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 acquired via compressive sensing by multiple sensors in parallel, but with unknown calibration parameters of the sensors. This inverse problem has an important application in pMRI reconstruction, where the calibration parameters of the receiver coils are often difficult and costly to obtain explicitly, but nonetheless are a fundamental requirement for high-precision reconstructions. Most auto-calibration strategies for this problem involve solving a challenging biconvex optimization problem, which lacks reconstruction guarantees. In this work, we transform the auto-calibrated parallel compressive sensing problem to 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. Our method provides robust and stable recovery guarantees that take into account the presence of noise and sparsity deficiencies in the signals. As such, it offers a theoretically guaranteed approach to auto-calibrated parallel imaging in MRI under appropriate assumptions. Applications in compressive sensing pMRI are discussed, and numerical experiments using real and simulated MRI data are presented to support our theoretical results.

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

1 Introduction

We frequently encounter challenges with imperfect sensing, where accurately calibrated sensors are critical for high-precision measurements and 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 both 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

---

*Department of Mathematics, University of California Davis, Davis, CA 95616-5270, USA yuani@ucdavis.edu

†Department of Mathematics and Center of Data Science and Artificial Intelligence Research, University of California Davis, Davis, CA 95616-5270, USA strohmer@math.ucdavis.edu

sensing in solving various auto-calibration tasks [1, 2, 3, 4]. A common scheme in such approaches includes first transforming the biconvex constraints to convex constraints using the concept of 'lifting' [2]; 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 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 belonging to a subset $\Omega \subset [N]$, $\mathbf{s}_i \in \mathbb{C}^{N \times 1}$ is the vector of unknown calibration parameters, $\mathbf{x} \in \mathbb{C}^{N \times 1}$ is 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 the parameters related to the sensor profiles (calibration parameters) $\mathbf{h}_i$.

Such a forward model appears, for instance, in blind parallel MRI (pMRI). Specifically, the i-th coil measurement $\mathbf{y}_i$ is a noisy sample of the spatial Fourier transform of an object $\mathbf{x}$, weighted by an unknown coil sensitivity $\mathbf{s}_i$ that corresponds to the i-th sensing coil. In parallel imaging, C coils are used to take measurements simultaneously for the same underlying object $\mathbf{x}$. The joint estimation of coil sensitivities and the object of interest leads to the nonlinear inverse problem of the form (1).

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 $\mathbf{x} \odot 1/\mathbf{v}$ are 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 number of 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 a low-order polynomial basis or a low-frequency sinusoidal basis. Such choices promote smoothness in the sensing parameters. 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 transformations.

Although the above model may appear straightforward and specific, providing a rigorous analysis is nontrivial. The primary challenge arises from carefully analyzing the incoherence of the forward matrix, which involves the interactions between $\mathbf{F}_\Omega$, $\mathbf{B}$, and $\boldsymbol{\Psi}$. Our goal is to understand the requirements on the number of samples L and the number of channels C for robust and stable recovery of all $\mathbf{s}_i$ and $\mathbf{x}$. We will discuss particular choices of $\mathbf{F}_\Omega$, $\mathbf{B}$, and $\boldsymbol{\Psi}$ that align with the pMRI application in detail in Section 2.

An important observation is that Parallel Imaging (PI) studies can sometimes appear "noisy". This is largely because the primary objective of PI techniques is to reduce imaging time, which leads to acquiring fewer data points per coil due to the need to reduce imaging time. Consequently, this reduction results in a lower signal-to-noise ratio (SNR). As a result, the SNR of PI sequences is typically lower than that of comparable

---

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

non-PI sequences [5, 6].

The SNR of PI sequences is affected by the same factors that influence non-PI sequences, such as field strength, magnet hardware, coil loading, the type of tissue being imaged, pulse sequence, timing parameters, voxel volume, the number of phase-encoding steps, receiver bandwidth, and the number of signals averaged. Additionally, the SNR for PI sequences ($\text{SNR}_{\text{parallel}}$) is further reduced by two extra factors, as shown in the equation below:

$$\text{SNR}_{\text{parallel}} = \text{SNR}_{\text{non-PI}}/(\sqrt{R} \cdot g),$$

where R denotes the reduction or acceleration factor for undersampling, and g represents a spatially dependent term known as the geometry factor. The term g relates to the number, size, and orientation of the surface coil elements and can be considered a measure of coil separation [5, 6].

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. We denote the circular convolution as $*$ and the Kronecker product as $\otimes$. For any integer N, we denote the set $\{1, \cdots, N\}$ as $[N]$. 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 $\mathbf{x}_S(i) = 0, \text{for } i \in S^c$). 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 $\mathbf{X}_T(i, j) = 0, \text{otherwise}$. 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. For any set S, we denote S^c as its complementary set and $|S| = \text{Card}(S)$. Informally, we refer to a matrix $X \in \mathbb{C}^{N_1 \times N_2}$ being incoherent if the 2-norm of its rows does not vary too much.

The paper is organized as follows: In Section 2, we briefly discuss our contributions in relation to the state of the art in auto-calibration and compressive sensing pMRI. 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

The work is inspired by papers related to bringing auto-calibration problems into the framework of biconvex compressive sensing. Many auto-calibration problems can be seen as jointly recovering the signal and the calibration parameters from bilinear measurements. Specifically, they can be seen as variations of recovering two unknown signals $\mathbf{x}$ and $\mathbf{z}$ from bilinear measurements,

$$y_i = \langle \mathbf{x}, \mathbf{b_i} \rangle \langle \mathbf{z}, \bar{\mathbf{a}}_i \rangle, i = 1 \cdots m, \tag{3}$$

where $\mathbf{x}, \mathbf{b}_i \in \mathbb{C}^{N_1 \times 1}, \mathbf{z}, \mathbf{a}_i \in \mathbb{C}^{N_2 \times 1}$.

Various auto-calibration problems of similar setups have been explored, including blind deconvolution in [3], direction-of-arrival estimation as seen in [1], blind deconvolution and demixing [4], phase retrieval via matrix completion [2], among others. The analysis of the auto-calibration problems generally differs in the following two aspects:

- Additional assumptions must be made about the signal, the calibration parameters, or both to address the ill-posed nature of the problem. Popular assumptions may vary in several ways, either regarding the signal, the sensing parameters, or both: whether they assume sparsity individually or jointly, sparsity in the natural parameter space or after transformations, subspace constraints, or restrictions imposed by sign constraints.

3

- Specific application cases impose different forward models due to physical or other constraints. For example, the choice of sensing matrices - whether sub-Gaussian (e.g., b_i or a_i are Gaussian), structured random (e.g., taking random samples from the Fourier transform) or deterministic.

Two primary approaches to solving the bilinear optimization problem are as follows: one involves convexifying it through methods such as linearization or lifting [3, 1, 7, 4, 2]. The other involves optimization in the natural parameter space using alternating minimization and gradient descent algorithms [8, 9, 10].

For a general discussion on the issue of injectivity and the principle of identifiability for bi-linear compressive sensing, one may refer to [11, 12].

2.1 Related work in auto-calibration and compressive sensing

It seems natural and tempting to recover $(\mathbf{h}_1, \cdots, \mathbf{h}_C, \mathbf{z})$ obeying (2) by solving the optimization problem

$$\min_{\mathbf{h}_1, \cdots, \mathbf{h}_C, \mathbf{z}} \sum_i \|\mathbf{y}_i - \mathbf{F}_\Omega(\mathbf{B}\mathbf{h}_i \odot \mathbf{\Psi}^* \mathbf{z})\|_2^2. \tag{4}$$

However, the above nonlinear least squares is a challenging optimization problem since it is highly nonconvex and most of the available algorithms, such as alternating minimization and gradient descent, may suffer from getting easily trapped in some local minima.

A popular convex optimization approach involves transforming the bi-convex problem (3) to the convex problem using the idea of 'lifting'. Define the 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,

$$\langle \mathbf{x}, \mathbf{b_i} \rangle \langle \mathbf{z}, \bar{\mathbf{a}}_\mathbf{i} \rangle = \mathbf{b}_i^* \mathbf{x} \mathbf{z}^T \mathbf{a}_i = \mathbf{b}_i^* \mathbf{X} \mathbf{a}_i = (\mathbf{a}_i^T \otimes \mathbf{b}_i^*) \operatorname{vec}(\mathbf{X}). \tag{5}$$

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

$$[y_1, \cdots, y_m] = \mathcal{A}(\mathbf{X}) := \{\mathbf{b}_i^* \mathbf{X} \mathbf{a}_i\}_{i=1}^m. \tag{6}$$

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

An important auto-calibration problem pertaining to sparse signals stems from the task of image reconstruction using randomly coded masks [13, 14]. The forward model for this auto-calibration problem is given by:

$$\mathbf{y} = \operatorname{diag}(\mathbf{B}\mathbf{h})\mathbf{A}\mathbf{x} + \boldsymbol{\omega}, \tag{7}$$

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 matrix $\mathbf{A}$ acts as the sensing matrix. The differences between the forward model (7) and the current work (1) are as follows: 1. The blind sensing (i.e., $\operatorname{diag}(\mathbf{B}\mathbf{h})$) is applied after compressed sensed signals (i.e., $\mathbf{A}\mathbf{x}$). 2. The signal is sparse in its natural domain rather than being transform-sparse.

The SparseLift framework [1] solves the auto-calibration problem in (7) by combining 'lifting' and bi-convex compressive sensing. The authors in [1] first transform the biconvex constraint (7) 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 discrete Fourier transformation (DFT) matrix, when the number of measurements L is of the order of $\mathcal{O}(kn \log^2(L))$. Note that the case where the signal is only approximately sparse is not analyzed. In a related work, the author in [4] extends the results of SparseLift to the case where one observes a summation of such measurements from

multiple sparse signals, $\mathbf{y} = \sum\limits_{i \in [r]} \mathrm{diag}(\mathbf{Bh}_i)\mathbf{Ax}_i + \boldsymbol{\omega}$. This extension improves recovery guarantees by solving a $\ell_{1,2}$ minimization by promoting block sparsity in the lifted signals.

The idea of 'lifting' also applies to another class of self-calibration problems, notably the blind deconvolution problem [3, 15]. In the general form of blind deconvolution, one measures the convolution of a signal with an unknown filter. That is, $\mathbf{y} = I_\Omega(\mathbf{s} * \mathbf{x}) + \mathbf{n}$ where both $\mathbf{s}$ and $\mathbf{x}$ are unknown. Notice that the forward model for our problem in (2) can be written in the form of a convolution $\mathbf{y} = \mathbf{I}_\Omega(\mathbf{FBh} * \mathbf{F\Psi^*z}) + \mathbf{n}$ using the convolution theorem. When the Fourier space is fully sampled (i.e. $\mathbf{I}_\Omega = \mathbf{I}$), the problem is equivalent to recovery from

$$\mathbf{F}^{-1}\mathbf{y} = \mathbf{Bh} * \mathbf{\Psi^*z} + \mathbf{w}, \tag{8}$$

where $\mathbf{w}$ is also Gaussian. The work by Ahmed et al. [3] provides a theoretically guaranteed solution to the fully sampled case in (8). Using lifting, the authors solve a nuclear norm minimization problem by exploiting low-rankness in the lifted signal. If the active coefficients of $\mathbf{z}$ are known and $\|\mathbf{z}\|_0 = n$, the main theorem of [3] guarantees a stable recovery when the number of measurements is, up to log factor, of the order $\mathcal{O}(n+k)$ when $\mathbf{\Psi}$ is a random Gaussian matrix and $\mathbf{B}$ is incoherent. For the more difficult case when the convolution is only subsampled, the authors in [16] provide rigorous analysis when both $\mathbf{s}$ and $\mathbf{x}$ are transform sparse (i.e., there exists $\mathbf{\Phi}$ and $\mathbf{\Theta}$ such that $\mathbf{\Phi x}$ and $\mathbf{\Theta s}$ are sparse). It provides an iterative algorithm for robust recovery guarantees with near optimal sample complexity. However, the recovery can only be guaranteed when both the sparsifying transformations (i.e. $\mathbf{\Phi}$ and $\mathbf{\Theta}$) are random Gaussians. As seen in many theories, subgaussian random matrices provide optimal measurement matrices for compressive sensing [17]. However, note that for our application case, $\mathbf{\Phi} = \mathbf{\Psi F}^{-1}$ and $\mathbf{\Theta} = \mathbf{B^*F}^{-1}$, both impose structures due to physical constraints.

Similar to the parallel imaging (i.e., $C > 1$ in (2)), from a practical viewpoint, a relevant scenario focuses on self-calibration from multiple snapshots. For example, blind image restoration from multiple filters [18] and self-calibration model for sensors [19, 7, 20, 21]. The goal here is to recover the unknown gains/phases $\mathbf{S} = \mathrm{diag}(\mathbf{s})$ 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

$$\mathbf{Y} = \mathbf{SAX}. \tag{9}$$

For this model, the sensing matrix $\mathbf{A}$ is fixed through the sensing process, just as in the pMRI case. 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 convexify the biconvex measurements (9) into a linear one. One crucial assumption for the 'linearization' scheme is that $\mathbf{s}$ does not have zero entries. Define $\mathbf{s} = \mathbf{1/d}$, the equations $\mathbf{y}_l = \mathrm{diag}(\mathbf{s})\mathbf{Ax_l}$ can be written as

$$\mathbf{y}_l\mathbf{d} = \mathbf{Ax_l}, \tag{10}$$

which is a linear equation in the concatenated vector $\begin{bmatrix} \mathbf{d} \\ \mathbf{x}_l \end{bmatrix}$. Compared to 'lifting', an advantage of 'linearization' is that the underlying problem dimension remains unchanged. The same idea of linearization can also be found in [20, 21, 19, 22]. In [7], the authors use 'linearization' and solve a least-square minimization problem for (9). Furthermore, theoretical recovery guarantee is provided when $\mathbf{A}$ is Gaussian.

As observant readers already may have noticed, the 'true dimension' of the problem of recovering $(\mathbf{h}_1, \cdots, \mathbf{h}_c, \mathbf{z})$ is $(kC+n)$. To the best of the authors' knowledge, currently there are no convex optimization schemes which succeed with such few measurements. In this context, the work by Oymak et al. [23] should be mentioned. The authors demonstrated nonconvex procedures (e.g., minimizing a linear combination of ℓ_0, $\ell_{0,2}$ or rank of the lifted matrix) that succeed with high probability, when the number of measurements is (up to log-factors) slightly larger than the information-theoretical limits. However, the theories apply only to measurement ensembles with specific properties, such as matrices with sub-Gaussian rows, subsampled standard basis (e.g.

in matrix completion), or quadratic measurements for phase retrieval. Also, these procedures are not efficiently computable. For other structured measurement ensembles, it is worth noting the works by Qu et al. and Ling et al. [10, 9]. These studies address certain blind deconvolution problems with efficient algorithms and theoretical guarantees when the number of measurements is (up to log-factors) slightly larger than the information-theoretical limits. However, to the best of our knowledge, none of these methods directly apply to our measurement ensemble or require additional assumptions that do not align with our application case.

2.2 Developments in compressive sensing methods in pMRI

In the literature on pMRI, combining compressive sensing and parallel imaging has gained significant importance due to the need for fast acquisition. Accelerated imaging with fewer samples using multiple receiver coils has been an active field of research since its beginning. In this section, we briefly discuss the development of compressive sensing approaches in pMRI that exploit transform sparsity in the time domain, with a focus on those most closely related to our approach.

2.2.1 Relevant problem formulations and the optimization problems

In parallel imaging, the data are acquired from multiple channels simultaneously, but each at a rate lower than the Nyquist rate such that the data acquisition time is reduced. If only $1/R$ of the Fourier space is sampled per channel, the acquisition time is reduced by a factor of R.

The measurement model for pMRI using C coils/channels is the following,

$$\mathbf{E}\mathbf{x} = \mathbf{y}, \tag{11}$$

where the encoding matrix $\mathbf{E}$ consists of the product of partial Fourier encoding with channel-specific sensitivity encoding over the image. Specifically, $\mathbf{E} = \begin{bmatrix} \mathbf{F}_\Omega \mathbf{D}_1 \\ \vdots \\ \mathbf{F}_\Omega \mathbf{D}_C \end{bmatrix}$ and $\mathbf{D}_i = \mathrm{diag}(\mathbf{s}_i)$ where $\mathbf{s}_i$ is the coil sensitivity for the i-th coil.

According to linear algebra and the generalized sampling theorem, the maximum reduction factor is equal to the number of channels if $\mathbf{E}$ is of full rank. Because these conditions are rarely guaranteed in practice, the possible reduction factor is much lower. In practice, the reduction factor is typically chosen between 2 and 6 [24, 25] for parallel imaging when the number of coils is large. Thus, at the boundary case, one would expect a threshold for the maximum reduction factor and the minimum number of channels necessary to ensure the invertibility of such a linear problem.

In addition to reducing sampling time through sensitivity encoding, further reduction in the number of samples can be achieved by exploiting redundancy in the signals. In this work, we specifically discuss time-domain sparsity, as magnetic resonance (MR) images typically exhibit sparse representation in some appropriate transform domain. Suppose it is assumed that the signal is n-sparse after some known transformation. One approach is to solve the following ℓ_0 norm constrained least squares problem to directly exploit the n-sparsity of the signal:

$$\min_{\mathbf{x}} \|\mathbf{F}_\Omega \mathbf{x} - \mathbf{y}\|_2^2 \quad \text{subject to } \|\mathbf{\Psi}\mathbf{x}\|_0 = n. \tag{12}$$

As a convex relaxation to ℓ_0 minimization, the following ℓ_1 minimization is often utilized for practical compressive sensing MRI applications. Since Lustig et al. [26] 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 original ℓ_1 norm minimization problem solves the following convex program,

$$\min_{\mathbf{x}} \|\mathbf{\Psi}\mathbf{x}\|_1 \quad \text{subject to } \mathbf{F}_\Omega \mathbf{x} = \mathbf{y}, \tag{13}$$

where the signal $\mathbf{x}$ is assumed to be transform sparse. In the MRI literature, it has been demonstrated that the discrete cosine transformation, the wavelet transform, and the finite difference transform are the most commonly utilized choices for sparsifying MR images [27, 28, 29]. The choice of such $\mathbf{\Psi}$ is related to the field of research in sparsifying transformations [30, 31]. More recently, attention has turned to learning the sparsifying transformations from data [32, 33].

In terms of combining CS with parallel imaging, the first applications were demonstrated in [34, 35, 36, 37]. They can be seen as a direct extension of (13) to the parallel case,

$$\min_{\mathbf{x}} \|\mathbf{\Psi x}\|_1 + \alpha \|\mathbf{x}\|_{TV} \quad \text{subject to } \mathbf{Ex} = \mathbf{y}, \tag{14}$$

where $\mathbf{D}_i = \text{diag}(\mathbf{s}_i)$ is the diagonal matrix of known calibration parameters of the i-th coil.

Note that both the convex relaxations in (13) (14) and nonconvex problem in (12) assume that the coil sensitivities are known. In practice, coil sensitivities are typically determined either through a separate pre-scan [38, 39, 40] or by interpolating from a fully sampled region of k-space (Fourier space) [41, 42, 43]. However, both methods have inherent drawbacks. The pre-scan approach is susceptible to motion artifacts, environmental fluctuations, and requires additional acquisition time. The interpolation method, while faster, may introduce errors due to imperfect k-space sampling or interpolation techniques. Moreover, both approaches may necessitate the use of filtering and kernel selection to obtain reasonable estimates of the coil sensitivities, which can lead to variations in the resulting reconstructions.

Given these challenges, auto-calibrated reconstruction methods have gained attention. In particular, it considers recovering all $\mathbf{s}_i$ and $\mathbf{x}$ from the measurement model in (11). However, the linear system is highly underdetermined and ill-posed in this case. Indeed, if $\mathbf{s}_i$ and $\mathbf{x}$ form a pair of solutions to (11), then for any complex vector $\mathbf{v}$ with nonzero entries, $\mathbf{s}_i \odot \mathbf{v}$ and $\mathbf{x} \odot 1/\mathbf{v}$ form a pair of solutions.

Prior knowledge about the object and the coil sensitivities must be incorporated to render reasonable solutions. As formulated in Joint Image Reconstruction and Sensitivity Estimation in SENSE (JSENSE) [44] and Regularized Nonlinear Inversion (NLINV) [45], the smoothness in the coil sensitivities is exploited. In JSENSE, the coil sensitivities are assumed to be expanded by a low-order polynomial basis (i.e. $\mathbf{B}$) and the following bi-convex optimization problem is solved via alternating minimization,

$$\arg \min_{\mathbf{x}, \mathbf{h_j}} \Sigma_{j=1}^C \|\mathbf{y} - \mathcal{F}_\Omega \{\mathbf{Bh}_j \odot \mathbf{x}\}\|_2^2. \tag{15}$$

In the NLINV formulation, they solve the following bi-convex optimization problem iteratively using a Gauss Newton type method,

$$\arg \min_{\mathbf{x}, \mathbf{s_j}} \Sigma_{j=1}^C \|\mathbf{y} - \mathcal{F}_\Omega \{\mathbf{s}_j \odot \mathbf{x}\}\|_2^2 + \alpha(\Sigma_{j=1}^C \|\mathbf{W s}_j\|_2^2 + \|\mathbf{x}\|_2^2), \tag{16}$$

where $\mathbf{W} \in \mathbb{C}^{N \times N}$ is an invertible preconditioning matrix that penalizes high frequencies, thereby enforcing smoothness in the coil sensitivities. Other methods to enforce smoothness in the coil include expanding the coil sensitivities using low-order polynomial or sinusoidal basis functions, a method first introduced in the original SENSE paper [38] and later adopted in the analytical phantom package [46].

An extension of the NLINV (16) framework is the ENLIVE [47], which solves the following relaxed joint nonconvex optimization problem over k sets of measurements,

$$\arg \min_{\mathbf{x^i}, \mathbf{s_j^i}} \Sigma_{j=1}^C \|\mathbf{y}_i - \mathcal{F}_\Omega \{\Sigma_{i=1}^k \mathbf{s}_j^i \odot \mathbf{x}^i\}\|_2^2 + \alpha \Sigma_{i=1}^k (\Sigma_{j=1}^C \|\mathbf{W s}_j^i\|_2^2 + \|\mathbf{x}^i\|_2^2). \tag{17}$$

Here, k sets of images $\mathbf{x}^i$ are used instead of one single image $\mathbf{x}$ as in (16) to account for model violations. In a post-processing step, an average image using the k sets of solutions is calculated as the solution.

Of particular interest to us is that ENLIVE can be related to a nuclear norm minimization problem that exploits the low-rankness in a 'lifted' domain. Specifically, by defining the matrix $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 & \cdots & \mathbf{y}_C \end{bmatrix}$, $\mathbf{U} = \begin{bmatrix} \mathbf{x}^1 & \cdots \mathbf{x}^k \end{bmatrix} \in \mathbb{C}^{N \times k}$, $\mathbf{V} = \begin{bmatrix} \mathbf{W}\mathbf{s}_1^1 & \cdots \mathbf{W}\mathbf{s}_1^k \\ \vdots & \vdots \\ \mathbf{W}\mathbf{s}_C^1 & \cdots \mathbf{W}\mathbf{s}_C^k \end{bmatrix} \in \mathbb{C}^{CN \times k}$, the bi-linear measurements can be transformed into linear measurements in terms of a rank-k matrix $\mathbf{U}\mathbf{V}^T$:

$$\arg \min_{\mathbf{U},\mathbf{V}} \|\mathbf{y} - \mathcal{A}(\mathbf{U}\mathbf{V}^T)\|_2^2 + \alpha(\|\mathbf{V}\|_F^2 + \|\mathbf{U}\|_F^2), \tag{18}$$

with a linear operator $\mathcal{A}$ that maps $\mathbf{U}\mathbf{V}^T$ to $\mathcal{F}_\Omega\{\Sigma_{i=1}^k \mathbf{s}_j^i \odot \mathbf{x}^i\}$. As shown in [47], the above formulation can be further relaxed to a nuclear norm minimization problem in terms of a matrix $\mathbf{Z}$,

$$\arg \min_{\mathbf{Z}} \|\mathbf{y} - \mathcal{A}(\mathbf{Z})\|_2^2 + 2\alpha\|\mathbf{Z}\|_*. \tag{19}$$

The authors claim that this connection provides an explanation for why ENLIVE favors low-rank solutions. However, it is worth noting that due to the computational cost associated with the nuclear norm minimization, the convex formulation (19) is not used for practical computations. Instead, the non-convex formulation (17) is employed to solve the auto-calibrated reconstruction problem. Notably, neither of the formulations in (15) nor (16) explicitly exploits sparsity in the object being imaged.

2.2.2 The relationship between L, C and n

The primary objective of parallel imaging (PI) techniques is to reduce imaging time, which is achieved by acquiring fewer data points per coil and taking multiple coil measurements at the same time. From compressive sensing theory [48, 49], it is well known that the incoherence in the sensing matrix is crucial for the successful solution of the problems in (13) and (14). And it is well accepted in the literature on pMRI that increasing the number of channels C improves the incoherence of the sensing matrix $\mathbf{E}$.

However, to understand the role of C and L versus the sparsity level n, a thorough analysis of the sensitivity encoding, partial Fourier sampling, and the sparsifying transformation in the overall measurement ensemble is required. As with linear inversion for solving (11), one would expect a threshold for the maximum reduction factor and a minimum number of coils for the compressive sensing approaches (e.g. ℓ_0, ℓ_1 or mixed norm minimization) to be successful. The problem becomes even more challenging, when the sensitivity encoding is unknown as in the auto-calibration framework.

In this context, the works by Otazo et al. [50, 51] should be mentioned, where they numerically assess the sufficient conditions on the number of measurements per coil L and the number of coils C for solving the ℓ_0 norm minimization problem in (12) or the ℓ_1 norm minimization problem in (13) to recover an n-sparse signal. The authors conclude that in the known calibration setup, when the number of coils is large enough, the required number of measurements per coil is very close to the signal sparsity for both the noiseless and noisy cases. To illustrate the numerical findings, we reproduced the reconstruction results from [50] by solving the ℓ_0 minimization problem (12), using the orthogonal matching pursuit (OMP) algorithm. The results can be found in Figure 1.

At first glance, it may seem surprising that when C is sufficiently large, increasing the number of coils does not yield any additional benefit in reducing the number of measurements per coil or the reconstruction quality. The following work may give us theoretical insights into the observation. Consider a more general setup in solving the following distributed compressed sensing problem:

$$\mathbf{y}_i = \mathbf{F}_{\Omega_i}(\mathbf{s}_i \odot \mathbf{x}), i = 1 \cdots C, |\Omega_i| = L, \tag{20}$$

where $\mathbf{s}_i$ are known and $\mathbf{x}$ is n-sparse.
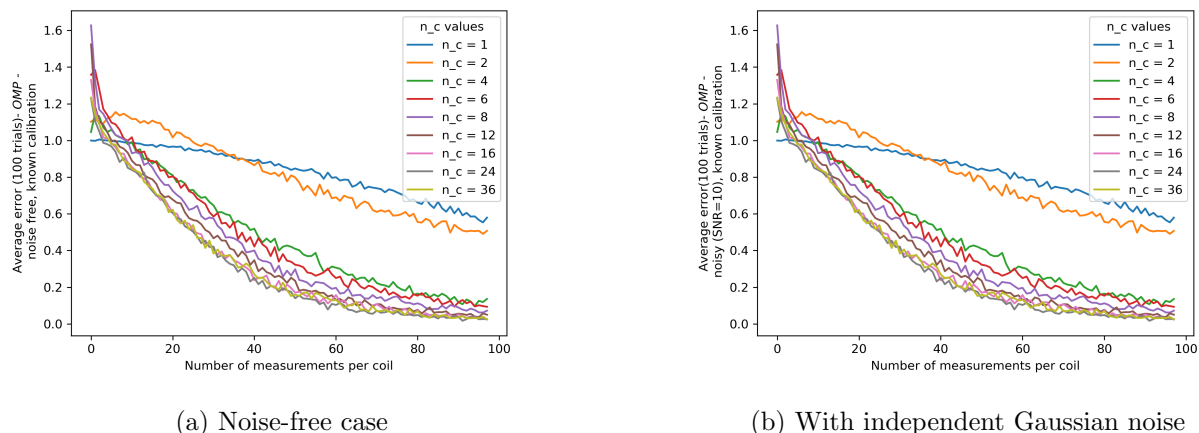
(a) Noise-free case

(b) With independent Gaussian noise

Figure 1: Two-dimensional pMRI simulations were performed, assuming truly transform n-sparse signals with randomly distributed non-zero components (N = 529, n = 32). Coils with sensitivities calculated using the Biot-Savart law were simulated for various numbers of coil elements ($n_c = 2, 4, 6, 8, 12, 16, 24, 36$). In the case of a single coil, a constant sensitivity was assumed. The multi-coil signal was generated by multiplying the signal with the coil sensitivities, with optional addition of Gaussian noise. The signal is assumed to be sparse after 2-dimensional DCT transforms. The average of the relative ℓ_2 error was computed over 100 different random undersampling realizations in Fourier-space and random realizations of transform sparse signals, using varying numbers of measurements and coil elements.

Note that a key difference between the above problem and the pMRI forward model is that the sensors use different random undersampling patterns. The authors in [52] have theoretically shown that, in such a distributed compressed sensing model with a large number of sensors, the required number of samples per sensor approaches the sparsity level, as solved by a greedy OMP-type algorithm. Recall that compressed sensing with ℓ_1-norm minimization (i.e. solving equation (13) with a single coil) typically requires L to be three to five times the number of sparse coefficients n (i.e., 3–5n where $\mathbf{x}$ is n-sparse) to achieve accurate reconstructions [26, 53]. Hence, a large number of sensors provides a constant improvement in the number of measurements per sensor, up to the sparsity level of the signal $\mathbf{x}$.

Note that this condition does not apply to multi-coil acquisition in pMRI, since all coils share the same undersampling pattern, albeit with distinct coil sensitivity profiles. Hence, one might expect a reduction in incoherence and a deviation from the theoretical recovery bounds discussed in [52]. Additionally, the measurement ensemble in the pMRI setting must adhere to additional physical constraints, which makes the analysis within the compressive sensing framework more complex. By combining the numerical experiments with the aforementioned theory for the known calibration case, one could reasonably expect that a sufficient condition for robust reconstruction would adopt a similar form to that of the best-case scenario for the more complex auto-calibration problem in (1).

2.2.3 Further discussions

Essentially, auto-calibration methods depend on data redundancy. In the literature on pMRI, other sources of redundancy have been utilized for auto-calibration, each under different assumptions. Data driven auto-calibration methods, such as GRAPPA [41] and SPIRiT [54] can be viewed as interpolation methods by estimating linear relationships within the fully sampled k-space (e.g., kernel calibration) and enforcing that relationship to synthesize data values in place of unacquired data (e.g., data reconstruction). More recently, a data-driven method SAKE [27] 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 [44, 45]. It is worth noting that the effectiveness of these methods depends on whether the coil and signal adhere to the assumptions made. As a result, they may produce varying quality of reconstructions for different datasets.

2.3   Our contributions

While the aforementioned auto-calibration methods in Section 2.2.1 undoubtedly offer utility and applicability by giving successful numerical results, the majority of them do not provide rigorous analysis or theoretical guarantees of recovery. Note that even in the known calibration case (14), the incoherence of the sensing matrix has not been properly analyzed for the popular ℓ_1 minimization scheme, particularly for measurement ensembles that are applicable to pMRI. In this work, we take a step forward to provide analysis for the more complicated auto-calibration problem in (1) and to provide sufficient conditions on L and C for stable and robust recovery of an approximate transform sparse signal $\mathbf{x}$.

We propose to solve the auto-calibration problem in (2) using convex optimization by leveraging the block sparsity structure across all channels simultaneously. Our approach efficiently solves a mixed norm minimization problem and provides robust and stable recovery with theoretical guarantees. Interestingly, our theoretical findings align with previous numerical observations from past literature, providing valuable explanatory insights. To provide additional evidence, we present numerical results using the proposed method to solve the auto-calibrated pMRI. We include simulations using the analytical phantom and simulated coil sensitivities, as well as reconstructions from real pMRI measurements.

3   Problem setup

We are concerned with the following auto-calibration problem:

$$\mathbf{y}_i = \mathbf{F}_\Omega \mathbf{D}_i \mathbf{x} + \mathbf{n}_i, i = 1 \cdots C, \tag{21}$$

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 Fourier matrix with its rows chosen from the N-point 2-dimensional discrete Fourier transformation (DFT) matrix which belongs to a subset $\Omega \subset [N]$, $\mathbf{D}_i = \mathrm{diag}(\mathbf{s}_i) \in \mathbb{C}^{N \times N}$ is the diagonal matrix of calibration parameters corresponding to the i-th channel/coil, $\mathbf{x} \in \mathbb{C}^{N \times 1}$ is the signal of interest and $\mathbf{n}_i \in \mathbb{C}^{N \times 1}$ is additive noise.

Our goal is to recover the common signal of interest $\mathbf{x}$ from the measurements $\mathbf{y}_i$ across all channels, while the calibration matrices $\mathbf{D}_i$ are unknown. As discussed in the previous Sections, such a task is impossible unless additional assumptions are made on the signal and/or the calibration parameters.

Hence, we adopt the following two assumptions on both the signal $\mathbf{x}$ and the calibration parameters $\mathbf{D}_i$, as justified in the previous discussions. 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 $\mathbf{z} = \mathbf{\Psi}\mathbf{x}$ 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 $\mathbf{D}_i = \mathrm{diag}(\mathbf{B}\mathbf{h}_i)$, 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

$$\mathbf{y}_i = \mathbf{F}_\Omega \mathrm{diag}(\mathbf{B}\mathbf{h}_i)\mathbf{\Psi}^*\mathbf{z}, i = 1 \cdots C. \tag{22}$$

10

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 [2, 3, 1], we can transform the biconvex constraint (22) into the problem of recovering a block-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:

$$\mathbf{y}_i = \mathbf{F}_\Omega \begin{bmatrix} \mathbf{b}_1^* \mathbf{h}_i \mathbf{z}^T \mathbf{q}_1 \\ \mathbf{b}_2^* \mathbf{h}_i \mathbf{z}^T \mathbf{q}_2 \\ \vdots \\ \mathbf{b}_N^* \mathbf{h}_i \mathbf{z}^T \mathbf{q}_N \end{bmatrix} = \mathbf{F}_\Omega \underbrace{\begin{bmatrix} (\bar{\mathbf{q}}_1 \otimes \mathbf{b}_1)^* \\ \vdots \\ \bar{\mathbf{q}}_N \otimes \mathbf{b}_N)^* \end{bmatrix}}_{\mathbf{\Phi}} \mathrm{vec}(\mathbf{h}_i\mathbf{z}^T) := \mathbf{F}_\Omega \mathbf{\Phi} \mathrm{vec}(\mathbf{h}_i\mathbf{z}^T), i = 1\cdots C, \qquad (23)$$

where

$$\mathbf{\Phi} := \begin{bmatrix} (\bar{\mathbf{q}}_1 \otimes \mathbf{b}_1)^* \\ \vdots \\ \bar{\mathbf{q}}_N \otimes \mathbf{b}_N)^* \end{bmatrix} \in \mathbb{C}^{N\times kN} \qquad (24)$$

.

Defining

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 | \mathbf{y}_2 | \cdots | \mathbf{y}_C \end{bmatrix},$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 | \mathbf{h}_2 | \cdots | \mathbf{h}_C \end{bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 | \mathbf{X}_2 | \cdots | \mathbf{X}_C \end{bmatrix} = \begin{bmatrix} \mathrm{vec}(\mathbf{h}_1\mathbf{z}^T) | \mathrm{vec}(\mathbf{h}_2\mathbf{z}^T) | \cdots | \mathrm{vec}(\mathbf{h}_C\mathbf{z}^T) \end{bmatrix} = \mathbf{z} \otimes \mathbf{H} \in \mathbf{C}^{kN\times C},$$

we can rewrite (23) in matrix form as follows:

$$\mathbf{Y} = \underbrace{\mathbf{F}_\Omega \mathbf{\Phi}}_{\mathbf{A}} \mathbf{X}, \qquad (25)$$

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 or measurement ensemble. Now the problem is to find the matrix $\mathbf{X}$ satisfying the linear constraint described by (25). To that end we will exploit the block sparsity of $\mathbf{X}$.

3.2 Block sparsity of X

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

Moreover, $\mathbf{X}$ is at most rank-k if C is large. Methods that exploit the low-rankness include [55, 56, 47] and also the the ENLIVE formulation (17) introduced in Section 2.2. Naturally, one could minimize over a linear combination $\|\tilde{\mathbf{X}}\|_{1,2} + \lambda\|\tilde{\mathbf{X}}\|_*, \lambda > 0$ in an attempt to simultaneously capture the low-rankness and sparse characteristics of $\tilde{\mathbf{X}}$. However, as demonstrated in [23], 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

$$\|\mathbf{X}\|_{1,2} = \sum_{j \in [N]} \|\mathbf{X}_{T_j \times [C]}\|_F, \tag{26}$$

which is the sum of the Frobenius 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 $\mathbf{X}$ [57, 58, 4], we solve the following convex program

$$\min \|\mathbf{X}\|_{1,2} \quad \text{subject to } \mathbf{Y} = \mathbf{F}_\Omega \mathbf{\Phi} \mathbf{X}. \tag{27}$$

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

$$\min \|\mathbf{X}\|_{1,2} \quad \text{subject to} \|\mathbf{Y} - \mathbf{F}_\Omega \mathbf{\Phi} \mathbf{X}\|_F \leq \sqrt{C}\sigma. \tag{28}$$

We also denote the forward operator as $\mathcal{A} : \mathbb{C}^{k \times N} \rightarrow \mathbb{C}^L$, 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 (27)(28), we calculate the average of the best rank-one estimations of its columns to retrieve the common 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 use the following Lemma from [1] to compute an average of the estimates from $\hat{\mathbf{X}}$ and to ensure that it is close to $\mathbf{z}$.

Lemma 1 *For any solution $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 | \cdots | \mathbf{x}_C \end{bmatrix}$ and the lifted matrix $\mathbf{X}_0 = \begin{bmatrix} vec(\mathbf{h}_1 \mathbf{z}_0^T) | vec(\mathbf{h}_2 \mathbf{z}_0^T) | \cdots | vec(\mathbf{h}_C \mathbf{z}_0^T)) \end{bmatrix}$, let $\hat{\sigma} \hat{\mathbf{u}} \hat{\mathbf{v}}^T$ be the best rank-one Frobenius norm approximation of $\sum_i \mathbf{x}_i / C$ reshaped into a matrix $\hat{\mathbf{X}} \in \mathbb{C}^{k \times N}$ in the column-wise order. If $\|\mathbf{X} - \mathbf{X}_0\|_F = \sqrt{C}\epsilon$, then there exists a scalar α and a constant C_0 such that,*

$$\|\mathbf{z}_0 - \alpha^{-1} \hat{\mathbf{v}}\| \leq C_0 \min(\epsilon / \|\mathbf{z}_0\|, \|\mathbf{z}_0\|).$$

The Lemma is a direct consequence of combining Corollary 3.3 from [1] and using norm equivalency. Also notice if for any $\alpha_0 \neq 0$, $\alpha_0 \mathbf{u}_0$ and $1/\alpha_0 \mathbf{v}_0$ is also a pair of solutions. However, the scalar ambiguity is not important for many applications, including pMRI. The reconstructed signal is retrieved as $\hat{\mathbf{x}} = \mathbf{\Psi}^* \hat{\mathbf{v}}$.

In the aspect of solving the optimization problem (27)(28), the following unconstrained form is often used,

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_{1,2}. \tag{29}$$

A comprehensive review of the algorithms to solve problems of the form (29) is not a focus of this paper. We refer to a few common approaches such as the forward-backward splitting (FBS) [59], ADMM [60] and FISTA [61] which have found numerous applications in the context of pMRI reconstruction. In the simulations part of this paper, we use a FISTA-type iterative method to solve the unconstrained problem. Detailed iterative updates are provided in the Appendix 7.2.

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 $S = \{s_1, s_2, \cdots\} \subset [N]$, 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 from 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 $\text{sgn}(\mathbf{X}) \in \mathbb{C}^{kN \times C}$ as a matrix formed by normalizing each block of $\mathbf{X}$ with length $k \times C$. To be precise,

$$\text{sgn}(\mathbf{X}) = \begin{cases} \frac{\mathbf{X}_{T_j \times [C]}}{\|\mathbf{X}_{T_j \times [C]}\|_F}, \text{if} \|\mathbf{X}_{T_j \times [C]}\|_F \neq 0 \\ \mathbf{0}, \text{otherwise} \end{cases} . \tag{30}$$

It will come in handy in the proofs to decompose a linear map to any given subset of its columns. To that end, 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

$$\mathbf{A}_{S_j}(:,k) = \begin{cases} \mathbf{A}(:,k), \text{if } k \in S_j \\ \mathbf{0}, \text{otherwise} \end{cases} \quad \text{and } \mathbf{A}_S(:,k) = \begin{cases} \mathbf{A}(:,k), \text{if } k \in T \\ \mathbf{0}, \text{otherwise} \end{cases} . \tag{31}$$

4.1   Assumptions and main theorem

Before presenting the main theorem, we first summarize the assumptions on the forward model in (25) and on the signals for easy reference.

(A1). The rows of $\mathbf{B}$ are chosen uniformly at random without replacement from the rows of a matrix $\mathbf{B}_0$ with $\mathbf{B}_0^* \mathbf{B}_0 = \mathbf{I}_k$. (Equivalently, the columns of $\mathbf{B}_0$ form a tight frame.)

(A2). $\boldsymbol{\Psi}$ is a known and fixed orthonormal basis with $\boldsymbol{\Psi}^* \boldsymbol{\Psi} = \boldsymbol{\Psi} \boldsymbol{\Psi}^* = \mathbf{I}_N$.

(A3). The rows of $\mathbf{F}_\Omega$ are chosen uniformly at random without replacement from the DFT matrix $\mathbf{F}$ with $\mathbf{F}^* \mathbf{F} = \frac{N}{L} \mathbf{I}_N$.

(A4). We analyze two probability models/distributions 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 $S_\mathbb{C}^{k-1}$

We briefly introduced some common choices of $\mathbf{B}$ and $\boldsymbol{\Psi}$ in Section 2, including the smoothness promoting polynomial or sinosudal basis for $\mathbf{B}$ and DCT or wavelet transformation for $\boldsymbol{\Psi}$. Here, we further comment on our choice of assumptions (A1)-(A4). (A1) and (A3) are regarding the sensing matrix. (A3) is a standard assumption in compressive sensing theory concerning the random sampling scheme for k-space/Fourier space measurements. For (A1), $\mathbf{B}$ 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 $\mathbf{B}_i^* \mathbf{B}_j = \sigma(\mathbf{B}_i)^* \sigma(\mathbf{B}_j)$. With $\mathbf{B}$ being a matrix with randomly permuted rows, for any fixed $\mathbf{h}$, the coil sensitivity $\mathbf{s} = \mathbf{B}\mathbf{h}$ will be randomly permuted copies of some underlying $\mathbf{s}_0 = \mathbf{B}_0 \mathbf{h}$. On the other hand, if $\mathbf{B}$ were deterministic, any zero entries in $\mathbf{s} = \mathbf{B}\mathbf{h}$ will make it impossible to recover uniquely the values of the image $\mathbf{z}$ at those locations. The randomness in $\mathbf{B}$ helps in providing an average case analysis for each

13

coil. Assumption (A4) on $\mathbf{H}$ is critical for developing an average case analysis across coils to explain the improved results when more channels are included.

Furthermore, we consider the underlying image $\mathbf{z} = \boldsymbol{\Psi}\mathbf{x}$ to be fixed and only impose some probability distribution on $\mathbf{h}_i$. That is to assume 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 $\mathbf{s}_i^* \mathbf{s}_i = (\mathbf{B}\mathbf{h}_i)^*(\mathbf{B}\mathbf{h}_i) = \mathbf{h}_i^* \mathbf{B}^* \mathbf{B}\mathbf{h}_i = \mathbf{h}_i^* \mathbf{h}_i = 1$.

Before moving on to stating the theorem, let us discuss the reason for imposing such a 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 channels, we are not gaining more information on $\mathbf{X}$ regardless of the number of coil used. The following proposition gives a formal statement of the worst case analysis. The proof follows exactly from Proposition 4.1 in [62].

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

However, one would expect that including multiple channels will provide more information to aid the recovery of the shared underlying object of interest. In a pMRI setting, it has been shown that parallel imaging (pMRI) outperforms single-coil MRI in many cases, and improved incoherence in the sensing matrix is often used as an explanation for such improvement. To understand the role of including multiple channels, and from the worst case analysis, it suggests a need to impose some probability model on the sensing parameters in order to prove some average case results.

Next, we state our main theorem that provides a sufficient condition on robust and stable recovery.

Theorem 1 *For any arbitrary subset* $S \in [N]$ *with* $card(S) = n$ *and* $\mathbf{X}_0 := \mathbf{z}_0 \otimes \mathbf{H}, \mathbf{z}_0 \in \mathbb{C}^{N \times 1}, \mathbf{H} \in \mathbb{C}^{k \times C}$, *consider the linear map* $\mathbf{A} \in \mathbb{C}^{L \times kN}$ *as defined in (25) that satisfies the randomness assumptions (A1)-(A4) and the noisy measurement* $\mathbf{Y} = \mathbf{A}\mathbf{X}_0 + \mathcal{N}$ *with* $\|\mathcal{N}\|_F \leq \sqrt{C}\sigma$. *The solution* $\mathbf{X}$ *to the* $\ell_{1,2}$ *minimization problem (28) satisfies*

$$\|\mathbf{X} - \mathbf{X}_0\|_F \leq C_1 \|\mathcal{P}_{S^c}\mathbf{X}_0\|_{1,2} + (C_2 + C_3\sqrt{s})\sqrt{C}\sigma \tag{32}$$

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

(C1.) the number of measurement satisfies

$$L \geq C_\alpha kn {\mu_{max}^B}^2 {\mu_{max}^\Psi}^2 \log(knN),$$

(C2.) $\mathbf{h}_l$ *satisfies the assumption (A4) and the number of coils satisfies*

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

Here, $\mu_{max}^B := \max_{i,j} \sqrt{L}|\mathbf{B}_{i,j}|$, $\mu_{max}^\Psi := \max_{i,j} \sqrt{N}|\boldsymbol{\Psi}_{i,j}|$, which measure the incoherence of the rows of $\mathbf{B}$ and of $\boldsymbol{\Psi}$, respectively. The constant C_α grows linearly with respect to α.

A detailed bound for the constants in the theorem can be found in the proofs in Section 6. Next, we use Lemma 1 to get rid of the $\sqrt{C}$ scaling in the error bound in (32).

Remark: The result provides a sufficient condition for stable and robust recovery in the presence of noise and sparse deficiency in the signal. It is important to emphasize that if the signal is truly sparse and there

is no noise present, perfect reconstruction is achieved (i.e., the error is zero). Hence, increasing the number of the coils does not provide any additional benefit when the number of measurements is sufficiently large. Furthermore, to understand the strength of the statement, note that if S is the true support of an n-sparse $\mathbf{z}_0$, $\|\mathcal{P}_{S^c}\mathbf{X}_0\|_{1,2} = 0$, then the number of required measurements L is of the order of the sparsity level of $\mathbf{X}_0$, up to log factors (i.e., $kn = \|P_S(\mathbf{X}_0)\|_0$). For the constants in the bound of (C1), it will be meaningful when the magnitudes of the entries in $\mathbf{B}$ and $\mathbf{\Psi}$ do not vary too much. Optimally, when the columns of $\mathbf{B}$ and $\mathbf{\Psi}$ are chosen from the DFT matrices, we optimally have $\mu_{max}^B = 1$, $\mu_{max}^\Psi = 1$.

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 [17]. Yet, we encountered some (at least to us) 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 of the bound and some discussions in the Appendix 7.1.

The proof of the main theorem follows a well-established proof outline of compressive sensing [17]. First, we refer to a sufficient condition from [4] to ensure a robust and stable recovery. Subsequently, we analyze the circumstances under which these conditions can be met. We show in 8 that under assumptions (A1)-(A4), the sufficient conditions for recovery are satisfied with a higher probability as the number of coils increases. This means that, in practice, parallel imaging is likely to perform better than single imaging recovery.

4.2 A sufficient condition

We state below a sufficient condition for solving the general the $\ell_{1,2}$ minimization problem for recovering block sparse signals.

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

$$\|(\boldsymbol{A}_S)^*\boldsymbol{A}_S - \boldsymbol{I}_S\|_{2\to 2} \leq \delta, \max_{j \in |S^c|} \|\boldsymbol{A}_S^*\boldsymbol{A}_{S_j^c}\|_{2\to 2} \leq \beta \tag{33}$$

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

$$\|\mathcal{P}_S\boldsymbol{\mathcal{Y}} - sgn(\boldsymbol{X}_0)\|_F \leq \eta, \max_{j \in |S^c|} \|\mathcal{P}_{S_j^c}\boldsymbol{\mathcal{Y}}\|_F \leq \theta \text{ and } \|\boldsymbol{V}\|_F \leq \tau\sqrt{s}. \tag{34}$$

Define $\rho = \theta + \frac{\eta\beta}{1-\delta}$, $\mu = \frac{\sqrt{1+\delta}}{1-\delta}$. If $\rho < 1$, then the solution $\boldsymbol{X}^*$ to the problem (27) obeys

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

The derivation of the lemma is a direct application of Lemma 7 of [4] which relies on the sub-differential of $\partial(\|\cdot\|_{(1,2)})$. Given $\mathbf{Z} \in \mathbb{C}^{kN \times C}$, the sub-differential $\partial_{\mathbf{Z}}(\|\cdot\|_{(1,2)})$ is the set

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

In the following sections, we will prove that with the assumptions we have made, the parameters δ and β will be small with high probability. We will also construct a dual certificate $\boldsymbol{\mathcal{Y}}$ with small η, θ and τ. Most of the proof procedure follows well-established methods in the compressive sensing literature, such as those described in [1] and [4].

15

4.3 Conditioning of $\mathbf{A}_S$

Next, we give the conditions on when our measurement matrix $\mathbf{A}$ satisfies the condition $\|\mathbf{A}_S^*\mathbf{A}_S - \mathbf{I}_S\| \leq \delta$. 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 (1) is zero and we would at the very least, require the condition (2) to be satisfied.

Lemma 4 *For operator $\mathbf{\Phi}$ defined in (23) which follows the assumptions induced by $\mathbf{\Psi}$ and $\mathbf{B}$ from (A1)-(A4) 4.1,*

$$\|\mathbf{\Phi}_S^*\mathbf{\Phi}_S - \boldsymbol{I}_S\| \leq \eta \tag{35}$$

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

$$N \geq C_\alpha {\mu_{max}^B}^2 {\mu_{max}^\Psi}^2 kn \log N / \eta^2, \tag{36}$$

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 [1], 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 $\mathbf{\Phi}$ (23), 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,

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

Combining Lemma 4 and Lemma 5, we have a probabilistic bound for $\|\mathbf{A}_S^*\mathbf{A}_S - \mathbf{I}_S\|_{2\to2}$. 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 provides a bound for the parameter β in $\max_{j\in|S^c|}\|\mathbf{A}_S^*\mathbf{A}_{S_j^c}\|_{2\to2} \leq \beta$ in Lemma 3.

Lemma 6 *Consider the forward matrix $\boldsymbol{A}$ defined in (23) and assume that $\|\boldsymbol{A}_S^*\boldsymbol{A}_S - \boldsymbol{I}_S\|_{2\to2} \leq \eta$,*

$$\max_{j\in[|S^c|]}\|\boldsymbol{A}_S^*\boldsymbol{A}_{S_j^c}\|_{2\to2} \leq \sqrt{\frac{1+\eta}{L}}\mu_{max}^\mathbf{\Psi}. \tag{37}$$

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 $\mathcal{Y} = \mathbf{A}^*\mathbf{A}_S(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\text{sgn}(\mathbf{X}_0)$ is the exact dual certificate. The next lemma deals with bounding the parameter τ in $\|\mathbf{V}\|_F \leq \tau\sqrt{s}$ in Lemma 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* $\|\boldsymbol{A}_S^*\boldsymbol{A}_S - \boldsymbol{I}_S\|_{2\to 2} \leq \eta$,

$$\|\boldsymbol{V}\|_F \leq \frac{\sqrt{1+\eta}}{1-\eta}\sqrt{n}. \tag{38}$$

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*

$$L > 6{\mu_{max}^{\Psi}}^2 \frac{1+\eta}{(1-\eta)^2}n, \tag{39}$$

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

Remark: Note that increasing the number of channels C increases the probability of recovery. Furthermore, the condition on L is automatically met due to the bound $L \sim O(kn\log(k^2n^2N))$ from Lemma 5.

Equivalently, we can state the result in Lemma 8 as $\max\limits_{j\in[|S^c|]} \|\mathcal{P}_{S_j^c}\mathcal{Y}\|_F < 1/2$ 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 Simulations

In the first part of the simulations, we demonstrate the sufficient conditions of L and C for exact recovery using simulated data with different underlying dimensions of k and n. We calculate the empirical probability of success for different values of k and n and show that the transition curves for L and C are consistent with the analytically derived bounds. In the second part, we present reconstruction results using simulated data and real pMRI measurements.

5.1 Empirical success rate for the proposed method

For the first set of experiments, we follow the same setup as in Ling and Strohmer [1] and calculate the empirical probability of success of the proposed method. Fix the number of measurements $L = 128$ and the underlying signal dimension $N = 256$, and solve the optimization problem (28) for varying $(k,n), k, n \in [15]$ and $C \in \{1, 2, 4\}$. 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 $\mathbf{H} \in \mathbb{C}^{k\times C}$ drawn from $\mathcal{N}(0,1)$. Each time, we draw L points uniformly from the Fourier space. 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 0.01%, i.e., $\frac{\|\hat{\mathbf{X}} - \mathbf{X}_0\|_F}{\|\mathbf{X}_0\|_F} \leq 10^{-4}$. And the empirical success rate is a fraction defined as (total number of successes) / (total number of experiments). We use the cvxpy [63, 64] package of Python with the MOSEK [65] solver to solve the mixed norm optimization problem (28).
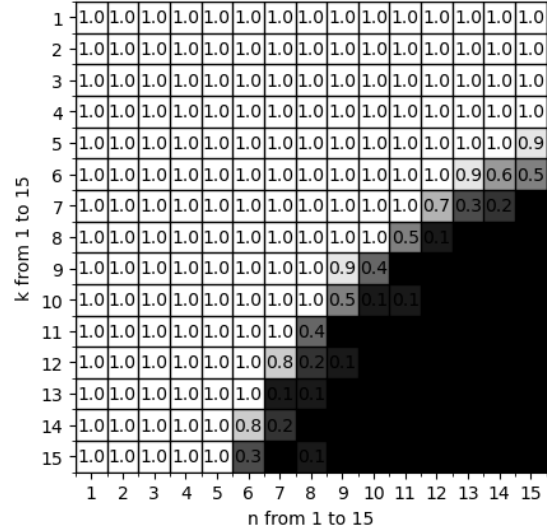
The Frequency of Success: L=128, N=256, C=1
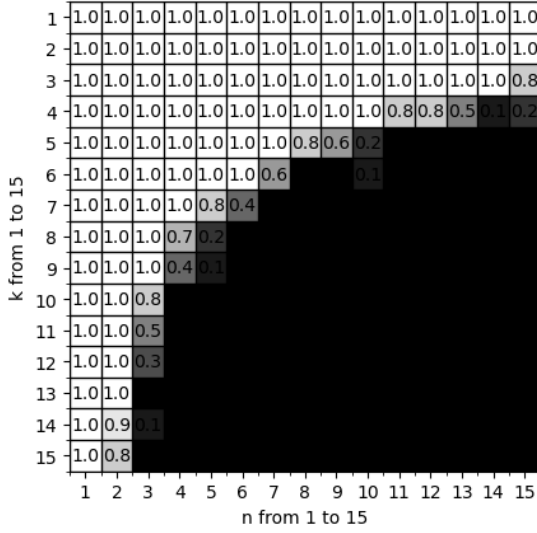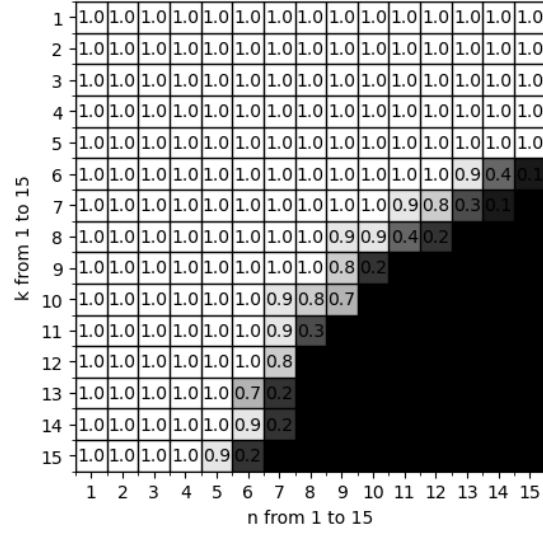
The Frequency of Success: L=128, N=256, C=2

(a) 1 coil

(b) 2 coil

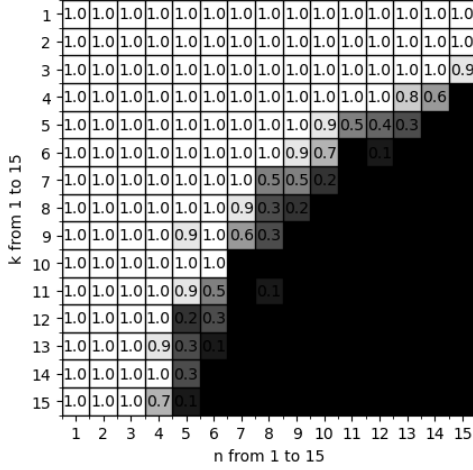The Frequency of Success: L=128, N=256, C=4
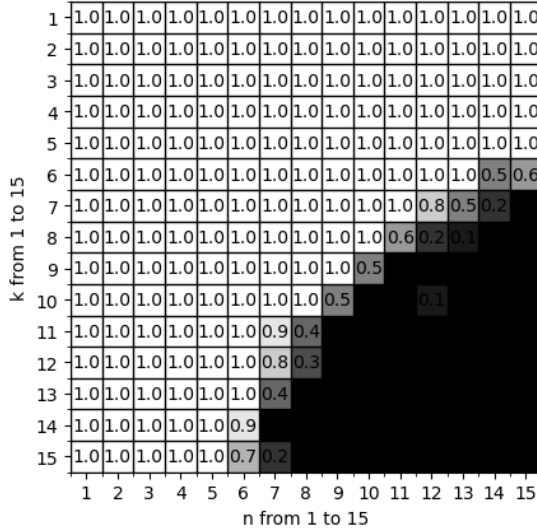
The Frequency of Success: L=128, N=256, C=8

(c) 4 coil

(d) 8 coil

Figure 2: Phase transition plot of performance by solving (28) 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.

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
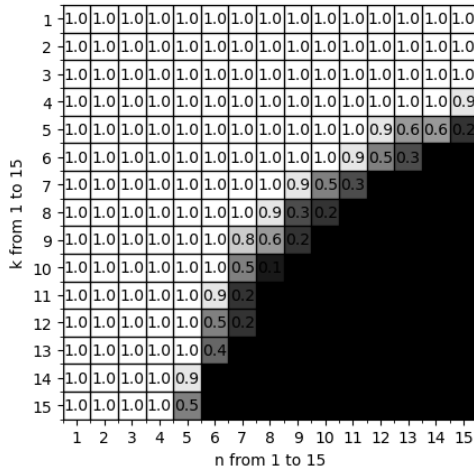
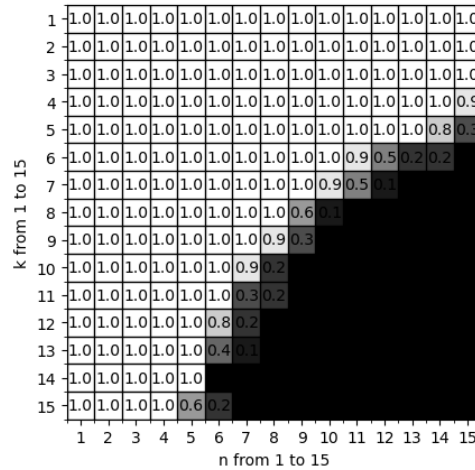The Frequency of Success: L=128, N=256, C=1, NSR=1%   The Frequency of Success: L=128, N=256, C=2, NSR=1%

(a) 1 coil
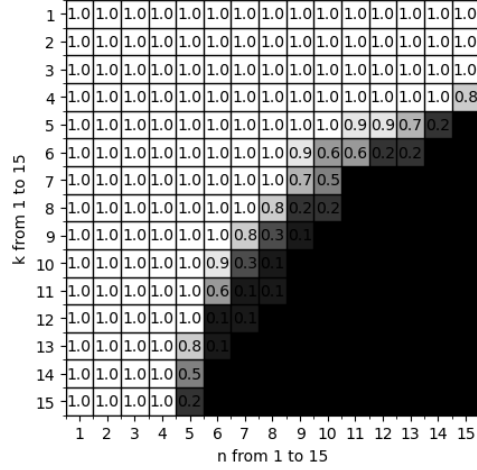
(b) 2 coil

The Frequency of Success: L=128, N=256, C=4, NSR=1%   The Frequency of Success: L=128, N=256, C=8, NSR=1%


(c) 4 coil

(d) 8 coil

Figure 3: Repeated the same experimental setup as in Fig:2, with 1% of noise added.

of a more relevant circumstance in practice. The results for choosing the DCT matrix as the sparsifying transformation (i.e., $\mathbf{\Psi}$) are shown in Fig:2 and Fig:3 for the noise-free and noisy cases, respectively.

Next, we illustrate the barrier for reconstruction with the proposed framework when $\mathbf{\Psi}$ is a 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, which can be used to sparsify MRI images [29], has high coherence. We followed the non-uniform sampling scheme in the Fourier measurements as suggested by [29, 66] as an attempt to improve the results. However, even when $L = N$ (when Fourier space is fully sampled), recovery is poor. We use the Daubechies 4 wavelet as an illustration, and the results are shown in Figure 4.
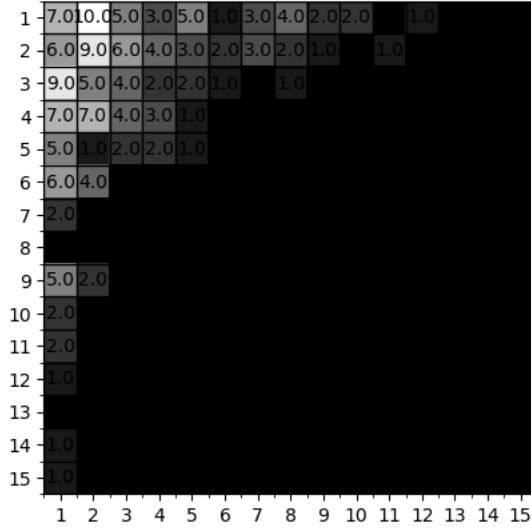
The Frequency of Success: L=128, N=256, C=4   The Frequency of Success: L=256, N=256, C=2
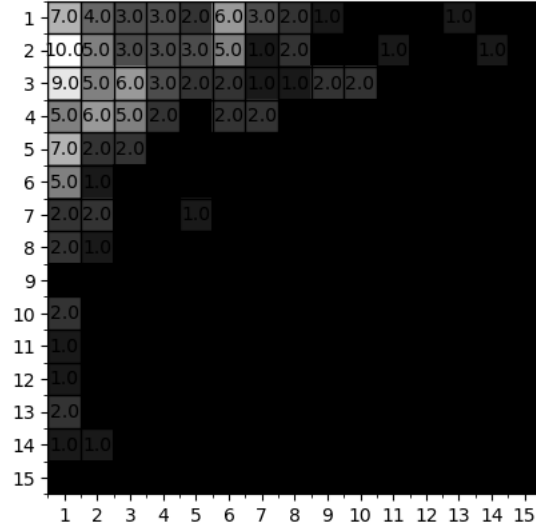
(a) 1 coil   (b) 2 coil

Figure 4: Phase transition plot of performance by solving (28) and using db4 wavelet transform as the sparsifying transformation matrix.

Nevertheless, since our method is designed to work with approximately sparse signals, and given that we will demonstrate the effectiveness of using the Discrete Cosine Transform (DCT) as the sparsifying transformation, it is not critical if the wavelet does not perform as well in this context.

5.1.1 Improved results compared to ℓ_1 minimization

Our proposed method minimizes $\sum\limits_{i \in [N]} \|\mathbf{X}_{S_i}\|_2$ to promote block sparsity within a lifted signal and between coils. We demonstrate its advantage over minimizing $\sum\limits_{i \in [C]} \|\mathbf{X}(:,i)\|_2$, which only promotes row sparsity between coils. We change the objective in (28) to minimize $\sum\limits_{i \in [C]} \|\mathbf{X}(:,i)\|_2$ and use the same experimental setup. The results are shown in Figure 5.

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 the sparsity level 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 Figure 6 and 7.
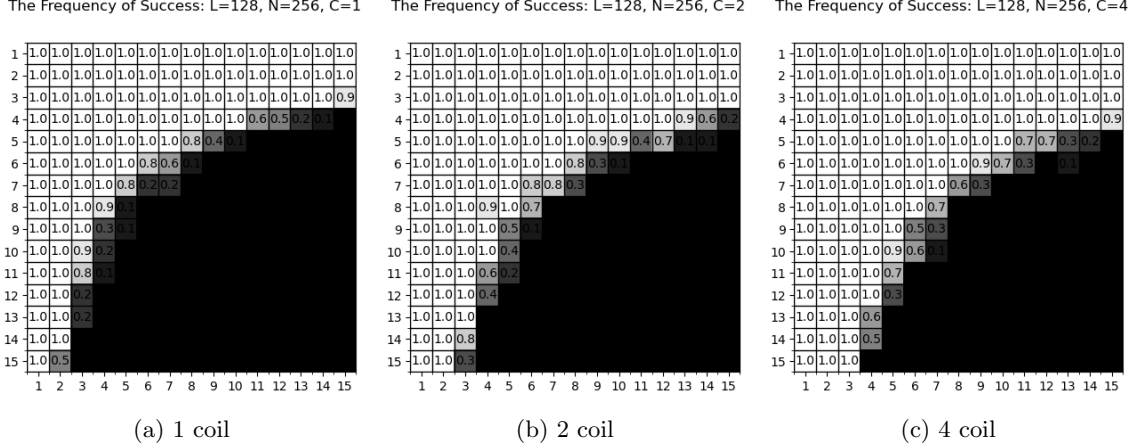
The Frequency of Success: L=128, N=256, C=1   The Frequency of Success: L=128, N=256, C=2   The Frequency of Success: L=128, N=256, C=4

(a) 1 coil   (b) 2 coil   (c) 4 coil

Figure 5: Phase transition plot of performance by minimizing $\sum\limits_{i\in[C]}\|\mathbf{X}(:,i)\|_2$ instead. Observe that compared to the proposed method, the improvement in the success rate is not obvious for including larger C.
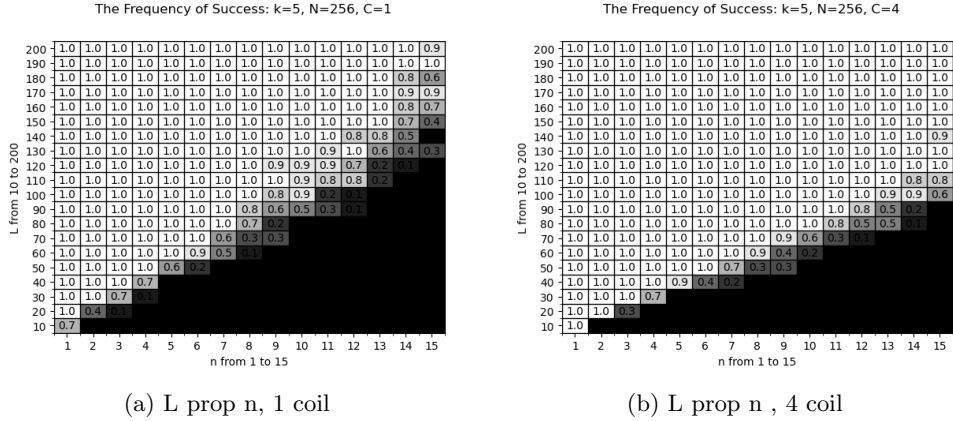
The Frequency of Success: k=5, N=256, C=1   The Frequency of Success: k=5, N=256, C=4

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

Figure 6: The empirical rate of success for a fixed k and varying (L, n, C).

5.2 Reconstruction of the analytical phantom

For the second part of the simulation, we show reconstruction results using the analytical phantom and simulated coils. We point out the major differences of the experimental setup from the last part:

- A crucial difference in the simulations is that we do not manually enforce the object of interest (i.e. $\mathbf{x}$ or $\mathbf{z}$) to be sparse/transform sparse. Hence, it allows for sparsity deficiency.

- Different levels of noise are added to the measurements.

- Instead of solving the constrained optimization problem (28), we solve the regularized problem (29) using an ISTA-type of updates as shown in Appendix 7.2.

We use the analytical phantom [46] of size 256 by 256 and randomly generated coil sensitivities following the 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
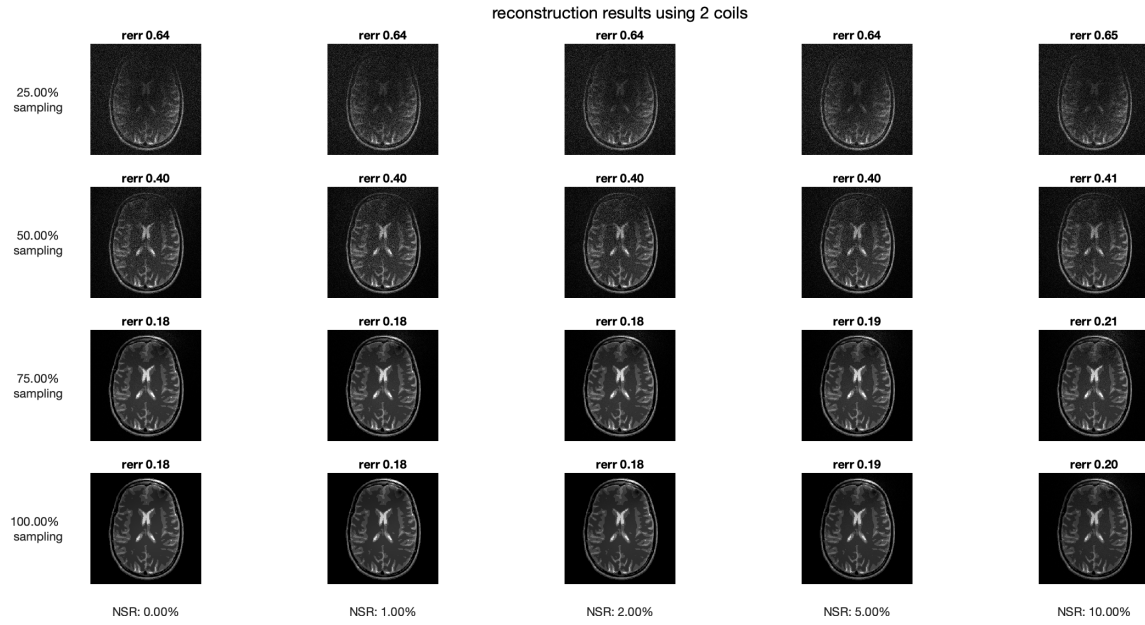
The Frequency of Success: n=5, N=256, C=1   The Frequency of Success: n=5, N=256, C=4
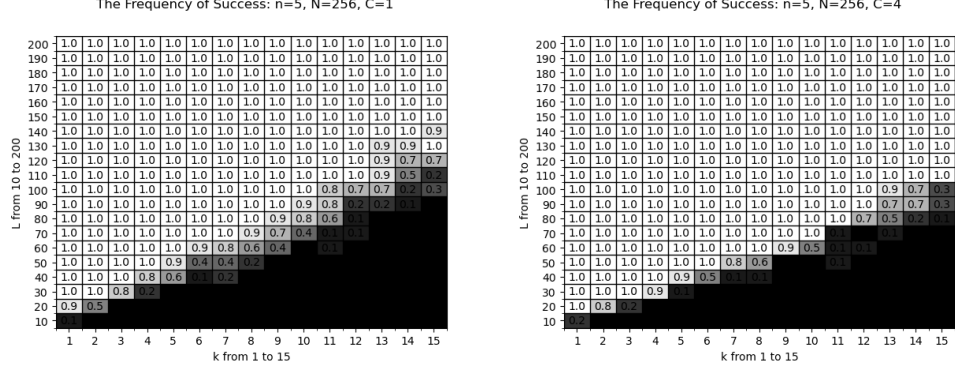
(a) L prop k, 1 coil   (b) L prop k , 4 coil

Figure 7: 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.

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 (rerr) of the reconstructed image and the true phantom. The reconstruction results are shown in Figure 8.

reconstruction results using 2 coils

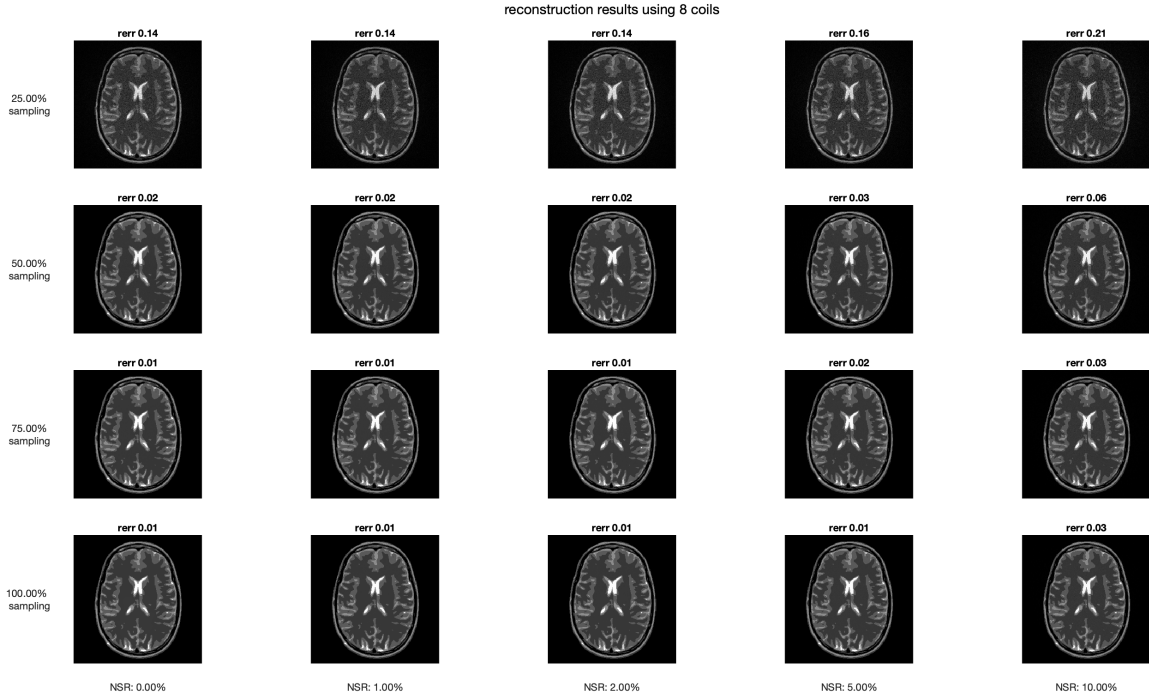|	rerr 0.64	rerr 0.64	rerr 0.64	rerr 0.64	rerr 0.65
25.00% sampling					
	rerr 0.40	rerr 0.40	rerr 0.40	rerr 0.40	rerr 0.41
50.00% sampling					
	rerr 0.18	rerr 0.18	rerr 0.18	rerr 0.19	rerr 0.21
75.00% sampling					
	rerr 0.18	rerr 0.18	rerr 0.18	rerr 0.19	rerr 0.20
100.00% sampling					
	NSR: 0.00%	NSR: 1.00%	NSR: 2.00%	NSR: 5.00%	NSR: 10.00%

(a) 2 coils

reconstruction results using 4 coils

	rerr 0.37	rerr 0.37	rerr 0.37	rerr 0.39	rerr 0.43
25.00% sampling					
	rerr 0.04	rerr 0.04	rerr 0.05	rerr 0.06	rerr 0.09
50.00% sampling					
	rerr 0.01	rerr 0.02	rerr 0.02	rerr 0.03	rerr 0.05
75.00% sampling					
	rerr 0.01	rerr 0.02	rerr 0.02	rerr 0.02	rerr 0.04
100.00% sampling					
	NSR: 0.00%	NSR: 1.00%	NSR: 2.00%	NSR: 5.00%	NSR: 10.00%

(b) 4 coils

23

(c) 8 coils

Figure 8: Reconstruction of the analytical phantom. 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 increased reduction factor.

To assess the performance of the proposed method with respect to the number of coils, we plot the average reconstruction error versus the number of measurements per coil, for different numbers of coils and levels of Gaussian noise. We adopted a similar setup as that in [50, 51] where the multi-coil signal was generated by multiplying a small analytical phantom (N = 529) by the coil sensitivities (k=6) and optionally adding Gaussian noise. The average reconstruction error (relative ℓ_2 distance) over 10 different random Fourier space undersampling realizations was computed for different numbers of measurements and coil elements. A DCT transform was employed to sparsify the image and no strict sparsity was enforced in the signal. The number of required measurements to achieve the noise floor reconstruction decreases with the number of coils for both the noise-free and the noisy case. These trends are consistent with previous numerical experiments for the known calibration pMRI reconstruction, e.g. [50]. The plots confirm that with a large number of coils, the number of measurements per coil sufficient to achieve a baseline level of reconstruction error is significantly smaller than the true dimensionality of the signal.
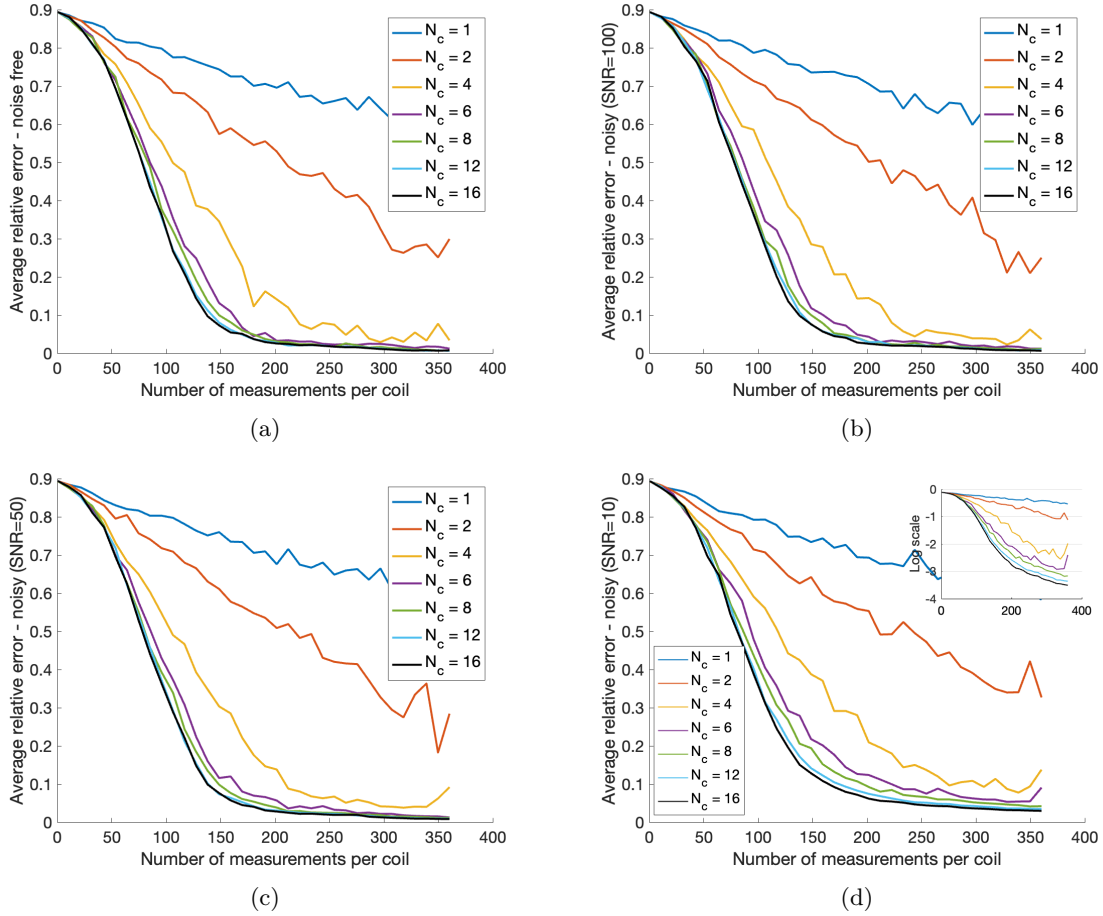

Figure 9: The four subplots correspond to the noise-free case and SNR values of 100, 50, and 10. The relative ℓ_2 error of the reconstruction is plotted against varying number of measurements per coil (L). For SNR=10, the log scale of the relative error is plotted in the inset of (d).

5.3 Comparison with the ENLIVE method

Next, we compare our reconstruction results with the ENLIVE formulation (17) in the paper by Holme et al. [47]. We reproduced the reconstructed brain and knee images exactly as shown in the original paper using the data and source code provided in `https://github.com/mrirecon/enlive`. We then compared these results with our own reconstructions. Figures 10 and 11 show calibrationless variable-density Poisson-disc undersampled reconstructions with differing undersampling factors comparing ENLIVE to our proposed method. Figure 12 shows the reconstructions of the simulated analytical phantom from uniformly subsampled Fourier space with different undersampling factors. As a reference, the corresponding sampling patterns are shown in the third columns of each of the plots.
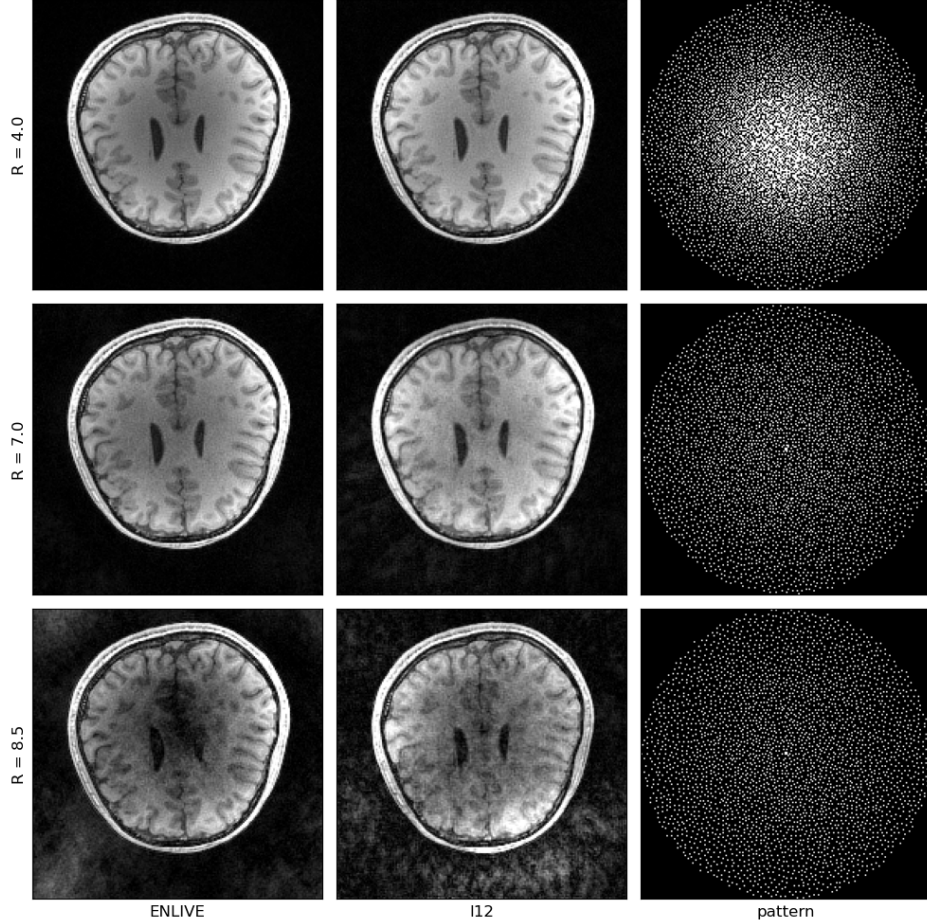

Figure 10: Variable-density Poisson-disc undersampled data with varying undersampling factors were reconstructed using ENLIVE (with $k = 2$ in (17)) as well as our proposed method. For undersampling factors up to $R = 7.0$, both methods produce images that are essentially realistic. However, at an undersampling factor of $R = 8.5$, the ENLIVE reconstruction becomes very noisy, and creates some features void in the center of the image. With higher undersampling factors, artifacts appear in the images reconstructed with our proposed method, but the brain retains more of its features and shows a more uniform contrast.
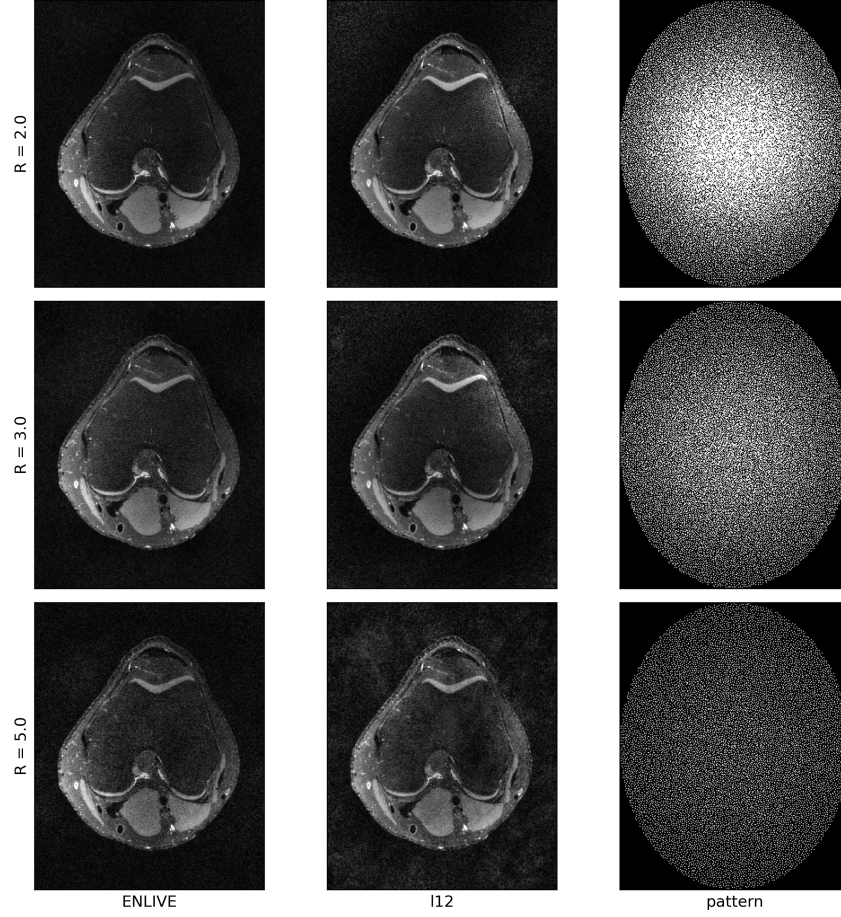
Figure 11: Variable-density Poisson-disc undersampled data of a human knee with varying undersampling factors reconstructed with ENLIVE allowing two sets of maps (i.e. with $k = 2$ in eq:(17)) and with our method. The proposed method provides reconstructions with higher and more consistent contrast. For R = 5.0, the reconstruction by ENLIVE seems preferable.
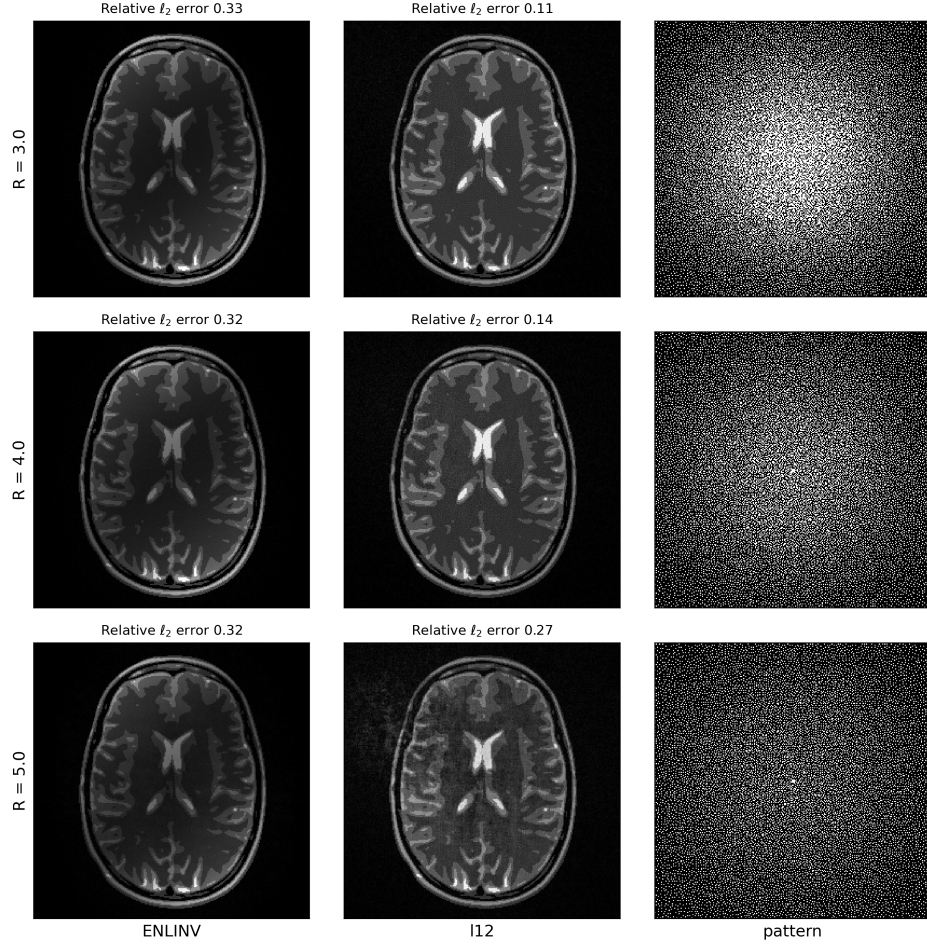
Figure 12: Uniformly undersampled noisy data of the analytical phantom and simulated coils with varying undersampling factors reconstructed with ENLIVE and our method. Relative ℓ_2 error of the reconstructed phantom and the true phantom is computed. For reduction factors up to $R = 5.0$, the ENLIVE produces reconstructions with low signal at the center, whereas our method produces images with uniform contrast throughout. The reconstructions of our method all have smaller ℓ_2 error than the ENLIVE reconstructions. For $R = 5.0$, artifacts start to appear in our reconstruction.

6 Proofs

6.1 Proof of Lemma 3

Recall that $\boldsymbol{\Phi} = \begin{bmatrix} (\bar{\mathbf{q}}_1 \otimes \mathbf{b}_1)^* \\ \vdots \\ \bar{\mathbf{q}}_N \otimes \mathbf{b}_N)^* \end{bmatrix}$, where each row of $\boldsymbol{\Phi}$ depends only on one row of $\mathbf{B}$ and $\boldsymbol{\Psi}^*$. Due to symmetry, the probability distribution of $\boldsymbol{\Phi}$ is unchanged if we assume $\mathbf{B}$ is fixed but the rows of $\boldsymbol{\Psi}^*$ are random. More precisely, we assume the rows of $\boldsymbol{\Psi}^* \in \mathbb{C}^{N \times M}$ are chosen uniformly at random without replacement from some $\boldsymbol{\Psi}_0^* \in \mathbb{C}^{N \times M}$ with $\boldsymbol{\Psi}_0^* \boldsymbol{\Psi}_0 = \mathbf{I}_N$. Under such a model, the rows of $\boldsymbol{\Psi}^*$ are not independent, making it difficult to analyze the probability model on $\boldsymbol{\Phi}$ directly. On the other hand, if the rows of $\boldsymbol{\Psi}$ are chosen uniformly at random with replacement from $\boldsymbol{\Psi}_0$, we can apply a matrix Bernstein inequality to estimate the operator norm of $\boldsymbol{\Psi}_S^* \boldsymbol{\Psi}_S - \mathbf{I}_S$. We first state the following Lemma which follows almost exactly from Lemma 4.3 in [1]. 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 $\boldsymbol{\Psi}^* \in \mathbb{C}^{N \times M}$ with its rows chosen uniformly at random with replacement from the rows of $\boldsymbol{\Psi}_0 \in \mathbb{C}^{N \times M}, \boldsymbol{\Psi}_0^* \boldsymbol{\Psi}_0 = \mathbf{I}_N$. For the matrix $\boldsymbol{\Phi}$ defined in (24),*

$$\|\boldsymbol{\Phi}_S^* \boldsymbol{\Phi}_S - \mathbf{I}_S\|_{F \to F} \leq \delta \tag{40}$$

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

The following lemma relates 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 $\boldsymbol{\Psi}^*$ with $\boldsymbol{\Psi}^* \boldsymbol{\Psi} = \boldsymbol{I}_N$, T' and T_m, define $\boldsymbol{\Psi}_{T'}^*$ and $\boldsymbol{\Psi}_{T_m}^*$ as matrices consisting of rows of $\boldsymbol{\Psi}^*$ indexed by T' and T_m respectively.
Given any fixed $\mathbf{B}$, consider the forward matrix $\boldsymbol{\Phi}_{T'}$ and $\boldsymbol{\Phi}_{T_m}$ formulated using $\boldsymbol{\Psi}_{T'}$ and $\boldsymbol{\Psi}_{T_m}$ respectively as defined in (23).
We have that

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

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

Proof:

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

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

$\square$

6.2 Proof of Lemma 4

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 $\boldsymbol{\Phi}_S$ with $\|\boldsymbol{\Phi}_S^* \boldsymbol{\Phi}_S - \mathbf{I}_S\| \leq \eta_1$,

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

provided,

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

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 $\mathbf{Z}_l := \mathbf{\Phi}_S^* \mathbf{a}_l^* \mathbf{\Phi}_S, l = 1 \cdots L$, 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 $p \geq 2$

$$\mathbb{E}\|\sum_{l=1}^{L} \mathbf{Z}_l - \mathbb{E}(\mathbf{Z}_l)\|_{2\to2}^p \leq 2^p \mathbb{E}\|\sum_{l=1}^{L} \epsilon_l \mathbf{Z}_l\|_{2\to2}^p,$$

where ϵ is a Rademacher sequence independent of Z.

Next, we use the following tail bound for matrix valued Rademacher sums,

$$\mathbb{P}_\epsilon(\|\sum_{l=1}^{L} \epsilon_l \mathbf{\Phi}_S^* \mathbf{a}_l \mathbf{a}_l^* \mathbf{\Phi}_S\|_{2\to2} \geq t\sigma) \leq 2kne^{-t^2/2}, t \geq 0,$$

where $\sigma = \|\sum_{l=1}^{L} \mathbf{\Phi}_S^* \mathbf{a}_l \mathbf{a}_l^* \mathbf{\Phi}_S \mathbf{\Phi}_S^* \mathbf{a}_l \mathbf{a}_l^* \mathbf{\Phi}_S\|_{2\to2}^{1/2} \leq \max_l\{|\mathbf{\Phi}_S^* \mathbf{a}_l|\}\|\mathbf{F}_{\Omega}\mathbf{\Phi}_S\|_{2\to2} = \|\mathbf{\Phi}_S^* \mathbf{F}_{\Omega}^*\|_{1\to2}\|\mathbf{F}_{\Omega}\mathbf{\Phi}_S\|_{2\to2}$. The derivation of the bound can be found, e.g. in Proposition 8.20 of [17].

By Proposition 7.13 in [17] regarding the moment estimation,

$$(\mathbb{E}_\epsilon\|\sum_{l=1}^{L} \epsilon_l \mathbf{Z}_l\|_{2\to2}^p)^{1/p} \leq C_1(C_2 kn)^{1/p}\sqrt{p}\|\mathbf{\Phi}_S^* \mathbf{F}_{\Omega}^*\|_{1\to2}\|\mathbf{F}_{\Omega}\mathbf{\Phi}_S\|_{2\to2},$$

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,

$$\mathbb{E}\|\sum_{l=1}^{L} \epsilon_l \mathbf{Z}_l\|_{2\to2}^p \leq C_1^p(C_2 kn)p^{p/2}\mathbb{E}(\|\mathbf{\Phi}_S^* \mathbf{F}_{\Omega}^*\|_{1\to2}^{2p})^{1/2}\mathbb{E}(\|\mathbf{F}_{\Omega}\mathbf{\Phi}_S\|_{2\to2}^{2p})^{1/2}$$

Combining the bounds, we have

$$(\mathbb{E}\|\sum_{l=1}^{L} \mathbf{Z}_l - \mathbb{E}(\mathbf{Z}_l)\|_{2\to2}^p)^{1/p} \leq 2(\mathbb{E}\|\sum_{l=1}^{L} \epsilon_l \mathbf{Z}_l\|_{2\to2}^p)^{1/p} \leq 2C_1(C_2 kn)^{1/p}\sqrt{p}\mathbb{E}(\|\mathbf{\Phi}_S^* \mathbf{F}_{\Omega}^*\|_{1\to2}^{2p})^{1/2p}\mathbb{E}(\|\mathbf{F}_{\Omega}\mathbf{\Phi}_S\|_{2\to2}^{2p})^{1/2p}.$$

Next, we estimate $\mathbb{E}(\|\mathbf{\Phi}_S^* \mathbf{F}_{\Omega}^*\|_{1\to2}^{2p})^{1/2p}$ and $\mathbb{E}(\|\mathbf{F}_{\Omega}\mathbf{\Phi}_S\|_{2\to2}^{2p})^{1/2p}$ respectively. For the first bound, observe that $\|\mathbf{\Phi}_S^* \mathbf{F}_{\Omega}^*\|_{1\to2} \leq \|\mathbf{\Phi}_S^*\|_{1\to2}\|\mathbf{F}_{\Omega}^*\|_{1\to1}$. Also, recall that $\mathbf{\Phi}_S = [\bar{\mathbf{q}}_{\mathbf{1,S}} \otimes \mathbf{b_1} \quad | \cdots \quad | \bar{\mathbf{q}}_{\mathbf{N,S}} \otimes \mathbf{b_N}]$. Hence, $\|\mathbf{\Phi}_S^*\|_{1\to2} = \max_i \|\mathbf{q}_{i,s} \otimes \mathbf{b}_i\|_2 = \max_i \|\mathbf{q}_{i,s}\|_2 \max_i \|\mathbf{b}_i\|_2$. By definition, $\mu_{max}^{\mathbf{B}} = \sqrt{N}\max_{i,j}|\mathbf{B}_{i,j}|$ and $\mu_{max}^{\Psi} = \sqrt{N}\max_{i,j}|\mathbf{\Psi}_{i,j}|$, we have $\|\mathbf{\Phi}_S^*\|_{1\to2} \leq \frac{\sqrt{kn}}{N}\mu_{max}^B\mu_{max}^{\Psi}$.

To bound $\|\mathbf{F}_{\Omega}^*\|_{1\to1}$, we explicitly write out that

$$\|\mathbf{F}_{\Omega}^*\|_{1\to1} = \max_{\|x\|_1=1}\sum_{j\in[N]}|\sum_{i\in[L]}\overline{\mathbf{F}}_{i,j}x_i| \leq \max_{\|x\|_1=1} N\max_j|\sum_i \overline{\mathbf{F}}_{i,j}x_i| \leq N\max_{i,j}|\mathbf{F}_{i,j}| \leq N/\sqrt{L},$$

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 $\mathbf{P} = \mathrm{diag}(\delta_1, \cdots, \delta_N)$ and δ_i are independent Bernoulli variables with $\mathbb{E}(\delta_i) = L/N$.

Apply Lemma 14.3 [17] on the non-square matrix $\mathbf{\Phi}_S^* \mathbf{F}_\Omega^*$, for any $p \geq 1$,

$$\mathbb{E}_\delta(\|\mathbf{\Phi}_S^*\mathbf{F}^*\mathbf{P}\|_{2\to 2}^{2p})^{1/2p} \leq \sqrt{2}C_1(C_2 kn)^{1/p}\sqrt{p}\mathbb{E}(\|\mathbf{\Phi}_S^*\mathbf{F}_\Omega^*\|_{1\to 2}^{2p})^{1/2p} + \sqrt{\frac{L}{N}}\|\mathbf{\Phi}_S^*\mathbf{F}_\Omega^*\|_{2\to 2}.$$

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

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

$$\leq 2C_1(C_2 kn)^{1/p}\sqrt{p}\sqrt{\frac{kn}{L}}\mu_{max}^B\mu_{max}^{\mathbf{\Psi}}(\sqrt{2}C_1(C_2 kn)^{1/p}\sqrt{p}\sqrt{\frac{kn}{L}}\mu_{max}^B\mu_{max}^{\mathbf{\Psi}} + \|\mathbf{\Phi}_S\|_{2\to 2})$$

$$\leq (C_2 kn)^{2/p}(2\sqrt{2}C_1^2\frac{knp}{L}{\mu_{max}^B}^2{\mu_{max}^{\mathbf{\Psi}}}^2 + 2C_1\sqrt{\frac{knp}{L}}\mu_{max}^B\mu_{max}^{\mathbf{\Psi}}\|\mathbf{\Phi}_S\|_{2\to 2}).$$

Then, we apply a tail bound from moment estimation, see Proposition 7.15 from [17]. 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 $\beta = C_2^2 k^2 n^2$, $\alpha_1 = 2\sqrt{2}C_1^2\frac{kn}{L}{\mu_{max}^B}^2{\mu_{max}^{\mathbf{\Psi}}}^2$, $\alpha_2 = 2C_1\sqrt{\frac{kn}{L}}\mu_{max}^B\mu_{max}^{\mathbf{\Psi}}\|\mathbf{\Phi}_s\|_{2\to 2}$.

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,

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

By the definition of α_1 and α_2,

$$e\alpha_1 \ln(C_2^2 k^2 n^2/\epsilon) \leq \eta/2 \to L \geq \frac{C_3{\mu_{max}^B}^2{\mu_{max}^{\mathbf{\Psi}}}^2 kn\ln(\frac{C_2^2 k^2 n^2}{\epsilon})}{\eta}, C_3 = 4\sqrt{2}eC_1^2.$$

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

$$\to L \geq \frac{C_4^2{\mu_{max}^B}^2{\mu_{max}^{\mathbf{\Psi}}}^2 kn\ln(C_2^2 k^2 n^2/\epsilon)\|\mathbf{\Phi}_S\|_{2\to 2}^2}{\eta^2}.$$

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,

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

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 [17], where it shows the probability for the uniform model is bounded by twice the one for the Bernoulli model derived here.

$\square$

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 $\mathbf{A}_i = \mathbf{F}_\Omega \mathbf{\Phi}_i$.

$$\max_{j \in [|S^c|]} \|\mathbf{A}_S^* \mathbf{A}_{S_j^c}\|_{2\to2}$$
$$\leq \|\mathbf{A}_S\|_{2\to2} \max_i \|\mathbf{A}_i\|_{2\to2}$$
$$\leq \sqrt{1+\eta} \max_i \|\mathbf{A}_i\|_{2\to2}.$$

The last inequality is implied by $\|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}_S\|_{2\to2} \leq \eta$.

Since $\mathbf{A}_i = \mathbf{F}_\Omega \mathbf{\Phi}_i = \mathbf{F}_\Omega \text{diag}(\mathbf{\Psi}_i)\mathbf{B}$, we have $\|\mathbf{A}_i\|_{2\to2} \leq \|\mathbf{F}_\Omega\|_{2\to2}\|\mathbf{\Phi}_i\|_{2\to2}$. 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 $\mathbf{F}_\Omega \mathbf{F}_\Omega^* = \frac{N}{L}\mathbf{I}_L \to \|\mathbf{F}_\Omega\|_{2\to2} = \sqrt{\frac{N}{L}}$. 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, $\|\mathbf{\Phi}_i\|_{2\to2}^2 = \|\mathbf{\Phi}_i^*\mathbf{\Phi}_i\| = \|\mathbf{B}^*\text{diag}([|\mathbf{\Psi}_{i,1}|^2, \cdots, |\mathbf{\Psi}_{i,N}|^2])\mathbf{B}\|_{2\to2} \leq \frac{1}{N}(\mu_{max}^{\mathbf{\Psi}})^2\|\mathbf{B}^*\mathbf{B}\|_{2\to2} = \frac{1}{N}(\mu_{max}^{\mathbf{\Psi}})^2$. The last equality is due to the assumption that $\mathbf{B}^*\mathbf{B} = I_k$.

Combining the bounds,

$$\max_{j \in [|S^c|]} \|\mathbf{A}_S^* \mathbf{A}_{S_j}\|_{2\to2} \leq \sqrt{\frac{1+\eta}{L}}\mu_{max}^{\mathbf{\Psi}}.$$

$\square$

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

$$\|\mathbf{V}\|_F = \|\mathbf{A}_S(\mathbf{A}_S^*\mathbf{A}_S)^{-1}sgn(\mathbf{X}_0)\|_F \leq \frac{\sqrt{1+\eta}}{1-\eta}\sqrt{n}. \tag{41}$$

The inequality comes from the implication by the conditioning of $\mathbf{A}_S$ that $\mathbf{A}_S^*\mathbf{A}_S$ is invertible with $\|(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\|_{2\to2} \leq \frac{1}{1-\eta}$ and $\|\mathbf{A}_S\|_{2\to2} \leq \sqrt{1+\eta}$.

$\square$

6.5 Proof of Lemma 8

We consider the case when the spanning coefficients are chosen uniformly at random from a fixed basis. The structure of the proof is similar to the Gaussian case. Suppose the columns of $\mathbf{H}$ are chosen uniformly at random 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 exists 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 $\tilde{\boldsymbol{X}} = \begin{bmatrix} \mathbf{h_1}sgn(z_1) & \mathbf{h_2}sgn(z_1) & \cdots & \mathbf{h_c}sgn(z_1) \\ \mathbf{h_1}sgn(z_2) & \mathbf{h_2}sgn(z_2) & \cdots & \mathbf{h_c}sgn(z_2) \\ \vdots & \vdots & \vdots & \\ \mathbf{h_1}sgn(z_N) & \mathbf{h_2}sgn(z_N) & \cdots & \mathbf{h_c}sgn(z_N) \end{bmatrix}$. Then $sgn(\mathbf{X}_0) = \tilde{\boldsymbol{X}}/\|\mathbf{H}\|_F \in \mathbb{C}^{kN \times C}$, where $\mathbf{H} = [\mathbf{h_1}|\mathbf{h_2}|\cdots|\mathbf{h_C}]$, $\mathbf{H} \in \mathbb{C}^{k \times C}$.

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

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

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

We want to bound the two terms, $\| \mathbf{H} \|_F$ and $\| \mathbf{H} \|_{2 \to 2}$. For $\| \mathbf{H} \|_F$, since each column of $\mathbf{H}$ 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 $\mathbb{E}(\mathbf{Z}_l) = 0$,

$$R = \max_l \| \mathbf{Z}_l \| = \| \mathbf{h}_l \mathbf{h}_l^* - \frac{1}{k} I \| \leq \max_l \max(\| \mathbf{h}_l \|^2, \frac{1}{k}) = 1, \tag{42}$$

$$\sigma^2 = \| \Sigma_{l=1}^C \mathbb{E}(\mathbf{Z}_l \mathbf{Z}_l^*) \| = \| \Sigma_{l=1}^C \mathbb{E}(\mathbf{h}_l \underbrace{\mathbf{h}_l^* \mathbf{h}_l}_{=1} \mathbf{h}_l^*) - \frac{2}{k} \mathbb{E}(\mathbf{h}_l \mathbf{h}_l^*) + \frac{I}{k^2} \| = \| \Sigma_{l=1}^C (\frac{I}{k} - \frac{I}{k^2}) \| \leq C/k. \tag{43}$$

Hence,

$$\mathbb{P}(\| \mathbf{H} \mathbf{H}^* - \frac{C}{k} I \| \geq t) \leq 2k \exp(-\frac{t^2/2}{\sigma^2 + Rt/3}) = 2k \exp(-\frac{t^2/2}{\frac{C}{k} + t/3}). \tag{44}$$

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

Combining the bounds, we have

$$\frac{\sqrt{1 + \eta}}{1 - \eta} \sqrt{\frac{1}{L}} \mu_{max}^{\mathbf{\Psi}} \sqrt{3n/2} < \theta < 1 \tag{45}$$

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 vectors drawn from $\frac{1}{\sqrt{2}} \mathcal{N}(0, 1) + i \frac{1}{\sqrt{2}} \mathcal{N}(0, 1)$. □

Acknowledgement

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 $\mathbf{\Phi}$. We omit the proof in the manuscript.

Lemma 11 *Suppose* $\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_1 & \cdots & \mathbf{\Phi}_N \end{bmatrix}$, $\mathbf{\Phi}_i \in \mathbb{C}^{N \times k}$ *and* $\mathbf{\Phi}_i^* \mathbf{\Phi}_i = \mathbf{I}_k$ *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 $(1 - \epsilon)$, $\|\mathbf{\Phi}_S^* \mathbf{\Phi}_S - \mathbf{I}_S\| \leq \eta$.

Note that the derivation of the Lemma does not depend on any structures of Φ. However, take into account the structure of our specific $\mathbf{\Phi}$, defined as $\mathbf{\Phi} = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\psi}_1)\mathbf{B} & \cdots & \mathrm{diag}(\boldsymbol{\psi}_N)\mathbf{B} \end{bmatrix}$. The condition on μ_{block} is strict. Note that the derivation of the Lemma does not depend on any structures of Φ. Recall that our specific $\mathbf{\Phi}$ is defined as $\mathbf{\Phi} = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\psi}_1)\mathbf{B} & \cdots & \mathrm{diag}(\boldsymbol{\psi}_N)\mathbf{B} \end{bmatrix}$. When examining the derived conditions for our specific $\mathbf{\Phi}$, we find that it does not readily satisfy the necessary requirements; this makes the Lemma less useful in this context.

7.2 Algorithm

Recall from (26), $\|\mathbf{X}\|_{1,2} = \sum_{j \in [N]} \|\mathbf{X}_{T_j \times [C]}\|_F$. Let $f(\mathbf{X}) = \frac{1}{2}\|\mathbf{Y} - \mathbf{AX}\|_F^2$ and $g(\mathbf{X}) = \lambda\|\mathbf{X}\|_{1,2}$, an basic update of ISTA [67] for problem (29) at iteration k using stepsize L is,

$$\mathbf{X}_k = \operatorname*{argmin}_{\mathbf{X}}\{g(\mathbf{X}) + \frac{L}{2}\|\mathbf{X} - [\mathbf{X}_{k-1} - \frac{1}{L}\nabla f(\mathbf{X}_{k-1})]\|_F^2\}. \tag{46}$$

Together with $\nabla f(\mathbf{X}) = \mathbf{A}^*(\mathbf{AX} - \mathbf{Y})$ and define $\mathbf{Z}_{k-1} = \mathbf{X}_{k-1} - \frac{1}{L}\mathbf{A}^*(\mathbf{AX}_{k-1} - \mathbf{Y})$, the update (46) becomes,

$$\mathbf{X}_k = \operatorname*{argmin}_{\mathbf{X}}\{g(\mathbf{X}) + \frac{L}{2}\|\mathbf{X} - \mathbf{Z}_{k-1}\|_F^2\}. \tag{47}$$

Now, consider the general problem of

$$\mathbf{X} = \operatorname*{argmin}_{\mathbf{X}}\{\lambda\|\mathbf{X}\|_{1,2} + \frac{L}{2}\|\mathbf{X} - \mathbf{Z}\|_F^2\}. \tag{48}$$

Note that the optimization problem (48) is separable. In particular, if we define $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}$, $\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_N \end{bmatrix}$,

$\operatorname*{argmin}_{\mathbf{X}}\{\lambda\|\mathbf{X}\|_{1,2} + \frac{L}{2}\|\mathbf{X} - \mathbf{Z}\|_F^2\} = \operatorname*{argmin}_{\mathbf{X}_1, \cdots, \mathbf{X}_N}\{\sum_{i \in [N]} \lambda\|\mathbf{X}_i\|_F + \frac{L}{2}\|\mathbf{X}_i - \mathbf{Z}_i\|_F^2\}$. Furthermore,

$$\operatorname*{argmin}_{\mathbf{X}_i}\{\sum_{i \in [N]} \lambda\|\mathbf{X}_i\|_F + \frac{L}{2}\|\mathbf{X}_i - \mathbf{Z}_i\|_F^2\}$$

$$= \operatorname*{argmin}_{\mathbf{X}_i}\{\sum_{i \in [N]} \frac{\lambda}{L}\|\mathbf{X}_i\|_F + \frac{1}{2}\|\mathbf{X}_i - \mathbf{Z}_i\|_F^2\} \tag{49}$$

$$= \mathrm{prox}_{\frac{\lambda}{L}\|\cdot\|_F}(\mathbf{X}_i) = \mathbf{Z}_i(1 - \frac{\lambda/L}{\max(\|\mathbf{Z}_i\|_F, \lambda/L)})$$

References

1. Ling S and Strohmer T. Self-calibration and Biconvex Compressive Sensing. Inverse Problems 2015; 31:115002. DOI: 10.1088/0266-5611/31/11/115002

2. Candès EJ, Eldar YC, Strohmer T, and Voroninski V. Phase Retrieval via Matrix Completion. SIAM Journal on Imaging Sciences 2013; 6:199–225. DOI: 10.1137/110848074

3. Ahmed A, Recht B, and Romberg J. Blind Deconvolution using Convex Programming. IEEE Transactions on Information Theory 2014; 60:1711–32. DOI: 10.1109/tit.2013.2294644

4. Flinth A. Sparse Blind Deconvolution and Demixing Through $\ell_{1,2}$-Minimization. Advances in Computational Mathematics 2017; 44:1–21. DOI: 10.1007/s10444-017-9533-0

5. Aja-Fernández S, Vegas-Sánchez-Ferrero G, and Tristán-Vega A. Noise Estimation in Parallel MRI: GRAPPA and SENSE. Magnetic Resonance Imaging 2014; 32:281–90. DOI: 10.1016/j.mri.2013.12.001

6. Robson PM, Grant AK, Madhuranthakam AJ, Lattanzi R, Sodickson DK, and McKenzie CA. Comprehensive Quantification of Signal-to-Noise Ratio and g-Factor for Image-Based and k-Space-Based Parallel Imaging Reconstructions. Magnetic Resonance in Medicine 2008; 60:895–907. DOI: 10.1002/mrm.21728

7. Ling S and Strohmer T. Self-calibration and Bilinear Inverse Problems via Linear Least Squares. SIAM Journal on Imaging Sciences 2018; 11:252–92. DOI: 10.1137/16m1103634

8. Lee K, Tian N, and Romberg J. Fast and Guaranteed Blind Multichannel Deconvolution under a Bilinear System Model. IEEE Transactions on Information Theory 2018; 64:4792–818. DOI: 10.1109/tit.2018.2840711

9. Li X, Ling S, Strohmer T, and Wei K. Rapid, Robust, and Reliable Blind Deconvolution via Nonconvex Optimization. Applied and Computational Harmonic Analysis 2019; 47:893–934. DOI: 10.1016/j.acha.2018.01.001

10. Qu Q, Li X, and Zhu Z. Exact and Efficient Multi-Channel Sparse Blind Deconvolution — A Nonconvex Approach. 2019 53rd Asilomar Conference on Signals, Systems, and Computers 2019. DOI: 10.1109/ieeeconf44664.2019.9049053

11. Kech M and Krahmer F. Optimal Injectivity Conditions for Bilinear Inverse Problems with Applications to Identifiability of Deconvolution Problems. SIAM Journal on Applied Algebra and Geometry 2017; 1:20–37. DOI: 10.1137/16m1067469

12. Choudhary S and Mitra U. On Identifiability in Bilinear Inverse Problems. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing 2013. DOI: 10.1109/icassp.2013.6638476

13. Bahmani S and Romberg J. Lifting for Blind Deconvolution in Random Mask Imaging: Identifiability and Convex Relaxation. SIAM Journal on Imaging Sciences 2015; 8:2203–38. DOI: 10.1137/141002165

14. Tang G and Recht B. Convex Blind Deconvolution with Random Masks. Classical Optics 2014 2014. DOI: 10.1364/cosi.2014.cw4c.1

15. Levin A, Weiss Y, Durand F, and Freeman WT. Understanding Blind Deconvolution Algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 2011; 33:2354–67. DOI: 10.1109/tpami.2011.148

16. Lee K, Li Y, Junge M, and Bresler Y. Blind Recovery of Sparse Signals from Subsampled Convolution. IEEE Transactions on Information Theory 2017; 63:802–21. DOI: 10.1109/tit.2016.2636204

17. Foucart S and Rauhut H. A Mathematical Introduction to Compressive Sensing. Springer New York, 2015

18. Harikumar G and Bresler Y. Perfect Blind Restoration of Images Blurred by Multiple Filters: Theory and Efficient Algorithms. IEEE Transactions on Image Processing 1999; 8:202–19. DOI: 10.1109/83.743855

19. Balzano L and Nowak R. Blind Calibration of Sensor Networks. 2007 6th International Symposium on Information Processing in Sensor Networks 2007. DOI: 10.1109/ipsn.2007.4379667

20. Gribonval R, Chardon G, and Daudet L. Blind Calibration for Compressed Sensing by Convex Optimization. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2012. DOI: 10.1109/icassp.2012.6288477

21. Bilen C, Puy G, Gribonval R, and Daudet L. Convex Optimization Approaches for Blind Sensor Calibration Using Sparsity. IEEE Transactions on Signal Processing 2014; 62:4847–56. DOI: 10.1109/tsp.2014.2342651

22. Wang L and Chi Y. Blind Deconvolution from Multiple Sparse Inputs. IEEE Signal Processing Letters 2016; 23:1384–8. DOI: 10.1109/lsp.2016.2599104

23. Oymak S, Jalali A, Fazel M, Eldar YC, and Hassibi B. Simultaneously Structured Models with Application to Sparse and Low-Rank Matrices. IEEE Transactions on Information Theory 2015; 61:2886–908. DOI: 10.1109/tit.2015.2401574

24. Deshmane A, Gulani V, Griswold MA, and Seiberlich N. Parallel MR Imaging. Journal of Magnetic Resonance Imaging 2012; 36:55–72. DOI: 10.1002/jmri.23639

25. Glockner JF, Hu HH, Stanley DW, Angelos L, and King K. Parallel MR Imaging: A User's Guide. RadioGraphics 2005; 25:1279–97. DOI: 10.1148/rg.255045202

26. Lustig M, Donoho D, and Pauly JM. Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging. Magnetic Resonance in Medicine 2007; 58:1182–95. DOI: 10.1002/mrm.21391

27. Shin PJ, Larson PE, Ohliger MA, Elad M, Pauly JM, Vigneron DB, and Lustig M. Calibrationless Parallel Imaging Reconstruction Based on Structured Low-Rank Matrix Completion. Magnetic Resonance in Medicine 2014; 72:959–70. DOI: 10.1002/mrm.24997

28. Ravishankar S and Bresler Y. Efficient Blind Compressed Sensing Using Sparsifying Transforms with Convergence Guarantees and Application to Magnetic Resonance Imaging. SIAM Journal on Imaging Sciences 2015; 8:2519–57. DOI: 10.1137/141002293

29. Chun IY, Adcock B, and Talavage TM. Efficient Compressed Sensing SENSE Parallel MRI Reconstruction with Joint Sparsity Promotion and Mutual Incoherence Enhancement. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2014. DOI: 10.1109/embc.2014.6944111

30. Elad M, Milanfar P, and Rubinstein R. Analysis versus Synthesis in Signal Priors. Inverse Problems 2007; 23:947–68. DOI: 10.1088/0266-5611/23/3/007

31. Pratt W, Kane J, and Andrews H. Hadamard Transform Image Coding. Proceedings of the IEEE 1969; 57:58–68. DOI: 10.1109/proc.1969.6869

32. Aharon M, Elad M, and Bruckstein A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. IEEE Transactions on Signal Processing 2006; 54:4311–22. DOI: 10.1109/tsp.2006.881199

33. Ravishankar S and Bresler Y. Learning Sparsifying Transforms for Image Processing. 2012 19th IEEE International Conference on Image Processing 2012. DOI: 10.1109/icip.2012.6466951

34. Liang D, Liu B, Wang J, and Ying L. Accelerating SENSE Using Compressed Sensing. Magnetic Resonance in Medicine 2009; 62:1574–84. DOI: 10.1002/mrm.22161

35. Liu B, Zou YM, and Ying L. SPARSESENSE: Application of Compressed Sensing in Parallel MRI. 2008 International Conference on Technology and Applications in Biomedicine 2008

36. Ji JX, Zhao C, and Lang T. Compressed Sensing Parallel Magnetic Resonance Imaging. 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2008

37. Wu B, Millane RP, Watts R, and Bones PJ. Prior Estimate-Based Compressed Sensing in Parallel MRI. Magnetic Resonance in Medicine 2010; 65:83–95. DOI: 10.1002/mrm.22545

38. Pruessmann KP, Weiger M, Scheidegger MB, and Boesiger P. SENSE: Sensitivity Encoding for Fast MRI. Magnetic Resonance in Medicine 1999; 42:952–62

39. Committee on the Mathematics and Physics of Emerging Dynamic Biomedical Imaging, National Research Council. Mathematics and Physics of Emerging Biomedical Imaging. National Academy Press, 1996

40. Liang ZP and Lauterbur PC. Principles of Magnetic Resonance Imaging. SPIE Optical Engineering Press, 2000

41. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, and Haase A. Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). Magnetic Resonance in Medicine 2002; 47:1202–10. DOI: 10.1002/mrm.10171

42. McKenzie CA, Yeh EN, Ohliger MA, Price MD, and Sodickson DK. Self-Calibrating Parallel Imaging with Automatic Coil Sensitivity Extraction. Magnetic Resonance in Medicine 2002; 47:529–38. DOI: 10.1002/mrm.10087

43. Morrison R, Jacob M, and Do M. Multichannel Estimation of Coil Sensitivities in Parallel MRI. 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro 2007. DOI: 10.1109/isbi.2007.356802

44. Ying L and Sheng J. Joint Image Reconstruction and Sensitivity Estimation in SENSE (JSENSE). Magnetic Resonance in Medicine 2007; 57:1196–202. DOI: 10.1002/mrm.21245

45. Uecker M, Hohage T, Block KT, and Frahm J. Image Reconstruction by Regularized Nonlinear Inversion—Joint Estimation of Coil Sensitivities and Image Content. Magnetic Resonance in Medicine 2008; 60:674–82. DOI: 10.1002/mrm.21691

46. Guerquin-Kern M, Lejeune L, Pruessmann KP, and Unser M. Realistic Analytical Phantoms for Parallel Magnetic Resonance Imaging. IEEE Transactions on Medical Imaging 2012; 31:626–36. DOI: 10.1109/tmi.2011.2174158

47. Holme HC, Rosenzweig S, Ong F, Wilke RN, Lustig M, and Uecker M. ENLIVE: An Efficient Nonlinear Method for Calibrationless and Robust Parallel Imaging. Scientific Reports 2019; 9. DOI: 10.1038/s41598-019-39888-7

48. Candes E, Romberg J, and Tao T. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. IEEE Transactions on Information Theory 2006; 52:489–509. DOI: 10.1109/tit.2005.862083

49. Recht B, Fazel M, and Parrilo PA. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. SIAM Review 2010; 52:471–501. DOI: 10.1137/070697835

50. Otazo R and Sodickson DK. Distributed Compressed Sensing for Accelerated MRI. *Proceedings of the 17th Annual Meeting of ISMRM.* 2009 :378

51. Otazo R, Kim D, Axel L, and Sodickson DK. Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI. Magnetic Resonance in Medicine 2010 Aug; 64:767–76. DOI: 10.1002/mrm.22463

52. Duarte M, Sarvotham S, Baron D, Wakin M, and Baraniuk R. Distributed compressed sensing of jointly sparse signals. Conference Record of the Thirty-Ninth Asilomar Conference onSignals, Systems and Computers 2005. DOI: 10.1109/acssc.2005.1600024

53. Tsaig Y and Donoho DL. Extensions of compressed sensing. Signal Processing 2006 Mar; 86:549–71. DOI: 10.1016/j.sigpro.2005.05.029

54. Lustig M and Pauly JM. SPIRiT: Iterative Self-Consistent Parallel Imaging Reconstruction from Arbitrary K-Space. Magnetic Resonance in Medicine 2010; 64:457–71. DOI: 10.1002/mrm.22428

55. Akçakaya M, Basha TA, Goddu B, Goepfert LA, Kissinger KV, Tarokh V, Manning WJ, and Nezafat R. Low-Dimensional-Structure Self-Learning and Thresholding: Regularization Beyond Compressed Sensing for MRI Reconstruction. Magnetic Resonance in Medicine 2011; 66:756–67. DOI: `10.1002/mrm.22841`

56. Trzasko JD and Manduca A. Calibrationless Parallel MRI Using CLEAR. 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASIlOMAR) 2011. DOI: `10.1109/acssc.2011.6189958`

57. Eldar YC, Kuppinger P, and Bolcskei H. Block-Sparse Signals: Uncertainty Relations and Efficient Recovery. IEEE Transactions on Signal Processing 2010; 58:3042–54. DOI: `10.1109/tsp.2010.2044837`

58. Eldar YC and Mishali M. Robust Recovery of Signals from a Structured Union of Subspaces. IEEE Transactions on Information Theory 2009; 55:5302–16. DOI: `10.1109/tit.2009.2030471`

59. Combettes Pl and Wajs VR. Signal Recovery by Proximal Forward-Backward Splitting. Multiscale Modeling & Simulation 2005; 4:1168–200. DOI: `10.1137/050626090`

60. Wang Y, Yang J, Yin W, and Zhang Y. A New Alternating Minimization Algorithm for Total Variation Image Reconstruction. SIAM Journal on Imaging Sciences 2008; 1:248–72. DOI: `10.1137/080724265`

61. Beck A and Teboulle M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. SIAM Journal on Imaging Sciences 2009; 2:183–202. DOI: `10.1137/080716542`

62. Eldar YC and Rauhut H. Average Case Analysis of Multichannel Sparse Recovery Using Convex Relaxation. IEEE Transactions on Information Theory 2010; 56:505–19. DOI: `10.1109/tit.2009.2034789`

63. Agrawal A, Verschueren R, Diamond S, and Boyd S. A Rewriting System for Convex Optimization Problems. Journal of Control and Decision 2018; 5:42–60

64. Diamond S and Boyd S. CVXPY: A Python-Embedded Modeling Language for Convex Optimization. Journal of Machine Learning Research 2016; 17:1–5

65. MOSEK ApS. MOSEK Optimizer API for Python 10.1.17. `https://docs.mosek.com/latest/pythonapi/index.html`. 2023

66. Adcock B, Hansen AC, Poon C, and Roman B. Breaking the Coherence Barrier: A New Theory for Compressed Sensing. Forum of Mathematics, Sigma 2017; 5. DOI: `10.1017/fms.2016.32`

67. Beck A and Teboulle M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. SIAM Journal on Imaging Sciences 2009; 2:183–202