NEW INVERSION ALGORITHM FOR THE COMPUTATION OF FOURIER TRANSFORM – EXAMINATION ON A SYNTHETIC DATA SET

HAJNALKA SZEGEDI–MIHÁLY DOBRÓKA–TAMÁS BODOKY
Department of Geophysics
University of Miskolc

ABSTRACT
In order to make Fourier transform more robust and noise resistant, the tools of geophysical inverse problem theory are used. Assuming that the Fourier spectrum is given in the form of series expansion and the expansion coefficients can be determined by solving an over-determined inverse problem. The Iteratively Reweighted Least Squares (IRLS) method was used during the inversion process. Using it the geophysical inversion methods show robustness. The behavior of the new inversion-based Fourier transform method is very effective against noises.

INTRODUCTION
With a view to changing over from time domain to frequency domain, the Fourier transform is applied most frequently. The Discrete Fourier Transform (DFT) algorithm is applicable to determine the discrete frequency spectrum, if we have a discrete time domain data set. As the measured data set always contains noise, which is also projected into the frequency domain, the calculated frequency spectrum will be noisy, too. In order to decrease the influence of noise, either the number of input data must be increased. If we have more input data than inversion parameters, the inverse problem is called over-determined. P. Vass in [1] treats the Fourier transformation as an over-determined inverse problem, where the influence of noise on the spectrum can be effectively reduced by solving it. This paper presents a new robust inversion method for performing one-dimensional Fourier transform. The new method is based on series expansion of the spectrum, where scaled Hermite functions are used as basis functions. Using the special feature of the basic Hermite functions (they are the eigen-functions of the Fourier transform) the Jacobian matrix can be determined easily. It means that the element of the Jacobian matrix can be calculated without integration. The new inversion method was numerically tested on a noisy synthetic data set. The inversion results show appropriate accuracy and high noise suppression capability of the inversion-based Fourier transform algorithm.

THE THEORETICAL BACKGROUND OF THE METHOD
The Fourier transform plays a very important role in geophysical data processing. Using it we can generate the frequency spectrum of the time domain signal

\[ U(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(t) \cdot e^{-j\omega t} dt, \quad (1) \]

where \( u(t) \) denotes the time dependent function and \( U(\omega) \) is the complex frequency spectrum.

By means of the inverse Fourier transform we can return from the frequency domain to the time domain

\[ u(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U(\omega) \cdot e^{j\omega t} d\omega. \quad (2) \]
In inverse problem theory various methods have been developed to extract information from noisy data sets. This can give the idea to use the method of inverse problem theory in the field of Fourier transform or in other word treatment of the Fourier transform as an over-determined inverse problem. In formulating the inverse problem the first step is the discretization. Assuming that the Fourier spectrum is given in the form of series expansion

\[ U(\omega) = \sum_{n=1}^{M} B_n \cdot \Psi_n(\omega), \]  

where \( B_n \) represents the complex expansion coefficients, \( \Psi_n(\omega) \) is the \( n \)-th known basis function and \( M \) is the number of unknown series expansion coefficients. Using the terminology of discrete inverse problem theory, the theoretical values of time domain data (forward modeling) can be given by the inverse Fourier transform

\[ u^{\text{theor}}(t_k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left( \sum_{n=1}^{M} B_n \cdot \Psi_n(\omega) \right) \cdot e^{i\omega t_k} \, d\omega = \sum_{n=1}^{M} B_n \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Psi_n(\omega) \cdot e^{i\omega t_k} \, d\omega, \]  

where \( t_k \) is the \( k \)-th sampled time. Thus, we have obtained a very simple direct problem. Let us introduce the Jacobian matrix

\[ G_{k,n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Psi_n(\omega) \cdot e^{i\omega t_k} \, d\omega = \mathcal{F}^{-1}\{\Psi_n(\omega)\}, \]  

which is the inverse Fourier transform of the basis function system \( \Psi_n(\omega) \). We can see that the direct problem make a relationship between the model parameters and the theoretically approximated measurement data. Then the theoretical data can be written in a linear form as

\[ u^{\text{theor}}(t_k) = \sum_{n=1}^{M} B_n \cdot G_{k,n}, \]  

where the unknown model parameters \( B_n \) are estimated by an inversion method.

**CHOOSING BASIS FUNCTION SYSTEM - THE HERMITE FUNCTIONS**

The basic Hermite functions have been chosen as basis functions because in case of inverse problems it is advised to use complete, orthogonal and normal function systems in order to reduce the number of unknown parameters and to improve the stability of the inversion procedure. The description of the basic Hermite functions can be found in [2–3]. The Hermite functions require proper scaling, because in geophysical applications the measurement of frequency covers a wide range. The scaled Hermite polynomials can be calculated by using the Rodriguez formula

\[ h_n(\omega, \alpha) = (-1)^n \, e^{\alpha^2 \omega^2} \left( \frac{d}{d\omega} \right)^n e^{-\alpha^2 \omega^2} \]  

and can be determined by the recursive formula

\[ h_{n+1}(\omega, \alpha) = 2\omega\alpha \, h_n(\omega, \alpha) - 2n\alpha \, h_{n-1}(\omega, \alpha), \]  

where \( \alpha \) is a scaling factor. The first and the second Hermite polynomials are shown below [4]:

\[ h_0(\omega, \alpha) = 1, \]  

\[ h_1(\omega, \alpha) = 2\alpha \omega. \]
The scaled Hermite polynomials fulfill the condition of orthogonality
\[ \int_{-\infty}^{\infty} e^{-\alpha \omega^2} \cdot h_n(\omega, \alpha) \cdot h_m(\omega, \alpha) d\omega = \sqrt{\frac{\pi}{\alpha}} \frac{(2\alpha)^n}{n!} \delta_{nm}, \delta_{nm} = \begin{cases} 0, & n \neq m \\ 1, & n = m. \end{cases} \] (11)

In the formula \( \delta_{nm} \) denotes the Kronecker-delta symbol. The scaled Hermite function can be defined as
\[ H_n(\omega, \alpha) = e^{-\frac{\alpha \omega^2}{2}} \cdot h_n(\omega, \alpha), \] (12)

where \( h_n(\omega, \alpha) \) denotes the scaled Hermite polynomials. Thus the introduced Hermite functions are orthonormal
\[ \int_{-\infty}^{\infty} H_n(\omega, \alpha) \cdot H_m(\omega, \alpha) d\omega = \delta_{nm}. \] (13)

In order to get a fast and simple formula for the calculation of the Jacobian matrix a special feature of the Hermite functions was used. As we can see earlier, the \( G_{k,n} \) is the inverse Fourier transformation of the basis function. This is the main reason of selected the Hermite functions, because their the non-scaled version are eigen-functions of the Fourier transform [5]
\[ \mathcal{F}\{ h_n^{(0)}(t) \} = (-j)^n h_n^{(0)}(\omega) \] (14)

and can be written for the inverse Fourier transform
\[ \mathcal{F}^{-1}\{ h_n^{(0)}(\omega) \} = (j)^n h_n^{(0)}(t). \] (15)

Based on Eq. (5) and by the scaled Hermite functions \( H_n(\omega, \alpha) \) can be written for the Jacobian matrix
\[ G_{k,n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H_n(\omega, \alpha) \cdot e^{j\omega t_1} d\omega = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{\alpha} H_n^{(0)}(\omega') \cdot e^{j\omega t_1} d\omega' \] (16)

Introducing the following notations
\[ \omega t = \omega t', \omega' = \sqrt{\alpha} \omega, t' = \frac{t}{\sqrt{\alpha}} \] (17)

it can be computed easily
\[ G_{k,n} = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty} H_n^{(0)}(\omega') \cdot e^{j\omega t_1} d\omega' = \frac{1}{\sqrt{\alpha}} \mathcal{F}^{-1}\{ h_n^{(0)}(\omega') \}. \] (18)

The Jacobian matrix can be written again based on Eq. (15)
\[ G_{k,n} = \frac{1}{\sqrt{\alpha}} \left( j \right)^n H_n^{(0)}(t') = \frac{1}{\sqrt{\alpha}} \left( j \right)^n H_n^{(0)} \left( \frac{t}{\sqrt{\alpha}} \right). \] (19)

This is a very important result, because the Jacobian matrix can be derived quickly without integration.

The discretized form of the spectrum can be written according to Eq. (3) using the scaled Hermite function system, where the expansion coefficients \( B_n \) are defined (including the expression of the Jacobian matrix, Eq. (19)) in the frame of the over-determined inverse
problem. This inverse problem is highly over-determined, because the number of measurement data is much more than that of the parameters \((N>M)\). In case of Least Squares method (LSQ) the \(L_2\) norm of the deviation vector is minimized

\[
E_2 = \sum_{k=1}^{N} e_k^2 = \sum_{k=1}^{N} (u_k^{(\text{measured})} - u_k^{(\text{theor})})^2 = \sum_{k=1}^{N} (u_k^{(\text{measured})} - \sum_{m=0}^{M} B_m \cdot G_{k,m})^2 = \text{min}.
\] (20)

The well-known normal equation can be derived from the above condition in the following form

\[
G^T G \tilde{B} = G^T \tilde{u}^{\text{measured}}.
\] (21)

After this we can estimate the complex series expansion coefficients

\[
\tilde{B} = (G^T G)^{-1} G^T \tilde{u}^{\text{measured}}
\] (22)

and by means of it, the real and imaginary part of the Fourier spectrum can be calculated at any frequency by using Eq. (3) [1]. The LSQ method gives optimal results in case of Gaussian distributed data set.

In order to make the Fourier transform more robust an Iteratively Reweighted Least Squares (IRLS [6]) method using Cauchy-weights was implemented. To achieve the optimal values of the unknown parameters \((B_m)\) the following weighted norm is minimized

\[
E_w = \sum_{k=1}^{N} w_k e_k^2,
\] (23)

where the weighting matrix (which is a diagonal matrix) defined as

\[
w_k = \frac{\epsilon^2}{\epsilon^2 + e_k^2},
\] (24)

and the \(k\)-th element of the deviation vector is

\[
e_k = u_k^{\text{measured}} - u_k^{\text{theor}}.
\] (25)

The scale parameter \(\epsilon\) of the Cauchy distribution is not a priori given because the data residuals change from iteration to iteration. The lower bound of the scale parameter is given by [7]. The weighted norm gives reliable results for inverse problems even if the measured data set contains outliers. The \(j\)-th iteration step of the normal equation is

\[
G^T W^{(j-1)} G B^{(j)} = G^T W^{(j-1)} \tilde{u}^{\text{measured}}.
\] (26)

This iteration is repeated until a proper stop criterion is met. Finally, the Fourier spectrum can be calculated at any frequency by using Eq. (3).

**TESTING THE METHOD**

In order to test the new inversion method a synthetic time domain data set was generated. The noiseless time function of the tested data was chosen in this form

\[
u(t) = \begin{cases} 
K t^n e^{-at} \sin(\omega t + \varphi), & t \geq 0 \\
0, & t < 0
\end{cases}
, \quad (27)
\]
where the Greek letters represents the parameters of the signal and these values were fixed during the procedure: \( \kappa \approx 738.91, \eta = 2, \lambda = 20, \omega = 2 \times 20\pi, \phi = \pi/4. \)

The noiseless time function is shown in Figure 1, where the time variable was sampled equidistantly to 256 points over the interval \([0, 1]\).

The DFT of this time domain signal was computed. The real and imaginary part of the ideal spectrum can be seen in Figure 2.

In order to test the noise sensitivity of DFT a Cauchy distributed noisy signal was generated with the location parameter 0 and the scale parameter 0.04. The noisy signal is shown in Figure 3 and the DFT of the noisy signal is represented in Figure 4.

As we can see the real and imaginary part of the frequency spectra are very noisy compared to the noiseless data set (see Figure 2), and it is hard to recognize the ideal spectrum from Figure 4.
The new inversion-based Fourier transform method was numerically tested using the signal contaminated with non-Gaussian noise (Figure 3). The inversion results are illustrated in Figures 5–6.

These figures show that the IRLS method provides much better estimates in both the time (Figure 6) and frequency domains (Figure 5).

In order to characterize the accuracy of the inversion results the distance between the measured and calculated data was computed

\[
    d_{RMS} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (u_k^{measured} - u_k^{calculated})^2}. \tag{28}
\]

In the frequency domain for characterizing the deviations the root mean square of the errors (RMS) can be written in the form below

\[
    D_{RMS} = \left[ \frac{1}{N_f} \sum_{i=1}^{N_f} \left( \text{Re} \left[ U^{calculated}(\omega_i) \right] - \text{Re} \left[ U^{measured}(\omega_i) \right] \right)^2 + \frac{1}{N_f} \sum_{i=1}^{N_f} \left( \text{Im} \left[ U^{calculated}(\omega_i) \right] - \text{Im} \left[ U^{measured}(\omega_i) \right] \right)^2 \right]^{\frac{1}{2}}. \tag{29}
\]

where \( U \) is the Fourier spectrum, \( N_f \) implies the number of compared data pairs in the frequency domain.

Table 1 summarizes the values of RMS for both domains. If we compare the values of the two columns, a significant difference can be noticed. The higher values in the first column show that the noise caused a great deviation from the noiseless signal. The IRLS algorithm could reduce the effect of noise which is mirrored by the smaller values of RMS in both domains.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>deviation between noisy and noiseless data</th>
<th>deviation between inverted and noiseless data</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_{RMS} ) (time domain)</td>
<td>0.103</td>
<td>0.035</td>
</tr>
<tr>
<td>( D_{RMS} ) (frequency domain)</td>
<td>0.0107</td>
<td>0.0055</td>
</tr>
</tbody>
</table>
CONCLUSIONS

A new inversion algorithm for the one-dimensional Fourier transformation was presented. Series expansion was used for the discretization of the continuous functions of the spectrum where the scaled Hermite functions were chosen as the set of basis functions. The unknown parameters are defined in an over-determined inverse problem. Using the special property of the basic Hermite functions (they are the eigen-functions of the Fourier transform) the Jacobian matrix can be determined easily.

The IRLS algorithm provided the best accuracy of the estimation in case of outliers which can be seen in Figure 5–6. The proposed new inversion method has reduced the noise sensitivity of Fourier transformation even in case of having outliers.

ACKNOWLEDGEMENTS

The described work was carried out as part of the TÁMOP-4.2.2/B-10/1-2010-0008 project in the framework of the New Hungary Development Plan. The realization of this project is supported by the European Union, co-financed by the European Social Fund.

The author is also grateful for the support of the research team of the Department of Geophysics, University of Miskolc.

LIST OF SYMBOLS

- \( B \) series expansion coefficients
- \( B_n \) complex expansion coefficients
- \( D_{\text{RMS}} \) deviations the root mean square of the errors in the frequency domain
- \( E_2 \) objective function
- \( E_w \) weighted norm
- \( G \) Jacobian matrix
- \( G_{k,n} \) Jacobian matrix
- \( (G^T G)^{-1} \) inverse matrix
- \( H_n(\omega, \alpha) \) scaled Hermite function
- \( M \) number of unknown series expansion coefficients
- \( N \) number of measured data
- \( N_f \) number of compared data pairs in the frequency domain
- \( \text{RMS} \) root mean square
- \( U \) Fourier spectrum
- \( U(\omega) \) complex frequency spectrum
- \( W^{(j-1)} \) j-th iteration step of the weight
- \( d_{\text{RMS}} \) distance between the measured and calculated data
- \( e_k \) k-th element of the deviation vector
- \( h_n(\omega, \alpha) \) scaled Hermite polynomial
- \( j \) imaginary unit
- \( t_k \) k-th sampled time
- \( u(t) \) time dependent function
- \( u_{\text{theor}} \) calculated data
- \( u_{\text{measured}} \) measured (original) data
- \( w_k \) weighting matrix
- \( \Psi_n(\omega) \) n-th known basis function
REFERENCES