

## Robust source signature deconvolution and the estimation of primaries by sparse inversion

Tim T.Y. Lin and Felix J. Herrmann, Dept. of Earth and Ocean Sciences, University of British Columbia

### INTRODUCTION

The past few years had seen some concentrated interest on a particular wavefield-inversion approach to the popular SRME multiple removal technique called Estimation of Primaries by Sparse Inversion (EPSI). EPSI promises greatly improved tolerance to noise, missing data, edge effect, and other physical phenomenon generally not described by the SRME relation (van Groenestijn and Verschuur, 2009a,b). It is based on the premise that it is possible to stably invert for both the primary impulse response and the source signature despite beforehand having no (or very limited) explicit knowledge of latter. The key to successful applications of EPSI, as shown in very recent works (Savels et al., 2010), is a robust way to reconstruct very sparse primary impulse response events as part of the inversion process. Based on the various successful demonstrations in literature, there is a very strong sense that EPSI will also play an important role in future developments of source signature deconvolution and the general recovering of wavefield spectrum.

### THE EPSI MODEL

EPSI describes the total up-going wavefield  $\mathbf{P}$  recorded at the surface as the sum of two terms. The first is the (surface-related) primary wavefield, described as the surface-free primary impulse response  $\mathbf{G}$  time-convolved with the source signature  $\mathbf{Q}$  (expressed in Berkhouit's detail-hiding monochromatic notation). The second term describes *all* subsequent multiples caused by the free surface, described as the multi-dimensional convolution of  $\mathbf{G}$  and  $\mathbf{P}$  via the SRME principle that primary reflections at the surface act as Huygen sources of the multiple wavefields, assuming that the free-surface is a perfect reflector. Even though the (monochromatic) EPSI modelling operator  $\mathcal{M}(\mathbf{G}, \mathbf{Q})$ , defined as

$$\mathbf{P} = \mathcal{M}(\mathbf{G}, \mathbf{Q}) := \mathbf{G}\mathbf{Q} + \mathbf{G}\mathbf{P}, \quad (1)$$

is nonlinear in the joint domain of the variables  $\mathbf{G}$  and  $\mathbf{Q}$  due to the term  $\mathbf{G}\mathbf{Q}$ , it is absolutely linear with respect to just  $\mathbf{G}$  and just  $\mathbf{Q}$ . This leads to the nice property that the block-coordinate descent method, also called the non-linear Gauss-Sidel method, to solve for  $\mathbf{G}$  and  $\mathbf{Q}$  exactly coincides with inverting  $\mathcal{M}(\mathbf{G}, \mathbf{Q})$  by successive linearizations around iteratively refined estimates  $\tilde{\mathbf{G}}$  and  $\tilde{\mathbf{Q}}$ . Thus the machinery required to perform EPSI reduces to iteratively solving two simple linear problems: a least-squares problem for  $\tilde{\mathbf{Q}}$ , and a sparsity-regularized solution of linear systems for  $\tilde{\mathbf{G}}$ . Both of these are well-studied problems and there exists many ways to solve them.

The development of EPSI lies at an interesting confluence of several advances in a multitude of areas. On one hand, the success of SRME has given us a fairly well-behaved and well-studied model of the relationship between multiple wavefields

and the primary impulse response. Further developmental attempt to overcome practical limitations of SRME as well as motivation from shallow water-bottom multiple prompted the consideration for an inversion form. The contribution of EPSI was realizing that the sparsity constraint should be used when inverting for these primary reflections, and this will in fact recover both the primary impulse response and a wavelet that describes global source properties. The key to this approach lies in robustly finding sparse solutions to the primary impulse-response, which basically translates to the ability to find the sparsest solutions of large linear systems. This was originally proposed to be solved using very sparse gradients (van Groenestijn and Verschuur, 2009a). However, as reported in Lin and Herrmann (2010) a much more robust approach is possible via a convexification from the sparsity minimization to a minimization of the 1-norm ( $\ell_1$ ) of the solution, which typically coincides with the sparsest solution (Donoho, 2006). This approach is convex, converges nicely, robust under noise, and is well-studied due to its critical role in the field of compressive sensing (Candès et al., 2006). Large-scale algorithms for this problem is a hot area of research interest and we are already utilizing some promising solvers. Examples include iterative-reweighted least-squares (IRLS), as well as such the Spectral Projected Gradient for  $\ell_1$ -minimization (SPG $\ell_1$ , see van den Berg and Friedlander (2008)) which we used to produce the examples in this work.

### A MORE PHYSICAL APPROACH TO SOURCE SIGNATURE DECONVOLUTION

Note that if we eliminate the  $\mathbf{GP}$  term in eq 1, then the problem of inverting  $\mathcal{M}(\mathbf{G}, \mathbf{Q})$  reduces immediately to blind spiky deconvolution (ignoring the multiples). If we for a moment forget about the goal of multiple removal and set out instead to design a spiky deconvolution algorithm, we can arrive at EPSI by making the following statement: the spiky events as a result of the deconvolution should also be able to predict any surface-related multiples in the recorded data via the SRME relation. Exploiting the free-surface, we have in fact introduced very useful physical constraints into what was traditionally regarded as a blind deconvolution problem. By complicating the problem, we have actually made it more amiable to nice solutions.

An example of the information provided by the multiple is as follows: If we take the multidimensional autocorrelation of the recorded up-going wavefield  $\mathbf{PP}^H$  (analogous to the match-filter approach of inverting  $\mathcal{M}(\mathbf{G}, \mathbf{Q})$  with  $\mathbf{Q} = 0$ ) and picked out the extremum point of the first reflected event while setting everywhere else to zeros, we have a roughly kinematically correct, spiky estimate of the first event in  $\mathbf{G}$ . Denote this as  $\mathbf{G}_0$ . To leading order, the correct amplitude of this event can be obtained by locating the scaling parameter  $a$  that minimizes  $\|\mathbf{P} - a\mathbf{G}_0\mathbf{P}\|_2$ . At this point, without invoking any spiky de-

convolution techniques and higher-order statistics, we have already a leading-order idea of the location and amplitude of the first spiky event based on the physics alone. This will benefit the deconvolution greatly as we can quickly identify the correct neighbourhood of the amplitude and shape of the wavelet without guesswork. Throughout the inversion process, the **GP** term will continually provide information on the correct location and amplitude of the spiky primary reflections.

A further advantage of using EPSI for source deconvolution is that this algorithm is necessarily a synthesizing one - the primary impulse response need to be specifically reconstructed along with the source wavelet estimate in order to calculate how well it predicts multiples with the multidimensional convolution. This leads to an especially stable way to accomplish spiky deconvolution. At no point in the EPSI algorithm was an inverse wavelet filter calculated, and any noise in the data (and other physical effect not related to reflections and thus not modelled by the **GP** term) will simply be rejected by the waveform inversion process into the residual. It should be evident that solving the EPSI problem is in fact also a stable, data-driven, physics-based way to deconvolve to the source signature.

## EXAMPLES ON SYNTHETIC AND REAL DATA

In Figure 1(b) we show the primary impulse response of the Pluto15 dataset obtained by solving an  $\ell_1$ -convexified version of EPSI. Although noise-free, this dataset was produced using elastic modelling code and thus contains P-S conversions and other effects outside of the range of  $\mathcal{M}(\mathbf{G}, \mathbf{Q})$ . We also directly used the pressure recording without up-down decomposition or deghosting. Nevertheless, the inversion process was stable and produced a wavefield with spiky appearance. The f-k spectra in Figure 2 also show a drastically increased frequency content, with the main limiting factor being the receiver ghost. As seen in Figure 3 we obtained a source wavelet (with source ghosts) that compared well with the direct waves. It is worth noting that the synthetic appearance of the estimated source was produced with minimum regularization; the least-squares solution was sought in the time domain over a short time window as the only constraint, and the wavelet was allowed to take any shape.

We can readily extend this method to account for continuities arising from fractional-order discontinuities, where impulse responses do not have a spiky appearance, by seeking  $\mathbf{G}$  in a transform domain that sparsely represents such signals. Figure 4(b) shows the primary impulse response of a Gulf of Suez dataset obtained by inverting for  $\mathbf{G}$  in a spline wavelet basis in the time direction, combined with a 2D-curvelet representation in the shot-receiver plane (Lin and Herrmann, 2011). In this case the  $\ell_1$ -convexified version of EPSI is required. No deghosting is performed on the data. As shown in Figure 5 we nicely recover a large portion of the frequency spectrum in addition to the effective multiple removal. We show the resulting estimated global source wavelet in Figure 6, where again no constraint is put on the shape besides a time-window that coincides with the abscissa range of the plot.

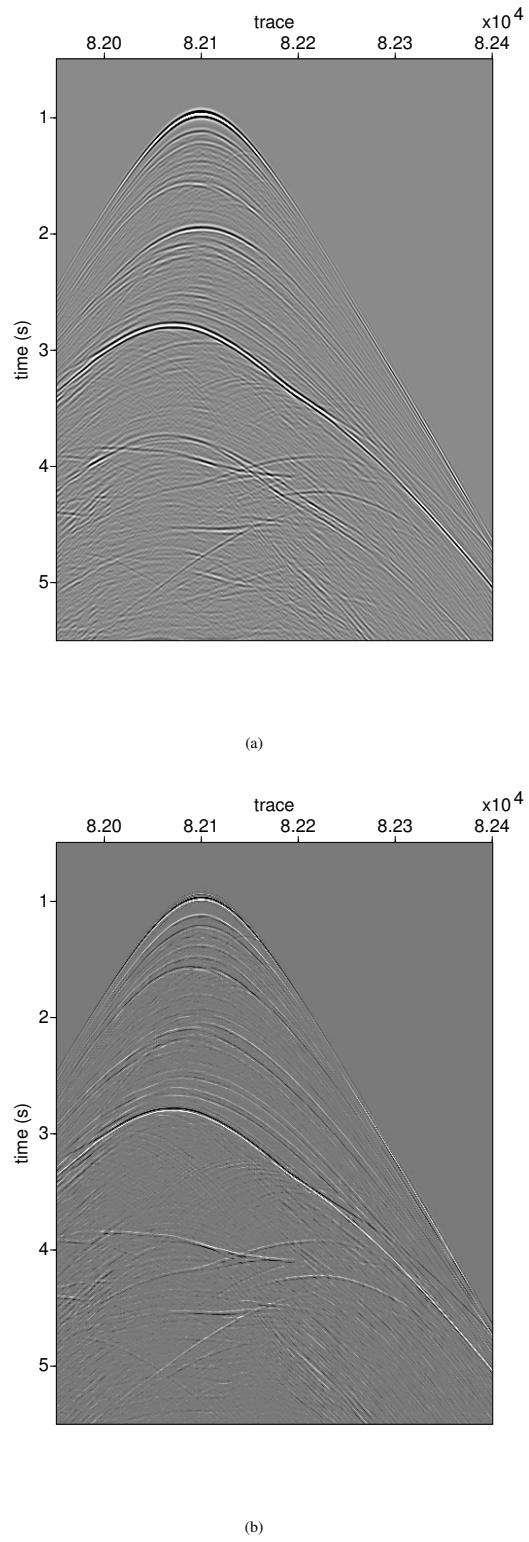


Figure 1: Shot-gather of Pluto15 data provided by SMAART JV. **(a)** Original data which was input into the  $\ell_1$ -convexified EPSI algorithm without deghosting or up-down decomposition. **(b)** Resulting primary impulse response of  $\ell_1$ -convexified EPSI algorithm after 90 gradient iterations.

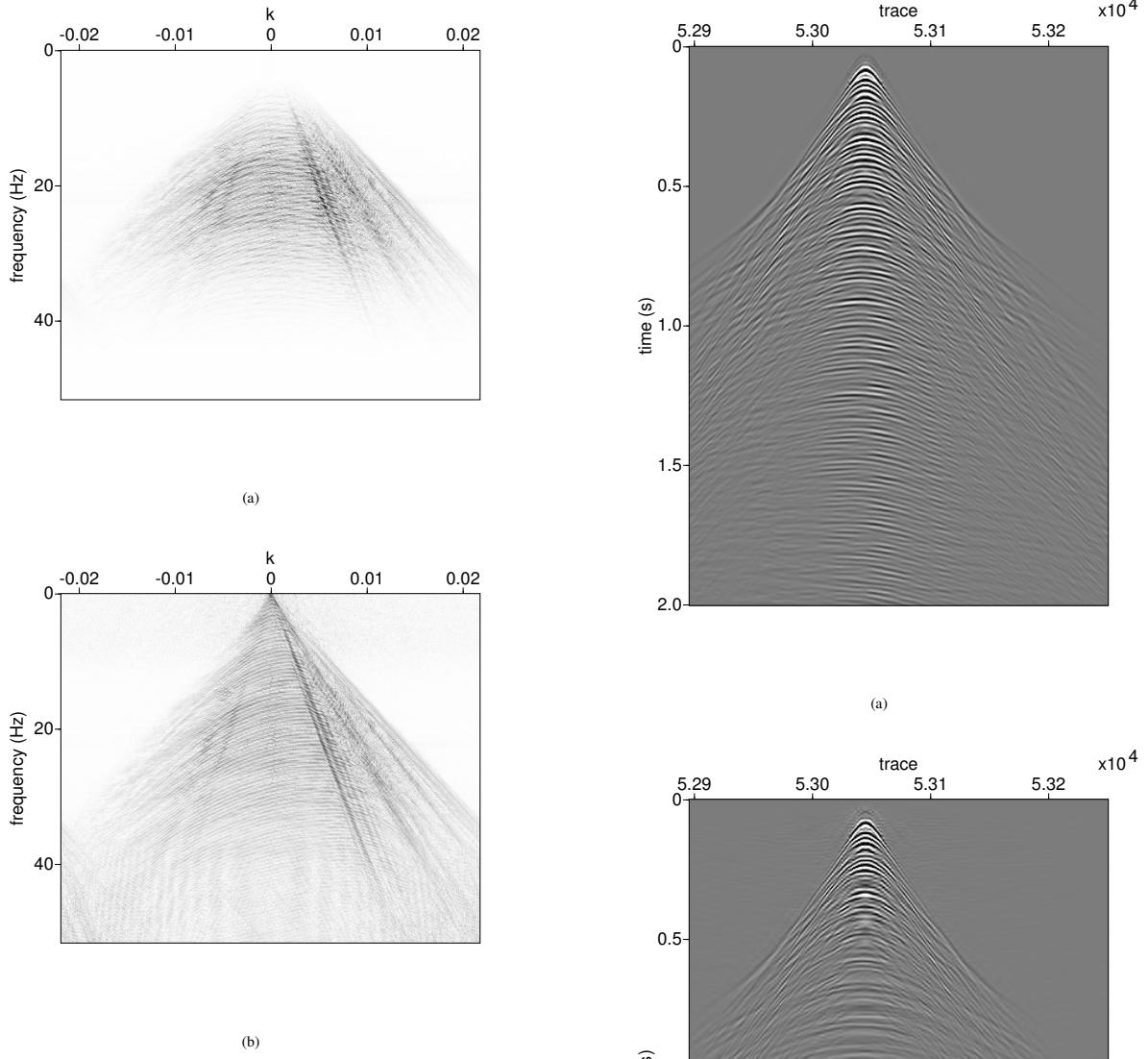


Figure 2: F-k spectrum of the shot-gathers of Pluto15 data shown in Figure 1. **(a)** Original data. **(b)** Result of  $\ell_1$ -convexified EPSI algorithm.

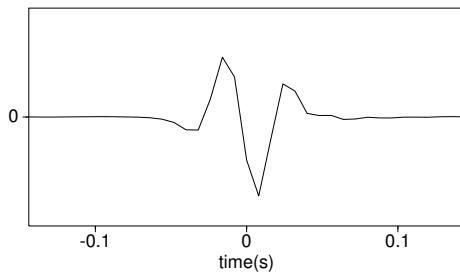


Figure 3: Final estimated source signature wavelet of the Pluto15 data shown in Figure 1 produced by the  $\ell_1$ -convexified EPSI algorithm.

Figure 4: Shot-gather of Gulf of Suez data provided by BP for academic purposes. Missing near-offsets were interpolated with Radon transform and negative offsets were produced by reciprocity. Shot spacing was 2x upsampled by interpolation. **(a)** Original data which was input into the  $\ell_1$ -convexified EPSI algorithm without deghosting or up-down decomposition. **(b)** Resulting primary impulse response of  $\ell_1$ -convexified EPSI algorithm after 100 gradient iterations.

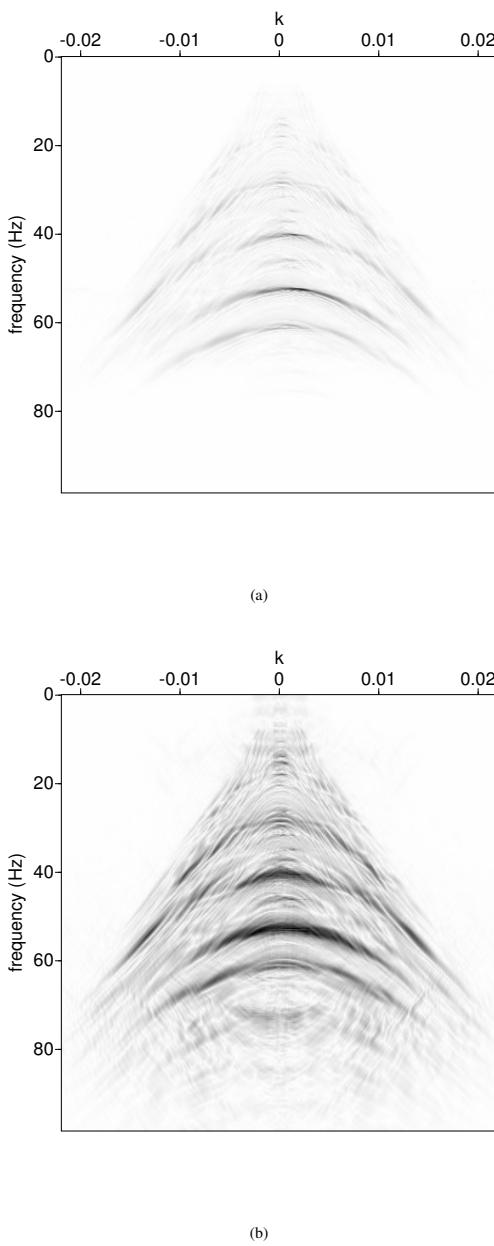


Figure 5: F-k spectrum of the shot-gathers of Gulf of Suez data shown in Figure 4. (a) Original data. (b) Result of  $\ell_1$ -convexified EPSI algorithm.

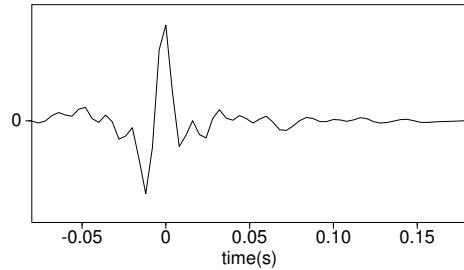


Figure 6: Final estimated source signature wavelet of the Gulf of Suez data shown in Figure 4 produced by the  $\ell_1$ -convexified EPSI algorithm.

## SUMMARY

Although originally motivated by the goal of improving multiple removal for shallow water-bottom regions under noise and missing near-offset data, the nature of EPSI makes it also very potent for the purpose of source signature deconvolution. In addition to estimating the source wavelet, this method also produces the deconvolved primary impulse response in a stable, synthesis-based fashion. Although EPSI is considered computationally expensive (typically involving 50 to 100 multidimensional wavefield convolution and correlations as gradient steps), it benefits directly from the effort to optimize SRME, being based on the same mathematical operation. Also, the algorithmic aspect of the large-scale sparse inversion nature of EPSI taps directly into the active community of researchers involved in compressive sensing and machine learning, many of whom are concerned with efficiently solving sparsity-regularized problems in industrial applications. It is also worth noting that the potential of EPSI as a valuable tool and the limiting effects of the receiver ghost on the physical model suggests that more serious efforts be put into up-down wavefield decompositions of marine data, to ensure that the EPSI modelling operator acts under the correct assumptions.

## ACKNOWLEDGEMENTS

We would like to thank Eric Verschuur and G.J.A. van Groenestijn for the genesis of our involvement in EPSI as well as valuable discussions. The Pluto15 dataset was provided by SMAART JV. The Gulf of Suez dataset originates from BP and was generously provided by Eric Verschuur. SPG $\ell_1$  is the work of Michael Friedlander and Ewout van den Berg. This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco.

## REFERENCES

- Candès, E. J., J. K. Romberg, and T. Tao, 2006, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Mathematics, **59**, 1207–1223.
- Donoho, D., 2006, For most large underdetermined systems of linear equations the minimal  $\ell(1)$ -norm solution is also the sparsest solution: Communications on Pure and Applied Mathematics, **59**, 797–829.
- Lin, T. T. Y., and F. J. Herrmann, 2010, Stabilized estimation of primaries via sparse inversion: Presented at the 72nd EAGE Conference Technical Program Expanded Abstracts, EAGE, EAGE.
- , 2011, Estimating primaries by sparse inversion in a curvelet-like representation domain: Presented at the 73rd EAGE Conference Technical Program Expanded Abstracts, EAGE, EAGE.
- Savels, T., K. de Vos, and W. de Maag, 2010, Surface multiple attenuation through sparse inversion: attenuation results for complex synthetics and real data: Presented at the 72nd EAGE Conference Technical Program Expanded Abstracts, EAGE, EAGE.
- van den Berg, E., and M. P. Friedlander, 2008, Probing the pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, **31**, 890–912.
- van Groenestijn, G. J. A., and D. J. Verschuur, 2009a, Estimating primaries by sparse inversion and application to near-offset data reconstruction: Geophysics, **74**, A23–A28.
- , 2009b, Estimation of primaries and near-offset reconstruction by sparse inversion: Marine data applications: Geophysics, **74**, R119–R128.