

A parameterization analysis for acoustic full-waveform inversion of sub-wavelength anomalies

Pawan Bharadwaj¹, Wim Mulder^{1,2} and Guy Drijkoningen¹

¹ Department of Geoscience & Engineering, Delft University of Technology

² Shell Global Solutions International B.V.

April 5, 2018

Abstract

With single-parameter full-waveform inversion (FWI), estimating the inverse of the Hessian matrix will accelerate the convergence but will increase the computational cost and memory requirements at each iteration. Therefore, it is often replaced by the inverse of an approximate Hessian that is easier to compute and serves as a preconditioning matrix. Alternatively, or in combination, a BFGS-type of optimization method that estimates the inverse Hessian from subsequent iterations can be applied. Similarly, in the case of multi-parameter full-waveform inversion, the computation of the additional Hessian terms that contain derivatives with respect to more than one type of parameter is necessary. If a simple gradient-based minimization with, for instance, just scalar weights for each of the parameter types is used, different choices of parameterization can be interpreted as different preconditioners that change the condition number of the Hessian. If the non-linear inverse problem is well-posed, then the inversion should converge to a band-limited version of the true solution irrespective of the parameterization choice, provided we start sufficiently close to the global minimum. However, the choice of parameterization will affect the rate of convergence to the exact solution and the ‘best’ choice of parameterization is the one with the fastest rate.

In this paper, we search for the best choice for acoustic full-waveform inversion, where 1. anomalies with a size less than a quarter of the dominant wavelength have to be estimated without the risk of converging to a local minimum; 2. the scattered wavefield is recorded at all the scattering angles; 3. a steepest-descent minimization scheme is used. Towards that end, we review two different conventional analysis methods, i.e., the point-scatterer analysis and diffraction-pattern analysis. The conventional methods consider only a few terms of the full Hessian matrix for parameterization analysis. To validate them, we consider single-component numerical examples, where the inversion estimates one of the following: 1. only the contrast of a point-shaped scatterer; 2. only the contrast of a Gaussian-shaped scatterer; 3. both the shape and contrast of a Gaussian-shaped scatterer. The numerical examples show that the suggestions of the conventional analyses are only valid while estimating the contrast of point-shaped anomalies at a known location. For extended anomalies, the examples suggest that the best choice of parameterization depends on the contrast of the subsurface scatterer that the inversion tries to estimate. Based on the results, we observe that there is no best parameterization choice for full-waveform inversion in the case that both the shape and the size of the anomalies have to be estimated. We also observe that a parameterization using the acoustic impedance and mass density has the worst convergence rate.

In addition to full-waveform inversion, we consider Born modelling and inversion to learn if the dependence on the contrast of subsurface scatterer is due to the non-linearity in the

full-waveform modelling and inversion. We again observe that there is no best parameterization choice for Born modelling and inversion and the parameterization using the acoustic impedance and mass density has the worst convergence in the general case.

Finally, we also show that the parameterization analysis during a hierarchical inversion, where the data have limited scattering angles, only helps to select a subspace for mono-parameter inversion. For multi-parameter hierarchical inversion, the search for the *best* parameterization in terms of the convergence speed might be obfuscated by non-uniqueness problems.

1 Introduction

Quantitative imaging of various near-surface elastic parameters is essential in many civil engineering applications as well as for hydrocarbon exploration. One approach is full-waveform inversion (FWI) of the recorded elastic wavefield [15, 16], which is sensitive to the shear and compressional properties of the subsurface. Multi-parameter FWI is a non-linear procedure that minimizes the least-squares misfit between the recorded and the modelled seismic data to estimate various subsurface parameters. Given the size of the seismic problem, i.e., estimating thousands or millions of parameters, it is only feasible in practice to use descent methods for optimization.

In order to reduce the number of iterations needed to reach an acceptable solution, the gradient at each iteration can be preconditioned using the inverse of the Hessian matrix [12]. The elements of the Hessian matrix are the second-order derivatives of the data misfit function with respect to the model parameters. If model parameter of kinds a and b at locations \mathbf{x}_i and \mathbf{x}_j are given by $m_a(\mathbf{x}_i)$ and $m_b(\mathbf{x}_j)$, respectively, the elements of the Hessian matrix \mathbf{H} are

$$\mathbf{H}_{(a,i),(b,j)} = \partial_{m_a(\mathbf{x}_i)} \partial_{m_b(\mathbf{x}_j)} J_{ls}, \quad (1)$$

where J_{ls} denotes the least-squares misfit functional.

We will refer to the Hessian terms with $a = b$ as *mono-parameter* terms. The terms of the Hessian on its band diagonal with $i \approx j$, $i \neq j$ and $a = b$ will be called *band-diagonal* mono-parameter terms, as sketched in Figure 1. During the preconditioning, they deconvolve the gradient such that it is less dependent on source bandwidth and acquisition geometry. The terms of the Hessian where $i = j$ and $a = b$, called *main-diagonal* mono-parameter terms, account for the amplitude effects in wave propagation [12, 16], for instance, due to geometrical spreading. In the presence of multiple scattering, the change in data due to a model perturbation at \mathbf{x}_i depends on the model perturbation at \mathbf{x}_j even if $i \neq j$ and $i \not\approx j$. This means that these non-band-diagonal terms of the Hessian, $\mathbf{H}_{(a,i),(b,j)}$ with $i \neq j$ and $i \not\approx j$, are non-zero. During the preconditioning, they correct the gradient such that multiple scattering is taken into account.

The non-zero Hessian terms with $a \neq b$ represent cross-talk between the different parameter types. We call them *cross-parameter* terms, sketched in Figure 1. The cross-parameter terms corresponding to a single subsurface position when $i = j$ are called *main* cross-parameter terms, whereas the cross-parameter terms with $i \approx j$ and $i \neq j$ are called *block* cross-parameter terms. When cross-parameter terms are smallest, a change in the data due to a perturbation of one parameter should be almost independent of the perturbation of another parameter at the same location.

The computation of the Hessian matrix or its inverse will increase the computational cost and memory requirements at each iteration for large-scale 3-D acoustic inverse problems [12, 3, 8]. Therefore, in this paper, we restrict ourselves to gradient-based optimisation without involving the Hessian. Since the terms of the Hessian matrix depend on the choice of the subsurface parameterization in the case of multi-parameter full-waveform inversion, the various

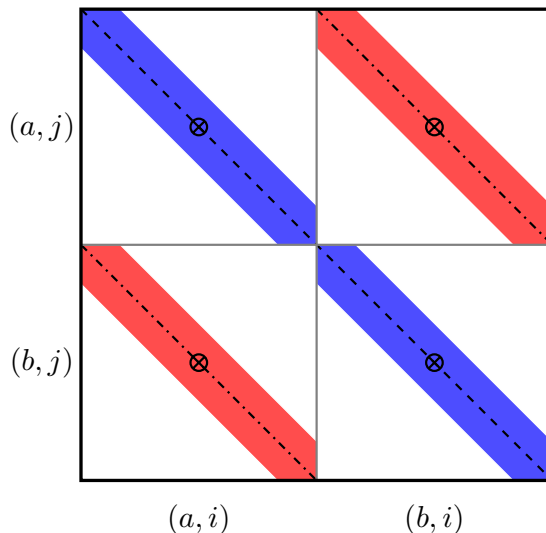


Figure 1: This sketch marks different terms of the Hessian matrix in equation 1. The main-diagonal mono-parameter terms are marked with a dashed line. The band-diagonal mono-parameter terms are marked in blue and the block cross-parameter terms in red. Dashed-dotted line mark the main cross-parameter terms. The terms of the reduced Hessian, in equation 17, are marked with \otimes .

parameterization choices are not equivalent. If the inverse problem is well-posed, the inversion should converge to a band-limited version of the true solution irrespective of the parameterization choice, provided we start close to the global minimum. The *best* choice of parameterization is the one that converges the fastest. It should be noted that the rate of convergence of a particular parameterization depends on the background subsurface model, without scatterers, and the acquisition geometry. The background subsurface model serves as the initial or starting model for inversion.

Hence, for a given acquisition geometry and background subsurface model, it is obvious that the fastest convergence will be obtained for a parameterization choice that provides zero cross-parameter terms in the Hessian. We notice, for an acoustic inverse problem, that both the main and block cross-parameter terms are non-zero for any of the parameterization choices involving wave velocity, density, bulk modulus or wave-impedance. Among these options, we, therefore, have to choose a parameterization that provides the smallest condition number of the Hessian matrix. Because the computation of the full Hessian matrix and its condition number for large-scale inverse problems is costly and requires large memory, we prefer a simpler method for parameterization analysis.

There are two conventional methods to analyse different subsurface parameterization choices with the aim to find the *best* one. The first method, called diffraction-pattern analysis, chooses the best parameterization by examining the radiation patterns of the scattered wavefield generated by different perturbations [15, 10, 13]. The second method, called point-scatterer analysis, considers the Hessian of a reduced inverse problem that only estimates the contrast of a point scatterer at a known location. The reduction will result in fewer unknowns compared to the original inverse problem, where both the contrast and shape of the scatterer have to be estimated. The block cross-parameter terms of the original non-reduced Hessian, the red terms in Figure 1,

Table 1: Various medium parameterizations that are used in this paper for the acoustic inverse problem.

	Parameterization	First Parameter	Second Parameter
(A)	$\mathbf{m}_{c_p, \rho}$	$\xi_{c_p} = (c_p - c_{po})/c_{po}$	$\xi_\rho = (\rho - \rho_o)/\rho_o$
(B)	$\mathbf{m}_{Z_p, \rho}$	$\xi_{Z_p} = (Z_p - Z_{po})/Z_{po}$	$\xi_\rho = (\rho - \rho_o)/\rho_o$
(C)	\mathbf{m}_{c_p, Z_p}	$\xi_{c_p} = (c_p - c_{po})/c_{po}$	$\xi_{Z_p} = (Z_p - Z_{po})/Z_{po}$
(D)	$\mathbf{m}_{K, \rho}$	$\xi_K = (K - K_o)/K_o$	$\xi_\rho = (\rho - \rho_o)/\rho_o$
(E)	$\mathbf{m}^{1/K, 1/\rho}$	$\xi_{1/K} = (1/K - 1/K_o)/(1/K_o)$	$\xi_{1/\rho} = (1/\rho - 1/\rho_o)/(1/\rho_o)$

are absent in the reduced Hessian. Therefore, this analysis ignores the block cross-parameter terms of the original non-reduced Hessian matrix and chooses a parameterization which has main cross-parameter terms of the least possible magnitude. For elastic waveform inversion, [9] numerically demonstrated that the choice of material parameters is more complicated than the current literature suggests.

In this paper, we consider an almost well-posed 2-D acoustic inversion example. In the numerical examples, the aim is to reconstruct seven different scatterers surrounded by sources and receivers using the steepest-descent minimization method, which does not involve any computation of the Hessian matrix at each iteration. The steepest-descent method, unlike the conjugate-gradient or L-BFGS quasi-Newton [2] methods that are preferred in practice, allows us to clearly observe the differences in convergence rates among the various parameterization choices. In this regard, we employ three different modelling and inversion schemes, where the first two schemes use the Born approximation for modelling and inversion, while the third scheme uses a time-domain acoustic staggered-grid finite-difference code. For Born modelling and inversion, we use a homogeneous background model and the analytical expressions of the corresponding 2-D Green functions. With this example, we show that the findings of the conventional parameterization analysis methods are not valid in general.

This paper is organised as follows. We start with five different parameterization choices for acoustic inversion whose performance will be investigated. Then, we outline the three modelling and inversion schemes that are employed in the numerical example. Next, we briefly review the two conventional parameterization analysis methods before validating them against the numerical examples. The last section concludes the paper.

2 Parameterization Choices

We parameterize the Earth model with two parameters at each point, involving density, ρ , and compressional-wave speed, c_p , or their combinations, for instance, compressional-wave impedance, $Z_p = \rho c_p$, or bulk modulus, $K = \rho c_p^2$. We denote the model parameter of kind a by m_a and write

$$m_a(\mathbf{x}) = m_{ao}(\mathbf{x}) [1 + \xi_{m_a}(\mathbf{x})], \quad (2)$$

where $\xi_{m_a}(\mathbf{x})$ is the contrast function and an additional subscript o is used to denote the parameters corresponding to the reference medium. The subsurface location is denoted by \mathbf{x} . When the subsurface is parameterized using ξ_{m_a} and ξ_{m_b} , the model vector at each point is denoted by \mathbf{m}_{m_a, m_b} and the choice of parameterization is indicated by the subscript ‘ m_a, m_b ’. For example, in the case of $\mathbf{m}_{c_p, \rho}$ -parameterization, the subsurface is parameterized using the

Table 2: Conversion formulas used for the non-linear re-parameterization.

	Parameterization	First Parameter	Second Parameter
(A)	$\mathbf{m}_{c_p, \rho}$	$\xi_{c_p} = \sqrt{\frac{(\xi_{1/\rho}+1)}{(\xi_{1/K}+1)}} - 1$	$\xi_\rho = -\frac{\xi_{1/\rho}}{\xi_{1/\rho}+1}$
(B)	$\mathbf{m}_{Z_p, \rho}$	$\xi_{Z_p} = \sqrt{\frac{1}{(\xi_{1/\rho}+1)(\xi_{1/K}+1)}} - 1$	$\xi_\rho = -\frac{\xi_{1/\rho}}{\xi_{1/\rho}+1}$
(C)	\mathbf{m}_{c_p, Z_p}	$\xi_{c_p} = \sqrt{\frac{(\xi_{1/\rho}+1)}{(\xi_{1/K}+1)}} - 1$	$\xi_{Z_p} = \sqrt{\frac{1}{(\xi_{1/\rho}+1)(\xi_{1/K}+1)}} - 1$
(D)	$\mathbf{m}_{K, \rho}$	$\xi_K = -\frac{\xi_{1/K}}{\xi_{1/K}+1}$	$\xi_\rho = -\frac{\xi_{1/\rho}}{\xi_{1/\rho}+1}$

contrast functions of the compressional-wave speed c_p and mass density ρ . The model vector for $\mathbf{m}_{c_p, \rho}$ -parameterization is

$$\mathbf{m}_{c, \rho} = \begin{bmatrix} \xi_c \\ \xi_\rho \end{bmatrix}. \quad (3)$$

In this paper, we consider the five different parameterization choices given in the Table 1. In the case of a \mathbf{m}_{c_p, Z_p} -parameterization, perturbations in the first subsurface parameter mainly affect the transmission of waves and perturbations in the second mainly their reflections. We consider the $\mathbf{m}_{1/K, 1/\rho}$ -parameterization because the acoustic wave-operator is linear in $\xi_{1/K}$ and $\xi_{1/\rho}$, providing zero second derivatives with respect to these medium parameters. The second derivatives, which are non-zero for other parameterization choices, are used in the formulae of the full Hessian matrix as shown by [3]. These are neglected during the conventional parameterization analysis methods. We consider the $\mathbf{m}_{K, \rho}$ -parameterization, even though it has similar first-order Hessian information as that of $\mathbf{m}_{1/K, 1/\rho}$ -parameterization, to understand the importance of the second derivatives. In other words, a significant difference in the convergence rates of $\mathbf{m}_{K, \rho}$ - and $\mathbf{m}_{1/K, 1/\rho}$ -parameterizations will suggest that the conventional analysis is not sufficient for multi-parameter FWI. Now, we discuss two approaches to re-parameterize the subsurface from one parameterization to another.

2.1 Re-parameterization

The non-linear re-parameterization approach involves the exact non-linear transformation. For example, in order to obtain ξ_K from $\xi_{1/K}$, we write

$$\xi_K = \frac{K - K_o}{K_o} = \left(\frac{1}{\xi_{1/K}^{1/K_o} + 1/K_o} - K_o \right) \frac{1}{K_o} = \frac{1}{\xi_{1/K} + 1} - 1 = -\frac{\xi_{1/K}}{\xi_{1/K} + 1}. \quad (4)$$

Similarly, Table 2 gives non-linear re-parameterization formulas for various parameterization choices.

The linear re-parameterization approach assumes small contrasts, ignoring higher-order terms in the exact non-linear conversion formula. As an example, we consider a case where the subsurface is initially parameterized by $\mathbf{m}_{K, \rho}$. In order to re-parameterize the subsurface to $\mathbf{m}_{1/K, 1/\rho}$,

Table 3: Conversion formulas used for the linear re-parameterization.

	Parameterization	First Parameter	Second Parameter
(A)	$\mathbf{m}_{c_p, \rho}$	$\xi_{c_p} = \frac{1}{2} (\xi_{1/\rho} - \xi_{1/K})$	$\xi_\rho = -\xi_{1/\rho}$
(B)	$\mathbf{m}_{Z_p, \rho}$	$\xi_{Z_p} = \xi_{1/\rho} - \frac{1}{2} \xi_{1/K}$	$\xi_\rho = -\xi_{1/\rho}$
(C)	\mathbf{m}_{c_p, Z_p}	$\xi_{c_p} = \frac{1}{2} (\xi_{1/\rho} - \xi_{1/K})$	$\xi_{Z_p} = \xi_{1/\rho} - \frac{1}{2} \xi_{1/K}$
(D)	$\mathbf{m}_{K, \rho}$	$\xi_K = -\xi_{1/K}$	$\xi_\rho = -\xi_{1/\rho}$

we write the model vector as

$$\begin{aligned} \mathbf{m}_{1/K, 1/\rho} = \begin{bmatrix} \xi_{1/K} \\ \xi_{1/\rho} \end{bmatrix} &= \begin{bmatrix} -\frac{K_o}{K} & 0 \\ 0 & -\frac{\rho_o}{\rho} \end{bmatrix} \begin{bmatrix} \xi_K \\ \xi_\rho \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \xi_K \\ \xi_\rho \end{bmatrix} + \begin{bmatrix} \frac{\xi_K}{1+\xi_K} & 0 \\ 0 & \frac{\xi_\rho}{1+\xi_\rho} \end{bmatrix} \begin{bmatrix} \xi_K \\ \xi_\rho \end{bmatrix}. \end{aligned} \quad (5)$$

Now, the second term on the right-hand side of the above equation is ignored since it is of the order ξ_K^2 and ξ_ρ^2 for small ξ_K and ξ_ρ . The linear re-parameterization formula becomes

$$\mathbf{m}_{1/K, 1/\rho} \approx \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{m}_{K, \rho} = -\mathbf{m}_{K, \rho}. \quad (6)$$

Similarly, Table 3 gives linear re-parameterization formulas for other parameterization choices.

3 Modelling and Inversion

We denote the 2-D spatial coordinates by $\mathbf{x} = (x, z)$, the origin by $\mathbf{x}_0 = (0, 0)$ and the positions of sources and receivers by \mathbf{x}_s and \mathbf{x}_r , respectively. We introduce the reference Green function, $G_o(\mathbf{x}, \omega; \mathbf{x}_s)$, satisfying the 2-D acoustic wave equation

$$L_o G_o = \delta(\mathbf{x} - \mathbf{x}_s). \quad (7)$$

Here, ω denotes angular frequency. The wave operator L_o is given by

$$L_o = \frac{\omega^2}{K_o} + \nabla \cdot \frac{1}{\rho_o} \nabla, \quad (8)$$

where K_o denotes the bulk modulus and ρ_o the mass density of the reference medium. The Green function corresponding to the actual inhomogeneous medium is denoted by $G(\mathbf{x}, \omega; \mathbf{x}_s)$ and satisfies the wave equation

$$L G = \delta(\mathbf{x} - \mathbf{x}_s), \quad (9)$$

with

$$L = \frac{\omega^2}{K} + \nabla \cdot \frac{1}{\rho} \nabla. \quad (10)$$

We write Green's function in the actual inhomogeneous medium as $G = G_o + G_s$, where G_s stands for the scattered component of the total pressure wavefield. The Lippmann-Schwinger

equation produces

$$\begin{aligned}
G(\mathbf{x}_r, \omega; \mathbf{x}_s) &= G_o(\mathbf{x}_r, \omega; \mathbf{x}_s) + \\
&\int_{\mathbf{x}} G_o(\mathbf{x}_r, \omega; \mathbf{x}) \omega^2 \left(\frac{1}{K} - \frac{1}{K_o} \right) G(\mathbf{x}, \omega; \mathbf{x}_s) d\mathbf{x} + \\
&\int_{\mathbf{x}} G_o(\mathbf{x}_r, \omega; \mathbf{x}) \left(\nabla \cdot \left(\frac{1}{\rho} - \frac{1}{\rho_o} \right) \nabla G(\mathbf{x}, \omega; \mathbf{x}_s) \right) d\mathbf{x}, \quad (11)
\end{aligned}$$

where $G_o(\mathbf{x}_r, \omega; \mathbf{x})$ denotes the scatterer-to-receiver field in the unperturbed medium and $G(\mathbf{x}, \omega; \mathbf{x}_s)$ denotes the source-to-scatterer field in the perturbed medium. We now employ the Born approximation to replace the actual Green function G in the right-hand side of equation 11 with G_o . Also, we use integration by parts to obtain the Born approximation, G_b , of Green's function as

$$\begin{aligned}
G_b(\mathbf{x}_r, \omega; \mathbf{x}_s) &= G_o(\mathbf{x}_r, \omega; \mathbf{x}_s) + \\
&\int_{\mathbf{x}} \frac{\omega^2}{K_o} G_o(\mathbf{x}_r, \omega; \mathbf{x}_s) G_o(\mathbf{x}_r, \omega; \mathbf{x}_s) \xi_{1/K} d\mathbf{x} - \\
&\int_{\mathbf{x}} \frac{1}{\rho_o} [\nabla G_o(\mathbf{x}_r, \omega; \mathbf{x}_s) \cdot \nabla G_o(\mathbf{x}_r, \omega; \mathbf{x}_s)] \xi_{1/\rho} d\mathbf{x}, \quad (12)
\end{aligned}$$

$$= G_o(\mathbf{x}_r, \omega; \mathbf{x}_s) + \int_{\mathbf{x}} \mathbf{W}_{1/K, 1/\rho}^T(\mathbf{x}, \omega; \mathbf{x}_r, \mathbf{x}_s) \mathbf{m}_{1/K, 1/\rho}(\mathbf{x}) d\mathbf{x}, \quad (13)$$

where the modelling vector $\mathbf{W}_{1/K, 1/\rho}$ is given by

$$\mathbf{W}_{1/K, 1/\rho}(\mathbf{x}, \omega; \mathbf{x}_r, \mathbf{x}_s) = \left[\frac{\omega^2}{K_o} G_o(\mathbf{x}_r, \mathbf{x}) G_o(\mathbf{x}, \mathbf{x}_s) \quad - \frac{1}{\rho_o} (\nabla G_o(\mathbf{x}_r, \mathbf{x})) \cdot (\nabla G_o(\mathbf{x}, \mathbf{x}_s)) \right]^T. \quad (14)$$

We now discuss different modelling and inversion schemes that are used in the numerical experiments in this paper. In each scheme, the least-squares misfit between the artificially-generated *observed* data $Q(\mathbf{x}_r, \omega; \mathbf{x}_s)$ and the modelled data $P(\mathbf{x}_r, \omega; \mathbf{x}_s)$,

$$J_{ls} = \frac{1}{2} \sum_{\omega \geq 0} \sum_{s, r} \|P(\mathbf{x}_r, \omega; \mathbf{x}_s) - Q(\mathbf{x}_r, \omega; \mathbf{x}_s)\|^2, \quad (15)$$

is minimized. If the subsurface is parameterized by a model vector other than $\mathbf{m}_{1/K, 1/\rho}$, either the linear or the non-linear re-parameterization approach is used. Then, for forward and adjoint modelling, we use either the Born or a full-waveform approach. After the adjoint modelling, the gradient of J_{ls} with respect to $\mathbf{m}_{1/K, 1/\rho}$ is obtained by chain rule. As already mentioned, we use a steepest-descent algorithm to minimize J_{ls} .

3.1 Scheme I

With this scheme, the modelled data are obtained by using the linear re-parameterization, listed in Table 3, and the Born-approximated Green functions G_b of equation 13:

$$P(\mathbf{x}_r, \omega; \mathbf{x}_s) = \Phi(\omega) G_b(\mathbf{x}_r, \omega; \mathbf{x}_s). \quad (16)$$

Here, $\Phi(\omega)$ denotes the source spectrum, which is assumed to be known. Also, the 'observed' data are generated with G_b , committing an inverse crime, but for those, the non-linear re-parameterization approach of Table 2 is used. The flowchart in Figure 2 illustrates this scheme.

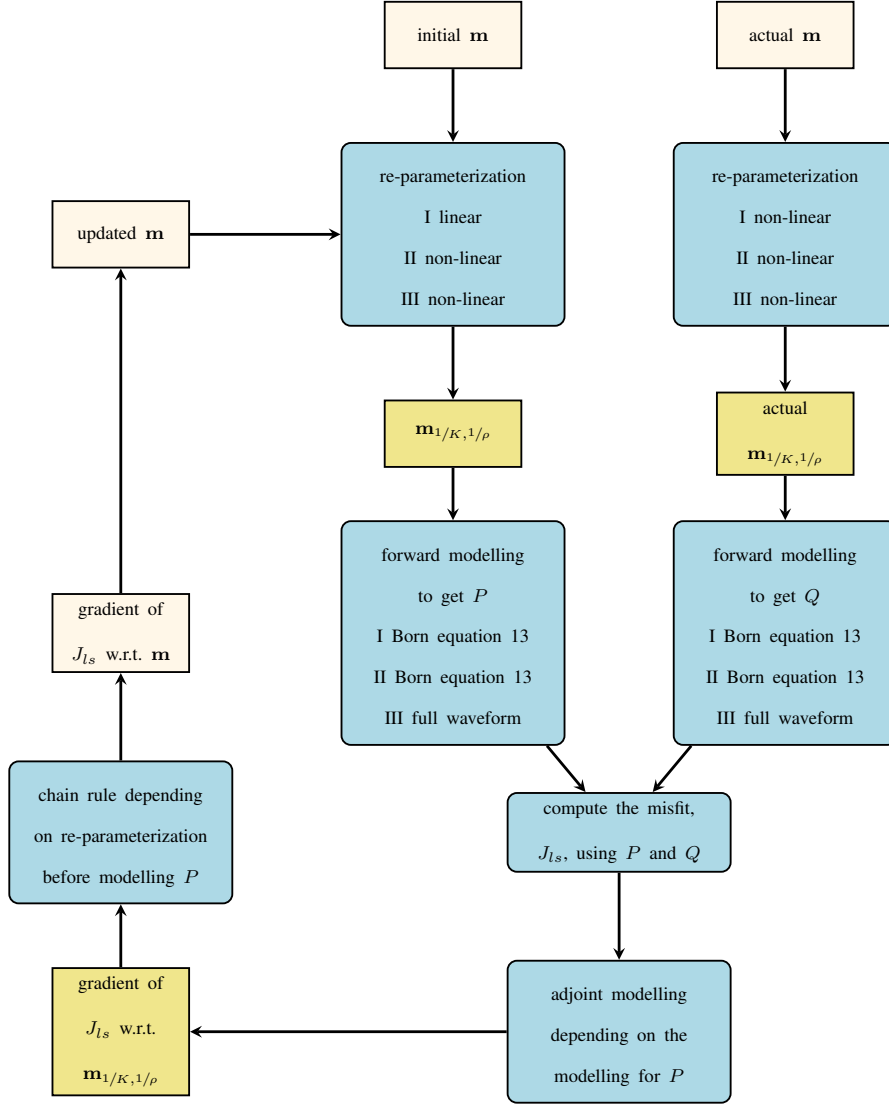


Figure 2: A flowchart to illustrate the modelling and inversion schemes I, II and III. Square boxes represent subsurface models and boxes with rounded corners represent an operation. Boxes that depend on the parameterization choice are white. Linear and non-linear re-parameterizations are described by the formulas in Tables 3 and 2, respectively.

3.2 Scheme II

As shown in Figure 2, this scheme is similar to the modelling and inversion scheme I, except that the non-linear re-parameterization, listed in Table 2, is adopted while generating both the modelled and the ‘observed’ data.

3.3 Scheme III

The numerical results obtained for this modelling and inversion scheme are decisive because it is identical to conventional acoustic full-waveform modelling and inversion. We used a time-domain acoustic staggered-grid finite-difference code to solve equation 9 for the forward as well as the adjoint wavefield computations required for the gradients [14, 15]. Absorbing boundary conditions are used on all sides of the computational domain. The flowchart in the Figure 2 also illustrates this scheme.

4 Diffraction-pattern Analysis

The most common way to analyse various parameterization choices is with diffraction patterns [17, 15, 7, 10, 5, 13, 6]. In this analysis, for each individual parameterization choice, a perturbation of both the first and the second parameter in the same scatter point at the centre, $\mathbf{x}_0 = (0, 0)$, of the model is considered, leading to an inverse problem with two parameters. When the primary wavefield, incident on these perturbations, is scattered, the contours of the scattered energy produce diffraction patterns. For example, in the case of a $\mathbf{m}_{c_p, \rho}$ -parameterization, contours of the scattered energy due to a point scatterer with velocity and density contrasts are plotted in Figures 3a and 3b, respectively. Here, the incident wavefield is generated by a point source at $(0, 490)$. Note that these plots are insensitive to the sign of the incident wavefield. The diffraction patterns for other parameterization choices are also plotted in Figure 3. Alternatively, analytical expressions derived in the framework of asymptotic ray+Born inversion [4] can be used to obtain the diffraction patterns.

Diffraction pattern analysis chooses subsurface parameterization depending on the recorded arrivals, at a particular scattering angle θ , which are being inverted or fit. A parameterization where the two diffraction patterns at θ differ as much as possible has the fastest convergence because the change in the data due to a perturbation in the first parameter is independent of that caused by the second parameter. For example, when mainly inverting arrivals recorded at short-to-intermediate scattering angles, the trade-off between the two parameters is minimum in the case of 1. \mathbf{m}_{c_p, Z_p} -parameterization — ρ doesn’t affect the data; 2. $\mathbf{m}_{Z_p, \rho}$ -parameterization — c_p doesn’t affect the data, as shown in the Figures 3c, 3d, 3e and 3f. This is due to the fact that there is an overlap in the diffraction patterns in the case of other parameterization choices (Figures 3a, 3b, 3g and 3h). Therefore, as shown by many authors, this analysis is useful to select the subspace for mono-parameter inversion in separate scattering regimes. However, in this paper, we focus on the case in which the recorded arrivals for all the scattering angles are considered simultaneously. In this case, the contrasts in both parameters can be uniquely determined as the inverse problem is almost well-posed. The diffraction-pattern analysis suggests that a \mathbf{m}_{c_p, Z_p} -parameterization is the *best* choice for an almost well-posed problem because its two diffraction patterns differ the most.

Later in this paper, we want to numerically validate this claim. In the next section, we plot the terms of the Hessian matrix to show that there is a dependence between the two parameters for all the parameterization choices, given that all the scattering angles are considered for inversion.

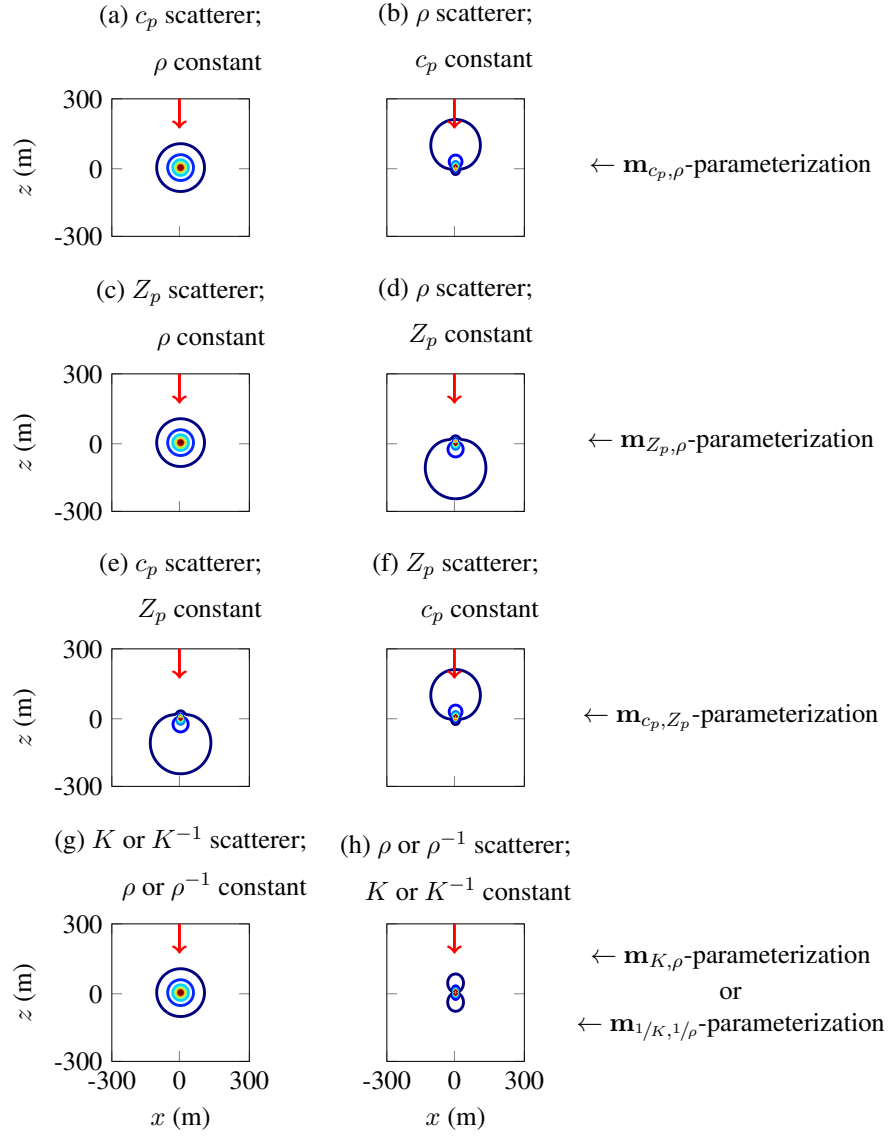


Figure 3: Diffraction patterns due to different point scatterers located at the center of the model $(0,0)$. The red arrow indicates the direction of the incident primary wavefield radiated by a source located outside the plotted domain at $(0,490)$. Point scatterers have (a) velocity and (b) density perturbations for a $\mathbf{m}_{c_p, \rho}$ -parameterization. (c), (d) For a $\mathbf{m}_{Z_p, \rho}$ -parameterization. (e), (f) For a \mathbf{m}_{c_p, Z_p} -parameterization. (g), (h) For a $\mathbf{m}_{K, \rho}$ -parameterization or $\mathbf{m}_{1/K, 1/\rho}$ -parameterization. Red-coloured contours are used for higher scattered energy than blue-coloured contours.

5 Point-scatterer Analysis

During this analysis, the inverse problem is simplified such that only the contrast of a point-shaped scatterer at a known location \mathbf{x}_0 has to be estimated, in our case $\mathbf{x}_0 = (0, 0)$. In this case, the terms of the original Hessian matrix in equation 1 are reduced to a 2 by 2 symmetric matrix since there are only two unknowns during the inversion:

$$\mathbb{H}_{a,b} = \begin{bmatrix} \mathbf{H}_{(a,0),(a,0)} & \mathbf{H}_{(a,0),(b,0)} \\ \mathbf{H}_{(b,0),(a,0)} & \mathbf{H}_{(b,0),(b,0)} \end{bmatrix}, \quad (17)$$

where \mathbf{H} was defined in equation 1. We call the matrix \mathbb{H} in equation 17 the reduced Hessian. It varies with the parameterization choice and acquisition geometry. As an example, when the modelling and inversion scheme I is employed and the subsurface is parameterized using $\mathbf{m}_{1/K,1/\rho}$, we can write \mathbb{H} in terms of the modelling vector \mathbf{W} as

$$\mathbb{H}_{1/K,1/\rho} = \sum_{\omega} \Phi(\omega)\Phi^*(\omega) \sum_{s,r} \mathbf{W}_{1/K,1/\rho}^*(\mathbf{x}_0, \omega; \mathbf{x}_r, \mathbf{x}_s) \mathbf{W}_{1/K,1/\rho}^T(\mathbf{x}_0, \omega; \mathbf{x}_r, \mathbf{x}_s). \quad (18)$$

The point-scatterer analysis compares the reduced Hessians \mathbb{H} for different parameterization choices with the objective to find the one with the fastest convergence. As the reduced Hessian also depends on the acquisition geometry, we choose, as an example, a circular acquisition geometry with sources and receivers along a circle with \mathbf{x}_0 as the centre. The eigenvalues and eigenvectors of the reduced Hessians [11, 10] for the various parameterization choices can be plotted to assess the relative convergence rate for each choice. This analysis was also employed for ray-based inversion by [4], where under the high-frequency approximation and in the absence of multiple scattering, the terms of the non-reduced Hessian matrix, $\mathbf{H}_{(a,i),(b,j)}$, are non-zero only if $i = j$. This analysis suggests that the $\mathbf{m}_{1/K,1/\rho}$ and $\mathbf{m}_{K,\rho}$ parameterization choices are equivalent and have the fastest convergence for the chosen circular acquisition geometry for the following reasons:

Condition Number. Convergence to the exact solution is possible in only one iteration step if all the eigenvalues of the reduced Hessian are equal. Intuitively, this corresponds to the case when the ellipsoidal contours of the objective function become circular. The condition number of $\mathbb{H}_{1/K,1/\rho}$ or $\mathbb{H}_{K,\rho}$, unlike the reduced Hessians for other parameterization choices, is one.

Cross-parameter Terms. The convergence is faster when the coordinate axes, which refer to the parameters, coincide with the eigenvector directions of the reduced Hessian. This corresponds to the case where the cross-parameter terms of the reduced Hessian are zero. This is true only if the subsurface is parameterized using either $\mathbf{m}_{K,\rho}$ or $\mathbf{m}_{1/K,1/\rho}$.

This analysis is not suited for realistic inverse problems, where both the shape and contrast of sub-wavelength scatterers have to be estimated, unlike the simplified one we considered with the point-shaped scatterer. To show this, we employ the modelling and inversion scheme I and plot some terms of the non-reduced Hessian matrix, $\mathbf{H}_{(a,i),(b,j)}$ with $\mathbf{x}_j = \mathbf{x}_0$, in Figure 4. Note that \mathbf{H} is a normal matrix and therefore has three unknown sub-matrices since $\mathbf{H}_{(a,i),(b,j)} = \mathbf{H}_{(b,i),(a,j)}$ for any given i and j . For this reason, only three plots are shown for each parameterization in Figure 4. It can be observed that the cross-parameter terms of the non-reduced Hessian matrix are non-zero for any given parameterization demonstrating the dependence between the two parameters. As discussed before, the cross-parameter terms of the reduced Hessian, i.e., the main cross-parameter terms of the non-reduced Hessian matrix, are zero (white colour in Figure 4k) for a $\mathbf{m}_{K,\rho}$ - or $\mathbf{m}_{1/K,1/\rho}$ -parameterization. However, the block cross-parameter terms

Table 4: Properties of different scatterers located at \mathbf{x}_0 .

Scatterer Case	Actual Properties (Shape & Contrast)	Shape Known?	Properties to be Estimated
(i)	Point; c_p -only	Yes	Contrast
(ii)	Gaussian-shaped; c_p -only	Yes	Contrast
(iii)	Gaussian-shaped; c_p -only	No	Shape & Contrast
(iv)	Point; ρ -only	Yes	Contrast
(v)	Gaussian-shaped; ρ -only	Yes	Contrast
(vi)	Gaussian-shaped; ρ -only	No	Shape & Contrast
(vii)	Gaussian-shaped; non- K	No	Shape & Contrast

are non-zero for these parameterization choices, as in Figure 4k, resulting in the dependence between an K or $1/K$ contrast at \mathbf{x}_0 and a ρ or $1/\rho$ contrast at any of the points neighbouring \mathbf{x}_0 . Hence, there is a dependency between the parameters for all the parameterization choices.

Anyway, our goal is to validate the suggestions of this analysis with the numerical examples later in this paper.

6 Almost Well-posed Example

We define an almost well-posed numerical example by placing 10 sources and 100 receivers evenly on a circle with a 490-m radius all around \mathbf{x}_0 to avoid ill-posedness problems related to incomplete illumination and angle-dependent information. Figure 5 shows the setup. In a realistic setting, with one-sided acquisition, a complete illumination of the target can never be achieved. This means, it is difficult to estimate the shape of the target without additional constraints. However, since a wide-angle one-sided acquisition can record all the angle-dependent information of the target, the inverse problem of estimating only the contrast is almost well posed. The starting model for all inversion examples is the same as the background homogeneous model with $c_p = 2000 \text{ m s}^{-1}$ and $\rho = 2000 \text{ kg m}^{-3}$. The peak frequency of the Ricker source wavelet, which is assumed to be known during inversion, is 5 Hz. The dominant wavelength of the acoustic wavefield used for imaging is 400 m.

If one aims to reconstruct model perturbations of a size comparable to the wavelength that corresponds to the dominant frequency in the data, then the problem may be considered as almost well-posed even with band-limited data. The standard deviation of the extended Gaussian-shaped scatterers is 50 m, corresponding to a size around one-fourth of the dominant wavelength. This means that the observed extended-scatterer data, as shown in the Figure 6, mainly differ from the point-scatterer data in scattering amplitudes, but not in kinematics. Note the difference in amplitude-versus-angle between Figures 6a and 6b or 6c and 6d, suggesting that the radiation-pattern analysis, which only considers point scatterers, is not valid for the extended Gaussian-shaped sub-wavelength scatterers.

For the parameterization analysis, we position one out of the seven different scatterers, listed in Table 4, at \mathbf{x}_0 . We employ the three modelling and inversion schemes to reconstruct the scatterer at \mathbf{x}_0 with different subsurface parameterizations (Table 1). During inversion, we examine the relative convergence rate of different parameterizations by displaying the least-squares data misfit as a function of the iteration count on a log-log scale. We want to observe if

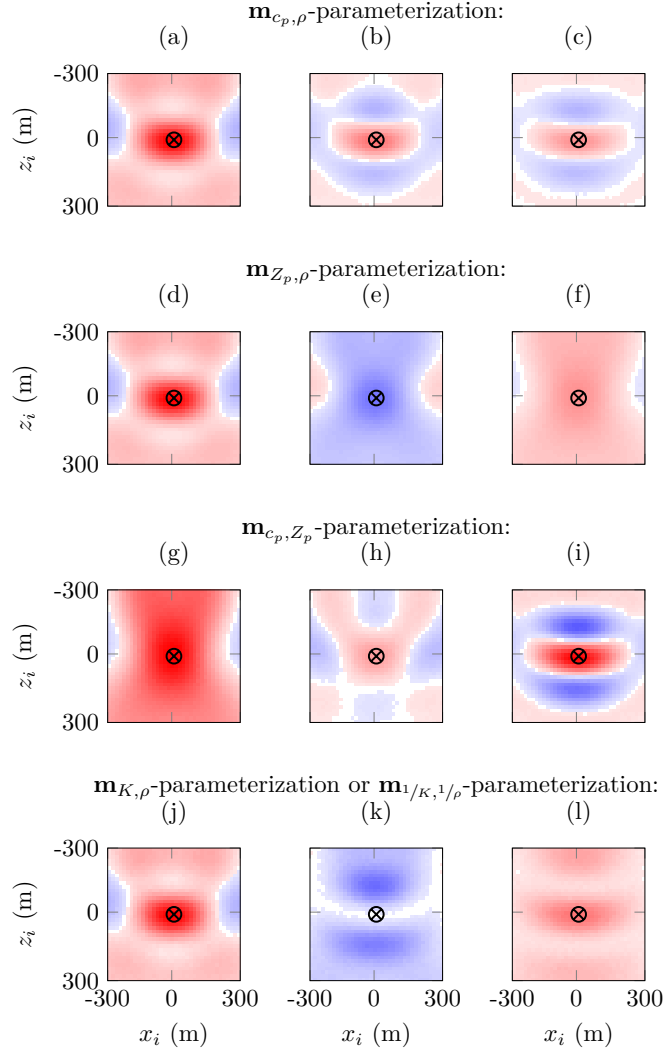


Figure 4: The terms of the non-reduced Hessian matrix, $\mathbf{H}_{(a,i),(b,j)}$ such that $\mathbf{x}_j = \mathbf{x}_0$, plotted as a function of $\mathbf{x}_i = (x_i, z_i)$ for the chosen, almost well-posed problem. The modelling and inversion scheme I is adopted. A source is located at $(0, -490)$ and receivers surround $\mathbf{x}_0 = (0, 0)$ along a circle of radius 490 m. Each row shows the sub-matrices of \mathbf{H} ; the first and third columns show the mono-parameter terms while the second column shows the cross-parameter terms, respectively. In $\mathbf{H}_{(a,i),(b,j)}$, a and b are chosen as: a) c_p and c_p ; b) c_p and ρ ; c) ρ and ρ ; d) Z_p and Z_p ; e) Z_p and ρ ; f) ρ and ρ ; g) c_p and c_p ; h) c_p and Z_p ; i) Z_p and Z_p ; j) K and K or $1/K$ and $1/K$; k) K and ρ or $1/K$ and $1/\rho$; l) ρ and ρ or $1/\rho$ and $1/\rho$. Observe that the block cross-parameter terms, plotted in the second column except at $(0, 0)$, of the non-reduced Hessian are non-zero for any given parameterization choice. For each parameterization choice, the terms of the reduced Hessian used for the point-scatterer analysis are marked with \otimes . In this figure, red and blue colours represent positive and negative values, respectively. White corresponds to values close to zero.

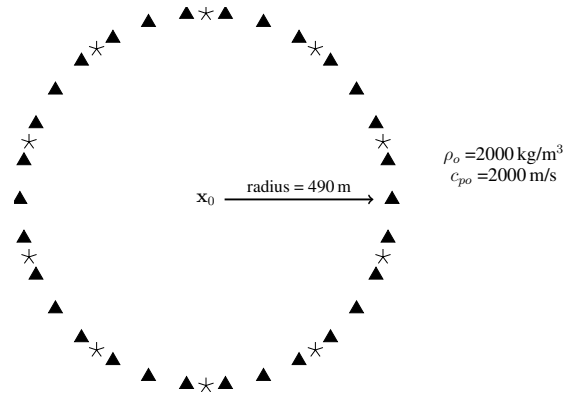


Figure 5: A sketch illustrating the almost well-posed example used for parameterization analysis. We place 10 sources, marked by stars, and 100 receivers, marked by triangles, on a circle with a 490-m radius. Different scatterers listed in the Table 4 are located at \mathbf{x}_0 .

- the suggestions of the point-scatterer analysis and/or radiation pattern analysis can be validated at least by one of the three modelling and inversion schemes used;
- the relative convergence rate of a particular parameterization choice is similar among different inversion schemes;
- the relative convergence rate of a particular parameterization choice depends on the actual contrast of the scatterer — we have chosen scatterers (i)–(iii) with a c_p -only contrast, (iv)–(vi) with a ρ -only contrast and scatterer (vii) such that there is no contrast in K ;
- the relative convergence rate of a particular parameterization choice depends on the scatterer properties that are to be estimated — only the contrast of the scatterers (i), (ii), (iv) and (v) has to be estimated, while both the contrast and shape are unknown for other scatterers;
- the relative convergence rate of a particular parameterization choice during reconstructing point-shaped scatterers is different when compared to Gaussian-shaped scatterers — scatterers (i) and (iv) are point-shaped while the others are Gaussian-shaped.

6.1 Error-bowl Analysis

Since there are only two unknown variables while reconstructing scatterers (i), (ii), (iv) and (v) with a known shape (see Table 4), we can plot two-dimensional logarithmic contours of the least-squares data misfit, i.e., contours of $-\log_{10}(J_{ls})$, as a function of the variables. These contours or error bowls can be used to assess the observed relative convergence rate of each parameterization choice. Similar to the analysis using the Hessian matrix, the relative convergence rate for each parameterization can also be assessed based on the (a) ellipticity of the misfit contours; (b) angle between the error vector and one of the principle axis of the contours. We call this an error-bowl analysis. In this analysis, a more elliptic or less circular error bowl means that it is more sensitive to one parameter than the other. The orientation angle of an error bowl determines the correlation between the two parameters. An orientation angle of 0° or 90° means that the parameters are least correlated while an angle of 45° or 135° means that they are most correlated.

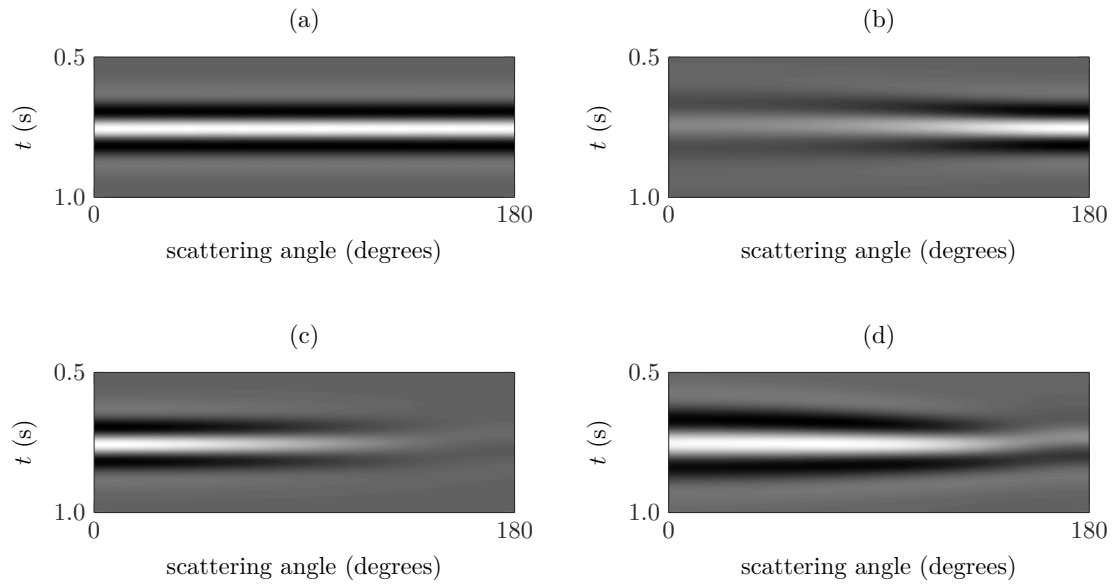


Figure 6: Observed data as a function of scattering angle due to: a) point c_p -only scatterer (i); b) Gaussian-shaped c_p -only scatterer (ii); c) point ρ -only scatterer (iv); d) Gaussian-shaped ρ -only scatterer (v). The data due to the Gaussian-shaped scatterers are different from those of the point scatterers, even though they have a size less than a quarter of the dominant wavelength. The scattering amplitude-versus-angle for point scatterers in these plots is determined by the radiation patterns in Figure 3.

So the best convergence rate is for a circular error bowl and the worst convergence rate is for an elliptical bowl oriented at either 45° or 135° angle.

We are analysing the same simplified inverse problem as that of the point-scatterer analysis, when the error bowls are plotted for the point-shaped scatterers (i) and (iv) after adopting scheme I. Therefore, the error-bowl analysis is the same as the point-scatterer analysis in this case, because the shape and orientation of the error bowls in Figure 7 are determined by the reduced Hessians. For example, the error bowls for \mathbf{m}_{K,ρ^-} or $\mathbf{m}_{1/K,1/\rho^-}$ -parameterization, in Figure 7, are circular due to the fact that the condition number of $\mathbb{H}_{1/K,1/\rho^-}$ or $\mathbb{H}_{K,\rho}$ is one. Furthermore, the error-bowl analysis can be seen as an extension of the point-scatterer analysis, in which different modelling and inversion schemes can be employed along with arbitrarily shaped scatterers. It has to be noted that it is impossible to derive analytical expressions of the reduced Hessians in the case of arbitrarily shaped scatterers, when full-waveform modelling is employed.

6.2 Point-shaped Scatterers

We first consider the inverse problems of reconstructing the point-shaped scatterer (i) with c_p -only contrast in Table 4. Appendix A, which reconstructs an equivalent scatterer (iv) with ρ -only contrast, provides additional support to the discussion in this subsection.

When scheme I is adopted for modelling and inversion, the error-bowl analysis, identical to the point-scatterer analysis, suggests that \mathbf{m}_{K,ρ^-} or $\mathbf{m}_{1/K,1/\rho^-}$ -parameterizations have the fastest convergence because of their corresponding circular error bowls, as plotted in Figures 7c and 7e. In addition to that, we observe that the bowls for \mathbf{m}_{c_p,Z_p^-} -parameterization are more circular than either the \mathbf{m}_{c_p,ρ^-} or \mathbf{m}_{Z_p,ρ^-} -parameterization. The orientation also explains the slightly slower convergence. The least-squares misfit plotted against the iteration count while estimating the c_p -only contrast, in Figure 7f shows that the \mathbf{m}_{c_p,Z_p^-} , \mathbf{m}_{K,ρ^-} and $\mathbf{m}_{1/K,1/\rho^-}$ -parameterizations have faster convergence than the others.

We now employ the modelling and inversion scheme II. Figure 8 shows that the ellipticity of the error bowls for all the parameterizations, except for the $\mathbf{m}_{1/K,1/\rho^-}$ -parameterization, are different from the previous case. Note that the \mathbf{m}_{K,ρ^-} -parameterization is no longer equivalent to the $\mathbf{m}_{1/K,1/\rho^-}$ -parameterization because of the non-linear re-parameterization. However, similar to the previous case with the scheme I, we observe that the \mathbf{m}_{c_p,Z_p^-} , \mathbf{m}_{K,ρ^-} and $\mathbf{m}_{1/K,1/\rho^-}$ -parameterizations converge faster.

Figure 9 display the error bowls when modelling and inversion with scheme III is employed. It can be seen that the suggestions of the point-scatterer analysis remain valid even when employing the modelling and inversion scheme II or III. This is because the error bowls (c)–(e) are more *circular* or less *elliptical* and more oriented towards 0° or 90° than the error bowls (a)–(b) in the Figures 8 and 9.

6.3 Gaussian-shaped Scatterers

We now want to see if the sub-wavelength Gaussian-shaped scatterer (ii) with c_p -only contrast at \mathbf{x}_0 will result in error bowls of a different shape compared to the point-shaped scatterers of the previous subsection. Appendix A, which reconstructs an equivalent scatterer (v) with ρ -only contrast, provides additional support to the discussion in this subsection. When the modelling and inversion scheme I is employed, in Figure 10, the error bowls are more elliptical for the \mathbf{m}_{c_p,ρ^-} , \mathbf{m}_{Z_p,ρ^-} and \mathbf{m}_{c_p,Z_p^-} -parameterization choices than those with point-shaped scatterers in Figure 7. The bowl of \mathbf{m}_{Z_p,ρ^-} -parameterization is maximally oriented towards a 45° angle compared to the others. Moreover, the error bowls for the \mathbf{m}_{K,ρ^-} and $\mathbf{m}_{1/K,1/\rho^-}$ -parameterizations are not circular

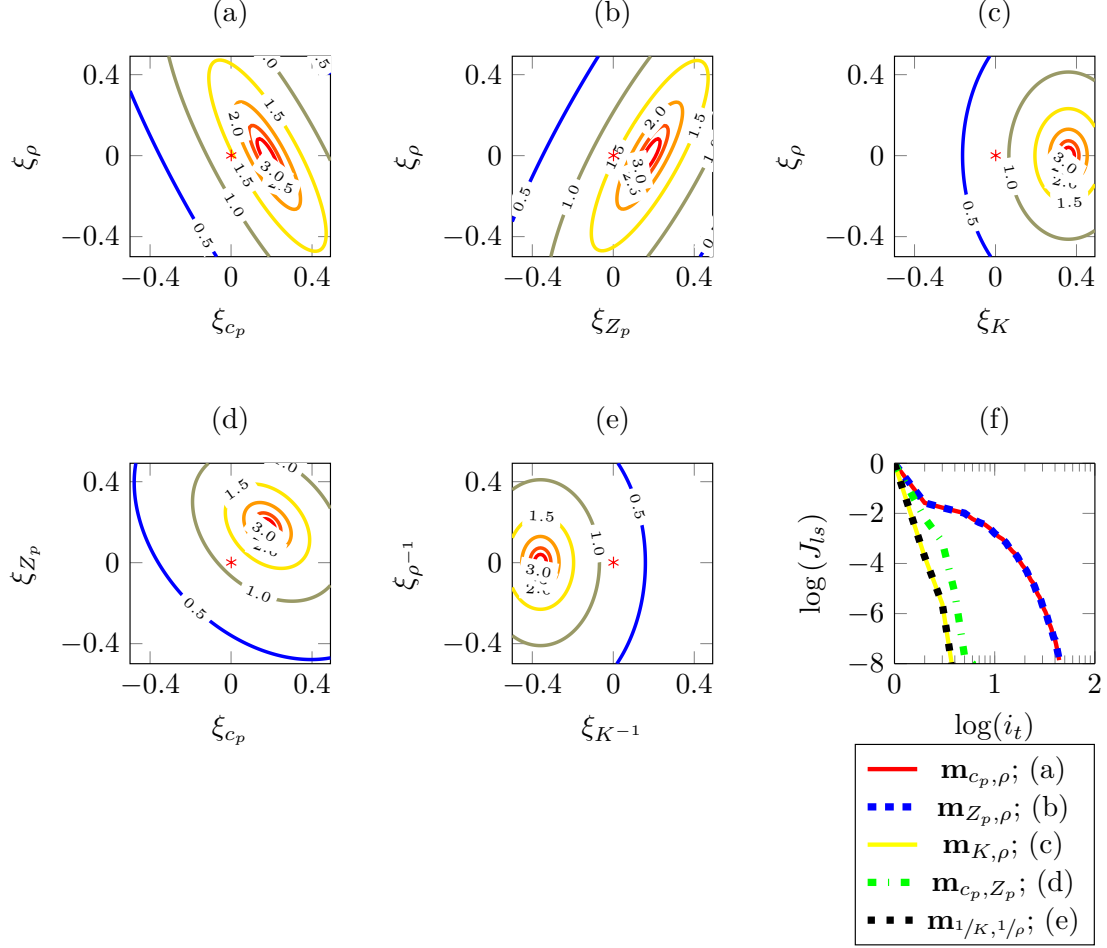


Figure 7: Reconstruction of the point-shaped c_p -only scatterer (i) when the modelling and inversion scheme I, with the Born approximation and linear re-parameterization, is adopted. Error-bowl analysis is performed by plotting the logarithmic contours of the least-squares data misfit as a function of the two subsurface parameters while using the following parameterization choices: a) $\mathbf{m}_{c_p, \rho}$ — slower convergence expected due to high ellipticity; b) $\mathbf{m}_{Z_p, \rho}$ — slower convergence expected due to high ellipticity; c) $\mathbf{m}_{K, \rho}$ —expect faster convergence due to circular contours; d) \mathbf{m}_{c_p, Z_p} —expect faster convergence due to almost circular contours; e) $\mathbf{m}_{1/K, 1/\rho}$ —expect faster convergence due to circular contours. In all the plots, the starting homogeneous model, $(0,0)^T$, is marked by the red star. f) The least-squares data misfit is plotted as a function of the iteration count on a log-log scale. This plot shows that the suggestions of the point-scatterer analysis are valid in this case.

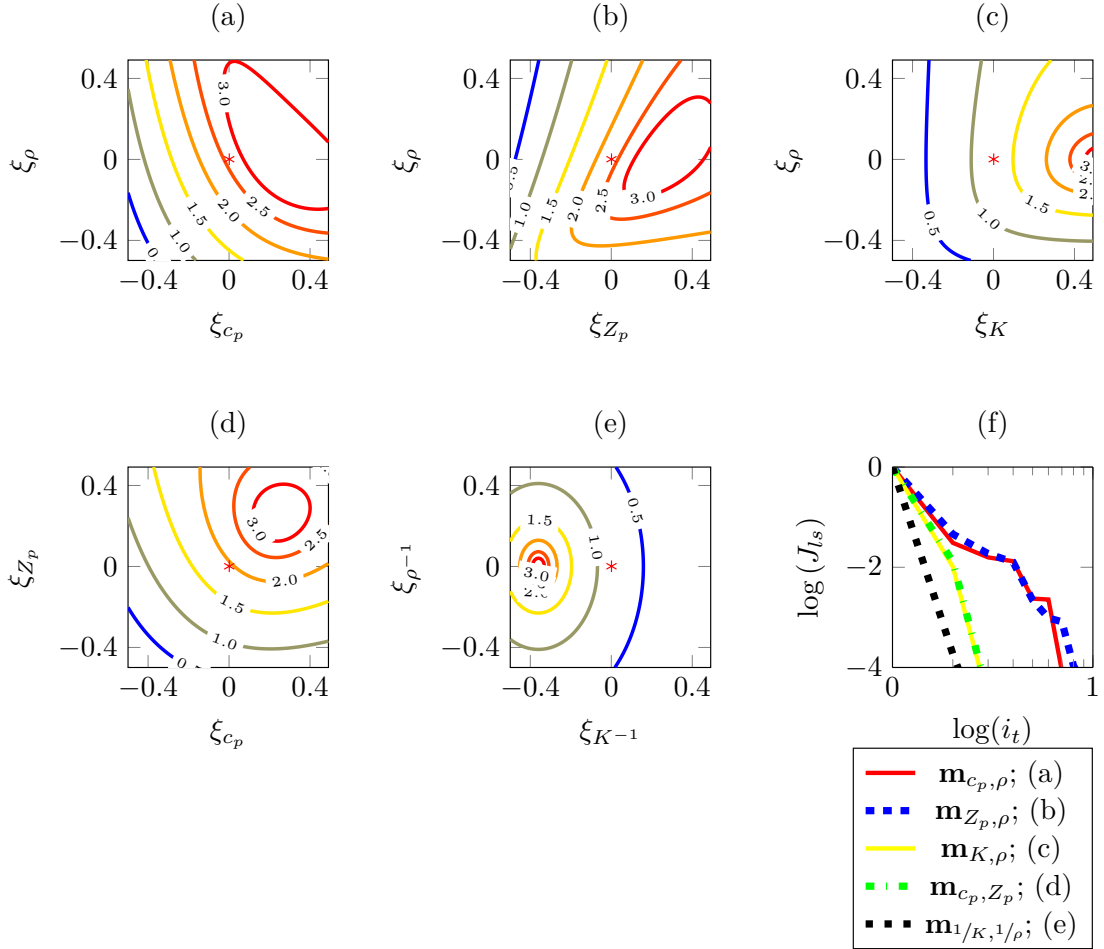


Figure 8: Same as Figure 7, except for adopting the modelling and inversion scheme II, with the Born approximation and non-linear re-parameterization. It can be seen that the error bowls for the $\mathbf{m}_{K, \rho}$ -parameterization are not equivalent to the $\mathbf{m}_{1/K, 1/\rho}$ -parameterization because of the non-linear re-parameterization. Also, the shapes of the error bowls (a)–(d) are different compared to that of Figure 7. It can be noted that the error bowls (c)–(e) are more *circular* than error bowls (a)–(b). The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are valid.

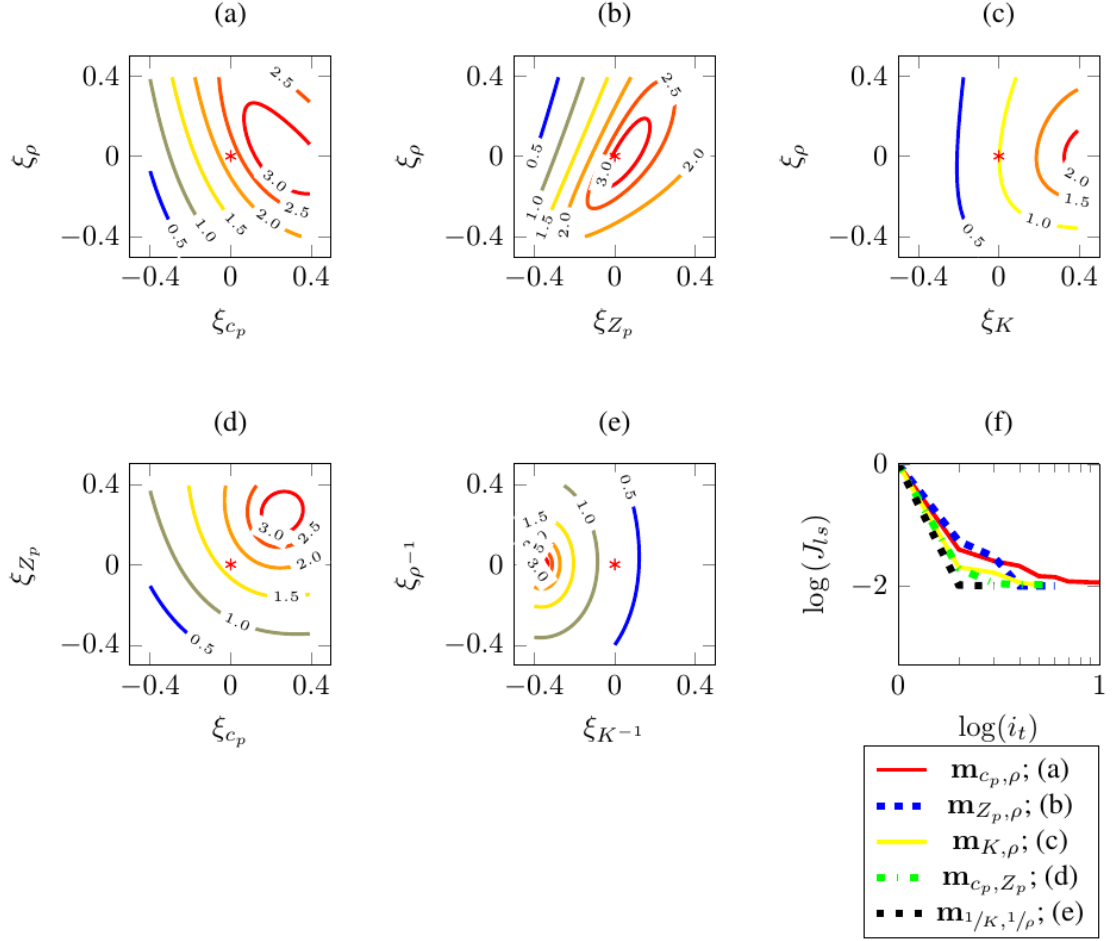


Figure 9: Same as Figure 7, except for adopting scheme III, with full-waveform modelling and inversion. We see that the shapes of the error bowls and the relative convergence rates of different parameterization choices are similar to that of Figure 8. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are valid.

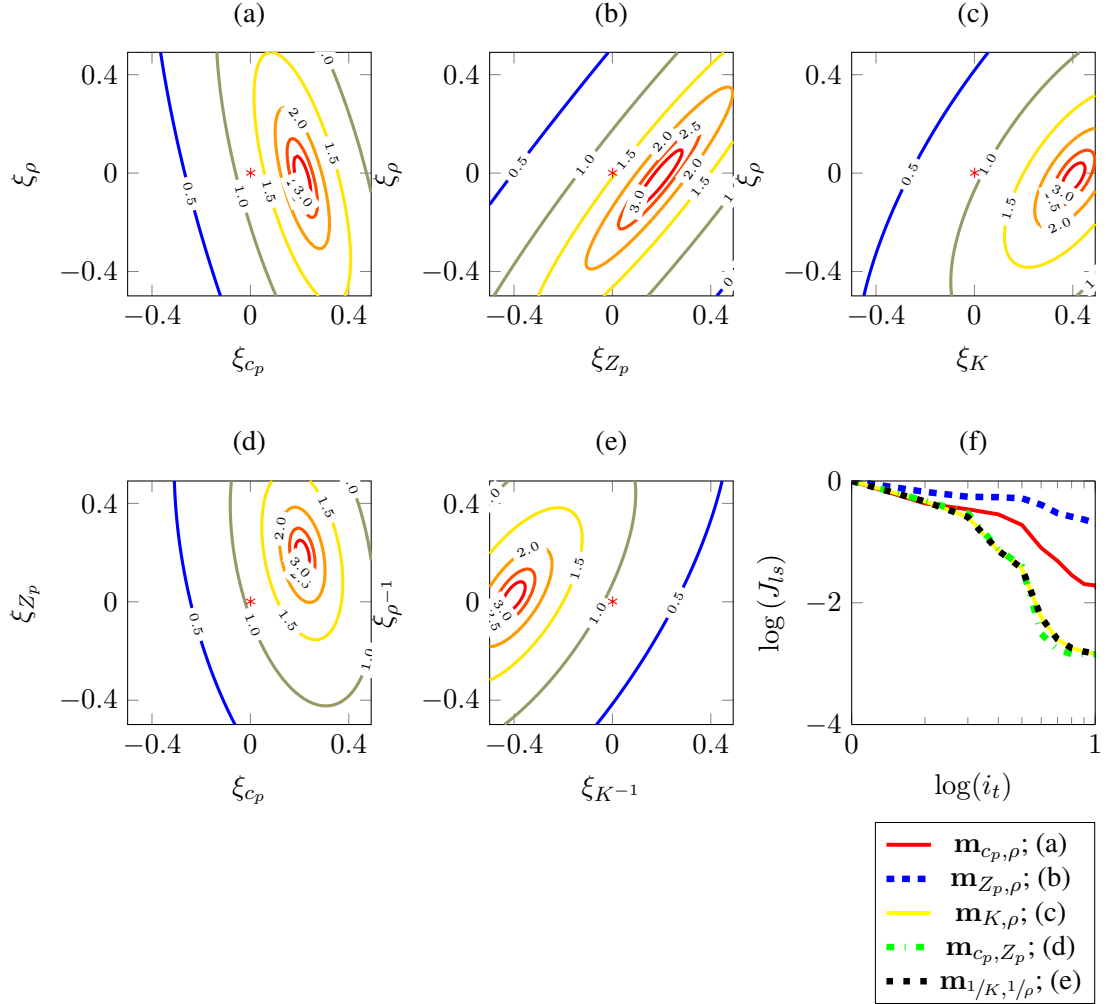


Figure 10: Same as Figure 7, except for the Gaussian-shaped c_p -only scatterer (ii). We observe that the errors bowls for all the parameterization choices are more elliptical compared to that of Figure 7. The least-squares misfit plot shows that the suggestion of the point-scatterer analysis is valid. The contours in the case $\mathbf{m}_{1/K, 1/\rho}$ - and $\mathbf{m}_{K, \rho}$ -parameterizations are not circular. The rate of convergence in the case of \mathbf{m}_{c_p, Z_p} -, $\mathbf{m}_{K, \rho}$ - and $\mathbf{m}_{1/K, 1/\rho}$ -parameterizations is higher as their corresponding error bowls are more circular compared to the others.

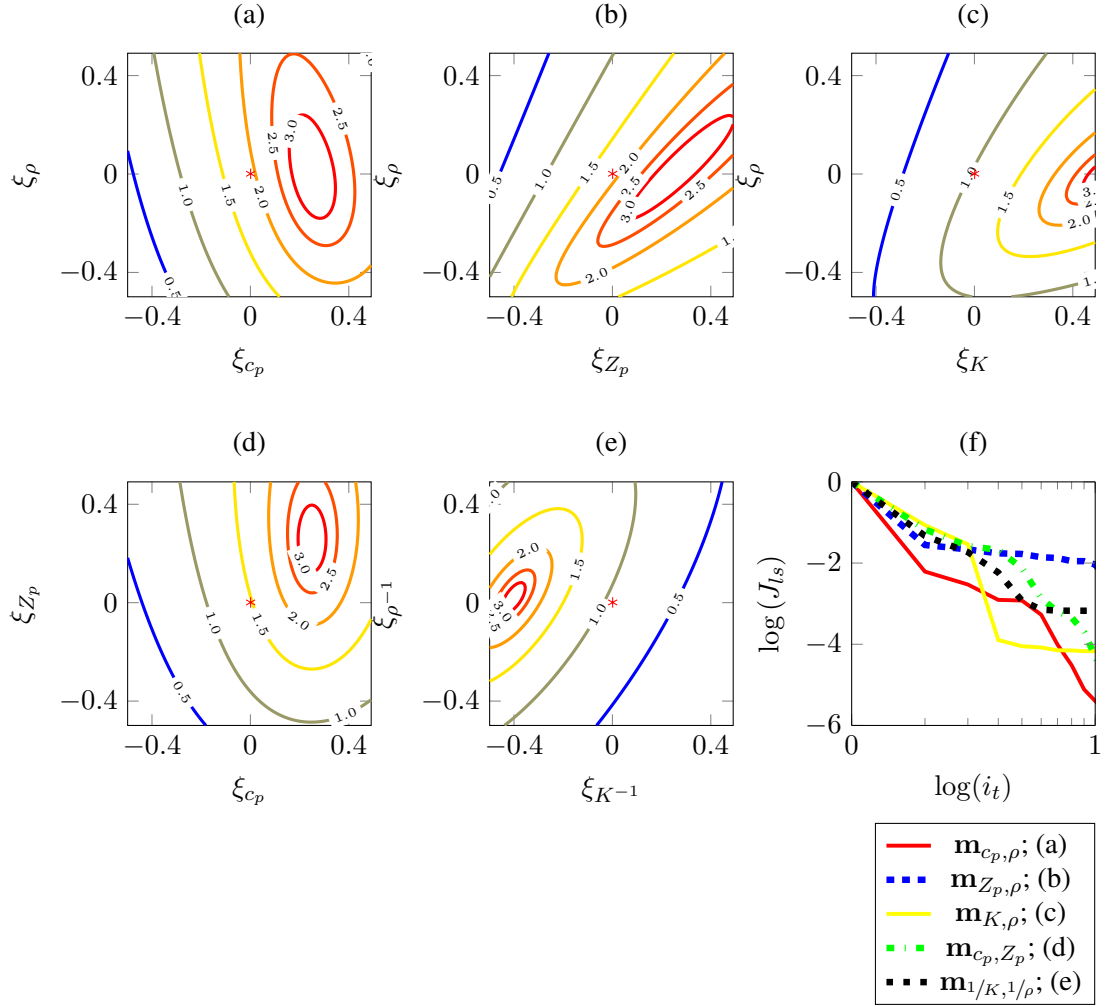


Figure 11: Same as Figure 7, except for scatterer (ii) and adopting scheme II, with the Born approximation and non-linear re-parameterization. Compared to the Figure 10, we see that the error bowls, in this case, have much different orientation. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are no longer valid.

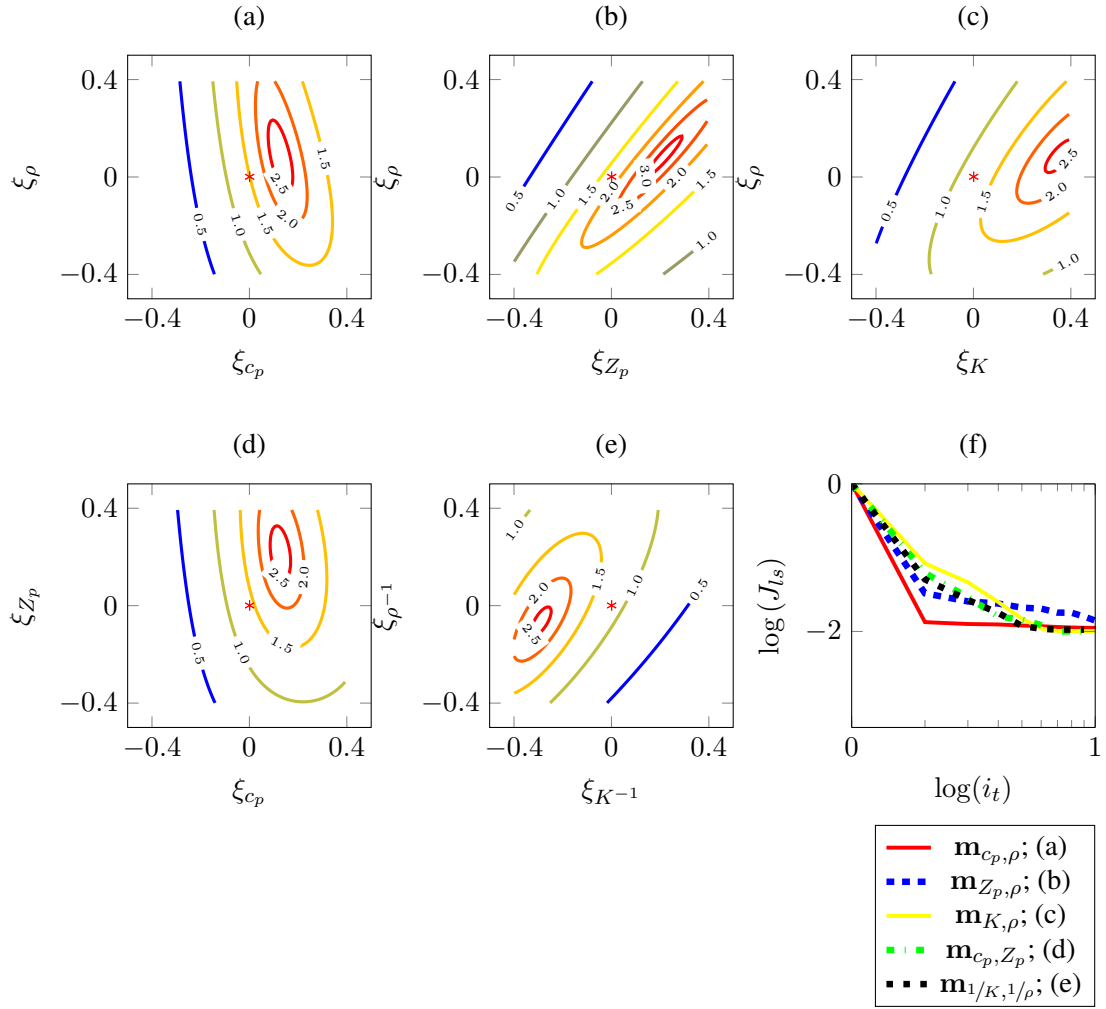


Figure 12: Same as Figure 7, except for scatterer (ii) and adopting scheme III, with full-waveform modelling and inversion. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are no longer valid. Parameterization using $\mathbf{m}_{c_p, \rho}$ has the best convergence rate.

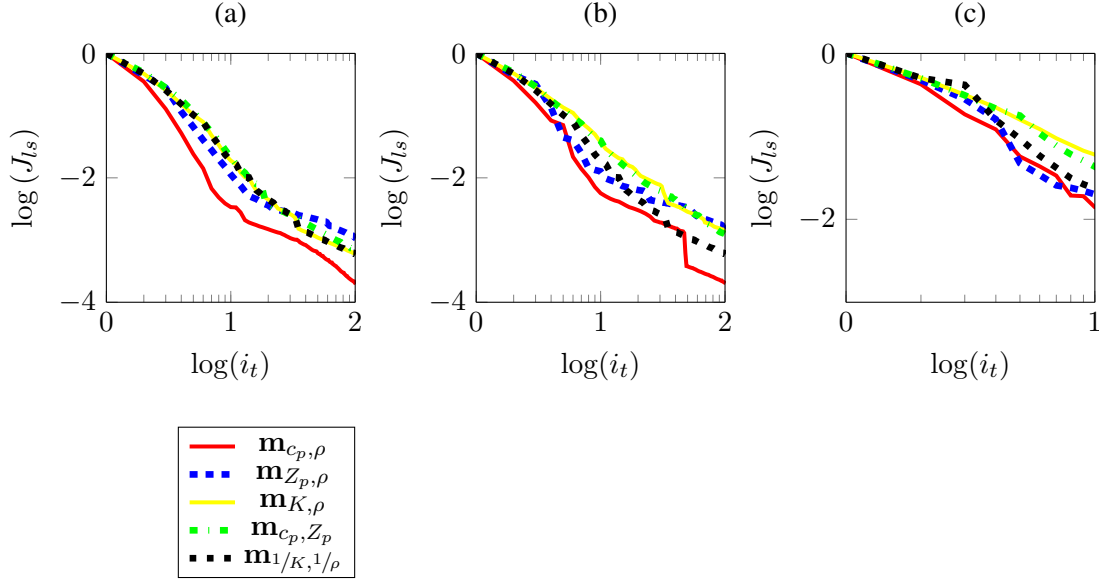


Figure 13: The least-squares misfit is plotted as a function of the iteration count on a log-log scale during the reconstruction of scatterer (iii), where both shape and contrast of a c_p -only scatterer are unknown. The modelling and inversion schemes (a) I, (b) II and (c) III are adopted. It can be observed that the parameterization using $\mathbf{m}_{K, \rho}$ has relatively lower convergence rate when all the three cases are taken into account.

any more. We still observe that the \mathbf{m}_{c_p, Z_p} -, $\mathbf{m}_{K, \rho}$ - and $\mathbf{m}_{1/K, 1/\rho}$ -parameterizations have a faster convergence because their corresponding error bowls are more circular.

When either modelling and inversion scheme II or III is employed, the error bowls for all parameterizations not only have different ellipticities but also different orientations. This indicates that the suggestions of the point-scatterer analysis are no longer valid. We observe that the $\mathbf{m}_{c_p, \rho}$ -parameterization is faster in reconstructing a c_p -only contrast (Figure 11f and 12f) and that the $\mathbf{m}_{1/K, 1/\rho}$ -parameterization has the fastest convergence when reconstructing a ρ -only contrast (see Appendix). Hence, when modelling and inversion schemes II and III are employed to reconstruct non-point shaped scatterers, the rate of convergence of a particular parameterization also depends on the type of contrast that has to be reconstructed.

6.4 Scatterers with Unknown Shape and Contrast

We now consider the inverse problems of estimating both the shape and contrast of the sub-wavelength Gaussian scatterers (iii), (vi) and (vii). The error bowls cannot be easily plotted in this case because there are more than two unknown variables. As shown in Figures 13, 14 and 15, the $\mathbf{m}_{c_p, \rho}$ -parameterization is the fastest while estimating a c_p -only scatterer, whereas the \mathbf{m}_{c_p, Z_p} -, $\mathbf{m}_{K, \rho}$ - and $\mathbf{m}_{1/K, 1/\rho}$ -parameterizations have a faster convergence while estimating a ρ -only or a non- K scatterer. This observation is similar to that of estimating the contrast of the Gaussian-shaped scatterers in the last subsection.

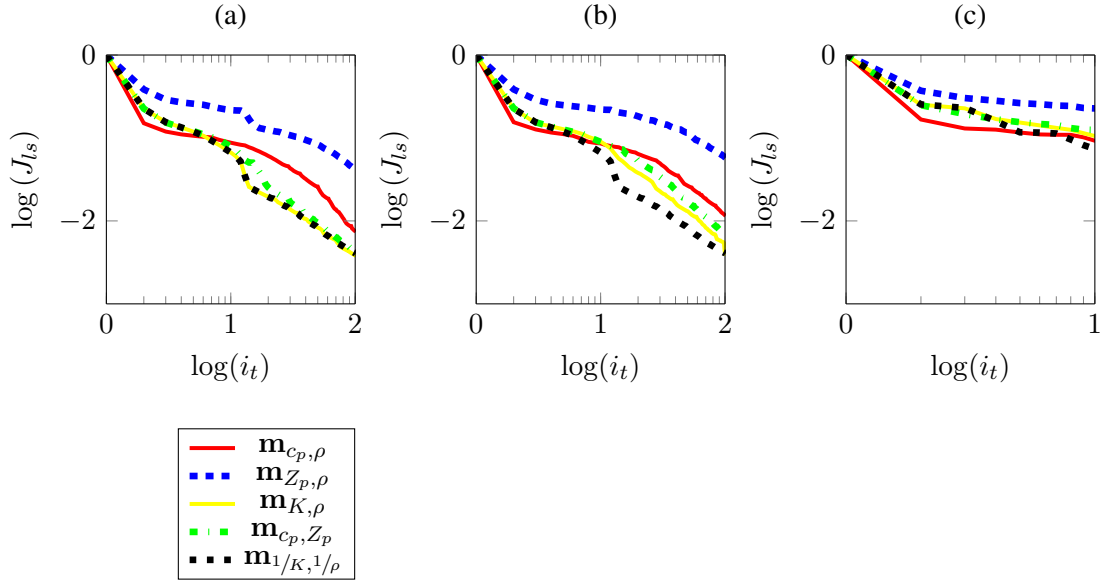


Figure 14: Same as Figure 13, except for scatterer (vi), with ρ -only contrast. Note that the relative rates of convergence of various parameterization choices are different compared to that of the Figure 13. It can be observed that the parameterization using $\mathbf{m}_{Z_p, \rho}$ has the worst convergence rate in all the cases.

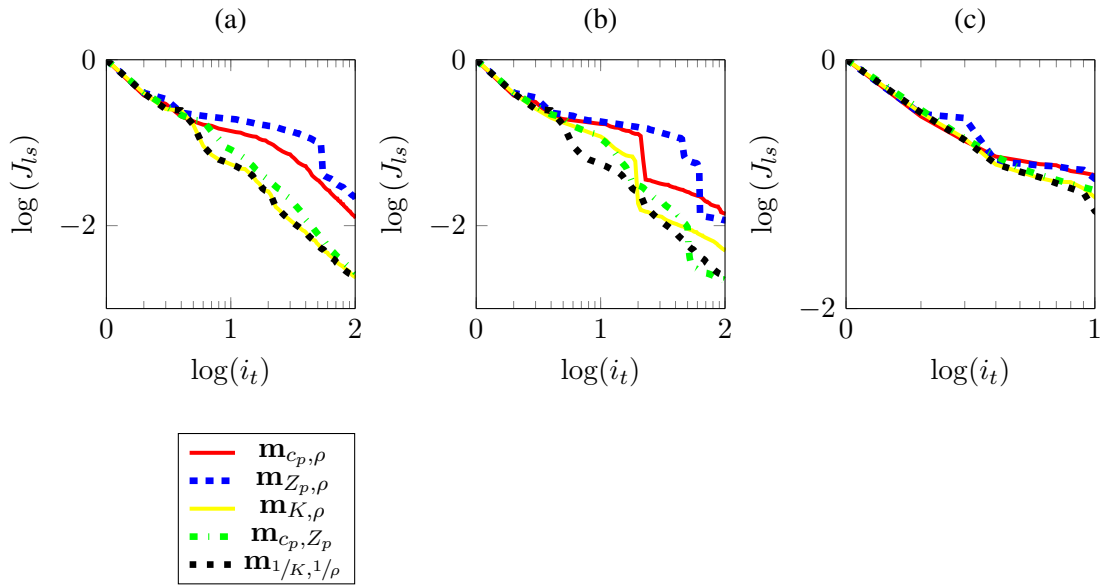


Figure 15: Same as Figure 13, except for scatterer (vii), with a non- K contrast. Note that relative rates of convergence of various parameterization choices are different compared to that of the Figures 13 and 14. It can be observed that the parameterization using $\mathbf{m}_{Z_p, \rho}$ has the worst convergence rate in all the cases.

7 Hierarchical Inversion

Let us consider the problem of finding the best parameterization choice for a multi-parameter inversion with limited-aperture acquisition. Such inversions are common during hierarchical data-fitting approaches, for example, from wide angles to short offsets. In Figure 16, we plotted the $\mathbf{m}_{c_p, \rho}$ and $\mathbf{m}_{1/\kappa, 1/\rho}$ error bowls similar to Figure 7, except in two separate regimes:

- short offsets with 0° – 90° scattering angles only;
- long offsets with 135° – 180° scattering angles only.

In both the regimes, it can be seen that the lack of angle-dependent information has resulted in error bowls with long *valleys* pointing in the *null-space* direction. Lack of angle-dependent information in the data makes it difficult to converge to the true solution. Depending on the starting model, the inversion just chooses the shortest possible path to reach the valley and stagnates. The acoustic inverse problem is non-unique when the source-receiver aperture is limited and/or the necessary frequencies are lacking from the data. It is known that in the presence of non-uniqueness, different choices of parameterization may lead to different inversion results (see, for example, [1]), each explaining the data at convergence. Hence, the search for the *best* parameterization in terms of convergence speed might be obfuscated by non-uniqueness problems for multi-parameter inversion in the separated regimes that are described above. However, one can still perform mono-parameter inversion in the subspace orthogonal to the null-space, by assuming a linearized problem, in each of these regimes.

8 Interpretation

We now interpret the least-squares misfit graphs in Figures 7–15 and 17–22, which are plotted as a function of the iteration count during the reconstruction of different scatterers in Table 4. In Table 5, the parameterization choices are divided into three classes depending on the relative convergence rate: the choices with the faster and slower convergence are coloured blue and red, respectively; the choices which neither have a faster nor a slower convergence are coloured in black. Table 5 also indicates if there is a significant difference in the convergence rate among the parameterizations.

We observe that there is no best parameterization in a general sense since parameterizations (C), (D) and (E) perform similarly good in most of the cases. Also, note that these parameterizations have a slower convergence when reconstructing scatterer (iii). However, on the other side of the spectrum, we observe that the $\mathbf{m}_{Z_p, \rho}$ -parameterization (case(B)) has the worst convergence rate, because it is coloured red or black almost in all the cases of Table 5. It can also be noticed that the $\mathbf{m}_{c_p, \rho}$ -parameterization (case (A)) is the second worst one, because there is no significant difference in its convergence even when it is the fastest.

9 Conclusions

We have briefly outlined the conventional point-scatterer and the diffraction-pattern parameterization-analysis methods for the 2-D acoustic inverse problem. Using almost well-posed-numerical examples, we have shown that the suggestions of these conventional methods are valid only when the contrasts of the point-shaped scatterers at a known location are estimated. The numerical examples employ three different modelling and inversion schemes using both Born and full-waveform modelling. As expected, for almost well-posed inverse problems, we observed that a change in

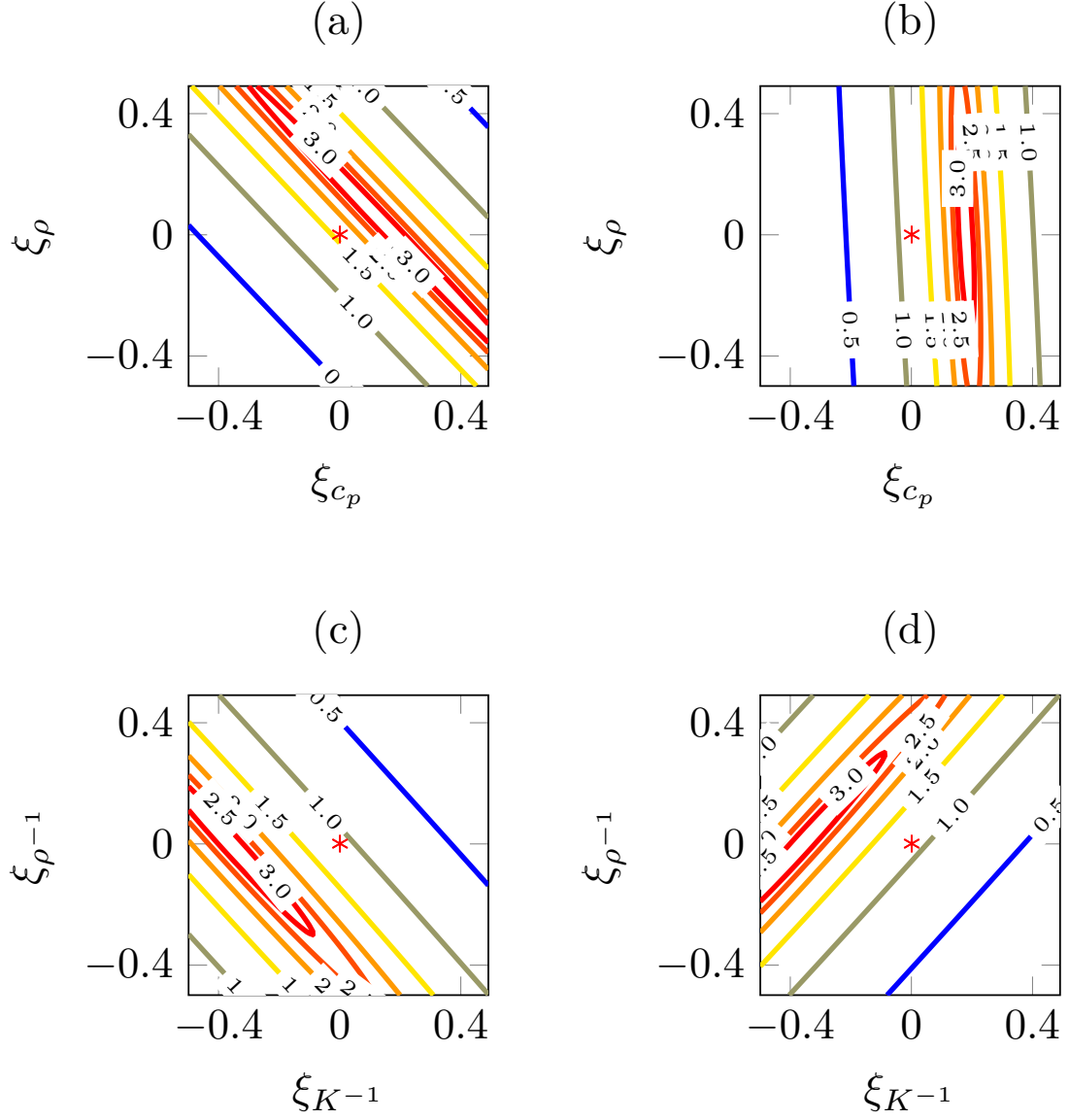


Figure 16: Similar to Figure 7, except a) using $\mathbf{m}_{c_p, \rho}$ -parameterization and short-offset regime with 0° – 90° scattering angles only; b) using $\mathbf{m}_{c_p, \rho}$ -parameterization and long-offset regime with 135° – 180° scattering angles only; c) same as (a), but for $\mathbf{m}_{1/K, 1/\rho}$ -parameterization; d) same as (b), but for $\mathbf{m}_{1/K, 1/\rho}$ -parameterization.

Table 5: Reconstruction of seven different scatterers in Table 4 when different modelling and inversion schemes are employed. The parameterizations, listed in Table 1, with faster and slower convergence relative to each other are in blue and red, respectively.

Scatterer	Scheme I	Scheme II	Scheme III
(i)	(A) (B) (C) (D) (E) Figure 7	(A) (B) (C) (D) (E) Figure 8	(A) (B) (C) (D) (E) Figure 9
(ii)	(A) (B) (C) (D) (E) Figure 10	(A) (B) (C) (D) (E) Figure 11	(A) (B) (C) (D) (E) (no significant difference) Figure 12
(iii)	(A) (B) (C) (D) (E) Figure 13a	(A) (B) (C) (D) (E) (no significant difference) Figure 13b	(A) (B) (C) (D) (E) (no significant difference) Figure 13c
(iv)	(A) (B) (C) (D) (E) Figure 17	(A) (B) (C) (D) (E) Figure 18	(A) (B) (C) (D) (E) Figure 19
(v)	(A) (B) (C) (D) (E) Figure 20	(A) (B) (C) (D) (E) Figure 21	(A) (B) (C) (D) (E) Figure 22
(vi)	(A) (B) (C) (D) (E) Figure 14a	(A) (B) (C) (D) (E) Figure 14b	(A) (B) (C) (D) (E) (no significant difference) Figure 14c
(vii)	(A) (B) (C) (D) (E) Figure 15a	(A) (B) (C) (D) (E) Figure 15b	(A) (B) (C) (D) (E) (no significant difference) Figure 15c

parameterization will result in a different convergence rate. Furthermore, the relative rate of convergence for a particular choice of parameterization depends on (a) the modelling and inversion scheme employed; (b) the contrast of subsurface scatterer that has to be reconstructed; (c) the shape of the sub-wavelength scatterer. We observed by considering all the cases that the $\mathbf{m}_{Z_p, \rho}$ -parameterization has the worst convergence rate, followed by the $\mathbf{m}_{c_p, \rho}$ -parameterization. Finally, our numerical examples show that, in general, there is no such thing as the best parameterization choice that provides the fastest convergence for acoustic inversion.

10 Acknowledgments

The first author thanks Prof. Kees Wapenaar (TU Delft) for helpful discussions. This work is a part of the NeTTUN project, funded from the European Commissions Seventh Framework Programme for Research, Technological Development and Demonstration (FP7 2007-2013) under Grant Agreement 280712. The use of supercomputer facilities was sponsored by the ‘Nederlandse Organisatie voor Wetenschappelijk Onderzoek’ (Netherlands Organisation for Scientific Research, NWO). The computations were carried out on the Dutch national e-infrastructure maintained by the SURF Foundation (www.surfsara.nl).

References

- [1] BHARADWAJ, P., MULDER, W., AND DRIJKONINGEN, G. Parametrization for 2-D SH full waveform inversion. In *76th EAGE Conference and Exhibition 2014* (2014).
- [2] BYRD, R. H., LU, P., NOCEDAL, J., AND ZHU, C. Y. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16, 5 (1995), 1190–1208.
- [3] FICHTNER, A., AND TRAMPERT, J. Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International* 185, 2 (2011), 775–798.
- [4] FORGUES, E., AND LAMBARÉ, G. Parameterization study for acoustic and elastic ray plus born inversion. *Journal of Seismic Exploration* 6, 2-3 (1997), 253–277.
- [5] GHOLAMI, Y., BROSSIER, R., OPERTO, S., RIBODETTI, A., AND VIRIEUX, J. Which parameterization is suitable for acoustic vertical transverse isotropic full waveform inversion? Part 1: Sensitivity and trade-off analysis. *Geophysics* 78, 2 (2013), R81–R105.
- [6] HE, W., AND PLESSIX, R.-É. Analysis of different parameterisations of waveform inversion of compressional body waves in an elastic transverse isotropic Earth with a vertical axis of symmetry. *Geophysical Prospecting* (2016), 1–21.
- [7] MALINOWSKI, M., OPERTO, S., AND RIBODETTI, A. High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full-waveform inversion. *Geophysical Journal International* 186, 3 (2011), 1179–1204.
- [8] MÉTIVIER, L., BRETAUDEAU, F., BROSSIER, R., OPERTO, S., AND VIRIEUX, J. Full waveform inversion and the truncated Newton method: Quantitative imaging of complex subsurface structures. *Geophysical Prospecting* 62, 6 (2014), 1353–1375.
- [9] MODRAK, R., TROMP, J., AND YUAN, Y. On the choice of material parameters for elastic waveform inversion. In *SEG Technical Program Expanded Abstracts 2016* (2016), Society of Exploration Geophysicists, pp. 1115–1119.

- [10] OPERTO, S., GHOLAMI, Y., PRIEUX, V., RIBODETTI, A., BROSSIER, R., METIVIER, L., AND VIRIEUX, J. A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice. *The Leading Edge* 32, 9 (2013), 1040–1054.
- [11] PLESSIX, R.-E., AND CAO, Q. A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium. *Geophysical Journal International* 185, 1 (2011), 539–556.
- [12] PRATT, R. G., C.SHIN, AND HICKS, G. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International* 133, 2 (1998), 341–362.
- [13] PRIEUX, V., BROSSIER, R., OPERTO, S., AND VIRIEUX, J. 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* 194, 3 (2013), 1640–1664.
- [14] TARANTOLA, A. Inversion of seismic reflection data in the acoustic approximation. *Geophysics* 49, 8 (1984), 1259–1266.
- [15] TARANTOLA, A. Strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics* 51 (1986), 1893–1903.
- [16] VIRIEUX, J., AND OPERTO, S. An overview of full-waveform inversion in exploration geophysics. *Geophysics* 74, 6 (2009), WCC1–WCC26.
- [17] WU, R., AND AKI, K. Scattering characteristics of elastic waves by an elastic heterogeneity. *Geophysics* 50, 4 (1985), 582–595.

A Error-bowls for Scatterers (iv) and (v)

Error-bowl analysis is performed by determining and inspecting the logarithmic contours of the least-squares misfit as a function of the two subsurface parameters. Reconstruction results of the point-shaped scatterer (iv) with ρ -only contrast adopting the modelling and inversion schemes I, II and II are given in Figures 17, 18 and 19, respectively. It can be observed that the \mathbf{m}_{c_p, Z_p} -, $\mathbf{m}_{K, \rho}$ - and $\mathbf{m}_{1/K, 1/\rho}$ -parameterizations have faster convergence than the others, irrespective of the scheme adopted. Point-scatterer analysis suggests that the $\mathbf{m}_{1/K, 1/\rho}$ and $\mathbf{m}_{K, \rho}$ parameterization choices are equivalent and have the fastest convergence for the chosen circular acquisition geometry. It can be seen that these suggestions are valid in this case.

Reconstruction results of the sub-wavelength Gaussian-shaped scatterer (v) with ρ -only contrast adopting the modelling and inversion schemes I, II and II are given in Figures 20, 21 and 22, respectively. In the case of schemes II and III, the error bowls have different ellipticities and orientations compared to scheme I. Here, the $\mathbf{m}_{1/K, 1/\rho}$ -parameterization has the fastest convergence and $\mathbf{m}_{c_p, \rho}$ - and $\mathbf{m}_{Z_p, \rho}$ - parameterizations have a slower convergence. We observe that the suggestions of the point-scatterer analysis are only valid when scheme I is adopted.

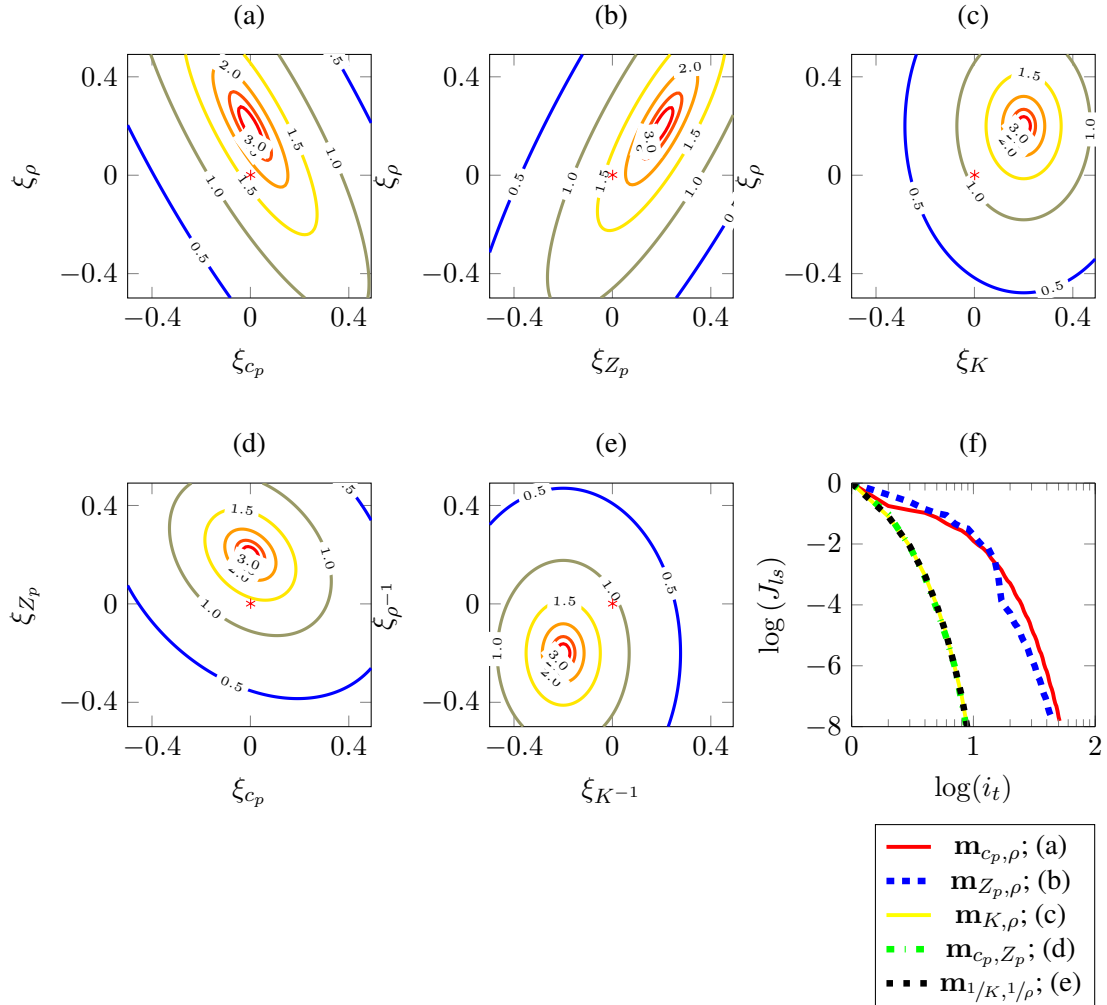


Figure 17: Reconstruction of scatterer (iv) using scheme I and the following parameterization choices: a) $\mathbf{m}_{c_p, \rho}$ — slower convergence expected because of high ellipticity; b) $\mathbf{m}_{Z_p, \rho}$ — slower convergence expected because of high ellipticity; c) $\mathbf{m}_{K, \rho}$ — faster convergence expected because of circular contours; d) \mathbf{m}_{c_p, Z_p} — faster convergence expected because of almost circular contours; e) $\mathbf{m}_{1/K, 1/\rho}$ — faster convergence expected because of circular contours. In all the plots, the starting homogeneous model, $(0, 0)^T$, is marked by the red star. f) The least-squares misfit is plotted as a function of the iteration count on a log-log scale. This plot shows that the suggestions of the point-scatterer analysis are valid in this case.

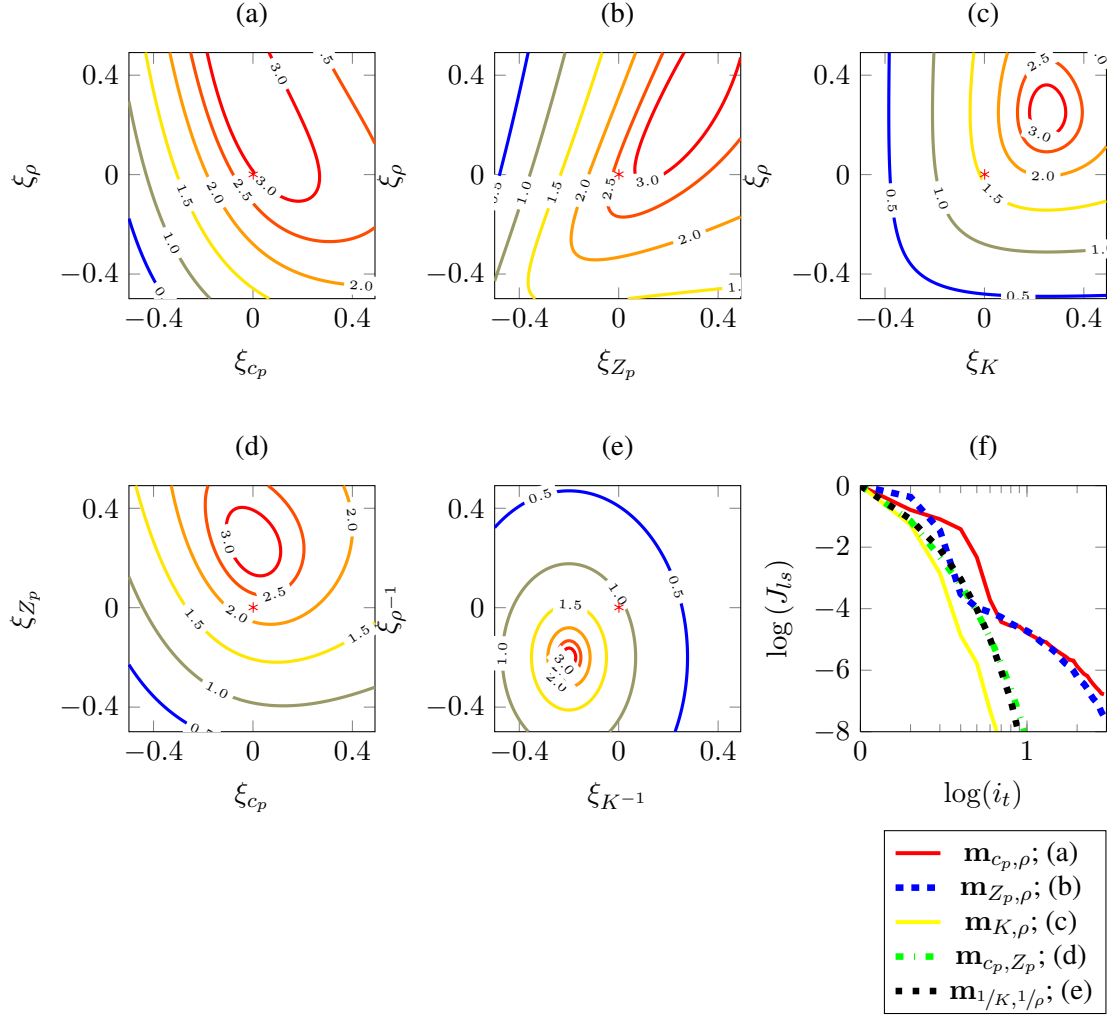


Figure 18: Same as Figure 17, except for adopting scheme II, with the Born approximation and non-linear re-parameterization. The least-squares misfit plot, similar to that of Figure 8, shows that the suggestions of the point-scatterer analysis are valid.

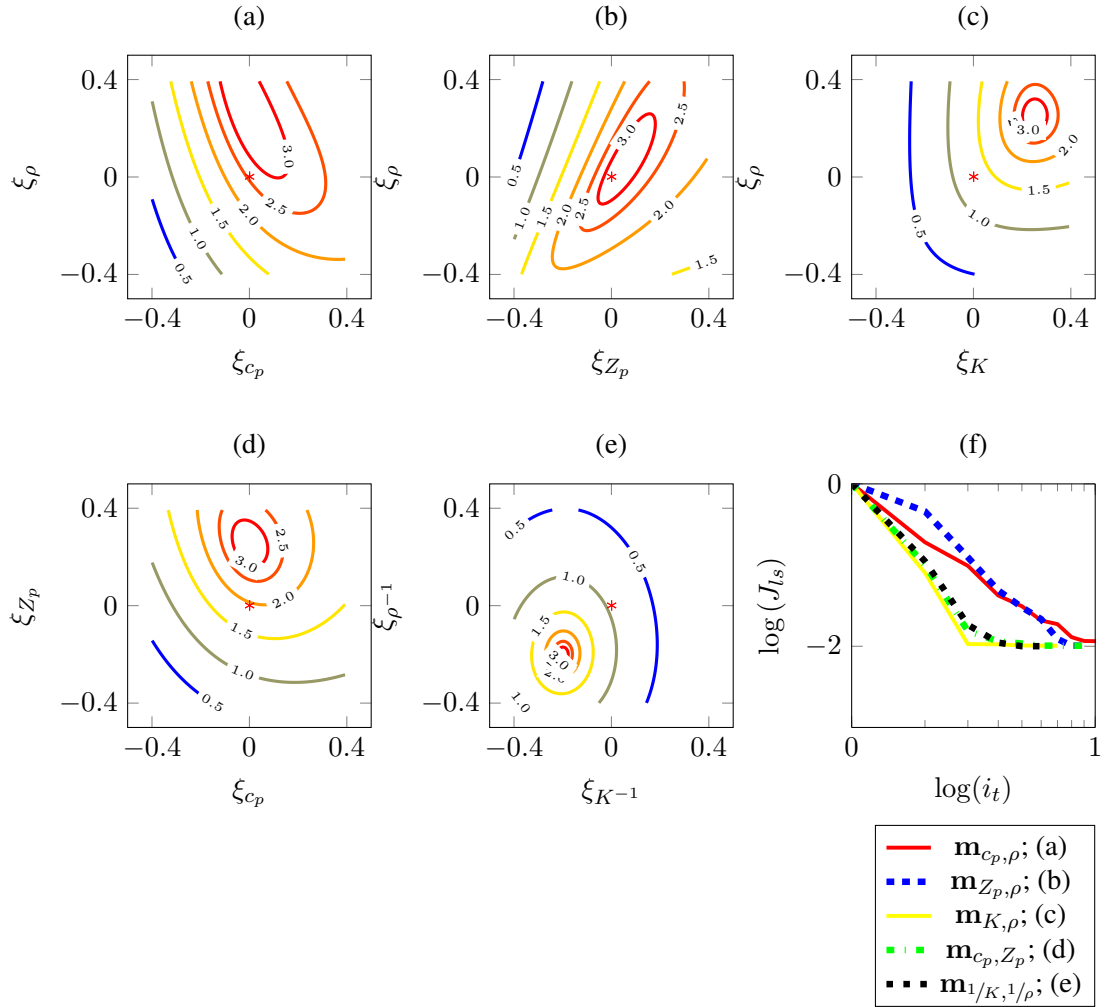


Figure 19: Same as Figure 17, except for adopting scheme III, with full-waveform modelling and inversion. It can be seen that the shapes of the error bowls and the relative convergence rates of different parameterization choices are similar to that of Figure 18. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are valid.

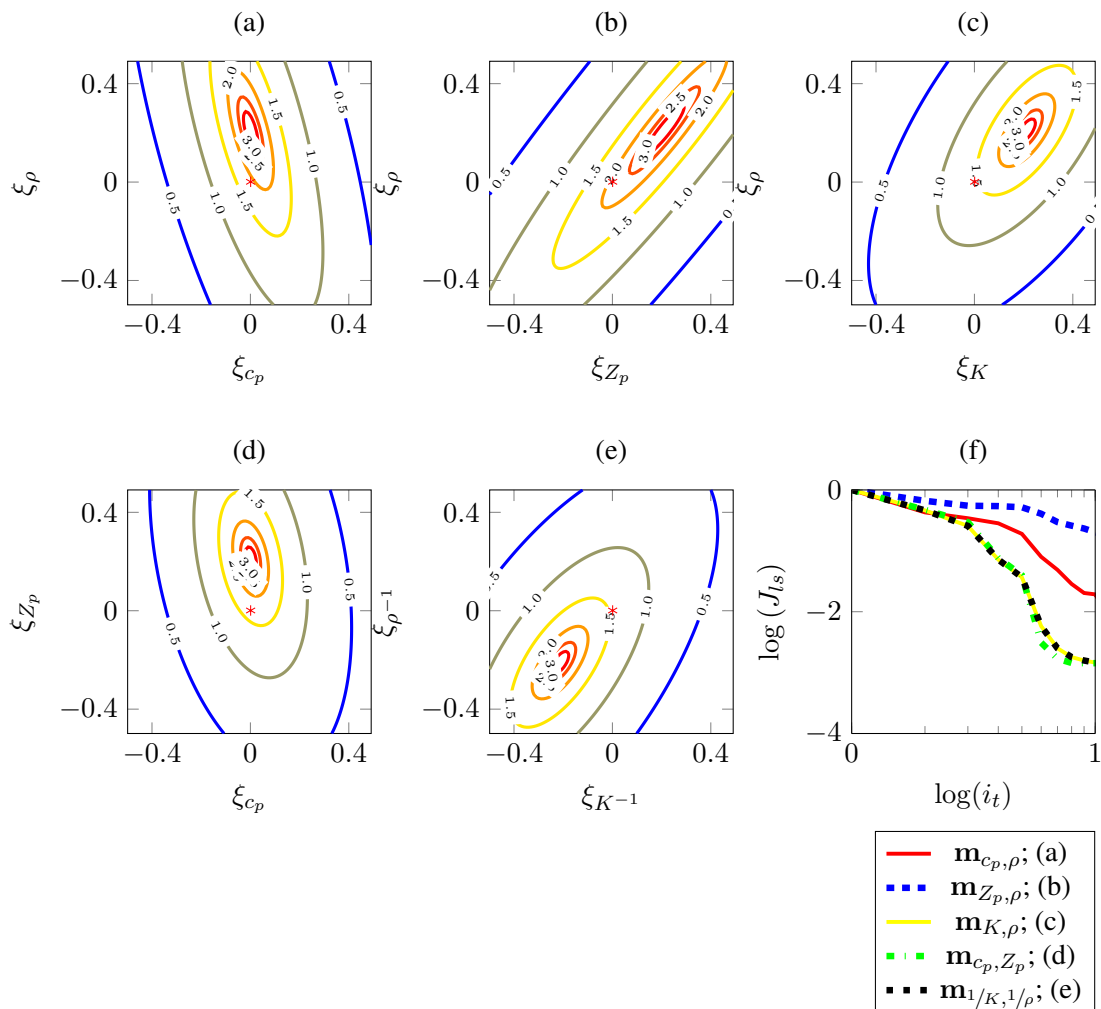


Figure 20: Same as Figure 17, except for the Gaussian-shaped ρ -only scatterer (v). We observe that the errors bowls for all the parameterization choices are more elliptical compared to that of Figure 17. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are valid. The rate of convergence in the case of \mathbf{m}_{c_p, Z_p} , $\mathbf{m}_{K, \rho}$ and $\mathbf{m}_{1/K, 1/\rho}$ -parameterizations is higher as their corresponding error bowls are more circular compared to the others.

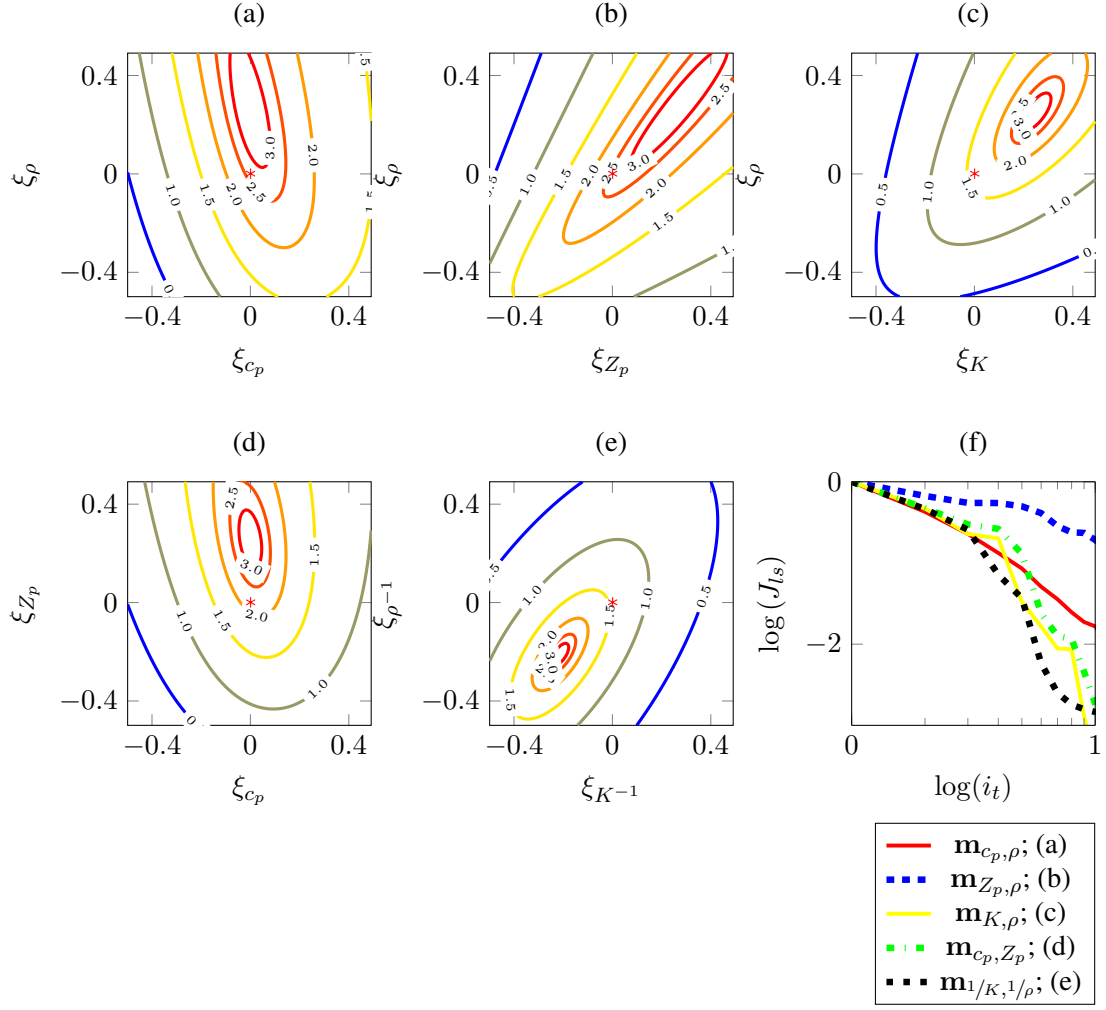


Figure 21: Same as Figure 17, except for scatterer (v) and adopting scheme II, with the Born approximation and non-linear re-parameterization.

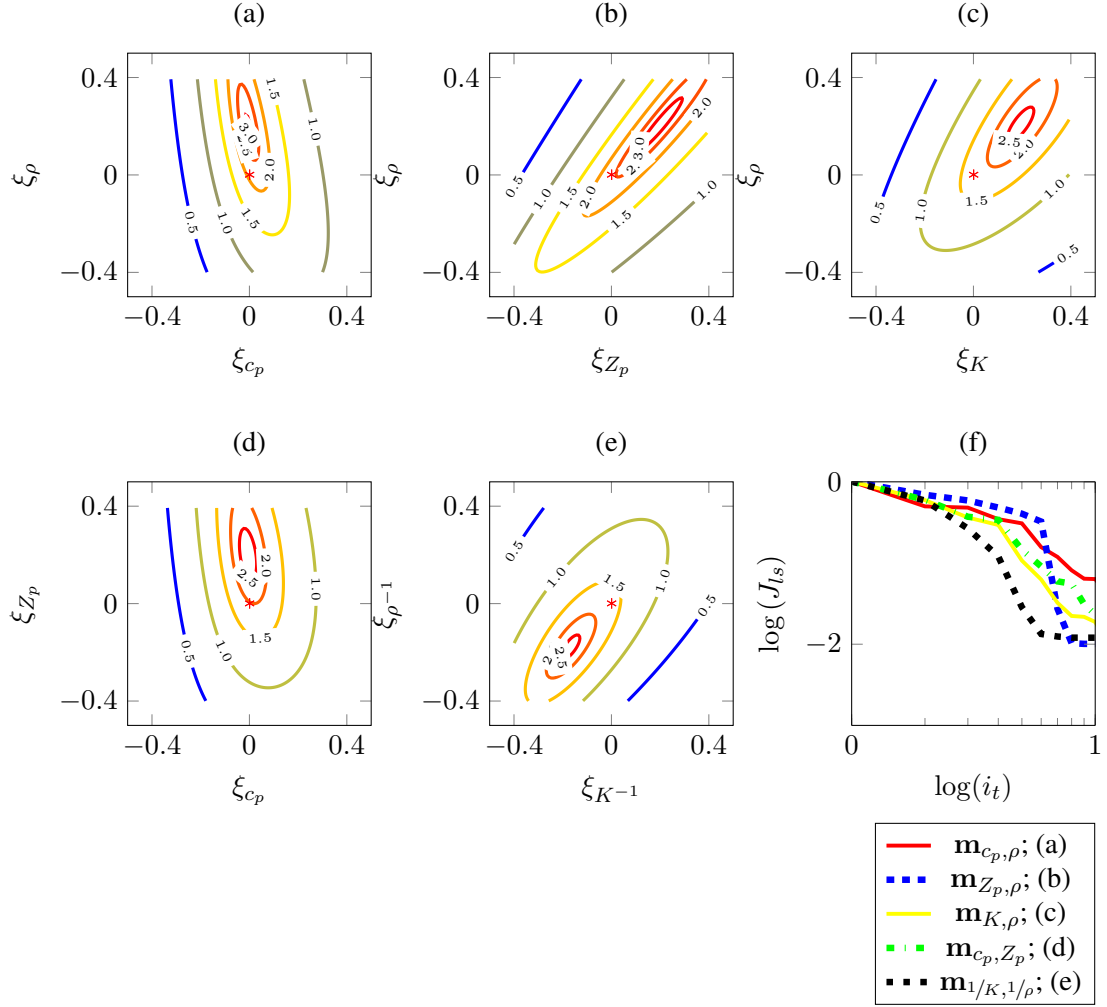


Figure 22: Same as Figure 17, except for scatterer (v) and adopting scheme III, with full-waveform modelling and inversion. Parameterization using $\mathbf{m}_{1/K, 1/\rho}$ has the best convergence rate.