LOCAL SPARSITY AND RECOVERY OF FUSION FRAME STRUCTURED SIGNALS

ROZA ACESKA, JEAN-LUC BOUCHOT, AND SHIDONG LI

ABSTRACT. The problem of recovering signals of high complexity from low quality sensing devices is analyzed via a combination of tools from signal processing and harmonic analysis. By using the rich structure offered by the recent development in fusion frames, we introduce a compressed sensing framework in which we split the dense information into sub-channel or local pieces and then fuse the local estimations. Each piece of information is measured by potentially low quality sensors, modeled by linear matrices and recovered via compressed sensing – when necessary. Finally, by a fusion process within the fusion frames, we are able to recover accurately the original signal.

Using our new method, we show, and illustrate on simple numerical examples, that it is possible, and sometimes necessary, to split a signal via local projections and / or filtering for accurate, stable, and robust estimation. In particular, we show that by increasing the size of the fusion frame, a certain robustness to noise can also be achieved. While the computational complexity remains relatively low, we achieve stronger recovery performance compared to usual single-device compressed sensing systems.

1. PROBLEM STATEMENT

In a traditional sampling and reconstruction system, the sensors are designed so that the recovery of the signal(s) of interest is possible. For instance, when considering the sparse recovery problem, one tries to find the sparsest solution $\hat{\mathbf{x}} \in \mathbb{K}^N$ from the noisy measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \in \mathbb{K}^m$ and $m \ll N$. Here $\mathbb{K}$ denotes the field $\mathbb{R}$ or $\mathbb{C}$. This is done by solving the mathematical program

$$\hat{\mathbf{x}} := \operatorname{argmin} \|\mathbf{z}\|_0, \quad \text{subject to } \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta. \tag{ℓ^0-min}$$

This problem is NP-Hard and usually only approximately solved, for instance by solving its convex relaxation, known as the Basis Pursuit denoising

$$\hat{\mathbf{x}} := \operatorname{argmin} \|\mathbf{z}\|_1, \quad \text{subject to } \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta. \tag{BPDN}$$

It is known that for a given *complexity* (usually measured as the sparsity) s of the signal $\mathbf{x}$, the number of random subgaussian linear measurements needs to grow as $m \gtrsim s \log(N/s)$ for $\hat{\mathbf{x}}$ to be a good enough approximation to $\mathbf{x}$. Said differently, if the design of a sensor can be made at will, then knowing the complexity of the signal, here characterized by the sparsity, is sufficient for a stable and robust recovery. This paper looks at the problem of sampling potentially highly non-sparse signals when the quality of the sensors is constrained. We emphasize in passing that throughout this paper, the sought after signal will always be considered of high-complexity. We use the sparsity or density as a measure of complexity, but one could consider other models. We investigate problems where the number of measurements m cannot be chosen based on the complexity of the signals to recover. In the context described above, one would have a limit on the sparsity of the vectors that can be recovered by $s \lesssim m/\log(N/m)$. These constraints can be due to many reasons such as cost – e.g. using 10 sensors at a coarser resolution is cheaper than one at the finest –, frequency rate – sensors at 2000 THz might not exists for a while –, legal regulation – e.g. in nuclear medicine where one should not expose a patient to too high radiations at once. Problems arise when the signals being sampled are *too dense* for the usual mathematical theories. When one thinks about compressed sensing, the size of the sensor required is driven by a certain measure of complexity of the signals considered. Allowing for the recovery of signals with higher level of complexity entails the use of better sensors. Here we look at the problem differently: first, we assume constraints on the sensor design which are fixed due to some outside reasons. Under these assumptions, we take on the following challenge: split the information carried by the signal in a clever way so that a mathematical recovery is possible.

This paper revisits the theory of fusion frames and applies it to the dense signal recovery problem. We show that by using advanced mathematical techniques stemming from applied harmonic analysis, it is possible to handle very high complexity signals in an efficient and stable manner. Before we dig into

---

the more technical details, we present some real-world scenarios where our framework appears useful, if not essential.

1.1. **Examples.**

1.1.1. *Unavailability of high quality observation devices.* A typical time-invariant bounded linear operator is always represented by a circulant matrix $\mathbf{A}$. So suppose $\mathbf{A}$ represents a sensing device whose number of rows m is physically limited by the sampling rate (or resolution) of the device. Suppose also that the sparsity of the sampled signal $\mathbf{x}$ is substantially larger than what a single observation by $\mathbf{A}$ could handle/recover by various compressed sensing techniques.

In this context, the limitations on the sensing devices combined with the (potentially) high number of non-zeros in the signals makes it impossible for a state-of-the-art algorithm to recover the unknown signal $\mathbf{x}$. As illustrated in Figure 1, we suggest to apply n such devices in parallel after prefiltering. The fused compressed sensing technique introduced later allows to resolve the problem that otherwise a single device can not! Such scenarios actually exist and show the necessity of the fused compressed sensing technique presented below.

For example, suppose an application requires a sensing device of capacity X, described by a sensing matrix $\tilde{\mathbf{A}}$. In case such a device is either very expensive, or not available, we may choose to combine n parallel projections $\{P_j\}_{j=1}^n$ prior to measuring, and use n low-quality sensing devices of capacity $\frac{1}{n}X$, each described by the sensing matrix $\mathbf{A}$. The (sparse) signal $\mathbf{x}$ is then subsequently recovered by various techniques via each channel and, through the theory of fusion frames [13, 14, 11], merged into a single vector. As long as $\{P_j\}_{j=1}^n$ are projections - or any filtering operations - with the property that $C\mathbf{I} \le \sum_i P_j^* P_j \le D\mathbf{I}$ for some $0 < C \le D < \infty$, such a fusion operation is always possible.
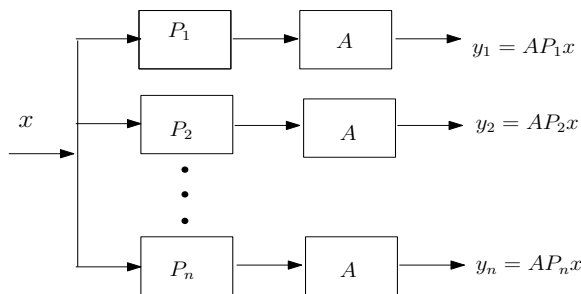

FIGURE 1. Sparse linear array geometry

In order for the fusion operator to be bounded away from 0, in our work we assume that the pre-filtering/projections P_i are *complete* in the sense that there won't be "holes" in the signal coverage for a whole class of signals; more precisely, we require that the collection of projections satisfy the fusion frame inequality (6).

Evidently, if we design such a fused compressed sensing technique, by enabling the subdivision of sparsity of $\mathbf{x}$ into individual subspaces, the sparse recovery problem becomes a feasible one and can be resolved by multiple sensing devices with low quality or resolution, which are not only widely available but also economical.

The value of fused compressed sensing techniques presented in this paper is clearly reflected in this situation where an otherwise too expensive or impossible problem can now be resolved by using a number of lower resolution/sampling rate devices and by making a *reasonable* number of observations, and processed by the fusion frame theory.

1.1.2. *SAR imaging and spatial filtering.* In the Synthetic Aperture Radar (SAR) imaging process, a flying carrier (an airplane or a satellite) emits a sequence of radio waves to the field of observation (and then detects the reflections by the objects in the field). Each radio wave is sent through an antenna with a fixed aperture/size which physically presents an antenna beam (magnetic field) pattern, say $\{F_i\}$.

Consequently, as seen in Figure 2, during the ith data collection, the antenna beam pattern F_i physi-cally implements a spatial filtering operator. And, naturally, between adjacent i's (even among several adjacent i's), the spatial antenna beams F_i's have intentional overlaps. The actual data observation, for each i, is modeled by $\mathbf{y}^{(i)} = \mathbf{A}F_i\mathbf{x}$, where $\mathbf{x}$ is the field image refection coefficients that eventually form the SAR image of the field, and $\mathbf{A}$ is a fixed observation matrix determined by the SAR imaging mechanism, which is typically chirping (linearly changing frequencies in time) functions in two different
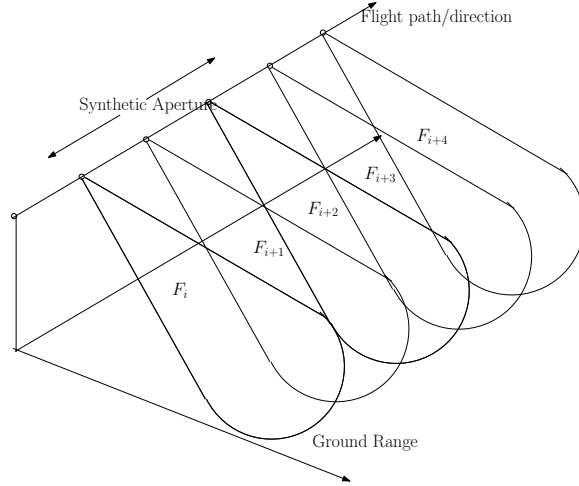
FIGURE 2. Synthetic Aperture Imaging beam and data collection mechanism

dimensions [19, 25, 18]. This formulation has a natural "distributed sparsity" due to the spatial filtering operations inherent to the physical beams.

Here, to consider the spatial filtering effect, and to avoid the conventional ad-hoc "alignment" in the flight direction, the fused SAR imaging process needs to be considered [35]. The new fused SAR imaging formulation is not only the should-be rigorous mathematical formulation of the SAR imaging process, but also potentially beneficial to image resolution and to the robustness of the (often) turbulent data collection process.

1.2. **Contributions.** In this paper we combine mathematical tools from compressed sensing and fusion frame theory to break the limitations on the signal complexity induced by traditional recovery methods.

Our general approach can be described as the two following steps, or summarized in Algorithm 1.

1) Estimate the local information $\widehat{\mathbf{x}^{(i)}}$, for $1 \leq i \leq n$ by any sparse recovery method, and
2) Approximate the fused solution $\hat{\mathbf{x}} = S^{-1} \left(\sum_{i=1}^{n} \widehat{\mathbf{x}^{(i)}} \right)$, either directly or by virtue of the frame algorithm.

Data: $(P_i)_{1 \leq i \leq n}$, n such that $C\mathbf{I} \leq \sum |P_i|_{2 \to 2}^2 \leq D\mathbf{I}$, an estimation of the measurement noise η
Result: An estimation $\hat{\mathbf{x}}$ of the vector $\mathbf{x}$
$k \leftarrow 1$;
$\widehat{\mathbf{x}} = 0$;
while $k \leq n$ **do**
    Measure k^{th} vector $\mathbf{y} \leftarrow \mathbf{A}P_k\mathbf{x} + \mathbf{e}^{(k)}$;
    Add local information $\widehat{\mathbf{x}} \leftarrow \widehat{\mathbf{x}} + \text{argmin} \|\mathbf{z}\|_1$ s.t. $\|\mathbf{A}P_k\mathbf{z} - \mathbf{y}^{(k)}\|_2 \leq \eta$;
    $k \leftarrow k + 1$
end
Fusion: $\widehat{\mathbf{x}} \leftarrow S^{-1}\widehat{\mathbf{x}}$
Algorithm 1: Fused distributed sensing recovery framework

Note that the fusion operation either requires the inverse frame operator S^{-1} or can be approximated using the (fusion-) frame algorithm, described at the end of Subsection 2.2. An important aspect of the framework is that it can either be used sequentially or in a parallel manner. In the sequential approach, we get the local pieces of information one after the other. In this case, as described in Algorithm 1, we have no need to save the local measurements and can update our guess in an online fashion. In the parallel approach, the local measurements are processed independently of each other, and the fusion is done by a central unit, once all the local information have been collected.

In particular, Proposition 4 later in the manuscript shows that the robustness is preserved independently from the number of subspaces considered, which can be phrased in simple terms as:

Proposition 1. *Let* $\mathbf{A} \in \mathbb{K}^{m \times N}$, *and let* $\mathcal{W} = (W_i, P_i)_{i=1}^{n}$, $n \geq 1$ *be a fusion frame. Assume the measurement vectors* $\mathbf{y}^{(i)} = \mathbf{A}P_i\mathbf{x} + \mathbf{e}^{(i)}$ *are corrupted by some (adversarial) independent noise uniformly bounded* $\|\mathbf{e}^{(i)}\|_2 \leq \eta$. *Then our solution* $\widehat{\mathbf{x}}$ *satisfies*

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2 \leq K\eta^2. \tag{1}$$

In the more precise statement (Proposition 4), we specify precisely the value of the constant $K = 1/C$ where C denotes the lower frame bound.

As a special case of Theorem 9, we proved the following

Theorem 1. *Let $\mathcal{W} = (W_i, P_i)_{i=1}^{n}$ be a fusion frame system for $\mathbb{K}^N$ with frame operator S and frame bounds $0 < C \leq D < \infty$. Let $\mathbf{A} \in \mathbb{K}^{m \times N}$ be a matrix satisfying some Partial RIP condition (see Definition 6). Any s distributed-sparse vector $\mathbf{x}$ whose sparsity is uniformly distributed among the n subspaces (i.e. $s_i = s/n$ is the sparsity of $P_i\mathbf{x}$) can be recovered as*

$$\widehat{\mathbf{x}} = S^{-1}\left(\sum_{i=1}^{n} \widehat{\mathbf{x}^{(i)}}\right),$$

where the $\widehat{\mathbf{x}^{(i)}}$ are obtained as solutions to the n local sparse recovery problems

$$\widehat{\mathbf{x}^{(i)}} := \operatorname{argmin} \|\mathbf{z}\|_0, \quad \text{subject to } \|\mathbf{A}P_i\mathbf{z} - \mathbf{y}^{(i)}\|_2 \leq \eta_i.$$

Moreover, assuming the errors to be uniformly bounded by η, and that the RIP constants $\delta_1 = \delta_2 = \cdots = \delta_n = \delta < 4/\sqrt{41}$ are the same in all the subspaces, the solution approximate the true vector $\mathbf{x}$ in the following sense:

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{n}{C}\mathcal{C}\eta$$

where $\eta := \max_i \eta_i$ and $\mathcal{C} \leq \frac{960\sqrt{2}}{(16 - 41\delta^2)^2}$ is a constant depending only on δ.

Note that under the usual compressed sensing assumptions, this allows to have a number of measurements per sampled vector that decay linearly with the number of sensors.

The paper is articulated as follows. We review basics from compressed sensing and fusion frames in Section 2. In Section 3 we explore several theoretical examples to motivate our research, and to emphasize the importance of our findings. Our claims are empirically verified by numerical results. In the flair of traditional compressed sensing, we extend the standard compressed sensing results to the case of recovery where we make explicit use of local redundancy in the fusion frame decomposition in Section 5. Finally, Section 4 derives a mathematical theory allowing to work with fusion frames where sparsity is exploited along a subspace decomposition. These findings extend the traditional compressed sensing problem [24] and show some similarities with recent developments in parallel acquisition [17] and structured sensing [10].

2. General tools and models

2.1. **(Traditional) Compressed sensing.** Compressed sensing (CS) relies (see [24, 22] and references therein) on the inherent sparsity of natural signals $\mathbf{x}$ for their recovery from seemingly few measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for some linear measurement matrix $\mathbf{A} \in \mathbb{K}^{m \times N}$, with $m \ll N$. Here the vector $\mathbf{e} \in \mathbb{K}^m$ contains the noise and is usually assumed to be bounded, $\|\mathbf{e}\|_2 \leq \eta$. With the sparsity assumption, CS aims at finding (approximate) solutions to the (ℓ^0-min) problem. General recovery guarantees ensure that the recovery is stable and robust, that is, the solution satisfies the following Lebesgue-type inequalities

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(\mathbf{x})_1 + D\eta, \tag{2}$$

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_1 \leq C\sigma_s(\mathbf{x})_1 + D\sqrt{s}\eta, \tag{3}$$

where $\sigma_s(\mathbf{x})_1 := \operatorname{argmin}_{\mathbf{z}:\|\mathbf{z}\|_0 \leq s} \|\mathbf{x} - \mathbf{z}\|_1$ defines the best s-term approximation of $\mathbf{x}$ in the ℓ^1 norm.

Plethora of conditions (some of which we look closer in Section 4) on $\mathbf{A}$ have been derived to ensure that the previous (or similar) estimates hold. These are based on restricted isometry constants ($\delta_{2s} < \sqrt{2} - 1$ [12] or $\delta_{2s} < 4/\sqrt{41}$ [24], $\delta_s < 1/3$), null space properties ($\mathbf{A}$ fulfills the robust and stable NSP(s, ρ, τ) if $\|\mathbf{v}_S\|_1 \leq \rho\|\mathbf{v}_{\overline{S}}\|_1 + \tau\|\mathbf{A}\mathbf{v}\|_2$ for any vector $\mathbf{v} \in \mathbb{K}^N$, and S any index set with $|S| \leq s$), coherence ($\mu_1(s) + \mu_1(s-1) < 1$), quotient property, and so on. Similar conditions and bounds can be found when greedy and thresholding algorithms are used to approximate (ℓ^0-min) (see [6, 23, 8, 7, 36, 31] for some relatively recent results in this direction).

However, to ensure such recovery bounds, there is still a need for building adequate sensing matrices $\mathbf{A}$. So far, no deterministic matrices $\mathbf{A}$ can be built with reasonable numbers of rows (i.e. with decent, small enough, number of measurements) and we have to rely on randomness to build measurement matrices. It has been shown, for instance, that matrices with independent random subgaussian entries fulfill the RIP(s, δ) provided that the number of measurements scales as $m \asymp \delta^{-2}s\log(N/s)$ [5]. Similar results have been obtained (with different log factors) for structured random matrices or matrices from bounded orthonormal systems, see [32] for instance.

2.2. **Frames, fusion frames, and distributed signal processing.** By definition, a sequence $\mathcal{F} = \{f_i\}_{i \in I}$ in a Hilbert space $\mathcal{H}$ is a frame [16] for $\mathcal{H}$ if there exist $0 < A \leq B < \infty$ (lower and upper frame bounds) such that

$$A\|f\|^2 \leq \sum_{i \in I} |\langle f, f_i \rangle|^2 \leq B\|f\|^2 \text{ for all } f \in \mathcal{H}. \tag{4}$$

The representation space associated with $\mathcal{F}$ is $\ell^2(I)$ and its analysis and synthesis operators are respectively given by

$$T(f) = \{\langle f, f_i \rangle\}_{i \in I} \text{ and } T^*(\{c_i\}_{i \in I}) = \sum_{i \in I} c_i f_i,$$

for all $f \in \mathcal{H}$ and $\{c_i\}_{i \in I} \in \ell^2(I)$. It shows that the frame operator $S := T^*T$ is a positive, self-adjoint and invertible operator; this means that recovery of any $f \in \mathcal{H}$ is possible, if Sf is known; however, computing S^{-1} can be computationally challenging. Luckily, each frame $\mathcal{F}$ is accompanied by at least one so-called dual frame $\mathcal{G} = \{g_i\}_{i \in I}$, which satisfies

$$f = \sum_{i \in I} \langle f, f_i \rangle g_i = \sum_{i \in I} \langle f, g_i \rangle f_i \text{ for all } f \in \mathcal{H}, \tag{5}$$

and ensures recovery of the function f. Whenever a frame is A-tight ($A = B$), the problem of function reconstruction is simplified, since the frame operator S in this case is a scalar multiple of the identity operator.

Fusion frames (FF) have been initially created [13] to model the setting of a wireless sensor network. Sensor networks are composed of wireless sensors with constraints in their processing power and transmission bandwidth, which reduces the costs but also affects the precision of the system. The sensors are distributed over a significantly large area of interest, to measure, for instance, pollution, temperature, sound, pressure, motion etc. The network is redundant, i.e., there is no orthogonality among sensors, so each sensor can be interpreted as a frame element. In addition, a large sensor network is split into redundant sub-networks; the local measurements within each sub-network are sent to a local sub-station, which submits the gathered information further to a central processing station for final reconstruction.

Every (local) sensor is represented by a single frame vector; that is, each sub-network is related to a frame for a subspace in a Hilbert space. The subspaces have to satisfy a certain overlapping property, which ensures that the overlaps are not too large. The reconstruction in such a system is done in two steps: first, within each subspace the conventional frame reconstruction is employed; then, the local pieces of information serve as the inputs for the fusion frame reconstruction, which reconstructs the initial signal completely.

Definition 1 (Fusion frames). *Given an index set I, let $\mathcal{W} := \{W_i \,|\, i \in I\}$ be a family of closed subspaces in $\mathcal{H}$. We denote the orthogonal projections[1] onto W_i by P_i. Then $\mathcal{W}$ is a fusion frame, if there exist $C, D > 0$ such that*

$$C\|f\|^2 \leq \sum_{i \in I} \|P_i(f)\|^2 \leq D\|f\|^2 \text{ for all } f \in \mathcal{H}. \tag{6}$$

Remark 1. Fusion frames are often accompanied by respective weights. In the weighted case, the frame condition reads $C\|f\|^2 \leq \sum_{i \in I} v_i^2 \|P_i(f)\|^2 \leq D\|f\|^2$ for all $f \in \mathcal{H}$ and some positive weights $(v_i)_{i \in I}$. All (unweighted) results derived below apply mutatis mutandis to the weighted case.

Given a fusion frame $\mathcal{W}$ for a Hilbert space $\mathcal{H}$, let $\mathcal{F}_i := \{f_{ij} \,|\, j \in J_i\}$ be a frame for W_i, $i \in I$. Then $\{(W_i, \mathcal{F}_i) \,|\, i \in I\}$ is a FF system for $\mathcal{H}$. The following theorem [13] illustrates the relationship between the local and global properties of a FF:

Theorem 2. *For each $i \in I$, let W_i be a closed subspace of $\mathcal{H}$, and let $\mathcal{F}_i = \{f_{ij} \,|\, j \in J_i\}$ be a frame for W_i, with frame bounds A_i, B_i. If $0 < A = \inf_{i \in I} A_i \leq \sup_{i \in I} B_i = B < \infty$, then the following conditions are equivalent:*

- *$\cup_{i \in I}\{f_{ij} \,|\, j \in J_i\}$ is a frame for $\mathcal{H}$.*
- *$\{W_i\}_{i \in I}$ is a fusion frame for $\mathcal{H}$.*

In particular, if $\{(W_i, \mathcal{F}_i)\}_{i \in I}$ is a fusion frame system for $\mathcal{H}$ with frame bounds C and D, then $\cup_{i \in I}\{f_{ij} \,|\, j \in J_i\}$ is a frame for $\mathcal{H}$ with frame bounds AC and BD.

---

[1]an operator P is an orthogonal projection if $P^2 = P$ and $P^* = P$

In the fusion frame theory, an input signal is represented by a collection of vector coefficients that represent the projection onto each subspace. The representation space used in this setting is

$$\left(\sum_{i \in I} \oplus W_i \right)_{\ell^2} = \{\{f_i\}_{i \in I} \mid f_i \in W_i\}$$

with $\{\|f_i\|\}_{i \in I} \in \ell^2(I)$.

The analysis operator T is then defined by

$$T(f) := \{P_i(f)\}_{i \in I} \text{ for all } f \in \mathcal{H},$$

while its adjoint operator is the synthesis operator $T^* : \left(\sum_{i \in I} \oplus W_i \right)_{\ell^2} \to \mathcal{H}$, defined by

$$T^*(f) = \sum_i f_i, \text{ where } f = \{f_i\}_{i \in I} \in \left(\sum_{i \in I} \oplus W_i \right)_{\ell^2}.$$

The fusion frame operator $S = T^*T$ is given by

$$S(f) = \sum_{i \in I} P_i(f).$$

It is easy to verify that S is a positive and invertible operator on $\mathcal{H}$. In particular, it holds $C\mathbf{I} \leq S \leq D\mathbf{I}$, with $\mathbf{I}$ denoting the identity.

If the dual frames $\mathcal{G}_i$, $i \in I$, for each local frames are known then the fusion frame operator can be expressed in terms of the local (dual) frames [14]:

$$S = \sum_{i \in I} T^*_{\mathcal{G}_i} T_{\mathcal{F}_i} = \sum_{i \in I} T^*_{\mathcal{F}_i} T_{\mathcal{G}_i}.$$

For computational needs, we may only consider the fusion frame operator in finite frame settings, where the fusion frame operator becomes a sum of matrices of each subspace frame operator. The evaluation of the fusion frame operator S and its inverse S^{-1} in finite frame settings are conveniently straightforward. By F_i we denote the frame matrices formed by the frame vectors from $\mathcal{F}_i$, in a column-by-column format. Let G_i be defined in the same way from the dual frame $\{g_{ij}\}_{j \in J_i}$. Then the fusion frame operator is

$$S(f) = \sum_{i \in I} F_i G_i^T f = \sum_{i \in I} G_i F_i^T f.$$

Hence a distributed fusion processing is feasible in an elegant way, since the reconstruction formula for all $f \in \mathcal{H}$ is

$$f = \sum_{i \in I} S^{-1} P_i(f).$$

The standard distributed fusion procedure uses the local projections of each subspace. In this procedure, the local reconstruction takes place first in each subspace W_i, and the inverse fusion frame is applied to each local reconstruction and combined together:

$$f = \sum_{i \in I} S^{-1} P_i(f) = \sum_{i \in I} S^{-1} \left(\sum_{j \in J_i} \langle f, f_{ij} \rangle g_{ij} \right) \text{ for all } f \in \mathcal{H}. \tag{7}$$

Alternatively, one may use a global-like reconstruction procedure, which is possible if the coefficients of signal/function decompositions are available:

$$f = \sum_{i \in I} \sum_{j \in J_i} \langle f, f_{ij} \rangle (S^{-1} g_{ij}) \text{ for all } f \in \mathcal{H}. \tag{8}$$

The difference in procedure (8), compared with a global frame reconstruction lies in the fact that the (global) dual frame $\{S^{-1} g_{ij}\}$ is first calculated at the local level, and then fused into the global dual frame by applying the inverse fusion frame operator. This potentially makes the evaluation of (global) duals much more efficient.

As stated above, computing the inverse (fusion) frame operator is often a challenging task. Instead, one can approximate the solution $\widehat{x}$ by employing the so-called frame algorithm. All we need to start the iterative algorithm is $S\widehat{x}$, which we already have. We recall the relevant result below for completeness.

Proposition 2. [14] *Let $(W_i)_{i \in I}$ be a fusion frame in $\mathbb{K}^N$, with fusion frame operator $S = S_W$, and fusion frame bounds C, D. Further, let $\mathbf{x} \in \mathbb{K}^N$, and define the sequence $(\mathbf{x}_k)$ by $\mathbf{x}_0 = 0$ and $\mathbf{x}_k = \mathbf{x}_{k-1} + \frac{2}{C+D} S(\mathbf{x} - \mathbf{x}_{k-1})$, $k \geq 1$. Then we have $\mathbf{x} = \lim_{k \to \infty} \mathbf{x}_k$, with the error estimate $\|\mathbf{x} - \mathbf{x}_k\| \leq \left(\frac{D-C}{D+C} \right)^k \|\mathbf{x}\|$.*

Concretely, using the fusion frame reconstruction, the updates read

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \frac{2}{C+D} \sum_{i=1}^{n} \widehat{\mathbf{x}^{(i)}} + \frac{2}{C+D} \sum_{i=1}^{n} P_i \mathbf{x}_{k-1}.$$

The middle term is computed once and for all. The following updates then only require some basic matrix-vector multiplications. Note that starting the algorithm with $\widehat{\mathbf{x}}_0 = 0$ yields $\widehat{\mathbf{x}}_1 = \frac{2}{C+D} S\widehat{\mathbf{x}}$.

Similarly, in Section 5, we have $\widehat{\mathbf{x}} = S^{-1} \left(\sum_j \widehat{\mathbf{x}^{(j)}} \right)$, and $\widehat{f} = D\widehat{\mathbf{x}}$. We can find $\widehat{\mathbf{x}}$ via the iterative approach of Proposition 2, by starting the algorithm with $\widehat{\mathbf{x}_0} = 0$, $\widehat{\mathbf{x}_1} = \frac{2}{C+D} S(\widehat{\mathbf{x}}) = \frac{2}{C+D} \sum_j S\widehat{\mathbf{x}^{(j)}}$, and so on. The $n-$th term approximation of $\widehat{f}$ is then estimated with $\widehat{f}_n = D\widehat{\mathbf{x}}_n$.

2.3. **Signal recovery in fusion frames.** This work describes an approach for sensing and reconstructing signals in a fusion frame structure. As presented above, given some local information $\mathbf{x}^{(i)} := P_i(\mathbf{x})$, for $1 \le i \le n$, a vector can easily be reconstructed by applying the inverse fusion frame operator

$$\mathbf{x} := S^{-1} S(\mathbf{x}) = S^{-1} \sum_{i=1}^{n} P_i(\mathbf{x}) = S^{-1} \left(\sum_{i=1}^{n} \mathbf{x}^{(i)} \right), \text{ where } \mathbf{x}^{(i)} := P_i(\mathbf{x}), \text{ for } 1 \le i \le n. \quad (9)$$

Throughout the work, we assume that the projected vectors are sampled independently from one another, with n devices modeled by the same sensing matrix $\mathbf{A}$. Formally, the problem is as follows *From the measurements* $\mathbf{y}^{(i)} = \mathbf{A}\mathbf{x}^{(i)} + \mathbf{e}^{(i)}$, *reconstruct an estimation* $\widehat{\mathbf{x}^{(i)}}$ *of the local information to compute the approximation* $\widehat{\mathbf{x}}$ *of the signal* $\mathbf{x}$.

From this point on, there are two ways of thinking about the problem. In a first scenario, the signals are measured in the subspace and the recovery of the $\mathbf{x}^{(i)}$ are done locally. In other words, it accounts for solving n (ℓ^0-min) problems (or their approximations via (BPDN) for instance) in the subspaces, and then transmitting the estimated local signals to a central unit taking care of the fusion via Equation (9). The other approach consists of transmitting the local observations $\mathbf{y}^{(i)} = \mathbf{A}P_i\mathbf{x}$ to a central processing station which takes care of the whole reconstruction process. In this case, the vector $\mathbf{x}$ can be recovered by solving a unique (ℓ^0-min) problem directly with, letting $\mathbf{I}_n$ denote the n dimensional identity matrix,

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(n)} \end{pmatrix} = \mathbf{I}_n \otimes \mathbf{A} \begin{bmatrix} P_1 \\ \vdots \\ P_n \end{bmatrix} \mathbf{x} \quad (10)$$

While the latter case is interesting on its own (see for instance [17]), we investigate here some results for the first case. Our results can be investigated and generalized further, integrating ideas where the measurements matrix (here, the sensors) vary locally, as is the case in [17] or driven with some structured acquisition (see for instance [10]).

We would like to put our work in context. This paper is not the first one to describe the use of fusion frames in sparse signal recovery. However, it is inherently different from previous works in [9, 3]. In [9] the authors provide a framework for recovering sparse fusion frame coefficients. In other words, given a fusion frame system $\{(W_i, P_i)\}_{i=1}^{n}$ a vector is represented on this fusion frame as a set of n vectors $(\mathbf{x}^{(i)})_{i=1}^{n}$ where each of the $\mathbf{x}^{(i)}$ corresponds to the coefficient vector in subspace W_i. The main idea of the authors is that the original signal may only lie in few of the n subspaces, implying that most of the $\mathbf{x}^{(i)}$ should be 0. To rephrase the problem, we can say that the vector has to lie in a sparse subset of the original fusion frame. While [9] is concerned with some recovery guarantees under (fusion-) RIP and average case analysis, the paper [3] gives uniform results for subgaussian measurements and derive results on the minimum number of vector measurements required for robust and stable recovery.

We, on the other hand, exploit structures (or sparsity) locally[2]. We do not ask that only a few of the subspaces be active for a given signal, but that the signal only has a few active components per subspace. This justifies the use of the local properties of fusion frame systems.

3. Examples and applications

This section introduces some theoretical examples where the application of our fusion frame-based recovery shows increased performance over traditional recovery techniques. In particular, we justify in Proposition 4 that the fusion frame recovery may provide better robustness against noise.

---

[2]This sparsity assumption is needed only when a given frame component has a large dimension. We hope that in most of the cases we can control this dimension to be small enough avoiding the necessity of sparse assumptions. See the following section for more details.

3.1. **Orthogonal canonical projections.** Consider the problem of recovering $\mathbf{x} \in \mathbb{K}^N$ from the measurements:

$$\mathbf{y}^{(i)} = \mathbf{A}_i \mathbf{x} + \mathbf{e}^{(i)}, \quad 1 \leq i \leq n,$$

where $\mathbf{A}_i$'s are defined as $\mathbf{A}_i = \mathbf{A} P_i$ and P_i is the orthogonal projection onto Ω_i, with $\Omega_i \subseteq \{1, \cdots, N\}$. We assume that the noise vectors $\mathbf{e}^{(i)}$ are uncorrelated and independent, with $\|\mathbf{e}^{(i)}\|_2 \leq \eta_i$. $\mathbf{A}$ corresponds to a matrix of linear measurements in $\mathbb{K}^{m \times N}$. Note that this framework is an example of recovery from Multiple Measurement Vectors (*MMV*) (see [20, 28, 33] and references therein) where the signal is the same in every measurements, and the matrices are different for each set of measurements. We consider the problem of recovering local vectors $\mathbf{x}^{(i)}$, for $1 \leq i \leq n$, where the $\mathbf{x}^{(i)}$ are defined as $\mathbf{x}^{(i)} := P_i \mathbf{x}$. In this case, we have that the sparsity of each local signal is at most $N_i := \mathrm{rank}(P_i) = |\Omega_i|$. Once the local pieces of information $\widehat{\mathbf{x}^{(i)}}$ are recovered, the original signal $\mathbf{x}$ can be estimated from a *fusion frame like* reconstruction:

$$\widehat{\mathbf{x}} = S^{-1} \left(\sum_{i=1}^n \widehat{\mathbf{x}^{(i)}} \right) = \sum_{i=1}^n S^{-1} \widehat{\mathbf{x}^{(i)}}, \tag{11}$$

where the fusion frame operator S is defined in the usual way as $S : \mathbb{K}^N \to \mathbb{K}^N$, $S(\mathbf{x}) = \sum_{i=1}^n P_i(\mathbf{x})$. The problem is to recover the unknown vector $\mathbf{x}$ from the measurements $\{\mathbf{y}^{(i)}\}_{i=1}^n$ by solving local problems. Clearly, a necessary condition for uniform recovery is that we have $\bigcup_{i=1}^n \Omega_i = \{1, \cdots, N\}$ in a deterministic setting, or with high probability in a probabilistic setting.

The main idea of our approach is to solve the (very) high-dimensional, high-complexity, and demanding problem $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ by combining results obtained from n problems that are much easier to solve. We investigate first the case where the orthogonal projections do not overlap, i.e. $\Omega_i \cap \Omega_j = \varnothing$, for any $i, j \in \{1, \cdots, n\}$ with $i \neq j$ and $\bigcup_{i=1}^n \Omega_i = \{1, \cdots, N\}$. The results developed are reminiscent of some work on partial sparse recovery, when part of the support is already known [4]. In a second step we analyze the use of random and deterministic projections that may overlap for the recovery of the signal $\mathbf{x}$.

Remark 2. It is important to pause here and understand the problem above. Knowing the support of the projections, one could work on the local vectors $\mathbf{x}^{(i)}$ of smaller sizes $N_i < N$. Doing this would allow faster computations of the solutions but also has the implication that the sparsity (per subspace) has to remain small. On the other hand, considering large vectors with many zero components allows to keep a sparsity rather large (compared to the subspace dimension). We choose to investigate mainly the latter case as we try to break the limitation on the complexity for a given sensing matrix.

3.1.1. *Decomposition via direct sums.* It is clear that the dimension N of the ambient space will play an important role. Indeed, if we consider no overlap between the projections, we have two drastically different extreme cases. These scenarios are important as they shed light on the main ideas behind our recovery results.

<u>Case $n = N$</u> This case is characterized by (up to a permutation of the indices) $\Omega_i = \{i\}$ for all $1 \leq i \leq N$. In other words, each (set of) measurements $\mathbf{y}^{(i)} = \mathbf{A} P_i \mathbf{x} + \mathbf{e}^{(i)}$ gives information about a single entry x_i of the input vector $\mathbf{x}$ via an overdetermined linear system. In this case, assuming $\mathbf{A}$ has full rank, we can compute an estimate $\widehat{x_i}$ of the entry x_i as the solution to the ℓ^2 minimization problem:

$$\widehat{x_i} := \mathrm{argmin}_{x \in \mathbb{R}} \|\mathbf{y}^{(i)} - x\mathbf{a}_i\|_2^2 = \frac{\mathbf{a}_i^T \mathbf{y}^{(i)}}{\mathbf{a}_i^T \mathbf{a}_i} \tag{12}$$

where $\mathbf{a}_i$ denotes the ith column of the matrix $\mathbf{A}$. The N independent ℓ^2 minimizations ensure that the final solution $\widehat{\mathbf{x}}$ satisfies the bound

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 = \sqrt{\sum_{i=1}^N \frac{|\langle \mathbf{a}_i, \mathbf{e}^{(i)} \rangle|^2}{\|\mathbf{a}_i\|_2^4}} \leq \sqrt{\sum_{i=1}^N \frac{\|\mathbf{e}^{(i)}\|_2^2}{\|\mathbf{a}_i\|_2^2}} \leq \sum_{i=1}^N \frac{\|\mathbf{e}^{(i)}\|_2}{\|\mathbf{a}_i\|_2}. \tag{13}$$

In particular, in the noiseless scenario ($\mathbf{e}^{(i)} = 0$, for all i), the recovery is exact. In terms of (normalized) matrices from Gaussian measurements, which ensures that $\mathbf{A}$ has full rank – even numerically – with high-probability [15, 34], it holds $\mathbb{E}\|\mathbf{a}_i\|_2 = 1$ and it follows $\mathbb{E}\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq \sum_{i=1}^N \|\mathbf{e}^{(i)}\|_2$.

Notice that even though we have only one fixed matrix $\mathbf{A} \in \mathbb{K}^{m \times N}$, (obviously) we are no longer limited to the recovery of sparse vectors, which is the motivation for using fusion frames as described below. However, this has a price as the (actual) number of measurements drastically increased to $m \cdot N$. This is clearly too many measurements in any practical scenarios, but with similar ideas, we can reach practical applications such as using multiple lower sampling rate devices in order to resolve a harder

problem, typically required for much higher rate of sampling devices (see Section 1.1.1). Note also that all the results here should be put in the context of MMV problems, and not with the usual CS setup.

<u>Case $n = 2$</u> We consider now two sets Ω_1 and Ω_2 such that $\Omega_1 \cup \Omega_2 = \{1, \cdots, N\}$ and $\Omega_1 \cap \Omega_2 = \varnothing$. Without loss of generality we can assume that $\Omega_1 = \{1, \cdots, N_1\}$ and $\Omega_2 = \{N_1 + 1, \cdots, N_1 + N_2\}$ with $N_1 + N_2 = N$. We measure twice the unknown vector $\mathbf{x}$ as $\mathbf{y}^{(1)} = \mathbf{A}_1 \mathbf{x} + \mathbf{e}^{(1)} = \mathbf{A} P_1 \mathbf{x} + \mathbf{e}^{(1)}$ and $\mathbf{y}^{(2)} = \mathbf{A}_2 \mathbf{x} + \mathbf{e}^{(2)} = \mathbf{A} P_2 \mathbf{x} + \mathbf{e}^{(2)}$. If both N_1 and N_2 are smaller than m, then two ℓ^2 minimizations recover $\widehat{\mathbf{x}^{(1)}}$ and $\widehat{\mathbf{x}^{(2)}}$ independently and it follows that the solution $\widehat{\mathbf{x}} = \widehat{\mathbf{x}^{(1)}} + \widehat{\mathbf{x}^{(2)}}$ obeys the following error bound:

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \|\mathbf{A}_{\Omega_1}^+ \mathbf{e}^{(1)}\|_2^2 + \|\mathbf{A}_{\Omega_2}^+ \mathbf{e}^{(2)}\|_2^2. \tag{14}$$

The reconstruction is again perfect in the noiseless case. Moreover, as in the previous case, there is no need for a sparsity assumption on the original vector $\mathbf{x}$. The oversampling ratio is not very large as we only have a total of $2m$ measurements.

If however one at least of N_1 or N_2 (say N_1) is larger than m then we need to use other tools as we are now solving an underdetermined linear system. Driven by ideas from CS, we can assume the vector to be sparse on Ω_1. The recovery problem becomes

$$\text{Find } \widehat{\mathbf{x}^{(1)}} \text{ that minimizes } \|\mathbf{z}\|_1, \quad \text{subject to } \|\mathbf{A}_1 \mathbf{z} - \mathbf{y}^{(1)}\|_2 \leq \eta_1 \tag{15}$$

The ℓ^1 minimization problem is introduced as a convex relaxation of the NP-hard ℓ^0 minimization. Note that the constraints apply only on the support Ω_1 of the unknown vector. Hence the usual sparsity requirements encountered in the CS literature need not to apply to the whole vector. Unfortunately, the sparsity assumption being applied independently on Ω_1 and/or Ω_2 restricts ourselves to non-uniform recovery guarantees only, at least when considering the full set of sparse signals. The following definition will become handy in the analysis.

Definition 2 (Partial null space property). *A matrix $\mathbf{A} \in \mathbb{K}^{m \times N}$ is said to satisfy a partial robust and stable null space property of order s with respect to a subset $\Omega \subset \{1, \cdots, N\}$ with $N \geq |\Omega| > m$, P the orthogonal projection onto Ω, and constants $0 < \rho < 1$, $0 < \tau$ if*

$$\|(P\mathbf{v})_S\|_1 \leq \rho \|(P\mathbf{v})_{\bar{S}}\|_1 + \tau \|\mathbf{A}\mathbf{v}\|_2 \tag{16}$$

holds for any vector $\mathbf{v}$ and any set $S \subset \Omega$ such that $|S| \leq s$.

Remark 3. The previous definition is a weakening of the usual robust null space property, where what happens in the complement of the set Ω is irrelevant. It is worth noticing that it also coincides with the usual definition when $\Omega = \{1, \cdots, N\}$.

The partial robust null space property of the measurement matrix $\mathbf{A}$ on Ω_1 ensures that the recovered local vector $\widehat{\mathbf{x}^{(1)}}$ obeys the following error bound [24, Theorem 4.19]:

$$\|\widehat{\mathbf{x}^{(1)}} - P_1 \mathbf{x}\|_1 \leq \frac{2(1 + \rho)}{1 - \rho} \sigma_s(\mathbf{x}^{(1)})_1 + \frac{4\tau}{1 - \rho} \|\mathbf{e}^{(1)}\|_2. \tag{17}$$

Equation (17) is obtained by modifying the proofs of [24, Theorem 4.19, Lemma 4.15] and adapting them to the presence of the projection P_1. A formal proof of a more general statement is given in the proof of the later Theorem 6. This yields the following direct consequence

Proposition 3. *Let N and m be positive integers with $N > m$ and $\mathbf{A} \in \mathbb{K}^{m \times N}$. There exist two integers N_1 and N_2 such that, up to a permutation, $\Omega_1 = \{1, \cdots, N_1\}$ and $\Omega_2 = \{N_1 + 1, \cdots N_1 + N_2\}$. Assume that $N_2 \leq m$ and $N_1 > m$. Assume that the matrix $\mathbf{A}$ satisfies the partial robust null space property of order s with respect to Ω_1 with constants $0 < \rho < 1$ and $\tau > 0$ and that $\mathbf{A}_{\Omega_2}$ has full rank. For $\mathbf{x} \in \mathbb{K}^N$, let $\widehat{\mathbf{x}} := \widehat{\mathbf{x}^{(1)}} + \widehat{\mathbf{x}^{(2)}}$ with $\widehat{\mathbf{x}^{(1)}}$ solution to Problem (15) and $\widehat{\mathbf{x}^{(2)}}$ solution to the overdetermined ℓ^2 minimization problem on Ω_2. Then the solution obeys:*

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq \|\mathbf{A}_{\Omega_2}^+ \mathbf{e}^{(2)}\|_2 + \frac{2(1 + \rho)}{1 - \rho} \sigma_s(\mathbf{x}^{(1)})_1 + \frac{4\tau}{1 - \rho} \|\mathbf{e}^{(1)}\|_2. \tag{18}$$

Moreover, the total number of measurements amounts to $2m$ for the recovery of an $N_2 + s$ sparse vector.

Recovering the vector $\mathbf{x}^{(2)}$ does not create any problem, as long as the pseudo-inverse has a reasonable ℓ^2 norm, i.e. as soon as the rank of the projection is reasonably small (see discussion below and Theorem 3 in particular). The recovery of the vector $\mathbf{x}^{(1)} \in \mathbb{K}^N$ is ensured provided the number of subgaussian measurements m scales as [24, Corollary 9.34]

$$m \gtrsim 2s \ln(eN/s) \left(1 + \rho^{-1}\right)^2. \tag{19}$$

Proposition 3 slightly reformulated yields the following result:

Corollary 1. *Let* $\mathbf{A} := \mathbf{A}P_1 + \mathbf{A}P_2 \in \mathbb{K}^{m \times N}$ *with* $\Omega_1 \cap \Omega_2 = \varnothing$ *and* $\Omega_1 \cup \Omega_2 = \{1, \cdots, N\}$ *and* $N_1 = |\Omega_1|, N_2 = |\Omega_2|$. *If* $\mathbf{A}$ *is a subgaussian matrix and* $\mathbf{A}_{\Omega_2}$ *has full rank, then provided that*

$$m \gtrsim 2s \ln(eN/s) \left(1 + \rho^{-1}\right)^2,$$

any vector $\mathbf{x} \in \Theta := \Sigma_s^{N_1} + \mathbb{K}^{N_2} \subset \Sigma_{s+N_2}^N$ *can be recovered with the bound* (18).

Here, $\Sigma_s^{N_1}$ denotes the set of s sparse vectors in $\mathbb{K}^{N_1}$. The previous Corollary provides a uniform recovery result, on the proviso on the model for the vector $\mathbf{x}$.

3.1.2. *Controlled deterministic projections.* We look here at a scenario where the sets Ω_i (and hence the projections P_i) are completely and deterministically controlled. We consider an integer $n > 0$ and subsets $\Omega_i \subset \{1, \cdots, N\}$ with $|\Omega_i| = N_i$, $\cup_{i=1}^n \Omega_i = \{1, \cdots, N\}$ and $N_i \leq m$, and assume again that $\mathbf{A}$ has full rank. In this case, we can always enforce the disjointness of the support $\Omega_i \cap \Omega_j = \varnothing$, for any $i \neq j$. This yields the trivial recovery of the input signal from its local information vectors:

$$\widehat{\mathbf{x}} := \sum_{i=1}^n \widehat{\mathbf{x}^{(i)}}. \tag{20}$$

Here, all the local vectors are recovered via generalized Moore-Penrose pseudo inverses of the sub-matrices $\mathbf{A}_{\Omega_i}$ composed of only the columns supported on Ω_i:

$$\widehat{\mathbf{x}^{(i)}} = \mathbf{A}_{\Omega_i}^+ \mathbf{y}^{(i)} = \mathbf{x}^{(i)} + \mathbf{A}_{\Omega_i}^+ \mathbf{e}^{(i)}. \tag{21}$$

The error estimate follows directly:

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2 \leq \sum_{i=1}^n \|\mathbf{A}_{\Omega_i}^+ \mathbf{e}^{(i)}\|_2^2 \tag{22}$$

Note that it holds $\|\mathbf{A}^+ \mathbf{e}\|_2 \leq \|\mathbf{A}^+\|_{2 \to 2} \|\mathbf{e}\|_2$ and that $\|\mathbf{A}^+\|_{2 \to 2} = 1/\min_{1 \leq i \leq r}(\sigma_i)$, with r the rank of $\mathbf{A}$ and $\{\sigma_i\}_{i=1}^r$ its singular values. As a consequence, when dealing with "nice" matrices, the bound (22) is reasonable. In a CS setup, normalized sensing matrices generated at random from a Gaussian distribution have the property of being *well-behaved*, as suggested by [24, Theorem 9.26]:

Theorem 3. *Let* $\widetilde{\mathbf{A}}$ *be an* $m \times s$ *Gaussian matrix with* $m > s$ *and* $\mathbf{A} := \frac{1}{\sqrt{m}} \widetilde{\mathbf{A}}$ *be its variance-normalized counterpart. Then for* $t > 0$,

$$\mathbb{P}\left(\sigma_{max}(\mathbf{A}) \geq 1 + \sqrt{s/m} + t\right) \leq e^{-\frac{mt^2}{2}}$$

$$\mathbb{P}\left(\sigma_{min}(\mathbf{A}) \leq 1 - \sqrt{s/m} - t\right) \leq e^{-\frac{mt^2}{2}},$$

where $\sigma_{max}(\Theta)$ *and* $\sigma_{min}(\Theta)$ *are the maximum and the minimum singular values of* Θ, *respectively.*

For small projection ranks, and $\mathbf{A}$ as in Theorem 3, it holds $s \leq r < m$, and therefore $\mathbb{P}(\|\mathbf{A}_{\Omega_i}^+\|_{2 \to 2} \geq \frac{1}{1 - \sqrt{\frac{r}{m}} - t}) \leq e^{-mt^2/2}$ which justifies that the bound (22) is small.

3.1.3. *Rank-controlled projections.* This scenario differs from the previous one by the fact that we may control only the ranks N_i's of the projections but let the support of projections be random. This example is motivated by the SAR applications where the rank of the projections is controlled by the aperture and resolution of the sensing device. In this case, for uniform recovery of any vector $\mathbf{x}$ we need to ensure that, with high probability, the whole support $\{1, \cdots, N\}$ is covered by the random projections. We assume that all the projections have the same rank r for simplicity of calculations, but similar ideas apply if variations in the rank of the projections were needed. We pick uniformly at random n sets Ω_i of size r in $\{1, \cdots, N\}$. It holds

$$\mathbb{P}\left[\exists i \in \{1, \cdots, N\} : i \notin \Omega := \cup_{j=1}^n \Omega_j\right] = N\mathbb{P}\left[1 \notin \Omega\right] = N\left(\mathbb{P}\left[1 \notin \Omega_1\right]\right)^n = N\left(\frac{N-r}{N}\right)^n. \tag{23}$$

This gives that $\mathbb{P}\left[\exists i \in \{1, \cdots, N\} : i \notin \Omega := \cup_{j=1}^n \Omega_j\right] \leq \varepsilon$ whenever

$$n \geq \frac{\log(N/\varepsilon)}{\log(N/(N-r))}. \tag{24}$$

This direct consequence follows.

Proposition 4. *Let* $\mathbf{A} \in \mathbb{K}^{m \times N}$ *with* $m < N$, $r \leq m$, $\varepsilon > 0$, *and* n *be such that Equation* (24) *holds. In addition, assume that every submatrix with r columns extracted from $\mathbf{A}$ has full rank. Let* $\Omega_1, \Omega_2, \cdots, \Omega_n$ *be* n *subsets in* $\{1, \cdots, N\}$ *of size r chosen independently and uniformly at random. Any vector $\mathbf{x}$ recovered from the fusion of local recovery as $\widehat{\mathbf{x}} := S^{-1} \sum_{i=1}^{n} \widehat{\mathbf{x}^{(i)}}$ from the measurements* $\mathbf{y}^{(i)} = \mathbf{A} P_i \mathbf{x} + \mathbf{e}^{(i)}$, $1 \leq i \leq n$ *satisfies, with probability at least* $1 - \varepsilon$

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2^2 \leq \frac{1}{C} \sum_{i=1}^{n} \|\mathbf{A}_{\Omega_i}^+ \mathbf{e}^{(i)}\|_2^2 \tag{25}$$

where $S(\mathbf{x}) := \sum_{i=1}^{n} P_i \mathbf{x}$ *and* $\widehat{\mathbf{x}^{(i)}} := A_{\Omega_i}^+ \mathbf{y}^{(i)}$ *and* C *denotes the lower frame bound of* S.

The lower frame bound is defined for general fusion frame operators as in Definition 1. The error bound above is applicable in case of any fusion/filtering scenario. In our particular set-up, we calculate exactly this bound.

Lemma 1. *Given n projections P_i of rank r, the lower frame bound of the fusion frame operator $S := \sum P_i$ is given by*

$$C := \min_{1 \leq k \leq N} \mathcal{M}(k) =: \mathcal{M}, \tag{26}$$

where $\mathcal{M}(k)$ is the multiplicity of index k. Formally, let $\Gamma_k := \{j : 1 \leq j \leq n : k \in \Omega_j\}$, then $\mathcal{M}(k) := |\Gamma_k|$.

Proof. That $C \leq \mathcal{M}$ is clear. Indeed, let $\mathbf{x} = \mathbf{e}_{k_0}$ where k_0 is one of the indices achieving the minimum $\mathbf{e}_{k_0}$ denotes the canonical vector. Then it holds $S(\mathbf{e}_{k_0}) = \sum_{j \in \Gamma_{k_0}} \mathbf{e}_{k_0}$ and it follows that $\|S(\mathbf{x})\|_2^2 \leq \mathcal{M}^2$.

On the other hand, for any $\mathbf{x} \in \mathbb{K}^N$, the fusion frame operator can be rewritten as $S(\mathbf{x}) = \sum_{i=1}^{N} \mathcal{M}(i) x_i$. Consequently it holds that for any $\mathbf{x} \in \mathbb{K}^N$, $\|S(\mathbf{x})\|_2^2 \geq \mathcal{M}^2 \sum_{i=1}^{N} |x_i|^2$ and hence $C^2 \geq \mathcal{M}^2$. $\qquad \square$

As a consequence, it is trivial to see that for any $1 \leq k \leq N$, $\mathcal{M}(k)$ is a non-decreasing function of the number of projections n. Hence, the greater the number of projections (the redundancy, in fusion frame terms) the greater the bound C and therefore the smaller the error in the recovery according to Equation (25). This however can only hold true, as long as the noise per measurement vectors remains small.

Note that in the case of subsets selected independently uniformly at random, we can have a precise statement. The lower frame bound is defined as the minimum number of occurrences of any index $k \in \{1, \cdots, N\}$. Let us consider $\mathcal{M}(k)$ and Γ_k as in the previous lemma. Let $k \in \{1, \cdots, N\}$ be any index and Ω_j denote a draw of a random set. It holds $\mathbb{P}[k \in \Omega_j] = r/N$. The subsets being independent of each other, $\mathcal{M}(k)$ is a binomial random variable with probability of success $P = r/N$ for each of the n trials. It follows that $\mathbb{P}[\mathcal{M}(k) = l] = \binom{n}{l} P^l (1-P)^{n-l}$ for all $1 \leq k \leq N$ and $1 \leq l \leq n$.

Putting everything together, we get that $C := \min_{k \in \{1, \cdots, N\}} \mathcal{M}(k)$ is a random variable such that

$$\mathbb{P}[C \geq l] = \mathbb{P}[\forall 1 \leq k \leq N, \mathcal{M}(k) \geq l] \tag{27}$$

$$= (\mathbb{P}[\mathcal{M}(k) \geq l])^N = \left(1 - \sum_{j=1}^{l-1} \mathbb{P}[\mathcal{M}(k) = j]\right)^N = (1 - F(l-1, n, P))^N. \tag{28}$$

These expressions resemble the calculations involved in Equations (23) and (24), however with a much more complicated probability distribution. To avoid unnecessarily tedious calculations that would infer the readability of the paper, we choose to leave the following result as a conjecture.

Conjecture 1. *Let $\Omega_1, \Omega_2, \cdots, \Omega_n$ be n subsets of r elements taken uniformly at random in $\{1, \cdots, N\}$. Let $\Gamma_k := \{j : 1 \leq j \leq n : k \in \Omega_j\}$, and $\mathcal{M}(k) := |\Gamma_k|$. Then $C := \min_{1 \leq k \leq N} \mathcal{M}(k) =: \mathcal{M}$ grows at least linearly with n, i.e., there exists constants $c > 0$ and $b \geq 0$, independent of n, such that*

$$\mathbb{E}[C] \geq cn - b. \tag{29}$$

A direct consequence of this is that one can always find a multiplicative constant $c' \leq c$ such that for a number of projections $n \geq b/(c - c')$, $\mathbb{E}[C] \geq c'n$ (the case $c = c'$ can only be true for $b = 0$, in which case, there is no need for this remark). Though the proof is not given, the result is backed up with numerical results shown in Figure 3.[3] The graphs always show the expected linear growth, independently of the parameters of the problem.

---

[3]It is also backed up with particular cases, when the ranks are exactly half the dimension, see http://math.stackexchange.com/questions/1135253/mean-value-of-minimum-of-binomial-variables

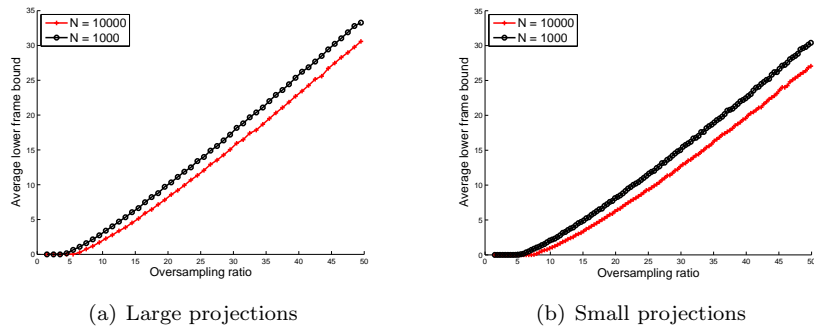(a) Large projections   (b) Small projections

FIGURE 3. Linear growth of the expectation of the lower frame bound for various use cases. The averages are calculated from 300 random experiments. The left graphs show the average lower frame bound when dealing with rather large projections (with rank $r = N/2$) while the right graphs are in the case of much smaller projections ($r = N/5$). The oversampling ratio is calculated as a linear function of the number of projections and is independent of the ambient dimension N; $\zeta = rn/N$.

Assuming the conjecture to be true, Equation (25) then simplifies to

$$\mathbb{E}\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \frac{1}{c'n} \sum_{i=1}^n \|\mathbf{A}_{\Omega_i}^+ \mathbf{e}^{(i)}\|_2^2 \leq \frac{\delta\nu}{c'}, \tag{30}$$

where the expectation is taken over the draw of the projections, for a certain constant $c' > 0$, where we let $\delta = \max_i \|\mathbf{A}_{\Omega_i}^+\|_{2\to 2}$ and $\nu := \max_i \|\mathbf{e}^{(i)}\|_2^2$. It shows that at least in expectation, the error should not grow as the number of measurements increases. Moreover, numerical results illustrated in Figure 5 suggest that the expected reconstruction error decreases. An intuition for this behavior is provided in a simple example in Equations (34) and (35). From a compressed sensing perspective, it has been shown that the least favorable case appears as the case when the vectors are repeated again and again. This translates in our scenario to the case where the increase of the measurements consist solely in repeating the same projections. Considering the example of a SAR imaging process developed in Section 1.1.2, this would correspond to a plane flying over a given region multiple times, with the spatial filtering being exactly the same.

Remark 4. It is important to note here that there is no sparsity assumption. It is possible to recover **any** vector $\mathbf{x}$ with high probability, as long as the number of measurement vector scales reasonably with the dimension of the input space.

3.1.4. *Constrained number of measurements and ranks of projections.* In certain scenarios, the physical measurement devices limit our freedom in choosing the ranks and size of the measurements. We hence deal here with fixed ranks N_i's for the projections P_i. Assume, without loss of generality, that the first $0 \leq k \leq n$ projections have ranked lower than the number of measurements m. For these subspaces, the usual ℓ^2 minimization procedure yields perfect (or optimal in the noisy case) recovery of the local information $\mathbf{x}^{(i)}$, for $1 \leq i \leq k$. For the remaining $n - k$ subspaces, the pseudo-inverse is not sufficient, and we use tools from CS. We here again use the partial robust null space property from Definition 2 on every subspace Ω_j, for $k + 1 \leq j \leq n$. Combining results from Propositions 4 and 3 yields the following direct corollary as a consequence:

Corollary 2. *Let $m < N$ and $A \in \mathbb{K}^{m\times N}$. Assume that we are given a set of n sets $\{\Omega_i\}_{i=1}^n$ in $\{1, \cdots, N\}$ of respective sizes $\{N_i\}_{i=1}^n$ with $N_1 \leq N_2 \leq \cdots N_n$. Assume in addition that $\cup_i \Omega_i = \{1, \cdots, N\}$. Moreover, there exists a unique $0 \leq k \leq n$ such that $N_k \leq m$ and $N_{k+1} > m$ (with the convention that $N_0 = 0$ and $N_{n+1} > m$). Assume that the submatrices $\mathbf{A}_{\Omega_i}$, for $1 \leq i \leq k$, have full rank. If in addition the measurement matrix satisfies a partial robust null space property of order s with respect to every subset Ω_i, $i \geq k + 1$, then, the approximation $\widehat{\mathbf{x}}$ defined as $\widehat{\mathbf{x}} := S^{-1} \sum_{i=1}^n \widehat{\mathbf{x}^{(i)}}$ with $\widehat{\mathbf{x}^{(i)}} := \mathbf{A}_{\Omega_i}^+ \mathbf{y}^{(i)}$, for $1 \leq i \leq k$ and $\widehat{\mathbf{x}^{(i)}}$ solutions to the ℓ^2 constrained ℓ^1 minimization problems (BPDN), for $k + 1 \leq i \leq n$ fulfills the following estimate:*

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq \frac{1}{\mathcal{M}} \left(\sum_{i=1}^k \|\mathbf{A}_{\Omega_i}^+ \mathbf{e}^{(i)}\|_2 + \frac{2(1+\rho)}{1-\rho} \sum_{i=k+1}^n \sigma_s(\mathbf{x}^{(i)})_1 + \frac{4\tau}{1-\rho} \sum_{i=k+1}^n \|\mathbf{e}^{(i)}\|_2 \right) \tag{31}$$

Note that for simplicity we have assumed that the NSP is valid on every subspaces for the same set of parameters. To avoid notational encumbrance we do not write results where the ρ, τ, and s may depend on the subspace considered.

Remark 5. This is a direct application of the previous results. It is obtained by recovering every local pieces of information independently from one another. It is possible to improve these estimates by a sequential procedure where we first estimate the x_j for $j \in \Omega_i$, for some $1 \leq i \leq k$, and then using this reliable estimate to improve the accuracy of the ℓ^1 minimization program.

A final remark considers the case where the rank of the projection is also random. We can think of the sets Ω_i, for $1 \leq i \leq n$ as Binomial random variables with probability of success p. In this case, the rank of the projection P_i if controlled by the expectation of this random variable. An analysis similar to the previous one can be carried over to ensure recovery of any vector with high probability.

3.2. **Numerical examples.** This Section describes numerical recovery results on the case described above. We first show some particular examples of recovery, when dealing with dense signals and show that we can break the traditional sparsity limit by adding a few sensors. The next example shows the behavior of the quality of the reconstruction when slowly adding sensors. Finally, the last subsection validates our Conjecture and show that the noise tends to decrease while adding projections, and that the overall quality of recovery scales linearly with the noise level. In another contribution, we have also verified that we can recover a very dense Fourier spectrum, by using our approach with Fourier measurement matrices, see [1].
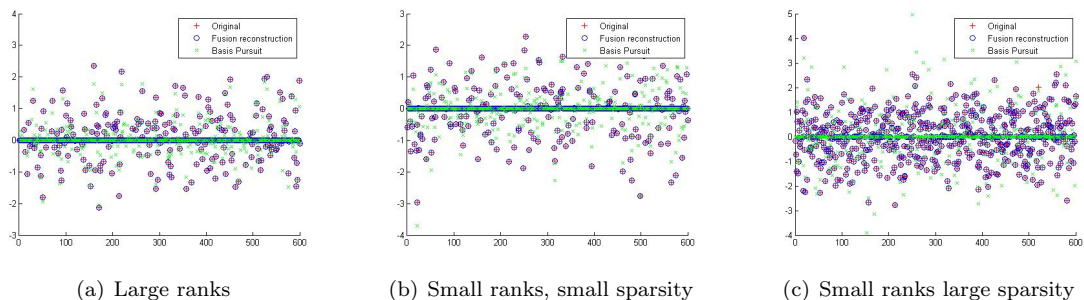

(a) Large ranks	(b) Small ranks, small sparsity	(c) Small ranks large sparsity

FIGURE 4. Examples of fusion reconstruction.

3.2.1. *Examples of recovery.* The examples of reconstruction depicted on Figures 4 were created by generating at random s sparse Gaussian vectors with $s = 200$ for Figures 4(a) and 4(b), and $s = 500$ for Figure 4(c). For each cases, the 250×600 matrix was generated at random with entries i.i.d. from a Gaussian distribution. In the first example, the rank is set to $r = 300 > m$ and hence, ℓ^1 minimization (traditional basis pursuit) is used on every subspaces. This yields a total of 13 projections generated at random. For the two last examples, the ranks of the projections are set to $r = 200 < m$. In this case, only a classical ℓ^2 inversion is needed on every subspace. Here, since the ranks are smaller, the number of projections has to be increased to 22 in order to ensure that the whole set $\{1, \cdots, N\}$ is covered with high probability. By looking carefully at the last example, one can see that an index with non-zero magnitude has not been selected (around index 520). These cases, however, are rare. In every figure, the red '+' crosses represent the true signal, the blue circles represent the reconstructed signal from our fusion approach, and the green 'x' correspond to the reconstruction with traditional basis pursuit. While the very last set-up (very high number of non-zero components) is clearly not suitable for usual compressed sensing, the advantage of projections and fusions can be seen even in the fist two (where the number of non-zeros remains relatively small). Another aspect to look at is that when dealing with small ranks, the solutions to the ℓ^2 minimization problems are computed efficiently. As a consequence, even when the number of projections increases, the calculation of the recovered $\widehat{\mathbf{x}}$ is still orders of magnitude faster.

3.2.2. *Recovery and number of projections.* Given any scenario introduced above, it is expected from Conjecture 1 that increasing the number of projections (of a given rank) will increase the lower frame bound (26) at least linearly. This linear increase of the lower frame bound compensates for the potential increase in the reconstruction error. One could however hope for better results according to the numerical

evidence illustrated in Figure 5, where the average case error seems to be reduced as the number of projections increases. Given a set of n projections generating a fusion frame, the least favorable case, studied in [21] for the case of a single filtering operation, when adding an extra n projections, is when the exact same Ω_i are repeated. The exact behavior of the reconstruction error with respect to the number of projections is still under research.
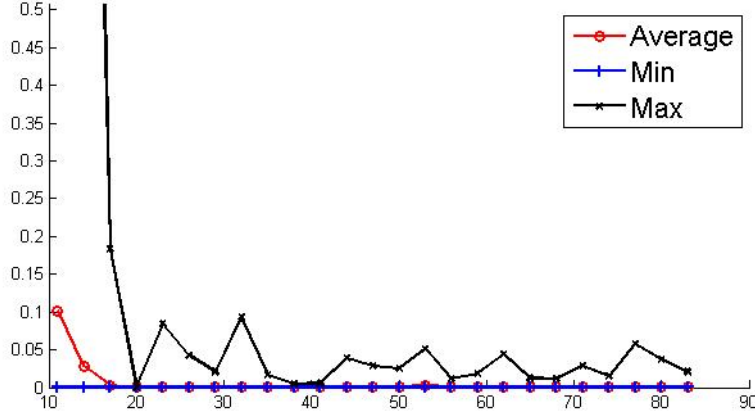

FIGURE 5. ℓ^2 error of recovery for various number of projections.

For a total of 10 generations of random Gaussian matrices of size 300×1000, and 10 random Gaussian 200 sparse vectors (note: that is a setting where usual compressed sensing has no recovery guarantees), we let the number of random projections of rank $r = 500$ go from 60% up to 5 times the minimum given by Equation (24) (17 in our setting, for a success of covering the whole input set of $\varepsilon = 99\%$). As can be seen from the figure, the recovery is unlikely to be correct as long as the number of projections remains small. The trend is generally towards the perfect recovery - though this can only be guarantee with high accuracy, which explains the little spikes. Moreover, while the maximum might be somewhat bigger, it appears 1) that it remains within reasonable bounds (depending on the application) and 2) the average of the error among the 100 tests per number of projections is very small, suggesting that only a few of the recoveries are off. A way to understand this is to consider the case, where we have two projections that are slightly overlapping. Given the global set $\Omega = \{1, \cdots, N\}$, assume we split it into Ω_1 and Ω_2 with some overlap, i.e. $\Omega = \Omega_1 \cup \Omega_2$ and $\Omega_{12} := \Omega_1 \cap \Omega_2 \neq \varnothing$. We independently solve

$$\widehat{\mathbf{x}^{(i)}} := \operatorname{argmin} \|\mathbf{x}\|_1 \tag{32}$$

$$\text{s.t. } \|\mathbf{A}P_i\mathbf{x} - \mathbf{y}^{(i)}\|_2 \leq \eta_i \tag{33}$$

for $1 \leq i \leq 2$. For now we consider only the noiseless case, i.e. $\varepsilon_1 = \varepsilon_2 = 0$ which is equivalent to solving two basis pursuit optimization problems. We can decompose the solutions recovered as $\widehat{\mathbf{x}^{(i)}} = \widehat{\mathbf{x}^{(i)}}|_{\Omega_i \setminus \Omega_{12}} + \widehat{\mathbf{x}^{(i)}}|_{\Omega_{12}}$. If we assume, for simplicity of understanding, that the first component is recovered exactly, we have that $\widehat{\mathbf{x}^{(1)}} = \mathbf{x}^{(1)}$. If however the second component is not recovered accurately, we can write $\widehat{\mathbf{x}^{(2)}} = \mathbf{x}^{(2)} + \mathbf{e}^{(2)}$ (for further generalizations, we can always write $\widehat{\mathbf{x}^{(i)}} = \mathbf{x}^{(i)} + \mathbf{e}^{(i)}$, for $1 \leq i \leq n$ with $\mathbf{e}^{(i)} = 0$ in case of successful recovery on the set Ω_i). In this very simple example, we have $S(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{x}^{(i)} = \mathbf{x}^{(1)} + \mathbf{x}^{(2)}$. Similarly, the inverse fusion frame operator can be seen as the point wise empirical average of the evidences (i.e. the local recoveries):

$$S^{-1}(\mathbf{x}) := \mathbf{x}^{(1)}|_{\Omega_1 \setminus \Omega_{12}} + \mathbf{x}^{(2)}|_{\Omega_2 \setminus \Omega_{12}} + \frac{\mathbf{x}^{(1)}|_{\Omega_{12}} + \mathbf{x}^{(2)}|_{\Omega_{12}}}{2}. \tag{34}$$

Note that the 2 in the denominator denotes the number of times that the indices in Ω_{12} are picked in the creation of the (random) projections. In other words, increasing the number of projections increases potentially the denominator under a given index and hence reduces the error.

$$\widehat{\mathbf{x}} - \mathbf{x} = S^{-1}\left(\widehat{\mathbf{x}^{(1)}} + \widehat{\mathbf{x}^{(2)}} - \mathbf{x}^{(1)} - \mathbf{x}^{(2)}\right) = \mathbf{e}^{(2)}|_{\Omega_2 \setminus \Omega_{12}} + \frac{\mathbf{e}^{(2)}|_{\Omega_{12}}}{2} \tag{35}$$

As a consequence, the error (on the set Ω_{12}) is bounded by the maximum of the error from every recovery problem divided by the number of times a particular index appears.
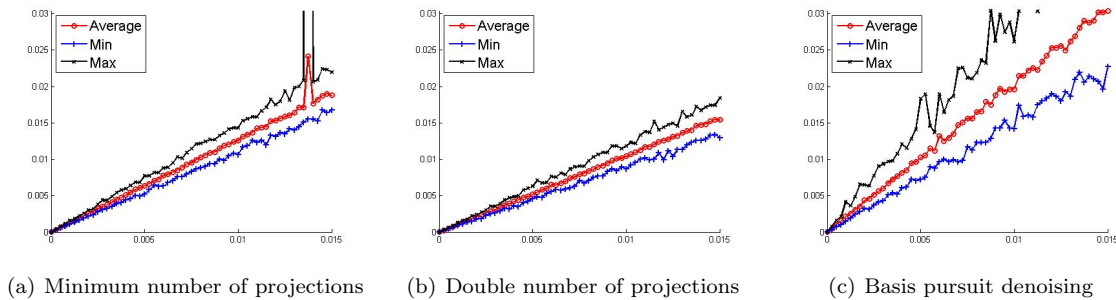
(a) Minimum number of projections   (b) Double number of projections   (c) Basis pursuit denoising

FIGURE 6. Behavior facing noise

3.2.3. *Robustness to noise.* In the set of experiments illustrated in Figures 6, we have compared the robustness of our algorithm compared to traditional basis pursuit denoising when dealing with noise. We have set the parameters in a regime where (BPDN) works for fair comparison. For a 500 dimensional ambient space and 300 dimensional measurement space, we generated 5 random 80 sparse Gaussian vectors for each of the 10 randomly generated Gaussian matrices. For every single test, some additive Gaussian noise with a given ℓ^2 norm θ has been added. This means that for the case of the distributed approach, n independent noise components of norm θ have been added in total; one for each channel. The figures depict the evolution of the error of the ℓ^2 norm when the energy of the noise component increases. The black line represents the maximum of the error, the blue one the minimum, and the red one the average over every samples. The left graph shows the result when the number of projections is exactly set as in Equation (24) while the second one doubles this number. The third figure shows the result when using the usual Basis Pursuit Denoising. It is important to notice that the higher peak appearing on the first figure is due to an index with non zero component from the support of the original vector $\mathbf{x}$ not being selected at all during the random projections. This however does not contradict the high probability of recovery. The improvement in the noise behavior from the first figure to the second shows how the fusion frame operator tends to average out the local errors to yield a better estimate (as described in the previous section). Finally, all of the algorithms scale linearly with the norm of the noise per measurements (as suggested by Equation (25)) and even if the total noise is increased as the number of projections is increased, the recovery tends to be improved by considering more projections.

4. AN EXTENSION OF TRADITIONAL COMPRESSED SENSING

Our results so far ensure that robust and stable recovery of dense signals is possible, by smartly combining local information. In this section we show that the ideas developed in Section 5 can also be considered as an extension of the traditional CS. In particular, it gives solid mathematical foundations to our work. It is important to note that similar ideas have been developed in parallel in [17]. There, the authors introduced a model similar to (10), albeit asking that the fusion frame be tight with $C = D = 1$, but reconstruct the whole signal globally, without the use of the fusion process. Other authors [10] looked at CS with structured acquisition and structure sparsity. In simple words, they prove that adapting the sampling matrices to some prior knowledge of the sparsity pattern allowed for larger applicability of the CS framework.

In this section we describe the recovery of a signal $\mathbf{x}$ by means of CS in the local subspaces and fusion processing. The local pieces of information are computed as solutions to the problems

$$\min_{\mathbf{x}\in\mathbb{K}^N} \quad \|\mathbf{z}\|_1 \qquad \text{s.t.} \quad \|\mathbf{A}P_i\mathbf{z} - \mathbf{y}^{(i)}\|_2 \le \eta_i. \tag{$\mathcal{P}_{1,\eta}$}$$

In the noiseless case, the problem is solved by the basis pursuit

$$\min_{\mathbf{x}\in\mathbb{K}^N} \quad \|\mathbf{z}\|_1 \qquad \text{s.t.} \quad \mathbf{A}P_i\mathbf{z} = \mathbf{y}^{(i)}. \tag{$\mathcal{P}_{1,0}$}$$

Lemma 2. *Let* $\widehat{\mathbf{z}^{(i)}}$ *be a solution to the noisy* ($\mathcal{P}_{1,\eta}$) *or noiseless* ($\mathcal{P}_{1,0}$) *basis pursuit problem. Then* $\widehat{\mathbf{z}^{(i)}} \in W_i$.

Proof. Let $\widehat{\mathbf{z}(i)}$ be a solution and let $\tilde{\mathbf{z}} = P_i\widehat{\mathbf{z}^{(i)}}$. Then $\|\tilde{\mathbf{z}}\|_1 \le \|\widehat{\mathbf{z}^{(i)}}\|_1 + \|(I - P_i)(\widehat{\mathbf{z}^{(i)}})\|_1$. $\tilde{\mathbf{z}}$ being also admissible, it follows that $\|\tilde{\mathbf{z}}\|_1 = \|\widehat{\mathbf{z}^{(i)}}\|_1$ and $\widehat{\mathbf{z}^{(i)}} = P_i(\widehat{\mathbf{z}^{(i)}})$. $\qquad\square$

4.1. Signal models and recovery conditions.

4.1.1. *Extension of the sparsity model.* The traditional sparsity model is not appropriate in this setting. As an example, let us consider that all the s nonzero components of a vector $\mathbf{x}$ fall within a certain subspace (say W_1), there is, a priori, no hope to improve the recovery performance compared to a single sensor/subspace problem. Indeed, in this case, the recovery is ensured (locally) by CS methods if the number of observations m scales as

$$m \asymp s \log(N/s).$$

Since we are dealing here only with an identical sensing model, this yields a total number of observations m_T scaling as

$$m_T \asymp ns \log(N/s).$$

This may be acceptable if we consider only very few subspaces but may explode in certain cases. Therefore the model of distributed sparsity may be more appropriate.

Definition 3. *A signal* $\mathbf{x} \in \mathbb{K}^N$ *is said to be* $\mathbf{s} = (s_1, \cdots s_n)$-*distributed sparse with respect to a fusion frame* $\mathcal{W} = (W_i, P_i)_{1 \leq i \leq n}$, *if* $\|P_i(\mathbf{x})\|_0 \leq s_i$, *for every* $1 \leq i \leq n$. $\mathbf{s}$ *is called the* sparsity pattern *of* $\mathbf{x}$ *with respect to* $\mathcal{W}$.

We denote by $\Sigma_{\mathbf{s}}^{(\mathcal{W})}$ the set of all $\mathbf{s}$-distributed sparse vectors with respect to the family of subspaces $(W_i)_i$. We let $s = \|\mathbf{s}\|_1$ denote the global sparsity of the vector, with respect to $\mathcal{W}$. In the case that the sparsity of the signal is uniformly distributed among the subspaces ($s_i = s/n$), the usual CS recovery guarantees ensure us that

$$m \asymp s_i \log(N/s_i)$$

observations per subspace are required for a stable and robust recovery of the signal. This accounts for a total number of measurements scaling as $m_T \asymp ns_i \log(N/s_i) = s \log(N/s_i)$. In other words, we are able to recover similar sparsities as in the classical CS framework. But in opposition to the classical theory most of the computations can be easily carried in a distributed setting, where only pieces of the information are available. Only the fusion process requires all the local information to compute the final estimation of a signal.

Locally, it only requires solving some very small CS system, which can be done faster than solving the original one. This is also the findings found in parallel in [17], where it is concluded that the number of measurements per sensor decreases linearly with the number of sensor. We describe a similar problem, while looking at it from a different perspective. In particular, we try to find the sparsity patterns that may be recovered for a given sensor design. The motivation for this problem comes from the applications in SAR imaging where the sensor is given and the same everywhere, and where we may not have any control on the number of observations per subspace. As it will become useful later, we also need to introduce the local best approximations.

Definition 4. *Let* $\mathcal{W} = (W_i, P_i)_{i=1}^n$ *be a fusion frame, and let* $\mathbf{x} \in \mathbb{K}^N$. *For* $p > 0$ *and a sparsity pattern* $\mathbf{s} = (s_1, \cdots, s_n)$ *with* $s_i \in \mathbb{N}$ *for all* $1 \leq i \leq n$, *the* ℓ^p *errors of best* $\mathbf{s}$-*term approximations are defined as the vector*

$$\sigma_{\mathbf{s}}^{\mathcal{W}}(\mathbf{x})_p := (\sigma_{s_1}(P_1\mathbf{x}), \sigma_{s_2}(P_2\mathbf{x}), \cdots, \sigma_{s_n}(P_n\mathbf{x}))^T$$

4.1.2. *Partial properties.* The *null space property* (NSP) has been used throughout the past decade in the CS literature as a necessary and sufficient condition for the sparse recovery problem via (ℓ^0-min). A matrix $\mathbf{A}$ is said to satisfy the (robust) null space property with parameters $\rho \in (0, 1)$ and $\tau > 0$ relative to a set $S \subset \{1, \cdots, N\}$ if

$$\|\mathbf{v}_S\|_1 \leq \rho \|\mathbf{v}_{\overline{S}}\|_1 + \tau \|\mathbf{A}\mathbf{v}\|_2, \quad \text{for all } \mathbf{v} \in \mathbb{K}^N.$$

More generally, we say that the matrix $\mathbf{A}$ satisfies the NSP of order s if it satisfies the NSP relative to all sets S such that $|S| \leq s$. We extend here this idea to the context of distributed sparsity with respect to fusion frames, as already mentioned in Definition 2. Here we talk about a sparsity pattern and ask that the NSP property be valid for all local subspaces up to a certain (local) sparsity level.

Definition 5 (Robust and stable partial null space property (RP-NSP)). *Let* n *be an integer and* $\mathcal{W} = (W_i, P_i)_{i=1}^n$ *be a frame for* $\mathbb{K}^N$. *Let* $\mathbf{s} = (s_1, \cdots, s_n)$ *be a sequence of non negative numbers representing the sparsity pattern with respect to* $\mathcal{W}$. *For a number* $q \geq 1$, *a sensing matrix* $\mathbf{A} \in \mathbb{K}^{m \times N}$ *is said to satisfy the* ℓ^q-*RP-NSP with pattern* $\mathbf{s}$ *with respect to* $\mathcal{W}$ *and with constants* $\rho_1, \cdots, \rho_n \in (0, 1)$ *and* $\tau_1, \cdots, \tau_n > 0$ *if*

$$\|(P_i\mathbf{v})_{S_i}\|_q \leq \frac{\rho_i}{s^{1-1/q}} \|(P_i\mathbf{v})_{overlineS_i}\|_1 + \tau_i \|\mathbf{A}\mathbf{v}\|_2, \quad \text{for all } \mathbf{v} \in \mathbb{K}^N, 1 \leq i \leq n, S_i \subset W_i, \text{ and } |S_i| \leq s_i.$$

This definition is reminiscent of the work on sparse recovery with partially known support [4]. The difference here being that there is no need to enforce a condition on the vector $\mathbf{v}$ to lie in the range of the other subspaces. In a sense, this is taken care of by the fusion process and the fact that we have multiple measurement vectors.

Remark 6. Note that we could simplify the definition by asking that the parameters be uniform and independent of the local subspace. Namely, introducing $\tau := \max_{1 \le i \le n} \tau_i$ and $\rho := \max_{1 \le i \le n} \rho_i$, the above definition becomes

$$\left\| (P_{W_i} \mathbf{v})_{S_i} \right\|_q \le \frac{\rho}{s^{1-1/q}} \left\| (P_{W_i} \mathbf{v})_{\overline{S_i}} \right\|_1 + \tau \|\mathbf{A}\mathbf{v}\|_2, \quad \text{for all } \mathbf{v} \in \mathbb{K}^N, 1 \le i \le n, S_i \subset W_i, \text{ and } |S_i| \le s_i.$$

A stronger, but easier to verify, condition often used as a sufficient recovery condition is the by-now well known *Restricted Isometry Property* (RIP). Informally speaking a matrix is said to satisfy the $RIP(s, \delta)$ if it behaves like an isometry (up to a constant δ) on every s-sparse vector $\mathbf{v} \in \Sigma_s$. Formally speaking, $\mathbf{A} \in \mathbb{K}^{m \times N}$ satisfies $RIP(s, \delta)$, for some $s \ge 2$ and $\delta \in (0, 1)$ if

$$(1 - \delta)\|\mathbf{v}\|_2^2 \le \|\mathbf{A}\mathbf{v}\|_2^2 \le (1 + \delta)\|\mathbf{v}\|_2^2, \quad \text{for every } \mathbf{v} \in \Sigma_s. \tag{36}$$

The lowest δ satisfying the inequalities is called the *restricted isometry constant*. Once again, we want to derive similar properties on our sensing matrix for the distributed sparse signal model.

Definition 6 (Partial-RIP (P-RIP)). *Let $\mathcal{W} = (W_i, P_i)_{i=1}^n$ be a fusion frame, and let $\mathbf{A} \in \mathbb{K}^{m \times N}$. Assume that $\mathbf{A}$ satisfies the $RIP(s_i, \delta_i)$ on W_i, with $\delta_i \in (0, 1)$, $i \in I = \{1, \cdots, n\}$. Then, we say that $\mathbf{A} \in \mathbb{K}^{m \times N}$ satisfies the partial RIP with respect to $\mathcal{W}$, with bounds $\delta_1, \cdots, \delta_n$ and sparsity pattern $\mathbf{s} = (s_1, \cdots, s_n)$.*

In other words, $\mathbf{A}$ satisfies a P-RIP conditions, if it satisfies RIP-like conditions on every subset of vectors in $\mathrm{range}(P_i)$.

Remark 7. This definition is consistent with the definition of the classical RIP in the sense that the case $n = 1$ (only one projection, one subspace) recovers the usual RIP.

The P-RIP can be written in a form similar to the traditional RIP, Equation (36).

Proposition 5. *Let $\mathcal{W} = (W_i)_{i \in I}$ be a fusion frame (with frame bounds $0 < C \le D < \infty$). Let $\mathbf{A} \in \mathbb{K}^{m \times N}$ satisfy the P-RIP with respect to $\mathcal{W}$, with bounds $\delta_1, \cdots, \delta_n$ and sparsity pattern $\mathbf{s} = (s_1, \cdots, s_n)$, and let $C_o = C \min_i \{1 - \delta_i\}$, $D_o = D \max_i \{1 + \delta_i\}$. Then, for any $\mathbf{v} \in \mathbb{K}^N$,*

$$C_o \|\mathbf{v}\|_2^2 \le \sum_i \|\mathbf{A}\mathbf{v}_i\|_2^2 \le D_o \|\mathbf{v}\|_2^2.$$

Proof. Using the fusion frame inequality, and inequalities (36) for all $i \in I$, we obtain

$$C \min_i \{1 - \delta_i\} \|\mathbf{v}\|_2^2 \le \min_i \{1 - \delta_i\} \sum_i \|\mathbf{v}_i\|_2^2 \le \sum_i (1 - \delta_i) \|\mathbf{v}_i\|_2^2 \le \sum_i \|\mathbf{A}\mathbf{v}_i\|_2^2$$

$$\le \sum_i (1 + \delta_i) \|\mathbf{v}_i\|_2^2 \le \max_i \{1 + \delta_i\} \sum_i \|\mathbf{v}_i\|_2^2 \le D \max_i \{1 + \delta_i\} \|\mathbf{v}\|_2^2.$$

$\square$

Theorem 4. *Let $\varepsilon > 0$. Let $\mathcal{W} = (W_i, P_i)_{i=1}^n$ be a fusion frame for $\mathbb{K}^N$, $N \ge 1$. Let $\mathbf{A} \in \mathbb{K}^{m \times N}$ be a subgaussian matrix with parameters β, k. Then, there exists a constant $C = C_{\beta,k}$ such that the P-RIP constants of $\frac{1}{\sqrt{m}} \mathbf{A}$ satisfy $\delta_{s_i} \le \delta_i$, for $1 \le i \le n$ with probability at least $1 - \varepsilon$, provided*

$$m \ge C \min(\delta)^{-2} \left(\max(s_i) \ln(eN/\max(s_i)) + \ln(2\varepsilon n) \right)$$

Before we prove this result, we recall a standard RIP result for subgaussian matrices (Theorem 9.2 in [24]):

Theorem 5. *Let $\mathbf{A}$ be an $m \times N$ subgaussian random matrix. Then there exists a constant $C > 0$ (depending only on subgaussian parameters β, k) such that the RIP constant of $\frac{1}{\sqrt{m}} \mathbf{A}$ satisfies $\delta_s \le \delta$ with probability at least $1 - \varepsilon$, if*

$$m \ge C \delta^{-2} \left(s \ln(eN/s) + \ln(2\epsilon^{-1}) \right).$$

Proof. (of Theorem 4)

For some constants $\delta_i \in (0, 1)$, let E be the event "$\frac{1}{\sqrt{m}}\mathbf{A}$ does not satisfy P-RIP with respect to $\mathcal{W}$ with constants $\delta_1, \cdots, \delta_n$". Applying a union bound it follows that

$$\mathbb{P}(E) = \mathbb{P}(\exists i \in \{1, \cdots, n\} : \delta_{s_i} > \delta_i) \leq \sum_{i=1}^n \mathbb{P}(\delta_{s_i} > \delta_i) = \sum_{i=1}^n \varepsilon_i$$

For some ε_i such that $\varepsilon_1 + \cdots + \varepsilon_n = \varepsilon \in (0, 1)$, since the sensing matrix is the same for every sensors, there exists a unique $C > 0$ (depending on β, k) such that $\frac{1}{\sqrt{m}}\mathbf{A}$ satisfies the RIP locally within the subset W_i and for a sparsity s_i with constant $\delta_{s_i} \leq \delta_i$, provided that $m \geq \max_{1 \leq i \leq n} m_i$, with

$$m_i \geq C\delta_i^{-2}(s_i \ln(eN/s_i) + \ln(2\epsilon_i^{-1})).$$

Additionally, since the function $s \to s \ln(eN/s)$ is monotonically increasing on $(0, N)$ and $s \ll N$, $\max_{1 \leq i \leq n} m_i \geq C \max_{1 \leq i \leq n} \delta_i^{-2} \left(\max_{1 \leq i \leq n} s_i \ln(eN/\max_{1 \leq i \leq n} s_i) + \ln(2\varepsilon_i^{-1}) \right)$. Considering $\varepsilon_1 = \cdots = \varepsilon_n = \varepsilon/n$ concludes the proof. $\qquad\square$

Note: all Gaussian and Bernoulli random matrices are subgaussian random matrices, so Theorem 4 holds true for Gaussian and Bernoulli random matrices.

4.2. **Recovery in general fusion frames settings.** With the tools introduced above, we show that any signals with sparsity pattern $\mathbf{s}$ can be recovered in a stable and robust manner via the fusion frame approach described in the previous sections.

4.2.1. *RP-NSP based results.* We give a first recovery guarantee based on the robust partial NSP, introduced in Definition 5.

Theorem 6. *Let $\mathbf{A} \in \mathbb{K}^{m \times N}$ and $\mathcal{W} = (W_i, P_i)_{i=1}^n$ a fusion frame with frame bounds $0 < C \leq D < \infty$ and frame operator S. Let $\left(\mathbf{y}^{(i)}\right)_{i=1}^n$ be the linear measurements $\mathbf{y}^{(i)} = \mathbf{A}P_i\mathbf{x} + \mathbf{e}^{(i)}$, $1 \leq i \leq n$ for some noise vectors $\mathbf{e}^{(i)}$ such that $\|\mathbf{e}^{(i)}\|_2 \leq \eta_i$. Denote by $\widehat{\mathbf{x}^{(i)}}$ the solution to the local Basis Pursuit problems $(\mathcal{P}_{1,\eta})$ and let $\widehat{\mathbf{x}} = S^{-1} \sum_i \widehat{\mathbf{x}^{(i)}}$. If the matrix $\mathbf{A}$ satisfies the ℓ^1-RP-NSP with sparsity pattern $\mathbf{s} = (s_1, \cdots, s_n)$ with constants $0 < \rho_1, \cdots, \rho_n < 1$ and $\tau_1, \cdots, \tau_n > 0$ with respect to $\mathcal{W}$, then the estimation $\widehat{\mathbf{x}}$ approximates $\mathbf{x}$ in the following sense:*

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2}{C} \left(\langle \vec{\rho}, \sigma_{\mathbf{s}}^{\mathcal{W}}(\mathbf{x})_1 \rangle + \langle \vec{\tau}, \vec{\eta} \rangle \right), \tag{37}$$

where $\vec{\rho} = \left(\frac{1+\rho_i}{1-\rho_i}\right)_{i=1}^n$, $\vec{\tau} = \left(\frac{2\tau_i}{1-\rho_i}\right)_{i=1}^n$, and $\vec{\eta} = (\eta_i)_{i=1}^n$.

Proof. The solution is given by the fusion process $\widehat{\mathbf{x}} = S^{-1} \left(\sum_{i=1}^n \widehat{\mathbf{x}^{(i)}} \right)$ with $\widehat{\mathbf{x}^{(i)}}$ the solutions to the local problems $(\mathcal{P}_{1,\eta})$. It holds

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 = \left\| S^{-1} \left(\sum_{i=1}^n P_i\mathbf{x} - \sum_{i=1}^n \widehat{\mathbf{x}^{(i)}} \right) \right\|_2 \leq C^{-1} \sum_{i=1}^n \left\| P_i\mathbf{x} - \widehat{\mathbf{x}^{(i)}} \right\|_2 \leq C^{-1} \sum_{i=1}^n \left\| P_i\mathbf{x} - \widehat{\mathbf{x}^{(i)}} \right\|_1.$$

For a particular $i \in \{1, \cdots, n\}$, we estimate the error on the subspace W_i in the ℓ^1 sense. We follow the proof techniques from [24, Section 4.3] with the adequate changes. With $\mathbf{v} := P_i\mathbf{x} - \widehat{\mathbf{x}^{(i)}}$ and $S_i \subset W_i$ the set of best s_i components of $\mathbf{x}$ supported on W_i, the ℓ^1-RP-NSP yields

$$\|(P_i\mathbf{v})_{S_i}\|_1 \leq \rho_i \left\| (P_i\mathbf{v})_{\overline{S_i}} \right\|_1 + \tau_i \|\mathbf{A}\mathbf{v}\|_2.$$

Combining with [24, Lemma 4.15] stating

$$\left\| (P_i\mathbf{v})_{\overline{S_i}} \right\|_1 \leq \left\| P_i\widehat{\mathbf{x}^{(i)}} \right\|_1 - \|P_i\mathbf{x}\|_1 + \left\| (P_i\mathbf{v})_{S_i} \right\|_1 + 2 \left\| (P_i\mathbf{x})_{\overline{S_i}} \right\|_1.$$

we arrive at

$$(1 - \rho_i) \left\| (P_i\mathbf{v})_{\overline{S_i}} \right\|_1 \leq \|P_i\widehat{\mathbf{x}^{(i)}}\|_1 - \|P_i\mathbf{x}\|_1 + 2 \left\| (P_i\mathbf{x})_{\overline{S_i}} \right\|_1 + \tau_i \|\mathbf{A}\mathbf{v}\|_2.$$

Applying once again the ℓ^1-RP-NSP, it holds

$$\|P_i\mathbf{v}\|_1 = \|(P_i\mathbf{v})_{S_i}\|_1 + \left\| (P_i\mathbf{v})_{\overline{S_i}} \right\|_1 \leq \rho_i \left\| (P_i\mathbf{v})_{\overline{S_i}} \right\|_1 + \tau_i \|\mathbf{A}\mathbf{v}\|_2 + \left\| (P_i\mathbf{v})_{\overline{S_i}} \right\|_1$$

$$\leq (1 + \rho_i) \left\| (P_i\mathbf{v})_{\overline{S_i}} \right\|_1 + \tau_i \|\mathbf{A}\mathbf{v}\|_2$$

$$\leq \frac{1 + \rho_i}{1 - \rho_i} \left(\|P_i\widehat{\mathbf{x}^{(i)}}\|_1 - \|P_i\mathbf{x}\|_1 + 2 \left\| (P_i\mathbf{x})_{\overline{S_i}} \right\|_1 \right) + \frac{4\tau_i}{1 - \rho_i} \|\mathbf{A}\mathbf{v}\|_2.$$

We now remember Lemma 2 and notice that $P_i \widehat{\mathbf{x}^{(i)}} = \widehat{\mathbf{x}^{(i)}}$. $\widehat{\mathbf{x}^{(i)}}$ being the optimal solution to $(\mathcal{P}_{1,\eta})$, it is clear that $\|\widehat{\mathbf{x}}\|_1 \leq \|P_i \mathbf{x}\|_1$ from what we can conclude that

$$\left\| P_i \mathbf{x} - \widehat{\mathbf{x}^{(i)}} \right\|_1 = \|P_i \mathbf{v}\|_1 \leq 2 \frac{1+\rho_i}{1-\rho_i} \sigma_{\mathbf{s}}^{\mathcal{W}}(\mathbf{x})_{1,i} + \frac{4\tau_i}{1-\rho_i} \|\mathbf{A}\mathbf{v}\|_2.$$

Summing up the contributions for all i in $\{1, \cdots, n\}$ and applying the inverse frame operator finishes the proof.   □

Similarly, assuming ℓ^q-RP-NSP, one can adapt the proof techniques from [24, Theorems 4.22, 4.25] to the local problems. This yields the following result

Theorem 7. *Let $\mathbf{A} \in \mathbb{K}^{m \times N}$ and $\mathcal{W} = (W_i, P_i)_{i=1}^n$ a fusion frame with frame bounds $0 < C \leq D < \infty$ and frame operator S. Let $\left(\mathbf{y}^{(i)}\right)_{i=1}^n$ be the linear measurements $\mathbf{y}^{(i)} = \mathbf{A}P_i\mathbf{x} + \mathbf{e}^{(i)}$, $1 \leq i \leq n$ for some noise vectors $\mathbf{e}^{(i)}$ such that $\|\mathbf{e}^{(i)}\|_2 \leq \eta_i$. Denote by $\widehat{\mathbf{x}^{(i)}}$ the solution to the local Basis Pursuit problems $(\mathcal{P}_{1,\eta})$ and let $\widehat{\mathbf{x}} = S^{-1} \sum_i \widehat{\mathbf{x}^{(i)}}$. If the matrix $\mathbf{A}$ satisfies the ℓ^2-RP-NSP with sparsity pattern $\mathbf{s} = (s_1, \cdots, s_n)$ with constants $0 < \rho_1, \cdots, \rho_n < 1$ and $\tau_1, \cdots, \tau_n > 0$ with respect to $\mathcal{W}$, then the estimation $\widehat{\mathbf{x}}$ approximates $\mathbf{x}$ in the following sense:*

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_p \leq \frac{1}{C} \left(\frac{\langle \vec{\rho}, \sigma_{\mathbf{s}}^{\mathcal{W}}(\mathbf{x})_1 \rangle}{s^{1-1/p}} + \frac{\langle \vec{\tau}, \vec{\eta} \rangle}{s^{1/2-1/p}} \right), \quad 1 \leq p \leq 2, \tag{38}$$

where $\vec{\rho} = \left(\frac{2(1+\rho_i)^2}{1-\rho_i} \right)_{i=1}^n$, $\vec{\tau} = \left(\frac{3-\rho_1}{1-\rho_i} \tau_i \right)_{i=1}^n$, and $\vec{\eta} = (\eta_i)_{i=1}^n$.

4.2.2. *P-RIP based recovery.* One can show that the P-RIP is sufficient for stable and robust recovery by combining Theorem 6 with the following result, showing the existence of random matrices satisfying the RP-NSP, from an RIP argument.

Theorem 8. *Let $\mathbf{A} \in \mathbb{K}^{m \times N}$ be a matrix satisfying the P-RIP(2$\mathbf{s}$,δ), with $\mathbf{s} = (s_1, \cdots, s_n)$ and $\delta = (\delta_1, \cdots, \delta_n)$ and $\delta_i < 4/\sqrt{41}$, for all $1 \leq i \leq n$. Then, $\mathbf{A}$ satisfies the ℓ^2-RP-NSP with constants $(\rho_i, \tau_i)_{i=1}^n$ where*

$$\begin{aligned} \rho_i &:= \frac{\delta_i}{\sqrt{1-\delta_i^2} - \delta_i/4} < 1 \\ \tau_i &:= \frac{\sqrt{1+\delta_i}}{\sqrt{1-\delta_i^2} - \delta_i/4}. \end{aligned} \tag{39}$$

Proof. The proof of this results consists in simply applying [24, Theorem 6.13] to every subspaces independently.   □

From this result and Theorem 7, it follows.

Theorem 9. *Let $\mathcal{W} = (W_i, P_i)_{i=1}^n$ be a fusion frame for $\mathbb{K}^N$ with frame operator S and frame bounds $0 < C \leq D < \infty$. Let $\mathbf{A} \in \mathbb{K}^{m \times N}$ be a matrix satisfying the P-RIP(2$\mathbf{s}$, δ) where $\mathbf{s} = (s_1, \cdots, s_n)$ and $\delta = (\delta_1, \cdots, \delta_n)$ with $\delta_i < 4/\sqrt{41}$, for all $1 \leq i \leq n$. Then any distributed-sparse vector $\mathbf{x} \in \Sigma_{\mathbf{s}}^{(\mathcal{W})}$ can be recovered by solving n (BPDN) problems.*

Assuming the noise in each (BPDN) problem is controlled by $\|\mathbf{e}^{(i)}\|_2 \leq \eta_i$, $1 \leq i \leq n$, and set $vec\eta = (\eta_i)_{i=1}^n$. Let $\widehat{\mathbf{x}} = S^{-1} \sum \widehat{\mathbf{x}^{(i)}}$. Then

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{1}{C} \sum_{i=1}^n \alpha_i \frac{\sigma_{\mathbf{s}}^{\mathcal{W}}(\mathbf{x})_{1,i}}{\sqrt{s_i}} + \beta_i \eta_i$$

where α_i and β_i depend only on the RIP constants δ_i.

5. LOCAL SPARSITY IN GENERAL DICTIONARIES AND FRAMES

The redundancy inherent to frame structures (and their generalization to frames of subspaces) makes them appealing to signal analysis task. So far, we have used the redundancy of the fusion frame process in order to increase the global sparsity of the original vector $\mathbf{x}$ as well as increase the robustness to noise. We investigate now the use of local dictionaries in order to use the redundancy within subspaces, using the local frames for the representation of the partial information. A common scenario in applications is when $f \in \mathcal{H}$ has a sparse frame representation $f = D\mathbf{x}$, i.e. $\mathbf{x}$ is sparse, and the multiple measurements are given by

$$\mathbf{y}^{(i)} = \mathbf{A}P_i f + \mathbf{e}^{(i)} = \mathbf{A}P_i D\mathbf{x} + \mathbf{e}^{(i)}, \quad 1 \leq i \leq n.$$

Here P_i can be any projection onto a subspace of $\mathcal{H}$. In practical applications such as in SAR radar imaging, P_i can just be a projection or spatial filter onto $W_i \equiv \text{span}\{\mathbf{d}_k\}_{k \in \Omega_i}$, where $\mathbf{d}_k$ is the kth

column of D. Such an operation can potentially reduce the number of nonzero entries of $\mathbf{x}$ in the ith observation, when the respective column vectors $\mathbf{d}_k$ are in $\ker P_i$. In particular, let us denote by $\Gamma_i \equiv \{k \mid k \notin \Omega_i, \ \mathbf{d}_k \in W_i\}$, and $\Lambda_i \equiv \{l \mid \mathbf{d}_l \in \ker P_i\}$.

$$
\begin{aligned}
P_i f &= P_i D\mathbf{x} = P_i\left(\{\mathbf{d}_k\}_{k \in \Omega_i}, \{\mathbf{d}_j\}_{j \in \Gamma_i}, \{\mathbf{d}_l\}_{l \in \Lambda_i}\right)\mathbf{x} \\
&= \left(\{\mathbf{d}_k\}_{k \in \Omega_i}, \{y_j = P_i\mathbf{d}_j\}_{j \in \Gamma_i}, \{0's\}_{l \in \Lambda_i}\right)\begin{pmatrix} \mathbf{x}_{\Omega_i} \\ \mathbf{x}_{\Gamma_i} \\ \mathbf{x}_{\Lambda_i} \end{pmatrix} \\
&= \left(\{\mathbf{d}_k\}_{k \in \Omega_i}, \{y_l = P_i\mathbf{d}_l\}_{l \in \Gamma_i}\right)\begin{pmatrix} \mathbf{x}_{\Omega_i} \\ \mathbf{x}_{\Gamma_i} \end{pmatrix} \\
&= D_i\mathbf{x}^{(i)},
\end{aligned}
$$

where $D_i \equiv (\{\mathbf{d}_k\}_{k \in \Omega_i}, \{y_k = P_i\mathbf{d}_k\}_{k \in \Gamma_i})$, and

$$
\mathbf{x}^{(i)} \equiv \begin{pmatrix} \mathbf{x}_{\Omega_i} \\ \mathbf{x}_{\Gamma_i} \end{pmatrix}.
$$

As a result, the ith measurement becomes

$$
\mathbf{y}^{(i)} = \mathbf{A}D_i\mathbf{x}^{(i)} + \mathbf{e}^{(i)}, \quad 1 \le i \le n,
$$

or

$$
\mathbf{y}^{(i)} = \mathbf{A}f^{(i)} + \mathbf{e}^{(i)}, \quad f^{(i)} = D_i\mathbf{x}^{(i)}, \quad 1 \le i \le n.
$$

Note that the first version suggests the use of ℓ^1 synthesis methods, while the second one looks at ℓ^1 analysis tools. ℓ^1 synthesis corresponds to the usual sparse recovery, via a dictionary D:

$$
\min_{\mathbf{z}} \|\mathbf{z}\|_0, \quad \text{subject to } \|\mathbf{A}D\mathbf{z} - \mathbf{y}\|_2 \le \eta.
$$

The solution $\widehat{f}$ is later computed as $\widehat{f} = D\widehat{\mathbf{x}}$. In the ℓ^1 analysis approach, we do not care for a particular (sparse) representation of f. We just ask for this representation to have high fidelity with the data:

$$
\min_{\mathbf{g}} \|D^*\mathbf{g}\|_0, \quad \text{subject to } \|\mathbf{A}g - \mathbf{y}\|_2 \le \eta \tag{40}
$$

Here, D^* denotes the canonical dual frame. Both approaches are further detailed in the next sections. We comment that if the choice of P_i is allowed, one strategy is again to use random projections by randomly selecting the index set Ω_i to set the subspaces $W_i = \text{span}\{\mathbf{d}_j\}_{j \in \Omega_i}$.

5.1. **Recovery via general ℓ^1-analysis method.** As introduced above, we try to recover f that has **a** sparse representation by solving Problem (40). While the problem is written in terms of the canonical dual frame, there is no obligation in using this particular dual frame. One may instead optimize the dual frame considered and use the sparsity-inducing dual frame [30], computed as part of the optimization problem:

$$
\widehat{f^{(i)}} = \text{argmin}_{g, D\tilde{D}_i^* = \mathbf{I}} \|\tilde{D}_i^* g\|_1 \quad \text{s.t. } \|\mathbf{A}g - \mathbf{y}^{(i)}\|_2 \le \eta_i, \quad 1 \le i \le n. \tag{41}
$$

The sparsity-inducing frame $\tilde{D}_i$ can be uniform across all i but not necessarily. The following result is known to hold for any dual frame [30].

Theorem 10. *Let D be a general frame of $\mathbb{R}^N$ with frame bounds $0 < A \le B < \infty$. Let $\tilde{D}$ be an alternative dual frame of D with frame bounds $0 < \tilde{A} \le \tilde{B} < \infty$, and let $\rho = s/b$. Suppose that the matrix $\mathbf{A}$ satisfies the following D-RIP condition*

$$
\left(1 - \sqrt{\rho B\tilde{B}}\right)^2 \cdot \delta_{s+a} + \rho B\tilde{B} \cdot \delta_b < 1 - 2\sqrt{\rho B\tilde{B}} \tag{42}
$$

for some positive integers a and b satisfying $0 < b - a \le 3a$. Let $\widehat{f}$ be the solution to the typical ℓ^1-analysis problem

$$
\widehat{f} = \text{argmin}_g \|\tilde{D}^* g\|_1 \quad \text{s.t. } \|\mathbf{A}g - \mathbf{y}\|_2 \le \eta.
$$

Then

$$
\|\widehat{f} - f\|_2 \le \alpha\eta + \beta\frac{\|\tilde{D}^* f - (\tilde{D}^* f)_{s_i}\|_1}{\sqrt{s}}, \tag{43}
$$

where α and β are some constants and $(\tilde{D}^* f)_s$ denotes the vector consisting the s largest entries in magnitude of $\tilde{D}^* f$.

In particular, the bound (43) applied to the local pieces of information $\widehat{f^{(i)}}$ analyzed with the local frames $\tilde{D}_i^*$, for $1 \leq i \leq n$, obtained as solution to Problem (41) yields the following error bound for the reconstruct signal $\widehat{f} = S^{-1} \sum_{i=1}^n \widehat{f^{(i)}}$:

$$\|\widehat{f} - f\|_2 \leq \|S^{-1}\|_{2\to2} \sum_{i=1}^n \|\widehat{f^{(i)}} - f^{(i)}\|_2 \leq \alpha\eta + \beta \left(\sum_{i=1}^n \frac{\|\tilde{D}_i^* f^{(i)} - (\tilde{D}_i^* f^{(i)})_{s_i}\|_1}{\sqrt{s_i}} \right),$$

where $\alpha \equiv \sup_i \alpha_i \|S^{-1}\|_{2\to2}$, $\eta \equiv \sum_i \eta_i$, and $\beta \equiv \sup_i \beta_i \|S^{-1}\|_{2\to2}$ This bound is reminiscent of the traditional bounds in CS where the (local) error decays as $\sigma_s(\mathbf{x})_1/\sqrt{s}$ except that the contributions of each subspace are added together and normalized by the norm of the inverse frame operator.

5.2. **Recovery via ℓ^1 synthesis method.** The gap between recovery via ℓ^1 synthesis and ℓ^1 analysis has long been studied. Interestingly, it can be shown [29] that, when using the sparsity-inducing frames described above (41), both approaches are equivalent. Denote by $\tilde{D}_{i,o}$ the resultant optimal dual frame and suppose that $\mathbf{A}$ satisfies a D-RIP property (42). Then it follows from [29]

$$\|\widehat{f^{(i)}} - f^{(i)}\|_2 \leq \alpha_i \eta_i + \beta_i \frac{\|\tilde{D}_{i,o}^* f^{(i)} - (\tilde{D}_{i,o}^* f^{(i)})_{s_i}\|_1}{\sqrt{s_i}}$$

for some positive constants α_i and β_i. Considering the fusion of the local information $\widehat{f} = S^{-1} \left(\sum_{i=1}^n \widehat{f^{(i)}} \right)$, the following result holds true:

Proposition 6. *Let S be the invertible fusion frame operator. Suppose $\mathbf{A}$ satisfies condition (42). Then the fused solution $\widehat{f}$ has an error bound given by*

$$\|\widehat{f} - f\|_2 \leq \alpha\eta + \beta \left(\sum_{i=1}^n \frac{\|\tilde{D}_{i,o}^* f^{(i)} - (\tilde{D}_{i,o}^* f^{(i)})_{s_i}\|_1}{\sqrt{s_i}} \right),$$

where $\alpha \equiv \sup_i \alpha_i \|S^{-1}\|_{2\to2}$, $\eta \equiv \sum_i \eta_i$, and $\beta \equiv \sup_i \beta_i \|S^{-1}\|_{2\to2}$.

Proof. Write $f = S^{-1} S f = S^{-1} \sum_{i=1}^n P_i f = S^{-1} \sum_{i=1}^n f^{(i)}$. Direct computation shows

$$
\begin{aligned}
\|\widehat{f} - f\|_2 &= \|S^{-1} \sum_{i=1}^n (\widehat{f^{(i)}} - f^{(i)})\|_2 \\
&\leq \|S^{-1}\|_{2\to2} \sum_{i=1}^n \|(\widehat{f^{(i)}} - f^{(i)})\|_2 \\
&\leq \|S^{-1}\|_{2\to2} \sum_{i=1}^n \left(\alpha_i \eta_i + \beta i \frac{\|\tilde{D}_{i,o}^* f^{(i)} - (\tilde{D}_{i,o}^* f^{(i)})_{s_i}\|_1}{\sqrt{s_i}} \right) \\
&\leq \sup_i \alpha_i \|S^{-1}\|_{2\to2} \sum_{i=1}^n \eta_i + \\
&\quad \sup_i \beta_j \|S^{-1}\|_{2\to2} \left(\sum_{i=1}^n \frac{\|\tilde{D}_{i,o}^* f^{(i)} - (\tilde{D}_{i,o}^* f^{(i)})_{s_i}\|_1}{\sqrt{s_i}} \right).
\end{aligned}
$$

The result follows directly by setting $\alpha \equiv \sup_i \alpha_i \|S^{-1}\|_{2\to2}$, $\eta \equiv \sum_i \eta_i$, and $\beta \equiv \sup_i \beta_j \|S^{-1}\|_{2\to2}$. $\square$

We comment that this result is not surprising due to the equivalence between the two problems described in this section when dealing with the sparsity-inducing dual frames.

5.3. **Wavelet frames and recovery of Doppler signals.** This section is intended as an illustrative example and proof-of-concept of the tools developed so far. In particular, we want to show that using a fairly *poor* quality device, we are capable to recover a signal of fairly high complexity. For the sake of reproducible research, all the experiments presented in this section can be obtained and reproduced from of the named authors' Github page[4]. In this experiment, we try to recover a noisy Doppler signal (see Fig. 7(a)) using sensors with very few measurements.

To produce the Figures, we have considered a 1024 dimensional noisy Doppler signal (obtained directly via Matlab's Wavelet toolbox). We compare a traditional ℓ^1 analysis approach, where the sparsifying

---

[4]See https://github.com/jlbouchot/FFCS for all the self implemented files. These files require, as described in the README file, to have the access to a Haar matrix function and to have CVX [26, 27] installed.
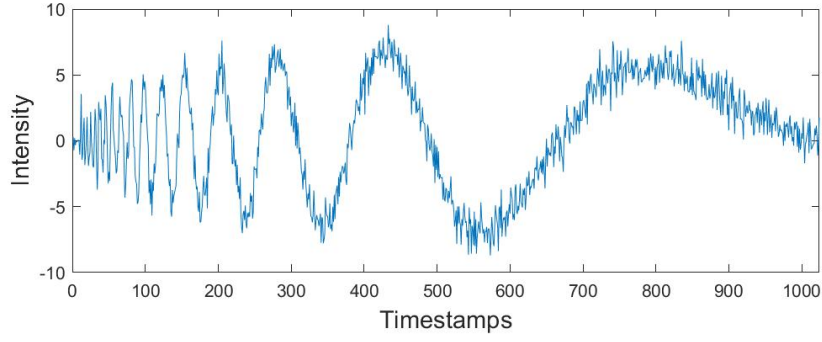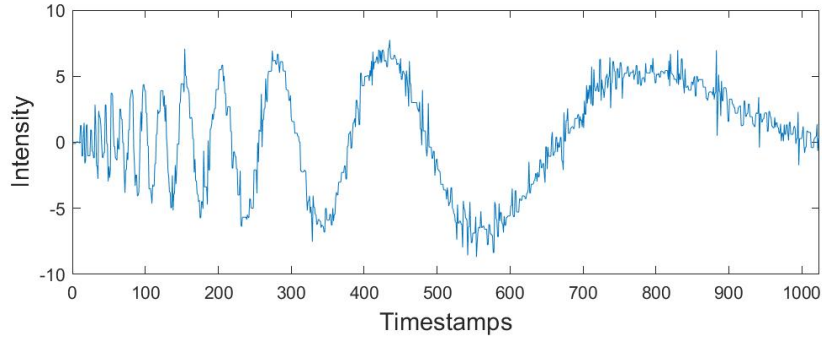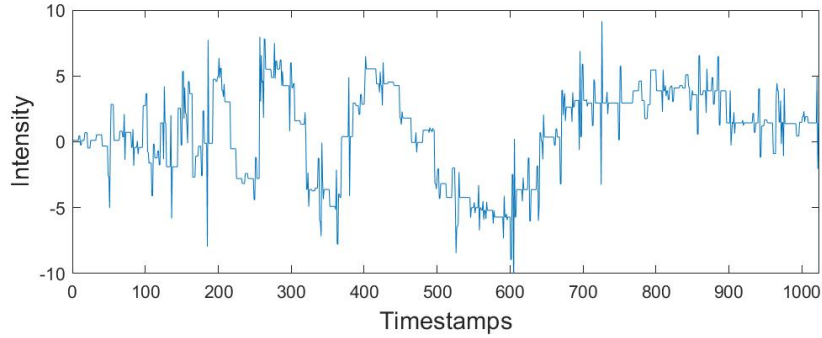
(a) Original noisy Doppler signal

(b) Recovery from a distributed sensing approach – L^2 error: 37.18

(c) Reconstruction with an ℓ^1 analysis method – L^2 error: 82.72

FIGURE 7. A noisy Doppler signal (7(a)) and its reconstructions via a Haar wavelet frame approach 7(b) and a direct ℓ^1 analysis on the Haar system 7(c)

basis is chosen as the Haar system. Similarly, we consider the Haar wavelet systems as subspaces for our fusion frame approach. Given that a signal $s(t)$ can be decomposed as

$$s(t) = a_l(t) + \sum_{j \geq l} d_j(t)$$

where a_l corresponds to the approximation coefficients at level l and d_j corresponds to the details at level j. We therefore consider our projections P_j to be onto the subspaces of details W_j at each level (using the usual definition from multiresolution analysis), and a remaining low frequency approximation part. This gives us (in our setting) a grand total of $n = 9$ projections, 8 of which $(P_1, \cdots, P_8)$ correspond to detail coefficients, and the last one (P_0) corresponds to the low-frequency approximation. This system being orthogonal, it gives us the easy reconstruction formula

$$\widehat{\mathbf{x}} = \sum_{i=0}^{8} \widehat{\mathbf{x}^{(i)}}.$$

All the local information are recovered with an ℓ^1 analysis procedure, where the sparsifying basis are precisely chosen to be the Haar wavelets at the appropriate levels. Finally, the sparsity is set uniformly

to 25 on each subspace and a random Gaussian matrix with $m = 174$ rows is created and some normal noise with variance 0.05 is added to the measurement vectors, all drawn independently from one another.

Fig. 7(b) shows the reconstructed signal using our ℓ^1-analysis fusion frame approach described in the previous section, while Fig. 7(c) is the reconstructed signal obtained from a traditional ℓ^1-analysis approach. Both of them have been reconstructed using measurements obtained from the exact same matrix. It is important to note however, that we have note tried to optimize our parameters in this example. In particular, bearing in mind that the subspaces of higher-frequency details have higher dimensions (in fact, $\text{rank}(P_j) = \mathcal{O}(2^j)$, for $1 \le j \le 8$), a better choice would be to set the local sparsity to be higher in these subspaces to improve our results. This level dependent-sampling is nothing new in the community and our results corroborate those from other approaches [2]. The Figures clearly show the reconstruction capabilities of our approach even with a sub-optimal setting and not making use of any kind of redundancy. Note that mostly the high frequency components of the signal are missing, which can be overcome by a better setting of the sparsity per subspace.

6. Conclusion

In applications, it may be the case the the sensors limit the complexity of the signals to be recovered. However, it cannot always be guaranteed that the signals fulfill this complexity requirements. We have here proposed a novel approach combining fusion frames with local recovery for highly-complex signal recovery from cheap, low-quality sensors. Motivated by such problems, we are able to prove stable and robust recovery in a uniform setting in the set of sparsely-distributed signals. This has some interesting implications. In particular, we verify that the number of measurements required locally is inversely proportional to the number of sensors. Moreover it is empirically verified that adding more sensors can add robustness to noise; this is a consequence of the redundancy in the (fusion) frame representation. By splitting the main high-complexity, dense problems into many subproblems, we recover the local information stably, robustly, and efficiently by solving a few low-dimensional problems.

Section 5 investigates redundant representations locally, by means of coherent frames. We gave results on signal recovery via the ℓ^1 synthesis and ℓ^1 analysis method, in two cases, varying in the use of the filter projection.

As a last Section, we extend the traditional single-sensor compressed sensing theory to embed this novel multi-sensor framework. It gives strong mathematical foundations to the approaches presented in this paper, and reduces to the usual compressed sensing approach when dealing with a single sensor. Our work differs from that of others as we do not have a requirement on our fusion frame to be tight, albeit solving the local problems independently.

This work can be further developed by integrating various sensors instead of a single one, repeated n times. Further real-world applications, for instance in SAR systems, can also be implemented.

References

[1] R. Aceska, J.-L. Bouchot, and S. Li. Fusion frames and distributed sparsity. Submitted to contemporary Mathematics, 2017.

[2] Ben Adcock, Anders C. Hansen, Clarice Poon, and Bogdan Roman. Breaking the coherence barrier: A new theory for compressed sensing. *In Forum of Mathematics, Sigma*, vol. 5. Cambridge University Press, 2017.

[3] U. Ayaz, S. Dirksen, and H. Rauhut. Uniform recovery of fusion frame structured sparse signals. *Applied and Computational Harmonic Analysis*, 41(2):341–361, 2016.

[4] A. S. Bandeira, K. Scheinberg, and L. N. Vicente. On partial sparse recovery. arXiv preprint arXiv:1304.2809, 2013.

[5] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[6] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.

[7] J.-L. Bouchot. A generalized class of hard thresholding algorithms for sparse signal recovery. In *Approximation Theory XIV: San Antonio 2013*, pages 45–63. Springer, 2014.

[8] J.-L. Bouchot, S. Foucart, and P. Hitczenko. Hard thresholding pursuit algorithms: number of iterations. *Applied and Computational Harmonic Analysis*, page pp in press, 2016.

[9] P. Boufounos, G. Kutyniok, and H. Rauhut. Sparse recovery from combined fusion frame measurements. *Information Theory, IEEE Transactions on*, 57(6):3864–3876, 2011.

[10] C. Boyer, J. Bigot, and P. Weiss. Compressed sensing with structured sparsity and structured acquisition. arXiv preprint arXiv:1505.01619, 2015.

[11] J. Cahill, P. Casazza, and S. Li. Non-orthogonal fusion frames and the sparsity of fusion frame operators. *Journal of Fourier Analysis and Applications*, **18**, no. 2: 287 – 308, 2012.

[12] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592, 2008.

[13] P. G. Casazza and G. Kutyniok. Frames of subspaces. *Contemporary Mathematics*, 345:87–113, 2004.

[14] P. G. Casazza, G. Kutyniok, and S. Li. Fusion frames and distributed processing. *Applied and Computational Harmonic Analysis*, 25:114–132, 2008.

[15] Z. Chen and J. J. Dongarra. Condition numbers of Gaussian random matrices. *SIAM Journal on Matrix Analysis and Applications*, 27(3):603–620, 2005.

[16] O. Christensen. *An introduction to frames and Riesz bases*. Springer Science & Business Media, 2003.

[17] I. Y. Chun and B. Adcock. Compressed sensing and parallel acquisition. *IEEE Transactions on Information Theory*, PP(99):1–1, 2017.

[18] I. G. Cumming and F. H.-c. Wong. *Digital processing of Synthetic Aperture Radar Data*. Artech House, 2005.

[19] J. Curlander and R. McDonough. *Synthetic Aperture Radar*. Wiley Interscience, New York, 1991.

[20] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk. Distributed compressed sensing of jointly sparse signals. In *Asilomar Conf. Signals, Sys., Comput*, pages 1537–1541, 2005.

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

[22] M. Fornasier and H. Rauhut. Compressive sensing. *Handbook of Mathematical Methods in Imaging*, pages 187–228, 2011.

[23] S. Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. *SIAM Journal on Numerical Analysis*, 49(6):2543–2563, 2011.

[24] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

[25] G. Franceschetti and R. Lanari. *Synthetic Aperture Radar Processing*. CRC Press, 1999.

[26] M. Grant and S. Boyd. *CVX: Matlab Software for Disciplined Convex Programming, version 2.1.* http://cvxr.com/cvx, 2014.

[27] M. Grant and S. Boyd. *Graph implementations for nonsmooth convex programs*, Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences, pp 95–110, 2008.

[28] 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. *Journal of Fourier analysis and Applications*, 14(5-6):655–687, 2008.

[29] Y. Liu, S. Li, T. Mi, H. Lei, and W. Yu. Performance analysis of ℓ_1-synthesis with coherent frames. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2042–2046. IEEE, 2012.

[30] Y. Liu, T. Mi, and S. Li. Compressed sensing with general frames via optimal-dual-based-analysis. *Information Theory, IEEE Transactions on*, 58(7):4201–4214, 2012.

[31] D. Needell and J. A. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[32] H. Rauhut. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92, 2010.

[33] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *Information Theory, IEEE Transactions on*, 54(5):2210–2219, 2008.

[34] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics*, 62(12):1707–1739, 2009.

[35] C. Zeng and S. Li. Sparse measurements and fused sar imaging. in preparation, 2015.

[36] T. Zhang. Sparse recovery with orthogonal matching pursuit under rip. *Information Theory, IEEE Transactions on*, 57(9):6215–6221, 2011.

DEPARTMENT OF MATHEMATICAL SCIENCES, BALL STATE UNIVERSITY
E-mail address: `raceska@bsu.edu`

CHAIR FOR MATHEMATICS C (ANALYSIS), RWTH AACHEN UNIVERSITY
E-mail address: `bouchot@mathc.rwth-aachen.de`

DEPARTMENT OF MATHEMATICS, SAN FRANCISCO STATE UNIVERSITY
E-mail address: `shidong@math.sfsu.edu`