Receiver-extension strategy for time-domain full waveform inversion using a relocalization approach

Ludovic Métivier^{1,2}, Romain Brossier²

¹CNRS, Univ. Grenoble Alpes, LJK, F-38000 Grenoble, France

²Univ. Grenoble Alpes, ISTerre, F-38000 Grenoble, France

(August 26, 2021)

Running head: **L. Métivier & R. Brossier**

ABSTRACT

A receiver-extension strategy is presented as an alternative to recently promoted source-extension strategies, in the framework of high resolution seismic imaging by full waveform inversion. This receiver-extension strategy is directly applicable in time-domain full waveform inversion, and unlike source-extension methods it incurs negligible extra computational cost. After connections between difference source-extension strategies are reviewed, the receiver-extension method is introduced and analyzed for single-arrival data. The method results in a misfit function convex with respect to the velocity model in this context. The method is then applied to three exploration scale synthetic case studies representative of different geological environment, based on: the Marmousi model, the BP 2004 salt model, and the Valhall model. In all three cases the receiver-extension strategy makes it possible to start full waveform inversion with crude initial models, and reconstruct meaningful subsurface velocity models. The good performance of the method even considering inaccurate amplitude prediction due to noise, imperfect modeling, and source wavelet estimation, bodes well for

₂₁ field data applications.

INTRODUCTION

22 Full waveform inversion (FWI) is a high resolution seismic imaging strategy. At the core of
23 the method is a partial-differential-equations (PDE) constrained optimization problem, which
24 is solved by iteratively reducing a misfit between calculated and observed data, as initially
25 introduced by Lailly (1983) and Tarantola (1984). Continuous progress in the understanding
26 of this geophysical imaging problem, as well as the design of wide azimuth/wide offset seismic
27 acquisition systems and the development of high performance computing platforms have led
28 to the current success of FWI. It is now routinely applied in the industry for exploration
29 scale targets (Sirgue et al., 2010; Plessix and Perkins, 2010; Warner et al., 2013; Vigh et al.,
30 2014; Operto et al., 2015; Raknes et al., 2015; Solano and Plessix, 2019), and in academia
31 for crustal, regional and global scale imaging, yielding unprecedented high resolution 3D
32 reconstruction of subsurface mechanical parameters (Fichtner and Villaseñor, 2015; Bozdağ
33 et al., 2016; Górszczyk et al., 2017; Beller et al., 2018; Lei et al., 2020; Lu et al., 2020). A
34 recent overview of FWI and its applications can be found in Virieux et al. (2017).

35 Despite this success, challenges remain for a wide and more automated application
36 of FWI, especially at the exploration scale. The main reason is the absence of sufficient
37 low frequency content in exploration data, yielding the well known cycle skipping problem
38 (Virieux and Operto, 2009). From a mathematical perspective, cycle skipping is due to the
39 non-convexity of the misfit function which is iteratively minimized. As FWI relies on local
40 optimization techniques, the presence of local minima in the misfit function is harmful: if
41 the starting model is not in the basin of attraction of the global minimum, the method
42 converges to a possibly non-informative local minimum.

43 In practice, this issue is overcome through the careful design of data-based hierarchical

44 schemes. The main ingredient is a multi-scale approach, leading to the interpretation of the
 45 data from low to high frequency (Bunks et al., 1995). Interpreting the low frequency content
 46 first reduces the number of phases in the data and thus enlarges the basin of attraction of
 47 the global minimum. This strategy is usually complemented with time-windowing and offset
 48 selection strategies, to foster the interpretation of specific arrivals, such as diving waves, to
 49 constrain a specific part of the medium and again reduce the risk of cycle skipping (Shipp
 50 and Singh, 2002; Wang and Rao, 2009; Brossier et al., 2009). This complex design requires
 51 human expertise, can be time-consuming, and can also question the robustness of the results
 52 while increasing the uncertainty attached to them. What is the sensitivity of the inversion
 53 to the different choices made to design the workflow?

54 For this reason, research efforts are still dedicated to the design of more robust and
 55 efficient full waveform inversion schemes. To give an overview of this research field, it
 56 is convenient to split the proposed methods into two categories. In the first group, the
 57 focus is on the misfit measurement. Alternative misfit functions are proposed, with a
 58 desired improved convexity with respect to time-shifts, seen as a good proxy for convexity
 59 with respect to velocities (Jannane et al., 1989). Cross-correlation, (Luo and Schuster,
 60 1991; van Leeuwen and Herrmann, 2013), deconvolution (Luo and Sava, 2011; Warner
 61 and Guasch, 2016), normalized integration (Donno et al., 2013), instantaneous envelope
 62 and phase (Fichtner et al., 2008; Bozdağ et al., 2011; Wu et al., 2014), optimal transport
 63 (Engquist and Froese, 2014; Métivier et al., 2016; Yang et al., 2018b; Métivier et al., 2019)
 64 are instances of the many methods which have been investigated in this frame. Some of
 65 these methods have been applied only on synthetic cases, while other have shown interesting
 66 properties in the frame of 3D field data applications. A common feature of all these methods
 67 is the presence of tuning parameters which might be sometimes difficult to control.

68 It is not our purpose here to elaborate on this first group. We focus instead on the second
69 group, which could be labeled as “extension strategies”. The fundamental idea is slightly
70 different. The non-convexity of least-squares FWI is linked to the increased nonlinearity of
71 the inverse problem with respect to the model parameters induced by the reduced space
72 approach used to solve the PDE-constrained optimization problem. To overcome this
73 difficulty, artificial degrees of freedom are injected in the problem, which shall be gradually
74 eliminated along the convergence path to recover a physical solution. These degrees of
75 freedom help to fit the data in the early iterations when the model estimate is poor.

76 When these extension methods are model based, they are generally known as migration
77 velocity analysis (MVA) methods. A quite complete overview of these techniques is proposed
78 in Symes (2008). Based on the scale separation assumption (subsurface parameters split in
79 a smooth background and a sharp reflectivity model), the artificial degrees of freedom are
80 introduced at the reflectivity level. The FWI problem is reformulated as the focusing of the
81 extended reflectivity model at zero time-lag or zero subsurface offset or alternatively as the
82 flattening of the extended reflectivity in the offset or angle direction. Mathematical analysis
83 shows that in a transmission regime, under specific mathematical conditions which can be
84 related to the absence of triplication the resulting problem asymptotically converges to a
85 travel-time tomography problem, known to be convex (Symes, 2014).

86 More recently, a class of source extension strategies has emerged, named matched source
87 waveform inversion (MSWI) (Huang et al., 2018a,b, 2019). Preliminary concepts on source
88 extension had already been proposed in Almomin (2016). In this approach, the artificial
89 degrees of freedom are introduced at the source instead of being introduced at the model
90 level. This overcomes a series of limitations encountered by MVA approaches. In practice,
91 the high computational cost for building extended reflectivity hypercubes makes it difficult

92 to apply MVA to 3D field data. A more fundamental difficulty is related to complex data
 93 with multi-arrival and multiple reflections (Cocher et al., 2017). As will be detailed further
 94 in this study, MSWI is equivalent to wavefield reconstruction inversion techniques (WRI),
 95 another class of methods previously introduced to relax cycle skipping in FWI van Leeuwen
 96 and Herrmann (2013, 2016); Aghamiry et al. (2019b). Note that the deconvolution approach
 97 introduced as adaptive waveform inversion (AWI) by Warner and Guasch (2016) can also
 98 be recast as a MSWI technique. This shows that the distinction between misfit function
 99 modification methods and extension strategies is not as clear as one could think, however it
 100 is convenient to draw a landscape of the numerous investigations performed in this field.

101 MSWI techniques have shown promising results on 2D synthetic applications in the frame
 102 of frequency-domain FWI. Theoretical results for a 1D transmission canonical case also
 103 show that, depending on the chosen formulation and particularly the choice of annihilator
 104 operator, MSWI can yield a convex misfit function. The use of the variable projection
 105 method to solve the extended inversion problem, detailed in the next subsection, seems also
 106 key to the success of such strategies (Symes et al., 2020). However, their implementation
 107 in the frame of time-domain FWI is still under development. Such an implementation is
 108 required to handle 3D field data applications. Indeed, frequency-domain FWI is for now
 109 limited to moderate size targets. This is due to the lack of scalability of the direct solvers
 110 on which they rely to solve harmonic equations (see Li et al., 2020, for a recent status on
 111 the capabilities of direct solvers to solve large scale harmonic wave equation problems). The
 112 reason why time-domain MSWI techniques are difficult to design is detailed in this study. In
 113 essence, MSWI requires the solution of a square wave propagation problem, which is possible
 114 in the harmonic case when a factorization of the wave propagation operator is available, but
 115 which is much more difficult to solve in the time-domain case through explicit time-stepping

116 algorithms.

117 This intrinsic difficulty for MSWI methods to be applied in the time-domain is the
118 motivation of this study. We propose here an alternative extension strategy, based on the
119 receivers rather than the source. We propose to introduce the receiver location as the
120 artificial degree of freedom in the inversion. As will be shown, this avoids the introduction
121 of a square wave propagation operator and thus makes this method applicable directly in
122 the time-domain at a reasonable computational cost. In addition, introducing the receiver
123 position as a new unknown makes possible to mitigate cycle skipping. The kinematic
124 mismatch is compensated by the repositioning of the receivers which is slowly relaxed to
125 the true receiver position. After presenting the method on a schematic cross-hole example,
126 we illustrate how our algorithm works on 2D synthetic (visco-)acoustic examples based on
127 the Marmousi, BP 2004 and 2D Valhall synthetic models. In all three cases, our receiver
128 relocalization strategy makes possible to start FWI with crude initial models, outperforming
129 standard least-squares based inversion.

130 The structure of the study is as follows. First, we give an overview of the theory behind
131 MSWI methods. Then, we introduce our receiver extension strategy. We illustrate the
132 fundamental properties of the algorithm on a schematic transmission case. We then present
133 the application of our algorithm to three synthetic benchmark models. We propose finally a
134 discussion, after what we conclude and we give some opening perspectives.

BACKGROUND AND STATE OF THE ART ON MSWI AND WRI

METHODS

FWI as a PDE-constrained optimization problem

FWI can be cast as the following PDE-constrained optimization problem

$$\min_m \frac{1}{2} \sum_{s=1}^{N_s} \|Ru_s - d_{obs,s}\|_{\mathcal{D}}^2, \quad s.t. \quad A(m)u_s = b_s, \quad s = 1, \dots, N_s, \quad (1)$$

where m denotes the subsurface model parameters which are to be reconstructed, $N_s \in \mathbb{N}$ is the number of source positions used to generate the data, $d_{obs,s}$ is the s -th shot gather, $A(m)$ is a general wave equation operator (from acoustic to visco-elastic), $u_s[m]$ is the synthetic wavefield solution of the wave equation for the s -th source position, $b_s(t)$ is the source term of the s -th wavefield, and R is a restriction operator mapping the wavefield u_s to the receivers location. Here and in the following, $\|\cdot\|_{\mathcal{D}}$ will refer to the following L^2 norm in the data space: for a shot gather d , we will have

$$\|d\|_{\mathcal{D}}^2 = \sum_{r=1}^{N_r} \int_0^T |d(x_r, t)|^2 dt, \quad (2)$$

where N_r corresponds to the number of receivers and x_r denotes the receiver positions.

The Lagrangian operator associated with this PDE-constrained optimization problem is

$$L(m, u, \lambda) = \frac{1}{2} \sum_{s=1}^{N_s} \|Ru_s - d_{obs,s}\|_{\mathcal{D}}^2 + \sum_{s=1}^{N_s} \langle \lambda_s, A(m)u_s - b_s \rangle_{\mathcal{W}} \quad (3)$$

where $u = (u_1, \dots, u_{N_s})$ gathers the N_s synthetic wavefields, $\lambda = (\lambda_1, \dots, \lambda_{N_s})$ gathers the N_s adjoint wavefields, and $\langle \cdot, \cdot \rangle_{\mathcal{W}}$ is the Euclidean scalar product in the wavefield space. For

148 two wavefields u, v we have

$$\langle u, v \rangle_{\mathcal{W}} = \int_0^T \int_{\Omega} u(x, t) v(x, t) dx dt, \quad (4)$$

149 where Ω represents the subsurface.

150 Finding a solution to the PDE-constrained optimization problem 1 is equivalent to find
 151 a saddle point of the Lagrangian operator by solving the min max problem

$$\min_{u, m} \max_{\lambda} L(m, u, \lambda). \quad (5)$$

152 However, the computational cost for solving the problem 5 through local optimization is
 153 prohibitive: aside the convergence rate, it would imply all incident and adjoint wavefields in
 154 space and time, which is not affordable for realistic size FWI application. The reduced space
 155 approach is thus conventionally used. The problem 5 is transformed into the unconstrained
 156 optimization problem

$$\min_m \frac{1}{2} \sum_{s=1}^{N_s} \|RA(m)^{-1} b_s - d_{obs, s}\|_{\mathcal{D}}^2. \quad (6)$$

157 This conventional form for FWI is known to exhibit local minima into which local optimization
 158 solvers can converge. Compared with the problem 5, the nonlinearity with respect to the
 159 model parameter becomes more apparent in the term $RA(m)^{-1} b_s$, which corresponds to the
 160 solution of the wave equation for a given model parameter m .

161 WRI and MSWI formalism

162 As noted by van Leeuwen and Herrmann (2013), the problem 5 is only “mildly” nonlinear.
 163 Indeed, the Lagrangian $L(m, u, \lambda)$ depends linearly on λ . In addition, because of the

164 bilinearity of the wave equation operator it also depends linearly on m and quadratically on
 165 u . We express this bilinearity by introducing the operator $F(m, u)$

$$F(m, u) = A(m)u, \quad (7)$$

166 and the identity

$$F(m, u) = A(m)u = B(u)m. \quad (8)$$

167 This identity shows that the wave propagation problem can be rewritten equally as a linear
 168 operator $A(m)$ acting on u or a linear operator $B(u)$ acting on m . This identity is useful in
 169 the following developments. This property is true for general elastic and visco-elastic wave
 170 propagation, up to the choice of the parameterization for m , as is discussed in Aghamiry
 171 et al. (2019a).

172 This apparent “well behaved” property motivates the design of WRI (van Leeuwen and
 173 Herrmann, 2013). With the idea to make the nonlinearity with respect to m less stringent,
 174 they propose to reformulate the FWI problem using a quadratic penalty method instead
 175 of using the reduced space approach (Nocedal and Wright, 2006). This method, coined as
 176 wavefield reconstruction inversion (WRI), is expressed as

$$\min_{m, u} \frac{1}{2} \sum_{s=1}^{N_s} \|Ru_s - d_{obs, s}\|_{\mathcal{D}}^2 + \eta \sum_{s=1}^{N_s} \|F(m, u_s) - b_s\|_{\mathcal{W}}^2. \quad (9)$$

177 where $\|\cdot\|_{\mathcal{W}}$ is the Euclidean norm associated with the scalar product $\langle \cdot, \cdot \rangle_{\mathcal{W}}$. The wave
 178 equation is not imposed as a strict constraint, instead it should be fitted in the least-squares
 179 sense. This reformulation implies a change of paradigm: from a parameter estimation
 180 problem posed on m only (reduced space approach), FWI becomes a compatibility problem

181 where both the wavefield u and the model parameter m are reconstructed from partial
182 observations d_{obs} and a priori knowledge of the physics of wave propagation (the operator
183 $A(m)$). In this frame, solving exactly for the wave equation to compute u at each iteration
184 while the model m is known to be only poorly approximated does not appear as a good
185 choice, hence the freedom added on the reconstruction of u . The level of accuracy for the
186 wavefield to satisfy the wave equation is controlled with the penalty parameter η .

187 Later on, Aghamiry et al. (2019b) have proposed an improvement of the WRI strategy
188 where the FWI problem is reformulated following an augmented Lagrangian approach, which
189 presents several advantages over the quadratic penalty method regarding convergence rate
190 issues and selection of the parameter η (Nocedal and Wright, 2006). The Iteratively-Refined
191 Wavefield Reconstruction Inversion (IR-WRI) is formulated as

$$\min_{m,u} \max_{\lambda} = \frac{1}{2} \sum_{s=1}^{N_s} \|Ru_s - d_{obs,s}\|_{\mathcal{D}}^2 + \sum_{s=1}^{N_s} \langle \lambda_s, F(m, u_s) - b_s \rangle_{\mathcal{W}} + \eta \sum_{s=1}^{N_s} \|F(m, u_s) - b_s\|_{\mathcal{W}}^2. \quad (10)$$

192 that is the standard Lagrangian augmented with the quadratic penalty term.

193 Please note however that there is no formal guarantee of the existence of a unique
194 solution to the problem 5. Such a proof would require the operator $F(m, u) = A(m)u$ to be
195 convex which is not the case (bilinearity does not imply convexity). A recent mathematical
196 analysis of WRI also shows that WRI asymptotically tends to standard FWI in the context
197 of pure 1D acoustic transmission and suffers from the same non-convexity problems in this
198 case (Symes, 2020).

199 In parallel, Huang et al. (2018a,b) have proposed a matched source waveform inversion
200 (MSWI) method. MSWI relies on an extended modeling operator making use of an extended
201 source. In Huang et al. (2018a) this extension is proposed in space and time while in Huang

et al. (2018b) the extension is performed only in space, with the time signature of the source supposed to be known *a priori* and treated by deconvolution. In the general case of space and time extension, the extended source can be denoted by $\tilde{b}(x, t) = (\tilde{b}_1, \dots, \tilde{b}_{N_s})$. MSWI is then formulated as

$$\min_{m, \tilde{b}} = \frac{1}{2} \sum_{s=1}^{N_s} \|S(m)\tilde{b}_s - d_{obs,s}\|_{\mathcal{D}}^2 + \eta \sum_{s=1}^{N_s} \|\tilde{b}_s - b_s\|_{\mathcal{W}}^2, \quad (11)$$

where $S(m) = RA(m)^{-1}$ is the forward problem operator.

The philosophy of MSWI relies on the frame of extended inversion. Unphysical degrees of freedom are added to the modeling operator to help fit the data. In the case of MSWI the source is not punctual in space, and possibly the time signature becomes also an unknown. An annihilator is added to the misfit function to constrain the additional degrees of freedom towards physical values at convergence. In the case of MSWI, the extended source shall be localized on the correct source location with the correct time signature at convergence. For simplicity we restrict this annihilator here as the least-squares misfit but more general annihilator can be used (Huang et al., 2018a,b).

Interestingly, as noted by Wang et al. (2016) and Huang et al. (2018a), the change of variables $\tilde{b}_s = F(m, u_s)$ yields

$$S(m)\tilde{b}_s = RA(m)^{-1}F(m, u_s) = RA(m)^{-1}A(m)u_s = Ru_s. \quad (12)$$

Using this identity, we see that MSWI with a least-squares annihilator is equivalent to WRI. The difference between MSWI and WRI relies on the choice of unknown: $\tilde{b}$ for MSWI, the source wavefield u for WRI.

220 Numerical solution and limitation for time-domain applications

221 We now explain the origin of the limitations of WRI, IR-WRI and MSWI when considering
 222 time-domain inversion. All three approaches rely on the minimization of a misfit function
 223 which depends on two parameters: the model parameter m and an additional parameter
 224 (wavefield u or extended source $\tilde{b}$). The minimization is achieved by defining an outer
 225 minimization loop over the model parameter m and an inner loop on the additional parameter.
 226 This method is often referred to as variable projection approach (Golub and Pereyra, 2003).
 227 We recall it formally as it will be used throughout the paper.

228 *Nested loop optimization*

229 Consider the joint problem

$$\min_{x_1, x_2} f(x_1, x_2). \quad (13)$$

230 Assuming f is twice differentiable with respect to x_1 and x_2 , the problem 13 is equivalent to

$$\min_{x_1} g(x_1), \quad (14)$$

231 where

$$g(x_1) = f(x_1, \overline{x_2}(x_1)), \quad \overline{x_2}(x_1) = \arg \min_{x_2} f(x_1, x_2). \quad (15)$$

232 The outer loop is the minimization of $g(x_1)$ and the computation of $\overline{x_2}(x_1)$ is the inner loop.
 233 This method is interesting in practice when the computation cost of the inner minimization
 234 over x_2 is cheap *i.e.* a quadratic problem with a closed form formula is solved. Gradient-
 235 based or quasi-Newton methods are then conventionally used to minimize $g(x_1)$ in the outer

236 loop. Interestingly, the gradient of $g(x_1)$ is given by

$$\nabla g(x_1) = \frac{\partial f}{\partial x_1}(x_1, \overline{x_2}(x_1)) + \frac{\partial f}{\partial x_2} \frac{\partial \overline{x_2}}{\partial x_1}, \quad (16)$$

237 however because of the definition of $\overline{x_2}(x_1)$ as a minimizer of $f(x_1, x_2)$ with respect to x_2

238 the second term in the right hand side vanishes and we have

$$\nabla g(x_1) = \frac{\partial f}{\partial x_1}(x_1, \overline{x_2}(x_1)). \quad (17)$$

239 This last equation shows that to compute the gradient of $g(x_1)$, one has only to solve the

240 inner problem for x_2 and inject the solution in the gradient formula for $g(x_1)$.

241 *WRI and IR-WRI*

242 In van Leeuwen and Herrmann (2013), the nested loop optimization is employed with

$$x_1 = m, \quad x_2 = u. \quad (18)$$

243 The inner loop corresponds to the reconstruction of the wavefield u , by solving the problems

$$\min_{u_s} \frac{1}{2} \|Ru_s - d_{obs,s}\|_{\mathcal{D}}^2 + \eta \|F(m, u_s) - b_s\|_{\mathcal{W}}^2, \quad s = 1, \dots, N_s. \quad (19)$$

244 Thanks to the bilinearity of the wave propagation operator, this problem is quadratic and a

245 closed-form formula for u_s exists:

$$(\eta A(m)^T A(m) + R^T R) u_s = R^T d_{obs,s} + \eta A(m)^T b_s \quad (20)$$

246 Interestingly, the bilinearity of $F(m, u)$ makes also the outer minimization problem quadratic
 247 with respect to m , making possible to use a Newton method to solve the outer loop in a
 248 single step.

$$\left(\sum_{s=1}^{N_s} B(u_s)^T B(u_s) \right) m = \sum_{s=1}^{N_s} B(u_s)^T b_s \quad (21)$$

249 Another level of iteration further consists in reducing the weight η step by step. This
 250 iterative reduction of the weight can be difficult to adjust for practical applications.

251 IR-WRI circumvents this difficulty. It relies on a more sophisticated optimization scheme
 252 (ADMM method, see Boyd and Vandenberghe (2004); Combettes and Pesquet (2011) for
 253 instance), where such reduction of the weight “by hand” is not required. However, the core
 254 of the iteration is based on the same alternate reconstruction of the wavefield and the model.
 255 The same equations are solved, only with different right-hand-sides. For more details, the
 256 reader is referred to Aghamiry et al. (2019b).

257 *MSWI*

258 MSWI also relies on an alternate reconstruction between the extended source and the model
 259 parameters, with this time

$$x_1 = m, \quad x_2 = \tilde{b} \quad (22)$$

260 As for WRI, the inner loop on $\tilde{b}$ is equivalent to the following quadratic problems

$$\min_{\tilde{b}_s} \frac{1}{2} \|S(m)\tilde{b}_s - d_{obs,s}\|_{\mathcal{D}}^2 + \eta \|\tilde{b}_s - b_s\|_{\mathcal{W}}^2, \quad s = 1, \dots, N_s. \quad (23)$$

261 Therefore, closed-form formula exist for $\tilde{b}_s$ such that

$$\tilde{b}_s = [S(m)^T S(m) + \eta I]^{-1} (S(m)^T d_{obs,s} + b_s), \quad s = 1, \dots, N_s. \quad (24)$$

262 Unlike WRI, the outer minimization problem is not quadratic with respect to m , therefore
 263 it should rely on a gradient-based algorithm (*i.e.* quasi-Newton methods). The gradient
 264 of the outer loop is computed following the adjoint state strategy (Plessix, 2006), as for
 265 conventional FWI. It is built as the correlation between incident and adjoint fields, where
 266 the adjoint is the backpropagation of the residuals at the receiver location. The difference is
 267 that the incident field and the residuals are computed using the extended source $\tilde{b}$.

268 *Extension to time-domain FWI*

269 It can be shown that the operator $B(u_s)^T B(u_s)$ in equation 21 is diagonal for the acoustic
 270 wave equation (van Leeuwen and Herrmann, 2013; Aghamiry et al., 2019b), and block
 271 diagonal for general elastodynamics equations. In time-domain, each element of the diagonal
 272 blocks is accumulated by summation in time. The system in equation 21 therefore does not
 273 present particular difficulties for time-domain formulation.

274 However, this is not the case for the system in equation 20. The latter implies the operator
 275 $\eta A(m)^T A(m) + R^T R$. In the frequency-domain, $A(m)$ is a matrix after spatial discretization.
 276 It can be decomposed as a LU product and the system in equation 20 can be easily solved.
 277 In the time-domain, such technique is not available and solving the corresponding system is
 278 a real challenge. The difficulty actually comes from the component $R^T R$ in the operator
 279 which makes impossible the use of explicit time-domains schemes required for time-domain
 280 FWI. Neglecting $R^T R$ indeed yields the operator $A(m)^T A(m)$ which can be solved in two

281 steps through explicit time-domain schemes. Consider for instance, for a given right hand
 282 side z ,

$$A^T A u = z \quad (25)$$

283 This can be solved using

$$A^T y = z, \quad y = A u \quad (26)$$

284 The computation of y would require the solution of the adjoint wave equation with the
 285 right-hand-side z , and the computation of u the solution of the wave equation with the
 286 right-hand-side y . However, neglecting $R^T R$ amounts to an infinite weight η which goes
 287 back to solving the wave equation with infinite accuracy, *i.e.* the reduced space approach.

288 The same problem arises for MSWI. The reconstruction of the extended source implies
 289 the operator $S(m)^T S(m) + \eta I$. For the same reason mentioned above, this operator cannot
 290 be solved straightforwardly using explicit time-domain schemes because of the term ηI .
 291 Without it, the operator $S(m)^T S(m)$ can be solved through explicit time schemes in two
 292 steps, as in the WRI case. Circumventing this difficulty could imply giving an infinite weight
 293 to η : in this case MSWI also comes back to the reduced space approach as the extended
 294 source needs to conform with infinite accuracy to the true source b_s . Another option would
 295 be to make η tends to 0. However, this implies no regularization term in the MSWI problem
 296 11, which is known to be an ill-posed problem because of the ambiguity between extended
 297 sources $\tilde{b}$ and the model parameter m (Huang et al., 2018a).

298 Recent work proposed by Aghamiry et al. (2020) in the frame of WRI shows that
 299 an accurate reconstruction of the time-domain wavefields is however possible following a
 300 sophisticated backward-forward recursion where each iteration requires the solution of a wave
 301 propagation problem. The number of required iterations should be larger at the beginning of

302 the inversion. However preliminary result shows a computational extra cost approximately
303 8 times the cost of a gradient in the early stages of the inversion, which questions the
304 feasibility of this strategy for field data application. The study by Aghamiry et al. (2020)
305 also shows that the time-domain extension proposed in Wang et al. (2016) relies on a very
306 crude approximation of the wavefield reconstruction step. Hence, the interest of the WRI
307 approach tends to be lost following this time-domain approximation.

308 The difficulty of applying WRI, IR-WRI or MSWI in the frame of time-domain FWI has
309 prompted us to investigate the alternative approach based on a receiver extension strategy
310 we present in the next Section.

RECEIVER EXTENSION STRATEGY

311 Theory

312 In the same spirit as extended method strategies, we add an artificial degree of freedom to
 313 help fit the data when the subsurface parameter m is too far from the exact model. The
 314 difference is that this artificial degree of freedom is introduced at the receiver level, instead
 315 of being introduced at the source level.

316 The degree of freedom we introduce is the receiver position. As illustrated in the sequel,
 317 moving the receiver away from its true position can compensate for kinematic mismatch
 318 due to wrong subsurface model m . Formally, denote by x_r , $r = 1, \dots, N_r$ the N_r receiver
 319 positions. Denote by $\Delta x_s \in \mathbb{R}^{N_r}$ a vector of N_r receiver corrections for receiver associated
 320 with source s , and $\Delta x = [\Delta x_1, \dots, \Delta x_{N_s}] \in \mathbb{R}^{N_r \times N_s}$ the vector gathering the receiver
 321 position correction for each source/receiver pair. The receiver extension strategy consists in
 322 solving the problem

$$\min_{m, \Delta x} f(m, \Delta x) = \frac{1}{2} \sum_{s=1}^{N_s} \|R(\Delta x_s)A(m)^{-1}b_s - d_{obs,s}\|_{\mathcal{D}}^2 + \frac{1}{2} \|\Delta x\|_{\eta}^2, \quad (27)$$

323 where $\|\cdot\|_{\eta}$ is a weighted least-squares norm

$$\|\Delta x\|_{\eta}^2 = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \eta_{s,r} \Delta x_{s,r}^2, \quad (28)$$

324 with $\eta \in \mathbb{R}^{N_r \times N_s}$ a vector of weights $\eta_{s,r}$ (one per source/receiver couple), and $R(\Delta x_s)$ an
 325 extraction operator returning the values of the wavefield at the corrected receiver position

326 $x_r + \Delta x_{s,r}$, following the convolution

$$R(\Delta x_s)u = \int_{\Omega} u(x, t) \delta(x - (x_r + \Delta x_{s,r})) dx. \quad (29)$$

327 The second term in the right-hand-side of equation 27 is a least-squares annihilator, specifying
 328 that the receiver position correction should not be too large and converge to 0 for the correct
 329 model m . Note how close problem 27 is from reduced space problem 6. The only difference
 330 is in the receiver position correction introduction as a variable and the annihilator terms
 331 associated with this correction.

332 Numerical solution and implementation

333 *Inner loop*

334 To solve the problem 27, we use the nested optimization approach (equations 13 to 17)
 335 shared by WRI, IR-WRI, and MSWI techniques, with

$$x_1 = m, \quad x_2 = \Delta x. \quad (30)$$

336 The inner loop problem thus consists in determining the receiver position correction for a
 337 given model m . We denote it by $\overline{\Delta x}(m)$. The key point for an efficient implementation is a
 338 fast solution of this inner problem. When using WRI, IR-WRI, or MSWI techniques, the
 339 inner problem is quadratic: it has a unique solution given by a closed-form formula. Using
 340 the receiver-extension strategy, the inner problem is highly non-linear. Thus, there is no
 341 closed-form formula for $\overline{\Delta x}(m)$. In addition, the associated misfit function presents local
 342 minima, condemning the use of local optimization methods. However, for the nested loop

343 optimization to be efficient, we need a fast and accurate solver for the solution of the inner
 344 problem.

345 Here, it is important to realize that, thanks to the use of L^2 norm both for data misfit
 346 and annihilator terms, the inner problem is separable for all source/receiver couples. It
 347 means the objective function in 27 can be decomposed as a sum of misfit function depending
 348 only on one source/receiver couple. Mathematically, we have

$$f(m, \Delta x) = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} f_{s,r}(m, \Delta x_{s,r}) \quad (31)$$

349 where

$$f_{s,r}(m, \Delta x_{s,r}) = \frac{1}{2} \int_0^T |u_s[m](x_r + \Delta x_{s,r}, t) - d_{obs,s}(x_r, t)|^2 dt + \frac{\eta_{s,r}}{2} |\Delta x_{s,r}|^2, \quad (32)$$

350 where $u_s[m] = A(m)^{-1}b_s$.

351 Hence, the solution of the inner loop can be obtained by solving independently for each
 352 receiver correction $\Delta x_{s,r}$ the subproblem

$$\min_{\Delta x_{s,r}} f_{s,r}(m, \Delta x_{s,r}) \quad (33)$$

353 The number of unknowns for each subproblem 33 is small: maximum 2 unknowns in 2D to
 354 and 3 unknowns in 3D to specify a receiver position correction. Global optimization methods
 355 can thus be employed to determine the optimal receiver position corrections $\overline{\Delta x_{s,r}}(m)$.

356 In practice, it is even possible and/or advisable to reduce this number of unknowns to a
 357 single parameter. For instance, in the 2D case, considering a seismic trace containing a single

358 event, there is an intrinsic ambiguity in the receiver correction making possible to fit the data.
 359 This ambiguity is related to the isochrones, which are 2D curves in the 2D approximation.
 360 This means that there would be an infinity of 2D receiver corrections (vertical and horizontal
 361 repositioning) yielding an equivalent data fit. To avoid this non-uniqueness, we consider in
 362 this study only horizontal repositioning. No vertical receiver position corrections are allowed.
 363 The additional benefit of this strategy is that the global optimization problems to be solved
 364 in the inner loop are single parameter problems.

365 In terms of implementation, we rely on a brute-force grid search approach. The misfit
 366 function in equation 33 is evaluated for different values of $\Delta x_{s,r}$ within bounds defined
 367 depending on the application. The time-history of the wavefield is stored on a line in 2D
 368 (or a plane in 3D) on which the receivers are confined. As explained above we restrict the
 369 receivers to move only laterally to avoid intrinsic ambiguity related to isochrones. From this
 370 stored time-history of the wavefield, the calculated data can be extracted at various receiver
 371 position without having to solve again the wave equation. For each receiver position, the
 372 misfit function is evaluated. We select the receiver position correction which provides the
 373 minimum misfit value. As we illustrate in the following, this provides an efficient method to
 374 solve the inner problem. In our 2D examples, the additional computational cost compared
 375 with conventional FWI is negligible.

376 *Outer loop*

377 We solve the outer problem by a conventional quasi-Newton strategy. We use a preconditioned
 378 l -BFGS method in this study (Nocedal, 1980). The gradient of the outer function can be
 379 computed, as in MSWI, following the adjoint source strategy (Plessix, 2006). We denote it

380 by $\nabla g(m)$, and in condensed form it can be expressed as

$$\nabla g(m) = \sum_{s=1}^{N_s} \left\langle \frac{\partial A}{\partial m} u_s, \lambda_s \right\rangle, \quad (34)$$

381 where $\langle \cdot, \cdot \rangle$ denotes the scalar product in time domain and

$$\begin{cases} A(m)u_s = b_s, & s = 1, \dots, N_s \\ A(m)^T \lambda_s = R(\overline{\Delta x_s}(m))^T (R(\overline{\Delta x_s}(m)) u_s - d_{obs,s}), & s = 1, \dots, N_s. \end{cases} \quad (35)$$

382 The difference with the conventional reduced space approach is that the calculated data
 383 and the adjoint wavefields are computed using corrected receiver positions, both for the
 384 extraction of the wavefield values to build the calculated data with the operator $R(\overline{\Delta x_s}(m))$
 385 and the injection of the adjoint source with the operator $R(\overline{\Delta x_s}(m))^T$. Using $\overline{\Delta x_s}(m) = 0$
 386 in the previous equations yields the conventional least-squares gradient for FWI based on
 387 the reduced space approach.

388 *Weight parameters $\eta_{s,r}$*

389 The weights $\eta_{s,r}$ are computed following

$$\eta_{s,r} = \alpha \frac{\|d_{obs,s,r}\|_\infty}{L} \quad (36)$$

390 where

$$\|d_{obs,s,r}\|_\infty = \max_{t \in [0, T]} |d_{obs,s,r}(t)|, \quad (37)$$

391 while L is the maximum value we allow for $|\Delta x_{s,r}|$. The parameter α is a tuning parameter
 392 to control the constraint on the receiver position correction. The choice $\alpha = 1$ corresponds

393 to a simple dimensioning of the two terms in the misfit function (data fitting term and
394 annihilator term). In the next numerical experiments, the sensitivity of the method to the
395 choice of α is investigated. In synthetic experiments with inverse crime settings, low values
396 of α (to the order of 10^{-2}) seem to yield satisfactory results (transmission case, Marmousi
397 and BP2004 studies). When the amplitude cannot be predicted with perfect accuracy, higher
398 values of α might be better adapted (Valhall case study).

A TRANSMISSION CASE ANALYSIS

399 We consider here a canonical transmission problem. We use a 2D cross-hole configuration,
400 with two wells located 50 m apart (Fig. 1).

401 [Figure 1 about here.]

402 The source is a Ricker pulse with 250 Hz central frequency. We consider a single
403 source/receiver couple at 50 m depth in each well. The source is in the left well, the receiver
404 in the right well. We compute a reference seismic trace in a homogeneous medium at 2000
405 m.s⁻¹. We use for that a 2D constant density acoustic wave propagation model. Using
406 this reference trace, we construct the misfit function $f(m, \Delta x)$ considering homogeneous
407 velocity models m varying from 1000 m.s⁻¹ to 3000 m.s⁻¹, and receiver position correction
408 Δx varying only horizontally (following the x axis) from -37.5 m to 37.5 m. We select the
409 weight α to be equal to 1. The resulting misfit function is presented in Figure 2.

410 [Figure 2 about here.]

411 We see that the misfit function $f(m, \Delta x)$ is not convex. Its minimum is hidden in a narrow
412 valley, at position $m = 2000$ m.s⁻¹ and $\Delta x = 0$, and surrounded by large barriers. The
413 shape of the valley is driven by the shape of the Ricker function used to build the data: the
414 lower frequency used, the wider the valley of attraction is.

415 Nevertheless, if we select, for each velocity value, the minimum reached in the receiver
416 extension direction Δx , we can represent the function $g(m)$ that we aim at minimizing in
417 the outer loop. This function is presented in Figure 3 for different values of the weight α .

418 [Figure 3 about here.]

419 We see in Figure 3 that for a proper selection of the weight α , a convex function depending
 420 on the velocity m can be obtained. The choice of α influences the size of the valley of
 421 attraction toward the global minimum.

422 To better understand why the receiver extension approach can yield a convex misfit
 423 function in this simple transmission case, we present in Figure 4 the synthetic traces computed
 424 for different values of velocity before and after the relocalization, and we compare it to the
 425 reference trace.

426 [Figure 4 about here.]

427 As expected, the relocalization of the receiver corrects for the kinematic mismatch. The
 428 relocalized synthetic traces are all in phases with the reference trace. However, the relo-
 429 calization cannot compensate for the amplitude mismatch. This amplitude mismatch is
 430 related to energy conservation rules of wave propagation: the amplitude of the recorded
 431 signal depends on the rigidity of the medium in which it propagates, hence on the velocity
 432 in the simple constant density acoustic approximation we use here.

433 The result of this is that in the context of this single arrival canonical case, the misfit
 434 measured by the function $g(m)$ is related to this amplitude mismatch only. This mismatch
 435 increases with the velocity mismatch between the reference medium and the synthetic
 436 medium. Hence, the misfit function $g(m)$ is convex with respect to the velocity in this case.
 437 Note that the use of an amplitude sensitive misfit function, such as the least-square norm,
 438 to define $g(m)$, is crucial. A receiver extension approach based on a misfit function not
 439 sensitive to amplitude mismatch would not yield a convex function in this canonical case.

440 To end with this simple transmission case, we analyze the shape of the gradient using
 441 the same single source/receiver couple. We compare the conventional least-squares gradient

442 and the receiver extension gradient in two different homogeneous media: one at 1500 m.s^{-1} ,
 443 the second at 2500 m.s^{-1} . The results are presented in Figures 5 and 6. In the least-squares
 444 case, the first Fresnel zone of the two kernels is negative, while we expect a change of sign:
 445 in one case the medium is slower than the reference one, in the other case it is faster. This
 446 is a clear indication of cycle skipping: starting from the faster medium, the least-squares
 447 gradient would produce a positive update (opposite of the gradient) of the velocity within
 448 the first Fresnel zone. Converging to the correct solution would require to slow down the
 449 velocity.

450 The receiver extension approach does not suffer from such inconsistencies. The sign
 451 of the first Fresnel zone is correct in both slower and faster media. This change of sign
 452 is directly related to the corresponding adjoint source. Let us remind that in the receiver
 453 extension approach, it is computed as the difference between observed and synthetic data
 454 *after* relocalization. As can be seen in Figure 6, the difference between observed and synthetic
 455 data changes of sign, depending on the velocity is faster or slower than the reference one,
 456 for the same reason as mentioned previously (energy conservation law). This explains the
 457 differences in the two kernels. Note also that the relocalization affects the shape of the
 458 kernels. In the slow medium, the receiver is relocalized closer from the source. As the adjoint
 459 source is injected at the corrected receiver position, the size of the kernel is smaller. In the
 460 faster medium, the receiver is relocalized farther from the source, and the size of the kernel
 461 is larger.

462 [Figure 5 about here.]

463 [Figure 6 about here.]

464 This simple experiment illustrates how the receiver extension approach can handle kinematic
465 mismatch in the frame of FWI.

2D SYNTHETIC EXAMPLES

Choice of three models

We investigate the performance of the receiver relocation approach on three benchmark models: Marmousi II (Martin et al., 2006), BP 2004 (Billette and Brandsberg-Dahl, 2004) and a 2D synthetic model built from 3D inversion results of the Valhall OBC data (North Sea) (Sirgue et al., 2010; Operto et al., 2015; Amestoy et al., 2016). We choose these three models to test the method in different geological contexts. Marmousi II is a useful framework to investigate in details the ability to mitigate cycle skipping issues. BP 2004 is representative of the gulf of Mexico geology and contains salt structures known to be challenging to reconstruct for seismic imaging methods, because of the high velocity contrasts between these structures and the surrounding water. The Valhall model contains an important gas cloud in its middle part, which significantly attenuates seismic wave energy and makes it difficult to image the reservoir located below.

Common framework

The three experiments we present in this Section are performed in the 2D (visco-) acoustic approximation. They rely on our 2D/3D (visco-)acoustic time-domain finite-difference based full waveform inversion code TOYxDAC_TIME, which implements the method described in Yang et al. (2018a). All are based on the reconstruction of the P-wave velocity $v_P(x)$.

The source which is used is a Ricker wavelet centered on 5 Hz and high pass filtered to remove all energy below 2.5 Hz. In the BP 2004 case study, we use an additional low-pass filter to remove energy above 8 Hz. The corresponding wavelets and their power spectrum are presented in Figure 7.

[Figure 7 about here.]

In all cases we use the bound constraint preconditioned l-BFGS solver from the SEISCOPE toolbox (Métivier and Brossier, 2016). The preconditioner chosen is either a simple linear scaling in depth for the Marmousi II case study, or a wavefield based pseudo-Hessian preconditioner (Choi and Shin, 2008) for BP 2004 and Valhall case studies. We also use a Gaussian smoothing of the gradient, with correlation lengths associated with the estimated local wavelength

$$\lambda(x) = \frac{v_P(x)}{f_{ref}} \quad (38)$$

where f_{ref} corresponds to the central frequency of the Ricker wavelet.

In all three experiments, a free surface condition is imposed on top of the model. Perfectly matched layers (PML) (Bérenger, 1994) (for Marmousi II and BP 2004 models) or sponge layers (Cerjan et al., 1985) (for the Valhall model) are applied on the other boundaries to mimic a medium of infinite extension.

While Marmousi II and BP 2004 experiments are performed in an “inverse crime” settings, using a constant density acoustic modeling, the Valhall case study intends to mimic a more realistic framework. In this case the observed data is computed using a variable density and variable quality factor under the visco-acoustic approximation. A Gaussian noise, filtered in the frequency band of the data, is added, with a signal to noise ratio (SNR) equal to 10. The mesh used to compute the observed data is finer than the inversion mesh.

For the relocation strategy, the selection of the parameter α is discussed for each experiments. Regarding the choice of the parameter L (maximum absolute value for the receiver shifts $\Delta x_{s,r}$, see equation 36), we select it equal to the surface length for the

508 Marmousi II case (the receivers are allowed to be relocalized on the whole surface), while
509 we take it equal to half of the surface length for BP 2004 and Valhall experiments. These
510 rather unrestrictive choices yield meaningful results, and it seems not necessary at this stage
511 to adapt L along the iterations.

512 **Marmousi II**

513 We use the Marmousi II P-wave velocity model introduced in (Martin et al., 2006), which
514 is 3.5 km deep and 17 km long. We generate observed data using a fixed spread surface
515 acquisition with 128 sources and 170 receivers. The source and receiver spacing is 125 m
516 and 100 m respectively. We use a 25 m discretization mesh.

517 We investigate how the receiver relocalization approach can help mitigate the sensitivity
518 to the initial model design. To this purpose we define four initial models, increasingly far
519 from the exact model. Initial model 1,2 and 3 are obtained by applying a 2D Gaussian
520 smoothing to the exact model, with correlation lengths equal to 1 km, 2 km, and 4 km in
521 both horizontal and vertical directions respectively. Initial model 4 is a simple 1D linearly
522 increasing model from the water bottom at 1500 m.s^{-1} to the bottom of the model at 4000
523 m.s^{-1} . For all initial models the correct water layer (same as exact model) is appended on
524 top of the model. The exact and initial models are presented in Figure 8.

525 [Figure 8 about here.]

526 We compare inversion results obtained using a conventional L^2 FWI and the receiver
527 relocalization approach starting from these 4 models in Figure 9 and 10. In this first
528 experiment, α is set to 5×10^{-2} . Starting from model 1, both methods reconstruct satisfactory

estimates of the true model. Note that the receiver relocalization approach corrects the up-bending of the bottom left part observed in the L^2 reconstruction (around 2.5 km depth between $x = 0$ and $x = 4$ km). Starting from model 2, L^2 reconstruction starts introducing artifacts in the left part of the model, plus a low velocity anomaly at $x = 6$ km, $z = 2.5$ km. The receiver relocalization approach is more stable: there is no such artifacts, and the low velocity anomaly appears further from the center of the model ($x=3$ km, $z=2.5$ km). Starting from model 3 and 4, the L^2 reconstructions are not meaningful anymore. The receiver relocalization approach is more stable, preserving a correct estimate of the true model in the zone of main illumination (down to 3 km depth and between $x = 2$ and $x = 15$ km approximately).

[Figure 9 about here.]

This is confirmed by the analysis of the data fit presented in Figure 10. We overlay the exact left shot gather in red/blue color with the final shot gather in black and white in the different estimated models. While we observe a degradation of the data fit using the conventional L^2 approach, we see that the receiver relocalization approach is able to maintain a similar level of data-fit starting from the 4 different initial models.

[Figure 10 about here.]

One interest for working with synthetic models is the ability to quantify the model error. We use here a relative L^1 model misfit measure. For a given v_P model, discretized on a M points mesh, it is computed as

$$E_{v_P} = \frac{100}{M} \sum_{i=1}^M \frac{|v_{P,i} - v_{P,i}^{true}|}{|v_{P,i}^{true}|} \quad (39)$$

549 where v_P^{true} is the true P-wave velocity model. In Figure 11a we compare the decrease of
 550 the misfit function along the inversion iterations for both L^2 and receiver relocalization
 551 approaches, starting from the four initial models. The same plot for the model error is
 552 presented in Figure 11b. Finally, we present the model error decrease with respect with the
 553 misfit function decrease in Figure 11c. Interestingly, we see that the receiver relocalization
 554 approach provides a systematic lower model misfit error, even in the case where there
 555 is no cycle skipping and L^2 FWI works well. Starting from initial models 3 and 4, the
 556 receiver relocalization approach is able to decrease the model error, which is not the case
 557 for conventional L^2 FWI. Except for initial model 4, receiver relocalization always provides
 558 a monotonic decrease of the model error with respect to the misfit function (which is the
 559 expected behavior for a stable inversion). In case of initial model 4, there is an initial phase
 560 where the model error increases before decreasing, which corresponds to the first iterations
 561 of the process. Remember that initial model 4 is a vertically increasing model, therefore
 562 significantly far from the exact model.

563 [Figure 11 about here.]

564 To foster the analysis of the receiver relocalization strategy in itself, we present in Figure
 565 12 the evolution through iterations of the relocalization error for the leftmost, central, and
 566 rightmost shot gathers, depending on the choice of initial model. This error, for a given shot
 567 gather s , corresponds to the quantity

$$E_{\Delta x} = \frac{1}{N_r} \sqrt{\sum_{r=1}^{N_r} |\Delta x_{s,r}|^2}. \quad (40)$$

568 This is an average over all the receivers of the relocalization error Δx for the shot s . We see

569 that this error tends to 0 along the iteration process. The speed of convergence depends
 570 on the initial model and on the shot gathers. For the central shot gathers, the convergence
 571 is much faster than for the leftmost and rightmost ones. For initial model 1 (easiest one)
 572 the convergence is also attained faster. For initial models 2 and 3, the speed of convergence
 573 is comparable. The values of the average relocalization are higher for initial model 3. The
 574 model being further from the exact one, stronger kinematic effects need to be accounted for
 575 through the relocalization process. This is even more visible for initial model 4. For this
 576 model, the convergence is the slowest, as well as the value of the mean relocalization error.
 577 As expected, stronger kinematic mismatch thus results in a higher compensation through
 578 relocalization of receivers.

579 [Figure 12 about here.]

580 A more qualitative visualization of the relocalization process is proposed in Figures 13,
 581 14 and 15. In these figures, we present the leftmost shot gather data-fit before and after
 582 the relocalization, in P-wave velocity models obtained at iteration 0 (Fig. 13), iteration 100
 583 (Fig. 14) and in the final model (Fig. 15). We have selected the experiment starting from
 584 the initial model 4 (1D linearly increasing model) for these Figures. The models with the
 585 receiver position represented as yellow ellipses are appended to the data. The effect of the
 586 relocalization step on the data-fit is strong: the receiver repositioning makes possible to
 587 compensate for the cycle-skipped diving wave visible on the left panel of Figure 13. Some
 588 events are not correctly matched: in particular we can see that the part of the diving waves in
 589 the synthetic data arriving at offset between 3 and 7 km are matched with strong reflections
 590 in the observed data. However, at further offset, the match seems better, and it certainly
 591 helps the method to mitigate this strong cycle skipping effect. We can also link this incorrect

592 initial matching to the rather slow convergence of the process in the initial iterations when
 593 starting from initial model 4. However, at iteration 100 (Fig. 14), the data fit is already
 594 much better, and we can see the same effect of the relocalization which compensates for the
 595 too fast diving wave. In the final model, the data fit is already very good, with much of the
 596 events in phase and correctly predicted. Therefore, as expected, the relocalization has very
 597 little effect on the data fit in the final stage of the iterations.

598 [Figure 13 about here.]

599 [Figure 14 about here.]

600 [Figure 15 about here.]

601 Finally, we analyze in Figure 16 the sensitivity of the relocalization error with respect
 602 to the choice of the regularization parameter α (equation 36). This parameter controls
 603 the weight on the annihilator term, which restrains the receivers from moving to far away
 604 from their true position. We present the model error evolution along the iteration of the
 605 inversion process. We vary α between 0.01 and 0.1 with 0.01 increment. Interestingly, we
 606 see that the model error follows the same trend for any of these parameters, with relatively
 607 few variations. This is encouraging toward a robust behavior of the receiver relocalization
 608 method regarding the tuning parameter α .

609 [Figure 16 about here.]

610 **BP 2004**

611 We use a rescaled version of the original BP 2004 model (rescaling by a factor 2), and focus
 612 on the left part of the model where the high velocity salt structures are the more complex.

613 The exact model we consider is almost 6 km deep and 16.2 km long (Fig.17a). We use a
614 fixed spread acquisition with 128 sources and 161 receivers at 50 m depth in the water layer,
615 from $x = 0.1$ km to $x = 16.1$ km. The source and receiver spacing is 125 m and 100 m
616 respectively. To design the initial model, we first remove the salt from the exact model. We
617 then smooth the resulting background model. The resulting initial model is presented in
618 Figure 17b.

619 [Figure 17 about here.]

620 The leftmost shot gather is presented in Figure 18. The salt structure, especially the
621 canyon structure at $x = 2$ km, generates energetic first order (red arrow) and higher order
622 reflections (orange arrows), also with interactions with the free surface at $z = 0$ km. The
623 blue arrows depict the refraction of the direct by the salt body. The event depicted by the
624 green arrows corresponds to the transmission of the direct wave within the salt structure.
625 Black arrows depict arrivals coming from below the salt after interacting with the canyon.
626 Correctly matching the events depicted by the red, blue and green arrows is crucial to
627 recover correctly the salt structure, especially starting from the model in Figure 17b.

628 [Figure 18 about here.]

629 To mitigate the complexity of the data, we use a time-windowing approach similar to the
630 one we designed in Métivier et al. (2016). The inversion is decomposed in 7 time windows of
631 increasing lengths: 6.9 s, 9.2 s, 10.35 s, 11.5 s, 12.65 s, 13.8 s and 14.95 s. We use such a
632 long recording time to investigate the ability to reconstruct the subsalt velocity. Exploiting
633 late events, which have traveled below the salt might help achieving this reconstruction.

634 Subsalt imaging is a knowledgeable challenge. As for the previous experiment, we select a
635 low value for α , such that $\alpha = 5 \times 10^{-2}$.

636 We compare the results obtained using the receiver relocalization approach and con-
637 ventional L^2 FWI. The comparison is shown for the 1st, 2nd, and last time-window. The
638 reconstructed models are presented in Figure 19. Interestingly, the receiver relocalization
639 method provides satisfactory reconstruction of the main salt body, including the canyon
640 zone around $x = 2$ km, already from the inversion of the two first time-windows. The final
641 results, obtained after the inversion of all the time windows, show that the subsalt velocity
642 in the zone between $x = 6$ km and $x = 10$ km is correctly reconstructed, down to 5 km
643 depth. We note also that the whole right part of the model, with no salt structure on top,
644 between $x = 10$ km and $x = 16$ km, is accurately reconstructed, down to 6 km depth. The
645 subsalt zone between $x = 0$ km and $x = 6$ km remains difficult to image.

646 Comparatively, results achieved using a conventional L^2 FWI are much less satisfactory.
647 Inverting for the first time-window only yields the top-salt structure. The whole salt structure
648 is reconstructed only after the last stage of inversion, with still a visibly incorrect recovery
649 of the canyon structure on the left. The whole subsalt target is not correctly imaged either.

650 [Figure 19 about here.]

651 To interpret these results, we present the data fit in the final models in Figure 20. The
652 true data in blue/red color is superposed with the synthetic data in black and white color. In
653 the correct data fit, no black and white events should appear. The L^2 data fit is correct for
654 the refracted and transmitted events depicted by blue and green arrows. However, the short
655 offset reflections depicted by the red arrow, and multiples of these reflections (orange arrows)
656 are not correctly matched. This is consistent with the incorrect geometry of the canyon

657 structure within the salt body which is recovered using the L^2 approach. Later arrivals
 658 coming from under the salt (black arrows) are also not correctly matched. On the contrary,
 659 the data fit achieved following the receiver relocalization strategy is more satisfactory. All
 660 the events depicted by the colored arrows are correctly matched. Even relatively late events
 661 ($t > 10$ s) are matched, which is consistent with the correct reconstruction of the subsalt
 662 part of the model. This experiment thus shows that the receiver relocalization strategy
 663 could be useful in the specific context of salt and subsalt imaging. The degree of freedom
 664 introduced on the receiver position level helps matching out of phase events, associated
 665 with complex paths within and below the salt structure, which cause strong artifacts in
 666 a conventional L^2 reconstruction. By progressively relaxing the receiver position towards
 667 their physical position, the receiver relocalization strategy makes it possible to improve the
 668 velocity model to match all these events and recover the correct geometry of the salt body,
 669 as well as information on the subsalt region.

670 [Figure 20 about here.]

671 **Valhall**

672 We end up this series of experiment with the synthetic Valhall case study. Here the model is
 673 representative of the North Sea geology, with shallow water, horizontally stratified structure,
 674 and gas bearing sediments. A layered gas cloud is located above a strong reflector with
 675 an anticlinal structure. The oil reservoir is located below. The presence of gas induces
 676 a rather strong attenuation effect (amplitude decrease and dispersion), which makes the
 677 reservoir imaging challenging. The exact P-wave velocity, density and quality factor model
 678 used to generate the data are presented in Figure 21. In the modeling, the quality factor

679 is considered independent of the frequency within the frequency band considered, which
 680 is approximately 2.5 - 15 Hz. This is enforced through the use of 3 standard linear solid
 681 (SLS) mechanisms (Yang et al., 2018a). The Valhall field is one of the first exploration scale
 682 target on which FWI has been applied successfully, yielding unprecedented high resolution
 683 images of the subsurface (Sirgue et al., 2010). Since then the Valhall data has served for
 684 testing different FWI methodologies including frequency-domain multiparameter FWI and
 685 time-domain visco-acoustic FWI (Operto et al., 2015; Operto and Miniussi, 2018; Kamath
 686 et al., 2021).

687 [Figure 21 about here.]

688 The initial models we consider are presented in Figure 22. The initial P-wave velocity
 689 model is obtained through a strong Gaussian smoothing of the exact model, with correlation
 690 lengths equal to 4 km. The initial density model is derived from a Gardner’s law from this
 691 initial P-wave velocity model

$$\rho(x) = 1741 \left(10^{-3}v_P(x)\right)^{0.25}, \quad (41)$$

692 with $\rho = 1000 \text{ kg.m}^{-3}$ in the water layer. The initial quality factor model is built by setting
 693 its value to 1000 in the water layer and 100 below. During the inversion, these initial density
 694 and quality factor models are kept unchanged (passive parameters).

695 [Figure 22 about here.]

696 As a first step, we estimate the source wavelet in this initial model, following the frequency-
 697 domain deconvolution of Pratt (1999). We assume here the same wavelet for all shots. The

698 resulting estimated source wavelet is presented in Figure 23, where it is compared with
 699 the true source wavelet. Both time signature and amplitude spectrum are presented. We
 700 see that despite the noise and the inaccurate starting velocity and density models, the
 701 estimated wavelet remains relatively close to the true one. Differences are however visible in
 702 the normalized amplitude spectrum.

703 [Figure 23 about here.]

704 The P-wave velocity models obtained using L^2 FWI and the receiver relocation
 705 approach are presented in Figure 24. As can be seen, the L^2 inversion fails to produce a
 706 meaningful estimate of the P-wave velocity model, except in the shallow part above 1 km
 707 depth. This part, sampled by diving and reflected waves, is relatively well reconstructed,
 708 except for the presence of high wavenumber artifacts around $x = 9$ km and $z = 0.8$ km.
 709 Below 1 km depth, a strong horizontally extended low velocity artifact is injected in the
 710 model reconstruction. The layered shape gas cloud below is not properly reconstructed. The
 711 continuity of the strong reflector at 2.5 km depth is broken, and its anticlinal shape is not
 712 reconstructed. All this indicates the convergence towards a non informative local minimum
 713 due to cycle skipping.

714 On the contrary, the P-wave velocity model obtained following the receiver relocation
 715 approach is much closer to the exact model. The successive gas layers are properly
 716 reconstructed, as well as the main reflector at 2.5 km depth, which appears continuous, and
 717 with an anticlinal shape. Below, the medium is not sufficiently sampled by waves to make
 718 it possible to reconstruct it from the initial model which is used here. We can also note
 719 the presence of artifacts on the lateral edges of the model, which are also due to a lack of
 720 illumination in these part of the model. Low velocity V-shape artifacts also appear on both

721 sides of the gas cloud, which indicate still the presence of cycle skipped events. However,
722 the overall estimation is correct down to 3 km depth.

723 [Figure 24 about here.]

724 These results indicate that the receiver relocation approach is robust to relatively
725 realistic settings where the amplitude of the data cannot be predicted to machine precision.
726 This is comforting for perspectives of application to field data. To better understand the
727 difference between the L^2 and receiver relocation reconstruction, we compare the final
728 data match using both approaches for the shot-gather associated with source position $x_S = 8$
729 km (Fig. 25). The superposition of exact (blue and red) and synthetic data (black and
730 white) in the final model is intriguing: the L^2 data match seems relatively good, especially
731 for diving waves. The receiver relocation approach provides also a good data match,
732 however less accurate regarding the larger offset arrivals. In Figure 26, we compare the
733 normalized residuals computed between the observed data without noise, and the synthetic
734 data in the final models provided by the two approaches. This comparison provides the
735 explanation of the difference between the two reconstructed models. The L^2 approach is
736 unable to correctly explain the short and medium offset reflections, associated with the
737 gas cloud layers. Conversely, the receiver relocation approach provides a model which
738 explains significantly better these reflections, while increasing slightly the misfit with respect
739 to largest offset diving waves.

740 The reason why the misfit related to these events remains large is that in the final model,
741 the receiver position has still not converged towards the true position of the receivers. The
742 average relocation error for the shot considered here ($x_S=8$ km) indicates a systematic
743 drift of 50 m even in the final model. This ambiguity shows that the weight associated to the

annihilator might not be not sufficiently high. We have tested different values of the weight α , however choosing a too large value prevents for adding sufficient freedom to the inversion in the early stage of the inversion to obtain a satisfactory reconstruction. The best results where achieved with $\alpha = 1000$, which is already a significantly higher value than what is used for Marmousi and BP 2004 case studies. This is in accordance with the presence of noise and the consequently higher value of the data matching term in the misfit function, which requires to strengthen the weight of the annihilator term in the relocalization approach.

[Figure 25 about here.]

[Figure 26 about here.]

Computational cost

We end this Section with a comparison of computational cost of the receiver relocalization approach for each case study. The results are presented in Table 1. For each case study, we provide the computational time for a gradient computation, and provide the extra computational time associated with the receiver relocation strategy. The reference time for the extra computational cost is the one which would be obtained with a L^2 approach. We see that in the three cases, the extra cost remains below 5% which makes the receiver relocalization strategy relatively inexpensive. Note also that the overall computational cost associated with the Valhall model (approximately the same size as the Marmousi model), is significantly higher: this is related to the visco-acoustic modeling.

[Table 1 about here.]

DISCUSSION

764 The three case studies investigated in the previous Section illustrate the interesting properties
765 of the receiver relocation approach. In all cases, the method makes possible to start from
766 crude initial models while still providing meaningful velocity estimations. The method is
767 applicable directly in the frame of time-domain FWI. In the 2D (visco-)acoustic approximation
768 considered here, the extra computational cost compared to a conventional least-squares
769 approach is negligible. Based on the separability of the least-squares misfit function, the
770 inner loop complexity depends linearly on the number of source/receiver pairs. We have
771 considered here fixed spread acquisition systems. The computational cost increase would be
772 even lower for corresponding streamer acquisition with constant offset, which would induce
773 less source/receiver pairs.

774 We discuss here practical aspect and potential extensions of the method. First, it is
775 important to control the design of the initial model with respect to the initial step of
776 receiver relocation. A too fast model could require to relocate the receivers outside
777 the computational box. It might thus be advisable to start with initial velocity models
778 underestimating the true velocity, or to adapt the computational box to the initial receiver
779 relocation step.

780 Second, we have used here a grid search algorithm for the solution of the inner loop
781 problem. Other possibilities could be considered if it becomes necessary to reduce the
782 computational cost. Markov-Chain Monte-Carlo method could be used instead, in particular
783 its recent Hamiltonian accelerated variant (Neal et al., 2011).

784 Third, we have observed that, for a given source s , the receiver relocation $\Delta x_{r,s}$
785 can change rather abruptly for neighboring traces r_k, r_{k+1} . It is possible that these rapid

786 changes slow down the convergence of the whole method, which is observed for instance
 787 on the Marmousi II experiment where several hundreds iterations are required to converge.
 788 For this reason, it might be advisable to add a regularization term and/or constraints in
 789 the misfit function to promote smoother variations. This could be done by penalizing the
 790 discrete difference of the corrections between two traces for a given source, or by smoothing
 791 directly in the receiver direction the receiver correction vector $\Delta x_{r,s}$ solution of the inner
 792 loop.

793 Fourth, the convexity analysis and the resulting implementation performed in this study
 794 is done in the frame of single arrival traces. While the three case studies of the preceding
 795 Section illustrate that the method works in the frame of complex multi-arrival data, it might
 796 still be interesting to extend the analysis and implementation of the method to the case
 797 of multi-arrival traces. In this frame, the receiver position correction which we consider
 798 could depend on time. For a workable method, time windows should be defined prior to the
 799 application of the method, and a receiver position correction could be computed for each
 800 time-window. This could be interesting for instance to avoid mismatch of events (diving
 801 interpreted as strong reflections) in the initial iterations of the process.

802 Finally, we discuss the application of the receiver relocation method in a 3D context.
 803 In 3D, isochrones are surfaces. Therefore, even if we restrict the receiver relocation to the
 804 surface (forbidding vertical relocation), an ambiguity would subsist in the case of single
 805 event traces. Again, this ambiguity can be prevented by restricting the receiver relocation
 806 correction to a single parameter, which could be in this case a surface repositioning r in
 807 the direction of the source/receiver axis. The correction would thus not be aligned with
 808 horizontal axis x and y . The additional benefit would be again to obtain inner loop problems
 809 depending on a single degree of freedom, making the solution through global optimization

810 almost negligible. Thus it seems 3D extension of the method might be feasible.

CONCLUSION

811 We propose in this study a receiver relocation method as a novel extension strategy,
812 which is directly applicable to time-domain FWI. The receiver position is introduced as a
813 degree of freedom in the FWI problem, which makes it possible to reduce the kinematic
814 mismatch which would lead conventional least-squares FWI to converge a local minimum.
815 Doing so, the data is fit progressively by the subsurface model as receivers converge towards
816 their true positions. The method is implemented similarly as source extension strategies,
817 using a variable projection approach, with an inner loop dedicated to the computation of
818 the optimal receiver position and an outer loop dedicated to the subsurface model update.
819 Our implementation solves the inner loop problem using a brute force grid search approach.
820 The outer loop problem is solved using a conventional quasi-Newton *l*-BFGS approach.

821 We illustrate the properties of this receiver relocation strategy first on a schematic
822 cross-hole experiment, exhibiting the robustness of the approach with respect to strong
823 kinematic mismatch and its resilience with respect to cycle skipping. Then we investigate
824 three synthetic case studies, representative of different geological context. In all three
825 cases, the receiver relocation strategy is shown to successfully converge toward a correct
826 estimation of the subsurface model starting from crude initial models, with a relatively
827 inexpensive additional computational cost (no more than 5% more expensive).

828 The good results obtained, in particular in the Valhall case, where noise, inexact source
829 wavelet, inexact density and attenuation models, make it not possible to predict the data
830 amplitude with arbitrary precision, are encouraging towards application to field data. Finally,
831 compared to misfit modification approaches based on optimal transport distances, which we
832 have recently studied, our experiments indicate that the receiver relocation approach

833 appears as a competitive alternative. The computational cost increase is of the same order
834 or even lower for these 2D experiments, and the robustness to cycle skipping seems also
835 comparable. Future studies will include comparisons between these different approaches and
836 applications to 3D field data.

ACKNOWLEDGMENTS

837 The authors would like to express their gratitude to William Symes for insightful discussions
838 on source-extension strategies. This study was partially funded by the SEISCOPE consortium
839 (<http://seiscope2.osug.fr>), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-
840 MOBIL, JGI, SHELL, SINOPEC, SISPROBE and TOTAL. This study was granted access to
841 the HPC resources of the Froggy platform of the CIMENT infrastructure ([https://ciment.ujf-](https://ciment.ujf-grenoble.fr)
842 [grenoble.fr](https://ciment.ujf-grenoble.fr)), which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA),
843 the OSUG@2020 labex (reference ANR10 LABX56) and the Equip@Meso project (reference
844 ANR-10-EQPX-29-01) of the programme Investissements d’Avenir supervised by the Agence
845 Nationale pour la Recherche, and the HPC resources of CINES/IDRIS/TGCC under the
846 allocation 046091 made by GENCI.”

REFERENCES

- 847 Aghamiry, H., A. Gholami, and S. Operto, 2019a, Admm-based multi-parameter wavefield
848 reconstruction inversion in VTI acoustic media with TV regularization: Geophysical
849 Journal International, **219**, 1316–1333.
- 850 ———, 2019b, Improving full-waveform inversion by wavefield reconstruction with alternating
851 direction method of multipliers: Geophysics, **84(1)**, R139–R162.
- 852 ———, 2020, Accurate and efficient wavefield reconstruction in the time domain: Geophysics,
853 **85(2)**, A7–A12.
- 854 Almomin, A., 2016, Tomographic full waveform inversion: PhD thesis, Stanford University.
- 855 Amestoy, P., R. Brossier, A. Buttari, J.-Y. L’Excellent, T. Mary, L. Métivier, A. Miniussi,
856 and S. Operto, 2016, Fast 3D frequency-domain full waveform inversion with a parallel
857 Block Low-Rank multifrontal direct solver: application to OBC data from the North Sea:
858 Geophysics, **81**, R363 – R383.
- 859 Beller, S., V. Monteiller, L. Combe, S. Operto, and G. Nolet, 2018, On the sensitivity of
860 teleseismic full waveform inversion to earth parametrisation, initial model and acquisition
861 design: Geophysical Journal International, **212(2)**, 1344–1368.
- 862 Bérenger, J.-P., 1994, A perfectly matched layer for absorption of electromagnetic waves:
863 Journal of Computational Physics, **114**, 185–200.
- 864 Billette, F. J., and S. Brandsberg-Dahl, 2004, The 2004 BP velocity benchmark: Extended
865 Abstracts, 67th Annual EAGE Conference & Exhibition, Madrid, Spain, B035.
- 866 Boyd, S. P., and L. Vandenberghe, 2004, Convex optimization: Cambridge university press.
- 867 Bozdağ, E., D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D.
868 Pugmire, 2016, Global adjoint tomography: first-generation model: Geophysical Journal
869 International, **207**, 1739–1766.

870 Bozdağ, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inver-
871 sion based on instantaneous phase and envelope measurements: *Geophysical Journal*
872 *International*, **185**, 845–870.

873 Brossier, R., S. Operto, and J. Virieux, 2009, Seismic imaging of complex onshore struc-
874 tures by 2D elastic frequency-domain full-waveform inversion: *Geophysics*, **74**, WCC105–
875 WCC118.

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

878 Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef, 1985, A nonreflecting boundary condition
879 for discrete acoustic and elastic wave equations: *Geophysics*, **50**, 2117–2131.

880 Choi, Y., and C. Shin, 2008, Frequency-Domain Elastic Full Waveform Inversion Using the
881 New Pseudo-Hessian Matrix: Experience Of Elastic Marmousi 2 Synthetic Data: *Bulletin*
882 *of the Seismological Society of America*, **98**, 2402–2415.

883 Cocher, E., H. Chauris, and R. Plessix, 2017, Seismic Iterative Migration Velocity Analysis:
884 two strategies to update the velocity model: *Computational Geosciences*, **21**, 759–780.

885 Combettes, P. L., and J.-C. Pesquet, 2011, Proximal splitting methods in signal processing,
886 *in Fixed-Point Algorithms for Inverse Problems in Science and Engineering*: Springer New
887 York, volume **49** of *Springer Optimization and Its Applications*, 185–212.

888 Donno, D., H. Chauris, and H. Calandra, 2013, Estimating the background velocity model
889 with the normalized integration method: *EAGE Technical Program Expanded Abstracts*
890 2013, Tu0704.

891 Engquist, B., and B. D. Froese, 2014, Application of the Wasserstein metric to seismic
892 signals: *Communications in Mathematical Science*, **12**, 979–988.

893 Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge, 2008, Theoretical background

894 for continental- and global-scale full-waveform inversion in the time-frequency domain:
895 Geophysical Journal International, **175**, 665–685.

896 Fichtner, A., and A. Villaseñor, 2015, Crust and upper mantle of the western mediterranean
897 – constraints from full-waveform inversion: Earth and Planetary Science Letters, **428**, 52 –
898 62.

899 Golub, G., and V. Pereyra, 2003, Separable nonlinear least squares: the variable projection
900 method and its applications: Inverse problems, **19**, R1.

901 Górszczyk, A., S. Operto, and M. Malinowski, 2017, Toward a robust workflow for deep
902 crustal imaging by FWI of OBS data: The eastern nankai trough revisited: Journal of
903 Geophysical Research: Solid Earth, **122**, 4601–4630.

904 Huang, G., R. Nammour, and W. W. Symes, 2018a, Source-independent extended wave-
905 form inversion based on space-time source extension: Frequency-domain implementation:
906 Geophysics, **83**, R449–R461.

907 ———, 2018b, Volume source-based extended waveform inversion: Geophysics, **83**, R369–387.

908 Huang, G., R. Nammour, W. W. Symes, and M. Dolliazal, 2019, Waveform inversion via
909 source extension, *in* SEG Technical Program Expanded Abstracts 2019: 4761–4766.

910 Jannane, M., W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M.
911 Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, and D. Trezeguet, 1989, Wavelengths
912 of Earth structures that can be resolved from seismic reflection data: Geophysics, **54**,
913 906–910.

914 Kamath, N., R. Brossier, L. Métivier, A. Pladys, and P. Yang, 2021, Multiparameter full-
915 waveform inversion of 3D ocean-bottom cable data from the Valhall field: Geophysics, **86**,
916 B15–B35.

917 Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Con-

ference on Inverse Scattering, Theory and application, Society for Industrial and Applied
Mathematics, Philadelphia, Conference on Inverse Scattering, Theory and application,
Society for Industrial and Applied Mathematics, Philadelphia, 206–220.

Lei, W., Y. Ruan, E. Bozdağ, D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill,
N. Podhorszki, and D. Pugmire, 2020, Global adjoint tomography—model glad-m25:
Geophysical Journal International, **223**, 1–21.

Li, Y., R. Brossier, and L. Métivier, 2020, On the comparison of MUMPS and STRUMPACK
for 3D frequency-domain elastic wave modeling: Presented at the Expanded Abstracts,
82nd Annual EAGE Meeting (Amsterdam).

Lu, Y., L. Stehly, R. Brossier, A. Paul, and A. W. ng Group, 2020, Imaging Alpine crust
using ambient noise wave-equation tomography: Geophysical Journal International, **222**,
69–85.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation
inversion: SEG Technical Program Expanded Abstracts, **30**, 2788–2792.

Luo, Y., and G. T. Schuster, 1991, Wave-equation travelttime inversion: Geophysics, **56**,
645–653.

Martin, G. S., R. Wiley, and K. J. Marfurt, 2006, Marmousi2: An elastic upgrade for
Marmousi: The Leading Edge, **25**, 156–166.

Métivier, L., and R. Brossier, 2016, The SEISCOPE optimization toolbox: A large-scale
nonlinear optimization library based on reverse communication: Geophysics, **81**, F11–F25.

Métivier, L., R. Brossier, Q. Mérigot, and E. Oudet, 2019, A graph space optimal transport
distance as a generalization of L^p distances: application to a seismic imaging inverse
problem: Inverse Problems, **35**, 085001.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016, Measuring the misfit

942 between seismograms using an optimal transport distance: Application to full waveform
 943 inversion: *Geophysical Journal International*, **205**, 345–377.

944 Neal, R. M., et al., 2011, Mcmc using hamiltonian dynamics: *HandBook of markov chain*
 945 monte carlo, **2**, 2.

946 Nocedal, J., 1980, Updating Quasi-Newton Matrices With Limited Storage: *Mathematics of*
 947 Computation, **35**, 773–782.

948 Nocedal, J., and S. J. Wright, 2006, *Numerical optimization*, 2nd ed.: Springer.

949 Operto, S., and A. Miniussi, 2018, On the role of density and attenuation in 3D multi-
 950 parameter visco-acoustic VTI frequency-domain FWI: an OBC case study from the North
 951 Sea: *Geophysical Journal International*, **213**, 2037–2059.

952 Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and
 953 J. Virieux, 2015, Efficient 3-D frequency-domain mono-parameter full-waveform inversion
 954 of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse
 955 isotropic approximation: *Geophysical Journal International*, **202**, 1362–1391.

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

958 Plessix, R. E., and C. Perkins, 2010, Full waveform inversion of a deep water ocean bottom
 959 seismometer dataset: *First Break*, **28**, 71–78.

960 Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part I: theory and
 961 verification in a physical scale model: *Geophysics*, **64**, 888–901.

962 Raknes, E. B., B. Arntsen, and W. Weibull, 2015, Three-dimensional elastic full waveform
 963 inversion using seismic data from the sleipner area: *Geophysical Journal International*,
 964 **202**, 1877–1894.

965 Shipp, R. M., and S. C. Singh, 2002, Two-dimensional full wavefield inversion of wide-aperture

966 marine seismic streamer data: *Geophysical Journal International*, **151**, 325–344.

967 Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010,

968 Full waveform inversion: the next leap forward in imaging at Valhall: *First Break*, **28**,

969 65–70.

970 Solano, C. P., and R.-É. Plessix, 2019, Velocity-model building with enhanced shallow resolu-

971 tion using elastic waveform inversion — an example from onshore oman: *GEOPHYSICS*,

972 **84**, R977–R988.

973 Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical*

974 *Prospecting*, **56**, 765–790.

975 ———, 2014, Seismic inverse problems: recent developments in theory and practice, *in*

976 *Proceedings of the Inverse Problems from Theory to Applications Conference (IPTA2014)*.

977 ———, 2020, Wavefield reconstruction inversion: an example: *Inverse Problems*, **36**, 105010.

978 Symes, W. W., H. Chen, and S. E. Minkoff, 2020, Full-waveform inversion by source extension:

979 Why it works, *in* *SEG Technical Program Expanded Abstracts 2020*: 765–769.

980 Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation:

981 *Geophysics*, **49**, 1259–1266.

982 van Leeuwen, T., and F. Herrmann, 2016, A penalty method for PDE-constrained optimiza-

983 tion in inverse problems: *Inverse Problems*, **32(1)**, 1–26.

984 van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform

985 inversion by expanding the search space: *Geophysical Journal International*, **195(1)**,

986 661–667.

987 Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application

988 using multicomponent measurements of seismic data collection: *Geophysics*, **79**, R63–R77.

989 Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, An

990 introduction to Full Waveform Inversion, *in* Encyclopedia of Exploration Geophysics:
991 Society of Exploration Geophysics, R1–1–R1–40.

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

994 Wang, C., D. Yingst, P. Farmer, and J. Leveille, 2016, Full-waveform inversion with the
995 reconstructed wavefield method, *in* SEG Technical Program Expanded Abstracts 2016:
996 1237–1241.

997 Wang, Y., and Y. Rao, 2009, Reflection seismic waveform tomography: *Journal of Geophysical*
998 *Research*, **114**, 1978–2012.

999 Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: *Geophysics*, **81**,
1000 R429–R445.

1001 Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl,
1002 L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform
1003 inversion: *Geophysics*, **78**, R59–R80.

1004 Wu, R.-S., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal
1005 model: *Geophysics*, **79**, WA13–WA24.

1006 Yang, P., R. Brossier, L. Métivier, J. Virieux, and W. Zhou, 2018a, A Time-Domain
1007 Preconditioned Truncated Newton Approach to Multiparameter Visco-acoustic Full
1008 Waveform Inversion: *SIAM Journal on Scientific Computing*, **40**, B1101–B1130.

1009 Yang, Y., B. Engquist, J. Sun, and B. F. Hamfeldt, 2018b, Application of optimal transport
1010 and the quadratic Wasserstein metric to full-waveform inversion: *Geophysics*, **83**, R43–
1011 R62.

LIST OF FIGURES

1012	1	Acquisition configuration for the transmission case.	58
1013	2	Map of the misfit function $f(m, \Delta x)$ in the cross-hole transmission case,	
1014		using a single source/receiver couple at 50 m depth. The misfit function is	
1015		normalized such that its maximum reaches 1. Its minimum is located in a	
1016		narrow valley of attraction at position $\Delta x = 0$ and $m = 2000 \text{ m.s}^{-1}$	59
1017	3	Profile of the misfit function $g(m)$ in the single couple source/receiver trans-	
1018		mission case, depending on the weight α . Small weights (0.1, 1) yield a convex	
1019		function with respect to the velocity m , while larger weights (10, 100) tend	
1020		to reduce the width of the valley of attraction. In all cases the minimum is	
1021		reached at the correct velocity $m = 2000 \text{ m.s}^{-1}$	60
1022	4	Comparison of the reference trace in black, obtained in the medium at 2000	
1023		m.s^{-1} , and synthetic traces before and after relocalization, obtained in media	
1024		at 1250 m.s^{-1} , 1500 m.s^{-1} , and 2500 m.s^{-1} respectively. The traces have	
1025		been normalized according to the maximum amplitude of the reference trace.	
1026		The receiver relocalization is able to correct for the wrong kinematic and put	
1027		the traces in phase with the reference trace. The amplitude difference remains	
1028		however incorrect. The amplitude mismatch is convex with respect to the	
1029		velocity: this explains why the receiver extension misfit function is convex	
1030		with respect to the velocity in this case.	61
1031	5	Transmission case, cross-hole configuration, single source/receiver pair. Ref-	
1032		erence, synthetic and adjoint traces in the least squares case for the slower	
1033		medium at 1500 m.s^{-1} (a), faster medium at 2500 m.s^{-1} (b). Least-squares	
1034		kernel in the slower medium at 1500 m.s^{-1} (c), faster medium at 2500 m.s^{-1}	
1035		(d). In both cases the first Fresnel zone exhibits the same negative value,	
1036		while a change of sign would be expected from a misfit function convex with	
1037		respect to the velocity.	62
1038	6	Transmission case, cross-hole configuration, single source/receiver pair. Ref-	
1039		erence, synthetic and adjoint traces in the receiver extension case for the	
1040		slower medium at 1500 m.s^{-1} (a), faster medium at 2500 m.s^{-1} (b). Receiver	
1041		extension kernel in the slower medium at 1500 m.s^{-1} (c), faster medium	
1042		at 2500 m.s^{-1} (d). The synthetic and reference traces are in phase, with	
1043		higher amplitude for the slower medium (a) and lower amplitude for the faster	
1044		medium (b). The adjoint source therefore changes its sign from one medium	
1045		to the other. The first Fresnel zone of the corresponding kernels thus exhibits	
1046		the same change of sign depending whether it is calculated in the slower (c)	
1047		or faster medium (d).	63
1048	7	Time signature of the source wavelet used to generate the data for Marmousi	
1049		and Valhall case (a), together with its power spectrum (b). No energy is	
1050		present below 2.5 Hz.	64
1051	8	Marmousi II case study. Exact (a) and initial models 1 (b), 2 (c), 3 (d), and	
1052		4 (e).	65

1053	9	Marmousi II case study. Reconstructed models starting from the initial models in Figure 8 using a conventional L^2 approach (left column) and the receiver relocalization approach (right column). L^2 results starting from models 1 (a), 2 (b), 3 (c), 4 (d). Receiver relocalization results starting from models 1 (e), 2 (f), 3 (g), 4 (h).	66
1054			
1055			
1056			
1057			
1058	10	Marmousi II case study. Final data fit using a conventional L^2 approach (top row) and the receiver relocalization approach (bottom row) starting from the initial models presented in Figure 8. The exact data appears in blue and red colors in transparency. The data in the final models is in black and white. L^2 data fit starting from model 1 (a), 2 (b), 3 (c), 4 (d). Receiver relocalization data fit starting from model 1 (e), 2 (f), 3 (g), 4 (h).	67
1059			
1060			
1061			
1062			
1063			
1064	11	Marmousi II case study. Quantitative analysis of error evolution. Relative decrease of the misfit function along iterations (a). Relative decrease of the model error along iterations (b). Relative decrease of the model error depending on the decrease of the misfit function (c).	68
1065			
1066			
1067			
1068	12	Marmousi II case study. Evolution of the relocalization error along iterations, depending on the choice of initial model for leftmost shot gather (a), central shot gather (b), right shot gather (c).	69
1069			
1070			
1071	13	Marmousi II case study. Data fit in the initial model before (a) and after (b) the receiver relocalization step. The receiver positions are represented in yellow ellipses.	70
1072			
1073			
1074	14	Marmousi II case study. Data fit in the iteration 100 model before (a) and after (b) the receiver relocalization step. The receiver positions are represented in yellow ellipses.	71
1075			
1076			
1077	15	Marmousi II case study. Data fit in the final model before (a) and after (b) the receiver relocalization step. The receiver positions are represented in yellow ellipses.	72
1078			
1079			
1080	16	Marmousi II case study. Sensitivity to the weighting parameter α . Model error along the iteration depending on the choice of α , ranging from 10^{-2} to 10^{-1}	73
1081			
1082			
1083	17	Exact (a) and initial (b) models used for the BP 2004 case study.	74
1084	18	BP 2004 case study. Left most common shot gather. The canyon structure at $x = 2$ km, generates energetic first order (red arrow) and higher order reflections (orange arrows). The blue arrows depict the refraction of the direct by the salt body. The event depicted by the green arrows corresponds to the transmission of the direct wave within the salt structure. Black arrows depict arrivals coming from below the salt after interacting with the canyon	75
1085			
1086			
1087			
1088			
1089			
1090	19	BP 2004 case study. P-wave velocity reconstruction using conventional FWI after 1st time-window (a), 2nd time-window (b), 3rd time-window (c). P-wave velocity reconstruction using receiver relocalization approach after 1st time-window (d), 2nd time-window (e), 3rd time-window (f).	76
1091			
1092			
1093			

1094	20	BP 2004 case study. Data fit in the P-wave velocity models reconstructed using	
1095		a conventional FWI approach (a), using the receiver relocalization approach	
1096		(b). The true data in blue/red color is superposed with the synthetic data	
1097		in black and white color. For a correct data fit no black and white color	
1098		should appear. Blue and green arrows depict refracted and transmitted	
1099		events. Red arrow depicts short offset reflection on the left (canyon shape)	
1100		part of the model. Multiples of these reflections are depicted by orange	
1101		arrows. Black arrow depict later arrivals coming from below the salt. While	
1102		using conventional FWI, mostly transmitted events are matched, using the	
1103		relocalization approach all of these events are correctly matched.	77
1104	21	Valhall case study. Exact P-wave velocity model (a), exact density model (b),	
1105		and exact quality factor (c) used to generate the data.	78
1106	22	Valhall case study. Initial P-wave velocity model (a), initial density model	
1107		(b), and initial quality factor (c).	79
1108	23	Valhall case study. Estimated source wavelet (blue) compared with the true	
1109		source wavelet (red). Time-signature (a), normalized amplitude spectrum (b).	80
1110	24	Valhall case study. Final models obtained using L^2 FWI (a), and the receiver	
1111		relocalization strategy (b).	81
1112	25	Valhall case study. Comparison of observed (blue and red) and synthetic	
1113		(black and white) data for the shot gather associated with the source located	
1114		at $x_S = 8$ km. L^2 result (a), receiver relocalization result (b). A good data	
1115		match is indicated by the dominance of purple and black color. The presence	
1116		of white and red indicates an incorrect data match. In this respect the L^2	
1117		data match is relatively good. The receiver relocalization data match is also	
1118		correct, except for large offset diving waves events.	82
1119	26	Valhall case study. Comparison of normalized residuals for the shot gather	
1120		associated with the source located at $x_S = 8$ km. L^2 result (a), receiver	
1121		relocalization result (b). The residuals are computed as the difference between	
1122		the synthetic data in the final model using both approaches and the observed	
1123		data without noise. The L^2 residuals exhibit a clear mismatch of the short	
1124		and medium offsets event associated with reflection on the gas cloud layers.	83

Figure 1: Acquisition configuration for the transmission case.

Figure 2: Map of the misfit function $f(m, \Delta x)$ in the cross-hole transmission case, using a single source/receiver couple at 50 m depth. The misfit function is normalized such that its maximum reaches 1. Its minimum is located in a narrow valley of attraction at position $\Delta x = 0$ and $m = 2000$ m.s⁻¹.

Figure 3: Profile of the misfit function $g(m)$ in the single couple source/receiver transmission case, depending on the weight α . Small weights (0.1, 1) yield a convex function with respect to the velocity m , while larger weights (10, 100) tend to reduce the width of the valley of attraction. In all cases the minimum is reached at the correct velocity $m = 2000 \text{ m.s}^{-1}$.

Figure 4: Comparison of the reference trace in black, obtained in the medium at 2000 m.s^{-1} , and synthetic traces before and after relocalization, obtained in media at 1250 m.s^{-1} , 1500 m.s^{-1} , and 2500 m.s^{-1} respectively. The traces have been normalized according to the maximum amplitude of the reference trace. The receiver relocalization is able to correct for the wrong kinematic and put the traces in phase with the reference trace. The amplitude difference remains however incorrect. The amplitude mismatch is convex with respect to the velocity: this explains why the receiver extension misfit function is convex with respect to the velocity in this case.

Figure 5: Transmission case, cross-hole configuration, single source/receiver pair. Reference, synthetic and adjoint traces in the least squares case for the slower medium at 1500 m.s⁻¹ (a), faster medium at 2500 m.s⁻¹ (b). Least-squares kernel in the slower medium at 1500 m.s⁻¹ (c), faster medium at 2500 m.s⁻¹ (d). In both cases the first Fresnel zone exhibits the same negative value, while a change of sign would be expected from a misfit function convex with respect to the velocity.

Figure 6: Transmission case, cross-hole configuration, single source/receiver pair. Reference, synthetic and adjoint traces in the receiver extension case for the slower medium at 1500 m.s^{-1} (a), faster medium at 2500 m.s^{-1} (b). Receiver extension kernel in the slower medium at 1500 m.s^{-1} (c), faster medium at 2500 m.s^{-1} (d). The synthetic and reference traces are in phase, with higher amplitude for the slower medium (a) and lower amplitude for the faster medium (b). The adjoint source therefore changes its sign from one medium to the other. The first Fresnel zone of the corresponding kernels thus exhibits the same change of sign depending whether it is calculated in the slower (c) or faster medium (d).

Figure 7: Time signature of the source wavelet used to generate the data for Marmousi and Valhall case (a), together with its power spectrum (b). No energy is present below 2.5 Hz.

Figure 8: Marmousi II case study. Exact (a) and initial models 1 (b), 2 (c), 3 (d), and 4 (e).

Figure 9: Marmousi II case study. Reconstructed models starting from the initial models in Figure 8 using a conventional L^2 approach (left column) and the receiver relocation approach (right column). L^2 results starting from models 1 (a), 2 (b), 3 (c), 4 (d). Receiver relocation results starting from models 1 (e), 2 (f), 3 (g), 4 (h).

Figure 10: Marmousi II case study. Final data fit using a conventional L^2 approach (top row) and the receiver relocation approach (bottom row) starting from the initial models presented in Figure 8. The exact data appears in blue and red colors in transparency. The data in the final models is in black and white. L^2 data fit starting from model 1 (a), 2 (b), 3 (c), 4 (d). Receiver relocation data fit starting from model 1 (e), 2 (f), 3 (g), 4 (h).

Figure 11: Marmousi II case study. Quantitative analysis of error evolution. Relative decrease of the misfit function along iterations (a). Relative decrease of the model error along iterations (b). Relative decrease of the model error depending on the decrease of the misfit function (c).

Figure 12: Marmousi II case study. Evolution of the relocalization error along iterations, depending on the choice of initial model for leftmost shot gather (a), central shot gather (b), right shot gather (c).

Figure 13: Marmousi II case study. Data fit in the initial model before (a) and after (b) the receiver relocalization step. The receiver positions are represented in yellow ellipses.

Figure 14: Marmousi II case study. Data fit in the iteration 100 model before (a) and after (b) the receiver relocation step. The receiver positions are represented in yellow ellipses.

Figure 15: Marmousi II case study. Data fit in the final model before (a) and after (b) the receiver relocation step. The receiver positions are represented in yellow ellipses.

Figure 16: Marmousi II case study. Sensitivity to the weighting parameter α . Model error along the iteration depending on the choice of α , ranging from 10^{-2} to 10^{-1} .

Figure 17: Exact (a) and initial (b) models used for the BP 2004 case study.

Figure 18: BP 2004 case study. Left most common shot gather. The canyon structure at $x = 2$ km, generates energetic first order (red arrow) and higher order reflections (orange arrows). The blue arrows depict the refraction of the direct by the salt body. The event depicted by the green arrows corresponds to the transmission of the direct wave within the salt structure. Black arrows depict arrivals coming from below the salt after interacting with the canyon

Figure 19: BP 2004 case study. P-wave velocity reconstruction using conventional FWI after 1st time-window (a), 2nd time-window (b), 3rd time-window (c). P-wave velocity reconstruction using receiver relocation approach after 1st time-window (d), 2nd time-window (e), 3rd time-window (f).

Figure 20: BP 2004 case study. Data fit in the P-wave velocity models reconstructed using a conventional FWI approach (a), using the receiver relocation approach (b). The true data in blue/red color is superposed with the synthetic data in black and white color. For a correct data fit no black and white color should appear. Blue and green arrows depict refracted and transmitted events. Red arrow depicts short offset reflection on the left (canyon shape) part of the model. Multiples of these reflections are depicted by orange arrows. Black arrow depict later arrivals coming from below the salt. While using conventional FWI, mostly transmitted events are matched, using the relocation approach all of these events are correctly matched.

Figure 21: Valhall case study. Exact P-wave velocity model (a), exact density model (b), and exact quality factor (c) used to generate the data.

Figure 22: Valhall case study. Initial P-wave velocity model (a), initial density model (b), and initial quality factor (c).

Figure 23: Valhall case study. Estimated source wavelet (blue) compared with the true source wavelet (red). Time-signature (a), normalized amplitude spectrum (b).

Figure 24: Valhall case study. Final models obtained using L^2 FWI (a), and the receiver relocation strategy (b).

Figure 25: Valhall case study. Comparison of observed (blue and red) and synthetic (black and white) data for the shot gather associated with the source located at $x_S = 8$ km. L^2 result (a), receiver relocation result (b). A good data match is indicated by the dominance of purple and black color. The presence of white and red indicates an incorrect data match. In this respect the L^2 data match is relatively good. The receiver relocation data match is also correct, except for large offset diving waves events.

Figure 26: Valhall case study. Comparison of normalized residuals for the shot gather associated with the source located at $x_S = 8$ km. L^2 result (a), receiver relocation result (b). The residuals are computed as the difference between the synthetic data in the final model using both approaches and the observed data without noise. The L^2 residuals exhibit a clear mismatch of the short and medium offsets event associated with reflection on the gas cloud layers.

LIST OF TABLES

1125	1	Computational time for different gradient building steps.	85
------	---	---	----

Case study	$n_z \times n_x$	Inc. field	Adj. field	Rec. reloc. loop	Other	Total	Extra cost
Marmousi	141×681	6.7 s	19.1 s	1.4 s	2.4 s	29.6 s	4.9%
BP 2004	237×651	10.6 s	29.9 s	1.1 s	2.9 s	44.5	2.53%
Valhall	160×679	11.5 s	44.5 s	1.1 s	10.6 s	67.7	1.65 %

Table 1: Computational time for different gradient building steps.