DECONVOLUTION ESTIMATORS FOR PARSIMONIOUS DECONVOLUTION

Jon F. Claerbout

In report 13 I give some working programs for parsimonious deconvolution along with the warning that stability was not yet understood. I should have added that lots of algorithms were tried and abandoned because of instability. Subsequent attempts to improve the algorithm are leading to better understanding of stability. The heart of the matter is how the data Y is deconvolved once a forward filter B is known. That is to say, given the model (in the frequency domain)

$$Y = X B + N \tag{1}$$

where X denotes the innovation series that we wish to estimate in the presence of noise N assuming that B is exactly known (although it isn't), how do we estimate X? An obvious estimator which we will term the *noise-free estimator* is

$$\hat{X} = Y / B \tag{E0}$$

This estimator is good when the filter is known (i.e., dB/B is zero) and the noise N is absent. The so-called *optimum* estimator $\hat{X}$ of X is obtained by minimizing the expected difference, i.e.,

$$\min = E\,\overline{(\hat{X} - X)}\,(\hat{X} - X) \tag{2}$$

To find this optimum estimator, we assume that it is a linear operator F on the observed data Y, say

$$\hat{X} = F\,Y \tag{3}$$

Substituting into (2) and setting to zero the derivative with respect to $\overline{F}$, we have

$$0 = E\,[\overline{Y}\,(FY - X)]$$

$$F = \frac{E(\overline{Y}X)}{E(\overline{Y}Y)} \tag{4}$$

Substituting the conjugate of (1) into the numerator and assuming the signal and noise are uncorrelated $E(\overline{N}\,X) = 0$, we get

$$F = \overline{B}\,\frac{E(\overline{X}X)}{E(\overline{Y}Y)}$$

using (3)

$$\hat{X} = \overline{B}\,\frac{E(\overline{X}X)}{E(\overline{Y}Y)}\,Y \tag{5}$$

Now if we are willing to assume that the unknown innovation series x_t has a white spectrum, then we can replace $E(\overline{X}X)$ by unity. The denominator expectation can be replaced by the appropriately smoothed observation spectrum $< \bar{Y}Y >$. Hence, we have the *optimum white signal estimator*

$$\hat{X} = \frac{\bar{B}}{< \bar{Y}Y >} Y \tag{E1}$$

As a practical matter it seemed to be expeditious to consider the estimator

$$\hat{X} = \frac{\bar{B}}{\varepsilon + < \bar{Y}Y >} Y \tag{E2}$$

To enable us to claim that this estimator is optimum, let us presume it was derived from the optimum estimator (5) by means of some assumption of the form of $E(\bar{X}X)$. By substitution, we see that the assumption

$$E(\bar{X}X) = \frac{E(\bar{Y}Y)}{\varepsilon + E(\bar{Y}Y)} \tag{6}$$

will do the job. The assumption (6) says that the spike series x_t is white for those frequencies at which $\bar{Y}Y >> \varepsilon$ and that the spectrum of the spike series drops off in proportion to the spectrum of the observed data y_t where $\bar{Y}Y << \varepsilon$. Thus, we may choose ε to characterize a signal-to-noise threshold. Where the spectrum of the observed data drops below this threshold, the input spike series is no longer considered to be white, it drops off with the observations. Thus we may call the E2 estimator the optimum *band limited white signal estimator*.

When you realize that B isn't absolutely known but is also being estimated by some external procedure, then you appreciate that the use of ε may allow B to avoid becoming infinite at frequencies for which $\overline{Y}Y$ tends to zero.

An appealing idea is to use the estimator (E1) to get $\hat{X}$. Then, find the spectrum of X and use it back in (5). This is a very inviting idea which can be applied again and again defining the *super gain estimator*

$$\hat{X}_{n+1} = \frac{\overline{B} \ \langle \overline{\hat{X}}_n \hat{X}_n \rangle}{\langle \overline{Y}Y \rangle} \ Y \tag{E3}$$

The *super gain estimator* provides a super trap for the unwary. Whether or not you fall into the trap depends on whether or not you do enough smoothing in the competition $\langle \overline{\hat{X}}_n \hat{X}_n \rangle$. If you don't do any smoothing, the final estimated spectrum will be a delta function at the frequency component which had largest value in the original estimate. How much smoothing is enough? That is a good question but not the subject of our present effort.

The last estimator to consider is the one which emerged in SEP 13 from the frantic search for a stable parsimonious deconvolution algorithm. This estimator which I will call the *lucky estimator* was

$$\hat{X} = \frac{\overline{B}}{\varepsilon + \overline{B}B} \cdot Y \tag{E4}$$

Let us suppose that this is an optimal estimator in the sense of equation (5). Then we must have

$$\frac{E\ (\overline{X}X)}{E\ (\overline{Y}Y)} = \frac{1}{\varepsilon + \overline{B}B} \tag{7}$$

inverting and expanding $E\ (\overline{Y}Y)$ we get

$$\frac{BB\ E(\overline{X}X) + E(\overline{N}N)}{E(\overline{X}X)} = \varepsilon + \overline{B}B \tag{8}$$

or

$$E(\overline{N}N) = \varepsilon\ E(\overline{X}X) \tag{9}$$

Now, we can see that the *lucky estimator* is optimum under the assumption that signal and noise have the same power spectral shape with a scale factor between them of ε.

No comparison of the above estimators will be made at this time because this problem is a part of a larger problem. The larger problem is that we don't know filter B, we have only estimates $\hat{B}$ of it. These estimates are being used in (E0 - E4) as though they were the true filter B.

Let us hazard the following general formulation of the parsimonious deconvolution problem: We have the estimated deconvolved traces given by

$$\hat{X} = F(\overline{B},B)\quad Y \tag{10}$$

we seek to maximize some parsimony function. In SEP 13, we minimized the various functions S where

$$\min = S[\hat{X}(B,B,Y)] \tag{11}$$

The negative parsimony function S is always a non-linear function of $|x_t|$ which is to say that given an $\hat{x}_t$ we can compute a gradient vector

$$g_t = \frac{\partial S}{\partial x_t} \text{ at } \hat{x}_t .$$

We wish to improve the estimate $\hat{x}_t$ by a small motion $d\hat{x}$ against the gradient, say

$$d\hat{x}_t = -\alpha g_t \tag{12}$$

From here on, we shall no longer need to distinguish $\hat{x}$ from x so will write x everywhere. We need to relate dX to a change in filter dB. From the *noise free estimator* E0, we get

$$dX = -\frac{Y}{B^2} dB = -\frac{X}{B} dB \tag{13}$$

Absorbing α into g and going to the Fourier transform (13) becomes

$$GB = X\, dB$$

Multiplying by $\overline{X}$ gives

$$\overline{X}X\, dB = \overline{X}\, GB$$

The filter incrementation in SEP 13 ignored the positive scaling function $\overline{X}X$ and was actually done with the increment

$$dB = \overline{X}\ GB \qquad (I0)$$

Besides experience, the rationale for omitting positive scaling factors is that we don't have an *a priori* good choice for α anyway nor do we have a good idea how much smoothing on $< \overline{X}X >$ is appropriate. Dropping $< \overline{X}X >$ is like massive smoothing of it.

For the white signal estimators (E1) and (E2), we get

$$dX = \frac{dB}{\varepsilon + < \overline{Y}Y >} Y = -G$$

multiplying by $\overline{Y}$, then taking the conjugate, we get

$$dB = -Y\overline{G} \frac{\varepsilon + < \overline{Y}Y >}{< \overline{Y}Y >} \qquad (I1,I2)$$

In current (Nov 77) experimentation, the right side scaling ratio is being omitted. This seems to be a good idea since it is large only for those frequencies which are *outside* the passband! The *supergain estimator* (E3) would lead to a filter increment similar to (I1,I2).

The most successful deconvolution estimator at present seems to be the *lucky estimator* E4. The algebra for its natural increment is somewhat more complicated. From (E4), we get

$$\frac{\partial X(\overline{B},B)}{\partial \overline{B}} = -\frac{\overline{B}Y}{(\varepsilon + \overline{B}B)^2} B + \frac{Y}{(\varepsilon + \overline{B}B)} \qquad (14)$$

$$dX = \frac{\partial X}{\partial \bar{B}} d\bar{B} + \frac{\partial X}{\partial B} dB \tag{15}$$

$$= \frac{Y\ \bar{B}(B\ d\bar{B} + \bar{B}\ dB)}{(\varepsilon + \bar{B}B)^2} + \frac{Y\ d\bar{B}}{(\varepsilon + \bar{B}B)} \tag{16}$$

$$dX = \frac{-Y\ \bar{B}(B\ d\bar{B} + \bar{B}\ dB) + (\varepsilon + \bar{B}B)Y\ d\bar{B}}{(\varepsilon + \bar{B}B)^2} \tag{17}$$

$$-G = \frac{-Y\ \bar{B}^2\ dB + \varepsilon\ Y\ d\bar{B}}{(\varepsilon + \bar{B}B)^2} \tag{18}$$

$$-(\varepsilon + \bar{B}B)^2\ G = -Y\ \bar{B}^2\ dB + \varepsilon\ Y\ d\bar{B} \tag{19}$$

taking the conjugate

$$-(\varepsilon + \bar{B}B)^2\ \bar{G} = \varepsilon\ \bar{Y}\ dB - \bar{Y}\ B^2\ dB \tag{20}$$

solving (19) and (20) simultaneously

$$\begin{bmatrix} dB \\ d\bar{B} \end{bmatrix} = \frac{(\varepsilon + \bar{B}B)^2}{\bar{Y}Y[(\bar{B}B)^2 - \varepsilon^2]} \begin{bmatrix} YB^2 & \varepsilon Y \\ \varepsilon Y & Y\bar{B}^2 \end{bmatrix} \begin{bmatrix} G \\ \bar{G} \end{bmatrix} \tag{21}$$

which gives the *lucky increment*

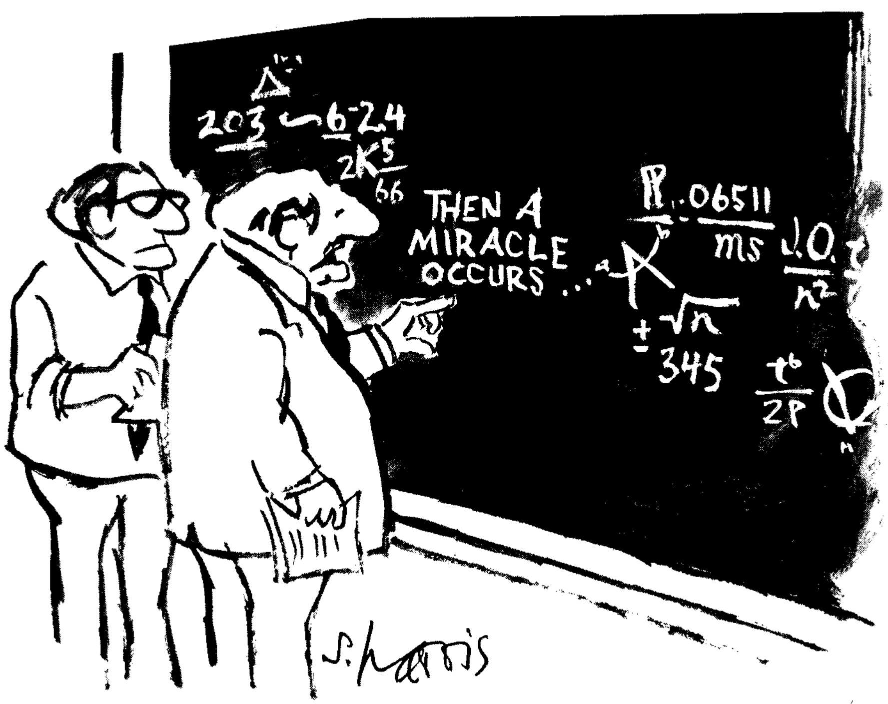
$$dB = \frac{1}{\bar{Y}Y} \frac{\varepsilon + \bar{B}B}{\varepsilon - \bar{B}B} (\bar{Y}\ B^2\ G + \varepsilon\ Y\bar{G}) \tag{I4a}$$

We now recognize a serious hazard with the increment (I4a). The scaling factor ratio contains $\varepsilon - \overline{B}B$ which may change sign depending on whether or not $\overline{B}B$ contains spectral components which are smaller than ε. Thus, the scaling factor really cannot be omitted without danger of positive feedback. One pragmatic way to drop the scaling factor and still prevent such a disaster is a reassignment of the type

$$B \leftarrow (2\varepsilon + \overline{B}B) \,/\, \overline{B} \qquad (22)$$

before the increment I4a is applied. We do indeed find such a reassignment in the SEP 13 program.

Before the reader jumps to the mistaken conclusion that we have shown that the SEP 13 program has a sound and stable theoretical basis, it should be recalled that this program really used the deconvolution estimator (E4) along with the filter increment (I0) and it wouldn't work at all without the reassignment (22).



"I think you should be more explicit here in step two."

American Scientist
1977 November-December 743