.2), but extended to the vector case. This extension is straightforward because the cost is additive in the components of the output $\hat{\mathbf{x}}$ and because the proofs for functional quantization rely on standard calculus arguments. Thus, the net sensitivity is simply the mean effect due to the sensitivities for each scalar function and is represented by (2.17).

We then determine the sensitivity for the application of DFSQ to QCS and present positive numerical results. For these results, we make specific choices on the source and sensing matrix distributions and on the reconstruction method. Also, we focus almost entirely on fixed-rate scalar quantization. However, the theory applies more generally and we provide pointers for later extensions.

This work has been published in [61] and [62].

3.3 Problem Model

Figure 3-1 presents the QCS model. We use the notation discussed in Section 2.1 and assume that η is Gaussian noise. The transmitter samples the input using $\mathbf{\Phi}$ and encodes the measurements $\mathbf{y}$ into a bitstream $\mathbf{b_y}$ using encoder S with total rate R. Next, a decoder $\hat{S}$ produces a quantized signal $\hat{\mathbf{y}}$ from $\mathbf{b_y}$. The overall quantizer is denoted $Q(\cdot) = \hat{S}(S(\cdot))$. Finally, a reconstruction algorithm G outputs an estimate $\hat{\mathbf{x}}$. The function G is a black box representing lasso, OMP or another CS reconstruction algorithm. Note that G takes as input a vector of length M and outputs a vector of length N.

We now present a probabilistic model for the input source and sensing matrix.

It is chosen to guarantee finite support on both the input and measurement vectors, and hence prevent overload errors in the quantizer. Although this does not need to hold in general, it will obviate discussions on overload and allow us to focus on the important aspects of the analysis.

Assume the K-sparse vector $\mathbf{x}$ has random sparsity pattern $\mathbf{J}$ chosen uniformly from all possibilities, and each nonzero component $\mathbf{x_i}$ is distributed iid $\mathcal{U}(-1,1)$. This corresponds to the least-informative prior for bounded and sparse random vectors. Let the additive noise vector $\boldsymbol{\eta}$ be distributed iid Gaussian with zero mean and variance σ^2. Finally, assume $\boldsymbol{\Phi}$ corresponds to random projections such that each column $\boldsymbol{\phi_j} \in \mathbb{R}^M$ has unit energy ($\|\boldsymbol{\phi_j}\|^2 = 1$). The columns of $\boldsymbol{\Phi}$ thus form a set of N random vectors chosen uniformly on the unit $(M-1)$-hypersphere. The cumulative distribution function (cdf) of each matrix entry $\boldsymbol{\Phi_{ij}}$ is described in the following lemma:

Lemma 3.1. *Assume $\boldsymbol{\phi_j} \in \mathbb{R}^M$ is a random vector uniformly chosen on a unit $(M-1)$-hypersphere for $M \geq 2$. Then the cdf of each entry $\boldsymbol{\Phi_{ij}}$ of the matrix $\boldsymbol{\Phi}$ is*

$$F_{\boldsymbol{\Phi_{ij}}}(v, M) = \begin{cases} 1 - T(v, M), & 0 \leq v \leq 1; \\ T(-v, M), & -1 \leq v < 0; \\ 0, & \text{otherwise}, \end{cases}$$

where

$$T(v, M) = \frac{\Gamma(\frac{M}{2})}{\sqrt{\pi}\,\Gamma(\frac{M-1}{2})} \int_0^{\arccos(v)} (\sin\theta)^{M-2}\, d\theta$$

and $\Gamma(\cdot)$ is the Gamma function.

Proof. See Appendix 3.A. □

We find the pdf of $\boldsymbol{\Phi_{ij}}$ by differentiating the cdf or using a tractable computational approximation. Since $\mathbf{y} = \Phi\mathbf{x}$, each measurement is

$$\mathbf{y_i} = \sum_{j=1}^{N} \boldsymbol{\Phi_{ij}}\mathbf{x_j} = \sum_{j \in \mathbf{J}} \underbrace{\boldsymbol{\Phi_{ij}}\mathbf{x_j}}_{\mathbf{z_{ij}}}.$$

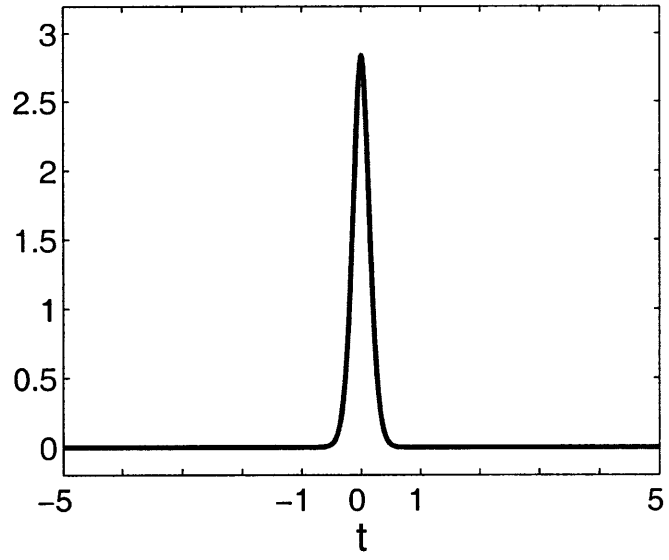


Figure 3-2: Distribution $f_{\mathbf{y_i}}(t)$ for $(K, M, N) = (5, 71, 100)$. The support of $\mathbf{y_i}$ is the range $[-K, K]$, where K is the sparsity of the input signal. However, the probability is only non-negligible for small $\mathbf{y_i}$.

The distribution of each $\mathbf{z_{ij}}$ is found using derived distributions. By symmetry, these pdfs are all identical and will be represented by $f_{\mathbf{z}}(t)$, which is bounded on $[-1, 1]$. The distribution of $\mathbf{y_i}$ is then the $(K-1)$-fold convolution cascade of $f_{\mathbf{z}}(t)$ with itself. Thus, the joint pdf $f_{\mathbf{y}}(y)$ is smooth and supported for $\{|\mathbf{y_i}| \leq K\}$, satisfying one of the three conditions of DFSQ listed in Section 2.2.2. Figure 3-2 illustrates the distribution of $\mathbf{y_i}$ for a particular choice of signal dimensions.

The reconstruction algorithm G is a function of the measurement vector $\mathbf{y}$ and sampling matrix $\mathbf{\Phi}$. For this work, we assume $G(\mathbf{y}, \mathbf{\Phi})$ is lasso with a proper relaxation variable μ, as formulated in (2.7). From the homotopy continuation view of lasso, as discussed in Section 2.1.3, we see $G(\mathbf{y}, \mathbf{\Phi})$ is a piecewise smooth function that is also piecewise monotonic with every $\mathbf{y_i}$ for any *fixed* μ. Moreover, for every μ the reconstruction is an affine function of the measurements through (2.8), so the partial derivative with respect to any entry y_i is piecewise defined and smooth (constant in this case). Hence, conditions C1 and C2 for DFSQ are satisfied and we can use the sensitivity equations discussed earlier.

3.4 Optimal Quantizer Design

We now pose the optimal fixed-rate quantizer design as a DFSQ problem. For a given noise variance σ^2, choose an appropriate μ to form the best reconstruction $\tilde{\mathbf{x}}$ from the unquantized random measurements $\mathbf{y}$. We produce M scalar quantizers for the entries of $\mathbf{y}$ such that the quantized measurements $\hat{\mathbf{y}}$ will minimize the distortion between $\tilde{\mathbf{x}} = G(\mathbf{y}, \mathbf{\Phi})$ and $\hat{\mathbf{x}} = G(\hat{\mathbf{y}}, \mathbf{\Phi})$ for a total rate R. Note G can be visualized as a set of N scalar functions $\hat{\mathbf{x}}_\mathbf{j} = G^{(j)}(\hat{\mathbf{y}}, \mathbf{\Phi})$ that are identical in distribution due to the randomness in $\mathbf{\Phi}$. Since the sparse input signal is assumed to have uniformly distributed sparsity and $\mathbf{\Phi}$ distributes energy equally to all measurements $\mathbf{y}_\mathbf{i}$ in expectation, we argue by symmetry that each measurement is allotted the same number of bits and that every measurement's quantizer is the same. Moreover, again by symmetry in $\mathbf{\Phi}$, the functions representing the reconstruction are identical in expectation and we argue using (2.17) that the overall sensitivity $\gamma_{\text{cs}}(\cdot)$ is the same as the sensitivity of any $G^{(j)}(\hat{\mathbf{y}}, \mathbf{\Phi})$. Computing (2.13) yields the point density $\lambda_{\text{cs}}(\cdot)$.

To determine a functional form of lasso, we use homotopy continuation. For a given realization of $\mathbf{\Phi}$ and $\boldsymbol{\eta}$, HC can find that appropriate sparsity pattern J_μ for a chosen μ. Equation (2.8) is then used to find the partial derivative $\partial G^{(j)}(y, \Phi)/\partial y_i$, which is needed to compute $\gamma_{\text{cs}}(\cdot)$. The resulting differentials can be defined as

$$G_i^{(j)}(y, \Phi) = \frac{\partial G^{(j)}(y, \Phi)}{\partial y_i} \tag{3.2}$$

$$= \left[\left(\Phi_{J_\mu}^T \Phi_{J_\mu} \right)^{-1} \Phi_{J_\mu}^T \right]_{ji}. \tag{3.3}$$

We now present the sensitivity through the following theorem.

Theorem 3.1. *Let the noise variance be σ^2 and choose an appropriate μ for the sparsity K. Define $\mathbf{y}_{\setminus\mathbf{i}}$ to be all the entries of vector $\mathbf{y}$ except $\mathbf{y}_\mathbf{i}$. The sensitivity of each entry $\mathbf{y}_\mathbf{i}$, defined as $\gamma_i^{(j)}(t)$, can be written as*

$$\left(E_{\mathbf{\Phi}} \left[\frac{f_{\mathbf{y_i}|\mathbf{\Phi}}(t|\mathbf{\Phi})}{f_{\mathbf{y_i}}(t)} E_{\mathbf{y}_{\setminus\mathbf{i}}} \left[\left| G_i^{(j)}(\mathbf{y}, \mathbf{\Phi}) \right|^2 \mid \mathbf{y_i} = t, \mathbf{\Phi} \right] \right] \right)^{\frac{1}{2}}.$$

For a realization Φ and J_μ found through HC, $f_{\mathbf{y_i}|\mathbf{\Phi}}(t|\Phi)$ is the convolution cascade of $\{\mathbf{z_{ij}} \sim \mathcal{U}(-\Phi_{ij}, \Phi_{ij})\}$ for $j \in J$. By symmetry arguments, $\gamma_{\text{cs}}(t) = \gamma_i^{(j)}(t)$ for any i and j.

Proof. By symmetry arguments, we can consider any i and j for the partial derivative in the sensitivity equation without loss of generality. Noting (3.2), we define

$$\Gamma_i^{(j)}(t, \Phi) = E_{\mathbf{y}\setminus\mathbf{i}}\left[\left|G_i^{(j)}(\mathbf{y}, \mathbf{\Phi})\right|^2 \mid \mathbf{y_i} = t\right],$$

and then modify (2.13) in the following steps:

$$\begin{aligned}
\gamma_i^{(j)}(t) &= \left(E\left[\left|G_i^{(j)}(\mathbf{y}, \mathbf{\Phi})\right|^2 \mid \mathbf{y_i} = t\right]\right)^{\frac{1}{2}} \\
&= \left(E_{\mathbf{\Phi}}\left[\Gamma_i^{(j)}(t, \mathbf{\Phi}) \mid \mathbf{y_i} = t\right]\right)^{\frac{1}{2}} \\
&= \left(\int f_{\mathbf{\Phi}|\mathbf{y_i}}(\Phi|t)\Gamma_i^{(j)}(t, \Phi)d\Phi\right)^{\frac{1}{2}} \\
&= \left(E_{\mathbf{\Phi}}\left[\frac{f_{\mathbf{y_i}|\mathbf{\Phi}}(t|\mathbf{\Phi})}{f_{\mathbf{y_i}}(t)}\Gamma_i^{(j)}(t, \mathbf{\Phi})\right]\right)^{\frac{1}{2}}.
\end{aligned}$$

Plugging in (3.3) will give us the final form of the theorem. Given a realization Φ, $\mathbf{y_i} = \sum_{j\in\mathbf{J}} \Phi_{ij}\mathbf{x_j} = \sum_{j\in J} \mathbf{z_{ij}}$, meaning $\mathbf{z_{ij}} \sim \mathcal{U}(-\Phi_{ij}, \Phi_{ij})$. The conditional probability $f_{\mathbf{y_i}|\mathbf{\Phi}}(y|\Phi)$ can be found by taking the $(K-1)$-fold convolution of the set of density functions representing the K nonzero $\mathbf{z_{ij}}$'s. □

The expectation in Theorem 3.1 is difficult to calculate but can be approached through L Monte Carlo trials on $\mathbf{\Phi}$, $\boldsymbol{\eta}$, and $\mathbf{x}$. For each trial, we can compute the partial derivative using (3.3). We denote the Monte Carlo approximation to that function to be $\gamma_{\text{cs}}^{(L)}(\cdot)$. Its form is

$$\gamma_{\text{cs}}^{(L)}(t) = \frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{f_{\mathbf{y_i}|\mathbf{\Phi}}(t|\Phi_\ell)}{f_{\mathbf{y_i}}(t)}\left[G_i^{(j)}(y_\ell, \Phi_\ell)\right]^2\right)^{\frac{1}{2}}, \tag{3.4}$$

with i and j arbitrarily chosen. By the weak law of large numbers, the empirical mean of L realizations of the random parameters should approach the true expectation for

L large.

We now substitute (3.4) into (2.14) to find the Monte Carlo approximation to the optimal quantizer for compressed sensing. It becomes

$$\lambda_{\text{cs}}^{(L)}(t) = C\left(\gamma_{\text{cs}}^{(L)}(t) f_{\mathbf{y_i}}(t)\right)^{1/3}, \tag{3.5}$$

for some normalization constant C. Again by law of large numbers arguments,

$$\lambda_{\text{cs}}^{(L)}(t) \xrightarrow{p} \lambda_{\text{cs}}(t) \tag{3.6}$$

for L large.

3.5 Experimental Results

We compare the CS-optimized quantizer, called the "sensitive" quantizer, to a uniform quantizer and "ordinary" quantizer which is optimized for the distribution of $\mathbf{y}$ through (2.10). The ordinary quantizer would be best if we want to minimize distortion between $\mathbf{y}$ and $\hat{\mathbf{y}}$, and hence has a flat sensitivity curve over the support of $\mathbf{y}$. The sensitive quantizer $\lambda_{\text{cs}}(t)$ is found using (3.5) and the uniform quantizer $\lambda_{\text{uni}}(t)$ is constant and normalized to integrate to 1.

If we restrict ourselves to fixed-rate scalar quantizers, the high-resolution approximation for quantization distortion (2.11) can be used. The distortion for an arbitrary quantizer $\lambda_q(t)$ with rate R is

$$\begin{aligned} D(R) &\approx 2^{-2R} E\left[\frac{\gamma_{\text{cs}}^2(\mathbf{y_i})}{12\lambda_q^2(\mathbf{y_i})}\right] \\ &= 2^{-2R} \int \frac{\gamma_{\text{cs}}^2(t) f_{\mathbf{y_i}}(t)}{12\lambda_q^2(t)} dt. \end{aligned} \tag{3.7}$$

Using 1000 Monte Carlo trials, we estimate $\gamma_{\text{cs}}(t)$ in Figure 3-3. Note that the estimate is found through importance sampling since there is low probability of getting samples for large $\mathbf{y_i}$ in Monte Carlo simulations. The sensitivity is symmetric and

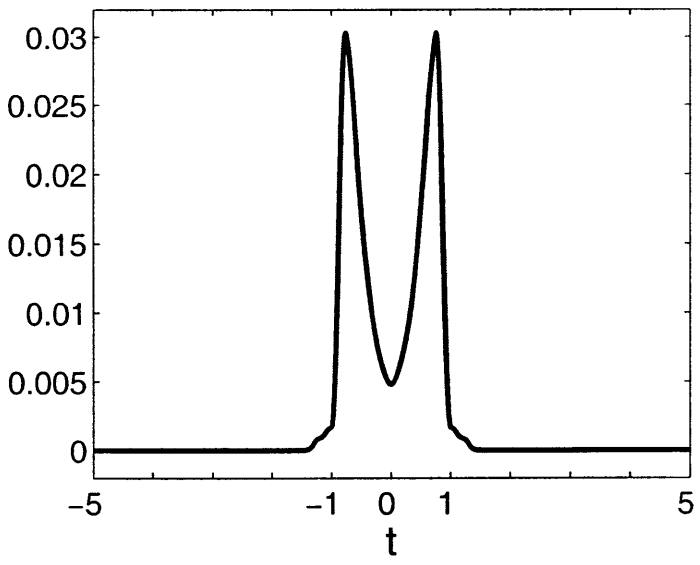


Figure 3-3: Estimated sensitivity $\gamma_{\text{cs}}(t)$ via Monte Carlo trials and importance sampling for $(K, M, N) = (5, 71, 100)$.

peaks away from zero because of the structure in (3.3). Some intuition is provided in Appendix 3.C for the scalar case. The point density functions for the three quantizers are illustrated in Figure 3-4.

Experimental results are performed on a Matlab testbench. Practical quantizers are designed by extracting codewords from the cdf of the normalized point densities. In the approximation, the ith codeword is the point t such that

$$\int_{-\infty}^{t} \lambda_{\text{cs}}(t')dt' = \frac{i - 1/2}{2^{R_i}},$$

where R_i is the rate for each measurement. The partition points are then chosen to be the midpoints between codewords.

We compare the sensitive quantizer to uniform and ordinary quantizers using the parameters $\sigma^2 = 0.3$ and $\mu = 0.1$. Results are shown in Figure 3-5.

We find the sensitive quantizer performs best in experimental trials for this combination of σ^2 and μ at sufficiently high rates. This makes sense because $\lambda_{\text{cs}}(t)$ is a high-resolution approximation and should not necessarily perform well at very low rates. Numerical comparisons between experimental data and the estimated quanti-

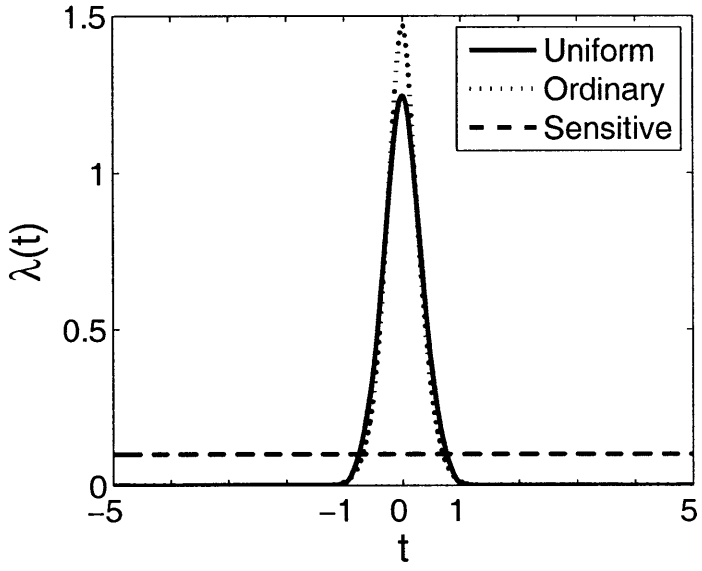


Figure 3-4: Estimated point density functions $\lambda_{\text{cs}}(t)$, $\lambda_{\text{ord}}(t)$, and $\lambda_{\text{uni}}(t)$ for $(K, M, N) = (5, 71, 100)$.

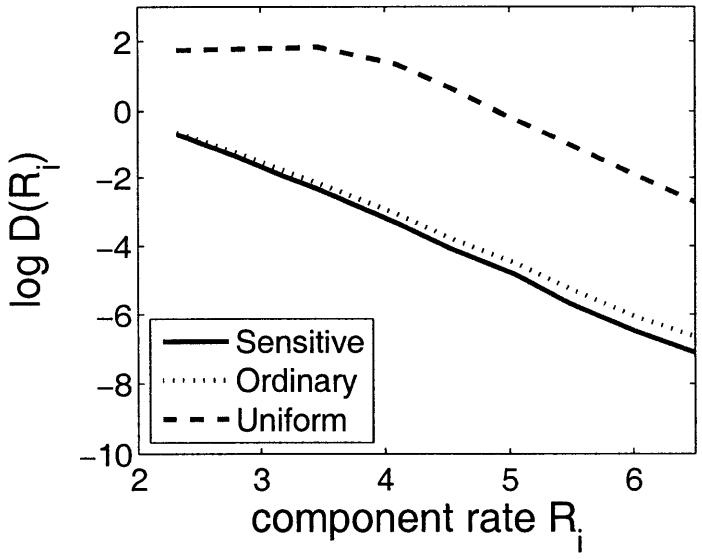


Figure 3-5: Results for distortion-rate for the three quantizers with $\sigma^2 = 0.3$ and $\mu = .01$. We see the sensitive quantizer has the least distortion.

zation distortion in (3.7) are similar.

3.A Proof of Lemma 3.1

Consider a unit hypersphere of dimension $M-1$ centered at origin in the space of $\mathbb{R}^M$. We draw N vectors $\boldsymbol{\phi}_\mathbf{j}$ uniformly from the surface of the hypersphere and form a matrix $\Phi \in \mathbb{R}^{M\times N}$. We show that every entry of $\boldsymbol{\Phi}$ is identically distributed and the cdf of each entry $\boldsymbol{\Phi}_{\mathbf{ij}}$ is

$$F_{\boldsymbol{\Phi}_{\mathbf{ij}}}(v,M) = \begin{cases} 1-T(v,M), & 0 \le v \le 1; \\ T(-v,M), & -1 \le v < 0; \\ 0, & \text{otherwise}, \end{cases}$$

where

$$T(v,M) = \frac{\Gamma(\frac{M}{2})}{\sqrt{\pi}\,\Gamma(\frac{M-1}{2})} \int_0^{\arccos(v)} (\sin\theta)^{M-2}\, d\theta$$

and $\Gamma(\cdot)$ is the Gamma function.

Proof. We begin by noting that an $(M-1)$ hypersphere (in $\mathbb{R}^M$) with radius R has surface area

$$S(M,R) = \frac{2\pi^{\frac{M}{2}} R^{M-1}}{\Gamma\left(\frac{M}{2}\right)}, \tag{3.8}$$

where $\Gamma(\cdot)$ is the Gamma function.

Because we are drawing uniformly over the shell, the fraction of the hypersphere in the region satisfying the constraint $\mathbf{v_i} > v$ is $\Pr(\mathbf{v_i} > v)$. This is found through the integration

$$\Pr(\mathbf{v_i} > v) = \frac{1}{S(M,1)} \int_0^{\arccos(v)} S(M-1, \sin\theta) d\theta$$

for $v \in [-1,1]$.

Geometrically, the integration is over every $(M-2)$ hypersphere in the region $\{v \le \mathbf{v_i} \le 1\}$ (or equivalently $\{\arccos(v) \ge \theta \ge 0\}$). Figure 3-6 visualizes the integration for the case $M=3$.

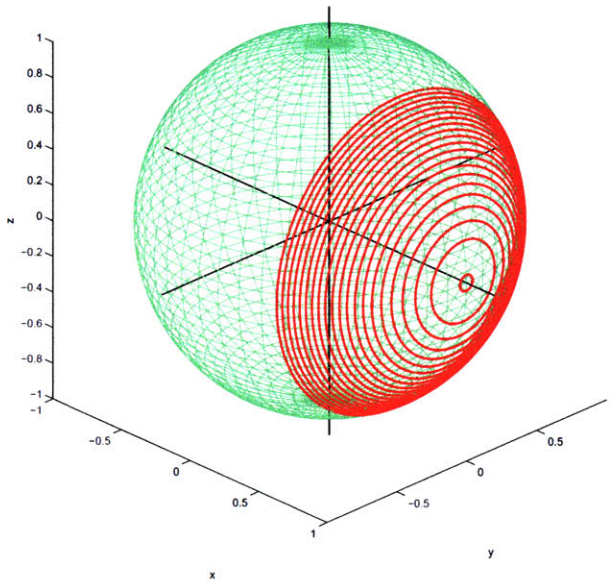


Figure 3-6: Integration over the circles that form the unit spherical shell for $0 \leq \theta \leq \arccos(v)$.

We simplify the integral using (3.8):

$$\begin{aligned}
\Pr(\mathbf{v_i} > v) &= \frac{\Gamma(\frac{M}{2})}{2\pi^{\frac{M}{2}}} \int_0^{\arccos(v)} \frac{2\pi^{\frac{M-1}{2}} (\sin\theta)^{M-2}}{\Gamma(\frac{M-1}{2})} d\theta \\
&= \frac{\Gamma(\frac{M}{2})}{\sqrt{\pi}\Gamma(\frac{M-1}{2})} \int_0^{\arccos(v)} (\sin\theta)^{M-2} \, d\theta \\
&\triangleq T(v, M).
\end{aligned}$$

By symmetry, $\Pr(\mathbf{v_i} > v) = \Pr(\mathbf{v_i} < -v)$ and hence the cdf is

$$F_{\mathbf{v_i}}(v, M) = \Pr(\mathbf{v_i} < v) = \begin{cases} 1 - T(v, M), & 0 \leq v \leq 1; \\ T(-v, M), & -1 \leq v < 0; \\ 0, & \text{o.w.} \end{cases}$$

□

We use symmetry arguments to show each entry in $\mathbf{v}$ has the same distribution.

Since $\mathbf{\Phi}$ is constructed from N independent $\mathbf{v}$ vectors, we can also use symmetry arguments to show every entry $\mathbf{\Phi_{ij}}$ is identically distributed. We now make some remarks about the distribution of $\mathbf{\Phi_{ij}}$:

1) Unfortunately, the distribution of $\mathbf{\Phi_{ij}}$ is difficult to compute analytically (it is a sum of hypergeometric functions). Instead, one can use numerical methods to find both $F_{\mathbf{v_i}}(v, M)$ and the probability density function (pdf) $f_{\mathbf{v_i}}(v, M)$.

2) The distribution is always symmetric around 0 and has a support of $[-1, 1]$.

3) As M increases, the distribution becomes more peaked around 0 and approaches a Dirac delta.

4) Because every column vector must have unit norm, the entries in a column of Φ are not independent. However, since the columns are chosen independently, entries across rows are independent.

One can easily generate $\mathbf{v}$ by just creating a random Gaussian $M \times N$ matrix and then rescaling the columns to have unit norm. Figure 3-7 compares the results from this lemma (red line) with the empirical distribution (blue bars) formed through Monte Carlo trials for several values of M. We see the empirical results match the theoretical distribution very well.

3.B Functional Quantization Example

We present a pedagogical fixed-rate scalar quantizer example to build intuition for functional quantization. Assume $\mathbf{y_1}$ and $\mathbf{y_2}$ are uniform random variables, iid $\mathcal{U}(0, 1)$. They are quantized separately and we wish to minimize the distortion of the function $g(\mathbf{y_1}, \mathbf{y_2}) = \mathbf{y_1}^2 + \mathbf{y_2}^2$.

By (2.10), the best quantizer to minimize distortion for each $\mathbf{y_i}$ is uniform. However, it is clear that this is not the optimal choice for $g(\mathbf{y_1}, \mathbf{y_2})$ since a small perturbation for larger $\mathbf{y_i}$ leads to a larger distortion penalty in the function g.

Instead, we can find the "sensitive" quantizer for this example. By (2.13), the

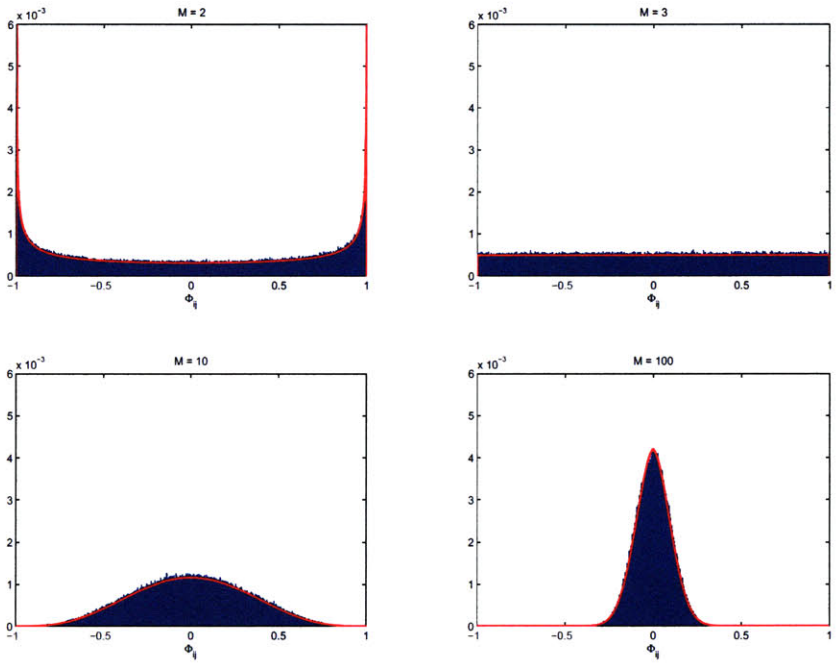


Figure 3-7: Theoretical versus empirical distribution for different values of M to validate Lemma 1.

sensitivities are $\gamma_i(t) = 2t$, meaning the optimal point density is

$$\lambda(t) = \begin{cases} \frac{5}{3}t^{2/3}, & t \in [0,1]; \\ 0, & \text{o.w.} \end{cases} \tag{3.9}$$

The quantization cells of this quantizer are visualized in Figure 3-8. As predicted, larger values of $\mathbf{y_i}$ have higher resolution because those values have more weight in the distortion calculation.

Using (2.15), the resulting distortion of the sensitive quantizer is approximately $\frac{18}{125}2^R$, where R is the rate for each quantizer. This is better than the ordinary quantizer (optimal for the observations), which leads to a distortion on g of $\frac{2}{3}2^R$.

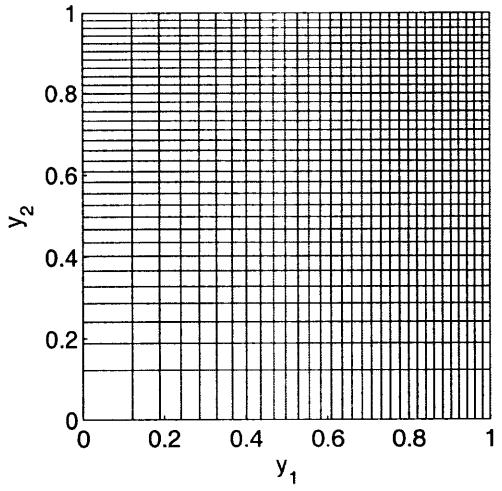


Figure 3-8: Quantization cells of a sensitive quantizer for the function $g(\mathbf{y_1}, \mathbf{y_2}) = \mathbf{y_1}^2 + \mathbf{y_1}^2$. The cells for larger values of $\mathbf{y_i}$ are smaller and hence have better resolution.

3.C Scalar Lasso Example

We consider a scalar version of the quantized compressed sensing problem. This example has little applicable purpose, but illuminates how the sensitive quantizer is shaped for lasso.

Assume we implement the system shown in Figure 3-9. The scalar random variable $\mathbf{x}$ is assumed to have the Laplacian pdf

$$f_{\mathbf{x}}(x; m, b) = \frac{1}{2b} \exp\left(-\frac{|x-m|}{b}\right).$$

Meanwhile, $\boldsymbol{\eta}$ is additive Gaussian noise and the reconstruction function g is the MAP estimator

$$\hat{x} = \arg\min_x \left(\|y - x\|_2^2 + \mu\|x\|_1\right) \tag{3.10}$$

that we call *scalar lasso*. Like lasso, it has a signal fidelity and sparsity tradeoff, with the regularization parameter μ determined by the noise variance.

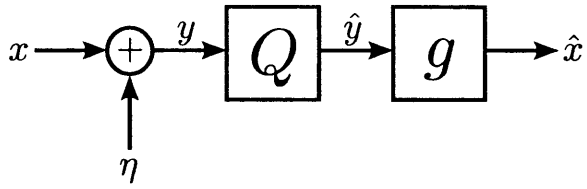


Figure 3-9: Model for scalar lasso example. Assume $\mathbf{x}$ is a compressible source and $\mathbf{y} = \mathbf{x} + \boldsymbol{\eta}$. The measurements are scalar quantized and then used to reconstruct $\hat{\mathbf{x}}$ through scalar lasso.

Functionally, the reconstruction is

$$g(x) = \begin{cases} x + \mu, & x < -\frac{\mu}{2}; \\ 0, & x < -\frac{\mu}{2} \leq x \leq \frac{\mu}{2}; \\ x - \mu, & x > \frac{\mu}{2}. \end{cases} \tag{3.11}$$

The scalar lasso function $\hat{\mathbf{x}} = g(\mathbf{x})$ is shown in Figure 3-10, along with the ordinary and sensitive quantizers. The ordinary scalar quantizer, where the point density is optimized for $f_{\mathbf{y}}(y)$, is represented by the diamonds. Meanwhile, the sensitive quantizer also takes into account the sensitivity and is represented by the dots on the same plot (using don't-care regions as described in [50]). Similar to the results presented in the vector case (Section 3.5), the sensitive quantizer puts less weight near zero due to lasso shrinkage.

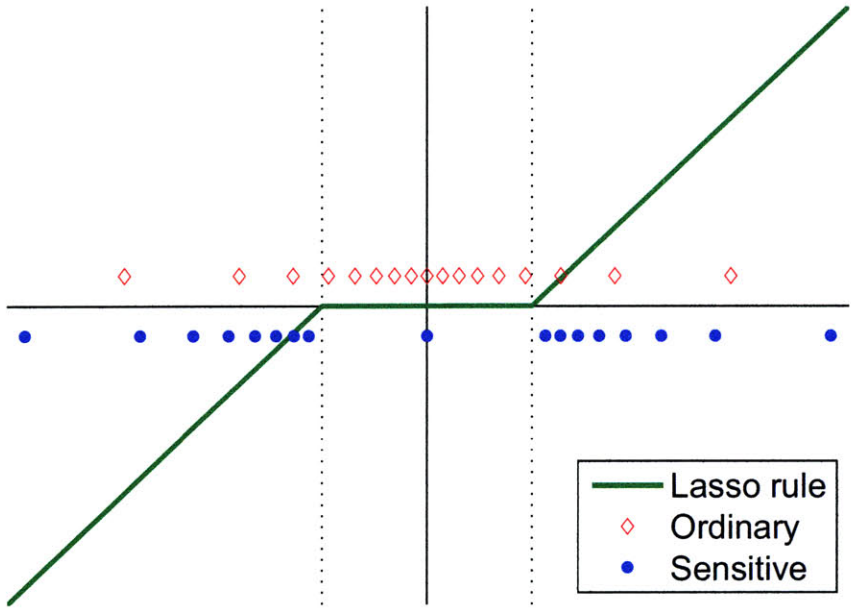


Figure 3-10: Scalar lasso and its quantizers. The functional form of lasso is represented by the solid green line and demonstrates lasso shrinkage in that the output has less energy than the input. The ordinary quantizer is shown in the red diamonds and the sensitive quantizer is represented by the blue dots.

Chapter 4

Performance Bounds for Estimation in CSN

One major application of sensor networks is surveillance and persistent monitoring for signals of interest (SOI). If these SOIs are sparse in some basis, then compressive sensor networks can be used to allow sampling much below Nyquist and a large reduction in information storage and communication. Two methods in the literature for geolocation of SOIs are time difference of arrival (TDOA) and angle of arrival (AoA). For TDOA, the architecture is usually single-sensor multi-platform (SSMP), and the data from each sensor is sent to a fusion center (FC) for location estimation. For AoA, we consider a multi-sensor single-platform (MSSP) architecture, where the platform does the computation onboard and outputs a direction estimate.

In this chapter, we derive performance bounds on estimators for both types of geolocation. This work is a subset of a joint MIT-Lincoln Laboratory project [63] that studies applications and theory for such networks. The algorithms developed in the project are used as comparisons to the bounds.

4.1 Related Work

Numerous papers have studied performance bounds in the context of both time delay [64, 65] and angle estimation [66] for geolocation. As discussed in Section 2.3, the

well-known SNR threshold effect of the Cramér-Rao bound (CRB) is observed for these scenarios due to the nonlinearity of the estimation, meaning the CRB is an overly optimistic bound for low-SNR regimes. More complex bounds such as Ziv-Zakai perform better, as illustrated in [67].

Other relevant related works include recent investigations on performance bounds for compressed sensing (CS) estimators [68, 69]. In both papers, the authors found that the oracle CRB (with knowledge of the sparsity pattern) is approached by maximum-likelihood (ML) estimators. However, there are no comparisons to more practical CS reconstruction algorithms.

4.2 Contribution

We consider performance bounds on CSN nodes being used as delay or angle estimators for geolocation and object tracking. Although related topics have been explored, this specific question has not been addressed in the literature.

We use the bounds as a benchmark for the practical algorithms developed in [63]. These algorithms are outside the scope of the thesis and we defer to the Lincoln Labs report for the details. However, numerical results are shown in Section 4.3.4 and 4.4.3 for comparison with bounds.

For TDOA and AoA, we find that there is a very large gap between the Cramér-Rao bound, as described in Section 2.3, and the practical algorithms used in numerical simulations, especially at low SNR. This agrees with previous work in estimation, which finds the CRB to be an optimistic lower bound, and recent research in performance bounds for CS, which has only shown the computationally-intractable ML estimator approaching the oracle CRB. Nevertheless, our work demonstrates some interesting relationships between model parameters such as the signal dimensions, sparsity ratio and delay factor in estimation performance. Also, it demonstrates that practical signal reconstruction algorithms can approach the CRB.

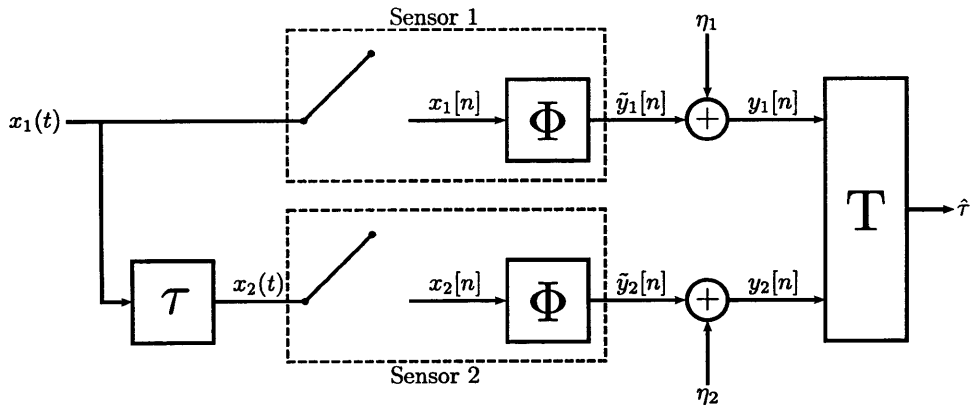


Figure 4-1: Problem model for time difference of arrival. The original signal $x_1(t)$ and its delayed version $x_2(t)$ are observed by different sensors, compressed using Φ, and transmitted in noise. An estimator T determines the predicted delay $\hat{\tau}$.

4.3 Time Difference of Arrival

4.3.1 Problem Model

Assume the SOI corresponds to a band-limited Fourier series with fundamental frequency f_o and $\Omega_{max} < B$. Further assume the signal is K-sparse in the Fourier basis, meaning only K Fourier series coefficients are nonzero. Each nonzero coefficient is random, and for the following derivations assume that the distribution is Gaussian with zero mean and variance σ_X^2. The SOI is observed at two locations with analog signals $x_1(t)$ and $x_2(t)$. The time delay between the two sensors is τ, meaning $x_2(t) = x_1(t-\tau)$. The two signals are then processed by identical and synced ADCs, as shown in Figure 4-1. The ADCs are sampled above the Nyquist sampling period $T = \frac{\pi}{B}$ and produce outputs $x_1[n]$ and $x_2[n]$, corresponding to the original and delayed discrete-time signals respectively.

Take two length-N data vectors $\mathbf{x_1}$ and $\mathbf{x_2}$ from time segments of $x_1[n]$ and $x_2[n]$ respectively. The two vectors are multiplied by an $M \times N$ matrix Φ corresponding to a particular downsampling method with factor $d = N/M$. The noiseless measurement vectors $\tilde{\mathbf{y}}_\mathbf{i}$ are then transmitted through a channel with additive noise $\boldsymbol{\eta}_\mathbf{i} \sim \mathcal{N}(0, \sigma_N^2 I)$. The observed measurement vector at the receiver is denoted $\mathbf{y_i} = \tilde{\mathbf{y}}_\mathbf{i} + \boldsymbol{\eta}_\mathbf{i} = \Phi \mathbf{x_i} + \boldsymbol{\eta}_\mathbf{i}$. The receiver then produces an estimate $\hat{\tau}$ of the timing delay between $x_1(t)$ and $x_2(t)$

using the observations $\mathbf{y_1}$ and $\mathbf{y_2}$. Note that the projection using Φ can be performed before sampling without a change in the analysis.

Clearly, $x_1[n]$ and $x_2[n]$ are highly correlated, and we can find the timing delay by estimating the phase shift between the two signals. More specifically, for any frequency component Ω, $x_2[n] = \cos(\omega n - \Omega\tau)$ if $x_1[n] = \cos(\omega n)$, where $\omega = \Omega T$. The delay τ is distinguishable up to the modulus of $\frac{2\pi}{\Omega}$ since the signal is periodic. However, the sum of signals is distinguishable up to the modulus of $\frac{2\pi}{\text{GCD}(\Omega_1,\Omega_2,\ldots,\Omega_k)}$, which allows a larger range of τ to be identifiable if the signal frequencies have a very small greatest common denominator (GCD).

Moreover, if the frequency components are harmonics of the fundamental frequency $\frac{1}{NT}$ (as assumed in the problem model), then the discrete Fourier transforms (DFTs) of N samples will be related by the following:

$$X_2[k] = X_1[k]e^{f(k)\tau}, \tag{4.1}$$

$$f(k) = \begin{cases} -j\frac{2\pi k}{NT}, & 0 < k < N/2; \\ -j\frac{2\pi(k-N)}{NT}, & N/2 < k < N; \\ 0, & \text{o.w.} \end{cases} \tag{4.2}$$

We can extract τ exactly from $\mathbf{X_1}$ and $\mathbf{X_2}$ in the noiseless, Nyquist-sampled case. Surprisingly, for large N and random sparsity, the above relationship can be used to find τ even in the presence of noise and aliasing.

4.3.2 CRB for Time Difference

We aim to find a Cramér-Rao bound for the delay estimator $\hat{\boldsymbol{\tau}} = T(\mathbf{y_1}, \mathbf{y_2})$ in Figure 4-1. Assume $\hat{\boldsymbol{\tau}}$ exploits (4.1) by taking the DFTs of $\mathbf{x_1}$ and $\mathbf{x_2}$, defined as $\mathbf{X_1}$ and $\mathbf{X_2}$ respectively. As specified in the problem model, $\mathbf{X_1}$ is K-sparse with some random sparsity pattern $\mathbf{J}$ chosen uniformly among all possibilities. As a reminder, the nonzero entries of $\mathbf{X_1}$ are assumed iid $\mathcal{N}(0, \sigma_X^2)$. Define the vector of parameters to be $\boldsymbol{\theta} = (\tau, \mathbf{X_1})^T$ and the observed data be $\mathbf{y} = (\mathbf{y_1}, \mathbf{y_2})^T$. The likelihood function

of $\mathbf{y}$ is

$$\begin{aligned} p(\mathbf{y};\boldsymbol{\theta}) &= p(\mathbf{y_2}|\mathbf{y_1},\boldsymbol{\theta})p(\mathbf{y_1}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &= p(\mathbf{y_2};\tau|\mathbf{y_1},\mathbf{X_1})p(\mathbf{y_1};\tau|\mathbf{X_1};)p(\mathbf{X_1}) \\ &= p(\mathbf{y_2};\tau|\mathbf{X_1})p(\mathbf{y_1}|\mathbf{X_1})p(\mathbf{X_1}). \end{aligned} \tag{4.3}$$

There is a nice decoupling of $\mathbf{y_1}$ and $\mathbf{y_2}$, which is surprising considering the correlation between the two signals. This is because the parameters $\mathbf{X_1}$ and τ provide all the information needed to determine the likelihood functions from the observed data. Since $\mathbf{x_i} = \Psi\mathbf{X_i}$, where Ψ is the inverse DFT matrix, each measurement vector is $\mathbf{y_i} = \Phi\Psi\mathbf{X_i} + \boldsymbol{\eta_i} = \tilde{\Phi}\mathbf{X_i} + \boldsymbol{\eta_i}$. The conditional distributions are

$$p(\mathbf{y_1}|\mathbf{X_1}) \sim \mathcal{N}(\tilde{\Phi}\mathbf{X_1}, \sigma_N^2 I); \tag{4.4}$$

$$p(\mathbf{y_2};\tau|\mathbf{X_1}) \sim \mathcal{N}(\tilde{\Phi}\mathbf{X_2}, \sigma_N^2 I). \tag{4.5}$$

where $\mathbf{X_2}$ is related to $\mathbf{X_1}$ through (4.1).

We now introduce some notation to simplify the following derivations. Define the rows of $\tilde{\Phi}$ to be $\tilde{\Phi}_m$, $P \equiv \text{diag}\left(\{f(k)\}_{k=0}^{N-1}\right)$, and $Q \equiv \text{diag}\left(\{e^{f(k)\tau}\}_{k=0}^{N-1}\right)$. The Cramér-Rao lower bound of the variance of $\hat{\boldsymbol{\tau}}$ is then shown in Theorem 4.1.

Theorem 4.1. *The variance of any unbiased delay estimator $\hat{\boldsymbol{\tau}}$ is bounded below by the inverse of the Fisher information $I_{\text{td}}(\tau)$, such that*

$$Var(\hat{\boldsymbol{\tau}}) \geq \binom{N}{K} \frac{\sigma_N^2}{\sigma_X^2} \left(\sum_{\mathbf{J_i} \in \mathbf{J}} \left(\sum_{m=1}^{M} (\tilde{\Phi}_m P Q \mathbf{J_i})^2 \right) \right)^{-1}.$$

Proof. Using Definition 2.3 in Section 2.3, the variance of any unbiased estimator is bounded by the Fisher information (FI). For this problem model, the FI here is found through the following steps:

$$
\begin{aligned}
I_{\text{td}}(\tau) &= E\left[\left(\frac{\partial}{\partial\tau}\ln p(\mathbf{y};\theta)\right)^2\right] \\
&\overset{(a)}{=} E\left[\left(\frac{\partial}{\partial\tau}\left\{\ln p(\mathbf{y_2};\tau|\mathbf{X_1}) + \ln p(\mathbf{y_1}|\mathbf{X_1}) + \ln p(\mathbf{X_1})\right\}\right)^2\right] \\
&\overset{(b)}{=} E\left[\left(\frac{\partial}{\partial\tau}\ln p(\mathbf{y_2};\tau|\mathbf{X_1})\right)^2\right] \\
&\overset{(c)}{=} E\left[\left(\frac{\partial}{\partial\tau}\left\{-\frac{M}{2}\ln 2\pi\sigma_N^2 - \frac{1}{2\sigma_N^2}\sum_{m=1}^{M}(y_{2m} - \tilde{\Phi}_m\mathbf{X_2})^2\right\}\right)^2\right] \\
&\overset{(d)}{=} E\left[\left(\frac{1}{\sigma_N^2}\sum_{m=1}^{M}(y_{2m} - \tilde{\Phi}_m\mathbf{X_2})(\tilde{\Phi}_m P\mathbf{X_2})\right)^2\right] \\
&\overset{(e)}{=} E_{\boldsymbol{\eta_2},\mathbf{X_1}}\left[\left(\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\boldsymbol{\eta_{2m}}(\tilde{\Phi}_m P\mathbf{X_2})\right)^2\right] \\
&\overset{(f)}{=} E_{\mathbf{X_1}}\left[\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\left(\tilde{\Phi}_m P\mathbf{X_2}\right)^2\right] \\
&\overset{(g)}{=} E_{\mathbf{J}}\left[E_{\mathbf{X_1}}\left[\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\left(\tilde{\Phi}_m P\mathbf{X_2}\right)^2|\mathbf{J}=J_i\right]\right] \\
&\overset{(h)}{=} E_{\mathbf{J}}\left[\frac{1}{\sigma_N^2}\sum_{m=1}^{M}(\tilde{\Phi}_m PQJ_i)^2 E[\mathbf{X_1}^2]|\mathbf{J}=J_i\right] \\
&\overset{(i)}{=} E_{\mathbf{J}}\left[\frac{\sigma_X^2}{\sigma_N^2}\sum_{m=1}^{M}(\tilde{\Phi}_m PQ\mathbf{J})^2\right] \\
&\overset{(j)}{=} \frac{1}{\binom{N}{K}}\sum_{\mathbf{J_i}\in\mathbf{J}}\left(\frac{\sigma_X^2}{\sigma_N^2}\sum_{m=1}^{M}(\tilde{\Phi}_m PQ\mathbf{J_i})^2\right),
\end{aligned}
$$

where (a) follows from taking the logarithm of (4.3); (b) follows from the fact that only the first term depends on τ; (c) follows from (4.5); (d) results from taking the partial derivative; (e) comes from noting $\mathbf{y_2} = \tilde{\Phi}\mathbf{X_2} + \boldsymbol{\eta_2}$; (f) follows from taking the expectation over $\boldsymbol{\eta_2}$; (g) uses the total probability theorem; (h) comes from moving the expectation into the sum; (i) follows the assumption that the entries of $\mathbf{X_1}$ have zero mean and variance σ_X^2; and (j) uses the fact that all sparsity patterns are equally

likely.

□

We make several observations using this theorem:

1) The expectation over $\mathbf{J}$ is computationally prohibitive to calculate. However, Monte Carlo trials will approach the true expectation quickly.

2) The Fisher information $I_{\text{td}}(\tau)$ scales linearly with the sparsity K. Since τ is extracted from the phase difference between $\mathbf{X_1}$ and $\mathbf{X_2}$, a DFT with more nonzero components will lead to more observations of τ, and hence a better estimator.

3) $I_{\text{td}}(\tau)$ decreases when the downsampling factor d increases. This agrees with our intuition that more downsampling creates larger signal distortion and a less reliable estimate $\hat{\boldsymbol{\tau}}$.

4) $I_{\text{td}}(\tau)$ does not vary with N for fixed M and K, as the Fisher information only depends on the nonzero entries of $\mathbf{X_1}$, not the total length.

5) Combining remarks 2 and 4, we conclude $I(\tau)$ will increase linearly with the sparsity ratio $\alpha = K/N$.

6) $I_{\text{td}}(\tau)$ increases linearly with SNR $= \sigma_X^2/\sigma_N^2$. It does not vary with the noise variance given a fixed SNR.

4.3.3 CRB for Time Difference and Signal Reconstruction

Now assume X_1 is a deterministic signal that is estimated along with the time difference. We will not only bound the variance of the estimator $\hat{\boldsymbol{\tau}}$ but also the estimator $\hat{\mathbf{X}}_\mathbf{1}$ for signal reconstruction. Assume the observed data is again $\mathbf{y} = (\mathbf{y_1}, \mathbf{y_2})^T$ and the vector of parameters is $\theta = (\tau, X_1)^T$, but this time all parameters are non-random. The likelihood function is

$$\begin{aligned} p(\mathbf{y};\theta) &= p(\mathbf{y_2};\theta|\mathbf{y_1})p(\mathbf{y_1};\theta) \\ &= p(\mathbf{y_2};\tau,X_1|\mathbf{y_1})p(\mathbf{y_1};\tau,X_1) \\ &= p(\mathbf{y_2};\tau,X_1)p(\mathbf{y_1};X_1). \end{aligned}$$

As discussed in Section 2.3, in this multi-parameter case we will need a Fisher information matrix (FIM) to find the CRB. To compute it, we need the following terms:

$$\begin{aligned}
\frac{\partial}{\partial X_{1k}} \ln p(\mathbf{y_1}|X_1) &= \frac{\partial}{\partial X_{1k}} \left\{ -\frac{M}{2} \ln 2\pi\sigma_N^2 - \frac{1}{2\sigma_N^2} \sum_{m=1}^{M} (y_{1m} - \tilde{\Phi}_m X_1)^2 \right\} \\
&= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} (y_{1m} - \tilde{\Phi}_m X_1) \tilde{\Phi}_{mk} \\
&= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \boldsymbol{\eta_{1m}} \tilde{\Phi}_{mk}, \\
\frac{\partial}{\partial X_{1k}} \ln p(\mathbf{y_2}; \tau|X_1) &= \frac{\partial}{\partial X_{1k}} \left\{ -\frac{M}{2} \ln 2\pi\sigma_N^2 - \frac{1}{2\sigma_N^2} \sum_{m=1}^{M} (y_{2m} - \tilde{\Phi}_m X_2)^2 \right\} \\
&= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} (y_{2m} - \tilde{\Phi}_m X_2) \tilde{\Phi}_{mk} e^{-j\frac{2\pi k\tau}{NT}} \\
&= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \boldsymbol{\eta_{2m}} \tilde{\Phi}_{mk} e^{-j\frac{2\pi k\tau}{NT}}, \\
\frac{\partial}{\partial \tau} \ln p(\mathbf{y_1}|X_1) &= 0, \\
\frac{\partial}{\partial \tau} \ln p(\mathbf{y_2}; \tau|X_1) &= \frac{\partial}{\partial \tau} \left\{ -\frac{M}{2} \ln 2\pi\sigma_N^2 - \frac{1}{2\sigma_N^2} \sum_{m=1}^{M} (y_{2m} - \tilde{\Phi}_m X_2)^2 \right\} \\
&= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} (y_{2m} - \tilde{\Phi}_m X_2) \left(\sum_{n=1}^{N} \tilde{\Phi}_{mn} X_{2n} \left(-j\frac{2\pi n}{NT} \right) \right) \\
&= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \boldsymbol{\eta_{2m}} \left(\sum_{n=1}^{N} \tilde{\Phi}_{mn} X_{2n} \left(-j\frac{2\pi n}{NT} \right) \right).
\end{aligned}$$

Putting all the components together, we get

$$\begin{aligned}
\frac{\partial}{\partial X_{1k}} \ln p(\mathbf{y}|\theta) &= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \boldsymbol{\eta_{1m}} \tilde{\Phi}_{mk} + \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \boldsymbol{\eta_{2m}} \tilde{\Phi}_{mk} e^{-j\frac{2\pi k\tau}{NT}}, \\
\frac{\partial}{\partial \tau} \ln p(\mathbf{y}|\theta) &= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \boldsymbol{\eta_{2m}} \left(\sum_{n=1}^{N} \tilde{\Phi}_{mn} X_{2n} \left(-j\frac{2\pi n}{NT} \right) \right).
\end{aligned}$$

The resulting Fisher information matrix $I_{\text{td/sr}}(\theta)$ for both time difference and

signal estimation then looks like $I_{\mathrm{td/sr}}(\theta) = \begin{pmatrix} A & C \\ B & D \end{pmatrix}$, where

$$
\begin{aligned}
A &= E\left[\frac{\partial}{\partial\tau}\ln p(\mathbf{y};\theta)\frac{\partial}{\partial\tau}\ln p(\mathbf{y};\theta)\right] \\
&= E\left[\left\{\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\boldsymbol{\eta_{2m}}\left(\sum_{n=1}^{N}\tilde{\Phi}_{mn}X_{2n}\left(-j\frac{2\pi n}{NT}\right)\right)\right\}\right. \\
&\qquad \left.\cdot\left\{\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\boldsymbol{\eta_{2m}}\left(\sum_{n=1}^{N}\tilde{\Phi}_{mn}X_{2n}\left(-j\frac{2\pi n}{NT}\right)\right)\right\}\right] \\
&= \frac{1}{\sigma^4}\sum_{m=1}^{M}E\left[\boldsymbol{\eta_{2m}}^2\right]\left(\sum_{n=1}^{N}\tilde{\Phi}_{mn}X_{2n}\left(-j\frac{2\pi n}{NT}\right)\right)^2 \\
&= \frac{1}{\sigma_N^2}\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\tilde{\Phi}_{mn}X_{1n}e^{-j\frac{2\pi n\tau}{NT}}\left(-j\frac{2\pi n}{NT}\right)\right)^2,
\end{aligned}
$$

$$
\begin{aligned}
(B)_{p1} &= E\left[\frac{\partial}{\partial X_{1p}}\ln p(\mathbf{y};\theta)\frac{\partial}{\partial\tau}\ln p(\mathbf{y};\theta)\right] \\
&= E\left[\left\{\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\boldsymbol{\eta_{1m}}\tilde{\Phi}_{mk}+\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\boldsymbol{\eta_{2m}}\tilde{\Phi}_{mk}e^{-j\frac{2\pi k\tau}{NT}}\right\}\right. \\
&\qquad \left.\cdot\left\{\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\boldsymbol{\eta_{2m}}\left(\sum_{n=1}^{N}\tilde{\Phi}_{mn}X_{2n}\left(-j\frac{2\pi n}{NT}\right)\right)\right\}\right] \\
&= \frac{1}{\sigma^4}\sum_{m=1}^{M}E\left[\boldsymbol{\eta_{2m}}^2\right]\tilde{\Phi}_{mp}e^{-j\frac{2\pi p\tau}{NT}}\left(\sum_{n=1}^{N}\tilde{\Phi}_{mn}X_{2n}\left(-j\frac{2\pi n}{NT}\right)\right) \\
&= \frac{1}{\sigma_N^2}\sum_{m=1}^{M}\tilde{\Phi}_{mp}e^{-j\frac{2\pi p\tau}{NT}}\left(\sum_{n=1}^{N}\tilde{\Phi}_{mn}X_{2n}\left(-j\frac{2\pi n}{NT}\right)\right),
\end{aligned}
$$

$$
\begin{aligned}
(C)_{1q} &= E\left[\frac{\partial}{\partial\tau}\ln p(\mathbf{y};\theta)\frac{\partial}{\partial X_{1q}}\ln p(\mathbf{y};\theta)\right] \\
&= E\left[\left\{\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\boldsymbol{\eta_{2m}}\left(\sum_{n=1}^{N}\tilde{\Phi}_{mn}X_{2n}\left(-j\frac{2\pi n}{NT}\right)\right)\right\}\right. \\
&\qquad \left.\cdot\left\{\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\boldsymbol{\eta_{1m}}\tilde{\Phi}_{mq}+\frac{1}{\sigma_N^2}\sum_{m=1}^{M}\boldsymbol{\eta_{2m}}\tilde{\Phi}_{mq}e^{-j\frac{2\pi q\tau}{NT}}\right\}\right]
\end{aligned}
$$

$$= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \tilde{\Phi}_{mq} e^{-j\frac{2\pi q\tau}{NT}} \left(\sum_{n=1}^{N} \tilde{\Phi}_{mn} X_{2n} \left(-j\frac{2\pi n}{NT} \right) \right),$$

$$\begin{aligned}
(D)_{pq} &= E\left[\frac{\partial}{\partial X_{1p}} \ln p(\mathbf{y};\theta) \frac{\partial}{\partial X_{1q}} \ln p(\mathbf{y};\theta) \right] \\
&= E\left[\left\{ \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \boldsymbol{\eta}_{\mathbf{1m}} \tilde{\Phi}_{mp} + \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \boldsymbol{\eta}_{\mathbf{2m}} \tilde{\Phi}_{mp} e^{-j\frac{2\pi k\tau}{NT}} \right\} \right. \\
&\qquad \left. \cdot \left\{ \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \boldsymbol{\eta}_{\mathbf{1m}} \tilde{\Phi}_{mq} + \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \boldsymbol{\eta}_{\mathbf{2m}} \tilde{\Phi}_{mq} e^{-j\frac{2\pi k\tau}{NT}} \right\} \right] \\
&= \frac{1}{\sigma^4} \sum_{m=1}^{M} E\left[\boldsymbol{\eta}_{\mathbf{1m}}{}^2 \right] \tilde{\Phi}_{mp} \tilde{\Phi}_{mq} + \frac{1}{\sigma^4} \sum_{m=1}^{M} E\left[\boldsymbol{\eta}_{\mathbf{2m}}{}^2 \right] \tilde{\Phi}_{mp} \tilde{\Phi}_{mq} e^{-j\frac{2\pi(p+q)\tau}{NT}} \\
&= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \tilde{\Phi}_{mp} \tilde{\Phi}_{mq} \left(1 + e^{-j\frac{2\pi(p+q)\tau}{NT}} \right).
\end{aligned}$$

To summarize,

$$\begin{aligned}
A &= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \left(\sum_{n=1}^{N} \tilde{\Phi}_{mn} X_{1n} e^{-j\frac{2\pi n\tau}{NT}} \left(-j\frac{2\pi n}{NT} \right) \right)^2, \\
(B)_{p1} &= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \tilde{\Phi}_{mp} e^{-j\frac{2\pi p\tau}{NT}} \left(\sum_{n=1}^{N} \tilde{\Phi}_{mn} X_{2n} \left(-j\frac{2\pi n}{NT} \right) \right), \\
(C)_{1q} &= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \tilde{\Phi}_{mq} e^{-j\frac{2\pi q\tau}{NT}} \left(\sum_{n=1}^{N} \tilde{\Phi}_{mn} X_{2n} \left(-j\frac{2\pi n}{NT} \right) \right), \\
(D)_{pq} &= \frac{1}{\sigma_N^2} \sum_{m=1}^{M} \tilde{\Phi}_{mp} \tilde{\Phi}_{mq} \left(1 + e^{-j\frac{2\pi(p+q)\tau}{NT}} \right).
\end{aligned}$$

Unfortunately, $I_{\text{td/sr}}(\theta)$ is not invertible for two reasons. First, Φ downsamples the original signal and thus D is not full rank. Second, since the signal is known a priori to be K-sparse, it does not make sense to consider the partial derivatives in the zero entries in X_1. Instead, we consider an oracle CRB by projecting $I_{\text{td/sr}}(\theta)$ to the K-dimensional subspace spanned by the sparsity of the signal. This is equivalent to taking a constrained Fisher information matrix described in [70] with constraint set $\{X_{1i} = 0, i \notin J\}$. The resulting D_{cons} would be a full-rank K-by-K matrix which

is invertible. Moreover, the constrained Fisher information matrix $I_{\text{cons}}(\theta)$ contains only partial derivative terms on the nonzero entries of X_1. Now, we can express the submatrices of $I_{\text{cons}}(\theta)$ using the notation $[A]_J$ for a submatrix of A that contains the J rows (or in the case of row vectors, the J entries).

$$A_{\text{cons}} = \frac{1}{\sigma_N^2} \sum_{m=1}^{M} ([\tilde{\Phi}_m]_J[P]_J[X_2]_J)^T ([\tilde{\Phi}_m]_J[P]_J[X_2]_J), \tag{4.6}$$

$$B_{\text{cons}} = \frac{1}{\sigma_N^2} \left([\tilde{\Phi}]_J[Q]_J\right)^T \left([\tilde{\Phi}_m]_J[P]_J[X_2]_J\right), \tag{4.7}$$

$$C_{\text{cons}} = \frac{1}{\sigma_N^2} \left([\tilde{\Phi}_m]_J[P]_J[X_2]_J\right)^T \left([\tilde{\Phi}]_J[Q]_J\right), \tag{4.8}$$

$$D_{\text{cons}} = \frac{1}{\sigma_N^2} \left[\left([\tilde{\Phi}]_J\right)^T \left([\tilde{\Phi}]_J\right) + \left([\tilde{\Phi}]_J[Q]_J\right)^T \left([\tilde{\Phi}]_J[Q]_J\right) \right]. \tag{4.9}$$

The subsequent variance bounds on delay and reconstruction can be found using $I_{\text{cons}}(\theta)$.

Theorem 4.2. *For a given sparsity, the variance of any unbiased delay/signal-reconstruction estimator $\hat{\theta} = (\hat{\tau}, \hat{\mathbf{X}}_1)^T$ is bounded below by the inverse of the constrained Fisher information matrix $I_{\text{cons}}(\theta)$ corresponding to the CRB of an oracle bound that has the sparsity pattern J as side information. This bound is described by*

$$\text{Var}\,(\hat{\tau}) \geq [I_{\text{cons}}(\theta)^{-1}]_{11},$$

and

$$\text{Var}\left(\hat{X}_{1k}\right) \geq [I_{\text{cons}}(\theta)^{-1}]_{(k+1)(k+1)},$$

where $\hat{\mathbf{X}}_{\mathbf{1k}}$ corresponds to the estimator for the k-th nonzero entry of $\mathbf{X_1}$.

Proof. We use Definition 2.3 in Section 2.3 to define the CRB. The FIM can be found through (4.6)-(4.9). □

We now present some remarks about the theorem.

1) The estimator $\hat{\mathbf{X}}_\mathbf{1}$ is assumed to know the sparsity pattern so there is no variance on the zero entries. Thus, Theorem 4.2 is forced to be an oracle bound, similar to [69].

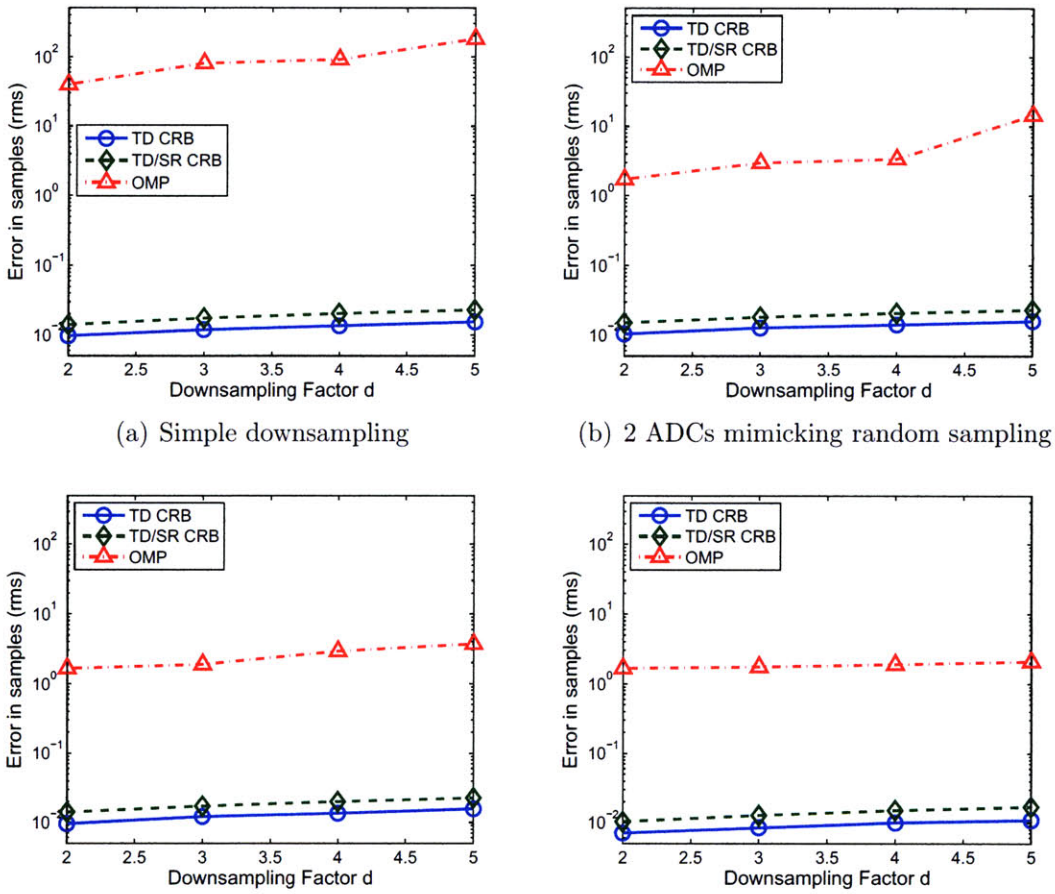
2) Both $\hat{\boldsymbol{\tau}}$ and $\hat{\mathbf{X}}_{\mathbf{1k}}$ satisfy the same scaling laws as in Theorem 4.1 for K, d, SNR and N. More specifically, the diagonal entries of the Fisher information increase with increasing K, decreasing d and increasing SNR. It is not affected by N.

4.3.4 Comparison to Experimental Results

In simulations, we assume an OFDM point source transmission with 2dB signal-to-noise ratio (SNR) sampled at 1GHz (demodulated with carrier frequency 2.2GHz). Using the discrete-frequency approximation, N is 4096, leading to a DFT that is 66-sparse. A bed of sensors receives the transmission, compresses the sampled data and transmits to a FC. The sensors are assumed to be on the same axis at a distance of 333.3m, corresponding to a maximum delay of 1.111 μs. The FC has an estimator $\hat{\boldsymbol{\tau}}$ that cross-correlates the reconstruction vectors $\hat{\mathbf{x}}_{\mathbf{1}}$ and $\hat{\mathbf{x}}_{\mathbf{2}}$ using an OMP variant [71] which takes as input measurement vectors $\mathbf{y_1}$ and $\mathbf{y_2}$. The time offset on $\hat{\mathbf{x}}_{\mathbf{2}}$ which yields the maximum correlation is chosen as the delay estimate. Thus, the estimate is discretized to the resolution of the sampling period T. We can increase the resolution of the τ estimation by upsampling the reconstruction signal vectors.

Four types of compression are considered in the experimental trials of TDOA, each represented by a matrix Φ as described in Section 4.3.1. The first method is simple downsampling and Φ is the identity matrix with all the rows taken out except every d-th (where d is the downsampling factor). The second is a sampler that consists of two ADCs with rates related by a ratio of 18/25. These ADCs are simple downsamplers but used together to mimic random sampling. The third method is actual random sampling, meaning Φ is an identity matrix with only a random set of rows not taken out. The final method is random projections where the matrix is full and the entries are random. Figure 4-2 illustrates the performance of the four estimators with the time difference CRBs found in Section 4.3.2 (blue solid lines) and Section 4.3.3 (green dashed lines).

We find a significant difference between both bounds and the experimental data. We also find that the an estimator that aims to perform signal recovery can be expected to do worse than just time difference estimation alone. This makes intuitive



(a) Simple downsampling

(b) 2 ADCs mimicking random sampling

(c) Random sampling

(d) Random projections

Figure 4-2: CRB for both time difference (TD) and time difference/signal reconstruction (TD/SR) versus OMP experimental error variance for each sampling matrix. We find a large gap between experimental results and the bound. This is because the CRB is optimistic for nonlinear estimation. Other reasons for the loose bounds are discussed.

sense because the problems of signal reconstruction and time difference estimation cannot be decoupled in this model. Hence, an estimator designed to have good signal reconstruction might not be the optimal estimator for τ alone.

We now provide some insight for the large difference between the lower bounds and experimental results. One reason is that the estimators can at best distinguish τ up to the sampling period T (or some fraction after upsampling). Thus, for large enough SNR, the estimator output will have no variance but some fixed error offset due to the discretization of τ. Another important factor for the disparity is that the estimators use Orthogonal Matching Pursuit (OMP) to reconstruct $x_1[n]$ and $x_2[n]$, and then infer τ, which may not be the optimal method.

There are also some issues with using a Cramér-Rao lower bound itself. In general, the Cramér-Rao bound is only tight for unbiased estimators in the asymptotic regime. Moreover, in the case where the source signal is a nonlinear function of the parameter τ, the CRB may not be tight even under those conditions. There is a well known threshold phenomenon in the literature such that for SNR below some level, the estimator deviates significantly from the CRB. In a cross-correlation estimator, this is because the noise effectively swamps out relevant information and the estimator chooses a value based on the a priori distribution of the parameter. In Figure 4-3, we illustrate this threshold effect in one estimator. Although the bound is not necessarily tight even at high SNR, the effect is more pronounced at low SNR.

Even though it is not tight, the CRB still proves to be useful in understanding the scaling of estimators in the TDOA model. Specifically, the trends mentioned in the remarks after Theorem 4.1 and Theorem 4.2 hold in simulation results. Tighter bounds for this nonlinear estimation problem include Ziv-Zakai and Weiss-Weinstein, which can be explored in future work.

The oracle bound derived in Section 4.3.3 also provides a lower bound on signal reconstruction error. Figure 4-4 compares these bounds to experimental results and reveals that reconstruction estimator is close to the bound. Since the observations are linear functions of the signal parameters, the CRB can be tight in this case. Our experiments validate the OMP algorithm is indeed close to the bound.

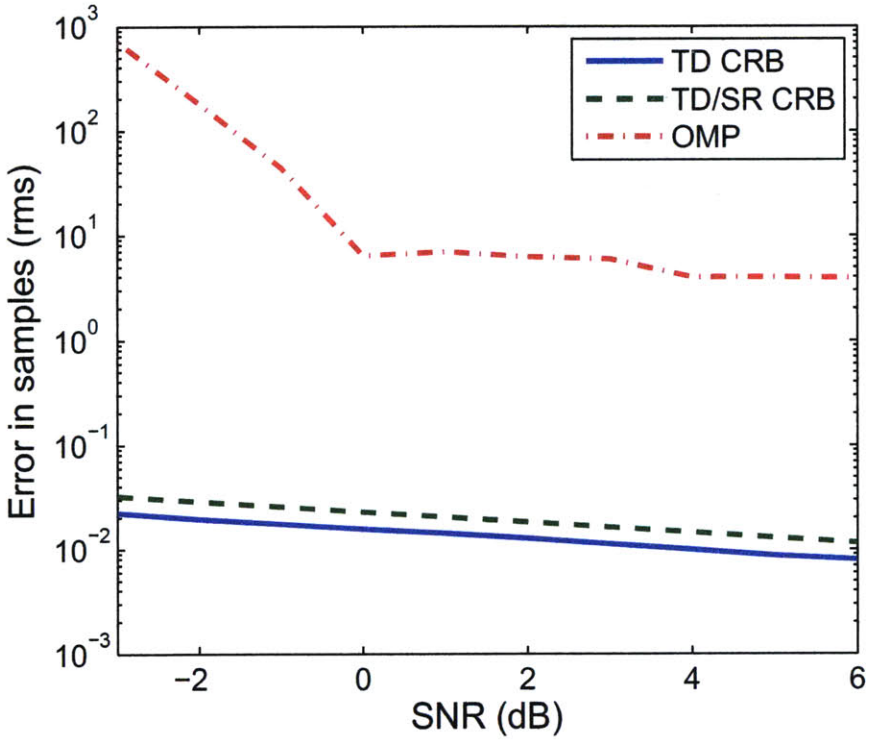


Figure 4-3: CRB and experimental results with SNR. There is a well-known phenomenon where the estimator performance dramatically deteriorates below a certain SNR. This corresponds to the point where the cross-correlation begins to fail.

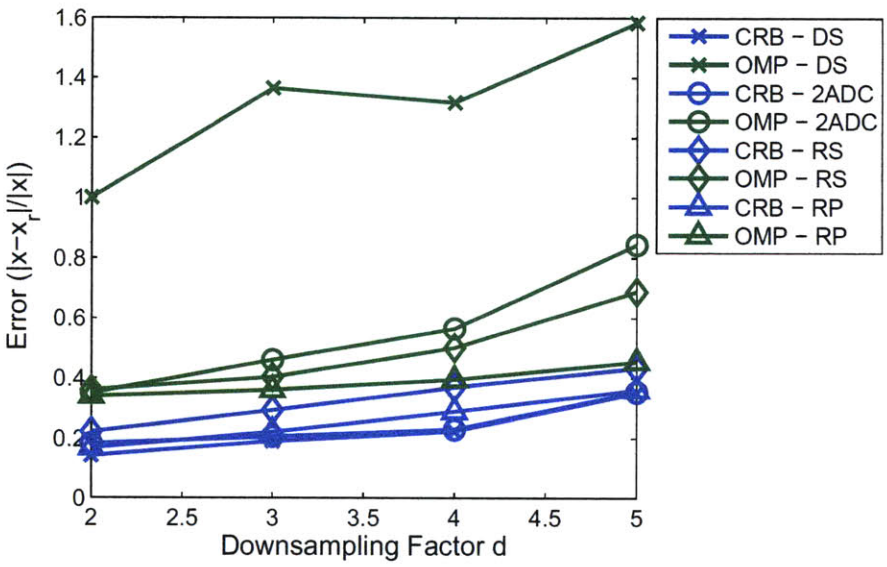


Figure 4-4: Experimental error variance versus CRB for signal reconstruction. Since this is a linear estimation problem, the bounds are much tighter than for time difference estimation.

4.4 Angle of Arrival

4.4.1 Problem Model

We now turn our attention to geolocation via angle of arrival, also called direction of arrival in the literature. In this case, the sensors are very close to each other on the same platform. Using far field and point source approximations, the SOI is modeled as a plane wave that passes through the sensor array with uniform power. Assume the platform contains S sensors and the transmission is coming from angle ϕ (relative to some defined coordinate space). The platform considers the data from each sensor and produces an estimate $\hat{\phi}$ of the direction of arrival.

Assume a reference sensor s_0 is located at $(0, 0)$ and every other sensor s_i is located at (x_{s_i}, y_{s_i}) for $1 \leq i \leq S-1$. The distance d_i the plane wave travels to get from the sensor s_i to s_0 is $x_{s_i}\cos(\phi) + y_{s_i}\sin(\phi)$. Note also that $d_i = c\tau_i$, where c is the speed of light and τ_i is the time difference between the transmission arriving at s_0 and s_i.

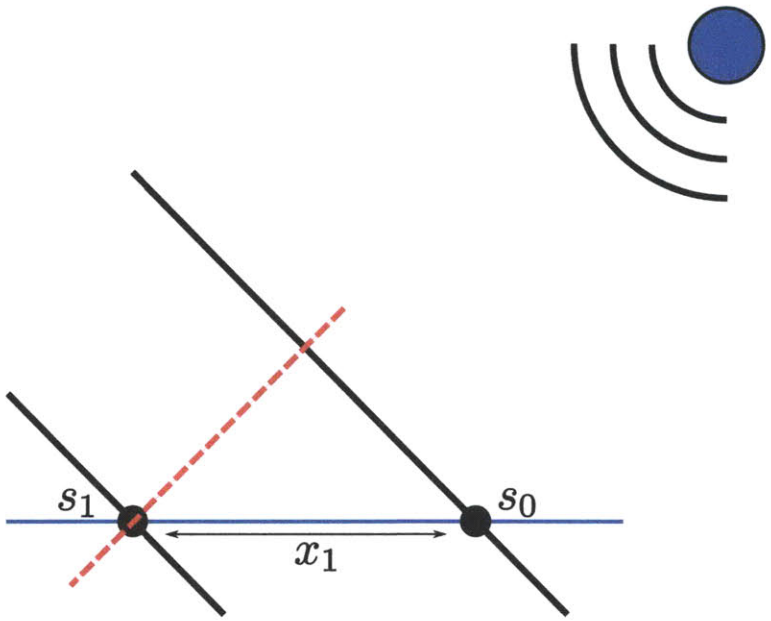


Figure 4-5: AoA problem model. We assume the transmission is sparse in a known basis and satisfy far field and point source approximations.

To simplify the problem, assume that all the sensors are on the x-axis. This reduces $d_i = c\tau_i = x_i \cos(\phi)$. The platform computes the estimated time difference $\hat{\boldsymbol{\tau}}_\mathbf{i}$ for each sensor and forms an AoA estimate $\hat{\boldsymbol{\phi}}_\mathbf{i} = \arccos\left(\frac{c}{x_i}\hat{\boldsymbol{\tau}}_\mathbf{i}\right)$. Figure 4-5 illustrates this setup for a pair of sensors.

4.4.2 CRB of Angle of Arrival

We previously formulated a CRB for TDOA in Section 4.3.2. We will now derive a similar bound for AoA estimation as a function of the bound for TDOA.

We begin by again pointing out $\hat{\boldsymbol{\phi}}_\mathbf{i} = \arccos\left(\frac{c}{x_i}\hat{\boldsymbol{\tau}}_\mathbf{i}\right)$. The variance of $\hat{\boldsymbol{\phi}}_\mathbf{i}$ can be approximated by the variance of $\hat{\boldsymbol{\tau}}_\mathbf{i}$ using a second-order Taylor expansion:

$$\operatorname{Var}\left(f(\mathbf{x})\right) \approx \left(f'(\mathrm{E}\left[\mathbf{x}\right])\right)^2 \operatorname{Var}\left(\mathbf{x}\right). \tag{4.10}$$

Theorem 4.3. *Given* $\hat{\boldsymbol{\phi}}_\mathbf{i} = \arccos\left(\frac{c}{x_i}\hat{\boldsymbol{\tau}}_\mathbf{i}\right)$, *the variance an estimator* $\hat{\boldsymbol{\phi}}_\mathbf{i}$ *is*

$$\mathrm{Var}\left(\hat{\boldsymbol{\phi}}_\mathbf{i}\right) \approx \left(\frac{\left(\frac{c}{x_i}\right)^2}{1-\left(\frac{c}{x_i}\tau_i\right)^2}\right)\mathrm{Var}(\hat{\boldsymbol{\tau}}_\mathbf{i})$$

assuming the estimator $\hat{\boldsymbol{\tau}}_\mathbf{i}$ *is unbiased.*

With S sensors, there will be $S-1$ estimates for ϕ. If the noise at each sensor is assumed to be independent, $\{\hat{\boldsymbol{\phi}}_\mathbf{i}\}_{i=1}^{S-1}$ are independent estimators of the same nonrandom parameter. By the weak law of large numbers and assuming each $\hat{\boldsymbol{\phi}}_\mathbf{i}$ is unbiased,

$$\frac{1}{S-1}\sum_{i=1}^{S-1}\hat{\boldsymbol{\phi}}_\mathbf{i} \xrightarrow{p} \phi.$$

This leads to the following theorem on a variance bound for the S-sensor AoA estimator:

Theorem 4.4. *In a platform with* S *sensors, the AoA estimator* $\hat{\boldsymbol{\phi}}$ *has a variance bounded by*

$$\mathrm{Var}\left(\hat{\boldsymbol{\phi}}\right) \geq \frac{1}{(S-1)^2}\sum_{i=1}^{S-1}\mathrm{Var}\left(\hat{\boldsymbol{\phi}}_\mathbf{i}\right) \approx \frac{1}{(S-1)^2}\sum_{i=1}^{S-1}\frac{\left(\frac{c}{x_i}\right)^2}{1-\left(\frac{c}{x_i}\tau_i\right)^2}\mathrm{Var}\left(\hat{\boldsymbol{\tau}}_\mathbf{i}\right)$$

where $\mathrm{Var}\left(\hat{\boldsymbol{\tau}}_\mathbf{i}\right)$ *is bounded in Theorem 2 and every* $\hat{\boldsymbol{\phi}}_\mathbf{i}$ *is unbiased.*

We present the following remarks for Theorem 4.4:

1) The scaling of the CRB mentioned in the remarks after Theorem 4.1 hold.

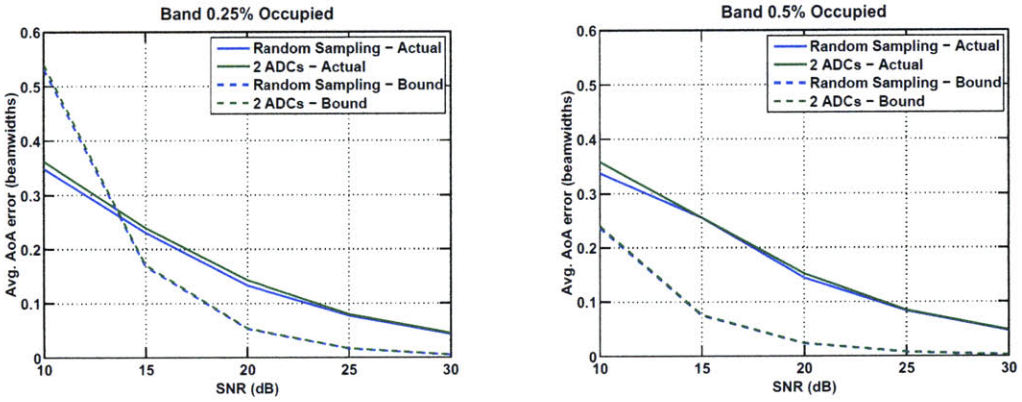
2) If we note $\hat{\boldsymbol{\tau}}_\mathbf{i} \approx \tau_i$, then $\mathrm{Var}\left(\hat{\boldsymbol{\phi}}_\mathbf{i}\right) \propto 1/(1-\cos\phi_i)^2$. This means that the estimator performs worse when ϕ is near 0 or π.

3) The approximation in Theorem 4.3 shows that the bound decreases quadratically with x_i.

4.4.3 Comparison to Experimental Results

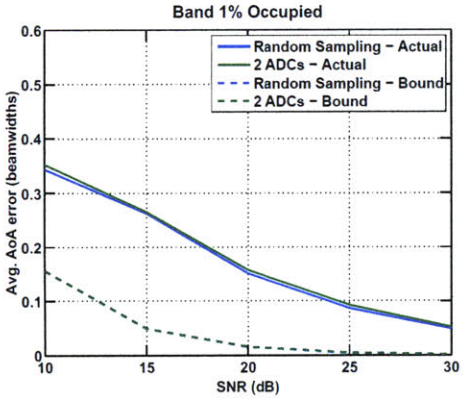
We find the CRB for the situation presented in Figure 4-5. We assume a frequency-sparse signal in a basis with 512 equally-spaced digital frequencies. The signal has a center frequency of 4GHz and bandwidth to center frequency ratio of 50%. The sensor array has four sensors on a line spaced out by $\lambda_{\min}/2$, where $\lambda_{\min}$ corresponds to the wavelength of the largest frequency component in the band of interest. We further assume the transmission is arriving from a random direction, meaning $\phi \sim \mathcal{U}(0, 2\pi)$.

Figure 4-6 presents performance lower bounds using Theorem 4.4. Unfortunately, two issues make this bound problematic. First, we found the AoA bound as a function of the TDOA bounds discussed in Section 4.3, which is known to be loose and flawed because it is too naive to capture reconstruction losses due to band density. Thus, the bound decreases with more of the band occupied while experimental results have increasing error variance. Second, we estimated $\hat{\boldsymbol{\phi}}$ as a function of the time differences between pairs of sensors on the platform, which may not be the best method to use, especially when ϕ is near 0 or π. In those regimes, the variance is very large and leads to bounds that are actually larger than the experimental results (which employs a different model).

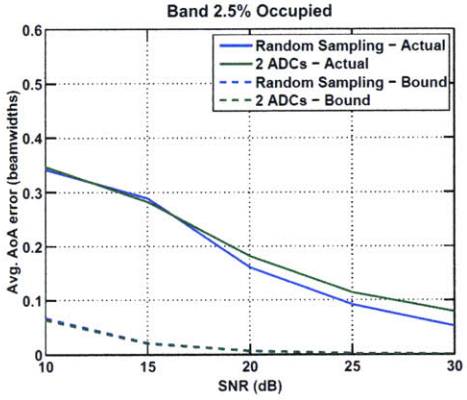


(a) 0.25%

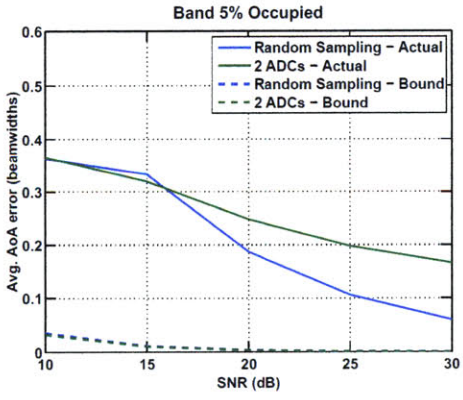
(b) 0.5%



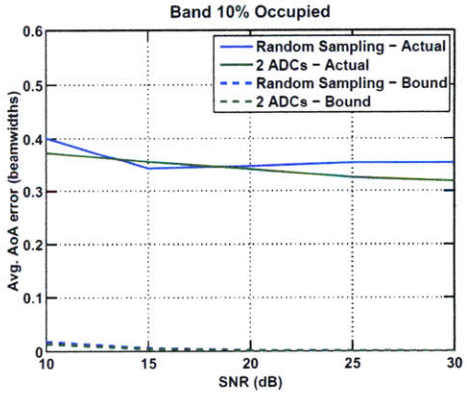
(c) 1%



(d) 2.5%



(e) 5%



(f) 10%

Figure 4-6: AoA reconstruction performance versus CRB for varying levels of sparsity.

Chapter 5

Scaling Laws for Discrete TDOA

The use of compressed sensing for geolocation of sparse transmissions in Chapter 4 can alleviate storage and communication loads while preserving estimation performance. It also exemplifies that many CS applications care about parameter estimation rather than signal recovery. In these situations, it may be possible to use fewer measurements than schemes proposed for stable signal reconstruction [2, 27] or support recovery [17, 19].

This chapter focuses on a discrete time difference of arrival (dTDOA) estimation problem, where the delay is from a countable set of possibilities. We also abstract out the practical scenario discussed in Section 4.3.1 and hence focus only on finite-dimensional vectors. In this scenario, we show measurement scaling laws are improved when the objective is perfect delay estimation rather than signal recovery. Our sufficient condition is based on a practical algorithm for partial support recovery.

5.1 Related Work

The discrete time difference estimation problem has been most notably studied as an information-theoretic formulation to TDOA. In [72], a maximum empirical mutual information (MMI) decoder is shown to have the same error exponent as a maximum-likelihood (ML) decoder for delay estimation in discrete channels. This is strongly related to work in image registration using MMI [73].

Also important for this problem is recent works on support recovery by Reeves [74, 18] and Aeron *et al.* [75]. They show partial support recovery is a much simpler problem than full support recovery and can lead to more favorable scaling laws on measurements while using a computationally intractable maximum-likelihood estimator. More practical algorithms are not discussed.

5.2 Contribution

This chapter shows improved scaling laws in the problem of dTDOA estimation compared to previously known results. In particular, we show through law of large numbers arguments that the problem model illustrated in Figure 5-1 can be approximated using a discrete memoryless channel and MMI decoder, and the error of the estimation will decay exponentially with the scaling of N. We show that this holds even for computationally cheap CS reconstruction algorithms.

Part of our proof extends previous work on partial support recovery. We complement Reeves' necessary and sufficient conditions for an ML estimator with a sufficient condition on an almost trivial thresholded correlation estimator (TCE) discussed in [19]. This result is essential to delay estimation but is also of independent interest.

A final point worth mentioning in this work is that it demonstrates the gains of processing information with a goal in mind. By knowing a priori that the computation is of parameter estimation rather than signal recovery, we can reduce the number of measurements and relax the requirements on the signal acquisition stage.

5.3 Problem Model

Let us first define the notation that $\hat{x} = S(x, n)$ corresponds to downward circular shift (or delay) of a vector x by n. Another way to describe this is in matrix form with $\hat{x} = D_n x$, where D_n is the identity matrix with rows circularly shifted down by n.

In the discrete TDOA problem shown in Figure 5-1, assume that the length-N

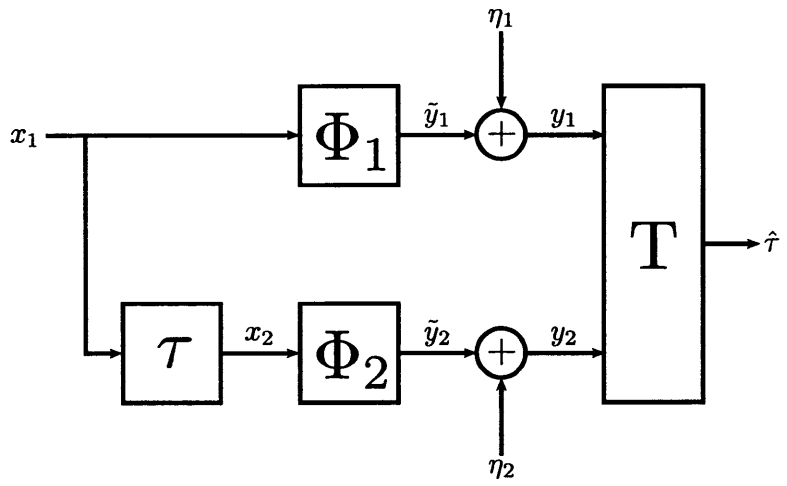


Figure 5-1: A dTDOA model where the input vectors are compressed through random projections. The delay τ is an element of a discrete set $\{0, 1, \ldots N\}$. The estimator T uses Φ_1 and Φ_2 as side information to find an estimate $\hat{\boldsymbol{\tau}}$.

input vector x_1 is K-sparse with sparsity pattern J. The vector x_2 is a shifted version of x_1 by a delay $\tau \in \{0, 1, \ldots N-1\}$, meaning $x_2 = S(x_1, \tau) = D_\tau x_1$. The two signals are then compressed separately through sampling matrices and transmitted in additive noise. The decoder T has the sampling matrices Φ_1 and Φ_2 as side information and estimates the time difference τ using the measurement vectors $\mathbf{y_i} = \Phi_i x_i + \boldsymbol{\eta_i}$.

One possible decoder, shown in Figure 5-2, performs lasso-and-threshold on each measurement matrix and produces K-sparse estimates $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$. Using the corresponding estimated sparsity vectors $\hat{\mathbf{J}}_1$ and $\hat{\mathbf{J}}_2$, a cross-correlator predicts the delay through the optimization

$$\hat{\boldsymbol{\tau}} = \arg\min_n \left| \langle S(\hat{\mathbf{J}}_1, n), \hat{\mathbf{J}}_2 \rangle \right|. \tag{5.1}$$

We exclusively consider decoders that follow this structure, producing K-sparse estimates for both signals through a block denoted T_s and running a delay estimator T_d to find $\hat{\boldsymbol{\tau}}$. The block T_s can be lasso-and-threshold, OMP or thresholded correlation, and T_d can be cross-correlation, maximum a-posteriori (MAP), or maximum mutual information (MMI).

An obvious scaling for the sample decoder is to require the number of measure-

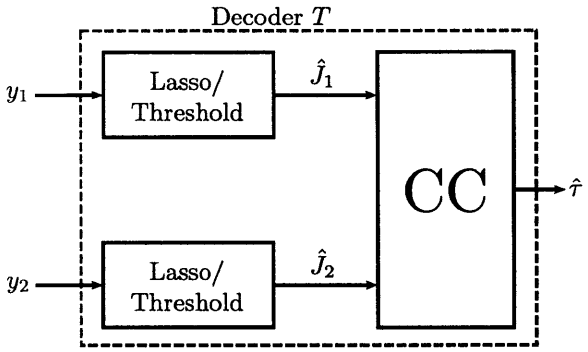


Figure 5-2: One possible decoder for dTDOA estimation. The lasso/threshold block provides K-sparse estimates with sparsity pattern vectors $\hat{\mathbf{J}}_\mathbf{i}$. The cross-correlation block then finds the delay estimate $\hat{\boldsymbol{\tau}}$. For scalings where lasso estimates the sparsity pattern correctly, the estimation error vanishes with increasing N.

ments to match Wainwright's sharp threshold of $M > 2K\log(N-K)+K+1$ for full support recovery [17]. With the support correctly estimated, the cross-correlation estimator succeeds almost surely for N large and random sparsity. In fact, the only class of sparsity pattern which causes the cross-correlation estimator to fail has nonzero entries at periodic intervals, and its probability of occurrence decreases quickly with increasing N.

However, perfect support recovery is not a necessary condition for proper estimation of $\boldsymbol{\tau}$. Indeed, the cross-correlation estimator can still succeed with a large number of errors in the sparsity pattern, provided that the errors are random and the error rate is controlled. We will explore partial support recovery in more depth in the next section, and then relate it to delay estimation in Section 5.5.

We now present a probabilistic model for the input source and sensing matrix that will be used for the analysis in this chapter. Assume the K-sparse vector $\mathbf{x_1}$ has random sparsity $\mathbf{J_1}$ chosen uniformly from all possibilities. The distribution of the nonzero entries are not necessary in the following derivations, but we need to bound the smallest nonzero entry to be at least $x_{\min}$. Let the additive noise vectors $\boldsymbol{\eta}_\mathbf{i}$ be distributed iid Gaussian with zero mean and variance σ^2. Finally, assume $\boldsymbol{\Phi}$ has iid Gaussian entries with zero mean and variance $1/M$.

It is worth mentioning that the following derivations will not apply when Ψ is not

identity and $\mathbf{x_1}$ is not sparse. This is because the delay estimator operates solely on the binary support vectors $\mathbf{J_i}$, which is required to be exactly K-sparse.

5.4 Partial Support Recovery

We consider the problem of partial support recovery, building on previous work by Reeves [18]. In these problems, assume the original signal x is K-sparse and has support J. We overload J to also represent the binary vector associated with x, such that

$$J_i = \begin{cases} 1, & x_i \neq 0; \\ 0, & x_i = 0. \end{cases} \tag{5.2}$$

As in any CS-flavored problem, the sparse input signal is compressed into a measurement y through multiplication with a sampling matrix Φ. Assume a reconstruction algorithm (e.g. lasso-and-threshold, thresholded correlation) recovers a K-sparse $\hat{x}$ with sparsity pattern $\hat{J}$. Then a natural metric for partial support recovery is

$$d(J, \hat{J}) = 1 - |J \cap \hat{J}|/K, \tag{5.3}$$

which corresponds to the percentage of indices of the true sparsity pattern that are missed.

Borrowing notation from previous work, we call a sparsity pattern estimator *asymptotically reliable* for an error rate β_{MD} if there exists a constant $c > 0$ such that the error probability satisfies

$$\mathrm{P}\left\{d(J, \hat{J}) > \beta_{\text{MD}}\right\} < e^{-nc}.$$

Hence, an asymptotically reliable estimator has controlled support error rate. Note that we do not care about the false alarm error rate because it is similarly bounded since the estimate is K-sparse.

A known result [18] says, under mild conditions on the problem model and linear sparsity ($K = \alpha N$ with α constant), a sufficient condition for *linear* partial support

recovery using maximum likelihood is $M \asymp \mathcal{O}(K)$, which is significantly better than the $M \asymp \mathcal{O}(K \log K)$ scaling needed for perfect support recovery. We show a similarly favorable scaling in the case of TCE, a much simpler decoder.

The thresholded correlation estimator works as follows. Define $\boldsymbol{\phi}_\mathbf{j}$ to be the j-th column of $\boldsymbol{\Phi}$. Let the normalized correlation vector $\boldsymbol{\rho}$ satisfy

$$\boldsymbol{\rho}_\mathbf{j} = \frac{|\boldsymbol{\phi}_\mathbf{j}' y|^2}{\|\boldsymbol{\phi}_\mathbf{j}\|^2}. \tag{5.4}$$

The result is then thresholded using a parameter μ such that only entries greater than μ are nonzero. The estimate's sparsity vector is then

$$\hat{\mathbf{J}}_\mathbf{j} = \begin{cases} 1, & \boldsymbol{\rho}_\mathbf{j} > \mu; \\ 0, & \text{o.w.} \end{cases} \tag{5.5}$$

A condition on μ is that it produces a K-sparse $\hat{J}$.

We now find the scaling of M so that TCE is asymptotically reliable up to an error rate β_{MD}. We first denote the cdf of a chi-square distribution with parameter n to be $F_{\chi^2,n}(t)$. We also define a function $F_{\text{true}}(t; M, \|x_{\setminus j_{\min}}\|^2, x_{\min})$ in (5.8) to be the pdf of $\boldsymbol{\rho}_\mathbf{j}$ for $\mathbf{J}_\mathbf{j} = 1$. Finally, the vector $x_{\setminus j_{\min}}$ is x with the smallest entry taken out. The resulting scaling of measurements is then shown in the following theorem.

Theorem 5.1. *Consider a sequence of dTDOA problems indexed by N with parameters $K = \alpha N$ and $M > M'(N)$. Assume the norm of $x \in \mathbb{R}^N$ is known and the minimum magnitude entry is bounded by $x_{\min}$. If $M'(N)$ is the solution to the transcendental equation*

$$\left(\sigma^2 + \frac{\|x\|^2}{M'(N)}\right) F^{-1}_{\chi^2,1}(1 - \beta_{\text{FA}}) = F^{-1}_{\text{true}}\left(\beta_{\text{MD}}; M'(N), \|x_{\setminus j_{\min}}\|^2, x_{\min}\right),$$

there exists a sequence of thresholds $\mu = \mu(N)$ such that the thresholded correlation estimator is asymptotically reliable for error probability less than β_{MD}.

Proof. See Appendix 5.A. □

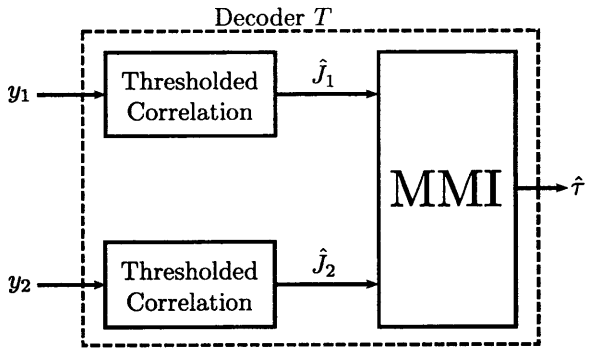


Figure 5-3: A decoder for dTDOA estimation. The thresholded correlation block has sparsity error rate bounded by β_{MD} through Theorem 5.1. This can then be approximated using a binary asymmetric channel. The delay estimator is a MMI decoder that succeeds with error vanishing exponentially.

The transcendental equation for each N is not possible to solve analytically but is shown in numerical simulations (c.f. Section 5.6) to scale $\mathcal{O}(K)$, meaning it is better than the sufficient condition for perfect support recovery in [19].

5.5 Time Difference Estimation

To relate partial support recovery to time difference estimation, we use the model shown in Figure 5-3. The signal reconstruction block T_s uses TCE to produce a K-sparse estimate $\hat{\mathbf{x}}$. Meanwhile, the delay estimation block T_d is a MMI decoder, which is defined in (5.6).

We begin by modeling the signal reconstruction block as a noisy channel. We express the estimated sparsity binary vector $\hat{\mathbf{J}}_\mathbf{i}$ as a function of the true sparsity binary vector $\mathbf{J}_\mathbf{i}$ and have shown this estimate is asymptotically reliable with measurements scaling linearly with N. By the symmetry of distributions on $\mathbf{J}_\mathbf{i}$ and Φ, each entry of $\hat{\mathbf{J}}_\mathbf{i}$ is equally likely to be in error. Hence, for N large and linear sparsity, this estimator can be approximated by a memoryless binary asymmetric channel with crossover probabilities that are *at most* β_{MD} and $\beta_{\text{FA}} = \beta_{\text{MD}}\alpha/(1-\alpha)$.

Using the noisy channel argument, we utilize a nice result from Stein *et al.* [72] on universal delay estimation. They show that, for any discrete memoryless channel, the average error of an MMI estimator will exponentially vanish with N. This MMI

estimator works regardless of the channel, making it universal.

We define the MMI estimator $T_{\text{mmi}}(\hat{J}_1, \hat{J}_2)$ as solving the minimization

$$\hat{\boldsymbol{\tau}}_{\mathbf{2}} = \arg\min_n \hat{H}\left(S(\hat{J}_1, n), \hat{J}_2\right). \tag{5.6}$$

Because the empirical entropy is the same for all circular shifts of a vector, maximizing mutual information is equivalent to minimizing joint entropy, which is shown in (5.6). As mentioned in Stein's work, the intuition behind the MMI estimator's success is that the two sparsity vectors are independent when they are not shifted to be aligned. The resulting empirical joint entropy is thus larger than when they are aligned and dependent.

We now present a sufficient condition to guarantee delay recovery almost surely using a CS encoder and the decoder from Figure 5-3.

Theorem 5.2. *Assume the decoder is the same as the one shown in Figure 5-3 and that the scaling of the number of measurements M is of that of Theorem 5.1. Then the MMI decoder succeeds in recovering the delay almost surely for large N. More specifically, the error probability decays exponentially with N.*

Proof. We use the above reasoning to model the signal estimation step as a noisy channel for the sparsity pattern, leading to the situation illustrated in Figure 5-4. This is simply a variation of the model from [72], and we use their results to show the MMI estimator is successful with error probability vanishing exponentially. □

5.6 Numerical Results

We validate the scalings for partial support recovery and demonstrate that delay estimation is successful with an MMI decoder. For simulations, we assume the noise variance is $1/M$, all nonzero entries of x have value 1, and linear sparsity with $\alpha = K/N = 0.04$. We find the scalings of M by assuming $\beta_{\text{MD}} = 0.4$ and solve the transcendental equation constrained on M being an integer. This scaling proves to be linear with K and a few sample points are shown in Table 5.1.

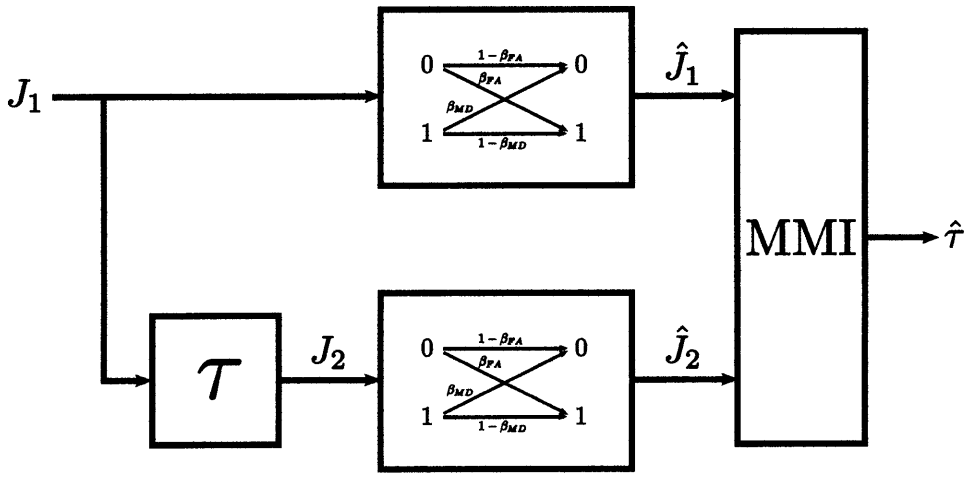


Figure 5-4: Modeling the decoder in Figure 5-3 as a noisy channel with MMI estimation. The resulting delay estimation error vanishes exponentially.

K	N	M
2	50	21
4	100	35
8	200	64
20	500	149

Table 5.1: Scaling of M necessary to ensure TCE is asymptotically reliable up to $\beta_{\text{MD}} = 0.4$. The noise variance is $1/M$ and the nonzero components of x have value 1.

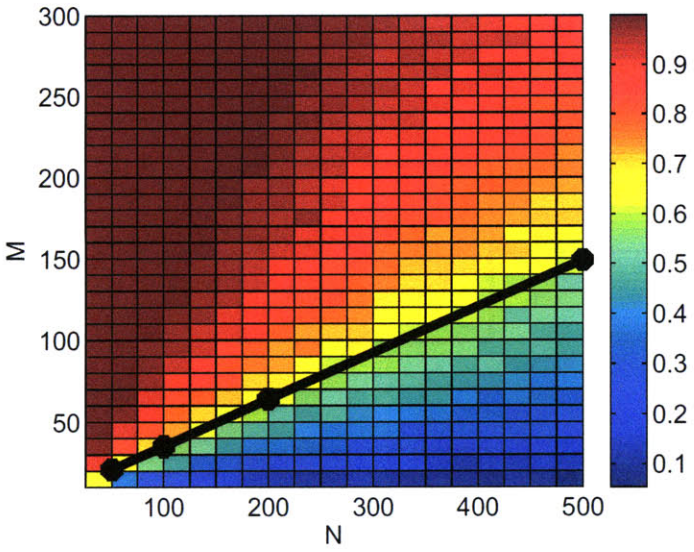


Figure 5-5: Numerical simulations for the success of support recovery using TCE. The color mapping indicates the expected percentage of the support predicted successfully for choices of M and N. The black line indicates the scalings for $\beta_{\text{MD}} = 0.4$ using Theorem 5.1.

We then compare the scalings to runs of TCE using realizations of $\mathbf{\Phi}$, $\mathbf{J}$ and $\boldsymbol{\eta}$. The percentage of the support correctly estimated is then averaged over trials and presented in Figure 5-5. We see a close match to the scalings derived from Theorem 5.1.

Finally, we implement the TCE-MMI delay estimator shown in Figure 5-3. Assuming random delay and the parameters used in the partial support recovery trials, simulations show that the probability of error decreases rapidly with N. This is illustrated in Figure 5-6.

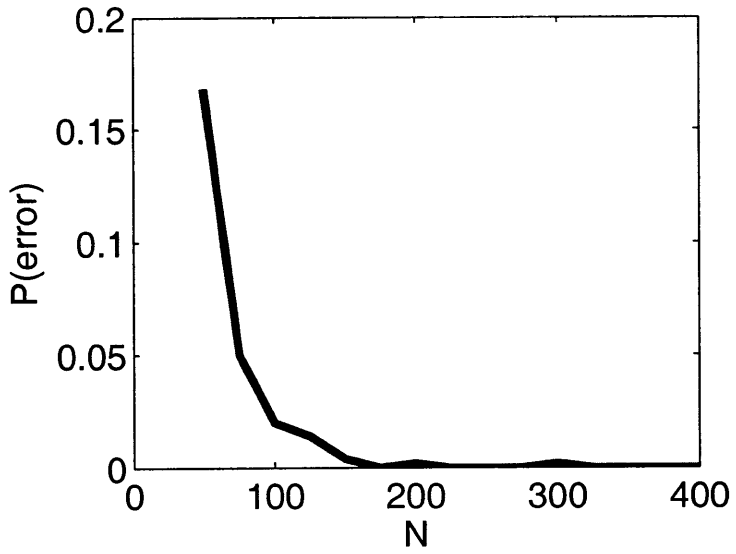


Figure 5-6: Numerical simulations of the TCE-MMI delay estimator for $\beta_{\text{MD}} = 0.4$. As N grows large, the estimator succeeds with higher probability.

5.A Proof of Theorem 5.1

We present the proof for Theorem 5.1, which finds a sufficient scaling of the number of measurements to allow the thresholded correlation estimator (TCE) to be asymptotically reliable up to an error rate β_{MD}. First, we acknowledge that this proof strongly parallels earlier work on full support recovery for TCE in the appendices of [19]. We recall that, for the sparse input x, J is defined to be sparsity binary vector. We will then define the set of nonzero indices to be $J_{\text{true}} = \{j \mid J(j) = 1\}$. We also recall that the sampling matrix is denoted $\mathbf{\Phi}$, and its j-th column is $\boldsymbol{\phi}_{\mathbf{j}}$.

Before we begin, some known results for chi-square random variables are reviewed. A chi-square with parameter n is the sum of the squares of n zero-mean unit-variance Gaussian random variables. Some simple properties of the chi-square is that it is non-negative and has mean n. We denote the pdf and cdf of the chi-square distribution to be $f_{\chi^2,n}(t)$ and $F_{\chi^2,n}(t)$ respectively. Note that if the iid Gaussian random variables $\{\mathbf{g_i}\}_{i=1}^{N}$ have variance V, then the resulting random variable $\sum \mathbf{g}_i^2$ has cdf $F_{\chi^2,n}(t/V)$. We call these *scaled chi-square* random variables.

To prove the theorem, we consider the distributions of the entries of the normalized

correlation vector $\boldsymbol{\rho}$, which is defined in (5.4). We define a vector $\mathbf{r}$ such that

$$\mathbf{r_j} = \frac{\boldsymbol{\phi_j}'\mathbf{y}}{\|\boldsymbol{\phi_j}\|},$$

meaning $\boldsymbol{\rho_j} = \mathbf{r_j}^2$. Each entry y_i is distributed $\mathcal{N}(0, \sigma^2 + \|x\|^2/M)$ conditioned on x. This is can be easily seen by noting $\mathbf{y} = \boldsymbol{\Phi}x + \boldsymbol{\eta}$ and that each entry of $\boldsymbol{\Phi}$ is iid $\mathcal{N}(0, 1/M)$.

For $j \notin J_{\text{true}}$,

$$\mathbf{r_j} = \frac{1}{\|\boldsymbol{\phi_j}\|} \sum_{i=1}^{M} (\boldsymbol{\phi_j})_i \mathbf{y_i}. \tag{5.7}$$

Conditioned on $\boldsymbol{\phi_j}$, $\mathbf{r_j}$ is distributed $\mathcal{N}(0, \sigma^2 + \|x\|^2/M)$. Since this distribution does not depend on $\boldsymbol{\phi_j}$, the unconditioned distribution of $\mathbf{r_j}$ is also $\mathcal{N}(0, \sigma^2 + \|x\|^2/M)$. For $j \neq k$, $\mathbf{r_j}$ and $\mathbf{r_k}$ are uncorrelated, which implies that they are also independent. Since $\boldsymbol{\rho_j} = \mathbf{r_j}^2$, $\boldsymbol{\rho_j}$ is a scaled chi-square random variable with parameter 1. Hence, the entries of $\boldsymbol{\rho}$ are iid $F_{\chi^2,1}(t/\sigma_r^2)$, where $\sigma_r^2 = \sigma^2 + \|x\|^2/M$.

For $j \in J_{\text{true}}$,

$$\begin{aligned}
\mathbf{r_j} &= \frac{1}{\|\boldsymbol{\phi_j}\|}\boldsymbol{\phi_j}'\mathbf{y} \\
&= \frac{1}{\|\boldsymbol{\phi_j}\|}\left[\boldsymbol{\phi_j}'\boldsymbol{\phi_j}x_j + \boldsymbol{\phi_j}'\left(\mathbf{y} - \boldsymbol{\phi_j}x_j\right)\right] \\
&= \|\boldsymbol{\phi_j}\|x_j + \frac{\boldsymbol{\phi_j}'\left(\mathbf{y_i} - \boldsymbol{\phi_j}x_j\right)}{\|\boldsymbol{\phi_j}\|} \\
&= \|\boldsymbol{\phi_j}\|x_j + \mathbf{e_j}.
\end{aligned}$$

The random variable $\|\boldsymbol{\phi_j}\|$ is the square root of a scaled chi-square random variable with cdf $F_{\chi^2,M}(Mt^2)$. Meanwhile, $\mathbf{e_j}$ will closely follow the computation above for $\mathbf{r_k}, k \notin J_{\text{true}}$ but has variance $\sigma^2 + \|x_{\setminus j}\|^2/M$, where the norm is of x with the j-th entry taken out. Similarly, $\mathbf{e_j}$ and $\mathbf{e_k}$ are uncorrelated for $j \neq k$, which implies that they are also independent. Also, $\|\boldsymbol{\phi_j}\|x_j$ and $\|\boldsymbol{\phi_k}\|x_k$ are also independent conditioned on x.

Since $\boldsymbol{\rho_j} = \mathbf{r_j}^2$, the distribution of $\boldsymbol{\rho_j}$ can easily be found. Let us define the cdf of

$\boldsymbol{\rho}_\mathbf{j}$ to be

$$F_{\text{true}}(t; M. \|x_{\setminus j}\|, x_j). \tag{5.8}$$

For $j \neq k$, $\boldsymbol{\rho}_\mathbf{j}$ and $\boldsymbol{\rho}_\mathbf{k}$ are independent since the underlying random variables are independent.

We will now use the following lemma to bound the effect of thresholding entries of $\boldsymbol{\rho}$. We define the conditional indicator function $I(\cdot)$ where the argument is a conditional statement.

Lemma 5.1. *Consider a set of iid random variables* $\{\mathbf{u_i}\}_{i=1}^N$ *with cdf* $F_\mathbf{u}(t)$. *Then, for realizations of the random variables and an error rate* β, *there exists a threshold* $\mu = F_\mathbf{u}^{-1}(1-\beta)$ *that satisfies*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} I(u_i > \mu) = \beta.$$

Similarly, for $\mu = F_\mathbf{u}^{-1}(\beta)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} I(u_i < \mu) = \beta.$$

Proof. This is a simple extension of the Gilvenko-Cantelli theorem. □

To be asymptotically reliable, we need the following two conditions to hold:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j\in J_{\text{true}}} I(\boldsymbol{\rho}_\mathbf{j} < \mu) < \beta_{\text{MD}}, \tag{5.9}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j\notin J_{\text{true}}} I(\boldsymbol{\rho}_\mathbf{j} > \mu) < \beta_{\text{FA}}. \tag{5.10}$$

In the case where $K = \alpha N$, the error rate β_{FA} is $\beta_{\text{MD}}\alpha/(1-\alpha)$. The threshold for the first condition $\mu_{\min}(x_j)$ will increase with increasing M, while the second threshold μ_{out} will decrease with increasing M. In order to be asymptotically reliable for the error rate β_{MD}, we need $\mu_{\text{out}} \leq \mu_{\min}(x_j)$ for every j. The smallest such M to achieve

this is when

$$\mu_{\text{out}} = \mu_{\text{min}}(x_{\text{min}}). \tag{5.11}$$

To find the scaling of M, we apply Lemma 5.1 to both conditions for x_{min} and set the resulting thresholds equal to each other. This leads to the transcendental equation in Theorem 5.1.

Chapter 6

Concluding Remarks

The introduction of compressed sensing has reinvigorated the study of sparsity and has persuaded many researchers and engineers to rethink what is necessary to acquire data. Specifically, less resources and processing are needed for data acquisition when framed in the domain in which the information is naturally sparse or compressible. This has far-reaching implications in a diverse range of fields and applications.

Loosely, this thesis explores the connection between compressed sensing and sensor networks. More fundamentally, it is about the limits on performance of compressed sensing systems for quantization and parameter estimation. Although particularly pertinent in compressive sensor networks, these topics have also considerable independent relevance.

In Chapter 3, we present a high-resolution approximation to the optimal quantizer for storage or transmission of measurements in a CS system. We integrate ideas from functional scalar quantization and the homotopy continuation view of lasso to find a sensitivity function $\gamma_{\text{cs}}(\cdot)$ that determines the optimal point density function $\lambda_{\text{cs}}(\cdot)$ of such a quantizer. Experimental results show that the operational distortion-rate is best when using this so called "sensitive" quantizer.

Our main finding is that proper quantization in compressed sensing is not simply a function of the distribution of random measurements (using either high-resolution approximation or practical algorithms like Lloyd-Max). Rather, optimal quantization takes into account the function to be performed on the data, and can be factored in the

quantizer design using the functional sensitivity. In the case of lasso reconstruction, the homotopy continuation method allows us to compute the sensitivity analytically or through Monte Carlo simulations.

A significant amount of work can still be done in this area. Parallel developments could be made for variable-rate quantizers. Also, this theory can be extended to other probabilistic signal and sensing models, and CS reconstruction methods that satisfy DFSQ conditions.

In the thesis, we also explore two fundamental limits for parameter estimation in CSN. In Chapter 4, we present performance bounds for time difference and angle of arrival estimation. In Chapter 5, we derive a sufficient condition in the number of measurements that will guarantee successful delay estimation in the case where the delay is from a countable set.

For TDOA and AoA performance bounds, we find the Cramér-Rao lower bound is overly optimistic and is unattainable by practical algorithms, especially in the low-SNR regime. However, we can still parse the bound to find interesting relationships between model parameters and estimator performance. Also, we show that signal reconstruction estimation through a block OMP algorithm is close to the CRB, which matches previous results that use much more computationally expensive methods.

For scaling laws, we extend previous work on partial support recovery to practically feasible reconstruction algorithms. We show that the simple thresholded correlation estimator can have controlled support error probability at much fewer measurements than needed for full support recovery. This can be bootstrapped to known results in discrete delay estimation techniques to show that the delay can be estimated correctly almost surely with increasing N.

More work can be done on both performance bounds and scaling laws for CSN. Future research into more complicated bound families such as Ziv-Zakai or Weiss-Weinstein may be useful in forming tighter bounds that will be a better indicator for the limits of TDOA and AoA estimation. Also, better delay estimators may also reduce the gap between practical algorithms and performance bounds under this model. In scaling laws, we can improve the analysis for partial support recovery by consid-

ering other types of signal reconstruction that might lead to analytical scalings. Of particular interest is how to bound support error rate for lasso reconstruction. A very recent and promising result that may answer this question is a recent investigation into the replica method [76] that breaks lasso into a set of scalar estimation problems. Other types of reconstruction may yield similarly interesting results.

As a whole, compressed sensing has progressed at an amazing rate. However, there remains the question of how practical applications such as sensor networks can best integrate CS ideas into their architectures without imposing serious restrictions. This thesis aims to address some of the fundamental limits of applying compressed sensing to systems that not only collect but also process information. A major theme of this thesis is that if one cares about some function of the data collected, then smart processing and a different cost criteria may yield improved performance and relaxed requirements on the hardware.

Bibliography

[1] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, pp. 21–30, Mar. 2008.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, Feb. 2006.

[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[4] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.

[5] D. W. Oldenburg, T. Scheuer, and S. Levy, "Recovery of the acoustic impedance from reflection seismograms," *Geophysics*, vol. 48, pp. 1318–1337, Oct. 1983.

[6] A. Papoulis and C. Chamzas, "Detection of hidden periodicities by adaptive extrapolation," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-27, pp. 492–500, Oct. 1979.

[7] D. Gabor, "Theory of communications," *Journ. IEE*, vol. 93, pp. 429–457, 1946.

[8] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, pp. 906–931, June 1989.

[9] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2845–2862, Nov. 2001.

[10] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2558–2567, Sept. 2002.

[11] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 52, pp. 6–18, Jan. 2006.

[12] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, pp. 1207–1223, Aug. 2006.

[13] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, pp. 4036–4048, Sept. 2006.

[14] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Stat.*, vol. 35, no. 6, pp. 2313–2351, 2007.

[15] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Prob.*, vol. 23, no. 3, pp. 969–985, 2007.

[16] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, pp. 4203–4215, Dec. 2005.

[17] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso)," *IEEE Trans. Inform. Theory*, vol. 55, pp. 2183–2202, May 2009.

[18] G. Reeves, "Sparse signal sampling using noisy linear projections," Tech. Rep. UCB/EECS-2008-3, Univ. of California, Berkeley, Dept. of Elec. Eng. and Comp. Sci., Jan. 2008.

[19] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inform. Theory*, vol. 55, Dec. 2009. To appear. Original submission available online [31].

[20] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1999.

[21] D. L. Donoho and J. Tanner, "Counting faces of randomly projected polytopes when the projection radically lowers dimension," *J. Am. Math. Soc.*, vol. 22, no. 1, pp. 1–53, 2009.

[22] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[23] M. Elad, "Why simple shrinkage is still relevant for redundant representation?," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5559–5569, Dec. 2006.

[24] D. M. Malioutov, M. Çetin, and A. S. Willsky, "Homotopy continuation for sparse signal representation," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, (Philadelphia, PA), pp. 733–736, Mar. 2006.

[25] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, pp. 3397–3415, Dec. 1993.

[26] J. A. Tropp, "Greed is good: Algorithmic results for sparse reconstruction," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2231–2242, Oct. 2004.

[27] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, pp. 4655–4666, Dec. 2007.

[28] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. Signal Process.*, vol. 56, pp. 2334–2345, June 2008.

[29] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces." arXiv:0806.3332v2 [cs.IT]., Mar. 2009.

[30] V. K. Goyal, A. K. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Processing Mag.*, vol. 25, pp. 48–56, Mar. 2008.

[31] A. K. Fletcher, S. Rangan, and V. K. Goyal. "Necessary and sufficient conditions on sparsity pattern recovery." arXiv:0804.1839v1 [cs.IT]., Apr. 2008.

[32] S. Aeron, M. Zhao, and V. Saligrama, "Information theoretic bounds to performance of compressed sensing and sensor networks." arXiv:0804.3439v2 [cs.IT]., Feb. 2009.

[33] D. Baron *et al*, "An information-theoretic approach to distributed compressed sensing," in *Proc. 41st Ann. Allerton Conf. on Commun., Control and Comp.*, pp. 16–21, Sept. 2008.

[34] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Processing Mag.*, vol. 25, pp. 92–101, Mar. 2008.

[35] V. Saligrama and M. Zhao, "Thresholded basis pursuit: Support recovery for sparse and approximately sparse signals." arXiv:0809.4883v2 [cs.IT]., Mar. 2009.

[36] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harm. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[37] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inform. Theory*, vol. 55, May 2009.

[38] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[39] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On-off random access channels: A compressed sensing framework." arXiv:0903.1022v2 [cs.IT]., Mar. 2009.

[40] M. Duarte *et al*, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, pp. 83–91, Mar. 2008.

[41] J. Starck and J. Bobin, "Astronomical data analysis and sparsity: From wavelets to compressed sensing." arXiv:0903.3383v1 [astro-ph.IM]., Mar. 2009.

[42] F. Koushanfar, P. T. Boufounos, and D. Shamsi, "Post-silicon timing characterization by compressed sensing," in *Proc. IEEE/ACM Computer-Aided Design*, pp. 185–189, Nov. 2008.

[43] W. K. Coulter, C. J. Hillar, and F. T. Sommer, "Adaptive compressed sensing - A new class of self-organizing coding models for neuroscience." arXiv:0906.1202v1 [q-bio.NC]., June 2009.

[44] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.

[45] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948. Continued 27:623–656, October 1948.

[46] B. M. Oliver, J. Pierce, and C. E. Shannon, "The philosophy of PCM," *Proc. IRE*, vol. 36, pp. 1324–1331, Nov. 1948.

[47] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, July 1948.

[48] R. M. Gray and A. H. G. Jr., "Asymptotically optimal quantizers," *IEEE Trans. Inform. Theory*, vol. 23, pp. 143–144, Feb. 1977.

[49] V. Misra, V. K. Goyal, and L. R. Varshney, "High-resolution functional quantization," in *Proc. IEEE Data Compression Conf.*, (Snowbird, Utah), pp. 113–122, IEEE Comp. Soc. Press, Mar. 2008.

[50] V. Misra, V. K. Goyal, and L. R. Varshney, "Distributed functional scalar quantization: High-resolution analysis and extensions." arXiv:0811.3617v1 [cs.IT]., Nov. 2008.

[51] S. M. Kay, *Fundamentals of Statistical Signal Processing.* Prentice Hall, 1993.

[52] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. V. Trees, "Extending Ziv–Zakai lower bound for vector parameter estimation," *IEEE Trans. Inform. Theory*, vol. 43, pp. 624–637, Mar. 1997.

[53] A. Renaux, P. L. P. Forster, C. D. Richmond, and A. Nehorai, "A fresh look at the Bayesian bounds of the Weiss–Weinstein family," *IEEE Trans. Inform. Theory*, vol. 56, pp. 5334–5352, Nov. 2008.

[54] E. Chaumette, J. Galy, A. Quinlin, and P. Larzabel, "A new Barankin bound approximation for the prediction of the thresholded region performance of maximum likelihood estimators," *IEEE Trans. Inform. Theory*, vol. 56, pp. 5319–5333, Nov. 2008.

[55] E. J. Candès and J. Romberg, "Encoding the ℓ_p ball from limited measurements," in *Proc. IEEE Data Compression Conf.*, (Snowbird, UT), pp. 33–42, Mar. 2006.

[56] P. T. Boufounos and R. G. Baraniuk, "Quantization of sparse representations," in *Proc. IEEE Data Compression Conf.*, (Snowbird, UT), pp. 378–378, Mar. 2007.

[57] W. Dai, H. V. Pham, and O. Milenkovic, "Quantized compressive sensing." arXiv:0901.0749v2 [cs.IT]., Mar. 2009.

[58] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *Proc. Conf. on Inform. Sci. & Sys.*, (Princeton, NJ), pp. 16–21, Mar. 2008.

[59] L. Jacques, D. K. Hammond, and M. J. Fadili, "Dequantized compressed sensing with non-Gaussian constraints." arXiv:0902.2367v2 [math.OC]., Feb. 2009.

[60] R. J. Pai, "Nonadaptive lossy encoding of sparse signals," Master's thesis, Massachusetts Inst. of Tech., Cambridge, MA, Aug. 2006.

[61] J. Z. Sun and V. K. Goyal, "Optimal quantization of random measurements in compressed sensing," in *Proc. IEEE Int. Symp. Inform. Th.*, (Seoul, Korea), June 2009.

[62] J. Z. Sun and V. K. Goyal, "Quantization for compressed sensing reconstruction," in *Samp. Th. Appl.*, (Marseille, France), IEEE Comp. Soc. Press, May 2009.

[63] K. W. Forsythe, J. I. Goodman, B. A. Miller, V. K. Goyal, J. Z. Sun, and A. Bolstad, "Compressive sensor networks." Tech. Rep. CSN-1, Lincoln Labs MIT, Mar. 2009.

[64] B. Friedlander, "On the Cramer-Rao bound for time delay and Doppler estimation," *IEEE Trans. Inform. Theory*, vol. 30, pp. 575–580, May 1984.

[65] K. Ho and Y. Chan, "Solution and performance analysis of geolocation by TDOA," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, pp. 1311–1322, Oct. 1993.

[66] K. Bell, Y. Ephraim, and H. Van Trees, "Explicit Ziv-Zakai lower bound for bearing estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, (Atlanta, GA), pp. 2852–2855, May 1996.

[67] H. Nguyen and H. Van Trees, "Comparison of performance bounds for DOA estimation," in *Proc. IEEE 7th SP Workshop Stat. Signal and Array Process.*, (Quebec, Canada), pp. 313–316, June 1994.

[68] B. Babadi, B. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramer-Rao bound for noisy compressive sampling," *IEEE Trans. Inform. Theory*, vol. 57, pp. 1233–1236, Mar. 2009.

[69] Z. Ben-Haim and Y. C. Eldar, "The Cramer-Rao bound for sparse estimation." arXiv:0905.4378v1 [math.ST]., June 2009.

[70] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1285–1301, Nov. 1990.

[71] A. Bolstad, "Compressive sensing approaches to TDOA estimation and signal reconstruction." MIT Lincoln Laboratory internal presentation, 2008.

[72] J. Stein, J. Ziv, and N. Merhav, "Universal delay estimation for discrete channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2085–2093, Nov. 1996.

[73] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a surevey," *IEEE Trans. Med. Imaging*, vol. 22, pp. 986–1004, Aug. 2003.

[74] G. Reeves and M. Gastpar, "Sampling bounds for sparse support recovery in the presence of noise," in *Proc. IEEE Int. Symp. Inform. Th.*, (Toronto, Canada), June 2008.

[75] S. Aeron, M. Zhao, and V. Saligrama, "Fundamental limits on sensing capacity for sensor networks and compressed sensing." arXiv:0804.3439v1 [cs.IT]., May 2009.

[76] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing." arXiv:0906.3234v1 [cs.IT]., June 2009.