Full-waveform inversion using an efficient preconditioning method for the gradient vector

Peterson Nogueira Santos *, CPGG/UFBA, Reynam C. Pestana, CPGG/UFBA and INCT-GP/CNPq.

This paper was prepared for presentation at the 15th International Congress of the Brazilian Geophysical Society, held in Rio de Janeiro, Brazil, 31 July to 3 August 2017, 2017.

Abstract

Full-waveform inversion (FWI) is a powerful method and it has been used successfully to invert subsurface parameters. It consists basically on the minimization of the difference between the predicted and observed data. However, its application using finite-difference schemes is limited to low frequency content and the increase of the range of higher frequency will demand a high computational cost of the wavefield propagation procedure and also the whole inversion scheme. To overcome this problem, we apply the rapid expansion method (REM) for numerical wavefield extrapolation inside the FWI workflow thus increasing the frequency content of the inversion process. Besides that, an efficient preconditioning method for computing the gradient vector in order to increase its resolution has also been proposed.
To test the efficiency of our proposed FWI approach, we apply it using a frequency multiscale scheme for a synthetic data set with a complex velocity model. The inversion results show satisfactory inverted velocity models which can be used to produce depth imaging of high quality. Thus we demonstrate the effectiveness and applicability of our FWI scheme using REM combined with a multiscale approach .

Introduction

The Full-waveform inversion has been used more frequently due to ability to estimate model parameters with higher resolution when compared with the traditional methods. Consequently, it can be apply to solve complex depth imaging problem and thus produce accurate subsurface images. FWI is used traditionally for acoustic case in order to estimate the velocity field. The best inverted velocity model is the one that can predict the observed data, using the complete wave equation, and it will be match the actual observed data in terms of traveltimes, phase and amplitude.

The efficient solution of the direct problem is of great relevance in FWI, because the difference between observed and estimated data has to contain only information about the model. That is, the modeling procedure can not produce events that do not exist in the observed data. This procedure is performed with the extrapolation of the wave field in a physical model through a direct modeling operator. This operator should be implemented to include elastic modulus and density (Virieux and Operto, 2009), but it is in the current moment still computational expensive. Thus, propagation of the wave can be understood by considering a purely acoustic medium in which no transverse waves are propagated.

Theory

We consider the following acoustic wave equation, which predicts only the propagation of longitudinal waves:

$$\frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} + L^2 u(\mathbf{x},t) = f(\mathbf{x},t) \tag{1}$$

where $-L^2 = c^2(\mathbf{x})\nabla^2$, $c(\mathbf{x})$ is the velocity of propagation, $\mathbf{x} = (x,y,z)$ is the position vector and $\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)$ is the Laplacian operator in cartesian coordinates and $f(\mathbf{x},t)$ is the source term.

The approach that we use to solve equation 1 is called variations of parameters (VOP). Thus, the general solution of $u(\mathbf{x},t)$ to equation (1) on $[0,t]$ is written as:

$$u(\mathbf{x},t) = u_0 \cos(Lt) + \frac{\dot{u_0}}{L} sin(Lt) + \frac{1}{L}\int_0^t f(\mathbf{x},s)\sin[L(t-s)]ds \tag{2}$$

where $u(\mathbf{x},t=0) = u_0$ and $\frac{\partial u(\mathbf{x},t)}{\partial t} \mid_{t=0} = \dot{u_0}$.

Equation (2) is the fundamental equation from which we derive the integration procedure. Now, if equation 2 is reevaluated using the intervals $[t, t+\Delta t]$, $[t, t-\Delta t]$ and by adding them and evaluating the resulting integral, we obtain the following complete solution of 1, which includes the source term and is given by:

$$u(\mathbf{x},t+\Delta t) + u(\mathbf{x},t-\Delta t) = 2\cos(L\Delta t)u(\mathbf{x},t) + S(\mathbf{x},t\pm\Delta t) \tag{3}$$

where $S(\mathbf{x},t\pm\Delta t) = \frac{\Delta t^2}{2}[f(\mathbf{x},t+\Delta t) + f(\mathbf{x},t-\Delta t)]$.

Rapid Expansion Method

The rapid expansion method is an efficient way of numerically solving the acoustic wave equation (Pestana et al., 2009). This technique can extrapolate the wave field with higher frequencies and larger sampling intervals in time and space, thus being more stable and less dispersive than the conventional finite difference scheme.

Following Kosloff et al. (1989) and based on the expansion method presented by Tal-Ezer et al. (1987) the cosine function can be expanded as

$$\cos(L\Delta t) = \sum_{k=0}^{\infty} C_{2k} J_{2k}(\Delta t R) Q_{2k}\left(\frac{iL}{R}\right), \tag{4}$$

where $C_{2k} = 1$ for $k = 0$ and $C_{2k} = 2$ for $k > 0$, J_{2k} represents the Bessel function of order $2k$ and $Q_{2k}(w)$ are the modified Chebyshev polynomials. The term R is a scalar larger than the range of eigenvalues of $-L^2$ and it is the same R which appeared in the original Tal-Ezer method (Tal-Ezer et al., 1987).

Since (4) contains only even polynomials, it is more convenient to use the relation,

$$Q_{k+2}(w) = 2(1+2w^2)Q_k(w) - Q_{k-2}(w). \quad (5)$$

The recursion is initiated by

$$Q_0(w) = 1 \quad \text{and} \quad Q_2(w) = 1+2w^2, \quad (6)$$

where we have replaced w by iL/R.

For 2D wave propagation, and considering the constant velocity case, R is given by $R = \pi c\sqrt{(1/\Delta x^2)+(1/\Delta z^2)}$. In general c should be replaced by c_{max}, the highest velocity in the grid, and Δx, Δy and Δz are the spatial grid spacing (Tal-Ezer et al., 1987).

The sum in (4) is known to converge exponentially for $k > \Delta tR$ and the summation can be safely truncated with a k value slightly greater than ΔtR.

In this way, the stability of the REM is ensured, since the number of terms used in the expansion is directly proportional to the used time sampling interval. Thus, any Δt can be used, provided that the number of terms calculated will be sufficient to guarantee the convergence and the stability of the method. Therefore, the use of REM makes the FWI process with wave propagation more stable and free of numerical dispersion when high frequency components are inserted in the modeling of these wave fields (dos Santos and Pestana, 2015).

Time-domain multiscale full-waveform inversion

The application of the multiscale scheme is very crucial for the FWI method specially because it can prevent the inversion method converges to a local minimum. Here, in our case, we have used during the inversion procedure higher frequency data and this allows us to obtain inverted models with more details.

The full-waveform inversion method consists of iteratively improving an initial velocity model $(\mathbf{m})$. This optimization is based on the modeling of seismic waves through the solution of the direct problem, which offers the possibility to compute simultaneously the amplitude and the traveltimes of the waves. The solution of the inverse problem involves minimizing the objective function in the space of the model parameters to measure the difference between the predicted $(\mathbf{d}_{cal})$ and observed data $(\mathbf{d}_{obs})$ (Virieux and Operto, 2009). The inversion of the directly observed data is computationally impractical. Thus, FWI is formulated as a least squares type optimization problem, in which the objective function of norm l_2 representing the problem can be defined by:

$$F(\mathbf{m}) = \frac{1}{2}||\mathbf{d}(\mathbf{m})_{cal} - \mathbf{d}_{obs}||_2^2 = \frac{1}{2}\sum_{ns}\int_0^T (d_{cal} - d_{obs})dt, \quad (7)$$

where F is a measure of error, T is the maximum record time. The summation is done in the shot domain, where ns is the total number of shots. The purpose of this formulation is to find the model $\mathbf{m}$ so that the functional $F(\mathbf{m})$ is minimum.

This minimization is done in a recurrent way, that is, given a current model $\mathbf{m}_k$, we search for a next model $\mathbf{m}_{k+1}$, which theoretically should be closer to the true model.

The iterative process can be deduced using the second order Taylor-Lagrange approximation (Virieux and Operto, 2009). Assuming that **m** can be written as the sum of an initial model $\mathbf{m}_0$ and a perturbation in the model $\Delta\mathbf{m}$, we have:

$$F(\mathbf{m}) = F(\mathbf{m}_0 + \Delta\mathbf{m}) = F(\mathbf{m}_0) + \textstyle\sum_{j=1}^{M} \frac{\partial F(\mathbf{m}_0)}{\partial m_j}\Delta m_j \quad (8)$$

$$+\tfrac{1}{2}\textstyle\sum_{j=1}^{M}\sum_{k=1}^{M} \frac{\partial F^2(\mathbf{m}_0)}{\partial m_j \partial m_k}\Delta m_j \Delta m_k + \Omega(\mathbf{m}^3). \quad (9)$$

By truncating the Taylor-Lagrange expansion in the first order and deriving from the model parameter m_l, we have:

$$\frac{\partial F(\mathbf{m})}{\partial m_l} = \frac{\partial F(\mathbf{m}_0)}{\partial m_l} + \sum_{j=1}^{M} \frac{\partial F^2(\mathbf{m}_0)}{\partial m_j \partial m_l}\Delta m_j \quad (10)$$

The minimum of the objective function around $\mathbf{m}_0$ occurs when its first derivative is zero. In this way, it is possible to obtain the perturbation of the model $\Delta\mathbf{m}$.

$$\Delta\mathbf{m} = -\left[\frac{\partial F^2(\mathbf{m}_0)}{\partial \mathbf{m}^2}\right]^{-1}\left[\frac{\partial F(\mathbf{m}_0)}{\partial \mathbf{m}}\right] = -\mathbf{H}^{-1}\nabla F, \quad (11)$$

∇F is the gradient vector of the objective function at point $\mathbf{m}_0$, defined as:

$$\nabla F = \left[\frac{\partial F(\mathbf{m}_0)}{\partial \mathbf{m}}\right] = \left[\frac{\partial F(\mathbf{m}_0)}{\partial m_1}, \frac{\partial F(\mathbf{m}_0)}{\partial m_2}, ..., \frac{\partial F(\mathbf{m}_0)}{\partial m_M}\right]^T. \quad (12)$$

The second derivatives of the objective function correspond to the Hessian matrix and define the curvature of F in $\mathbf{m}_0$ and is defined as:

$$\mathbf{H} = \frac{\partial^2 F(\mathbf{m}_0)}{\partial m^2} = \frac{\partial^2 F(\mathbf{m}_0)}{\partial m_j m_l}. \quad (13)$$

The calculation of $\Delta\mathbf{m}$ in the equation (11) would lead to the minimum value in a single (Newton's method if $F(\mathbf{m})$ is quadratic and the term $\Omega(\mathbf{m3})$ in 9 is neglected. However, FWI is a strongly non-linear inversion, and it is necessary to solve the problem recurrently (Virieux and Operto, 2009).

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \mathbf{H}_k^{-1}\nabla F_k. \quad (14)$$

Therefore, the solution of the inverse problem is obtained in an iterative way, where in each iteration the parameters of the model obtained in the previous iteration are updated. The velocity model is updated by following the direction of the gradient vector in order to reach the minimum value of the objective function and obtain the inverted model $\mathbf{m}_{k+1}$ that best approximates the true model.

Preconditioning with source illumination - receiver applied to the gradient

A good alternative to perform the gradient computation is to make use of the adjoint method. This method can calculate the objective function gradient without requiring the explicit numerical calculation of the partial derivatives of $F(\mathbf{m})$ relative to the parameters of the model (Plessix, 2006). The gradient can be then calculated in time domain (Bunks et al., 1995) using the following formulation:

$$\nabla F = \frac{1}{c^3} \sum_{ns} \sum_{t=0}^{t_{max}} \ddot{u}_s u_r, \tag{15}$$

where ns is the total number of shots. $\ddot{u}_s$ corresponds to the second time derivative and can be obtained from:

$$\ddot{u}_s = \frac{\partial^2 u}{\partial t^2} = \frac{u^{n+1} - 2u^n + u^{n-1}}{\Delta t^2}, \tag{16}$$

where, u^{n+1}, u^n and u^{n-1} are the future, present and past wavefields, respectively.

By using the following relation

$$u^{n+1} + u^{n-1} = 2\cos(L\Delta t)u^n, \tag{17}$$

where $\cos(L\Delta t)$ is given by equation (4), equation (16) can be expressed as Thus,

$$\ddot{u}_s = \frac{2u^n \left[\sum_{k=0}^{\infty} C_{2k} J_{2k}(\Delta t R) Q_{2k}\left(\frac{iL}{R}\right) - 1\right]}{\Delta t^2} \tag{18}$$

Additionally, we need to compute u_r which is the field resulting from the reverse propagation of the residue $(\mathbf{d}_{calc} - \mathbf{d}_{obs})$ in the current velocity model. The residue wavefield, u_r, is called the adjoint state variable and it can be computed by the following equation:

$$\frac{1}{c(\mathbf{x})^2} \frac{\partial^2 u_r(x,t)}{\partial t^2} = \nabla^2 u_r(x,t) + (\mathbf{d}_{calc} - \mathbf{d}_{obs}). \tag{19}$$

The source term, which is propagated to generate these wavefields, depends on the formulation of the objective function that is minimized in the FWI procedure. That is, the adjoint source is the derivative of the objective function with respect to the modeled wavefield $\left(\frac{\partial F(\mathbf{m})}{\mathbf{u}_s}\right)$.

In order to improve the computation of the gradient and suppress the noises, Kaelin et al. (2007) proposed to divide equation (15) by the illumination of the source and the receivers and they can be expressed as:

$$\nabla F_s = \frac{1}{c^3} \frac{\sum_{ns} \sum_{t=0}^{t_{max}} \ddot{u}_s u_r}{\sum_{ns} \sum_t \ddot{u}_s^2}, \tag{20}$$

and

$$\nabla F_r = \frac{1}{c^3} \frac{\sum_{ns} \sum_{t=0}^{t_{max}} \ddot{u}_s u_r}{\sum_{ns} \sum_t u_r^2}, \tag{21}$$

Equation (20) is intended to highlight the shallower reflectors and (21) tries to highlight the deepest reflectors.

Thus, by combining the two previous imaging conditions (20) and (21), we can obtain equal illumination for all reflectors:

$$\nabla F_{sr} = \nabla F_s + \nabla F_r. \tag{22}$$

In the FWI procedure, equation (22) proved to be very efficient, improving the amplitude values of the conventional gradient (equation 15) and consequently increasing the resolution of the inverted velocity models.

Results and Conclusions

In the FWI procedure, the observed data was generated using the REM with the true velocity field of the model Marmousi (Figure 1). The input velocity field (Figure 2) that initiated the inversion is a smoothed version of the original model. The observed data was filtered for each frequency band and the calculated data was modeled with these corresponding frequencies using the updated fields and then compared with the observed data.

The inversion procedure starts by updating the observed low frequency seismic data in the ranges of $0-2.5Hz$ and $0-5.0Hz$, which will result in the recover of the large structures of the model as we can notice in Figure 3 and Figure 4, respectively. Afterwards, taken a fixed number of iterations, we increased the range of frequency, in range $0-7.5Hz$ and $0-10.0Hz$ and continue the inversion small structure start to show up. The results in Figure 5 and Figure 6, for these intermediate range of frequency, show that the resolution has increased and the inverted model is comparable with the true one present in Figure 1. Finally the process ends with the inversion of the seismic high frequency data, which inserts the high resolution details on the inverted final model (Figures 7 and 8).

Considering the previously mentioned results, it was possible verify that FWI with the rapid expansion method and the proposed preconditioning method for the gradient vector led to an overall improvement of the procedure and has produced velocity models with high resolution. The application of the multiscale approach is crucial to avoid local minimums and also to obtain models with more resolution.

The preconditioning method used here for computing the gradient vector consists of a light compensation which can be considered an approximation of the inverse of the Hessian matrix. The graphs of the objective functions (Figure 9) and velocity profile (Figure 10) show the efficiency of the method (Figures: 3-8).

Acknowledgements

This research was supported by Fapesb and INCT-GP/CNPq. The facility support from CPGG/UFBA is also acknowledged.

References

Araújo, E. S. d., 2009, Análise dos métodos de diferenças finitas e expansão rápida na migração reversa no tempo.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

dos Santos, A. W. and R. C. Pestana, 2015, Time-domain multiscale full-waveform inversion using the rapid expansion method and efficient step-length estimation: Geophysics, **80**, R203–R216.

Kaelin, B., A. Guitton, and F. Ortigosa, 2007, Illumination effects in rreverese time migration: Presented at the 69th EAGE Conference and Exhibition incorporating SPE EUROPEC 2007.

Kosloff, D., A. Filho, E. Tessmer, and A. Behle, 1989, Numerical solution of the acoustic and elastic wave equation by new rapid expansion method: Geoph. Prosp., **37**, 383–394.

Pestana, R. C., P. L. Stoffa, et al., 2009, Rapid expansion method (rem) for time-stepping in reverse time migration (rtm): Presented at the 2009 SEG Annual Meeting.

Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

Tal-Ezer, H., D. D. Kosloff, and Z. Koren, 1987, An accurate scheme for seismic forward modeling: Geophysical Prospecting, **35**, 479–490.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Figures

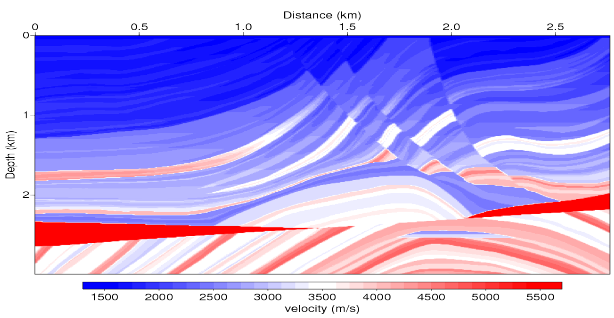


Figure 1: True Marmousi model

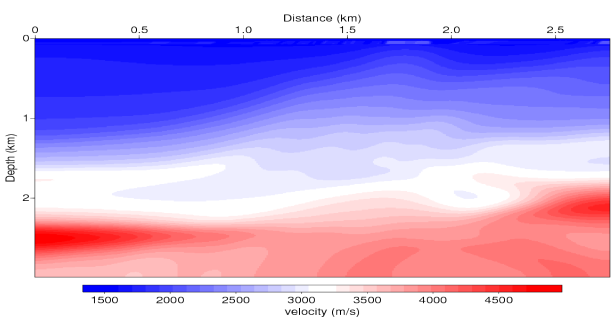


Figure 2: Initial model of the Marmousi model for inversion procedure

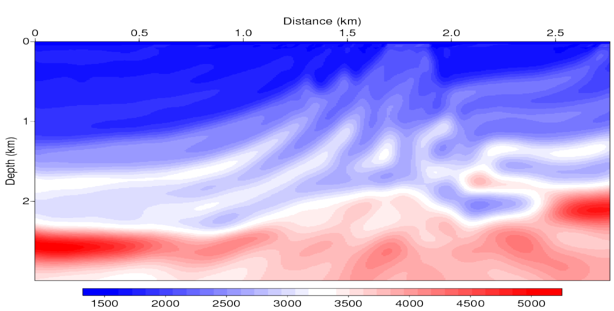


Figure 3: Estimated model using gradient preconditioning method for peak frequency of 2.5 Hz.

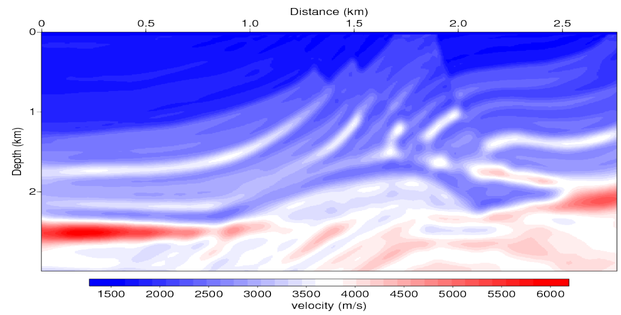


Figure 4: Estimated model using gradient preconditioning method for peak frequency of 5 Hz.

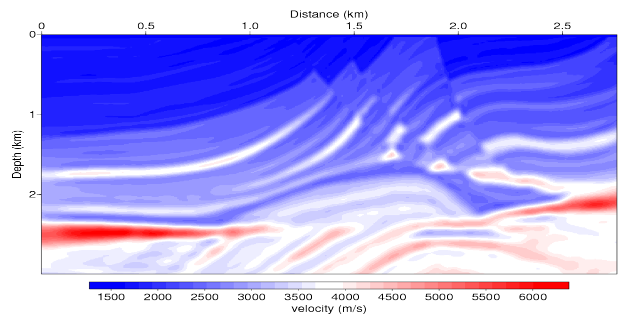


Figure 5: Estimated model using gradient preconditioning method for peak frequency of 7.5 Hz.

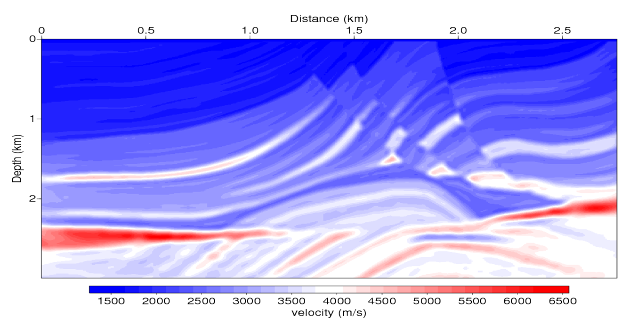


Figure 6: Estimated model using gradient preconditioning method for peak frequency of 10 Hz.

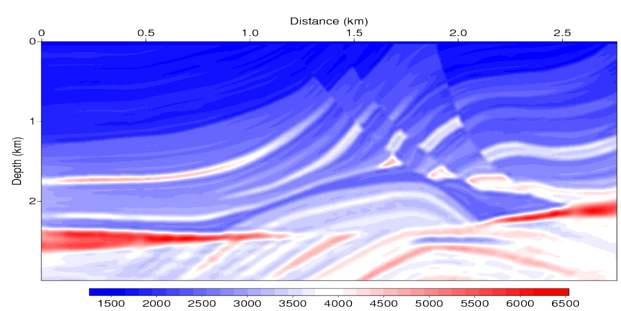


Figure 7: Estimated model using gradient preconditioning method for peak frequency of 12.5 Hz.

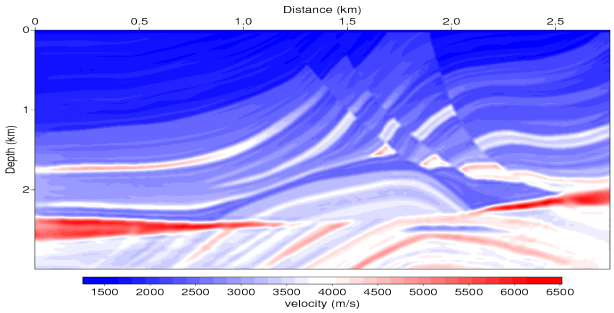


Figure 8: Estimated model using gradient preconditioning method for peak frequency of 15 Hz.

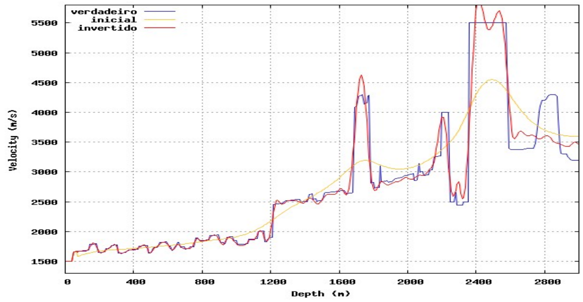
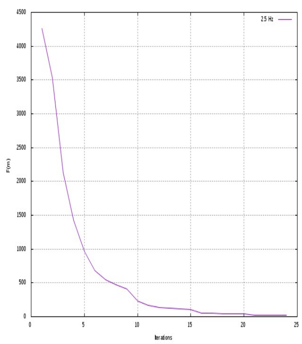
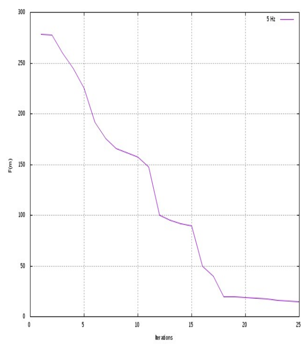


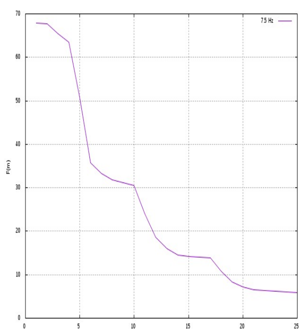
Figure 10: Profile of the estimated model for the preconditioning method for the gradient. The profile corresponds to position $x = 1250m$



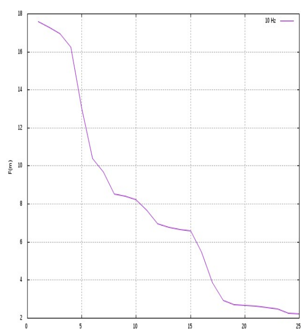
(a)



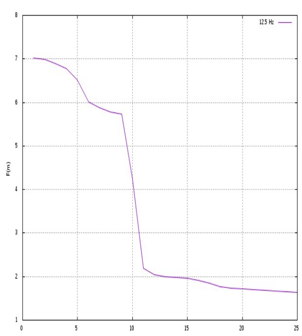
(b)



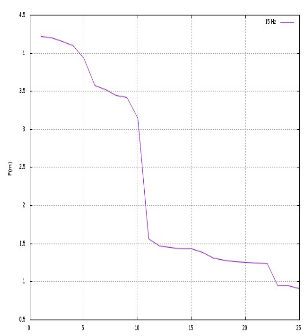
(c)



(d)



(e)



(f)

Figure 9: Convergence of the objective function using the gradient preconditioning method for peak frequency of: (a) 2.5 Hz, (b) 5 Hz, (c)7.5 Hz, (d) 10 Hz, (e) 12.5 Hz, (f) 15 Hz.