

# Sensitivity to Basis Mismatch in Compressed Sensing

Yuejie Chi, *Student Member, IEEE*, Louis L. Scharf, *Life Fellow, IEEE*, Ali Pezeshki, *Member, IEEE*, and A. Robert Calderbank, *Fellow, IEEE*

**Abstract**—The theory of compressed sensing suggests that successful inversion of an image of the physical world (broadly defined to include speech signals, radar/sonar returns, vibration records, sensor array snapshot vectors, 2-D images, and so on) for its source modes and amplitudes can be achieved at measurement dimensions far lower than what might be expected from the classical theories of spectrum or modal analysis, provided that the image is sparse in an *a priori known* basis. For imaging problems in spectrum analysis, and passive and active radar/sonar, this basis is usually taken to be a DFT basis. However, in reality no physical field is sparse in the DFT basis or in any *a priori known* basis. No matter how finely we grid the parameter space the sources may not lie in the center of the grid cells and consequently there is *mismatch* between the assumed and the actual bases for sparsity. In this paper, we study the sensitivity of compressed sensing to mismatch between the assumed and the actual sparsity bases. We start by analyzing the effect of basis mismatch on the best  $k$ -term approximation error, which is central to providing exact sparse recovery guarantees. We establish achievable bounds for the  $\ell_1$  error of the best  $k$ -term approximation and show that these bounds grow linearly with the image (or grid) dimension and the mismatch level between the assumed and actual bases for sparsity. We then derive bounds, with similar growth behavior, for the basis pursuit  $\ell_1$  recovery error, indicating that the sparse recovery *may* suffer large errors in the presence of basis mismatch. Although, we present our results in the context of basis pursuit, our analysis applies to any sparse recovery principle that relies on the accuracy of best  $k$ -term approximations for its performance guarantees. We particularly highlight the problematic nature of basis mismatch in Fourier imaging, where spillage from off-grid DFT components turns a sparse representation into an incompressible one. We substantiate our mathematical analysis by numerical examples that demonstrate a considerable performance degradation for image inversion from compressed sensing measurements in the presence of basis mismatch, for problem sizes common to radar and sonar.

**Index Terms**—Compressed sensing, image inversion, modal analysis, sensitivity to basis mismatch, sparse recovery.

Manuscript received May 15, 2010; revised October 11, 2010; accepted January 14, 2011. Date of publication February 10, 2011; date of current version April 13, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Konstantinos I. Diamantaras. This work was supported in part by the NSF by Grants CCF-1018472, CCF-1017431, CCF-0916314, CCF-0915299, and DMS-0701226, by the ONR under Grant N00173-06-1-G006, and by the AFOSR under Grants FA9550-05-1-0443 and FA9550-10-1-0241. The material in this paper was presented at the Joint US-Australia Workshop on Defence Applications of Signal Processing (DASP), Lihue, HI, Sep. 26–30, 2009, and at the International Conference on Acoustic, Speech, and Signal Processing (ICASSP), Dallas, TX, Mar. 16–20, 2010.

Y. Chi is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: ychi@princeton.edu).

L. L. Scharf and A. Pezeshki are with the Department of Electrical and Computer Engineering, and the Department of Statistics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: Louis.Scharf@colostate.edu; Ali.Pezeshki@colostate.edu).

A. R. Calderbank is with the Department of Computer Science, Duke University, Durham, NC 27708 USA (e-mail: robert.calderbank@duke.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2011.2112650

## I. INTRODUCTION

### A. Motivation

IN a great number of fields of engineering and applied science the problem confronting the designer is to *invert an image*, acquired from a sensor suite, for the underlying field that produced the image. And typically the desired resolution for the underlying field exceeds the temporal or spatial resolution of the image itself. Here we give *image* its most general meaning to encompass a times series, a space series, a space-time series, a 2-D image, and so on. Similarly, we give *field* its most general meaning to encompass complex-exponential modes, radiating modes, coded modulations, multipath components, and the like. Certainly this interpretation includes the problem of identifying field elements from electromagnetic and acoustic images, multipath components in wireless communication, radiating sources in radar and sonar, and light sources in optical imaging and spectrometry.

Broadly speaking there are two main (classical) principles for inverting the kinds of images that are measured in speech, communication, radar, sonar, and optics. The first principle is one of matched filtering, wherein a sequence of test images is matched to the measured image. The test images are generated by scanning a prototype image (e.g., a waveform or a steering vector) through frequency, wavenumber, doppler, and/or delay. In time series analysis, this amounts to classical spectrum analysis to identify the frequency modes, and the corresponding mode amplitudes or powers, of the signal [1], [2]. In phased-array processing, it amounts to spectrum analysis in frequency and wavenumber to identify the frequency-wavenumber coordinates of source radiations impinging on the array [3]–[6]. In Space-Time Adaptive Processing (STAP) for radar and sonar, it amounts to spectrum analysis in delay, frequency, and wavenumber to reconstruct the radar/sonar field [7], [8]. The second principle is one of parameter estimation in a separable linear model, wherein a sparse *modal* representation for the field is posited and estimates of linear parameters (complex amplitudes of modes) and nonlinear mode parameters (frequency, wavenumber, delay, and/or doppler) are extracted, usually based on maximum likelihood, or some variation on linear prediction, using  $\ell_2$  minimization (see, e.g., [2], [9], [10]). There is a comprehensive literature in electrical engineering, physics, and chemistry on the performance and limitations of these two classical principles (see, e.g., [1]–[13]). One important limitation is that any subsampling of the measured image has consequences for resolution (or bias) and for variability (or variance).

The recent advent of compressed sensing theory has revolutionized our view of imaging, as it demonstrates that

subsampling (actually *subrecording*<sup>1</sup>) has manageable consequences for image inversion, provided that the image is sparse in an *a priori known* basis (see, e.g., [14]–[30]). For imaging problems in spectrum analysis (estimating complex exponential modes), and passive and active radar/sonar (estimating Doppler and angle of arrival), this basis is usually *taken* to be a Fourier basis (actually a DFT basis) constructed for resolution of  $2\pi/N$ , with  $N$  a window length, array length, or pulse-to-pulse processing length. Several articles (see, e.g., [31]–[37]) consider the use of compressed sensing theory for discrete radar/sonar imaging, and sensor array processing, when the targets are taken to be on a regular grid in delay, Doppler, and wavenumber, and study this theory as a new high resolution imaging principle. But no matter how large the size  $N$  of the grid is, the actual field will not place its sources on the center of the grid points  $\{2\pi n/N\}$  in frequency or wavenumber, or on the center of the grid points in delay-Doppler-wavenumber. This means the image is actually *not sparse* in the DFT basis or the basis defined by the grid. In fact any target that lies between two cells of a discretely-resolved range-doppler plane or frequency-wavenumber plane will spill non-zero values into all cells, with the amplitude of the spillage following a Dirichlet kernel, decaying as  $1/f$ , where  $f$  is frequency or wavenumber. This spillage turns a sparse representation into an incompressible one in the DFT basis. Grid misalignment problems also arise in many other applications including channel estimation [38]–[41]. These observations raise the following question: What is the sensitivity of compressed sensing for image inversion to *mismatch* between the assumed basis for sparsity and the actual basis in which the image is sparse? In this paper, we aim to answer this question and examine its consequences for problems in spectrum analysis, beamforming, modal analysis, and radar/sonar imaging.

## B. Basis Mismatch

In order to frame our question more precisely, let us begin with two models for a measured image  $\mathbf{s} \in \mathbb{C}^N$ . In the mathematical model to be *assumed* in the compressed sensing procedure, the image is composed as  $\mathbf{s} = \Psi_0 \mathbf{x}$ , where the basis  $\Psi_0 \in \mathbb{C}^{N \times N}$  is known, and is typically a gridded imaging matrix (e.g., the  $N$ -point DFT matrix), and  $\mathbf{x} \in \mathbb{C}^N$  is a sparse or compressible vector<sup>2</sup> of field parameters that compose the image as a linear combination of columns of  $\Psi_0$ . But, as a matter of fact, the image  $\mathbf{s}$  is composed by the physics as  $\mathbf{s} = \Psi_1 \boldsymbol{\theta}$ , where the basis  $\Psi_1 \in \mathbb{C}^{N \times N}$  is determined by a point spread function, a Green's function, or an impulse response, and the field parameter vector  $\boldsymbol{\theta}$  is sparse. Typically  $\Psi_1$  is determined by frequency, wavenumber, delay, and/or doppler parameters that are

<sup>1</sup>A single measurement in compressed sensing is an inner product between the high-dimensional signal vector and a sensing (recording) vector that typically consists of i.i.d. Gaussian or Bernoulli entries. Computing this inner product still requires access to *all* elements (typically obtained by sampling a full field of view imaging system) in the high-dimensional vector. But we only *record* a small number of such inner products as samples.

<sup>2</sup>If  $\mathbf{x}$  is sparse, then the cardinality of its support  $T_{\mathbf{x}} = \{k : x_k \neq 0\}$  is assumed to be small. If  $\mathbf{x}$  is compressible, then its entries obey a power law i.e. the  $k$ th largest entry of absolute values satisfies  $|x_{(k)}| \leq C_r \cdot k^{-r}$ ,  $r > 1$  and  $C_r$  is a constant depending only on  $r$  (see, e.g., [22]). Then  $\|\mathbf{x} - \mathbf{x}_k\|_1 \leq \sqrt{k} C_r' \cdot k^{-r+1}$ , where  $\mathbf{x}_k$  is the best  $k$ -term approximation of  $\mathbf{x}$ .

*unknown* a priori. More importantly, these parameters do *not* lie exactly on the gridding points of  $\Psi_0$ , e.g., a DFT matrix or an identity matrix. So  $\Psi_0 \neq \Psi_1$ . We call this *basis mismatch*, and note that it is present in all imaging problems, no matter how large  $N$  is, or equivalently no matter how fine-grained the gridding procedure is.

Each of the hypothesized models for the image  $\mathbf{s}$  may be inverted for its field parameters:

$$\mathbf{x} = \Psi_0^{-1} \mathbf{s} \quad \text{and} \quad \boldsymbol{\theta} = \Psi_1^{-1} \mathbf{s}. \quad (1)$$

These inversions determine the coordinate transformation

$$\mathbf{x} = \Psi \boldsymbol{\theta} \quad \text{and} \quad \boldsymbol{\theta} = \Psi^{-1} \mathbf{x} \quad (2)$$

where  $\Psi = \Psi_0^{-1} \Psi_1 \in \mathbb{C}^{N \times N}$ .

If  $\mathbf{s}$  is sparse in  $\Psi_1$ , then the field parameters  $\boldsymbol{\theta}$  will be sparse in the identity basis, denoted by  $\mathbf{I}$ . The field parameters  $\mathbf{x} = \Psi \boldsymbol{\theta}$  in the model  $\mathbf{s} = \Psi_0 \mathbf{x}$  will be sparse in the  $\Psi$  basis, *but not in the identity basis*. So the question is, “what is the consequence of assuming that  $\mathbf{x}$  is sparse in  $\mathbf{I}$ , when in fact it is only sparse in an *unknown* basis  $\Psi$ , which is determined by the mismatch between  $\Psi_0$  and  $\Psi_1$ ?”

We answer this question by deriving bounds on the  $\ell_1$ -norm (and also the  $\ell_2$ -norm) of the error in approximating the presumably sparse parameter vector  $\mathbf{x}$  from its compressed sensing measurements. We start by analyzing the effect of basis mismatch on the best  $k$ -term approximation error  $\mathbf{x} - \mathbf{x}_k$ , which is central to arguing for the accuracy of the basis pursuit solution. Here, we derive bounds for  $\|\mathbf{x} - \mathbf{x}_k\|_1$  in terms of the mismatch level between  $\Psi$  and  $\mathbf{I}$  and the image dimension  $N$ . More specifically, we find a tight upper bound for  $\|\mathbf{x} - \mathbf{x}_k\|_1$  and show that this bound grows as  $(N - k)\beta\|\boldsymbol{\theta}\|_q$ , where  $\beta$  upper bounds the degree of mismatch between the rows of  $\Psi$  and  $\mathbf{I}$ , and  $q \geq 1$ . We also establish a tight lower bound on the worst-case  $\ell_1$  error in the best  $k$ -term approximation  $\mathbf{x} - \mathbf{x}_k$  and show that it grows as  $(N - k)\eta\|\boldsymbol{\theta}\|_q$ , with  $\eta$  representing the lower bound on the degree of mismatch between the rows of  $\Psi$  and  $\mathbf{I}$ . Subsequently, we derive bounds on the  $\ell_1$  and  $\ell_2$  norms of the errors in estimating the parameter vector  $\mathbf{x}$  and the image  $\mathbf{s} = \Psi_0 \mathbf{x}$  using basis pursuit. We refer to the former error as *image inversion* error, which speaks to how well the basis pursuit solution approximates the presumably sparse parameter vector  $\mathbf{x}$  in the mathematical model  $\mathbf{s} = \Psi_0 \mathbf{x}$ , when in fact the representation  $\mathbf{x}$  in the the basis  $\Psi_0$  is non-sparse or incompressible. The latter error is the *image reconstruction* error, which speaks to how well the basis pursuit approximation to  $\mathbf{x}$  recomposes the image  $\mathbf{s}$  in the a priori selected basis  $\Psi_0$ . All these bounds exhibit similar growth behavior as those for the best  $k$ -term approximation error, but they are not tight. Our analysis shows that, in the presence of basis mismatch, exact or near-exact (within noise levels) recovery cannot be guaranteed from bounds that involve best  $k$ -term approximation errors and suggests that the basis pursuit recovery *may* suffer large errors. Our numerical examples demonstrate a considerable performance degradation in recovering  $\mathbf{x}$  from compressed sensing measurements, when the assumed basis for sparsity is a DFT basis but the actual sparsity basis does not align with the DFT basis. The inaccuracy in field reconstruction persists even when the number of compressed sensing measurements is increased

to the full image dimension. Comparisons show that classical image inversion approaches, such as reduced rank linear prediction, can provide more reliable reconstructions of the field than basis pursuit with a similar number of measurements in the presence of basis mismatch.

We note that although we present our results in the context of basis pursuit, our mathematical analysis is applicable to any sparse recovery principle that relies on the accuracy of best  $k$ -term approximations for performance guarantees. These include greedy recovery algorithms, such as Regularized Orthogonal Matching Pursuit (ROMP) [42] and Compressive Sampling Matching Pursuit (CoSaMP) [43]. The implication of our results is that, at least for problem sizes typical of speech processing, communication, spectrum analysis, and radar/sonar imaging, extra care may be needed to account for the effects of basis mismatch.

### C. Related Work

A parallel and equally fundamental study of perturbation effects in compressed sensing is presented in [44] and [45] (see also [46] and [47]), where the authors consider the following model mismatch problem. Suppose the low-dimensional compressed sensing measurements are made according to the model

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{b} \quad (3)$$

where  $\boldsymbol{\theta}$  is a  $k$ -sparse vector,  $\mathbf{A}$  is a compressed measurement matrix that produces the measurement vector  $\mathbf{y}$ , and  $\mathbf{b}$  is a noise vector. The sparse recovery algorithm, however, presumes the measurements  $\mathbf{y}$  are produced from a different model<sup>3</sup>

$$\mathbf{y} = \hat{\mathbf{A}}\mathbf{x} + \mathbf{b} \quad (4)$$

in which  $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{P}$  differs from the actual, but unknown, measurement matrix  $\mathbf{A}$  in a perturbation matrix  $\mathbf{P}$ . Under certain assumptions on the perturbation matrix  $\mathbf{P}$ , [44] and [45] show that the solution  $\mathbf{x}^*$  to

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad s.t. \quad \|\mathbf{y} - \hat{\mathbf{A}}\mathbf{x}\|_2 \leq \epsilon_{\mathbf{A},k,\mathbf{A}\boldsymbol{\theta}} \quad (5)$$

will satisfy

$$\|\mathbf{x}^* - \boldsymbol{\theta}\|_2 \leq \frac{C_0}{\sqrt{k}} \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 + C_1 \epsilon_{\mathbf{A},k,\mathbf{A}\boldsymbol{\theta}} \quad (6)$$

if the noise level  $\epsilon_{\mathbf{A},k,\mathbf{A}\boldsymbol{\theta}}$  is properly set. The proper value of  $\epsilon_{\mathbf{A},k,\mathbf{A}\boldsymbol{\theta}}$  depends on the restricted isometry constant (RIC) and the 2-norm of the true measurement matrix  $\mathbf{A}$ .  $C_0$  and  $C_1$  are constants for a given  $\mathbf{A}$ . We refer the reader to [44] and [45] for more details and simply focus on contrasting their analysis to ours in the context of basis mismatch.

The perturbed compressed sensing problem considered in [44] and [45] can reflect basis mismatch. In such a case, the actual measurement matrix  $\mathbf{A}$  may be viewed as the product  $\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}_1$  of a compressed sensing matrix  $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ , where  $M < N$ , and the unknown basis  $\boldsymbol{\Psi}_1$  in which the image  $\mathbf{s} = \boldsymbol{\Psi}_1\boldsymbol{\theta}$  is sparse. The measurement matrix  $\hat{\mathbf{A}}$  in the presumed

model (4) is related to  $\hat{\mathbf{A}}$  through the transformation  $\mathbf{A} = \hat{\mathbf{A}}\boldsymbol{\Psi}$ , where  $\boldsymbol{\Psi} = \boldsymbol{\Psi}_0^{-1}\boldsymbol{\Psi}_1$ . This can easily be verified by combining (2) and (4). Without loss of generality, we can express  $\boldsymbol{\Psi}$  as a perturbation term plus the identity matrix, in which case the additive perturbation matrix  $\mathbf{P}$  that relates  $\hat{\mathbf{A}}$  to  $\mathbf{A}$  reflects basis mismatch. Equation (6) then provides a comparison (in 2-norm of the difference) between the estimated coordinate vector  $\mathbf{x}^*$  in the basis  $\boldsymbol{\Psi}_0$  and the coordinate vector  $\boldsymbol{\theta}$  in the unknown basis  $\boldsymbol{\Psi}_1$ . If the mismatch between  $\boldsymbol{\Psi}_1$  and  $\boldsymbol{\Psi}_0$  is relatively modest (as defined in [44] and [45]) then the analysis of [44] and [45] shows that the coordinates  $\boldsymbol{\theta}$  in the unknown basis  $\boldsymbol{\Psi}_1$  can be estimated by  $\mathbf{x}^*$ , by solving the compressed sensing problem (5) in the nearby basis  $\boldsymbol{\Psi}_0$ .

Our analysis, on the other hand, compares the compressed sensing solution  $\mathbf{x}^*$  obtained using the *a priori* selected basis  $\boldsymbol{\Psi}_0$ , with the actual representation vector  $\mathbf{x}$  in  $\boldsymbol{\Psi}_0$ . We wish to investigate the consequences of assuming that  $\mathbf{s} = \boldsymbol{\Psi}_0\mathbf{x}$  is sparse in  $\boldsymbol{\Psi}_0$  on estimating  $\mathbf{x}$ , when in fact  $\mathbf{s}$  is sparse in the model  $\mathbf{s} = \boldsymbol{\Psi}_1\boldsymbol{\theta}$ . We consider an  $\ell_1$  minimization problem of the form (5), with  $\hat{\mathbf{A}} = \boldsymbol{\Phi}\boldsymbol{\Psi}_0$  and some noise level  $\epsilon$ , but unlike [44] and [45], we analyze  $\|\mathbf{x}^* - \mathbf{x}\|_p$ ,  $p = 1, 2$ , not  $\|\mathbf{x}^* - \boldsymbol{\theta}\|_p$ . We show that bounds on  $\|\mathbf{x}^* - \mathbf{x}\|_p$  grow linearly with the image dimension  $N$  and the degree of mismatch between  $\boldsymbol{\Psi}_0$  and  $\boldsymbol{\Psi}_1$ . This suggests that even a small mismatch may result in a large error in estimating  $\mathbf{x}$ . Our numerical examples demonstrate that this can indeed happen. The numerical examples presented in Section V-B for modal analysis adds another new aspect to our paper by providing comparisons between the true modal structure ( $\boldsymbol{\Psi}_1, \boldsymbol{\theta}$ ) and the modal structure ( $\boldsymbol{\Psi}_0, \mathbf{x}^*$ ) obtained from compressed sensing. Comparisons with conventional DFT processing and reduced-rank linear prediction are also provided. These examples demonstrate the effect of basis mismatch on determining *both* the modes and mode amplitudes, which cannot be observed from bounds on  $\|\mathbf{x}^* - \mathbf{x}\|_p$  or  $\|\mathbf{x}^* - \boldsymbol{\theta}\|_p$ .

Another relevant study is presented in [48]. The authors consider the problem of recovering a frequency sparse signal, i.e., a signal that can be expressed as a sparse linear combination of sinusoids, but the sinusoid frequencies are not necessarily aligned with DFT frequencies. However, the focus of [48] is on developing new algorithms for recovering frequency sparse signals rather than basis mismatch analysis. The assumption in [48] is that the signal is sparse in an oversampled DFT frame, which again gives rise to off-grid mismatch problems. However, the authors assume a so-called ‘‘inhibition model’’ for the signal to ensure that closely spaced columns from the oversampled DFT matrix are not simultaneously present in the signal representation. Under this assumption, they develop algorithms for recovering frequency sparse signals from compressed sensing measurements. These algorithms are obtained by integrating classical methods from parameter estimation (e.g., MUSIC) with model-based compressed sensing and greedy recovery algorithms. The developed methods outperform the compressed sensing algorithms that rely on signal recovery in the DFT basis.

## II. PROBLEM FORMULATION

There are two steps in compressed sensing, namely compressed recording and inversion (or recovery) for parameters

<sup>3</sup>In [44], [45], the models (3) and (4) also differ in the additive noise term  $\mathbf{b}$ , but since this is not important for contrasting our analysis to that of [44], [45] we take them to be the same here.

(see, e.g., [14]–[18]). In recording, we make linear measurements of  $\mathbf{s}$  as in (1), with possible additive noise  $\mathbf{b}$ , so the low-dimensional observation  $\mathbf{y}$  is *assumed* to be

$$\mathbf{y} = \Phi \mathbf{s} + \mathbf{b} = (\Phi \Psi_0) \mathbf{x} + \mathbf{b} \triangleq \mathbf{A} \mathbf{x} + \mathbf{b} \quad (7)$$

where  $\Phi \in \mathbb{C}^{M \times N}$  is the compressed sensing matrix (typically a matrix with i.i.d. Gaussian or i.i.d. Bernoulli entries),  $M$  is the number of measurements, and  $\Psi_0$  is the assumed basis that sparsely composes  $\mathbf{s}$  as  $\mathbf{s} = \Psi_0 \mathbf{x}$ . We define  $\mathbf{A} = \Phi \Psi_0 \in \mathbb{C}^{M \times N}$  as the measurement matrix. Without loss of generality, we need only deal with  $\mathbf{A}$  and  $\Psi$  in the following discussions.

In inversion, we seek the sparsest solution to  $\mathbf{x}$  given the observation  $\mathbf{y}$ . For the noise-free case ( $\mathbf{b} = \mathbf{0}$ ), we wish to solve

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A} \mathbf{x}. \quad (8)$$

This is in general an NP-hard problem. However, if the measurement matrix  $\mathbf{A}$  satisfies the so-called *Restricted Isometry Property* (RIP) [22], [26] the solution to (8) can be obtained using linear programming by solving the following  $\ell_1$  minimization problem:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A} \mathbf{x}. \quad (9)$$

This is referred to as *Basis Pursuit* (BP). In the noisy case, the problem is modified as

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A} \mathbf{x}\|_2 \leq \epsilon \quad (10)$$

where  $\|\mathbf{b}\|_2 \leq \epsilon$  is the bounded noise.

From [22], [23], [26], when the RIC of the measurement matrix  $\mathbf{A}$  satisfies  $\delta_{2k}^{\mathbf{A}} < \sqrt{2} - 1$ , that is, for any  $2k$ -sparse vector  $\mathbf{u}$ ,

$$(1 - \delta_{2k}^{\mathbf{A}}) \|\mathbf{u}\|_2^2 \leq \|\mathbf{A} \mathbf{u}\|_2^2 \leq (1 + \delta_{2k}^{\mathbf{A}}) \|\mathbf{u}\|_2^2, \quad (11)$$

the solution  $\mathbf{x}^*$  to (9) approximates  $\mathbf{x}$  as

$$\|\mathbf{x} - \mathbf{x}^*\|_1 \leq C_0 \|\mathbf{x} - \mathbf{x}_k\|_1 \quad (12)$$

and

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq C_0 k^{-1/2} \|\mathbf{x} - \mathbf{x}_k\|_1 \quad (13)$$

where  $\mathbf{x}_k$  is the best  $k$ -term approximation of  $\mathbf{x}$ . That is,

$$\mathbf{x}_k = \arg \min_{\mathbf{z} \in \Sigma_k} \|\mathbf{x} - \mathbf{z}\|_1$$

where  $\Sigma_k = \{\mathbf{z} \in \mathbb{C}^N : |T_{\mathbf{z}}| \leq k\}$  and  $T_{\mathbf{z}}$  is the support of  $\mathbf{z}$ . For the noisy case (10), we have [23], [26]

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq C_0 k^{-1/2} \|\mathbf{x} - \mathbf{x}_k\|_1 + C_1 \epsilon. \quad (14)$$

The constants  $C_0$  and  $C_1$  are given by

$$C_0 = \frac{2(1 + \alpha_0)}{1 - \alpha_0} \quad \text{and} \quad C_1 = \frac{2\alpha_1}{1 - \alpha_0} \quad (15)$$

where

$$\alpha_0 = \frac{\sqrt{2}\delta_{2k}^{\mathbf{A}}}{1 - \delta_{2k}^{\mathbf{A}}} \quad \text{and} \quad \alpha_1 = \frac{2\sqrt{1 + \delta_{2k}^{\mathbf{A}}}}{1 - \delta_{2k}^{\mathbf{A}}}. \quad (16)$$

In the matched case where the hypothesized basis  $\Psi_0$  coincides with the actual basis  $\Psi_1$ , the mismatched basis  $\Psi = \Psi_0^{-1} \Psi_1$  reduces to  $\mathbf{I}$  and  $\mathbf{x} = \boldsymbol{\theta}$  is sparse in the identity basis  $\mathbf{I}$ . Then, for  $k$ -sparse  $\boldsymbol{\theta}$  the bound  $\|\mathbf{x} - \mathbf{x}_k\|_1$  is zero in (12) and the solution  $\mathbf{x}^*$  is an exact recovery of  $\mathbf{x}$  in the noise-free case.

However, in the mismatched case where  $\Psi_0 \neq \Psi_1$ ,  $\mathbf{x} = \Psi \boldsymbol{\theta}$  is actually sparse in the  $\Psi$  basis, rather than the  $\mathbf{I}$  basis. The question is, “what is the consequence of minimizing  $\|\mathbf{x}\|_1$  under the constraint  $\mathbf{y} = \mathbf{A} \mathbf{x}$ , when in fact the correct problem is to minimize  $\|\boldsymbol{\theta}\|_1$  under the constraint  $\mathbf{y} = \mathbf{A} \Psi \boldsymbol{\theta} = \Phi \Psi_1 \boldsymbol{\theta}$ ?”

*Remark 1:* ROMP [42] and CoSaMP [43] have similar universal performance bounds as BP. For example, under the RIP of the measurement matrix, ROMP can approximately recover any compressible signal from noisy observations. Given noisy observations  $\mathbf{y} = \mathbf{A} \mathbf{x} + \mathbf{b}$  with  $\|\mathbf{b}\|_2 \leq \epsilon$ , ROMP produces a  $2k$ -sparse signal approximation  $\mathbf{x}^*$  that satisfies the bound

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq C \sqrt{\log 2k} \left( \frac{1}{\sqrt{k}} \|\mathbf{x} - \mathbf{x}_k\|_1 + \epsilon \right). \quad (17)$$

where  $C = 160$  [42]. Similarly, for a given precision parameter  $\zeta$ , CoSaMP produces a  $2k$ -sparse signal approximation  $\mathbf{x}^*$  that satisfies

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq C \max \left\{ \zeta, \frac{1}{\sqrt{2k}} \|\mathbf{x} - \mathbf{x}_k\|_1 + \epsilon \right\}. \quad (18)$$

where  $C = 3.42$  [43]. Therefore, our analysis of the best  $k$ -term approximation error is also relevant to these algorithms.

### III. MAIN RESULTS

#### A. Degeneration of the Best $k$ -Term Approximation

Let us express the mismatched basis  $\Psi = \Psi_0^{-1} \Psi_1 \in \mathbb{C}^{N \times N}$  as  $\Psi = \mathbf{I} + \mathbf{E}$ , where  $\mathbf{E}$  is a perturbation matrix with respect to the identity basis and  $\mathbf{x} = \Psi \boldsymbol{\theta}$ . Let  $\mathbf{x}_k$  and  $\boldsymbol{\theta}_k$  denote the best  $k$ -term approximations to  $\mathbf{x}$  and  $\boldsymbol{\theta}$ , respectively. We have the following theorem.

*Theorem 1 (Best  $k$ -Term Approximation Error):* Let  $\Psi = \mathbf{I} + \mathbf{E}$  and  $\mathbf{x} = \Psi \boldsymbol{\theta}$ . Let  $1 \leq p, q \leq \infty$  and  $1/p + 1/q = 1$ . If the rows  $\mathbf{e}_m^T \in \mathbb{C}^{1 \times N}$  of  $\mathbf{E}$  are bounded as  $\|\mathbf{e}_m\|_p \leq \beta$  for  $1 \leq m \leq N$ , then

$$\|\|\mathbf{x} - \mathbf{x}_k\|_1 - \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1\| \leq (N - k)\beta \|\boldsymbol{\theta}\|_q. \quad (19)$$

The bound is achieved when the entries of  $\mathbf{E}$  satisfy

$$e_{mn} = \pm \beta \cdot e^{j(\arg(\theta_m) - \arg(\theta_n))} \cdot (|\theta_n| / \|\boldsymbol{\theta}\|_q)^{q/p} \quad (20)$$

for  $n \in T_{\boldsymbol{\theta}}$  and  $1 \leq m \leq N$ .

*Proof:* See Appendix A.

When  $\boldsymbol{\theta}$  is  $k$ -sparse in  $\mathbf{I}$ , i.e.,  $\boldsymbol{\theta} = \boldsymbol{\theta}_k$ , then (19) reduces to

$$\|\mathbf{x} - \mathbf{x}_k\|_1 \leq (N - k)\beta \|\boldsymbol{\theta}\|_q, \quad (21)$$

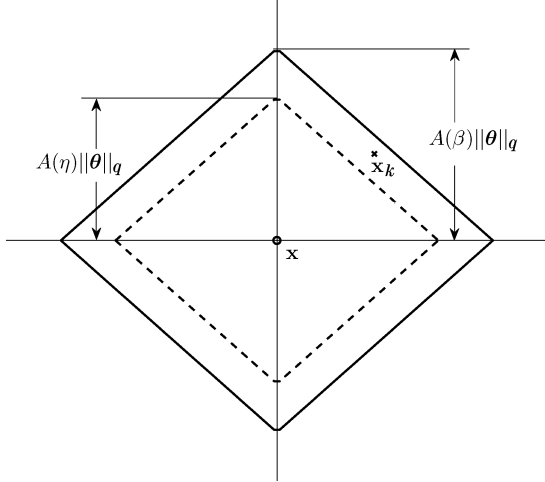


Fig. 1. Confidence  $\ell_1$  balls for  $\mathbf{x}_k$  when  $\eta \leq \|\mathbf{e}_m\|_p \leq \beta$  and  $\boldsymbol{\theta}$  is  $k$ -sparse.

which shows that the upper bound on the  $\ell_1$  error  $\|\mathbf{x} - \mathbf{x}_k\|_1$  is linearly increasing in  $N$ ,  $\beta$  and  $\|\boldsymbol{\theta}\|_q$ , and linearly decreasing in  $k$ .

*Remark 2:* When  $p = \infty$  and  $q = 1$ , (19) corresponds to the case where the entries themselves of the perturbation matrix  $\mathbf{E}$  are upper bounded, that is, when  $|e_{mn}| \leq \beta$  for  $n \in T_\theta$ . When  $p = q = 2$ , (19) corresponds to the case where the row 2-norms of  $\mathbf{E}$  are bounded as  $\|\mathbf{e}_m\|_2 \leq \beta$ .

*Remark 3:* Theorem 1 can be easily generalized to have  $\ell_2$  norms on the left-hand side (LHS) of (19), i.e.

$$\|\|\mathbf{x} - \mathbf{x}_k\|_2 - \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_2\| \leq (N - k)^{1/2} \beta \|\boldsymbol{\theta}\|_q. \quad (22)$$

This follows by taking similar steps as in the proof of Theorem 1, but using the Minkowski inequality instead of the triangle inequality.

*Corollary 1 (Normalized  $\ell_1$  Error):* Let  $\boldsymbol{\Psi} = \mathbf{I} + \mathbf{E}$  and  $\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\theta}$ . Let  $1 \leq p, q \leq \infty$  and  $1/p + 1/q = 1$ . If the rows  $\mathbf{e}_m^T \in \mathbb{C}^{1 \times N}$  of  $\mathbf{E}$  are bounded as  $\|\mathbf{e}_m\|_p \leq \beta$  for  $1 \leq m \leq N$ , then the  $\ell_1$  norm of  $\mathbf{x}$  is bounded as

$$(1 - N\beta)\|\boldsymbol{\theta}\|_1 \leq \|\mathbf{x}\|_1 \leq (1 + N\beta)\|\boldsymbol{\theta}\|_1 \quad (23)$$

and the normalized  $\ell_1$  error for approximating  $\mathbf{x}$  by  $\mathbf{x}_k$  is bounded as

$$\begin{aligned} & \frac{1}{1 + N\beta} \frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1}{\|\boldsymbol{\theta}\|_1} - \frac{(N - k)\beta}{1 + N\beta} \\ & \leq \frac{\|\mathbf{x} - \mathbf{x}_k\|_1}{\|\mathbf{x}\|_1} \\ & \leq \frac{1}{1 - N\beta} \frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1}{\|\boldsymbol{\theta}\|_1} + \frac{(N - k)\beta}{1 - N\beta} \end{aligned} \quad (24)$$

where the upper bound in (24) is valid if  $N\beta < 1$ .

*Proof:* The inequalities in (23) follow from Theorem 1 by setting  $k = 0$  and  $q = 1$ , and (24) follows by combining (19) and (23).

*Remark 4:* In Theorem 1, we have characterized the mismatch between the assumed basis  $\boldsymbol{\Psi}_0$  and the actual basis  $\boldsymbol{\Psi}_1$  by considering bounds on the  $p$ -norm of the rows of perturbation matrix  $\mathbf{E} = \boldsymbol{\Psi}_0^{-1}(\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_0)$  that captures the deviation of  $\boldsymbol{\Psi} = \boldsymbol{\Psi}_0^{-1}\boldsymbol{\Psi}_1$  from the identity matrix  $\mathbf{I}$ . But we may think of

the incoherence  $\mu(\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1)$  between  $\boldsymbol{\Psi}_0$  and  $\boldsymbol{\Psi}_1$  as another way of characterizing the degree of mismatch between  $\boldsymbol{\Psi}_0$  and  $\boldsymbol{\Psi}_1$ . And in fact  $\mu(\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1)$  and the “max norm” of  $\|\mathbf{E}\|$  are related, as we now show. Let  $\boldsymbol{\psi}_i(j)$  denote the  $j$ th column of the basis  $\boldsymbol{\Psi}_i$ ,  $i = 0, 1$ . The incoherence between the bases  $\boldsymbol{\Psi}_0$  and  $\boldsymbol{\Psi}_1$  is defined as

$$\mu(\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1) = \max_{i,j} |\langle \boldsymbol{\psi}_0(i), \boldsymbol{\psi}_1(j) \rangle|. \quad (25)$$

Let  $\|\mathbf{M}\|_{\max} = \max_{i,j} |M_{i,j}|$  denote the max-norm of a matrix  $\mathbf{M}$ . Then, we can express  $\mu(\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1)$  as

$$\begin{aligned} \mu(\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1) &= \left\| \boldsymbol{\Psi}_0^H \boldsymbol{\Psi}_1 \right\|_{\max} \\ &= \left\| \boldsymbol{\Psi}_0^H \boldsymbol{\Psi}_0 \boldsymbol{\Psi}_0^{-1} \boldsymbol{\Psi}_1 \right\|_{\max} \\ &\leq \left\| \boldsymbol{\Psi}_0^H \boldsymbol{\Psi}_0 \right\|_{\max} \|\boldsymbol{\Psi}\|_{\max} \end{aligned} \quad (26)$$

where  $\boldsymbol{\Psi} = \boldsymbol{\Psi}_0^{-1}\boldsymbol{\Psi}_1$ . When  $\boldsymbol{\Psi}_0$  is unitary, we have  $\mu(\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1) = \|\boldsymbol{\Psi}\|_{\max}$ . Substituting  $\boldsymbol{\Psi} = \mathbf{I} + \mathbf{E}$  and applying the triangular inequality gives

$$\mu(\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1) \leq \left\| \boldsymbol{\Psi}_0^H \boldsymbol{\Psi}_0 \right\|_{\max} (1 + \|\mathbf{E}\|_{\max}). \quad (27)$$

Theorem 1 establishes an upper bound for the  $\ell_1$  error  $\|\mathbf{x} - \mathbf{x}_k\|_1$  in the best  $k$ -term approximation of  $\mathbf{x}$  for the case where the degree of mismatch is upper bounded. The following theorem considers the case where the level of mismatch between the bases  $\boldsymbol{\Psi}_0$  and  $\boldsymbol{\Psi}_1$  is lower bounded and establishes a lower bound for the worst-case  $\ell_1$  error  $\|\mathbf{x} - \mathbf{x}_k\|_1$ .

*Theorem 2 (Worst-Case Best  $k$ -Term Approximation Error):* Let  $\boldsymbol{\Psi} = \mathbf{I} + \mathbf{E}$  and  $\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\theta}$ . Let  $1 \leq p, q \leq \infty$  and  $1/p + 1/q = 1$ . If the rows of  $\mathbf{E}$  are lower bounded as  $\|\mathbf{e}_m\|_p \geq \eta$ , then

$$\max_{\mathbf{E}: \|\mathbf{e}_m\|_p \geq \eta} \|\mathbf{x} - \mathbf{x}_k\|_1 \geq \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 + (N - k)\eta\|\boldsymbol{\theta}\|_q. \quad (28)$$

*Proof:* See Appendix B.

Theorem 1 shows that the  $\ell_1$  norm  $\|\mathbf{x} - \mathbf{x}_k\|_1$  of the best- $k$  term approximation error will be no worse than the upper bound deduced from (19) if the mismatch level between  $\boldsymbol{\Psi}_0$  and  $\boldsymbol{\Psi}_1$  (measured by  $\|\mathbf{e}_m\|_p$ ) is upper bounded by  $\beta$ . In contrast, Theorem 2 shows that if the mismatch level is lower bounded by  $\eta$  then there exists a mismatch scenario where  $\|\mathbf{x} - \mathbf{x}_k\|_1$  is no better than the bound in (28). Both bounds grow linearly with the grid size  $N$  and the mismatch levels. Fig. 1 illustrates the interplay between Theorems 1 and 2 when  $\boldsymbol{\theta}$  is  $k$ -sparse, i.e., when  $\|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 = 0$ . The outer ball has radius  $A(\beta)\|\boldsymbol{\theta}\|_q$ , with  $A(\beta) = (N - k)\beta$  and  $q \geq 1$ , and corresponds to the upper bound in Theorem 1. The inner ball, with radius  $A(\eta)\|\boldsymbol{\theta}\|_q$ , corresponds to Theorem 2. The figure illustrates that there always exists a mismatched basis  $\boldsymbol{\Psi} = \mathbf{I} + \mathbf{E}$ , with mismatch level  $\eta \leq \|\mathbf{e}_m\|_p \leq \beta$ , for which the best  $k$ -term approximation  $\mathbf{x}_k$  lies between the two balls, away from zero, and thus no guarantee can be provided for the performance of basis pursuit.

## B. Confidence Bounds for Image Inversion

We now derive bounds for the  $\ell_1$  and  $\ell_2$  norms of the image inversion error  $\mathbf{x} - \mathbf{x}^*$ . Here we invert the image  $\mathbf{s} = \boldsymbol{\Psi}_0\mathbf{x}$  for its field parameter vector  $\mathbf{x}$ , using basis pursuit, under the

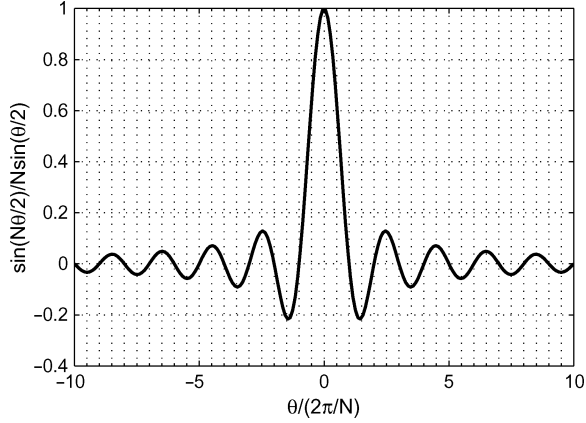


Fig. 2. The function  $(1/N)(\sin(N\theta/2)/\sin(\theta/2))$  versus  $\theta/(2\pi/N)$ .

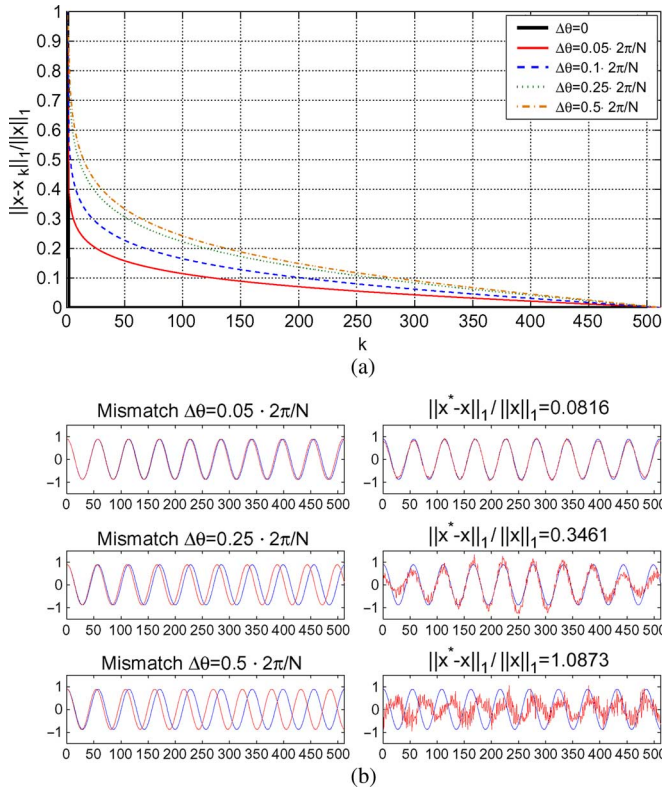


Fig. 3. (a)  $\|\mathbf{x} - \mathbf{x}_k\|_1 / \|\mathbf{x}\|_1$  versus  $\Delta\theta/(2\pi/N)$  for various  $k$ . (b) Left column: the actual tone (blue) superimposed on the closest DFT tone; Right column: the reconstructed tone (red) superimposed on the actual tone (blue). The frequency mismatch  $\Delta\theta/(2\pi/N)$  is 0.05 for the plots in the top row, 0.25 for the plots in the middle row, and 0.5 for the plots in the bottom row.

assumption that  $\mathbf{s}$  has a sparse representation in  $\Psi_0$ , when in fact this representation is non-sparse or incompressible. These bounds can then be used to find bounds on the  $\ell_1$  and  $\ell_2$  norms of the image reconstruction error  $\mathbf{s} - \mathbf{s}^*$ , which speaks to how well the image  $\mathbf{s}$  can be recomposed in the basis  $\Psi_0$  using the mismatched basis pursuit approximation  $\mathbf{x}^*$ .

**Theorem 3 (Image Inversion Error):** Let  $\mathbf{A}$  be fixed and satisfy  $\delta_{2k}^{\mathbf{A}} < \sqrt{2}-1$  at the assumed sparsity level  $k$ . Let  $1 \leq p, q \leq \infty$  and  $1/p + 1/q = 1$ . If the rows of  $\mathbf{E}$  satisfy  $\|\mathbf{e}_m\|_p \leq \beta$ ,

then the image inversion error  $\|\mathbf{x} - \mathbf{x}^*\|_1$  is bounded as

$$\|\mathbf{x} - \mathbf{x}^*\|_1 \leq C_0 \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 + C_0(N-k)\beta \|\boldsymbol{\theta}\|_q. \quad (29)$$

For noisy recovery (10), with  $\|\mathbf{b}\|_2 \leq \epsilon$ , we have

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq C_0 k^{-1/2} \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 + C_0(N-k)k^{-1/2}\beta \|\boldsymbol{\theta}\|_q + C_1\epsilon \quad (30)$$

where  $C_0$  and  $C_1$  are given in (15).

*Proof:* This follows easily by combining Theorem 1 and the performance bounds of basis pursuit.

**Remark 5:** Theorem 3 is a direct consequence of Theorem 1 and therefore the discussion about the connection between the max-norm of the perturbation matrix  $\mathbf{E}$  and the incoherence measure  $\mu(\Psi_0, \Psi_1)$  also applies here.

**Remark 6:** The  $\ell_1$  and  $\ell_2$  norms of the image reconstruction error  $\mathbf{s} - \mathbf{s}^*$  are bounded as

$$\|\mathbf{s} - \mathbf{s}^*\|_{1,2} = \|\Psi_0 \mathbf{x} - \Psi_0 \mathbf{x}^*\|_{1,2} \leq \|\Psi_0\|_{1,2} \|\mathbf{x} - \mathbf{x}^*\|_{1,2}. \quad (31)$$

A simple application of Theorem 3 will then yield an upper bound for  $\|\mathbf{s} - \mathbf{s}^*\|_{1,2}$  in terms of  $N$ ,  $k$ ,  $\beta$ , and  $\|\boldsymbol{\theta}\|_q$ .

Theorem 3 shows that the upper bound on  $\|\mathbf{x} - \mathbf{x}^*\|_1$  is linearly increasing in the image dimension  $N$  and in the mismatch level  $\beta$ , and linearly decreasing in  $k$ . The upper bound on the  $\ell_1$  error in image reconstruction has similar growth behavior following Remark 6. The maximum allowable value for  $k$ , for which the basis pursuit bounds still hold, depends on the measurement matrix  $\mathbf{A}$ . For  $\mathbf{A} \in \mathbb{R}^{M \times N}$  to be RIP with constant  $\delta_{2k}^{\mathbf{A}} < \sqrt{2} - 1$  we typically require  $M = \mathcal{O}(k \log(N/k))$ .

#### IV. FOURIER IMAGING AND DFT GRID MISMATCH

A mismatch case of particular interest arises in Fourier imaging when a sparse signal with arbitrary frequency components is taken to be sparse in a DFT basis. Our objective in this section is to highlight the particularly problematic nature of basis mismatch in this application.

Suppose the sparsity basis  $\Psi_0$  in the mathematical model  $\mathbf{s} = \Psi_0 \mathbf{x}$ , assumed by the compressed sensing procedure, is the unitary  $N$ -point DFT basis. Then the  $\ell$ th column of  $\Psi_0$  is a Vandermonde vector of the form  $\boldsymbol{\psi}_{0,\ell} = [1, e^{j2\pi\ell/N}, \dots, e^{j2\pi\ell(N-1)/N}]^T$  and the basis  $\Psi_0$  is

$$\Psi_0 = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & e^{j2\pi/N} & \dots & e^{j2\pi(N-1)/N} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j2\pi(N-1)/N} & \dots & e^{j2\pi(N-1)^2/N} \end{bmatrix}. \quad (32)$$

Without loss of generality, let us assume that the  $\ell$ th column  $\boldsymbol{\psi}_{1,\ell}$  of the actual sparsity basis  $\Psi_1$  is mismatched to the  $\ell$ th column of  $\Psi_0$  by  $\Delta\theta_\ell$  (in normalized) frequency, where  $0 \leq \Delta\theta_\ell < (2\pi/N)$ . Then,  $\Psi_1$  is given by (33), shown at the bottom of the next page.

The mismatched basis  $\Psi = \Psi_0^{-1} \Psi_1$  can be written as (34), shown at the bottom of the next page, where  $L(\theta)$  is the Dirichlet

kernel given by

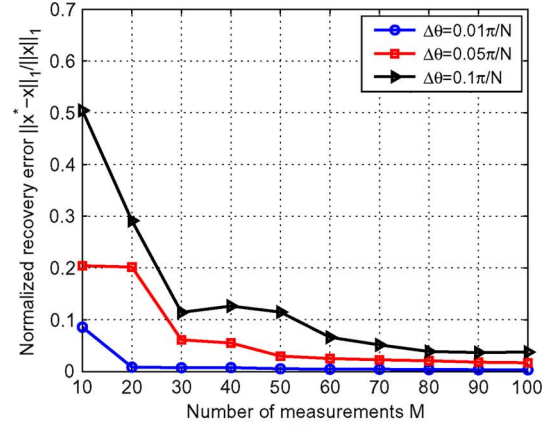
$$L(\theta) = \frac{1}{N} \sum_{n=0}^{N-1} e^{jn\theta} = \frac{1}{N} e^{j\frac{\theta(N-1)}{2}} \frac{\sin(\theta N/2)}{\sin(\theta/2)}. \quad (35)$$

where  $L(0) = 1$ . The  $(m, \ell)$ th element of the mismatched basis  $\Psi$  is a sample of the Dirichlet kernel  $L(\theta)$  at  $\theta = \Delta\theta_\ell - (2\pi/N)(m - \ell)$ , where  $m = 0, 1, \dots, N - 1$  and  $\ell = 0, 1, \dots, N - 1$ .

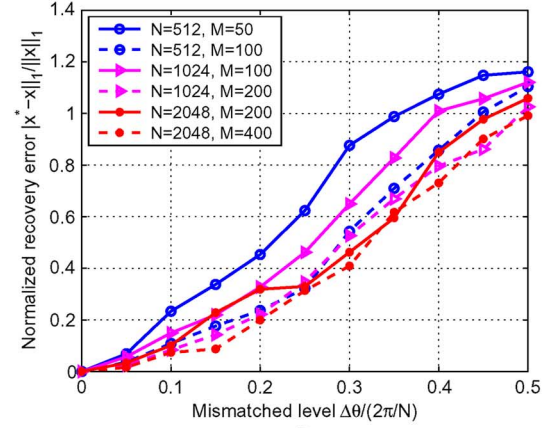
The Dirichlet kernel  $L(\theta)$ , shown in Fig. 2 (ignoring the unimodular phasing term) for  $N = 64$ , decays slowly as  $|L(\theta)| \leq (N\theta/2\pi)^{-1}$  for  $|\theta| \leq \pi$ , with  $L(0) = 1$ . This decay behavior follows from the fact that  $|\sin(\theta/2)| \geq 2|\theta/2\pi|$  for  $|\theta| \leq \pi$ , where the equality holds when  $|\theta| = \pi$ . This means that  $(N\theta/2\pi)^{-1}$  is in fact the envelope of  $|L(\theta)|$ . Therefore, every mismatch between a physical frequency  $\theta_\ell$  and the corresponding DFT frequency  $2\pi\ell/N$  produces a column in the mismatched basis  $\Psi$  for which the entries vanish slowly as each column is traversed. The consequence of this is that the parameter vector  $\mathbf{x}$  in the mathematical model  $\mathbf{s} = \Psi_0\mathbf{x}$ , for which the compressed sensing procedure is seeking a sparse solution, is in fact incompressible. This follows from the coordinate transformation  $\mathbf{x} = \Psi\boldsymbol{\theta}$  between the true sparse parameter vector  $\boldsymbol{\theta}$  and the presumed sparse vector  $\mathbf{x}$ . The few nonzero elements of  $\boldsymbol{\theta}$  leak in to all locations through  $\Psi$  and the slow decay of elements of  $\Psi$  makes  $\mathbf{x}$  incompressible.

*Remark 7:* Even if only a single physical frequency, say  $\theta_\ell$ , is not on the DFT grid the matrix  $\Psi$  will have a column whose entries vanish slowly when traversed and the mismatch problem remains.

*Remark 8:* In addition, to frequency mismatch, the columns of  $\Psi_1$  may also be mismatched to those of  $\Psi_0$  by damping factors  $\lambda_\ell \leq 0$ . In such a case, the basis  $\Psi_1$  is given by (36), shown at the bottom of the page. The  $(m, \ell)$ th element of the mismatch



(a)



(b)

Fig. 4. (a) Normalized image inversion error  $\|\mathbf{x} - \mathbf{x}^*\|_1 / \|\mathbf{x}\|_1$  versus the number of measurements  $M$  for different frequency mismatch levels  $\Delta\theta = 0.005 \cdot 2\pi/N$ ,  $0.05 \cdot 2\pi/N$  and  $0.1 \cdot 2\pi/N$ , with  $N = 512$ . (b)  $\|\mathbf{x} - \mathbf{x}^*\|_1 / \|\mathbf{x}\|_1$  versus the frequency mismatch level  $\Delta\theta / (2\pi/N)$  for different  $(M, N)$  pairs.

basis  $\Psi$  is

$$\psi_{m,\ell} = \frac{1}{N} \sum_{n=0}^{N-1} e^{n[\lambda_\ell + j(\Delta\theta_\ell - \frac{2\pi(m-\ell)}{N})]}. \quad (37)$$

$$\Psi_1 = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\Delta\theta_0} & e^{j(\frac{2\pi}{N} + \Delta\theta_1)} & \cdots & e^{j(\frac{2\pi(N-1)}{N} + \Delta\theta_{N-1})} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j\Delta\theta_0(N-1)} & e^{j(\frac{2\pi}{N} + \Delta\theta_1)(N-1)} & \cdots & e^{j(\frac{2\pi(N-1)}{N} + \Delta\theta_{N-1})(N-1)} \end{bmatrix}. \quad (38)$$

$$\Psi = \Psi_0^{-1} \Psi_1 = \begin{bmatrix} L(\Delta\theta_0 - 0) & L(\Delta\theta_1 - \frac{2\pi(N-1)}{N}) & \cdots & L(\Delta\theta_{N-1} - \frac{2\pi}{N}) \\ L(\Delta\theta_0 - \frac{2\pi}{N}) & L(\Delta\theta_1 - 0) & \cdots & L(\Delta\theta_{N-1} - \frac{2\pi(N-2)}{N}) \\ \vdots & \vdots & \ddots & \vdots \\ L(\Delta\theta_0 - \frac{2\pi(N-1)}{N}) & L(\Delta\theta_1 - \frac{2\pi(N-2)}{N}) & \cdots & L(\Delta\theta_{N-1} - 0) \end{bmatrix} \triangleq \mathbf{I} + \mathbf{E} \quad (39)$$

$$\Psi_1 = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{[\lambda_0 + j\Delta\theta_0]} & e^{[\lambda_1 + j(\frac{2\pi}{N} + \Delta\theta_1)]} & \cdots & e^{[\lambda_{N-1} + j(\frac{2\pi(N-1)}{N} + \Delta\theta_{N-1})]} \\ \vdots & \vdots & \ddots & \vdots \\ e^{[\lambda_0 + j\Delta\theta_0](N-1)} & e^{[\lambda_1 + j(\frac{2\pi}{N} + \Delta\theta_1)](N-1)} & \cdots & e^{[\lambda_{N-1} + j(\frac{2\pi(N-1)}{N} + \Delta\theta_{N-1})](N-1)} \end{bmatrix}. \quad (40)$$

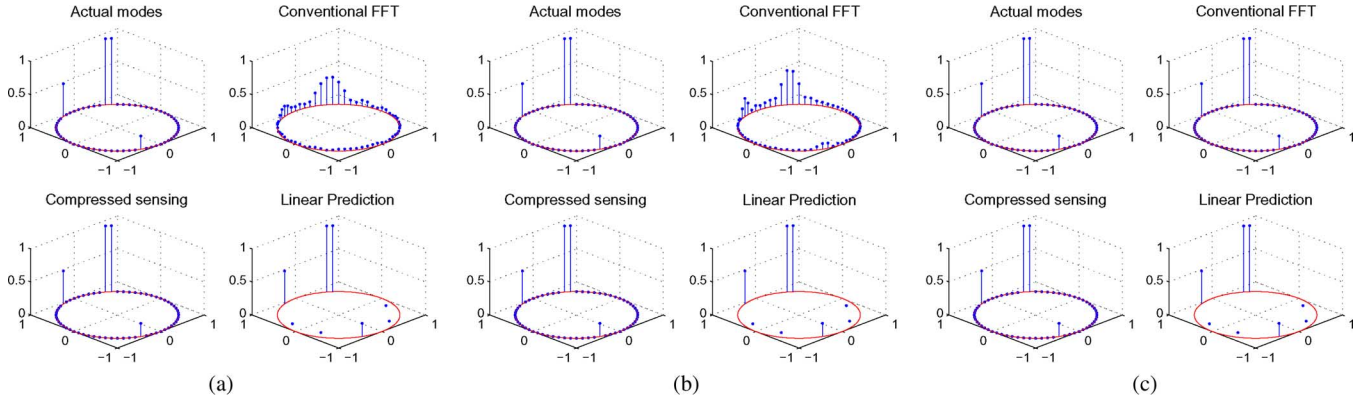


Fig. 5. Comparison of DFT, CS, and LP inversions in the absence of basis mismatch (a)  $M = N/4 = 16$ . (b)  $M = N/2 = 32$ . (c)  $M = N/1 = 64$ .

In general, the basis mismatch problem exists in almost all applications and is not limited to Fourier imaging. However, we emphasize Fourier imaging in this paper as a ubiquitous imaging problem, where basis mismatch seems to have a particularly destructive effect.

## V. NUMERICAL EXAMPLES

We now present three simple numerical examples to demonstrate the effect of basis mismatch on the performance of compressed sensing. The first example considers single tone mismatch to the DFT grid, the second considers both frequency and damping mismatch, and the third example considers a synthetic mismatch where the bound established in Theorem 1 is achieved.

### A. Example 1: Tone Reconstruction With Basis Pursuit

Here we assume that the image  $\mathbf{s}$  is a single tone (a 1-sparse signal) with an unknown frequency. The actual sparsity basis  $\Psi_1$  is a matrix, in which the  $\ell$ th column is mismatched in frequency by  $0 \leq \Delta\theta < (2\pi/N)$  with respect to the  $\ell$ th column of the  $N$ -point DFT matrix, which the compressed sensing procedure takes as the sparsity basis  $\Psi_0$ . The rest of the columns of  $\Psi_1$  and  $\Psi_0$  are assumed to be identical. Damping mismatch is not considered in this example.

Fig. 3(a) shows the normalized  $\ell_1$  error  $\|\mathbf{x} - \mathbf{x}_k\|_1 / \|\mathbf{x}\|_1$  for different frequency mismatch levels  $\Delta\theta$  and different values of  $k$ , when the single tone is located off the ( $\ell = 10$ )th DFT frequency in the ( $N = 512$ )-point DFT grid. The plot shows that the normalized best  $k$ -term approximation error is considerable even at moderate mismatch levels, and speaks to the fact that, even for a slight mismatch with respect to the DFT mode, the presumably sparse representation is in fact incompressible due to the slow decay of the Dirichlet kernel.

Fig. 3(b) compares the reconstructed tone to the actual tone for different mismatch levels. In the left column, the blue plots show the actual tone and the red plots show the closest tone on the DFT grid to the actual tone, for the corresponding mismatch level. On the right-hand side (RHS), the red plots show the reconstructed tone  $\mathbf{s}^* = \Psi_0 \mathbf{x}^*$ , where  $\Psi_0$  is the DFT basis, and the blue plots again show the actual tone. The number of compressed sensing measurements used for reconstruction is

$M = 64$ . The frequency mismatch  $\Delta\theta / (2\pi/N)$  is 0.05 for the plots in the top row, 0.25 for the plots in the middle row, and 0.5 for the plots in the bottom row. The inaccuracy in reconstruction is noticeable for 25% (i.e.,  $\Delta\theta = 0.25(2\pi/N)$ ) and 50% frequency mismatch.

Fig. 4(a) shows how the normalized image inversion error  $\|\mathbf{x} - \mathbf{x}^*\|_1 / \|\mathbf{x}\|_1$  varies with the number of measurements  $M$  in  $N = 512$  dimensions for different mismatch levels  $\Delta\theta = 0.005 \cdot 2\pi/N$ ,  $0.05 \cdot 2\pi/N$  and  $0.1 \cdot 2\pi/N$ . The measurement matrix  $\mathbf{A}$  is generated with random Gaussian entries satisfying RIP conditions, and the experiment takes 50 runs and returns the average error. We observe that without increasing the number of measurements beyond what is required in the mismatch-free case the normalized error can be large, even for a small frequency mismatch  $\Delta\theta = 0.005 \cdot 2\pi/N$ . This threshold is approximately  $3k \log N/k = 27$  in this example. For measurement dimensions smaller than 27 the normalized error is large, but it decreases relatively fast as the number of measurements increases. However, after this threshold the normalized error decays slowly and only when the number of measurements is increased to  $N$  the error goes to zero. Fig. 4(b) shows the normalized image inversion error  $\|\mathbf{x} - \mathbf{x}^*\|_1 / \|\mathbf{x}\|_1$  versus the frequency mismatch level  $\Delta\theta / (2\pi/N)$  for several  $(M, N)$  pairs, where  $M$  is the number of compressed sensing measurements and  $N$  is the grid dimension.

### B. Example 2: Modal Analysis

We now give examples to demonstrate the effect of DFT grid mismatch on modal analysis based on compressed sensing measurements and compare the results with those obtained using classical image inversion principles, namely standard DFT imaging and linear prediction (LP). The reader is referred to [2], [9], [10] for a description of linear prediction. In all the experiments, the dimension of the image/signal is  $N = 64$ . The number of recordings or measurements  $M$  used for inversion is the same for all methods and we report results for  $M = N/4 = 16$  to  $M = N/2 = 32$  to  $M = N/1 = 64$ .

We first consider the case where the field we wish to invert for contains only modes that are aligned with the DFT frequencies. This is to demonstrate that CS and LP both provide perfect field recovery when there is no mismatch, as shown in Fig. 5(a)–(c). In each subfigure (a) through (c) there are four

panels. In the top-left panel the true underlying modes are illustrated with stems whose locations on the unit disc indicate the frequencies of the modes, and whose heights illustrate the mode amplitudes. The phases of the modes are randomly chosen, and not indicated on the figures. The frequencies at which modes are placed, and their amplitudes, are  $(2\pi \cdot 9/N, 1)$ ,  $(2\pi \cdot 10/N, 1)$ ,  $(2\pi \cdot 20/N, .5)$ , and  $(2\pi \cdot 45/N, .2)$ . These frequencies are perfectly aligned with the DFT frequencies. No noise is considered for now. We observe that both CS and LP provide perfect recovery. The DFT processing however has leakage according to the Dirichlet kernel unless the measurement dimension is increased to the full dimension  $N = 64$ . This was of course expected.

But what is the connection between the circular plots in Fig. 5(a)–(c) and the models  $\Psi_0 \mathbf{x}$  and  $\Psi_1 \boldsymbol{\theta}$ ? The top-left panel (actual modes) in each subplot is an “illustration” of  $(\Psi_1, \boldsymbol{\theta})$ , with the locations of the bars on the unit disc corresponding to active modes from  $\Psi_1$  and the heights of the bars corresponding to the values of nonzero entries in  $\boldsymbol{\theta}$ . The top-right panel (conventional FFT) illustrates  $(\Psi_0, \hat{\mathbf{x}})$ , where  $\hat{\mathbf{x}}$  is the estimate of  $\mathbf{x}$  obtained by DFT processing the measurement vector  $\mathbf{y}$ . The bottom-left (compressed sensing) illustrates  $(\Psi_0, \mathbf{x}^*)$ , where  $\mathbf{x}^*$  is the solution to (9) (or to (10) in the noisy cases to follow). The bottom-right panel (LP) illustrates  $(\hat{\Psi}_1, \hat{\boldsymbol{\theta}})$ , where  $\hat{\Psi}_1$  and  $\hat{\boldsymbol{\theta}}$  are, respectively, estimates of  $\Psi_1$  and  $\boldsymbol{\theta}$  obtained by LP (or reduced-rank LP). When there is no mismatch and noise,  $(\Psi_0, \mathbf{x}^*)$  and  $(\hat{\Psi}_1, \hat{\boldsymbol{\theta}})$  match  $(\Psi_1, \boldsymbol{\theta})$ . However, they significantly differ when mismatch is introduced.

We now introduce basis mismatch either by moving some of the modes off the DFT grid or by damping them. For frequency mismatch, the first two modes are moved to  $(2\pi \cdot 9.25/N, 1)$  and  $(2\pi \cdot 9.75/N, 1)$ . For damping mismatch the mode at  $(2\pi \cdot 9/N, 1)$  is drawn off the unit circle to radius 0.95, so that the mode is damped as  $(0.95)^n$ . The rest of the modes are the same as in the mismatch free case. Fig. 6(a)–(f) shows the inversion results for DFT, CS, and LP (order 8) for  $M = N/4 = 16$  to  $M = N/2 = 32$  to  $M = N/1 = 64$ . In all cases, DFT and CS result in erroneous inversion. The inaccuracy in inversion persists even when the number of measurements is increased to the full dimension. However, we observe that LP is always exact. These are all noise free cases.

But can the mismatch effect be compensated for by replacing the observation constraint  $\mathbf{y} = \mathbf{A}\mathbf{x}$  in basis pursuit with a quadratic constraint  $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon$ ? The plots in Fig. 7 suggest that this is not the case. These plots show CS inversions for three different values of  $\epsilon$ . The matched vs. mismatched modes and mode amplitudes in these plots are exactly the same as those in Fig. 6 and the number of measurements is  $M = 16$ . These results are not an artifact of the quadratic constraint allowed or the choice of rows in the compressed recording matrix, as we have experimented with many values and reported typical plots. For large values of  $\epsilon$  the inversion returns a zero vector, as seen in the fourth subplot in Fig. 7(a), (b). The reason is that when  $\epsilon > \|\mathbf{y}\|_2$  the sparsest solution satisfying  $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon$  is  $\mathbf{x} = 0$ .

Finally, we consider noisy observations for both mismatched and mismatch-free cases. In the mismatch-free case, the frequencies at which modes are placed, and their amplitudes, are

$(2\pi \cdot 9/N, 1)$ ,  $(2\pi \cdot 11/N, 1)$ ,  $(2\pi \cdot 20/N, .5)$ , and  $(2\pi \cdot 45/N, .2)$ . For frequency mismatch, the first two modes are moved to  $(2\pi \cdot 9.25/N, 1)$  and  $(2\pi \cdot 10.75/N, 1)$ . For damping mismatch the mode at  $(2\pi \cdot 9/N, 1)$  is drawn off the unit circle to radius 0.95. The number of measurements is  $M = N/2 = 32$ . The LP order is changed to 16, but rank reduction [9], [10] is applied to reduce the order back to 8 as is typical in noisy cases. The inversion results are shown in Fig. 8.

### C. Example 3: A Synthetic Worst-Case Example

Let  $\Psi_0 = \mathbf{I}$  be the sparsity basis assumed in the compressed sensing procedure and let  $\Psi_1 = \mathbf{I} + \beta \mathbf{1}\mathbf{1}^T$  be the actual sparsity basis, where  $\mathbf{1} = [1, 1, \dots, 1]^T$  is an  $N$ -dimensional vector of ones. Then, the mismatched basis is  $\Psi = \Psi_0^{-1} \Psi_1 = \mathbf{I} + \beta \mathbf{1}\mathbf{1}^T$ . Assume that the true parameter vector  $\boldsymbol{\theta}$  is a sparse vector in which all nonzero entries are equal to one and the positions of the nonzero entries are uniformly distributed. The coordinate transformation  $\mathbf{x} = \Psi \boldsymbol{\theta}$  has a simple form in this example and is given by

$$\mathbf{x} = (\mathbf{I} + \beta \mathbf{1}\mathbf{1}^T) \boldsymbol{\theta} = \boldsymbol{\theta} + k\beta \mathbf{1}. \quad (38)$$

Thus, a nonzero entry of 1 in  $\boldsymbol{\theta}$  results in a  $1 + k\beta$  entry in  $\mathbf{x}$  and a zero entry in  $\boldsymbol{\theta}$  produces a  $k\beta$  entry in  $\mathbf{x}$ . This example amounts to the worst case mismatch scenario in Theorem 1, with  $q = 1$ , where the upper bound is achieved. In this case,  $\|\mathbf{x}\|_1 = (1 + N\beta)\|\boldsymbol{\theta}\|_q$ , and the normalized error bound can be written as

$$\frac{\|\mathbf{x} - \mathbf{x}^*\|_1}{\|\mathbf{x}\|_1} \leq \frac{C_0}{1 + N\beta} \frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1}{\|\boldsymbol{\theta}\|_1} + C_0 \frac{(N - k)\beta}{1 + N\beta}. \quad (39)$$

Figs. 9(a)–(f) show the image inversion error  $\mathbf{x} - \mathbf{x}^*$  for three different values  $\beta = 0.001, 0.01, 0.1$  of the mismatch parameter. The plots in (a)–(c) correspond to one realization of the  $(k = 10)$ -sparse vector  $\boldsymbol{\theta}$  and the plots in (d)–(f) to another. The bars with solid circles show the value of error at an index where  $\boldsymbol{\theta}$  has a nonzero entry, which means the corresponding entry in  $\mathbf{x}$  is  $1 + k\beta$ . The bars with cross signs correspond to the zero elements in  $\boldsymbol{\theta}$  and their heights show the error in approximating entries in  $\mathbf{x}$  that have value  $k\beta$ .

Fig. 10(a)–(c) shows the histograms of the elements of  $(\mathbf{x}^* - \mathbf{x})/k\beta$ , averaged over all indices of  $\mathbf{x}$  and over 50 independent realizations of  $\boldsymbol{\theta}$ , for each of  $\beta = 0.001, 0.01, 0.1$ . The mean value for each histogram is approximately one. Since we have normalized the errors by  $k\beta$ , this means that the mean value of an element-wise error is approximately  $k\beta$ . Evidently, the realizations in Fig. 9(a)–(f) is typical error realization, where the majority of the element-wise errors are close to the mean value  $k\beta$ . The much larger values of error populate the tails of the histograms.

Fig. 11(a) and (b), respectively, shows the  $\ell_1$  error  $\|\mathbf{x} - \mathbf{x}^*\|_1$  and the normalized  $\ell_1$  error  $\|\mathbf{x} - \mathbf{x}^*\|_1 / \|\mathbf{x}\|_1$  as a function of the mismatch parameter  $\beta$  for dimensions  $N = 512$ ,  $M = 64$ . We notice that the  $\ell_1$  error  $\|\mathbf{x} - \mathbf{x}^*\|_1$  grows linearly with the increase in the mismatched level  $\beta$ , which agrees with Theorem 3. We also observe that even for moderate amounts of mismatch the normalized  $\ell_1$  error is considerable. When the mismatch level  $\beta$  goes to infinity, the normalized error bound converges

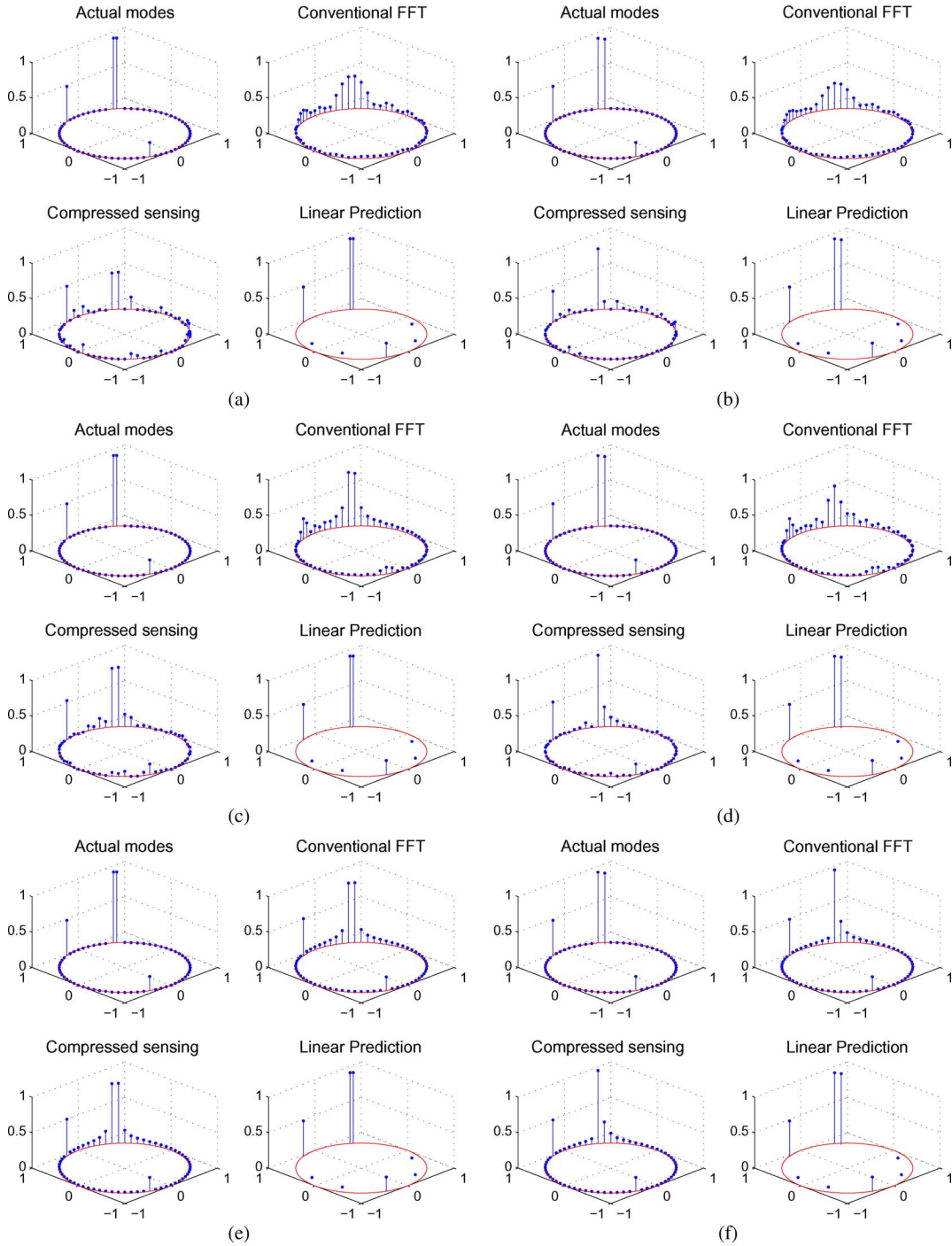


Fig. 6. Comparison of DFT, CS, and LP inversions in the presence of basis mismatch. Moving from the top to middle to bottom row, the number of measurements changes from  $M = N/4 = 16$  to  $M = N/2 = 32$  to  $M = N/1 = 64$ . (a) Frequency mismatch. (b) Damping mismatch. (c) Frequency mismatch. (d) Damping mismatch. (e) Frequency mismatch. (f) Damping mismatch.

to  $C_0((N - k)/N)$ . In Fig. 11(b), the normalized error curve becomes flat when  $\beta$  is above  $\mathcal{O}(10^{-2})$ .

### VI. CONCLUSIONS

The theory of compressed sensing suggests that compressed recording has manageable consequences for image inversion, provided the image is sparse in an apriori known basis, e.g., a DFT basis or a basis associated with a range-Doppler-wavenumber grid. But no physical field is

sparse in the DFT basis or in any apriori known basis defined by a regular grid in delay, doppler, frequency, and/or wavenumber, and there is always mismatch between the mathematical model for sparsity and the physical model for sparsity.

In this paper, we have investigated the sensitivity of compressed sensing (specifically basis pursuit) to mismatch between the assumed basis for sparsity and the actual sparsity basis. Our mathematical analysis and numerical examples indicate that the performance of compressed sensing for approximating a sparse

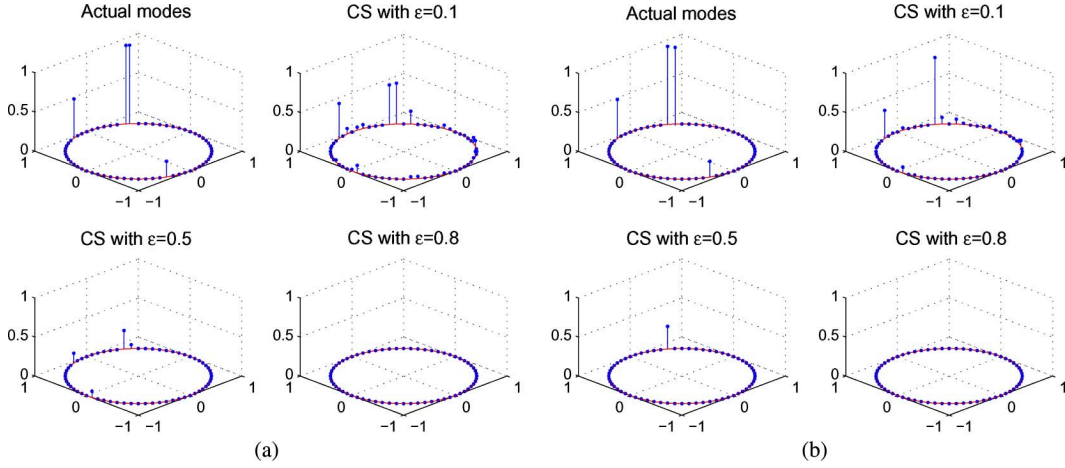


Fig. 7. Compressed sensing performance as the quadratic constraint  $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon$  is relaxed from  $\epsilon = 0.1$  to  $\epsilon = 0.5$  to  $\epsilon = 0.8$ ; (a) with frequency mismatch (b) with damping mismatch.

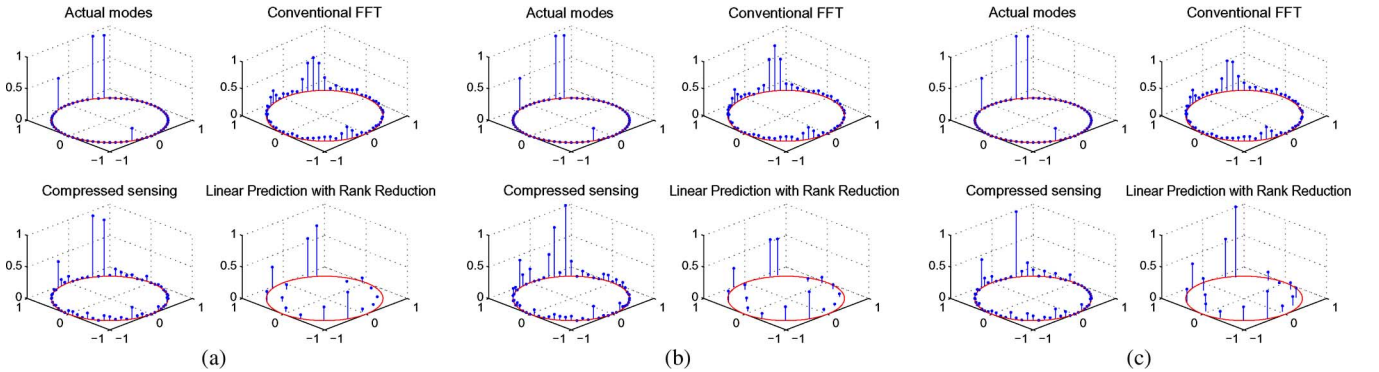


Fig. 8. Comparison of DFT, CS, and LP inversions with noisy observations: (a) no mismatch, (b) frequency mismatch, and (c) damping mismatch.

physical field may degrade considerably in the presence of basis mismatch, even when the assumed basis corresponds to a fine-grained discretization of the parameter space. Our analysis suggests that for high resolution spectrum analysis, DOA estimation, or delay-doppler imaging, where the problem is to identify a small number of modal parameters, extra care may be needed to account for the effects of basis mismatch.

## APPENDIX

### A. Proof of Theorem 1

Without loss of generality, let  $|\theta_1| \geq |\theta_2| \geq \dots \geq |\theta_N|$ . We have

$$\mathbf{x} = \Psi\boldsymbol{\theta} = (\mathbf{I} + \mathbf{E})\boldsymbol{\theta} = \boldsymbol{\theta} + \mathbf{E}\boldsymbol{\theta} = \begin{bmatrix} \theta_1 + \mathbf{e}_1^T \boldsymbol{\theta} \\ \theta_2 + \mathbf{e}_2^T \boldsymbol{\theta} \\ \dots \\ \theta_N + \mathbf{e}_N^T \boldsymbol{\theta} \end{bmatrix}. \quad (\text{A.1})$$

By the triangle inequality and Hölder's inequality we have

$$|\mathbf{e}_m^T \boldsymbol{\theta}| \leq \sum_{n=1}^N |e_{mn} \theta_n| \leq \|\mathbf{e}_m\|_p \|\boldsymbol{\theta}\|_q \leq \beta \|\boldsymbol{\theta}\|_q \quad (\text{A.2})$$

for all  $1 \leq m \leq N$ . Also, by the triangle inequality we have

$$|\theta_m| - |\mathbf{e}_m^T \boldsymbol{\theta}| \leq |\theta_m + \mathbf{e}_m^T \boldsymbol{\theta}| \leq |\theta_m| + |\mathbf{e}_m^T \boldsymbol{\theta}|. \quad (\text{A.3})$$

By combining the above two inequalities, we obtain

$$|\theta_m| - \beta \|\boldsymbol{\theta}\|_q \leq |\theta_m + \mathbf{e}_m^T \boldsymbol{\theta}| \leq |\theta_m| + \beta \|\boldsymbol{\theta}\|_q. \quad (\text{A.4})$$

Let  $\Sigma_{N-k}$  denote any subset of  $\{1, \dots, N\}$  with cardinality  $(N - k)$ . To get the upper bound, we can write

$$\|\mathbf{x} - \mathbf{x}_k\|_1 = \min_{\Sigma_{N-k}} \sum_{m \in \Sigma_{N-k}} |x_m| \quad (\text{A.5})$$

$$= \min_{\Sigma_{N-k}} \sum_{m \in \Sigma_{N-k}} |\theta_m + \mathbf{e}_m^T \boldsymbol{\theta}| \quad (\text{A.6})$$

$$\leq \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 + (N - k)\beta \|\boldsymbol{\theta}\|_q \quad (\text{A.7})$$

where (A.7) follows from (A.4). The upper bound is achieved when (20) holds with the positive sign, as shown in (A.8)–(A.11).

$$\min_{\Sigma_{N-k}} \sum_{m \in \Sigma_{N-k}} |\theta_m + \mathbf{e}_m^T \boldsymbol{\theta}| \quad (\text{A.8})$$

$$= \min_{\Sigma_{N-k}} \sum_{m \in \Sigma_{N-k}} \left| \theta_m + e^{j \arg(\theta_m)} \beta \right| \times \left( \sum_{n=1}^N \left( \frac{|\theta_n|}{\|\boldsymbol{\theta}\|_q} \right)^{q/p} \times e^{-j \arg(\theta_n)} \cdot \theta_n \right) \quad (\text{A.9})$$

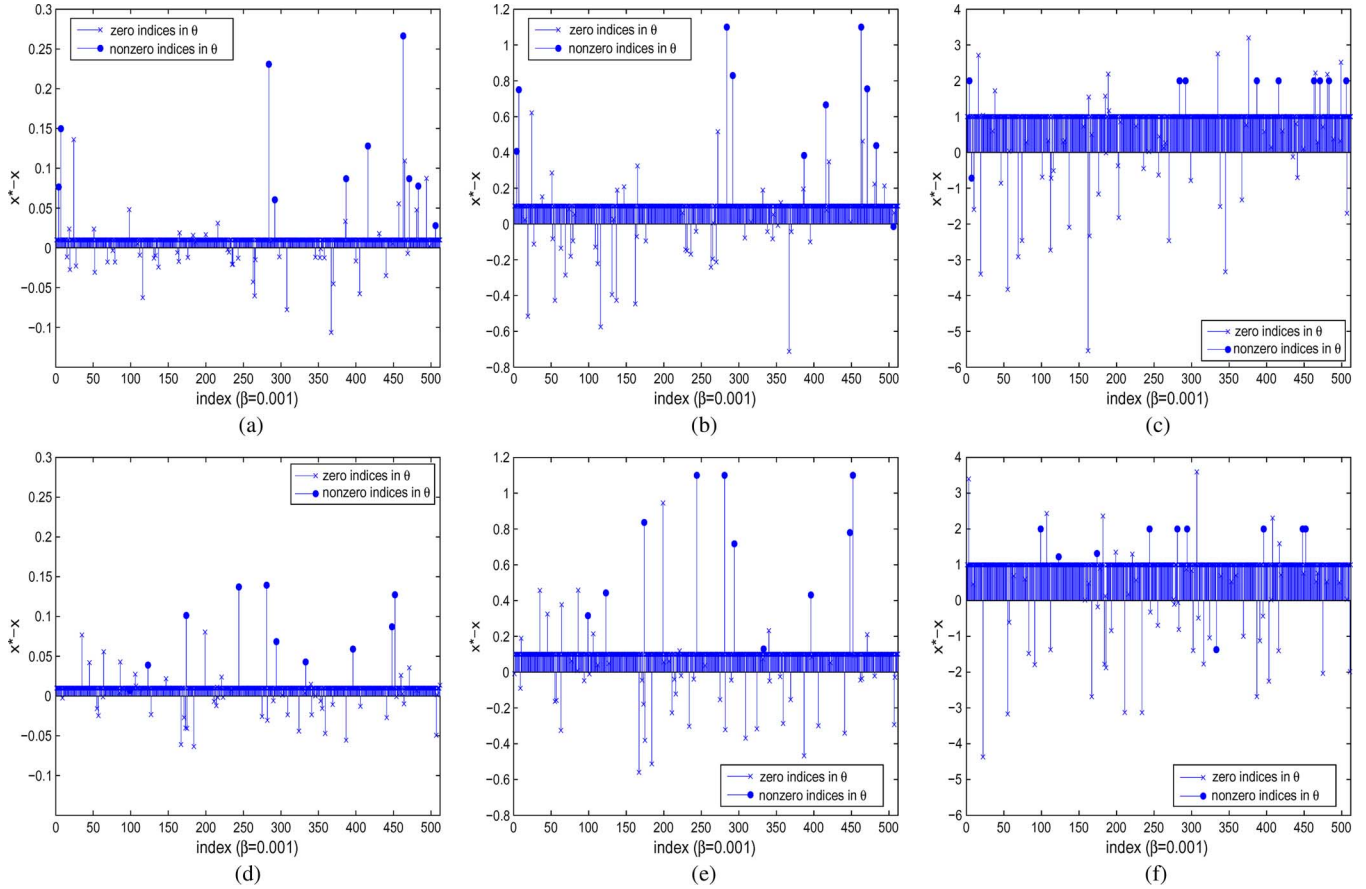


Fig. 9. The recovery error  $\mathbf{x} - \mathbf{x}^*$  from basis pursuit for  $N = 512$ ,  $M = 64$ ,  $k = 10$  and  $\beta = 0.001, 0.01, 0.1$ , respectively. The plots in (a)–(c) correspond to one realization of  $\boldsymbol{\theta}$  and the plots in (d)–(f) to another. (a)  $\beta = 0.001$ . (b)  $\beta = 0.01$ . (c)  $\beta = 0.1$ . (d)  $\beta = 0.001$ . (e)  $\beta = 0.01$ . (f)  $\beta = 0.1$ .

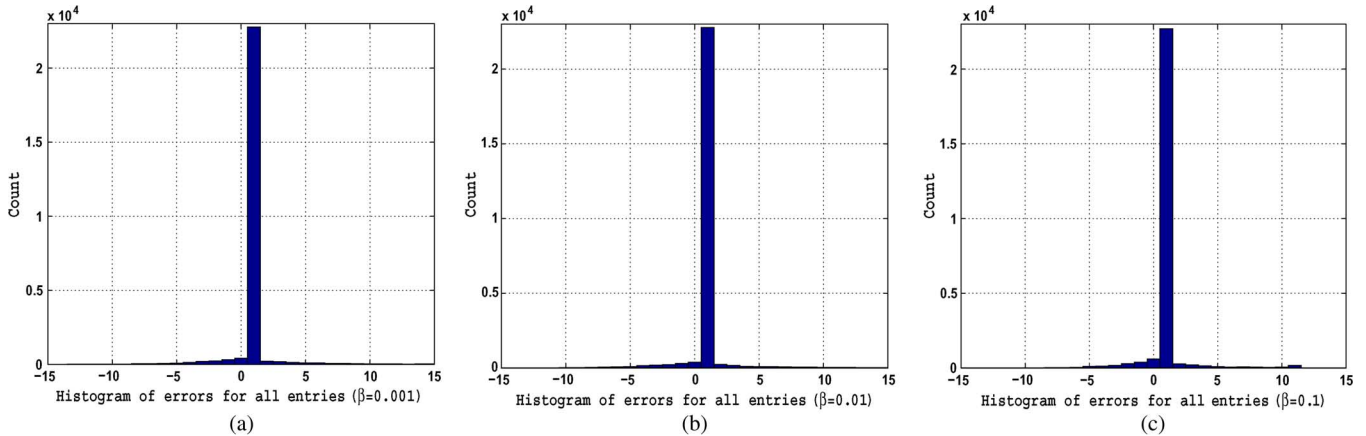


Fig. 10. Histograms of the elements of  $(\mathbf{x}^* - \mathbf{x})/k\beta$ , averaged over all indices of  $\mathbf{x}$  and over 50 independent realizations of  $\boldsymbol{\theta}$ , for  $\beta = 0.001, 0.01, 0.1$ . (a)  $\beta = 0.001$ . (b)  $\beta = 0.01$ . (c)  $\beta = 0.1$ .

$$\begin{aligned}
 &= \min_{\Sigma_{N-k}} \sum_{m \in \Sigma_{N-k}} |\theta_m| \\
 &+ (N - k)\beta \left( \sum_{n=1}^N \left( \frac{|\theta_n|}{\|\boldsymbol{\theta}\|_q} \right)^{q/p} |\theta_n| \right) \\
 &= \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 + (N - k)\beta \frac{\sum_{n=1}^N |\theta_n|^q}{\|\boldsymbol{\theta}\|_q^{q/p}} \quad (\text{A.10}) \\
 &= \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 + (N - k)\beta \|\boldsymbol{\theta}\|_q \quad (\text{A.11})
 \end{aligned}$$

where in writing (A.10) and (A.11) we have used the fact that  $1/p + 1/q = 1$ .

To get the lower bound, note that

$$\sum_{m \in \Sigma_{N-k}} |x_m| = \sum_{m \in \Sigma_{N-k}} |\theta_m + \mathbf{e}_m^T \boldsymbol{\theta}| \quad (\text{A.12})$$

$$\geq \sum_{m \in \Sigma_{N-k}} |\theta_m| - (N - k)\beta \|\boldsymbol{\theta}\|_q \quad (\text{A.13})$$

$$\geq \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 - (N - k)\beta \|\boldsymbol{\theta}\|_q \quad (\text{A.14})$$

for all  $\Sigma_{N-k}$ . Then,

$$\|\mathbf{x} - \mathbf{x}_k\|_1 = \min_{\Sigma_{N-k}} \sum_{m \in \Sigma_{N-k}} |x_m| \quad (\text{A.15})$$

$$\geq \|\boldsymbol{\theta} - \boldsymbol{\theta}_k\|_1 - (N - k)\beta \|\boldsymbol{\theta}\|_q, \quad (\text{A.16})$$

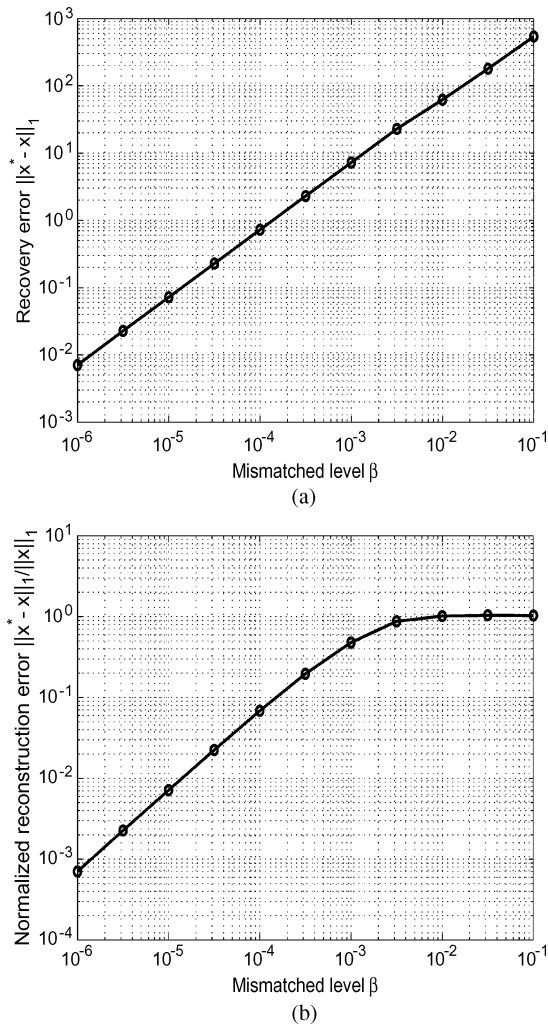


Fig. 11. (a) The recovery error  $\|x - x^*\|_1$  from basis pursuit versus the mismatched level  $\beta$  for  $N = 512$ ,  $M = 64$  and  $k = 10$ . (b) The normalized recovery error  $\|x - x^*\|_1 / \|x\|_1$  from basis pursuit versus the mismatched level  $\beta$  for  $N = 512$ ,  $M = 64$  and  $k = 10$ .

and it is easy to see that the bound is achieved when (20) holds with the negative sign.

### B. Proof of Theorem 2

By taking  $\beta = \eta$  in Theorem 1 and letting  $\mathbf{E}$  be the perturbation matrix with  $\|e_m\|_p = \beta = \eta$ , we have

$$\|x - x_k\|_1 = \|\theta - \theta_k\|_1 + (N - k)\eta\|\theta\|_q \quad (\text{B.1})$$

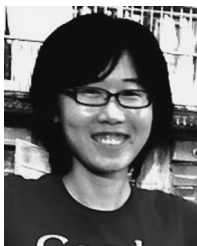
for  $x = (\mathbf{I} + \mathbf{E})\theta$ . Since  $\mathbf{E} \in \{\tilde{\mathbf{E}} : \|\tilde{e}_i\|_p \geq \eta\}$ , then

$$\max_{\tilde{\mathbf{E}}: \|\tilde{e}_i\|_p \geq \eta} \|x - x_k\|_1 \geq \|\theta - \theta_k\|_1 + (N - k)\eta\|\theta\|_q. \quad (\text{B.2})$$

### REFERENCES

- [1] C. T. Mullis and L. L. Scharf, "Quadratic estimators of the power spectrum," in *Advances in Spectrum Estimation*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1990, vol. 1, ch. 1, pp. 1–57.
- [2] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.
- [3] H. L. Van Trees, *Optimum Array Processing*. New York: Wiley Intersci., 2002.
- [4] H. Karim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.
- [5] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [6] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT-A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1340–1344, Oct. 1986.
- [7] J. Ward, "Maximum likelihood angle and velocity estimation with space-time adaptive processing radar," in *Proc. Conf. Rec. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1996, pp. 1265–1267.
- [8] R. Klemm, *Space-Time Adaptive Processing*. New York: IEEE, 1998.
- [9] D. W. Tufts and R. Kumaresan, "Singular value decomposition and improved frequency estimation using linear prediction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, no. 4, pp. 671–675, Aug. 1982.
- [10] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975–989, 1982.
- [11] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: John Hopkins Univ. Press, 1996.
- [12] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering and Tracking*. New York: IEEE, 2007.
- [13] L. T. McWhorter and L. L. Scharf, "Cramer-Rao bounds for deterministic modal analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 5, pp. 1847–1862, May 1993.
- [14] E. J. Candés, "Compressive sampling," in *Proc. Int. Congress Math.*, 2006, vol. 3, pp. 1433–1452.
- [15] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [16] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, 2007.
- [17] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 14–20, Mar. 2008.
- [18] E. J. Candés and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [19] D. L. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via  $\ell_1$  minimization," *Proc. Natl. Acad. Sci.*, vol. 100, pp. 2197–2202, 2002.
- [20] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [21] D. Donoho and J. Tanner, "Neighborliness of randomly-projected simplices in high dimensions," *Proc. Natl. Acad. Sci.*, vol. 102, no. 27, pp. 9452–9457, 2005.
- [22] E. J. Candés and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, Dec. 2005.
- [23] E. J. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [24] E. J. Candés and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [25] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.
- [26] E. J. Candés, "The restricted isometry property and its implications for compressed sensing," *Acad. des Sci.*, vol. 1, no. 346, pp. 589–592, 2008.
- [27] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing of approximately sparse signals," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, CA, Jul. 6–11, 2008, pp. 2182–2186.
- [28] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construct. Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.
- [29] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient and robust compressed sensing using optimized expander graphs," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4299–4308, Sep. 2009.
- [30] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic matrices that satisfy a statistical isometry property," *IEEE J. Sel. Topics in Signal Process., Special Issue on Compressed Sens.*, vol. 4, no. 2, pp. 358–374, Apr. 2010.
- [31] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. 2007 IEEE Radar Conf.*, Waltham, MA, Apr. 2007, pp. 128–133.
- [32] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, Jun. 2009.
- [33] M. A. Herman and T. Strohmer, "Compressed sensing radar," in *Proc. IEEE Radar Conf.*, Rome, Italy, May 2008, pp. 1–6.

- [34] M. A. Herman and T. Strohmer, "Compressed sensing radar," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 1509–1512.
- [35] A. Gurbuz, J. McClellan, and V. Cevher, "A compressive beamforming method," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 2617–2620.
- [36] V. Cevher, A. Gurbuz, J. McClellan, and R. Chellappa, "Compressive wireless arrays for bearing estimation of sparse sources in angle domain," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 2497–2500.
- [37] A. Fannjiang, P. Yan, and T. Strohmer, "Compressed remote sensing of sparse objects," *SIAM J. Imaging Sci.* 3, pp. 595–618, May 2010, arXiv:0904.3999v2.
- [38] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
- [39] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, Nov. 2010.
- [40] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," in *Proc. 46th Ann. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Sep. 23–26, 2008, pp. 575–582.
- [41] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. 42nd Ann. Conf. Inf. Sci. Syst.*, Princeton, NJ, Mar. 19–21, 2008, pp. 5–10.
- [42] D. Needell and R. Vershynin, "Signal recovery from inaccurate and incomplete measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics in Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.
- [43] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Computat. Harmon. Anal.*, vol. 26, pp. 301–321, 2008.
- [44] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE J. Sel. Topics in Signal Process.: Special Issue on Compressive Sens.*, vol. 4, no. 2, pp. 342–349, Apr. 2010.
- [45] M. A. Herman and D. Needell, "Mixed operators in compressed sensing," in *Proc. 44th Ann. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, Mar. 2010.
- [46] M. A. Herman and T. Strohmer, "General perturbations in compressed sensing," in *Workshop on Signal Process. With Adapt. Sparse Structured Represent. (SPARS)*, Saint-Malo, France, Apr. 2009.
- [47] M. A. Herman and T. Strohmer, "General perturbations of sparse signals in compressed sensing," in *Proc. Int. Conf. Sampling Theory and Appl. (SampTA)*, Marseille, France, May 2009.
- [48] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *IEEE Trans. Signal Process.*, 2010, submitted for publication.



**Yuejie Chi** (S'09) received the B.E. (Hon.) degree in electrical engineering from Tsinghua University, Beijing, China, in 2007.

She is currently working toward the Ph.D. degree in the Department of Electrical Engineering, Princeton University, NJ. Her research interests include compressed sensing, statistical signal processing, machine learning and their applications to wireless communications, and active sensing.



**Louis L. Scharf** (S'67–M'69–SM'77–F'86–LF'07) received the Ph.D. degree from the University of Washington, Seattle.

From 1971 to 1982, he served as Professor of Electrical Engineering and Statistics with Colorado State University (CSU), Ft. Collins. From 1982 to 1985, he was Professor and Chairman of Electrical and Computer Engineering, University of Rhode Island, Kingston. From 1985 to 2000, he was Professor of Electrical and Computer Engineering, University of Colorado, Boulder. In January 2001,

he rejoined CSU as Professor of Electrical and Computer Engineering and Statistics. He has held several visiting positions here and abroad, including the Ecole Supérieure d'Électricité, Gif-sur-Yvette, France; Ecole Nationale Supérieure des Télécommunications, Paris, France; EURECOM, Nice, Italy;

the University of La Plata, La Plata, Argentina; Duke University, Durham, NC; the University of Wisconsin, Madison; and the University of Tromsø, Tromsø, Norway. His interests are in statistical signal processing, as it applies to adaptive radar, sonar, and wireless communication. His most important contributions to date are to invariance theories for detection and estimation; matched and adaptive subspace detectors and estimators for radar, sonar, and data communication; and canonical decompositions for reduced dimensional filtering and quantizing. His current interests are in rapidly adaptive receiver design for space-time and frequency-time signal processing in the radar/sonar and wireless communication channels.

Prof. Scharf was Technical Program Chair for the 1980 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Denver, CO; Tutorials Chair for ICASSP 2001, Salt Lake City, UT; and Technical Program Chair for the Asilomar Conference on Signals, Systems, and Computers 2002. He is past-Chair of the Fellow Committee for the IEEE Signal Processing Society and serves on the Technical Committee for Sensor Arrays and Multichannel Signal Processing. He has received numerous awards for his research contributions to statistical signal processing, including a College Research Award, an IEEE Distinguished Lectureship, an IEEE Third Millennium Medal, and the Technical Achievement and Society Awards from the IEEE Signal Processing Society.



**Ali Pezeshki** (S'95–M'05) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Tehran, Tehran, Iran, in 1999 and 2001, respectively. He received the Ph.D. degree in electrical engineering from Colorado State University, Fort Collins, in 2004.

In 2005, he was a Postdoctoral Research Associate with the Electrical and Computer Engineering Department, Colorado State University. From January 2006 to August 2008, he was a Postdoctoral Research Associate with The Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ. Since August 2008, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, Colorado State University. His research interests are in statistical signal processing and coding theory and their applications to distributed sensing, wireless communication, active/passive sensing, data networking, and bioimaging.



**A. Robert Calderbank** (M'89–SM'97–F'98) received the B.Sc. degree in 1975 from Warwick University, England, the M.Sc. degree in 1976 from Oxford University, England, and the Ph.D. degree in 1980 from the California Institute of Technology, Pasadena, all in mathematics.

He is Dean of Natural Sciences at Duke University, Durham, NC. He was previously Professor of Electrical Engineering and Mathematics at Princeton University, Princeton, NJ, where he directed the Program in Applied and Computational Mathematics. Prior to joining Princeton University in 2004, he was Vice President for Research at AT&T, responsible for directing the first industrial research lab in the world where the primary focus is data at scale. At the start of his career at Bell Labs, his innovations were incorporated in a progression of voiceband modem standards that moved communications practice close to the Shannon limit. Together with Peter Shor and colleagues at AT&T Labs, he showed that good quantum error correcting codes exist and developed the group theoretic framework for quantum error correction. He is a co-inventor of space-time codes for wireless communication, where correlation of signals across different transmit antennas is the key to reliable transmission.

Dr. Calderbank served as Editor-in-Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY from 1995 to 1998, and as an Associate Editor for Coding Techniques from 1986 to 1989. He was a member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996 and from 2006 to 2008. He was honored by the IEEE Information Theory Prize Paper Award in 1995 for his work on the Z4 linearity of Kerdock and Preparata Codes (joint with A. R. Hammons Jr., P. V. Kumar, N. J. A. Sloane, and P. Sole), and again in 1999 for the invention of space-time codes (joint with V. Tarokh and N. Seshadri). He received the 2006 IEEE Donald G. Fink Prize Paper Award and the IEEE Millennium Medal, and was elected to the US National Academy of Engineering in 2005.