Analysis of data separation and recovery problems using clustered sparsity

Emily J. King[a], Gitta Kutyniok[a], and Xiaosheng Zhuang[a]

[a] Institute of Mathematics, Universität Osnabrück, Osnabrück, Germany

ABSTRACT

Data often have two or more fundamental components, like cartoon-like and textured elements in images; point, filament, and sheet clusters in astronomical data; and tonal and transient layers in audio signals. For many applications, separating these components is of interest. Another issue in data analysis is that of incomplete data, for example a photograph with scratches or seismic data collected with fewer than necessary sensors. There exists a unified approach to solving these problems which is minimizing the ℓ_1 norm of the analysis coefficients with respect to particular frame(s). This approach using the concept of clustered sparsity leads to similar theoretical bounds and results, which are presented here. Furthermore, necessary conditions for the frames to lead to sufficiently good solutions are also shown.

Keywords: ℓ_1 minimization, cluster coherence, geometric separation, inpainting, Parseval frames, sparse representation, data recovery, clustered sparsity

1. INTRODUCTION

Data analysts in varied fields often face the task of *geometric separation*. Namely, data may be superpositions of various types of structures which the scientist would like to separate. For example, gravitation causes 3-d data to concentrate near lower-dimensional structures such as points, filaments, and sheets. One aspiration of cosmological data analysis is to be able to extract these three "pure" elements of matter density [1, 2]. Separating texture and the piecewise smooth parts of images [3–5] and decomposing a single-channel audio signal into tonal and transient layers [6, 7] are both examples of similar problems in other fields. Astronomers have recently presented empirical evidence that geometric separation can be achieved by using two or more overcomplete frames [1, 2].

Another issue that arises in data analysis is that of missing data. Due to land development and bodies of water, it is not always possible to place sensors at all necessary locations when making seismic measurements [8,9]; however, the complete set of "actual" data is desired. Data recovery in images and videos is called *inpainting*, a term used by conservators working with damaged paintings. The removal of overlaid text in images, the repair of scratched photos and audio recordings, and the recovery of missing blocks in a streamed video are all examples of inpainting. Variational inpainting methods are commonly used [10–13], but we shall prove results about a technique based on sparse representations as in [14] and [15].

Surprisingly, these seemingly disparate problems share common approaches and very similar theoretical results. These similarities will become clear in what follows. We shall explicitly set up the

Send correspondence to Emily J. King, eking@math.umd.edu, +49 (0)541 969 2471.
Further contact information: Gitta Kutyniok, kutyniok@uos.de, +49 (0)541 969 3516; Xiaosheng Zhuang, xzhuang@uos.de, +49 (0)541 969-3508

Wavelets and Sparsity XIV, edited by Manos Papadakis, Dimitri Van De Ville, Vivek K. Goyal,
Proc. of SPIE Vol. 8138, 813818 · © 2011 SPIE · CCC code: 0277-786X/11/$18
doi: 10.1117/12.892723

problems of geometric separation and data recovery in Section 2. In Section 3 the auxiliary notions of cluster coherence and (joint) concentration are presented. These notions will be used to formulate the bounds of the solutions to the geometric separation (Section 4) and data recovery (Section 5) problems.

2. PROBLEM FORMULATION

2.1 Notation

We comment here on our notation. A collection of vectors $\Phi = \{\varphi_i\}_{i\in I}$ in a separable Hilbert space $\mathcal{H}$ forms a *Parseval frame* for $\mathcal{H}$ if for all $x \in \mathcal{H}$,

$$\sum_{i\in I} |\langle x, \varphi_i\rangle|^2 = \|x\|^2 .$$

With a slight abuse of notation, given a Parseval frame Φ, we also use Φ to denote the *synthesis operator*

$$\Phi : \ell_2(I) \to \mathcal{H}, \quad \Phi(\{c_i\}_{i\in I}) = \sum_{i\in I} c_i \varphi_i.$$

With this notation, Φ^* is called the *analysis operator.*

Although our results concern Parseval frames, the following related definitions will be used when discussing prior results. With notation as above, Φ is a *frame* for $\mathcal{H}$ if there exist $0 < A \le B < \infty$ such that for all $x \in \mathcal{H}$,

$$A\|x\|^2 \le \sum_{i\in I} |\langle x, \varphi_i\rangle|^2 \le B\|x\|^2 .$$

If there exists some $c \neq 0$ such that $\|\varphi_i\|_2 = c$ for all $i \in I$, then we call Φ a *normalized frame.*

Given a space X and a subset $A \subseteq X$, we will use the notation A^c to denote $X\backslash A$. Also, the *indicator function* 1_A is defined to take the value 1 on A and 0 on A^c.

2.2 Geometric Separation

Let x be our signal of interest, which belongs to some Hilbert space $\mathcal{H}$, and assume that

$$x = x_1^0 + x_2^0.$$

We assume that although we are not given x_1^0 and x_2^0, certain "characteristics" of those components are known to us. Such "characteristics" might be, for instance, the pointlike structure of stars and the curvelike structure of filaments in astronomical imaging. This knowledge now enables us to choose two representation systems, say Φ_1 and Φ_2, which allow sparse expansions of x_1^0 and x_2^0, respectively (see also [16, 17]). The measure of sparsity is $x \mapsto \|x\|_0$, which counts the number of non-zero entries of a vector x. Such representation systems might be chosen from the collection of well-known systems such as wavelets or shearlets; alternatively, one could choose the systems adaptively through dictionary learning procedures such as K-SVD or MOD [18–20]. However, this approach requires training data sets for the two components x_1^0 and x_2^0 and also does not always yield representation systems which are frames, which is a trait that we desire in the Φ_i. Of course, real data could be composed of more than two "natural" components. We will focus on the two-component situation for clarity but mention that most of the theoretical results presented can be extended to the multiple-component situation (see also [21]).

Given two appropriate representation systems Φ_1 and Φ_2, we can write x as

$$x = x_1^0 + x_2^0 = \Phi_1 c_1^0 + \Phi_2 c_2^0 = [\, \Phi_1 \mid \Phi_2 \,][\, c_1 \mid c_2 \,]^T$$

with $\|c_1^0\|_0$ and $\|c_2^0\|_0$ "sufficiently small." Thus, the data separation problem has been reduced to solving an underdetermined linear system. Unique recovery of the original vector $[c_1^0, c_2^0]^T$ automatically extracts the correct two components x_1^0 and x_2^0 from x, since

$$x_1^0 = \Phi_1 c_1^0 \quad \text{and} \quad x_2^0 = \Phi_2 c_2^0.$$

Ideally, one might want to solve

$$\min_{c_1, c_2} \|c_1\|_0 + \|c_2\|_0 \quad \text{s.t.} \quad x = [\, \Phi_1 \mid \Phi_2 \,][\, c_1 \mid c_2 \,]^T, \tag{1}$$

but this is an NP-hard problem. Under certain circumstances (see, for example, [22, 23]), the ℓ_1 minimization problem

$$\min_{c_1, c_2} \|c_1\|_1 + \|c_2\|_1 \quad \text{s.t.} \quad x = [\, \Phi_1 \mid \Phi_2 \,][\, c_1 \mid c_2 \,]^T \tag{2}$$

yields the same or appropriately close solution. Note that the ℓ_1 norm is placed on the synthesis side. However, (2) is still not the final form of the geometric separation optimization problem. There is no reason to assume that the Φ_i are bases. Some well-known representation systems are in fact redundant and typically constitute Parseval frames such as wavelets, shearlets, or curvelets. Also, systems generated by dictionary learning are typically highly redundant. In this situation, for each possible separation

$$x = x_1 + x_2, \tag{3}$$

there exist infinitely many coefficient sequences $[c_1, c_2]^T$ satisfying

$$x_1 = \Phi_1 c_1 \quad \text{and} \quad x_2 = \Phi_2 c_2. \tag{4}$$

Solving (2) can be numerically unstable for certain representation systems. Since we are *only* interested in the correct separation and *not* in computing the sparsest expansion, we can circumvent possible problems by solving the separation problem by selecting particular coefficient sequences $\tilde{c}_i$ which expand out to the x_i for each separation. Assuming Φ_1 and Φ_2 are Parseval frames, we can exploit this structure and rewrite (4) as

$$x_1 = \Phi_1(\Phi_1^* x_1) \quad \text{and} \quad x_2 = \Phi_2(\Phi_2^* x_2).$$

Thus, for each separation (3), we choose a *specific* coefficient sequence when expanding the components in the Parseval frames, in fact, we choose the *analysis sequence*. This leads to the following different ℓ_1 minimization problem in which the ℓ_1 norm is placed on the *analysis* rather than the *synthesis* side:

$$(\textsc{Sep}) \qquad (x_1^\star, x_2^\star) = \operatorname{argmin}_{x_1, x_2} \|\Phi_1^* x_1\|_1 + \|\Phi_2^* x_2\|_1 \quad \text{s.t.} \quad x = x_1 + x_2. \tag{5}$$

This new minimization problem can be also regarded as a mixed ℓ_1-ℓ_2 problem [16], since the analysis coefficient sequence is exactly the coefficient sequence which is minimal in the ℓ_2 norm. This minimization on the analysis side is also called optimization of *cosparsity* [24].

Real data are typically not free of noise. We now assume that we only know $\tilde{x} = x + n$, where n is unknown but assumed to be small; that is, for $i = 1$ or 2, $\|\Phi_i^* n\|_1 \le \epsilon$. The formulation of (5) with noise is

$$(\textsc{SepNoise}) \qquad (\tilde{x}_1^\star, \tilde{x}_2^\star) = \operatorname{argmin}_{x_1, x_2} \|\Phi_1^* x_1\|_1 + \|\Phi_2^* x_2\|_1 \quad \text{s.t.} \quad \tilde{x} = x_1 + x_2. \tag{6}$$

2.3 Recovery of Missing Data

Here x^0 is our signal of interest, which belongs to some Hilbert space $\mathcal{H}$. We assume that $\mathcal{H}$ can be decomposed into a direct sum $\mathcal{H} = \mathcal{H}_M \oplus \mathcal{H}_K$, where $\mathcal{H}_M$ is associated to the missing part of the signal x^0 and $\mathcal{H}_K$ to the known part. The respective orthogonal projections onto these subspaces are P_K and P_M. We assume that only $P_K x^0$ is known to us. We would like to reconstruct the full signal x^0. In order to do so, given a frame Φ for $\mathcal{H}$, we could consider the following optimization problem

$$x^\star = \Phi c^\star, \quad c^\star = \operatorname{argmin}_c \|c\|_1 \quad \text{s.t.} \quad P_K \Phi c = P_K x^0.$$

However, as with the geometric separation problem, we are concerned with the end product – namely, the complete signal x^0 – and not how we get there. Thus, we solve the optimization problem on the analysis side:

$$(\textsc{Rec}) \quad x^\star = \operatorname{argmin}_x \|\Phi^* x\|_1 \quad \text{s.t.} \quad P_K x = P_K x^0. \tag{7}$$

A similar approach is taken to inpainting in [14]. Assume now that we know $\tilde{x} = P_K x^0 + n$, where x^0 and n are unknown but n is assumed to be small. That is, $\|\Phi^* n\|_1 \leq \epsilon$. Also, clearly $n = P_K n$. Then we solve

$$(\textsc{RecNoise}) \quad \tilde{x}^\star = \operatorname{argmin}_x \|\Phi^* x\|_1 \quad \text{s.t.} \quad P_K x = \tilde{x}. \tag{8}$$

With both the geometric separation and data recovery problems, using arbitrary frames will not lead to desirable outcomes. However, we quantify in the next section which frames will lead to meaningful results in (Rec) and (Sep). We note here that numerical algorithms have been presented which combine these two problems by separating and inpainting on each component [14, 25, 26]. While the algorithms perform well on the test images, the papers do not address the theoretical issues of which frames to employ and how close the solution of the minimization problem will be to the actual image.

3. COHERENCE AND CONCENTRATION

One possibility for capturing the appropriateness of using particular frames in (Sep) and (Rec) is *(joint) concentration.*

Definition 3.1. *Let $\Phi_j = \{\varphi_{ji}\}_{i \in I_j}$ for $j = 1 \ldots N$ be a finite collection of Parseval frames for a Hilbert space $\mathcal{H}$. Also assume that $\mathcal{G}$ is a subspace of $\mathcal{H}$. Further, let $\Lambda_j \subseteq I_j$ for each $j = 1 \ldots N$. Then the* joint concentration κ on $\mathcal{G}$ *is defined by*

$$\kappa = \kappa(\Lambda_1, \Phi_1; \ldots; \Lambda_N, \Phi_N : \mathcal{G}) = \sup_{x \in \mathcal{G}} \frac{\|1_{\Lambda_1} \Phi_1^* x\|_1 + \ldots \|1_{\Lambda_N} \Phi_N^* x\|_1}{\|\Phi_1^* x\|_1 + \ldots + \|\Phi_N^* x\|_1}.$$

When $\mathcal{G} = \mathcal{H}$, we write joint concentration *and use the notation $\kappa(\Lambda_1, \Phi_1; \ldots; \Lambda_N, \Phi_N)$. Also, when $N = 1$, we refer to κ as the* concentration on $\mathcal{G}$.

This notion was introduced for the geometric separation problem in [27] with concepts going back to [28] and [29], and the data recovery analog was first presented in [15]. The joint concentration measures the maximal fraction of the total ℓ_1 norm which can be concentrated on the index set $\Lambda_1 \cup \cdots \cup \Lambda_N$ of the combined dictionary. Adequate control of joint concentration ensures that in principle (5) gives a successful approximate separation. Similarly, the concentration is the maximal fraction of the total ℓ_1 norm which can be concentrated to the index set Λ restricted to functions in $\mathcal{H}_M$.

In many studies of ℓ_1 optimization, one utilizes the *mutual coherence*

$$\mu(\Phi_1, \Phi_2) = \max_j \max_i |\langle \varphi_{1i}, \varphi_{2j} \rangle|, \tag{9}$$

whose importance was shown by [29]. This may be called the *singleton coherence*. However, in a concrete situation, we often have more information about the geometry of the to-be-separated components x_i^0 related to the Φ_i. This information is typically encoded in a particular clustering of the non-zero coefficients in a suitable basis or frame for the expansion of one of the x_i^0. For example, wavelet coefficients of a point singularity cluster. Thus, it seems conceivable that the morphological difference is encoded not only in the incoherence of the Φ_i but in the interaction of the elements of the Φ_i associated with the clusters of significant coefficients. This leads to the following definition.

DEFINITION 3.2. *Let $\Phi_1 = \{\varphi_{1i}\}_{i\in I}$ and $\Phi_2 = \{\varphi_{2j}\}_{j\in J}$ be two Parseval frames for a Hilbert space $\mathcal{H}$ and let $\Lambda \subseteq I$. Then the* cluster coherence *$\mu_c(\Lambda, \Phi_1; \Phi_2)$ of Φ_1 and Φ_2 with respect to Λ is defined by*

$$\mu_c(\Lambda, \Phi_1; \Phi_2) = \max_{j\in J} \sum_{i\in\Lambda} |\langle \varphi_{1i}, \varphi_{2j}\rangle|.$$

An early notion of coherence adapted to the clustering of frame vectors was the *Babel function*, first introduced in [22] and later in [30] as the *cumulative coherence function*, which, for a normalized frame $\Phi = \{\varphi_i\}_{i\in I}$ and some $m \in \{1, \dots, |I|\}$ is defined by

$$\mu_B(m, \Phi) = \max_{\Lambda\subset I, |\Lambda|=m} \max_{j\notin\Lambda} \sum_{i\in I} |\langle \varphi_i, \varphi_j\rangle|.$$

This notion was later refined in [31] by considering the so-called *structured p-Babel function*, defined for some family $\mathcal{S}$ of subsets of I and some $1 \le p < \infty$ by

$$\mu_{sB}(\mathcal{S}, \Phi) = \max_{\Lambda\in\mathcal{S}} \left(\max_{j\notin\Lambda} \sum_{i\in I} |\langle \varphi_i, \varphi_j\rangle|^p \right)^{1/p}.$$

These other notions of coherence maximize over subsets Λ of a given size, whereas for cluster coherence we fix a specific set Λ of indices. In our related work, we select Λ's which have specific geometric interpretations [15, 27]. Maximizing over all subsets of a given size would give very loose bounds and would not be suitable for our purposes. Several other coherence measures involving subsets appear in the literature, e.g., [32] and [33] but do not seem to be strongly related to cluster coherence.

The relation between (joint) concentration and cluster coherence is made precise in the following result originally from [27] and generalized here. By abuse of notation we write $P_\mathcal{G}\Phi = \{P_\mathcal{G}\varphi_i\}_{i\in I}$.

PROPOSITION 3.1. *Let $\Phi_j = \{\varphi_{ji}\}_{i\in I_j}$ for $j = 1 \dots N$ be a finite collection of Parseval frames for a Hilbert space $\mathcal{H}$. Also assume that $\mathcal{G}$ is a subspace of $\mathcal{H}$ with orthogonal projection $P_\mathcal{G}$. Further, choose $\Lambda_j \subseteq I_j$ for each $j = 1 \dots N$. Let S_N denote the symmetric group on $\{1, \dots, N\}$. Then*

$$\kappa(\Lambda_1, \Phi_1; \dots; \Lambda_N, \Phi_N : \mathcal{G}) \le \min_{\sigma\in S_N} \max_{1\le j\le N} \{\mu_c(\Lambda_j, P_\mathcal{G}\Phi_j; P_\mathcal{G}\Phi_{\sigma(j)})\} = \min_{\sigma\in S_N} \max_{1\le j\le N} \{\mu_c(\Lambda_j, P_\mathcal{G}\Phi_j; \Phi_{\sigma(j)})\}.$$

Proof. Let $\sigma \in S_N$ be arbitrary. For each $f \in \mathcal{G}$, we choose for $1 \le j \le N$ coefficient sequences α_j such that $f = \Phi_j\alpha_i$ and $\|\alpha_j\|_1 \le \|\beta_j\|$ for all β_j satisfying $f = \Phi_j\beta_j$. Since each Φ_j, $1 \le j \le N$, is

Parseval, $f = \Phi_j \Phi_j^* \Phi_j \alpha_j$ and $f = (P_{\mathcal{G}} \Phi_j)\alpha_j$. We calculate

$$\begin{aligned}
\sum_{j=1}^N \left\| 1_{\Lambda_j} \Phi_j^* f \right\|_1 &= \sum_{j=1}^N \left\| 1_{\Lambda_j} (P_{\mathcal{G}} \Phi_j)^* f \right\|_1 = \sum_{j=1}^N \left\| 1_{\Lambda_j} (P_{\mathcal{G}} \Phi_j)^* (P_{\mathcal{G}} \Phi_{\sigma(j)}) \alpha_{\sigma(j)} \right\|_1 \\
&\leq \sum_{j=1}^N \left[\sum_{i \in \Lambda_j} \left(\sum_k \left| \langle P_{\mathcal{G}} \varphi_{j,i}, P_{\mathcal{G}} \varphi_{\sigma(j),k} \rangle \right| |\alpha_{\sigma(j),k}| \right) \right] \\
&= \sum_{j=1}^N \sum_k \left(\sum_{i \in \Lambda_j} \left| \langle P_{\mathcal{G}} \varphi_{j,i}, P_{\mathcal{G}} \varphi_{\sigma(j),k} \rangle \right| \right) |\alpha_{\sigma(j),k}| \\
&\leq \sum_{j=1}^N \mu_c(\Lambda_j, P_{\mathcal{G}} \Phi_j; P_{\mathcal{G}} \Phi_{\sigma(j)}) \left\| \alpha_{\sigma(j)} \right\|_1 \\
&\leq \max_{1 \leq j \leq N} \mu_c(\Lambda_j, P_{\mathcal{G}} \Phi_j; P_{\mathcal{G}} \Phi_{\sigma(j)}) \sum_{j=1}^N \left\| \alpha_{\sigma(j)} \right\|_1 \\
&= \max_{1 \leq j \leq N} \mu_c(\Lambda_j, P_{\mathcal{G}} \Phi_j; P_{\mathcal{G}} \Phi_{\sigma(j)}) \sum_{j=1}^N \left\| \Phi_j^* f \right\|_1 .
\end{aligned}$$

Since σ was arbitrary and $P_{\mathcal{G}}$ is an orthogonal projection, the proof is complete.

4. SEPARATION ESTIMATES

We now present general estimates on the separability of composed data. For real data "true sparsity" is unrealistic. Instead we present a modified idea which makes use of the clustering of significant coefficients. This notion was first utilized in [34] and is sufficient to show that the solution $(x_1^\star, x_2^\star)$ of (5) is a "good" approximation of the actual components x_i^0 of x.

Definition 4.1. *Let $\Phi_1 = \{\varphi_{1i}\}_{i \in I}$ and $\Phi_2 = \{\varphi_{2j}\}_{j \in J}$ be two Parseval frames for a Hilbert space $\mathcal{H}$, and let $\Lambda_1 \subseteq I$ and $\Lambda_2 \subseteq J$. Further, suppose that $x \in \mathcal{H}$ can be decomposed as $x = x_1^0 + x_2^0$. Then the components x_1^0 and x_2^0 are called* δ-relatively sparse in Φ_1 and Φ_2 with respect to Λ_1 and Λ_2, *if*

$$\left\| 1_{\Lambda_1^c} \Phi_1^* x_1^0 \right\|_1 + \left\| 1_{\Lambda_2^c} \Phi_2^* x_2^0 \right\|_1 \leq \delta.$$

δ-relative sparsity is a type of *clustered sparsity.* We now have all ingredients to state the data separation result from [27].

Theorem 4.1 ([27]). *Let $\Phi_1 = \{\varphi_{1i}\}_{i \in I}$ and $\Phi_2 = \{\varphi_{2j}\}_{j \in J}$ be two Parseval frames for a Hilbert space $\mathcal{H}$, and suppose that $x \in \mathcal{H}$ can be decomposed as $x = x_1^0 + x_2^0$. Further, let $\Lambda_1 \subseteq I$ and $\Lambda_2 \subseteq J$ be chosen such that x_1^0 and x_2^0 are δ-relatively sparse in Φ_1 and Φ_2 with respect to Λ_1 and Λ_2. Then the solution $(x_1^\star, x_2^\star)$ of the ℓ_1 minimization problem* (Sep) *stated in* (5) *satisfies*

$$\left\| x_1^\star - x_1^0 \right\|_2 + \left\| x_2^\star - x_2^0 \right\|_2 \leq \frac{2\delta}{1 - 2\kappa}. \tag{10}$$

Using Proposition 3.1 this result can also be stated in terms of cluster coherence, which on one hand provides an easier estimate and allows a better comparison with results using mutual coherence but on the other hand poses a slightly weaker estimate.

THEOREM 4.2 ([27]). *Let $\Phi_1 = \{\varphi_{1i}\}_{i\in I}$ and $\Phi_2 = \{\varphi_{2j}\}_{j\in J}$ be two Parseval frames for a Hilbert space $\mathcal{H}$, and suppose that $x \in \mathcal{H}$ can be decomposed as $x = x_1^0 + x_2^0$. Further, let $\Lambda_1 \subseteq I$ and $\Lambda_2 \subseteq J$ be chosen such that x_1^0 and x_2^0 are δ-relatively sparse in Φ_1 and Φ_2 with respect to Λ_1 and Λ_2. Then the solution $(x_1^\star, x_2^\star)$ of the minimization problem* (SEP) *stated in* (5) *satisfies*

$$\left\|x_1^\star - x_1^0\right\|_2 + \left\|x_2^\star - x_2^0\right\|_2 \leq \frac{2\delta}{1-2\mu_c},$$

with

$$\mu_c = \max\{\mu_c(\Lambda_1, \Phi_1; \Phi_2), \mu_c(\Lambda_2, \Phi_2; \Phi_1)\}.$$

To thoroughly understand this estimate, it is important to notice the various dependencies of the relative sparsity δ and the joint concentration κ on the Λ_i and x_i^0. In general, replacing either of the Λ_i with a superset will increase the value of κ and decrease the possible values of δ and *vice versa* when one or both Λ_i is replaced with a subset. As κ increases to $\frac{1}{2}$, the denominator of the error estimate in (10) approaches 0. However, note that while the value of κ depends on all $x \in \mathcal{H}$, δ is a bound only on the x_i^0. Thus, if each $\left\|\Phi_i^* x_i^0\right\|_1$ is truly small, then the relative sparsity will remain small regardless of the choice of the Λ_i. In [27] the Λ_i are selected based on geometric information. It is also important to realize that the sets Λ_1 and Λ_2 serve as a mere analysis tool; they do not appear in the minimization problem (SEP). Thus, the algorithm does not depend on this choice at all; however, the estimate for accuracy of separation does. We comment here that in similar theoretical results applied to a basic thresholding algorithm, the Λ_i are selected by the algorithm [27].

Theorems 4.1 and 4.2 can be generalized to include noise. The following result was presented in [27] and is tightened here.

THEOREM 4.3. *Let $\Phi_1 = \{\varphi_{1i}\}_{i\in I}$ and $\Phi_2 = \{\varphi_{2j}\}_{j\in J}$ be two Parseval frames for a Hilbert space $\mathcal{H}$, and suppose that $x \in \mathcal{H}$ can be decomposed as $x = x_1^0 + x_2^0$. Further, let $\Lambda_1 \subseteq I$ and $\Lambda_2 \subseteq J$ be chosen such that x_1^0 and x_2^0 are δ-relatively sparse in Φ_1 and Φ_2 with respect to Λ_1 and Λ_2. Further assume that only $\tilde{x} = x + n$ is known, where n satisfies $\|\Phi_i^* n\|_1 \leq \epsilon$ for either $i = 1$ or $i = 2$. Then the solution $(\tilde{x}_1^\star, \tilde{x}_2^\star)$ of the ℓ_1 minimization problem* (SEPNOISE) *stated in* (6) *satisfies*

$$\left\|\tilde{x}_1^\star - x_1^0\right\|_2 + \left\|\tilde{x}_2^\star - x_2^0\right\|_2 \leq \frac{2\delta + (5-2\kappa)\epsilon}{1-2\kappa} \leq \frac{2\delta + (5-2\kappa)\epsilon}{1-2\mu_c} \tag{11}$$

with

$$\mu_c = \max\{\mu_c(\Lambda_1, \Phi_1; \Phi_2), \mu_c(\Lambda_2, \Phi_2; \Phi_1)\}.$$

Proof. By symmetry, we can assume without loss of generality that $\|\Phi_2^* n\|_1 \leq \epsilon$. Since the Φ_i are Parseval,

$$\left\|\tilde{x}_1^\star - x_1^0\right\|_2 + \left\|\tilde{x}_2^\star - x_2^0\right\|_2 \leq \left\|\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|\Phi_2^*(\tilde{x}_2^\star - x_2^0)\right\|_1. \tag{12}$$

By the construction of the solution,

$$\tilde{x}_1^\star + \tilde{x}_2^\star = x_1^0 + x_2^0 + n \quad \Rightarrow \quad \tilde{x}_1^\star - x_1^0 = x_2^0 - \tilde{x}_2^\star + n. \tag{13}$$

Thus,

$$\left\|\Phi_2^*(\tilde{x}_2^\star - x_2^0)\right\|_1 \leq \left\|\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \|\Phi_2^* n\|_1 \leq \left\|\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \epsilon. \tag{14}$$

By the definition of κ and applying (13) again, we calculate

$$\begin{aligned}
& \left\|\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1 \qquad (15)\\
= \;& \left\|1_{\Lambda_1}\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_2}\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_1^c}\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_2^c}\Phi_2^*(\tilde{x}_2^\star - x_2^0 - n)\right\|_1\\
\leq \;& \left\|1_{\Lambda_1}\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_2}\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_1^c}\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_2^c}\Phi_2^*(\tilde{x}_2^\star - x_2^0)\right\|_1 + \|\Phi_2^* n\|_1\\
\leq \;& \kappa\left(\left\|\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1\right) + \left\|1_{\Lambda_1^c}\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_2^c}\Phi_2^*(\tilde{x}_2^\star - x_2^0)\right\|_1 + \epsilon. \qquad (16)
\end{aligned}$$

Rearranging (16) and using the relative sparsity of the systems, we get

$$\begin{aligned}
\left\|\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1 &= \frac{1}{1-\kappa}\left(\left\|1_{\Lambda_1^c}\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_2^c}\Phi_2^*(\tilde{x}_2^\star - x_2^0)\right\|_1 + \epsilon\right)\\
&\leq \frac{1}{1-\kappa}\left(\left\|1_{\Lambda_1^c}\Phi_1^*\tilde{x}_1^\star\right\|_1 + \left\|1_{\Lambda_2^c}\Phi_2^*\tilde{x}_2^\star\right\|_1 + \delta + \epsilon\right). \qquad (17)
\end{aligned}$$

Recall that the $\tilde{x}_i^\star$ are minimal solutions to (6). Using this fact and relatively sparsity, we obtain

$$\begin{aligned}
& \left\|1_{\Lambda_1^c}\Phi_1^*\tilde{x}_1^\star\right\|_1 + \left\|1_{\Lambda_2^c}\Phi_2^*\tilde{x}_2^\star\right\|_1\\
= \;& \|\Phi_1^*\tilde{x}_1^\star\|_1 + \|\Phi_2^*\tilde{x}_2^\star\|_1 - \|1_{\Lambda_1}\Phi_1^*\tilde{x}_1^\star\|_1 - \|1_{\Lambda_2}\Phi_2^*\tilde{x}_2^\star\|_1\\
\leq \;& \left\|\Phi_1^* x_1^0\right\|_1 + \left\|\Phi_2^* x_2^0\right\|_1 + \|\Phi_2^* n\|_1 - \|1_{\Lambda_1}\Phi_1^*\tilde{x}_1^\star\|_1 - \|1_{\Lambda_2}\Phi_2^*\tilde{x}_2^\star\|_1\\
\leq \;& \left\|\Phi_1^* x_1^0\right\|_1 + \left\|\Phi_2^* x_2^0\right\|_1 + \epsilon + \left\|1_{\Lambda_1}\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 - \left\|1_{\Lambda_1}\Phi_1^* x_1^0\right\|_1 + \left\|1_{\Lambda_2}\Phi_2^*(\tilde{x}_2^\star - x_2^0)\right\|_1 - \left\|1_{\Lambda_2}\Phi_2^* x_2^0\right\|_1\\
\leq \;& \left\|1_{\Lambda_1}\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_2}\Phi_2^*(\tilde{x}_2^\star - x_2^0)\right\|_1 + \delta + \epsilon. \qquad (18)
\end{aligned}$$

We combine (17) and (18) with the definition of κ in the calculation

$$\begin{aligned}
\left\|\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1 &\leq \frac{1}{1-\kappa}\left(\left\|1_{\Lambda_1^c}\Phi_1^*\tilde{x}_1^\star\right\|_1 + \left\|1_{\Lambda_2^c}\Phi_2^*\tilde{x}_2^\star\right\|_1 + \delta + \epsilon\right)\\
&\leq \frac{1}{1-\kappa}\left(\left\|1_{\Lambda_1}\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_2}\Phi_2^*(\tilde{x}_2^\star - x_2^0)\right\|_1 + 2\delta + 2\epsilon\right)\\
&\leq \frac{1}{1-\kappa}\left(\left\|1_{\Lambda_1}\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|1_{\Lambda_2}\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + 2\delta + 3\epsilon\right)\\
&\leq \frac{1}{1-\kappa}\left(\kappa\left(\left\|\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1\right) + 2\delta + 3\epsilon\right)\\
&\leq \frac{1}{1-\kappa}\left(\kappa\left(\left\|\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|\Phi_2^*(\tilde{x}_2^\star - x_2^0)\right\|_1\right) + 2\delta + 4\epsilon\right).
\end{aligned}$$

Merging this result with (12) and (14) yields

$$\begin{aligned}
\left\|\tilde{x}_1^\star - x_1^0\right\|_2 + \left\|\tilde{x}_2^\star - x_2^0\right\|_2 &\leq \left\|\Phi_1^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \left\|\Phi_2^*(\tilde{x}_1^\star - x_1^0)\right\|_1 + \epsilon\\
&\leq \left(1 - \frac{\kappa}{1-\kappa}\right)^{-1}\left(\frac{2\delta + 4\epsilon}{1-\kappa}\right) + \epsilon\\
&= \frac{2\delta + (5-2\kappa)\epsilon}{1-2\kappa},
\end{aligned}$$

as desired.

Also note that these results can be easily generalized to general frames instead of Parseval frames, which then changes the separation estimates by invoking the lower frame bound.

5. RECOVERY OF MISSING DATA ESTIMATES

The bounds for the accuracy of the solution of (REC) presented in this section are remarkably similar to the results in the preceding section, showing the versatility of ℓ_1 minimization on the analysis side. The proofs for the following results may be found in [15].

We begin with the data recovery analog of Definition 4.1.

DEFINITION 5.1. *Let $\Phi = \{\varphi_i\}_{i\in I}$ be a Parseval frame for a Hilbert space $\mathcal{H}$, and let $\Lambda \subseteq I$. Then $x^0 \in \mathcal{H}$ is called* δ-relatively sparse in Φ with respect to Λ *if*

$$\left\|1_{\Lambda^c}\Phi^* x^0\right\|_1 \leq \delta.$$

Comparing this definition with Definition 4.1, we note that if $x = x_1 + x_2$ is δ-relatively sparse in Φ_1 and Φ_2 with respect to Λ_1 and Λ_2, then there exists some $0 \leq \eta \leq \delta$ such that x_1 is η-relatively sparse in Φ_1 with respect to Λ_1 and x_2 is is $(\delta - \eta)$-relatively sparse in Φ_2 with respect to Λ_2. Also, as with Definition 4.1, δ-relative sparsity is a clustered sparsity.

The data recovery versions of Theorems 4.1 and 4.2 follow. Note that the κ used here is the concentration on $\mathcal{H}_M$, not the joint concentration.

THEOREM 5.1 ([15]). *Let $\Phi = \{\varphi_i\}_{i\in I}$ be a Parseval frame for a Hilbert space $\mathcal{H} = \mathcal{H}_M \oplus \mathcal{H}_K$, and suppose that for $x^0 \in \mathcal{H}$, only $P_K x^0$ is known. Further, let $\Lambda \subseteq I$ be chosen such that x^0 is δ-relatively sparse in Φ with respect to Λ. Then the solution $x^\star$ of the ℓ_1 minimization problem* (REC) *stated in (7) satisfies*

$$\left\|x^\star - x^0\right\|_2 \leq \frac{2\delta}{1-2\kappa}. \tag{19}$$

THEOREM 5.2 ([15]). *Let $\Phi = \{\varphi_i\}_{i\in I}$ be a Parseval frame for a Hilbert space $\mathcal{H} = \mathcal{H}_M \oplus \mathcal{H}_K$, and suppose that for $x^0 \in \mathcal{H}$, only $P_K x^0$ is known. Further, let $\Lambda \subseteq I$ be chosen such that x^0 is δ-relatively sparse in Φ with respect to Λ. Then the solution $x^\star$ of the ℓ_1 minimization problem (*REC*) stated in (7) satisfies*

$$\left\|x^\star - x^0\right\|_2 \leq \frac{2\delta}{1-2\mu_c(\Lambda, P_M\Phi; \Phi)}.$$

The reader should notice that the considered error $\left\|x^\star - x^0\right\|_2$ is only measured on $\mathcal{H}_M$, the masked space, since $P_K x^\star = P_K x^0$ due to the constraint in (REC). We also point out the difference in dependencies of δ and κ in these results. The relative sparsity here is unrelated to how much data is missing from the measurement, while κ and μ_c are.

Adding noise to the hypotheses of Theorems 5.1 and 5.2, we obtain the following theorem.

THEOREM 5.3 ([15]). *Let $\Phi = \{\varphi_i\}_{i\in I}$ be a Parseval frame for a Hilbert space $\mathcal{H} = \mathcal{H}_M \oplus \mathcal{H}_K$, and suppose that for $x^0 \in \mathcal{H}$, only $\tilde{x} = P_K x^0 + n$ is known, where n satisfies $\|\Phi^* n\|_1 \leq \epsilon$ and $P_K n = n$. Further, let $\Lambda \subseteq I$ be chosen such that x^0 is δ-relatively sparse in Φ with respect to Λ. Then the solution $x^\star$ of the ℓ_1 minimization problem* (RECNOISE) *stated in (8) satisfies*

$$\left\|\tilde{x}^\star - x^0\right\|_2 \leq \frac{2\delta + (3+2\kappa)\epsilon}{1-2\kappa}$$

and more loosely satisfies

$$\left\|\tilde{x}^\star - x^0\right\|_2 \leq \frac{2\delta + (3+2\kappa)\epsilon}{1-2\mu_c(\Lambda, P_M\Phi; \Phi)}.$$

Acknowledgements

Emily J. King is supported by a fellowship for postdoctoral researchers from the Alexander von Humboldt Foundation. Gitta Kutyniok acknowledges support from DFG Grant SPP-1324, KU 1446/13 and DFG Grant KU 1446/14. Xiaosheng Zhuang is supported by DFG Grant KU 1446/14.

REFERENCES

1. J. Starck, D. Donoho, and E. Candes, "Astronomical image representation by the curvelet transform," *Astronomy and Astrophysics* **398**, pp. 785–800, 2003.
2. J.-L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component separation," *Advances in Imaging and Electron Physics* **132**, pp. 287 – 348, Elsevier, 2004.
3. J. Bobin, J.-L. Starck, J. M. Fadili, Y. Moudden, and D. L. Donoho, "Morphological component analysis: an adaptive thresholding strategy," *IEEE Trans. Image Process.* **16**(11), pp. 2675–2681, 2007.
4. F. G. Meyer, A. Z. Averbuch, and R. R. Coifman, "Multilayered image representation: application to image compression," *IEEE Trans. Image Process.* **11**(9), pp. 1072–1080, 2002.
5. J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Process.* **14**(10), pp. 1570–1582, 2005.
6. C. Fevotte, B. Torresani, L. Daudet, and S. Godsill, "Sparse linear regression with structured priors and application to denoising of musical audio," *Audio, Speech, and Language Processing, IEEE Transactions on* **16**, pp. 174 –185, Jan. 2008.
7. R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Process.* **51**(1), pp. 101–111, 2003.
8. G. Hennenfent, L. Fenelon, and F. J. Herrmann, "Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach," *Geophysics* **75**(6), pp. WB203–WB210, 2010.
9. F. J. Herrmann and G. Hennenfent, "Non-parametric seismic data recovery with curvelet frames," *Geophysical Journal International* **173**, pp. 233–248, 2008.
10. C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, "Filling-in by joint interpolation of vector fields and gray levels," *IEEE Trans. on Image Process.* **10**(8), pp. 1200–1211, 2001.
11. M. Bertalmio, A. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001 (CVPR 2001)*, pp. I–355–I362, IEEE, 2001.
12. M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of SIGGRAPH 2000, New Orleans*, pp. 417–424, July 2000.
13. T. F. Chan and J. Shen, "Mathematical models for local nontexture inpaintings," *SIAM J. Appl. Math.* **62**(3), pp. 1019–1043 (electronic), 2001/02.
14. M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Appl. Comput. Harmon. Anal.* **19**(3), pp. 340–358, 2005.
15. E. J. King, G. Kutyniok, and X. Zhuang, "Using clustered sparsity and microlocal analysis to analyze data recovery." in preparation, 2011.
16. M. Kowalski and B. Torrésani, "Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients," *Signal, Image and Video Processing* **3**, pp. 251–264, 2009. 10.1007/s11760-008-0076-1.
17. G. Teschke, "Multi-frame representations in linear inverse problems with mixed multi-constraints," *Appl. Comput. Harmon. Anal.* **22**(1), pp. 43–60, 2007.
18. M. Aharon, M. Elad, and A. Bruckstein, "K -SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Process.* **54**, pp. 4311–4322, Nov. 2006.
19. K. Engan, S. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. (ICASSP '99)*, **5**, pp. 2443–2446, 1999.

20. B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vision Research* **37**(23), pp. 3311–3325, 1997.
21. J.-L. Starck, Y. Moudden, J. Bobin, M. Elad, and D. Donoho, "Morphological component analysis," in *Wavelets XI (San Diego, CA, 2005)*, M. Papadakis, A. F. Laine, and M. A. Unser, eds., *SPIE Proceedings*(5914), SPIE, (Bellingham, WA), 2005.
22. D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization," *Proc. Natl. Acad. Sci. USA* **100**(5), pp. 2197–2202 (electronic), 2003.
23. M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory* **48**(9), pp. 2558–2567, 2002.
24. S. Nam, M. Davies, M. Elad, and R. Gribonval, "Cosparse analysis modeling - uniqueness and algorithms," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, IEEE, 2011.
25. J.-F. Cai, R. H. Cha, and Z. Shen, "Simultaneous cartoon and texture inpaintingn," *Inverse Probl. Imag.* **4**(3), pp. 379 – 395, 2010.
26. B. Dong, H. Ji, J. Li, Z. Shen, and Y. Xu, "Wavelet frame based blind image inpainting," *Appl. Comput. Harmon. Anal.* , 2011. In press.
27. D. L. Donoho and G. Kutyniok, "Microlocal analysis of the geometric separation problem." preprint, 2010.
28. D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.* **49**(3), pp. 906–931, 1989.
29. D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory* **47**(7), pp. 2845–2862, 2001.
30. J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory* **50**(10), pp. 2231–2242, 2004.
31. L. Borup, R. Gribonval, and M. Nielsen, "Beyond coherence: recovering structured time-frequency representations," *Appl. Comput. Harmon. Anal.* **24**(1), pp. 120–128, 2008.
32. F. Bunea, A. B. Tsybakov, and M. H. Wegkamp, "Aggregation for Gaussian regression," *Ann. Statist.* **35**(4), pp. 1674–1697, 2007.
33. M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68**(1), pp. 49–67, 2006.
34. D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory* **52**(4), pp. 1289–1306, 2006.