Using the Principal Component Analysis to Detect the Effective Information in Noise

Tenglong Wang[1,a] , Xiang-e Sun[2,b] *

[1,2] School of Electronics and Information, Yangtze University , Jingzhou ,Hubei , China

[a]wangtenglong99@163.com, [b]286693443@qq.com

*Corresponding Author: Xiang-e Sun, Professor, School of Electronics and Information, Yangtze University, Jingzhou, Hubei.

Keywords: PCA noisy signals the effective information two-dimensional seismic reflection signal

Abstract. Faced with the massive information and data, signal and information processing need to solve the problem of removing redundant information and obtain useful information. The PCA method is introduced, which can detect the effective information in noise. Taking the simulated seismic data as targets, the principle component is used to analysis the seismic data in different speeds and slopes in the first place . And then using the principle component to recover the original signals is studied. After the signal simulation experiment, the principle component effectiveness is verified further after comparing the original signals and the recovered signals not adding noise.

Introduction

As signals are interfered by a variety of noise in the transmission, it need to find a good method to detect the effective signal from the noise. The PCA method is able to solve this puzzle. As the multi-dimensional characteristics can be compressed into a few dimensions by the PCA method[2,5]., the PCA method has good ability of reducing dimensions. Besides, the PCA method can get the important elements and structures from the large effective information, and remove the redundance and noise in data. In addition, it can reduce dimensions in the original complex data and reveal simple structure hidden behind the complicated data.

Principle of PCA algorithm

The PCA method is the process of mapping the data space to low-dimensional subspace by orthogonal transformation. The corresponding vector group should meet orthogonality , and low-dimensional subspace made up of basis vector group should consider data's correlation optimally[1,3] . Besides, the mutual correlation of single data sample should be reduced to the lowest point. Following is the steps of the PCA method.

(1) The original data should be standardized treatment

$$x_{ij}^{'} = (x_{ij} - x_i) / d_i (i = 1,2, \mathbf{L}\ p.j = 1,2, \mathbf{L}\ n) \qquad (1)$$

x_{ij} means the original data of the i-th index and the j-th sample , x_i and d_i is the sample mean and standard deviation of the i-th index respectively.

(2) According to a standardized data sheet($x_{ij}^{'}$)$_{p \times n}$, calculate the correlation coefficient matrix

$$R = (r_{ij})_{p \times n}$$

$$r_{ij} = \frac{1}{n} \sum_{k=1}^{n} (x_{ki} - x_i)(x_{kj} - x_j) / d_i d_j \qquad (2)$$

(3) Calculate R eigenvalues and eigenvectors。 According to the characteristic equation $| R - 1i | = 0$, calculate characteristic root $1i$, and make the characteristic root range from small

to big, $l_1 \geq l_2 \geq \mathbf{L} \, l_p$, get the corresponding eigenvectors $u_1, u_2, \mathbf{L} \, u_p$. and these eigenvectors are standard orthogonal vectors.

(4) Calculate the contribution rate $e_i = l_i \, / \, \sum_{i=1}^{p} l_i$, and add up the contribution rate

$$E_m = \sum_{j=1}^{m} l_j \, / \, \sum_{i=1}^{p} l_i \qquad (3)$$

(5) Select the number of principle components. It mainly balances the number and the accumulative contribution rate . and the aim is to use less number of the principle components to obtain more original information[4,6].

Seismic data principle component analysis

The relationship of each principle component in different speeds. Set depth fixed h=1500m, take four different speeds V1=2000m/s, V2=4000m/s, V3=8000m/s, V4=16000m/s respectively , time space graph and each eigenvalue size are shown in Fig.1. With the increase of the reflected wave velocity , the number of the effective eigenvalues which the time-distance curve corresponds to the matrix decreased.
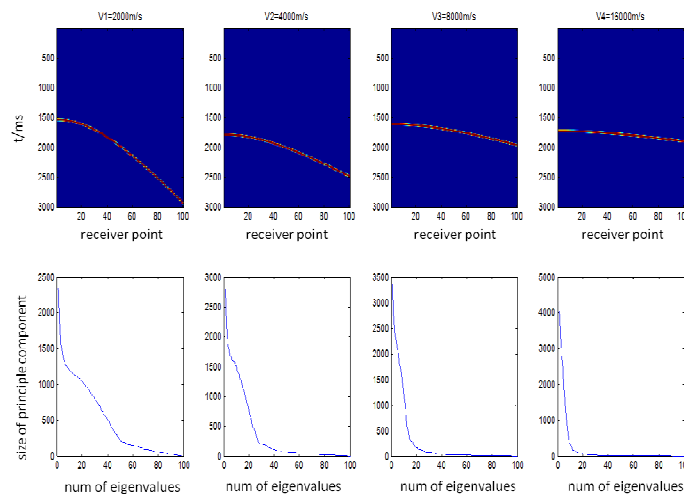

Fig.1 time space graph in different speeds and its eigenvalues

Relationship of each principle component in different slopes. It is supposed that the time-distance curves of the received signals are all straight line, while their slopes are different, then their relationship of eigenvalues are shown in Fig. 2 . With the increase of slope, the number of effective eigenvalues increases, and the value of the principle component decreases accordingly. For the straight line in different slopes , Fig.2 shows that the number of effective eigenvalues would decrease if the slope decreases and the previous eigenvalues would increase.
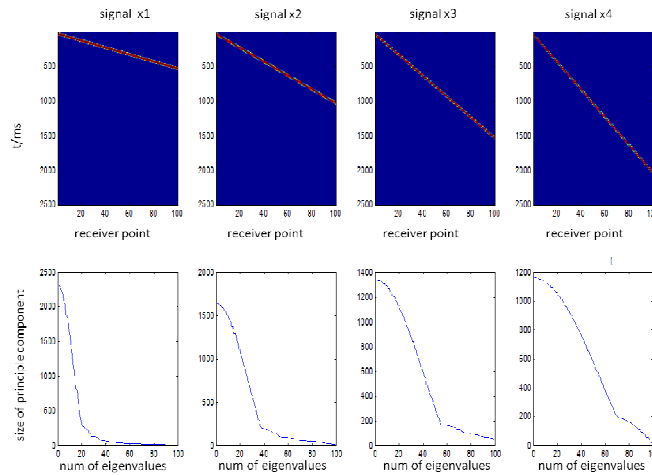
Fig.2 Each principle compoment's relationship graph in different slopes

Recovery of seismic data information

Recovery of the reflected wave information in no noise. The value of the first principle component is relatively small when time space graph is curve. While the slope of straight line turn zero, the value of the first principle component is very large, and it covers all the signal information[7]. Therefore the time-distance curve of the reflected wave can be corrected in a straight line by the method of time and speed scanning gradually. If the common reflection point gather can be corrected in a straight line, the wave of each gather has no phase difference and the stack wave energy will strengthen. Original signal, signal processed by NMO and the recovered signal are shown in Fig.3.
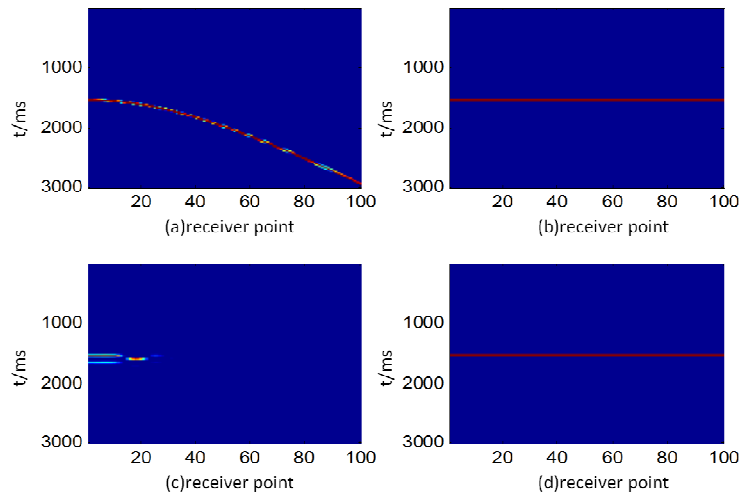

Fig. 3 original signal, signal processed by NMO and the recovered signal

Fig.3(a) is the original signal, Fig.3(c) shows the signal that the original signal is directly recovered by the first principle component, obviously the original signal can't be recovered by the first principle component. Fig.3(b) is the signal processed by NMO, Fig.3(d) is the signal that the signal processed by NMO is recovered by the first principle component. Apparently Fig.3(b) and Fig.3(d) are approximately the same. The following conclusion can be got after comparing the four figure, the speed of seismic wave is received, and the original signal can be extracted by the PCA method

Recovery of the effective signal in different noise. In order to analysis PCA decomposition characteristics, Gaussian noise, salt and pepper noise , random noise and uniform noise are added to the original signal after NMO. Four kinds of noise are corresponding to Fig.4 (a), Fig.4 (b), Fig.4 (c) and Fig.4 (d) respectively. Then the PCA method is used to recover the signal after NMO from the different noisy signals. The recovery signals shows in Fig.4 (e), Fig.4 (f), Fig.4 (g) and Fig.4 (h), which

are corresponding to Fig.4 (a), Fig.4 (b), Fig.4 (c) and Fig.4 (d) respectively. Obviously signals can be recovered from different kinds of noise by the PCA method.
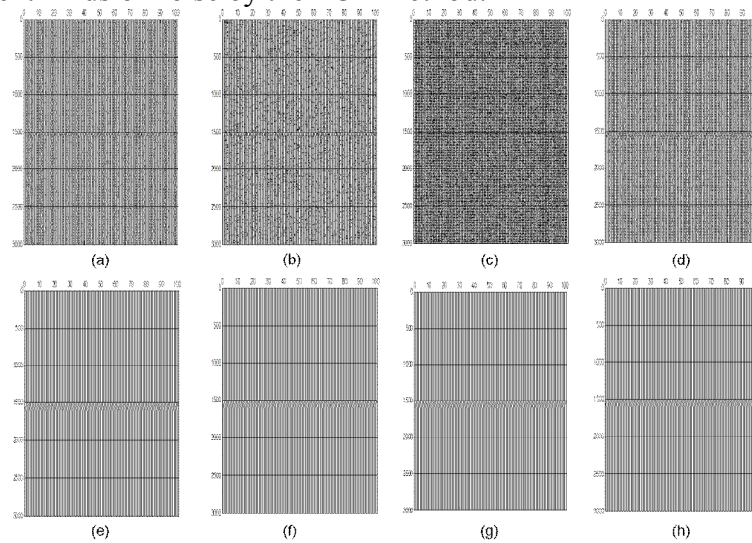

Fig. 4 adding noise to original signal after NMO and signal recovery by PCA
(a)adding Gaussian noise that the variance is 0.04 to original signal after NMO (b) adding salt and pepper noise that the density is 0.04 to original signal after NMO (c)adding random noise that the SNR is 0.1 to original signal after NMO (d)adding uniform nise to original signal after NMO (e) signal recovers from the Gaussian noise (f) signal recovers from the salt and pepper noise (g) signal recovers from the random noise (h) signal recovers from the uniform noise

Conclusions

 As the first principle component contains the max variance of the original data, the effective signal can be recovered by using the first principle component from different kinds of noise. Apparently the PCA method has great superiority and reliability in detecting the effective information in noise. Although the PCA method has been widely used , it still has a large space for development. The practical problems is diverse and complicated, so it is a good way to solve problems by combining the PCA method.

References

[1] Yin Yu, Advanced engineering mathematics (Third Edition)［M］, Huazhong University of Science and Technology Press ,2001, p.196-205 in Chinese.

[2] Haiqing Wang, Zhihuan Song, Ping Li, Study on the fault detectability of principle component analysis［J］, Chinese Journal of Scientific Instrument, 2002, 23(31): 12-15.

[3] Wei Wang, Ming Lin, Qinzhong Ma, The primary component analysis method and its application in earthquake prediction［J］, Earthquake Research in China, 2005, 21(3): 409-416.

[4] Xiang-e Sun, High resolution stacking velocity analysis based on the principle component analysis ［J］, Journal of Oil and Gas Technology, 2006, 28(4): 85-88.

[5] Zhaoling Wang, Wan Lei ,Hongliang Sun, Tao Zhong ,Application of seismic reflection wave method in exploration of national key project［J］, Site Investigation Science and Tecnology, 2007, (3): 62-65.

[6] Xiang-e Sun, Study on some problems in obtaining seismic wave velocity［D］, Chengdu , Chengdu University of Technology, 2007.

 [7] Xuan Cui, Liang Wu, Applications of principal component analysis to geosciences review and project 〔J〕, Journal of the Graduates Sun YAT-SEN University(Natural Sciences, Medicine), 2009, 30(4): 41-47.