A hitchhiker's guide to the galaxy of transform-domain sparsification

Evgeniy Lebed and Felix J. Herrmann*
Department of Mathematics and Department of Earth & Ocean Sciences,
University of British Columbia

SUMMARY

The ability to efficiently and sparsely represent seismic data is becoming an increasingly important problem in geophysics. Over the last decade many transforms such as wavelets, curvelets, contourlets, surfacelets, shearlets, and many other types of *'x-lets'* have been developed to try to resolve this issue. In this abstract we compare the properties of four of these commonly used transforms, namely the shift-invariant wavelets, complex wavelets, curvelets and surfacelets. We also briefly explore the performance of these transforms for the problem of recovering seismic wavefields from incomplete measurements.

INTRODUCTION

There exist many different types of transforms which try to achieve a sparse representation of a signal. Numerous sparsifying transforms have been invented and extensively studied, and in this work we compare some of the important properties that these transforms possess. A sparsifying transform is a linear operator mapping a vector of image data to a sparse vector. Different types of images will have sparse representation under different transforms - i.e., there is no single transform that will sparsely represent all types of images. For example, piecewise constant images can be sparsely represented by spatial finite differences. Natural, real-life images are known to have a sparse representation in the discrete cosine transform (DCT) and wavelet transform domains. It has been recently shown that the curvelet transform has a sparse representation of seismic data (the curvelet dictionary consists of prototype waveform, and since seismic data consists of solutions to the wave equation, will we have a good sparse representation in curvelets, see Candès and Demanet (2004)). Surfacelets also try to achieve the same result, but with a much different construction. The newly developed complex-wavelet transform is of interest because it has a fixed and low redundancy factor and has the ability to give signals a multidirectional representation. Classical shift-invariant wavelets are of interest because computationally they provide a very fast multiscale representation of a signal although at a price of a high redundancy factor. The coherence of the transforms is important from a compressive sampling point of view. All of these factors - redundancy, compressibility, computational complexity and coherence play an important role in recovering seismic wavefields from incomplete measurements.

THE FOUR TRANSFORMS

In this section we investigate the construction and properties of shift-invariant wavelets, complex wavelets, surfacelets and curvelets.

The shift-invariant wavelet transform

Numerous variations of the wavelet transforms have been extensively studied and have appeared in literature as early as 1975. For all classes of wavelets the definition is similar; wavelets are scaled and shifted versions of a real valued bandpass wavelet $\psi(t)$, also know as the mother wavelet. When combined with a carefully chosen lowpass scaling function $\phi(t)$, wavelets form an orthonormal basis expansion for arbitrary signals. We can use these wavelet and scaling functions to decompose any signal $x(t)$ as follows

$$x(t) = \sum_{n=-\infty}^{\infty} c(n)\phi(t-n) + \sum_{j=0}^{\infty}\sum_{n=-\infty}^{\infty} d(j,n)2^{j/2}\psi(2^j t - n) \tag{1}$$

where $c(n)$ and $d(j,n)$ are the scaling and wavelet coefficients which are computed by

$$c(n) = \int_{-\infty}^{\infty} x(t)\phi(t-n)dt \tag{2}$$

$$d(j,n) = 2^{j/2}\int_{-\infty}^{\infty} x(t)\psi(2^j t - n)dt \tag{3}$$

These coefficients provide a multiscale analysis of the signal at scale j at different times n. The shift-invariant wavelet transform (sometimes called the undecimated discrete wavelet transform UDWT) was independently discovered by several authors, see for example Shensa (1992). From a filter bank point of view both the odd and even downsamples are kept and the lowpass bands are further split to provide more scales in the decomposition. In the synthesis the odd and even parts are inverted and the result is averaged. The key properties of the UDWT are given below;

1. The UDWT is shift-invariant. i.e. a small shift in the signal does not perturb UDWT's coefficient oscillation pattern around singularities.
2. In 2D the UDWT has a redundancy factor of $3L+1$ where L is the number of scales in the decomposition.

It is important to note that although computational complexity of UDWT is $O(LN)$ (which is at most $O(NlogN)$) ,the redundancy factor of UDWT is exceedingly high for practical applications. If we are dealing with a moderatly-sized dataset that we wish to analyze, perhaps with say scales, this already amounts to a redundancy factor of 19, which is unpractical for a lot of applications.

The dual-tree complex wavelet transform ($\mathbb{C}$WT)

The complex dualt-tree framework proposed by Selesnick et al. (2005) extends the work of Kingsbury (2560) to use the the filter bank implementation along with two real desecrate wavelet transforms (DWTs) which use different sets of filters and satisfy the perfect reconstruction conditions. One of the DWTs

gives the real part of the ℂWT's coefficients while the other DWT gives the imaginary part of the coefficients. These filters are designed in such a way so that the overall ℂWT is approximately analytic. If we let $h_0(n)$ and $h_1(n)$ be the lowpass/highpass pair for one of the filterbanks and $g_0(n)$ and $g_1(n)$ be the lowpass/highpass pair for the other filterbank than we can display the analysis/synthesis for the ℂWT in schematical format as shown in Figure 1.

These filters are designed in such a way so that the wavelet

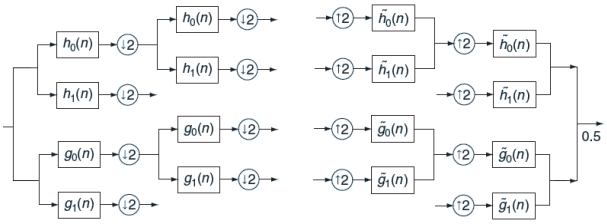


Figure 1: Left: two levels of the analysis filter bank of ℂWT. Right: synthesis filter bank of ℂWT.

$\psi(t) := \psi_h(h) + i\phi_g(t)$, where $\psi_g(t)$ is the Hilbert transform of $\psi_h(t)$, is approximately analytic. The key properties of the ℂWT are outlined below;

1. The ℂWT is nearly shift invariant.
2. The ℂWT is a multiscale transform with 12 distinct directions at every scale.
3. A complex wavelet is strictly localized in the spatial domain.
4. Regardless of the number of scales used in the decomposition, the redundancy factor of ℂWT is always 4 in 2D.

The curvelet transform

The shift-invariant and complex wavelet transforms provide a multiscale representation of signals. The ℂWT provides a limited number of directions in the representation. We now explore two transforms which provide an arbitrary number of directions in the decomposition.

The 2 dimensional discrete curvelet transform, proposed by Candès and Donoho (2002), takes a 2D signal of the form $f(n_1,n_2)$, $0 \le n_1,n_2 < n$ as input and outputs a collection of coefficients that are obtained by computing inner products of the signal with curvelet window functions at different scales and directions. The curvelet coefficients $c(j,l,k)$ at scale j, orientation l and spatial location k are defined by

$$c(j,l,k) = \sum_{n_1,n_2} \overline{\varphi_{j,l,k}(n_1,n_2)} f(n_1,n_2), \qquad (4)$$

where $\varphi_{j,l,k}(n_1,n_2)$ is the curvelet function. The outline of the curvelet transform algorithm is as follows; A 2D fast Fourier transform is applied to $f(n_1,n_2)$ to get $\hat{f}(\omega_1,\omega_2)$ with $-n/2 \le \omega_1,\omega_2 < n/2$. Next the Fourier samples of the input image are multiplied with the curvelet window functions $\tilde{U}_{j,l}(\omega_1,\omega_2)$ at different scales and directions to obtain the product $a_{j,l} = \tilde{U}_{j,l}(\omega_1,\omega_2)\hat{f}(\omega_1,\omega_2)$. This product is wrapped around the origin to produce $W(a_{j,l})(\omega_1,\omega_2)$. No wrapping is required at coarsest and finest scales. Lastly, a windowed 2D inverse fast Fourier transform is applied and the curvelet coefficients $c(j,l,k)$ are collected.

We list some of the main properties of the curvelet transform below.

1. The curvelet transform provides a tight-frame expansion for any function $f(x_1,x_2) \in L^2(\mathbf{R}^2)$ as a series of curvelets $f = \sum_{j,l,k} \langle \varphi_{j,l,k}, f \rangle \varphi_{j,l,k}$
2. A single curvelet is strictly localized to one angular wedge in the frequency domain and has fast decay in the spatial domain. The effective support of $\varphi_{j,l,k}(x)$ obeys a parabolic scaling relation with $width \propto length^2$. At scale j a curvelet is a 'fat needle' of length $2^{-j/2}$ and width 2^{-j}.
3. A curvelet is smooth along its major axis and is oscillatory along its minor axis.
4. The curvelet transform is approximately 8 times redundant in 2D.

The surfacelet transform

Similar to the curvelet transform, the surfacelet transform, proposed by Do and Lu (2007) is another redundant multiscale and multidirectional signal decomposition system for signal of 2 or more dimensions. The surfacelet transforms conceptually begins from a much different construction than the curvelet transform. The surfacelet construction begins by extending the directional filter bank (DFB) that was proposed by Bamberger and Smith (1992) to an arbitrary number ($N \ge 2$) of dimensions. The DFB partitions the frequency plane into $2^l, l \in \mathbb{N}$ triangular wedges of equal area by radial lines passing through the origin of the (ω_1,ω_2) plane. In the case for a general N, the n-dimensional filter bank (NDFB) will partition the frequency spectrum by rectangular based pyramids radiating from the origin. Lets consider a 3D example; to achieve the desired frequency partitioning, the NDFB (now $N = 3$) partitions the frequency cube into three hourglass-shaped subbands that are aligned the the ω_1,ω_2 and ω_3 axes, respectively. Next, two Iteratively Resampled Checkerboard filterbanks (IRC), that act along the (n_1,n_2) and (n_1,n_3) planes, are applied to each of the hour-glass shaped frequency subbands. The resulting output is $3 \cdot 2^l$ directional subbands. This NDFB is a critical part of the surfacelet transform; it efficiently captures surface singularities by a tree-stuctured bank. One desirable aspect of the transform is the multiresolution property. We would like to be able to decompose signals not just into multidirectional components but also into a multiscale structure to be able to capture singularities of different size. The combination of the NDFB with a multiscale structure is similar to the contourlet construction, proposed by Do and Vetterli (2005). The resulting algorithmic construction is the following; the frequency spectrum of the signal gets inputed into first level of the multiscale pyramid where a highpass filter $H(\omega)$ and a lowpass filter $L(\omega)$ act on the signal. The output of $H(\omega)$ is inputed into the NDFB, and this returns the fine scale directional approximation to the signal. At the lowpass branch of the pyramid the signal is upsampled by a factor of two followed by an anti-alliasing filter $S(\omega)$, and then the signal is downsampled by a factor of 3. The output of this branch is a coarse approximation to the signal, which gets inputed into the next level of the

multiscale pyramid. At the subsequent lowpass branches $S(\omega)$ is no longer applied and the signal only gets downsampled by a factor of 2. Once again, we list the essential properties of the surfacelet transforms;

1. Just like the curvelet transform, the 2D surfacelet transform provides a redundant, multiscale and multidirectional decomposition for functions $f(x_1,x_2) \in L^2(\mathbf{R}^2)$. The transform is tight frame.
2. Unlike curvelets, which are strictly localized to one angular wedge, a surfacelet is spread across an entire scale in the frequency domain. In the spatial domain surfacelets also obey a parabolic scaling relation of $width \propto length^2$, with most of the oscillations occurring along the minor axis.
3. In the 2D case the surfacelet transform is approximately 5 times redundant.

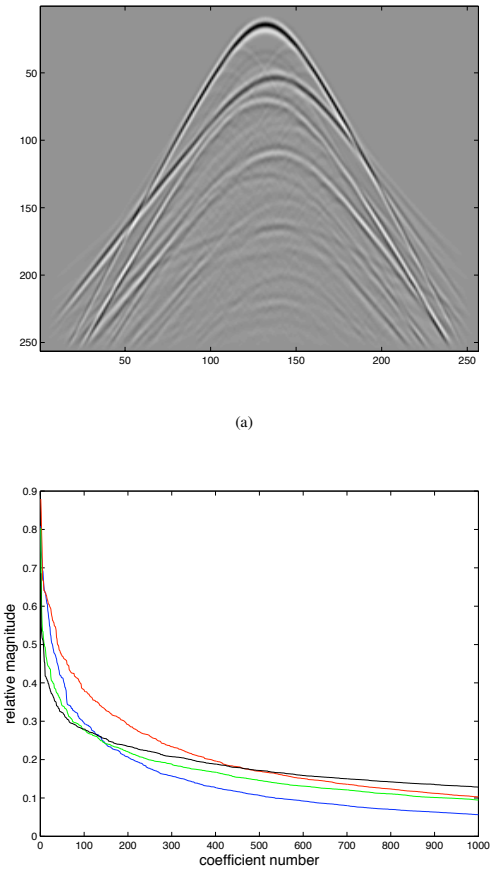
Nonlinear approximation

The ability to sparsely represent seismic data is often a necessary condition for many ℓ_1 solvers that are used either for denoising or for signal recovery and in this section we compare the coefficient decay rate of the four different transforms. We take the seismic signal shown in Figure 2(a) as input and plot it's normalized transform coefficients in Figure 2(b). To make a fare comparison of the decay rates we must account for the different redundancies of the transforms. To remedy this problem we randomly sample the coefficients - i.e. if the curvelet transform is 8 times redundant, we randomly remove 7 out of 8 coefficients, and plot the remaining ones in decreasing order. Similar procedure is done to the other transforms' coefficients.

Coherence

The coherence of a measurement basis $\mathbf{M}$ and a transform matrix $\mathbf{S}$ is defined by $\sqrt{m} \cdot \max_{k,l} |\mathbf{m}_k(\mathbf{s}_l)^H|$ with $\mathbf{m}_k$ the k^{th} row of $\mathbf{M}$ and $\mathbf{s}_l$ the l^{th} row of $\mathbf{S}$. In this section we investigate the coherence of the four transforms from a qualitative point of view. In the problem of recovering seismic wavefields from incomplete data (either from a subset of traces or temporal frequencies), it is important to have incoherency between the measurement basis and the sparsifying tranform. Figure 4 shows the frequency response of the four transforms at the same scale. From Figures 3 and 4 we can see that a curvelet is strictly localized in the frequency domain. As for surfacelets, although most of the energy is localized to one angular wedge, we can see that there is some energy leakage occurring. Unlike a single curvelet, which is localized to one angular wedge, a surfacelet is spread across all wedges at a given scale. As for the other two transforms, due to their construction, it turns out that complex wavelets and shift-invariant wavelets are strictly localized in the spatial domain, and in the frequency domain they are spread out.

In the problem of recovering seismic signals from missing temporal frequencies we can expect to see better results from transforms that are more spread in the frequency domain and hence more incoherent (i.e. shift invariant wavelets) while in the problem of recovering signals with missing traces we expect to see the best results from transforms that are more incoherent in the physical domain (i.e. curvelets).



(a)

(b)

Figure 2: (a) Seismic signal; (b) Randomly sampled transform coefficients of the four transforms; blue-curvelets, red-surfacelets, black-wavelets, green- complex wavelets

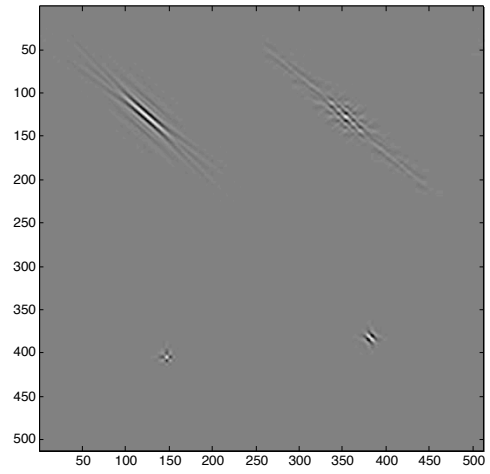


Figure 3: Top left- curvelet; top right- surfacelet; bottom left-wavelet; bottom right- complex wavelet in the spatial domain

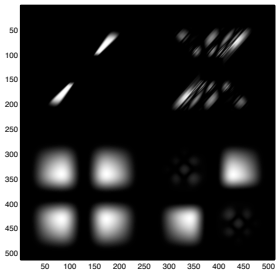


Figure 4: Frequency response of one curvelet, sufacelet, wavelet and complex wavelet. Top left- curvelet; top right- surfacelet; bottom left-wavelet; bottom right- complex wavelet

APPLICATIONS

The application of the above-descrived transforms that we are considering in this section is recovering seismic wavefields from (1) subset of traces and (2) a subset of temporal frequencies. For more applications see other contributions by the authors in the proceedings of this conference. To briefly describe our problem we assume the following foreword model: $\mathbf{y} = \mathbf{RMf}_0 + \mathbf{n}$ where $\mathbf{y}$ is a subgroup of the full data $\mathbf{f}_0$. The matrix $\mathbf{M}$ refers to the measurement basis (in problem (1) $\mathbf{M}$ is the Dirac measurement basis and in (2) $\mathbf{M}$ is the Fourier measurement basis) and the matrix $\mathbf{R}$ is a restriction operator; it extracts those rows from $\mathbf{M}$ that represent the samples the are actually acquired.
We want to find an approximation for $\mathbf{f}$ by a sparse superposition of transform coefficients. The approximation to $\mathbf{f}$ is given by $\tilde{\mathbf{f}} = \mathbf{S}^H\tilde{\mathbf{x}}$ where $\mathbf{x}$ is obtained by solving the following optimization problem

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{RMS}^H\mathbf{x}\|_2 \leq \varepsilon \qquad (5)$$

and $\mathbf{S}^H$ refers to the synthesis matrix of one of the four transforms domains described above. Three key ingredients are required to successfully resolve this problem. They are (1) a good sampling scheme and for this we use the jitter sampling proposed by Hennenfent and Herrmann (2008) (2) a good ℓ_1 solver and for this we use iterative soft thresholding, proposed by Daubechies et al. (2004) and lastly (3) a good sparsifying transforms for the seismic data. We use these sparsifying domains to solve two problems; signal recovery from (1) missing traces and (2) missing temporal frequencies.

The results in Figure 5 show the recovered signals by the four different transforms when we restrict 60% of the traces. In this case the SNRs are: curvelets: 6.42 dB, surfacelets: 4.72 dB; shift-invariant wavelets: 3.54 dB, and complex wavelets: 3.88 dB.Figure 6 shows the recovered results when we restrict 60% of the temporal frequencies. In this case the SNRs are:curvelets: 5.45 dB, surfacelets: 6.16 dB; shift-invariant wavelets: 9.03 dB, and complex wavelets: 7.04 dB.

CONCLUSIONS

Without claiming to be exhaustive, we have presented a brief comparison on construction and properties of four different transforms. Although all the constructions are conceptually

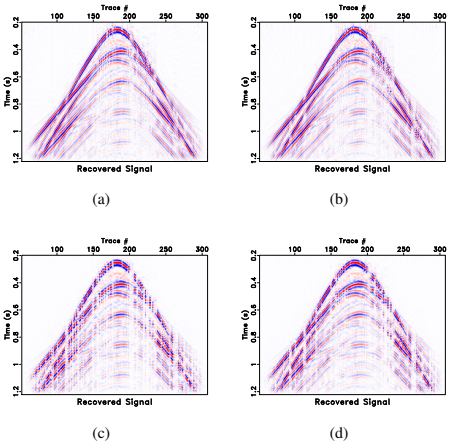


Figure 5: (a) Recovered signal from missing traces by (a) curvelets (b) surfacelets (c) shift-invariant wavelets (d) complex wavelets

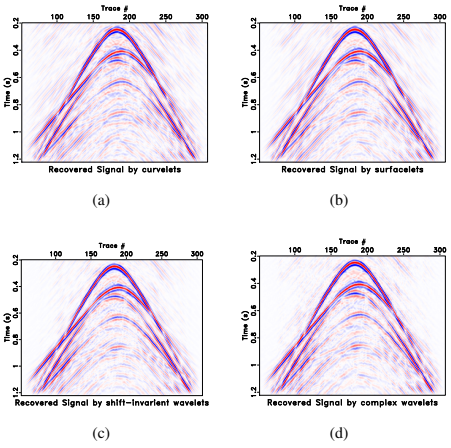


Figure 6: (a) Recovered signal from missing temporal frequencies by. (a) curvelets (b) surfacelets (c) shift-invariant wavelets (d) complex wavelets

different, they all try to achieve the same results - to be able to sparsely represent and efficiently capture certain aspects (like singularities) of 2D signals. In the problem of recovering seismic wavefields from missing temporal frequencies we observed the best results from the shift-invariant wavelet transform, although at a cost of high redundancy. Out of the four transforms, shift-invariant wavelets are most incoherent with the Fourier measurement basis. When recovering the signal from missing traces, the best results were observed by using curvelets, which are most incoherent with the Dirac measurement basis. Therefore The success of of the recovery of the signal appears to be an intricate play of mutual coherence, redundancy and compressibility of a signal in some transform domain.

REFERENCES

Bamberger, R. H. and M. J. T. Smith, 1992, A filter bank for the directional decomposition of images: theory and design: IEEE Transactions. Image processing, **40**, 882–893.

Candès, E. J. and L. Demanet, 2004, The curvelet representation of wave propagators is optimally sparse: Comm. Pure Appl. Math, **58**, 1472–1528.

Candès, E. J. and D. L. Donoho, 2002, New tight frames of curvelets and optimal representations of objects with C^2 singularities: Technical report, Caltech.

Daubechies, I., M. Defrise, and C. de Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint: Comm. Pure Appl. Math., 1413–1457.

Do, M. and Y. Lu, 2007, Multidimensional directional filter banks and surfacelets: IEEE Transactions. Image processing, **16**, 918–931.

Do, M. N. and M. Vetterli, 2005, The contourlet transform: an efficient directional multiresolution image representation: IEEE Transactions. Image processing, **14**, 2019–2106.

Hennenfent, G. and F. J. Herrmann, 2008, Simply denoise: wavefield reconstruction via jittered undersampling: Geophysics, **73**.

Kingsbury, N. G., 2560, Image procesing with complex wavelets: Philos. Trans. R. Soc. London A, Math. Phys. Sci., **357**, 2543.

Selesnick, I. W., R. G. Baraniuk, and N. G. Kingsbury, 2005, The dual-tree complex wavelet transform: IEEE Transactions. Signal processing, 123–151.

Shensa, M., 1992, The discrete wavelet transform: Wedding the à torus and mallat algorithms: IEEE Trans. Inform. Theory, **40**, 2464–2482.