Introduction

Recent works on surface-related multiple removal include a direct estimation method closely related to Amundsen inversion (Amundsen, 2001) proposed by van Groenestijn and Verschuur (2009a), where under a sparsity assumption the primary impulse response is determined directly from a data-driven wavefield inversion process called Estimation of Primaries by Sparse Inversion (EPSI). The authors have shown in van Groenestijn and Verschuur (2009b) that for shallow bottom marine data this approach improves on traditional estimation-subtraction processes such as SRME, where by expanding the model to simultaneously invert for the near-offset traces – not directly available in most marine surveys but are observable in the surface-related multiples multiples – a large improvement over Radon interpolation is demonstrated.

One of the major difficulties in the practical adoption of EPSI is that one must have precise knowledge of a time-window that contains multiple-free primaries during each update. There is some anecdotal evidence that the inversion result is unstable under errors in the time-window length, a behaviour that runs contrary to the strengths of EPSI and diminishes its effectiveness for shallow-bottom marine data where multiples are closely spaced. Moreover, due to the nuances involved in regularizing the model impulse response in the inverse problem, the EPSI approach has an additional number of inversion parameters where it may be difficult to choose a reasonable value. We show that the specific sparsity constraint on the EPSI updates lead to an inherently intractable problem, and that the time-window and other inversion variables arise in the context of additional regularizations that attempts to drive towards a meaningful solution. We then suggest a way to remove almost all of these parameters via a $\ell 0$ to $\ell 1$ convexification, which stabilizes the inversion while preserving the crucial sparsity assumption in the primary impulse response model.

EPSI as an optimization problem

The EPSI modeling operator is formulated from an underlying principle that relates the primary impulse response to the total up-going wavefield that includes the source signature and surface-related multiples. Expressed mathematically, this is written as $\hat{\mathbf{P}} = \hat{\mathbf{G}}(\hat{\mathbf{Q}} + \mathbf{R}\hat{\mathbf{P}})$ where hatted quantities represent monochromatic representations of wavefields arranged into a matrix that have columns representing common shot gathers similar to the detail-hiding notation of Berkhout and Pao (1982), such that the matrix multiplication of two hatted wavefield quantities become non-stationary convolutions in the time domain. $\mathbf{G}$ represents the primary impulse response, $\mathbf{P}$ the total up-going wavefield, and $\mathbf{Q}$ a (possibly shot-dependent) source signature function, which depending on the data may not exactly be the actual seismic source signature without additional regularization. $\mathbf{R}$ is the reflection coefficient at the surface that is approximated to $-\mathbf{I}$ for the rest of this text.

In order to be consistent with the usual notations of general optimization problems we introduce vectored wavefields in lower case, e.g. $\mathbf{p} = \mathrm{vec}(\mathbf{P})$. We can then express the above wavefield relation using a linear operator $\mathbf{A}[\mathbf{Q}]$ acting on vectorized primary impulse response $\mathbf{g}$:

$$\mathbf{A}[\mathbf{Q}]\mathbf{g} := \mathcal{F}_t^* \mathrm{BlockDiag}_f[(\hat{\mathbf{Q}} - \hat{\mathbf{P}}^-)^* \otimes \mathbf{I}]\mathcal{F}_t \mathbf{g} = \mathbf{p}, \tag{1}$$

where the block diagonal elements varies over frequency, and $\otimes$ defines a Kronecker product that reformulates matrix multiplication into matrix-vector products. $\mathcal{F}_t$ is a Fourier transform in the time axis that also lays out the different frequency "slices" of the wavefield in a vectorized manner, such that $\mathcal{F}_t \mathbf{g} = \hat{\mathbf{g}} := [\hat{\mathbf{g}}_{f1}, \hat{\mathbf{g}}_{f2}, ..., \hat{\mathbf{g}}_{fn}]^T$, while the adjoint operation $\mathcal{F}_t^*$ on the left brings the wavefield back to the time domain. Note that $\mathbf{A}$ is a simple linear operator that depends both on a source signature estimate $\mathbf{Q}$ and the (recorded) up going wavefield $\mathbf{P}$, but since

EPSI gives a description of how a reasonable estimate of the primary impulse response can be obtained by a steepest-descent inversion process. The gradients of the objective function $f(\mathbf{g}) = \|\mathbf{p} - \mathbf{A}\mathbf{g}\|_2^2$, is evaluated at $\tilde{\mathbf{g}}$ (an estimate on $\mathbf{g}$) according to

$$\nabla f|_{\tilde{\mathbf{g}}} = 2\left(\mathbf{p} - \mathbf{A}\tilde{\mathbf{g}}\right)\mathbf{A}^*. \tag{2}$$

Since the assumptions of EPSI calls for the solution to be regularized by its sparsity, a sparse update $\delta\tilde{\mathbf{g}}$ on $\tilde{\mathbf{g}}$ is then obtained by picking the τ-th largest elements of the gradient and setting the rest to zero, followed by a scaling factor determined by a simple line search. The next update will then be calculated on the gradient of $(\tilde{\mathbf{g}} + \delta\tilde{\mathbf{g}})$. The process is repeatedly carried out until a desired image of the primary is formed. Its previously reported (van Groenestijn and Verschuur, 2009a) that for the case of synthetic marine data a reasonable estimate can typically be obtained on the order of 100 such steps.

Since steepest decent methods can be generalized as a gradient method in optimization, it is beneficial to look at the EPSI problem in light of the notations inherent of that field. If we make the simplifying assumption that $\mathbf{Q}$ is known, then the goal of EPSI is to solve an instance of the following *non-convex* and NP-hard optimization problem:

$$\tilde{\mathbf{g}} = \underset{\mathbf{g}}{\operatorname{argmin}} \|\mathbf{p} - \mathbf{A}\mathbf{g}\|_2 \quad \text{subject to} \quad \|\mathbf{g}\|_0 \leq k\tau, \tag{3}$$

where k is the number of iterations taken in the EPSI process. The $\ell 0$ pseudonorm $\|\mathbf{g}\|_0$ measures the *cardinality* – the number of non-zero elements – in $\mathbf{g}$.

In general, finding the solution of cardinality-based optimization problems is notoriously difficult with a myriad of convergence and stability issues. As such, the constraint $\|\mathbf{g}\|_0 < k\tau$ limiting the number of non-zeros in the solution makes the problem in Eq. 3 hopeless for simple gradient approaches, and furthermore implies that EPSI is not theoretically guaranteed to converge to a global solution. A glimpse of this fragility is provided by van Groenestijn and Verschuur (2009a): if the cardinality constraint for fixed k is instead naively imposed on the entire current estimate $\tilde{\mathbf{x}}$ at each step, then ESPI ceases to converge in a reasonable amount of time. A workaround is to severely limit the size of the feasible set at every iteration, such as imposing constraint $\|\delta\tilde{\mathbf{x}}\|_0 = \tau$, so that the update making the largest possible progress to the minimization objective in Eq. 3 can be found by simple searching over the whole set. It is easy to show that this is done by zeroing everything in the gradient except the τ-th largest elements, leading exactly to the updates used in EPSI. Note that final solution by definition will remain feasible for the original constraint $\|\mathbf{g}\|_0 < k\tau$. One can also see that the time-window normally used on each update is similarly imposed to prevent the inversion from being trapped in a multiples-included local minima solution that is hard to escape with steepest descent updates.

Stability via convexification

Instead of using various regularizations on the update to stabilize the ill-posed EPSI problem, one can postulate a similar but much more tractable problem by performing a convexification. A very well-known strategy when tackling cardinality-constrained problems is to replace the $\ell 0$ term with the $\ell 1$ norm $\|\mathbf{g}\|_1$, the sum over the element-wise absolute value of $\mathbf{g}$, leading to

$$\tilde{\mathbf{g}} = \underset{\mathbf{g}}{\operatorname{argmin}} \|\mathbf{p} - \mathbf{A}\mathbf{g}\|_2 \quad \text{subject to} \quad \|\mathbf{g}\|_1 \leq \tau, \tag{4}$$

where the τ in this expression is overloaded to be an $\ell 1$ norm constraint on the model. This approach has been both theoretically and experimentally justified in the optimization community as being a very effective heuristic for sparsity regularization on the model, and play a significant role in current developments in large-scale machine learning techniques. Critically, eq. 4 is a convex problem, implicating that rapid convergence to the global solution is guaranteed using several highly efficient existing gradient-based methods such as Nesterov updates (Nesterov, 2005) or Spectral Gradients (Birgin et al., 1999). A convex problem in general is also stable with no local minima in the objective, which strongly suggests that the time-windowing performed on EPSI updates can be removed with no impact on the final solution.

The convexification can also deal with the issue of determining a good value for the sparsity restriction τ, which is the maximum allowed cardinality (or in the case of eq. 4 the L1 norm) of the model. Overestimating the value of τ leads to a weak sparsity regularization of the primary impulse response, while underestimation leads to artifacts from an incomplete representation of the primaries. Exploiting a well known duality result, one can (non-trivially) associate a certain value of τ in eq. 4 with a certain value of σ in the following problem:

$$\tilde{\mathbf{g}} = \underset{\mathbf{g}}{\operatorname{argmin}} \|\mathbf{g}\|_1 \quad \text{subject to} \quad \|\mathbf{p} - \mathbf{A}\mathbf{g}\|_2 \leq \sigma, \tag{5}$$

such that both problems lead to the same solution. Here σ is seen as the residual between the recorded data and the total up-going wavefield predicted by the estimated primaries, which is highly linked to the noise level of the shot record. Due to its physical significance, a good estimate for σ should be more easily determined compared to τ. Eq. 5 can be solved quite efficiently using methods such as SPGℓ_1, which is a gradient-based method using spectral step-lengths and pareto root-finding described in van den Berg and Friedlander (2008). Typically we find the computation cost of eq. 5 to be similar to traditional EPSI, using a comparable number of gradient calculations usually quoted for EPSI, while having neither any time-window for primary only regions nor any knowledge of the sparsity level of the primaries.

The sparsfying transform: Inspired by recent efforts to exploit Compressive Sensing (CS) principles in seismic processing, one can furthermore reinforce the sparsity assumption of primary impulse response by using a suitable sparsity frame (Herrmann et al., 2008, 2009). We accomplish this by defining the sparsity transform as the Kronecker product between the 2-D discrete curvelet transform (Candès et al., 2006) along the source-receiver coordinates, and the discrete wavelet transform along the time coordinate—i.e., $\mathbf{S} := \mathbf{C} \otimes \mathbf{W}$ with $\mathbf{C}$, $\mathbf{W}$ the curvelet- and wavelet-transform matrices, respectively. The associated synthesis operator $\mathbf{S}^*$ is then inserted between $\mathbf{A}[\mathbf{Q}]$ and the variable $\mathbf{g}$, and instead of obtaining $\tilde{\mathbf{g}}$ as a solution we obtain the analysis coefficients $\mathbf{S}\tilde{\mathbf{g}}$

Informed blind deconvolution of source signature: We can now relax the assumption of a known source signature such that $\mathbf{A}$ becomes a function of $\mathbf{Q}$. This yield the final form of the Stabilized EPSI problem:

$$(\mathbf{S}\tilde{\mathbf{g}}, \tilde{\mathbf{Q}}) = \underset{\mathbf{g},\mathbf{Q}}{\operatorname{argmin}} \|\mathbf{g}\|_1 \quad \text{subject to} \quad \|\mathbf{p} - \mathbf{A}(\mathbf{Q})\mathbf{S}^*\mathbf{g}\|_2 \leq \sigma, \tag{6}$$

which is a *biconvex* problem – fixing either of the variables $\mathbf{g}$ or $\mathbf{Q}$ leads to a purely convex problem in the other variable. Note that fixing $\mathbf{g}$ leads to a simple least-squres problem in $\mathbf{Q}$, which we will name $\mathrm{OP_q}$. Fixing $\mathbf{Q}$ will instead lead directly to eq. 5, which will now be named $\mathrm{OP_g}$.

Algorithm 1: Stabilized Estimation of Primaries by Sparse Inversion

Result: Estimate for $\mathbf{Sg}$ and $\mathbf{Q}$
choose noise level σ, wavelet window length, iteration increment ρ
$\mathbf{g}_0 \longleftarrow 0$, $\mathbf{Q}_0 \longleftarrow 0$, $\mathbf{Q}_{\Lambda 0} \longleftarrow 0$, $k \longleftarrow 1$
while $\|\mathbf{p} - \mathbf{Ag}\|_2 \geq \sigma$ **do**
  $(\mathbf{g}_k, \tau_k) \leftarrow$ solve $\mathrm{OP_g}$ using initial guess $\mathbf{g}_{k-1}, \tau_{k-1}$, $\mathbf{Q}_{\Lambda k-1}$ with SPGℓ_1 at least ρk iterations
  $\mathbf{Q}_k \leftarrow$ solve $\mathrm{OP_q}$ with LSQR
  $\mathbf{Q}_{\Lambda k} \leftarrow$ window $\mathbf{s}_k$ around $t = 0$
  if $||\mathbf{Q}_{\Lambda k} - \mathbf{Q}_k||_2 < \epsilon$ **then**
    break
  end
  $k \leftarrow k + 1$
end
$\tilde{\mathbf{Q}} \leftarrow \mathbf{Q}_{\Lambda k}$
$\mathbf{S}\tilde{\mathbf{g}} \leftarrow$ solve $\mathrm{OP_g}$ using initial guess $\mathbf{g}_k, \tau_k$, $\tilde{\mathbf{Q}}$ with SPGl1 upto σ

References

Amundsen, L. [2001] Elimination of free-surface related multiples without need of the source wavelet. *Geophysics*, **66**(1), 327–341, doi:10.1190/1.1444912.

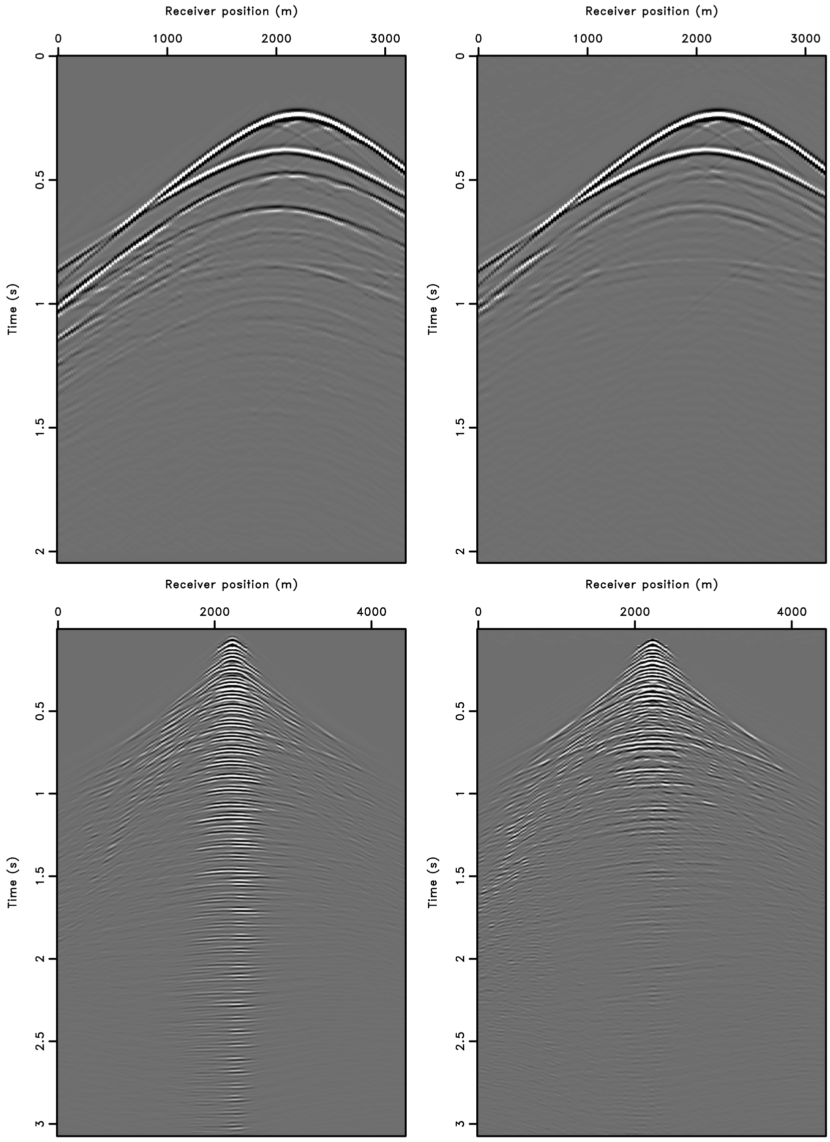

Figure 1 *Examples of primary estimation with stabilized EPSI according to Algorithm 1.* ***(top left)*** *A single shot-record of synthetic marine data* ***(top right)*** *Estimated primaries form synthetic data.* ***(bottom left)*** *A single shot-record of processed shallow-bottom marine data from Gulf of Suez.* ***(bottom right)*** *Estimated primaries form real data.*

Berkhout, A.J. and Pao, Y.H. [1982] Seismic migration—imaging of acoustic energy by wave field extrapolation. *Journal of Applied Mechanics*, **49**(3), 682–683, doi:10.1115/1.3162563.

Birgin, E.G., Martínez, J.M. and Raydan, M. [1999] Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. on Optimization*, **10**, 1196–1211, ISSN 1052-6234, doi:10.1137/S1052623497330963.

Candès, E.J., Demanet, L., Donoho, D.L. and Ying, L. [2006] Fast discrete curvelet transforms. *Multiscale Modeling and Simulation*, **5**, 861–899, doi:10.1137/05064182X.

Herrmann, F.J., Moghaddam, P.P. and Stolk, C.C. [2008] Sparsity- and continuity-promoting seismic image recovery with curvelet frames. *Applied and Computational Harmonic Analysis*, **24**(2), 150–173, doi:10.1016/j.acha.2007.06.007.

Herrmann, F.J., Erlangga, Y.A. and Lin, T.T.Y. [2009] Compressive simultaneous full-waveform simulation. *Geophysics*, **74**(4), A35–A40, doi:10.1190/1.3115122.

Nesterov, Y. [2005] Smooth minimization of non-smooth functions. *Math. Program.*, **103**(1), 127–152, ISSN 0025-5610, doi: http://dx.doi.org/10.1007/s10107-004-0552-5.

van den Berg, E. and Friedlander, M.P. [2008] Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, **31**(2), 890–912, doi:10.1137/080714488.

van Groenestijn, G.J.A. and Verschuur, D.J. [2009a] Estimating primaries by sparse inversion and application to near-offset data reconstruction. *Geophysics*, **74**(3), A23–A28, doi:10.1190/1.3111115.

van Groenestijn, G.J.A. and Verschuur, D.J. [2009b] Estimation of primaries and near-offset reconstruction by sparse inversion: Marine data applications. *Geophysics*, **74**(6), R119–R128, doi:10.1190/1.3213532.