Deconvolution of Noisy Experimental Data

GEORGE A. BAKER, JR. AND JOHN E. BROLLEY

*Theoretical and Physics Divisions, University of California,
Los Alamos National Laboratory, Los Alamos, New Mexico 87545*

For a given impulse function, which describes the influence of apparatus on a data stream, we deconvolved the noisy data by two procedures. The first method develops an iterative scheme that searches for a solution that simultaneously is a best least-squares fit to the data and has minimum length. The second procedure applies the method of circulants to obtain a very fast approximation to the first. Good results are obtained for a synthetic problem utilizing a digital computer.

1. INTRODUCTION AND SUMMARY

Many observations of physical phenomena are burdened with noise, which interferes with the reduction of the data. Additionally, the data analysis is further complicated by the modification of the data stream by the observing equipment. The equipment is considered to be the concatenation of hardware, observer, and reduction procedures and associated computing machinery. We do not consider quantum effects [1].

Deconvolution of the signatures of physical phenomena from the data stream has been a major effort for many years [2–5]. From this effort two factors emerge. An estimation procedure must take account of noise and the effects of the observing equipment. The presence of noise requires that we abandon the concepts of an exact deconvolution and search for an estimator. In one guise or another, the effect of noise is treated by filtering the data stream. The effects of the observing equipment are described in the deconvolution procedure by an impulse function which may or may not consolidate several different characteristics of the equipment. In the following analysis the impulse function will be presumed to have either been measured or calculated from known properties of the observing system. The sequencing of the two factors in a deconvolution procedure may be either in tandem or parallel.

We shall consider digitized data streams and apply contemporary computer techniques. We have already treated the non-noisy case by cepstral operators [6] on a digital computer. The cepstral operator, acting on a time series separates convolved factors $a_1(t) * a_2(t)...$ to a sum $\hat{a}_1(T) + \hat{a}_2(T) + ...$ This procedure requires miniscule computer time. In the following we will show that the noisy case can also be treated with the expenditure of very little computer time.

227

In the second section we describe a method which is well justified theoretically and give a procedure to solve the atrociously ill-conditioned equations which result. Considerable computational effort is required. In the third section we show how to modify that method to one for which, although not so well-justified theoretically, the basic equations can be easily solved by fast Fourier transform methods. In the fourth section we illustrate the methods of Sections 2 and 3 with a numerical example. We find that the quality of the deconvolution we obtain with the quick method of Section 3 is not degraded with respect to the computationally more costly method of Section 2. In the final section, we compare the quality of our deconvolution with that of another method now in use.

2. METHOD OF MINIMUM LENGTH

The general problem we are now addressing refers to the case where experimental observations are taken at equally spaced intervals. The instrumental convolution factor is assumed to be known and uncorrelated random noise is present in addition to the convolved signal in the experimental observations. We contemplated the case where the observations are taken sufficiently often so that the noise-free curve is a relatively smooth function of time. Although various criteria for a smoothing fit to the data, $f(x)$, have been proposed, we prefer following Baker [7]

$$\delta \left\{ \int_a^b (f')^2 \, dx \right\} = 0 \tag{2.1}$$

because we feel instinctively that the shortest curve would be the best one. The length of a curve is well known to be

$$\int_{x=a}^b ds = \int_a^b [1 + (f')^2]^{1/2} \, dx. \tag{2.2}$$

Plainly, however, this condition is not independent of the scale for f. To derive a related, scale-free criterion we introduce a scale factor, ε, which we think of as very small. Then (2.2) becomes

$$\int_a^b [1 + \varepsilon(f')^2]^{1/2} \, dx = b - a + \tfrac{1}{2}\varepsilon^2 \int_a^b (f')^2 \, dx + O(\varepsilon^4). \tag{2.3}$$

As the variation $\delta\{b - a\} = 0$ for this problem, we obtain, to leading order in ε, the scale-free criterion (2.1) for the smoothest curve.

For physically reasonable instrumental factors, the output signal comes either after or coincidentally with the input signal, and never before. Further, the absolute time does not matter, but only the delay between the input and the output signals. Thus for

our discrete data with y_i the observations, f_j the true signals, we can represent the instrumental factor as M_{ij} and we have the relation

$$y_i = M_{ij}f_j + \varepsilon_i, \tag{2.4}$$

where ε_i is the random noise, and as explained above M_{ij} is a lower triangular matrix of the form

$$M_{ij} = T(i - j) \tag{2.5}$$

and where

$$T(k) = 0 \quad \text{if} \quad k < 0 \text{ or } k > K. \tag{2.6}$$

The latter requirement says that the output signal drops to zero a finite period of time after the input ceases. We will employ the convention that repeated subscripts are to be summed over.

In addition to the requirement that the function f_i be smooth, we certainly wish to require that it fit the experimental data. Thus we are lead to impose

$$\sum_i (y_i - M_{ij}f_j)^2 \leqslant S_0, \tag{2.7}$$

a least-squares procedure. If available, a knowledge of the expected value of ε_i^2 could be used to determine an acceptable value of S_0. We now impose that the discrete analogue of (2.1) hold, subject to the constraint (2.7) by the standard method of Lagrange multipliers. Define

$$Q = \sum_{i=1}^{N} \frac{(f_{i+1} - f_i)^2}{\Delta} + \lambda \sum_{i=1}^{N} (y_i - M_{ij}f_j)^2 \tag{2.8}$$

with the boundary conditions $f_0 = f_{N+1} = 0$ for added stability. The minimizing equations are

$$\partial(Q)/\partial f = 0, \tag{2.9}$$

where λ is adjusted to impose (2.7). Working out (2.9) we obtain

$$f_{i+1} - 2f_i + f_{i-1} + 2\lambda\Delta M_{ij}^T(y_j - M_{jk}f_k) = 0, \qquad i = 1,..., N; \tag{2.10}$$

M_{ij}^T is the transpose matrix to M_{ij} and Δ is the spacing in x.

In order to solve (2.10), which is found to be a rather ill-conditioned set of equations, we have taken a somewhat round-about approach. First we note that although M_{ij} has all positive or zero entries, this fact alone does not make it a positive definite matrix. For ease in computation we select a μ such that

$$\mathcal{M}_{jk} = M_{jk} + \mu\delta_{jk} \tag{2.11}$$

is a positive definite matrix. In terms of it, we can rewrite (2.10) as

$$2f_i - f_{i+1} - f_{i-1} + 2\lambda\Delta(\mathscr{M}_{ij} - \mu\delta_{ij})^T (\mathscr{M}_{jk} - \mu\delta_{jk})f_k$$
$$= 2\lambda\Delta M_{ij}^T y_j. \tag{2.12}$$

By adding common terms to both sides of (2.12), it can be modified to yield

$$\gamma\mathscr{M}_{ij}^T f_j + (2\lambda\Delta\mu^2 + \Delta_{i-1} - \Delta_i)f_i + 2\lambda\Delta\mathscr{M}_{ij}\mathscr{M}_{jk}f_k$$

$$+ \frac{2\lambda\Delta}{\gamma}(2\lambda\Delta\mu^2 + \Delta_{i-1} - \Delta_i)\cdot\mathscr{M}_{ij}f_j = 2\lambda\Delta M_{ij}^T y_j + \gamma\mathscr{M}_{ij}^T f_i$$

$$+ \frac{2\lambda\Delta}{\gamma}(2\lambda\Delta\mu^2 + \Delta_{i-1} - \Delta_i)\cdot\mathscr{M}_{ij}f_j + 2\lambda\Delta\mu\cdot\mathscr{M}_{ik}f_k + 2\lambda\Delta\mu\cdot\mathscr{M}_{ij}^T f_j, \tag{2.13}$$

where $\Delta_i f = f_{i+1} - f_i$.

The purpose of form (2.13) is to allow us to follow the method of Baker and Oliphant [8]. We propose to solve (2.13) by an iteration in an artificial time. The left-hand side will be at the present artificial time and the right-hand side will be constructed from values f at previous artifical times. We now use the first term on the left-hand side taken together with the second term on the right-hand side to represent

$$\mathscr{M}_{ij}^T \frac{\partial f_i}{\partial t}. \tag{2.14}$$

It is at this step that $\mathscr{M}_{ij}^T$ be positive definite is important as otherwise some modes in our iterative procedures would have turned out to diverge instead of converge. If we used the standard, 3-point forward difference approximation for the artificial time derivative

$$\frac{\partial f^{(n)}}{\partial t} = \left(\frac{3}{2}f^{(n)} - 2f^{(n-1)} + \frac{1}{2}f^{(n-2)}\right)\bigg/\Delta t, \tag{2.15}$$

then we are lead to select the free parameter γ which we introduced as

$$\gamma = \tfrac{3}{2}(\Delta t)^{-1}, \qquad \hat{f}_i = \tfrac{4}{3}f^{(n-1)} - \tfrac{1}{3}f^{(n-2)}, \tag{2.16}$$

where we will substitute $\hat{f}_i$ for f_i on the right side of (2.13). Since (2.13) was specially constructed to factorize, we may write it as

$$\left(M_{ij}^T + \frac{1}{\gamma}(2\lambda\Delta\mu^2 + \Delta_{i-1} - \Delta_i)\delta_{ij}\right)(\gamma\delta_{jk} + 2\lambda\Delta M_{jk})f_k$$

$$= 2\lambda\Delta M_{ij}^T y_j + (\gamma + 2\lambda\Delta\mu)\mathscr{M}_{ij}^T \hat{f}_j$$

$$+ \frac{2\lambda\Delta}{\gamma}(\gamma\mu + 2\lambda\Delta\mu^2 + \Delta_{i-1} - \Delta_i)\cdot\mathscr{M}_{ij}\hat{f}_j. \tag{2.17}$$

The solution procedure is now to write (2.17) as two equations

$$\left(\mathcal{M}_{ij}^T + \frac{1}{\gamma} \left(2\lambda\Delta\mu^2 + \Delta_{i-1} - \Delta_i \right) \delta_{ij} \right) g_j^{(n)}$$

$$= 2\lambda\Delta \mathcal{M}_{ij}^T y_j + (\gamma + 2\lambda\Delta\mu) \mathcal{M}_{ij}^T \hat{f}_j$$

$$+ \frac{2\lambda\Delta}{\gamma} \left(\gamma\mu + 2\lambda\Delta\mu^2 + \Delta_{i-1} - \Delta_i \right) \cdot \mathcal{M}_{ij}^T \hat{f}_j, \qquad (2.18)$$

$$(\gamma\delta_{jk} + 2\lambda\Delta \mathcal{M}_{jk}) f_k^{(n)} = g_j^{(n)}. \qquad (2.19)$$

Now Eq. (2.19) is of lower triangular form so that it can be solved directly by elimination. Equation (2.18) is an upper triangular system, plus one subdiagonal. This system by itself is badly conditioned and its direct inversion is wildly impracticable. Instead, we use an iteration procedure and break (2.18) down as

$$\left[\mathcal{M}_{ij}^T + \frac{1}{\gamma} \left(2\lambda\Delta\mu^2 + 2 \right) \delta_{ij} \right] \tilde{g}_j = \gamma^{-1} \left(g_{i-1}^* + g_{i+1}^* \right)$$

$$+ \text{RHS of (2.18)}, \qquad (2.20)$$

where $g_0^* = g_N^* = 0$. To solve (2.18) we first guess

$$g_j^* = (\gamma\delta_{jk} + 2\lambda\Delta \mathcal{M}_{jk}) \hat{f}_k \qquad (2.21)$$

and solve the upper triangular set of Eqs. (2.20) by direct elimination for $\tilde{g}_i$. We next form by linear extrapolation a new g_j^* from $\tilde{g}_j$ and the old g_j^* and resolve (2.20). We iterate this "minor cycle" until the desired accuracy of solution is obtained as measured by agreement between g^* and $\tilde{g}$. Using this value of g as $g^{(n)}$ we solve (2.19) for $f^{(n)}$. This step completes a "major cycle": to begin a new major cycle we compute a new $\hat{f}$ from (2.16). To start the procedure off we have either used

$$f_i^{(1)} = M_{ij}^{-1} y_j \qquad (2.22)$$

or cepstral operations [6]. Note is made that to use (2.22) we need $T(1) \neq 0$. The major cycle iterations, advancing toward the asymptotic limit in artifical time, are continued until the desired accuracy, as measured by the difference between $f^{(n)}$ and $f^{(n-1)}$, is obtained.

The reader will have noticed that we have introduced a number of parameters, some arbitrary, into our procedure, namely Δ, the time interval spacing in observations, λ the Lagrange multiplier, μ, the modifier of the convolution matrix, and γ the inverse artificial time step. First the equations in fact depend only on $(\lambda\Delta)$ and not λ and Δ separately. As mentioned above λ and hence $(\lambda\Delta)$ is to be adjusted to impose (2.7). We have adjusted μ so that the minor cycle iterations reduced the sum of the squares of the errors by about a factor 10 for every cycle. The faster we make the minor cycle run however, the slower the major cycle runs so this adjustment must

be selected with an eye to overall minimization of computational effort. Finally, guided by the results of Baker and Oliphant [8] we have selected

$$\gamma = 2\lambda\Delta\mu + 50/N. \tag{2.23}$$

As the choices of γ and μ are linked, and we have not been able to make an exhaustive investigation, our selections may not be quite optimum, but we believe that they are not far off.

3. METHOD OF CIRCULANTS

The method described in the previous section is in principle an excellent one; however, quite a bit of computational effort is required to implement it. In this section we describe a variant of it which, though not quite so well based in principle, is computationally much simpler and in our experience gives virtually as good results.

The basis of the method of this section is the observation that if M_{ij} were a circulant then the basic equation (2.10) of the method of the previous section could be solved simply by finite Fourier transform methods, which can now [9] be performed very rapidly. A matrix is called a circulant if any row can be obtained from the proceeding one by passing the last element over the others to the first position. If K [Eq. (2.6)] is considerably smaller than the number of data points N, then if we expand M_{ij} to be a circulant C_{ij}, we have

$$C_{ij} = T((i-j)_{\text{modulo} N}) \tag{3.1}$$

so that C_{ij} differs from the lower triangular matrix M by some non-zero entries in the upper right-hand corner. If enough data has been taken so that all the y_i [eq. (2.4)] which are significantly different from zero have been recorded then, we expect that the f_j for $j = N - K + 1,..., N$ are substantial zero so that the f_j which (2.10) yields will come very close to satisfying (2.10) with M_{ij} replaced by C_{ij}. The situation can be improved further if we expand the data set as

$$
\begin{aligned}
z_j &= 0, & j &= 1,..., K, \\
z_j &= y_{j-K}, & j &= K + 1,..., K + N, \\
z_j &= 0, & j &= K + N + 1,..., N + 2K + L \equiv J,
\end{aligned}
\tag{3.2}
$$

which amounts to introducing a null signal portion before and after the data. The effect of the change from M_{ij} to C_{ij} over this extended data set is to modify the first and last K equations and to leave the central $N + L$ unmodified. Since the deviation of the first K equations is driven by the last K f's we expect that for a large enough extension (L) of the data set that the last f's and hence the first f's will go substantially to zero. The vanishing of the first K f's can be easily verified to be a sufficient condition that the solution of this problem is equal to that of the previous section.

Formally, the problem we wish to solve in this section is

$$f_{i+1} - 2f_i + f_{i-1} + 2\lambda \Delta C_{ij}^T (y_j - C_{jk} f_k) = 0, \qquad i = 1,\dots,J, \qquad (3.3)$$

where $f_{i+J} = f_i$ defines f_0 and f_{J+1}, and the definition of C_{jk} is given by (3.1) except it is modulo J instead of modulo N. Next we introduce the Fourier transform of f

$$g_n = \frac{1}{\sqrt{J}} \sum_{j=1}^{J} \exp(2\pi i n j / J) f_j, \qquad n = 1,\dots,J, \qquad (3.4)$$

where by standard theory,

$$f_j = \frac{1}{\sqrt{J}} \sum_{n=1}^{J} \exp(-2\pi i n j / J) g_n, \qquad j = 1,\dots,J. \qquad (3.5)$$

If we substitute (3.5) in (3.3) we obtain, after some manipulation,

$$g_n = \frac{(\lambda \Delta / \sqrt{J})[t(n)]^* \sum_{j=1}^{J} \exp(2\pi i n j / J) z_j}{4 \sin^2(\pi n / J) + \lambda \Delta |t(n)|^2}, \qquad (3.6)$$

where $*$ denotes complex conjugate, $| \, |$ denotes absolute value and we have defined

$$t(n) = \sum_{l=0}^{K} T(l) \exp(2\pi i n l / J). \qquad (3.7)$$

The solution of (3.3) is now given by substitution of the values of g_n in (3.5) to recover the f_j.

A convenient formula for the overall error is given by the Pareseval type relation.

$$\sum_{l=1}^{J} \left(z_l - \sum_{j=1}^{J} C_{lj} f_j \right)^2$$

$$= J^{-1} \sum_{n=1}^{J} \frac{16 \sin^4(\pi n / J) | \sum_{j=1}^{J} \exp(2\pi i n j / J) z_j |}{[4 \sin^2(\pi n / J) + \lambda \Delta |t(n)|^2]^2} \qquad (3.8)$$

This error, of course, is not identical with that of (2.7) owing to the presence of the $2K + L$ extra zeros being fitted, but that error could be computed directly from the f_j's obtained here.

4. NUMERICAL RESULTS

We now consider a common numerical example for both procedures. The object data is a 1024 point data sequence containing the letter **M**, whose central depth is $\frac{1}{2}$ its height.

This sequence was convolved with the 1024 point sequence, Fig. 1, representing an
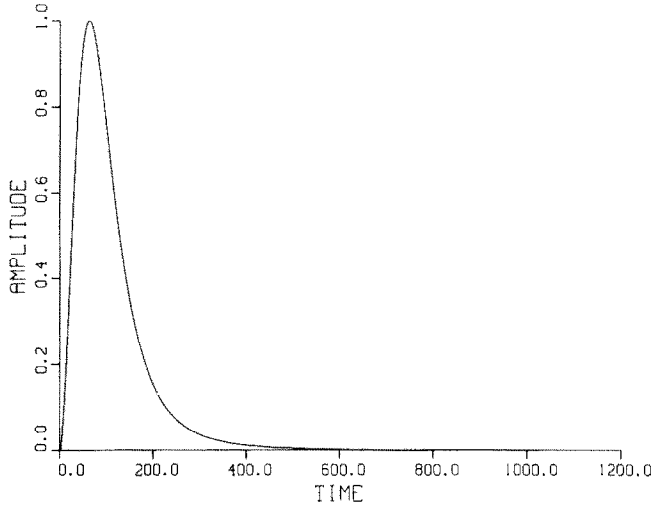
FIG. 1.   1024 point sequence representing the impulse function (07/02/82, 14:53:51).

impulse function. To this convolution we have added approximately 0.1% rms noise. The contaminated data stream, shown in Fig. 2, can no longer be deconvolved by a direct application of the cepstral operators [6].

The impulse and data sequences comprise the inputs for a deconvolution procedure. We consider first the minimum length technique. There are several possibilities for providing a starting solution. One rapid computational technique
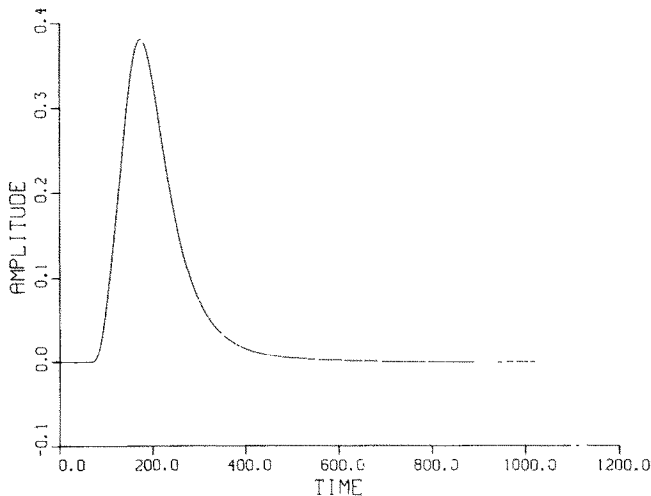
FIG. 2.   1024 point sequence representing the contaminated data stream (07/02/82, 14:52:21).
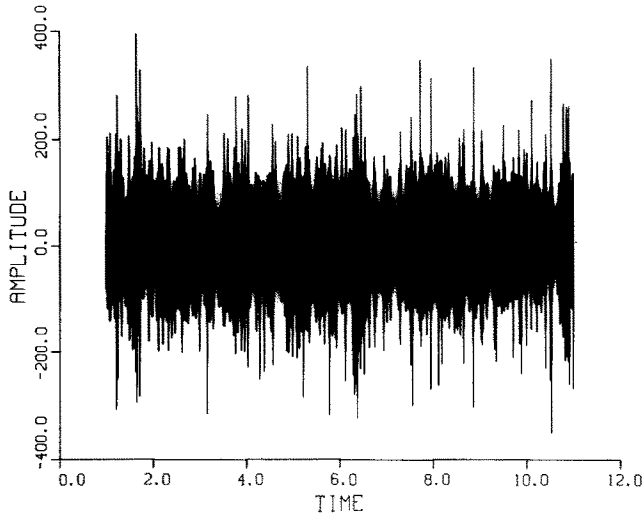
FIG. 3. Starting solution produced by deconcolving the data with the impulse function (07/02/82; 14:56:06).

would be to suitably filter the noisy data and apply the cepstral operators. Another approach which we now apply is to simply deconvolve the data stream with the impulse function. The result is the starting solution shown in Fig. 3. The iteration procedure is then initiated. Figures 4, 5, and 6 illustrate the state of deconvolution at 50, 100, and 150 major cycles for $\lambda = 100$. The filtering action, driven by the
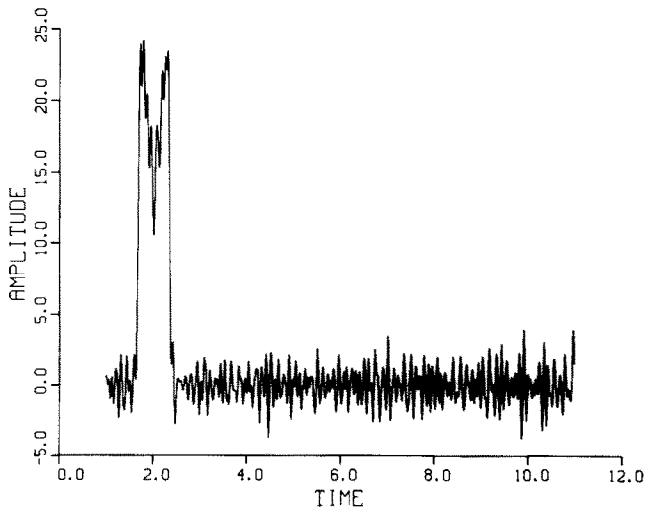
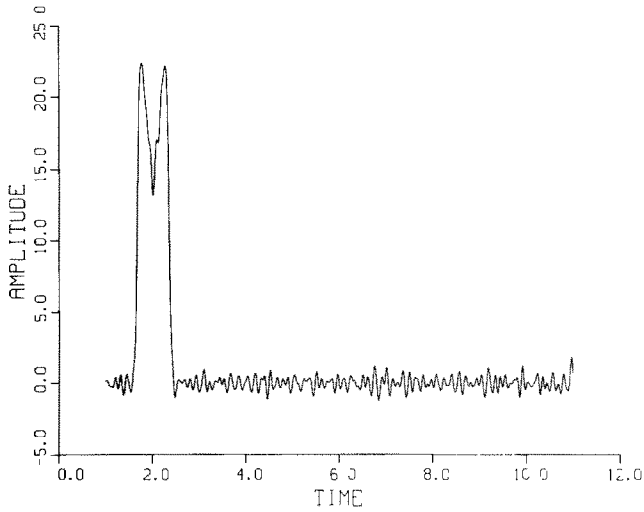FIG. 4. Deconvolution at 50 cycles (07/02/82, 15:36:29).

FIG. 5.   Deconvolution at 100 cycles (07/02/82, 15:40:31).

minimum length consideration, is clearly discernable. The effective tension in the filtering process is altered by changing the Lagrange multiplier.

We now apply the method of circulants to the same input data. Comparison of Eqs. (2.8) and (3.3) indicates the modified role of λ. We let the number of trailing zeros be 100 and deconvolved with $\lambda = 1$, 10, and 100. The respective results are shown in Figs. 7, 8, and 9. A different normalization owing to technical reasons was
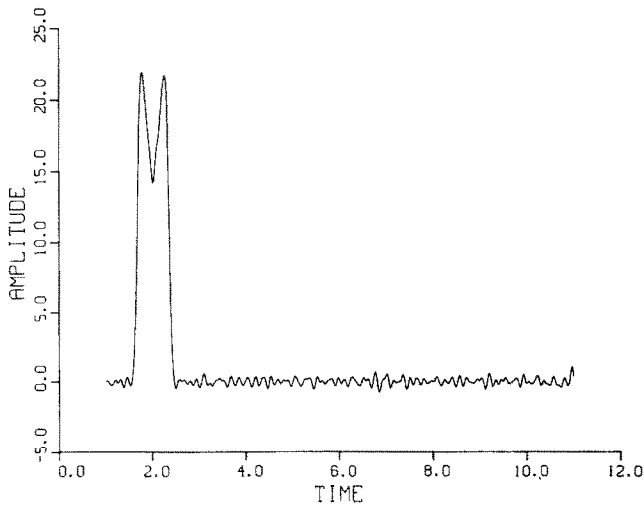

FIG. 6.   Deconvolution at 150 cycles (07/02/82, 15:43:14).
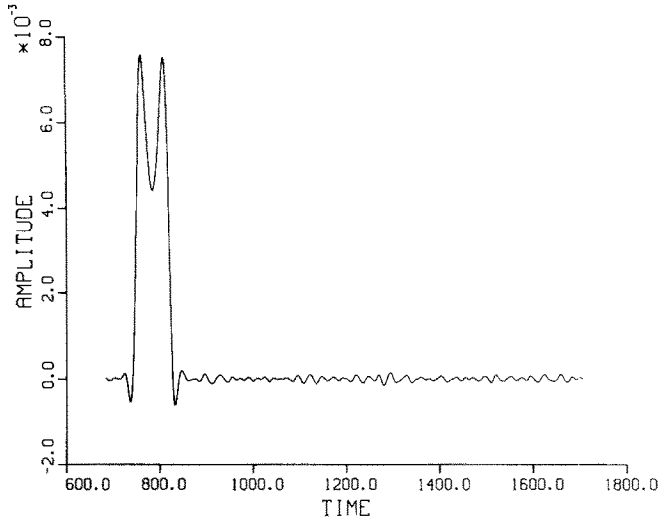
FIG. 7.   Deconvolution by the method of circulants with $\lambda = 1$ (07/02/82, 15:51:53, KZ = 100).

used to draw the figures by the two methods and is of no significance. Changing the number of trailing zeros to 200 does not noticeably alter these results, although for a different type of impulse function or initial convolved data it might well.

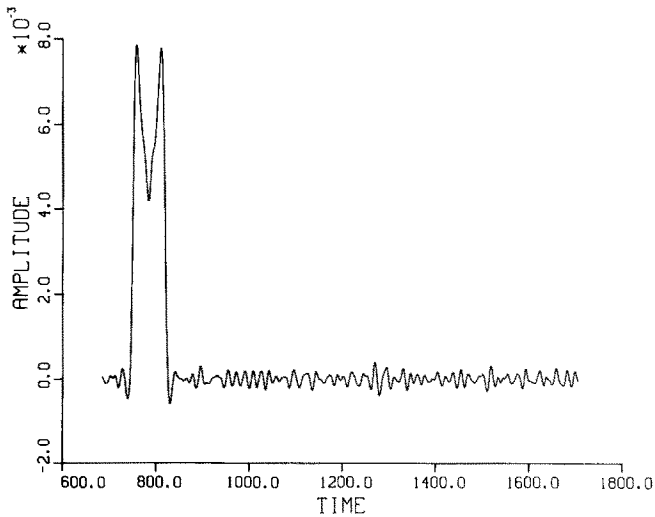The minimum length procedure requires considerably more time on a digital

FIG. 8.   Deconvolution by the method of circulants with $\lambda = 10$ (07/02/82, 15:57:53, KZ = 100).
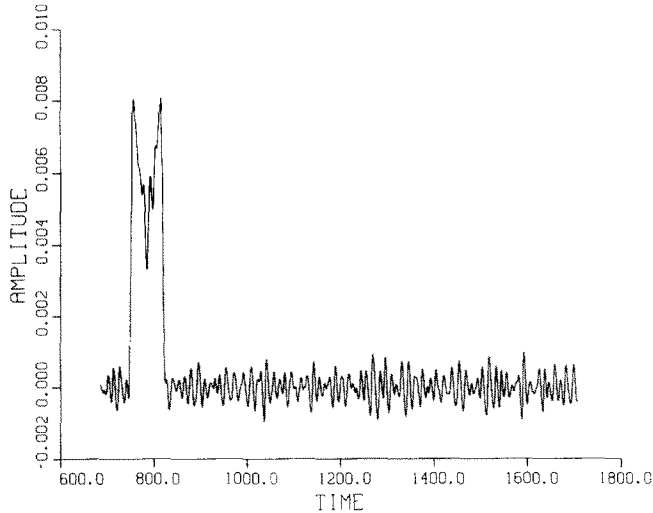
FIG. 9. Deconvolution by the method of circulants with $\lambda = 100$ (07/02/82, 15:35:19, KZ = 100).

computer (CRAY in the present work) than the method of circulants where only a fraction of a second was required per case for our example. We expect that the ratio of times is problem-specific.

5. CONCLUSIONS

While the problem of deconvolution has confronted the experimentalist for a long time, it would appear that the first formal analysis of the problem was given by Van Cittert [10] in 1931. Variations on his iterative procedure have appeared from time to time and his method has recently been adapted to noisy data by Thomas [11]. Since 1931, numerous other procedures [12], differing from those discussed in this paper, have been developed. It is manifestly impossible to compare the present work with this large corpus of papers.

It is useful, however, to compare the present results with a contemporary efficacious technique operating on the same impulse function and noisy data stream used in this paper. E. Hodson and J. Canada[1] of this laboratory have applied the computer program CONFOLD [13] with the result shown in Fig. 10. CONFOLD is a program which determines the minimum required truncation of the impulse function by iteration, and unfolds a Gaussian filter, producing a constrained inverse which is then folded with the data. The reader should notice the relative degrees of smearing at

---

[1] Private communication. We thank these authors for their help.
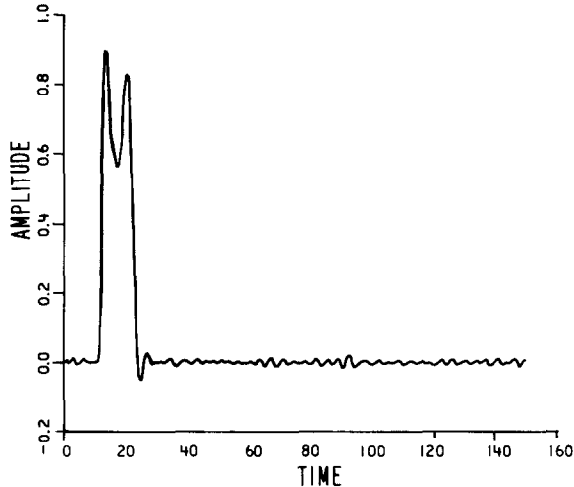
FIG. 10.   Deconvolution by CONFOLD.

the peaks and the relative degrees of filling in at the minimum of the unfold data by the different methods reported here.

We feel that the minimum length constraint on a deconvolution procedure is a viable technique that is computationally robust and makes minimum demands on a starting solution. By synthetic numerical example we have demonstrated, by this technique, useful recovery of information from a noisy data stream. We have further provided a rapid approximation procedure which is very likely considerably faster than many extant iterative procedures and is well adapted to small computers.

REFERENCES

1. C. W. HELSTROM, "Quantum Detection and Estimation Theory," Academic Press, New York, 1976.
2. A. PAPOULIS, "Signal Analysis," McGraw–Hill, New York, 1977.
3. A. V. OPPENHEIM AND R. W. SCHAFER, "Digital Signal Processing," Prentice–Hall, Englewood Cliffs, N.J., 1975.
4. E. A. ROBINSON, "Statistical Communication and Detection," Griffin, London, 1967.
5. C. W. HELSTROM, "Statistical Theory of Signal Detection," Pergamon, Oxford, 1968.
6. J. E. BROLLEY AND G. A. BAKER, "Deconvolution of Diagnostic Data," Los Alamos preprint LA-UR-81-3126 available from Technical Information Center, Oak Ridge, Tenn., see also J. M. Tribolet, "Seismic Applications of Homomorphic Signal Processing," Prentice–Hall, Englewood Cliffs, N. J., 1979.
7. G. A. BAKER, JR., J. Comp. Phys. 3 (1969), 486–493.
8. G. A. BAKER, JR. AND T. OLIPHANT, Quart. Appl. Math. 17 (1960), 361–373.
9. E. BRIGHAM, "The Fast Fourier Transform," Prentice–Hall, Englewood Cliffs, N. J., 1974.
10. P. H. VAN CITTERT, Z. Phys. 69 (1931), 298.

11. G. THOMAS, *in* "Proceedings, 1981 IEEE International Conference on Acoustics. Speech and Signal Processing," Vol. 1, p. 47, 1981.
12. "Deconvolution," Vols. I, II, Geophysics Reprint Series. Soc. Explor. Geophysicists, 1978, Tulsa. Oklahoma.
13. E. K. HODSON, "Predictable Unfolding in the Time Domain," Los Alamos Report LA-3830, 1968, available from Technical Information Center, Oak Ridge, Tennessee.