



HAL
open science

Robust wavefield inversion via phase retrieval

H. Aghamiry, A. Gholami, S. Operto

► **To cite this version:**

H. Aghamiry, A. Gholami, S. Operto. Robust wavefield inversion via phase retrieval. *Geophysical Journal International*, 2020, 221 (2), pp.1327-1340. 10.1093/gji/ggaa035 . hal-02505816

HAL Id: hal-02505816

<https://hal.science/hal-02505816>

Submitted on 30 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust wavefield inversion via phase retrieval

H.S. Aghamiry^{1,2}, A. Gholami¹ and S. Operto²

¹*Institute of Geophysics, University of Tehran, Tehran, Tehran, Iran. E-mail: agholami@ut.ac.ir*

²*Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, IRD, Géoazur, Valbonne, France.*

Accepted 2020 January 16. Received 2019 November 21; in original form 2019 July 23

SUMMARY

Extended formulation of full waveform inversion (FWI), called wavefield reconstruction inversion (WRI), offers potential benefits of decreasing the non-linearity of the inverse problem by replacing the explicit inverse of the wave-equation operator of classical FWI (the oscillating Green functions) with a suitably defined data-driven regularized inverse. This regularization relaxes the wave-equation constraint to reconstruct wavefields that match the data, hence mitigating the risk of cycle skipping. The subsurface model parameters are then updated in a direction that reduces these constraint violations. However, in the case of a rough initial model, the phase errors in the reconstructed wavefields may trap the waveform inversion in a local minimum leading to inaccurate subsurface models. In this paper, in order to avoid matching such incorrect phase information during the early WRI iterations, we design a new cost function based upon phase retrieval, namely a process which seeks to reconstruct a signal from the amplitude of linear measurements. This new formulation, called wavefield inversion with phase retrieval (WIPR), further improves the robustness of the parameter estimation subproblem by a suitable phase correction. We implement the resulting WIPR problem with an alternating-direction approach, which combines the majorization-minimization (MM) algorithm to linearise the phase-retrieval term and a variable splitting technique based upon the alternating direction method of multipliers (ADMM). This new workflow equipped with Tikhonov-Total variation (TT) regularization, which is the combination of second-order Tikhonov and total variation regularizations and bound constraints, successfully reconstructs the 2004 BP salt model from a sparse fixed-spread acquisition using a 3 Hz starting frequency and a homogeneous initial velocity model.

Key words: Inverse theory; Numerical modelling; Waveform inversion; Controlled source seismology.

1 INTRODUCTION

Full waveform inversion (FWI) is a waveform matching procedure which provides a subsurface model with a resolution reaching to the half of the smallest propagated wavelength (Virieux & Operto 2009). Ultra-long offset wide-azimuth stationary-recording acquisitions provide a varied angular illumination of the subsurface, which is amenable to the development of the broadband velocity model (Sirgue & Pratt 2004). However, the large number of propagated wavelengths generated by these acquisitions makes FWI prone to cycle skipping or phase wrapping (Shah 2014; Choi & Alkhalifah 2015). Various continuation strategies in frequencies, traveltimes and offsets can be used to mitigate this pathology (e.g. Shipp & Singh 2002; Górszczyk *et al.* 2017). However, these multiscale approaches can be tedious to implement and still require quite accurate initial velocity models and low frequencies to prevent cycle skipping at long offsets.

To increase the resilience of FWI to cycle skipping, frequency-domain wavefield reconstruction inversion (WRI) has been proposed by van Leeuwen & Herrmann (2013, 2016). The governing idea of the method relies on the following statement: if we would be able to record a monochromatic wavefield generated by a single source everywhere in the subsurface, then it would be straightforward to reconstruct the subsurface model exactly in virtue of the wave equation bilinearity (Fig. 1). Since we cannot record wavefields everywhere, the best we can do to approach the true wavefields is to find those that best match the sparse observations and satisfy the wave equation in a least-squares sense. This least-squares problem can be implemented by solving an overdetermined linear system gathering the weighted wave equation and the observation equation (the equation relating the data to the wavefield through a sparse sampling operator). The observation equation generates a wave equation relaxation whose strength depends on the accuracy of the available subsurface model and the weight assigned to

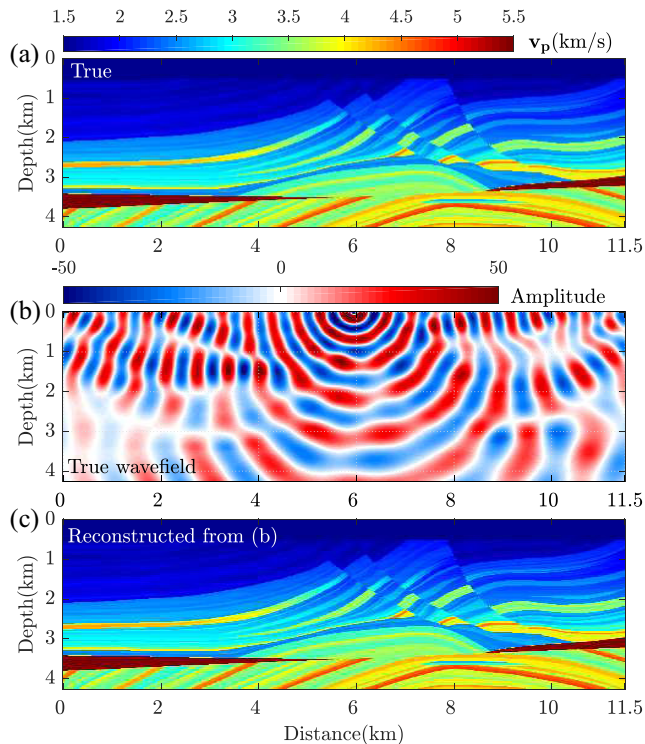


Figure 1. Illustration of the wave-equation bilinearity. (a) True Marmousi II velocity model. (b) 3-Hz wavefield (real part) for the source located at the surface and a distance of 6 km. (c) Velocity model inferred from the amplitude and phase of the monochromatic wavefield, eq. (24). The same model can be inferred from the sole amplitude of the monochromatic wavefield, eq. (25). The reconstructed velocity model exactly matches the true velocity model.

the wave equation. Then, the subsurface parameters are updated from the reconstructed wavefields by the least-squares minimization of the wave equation errors (source residuals), and this two-step cycle is iterated until both the observation equation and the wave equation are satisfied with a prescribed accuracy. Typically, a small weight is assigned to the wave equation during early iterations to foster data fitting and prevent cycle skipping accordingly. This weight is progressively increased to guarantee that the wave equation is fulfilled near the convergence point. The parameter-estimation subproblem can be solved through a variable projection approach (van Leeuwen & Herrmann 2016) or through alternating optimization (van Leeuwen & Herrmann 2013). In this latter case, the parameter-estimation subproblem is linearized around the reconstructed wavefield according to the wave-equation bilinearity.

WRI has been originally implemented with a penalty method, which requires a tedious dynamic control of the penalty parameter as above mentioned (Fu & Symes 2017). To overcome this issue, Aghamiry *et al.* (2019d) proposed to replace the penalty method with an augmented Lagrangian method equipped with variable splitting, namely the alternating-direction method of multiplier (ADMM, Boyd *et al.* 2010). Contrary to penalty methods, augmented Lagrangian methods converge to an accurate solution with a fixed penalty parameter by iteratively updating the Lagrange multipliers with the running sum of the constraint violations (Nocedal & Wright 2006, chapter 17). This defines what is sometimes referred to as an iterative refinement or defect correction algorithm. Accordingly, Aghamiry *et al.* (2019d) called their method iteratively refined (IR-) WRI.

ADMM also provides a suitable framework to implement bound constraints and ℓ_1 -based non-smooth regularizations [such as total variation (TV) regularization] in IR-WRI (Aghamiry *et al.* 2019c). IR-WRI with TV regularization was further improved by using more versatile Tikhonov-TV (TT) regularization combining second-order Tikhonov and TV regularizers through infimal convolution (Aghamiry *et al.* 2019b). Infimal convolution means that the TT regularization explicitly decomposes the model into two components of different statistical properties (a smooth one and a blocky one) such that a suitable regularization can be tailored to each component (Tikhonov and TV, Gholami & Hosseini 2013).

Although IR-WRI, when equipped with basic frequency continuation strategies and advanced regularization techniques, mitigates cycle skipping and helps to overcome non-linearity issues broadly speaking, still it can be trapped in a local minimum when the initial model is very far from the true model. In this case, the phase and amplitude of the reconstructed wavefields match those of the true wavefields only in the vicinity of the receivers, while they can be quite inaccurate elsewhere. In IR-WRI, the inaccuracies of the reconstructed wavefields are directly mapped into the subsurface model through a linear deconvolution-like processing. This implies that only the shallow part of the subsurface model is expected to be reliably updated during the early iterations of IR-WRI for surface acquisitions. It remains unclear from a theoretical viewpoint how these inaccuracies are progressively canceled out by IR-WRI in iterations through a depth-continuation process.

It is well acknowledged that the phase often has a more dominant role than amplitude in image processing and waveform inversion methods (Oppenheim & Lim 1981; Shechtman *et al.* 2015). Therefore, the inaccuracies in the phase of the reconstructed wavefields may be those which can more likely drive the inversion toward spurious local minima. Accordingly, the objective of this study is to assess whether a reformulation of the parameter-estimation subproblem that mitigates the role of the phase during the early stages of the IR-WRI at the benefit of the more robust amplitude counterpart contributes to stabilizing the inversion.

To achieve this goal, we recast the parameter-estimation subproblem of IR-WRI as a phase retrieval problem (Fienup 1982). Phase retrieval considers the fundamental problem of how to reconstruct a signal from the magnitude of linear measurements (e.g. the magnitude of its Fourier transform). The name phase retrieval arose because if the phase of the linear measurements can be retrieved, then it is easy to reconstruct the signal by solving a linear problem (Waldspurger *et al.* 2015). It has a long history of applications in science and engineering when the phase of the linear measurements cannot be recorded or is inaccurate, for example, optics (Walther 1963), X-ray crystallography (Millane 1990; Harrison 1993), astronomy (Fienup 1982). Also, it has recently attracted renewed interest in many imaging problems (Fogel *et al.* 2016), such as X-ray medical imaging (Pfeiffer *et al.* 2006; Burvall *et al.* 2011), optical imaging (Shechtman *et al.* 2015) and seismic processing (Gholami 2014).

In IR-WRI, the linear operator of the parameter estimation subproblem is formed by the so-called virtual sources (Pratt *et al.* 1998), while the right-hand sides (the linear measurements) depend on the source and the Laplacian of the wavefields (considering the Helmholtz equation as wave equation). According to the above definition of phase retrieval, we update the subsurface parameters from the magnitude of the right-hand sides rather than from their phase and amplitude. A potential downside of the phase-retrieval algorithm is ill-posedness generated by the rapidly decreasing sensitivity of the inversion with depth, in particular in large contrast or attenuating media when a small amount of the seismic energy

is transmitted in the deep subsurface. This makes the use of efficient regularization necessary, as we will show. As in IR-WRI, we implement the iterative refinement procedure in the phase retrieval algorithm to design an adaptive control on the regularization parameters and achieve a faster convergence to the desired model. Once the velocity model estimation has been stabilized by phase retrieval during a first low-frequency band inversion, we switch to classical IR-WRI to process the higher frequencies with the phase information.

We assess our phase-retrieval based waveform inversion, referred to as WIPR, against two complex synthetic case studies, the Marmousi II and the large contrast 2004 BP salt models with a sparse long-offset fixed-spread acquisition. We show how WIPR helps to reconstruct more accurately the phase of the reconstructed wavefields, which is translated into more accurate velocity estimation. Then, we show the impact of this more accurate early-stage velocity reconstruction on the convergence of IR-WRI at higher frequencies.

This paper is organized in a method, numerical and discussion sections. In the method section, we first introduce phase retrieval in a general context. Secondly, we review the principles of WRI with iterative refinement and show how we implement phase retrieval in the underlying parameter-estimation subproblem. Then, we briefly review how compound regularizations can be implemented in WIPR. In the numerical experiment section, we show the application on the Marmousi and 2004 BP salt models. In the 'Discussion' section, we further discuss the resolution power of amplitudes in WIPR and show its impact on the phase reconstruction.

2 NOTATION

The mathematical symbols adopted in this paper are as follows. We use italics for scalar quantities, boldface lower-case letters for vectors, and boldface capital letters for matrices and tensors. We use the superscript T to denote the adjoint of an operator and $*$ to show the complex conjugate. We use $\text{diag}(\bullet)$ to show a square diagonal matrix which includes vector \bullet on its main diagonal. The i th component of the column vector \mathbf{x} is shown by x_i . For a complex number $x = ae^{jb}$ with $j = \sqrt{-1}$, $|x| = a$ denotes the magnitude of x and $\angle x = b$ denotes its phase. For the n -length column vectors \mathbf{x} and \mathbf{y} , the dot product is defined by $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i^* y_i$ and their Hadamard product, denoted by $\mathbf{x} \circ \mathbf{y}$, is another vector made up of their component-wise products, that is $(\mathbf{x} \circ \mathbf{y})_i = x_i y_i$. The ℓ_1 - and ℓ_2 -norms of \mathbf{x} are, respectively, defined by $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$ and $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^n |x_i|^2}$.

3 METHOD

In this section, we briefly review the phase retrieval problem (Gerchberg 1972) and a simple majorization-minimization (MM, Lange 2016) algorithm to solve this non-convex problem. Then we review the frequency-domain FWI (Pratt *et al.* 1998), its WRI alternative (van Leeuwen & Herrmann 2016), and its improved version called iteratively refined WRI (IR-WRI) (Aghamiry *et al.* 2019d). Finally, we introduce phase retrieval in the parameter-estimation subproblem of IR-WRI. In the following of this study, we refer to IR-WRI with phase retrieval as WIPR. WIPR is solved efficiently with the alternating direction method of multipliers (ADMM) (Boyd *et al.* 2010) and MM algorithms, which provide a suitable framework to implement bound constraints and TT regularization in the parameter estimation subproblem and tackle large data sets.

3.1 Phase retrieval

Phase retrieval refers to the reconstruction of a complex signal from the amplitude of linear measurements. Consider the complex-valued linear problem $\mathbf{L}\mathbf{x} = \mathbf{y}$ with $\mathbf{L} \in \mathbb{C}^{m \times n}$, $\mathbf{x} \in \mathbb{C}^{n \times 1}$, and $\mathbf{y} \in \mathbb{C}^{m \times 1}$ and assume that only the amplitude of the right-hand side, $|\mathbf{y}|$, can be measured or is reliable. The corresponding phase-retrieval problem can be written as (Gholami 2014)

$$\min_{\mathbf{x}} f(\mathbf{x}) = \min_{\mathbf{x}} \frac{1}{2} \|\|\mathbf{L}\mathbf{x}| - |\mathbf{y}|\|_2^2. \quad (1)$$

Problem (1) is non-convex where this non-convexity finds its root in removing the phase of the right-hand side ($\|\mathbf{L}\mathbf{x}| = |\mathbf{y}|$ is non-unique and there are many \mathbf{x} that can fit the magnitude of \mathbf{y}).

Several algorithms have been proposed to solve the problem ranging from the Gerchberg–Saxton algorithm (Gerchberg 1972) as a first algorithm to more modern optimization techniques (Candes *et al.* 2013; Netrapalli *et al.* 2013; Eldar *et al.* 2016). In this paper, we present a simple algorithm based on the MM technique (Appendix A) to find a local minimum of $f(\mathbf{x})$ iteratively. Based on the proposed algorithm, the minimizer of the convex surrogate function

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{L}\mathbf{x} - |\mathbf{y}|e^{j\angle \mathbf{L}\mathbf{x}^k}\|_2^2. \quad (2)$$

converges to the minimizer of $f(\mathbf{x})$ as iteration number, k , tends to infinity. The quadratic problem 2 is the same as the least-squares norm of the original system $\mathbf{L}\mathbf{x} = \mathbf{y}$, when the phase of \mathbf{y} is replaced with the phase of $\mathbf{L}\mathbf{x}^k$, which is extracted from the solution of k 'th iteration.

3.2 FWI versus wavefield inversion

The reduced formulation of frequency-domain FWI can be written as (Pratt *et al.* 1998; Plessix 2006)

$$\min_{\mathbf{m}} \|\mathbf{P}\mathbf{u}(\mathbf{m}) - \mathbf{d}\|_2^2, \quad (3)$$

where \mathbf{d} denotes the recorded seismic data, \mathbf{P} is the linear observation operator that samples the wavefield $\mathbf{u}(\mathbf{m})$ at the receiver positions, and \mathbf{m} denotes the subsurface parameters. The wavefield $\mathbf{u}(\mathbf{m})$ is the solution of the wave equation

$$\mathbf{u}(\mathbf{m}) = \mathbf{A}(\mathbf{m})^{-1}\mathbf{b}, \quad (4)$$

where \mathbf{b} is the source term and $\mathbf{A}(\mathbf{m}) \in \mathbb{C}^{n \times n}$ is the discretized wave-equation operator. In this study, the wave equation is the Helmholtz equation, whose operator $\mathbf{A}(\mathbf{m})$ is given by

$$\mathbf{A}(\mathbf{m}) = \mathbf{\Delta} + \omega^2 \mathbf{C} \text{diag}(\mathbf{m}) \mathbf{B}, \quad (5)$$

where ω is the angular frequency, \mathbf{m} contains the squared slownesses, $\mathbf{\Delta}$ is the discretized Laplace operator, \mathbf{C} introduces boundary conditions (e.g. sponge-like absorbing boundary conditions such as perfectly matched layers (Bérenger 1994)) and \mathbf{B} spreads the 'mass' term $\omega^2 \mathbf{C} \text{diag}(\mathbf{m})$ over all the coefficients of the stencil to improve its accuracy following an antilumped mass strategy (Chen *et al.* 2013).

A main drawback of the reduced formulation in eq. (3) is that the objective function depends on the model parameters \mathbf{m} through the oscillating inverse operator $\mathbf{A}^{-1}(\mathbf{m})$. This makes the inverse problem highly non-linear, and hence prone to convergence to inaccurate minimizer when the initial \mathbf{m} is not accurate enough.

An alternative objective function as proposed by van Leeuwen & Herrmann (2016) is defined as

$$\min_{\mathbf{m}} \|\mathbf{A}(\mathbf{m})\mathbf{u}(\mathbf{m}) - \mathbf{b}\|_2^2 \quad (6)$$

in which the wavefield is given by

$$\mathbf{u}(\mathbf{m}) = (\lambda\mathbf{A}(\mathbf{m})^T\mathbf{A}(\mathbf{m}) + \mathbf{P}^T\mathbf{P})^{-1} (\lambda\mathbf{A}(\mathbf{m})^T\mathbf{b} + \mathbf{P}^T\mathbf{d}). \quad (7)$$

The wavefield $\mathbf{u}(\mathbf{m})$ is the solution of an augmented wave equation system, which gathers the wave equation $\mathbf{A}(\mathbf{m})\mathbf{u}(\mathbf{m}) = \mathbf{b}$ weighted by the scalar penalty parameter $\lambda > 0$ and the observation equation $\mathbf{P}\mathbf{u}(\mathbf{m}) = \mathbf{d}$ (van Leeuwen & Herrmann 2013, their eq. 6)).

A main advantage of eq. (6) over eq. (3) is that its objective function depends on the model parameters \mathbf{m} through a regularized inverse of the wave-equation operator $(\lambda\mathbf{A}(\mathbf{m})^T\mathbf{A}(\mathbf{m}) + \mathbf{P}^T\mathbf{P})^{-1}$, where the regularizer injects the prior information on the true wavefields, namely their sparse measurements \mathbf{d} , through the sampling operator \mathbf{P} . This wavefield reconstruction driven by the observations introduces a relaxation of the wave equation, namely source residuals, which are minimized to update the model parameters, eq. (6), such that the wavefields, eq. (7), are pushed back towards the wave equation.

The minimization problem, eq. (6), can be solved with a variable projection approach by enforcing the closed-form expression of \mathbf{u} , eq. (7), in the objective function, eq. (6) (van Leeuwen & Herrmann 2016), or with an alternating-direction strategy for \mathbf{u} and \mathbf{m} through a Gauss–Seidel like iteration (van Leeuwen & Herrmann 2013). We will follow the latter option, where the model estimation problem, eq. (6), reduces to a quadratic optimization problem

$$\mathbf{m}^{k+1} = \arg \min_{\mathbf{m}} \|\mathbf{A}(\mathbf{m})\mathbf{u}^k - \mathbf{b}\|_2^2, \quad (8)$$

which, in virtue of the wave equation bilinearity (van Leeuwen & Herrmann 2013; Aghamiry *et al.* 2019d), can be recast as

$$\mathbf{m}^{k+1} = \min_{\mathbf{m}} \|\mathbf{L}(\mathbf{u}^k)\mathbf{m} - \mathbf{y}(\mathbf{u}^k)\|_2^2, \quad (9)$$

with

$$\begin{cases} \mathbf{L}(\mathbf{u}^k) = \omega^2 \mathbf{C} \text{diag}(\mathbf{B}\mathbf{u}^k), \\ \mathbf{y}(\mathbf{u}^k) = \mathbf{b} - \Delta\mathbf{u}^k. \end{cases} \quad (10)$$

and $\mathbf{u}^k \equiv \mathbf{u}(\mathbf{m}^k)$.

During wavefield reconstruction, eq. (7), a small value of λ is used during the early iterations to foster the data fit, hence mitigating cycle skipping, at the expense of the fidelity with which the wave equation is satisfied. Then, λ is progressively increased in iterations such that the wave equation is satisfied at the convergence point (Fu & Symes 2017). To avoid the tedious dynamic control of penalty parameters and converge in an automatic way towards accurate solution with fixed λ , Aghamiry *et al.* (2019d) introduced iteratively refined WRI (IR-WRI). IR-WRI relies on an iterative refinement procedure, which finds its root in augmented Lagrangian method (Nocedal & Wright 2006, chapter 17). In the end, it simply consists of updating at the end of each iteration the right-hand sides of the wave equation and the observation equation, namely the data \mathbf{d} and the source \mathbf{b} , with the running sum of the data and source residuals in iteration

$$\mathbf{b}^{k+1} = \mathbf{b}^k + \mathbf{b} - \mathbf{A}(\mathbf{m}^{k+1})\mathbf{u}^{k+1}, \quad (11a)$$

$$\mathbf{d}^{k+1} = \mathbf{d}^k + \mathbf{d} - \mathbf{P}\mathbf{u}^{k+1}. \quad (11b)$$

In the framework of linear inverse problems, \mathbf{d}^{k+1} and \mathbf{b}^{k+1} record the wavefield and model refinements performed at previous iterations to refine the wavefields and the model parameters at the current

iteration from the residual errors only. This iterative refinement or defect correction (Böhmer & Stetter 1984) procedure, which is lacking in WRI, is the key ingredient which controls the accuracy of the wavefield and model solutions at the convergence point when fixed λ is used. The reader is also referred to Gholami *et al.* (2018); Gholami & Naeini (2019) for other recent applications of iterative refinement in the field of geophysics.

IR-WRI has shown promising results even for complicated velocity models when equipped with frequency continuation strategies and appropriate regularization (Aghamiry *et al.* 2019b,c). However, when the initial model is far from the true model, the phase of the reconstructed wavefields becomes inaccurate in areas that are located far away from the receivers. These phase inaccuracies can lead to an inaccurate model reconstruction of these regions, hence trapping the inversion into a local minimum. In order to mitigate this issue, we show in the next section how to solve the quadratic optimization problem, eq. (9), with phase retrieval.

3.3 Wavefield inversion with phase retrieval (WIPR)

To estimate the model parameters without involving inaccurate phase information, we propose to replace the IR-WRI parameter-estimation problem given by eq. (9) by the following phase retrieval problem

$$\mathbf{m}^{k+1} = \arg \min_{\mathbf{m}} \|\mathbf{L}(\mathbf{u}^k)\mathbf{m} - |\mathbf{y}(\mathbf{u}^k)|\|_2^2. \quad (12)$$

In order to solve (12), we replace it by a surrogate majorizing function (as described in Section 3.1 and Appendix A)

$$\mathbf{m}^{k+1} = \arg \min_{\mathbf{m}} \|\mathbf{L}(\mathbf{u}^k)\mathbf{m} - |\mathbf{y}(\mathbf{u}^k)|e^{j\angle\mathbf{L}(\mathbf{u}^k)\mathbf{m}^k}\|_2^2. \quad (13)$$

This objective function is now quadratic and admits a closed form solution.

In order to better understand why phase retrieval helps to improve the solution, we give a tentative interpretation hereafter. From eq. (7) it is seen that, at each iteration, the wavefield approximately satisfies the wave equation

$$\mathbf{A}(\mathbf{m}^k)\mathbf{u}^k \approx \mathbf{b} \quad (14)$$

or

$$\mathbf{L}(\mathbf{u}^k)\mathbf{m}^k \approx \mathbf{y}(\mathbf{u}^k), \quad (15)$$

where the approximation level is controlled by the parameter λ and the approximation becomes an equality as $\lambda \rightarrow \infty$. The next iterate \mathbf{m}^{k+1} is then found such that it minimizes the wave-equation errors generated by the wave-equation relaxation, eq. (8). In this way, simple minimization via eq. (8) forces the model to match both the amplitude and phase information of $\mathbf{y}(\mathbf{u}^k)$. In order to reduce the imprint of the phase errors, a phase correction or phase alignment step can be applied to eq. (15) before updating the model parameters:

$$\min_{\phi} \|\mathbf{L}(\mathbf{u}^k)\mathbf{m}^k - |\mathbf{y}(\mathbf{u}^k)|e^{j\phi}\|_2^2. \quad (16)$$

The optimal minimizer of eq. (16) is given by $\phi = \angle\mathbf{L}(\mathbf{u}^k)\mathbf{m}^k$. Using the optimal ϕ , the next iterate is found by minimizing the corrected quadratic problem, as presented in eq. (13). In this way, the solution of the inverse problem is less affected by the phase errors in \mathbf{u} and thus hopefully less prone to convergence to a local minimum. The reader is referred to Jiang *et al.* (2017) and Qian *et al.* (2017) for other applications of phase retrieval with alternating optimization.

Furthermore, we apply the iterative refinement procedure to WIPR in the same way as for IR-WRI by updating \mathbf{b} and \mathbf{d} according to eqs (11a) and (11b) as outlined in the Algorithm 1. This implies that, although we recast the parameter-estimation subproblem as a phase retrieval problem (Algorithm 1, Line 5), we use both the phase and amplitude of the source and data residuals to update the right-hand sides of the wavefield-reconstruction and parameter-estimation subproblems (Algorithm 1, Lines 8 and 9).

Algorithm 1 WIPR algorithm

```

1: Input:  $\mathbf{m}^0$ 
2: initialize:  $\mathbf{A} \leftarrow \Delta + \omega^2 \mathbf{C} \text{diag}(\mathbf{m}^0) \mathbf{B}, \tilde{\mathbf{b}}^0 \leftarrow \mathbf{0}, \mathbf{d}^0 \leftarrow \mathbf{0}$ 
3: while convergence conditions not satisfied do
4:    $\mathbf{u}^{k+1} \leftarrow [\mathbf{P}^T \mathbf{P} + \lambda \mathbf{A}^T \mathbf{A}]^{-1} [\mathbf{P}^T (\mathbf{d} + \mathbf{d}^k) + \lambda \mathbf{A}^T (\mathbf{b} + \mathbf{b}^k)]$ 
5:    $\mathbf{L} \leftarrow \omega^2 \mathbf{C} \text{diag}(\mathbf{B} \mathbf{u}^{k+1})$ 
      $\tilde{\mathbf{y}} \leftarrow |\mathbf{b} + \mathbf{b}^k - \Delta \mathbf{u}^{k+1}| e^{j \angle \mathbf{L} \mathbf{m}^k}$ 
6:    $\mathbf{m}^{k+1} \leftarrow [\mathbf{L}^T \mathbf{L}]^{-1} \mathbf{L}^T \tilde{\mathbf{y}}$ 
7:    $\mathbf{A} \leftarrow \Delta + \omega^2 \mathbf{C} \text{diag}(\mathbf{B} \mathbf{m}^{k+1}) \mathbf{B}$ 
8:    $\mathbf{b}^{k+1} \leftarrow \mathbf{b}^k + \mathbf{b} - \mathbf{A} \mathbf{u}^{k+1}$ 
9:    $\mathbf{d}^{k+1} \leftarrow \mathbf{d}^k + \mathbf{d} - \mathbf{P} \mathbf{u}^{k+1}$ 
10: end while

```

3.4 Tikhonov-Total variation (TT) regularized WIPR with bounding constraints

In this section, we implement TT regularization and bounding constraints in WIPR to stabilize the updates of the model parameters. By adding a convex regularization term $\|\mathbf{m}\|_{\text{TT}}$ to the cost function of WIPR, we get the following optimization for the \mathbf{m} -subproblem

$$\mathbf{m}^{k+1} = \arg \min_{\mathbf{m} \in \mathcal{C}} \|\mathbf{m}\|_{\text{TT}} + \lambda \|\mathbf{L}(\mathbf{u}^k) \mathbf{m} - |\mathbf{y}(\mathbf{u}^k)| e^{j \angle \mathbf{L}(\mathbf{u}^k) \mathbf{m}^k}\|_2^2, \quad (17)$$

where $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^{n \times 1} \mid \mathbf{m}_l \leq \mathbf{x} \leq \mathbf{m}_u\}$ is the set of all feasible models bounded by the lower bound \mathbf{m}_l and the upper bound \mathbf{m}_u and $\|\mathbf{m}\|_{\text{TT}}$ is the TT regularization functional as defined by Aghamiry *et al.* (2019b,c)

$$\|\mathbf{m}\|_{\text{TT}} = \min_{\mathbf{m}=\mathbf{m}_1+\mathbf{m}_2} \|\mathbf{m}_1\|_{\text{TV}} + \alpha \|\mathbf{m}_2\|_{\text{Tikh}}, \quad (18)$$

with

$$\begin{cases} \|\mathbf{m}\|_{\text{TV}} = \sum \sqrt{|\nabla_x \mathbf{m}|^2 + |\nabla_z \mathbf{m}|^2}, \\ \|\mathbf{m}\|_{\text{Tikh}} = \sum (|\nabla_{xx} \mathbf{m}|^2 + 2|\nabla_{xz} \mathbf{m}|^2 + |\nabla_{zz} \mathbf{m}|^2), \end{cases} \quad (19)$$

in which the sum runs over all element, ∇_i are first order difference operators in direction i and ∇_{ij} are second-order differential operators in directions i and j .

The TT regularization is an infimal convolution-based combination of the second-order Tikhonov and TV regularizations and is suitable for recovering piecewise-smooth models (Gholami & Hosseini 2013; Aghamiry *et al.* 2019b). It decomposes the model \mathbf{m} explicitly into two components, \mathbf{m}_1 and \mathbf{m}_2 , which have different statistical properties. Here, the blocky component \mathbf{m}_1 and the smooth component \mathbf{m}_2 are captured by the TV and the Tikhonov regularizations, respectively.

Subproblem (17) is solved with ADMM and the splitting scheme, which decouples the ℓ_1 and ℓ_2 components and bound constraints of the function through the introduction of auxiliary variables and solves each related subproblem in sequence (Goldstein & Osher 2009). By doing this, we come up with a least-squares problem with a closed-form expression to update \mathbf{m}_1 and \mathbf{m}_2 , a proximity subproblem to update the auxiliary variables and a gradient ascent

step to update dual variables. The reader is referred to Aghamiry *et al.* (2019b, c, section 2.2.2) for the detailed derivation of the algorithm.

4 NUMERICAL EXAMPLES

We assess our workflow against the Marmousi II model (Martin *et al.* 2006) and the large-contrast 2004 BP salt model (Billette & Brandsberg-Dahl 2004) when the inversion is started from a homogeneous velocity model and a 3 Hz frequency such that a considerable phase error is generated. In both cases, we consider a long-offset fixed-spread acquisition and perform the frequency-domain inversion in the 3–13 Hz frequency band with a frequency interval of 0.5 Hz. Small batches of two frequencies with one frequency overlapping between two consecutive batches are successively inverted following a classical frequency continuation strategy. We apply the phase retrieval algorithm only during the first frequency batch processing to stabilize the inversion. Then, we proceed with regular IR-WRI at higher frequencies. The IR-WRI results obtained from an early WIPR stage will be referred to as IR-WRI_{pr} to prevent confusions with results obtained when IR-WRI is applied from the first frequency-batch inversion. During IR-WRI/IR-WRI_{pr}, we perform three paths through the frequency batches to improve the inversion results, using the final model of one path as the initial model of the next one (these cycles can be viewed as outer iterations of the algorithms). The starting and finishing frequencies of the paths are [3.5, 6], [4, 8.5], [6, 13] Hz, respectively, where the first element of each pair shows the starting frequency and the second one is the finishing frequency. We will compare the results of IR-WRI_{pr} and IR-WRI to assess the improvements resulting from the initial WIPR step. We will also compare the results of IR-WRI and IR-WRI_{pr} when regularization is applied or not such that the relative role of phase retrieval and regularization can be discriminated. Namely, we will illustrate that the role of phase retrieval is to extend the linear regime of the waveform inversion further when an inaccurate starting model is used, while regularization injects suitable prior to manage non-uniqueness of the solution in the poorly illuminated area. When regularization is applied, the same tuning is used during WIPR and IR-WRI, and during IR-WRI and IR-WRI_{pr}. For all the numerical tests, we use a 9-point finite-difference staggered-grid stencil with PML boundary condition along the four edges of the model (no free-surface boundary condition is used) and anti-lumped mass to solve the Helmholtz equation, where the stencil coefficients are optimized to the frequency (Chen *et al.* 2013).

4.1 Marmousi II model

The Marmousi II model covers a 11 500 m \times 4200 m spatial domain (Fig. 1a). The fixed-spread surface acquisition consists of 56 sources spaced 200 m apart, and 230 receivers spaced 50 m apart at the surface. The source signature is a 10 Hz Ricker wavelet, and the wavespeed in the homogeneous starting model is 3 km s⁻¹.

We first compare the results of WIPR and IR-WRI after 45 iterations of the [3,3.5] Hz frequency batch inversion when bound constraints are applied without any additional regularization (Figs 2a and b). Bound constraints are used for all of the tests of this section from the first iteration. WIPR manages to reconstruct an accurate long-wavelength approximation of the Marmousi model (Fig. 2b), while IR-WRI clear fails to capture the kinematic trend of the model (Fig. 2a). This statement is clearly illustrated by the direct comparison between the true model, the initial model, and the reconstructed

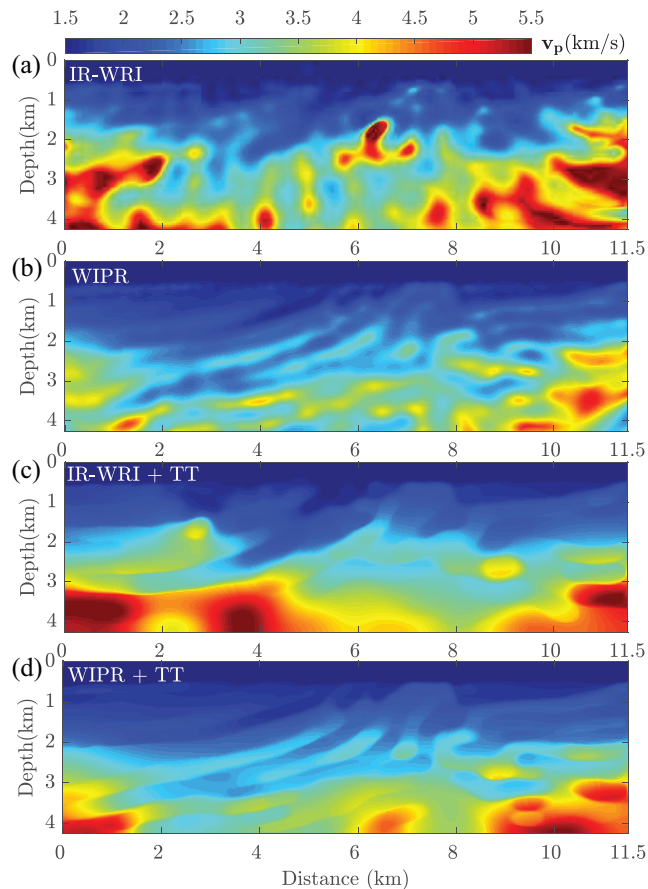


Figure 2. Marmousi II example. Velocity models after the [3,3.5] Hz inversion. (a) IR-WRI, (b) WIPR, (c) TT regularized IR-WRI and (d) TT regularized WIPR. Bound constraints are used for all of the tests from the first iteration.

IR-WRI and WIPR models along with three vertical logs at horizontal distances of 4, 8.2 and 10 km (Fig. 3a). We also apply TT regularization during WIPR and IR-WRI (Figs 2c–d and 3b). The TT regularization improves the WIPR and IR-WRI results in the poorly illuminated area, although it does not remove high-velocity artefacts near the left-bottom end of the IR-WRI velocity model (Fig. 3b, $x = 4$ km).

Since the model parameters are updated from the reconstructed wavefields, the accuracy of these wavefields is a visible indicator of the reliability of the estimated model parameters. We compare the real part of the 3 Hz wavefields reconstructed by IR-WRI and WIPR without and with TT regularization (Fig. 4) with the true wavefield (Fig. 1b) for a source located at the horizontal distance of 6 km. Both IR-WRI and WIPR wavefields are in phase with the true wavefield along the receiver line since we tune the penalty parameter such that the data are matched from the first iteration. However, the WIPR wavefield matches much better the true wavefield than the IR-WRI counterpart in depth. When regularization is used, the accuracy of the wavefield reconstruction is mostly improved in the poorly illuminated zones, mainly near the bottom ends of the model.

We continue with IR-WRI at higher frequencies using the final IR-WRI and WIPR models of the [3, 3.5] Hz inversion (Fig. 2) as initial models. The stopping criterion of iterations is a maximum of

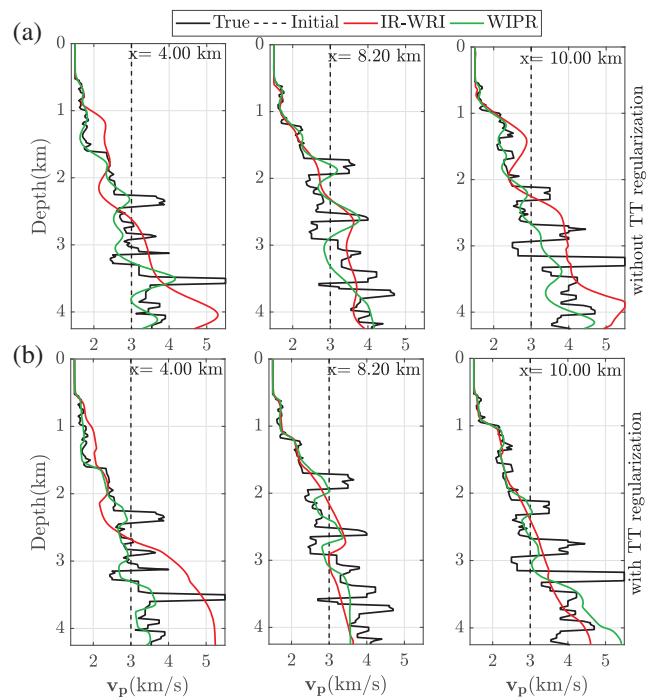


Figure 3. Marmousi II example. Direct comparison along with three vertical logs at $x = 4, 8.2,$ and 10 km between the true velocity model (black), the initial model (dashed black) and the IR-WRI (red) and WIPR (green) models shown in Fig. 2. (a) Without TT regularization. (b) With TT regularization.

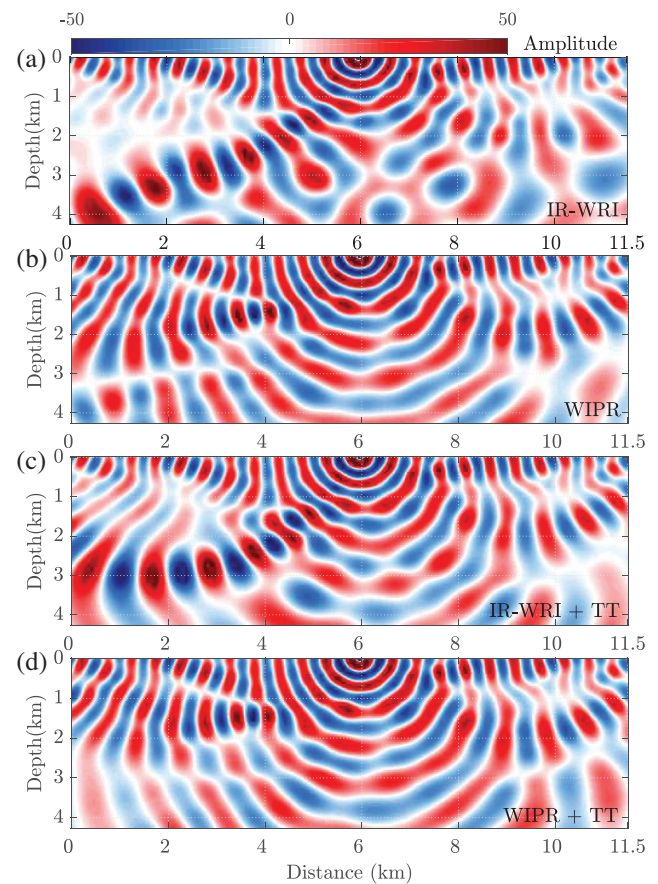


Figure 4. Marmousi II example. Three-Hz wavefields (real part) reconstructed after the first frequency batch inversion for the source located at 6 km. (a) IR-WRI, (b) WIPR, (c) TT regularized IR-WRI and (d) TT regularized WIPR. The corresponding velocity models are shown in Fig. 2. These wavefields can be compared with the true one shown in Fig. 1(b).

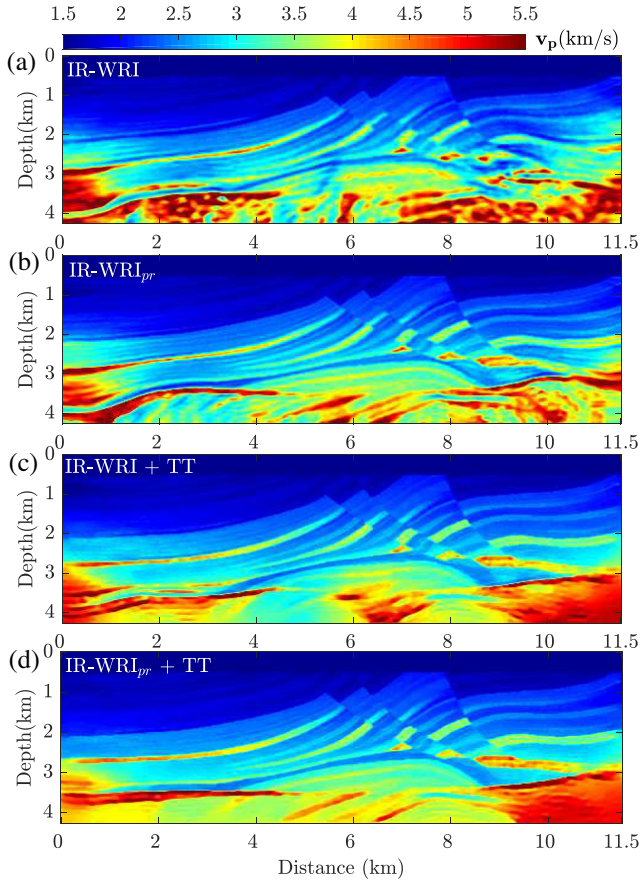


Figure 5. Marmousi II example. Final velocity models obtained by IR-WRI (a and c) and IR-WRI_{pr} (b and d). (a and b) Without TT regularization. (c and d) With TT regularization.

30 iterations per batch or

$$\left(\sum \| \mathbf{A}(\mathbf{m}^k) \mathbf{u}^k - \mathbf{b} \|_2 \leq 10^{-3} \text{ and } \sum \| \mathbf{P} \mathbf{u}^k - \mathbf{d} \|_2 \leq 10^{-5} \right), (20)$$

where summation is over all the sources and frequencies of the batch. The final IR-WRI and IR-WRI_{pr} velocity models are shown in Fig. 5, while a direct comparison between them and the true model are plotted in Fig. 6 at distances 4, 8.2 and 10 km. IR-WRI performed a total of 382 and 369 iterations without and with TT regularization, respectively (Figs 5a and c), while IR-WRI_{pr} performed a total of 201 and 234 iterations (Figs 5b and d). The smaller number of iteration required to satisfy the stopping criterion of iterations performed by IR-WRI_{pr} compared to IR-WRI highlights the role of the initial model built by WIPR during the [3, 3.5] Hz inversion (Figs 2b and d). Without regularization, IR-WRI_{pr} outperforms IR-WRI in the deep part, namely near the reservoir at around 8 km distance and near the salt layers near the ends of the models. Regularization improves significantly both IR-WRI and IR-WRI_{pr} results in these areas. However, IR-WRI_{pr} still provides superior results in the deep part.

4.2 The large contrast 2004 BP salt model

We consider now a target of the large contrast 2004 BP salt model located on the left side of the model (Fig. 7a). The 2004 BP salt model mainly consists of a simple sediment background with a complex rugose multivalued salt body, subsalt slow velocity anomalies related to overpressure zones, and a fast velocity anomaly to the

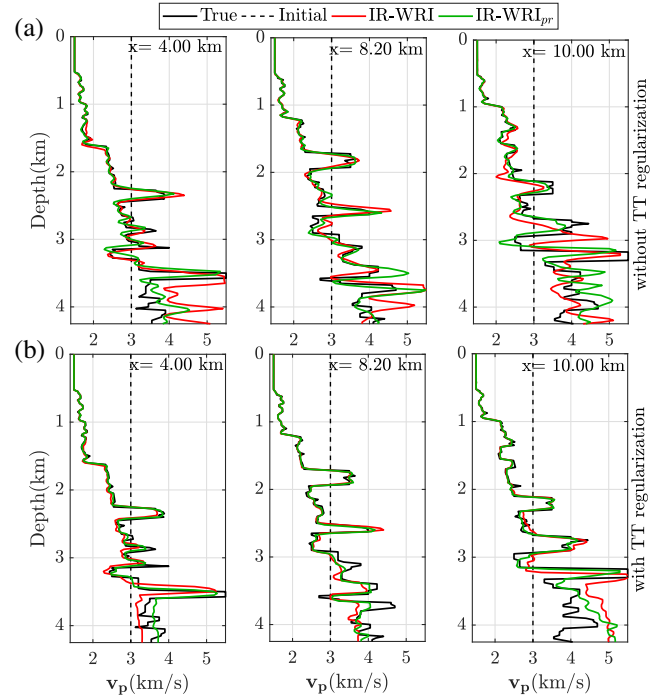


Figure 6. Marmousi II example. Direct comparison at $x = 4, 8.2$ and 10 km between the true velocity model (black), the initial model (dashed line) and the models estimated by IR-WRI (red) and IR-WRI_{pr} (green) (Fig. 5). (a) Without TT regularization. (b) With TT regularization.

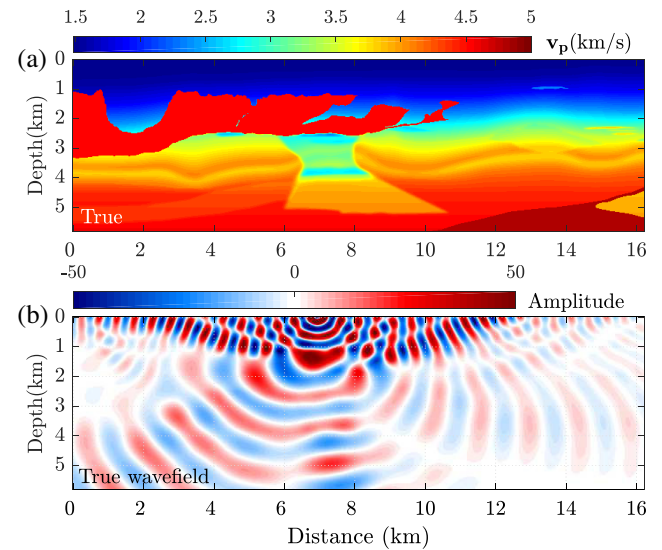


Figure 7. (a) True 2004 BP salt velocity model. (b) Real part of the 3-Hz wavefield computed in (a) for the source located at 7 km.

right of the salt body. The selected subsurface model is 16 250 m wide and 5825 m deep and is discretized with a 25 m grid interval. The surface fixed-spread acquisition consists of 66 sources spaced 250 m apart, and 131 receivers spaced 125 m apart. We perform forward modelling using a 10 Hz Ricker wavelet as a source signature. The starting velocity model for inversion is homogeneous with a velocity of 3 km s^{-1} .

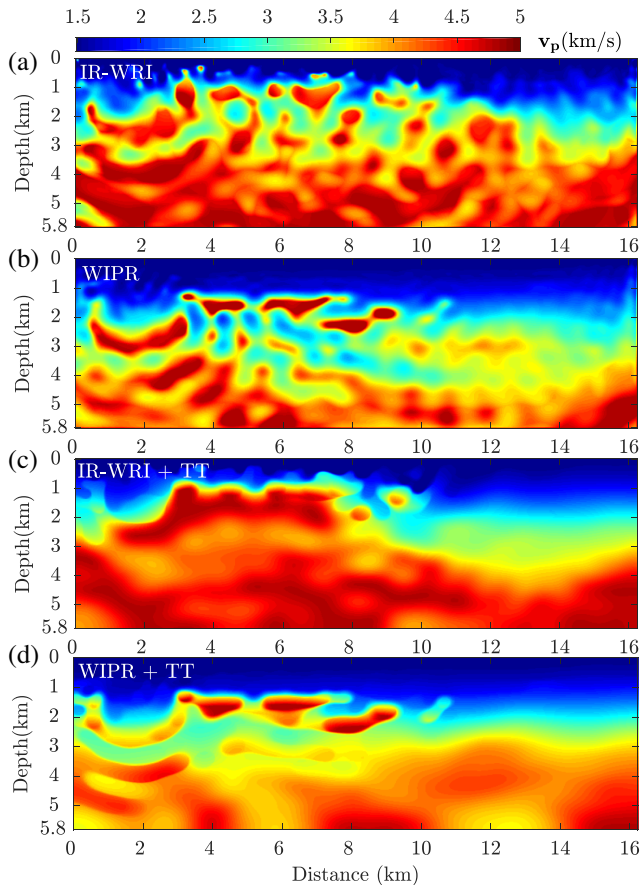


Figure 8. BP salt example. Velocity models inferred from the inversion of the [3,3.5] Hz frequencies. (a) IR-WRI, (b) WIPR, (c) IR-WRI with TT regularization and (d) WIPR with TT regularization.

4.2.1 IR-WRI versus WIPR at low frequencies

We first perform 45 iterations of bound-constrained WIPR and IR-WRI for a starting [3,3.5] Hz frequency batch. Note that we turn on the bound constraints after 20 iterations to discriminate their role from those of the TT regularization and optimization scheme (see later discussion in the text). The velocity models reconstructed by bound-constrained IR-WRI and WIPR with and without TT regularization are shown in Fig. 8, while the direct comparison between the true model, the initial model, and the models reconstructed by IR-WRI and WIPR are shown in Fig. 9 along three vertical logs at horizontal distances of 3.5, 7.7 and 12 km. The most striking feature is that IR-WRI overestimates velocities above the salt and mispositions the top of the salt layers accordingly, while WIPR captures the top salt much more accurately (Fig. 9a, green versus red curves, $x = 3.5$ km and 7.7 km). This overestimation of the shallow velocities by IR-WRI is also shown in the right part of the model away from the salt (Fig. 9a, red curve, $x = 12$ km). On the other hand, WIPR fails to reconstruct the bottom part of the thick salt body and overprints unrealistic variations on the subsalt area when TT regularization is not used (Figs 8b and 9a, green curve, $x = 3.5$ km). These artefacts are likely generated by the increased ill-posedness of WIPR with depth, which itself results from the lack of phase information. This ill-posedness is however, efficiently mitigated by the TT regularization (Fig. 8d). The TV component of the TT regularization helps to sharpen the top and bottom of the salt layers (Fig. 9b, green curve, $x = 7.7$ km), recovers better the

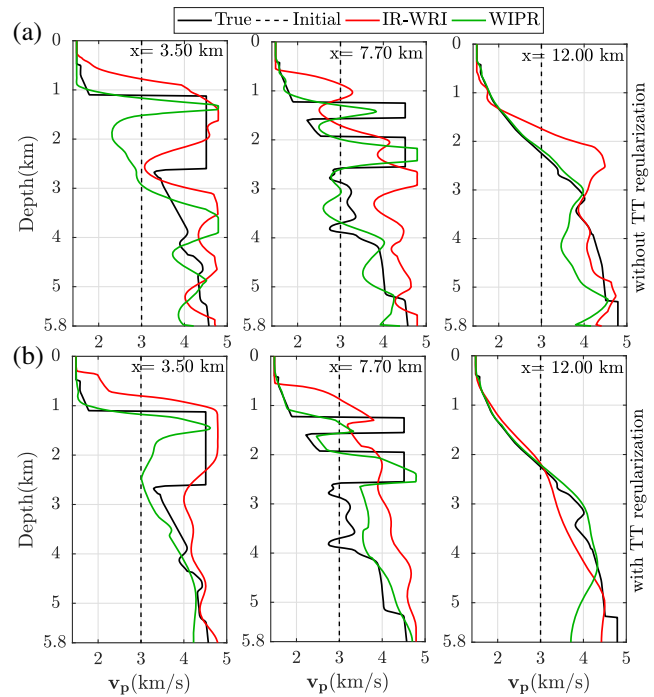


Figure 9. BP salt example. Direct comparison along logs located at $x = 3.5$, 7.7 and 12 km between the true velocity model (black), the initial model (dashed line) and the models estimated by IR-WRI (red) and WIPR (green) shown in Fig. 8. (a) Without TT regularization. (b) With TT regularization.

geometry of the salt layer (Fig. 9b, green curve, $x = 3.5$ km), while the Tikhonov component stabilizes the inversion in-depth and drives it toward smooth reconstruction of the poorly illuminated subsalt structures (Fig. 9b, green curve, $x = 12$ km). During IR-WRI, TT regularization fails to achieve these goals (Fig. 8c) because the mispositioning in depth of the velocity structures generated by the increased non-linearity of IR-WRI are too heavy to be corrected by the regularization (Fig. 9b, red curve).

We compare the true 3-Hz wavefield shown in Fig. 7(b) with the wavefields reconstructed by IR-WRI and WIPR without and with TT regularization (Fig. 10). Like the Marmousi test, both IR-WRI and WIPR wavefields are in phase with the true wavefield along the receiver line since we tune the penalty parameter such that the data are matched from the first iteration. However, the WIPR wavefields match much better the true wavefield than the IR-WRI counterpart in the first two kilometers of the model. This is pointed by the arrows in Fig. 10, which show that the IR-WRI wavefield transmitted into the salt body is out-of-phase with respect to the true wavefield, unlike the WIPR wavefield.

We also computed time-domain seismograms in the TT-regularized IR-WRI and WIPR models (Figs 8c and d) with a 10-Hz Ricker wavelet for a source located at 15.8 km (Fig. 11). We show quite significant traveltimes mismatches at long offsets between the true seismograms and those computed in the IR-WRI model (Fig. 11a), which result from the overestimated velocities above the salt shown in Fig. 8(c), while the match between the true seismograms and those computed in the WIPR model is far better. This poor data fit achieved by IR-WRI highlights both the inaccuracies of the reconstructed velocity model and the limited accuracy with which the wave equation has been satisfied at the convergence point.

This is further supported by Fig. 12, which shows the relative model error (Fig. 12a) and the joint evolution of the data misfit

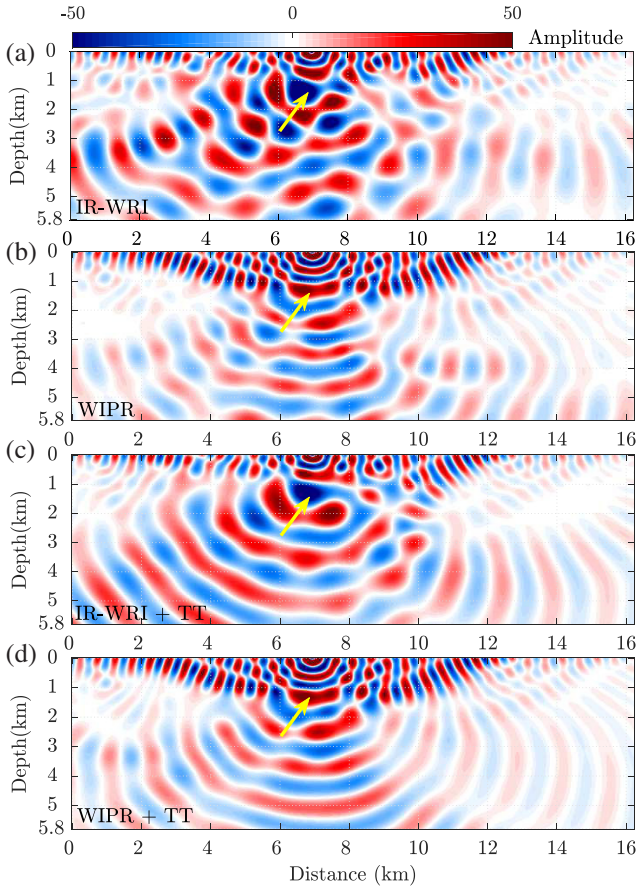


Figure 10. BP salt example. Three-Hz wavefields (real part) reconstructed after the first frequency batch inversion. (a) IR-WRI, (b) WIPR, (c) TT regularized IR-WRI and (d) TT regularized WIPR. The corresponding velocity models are shown in Fig. 8. The source is located at 7 km distance. These wavefields can be compared with the true wavefield shown in Fig. 7(b). The arrows point where the incident wavefields interact with the top salt.

and wave-equation error in iterations (Fig. 12b). Also, the evolution of the data misfit and wave-equation error in iterations are shown separately in Figs 12(c) and (d), respectively. First, these error curves show that WIPR converges toward a more accurate velocity model than IR-WRI (Fig. 12a). Accordingly, it shows that WIPR satisfies more accurately the wave-equation constraint than IR-WRI (Fig. 12d), although the parameter estimation has been recast as a phase retrieval problem rather than as a wave-equation error minimization. Fig. 12a also shows that bound constraints have the least impact on the TT regularized WIPR (blue curve) since the descent direction is rather continuous around iteration 20 (when bound constraints are activated). This highlights the more stable descent direction followed by TT regularized WIPR relative to IR-WRI.

4.2.2 IR-WRI versus IR-WRI_{pr}

We continue with IR-WRI at higher frequencies using the final IR-WRI and WIPR models of the [3, 3.5] Hz inversion (Fig. 8) as initial models. The stopping criterion of iterations is 10 iterations per batch. The final velocity models built by IR-WRI and IR-WRI_{pr} are shown in Fig. 13, while a direct comparison between them and the true model are plotted in Fig. 14 at distances 3.5, 7.7 and 12 km. Let's first focus on the results obtained without TT regularization (Figs 13a and b). IR-WRI apparently reconstructs the salt body with

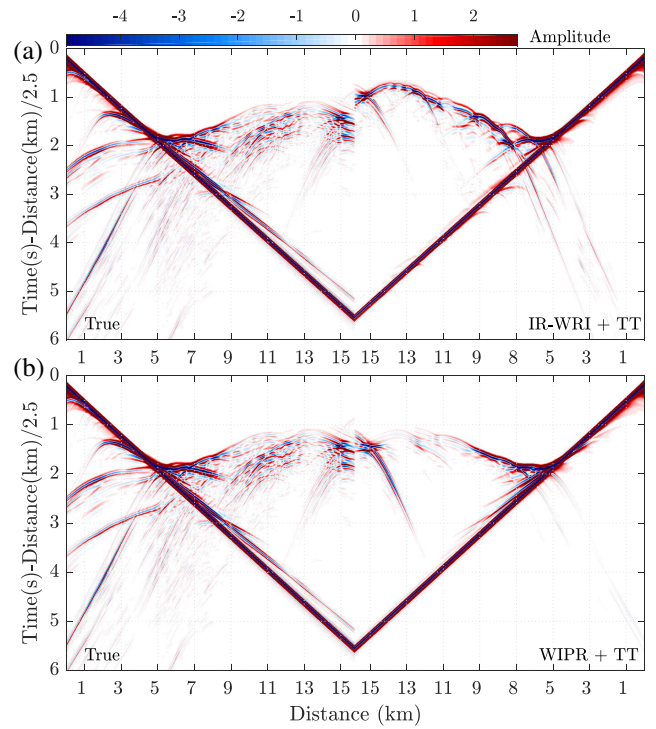


Figure 11. BP salt example. Comparison between time-domain seismograms computed in the true velocity model (left-hand panel) and those computed in the velocity models obtained by (a) IR-WRI (Fig. 8c) and (b) WIPR (Fig. 8d). The seismograms are plotted with a reduction velocity of 2.5 km s^{-1} . Note the significant traveltimes mismatch in (a) at long offset.

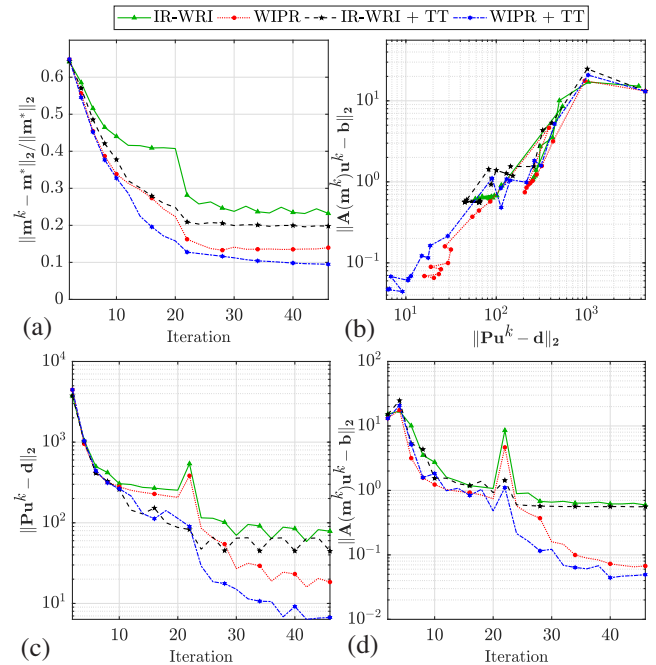


Figure 12. BP salt example. Convergence path of [3,3.5] Hz inversion. (a) Evolution of $\| \mathbf{m}^k - \mathbf{m}^* \|_2 / \| \mathbf{m}^* \|_2$ during iterations (\mathbf{m}^* denotes the true model). (b) Convergence history of the algorithm in the $(\| \mathbf{P}\mathbf{u}^k - \mathbf{d} \|_2 - \| \mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b} \|_2)$ plane. (c and d) Evolution of (c) $\| \mathbf{P}\mathbf{u}^k - \mathbf{d} \|_2$, (d) $\| \mathbf{A}(\mathbf{m}^k)\mathbf{u}^k - \mathbf{b} \|_2$ during iterations for IR-WRI (green), WIPR (red), TT regularized IR-WRI (black) and TT regularized WIPR (blue). Note that the data and PDE misfit values are summed over all the sources and frequencies.

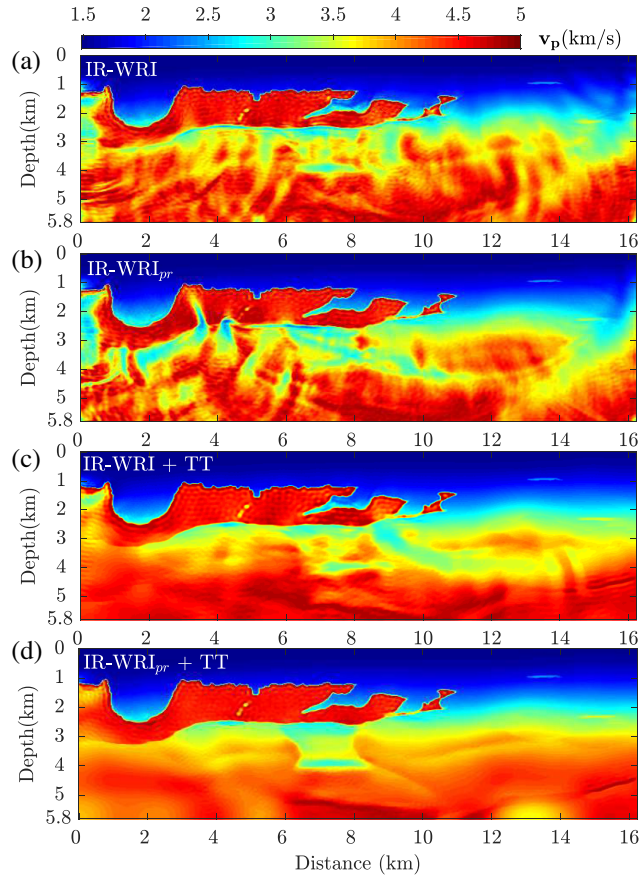


Figure 13. BP salt example. Final velocity models estimated by IR-WRI (a, c) and IR-WRI_{pr} (b, d). (a–b) Without TT regularization. (c–d) With TT regularization.

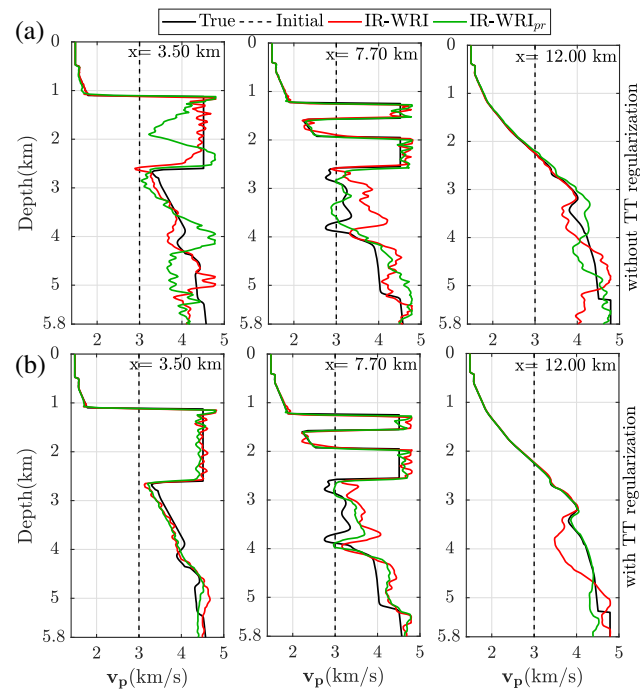


Figure 14. BP salt example. Direct comparison along logs located at $x = 3.5, 7.7$ and 12 km between the true velocity model (black), the initial model (dashed line) and the models estimated by IR-WRI (red) and IR-WRI_{pr} (green) shown in Fig. 13. (a) Without TT. (b) With TT regularization.

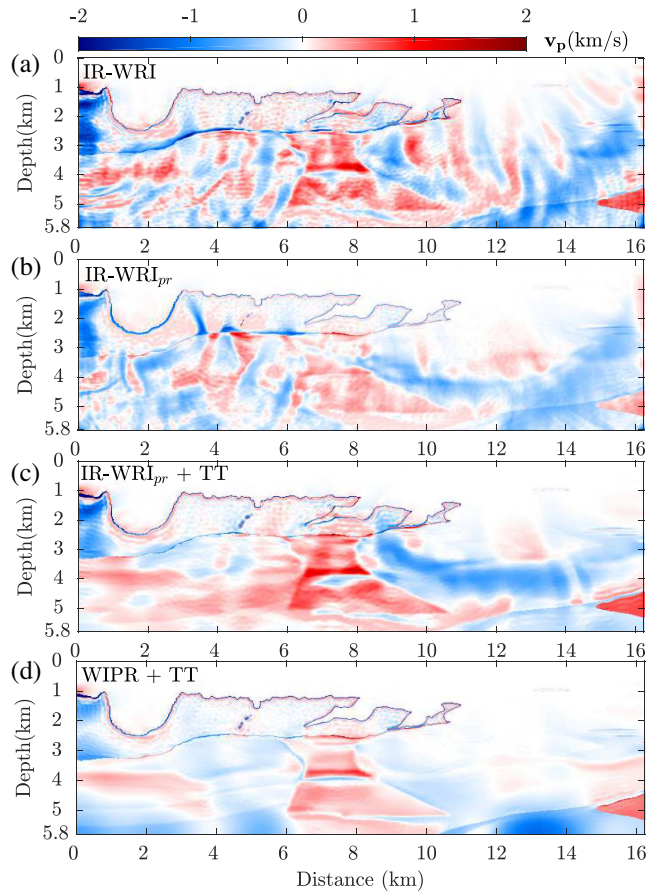


Figure 15. BP salt example. Difference between the true velocity model and the velocity models shown in Fig. 13. (a, c) IR-WRI. (b, d) IR-WRI_{pr}. (a–b) Without TT regularization. (c–d) With TT regularization.

good accuracy (Fig. 13a), although the shallow kinematic inaccuracies of the starting IR-WRI model shown in Figs 8(a) and 9, red curves. This highlights the potential of the search-space expansion implemented in IR-WRI to manage cycle skipping. The final IR-WRI_{pr} velocity model shows low-velocity artefacts in the salt body at around 3 km distance (Fig. 13b), which are probably inherited from the underestimated velocities below the top of the salt shown in Figs 8(b) and 9(a), green curve, $x = 3.5$ km. For both inversions, the reconstruction of the subsalt structures are quite noisy. To gain more quantitative insights on the relative accuracy of the IR-WRI and IR-WRI_{pr} models, we show the difference between these velocity models and the true model in Figs 15(a) and (b) and outline the model error (ME) for the velocity models as well as the least-squares data and source misfit at the convergence point in Table 1. Figs 15(a) and (b) clearly shows that the IR-WRI_{pr} model is as a whole more accurate than the IR-WRI counterpart, although the above-mentioned local artefacts.

The TT regularization removes the local artefacts shown in the salt body of the IR-WRI_{pr} model and improves the subsalt imaging (Fig. 13d) more significantly than for the IR-WRI model (Fig. 13c). The difference between the true model and the IR-WRI/IR-WRI_{pr} models (Figs 15c–d) and the model error (ME) provided in Table 1 outline the superior accuracy of the IR-WRI_{pr} model relative to the IR-WRI counterpart. These relative model accuracies are consistent with the data and source misfit achieved by IR-WRI_{pr} and IR-WRI at the convergence point (Table 1).

Table 1. Model error (ME) for the velocity models and data and source misfit (summed over all the sources and frequencies) at the end of the first frequency batch inversion and at the convergence point of the four inversion tests (with or without phase retrieval, with or without TT regularization). Here, ME is defined as $100 \frac{\|\mathbf{m} - \mathbf{m}_*\|_1}{\|\mathbf{m}_*\|_1}$, where \mathbf{m}_* denotes the true model.

		ME	$\ \mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{b}\ _2$	$\ \mathbf{P}\mathbf{u} - \mathbf{d}\ _2$
Iteration No. 45	IR-WRI	23.23	0.5954	7.8550
	WIPR	13.97	0.0672	1.8450
	IR-WRI+TT	19.81	0.5640	4.4630
	WIPR+TT	9.49	0.04926	0.6706
Final iteration	IR-WRI	7.57	0.2046	2.8911
	IR-WRI _{pr}	6.50	0.0647	1.6009
	IR-WRI+TT	4.89	0.1272	1.6470
	IR-WRI _{pr} +TT	3.16	0.0359	0.6350

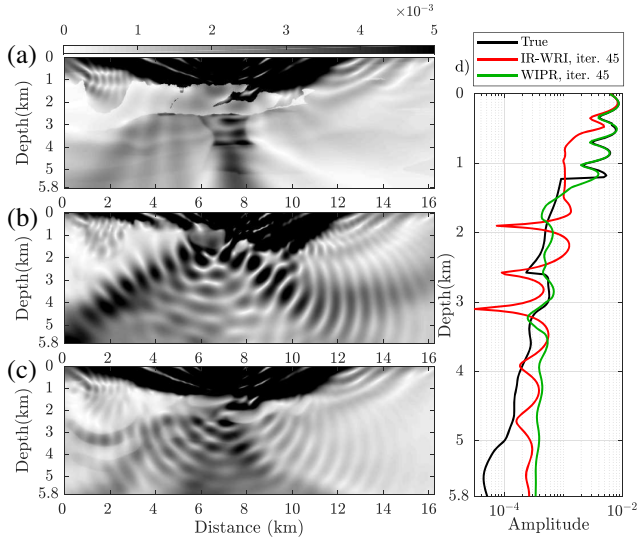


Figure 16. Amplitude of $\Delta\mathbf{u}$ for 2004 BP model. (a) The wavefield \mathbf{u} is computed in the true velocity model for the 3 Hz frequency. (b and c) The wavefields are computed in the velocity models inferred from TT-regularized (b) IR-WRI (Fig. 8c) and (c) WIPR (Fig. 8d) for the first frequency batch. (d) Comparison between vertical profiles extracted from (a) (black line), (b) (red line), (c) (green line) at a distance of 6.5 km.

5 DISCUSSION

With the large contrast BP salt model, we have shown that the phase inaccuracies of the reconstructed wavefields prevent accurate estimation of the shallow sediments and accurate positioning in depth of the top of the salt after the first frequency batch inversion. In contrast, WIPR reconstructs more accurately the velocity gradient in the sedimentary cover as well as the sharp velocity contrast on top of the salt. This accurate positioning of the top salt is indeed important to speed up the convergence and improve the solution at higher frequencies.

To further illustrate the sensitivity of IR-WRI to wavefield amplitudes and phases in large contrast media, we show in Figs 16 and 17 the amplitude and phase of $\Delta\mathbf{u}$ for the 3 Hz frequency and a source located at 7 km of distance. In these figures, $\Delta\mathbf{u}$ have been computed in the true velocity model, and in the TT regularized IR-WRI and WIPR velocity models obtained after the first frequency batch inversion (Figs 8c and d). We remind that $\Delta\mathbf{u}$ is one of the terms forming the right-hand side $\tilde{\mathbf{y}}$ of the quadratic parameter-estimation subproblem together with the updated source $\mathbf{b} + \mathbf{b}^k$,

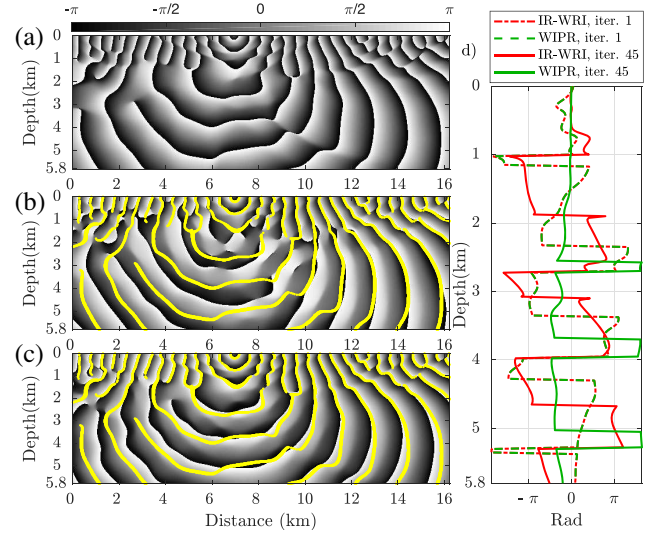


Figure 17. Phase of $\Delta\mathbf{u}$ for 2004 BP model. Same as Fig. 16 for the phase of $\Delta\mathbf{u}$. (d) Difference between the phases computed in the true model and the reconstructed models at the first iteration (dashed red and green lines), after the first frequency batch inversion for IR-WRI (red line) and WIPR (green line) at a distance of 6.5 km. Note the phase mismatch between the phase computed in the true model and in the IR-WRI model below 0.5 km depth.

Algorithm 1, Line 5. As such, $\Delta\mathbf{u}$ can be viewed as the observable of the parameter-estimation subproblem.

The amplitudes of $\Delta\mathbf{u}$ computed in the true velocity model show the dominant imprint of the reflection and the refraction from the top of the salt (Fig. 16a, shallow black zone). Also, the geometry of the seismically transparent salt body is almost perfectly delineated. This shows that the amplitudes of $\Delta\mathbf{u}$ intrinsically embed high-resolution information on the subsurface.

This high-resolution potential can be simply illustrated with the wave-equation bilinearity as discussed below. Let us remind the wave equation for a single source and a single frequency

$$\Delta\mathbf{u} + \omega^2 \text{diag}(\mathbf{m})\mathbf{u} = \mathbf{b} \quad (21)$$

which leads to the following linear equation of the parameter-estimation subproblem:

$$\underbrace{\omega^2 \text{diag}(\mathbf{m})\mathbf{u}}_{\mathbf{L}(\mathbf{u})\mathbf{m}} = \underbrace{\mathbf{b} - \Delta\mathbf{u}}_{\mathbf{y}(\mathbf{u})}. \quad (22)$$

Compared to eq. (5), we drop the matrix \mathbf{B} and \mathbf{C} which are not important for the discussion. The equality implies that $|\mathbf{L}(\mathbf{u})\mathbf{m}| = |\mathbf{y}(\mathbf{u})|$ and $\angle\mathbf{L}(\mathbf{u})\mathbf{m} = \angle\mathbf{y}(\mathbf{u})$. If \mathbf{u} is the true wavefield (if it would be recorded everywhere) then we would be able to recover the true model exactly by minimizing

$$\begin{aligned} \|\mathbf{L}(\mathbf{u})\mathbf{m} - \mathbf{y}(\mathbf{u})\|_2^2 &= \|(|\mathbf{L}(\mathbf{u})\mathbf{m}| - |\mathbf{y}(\mathbf{u})|)e^{j\angle\mathbf{L}(\mathbf{u})\mathbf{m}}\|_2^2 \\ &= \|\mathbf{L}(\mathbf{u})\mathbf{m} - |\mathbf{y}(\mathbf{u})|e^{j\angle\mathbf{L}(\mathbf{u})\mathbf{m}}\|_2^2. \end{aligned} \quad (23)$$

It is seen that, basically, the true model is the solution of a weighted phase retrieval problem (as indicated by the middle term) where the complex exponential of the phase serves as the weight.

Indeed, for an approximate wavefield which is obtained from eq. (7) with a rough velocity model or a large penalty parameter the equality in eq. (23) is not necessarily satisfied because, in this case, $\angle\mathbf{L}(\mathbf{u})\mathbf{m} \neq \angle\mathbf{y}(\mathbf{u})$. The proposed WIPR algorithm simply forces this equality to be satisfied at each iteration by aligning the phase of \mathbf{y} with that of $\mathbf{L}(\mathbf{u})\mathbf{m}$ (right term in eq. 23).

The ability to reconstruct the true model from amplitude information only can be illustrated more explicitly if we assume that \mathbf{m} is real in eq. (21) (non-attenuating medium). The matrix $\mathbf{L}(\mathbf{u})$ is diagonal and contains the so-called virtual source (eq. 22).

$$\mathbf{m} = \frac{1}{\omega^2} \left(\frac{1}{\mathbf{u}} \right) \circ \mathbf{y}(\mathbf{u}), \quad (24)$$

where \circ denotes element wise multiplication. Indeed, we can equally reconstruct the true \mathbf{m} from the magnitude of \mathbf{y} as

$$\mathbf{m} = \frac{1}{\omega^2} \left(\frac{1}{|\mathbf{u}|} \right) \circ |\mathbf{y}(\mathbf{u})|, \quad (25)$$

where $1/\mathbf{u}$ and $1/|\mathbf{u}|$ are element-wise reciprocal of \mathbf{u} and $|\mathbf{u}|$, respectively. The Marmousi II model computed with eqs (24) and (25) by using the true wavefield (Fig. 1b) is shown in Figs 1(c) and matches exactly the true one (Fig. 1a).

The Fig. 16(c), as well as the vertical profile extracted at 6.5 km of distance, confirm that WIPR manages to match the amplitudes of the true $\Delta\mathbf{u}$ down to the top of the salt at 1.25 km depth, while IR-WRI (Fig. 16b) allows for an amplitude fit down to a maximum depth of 0.5 km only, consistently with the accuracy of the velocity fields reconstructed by IR-WRI and WIPR (Figs 8c and d).

It is also instructive to look at the phase of $\Delta\mathbf{u}$ (Fig. 17). To facilitate the comparison of the phases computed in the true model (Fig. 17a) and in the reconstructed models (Figs 17b and c), we superimpose in Figs 17(b) and (c) the contours of the phase computed in the true model at the spatial positions where the phase wraparounds, namely when it jumps from $\pm\pi$ to $\mp\pi$ (Figs 17b and c). We clearly show that WIPR reconstructs the phase of $\Delta\mathbf{u}$ much more accurately than IR-WRI around the salt body. This is further illustrated by the difference between the phases computed in the true model and the reconstructed models along the vertical profile at 6.5 km distance (Fig. 17d). The phase difference associated with IR-WRI (solid red curve) reaches a value of $-\pi$ at the top of the salt (1.25 km depth) showing that at these depths the two wavefields are cycle skipped, while the phase differences associated with WIPR (solid green curve) remains far below this limit down to around 2.7 km depth.

Conversely, it is worth noting that IR-WRI matches better the phase of the true wavefield than WIPR in the smooth bottom-right part of the model (Compare Figs 17b and c). This highlights that the sensitivity of the inversion to amplitudes decreases rapidly with depth due to lack of illumination, geometrical spreading effects and energy partitioning at interfaces. We also want to stress that, although we update the wave speeds from the amplitudes of \mathbf{y} , we update \mathbf{y} at each iteration with the phase and amplitude of the wave equation error (the source residuals) in the framework of the ADMM optimization (Algorithm 1). This right-hand side updating re-injects at each iteration the phase error as a defect correction term in the optimization.

For sake of completeness, we show also the amplitude and phase of $\Delta\mathbf{u}$ for the Marmousi case study in Figs 18 and 19. Fig. 18 shows how WIPR reproduces more accurately the wide-angle reflection from the dipping layer on the left-hand side of the source than IR-WRI. Fig. 19 shows that WIPR better reconstructs the true phase of $\Delta\mathbf{u}$ than IR-WRI.

Although the application of the phase retrieval algorithm has been limited to the scalar Helmholtz equation in this study, the method has been formulated such that more complex physics and multi-parameter reconstruction can be viewed (the reader is referred to Aghamiry *et al.* (2019e) and Aghamiry *et al.* (2019a) for application of IR-WRI in visco-acoustic VTI media). The key feature allowing

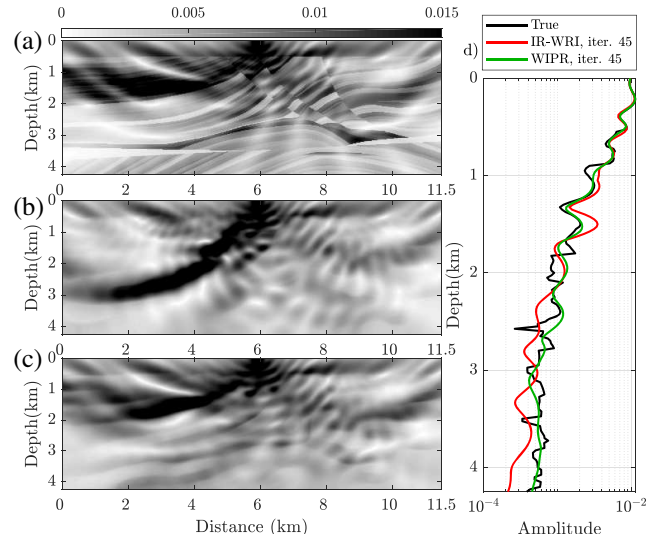


Figure 18. Amplitude of $\Delta\mathbf{u}$ for Marmousi II model. (a) The wavefield \mathbf{u} is computed in the true velocity model for the 3 Hz frequency. (b–c) The wavefields are computed in the velocity models inferred from TT-regularized (b) IR-WRI (Fig. 2c) and (c) WIPR (Fig. 2d) for the first frequency batch. (d) Comparison between vertical profiles extracted from (a) (black line), (b) (red line), (c) (green line) at a distance of 6.0 km.

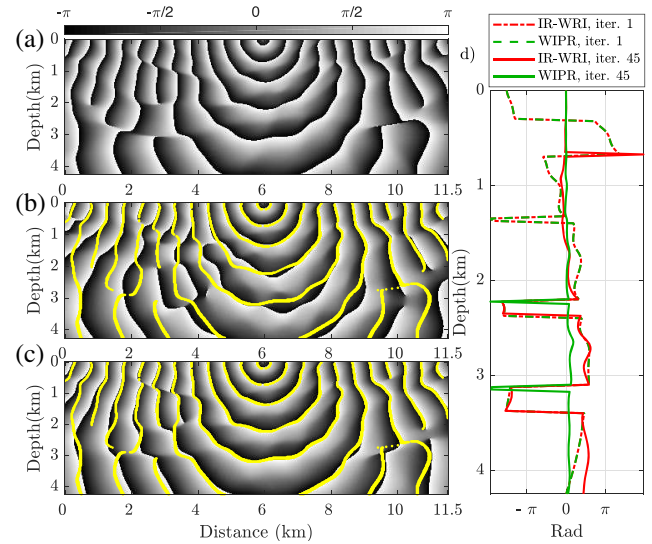


Figure 19. Phase of $\Delta\mathbf{u}$ for Marmousi II model. Same as Fig. 18 for the phase of $\Delta\mathbf{u}$. (d) Difference between the phases computed in the true model and the reconstructed models at the first iteration (dashed red and green lines), after the first frequency batch inversion for IR-WRI (red line) and WIPR (green line) at a distance of 6.0 km.

for these extensions is the linearization of the non-convex phase retrieval problem performed by the MM approach, as reviewed in Appendix A.

6 CONCLUSIONS

We present a preliminary application of bound-constrained and TT regularized WIPR. Using the large contrast BP salt model, we show that, when the inversion is started from scratch (homogeneous initial velocity model), WIPR improves significantly the reconstruction of the sedimentary background and the top of the salt during the early

iterations of the inversion (that is, when the wavefields are not yet accurately reconstructed far away from the receivers). Sparsity-promoting regularization is a necessary ingredient to stabilize the phase retrieval inversion at greater depths where the amplitude information becomes more challenging to extract. Sparsity-promoting regularization combined with phase retrieval maybe also two complementary tools to relax the need for finely sampled stationary-recording acquisitions. Phase retrieval can also be implemented in classical FWI. Also, WIPR can be theoretically extended to multiparameter reconstruction in attenuating media. Further theoretical and numerical works are, however, necessary to flesh out the potential and limits of phase retrieval in the FWI technology.

ACKNOWLEDGEMENTS

We would like to thank editors Huajian Yao, Louise Alexander and reviewers Hejun Zhu and an anonymous for their comments which help improving the paper. This study was partially funded by the SEISCOPE consortium (<http://seiscope2.osug.fr>), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC and TOTAL. This study was granted access to the HPC resources of SIGAMM infrastructure (<https://www.oca.eu/fr/mesocentre-sigamm>) and CINES/IDRIS/TGCC under the allocation A0050410596 made by GENCI.

REFERENCES

- Aghamiry, H., Gholami, A. & Operto, S., 2019a. ADMM-based multiparameter wavefield reconstruction inversion in VTI acoustic media with TV regularization, *Geophys. J. Int.*, **219**(2), 1316–1333.
- Aghamiry, H., Gholami, A. & Operto, S., 2019b. Compound regularization of full-waveform inversion for imaging piecewise media, *IEEE Trans. Geosci. Remote Sens.*, **58**(2), 1192–1204.
- Aghamiry, H., Gholami, A. & Operto, S., 2019c. Implementing bound constraints and total-variation regularization in extended full waveform inversion with the alternating direction method of multiplier: application to large contrast media, *Geophys. J. Int.*, **218**(2), 855–872.
- Aghamiry, H., Gholami, A. & Operto, S., 2019d. Improving full-waveform inversion by wavefield reconstruction with alternating direction method of multipliers, *Geophysics*, **84**(1), R139–R162.
- Aghamiry, H., Gholami, A. & Operto, S., 2019e. Joint estimation of velocity and attenuation by frequency-domain TV-regularized wavefield reconstruction inversion, in *Proceedings of the 81th Annual EAGE Meeting (London) - WS01: Attenuation: Challenges in Modelling and Imaging at the Exploration Scale*.
- Béranger, J.-P., 1994. A perfectly matched layer for absorption of electromagnetic waves, *J. Comput. Phys.*, **114**, 185–200.
- Billette, F.J. & Brandsberg-Dahl, S., 2004. The 2004 BP velocity benchmark, in *Extended Abstracts, 67th Annual EAGE, Conference & Exhibition, Madrid, Spain*, p. B035.
- Böhmer, K. & Stetter, H., 1984. *Defect Correction Methods: Theory and Applications*, Vol. 5, Springer-Verlag Wien.
- Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J., 2010. Distributed optimization and statistical learning via the alternating direction of multipliers, *Found. Trends Mach. Learn.*, **3**(1), 1–122.
- Burvall, A., Lundstrom, U., Takman, P.A.C., Larsson, D.H. & Hertz, H.M., 2011. Phase retrieval in X-ray phase-contrast imaging suitable for tomography, *Opt. Express*, **19**(11), 10 359–10 376.
- Candes, E.J., Strohmer, T. & Vershynski, V., 2013. Phaselift: exact and stable signal recovery from magnitude measurements via convex programming, *Commun. Pure appl. Math.*, **66**(8), 1241–1274.
- Chen, Z., Cheng, D., Feng, W. & Wu, T., 2013. An optimal 9-point finite difference scheme for the Helmholtz equation with PML, *Int. J. Numer. Anal. Model.*, **10**(2), 389–410.
- Choi, Y. & Alkhalifah, T., 2015. Unwrapped phase inversion with an exponential damping, *Geophysics*, **80**(5), 251–264.
- Eldar, Y.C., Hammen, N. & Mixon, D.G., 2016. Recent advances in phase retrieval, *IEEE Signal Process. Mag., Lecture Notes*, **33**(5), 158–162.
- Fienup, J.R., 1982. Phase retrieval algorithms: a comparison, *Appl. Opt.*, **21**(15), 2758–2769.
- Fogel, F., Waldspurger, I. & d’Aspremont, A., 2016. Phase retrieval for imaging problems, *Math. Program. Comput.*, **8**(3), 311–335.
- Fu, L. & Symes, W.W., 2017. A discrepancy-based penalty method for extended waveform inversion, *Geophysics*, **R282–R298**, 78–82.
- Gerchberg, R.W., 1972. A practical algorithm for the determination of phase from image and diffraction plane pictures, *Optik*, **35**, 237–246.
- Gholami, A., 2014. Phase retrieval through regularization for seismic problems, *Geophysics*, **79**(5), V153–V164.
- Gholami, A. & Hosseini, S.M., 2013. A balanced combination of Tikhonov and total variation regularizations for reconstruction of piecewise-smooth signals, *Signal Process.*, **93**, 1945–1960.
- Gholami, A. & Zabihi Naeini, E., 2019. 3D Dix inversion using bound-constrained TV regularization, *Geophysics*, **84**(3), 1–43.
- Gholami, A., Aghamiry, H. & Abbasi, M., 2018. Constrained nonlinear AVO inversion using Zoeppritz equations, *Geophysics*, **83**(3), R245–R255.
- Goldstein, T. & Osher, S., 2009. The split Bregman method for L1-regularized problems, *SIAM J. Imag. Sci.*, **2**(2), 323–343.
- Górszczyk, A., Operto, S. & Malinowski, M., 2017. Toward a robust workflow for deep crustal imaging by FWI of OBS data: the eastern Nankai trough revisited, *J. geophys. Res.*, **122**(6), 4601–4630.
- Harrison, R.W., 1993. Phase problem in crystallography, *J. Opt. Soc. Am. A*, **10**(5), 1046–1055.
- Jiang, X., So, H.C. & Liu, X., 2017. Robust phase retrieval via ADMM with outliers, arXiv:1702.06157v1.
- Kreutz-Delgado, K., 2009. The complex gradient operator and the calculus, preprint arXiv:0906.4835.
- Lange, K., 2016. *MM Optimization Algorithms*, Vol. **147**, SIAM.
- Martin, G.S., Wiley, R. & Marfurt, K.J., 2006. Marmousi2: an elastic upgrade for Marmousi, *Leading Edge*, **25**(2), 156–166.
- Meyer, C.D., 2000. *Matrix Analysis and applied Linear Algebra*, Vol. **71**, SIAM.
- Millane, R.P., 1990. Phase retrieval in crystallography and optics, *J. Opt. Soc. Am. A*, **7**(3), 394–411.
- Netrapalli, P., Jain, P. & Sanghavi, S., 2013. Phase retrieval using alternating minimization, in *Advances in Neural Information Processing Systems*, pp. 2796–2804.
- Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization*, 2nd Edition, Springer.
- Oppenheim, A.V. & Lim, J.S., 1981. The importance of phase in signals, *Proc. IEEE*, **69**(5), 529–541.
- Pfeiffer, F., Weitkamp, T., Bunk, O. & David, C., 2006. Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources, *Nat. Phys.*, **2**, 258–261.
- Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.*, **167**(2), 495–503.
- Pratt, R.G., Shin, C. & Hicks, G.J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.*, **133**, 341–362.
- Qian, C., Fu, X., Sidiropoulos, N.D., Huang, L. & Xie, J., 2017. Inexact alternating optimization for phase retrieval in the presence of outliers, *IEEE Trans. Signal Process.*, **65**(22), 6069–6082.
- Shah, N.K., 2014. Seismic full waveform inversion for wrapped and unwrapped phase, *PhD thesis*, Imperial College London.
- Shechtman, Y., Eldar, Y.C., Cohen, O., Chapman, H.N., Miao, J. & Segev, M., 2015. Phase retrieval with application to optical imaging: a contemporary overview, *IEEE Signal Process. Mag.*, **32**(3), 87–109.
- Shipp, R.M. & Singh, S.C., 2002. Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data, *Geophys. J. Int.*, **151**, 325–344.
- Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics*, **69**(1), 231–248.

- van Leeuwen, T. & Herrmann, F., 2016. A penalty method for PDE-constrained optimization in inverse problems, *Inverse Problems*, **32**(1), 1–26.
- van Leeuwen, T. & Herrmann, F.J., 2013. Mitigating local minima in full-waveform inversion by expanding the search space, *Geophys. J. Int.*, **195**(1), 661–667.
- Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Waldspurger, I., d’Aspremont, A. & Mallat, S., 2015. Phase recovery, maxcut and complex semidefinite programming, *Math. Program.*, **149**(1–2), 47–81.
- Walther, A., 1963. The question of phase retrieval in optics, *Opt. Acta: Int. J. Opt.*, **10**(1), 41–49.
- Wu, C.J. *et al.*, 1983. On the convergence properties of the EM algorithm, *Anna. Stat.*, **11**(1), 95–103.

APPENDIX A: MAJORIZATION-MINIMIZATION TO SOLVE PHASE RETRIEVAL PROBLEM

Here, we present a simple algorithm to solve phase retrieval problem based on the majorization-minimization (MM, Lange 2016) technique. The governing idea of MM is to find the minimum of non-convex/convex function $f(\mathbf{x})$ via the iterative minimization of a simpler convex surrogate function $g(\mathbf{x}, \mathbf{x}^k)$ that majorizes $f(\mathbf{x})$ at step k (i.e. $g(\mathbf{x}, \mathbf{x}^k) \geq f(\mathbf{x})$). Fig. A1 shows a schematic of the MM process. The non-convex function $f(\mathbf{x})$ is shown in blue while a few surrogate functions $g(\mathbf{x}, \mathbf{x}^k)$ for points $\mathbf{x}^k, k \in \{0, 1, 2, 3\}$, are shown in orange. This figure shows how the iterative MM algorithm approaches a local minimum of $f(\mathbf{x})$ through the minimum of easy to minimize surrogate functions $g(\mathbf{x}, \mathbf{x}^k)$.

The phase-retrieval problem for linear system $\mathbf{L}\mathbf{x} = \mathbf{y}$ can be written as (Gholami 2014)

$$\min_{\mathbf{x}} f(\mathbf{x}) = \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{L}\mathbf{x} - \mathbf{y}\|_2^2, \quad (\text{A1})$$

where $\mathbf{L} \in \mathbb{C}^{m \times n}$, $\mathbf{x} \in \mathbb{C}^{n \times 1}$, and $\mathbf{y} \in \mathbb{C}^{m \times 1}$. Problem (A1) is non-convex where this non-convexity finds its root in removing the phase of the right hand side.

Let us introduce the auxiliary variable $\mathbf{z} = \mathbf{L}\mathbf{x}$, then the objective function in (A1) can be written as

$$\tilde{f}(\mathbf{z}) = \frac{1}{2} \|\sqrt{\mathbf{z}\mathbf{z}^*} - \mathbf{y}\|_2^2. \quad (\text{A2})$$

Using the Wirtinger calculus (Kreutz-Delgado 2009), the quadratic approximation of $f(\mathbf{x})$ around an initial guess \mathbf{x}^k is

$$g(\mathbf{x}, \mathbf{x}^k) = f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)^T \Delta \mathbf{x} + \Delta \mathbf{x}^T \mathbf{H}(\mathbf{x}^k) \Delta \mathbf{x}, \quad (\text{A3})$$

where $\Delta \mathbf{x} = \mathbf{x} - \mathbf{x}^k$, ∇f is the gradient of $f(\mathbf{x})$ and \mathbf{H} denotes the Hessian, both evaluated at \mathbf{x}^k . Simple algebra shows that the gradient of $f(\mathbf{x})$ is given by

$$\nabla f(\mathbf{x}) = 2\mathbf{L}^T \frac{\partial \tilde{f}(\mathbf{z})}{\partial \mathbf{z}^*} \quad (\text{A4})$$

$$= \mathbf{L}^T \text{diag} \left(\mathbf{1} - \frac{|\mathbf{y}|}{|\mathbf{L}\mathbf{x}|} \right) \mathbf{L}\mathbf{x} \quad (\text{A5})$$

$$= \mathbf{L}^T (\mathbf{L}\mathbf{x} - |\mathbf{y}|e^{j\angle \mathbf{L}\mathbf{x}}), \quad (\text{A6})$$

where $\mathbf{1}$ is an all-ones vector, and the Hessian operator is given by

$$\begin{aligned} \mathbf{H}(\mathbf{x}) &= 2\mathbf{L}^T \left(\frac{\partial^2 \tilde{f}(\mathbf{z})}{\partial \mathbf{z} \partial \mathbf{z}^*} + \frac{\partial^2 \tilde{f}(\mathbf{z})}{\partial \mathbf{z}^* \partial \mathbf{z}^*} \Omega \right) \mathbf{L} \\ &= \mathbf{L}^T \left(\text{diag} \left(\mathbf{1} - \frac{|\mathbf{y}|}{2|\mathbf{L}\mathbf{x}|} \right) + \text{diag} \left(\frac{|\mathbf{y}|e^{j2\angle \mathbf{L}\mathbf{x}}}{2|\mathbf{L}\mathbf{x}|} \right) \Omega \right) \mathbf{L}, \end{aligned}$$

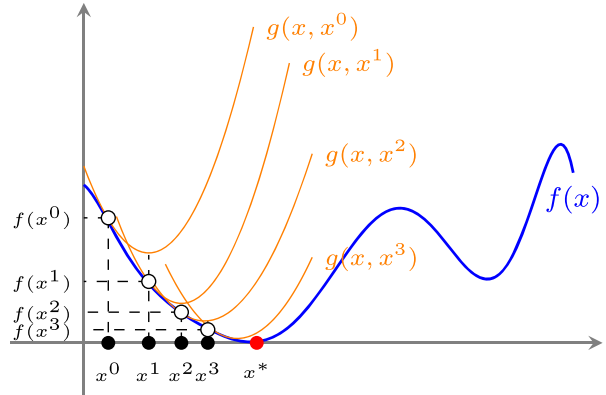


Figure A1. Sketch of the iterative MM algorithm. The function $f(\mathbf{x})$ is shown in blue, while a few surrogate functions $g(\mathbf{x}, \mathbf{x}^k), k \in \{0, 1, 2, 3\}$, are shown in orange. The MM algorithm seeks to find a local minimizer of $f(\mathbf{x})$, red point, by iteratively minimizing easy to minimize functions $g(\mathbf{x}, \mathbf{x}^k)$. The surrogate function $g(\mathbf{x}, \mathbf{x}^k)$ is greater than or equal to $f(\mathbf{x})$, and the equality holds at the current optimal point \mathbf{x}^k . Minimization of $g(\mathbf{x}, \mathbf{x}^k)$ gives a new optimal point and it reaches to \mathbf{x}^* when k tends to infinity.

where $\Omega(\mathbf{z}) = \mathbf{z}^*$ denotes the complex conjugate operator.

Ignoring the nonlinear terms of Hessian leads to the following Gauss–Newton approximation

$$g(\mathbf{x}, \mathbf{x}^k) = f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)^T \Delta \mathbf{x} + \Delta \mathbf{x}^T \mathbf{L}^T \mathbf{L} \Delta \mathbf{x}, \quad (\text{A8})$$

or equivalently

$$g(\mathbf{x}, \mathbf{x}^k) = \frac{1}{2} \|\mathbf{L}\mathbf{x} - |\mathbf{y}|e^{j\angle \mathbf{L}\mathbf{x}^k}\|_2^2. \quad (\text{A9})$$

The quadratic function $g(\mathbf{x}, \mathbf{x}^k)$ majorizes $f(\mathbf{x})$ at the point \mathbf{x}^k provided that the following MM conditions are satisfied:

$$\begin{cases} g(\mathbf{x}^k, \mathbf{x}^k) = f(\mathbf{x}^k) & (\text{A10a}) \\ g(\mathbf{x}, \mathbf{x}^k) \geq f(\mathbf{x}) \quad \text{for all } \mathbf{x}. & (\text{A10b}) \end{cases}$$

Equality (A10a) can easily be confirmed by substituting \mathbf{x} by \mathbf{x}^k in $f(\mathbf{x})$ and $g(\mathbf{x}, \mathbf{x}^k)$. Furthermore, let $z_i = (\mathbf{L}\mathbf{x})_i$ and $\theta_i = \angle(\mathbf{L}\mathbf{x}^k)_i$ then according to the backward triangle inequality (Meyer 2000)

$$\begin{aligned} \frac{1}{2} |z_i - |y_i|e^{j\theta_i}|^2 &\geq \frac{1}{2} (|z_i| - ||y_i|e^{j\theta_i}|)^2 \\ &= \frac{1}{2} (|z_i| - |y_i||e^{j\theta_i}|)^2 \\ &= \frac{1}{2} (|z_i| - |y_i|)^2. \end{aligned}$$

Hence,

$$\frac{1}{2} \sum_i |z_i - |y_i|e^{j\theta_i}|^2 \geq \frac{1}{2} \sum_i (|z_i| - |y_i|)^2, \quad (\text{A12})$$

which confirms that the inequality (A10b) is also satisfied and thus $g(\mathbf{x}, \mathbf{x}^k)$ majorizes $f(\mathbf{x})$. These conditions guarantee that iterative minimization of $g(\mathbf{x}, \mathbf{x}^k)$ converge to a local minimum of $f(\mathbf{x})$ as k tends to infinity (Wu *et al.* 1983).