Average Case Analysis of Sparse Recovery from Combined Fusion Frame Measurements

Petros Boufounos
Mitsubishi Electric Research Laboratories
201 Broadway
Cambridge, MA 02139, USA
Email: petrosb@merl.com

Gitta Kutyniok
Institute of Mathematics
University of Osnabrück
49069 Osnabrück, Germany
Email: kutyniok@math.uos.de

Holger Rauhut
Hausdorff Center for Mathematics
University of Bonn
53012 Bonn, Germany
Email: rauhut@hcm.uni-bonn.de

***Abstract*—Sparse representations have emerged as a powerful tool in signal and information processing, culminated by the success of new acquisition and processing techniques such as Compressed Sensing (CS). Fusion frames are very rich new signal representation methods that use collections of subspaces instead of vectors to represent signals. These exciting fields have been recently combined to introduce a new sparsity model for fusion frames. Signals that are sparse under the new model can be compressively sampled and uniquely reconstructed in ways similar to sparse signals using standard CS. The combination provides a promising new set of mathematical tools and signal models useful in a variety of applications. With the new model, a sparse signal has energy in very few of the subspaces of the fusion frame, although it does not need to be sparse within each of the subspaces it occupies.**

In this paper we demonstrate that although a worst-case analysis of recovery under the new model can often be quite pessimistic, an average case analysis is not and provides significantly more insight. Using a probability model on the sparse signal we show that under very mild conditions the probability of recovery failure decays exponentially with increasing dimension of the subspaces.

Index Terms—**Compressed sensing. ℓ_1 Minimization. $\ell_{1,2}$ Minimization. Sparse Recovery. Fusion Frames. Random Matrices.**

I. Introduction

Compressed Sensing (CS) has recently emerged as a very powerful field in signal processing, enabling the acquisition of signals at rates much lower than previously thought possible [1], [2]. To achieve such performance, CS exploits the structure inherent in many naturally occurring and man-made signals. Specifically, CS uses classical signal representations and imposes a sparsity model on the signal of interest. The sparsity model, combined with randomized linear acquisition, guarantees that non-linear reconstruction can be used to efficiently and accurately recover the signal.

Fusion frames are recently emerged mathematical structures that can better capture the richness of the natural and man-made signals compared to classically used representations [3]. In particular, fusion frames generalize frame theory by using subspaces in the place of vectors as signal building blocks. Thus signals can be represented as linear combinations of components that lie in particular, and often overlapping, signal subspaces. Such a representation provides significant flexibility in representing signals of interest compared to classical frame representations.

In this paper we extend the concepts and methods of Compressed Sensing to fusion frames (for an extended version with complete proofs we refer to [4]). In doing so we demonstrate that it is possible to recover signals from underdetermined measurements if the signals lie only in very few subspaces of the fusion frame. Our generalized model does not require that the signals are sparse within each subspace. The rich structure of the fusion frame framework allows us to characterize more complicated signal models than the standard sparse or compressible signals used in compressed sensing techniques.

We provide results using an average-case approach. Two worst case analyses are contained in the aforementioned paper [4]. The average case analysis discussed in the present paper provides a framework to discern which assumptions of the worst case model become irrelevant and which are critical. It is based on the fundamental work [5], which develops an average case analysis for multichannel signals exhibiting joint sparsity patterns.

In the remainder of this section we provide the motivation behind our work and describe some possible applications. Section II provides some background on Compressed Sensing and on fusion frames to serve as a quick reference for the fundamental concepts and our basic notation. In Section III we formulate the problem and establish the additional notation necessary in our development. We further explore the connections with existing research in the field, as well as possible extensions. Section IV then presents an average case analysis of our methods which is more appropriate for typical usage scenarios. We conclude with a discussion of our results.

A. Motivation

As technology progresses, signals and computational sensing equipment becomes increasingly multidimensional. Sensors are being replaced by sensor arrays and samples are being replaced by multidimensional measurements. Yet, modern signal acquisition theory has not fully embraced the new computational sensing paradigm. Multidimensional measurements are often treated as collections of one-dimensional ones due to the mathematical simplicity of such treatment. This approach ignores the potential information and structure embedded in multidimensional signal and measurement models.

Our ultimate motivation is to provide a better understanding of more general mathematical objects, such as vector-valued data points [6]. Generalizing the notion of sparsity is part of such understanding. Towards that goal, we demonstrate that the generalization we present in this paper encompasses joint sparsity models [7], [8] as a special case. Furthermore, it is itself a special case of block-sparsity models [7]–[9], with significant additional structure that enhances existing results.

B. Applications

Although the development in this paper provides a general theoretical perspective, the principles and the methods we develop are widely applicable. In particular, the special case of joint (or simultaneous) sparsity has already been widely used in radar [10], sensor arrays [11], and MRI pulse design [12]. In these applications a mixed ℓ_1/ℓ_2 norm was used heuristically as a sparsity proxy. Part of our goals in this paper is to provide a solid theoretical understanding of such methods.

In addition, the richness of fusion frames allows the application of this work to other cases, such as target recognition and music segmentation. The goal in such applications is to identify, measure and track targets that are not well described by a single vector but by a whole subspace. In music segmentation, for example, each note is not characterized by a single frequency, but by the subspace spanned by the fundamental frequency of the instrument and its harmonics [13]. Furthermore, depending on the type of instrument in use, certain harmonics might or might not be present in the subspace. Similarly, in vehicle tracking and identification, the subspace of a vehicle's acoustic signature depends on the type of vehicle, its engine and its tires [14]. Note that in both applications, there might be some overlap in the subspaces which distinct instruments or vehicles occupy.

Fusion frames are quite suitable for such representations. The subspaces defined by each note and each instrument or each tracked vehicle generate a fusion frame for the whole space. Thus the fusion frame serves as a dictionary of targets to be acquired, tracked, and identified. The fusion frame structure further enables the use of sensor arrays to perform joint source identification and localization using far fewer measurements than a classical sampling framework.

Fusion frames and vector-based signal models also play a key role in video acquisition, reconstruction and compression applications such as [15], [16]. Nearby pixels in a video exhibit similar sparsity structure locally, but not globally. A block- or joint- sparsity model such as [7]–[9] can be very constraining in such cases. On the other hand, subspace-based models for different parts of an image significantly improve the modeling ability compared to the standard compressed sensing model.

C. Notation

Throughout this paper $\|\mathbf{x}\|_p = \left(\sum_i x_i^p\right)^{1/p},\ p > 0$ denotes the standard ℓ_p norm. The operator norm of a matrix A from ℓ_p into ℓ_p is written as $\|A\|_{p\to p} = \max_{\|x\|_p \le 1} \|Ax\|_p$.

II. Background

A. Compressed Sensing

Compressed Sensing (CS) is a recently emerged field in signal processing that enables signal acquisition using very few measurements compared to the signal dimension, as long as the signal is sparse in some basis. It predicts that a signal $\mathbf{x} \in \mathbb{R}^N$ with only k non-zero coefficients can be recovered from only $n = \mathcal{O}(k \log(N/k))$ suitably chosen linear non-adaptive measurements, compactly represented using

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad \mathbf{y} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{n\times N}.$$

A necessary condition for exact signal recovery of all k-sparse $\mathbf{x}$ is that

$$\mathbf{A}\mathbf{z} \neq 0 \quad \text{for all } \mathbf{z} \neq 0, \|\mathbf{z}\|_0 \le 2k,$$

where the ℓ_0 'norm,' $\|\mathbf{x}\|_0$, counts the number of non-zero coefficients in $\mathbf{x}$. In this case, recovery is possible using the following combinatorial optimization,

$$\widehat{\mathbf{x}} = \text{argmin}_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}.$$

Unfortunately this is an NP-hard problem [17] in general, hence is infeasible.

Exact signal recovery using computationally tractable methods even in the presence of noise can be guaranteed if, e.g., the coherence of the measurement matrix $\mathbf{A}$ is sufficiently small [6], [18] or it satisfies a restricted isometry property (RIP) [1]. A large body of literature extends these results to measurements of signals in the presence of noise, to signals that are not exactly sparse but compressible [1], to several types of measurement matrices [19]–[23] and to measurement models beyond simple sparsity [24].

B. Fusion Frames

Fusion frames are generalizations of frames that provide a richer description of signal spaces. A *fusion frame* for $\mathbb{R}^M$ is a collection of subspaces $\mathcal{W}_j \subseteq \mathbb{R}^M$ and associated weights v_j, compactly denoted using $(\mathcal{W}_j, v_j)_{j=1}^N$, that satisfies

$$A\|\mathbf{x}\|_2^2 \le \sum_{j=1}^N v_j^2 \, \|\mathbf{P}_j\mathbf{x}\|_2^2 \le B\|\mathbf{x}\|_2^2$$

for some universal fusion frame bounds $0 < A \le B < \infty$ and for all $\mathbf{x} \in \mathbb{R}^M$, where $\mathbf{P}_j$ denotes the orthogonal projection onto the subspace $\mathcal{W}_j$. We use m_j to denote the dimension of the jth subspace $\mathcal{W}_j$, $j = 1, \dots, N$. A frame is a special case of a fusion frame in which all the subspaces $\mathcal{W}_j$ are one-dimensional (i.e., $m_j = 1,\ j = 1, \dots, N$), and the weights v_j are the norms of the frame vectors.

The generalization to fusion frames allows us to capture interactions between frame vectors to form specific subspaces that are not possible in classical frame theory. Similar to classical frame theory, we call the fusion frame *tight* if the frame bounds are equal, $A = B$. If the fusion frame has $v_j = 1,\ j = 1, \dots, N$, we call it a *unit-norm* fusion frame. In this paper, we will in fact restrict to the situation of unit-norm fusion frames, since the anticipated applications are only concerned with membership in the subspaces and do not necessitate a particular weighting.

Dependent on a fusion frame $(\mathcal{W}_j, v_j)_{j=1}^N$ we define the Hilbert space $\mathcal{H}$ as

$$\mathcal{H} = \{(\mathbf{x}_j)_{j=1}^N : \mathbf{x}_j \in \mathcal{W}_j \text{ for all } j = 1, \dots, N\} \subseteq \mathbb{R}^{M \times N}.$$

Finally, let $\mathbf{U}_j \in \mathbb{R}^{M \times m_j}$ be a known but otherwise arbitrary matrix, the columns of which form an orthonormal basis for $\mathcal{W}_j$, $j = 1, \dots, N$, that is $\mathbf{U}_j^T \mathbf{U}_j = \mathbf{I}_{m_j}$, where $\mathbf{I}_{m_j}$ is the $m_j \times m_j$ identity matrix, and $\mathbf{U}_j \mathbf{U}_j^T = \mathbf{P}_j$.

The *fusion frame mixed $\ell_{q,p}$ norm* is defined as

$$\left\|(\mathbf{x}_j)_{j=1}^N\right\|_{q,p} \equiv \left(\sum_{j=1}^N (v_j \|\mathbf{x}_j\|_q)^p\right)^{1/p}, \tag{1}$$

where $(v_j)_{j=1}^N$ are the fusion frame weights. When the parameter q of the norm is omitted, it is implied to be $q = 2$:

$$\left\|(\mathbf{x}_j)_{j=1}^N\right\|_p \equiv \left(\sum_{j=1}^N (v_j \|\mathbf{x}_j\|_2)^p\right)^{1/p}.$$

Furthermore, for a sequence $\mathbf{c} = (\mathbf{c}_j)_{j=1}^N$, $\mathbf{c}_j \in \mathbf{R}^{m_j}$, we similarly define the mixed norm

$$\|\mathbf{c}\|_{2,1} = \sum_{j=1}^N \|\mathbf{c}_j\|_2.$$

The ℓ_0–'norm' (which is actually not even a quasi-norm) is defined as

$$\|\mathbf{x}\|_0 = \#\{j : \mathbf{x}_j \neq 0\}.$$

We call a vector $\mathbf{x} \in \mathcal{H}$ *k-sparse*, if $\|\mathbf{x}\|_0 \leq k$.

III. Sparse Recovery of Fusion Frame Vectors

We now consider the following scenario. Let $\mathbf{x}^0 = (\mathbf{x}_j^0)_{j=1}^N \in \mathcal{H}$, and assume that we only observe n linear combinations of those vectors, i.e., there exist some scalars a_{ij} satisfying that $\|(a_{ij})_{i=1}^n\|_2 = 1$ for all $j = 1, \dots, N$ such that we observe

$$\mathbf{y} = (\mathbf{y}_i)_{i=1}^n = \left(\sum_{j=1}^N a_{ij} \mathbf{x}_j^0\right)_{i=1}^n \in \mathcal{K}, \tag{2}$$

where $\mathcal{K}$ denotes the Hilbert space

$$\mathcal{K} = \{(\mathbf{y}_i)_{i=1}^n : \mathbf{y}_i \in \mathbf{R}^M \text{ for all } i = 1, \dots, n\}.$$

We first notice that (2) can be rewritten as

$$\mathbf{y} = \mathbf{A_I}\mathbf{x}^0, \quad \text{where } \mathbf{A_I} = (a_{ij}\mathbf{I}_M)_{1 \leq i \leq n,\, 1 \leq j \leq N},$$

i.e., $\mathbf{A_I}$ is the matrix consisting of the blocks $a_{ij}\mathbf{I}_M$.

We now wish to recover $\mathbf{x}^0$ from those measurements. If we impose conditions on the sparsity of $\mathbf{x}^0$, it is suggestive to consider the following minimization problem,

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{H}} \|\mathbf{x}\|_0 \text{ subject to} \sum_{j=1}^N a_{ij}\mathbf{x}_j = \mathbf{y}_i \forall i = 1, \dots, n.$$

Using the matrix $\mathbf{A_I}$, we can rewrite this optimization problem as

$$(P_0) \quad \hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{H}} \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{A_I}\mathbf{x} = \mathbf{y}.$$

However, this problem is NP-hard [17] and, as proposed in numerous publications initiated by [25], we prefer to employ ℓ_1 minimization techniques. This leads to the investigation of the following minimization problem,

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{H}} \|\mathbf{x}\|_1 \quad \text{subject to } \mathbf{A_I}\mathbf{x} = \mathbf{y}.$$

Since we minimize over all $\mathbf{x} = (\mathbf{x}_j)_{j=1}^N \in \mathcal{H}$ and certainly $\mathbf{P}_j\mathbf{x}_j = \mathbf{x}_j$ by definition, we can rewrite this minimization problem as

$$(\tilde{P}_1) \quad \hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{H}} \|\mathbf{x}\|_1 \quad \text{subject to } \mathbf{A_P}\mathbf{x} = \mathbf{y},$$

where

$$\mathbf{A_P} = (a_{ij}\mathbf{P}_j)_{1 \leq i \leq n,\, 1 \leq j \leq N}. \tag{3}$$

Problem $(\tilde{P}_1)$ bears difficulties to implement since minimization runs over $\mathcal{H}$. Still, it is easy to see that $(\tilde{P}_1)$ is equivalent to the optimization problem

$$(P_1) \quad (\hat{\mathbf{c}}_j)_j = \operatorname{argmin}_{\mathbf{c}_j \in \mathbb{R}^{m_j}} \|(\mathbf{U}_j\mathbf{c}_j)_{j=1}^N\|_1 \quad \text{subject to } \mathbf{A_I}(\mathbf{U}_j\mathbf{c}_j)_j = \mathbf{y}, \tag{4}$$

where then $\hat{\mathbf{x}} = (\mathbf{U}_j\hat{\mathbf{c}}_j)_{j=1}^N$. This particular form ensures that the minimizer lies in the collection of subspaces $(\mathcal{W}_j)_{j=1}^N$ while minimization is performed over $\mathbf{c}_j \in \mathbb{R}^{m_j}, j = 1, \dots, N$, hence feasible.

Finally, by rearranging (4), the optimization problems, invoking the ℓ_0-'norm' and ℓ_1-norm, can be rewritten using matrix-only notation as

$$(P_0) \quad \hat{\mathbf{c}} = \operatorname{argmin}_{\mathbf{c}} \|\mathbf{c}\|_0 \text{ subject to } \mathbf{Y} = \mathbf{A}\mathbf{U}(\mathbf{c})$$

and

$$(P_1) \quad \hat{\mathbf{c}} = \operatorname{argmin}_{\mathbf{c}} \|\mathbf{c}\|_{2,1} \text{ subject to } \mathbf{Y} = \mathbf{A}\mathbf{U}(\mathbf{c}),$$

in which

$$\mathbf{U}(\mathbf{c}) = \begin{pmatrix} \mathbf{c}_1^T\mathbf{U}_1^T \\ \hline \vdots \\ \hline \mathbf{c}_N^T\mathbf{U}_N^T \end{pmatrix} \in \mathbb{R}^{N \times M}, \; \mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \hline \vdots \\ \hline \mathbf{y}_n \end{pmatrix} \in \mathbb{R}^{n \times M},$$

where $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n \times N}$, $\mathbf{c}_j \in \mathbb{R}^{m_j}$, and $\mathbf{y}_i \in \mathbb{R}^M$. Hereby, we additionally used that $\|\mathbf{U}_j\mathbf{c}_j\|_2 = \|\mathbf{c}_j\|_2$ by orthonormality of the columns of $\mathbf{U}_j$. We follow this notation for the remainder of the paper.

A. Relation with Previous Work

A special case of the problem above appears when all subspaces $(\mathcal{W}_j)_{j=1}^N$ are equal and also equal to the ambient space $\mathcal{W}_j = \mathbb{R}^M$ for all j. Thus, $\mathbf{P}_j = \mathbf{I}_M$ and the observation setup of Eq. (2) is identical to the matrix product

$$\mathbf{Y} = \mathbf{A}\mathbf{X}, \quad \text{where } \mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \hline \vdots \\ \hline \mathbf{x}_N \end{pmatrix} \in \mathbb{R}^{N \times M}.$$

This special case is the same as the well studied joint-sparsity setup of [5], [7], [8], [26], [27] in which a collection of M sparse vectors in $\mathbb{R}^N$ is observed through the same measurement matrix $\mathbf{A}$, and the recovery assumes that all the vectors have the same sparsity structure. The use of mixed ℓ_1/ℓ_2 optimization has been proposed and widely used in this case.

Our formulation is a special case of the block sparsity problem [9], [28], [29], where we impose a particular structure on the measurement matrix $\mathbf{A}$. This relationship is already known for the joint sparsity model, which is also a special case of block sparsity. In other words, the fusion frames formulation we examine here specializes block sparsity problems and generalizes joint sparsity ones.

IV. Average Case Analysis

In this section we study the effect of the dimension of subspaces of the considered fusion frame on recoverability of a vector $\mathbf{x} \in \mathcal{H}$. Intuitively, it seems that the higher the dimension, the 'easier' the recovery via the ℓ_1 minimization problem should be. However, it turns out that this intuition only holds true if we consider an average case analysis, averaging over the vectors to be recovered. Adding dimensions can be interpreted as adding more channels to a joint sparsity problem, which is known to not improve recoverability in the worst case [5].

A. General Recovery Condition

We start our analysis by deriving a recovery condition which the reader might want to compare with [30], [31]. Given a matrix $\mathbf{X} \in \mathbb{R}^{N\times M}$, we let $\mathrm{sgn}(\mathbf{X}) \in \mathbb{R}^{N\times M}$ denote the matrix which is generated from $\mathbf{X}$ by normalizing each entry X_{ji} by the norm of the corresponding row $\mathbf{X}_{j,\cdot}$. More precisely,

$$\mathrm{sgn}(\mathbf{X})_{ji} = \begin{cases} \frac{X_{ji}}{\|\mathbf{X}_{j,\cdot}\|_2} & : \quad \|\mathbf{X}_{j,\cdot}\|_2 \neq 0, \\ 0 & : \quad \|\mathbf{X}_{j,\cdot}\|_2 = 0. \end{cases}$$

Column vectors are defined similarly by $\mathbf{X}_{\cdot,i}$.

Under a certain condition on A, which is dependent on the support of the solution, we derive the result below on unique recovery. To phrase it, let $(\mathcal{W}_j)_{j=1}^N$ be a fusion frame with associated orthogonal bases $(\mathbf{U}_j)_{j=1}^N$ and orthogonal projections $(\mathbf{P}_j)_{j=1}^N$, and recall the definition of the notion $\mathbf{U}(\mathbf{c})$ in Section III. Then, for some support set $S = \{j_1, \ldots, j_{|S|}\} \subseteq \{1, \ldots, N\}$ of $\mathbf{U}(\mathbf{c})$, we let

$$\mathbf{A}_S = (\mathbf{A}_{\cdot,j_1} \cdots \mathbf{A}_{\cdot,j_{|S|}}) \in \mathbb{R}^{n\times|S|}$$

and

$$\mathbf{U}(\mathbf{c})_S = \begin{pmatrix} \mathbf{c}_{j_1}^T \mathbf{U}_{j_1}^T \\ \hline \vdots \\ \hline \mathbf{c}_{j_{|S|}}^T \mathbf{U}_{j_{|S|}}^T \end{pmatrix} \in \mathbb{R}^{|S|\times M}.$$

Before stating the theorem, we wish to remark that its proof uses similar ideas as the analog proof in [5].

Theorem 4.1 ([4]): Retaining the notions from the beginning of this section, we let $\mathbf{c}_j \in \mathbb{R}^{m_j}$, $j = 1\ldots, N$ with $S = \mathrm{supp}(\mathbf{c}) = \{j : \mathbf{c}_j \neq 0\}$. If $\mathbf{A}_S$ is non-singular and there exists a matrix $\mathbf{H} \in \mathbb{R}^{n\times M}$ such that

$$\mathbf{A}_S^T \mathbf{H} = \mathrm{sgn}(\mathbf{U}(\mathbf{c})_S) \tag{5}$$

and

$$\|\mathbf{H}^T \mathbf{A}_{\cdot,j}\|_2 < 1 \quad \text{for all } j \notin S, \tag{6}$$

then $\mathbf{U}(\mathbf{c})$ is the unique solution of (P_1).

The matrix $\mathbf{H}$ exploited in Theorem 4.1 might be chosen as

$$\mathbf{H} = (\mathbf{A}_S^\dagger)^T \mathrm{sgn}(\mathbf{U}(\mathbf{c})_S)$$

to satisfy (5). This particular choice will in fact be instrumental for the average case result we are aiming for. For now, we obtain the following result as a corollary from Theorem 4.1.

Corollary 4.2: Retaining the notions from the beginning of this section, we let $\mathbf{c}_j \in \mathbb{R}^{m_j}$, $j = 1\ldots, N$ with $S = \mathrm{supp}(\mathbf{U}(\mathbf{c}))$. If $\mathbf{A}_S$ is non-singular and

$$\|\mathrm{sgn}(\mathbf{U}(\mathbf{c})_S)^T \mathbf{A}_S^\dagger \mathbf{A}_{\cdot,j}\|_2 < 1 \quad \text{for all } j \notin S, \tag{7}$$

then $\mathbf{U}(\mathbf{c})$ is the unique solution of (P_1).

B. Probability Model

To derive a result in the average case, we require a probability model on the k-sparse $\mathbf{U}(\mathbf{c})$, more precisely, on the associated vectors $\mathbf{c}_j \in \mathbb{R}^{m_j}$, $j = 1\ldots, N$. From now on, we assume that the dimensions of all subspaces are the same, i.e., that

$$m = m_j \quad \text{for all } j = 1\ldots, N.$$

Inspired by the probability model in [5], we will assume that on the k-element support set $S = \mathrm{supp}(\mathbf{U}(\mathbf{c})) = \{j_1, \ldots, j_k\}$ the entries of each vector $\mathbf{c}_j$ are independent and follow a normal distribution,

$$\mathbf{U}(\mathbf{c})_S = \begin{pmatrix} \mathbf{X}_1^T \mathbf{U}_{j_1}^T \\ \hline \vdots \\ \hline \mathbf{X}_k^T \mathbf{U}_{j_k}^T \end{pmatrix} \in \mathbb{R}^{k\times M}, \tag{8}$$

where $\mathbf{X} = (\mathbf{X}_1^T \ldots \mathbf{X}_k^T)^T \in \mathbb{R}^{Nm}$ is a Gaussian random vector, i.e., all entries are independent standard normal random variables.

For later use, we will introduce the matrices $\tilde{\mathbf{U}}_j \in \mathbb{R}^{M\times Nm}$ defined by

$$\tilde{\mathbf{U}}_j = (\mathbf{0}_{M\times m}|\cdots|\mathbf{0}_{M\times m}|\mathbf{U}_j|\mathbf{0}_{M\times m}|\cdots|\mathbf{0}_{M\times m}),$$

where $\mathbf{U}_j$ is the jth block. For some $\mathbf{b} = (b_1, \ldots, b_k)^T \in \mathbb{R}^k$, we can then write $\mathbf{U}(\mathbf{c})_S^T \mathbf{b}$ as

$$\sum_{\ell=1}^{k} b_\ell \tilde{\mathbf{U}}_{j_\ell} \mathbf{X} \in \mathbb{R}^M. \tag{9}$$

C. Average Case Recovery for Fusion Frames

Our main result shows that the failure probability for recovering $\mathbf{U}(\mathbf{c})$ decays exponentially fast with growing dimension m of the subspaces. Interestingly, the quantity θ involved in the estimate is dependent on the 'angles' between subspaces.

Theorem 4.3: Let $S \subseteq \{1, \ldots, N\}$ be a set of cardinality k and suppose that $\mathbf{A} \in \mathbb{R}^{n \times N}$ satisfies

$$\|\mathbf{A}_S^\dagger \mathbf{A}_{\cdot,j}\|_2 \leq \alpha < 1 \quad \text{for all } j \notin S. \tag{10}$$

Let $(\mathcal{W}_j)_{j=1}^N$ be a fusion frame with associated orthogonal bases $(\mathbf{U}_j)_{j=1}^N$ and orthogonal projections $(\mathbf{P}_j)_{j=1}^N$, and let $\mathbf{Y} \in \mathbf{R}^{n \times M}$. Further, let $\mathbf{c}_j \in \mathbb{R}^{m_j}$, $j = 1 \ldots, N$ with $S = \text{supp}(\mathbf{U}(\mathbf{c}))$ such that the coefficients on S are given by (8), and let θ be defined by

$$\theta = 1 + \max_i \sum_{j \neq i} \lambda_{\max}(\mathbf{P}_i \mathbf{P}_j)^{1/2}.$$

Choose $\delta \in (0, 1 - \alpha^2)$. Then with probability at least

$$1 - (N-k)\exp\left(-\frac{(\sqrt{1-\delta} - \alpha)^2}{2\alpha^2\theta} m\right) - k\exp\left(-\frac{\delta^2}{4} m\right)$$

the minimization problem (P_1) recovers $\mathbf{U}(\mathbf{c})$ from $\mathbf{Y} = \mathbf{A}\mathbf{U}(\mathbf{c})$. In particular, the failure probability can be estimated by

$$N \exp\left(-\left(\max_{\delta \in (0, 1-\alpha^2)} \min\left\{\frac{(\sqrt{1-\delta} - \alpha)^2}{2\alpha^2\theta}, \frac{\delta^2}{4}\right\}\right) m\right).$$

Let us note that [5] provides several mild conditions that imply (10).

The proof of the above result is developed in several steps. A key ingredient is a concentration of measure result: If f is a Lipschitz function on $\mathbb{R}^K$ with Lipschitz constant L, i.e., $|f(x) - f(y)| \leq L\|x - y\|_2$ for all $x, y \in \mathbb{R}^K$, and $\mathbf{X}$ is a K-dimensional vector of independent standard normal random variables then [32, eq. (2.35)]

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq u) \leq 2e^{-u^2/(2L^2)} \quad \text{for all } u > 0. \tag{11}$$

Our first lemma investigates the properties of a function related to (9) that are needed to apply the above inequality.

Lemma 4.4 ([4]): Let $\mathbf{b} = (b_1, \ldots, b_k)^T \in \mathbb{R}^k$ and $S = \{j_1, \ldots, j_k\} \subseteq \{1, \ldots, N\}$. Define the function f by

$$f(\mathbf{X}) = \|\sum_{\ell=1}^k b_\ell \tilde{\mathbf{U}}_{j_\ell} \mathbf{X}\|_2, \quad \mathbf{X} \in \mathbf{R}^{Nm}.$$

Then the following holds.

(i) f is Lipschitz with constant $\|\sum_{\ell=1}^k b_\ell \tilde{\mathbf{U}}_{j_\ell}\|_{2\to2}$.

(ii) For a standard Gaussian vector $\mathbf{X} \in \mathbb{R}^{Nm}$ we have $\mathbb{E}[f(\mathbf{X})] \leq \sqrt{m}\|\mathbf{b}\|_2$.

Next we estimate the Lipschitz constant of the function f in the previous lemma.

Lemma 4.5 ([4]): Let $\mathbf{b} = (b_1, \ldots, b_k)^T \in \mathbb{R}^k$ and $S = \{j_1, \ldots, j_k\} \subseteq \{1, \ldots, N\}$. Then

$$\|\sum_{\ell=1}^k b_\ell \tilde{\mathbf{U}}_{j_\ell}\|_{2\to2} \leq \|\mathbf{b}\|_\infty \sqrt{1 + \max_{i \in S} \sum_{j \in S, j \neq i} \lambda_{\max}(\mathbf{P}_i \mathbf{P}_j)^{1/2}}.$$

Now we have collected all ingredients to prove our main result.

D. Proof of Theorem 4.3

Denote $\mathbf{b}^{(j)} = (b_1^{(j)}, \ldots, b_k^{(j)})^T = \mathbf{A}_S^\dagger \mathbf{A}_{\cdot,j} \in \mathbb{R}^k$ for all $j \notin S$ and choose $\delta \in (0, 1 - \alpha^2)$. By Corollary 4.2, the probability that the minimization problem (P_1) fails to recover $\mathbf{U}(\mathbf{c})$ from $\mathbf{Y} = \mathbf{A}\mathbf{U}(\mathbf{c})$ can be estimated as

$$\begin{aligned}
&\mathbb{P}\left(\max_{j \notin S} \|\text{sgn}(\mathbf{U}(\mathbf{c})_S)^T \mathbf{b}^{(j)}\|_2 > 1\right) \\
&= \mathbb{P}\left(\max_{j \notin S} \|\sum_{\ell=1}^k b_\ell^{(j)} \|U_{j_\ell}\mathbf{X}_\ell\|_2^{-1} \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 > 1\right) \\
&\leq \mathbb{P}\left(\max_{j \notin S} \|\sum_{\ell=1}^k b_\ell^{(j)} \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 > \sqrt{(1-\delta)m}\right) \\
&\quad + \mathbb{P}\left(\max_{\ell=1,\ldots,k} \|U_{j_\ell}\mathbf{X}_\ell\|_2 < \sqrt{(1-\delta)m}\right) \\
&\leq \sum_{j \notin S} \mathbb{P}\left(\|\sum_{\ell=1}^k b_\ell^{(j)} \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 > \sqrt{(1-\delta)m}\right) \\
&\quad + \sum_{\ell=1}^k \mathbb{P}\left(\|\mathbf{X}_\ell\|_2^2 \leq (1-\delta)m\right).
\end{aligned}$$

Since X_ℓ is a standard Gaussian vector in $\mathbb{R}^m$ [33, Corollary 3] gives

$$\mathbb{P}\left(\|\mathbf{X}_\ell\|_2^2 \leq (1-\delta)m\right) \leq \exp(-\delta^2 m/4).$$

Furthermore, the concentration inequality (11) combined with Lemmas 4.4 and Lemma 4.5 yields

$$\begin{aligned}
&\mathbb{P}\left(\|\sum_{\ell=1}^k b_\ell^{(j)} \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 > \sqrt{(1-\delta)m}\right) \\
&= \mathbb{P}\left(\|\sum_{\ell=1}^k b_\ell^{(j)} \tilde{\mathbf{U}}_{j_\ell}\mathbf{X}\|_2 > \|b^{(j)}\|_2\sqrt{m}\right. \\
&\quad \left. + (\sqrt{1-\delta} - \|b^{(j)}\|_2)\sqrt{m}\right) \\
&\leq \exp\left(-\frac{(\sqrt{1-\delta} - \|b^{(j)}\|_2)^2 m}{2\|b^{(j)}\|_\infty^2 \theta}\right) \\
&\leq \exp\left(-\frac{(\sqrt{1-\delta} - \|b^{(j)}\|_2)^2 m}{2\|b^{(j)}\|_2^2 \theta}\right) \\
&\leq \exp\left(-\frac{(\sqrt{1-\delta} - \alpha)^2 m}{2\alpha^2\theta}\right).
\end{aligned}$$

Combining the above estimates yields the statement of the Theorem.

V. Conclusions and Discussion

The main contribution in this paper is the generalization of standard Compressed Sensing results for sparse signals to signals that have a sparse fusion frame representation. The key result in our work shows that the inherent structure of fusion frames provides additional information that can be exploited to derive stronger recovery results. Indeed if the signal lies in subspaces with very little overlap (i.e., where $\|\mathbf{P}_j\mathbf{P}_k\|_2$ is small), the probability of recovery increases.

Our average case analysis also demonstrates that as the sparsity structure of the problem becomes more intricate, the worst case analysis can become too pessimistic for many practical cases. The average case analysis provides reassurance that typical behavior is as expected; significantly better compared to the worst case. Our results corroborate and extend similar findings for the special case of joint sparsity in [5].

ACKNOWLEDGEMENT

The first author did part of this work while at Rice university. This part was supported by the grants NSF CCF-0431150, CCF-0728867, CNS-0435425, and CNS-0520280, DARPA/ONR N66001-08-1-2065, ONR N00014-07-1-0936, N00014-08-1-1067, N00014-08-1-1112, and N00014-08-1-1066, AFOSR FA9550-07-1-0301, ARO MURI W311NF-07-1-0185, and the Texas Instruments Leadership University Program. The author would also like acknowledge Mitsubishi Electric Research Laboratories (MERL) for their support. The second author would like to thank Peter Casazza, David Donoho, and Ali Pezeshki for inspiring discussions on ℓ_1 minimization and fusion frames. The second author would also like to thank the Department of Statistics at Stanford University and the Department of Mathematics at Yale University for their hospitality and support during her visits. This work was partially supported by Deutsche Forschungsgemeinschaft (DFG) Heisenberg Fellowship KU 1446/8-1. The third author acknowledges support by the Hausdorff Center for Mathematics and by the WWTF project SPORTS (MA 07-004).

REFERENCES

[1] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] P. G. Casazza, G. Kutyniok, and S. Li, "Fusion Frames and Distributed Processing," *Appl. Comput. Harmon. Anal.*, vol. 25, pp. 114–132, 2008.

[4] P. Boufounos, G. Kutyniok, and H. Rauhut, "Sparse Recovery from Combined Fusion Frame Measurements," 2009, preprint.

[5] Y. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inform. Theory*, to appear.

[6] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[7] M. Fornasier and H. Rauhut, "Recovery algorithms for vector valued data with joint sparsity constraints," *SIAM J. Numer. Anal.*, vol. 46, no. 2, pp. 577–613, 2008.

[8] J. A. Tropp, "Algorithms for simultaneous sparse approximation: part II: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[9] Y. C. Eldar and H. Bolcskei, "Block-sparsity: Coherence and efficient recovery," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing, 2009 (ICASSP 2009)*, April 2009, pp. 2885–2888.

[10] D. Model and M. Zibulevsky, "Signal reconstruction in sensor arrays using sparse representations," *Signal Processing*, vol. 86, no. 3, pp. 624–638, 2006.

[11] D. Malioutov, "A sparse signal reconstruction perspective for source localization with sensor arrays," Master's thesis, MIT, Cambridge, MA, July 2003.

[12] A. C. Zelinski, V. K. Goyal, E. Adalsteinsson, and L. L. Wald, "Sparsity in MRI RF excitation pulse design," in *Proc. 42nd Annual Conference on Information Sciences and Systems (CISS 2008)*, March 2008, pp. 252–257.

[13] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1808–1816, 2006.

[14] V. Cevher, R. Chellappa, and J. H. McClellan, "Vehicle Speed Estimation Using Acoustic Wave Patterns," *IEEE Trans. Signal Processing*, vol. 57, no. 1, pp. 30–47, Jan 2009.

[15] A. Veeraraghavan, D. Reddy, and R. Raskar, "Coded Strobing Photography for High Speed Periodic Events," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2009, submitted.

[16] M. S. Asif, D. Reddy, P. T. Boufounos, and A. Veeraraghavan, "Streaming Compressive Sensing for high-speed periodic videos," 2009, submitted.

[17] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, pp. 227–234, 1995.

[18] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2331–2242, 2004.

[19] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[20] H. Rauhut, "Stability results for random sampling of sparse trigonometric polynomials," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5661–5670, 2008.

[21] G. E. Pfander and H. Rauhut, "Sparsity in time-frequency representations," *J. Fourier Anal. Appl.*, to appear.

[22] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse, bandlimited signals," *IEEE Trans. Inform. Theory*, to appear.

[23] H. Rauhut, "Circulant and Toeplitz matrices in compressed sensing," in *Proc. SPARS '09*, Saint-Malo, France, 2009.

[24] R. Baraniuk, V. Cevher, M. Duarte, and C. Hedge, "Model-based compressive sensing," 2008, preprint.

[25] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, pp. 129–159, 2001.

[26] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," 2005, preprint.

[27] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 655–687, 2008.

[28] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 5302–5316, 2006.

[29] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Compressed Sensing of Block-Sparse Signals: Uncertainty Relations and Efficient Recovery," 2009, preprint.

[30] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.

[31] J. A. Tropp, "Recovery of short, complex linear combinations via l_1 minimization," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1568–1570, 2005.

[32] M. Ledoux, *The Concentration of Measure Phenomenon*. AMS, 2001.

[33] A. Barvinok, "Measure concentration," 2005, lecture notes.