Compressive Sensing for Gauss-Gauss Detection

J. Derek Tucker
Naval Surface Warfare Center
Panama City Division - Code X13
Panama City, FL 32428 USA
james.d.tucker@navy.mil

Nick Klausner
Department of Electrical and Computer Engineering
Colorado State University
Fort Collins, Colorado, 80523, USA
nklausne@engr.colostate.edu

***Abstract*—The recently introduced theory of compressed sensing (CS) enables the reconstruction of sparse signals from a small set of linear measurements. If properly chosen, the number of measurements can be much smaller than the number of Nyquist rate samples. However, despite the intense focus on the reconstruction of signals, many signal processing problems do not require a full reconstruction of the signal and little attention has been paid to doing inference in the CS domain. In this paper we show the performance of CS for the problem of signal detection using Gauss-Gauss detection. We investigate how the J-divergence and Fisher Discriminant are affected when used in the CS domain. In particular, we demonstrate how to perform detection given the measurements without ever reconstructing the signals themselves and provide theoretical bounds on the performance. A numerical example is provided to demonstrate the effectiveness of CS under Gauss-Gauss detection.**

***Index Terms*—binary hypothesis testing, compressive sensing, Fisher Discriminant, J-divergence, signal detection**

I. Introduction

The up and coming area of compressive sensing (CS) has become a hot topic in many signal processing applications today. The theory behind CS provides a generalization of the point-wise conventional sampling theorem where samples are theoretically interpreted as inner products of an unknown signal vector with a set of user-defined basis vectors. This framework has many applications where standard Nyquist sampling theory is feasibly impossible as sparse signals can be reconstructed from a smaller subset of linear measurements when certain criterion are met. Most of the work in signal processing revolving around this area is concerned with reliable signal recovery while not much attention has been paid to use of the compressed signals for applications such as detection, classification, and general inferences in the CS domain. Moreover, many times the theoretical foundation of CS lacks the consideration of the effects additive or multiplicative noise can have when we make compressive measurements of the environment, allowing one to question the robustness of this architecture.

Examples of applications where detection and classification performed in the CS domain may become desirable are abundant. For instance, suppose we are interested in detecting underwater objects using electro-optical sensing devices in an environment without sunlight and therefore do not require a large number of samples for inference purposes as most of the image will be dark. Or imagine the patient who must wait extended periods of time while medical imaging devices scan their body in the hope of detecting malignant tumors when really we could take a smaller number of samples thereby alleviating the process. In either case, it becomes necessary to understand the effects compressive sampling schemes will have on the performance of these automated detection and classification techniques.

The Restricted Isometry Property (RIP) from Candés and Tao [1] is the fundamental theorem in CS theory and states the fact that some classes of sampling vectors exhibit a degree of incoherence with any sparse signal and thereby preserve the distance between all sparse vectors. This theory will also play a key part in the analysis given in this paper. This important theorem is generalized in [2] to show that for any two sparse signals separated by a given angle and satisfying the conditions stated by the RIP, the angle separating their compressively sampled versions will also be bounded above and below by constants related to those given by the RIP condition. This work then goes on to consider several detection problems and it is shown that a universal CS matched filtering scheme is an effective way to detect any sparse signal both with the robustness and performance of the ideal detector. This is ultimately related back to the generalized RIP property and the fact that compressive measurements will still maintain the angles separating any set of sparse signals.

Likewise, signal detection, classification, and estimation in the CS domain are considered in [3]. Throughout the course of this work a bound is derived for the probability of detection showing it to approach one exponentially fast in terms of how many compressive measurements are made and that the exponential rate depends on the signal-to-noise ratio (SNR). Next, through the use of a simple classification problem, it is shown that the probability of error of the classifier will increase upon projection to a lower-dimensional space in a way that depends on the SNR. Finally signal estimation is considered and again bounds are derived which decay sub-linearly as a function of the number of compressive measurements that are made, illustrating the fact that estimation is in a sense a harder problem than detection and classification in the CS domain.

Through the use of looking at the application of soft margin Support Vector Machine (SVM) classification in the compressed domain, [4] ultimately shows that learning in the compressive measurement domain is possible by showing that a family of matrices satisfying the RIP condition preserves the learnability of the data set. The main result of the paper is that the accuracy of the soft margin SVM's classifier in the measurement domain is at most $O\left(\sqrt{\delta}\right)$ worse than the accuracy of the classifier in the original, high-dimensional space where the constant δ is determined from the RIP condition. Experimental results are then formulated for a synthesized data set.

In this work, we consider the standard signal detection problem [5] and develop error bounds for the J-divergence as well as the Fisher distance in the compressive domain. Both the J-divergence and Fisher Discriminant in some manner measure the distance between two hypotheses for a given detection problem and therefore give one some sense of the amount of discriminatory information present for the purposes of detection and classification. Throughout the analysis, we will consider two different scenarios, namely where compressive measurements of an unknown signal are made which is subsequently corrupted by additive noise and where compressive measurements are made of a sparse signal corrupted by additive noise. By providing bounds on the J-divergence and Fisher distance for both cases, it is the goal of this work to develop quantitative inferences of how discriminatory information is affected by performing these compressive measurements.

This paper is organized as follows. Section II will provide a brief review of compressive sensing. Section III briefly reviews linear Gauss-Gauss signal detection. Section IV introduces the detector in the CS domain and bounds on the performance of J-divergence and Fisher Discriminant for both of the cases considered. In Section V, a numerical example is given to analyze the CS for Gauss-Gauss detection. Finally, concluding remarks will then be given in Section VI.

II. Review of Compressive Sensing

We now provide a brief review of compressive sensing and its applicability to areas such as detection and classification. Suppose now that the signal we are interested in sampling or sensing is in some sense "simple" enough that we can get away with a smaller number of samples than that of conventional, point-wise sampling schemes. In other words, we assume the signal $\mathbf{x} \in \mathbb{R}^n$ to be r-sparse in some domain meaning that it is well approximated by a linear combination of r vectors from a basis of $\mathbb{R}^n$, i.e.,

$$\mathbf{x} \approx \sum_{i=1}^{r} \theta_i \boldsymbol{\phi}_i \tag{1}$$

with $r \ll n$ and $\boldsymbol{\phi} = \{\phi_1, \cdots, \phi_N\}$ a collection of basis vectors in $\mathbb{R}^n$. The natural approach to sampling would be that each column of a sampling matrix $\Psi \in \mathbb{R}^{m \times n}$ would be a vector whose entries are all zero except for the entry corresponding the desired sample location in $\mathbf{x}$.

Since it is assumed that the signal $\mathbf{x}$ is r-sparse, the linear measurement matrix Ψ whose rows are incoherent with the columns of Φ, the matrix containing the basis elements $\boldsymbol{\phi}_i$, we know from the CS theory that there exists an over-measuring factor $c > 1$ such that only $M := cK$ incoherent measurements, $\mathbf{y}$, are required to reconstruct $\mathbf{x}$ with a high degree of confidence [6], [7]. Therefore reducing the amount of computations one needs to make to effectively reconstruct the data.

One class of measurement matrices that are used in compressed sensing are matrices that satisfy the Restricted Isometry Property (RIP) which was proposed by Candés and Tao [1].

Definition 1 (Restricted Isometry Property)**.** *For each integer $r = 1, 2, \ldots,$ define the isometry constant δ of a matrix Ψ as the smallest number such that*

$$(1-\delta)||\mathbf{x}||_{l_2}^2 \leq ||\Psi\mathbf{x}||_{l_2}^2 \leq (1+\delta)||\mathbf{x}||_{l_2}^2 \tag{2}$$

holds for all r-sparse vectors $\mathbf{x}$.

This property implies that a matrix Ψ obeys the RIP of order r if δ is not too close to one and when this property holds, Ψ approximately preserves the Euclidean length of r-sparse signals.

The following theorem [8] shows that a large family of random matrices that satisfy the RIP condition.

Theorem 1. *If entries of $\sqrt{m}\Psi$ are sampled i.i.d from either*

- *Gaussian distribution: $\mathcal{N}(0,1)$, or*
- *Bernoulli distribution: $\mathcal{BE}(-1,1)$,*

and $m = \Omega(r\log(n/r))$ then except with probability $e^{-c(\delta)m}$, Ψ satisfies the restricted isometry property.

Throughout the rest of this paper we assume that Ψ is a linear measurement matrix satisfying the RIP condition and are generated using either method stated in Theorem 1.

III. Review of Standard Gauss-Gauss Signal Detection

In this section, we provide a brief review of standard Gauss-Gauss signal detection. Assume we have an observation $\mathbf{y} \in \mathbb{R}^{n\times 1}$ which is a normal random vector that is distributed with mean $\mathbf{m} = \mu\mathbf{x}$ and covariance matrix R. A classical detection problem, [5] is to test the hypothesis $H_0 : \mu = 0$ i.e., noise alone $\mathbf{y} = \mathbf{n}$, versus $H_1 : \mu = 1$ i.e., signal plus noise $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where R is the covariance matrix of the noise. It is assumed that noise and signal are uncorrelated.

The log-likelihood ratio test (LRT) that minimizes the risk involved in deciding between H_0 and H_1 leads to

$$\gamma(\mathbf{y}) \quad = \quad \begin{cases} 1 \sim H_1, & \text{when } l(\mathbf{y}) > \eta \\ 0 \sim H_0, & \text{when } l(\mathbf{y}) \leq \eta \end{cases}$$

where $l(\mathbf{y}) = \mathbf{x}^H Q \mathbf{y}$ is the log-likelihood ratio. The matrix $Q = R^{-1}$, is the inverse of the covariance matrix under both hypotheses, and can be decomposed as

$$Q = U \Lambda U^H, \tag{3}$$

where Λ is a diagonal matrix with diagonal elements λ_i, which are the eigenvalues of Q, and U is an eigenvector matrix containing the corresponding eigenvectors in its column space. We can therefore rewrite the log-likelihood as

$$l(\mathbf{y}) = \mathbf{x}^H U \Lambda U^H \mathbf{y}. \tag{4}$$

A. J-divergence

The J-divergence [9] between the two hypotheses, which is a global measure of the separability or detectability, is defined as $J = E_{H_1}[l(\mathbf{y})] - E_{H_0}[l(\mathbf{y})]$, where $E_{H_0}[\cdot]$ and $E_{H_1}[\cdot]$ are the expectation operation under the H_0 and H_1 hypothesis, respectively. Using the cyclic property of the trace we can write the J-divergence as,

$$\begin{aligned} J &= tr(\mathbf{x}^H U\Lambda U^H E_{H_1}[\mathbf{y}]) - tr(\mathbf{x}^H U\Lambda U^H E_{H_0}[\mathbf{y}]) \\ &= tr(\mathbf{x}^H U\Lambda U^H \mathbf{x}) = \mathbf{x}^H \Lambda \mathbf{x}, \end{aligned} \tag{5}$$

Thus, we only need to solve the eigenvalue problem in (3) to form the J-divergence in (5) and the log-likelihood function in (4).

B. Fisher Discriminant

The Fisher Discriminant provides an alternative method to measuring the separability of two hypotheses by measuring the squared distance between the means normalized by the sum of the variances, i.e.,

$$\begin{aligned} F &= \frac{(E_{H_1}[l(\mathbf{y})] - E_{H_0}[l(\mathbf{y})])^2}{\mathrm{Var}_{H_1}(l(\mathbf{y})) + \mathrm{Var}_{H_0}(l(\mathbf{y}))} \\ &= \frac{J^2}{\mathrm{Var}_{H_1}(l(\mathbf{y})) + \mathrm{Var}_{H_0}(l(\mathbf{y}))}, \end{aligned} \tag{6}$$

where again the subscript notation refers to the operation under its respective hypothesis. It is clear that the variance of the log-likelihood under the H_1 hypothesis is given as

$$\begin{aligned} \mathrm{Var}_{H_1}(l(\mathbf{y})) &= E_{H_1}\left[l(\mathbf{y})^2\right] - (E_{H_1}[l(\mathbf{y})])^2 \\ &= \mathbf{x}^H Q E_{H_1}\left[\mathbf{y}\mathbf{y}^H\right] Q\mathbf{x} - \left(\mathbf{x}^H Q\mathbf{x}\right)^2, \end{aligned} \tag{7}$$

while that for the H_0 hypothesis is

$$\begin{aligned} \mathrm{Var}_{H_0}(l(\mathbf{y})) &= E_{H_0}\left[l(\mathbf{y})^2\right] - (E_{H_0}[l(\mathbf{y})])^2 \\ &= \mathbf{x}^H Q E_{H_0}\left[\mathbf{y}\mathbf{y}^H\right] Q\mathbf{x}. \end{aligned} \tag{8}$$

From the statement of the problem, it is assumed that both hypotheses share the same covariance structure so that $E_{H_1}\left[\mathbf{y}\mathbf{y}^H\right] = R + \mathbf{x}\mathbf{x}^H$ and $E_{H_0}\left[\mathbf{y}\mathbf{y}^H\right] = R$. From Equations (7) and (8) it can easily be seen that $\mathrm{Var}_{H_1}(l(\mathbf{y})) = \mathrm{Var}_{H_0}(l(\mathbf{y})) = \mathbf{x}^H Q\mathbf{x}$ and thus the Fisher Discriminant posed in the framework of this simple detection problem is given as

$$F = \frac{1}{2}\mathbf{x}^H Q\mathbf{x} = \frac{1}{2}\mathbf{x}^H \Lambda\mathbf{x} = \frac{1}{2}J. \tag{9}$$

IV. Detection in Compressive Sampling Domain

We now assume that the observer is allowed to make only a limited number of observations, where each observation is the inner product between the sparse signal vector, $\mathbf{x}$ and a sampling matrix chosen *a priori*. These observations can be described in two cases. The first case where a sparse signal is sampled with additive noise, which can be viewed as,

$$\mathbf{y} = \Psi\mathbf{x} + \mathbf{n} \tag{10}$$

The second case where we make a compressed measurement of a noise signal, $\mathbf{y} \in \mathbb{R}^{m\times 1}$, i.e.,

$$\mathbf{y} = \Psi(\mathbf{x} + \mathbf{n}). \tag{11}$$

We will now explore the effects this compressive sampling scheme has on the J-divergence and Fisher Distance for both cases described above.

A. Compressed Signal Measurements

We now turn our attention to the detection problem in the CS domain for compressive signal measurements corrupted by additive noise. In this case, assume we have an observation $\mathbf{y} \in \mathbb{R}^{m\times 1}$ which is a normal random vector with mean $\mathbf{m} = \mu\Psi\mathbf{x}$ and covariance matrix R and is a compressive sensing measurement described $\mathbf{y} = \mu\Psi\mathbf{x} + \mathbf{n}$. The detection problem is then to test the hypothesis $H_0 : \mu = 0$ i.e., noise alone $\mathbf{y} = \mathbf{n}$, versus $H_1 : \mu = 1$ i.e., signal plus noise $\mathbf{y} = \Psi\mathbf{x} + \mathbf{n}$, where R is the covariance matrix of the noise. It is assumed that noise and signal are uncorrelated.

The LRT that minimizes the risk involved in deciding between H_0 and H_1 leads to

$$\gamma(\mathbf{y}) = \begin{cases} 1 \sim H_1, & \text{when } l(\mathbf{y}) > \eta \\ 0 \sim H_0, & \text{when } l(\mathbf{y}) \le \eta \end{cases}$$

where $l(\mathbf{y}) = \mathbf{x}^H \Psi^H Q\mathbf{y}$ is the log-likelihood ratio in the CS domain. The matrix Q is defined as before, $Q = R^{-1}$, and can be decomposed as

$$Q = U\Lambda U^H, \tag{12}$$

where Λ is a diagonal matrix with diagonal elements λ_i, which are the eigenvalues of Q, and U is an eigenvector matrix containing the corresponding eigenvectors in its column space. We can therefore rewrite the log-likelihood as

$$l(\mathbf{y}) = \mathbf{x}^H \Psi^H U\Lambda U^H \mathbf{y}. \tag{13}$$

1) J-divergence: The J-divergence [9] between the two hypotheses, which is a global measure of the separability or detectability of the two hypothesis again is defined as $J = E_{H_1}[l(\mathbf{y})] - E_{H_0}[l(\mathbf{y})]$, where $E_{H_0}[\cdot]$ and $E_{H_1}[\cdot]$ are the expectation operation under H_0 and H_1 hypothesis, respectively.

Using the cyclic property of the trace as before we can write the J-divergence as,

$$\begin{aligned} J &= tr(\mathbf{x}^H \Psi^H U\Lambda U^H E_{H_1}[\mathbf{y}]) - tr(\mathbf{x}^H \Psi^H U\Lambda U^H E_{H_0}[\mathbf{y}]) \\ &= \mathbf{x}^H \Psi^H \Lambda \Psi\mathbf{x}, \end{aligned} \tag{14}$$

We know that for a matrix $A \in \mathbb{R}^{m\times m}$ that is symmetric with eigenvalues $\lambda_1(A) \le \cdots \le \lambda_m(A)$ and a matrix $V \in \mathbb{R}^{m\times n}$ with $V^H V = I$ and $\lambda_1(V^H AV) \le \cdots \le \lambda_n(V^H AV)$. Then from Poincaré's Separation Theorem we know that the following inequality is true [10]

$$\lambda_i(A) \le \lambda_i(V^H AV) \le \lambda_{m-n+i}(A). \tag{15}$$

Since, we have assumed that the matrix Ψ satisfies RIP with constraints δ and r, then it follows that [1]

$$1 - \delta \le \lambda_{min}(\Psi_t^H \Psi_t) \le \lambda_{max}(\Psi_t^H \Psi_t) \le 1 + \delta \tag{16}$$

or in other words the eigenvalues of $\Psi_t^H \Psi_t$ lie between $(1-\delta)$ and $(1+\delta)$. Where t are the indices of the columns of Ψ that are picked from the r non-zero entries of $\mathbf{x}$. Using (16) with (15) we can bound the matrix $\Psi^H Q \Psi$ in (14) by

$$(1-\delta)\lambda_i(Q) \leq \lambda_i(\Psi^H Q \Psi) \leq (1+\delta)\lambda_{m-n+i}(Q) \quad (17)$$

Then using Rayleigh's Inequality [11] we can bound the J-divergence in (14) by

$$||\mathbf{x}||^2(1-\delta)\lambda_1(Q) \leq \mathbf{x}^H \Psi^H Q \Psi \mathbf{x} \leq ||\mathbf{x}||^2(1+\delta)\lambda_m(Q) \quad (18)$$

Recalling the fact that $Q = R^{-1}$, we can write the above relationship in terms of the covariance matrix R as

$$||\mathbf{x}||^2 \frac{1-\delta}{\lambda_m(R)} \leq J \leq ||\mathbf{x}||^2 \frac{1+\delta}{\lambda_1(R)} \quad (19)$$

Comparing this to the J-divergence for the standard case we see that when δ is small we have nearly the same separation between the hypothesis. Therefore the J-divergence in the CS domain for this case is affected by δ which is the price paid for taking smaller samples. Moreover, as $\delta \to 0$ we have the same performance as the standard detector.

2) Fisher Discriminant: In the same manner as the previous subsection, we wish to investigate the effects compressive sensing measurements can have on the Fisher Discriminant. From the definition of the detection problem we can see that

$$\begin{aligned} \text{Var}_{H_1}(l(\mathbf{y})) &= E_{H_1}\left[l(\mathbf{y})^2\right] - (E_{H_1}\left[l(\mathbf{y})\right])^2 \\ &= \mathbf{x}^H \Psi^H Q E_{H_1}\left[\mathbf{y}\mathbf{y}^H\right] Q \Psi \mathbf{x} \\ &\quad -(E_{H_1}\left[l(\mathbf{y})\right])^2 \end{aligned} \quad (20)$$

and

$$\begin{aligned} \text{Var}_{H_0}(l(\mathbf{y})) &= E_{H_0}\left[l(\mathbf{y})^2\right] - (E_{H_0}\left[l(\mathbf{y})\right])^2 \\ &= \mathbf{x}^H \Psi^H Q E_{H_0}\left[\mathbf{y}\mathbf{y}^H\right] Q \Psi \mathbf{x}. \end{aligned} \quad (21)$$

Again because both hypotheses share the same covariance structure, we can see that $E_{H_1}\left[\mathbf{y}\mathbf{y}^H\right] = R + \Psi \mathbf{x}\mathbf{x}^H \Psi^H$ and $E_{H_0}\left[\mathbf{y}\mathbf{y}^H\right] = R$. Therefore,

$$\text{Var}_{H_1}(l(y)) = \text{Var}_{H_0}(l(y)) = \mathbf{x}^H \Psi^H Q \Psi \mathbf{x} \quad (22)$$

and the Fisher Distance becomes

$$F = \frac{\left(\mathbf{x}^H \Psi^H Q \Psi \mathbf{x}\right)^2}{2\mathbf{x}^H \Psi^H Q \Psi \mathbf{x}} = \frac{1}{2}\mathbf{x}^H \Psi^H Q \Psi \mathbf{x}. \quad (23)$$

Clearly in this simple detection problem, the Fisher distance shares bounds with that of the J-divergence under the same assumptions, i.e.,

$$\frac{1}{2}||\mathbf{x}||^2 \frac{1-\delta}{\lambda_m(R)} \leq F \leq \frac{1}{2}||\mathbf{x}||^2 \frac{1+\delta}{\lambda_1(R)}. \quad (24)$$

Again we can see that the ability to discriminate among two hypotheses is ultimately affected by the constant δ in this compressive sensing framework.

B. Compressed Noisy Signal Measurements

We now turn our attention to the detection problem in the CS domain when compression measurements are made of a noisy sparse signal. In this case, assume we have an observation $\mathbf{y} \in \mathbb{R}^{m \times 1}$ which is a normal random vector with mean $\mathbf{m} = \mu \Psi \mathbf{x}$ and covariance matrix $\Psi R \Psi^H$ and is a compressive sensing measurement of a noisy sparse signal, i.e., $\mathbf{y} = \Psi(\mu\mathbf{x} + \mathbf{n})$. The detection problem is then to test the hypothesis $H_0 : \mu = 0$ i.e., noise alone $\mathbf{y} = \Psi\mathbf{n}$, versus $H_1 : \mu = 1$ i.e signal plus noise $\mathbf{y} = \Psi(\mathbf{x}+\mathbf{n})$, where $\Psi R \Psi^H$ is the covariance matrix of the compressed sampled noise. It is again assumed that noise and signal are uncorrelated.

The LRT that minimizes the risk involved in deciding between H_0 and H_1 leads to

$$\gamma(\mathbf{y}) = \begin{cases} 1 \sim H_1, & \text{when } l(\mathbf{y}) > \eta \\ 0 \sim H_0, & \text{when } l(\mathbf{y}) \leq \eta \end{cases}$$

where $l(\mathbf{y}) = \mathbf{x}^H \Psi^H Q \mathbf{y}$ is the log-likelihood ratio and $Q = (\Psi R \Psi^H)^{-1}$.

1) J-divergence: In this case we can write the J-divergence as

$$\begin{aligned} J &= \mathbf{x}^H \Psi^H (\Psi R \Psi^H)^{-1} \Psi \mathbf{x} \\ &= \mathbf{x}^H M \mathbf{x} \end{aligned} \quad (25)$$

where $M = \Psi^H(\Psi R \Psi^H)^{-1}\Psi$. We can take the singular value decomposition (SVD) of $\Psi = U\Sigma V^H$, where U and V are unitary matrices, i.e. $U^H U = I$ and $V^H V = I$ and the matrix Σ is a block diagonal matrix containing the singular values of Ψ. Inserting the SVD of Ψ into M we get

$$M = \Sigma^H(\Sigma V^H R V \Sigma^H)^{-1}\Sigma \quad (26)$$

Using the bounds for $\lambda(AB)$ [11] and the fact that $\Sigma^H\Sigma$ is bounded by (16) we can bound $N = (\Sigma V^H R V \Sigma^H)^{-1}$ by

$$\frac{1}{(1+\delta)\lambda_n(R)} \leq \lambda_i(N) \leq \frac{1}{(1-\delta)\lambda_1(R)} \quad (27)$$

Using (16) again we can bound M by

$$\frac{1-\delta}{(1+\delta)\lambda_n(R)} \leq \lambda_i(M) \leq \frac{1+\delta}{(1-\delta)\lambda_1(R)} \quad (28)$$

Then using Rayleigh's Inequality [11] we can bound the J-divergence in (25) by

$$||\mathbf{x}||^2 \frac{1-\delta}{(1+\delta)\lambda_n(R)} \leq \mathbf{x}^H M \mathbf{x} \leq ||\mathbf{x}||^2 \frac{1+\delta}{(1-\delta)\lambda_1(R)} \quad (29)$$

Comparing this to the J-divergence for the standard case we see that when δ is small we have nearly the same separation between the hypothesis. Therefore the J-divergence in the CS domain for this case is affected by δ which is price paid taking smaller samples.

2) Fisher Discriminant: Given the the fact that both hypotheses share the same covariance structure, the variance of the log-likelihood ratio becomes identical under both the alternative and null hypothesis and is given as

$$\begin{aligned} \mathrm{Var}_{H_1}\left(l(\mathbf{y})\right) &= \mathrm{Var}_{H_0}\left(l(\mathbf{y})\right) \\ &= \mathbf{x}^H \Psi^H \left(\Psi R \Psi^H\right)^{-1} \Psi \mathbf{x}. \end{aligned} \tag{30}$$

Therefore, the Fisher Discriminant becomes

$$F = \frac{1}{2}\mathbf{x}^H \Psi^H \left(\Psi R \Psi^H\right)^{-1} \Psi \mathbf{x} \tag{31}$$

and we can bound F similar to (29) as

$$\frac{1}{2}||\mathbf{x}||^2 \frac{1-\delta}{(1+\delta)\lambda_n(R)} \leq F \leq \frac{1}{2}||\mathbf{x}||^2 \frac{1+\delta}{(1-\delta)\lambda_1(R)}. \tag{32}$$

For both cases we see that the ability to discriminate between two hypotheses is directly related to the constant δ. We also see that inducing compressive measurements of noise-corrupted signals generally leads to a loosening of the bounds as it is obvious that for any $\delta \in (0,1)$

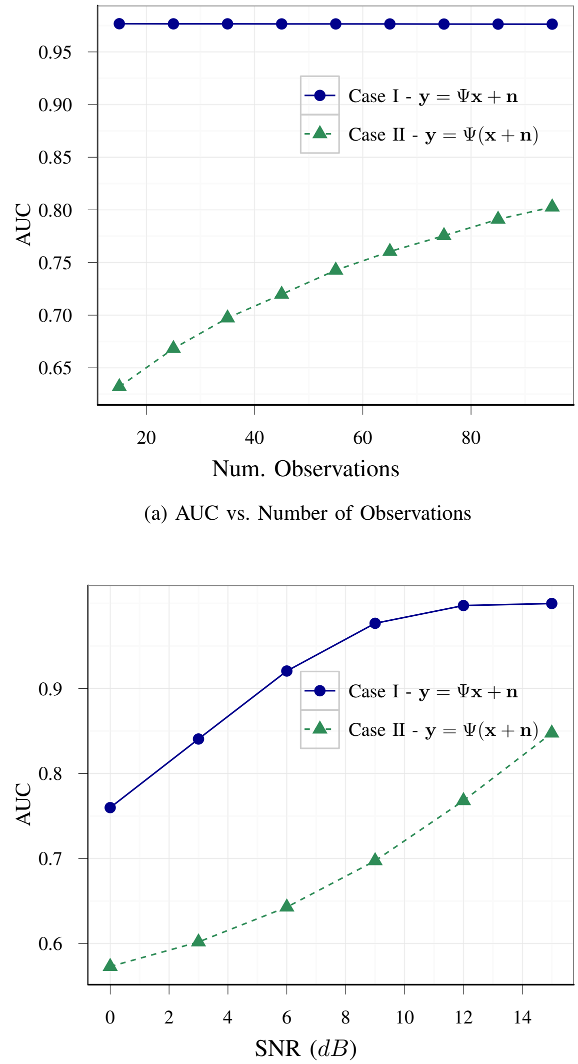
$$\frac{1-\delta}{1+\delta} < 1-\delta$$
$$\frac{1+\delta}{1-\delta} > 1+\delta.$$

Again we can make the same observation that as $\delta \to 0$ we have the same performance as the standard detector. It also needs to be mentioned here that for each of these detection methods in the compressive sensing framework is not necessarily optimal as once could design a compressed detector for the sparse signal that would be optimal. However, with the rise in interest in the CS Theory one needs to understand the performance of the detector in this domain. One example of using these detection methods in the CS-domain is if we want to know if the signal is present before we reconstruct the signal. Therefore by knowing the performance bounds of J-divergence and Fisher Discriminant one will know what to expect if one needs to perform detection in this domain.

V. Numerical Simulation

An experiment is conducted to show the effectiveness of the Gauss-Gauss detector under compressive sensing measurements for case I and case II. Where case I is compressed signal measurements ($\mathbf{y} = \Psi\mathbf{x} + \mathbf{n}$) and case II is compressed noisy signal measurements ($\mathbf{y} = \Psi(\mathbf{x} + \mathbf{n})$). For both cases under H_0, $\mathbf{y} = \mathbf{n}$ is a zero-mean white Gaussian vector process with covariance matrix R_0. Under H_1, $\mathbf{x}$ is a signal of length 512 with a sparsity condition of 10, i.e., there are only 10 non-zero entries in $\mathbf{x}$. The sensing matrix (Ψ) was generated using a uniform spherical ensemble where the columns are uniformly distributed on the sphere $\mathbb{S}^{n-1}$.

The sample data matrices are formed for the two cases using number of samples ranging from 15 to 100 samples in 10 sample increments. The matrices are also formed under varying signal-to-noise ratio (SNR) conditions ranging from $0dB$ to $16dB$ in $3dB$ increments for each sample size. With



(a) AUC vs. Number of Observations

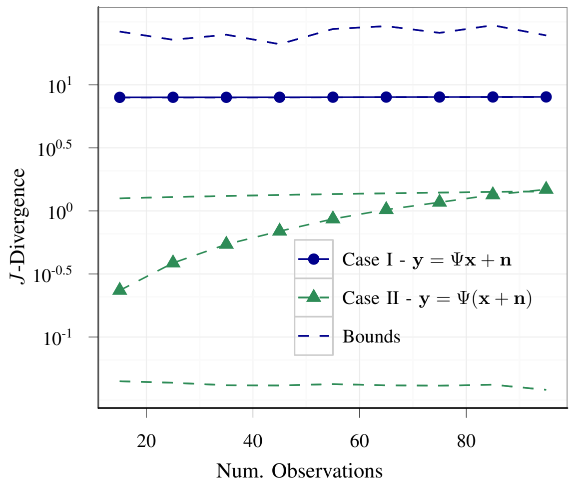
(b) AUC vs. SNR

Fig. 1: Area under the curve performance for case I and II.

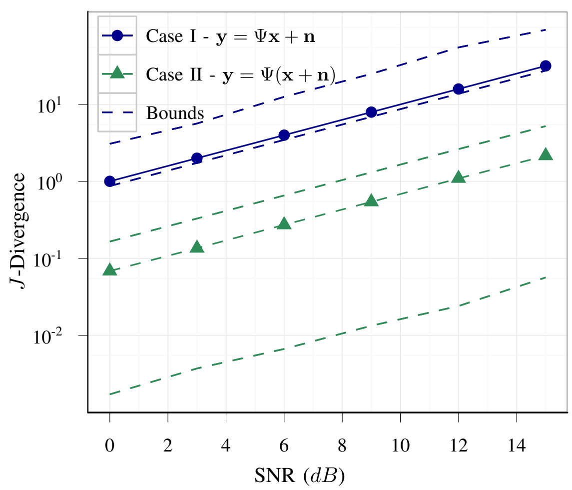
this data we could study the performance of Gauss-Gauss detection under CS for varying sample sizes and varying SNR.

The eigenvalues for the matrix Q for both cases were estimated and then the J-divergence and log-likelihood ratio were computed. For each case, the J-divergence and log-likelihood functions are formed and a separate test set of 100 samples is then applied to the detector. The experiment was repeated for 1000 Monte Carlo trials.

The plots of the area under the receiver operating characteristic (ROC) curve (AUC) versus the number of observations and versus SNR are presented in Figure 1a and 1b, respectively for cases I and II. These plots represent the averaged results for 1000 Monte Carlo trials. Figure 1a is calculated at a fixed SNR of 9 dB and we see that the performance is constant for case I which is to be expected from the CS theory. For case II we see an increase in performance as the number of samples increase, however the performance is relatively similar over



(a) J-Divergence vs. Number of Observations



(b) J-Divergence vs. SNR

Fig. 2: J-divergence performance for case I and II.

the increase in the number of observations. Figure 1b shows an increase in AUC as the SNR increases for both cases which is to be expected with case I having a sharper increase.

The plots of the J-divergence versus the number of observations and versus SNR are presented in Figure 2a and 2b, respectively for cases I and II. These plots represent the averaged results for 1000 Monte Carlo trials. Figure 2a is calculated at a fixed SNR of 9 dB which is the same as the AUC calculation and we see that the performance is constant for case I which is to be expected from the results presented in Figure 2a. The bounds that were defined in (19) are shown in the figure by the dashed lines and show a good fit around the empirical calculation. For case II we see a slight increase in performance as the number of samples increase and the bounds that were defined in (29) are shown by the dashed lines and show a good fit around the empirical calculations. Figure 2b shows an increase in J-divergence as SNR increases for both cases with case I having a much sharper increase. The calculated bounds for case II are wider than case I as expected.

VI. Conclusion

In this paper, we began by considering a simple, linear detection problem and derived expressions for the J-divergence and Fisher Discriminant, both of which give one a sense of the amount of available discriminatory information for the purposes of detection and classification. Then using linear measurement matrices commonly used in CS, we considered a case of this detection problem where we make compressive measurements of the signal which is then corrupted by additive noise and derived bounds for both measures. We next concerned ourselves with the case where compressive measurements are taken of a noise corrupted version of the signal and again bounds were derived. Overall, we have shown that for both cases the derived bounds depend on the energy present in the signal $\left(||\mathbf{x}||^2\right)$, the largest and smallest eigenvalues of the covariance matrix $(\lambda_{max}(R), \lambda_{min}(R))$, and the constant δ described by the RIP condition. Ultimately however, the bound on the amount of discriminatory information projected into the CS domain becomes solely determined by the constant δ. We have also shown that inducing compressive measurements on noise corrupted signals tends to loosen the error bounds on these measures thereby decreasing the certainty of the discriminatory information available in this case. Again, given a reasonable value of δ, the amount of discriminatory information available for detection and classification closely resembles that in the original, high-dimensional original data domain. Numerical simulations were also provided confirming the developed bounds.

Acknowledgment

The authors would like to thank Dr. Ali Pezeshki at Colorado State University for his guidance during this work. Approved for public release; distribution is unlimited.

References

[1] E. J. Candés and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory.*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[2] J. Haupt and R. Nowak, "Compressive sampling for signal detection," *Proc. of ICASSP 2007*, vol. 3, pp. 1509–1512, April 2007.
[3] M. Davenport and M. W. R. Baraniuk, "Detection and estimation with compressive measurements," *Rice ECE Department Technical Report*, vol. TREE 0610, November 2006.
[4] R. Calderbank, S. Jafarpour, and R. Schapire, "Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain," *Preprint*, 2009.
[5] H. L. Van Trees, *Detection, Estimation, and Modulation Theory Part I.* John Wiley and Sons, 1968.
[6] E. J. Candés, J. K. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, p. 2006, 2006.
[7] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr 2006.
[8] E. J. Candés and T. Tao, "Near optimal signal recovery from random projections: universal encoding stragies," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
[9] S. Kullaback, *Information Theory and Statistics*. New York: Dover, 1968.
[10] H. Lutkepohl, *Handbook of Matrices*. New York : Wiley, 1996.
[11] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. Academic Press: New York, 1979.