Geometric Separation using a Wavelet-Shearlet Dictionary

David L. Donoho [(1)] and Gitta Kutyniok [(2)]

(1) Department of Statistics, Stanford University, Stanford, CA 94305, USA.
(2) Institute of Mathematics, University of Osnabrück, 49069 Osnabrück, Germany.
donoho@stanford.edu, kutyniok@uni-osnabrueck.de

Abstract:

Astronomical images of galaxies can be modeled as a superposition of pointlike and curvelike structures. Astronomers typically face the problem of extracting those components as accurate as possible. Although this problem seems unsolvable – as there are two unknowns for every datum – suggestive empirical results have been achieved by employing a dictionary consisting of wavelets and curvelets combined with ℓ_1 minimization techniques. In this paper we present a theoretical analysis in a model problem showing that accurate geometric separation can be achieved by ℓ_1 minimization. We introduce the notions of *cluster coherence* and clustered sparse objects as a machinery to show that the underdetermined system of equations can be stably solved by ℓ_1 minimization. We prove that not only a radial wavelet-curvelet dictionary achieves nearly-perfect separation at all sufficiently fine scales, but, in particular, also an orthonormal wavelet-shearlet dictionary, thereby proposing this dictionary as an interesting alternative for geometric separation of pointlike and curvelike structures. To derive this final result we show that curvelets and shearlets are sparsity equivalent in the sense of a finite p-norm ($0 < p \leq 1$) of the cross-Grammian matrix.

1. Introduction

Cosmological data analysts face tasks of *geometric separation*. Gravitation, acting over time, drives an initially quasi-uniform distribution of matter in 3D to concentrate near lower-dimensional structures: points, filaments, and sheets. It would be desirable to process single 'maps' of matter density and somehow extract three 'pure' maps containing just the points, just the filaments, and just the sheets around which matter is concentrating. However, this problem contains three unknowns for every datum which seems impossible to solve on mathematical grounds.

Surprisingly, astronomer Jean-Luc Starck and collaborators have recently been empirically successful in numerical experiments with component separation. They used two or more overcomplete frames, each one specially adapted to particular geometric structures, and were able to obtain separation despite the fact that the underlying system of equations is highly underdetermined.

Here we analyze such approaches in a mathematical framework where we can show that success stems from an interplay between geometric properties of the objects to be separated, and the harmonic analysis for singularities of various geometric types.

1.1 Singularities and Sparsity

As a mathematical idealization of 'image', consider a Schwartz distribution f with domain $\mathbf{R}^2$. The distribution f will be given singularities with specified geometry: points and curves.

We plan to represent such an 'image' using tools of harmonic analysis; in particular bases and frames. While many such representations are conceivable, we are interested here just in those bases or frames which can sparsely represent f.

The type of basis which best sparsifies f depends on the geometry of its singularities. If the singularities occur at a finite number of (variable) points, then *wavelets* give what is, roughly speaking, an optimally sparse representation. If the singularities occur at a finite number of smooth curves, then one of the recently studied directional multiscale representations (*curvelets* or *shearlets*) will do the best job of sparsification.

Since we are concerned with f being a mixture of content types, i.e., points and curves, presumably *both* systems are needed to represent f sparsely.

1.2 Minimum ℓ_1 Decomposition and Perfect Separation

In the early 1990's, R. R. Coifman, Wickerhauser and coworkers became interested in the problem of representing signals using more than one basis and started a first heuristic exploration motivated intuitively, see [5]. A few years later, one of us worked with S .S. Chen to develop a formal, optimization-based approach to the multiple-basis representation problem [4]. Given bases Φ_i, $i = 1, 2$, one solves the following problem

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

thereby exploiting that the ℓ_1 norm has a tendency to find sparse solutions when they exist. This can be regarded as the starting point for ℓ_1 decomposition techniques. For *theoretical work* on this topic we refer to, e.g., [6, 10, 15, 16], and for *empirical work* see, for instance, [9,12,14,15].

For further references we would like to mention the survey paper [1].

1.3 A Geometric Separation Problem

The work just cited, while suggestive and inspiring, concerns discretely indexed signal/image processing, and so is either empirical or else rigorously analytical but not directly relevant to *geometric* separation tasks, which will involve always continuum ideas.
In this paper we develop related methods in a mathematical setting where the notion of successful separation can be made definitionally precise and can be established by mathematical analysis. For this, we pose a simple but clear model problem of geometric separation.
Consider a 'pointlike' object $\mathcal{P}$ made of point singularities:

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

Consider as well a curvelike object $\mathcal{C}$, a singularity along a closed curve $\tau : [0,1] \mapsto \mathbf{R}^2$:

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

where δ_x is the usual Dirac Delta at x. By this choice, we arrange that one of the two distributions does not become dramatically larger than the other as we go to finer and finer scales; rather the ratio of energies is more or less independent of scale. This makes the separation problem challenging at every scale.
Now assume that we observe the 'Signal'

$$f = \mathcal{P} + \mathcal{C}, \tag{1}$$

however, the distributions $\mathcal{P}$ and $\mathcal{C}$ are unknown to us. The *Geometric Separation Problem* now consists in recovering $\mathcal{P}$ and $\mathcal{C}$ from knowledge of f.

1.4 Two Geometric Frames

We focus on two pairs of overcomplete systems for representing the object f:

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

as well as the pair

- *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 and a unified treatment of both the continuous and digital setting.

We pick these because, as is well known, point singularities are coherent in wavelets and curvilinear singularities are coherent in curvelets/shearlets. For the precise definitions we refer to [2, 3], [11, 13], as well as [7].

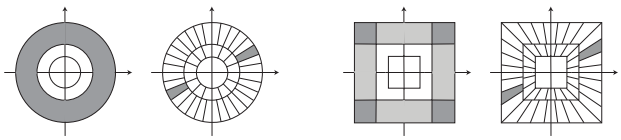

Figure 1: Frequency tilings of radial wavelets and curvelets as well as of orthonormal wavelets and shearlets (from left to right).

Since the scaling subband of each pair are similar as illustrated in Figure 1, we can define two families of filters $(F_j^C)_j$ and $(F_j^S)_j$ which allows to decompose a function f into pieces f_j^C (resp. f_j^S) with different scales j. The piece f_j^C (resp. f_j^S) at subband j arises from filtering f using F_j^C (resp. F_j^S):

$$f_j^C = F_j^C \star f \text{ and } f_j^S = F_j^S \star f,$$

so that the Fourier transform $\hat{f}_i^C$ (resp. $\hat{f}_j^S$) is supported in the scaling subband of scale j of the associated pair of tight frames. The filters are defined in such as way, that we can reconstruct the original function from these pieces using the formula

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

For the precise construction of those filters and further properties, we refer to [7].
We can now use these tools to attack the Geometric Separation Problem scale-by-scale. For this, we filter the model problem (1) to derive the sequences of filtered images

$$f_j^C = \mathcal{P}_j^C + \mathcal{C}_j^C \text{ and } f_j^S = \mathcal{P}_j^S + \mathcal{C}_j^S \text{ for all scales } j. \tag{2}$$

1.5 Outline

In Section 2 we will develop and analyze the decomposition technique based on ℓ_1 minimization we intend to employ, first in a very general Hilbert space setting. These results will then be applied to the scale-dependent Geometric Separation Problem (2) proving that the radial wavelet-curvelet as well as the orthonormal wavelet-shearlet dictionary achieves nearly-perfect separation at all sufficient fine scales (Theorems 1 and 3). The sparsity equivalence between curvelets and shearlets we derive in Subsection 3.2 thereby allows transference of this result from the radial wavelet-curvelet to the orthonormal wavelet-shearlet dictionary.

2. General Component Separation

We now first study the behavior of ℓ^1 minimization in the general two-frame case. 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. Our analysis will center on the use of cluster coherence to exploit the geometric structure of the sparse expansions rather than merely the fact that the vector is sparse.

2.1 Cluster Coherence

Typically, separation results employ the notion of mutual coherence between two tight frames $\Phi = (\phi_i)_i$ and $\Psi = (\psi_j)_j$,

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

whose importance was shown by [6], as a means to impose conditions on the interactions between the dictionary elements. However, this notion is too weak for our purposes. Our novel contribution to sparse recovery and ℓ_1 minimization consists in exploiting the facts that

- the nonzeros of sparse vectors often do not arise in arbitrary patterns, but are rather highly structured, and that
- the interactions between the dictionary elements in ill-posed problems are not arbitrary, but rather geometrically driven.

These key observations lead to the following new notion.

Definition 1. *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**

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

Thus cluster coherence bounds between a single member of frame Ψ and a cluster of members of frame Φ, clustered at $\mathcal{S}$, in contrast to mutual coherence, which can be thought of as singleton coherence.
A related notion called 'cumulative coherence' was introduced in [16], but notice that here we fix a specific set of significant coefficients and do not maximize over all such subsets. The key idea for our analysis is that the index subsets we consider are not abstract, but have a specific geometric interpretation. Maximizing over all subsets with a common combinatorial property would prohibit utilizing this interpretation, hence cumulative coherence is not suitable for our purposes.

2.2 Component Separation by ℓ_1 Minimization

Now consider the following optimization problem:

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

Notice that in this problem, the norm is placed on the **analysis** coefficients rather than on the **synthesis** coefficients as in (BP) to avoid 'self-terms' in the frame expansions. The introduction of cluster coherence now ensures that the principle (SEP) gives a successful approximate separation.

Proposition 1 ([7]). *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.,*

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

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

$$\|S_1^\star - S_1^0\|_2 + \|S_2^\star - S_2^0\|_2 \leq \frac{2\delta}{1 - 2\mu_c},$$

where

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

3. Geometric Separation of Pointlike and Curvelike Structures

3.1 Radial Wavelet-Curvelet Dictionary

The concepts of the previous section will now be applied to $S = f_j^C = \mathcal{P}_j^C + \mathcal{C}_j^C$, our signal of interest from (2). The tight frames are Φ_1, the full radial wavelet frame, and Φ_2, the full curvelet tight frame. The subsignals $S_1^\star, S_2^\star$ we derive by applying the optimization problem (SEP) will be relabel to W_j, the wavelet component, and C_j, the curvelet component.
The main difficulty in applying Proposition 1 consists in choosing the sets of significant coefficients suitably. We achieve this by using microlocal analysis to understand heuristically the location of the significant coefficients in phase space. Roughly speaking, we then employ the Hart-Smith phase space metric defined by

$$\begin{aligned} d((b,\theta);(b',\theta')) &= |\langle e_\theta, b - b' \rangle| + |\langle e_{\theta'}, b - b' \rangle| \\ &\quad + |b - b'|^2 + |\theta - \theta'|^2 \end{aligned}$$

to define an 'approximate' set of significant wavelet coefficients

$$\begin{aligned} \Lambda_{1,j} &= \{\text{wavelet lattice}\} \\ &\quad \cap \{(b,\theta) : d((b,\theta); WF(\mathcal{P})) \leq \eta_j a_j\} \end{aligned}$$

and an 'approximate' set of significant curvelet coefficients

$$\begin{aligned} \Lambda_{2,j} &= \{\text{curvelet lattice}\} \\ &\quad \cap \{(b,\theta) : d((b,\theta); WF(\mathcal{C})) \leq \eta_j a_j\} \end{aligned}$$

for carefully chosen η_j; WF denotes the wavefront set.
Tedious, highly technical estimates then lead to the following separation result:

Theorem 1 ([7]). ASYMPTOTIC SEPARATION USING A RADIAL WAVELET-CURVELET DICTIONARY.

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

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.

3.2 Sparsity Equivalence

We now aim to show that curvelets and shearlets are *sparsity equivalent* in the sense that, for $0 < p \leq 1$, the ℓ_p norm of the curvelet coefficient sequence is finite if and only if the same is true for the shearlet coefficient sequence.
First we observe that for two tight frames $\Phi = (\phi_i)_i$ and $\Psi = (\psi_j)_j$, their cross-Grammian matrix

$$M(i,j) = \langle \phi_i, \psi_j \rangle$$

contains all information on the relation between coefficient sequences $\Phi^T S$ and $\Psi^T S$ for some signal S. Sparsity equivalence can therefore be proven by analyzing the p-norm, $0 < p \leq 1$ defined by

$$\begin{aligned} \|M\|_p &= \max\Big((\sup_i \sum_j |M(i,j)|^p)^{1/p}, \\ &\qquad (\sup_j \sum_i |M(i,j)|^p)^{1/p}\Big) \end{aligned}$$

of a cross-Grammian matrix M.
Now setting $(\sigma_\eta)_\eta$ to be the shearlet tight frame and $(\gamma_\mu)_\mu$ to be the curvelet tight frame, we derive the following result. We remark that the low frequency part has to be dealt with particular care, but for these technicalities we refer to [7].

Proposition 2 ([8]). *For all* $0 < p \leq 1$,

$$\|(\langle \sigma_\eta, \gamma_\mu \rangle)_{\eta,\mu}\|_p < \infty.$$

Using basic estimates from frame theory and the previous proposition, we can show that shearlets and curvelets are indeed sparsity equivalent, thereby allowing us to easily transfer results about sparsity from one system to the other.

Theorem 2 ([8]). *Let* $f \in L^2(\mathbf{R}^2)$ *and* $0 < p \leq 1$. *Then* $\|(\langle f, \sigma_\eta \rangle)_\eta\|_p < \infty$ *if and only if* $\|(\langle f, \gamma_\mu \rangle)_\mu\|_p < \infty$.

3.3 Orthonormal Wavelet-Shearlet Dictionary

Similar to Subsection 3.1, $S = f_j^S = \mathcal{P}_j^S + \mathcal{C}_j^S$ (see (2)) is now our signal of interest, and the tight frames are Φ_1, the full orthonormal wavelet frame, and Φ_2 the full shearlet tight frame. The subsignals $S_1^\star, S_2^\star$, we derive by applying the optimization problem (SEP) will be relabel to W_j, the wavelet component, and S_j, the shearlet component.
The results from Subsection 3.2 as well as similar correspondences between radial wavelets and orthonormal wavelets now form the backbone for the transfer of Theorem 1 to the orthonormal wavelet-shearlet dictionary. Careful application of those to the key estimates in the proof of Theorem 1 leads to a similar result for the orthonormal wavelet-shearlet dictionary.

Theorem 3 ([7]). ASYMPTOTIC SEPARATION USING AN ORTHONORMAL WAVELET-SHEARLET DICTIONARY.

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

4. Conclusion

We first considered signals, being a superposition of two subsignals, each of which is relatively sparse with respect to some tight frame. As a model procedure for separation we considered ℓ_1 minimization of the analysis (rather than synthesis) frame coefficients. By introducing cluster coherence as a new concept for analyzing the interaction of the two tight frames by taking the geometry of the sparse component expansions into account, we derived an estimate for the ℓ_2 norm of the separation error. We then considered signals, which are a superposition of pointlike and curvelike structures. Using the previously derived estimate, we proved that for both pairs of tight frames (radial wavelets/curvelets) as well as (orthonormal wavelets/shearlets) at sufficiently fine scale, nearly-perfect separation is achieved using the model procedure, thereby proposing the orthonormal wavelet-shearlet dictionary as an interesting alternative for geometric separation of pointlike and curvelike structures. The sparsity equivalence between curvelets and shearlets we further proved thereby allows to derive this separation result only for one dictionary and easily transfer it to the other one.

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 and the Department of Mathematics at Yale University for their hospitality and support during her long-term visits. 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. We further thank the anonymous referee for useful comments and suggestions.

References:

[1] A.M. Bruckstein, D.L. Donoho, and M. Elad. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Review* 51:34–81, 2009.

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

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

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

[5] R. R. Coifman and M. V. Wickerhauser. Wavelets and adapted waveform analysis. A toolkit for signal processing and numerical analysis, In *Different*

perspectives on wavelets (San Antonio, TX, 1993), 47:119–153, Proc. Sympos. Appl. Math., Amer. Math. Soc., Providence, RI, 1993.

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

[7] D. L. Donoho and G. Kutyniok. Microlocal Analysis of the Geometric Separation Problem. Preprint, 2009.

[8] D. L. Donoho and G. Kutyniok. Sparsity Equivalence of Anisotropic Decompositions. Preprint, 2009.

[9] 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.* 19:340–358, 2005.

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

[11] 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.

[12] M. Kowalski and B. Torrésani. Sparsity and Persistence: mixed norms provide simple signal models with dependent coefficients. *Signal, Image and Video Processing*, to appear.

[13] G. Kutyniok and D. Labate. Resolution of the Wavefront Set using Continuous Shearlets. *Trans. Amer. Math. Soc.* 361:2719–2754, 2009.

[14] F. G. Meyer, A. Averbuch, and R. R. Coifman. Multilayered Image Representation: Application to Image Compression. *IEEE Trans. Image Proc.* 11:1072–1080, 2002.

[15] 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.* 14:1570–1582, 2005.

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