-statistics approach to optimal transport waveform inversion

κ-statistics approach to optimal transport waveform inversion

Sérgio Luiz E. F. da Silva [*] and G. Kaniadakis [†]
Department of Applied Science and Technology, Politecnico di Torino, 10129 Torino, Italy

(Received 9 February 2022; accepted 29 August 2022; published 8 September 2022)

Extracting physical parameters that cannot be directly measured from an observed data set remains a great challenge in several fields of science and physics. In many of these problems, the construction of a physical model from waveforms is hampered by the phase ambiguity of the recorded wave fronts. In this work, we present an approach for mitigating the effect of phase ambiguity in waveform-driven issues. Our proposal combines the optimal transport theory with the κ-statistical thermodynamics approach. We construct an energy function from the most probable state of a system described by a finite-variance κ-Gaussian distribution to introduce an optimal transport metric. We demonstrate that our proposal outperforms the classical frameworks by considering a nonlinear geophysical data-driven problem based on a wave equation numerical solution. The κ-generalized optimal transport metric is easily adapted to various inverse problems, from estimating power-law exponents to machine learning approaches in quantum mechanics.

DOI: 10.1103/PhysRevE.106.034113

.The transportation theory was first formalized by Monge [1] to study the optimal allocation of resources by redistributing mass [2]. However, Monge's formulation of the optimal transport (OT) problem is an ill-posed problem, which means that the solution (if it exists) is not unique. Later, Kantorovich introduced a well-posed relaxation of the OT problem proposed by Monge by defining the Wasserstein distance (also known as Kantorovich-Rubinstein metric) between two probability distributions [3]. In recent years, transport optimization has been applied in several areas such as quantum mechanics [4–6], statistical physics [7–9], machine learning [10–12], geophysics [13–15], and ecology [16–18].

To illustrate, in a general context, the idea behind the OT framework, suppose that there are N_1 tungsten mines in Rio Grande do Norte, a state of Brazil, that serve N_2 factories within Brazilian territory to produce, for instance, efficient electrical conductors. Let us assume that the N_2 factories use all the tungsten ore produced by the N_1 tungsten mines. What is the *optimal transport plan* for the distribution of the tungsten ore between each mine and each factory that leads to a minimal transport cost? This practical problem, although simple to pose, is remarkably difficult to solve.

In the Kantorovich-Rubinstein formulation, if we consider two sets of points $\Omega_1 = \{x_i, i = 1, 2, \ldots, N_1\}$ and $\Omega_2 = \{y_j, j = 1, 2, \ldots, N_2\}$, each point $x_i \in \Omega_1$ ($y_j \in \Omega_2$) is represented by a "mass" function, namely, $\mu(x_i)$ [$\upsilon(y_j)$]. In this approach, it is assumed that the mass is conserved and that the total mass is equal to 1,

$$\sum_{i=1}^{N_1} \mu(x_i) = \sum_{j=1}^{N_2} \upsilon(y_j) = 1, \tag{1}$$

to satisfy the probability axioms. So the OT problem is formulated as the following optimization task [2]:

$$\mathcal{W}(\mu, \upsilon) = \min_{G \in \Lambda(\mu,\upsilon)} \sum_{i,j} G_{i,j} C_{i,j}, \tag{2}$$

where $\mathcal{W}$ is the transport cost (also known as the optimal total cost), $\Lambda(\mu, \upsilon)$ is the set of transport plans G given by

$$\Lambda(\mu, \upsilon) = \begin{cases} G_{i,j} \geqslant 0 & \forall\,(i, j), \\ \sum_{j=1}^{N_2} G_{i,j} = \mu(x_i) & \forall\, i, \\ \sum_{i=1}^{N_1} G_{i,j} = \upsilon(y_j) & \forall\, j, \end{cases} \tag{3}$$

and C is a cost function which maps each pair (x_i, y_j) to $[0; +\infty]$. The transport plan G specifies for each pair (x, y) how many "tungsten particles" from $\mu(x)$ should be transported to $\upsilon(y)$. Thus, solving the OT problem (2) is equivalent to finding the optimal transport plan G_{opt} that minimizes the optimal total cost $\mathcal{W}$ between μ and υ for a given cost function C.

It is worth emphasizing that for two complete and separable metric spaces $\mathcal{X}$ and $\mathcal{Y}$ whose marginal distributions are μ and υ, respectively, the set of transport plans $\Lambda(\mu, \upsilon)$ represents the set of all joint probability measures on $\mathcal{X} \times \mathcal{Y}$ for $\mu(x) \in \mathcal{P}(\mathcal{X})$ and $\upsilon(y) \in \mathcal{P}(\mathcal{Y})$. The well-posed relaxation of Monge's problem presented in (2) is usually referred to as the Wasserstein distance $\mathcal{W}(\mu, \upsilon)$, which is a metric defined between probability measures in a space $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$. For instance, the metric space $\mathcal{P}(X \times Y)$ formed by the sets of probability measures μ and υ with finite moments of order $p \in [1, \infty)$ for $N_1 = N_2 = N$) has been used in physical problems, such as the $p = 1$ and $p = 2$ cases, respectively, in Refs. [19,20], for studying the energy flow between collider events. Formally, the so-called p-Wasserstein distance is given

[*] sergio.dasilva@polito.it
[†] giorgio.kaniadakis@polito.it

by [2]

$$\mathcal{W}_p(\mu,\upsilon)=\min_{G\in\Lambda(\mu,\upsilon)}\left(\sum_{i,j}G_{i,j}||x_i-y_j||^p\right)^{1/p}, \tag{4}$$

where $||\cdot||$ denotes the norm.

The OT theory has also been used in various physical applications for estimating model parameters [21–24] in which the cost function is linked to a law of error. For example, the cost function associated with the quadratic p-Wasserstein metric with $p=2$, $\mathcal{W}_2^2(\mu,\upsilon)$, is obtained by maximizing a Gaussian distribution by assuming that independent errors $\varepsilon_{i,j}=x_i-y_j$ obey Gaussian statistics [25]. In fact, a cost function is closely linked to the statistical interpretation given to the variables [26–28]. From a statistical physics point of view, this is because the most probable state is the one with the highest multiplicity (i.e., the lowest-energy state). Thus, determining the most likely state of a system is equivalent to minimizing an energy function (or cost function).

In this work, we introduce a Wasserstein distance based on a non-Gaussian cost function that arises from the κ-statistical thermodynamics approach [29–33]. In this regard, we assume that the errors $\vec{\varepsilon}=\{\varepsilon_1,\varepsilon_2,\varepsilon_3,\ldots,\varepsilon_N\}$ are independent and identically distributed (iid) by a κ-generalized Gaussian distribution with finite variance (or κ-Gaussian distribution) as follows [34]:

$$P_\kappa(\varepsilon_i)=\frac{1}{Z_\kappa}\exp_\kappa\left(-\beta_\kappa\varepsilon_i^2\right), \tag{5}$$

where

$$\exp_\kappa(y)=\left(\sqrt{1+\kappa^2y^2}+\kappa y\right)^{\frac{1}{\kappa}} \tag{6}$$

is the κ-exponential function [29] and

$$Z_\kappa=\left(\frac{2}{2+|\kappa|}\right)\frac{\Gamma(1/|2\kappa|-1/4)}{\Gamma(1/|2\kappa|+1/4)}\sqrt{\frac{\pi}{2|\kappa|\beta_\kappa}} \tag{7}$$

and [35]

$$\beta_\kappa=\frac{(1+|\kappa|/2)}{|2\kappa|(2+3|\kappa|)}\frac{\Gamma(1/|2\kappa|-3/4)\Gamma(1/|2\kappa|+1/4)}{\Gamma(1/|2\kappa|+3/4)\Gamma(1/|2\kappa|-1/4)} \tag{8}$$

are defined for $|\kappa|<2/3$. In the limit $\kappa\to0$, the κ-Gaussian distribution (5) becomes the standard Gaussian distribution ($\beta_0\approx1/2$ and $Z_0\approx\sqrt{2\pi}$). In Fig. 1, we present plots of the κ-Gaussian distribution for typical κ values, where the black curve refers to the standard Gaussian distribution ($\kappa\to0$).

Since the errors are assumed to be iid by the power-law tailed distribution in Eq. (5), we can obtain the κ-cost function by computing the most probable state using the maximum likelihood:

$$C_\kappa(m)\propto\max_m\mathcal{L}_\kappa(m):=\prod_{i=1}^{N}P_\kappa(\varepsilon_i(m)), \tag{9}$$

where $\mathcal{L}_\kappa$ denotes the likelihood function and m is the model parameter. We notice that maximizing the likelihood function is equivalent to minimizing the negative log likelihood:

$$C_\kappa(m)\propto\min_m\Theta_\kappa(m):=-\ln[\mathcal{L}_\kappa(m)], \tag{10}$$

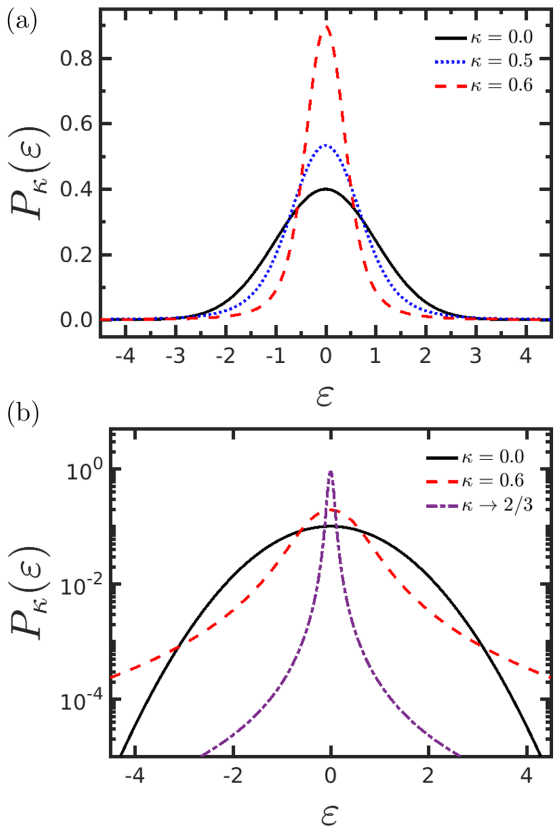

FIG. 1. Probability plots of the κ-Gaussian distribution (5) for some κ values, in which the black curve depicts the standard Gaussian distribution ($\kappa\to0$), in (a) a linear plot and (b) a semilog plot.

where

$$\Theta_\kappa(m)=N\ln(Z_\kappa)-\sum_{i=1}^{N}\ln\left\{\exp_\kappa\left[-\beta_\kappa\varepsilon_i^2(m)\right]\right\}. \tag{11}$$

Once $N\ln(Z_\kappa)$ is constant, note that minimizing the negative log-likelihood function is equivalent to minimizing the second term of Θ_κ, so we can define the κ-cost function for any data-driven problem as follows:

$$C_\kappa(m)=-\sum_{i=1}^{N}\ln\left\{\exp_\kappa\left[-\beta_\kappa\varepsilon_i^2(m)\right]\right\}, \tag{12}$$

where ε denotes the error (the difference between modeled data and observed data). We notice that in the classical limit $\kappa\to0$, the κ-cost function converges to the classical energy function: $C_0(m)=\frac{1}{2}\sum_{i=1}^{N}\varepsilon_i^2(m)$.

In addition to representing a non-Gaussian behavior, the κ-cost function (12) is robust concerning the existence of possible erratic measurements and discrepant observations (outliers). To verify the sensitivity of the κ-cost function to erratic data, we compute the influence function, which is defined as [36]

$$\Upsilon_\kappa:=\frac{\partial C_\kappa(\varepsilon;m_j)}{\partial\varepsilon}, \tag{13}$$

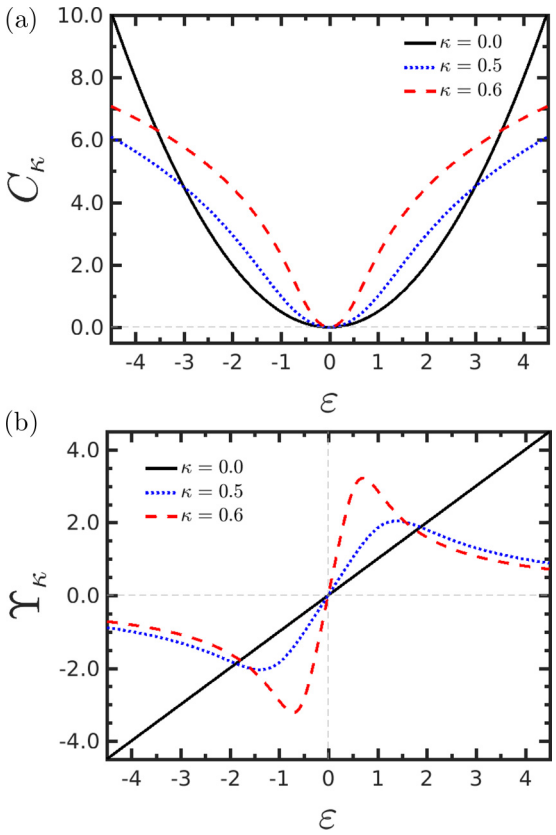

FIG. 2. (a) κ-cost function (14) for typical κ values, in which the black curve depicts the classical cost function ($\kappa \to 0$). (b) The plot of the derivative of the κ-cost function C_κ (also known as the influence function).

where $C_\kappa(\varepsilon; m_j)$ represents the κ-cost function for a given model m_j, which can be written as

$$C_\kappa(\varepsilon; m_j) = C_\kappa(\varepsilon) = -\sum_{i=1}^{N} \ln\left[\exp_\kappa\left(-\beta_\kappa \varepsilon_i^2\right)\right], \tag{14}$$

which is depicted in Fig. 2(a). A cost function is said to be robust if $\Upsilon \to 0$ under $\varepsilon \to \pm\infty$ and not robust if $\Upsilon \to \pm\infty$ under $\varepsilon \to \pm\infty$ [36]. The influence function associated with the κ-cost function for a given ε is then given by

$$\Upsilon_\kappa = \begin{cases} \frac{2\beta_\kappa \varepsilon}{\sqrt{1+\kappa^2 \beta_\kappa^2 \varepsilon^4}} & \text{for } 0 < \kappa < 2/3, \\ \varepsilon & \text{for } \kappa = 0. \end{cases} \tag{15}$$

Analyzing Eq. (15), we notice that if there is an outlier ($\varepsilon \to \infty$), the influence function tends to infinity in the classical limit $\kappa \to 0$, as depicted by the black curve in Fig. 2(b). In contrast, the influence function in the case $0 < \kappa < 2/3$ is robust since $\Upsilon \to 0$ when $\varepsilon \to \infty$, illustrated by the dotted blue and dashed red lines in Fig. 2(b).

For simplicity, from now on, let us consider the space $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ constituted by a set of probability measures with $\mathcal{X} = \mathcal{Y} \subset \mathbb{R}^N$ ($N \in \mathbb{N}$), with $N_1 = N_2 = N$. Considering the κ-cost function (12) and the classical Wasserstein distance (2), we define a criterion in a general setting named the κ-Wasserstein metric, given by

$$\mathcal{W}_\kappa(\mu, \upsilon) = \min_{G_{i,j}} \sum_{i,j=1}^{N} G_{i,j} C_{i,j,\kappa} \tag{16}$$

subject to

$$G_{i,j} \geqslant 0, \quad \sum_{j=1}^{N} G_{i,j} = \mu(x_i), \quad \sum_{i=1}^{N} G_{i,j} = \upsilon(y_j), \tag{17}$$

with

$$C_{i,j,\kappa} = -\ln\{\exp_\kappa[-\beta_\kappa (x_i - y_j)^2]\}. \tag{18}$$

In its most common formulation, the OT metric is expressed as a function of probability distributions. Let us represent the probability distributions μ and υ in terms of the Dirac delta function:

$$\mu(x) = \frac{1}{N} \sum_{l=1}^{N} \delta(x - u_l), \tag{19a}$$

$$\upsilon(y) = \frac{1}{N} \sum_{l=1}^{N} \delta(y - w_l), \tag{19b}$$

where $u_l \in \Omega$ and $w_l \in \Omega$ indicate the location of the data points defining $\mu(x)$ and $\upsilon(x)$. Furthermore, to satisfy the constraint $G_{i,j} \geqslant 0$ in Eq. (17), the coefficients of $G_{i,j}$ for $i = j$ must be zero. Thus, the constrained optimization problem in (16) can be reformulated as

$$\mathcal{W}_\kappa(\mu, \upsilon) = \min_{G_{i,j}} -\frac{1}{N} \sum_{i,j=1}^{N} G_{i,j} h_{i,j,\kappa} \tag{20}$$

subject to

$$G_{i,j} \geqslant 0, \quad \sum_{j=1}^{N} G_{i,j} = 1, \quad \sum_{i=1}^{N} G_{i,j} = 1, \tag{21}$$

where

$$h_{i,j,\kappa} = \ln(\exp_\kappa\{-\beta_\kappa[\mu(x_i) - \upsilon(y_j)]^2\}). \tag{22}$$

Obtaining the solution of the optimization problem in (20) consists of computing a transport plan G that connects points in $\mathcal{P}(\mathcal{X})$ to correspondents points in $\mathcal{P}(\mathcal{Y})$, which minimizes optimal total cost $\mathcal{W}_\kappa$ [37].

Hence, the OT in the context of κ statistics is formulated as a linear sum assignment problem (LSAP) [38], which is, in essence, a combinatorial optimization task. Therefore, the κ-Wasserstein distance between two probability distributions is given by

$$\mathcal{W}_\kappa(\mu, \upsilon) = \min_{\sigma \in \Omega(N)} -\frac{1}{N} \sum_{i=1}^{N} \mathcal{H}_{i,\sigma,\kappa}, \tag{23}$$

with

$$\mathcal{H}_{i,\sigma,\kappa} = \ln(\exp_\kappa\{-\beta_\kappa[\mu(x_i) - \upsilon(y_{\sigma(i)})]^2\}), \tag{24}$$

where σ is a permutation solution of the LSAP problem in (20) associated with the transport map G and $\Omega(N) = \{1, 2, \ldots, N\}$ denotes an ensemble of permutations.

The κ-statistics-based OT (23) is more flexible regarding standard approaches since it appreciates the improvements provided by κ-Gaussian statistics. Indeed, the κ-Wasserstein distance (23) suppresses the contributions of large errors $\varepsilon_i = \mu(x_i) - \upsilon(y_{\sigma(i)})$ and magnifies the contributions of small errors for the optimal total cost $\mathcal{W}_\kappa$ since the influence function associated with $\mathcal{W}_\kappa$ is $\Upsilon_\kappa \propto 1/\varepsilon$ for large errors and $\Upsilon \propto \varepsilon$ for small errors, as discussed earlier.

However, the κ-OT metric is defined as matching two probability distributions, and therefore, this framework (in this format) is valid only for comparing normalized and non-negative quantities. Usually, nonpositive measures are distorted through nonlinear transformations to represent them as probability functions, which may "manufacture" superfluous information. In fact, several applied physics problems deal with oscillatory and non-normalized observable quantities, such as in the analysis of measured signals to estimate physical parameters in a wide variety of physics problems [39–43]. In these applications, namely, inverse problems [44], physical parameters are obtained from the minimization of the difference between the modeled data and the observed data (or error).

In this work, to demonstrate the potential of our proposal, we apply the κ-Wasserstein metric to solving a nonlinear inverse problem used in imaging issues known as full-waveform inversion (FWI) [45–47]. The main goal of FWI consists of estimating a quantitative model by matching modeled waveforms to observed waveforms [48,49]. The model parameters and the modeled waveforms are the coefficients and the solution of the wave equation, respectively. Thanks to its ability to explore the physics provided by wave equations, FWI has been applied in several areas such as astrophysics [50,51] and biomedical imaging [52,53].

The OT-related metrics have been verified to be a powerful tool to mitigate a critical problem in FWI, namely, cycle skipping (or phase ambiguity) [54]. Cycle skipping occurs due to the low-frequency lacuna in the observed data set and the absence of an adequate subsurface model used as an initial guess in the FWI process, which generates a misfit greater than half a wavelength when comparing observed and modeled waveforms (see, for instance, Eq. (31) of Ref. [45]). In this context, the use of OT distance has provided notable progress in image reconstruction via FWI by transforming the amplitude of the seismic signals into positive quantities and has been normalized through linear [55,56], quadratic [57], and exponential-based relationships [58], in addition to other quite interesting ones [59–61] such as the graph-space transformation [62–65], the approach used in this work. Such OT approaches assume that the errors obey Gaussian statistics, while our proposal is suitable to handle non-Gaussian (and also Gaussian) errors that arise naturally in nonlinear problems like FWI.

We consider the two-dimensional acoustic wave equation, where the modeled waveform is the pressure field ψ, which satisfies the following equation:

$$\frac{1}{c^2(\vec{z})}\frac{\partial^2\psi_s(\vec{z},t)}{\partial t^2} - \rho(\vec{z})\nabla\cdot\left(\frac{1}{\rho(\vec{z})}\nabla\psi_s(\vec{z},t)\right) = f_s(\vec{z}_s,t), \tag{25}$$

where $\vec{z}\in\mathbb{R}^2$ and t denote spatial coordinates and time, respectively, ψ_s is the pressure field generated by the source f_s at the position $\vec{z}=\vec{z}_s$, c is the P-wave velocity model, and ρ represents the density model. The modeled data d^{mod} are defined by $d^{\mathrm{mod}}_{s,r}(t) = \mathcal{S}_{s,r}\psi_s(\vec{z},t) = \psi_s(\vec{z}_r,t)$, where $\mathcal{S}_{s,r}$ is a sampling operator and the r index represents a receiver employed in the data acquisition. That is, the modeled data are the pressure field from Eq. (25) at the position $\vec{z}=\vec{z}_r$. Indeed, since FWI compares the modeled wave field with measured data in a seismic survey (observed data), the sampling operator represents a measurement process (onto receiver r of source s) since the solution of the wave equation (25) is computed in the entire physical domain [45].

The comparison between modeled waveforms (modeled data) and observed waveforms (observed data) is performed by means of a misfit function. In the classical approach, FWI minimizes the error in the least-squares sense, in which the residual data are assumed to obey Gaussian statistics. However, in various data analysis problems, the errors are non-Gaussian, for instance, in geomagnetic data sets [66], vehicle state data [67], and seismic data [68]. Indeed, errors are seldom Gaussian in nonlinear data-driven problems [69]. In fact, non-Gaussian cost functions have been shown to be efficient tools to deal with data inversion problems. For this reason, a wide variety of criteria based on non-Gaussian behavior have been introduced in the literature. For instance, Ref. [70] verified that cost functions based on hyperbolic secant, Laplace, and Cauchy distributions are less sensitive to non-Gaussian errors in comparison to the classical approach based on the Gaussian criterion. The success of each above-mentioned case is associated with the statistical interpretation of the long tails of these distributions. Thus, cost functions based on long-tailed distributions have been applied to solve many problems, such as the Student's t distribution [71], nonparametric statistical distributions [72], and generalized distributions [73].

In addition, the classical approach suffers from the so-called cycle skipping (or phase ambiguity) [54]. Such an issue occurs because the classical approach measure pointwise the discrepancy between modeled and observed data, and therefore [as depicted by the black bars in Fig. 3(a)], this framework is not convex concerning time shifts between the waveforms caused by kinematic inconsistencies [74].

To employ our proposal to analyze full waveforms (which are oscillatory and non-normalized) without distorting them, we represent the waveforms in the graph space. The graph space is constituted by mathematical structures, namely, graphs, which are ordered pairs of disjoint sets (V,E), where V represents a set of elements of a discrete point cloud called vertices and E is a subset of V composed of paired vertices, namely, edges (see, for instance, Ref. [75] for more details). Here, to solve the κ-OT problem, we consider a waveform (seismic trace) $d(t)$ to be a set of ordered pairs $\{(t_i,d_i)\in\mathbb{R}^2, i=1,2,\ldots,N\}$ for $d_i = d(t_i)$, where t_i denotes the time discretization. In this regard, the graph-transformed representation of the seismic trace is defined by the following transformation:

$$\mathcal{G}:\mathbf{d}\to\mathcal{G}(\mathbf{d}) = d^{\mathcal{G}}(y,t),$$
$$\mathbb{R}^N\to\mathcal{D}(\mathbb{R}^2), \tag{26}$$

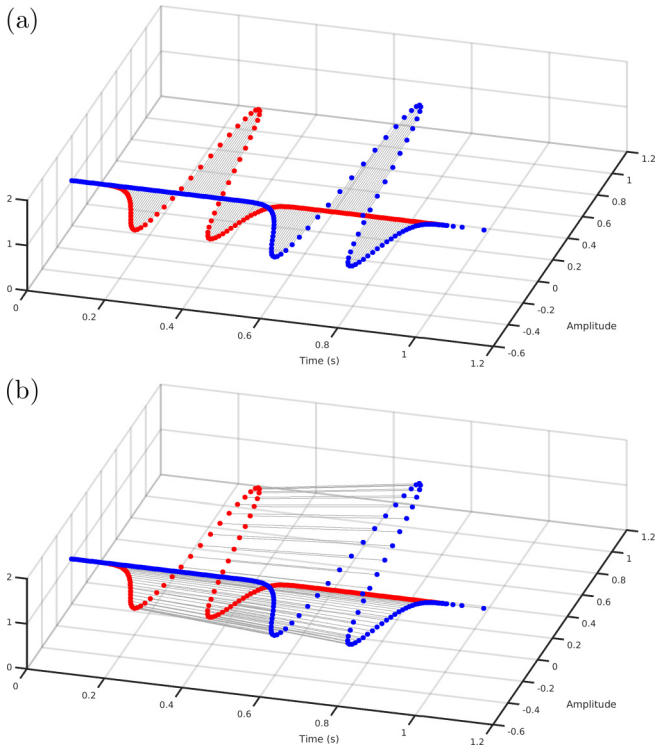

FIG. 3. Representation of two discretized waveforms (red and blue dots), in which the black bars represent the way to compare them. (a) Point-by-point comparison, on the time axis, between the waveforms. (b) The link of each vertex from one waveform to another is computed, considering amplitude and time, by solving a linear sum assignment problem.

where $\mathcal{G}$ represents the graph transformation, $\mathcal{D}(\mathbb{R}^2)$ denotes the space of probability distributions on $\mathbb{R}^2$, $\mathbf{d} = \{d_i, i = 1, 2, \ldots, N\}$ is the discretized waveform, and $d^{\mathcal{G}}(y,t)$ is the graph-transformed signal which is represented by a sum of Dirac delta functions:

$$d^{\mathcal{G}}(y,t) = \frac{1}{N}\sum_{i=1}^{N}\delta(t-t_i)\delta(y-d_i), \tag{27}$$

where y is linked to the amplitude. Figure 3 shows two discretized waveforms (one represented by red dots and the other by blue dots) in which each point represents a vertex given by pairs (t, y). Figure 3(a) illustrates the classical approach in which the two waveforms are compared point by point on the time axis, as depicted by the black bars parallel to the amplitude axis and perpendicular to the time axis. In contrast, optimal-transport based FWI computes the minimum cost to link each vertex from one waveform to another, considering amplitude and time, thus comparing the waveform peak to peak and valley to valley, among other things, as depicted in Fig. 3(b).

From a practical point of view, FWI is formulated as a local optimization problem in which the model parameters (in our case, the P-wave velocities) are estimated iteratively using Newton-based methods [45]. Local optimization issues consist of finding the optimal solution within a neighboring set of an initial guess (initial model) [76]. Thus, using Dirac distributions is not suitable because $\mathcal{G}$ is not a differentiable operator. For this reason, we propose a smooth version of the graph transform based on the approximation of Dirac functions through κ-Gaussian distributions (5) defined as

$$\begin{aligned}\mathcal{G}_\kappa : \mathbf{d} \to \mathcal{G}_\kappa(\mathbf{d}) = d^{\mathcal{G}_\kappa}(y,t),\\ \mathbb{R}^N \to \mathbb{C}^\infty(\mathbb{R}, \mathbb{R}_*^+),\end{aligned} \tag{28}$$

where $\mathbb{C}^\infty(\mathbb{R}, \mathbb{R}_*^+)$ is the set of strictly positive and infinitely differentiable functions of $(\mathbb{R})$ and

$$d^{\mathcal{G}_\kappa}(y,t) = \frac{1}{Z_\kappa}\sum_{i=1}^{N}\exp_\kappa[-\beta_\kappa(t-t_i)^2]\exp_\kappa[-\beta_\kappa(y-d_i)^2]. \tag{29}$$

Finally, based on the κ-Wasserstein distance and the graph representation of a signal in the κ-Gaussian statistics sense, we introduce the graph-space κ-OT misfit function $\phi_{\mathcal{W}_\kappa^{\mathcal{G}_\kappa}}$ as follows:

$$\min_m \quad \phi_{\mathcal{W}_\kappa^{\mathcal{G}_\kappa}}(m) := \sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\mathcal{C}_\kappa\left(d_{s,r}^{\mathrm{mod}}(m), d_{s,r}^{\mathrm{obs}}\right), \tag{30}$$

with

$$\mathcal{C}_\kappa(d^{\mathrm{mod}}, d^{\mathrm{obs}}) = \mathcal{W}_\kappa^{\mathcal{G}_\kappa}\left(d_{\mathrm{mod}}^{\mathcal{G}_\kappa}, d_{\mathrm{obs}}^{\mathcal{G}_\kappa}\right), \tag{31}$$

where $d_{\mathrm{mod},i}^{\mathcal{G}_\kappa} = (t_i, d_{\mathrm{mod},i})$, $d_{\mathrm{obs},i}^{\mathcal{G}_\kappa} = (t_i, d_{\mathrm{obs},i})$, and

$$\mathcal{W}_\kappa^{\mathcal{G}_\kappa}\left(d_{\mathrm{mod}}^{\mathcal{G}_\kappa}, d_{\mathrm{obs}}^{\mathcal{G}_\kappa}\right) = \min_{\sigma\in\Omega(N)} \quad -\sum_{i=1}^{N_t}\ln\left\{\exp_\kappa[-\beta_\kappa(t_i - t_{\sigma(i)})^2]\ \exp_\kappa\left[-\beta_\kappa\left(d_i^{\mathrm{mod}} - d_{\sigma(i)}^{\mathrm{obs}}\right)^2\right]\right\}, \tag{32}$$

where σ is a solution of the combinatorial problem in (23), N_s and N_r are the numbers of seismic sources and receivers employed in a seismic survey, and N_t is the number of time samples. It is worth noting that we multiply $\mathcal{W}_\kappa$ (23) by the factor NZ_κ since minimizing $\mathcal{W}_\kappa$ is equivalent to minimizing the product $NZ_\kappa\mathcal{W}_\kappa$. From now on, we call our proposal the κ-graph-space optimal-transport FWI (or κ-GSOT-FWI for short).

To demonstrate the potentiality of κ-GSOT-FWI to overcome phase ambiguity, we consider, as the true model, a P-wave velocity model inspired by a typical Brazilian pre-salt oil region, namely, the *Chalda model*, which is an area of great economic interest. Such an Earth model, depicted in Fig. 4(a), contains a water layer 2 km thick, followed by post-salt rocks, a salt body, pre-salt layers with an oil region, and bedrock below. The model consists of 562 and 1282 grid cells in

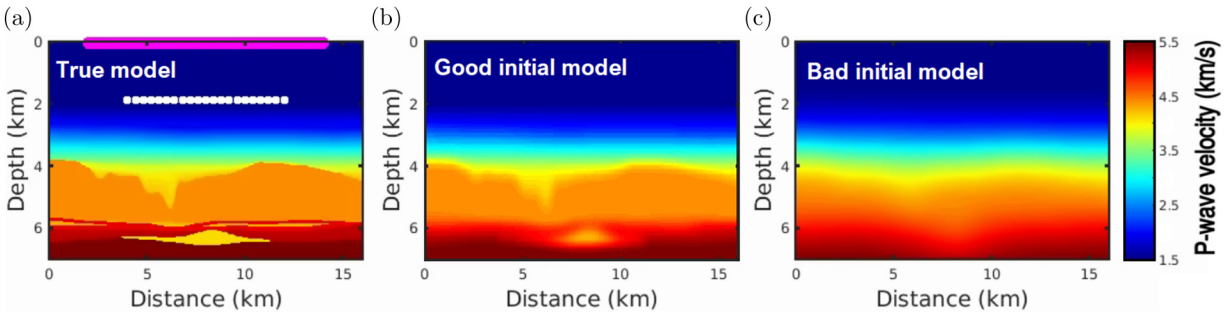

FIG. 4. (a) *P*-wave velocity model used in this study as the true model. The magenta line and the white dots denote the seismic sources and receiver positions, respectively. The initial model employed in the waveform inversion in the (b) first scenario (good initial model) and in the (c) second scenario (bad initial model).

the vertical and horizontal directions, which means that our problem has 720 484 unknown variables.

As a seismic source, we consider a Ricker wavelet, mathematically described by $f(t) = (1 - 2\pi^2\mu_p^2 t^2)\exp(-\pi^2\mu_p^2 t^2)$, where μ_p is the peak frequency. In all experiments, we employ 161 seismic sources, in line and equally spaced every 75 m, at 12.5 m in depth [see the magenta line in Fig. 4(a)]. Furthermore, we consider a Ricker wavelet with $\mu_p = 5$ Hz, which has been high-pass filtered to remove energy less than 2.5 Hz to simulate a realistic scenario. For each source, 21 receivers located every 400 m are implanted on the ocean floor [see the white dots in Fig. 4(a)] to simulate an ocean bottom node survey [77]. This type of geometry acquisition is not affected by systematic noise caused, for instance, by boat traffic, water waves, or wind [78]. We set the simulation time to 10 s.

We carried out waveform inversion by employing a quasi-Newton algorithm known as limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS), a powerful method to solve large-scale optimization issues [79]. In this approach, the optimization problem formulated in Eq. (30) is solved by updating the model m from an initial model m_0 according to

$$m_{j+1} = m_j - \alpha_j H^{-1}(m_j)\nabla\phi_{\mathcal{W}_\kappa^{\mathcal{G}_\kappa}}(m_j), \tag{33}$$

for $j = 0, 1, \ldots, N_{\text{iter}}$, with N_{iter} being the total number of iterations, where α is a step length computed through a line search method [76], H^{-1} is the inverse of the Hessian matrix (second-order derivative of $\phi_{\mathcal{W}_\kappa^{\mathcal{G}_\kappa}}$), and $\nabla\phi_{\mathcal{W}_\kappa^{\mathcal{G}_\kappa}}(m_j)$ is the gradient of $\phi_{\mathcal{W}_\kappa^{\mathcal{G}_\kappa}}$ with respect to the model parameter m_j.

Since computing the gradient of the misfit function is computationally very expensive, we consider the adjoint-state method to efficiently calculate it [80]. In this approach, $\nabla\phi_{\mathcal{W}_\kappa^{\mathcal{G}_\kappa}}(m_j)$ is computed by cross correlating the forward pressure wave field ψ_s [the solution of Eq. (25)] with the so-called adjoint wave field q_s as follows:

$$\nabla\phi_{\mathcal{W}_\kappa^{\mathcal{G}_\kappa}}(m) = -\sum_s \int_0^T \left\langle q_s(\vec{z}, T-t), \frac{\partial^2\psi_s(\vec{z},t)}{\partial t^2}\right\rangle_{\vec{z}} dt, \tag{34}$$

where T denotes the recording time and q_s is the solution of the adjoint-state equation [80]:

$$\frac{1}{c^2(\vec{z})}\frac{\partial^2 q_s(\vec{z},t)}{\partial t^2} - \rho(\vec{z})\nabla\cdot\left(\frac{1}{\rho(\vec{z})}\nabla q_s(\vec{z},t)\right) = \sum_{r=1}^{N_r} F_{s,r}(\vec{z}_r, t), \tag{35}$$

where $F_{s,r}$ is the known adjoint source. For a receiver r and a source s, the κ-related adjoint source is given by

$$F_{s,r}(\vec{z}_r, t_i) = \frac{2\beta_\kappa\left[d_{s,r}^{\text{mod}}(t_i) - d_{s,r}^{\text{obs}}(t_{\sigma(i)})\right]}{\sqrt{1 + \beta_\kappa^2\kappa^2\left[d_{s,r}^{\text{mod}}(t_i) - d_{s,r}^{\text{obs}}(t_{\sigma(i)})\right]^4}}, \tag{36}$$

with $i = 1, 2, \ldots, N_t$. We remark that in the classical limit $\kappa \to 0$, the latter equation becomes the adjoint-source based on Gaussian statistics [64].

In this context, we consider two scenarios regarding the initial guess (initial model): (i) In the first one, we use the *P*-wave velocity model illustrated in Fig. 4(b) as the initial model, which was constructed by applying a Gaussian filter with a standard deviation of 200 m on the true model, namely, the "good initial model." (ii) In the second scenario, we consider an initial model very far from the true one, namely, the "bad initial model," which was constructed by applying a Gaussian filter with a standard deviation of 800 m, which is depicted in Fig. 4(c). The bad initial model does not present the main geological structures of the true model, generating cycle-skipped data. Figure 5 shows some seismograms (also known as receiver gather) which are formed by sets of seismic traces associated with the first receiver on the left and all the seismic sources. Figure 5(a) shows the observed data, and Figs. 5(b) and 5(c) show modeled data related to the bad initial model and the good initial model. We notice that in bad initial

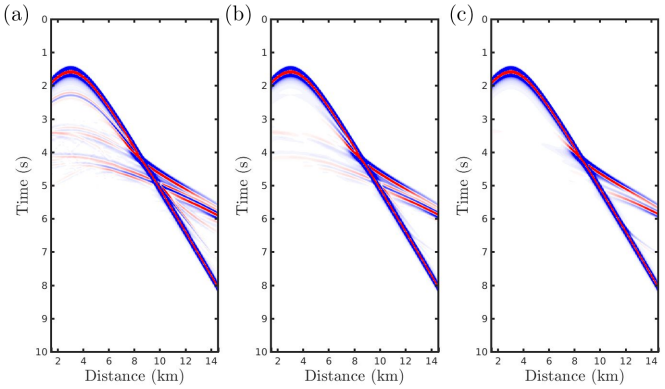

FIG. 5. Seismograms of the first receiver on the left and all the seismic sources. (a) shows the observed data, and (b) and (c) show modeled data related to the bad initial model and the good initial model, respectively.

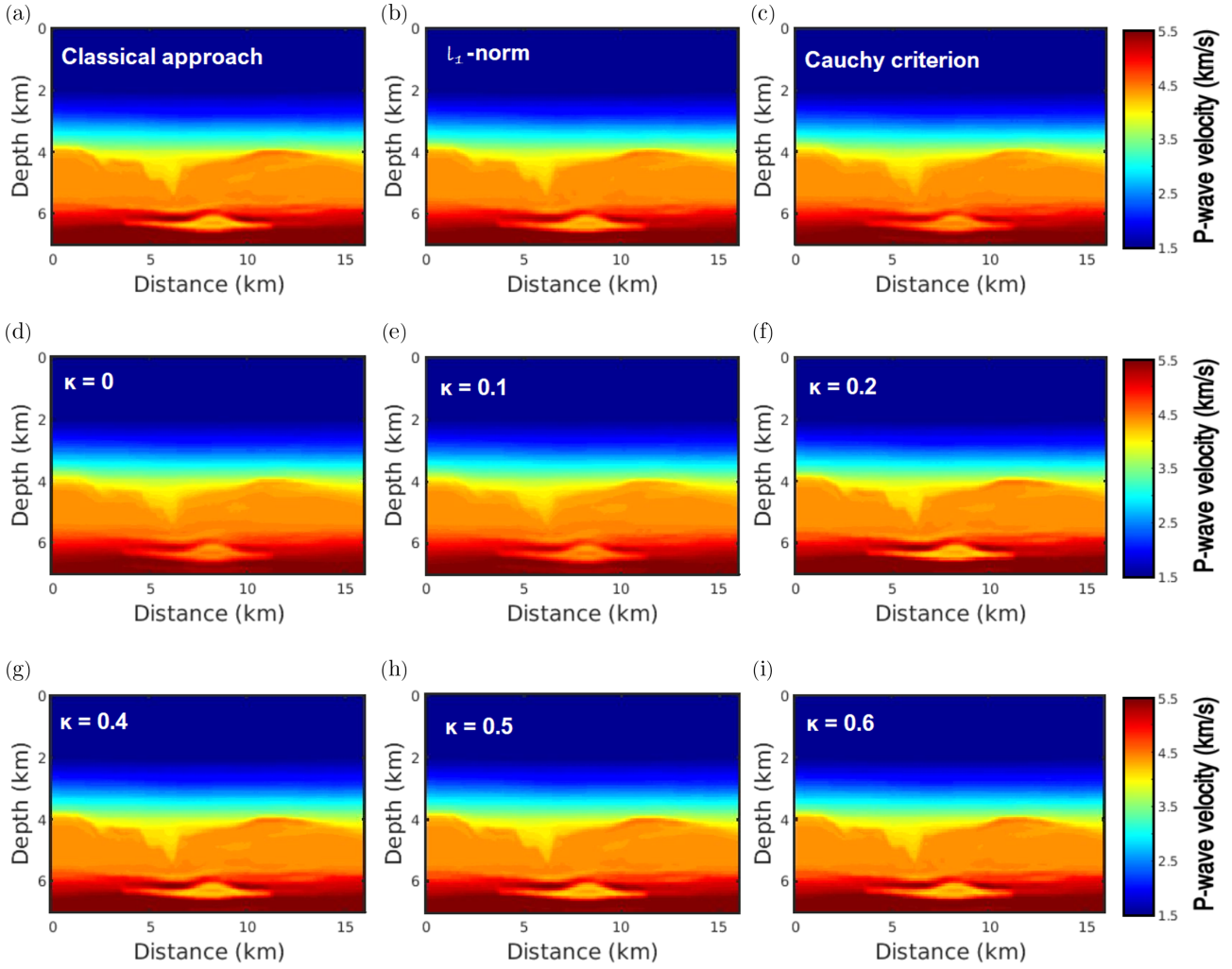

FIG. 6. First scenario. Reconstructed models, after 50 L-BFGS iterations, from (a) the classical approach, (b) l_1-norm framework, (c) Cauchy criterion, and our proposal with (d) $\kappa = 0$, (e) $\kappa = 0.1$, (f) $\kappa = 0.2$, (g) $\kappa = 0.4$, (h) $\kappa = 0.5$, and (i) $\kappa = 0.6$.

model case [Fig. 5(c)], several seismic events are missing, which causes cycle skipping.

In all experiments, we consider 50 L-BFGS iterations. For each scenario, we applied our proposal by considering $\kappa = 0.0$, 0.1, 0.2, 0.4, 0.5, and 0.6. In addition, we also consider the classical approach based on the least-squares method and robust methods based on the l_1 norm [81] and Cauchy criterion [70].

Figure 6 shows the waveform inversion results of the first scenario, in which Figs. 6(a)– 6(c) refer, respectively, to the classical approach and l_1-norm-based and Cauchy-based FWI, while Figs. 6(d)–6(i) show the reconstructed models using our proposal with $\kappa = 0.0$, 0.1, 0.2, 0.4, 0.5, and 0.6. From a visual examination of Fig. 6, we notice that all approaches provide satisfactory velocity models, which means they are very close to the true model [Fig 4(a)]. Such results were already expected since the good initial model [Fig 4(b)] is very close to the true one. However, this experiment demonstrates that all the approaches and algorithms used in this study work well. Thus, if the initial model is quite accurate, using our proposal or the classical approach is immaterial in analyzing no-cycle-skipped data sets.

Since the differences between the reconstructed models, in the first scenario, are not remarkable, we quantitatively compare the reconstructed models c_p^{rec} with the true model c_p^{true}, using two statistical measures: the linear correlation coefficient (R) and the normalized root mean square (NRMS) r_N. The value of R varies between -1 (bad similarity) and 1

TABLE I. Comparative main statistics between the reconstructed models in the first scenario and the true model

Framework	R	NRMS
Classical approach	0.9991	0.0082
l_1-norm approach	0.9989	0.0089
Cauchy criterion	0.9985	0.0088
Our proposal ($\kappa \to 0$)	0.9983	0.0097
Our proposal ($\kappa = 0.1$)	0.9985	0.0095
Our proposal ($\kappa = 0.2$)	0.9988	0.0091
Our proposal ($\kappa = 0.4$)	0.9989	0.0087
Our proposal ($\kappa = 0.5$)	0.9989	0.0088
Our proposal ($\kappa = 0.6$)	0.9989	0.0085

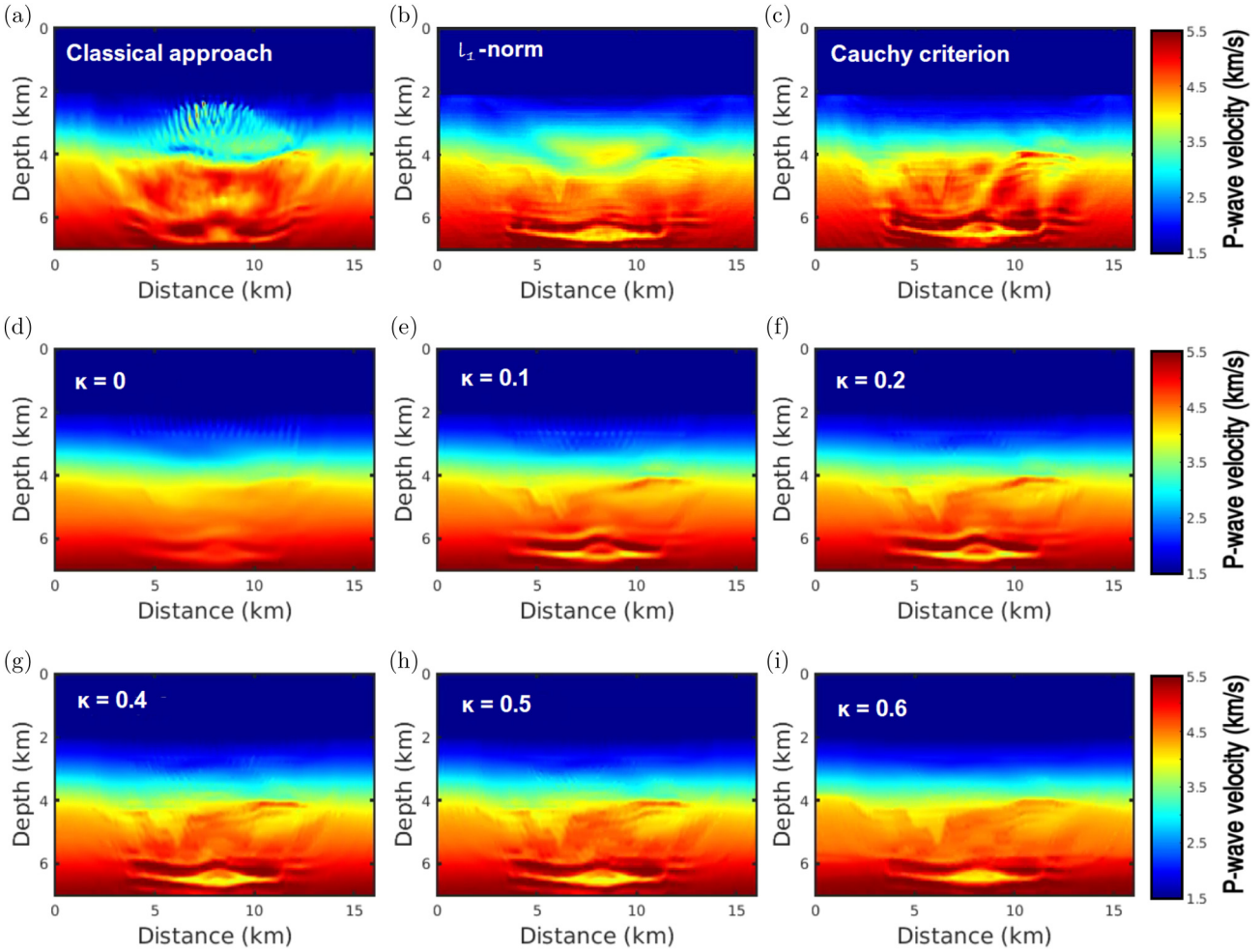

FIG. 7. Second scenario. Reconstructed models, after 50 L-BFGS iterations, from (a) classical approach, (b) l_1-norm framework, (c) Cauchy criterion, and our proposal with (d) $\kappa = 0$, (e) $\kappa = 0.1$, (f) $\kappa = 0.2$, (g) $\kappa = 0.4$, (h) $\kappa = 0.5$, and (i) $\kappa = 0.6$.

(perfect similarity). The NRMS is an error measure defined by

$$r_N = \left[\frac{\sum_i \left(c_{p_i}^{\text{true}} - c_{p_i}^{\text{rec}} \right)^2}{\sum_i \left(c_{p_i}^{\text{true}} \right)^2} \right]^{1/2}, \tag{37}$$

which varies from zero (perfect result model) to ∞ (bad result model). These statistical measures are summarized in Table I, where it is notable that the classical approach presents a reconstructed model with greater correlation ($\mathtt{R} \to 1$) and less error ($r_N \to 0$) in relation to the true model. In fact, the statistical measurements reveal that the reconstructed models are similar to each other.

Figure 7 shows the waveform inversion results of the second scenario, in which Figs. 7(a)– 7(c) refer, respectively, to the classical approach and l_1-norm-based and Cauchy-based FWI, while Figs. 7(d)–7(i) show the reconstructed models using our proposal with $\kappa = 0.0$, 0.1, 0.2, 0.4, 0.5, and 0.6. From a visual examination of Fig. 7(a), it is remarkable that the classical approach fails, as expected, to obtain a satisfactory P-wave model. In fact, the reconstructed model shows many artifacts in the structures of the first half of the model (depth less than 4 km), in addition to being unable to identify the deeper structures, especially in the pre-salt region (around 6 km in depth). Regarding the waveform inversion based on robust criteria, the reconstructed models, in the second scenario, are better than the classical approach [see Figs. 7(b) and 7(c)]. However, such resulting models present many artifacts, and therefore, cases based on the l_1 norm and Cauchy criterion produce unsatisfactory models. In contrast, the κ-GSOT-FWI gives better reconstructed models [see Figs. 7(d)–7(i)]. Indeed, the reconstructed model comes closer to the true model as the κ parameter moves away from zero, which means a greater deviation from Gaussian behavior. In this way, the $\kappa = 0.6$ case presents a reconstructed P-wave velocity model [Fig. 7(i)] closer to the true model [Fig. 7(a)].

We present in Fig. 8 the first gradients related to the bad initial model case for the classical approach, l_1-norm framework, Cauchy criterion, and our proposal with $\kappa = 0$, 0.1, 0.2, 0.4, 0.5, and 0.6. In Fig. 8(a), we remark that although the classical approach identifies the deep structures, the classical gradient is dominated by the imprints of the wave paths in the shallow regions (up to approximately 4 km depth), badly conditioning the FWI process. Similar behaviors are also identified in the case of robust criteria, as depicted in Figs. 8(b) and 8(c). However, as the κ-value increases, the shallow wave path related marks are progressively mitigated, and the gradient information is directed towards updates of all reflective layers,

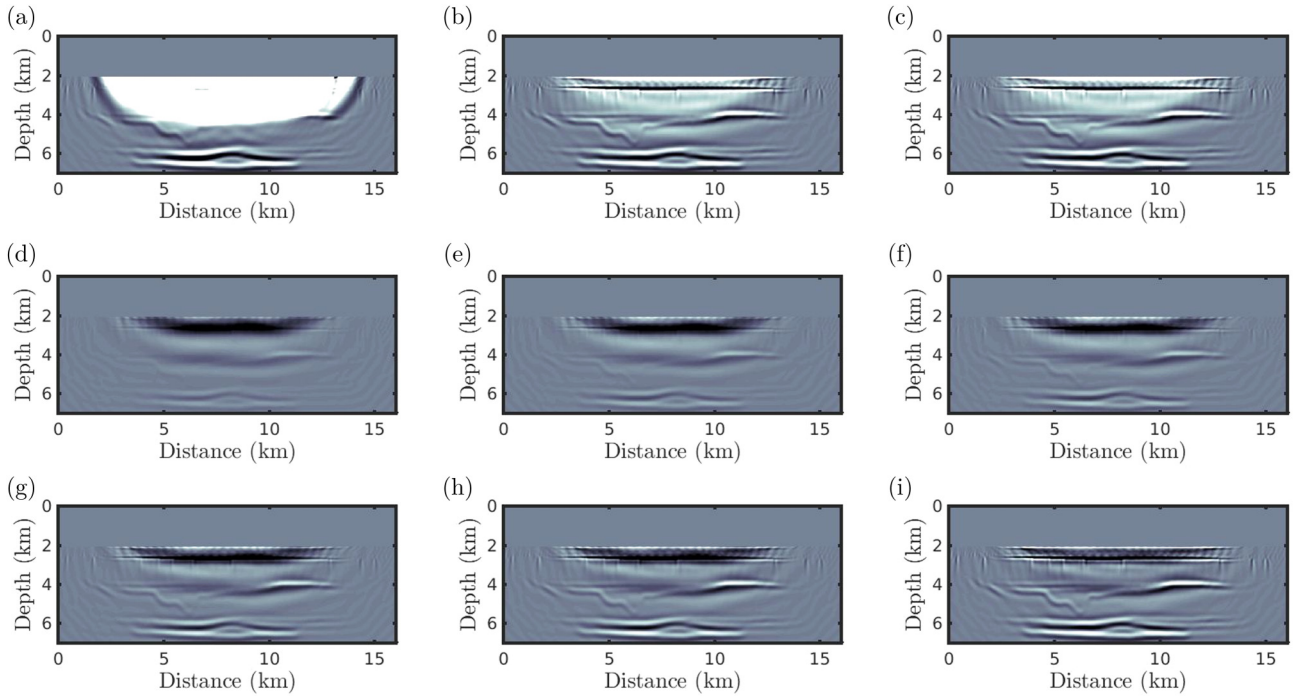

FIG. 8. Gradients of the misfit functions employed in the first FWI iteration related to the bad initial model case for the (a) classical approach, (b) l_1-norm framework, (c) Cauchy criterion, and our proposal with (d) $\kappa = 0$, (e) $\kappa = 0.1$, (f) $\kappa = 0.2$, (g) $\kappa = 0.4$, (h) $\kappa = 0.5$, and (i) $\kappa = 0.6$.

providing a high-resolution image [see, for instance, Figs. 8(h) and 8(i)].

The black points in Figs. 9(a)– 9(e) show the semilog plot statistical distributions of the errors (difference between modeled and observed data) before the graph transformation in the second scenario, which indicates that errors do not, in fact, follow Gaussian statistics, as depicted by the discrepancy between the black dots and magenta curves. The red dots refer to

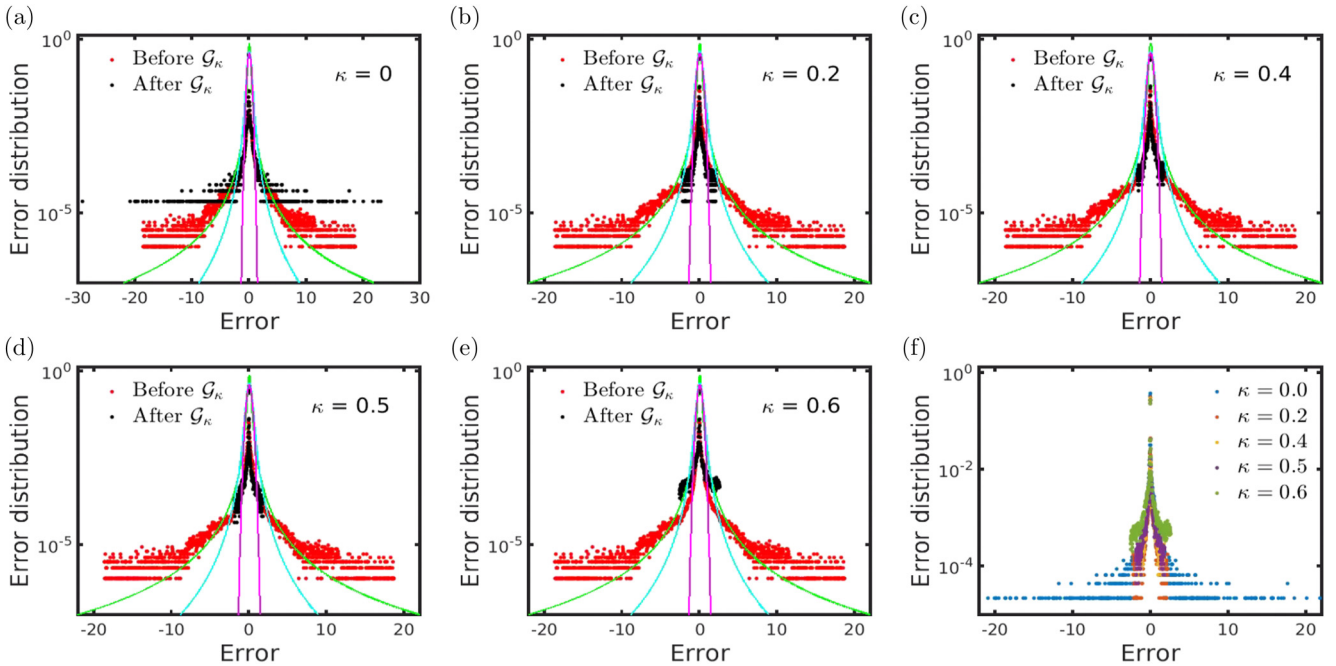

FIG. 9. Statistical distribution of errors associated, in the second scenario, before (red points) and after the graph transformation (black points) for the κ-GSOT-FWI, with (a) $\kappa = 0$, (b) $\kappa = 0.2$, (c) $\kappa = 0.4$, (d) $\kappa = 0.5$, and (e) $\kappa = 0.6$, in which the magenta, cyan, and green curves represent, respectively, the ordinary Gaussian distribution ($\kappa = 0$) and κ-Gaussian distribution with $\kappa = 0.4$ and $\kappa = 0.6$. (f) is a zoom of the other panels, considering only the error distributions after the graph transformation [black dots in (a)–(e)].

TABLE II. Comparative main statistics between the reconstructed models in the second scenario and the true model.

Framework	R	NRMS
Classical approach	0.9582	0.1083
l_1-norm approach	0.9788	0.0825
Cauchy criterion	0.9601	0.0973
Our proposal ($\kappa \to 0$)	0.9808	0.0684
Our proposal ($\kappa = 0.1$)	0.9878	0.0664
Our proposal ($\kappa = 0.2$)	0.9881	0.0662
Our proposal ($\kappa = 0.4$)	0.9883	0.0648
Our proposal ($\kappa = 0.5$)	0.9891	0.0624
Our proposal ($\kappa = 0.6$)	0.9950	0.0426

the error distributions after graph transformation. In regard to this point, we remark that the shapes of the error distributions before and after the graph transformation are very similar, and therefore, the Gaussian distribution is not the most suitable for this one. Furthermore, regardless of the κ value used to compute the optimal transport, the error distribution after the graph transformation is "more likely" to be a κ-Gaussian distribution with $\kappa = 0.6$ [see green curves in Figs. 9(a)–9(e)]. Figure 9(f) is a zoom of the other panels, considering only the error distributions after the graph transformation [black dots in Figs. 9(a)–9(e)] for a better comparison between them (we consider different colors for each κ case). We notice that the error distributions are very similar around the error equal to zero, which refer to the non-cycle-skipped waveforms. The main differences are in the tails of the distributions, which indicate seismic phases that skip the cycle. Analyzing Fig. 9, we observe that the errors are smaller after the graph transformation, especially in the case with $\kappa = 0.6$, where the errors are more accurately represented (see the good agreement between the green curves and the black points). In real applications, testing several κ values to search for the optimal κ parameter is time-consuming. In this context, we suggest determining the optimal κ value from the histogram of the residual data, similarly to what is depicted in Fig. 9.

In this work, we have explored the portability of the κ-statistical thermodynamics approach in the context of OT problems for a robust solution of data-driven issues. In this regard, from the computation of the most probable state using a κ-Gaussian distribution, we have formulated the OT issue to deal with oscillatory and non-normalized physical quantities by employing a graph-transformed representation of temporal series. To demonstrate the effectiveness of our proposal, we considered a nonlinear geophysical data-driven problem in a realistic setting. The results reinforce the fact that the classical approach fails to obtain good physical models, as well as frameworks based on thel_1 norm and Cauchy criterion. By contrast, the results summarized in Fig. 7 and Table II revealed that our proposal mitigates the phase ambiguity effect on waveform inversion. They also showed that a greater deviation of the Gaussian behavior better indicates the estimated model, which in our applications was represented by the $\kappa = 0.6$.

From a practical point of view, long and tedious data processing is necessary to construct a good initial model to mitigate the phase ambiguity problem, increasing human subjectivity. In this context, our proposal reduces the dependence on human expertise, which is interesting for automated approaches for analyzing big data sets in several areas. In this way, the κ-generalized OT shows great promise to deal with modern data-driven issues. To conclude, we emphasize that our proposal is easily adapted to various other inverse problems, from estimating power-law exponents to machine learning approaches, as well as to facial image comparison [8] and collective motion [82].

[1] G. Monge, *Mémoire sur la théorie des déblais et des remblais* (Histoire de l'Académie Royale des Sciences, Paris, 1781), p. 666.

[2] C. Villani, *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften (Springer, Berlin, 2008)

[3] L. Kantorovich, J. Manage. Sci. **5**, 1 (1958).

[4] Y. Chen, T. T. Georgiou, and A. Tannenbaum, IEEE Trans. Autom. Control **63**, 2612 (2018).

[5] K. Ikeda, Quantum Inf. Process. **19**, 25 (2020).

[6] G. De Palma and D. Trevisan, Ann. Henri Poincare **22**, 3199 (2021).

[7] B. Danila, Y. Yu, J. A. Marsh, and K. E. Bassler, Phys. Rev. E **74**, 046106 (2006).

[8] P. Koehl, M. Delarue, and H. Orland, Phys. Rev. Lett. **123**, 040603 (2019).

[9] P. Koehl, M. Delarue, and H. Orland, Phys. Rev. E **103**, 012113 (2021).

[10] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher, in *Proceedings of the 31st International Conference on Machine Learning* (PMLR, Bejing, China, 2014), p. 306.

[11] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, IEEE Signal Process. Mag. **34**, 43 (2017).

[12] G. Khan and J. Zhang, Info. Geo. **5**, 47 (2022).

[13] J. Chen, Y. Chen, H. Wu, and D. Yang, J. Comput. Phys. **373**, 188 (2018).

[14] N. Hedjazian, T. Bodin, and L. Métivier, Geophys. J. Int. **216**, 130 (2018).

[15] J. Messud, R. Poncet, and G. Lambaré, Inverse Probl. **37**, 065012 (2021).

[16] J. R. Potts, M. Auger-Méthé, K. Mokross, and M. A. Lewis, Methods Ecol. Evol. **5**, 1012 (2014).

[17] B. Muzellec, R. Nock, G. Patrini, and F. Nielsen, in *Thirty-First AAAI Conference on Artificial Intelligence* (AAAI Press, Palo Alto, California USA, 2017), p. 1.

[18] M. Stock, T. Poisot, and B. De Baets, Ecol. Evol. **11**, 3841 (2021).

[19] P. T. Komiske, E. M. Metodiev, and J. Thaler, Phys. Rev. Lett. **123**, 041801 (2019).

[20] T. Cai, J. Cheng, N. Craig, and K. Craig, Phys. Rev. D **102**, 116019 (2020).

[21] Q. Mérigot, Comput. Graphics Forum **30**, 1583 (2011).

[22] G. Buttazzo, L. De Pascale, and P. Gori-Giorgi, Phys. Rev. A **85**, 062502 (2012).

[23] W. Gangbo, W. Li, S. Osher, and M. Puthawala, J. Comput. Phys. **399**, 108940 (2019).

[24] A. M. Stuart and M.-T. Wolfram, SIAM J. Appl. Math. **80**, 599 (2020).

[25] J. V. T. de Lima, S. L. E. F. da Silva, J. M. de Araújo, G. Corso, and G. Z. dos Santos Lima, Eur. Phys. J. Plus **136**, 551 (2021).

[26] H. Suyari and M. Tsukada, IEEE Trans. Inf. Theory **51**, 753 (2005).

[27] S. L. E. F. da Silva, G. Z. dos Santos Lima, J. M. de Araújo, and G. Corso, Phys. A (Amsterdam, Neth.) **563**, 125496 (2021).

[28] S. L. E. F. da Silva and G. Kaniadakis, Phys. Rev. E **104**, 024107 (2021).

[29] G. Kaniadakis, Phys. A (Amsterdam, Neth.) **296**, 405 (2001).

[30] G. Kaniadakis, Phys. Rev. E **66**, 056125 (2002).

[31] G. Kaniadakis, Phys. Rev. E **72**, 036108 (2005).

[32] G. Kaniadakis, Entropy **15**, 3983 (2013).

[33] G. Kaniadakis, Europhys. Lett. **133**, 10002 (2021).

[34] S. L. E. F. da Silva, P. T. C. Carvalho, J. M. de Araújo, and G. Corso, Phys. Rev. E **101**, 053311 (2020).

[35] S. L. E. F. da Silva, R. Silva, G. Z. dos Santos Lima, J. M. de Araújo, and G. Corso, Phys. A (Amsterdam, Neth.) **600**, 127554 (2022).

[36] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions* (Wiley-Interscience, Hoboken, New Jersey, 2005).

[37] C. Villani, *Topics in Optimal Transportation* (American Mathematical Society, Providence, Rhode Island, 2003).

[38] J. D. Benamou and Y. Brenier, Numer. Math. **84**, 375 (2000).

[39] E. Ayón-Beato, A. García, R. Mansilla, and C. A. Terrero-Escalante, Phys. Rev. D **62**, 103513 (2000).

[40] V. I. Geyko and N. J. Fisch, Phys. Rev. Lett. **110**, 150604 (2013).

[41] S. L. E. F. da Silva, J. Julià, and F. H. R. Bezerra, Bull. Seismol. Soc. Am. **107**, 1495 (2017).

[42] S. H. Völkel, C. J. Krüger, and K. D. Kokkotas, Phys. Rev. D **103**, 083008 (2021).

[43] T. Robins, J. Camacho, O. C. Agudo, J. L. Herraiz, and L. Guasch, Sensors **21**, 4570 (2021).

[44] M. Razavy, *An Introduction to Inverse Problems in Physics* (World Scientific, Singapore, 2020).

[45] J. Virieux and S. Operto, Geophysics **74**, WCC1 (2009).

[46] A. Fichtner, *Full Seismic Waveform Modelling and Inversion* (Springer, New York, 2010).

[47] S. Hanasoge, *Imaging Convection and Magnetism in the Sun* (Springer, Cham, 2015).

[48] G. Sun, Q. Chang, and P. Sheng, Phys. Rev. Lett. **90**, 104301 (2003).

[49] S. L. E. F. da Silva, C. A. da Costa, P. T. C. Carvalho, J. M. de Araújo, L. dos Santos Lucena, and G. Corso, Phys. A (Amsterdam, Neth.) **548**, 124473 (2020).

[50] S. M. Hanasoge, Astrophys. J. **797**, 23 (2014).

[51] J. Bhattacharya and S. M. Hanasoge, Astrophys. J. **826**, 105 (2016).

[52] J. Wiskin, B. Malik, D. Borup, N. Pirshafiey, and J. Klock, Sci. Rep. **10**, 20166 (2020).

[53] L. Guasch, O. C. Agudo, M.-X. Tang, P. Nachev, and M. Warner, npj Digital Med. **3**, 28 (2020).

[54] W. Hu, J. Chen, J. Liu, and A. Abubakar, IEEE Signal Process. Mag. **35**, 132 (2018).

[55] Y. Yang, B. Engquist, J. Sun, and B. F. Hamfeldt, Geophysics **83**, R43 (2018).

[56] B. Sun and T. Alkhalifah, Geophysics **84**, R923 (2019).

[57] Y. Yang and B. Engquist, Geophysics **83**, A7 (2018).

[58] L. Qiu, J. Ramos-Martínez, A. Valenciano, Y. Yang, and B. Engquist, in *Proceedings of the SEG Technical Program Expanded Abstracts* (Society of Exploration Geophysicists, Houston, TX, USA, 2017)pp. 1286–1290.

[59] L. Métivier, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, Leading Edge **35**, 1060 (2016).

[60] B. Sun and T. Alkhalifah, Leading Edge **38**, 185 (2019).

[61] J. Messud, D. Carotti, O. Hermant, A. Sedova, and G. Lambaré, First Break **39**, 45 (2021).

[62] L. Métivier, A. Allain, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, Geophysics **83**, R515 (2018).

[63] A. Górszczyk, R. Brossier, and L. Metivier, J. Geophys. Res. Solid Earth **126**, e2020JB021504 (2021).

[64] L. Métivier, R. Brossier, Q. Mérigot, and E. Oudet, Inverse Probl. **35**, 085001 (2019).

[65] A. Pladys, R. Brossier, N. Kamath, and L. Métivier, Geophysics **87**, R261 (2022).

[66] C. G. Constable, Geophys. J. Int. **94**, 131 (1988).

[67] Z. Xiao, D. Xiao, V. Havyarimana, H. Jiang, D. Liu, D. Wang, and F. Zeng, IEEE Internet Things J. **6**, 10652 (2019).

[68] J. Li and M. D. Sacchi, Geophysics **86**, V171 (2021).

[69] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Society for Industrial and Applied Mathematics, Philadelphia, 2005).

[70] E. Crase, A. Pica, M. Noble, J. McDonald, and A. Tarantola, Geophysics **55**, 527 (1990).

[71] A. Y. Aravkin, M. P. Friedlander, F. J. Herrmann, and T. van Leeuwen, Math. Program. **134**, 101 (2012).

[72] P. T. C. Carvalho, S. L. E. F. da Silva, E. F. Duarte, R. Brossier, G. Corso, and J. M. de Araújo, Geophys. J. Int. **229**, 35 (2021).

[73] J. T. Barron, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2019), p. 4331.

[74] A. Pladys, R. Brossier, Y. Li, and L. Métivier, Geophysics **86**, R563 (2021).

[75] B. Bollobás, *Modern Graph Theory*, Graduate Texts in Mathematics (Springer, New York, NY, 2002).

[76] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, New York, 2006).

[77] B. Olofsson, P. Mitchell, and R. Doychev, Leading Edge **31**, 457 (2012).

[78] S. da Silva, A. Karsou, R. Moreira, J. Lopez, and M. Cetale, Klein-Gordon Equation and Variable Density Effects on Acoustic Wave Propagation in Brazilian Pre-Salt Fields, in *Conference Proceedings of 83rd EAGE Annual Conference & Exhibition* (European Association of Geoscientists & Engineers, 2022).

[79] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, J. Sci. Comput. **16**, 1190 (1995).

[80] R.-E. Plessix, Geophys. J. Int. **167**, 495 (2006).

[81] R. Brossier, S. Operto, and J. Virieux, Geophysics **75**, R37 (2010).

[82] R. Ishiwata, R. Kinukawa, and Y. Sugiyama, Sci. Rep. **8**, 6367 (2018).