저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer
Robust Elastic Full Waveform Inversion using Student’s t-distribution in the Frequency domain

스튜던트의 티 분포를 이용한 주파수영역 탄성파 완전파형역산

2013 년 2 월

서울대학교 대학원
에너지시스템공학부
강 민 지
Robust Elastic Full Waveform Inversion using Student’s t-distribution in the Frequency domain

스튜던트의 티 분포를 이용한 주파수영역 탄성파 완전파형역산

지도 교수 민 동 주

이 논문을 공학석사 학위논문으로 제출함
2012 년 12 월

서울대학교 대학원
에너지시스템공학부
강 민 지

강민지의 공학석사 학위논문을 인준함
2013 년 1 월

위원 장 _______________________(인)

부위원장 ______________________(인)

위원 ______________________(인)
ABSTRACT

Seismic full waveform inversion (FWI) is a numerical technique that estimates subsurface parameters. FWI is usually based on a nonlinear least-squares optimization problem. However, it has been known that the least-squares objective function cannot properly estimate subsurface material properties when field data are contaminated with noise such as outliers. In this study, we propose a 2D elastic FWI algorithm based on Student’s t-distribution, which has an overdispersed density compared to the Normal distribution and can be useful for data with outliers. To apply t-distribution to the elastic FWI, the statistical techniques such as Maximum a posteriori (MAP) and Maximum likelihood (ML) are used. The inversion algorithm is based on the finite-element modeling and the adjoint state of the wave equation. To calculate gradient directions efficiently, the gradients were computed by the cross-correlation of back-propagated residuals and virtual sources and the pseudo-Hessian matrix was applied. Also, the conjugate gradient method was used to accelerate the convergence rate of inversion.

The elastic FWI using Student’s t-distribution is demonstrated for 2D synthetic data set for the Modified elastic Marmousi-2 model. For comparison, the $l_2$- and $l_1$-norm-based FWI have also been applied to the model. For noise-free data, all the inversion results obtained by the three objective functions are in good agreement with the true velocities. For data with 10 outliers, the magnitude of outliers is 150% of the maximum amplitude of signal in each frequency. While the velocity model inverted by the $l_2$-norm FWI is severely distorted by the outliers, the $l_1$-norm and Student’s t misfit yield reliable results. When both outliers and random noises are applied, inversion results obtained by the $l_2$-norm are much poorer than those obtained for the data with only outliers. It seems like that the $l_1$-norm FWI is less influenced by random noise compared to the other methods. Although the RMS errors of Student’s t-distribution are lower than those of the other methods and yields better inversion results than the $l_2$- and $l_1$-norm objective functions, the distortions caused by random noise appear throughout the entire P-wave velocity model. From these results, we note that Student’s t misfit can decrease the influence of large outliers on inversion results, in particular for deep structures. We expect that other statistical distributions can be applied to seismic FWI.

KEYWORDS: Seismic full waveform inversion, elastic wave equations, Student’s t-distribution, gradient direction

STUDENT NUMBER: 2011-21089
CONTENTS

ABSTRACT ........................................................................................................... I

CONTENTS ........................................................................................................ II

LIST OF FIGURES ........................................................................................ III

INTRODUCTION ................................................................................................. 1

1. FREQUENCY-DOMAIN ELASTIC WAVE MODELING ........................... 3
   1.1. Frequency-domain elastic wave equations ............................................. 3
   1.2. Modeling using the finite element method ............................................. 6

2. METHODS FOR PARAMETER ESTIMATION ........................................ 8
   2.1. Objective functions ............................................................................. 8
   2.2. Maximum likelihood method............................................................... 9
   2.3. Long-tailed Student’s t misfit ............................................................. 11

3. FULL WAVEFORM INVERSION ALGORITHM ................................... 15
   3.1. Gradient direction............................................................................ 15
   3.2. Scaling method............................................................................... 17
   3.3. Conjugate gradient method............................................................ 17

4. NUMERICAL EXAMPLES ........................................................................ 18
   4.1. Elastic Marmousi-2 model without noises ......................................... 18
   4.2. Elastic Marmousi-2 model with outliers ............................................ 24
   4.3. Elastic Marmousi-2 model with outliers and random noises ............ 30

CONCLUSIONS ............................................................................................. 35

REFERENCES ............................................................................................... 36
LIST OF FIGURES

**Figure 1** P-wave velocity of the 2D elastic Marmousi-2 model. .......... 19

**Figure 2** Seismograms obtained for the 2D elastic Marmousi-2 model: (a) horizontal and (b) vertical displacements. .................................................. 19

**Figure 3** Initial model and P-wave velocities inverted by the (b) $l_2$-norm, (c) $l_1$-norm, and (d) Student’s t-distribution at the 300th iteration for synthetic data without noise for the Marmousi-2 model. ............... 20

**Figure 4** Depth profiles at a distance of 3 km extracted from the true and inverted models (Figure 3) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration. ......................... 22

**Figure 5** Depth profiles at a distance of 6 km extracted from the true and inverted models (Figure 3) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration. ......................... 23

**Figure 6** Seismograms with 10 outliers randomly distributed over the seismogram for the Marmousi-2 model: (a) horizontal and (b) vertical displacements................................................................. 25

**Figure 7** Seismogram with outliers recorded at a distance of 1.82 km for (a) horizontal and (b) vertical displacements. Red arrows indicate outliers. ............................................................................ 26

**Figure 8** P-wave velocities inverted by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration for synthetic data with outliers for the Marmousi-2 model. .................................................. 27

**Figure 9** Depth profiles at a distance of 3 km extracted from the true and inverted models (Figure 8) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration. ......................... 28

**Figure 10** Depth profiles at a distance of 6 km extracted from the true and inverted models (Figure 8) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration. ......................... 29

**Figure 11** The real part of forward data when 0.25% random noise was added to background including outliers at 10 Hz: (a) horizontal and (b) vertical displacements. ................................................................. 31
Figure 12 P-wave velocities inverted by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration for synthetic data with outliers and random noises for the Marmousi-2 model. .................... 32

Figure 13 Depth profiles at a distance of 3 km extracted from the true and inverted models (Figure 12) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration. .................... 33

Figure 14 Depth profiles at a distance of 6 km extracted from the true and inverted models (Figure 12) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration. .................... 34
INTRODUCTION

Seismic full waveform inversion is a numerical technique that can estimate subsurface parameters. By updating parameters, the error which indicates the difference between recorded data and modeling wavefields can be reduced and this process makes it possible to estimate unknown parameters such as the bulk modulus, the shear modulus, and density. Waveform inversion is normally based on a nonlinear least squares optimization problem (Virieux and Operto, 2009). But, the least-squares objective function has a weakness in the presence of large isolated and non-Gaussian errors (Brossier et al., 2010). Therefore, other norms have been considered such as the least-absolute-values $l_1$-norm by Crase et al. (1990), the Huber criterion by Huber (1973) and Guitton and Symes (2003), and the hybrid $l_1/l_2$ criterion by Bube and Langan (1997).

Waveform inversion can be interpreted by statistical approach. As statistical interpretation techniques, various methods of parameter estimation have been developed from statistics such as Bayesian methods including Maximum a posteriori (MAP) and Maximum likelihood (ML). Gauvain and Lee (1994) studied a framework for MAP estimation of hidden Markov models. Mecklenbräuker and Gerstoft (2000) investigated objective functions derived from a unified approach based on ML estimation in a Bayesian point of view. Press et al. (1993) explained the least squares using a ML estimator and emphasized that robust statistics can deal with the cases where the normal or the Gaussian model is a bad approximation. So, other prior distribution is needed when the data include dominant outliers.

Among many distributions, Student’s t-distribution can be considered for a prior distribution of recorded data that are out of range. Student’s t-distribution has an overdispersed density compared to the Normal density and is useful for the case of suspected outliers in the data set. Aravkin et al. (2011) studied the full-waveform inversion using Student’s t-distribution in acoustic media and compared the inverted results of student’s t misfit with that of least-squares and the Huber misfits.

In this study, we apply Student’s t-distribution to the elastic waveform inversion into the statistical approach. Numerical examples for the modified elastic Marmousi-2 model are performed to investigate the feasibility of Student’s t-distribution for elastic waveform inversion. Also, the results inverted by Student’s t misfit are compared with those obtained by the least-
squares and the least-absolute FWI under the same condition. For various cases, a few of noises including outliers and random noises are tested for the Marmousi-2 model.
1. FREQUENCY-DOMAIN ELASTIC WAVE MODELING

1.1. Frequency-domain elastic wave equations

The propagation of elastic waves can be described based on two equations: the generalized Hooke’s law and the equation of motion. Hooke’s law is formulated as the proportionality between stress $\sigma_{ij}$ and strain $\varepsilon_{kl}$, with the 4th order tensor $c_{ijkl}$ of elastic constants:

$$\sigma_{ij} = c_{ijkl} \varepsilon_{kl},$$  \hspace{1cm} (1)

This form is known as the generalized Hooke’s law and $c_{ijkl}$ is also known as the stiffness tensor. Because the elastic constants have symmetry like equation (3), the stiffness tensor $c_{ijkl}$ can be written as $C_{IJ}$ by corresponding index $ij$ and $kl$ to $IJ$ such as:

$$C_{ijkl} = C_{ijlk} = C_{jikl} = C_{klij},$$  \hspace{1cm} (3)

$$C = \begin{bmatrix}
C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\
C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\
C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} \\
C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} \\
C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56} \\
C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66}
\end{bmatrix}.$$  \hspace{1cm} (4)
Also, in one dimension, the equation of motion can be formulated using the stress tensor by

\[ \sum F_i = \int_V f_i dV + \sum S_{ij} n_j dS = \int_V \rho \frac{\partial^2 u_i}{\partial t^2} dV = ma_i, \]  

(5)

where \( u \) represents the displacement, \( m \) is the mass, \( a \) is the acceleration, \( t \) is the time, and \( \rho \) is the density. The force \( F \) is acting on the surface element \( dS \), and the unit vector \( n \) is normal to the surface element \( dS \). If we apply Gauss’ divergence theorem and ignore the body forces \( f_i \), Navier’s equation can be obtained from the equation of motion as

\[ \rho \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial \sigma_{ij}}{\partial x_j} \quad \text{or} \quad \rho \ddot{u}_i = \sigma_{ij,j}, \]  

(6)

where \( x \) is the location, and \( \ddot{u} \) denotes the 2nd derivative of the displacement. Substitution of equation (1) in equation (6) gives the wave equation in a general elastic, homogeneous medium:

\[ \rho \frac{\partial^2 u_i}{\partial t^2} = c_{ijkl} \frac{\partial}{\partial x_j} \frac{\partial u_k}{\partial x_l}. \]  

(7)

For 2D heterogeneous isotropic media, elastic constants can be written by

\[
C = \begin{bmatrix}
    C_{11} & C_{13} & 0 \\
    C_{13} & C_{11} & 0 \\
    0 & 0 & C_{44}
\end{bmatrix},
\]

(8)

and the elastic wave equation can be expressed by

\[
\rho \frac{\partial^2 u_x}{\partial t^2} = \frac{\partial}{\partial x} \left( C_{11} \frac{\partial u_x}{\partial x} + C_{13} \frac{\partial u_z}{\partial z} \right) + \frac{\partial}{\partial z} \left( C_{44} \left( \frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z} \right) \right),
\]

(9)

\[
\rho \frac{\partial^2 u_z}{\partial t^2} = \frac{\partial}{\partial x} \left( C_{44} \left( \frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z} \right) \right) + \frac{\partial}{\partial z} \left( C_{13} \frac{\partial u_x}{\partial x} + C_{11} \frac{\partial u_z}{\partial z} \right).
\]

(10)
Taking the Fourier transform of these formulae yields the 2D frequency-domain elastic wave equation is written as

$$-\rho \omega^2 \ddot{u}_x = \frac{\partial}{\partial x} \left( C_{11} \frac{\partial \ddot{u}_x}{\partial x} + C_{13} \frac{\partial \ddot{u}_z}{\partial z} \right) + \frac{\partial}{\partial z} \left( C_{44} \left( \frac{\partial \ddot{u}_x}{\partial x} + \frac{\partial \ddot{u}_z}{\partial z} \right) \right),$$  \hspace{1cm} (11)

$$-\rho \omega^2 \ddot{u}_z = \frac{\partial}{\partial x} \left( C_{44} \left( \frac{\partial \ddot{u}_z}{\partial x} + \frac{\partial \ddot{u}_x}{\partial z} \right) \right) + \frac{\partial}{\partial z} \left( C_{13} \frac{\partial \ddot{u}_x}{\partial x} + C_{11} \frac{\partial \ddot{u}_z}{\partial z} \right).$$  \hspace{1cm} (12)

If we replace elastic constants with Lamé constants, 2D frequency-domain elastic wave equation can be expressed by

$$-\rho \omega^2 \ddot{u}_x = \frac{\partial}{\partial x} \left( (\lambda + 2\mu) \frac{\partial \ddot{u}_x}{\partial x} + \lambda \frac{\partial \ddot{u}_z}{\partial z} \right) + \frac{\partial}{\partial z} \left( \mu \left( \frac{\partial \ddot{u}_x}{\partial x} + \frac{\partial \ddot{u}_z}{\partial z} \right) \right),$$  \hspace{1cm} (13)

$$-\rho \omega^2 \ddot{u}_z = \frac{\partial}{\partial x} \left( \mu \left( \frac{\partial \ddot{u}_x}{\partial x} + \frac{\partial \ddot{u}_z}{\partial z} \right) \right) + \frac{\partial}{\partial z} \left( \lambda \frac{\partial \ddot{u}_x}{\partial x} + (\lambda + 2\mu) \frac{\partial \ddot{u}_z}{\partial z} \right).$$  \hspace{1cm} (14)
1.2. Modeling using the finite element method

To solve the elastic wave equation numerically, the finite element method can be applied to equation (13) and (14). The finite element method is good for solving partial differential equations by dividing up a complicated model into small elements. The purpose of weighted residual methods in the finite element concept is to find the coefficients so that the residual \( R \) becomes small over a chose domain \( \Omega \), and it can be acquired from the integral form as

\[
\int \int \int_\Omega W_m \times R \, dV = 0, \quad (15)
\]

where \( W_m \) is a set of weighted functions \((m=1, \ldots, M)\).

\[
\int_\Omega q \left[ -\rho \omega^2 \ddot{u}_z - \frac{\partial}{\partial x} \left( (\lambda + 2\mu) \frac{\partial \ddot{u}_x}{\partial x} + \lambda \frac{\partial \ddot{u}_z}{\partial z} \right) - \frac{\partial}{\partial z} \left( \mu \left( \frac{\partial \ddot{u}_x}{\partial x} + \frac{\partial \ddot{u}_z}{\partial z} \right) \right) \right] \, dx \, dz = 0, \quad (16)
\]

\[
\int_\Omega q \left[ -\rho \omega^2 \ddot{u}_z - \frac{\partial}{\partial x} \left( \mu \left( \frac{\partial \ddot{u}_x}{\partial x} + \frac{\partial \ddot{u}_z}{\partial z} \right) \right) - \frac{\partial}{\partial z} \left( \lambda \frac{\partial \ddot{u}_x}{\partial x} + (\lambda + 2\mu) \frac{\partial \ddot{u}_z}{\partial z} \right) \right] \, dx \, dz = 0, \quad (17)
\]

where \( q \) means a weighted function expressed as \( q = \sum_{j=1}^{N} q_j \phi_j \), and \( \Omega \) is the 2-D elastic domain. Also, a solution is approximated piecewise analytically and can be expressed as a superposition of basis functions as

\[
u = \sum_{i=1}^{N} h_i \phi_i, \quad (18)
\]

\[
u = \sum_{i=1}^{N} v_i \phi_i, \quad (19)
\]

where \( h \) denotes the horizontal displacements and \( v \) is the vertical displacements. Therefore, the elastic wave equations applied finite element method can be expressed by
These equations can be written in matrix-vector form with the source vector $f$ by,

$$\begin{bmatrix} M + K_{xx} & K_{xz} \\ K_{zx} & M + K_{zz} \end{bmatrix} \begin{bmatrix} h \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ f \end{bmatrix},$$

where

$$K_{xx} = \sum_{i=1}^{N} \sum_{j=1}^{N} \int_{\Omega} \left[ \left( \lambda + 2\mu \right) \frac{\partial \phi_i}{\partial x} \frac{\partial \phi_j}{\partial x} + \mu \frac{\partial \phi_i}{\partial x} \frac{\partial \phi_j}{\partial z} \right] dxdz,$$

$$K_{xz} = \sum_{i=1}^{N} \sum_{j=1}^{N} \int_{\Omega} \left[ \left( \lambda \frac{\partial \phi_i}{\partial z} \frac{\partial \phi_j}{\partial x} + \mu \frac{\partial \phi_i}{\partial x} \frac{\partial \phi_j}{\partial z} \right) \right] dxdz,$$

$$K_{zx} = \sum_{i=1}^{N} \sum_{j=1}^{N} \int_{\Omega} \left[ \left( \mu \frac{\partial \phi_i}{\partial z} \frac{\partial \phi_j}{\partial x} + \lambda \frac{\partial \phi_i}{\partial x} \frac{\partial \phi_j}{\partial z} \right) \right] dxdz,$$

$$K_{zz} = \sum_{i=1}^{N} \sum_{j=1}^{N} \int_{\Omega} \left[ \left( \mu \frac{\partial \phi_i}{\partial z} \frac{\partial \phi_j}{\partial x} + \left( \lambda + 2\mu \right) \frac{\partial \phi_i}{\partial z} \frac{\partial \phi_j}{\partial z} \right) \right] dxdz,$$

$$M = \sum_{i=1}^{N} \sum_{j=1}^{N} \int_{\Omega} \left[ -\rho \omega^2 \phi_i \phi_j \right] dxdz.$$

To remove the unwanted reflections from the artificial boundaries, Perfectly matched layer (PML) was used (Komatitsch and Tromp, 2003).
2. METHODS FOR PARAMETER ESTIMATION

2.1. Objective functions

There have been developed various methods of parameter estimation. Among them, least-squares estimation (LSE) has been a common choice of inversion theory in seismic modeling and inversion (Virieux and Operto, 2009). The least-squares method (or simply $l_2$-norm) is given by

$$C_{L2}(\mathbf{m}) = \frac{1}{2} \left[ \mathbf{d}_{obs} - \mathbf{d}_{cal}(\mathbf{m}) \right]^2,$$  \hspace{1cm} (28)

where $\mathbf{m}$ is the model parameter vector, $\mathbf{d}_{obs}$ is observed seismic data, and $\mathbf{d}_{cal}$ is the modeled seismic data. In equation (28), we try to find the parameter vector $\mathbf{m}$ minimizing misfit function $C(\mathbf{m})$.

Another way to estimate parameters is to use the least absolute deviations (or simply $l_1$-norm). The $l_1$-norm is to measure the discrepancy between the model and the observations as follows:

$$C_{L1}(\mathbf{m}) = \frac{1}{2} |\mathbf{d}_{obs} - \mathbf{d}_{cal}(\mathbf{m})|,$$  \hspace{1cm} (29)

where $|x| = (x^2)^{1/2}$ for one source, and one frequency (Brossier et al., 2010).
2.2. Maximum likelihood method

The aforementioned classical methods can be considered from the point of statistical view. In a statistical approach, there are many methods for estimating parameters such as MAP (Maximum a posteriori) and ML (Maximum likelihood). ML is used to estimate the parameters of model when observations of the variables are given in the model. Concretely, ML finds the most likely value of parameter \( m_j \) by maximizing the likelihood function \( l(d;m) \) as

\[
\max_{m \in \Theta} l(d;m),
\]

and the resulting value of \( m \) is called the maximum likelihood estimate (Nocedal and Wright, 2006). The parameter \( \Theta \) denotes the parameter space, where \( \theta \in \Theta \subset R^n \). The difference between MAP and ML estimation is the assumption of an appropriate prior distribution of the parameters to be estimated (Gauvain and Lee, 1994).

Press et al. (1992) referred to least squares as a maximum likelihood estimator. Supposed that \( N \) data points \( (x_i, y_i), \ i = 1, \ldots, N \) are fitted to a model that has \( M \) adjustable parameters \( a_j, j = 1, \ldots, M \), there exists a functional relationship between the measured variables as follows:

\[
y_i = y(x_i; a_1, \ldots, a_M) + \varepsilon,
\]

where \( \varepsilon \) denotes errors. In MAP method, it is supposed that errors for data points are independent of each other and identically distributed with a certain standard deviations of probability density function for all the points. Under these assumptions, the likelihood of Gaussian density for all data set satisfies

\[
P \propto \prod_{i=1}^{N} \exp \left\{ -\frac{1}{2} \left( \frac{y_i - y(x_i)}{\sigma} \right)^2 \right\} \Delta y_i.
\]
Maximizing equation (32) is equivalent to maximizing its logarithm, and it is also equal to minimizing its negative logarithm. Therefore, final function for minimizing discrepancies between observed data and model values is expressed by

\[
\sum_{i=1}^{N} \frac{[y_i - y(x_i)]^2}{2\sigma^2} - N \log \Delta y.
\]  

(33)

Note that \(N\), \(\sigma\), and \(\Delta y\) are all constants. Therefore, it indicates that this formulation obtained by the maximum likelihood estimation is equivalent to least squares norms in respect of minimizing the sum of squares.
2.3. Long-tailed Student’s t misfit

Student’s t-distribution is a continuous probability distribution that is applied for small sample size or unknown standard deviation. It is used for many statistical analysis, including the Student’s t-test, the construction of confidence intervals, and linear regression analysis. In this study, we introduce Student’s t-density to construct an objective function for inversion algorithm.

The t-density is a heavier tailed or ‘overdispersed’ compared to the Normal density, and has higher chance of extreme values in the tails. This makes it useful for a robust alternative to the Normal in case of suspected outliers in the data set (Congdon, 2001). Student's t-density has the form as follows:

\[
f(t) = \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\sqrt{\nu \pi \Gamma\left(\frac{\nu}{2}\right)}} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{34}\]

where \(\nu\) is the parameter of degrees of freedom and \(\Gamma\) is the Gamma function. This may also be written as

\[
f(t) = \frac{1}{\sqrt{\nu} B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{35}\]

where \(B\) is the Beta function.
Under Student’s t-distribution, unstandardized Student’s t-distribution has a density as

\[ P(x | \nu, \mu, \sigma) = \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sigma\sqrt{\nu\pi}} \left(1 + \frac{1}{\nu}\left(\frac{x - \mu}{\sigma}\right)^2\right)^{-\frac{\nu + 1}{2}}. \] (36)

It is conventionally expressed as \( x \sim t_\nu(\mu; \sigma^2) \), where \( \mu \) is a location parameter, \( \sigma \) is a scale parameter (Jackman, 2009). When \( \mu \) is zero and \( \sigma \) is 1 in an unstandardized t-density, it approaches to the standardized version of the t-density. The degrees of freedom parameter determines the extent of overdispersion. As \( \nu \) increases, Student’s t-distribution tends to be Normal approximations decreasing its variance. In other words, when \( \nu \) approaches infinity, t-density converges pointwise to the density of \( N(0,1) \). Also when \( \nu \) equals 1, t-density behaves the Cauchy density. The t-distribution itself is not log-concave, but a truncated t-distribution will be log-concave on the central interval \([ -\sqrt{\nu}, \sqrt{\nu}] \) (Bagnoli and Bergstrom, 2004).

Now, Student’s t-distribution can be considered to be the same approach as least-squares method from the point of statistical view. In the same assumptions, modeled data can be written by

\[ \mathbf{d} = \mathbf{u}(\mathbf{x}; \mathbf{m}) + \mathbf{\varepsilon}, \] (37)

where \( \mathbf{d} \) is a vector of \( N \) observations, \( \mathbf{x} \) is a vector of predictor variables, \( \mathbf{m} \) is a vector of \( M \) unknown parameters, and \( \mathbf{\varepsilon} \) is a vector of errors.

As parameters are unknown, the residuals can be calculated to see how far our predictions deviate from the data:

\[ \mathbf{r} = \mathbf{d} - \mathbf{u}(\mathbf{x}; \mathbf{m}). \] (38)

The residuals are used for evaluating any particular set of parameter estimates.
To apply the ML method, probability distribution for the residuals should be assumed. For unstandardized Student’s t-distribution, supposed that the discrepancies follow Student’s t distribution, a density function for the residuals is written by:

$$
f(r) = \frac{\Gamma\left(\frac{v + 1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sigma\sqrt{v\pi}} \left(1 + \frac{1}{v}\left(\frac{r - \mu}{\sigma}\right)^2\right)^{-\frac{v+1}{2}}. \quad (39)
$$

Total likelihood can be calculated with its density function by substituting to:

$$
l(r) = l(d - u(x;m)) = \prod_{j=1}^{n} f(r_j;m). \quad (40)
$$

For convenience, logarithm function can be replaced with its original form. That is why maximizing original forms is equivalent to maximizing its logarithm. Also if the observations are independent and identically distributed with density for each observation, the likelihood function can be written by

$$
l(r;m) = \prod_{j=1}^{n} f(r_j;m)
\Rightarrow L(r;m) = \sum_{j=1}^{n} \log f(r_j;m). \quad (41)
$$

Notice that the natural logarithm of the likelihood function $L(r;m)$ is commonly used, which is called log-likelihood function (Nocedal and Wright, 2006).

However, we seek model parameter values that minimize the differences between model responses and observations. Hence, we take negative log likelihood as

$$
m = \arg \max_{m \in \Theta} L(r;m) = \arg \min_{m \in \Theta} \left[-\log l(r;m)\right], \quad (42)
$$
Substituting Student’s t-density for the residuals to negative log likelihood gives

\[ L(r; m) = \sum_{j=1}^{n} -\log f(r_j) = \sum_{j=1}^{n} -\log \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sigma\sqrt{\pi}} \left(1 + \frac{1}{\nu} \left(\frac{r_j - \mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}. \]  

Also, we assume that the mean \( \mu \) is zero as

\[ L(r; m) = \sum_{j=1}^{n} -\log \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sigma\sqrt{\pi}} \left(1 + \frac{1}{\nu} \left(\frac{r_j}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}. \]  

From this, negative log likelihood of Student’s t-density can be used as an objective function for inversion in the same way as other norms:

\[ E(m) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left[-\log \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sigma\sqrt{\pi}} + \frac{\nu+1}{2} \log \left(1 + \frac{1}{\nu\sigma^2} (u_{ij}(m) - d_{ij})(u_{ij}(m) - d_{ij})\right)\right]. \]  

In this objective function, the degrees of freedom parameter is chosen as a small value, such as 2 or 3, to make the tails as thick as possible.
3. FULL WAVEFORM INVERSION ALGORITHM

3.1. Gradient direction

The waveform inversion was performed based on a 2D elastic wave modeling using the finite element method in the frequency domain. In the gradient method, the direction of each parameter is determined by the gradients of the misfit function at each iteration (Pratt, 1998). The parameter vector was updated iteratively as follows:

$$\mathbf{m}^{(p+1)} = \mathbf{m}^{(p)} - \alpha^{(p)} \nabla_m E^{(p)},$$  \hspace{1cm} (47)

where $p$ is the iteration number, and $\alpha$ is the step length. There are several objective functions that can be minimized for one parameter to decide the gradients. Generally, the gradient direction can be calculated by taking partial derivatives of the objective function.

From the previous mention of three objective functions, the gradients can be written by

$$-\nabla_{m_{\ell}} E_{L^2} (\mathbf{m}) = - \text{Re} \left[ \sum_{i=1}^{n_r} \sum_{j=1}^{n_r} \frac{\partial u_{ij}(\mathbf{m})}{\partial m_k} (u_{ij}(\mathbf{m}) - d_{ij}) \right]$$ \hspace{1cm} (48)

for the conventional $L_2$-norm, and

$$-\nabla_{m_{\ell}} E_{L^1} (\mathbf{m}) = - \text{Re} \left[ \sum_{i=1}^{n_r} \sum_{j=1}^{n_r} \frac{\partial u_{ij}(\mathbf{m})}{\partial m_k} \left\{ \text{sgn} \left( \text{Re} \left[ u_{ij}(\mathbf{m}) - d_{ij} \right] \right) \right\} \right]$$ \hspace{1cm} (49)

for the least-absolute inversion using the $L_1$-norm criterion, and
for the ML estimation using Student’s t-distribution, where \( m_k \) is the \( k \)-th element of the model parameter \( m \) and \( \text{sgn}(\cdot) \) is the signum function. To calculate gradient directions efficiently, the gradients are computed by the cross-correlation of back-propagated residuals and virtual sources:

\[
\nabla_{m_k} E = \sum_k \Re \left[ (F^v)^T (S^{-1})^T r^* \right],
\]

where

\[
F^v = \begin{bmatrix} f^v_1 & f^v_2 & \cdots & f^v_k & \cdots & f^v_{w_p} \end{bmatrix},
\]

\[
f^v_{ij} = -\frac{\partial S}{\partial m_{ij}} u_{ij}.
\]

Finally, the residuals for each norm can be represented as follows:

\[
r_{L_2,ij} = u_{ij}(m) - d_{ij},
\]

\[
r_{L_1,ij} = \text{sgn} \left( \Re \left[ u_{ij}(m) - d_{ij} \right] \right) + i \text{sgn} \left( \Im \left[ u_{ij}(m) - d_{ij} \right] \right),
\]

\[
r_{T,ij} = \frac{1}{2} \frac{2(u_{ij}(m) - d_{ij})}{\nu \sigma^2}.
\]

\[
\nu + 1 \sum_{i=1}^{w} \sum_{j=1}^{w} \frac{\partial u_{ij}(m)}{\partial m_k} \frac{1}{1 + \frac{(u_{ij}(m) - d_{ij})(u_{ij}(m) - d_{ij})}{\nu \sigma^2}} \]

\[
(50)
\]

where
3.2. Scaling method

To improve convergence, the gradient needs to be scaled. The approximate Hessian matrix is used in the Gauss-Newton method, and the full Hessian matrix is used in the full Newton method in order to accelerate convergence of iterative solutions (Pratt et al., 1998). However, these methods have some disadvantages that they require a great amount of computer memory. Instead of the approximate Hessian, pseudo-Hessian matrix can be applied for scaling the gradient direction (Shin et al., 2001). The pseudo-Hessian matrix shows a similar pattern to the approximate Hessian, which only uses the virtual sources.

Gradient direction scaled by the pseudo-Hessian matrix can be expressed as

\[
\nabla E = \sum_{s} \left( \frac{\sum_i \text{Re} \left[ \left( F_s^T \left( S_i^{-1} \right)^T r_s^* \right] \right)}{\sum_s \text{Re} \left[ \text{diag} \left( \text{Re} \left[ \left( F_s^T F_s + \lambda I \right) \right] \right) \right]} \right),
\]

where \( \lambda \) is the damping factor, and \( I \) is the identity matrix.

3.3. Conjugate gradient method

In order to accelerate the convergence of inversion, the conjugate gradient method is used (Ha et al., 2009; Brandsberg-Dahl et al., 2003). At each iteration \( k \), the search direction \( p_k \) is determined by previous search direction \( p_{k-1} \) and current gradient vector \( g_k = \nabla E \) as follows:

\[
p_k = -g_k + \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} p_{k-1}.
\]

Finally, the updated model parameter \( m_{k+1} \) can be expressed by

\[
m_{k+1} = m_k + \alpha_k p_k,
\]

where \( \alpha \) is the step length and \( m \) is the model parameter.
4. NUMERICAL EXAMPLES

4.1. Elastic Marmousi-2 model without noises

To compare the three objective functions (the $l_2$- and $l_1$-norms and Student’s t-distribution) with each other, we perform the elastic waveform inversion for the 2D Elastic Marmousi-2 data set (Martin et al., 2002). The P-wave velocity of the Marmousi-2 model is shown in Figure 1. The size of the model is 9.2 km × 3 km and the grid interval for both vertical and horizontal directions is 0.02 km. The density and the Poisson’s ratio are fixed at 2 g·cm$^{-3}$ and 0.25. Receivers are located at all the nodal points of the surface and the number of shots is 231 with an interval of 0.04 km. The first derivative of a Gaussian function with maximum frequency of 10 Hz is used as a source wavelet. The total recording duration is 5 seconds. Synthetic seismograms for the Marmousi-2 model are generated using the finite-element method in the frequency domain (Figure 2).

In waveform inversion, an initial model is chosen by a linearly increasing velocity model ranging from 1.5 km/s to 4.5 km/s with depth (Figure 3a). Step length is 0.05. We carry out the elastic waveform inversion using the $l_2$-norm, the $l_1$-norm, and Student’s t misfit to the Marmousi-2 model, and each result is displayed in figure 3b, 3c, and 3d. All inversion results obtained by the three objective functions are in good agreement with the true P-wave velocity. Figure 4 shows depth profiles extracted at a distance of 3 km of the true and the inverted models. In figure 4, it is noted that the inverted P-wave velocities are very close to true values of the Marmousi-2 model (Figure 4).
Figure 1 P-wave velocity of the 2D elastic Marmousi-2 model (Martin et al., 2002).

Figure 2 Seismograms obtained for the 2D elastic Marmousi-2 model: (a) horizontal and (b) vertical displacements.
Figure 3 (a) Initial model and P-wave velocities inverted by the (b) $l_2$-norm, (c) $l_1$-norm, and (d) Student’s t-distribution at the 300th iteration for synthetic data without noise for the Marmousi-2 model.
Figure 3 (Continued)
**Figure 4** Depth profiles at a distance of 3 km extracted from the true and inverted models (Figure 3) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration.
Figure 5 Depth profiles at a distance of 6 km extracted from the true and inverted models (Figure 3) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration.
4.2. Elastic Marmousi-2 model with outliers

According to Bayesian estimation theory, the choice of t-density is useful for data sets with extreme values included in the tails. We generated data for the Marmousi-2 model with outliers added. Ten outliers, whose amplitude is 150% of the maximum amplitude of signals, are randomly added to the original data set (Figure 6). Figure 7 shows traces extracted at a distance of 1.82 km. In figure 7, outliers are observed around 4 seconds.

Inversion results for the $l_2$-norm, $l_1$-norm, and Student’s t objective function are shown in figure 8. The velocity inverted by the $l_2$-norm is severely distorted by the outliers at the location where outliers were inserted. In contrast, the other two results generated by the $l_1$-norm and Student’s t misfit are comparable to the Marmousi-2 model, which are very similar to those for outlier-free data. Compared to the $l_1$-norm, Student’s t misfit yields better results especially for the deep parts (Figure 8b and 8c). It means that the results obtained by Student’s t misfit are less influenced by outliers. Figure 8 shows the depth profiles extracted from the models inverted by the $l_2$-norm, $l_1$-norm, and Student’s t misfit. Depth profiles support the aforementioned phenomena.
Figure 6 Seismograms with 10 outliers randomly distributed over the seismogram for the Marmousi-2 model: (a) horizontal and (b) vertical displacements.
Figure 7 Seismogram with outliers recorded at a distance of 1.82 km for (a) horizontal and (b) vertical displacements. Red arrows indicate outliers.
Figure 8 P-wave velocities inverted by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration for synthetic data with outliers for the Marmousi-2 model.
Figure 9 Depth profiles at a distance of 3 km extracted from the true and inverted models (Figure 8) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration.
Figure 10 Depth profiles at a distance of 6 km extracted from the true and inverted models (Figure 8) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration.
4.3. Elastic Marmousi-2 model with outliers and random noises

We apply the $l_2$- and $l_1$-norm and Student’s t misfit for data with both outliers and random noise. In the frequency domain, random noise was applied to the forward data sets as a background noise. Random noise ranges from zero to 0.25 % of the maximum energy of the data. Figure 11 shows the seismograms with random noise.

Figure 12 shows P-wave velocity results inverted by the $l_2$-norm, $l_1$-norm, and Student’s t misfit when both outliers and random noise are applied. All the results are distorted by outliers and random noise. For $l_2$-norm, partial distortions affected by outliers are dominant in figure 12(a). Figure 12(a) is very similar to figure 8(a). Also, the results for $l_1$-norm are better than those of the $l_2$-norm. However, due to random noise, the image is poorer than figure 8(b), but the $l_1$-norm is seen to be less influenced by random noise compared to the $l_2$-norm and Student’s t misfit. In Student’s t misfit, the distortions caused by random noise appear over the entire P-wave velocity (Figure 8c). On the other hand, the results are rarely affected by outliers. The depth profiles extracted at distances of 3 km and 6 km are shown at figure 13 and figure 14.
Figure 11 The real part of forward data when 0.25% random noise was added to background including outliers at 10 Hz: (a) horizontal and (b) vertical displacements.
Figure 12 P-wave velocities inverted by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration for synthetic data with outliers and random noises for the Marmousi-2 model.
Figure 13 Depth profiles at a distance of 3 km extracted from the true and inverted models (Figure 12) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration.
Figure 14 Depth profiles at a distance of 6 km extracted from the true and inverted models (Figure 12) obtained by the (a) $l_2$-norm, (b) $l_1$-norm, and (c) Student’s t-distribution at the 300th iteration.
CONCLUSIONS

We proposed the elastic waveform inversion based on Student’s t-distribution. Student’s t density has the similar distribution form to Normal density, but its heavier tail enables Student’s t-distribution to include a higher chance of extreme values. In order to apply Student’s t-distribution to inversion algorithm, MAP and ML techniques were used to build the objective function. For the waveform inversion algorithm, gradient method was basically chosen. For the scaling method, the pseudo-Hessian was used, and conjugate gradient method was applied to accelerate the convergence.

Numerical examples for the Marmousi-2 model showed that Student’s t misfit was able to be successfully applied for elastic inversion for the data with and without outliers. To investigate the reliability of the inversion results, we compare inversion results obtained by Student’s t results to those of $l_2$-norm and $l_1$-norm FWI. For the data set without outliers, all the objective functions yield good results. However, the $l_2$-norm yielded the distorted results around location where outliers were added. However, the $l_1$-norm and Student’s t misfit gave reliable results. In particular, Student’s t misfit decreased the influence of large outliers on inversion results for the deep structures.

From those results, it can be expected that other statistical distributions can be applied to inversion algorithm like Student’s t distribution. There are many statistical distributions which have their own advantages and disadvantages. They could be utilized for the inversion of many kinds of data set if we would know each distinct property as new objective functions.
REFERENCES


국문 초록
스튜던트의 티 분포를 이용한 주파수영역 탄성파 완전파형역산

강 민 지
공과대학 에너지시스템공학부
서울대학교 대학원

탄성파 완전파형역산은 임의로 가정한 초기모델에 대하여 수치모델링으로 계산한 이론값과 실제 탐사에서 기록된 값들의 오차가 최소가 되는 방향으로 지하지질모델을 수정해 나갔으며, 현장자료에 가장 잘 부합하는 지하모델을 유추해 내는 과정이다. 이러한 탄성파 완전파형역산에는 일반적으로 수치모델링과 현장자료의 오차의 최소자승값을 목적함수 ($l_2$-norm)로 이용하는 방법이 주로 이용되어왔으나 이 방법은 진폭이 큰 스파이크 형태의 잡음 (outliers)이 존재하는 자료에 대해서 안 좋은 역산결과를 제시하는 것으로 알려져 있다. 이에 이 연구에서는 스튜던트의 티 분포 (Student’s $t$-distribution)에 기초한 탄성파 완전파형역산을 제안한다. 티 분포는 정규분포와 비슷한 모양을 가지지만 더 넓게 퍼져있는 모양의 밀도함수를 가지며, 스파이크 형태의 잡음이 포함된 자료에서 통계적으로 유용하게 사용될 수 있는 특징이 있다. 이 연구에서 제안한 티 목적함수에는 통계적 기법인 최대우도추정법이 적용되었고, 탄성파 완전파형 역산 알고리즘에는 유한요소법에 기초한 모델링 알고리즘, 역산의 효율성 및 수렴속도를 높이기 위한 역전과 알고리즘, 유사해시안 행렬 (pseudo-Hessian matrix), 복소구배법 (conjugate gradient method)이 이용되었다.
티 목적함수를 이용한 과정역산 알고리즘의 검증을 위해 마모시-2 모델에 대해 수치모델링을 수행하였다. 밀도는 2 \( g \cdot cm^{-3} \)로, 포아송비는 0.25로 고정하였으며 초기모델은 깊이에 따라 값이 증가되도록 선형적으로 설정하였다. 비교를 위해 \( l_2\)-norm과 이론적 계산값과 실제자료의 오차의 절댓값을 이용하는 \( l_1\)-norm을 이용한 역산을 함께 수행하였다. 역산 결과 이상값이 포함되어 있지 않은 경우 세 목적함수 모두 큰 차이 없이 실제 모델값을 잘 찾는 것으로 나타났다. 이상값이 포함되어 있는 자료에 대해서는 \( l_1\)-norm과 티-목적함수의 결과가 \( P \)파 속도 분포를 잘 유추해낸 반면 \( l_2\)-norm은 역산결과가 심하게 왜곡되었다. 또한 이상값뿐만 아니라 무작위잡음도 함께 포함된 경우에서는 티 목적함수를 이용한 결과가 \( l_1\)-norm보다 속도값을 더 정확하게 근사하는 것으로 나타났다. 이 연구를 통해 티 목적함수가 다른 목적함수에 비해 잡음의 영향을 덜 받는다는 것을 알 수 있었다. 추후 티 분포뿐만 아니라 여러 분포함수를 보다 다양한 통계학적 기법을 이용하여 역산 알고리즘에 적용할 수 있을 것으로 기대된다.

주요어: 탄성파 완전과정역산, 탄성파동방정식, 스튜던트의 티 분포, 기울기 방향
학번: 2011-21089