

Curvelet-based primary-multiple separation from a Bayesian perspective

Rayan Saab, ECE-UBC, Deli Wang, Jilin University. Özgür Yılmaz, Math-UBC and Felix J. Herrmann, EOS-UBC

SUMMARY

In this abstract, we present a novel primary-multiple separation scheme which makes use of the sparsity of both primaries and multiples in a transform domain, such as the curvelet transform, to provide estimates of each. The proposed algorithm utilizes seismic data as well as the output of a preliminary step that provides (possibly) erroneous predictions of the multiples. The algorithm separates the signal components, i.e., the primaries and multiples, by solving an optimization problem that assumes noisy input data and can be derived from a Bayesian perspective. More precisely, the optimization problem can be arrived at via an assumption of a weighted Laplacian distribution for the primary and multiple coefficients in the transform domain and of white Gaussian noise contaminating both the seismic data and the preliminary prediction of the multiples, which both serve as input to the algorithm.

INTRODUCTION

Predictive multiple suppression methods consist of two main steps: a prediction step, during which multiples are predicted from seismic data, and a primary-multiple separation step, during which the predicted multiples are 'matched' with the true multiples in the data and subsequently removed. This second separation step, which we will call the estimation step, is crucial in practice. An incorrect separation will cause residual multiple energy in the result or may lead to a distortion of the primaries, or both. To reduce these adverse effects, a new transform-domain method was proposed by Herrmann et al. (2007), where primaries and multiples are separated in the curvelet domain rather than matched.

This separation method is carried out on the basis of differences in the multiscale and multidirectional characteristics of the primaries and multiples. The sparsity in the curvelet domain of these two signal components is exploited by solving a weighted ℓ_1 -norm minimization problem that is designed to drive the two signal components apart. This separation is accomplished by using the (SRME) predicted primaries and multiples as weights. During the optimization, the two signal components are separated by enhancing sparseness (through weighted ℓ_1 -norms) in the transform domain subject to fitting the observed data as the sum of the separated components to within a user-defined tolerance level.

Even though good results were obtained with this method, problems with the appropriate balancing of the weights were experienced. Lack of a good balance yielded a 'separation' where all the signal energy ends up in one of the two signal components. To tackle this issue, we propose an alternative formulation for the primary-multiple separation where an additional constraint is imposed on the coefficients for the multiples. This formulation follows from a Bayesian approach where sparsity

for the two signal components as well as data misfits are taken into account.

Our approach derives from techniques employed in the field of signal processing and geophysics where Bayesian approaches supplemented with sparsity-promoting Laplace/Cauchy probability density functions are frequently used. Our method differs from the earlier results in the literature by (i) using sparsity promoting curvelets (see, e.g., Candes et al., 2006); (ii) using inaccurate predictions for the to-be-separated signal components and (iii) providing a new iterative solution technique based on descent projections and soft thresholding.

The outline of our abstract is as follows. First, we briefly discuss curvelets, our transform domain of choice, which has been shown to admit sparse coefficients for seismic data (Candes et al. (2006)). Next, Bayesian formulation of the primary-multiple separation problem is introduced, leading to the objective function to be minimized. Finally, a separation algorithm that *provably* converges to a minimizer of the objective function is given. Results of our algorithm applied to a synthetic shot with 3-D complexity are presented as well.

CURVELETS

Curvelets are localized 'little plane-waves' (see e.g. Hennenfent and Herrmann, 2006, and ancillary electronic material) that are oscillatory in one direction and smooth in the other direction(s). They are multiscale and multi-directional. Curvelets have an anisotropic shape – they obey the so-called parabolic scaling relationship, yielding a width $\propto \text{length}^2$ for the support of curvelets. This anisotropic scaling is optimal for detecting wavefronts and explains their high compression rates on seismic data and images (Candes et al., 2006; Hennenfent and Herrmann, 2006; Herrmann et al., 2007).

Curvelets represent a specific tiling of the 2-D/3-D frequency plane into strictly localized multiscale and multi-angular wedges. Fig. 1(b)). Because the directional sampling increases every-other scale-doubling, curvelets become more anisotropic at finer scales. They become 'needle-like' as illustrated in Fig. 1(a). The smoothness in the Fourier domain leads to a rapid decay in the physical domain. Their effective support is given by ellipsoids with a width $\propto 2^{j/2}$ and length $\propto 2^j$ and an angle $\theta_{jl} = 2\pi l 2^{\lfloor j/2 \rfloor}$ with j the scale, l the angular index with $\lfloor \cdot \rfloor$ a rounding off towards the nearest smaller integer. Curvelets are indexed by the multi-index $\mu := (j, l, k) \in \mathcal{M}$ with \mathcal{M} the multi-index set that runs over the scales, j , angles, l , and positions k (see for details Candes et al., 2006; Ying et al., 2005). The number of wedges doubles every other scale-doubling and such that directional selectivity increases for finer scales (see Fig. 1(b)). Curvelets compose an arbitrary column vector \mathbf{f} , with the reordered samples, according to $\mathbf{f} = \mathbf{C}^T \mathbf{C} \mathbf{f}$ with \mathbf{C} and \mathbf{C}^T , the forward/inverse discrete curvelet transform matrices (defined by the fast discrete curvelet transform, FDCT, with

Curvelet-based primary multiple separation

wrapping Candes et al., 2006; Ying et al., 2005). C^T denotes the transpose of C , which is equal to the pseudoinverse of C for our choice of discrete curvelet transform: the rows of C constitute a tight frame with frame bound 1 and with a moderate redundancy (a factor of roughly 8 for $d = 2$ and 24 for $d = 3$ with d the number of dimensions). Such tight frames (see e.g. Daubechies, 1992) are signal representations that preserve energy. Consequently, $C^T C = \mathbf{I}$ with \mathbf{I} the identity matrix. However, $CC^T \neq \mathbf{I}$.

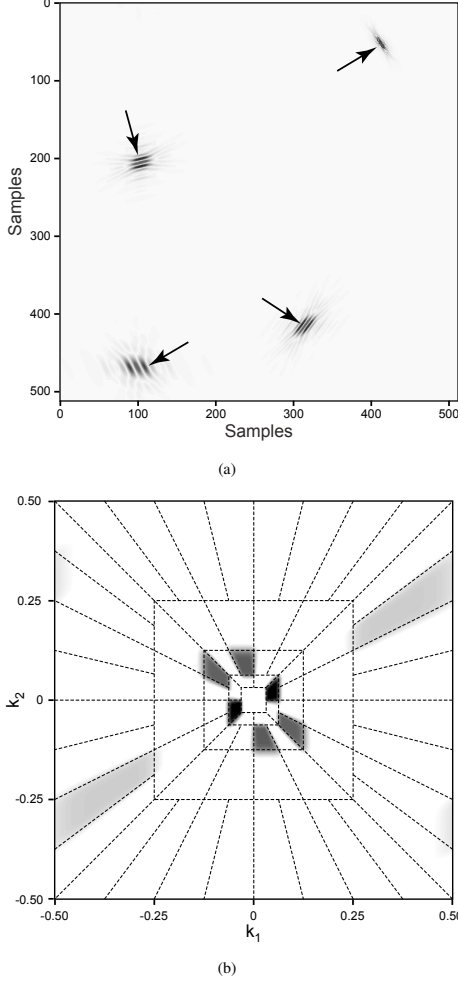


Figure 1: Curvelet examples. **(a)-(b)** spatial and frequency representation of four different curvelets in the spatial domain at three different scales and in the Fourier domain **(b)**. The Fourier spectrum is overlaid with the dyadic partitioning of curvelets in the frequency. Each wedge corresponds to the compact frequency support of a curvelet. The curvelet transform itself consists for each wedge of a collection of shifted curvelets on a scale-dependent grid. Curvelets are not strictly localized in the space but they are of rapid decay. They oscillate in one direction and are smooth in the other(s).

BAYESIAN FORMULATION

Here, we approach the primary-multiple separation problem from a probabilistic perspective, where we impose an a priori distribution for the unknown coefficients of the primaries and multiples. Estimates for the curvelet coefficients of the primaries and multiples are obtained by using a Laplacian-type probability density function for the curvelet coefficients and weights defined in terms of (SRME) predictions for the two signal components. The estimates are calculated within a Bayesian framework which leads to the minimization of a weighted sparsity-promoting functional. The choice of a Laplacian-type density function for the curvelet coefficients also serves as a sparsity promoting prior (see e.g. Zibulevsky and Pearlmutter, 2001; Li et al., 2004) and is consistent with the high compression rates that curvelets display for seismic data (Candes et al., 2006; Hennenfent and Herrmann, 2006; Herrmann et al., 2007).

More precisely, let

$$\mathbf{b} = \mathbf{s}_1 + \mathbf{s}_2 + \mathbf{n} \quad (1)$$

be the recorded total data set comprised of a noisy sum of primaries \mathbf{s}_1 and multiples \mathbf{s}_2 and let

$$\mathbf{b}_2 = \mathbf{s}_2 + \mathbf{n}_2 \quad (2)$$

be the predicted multiples, assumed to be erroneous versions of the true underlying multiples. Moreover, let the multiples and primaries admit (curvelet) transform coefficients satisfying

$$\mathbf{s}_i = \mathbf{A}\mathbf{x}_i, \quad \{i = 1, 2\}, \quad (3)$$

where \mathbf{A} is the inverse curvelet transform operator (defined by the fast discrete curvelet transform (FDCT by wrapping, see e.g. Candes et al., 2006; Ying et al., 2005)), which is linear. We can now rewrite the above equations as

$$\begin{aligned} \mathbf{b}_2 &= \mathbf{A}\mathbf{x}_2 + \mathbf{n}_2 \\ \mathbf{b}_1 &= \mathbf{A}\mathbf{x}_1 + \mathbf{n} - \mathbf{n}_2. \end{aligned} \quad (4)$$

Thus, our objective is to estimate the curvelet coefficients of the primaries \mathbf{x}_1 and multiples \mathbf{x}_2 , given the predictions \mathbf{b}_1 and \mathbf{b}_2 . These predictions could be given by SRME (see e.g. Berkhouit and Verschuur, 1997) or by other means.

From a Bayesian point of view, this means that our objective is to find \mathbf{x}_1 and \mathbf{x}_2 such that the conditional probability, $P(\mathbf{x}_1, \mathbf{x}_2 | \mathbf{b}_1, \mathbf{b}_2)$, is maximized. In other words, we aim to maximize

$$\begin{aligned} P(\mathbf{x}_1, \mathbf{x}_2 | \mathbf{b}_1, \mathbf{b}_2) &= \frac{P(\mathbf{x}_1, \mathbf{x}_2)P(\mathbf{b}_1 | \mathbf{x}_1, \mathbf{x}_2)P(\mathbf{b}_2 | \mathbf{b}_1, \mathbf{x}_1, \mathbf{x}_2)}{P(\mathbf{b}_1, \mathbf{b}_2)} \\ &= \frac{P(\mathbf{x}_1, \mathbf{x}_2)P(\mathbf{n})P(\mathbf{n}_2)}{P(\mathbf{b}_1, \mathbf{b}_2)}. \end{aligned} \quad (5)$$

Since both \mathbf{b}_1 and \mathbf{b}_2 are known, we try to find the curvelet coefficients for the primaries and multiples that maximize the posterior probability in Eq. 5, i.e.,

$$\arg \max_{\mathbf{x}_1, \mathbf{x}_2} P(\mathbf{x}_1, \mathbf{x}_2 | \mathbf{b}_1, \mathbf{b}_2) = \arg \max_{\mathbf{x}_1, \mathbf{x}_2} P(\mathbf{x}_1, \mathbf{x}_2)P(\mathbf{n})P(\mathbf{n}_2). \quad (6)$$

If we assume independent identically distributed white Gaussian noise distributions for \mathbf{n} and \mathbf{n}_2 , with possibly different

Curvelet-based primary multiple separation

variances, and independent weighted Laplacian priors for \mathbf{x}_1 and \mathbf{x}_2 , then it can be shown that equation (6) reduces to

$$\arg \max_{\mathbf{x}_1, \mathbf{x}_2} P(\mathbf{x}_1, \mathbf{x}_2 | \mathbf{b}_1, \mathbf{b}_2) = \arg \min_{\mathbf{x}_1, \mathbf{x}_2} f(\mathbf{x}_1, \mathbf{x}_2)$$

with

$$f(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1\|_{1, \mathbf{w}_1} + \|\mathbf{x}_2\|_{1, \mathbf{w}_2} + \|\mathbf{A}\mathbf{x}_2 - \mathbf{b}_2\|_2^2 + \eta \|\mathbf{A}(\mathbf{x}_1 + \mathbf{x}_2) - (\mathbf{b}_1 + \mathbf{b}_2)\|_2^2, \quad (7)$$

where $\|\mathbf{x}_i\|_{1, \mathbf{w}_i} = \sum_{\mu} |w_{i, \mu} x_{i, \mu}|$, $\mu \in \mathcal{M}$ is a weighted ℓ_1 -norm. The weights are given by the predictions for the two signal components, i.e., $\mathbf{w}_1 = \lambda_1 |\mathbf{A}^T \mathbf{b}_2|$ and $\mathbf{w}_2 = \lambda_2 |\mathbf{A}^T \mathbf{b}_1|$ with \mathbf{A}^T the forward curvelet transform, and λ_1 and λ_2 parameters that control the sparsity of the coefficient vectors. The parameter η controls the tradeoff between fitting the total data and fitting the predicted multiples.

Hence, the maximization of the posterior probability, $P(\mathbf{x}_1, \mathbf{x}_2 | \mathbf{b}_1, \mathbf{b}_2)$, leads to a minimization problem involving the weighted ℓ_1 -norms of the coefficients and the following data misfits: the misfit between the predicted and estimated multiples and between the sum of the estimated primaries and multiples and the observed total data. As previously mentioned, the above formalism is similar to approaches in the field of signal processing where Laplacian priors have been used extensively (see e.g. Zibulevsky and Pearlmutter, 2001; Li et al., 2004). Long-tailed Laplacian or Cauchy-distributions also have extensively been used in (exploration) seismology (see e.g. Sacchi and Ulrych, 1996). What is new in our approach, is the use of the sparsity promoting curvelet transform and the use of weights to formulate the primary-multiple separation problem. Moreover, the above approach differs from earlier work (see e.g. Herrmann and Verschuur, 2005; Herrmann et al., 2007) since it adds additional constraints that guarantee the consistency of the estimated multiples with the predicted multiples which are the input to our algorithm. This additional constraint prevents all the signal's energy from ending up in one of the two unknown coefficients vectors.

SEPARATION ALGORITHM

The primary-multiple separation thus entails the minimization of the objective function $f(\mathbf{x}_1, \mathbf{x}_2)$ for appropriately chosen $\lambda_1, \lambda_2, \eta$. To minimize $f(\mathbf{x}_1, \mathbf{x}_2)$ in equation 7 above, we devised an iterative thresholding algorithm in the spirit of the one developed by Daubechies et al. (2003). The proposed algorithm provably converges to the minimizer of $f(\mathbf{x}_1, \mathbf{x}_2)$, provided that all the weights in the vectors $\mathbf{w}_i, i \in \{1, 2\}$ are strictly positive. Starting from any initial estimates \mathbf{x}_1^0 and \mathbf{x}_2^0 of \mathbf{x}_1 and \mathbf{x}_2 , the iterations of the algorithm proceed as

$$\begin{aligned} \mathbf{x}_1^{n+1} &= \mathbf{S}_{\frac{\mathbf{w}_1}{2\eta}} (\mathbf{A}^T \mathbf{b}_2 - \mathbf{A}^T \mathbf{A} \mathbf{x}_2^n + \mathbf{A}^T \mathbf{b}_1 - \mathbf{A}^T \mathbf{A} \mathbf{x}_1^n + \mathbf{x}_1^n) \\ \mathbf{x}_2^{n+1} &= \mathbf{S}_{\frac{\mathbf{w}_2}{2(1+\eta)}} \left[\mathbf{A}^T \mathbf{b}_2 - \mathbf{A}^T \mathbf{A} \mathbf{x}_2^n + \mathbf{x}_2^n + \frac{\eta}{\eta+1} (\mathbf{A}^T \mathbf{b}_1 - \mathbf{A}^T \mathbf{A} \mathbf{x}_1^n) \right] \end{aligned} \quad (8)$$

where \mathbf{S}_{α} is the soft-thresholding operator acting *elementwise* as

$$S_{\alpha_{\mu}}(v_{\mu}) = \text{sgn}(v_{\mu}) \cdot \max(0, |v_{\mu}| - |\alpha_{\mu}|). \quad (9)$$

EXPERIMENT AND RESULTS

To test the performance of our algorithm, it is applied on one shot of a three-dimensional data volume generated by a subsurface velocity model with two-dimensional inhomogeneities. This velocity model consists of a high-velocity layer, which represents salt, surrounded by sedimentary layers and a water bottom that is not completely flat. Applying an acoustic finite-difference modeling algorithm, 361 shots with 361 receivers are simulated on a fixed receiver spread with receivers located from 0 to 5400 m with steps of 15 m. The complete prestack dataset can be represented as a three-dimensional volume along the shot, receiver and time coordinates from which multiples are predicted and removed with the iterative matched-filter method introduced by Berkhou and Verschuur (1997) (see figure 2). The results of the above method are compared to those obtained by using the method of Herrmann et al. (2007) as well as to those obtained by the proposed algorithm (applied here to one shot of the data). The results can be seen in figure 3. The comparison shows an improved removal of late arriving multiple energy by using the curvelet-based algorithms. However, the thresholded result of Herrmann et al. (2007) shows a large remaining multiple around 0.6s and an offset of 2400m which is no longer visible in the result obtained with our iterative method. Moreover, the artifacts that the thresholded method generates appear much less severe in the proposed method.

DISCUSSION AND CONCLUSION

We have presented a primary-multiple separation algorithm which uses seismic data as well as predictions of the primaries and multiples as input and exploits the sparsity of the components in the curvelet domain to perform the separation. The algorithm, which minimizes an objective function involving weighted ℓ_1 -norms of the curvelet coefficients as well as misfits with the total data and the predicted multiples was derived from a Bayesian formulation of the separation problem. Moreover, the algorithm was applied to a synthetic data set and the results were compared to those of SRME and of the first iteration of curvelet-based method proposed by Herrmann et al. (2007). Our algorithm was shown to produce favorable results, by removing late arriving multiple energy when compared to SRME results. It was also shown to remove large multiples that remained upon applying the single-threshold algorithm and also to produce less artifacts. The improved performance relative to the SRME method can be partly attributed to the ability of the curvelet transform to represent seismic data, while the improved results of the proposed algorithm relative to the thresholded results can be attributed to the use of the new cost function $f(\mathbf{x}_1, \mathbf{x}_2)$ which mitigates the effect of artifacts and is simultaneously relatively robust to the choice of control parameters.

Curvelet-based primary multiple separation

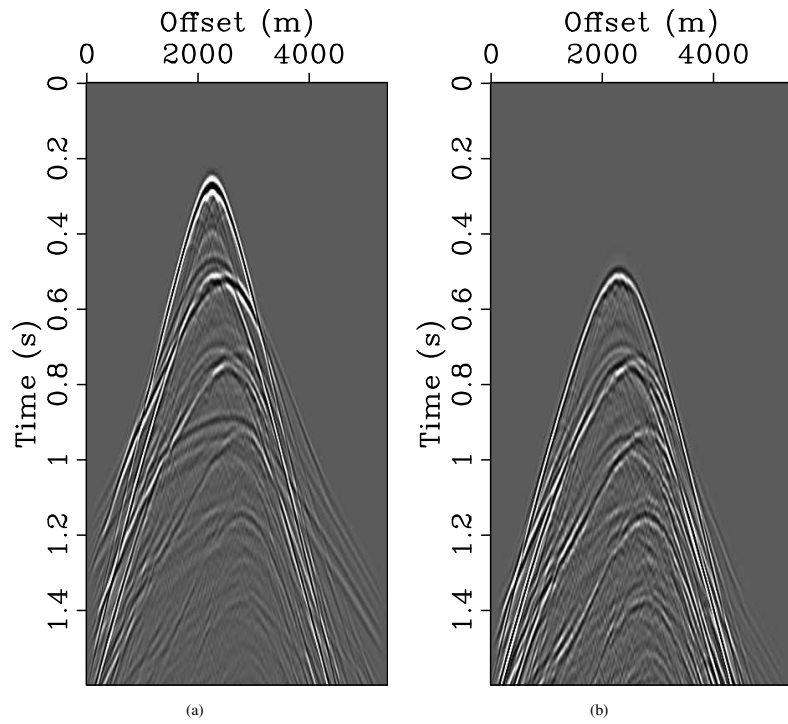


Figure 2: Synthetic data and SRME-predicted multiples. (a) Total data. (b) SRME-predicted multiples .

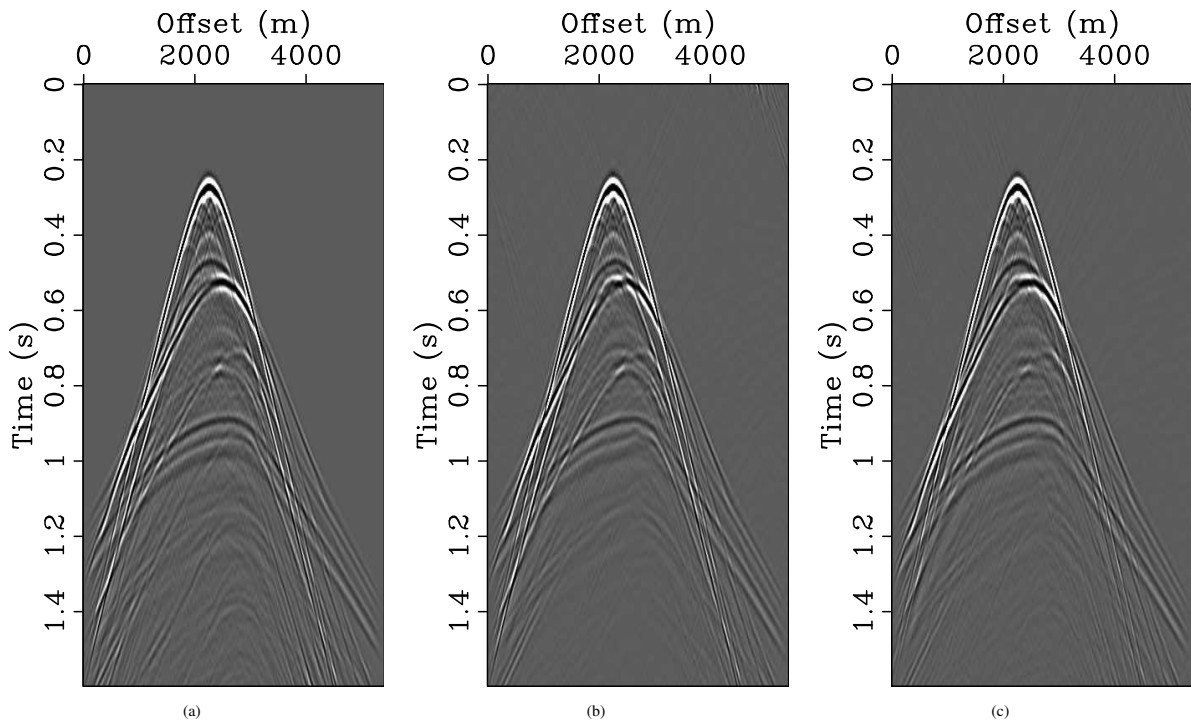


Figure 3: Curvelet-based primary estimation. (a) SRME-predicted primaries (b) Estimated primaries by a single thresholding as in Herrmann et al. (2007). (c) Estimated primaries after 10 iterations of our separation algorithm (equation 8) with $\lambda_1 = 0.8$, $\lambda_2 = 1.2$ and $\eta = 1.2$. Comparing the SRME predicted primaries with the curvelet-based estimates show an improved removal of late arriving multiple energy. In addition, the thresholded result gives rise to a large remaining multiple around 0.6s and an offset of 2400m which is no longer visible in the result obtained with our iterative method. Also the artifacts in this region are less severe.

Curvelet-based primary multiple separation

REFERENCES

- Berkhout, A. J. and D. J. Verschuur, 1997, Estimation of multiple scattering by iterative inversion, part I: theoretical considerations: *Geophysics*, **62**, 1586–1595.
- Candes, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: *SIAM Multiscale Model. Simul.*, **5**, 861–899.
- Daubechies, I., 1992, *Ten lectures on wavelets*: SIAM.
- Daubechies, I., M. Defrise, and C. De Mol, 2003, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.
- Hennenfent, G. and F. J. Herrmann, 2006, Seismic denoising with non-uniformly sampled curvelets: *IEEE Comp. in Sci. and Eng.*, **8**, 16–25.
- Herrmann, F. J., U. Boeniger, and D.-J. E. Verschuur, 2007, Nonlinear primary-multiple separation with directional curvelet frames: *Geoph. J. Int.* To appear.
- Herrmann, F. J. and D. J. Verschuur, 2005, Robust curvelet-domain primary-multiple separation with sparseness constraints: Presented at the EAGE 67th Conference & Exhibition Proceedings.
- Li, Y., A. Cichocki, and S. ichi Amari, 2004, Analysis of sparse representation and blind source separation: *Neural Comput.*, **16**, 1193–1234.
- Sacchi, M. and T. Ulrych, 1996, Estimation of the discrete fourier transform, a linear inversion approach: *Geophysics*, **61**, 1128–1136.
- Ying, L., L. Demanet, and E. J. Candés, 2005, 3D discrete curvelet transform: *Wavelets XI*, Expanded Abstracts, 591413, SPIE.
- Zibulevsky, M. and B. A. Pearlmutter, 2001, Blind source separation by sparse decomposition in a signal dictionary: *Neural Computation*, **13**, 863–882.