

Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures

L. Métivier^{1,2*}, F. Bretaudeau², R. Brossier², S. Operto³ and J. Virieux²

¹LJK, CNRS, Université de Grenoble I, BP 53, 38041 Grenoble cedex 09, France, ²ISTerre, Université de Grenoble I, BP 53, 38041 Grenoble Cedex 09 France, and ³Géoazur, Université de Nice Sophia Antipolis, CNRS, IRD, OCA, Villefranche-sur-mer, France

Received August 2013, revision accepted October 2013

ABSTRACT

Full waveform inversion is a powerful tool for quantitative seismic imaging from wide-azimuth seismic data. The method is based on the minimization of the misfit between observed and simulated data. This amounts to the solution of a large-scale nonlinear minimization problem. The inverse Hessian operator plays a crucial role in this reconstruction process. Accounting accurately for the effect of this operator within the minimization scheme should correct for illumination deficits, restore the amplitude of the subsurface parameters, and help to remove artefacts generated by energetic multiple reflections. Conventional minimization methods (nonlinear conjugate gradient, quasi-Newton methods) only roughly approximate the effect of this operator. In this study, we are interested in the truncated Newton minimization method. These methods are based on the computation of the model update through a matrix-free conjugate gradient solution of the Newton linear system. We present a feasible implementation of this method for the full waveform inversion problem, based on a second-order adjoint state formulation for the computation of Hessian-vector products. We compare this method with conventional methods within the context of 2D acoustic frequency full waveform inversion for the reconstruction of P-wave velocity models. Two test cases are investigated. The first is the synthetic BP 2004 model, representative of the Gulf of Mexico geology with high velocity contrasts associated with the presence of salt structures. The second is a 2D real data-set from the Valhall oil field in North sea. Although, from a computational cost point of view, the truncated Newton method appears to be more expensive than conventional optimization algorithms, the results emphasize its increased robustness. A better reconstruction of the P-wave velocity model is provided when energetic multiple reflections make it difficult to interpret the seismic data. A better trade-off between regularization and resolution is obtained when noise contamination of the data requires one to regularize the solution of the inverse problem.

Key words: computing aspects, full waveform, imaging, numerical study, theory.

INTRODUCTION

Full waveform inversion (FWI) is a powerful seismic imaging tool, dedicated to quantitative estimations of subsurface parameters, such as P-wave and S-wave velocities, density,

impedance or anisotropy. The method is based on the minimization of a misfit function that measures the difference between recorded seismic data and predicted data computed through the numerical simulation of wave propagation. An initial subsurface model is iteratively updated to produce the final estimation.

*E-mail: ludovic.metivier@ujf-grenoble.fr

The formalism of the FWI method has been introduced by Lailly (1983) and Tarantola (1984), based on a time domain discretization of the wave equation. Its first application to 2D synthetic data in the acoustic approximation was performed by Gauthier *et al.* (1986). Later on, a hierarchical frequency domain approach has been introduced by Pratt for cross-hole tomography (Pratt 1990; Pratt and Worthington 1990). During the past ten years, the simultaneous advances in acquisition systems (development of wide-azimuth seismic surveys, for instance) and high-performance computing facilities have made possible the successful application of FWI to surface data, in both the 2D acoustic and the 2D elastic approximations, to reconstruct one or several parameters (Operto *et al.* 2004, 2005; Ravaut *et al.* 2004; Gao *et al.* 2006; Brossier *et al.* 2009; Prieux *et al.* 2011; Plessix *et al.* 2012; Prieux *et al.* 2013a,b). Applications of FWI to real surface data in the 3D acoustic approximation have also been performed (Sirgue *et al.* 2008; Plessix and Perkins 2010; Vigh *et al.* 2010). For an overview of the FWI methodology and its applications to synthetic and real case studies, the reader is referred to the review of Virieux and Operto (2009).

The simplest optimization methods used in the context of FWI are gradient-based algorithms, such as the nonlinear conjugate gradient algorithm. From a given initial model, the sequence of updates yielding the final model is defined through evaluations of the gradient of the misfit function.

To improve the convergence of these methods and the quality of the model estimation, the information carried out by the inverse Hessian operator (the matrix of second-order derivatives) is crucial. The importance of this operator in the context of FWI has been emphasized by Pratt *et al.* (1998). The inverse Hessian operator acts as a deconvolution operator that accounts for the limited bandwidth of the seismic data and corrects for the loss of amplitude of poorly illuminated subsurface parameters. In addition, it helps to remove artefacts that the second-order reflected waves may generate on the model update. Accounting for this operator leads to the use of Newton-based methods.

However, because of the large-scale aspect of the FWI problem, which easily involves millions of discrete unknowns in 2D and up to billions of discrete unknowns in 3D, explicit computation of the inverse Hessian operator is beyond current computational capabilities. As a consequence, research efforts have been mainly directed toward direct approximation of this operator.

A first possibility involves approximating the diagonal of the Hessian. For instance, Operto *et al.* (2006) compute the diagonal terms of the Gauss–Newton approximation of the

Hessian, which requires some extra computation. A cheaper strategy, based on the so-called pseudo-Hessian operator, is also proposed by Shin *et al.* (2001).

A second possibility involves approximating the inverse Hessian operator using previous values of the gradient of the misfit function. Among this class of methods, known as quasi-Newton methods, the *l*-BFGS method is quite popular (Byrd *et al.* 1995; Nocedal and Wright 2006). Instead of approximating only the diagonal elements, a positive definite approximation of the full inverse Hessian is computed.

The approximation of the diagonal elements of the Hessian and the *l*-BFGS strategy can be combined to produce a more accurate approximation of the inverse Hessian operator. Indeed, the accuracy of the *l*-BFGS approximation of the inverse Hessian operator can be improved when based on a first estimation of the inverse Hessian operator. This method has been applied in the framework of 2D elastic FWI: Brossier *et al.* (2009) implemented an *l*-BFGS optimization using a diagonal pseudo-Hessian as the initial guess. The diagonal estimation can even be updated alongside the iterations (Nocedal and Wright 2006). More recently, Ma and Hale (2012) have proposed the use of a complete BFGS approximation, through a sparse description of the subsurface model on an adapted basis. The *l*-BFGS method shows good convergence properties compared with preconditioned non-linear conjugate-gradient.

The truncated Newton method represents an alternative to these already described optimization methods. At each iteration, the model update is computed as an approximate solution of the Newton equations through a linear iterative solver (namely a conjugate gradient solver) (Nash 2000). Implemented in a “matrix-free” fashion, this iterative solver only requires one to compute Hessian-vector products. It is not necessary to form the Hessian operator explicitly.

Although this class of methods is well known in the numerical optimization community, the application of truncated Newton method in the FWI context has still not been fully investigated. Given the importance of the inverse Hessian operator in the FWI reconstruction scheme, we believe that this method could benefit from a better approximation of the inverse Hessian effect, and provides more accurate subsurface parameter estimations than standard optimization schemes. Therefore, the aim of this study is to focus on the following two points:

- To design a feasible implementation of the truncated Newton method for FWI in terms of computational time;
- To compare the performance of this method with conventional methods (nonlinear conjugate gradient, *l*-BFGS) on two realistic 2D test cases.

In Section 2, we describe in more detail the principle of the truncated Newton method compared with preconditioned gradient-based methods. In Section 3, we explain how the method can be implemented efficiently in the FWI context. In Section 4, we present two application cases. The first case is based on the synthetic BP 2004 model, partly inspired from the deep water Gulf of Mexico geology. The presence of salt structures in a marine environment is responsible for high velocity contrasts, which make the seismic imaging task difficult. The second test case concerns a 2D line of a real ocean bottom cable (OBC) data-set, acquired in a shallow-water environment at the Valhall oil field, in the North Sea. This test case is investigated to emphasize how the truncated Newton method behaves when the data is noise-contaminated. A discussion and concluding remarks are given in Section 5.

THE TRUNCATED NEWTON SCHEME

For the sake of clarity, the mathematical results presented in this section are formulated for a number of sources equal to 1. The extension to a multi-source context is straightforward, as only explicit summations over the sources are required.

Problem Settings

We consider the frequency-domain forward problem

$$A(m)u = s, \quad (1)$$

where

- $m \in \mathcal{M}$ denotes the subsurface model;
- $u \in \mathcal{W}$ is the complex-valued seismic wavefield;
- s is a source term;
- $A(m)$ is a discretized partial differential operator related to the wave equation (from the acoustic dynamics to the viscoelastic anisotropic dynamics).

The FWI problem is defined as the minimization over the parameter space of a distance between the data predicted by the forward problem and the recorded data.

$$\min_{m \in \mathcal{M}} f(m) = \frac{1}{2} \|Ru(m) - d\|^2, \quad (2)$$

where

- $u(m)$ is the solution of the forward problem (1) for the source term and the subsurface parameter s and m ;
- R is a mapping of the wavefield to the receivers' locations;
- d is the data set associated with the source s ;
- $\|\cdot\|$ is a norm in the data space \mathcal{D} .

For practical reasons, the use of the L^2 norm is common. However, the more general L^p norm could also be selected (Tarantola 2005). The L^1 norm is, for instance, a good choice when high-amplitude noise (outliers) corrupts the data (Brossier *et al.* 2010). More complex measurements of the distance between data sets can also be proposed to mitigate the sensitivity of FWI to the initial model. This is, however, beyond the scope of the work presented here.

Preconditioned Gradient-Based Methods

From a numerical point of view, FWI is a large-scale nonlinear minimization problem. The high number of discrete parameters prevents us from using global or semi-global optimization techniques to solve this problem. Therefore, we focus on local optimization methods, which are based on the following recurrence: from an initial guess m_0 , a sequence m_k is computed, such that

$$m_{k+1} = m_k + \gamma_k \Delta m_k, \quad (3)$$

where Δm_k is the model update and γ_k is a scalar parameter computed through a linesearch or a trust-region procedure (Bonnans *et al.* 2006; Nocedal and Wright 2006).

Within the framework of Newton algorithms, the increment Δm_k is defined as

$$H(m_k) \Delta m_k = -\nabla f(m_k), \quad (4)$$

where $H(m)$ denotes the Hessian operator and $\nabla f(m)$ is the gradient of the misfit function with respect to the model parameter m .

However, the exact solution of this linear system at each iteration is beyond current computational capacities in the context of FWI. This is why many large-scale optimization schemes rely on an approximation of the inverse Hessian operator. We recall here the principle of two of these methods: the preconditioned nonlinear conjugate gradient and the l -BFGS method.

The preconditioned nonlinear conjugate gradient algorithm is based on the computation of Δm_k as

$$\Delta m_k = -P_k \nabla f(m_k) + \Delta m_{k-1}, \quad (5)$$

where P_k is a preconditioner approximating $H(m_k)^{-1}$.

The l -BFGS algorithm is based on the computation of Δm_k as

$$\Delta m_k = -Q_k \nabla f(m_k), \quad (6)$$

where Q_k is the l -BFGS approximation of the inverse Hessian operator $H(m_k)^{-1}$. This approximation is built through an initial estimation Q_0 and successive rank-2 updates.

We shall refer to these schemes as preconditioned gradient-based methods. The nonlinear conjugate gradient uses a “direct” approximation of the inverse Hessian, for instance an approximation of its diagonal (Shin *et al.* 2001; Operto *et al.* 2006). The l -BFGS method (Byrd *et al.* 1995) uses l former values of the gradient to estimate the inverse Hessian operator. As mentioned in the introduction, the preconditioning matrix P_k used in the context of the nonlinear conjugate gradient can be incorporated in the l -BFGS algorithm as the initial estimation Q_0 (Nocedal and Wright 2006).

Truncated Newton algorithm

Instead of building an approximation P_k , the linear system (4) can be solved using a matrix-free version of the conjugate gradient algorithm (Saad 2003). This only requires the capability of computing Hessian-vector products $H(m_k)v$ where v is an arbitrary vector in the model space \mathcal{M} . The truncated Newton method thus results in a two-nested-loops algorithm:

- The external loop consists of the iterative update of the current subsurface parameter estimation, following (3);
- The internal loop consists of the iterative solution of the linear system (4), in order to compute the model update Δm_k .

The advantage of using a truncated Newton method is two-fold:

- The approximation of the inverse Hessian operator depends only on the current model estimation, while the approximation computed using the l -BFGS or nonlinear conjugate gradient methods depends on models estimated at previous iterations. When minimizing strongly nonlinear misfit functions, this can be advantageous, since the information on the local curvature carried out by previous iterations can be erroneous. Setting the memory parameter l is thus known to be difficult, as it is strongly problem dependent (Nocedal and Wright 2006).
- The truncation strategy in the inner loop involves accounting only for the higher eigenvalues of the inverse Hessian operator. This has an intrinsic regularization effect on the computation of the model update, as demonstrated by Kaltenbacher *et al.* (2008).

In addition, the truncated Newton method offers the possibility of using the approximations of the inverse Hessian operator which have been developed in the context of preconditioned gradient-based methods as preconditioners of the in-

ner linear systems, yielding the solution of the preconditioned linear system

$$P_k H(m_k) \Delta m_k = -P_k \nabla f(m_k). \quad (7)$$

However, compared with preconditioned gradient-based methods, the truncated Newton method involves an additional computational expense associated with the iterative solution of the linear system (4). This should be balanced by an improvement of the convergence speed of the external loop. Therefore, as mentioned by Nash (2000), an efficient implementation of the truncated Newton method relies on the reduction of this additional cost. This can be achieved in the context of FWI by:

- Defining second-order adjoint formulae for the efficient computation of Hessian vector products $H(m_k)v$ for any $v \in \mathcal{M}$;
- Defining an adapted stopping criterion for the approximate solution of the linear system (4) to limit as much as possible the number of iterations of the conjugate gradient algorithm required to compute the descent direction;
- Using an appropriate preconditioner to accelerate the convergence of the solution of the linear system (4).

The truncated Newton strategy can also be implemented using the Gauss–Newton approximation $B(m)$ of the Hessian operator $H(m)$, which consists in neglecting the second-order terms of the Hessian operator. The Gauss–Newton approximation is

$$B(m) = J(m)^\dagger R^\dagger R J(m), \quad (8)$$

where $J(m)$ is the Jacobian matrix:

$$J(m) = \frac{\partial u(m)}{\partial m}, \quad (9)$$

and † denotes the transpose conjugate operator.

The Gauss–Newton method is appealing for the following reasons:

- $B(m)$ is positive definite by construction, therefore the conjugate gradient algorithm is well adapted for the solution of the linear system (4);
- Close to the solution, the residuals are small, therefore the matrix $B(m)$ should be a good approximation of the full Hessian matrix $H(m)$, since the second-order part that is neglected in the Gauss–Newton approximation is proportional to the residuals.

From the implementation point of view, the truncated Gauss–Newton procedure differs only from the truncated Newton method in the computation of matrix-vector products $B(m)v$ instead of $H(m)v$. In the next section, we investigate

how the truncated Newton and Gauss–Newton methods can be efficiently implemented in the context of FWI.

IMPLEMENTATION OF THE TRUNCATED NEWTON METHOD

Computation of the model update

The solution of the linear systems (4) first requires computation of the right-hand side, $-\nabla f(m_k)$. The computation of $\nabla f(m_k)$ is efficiently achieved through the first-order adjoint-state method, introduced by Lions (1968). For the sake of generality, we will use here the notation introduced by Plessix (2006) in his review of the adjoint-state technique for seismic imaging. In this framework, the forward problem should be rewritten as

$$F(m, u) = 0, \quad (10)$$

where

$$F(m, u) = A(m)u - s. \quad (11)$$

Computation of the gradient and first-order adjoint method

In the context of FWI, the first-order adjoint method amounts to computing $\nabla f(m_k)$ as the zero-lag cross-correlation of the incident wavefield and the adjoint wavefield (Chavent 1974). The adjoint wavefield is defined as the solution of

$$\frac{\partial F(m, u)^\dagger}{\partial u} \lambda = R^\dagger(d - Ru), \quad (12)$$

where $u(m)$ is the solution of (1). Based on the definition of the adjoint state λ , the gradient can be expressed as

$$\nabla f(m) = \mathbf{Re} \left(\frac{\partial F(m, u)^\dagger}{\partial m} \lambda \right), \quad (13)$$

where \mathbf{Re} denotes the real part operator.

Using this method, the computation cost of the gradient amounts to the solution of two wave propagation problems per shot: one forward problem (1) and one adjoint problem (12).

Computation of $H(m)v$ through the second-order adjoint method

The computation of Hessian-vector products through second-order adjoint methods is a topic that has already been investigated in the field of data assimilation and weather forecasting (Wang *et al.* 1992). However, the control variable in data assimilation is an initial condition for the system, whereas in seismic imaging, the control variable is a coefficient of

the partial differential equation that describes the system. A formula for the computation of Hessian-vector products has been given by Pratt *et al.* (1998) in the seismic imaging context, for the Gauss–Newton approximation in the discrete frequency domain. Fichtner and Trampert (2011) also propose more general formulae for the computation of Hessian kernels. Epanomeritakis *et al.* (2008) give the formulae corresponding to the elastic case.

We propose here a general framework in the frequency domain for deriving these formulae, with no assumption on the discretization and the kind of partial differential equations that are used for the wave propagation description. The method can be straightforwardly adapted to the time-domain formulation by adding proper initial and final conditions, and boundary conditions. In addition, no prior assumption on the linearity of the forward problem is required, as was the case in Pratt *et al.* (1998) and a previous work (Métivier *et al.* 2013).

We first define the functional $h_v(m)$ as

$$h_v(m) = (\nabla f(m), v)_{\mathcal{M}}, \quad (14)$$

where $(\cdot, \cdot)_{\mathcal{M}}$ denotes the scalar product on the parameter space \mathcal{M} . By definition, the gradient of the functional $h_v(m)$ is

$$\nabla h_v(m) = H(m)v. \quad (15)$$

We use the Lagrangian formalism to compute $\nabla h_v(m)$. We introduce the Lagrangian operator $L(m, u, \lambda, g, \mu_1, \mu_2, \mu_3)$ such that

$$\begin{aligned} L_v(m, u, \lambda, g, \mu_1, \mu_2, \mu_3)_{\mathcal{W}} &= (g, v)_{\mathcal{M}} \\ &+ \mathbf{Re} \left(g - \frac{\partial F(m, u)^\dagger}{\partial m} \lambda, \mu_1 \right)_{\mathcal{W}} \\ &+ \mathbf{Re} \left(\frac{\partial F(m, u)^\dagger}{\partial u} \lambda - R^\dagger(d - Ru), \mu_2 \right)_{\mathcal{W}} \\ &+ \mathbf{Re} (F(m, u), \mu_3)_{\mathcal{W}}. \end{aligned} \quad (16)$$

In this expression, $(\cdot, \cdot)_{\mathcal{W}}$ denotes the scalar product in the wavefield space \mathcal{W} . Let $\bar{u}(m), \bar{\lambda}(m), \bar{g}(m)$ satisfy the constraints

$$\begin{aligned} F(m, \bar{u}) &= 0, \quad \frac{\partial F(m, \bar{u})^\dagger}{\partial u} \bar{\lambda} = R^\dagger(d - R\bar{u}) \quad \text{and} \\ \bar{g} &= \frac{\partial F(m, \bar{u})^\dagger}{\partial m} \bar{\lambda}. \end{aligned} \quad (17)$$

We have

$$L_v(m, \bar{u}, \bar{\lambda}, \bar{g}, \mu_1, \mu_2, \mu_3) = h_v(m), \quad (18)$$

and

$$\frac{\partial L_v}{\partial m}(m, \bar{u}, \bar{\lambda}, \bar{g}, \mu_1, \mu_2, \mu_3) = \nabla h_v(m). \quad (19)$$

In addition,

$$\begin{aligned} \frac{\partial L_v}{\partial m}(m, \bar{u}, \bar{\lambda}, \bar{g}, \mu_1, \mu_2, \mu_3) &= -\operatorname{Re} \left(\left(\frac{\partial^2 F(m, u)^\dagger}{\partial m^2} \lambda \right)^\dagger \mu_1 \right) \\ &+ \operatorname{Re} \left(\left(\frac{\partial^2 F(m, u)^\dagger}{\partial m \partial u} \lambda \right)^\dagger \mu_2 \right) + \operatorname{Re} \left(\frac{\partial F(m, u)^\dagger}{\partial m} \mu_3 \right) \\ &+ \frac{\partial L_v}{\partial g}(\bar{m}, \bar{u}, \bar{g}, \mu_1, \mu_2, \mu_3) \frac{\partial \bar{g}(m)}{\partial m} \\ &+ \frac{\partial L_v}{\partial \lambda}(\bar{m}, \bar{u}, \bar{g}, \mu_1, \mu_2, \mu_3) \frac{\partial \bar{\lambda}(m)}{\partial m} \\ &+ \frac{\partial L_v}{\partial u}(\bar{m}, \bar{u}, \bar{g}, \mu_1, \mu_2, \mu_3) \frac{\partial \bar{u}(m)}{\partial m}. \end{aligned} \quad (20)$$

We can choose $\mu_i, i = 1, 2, 3$ such that for any perturbations $dg, d\lambda, du$ we have

$$\begin{cases} \frac{\partial L_v}{\partial g}(\bar{m}, \bar{u}, \bar{g}, \mu_1, \mu_2, \mu_3) dg = 0 \\ \frac{\partial L_v}{\partial \lambda}(\bar{m}, \bar{u}, \bar{g}, \mu_1, \mu_2, \mu_3) d\lambda = 0 \\ \frac{\partial L_v}{\partial u}(\bar{m}, \bar{u}, \bar{g}, \mu_1, \mu_2, \mu_3) du = 0. \end{cases} \quad (21)$$

This yields

$$\begin{cases} \mu_1 = -v \\ \frac{\partial F(m, u)}{\partial u} \mu_2 = \frac{\partial F(m, u)}{\partial m} \mu_1 \\ \frac{\partial F(m, u)^\dagger}{\partial u} \mu_3 = \left(\frac{\partial^2 F(m, u)^\dagger}{\partial u^2} \lambda \right)^\dagger \mu_2 \\ \quad - \left(\frac{\partial^2 F(m, u)^\dagger}{\partial u \partial m} \lambda \right)^\dagger \mu_1 - R^\dagger R \mu_2. \end{cases} \quad (22)$$

Using the relation (11) which defines our forward problem, we see that

$$\frac{\partial F(m, u)}{\partial u} = A(m), \quad \frac{\partial^2 F(m, u)}{\partial u^2} = 0. \quad (23)$$

This yields

$$\begin{cases} A(m) \mu_2 = -\frac{\partial F(m, u)}{\partial m} v \\ A(m)^\dagger \mu_3 = -\left(\frac{\partial^2 F(m, u)^\dagger}{\partial u \partial m} \lambda \right)^\dagger \mu_1 - R^\dagger R \mu_2. \end{cases} \quad (24)$$

We thus obtain the three-term Hessian-vector product formula

$$\begin{aligned} H(m)v &= -\operatorname{Re} \left(\left(\frac{\partial^2 F(m, u)^\dagger}{\partial m^2} \lambda \right)^\dagger \mu_1 \right) \\ &+ \operatorname{Re} \left(\left(\frac{\partial^2 F(m, u)^\dagger}{\partial u \partial m} \lambda \right)^\dagger \mu_2 \right) \\ &+ \operatorname{Re} \left(\frac{\partial F(m, u)^\dagger}{\partial m} \mu_3 \right). \end{aligned} \quad (25)$$

Note that the term

$$\frac{\partial^2 F(m, u)}{\partial u \partial m} = \frac{\partial A}{\partial m} \quad (26)$$

corresponds to the radiation pattern of the virtual secondary source, which emits the partial derivative wavefield from the position of the model parameter (Pratt *et al.* 1998).

The adjoint wavefield $\bar{\mu}_1$ is equal to $-v$. The computation of one Hessian-vector products thus requires computation of the incident wavefield $\bar{u}(m)$ and the adjoint wavefield $\bar{\lambda}(m)$, as well as two additional wavefields, $\bar{\mu}_2(m)$ and $\bar{\mu}_3(m)$. The two latter wavefields are computed through the solution of one forward and one adjoint problem, respectively, with new source terms.

Gauss–Newton approximation

From the definition of the Gauss–Newton operator $B(m)$ (8), we see that only first-order derivatives of the wavefield $\bar{u}(m)$ with respect to the model parameter m (also called the Jacobian matrix or Fréchet derivatives) are taken into account. The computation of $B(m)v$ can thus be derived from (24) and (25) by neglecting the contribution of all the second-order terms. This yields the following simplifications:

$$\begin{cases} A(m) \mu_2 = -\frac{\partial F(m, u)}{\partial m} v \\ A(m)^\dagger \mu_3 = -R^\dagger R \mu_2, \end{cases} \quad (27)$$

and the following formula for the Gauss–Newton approximation

$$B(m)v = \operatorname{Re} \left(\frac{\partial F(m, u)^\dagger}{\partial m} \mu_3 \right). \quad (28)$$

This formula is consistent with the one derived in Pratt *et al.* (1998) and Métivier *et al.* (2013). Note, however, that in these two articles, the method used to derive this formula relies on the assumption of a linear forward problem, which is not the case here.

Computation cost

Consider the solution of the linear system (4) for the computation of the model update Δm_k . The right hand side in (4) is the opposite gradient $-\nabla f(m)$. Using the first-order adjoint state method, it can be computed at the expense of one forward problem for $\bar{u}(m)$ and one adjoint problem for $\bar{\lambda}(m)$.

Provided these wavefields can be stored, the computation of $H(m)v$ or $B(m)v$ only requires the solution of two additional problems: one forward problem for the

computation of $\bar{\mu}_2(m)$, and one adjoint problem for the computation of $\bar{\mu}_3(m)$.

Note that in practice, the computation of $B(m)v$ does not make use of the adjoint wavefield $\bar{\lambda}(m)$. However, its computation is imposed by the computation of the gradient, which is the right-hand side of the linear system (4). The computation cost of the action of the Hessian operator or its Gauss–Newton approximation on an arbitrary vector is thus *the same* in terms of number of wave equations to be solved. Nonetheless, only the incident wavefield $\bar{u}(m)$ has to be stored in the Gauss–Newton approximation.

The overall computation cost of the truncated Newton method in terms of wave propagation simulation is thus given by

$$C = \sum_{k=1}^{N_{\text{ext}}} (2 + 2 \times N_{\text{ls},k} + 2 \times N_{\text{CG},k}) \quad (29)$$

where N_{ext} is the total number of iterations of the external loop, $N_{\text{ls},k}$ is the number of linesearch iteration at iteration k and $N_{\text{CG},k}$ is the number of conjugate gradient iterations performed at iteration k . The choice of an appropriate stopping criterion for the conjugate gradient helps to reduce $N_{\text{int},k}$. Note that the computational cost estimation C depends on the ability of storing $\bar{u}(m)$ and $\bar{\lambda}(m)$. This is a reasonable assumption for 2D frequency-domain applications. It is a more questionable issue for 3D frequency-domain FWI, and time-domain FWI (2D and 3D); more sophisticated memory and I/O management methods should be required. In particular, the increase of the number of sources is critical, as one incident and one adjoint wavefield, $\bar{u}(m)$ and $\bar{\lambda}(m)$, have to be stored per source. Time-domain and 3D implementations will therefore require particular attention. Among different possibilities, we may consider

- The use of source-encoding strategies (Krebs *et al.* 2009);
- The re-computation of $\bar{u}(m)$ and $\bar{\lambda}(m)$;
- The estimation of $H(m)v$ through finite differences of the gradient.

Definition of an adapted stopping criterion for the inner loop

The Newton method is an iterative minimization of local quadratic expansions of the misfit function. Indeed, the solution of the system (4) amounts to the minimization of the quadratic form

$$q_k(\Delta m) = f(m_k) + (\nabla f(m_k), \Delta m) + \frac{1}{2}(H(m_k)\Delta m, \Delta m). \quad (30)$$

The definition of the stopping criterion for the truncated Newton method is related to the accuracy of these quadratic expansions. The idea explored by Eisenstat and Walker (1994) is as follows. Consider a stopping criterion for the CG iterations of the form

$$\|H(m_k)\Delta m_k + \nabla f(m_k)\| \leq \eta_k \|\nabla f(m_k)\|, \quad (31)$$

where η_k is the forcing term, which accounts for the accuracy of the local quadratic approximation. When this accuracy increases, η_k should decrease, so as to require the linear system (4) to be solved more accurately. Conversely, when the accuracy decreases, η_k should increase, allowing a less precise solution of the linear system (4). This is achieved by defining η_k as the measure of the distance between the first-order Taylor expansion of the gradient at the iteration $k-1$ and the gradient at iteration k :

$$\eta_k = \frac{\|\nabla f(m_k) - \nabla f(m_{k-1}) - \gamma_{k-1}H(m_{k-1})\Delta m_{k-1}\|}{\|\nabla f(m_{k-1})\|}. \quad (32)$$

The definition of the stopping criterion is complemented by an appropriate strategy for dealing with the detection of negative eigenvalues of the Hessian operator. The conjugate gradient algorithm is designed for the solution of symmetric positive definite systems. However, far from the solution, the full Hessian operator $H(m_k)$ may be indefinite. Therefore, the iterative construction of the solution of (4) in the Krylov space

$$\{r_0, H(m_k)r_0, H(m_k)^2r_0, \dots\}, \quad (33)$$

where r_0 is the initial residual

$$H(m_k)\Delta m_k^0 + \nabla f(m_k), \quad (34)$$

may use an ascent direction associated with a negative eigenvalue of the operator $H(m_k)$. In this case, the linear iterations are stopped and the last value of the model update Δm_k that is computed is returned. If this ascent direction is met at the very first linear iteration, the steepest-descent direction is returned. This strategy, proposed by Eisenstat and Walker (1994), ensures superlinear convergence properties far from the solution, and quadratic convergence when entering the attraction basin of the minimum.

Preconditioning

To speed up the convergence of the solution of the linear system (4), it is natural to introduce a preconditioning matrix. In this study, we focus on the special preconditioner related to the FWI problem proposed by Shin *et al.* (2001). The

diagonal elements of the Gauss–Newton part of the Hessian $B(m)$ are approximated using the pseudo-Hessian approach, which appears to be relevant for fixed-spread seismic survey. Let us denote $\alpha_j(m)$ the column j of the Jacobian matrix:

$$\alpha_j(m) = \frac{\partial u(m)}{\partial m_j}. \quad (35)$$

Deriving the forward problem with respect to the j th component of the model parameter m_j yields

$$A(m) \frac{\partial u}{\partial m_j} + A(m) \frac{\partial A}{\partial m_j} u = 0. \quad (36)$$

Thus, $\alpha_j(m)$ is the solution of the forward problem

$$A(m) \alpha_j = - \frac{\partial A(m)}{\partial m_j} u. \quad (37)$$

Using this formula, an exact computation of the entire Jacobian matrix $J(m)$ would thus require the solution of m forward problems, which is intractable from a computational-cost point of view. Nonetheless, a cheap approximation can be built by approximating the forward problem operator $A(m)$ as the identity matrix I in the left-hand side of (37). This leads to the definition of the pseudo-Hessian matrix entries:

$$\tilde{H}_{ij}(m) = \left(\left[\frac{\partial A(m)}{\partial m_i} u(m) \right]^T \left[\frac{\partial A(m)}{\partial m_j} u(m) \right] \right), \quad i, j = 1, \dots, M. \quad (38)$$

The preconditioner used by Shin *et al.* (2001) is defined by

$$P_k = \text{diag} \left(\frac{1}{\tilde{H}_{ii}(m_k)} \right) \quad i = 1, \dots, M. \quad (39)$$

However, because of the fast decrease of the wavefield with depth, very small values appear on the diagonal entries of $\tilde{H}(m)$, corresponding to deep subsurface parameters. Therefore, using P_k directly as a preconditioner may yield numerical instabilities. We thus introduce the constant $C_k \in \mathbb{R}$, such that

$$C_k = \max_j \tilde{H}_{jj}(m_k), \quad (40)$$

and a threshold parameter $\theta \in \mathbb{R}$ to define the matrix P_k^θ , such that

$$P_k^\theta = \text{diag} \left(\frac{1}{\tilde{H}_{ii}(m_k) + \theta C_k} \right), \quad i = 1, \dots, M. \quad (41)$$

Finally the norm of the misfit gradient $\nabla f(m)$ should be preserved by the preconditioner. Since the stopping criterion for the linear system (4) is based on the reduction of the linear

residuals with respect to the norm of the gradient, it appears natural that the preconditioner conserves the gradient norm. Therefore, we introduce

$$P_k^{v,\theta} = v P_k^\theta, \quad (42)$$

such that

$$v = \frac{\|\nabla f(m_k)\|}{\|P_k^\theta \nabla f(m_k)\|}. \quad (43)$$

We use $P_k^{v,\theta}$ as a preconditioner of the linear system (4) at each iteration of the external loop. The preconditioner $P_k^{v,\theta}$ can also be used as a preconditioner of the nonlinear conjugate gradient method and the l -BFGS method.

CASE STUDIES

Numerical framework

Forward problem

The numerical tests we present are performed in the 2D frequency domain. In the first case study, we use an isotropic acoustic approximation of the wave propagation. In the second case study, we consider the propagation of acoustic anisotropic waves in a VTI media. In both cases, an optimized second-order finite differences scheme with a compact stencil is used (Hustedt *et al.* 2004; Operto *et al.* 2009). Perfectly matched layers (PMLs) (Bérenger 1994; Métivier 2011) are introduced, to avoid fictitious reflections on the boundaries of the computation domain, except the top, where a free surface condition is implemented.

The numerical solution of the forward problem amounts to the solution of a sparse linear system. This is performed through a parallel LU factorization using the MUMPS algorithm (Amestoy *et al.* 2000). The LU factorization of the impedance matrix associated with the discretization of the forward problem is reused to solve the adjoint problems, as MUMPS offers the possibility of solving the adjoint system once the factorization has been performed. This is especially important when the number of sources is large: the same LU factorization is used to solve the forward and adjoint problems associated with each source. This interesting feature is one of the reasons for working in the frequency domain: provided the LU factorization of the impedance matrix can be stored, this approach significantly reduces the computational cost, compared with the time-domain approach. In the particular case of the truncated Newton method, this also enables use of the factorization of the impedance matrix for the computation of the Hessian-vector products. The solution of the extra

forward and adjoint problems induced by the truncated Newton method can thus be performed without a new factorization of the impedance matrix.

Minimization scheme settings

In the following two tests an estimation of the P-wave velocity model is computed using a FWI scheme. For each of these two tests, we compare the performances of four minimization strategies:

- Nonlinear conjugate gradient method;
- l -BFGS method;
- Truncated Newton method using the full Hessian operator;
- Truncated Newton method using the Gauss–Newton approximation.

In our implementation of the truncated Newton algorithm, we complement the Eisenstat stopping criterion by setting the maximum number of inner iterations that can be performed in the first case study to 30, and only 3 in the second case study. This rather small value is chosen to enhance the smoothing effect related to the truncation strategy, which is appropriate for the inversion of noisy data.

The stopping criterion we use is as follows: the iterations end as soon as the relative decrease of the misfit function becomes smaller than ε

$$f(m_k)/f(m_0) < \varepsilon. \quad (44)$$

The quantity ε is set to 10^{-2} for the first experiment. Such an accuracy can be reached since in this case we work with synthetic data without noise. In the second experiment, we use real noisy data, and ε is set to 6×10^{-1} , which is adapted to the expected decrease of the misfit function.

In addition, if no acceptable step length is found after 20 linesearch iterations, the minimization is stopped and an error flag is returned.

Finally, note that the memory parameter l for the two l -BFGS methods, which corresponds to the number of gradient stored to compute the approximation of the inverse Hessian, is set to $l = 20$ in the first case, which involves synthetic seismic data without noise, and reduced to $l = 3$ in the second case, owing to the presence of noise.

Linesearch algorithm

The linesearch algorithm we use is designed so that the step γ_k (in (4)) computed at each nonlinear iteration satisfies the

two Wolfe conditions (Nocedal and Wright 2006). The aim of these conditions is to ensure the convergence to the nearest local minimum from any initial guess m_0 . A bracketing strategy is implemented to compute γ_k . Each iteration of the linesearch procedure requires computation of the gradient of the misfit function; therefore, it is important to control this additional computational cost. In practice, we implement the following strategy: at first iteration, the step γ_0 is computed so that the norm of the first model update reaches a few hundred meters per second. Then, at each nonlinear iteration k , the first estimation for γ_k is the step γ_{k-1} computed at the previous nonlinear iteration.

For the four minimization methods we use here we have observed that the estimation of γ_0 may require several linesearch iterations (up to ten). Once γ_0 has been computed, at a given iteration k , the previous estimation γ_{k-1} satisfies the Wolfe conditions most of the time. Therefore, very few additional linesearch iterations are required, and the computational burden associated with the linesearch strategy is mainly reported at the first nonlinear iteration.

Nonlinear conjugate gradient implementation

Because we want the four minimization methods to use the same linesearch algorithm, we implement a particular form of the nonlinear conjugate gradient method. Indeed, as mentioned by Nocedal and Wright (2006), standard implementations of this algorithm (such as the Fletcher–Reeves or Polak–Ribière implementations) require us to use a linesearch algorithm satisfying the *strong* Wolfe conditions to guarantee global convergence toward local minima. We select instead the nonlinear conjugate gradient algorithm proposed by Dai and Yuan (1999), which is compatible with a linesearch process that only enforces the *standard* Wolfe conditions.

Preconditioner

To design the preconditioner $P_k^{v,\theta}$, the threshold parameter θ has to be set. Because we want to compare the minimization methods using the same experimental settings, we have chosen a common value for θ such that $\theta = 10^{-2}$. This value should be considered as an average value, yielding a preconditioner with good properties for the four minimization methods. Of course, it is possible to optimize this value for each of the minimization methods, but this would not produce drastic changes of the results presented in this study.

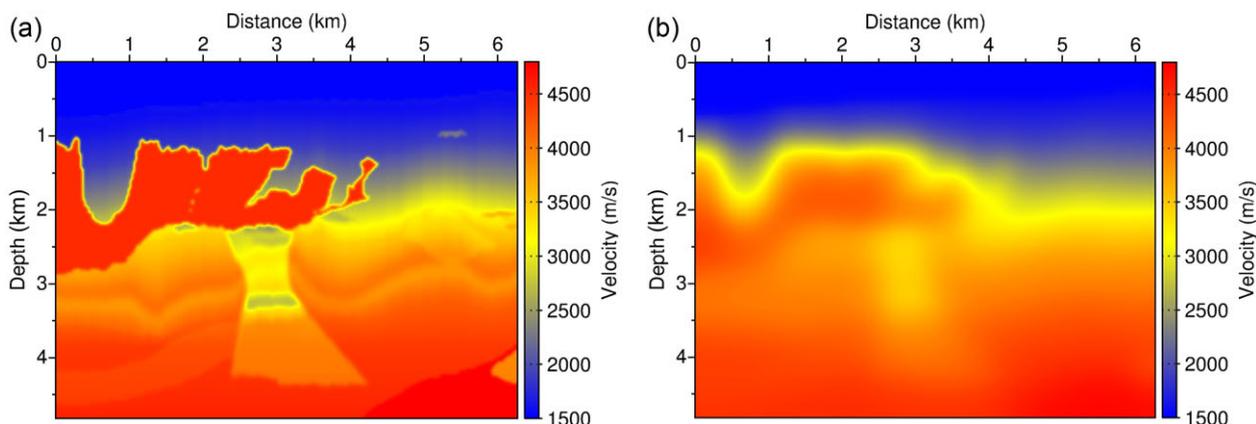


Figure 1 BP 2004 synthetic exact model (a), initial model (b).

The 2004 BP model

Presentation

The 2004 BP model has been originally designed as a benchmark model for testing sub-salt seismic imaging methods (Billlette and Brandsberg-Dahl 2004). We perform a decimation of the original model taking one parameter value for each ten grid points, and we choose a 25 m discretization grid. We end up with a reduced model 6.2 km wide and 4.8 km deep, described by approximately 5×10^4 discrete parameters.

The resulting model is presented in Figure 1a. It presents a complex rugose salt body, and sub-salt slow velocity anomalies that represent over-pressured zones. This tends to mimic the geology that can be found in the Gulf of Mexico. The main challenges in this area are related to the definition of a precise delineation of the salt and recovering information on the sub-salt velocity variations. The P-wave velocity in the salt reaches 4790 m.s^{-1} , while it is equal to 1486 m.s^{-1} in the water. The discrepancy between these two values is responsible for high-amplitude reflections. These energetic reflected waves are reflected back at the top of the water layer through the free-surface condition. The proximity of these two reflectors generates multiple scattering.

We use a surface acquisition configuration with 62 sources and 248 receivers, from $x = 50 \text{ m}$ to $x = 6225 \text{ m}$ at 25 m below the sea-level. The spatial sampling of the receivers and the sources is set up at 25 m and 100 m, respectively. The water layer is kept constant throughout the iterations, so as to stabilize the problem. The bathymetry of the sea bottom is respected. No regularization strategy is used here, as we consider synthetic data without additional noise.

The cycle of FWI starts with an initial model computed as a smooth version of the exact model using Gaussian smooth-

Table 1 Frequency group strategy for the BP 2004 case study.

Group 1	2 Hz	2.25 Hz	2.5 Hz	2.75 Hz	
Group 2	2.5 Hz	3 Hz	3.5 Hz	4 Hz	
Group 3	4 Hz	4.5 Hz	5 Hz	5.5 Hz	
Group 4	5.5 Hz	6 Hz	6.5 Hz	7 Hz	
Group 5	7 Hz	7.5 Hz	8 Hz	8.5 Hz	
Group 6	8.5 Hz	9.5 Hz	10.5 Hz	11.5 Hz	
Group 7	11.5 Hz	12.5 Hz	13.5 Hz	14.5 Hz	15.5 Hz
Group 8	15.5 Hz	16.5 Hz	17.5 Hz	18.5 Hz	19.5 Hz

ing. Compared with previous work (Métivier *et al.* 2012), we use here a smoother initial model. The characteristic length used for this smoothing is set to 500 m instead of 375 m for the former study. The resulting initial model is presented in Figure 1b. The use of this smoother initial model requires a careful hierarchical frequency strategy. We generate 27 data sets, of frequencies from 2 Hz to 19.5 Hz gathered into eight overlapping subgroups, as presented in Table 1.

Estimated models

The four final models obtained for the last frequency group are presented in Figure 2. From 1 km to 6 km in the horizontal direction, the top salt-structure is correctly delineated in the four estimates. The reconstruction of the basin between $x = 0 \text{ km}$ and $x = 1 \text{ km}$ seems more difficult. This basin is responsible for high amplitude multi-scattered waves that are difficult to interpret, and is located at one extremity of the acquisition. The nonlinear conjugate gradient method and the *l*-BFGS method seem to be most affected by this particular configuration. The geometry of the basin is not recovered and it is filled with high amplitude

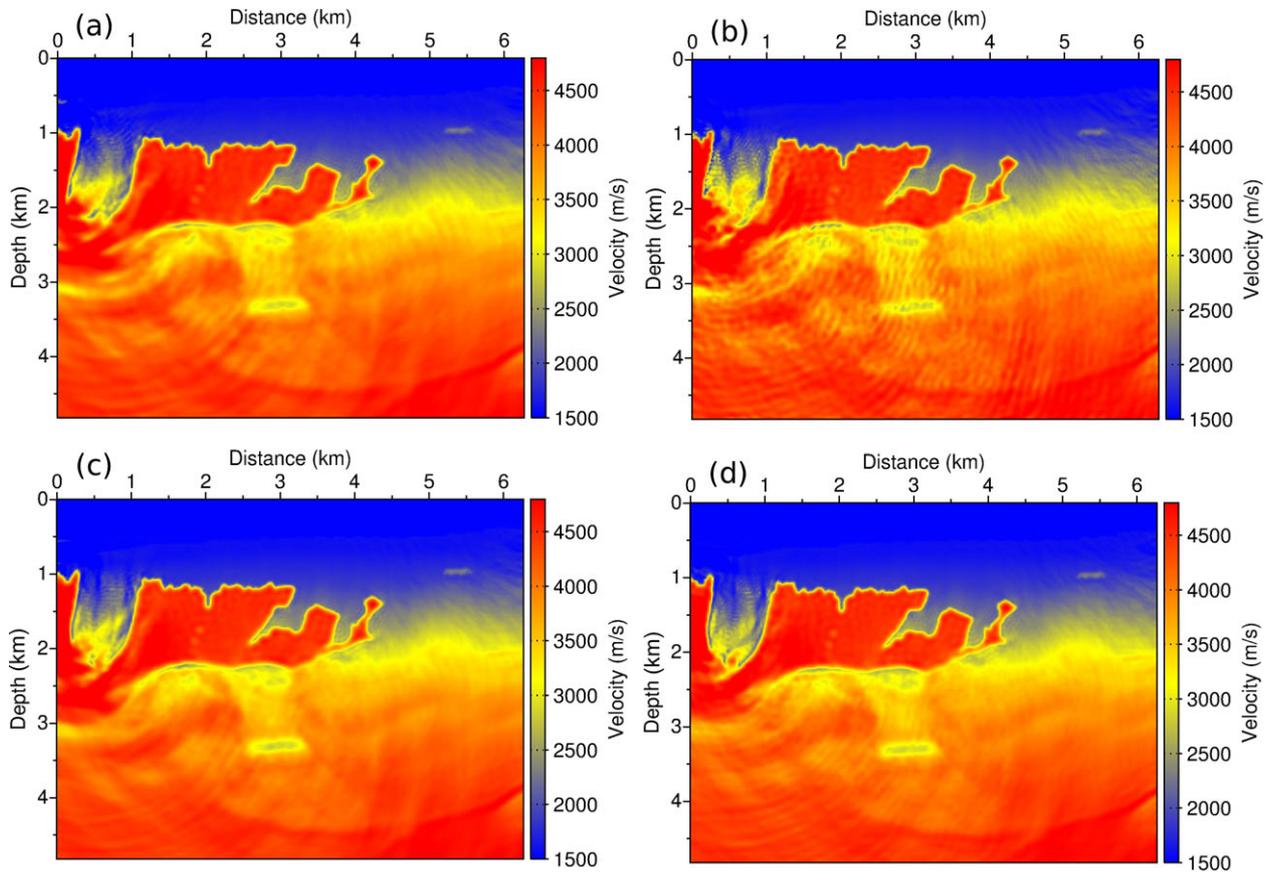


Figure 2 Estimated models for the BP case study. Nonlinear conjugate gradient (a), l -BFGS method (b), truncated Gauss–Newton method (c), truncated Newton method (d).

velocities. These perturbations are also responsible for obscuring the sub-salt targets, and creating erroneous slow velocity anomalies.

Conversely, the results provided by the truncated Gauss–Newton or the truncated Newton method are more reliable. The best estimate is provided by the truncated Newton method. The geometry of the basin is better recovered, and the sub-salt slow velocity anomalies better reconstructed. The possible enhancement of the inverse Hessian approximation yielded by the truncated Newton method may explain this improvement in the resolution and the stability of the inversion. The regularization effect of the truncation strategy (Kaltenbacher *et al.* 2008) may also contribute to the better quality of the subsurface BP 2004 model estimates. We see particularly that the l -BFGS estimate (and, to a lesser extent, the nonlinear conjugate gradient estimate) are affected by high frequency artefacts. Conversely, the truncated Newton and Gauss–Newton estimates are significantly smoother.

Convergence profiles and computational cost

The convergence profiles of the four methods are presented for frequency groups 1, 4 and 8 in Figure 3. The convergence profiles in terms of nonlinear iterations clearly show the higher convergence rate of the truncated Newton methods (Gauss–Newton and full Newton) compared with the l -BFGS method and the nonlinear conjugate gradient method. This higher convergence rate is less pronounced for the lowest frequency group and becomes more obvious at higher frequencies. In particular, for the last frequency group, the nonlinear conjugate gradient and the l -BFGS method does not generate a model that satisfies the misfit reduction set to 10^{-2} . These two methods stop because of a linesearch failure after more than 2000 nonlinear iterations. This linesearch failure indicates that the model update that is computed at the final iteration fails to further minimize the misfit function. This behaviour may be caused by the stronger imprint of multiple reflections in the data at high frequency.

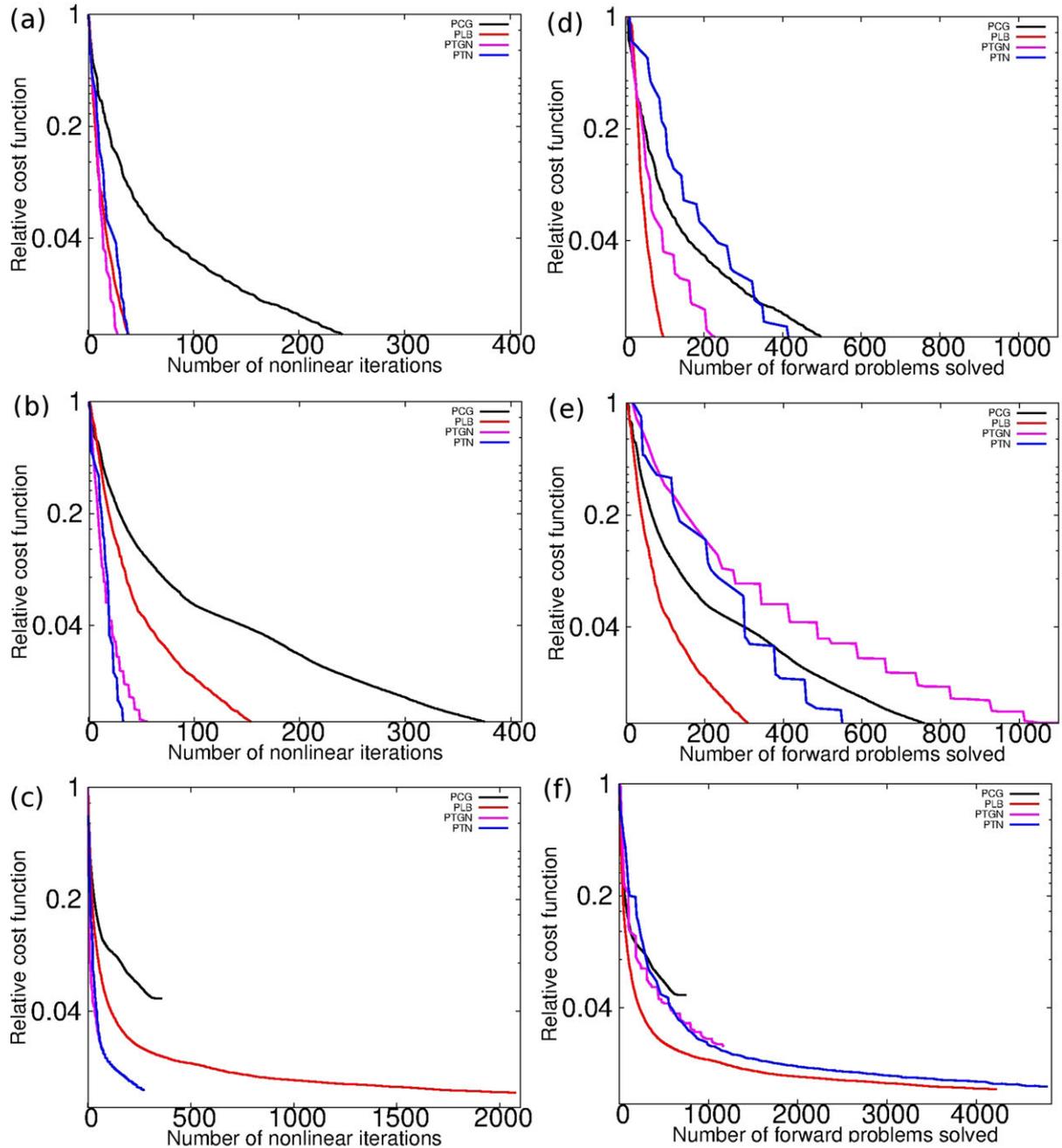


Figure 3 Misfit function decrease for the BP 2004 case study. (a–c) Decrease with respect to the number of nonlinear iterations: first frequency group (a), fourth frequency group (b), eighth frequency group (c). (d–f) Decrease with respect to the number of wave equation problem solved: first frequency group (d), fourth frequency group (e), eighth frequency group (f). PCG, nonlinear conjugate gradient; PLB, *l*-BFGS; PTN, truncated Newton; PTGN, truncated Gauss–Newton method.

However, for frequency groups 1 and 4, the speed-up provided by the truncated Newton method in terms of nonlinear iterations does not compensate for the extra computation cost required to compute the descent direction. Indeed, we can detect plateaux on the convergence profiles (especially

with respect to the number of forward problem solutions), which correspond to slowly converging inner iterations. The maximum allowed number of iterations, set to 30 for this experiment, is reached. Hence, reducing the total number of authorized internal iterations could possibly improve the

Table 2 Frequency group strategy for the Valhall case study.

Group	3.54 Hz	3.78 Hz	4.03 Hz
Group 1	3.54 Hz	3.78 Hz	4.03 Hz
Group 2	4.03 Hz	4.28 Hz	4.64 Hz
Group 3	4.64 Hz	5.00 Hz	5.25 Hz
Group 4	5.25 Hz	5.62 Hz	5.99 Hz

convergence of the truncated Newton methods in terms of forward problem solutions. In this noise-free synthetic case study, compared with the preconditioned gradient-based method, the truncated Newton method, however, appears to provide better subsurface estimates. We investigate in the next section how this method behaves on a real noise contaminated data case study.

2D Valhall case study

Description

The Valhall oil field is located in the North Sea, and has been in production since 1982. In this shallow water environment field, the water depth hardly reaches 100 m. This particular configuration is adapted for the use of ocean bottom cables (OBC) seismic recorders. Four component sensors have thus been placed at the sea bottom. Several 3D seismic surveys have been performed to follow the time evolution of the oil field. The Valhall seismic data have been investigated by several authors in the context of 2D multi-parameter FWI (Prioux *et al.* 2011, 2013a,b) and 3D mono-parameter FWI (Sirgue *et al.* 2010; Etienne *et al.* 2012).

In this study, we are interested in the mono-parameter inversion of a 2D line from the Valhall OBC data set. This data set involves 320 sources located at a depth of 5 m in the water layer, and 210 receivers located on the sea bottom, between 68 and 72 m deep. Sources and receivers are equally spaced each 50 m. Only the hydrophone component is used. The lowest frequency inverted is 3.5 Hz. The signal-to-noise ratio (SNR) for lower frequencies is too weak for the data to be used in this frequency band.

Below the sea level, the presence of soft shale sediments is responsible for strong attenuation and anisotropy of the wave propagation. As demonstrated in the work of Prioux *et al.* (2011), it is necessary to account for this anisotropy in the modelling to correctly invert the Valhall data set. For this case study, we thus change the forward problem from standard isotropic acoustic modelling to the VTI acoustic modelling proposed by Operto *et al.* (2009).

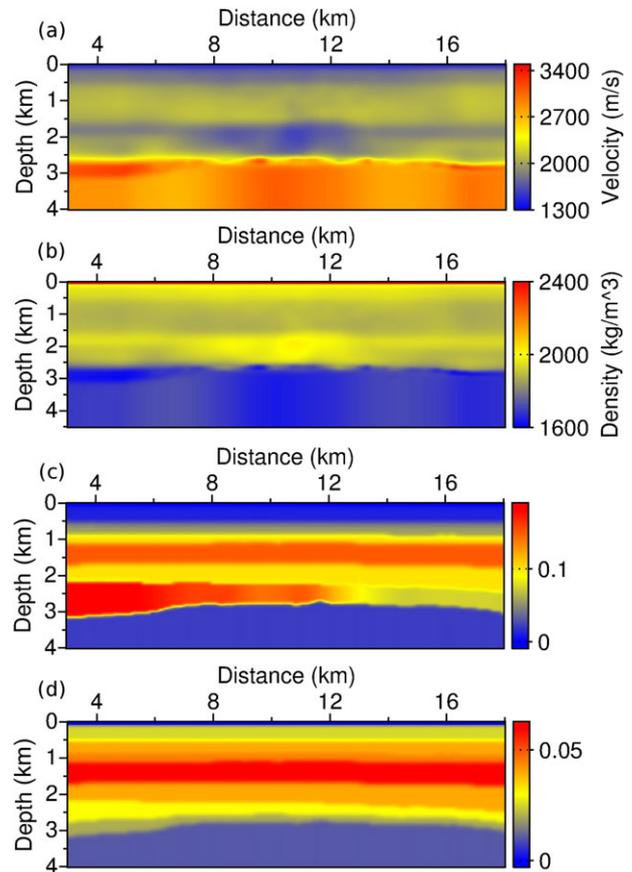


Figure 4 Initial models for P-wave velocity (a), density (b); dimensionless Thomsen parameters ϵ (c), δ (d).

From 3.5 Hz to 6 Hz, we define four overlapping frequency groups to be inverted sequentially (Table 2). The VTI forward modelling requires us to define initial models not only for P-wave velocity, density and attenuation, but also for the Thomsen anisotropy parameters ϵ and δ . The initial models for P-wave velocity and Thomsen parameters are determined by reflection travel-time tomography (courtesy of BP). The density initial model is derived from the P-wave velocity initial model through the Gardner law. The quality factor model is taken constant equal to 200. The corresponding models are presented in Figure 4. In our experiments, the initial models for density, attenuation and Thomsen parameters remain the same, and we only focus on the reconstruction of the P-wave velocity model (Gholami *et al.* 2013).

Regularization strategy

As the real seismic data is noise contaminated, an appropriate regularization strategy has to be designed. A trade-off has to

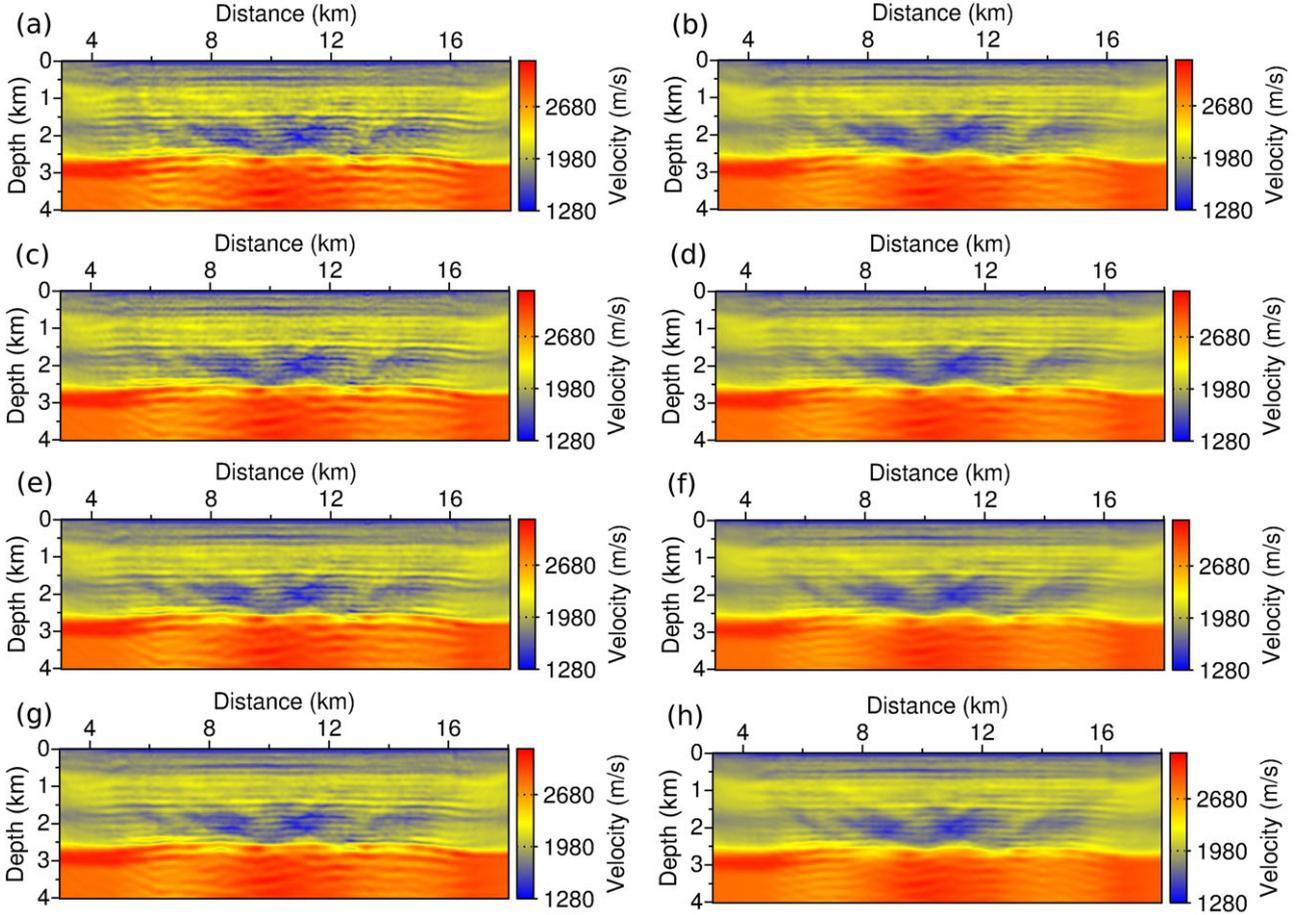


Figure 5 Computed P-wave velocity models. Nonlinear conjugate gradient with β_1 (a), with β_2 (b). *l*-BFGS with β_1 (c), with β_2 (d). Truncated Gauss–Newton with β_1 (e), with β_2 (f). Truncated Newton with β_1 (g), with β_2 (h).

be found between the model regularity and the reduction of the misfit. The first parameter that is used to control the noise is thus the stopping criterion ε , which defines the expected decrease of the relative misfit function (data driven strategy). For the three first frequency groups, setting $\varepsilon = 0.6$ yields an acceptable trade-off. For the fourth frequency group, however, the noise contamination has a stronger impact on the FWI results. To mitigate this difficulty, one can either set a higher value for the parameter ε or introduce a regularization term in the misfit function (model-driven strategy). In this case, the first option is not satisfactory: without additional regularization, the quality of the reconstructed model is degraded at the first iterations. The trade-off to be found between the model regularity and the misfit is thus not acceptable. We therefore choose to combine the data driven strategy to the model driven strategy. We set $\varepsilon = 0.7$ and we add a standard Tikhonov regularization term $T(m)$ to the misfit function $f(m)$. This defines the regularized misfit function $f_T(m)$, such that

$$f_T(m) = f(m) + \frac{\beta}{2} T(m), \quad (45)$$

where

$$T(m) = \left(\beta_x \left\| \frac{\partial m}{\partial x} \right\|^2 + \beta_z \left\| \frac{\partial m}{\partial z} \right\|^2 \right). \quad (46)$$

The parameter β accounts for the influence of the regularization term. The parameters β_x and β_z can be used to enforce a stronger regularization in the direction x or z . In the experiments presented in what follows, we use the settings

$$\beta_x = 1, \quad \beta_z = 0.5, \quad (47)$$

to enforce smoother variations in the horizontal direction. The introduction of $T(m)$ into the misfit function $f(m)$ also results in additive terms in the gradient and the Hessian-vector products. We account for these additive terms.

In addition, the truncation strategy for the computation of the descent direction within the truncated Newton

or Gauss–Newton methods is used as an additional regularization (see the work of Kaltenbacher *et al.* (2008) on this particular topic). The maximum number of inner conjugate gradient iterations is set to three.

To emphasize the sensitivity of the methods to the regularization parameter β , we have performed experiments for two different values of β , reflecting what we may call a “weak” β_1 and a “strong” β_2 regularization.

$$\beta_1 = 10^{-2}, \quad \beta_2 = 5 \times 10^{-2}. \quad (48)$$

FWI results

We present the results obtained by our four optimization schemes in Figure 5. Below shallow layers of shale sediments (from $z = 0$ to $z = 1.5$ km), we can see two low velocity zones, corresponding to the presence of gas layers (between $z = 1.5$ km and $z = 2.5$ km, $x = 8$ km and $x = 12$ km). Below these gas layers (between $z = 2.5$ km and $z = 4$ km), we can locate the cap rock of the reservoir and stronger reflectors.

The left column of Figure 5 shows the results obtained with the “weak” regularization parameter β_1 . The right column of Figure 5 shows the results obtained with the “strong” regularization parameter β_2 . The four optimization methods yield very similar results. The increase of the regularization parameter β from β_1 to β_2 significantly smooths out the results in all cases. Compared with the nonlinear conjugate gradient and the *l*-BFGS method, it seems that the truncated Gauss–Newton and Newton methods yield slightly smoother results.

To better visualize the smoothing effect of the truncated Newton method strategy, we compare the results obtained for $\beta = \beta_1$ using the *l*-BFGS method and the truncated Newton method (Figure 6). We remove the initial smooth velocity model from the two model estimates and compare the updates that have been provided by the two optimization methods. This comparison emphasizes the introduction of noise in the model estimate using the *l*-BFGS method, compared with the truncated Newton method result. More particularly, we see that the *l*-BFGS method incorporates noise in the top sediment layers. This noise is likely to generate artefacts in deeper zones, for instance in the gas layer zones at the centre of the model. The truncated Newton method thus appears more robust than the *l*-BFGS method when dealing with noise-contaminated data. As these results are presented for the same regularization weight β_1 , this indicates that better trade-offs between resolution and regularization can be found using the truncated Newton method.

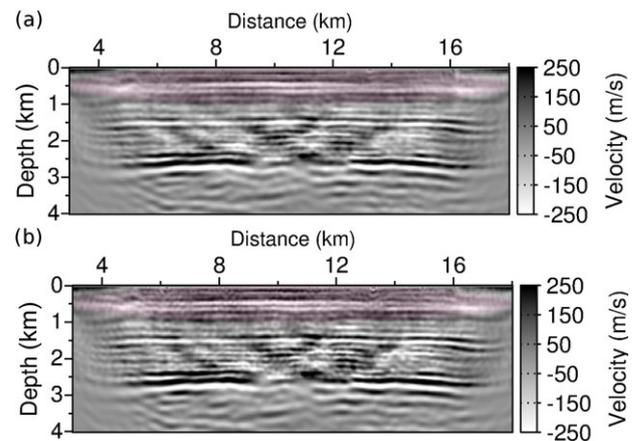


Figure 6 Computed P-wave velocity updates for the regularization weight β_1 . The initial velocity model has been removed from the final estimations. Truncated Newton method (a), *l*-BFGS method (b). The red ellipses focus on the shallow part of the updates: the *l*-BFGS update is more affected by the noise.

As a quality control of the results, we compare in Figure 7 the well-logs corresponding to the results. This well-log is located at $x = 9.5$ km. The results yielded by the four optimizations are quite similar. As expected, the results corresponding to the choice of a stronger regularization $\beta = \beta_2$ are smoother. We are particularly interested in the jump in depth at $z = 3.5$ km (green circles). This jump corresponds to the presence of an actual reflector (see the migrated sections in Figure 8). This jump can be clearly identified on the results obtained with the weak regularization parameter $\beta = \beta_1$. For the stronger regularization case $\beta = \beta_2$, the identification of the reflector is more difficult. This illustrates the classical trade-off between the resolution and the regularity expected from the solution.

We also compare in Figure 8 the migrated images obtained using the initial velocity model and the four final estimates provided by the four optimization methods, in the “weak” regularization case ($\beta = \beta_1$). As already reported (Gholami *et al.* 2013; Prioux *et al.* 2013a,b), FWI velocity models improve the migrated image in the shallow part (up to 1 km deep). In this area, the common image gathers (CIGs) are flattened, compared with the CIGs associated with the migration in the initial velocity model. The four final FWI models also systematically improve the image of the deep reflector at $z = 3.5$ km and $x = 11$ km. In addition, the two reflectors that delineate the top and the base of the reservoir (between $z = 2.5$ km and $z = 3$ km) are also better imaged using the final FWI models, particularly on the lateral extremities of the model ($x = 5$ km and $x = 13$ km). This is

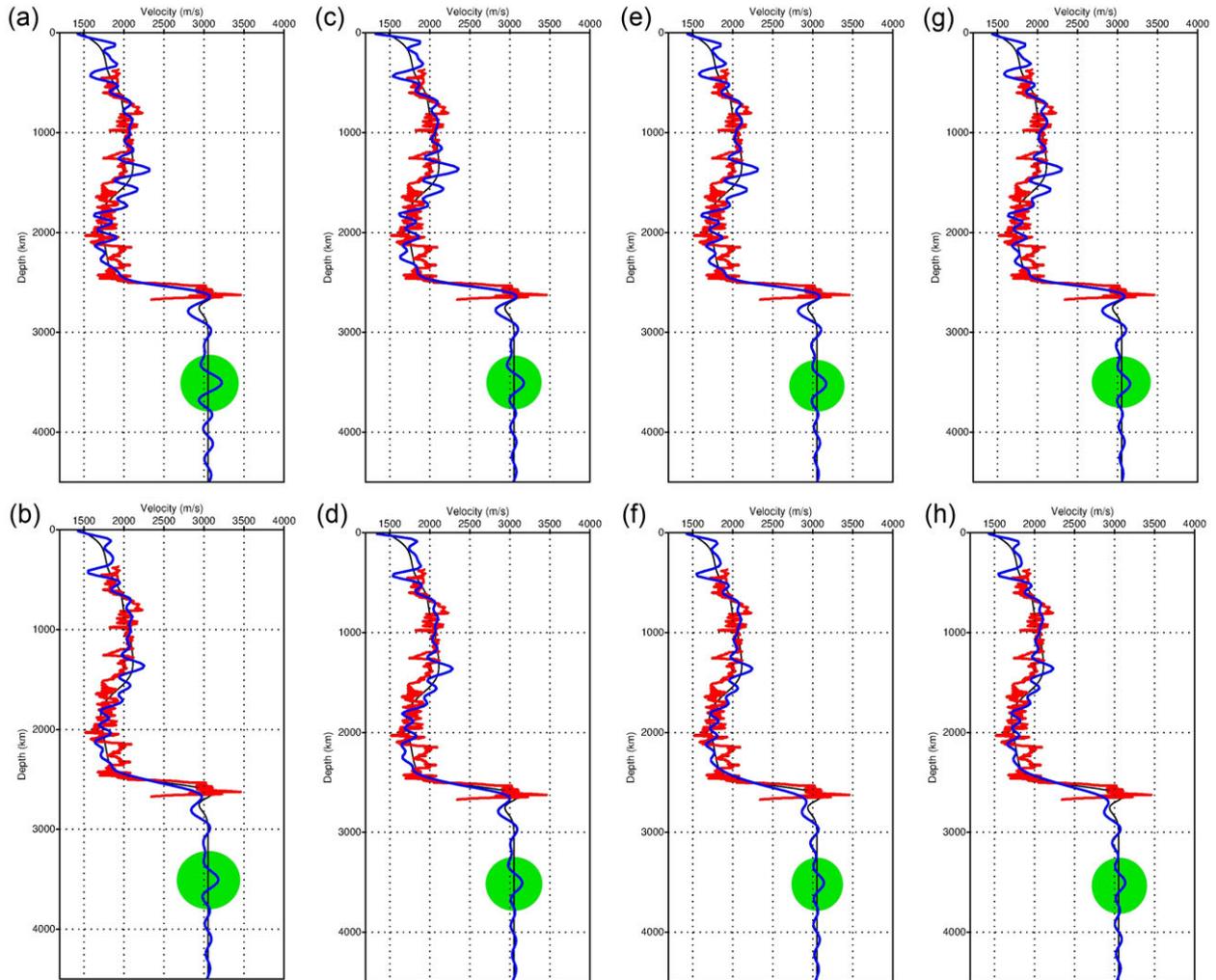


Figure 7 P-wave velocity logs compared with a reference well-log velocity at $x = 9.5$ km. Black: initial model. Blue: estimated models. Red: well-log velocity. Nonlinear conjugate gradient with β_1 (a), with β_2 (b). l -BFGS with β_1 (c), with β_2 (d). Truncated Gauss–Newton with β_1 (e), with β_2 (f). Truncated Newton with β_1 (g), with β_2 (h). The green circles focus on the reconstruction of the deep reflector visible on the migrated sections in Figure 8.

correlated by the CIGs (Figure 9). We see in particular that the velocity models provided by the truncated Newton and Gauss–Newton methods provide a good improvement of the flatness of the CIGs at $x = 5.1$ km compared with the initial velocity model.

Convergence profiles and computational cost

The convergence profiles for the Valhall case study are displayed in Figures 10 and 11. In Figure 10, we see that the convergence of the four optimization methods in terms of nonlinear iterations is very similar for the frequency

1 and 3. As each nonlinear iteration requires us to solve additional forward problems for the truncated Gauss–Newton and Newton methods, the computational cost of these two methods is higher than for the gradient-based methods. This is clearly shown by the convergence curves in terms of forward problem solutions.

Figure 11 presents the convergence curves for frequency group 4, with the two different regularization weights β_1 and β_2 . The truncated Gauss–Newton and Newton methods seem to be more sensitive to the choice of the regularization parameters. For $\beta = \beta_1$ the convergence rate of these two methods, in terms of nonlinear iterations, is similar to the

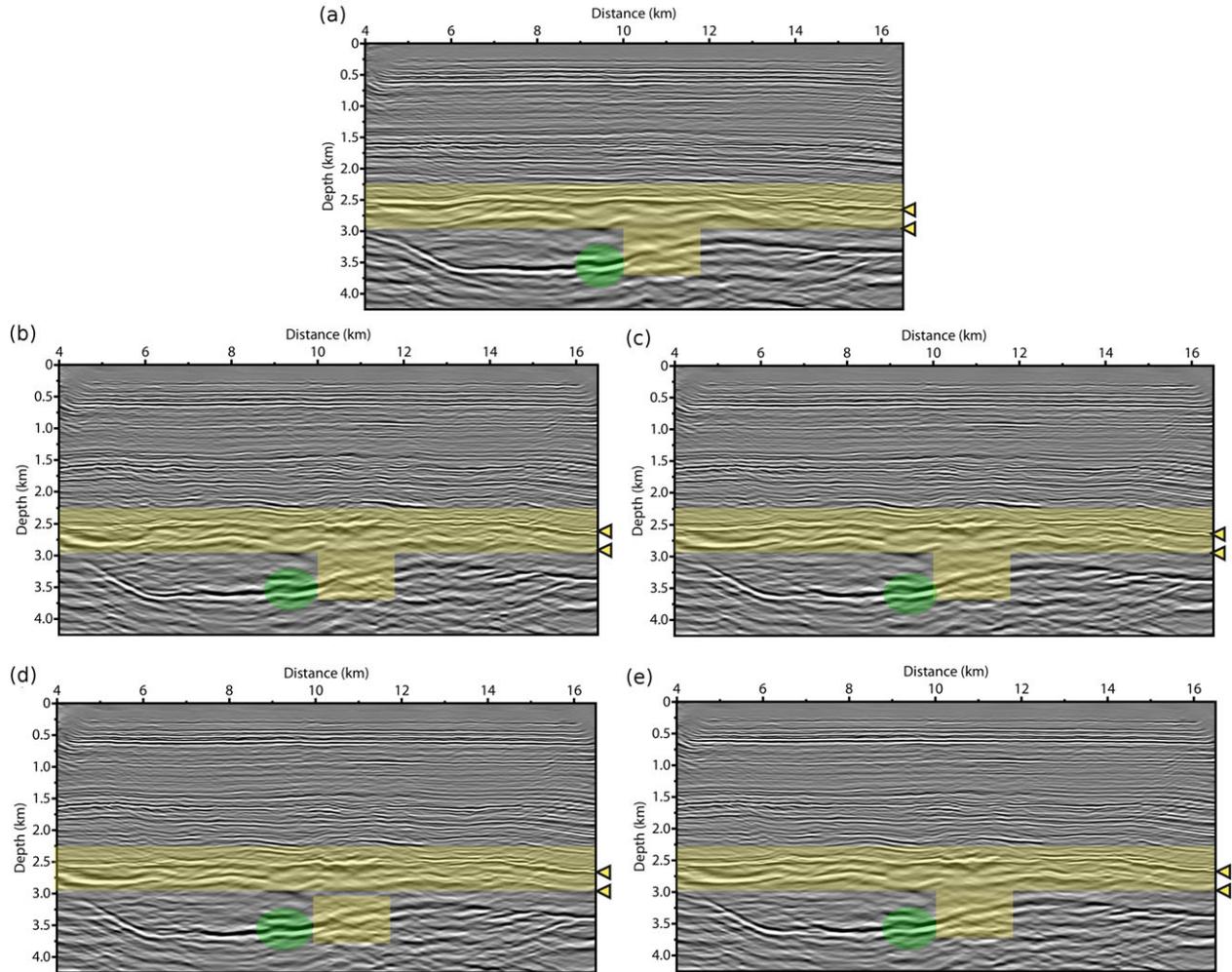


Figure 8 Reverse time migration results using the initial model (a), and FWI results provided by the four optimization methods using the “weak” regularization weight β_1 (b-e). Nonlinear conjugate gradient (b), l -BFGS (c), truncated Gauss–Newton (d), truncated Newton (e). The green circles focus on the deep reflector at $z = 3.5$ km highlighted in the P-wave velocity logs in Figure 7. The yellow regions focus on the right part of this deep reflector and on two reflectors indicated by the arrows. These reflectors delineate the top and the bottom of the reservoir.

nonlinear conjugate gradient or the l -BFGS method, as is the case for the frequency groups 1 and 3. However, for $\beta = \beta_2$ the convergence rate of the truncated Gauss–Newton and Newton methods decreases, and is smaller than for the gradient-based methods. In terms of computational cost, therefore, the truncated Gauss–Newton and Newton require more forward problem resolutions, as can be seen on the right column of Figure 11.

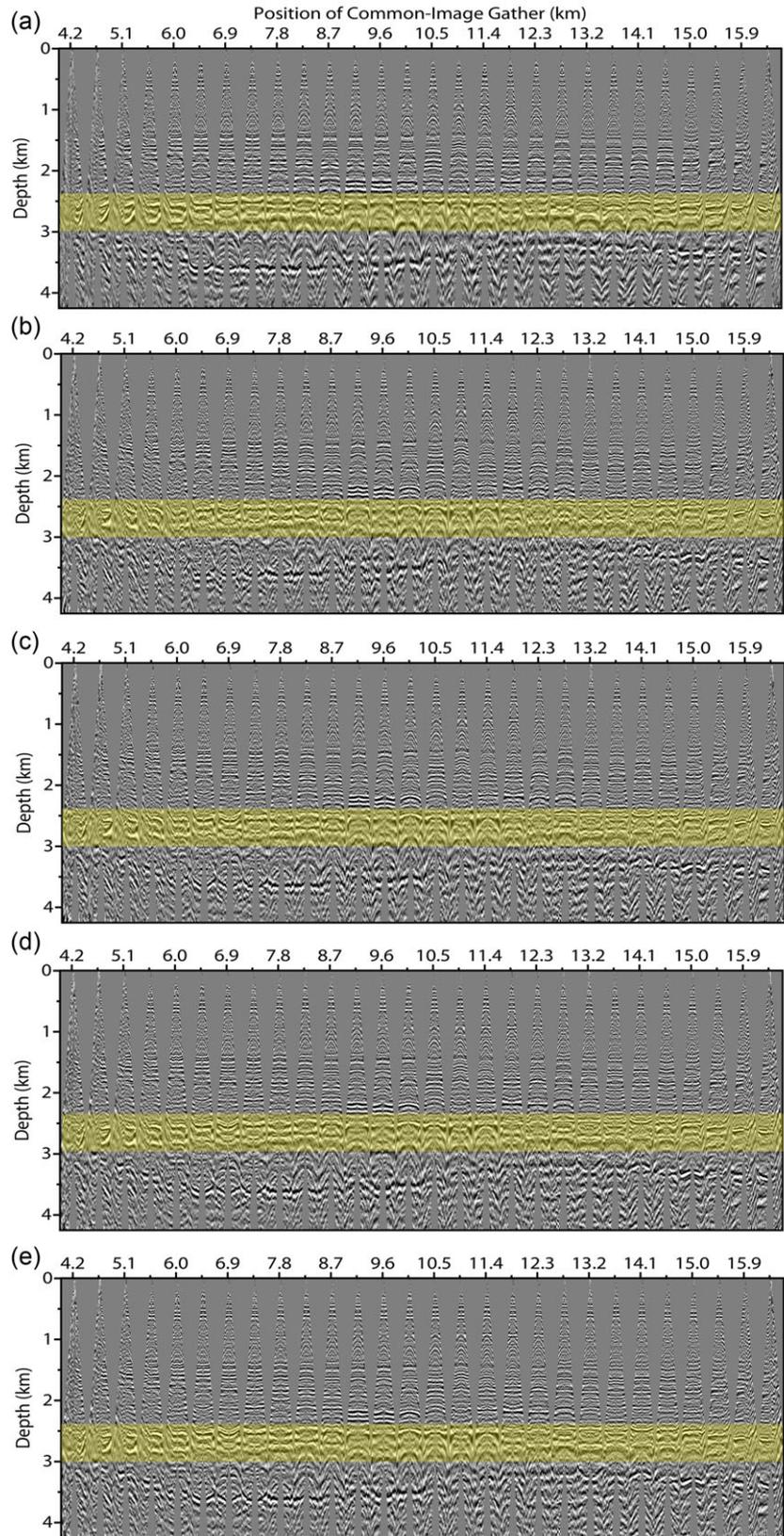
For the Valhall case study, it thus seems that the different type of optimization methods perform similarly. The best computation costs are provided by the l -BFGS method; however, the difference with the nonlinear conjugate gradient method is small. This should be related to the choice of a small

memory parameter ($l = 3$). The truncated Gauss–Newton and Newton methods appear to be relatively more costly. In addition, it is difficult in this case to discriminate between these two methods, the reconstructed models and the convergence rates are similar.

Resolution analysis

To conclude the comparison between gradient-based and truncated Newton methods, we perform a numerical resolution test (spike test). We add a positive perturbation of 200 m.s^{-1} in the gas cloud localized around $x = 11$ km, $z = 2$ km. We compute synthetic data using the same

Figure 9 Common image gather corresponding to the migrated image using the starting velocity model (a) provided and FWI results provided by the four optimization methods using the “weak” regularization weight β_1 (b–e). Nonlinear conjugate gradient (b), *l*-BFGS (c), truncated Gauss–Newton (d), truncated Newton (e). The yellow regions correspond to those of the migrated images in Figure 8.



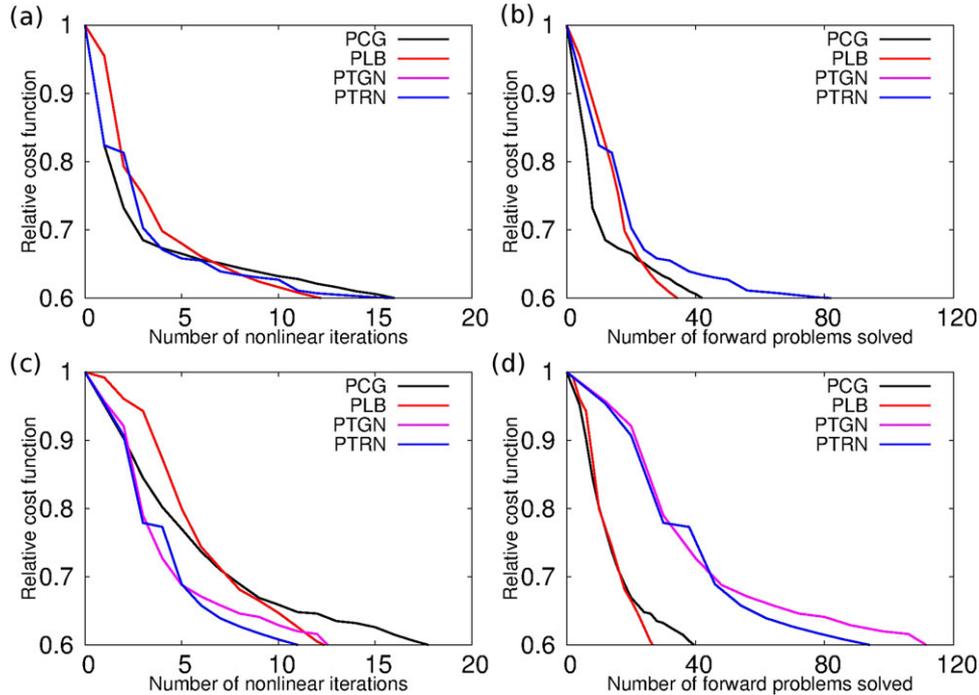


Figure 10 Misfit function decrease for the Valhall case study for frequency groups 1 and 3. Decrease with respect to the number of nonlinear iterations: frequency group 1 (a), frequency group 3 (c). Decrease with respect to the number of wave equation problems solved: frequency group 1 (b), frequency group 3 (d). PCG, nonlinear conjugate gradient; PLB, *l*-BFGS; PTGN, truncated Gauss–Newton method; PTRN, truncated Newton.

surface acquisition geometry in this perturbed model. We invert the synthetic data starting from the unperturbed model, and we perform only one nonlinear iteration. We do not use regularization. In this case, the nonlinear conjugate gradient and the *l*-BFGS method are equivalent: at the first iteration, the *l*-BFGS approximation is just given by the preconditioning matrix.

We compare the reconstruction of the perturbation for the nonlinear conjugate gradient model and for the truncated Newton model. For the truncated Newton experiment, we compute the perturbation using 5, 10 and 20 inner iterations.

The results are presented in Figure 12. Not surprisingly, the focusing achieved by the nonlinear conjugate gradient method is poorer than the one obtained using the truncated Newton method. The amplitude of the perturbation reconstructed by the nonlinear conjugate gradient (20 m.s^{-1}) is slightly smaller than the one reached by the truncated Newton method (30 m.s^{-1}). More interestingly, we can see the progressive refocusing effect associated with the inverse Hessian operator on the model update obtained depending on the number of inner linear iterations performed. The am-

plitude of the reconstructed perturbation is increased, and the artefacts around the perturbation vanish. This demonstrates the theoretical interest of using the local curvature information through the truncated Newton process. Note that the computational cost of performing such a resolution analysis in a single point of the medium requires 10 to 20 Hessian-vector products, which implies 20 to 40 forward problem resolutions per sources provided the incident and first-order adjoint wavefields can be stored. Assuming an average cost of 50 forward problems per sources to solve per frequency group, the solution of the entire inverse problem requires 200 forward problems. Therefore, performing the resolution analysis in 10 different locations using second-order adjoint formulae requires the same computational effort as solving one or two entire inverse problems.

DISCUSSION

In this study, we are interested in the implementation of the truncated Newton minimization scheme for FWI. The algorithm is based on a two-nested-loops architecture: the external

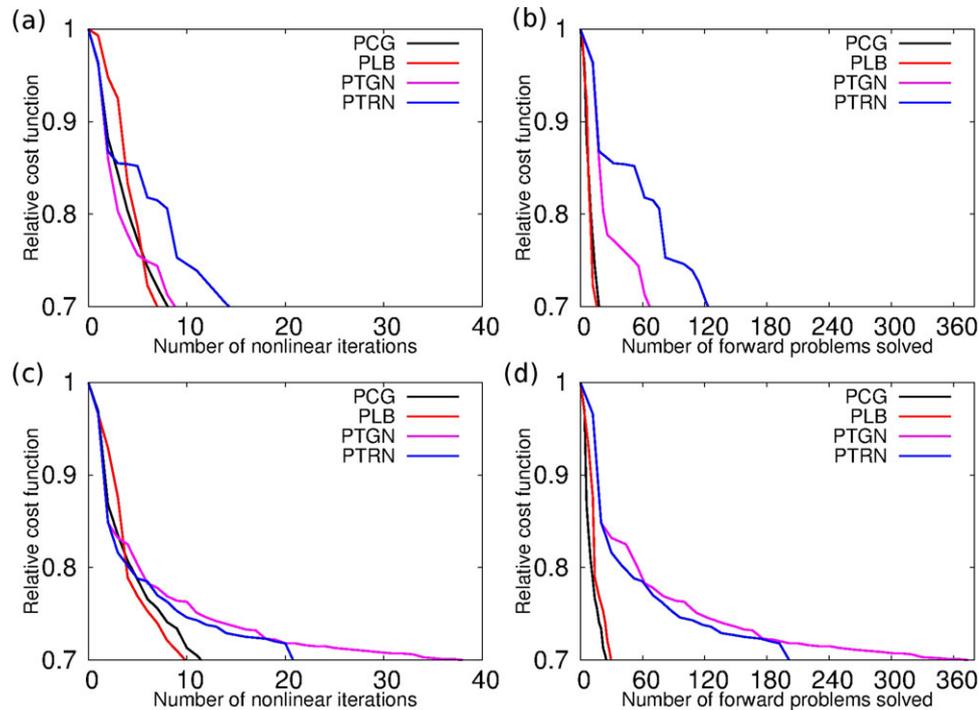


Figure 11 Misfit function decrease for the Valhall case study for frequency group 4. Decrease with respect to the number of nonlinear iterations with regularization parameter β_1 (a), β_2 (c). Decrease with respect to the number of forward problems solved with regularization parameter β_1 (b), β_2 (d). PCG, nonlinear conjugate gradient; PLB, l -BFGS; PTGN, truncated Gauss–Newton method; PTRN, truncated Newton.

loop involves updating an initial model up to the final estimation. The inner loop involves the computation of the model updates from an inexact solution of the linear system associated with the computation of the Newton descent direction. This inexact solution is computed with a matrix-free conjugate gradient solver. Only the action of the Hessian operator on an arbitrary vector has to be computed. This can be achieved efficiently by using second-order adjoint state formulae: only two additional wave propagation problems have to be solved to compute this quantity. The use of an adaptive stopping criterion for the inner loop and a suitable preconditioner enhance the efficiency of the algorithm.

Compared with conventional preconditioned gradient-based methods, such as the nonlinear conjugate gradient and l -BFGS methods, the truncated Newton method offers the possibility of estimating the curvature of the misfit function locally and accounting for it in the minimization scheme. The synthetic BP 2004 case study emphasizes the difficulty of salt and sub-salt imaging. The high velocity contrasts caused by the presence of large salt structures is responsible for the presence of energetic multiple reflections in the data. This cause difficulties in standard minimization methods to provide correct images of the salt and sub-salt targets. In this case, the

truncated Newton method seems to be more robust than the nonlinear conjugate gradient and the l -BFGS method. In particular, the truncated Newton method provide better results than the truncated Gauss–Newton method. A possible explanation is that the part of the Hessian operator that is neglected in the Gauss–Newton approximation is related to double scattered waves. In this context of multiple scattering between the salt and the free surface, it could be crucial to account for the whole Hessian operator.

The behaviour of the four minimization methods on the Valhall real data case study is different. The convergence rate in terms of nonlinear iterations is approximately the same for the four optimization methods. This indicates that the additional computational cost performed at each nonlinear iteration by the truncated Newton method is not compensated for by an increased convergence rate. In this case, the method thus becomes more expensive in terms of overall computational cost (number of forward problems to be solved). The fact that the four minimization methods converge at the same rate seems to emphasize that, in this case, the inverse Hessian estimation, whether performed through the l -BFGS approximation or the truncated Newton strategy, does not provide any speed-up. This may be due to the regularization

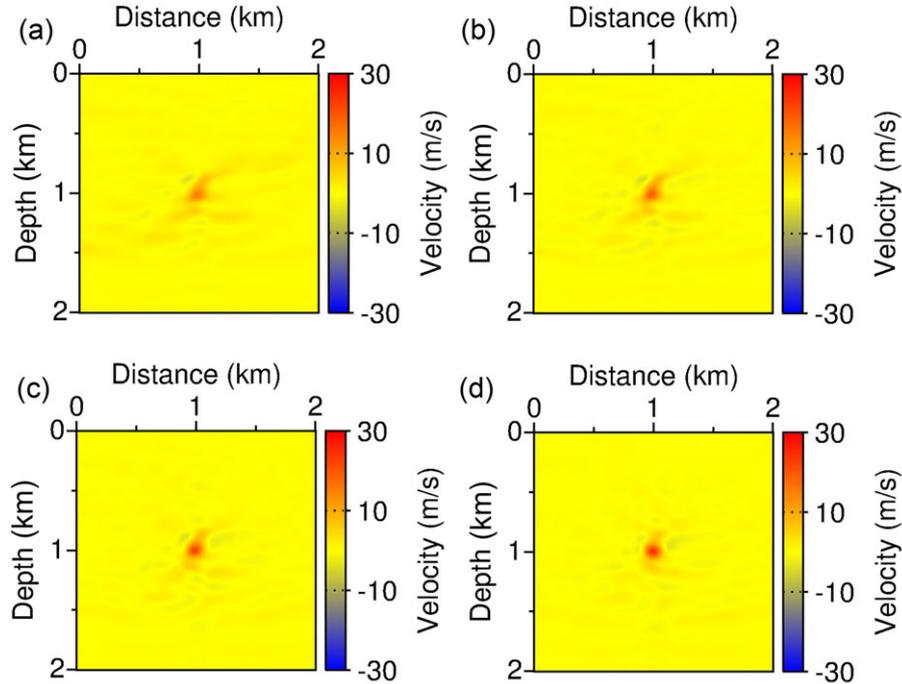


Figure 12 Reconstruction of a small amplitude perturbation in the final estimation. Zoom on the perturbation. Nonlinear conjugate gradient (a), truncated Newton method with 5 inner linear iterations (b), 10 inner linear iterations (c), 20 inner linear iterations (d).

needed to control the noise contamination of the data. The deconvolution effect of the Hessian is counterbalanced by either the truncation process or the damping, which is added as a Tikhonov regularization term. From a qualitative point of view, however, the results provided by the truncated Gauss–Newton and Newton methods seem to be smoother than those provided by gradient-based methods, for a given level of minimization and regularization. This may be related to the smoothing effect of the truncation strategy (Kaltenbacher *et al.* 2008). A possible consequence is that the truncated Newton method may be used with smaller Tikhonov regularization weights than the gradient-based methods, which would allow one to reach a better trade-off between resolution and regularization. Indeed, the resolution power of the truncated Newton method is, at least in principle, better than gradient-based methods, as emphasized by the resolution analysis case study.

The numerical performances of the different minimization schemes that we have compared depend strongly on an accurate estimate of the inverse Hessian operator as a preconditioner. For now, we have only used the diagonal preconditioner from the pseudo-Hessian approach developed by Shin *et al.* (2001). The accuracy of this approximation is, however, limited and the use of this preconditioner is restricted to sur-

face acquisition configurations. Therefore, we are interested in the development of better approximation of the inverse Hessian; see for instance Benzi (2002); Bekas *et al.* (2007); Chiu and Demanet (2012).

Another way of improving the computation speed of the solution of the inner linear systems is related to deflation strategies (Saad 2003). From the Krylov subspace created during the solution of the inner linear system at iteration $k - 1$, an estimation of the smallest eigenvectors of the Hessian operator can be computed. These eigenvectors can be inserted in the starting Krylov subspace at iteration k , to improve the convergence. Related methods developed within the data assimilation community could also be investigated. Gratton *et al.* (2011) propose, for instance, to compute an approximate initial guess for the inner linear systems in a reduced space given by a spectral decomposition of the initial Hessian operator.

An accurate estimation of the inverse Hessian operator should be even more important in the context of multi-parameter FWI (Operto *et al.* 2013). The simultaneous reconstruction of such parameters as, for instance, the P-wave velocity and the density is affected by strong trade-offs. The perturbation in the P-wave models can be erroneously interpreted as perturbations in the density model, and *vice versa*. As the Hessian operator expresses the correlated effects of

variations between different classes of parameter on the misfit function, a more accurate estimation of the inverse Hessian operator should help to mitigate these trade-off effects. In this context, the truncated Newton method could thus produce more reliable results. This topic will be investigated in future studies.

ACKNOWLEDGEMENTS

This study was funded by the SEISCOPE II consortium (<http://seiscope2.osug.fr>), sponsored by BP, CGG, Chevron, Exxon Mobil, JGI, Petrobras, Saudi Aramco, Shell, Statoil, Total and WesternGeco. This study was granted access to the HPC facilities of CIMENT (Université Joseph Fourier Grenoble), and of GENCI-CINES under Grant 2011-046091 of GENCI (Grand Equipement National de Calcul Intensif.) The authors would like to thank V. Tcheverda for his invitation to publish this study. They also thank BP Norge AS and their Valhall partner Hess Norge AS, for allowing access to the Valhall data set as well as the well-log velocities.

REFERENCES

- Amestoy P., Duff I.S. and L'Excellent J.Y., 2000. Multifrontal parallel distributed symmetric and unsymmetric solvers. *Computer Methods in Applied Mechanics and Engineering* 184(2-4), 501–520.
- Bekas C., Kokiopoulou E. and Saad Y. 2007. An estimator for the diagonal of a matrix. *Applied Numerical Mathematics* 57(11-12), 1214–1229.
- Benzi M. 2002. Preconditioning techniques for large linear systems: a survey. *Journal of Computational Physics* 182, 418–477.
- Bérenger J.-P. 1994. A perfectly matched layer for absorption of electromagnetic waves. *Journal of Computational Physics* 114, 185–200.
- Billette F.J. and Brandsberg-Dahl S. 2004. The 2004 BP velocity benchmark. *67th Annual EAGE Conference & Exhibition, Madrid, Spain, Extended Abstracts*, B035.
- Bonnans J.F., Gilbert J.C., Lemaréchal C. and Sagastizábal C.A. 2006. *Numerical Optimization, Theoretical and Practical Aspects*. Springer series. Universitext.
- Brossier R., Operto S. and Virieux J. 2009. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics* 74(6), WCC105–WCC118.
- Brossier R., Operto S. and Virieux J. 2010. Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*. 75(3), R37–R46.
- Byrd R.H., Lu P. and Nocedal J. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*. 16, 1190–1208.
- Chavent G. 1974. Identification of parameter distributed systems. in: *Identification of function Parameters in Partial Differential Equations*, (ed. R. Goodson and M. Polis), American Society of Mechanical Engineers, pp. 31–48.
- Chiu J. and Demanet L. 2012. Matrix probing and its conditioning. *SIAM Journal on Numerical Analysis* 50(1), 171–193.
- Dai Y. and Yuan Y. 1999. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization* 10, 177–182.
- Eisenstat S.C. and Walker H.F. 1994. Choosing the forcing terms in an inexact Newton method. *SIAM Journal on Scientific Computing* 17, 16–32.
- Epanomeritakis I., Akçelik V., Ghattas O. and Bielak J. 2008. A Newton-CG method for large-scale three-dimensional elastic full waveform seismic inversion. *Inverse Problems* 24, 1–26.
- Etienne V., Hu G., Operto S., Virieux J., Barkved O.I. and Kommedal J. 2012. Three-dimensional acoustic full waveform inversion: algorithm and application to Valhall. *74th Annual EAGE Conference & Exhibition, Copenhagen Denmark, Expanded Abstracts*.
- Fichtner A. and Trampert J. 2011. Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International* 185(2), 775–798.
- Gao F., Levander A.R., Pratt R.G., Zelt C.A. and Fradelizio G.L. 2006. Waveform tomography at a groundwater contamination site: surface reflection data. *Geophysics* 72(5), G45–G55.
- Gauthier O., Virieux J. and Tarantola A. 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results. *Geophysics* 51(7), 1387–1403.
- Gholami Y., Brossier R., Operto S., Prioux V., Ribodetti A. and Virieux J. 2013. Which parametrization is suitable for acoustic VTI full waveform inversion? Part 2: application to Valhall. *Geophysics* 78(2), R107–R124.
- Gratton S., Laloyaux P., Sartenaer A. and Tshimanga J. 2011. A reduced and limited-memory preconditioned approach for the 4d-var data-assimilation problem. *Quarterly Journal of the Royal Meteorological Society* 137, 452–466.
- Hustedt B., Operto S. and Virieux J. 2004. Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling. *Geophysical Journal International* 157, 1269–1296.
- Kaltenbacher B., Neubauer A. and Scherzer O. 2008. *Iterative Regularization Methods for Nonlinear Problems*. de Gruyter.
- Krebs J., Anderson J., Hinkley D., Neelamani R., Lee S., Baumstein A. and Lacasse M.D. 2009. Fast full-wavefield seismic inversion using encoded sources. *Geophysics* 74(6), WCC105–WCC116.
- Lailly P. 1983. The seismic inverse problem as a sequence of before stack migrations. in: *Conference on Inverse Scattering, Theory and Application*, Society for Industrial and Applied Mathematics, Philadelphia (ed. R. Bednar and Weglein), pp. 206–220.
- Lions J.L. 1968. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*. Dunod.
- Ma Y. and Hale D. 2012. Quasi-newton full-waveform inversion with a projected Hessian matrix. *Geophysics* 77(5), R207–R216.
- Métivier L. 2011. Interlocked optimization and fast gradient algorithm for a seismic inverse problem. *Journal of Computational Physics* 230(19), 7502–7518.
- Métivier L., Brossier R., Virieux J. and Operto S. 2012. Optimization schemes in FWI: the truncated Newton method. in: *2012 SEG Abstracts*.

- Métivier L., Brossier R., Virieux J. and Operto S. 2013. Full waveform inversion and the truncated newton method. *SIAM Journal On Scientific Computing* 35(2), B401–B437.
- Nash S.G. 2000. A survey of truncated Newton methods. *Journal of Computational and Applied Mathematics* 124, 45–59.
- Nocedal J. and Wright S.J. 2006. *Numerical Optimization* Springer, 2nd edn.
- Operto S., Ravaut C., Improta L., Virieux J., Herrero A. and Dell'Aversana P. 2004. Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltime and waveform inversions: a case study. *Geophysical Prospecting* 52, 625–651.
- Operto S., Virieux J. and Dessa J.X. 2005. High-resolution crustal seismic imaging from OBS data by full-waveform inversion: application to the eastern-Nankai trough. in: *EOS Trans. AGU* vol. 86. American Geophysical Union.
- Operto S., Virieux J., Dessa J.X. and Pascal G. 2006. Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough. *Journal of Geophysical Research* 111(B09306), doi:10.1029/2005JB003835.
- Operto S., Virieux J., Ribodetti A. and Anderson J.E. 2009. Finite-difference frequency-domain modeling of visco-acoustic wave propagation in two-dimensional TTI media. *Geophysics* 74 (5), T75–T95.
- Operto S., Brossier R., Gholami Y., Métivier L., Prieux V., Ribodetti A. and Virieux J. 2013. A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice. *The Leading Edge*, September (Special section Full Waveform Inversion), 1040–1054.
- 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(2), 495–503.
- Plessix R.E. and Perkins C. 2010. Full waveform inversion of a deep water ocean bottom seismometer dataset. *First Break* 28, 71–78.
- Plessix R.-E., Baeten G., de Maag J.W. and ten Kroode F. 2012. Full waveform inversion and distance separated simultaneous sweeping: a study with a land seismic data set. *Geophysical Prospecting* 60, 733–747.
- Pratt R.G. 1990. Inverse theory applied to multi-source cross-hole tomography. part II : elastic wave-equation method. *Geophysical Prospecting* 38, 311–330.
- Pratt R.G. and Worthington M.H. 1990. Inverse theory applied to multi-source cross-hole tomography. Part I: acoustic wave-equation method. *Geophysical Prospecting* 38, 287–310.
- Pratt R.G., Shin C. and Hicks G.J. 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International* 133, 341–362.
- Prieux V., Brossier R., Gholami Y., Operto S., Virieux J., Barkved O.I. and Kommedal J.H. 2011. On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study. *Geophysical Journal International* 187, 1495–1515.
- Prieux V., Brossier R., Operto S. and Virieux J. 2013. Multiparameter full waveform inversion of multicomponent OBC data from valhall. Part 1: imaging compressional wavespeed, density and attenuation. *Geophysical Journal International* 194(3), 1640–1664.
- Prieux V., Brossier R., Operto S. and Virieux, J. 2013. Multiparameter full waveform inversion of multicomponent OBC data from valhall. Part 2: imaging compressional and shear-wave velocities. *Geophysical Journal International* 194(3), 1665–1681.
- Ravaut C., Operto S., Improta L., Virieux J., Herrero A. and dell'Aversana P. 2004. Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt. *Geophysical Journal International* 159, 1032–1056.
- Saad Y. 2003. *Iterative Methods for Sparse Linear Systems*. SIAM,
- Shin C., Jang S. and Min D.J. 2001. Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting* 49, 592–606.
- Sirgue L., Etgen J.T. and Albertin U. 2008. 3D Frequency Domain Waveform Inversion using Time Domain Finite Difference Methods. *70th EAGE, Conference and Exhibition, Roma, Italy*, Expanded Abstracts, F022.
- Sirgue L., Barkved O.I., Dellinger J., Etgen J., Albertin U. and Kommedal J.H. 2010. Full waveform inversion: the next leap forward in imaging at Valhall. *First Break* 28, 65–70.
- Tarantola A. 1984. Inversion of seismic reflection data in the acoustic approximation. *Geophysics* 49(8), 1259–1266.
- Tarantola A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.
- Vigh D., Starr B., Kapoor J. and Li H. 2010. 3d full waveform inversion on a gulf of mexico waz data set. *SEG Technical Program Expanded Abstracts* 29(1), 957–961.
- Virieux J. and Operto S. 2009. An overview of full waveform inversion in exploration geophysics. *Geophysics* 74(6), WCC1–WCC26.
- Wang Z., Navon I.M., Dimet F.X. and Zou X. 1992. The second order adjoint analysis: Theory and applications. *Meteorology and Atmospheric Physics* 50(1-3), 3–20.