Decomposition by Conjugate Gradients

Jon F. Claerbout

ABSTRACT

By means of conjugate gradients, a field profile is decomposed into frequency components.

INTRODUCTION

One way to decompose a seismogram into frequency components is to put the seismogram through a bank of narrow-band filters. The trouble with this way is that the sum of the outputs of the filters would be very unlikely to sum to the original data. One way for the filter outputs to sum to the input would be for the narrow-band filters to be a carefully chosen family of *sinc* functions.

Another way to decompose a function is the decomposition filter pairs in IEI page 122-124. These filters have 6db cutoffs. But they can be compounded in various ways. I have put some such decompositions on previous SEP video tapes. Unfortunately I never found an appealing systematic analysis.

Still another way to decompose a seismogram into frequency components is to solve an underdetermined inverse problem. Given *any* bank of filters, say F_j, we can find the inputs, say $x_j(t)$, such that the outputs sum to our given data, say $d(t)$. The filters may be bandpass filters with spectral overlap.

Frequency decomposition is just one of many types of decomposition. Data can be decomposed by other characteristics such as dip or velocity. This study takes up decomposition by frequency components, but it is also a prototype for other decompositions such as those of Thorson and Harlan.

DECOMPOSITION CONCEPTS

For theoretical analysis let us consider a two component decomposition. The program to be developed will be generalized to multicomponents. We have the constraint equation:

$$d \quad = \quad F_1 x_1 + F_2 x_2 \tag{1}$$

Equation (1) is underdetermined. One solution is to set $x_2 = 0$ and solve for x_1. Another solution would do the reverse.

Quadratic examples

The solution of an underdetermined problem can be made unique by minimizing some quadratic, for example:

$$E_2 \quad = \quad |\, x_1 \,|^2 + \,|\, x_2 \,|^2 \tag{2}$$

Since (2) is just one of many functions that could be minimized we need to consider some of the various possibilities, and the answers they would lead to. For example, we could minimize

$$E_F \quad = \quad |\, F_1 x_1 \,|^2 + \,|\, F_2 x_2 \,|^2 \tag{3}$$

subject to the constraint (1). To see the answer to the minimization (1), let $y_i = F_i x_i$. The data gets divided into two halves, $y_1 = d/2$ and $y_2 = d/2$. This is not a very interesting decomposition. The two terms in (2) can be weighted arbitrarily. Weighting one term more than the other serves to push more energy from one term to the other. Likewise, the weighting could be frequency dependent (as (3) is). So you see, you really choose the answer by chosing the weights. It would be nice to use probability or parsimony concepts to be able to chose a decomposition that is somehow "best".

Pattern recognition inherent to L_1

As we have seen, the minimization of quadratic functions tends to distribute energy uniformly among its available places. This is not the case for the minimization of sums of absolute values. In some cases the absolute value minimizations almost magically give the "human" answer. I have devised the following illustration: Consider the decomposition

$$\begin{bmatrix} 12 \\ 10 \end{bmatrix} \quad = \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} x_1 \; + \; \begin{bmatrix} 2 \\ 0 \end{bmatrix} x_2 \; + \; \begin{bmatrix} 1 \\ -1 \end{bmatrix} x_3 \tag{4}$$

Before you look at any mathematics, think about what numerical choice you would like for x_1, x_2, and x_3.

Minimization of $x_1^2 + x_2^2 + x_3^2$ tends to smear the energy among the x_i as uniformly as the constraint (4) will allow. The methods in FGDP show this answer to be $(x_1, x_2, x_3) = (7, 4, -3)$.

Minimizing the sum of absolute values

$$E_1 \quad = \quad | x_1 | + | x_2 | + | x_3 | \tag{5}$$

yields the answer $(x_1, x_2, x_3) = (10, 1, 0)$. This answer fits my intuition better. There is no prejudice against one component absorbing most of the data. The poorest fitting component is rejected entirely.

Best weights

We may prefer absolute-value optimizations over quadratic optimizations, but we have no techniques for solving absolute-value optimizations with a million unknowns. With the conjugate gradient technique we can often get usable solutions of weighted quadratic optimizations. An absolute value is the same as a square divided by an absolute value. So with luck, we may be able to introduce appropriate weights to get a quadratic solution to look like an absolute value solution.

There are two different philosophies on what to minimize. Suppose that some statistical analysis tells you that the matrix M is the inverse of the covariance matrix of $x^T x$. If you don't know anything about crosscovariance, then you take the matrix M to be diagonal. One philosophy is to minimize the dimensionless quantity $x^T M x$. The other philosophy is to use the square root of M, because it is more like the L_1 norm problem.

DECOMPOSITION EXAMPLE

A well known way to deal with constraints is to use a penalty function. So we set up the regression:

$$\begin{pmatrix} 0 \\ 0 \\ d \end{pmatrix} \quad \approx \quad \begin{pmatrix} \lambda_1 I & 0 \\ 0 & \lambda_2 I \\ F_1 & F_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{6}$$

In the limit $\lambda_1 = \lambda_2 \to 0$ the regression (6) minimizes (2) or (5) subject to the constraints (1). The regression can be done by the method of the paper "Conjugate Gradients for Beginners" found elsewhere in this report.

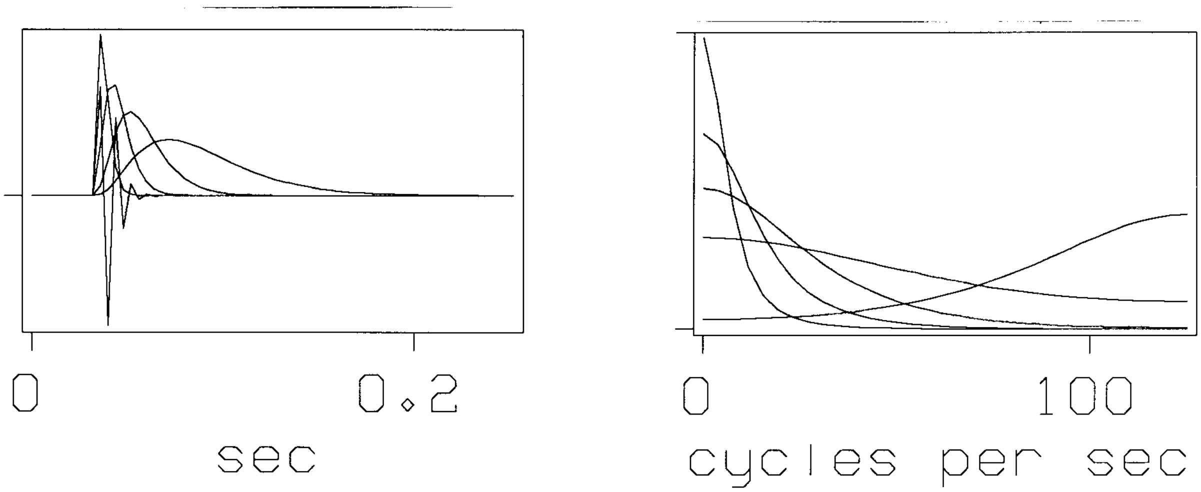

FIG. 1. Time domain response and frequency domain response of filter of equation (7) for values of $\rho_i = -.2, .1, .35, .6, .8$.

For our present example the matrix F is a filter matrix. (See Claerbout "What is ..."). In this trial the filter will be the recursive filter given in Z-transform notation by

$$F_i = \left(\frac{1 + \rho_i Z}{1 - \rho_i Z} \right)^3 \tag{7}$$

This filter is a leaky triple integrator. So the time domain response is something like $t^2 e^{-\alpha_i t}$ and the frequency domain response is something like $1/(\omega^2 + \beta_i)^{3/2}$. Plots are shown in figure 1. The filters appear to form a reasonable set of "basis functions" to use to decompose a seismogram. These basis functions, however, are clearly far from being orthogonal.

A field profile (Yilmaz and Cumro #32) was selected because of its broad spectral band. Linear moveout was performed with a velocity of 3 km/s. This profile is shown in figure 2. Several prominent head waves are over moved out. Figure 2 also shows the residual after a few descent steps. Figure 3 shows the $F_i x_i$ for various ρ_i. Figure 4 shows the x_i for various ρ_i.

CONCLUSIONS

There are a great number of loose ends. I am only writing this up because of the progress report deadline. Personally I found it rewarding only because it was the first time I solved an optimization problem with $400 \times 48 \times 5$ unknowns. I was planning to include my program for its tutorial value, but after studying the results I decided that this frequency decomposition had no apparent practical utility, so I decided not to.

REFERENCES

Claerbout, Jon F., 1984, Decomposition by markov processes, SEP 42 pp 65-68.

Claerbout, Jon F., 1984, What is the transpose operation?, SEP 42, pp 113-128

Harlan, Bill, numerous SEP report papers.

IEI, refers to Imaging the earth's interior, SEP 40.

Thorson, Jeff, 1984, Velocity and slant stack inversion, SEP 39.

Yilmaz, O., and Cumro, D., 1983, Worldwide assortment of field seismic records, released by Western Geophysical Company of America, Houston

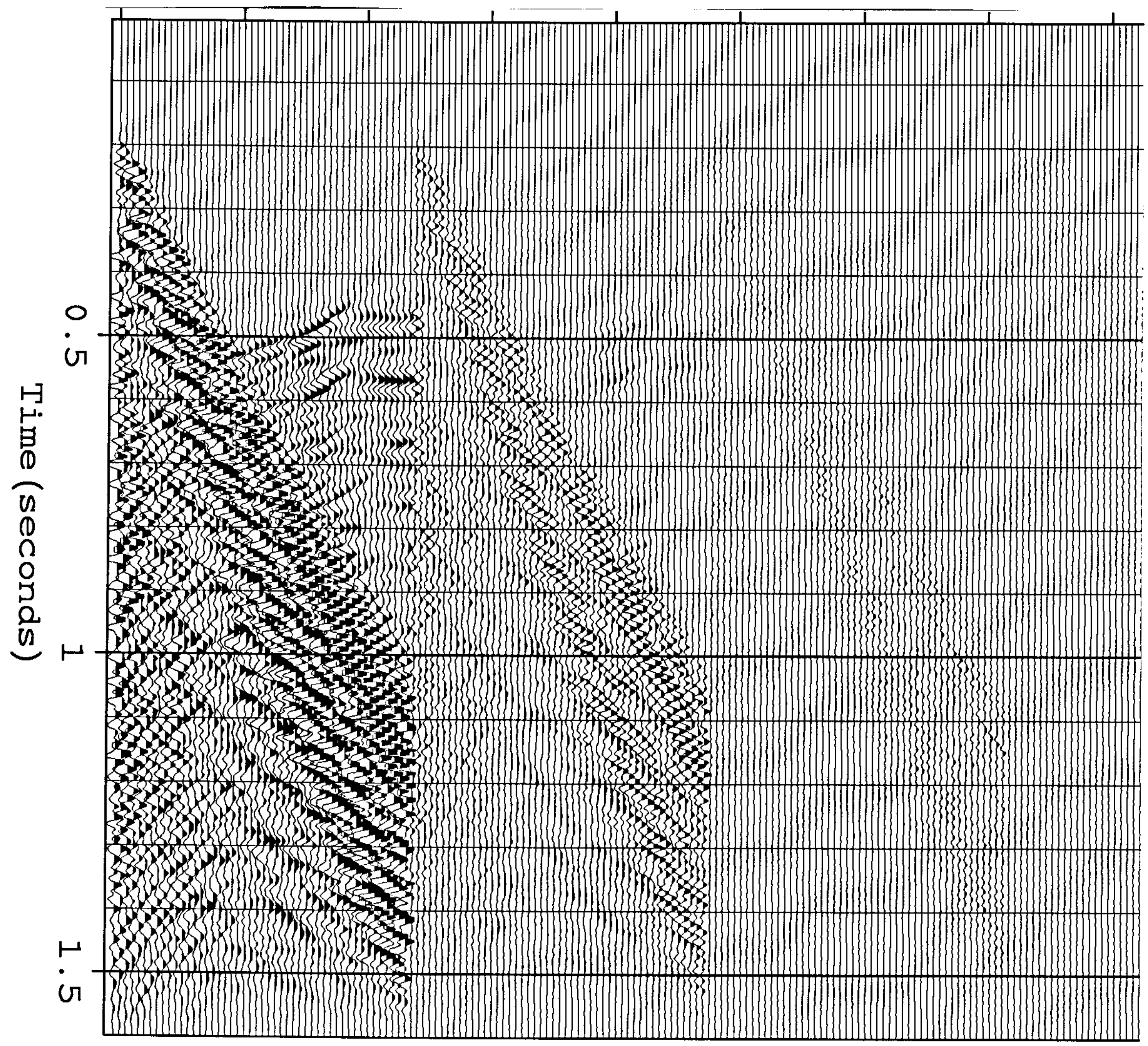

FIG. 2. A field profile (Yilmaz and Cumro #32) was multiplied by t^2, linearly moved out with a velocity of 3 km/s, and plotted on the left. (Head waves appear over moved out). Subsequent panels show the residual $d - \sum_i F_i x_i$. Note that the residual drops rapidly with iteration.

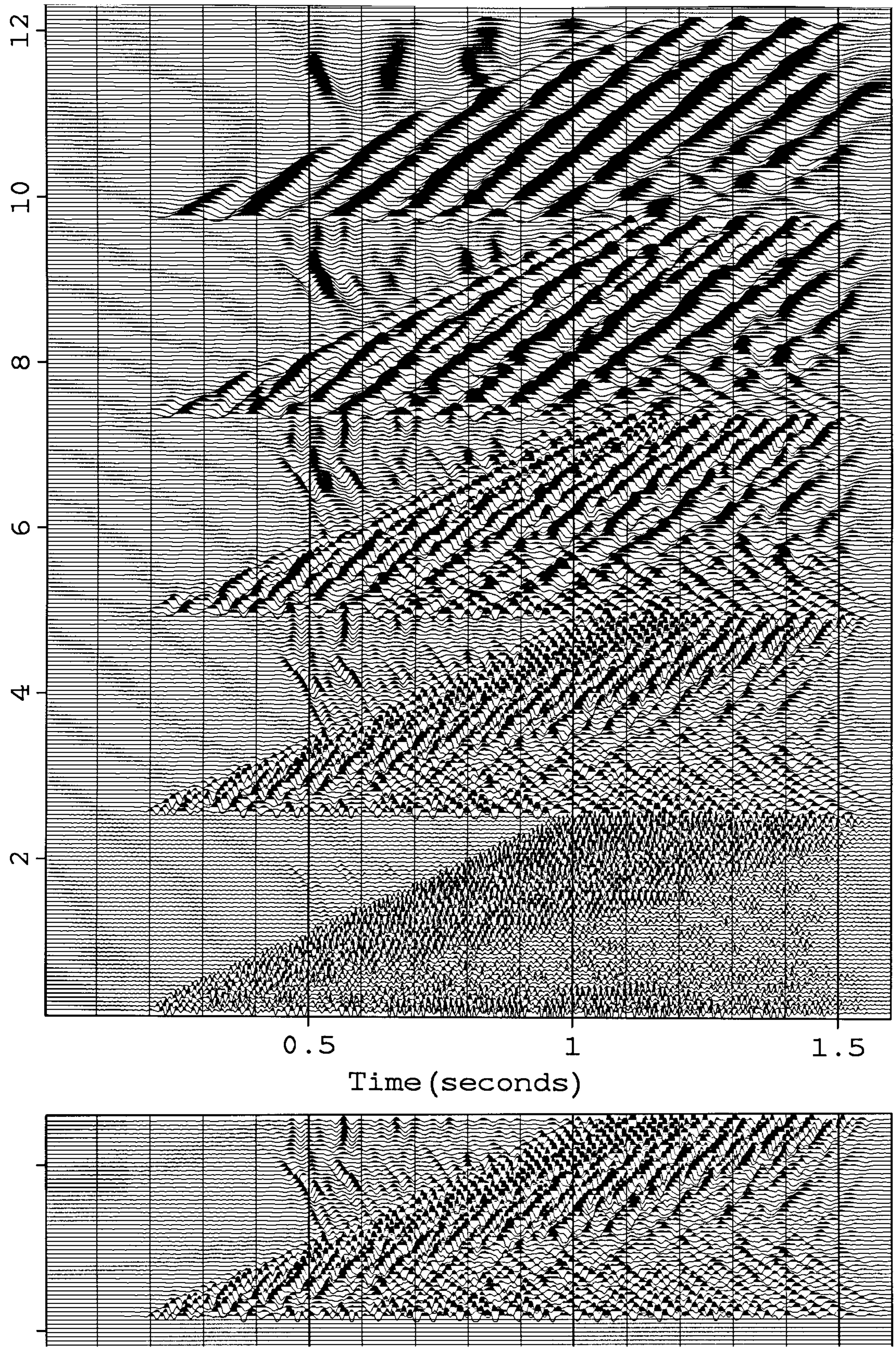

FIG. 3. Moved out field profile (left) and its decomposition into 5 frequency components (right). Each panel is scaled for plotting.

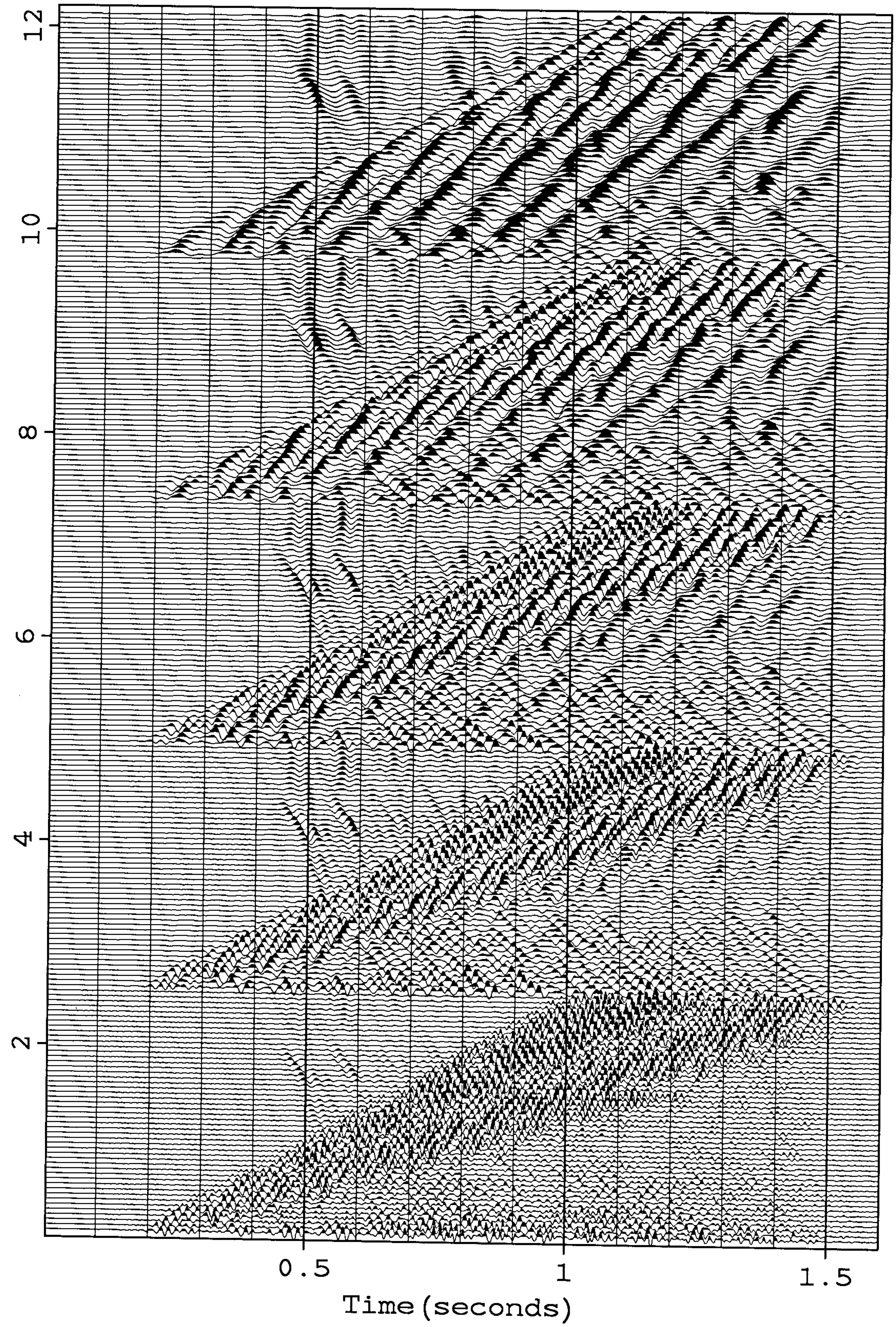

FIG. 4. The decomposition components x_i. Each panel is rescaled for plotting. Observe that the panels are not white. Also, some outputs can be seen before $t = 0$.