

A Generalized Restricted Isometry Property

Jarvis Haupt and Robert Nowak

Department of Electrical and Computer Engineering, University of Wisconsin – Madison

University of Wisconsin Technical Report ECE-07-1

May 2007, Revised December 2007

Abstract

Compressive Sampling (CS) describes a method for reconstructing high-dimensional sparse signals from a small number of linear measurements. Fundamental to the success of CS is the existence of special measurement matrices which satisfy the so-called Restricted Isometry Property (RIP). In essence, a matrix satisfying RIP is such that the lengths of all sufficiently sparse vectors are approximately preserved under transformation by the matrix. In this paper we describe a natural consequence of this property – if a matrix satisfies RIP, then acute angles between sparse vectors are also approximately preserved. We formulate this property as a Generalized Restricted Isometry Property (GRIP) and describe one application in robust signal detection.

Index Terms

Restricted isometry property, compressive sampling

I. INTRODUCTION

The problem of extracting essential information from a large data set is a fundamental goal in many applications. For example, natural images are often highly compressible, in the sense that a few discrete cosine transform (DCT) coefficients or wavelet coefficients are sufficient to capture most of the salient information in the signal. Indeed, the application of these ideas leads to effective image coding standards, such as JPEG and JPEG2000.

Despite the efficiency of representation that these methods provide, a fundamental limitation is that *all* of the data (for example, pixels, in the case of images) must be acquired before the compression can be performed, after which time the data are described by a relatively small number of transform coefficients. A natural question arises: for general signals is it possible to *streamline* the sampling process by directly measuring only the relevant coefficients? This question was posed in [1], and answered in various contexts in a variety of works (see, for example, [1]–[5]), giving rise to the theory of Compressive Sampling.

Compressive Sampling (CS), also known as Compressed Sensing, is a generalization of conventional point-sampling where observations are inner products between an unknown signal and a set of user-defined test vectors. Recent theoretical results show that, for certain ensembles of test vectors, CS projections provide an effective method of encoding the salient information in *any* sparse (or nearly sparse) signal. Further, these projection samples can be

E-mails: jdhaupt@wisc.edu, nowak@engr.wisc.edu. This research was supported in part by the DARPA Analog-to-Information Program.

used to obtain a consistent estimate of the unknown signal even in the presence of noise [1]–[5]. These results are remarkable because the number of samples required for low-distortion reconstruction is on the order of the number of relevant signal coefficients, which is often far fewer than the ambient dimension in which the signal is observed. This huge reduction in sampling makes CS a practical and viable option in many resource constrained applications.

One popular and computationally feasible reconstruction approach is ℓ_1 -constrained minimization, proposed and analyzed in [1]–[3]. Consider a vector of observations $\mathbf{y} \in \mathbb{R}^k$ described by $\mathbf{y} = \mathbf{A}\mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^n$ is the original signal and \mathbf{A} is a $k \times n$ observation matrix where $k \ll n$. Under certain conditions on the matrix \mathbf{A} and the class of “recoverable” signals \mathbf{x} (described below), the original signal can be reconstructed *exactly* as the unique solution of

$$\arg \min_{\mathbf{v}} \|\mathbf{v}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{v}, \quad (1)$$

where $\|\mathbf{v}\|_{\ell_1} \triangleq \sum_{i=1}^n |v(i)|$ describes the vector ℓ_1 -norm. In other words, of all of the length- n vectors that agree with the k -dimensional vector of observations, the actual signal is the one having minimum ℓ_1 -norm.

Before describing the conditions under which CS succeeds, it will be helpful to fix notation and provide several definitions that will help clarify the presentation in the sequel. For a vector $\mathbf{v} = [v(1) \ v(2) \ \dots \ v(n)]^T \in \mathbb{R}^n$, we let $\|\mathbf{v}\|^2 \triangleq \sum_{i=1}^n v^2(i)$ denote the (squared) Euclidean ℓ_2 -norm. For $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, we define the inner product by $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle \triangleq \sum_{i=1}^n v_1(i)v_2(i)$. The span of two vectors \mathbf{v}_1 and \mathbf{v}_2 is defined as $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \triangleq \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} = a\mathbf{v}_1 + b\mathbf{v}_2, \ a, b \in \mathbb{R}\}$. Geometrically, the span of two non-parallel (non-zero) vectors in \mathbb{R}^n ($n \geq 2$) is a plane, and we may rewrite the inner product in terms of the planar angle α separating the vectors as $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \|\mathbf{v}_1\| \cdot \|\mathbf{v}_2\| \cos(\alpha)$. We will use the notation $\hat{\mathbf{v}}$ for the unit vector in the direction of \mathbf{v} , given by $\mathbf{v}/\|\mathbf{v}\|$. In what follows we will at times be working with the subclass of sparse vectors. Formally, we let $T \subset \{1, 2, \dots, n\}$ and define $\mathcal{V}_T = \{\mathbf{v} \in \mathbb{R}^n : v(i) = 0 \ \forall i \in T^c\}$ to be the set of all vectors having at most $|T|$ nonzero entries, where the index of any nonzero entry must be contained in the set T . We say that such vectors are *supported on T* , and that any vector $\mathbf{v} \in \mathcal{V}_T$ is $|T|$ -sparse.

We are now in a position to quantify the conditions under which the ℓ_1 -minimization approach in (1) yields the correct reconstruction. The main enabling property is known as the Restricted Isometry Property (RIP) [3]. Matrices \mathbf{A} satisfy RIP if

$$(1 - \epsilon_m)\|\mathbf{x}\|^2 \leq \|\mathbf{A}\mathbf{x}\|^2 \leq (1 + \epsilon_m)\|\mathbf{x}\|^2 \quad (2)$$

holds simultaneously for all m -sparse vectors \mathbf{x} , for some $\epsilon_m \in (0, 1)$.¹

A typical CS result that leverages RIP is the following. Consider the problem of recovering an unknown m -sparse vector $\mathbf{x} \in \mathbb{R}^n$ using the k -dimensional observation vector $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $m < k \ll n$. When the $k \times n$ sampling matrix \mathbf{A} satisfies RIP conditions such that $\epsilon_m + \epsilon_{2m} + \epsilon_{3m} < 1$, the solution to (1) is unique and equal to \mathbf{x} [6].

¹In some settings the entries of \mathbf{A} are scaled to meet energy requirements (*i.e.*, unit-norm rows), resulting in a uniform attenuation by the constant factor $C = C(k, n)$. In those cases, the RIP statement becomes

$$(1 - \epsilon_m)C\|\mathbf{x}\|^2 \leq \|\mathbf{A}\mathbf{x}\|^2 \leq (1 + \epsilon_m)C\|\mathbf{x}\|^2. \quad (3)$$

To understand the condition $\epsilon_m + \epsilon_{2m} + \epsilon_{3m} < 1$, consider the following scenarios. Suppose that for some observation matrix \mathbf{A} there is a nonzero m -sparse signal \mathbf{x} such that the observations $\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{0}$. One could not possibly hope to recover \mathbf{x} in this setting, since the observations do not provide any information about the signal. The condition $\epsilon_m < 1$ prevents this ambiguity. Another similar problem might arise if two distinct m -sparse signals, say \mathbf{x} and \mathbf{x}' , (which may not be supported on the same set) are mapped to the same compressed data (i.e., $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}'$). The condition $\epsilon_{2m} < 1$ is sufficient to prevent this. Overall, the condition $\epsilon_m + \epsilon_{2m} + \epsilon_{3m} < 1$ is only slightly stronger than the conditions described here, and can be thought of as preventing such ambiguities.

Prior works have verified that RIP holds with high probability for sampling matrices \mathbf{A} whose entries are independent and identically distributed (i.i.d.) realizations of certain random variables provided the number of rows of the matrix is large enough [1]–[3], [7]. For example, let $\epsilon_m \in (0, 1)$ be fixed. Then, a matrix whose entries are generated as i.i.d. realizations of $\mathcal{N}(0, 1/k)$ random variables satisfies

$$\text{Prob}(RIP) \geq 1 - e^{-c_0 k + m \log(en/m) + m \log(12/\epsilon_m) + \log(2)} \quad (4)$$

where $c_0 = c_0(\epsilon_m) = \epsilon_m^2/16 - \epsilon_m^3/48$ [7]. This implies, for example, that if $c_0 k \geq m \log(en/m) + m \log(12/\epsilon_m) + \log(2)$, then the probability of RIP not being satisfied is less than 1, and can be made arbitrarily small by increasing the number of rows in the sampling matrix, k .

II. MAIN RESULTS

In addition to the preservation of ℓ_2 -norm, matrices that satisfy RIP (with sufficiently small ϵ) also approximately preserve angles between sparse vectors. This result is stated here as a theorem.

Theorem 1 *Suppose a matrix \mathbf{A} satisfies RIP for m -sparse vectors with $\epsilon_m \in [0, 1/3]$. Then, for any vectors having sparsity at most m , supported on the same set of indices and separated by an angle $\alpha \in [0, \pi/2]$, the angle α_p between the projected vectors obeys the bound*

$$(1 - \sqrt{3\epsilon_m})\alpha \leq \alpha_p \leq (1 + 3\epsilon_m)\alpha. \quad (5)$$

Remark 1 *A slight weakening of the angular upper bound yields the symmetric statement*

$$(1 - \sqrt{3\epsilon})\alpha \leq \alpha_p \leq (1 + \sqrt{3\epsilon})\alpha, \quad (6)$$

which follows from the fact that $x \leq \sqrt{x}$ for $x \in [0, 1]$. All corollaries can be obtained using this weaker bound as well.

Taking this result in conjunction with the RIP condition yields the *Generalized Restricted Isometry Property* (GRIP).

Corollary 1 (GRIP) *Under the conditions of Theorem 1 the projected inner product between two sparse vectors \mathbf{x} and \mathbf{y} , supported on the same set and separated by an acute angle α , satisfies*

$$(1 - \epsilon)\|\mathbf{x}\| \cdot \|\mathbf{y}\| \cos[(1 + 3\epsilon_m)\alpha] \leq \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{y} \rangle \leq (1 + \epsilon)\|\mathbf{x}\| \cdot \|\mathbf{y}\| \cos[(1 - \sqrt{3\epsilon_m})\alpha]. \quad (7)$$

This is a simple consequence of the fact that the function $\cos(x)$ function is monotone decreasing when $x \in [0, \pi/2]$. Applying the definition of the inner product (and using the length preservation assumptions) yields the result.

The main results are a direct consequence of the following lemma, which first appeared in [8].

Lemma 1 Consider two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ($n \geq 2$) separated by an angle $\alpha \in [0, \pi/2]$. In addition, consider three additional “supporting” vectors $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ chosen so that \mathbf{s}_1 completes the triangle formed by \mathbf{x} and its projection onto \mathbf{y} , \mathbf{s}_2 is the hypotenuse of the right-angle isosceles triangle formed by \mathbf{s}_1 and $\|\mathbf{s}_1\|\hat{\mathbf{y}}$, and \mathbf{s}_3 is the third side of an isosceles triangle having sides \mathbf{x} and $\|\mathbf{x}\|\hat{\mathbf{y}}$. A typical such construct is depicted in Fig. 1. Suppose that under projection by the matrix \mathbf{A} , for each vector $\mathbf{v} \in \{\mathbf{x}, \mathbf{y}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\}$

$$(1 - \epsilon)C\|\mathbf{v}\|^2 \leq \|\mathbf{A}\mathbf{v}\|^2 \leq (1 + \epsilon)C\|\mathbf{v}\|^2, \quad (8)$$

for some $C > 0$ that, in general, will depend on the dimensions of \mathbf{A} . Then, whenever $\epsilon \in [0, 1/3]$, the angle α_p between the projected vectors $\mathbf{A}\mathbf{x}$ and $\mathbf{A}\mathbf{y}$ satisfies

$$(1 - \sqrt{3\epsilon})\alpha \leq \alpha_p \leq (1 + 3\epsilon)\alpha. \quad (9)$$

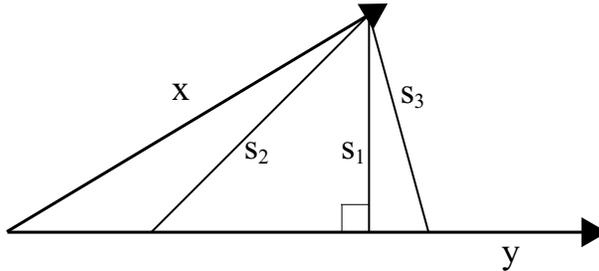


Fig. 1. Two vectors separated by an acute angle, shown with three additional “supporting” vectors.

The innovation in this paper is that the supporting vectors are themselves sparse, so RIP actually implies the conditions of Lemma 1. This idea, while obvious, is formalized here as a lemma.

Lemma 2 All vectors in the span of two non-parallel sparse vectors supported on a set T are themselves sparse and supported on the same set.

Proof: We prove the lemma by contradiction. Let \mathbf{v}_1 and \mathbf{v}_2 be any two non-parallel sparse vectors supported on a set T . If the claim is false, then there exists some vector $\mathbf{v}^* \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ having at least one nonzero entry at an index not contained in T . In other words, for some $i \in T^c$, $v(i) \neq 0$. But, by linearity of vector addition, this implies that either $v_1(i) \neq 0$ or $v_2(i) \neq 0$, or both. This contradicts the assumption that \mathbf{v}_1 and \mathbf{v}_2 are both supported on T , and the claim is established. ■

Several prior works consider the approximation of an inner product between two high-dimensional vectors by its low-dimensional projected surrogate. In the context of kernel methods, the bounds

$$\langle \mathbf{x}, \mathbf{y} \rangle - \epsilon (\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2) \leq \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{y} \rangle \leq \langle \mathbf{x}, \mathbf{y} \rangle + \epsilon (\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2) \quad (10)$$

were shown to hold for any two arbitrary vectors \mathbf{x} and \mathbf{y} with high probability when the elements of \mathbf{A} are i.i.d. realizations of $\pm 1/\sqrt{k}$ random variables, and the number of observations k is “large enough” [9]. Similar bounds were derived later in [10]. In that work, it is shown that

$$\langle \mathbf{x}, \mathbf{y} \rangle - 2\epsilon\|\mathbf{x}\|^2\|\mathbf{y}\|^2 \leq \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{y} \rangle \leq \langle \mathbf{x}, \mathbf{y} \rangle + 2\epsilon\|\mathbf{x}\|^2\|\mathbf{y}\|^2. \quad (11)$$

While either of these results could be used to obtain essentially the same results in the example in Sec. IV, it is interesting to note that these existing bounds give less control over the projected angle. Fig. 2 shows the upper and lower bounds of (5) on the projected angle, as well as the bounds implied by (11) for several values of ϵ .

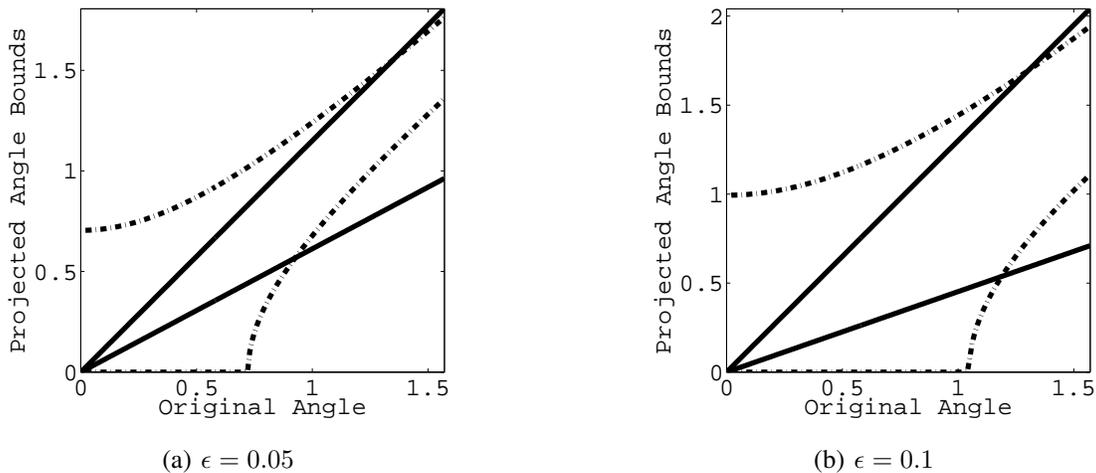


Fig. 2. Comparison of the angular bounds in (5), shown with solid lines, and (11), shown with dashed lines for varying values of ϵ . The bounds given by (5) do not depend on the lengths of the original vectors, but the bounds in (11) do – for this comparison, we let $\|\mathbf{x}\| = 1$ and $\|\mathbf{y}\| = 2$.

The remainder of this report is organized as follows. Using the essential ideas from [8], a proof of Lemma 1 is given in Section III. In Section IV, a brief introduction to signal detection preliminaries is provided, and an application of GRIP in the context of robust signal detection is described. Detailed derivations used in the proofs are established as lemmata in the Appendix.

III. PROOFS OF MAIN RESULTS

A. Main Result

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ($n \geq 2$) denote two vectors separated by an angle $\alpha \in [0, \pi/2]$. In addition, three “supporting” vectors $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ are chosen so that \mathbf{s}_1 completes the triangle formed by \mathbf{x} and its projection onto \mathbf{y} , \mathbf{s}_2 is the

hypotenuse of the right-angle isosceles triangle formed by s_1 and $\|s_1\|\hat{y}$, and s_3 is the third side of an isosceles triangle having sides x and $\|x\|\hat{y}$. (See Fig. 1) Suppose that under projection by the matrix \mathbf{A} , for each vector $v \in \{x, y, s_1, s_2, s_3\}$

$$(1 - \epsilon)C\|v\|^2 \leq \|\mathbf{A}v\|^2 \leq (1 + \epsilon)C\|v\|^2, \quad (12)$$

for some $C > 0$ and $\epsilon \leq 1/3$.

The goal is to establish the angular bounds

$$(1 - \sqrt{3\epsilon})\alpha \leq \alpha_p \leq (1 + 3\epsilon)\alpha. \quad (13)$$

We first establish the angular lower bound, making use of the following lemma.

Lemma 3 Consider a right angle isosceles triangle with sides of length h and hypotenuse of length $\sqrt{2}h$ (denoted by the vectors s_1, s'_1 , and s_2 , respectively, in Fig. 3). Let \mathbf{A} be a linear transformation such that for each vector $s \in \{s_1, s'_1, s_2\}$,

$$(1 - \epsilon)C\|s\|^2 \leq \|\mathbf{A}s\|^2 \leq (1 + \epsilon)C\|s\|^2 \quad (14)$$

holds simultaneously for some $\epsilon \in [0, 1/3]$. Then the images of the perpendicular vectors s_1 and s'_1 satisfy

$$|\langle \mathbf{A}s_1, \mathbf{A}s'_1 \rangle| \leq 2\epsilon Ch^2, \quad (15)$$

and the height h_p of the projected triangle satisfies

$$\left(1 - \epsilon - \frac{4\epsilon^2}{1 - \epsilon}\right)Ch^2 \leq h_p^2 \leq (1 + \epsilon)Ch^2. \quad (16)$$

The first statement says that perpendicular vectors remain nearly perpendicular under transformation by \mathbf{A} . This result was shown independently (in various contexts) in [6], [8], [9]. The second statement says that the height of the projected triangle is approximately preserved as well, and will provide the key to establishing the angular lower bound. The proof presented here is essentially due to Magen.

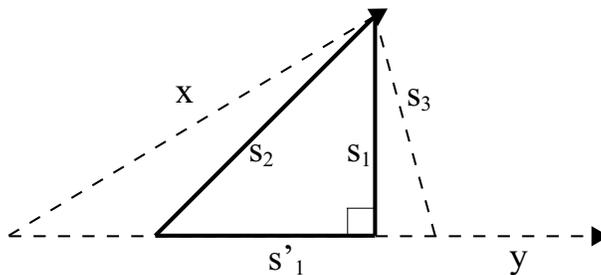


Fig. 3. The right angle isosceles triangle described in Lemma 3.

Proof: The first statement of the lemma makes use of the Law of Cosines, which states that

$$\|\mathbf{A}s_2\|^2 = \|\mathbf{A}s_1\|^2 + \|\mathbf{A}s'_1\|^2 - 2\langle \mathbf{A}s_1, \mathbf{A}s'_1 \rangle \quad (17)$$

implying

$$|\langle \mathbf{A}\mathbf{s}_1, \mathbf{A}\mathbf{s}'_1 \rangle| \leq \max \left\{ \frac{\|\mathbf{A}\mathbf{s}_1\|^2 + \|\mathbf{A}\mathbf{s}'_1\|^2 - \|\mathbf{A}\mathbf{s}_2\|^2}{2} \right\}, \quad (18)$$

where the maximum is taken over the restrictions on the lengths of the projected sides, as given in (14). Because the objective to be maximized is monotone in each term, the maximum must occur when each term takes one of its two possible extreme values. Further, the objective is symmetric in the terms $\|\mathbf{A}\mathbf{s}_1\|$ and $\|\mathbf{A}\mathbf{s}_2\|$, so it suffices to set those terms to be equal and test only four combinations. The maximum is found to occur when $\|\mathbf{A}\mathbf{s}_1\|^2 = \|\mathbf{A}\mathbf{s}_2\|^2 = (1 - \epsilon)Ch^2$ and $\|\mathbf{A}\mathbf{s}_3\|^2 = (1 + \epsilon)2Ch^2$, yielding the stated bound.

To establish the second claim, we first let α_p denote the angle between the projected vectors. Proceeding as above, we find

$$|\cos \alpha_p| = \left| \frac{\langle \mathbf{A}\mathbf{s}_1, \mathbf{A}\mathbf{s}'_1 \rangle}{2\|\mathbf{A}\mathbf{s}_1\| \cdot \|\mathbf{A}\mathbf{s}'_1\|} \right| \leq \max \left\{ \frac{\|\mathbf{A}\mathbf{s}_1\|^2 + \|\mathbf{A}\mathbf{s}'_1\|^2 - \|\mathbf{A}\mathbf{s}_2\|^2}{2\|\mathbf{A}\mathbf{s}_1\| \cdot \|\mathbf{A}\mathbf{s}'_1\|} \right\} = \frac{2\epsilon}{1 - \epsilon}, \quad (19)$$

where the condition $\epsilon \in [0, 1/3]$ ensures that $2\epsilon/(1 - \epsilon) \leq 1$. Since $\sin \alpha = \sin(\pi - \alpha)$ whenever $\alpha \in [0, \pi]$, the height of the projected triangle satisfies $h_p^2 = \|\mathbf{A}\mathbf{s}_1\|^2 \sin^2 \alpha_p$. The stated bound follows by considering worst-case expansions and contractions of $\|\mathbf{A}\mathbf{s}_1\|^2$, along with the implied bounds on $\sin^2 \alpha_p = 1 - \cos^2 \alpha_p$. ■

Now, since \mathbf{A} is a linear transformation, the relative length preservation of \mathbf{y} implies the relative length preservation of \mathbf{s}'_1 since they are parallel vectors. As such, the assumptions of the lemma are verified in our setting. Refer again to Fig. 3 and notice that $\sin(\alpha) = \|\mathbf{s}_1\|/\|\mathbf{x}\| = h/\|\mathbf{x}\|$ by definition. Further, the projected angle satisfies $\sin(\alpha_p) = h_p/\|\mathbf{A}\mathbf{x}\|$. Along with the above lemma which provides bounds on h_p , as well as the relative length preservation assumptions, this implies

$$\sqrt{\frac{1 - 3\epsilon}{1 - \epsilon}} \leq \frac{\sin(\alpha_p)}{\sin(\alpha)} \leq \sqrt{\frac{1 + \epsilon}{1 - \epsilon}}. \quad (20)$$

Notice that the lower bound is less or equal to 1 and the upper bound is greater or equal to 1 whenever $\epsilon \in [0, 1/3]$.

First consider the lower bound. We can write

$$\sqrt{\frac{1 - 3\epsilon}{1 - \epsilon}} \sin(\alpha) \leq \sin(\alpha_p), \quad (21)$$

or

$$\sin^{-1} \left(\sqrt{\frac{1 - 3\epsilon}{1 - \epsilon}} \sin(\alpha) \right) \leq \alpha_p, \quad (22)$$

which is well defined whenever $\alpha \in [0, \pi/2]$. To establish lower bounds that are linear in α , we note that the lower bound is a concave function of α over the domain $\alpha \in [0, \pi/2]$ and for $\epsilon \in [0, 1/3]$ (see Appendix), so it can be lower bounded by a line between any two points. Specifically, for a fixed ϵ , the worst-case lower bound is the line connecting the point 0 (achieved when $\alpha = 0$) to the point $\sin^{-1} \left(\sqrt{(1 - 3\epsilon)/(1 - \epsilon)} \right)$ (achieved when $\alpha = \pi/2$) to obtain

$$\frac{2}{\pi} \sin^{-1} \left(\sqrt{\frac{1 - 3\epsilon}{1 - \epsilon}} \right) \alpha \leq \sin^{-1} \left(\sqrt{\frac{1 - 3\epsilon}{1 - \epsilon}} \sin(\alpha) \right). \quad (23)$$

Unfortunately,

$$\sin^{-1} \left(\sqrt{\frac{1 - 3\epsilon}{1 - \epsilon}} \right) \quad (24)$$

is not a concave (or convex) function of ϵ , so we must employ a different bounding technique. We claim

$$1 - \sqrt{3\epsilon} \leq \frac{2}{\pi} \sin^{-1} \left(\sqrt{\frac{1-3\epsilon}{1-\epsilon}} \right). \quad (25)$$

To verify this, note that an equivalent statement is

$$\cos \left(\frac{\pi}{2} \sqrt{3\epsilon} \right) = \sin \left(\frac{\pi}{2} (1 - \sqrt{3\epsilon}) \right) \leq \sqrt{\frac{1-3\epsilon}{1-\epsilon}}. \quad (26)$$

But since $\sqrt{1-3\epsilon} \leq \sqrt{(1-3\epsilon)/(1-\epsilon)}$ and $1-3\epsilon \leq \sqrt{1-3\epsilon}$ when $\epsilon \in [0, 1/3]$, it suffices to show

$$\cos \left(\frac{\pi}{2} \sqrt{3\epsilon} \right) \leq 1 - 3\epsilon. \quad (27)$$

Here we can use the fact that $\cos(\sqrt{3\epsilon}\pi/2)$ is a convex function of ϵ on the domain we consider (see Appendix), and since it has the same endpoints as $\sqrt{1-3\epsilon}$, the claim is established. Thus, we have shown that, for $\alpha \in [0, \pi/2]$, the projected angle obeys

$$(1 - \sqrt{3\epsilon})\alpha \leq \alpha_p. \quad (28)$$

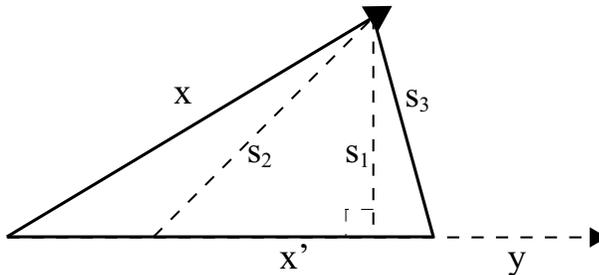


Fig. 4. The isosceles triangle used to establish the angular upper bound.

We now establish the angular upper bound for the case $\alpha \in [0, \pi/2]$. We will analyze the projection of the isosceles triangle having sides x and $\|x\|\hat{y}$ (See Fig. 4). Using standard trigonometry, it is straightforward to show that $\|s_3\|^2 = 2\|x\|^2(1 - \cos \alpha)$. Again, we appeal to the Law of Cosines, to find that

$$\cos \alpha_p = \frac{\langle \mathbf{Ax}, \mathbf{Ax}' \rangle}{2\|\mathbf{Ax}\| \cdot \|\mathbf{Ax}'\|} \geq \min \left\{ \frac{\|\mathbf{Ax}\|^2 + \|\mathbf{Ax}'\|^2 - \|\mathbf{As}_3\|^2}{2\|\mathbf{Ax}\| \cdot \|\mathbf{Ax}'\|} \right\} = \frac{(1 + \epsilon) \cos \alpha - 2\epsilon}{1 - \epsilon}, \quad (29)$$

which is equivalent to

$$\alpha_p \leq \cos^{-1} \left(\frac{(1 + \epsilon) \cos \alpha - 2\epsilon}{1 - \epsilon} \right). \quad (30)$$

This is a convex function of α over the domain we consider (see Appendix), so

$$\alpha_p \leq \frac{2}{\pi} \cos^{-1} \left(\frac{-2\epsilon}{1 - \epsilon} \right) \alpha. \quad (31)$$

Further, by the convexity of $\cos^{-1}(-2\epsilon/(1-\epsilon))$ (see Appendix), we have

$$\alpha_p \leq (1 + 3\epsilon)\alpha, \quad (32)$$

and so overall we have shown that, for $\alpha \in [0, \pi/2]$ and $\epsilon \in [0, 1/3]$,

$$(1 - \sqrt{3\epsilon})\alpha \leq \alpha_p \leq (1 + 3\epsilon)\alpha, \quad (33)$$

which proves the Theorem.

IV. APPLICATION – ROBUST DETECTION

In this section, the GRIP is used to establish performance bounds when detecting certain signals in the presence of white Gaussian noise. We begin with a brief summary of the canonical signal detection problem.

A. Signal Detection Preliminaries

This section gives a brief review of the basics of signal detection. There are many valid references for the material given here, such as [11]. The primary goal of signal detection is to reliably decide whether a signal is present from a set of (usually noisy) observations $\{y_j\}_{j=1}^k$. The case when the signal is absent is called the null hypothesis and is referred to by the notation H_0 . Similarly, H_1 refers to the case when the signal is present, also called the alternative hypothesis.

Each observation under H_0 or H_1 is a random variable with a probability distribution that depends on the observation model. We denote the distribution of each observation y_j under the null and alternative hypotheses by $P_0(y_j)$ and $P_1(y_j)$, respectively. When the observations are independent, we can write the distribution of $\mathbf{y} = [y_1 \dots y_k]$ as a product of the individual distributions.

A detector operates by testing whether the observation vector \mathbf{y} falls within some prescribed acceptance region \mathcal{A} . The detector announces that the signal is present if $\mathbf{y} \in \mathcal{A}$, and absent otherwise. The false alarm probability (the probability that the detector announces the signal is present when it is not) is $P_{\text{FA}} = \int_{\mathcal{A}} P_0(\mathbf{y}) d\mathbf{y}$, and the miss probability (the probability that the detector announces the signal is not present when it is) is given by $P_{\text{M}} = \int_{\mathcal{A}^c} P_1(\mathbf{y}) d\mathbf{y}$.

The Neyman-Pearson test is derived by computing the likelihood ratio between the probability distributions under H_1 and H_0 . The detector decides H_1 if $P_1(\mathbf{y})/P_0(\mathbf{y}) > \gamma$ for some threshold γ , and decides H_0 otherwise. It can be shown that the Neyman-Pearson test is the most powerful test in the sense that it minimizes P_{M} subject to $P_{\text{FA}} \leq \alpha$. When the goal is to reduce both errors simultaneously, the total probability of error is the preferred metric. Define P_{ERROR} to be the sum of the false alarm probability P_{FA} and the miss probability P_{M} . To optimize this total error, the contributions from each error term are balanced. When the observations are i.i.d., this optimal error obeys $P_{\text{ERROR}} \leq \prod_{j=1}^k \Gamma_j^*$, where $\Gamma_j^* = \min_{\lambda \in [0,1]} \Gamma_j(\lambda)$ and

$$\Gamma_j(\lambda) = \int_{-\infty}^{\infty} P_0^\lambda(y_j) P_1^{1-\lambda}(y_j) dx,$$

as shown in [12].

B. Problem Statement

Suppose that an observer is allowed to make only a limited number of noisy observations of an unknown vector $\mathbf{x} \in \mathbb{R}^n$, where each observation is the inner product between the signal and a sampling vector chosen by and known to the observer. These *projection samples* are described by

$$y_j = \langle \mathbf{x}, \mathbf{A}_j \rangle + w_j, \quad (34)$$

for $j = 1, \dots, k$, where $k < n$, $\mathbf{A}_j \in \mathbb{R}^n$ are the sampling vectors, and $\{w_j\}$ is a collection of independent and identically distributed (i.i.d.) $\mathcal{N}(0, \sigma^2)$ noises. While the first works in CS dealt with signal reconstruction [1]–[5], our goal here will be detection – to reliably determine, from the observed data $\{y_j\}$ and the known projection vectors $\{\mathbf{A}_j\}$, whether the signal is present.

C. Detection from Projection Samples Without Specific Signal Information

We first review prior results on detecting an arbitrary but unknown signal in the presence of additive white Gaussian noise [13]. Assume that the entries of the test functions \mathbf{A}_j defined in (34) are i.i.d. $\mathcal{N}(0, 1/n)$. Under the null hypothesis, observations will be noise only, independent, and distributed as

$$H_0 : P_0(y_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y_j^2}{2\sigma^2}\right).$$

Because the signal is unknown, exact knowledge of \mathbf{A} is not necessary. The presence of the signal amounts to a variance shift in the observations, which under the alternative hypothesis are independent and distributed as

$$H_1 : P_1(y_j) = \frac{1}{\sqrt{2\pi(\sigma^2 + \|\mathbf{x}\|^2/n)}} \exp\left(-\frac{y_j^2}{2(\sigma^2 + \|\mathbf{x}\|^2/n)}\right),$$

for $j = 1, \dots, k$. In this setting, it is straightforward to show that the optimal detector is an energy detector of the form $\|\mathbf{y}\|^2$. Following the approach in [12], we obtain

$$P_{\text{ERROR}} \leq \left[\exp(1/2) \sqrt{\frac{\log(1+S)}{S(1+S)^{1/S}}} \right]^k, \quad (35)$$

where $S = \|\mathbf{x}\|^2/n\sigma^2$ is the signal-to-noise ratio (SNR). This result illustrates that any unknown signal \mathbf{x} can be reliably detected using CS without using any prior knowledge of the signal. As we will see, though, the error decay rate can be improved if more is known about the signal.

D. Detection from Projection Samples Using Prior Signal Information

1) *Signal Detection From Matched Samples:* For fixed sensing matrices \mathbf{A} , the observations when the signal is absent are noise only, independent, and distributed as

$$H_0 : P_0(y_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y_j^2}{2\sigma^2}\right),$$

and when the signal is present, the observations are independent and distributed as

$$H_1 : P_1(y_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_j - \langle \mathbf{A}_j, \mathbf{x} \rangle)^2}{2\sigma^2}\right],$$

for $j = 1, \dots, k$. Using standard techniques from classical detection theory, it is straightforward to show that the optimal test statistic in this setting is of the form $\mathbf{y}^T \mathbf{A} \mathbf{x}$, which means that for some threshold τ , the detector decides that the signal is present if $\mathbf{y}^T \mathbf{A} \mathbf{x} > \tau$ and decides the signal is absent otherwise. In this case, we have $P_{\text{ERROR}} \leq \exp(-\|\mathbf{A} \mathbf{x}\|^2 / 8\sigma^2)$. If the rows of \mathbf{A} are given by $\tilde{\mathbf{x}} / \|\tilde{\mathbf{x}}\|$ for some approximation $\tilde{\mathbf{x}}$ such that $\langle \mathbf{x}, \tilde{\mathbf{x}} \rangle \geq 0$, then

$$P_{\text{ERROR}} \leq \exp \left[\frac{-k \|\mathbf{x}\|^2 \cos^2(\alpha)}{8} \right] \quad (36)$$

$$= \exp \left[-knS \cos^2(\alpha) / 8 \right], \quad (37)$$

where α is the angle between \mathbf{x} and $\tilde{\mathbf{x}}$. This result illustrates the robustness of the detector that employs matched sampling – the error performance degrades gracefully as the quality of the approximation decreases (*i.e.*, as α increases). The case $\alpha = 0$ is called the matched filter – the observations are projections onto (unit-normed versions of) the signal to be detected.

2) *Sparse Signal Detection From Compressive Samples*: Motivated by the idea that knowledge of the signal to be detected results in a detector whose error probability decays more rapidly than detectors that measure energy only, we propose a CS matched filtering approach that exploits the sparsity of the signal to be detected along with some amount “future knowledge” to offer both universality and robustness. The universality comes from the fact that prior signal information is not needed at the time the observations are collected. The “future knowledge” required is some reliable approximation of the signal which is not derived from the observations themselves. This could come, for example, from an auxiliary intelligence channel.

We proceed by utilizing the form of the ideal detector, but instead of correlating the observation vector with the projection of the true signal we instead correlate the observation vector with the projection of an approximation of the true signal. Specifically, let $\tilde{\mathbf{x}}$ denote our approximation of \mathbf{x} obtained through auxiliary methods. We will examine the error performance of the detector given by $\mathbf{y}^T \mathbf{A} \tilde{\mathbf{x}}$. Again, we assume that the approximation $\tilde{\mathbf{x}}$ is aligned with the true signal \mathbf{x} so that $\langle \mathbf{x}, \tilde{\mathbf{x}} \rangle \geq 0$ (*i.e.*, α is acute).

We examine the error performance by considering the distribution of the test statistic $T = \mathbf{y}^T \mathbf{A} \tilde{\mathbf{x}}$ directly. Under the null hypothesis, $T \sim \mathcal{N}(0, \|\mathbf{A} \tilde{\mathbf{x}}\|^2 \sigma^2)$, and when the signal is present, $T \sim \mathcal{N}(\langle \mathbf{A} \mathbf{x}, \mathbf{A} \tilde{\mathbf{x}} \rangle, \|\mathbf{A} \tilde{\mathbf{x}}\|^2 \sigma^2)$. Thus, the error probabilities are obtained by integrating tails of the appropriate Gaussian distributions. Now assume that the vectors \mathbf{x} and $\tilde{\mathbf{x}}$ are sufficiently sparse. Assume also that the matrix \mathbf{A} satisfies RIP (and GRIP) with $C = k/n$ and some sufficiently small ϵ (for example, $A_{i,j} \sim \mathcal{N}(0, 1/n)$, and $k \geq \text{const} \cdot m \log n$). We use the distributions that give the worst separation and error behavior by choosing those with the lowest mean and highest variance. Thus, under the null hypothesis we have $T \sim \mathcal{N}(0, (1 + \epsilon)(k/n) \|\tilde{\mathbf{x}}\|^2 \sigma^2)$, and when the signal is present, $T \sim \mathcal{N}((1 - \epsilon)(k/n) \|\mathbf{x}\| \|\tilde{\mathbf{x}}\| \cos[(1 + 3\epsilon)\alpha], (1 + \epsilon)(k/n) \|\tilde{\mathbf{x}}\|^2 \sigma^2)$, where α denotes the acute angle between \mathbf{x} and $\tilde{\mathbf{x}}$. For a positive detection threshold $\tau < (1 - \epsilon)(k/n) \|\mathbf{x}\| \|\tilde{\mathbf{x}}\| \cos[(1 + 3\epsilon)\alpha]$ we utilize a bound on the the integral of Gaussian tails [14] to obtain

$$P_{\text{FA}} \leq \frac{1}{2} \exp \left[\frac{-\tau^2}{2\sigma^2(1 + \epsilon)(k/n) \|\tilde{\mathbf{x}}\|^2} \right]$$

and

$$P_M \leq \frac{1}{2} \exp \left[\frac{-((1-\epsilon)(k/n)\|\mathbf{x}\|\|\tilde{\mathbf{x}}\| \cos[(1+3\epsilon)\alpha] - \tau)^2}{2\sigma^2(1+\epsilon)(k/n)\|\tilde{\mathbf{x}}\|^2} \right].$$

Setting $\tau = (1/2)(1-\epsilon)(k/n)\|\mathbf{x}\|\|\tilde{\mathbf{x}}\| \cos[(1+3\epsilon)\alpha]$ makes the two errors equal, giving

$$\begin{aligned} P_{\text{ERROR}} &\leq \exp \left[-\frac{(1-\epsilon)^2(k/n)\|\mathbf{x}\|^2 \cos^2[(1+3\epsilon)\alpha]}{8\sigma^2(1+\epsilon)} \right] \\ &= \exp \left[-\frac{kS(1-\epsilon)^2 \cos^2[(1+3\epsilon)\alpha]}{8(1+\epsilon)} \right], \end{aligned}$$

valid for all angles for which $\tau = (1/2)(1-\epsilon)(k/n)\|\mathbf{x}\|\|\tilde{\mathbf{x}}\| \cos[(1+3\epsilon)\alpha]$ is positive.

Of course, the above bounds are conditioned on the event that the measurement matrix satisfies RIP (and GRIP).

The unconditional probability of error is given by

$$\begin{aligned} P_{\text{ERROR}} &\leq \exp \left[-\frac{kS(1-\epsilon)^2 \cos^2[(1+3\epsilon)\alpha]}{8(1+\epsilon)} \right] \cdot P(\text{RIP}) + P(\text{RIP}) \\ &\leq \exp \left[-\frac{kS(1-\epsilon)^2 \cos^2[(1+3\epsilon)\alpha]}{8(1+\epsilon)} \right] + \exp \left[-c_0k + m \log \left(\frac{en}{m} \right) + m \log \left(\frac{12}{\epsilon} \right) + \log(2) \right] \\ &\approx \exp \left[-\frac{k\|\mathbf{x}\|^2}{8n} \cos^2(\alpha) \right] + C'n^m \exp[-c_0k], \end{aligned}$$

where $C' > 1$ is a factor that does not depend on k or n and the approximation is valid for small ϵ . The main point to note here is that the error decay rate (as a function of problem dimensions k and n) is dominated by the first term where the exponential scaling depends on k/n instead of the dependence on k alone in the second term. Further, note that the second term is negligible and can be ignored when $k \geq \text{const} \cdot m \log n$. Comparing the resulting approximation with (36), we see that the only difference is a factor of n in the rate. This arises because the amount of signal energy captured by each CS observation of the signal \mathbf{x} , in expectation, is given by $\|\mathbf{x}\|^2/n$, while on the contrary, each optimal sample captures the full signal energy $\|\mathbf{x}\|^2$. The loss by a factor of n illustrates a price to pay for the universality that CS provides.

V. SIMULATION RESULTS

In this section we examine the dependence of the minimum probability of error on the number of observations and SNR for the detectors that employ CS. To illustrate the robustness of the CS matched filter detector we consider two approximations, the best one-term approximation of \mathbf{x} and \mathbf{x} itself. For the realizations used in the simulations, the quality of the one term approximation $\tilde{\mathbf{x}}$ is such that $\cos(\alpha) = 0.85$.

In the first simulation, $n = 1000$ and $\mathbf{x} \in \mathbb{R}^n$ is a vector with three nonzero entries chosen at random, but normalized so that $\|\mathbf{x}\|^2 = 1$. For each detector, and for each value of k (the number of observations), 4000 realizations of the test statistic for the signal-absent and signal-present cases were computed, each using a randomly generated sampling matrix [with i.i.d. $\mathcal{N}(0, 1/n)$ entries] and noise vector [with i.i.d. $\mathcal{N}(0, 0.025)$ entries]. For each case, we used the empirical histograms of the distributions under H_0 and H_1 to compute the minimum total error probability.

The results are shown in Fig. 5(a), where the curves drawn with diamond markers indicate the theoretical bounds, and the curves with x markers show the simulation results. Errors corresponding to the energy detector, CS matched filter with approximate knowledge, and CS matched filter with exact knowledge are drawn using the dash-dot, dashed, and solid lines, respectively. We see that the empirical results agree with the theory.

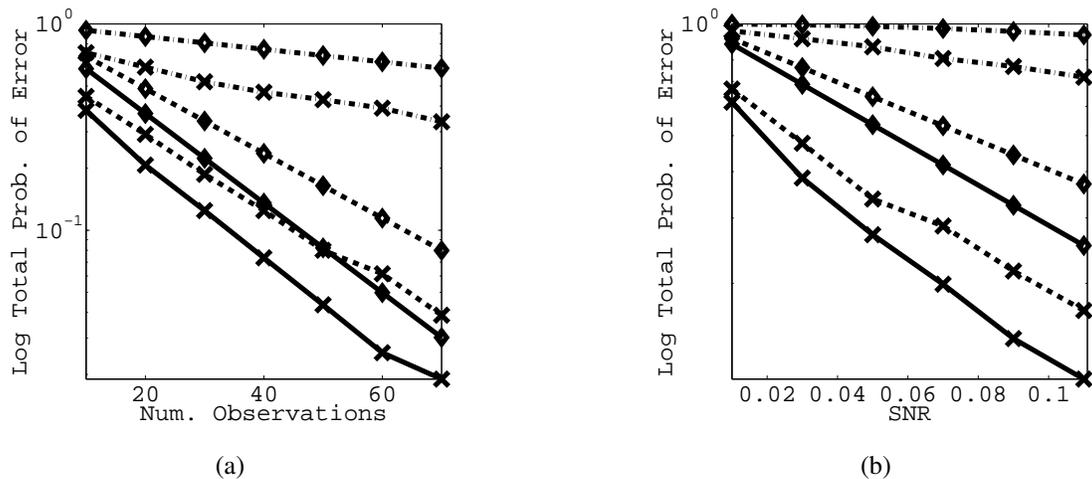


Fig. 5. Empirical probabilities of error for the robust detector, shown as a function of (a) number of observations, and (b) SNR.

The second simulation illustrates the scaling behavior with respect to SNR. Again, $n = 1000$, and this time the number of observations was fixed at 100. For each value of σ^2 , 1000 realizations of sampling matrices with i.i.d. $\mathcal{N}(0, 1/n)$ entries and vectors of i.i.d. $\mathcal{N}(0, \sigma^2)$ additive noise were generated, and test statistics for each hypothesis were computed. The minimum average probability of error was again determined empirically. The results depicted in Fig. 5(b) illustrate that the theoretical results again agree with the predicted error behavior. It is interesting to note that the dependence of the detector performance on the quality of approximation is noticeable in both simulations – the better approximation yields faster error decay.

VI. ACKNOWLEDGEMENT

J.H. would like to thank Steven Meitner for some enlightening geometrical discussions which were helpful in streamlining the proofs.

APPENDIX

A series of lemmata are established here.

Lemma 4 *The function*

$$f(\epsilon, \alpha) = \sin^{-1} \left(\sqrt{\frac{1-3\epsilon}{1-\epsilon}} \sin(\alpha) \right) \quad (38)$$

is a concave function of α when $\alpha \in [0, \pi/2]$ and $\epsilon \in [0, 1/3]$.

Proof: The second (partial) derivative is

$$\frac{\partial^2 f}{\partial \alpha^2} = \frac{\sqrt{(1-3\epsilon)/(1-\epsilon)} \sin(\alpha)}{\sqrt{1-(1-3\epsilon)\sin^2(\alpha)/(1-\epsilon)}} \left[\frac{\cos^2(\alpha)}{(1-(1-3\epsilon)\sin^2(\alpha)/(1-\epsilon))} \left(\frac{1-3\epsilon}{1-\epsilon} \right) - 1 \right]. \quad (39)$$

Now, $(1-3\epsilon)/(1-\epsilon) \leq 1$, so it is easy to see that

$$\cos^2(\alpha) \leq 1 - \frac{(1-3\epsilon)\sin^2(\alpha)}{1-\epsilon} \quad (40)$$

since

$$\frac{(1-3\epsilon)\sin^2(\alpha)}{1-\epsilon} \leq 1 - \cos^2(\alpha) = \sin^2(\alpha). \quad (41)$$

Therefore, the term in square brackets is negative. Further, the leading term is positive, so the function is concave as claimed. ■

Lemma 5 (Harmonic Addition) *The weighted sum of a sine and a cosine function (of the same argument) can be written as a single sinusoid. That is,*

$$a \sin x + b \cos x = \pm \sqrt{a^2 + b^2} \cos \left[x + \tan^{-1} \left(-\frac{b}{a} \right) \right]. \quad (42)$$

The statement follows from a straightforward application of the identity $\cos(a+b) = \cos a \cos b - \sin a \sin b$, and by equating terms.

Lemma 6 *The function $f(\epsilon) = \cos(\pi\sqrt{3\epsilon}/2)$ is a convex function of ϵ when $\epsilon \in (0, 1/3)$.*

Proof: Generically, let $f(\epsilon) = \cos(c\sqrt{\epsilon})$, then

$$\frac{d^2 f}{d\epsilon^2} = \frac{c}{4\epsilon^{3/2}} \sin(c\sqrt{\epsilon}) - \frac{c^2}{4\epsilon} \cos(c\sqrt{\epsilon}) = \pm \frac{c}{4\epsilon} \sqrt{\frac{1+c^2\epsilon}{\epsilon}} \cos \left[c\sqrt{\epsilon} + \tan^{-1} \left(\frac{1}{c\sqrt{\epsilon}} \right) \right] \quad (43)$$

using harmonic addition. In our setting, $c = \pi\sqrt{3}/2$. We can evaluate the expression at $\epsilon = 1/3$ to remove the sign ambiguity, giving

$$\frac{d^2 f}{d\epsilon^2} = -\frac{\pi\sqrt{3}}{8\epsilon} \sqrt{\frac{1}{\epsilon} + \frac{3\pi^2}{4}} \cos \left[\frac{\pi\sqrt{3\epsilon}}{2} + \tan^{-1} \left(\frac{2}{\pi\sqrt{3\epsilon}} \right) \right]. \quad (44)$$

Now, the argument of the \cos function is always between $\pi/2$ and π over the range of ϵ considered. To see this, first consider the lower limit. The claimed lower bound can be established by showing

$$\tan^{-1} \left(\frac{2}{\pi x} \right) \geq \frac{\pi(1-x)}{2} \quad (45)$$

for $x \in [0, 1]$. This is equivalent to

$$\tan \left(\frac{\pi x}{2} \right) \geq \frac{\pi x}{2}, \quad (46)$$

and this statement can be shown to hold using a Taylor series expansion of the \tan function. Now, the upper bound is nearly trivial, since

$$\tan^{-1} \left(\frac{2}{\pi\sqrt{3\epsilon}} \right) \leq \pi/2 \leq \pi \left(1 - \frac{\sqrt{3\epsilon}}{2} \right) \quad (47)$$

over the domain we consider. Therefore, the second derivative is positive over the domain considered and the convexity is shown. ■

Lemma 7 *The function*

$$f(\epsilon, \alpha) = \cos^{-1} \left(\frac{(1 + \epsilon) \cos(\alpha) - 2\epsilon}{1 - \epsilon} \right) \quad (48)$$

is a convex function of α when $\alpha \in [0, \pi/2]$ and $\epsilon \in [0, 1/3]$.

Proof: The second (partial) derivative is

$$\frac{\partial^2 f}{\partial \alpha^2} = \frac{2\epsilon(\cos(\alpha) - 1)(1 + \epsilon)}{(\epsilon - 1)[\cos(\alpha)(1 + \epsilon) + (1 - 3\epsilon)]} \left[\frac{(1 - \cos(\alpha)) [\cos(\alpha) (\epsilon^2 + 2\epsilon + 1) + (-3\epsilon^2 - 2\epsilon + 1)]}{(\epsilon - 1)^2} \right]^{-1/2}. \quad (49)$$

Analyzing the sign of each term, we see

$$\frac{\partial^2 f}{\partial \alpha^2} = \frac{\underbrace{2\epsilon}_{pos} \underbrace{(1 - \cos(\alpha))}_{pos} \underbrace{(1 + \epsilon)}_{pos}}{\underbrace{(1 - \epsilon)}_{pos} \underbrace{[\cos(\alpha)(1 + \epsilon) + (1 - 3\epsilon)]}_{pos}} \left[\frac{\overbrace{(1 - \cos(\alpha))}^{pos} \left[\overbrace{\cos(\alpha) (\epsilon^2 + 2\epsilon + 1)}^{pos} + \overbrace{(-3\epsilon^2 - 2\epsilon + 1)}^{pos\ddagger} \right]}{\underbrace{(\epsilon - 1)^2}_{pos}} \right]^{-1/2}, \quad (50)$$

where most of the claims are obvious. The term marked with \ddagger is positive because it is a parabola with negative leading term having roots -1 and $1/3$. Thus the claim is established. ■

Lemma 8 *The function*

$$f(\epsilon) = \cos^{-1} \left(\frac{-2\epsilon}{1 - \epsilon} \right) \quad (51)$$

is a convex function of ϵ when $\epsilon \in [0, 1/3]$.

Proof: The second derivative is

$$\frac{d^2 f}{d\epsilon^2} = \frac{4(1 - 3\epsilon^2)}{(\epsilon - 1)^3 (3\epsilon^2 + 2\epsilon - 1) \sqrt{-(3\epsilon^2 + 2\epsilon - 1)/(\epsilon - 1)^2}}. \quad (52)$$

Inspecting the sign of each term gives

$$\frac{d^2 f}{d\epsilon^2} = \frac{4 \overbrace{(1 - 3\epsilon^2)}^{pos}}{\underbrace{(\epsilon - 1)^3}_{neg} \underbrace{(3\epsilon^2 + 2\epsilon - 1)}_{neg} \sqrt{-(3\epsilon^2 + 2\epsilon - 1)/(\epsilon - 1)^2}}, \quad (53)$$

where the quadratic term is negative since it is a parabola with positive leading term having roots -1 and $1/3$. Since the second derivative is positive throughout, the function is convex as claimed. ■

REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [3] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [4] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, Sept. 2006.
- [5] E. Candès and T. Tao, "The Dantzig selector: statistical estimation when p is much larger than n ," *Annals of Statistics*, to appear.
- [6] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [7] R. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, to appear.
- [8] A. Magen, "Dimensionality reductions that preserve volumes and distance to affine spaces, and their algorithmic applications," in *RANDOM '02: Proceedings of the 6th International Workshop on Randomization and Approximation Techniques*. London, UK: Springer-Verlag, 2002, pp. 239–253.
- [9] D. Achlioptas, F. McSherry, and B. Schölkopf, "Sampling techniques for kernel methods," in *Annual Advances in Neural Inform. Processing Systems 14*, 2002, pp. 335–342.
- [10] M. Davenport, M. Wakin, and R. Baraniuk, "Detection and estimation with compressive measurements," Rice University, Tech. Rep. TREE 0610, 2006.
- [11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall, 1998.
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.
- [13] R. Castro, J. Haupt, and R. Nowak, "Active learning vs compressive sampling," in *Proc. IS&T/SPIE Defense and Security Symposium*, Orlando, FL, 2006.
- [14] S. Verdu, *Multiuser Detection*. Cambridge University Press, 1998.