Weighted Iteration Optimization of PCA Algorithm in the Identification of Seismic Prospecting Signal

Qiao Wang

*College of Earth Science, Jilin University,
Changchun 130061, Liaoning, China*

Yue Liu

*Institute of Information Engineering, Kaifeng University,
Kaifeng 475004 Henan, China*

Abstract
In light of low accuracy of standard PCA algorithm in seismic prospecting signal identification, this paper proposed an improved model that optimizes the PCA algorithm by weighted iteration of kernel function. Firstly, it takes radial basis function as the kernel function to build the test model, realizes the transformation from input space to characteristic space through non-linear mapping, conducts the linear PCA weighted iteration on the mapping data so as to improve the non-linear processing ability of PCA algorithm, and finally constructs the model based on the optimized PCA algorithm by kernel function. Simulation experiments show that the improved PCA algorithm has better performance than standard PCA algorithm and wavelet transform algorithm in seismic prospecting signal identification.
Key words: PCA ALGORITHM, SEISMIC PROSPECTING, SIGNAL IDENTIFICATION, WEIGHTED ITERAITON, KERNEL FUNCTION

1. Introduction

With the rapid growth of the national economy, the demand for energy is increasing. Oil not only has an important position as energy consumption in the national economy, but also occupies an important position in the national development strategy and the political and economic security. Therefore, the prospecting must be intensified to increase the oil and gas reserves. But with the deep degree of prospecting, it becomes more difficult to easily find a relatively simple type of oil and gas reservoir, and the main focus of the work is transferred to the west with complex surface geological structure in order to increase oil and gas reserves and improve the yield. The exploration targets are more complex with exploration domain extending from the original structural pools to subtle reservoirs [1]. The accuracy of current seismic acquisition, processing and structural interpretation techniques cannot meet the needs of increasingly complex exploration targets, especially the seismic signal processing techniques.

In order to meet the requirement of petroleum and mineral exploration and development, the seismic prospecting of our country further develop three technologies: 3D seismic exploration technology, three high processing technology and reservoir prediction technique [2]. Weak signal detection is a newly-developing science and technology, which uses modern electronics and signal processing methods to extract useful signal from noise. It mainly analyzes the causes and laws of the noise, studies the characteristics and the coherence of the measured signal and detects the weak signal submersed by background noise. Weak signal detection and estimation has made a lot of achievements since the research from Johnson according to the noise generated by random thermal motion of the electron [3]. Many scientists make important contributions to the signal detection, especially over the past 40 years when the limit of the measurement is continuously below the level of noise. As a result, weak signal embodied by the micro phenomenon or weak interaction, which was immeasurable in past, is now possible. The weak signal detection in the signal processing have been developed in the electronic radar, space technology and other fields, including the weak signal detection using the chaotic oscillator system [4], weak signal extraction based on correlation [5], independent component analysis method [6], weak signal extraction based on wavelet transform [7] and so on. Currently weak signal detection, the de-nosing and amplitude preservation has made some progress in the field of earth physics. Yang et al. proposed the wavelet decomposition of seismic signal which separated the strong background signal from weak oil and gas signal, and then conducted spectrum analysis on the basis of the effective combination of weak signal to finally get the easily explanatory results [8]. Wang Zhenglei decomposed the weak signal by wavelet analysis and designed the threshold value of each sub band independently to realize the purpose of extracting weak signal from strong noise signal [9]. Li et al. started from the quasi periodic or periodic signal under the background of strong random noise in chaotic oscillator system which is sensitive to the large-scale periodic phase response, and tested the detectability of chaotic oscillator system by using simulation experiments to explore the common-shot point seismic data in weak effective seismic signal detection [10]. However, most of the weak signal detection is not related to high density, and the majority of domestic and foreign articles focus on the field acquisition.

In view of the defect of PCA algorithm in seismic prospecting signal identification, this paper proposes a signal identification model based on PCA algorithm optimized by weighted iterative kernel function, and conducts the simulation experiments to prove the effective of this improved strategy.

2. Noisy seismic prospecting signal identification model based on PCA algorithm

2.1. Noisy seismic signal model

Firstly, the equation of seismic signal with noise is formed.

$$x(t) + kx(t) - x(t) + x^3(t) = a\cos\omega t \qquad (1)$$

where k is damping ratio, a is periodic driving force amplitude, $-x + x^3$ is non-linear elastic force, and ω is periodic driving force. This equation can be adjusted to first order differential systems.

$$\begin{cases} x = y \\ y = x - x^3 + a\cos\omega t - ky \end{cases} \qquad (2)$$

Because the frequency of the periodic signal to be detected is varying in the seismic signal detection, the equation of noisy seismic signal requires to be improved to measure the periodic signal of arbitrary frequency. The improved equation about noisy seismic signal is,

$$\ddot{x} + k\dot{x} - x + x^3 = \gamma\cos\omega t \qquad (3)$$

To set $t = \omega\lambda$, then $x(t) = x(\omega\lambda)$, and

$$x(t) = x(\omega\lambda) \qquad (4)$$

$$\begin{aligned} \dot{x}(t) &= \frac{dx(t)}{dt} \\ &= \frac{dx(\omega\lambda)}{d(\omega\lambda)} \\ &= \frac{1}{\omega}\dot{x}(\omega\lambda) \end{aligned} \qquad (5)$$

$$\begin{aligned} \dot{x}(t) &= d\left[\frac{1}{\omega}\cdot\dot{x}(\omega\lambda)\right] / d(\omega\lambda) \\ &= \frac{1}{\omega^2}\ddot{x}(\omega\lambda) \end{aligned} \qquad (6)$$

Adding equation (4) and (5) to the equation of noisy seismic signal, then we get

$$\frac{1}{\omega^2}\ddot{x}(\omega\lambda) + \frac{k}{\omega}\dot{x}(\omega\lambda) - x(\omega\lambda) + x^3(\omega\lambda) = \gamma\cos(\omega\lambda) \qquad (7)$$

It is a systematic equation with the independent variable λ which can be transformed to,

$$\ddot{x} = -\omega k\dot{x} + \omega^2(x - x^3 + \gamma\cos\omega\lambda) \qquad (8)$$

From the derivation result, it can be seen that the threshold of this equation has no relation to parameters of the system except ω . When the frequency of the detected signal changes, only ω requires to be adjusted.

The boundary of the periodic orbital of the noisy seismic signal is rough. Here, using $\Delta x(t)$ to represent the minor distribution of noise to $x(t)$, the differential equation of the system with the existence of noise is obtained,

$$(\ddot{x}+\Delta\ddot{x})+k(\dot{x}+\Delta\dot{x})-(x+\Delta x)^3+(x+\Delta x)^5=\gamma\cos(\omega t)+n(t) \quad (9)$$

where $n(t)$ is noise, $E\{n(t)\}=0$. Because Δx is so small that we can ignore the higher order term, equation (9) subtracts equation (3) to get,

$$\Delta\ddot{x}+k\Delta\dot{x}+3x^2\Delta x-5x^4\Delta x=n(t) \quad (10)$$

And $c(t)=5x^4-3x^2$,

$$\Delta\ddot{x}+k\Delta x-c(t)\Delta x=n(t) \quad (11)$$

It can be written into the form of vector differential equation,

$$\dot{X}(t)=A(t)X(t)+N(t) \quad (12)$$

Here,

$$\begin{cases} X(t)=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\begin{bmatrix} \Delta x(t) \\ \Delta\dot{x}(t) \end{bmatrix} \\ A(t)=\begin{bmatrix} 0 & 1 \\ c(t) & -k \end{bmatrix} \\ N(t)=\begin{bmatrix} 0 \\ n(t) \end{bmatrix} \end{cases} \quad (13)$$

The solution is $X(t)=\Phi(t,t_0)X_0+\int_{t_0}^{t}\Phi(t,u)N(u)du$. Φ is the state-transition matrix of the system. The first term is transient state solution which will decay to zero quickly, therefore only the second term is considered.

$$\begin{cases} X(t)=\int_{t_0}^{t}\Phi(t,u)N(u)du \\ E\{X(t)\}==\int_{t_0}^{t}\Phi(t,u)E\{N(u)\}du=0 \end{cases} \quad (14)$$

In the sense of statistics, any zero-mean noise cannot change the original moving trajectory of the system. It only makes the trajectory rougher and oscillating around the ideal trajectory.

2.2. Seismic signal identification based on PCA algorithm

Principal component analysis (PCA) is to select several main indexes from the original indexes to represent the main information.

If there is n samples, each sample has p variables, and then $n\times p$ sample data matrix is formed,

$$X=\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (15)$$

The i column vector in X can be expressed as $X_i(i=1,2,...,p)$, then p vectors of matrix X, $X_1,X_2,...X_p$ is combined linearly with the formula as follows,

$$F_i=a_{1i}X_1+a_{2i}X_2+...+a_{pi}X_p, i=1,2,...,p \quad (16)$$

Setting $a_i=(a_{1i},a_{2i},...,a_{pi})^T$, the combination coefficient a_i is required due to constraints,

$$a_i^T=1, i=1,2,...,p \quad (17)$$

Namely, a_i is the unit vector and determined by following rules.

(1) $F_j(i\neq j, i=1,2,...,p)$ is irrelevant, namely

$$\operatorname{cov}(F_i,F_j)=0 \quad (18)$$

(2) If F_1 represents the item with maximum variance among all the linear combination of $X_1,X_2,...X_p$, then F_1 is called as the first principal component. F_2 represents the item with maximum variance among all the linear combination of $X_1,X_2,...X_p$ and F_2 is irrelevant with F_1, then F_2 is called as the second principal component. By this way, p principal components can be determined.

Suppose the quantity of information of p principal components decreases in sequence and cam be expressed by variance. The variance contribution of each p is equal to the eigenvalue λ_i of original index. The eigenvector of eigenvalue λ_i can be expressed by combination coefficient a_i, then the expression of variance contribution ratio of F_i is $\lambda_i/\sum_{i=1}^{p}\lambda_i$. The higher the value of this expression is, the more amount of information the principal component contains.

Taking the seismic signal identification for example, its operation can be formulated as follows. There are M seismic signal, $i_1,i_2,...,i_M$, and the position of mean seismic signal is defined as,

$$\bar{i}=\frac{1}{M}\sum_{j=1}^{M}i_j \quad (19)$$

Each seismic signal is different with mean vector and has the difference $\phi_n=i_n-\bar{i}$. The covariance matrix $C=\sum_{j=1}^{M}\phi_j\phi_j^T$, then the eigenvector v_k, eigenvalue λ_k and covariance matrix C are calculated. Eigenvector determines the linear space with ϕ characteristic signals of M different images.

$$b_l=\sum_{k=1}^{M}v_{lk}\phi_k \quad (20)$$

In these characteristic signals, $K(K<M)$ signals are identical to the K maximum eigenvalues. The seismic signals are mapped into subspace, namely projected to characteristic signals with following transformation.

$$\omega_{nk}=b_k(i_n-\bar{i}) \quad (21)$$

Where projection coefficient is obtained from seismic signal vector $\Omega_n = [\omega_{n1}, \omega_{n2}, ..., \omega_{nk}]$ presented by each characteristic signal. When only the PCA method is used for signal identification, a new signal is required to project the sub-space generated by the characteristic signals of the signals sample and compare their position in the sub-space and known signal position in image.

PCA method belongs to algebraic feature analysis method, a traditional feature extraction and descending dimension method in pattern recognition. However, PCA is linear mapping method meaning the expression after descending dimension is generated linearly, which ignores the higher than second order mutual relation between data. Therefore, the feature it extracts is not the optimum. In addition, classifier is still considered in PCA method and right classification design can effectively improve the identification result and identification rate. The essence of PCA determines the feature is not the optimum classification feature but description feature, which goes against classification matching.

3. Weighted iteration optimization of PCA algorithm

3.1 Weighted iteration optimization based on kernel function

Polynomial kernel function, Gaussian radial basis function and Sigmoid function are three common functions. Other kernel functions include exponential type radial basis function, Fourier series, B-spline kernel function and tensor product kernel function.

(1)Polynomial kernel function

$$K(x, x_i) = \left[(x \cdot x_i) + 1\right]^q \quad (22)$$

q is the order of the polynomial.

(2)Gaussian radial basis function

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{\sigma^2}\right) \quad (23)$$

(3)Sigmoid function

$$K(x, x_i) = \tanh\left[v_k (x \cdot x_i) + c_i\right] \quad (24)$$

(4)Exponential type radial basis function

$$K(x, x') = \exp\left(-\frac{\|x - x'\|}{2\sigma^2}\right) \quad (25)$$

where x, x' input vectors.

(5)Fourier series

$$K(x, x') = \frac{\sin(N + \frac{1}{2})(x - x')}{\sin(\frac{1}{2}(x - x'))} \quad (26)$$

(6)B-spline kernel function

$$K(x, x') = B_{2N+1}(x - x') \quad (27)$$

(7)Tensor product kernel function

$$K(x, x') = \prod_i K_i(x_i, x_i') \quad (28)$$

We take the radial base function as kernel function to build test model. (If meeting Mercer theorem, also named as Gaussian kernel, shortly RBF). RBF has the form,

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{\sigma^2}\right) \quad (29)$$

In the classification analysis, the final decision function is solved based on RBF,

$$\begin{aligned} M(x) &= \operatorname{sgn}\left(\sum_{SVM} a_i y_i (x, x_i) + b\right) \\ &= \operatorname{sgn}\left(\sum_{SVM} a_i y_i \exp\left(-\frac{|x - x_i|^2}{\sigma^2}\right) + b\right) \end{aligned} \quad (30)$$

x_i is the sample factor vector of the support vector; x is the vector of predictor; a_i and b are the parameters to be determined; r is the kernel function. The sum operation works only on the support vector.

The PCA algorithm based on kernel function with weighted iteration optimization realizes the transformation from input space to characteristic space through non-linear mapping ϕ. The data after mapping is treated with linear PCA. Therefore, the improved algorithm has strong non-linear processing ability.

For the M samples x_k, $x_k \in R^N$ in input space, $\sum_{k=1}^{M} x_k = 0$, the covariance matrix is,

$$C = \frac{1}{M} \sum_{k=1}^{M} x_j x_j^T \quad (31)$$

Common PCA methods solve the characteristic function,

$$\lambda v = Cv \quad (32)$$

to get the eigenvalue with higher contribution and corresponding eigenvector. Here, the non-linear mapping function is introduced to transform the sample points $x_1, x_2, ..., x_M$ in input space into the points $\phi(x_1), \phi(x_2), ..., \phi(x_M)$ in characteristic space, and suppose that,

$$\sum_{k=1}^{M} \phi(x_k) = 0 \quad (33)$$

The covariance matrix in characteristic space F is,

$$\bar{C} = \frac{1}{M} \sum_{j=1}^{M} \phi(x_j) \phi(x_j)^T \quad (34)$$

Therefore, the PCA in characteristic space is to solve the eigenvalue λ and eigenvector $v \in F / \{0\}$ from equation,

$$\lambda v = \bar{C} v \quad (35)$$

Then there is

$$\lambda(\phi(x_k) v) = \phi(x_k) \bar{C} v \quad (36)$$

It is noticed that v can be linearly expressed by $\phi(x_i)$, namely,

$$v = \sum_{j=1}^{M} a_i(x_i) \quad (37)$$

$M \times M$ matrix K is defined,

$$K_{ij} = \phi(x_i)\phi(x_j) \quad (38)$$

To simplify it,

$$M\lambda K\alpha = K^2\alpha \quad (39)$$

Evidently,

$$M\lambda\alpha = K\alpha \quad (40)$$

Through the solution to these expressions, the required eigenvalue and eigenvector is obtained. The projection of sample in vector V^k in F space is,

$$(V^k\phi(x)) = \sum_{i=1}^{M} \alpha_i^k \phi(x_i)\phi(x) \quad (41)$$

At this time, the K in equation (39) is replaced with $\tilde{K}$.

3.2 Seismic prospecting signal identification based on improved PCA algorithm

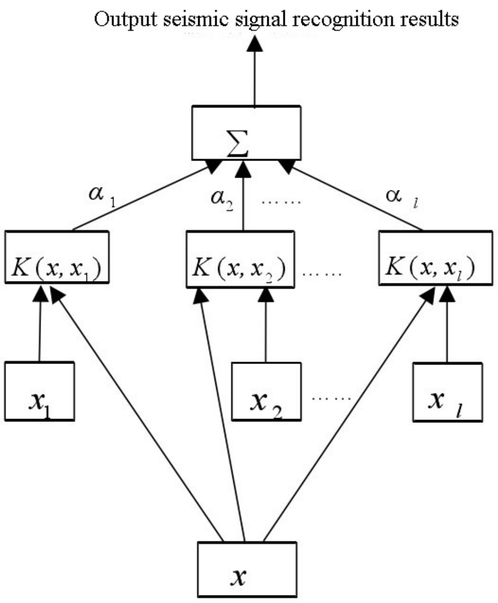

Figure 1. Seismic signal recognition model flow based on improved PCA algorithm

To summarize the discussion above, the seismic prospecting signal identification model based on improved PCA algorithm can be divided into three steps.

(1)To select kernel function $K(x, y)$, centralize it in high-dimensional space, and then calculate matrix $\tilde{K}$;

(2)To get the eigenvalue and eigenvector and normalize it in high-dimensional space;

(3)To calculate the non-linear principal component for seismic prospecting signal sample

After obtaining the non-linear principal component, appropriate classifier can be chosen to classify them. Figure 1 gives the specific process.

4. Simulation experiments

To test the performance of improved algorithm, this paper selected a group of seismic prospecting signals and simulated the standard PCA algorithm, wavelet transform algorithm and improved PCA algorithm, as shown from Figure 2 to Figure 4.

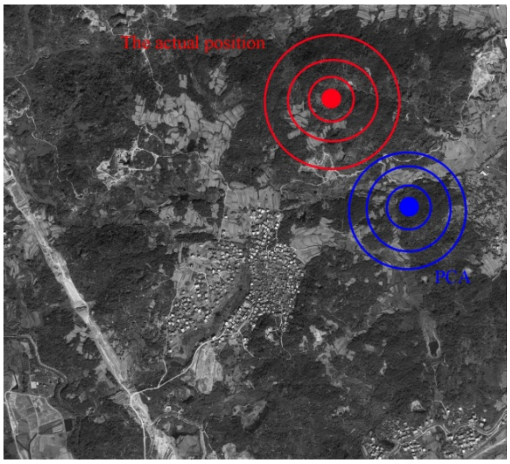

Figure 2. Signal recognition result of standard PCA algorithm

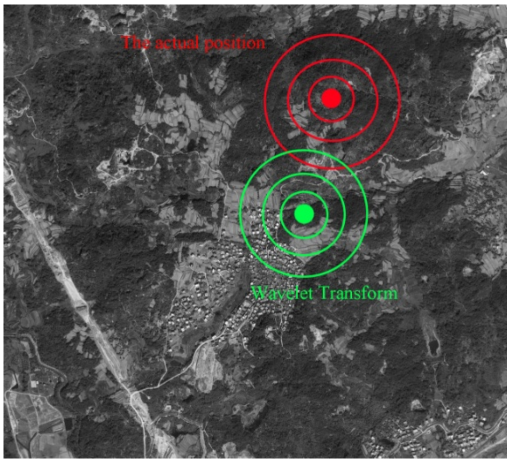

Figure 3. Signal recognition result of Wavelet Transform algorithm

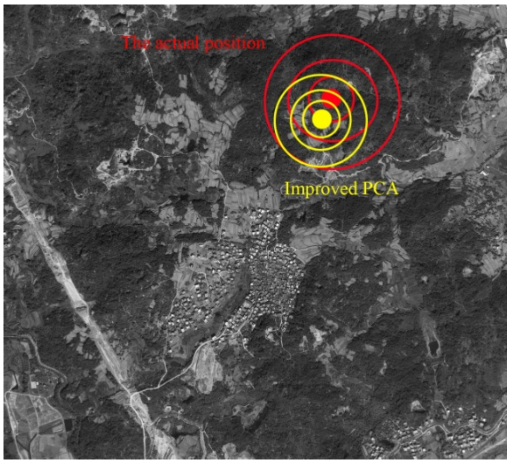

Figure 4. Signal recognition result of improved PCA algorithm

Comparing these figures, it is evident that the improved PCA algorithm based on kernel function weighted iteration optimization has better performance than standard PCA algorithm and wavelet transform algorithm in seismic prospecting signal identification.

5. Conclusions

In seismic prospecting signal analysis, increasing the signal to noise ratio of seismic image enhances the seismic interpretation of essential seismic events and its significant geologic structure, which provides the good basis for accurate geologic structure interpretation and deployment of the position of well. This paper takes into consideration the defects of traditional PCA algorithm in seismic prospecting signal identification, proposes an identification model based on improved PCA algorithm by weighted iteration optimization of kernel function. Simulation experiments prove that the improved PCA algorithm has better performance than standard PCA algorithm and wavelet transform algorithm in seismic prospecting signal identification.

References

1. Pan Jun (2015) The application of SRME in high resolution seismic exploration of shallow sea environment. *Progress in Geophysics*, 10(1), p.p. 429-434.
2. Hu Fengjun (2013) A rapid eye-to-hand coordination method of industrial robots. *Journal of Information and Computational Science*,10(5), p.p. 1489-1496.
3. Zhu Xufang,Yan Bing (2014) An Improved Method for Denoising Seismic Wave. *Laboratory Research and Exploration*, 33(2), p.p. 31-34.
4. Wei Sui (2015) 3-D Seismic Exploration Effect of Faer Colliery under Complex Condition. *Coal Mining Technology*, 20(1), p.p. 23-25.
5. Fang Shouchuan, Zhao Jianhu (2014) Research on the Gun Array Positioning Method in Marine Seismic Exploration. *Ocean Technology*, 33(5), p.p. 96-100.
6. Shen Hongyan (2014) Seismic scattering wave imaging based on seismic-earth model of point scattering. *Journal of Guilin University of Technology*, 34(4), p.p. 615-622.
7. Tian Lin (2014) The Improved Algorithm of Cohen Class Time-frequency Distributions and Its Application in Seismic Signal. *Science Technology and Engineering*, No.31, p.p. 9413-9416.
8. Rao Heqing, Li Pingjun (2014) Research and Application of the Wide Azimuth in 3D Seismic Exploration. *Metal Mine*, No.12, p.p.143-146.
9. Chen Xun, Liu Shucong (2014) A Method Based on Wavelet Analysis Seismic Signal Feature Vector Extraction Method. *Coal Technology*, 32(11), p.p. 143-144.
10. Yao Jiang (2014) 3D geometry optimization design based on the properties evaluation and analysis. *Geophysical Prospecting For Petroleum*, 53(4), p.p. 384-390.