



The role of parametrization in acoustic full waveform inversion

*Marília da Silva Ramos Carneiro¹, André Bulcão², Bruno Pereira-Dias², Djalma M. Soares Filho²,

Felipe de Souza Duarte¹, Luiz Landau¹,

¹LAB2M/COPPE/UFRJ, ²PETROBRAS, Brazil

Copyright 2014, SBGf - Sociedade Brasileira de Geofísica.

Este texto foi preparado para a apresentação no VI Simpósio Brasileiro de Geofísica, Porto Alegre, 14 a 16 de outubro de 2014. Seu conteúdo foi revisado pelo Comitê Técnico do VI SimBGf, mas não necessariamente representa a opinião da SBGf ou de seus associados. É proibida a reprodução total ou parcial deste material para propósitos comerciais sem prévia autorização da SBGf.

Abstract

Being an inverse problem, Full Waveform Inversion (FWI) is subject to scaling problems due to inappropriate choice of model parametrization. This work puts in evidence that the choice of compressional velocity or squared slowness parameters leads to very different gradient expressions (for acoustic FWI), and thus must have distinct convergence properties (for a gradient-based algorithm). Also, an interesting result is related to the role of the pseudo-Hessian, which shows that either compressional velocity or squared slowness yields the same update for small step-lengths. This is an interesting feature expected from a Hessian-based preconditioner, which makes inversion result more robust and less dependent of parametrization choice. These features are illustrated with a numerical example using the Marmousi-2 model.

Introduction

The full waveform inversion (FWI) is becoming an important tool to improve quantitative model parameters of the subsurface in high resolution. It aims to use of all information contained in the seismic data, such as traveltimes and amplitude (Tarantola, 1984). This ensures that the seismic wave propagations, through heterogeneous media, is modeled contemplating all wave-modes. The numerical solution provides seismic waveform that could be explored for the benefit for better tomographic solutions, without the requirement to specify the seismic phases. The results developed through this method can be used to the improvement of migration velocity for the seismic imaging (depth seismic migration) and possibly to the direct interpretation of lithology. This technique is a data-fitting procedure, in which updates iteratively the model parameters of the subsurface, in order to minimize the misfit between the registered data (\mathbf{d}) and the simulated data (\mathbf{u}).

The FWI method relies on two main steps: a forward problem (the algorithm is based on full waveform two-way modeling) and a solution of an inverse problem by the computation of an update direction, a step length and the evaluation of the objective function. FWI is a highly non-linear inverse problem subject to non-uniqueness issue and convergence to a spurious model (i.e., a local minimum

in the objective function). On top of that, the high computational cost puts great challenges for the FWI in industrial scale, which requires 3D simulation of large seismic surveys.

Therefore, due to the computational power available nowadays, FWI application on large-scale datasets has been mainly restricted to gradient-based methods. On the other hand, these methods, which does not include a Hessian correction, are known to face scaling problems (Nocedal, 2006). Thus, it is preferable to choose an algorithm that is not sensitive to scaling, such as the Newton method, because they can handle poor problem formulations in a more robust fashion (Nocedal, 2006). The problem of choice of parameters in FWI goes back to Tarantola (1986), who has already drawn attention to the necessity of adequately chosen parameters. The reason is that even theoretically equivalent parametrization can lead to distinct convergence rates in the inversion.

It is shown in this work, for the acoustic FWI, a comparison of the inversion results using different parameterizations to update the compressional velocity model. It is also discussed that the preconditioning by the pseudo-Hessian diagonal, leads to a less sensitive algorithm to the parametrization, and a better scaling of the gradient. In the result section, these features are illustrated with a numerical example using the Marmousi-2 model.

Theory

The core of the FWI algorithm relies on the definition of the forward modeling operator for the wavefield \mathbf{L} . It can be represented in a general form by

$$\mathbf{L}(\mathbf{u}; \mathbf{p}) = \mathbf{f}, \quad (1)$$

where \mathbf{f} are the external sources, \mathbf{u} is the wavefield, that relies on the model parameters \mathbf{p} .

For the sake of simplicity, we will focus mainly on the acoustic constant density case. For this simpler problem, the forward problem operator is given by:

$$L(u; v_p) = \frac{1}{v_p^2} \frac{\partial^2 u}{\partial t^2} - \nabla^2 u. \quad (2)$$

in which u represents the pressure wavefield and v_p is the compressional velocity model.

Once the physics of the problem is defined (i.e., the modeling operator), the FWI problem may be formulated as a local optimization problem that seeks a model \mathbf{p} which minimizes an objective function value $E(\mathbf{p})$. A very popular choice for the objective function is the \mathcal{L}_2 -norm of the residual, that is, the difference between the observed

dataset \mathbf{d}_s with the modeled wavefield \mathbf{u}_s , at the shot point \mathbf{x}^s and receiver points \mathbf{x}^r , integrated in the time-domain:

$$E(\mathbf{p}) = \frac{1}{2} \sum_{s \in \text{src}} \sum_{r \in \text{rcv}} \int_T dt \int_G d^3\mathbf{x} |\mathbf{u}_s(\mathbf{p}; \mathbf{x}, t) - \mathbf{d}_s|^2 \delta(\mathbf{x} - \mathbf{x}^r). \quad (3)$$

The model parameter \mathbf{p} can be updated through a gradient-based method so to find a model that reduces the value of the objective function. The gradient of the objective function is defined as

$$\nabla_p E(\mathbf{p}) \delta \mathbf{p} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} [E(\mathbf{p} + \varepsilon \delta \mathbf{p}) - E(\mathbf{p})]. \quad (4)$$

However, the explicit evaluation of the right side Eq. (4) by the finite-difference method has a prohibitive computational cost, because it is necessary the evaluation of each derivative model component.

An effective approach, fully described in Fichtner (2011), Chapter 8, is to make use of the adjoint method. This method allows to evaluate the gradient of the objective function in a very computationally efficient way.

It can be shown, without the use of any approximation, that the gradient of the objective function may be obtained by

$$\nabla_p E(\mathbf{p}) \delta \mathbf{p} = \int_T dt \int_G d^3\mathbf{x} \mathbf{u}^\dagger \cdot \nabla_p \mathbf{L}(\mathbf{u}; \mathbf{p}) \delta \mathbf{p}. \quad (5)$$

This means that the computation of the gradient can be done with a modeling of \mathbf{u} and \mathbf{u}^\dagger for each source.

The adjoint wavefield \mathbf{u}^\dagger is modeled through the adjoint equation of Eq. (2):

$$\nabla_u \mathbf{L}^\dagger \mathbf{u}^\dagger = -\nabla_u E_1^\dagger. \quad (6)$$

where E_1 is the integrand of the objective function. For the special case of \mathcal{L}_2 -norm of the residual (Eq. (3)), the right-hand side of the Eq. (6) has as the adjoint source residual data at the receivers position.

The local optimization methods consists in estimating the objective function derivatives in the neighborhood of an initial model \mathbf{p}_0 . In each iteration, it is computed a search direction \mathbf{h}_k and a step-length α_k , in a way that the model parameters update is represented by the following expression:

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \alpha_k \mathbf{h}_k \quad (7)$$

where \mathbf{p}_k is the model parameter for the k^{th} iteration. The step length is determined by a linear search (Nocedal, 2006). The update must guarantee the reduction of the objective function value.

In this work, we used the steepest-descent method to determine the search direction \mathbf{h}_k . However, one should be aware that the gradient scaling is very dependent on the choice of model parametrization. Therewith, we present two distinct parameterizations for the computation of the update of compressional velocity model (v_p) in the acoustic case.

To evaluate the gradient computation, we considered two distinct parameterizations, described as follow:

- P-wave squared slowness (σ);
- P-wave velocity (v_p);

P-wave squared slowness (σ)

The P-wave squared slowness is defined by $\sigma = \frac{1}{v_p^2}$, so we have the gradient formulation

$$\nabla_\sigma E \delta \sigma = \int_T dt \int_G d^3\mathbf{x} u^\dagger \ddot{u} \delta \sigma, \quad (8)$$

Therefore, in the k^{th} updated iteration model is given by

$$\sigma_{k+1} = \sigma_k - \alpha_k \nabla_\sigma E_k. \quad (9)$$

In terms of v_p , this update is given by

$$\frac{1}{v_{p_{k+1}}^2} = \frac{1}{v_{p_k}^2} - \alpha_k \nabla_\sigma E_k, \quad (10)$$

that can be expressed as follows:

$$v_{p_{k+1}} = \frac{v_{p_k}}{\sqrt{1 - v_{p_k}^2 \alpha \nabla_\sigma E_k}}. \quad (11)$$

For small updates, $\alpha \nabla_\sigma E_k \sim 0.05 * (1/v_{p_k}^2)$, it is possible to approximate the expression (11) in Taylor's expansion, such as

$$v_{p_{k+1}} = v_{p_k} + \frac{v_{p_k}^3}{2} \alpha \nabla_\sigma E_k. \quad (12)$$

P-wave velocity (v_p)

Using v_p velocity as the parameter, we have

$$\nabla_{v_p} E \delta v_p = \int_T dt \int_G d^3\mathbf{x} \left(-\frac{2}{v_p^3} \right) u^\dagger \ddot{u} \delta v_p. \quad (13)$$

The gradient can be represented as follows

$$\nabla_{v_p} E = -\frac{2}{v_p^3} \nabla_\sigma E. \quad (14)$$

Hence, the k^{th} update for the velocity parameter is

$$v_{p_{k+1}} = v_{p_k} + \alpha \frac{2}{v_{p_k}^3} \nabla_\sigma E_k. \quad (15)$$

In this way, it is apparently obvious that the gradients in terms of squared slowness (11) and with velocity (15) are distinct. This certainly implies in different convergence properties for each parametrization.

Pseudo-Hessian diagonal pre-conditioner

A secure way to mathematically verify which parametrization (v_p or σ) leads to a better scaled gradient is using the pseudo-Hessian as a pre-conditioner. This approach is expected to deliver a better update the model parameter.

The pseudo-Hessian diagonal $\tilde{\mathbf{H}}_a$ is a rough approximation for the complete Hessian matrix. It is given as follow

$$\tilde{\mathbf{H}}_{a;p} = \sum_{s \in \text{src}} \int_T dt \left| \mathbf{f}_{s,p}^{\text{virt}} \right|^2 \quad (16)$$

and it was suggested in Shin et al. (2001) to improve the computation of prestack migration. Here s is the source subscript, N_s is total number of shots and $\mathbf{f}_{s,p}^{\text{virt}}$ is the virtual source, defined as

$$\mathbf{f}_{s,p}^{\text{virt}} = - \frac{\partial L}{\partial p_i} (u_s). \quad (17)$$

This structure modify the update for different model parametrization. For the parametrization for the inversion problem with v_p and σ , the virtual sources are given by

$$\mathbf{f}_{\sigma}^{\text{virt}} = - \frac{\partial L}{\partial \sigma} (u) = -\ddot{u}, \quad (18)$$

$$\mathbf{f}_{v_p}^{\text{virt}} = - \frac{\partial L}{\partial v_p} (u) = - \frac{2}{v_p^3} \ddot{u}, \quad (19)$$

respectively.

Hence, the update direction for σ is given by

$$\tilde{\mathbf{H}}_{a;\sigma}^{-1} \nabla_{\sigma} E = \frac{\int_T dt u^{\dagger} \ddot{u}}{\int_T dt \ddot{u} \ddot{u}}, \quad (20)$$

and for v_p ,

$$\tilde{\mathbf{H}}_{a;v_p}^{-1} \nabla_{v_p} E = \frac{-\frac{2}{v_p^3} \int_T dt u^{\dagger} \ddot{u}}{\left(-\frac{2}{v_p^3}\right)^2 \int_T dt \ddot{u} \ddot{u}} = -\frac{v_p^3}{2} \frac{\int_T dt u^{\dagger} \ddot{u}}{\int_T dt \ddot{u} \ddot{u}}. \quad (21)$$

In this way, the expressions (21) and (20) gives the following update in terms of compressional velocity (for small updates):

$$v_{pk+1} = v_{pk} \left(1 - v_{pk}^2 \alpha \tilde{\mathbf{H}}_{a;\sigma}^{-1} \nabla_{\sigma} E_k \right)^{-\frac{1}{2}} \quad (22)$$

$$\approx v_{pk} + \alpha \frac{v_p^3}{2} \frac{\int_T dt u^{\dagger} \ddot{u}}{\int_T dt \ddot{u} \ddot{u}} \quad (23)$$

$$= v_{pk} - \alpha \tilde{\mathbf{H}}_{a;v_p}^{-1} \nabla_{v_p} E_k. \quad (24)$$

The importance of this result is that independently of the parametrization (v_p or σ) the gradient scaled by the pseudo-Hessian diagonal gives the same model update (for small updates).

Results for the Marmousi-2 Model

The target model used for the proposed methodology is the Marmousi-2 (Martin et al., 2002), shown in Fig. 1a, representing a complex acoustic isotropic media for P-wave velocity. We have tested the inversion scheme for a 2D noise-free synthetic dataset. It was modeled in the time domain, considering a fixed-spread acquisition geometry for the receivers, with 1360 receivers along the surface of the model. In Table 1 it is presented the parameters used for modeling and inversion.

Table 1: Modeling and inversion parameters

Model dimension	1361x281 grid points
Grid point interval (h)	10 m
Time step	1 ms
Total time step	3000 ms
Max. frequency	14.0 Hz
Initial frequency	8.0 Hz
Frequency interval	2.0 Hz
Number of shot gathers	68
Receiver interval	10 m

The inversion process was performed in the time domain. It was used a multi-scale approach by inverting from the lowest to a higher frequency. This is important to avoid the convergence toward a local minimum, due to cycle skipping (Bunks et al., 1995). It was used 4 bandwidth of frequencies, starting from maximum frequency 8 Hz up to maximum 14 Hz, with intervals of 2 Hz, and the velocity model generated for a lower frequency inversion became the input velocity model for the next frequency. For each frequency, it was performed 10 iterations. The initial model used for input in the inversion is represented in Fig. 1b. It corresponds to a smoothed version of the true model by a filter of 300m x 300m radius.

Three difference inversion schemes to update the model parameters were realized:

- Based on the v_p parametrization with the update shown in Eq. (15);
- Based on σ parametrization with the update shown in Eq. (12);
- Based on pseudo-Hessian diagonal correction with the update shown in Eq. (24);

Figure 1c-e contains the final results of the FWI for different parameterizations and in Fig. 1g is a comparison of the velocity profiles provided by each inversion result at the position $x=8.5$ km over the true and initial models. By these results:

1. It is possible to check in Figs. 1c and 1d huge difference of the updates below 1 km in depth comparing the v_p and σ parametrization. The σ parametrization favored better update throughout the whole of the subsurface, whereas the deeper part of the model for the v_p parametrization remained almost the same as the initial model.

2. With parametrization on σ and with pseudo-Hessian diagonal correction $\tilde{\mathbf{H}}_a$ the velocity model Marmousi-2 was greatly recovered, being both quite similar.
3. These conclusions are seen with greater detail in the velocity profile (Fig. 1g). All the inverted results have a good match with the true model for depth shallower than 1km. However, for the deeper part, the parametrization with v_p has the weaker match with the true model and a profile closer to the initial model. On the other hand, using parametrization on σ and $\tilde{\mathbf{H}}_a$ are very similar with each other and better matched with the true model, despite of the limited iteration number and limited bandwidth used.

These numerical results put in evidence the importance of the choice of parametrization for an adequate scaling of the gradient. It is clear that the parametrization with σ yields better inversion results and a faster convergence than the v_p parametrization. This, can be understood by seeing the only difference of expressions (12) and (23) is the autocorrelation term of the forward wavefield ($\int_T dt \ddot{u} \ddot{u}$). In other words, the σ parametrization has the same scaling property of the parameter as the pseudo-Hessian update, although it does not have the wavefield illumination factor correction.

In fact, a possible explanation for the very close results of the parametrization with σ and the correction by the pseudo-Hessian is by analyzing the plot of the autocorrelation of the forward wavefield for the last iteration (Fig. 1f). Since, the chosen geometry of acquisition leads to a complete illumination of the whole model, the term ($\int_T dt \ddot{u} \ddot{u}$) has influence only in the very boundary part of the model. This small difference can be seen in the very deep part of the velocity profile (around 2.7 km). For incomplete acquisition geometry (not considered in this work), the illumination factor should have a greater importance. This is fully discussed in the original article of Shin et al. (2001).

Conclusions and Discussion

In this work it was evaluated the gradient computation for acoustic FWI to two distinct parameterizations: the compressional velocity v_p and the squared slowness σ . The resulting different expressions implied in distinct convergence properties for them. It could be concluded, that the choice of parametrization is truly important to better update the velocity models from the inversion.

The mathematical explanation for two theoretically different parametrization give different inversion result is due to bad scaling property of gradient-based methods. A more robust way to approach the problem was done by updating the model with a pseudo-Hessian correction. With this preconditioning, it could be shown that the update for compressional velocity remains the same (for small updates) using either compressional velocity or squared slowness as the inversion parameters.

This study puts in perspective the huge role that parametrization considerations should have in multi-parameter inversion. In fact, with more parameters at stake greater problems are expected to arise due to the possibility of trade-off and cross-talk between the gradients of different parameters (Prioux 2013 a,b). Therefore, studies as this one should be pursued to enable gradient-

based methods (which are computationally cheap) to be used, but with appropriate scaling for different classes of parameters. The role and structure of the Hessian should be relevant in this future work.

Acknowledgments

The authors would like to thank PETROBRAS for authorizing this publication. M.S.R.C acknowledges PRH-ANP for the financial support.

References

- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, 60, 1457-1473.
- Fichtner, A., 2011, Full Seismic Waveform Modelling and Inversion. *Advances in Geophysical and Environmental Mechanics and Mathematics*. Springer.
- Martin, G. S., Marfurt, K. J., Larsen, S., 2002, Marmousi-2: An Updated Model for Investigation of AVO in Structurally Complex Area: *Society of Exploration Geophysics*.
- Nocedal, J., Wright, S. J., 2006, *Numerical Optimization*. Springer.
- Prioux, V., Brossier, R., Operto, S., et al., 2013a, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall Field. Part 1: imaging compressional wave speed, density and attenuation, *Geophysical Journal International*, v. 194, n. 3, pp. 1640-1664.
- Prioux, V., Brossier, R., Operto, S., et al., 2013b, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall Field. Part 2: imaging compressive-wave and shear-wave velocities, *Geophysical Journal International*, v. 194, n. 3, pp. 1665-1681.
- Shin, C., Jang, S., Min, D.-J., 2001, "Improved amplitude preservation for prestack depth migration by inverse scattering theory", *Geophysical Prospecting*, v. 49, pp. 592-606.
- Tarantola, A., 1984, "Inversion of seismic reflection data in the acoustic approximation", *Geophysics*, v. 49, pp. 1259-1903.
- Tarantola, A., 1986, "A strategy for nonlinear elastic inversion of seismic reflection data", *Geophysics*, v. 51, pp. 1893-1903.

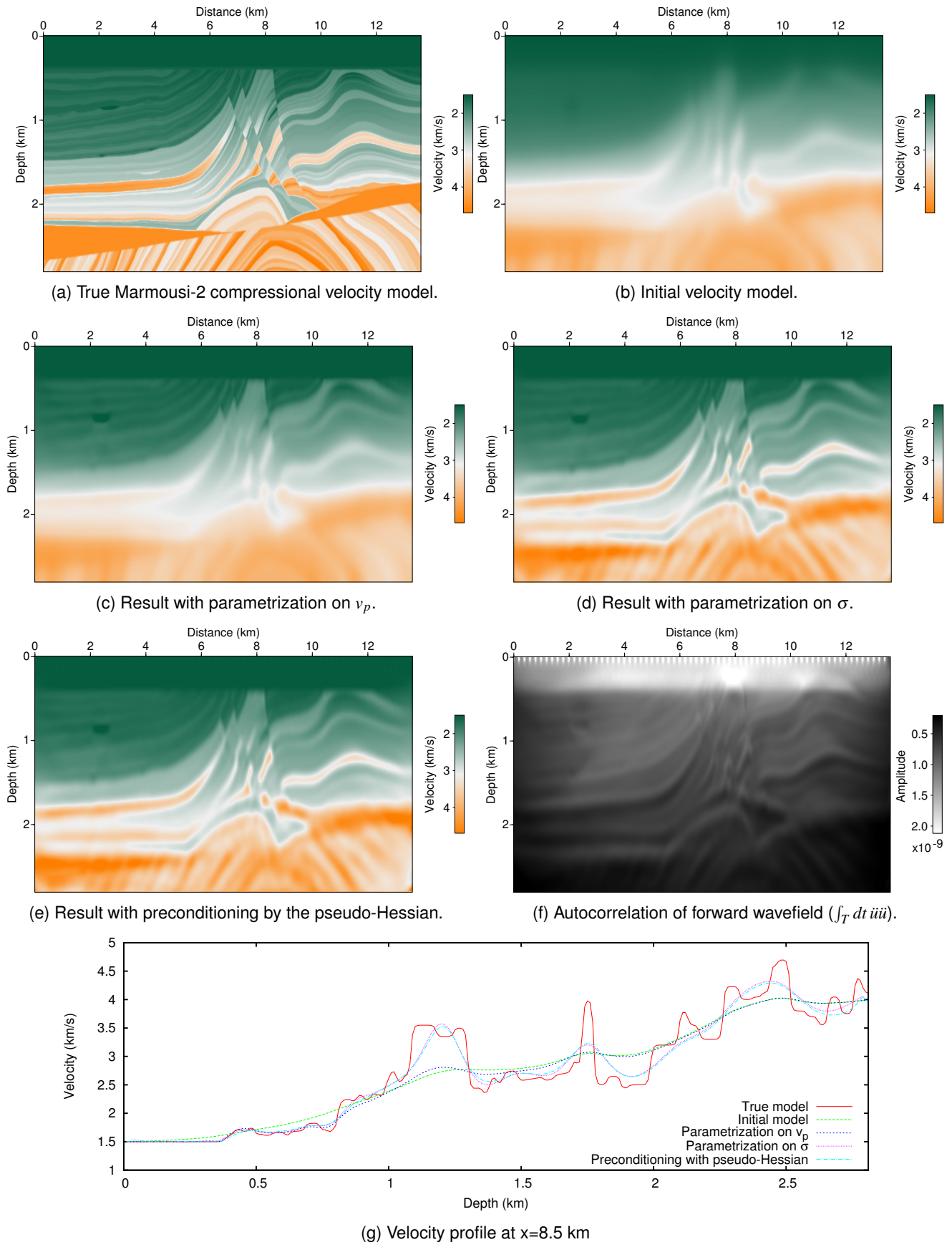


Figure 1: The true Marmousi-2 velocity model is given on (a). The initial model for the inversion is depicted on (b). In figure (c), is the result for the inversion using the parametrization on v_p . Figure (d), is the result for the inversion using the parametrization on σ . Figure (e), the result using preconditioning with pseudo-Hessian. Figure (f), is the autocorrelation of forward wavefield for the last iteration, shown in Eq. (20). And Fig. (g) a velocity profile for the estimated models at the position $x=8.5$ km.