Analysis of ℓ_1 Minimization in the Geometric Separation Problem

David L. Donoho
Department of Statistics
Stanford University
Stanford, CA 94305
Email: donoho@stanford.edu

Gitta Kutyniok
Department of Statistics
Stanford University
Stanford, CA 94305
Email: kutyniok@stanford.edu

***Abstract*— Modern data are often composed of two (or more) morphologically distinct constituents – for instance, pointlike and curvelike structures in astronomical imaging of galaxies. Although it seems impossible to extract those components – as there are two unknowns for every datum – suggestive empirical results have already been obtained especially by Jean-Luc Starck and collaborators. In this paper we develop a theoretical viewpoint, defining a Geometric Separation Problem and analyzing a model procedure. This procedure is inspired by work relating ℓ_1 minimization and sparsity. The procedure uses two deliberately overcomplete systems which sparsify the different components and decomposes by ℓ_1 minimization of the analysis (rather than synthesis) frame coefficients. We formalize two concepts – cluster coherence in place of the now-traditional singleton coherence and ℓ_1 minimization in frame settings, including those where singleton coherence within one frame may be high – and develop all the needed machinery to make these into fruitful tools. Our general approach applies to the problem of geometric separation of pointlike and curvelike structures in images by employing frames of radial wavelets and curvelets or orthonormal wavelets and shearlets. Our theoretical results show that at all sufficiently fine scales, nearly-perfect separation is achieved. We use microlocal analysis to understand heuristically why separation might be possible and to organize a rigorous analysis.**

Index Terms—ℓ_1 **minimization. Sparse Representation. Mutual Coherence. Tight Frames. Curvelets, Shearlets, Radial Wavelets.**

I. Introduction

Consider a distribution f with domain $\mathbf{R}^2$ which is smooth away from singularities (discontinuities, say). We are interested in developing a basis or frame which can sparsely represent f. Such sparsity is useful for data compression and noise removal, as well as compressed sensing. The type of basis which best sparsifies depends on the geometry of the singularities in f. If the singularities occur at a finite number of isolated points, then *wavelets* give an optimally sparse representation – one with the fewest significantly nonzero coefficients; if the singularities occur at a finite number of smooth curves, then a parabolic-scaling based multiscale representation (*curvelets* or *shearlets*; here we shall mostly use curvelets) will be optimally sparse.

In fact, real-world signals are generally speaking a mixture of content types and, correspondingly, a model where singularities are of only one geometric type is overly narrow. If f is actually a nontrivial superposition $\mathcal{P} + \mathcal{C}$ where $\mathcal{P}$ has only point singularities and $\mathcal{C}$ has only curvilinear singularities, then two things happen:

- Neither wavelets alone nor curvelets alone is very good – the sparsity either achieves drops dramatically.
- In fact, no basis or traditional linear representation is very good compared to the sparsest achievable representation.

In this paper we consider the problem of developing sparse representations by combining both wavelets and curvelets and using a nonlinear representation based on ℓ_1 minimization. The problem we solve is an abstraction of a problem with considerable practical interest; this problem has driven extensive empirical work over the last two decades, which we now relate. We find it especially useful and enlightening to first study this problem in a much more abstract setting by considering a signal which is composed out of two constituents each one being relatively sparse in a tight frame. Our approach to exploiting ℓ_1 minimization for separation in this setting by using the geometry of the problem as the driving force reaches conclusions reminiscent of those which have been obtained previously using randomness of the sparsity pattern to be recovered; however here, the pattern is not random but highly organized. The idea that ℓ_1 solves underdetermined problems when the important coefficients are *clustered geometrically* in particular ways should have many applications in other settings.

Finally, we note that this paper is really a summary of results. Here we omit proofs and derivations.

A. The Multiple-Basis Representation Problem

Sometime in the early 1990's, Raphy Coifman became interested in the problem of representing signals using more than one basis. Coifman, Wickerhauser and co-workers at the time made a sort of heuristic exploration motivated intuitively, see [5, Figure 26, Panels a-h]. They tried to represent a signal using superpositions of simple subsignals, i.e., signals that have relatively few significant coefficients in a particular basis. François Meyer and Raphy Coifman raised public awareness of the possibility of multiple-basis representation with a poster they produced for 'Mathematics Awareness Month' (1998) showing the decomposition of the famous 'Mandrill' picture into several bases at once.

In retrospect, using modern formulations, these early explorations suggest that there is an interesting class of real signals with these characteristics:

- An observed signal $S \in \mathbf{R}^n$ is a superposition of subsignals S_i, $i = 1, 2$.
- Each subsignal S_i is 'simple' in an 'appropriate' basis Φ_i, $i = 1, 2$.
- Each subsignal 'looks like noise' in an 'inappropriate' basis. Here Φ_2 is inappropriate for S_1, and Φ_1 is inappropriate for S_2.

This set of observations raises an ambitious question:

can one extract the two simple subsignals and express each one simply in its own appropriate basis?

Formally: can one recover 'simple' signals S_1 and S_2 given knowledge of $S = S_1 + S_2$ only? But people at the time said in conversation, that, when put this starkly, the answer was simply 'no', since there are twice as many unknowns as knowns. Nevertheless, some of the heuristic results at the time were very suggestive.

B. Minimum ℓ_1 Decomposition and Perfect Separation

A few years later, one of us worked with Scott Shaobing Chen to develop a formal, optimization-based approach to the multiple-basis representation problem. Given bases Φ_i, $i = 1, 2$, one solves the following problem

$$(\mathrm{BP}) \qquad \min \|\alpha_1\|_1 + \|\alpha_2\|_1 \text{ subject to } S = \Phi_1\alpha_1 + \Phi_2\alpha_2.$$

Here $\|\cdot\|_1$ denotes the usual ℓ_1 norm. Note that here there are $2n$ unknowns in α_1 and α_2 and only n knowns in S, but that an optimization principle is being used to select a particular element from the n-dimensional space of all possible solutions.

Inspired by earlier work of the first author (see [11], [12]), it was thought at the time that the ℓ_1 norm would find sparse solutions if they exist. And indeed, Chen's thesis showed that there was perfect separation of sinusoids from spikes, and the true underlying simplicity of the signal was revealed – even though there were more unknowns than equations.

In the years since that work, two streams of research emerged.

- *Theoretical work*, which showed that, indeed, one could often obtain the sparsest possible representations to an underdetermined problem by ℓ_1 optimization; thereby perfectly separating two simple subsignals that had been superposed to produce the complex observed signal. See [3], [6], [7], [23] for a selection of general work concerning ℓ_1 minimization, and [8], [9], [13], [15] for focused work on signal separation.
- *Empirical work*, which showed that combined representations such as wavelets with curvelets or wavelets with sinusoids often gave very compelling separations of real signals and images, see, for instance, [4], [14], [20], [22].

Jean-Luc Starck and Michael Elad found in their empirical work that ℓ_1 minimization per se needs an important modification for component separation, which is also important in this paper (see (CSEP) and (SEP)).

C. Geometric Separation Problem

We now pose a very stark geometric separation problem. Suppose there is a pointlike object $\mathcal{P}$ made of point singularities:

$$\mathcal{P} = \sum_{i=1}^{P} |x - x_i|^{-3/2}. \tag{1}$$

At the same time, there is a curvelike object $\mathcal{C}$, a singularity along a curve τ:

$$\mathcal{C} = \int \delta_{\tau(t)} dt, \tag{2}$$

where δ_x is the usual Dirac Delta at x. These two objects are geometrically quite different, but their energy distribution across scales is similar; if $\mathcal{A}_r$ denotes the annular region $r < |\xi| < 4r$,

$$\int_{\mathcal{A}_r} |\hat{\mathcal{P}}|^2(\xi) \asymp r, \qquad \int_{\mathcal{A}_r} |\hat{\mathcal{C}}|^2(\xi) \asymp r, \qquad r \to \infty.$$

Thus we cannot simply bandpass-filter the signal $\mathcal{P} + \mathcal{C}$ to extract $\mathcal{P}$ and $\mathcal{C}$ from it.

Now assume that we observe the 'signal'

$$f = \mathcal{P} + \mathcal{C}, \tag{3}$$

however, the distributions $\mathcal{P}$ and $\mathcal{C}$ are unknown to us.

Definition 1.1: The *Geometric Separation Problem* requires to recover $\mathcal{P}$ and $\mathcal{C}$ from knowledge only of f; here $\mathcal{P}$ and $\mathcal{C}$ are unknown to us, but obey (1), (2) and certain regularity conditions on the curve τ.

As there are two unknowns ($\mathcal{P}$ and $\mathcal{C}$) and only one observation (f), the problem seems improperly posed. Yet we will be able to go quite far in the direction of solving it.

D. Two Geometric Frames

Two overcomplete systems will be of interest to us at this point:

- *Radial Wavelets* – a tight frame with perfectly isotropic generating elements.
- *Curvelets* – a highly directional tight frame with increasingly anisotropic elements at fine scales.

We construct these as follows. Let $W(r)$ be an 'appropriate' window function. Then we define the radial wavelets at scale index j and spatial position index $k = (k_1, k_2)'$ by the Fourier transforms

$$\hat{\psi}_\lambda(\xi) = 2^{-j} \cdot W(|\xi|/2^j) \cdot e^{ik'\xi/2^j},$$

where we let $\lambda = (j, k)$ index position and scale. For the *same* window function W and a 'bump function' V, we define the curvelet at scale j, orientation ℓ, and spatial position $k = (k_1, k_2)$ by the Fourier transforms

$$\hat{\gamma}_\eta(\xi) = 2^{-j\frac{3}{4}} \cdot W(|\xi|/2^j) V((\omega - \theta_{j,\ell}) 2^{j/2}) \cdot e^{i(R_{\theta_{j,\ell}} D_{2^{-j}} k)'\xi},$$

where here $\theta_{j,\ell} = 2\pi\ell/2^{j/2}$, R_θ is planar rotation by $-\theta$ radians, D_a is anisotropic scaling with diagonal $(a, \sqrt{a})$, and we let $\eta = (j, \ell, k)$ index scale, orientation, and scale. See [1] for more details.

Roughly speaking, the radial wavelets are 'radial bumps' with position $k/2^j$ and scale 2^{-j} and provide an optimally sparse representation for point singularities, while the curvelets live on anisotropic regions of width $a = 2^{-j}$ and length $\sqrt{a}$ and provide an optimally sparse representation for curvilinear singularities.

Using the *same* window W, we can construct a family of filters F_j with transfer functions

$$\hat{F}_j(\xi) = W(|\xi|/2^j), \qquad \xi \in \mathbf{R}^2.$$

These filters allow us to decompose a function f into pieces f_j with different scales, the piece f_j at subband j arises from filtering f using F_j:

$$f_j = F_j \star f;$$

the Fourier transform $\hat{f}_j$ is supported in $\mathcal{A}_{2^j}$. Because of our assumption on W, we can reconstruct the original function from these pieces using the formula

$$f = \sum_j F_j \star f_j, \qquad f \in L^2(\mathbf{R}^2).$$

The tight frames of curvelets and radial wavelets discussed above interact in a very local way with the filtering F_j. Letting $\mathcal{F}_j$ denote the range of the operator of convolution with F_j, it is easy to see that radial wavelets (curvelets) at level j' are orthogonal to $\mathcal{F}_j$ unless $|j' - j| \leq 1$.

For future use, let Λ_j denote the collection of indices $\lambda = (j,k)$ of wavelets at level j, and $\Lambda_j^{\pm} = \Lambda_{j-1} \cup \Lambda_j \cup \Lambda_{j+1}$. Similarly, let Δ_j denote the indices $\eta = (j,k,\ell)$ of curvelets at level j, and let $\Delta_j^{\pm} = \Delta_{j-1} \cup \Delta_j \cup \Delta_{j+1}$. We conclude that elements of $\mathcal{F}_j$ can be represented using either curvelets $(\gamma_\eta : \eta \in \Delta_j^{\pm})$ or radial wavelets $(\psi_\lambda : \lambda \in \Lambda_j^{\pm})$.

We would like to briefly discuss a different pair of representation systems which also separate pointlike and curvelike structures. In contrast to the pair considered before, surprisingly, one system even forms an orthonormal basis:

- *Orthonormal Separable Meyer Wavelets* – an orthonormal basis of perfectly isotropic generating elements.
- *Shearlets* – a highly directional tight frame with increasingly anisotropic elements at fine scales.

The frequency supports of the frame elements for each scale match perfectly, thereby allowing us to proceed by decomposing a signal band by band in a similar fashion as above. We would like to remark that shearlets differ from curvelets by the fact that they exploit shear matrices rather than rotation matrices, hence parameterize directions by slope rather than angle, thereby providing a complete methodology for the continuous and discrete setting [16], [18] as well as for algorithmic realizations [19].

For the remainder of the paper we will now focus on the pair radial wavelets/curvelets. However, all results hold also true in a similar way for the pair orthonormal wavelets/shearlets.

E. Sparse Multiple Frame Expansions

We have two complete representations for $\mathcal{F}_j$, yielding two ways of representing the subband component f_j: in terms of its wavelet expansion or in terms of its curvelet expansion. Each frame uses a single geometric tendency – either highly nondirectional or highly directional – to represent f_j. However, f_j may have both isotropic and directional features. We therefore seek a combined representation

$$f_j = \sum_{\lambda \in \Lambda_j^{\pm}} w_\lambda \psi_\lambda + \sum_{\eta \in \Delta_j^{\pm}} c_\eta \gamma_\eta.$$

Because the combined frame formed by concatenating the two frames is overcomplete, there are many possible ways this decomposition can be done. Some of them may be geometrically motivated, many are not.

We propose to obtain a decomposition using a sparsity-promoting approach based on ℓ_1 minimization. Consider the following dual-frame **Component Separation** problem:

$$\begin{aligned}(\text{CSEP}) \quad (W_j, C_j) \;=\; & \operatorname{argmin} \|w\|_1 + \|c\|_1 \\ & \text{subject to } f_j = W_j + C_j \\ & \text{and } w_\lambda = \langle W_j, \psi_\lambda \rangle,\ \lambda \in \Lambda_j^{\pm} \\ & \text{and } c_\eta = \langle C_j, \gamma_\eta \rangle,\ \eta \in \Delta_j^{\pm}.\end{aligned}$$

At first glance (CSEP) seems identical to (BP) in the dual-basis case, however there is an important distinction to be made. Here the ℓ_1 norm is being applied on the *analysis* coefficients of the two different 'components' rather than on the individual *synthesis* coefficients. The hope in (BP) is to get exactly the right nonzero coefficients. The hope in (CSEP) is to separate components rather than identify the true nonzero coefficients.

Here is the reason for the name 'component separation'. Armed with the separation at each scale subband, we define the *pointlike* component as the superposition of all the wavelet terms

$$\tilde{P} = \sum_j F_j \star W_j$$

and the *curvelike* component as the superposition of all the curvelet terms

$$\tilde{C} = \sum_j F_j \star C_j$$

yielding the decomposition

$$f = \tilde{P} + \tilde{C}.$$

The justification for calling $\tilde{P}$ pointlike and $\tilde{C}$ curvelike is provided by our results below.

Letting f be defined as in (3), i.e., truly, a superposition of a pointlike distribution and a curvelike distribution, we will show that, at fine scales, the wavelet component picks up the pointlike terms, and the curvelet component picks up the curvelike terms. We now define the scale subbands of the geometric components by:

$$\mathcal{P}_j = F_j \star \mathcal{P} \quad \text{and} \quad \mathcal{C}_j = F_j \star \mathcal{C}.$$

Of course

$$f_j = \mathcal{P}_j + \mathcal{C}_j,$$

but we assume f_j is known and $\mathcal{P}_j$ and $\mathcal{C}_j$ are unknown to us. The following result shows that in a certain sense they are recoverable by solving (CSep).

Theorem 1.1 ([10]): ASYMPTOTIC SEPARATION.

$$\frac{\|W_j - \mathcal{P}_j\|_2 + \|C_j - \mathcal{C}_j\|_2}{\|\mathcal{P}_j\|_2 + \|\mathcal{C}_j\|_2} \to 0, \qquad j \to \infty.$$

This result shows that components are recovered asymptotically: at fine scales, the energy in the curvelike component is all captured by the curvelet coefficients and the energy in the pointlike component is all captured by the wavelet coefficients.

II. COMPONENT SEPARATION BY ℓ_1 MINIMIZATION

We now study the behavior of ℓ^1 minimization in the general two-frame case. Our analysis centers on the use of cluster coherence to control joint concentration.

A. ℓ_1 Minimization for Separation of Two Tight Frames

Suppose we have two tight frames Φ_1, Φ_2 in a Hilbert space $\mathcal{H}$, and a signal vector $S \in \mathcal{H}$. We know *a priori* that there exists a decomposition

$$S = S_1^0 + S_2^0,$$

where S_1^0 is sparse in Φ_1 and S_2^0 is sparsely represented by Φ_2.

Consider the following optimization problem

$$(\text{SEP}) \quad (S_1^\star, S_2^\star) = \text{argmin}_{S_1,S_2} \|\Phi_1^T S_1\|_1 + \|\Phi_2^T S_2\|_1 \quad \text{subject to } S = S_1 + S_2.$$

In this problem, the norm is placed on the **analysis** coefficients rather than on the **synthesis** coefficients as in (BP). To analyze this we need the following notion.

Definition 2.1: Let Φ_1 and Φ_2 be two tight frames. Given two sets of coefficients $\mathcal{S}_1$ and $\mathcal{S}_2$, define the *joint concentration* $\kappa = \kappa(\mathcal{S}_1, \mathcal{S}_2)$ by

$$\kappa(\mathcal{S}_1, \mathcal{S}_2) = \sup_f \frac{\|1_{\mathcal{S}_1} \Phi_1^T f\|_1 + \|1_{\mathcal{S}_2} \Phi_2^T f\|_1}{\|\Phi_1^T f\|_1 + \|\Phi_2^T f\|_1}.$$

In words, we consider the maximal fraction of total ℓ_1 norm which can be concentrated to the combined index set $\mathcal{S}_1 \cup \mathcal{S}_2$. Concepts of this kind go back to [12]. Adequate control of joint concentration ensures that the principle (SEP) gives a successful approximate separation.

Proposition 2.1 ([10]): Suppose that S can be decomposed as $S = S_1^0 + S_2^0$ so that each component S_i^0 is relatively sparse in Φ_i, $i = 1, 2$, i.e.,

$$\|1_{\mathcal{S}_1^c} \Phi_1^T S_1^0\|_1 + \|1_{\mathcal{S}_2^c} \Phi_2^T S_2^0\|_1 \le \delta.$$

Let $(S_1^\star, S_2^\star)$ solve (SEP). Then

$$\|S_1^\star - S_1^0\|_2 + \|S_2^\star - S_2^0\|_2 \le \frac{2\delta}{1 - 2\kappa}.$$

B. Cluster Coherence

Our novel contribution to the analysis of ℓ_1 minimization starts from this concept:

Definition 2.2: Given tight frames $\Phi = (\phi_i)_i$ and $\Psi = (\psi_j)_j$ and an index subset $\mathcal{S}$ associated with expansions in frame Φ, we define the *cluster coherence*

$$\mu_c(\mathcal{S}; \Phi, \Psi) = \max_j \sum_{i \in \mathcal{S}} |\langle \phi_i, \psi_j \rangle|.$$

In many studies of ℓ_1 optimization, one studies instead the mutual coherence

$$\mu(\Phi, \Psi) = \max_j \max_i |\langle \phi_i, \psi_j \rangle|,$$

whose importance was shown by [9]. This might be thought of as singleton coherence. In contrast, cluster coherence bounds coherence between a single member of frame Ψ and a cluster of members of frame Φ, clustered at $\mathcal{S}$.

Proposition 2.2 ([10]): We have

$$\kappa(\mathcal{S}_1, \mathcal{S}_2) \le \max(\mu_c(\mathcal{S}_1; \Phi_1, \Phi_2), \mu_c(\mathcal{S}_2; \Phi_2, \Phi_1)).$$

Of course, the concept of cluster coherence alone is barren, but with the right supporting infrastructure it becomes fruitful. Building this infrastructure in the geometric separation case is the core of our work.

III. GEOMETRIC SEPARATION OF POINTLIKE AND CURVELIKE STRUCTURES

A. Main Result

The concepts of the previous section will now be applied to the problem (CSEP), at each scale j separately. Thus $f = \mathcal{P} + \mathcal{C}$ (see (3)) is our distribution of interest, and F_j our bandpass filter. The object referred to in this section is $S = f_j$ and the tight frames are Φ_1, the full radial wavelet frame, and Φ_2, the full curvelet tight frame. We apply the optimization problem (SEP), getting subsignal components $S_1^\star, S_2^\star$, which we then relabel as the wavelet component W_j and curvelet component C_j.

We now assume that we have appropriately defined the cluster of significant wavelet coefficients $\mathcal{S}_{1,j}$ and the cluster of significant curvelet coefficients $\mathcal{S}_{2,j}$. One first idea might be to carefully choose an increasing series of thresholds $\epsilon_{j,1}$ and $\epsilon_{j,2}$ and define

$$\mathcal{S}_{1,j} = \{\lambda \in \Lambda_j^\pm : |\langle \psi_\lambda, \mathcal{P}_j \rangle| > \epsilon_{j,1} \|(\langle \psi_\lambda, \mathcal{P}_j \rangle)_\lambda\|_{\ell_\infty(\Lambda_j^\pm)}\}$$

and

$$\mathcal{S}_{2,j} = \{\eta \in \Delta_j^\pm : |\langle \gamma_\eta, \mathcal{C}_j \rangle| > \epsilon_{j,2} \|(\langle \gamma_\eta, \mathcal{C}_j \rangle)_\eta\|_{\ell_\infty(\Delta_j^\pm)}\}.$$

We will provide the reader with more insight in the process of choosing $\mathcal{S}_{1,j}$ and $\mathcal{S}_{2,j}$ in the following subsection. For the precise, technically quite involved definition we would like to refer to [10].

We then let δ_j denote the degree of approximation by significant coefficients; the sum of the wavelet approximation error to the point singularity:

$$\delta_{j,1} = \sum_{\lambda \in \mathcal{S}_{1,j}^c} |\langle \psi_\lambda, \mathcal{P}_j \rangle|;$$

and the curvelet approximation error to the curvilinear singularity:

$$\delta_{j,2} = \sum_{\eta \in \mathcal{S}_{2,j}^c} |\langle \gamma_\eta, \mathcal{C}_j \rangle|.$$

Finally, $\kappa(\mathcal{S}_{1,j}, \mathcal{S}_{2,j})$, the degree of joint wavelet-curvelet concentration at the significant subsets, will be controlled by cluster coherence $\mu_c(\mathcal{S}_{1,j}; \Phi_1, \Phi_2)$, which is the maximal coherence of a curvelet to a cluster of significant wavelet coefficients; and by $\mu_c(\mathcal{S}_{2,j}; \Phi_2, \Phi_1)$, which is the maximal coherence of a wavelet to a cluster of significant curvelet coefficients. We have

Corollary 3.1 ([10]): Suppose that the sequence of transform-space clusters $(\mathcal{S}_{1,j})$, and $(\mathcal{S}_{2,j})$ has **both** of the following two properties: (i) asymptotically negligible cluster coherences:

$$\mu_c(\mathcal{S}_{1,j}; \Phi_1, \Phi_2), \mu_c(\mathcal{S}_{2,j}; \Phi_2, \Phi_1) \to 0, \qquad j \to \infty, \tag{4}$$

and (ii) asymptotically negligible cluster approximation errors:

$$\delta_j = \delta_{1,j} + \delta_{2,j} = o(||f_j||_2), \qquad j \to \infty. \tag{5}$$

Then we have asymptotically near-perfect separation:

$$\frac{\|W_j - \mathcal{P}_j\|_2 + \|C_j - \mathcal{C}_j\|_2}{\|f_j\|_2} \to 0, \qquad j \to \infty.$$

Hence our main result, Theorem 1.1, depends on obtaining sufficiently good estimates for cluster coherence for clusters defined as sufficiently good approximations to the objects of interest. It becomes evident that there is no real need to define the clusters by thresholding; what is instead needed is that, however defined, the clusters obey conditions (i) and (ii) of the Corollary.

B. Heuristics via a Microlocal Analysis Viewpoint

The singular support of a distribution f, sing supp(f) is the set of points where f is not locally C^∞. In the geometric separation case, we have

$$\text{sing supp}(f) = \text{sing supp}(\mathcal{P} + \mathcal{C}) = \{x_i\} \cup image(\tau).$$

Note that the points x_i can intersect the image of the curve τ – we make no separation hypothesis asking the point singularities to 'stay away' from the curvilinear singularities. To properly separate between pointlike and curvelike singularities we need to consider a *phase space* for microlocal analysis indexed by position-direction pairs (b, θ).

Living in this phase space is the wavefront set $WF(f)$; roughly, this is the set of position-orientation pairs at which f is nonsmooth; for more details, see [1], [17], [18]. Under the geometric separation model of Section 1, we have

$$WF(\mathcal{P}) = \text{supp}(\mathcal{P}) \times [0, 2\pi),$$

as a point singularity is singular in all directions at its support, and nowhere else; while

$$WF(\mathcal{C}) = \{(\tau(t), \theta(t)) : t \in [0, L(\tau)]\},$$

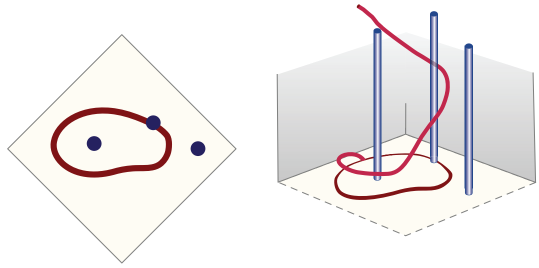

Fig. 1. The singular supports of some sample singularities $\mathcal{P}$ (dots) and $\mathcal{C}$ (curve) are pictured on the left hand side. On the right hand side it can be seen that their wavefront sets are in some sense better separated than the corresponding singular supports.

where $\tau(t)$ is a unit-speed parametrization of $\mathcal{C}$ and $\theta(t)$ is the normal direction to $\mathcal{C}$ at $\tau(t)$.

Figure 1 illustrates that even if x_1 meets $image(\tau)$, so the singular support of the point singularity and a curvilinear singularity overlap at x_1, they behave quite differently as wavefront sets in the full 3D parameter space, which gives us hope for separation.

Now turning our attention to the considered two frames, we observe that 'morally' radial wavelets are supported in spatial balls, whereas curvelets are 'morally' supported inside anisotropic spatial ellipses. Instead of studying the wavefront set, the continuous curvelet transform (CCT) (cf. [1]) now allows us to examine the location-orientation of those frame elements, i.e., their behavior in phase space. For a fixed sufficiently small scale, we see that the CCT of each radial wavelet is 'morally' supported in a vertical tube, whereas the CCT of a curvelet has a vaguely ellipsoidal structure exhibiting an anisotropic footprint (see Figure 2).

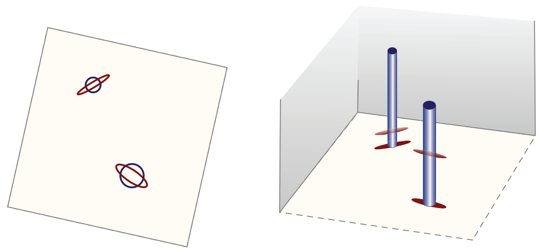

Fig. 2. The right hand side illustrates the 3D phase space portrait - obtained by CCT - of the radial wavelets and curvelets whose supports are pictured on the left hand side at different scales. The portrait conveys the intuition that curvelets and wavelets have less overlap at finer scales.

This visual analogy suggests to us what the likely configuration of cluster sets $\mathcal{S}_1$ and $\mathcal{S}_2$ will be and what computations will provide the needed cluster coherence estimates (4) and approximation error estimates (5). The following assertions seem quite plausible on visual grounds:

- Wavelets in $\mathcal{S}_1$ are associated to vertical tubes clustering around the point singularities in $\mathcal{P}$.
- Curvelets in $\mathcal{S}_2$ are associated with tubes clustering around the curvilinear phase portrait of $\mathcal{C}$.
- Curvelets are 'adapted' to curvilinear singularities (similar phase-space shape).

- Wavelets are ‘adapted’ to point singularities (similar phase-space shape).
- No single wavelet is coherent with anything built from the ensemble of curvelets in $\mathcal{S}_2$ (low overlap in phase space).
- No single curvelet is coherent with anything built from the ensemble of wavelets in $\mathcal{S}_1$ (low overlap in phase space).

The paper [10] designs clusters of wavelet coefficients $\mathcal{S}_{1,j}$ and curvelet coefficients $\mathcal{S}_{2,j}$ inspired by these intuitions. The following result shows that the hypothesis of Corollary 3.1 holds true, which proves Theorem 1.1. For the tedious and involved proof we refer to [10].

Lemma 3.1 ([10]): The sequence of transform-space clusters $(\mathcal{S}_{1,j})$, and $(\mathcal{S}_{2,j})$ provides (i) asymptotically negligible cluster coherences:

$$\mu_c(\mathcal{S}_{1,j}; \Phi_1, \Phi_2), \mu_c(\mathcal{S}_{2,j}; \Phi_2, \Phi_1) \to 0, \qquad j \to \infty,$$

and (ii) asymptotically negligible cluster approximation errors:

$$\delta_j = \delta_{1,j} + \delta_{2,j} = o(||f_j||_2) = o(2^{j/2}), \qquad j \to \infty.$$

IV. Conclusion

We considered signals decomposed of two subsignals each being relatively sparse in an appropriate basis or tight frame, and studied ℓ_1 minimization of the analysis (rather than synthesis) frame coefficients as the decomposition technique of such signals. We introduced the new notion of cluster coherence in contrast to the commonly considered singleton coherence. We then derived an upper estimate for the decomposition error based on cluster coherence and the cluster approximation error of the significant coefficients of the decomposition of the two subsignals, thereby extensively utilizing the geometry of the initial problem. Surprisingly, we reach conclusions reminiscent of those which have been obtained previously using randomness of the sparsity pattern to be recovered; however here, the pattern is not random but highly organized. Our results in fact open a new direction distinct from the recent avalanche of results on sparsity and ℓ_1 minimization by using the geometry of a problem as the driving force. We then applied our theoretical findings to the situation of images comprising point and curve singularities. Choosing the sparsifying frames of radial wavelets and curvelets or orthonormal wavelets and shearlets and employing methods from microlocal analysis, we prove that at all sufficiently fine scales, nearly-perfect separation is achieved.

Acknowledgment

The authors would like to thank Emmanuel Candès, Michael Elad, and Jean-Luc Starck, for numerous discussions on related topics. The second author would like to thank the Department of Statistics at Stanford University for its hospitality and support during her long-term visit. The authors would also like to thank the Newton Institute of Mathematics in Cambridge, UK for providing an inspiring research environment which led to the completion of a significant part of this work during their stay. This work was partially supported by NSF DMS 05-05303 and DMS 01-40698 (FRG), and by Deutsche Forschungsgemeinschaft (DFG) Heisenberg Fellowship KU 1446/8-1.

References

[1] E. J. Candès and D. L. Donoho, *Continuous curvelet transform: I. Resolution of the wavefront set*, Appl. Comput. Harmon. Anal. **19(2)** (2005), 162–197.

[2] E. J. Candès and D. L. Donoho, *Continuous curvelet transform: II. Discretization of frames*, Appl. Comput. Harmon. Anal. **19(2)** (2005), 198–222.

[3] E. J. Candès, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59(8)** (2006), 1207–1223.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Rev. **43** (2001), 129–159.

[5] R. R. Coifman and M. V. Wickerhauser, *Wavelets and adapted waveform analysis. A toolkit for signal processing and numerical analysis*, Different perspectives on wavelets (San Antonio, TX, 1993), 119–153, Proc. Sympos. Appl. Math., **47**, Amer. Math. Soc., Providence, RI, 1993.

[6] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory **52(4)** (2006), 1289–1306.

[7] D. L. Donoho, *For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution*, Comm. Pure Appl. Math. **59(6)** (2006), 797–829.

[8] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization*, Proc. Natl. Acad. Sci. USA **100(5)** (2003), 2197–2202.

[9] D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition,* IEEE Trans. Inform. Theory **47(7)** (2001), 2845–2862.

[10] D. L. Donoho and G. Kutyniok, *Microlocal Analysis of the Geometric Separation Problem*, preprint (2008).

[11] D. L. Donoho and B. F. Logan, *Signal recovery and the large sieve,* SIAM J. Appl. Math. **52(2)** (1992), 577–591.

[12] D. L. Donoho and P. B. Stark, *Uncertainty principles and signal recovery,* SIAM J. Appl. Math. **49(3)** (1989), 906–931.

[13] M. Elad and A. M. Bruckstein, *A Generalized Uncertainty Principle and Sparse Representation in Pairs of Bases,* IEEE Trans. Inform. Theory **48(9)** (2002), 2558–2567.

[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)** (2005), 340–358.

[15] R. Gribonval and M. Nielsen, *Sparse representations in unions of bases*, IEEE Trans. Inform. Theory **49(12)** (2003), 3320–3325.

[16] K. Guo, G. Kutyniok, and D. Labate, *Sparse Multidimensional Representations using Anisotropic Dilation und Shear Operators,* in Wavelets und Splines (Athens, GA, 2005), G. Chen und M. J. Lai, eds., Nashboro Press, Nashville, TN (2006), 189–201.

[17] L. Hörmander, *The analysis of linear partial differential operators. I. Distribution theory and Fourier analysis.* Springer-Verlag, Berlin, 2003.

[18] G. Kutyniok and D. Labate, *Resolution of the Wavefront Set using Continuous Shearlets,* Trans. Amer. Math. Soc., to appear.

[19] G. Kutyniok and T. Sauer, *Adaptive Directional Subdivision Schemes and Shearlet Multiresolution Analysis*, preprint (2007).

[20] F. G. Meyer, A. Averbuch, and R. R. Coifman, *Multi-layered Image Representation: Application to Image Compression*, IEEE Trans. on Image Processing **11(9)** (2002), 1072–1080.

[21] F. G. Meyer and R. R. Coifman, *Brushlets: a tool for directional image analysis and image compression*, Appl. Comput. Harmon. Anal. **4(2)** (1997), 147–187.

[22] J.-L. Starck, M. Elad, and D. L. Donoho, *Image decomposition via the combination of sparse representations and a variational approach,* IEEE Trans. Image Proc. **14(10)** (2005), 1570–1582.

[23] J. A. Tropp, *Greed is good: algorithmic results for sparse approximation*, IEEE Trans. Inform. Theory **50(10)** (2004), 2231–2242.