The Need for L_1 Norm Algorithms for Seismic Decomposition

by Jon F. Claerbout

In a paper "Robust Modeling with Erratic Data", (Geophysics, vol. 38, No. 5, p. 827) Francis Muir and I demonstrated a number of attractive features of the L_1 norm in seismic data analysis. Another possibility is more informative deconvolutions. I'll never forget my disappointment back in 1961 when I deconvolved my first seismic data (M.S. thesis page 53 and 54). I had expected the deconvolution to leave me with a train of spikes, each with physical meaning. What happens is that you get a spike at every point in time. The denser you sample the data, the more spikes. There is no way, within the L_2 mathematical formalism, to turn off all but two or three spikes per wavelength. By bandwidth considerations you know that an average of more than that can't be meaningful. One way to reduce the number of spikes is to reduce the data sampling rate, but thinking about bright spots, we sure don't want to go that route. Bright spots are produced by thin sands containing high pressure natural gas so that we expect strong pulses of opposite polarity closely separated in time.

An attractive alternative is the L_1 norm which more naturally fits data to the kind of models we have in our heads. Let's look at one of the simplest cases which arises. Ideally, we know the source waveform s_t . From it we can construct the usual convolution matrix

$$A = \begin{bmatrix} s_0 & & & & \\ s_1 & s_0 & & & \\ s_2 & s_1 & s_0 & & \\ s_3 & s_2 & s_1 & s_0 & \ddots \\ & s_3 & s_2 & & \\ & & s_3 & \ddots & \end{bmatrix}$$

Using the vector d to denote a data seismogram, a vector x to denote a model of reflection coefficients as a function of travel time, the vector e to denote additive noise (error) as a function of time, we have

$$A\,x = d + e \tag{1}$$

where we have neglected multiple reflections and made many other of the usual assumptions in reflection seismic data interpretation. The traditional L_2 norm approach is to minimize $e^T e$. Commonly, one will "add a small amount of white noise" by minimizing $e^T e + \lambda\, x^T x$ where λ is a small number. We can take the same approach but minimize the sum of absolute values (L_1) instead of sums of squares, i.e.,

$$\min_x \sum_t |e_t| + \lambda\,|x_t| = \min_x \sum_t \left| \sum_s A_{ts}\, x_s - d_t \right| + \lambda\,|x_t| \tag{2}$$

As an aside, I might mention that both the L_2 and L_1 frameworks allow much more general models, for example by (1) weighting e_t and x_t by some time function to account for signal amplitude variations, and (2) allowing for frequency dependent attenuation by a more general A matrix. Neglecting these generalizations here, let us think about (1) and (2) as they stand. Why do I assert that the L_1 approach "fits the kind of models which we have in our heads"? It is because seismic data interpreters frequently think of earth models which consist of relatively homogeneous regions, shale, sand, salt, limestone, or pressurized gas sands, with reflections coming from the interfaces. Obviously sand/shale mixtures and marls (shale/lime mixtures) occur, but well logs and bright spot studies clearly show that the sharp interface situations also occur and are important. The problem with the L_2 analysis is that the presence of any noise at all causes rapid degradation of interface information (because squared deviation in x is minimized) and L_2 quickly introduces fuzz inside homogeneous regions (again because squared deviation in x is minimized). Theoretically from the Robust Modeling paper we know that the L_1 situation is very different. With L_1 the answer x_t will be identically zero at many time points. The bigger λ, the more time points over which x_t will come out zero. Presumably λ should be chosen to keep an average of all but two or so points per wavelength equal zero. Notice that there is no reason why very small separations between "turned on" time points in x_t cannot occur. It is only the total number of turned on time points which is limited by bandwidth considerations. Hopefully this is just exactly the kind of analytical tool which is required to make high resolution (less than a half wavelength) measurements in bright spot gas resevoir volume calculations where the high resolution theoretically permits us to measure the thickness of the gas sand even though it may be thinner than the usual resolution limits.

Before we can determine whether such high resolution can be achieved in practice we need some synthetic studies. But before we can do them we need a sensible algorithm for the minimization of (2). Some very economical algorithms exist for some L_1 problems but, to my knowledge, not to (2). I spent a couple of months trying to think up sensible algorithms for the minimization of (2) but there was no result worthy of reporting. None-the-less, it seems that more effort in this area might be well rewarded. Even a "brute force" approach might be worthwhile.