STUDIES OF SYSTEM IDENTIFICATION AND DENOISING FOR GEOPHYSICAL SIGNALS

KARTHIKEYAN E

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2018

© Indian Institute of Technology Delhi (IITD), New Delhi, 2018

STUDIES OF SYSTEM IDENTIFICATION AND DENOISING FOR GEOPHYSICAL SIGNALS

by

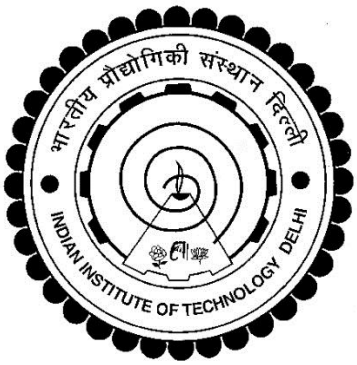
KARTHIKEYAN E

DEPATMENT OF ELECTRICAL ENGINEERING

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2018

Dedicated to my beloved
Mother & Brothers

Certificate

This is to certify that the thesis entitled " **Studies of system identification and denoising for geophysical signals**" being submitted by **Mr. Karthikeyan E** to the **Department of Electrical Engineering, Indian Institute of Technology Delhi**, for the award of the degree of **Doctor of Philosophy** is the record of the bona-fide research work carried out by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted either in part or in full to any other university or institute for the award of any degree or diploma.

Prof. Brejseh Lall
Professor
Department of Electrical Engineering
Indian Institute of Technology, Delhi
Hauz Khas New Delhi 110016

Prof. R K Patney
Professor
Department of Electrical Engineering
Indian Institute of Technology, Delhi
Hauz Khas New Delhi 110016

Date:

Place: New Delhi

Acknowledgments

First of all, I would like to thank the Indian Institute of Technology, Delhi for giving an opportunity to carry out this research work. Equally, I would like to acknowledge my advisors Prof. R K Patney and Prof. Brejesh Lall for their immense support and guidance for my research. Their profound technical knowledge and passion for work have always inspired me to work hard.

I am thankful to my student research committee members Prof. S. D. Joshi, Prof. R K P Bhaat, and Prof. Arun Kumar for many useful interactions and their valuable comments and suggestions on my work. I am also grateful to all the faculty members at IIT Delhi from whom I have learned a lot during the course work and the period of my research.

I feel fortunate to have many friends and well-wishers during my PhD with whom I have shared my joy and sorrows. I thank each and every one for their support, trust and encouragement all the time.

Last but not the least, I owe a lot to my mother and brothers for supporting me throughout the course of my PhD and in my personal life.

Karthikeyan E

Abstract

The seismic exploration method is a cost-effective approach to analyze the earth's subsurface layer up to certain depth at a particular place. In seismic exploration, a short duration seismic wavelet (pulse) is transmitted from the surface, which is reflected from the boundaries between earth's subsurface layers. The reflected waves are observed and recorded by an array of sensors at the surface of the earth. The sensor output (seismic trace) contains information about the seismic wavelet, the reflection coefficients (ratio of incident and reflected wave energy) and noise. This sequence of reflection coefficients contains the earth subsurface layer's information. To obtain the reflection coefficients from a noisy trace, seismic deconvolution and denoising are needed to mitigate the seismic wavelet and the noise. In order to perform deconvolution, the seismic wavelet must be known a priori. If it is unknown, then the seismic wavelet needs to be estimated accurately from the seismic trace. Otherwise, the obtained noisy reflection coefficients may lead to misinterpretation about the earth layers.

The objective is to estimate the seismic wavelet from noisy trace based on the statistical assumption about the reflection coefficients and noise. The seismic wavelet to be determined is modeled as non-minimum phase finite impulse response (FIR) system and reflection coefficients

assumed to be zero mean, non-Gaussian statistically independent and identically distributed (IID) process. In order to estimate the system parameters from the observed noisy seismic signal, we pose this as a Blind System Identification (BSI) problem. To estimate the system parameters, we derive a set of equations by matching the higher-order cumulant of observed noisy data with a higher order moment of blind system's impulse response. In order to solve those equations to obtain the desired system parameters, this thesis deals with two different methods to estimate the seismic wavelet from the noisy trace.

The first contribution presented in this thesis is that the obtained system of equations is mapped to a multivariate polynomial function. The Groebner bases that form the solution to this set of polynomial function is obtained using the proposed algorithm. The Groebner basis finds all the possible solution in the cumulant matching method. Since the polynomial function is derived using single order cumulant; this improves the accuracy of the estimated system parameters. In order to test the efficiency of the proposed blind system identification method, it is applied to the synthetic data set to estimate the desired system parameters. For the synthetic data, the system parameters are known beforehand. The proposed method estimates the desired system parameters blindly from the synthetic data set. In order to measure the

performance of the proposed method, the mean squared error (MSE) is used. It measures the accuracy between the estimated and true parameters. From the simulation results, we conclude that the accuracy of wavelet estimation is improved 10 times than the existing method.
Next, if the system parameters are large, then the set of non-linear equations obtained by cumulant matching has a large number of equations. It is very difficult to solve using the algebraic method. We proposed a new objective function to solve a system of nonlinear equations. The Particle Swarm Optimization (PSO) algorithm is used to minimize the objective function and obtain the desired seismic wavelet parameters. The higher order statistics are insensitive to Gaussian noise. Therefore the error between estimated and true parameters is small even for low signal to noise ratio. From the simulation results, conclude that the accuracy of wavelet estimation improved 10 times than the existing method.
The next contribution of the thesis is denoising the seismic data. There are various methods existing in the literature to eliminate the random noise in seismic data. In those methods, the data is transformed to another domain and used a threshold to eliminate the noise and inverse transform to the original domain. But those methods require an appropriate threshold, which is difficult to estimate. In order to over-

come this problem, we use a dictionary learning method for denoising the seismic data. The set of traces can be modeled as belonging to a low dimensional subspace of ambient signal space. This allows for construction of sparse representations of each trace in terms of other traces in the commom mid-point (CMP) gathered and normal move-out (NMO) corrected traces. The sparse representations provide us a way to construct approximations of individual traces in which the noise is suppressed. We use OMP to construct these approximations. If the number of interfaces (sparsity) is K, we assume that the signal matrix (seismic data) is of rank K and we compute the K term global approximation of each trace in terms of other traces. If the sparsity K of the reflection coefficients matrix is known, then we can use K as the number of iterations for which OMP should run for each trace. The signal can then be recovered exactly using the OMP algorithm. The K can be choose approximately from the seismic data, here K represents number of layers. We demonstrate, with applications to synthetic and field seismic data, that the proposed technique performs better on established benchmarks while capturing the true locations of weak reflections and effectively attenuating the random noise. In order to numerically compute the performance of the proposed approach, we use the Signal-to-Noise Ratio(SNR) and the Mean Square Error (MSE).

The proposed method improves the SNR of data and reduces the mean square error between the true and estimated data. From the simulation results, conclude that the SNR of seismic data is improved by 6dB as compared to the existing method.

सार

भूकंपीय अन्वेषण विधि किसी विशेष स्थान पर कुछ गहराई तक पृथ्वी की उप-सतह परत का विश्लेषण करने के लिए एक लागत प्रभावी तरीका है। भूकंपीय अन्वेषण में, सतह से एक छोटी अवधि भूकंपीय तरंग (नाड़ी) फैलती है, जो पृथ्वी की उप-सतह परतों के बीच की सीमाओं से प्रतिबिंबित होती है। परावर्तित तरंगें पृथ्वी की सतह पर सेंसर की एक सरणी द्वारा देखी और दर्ज की जाती हैं। सेंसर आउटपुट (भूकंपीय निशान) में भूकंपीय वेवलेट, प्रतिबिंब गुणांक (घटना का अनुपात और प्रतिबिंबित लहर ऊर्जा) और शोर के बारे में जानकारी शामिल है। प्रतिबिंब गुणांक के इस अनुक्रम में पृथ्वी उप-सतह परत की जानकारी शामिल है। एक शोर ट्रेस से प्रतिबिंब गुणांक प्राप्त करने के लिए, भूकंपीय तरंगों और शोर को कम करने के लिए भूकंपीय deconvolution और denoising की जरूरत है। Deconvolution करने के लिए, भूकंपीय तरंग एक prei जाना जाना चाहिए। यदि यह अज्ञात है, तो भूकंपीय तरंगों को भूकंपीय निशान से सटीक रूप से अनुमानित करने की आवश्यकता है। अन्यथा, प्राप्त शोर प्रतिबिंब गुणांक पृथ्वी परतों के बारे में गलत व्याख्या कर सकते हैं।

इसका उद्देश्य प्रतिबिंब गुणांक और शोर के बारे में सांख्यिकीय धारणा के आधार पर शोर से निकलने वाली भूकंपीय लहर को अनुमानित करना है। निर्धारित करने के लिए भूकंपीय वेवलेट को गैर-न्यूनतम चरण परिमित आवेग प्रतिक्रिया (एफआईआर) प्रणाली के रूप में मॉडलिंग किया गया है और प्रतिबिंब गुणांक शून्य माध्य, गैर-गॉसियन सांख्यिकीय रूप से स्वतंत्र और समान रूप से वितरित (आईआईडी) प्रक्रिया माना जाता है। मनाए गए शोर भूकंपीय संकेत से सिस्टम पैरामीटर का अनुमान लगाने के लिए, हम इसे एक अंधेरे सिस्टम पहचान (बीएसआई) समस्या के रूप में पेश करते हैं। सिस्टम पैरामीटर का अनुमान लगाने के लिए, हम अंधेरे प्रणाली के आवेग प्रतिक्रिया के उच्च क्रम पल के साथ मनाए गए शोर डेटा के उच्च-आदेश संचयी से मेल करके समीकरणों का एक सेट प्राप्त करते हैं। वांछित सिस्टम पैरामीटर प्राप्त करने के लिए उन समीकरणों को हल करने के लिए, यह थीसिस शोर ट्रेस से भूकंपीय तरंग का अनुमान लगाने के लिए दो अलग-अलग तरीकों से संबंधित है।

इस थीसिस में प्रस्तुत पहला योगदान यह है कि समीकरणों की प्राप्त प्रणाली को बहुआयामी बहुपद कार्य में मैप किया जाता है। ग्रोबनेर बेस जो बहुपद कार्य के इस सेट के समाधान का निर्माण करते हैं, प्रस्तावित एल्गोरिदम का उपयोग करके प्राप्त किया जाता है। ग्रोबेनर आधार संचयी मिलान विधि में सभी संभावित समाधान पाता है। चूंकि बहुपद कार्य एकल आदेश संचयी का उपयोग करके लिया गया है; यह अनुमानित सिस्टम पैरामीटर की सटीकता में सुधार करता है। प्रस्तावित अंधेरे प्रणाली पहचान विधि की दक्षता का परीक्षण करने के लिए, यह वांछित सिस्टम पैरामीटर का अनुमान लगाने के लिए सिंथेटिक डेटा सेट पर लागू होता है। सिंथेटिक डेटा के लिए, सिस्टम पैरामीटर पहले से ज्ञात हैं। प्रस्तावित विधि सिंथेटिक डेटा सेट से वांछित सिस्टम पैरामीटर का आकलन करती है। प्रस्तावित विधि के प्रदर्शन को मापने के लिए, माध्य वर्ग त्रुटि (एमएसई) का उपयोग किया जाता है। यह अनुमानित और सही मानकों के बीच सटीकता को मापता है। अनुकरण परिणामों से, हम निष्कर्ष निकालते हैं कि मौजूदा विधि की तुलना में वेवलेट अनुमान की शुद्धता 10 गुना बेहतर है।

अगला, यदि सिस्टम पैरामीटर बड़े हैं, तो संचयी मिलान द्वारा प्राप्त गैर-रैखिक समीकरणों का सेट समीकरणों की एक बड़ी संख्या है। बीजगणितीय विधि का उपयोग करके हल करना बहुत मुश्किल है। हमने nonlinear समीकरणों की एक प्रणाली को हल करने के लिए एक नया उद्देश्य समारोह प्रस्तावित किया। कण स्वार अनुकूलन (पीएसओ) एल्गोरिदम का उपयोग उद्देश्य कार्य को कम करने और वांछित भूकंपीय वेवलेट पैरामीटर प्राप्त करने के लिए किया जाता है। उच्च आदेश आंकड़े गॉसियन शोर के लिए असंवेदनशील हैं। इसलिए अनुमानित और सही पैरामीटर के बीच त्रुटि शोर अनुपात के लिए कम सिग्नल के लिए भी छोटी है। सिमुलेशन परिणामों से, निष्कर्ष निकाला है कि वेवलेट अनुमान की शुद्धता मौजूदा विधि से 10 गुना बेहतर हुई है।

थीसिस का अगला योगदान भूकंपीय डेटा को denoising है। भूकंपीय डेटा में यादृच्छिक शोर को खत्म करने के लिए साहित्य में मौजूद विभिन्न विधियां हैं। उन तरीकों से, डेटा दूसरे डोमेन में परिवर्तित हो जाता है और शोर को समाप्त करने और मूल डोमेन में उलटा परिवर्तन करने के लिए थ्रेसहोल्ड का उपयोग किया जाता है। लेकिन उन तरीकों के लिए उचित सीमा की आवश्यकता होती है, जो अनुमान लगाना मुश्किल है। इस समस्या को दूर करने

के लिए, हम भूकंपीय डेटा को नकारने के लिए एक शब्दकोश सीखने की विधि का उपयोग करते हैं। निशान का सेट परिवेश सिग्नल स्पेस के निम्न आयामी उप-स्पेस से संबंधित के रूप में मॉडलिंग किया जा सकता है। यह एकत्रित कमांड मिड-पॉइंट (सीएमपी) में अन्य निशानों के मामले में प्रत्येक ट्रेस के स्पैस प्रस्तुतियों के निर्माण की अनुमति देता है और सामान्य चाल-आउट (एनएमओ) सही निशान का पता लगाता है। स्पैस प्रस्तुतियां हमें व्यक्तिगत निशान के अनुमान बनाने का एक तरीका प्रदान करती हैं जिसमें शोर दबाया जाता है। हम इन अनुमानों को बनाने के लिए ओएमपी का उपयोग करते हैं। यदि इंटरफेस (स्पैरसिटी) की संख्या के है, तो हम मानते हैं कि सिग्नल मैट्रिक्स (भूकंपीय डेटा) रैंक के है और हम अन्य निशानों के संदर्भ में के ट्रेस के प्रत्येक अनुमान के वैश्विक अनुमान को गणना करते हैं। यदि प्रतिबिंब के sparsity के

Contents

Abstract **iii**

List of Figures **xii**

List of Tables **xiv**

1 Introduction **1**

1.1 Brief overview 1

1.2 Challenges in seismic data processing 6

1.3 Seismic signal model 7

1.4 Literature 11

1.5 Motivation 20

1.6 Key contributions 23

1.7 Outline of thesis 24

2 Mathematical preliminaries **27**

2.1 Preliminaries of higher order statistics 28

2.2 Groebner Bases 30
2.3 Preliminaries for subspace methods 39

3 Seismic wavelet estimation using Groebner basis 41
3.1 Introduction 41
3.2 Blind system identification 43
3.3 Results and discussion 48
3.4 Real seismic data 54
3.5 Summary 57

4 Seismic wavelet estimation using particle swarm optimization 62
4.1 Introduction 62
4.2 System of non-linear equations 63
4.2.1 Proposed model 65
4.2.2 Particle swarm optimization algorithm 66
4.3 Results and discussion 72
4.4 Summary 77

5 Pre-stack denoising using subspace estimation method 81
5.1 Introduction 81
5.2 Modeling seismic data 85
5.3 Subspace estimation methods 87

5.3.1 Low rank minimization 87
5.3.2 Principal component pursuit 89
5.3.3 Sparse representations 90
5.3.4 Sparse representations by orthogonal matching pursuit 91
5.4 Results and discussion 93
5.4.1 Synthetic seismic data: Case study 1 Data before deconvolution 93
5.4.2 Synthetic seismic data: Case study 2 Deconvolved data 95
5.4.3 Application to field data 96
5.5 Summary 99

6 Conclusions and Future work **100**
6.1 Contributions of this research 100
6.2 Future work 103
6.3 List of publications 104

Bibliography **104**

Appendix A **119**

List of Figures

1.1 Seismic data acquisition [13] 4
1.2 Block diagram of seismic data processing 4
1.3 Block diagram for deconvolution 11

3.1 Mean Square Error of the obtained system parameters vs SNR 55
3.2 The real seismic data before deconvolution 59
3.3 The seismic wavelet estimated from traces 60
3.4 The real seismic data after deconvolution 61

4.1 PSO Flowchart 68
4.2 The synthetic seismic data consist of 50 traces each trace generated by convolving seismic wavelet with input and corrupted by additive Gaussian noise. 78
4.3 The unknown wavelet estimated from seismic trace for different data length and compared with true wavelet . . 78
4.4 The seismic wavelet estimated from traces 79

4.5 Real seismic data after deconvolution 80

5.1 (a) Synthetic seismic data corrupted by white Gaussian noise. (b) Denoised using Predictive filtering method (c) Denoised using wavelet method (d) denoised using proposed method 94

5.2 (a) A deconvolved synthetic seismic data corrupted by Gaussian noise. (b) Denoised using Predictive filtering method (c) Denoised using wavelet method (d) denoised using proposed method 97

5.3 (a) one CMP gather from synthetic data set. (b) NMO-corrected gather. (c) Result of proposed method. 98

List of Tables

3.1 The minimum phase FIR parameters estimated using third order statistics for noise free case and 50 Monte Carlo runs 50

3.2 The minimum phase FIR system parameter estimated using third oder statistics for different SNR, N=2000 and for 50 monte carlo runs 51

3.3 The non-minimum phase channel system parameter estimated for different SNR (50 monte carlo runs, N=640) 53

3.4 The non-minimum phase channel system parameter estimated for various data length for 50 monte carlo runs . 53

4.1 The non-minimum phase FIR system parameters estimated using third order statistics for noise free case and 50 Monte Carlo runs. 72

4.2 The non-minimum phase FIR system parameter estimated using third oder statistics for different SNR, N=1000 and for 50 monte carlo runs 73

5.1 The SNR and MSE comparison for deconvolved data based on predictive filtering, wavelet transform and the proposed method 95

5.2 The SNR and MSE comparison for deconvolved data based on predictive filtering, wavelet transform and the proposed method 96