

# Compressed Sensing for Fusion Frames

Petros Boufounos<sup>a</sup>, Gitta Kutyniok<sup>b</sup>, and Holger Rauhut<sup>c</sup>

<sup>a</sup>Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139 USA ;

<sup>b</sup>Institute of Mathematics, University of Osnabrück, 49069 Osnabrück, Germany;

<sup>c</sup>Hausdorff Center for Mathematics & Institute for Numerical Simulation, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany

## ABSTRACT

Compressed Sensing (CS) is a new signal acquisition technique that allows sampling of sparse signal using significantly fewer measurements than previously thought possible. On the other hand, a fusion frame is a new signal representation method that uses collections of subspaces instead of vectors to represent signals. This work combines these exciting new fields 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 needs not be sparse within each of the subspaces it occupies. We define a mixed  $\ell_1/\ell_2$  norm for fusion frames. A signal sparse in the subspaces of the fusion frame can thus be sampled using very few random projections and exactly reconstructed using a convex optimization that minimizes this mixed  $\ell_1/\ell_2$  norm. The sampling conditions we derive are very similar to the coherence and RIP conditions used in standard CS theory.

**Keywords:**  $\ell_1$  minimization. Sparse Recovery. Mutual Coherence. Fusion Frames.

## 1. INTRODUCTION AND BACKGROUND

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. To achieve such performance, CS exploits the structure inherent in many naturally occurring and man-made signals in the signal reconstruction process. Specifically, CS uses classical signal representations and imposes a sparsity model on the signal of interest. The sparsity model is combined with randomized acquisition and non-linear reconstruction.

Fusion frames are recently emerged mathematical structures that can better capture the richness of the natural and man-made signals compared to traditionally used representations. 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 compared to classical frame representations.

In this paper we extend the concepts and methods of Compressed Sensing to Fusion Frames. In doing so we demonstrate that it is possible to recover signals from underdetermined measurements if the signals lie only on very few subspaces of the fusion frame. Our generalized model does not require that the signals are sparse within each subspace.

Our ultimate motivation is to generalize the notion of sparsity to more general mathematical objects, such as vector-valued data points.<sup>4</sup> Toward that goal, we demonstrate that the generalization we present in this paper encompasses joint sparsity models<sup>15,26</sup> as a special case. Furthermore, it is itself a special case of block-sparsity models,<sup>13</sup> with significant additional structure that enhances existing results.

The rich structure of the fusion frames framework allows us to capture more complicated signal modes than the simple sparse or compressible signals captured using standard compressive sensing techniques. Our results are applicable in target detection and tracking, and in audio segmentation applications, just to name a few.

---

E-mails: petrosb@merl.com, kutyniok@uni-osnabrueck.de, rauhut@hcm.uni-bonn.de

In the remainder of this section we describe some possible applications and provide some background on Compressed Sensing and on Fusion Frames. In Section 2 we formulate the problem and establish notation. We further explore the connections with existing research in the field, as well as possible extensions. In Section 3 we prove recovery guarantees using the coherence properties of the sampling matrix. In Section 4 we prove similar guarantees using the restricted isometry properties (RIP) of the sampling matrix. We conclude with a discussion of our results.

## 1.1 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,<sup>19</sup> sensor arrays,<sup>18</sup> and MRI pulse design.<sup>28</sup> In these applications a mixed  $\ell_1/\ell_2$  norm was used heuristically as a sparsity proxy.

The richness of Fusion Frames, however, allow 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.<sup>10</sup> 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.<sup>8</sup> Note that in both applications, there might be some overlap in the subspaces that distinct instruments or vehicles occupy.

Fusion frames are quite suitable for such representations. The subspaces defined by the particular instrument or the particular 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.

## 1.2 Compressed Sensing Background

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

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

A necessary condition for exact signal recovery is that

$$\mathbf{A}\mathbf{x} \neq 0 \text{ for all } \mathbf{x} \neq 0, \|\mathbf{x}\|_0 \leq 2k,$$

where the  $\ell_0$  '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,

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

Unfortunately this is an NP-hard problem<sup>20</sup> in general, hence becomes infeasible in high dimensions.

Exact signal recovery using computationally tractable methods can be guaranteed if the coherence of the measurement matrix  $\mathbf{A}$  is sufficiently small.<sup>4,25</sup> The coherence of a matrix  $\mathbf{A}$  with unit norm columns  $a_i$ ,  $\|a_i\|_2 = 1$ , is defined as

$$\mu = \max_{i \neq j} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|,$$

Exact signal recovery is also guaranteed if  $\mathbf{A}$  obeys a restricted isometry property (RIP) of order  $2k$ , i.e., if there exists a constant  $\delta_{2k}$  such that for all  $2k$ -sparse signals  $\mathbf{x}$

$$(1 - \delta_{2k})\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_{2k})\|\mathbf{x}\|_2^2.$$

If the coherence of  $\mathbf{A}$  is low or if  $\mathbf{A}$  has a small RIP constant (see below), then the following convex optimization exactly recovers the signal from the measurement vector  $\mathbf{y}$ ,<sup>5</sup>

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}.$$

A surprising result is that random matrices with sufficient number of rows can achieve low coherence and low RIP constants with overwhelming probability.

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,<sup>5</sup> to several types of measurement matrices<sup>6, 21–23, 27</sup> and to measurement models beyond simple sparsity.<sup>2</sup>

### 1.3 Background on Fusion Frames

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

$$A\|\mathbf{f}\|_2^2 \leq \sum_{j=1}^N v_j^2 \|\mathbf{P}_j(\mathbf{f})\|_2^2 \leq B\|\mathbf{f}\|_2^2$$

for some universal fusion frame bounds  $0 < A \leq B < \infty$  and for all  $\mathbf{f} \in \mathbb{R}^M$ , where  $\mathbf{P}_j(\cdot)$  denotes orthogonal projection to the subspace  $\mathcal{W}_j$ . We use  $m_j$  to denote the dimension of the  $j$ th 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.,  $\mathbf{m}_j = 1$ ,  $j = 1, \dots, N$ ), 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.

Using the fusion frame we define the Hilbert space  $\mathcal{H}$  as

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

Finally, we use  $\mathbf{U}_j \in \mathbb{R}^{M \times m_j}$  to denote a known but otherwise arbitrary *left-orthogonal* 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

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

where  $\{v_j\}$  are the frame weights. When the  $q$  parameter of the norm is omitted, it is implied to be  $q = 2$ :

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

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  $\ell_0$ -‘norm’ (which is actually not even a quasi-norm) is defined as

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

## 2. SPARSE RECOVERY PROBLEM

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  $\|(a_{ij})_i\|_2 = 1$  for all  $j$  such that we observe

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

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

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

i.e.,  $\mathbf{A}_\mathbf{I}$  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 very suggestive to consider the following minimization problem,

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

Using the matrix  $\mathbf{A}_\mathbf{I}$ , we can rewrite these optimization problems as

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

However, this problem is NP-hard,<sup>20</sup> and therefore – as proposed in numerous publications and initiated by<sup>9</sup> – we will employ  $\ell_1$  minimization techniques. This leads to the investigation of the following minimization problem,

$$(P_1) \quad \hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{H}} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{A}_\mathbf{I} \mathbf{x} = \mathbf{y}.$$

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 wish to remark that we can rewrite this minimization problem as

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{H}} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{A}_\mathbf{P} \mathbf{x} = \mathbf{y},$$

where

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

However, problem  $(P_1)$  bears difficulties to implement, since minimization runs over  $\mathcal{H}$ . It is easy to see that in fact  $(P_1)$  is equivalent to the optimization problem

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

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  $(\mathcal{W}_i)_i$  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 can be rewritten using matrix-only notation

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

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

$$\mathbf{U}(\mathbf{c}) = \begin{pmatrix} \mathbf{c}_1^T \mathbf{U}_1^T \\ \vdots \\ \mathbf{c}_N^T \mathbf{U}_N^T \end{pmatrix} \in \mathbb{R}^{N \times M}, \quad \mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{pmatrix} \in \mathbb{R}^{n \times M}, \quad (\mathbf{A})_{ij} = a_{ij} \in \mathbb{R}^{n \times N}, \quad \mathbf{c}_j \in \mathbb{R}^{m_j}, \quad \text{and } \mathbf{y}_i \in \mathbb{R}^M.$$

Hereby, we additionally used that  $\|\mathbf{U}_j \mathbf{c}_j\|_2 = \|\mathbf{c}_j\|_2$  by (left) orthogonality of  $\mathbf{U}_j$ . We follow this notation for the remainder of the paper.

## 2.1 Extensions

Several extensions of this formulation and the work in this paper are possible. First, the analysis we provide is on the exactly sparse, noiseless case. As with classical compressed sensing, it is possible to accommodate sampling in the presence of noise. It is also natural to consider the extension of this work to sampling signals that are not  $k$ -sparse in a fusion frame representation but can be very well approximated by a  $k$ -sparse representation.

The richness of fusion frames allows us to consider richer sampling matrices. Specifically, it is possible to consider sampling matrices with matrix entries, each operating on a different subspace of the fusion frame. Such extensions open the use of  $\ell_1$  methods to general vector-valued mathematical objects and to the general problem of sampling such objects.<sup>4</sup> Of course such extensions are beyond the scope of this paper.

We should also note that in this paper we only consider a worst-case analysis. Such an analysis is quite pessimistic in practice. An average case analysis, similar to the one in,<sup>14</sup> can be more informative for several practical applications. We defer the average case analysis to a future publication.

## 2.2 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

$$\mathbf{Y} = \mathbf{A}\mathbf{X},$$

where

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{pmatrix} \in \mathbb{R}^{N \times M}.$$

This special case is the same as the well studied joint-sparsity setup of<sup>3,15,16,26</sup> 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  $\ell_1/\ell_2$  optimization has been proposed and widely used in this case.

Our formulation is also special case of the block sparsity problem,<sup>13</sup> 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. The fusion frames formulation we examine here specializes block sparsity problems and generalizes joint sparsity. As we discuss in the introduction, the fusion frames structure provides significant structure to enhance the existing block-sparsity results, especially in the form of the fusion frame coherence, which we discuss in the next section.

## 3. SPARSE RECOVERY VIA COHERENCE BOUNDS

In this section we derive conditions on  $\mathbf{c}^0$  and  $\mathbf{A}$  so that  $\mathbf{c}^0$  is the unique solution of  $(P_0)$  as well as of  $(P_1)$ . Our approach generalizes the notion of *mutual coherence*, a commonly used measure of morphological difference between the vectors of a measuring matrix. An excellent survey of the role of mutual coherence in sparse representations can be found in.<sup>4</sup>

### 3.1 Fusion Coherence

First, we require some analog of mutual coherence. In particular, we need to consider an adaptation of this notion to our more complicated situation involving the *angles* between the subspaces generated by the bases  $\mathbf{U}_j$ ,  $j = 1, \dots, N$ . In other words, here we face the problem of recovery of *vector-valued* (instead of scalar-valued) components. This leads to the following definition.

**DEFINITION 3.1.** *The fusion coherence of a matrix  $\mathbf{A} \in \mathbf{R}^{n \times N}$  with normalized ‘columns’  $(\mathbf{a}_j = \mathbf{a}_{\cdot j})_{j=1}^N$  and a collection of orthogonal projection matrices  $(\mathbf{P}_i)_{i=1}^n$  in  $\mathbf{R}^{M \times M}$  is given by*

$$\mu_f = \mu_f(\mathbf{A}, (\mathbf{P}_i)_{i=1}^n) = \max_{j \neq k} [|\langle \mathbf{a}_j, \mathbf{a}_k \rangle| \cdot \|\mathbf{P}_j \mathbf{P}_k\|_2].$$

Since the  $\mathbf{P}_j$ 's are projection matrices, we can also rewrite the definition of fusion coherence as

$$\mu_f = \max_{j \neq k} \left[ |\langle \mathbf{a}_j, \mathbf{a}_k \rangle| \cdot |\lambda_{\max}(\mathbf{P}_j \mathbf{P}_k)|^{1/2} \right]$$

with  $\lambda_{\max}$  denoting the largest eigenvalue, simply due to the fact that the eigenvalues of  $\mathbf{P}_k \mathbf{P}_j \mathbf{P}_k$  and  $\mathbf{P}_j \mathbf{P}_k$  coincide. Let us also remark that  $|\lambda_{\max}(\mathbf{P}_j \mathbf{P}_k)|^{1/2}$  equals the largest absolute value of the cosines of the principle angles between  $\mathcal{W}_j$  and  $\mathcal{W}_k$ .

### 3.2 Main Result

We first formulate the main result of this section using the new notions previously developed.

**THEOREM 3.2.** *Let  $\mathbf{A} \in \mathbf{R}^{n \times N}$  with normalized columns  $(\mathbf{a}_j)_{j=1}^N$ , let  $(\mathbf{P}_i)_{i=1}^n$  be a collection of orthogonal projection matrices in  $\mathbf{R}^{M \times M}$ , and let  $\mathbf{Y} \in \mathbf{R}^{n \times M}$ . If there exists a solution  $\mathbf{c}^0$  of the system  $\mathbf{A}\mathbf{U}(\mathbf{c}) = \mathbf{Y}$  satisfying*

$$\|\mathbf{c}^0\|_0 < \frac{1}{2}(1 + \mu_f^{-1}), \quad (5)$$

*then this solution is the unique solution of  $(P_0)$  as well as of  $(P_1)$ .*

Before we continue with the proof, let us for a moment consider the following special cases of this theorem.

#### 3.2.1 Case $M = 1$

In this case the projection matrices equal 1, and hence the problem reduces to the classical recovery problem  $\mathbf{A}\mathbf{x} = \mathbf{y}$  with  $\mathbf{x} \in \mathbf{R}^N$  and  $\mathbf{y} \in \mathbf{R}^n$ . Thus our result reduces to the result obtained in,<sup>11</sup> and the fusion coherence coincides with the commonly used mutual coherence, i.e.,  $\mu_f = \max_{j \neq k} |\langle \mathbf{a}_j, \mathbf{a}_k \rangle|$ .

#### 3.2.2 Case $\mathbf{P}_i = \mathbf{I}_M$ for all $i$

In this case we are in the general 'joint sparsity situation', where we recover a matrix  $\mathbf{X}^0 \in \mathbf{R}^{N \times M}$  from knowledge of  $\mathbf{A}\mathbf{X}^0 \in \mathbf{R}^{n \times M}$  without any constraints on the  $\mathbf{X}^0$  – before we had the constraint that  $\mathbf{X}^0$  is required to be of the form  $\mathbf{U}(\mathbf{c}^0)$ . Again fusion coherence coincides with the commonly used mutual coherence, i.e.,  $\mu_f = \max_{j \neq k} |\langle \mathbf{a}_j, \mathbf{a}_k \rangle|$ .

#### 3.2.3 Case $\mathcal{W}_i \perp \mathcal{W}_j$ for all $i, j$

Here the fusion coherence becomes 0. And this is also the correct answer, since in this case there exists precisely one solution of the system  $\mathbf{A}\mathbf{U}(\mathbf{c}) = \mathbf{Y}$  for a given  $\mathbf{Y}$ . Hence the condition (5) becomes meaningless.

#### 3.2.4 General Case

In the general case we can consider two scenarios: either we are given the subspaces  $(\mathcal{W}_i)_i$  or we are given the measuring matrix  $\mathbf{A}$ . In the first situation we face the task of choosing the measuring matrix such that  $\mu_f$  is as small as possible. Intuitively, we would choose the vectors  $(\mathbf{a}_j)_j$  so that a pair  $(\mathbf{a}_i, \mathbf{a}_j)$  has a large angle if the associated two subspaces  $(\mathcal{W}_i, \mathcal{W}_j)$  have a small angle, hence balancing the two factors and try to reduce the maximum. In the second situation, we can use a similar strategy now designing the subspaces  $(\mathcal{W}_i)_i$  accordingly.

### 3.3 Proof of Theorem 3.2

We first derive a reformulation of the equation  $\mathbf{A}\mathbf{U}(\mathbf{c}) = \mathbf{Y}$ , which will turn out to be useful in proving Theorem 3.2. Setting

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

and defining the map  $\varphi_k : \mathbf{R}^{k \times M} \rightarrow \mathbf{R}^{kM}$ ,  $k \geq 1$  by

$$\varphi_k(\mathbf{Z}) = \varphi_k \begin{pmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_k \end{pmatrix} = (\mathbf{z}_1 \dots \mathbf{z}_k)^T,$$

i.e., the concatenation of the rows. Then it is easy to see that

$$\mathbf{A}\mathbf{U}(\mathbf{c}) = \mathbf{Y} \quad \Leftrightarrow \quad \mathbf{A}_{\mathbf{P}}\varphi_N(\mathbf{U}(\mathbf{c})) = \varphi_n(\mathbf{Y}). \quad (6)$$

We now split the proof of Theorem 3.2 into two lemmata, and wish to remark that many parts are closely inspired by the techniques employed in.<sup>11</sup>

We first show that  $\mathbf{c}^0$  satisfying (5) is the *unique* solution of  $(P_1)$ .

**LEMMA 3.3.** *If there exists a solution  $\mathbf{U}(\mathbf{c}^0) \in \mathbb{R}^{N \times M}$  of the system  $\mathbf{A}\mathbf{U}(\mathbf{c}) = \mathbf{Y}$  with  $\mathbf{c}^0$  satisfying (5), then  $\mathbf{c}^0$  is the unique solution of  $(P_1)$ .*

*Proof.* In the following we will refer to

$$\text{supp}(\mathbf{U}(\mathbf{c}^0)) = \{j \in \{1, \dots, N\} : (\mathbf{c}_j^0)^T \mathbf{U}_j^T \neq 0\}$$

as the *support* of  $\mathbf{U}(\mathbf{c}^0)$  (as well as of  $\mathbf{c}^0$ ) and denote it by  $S$ . Let  $\mathbf{c}^1$  be an arbitrary solution of the system  $\mathbf{A}\mathbf{U}(\mathbf{c}) = \mathbf{Y}$ , and set

$$\mathbf{h} = \mathbf{c}^0 - \mathbf{c}^1.$$

Denoting by a subscript  $S$  a vector restricted to the set  $S$ , we obtain

$$\|\mathbf{c}^0\|_{2,1} - \|\mathbf{c}^1\|_{2,1} = \|\mathbf{c}_{S^c}^1\|_{2,1} + \|\mathbf{c}_S^1\|_{2,1} - \|\mathbf{c}_S^0\|_{2,1} \geq \|\mathbf{h}_{S^c}\|_{2,1} - \|\mathbf{h}_S\|_{2,1}.$$

We have to show that this term is greater than zero for any  $\mathbf{h} \neq 0$ , which is the case provided that

$$\|\mathbf{h}_{S^c}\|_{2,1} > \|\mathbf{h}_S\|_{2,1} \quad (7)$$

or, in other words,

$$\frac{1}{2}\|\mathbf{h}\|_{2,1} > \|\mathbf{h}_S\|_{2,1}, \quad (8)$$

which we next aim to prove.

Since  $\mathbf{h}$  satisfies  $\mathbf{A}\mathbf{U}(\mathbf{h}) = 0$ , by using the reformulation (6), it follows that

$$\mathbf{A}_{\mathbf{P}}\varphi_N(\mathbf{U}(\mathbf{h})) = 0.$$

This implies that

$$\mathbf{A}_{\mathbf{P}}^* \mathbf{A}_{\mathbf{P}} \varphi_N(\mathbf{U}(\mathbf{h})) = 0.$$

Defining  $\mathbf{a}_j$  by  $\mathbf{a}_j = (\mathbf{a}_{ij})_i$  for each  $j$ , the previous equality can be computed to be

$$\langle \mathbf{a}_i, \mathbf{a}_j \rangle \mathbf{P}_i \mathbf{P}_j \varphi_N(\mathbf{U}(\mathbf{h})) = 0.$$

Recall that we have required the vectors  $\mathbf{a}_j$  to be normalized. Hence, for each  $i$ ,

$$\mathbf{U}_i \mathbf{h}_i = - \sum_{j \neq i} \langle \mathbf{a}_i, \mathbf{a}_j \rangle \mathbf{P}_i \mathbf{P}_j \mathbf{U}_j \mathbf{h}_j.$$

Since  $\|\mathbf{U}_i \mathbf{h}_i\|_2 = \|\mathbf{h}_i\|_2$  for any  $i$ , this gives

$$\|\mathbf{h}_i\|_2 \leq \sum_{j \neq i} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle| \cdot \|\mathbf{P}_i \mathbf{P}_j\|_2 \|\mathbf{h}_j\|_2 \leq \mu_f (\|\mathbf{h}\|_{2,1} - \|\mathbf{h}_i\|_2),$$

which implies

$$\|\mathbf{h}_i\|_2 \leq (1 + \mu_f^{-1})^{-1} \|\mathbf{h}\|_{2,1}.$$

Thus, we have

$$\|\mathbf{h}_S\|_{2,1} \leq \#(S) \cdot (1 + \mu_f^{-1})^{-1} \|\mathbf{h}\|_{2,1} = \|\mathbf{c}^0\|_0 \cdot (1 + \mu_f^{-1})^{-1} \|\mathbf{h}\|_{2,1}.$$

Concluding, (5) and (8) show that  $\mathbf{h}$  satisfies (8) unless  $\mathbf{h} = 0$ , which implies that  $\mathbf{c}^0$  is the unique minimizer of  $(P_1)$  as claimed.  $\square$

Using Lemma 3.3 it is easy to show the following lemma. We omit the proof in the interest of space.

**LEMMA 3.4.** *If there exists a solution  $\mathbf{U}(\mathbf{c}^0) \in \mathbb{R}^{N \times M}$  of the system  $\mathbf{A}\mathbf{U}(\mathbf{c}) = \mathbf{Y}$  with  $\mathbf{c}^0$  satisfying (5), then  $\mathbf{c}^0$  is the unique solution of  $(P_0)$ .*

We observe that Theorem 3.2 now follows immediately from Lemmata 3.4 and 3.3.

#### 4. SPARSE RECOVERY USING THE RESTRICTED ISOMETRY PROPERTY (RIP)

In this section we consider an alternative condition for sparse recovery using the restricted isometry property (RIP) of the sampling matrix. This property on the sampling matrix, first introduced in,<sup>5</sup> complements the mutual coherence conditions and is often preferred in the literature.

**DEFINITION 4.1.** *A matrix  $\mathbf{A} \in \mathbb{R}^{n \times N}$  satisfies a restricted isometry property (RIP) of order  $k$  if there exists constant  $\delta_k$  such that for all  $\mathbf{x} \in \mathbb{R}^N$ ,  $\|\mathbf{x}\|_0 \leq k$*

$$(1 - \delta_k) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_k) \|\mathbf{x}\|_2^2.$$

Note that this is the classical definition of the RIP, used in the classical CS literature to measure a single sparse vector. However, we use that property to recover signals that have a sparse fusion frame representation. Specifically, in the remainder of this section we demonstrate that if our matrix  $\mathbf{A}$  satisfies the RIP, we can recover a  $k$ -sparse signal  $\mathbf{X} \in \mathcal{H}$  from its fusion frame measurements  $\mathbf{Y} = \mathbf{A}\mathbf{X}$  by solving the  $\ell_1$  minimization  $(P_1)$ . Notice that we follow the line of proof from,<sup>5</sup> but are required to make changes to adapt it to our situation.

We first show that imposing the RIP condition on the matrix  $\mathbf{A}$  implies a form of RIP we require for our modified situation.

**LEMMA 4.2.** *Suppose that the matrix  $\mathbf{A} \in \mathbb{R}^{n \times N}$  satisfies the RIP of order  $k$  with constant  $\delta_k$ . Then*

$$(1 - \delta_k) \|\mathbf{c}\|_{2,2}^2 \leq \|\mathbf{A}\mathbf{U}(\mathbf{c})\|_{2,2}^2 \leq (1 + \delta_k) \|\mathbf{c}\|_{2,2}^2$$

for all  $\mathbf{c}$  satisfying  $\|\mathbf{c}\|_0 \leq k$ .

**THEOREM 4.3.** *Let  $\mathbf{A} \in \mathbb{R}^{n \times N}$  with normalized columns  $(\mathbf{a}_j)_{j=1}^N$ , and suppose that it satisfies  $k$ -RIP with  $k$  such that  $\delta_{3k} + 3\delta_{4k} < 2$ . Let  $(\mathbf{P}_i)_{i=1}^n$  be a collection of projection matrices in  $\mathbb{R}^{M \times M}$ , and let  $\mathbf{Y} \in \mathbb{R}^{n \times M}$ . If there exists a solution  $\mathbf{c}^0$  of the system  $\mathbf{A}\mathbf{U}(\mathbf{c}) = \mathbf{Y}$ , then this solution is both the unique solution of  $(P_0)$  and  $(P_1)$ .*

*Proof.* Let the support of  $\mathbf{U}(\mathbf{c}^0)$  be denoted by  $\mathcal{S}$ , set  $|\mathcal{S}| = k$ , and define  $\mathbf{h} := \mathbf{c}^1 - \mathbf{c}^0$ , where  $\mathbf{c}^1$  is a solution to the system. We then partition the indexing set  $\{1, \dots, N\}$  into subsets  $\mathcal{S}_\ell$ ,  $\ell = 1, \dots, \rho^{-1} \cdot (N/k - 1)$  for some  $\rho$  to be chosen later such that

- (i)  $|\mathcal{S}_\ell| = \rho \cdot k$  for all  $\ell$ ,
- (ii)  $\{1, \dots, N\} = \mathcal{S} \cup \bigcup_\ell \mathcal{S}_\ell$ ,
- (iii)  $\|\mathbf{h}_{j_1}\|_2 \geq \|\mathbf{h}_{j_2}\|_2$  if and only if  $j_1 \in \mathcal{S}_{\ell_1}$  and  $j_2 \in \mathcal{S}_{\ell_2}$  with  $\ell_1 \leq \ell_2$ .



We would like the reader to note that this selection according to the norms of the elements  $\mathbf{h}_j$  of  $\mathbf{h} = (\mathbf{h}_j)_j$  belonging to the different fusion frame subspaces differs from,<sup>5</sup> since here we are concerned with sparsity of objects consisting of vectors.

With this partition at hand we now proceed along the lines of<sup>5</sup> as follows: Defining  $\tilde{\mathcal{S}} := \mathcal{S} \cup \mathcal{S}_1$ , noting that the  $j$  largest value of  $(\|\mathbf{h}_j\|_2^2)_{j \in \mathcal{S}^c}$  is  $\leq \|1_{\mathcal{S}^c} \mathbf{h}\|_1^2 / j$ , and recalling (7), we obtain

$$\|1_{\tilde{\mathcal{S}}^c} \mathbf{h}\|_2^2 \leq \|1_{\mathcal{S}^c} \mathbf{h}\|_1^2 \sum_{j=\rho \cdot k+1}^N 1/j^2 \leq \|1_{\mathcal{S}^c} \mathbf{h}\|_1^2 / (\rho \cdot k) \leq \|1_{\mathcal{S}} \mathbf{h}\|_1^2 / (\rho \cdot k) \leq \|1_{\mathcal{S}} \mathbf{h}\|_2^2 / \rho.$$

This implies

$$\|\mathbf{h}\|_2^2 \leq (1 + \rho) \|1_{\tilde{\mathcal{S}}} \mathbf{h}\|_2^2. \quad (9)$$

Next, observing that for  $j \in \mathcal{S}_{\ell+1}$  we have  $\|\mathbf{h}_j\|_2^2 \leq \|1_{\mathcal{S}_\ell} \mathbf{h}\|_1 \cdot (\rho/k)$ , it follows that

$$\sum_{\ell \geq 2} \|1_{\mathcal{S}_\ell} \mathbf{h}\|_2 \leq \sum_{\ell \geq 1} \|1_{\mathcal{S}_\ell} \mathbf{h}\|_1 \cdot \sqrt{\rho/k} \leq \|1_{\mathcal{S}} \mathbf{h}\|_1 \cdot \sqrt{\rho/k} \leq \sqrt{\rho} \cdot \|1_{\mathcal{S}} \mathbf{h}\|_2. \quad (10)$$

Now, we use the RIP-property and (10) to obtain

$$\begin{aligned} \|\mathbf{A}\mathbf{U}(\mathbf{h})\|_2 &= \|\mathbf{A}\mathbf{U}(1_{\tilde{\mathcal{S}}}\mathbf{h}) + \sum_{\ell \geq 2} \mathbf{A}\mathbf{U}(1_{\mathcal{S}_\ell}\mathbf{h})\|_2 \\ &\geq \|\mathbf{A}\mathbf{U}(1_{\tilde{\mathcal{S}}}\mathbf{h})\|_2 - \sum_{\ell \geq 2} \|\mathbf{A}\mathbf{U}(1_{\mathcal{S}_\ell}\mathbf{h})\|_2 \\ &\geq \sqrt{1 - \delta_{(1+1/\rho)k}} \|1_{\tilde{\mathcal{S}}}\mathbf{h}\|_2 - \sqrt{1 + \delta_{(k/\rho)}} \sum_{\ell \geq 2} \|1_{\mathcal{S}_\ell}\mathbf{h}\|_2 \\ &\geq \left( \sqrt{1 - \delta_{(1+1/\rho)k}} - \sqrt{\rho \cdot (1 + \delta_{(k/\rho)})} \right) \|1_{\tilde{\mathcal{S}}}\mathbf{h}\|_2. \end{aligned} \quad (11)$$

Setting  $C_{\rho,k} = \sqrt{1 - \delta_{(1+1/\rho)k}} - \sqrt{\rho \cdot (1 + \delta_{(k/\rho)})}$ , applying (9), and the fact that

$$\|\mathbf{A}\mathbf{U}(\mathbf{h})\|_2 \leq \|\mathbf{A}\mathbf{U}(\mathbf{c}^1) - \mathbf{Y}\|_2 + \|\mathbf{A}\mathbf{U}(\mathbf{c}^0) - \mathbf{Y}\|_2 = 0,$$

yields

$$\|\mathbf{h}\|_2 = \sqrt{1 + \rho} \|1_{\tilde{\mathcal{S}}}\mathbf{h}\|_2 \leq (\sqrt{1 + \rho}) / C_{\rho,k} \cdot \|\mathbf{A}\mathbf{U}(\mathbf{h})\|_2 = 0.$$

The proof is finished by noticing that we require  $C_{\rho,k} > 0$  for the last argument to be valid, and choosing  $\rho = 1/3$ .  $\square$

Of course, it is possible to extend this proof in a similar manner to<sup>5</sup> such that we can accommodate measurement noise and signals that are well approximated by sparse fusion frame representation. We defer this generalization to a future paper.

## 5. CONCLUSIONS AND DISCUSSION

Need to fill this in.

Main conclusions:

- FFs are very powerful in describing sparsity and structure.
- Standard CS results extend to FFs in very nice way
- At the current state mutual coherence provides a better understanding in terms of the FF properties. Incorporating the projections into the RIP would be desirable
- Our analysis is worst case. The average case provides significant more insight, but we defer that for a future publication.

## Acknowledgements

The second author would like to thank Peter Casazza, David Donoho, and Ali Pezeshki for inspiring discussions on  $\ell_1$  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] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin *A Simple Proof of the Restricted Isometry Property of Random Matrices*, *Constr. Approx.* **28(3)** (2007), 253–263.
- [2] R.G. Baraniuk, V. Cevher, M. Duarte, and C. Hedge. Model-based compressive sensing. *preprint*, 2008.
- [3] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk. Distributed compressed sensing. *preprint*, 2005.
- [4] 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*, to appear.
- [5] E. J. Candès, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, *Comm. Pure Appl. Math.* **59(8)** (2006), 1207–1223.
- [6] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406–5425, 2006.
- [7] P. G. Casazza, G. Kutyniok, and S. Li, *Fusion Frames and Distributed Processing*, *Appl. Comput. Harmon. Anal.* **25** (2008), 114–132.
- [8] V. Cevher, R. Chellappa, and J. H. McClellan, “Vehicle Speed Estimation Using Acoustic Wave Patterns,” *IEEE Trans. Signal Processing*, 57(1):30–47, Jan. 2009.
- [9] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, *SIAM Rev.* **43** (2001), 129–159.
- [10] L. Daudet, “Sparse and structured decompositions of signals with the molecular matching pursuit,” *IEEE Trans. Audio, Speech, and Language Processing*, 14(5):1808–1816, Sept. 2006.
- [11] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell^1$  minimization*, *Proc. Natl. Acad. Sci. USA* **100(5)** (2003), 2197–2202.
- [12] D. L. Donoho, M. Elad, and V. N. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, *IEEE Trans. Inform. Theory* **52(1)** (2006), 6–18.
- [13] Y. C. Eldar and H. Bölcskei. Block-sparsity: Coherence and efficient recovery. to appear in *ICASSP09*.
- [14] Y. C. Eldar and H. Rauhut. Average case analysis of multichannel sparse recovery using convex relaxation. *preprint*, 2009.
- [15] M. Fornasier and H. Rauhut. Recovery algorithms for vector valued data with joint sparsity constraints. *SIAM J. Numer. Anal.*, 46(2):577–613, 2008.
- [16] 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.*, 14(5):655–687, 2008.
- [17] G. Kutyniok, A. Pezeshki, A. R. Calderbank, and T. Liu, *Robust Dimension Reduction, Fusion Frames, and Grassmannian Packings*, *Appl. Comput. Harmon. Anal.*, to appear.
- [18] D. Malioutov, *A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays*, MIT Master Thesis, Cambridge, MA July 2003.
- [19] D. Model and M. Zibulevsky, *Signal reconstruction in sensor arrays using sparse representations*, *Signal Processing* **86** (2006), 624–638.
- [20] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24:227–234, 1995.
- [21] G. E. Pfander and H. Rauhut. Sparsity in time-frequency representations. *preprint*, 2007.
- [22] H. Rauhut. Stability results for random sampling of sparse trigonometric polynomials. *IEEE Trans. Information Theory*, 54(12):5661–5670, 2008.

- [23] H. Rauhut. Circulant and Toeplitz matrices in compressed sensing. In *Proc. SPARS'09*, Saint-Malo, France, 2009.
- [24] T. Strohmer and R. W. Heath, Jr., *Grassmannian frames with applications to coding and communication*, Appl. Comput. Harmon. Anal. **14** (2003), 257–275.
- [25] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50(10):2231–2242, 2004.
- [26] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation: part II: Convex relaxation. *Signal Processing*, 86(3):589 – 602, 2006.
- [27] Joel Tropp, J.N. Laska, M. Duarte, J. Romberg, and R.G. Baraniuk. Beyond Nyquist: Efficient sampling of sparse, bandlimited signals. *preprint*, 2009.
- [28] A. C. Zelinski, V. K. Goyal, E. Adalsteinsson, and L. L. Wald, “Sparsity in MRI RF excitation pulse design,” *42nd Annual Conference on Information Sciences and Systems 2008 (CISS 2008)* pp.252–257, 19-21 March 2008