Dealing with Frequency Perturbations in Compressive Reconstructions with Fourier Sensing Matrices

Himanshu Pandotra[a], Eeshan Malhotra[b], Ajit Rajwade[b,*], Karthik S. Gurumoorthy[c]

[a]*Department of Electrical Engineering, IIT Bombay*
[b]*Department of Computer Science and Engineering, IIT Bombay*
[c]*International Center for Theoretical Sciences, TIFR (ICTS-TIFR), Bangalore*

Abstract

In many applications in compressed sensing, the measurement matrix is a Fourier matrix, *i.e.*, it measures the Fourier transform of the underlying signal at some specified 'base' frequencies $\{u_i\}_{i=1}^M$, where M is the number of measurements. However due to system calibration errors, the system may measure the Fourier transform at frequencies $\{u_i + \delta_i\}_{i=1}^M$ that are different from the base frequencies and where $\{\delta_i\}_{i=1}^M$ are unknown frequency perturbations. Ignoring such perturbations can lead to major errors in signal recovery. In this paper, we present a simple but effective alternating minimization algorithm to recover the perturbations in the frequencies *in situ* with the signal, which we assume is sparse or compressible in some known basis. In many cases, the perturbations $\{\delta_i\}_{i=1}^M$ can be expressed in terms of a small number of unique parameters $P \ll M$. We demonstrate that in such cases, the method leads to excellent quality results that are several times better than baseline algorithms (which are based on existing off-grid methods in the recent literature on direction of arrival (DOA) estimation, modified to suit the computational problem in this paper). Our results are also robust to noise in the measurement values. We also provide theoretical results for (1) the conditional convergence of our algorithm, and (2) the uniqueness of the solution for this computational problem, under some restrictions.

Keywords: Compressed sensing, Fourier measurements, Frequency Perturbation

1. Introduction

Compressed sensing (CS) is today a very widely researched branch of signal and image processing. Consider a vector of compressive measurements $\boldsymbol{y} \in \mathbb{C}^M, \boldsymbol{y} = \boldsymbol{\Phi x}$ for signal $\boldsymbol{x} \in \mathbb{C}^N$, acquired through a

*Corresponding author
Email addresses: pandotra.himanshu@gmail.com (Himanshu Pandotra), eeshan@gmail.com (Eeshan Malhotra), ajitvr@cse.iitb.ac.in (Ajit Rajwade), karthik.gurumoorthy@icst.res.in (Karthik S. Gurumoorthy)

[1]Ajit Rajwade gratefully acknowledges support from IIT Bombay Seed Grant number 14IRCCSG012, and from NVIDIA Corporation for the donation of Titan Xp GPUs.
[2]Karthik S. Gurumoorthy thanks the AIRBUS Group Corporate Foundation Chair in Mathematics of Complex Systems established in ICTS-TIFR.

sensing matrix $\mathbf{\Phi} \in \mathbb{C}^{M \times N}, M < N$. CS theory offers guarantees on the error of reconstruction of $\boldsymbol{x}$ that is sparse or compressible in a given orthonormal basis $\mathbf{\Psi} \in \mathbb{C}^{N \times N}$, assuming that the sensing matrix (also called measurement matrix) $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$ (and hence the product matrix $\mathbf{\Phi\Psi}$) obeys some properties such as the restricted isometry (RIP) [1]. Moreover, the guarantees apply to efficient algorithms such as basis pursuit. However the underlying assumption is that the sensing matrix $\mathbf{\Phi}$ is known accurately. If $\mathbf{\Phi}$ is known inaccurately, then signal-dependent noise will be introduced in the system causing substantial loss in reconstruction accuracy.

Of particular interest in many imaging applications such as magnetic resonance imaging (MRI), tomography or Fourier optics [2, 3, 4, 5], is the case where the measurement matrix is a row-subsampled version of the Fourier matrix, where the frequencies may or may not lie on a Cartesian grid of frequencies used in defining the Discrete Fourier Transform (DFT). However, it is well-known that such Fourier measurements are prone to inaccuracies in the acquisition frequencies. This may be due to an imperfectly calibrated sensor. In case of specific applications such as MRI, this is due to perturbations introduced by gradient delays in the MRI machine [6, 7, 8]. In case of computed tomography (CT), it may be due to errors in specification of the angles of tomographic acquisition due to geometric calibration errors in a CT machine [5], or in the problem of tomographic reconstruction under unknown angles [9].

1.1. Relation to Previous Work

The problem we deal with in this paper is a special case of the problem of 'blind calibration' (also termed 'self-calibration') where perturbations in the sensing matrix are estimated *in situ* along with the signal. Here, we expressly deal with the case of Fourier sensing matrices with imperfectly known frequencies. There exists a decent-sized body of earlier literature on the general blind calibration problem (not applied to Fourier matrices) beginning with theoretical bounds derived in [10]. Further on, [11, 12] analyze a structured perturbation model of the form $\boldsymbol{y} = (\boldsymbol{A} + \boldsymbol{B}\mathbf{\Delta})\boldsymbol{x}$ where $\boldsymbol{x}, \mathbf{\Delta}$ are the unknown signal and diagonal matrix of perturbation values respectively, and $\boldsymbol{A}, \boldsymbol{B}$ are the fully known original sensing matrix and perturbation matrix respectively. The theory is then applied to direction of arrival (DOA) estimation in signal processing. Further work in [13] uses the notion of group-sparsity to infer the signal $\boldsymbol{x}$ and the perturbations $\mathbf{\Delta}$ using a convex program based on a first order Taylor expansion of the parametric DOA matrix. A total least squares framework that also accounts for sparsity of the signal is explored in [14] for a perturbation model of the form $\boldsymbol{y} + \boldsymbol{e} = (\boldsymbol{A} + \boldsymbol{E})\boldsymbol{x}$ where $\boldsymbol{e}, \boldsymbol{E}$ are the additive errors in the measurement vector $\boldsymbol{y}$ and measurement matrix $\boldsymbol{A}$ respectively. The total least squares framework is further extended to include an ℓ_1-regularized Rayleigh quotient term in [15]. Error bounds for compressed sensing in an ℓ_p minimization framework $(0 < p < 1)$ for sensing matrices with additive perturbations are developed in [16]. In [17], performance bounds are developed for the simultaneous orthogonal matching pursuit (SOMP) algorithm, for the case when the signals have similar supports and are measured with additively perturbed sensing matrices. In [18], [19], [20],[21], the

following framework is considered: $\boldsymbol{y} = \boldsymbol{\Delta A x}$, where $\boldsymbol{\Delta}$ is a diagonal matrix containing the unknown sensor gains which may be complex, $\boldsymbol{x}$ is the unknown sparse signal, and $\boldsymbol{A}$ is the known sensing matrix. Both $\boldsymbol{x}$ and $\boldsymbol{\Delta}$ are recovered together via linear least squares in [18], via the lifting technique on a biconvex problem in [19], via a variety of convex optimization tools in [20], and in [21] via a non-convex method. The problem we deal with in this paper cannot be framed as a single (per measurement) unknown phase or amplitude shift/gain unlike these techniques, and hence is considerably different.

Related to (but still very different from) the aforementioned problem of a perturbed sensing matrix, is the problem of a perturbed or mismatched signal representation matrix $\boldsymbol{\Psi}$ which can also cause significant errors in compressive recovery [22]. This has been explored via alternating minimization in [23], via a perturbed form of orthogonal matching pursuit (OMP) in [24], and via group-sparsity in [13]. The problem of estimating a small number of complex sinusoids with off-the-grid frequencies from a subset of regularly spaced samples has been explored in [25]. Note that in [22, 25, 13, 23], the emphasis is on mismatch in the *representation* matrix $\boldsymbol{\Psi}$ and *not* in the sensing matrix $\boldsymbol{\Phi}$ - see Section 4 for more details. The problem of *additive* perturbations in both the sensing matrix as well as the representation matrix has been analyzed in [26], using several assumptions on both perturbations. Note that the perturbations in the Fourier sensing matrix do not possess such an additive nature.

To the best of our knowledge, there is no previous work on the analysis of perturbations in a Fourier *measurement* matrix in a *compressive sensing* framework. Some attempts have been made to account for frequency specification errors in MRI, however, most of these require a separate off-line calibration step where the perturbations are measured. However in practice, the perturbations in frequencies may be common to only subsets of measurements (or even vary with each measurement), and need not be static. Hence the measurement matrix inaccuracies cannot be pre-determined, and must be estimated *in situ* along with the signal. In cases where the correction is made alongside the recovery step, a large number of measurements may be required [27], as the signal reconstruction does not deal with a compressed sensing framework involving ℓ_q ($q \leq 1$) minimization. The problem of perturbations in the Fourier matrix also occurs in computed tomography (CT). This happens in an indirect way via the Fourier slice theorem, since the 1D Fourier transform of a parallel beam tomographic projection in some acquisition angle α is known to be equal to a slice of the Fourier transform of the underlying 2D image at angle α. In CT, the angles for tomographic projection may be incorrectly known due to geometric errors [5], and uncertainty in the angle will manifest as inaccuracy of the Fourier measurements. While there exist approaches to determine even the completely unknown angles of projection, they require a large number of projections, and also the knowledge of the distribution of the angles [28, 29]. Our group has presented a method [9] which does not require this knowledge, but in [9], the angles are estimated only along with the image moments. The image itself is estimated after determining the angles. In contrast, in this paper, the errors in frequency are determined

along with the underlying signal.

A large body of existing work is also lacking in theoretical backing. Some existing approaches to handle perturbations in $\mathbf{\Psi}$ simplify the problem using a Taylor approximation [11, 30]. However, when such an approach is tailored to the problem of perturbation in $\mathbf{\Phi}$, it proves to be adequate only at extremely small perturbation levels in our case, rendering the adjustment for the perturbation to be much less effective (See Section 5).

Contributions: A method for simultaneous recovery of the perturbations and the signal in a perturbed Fourier compressed sensing structure is proposed in this paper. The algorithm is verified empirically over a large range of simulated data under noise-free and noisy cases. Further, we analyze the convergence of the algorithm, as well as the uniqueness of the solution to our specific computational problem under specific but realistic assumptions about the measurement perturbations. A preliminary version of this work has earlier appeared in [31]. In this paper, we provide more extensive experimental comparison, problem motivation as well as uniqueness results.

1.2. Organization of the Paper

This paper is organized as follows. Section 2 defines the problem statement. The recovery algorithm is presented in Section 3, followed by a comparison with related computational problems in Sec. 4, and extensive numerical results in Section 5. The theoretical treatment is covered in Section 6, followed by a conclusion in Section 7

2. Problem Definition

Formally, let $\boldsymbol{F} \in \mathbb{C}^{M\times N}$ be a Fourier matrix using a known (possibly, but not necessarily on-grid) frequency set $\boldsymbol{u} \triangleq \{u_i\}_{i=1}^{M} \in \mathbb{R}^M$, $\boldsymbol{x} \in \mathbb{R}^N$ be a signal that is sparse (with at most s non-zero values) or compressible, measured using a perturbed Fourier matrix $\boldsymbol{F}(\boldsymbol{\delta}) \in \mathbb{C}^{M\times N}$. That is,

$$\boldsymbol{y} = \boldsymbol{F}(\boldsymbol{\delta})\boldsymbol{x} + \boldsymbol{\eta}, \tag{1}$$

where, $\boldsymbol{\eta}$ is a signal-independent noise vector, $\boldsymbol{F}(\boldsymbol{\delta})$ is a Fourier measurement matrix at the set of unknown frequencies $\boldsymbol{u} + \boldsymbol{\delta} \triangleq \{u_i + \delta_i\}_{i=1}^{M}$, with $\forall i, \delta_i \in \mathbb{R}, |\delta_i| \leq r, r \geq 0, \boldsymbol{\delta} \triangleq \{\delta_i\}_{i=1}^{M}$. For the i^{th} measurement in particular, we have:

$$y_i = \sum_l x(l)\exp(-\iota 2\pi(u_i+\delta_i)l/N) = \sum_l (x(l)\exp(-\iota 2\pi\delta_i l/N))\exp(-\iota 2\pi u_i l/N), \tag{2}$$

where $\iota \triangleq \sqrt{-1}$. This is the Fourier transform at frequency u_i of a phase-modulated version of $\boldsymbol{x}$. Note that we assume full knowledge of $\{u_i\}_{i=1}^{M}$, *i.e.*, the base frequencies. The problem is to recover *both*, the sparse

signal $\boldsymbol{x}$, and the unknown perturbations in the frequencies, $\boldsymbol{\delta}$. This is formalized as the following:

$$\min_{\hat{\boldsymbol{x}},\hat{\boldsymbol{\delta}}\in[-r,r]^M} J(\hat{\boldsymbol{x}},\hat{\boldsymbol{\delta}}) \triangleq \|\hat{\boldsymbol{x}}\|_1 + \lambda\|\boldsymbol{y} - \boldsymbol{F}(\hat{\boldsymbol{\delta}})\hat{\boldsymbol{x}}\|_2 \tag{3}$$

where $\boldsymbol{F}(\hat{\boldsymbol{\delta}})$ is the Fourier measurement matrix at frequencies $\boldsymbol{u} + \hat{\boldsymbol{\delta}}$, and $\hat{\boldsymbol{\delta}}$ denotes the estimate of $\boldsymbol{\delta}$. Note that the above problem is a perturbed version of the so-called square-root LASSO (SQ-LASSO), since the second term involves an ℓ_2 norm and not its square. We used the SQ-LASSO due to its advantages over the LASSO in terms of parameter tuning, as mentioned in [32].

Eqn. 3 presents the most general formulation of the problem. The signal may be sparse in a non-canonical basis, say the Discrete Wavelet transform (DWT), in which case the objective function in Eqn. 3 can be changed, leading to the following problem:

$$\min_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\delta}}\in[-r,r]^M} J(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\delta}}) \triangleq \|\hat{\boldsymbol{\theta}}\|_1 + \lambda\|\boldsymbol{y} - \boldsymbol{F}(\hat{\boldsymbol{\delta}})\boldsymbol{\Psi}\hat{\boldsymbol{\theta}}\|_2, \tag{4}$$

where $\boldsymbol{\theta} = \boldsymbol{\Psi}^T\boldsymbol{x}$ are the wavelet coefficients of $\boldsymbol{x}$. We also discuss an important modification. In Eqn. 3, we have assumed that all perturbations, i.e. entries in $\boldsymbol{\delta}$, are independent. However, this may not necessarily be the case in many applications. For example, consider the following three cases (though the applicability of our technique and analysis is not restricted to just these):

1. Consider parallel beam tomographic reconstruction of a 2D signal $f(x, y)$ with incorrectly specified angles. The 1D Fourier transform of the tomographic projection of f acquired at some angle α is equal to a slice through the 2D Fourier transform of f at angle α and passing through the origin of the Fourier plane. The frequencies along this slice can be expressed in the form $u^{(1)} = \rho\cos\alpha, u^{(2)} = \rho\sin\alpha$ where $\rho = \sqrt{(u^{(1)})^2 + (u^{(2)})^2}$. If the specified angle has an error $\bar{\alpha}$, the effective Fourier measurements are at frequencies $\bar{u}^{(1)} = \rho\cos(\alpha + \bar{\alpha}), \bar{u}^{(2)} = \rho\sin(\alpha + \bar{\alpha})$. In such a case, the perturbations in all the frequencies along a single slice are governed by a *single* parameter $\bar{\alpha}$ which is unknown. (The parameter ρ is known since the base frequencies $(u^{(1)}, u^{(2)})$ are known.) This basic principle also extends to other projection methods such as cone-beam and to higher dimensions. (See Fig. 11 for sample reconstructions for this application).

2. The problem of tomography under unknown angles is of interest in cryo-electron microscopy to determine the structure of virus particles [33]. Here the angles of tomographic projection as well as the underlying image are both unknown. In some techniques, the angles of projection are estimated first using techniques from dimensionality reduction [29] or geometric relationships [9, 28]. Any error in the angle estimates affects the estimate of the underlying image in a manner similar to that described in the previous point.

3. In MRI, gradient delays can cause errors in the specified set of frequencies at which the Fourier transform is measured [8]. The gradient delays are essentially the difference between the programmed or

specified start time and the start time which the machine uses for the measurement. For a single axis, the gradient $G(t)$ would produce a trajectory of measurements of the form $k(t) = K\int_{\tau=0}^{t} G(\tau)d\tau$ at time t where K is a hardware-related proportionality constant and $u(t) \triangleq (u^{(1)}(t), u^{(2)}(t))$ for 2D measurements. Given a gradient delay of $\bar{t}$, the actual trajectory would be $k'(t) = K\int_{\tau=0}^{t} G(\tau - \bar{t})d\tau$. For small-valued $\bar{t}$, this leads to a trajectory error proportional to $G(t)\bar{t}$ [33]. Thus frequency perturbations in MRI measurements for a single axis are governed by a single parameter $\bar{t}$. In some specific MRI sampling schemes such as radial, a single global trajectory error is assumed for all frequencies in one or all radial spokes (see Eqn. 3 of [34], and 'Methods section' in [35]). This global error arises due to gradient delays, which again presents a case of perturbations in multiple measurements being expressed in terms of a single parameter.

Handling cases such as these in fact makes the recovery problem more tractable, since the number of unknowns is essentially reduced. We now present our recovery algorithm and its modified version for handling cases where many measurements share a common set of 'perturbation parameters', in the following section. The convergence of the algorithm is analyzed in Section 6.1.

3. Recovery Algorithm

We present an algorithm to determine $\boldsymbol{x}$ and $\boldsymbol{\delta}$ by using an alternation between two sub-problems. Starting with a guess $\hat{\boldsymbol{\delta}}$ for the perturbations $\boldsymbol{\delta}$, we recover $\hat{\boldsymbol{x}}$, an estimate for $\boldsymbol{x}$, using the SQ-LASSO mentioned before, which is essentially an unconstrained l_1 norm minimization approach common in compressive sensing. Next, using this first estimate $\hat{\boldsymbol{x}}$, we update $\hat{\boldsymbol{\delta}}$ to be the best estimate, assuming $\hat{\boldsymbol{x}}$ to be the truth, using a linear brute force search in the range $-r$ to r. A linear search is possible because each measurement y_i is the dot product of a single row of $\boldsymbol{F}(\boldsymbol{\delta})$ with $\boldsymbol{x}$, and hence a single (u_i, δ_i) value is involved. Consequently, the different δ_i values can be recovered through *independent* parallel searches (see Section 4 for a comparison to related computational problems). From here on, we alternate between the two steps - recovery of $\hat{\boldsymbol{x}}$ and recovery of $\hat{\boldsymbol{\delta}}$, till convergence is achieved.

Since the search space is highly non-convex, we also employ a multi-start strategy, where, we perform multiple runs of the alternating algorithm to recover $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{x}}$, each time, initializing the first guess for $\hat{\boldsymbol{\delta}}$ randomly. We ultimately select the solution that minimizes the objective function $J(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\delta}})$. In practice, we have observed that the number of starts required for a good quality solution is rather small (around 10).

The full algorithm, including the optimization for multi-start is presented in Algorithm 1. Note that $\boldsymbol{F_k}(\hat{\delta_k})$ denotes the k^{th} row of $\boldsymbol{F}(\hat{\boldsymbol{\delta}})$. We now consider the important and realistic cases where values in $\hat{\boldsymbol{\delta}}$ can be expressed in terms of a small number of unique parameters $\boldsymbol{\beta} \triangleq \{\beta_i\}_{i=1}^{P}$ where $P \ll M$. We henceforth term these 'perturbation parameters'. In other words, there are subsets of measurements whose

Algorithm 1 Alternating Minimization Algorithm

procedure AlternativeRecovery
  $converged \leftarrow False, \chi \leftarrow 0.0001$
  $\hat{\boldsymbol{\delta}} \leftarrow$ sample from Uniform$[-r, +r]$
  while $converged == False$ **do**
    $\boldsymbol{F}(\hat{\boldsymbol{\delta}}) \leftarrow$ Fourier matrix at $(\boldsymbol{u} + \hat{\boldsymbol{\delta}})$
    Estimate $\hat{\boldsymbol{x}}$ as:
    $\min_{\hat{\boldsymbol{x}}} \|\hat{\boldsymbol{x}}\|_1 + \lambda \|\boldsymbol{y} - \boldsymbol{F}(\hat{\boldsymbol{\delta}})\hat{\boldsymbol{x}}\|_2$

    for k in $1 \rightarrow M$ **do**
      Test each discretized value of $\hat{\delta}_k$ in range $-r$ to r and select the value to achieve $\min_{\hat{\delta}_k} \|y_k - \boldsymbol{F_k}(\hat{\delta}_k)\hat{\boldsymbol{x}}\|^2$
    end for
    if $\|\hat{\boldsymbol{\delta}} - \hat{\boldsymbol{\delta}}_{prev}\|_2 < \chi$ and $\|\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}_{prev}\|_2 < \chi$ **then**
      $converged \leftarrow True$
  end while
  return $\hat{\boldsymbol{x}}, \hat{\boldsymbol{\delta}}$

procedure Multistart
  $minobjective \leftarrow \infty$
  $\hat{\boldsymbol{x}}_{best} \leftarrow null$
  $\hat{\boldsymbol{\delta}}_{best} \leftarrow null$
  for start in $1 \rightarrow$ numstarts **do**
    $\hat{\boldsymbol{x}}, \hat{\boldsymbol{\delta}} \leftarrow$ AlternatingRecovery()
    $\boldsymbol{F}(\hat{\boldsymbol{\delta}}) \leftarrow$ Fourier matrix at $(u + \hat{\boldsymbol{\delta}})$
    $objective \leftarrow \|\hat{\boldsymbol{x}}\|_1 + \lambda \|y - \boldsymbol{F}(\hat{\boldsymbol{\delta}})\hat{\boldsymbol{x}}\|_2$
    if $objective < minobjective$ **then**
      $\hat{\boldsymbol{x}}_{best} \leftarrow \hat{\boldsymbol{x}}$
      $\hat{\boldsymbol{\delta}}_{best} \leftarrow \hat{\boldsymbol{\delta}}$
      $minobjective \leftarrow objective$
  end for
  return $\hat{\boldsymbol{x}}_{best}, \hat{\boldsymbol{\delta}}_{best}$

frequency perturbation values are expressed fully in terms of a *single* perturbation parameter from $\boldsymbol{\beta}$ (besides the base frequency itself). We assume that $\forall k, 1 \leq k \leq P, |\beta_k| \leq r$, where $r > 0$ is known. Let the k^{th} unique value in $\boldsymbol{\beta}$ correspond to the perturbation parameter for measurements in a set L_k, indexing into the measurement vector $\boldsymbol{y}$. Thus $\forall i \in L_k, \delta_i = h(\beta_k, u_i)$ where h is a known function of the perturbation parameter β_k and base frequency u_i. The exact formula for h is dictated by the specific application.

For example, in the CT application cited at the end of the previous section, let us define set L_k to contain indices of all frequencies along the k^{th} radial spoke at some angle α_k. The perturbation values δ_i for all base frequencies u_i in L_k can be expressed in terms of a single parameter - the error β_k in specifying the angle. Here, for frequency $u_i = (u_i^{(1)}, u_i^{(2)})$, we would have $\delta_i = h(\beta_k, u_i) \triangleq (\rho_i(\cos(\alpha_k + \beta_k) - \cos\alpha_k), \rho_i(\sin(\alpha_k + \beta_k) - \sin\alpha_k))$ where $\rho_i = \sqrt{(u_i^{(1)})^2 + (u_i^{(2)})^2}, u_i^{(1)} = \rho_i \cos\alpha_k, u_i^{(2)} = \rho_i \sin\alpha_k$. In the MRI example, the perturbation values for all base frequencies u_i along the k^{th} axis can be expressed in terms of a single perturbation parameter β_k, which stands for the gradient delay for the k^{th} axis. In this case, $\delta_i = h(\beta_k, u_i) \triangleq (K'\beta_k G_x(t), K'\beta_k G_y(t))$ for hardware-related proportionality constant K' and where $G_x(t), G_y(t)$ are the x, y components of the gradient at time t (at which the Fourier transform at frequency $u_i + \delta_i$ was measured). In the case of radial MRI, the parallel and perpendicular components of the error at every frequency in the trajectory along the radial spoke at angle α are expressed as $\delta_{par} = K(t_x \cos^2\alpha + t_y \sin^2\alpha), \delta_{perp} = K(-t_x \cos\alpha \sin\alpha + t_y \sin\alpha \cos\alpha)$ where t_x, t_y represent gradient delays [35] and K is a hardware-related constant. Here, the perturbation parameters are $\beta_1 = t_x, \beta_2 = t_y$, and they are common to all radial spokes.

To suit these cases of perturbation parameters common to many measurements, we modify Algorithm 1, for which Step 9 can then be replaced by:

for k in $1 \rightarrow P$ **do**

Test each discretized value of $\hat{d}_k$ in range $-r$ to r

$\beta_k = \underset{\hat{d}_k}{\text{argmin}} \|\boldsymbol{y}_{\boldsymbol{L_k}} - \boldsymbol{F}_{\boldsymbol{L_k}}(\hat{d}_k)\hat{\boldsymbol{x}}\|_2$

for each i in L_k **do**

Compute δ_i from β_k using $\delta_i = h(\beta_k, u_i)$

end for

end for

In the above steps, $\boldsymbol{y}_{\boldsymbol{L_k}}$ is a subvector of $\boldsymbol{y}$, containing measurements for frequencies at indices only in L_k, and $\boldsymbol{F}_{\boldsymbol{L_k}}(\hat{d}_k)$ denotes a sub-matrix of $\boldsymbol{F}$ containing only those rows with indices in L_k and assuming perturbation parameter $\hat{d}_k$. Note that the modification to the main algorithm essentially computes only each *unique* value in $\boldsymbol{\beta}$ separately. Convergence results for Algorithm 1 (or its modification) are analyzed in

Section 6.1.

4. Comparison with Related Computational Problems

We emphasize that our computational problem is *very different* from the basis mismatch problem [23, 22, 24]. There, the signal is to be represented as a linear combination of (possibly sinsuoidal) bases whose frequencies are assumed to lie on a discrete grid, i.e. $\boldsymbol{x} = \boldsymbol{\Phi\Psi\theta} = \boldsymbol{\Phi}\sum_{k=0}^{K-1}\boldsymbol{\Psi_k}\theta_k$, where $\boldsymbol{\Psi_k} \in \mathbb{C}^N$ is the basis vector at discrete frequency k, and $\boldsymbol{\theta} \in \mathbb{C}^N$. However in many applications, the signals may be sparse linear combinations of bases whose frequencies lie off the grid. Hence the representation problem involves solving for the frequency perturbations δ_k along with $\boldsymbol{\theta}$ given $\boldsymbol{x}$, where $\boldsymbol{x} = \boldsymbol{\Psi_\delta\theta} = \sum_{k=0}^{K-1}\boldsymbol{\Psi_{\delta_k}}\theta_k$. Here $\boldsymbol{\Psi_\delta}$ is a perturbed form of $\boldsymbol{\Psi}$, and δ_k denotes the difference between the k^{th} off-grid frequency and its nearest grid-point. The problem can be extended to a compressive setting, where we have measurements of the form $\boldsymbol{y} = \boldsymbol{\Phi}\sum_{k=0}^{K-1}\boldsymbol{\Psi_{\delta_k}}\theta_k$. In this (compressive) basis mismatch problem, the perturbations are in $\boldsymbol{\Psi}$ and not in $\boldsymbol{\Phi}$, unlike in our paper where the perturbations are in $\boldsymbol{\Phi}$. This leads to the following major points of difference:

1. In the basis mismatch problem, the number of δ values is equal to the signal dimension N (or in some variants, equal to $\|\boldsymbol{\theta}\|_0$), unlike the problem in this paper where it is equal to M (or P if we count perturbation parameters in $\boldsymbol{\beta}$).

2. Moreover, unless $\boldsymbol{\Phi\Psi}$ is orthonormal (which is not possible in a compressive setting), the different δ values cannot be solved through independent searches in the basis mismatch problem and require block coordinate descent for optimization. This is in contrast to the problem in this paper (See Algorithm 1 and its modification).

3. In the basis mismatch problem, the performance is affected by the minimal separation between the frequencies of the columns of $\boldsymbol{\Psi}$ [25]. Moreover, using a dictionary $\boldsymbol{\Psi}$ with an increased frequency resolution makes the problem more under-determined and increases the mutual coherence of the matrix $\boldsymbol{\Phi\Psi}$), adversely affecting the bounds on compressive recovery. There is no such specification for minimal frequency separation in our problem.

4. A Taylor approximation approach in the basis mismatch problem would yield a system of equations of the form

$$\boldsymbol{y} = (\boldsymbol{F} + \boldsymbol{F'\Delta})\boldsymbol{x} + \boldsymbol{\eta_{Taylor}}, \tag{5}$$

where $\boldsymbol{x}$ and $\boldsymbol{\Delta x}$ are vectors with the same support, $\boldsymbol{F}$ represents the Fourier measurement matrix at known frequency set $\{u_i\}_{i=1}^M$, $\boldsymbol{F'}$ is the first derivative of the Fourier matrix w.r.t. $\boldsymbol{\delta}$, $\boldsymbol{\Delta} \triangleq \text{diag}(\boldsymbol{\delta})$ and $\boldsymbol{\eta_{Taylor}}$ represents error due to truncation of the Taylor series. This allows for simultaneous estimation

of $\boldsymbol{x}$ and $\boldsymbol{\Delta x}$ using joint sparsity [11, 12]. For our problem, the Taylor expansion leads to equations of the form:

$$\boldsymbol{y} = \boldsymbol{F_t x} \approx (\boldsymbol{F} + \boldsymbol{\Delta F'})\boldsymbol{x} + \boldsymbol{\eta_{Taylor}}. \tag{6}$$

Here, we notice that even if $\boldsymbol{x}$ is sparse, the vector $\boldsymbol{F'x}$ (and hence $\boldsymbol{\Delta F'x}$) is not sparse. Hence a joint-sparsity model cannot be directly used for our problem.

The DOA estimation techniques in [11, 13, 12, 36] and the synthetic aperture radar (SAR) target location estimation technique in [30] (see Eqns. (3) and (13) of [30]) are also related to the basis mismatch problem, and use the aforementioned joint sparsity. The DOA estimation technique follows the model $\boldsymbol{y} = \boldsymbol{A_{(d+\delta)}\theta}$ where $\boldsymbol{d}$ is a vector that contains parameters that represent the N different grid-aligned directions. The j^{th} column of $\boldsymbol{A_{(d+\delta)}}$ is given as $a_l(d_j, \delta_j) = \frac{1}{\sqrt{n}} \exp(\iota\pi(d_j + \delta_j)(l - (M+1)/2))$ where $l = 0, ..., M-1$ and $j = 0, ..., N-1$ and $\iota \triangleq \sqrt{-1}$ (see for example, Section III-F of [11]). Here again, the number of δ values is equal to N similar to the basis mismatch problem.

Likewise, consider the problem of Fourier compressive recovery under correctly specified frequencies, but with *unknown motion* of the scene $\boldsymbol{x}$, as considered in [37, 38]. We refer to this as the 'unknown motion problem' henceforth. Approaches such as [39] to solving the unknown motion problem have been proposed in the literature, but they do not exploit signal sparsity which is a key ingredient of most compressive recovery techniques, which is what we are specifically interested in. Assume for simplicity that $\boldsymbol{x}$ is in 1D and is undergoing translational motion δ_t in the t^{th} frame, where $1 \leq t \leq T$. Now, the i^{th} Fourier measurement of the t^{th} frame is given as follows (using the Fourier-shift theorem):

$$y_{it} = \sum_l x(l + \delta_t) \exp(-\iota 2\pi u_i l/N) = \exp(-\iota 2\pi u_i \delta_t/N) \sum_l x(l) \exp(-\iota 2\pi u_i l/N). \tag{7}$$

Comparing Eqn. 7 with the earlier Eqn. 2, the difference between the two computational problems is evident. In the unknown motion problem, the number of unknown perturbation parameters is equal to the number of frames T and each perturbation parameter δ_t is common to all Fourier measurements of the t^{th} frame. By the Fourier shift theorem, each frequency measurement is a phase-shifted version of the Fourier transform of the scene $\boldsymbol{x}$ in some reference or 'canonical' position. These phase-shifts can be interpreted as unknown sensor gains. Hence the unknown motion problem is related to the gain calibration problem from [19, 20, 21, 18]. As against this, the measurements in the frequency perturbations problem considered in this paper are the Fourier transforms of phase-modulated versions of the original signal $\boldsymbol{x}$, as is evident from Eqn. 2. These *cannot* be expressed as a sensor gain calibration problem. Also the number of perturbation parameters in the problem considered in this paper, is proportional to M, or to the number of groups of measurements (P). Lastly, the unknown motion problem can benefit from additional priors such as smoothness of motion, which are not relevant for the problem considered here. We summarily emphasize that the frequency perturbations problem cannot be posed as a problem of compressed sensing with unknown object motion.

5. Experimental Results

5.1. Recovery of 1-D signals

We present recovery results on signals in a multitude of cases below, using the modified version of Algorithm 1 (i.e. with a replacement of step 9 as described in the previous section). In each chart (see Figures 1,2,3,4), 1D signals of $N = 101$ elements were used, the sparsity $s \triangleq \|\boldsymbol{x}\|_0$ of the signal was varied along the x-axis, and the number of measurements M was varied along the y-axis. The cell at the intersection depicts the relative recovery error (RRMSE), $\frac{\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}\|_2}$, averaged across 5 different signals, where $\hat{\boldsymbol{x}}$ denotes the estimate of the signal $\boldsymbol{x}$. For any sparsity level, the signals were generated using randomly chosen supports with random values at each index in the support. The supports were chosen uniformly at random i.e. we generated permutations of indices from 1 to N where N is the number of elements in the signal, and then chose the first s elements of the permutation for the signal support. Thus, different signals had different supports. The value at each index in the support was drawn from Unif$(0, 1)$. The base frequencies $\boldsymbol{u}$ for the M Fourier compressive measurements for each signal were chosen uniformly randomly from $\{-N/2, -N/2+1, ..., N/2\}$. Each base frequency was subjected to a single perturbation chosen from Uniform$[-r, +r]$, for two separate cases with $r = 1$ and $r = 0.5$ respectively. (See Section 2 for the meaning of r.) Note that the same M base frequencies $\boldsymbol{u}$ for the Fourier sensing matrix were chosen for each signal, but the specific perturbations $\boldsymbol{\delta}$ were chosen differently for each signal. In Figures 1,2,3,4, black (RGB (0,0,0)) indicates perfect recovery, and white (RGB (1,1,1)) indicates recovery error of 100% or higher. Note that all the figures show error values plotted on the same scale, and hence the shades are comparable within and across figures. In all experiments, a multi-start strategy with 10 starts was adopted. In principle, we can avoid ambiguity in the estimation of the $\boldsymbol{\delta}$ values only if r is less than half the smallest difference between the selected base frequencies. However even relaxation of this condition did not have any major adverse effect on the signal reconstruction. Note that the regularization parameter λ in Eqn. 3 was chosen by cross-validation on a small 'training set' of signals. The same λ was used in all experiments. For our implementation, we used the CVX package[3].

Fig. 1 shows results for two different cases (top and bottom figures, for both $r = 1$ and $r = 0.5$), where the number of unique values in $\boldsymbol{\delta}$ is 2 and 10 respectively (this is henceforth denoted as $\delta_{(u)}$), although there are M measurements. (In this experiment, the perturbation parameters in $\boldsymbol{\beta}$ are the same as the perturbation values in $\boldsymbol{\delta}$.) In both cases, no external noise was added to the measurements. One can see that the average recovery error decreases with the number of measurements and increases with s, although the relationship is not strictly monotonic. Fig. 2 shows the same two cases as in Fig. 1, but with an addition of zero mean i.i.d. Gaussian noise with $\sigma = 5\%$ of the average magnitude of the individual (noiseless) measurements. The same trend of decrease in error with increased number of measurements and increase in error with increased

[3] `http://cvxr.com/cvx/`

s is observed here as well. For reference, we also include a typical sample reconstruction in 1D canonical basis for a signal of length 101, which is 10-sparse, in Fig. 9. Fig. 3 shows similar results as in Fig. 2 but using signals that are sparse in the Haar wavelet basis instead of the canonical basis.

Note that Algorithm 1 neither uses nor requires any assumptions about the distribution of the values in $\boldsymbol{\delta}$ or $\boldsymbol{\beta}$. The uniform distribution for $\boldsymbol{\delta}$ was used only to enable generation of *some* values in $\boldsymbol{\delta}$. We could as well have chosen any other distribution without affecting Algorithm 1.

5.2. Baselines for Recovery of 1-D signals

For comparison, we also establish two baselines:

1. A naive reconstruction algorithm (termed 'Baseline 1'), which ignores the perturbations and recovers the signal using a straightforward basis pursuit approach, with the *unperturbed, on-grid* Fourier matrix as the measurement matrix, i.e. assuming $\boldsymbol{\delta} = \mathbf{0}$. Results in similar settings as in Fig. 1 are shown in Fig. 4. The parameter λ for this approach was set using cross-validation on a training set of signals.

2. A Taylor approximation approach (termed 'Baseline 2'): Here, the signal as well as the perturbations are recovered using an alternating minimization algorithm based on a first order Taylor approximated formulation, from Eqn. 6. Results in similar settings as in Fig. 1 are shown in Fig. 5 for two cases: one where $\delta_{(u)} = 2$ and another where $\delta_{(u)} = 10$ (recall that $\delta_{(u)}$ is the number of unique values in $\boldsymbol{\delta}$). This baseline is similar in spirit to the truncated Taylor series approach presented in [11, 30] but modified for our (very different) computational problem. The parameter λ for this approach was again set using cross-validation on a training set of signals.

As is clear from the figures, Baseline 1 performs considerably worse, since inaccurate frequencies are trusted to be accurate. Baseline 2 *also* performs badly because the first order Taylor error, $\boldsymbol{\eta_{Taylor}}$, can be overwhelmingly large since it is directly proportional to the unknown $\|\boldsymbol{x}\|_2$, and consequently, the signal recovered is also inferior. In fact, a comparison between Figures 4 and 5 reveals that in case of Taylor approximations to a perturbed Fourier matrix, the results obtained are often as bad as those obtained when assuming $\boldsymbol{\delta} = \mathbf{0}$. Baseline 2 is akin to a strategy used in [11, 30] and applied to DOA estimation or in target detection in radar. However the specific inverse problem to be solved in these papers is similar to a problem of mismatched representation bases, which fortuitously allows for joint sparsity of $\boldsymbol{x}$ and $\boldsymbol{\delta} \cdot \boldsymbol{x}$ (see Eqn.5 in Section 4), but which cannot be achieved in the problem we attempt to solve in this paper.

5.3. Performance w.r.t. Perturbation Value

Here, we consider the case of a *single* perturbation value δ, and perform an empirical study of the variation of reconstruction quality (RRMSE) for Algorithm 1 w.r.t. change in M as well as $|\delta|$. The results

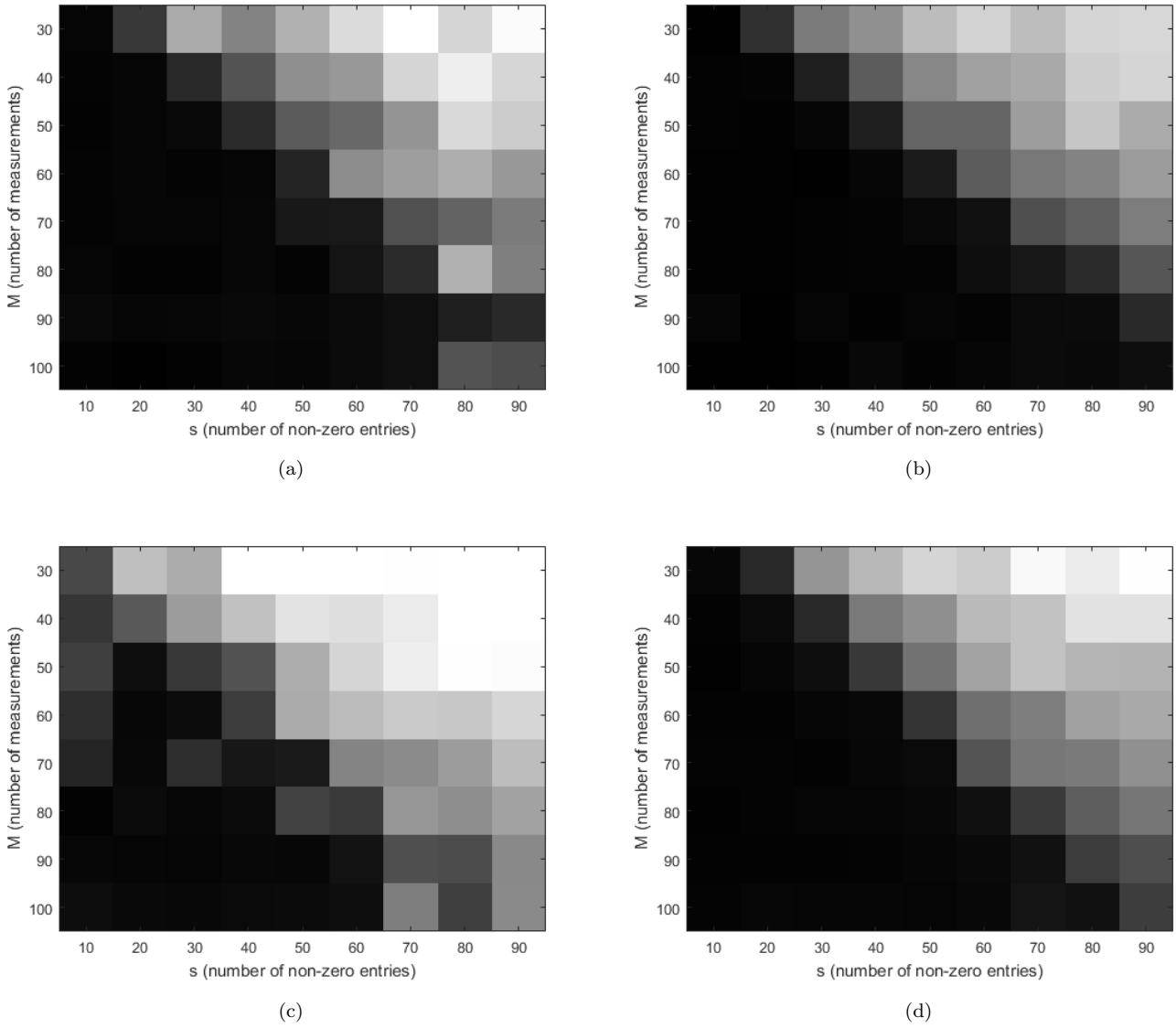


Figure 1: Relative recovery errors (RRMSE) with Algorithm 1 for a 1D signal with 101 elements, sparse in canonical basis, no measurement noise added, for: (a) $r = 1$, $\delta_{(u)} = 2$, (b) $r = 0.5$, $\delta_{(u)} = 2$, (c) $r = 1$, $\delta_{(u)} = 10$, (d) $r = 0.5, \delta_{(u)} = 10$, where $\delta_{(u)}$ represents number of unique values in $\boldsymbol{\delta}$ and r is the maximum allowed absolute value of δ.

are reported on a signal with 128 elements, expressed as a sparse linear combination of Haar wavelet bases. The signal sparsity in the wavelet basis was $\approx 0.15 \times 128$, M was varied from 20 to 128, and $|\delta|$ was varied from 0.05 to 2.5. These results are computed for 10 starts with different initial conditions. The recovery errors for reconstructed signals corresponding to the least value of the objective function (from the 10 starts) are shown in Fig. 6 for the case of (i) no measurement noise, and (ii) noise from from $N(0, \sigma^2)$ with $\sigma = 0.05\zeta$

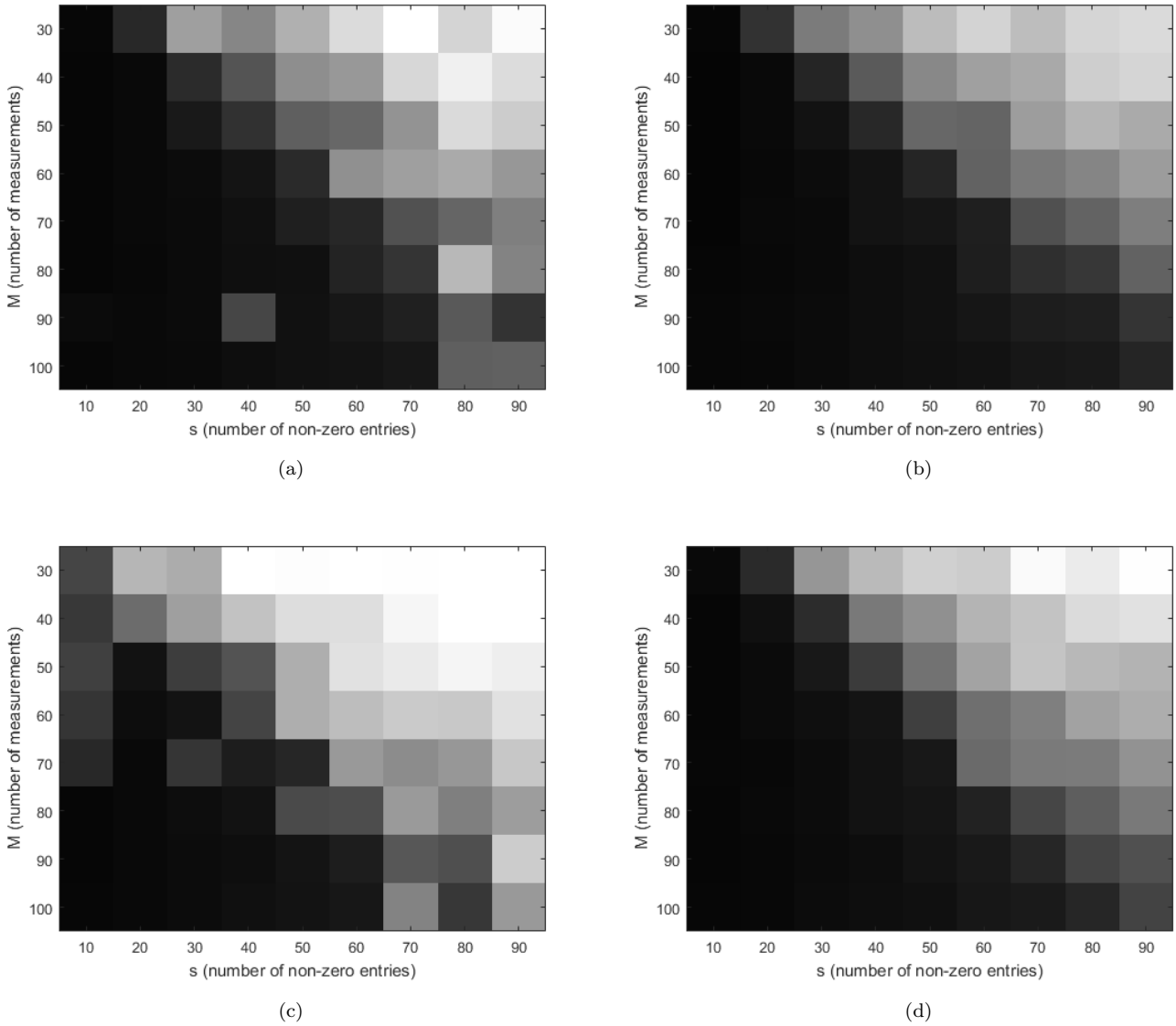


Figure 2: Relative recovery errors (RRMSE) with Algorithm 1 for a 1D signal with 101 elements, sparse in canonical basis, noise from $\mathcal{N}(0, \sigma^2)$ added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements). (a) $r = 1$, $\delta_{(u)} = 2$, (b) $r = 0.5$, $\delta_{(u)} = 2$, (c) $r = 1$, $\delta_{(u)} = 10$, (d) $r = 0.5$, $\delta_{(u)} = 10$, where $\delta_{(u)}$ represents number of unique values in $\boldsymbol{\delta}$ and allowed absolute value of δ.

where $\zeta =$ maximum of (average absolute value of real part of noiseless measurements, average absolute value of complex part of noiseless measurements). These results show good quality recovery until $\delta = 2$, beyond which we see increase in error. Likewise, there is good quality recovery with $M \geq 50$, below which there is an increase in error.

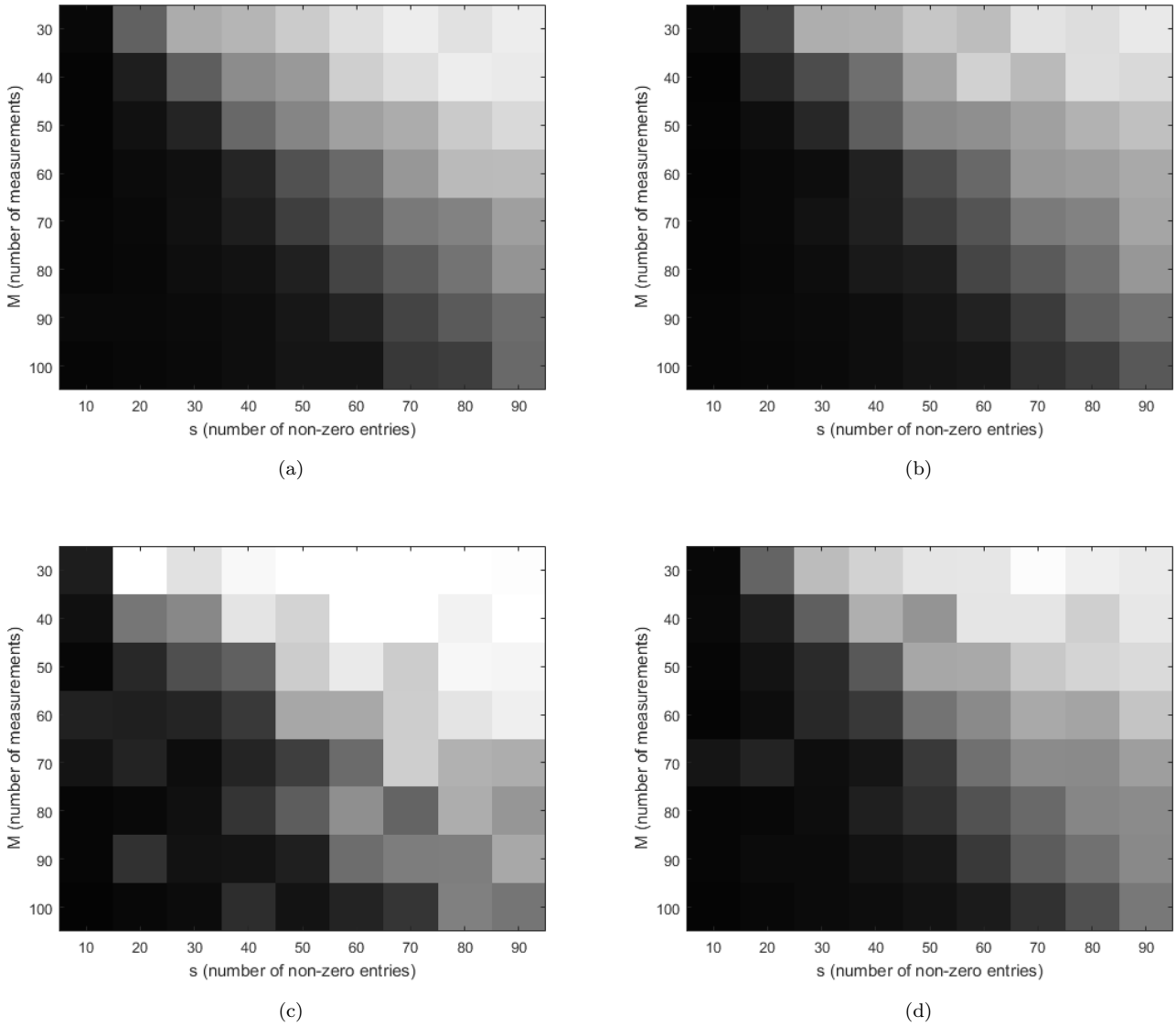


Figure 3: Relative recovery errors (RRMSE) with Algorithm 1 for a 1D signal with 128 elements, sparse in Haar DWT basis, noise from $\mathcal{N}(0, \sigma^2)$ added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements). (a) $r = 1$, $\delta_{(u)} = 2$ (b) $r = 0.5$, $\delta_{(u)} = 2$ (c) $r = 1$, $\delta_{(u)} = 10$, (d) $r = 0.5$, $\delta_{(u)} = 10$, where $\delta_{(u)}$ represents number of unique values in $\boldsymbol{\delta}$ and r is the maximum allowed absolute value of δ.

These experiments were repeated for the case when the signal was sparse in the canonical basis, keeping all other settings exactly the same. The results for this are reported in Fig. 7. Comparing with those in Fig. 6, we see that the reconstruction for the case of sparsity in the canonical basis is more resilient to larger values of δ. (Also see Lemma 1 of Sec. 6).

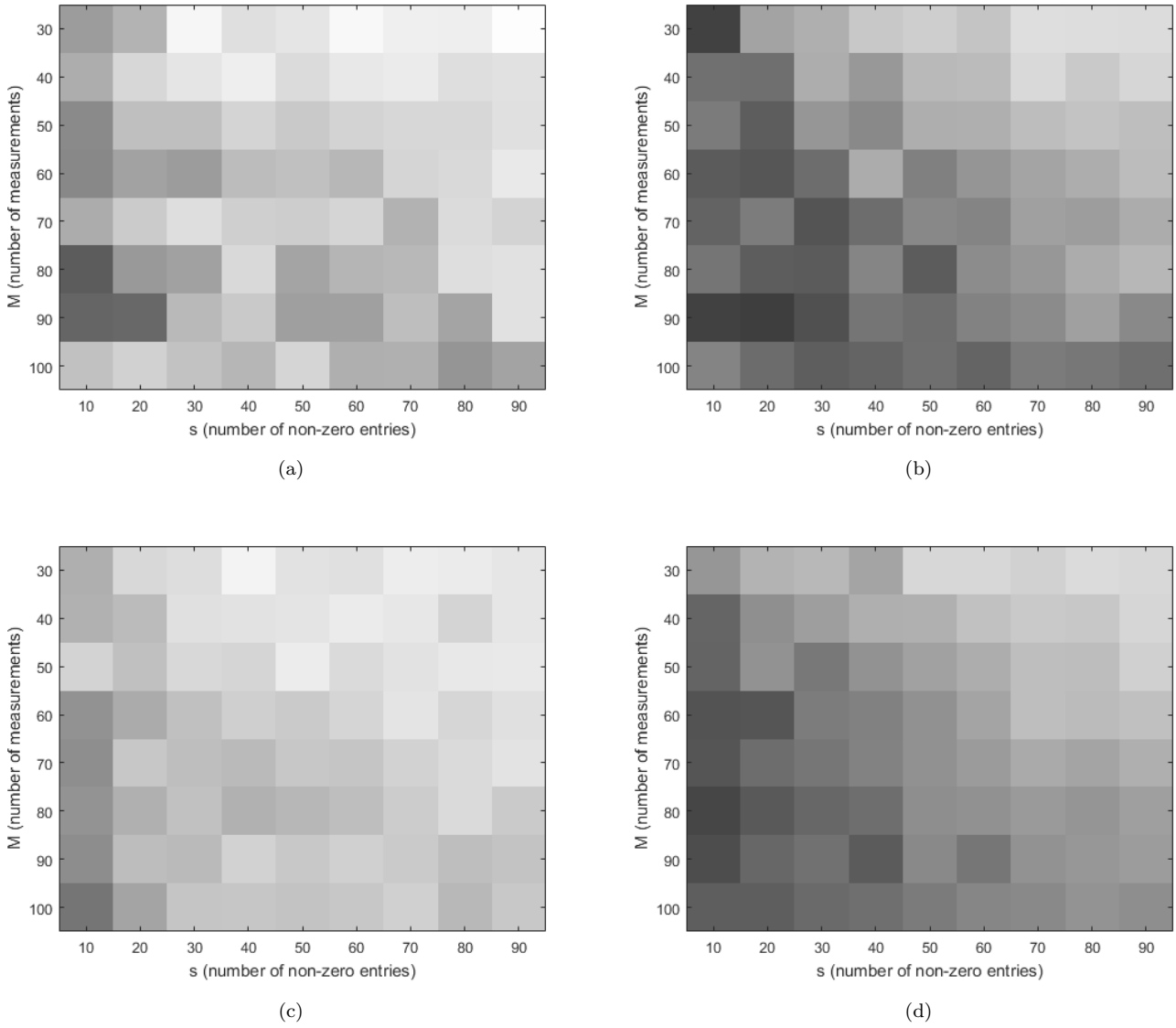


Figure 4: Relative recovery errors (RRMSE) with Baseline 1 algorithm (see text) for a 1D signal with 101 elements, sparse in canonical basis, noise from $\mathcal{N}(0, \sigma^2)$ added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements). (a) $r = 1$, $\delta_{(u)} = 2$, (b) $r = 0.5$, $\delta_{(u)} = 2$, (c) $r = 1$, $\delta_{(u)} = 10$, (d) $r = 0.5$, $\delta_{(u)} = 10$, where $\delta_{(u)}$ represents number of unique values in $\boldsymbol{\delta}$, and r is the maximum allowed absolute value allowed in $\boldsymbol{\delta}$. Compare with Fig. 2.

5.4. The case of M independent perturbations

All the experiments so far were conducted in the setting where the number of unique values in $\boldsymbol{\delta}$ was much less than M. The motivation for this setting has already been described in previous sections. In the case when each measurement has an independent perturbation, we expect the recovery error to be relatively

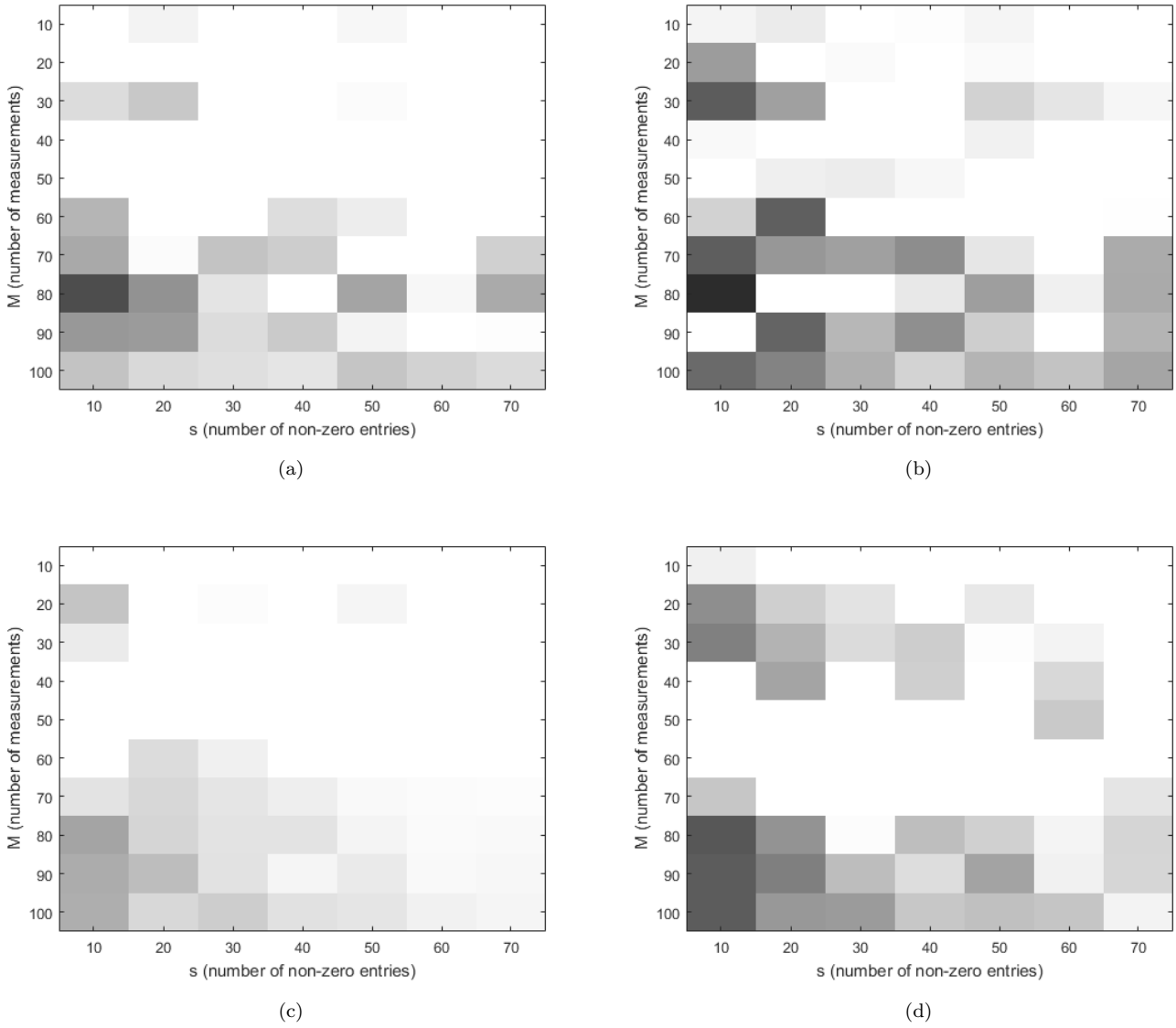
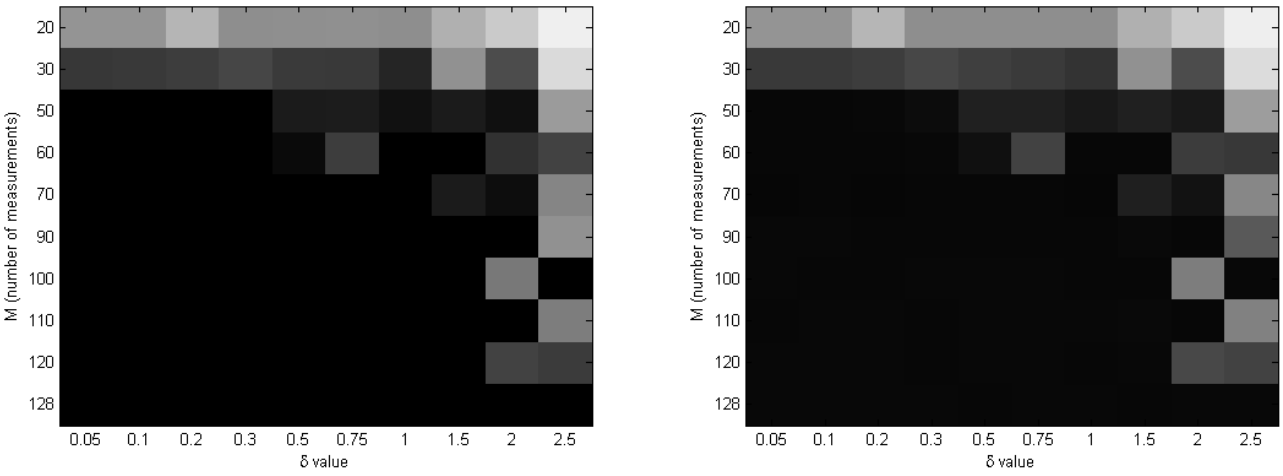


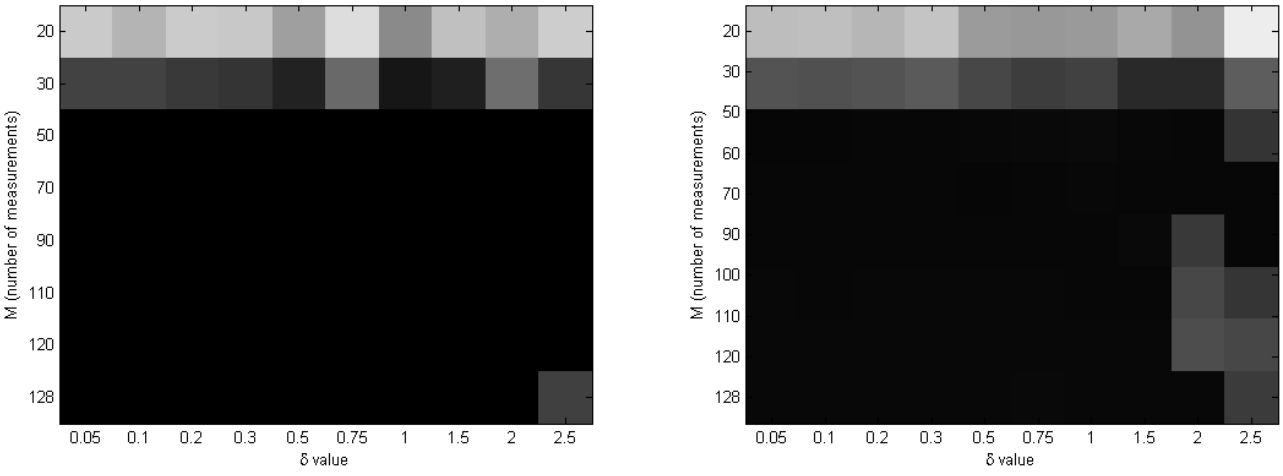
Figure 5: Relative recovery errors (RRMSE) with Baseline 2 algorithm (see text) for a 1D signal with 101 elements, sparse in canonical basis, noise from $\mathcal{N}(0, \sigma^2)$ added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements). (a) $r = 1$, $\delta_{(u)} = 2$, (b) $r = 0.5$, $\delta_{(u)} = 2$, (c) $r = 1$, $\delta_{(u)} = 10$, (d) $r = 0.5$, $\delta_{(u)} = 10$, where $\delta_{(u)}$ represents number of unique values in $\boldsymbol{\delta}$. Compare with Fig. 2.

higher, especially in the presence of measurement noise, as the number of unknowns increases significantly. For completeness, we perform similar experiments in the case when $\delta_{(u)} = M$ and plot the reconstruction errors. We observe that even with a large number of unique δ values, the errors are still low when the signal is very sparse (see Fig 8). However, the error increases when the signal is less sparse, and the error is much



(a) (b)

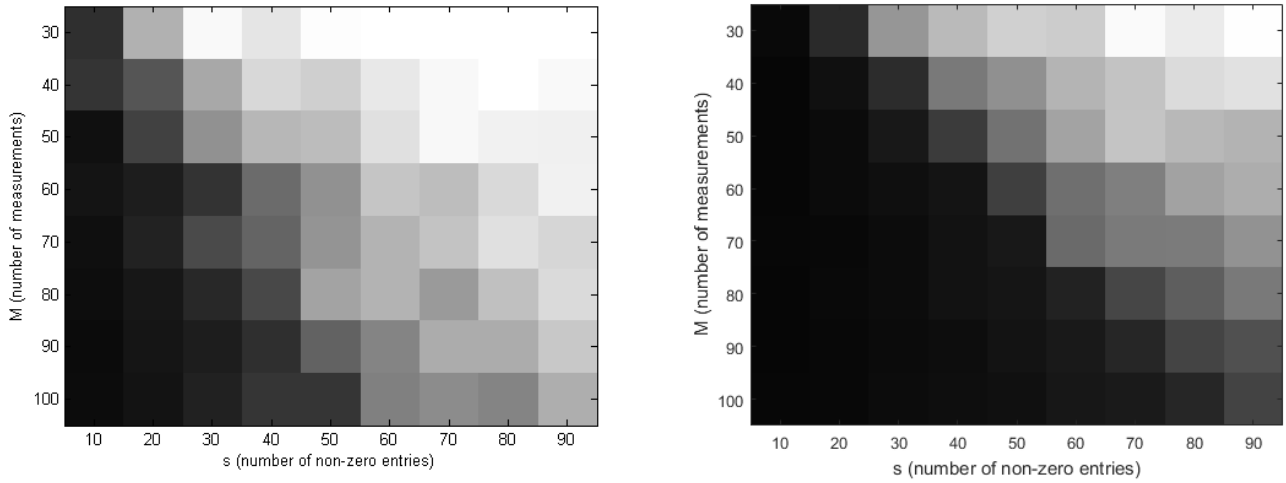
Figure 6: Variation in RRMSE w.r.t. (M, δ), reported for a 128×1 signal represented as a sparse linear combination of Haar wavelet basis vectors. Sparsity of wavelet coefficients $= 0.15 \times 128$. (a) Case with no measurement noise, (b) Case with noise drawn from $N(0, \sigma^2)$ with $\sigma = 0.05\zeta$ where $\zeta =$ maximum of (average absolute value of real part of noiseless measurements, average absolute value of complex part of noiseless measurements).



(a) (b)

Figure 7: Variation in RRMSE w.r.t. (M, δ), reported for a 128×1 signal sparse in the canonical basis, with ℓ_0 norm $= 0.15 \times 128$. (a) Case with no measurement noise, (b) Case with noise drawn from $N(0, \sigma^2)$ with $\sigma = 0.05\zeta$ where $\zeta =$ maximum of (average absolute value of real part of noiseless measurements, average absolute value of complex part of noiseless measurements).

higher than the case of a small $\delta_{(u)}$ as seen in Fig. 8.



(a) (b)

Figure 8: Relative recovery errors (RRMSE) with Proposed Alternating Minimization algorithm for a 1D signal with 128 elements, sparse in Haar DWT basis, noise from $\mathcal{N}(0, \sigma^2)$ added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements). Left: $r = 0.5, \delta_{(u)} = M$, where M is the number of measurements, Right: $r = 0.5, \delta_{(u)} = 10$.

5.5. Recovery of 2-D signals

Application of our algorithms to 2D images is natural and more immediately applicable in imaging scenarios. We first present results with a similar set of experiments using 2D images (as the signal $\boldsymbol{x}$). For this experiment, 30×30 images were used. The images were generated using a sparse linear combination of Haar wavelet bases. We used a radial sampling approach in the Fourier domain (equivalent to taking a Fourier transform of the Radon projections), taking a fixed number of measurements along each spoke, but varying the number of angles used and the sparsity of the image in the Haar basis. The angles for the spokes were incorrectly specified (which is typical in mis-calibrated tomography) with each angle error chosen from Uniform$[-2^\circ, +2^\circ]$ leading to *significant* perturbations in the frequencies. The base frequencies $\boldsymbol{u}$ were spaced uniformly along each spoke. In addition, noise from $\mathcal{N}(0, \sigma^2)$ was added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements). We used the YALL1[4] solver for optimization of $\boldsymbol{x}$ and the NUFFT package[5] for computing Fourier transforms at non-integer frequencies.

[4] http://yall1.blogs.rice.edu/

[5] `https://www-user.tu-chemnitz.de/~potts/nfft/`

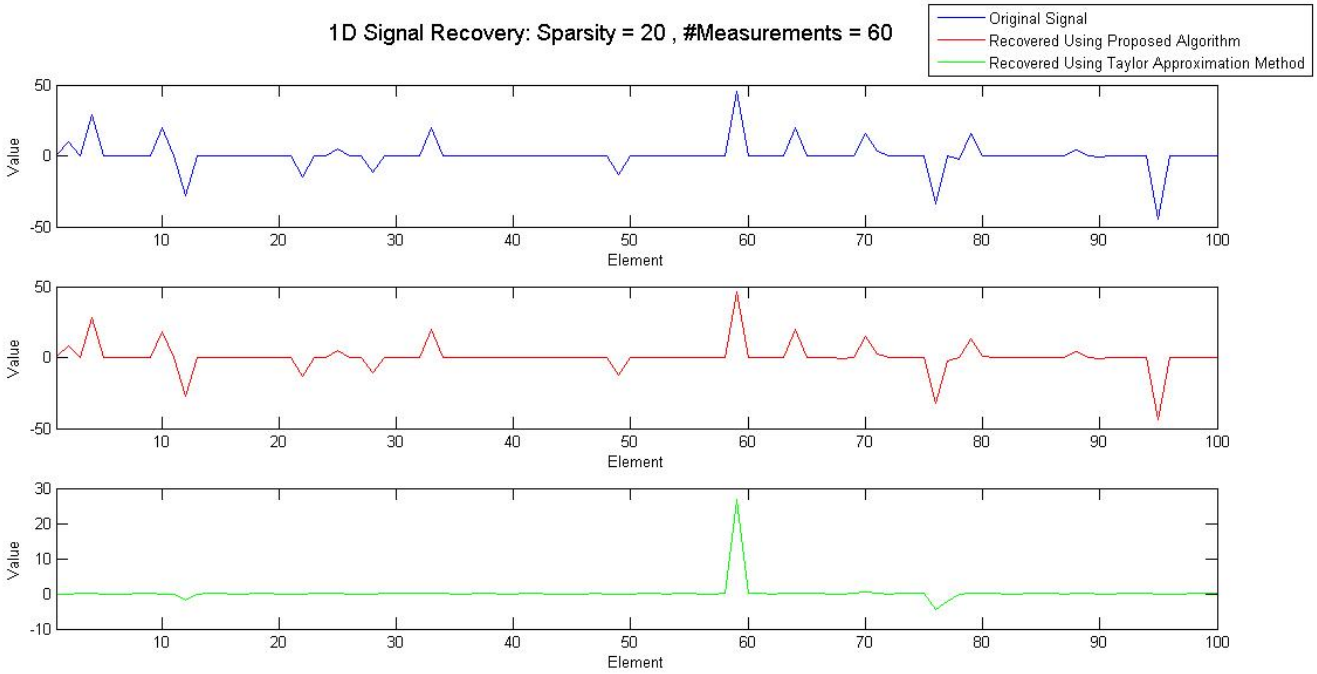


Figure 9: Sample recovery for 1D signal sparse canonical basis, $N = 100, M = 60, s = 20$, noise from $\mathcal{N}(0, \sigma^2)$ added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements). Relative reconstruction error by proposed algorithm: 5.5%. Relative reconstruction error by Baseline 2 (Taylor approximation): 88.7%.

The results are summarized in a chart shown in Fig. 10. As Fig. 10 shows, the recovery error was small, even for a reasonably small number of measurements, and the method was robust to noise in the measurements. Errors with the baseline algorithms were significantly larger and are not reported here.

In the second set of experiments, we show reconstruction results on three images each of size 200×200. Fourier measurements were simulated along 140 radial spokes with erroneously specified angles (which is typical in tomography with angle errors or unknown angles). The angle error for each spoke was chosen independently from Uniform$[-1°, +1°]$ - leading to *significant* perturbations in the frequencies. Noise from $\mathcal{N}(0, \sigma^2)$ where $\sigma \triangleq 0.05\times$ average (noiseless) measurement magnitude, was added to the real and complex parts of the measurements. During reconstruction, we exploited image sparsity in a Haar wavelet basis. Reconstruction results with the modified version of Algorithm 1 are presented in Fig. 11. In comparison with Baseline 1, we see that our algorithm performs significantly better, both visually and in terms of RRMSE values, as seen in Fig. 11. Results with a similar experiment for angle errors chosen independently from Uniform$[-2°, +2°]$ and Uniform$[-3°, +3°]$ are shown in Fig. 12 and Fig. 13 respectively, showing clear performance improvement of our method over Baseline 1. Errors with Baseline 2 were very high and hence

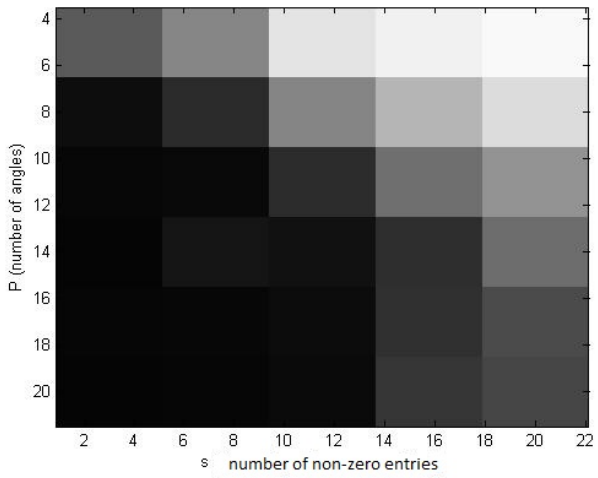


Figure 10: Relative recovery errors (RRMSE) error for 30×30 2D image, sparse in 2D Haar Wavelet basis, with noise from $\mathcal{N}(0, \sigma^2)$ added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements); and angle errors from Uniform$[-2°, +2°]$

not reported here.

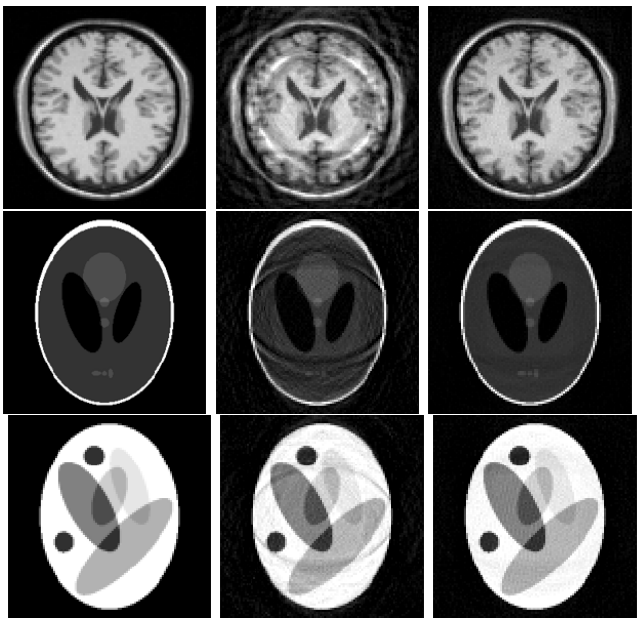


Figure 11: Reconstruction for 200×200 images with noise from $\mathcal{N}(0, \sigma^2)$ added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements), 70% compressive measurements, angle error from Uniform$[-1°, +1°]$. In each row, left: original image, middle: reconstruction using Baseline 1 (RRMSE $25\%, 23.36\%, 8.82\%$), right: reconstruction using modified version of Algorithm 1 (RRMSE $6.76\%, 5.27\%, 4.5\%$).

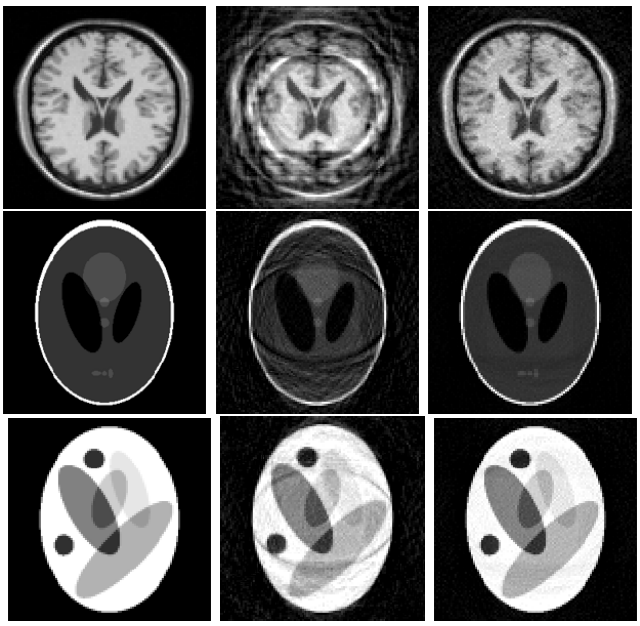


Figure 12: Reconstruction for 200×200 images with noise from $\mathcal{N}(0, \sigma^2)$ added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements), 70% compressive measurements, angle error from Uniform$[-2^\circ, +2^\circ]$. In each row, left: original image, middle: reconstruction using Baseline 1 (RRMSE $38.7\%, 30.98\%, 12.63\%$), right: reconstruction using modified version of Algorithm 1 (RRMSE $10.65\%, 5.22\%, 4.87\%$).

6. Theoretical Results

While the empirical results show Algorithm 1 working well across a large number of simulated scenarios, we also characterize the formulation by providing theoretical analysis for its convergence. Moreover, Algorithm 1 is one possible method to obtain the solution for the underlying *computational problem*. Hence, there are larger questions such as conditions for the uniqueness of the solution to the main *computational problem*, which we analyze in this section under some conditions.

6.1. Convergence of Algorithm 1

Here we analyze the convergence of Algorithm 1 (or its modified version) under a specific condition mentioned further. Let $\boldsymbol{F}(\boldsymbol{\delta})$ denote the Fourier transform computed at the frequencies values $\boldsymbol{u} + \boldsymbol{\delta}$ where $\boldsymbol{\delta} = h(\boldsymbol{\beta}, \boldsymbol{u})$. Assign $\boldsymbol{z} = \{\boldsymbol{x}, \boldsymbol{\beta}\}$. Recall that our objective is to determine the solution $\boldsymbol{z}^*$ that minimizes the objective function $J(\boldsymbol{z}) \triangleq \|\boldsymbol{x}\|_1 + \lambda \|\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{\delta})\boldsymbol{x}\|_2$, namely $\boldsymbol{z}^* = \text{argmin}_{\boldsymbol{z}} J(\boldsymbol{z})$.

Let $\boldsymbol{z_t} = \{\boldsymbol{x_t}, \boldsymbol{\beta_t}\}$ be the present solution of our alternating search algorithm at iteration t. Our alternating search algorithm ensures that the sequence of function values $\{J(\boldsymbol{z_t})\}_{t \in \mathbb{N}}$ is monotonically decreasing. As J is bounded below by 0, the sequence $\{J(\boldsymbol{z_t})\}_{t \in \mathbb{N}}$ converges to a limit value $E \in \mathbb{R}^+$ by the monotone convergence theorem.

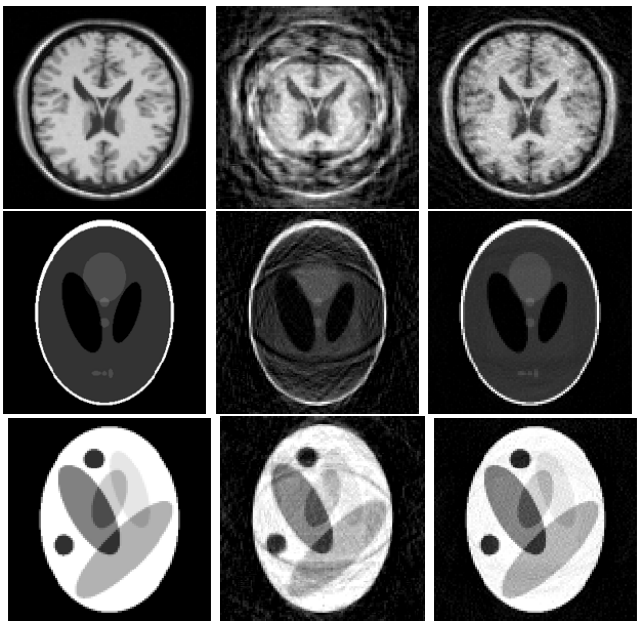


Figure 13: Reconstruction for 200×200 images with noise from $\mathcal{N}(0, \sigma^2)$ added to measurements where $\sigma = 0.05 \times \zeta$ where $\zeta \triangleq$ max(average absolute value of real component of noiseless measurements, average absolute value of complex component of noiseless measurements), 70% compressive measurements, angle error from Uniform$[-3°, +3°]$. In each row, left: original image, middle: reconstruction using Baseline 1 (RRMSE $38.75\%, 35.48\%, 14.59\%$), right: reconstruction using modified version of Algorithm 1 (RRMSE $13.15\%, 5.29\%, 5.85\%$).

However, this does not yet establish the convergence of the solution sequence $\{\boldsymbol{z_t}\}$, which is much more involved. We refer the reader to the supplemental material for a partial proof of the convergence of the solution iterates.

6.2. Uniqueness of Solution

It is quite natural to question whether the recovery of $\boldsymbol{x}$ from compressive measurements of the form $\boldsymbol{y} = \boldsymbol{F}(\boldsymbol{\delta})\boldsymbol{x}$ is unique, where $\boldsymbol{F}(\boldsymbol{\delta})$ is as defined in Eqn. 1. We answer this question in the affirmative (in the noiseless case, of course) for real-valued $\boldsymbol{x}$ under the condition that the perturbation parameters $\boldsymbol{\beta}$ are bounded and also independent of the base frequencies $\boldsymbol{u}$, i.e. $\forall k, 1 \leq k \leq P, \delta_k = \tilde{h}(\beta_k)$ where $\tilde{h}$ is a known function of only β_k. We comment on the effect of relaxing this condition, at the end of the section.

First consider real-valued $\boldsymbol{x}$, which is typical in tomography and certain protocols in MR (if the magnetization is proportional to the contrast-weighted proton density [40]). Consider the case where there is only a single unknown perturbation parameter value β in all measurements and where $\boldsymbol{x}$ is a 1D signal. Then, we state and prove the following Lemma.

Lemma 1. *Consider measurements of the form $\boldsymbol{y} = \boldsymbol{F}(\boldsymbol{\delta})\boldsymbol{x}$, where (i) $\boldsymbol{x} \in \mathbb{R}^N$ is an s-sparse vector, (ii) $\boldsymbol{y}$ contains at least $\mathcal{O}(s \log^2 s \log N)$ measurements, and (iii) $\boldsymbol{F}(\boldsymbol{\delta})$ has a single perturbation parameter β such*

that frequency perturbation $\delta = \tilde{h}(\beta)$ for a bijective function $\tilde{h}$. Then $\boldsymbol{x}$ and β can be uniquely *recovered from $\boldsymbol{y}, \boldsymbol{F}(\boldsymbol{\delta})$ with high probability, if $\delta < N/4$.* ♢

Proof: The measurement vector can be expressed as $\boldsymbol{y} = \boldsymbol{F}(\boldsymbol{\delta})\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{x} \cdot \boldsymbol{v}_\beta)$, where '$\cdot$' refers to an element-wise product, $\boldsymbol{v}_\beta$ is a vector in $\mathbb{C}^N$ whose l^{th} entry is equal to $\exp(-\iota 2\pi \tilde{h}(\beta) l/N)$ where l is a spatial/time index and $\iota = \sqrt{-1}$. To see this, consider the i^{th} measurement as follows:

$$y_i = \frac{1}{\sqrt{M}} \sum_l \exp(-\iota 2\pi (u+\delta) l/N) x(l) \tag{8}$$
$$= \frac{1}{\sqrt{M}} \sum_l \exp(-\iota 2\pi u l/N)(x(l) \exp(-\iota 2\pi \tilde{h}(\beta) l/N)).$$

Let $\boldsymbol{x}_\beta \triangleq \boldsymbol{x} \cdot \boldsymbol{v}_\beta$. Using standard compressive sensing results from [1], we can prove unique recovery of $\boldsymbol{x}_\beta$ using standard basis pursuit, i.e. by solving the problem $\min \|\boldsymbol{x}\|_1$ such that $\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x}_\beta$. This holds for an RIP-obeying $\boldsymbol{F}$ (which is true with high probability if the M base frequencies were chosen uniformly at random, and $M \geq s \log^2 s \log N$ for s-sparse $\boldsymbol{x}$ [41]).

Since $\boldsymbol{x}$ is real-valued, we now show that $\boldsymbol{x}$ and β can be uniquely recovered if $\delta < N/4$. We do so by contradiction. Suppose that $\boldsymbol{x}_\beta$ can be decomposed into two solutions of the form $x_1(l) \exp(-\iota 2\pi \tilde{h}(\beta_1) l/N) = x_2(l) \exp(-\iota 2\pi \tilde{h}(\beta_2) l/N), \forall l, 0 \leq l < N$. Let $\delta_1 \triangleq \tilde{h}(\beta_1), \delta_2 \triangleq \tilde{h}(\beta_2)$. Then we can write:

$$\forall l, x_2(l) = x_1(l) \exp(-\iota 2\pi l(\delta_1 - \delta_2)/N). \tag{9}$$

As both $x_1(l)$ and $x_2(l)$ are real-valued, this implies that the phase factor on the RHS is 0. This implies the following, given some integer k_l:

$$2\pi l(\delta_1 - \delta_2)/N = k_l \pi \implies l(\delta_1 - \delta_2)/N = k_l N/2. \tag{10}$$

When $l = 0$, we clearly have $x_2(l) = x_1(l)$. For any other l, we have $\delta_1 - \delta_2 = k_l N/(2l)$. When $|\delta_1| < N/(4l)$ and $|\delta_2| < N/(4l)$, then $|\delta_1 - \delta_2| < N/(2l)$. But from Eqn. 10, we have $|\delta_1 - \delta_2| = |k_l| N/(2l)$. This in turn implies that $|k_l| N/(2l) < N/(2l)$. As k_l is an integer, this means that $k_l = 0$. But we need to establish that this is true for all l. Note that in particular for $l = 1$, we see that when $|\delta_1| < N/4$ and $|\delta_2| < N/4$, we have $\delta_1 = \delta_2$. It is easy to see that $l = 1$ yields the least stringent condition on δ. With $\delta_1 = \delta_2$, i.e. $\tilde{h}(\beta_1) = \tilde{h}(\beta_2)$, we see that $x_2(l) = x_1(l)$ for all l. Thus we have proved the unique recovery of $\boldsymbol{x}$ and δ under the given condition. If $\tilde{h}$ is a bijective function, we can recover β from δ. This proves the lemma. ♢

Comments about Lemma 1:

1. From Lemma 1, it is clear that a standard basis pursuit algorithm can be used for the recovery of $\boldsymbol{x}_\beta$. This recovery is robust to measurement noise and compressibility (instead of strict sparsity) of $\boldsymbol{x}_{\boldsymbol{\beta}}$ and the bounds from [1] would follow.

2. This result extends to $\boldsymbol{x}$ in higher dimensions as well.

3. If $\boldsymbol{x} \in \mathbb{C}^N$, although $\boldsymbol{x}_\beta$ can be recovered uniquely, there is an inevitable phase ambiguity in estimating $\boldsymbol{x}$. However, the magnitude of each element of $\boldsymbol{x}$, i.e. $|\boldsymbol{x}|$, can be recovered uniquely under the aforestated conditions.

4. Now consider the case of $P \ll M$ unique perturbation parameter values in $\boldsymbol{\beta}$, and as before $\boldsymbol{x} \in \mathbb{R}^N$ where $\boldsymbol{x}$ is s-sparse. Let $\boldsymbol{F_{L_k}}$ be the sub-matrix of measurements corresponding to a particular value β_k. Under the conditions stated earlier, unique recovery of $\boldsymbol{x}, \boldsymbol{\delta}$ can be guaranteed if for at least one $k \in \{1, 2, ..., P\}$, the matrix $\boldsymbol{F_{L_k}}$ obeys the RIP of order s. This easily follows from Lemma 1.

5. The number of measurements to guarantee unique recovery is independent of the specific value of δ as long as $\delta < N/4$. An *indication* of this lack of dependence on δ can be seen from the performance of Algorithm 1 in Fig. 7 in Sec. 5.3. Nevertheless, there are subtle but important differences between using basis pursuit and Algorithm 1. First, basis pursuit is not susceptible to local minima unlike Algorithm 1. However, basis pursuit can be used only for the *specific* case of a single δ value and signal sparsity in the canonical basis. On the other hand, Algorithm 1 can achieve recovery for arbitrary $\delta_{(u)}$ and signal sparsity in any orthonormal basis.

These uniqueness results can be further strengthened (i.e. in terms of weaker conditions on the number of measurements), by observing that in the case of $P > 1$ unique values in $\boldsymbol{\beta}$, we need to recover different (complex) signals $\boldsymbol{x}_{\beta_1}, \boldsymbol{x}_{\beta_2}, ..., \boldsymbol{x}_{\beta_P}$ where $\forall i, 1 \leq i \leq P, \boldsymbol{x}_{\beta_i} \triangleq \boldsymbol{x} \cdot \boldsymbol{v}_{\beta_i}$ and the l^{th} entry of $\boldsymbol{v}_{\beta_i}$ equals $\exp(-\iota 2\pi \tilde{h}(\beta_i) l/N)$. All these signals are s-sparse if $\boldsymbol{x}$ is s-sparse, and they have the same support. This recovery problem is therefore an example of multiple measurement vectors (MMV) [42]. However, our computational problem has further refinements to MMV: the sensing sub-matrices corresponding to the different values in $\boldsymbol{\beta}$ are necessarily different, which is termed the generalized MMV (GMMV) problem [43],[44], for which stronger results exist. For example, we modify Theorem 1 of [43] which guarantees unique recovery of the sparsity pattern of the signals, to state the following Lemma:

Lemma 2. *Consider measurements $\boldsymbol{y_{L_k}}$ where $\forall k, 1 \leq k \leq P, \boldsymbol{y_{L_k}} = \boldsymbol{F_{L_k}} \boldsymbol{x}_{\beta_k}$ where $\boldsymbol{x}_{\beta_k} = \boldsymbol{x} \cdot \boldsymbol{v}_{\beta_k}$ and $v_{\beta_k}(l) = \exp(-\iota 2\pi \tilde{h}(\beta_k) l/N)$. Assume that $\boldsymbol{x}$ is s-sparse with support set denoted $\boldsymbol{\mathcal{S}}$ and has sub-Gaussian entries*[6]*. Assume that the following conditions hold:*

$$\forall j \notin \mathcal{S}, \left(\frac{1}{P} \sum_{k=1}^{P} \| \boldsymbol{F}^{\dagger}_{\boldsymbol{L_k}, \boldsymbol{\mathcal{S}}} \boldsymbol{F}_{\boldsymbol{L_k}, \boldsymbol{j}} \|^2 \right)^{0.5} \leq \alpha_1 < 1 \tag{11}$$

$$\forall j \notin \mathcal{S}, max_{k \in \{1,...,P\}} \| \boldsymbol{F}^{\dagger}_{\boldsymbol{L_k}, \boldsymbol{\mathcal{S}}} \boldsymbol{F}_{\boldsymbol{L_k}, \boldsymbol{j}} \|_2 \leq \alpha_2 > 0,$$

[6] This includes bounded random variables as a special case.

where † denotes the pseudo-inverse, $\boldsymbol{F}_{\boldsymbol{L_k},\boldsymbol{j}}$ is the j^{th} column of $\boldsymbol{F}_{\boldsymbol{L_k}}$ and $\boldsymbol{F}_{\boldsymbol{L_k},\boldsymbol{\mathcal{S}}}$ is a sub-matrix of $\boldsymbol{F}_{\boldsymbol{L_k}}$ with columns corresponding to entries in $\mathcal{S}$. Then the solution to the following optimization problem (Q1) is able to recover the exact solution for the signals $\boldsymbol{x}_{\beta_1}, \boldsymbol{x}_{\beta_2}, ..., \boldsymbol{x}_{\beta_P}$ with high probability decreasing in α_1, α_2 and increasing in P. The problem (Q1) is defined as follows: $min \|\boldsymbol{x}\|_1$ s. t. $\forall k \in \{1, ..., P\}$ $\boldsymbol{y}_{\boldsymbol{L_k}} = \boldsymbol{F}_{\boldsymbol{L_k}} \boldsymbol{x}_{\beta_k}$. $\diamondsuit$

We refer to [43] and [45] (Theorem 23, Chapter 5) for the proof of Lemma 2.

Comments about Lemma 2:

1. The two conditions on $\boldsymbol{F}_{\boldsymbol{L_k},\boldsymbol{\mathcal{S}}}$ essentially imply that the norm of $\boldsymbol{F}^{\dagger}_{\boldsymbol{L_k},\boldsymbol{\mathcal{S}}} \boldsymbol{F}_{\boldsymbol{L_k},\boldsymbol{j}}$ should be small, on an average over k. The sufficient conditions under which this holds true are mentioned in Proposition 5.2 of [46]. These sufficient but not necessary conditions are that every $\boldsymbol{F}_{\boldsymbol{L_k}}$ should obey RIP of order $s+1$ with a RIC (restricted isometry constant) of less than 0.5. It has been shown that Fourier sensing matrices with at least $\mathcal{O}(s \log^2 s \log n)$ randomly chosen frequencies obey the RIP with a high probability of $1 - 2^{-\Omega(\log N \log 2s)}$ and an RIC of less than or equal to 0.5 (see Theorem 4.5 of [41]).

2. The GMMV results require weaker conditions than MMV (see eqn. 12 of [43]). The MMV results require the one and only measurement matrix $\boldsymbol{F}$ to obey the RIP. However, the GMMV results state that among the different measurement matrices $\{\boldsymbol{F}_{\boldsymbol{L_1}}, \boldsymbol{F}_{\boldsymbol{L_2}}, ..., \boldsymbol{F}_{\boldsymbol{L_P}}\}$, a few can be 'bad' (i.e. not obey RIP), as long as the ensemble of measurement matrices is 'well-behaved', i.e. obeys the two conditions in Eqn. 12.

3. Note that the specific optimization problem used in [43] for GMMV (termed there as 'LOPT') requires penalization of $\sum_{i=0}^{n-1} \sqrt{\sum_{k=1}^{P} |\boldsymbol{x}_{\beta_k}(i)|^2}$, i.e. an $\ell_{2,1}$ norm. It is proportional to $\|\boldsymbol{x}\|_1$ in our case, given the specific definition of $\boldsymbol{x}_{\beta_k}$ here.

4. With this, we have established recovery of the signals $\boldsymbol{x}_{\beta_1}, \boldsymbol{x}_{\beta_2}, ..., \boldsymbol{x}_{\beta_P}$ (and thereby the recovery of $\boldsymbol{x}$ if it is real-valued, or the recovery of $|\boldsymbol{x}|$ for complex-valued $\boldsymbol{x}$) with high probability.

However, our computational problem in fact has further structure over and above GMMV. This is explained as follows:

1. Property C1: First, $\forall i, 1 \leq i \leq N, |\boldsymbol{x}_{\beta_1}(i)| = |\boldsymbol{x}_{\beta_2}(i)| = ... = |\boldsymbol{x}_{\beta_P}(i)|$.

2. Property C2: Second, the phase factors of all elements of $\boldsymbol{x}_{\beta_1}, \boldsymbol{x}_{\beta_2}, ..., \boldsymbol{x}_{\beta_P}$ are *completely determined* by just the P values in $\boldsymbol{\beta}$.

The modified version of Algorithm 1 imposes this structure by design. At this point, we conjecture that the lower bound on the required number of measurements is actually much lower, if we use Algorithm 1 for estimation of $\boldsymbol{x}, \boldsymbol{\beta}$, as compared to the predictions from the aforementioned Lemmas and their associated convex (ℓ_1) estimators. Moreover, we conjecture that Algorithm 1 is also more robust to measurement noise

(i.e. noise in $\boldsymbol{y}$) by design, as compared to these approaches. Most importantly, standard MMV/GMMV approaches *will not preserve the aforementioned solution structure (properties C1 and C2)* if there is noise in $\boldsymbol{y}$. Hence these techniques *cannot* be adopted to our computational problem in the presence of noise. They are only useful in specifying the conditions on the $\{\boldsymbol{F_{L_k}}\}_{k=1}^{P}$ matrices, for which the recovered solution from noiseless $\boldsymbol{y}$ will be unique.

Lastly, we consider the case when the values in $\boldsymbol{\delta}$ are functions of the base frequencies in addition to the values in $\boldsymbol{\beta}$ (i.e. $\forall i \in \{1, 2, ..., M\}, \exists! k \in \{1, 2, ..., P\}$ s. t. $\delta_i = h(\beta_k, u_i)$ where $\exists!$ is the unique existential quantifier), which is more challenging. This is because our theoretical treatment using MMV/GMMV-based algorithms now requires the estimation of M (as opposed to $P \ll M$) signals, albeit all with common support and with the aforementioned structure. Moreover, we no more have groups of measurements with the same parameter, and each $\boldsymbol{F_{L_k}}$ matrix will now contain only a single row. All these factors make the theoretical treatment in this case much more difficult. Empirically however, we have observed success of Algorithm 1 even in such a scenario. For example, see Fig. 11, where $\delta_i = h(\beta_k, u_i) \triangleq (\rho_i(\cos(\alpha_k + \beta_k) - \cos\alpha_k), \rho_i(\sin(\alpha_k + \beta_k) - \sin\alpha_k))$, as described in an example in Sec. 3. The primary reason for the success of the algorithm is that it readily exploits additional structure (due to the specific known function $h(\beta_k, u_i)$), which MMV/GMMV estimators ignore.

7. Conclusions and Discussion

We have presented a method to correct for perturbations in a compressive Fourier sensing matrix *in situ* during signal reconstruction. Our method is simple to implement, robust to noise and well grounded in theory. We have discussed several applications of our framework. Moreover, we have analyzed conditional convergence of our algorithm to a local optimum, and shown that the basic computational problem has a unique solution under reasonable conditions. We conjecture that due to the special structure of our problem, the requirements on the number of measurements is much below what is predicted by the theoretical development so far. Note that for the main algorithm and its analysis, we have consciously avoided using a Taylor approximation (Baseline 2) unlike the work in [11, 13], even though the Taylor approximation may cursorily appear to simplify the problem considerably. The primary reason for this is to avoid introduction of modeling error due to the Lagrange remainder term which can be quite significant except at small values of r. Our experimental results justify this choice.

Possible avenues for future work include attempting to prove analytical bounds for the global optimum of Algorithm 1, which we believe will be stronger than those provided by results from standard CS [1], MMV [42] or GMMV [43, 46]. The uniqueness results could also be extended (1) for the case when the signals are sparse in an orthonormal basis $\boldsymbol{\Psi}$ apart from the identity basis, or (2) for non-uniform recovery, i.e. recovery of a specific signal $\boldsymbol{x}$ for every random draw of sensing matrix $\boldsymbol{F}$ and perturbation vector $\boldsymbol{\delta}$. We also aim to

explore our algorithm in the context of different sampling strategies in practical MRI acquisition or various modes of tomographic acquisition. Furthermore, it would be interesting to explore a Bayesian algorithm for signal reconstruction which assumes a meaningful and application-specific distribution on $\boldsymbol{x}$ and $\boldsymbol{\delta}$. Lastly, the problem of mismatch of *both*, the Fourier sensing matrix and the signal representation matrix, is a useful avenue for research.

References

[1] E. Candès, M. Wakin, An introduction to compressive sampling, IEEE signal processing magazine 25 (2) (2008) 21–30.

[2] O. Katz, J. M. Levitt, Y. Silberberg, Compressive fourier transform spectroscopy, Optical Society of America, 2010.

[3] P. Schniter, S. Rangan, Compressive phase retrieval via generalized approximate message passing, IEEE Trans. Signal Processing 63 (4) (2015) 1043–1055. doi:10.1109/TSP.2014.2386294.
URL https://doi.org/10.1109/TSP.2014.2386294

[4] M. Lustig, Compressed sensing MRI, IEEE Signal Processing Magazine.

[5] M. Ferrucci, R. Leach, C. Giusca, S. Carmignato, W. Dewulf, Towards geometrical calibration of X-ray computed tomography systems: a review, Measurement Science and Technology 26 (9) (2015) 092003.

[6] H. Jang, A. B. McMillan, A rapid and robust gradient measurement technique using dynamic single-point imaging, Magnetic Resonance in Medicine.

[7] R. K. Robison, A. Devaraj, J. G. Pipe, Fast, simple gradient delay estimation for spiral MRI, Magnetic resonance in medicine 63 (6) (2010) 1683–1690.

[8] E. Brodsky, A. Samsonov, W. Block, Characterizing and correcting gradient errors in non-cartesian imaging: Are gradient errors linear-time-invariant?, Magnetic Resonance Imaging 62 (2009) 1466–1476.

[9] E. Malhotra, A. Rajwade, Tomographic reconstruction from projections with unknown view angles exploiting moment-based relationships, in: ICIP, 2016, pp. 1759–1763.

[10] M. A. Herman, T. Strohmer, General deviants: An analysis of perturbations in compressed sensing, J. Sel. Topics Signal Processing 4 (2) (2010) 342–349.

[11] Z. Yang, C. Zhang, L. Xie, Robustly stable signal recovery in compressed sensing with structured matrix perturbation, IEEE Transactions on Signal Processing 60 (9) (2012) 4658–4671.

[12] Q. Liu, H.-C. So, Y. Gu, Off-grid DOA estimation with nonconvex regularization via joint sparse representation, Signal Processing 140 (2017) 171 – 176.

[13] T. Zhao, Y. Peng, A. Nehorai, Joint sparse recovery method for compressed sensing with structured dictionary mismatches, IEEE Transactions on Signal Processing 62 (19) (2014) 4997–5008.

[14] H. Zhu, G. Leus, G. Giannakis, Sparsity-cognizant total least-squares for perturbed compressive sampling, IEEE Transactions on Signal Processing 59 (11).

[15] R. Arablouei, Fast reconstruction algorithm for perturbed compressive sensing based on total least-squares and proximal splitting, Signal Processing 130 (2017) 57 – 63.

[16] T. Ince, A. Nacaroglu, On the perturbation of measurement matrix in non-convex compressed sensing, Signal Processing 98 (2014) 143 – 149.

[17] W. Xu, Z. Li, Y. Tian, Y. Wang, J. Lin, Perturbation analysis of simultaneous orthogonal matching pursuit, Signal Processing 116 (2015) 91 – 100.

[18] S. Ling, T. Strohmer, Self-calibration via linear least squares, CoRR abs/1611.04196.
URL http://arxiv.org/abs/1611.04196

[19] S. Ling, T. Strohmer, Self-calibration and biconvex compressive sensing, Inverse Problems 31.

[20] C. Bilen, G. Puy, R. Gribonval, L. Daudet, Convex optimization approaches for blind sensor calibration using sparsity, IEEE Transactions on Signal Processing 62 (8) (2014) 4847–4856.

[21] V. Cambareri, L. Jacques, A non-convex blind calibration method for randomised sensing strategies, in: 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa), 2016, p. 16–20.

[22] Y. Chi, L. L. Scharf, A. Pezeshki, A. R. Calderbank, Sensitivity to basis mismatch in compressed sensing, IEEE Trans. Signal Processing 59 (5) (2011) 2182–2195.

[23] J. M. Nichols, A. K. Oh, R. M. Willett, Reducing basis mismatch in harmonic signal recovery via alternating convex search, IEEE Signal Process. Lett. 21 (8) (2014) 1007–1011.

[24] O. Teke, A. C. Gurbuz, O. Arikan, Perturbed orthogonal matching pursuit, IEEE Transactions on Signal Processing 61 (24) (2013) 6220–6231.

[25] G. Tang, B. N. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, IEEE Trans. Information Theory 59 (11) (2013) 7465–7490.

[26] A. Aldroubi, X. Chen, A. M. Powell, Perturbations of measurement matrices and dictionaries in compressed sensing, Applied and Computational Harmonic Analysis 33 (2) (2012) 282–291.

[27] J. D. Ianni, W. A. Grissom, Trajectory auto-corrected image reconstruction, Magnetic resonance in medicine 76 (3) (2016) 757–768.

[28] S. Basu, Y. Bresler, Uniqueness of tomography with unknown view angles, IEEE Transactions on Image Processing 9 (6) (2000) 1094–1106.

[29] Y. Fang, M. Sun, S. Vishwanathan, K. Ramani, sLLE: Spherical locally linear embedding with applications to tomography, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 1129–1136.

[30] A. Fannjiang, H.-C. Tseng, Compressive radar with off-grid targets: a perturbation approach, Inverse Problems 29 (5).

[31] H. Pandotra, E. Malhotra, A. Rajwade, K. S. Gurumoorthy, Signal recovery in perturbed fourier compressed sensing, in: 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2018.

[32] A. Belloni, V. Chernozhukov, L. Wang, Square-root LASSO: pivotal recovery of sparse signals via conic programming, Biometrika 98 (4) (2011) 791.

[33] V. Lučić, A. Rigort, W. Baumeister, Cryo-electron tomography: The challenge of doing structural biology in situ, The Journal of Cell Biology 202 (3) (2013) 407–419. doi:10.1083/jcb.201304193.

[34] A. Moussavi, M. Untenberger, M. Uecker, J. Frahm, Correction of gradient-induced phase errors in radial MRI, Magnetic Resonance in Medicine 71 (1).

[35] A. Deshmane, M. Blaimer, F. Breuer, P. Jakob, J. Duerk, N. Seiberlich, M. Griswold, Self-calibrated trajectory estimation and signal correction method for robust radial imaging using GRAPPA operator gridding, Magnetic Resonance in Medicine 75 (2) (2016) 883–896.

[36] Z. Yang, L. Xie, , C. Zhang, Off-grid direction of arrival estimation using sparse bayesian inference, IEEE Transactions on Signal Processing 61 (1) (2013) 38–43.

[37] G. Puy, P. Vandergheynst, Robust image reconstruction from multiview measurements, SIAM Journal on Imaging Sciences 7(1) (2014) 128 – 156.

[38] Z. Yang, C. Zhang, L. Xie, Sparse MRI for motion correction, in: International Symposium on Biomedical Imaging, 2013, pp. 962–965.

[39] A. Loktyushin, H. Nickisch, R. Pohmann, B. Scholkopf, Blind retrospective motion correction of MR images, Magnetic Resonance in Medicine 70 (6) (2013) 1608–1618.

[40] Wikipedia article on k-space in magnetic resonance imaging, https://tinyurl.com/yd4u7eap, online; accessed Jan 2018.

[41] I. Haviv, O. Regev, The restricted isometry property of subsampled fourier matrices, in: ACM-SIAM Symposium on Discrete Algorithms, 2016, pp. 288–297.

[42] M. F. Duarte, Y. C. Eldar, Structured compressed sensing: From theory to applications, Trans. Sig. Proc. 59 (9) (2011) 4053–4085.

[43] R. Heckel, H. Bölcskei, Joint sparsity with different measurement matrices, in: Allerton Conference on Communication, Control, and Computing, 2012, pp. 698–702.

[44] N. Rajamohan, A. Joshi, A. P. Kannu, Joint block sparse signal recovery problem and applications in LTE cell search, IEEE Transactions on Vehicular Technology 66 (2) (2017) 1130–1143.

[45] R. Heckel, Sparse signal processing: subspace clustering and system identification, Ph.D. thesis, ETH Zurich (2014).

[46] Y. C. Eldar, H. Rauhut, Average case analysis of multichannel sparse recovery using convex relaxation, IEEE Transactions on Information Theory 56 (1) (2010) 505–519.