Modeling data error during deconvolution

Jon Claerbout and Antoine Guitton

ABSTRACT

Our current decons take the data sacrosanct and find the best noncausal wavelet to deconvolve it with. We propose allowing the data to include an explicit noise that does not fit the convolutional model. We write regressions to define this noise, and develop an expression for the gradient needed to fit the regressions.

INTRODUCTION

Data fitting with the ℓ_1 norm has well-known remarkable qualities. Even more suitable to seismic problems is the hyperbolic penalty function $h(r) = \sqrt{1+r^2} - 1$ (Li et al. (2010)). Applying it to deconvolution of 2-D synthetic data easily solved a complicated case not solveable by ℓ_2 decons (Zhang and Claerbout (2010a)). Unfortunately, application to field data was not successful. Returning to synthetic data, the culprit turned out to be the minimum-phase assumption. This can be understood by examining the Ricker (non-minimum phase) wavelet, a long-standing example of a wavelet hard to spike.

Solving for a product of forward and backward PEFs achieved an excellent result on synthetic data and a spectacular result on a Gulf of Mexico data set (Zhang and Claerbout (2010b)). By deconvolving properly we were delighted to find we had made reflection coefficient polarity much more clear. We were highly motivated to improve on this and integrate it with impedance estimation. Not well documented were difficulties connected with polarity reversal and apparent time shifts. They were attributed to the non-linearity of the method.

Claerbout et al. (2011) reformulated the problem in the frequency domain with the unknown parameters being the values at lags defining the log spectrum. This avoided many problems, but extensive testing by Qiang Fu and Yi Shen revealed reliability issues much like those identified by Zhang. For a long while we understood our difficulty to be a need for preconditioning to guide the non-linear problem closer to the desired solution. Suddenly we came to realize the problem is more like a null space, though not exactly that because of the nonlinearity. The apparent polarity reversals and time shifts resulted from spiking the first or the third lobe of the Ricker wavelet instead of consistently spiking the middle lobe. Claerbout (2012) resolved these problems by means of a regularization (called the Ricker regularization) that ensures Ricker-like wavelets. Unfortunately, like all regularizations, you can never be sure how much to add, leading to degraded results when you add too much.

Then the non commutivity of gain with filtering was theoretically resolved leading to small but noticable improvements (Claerbout et al. (2012)). Guitton and Claerbout (2012) also added a regularization term that penalizes long positive or negative lags of the filter. This penalty ensures that the estimated wavelet does not shift in masking areas that can be present in the gain function (i.e., preventing local minima).

With stability now under much better control (we still need to experiment with strength of the regularizations) we set out to demonstrate that sparse decon principles could find natural cutoffs for high and low frequencies in data. We seemed to be seeing frequencies dangerously close to the 125Hz Nyquist on our available 4ms data so we ordered and waited to obtain 2ms data to boost the Nyquist to 250 Hz. Much to our horror (Guitton and Claerbout (2012)), sparseness decon, like old fashioned l_2 decon, boosts energy up to near the new much higher Nyquist. Also discovered in that paper is that our shot wavelets are picking up sea swell noise. We do not wish to filter out sea swell as a preprocess because we do not wish to lose low frequency information that could be essential to impedance estimation. Swell noise modeling has been done by Parrish (2005). Subtracting such models should work better than filtering.

The formulation of this paper integrates sea swell modeling with our non minimum-phase, sparseness goaled, shot waveform estimation and data deconvolution. The experimental results mentioned above led to the theory you find here. What else might we find? We expect the noise to contain any bits of the data with non-typical spectra, both amplitude and phase. Besides the low-frequency sea swell, we might find the water bottom itself and its multiples contain the very high frequencies that we do not expect in waves that penetrate the earth.

INTRODUCING NOISE AS ITSELF A MODEL

The idea of this paper is that we should not try push all our data into the convolutional model. We should explicitly solve for an unknown part of the data that poorly fits this model. I call this part noise and define it negatively $-N$ (so the minus sign is missing from all the analysis and code).

The decon filter $C = e^U$, parameterized by U, we take as noncausal. The constraint is no longer a spike at zero lag, but a filter whose log spectrum vanishes at zero lag, $0 = u_0 = \int \ln C(\omega)\, d\omega$, so we are now constraining the mean of the log spectrum. This is a fundamental change which we confess to being somewhat mysterious.

The single regression for U including noise N now becomes two.

$$0 \approx_h (D+N)e^U = (D+N)C \tag{1}$$

$$0 \approx_2 N \tag{2}$$

The notation $\approx_h$ means the data fitting is done under a hyperbolic penalty function. The regularization need not be ℓ_2. To save clutter I leave it as ℓ_2 until the last step when I remind how it can easily be made hyperbolic.

Under the constraint of a causal filter with $c_0 = 1$, traditional auto regression for $c_t = \mathrm{FT}^{-1} C$ with its regularization looks more like

$$0 \approx N = Dc \tag{3}$$

$$0 \approx c \tag{4}$$

Comparing equations 1-2 with 3-4 you see we are not simply rehashing traditional methodology but seem to be off in a wholly new direction! We are here because $C = e^U$ solved our non-minimum phase problem, and seeing sea swell in our estimated shot wavelets told us we need to replace D by $D + N$.

Antoine noticed the quasi-Newton method of data fitting requires gradients but not knowledge of how to update residuals $\Delta\mathbf{r}$ so the only thing we really need to think about is getting the gradient. The gradient wrt U is the same as before (Claerbout et al. (2011)) except that $D + N$ replaces D. The gradient wrt N is the new element here.

Let d, n, and c be time functions (data, noise, and filter). Let $r = (d-n)*c$ be the residual. Let $h_t = h(r_t) =$ hyperbolic stretch of r. Expressing our two regressions in the time domain we minimize

$$\min_n \quad \sum_t n^2/2 \; + \; h((d+n)*c) \tag{5}$$

A scaling factor is required between the terms. We expect to learn it by experimentation.

Now we go after the gradient, the derivative of the penalty function wrt each component of noise n_s. Let the derivative of the penalty function $h(r_t)$ wrt its argument r_t be called the softclip and be denoted $h'_t = h'(r_t)$. Let H' denote the FT of h'. Let $c'(t)$ be the time reverse of $c(t)$ while in Fourier space C' is the conjugate of C.

$$\Delta n_s = n_s + \frac{\partial}{\partial n_s} \sum_t h(r_t) \tag{6}$$

$$= n_s + \sum_t h'(r_t) \frac{\partial}{\partial n_s} (d+n)*c \tag{7}$$

$$= n_s + \sum_t h'_t \frac{\partial}{\partial n_s} \sum_\tau n_{t-\tau} c_\tau \tag{8}$$

$$= n_s + \sum_t h'_t \frac{\partial}{\partial n_s} \sum_j n_j c_{t-j} \tag{9}$$

$$= n_s + \sum_t h'_t \; c_{t-s} \tag{10}$$

$$= n_s + \sum_t h'_t \; c'_{s-t} \tag{11}$$

$$\Delta N = N + C'H' \tag{12}$$

For simplicity I set out with a quadratic penalty function on the noise, but it is easy to make it hyperbolic. Simply use softclip on n. Change $\Delta n = n + \cdots$ to $\Delta n = h'(n) + \cdots$.

Now having the gradient we should be ready to code.

ALGORITHM

Before altering the old algorithm we need to be careful about a couple things. We may need different gain functions for $(d+n)*c$ and for n. Sea swell is quite stationary in its physics, but the hyperbolic penalty function applies to the statistical perspective which is one where images are boosted in time from their physical form. We also need to be careful not to mix up $h(n)$ with $h(r = (d+n)*c)$. We will need to scale the regularization with the fitting by experimentation.

We could update the old algorithm (Claerbout et al. (2011)) with the new noise parts. Alternately, we could follow the suggestion of Antoine and switch to the quasi Newton method. In either case we'll need to introduce a scale factor (learned from practice) to choose how much of D ends out in N.

INTERNET HUMOR

Theory is when you know everything but nothing works.

Practice is when everything works, but you don't know why.

In our lab, theory and practice are combined. Nothing works and nobody knows why.

REFERENCES

Claerbout, J., 2012, Polarity preserving decon in "N log N" time: SEP-Report, **147**, 305–312.

Claerbout, J., Q. Fu, and Y. Shen, 2011, A log spectral approach to bidirectional deconvolution: SEP-Report, **143**, 297–300.

Claerbout, J., A. Guitton, and Q. Fu, 2012, Decon in the log domain with variable gain: SEP-Report, **147**, 313–322.

Guitton, A. and J. Claerbout, 2012, Six tests of sparse log decon: SEP-Report, **148**, 35–48.

Li, Y., Y. Zhang, and J. Claerbout, 2010, Geophysical applications of a novel and robust l1 solver: SEP-Report, **140**, 155–164.

Parrish, J. F., 2005, Streamer string waves and swell noise: SEG Technical Program Expanded Abstracts, **24**, 72–75.

Zhang, Y. and J. Claerbout, 2010a, Least-squares imaging and deconvolution using the hybrid norm conjugate-direction solver: SEP-Report, **140**, 129–142.

———, 2010b, A new bidirectional deconvolution method that overcomes the minimum phase assumption: SEP-Report, **142**, 93–104.