NEAR LOSSLESS SEISMIC DATA COMPRESSION USING SIGNAL PROJECTION TECHNIQUE

*Hilal H. Nuha , **M. Mohandes, ***M. Deriche, ****Naveed Iqbal

Center for Energy and Geo Processing (CeGP)
King Fahd University of Petroleum & Minerals
Saudi Arabia

*hilalh@kfupm.edu.sa, ** mohandes@kfupm.edu.sa , *** mderiche@kfupm.edu.sa ,
**** naveediqbal@kfupm.edu.sa

ABSTRACT

Seismic data files in SEGY format can be of substantial size as these contain generally hundreds of traces collected from multiple shots. The data is usually transmitted through direct wires to different processing centers. It is important to preserve the integrity of the data in transmission and for storage, however, with very little loss, it is possible to compress the data with large compression factors. To this extend we propose a computationally efficient and robust technique for compressing multiple traces from multiple shots using Principle Component Analysis (PCA). Here, we use PCA to reduce the dimensionality of the data. The basic concept relies on finding a linear transformation that could project the original data over a set of orthogonal basis. The transformation is obtained from Singular Value Decomposition (SVD) of the estimated autocorrelation matrix. The autocorrelation matrix represents the dependencies across the traces from the same sensors and across different sensors. In our experiments we used publically available data from the Texas Seismic database. The data is gathered from 18 shots recorded by 33 sensors. The shots are generated by dynamite in 80-100ft depth holes. Each sensor is located 220 feet from another sensor. The data consists of 18 traces per sensor each trace contains 1501 data points sampled with 2 ms of sampling period and are filtered by 8-64Hz bandpass filter. The experiment covered 18 shots. Using the developed algorithm we reduce the dimensionality of the data from 18 to 3. The dimension of the output feature vector (3) was obtained by preserving more than 95% of the total energy in the original data which can be seen as almost a lossless compression.

Keywords— seismic data; PCA; SVD; compression

1. INTRODUCTION

Seismic data file in SEGY format is quite large. Number of trace generated by 1 shot only can reach $\pm 400K$ traces or equals to 2.5 GB. One SEGY file contains multiple shots making the file size can reach 50TB[1]. Such huge file size is a challenging problem to solve. Each trace contains similarity with other traces since the traces are either generated by the same shot or recorded by the same recorder.

Some works on seismic data compression have been proposed ranging from compression based on quantization at acquisition[2], transform on quantization[3], and Wavelet transform at seismic image level[4]. Most of the works, however, did not use public data for experiment.

2. SEISMIC DATA AND PROPOSED WORK

2.1 Seismic Data

We use Seismic data from east Texas, USA [5]. Seismic data file contains seismic traces sampled at Ts=2ms or Fs=500Hz . The files uses number of sample per trace $\mathrm{N_t} = 1501$. Seismic data for our experiment, consist of Number of shot 18, each , number of channel per slot 33, making number of traces $\mathrm{N_n} = 18\times 33 = 594$. Sensor location is determined by the gx field at header file. Each sensor records 18 shots (Figure 1). Data are recorded using receivers with interval 220 ft=67.056 m, (distance between traces). Each sample is represented by double-floating point type data(64 bits=8Bytes). Total sample points are $N_s = N_t \times N_n = 1501\times 594 = 891594$ sample points. Data size $N_s \times 8B = 7132752B = 6.802\ MB$. The traces are filtered by 8-64Hz bandpass filter.

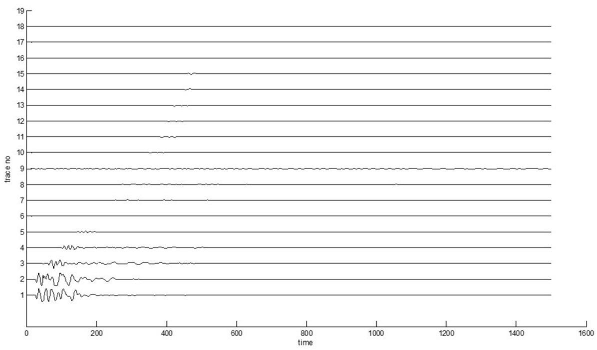

Figure 2.1. Trace plot of 18 traces recorded by a sensor located at 220ft

2.2 Singular Value Decomposition

Dimensionality reduction is a well-known technique to represent multidimensional signal into smaller dimension with least distortion. The idea is to project the data to a set of vector where each vector is perpendicular against other vectors. The data are projected onto these vectors orthogonally such that the distortion, which is the distance of the original position and the projection on the vector, is minimum. The figure below illustrates the technique.

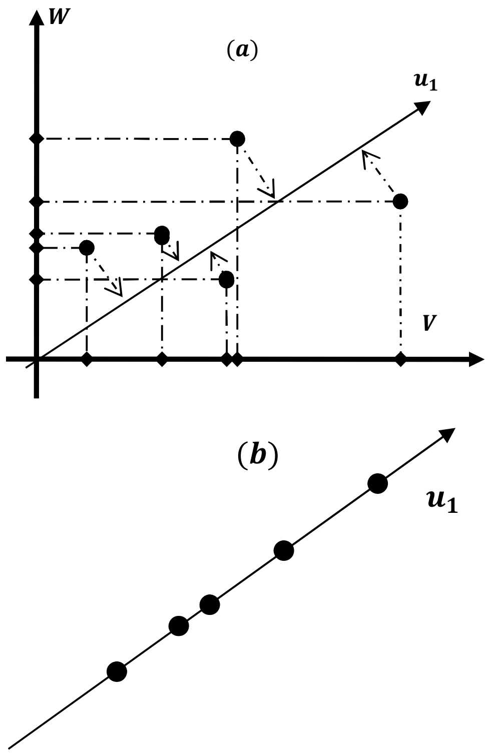

Figure 2.2 (a) 2-D data projected onto vector u_1 (b) 1-D Orthogonal Projection of 2-D data to u_1

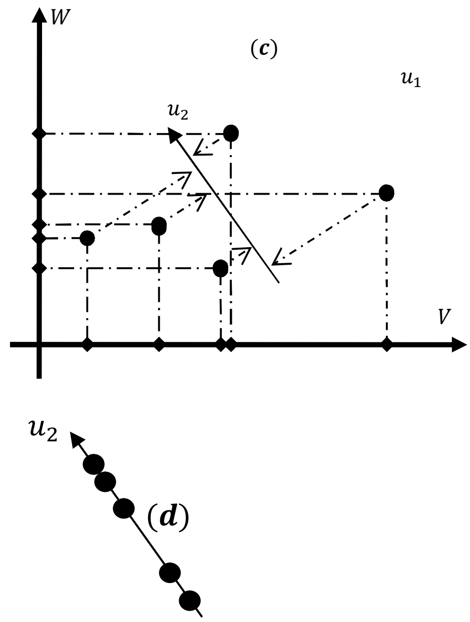

Figure 2.2 (c) 2-D data projected onto vector u_2 (d) 1-D Orthogonal Projection of 2-D data to u_2

Figure above illustrates how signal projection can reduce the number of dimension needed to represent the data. Figure (a) shows how original data on 2-Dimension projected onto vector u_1 and figure (b) shows 1-D representation of the data after are orthogonally projected to vector u_1. Figures (c) and (d) show the same mechanism, respectively, but the data are orthogonally projected to vector u_2 instead of vector u_1. Vectors $u_1$and u_2 are orthogonal each other. We can fully reconstruct the data by projecting both projections on the vectors onto original basis vector. For compression purpose, reducing dimension is a method to reduce number of data representation. To reconstruct the original data from the projection, we need to keep the projection vectors (u_1 and or u_2). Using less number of eigenvectors will generate distortion to the reconstructed data.

Please open http://arsgiso.org/journals.php?DoGetPublication=DoGet&varGetVolume=3&varGetIssue=1&varGetIdJournal=6# to see full version.

2.3 Principal Component Analysis

The traces above can be represented by a 18×1501 matrix T. for compression purpose we use Principal Component Analysis(PCA) for dimensionality reduction. We use each trace as feature such that we have 18 features. After averaging, we have zero average matrix for each column T_0 and we have average row vector $\bar{x}_T$ containing average for each column of T. we can obtain covariance matrix of T, $C_T = T_0^T T_0$. covariance matrix of T can be decomposed into $C_T = W^T \Lambda W$ where Λ is diagonal matrix consisting of eigenvalues of C_T sorted descending and W is a matrix consisting of eigenvectors corresponding to Λ.

Since we have 18 features, we get 18 eigenvectors. For compression purpose we can choose the N vectors to represent the most principal components(PCs) and compose the N largest eigenvectors to be the feature matrix W_N. Final data $T_0{}'$ can be obtained from

$$\boldsymbol{T_0' = W_N^T T_0^T}$$

To recover the data T', we need W_N, T_0', and $\bar{x}_T$. Following equation shows how to obtain T'.

$$\boldsymbol{T' = T_0'^{\,T} W_N^T + \bar{x}_T}$$

Recovered data T' equals to the original data T if we choose all eigenvectors instead of N largest eigenvectors such that $W_N = W$. However, the compression ratio of this technique mainly depends on number of eigenvector that we choose W_N.

2.4 Compression Ratio

Let us assume that T is $K{\times}M$ matrix, then W_N is a $M{\times}N$ matrix, T_0' is a $N{\times}K$, and $\bar{x}_T$is a $1{\times}M$ matrix. We can estimate the compression ratio as follows.

$$\boldsymbol{r = \frac{K\times M}{M\times N + N\times K + 1\times M} = \frac{KM}{N(K+M)+M}}$$

If we define $r_N = \frac{N}{M}$, then $N = r_N M$. Equation above becomes.

$$\boldsymbol{r = \frac{KM}{r_N M(K+M)+M}}$$

$$\boldsymbol{= \frac{KM}{(r_N(K+M)+1)M}}$$

$$\boldsymbol{= \frac{K}{r_N(K+M)+1}}$$

To obtain compression ratio $r > 1$, we can derive the requirements.

$$\boldsymbol{\frac{K}{r_N(K+M)+1} > 1}$$

$$\boldsymbol{r_N < \frac{K-1}{(K+M)}}$$

In case of a network of sensor where the fusion center kept the estimation of covariance matrix of traces from each sensor, we can directly obtain the compression ratio by evaluating the size of projected data(reduced dimensionality) and the original data. The compression ratio can be expressed as follows.

$$\boldsymbol{r = \frac{KM}{KN} = \frac{M}{N}}$$

Expression above clearly leads to higher compression ratio. Since there is no need to keep the eigenvectors or covariance matrix.

3. EXPERIMENTAL RESULT

In this section we will see the performance of PCA on the seismic dataset. The following figures show the performance of the technique to reduce the data representation size and its impact to the information conveyed by the seismic traces.

4. CONCLUSION AND FUTURE WORKS

We implement PCA to do compression by reducing the dimensionality of a seismic dataset. The experiment shows that PCA effectively preserves the integrity of seismic traces by 95 % while reduce the size by 6 times. In the future, we wish to evaluate some lossless technique to compress the lossy part of the data and increase the data integrity while maintain the compression rate

5. REFERENCES

[1] Breunig, Peter.2011. Managing More Bits Than Barrels. Chevron.

[2] Davis, Aaron J. "Linear prediction coding for compressing of seismic data." U.S. Patent No. 4,509,150. 2 Apr. 1985.

[3] Spanias, Andreas S., Stefan B. Jonsson, and Samuel D. Stearns. "Transform methods for seismic data compression." Geoscience and Remote Sensing, IEEE Transactions on 29.3 (1991): 407-416.

[4] Villasenor, John D., R. A. Ergas, and P. L. Donoho. "Seismic data compression using high-dimensional wavelet transforms." Data Compression Conference, 1996. DCC'96. Proceedings. IEEE, 1996.

[5] Al-Shuhail, Abdullatif A., and Wail A. Mousa. "Processing of Seismic Reflection Data Using MATLAB." 74th EAGE Conference & Exhibition-Workshops. 2012