THE PURIFICATION OF BINARY MIXTURES

Jon F. Claerbout

We solve problems all the time without being able to explain how we solved them. Who can explain how to ride a bicycle, for example? How are voices in a noisy crowd separated? Why is it that we listen to only one voice when two equally strong voices are present? Is the separation made at a cognitive or a purely physical level?

Why does the best of industrial multiple suppression processes often leave a residual of multiples that the first year student easily recognizes? ***Information theory*** must have something to do with it. Unfortunately, much of the academic discipline of information theory is based on the assumption of stationary Gaussian processes, which would be highly applicable to the study of a universe that had undergone "thermal grey death," but can be very misleading in an analysis of real life. Information theory tells us that independent variables must be uncorrelated, but everyday life says the opposite. The orthogonalization of two observation vectors based on the sample correlation will more easily mix independent variables than decompose them into "independent" causes.

Sand and Gravel

Take a volume V_1 of gravel, and add a volume V_2 of sand. The total volume is now less than $V_1 + V_2$ because some sand grains fit in the holes between the gravel pebbles. Whether mixing produces a volume change is an interesting thermodynamic problem. Water and oil are immiscible, but oil and gasoline are miscible. (***Miscible*** apparently means mixable in some language other than English.) In a technical sense the word ***mixture*** may be reserved for the case when mixing results in a volume change.

We are very far from equilibrium if we have most of the sand in one volume and most of the gravel in the other volume. Imagine that each grain and pebble were denoted by a number but that the numbers did not distinguish grains from pebbles. Suppose some mechanism enabled us to move numbers from one volume to the other and then measure the sum of the resulting volumes. This is all we need to purify the mixtures. Arbitrarily moving numbers from one volume to the other, we have only to reverse those moves that decrease total volume. It would also help a lot if the initial two volumes were somewhat removed from equilibrium. The process might work if the initial two volumes were homogeneous, but it might also be hard to get started. And we would not know which of the final volumes was sand and which was gravel.

Near Trace - Far Trace Problem

The sand-gravel problem nicely illustrates the advantages of starting far from equilibrium, but it is too vague to be directly interpretable as a seismic problem. Let us devise a seismological problem which is analogous to the sand-gravel problem but which has only a single adjustable parameter. Ultimately I am interested in solving a multiple reflection problem by adjusting a reflection coefficient to separate primary reflections from multiples. But the multiple problem carries with it so many peripheral problems that we will invent an analogous, simplified problem. Analogous to the gravel we will have an ideal far-end-of-cable seismogram, and analogous to sand we will have an ideal near-end-of-cable seismogram. Because of some poorly understood electrical cross-talk phenomena, our observations consist only of

$$x \;=\; \text{near} + \epsilon_1 \;\text{far} \qquad (1a)$$

$$y \;=\; \text{far} + \epsilon_2 \;\text{near} \qquad (1b)$$

where ϵ_1 and ϵ_2 are small unknown numbers. Notice that the discovery of some magic parameter ϵ will enable us to mix our observations x and y in the right proportion to get pure sand.

$$\text{Mixture}(\epsilon) = (\text{near} + \epsilon_1 \text{ far}) + \epsilon(\text{far} + \epsilon_2 \text{ near}) \qquad (2)$$

Clearly we want $\epsilon = -\epsilon_1$, but we have no way of knowing what ϵ_1 is. The procedure will be to try various values of ϵ and observe some macroscopic property like volume. Of course we are all well aware of the "grey-death" solution, which is to define ϵ to be the negative of the sample correlation coefficient $(x \cdot y)/(y \cdot y)$. But the grey-death solution arises when we minimize the energy of the mixture, and the trouble with energy is that is quite different from information: information is often carried by tiny amounts of energy.

A Measure of Homogeneity

The geometric inequality is closely related to information theory. This subject was developed with a deconvolution application in SEP-15, p. 104. The basic principle is that for N positive variables p_j we have the arithmetic average exceeding the geometric average:

$$\frac{1}{N} \sum_{j=1}^{N} p_j \ \ge \ \prod_{j=1}^{N} p_j^{1/N} \qquad (3)$$

If all the p_j are equal to one another, the two averages become the same. The ratio of the sum to the product always exceeds unity, so the natural logarithm of that ratio is always positive. Denote this by S.

$$S = \ln (\Sigma / \Pi)$$

$$S = \ln \frac{1}{N} \sum_{j=1}^{N} p_j - \frac{1}{N} \sum_{j=1}^{N} \ln p_j \ \ge \ 0 \qquad (4)$$

The most homogeneous situation occurs when all the p_j are the same.

Then $S = 0$. When the p_j become very different from one another S increases. If our sand-gravel mixture were divided into N equal masses, then we could measure the volume p_j of each mass. The minimum S value would be $S = 0$ when each mass was mixed in the same proportions as every other mass. The maximum S value would occur when some P_j contained all sand and others contained all gravel.

Defining Positive Variables

Before we can apply the geometric inequality measure of simplicity to seismograms, we need a way to convert oscillatory zero-mean variables to positive variables. As in the previously mentioned SEP-15 paper, three possible ways present themselves.

First, data can be sorted according to size. Then the difference between successive sizes is a positive variable that can be used in the geometric inequality. In SEP-15 this method is shown to be related to Shannon's definition of information in terms of probability.

We can also use the seismogram envelope function p_j in (4) - a method related to measuring information by counting bits (also as in SEP-15). The envelope approach may be implemented in either of two ways: (1) using the Hilbert transform, complex trace approach, or (2) using the local averaging approach.

Real world seismograms have envelope functions that exhibit quite a range of amplitudes, so the simplicity measure S is often very large. The near trace and far trace seismograms show a large burst in the envelope when the first arrivals come. Since the arrivals come at different times, crosstalk will tend to homogenize the envelopes, destroying simplicity. The low level noise before the first signal arrival contributes a lot to the simplicity of a seismogram as defined by equation (4). So even a small amount of crosstalk from the near trace onto the far trace can destroy much of the simplicity. Figure 1 depicts the analogy between the volume of sand-gravel mixtures and the simplicity of near and far trace seismogram mixtures.

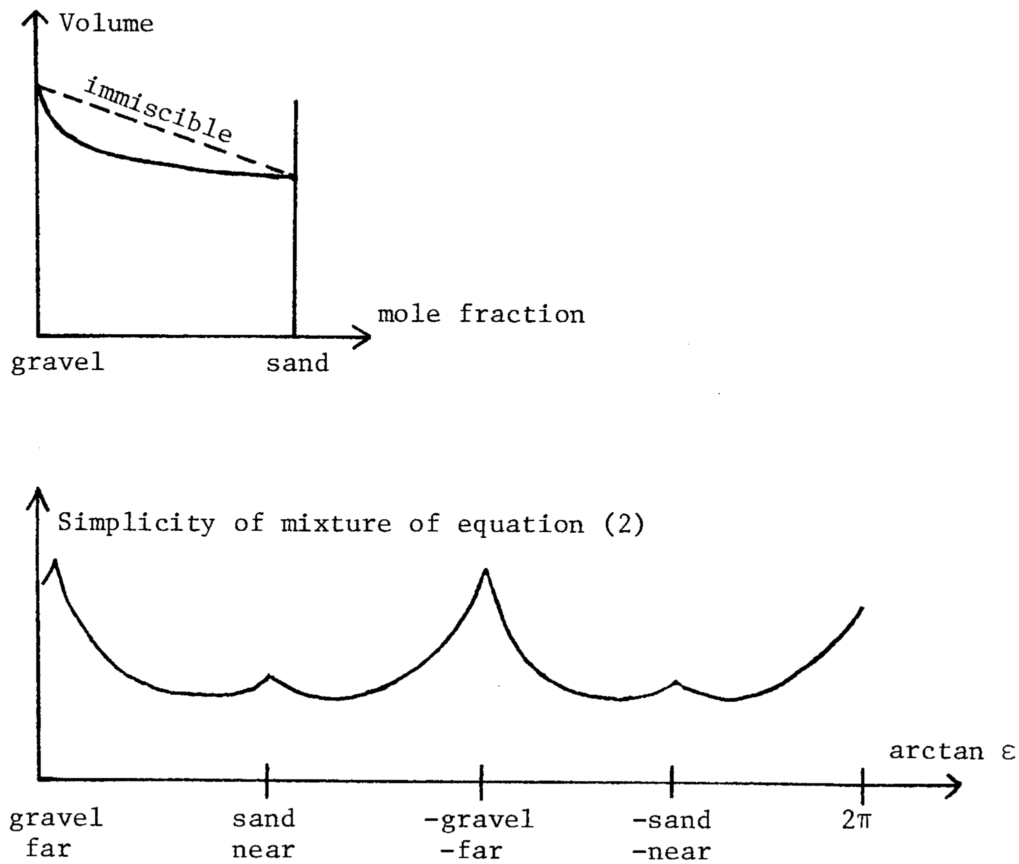

FIG. 1. Analogy of volume of sand-gravel mixture with simplicity of near trace-far trace mixture.

Purification

The cuspoid appearance of figure 1 may have some validity for the sand-gravel mixture but we can expect the seismogram mixture to be somewhat more continuous. This being the case we can hope that a Taylor series of the simplicity S about the origin $\epsilon = 0$ would enable us to go to maximum simplicity in one jump. My previous work on bitcount deconvolution suggested a method that proceeded to the extremum in a clumsy, iterative way. It was quite safe when the starting point was far from the maximum, but was slow to arrive at the final result. Now we will try for one jump.

Let us abbreviate (2) by

$$z_t = x_t + \epsilon y_t \tag{5}$$

For the envelope constructed with complex traces we have

$$p_t = (\bar{x}_t + \epsilon \bar{y}_t)(x_t + \epsilon y_t) \tag{6a}$$

At $\epsilon = 0$ we will need first and second derivatives of p_t with respect to ϵ. Denoting these by $\dot{p}_t$ and p_t we have

$$\dot{p}_t = \bar{x}_t y_t + \bar{y}_t x_t \tag{6b}$$

$$p_t = 2\, \bar{y}_t y_t \tag{6c}$$

The alternative form of the envelope is to average neighboring points. A convenient notation treats x_t as a matrix x_{ij} in exactly the same fashion as FORTRAN does. The product of the range of i multipled by the range of j equals the range of t. The subscript i ranges over neighboring points on the time series. Parallel to equation (6a), the envelope and its derivatives are given by

$$p_j = \sum_i (x_{ij} + \epsilon y_{ij})^2 \tag{7a}$$

$$\dot{p}_j = \sum_i 2\, x_{ij}\, y_{ij} \tag{7b}$$

$$p_j = \sum_i 2\, y_{ij}\, y_{ij} \tag{7c}$$

Now, recalling equation (4),

$$S = \ln \frac{1}{N} \sum_{j=1}^{N} p_j - \frac{1}{N} \sum_{j=1}^{N} \ln p_j \tag{8a}$$

The first derivative at $\epsilon = 0$ is

$$\frac{dS}{d\epsilon} = \frac{\Sigma\ \dot{p}_j}{\Sigma\ p_j} - \frac{1}{N}\Sigma\frac{\dot{p}_j}{p_j} \tag{8b}$$

The second derivative at $\epsilon = 0$ is

$$\frac{d^2S}{d\epsilon^2} = \left[\frac{\Sigma\ p_j}{\Sigma\ p_j} - \left(\frac{\Sigma\ \dot{p}_j}{\Sigma\ p_j}\right)^2\right] - \frac{1}{N}\Sigma\left[\frac{p_j}{p_j} - \frac{\dot{p}_j^2}{p_j^2}\right] \tag{8c}$$

By Newton's method, the value of ϵ for maximum S is

$$\epsilon = -\frac{\frac{dS}{d\epsilon}}{\frac{d^2S}{d\epsilon^2}} \tag{9}$$

Example: Separation of Near and Far Trace

Needs to be tried.

Example: Speed Convergence of Bitcount Debubble

Needs to be worked out.

Example: Improved Noah Demultiple

Needs to be tried.

Example: Separation of P Wave and S Wave Images

Good luck!