Neural Networks for Seismic Principal Components Analysis

Kou-Yuan Huang, *Senior Member, IEEE*

Abstract—The neural network, using an unsupervised generalized Hebbian algorithm (GHA), is adopted to find the principal eigenvectors of a covariance matrix in different kinds of seismograms. We have shown that the extensive computer results of the principal components analysis (PCA) using the neural net of GHA can extract the information of seismic reflection layers and uniform neighboring traces. The analyzed seismic data are the seismic traces with 20-, 25-, and 30-Hz Ricker wavelets, the fault, the reflection and diffraction patterns after normal moveout (NMO) correction, the bright spot pattern, and the real seismogram at Mississippi Canyon. The properties of high amplitude, low frequency, and polarity reversal can be shown from the projections on the principal eigenvectors. For PCA, a theorem is proposed, which states that adding an extra point along the direction of the existing eigenvector can enhance that eigenvector. The theorem is applied to the interpretation of a fault seismogram and the uniform property of other seismograms. The PCA also provides a significant seismic data compression.

Index Terms— Data compression, eigenvectors, generalized Hebbian algorithm, neural network, principal component analysis (PCA), Ricker wavelets, seismic interpretation.

I. INTRODUCTION

THE PRINCIPAL components analysis (PCA), also known as the Karhunen–Loeve transformation, has been investigated and used in many applications [1]–[8]. Given a set of random data $\bar{X}$ with dimension N and $\bar{M} = E[\bar{X}] = \bar{0}$, we can compute the correlation matrix (covariance matrix) $Q = E[(\bar{X} - \bar{M})(\bar{X} - \bar{M})^T] = E[\bar{X}\bar{X}^T]$ and find the eigenvalues and the corresponding eigenvectors. The principal eigenvectors can point to the principal directions of the distribution of the data.

The PCA was applied to a seismic data set by Hagen [7] and Jones [8]. Hagen adopted the input data vector in the vertical trace direction and computed the principal components to evaluate the subtle character changes of porosity in the neighboring uniform seismic traces. Jones adopted the input data vector in the horizontal direction and computed principal component to separate the diffraction and reflection patterns from seismic data after normal moveout correction (NMO).

Several neural network algorithms have been proposed for PCA [9]–[20]. Oja's learning rule could find one principal eigenvector of a covariance matrix [9]–[13]. Sanger's unsu-

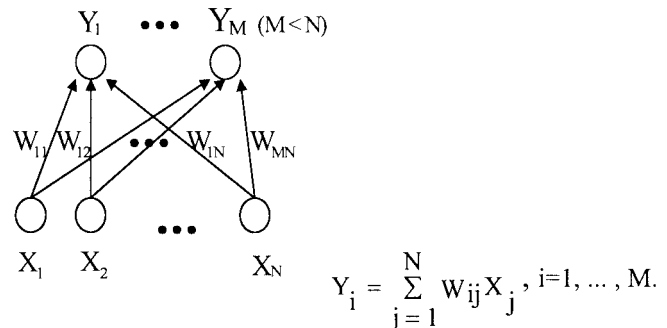$$Y_i = \sum_{j=1}^{N} W_{ij} X_j, \quad i = 1, \ldots, M.$$

Fig. 1. Sanger's neural net.

pervised generalized Hebbian learning algorithm (GHA) could find many principal eigenvectors of a covariance matrix in decreasing eigenvalue order [14], [15].

In this paper, the neural net with Sanger's unsupervised GHA [14] is adopted to find the principal eigenvectors of a covariance matrix in different kinds of seismograms. The neural net is shown in Fig. 1. The input data were fed into the neural net iteratively to find the principal eigenvectors. The advantage of learning is the ability to find the principal eigenvectors incrementally from data as it becomes available. It is not necessary to compute the covariance matrix from the complete data set in advance; the eigenvectors can be derived directly from the data. If we use the traditional method of eigenvector analysis to seismic data with 512 input dimensions, the covariance matrix with dimension 512×512 may be over the memory limitation of some computers. And we cannot solve the eigenvectors by the traditional method. However, using neural net with GHA can decrease the computational requirement and storage for the small number of output eigenvectors. For example, using the notation in the following Section II, if there are five output eigenvectors, $\vec{Y}\vec{X}^T$ will have only $5 \times 512 = 2560$ elements and $\vec{Y}\vec{Y}^T$ will have only 25 elements. The required memory is less. Sanger already stated that "when the number of inputs is large and the number of required outputs is small, GHA provides a practical and useful procedure for finding eigenvectors" [14]. Seismic data always have a large data set with large dimensions, and we want to extract the information of a small number of reflection layers and uniform traces to improve the seismic interpretations. So from the advantages stated above, PCA using GHA might be needed for seismic data.

Fig. 2 shows the processing steps using the neural net based on Sanger's GHA learning rule for the seismic PCA with the
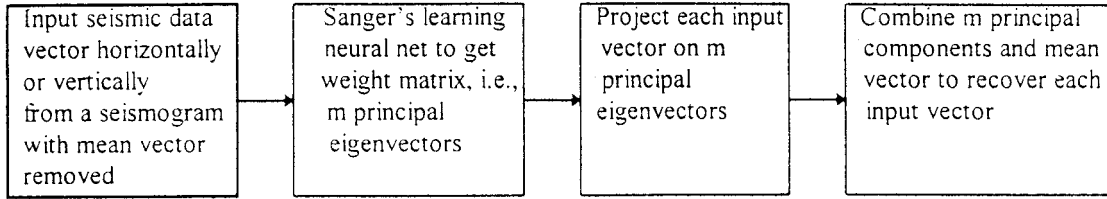
Input seismic data vector horizontally or vertically from a seismogram with mean vector removed	Sanger's learning neural net to get weight matrix, i.e., m principal eigenvectors	Project each input vector on m principal eigenvectors	Combine m principal components and mean vector to recover each input vector

Fig.2.   System of Sanger's learning net for seismic PCA.

input data vectors from horizontal or vertical directions. The analyzed seismic data are the seismic traces with 20-, 25-, and 30-Hz Ricker wavelets, the fault, the reflection and diffraction patterns after NMO correction, and the bright spot pattern. The real seismogram at Mississippi Canyon is also used in the experiment.

II. NEURAL NETWORK OF GENERALIZED HEBBIAN LEARNING RULE

In Fig. 1, the neural net with Sanger's GHA has a single weighting layer. The input N-dimensional column vector is $\bar{X}$, $E[\bar{X}] = \bar{0}$, the weight matrix W is $M \times N$, and the output M-dimensional column vector is $\bar{Y} = W\bar{X}$ with $M < N$, i.e., $y_i = \sum_{j=1}^{N} W_{ij}x_j$, $i = 1, \cdots, M$. The input data are fed into the net iteratively to find the principal eigenvectors of the correlation (covariance) matrix $Q = E[(\bar{X} - \bar{M})(\bar{X} - \bar{M})^T] = E[\bar{X}\bar{X}^T]$. The weight matrix W needs to be updated for each iteration step.

Sanger [14] proved that the GHA neural net converges from random initial weights with probability one to find the principal eigenvectors of the input covariance matrix in decreasing eigenvalue order. The GHA learning rule from Sanger [14] is given by

$$W_{ij}(t+1) = W_{ij}(t) + \gamma(t)y_i\left(x_j - \sum_{k=1}^{i} y_k W_{kj}\right)$$

where j is the index of input, i is the index of output, or in matrix form as

$$\Delta W(t) = \gamma(t)(\bar{Y}\bar{X}^T - LT[\bar{Y}\bar{Y}^T]W(t))$$

where $\bar{X} = [x_1, x_2, \cdots, x_N]^T$, $\bar{Y} = [y_1, y_2, \cdots, y_M]^T$ and $LT[.]$ is the lower triangular matrix, i.e., all elements above the diagonal of its matrix argument are zeros. $\lim_{t\to\infty}\gamma(t) = 0$ and $\sum_{t=0}^{t=\infty}\gamma(t) = \infty$. The ith eigenvector $\mathbf{e}_i$ is W_{ij}, $j = 1, \cdots, N$, and the corresponding eigenvalues have the relation $\lambda_1 > \lambda_2 > \cdots > \lambda_i > \cdots > \lambda_M$.

The "projection values" of the input data vector $\bar{X}$ onto unit eigenvectors $\mathbf{e}_i$ are $a_i = \bar{X}^T\mathbf{e}_i = \mathbf{e}_i^T\bar{X}$, $i = 1, \cdots, M$. We define $a_i\mathbf{e}_i$ as the "projection vectors" or "principal components" of input data $\bar{X}$ onto $\mathbf{e}_i$. We use a small number of a_i and $a_i\mathbf{e}_i$ in the interpretations of seismic principal components.

In our seismic experiments, the definition of convergence is, if all absolute values of $W_{ij}(t+1) - W_{ij}(t)$ are less than a constant 0.0001, the iteration stops. For the examples in [14], Sanger chose $\gamma(t)$ empirically at a fixed value between 0.1

and 0.01, which provides good convergence. Here we adopt Sanger's empirical values on $\gamma(t)$ that $\gamma(t)$ is held fixed at a value between 0.1 and 0.01. From our seismic experiments, the number of iteration is increased as the value of $\gamma(t)$ is decreased. Also, we have other experiments on $\gamma(t)$ that decreases linearly from 0.1 to 0.01, then keeps a small fixed value after some iteration. The designed formula is

$$\gamma(t) = -0.1 * (\text{# of iteration - 1000})/1000, \text{ when # of}$$
$$\text{iteration is less than or equal to 900}$$
$$\gamma(t) = 0.01, \text{ when # of iteration is greater than 900.}$$

In order to have the criterion for comparison, we calculate the mean-squared error (MSE) and the normalized mean-squared error (NMSE) [14]. The NMSE is the ratio of the mean-squared error (MSE) to the data variance. For the seismic intensity I at position m, n with removing the average, and the recovered intensity $\hat{I}$,

$$\text{NMSE} = E[(I_{m,n} - \hat{I}_{m,n})^2]/E[I_{m,n}^2].$$

III. DATA COMPRESSION

We can select the p largest eigenvalues and their corresponding p eigenvectors and discard the smallest $(M-p)$ eigenvalues in the data representation. The data representation for $\bar{X}$ becomes $\bar{X}' = \sum_{i=1}^{p} a_i\mathbf{e}_i$. The MSE value between random vector $\bar{X}$ and $\bar{X}'$ is the summation of the discarding $M - p$ eigenvalues [1]. There are two meanings in this technique of dimensionality reduction: 1) we can save a lot of memory space in data representation and 2) we can select the effective features in pattern recognition.

IV. EFFECT OF ADDING ONE EXTRA POINT ALONG THE DIRECTION OF EXISTING EIGENVECTOR

For PCA, a theorem is proposed, which states that adding an extra point along the direction of the existing eigenvector can enhance that eigenvector.

Theorem 1 (Effect of Adding One Extra Point Along the Direction of Existing Eigenvector of a Covariance Matrix): Given a covariance matrix Q computed from N data and its eigenvalues λ_i and eigenvectors $\mathbf{e}_i$, if adding one extra data vector (mean vector is removed) along the direction of existing eigenvector $\mathbf{e}_i$, then 1) the new eigenvalue λ_i' is larger than the original eigenvalue λ_i if the square length of the extra data vector is greater than λ_i, otherwise λ_i' is less than or equal to λ_i and 2) the new eigenvalue λ_i' is increased more than other new eigenvalues $\lambda_j'(j \neq i)$.

Proof: The extra data vector (mean vector is removed) is represented as $c\mathbf{e}_i$, c is the length of the vector $c \neq 0$. The set of the new eigenvalues and eigenvectors computed from the $N + 1$ data is as follows.

The new covariance matrix is $\mathbf{Q}' = (1/N+1)[N\mathbf{Q} + (c\mathbf{e}_i)(c\mathbf{e}_i)^T]$. For Q, its eigenvalues are λ_i and its eigenvectors are $\mathbf{e}_i$. For matrix $\mathbf{e}_i\mathbf{e}_i^T$ because $\mathbf{e}_i\mathbf{e}_i^T\mathbf{e}_i = \mathbf{e}_i$, $(\mathbf{e}_i^T\mathbf{e}_i = 1)$, the eigenvector is $\mathbf{e}_i$. So the eigenvectors of new Q' are the same as those of Q.

For eigenvector $\mathbf{e}_i$, the new eigenvalue is λ_i', $(1/N+1)[N\mathbf{Q} + (c\mathbf{e}_i)(c\mathbf{e}_i)^T]\mathbf{e}_i = \lambda_i'\mathbf{e}_i$.

$$\frac{N}{N+1}\mathbf{Q}\mathbf{e}_i + \frac{1}{N+1}c^2\mathbf{e}_i(\mathbf{e}_i^T\mathbf{e}_i) = \lambda_i'\mathbf{e}_i$$
$$\frac{N}{N+1}\lambda_i\mathbf{e}_i + \frac{1}{N+1}c^2\mathbf{e}_i = \lambda_i'\mathbf{e}_i, \quad \mathbf{e}_i^T\mathbf{e}_i = 1$$
$$\left(\frac{N}{N+1}\lambda_i + \frac{1}{N+1}c^2\right)\mathbf{e}_i = \lambda_i'\mathbf{e}_i.$$

New eigenvalue $\frac{N}{N+1}\lambda_i + \frac{1}{N+1}c^2 = \lambda_i'.$ (1)

From (1)

$$\lambda_i + \frac{1}{N+1}(c^2 - \lambda_i) = \lambda_i'. \quad (2)$$

From (2), if $c^2 > \lambda_i$, then $\lambda_i' > \lambda_i$. Otherwise $\lambda_i' < \lambda_i$ or $\lambda_i' = \lambda_i$. This proves 1). For eigenvector $\mathbf{e}_j(j \neq i)$, the new eigenvalue is λ_j', $(1/N+1)[N\mathbf{Q} + (c\mathbf{e}_i)(c\mathbf{e}_i)^T]\mathbf{e}_j = \lambda_j'\mathbf{e}_j(\mathbf{e}_i^T\mathbf{e}_j = 0)$.

$$\frac{N}{N+1}\mathbf{Q}\mathbf{e}_j = \lambda_j'\mathbf{e}_j, \quad \frac{N}{N+1}\lambda_j\mathbf{e}_j = \lambda_j'\mathbf{e}_j, \quad \frac{N}{N+1}\lambda_j = \lambda_j'. \quad (3)$$

From (1) and (3), the new λ_i' is enhanced more than the new $\lambda_j'(j \neq i)$. This proves 2).

Property 1: Given a covariance matrix Q computed from a set of N data, if we add these same set of N data to compute the eigenvalues and eigenvectors, the new ones are not changed.

Proof: The new covariance matrix is $\mathbf{Q}' = (1/N+N)[N\mathbf{Q}+N\mathbf{Q}] = \mathbf{Q}$, the same as the original covariance matrix Q, so new eigenvalues and eigenvectors are not changed.

Example 1: Given the samples: $\bar{X}_1 = [2, 2]^T$, $\bar{X}_2 = [-1, 1]^T$, $\bar{X}_3 = [-2, -2]^T$, $\bar{X}_4 = [1, -1]^T$.

Mean vector:
$$\bar{M} = E[\bar{X}] = \tfrac{1}{4}(\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \bar{X}_4) = \vec{0}.$$
Covariance matrix:
$$\mathbf{Q} = E[(\bar{X} - \bar{M})(\bar{X} - \bar{M})^T] = E[\bar{X}\bar{X}^T]$$
$$= \tfrac{1}{4}(\bar{X}_1\bar{X}_1^T + \bar{X}_2\bar{X}_2^T + \bar{X}_3\bar{X}_3^T + \bar{X}_4\bar{X}_4^T) = \begin{bmatrix} 2.5 & 1.5 \\ 1.5 & 2.5 \end{bmatrix}$$
$$\lambda_1 = 4, \quad \mathbf{e}_1 = \left[1/\sqrt{2}, 1/\sqrt{2}\right],$$
$$\lambda_2 = 1, \quad \mathbf{e}_2 = \left[-1/\sqrt{2}, 1/\sqrt{2}\right]^T.$$

TABLE I
EXPERIMENTAL RESULTS OF FIGS. 3–18 FOR DIFFERENT LEARNING RATES $\gamma(t)$. SYMBOLS: DIMENSION OF INPUT VECTORS (#1), NUMBER OF INPUT VECTORS (#2), NUMBER OF STEPS TO CONVERGENCE (#3), MSE, AND NMSE

Convergence steps #3, MSE, NMSE	Learning rate $\gamma(t)$ is a decreasing function	Learning rate $\gamma(t)$ is a constant 0.045
Figure 3. #1=2. #2=128. σ^2=.00182582	#3= 469 MSE=.00000136 NMSE=.00074489	#3= 697 MSE=.00000170 NMSE=.00093188
Figure 4. #1=2. #2=128. σ^2=.00509162	#3= 237 MSE=.00000011 NMSE=.00002089	#3= 432 MSE=.00000021 NMSE=.00004053
Figure 5. #1=3. #2=128. σ^2=.001168	#3= 811 MSE=.000004662 NMSE=.003989943	#3=1573 MSE=.000001545 NMSE=.001322128
Figure 6. #1=128. #2=3. σ^2=.001168	#3= 654 MSE=.00001847 NMSE=.01580673	#3=1104 MSE=.00001356 NMSE=.01160910
Figure 7. #1=3. #2=128. σ^2=.000961	#3= 864 MSE=.000006476 NMSE=.006740044	#3=3216 MSE=.000000281 NMSE=.000292096
Figure 8. #1=128. #2=3. σ^2=.000961	#3= 815 MSE=.00002695 NMSE=.02804647	#3= 1710 MSE=.00000679 NMSE=.00706588
Figure 9. #1=32. #2=128. σ^2=.004551	#3= 253 MSE=.002908 NMSE=.639098	#3= 445 MSE=.002908 NMSE=.639040
Figure 10. #1=128. #2=32. σ^2=.004551	#3= 556 MSE=.002675 NMSE=.587913	#3= 854 MSE=.002675 NMSE=.587908
Figure 11. #1=128. #2=28. σ^2=.0040	#3= 273 MSE=.0026 NMSE=.6549	#3= 435 MSE=.0026 NMSE=.6548
Figure 12. #1=128. #2=32. σ^2=.0042	#3= 110 MSE=.0027 NMSE=.6522	#3= 189 MSE=.0027 NMSE=.6520
Figure 13. #1=64. #2=512. σ^2=.0044	#3= 432 MSE=.0035 NMSE=.8017	#3= 768 MSE=.0035 NMSE=.8013
Figure 14. #1=512. #2=64. σ^2=.0044	#3= 255 MSE=.0034 NMSE=.7773	#3= 394 MSE=.0034 NMSE=.7768
Figure 15. #1=64. #2=512. σ^2=.004438	#3= 439 MSE=.003502 NMSE=.789089	#3= 432 MSE=.003501 NMSE=.788915
Figure 16. #1=512. #2=64. σ^2=.004438	#3= 237 MSE=.003511 NMSE=.791054	#3= 322 MSE=.003504 NMSE=.789480
Figure 17. #1=64. #2=512. σ^2=.000651	#3= 70 MSE=.000212 NMSE=.325803	#3= 127 MSE=.000212 NMSE=.325218
Figure 18. #1=512. #2=64. σ^2=.000651	#3= 80 MSE=.000166 NMSE=.255123	#3= 137 MSE=.000165 NMSE=.253560

Example 2: Adding two extra samples, $\bar{X}_5 = [5, -5]^T$, $\bar{X}_6 = [-5, 5]^T$, on Example 1 along the direction of the second eigenvector $\mathbf{e}_2$

Mean vector:
$$\bar{M} = E[\bar{X}] = \tfrac{1}{6}(\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \bar{X}_4 + \bar{X}_5 + \bar{X}_6) = \vec{0}.$$
Covariance matrix:
$$\mathbf{Q} = [(\bar{X} - \bar{M})(\bar{X} - \bar{M})^T]$$
$$= \tfrac{1}{6}(\bar{X}_1\bar{X}_1^T + \bar{X}_2\bar{X}_2^T + \bar{X}_3\bar{X}_3^T + \bar{X}_4\bar{X}_4^T$$
$$+ \bar{X}_5\bar{X}_5^T + \bar{X}_6\bar{X}_6^T)$$
$$= \begin{bmatrix} 10 & -22/3 \\ -22/3 & 10 \end{bmatrix},$$
$$\lambda_1 = 17.3348, \quad \mathbf{e}_1 = \left[-1/\sqrt{2}, 1/\sqrt{2}\right]^T,$$
$$\lambda_2 = 2.6652, \quad \mathbf{e}_2 = \left[1/\sqrt{2}, 1/\sqrt{2}\right]^T.$$

The second eigenvector on Example 1 becomes the first eigenvector on Example 2.

Using Theorem 1, if we add more and more points along the direction of some existing eigenvector, that eigenvector may become the first principal eigenvector. The seismic example of the Theorem 1 is shown in the following experiment of fault seismogram using vertical seismic traces as the inputs. This theorem can improve the seismic interpretation that the seismic data with uniform or consistent property can enhance the principal eigenvectors.
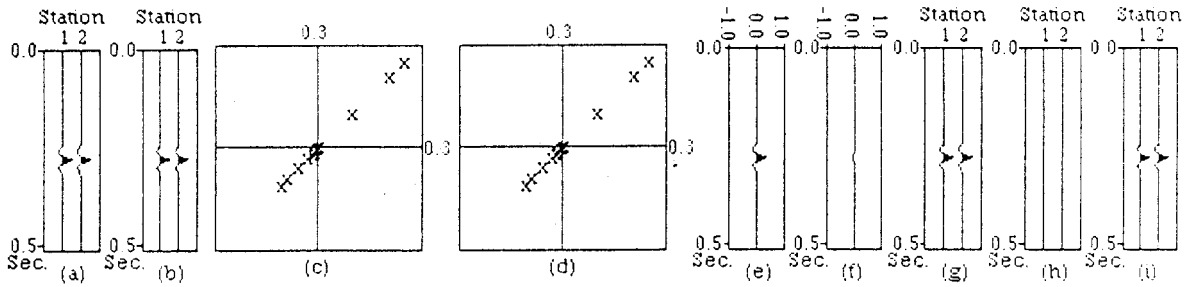
Fig. 3. Seismic data (20-Hz Rocker wavelet). (b) Seismic data (horizontal mean vector removed). (c) Distribution from (a). (d) Distribution from (b). (e) and (f) On first and second projection values. (g) and (h) On first and second components. (i) Sum of two components and mean vector.
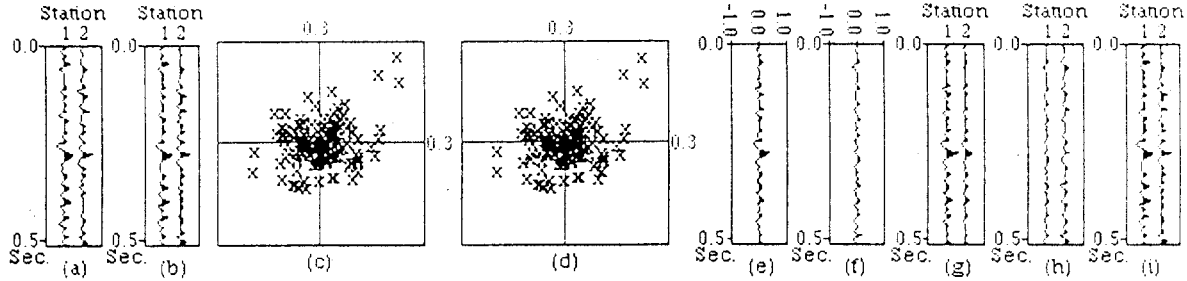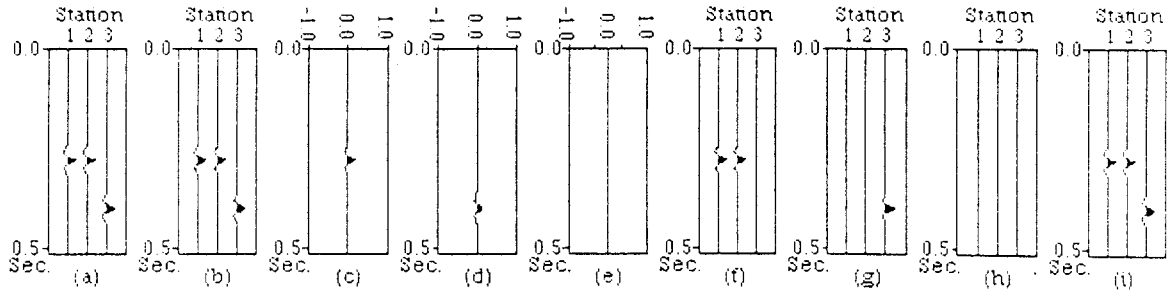

Fig. 4. (a) Seismic data (20-Hz Ricker wavelet). (b) Seismic data (horizontal mean vector removed). (c) Distribution from (a). (d) Distribution from (b). (e) and (f) On first and second projection values. (g) and (h) On first and second components. (i) Sum of two components and mean vector.

Fig. 5. (a) Seismic data (20-Hz Ricker wavelet). (b) Data (horizontal mean vector removed). (c), (d), and (e) On first, second, and third projection values. (f), (g), and (h) On first, second, and third component. (i) Sum of three components and mean vector.

V. PCA IN SEISMIC RICKER WAVELET ANALYSES

We start the basic seismic PCA experiments on the different layers with different Ricker wavelets. Then we extend the PCA applications to simulated seismograms and real seismic data. We list the results for each seismic experiment on Figs. 3–18 in Table I.

A. Two-Dimensional (2-D) PCA of Ricker Wavelets in One Seismic Layer using Input in Horizontal Direction

Fig. 3(a) shows one layer with the 20-Hz zero-phase Ricker wavelets with reflection coefficient -0.25 in two seismic traces. The sampling rate is 0.004 s. Using horizontal input data vector, the data vector form of 20-Hz Ricker wavelets at the layer is $[c, c]^T$ and can be in one line with high correlation in Fig. 3(c), which is the 2-D scatter diagram from Fig. 3(a). Fig. 3(b) is the traces with horizontal mean vector removed. Fig. 3(d) is the 2-D scatter diagram of data from Fig. 3(b). The neural net can find the first eigenvector corresponding to the layer's Ricker wavelets. Fig. 3(e) is the projection value on first eigenvector for each data vector. Fig. 3(f) is the projection

value on the second eigenvector for each data vector. Fig. 3(g) is the first component. Fig. 3(h) is the second component. Fig. 3(i) is the recovered portion from the first and the second components plus mean vector, $\overline{X}' = \sum_{i=1}^{2} a_i e_i + \overline{M}$. The sum of several components plus mean vector in the later analyses of Figs. 4–19 is similar to this explanation. Fig. 4(a) is the addition of 10–56-Hz Gaussian white band noise (mean $= 0$, standard deviation $= 0.1$) to the Fig. 3(a). In Fig. 4, the neural net can also find the first eigenvector.

B. PCA of Two Seismic Layers

1) Input in Horizontal Direction: In Fig. 5(a), the 20-Hz zero-phase Ricker wavelets are on the three traces. One horizontal layer is on traces 1 and 2, the other layer is on trace 3. We use input data in the horizontal direction, so the data are three-dimensional (3-D). The input data vector type at the first layer is $[c, c, 0]^T$, and the input data vector type at the second layer is $[0, 0, c]^T$. Fig. 5(c)–(e) show the three projection values. The projection component on the first eigenvector can recover the first layer in Fig. 5(f). The projection component
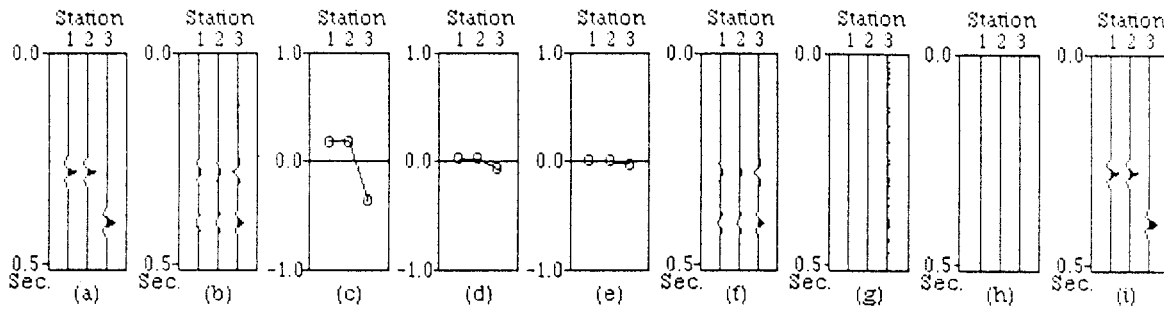
Fig. 6. (a) Seismic data (20-Hz Ricker wavelet). (b) Data (vertical mean vector removed). (c), (d), and (e) On first, second, and third projection values. (f), (g), and (h) On first, second, and third component. (i) Sum of three components and mean vector.
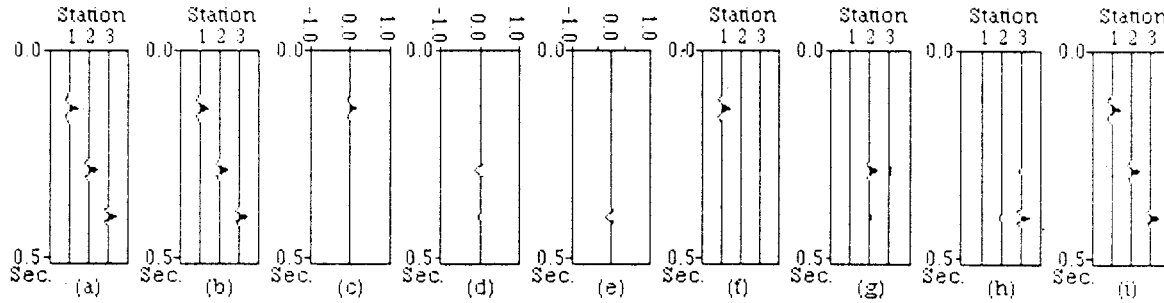

Fig. 7. (a) Seismic data (20-, 25-, and 30-Hz Ricker wavelet). (b) Data (horizontal mean vector removed). (c), (d), and (e) On first, second, and third projection values. (f), (g), and (h) On first, second, and third component. (i) Sum of three components and mean vector.

on the second eigenvector can recover the second layer in Fig. 5(g). From the distances between the origin $[0, 0, 0]^T$ and these 3-D data vectors, the distribution of the data in the first layer is longer than that of the second layer, so the largest component is corresponding to the first layer. Elongated data distribution can generate the first eigenvector.

2) Input in Vertical Direction: Fig. 6(a) is the same signal as Fig. 5(a). We take input data vector from vertical direction. The number of dimensions is 128. One trace is one sample. The samples are too less, and the mean vector is not $\overrightarrow{0}$. Fig. 6(b) is the traces with mean vector removed. After removing the mean vector, the shapes of wavelets are changed. In Fig. 6(i), through PCA and adding mean vector, the wavelets of the layers can be recovered.

C. Analysis of Three Class Ricker Wavelets in Three Seismic Layers, Respectively

1) Input in Horizontal Direction: In Fig. 7(a), the 20-, 25-, and 30-Hz zero-phase Ricker wavelets are on the first, second, and third traces, respectively. They have the same maximum amplitude 0.25. The difference is the duration, i.e., number of points, at three Ricker wavelets. We use input data in the horizontal direction. The input data vector types are $[c, 0, 0]^T$ at the first layer, $[0, d, 0]^T$ at the second layer, and $[0, 0, c]^T$ at the third layer. The first layer has the largest number of points in 3-D space corresponding to 20-Hz Ricker wavelet. Fig. 7(c)–(e) show the three projection values. The largest component is corresponding to the first layer with 20-Hz Ricker wavelet. Fig. 7(f) shows that the projection component on the first eigenvector can recover the first-layer 20-Hz Ricker wavelet. Fig. 7(g) shows that the projection component on the second eigenvector can recover the second-layer 25-Hz Ricker

wavelet. Fig. 7(h) shows that the projection component on the third eigenvector can recover the third-layer 30-Hz Ricker wavelet. So if the layer with Ricker wavelet in the same amplitude range can contribute more points, the direction of the data is more significant, and the eigenvector can dominate. We can use the above Theorem 1 in this interpretation.

2) Input in Vertical Direction: Fig. 8(a) is the same signal as Fig. 7(a). We take input data from the vertical direction. The mean vector is not $\overline{0}$. Fig. 8(b) is the traces with mean vector removed. After removing the mean vector, the wavelets are changed. In Fig. 8(i), through PCA and adding mean vector, the wavelets can be recovered.

VI. PCA IN SIMULATED AND REAL SEISMOGRAMS

We apply the PCA based on the neural net to the analysis of simulated and real seismograms. Finally, we use the traditional power numerical method of PCA in the simulated bright spot seismogram for comparison of the performance.

A. PCA of a Fault

1) Input in Horizontal Direction: The seismogram in Fig. 9(a) shows a fault. The left-hand side has 24 uniform traces, and the right-hand side has eight uniform traces. The seismic trace has the 20-Hz zero-phase Ricker wavelet with reflection coefficient -0.2, 4-ms sampling interval, and 10–56-Hz Gaussian white band noise (mean = 0, standard deviation = 0.1). Using the input data vector in the horizontal direction, the data type of the layer on the left side is $[c, c, c, \cdots, 0, 0, \cdots, 0]^T$ and the data type of the layer on the right side is $[0, 0, \cdots, c, c, \cdots, c]^T$. The two sides of a fault are corresponding to two principal eigenvectors.
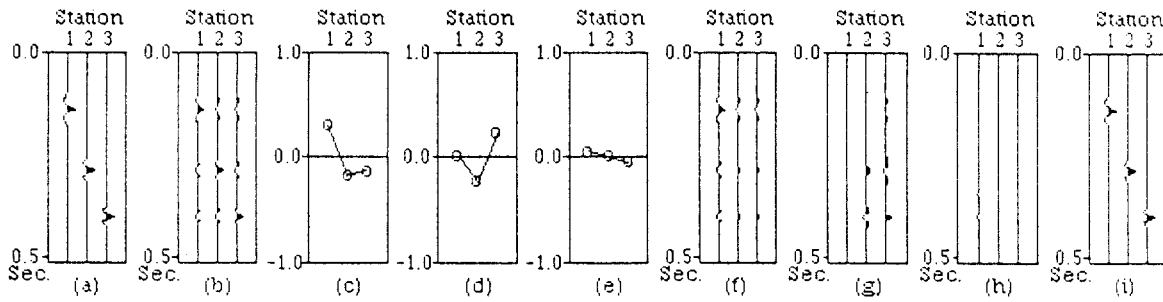
Fig. 8. (a) Seismic data (20-, 25-, and 30-Hz Ricker wavelet). (b) Data (vertical mean vector removed). (c), (d), and (e) On first, second, and third projection values. (f), (g), and (h) On first, second, and third component. (i) Sum of three components and mean vector.
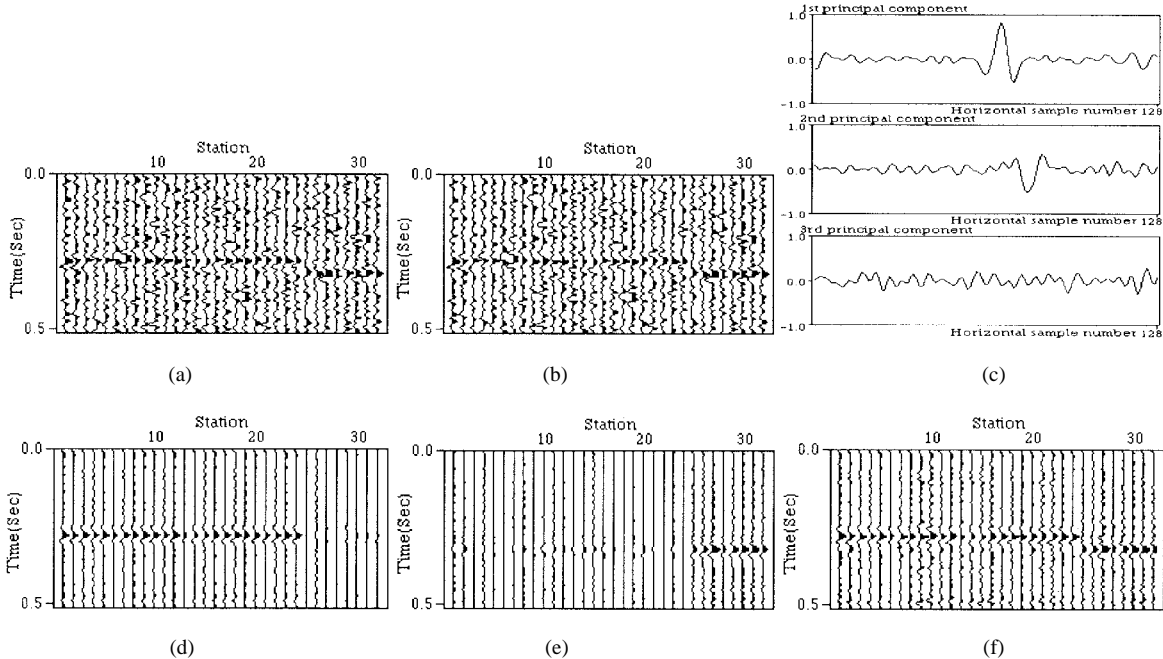

Fig. 9. (a) A fault. (b) Data (horizontal mean vector removed). (c) Three projection values. (d) First component. (e) Second component. (f) Sum of three components and mean vector.

Fig. 9(b) is the data with mean vector removed. Fig. 9(c) shows the projection values of each horizontal data vector on three eigenvectors. Fig. 9(d) shows the projection component of each horizontal data vector on the first eigenvector corresponding to the first layer. Fig. 9(e) shows the projection component of each horizontal data vector on the second eigenvector corresponding to the second layer. Fig. 9(f) shows the summation of three principal components and mean vector. The two sides of a fault are corresponding to two principal eigenvectors. The 24 uniform wavelets on the left-hand side reflect the largest projection value on the first eigenvector. And the eight uniform wavelets on the right-hand side reflect the largest (negative) projection value on the second eigenvector. The reason is the same as the above experiment (Section V-B). Because the distances between data vectors $[c, c, c, \cdots, 0, 0, \cdots, 0]^T$ and the origin are longer than the distances between data vectors $[0, 0, \cdots, c, c, \cdots, c]^T$ and the origin in horizontal data vector.

2) Input in Vertical Direction: We take input from the vertical direction. One trace is one sample. The mean vector is not

$\bar{0}$. Fig. 10(a) is the traces with mean vector removed. After removing the mean vector, the wavelets on the first layer are changed. In Fig. 10(e), through PCA and adding mean vector, the wavelets can be recovered.

3) Input in Vertical Direction in Another Fault Seismogram: Another fault seismogram is shown in Fig. 11(a). There are 28 traces. Traces #19–24 (six traces) show the left side of the fault, and traces #25–28 (four traces) show the right side of the fault. Using input from the vertical direction, one trace is one sample and the mean vector is approximate to $\bar{0}$. Fig. 11(b) is the traces with mean vector removed. After removing the mean vector, the shapes of the wavelets on the first and second layers are kept. The first eigenvector is on the direction of the traces #19–24, and the second eigenvector is on the direction of the traces #25–28. In Fig. 11(c), the uniform property on the first and second layers is shown from the projection values on the first and second eigenvectors.

4) Input Extra Vertical Seismic Traces to Show the Seismic Example of the Theorem 1: Fig. 12(a) is the seismogram with four extra traces on Fig. 11(a). The four extra traces #29–32
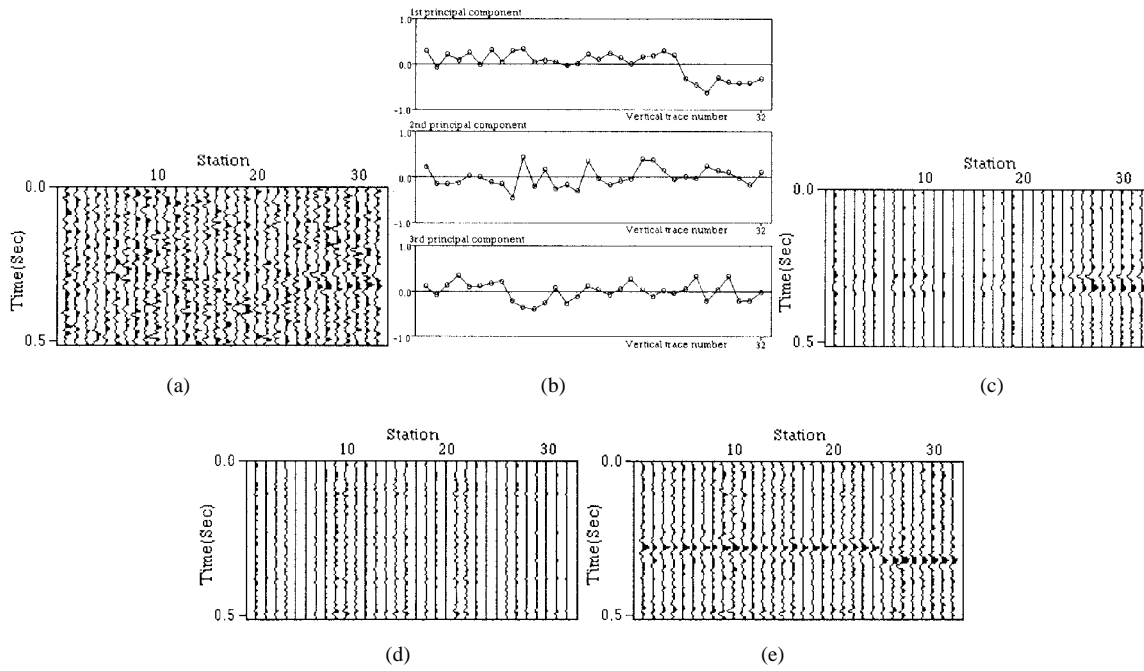
Fig. 10. (a) Data (vertical mean vector removed). (b) Three projection values. (c) First component. (d) Second component. (e) Sum of three components and mean vector.
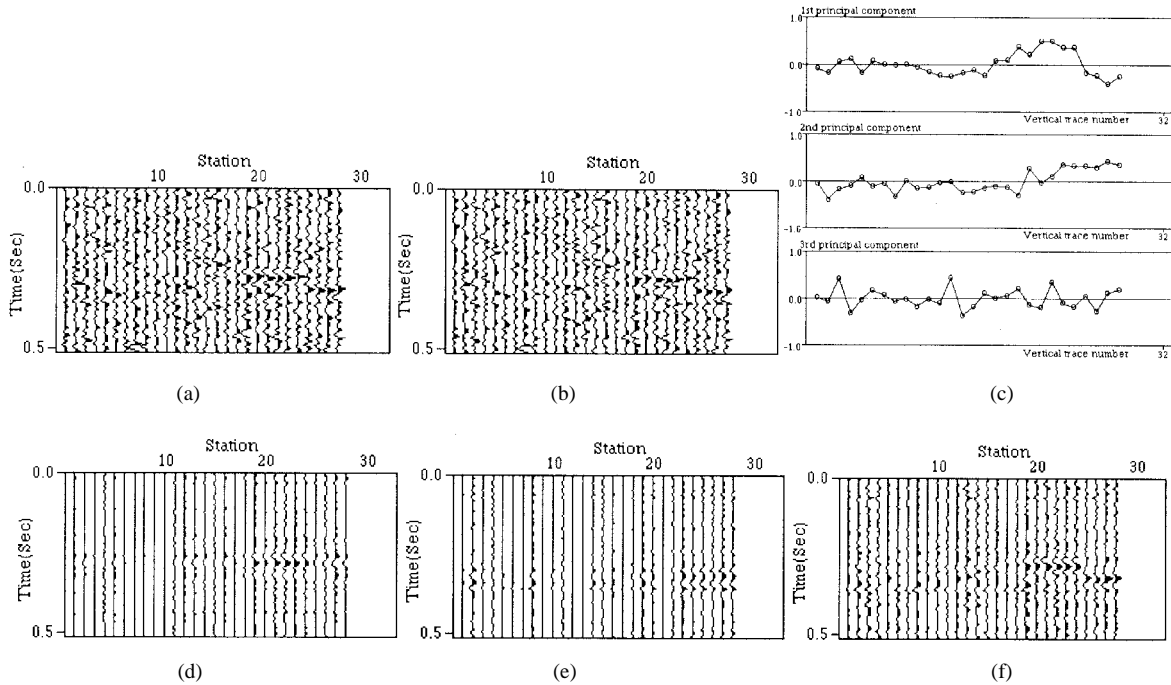

Fig. 11. (a) A fault. (b) Data (vertical mean vector removed). (c) Three projection values. (d) First component. (e) Second component. (f) Sum of three components and mean vector.

have the same uniform layer property with the right side traces #25–28. It means that four extra samples are added along the direction of the second eigenvector in Fig. 11. Because the number of uniform traces is increased, the second eigenvector in Fig. 11 becomes the first eigenvector in Fig. 12. This shows the property of the Theorem 1. We can infer that the more uniform samples can enhance the principal eigenvector.

B. PCA for Filtering of Diffraction Pattern

The simulated horizontal geological layer with termination is shown in Fig. 13(a). The depth of the layer is 500 m, the seismic P-wave velocity is 2500 m/s, and the receiving station interval is 50 m. The generated seismogram after NMO has reflection and diffraction patterns in Fig. 13(b). The source signal is a 20-Hz zero-phase Ricker wavelet with reflection
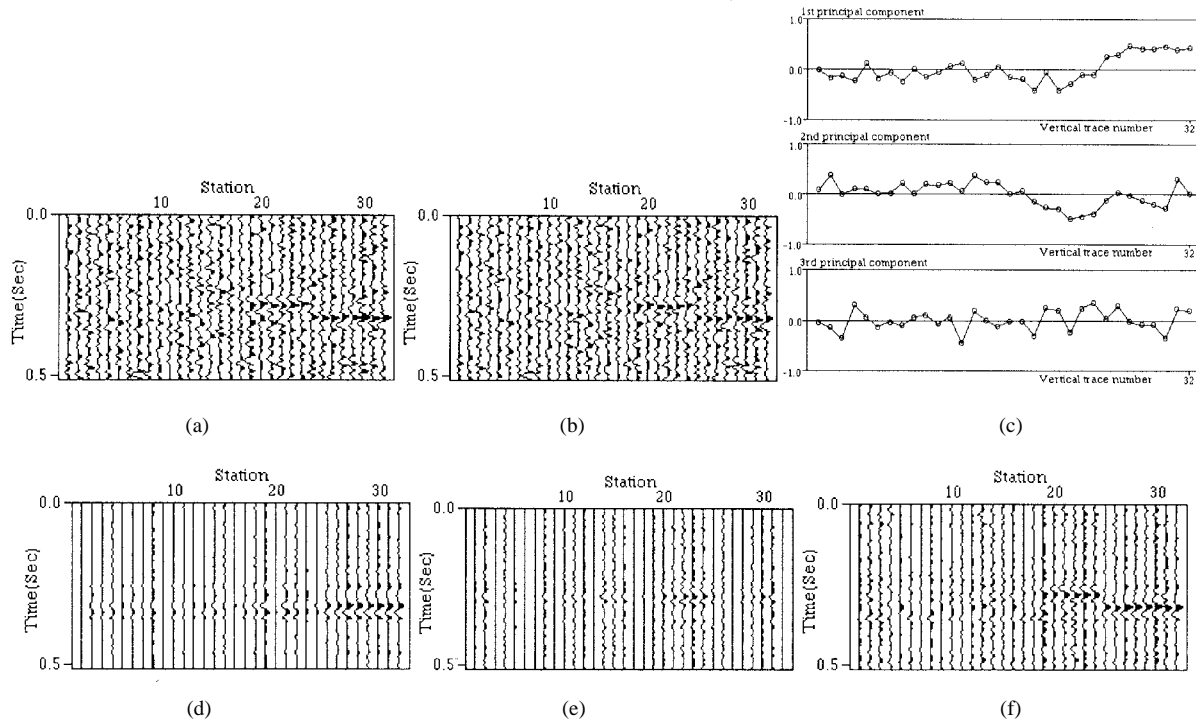
Fig. 12. (a) A fault with extra traces #29–32. (b) Data (vertical mean vector removed). (c) Three projection values. (d) First component. (e) Second component. (f) Sum of three components and mean vector.
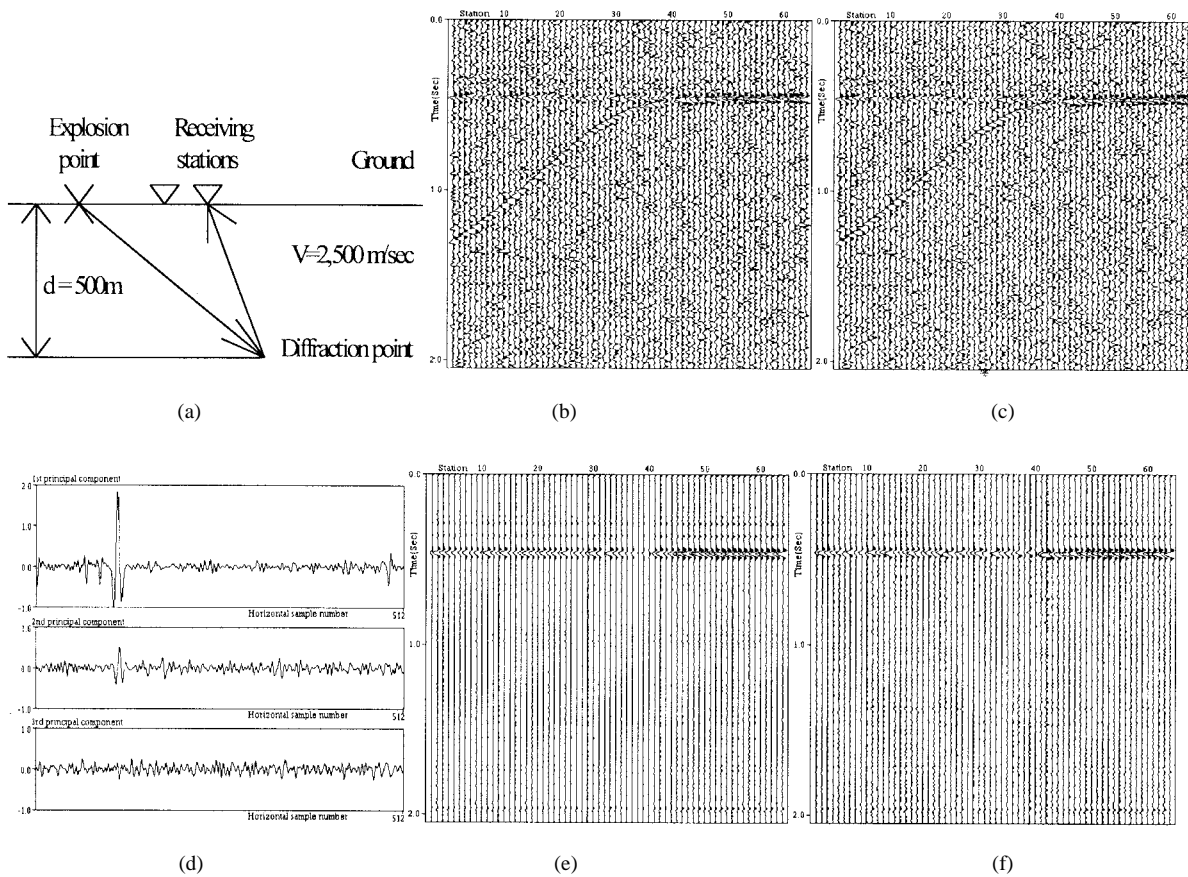

Fig. 13. (a) Geological model. (b) Reflection and diffraction patterns. (c) Data (horizontal mean vector removed). (d) Three projection values. (e) First component. (f) Sum of three components and mean vector.
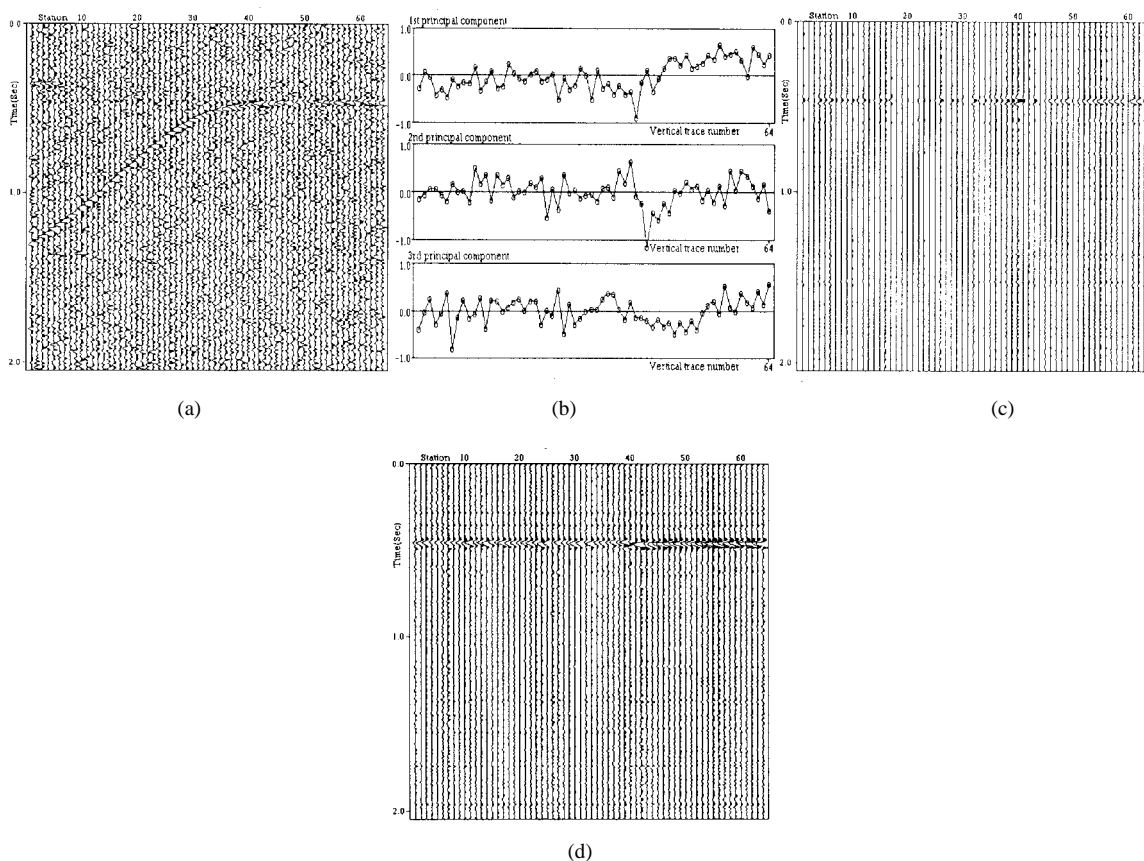
Fig. 14.   (a) Data (vertical mean vector removed). (b) Three projection values. (c) First component. (d) Sum of three components and mean vector.

coefficient 0.2. The Gaussian white band 10–56-Hz noises are added to the seismogram.

1) Input in Horizontal Direction: Using the input data vector in the horizontal direction, the data type is $[c, c, \cdots, c, 2c, 2c, \cdots, 2c]^T$ in the whole uniform layer. Fig. 13(e) shows that the diffraction seismic pattern is filtered from horizontal reflection layer in the projection component of each horizontal input data vector on the first eigenvector.

2) Input in Vertical Direction: We take input from the vertical direction. One trace is one sample. The mean vector is not $\bar{0}$. Fig. 14(a) is the traces with mean vector removed. The shapes of the Ricker wavelets on the right-hand side of the layer are kept, but amplitudes are reduced. The projection value on the first eigenvector in Fig. 14(b) shows the uniform property on the right-hand side of the traces. In Fig. 14(c), the diffraction pattern is filtered using one principal component.

C. PCA in Bright Spots Seismogram

1) Input in Horizontal Direction: Fig. 15(a) shows that there is a structure of bright spots that indicate the gas and oil sand zones with large (negative) reflection coefficient −0.29 at the top of the gas sand zone [21]. The seismogram has 64 traces with 512 data per trace. Using the input data vector in the horizontal direction to compute the five principal eigenvectors of the covariance matrix, Fig. 15(c) shows that the projection value of each horizontal vector on the first eigenvector can show the high amplitude content and polarity

reversal of the wavelets. Fig. 15(d) shows the projection of each horizontal data vector on the first eigenvector. The central part of the first layer at time 0.3 s and the central flat part of the bright spot structure, totally four layers, appear in the projection component on the first eigenvector. But the whole first layer at time 0.3 s is not corresponding to the first eigenvector. The reason is that the number of data vectors in the central flat three layers of the bright spot structure is three times more than that in the one layer at 0.3 s. Fig. 15(i) shows the recovered seismogram using five principal components. Because there are nonflat layers in the bright spot pattern, we need to choose more principal components in the recovered seismogram.

2) Input in Vertical Direction: For the input data vector in the vertical direction, one trace is one sample. In the first principal projection values of Fig. 16(b), the traces in the central part show the uniform property and the neighboring traces at two sides show the polarity reverse property.

D. PCA in Real Seismogram at Mississippi Canyon

1) Input in Horizontal Direction: The real seismogram at Mississippi Canyon is shown in Fig. 17(a). Fig. 17(c) shows the five projection values. The central part reflects strong to the projection on the first eigenvector in Fig. 17(d). The left part reflects strong to the projection on the second eigenvector in Fig. 17(e). The right part reflects strong to the projection on the third eigenvector in Fig. 17(f). One structure can
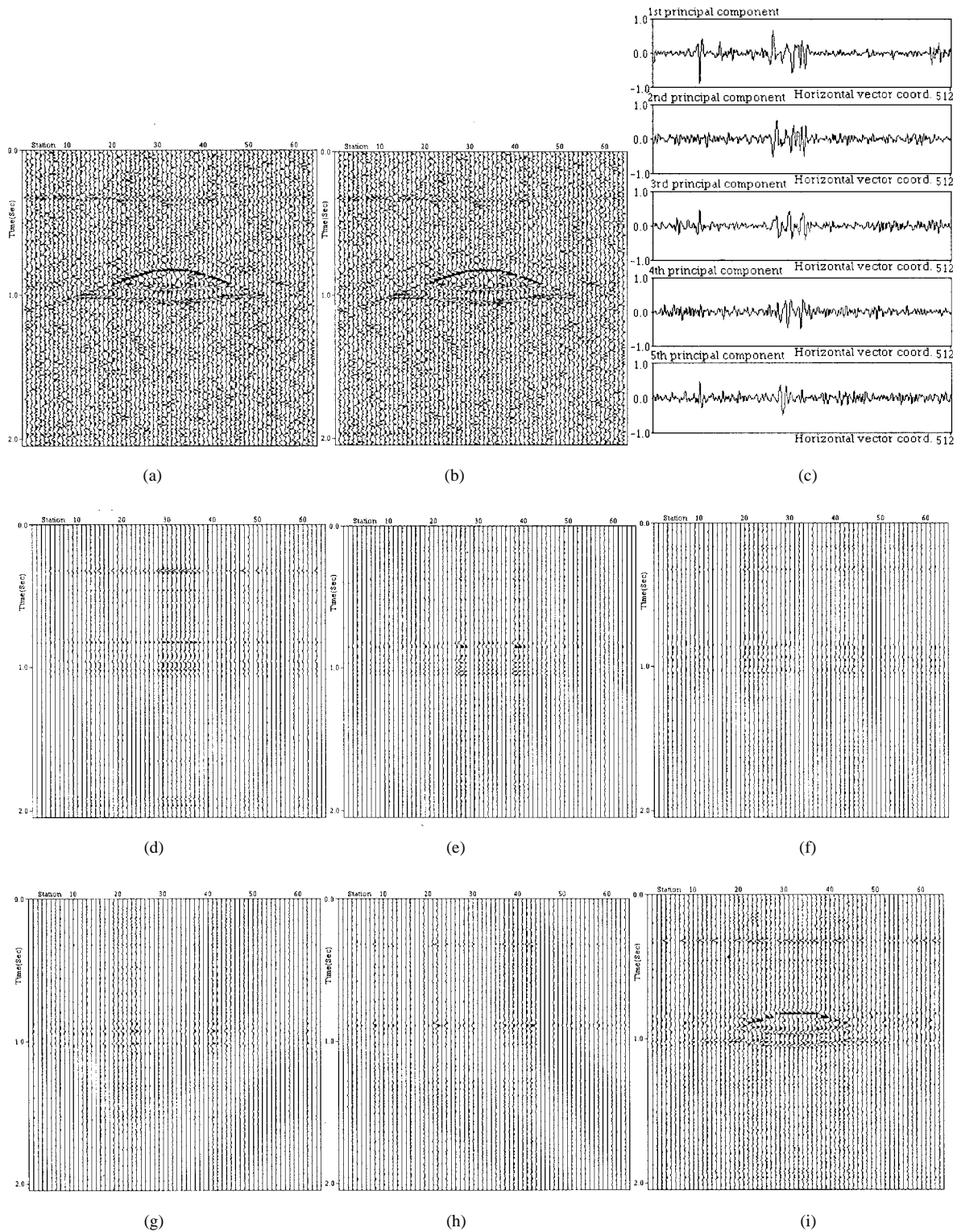
Fig. 15.   (a) Bright spot pattern. (b) Data (horizontal mean vector removed). (c) Five projection values. (d) First component. (e) Second component. (f) Third component. (g) Fourth component. (h) Fifth component. (i) Sum of five components and mean vector.

be decomposed into three significant principal components. Fig. 17(g) shows the recovered data using three principal components.

2) Input in Vertical Direction: We use the input data vector in the vertical direction in Fig. 18(a). Fig. 18(b) shows the three projection values. The first projection value can show the high amplitude and polarity reversal at the central part and both sides. The seismic traces at the central part of the seismogram show the uniform property. Fig. 18(c) shows the projection component of each trace on the first eigenvector. Fig. 18(f) shows the recovered data using three principal components.
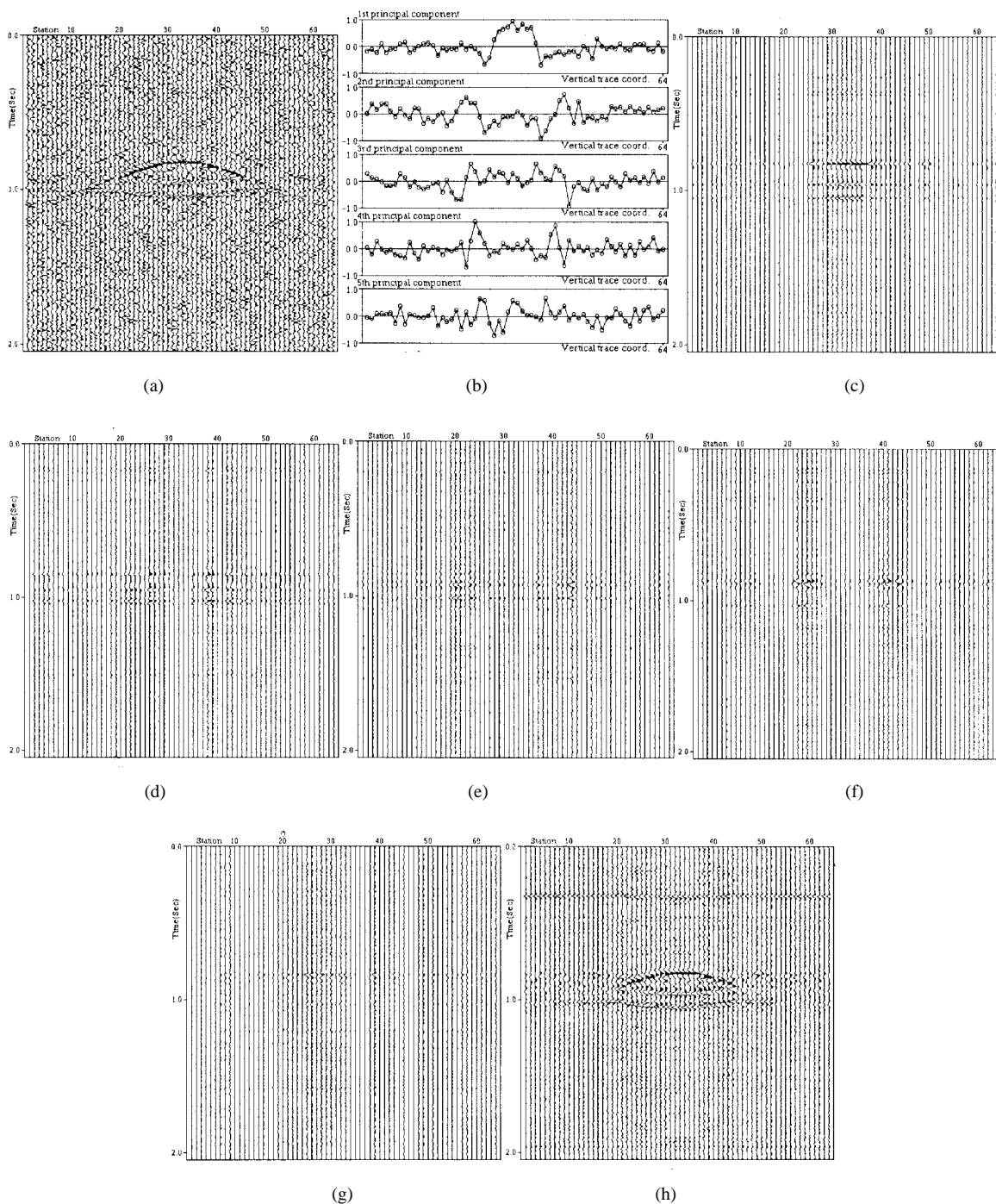
Fig. 16. (a) Data (vertical mean vector removed). (b) Five projection values. (c) First component. (d) Second component. (e) Third component. (f) Fourth component. (g) Fifth component. (h) Sum of five components and mean vector.

E. Traditional PCA in Bright Spots Seismogram

1) Input in Horizontal Direction: We use traditional PCA in the bright spots in Fig. 15(a). The input data vector is in the horizontal direction and the dimension is 64. The first step is to find the covariance matrix 64×64 from the complete data set. The second step is to find the five principal eigenvectors by the traditional power numerical method, which can find the several dominant eigenvectors [22]. The third step is to find the five principal components and the recovered seismogram. Fig. 19 shows the recovered data using five principal components.

In order to avoid the problem of large dimension in DOS-based computer memory, the calculation is one step by one step sequentially. The MSE is 0.003 551, and the NMSE is 0.800 152. They are close to the results of Fig. 15 in Table I.

2) Input in Vertical Direction: We use the input data vector in the vertical direction in Fig. 15(a). However, the dimension 512×512 of the covariance matrix causes the memory problem in the DOS-based computer. So it is not feasible to get the covariance matrix and the eigenvectors.
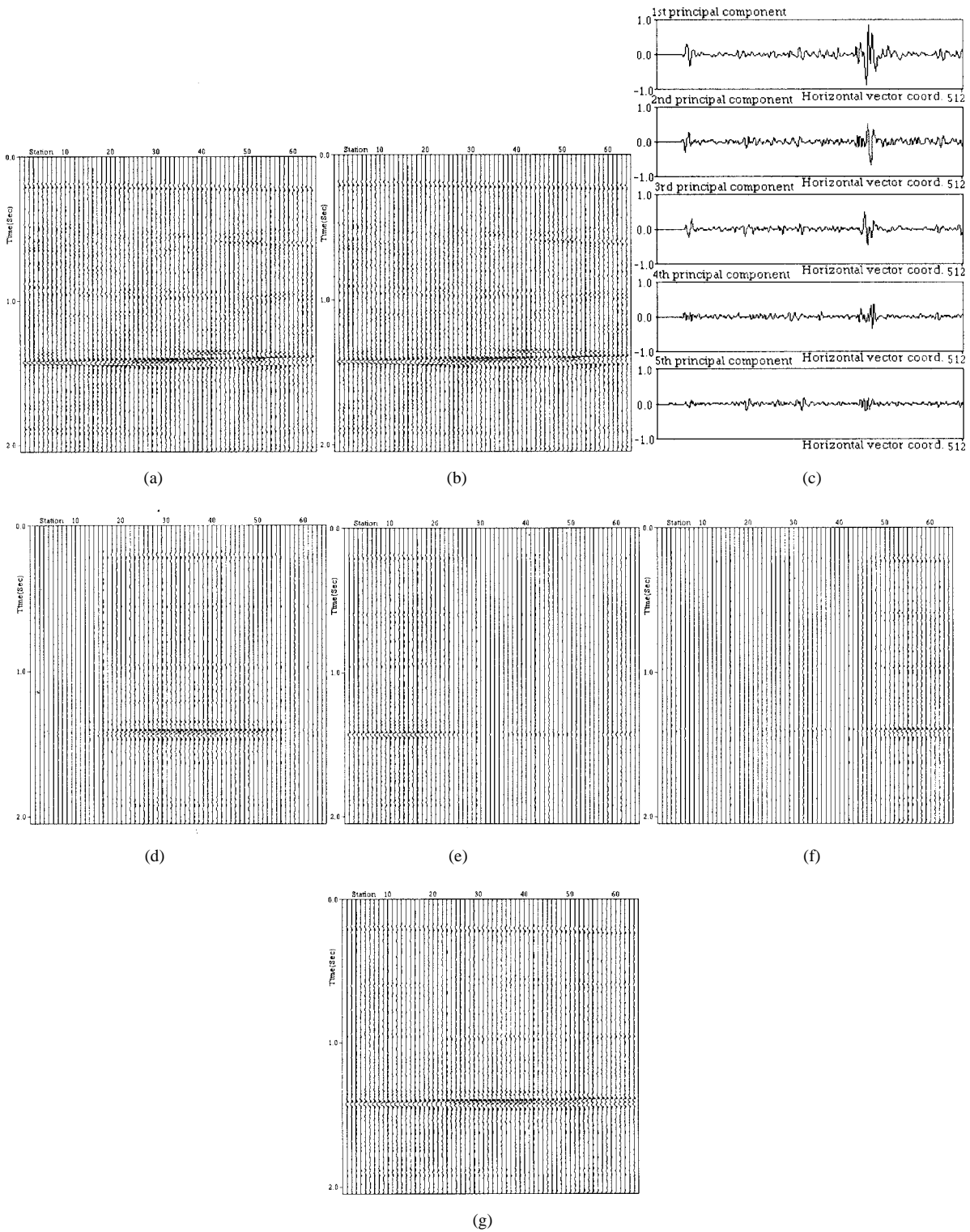
Fig. 17.   (a) Seismogram at Mississippi Canyon. (b) Data (horizontal mean vector removed). (c) Five projection values. (d) First component. (e) Second component. (f) Third component. (g) Sum of three components and mean vector.

VII. Discussions

1) Bright spot seismogram in Fig. 15(a) has magnitude between -0.4036 and 0.4622. The real seismogram in Fig. 17(a) has magnitude between -0.2075 and 0.2316. The dimension of the input data vector can reach to the scale 512. Although the number of seismic data is large and the magnitude of the seismic data is small, the information of the seismic layers and the uniform seismic traces can be extracted by the PCA using GHA.

2) For the input vector with dimension N, the traditional method uses memory space order of N^2 in computation because of the covariance matrix, but the neural net based on Sanger's learning rule uses a memory
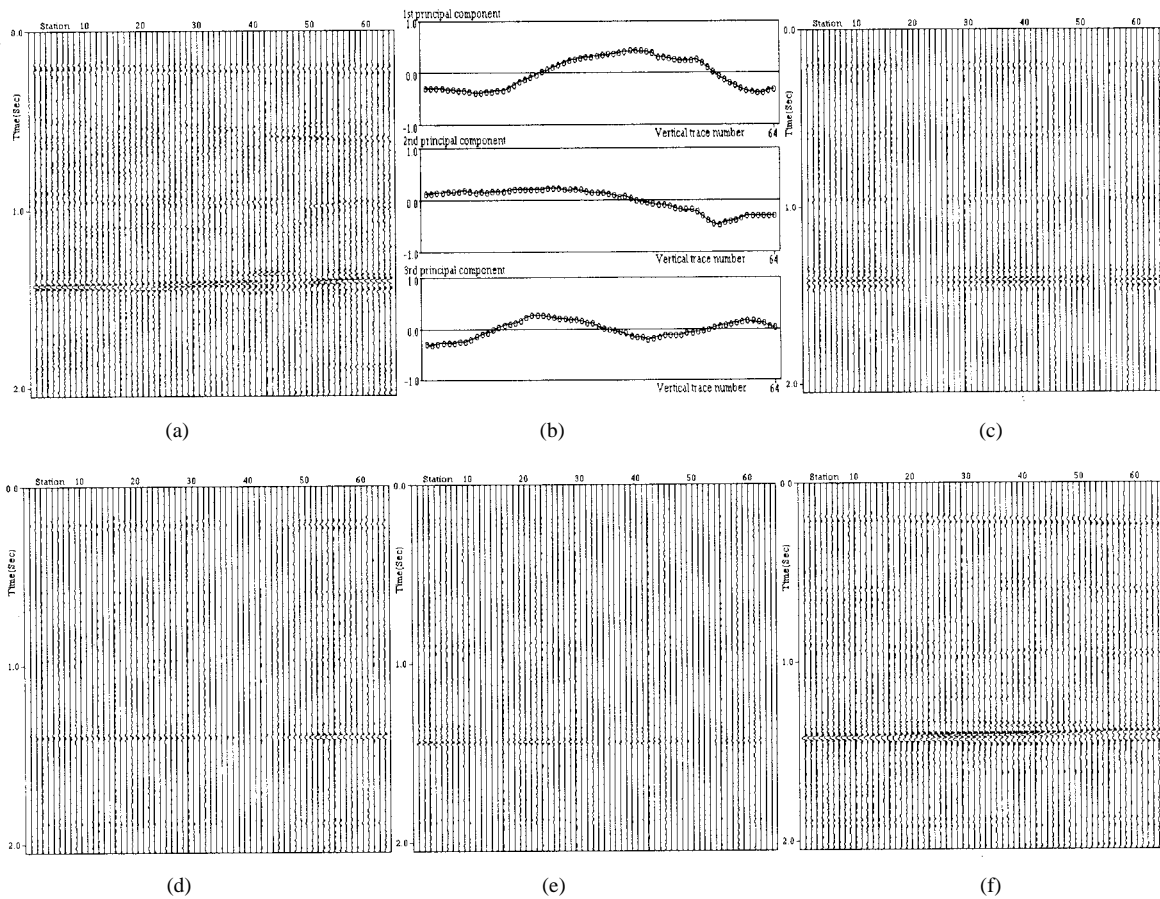
Fig. 18. (a) Data (vertical mean vector removed). (b) Three projection values. (c) First component. (d) Second component. (e) Third component. (f) Sum of three components and mean vector.
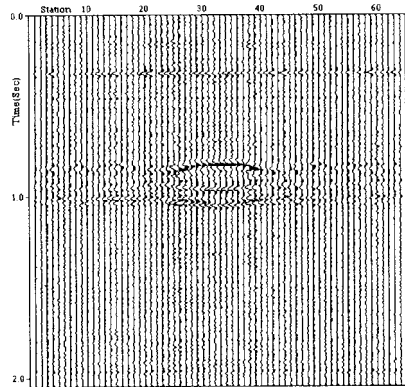

Fig. 19. Input vector in the horizontal direction. Sum of five components and mean vector by the traditional power numerical method.

space order of N to find the eigenvectors. For the input seismic data vector in the vertical direction with 512 dimensions, covariance matrix is 512×512. The memory space of real number is $4 \times 512 \times 512$ bytes $= 1\,048\,576$ bytes and is over the restriction of 64 K bytes of DOS-based computer memory space. In this study, the computation of our experiment cannot work on a DOS-based computer by the traditional method. However, using Sanger's learning rule with feeding data iteratively, we can find the eigenvectors running on Microsoft FORTRAN in the DOS-based computer without the memory space problem. The computation of eigenvectors in seismic data becomes feasible through the neural net based on Sanger's learning rule in a DOS-based computer.

3) Seismic data are compressed significantly. For a seismogram with 64 traces and 512 data per trace, there are $32\,768$ data. If we use input data in the vertical direction to find three principal eigenvectors, each eigenvector has 512 dimensions. Each seismic trace has three projection values, so there are 3×64 data for the projection values of 64 traces. We can store one mean vector (512 dimensions), three eigenvectors with 3×512 data, and projection values of 64 traces. The total stored data are $512 + 3 \times 512 + 3 \times 64 = 2240$. That indicates $32\,768/2240 = 14.6$ fold information redundancy. If we use input data in the horizontal direction to find three principal eigenvectors, each eigenvector has 64 dimensions. Each seismic horizontal data vector has three projection values, so there are 3×512 data for the projection values of 512 vectors. We can store one mean vector (64 dimensions), three eigenvectors with 3×64 data, and projection values of 512 vectors. The total stored data are $64 + 3 \times 64 + 3 \times 512 = 1792$. That indicates $32\,768/1792 = 18.3$ fold information redundancy. If the seismic layers are most flat, the recovered

seismogram is acceptable by using several principal components. From a data representation point of view, the principal components net provides a technique to seismic data compression.

4) Initial setting of coefficients of the weighting matrix W may affect the final direction of the first eigenvector to the positive or negative direction. To overcome this problem, at first, we randomly set the initial weight coefficients W_{ij}. After convergence, we check the first principal projection value. If the projection value is positive, we can accept the set of the eigenvectors. Otherwise, we change the initial weight coefficients until the first principal projection value is positive.

5) In 2-D, if the data distribution is highly correlated along one line, the first eigenvector can be computed from the net, but the remaining second eigenvector cannot be derived to the correct value. And it is the same in N-dimensional case.

6) How many eigenvectors can we choose? In the real seismogram, we can check the first to the fifth principal projection values of each data, then decide how many eigenvectors are still significant, i.e., large projection values. Or, we may project each data on each eigenvector and calculate the variance of the data on each eigenvector. The variances are the eigenvalues, and $\lambda_1 > \lambda_2 > \cdots > \lambda_i > \cdots > \lambda_M$ [1]. And we can choose eigenvectors with larger variances and neglect the eigenvectors with smaller variances.

7) $E[\bar{X}] = \bar{0}$ is easy to get in the horizontal input direction, but not easy to get in the vertical input direction because of the layer effect. In the vertical input direction, if $E[\bar{X}]$ is not equal to $\bar{0}$, the recovered components can not match the original seismogram until mean vector is added.

8) We may subtract the principal components from the original seismogram, then the remaining are the nonuniform property patterns: the slope pattern, the curve pattern like diffraction pattern, and the noise.

VIII. CONCLUSIONS

1) PCA can improve seismic interpretations and show how many uniform classes and where the locations are in the seismogram in both input directions, respectively, which are not easy to identify by human eyes or by experience. The real seismogram at Mississippi Canyon is a good example.

2) For PCA, a theorem is proposed, which states that adding an extra point along the direction of the existing eigenvector can enhance that eigenvector. The theorem is applied to the interpretation of a fault seismogram and the uniform property of other seismograms. The uniform property of seismic data is corresponding to high correlation.

3) Using the input data vector in the horizontal direction, the principal components can show the uniform property of the wavelets from the same horizontal layer. The projection values on the first eigenvector can show

the high amplitude content and polarity reversal for wavelets at different layers. These properties appear in the analyses of the seismogram of bright spot pattern and the real seismogram at Mississippi Canyon. From simulation on the Ricker wavelets, for different layers, the projection on the first eigenvector can also show the large duration (low frequency) content of the wavelet.

4) Using the input data vector in the vertical direction, the principal components can show the uniform property of the neighboring seismic traces. The projection on the first eigenvector can show the uniform and reverse properties of the traces. In the analyses of the seismogram of bright spot pattern and the real seismogram at Mississippi Canyon, the traces in the central part show the uniform property and the neighboring traces at two sides show the polarity reverse property.

5) Diffraction seismic pattern after NMO is separated from horizontal reflection layer using the first principal component for the input data in both horizontal and vertical directions.

6) Two sides of a fault correspond to two principal eigenvectors in both direction analyses.

7) We can use PCA in the seismic data compression.

ACKNOWLEDGMENT

REFERENCES

[1] Y. T. Chien and K. S. Fu, "On the generalized Karhunen–Loeve expansion," *IEEE Trans. Inform. Theory,* vol. IT-13, pp. 518–520, Apr. 1967.

[2] K. Fukunaga and W. L. G. Koontz, "Application of the Karhunen–Loeve expansion to feature selection and ordering," *IEEE Trans. Comput.,* vol. C-19, pp. 311–-318, Jan. 1969.

[3] D. Pelat, "Karhunen–Loeve series expansion: A new approach for studying astrophysical data," *Astron. Astrophys.,* vol. 33, pp. 321–329, 1974.

[4] A. K. Jain, "A fast Karhunen–Loeve transform for a class of random processes," *IEEE Trans. Commun.,* vol. COMM-26, pp. 1023–1029, Sept. 1976.

[5] ———, "A fast Karhunen–Loeve transform for digital restoration of images by white and colored noise," *IEEE Trans. Comput.,* vol. C-26, pp. 560–571, Apr. 1977.

[6] H. Hotelling, "Analysis of a complex statistical variables into principal components," *J. Educ. Psychol.,* vol. 24, pp. 417–441 and pp. 498–520, 1933.

[7] D. C. Hagen, "The application of principal components analysis to seismic data sets," *Geoexploration,* vol. 20, pp. 93–111, 1982.

[8] I. F. Jones, "Applications of Karhunen–Loeve transform in reflection seismology," Ph.D. dissertation, Univ. British Columbia, Vancouver, B.C., Canada, 1985.

[9] E. Oja and J. Karhunen, "Recursive construction of Karhunen–Loeve expansions for pattern recognition purposes," in *Proc. 5th Int. Conf. Pattern Recognit.,* Miami Beach, FL, 1980, pp. 1215–1218.

[10] J. Karhunen and E. Oja, "Optimal adaptive compression for high-dimensional data," in *Proc. 2nd Scand. Conf. Image Anal.,* Helsinki, Finland, 1981, pp. 152–157.

[11] ———, "New methods for stochastic approximation of truncated Karhunen–Loeve expansions," in *Proc. 6th Int. Conf. Pattern Recognit.,* Munich, Germany, 1982, pp. 550–553.

[12] E. Oja, "A simplified neuron model as a principal component analyzer," *J. Math. Biol.,* vol. 15, pp. 267–273, 1982.

[13] ———, "On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix," *J. Math. Anal. Applicat.,* vol. 106, pp. 69–84, 1985.

[14] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network," *Neural Networks,* vol. 2, pp. 459–473, 1989.

[15] ——, "An optimality principle for unsupervised learning," *Advances in Neural Information Processing Systems,* vol. I, D. S. Touretzky, Ed. San Mateo, CA: Morgan Kaufmann, 1989, pp. 11–19.

[16] P. Baldi and K. Hornik, "Neural network and principal component analysis: Learning from examples without local minima," *Neural Networks,* vol. 2, pp. 53–58, 1989.

[17] A. Krogh and J. A. Hertz, "Hebbian learning of principal components," *Parallel Processing in Neural Systems and Computers,* R. Eckmiller, G. Hartmann, and G. Hauske, Eds. Amsterdam, The Netherlands: Elsevier, 1990, pp. 183–186.

[18] S. Y. Kung, *Digital Neural Networks.* Englewood Cliffs, NJ: Prentice-Hall, 1993, ch. 8.

[19] S. Bannour and M. R. Azimi-Sadjadi, "Principal component extraction using recursive least squares learning," *IEEE Trans. Neural Networks,* vol. 6, pp. 457–469, Mar. 1995.

[20] F. Peper and H. Noda, "A symmetric linear neural network that learns principal components and their variances," *IEEE Trans. Neural Networks,* vol. 7, pp. 1042–1047, Sept. 1996.

[21] M. B. Dobrin, *Introduction to Geophysical Prospecting,* 3rd ed. New York: McGraw-Hill, 1976, ch. 10.

[22] B. Carnahan, H. A. Luther, and J. O. Wilkes, *Applied Numerical Methods.* New York: Wiley, 1969.

Kou-Yuan Huang (S'81–M'83–SM'94) received the B.S. degree in physics and the M.S. degree in geophysics from the National Central University, Taiwan, in 1973 and 1977, respectively, and the M.S.E.E. and Ph.D. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, in 1980 and 1983, respectively.

He was a Graduate Research Assistant at Purdue University in 1978. From 1978 to 1979, he was with the Department of Geoscience, Purdue University, where he worked with geophysics. He then joined the School of Electrical and Computer Engineering and the Laboratory for Applications of Remote Sensing (LARS), Purdue University, in 1979. From 1981 to 1983, he was with the Advanced Automation Research Laboratory, Purdue University. From September 1983 to August 1988, he was with the Department of Computer Science, University of Houston, Houston, TX. From August 1992 to July 1993, he was the Visiting Scholar at the University of Texas at Austin for one semester and later at Princeton University, Princeton, NJ. From August 1996 to July 1997, he took his sabbatical at Rice University, Houston, and the University of Houston. He is currently a Professor in the Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan. He has published papers in numerous journals, including *Geophysics, Geoexploration, Pattern Recognition,* IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, etc. His major contributions are in the areas of seismic pattern recognition using image processing, statistical, syntactic, neural networks, and fuzzy logic methods. His research interests also include character recognition and real-time updating of video servers.