Auto-Calibration and Biconvex Compressive Sensing with Applications to Parallel MRI

Yuan Ni and Thomas Strohmer

Abstract

We study an auto-calibration problem in which a transform-sparse signal is compressive-sensed by multiple sensors in parallel with unknown sensing parameters. The problem has an important application in pMRI reconstruction, where explicit coil calibrations are often difficult and costly to achieve in practice, but nevertheless a fundamental requirement for high-precision reconstructions. Most auto-calibrated strategies result in reconstruction that corresponds to solving a challenging biconvex optimization problem. We transform the auto-calibrated parallel sensing as a convex optimization problem using the idea of 'lifting'. By exploiting sparsity structures in the signal and the redundancy introduced by multiple sensors, we solve a mixed-norm minimization problem to recover the underlying signal and the sensing parameters simultaneously. Robust and stable recovery guarantees are derived in the presence of noise and sparsity deficiencies in the signals. For the pMRI application, our method provides a theoretically guaranteed approach to self-calibrated parallel imaging to accelerate MRI acquisitions under appropriate assumptions. Developments in MRI are discussed, and numerical simulations using the analytical phantom and simulated coil sensitives are presented to support our theoretical results.

Keywords: Self-calibration, Compressive sensing, Convex optimization, Random matrices, MRI.

1 Introduction

We frequently encounter challenges with imperfect sensing, where accurately calibrated sensors are critical for high-precision measurements or reconstructions. For many applications, explicit calibration is often difficult and expensive to carry out in practice, posing a major roadblock in many scientific and technological endeavors where accurate and reliable measurements are essential. The idea of auto-calibration (calibration-less/self-calibration) avoids the difficulties and inaccuracy associated with explicit estimations by equipping the sensors and systems with the capability to automatically derive the sensing information from its collected signals while performing the intended system function. Auto-calibration is a field of research that finds successful applications across diverse fields such as wireless communication, computer vision, remote sensing, biomedical imaging and more. This field encompasses a wide range of techniques, including blind deconvolution, blind phase retrieval, blind rotation estimation, and direction-of-arrival estimation, among others.

The inverse problem for auto-calibration is generally difficult to solve. In practice, most self-calibration algorithms rely on joint estimation of the signals of interest and the calibration parameters using standard techniques such as maximum likelihood estimation, alternating minimization or convex/nonconvex optimization. In this work, we extend the line of work that combines convex optimization and compression sensing in solving various auto-calibration tasks [36][11][4][20]. A common scheme in such approaches includes first transforming the biconvex constraints to convex constraints using the idea of 'lifting' [11]; then, drawing on ideas from compressive sensing to exploit redundancy/sparsity priors in the signal model and/or the calibration model while being consistent with the convex constraints.

In this work, we consider a type of auto-calibration problem where the same underlying signal is being sensed by multiple sensors in parallel with unknown calibration parameters. Specifically, we are concerned with the following problem:

$$\mathbf{y}_i = \mathbf{F}_\Omega(\mathbf{s}_i \odot \mathbf{x}) + \boldsymbol{\omega}_i, i = 1 \cdots C, \tag{1}$$

where $\mathbf{y}_i \in \mathbb{C}^{L\times 1}$ are the measurements, $\mathbf{F}_\Omega \in \mathbb{C}^{L\times N}$ with $L \leq N$ is the partial Fourier matrix with rows belongs to a subset $\Omega \in [N]$, $\mathbf{s}_i \in \mathbb{C}^{N\times 1}$ is the vector of unknown calibration parameters, $\mathbf{x} \in \mathbb{C}^{N\times 1}$ is the our signal of interest, $\boldsymbol{\omega}_i \in \mathbb{C}^{L\times 1}$ is the additive noise. We have C sets of measurements. Here $\odot$ is the point-wise multiplication. Our goal is to simultaneously reconstruct the signal $\mathbf{x}$ and sensor profiles (calibration parameters) $\mathbf{h}_i$. Such a forward model has important applications e.g. in parallel MRI reconstruction where one collects partial Fourier (k-space) measurements of an underlying image using multiple receiver coils where each one has nonuniform responses.

Before proceeding with further analyzing the model, we briefly discuss the identifiability and dimension of the problem. First, if $\mathbf{s}_i$ and $\mathbf{x}$ form a pair of solutions to (1), then for any complex vector $\mathbf{v}$ with nonzero entries, $\mathbf{s}_i \odot \mathbf{v}$ and $\mathbf{x} \odot 1/\mathbf{v}$ is a pair of solutions. Furthermore, with $L \leq N$, (1) has CL number of measurements and $(C + 1)N$ number of unknowns. Thus, it is impossible to recover all $\mathbf{s}_i$ and $\mathbf{x}$ without making further assumptions.

Of particular interest is the case when $\mathbf{x}$ is *transform-sparse*, a characteristic often observed in real-world signals[1]. That is, there is some known sparsifying transformation $\boldsymbol{\Psi} \in \mathbb{C}^{N\times N}$ such that $\boldsymbol{\Psi}\mathbf{x}$ is n-sparse with $n < N$. Common choices for $\boldsymbol{\Psi}$ include the well-known DCT basis and the wavelet basis. Despite the sparsity assumption, the available measurements are still fewer than the unknowns (i.e., $CL \leq CN + n$). We impose an additional subspace assumption, $\mathbf{s}_i = \mathbf{B}\mathbf{h}_i$, which can be interpreted as that the calibration parameters all lie in some known subspace spanned by a tall matrix $\mathbf{B}$. Practical choices for such $\mathbf{B}$ can be low-order polynomial basis or low-frequency sinusoidal basis to promote smoothness in the coil profiles. Altogether, the forward model becomes:

$$\mathbf{y_i} = \mathbf{F}_\Omega(\mathbf{s_i} \odot \mathbf{x}) + \boldsymbol{\omega}_i, \mathbf{s_i} = \mathbf{B}\mathbf{h_i}, \boldsymbol{\Psi}\mathbf{x} = \mathbf{z}, \mathbf{z} \text{ is sparse }, i = 1 \cdots C. \tag{2}$$

Here, $\mathbf{B} \in \mathbb{C}^{N\times k}, k < N$, $\boldsymbol{\Psi} \in \mathbb{C}^{N\times N}$ are known.

The forward model holds significant relevance in blind pMRI reconstructions. Additionally, although the above model may appear straightforward and specific, providing a rigorous analysis on it proves to be non-trivial. Our goal is to understand the requirements on the number of samples L and the number of parallel measurements C for a robust and stable recovery of all $\mathbf{h}_i$ and $\mathbf{x}$. The primary challenge for analysis comes from the inherent structure of the forward matrix, which requires us to make reasonable and careful assumptions regarding randomness that align with the specific application in mind.

1.1 Notation and outline

We denote vectors and matrices by bold font letters (e.g. $\mathbf{x}$ or $\mathbf{X}$) and scalars by regular font or Greek symbols. For any vector $\mathbf{x} \in \mathbb{C}^N$ and set $S \subset [N]$, define $\mathbf{x}_S = \mathcal{P}_S\mathbf{x} \in \mathbb{C}^N$ the orthogonal projection of $\mathbf{x}$ onto set S. (e.g. $\mathbf{x}_S(i) = \mathbf{x}(i), i \in S$ and $\mathbf{x}_S(i) = 0$, for $i \in S^c$). Similarly, define the orthogonal projection of matrix $\mathbf{X} \in \mathbb{C}^{N_1\times N_2}$ to set $T \subset [N_1] \times [N_2]$ by $\mathbf{X}_T = \mathcal{P}_T\mathbf{X} \in \mathbb{C}^{N_1\times N_2}$, such that $\mathbf{X}_T(i, j) = \mathbf{X}(i, j), (i, j) \in T$ and $\mathbf{X}_T(i, j) = 0$, otherwise. We denote the N by N identity matrix as $\mathbf{I}_N$. For any complex vector $\mathbf{x}$ or matrix $\mathbf{X}$, $\bar{\mathbf{x}}$ or $\bar{\mathbf{X}}$ means their conjugate respectively. $vec(\mathbf{X})$ means the vectorization of a matrix $\mathbf{X}$ in the column-wise order into a vector. We denote the circular convolution as $*$ and the kronecker product as $\otimes$. For any set S, we denote S^c as its complementary set and $|S| = \text{Card}(S)$. When we refer to a matrix as generic, we imply that the entries within that matrix/vector are independently sampled from a standard Gaussian distribution. Informally, we refer to a matrix $X \in \mathbb{C}^{N_1\times N_2}$ being incoherent if the 2-norm of its

[1]This notion can be easily generalized to *approximately* transform-sparse.

rows does not vary too much.

The paper is organized as following: In section 2 we briefly discuss our contributions in relation to the state of the art in self-calibration and compressive sensing MRI. In section 3 we introduce the problem setup. We present our main results in section 4 and provide the details of the proofs in section 6. Numerical experiments are shown in section 5.

2 Related work

Our work is grounded in a real-world application concerning auto-calibrated pMRI. We draw general ideas from compressive sensing, convex optimization, self-calibation and applications in MRI reconstruction. The work is inspired by papers related to bringing self-calibration problems into the framework of biconvex compressive sensing, such as PhaseLift, SparseLift [36][20], and by blind deconvolution using lifting by [4].

We first introduce the concept of 'lifting', which serves as an important tool to in our proposed method. Consider any bi-linear measurements of two unknown signals $\mathbf{x}$ and $\mathbf{z}$ of the form

$$y_i =< \mathbf{x}, \mathbf{b_i} >< \mathbf{z}, \bar{\mathbf{a}}_\mathbf{i} >, i = 1 \cdots m, \tag{3}$$

where $\mathbf{x}, \mathbf{b}_i \in \mathbb{C}^{N_1 \times 1}, \mathbf{z}, \mathbf{a}_i \in \mathbb{C}^{N_2 \times 1}$. Define matrix $\mathbf{X} := \mathbf{x}\mathbf{z}^T$, we can write the bilinear equations as a linear equation with respect to $\mathbf{X}$ using some linear algebra,

$$< \mathbf{x}, \mathbf{b_i} >< \mathbf{z}, \bar{\mathbf{a}}_\mathbf{i} >= \mathbf{b}_i^* \mathbf{x}\mathbf{z}^T \mathbf{a}_i = \mathbf{b}_i^* \mathbf{X} \mathbf{a}_i = (\mathbf{a}_i^T \otimes \mathbf{b}_i^*) \operatorname{vec}(\mathbf{X}). \tag{4}$$

Hence, by defining linear operator $\mathcal{A} : \mathbb{C}^{N_1 \times N_2} \to \mathbb{C}^m$, the systems of equations becomes

$$[y_1, \cdots, y_m] = \mathcal{A}(\mathbf{X}) := \{\mathbf{b}_i^* \mathbf{X} \mathbf{a}_i\}_{i=1}^m. \tag{5}$$

The original bilinear equations of dimension $(N_1 + N_2)$ is 'lifted' to linear equations of underlying dimension $N_1 \times N_2$.

2.1 Related work in auto-calibration and compressive sensing

There exist two primary approaches to address the bilinear optimization problems: one involves convexifying it through methods such as linearization or lifting [4][36][35][20][11]. The other involves optimization in the natural parameters space using alternating minimization and gradient descent algorithms[30][33][46]. Various auto-calibration problems of similar setups have been explored, mainly differing from each other in the following two assumptions:

- The assumptions made about the signal and calibration parameters: whether they are sparse individually or jointly, sparse in the natural parameter space or after transformations, subspace constrains, or restricted by sign constraints.
- The selection of the forward model: whether it employs generic sensing matrices (e.g., Gaussian) or structured sensing matrices implied by the specific application cases.

We briefly survey several works that have influenced our work, employing a combination of 'lifting', sparsity priors, and subspace assumptions. The SparseLift framework [36] is concerned with another type of bilinear self-calibration problem pertaining sparse signals. The authors consider recovery from the following bi-convex measurements:

$$\mathbf{y} = \operatorname{diag}(\mathbf{B}\mathbf{h})\mathbf{A}\mathbf{x} + \boldsymbol{\omega}, \tag{6}$$

where $\mathbf{h} \in \mathbb{C}^k$, $\mathbf{y} \in \mathbb{C}^L$, $\mathbf{A} \in \mathbb{C}^{L \times N}, L \leq N$ and $\|\mathbf{x}\|_0 = n$. In this model, the unknown sensing parameters come after compression (e.g. matrix $\mathbf{A}$ is Gaussian or partial Fourier). Using the idea of lifting, Sparselift

[36] first transforms the bi-convex constraint (6) to a linear constraint with respect to $\mathbf{X} = \mathbf{h}\mathbf{x}^T$. Exploiting the sparsity in $\mathbf{x}$ and consequently within $\mathbf{X}$, the recovery solves a convex relaxation by minimizing the $\|\mathbf{X}\|_1$ subject to the linear constraints. The authors show the recovery of $\mathbf{X}_0 = \mathbf{h}_0\mathbf{x}_0^T$ is guaranteed with high probability when the sensing matrix $\mathbf{A}$ is either a Gaussian random matrix or with its rows chosen uniformly at random with replacement from the DFT matrix, when the number of measurements L is of the order of $\mathcal{O}(kn\log^2(L))$. In a related work, [20] extends the results of SparseLift to the case where one observes $\mathbf{y} = \sum_{i\in[r]} \text{diag}(\mathbf{B}\mathbf{h}_i)\mathbf{A}\mathbf{x}_i + \boldsymbol{\omega}$. This extension improves recovery guarantees by solving a $l_{1,2}$ minimization by promoting block sparsity in the lifted signals. One important application of the auto-calibration problem in imaging includes reconstruction using randomly coded masks [6][51].

The idea of 'lifting' also applies to a particular class of self-calibration problems, notably the blind deconvolution problem [4][32], where one measures the convolutions of the signal with an unknown filter (i.e., $\mathbf{y} = \mathbf{h} * \mathbf{x}$). Notice that the forward model (2) can be written as a convolution $\mathbf{y} = I_\Omega(\mathbf{F}\mathbf{B}\mathbf{h} * \mathbf{F}\boldsymbol{\Psi}^*\mathbf{z})$. The work [4] analyzes the case when the Fourier space is fully sampled, it also uses lifting and solves a nuclear norm minimization problem by exploiting low-rankness in the lifted signal $\mathbf{X}$. If the active coefficients of $\mathbf{z}$ are known and $\|\mathbf{z}\|_0 = n$, the main theorem of [4] guarantees a stable recovery when the number of measurements is of the order $\mathcal{O}(n + k)$ when $\boldsymbol{\Psi}$ is generic and $\mathbf{B}$ is incoherent. [31] analyzes the case when the convolution is only subsampled, the authors provide an iterative algorithm for recovering bi-sparse signals $\mathbf{s}$ and $\mathbf{x}$. The subspace assumption that we adopt in this work $\mathbf{s} = \mathbf{B}\mathbf{h}$ is a special case when the support of a sparse calibration parameters is known. The authors provide robust recovery guarantees near optimal sample complexity when the sparsifying matrices are generic.

On the other hand, parallel imaging (i.e., $C > 1$ in (2)) consistently yields superior reconstruction results when compared to the single-coil ($C = 1$) scenario [34][45][28]. We aim to establish a framework for comprehensively analyzing the role of multiple coils and the underlying reasons for its improved recovery.

From a practical viewpoint a relevant scenario focuses on self-calibration from multiple snapshots. For example, blind image restoration from multiple filters [26] and self-calibration model for sensors [7][35][22][9]. The goal here is to recover the unknown gains/phases $\mathbf{D} = \text{diag}(\mathbf{d})$ and a signal matrix $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_p]$ from the measurement matrix $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_p]$ and $\mathbf{Y} = \mathbf{D}\mathbf{A}\mathbf{X}$. In general, the method does not require sparse priors on the signal or the calibrations. Among this line of research, an interesting and somewhat related method involves using the idea of 'linearization' to transform the biconvex constraint into a linear one. One crucial assumption is that $\mathbf{s}$ does not have zero entries and define $\mathbf{d} = \mathbf{1}/\mathbf{s}$, observations of the form $\mathbf{y}_l = \text{diag}(\mathbf{s})\mathbf{A}\mathbf{x}_\mathbf{l}$ can be written as

$$\mathbf{y}_l\mathbf{d} = \mathbf{A}\mathbf{x}_\mathbf{l}, \tag{7}$$

which is a linear equation in $[\mathbf{s}, \mathbf{x}_l]$. Compared to 'Lifting', the underlying problem dimension is unchanged. The same idea of linearization can also be found in [22][9][7][54]. In [35], the authors solve the problem using least-square minimization and give a theoretical recovery guarantee when $\mathcal{A}$ is generic.

For a general discussion on the issue of injectivity and the principle of identifiability for bi-linear compressive sensing, see e.g. [29][12]. It is also worth noting that [22][11] have developed self- calibration algorithms by using sparsity, convex optimization, and/or ideas from compressive sensing.

2.2 Related work in MRI

In this section, we show the connection of our problem to magnetic resonance imaging (MRI) and discuss auto-calibration methods within the specific application. MRI is a widely used imaging modality that enables high-contrast imaging. In a parallel MRI setting, multiple receiver coils are used to collect data from the same underlying image/3D volume. Data collected by the sensors are samples of the spatial Fourier transform of the underlying image multiplied by the complex receiver coil profile. We refer to the book [39] for an

introduction to the physics of MRI. The need for precise calibrations of the coil profiles arises due to their non-uniformity, including:

- Nonuniform sensitivity: coils are most sensitive near the body surface, diminishing with distance from the imaged object.
- Phase modulation: coils induce point-wise phase shifts in the sensed signal.

The coil profile can be regarded as a complex vector $\mathbf{s} \in \mathbb{C}^N$ of the same length as the object to be imaged $\mathbf{x} \in \mathbb{R}^N$ and the forward model of pMRI can be written in the form of equation (1). The Fourier measurement (k space) is subsampled to achieve faster imaging.

MR images often give rise to sparse representations, either in the spatial domain or, more commonly, in some transform domain [37]. Since Lustig et al. [37] first demonstrated the compressive sensing approach for MRI reconstruction, it has become one of the most important tools used for modern MR imaging research. The redundancy in MRI can come in four ways: a) transform sparsity: natural images are sparse in transform domains, b) k-space (Fourier space) redundancy: exploit structured low-rankness of some formed measurements matrix which rely on the Fourier convolutional relationship, c) coil redundancy: sensitivity encoding by using multiple coils and their pairwise relationships, d) temporal redundancy: dynamic MRI. Coil domain redundancy uses the pairwise relationship between coils. The two most famous methods include SENSE [45] and GRAPPA [23] and their references within. Compared to the spatial/coil domain sparsity, methods exploiting structured k-space redundancy are relatively new. These approaches are derived from the k-space convolution relationship. We refer to SAKE [50], LORAKS[25], ALOHA [28] for this line of research. In this paper, we exploit both transform sparsity and coil redundancy. A similar setup, though in the perfectly calibrated case, is used in sparse SENSE [13] where the authors exploit joint sparsity across channels by solving a $l_{1,2}$ minimization problem.

From compressive sensing theory we know that by exploiting sparsity, under certain conditions images with a sparse representation can be recovered from randomly undersampled k-space (Fourier space) [10][49]. In our model, we assume our signal is sparse in some transform domain $\mathbf{\Psi}$. The choice of such $\mathbf{\Psi}$ is related to the field of research in sparsifying transformations [16][44]. More recently, attention has turned to adapting these models to data [3][48]. In this paper, we only consider a known sparsifying transformation $\mathbf{\Psi}$ satisfying $\mathbf{\Psi}^*\mathbf{\Psi} = \mathbf{I}_N$. In the MRI literature, it has been demonstrated that the discrete cosine transformation, the wavelet transform and the finite difference transform are the mainly utilized choices for sparsifying MR images [50][47][13].

In practice, the coil profiles are often assumed to lie in a low dimensional subspace. Hence, $\mathbf{B}$ in (2) is commonly chosen as a polynomial basis or a sinusoidal basis as seen in practice [45] [24]. As for common practices for calibrating the coil profiles, typically, these profiles are estimated through a separate pre-scan. However, this approach is susceptible to motion errors, environmental fluctuations, and requires longer acquisition times, among other challenges. For auto-calibration methods, a line of research relies on a fully sampled k-space region [40][41] to extract the sensitivity parameters. Data driven auto-calibration methods, such as GRAPPA [23] and SPIRiT [38] can be viewed as interpolation methods by estimating linear relationships within the fully sampled k-space (e.g., kernel calibration) and enforce that relationship to synthesize data values in place of unacquired data (e.g., data reconstruction). More recently, a data-driven method SAKE [50] structures the multi-coil dataset into a new data matrix that is designed to have low-rankness property. Then it solves a matrix completion problem to fill in the missing data. Joint estimation techniques attempt to iteratively estimate both the coil sensitivities and image contents while imposing some smoothness constraints on the sensitivity profiles, some examples include [56][53]. Of particular interest is ENLIVE [27] where the author shows a connection between the original non-convex formulation to a nuclear norm minimization problem by using the idea of 'Lifting'.

ENLIVE solves the following relaxed joint optimization problem over k sets of coil profiles $\mathbf{s}_j^i$ and images $\mathbf{x}^i$ from measurements $\mathbf{y}_i$ over C coils,

$$\arg\min_{\mathbf{x^i},\mathbf{s_j^i}} \Sigma_{j=1}^{C}\|\mathbf{y}_i - \mathcal{F}_\Omega\{\Sigma_{i=1}^{k}\mathbf{s}_j^i \odot \mathbf{x}^i\}\|_2^2 + \alpha\Sigma_{i=1}^{k}(\Sigma_{j=1}^{C}\|\mathbf{W}\mathbf{s}_j^i\|_2^2 + \|\mathbf{x}^i\|_2^2). \tag{8}$$

The operator $\mathbf{W}$ enforces smoothness in the coil profiles. Here, k sets of images $\mathbf{x}^i$ are used instead of one single image $\mathbf{x}$ to account for model violations. In a post-processing step, an average image is calculated as the solution.

The author shows an equivalence relationship between the formulation to a linearly constraint nuclear norm minimization problem by using the idea of 'Lifting'. By defining the matrix $\mathbf{y} = \begin{bmatrix}\mathbf{y}_1 & \cdots & \mathbf{y}_C\end{bmatrix}$, $\mathbf{U} = \begin{bmatrix}\mathbf{x}^1 & \cdots \mathbf{x}^k\end{bmatrix}$, $\mathbf{V} = \begin{bmatrix}\mathbf{W}\mathbf{s}_1^1 & \cdots \mathbf{W}\mathbf{s}_j^k\end{bmatrix}$, the problem is lifted into

$$\arg\min_{\mathbf{U},\mathbf{V}} \|\mathbf{y} - \mathcal{A}(\mathbf{U}\mathbf{V}^T)\|_2^2 + \alpha(\|\mathbf{V}\|_F^2 + \|\mathbf{U}\|_F^2), \tag{9}$$

which is further related to the nuclear norm minimization problem with respect to the tensor product $\mathbf{Z} = \mathbf{U}\mathbf{V}^T$,

$$\arg\min_{\mathbf{Z}} \|\mathbf{y} - \mathcal{A}(\mathbf{Z})\|_2^2 + 2\alpha\|\mathbf{Z}\|_*. \tag{10}$$

ENLIVE is related to our approach as it uses lifting and exploits the low-rankness property of the lifted signal, whereas our method exploits the block-sparsity of the lifted signal. However, since nuclear norm minimization is computationally expensive, (10) is not utilized for practical computations but rather to provide an analytical explanation; In practice, the non-convex formulation (8) is used to solve the auto-calibrated reconstruction problem instead.

Although these algorithms undoubtedly offer utility and applicability by giving successful numerical results, the majority of them do not provide sufficient theoretical guarantees of recovery.

2.3 Our contributions

We propose a new method that combines convex optimization and compressive sensing to solve the auto-calibration problem in the form of (2) by exploiting the block sparsity structure within the lifted signals among all parallel sensors and solving a convenient mixed-norm optimization problem. The main contribution is to provide a robust and stable theoretical guarantee of recovery. Additionally, existing theoretical development in calibrationless compressive sensing methods often lacks assumptions aligned with those suitable for the pMRI application. We make an effort to incorporate appropriate assumptions throughout our analysis.

3 Problem setup

We are concerned with the following auto-calibration problem:

$$\mathbf{y}_i = \mathbf{F}_\Omega\mathbf{D}_i\mathbf{x} + \mathbf{n}_i, i = 1\cdots C, \tag{11}$$

where $\mathbf{y}_i \in \mathbb{C}^{L\times 1}$ are the measurements, $\mathbf{F}_\Omega \in \mathbb{C}^{L\times N}$ is the partial Tourier matrix of the N-point DFT matrix with L rows chosen from the subset $\Omega \in [N]$, $\mathbf{D}_i \in \mathbb{C}^{N\times N}$ is the diagonal matrices of calibration parameters, $\mathbf{x} \in \mathbb{C}^{N\times 1}$ the underlying signal, $\mathbf{n}_i \in \mathbb{C}^{N\times 1}$ being additive noise.

Our goal is to recover the common signal of interest $\mathbf{x}$ from the measurements $\mathbf{y}_i$ where the calibration matrices $\mathbf{D}_i$ are unknown. As discussed in the introduction, it is in general not possible to recover $\mathbf{x}$ (and

$\mathbf{D}_i$) from the measurements since the number of unknowns is larger than the number of measurements, regardless of how many independent experiments are taken (reflected in the number C).

Hence, we impose two assumptions on both the signal $\mathbf{x}$ and the calibration parameters $\mathbf{D}_i$. In particular,

1. $\mathbf{x}$ has a sparse representation: there is a known sparsifying transformation $\mathbf{\Psi} \in \mathbb{C}^{N\times N}$ such that $\mathbf{z} = \mathbf{\Psi x}$ is n-sparse. The location of the n active coefficient of $\mathbf{z}$ is unknown.

2. The calibration matrices belong to a common and known subspace: there is a known tall matrix $\mathbf{B} \in \mathbb{C}^{N\times k}$ with $k < N$ such that $\mathbf{D}_i = \text{diag}(\mathbf{Bh}_i)$, for all $i = 1 \cdots C$. Here $\mathbf{h_i} \in \mathbb{C}^k$ are unknown.

Under these assumptions, the recovery problem is to find $\mathbf{h}_i, i = 1 \cdots C$ and a sparse $\mathbf{z}$ that satisfies

$$\mathbf{y}_i = \mathbf{F}_\Omega \text{diag}(\mathbf{Bh}_i)\mathbf{\Psi}^*\mathbf{z}, i = 1 \cdots C. \tag{12}$$

Next, we apply the idea of 'Lifting' to transform the bi-linear constraint in both $\mathbf{h}_i$ and $\mathbf{z}$ to a linear constraint with respect to their outer product $\mathbf{h}_i\mathbf{z}^T$

3.1 Lifting

Using the lifting approach [11, 4, 36], we can transform the biconvex problem into the problem of recovering a sparse matrix from linear measurements.

In particular, denote $\mathbf{B} = \begin{bmatrix} \mathbf{b}_1^* \\ \vdots \\ \mathbf{b}_N^* \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_k \end{bmatrix} \in \mathbb{C}^{N\times k}$, $\mathbf{\Psi}^* = \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_N^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{\psi}_1 & \cdots & \boldsymbol{\psi}_N \end{bmatrix} \in \mathbb{C}^{N\times N}$.

A little linear algebra yields that the measurements $\mathbf{y}_i \in \mathbb{C}^{L\times 1}$ from each coil obey:

$$\mathbf{y}_i = \mathbf{F}_\Omega \begin{bmatrix} \mathbf{b}_1^*\mathbf{h}_i\mathbf{z}^T\mathbf{q}_1 \\ \mathbf{b}_2^*\mathbf{h}_i\mathbf{z}^T\mathbf{q}_2 \\ \vdots \\ \mathbf{b}_N^*\mathbf{h}_i\mathbf{z}^T\mathbf{q}_N \end{bmatrix} = \mathbf{F_\Omega} \underbrace{\begin{bmatrix} (\bar{\mathbf{q}}_1 \bigotimes \mathbf{b}_1)^* \\ \vdots \\ \bar{\mathbf{q}}_N \bigotimes \mathbf{b}_N)^* \end{bmatrix}}_{\mathbf{\Phi}} \text{vec}(\mathbf{h}_i\mathbf{z}^T) := \mathbf{F}_\Omega\mathbf{\Phi}\text{vec}(\mathbf{h}_i\mathbf{z}^T), i = 1 \cdots C, \tag{13}$$

where $\mathbf{\Phi} = \begin{bmatrix} (\bar{\mathbf{q}}_1 \bigotimes \mathbf{b}_1)^* \\ \vdots \\ \bar{\mathbf{q}}_N \bigotimes \mathbf{b}_N)^* \end{bmatrix} \in \mathbb{C}^{N\times kN}$.

Define $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 | \mathbf{y}_2 | \cdots | \mathbf{y}_C \end{bmatrix}$, $\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 | \mathbf{h}_2 | \cdots | \mathbf{h}_C \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 | \mathbf{X}_2 | \cdots | \mathbf{X}_C \end{bmatrix} = \begin{bmatrix} \text{vec}(\mathbf{h}_1\mathbf{z}^T) | \text{vec}(\mathbf{h}_2\mathbf{z}^T) | \cdots | \text{vec}(\mathbf{h}_C\mathbf{z}^T) \end{bmatrix} = \mathbf{z} \otimes \mathbf{H} \in \mathbf{C}^{kN\times C}$, we can rewrite (13) in matrix form as follows:

$$\mathbf{Y} = \underbrace{\mathbf{F}_\Omega\mathbf{\Phi}}_{\mathbf{A}} \mathbf{X}, \tag{14}$$

where $\mathbf{Y} \in \mathbb{C}^{L\times C}$, $\mathbf{F}_\Omega \in \mathbb{C}^{L\times N}$, $\mathbf{\Phi} \in \mathbb{C}^{N\times kN}$, $\mathbf{X} \in \mathbb{C}^{KN\times 1}$. We refer to $\mathbf{A} = \mathbf{F}_\Omega\mathbf{\Phi}$ as the forward matrix. Now the problem is to find the matrix $\mathbf{X}$ satisfying the linear constraint (14).

3.2 Block sparsity of X

Assume $\mathbf{z}$ is n-sparse (i.e., $\|\mathbf{z}\|_0 = n$) and that $\text{supp}(\mathbf{z}) = S$. For any i, $\text{vec}(\mathbf{h}_i\mathbf{z}^T)$ is block n-sparse with blocks of length k. More precisely, define the indices of the j-th block as $T_j = \{(j-1)k+1, \cdots, jk\}, j \in [N]$,

$\text{supp}(\mathbf{h}_i\mathbf{z}^T) = \cup_{j\in S}\{(j-1)k+1,\cdots,jk\}$. Concerning the sparsity of $\mathbf{X}$, since its columns share the same indexes for supports, it consists of at most n consecutive blocks of rows that are non-zero. More precisely, $\text{supp}(\mathbf{X}) = (\cup_{j\in S} T_j)\times[C]$.

Moreover, $\mathbf{X}$ is also low rank. Consider a reshaped matrix $\tilde{\mathbf{X}} = \begin{bmatrix}\mathbf{h}_1\mathbf{z}^T & \cdots & \mathbf{h}_c\mathbf{z}^T\end{bmatrix}^T \in \mathbb{C}^{kC\times N}$, which has rank 1. Methods that exploit the low-rankness are introduced in [5][52][27] and include the ENLIVE formulation (8). Naturally, one could minimize over a linear combination $\|\mathbf{X}\|_{1,2}+\lambda\|\mathbf{X}\|_*, \lambda>0$ in an attempt to simultaneously capture the low-rankness and sparse characteristics of $\mathbf{X}$. However, as demonstrated in reference [43], this approach is no more effective than using a single norm in terms of the number of measurements required. Theoretically, only recovering for a single coil case by minimizing $\|\cdot\|_*$ requires $L = O(N+k)$. Also $l_{1,2}$ minimization is more computationally efficient than the nuclear norm minimization problem. Hence, we choose to only minimize with respect to $\|\cdot\|_{1,2}$, which (as we will prove) may already yield a solution with sparse structure and small nuclear norm.

3.3 Optimization problem

Given (k,N,C), for any $\mathbf{X}\in\mathbb{C}^{kN\times C}$, define

$$\|\mathbf{X}\|_{1,2} = \sum_{j\in[N]} \|\mathbf{X}_{T_j\times[C]}\|_F, \tag{15}$$

which is the sum of the Frobenious norm of the k by C blocks of $\mathbf{X}$.

Since it is well known that the $\ell_{1,2}$-norm promotes the block sparse structure in $\mathbf{X}$ [17][18][20], we solve the following convex program

$$\min \|\mathbf{X}\|_{1,2} \quad \text{subject to} \mathbf{Y} = \mathbf{F}_\Omega\mathbf{\Phi}\mathbf{X}. \tag{16}$$

If the measurement are noisy, i.e., $\mathbf{Y} = \mathbf{F}_\Omega\mathbf{\Phi}\mathbf{X}+\mathbf{N}$ with $\|\mathbf{N}\|_F\le\sigma$, we solve the following convex program

$$\min \|\mathbf{X}\|_{1,2} \quad \text{subject to} \|\mathbf{Y}-\mathbf{F}_\Omega\mathbf{\Phi}\mathbf{X}\|_F\le\sigma. \tag{17}$$

We also denote the forward operator as $\mathcal{A}: \mathbb{C}^{k\times N}\to\mathbb{C}^L$, or via a matrix $\mathbf{A}\in\mathbb{C}^{L\times kN}$, such that $\mathbf{Y} = \mathcal{A}(X) = \mathbf{A}\mathbf{X}$. In our problem, $\mathbf{A} = \mathbf{F}_\Omega\mathbf{\Phi}$.

Once we get a solution $\hat{\mathbf{X}}$ to the $\ell_{1,2}$ minimization problem (16)(17), we calculate the average of the best rank-one estimations of its columns to retrieve the underlying signal. Recall the columns of $\hat{\mathbf{X}} = \begin{bmatrix}\hat{\mathbf{x}}_1 & \cdots\hat{\mathbf{x}}_C\end{bmatrix}$ and the truth $\mathbf{X}^0 = \begin{bmatrix}\mathbf{X}_1^0 & \cdots\mathbf{X}_C^0\end{bmatrix} = \begin{bmatrix}\text{vec}(\mathbf{h}_1\mathbf{z}^T)|\text{vec}(\mathbf{h}_2\mathbf{z}^T)|\cdots|\text{vec}(\mathbf{h}_C\mathbf{z}^T)\end{bmatrix}$. We first reshape the columns $\hat{\mathbf{x}}_i$ into a matrix $\hat{\mathbf{X}}_i\in\mathbb{C}^{k\times N}$ in the column-wise order, i.e., $\text{vec}(\hat{\mathbf{X}}_i) = \hat{\mathbf{x}}_i$. Consider the best rank-one estimation of $\hat{\mathbf{X}}_i$, $\sigma_i\hat{\mathbf{u}}_i\hat{\mathbf{v}}_i^T$, where σ_i, $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$ are its leading singular value, left and right singular vectors respectively. Let $\hat{\mathbf{h}}_i = \sqrt{\sigma_i}\hat{\mathbf{u}}_i$ and $\hat{\mathbf{z}}_i = \sqrt{\sigma_i}\hat{\mathbf{v}}_i$. We use the following lemma from [36] to ensure that $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{z}}_i$ are close to $\mathbf{h}_i$ and $\mathbf{z}$.

Lemma 1 *For any matrix $\boldsymbol{X}$ and $\boldsymbol{X}_0 = \boldsymbol{u}_0\boldsymbol{v}_0^T$, let $\hat{\sigma}\hat{\boldsymbol{u}}\hat{\boldsymbol{v}}^T$ be the best rank-one Frobenius norm approximation of $\boldsymbol{X}$. If $\|\boldsymbol{X}-\boldsymbol{X}_0\|_F = \epsilon$, then there exists a scalar α_0 and a constant C_0 such that,*

$$\|\boldsymbol{u}_0-\alpha_0\hat{\boldsymbol{u}}\|\le C_0\min(\epsilon/\|\boldsymbol{u}_0\|,\|\boldsymbol{u}_0\|), \|\boldsymbol{v}_0-\alpha_0^{-1}\hat{\boldsymbol{v}}\|\le C_0\min(\epsilon/\|\boldsymbol{v}_0\|,\|\boldsymbol{v}_0\|).$$

Notice if $\mathbf{X}_0 = \mathbf{u}_0\mathbf{v}_0^T$, then $\alpha_0\mathbf{u}_0$ and $1/\alpha_0\mathbf{v}_0$ is also a pair of solutions for any $\alpha_0\neq 0$. However, the scalar ambiguity is not important for many applications, including MRI. We set our estimation of $\mathbf{z}$ as an average $\hat{\mathbf{z}} = \sum_{i=1\cdots C}\hat{\mathbf{z}}_i/C$. The reconstructed signal is $\hat{\mathbf{x}} = \mathbf{\Psi}^*\hat{\mathbf{z}}$.

In the aspect of solving the optimization problem (16)(17), the following unconstrained form is often used,

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_2^2 + \lambda\|\mathbf{X}\|_{1,2}. \tag{18}$$

A comprehensive review of the algorithms to solve problems of the form (18) is not a focus of this paper; we refer to a few common approaches such as the forward-backward splitting (FBS) [14], ADMM [55] and FISTA [8] which have found numerous applications in the context of pMRI reconstruction.

4 Main Result

Given any integer k and N, we define $T_j \subset [kN]$ by $T_j = \{(j-1)k+1, \cdots, jk\}, j \in [N]$. Given an index set $S = \{s_1, s_2, \cdots\} \in [N]$, define $S_j = T_{s_j}$ for $j \in |S|$ and $T := \cup_{j\in|S|} S_j$. With slight abuse of notation, we will not distinguish $\mathcal{P}_S$, $\mathcal{P}_T$ and $\mathcal{P}_{T\times[C]}$ when it is clear form the context. For example, for $\mathbf{x} \in C^N$, $\mathbf{x} \in C^{kN}$ and $\mathbf{X} \in C^{kN\times C}$, $\mathcal{P}_S$ means the projection onto S, T and $T \times [C]$ respectively. We will also sometimes write the projections $\mathcal{P}_S$ as subscripts, for example, $\mathbf{x}_S$ or $\mathbf{X}_S$, when it is convenient.

For any $\mathbf{X} \in \mathbb{C}^{kN\times C}$, define $\mathrm{sgn}(\mathbf{X}) \in \mathbb{C}^{kN\times C}$ as a matrix formed by normalizing each block of $\mathbf{X}$ with length $k \times C$. To be precise,

$$\mathrm{sgn}(\mathbf{X}) = \begin{cases} \frac{\mathbf{X}_{T_j\times[C]}}{\|\mathbf{X}_{T_j\times[C]}\|_F}, \text{if} \|\mathbf{X}_{T_j\times[C]}\|_F \neq 0 \\ \mathbf{0}, \text{otherwise} \end{cases}. \tag{19}$$

It will come handy in the proofs to decompose a linear maps to any given subset of its columns. For a matrix $\mathbf{A} \in \mathbb{C}^{L\times kN}$ and a given set S, we define the projections onto its blocks of columns induced by S as

$$\mathbf{A}_{S_j}(:,k) = \begin{cases} \mathbf{A}(:,k), \text{if } k \in S_j \\ \mathbf{0}, \text{otherwise} \end{cases} \quad \text{and } \mathbf{A}_S(:,k) = \begin{cases} \mathbf{A}(:,k), \text{if } k \in T \\ \mathbf{0}, \text{otherwise} \end{cases}. \tag{20}$$

4.1 Assumptions and main theorem

Before we move on to present our main theoretical findings, for convenience we summarize all the assumptions on the model (14) and the signals here:

1. The rows of $\mathbf{B}$ are chosen uniformly at random without replacement from the rows of a fixed matrix $\mathbf{B}_0$ and $\mathbf{B}_0^*\mathbf{B}_0 = \mathbf{I}_k$.
2. $\mathbf{\Psi}$ is a known and fixed orthonormal basis with $\mathbf{\Psi}^*\mathbf{\Psi} = \mathbf{\Psi}\mathbf{\Psi}^* = \mathbf{I}$.
3. The rows of $\mathbf{F}_\Omega$ are chosen uniformly at random without replacement from the DFT matrix with $\mathbf{F}^*\mathbf{F} = \frac{N}{L}\mathbf{I}_N$.
4. We analyze two probability models for the sensitivity parameters:
   - **Sampled from some basis of** $\mathbb{C}^k$: Each column $\mathbf{h}_l, l \in [C]$ are chosen independently and uniformly at random with replacement from the columns of an orthonormal basis $\mathbf{W} \in \mathbb{C}^{k\times k}$.
   - **Complex spherical**: Each spanning coefficients $\mathbf{h}_l \in \mathbb{C}^k, l \in [C]$ are chosen independently and uniformly at random from the complex sphere $S_{\mathbb{C}}^{k-1}$

Theorem 1 *For any arbitrary subset $S \in [N]$ with card$(S) = n$ and $\boldsymbol{X}_0 := \boldsymbol{z}_0 \otimes \boldsymbol{H}, \boldsymbol{z}_0 \in \mathbb{C}^{N\times 1}, \boldsymbol{H} \in \mathbb{C}^{k\times C}$, consider the linear map $\boldsymbol{A} \in \mathbb{C}^{L\times kN}$ as defined in (14) that satisfies the randomness assumptions 4.1 and the noisy measurement $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}_0 + \boldsymbol{\mathcal{N}}$ with $\|\boldsymbol{\mathcal{N}}\|_F \leq \sigma$. The solution $\boldsymbol{X}$ to the $l_{1,2}$ minimization problem (17) satisfies*

$$\|\boldsymbol{X} - \boldsymbol{X}_0\|_F \leq C_1 \|\mathcal{P}_{S^c} \boldsymbol{X}_0\|_{1,2} + (C_2 + C_3\sqrt{s})\sigma \tag{21}$$

with probability $1 - 4N^{-\alpha}$ if

1. *the number of measurement satisfies*

$$L \geq C_\alpha kn {\mu^B_{max}}^2 {\mu^\Psi_{max}}^2 \log(knN).$$

2. *If $\boldsymbol{h}_l$ is independently chosen from a basis or lie on the complex sphere and the number of coils satisfy*

$$C \geq C_\alpha \log(N)k.$$

Here, $\mu^B_{max} := \max_{i,j} \sqrt{L}|\mathbf{B}_{i,j}|$, $\mu^\Psi_{max} := \max_{i,j} \sqrt{N}|\mathbf{\Psi}_{i,j}|$, which measures the incoherence of the rows of $\mathbf{B}$ and $\mathbf{\Psi}$ respectively. The constant C_α grows linearly with respect to α. A detailed bound for the constants can be found in the corresponding proofs, see Section 5.

Remark: The recovery is stable and robust. Note that in the bound, $kn = \|P_S(\mathbf{X}_0)\|_0$ and when S is the true support of $\mathbf{z}_0$, $\mathcal{P}_{S^c}\mathbf{X}_0 = 0$. For the bound regarding L, we expect that the magnitude of each entry in $\mathbf{B}$ and $\mathbf{\Psi}$ does not vary too much for it to be useful. For example, when the columns of $\mathbf{B}$ and $\mathbf{\Psi}$ are chosen from DFT matrices, we optimally have $\mu^B_{max} = 1$, $\mu^\Psi_{max} = 1$.

We would like to develop a non-uniform recovery result to take advantage of the randomness introduced by using more parallel measurements. We first comment on our choice of assumptions. The first and third assumptions are on the sensing matrix. Note that $\mathbf{B}$ is chosen uniformly at random from all possible row permutations of a fixed matrix $\mathbf{B}_0$. The orthogonality of the columns of $\mathbf{B}$ is not affected by row permutations since $\mathbf{B}_i^*\mathbf{B}_j = \sigma(\mathbf{B}_i)^*\sigma(\mathbf{B}_j)$. Consider when the rows of $\mathbf{B}$ are randomly permuted. For any $\mathbf{h}$ fixed, we have the coil sensitivity $\mathbf{s} = \mathbf{B}\mathbf{h}$ to be randomly permuted copies of some underlying $\mathbf{s}_0 = \mathbf{B}_0\mathbf{h}$. On the other hand, if we assume no randomness in $\mathbf{B}$, if there are zero entries in $\mathbf{s} = \mathbf{B}\mathbf{h}$, it is impossible to recover uniquely the values of the image $\mathbf{z}$ at those locations. The randomness helps in providing an average case analysis for each coil. The third assumption is a standard one in compressive sensing theory on the random sampling scheme for k-space/Fourier space measurements. The assumption on $\mathbf{H}$ is critical for developing an average cage analysis to explain the improved results when more parallel measurements are taken.

Ideally, we would like to derive a proof for the setup when both $\mathbf{B}$ and $\mathbf{\Psi}$ are fixed, and the randomness comes from the support set S of $\mathbf{z}$ alone. The result would be analogous to the conditioning of the random sub-matrices of a fixed sampling matrix, e.g. see Chapter 14 of [21]. Yet, we encountered some unsolvable difficulties in finding a realistic bound for the incoherence measure between the sub-blocks of the sampling matrix $\mathbf{A}$. Nevertheless, we still provide a formal statement and some discussions in the Appendix 7. The main theorem enables self-calibration and ensures stable and robust recovery in the presence of sparsity defects and noise.

Next, let us discuss the reason of including multiple measurements/coils and the probability model on $\mathbf{H}$. Consider the worst case scenario where all the $\mathbf{h}_i$ are identical, then the measurements $\mathbf{Y} = \mathbf{A}\mathbf{X}$ consist of identical columns. By including multiple measurements, we are not gaining more information on $\mathbf{X}$ regardless of the number of repeated measurements taken. Hence, employing multiple measurements does not necessarily outperform the single measurement case. The following proposition gives a formal statement. The proof follows exactly from Proposition 4.1 in [19].

Lemma 2 *(Worst case analysis) Suppose there exist a* $\boldsymbol{X}_1 \in \mathbb{C}^{kN\times 1}$ *that the* $\ell_{1,2}$ *minimization fails to receiver from* $\boldsymbol{Y}_1 = \boldsymbol{A}\boldsymbol{X}_1$*. Then the* $\ell_{1,2}$ *minimization fails to recover* $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{X}_1 & \boldsymbol{X}_1 & \cdots \boldsymbol{X}_1\end{bmatrix} \in \mathbb{C}^{kN\times C}$ *from* $[\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_1] = \boldsymbol{A}\boldsymbol{X}$*.*

However, as observed in many real world applications, one would expect multiple measurements to provide more information that should aid the recovery tasks in mind. In an MRI setting, it has been shown that parallel imaging (pMRI) outperforms single-coil MRI in many cases, and the incoherent coils are often used as an explanation for the improved results. To understand the role of including multiple measurements, and from the worst case analysis, it suggests a need to impose some probability model on the signals to be recovered in order to prove some average case results.

We consider the underlying image $\mathbf{z} = \mathbf{\Psi}\mathbf{x}$ to be fixed and only impose some probability distribution on $\mathbf{h}_i$. That is to say we are imaging a fixed object with varying coil sensitivities where the spanning coefficient of the coils follows a certain probability distribution. For the two choices on the probability models on $\mathbf{H}$, they all ensure incoherence between the coils and that the coil sensitivities have the same energy. Observe first that, the incoherence between coils $\mathbf{s}_i^*\mathbf{s}_j = (\mathbf{B}\mathbf{h}_i)^*(\mathbf{B}\mathbf{h}_j) = \mathbf{h}_i^*\mathbf{B}^*\mathbf{B}\mathbf{h}_j = \mathbf{h}_i^*\mathbf{h}_j$ is small for k large, and the coils all have the same energy $\mathbf{s}_i^*\mathbf{s}_i = (\mathbf{B}\mathbf{h}_i)^*(\mathbf{B}\mathbf{h}_i) = \mathbf{h}_i^*\mathbf{B}^*\mathbf{B}\mathbf{h}_i = \mathbf{h}_i^*\mathbf{h}_i = 1$.

The proof of the main theorem follows a well-established proof outline of compressive sensing [21]. First, we refer to a sufficient condition from [20] to ensure a robust and stable recovery. Subsequently, we analyze the circumstances under which these conditions can be met. We show that under a suitable probability model of the calibration parameters, the sufficient conditions for recovery are satisfied with a higher probability with increasing number of coils. This means that, in practice, parallel imaging is likely to perform better than single imaging recovery.

4.2 A sufficient condition

Lemma 3 *(Uniqueness Results) Let* $\mathbf{X}_0 \in \mathbb{C}^{kN\times C}$ *and* $S \subset [N]$ *be arbitrary. Consider a linear map* $\boldsymbol{A}$ *from* $\mathbb{C}^{kN}$ *to* $\mathbb{C}^L$ *and noisy measurements* $\boldsymbol{Y} = \boldsymbol{A}\mathbf{X}_0 + \boldsymbol{\mathcal{N}}$ *with* $\|\boldsymbol{\mathcal{N}}\|_F \le \sigma$*.*
Suppose that

$$\|(\boldsymbol{A}_S)^*\boldsymbol{A}_S - \boldsymbol{I}_S\|_{2\to 2} \le \delta, \max_{j\in|S^c|} \|\boldsymbol{A}_S^*\boldsymbol{A}_{S_j^c}\|_{2\to 2} \le \beta \tag{22}$$

for some $\delta \in [0,1)$ *and* $\beta \ge 0$*.*
Moreover, suppose that there exists $\mathbf{V} \in \mathbb{C}^{L\times C}$ *such that the approximate dual certificate* $\boldsymbol{\mathcal{Y}} = \boldsymbol{A}^*\boldsymbol{V}$ *satisfies the following conditions*

$$\|\mathcal{P}_S\boldsymbol{\mathcal{Y}} - sgn(\boldsymbol{X}_0)\|_F \le \eta, \max_{j\in|S^c|} \|\mathcal{P}_{S_j^c}\boldsymbol{\mathcal{Y}}\|_F \le \theta \text{ and } \|\boldsymbol{V}\|_F \le \tau\sqrt{s}. \tag{23}$$

If $\rho = \theta + \frac{\eta\beta}{1-\delta} < 1$*,* $\mu = \frac{\sqrt{1+\delta}}{1-\delta}$*, then any solution* $\boldsymbol{X}^*$ *of the problem* (16) *obeys*

$$\|\boldsymbol{X}^* - \boldsymbol{X}_0\|_F \le 2\mu\sigma + (1 + \frac{\beta}{1-\delta})(\frac{1}{1-\rho})(2\|\mathcal{P}_{S^c}\boldsymbol{X}_0\|_{1,2} + 2\eta\mu\delta + 2\tau\sigma\sqrt{s}).$$

The derivation of the lemma is a direct application of Lemma 7 of [20] which relies on the sub-differential of $\partial(\|\cdot\|_{(1,2)})$. Given $\mathbf{Z} \in \mathbb{C}^{kN\times C}$, the sub-differential $\partial_{\mathbf{Z}}(\|\cdot\|_{(1,2)})$ being

$$\{\mathbf{V} \in \mathbb{C}^{kN\times C} | \mathbf{V}_{T_i} := \frac{\mathbf{Z}_{T_i\times[C]}}{\|\mathbf{Z}_{T_i\times[C]}\|_F}, \text{if } \|\mathbf{Z}_{T_i\times[C]}\|_F \ne 0, \|\mathbf{V}_{T_i}\|_F \le 1, otherwise\}.$$

4.3 Conditioning of $\mathbf{A}_S$

Next, we give the conditions on when our matrix $\mathbf{A}$ satisfies the condition $\|\mathbf{A}_S^*\mathbf{A}_S - \mathbf{I}_S\| \le \delta$. Recall our forward matrix $\mathbf{A}_S = \mathbf{F}_\Omega\mathbf{\Phi}_S$, the challenges come from the fact that its rows are dependent. Instead of

bounding the norm directly, we apply the triangular inequality and work on bounding the two terms in $\underbrace{\|\mathbf{\Phi}_S^*\mathbf{F}_\Omega^*\mathbf{F}_\Omega\mathbf{\Phi}_S - \mathbf{\Phi}_S^*\mathbf{\Phi}_S\|}_{(1)} + \underbrace{\|\mathbf{\Phi}_S^*\mathbf{\Phi}_S - \mathbf{I}_S\|}_{(2)}$ separately. Indeed, even when the full Fourier space is sampled, i.e., $\mathbf{F}_\Omega = \mathbf{F}$, the first term is zero and we would at the very least, require the condition (2) to be satisfied.

Lemma 4 *For operator $\mathbf{\Phi}$ defined in* (13) *which follows the randomness assumptions 4.1,*

$$\|\mathbf{\Phi}_S^*\mathbf{\Phi}_S - \boldsymbol{I}_S\| \leq \eta \tag{24}$$

with probability at least $1 - N^{-\alpha}$ if

$$N \geq C_\alpha {\mu_{max}^B}^2 {\mu_{max}^\Psi}^2 kn \log N/\eta^2, \tag{25}$$

where C_α grows linearly with respect to α. $\mu_{max}^B := \max_{i,j} \sqrt{L}|\mathbf{B}_{i,j}|$, $\mu_{max}^\Psi := \max_{i,j} \sqrt{N}|\mathbf{\Psi}_{i,j}|$.

The condition given is mild since in practice we can properly assume $N \gg kn$.

The probability is over the randomness in $\mathbf{B}$. The proof of the Lemma is derived from Lemma 4.3 of [36], where the authors consider the case $\mathbf{B}$ to be fixed but the rows of $\mathbf{\Psi}_S$ are chosen uniformly at random with replacement from the DFT matrix.

Remark: Due to the symmetry of $\mathbf{\Psi}$ and $\mathbf{B}$ in constructing the forward matrix $\mathbf{\Phi}$ (13), the result of Lemma (4) also holds true when $\mathbf{B}$ is fixed but the rows of $\mathbf{\Psi}$ chosen uniformly from all possible row permutations of a fixed matrix.

The next lemma is concerned with the tail bound for $\|\mathbf{\Phi}_S^*\mathbf{F}_\Omega^*\mathbf{F}_\Omega\mathbf{\Phi}_S - \mathbf{\Phi}_S^*\mathbf{\Phi}_S\|$.

Lemma 5 *Suppose that the rows of $\mathbf{F}_\Omega$ are chosen uniformly at random without replacement from $\mathbf{F}$, where $\mathbf{F}^*\mathbf{F} = \mathbf{F}\mathbf{F}^* = \frac{N}{L}$. For a realization of $\mathbf{\Phi}_S$ with $\|\mathbf{\Phi}_S^*\mathbf{\Phi}_S - \mathbf{I}_S\| \leq \eta_1$,*

$$\|\mathbf{\Phi}_S^*\boldsymbol{F}_\Omega^*\boldsymbol{F}_\Omega\mathbf{\Phi}_S - \mathbf{\Phi}_S^*\mathbf{\Phi}_S\| \leq \eta, \text{ with probability at least } 1-\epsilon$$

provided,

$$L \geq \max\Big\{C_4^2\frac{(1+\eta_1)}{\eta^2}, C_3\frac{1}{\eta}\Big\}{\mu_{max}^B}^2{\mu_{max}^\Psi}^2 kn \ln\Big(\frac{2C_2^2k^2n^2}{\epsilon}\Big).$$

Note that both the bounds are useful only when $\mathbf{B}$ and $\mathbf{\Psi}$ are "flat", i.e., $\mu_{max}^{\mathbf{B}} \sim O(1)$ and $\mu_{max}^{\mathbf{\Psi}} \sim O(1)$.

The next lemma addresses a bound for the parameter β in $\max_{j\in|S^c|} \|\mathbf{A}_S^*\mathbf{A}_{S_j^c}\|_{2\to 2} \leq \beta$ in 3.

Lemma 6 *Consider the forward matrix $\boldsymbol{A}$ defined in* (13) *and asssume that $\|\boldsymbol{A}_S^*\boldsymbol{A}_S - \boldsymbol{I}_S\|_{2\to 2} \leq \eta$,*

$$\max_{j\in[|S^c|]} \|\boldsymbol{A}_S^*\boldsymbol{A}_{S_j^c}\|_{2\to 2} \leq \sqrt{\frac{1+\eta}{L}}\mu_{max}^{\mathbf{\Psi}}. \tag{26}$$

4.4 Dual Certificate

Regarding the exact dual certificate, let $\mathbf{V} = \mathbf{A}_S(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\text{sgn}(\mathbf{X}_0)$, we have $\boldsymbol{\mathcal{Y}} = \mathbf{A}^*\mathbf{A}_S(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\text{sgn}(\mathbf{X}_0)$ is the exact dual certificate. The next lemma deals with bounding the parameter τ in $\|\mathbf{V}\|_F \leq \tau\sqrt{s}$ in 3. The bound obtained in the following lemma does not depend on the randomness of $\mathbf{H}$ or $sgn(\mathbf{X}_0)$.

Lemma 7 *Let $\boldsymbol{V} = \boldsymbol{A}_S(\boldsymbol{A}_S^*\boldsymbol{A}_S)^{-1}sgn(\boldsymbol{X}_0)$. Conditioned on $\|\boldsymbol{A}_S^*\boldsymbol{A}_S - \boldsymbol{I}_S\|_{2\to 2} \leq \eta$,*

$$\|\boldsymbol{V}\|_F \leq \frac{\sqrt{1+\eta}}{1-\eta}\sqrt{n}. \tag{27}$$

Lemma 8 *Suppose that either each column of $\boldsymbol{H}$ is chosen independently and uniformly at random with replacement from the columns of an orthonormal basis $\boldsymbol{W} \in \mathbb{C}^{k\times k}$ or that each column of $\boldsymbol{H}$ is chosen independently and uniformly at random from the complex sphere $S_{\mathbb{C}}^{k-1}$ (e.g., $h_l \in \mathbb{C}^k$ with $\|h_l\| = 1$). Conditioned on $\|\boldsymbol{A}_S^*\boldsymbol{A}_S - \boldsymbol{I}\| \leq \eta$, for any $s > 0$, if*

$$L > 6{\mu_{max}^{\Psi}}^2 \frac{1+\eta}{(1-\eta)^2} n, \tag{28}$$

then $\max_{j\in[|S^c|]} \|\mathcal{P}_{S_j^c}\boldsymbol{\mathcal{Y}}\|_F < 1/2$ *with probability at least* $1 - 2k\exp(-\frac{3C}{28k})$.

Remark: The condition on L is automatically met due to the bound $L \sim O(kn\log(k^2n^2N))$ from Lemma (5).

Hence, $\max_{j\in[|S^c|]} \|\mathcal{P}_{S_j^c}\boldsymbol{\mathcal{Y}}\|_F < 1/2$ with probability at least $1 - N^{-\alpha}$ if

$$C \geq \tilde{C}_\alpha \log(N)k,$$

where $\tilde{C}_\alpha$ grows linearly with respect to α.

5 Simulation

In the first part of the simulations, we demonstrate the requirements of L and C for exact recovery in terms of different values of k and n. We calculate the empirical probability of success for different values of k and n and show that the cutoffs for L and C are consistent with the analytically derived bounds.

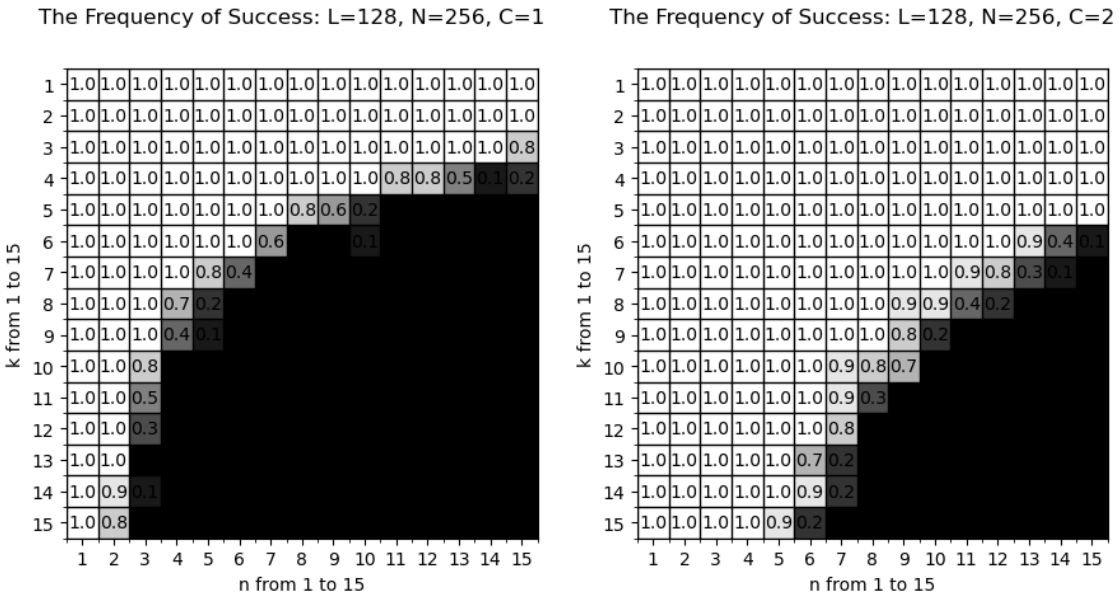
5.1 Empirical Success rate for the proposed method

For the first set of experiments, we fix the number of measurements $L = 128$ and the underlying signal dimension $N = 256$, and solve the optimization problem 17 for varying $(k, n), k, n \in [15]$ and $C \in \{1, 2, 4\}$. We follow the same setup as in the paper [36] and calculate the empirical probability of success of the proposed method. For each C and (k, n), we generate $\mathbf{z} \in \mathbb{C}^N$ with random support with cardinality n and the nonzero entries of $\mathbf{z}$ and $\mathbf{H} \in \mathbb{C}^{k\times C}$ and $\mathbf{z} \in \mathbb{C}^N$ yield $\mathcal{N}(0, 1)$. We repeat the experiment ten times, and each time new $\mathbf{H}$ and $\mathbf{z}$ are generated.

We count an experiment as a success if the relative error between the solution of our optimization problem $\hat{\mathbf{X}}$ and the truth $\mathbf{X}_0$ is less than 1%, i.e., $\frac{\|\hat{\mathbf{X}}-\mathbf{X}\|_F}{\|\mathbf{X}\|_F} \leq 1\%$. And the empirical success rate is a fraction defined as (total number of success) / (total number of experiments). We use the cvxpy [2][15] package of Python with the MOSEK [42] solver to solve the mixed norm optimization problem (17).

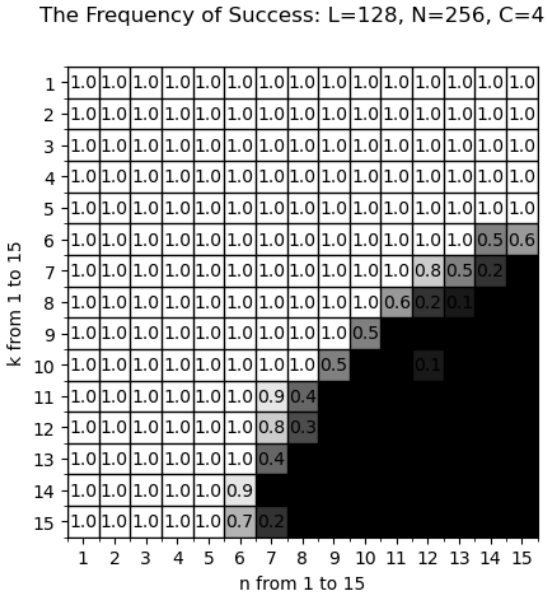
Even though in the theorems we treated $\mathbf{B}$ as a random matrix with row permutations, in the experiments we use a fixed $\mathbf{B}$. We choose $\mathbf{B}$ as a fixed matrix consisting of the first k columns of a random orthonormal matrix drawn from the $O(N)$ Haar distribution. It represents a stronger experiment than the theoretical treatment, as the theory only ensures recovery for most $\mathbf{B}$, while experimentally, we show the performance of a more relevant circumstance in practice. We test the performance of our proposed method $\mathbf{\Psi}$ as the DCT matrix or wavelet basis. The results for the sparsifying basis as the DCT are shown in Figure 1.

Next, we illustrate the barrier for reconstruction when $\mathbf{\Psi}$ is the wavelet transform. From the theory, a small mutual coherence, i.e., $\mu_{max}^{\Psi} \sim O(1)$, is required for the performance of the compressive sensing method. However, the Daubechies wavelet, as the most widely used wavelet in sparsifying MRI images [13], has high coherence, indicating a barrier in the proposed method. We followed the non-uniform sampling scheme in the Fourier measurements as suggested by [13][1] for attempt to mitigate the problem. However, even when



(a) 1 coil

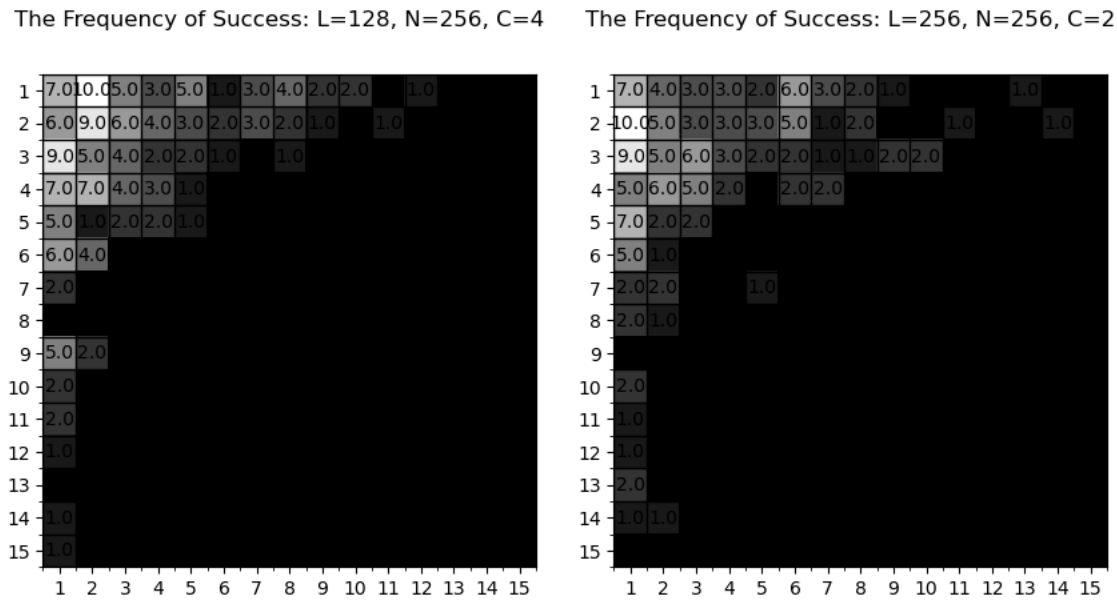
(b) 2 coil



(c) 4 coil

Figure 1: Phase transition plot of performance by solving (17) directly. The figures show empirical rate of success for a fixed sampling rate L and different pairs of (C, k, n). The numbers 1.0 means 100% rate of success and 0.0 means 0% rate of success. Observe that the transitional curve is improved for more C. However, when C reaches a certain level the improvements in the empirical rate of success saturates.

$L = N$ (when Fourier space is fully sampled), recovery is poor, as expected. We use the db4 wavelet as an illustration and the results are shown in Figure 2.



(a) 1 coil

(b) 2 coil

Figure 2

5.1.1 Improved results compared to l_1 minimization

Our proposed method minimizes $\sum_{i\in[N]} \|\mathbf{X}_{S_i}\|_2$ to promote block sparsity within a lifted signal and between coils. We demonstrate its advantage over minimizing $\sum_{i\in[C]} \|\mathbf{X}(:,i)\|_2$, which promotes row sparsity between coils. We minimize $\sum_{i\in[C]} \|\mathbf{X}(:,i)\|_2$ in (17) and use the same experimental setup. The results are shown in Figure 3.

5.1.2 Minimal L required for exact recovery is proportional to kn

Next, we perform two sets of experiments to show that the minimal measurements L required for exact recovery are proportional to kn.

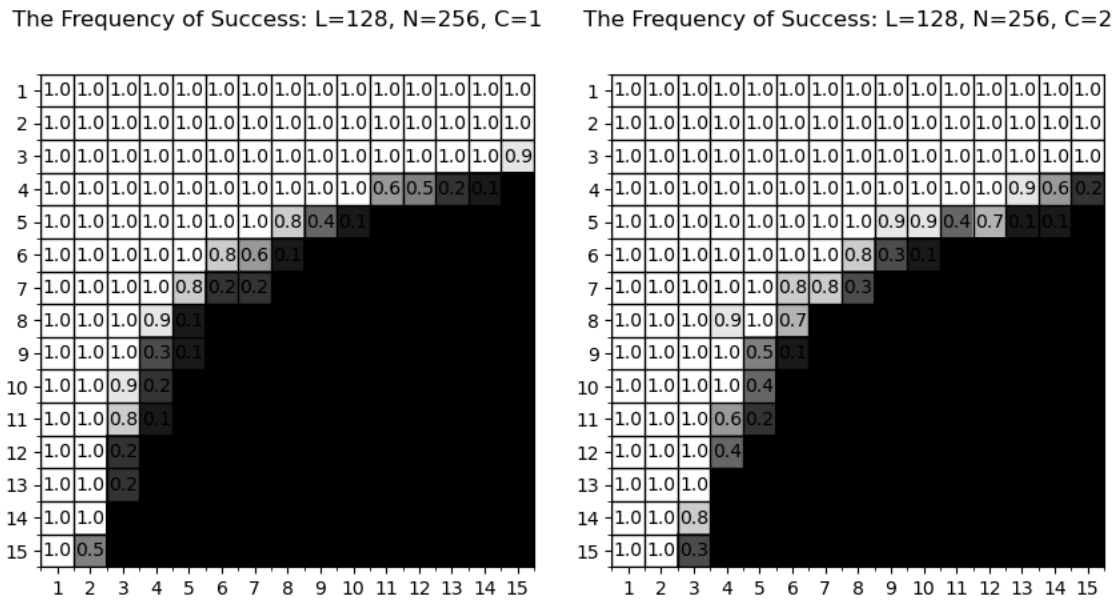
- For a fix $N = 256$, $n = 5$, we let k ranges from 1 to 15 and L varies from 10 to 200. We run the simulation for 10 times and calculate the empirical success rate
- For a fix $N = 256$, $k = 5$, we let n ranges from 1 to 15 and L varies from 10 to 200. We run the simulation for 10 times and calculate the empirical success rate.

We also repeat the same experiments for different $C = 1, 4$. The results are shown in the Figure 4.

5.2 Reconstruction of the analytical phantom

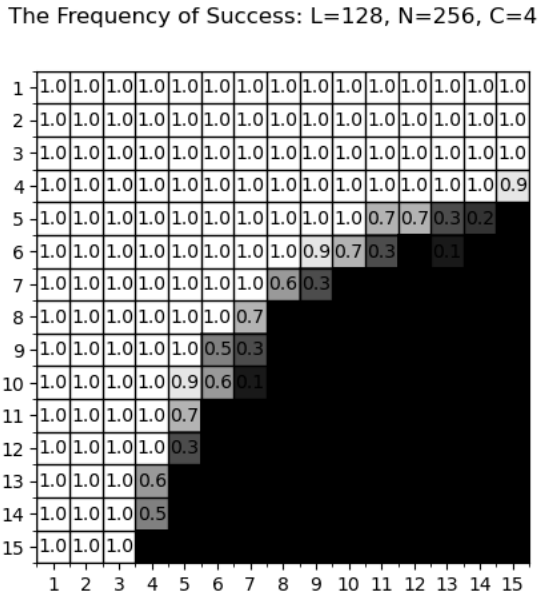
For the second part of the simulation, we focus on the MRI application. We point out two major differences from the last section:

- As a crutial difference to the simulation, we do not manually enforce the image to be sparse/transform sparse so to allow for sparsity deficiency.



(a) 1 coil

(b) 2 coil

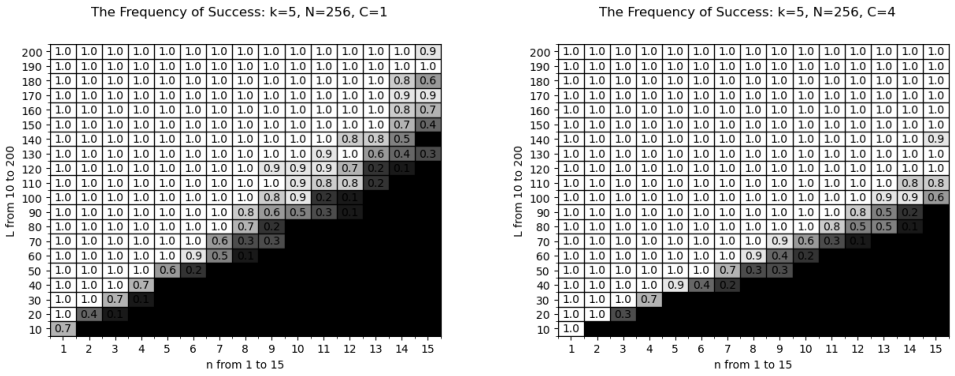


(c) 4 coil

Figure 3: Phase transition plot of performance by minimizing $\sum_{i\in[C]} \|\mathbf{X}(:,i)\|_2$. Observe that the improvement in the success rate is not obvious for including larger C. Compared to 1, the proposed optimization problem shows more improvement for larger C.

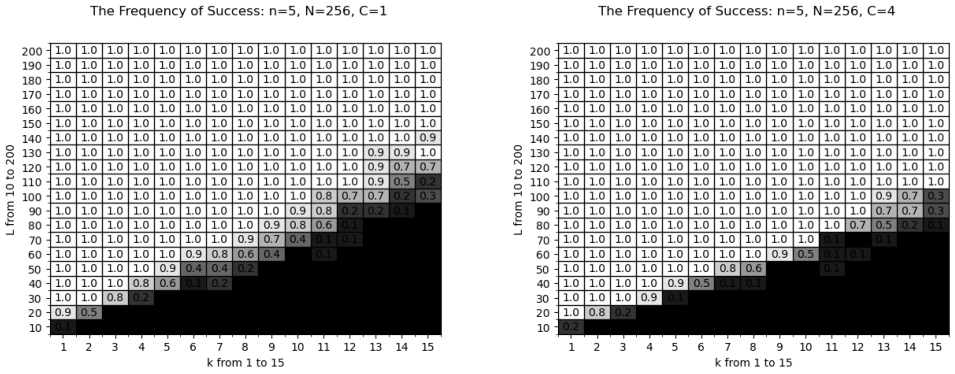
- We account for the presence of noise.
- Instead of solving the constraint optimization problem (17), we solve the regularized problem (18).

We use the analytical phantom of size 256 by 256 and randomly generated coil sensitivities following the



(a) L prop n, 1 coil (b) L prop n , 4 coil

Figure 4: The empirical rate of success for a fixed k and varying (L, n, C).

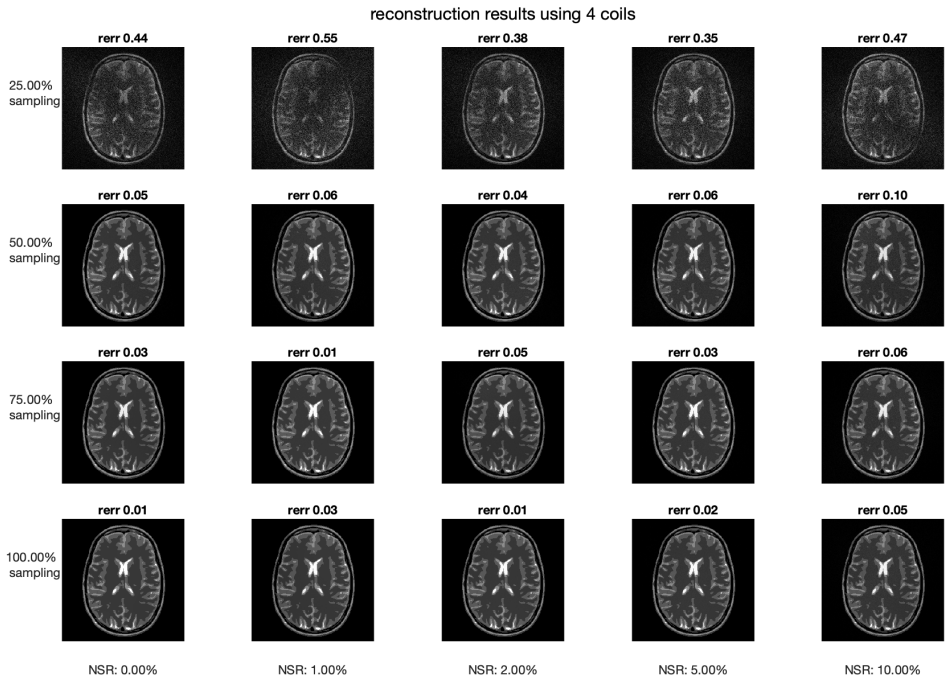


(a) L prop k, 1 coil (b) L prop k , 4 coil

Figure 5: The empirical rate of success for a fixed n and varying (L, k, C). Observe that the transitional curves for all four figures shows a linear trend.

Biot–Savart law. It is important to note that we do not manually enforce sparsity in our signal. For example, one may manually use a sparse estimation of the phantom $\mathbf{x}$, e.g., by crafting the signal by thresholding $\mathbf{z} = \mathbf{\Psi x} = 0, if|\mathbf{\Psi x}| < \tau$ and set $\mathbf{x} = \mathbf{\Psi}^*\mathbf{z}$. The thresholding effectively reduces the dimensionality of the problem and would favor the reconstruction results. However, such manipulation is only possible in simulations when the true signal is known beforehand. Our approach is more flexible and allows for sparsity deficiency in the underlying signal.

We use different number of coils to show the efficiency of the proposed method. We plot reconstruction results using different levels of sub-sampling rate and Gaussian noise level (i.e., noise to signal ratio). We report the relative error of the reconstructed image and the true phantom.



(a) 1 coil

Figure 6: The x-axis and y-axis represents varying noise-to-signal ratios and the sub-sampling rate respectively. Our observation indicates a gradual degradation in reconstruction quality with increased noise and reduced sampling.

6 Proofs

6.1 Proof of Lemma (4)

Recall that $\mathbf{\Phi} = \begin{bmatrix} (\bar{\mathbf{q}}_\mathbf{1} \bigotimes \mathbf{b_1})^* \\ \vdots \\ \bar{\mathbf{q}}_\mathbf{N} \bigotimes \mathbf{b_N})^* \end{bmatrix}$, where each row of $\mathbf{\Phi}$ depends only on one row of $\mathbf{B}$ and $\mathbf{\Psi}^*$. Due to symmetry, the probability distribution of $\mathbf{\Phi}$ is unchanged if we assume $\mathbf{B}$ is fixed but the rows of $\mathbf{\Psi}^*$ are random. More precisely, we assume the rows of $\mathbf{\Psi}^* \in \mathbb{C}^{N\times M}$ are chosen uniformly at random without replacement from some $\mathbf{\Psi}_0^* \in \mathbb{C}^{N\times M}$ with $\mathbf{\Psi}_0^*\mathbf{\Psi}_0 = \mathbf{I}_N$. Under such a model, the rows of $\mathbf{\Psi}^*$ are not independent, making it difficult to analyze the probability model on $\mathbf{\Phi}$ directly. On the other hand, if the rows of $\mathbf{\Psi}$ are chosen uniformly at random with replacement from $\mathbf{\Psi}_0$, we can apply a matrix Bernstein inequality to estimate the operator norm of $\mathbf{\Psi}_S^*\mathbf{\Psi}_S - \mathbf{I}_S$. We first state the following Lemma which follows almost exactly from Lemma 4.3 in [36]. The only minor change in the proof is in the estimation of the bounds of R and δ^2 for the matrix Bernstein inequality.

Lemma 9 *For any fixed matrix $\mathbf{B} \in \mathbb{C}^{N\times k}$ and $\mathbf{\Psi}^* \in \mathbb{C}^{N\times M}$ with its rows chosen uniformly at random with replacement from the rows of $\mathbf{\Psi}_0 \in \mathbb{C}^{N\times M}$, $\mathbf{\Psi}_0^*\mathbf{\Psi}_0 = \mathbf{I}_N$. For the matrix $\mathbf{\Phi}$ defined in 3.1,*

$$\|\mathbf{\Phi}_S^*\mathbf{\Phi}_S - \mathbf{I}_S\|_{F\to F} \leq \delta \tag{29}$$

with probability at least $1 - N^{-\alpha}$ if $N \geq C_\alpha {\mu_{max}^B}^2 {\mu_{max}^{\mathbf{\Psi}}}^2 kn \log N/\delta^2$.

The following Lemma relate Lemma (9) to the desired result.

Lemma 10 *For any integer $m \leq N$, let the random set $T' := \{t_1', \cdots, t_m'\}$, where $t_l' \in [N]$ are selected independently and uniformly at random from $[N]$. Let T_m be the random subset of $[N]$ chosen uniformly at random among all subsets of cardinality m.*
Given a $\mathbf{\Psi}^*$ with $\mathbf{\Psi}^*\mathbf{\Psi} = \boldsymbol{I}_N$, T' and T_m, define $\mathbf{\Psi}_{T'}^*$ and $\mathbf{\Psi}_{T_m}^*$ as matrices consisting of rows of $\mathbf{\Psi}^*$ indexed by T' and T_m respectively.
Given any fixed $\boldsymbol{B}$, consider the forward matrix $\mathbf{\Phi}_{T'}$ and $\mathbf{\Phi}_{T_m}$ formulated using $\mathbf{\Psi}_{T'}$ and $\mathbf{\Psi}_{T_m}$ respectively as defined in (13).
We have that

$$\mathbb{P}(\lambda_{min}(\mathbf{\Phi}_{T'}^*\mathbf{\Phi}_{T'}) = 0) \geq \mathbb{P}(\lambda_{min}(\mathbf{\Phi}_{T_m}^*\mathbf{\Phi}_{T_m}) = 0),$$

where $\lambda_{\min}(\mathbf{\Phi}_{T_m}^*\mathbf{\Phi}_{T_m})$ is the smallest eigenvalue of the positive semi-definite matrix.

Proof:

$$\begin{aligned}
&\mathbb{P}(\lambda_{min}(\mathbf{\Phi}_{T'}^*\mathbf{\Phi}_{T'}) = 0) \\
&= \Sigma_{k=1}^m \mathbb{P}(\lambda_{min}(\mathbf{\Phi}_{T'}^*\mathbf{\Phi}_{T'}) = 0 | card(T') = k)\mathbb{P}(card(T') = k) \\
&= \Sigma_{k=1}^m \mathbb{P}(\lambda_{min}(\mathbf{\Phi}_{T_k}^*\mathbf{\Phi}_{T_k}) = 0)\mathbb{P}(card(T') = k) \\
&\geq \mathbb{P}(\lambda_{min}(\mathbf{\Phi}_{T_m}^*\mathbf{\Phi}_{T_m}) = 0)\Sigma_{k=1}^m \mathbb{P}(card(T') = k) \\
&= \mathbb{P}(\lambda_{min}(\mathbf{\Phi}_{T_m}^*\mathbf{\Phi}_{T_m}) = 0).
\end{aligned}$$

The inequality holds because whenever k is smaller than m as, $T_k \subseteq T_m$. Hence, $\mathbf{\Phi}_{T_k}^*\mathbf{\Phi}_{T_k} \preccurlyeq \mathbf{\Phi}_{T_m}^*\mathbf{\Phi}_{T_m}$.

□

6.2 Proof of Lemma (5)

Suppose that the rows of $\mathbf{F}_\Omega$ are chosen uniformly at random without replacement from $\mathbf{F}$, where $\mathbf{F}^*\mathbf{F} = \mathbf{F}\mathbf{F}^* = \frac{N}{L}$. For a realization of $\mathbf{\Phi}_S$ with $\|\mathbf{\Phi}_S^*\mathbf{\Phi}_S - \mathbf{I}_S\| \leq \eta_1$,

$$\|\mathbf{\Phi}_S^*\mathbf{F}_\Omega^*\mathbf{F}_\Omega\mathbf{\Phi}_S - \mathbf{\Phi}_S^*\mathbf{\Phi}_S\| \leq \eta, \text{ with probability at least } 1-\epsilon$$

provided,

$$L \geq \max\{C_4^2 \frac{(1+\eta_1)}{\eta^2}, C_3 \frac{1}{\eta}\} {\mu_{max}^B}^2 {\mu_{max}^{\mathbf{\Psi}}}^2 kn \ln(\frac{2C_2^2 k^2 n^2}{\epsilon}).$$

Proof: Denote the rows of $\mathbf{F}_\Omega$ as $\mathbf{a}_l^*$ where $\mathbf{a}_l^*$ are chosen uniformly at random without replacement from the rows of the N point DFT matrix $\mathbf{F}$ with $\mathbf{F}^*\mathbf{F} = \frac{N}{L} I_N$. Define $\mathbf{Z}_l := \mathbf{\Phi}_S^* \mathbf{a}_l \mathbf{a}_l^* \mathbf{\Phi}_S, l = 1 \cdots L$, we have $\mathbb{E}(\mathbf{Z}_l) = \mathbf{\Phi}_S^* \mathbf{\Phi}_S / L$ and $\sum_{l=1}^{L} \mathbf{Z}_l - \mathbb{E}(\mathbf{Z}_l) = \mathbf{\Phi}_S^* \mathbf{F}_\Omega^* \mathbf{F}_\Omega \mathbf{\Phi}_S - \mathbf{\Phi}_S^* \mathbf{\Phi}_S$
By symmetrization, for any $p \geq 2$

$$\mathbb{E}\| \sum_{l=1}^{L} \mathbf{Z}_l - \mathbb{E}(\mathbf{Z}_l)\|_{2\to 2}^p \leq 2^p \mathbb{E}\| \sum_{l=1}^{L} \epsilon_l \mathbf{Z}_l\|_{2\to 2}^p,$$

where ϵ is a Rademacher sequence independent of Z.
Next, we use the following tail bound for matrix valued Rademacher sums,

$$\mathbb{P}_\epsilon(\| \sum_{l=1}^{L} \epsilon_l \mathbf{\Phi}_S^* \mathbf{a}_l \mathbf{a}_l^* \mathbf{\Phi}_S\|_{2\to 2} \geq t\sigma) \leq 2kne^{-t^2/2}, t \geq 0,$$

where $\sigma = \| \sum_{l=1}^{L} \mathbf{\Phi}_S^* \mathbf{a}_l \mathbf{a}_l^* \mathbf{\Phi}_S \mathbf{\Phi}_S^* \mathbf{a}_l \mathbf{a}_l^* \mathbf{\Phi}_S\|_{2\to 2}^{1/2} \leq \max_l\{|\mathbf{\Phi}_S^* \mathbf{a}_l|\} \|\mathbf{F}_\Omega \mathbf{\Phi}_S\|_{2\to 2} = \|\mathbf{\Phi}_S^* \mathbf{F}_\Omega^*\|_{1\to 2} \|\mathbf{F}_\Omega \mathbf{\Phi}_S\|_{2\to 2}$. The derivation of the bound can be found, e.g. in Proposition 8.20 of [21].
By Proposition 7.13 in [21] regarding the moment estimation,

$$(\mathbb{E}_\epsilon \| \sum_{l=1}^{L} \epsilon_l \mathbf{Z}_l\|_{2\to 2}^p)^{1/p} \leq C_1 (C_2 kn)^{1/p} \sqrt{p} \|\mathbf{\Phi}_S^* \mathbf{F}_\Omega^*\|_{1\to 2} \|\mathbf{F}_\Omega \mathbf{\Phi}_S\|_{2\to 2},$$

where $C_1 = e^{1/(2e)-1/2} \approx 0.73$, $C_2 = 2\sqrt{\pi} e^{1/6} \approx 4.19$. Hence, by Hölder's inequality,

$$\mathbb{E}\| \sum_{l=1}^{L} \epsilon_l \mathbf{Z}_l\|_{2\to 2}^p) \leq C_1^p (C_2 kn) p^{p/2} \mathbb{E}(\|\mathbf{\Phi}_S^* \mathbf{F}_\Omega^*\|_{1\to 2}^{2p})^{1/2} \mathbb{E}(\|\mathbf{F}_\Omega \mathbf{\Phi}_S\|_{2\to 2}^{2p})^{1/2}$$

Combining the bounds, we have

$$(\mathbb{E}\| \sum_{l=1}^{L} \mathbf{Z}_l - \mathbb{E}(\mathbf{Z}_l)\|_{2\to 2}^p)^{1/p} \leq 2(\mathbb{E}\| \sum_{l=1}^{L} \epsilon_l \mathbf{Z}_l\|_{2\to 2}^p)^{1/p} \leq 2C_1 (C_2 kn)^{1/p} \sqrt{p} \mathbb{E}(\|\mathbf{\Phi}_S^* \mathbf{F}_\Omega^*\|_{1\to 2}^{2p})^{1/2p} \mathbb{E}(\|\mathbf{F}_\Omega \mathbf{\Phi}_S\|_{2\to 2}^{2p})^{1/2p}.$$

Next, we estimate $\mathbb{E}(\|\mathbf{\Phi}_S^* \mathbf{F}_\Omega^*\|_{1\to 2}^{2p})^{1/2p}$ and $\mathbb{E}(\|\mathbf{F}_\Omega \mathbf{\Phi}_S\|_{2\to 2}^{2p})^{1/2p}$ respectively. For the first bound, observe that $\|\mathbf{\Phi}_S^* \mathbf{F}_\Omega^*\|_{1\to 2} \leq \|\mathbf{\Phi}_S^*\|_{1\to 2} \|\mathbf{F}_\Omega^*\|_{1\to 1}$. Also, recall that $\mathbf{\Phi}_S^* = \begin{bmatrix} \bar{\mathbf{q}}_{\mathbf{1},\mathbf{S}} \bigotimes \mathbf{b_1} & | \cdots & | \bar{\mathbf{q}}_{\mathbf{N},\mathbf{S}} \bigotimes \mathbf{b_N} \end{bmatrix}$. Hence, $\|\mathbf{\Phi}_S^*\|_{1\to 2} = \max_i \|\mathbf{q}_{i,s} \otimes \mathbf{b}_i\|_2 = \max_i \|\mathbf{q}_{i,s}\|_2 \max_i \|\mathbf{b}_i\|_2$. By definition, $\mu_{max}^{\mathbf{B}} = \sqrt{N} \max_{i,j} |\mathbf{B}_{i,j}|$ and $\mu_{max}^{\mathbf{\Psi}} = \sqrt{N} \max_{i,j} |\mathbf{\Psi}_{i,j}|$, we have $\|\mathbf{\Phi}_S^*\|_{1\to 2} \leq \frac{\sqrt{kn}}{N} \mu_{max}^B \mu_{max}^{\mathbf{\Psi}}$.

To bound $\|\mathbf{F}_\Omega^*\|_{1\to 1}$, we explicitly write out that

$$\|\mathbf{F}_\Omega^*\|_{1\to 1} = \max_{\|x\|_1=1} \sum_{j\in[N]} |\sum_{i\in[L]} \overline{\mathbf{F}}_{i,j} x_i| \leq \max_{\|x\|_1=1} N \max_j |\sum_i \overline{\mathbf{F}}_{i,j} x_i| \leq N \max_{i,j} |\mathbf{F}_{i,j}| \leq N/\sqrt{L},$$

since the rows of $\mathbf{F}_\Omega$ are chosen uniformly from the N point DFT matrix $\mathbf{F}$ for which $\mathbf{F}^*\mathbf{F} = \frac{N}{L} I$.

For the term $\mathbb{E}(\|\mathbf{F}_\Omega\mathbf{\Phi}_S\|_{2\to 2}^{2p})^{1/2p}$, we first derive a Bernoulli version where each rows of $\mathbf{F}_\Omega$ are the rows of $\mathbf{F}$ multiplied by independent Bernoulli selectors. More precisely, consider $\mathbf{F}_\Omega = \mathbf{PF}$, where $\mathbf{P} = \text{diag}(\delta_1, \cdots, \delta_N)$ and δ_i are independent Bernoulli variables with $\mathbb{E}(\delta_i) = L/N$.

Apply Lemma 14.3 [21] on the non-square matrix $\mathbf{\Phi}_S^*\mathbf{F}_\Omega^*$, for any $p \geq 1$,

$$\mathbb{E}_\delta(\|\mathbf{\Phi}_S^*\mathbf{F}^*\mathbf{P}\|_{2\to 2}^{2p})^{1/2p} \leq \sqrt{2}C_1(C_2kn)^{1/p}\sqrt{p}\mathbb{E}(\|\mathbf{\Phi}_S^*\mathbf{F}_\Omega^*\|_{1\to 2}^{2p})^{1/2p} + \sqrt{\frac{L}{N}}\|\mathbf{\Phi}_S^*\mathbf{F}_\Omega^*\|_{2\to 2}.$$

Also since $\|\mathbf{F}_\Omega\mathbf{\Phi}_S\|_{2\to 2} \leq \|\mathbf{F}_\Omega\|_{2\to 2}\|\mathbf{\Phi}_S\|_{2\to 2} \leq \sqrt{\frac{N}{L}}\|\mathbf{\Phi}_S\|_{2\to 2}$.Combining the bounds, we have

$$\begin{aligned}
&(\mathbb{E}\|\sum_{l=1}^{L}\mathbf{Z}_l - \mathbb{E}(\mathbf{Z}_l)\|_{2\to 2}^p)^{1/p} \\
&\leq 2C_1(C_2kn)^{1/p}\sqrt{p}\sqrt{\frac{kn}{L}}\mu_{max}^B\mu_{max}^{\mathbf{\Psi}}(\sqrt{2}C_1(C_2kn)^{1/p}\sqrt{p}\sqrt{\frac{kn}{L}}\mu_{max}^B\mu_{max}^{\mathbf{\Psi}} + \|\mathbf{\Phi}_S\|_{2\to 2}) \\
&\leq (C_2kn)^{2/p}(2\sqrt{2}C_1^2\frac{knp}{L}{\mu_{max}^B}^2{\mu_{max}^{\mathbf{\Psi}}}^2 + 2C_1\sqrt{\frac{knp}{L}}\mu_{max}^B\mu_{max}^{\mathbf{\Psi}}\|\mathbf{\Phi}_S\|_{2\to 2}).
\end{aligned}$$

Then, we apply a tail bound from moment estimation, see Proposition 7.15 from [21]. For $u \geq 1$,

$$\mathbb{P}(\|\sum_{l=1}^{L}\mathbf{Z}_l - \mathbb{E}(\mathbf{Z}_l)\| > e(\alpha_1 u + \alpha_2\sqrt{u})) \leq \beta e^{-u},$$

where $\beta = C_2^2k^2n^2$, $\alpha_1 = 2\sqrt{2}C_1^2\frac{kn}{L}{\mu_{max}^B}^2{\mu_{max}^{\mathbf{\Psi}}}^2$, $\alpha_2 = 2C_1\sqrt{\frac{kn}{L}}\mu_{max}^B\mu_{max}^{\mathbf{\Psi}}\|\mathbf{\Phi}_s\|_{2\to 2}$.

The result implies, for any realization of $\mathbf{\Phi}_S$, $\|\mathbf{\Phi}_s^*\mathbf{F}_\Omega^*\mathbf{F}_\Omega\mathbf{\Phi}_S - \mathbf{\Phi}_s^*\mathbf{\Phi}_s\| \leq \eta$ with probability at least $1-\epsilon$ provided,

$$e\alpha_1\ln(C_2^2k^2n^2/\epsilon) \leq \eta/2 \text{ and } e\alpha_2\sqrt{\ln(C_2^2k^2n^2/\epsilon)} \leq \eta/2.$$

By the definition of α_1 and α_2,

$$\begin{aligned}
e\alpha_1\ln(C_2^2k^2n^2/\epsilon) \leq \eta/2 \to & L \geq \frac{C_3{\mu_{max}^B}^2{\mu_{max}^{\mathbf{\Psi}}}^2kn\ln(\frac{C_2^2k^2n^2}{\epsilon})}{\eta}, C_3 = 4\sqrt{2}eC_1^2. \\
e\alpha_2\sqrt{\ln(C_2^2k^2n^2/\epsilon)} \leq \eta/2 \to & \sqrt{\frac{kn}{L}}\|\mathbf{\Phi}_S\| \leq \frac{\eta}{C_4\mu_{max}^B\mu_{max}^{\mathbf{\Psi}}\sqrt{\ln(C_2^2k^2n^2/\epsilon)}}, C_4 = 4eC_1 \\
& \to L \geq \frac{C_4^2{\mu_{max}^B}^2{\mu_{max}^{\mathbf{\Psi}}}^2kn\ln(C_2^2k^2n^2/\epsilon)\|\mathbf{\Phi}_S\|_{2\to 2}^2}{\eta^2}.
\end{aligned}$$

Then conditioned on $\|\mathbf{\Phi}_S^*\mathbf{\Phi}_S - \mathbf{I}_S\| \leq \eta_1$, we have $\|\mathbf{\Phi}_S\|_{2\to 2}^2 \leq 1+\eta_1$, the two conditions are equivalent to requiring,

$$L \geq \max\{C_4^2\frac{(1+\eta_1)}{\eta^2}, C_3\frac{1}{\eta}\}{\mu_{max}^B}^2{\mu_{max}^{\mathbf{\Psi}}}^2kn\ln(\frac{C_2^2k^2n^2}{\epsilon}).$$

This completes the proof for the Bernoulli case. For the case when the rows of $\mathbf{F}_\Omega$ are chosen uniformly from F, use a similar argument as in the later part of the proof of Theorem 14.1 [21], where it shows the probability for the uniform model is bounded by twice the one for the Bernoulli model derived here.

□

6.3 Proof of Lemma (6)

Proof: Partition $\mathbf{A}$ into blocks of one row and N columns. $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_N]$, where by definition $\mathbf{A}_i = \mathbf{F}_\Omega \mathbf{\Phi}_i$.

$$\begin{aligned}
&\max_{j\in[|S^c|]} \|\mathbf{A}_S^* \mathbf{A}_{S_j^c}\|_{2\to 2} \\
&\leq \|\mathbf{A}_S\|_{2\to 2} \max_i \|\mathbf{A}_i\|_{2\to 2} \\
&\leq \sqrt{1+\eta} \max_i \|\mathbf{A}_i\|_{2\to 2}.
\end{aligned}$$

The last inequality is implied by $\|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}_S\|_{2\to 2} \leq \ \eta$.

Since $\mathbf{A}_i = \mathbf{F}_\Omega \mathbf{\Phi}_i = \mathbf{F}_\Omega \text{diag}(\mathbf{\Psi}_i)\mathbf{B}$, we have $\|\mathbf{A}_i\|_{2\to 2} \leq \|\mathbf{F}_\Omega\|_{2\to 2}\|\mathbf{\Phi}_i\|_{2\to 2}$. Under the assumptions, the rows of $\mathbf{F}_\Omega$ are chosen from the DFT matrix $\mathbf{F}$ with $\mathbf{F}^*\mathbf{F} = \frac{N}{L}\mathbf{I}_N$, we have $\mathbf{F}_\Omega \mathbf{F}_\Omega^* = \frac{N}{L}\mathbf{I}_L \to \|\mathbf{F}_\Omega\|_{2\to 2} = \sqrt{\frac{N}{L}}$. Recall that $\mu_{max}^{\mathbf{\Psi}} = \sqrt{N}\max_{i,j}|\mathbf{\Psi}_{i,j}|$, we have for any i, j, $|\mathbf{\Psi}_{i,j}|^2 \leq \frac{1}{N}(\mu_{max}^{\mathbf{\Psi}})^2$. Hence, for any i, $\|\mathbf{\Phi}_i\|_{2\to 2}^2 = \|\mathbf{\Phi}_i^*\mathbf{\Phi}_i\| = \|\mathbf{B}^* \text{diag}([|\mathbf{\Psi}_{i,1}|^2, \cdots, |\mathbf{\Psi}_{i,N}|^2])\mathbf{B}\|_{2\to 2} \leq \frac{1}{N}(\mu_{max}^{\mathbf{\Psi}})^2 \|\mathbf{B}^*\mathbf{B}\|_{2\to 2} = \frac{1}{N}(\mu_{max}^{\mathbf{\Psi}})^2$. The last equality is due to the assumption that $\mathbf{B}^*\mathbf{B} = I_k$.

Combining the bounds,

$$\max_{j\in[|S^c|]} \|\mathbf{A}_S^* \mathbf{A}_{S_j}\|_{2\to 2} \leq \sqrt{\frac{1+\eta}{L}}\mu_{max}^{\mathbf{\Psi}}.$$

□

6.4 Proof of Lemma (7)

Proof: Let $\mathbf{V} = \mathbf{A}_S(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\text{sgn}(\mathbf{X}_0)$, define

$$\boldsymbol{\mathcal{Y}} := \mathbf{A}^*\mathbf{A}_S(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\text{sgn}(\mathbf{X}_0).$$

$\boldsymbol{\mathcal{Y}}$ is the exact dual certificate since $\mathcal{P}_S\boldsymbol{\mathcal{Y}} = \text{sgn}(\mathbf{X}_0)$. Next, we have

$$\|\mathbf{V}\|_F = \|\mathbf{A}_S(\mathbf{A}_S^*\mathbf{A}_S)^{-1} sgn(\mathbf{X}_0)\|_F \leq \frac{\sqrt{1+\eta}}{1-\eta}\sqrt{n}. \tag{30}$$

The inequality comes from the implication by the conditioning of $\mathbf{A}_S$ that $\mathbf{A}_S^*\mathbf{A}_S$ is invertible with $\|(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\|_{2\to 2} \leq \frac{1}{1-\eta}$ and $\|\mathbf{A}_S\|_{2\to 2} \leq \sqrt{1+\eta}$.

□

6.5 Proof of Lemma (8)

We consider the case when the spanning coefficients are chosen uniformly random from a fixed basis. The structure of the proof is similar to the Gaussian case. Suppose the columns of $\mathbf{H}$ are chosen uniform randomly with replacement from the columns of $\mathbf{W} \in \mathbb{C}^{k\times k}$ where $\mathbf{W}^*\mathbf{W} = \mathbf{I}_k$.

Proof: Let $\mathbf{A}^\dagger = (\mathbf{A}_S^*\mathbf{A}_S)^{-1}\mathbf{A}_S^*$. Suppose there exist an α such that $\max_{j\in S^c} \|\mathbf{A}^\dagger \mathbf{A}_{S_j}\|_F \leq \alpha$. We postpone the estimate of the α to the latter part of the proof.

Define $\tilde{\boldsymbol{X}} = \left[\begin{array}{c|c|c|c} \mathbf{h_1} sgn(z_1) & \mathbf{h_2} sgn(z_1) & \cdots & \mathbf{h_c} sgn(z_1) \\ \hline \mathbf{h_1} sgn(z_2) & \mathbf{h_2} sgn(z_2) & \cdots & \mathbf{h_c} sgn(z_2) \\ \hline \vdots & \vdots & \vdots & \\ \mathbf{h_1} sgn(z_N) & \mathbf{h_2} sgn(z_N) & \cdots & \mathbf{h_c} sgn(z_N) \end{array}\right]$. Then $sgn(\mathbf{X}_0) = \tilde{\boldsymbol{X}}/\|\mathbf{H}\|_F \in \mathbb{C}^{kN\times C}$, where $\mathbf{H} = \big[\mathbf{h_1}|\mathbf{h_2}|\cdots|\mathbf{h_C}\big]$, $\mathbf{H} \in \mathbb{C}^{k\times C}$.

We want to estimate the norm of $\max_{l\in S^c} \|\operatorname{sgn}(\mathbf{X}_0)^*\mathbf{A}^\dagger\mathbf{A}_{S^l}\|_F$,

$$\max_{l\in S^c} \|\operatorname{sgn}(\mathbf{X}_0)^*\mathbf{A}^\dagger\mathbf{A}_{S^l}\|_F \le \max_{l\in S^c} \|\mathbf{A}^\dagger\mathbf{A}_{S^l}\|_F \|\operatorname{sgn}(\mathbf{X}_0)\|_{2\to 2} \le \alpha \|\tilde{\boldsymbol{X}}\|_{2\to 2}/\|\mathbf{H}\|_F.$$

By the Cauchy Schwartz inequality, for any $\mathbf{y}$, $\|\tilde{\boldsymbol{X}}\mathbf{y}\| \le \|\mathbf{H}\|_{2\to 2}\|\operatorname{sgn}(\mathbf{z})\|_2\|\mathbf{y}\|_2 \le \sqrt{n}\|\mathbf{H}\|_{2\to 2}$.

We want to bound the two terms, $\|\mathbf{H}\|_F$ and $\|\mathbf{H}\|_{2\to 2}$. For $\|\mathbf{H}\|_F$, since each columns of $\mathbf{H}$ is has norm 1, we have $\|\mathbf{H}\|_F = \sqrt{C}$.

To bound $\|\mathbf{H}\|_{2\to 2}$, we apply the matrix Bernstein inequality to $\mathbf{H}^*\mathbf{H} = \Sigma_{l=1}^C \mathbf{h}_l\mathbf{h}_l^*$. Define $\mathbf{Z}_l = \mathbf{h}_l\mathbf{h}_l^* - \frac{1}{k}\mathbf{I}_k$, $\mathbf{Z}_l$ are independent with $\mathbb{E}(\mathbf{Z}_l) = 0$,

$$R = \max_l \|\mathbf{Z}_l\| = \|\mathbf{h}_l\mathbf{h}_l^* - \frac{1}{k}I\| \le \max_l \max(\|\mathbf{h}_l\|^2, \frac{1}{k}) = 1, \tag{31}$$

$$\sigma^2 = \|\Sigma_{l=1}^C \mathbb{E}(\mathbf{Z}_l\mathbf{Z}_l^*)\| = \|\Sigma_{l=1}^C \mathbb{E}(\mathbf{h}_l \underbrace{\mathbf{h}_l^*\mathbf{h}_l}_{=1}\mathbf{h}_l^*) - \frac{2}{k}\mathbb{E}(\mathbf{h}_l\mathbf{h}_l^*) + \frac{I}{k^2}\| = \|\Sigma_{l=1}^C(\frac{I}{k} - \frac{I}{k^2})\| \le C/k. \tag{32}$$

Hence,

$$\mathbb{P}(\|\mathbf{H}\mathbf{H}^* - \frac{C}{k}I\| \ge t) \le 2k\exp(-\frac{t^2/2}{\sigma^2 + Rt/3}) = 2k\exp(-\frac{t^2/2}{\frac{C}{k} + t/3}). \tag{33}$$

Take $t = \frac{C}{2k}$, we have $\|\mathbf{H}\| \le \sqrt{3C/2k}$ with probability at least $1 - 2k\exp(-\frac{3C}{28k})$.

Combining the bounds, we have

$$\frac{\sqrt{1+\eta}}{1-\eta}\sqrt{\frac{1}{L}}\mu_{max}^{\boldsymbol{\Psi}}\sqrt{3n/2} < \theta < 1 \tag{34}$$

with probability at least $1 - 2k\exp(-\frac{3C}{28k})$.

The case when the coefficients $\mathbf{h}_l, l \in [C]$ are chosen independently and uniformly at random from the complex sphere $S_{\mathbb{C}}^{k-1}$ follows the exact argument by noticing that $\mathbf{h}_l = \frac{\mathbf{s}_l}{\|\mathbf{s}_l\|_2}$ where $\mathbf{s}_l$ are independent Gaussian vector drawn from $\frac{1}{\sqrt{2}}\mathcal{N}(0,1) + i\frac{1}{\sqrt{2}}\mathcal{N}(0,1)$. □

Acknowledgement

The authors acknowledge support from NSF DMS-2208356, P41EB032840, and NIH R01HL16351.

7 Appendix

7.1 Conditioning of random submatrices

Ideally, we would like to consider the setting in which all the matrices in the forward model are fixed and the randomness comes from the support of the signals $\mathbf{x} \in \mathbb{C}^N$. We consider the setup where the support S selected uniformly at random among all subsets of $[N]$ of cardinality n.

The next lemma is related to the conditioning of random block submatrices for any fixed matrix $\boldsymbol{\Phi}$.

Lemma 11 *Suppose* $\boldsymbol{\Phi} = \begin{bmatrix}\boldsymbol{\Phi}_1 & \cdots & \boldsymbol{\Phi}_N\end{bmatrix}$, $\boldsymbol{\Phi}_i \in \mathbb{C}^{N\times k}$ *and* $\boldsymbol{\Phi}_i^*\boldsymbol{\Phi}_i = \mathbf{I}_k$ *for all i. Let S be a subset of* $[N]$ *selected at random according to the Bernoulli model with* $\mathbb{E}Card(S) = n$ *or uniformly random from all subset of* $[N]$ *of cardinality n.Let* $P = diag[\delta_1, \ldots, \delta_N]$, *where* δ_i *are i.i. d Bernoulli variables with* $\mathbb{E}\delta_i = \frac{n}{N}$.

If there exists a positive constant c, and for η, $\epsilon \in (0,1)$,

$$\mu_{block} \leq \frac{c\eta}{\ln(kN/\epsilon)}, \mu_{block} = \max_{i \neq k} \|\mathbf{\Phi}_i^* \mathbf{\Phi}_k\|_2,$$
$$\frac{n}{N}\|\mathbf{\Phi}\|_2^2 \leq \frac{c\eta^2}{\ln(kN/\epsilon)}.$$

Then, with probability at least $(1-\epsilon)$, $\|\mathbf{\Phi}_S^* \mathbf{\Phi}_S - \mathbf{I}_S\| \leq \eta$.

However, take into account the structure of our specific $\mathbf{\Phi}$, $\mathbf{\Phi} = \begin{bmatrix}\text{diag}(\boldsymbol{\psi}_1)\mathbf{B} & \cdots & \text{diag}(\boldsymbol{\psi}_N)\mathbf{B}\end{bmatrix}$. The condition on μ_{block} is strict.

References

[1] BEN ADCOCK et al. "Breaking the coherence barrier: A new theory for compressed sensing". In: *Forum of Mathematics, Sigma* 5 (2017). DOI: 10.1017/fms.2016.32.

[2] Akshay Agrawal et al. "A rewriting system for convex optimization problems". In: *Journal of Control and Decision* 5.1 (2018), pp. 42–60.

[3] M. Aharon, M. Elad, and A. Bruckstein. "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation". In: *IEEE Transactions on Signal Processing* 54.11 (2006), pp. 4311–4322. DOI: 10.1109/tsp.2006.881199.

[4] Ali Ahmed, Benjamin Recht, and Justin Romberg. "Blind deconvolution using convex programming". In: *IEEE Transactions on Information Theory* 60.3 (2014), pp. 1711–1732. DOI: 10.1109/tit.2013.2294644.

[5] Mehmet Akçakaya et al. "Low-dimensional-structure self-learning and Thresholding: Regularization beyond compressed sensing for MRI reconstruction". In: *Magnetic Resonance in Medicine* 66.3 (2011), pp. 756–767. DOI: 10.1002/mrm.22841.

[6] Sohail Bahmani and Justin Romberg. "Lifting for blind deconvolution in random mask imaging: Identifiability and convex relaxation". In: *SIAM Journal on Imaging Sciences* 8.4 (2015), pp. 2203–2238. DOI: 10.1137/141002165.

[7] Laura Balzano and Robert Nowak. "Blind calibration of Sensor Networks". In: *2007 6th International Symposium on Information Processing in Sensor Networks* (2007). DOI: 10.1109/ipsn.2007.4379667.

[8] Amir Beck and Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". In: *SIAM Journal on Imaging Sciences* 2.1 (2009), pp. 183–202. DOI: 10.1137/080716542.

[9] Cagdas Bilen et al. "Convex optimization approaches for blind sensor calibration using sparsity". In: *IEEE Transactions on Signal Processing* 62.18 (2014), pp. 4847–4856. DOI: 10.1109/tsp.2014.2342651.

[10] E.J. Candes, J. Romberg, and T. Tao. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information". In: *IEEE Transactions on Information Theory* 52.2 (2006), pp. 489–509. DOI: 10.1109/tit.2005.862083.

[11] Emmanuel J. Candès et al. "Phase retrieval via matrix completion". In: *SIAM Journal on Imaging Sciences* 6.1 (2013), pp. 199–225. DOI: 10.1137/110848074.

[12] Sunav Choudhary and Urbashi Mitra. "On identifiability in bilinear inverse problems". In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (2013). DOI: 10.1109/icassp.2013.6638476.

[13] Il Yong Chun, Ben Adcock, and Thomas M. Talavage. "Efficient compressed sensing sense parallel MRI reconstruction with joint sparsity promotion and mutual incoherence enhancement". In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2014). DOI: 10.1109/embc.2014.6944111.

[14] Patrick l. Combettes and Valérie R. Wajs. "Signal Recovery by Proximal Forward-Backward Splitting". In: *Multiscale Modeling & Simulation* 4.4 (2005), pp. 1168–1200. DOI: 10.1137/050626090.

[15] Steven Diamond and Stephen Boyd. "CVXPY: A Python-embedded modeling language for convex optimization". In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5.

[16] Michael Elad, Peyman Milanfar, and Ron Rubinstein. "Analysis versus synthesis in Signal Priors". In: *Inverse Problems* 23.3 (2007), pp. 947–968. DOI: 10.1088/0266-5611/23/3/007.

[17] Yonina C. Eldar, Patrick Kuppinger, and Helmut Bolcskei. "Block-sparse signals: Uncertainty Relations and Efficient Recovery". In: *IEEE Transactions on Signal Processing* 58.6 (2010), pp. 3042–3054. DOI: 10.1109/tsp.2010.2044837.

[18] Yonina C. Eldar and Moshe Mishali. "Robust recovery of signals from a structured union of subspaces". In: *IEEE Transactions on Information Theory* 55.11 (2009), pp. 5302–5316. DOI: 10.1109/tit.2009.2030471.

[19] Yonina C. Eldar and Holger Rauhut. "Average case analysis of multichannel sparse recovery using convex relaxation". In: *IEEE Transactions on Information Theory* 56.1 (2010), pp. 505–519. DOI: 10.1109/tit.2009.2034789.

[20] Axel Flinth. "Sparse blind deconvolution and demixing through l1,2-minimization". In: *Advances in Computational Mathematics* 44.1 (2017), pp. 1–21. DOI: 10.1007/s10444-017-9533-0.

[21] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer New York, 2015.

[22] R. Gribonval, G. Chardon, and L. Daudet. "Blind calibration for compressed sensing by convex optimization". In: *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2012). DOI: 10.1109/icassp.2012.6288477.

[23] Mark A. Griswold et al. "Generalized autocalibrating partially parallel acquisitions (grappa)". In: *Magnetic Resonance in Medicine* 47.6 (2002), pp. 1202–1210. DOI: 10.1002/mrm.10171.

[24] M. Guerquin-Kern et al. "Realistic analytical phantoms for parallel magnetic resonance imaging". In: *IEEE Transactions on Medical Imaging* 31.3 (2012), pp. 626–636. DOI: 10.1109/tmi.2011.2174158.

[25] Justin P Haldar. "low-rank modeling of local K-space neighborhoods (lORAKS) for constrained MRI". In: *IEEE transactions on medical imaging* 33.3 (2014), pp. 668–681. URL: https://pubmed.ncbi.nlm.nih.gov/24595341/.

[26] G. Harikumar and Y. Bresler. "Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms". In: *IEEE Transactions on Image Processing* 8.2 (1999), pp. 202–219. DOI: 10.1109/83.743855.

[27] H. Christian Holme et al. "ENLIVE: An efficient nonlinear method for calibrationless and robust parallel imaging". In: *Scientific Reports* 9.1 (2019). DOI: 10.1038/s41598-019-39888-7.

[28] Kyong Hwan Jin, Dongwook lee, and Jong Chul Ye. "A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix". In: *IEEE Transactions on Computational Imaging* 2.4 (2016), pp. 480–495. DOI: 10.1109/tci.2016.2601296.

[29] Michael Kech and Felix Krahmer. "Optimal injectivity conditions for bilinear inverse problems with applications to identifiability of Deconvolution Problems". In: *SIAM Journal on Applied Algebra and Geometry* 1.1 (2017), pp. 20–37. DOI: 10.1137/16m1067469.

[30] Kiryung Lee, Ning Tian, and Justin Romberg. "Fast and guaranteed blind multichannel deconvolution under a bilinear system model". In: *IEEE Transactions on Information Theory* 64.7 (2018), pp. 4792–4818. DOI: 10.1109/tit.2018.2840711.

[31] Kiryung Lee et al. "Blind recovery of sparse signals from subsampled convolution". In: *IEEE Transactions on Information Theory* 63.2 (2017), pp. 802–821. DOI: 10.1109/tit.2016.2636204.

[32] A. Levin et al. "Understanding blind deconvolution algorithms". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.12 (2011), pp. 2354–2367. DOI: 10.1109/tpami.2011.148.

[33] Xiaodong Li et al. "Rapid, robust, and reliable blind deconvolution via nonconvex optimization". In: *Applied and Computational Harmonic Analysis* 47.3 (2019), pp. 893–934. DOI: 10.1016/j.acha.2018.01.001.

[34] Zhi-Pei Liang and Paul C. Lauterbur. *Principles of Magnetic Resonance Imaging*. SPIE Optical Engineering Press, 2000.

[35] Shuyang Ling and Thomas Strohmer. "Self-calibration and bilinear inverse problems via linear least squares". In: *SIAM Journal on Imaging Sciences* 11.1 (2018), pp. 252–292. DOI: 10.1137/16m1103634.

[36] Shuyang ling and Thomas Strohmer. "Self-calibration and biconvex compressive sensing". In: *Inverse Problems* 31.11 (2015), p. 115002. DOI: 10.1088/0266-5611/31/11/115002.

[37] Michael lustig, David Donoho, and John M. Pauly. "Sparse MRI: The application of compressed sensing for rapid Mr Imaging". In: *Magnetic Resonance in Medicine* 58.6 (2007), pp. 1182–1195. DOI: 10.1002/mrm.21391.

[38] Michael Lustig and John M. Pauly. "Spirit: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space". In: *Magnetic Resonance in Medicine* 64.2 (2010), pp. 457–471. DOI: 10.1002/mrm.22428.

[39] Committee on the Mathematics and National Research Council Physics of Emerging Dynamic Biomedical Imaging. *Mathematics and physics of emerging biomedical imaging*. National Academy Press, 1996.

[40] Charles A. McKenzie et al. "Self-calibrating parallel imaging with automatic coil sensitivity extraction". In: *Magnetic Resonance in Medicine* 47.3 (2002), pp. 529–538. DOI: 10.1002/mrm.10087.

[41] Robert Morrison, Mathews Jacob, and Minh Do. "Multichannel estimation of coil sensitivities in Parallel MRI". In: *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2007). DOI: 10.1109/isbi.2007.356802.

[42] MOSEK ApS. *MOSEK Optimizer API for Python 10.1.17*. https://docs.mosek.com/latest/pythonapi/index.html. 2023.

[43] Samet Oymak et al. "Simultaneously structured models with application to sparse and low-rank matrices". In: *IEEE Transactions on Information Theory* 61.5 (2015), pp. 2886–2908. DOI: 10.1109/tit.2015.2401574.

[44] W.K. Pratt, J. Kane, and H.C. Andrews. "Hadamard transform image coding". In: *Proceedings of the IEEE* 57.1 (1969), pp. 58–68. DOI: 10.1109/proc.1969.6869.

[45] Klaas P. Pruessmann et al. "Sense: Sensitivity encoding for Fast Mri". In: *Magnetic Resonance in Medicine* 42.5 (1999), pp. 952–962. DOI: 10.1002/(sici)1522-2594(199911)42:5<952::aid-mrm16>3.0.co;2-s.

[46] Qing Qu, Xiao Li, and Zhihui Zhu. "Exact and efficient multi-channel sparse blind deconvolution — a nonconvex approach". In: *2019 53rd Asilomar Conference on Signals, Systems, and Computers* (2019). DOI: 10.1109/ieeeconf44664.2019.9049053.

[47] Saiprasad Ravishankar and Yoram Bresler. "Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging". In: *SIAM Journal on Imaging Sciences* 8.4 (2015), pp. 2519–2557. DOI: 10.1137/141002293.

[48] Saiprasad Ravishankar and Yoram Bresler. "Learning sparsifying transforms for Image Processing". In: *2012 19th IEEE International Conference on Image Processing* (2012). DOI: 10.1109/icip.2012.6466951.

[49] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization". In: *SIAM Review* 52.3 (2010), pp. 471–501. DOI: 10.1137/070697835.

[50] Peter J. Shin et al. “Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion”. In: *Magnetic Resonance in Medicine* 72.4 (2014), pp. 959–970. DOI: 10.1002/mrm.24997.

[51] Gongguo Tang and Benjamin Recht. “Convex blind deconvolution with random masks”. In: *Classical Optics 2014* (2014). DOI: 10.1364/cosi.2014.cw4c.1.

[52] Joshua D. Trzasko and Armando Manduca. “Calibrationless parallel MRI using clear”. In: *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)* (2011). DOI: 10.1109/acssc.2011.6189958.

[53] Martin Uecker et al. “Image reconstruction by regularized nonlinear inversion—joint estimation of coil sensitivities and image content”. In: *Magnetic Resonance in Medicine* 60.3 (2008), pp. 674–682. DOI: 10.1002/mrm.21691.

[54] Liming Wang and Yuejie Chi. “Blind deconvolution from multiple sparse inputs”. In: *IEEE Signal Processing Letters* 23.10 (2016), pp. 1384–1388. DOI: 10.1109/lsp.2016.2599104.

[55] Yilun Wang et al. “A new alternating minimization algorithm for total variation image reconstruction”. In: *SIAM Journal on Imaging Sciences* 1.3 (2008), pp. 248–272. DOI: 10.1137/080724265.

[56] Leslie Ying and Jinhua Sheng. “Joint Image Reconstruction and sensitivity estimation in sense (JSENSE)”. In: *Magnetic Resonance in Medicine* 57.6 (2007), pp. 1196–1202. DOI: 10.1002/mrm.21245.