Physica A 548 (2020) 124473

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Robust full-waveform inversion using q-statistics

Sérgio Luiz Eduardo Ferreira da Silva [a,*], Carlos A.N. da Costa [a], Pedro Tiago C. Carvalho [a], João Medeiros de Araújo [a], Liacir dos Santos Lucena [a], Gilberto Corso [b]

[a] Departamento de Física Teórica e Experimental, Universidade Federal do Rio Grande do Norte, 59078-970, Natal, RN, Brazil
[b] Departamento de Biofísica e Farmacologia, Universidade Federal do Rio Grande do Norte, 59078-970, Natal, RN, Brazil

ARTICLE INFO

Article history:
Received 7 October 2019
Received in revised form 23 February 2020
Available online 17 March 2020

Keywords:
Seismic imaging
q-Gaussian
Inverse theory
FWI
Tsallis entropy

ABSTRACT

Imaging of the subsurface is central in seismic exploration and a topic of great economic interest. A promising technique for seismic imaging is a wave-equation-based method called full-waveform inversion (FWI). FWI is a data-fitting technique that minimises the difference between the observed data in the seismic records and the simulated data, which is extracted from the solution of the wave equation. Usually, FWI is formulated as an optimisation problem that minimises the least-squares distance. In the perspective of likelihood theory, the minimisation of the least-squares distance assumes a Gaussian distribution of the residual data. In this work, we deal with the q-Gaussian distribution associated with the Tsallis statistics to construct a robust optimisation problem, which we call q-FWI. We tested our method in a typical geophysics velocity model with noisy data. Our results show that q-FWI, based on the q-statistics, is a powerful methodology in noisy environments, especially in the presence of outliers. The long tail of the q-distribution exploits the outliers' information, which helps in the image reconstruction. Furthermore, q-FWI provides better image reconstruction without additional computational cost compared to the traditional approach of using the Gaussian distribution for the residuals.

1. Introduction

The imaging of underground structures is central in seismic exploration and a topic of great economic interest [1]. The most-employed geophysical data acquisition method uses acoustic sources that produce mechanical waves that reflect and refract in the subsurface layers to be recorded in seismic sensors (receivers). A promising technique for seismic imaging is a wave-equation-based method called full-waveform inversion (FWI) [2,3]. FWI is a data-fitting technique for obtaining high-resolution images of the subsurface by minimising the difference between the observed and simulated seismic data (residuals) [4,5]. FWI is a technique that has received increasing attention from both exploration and global geophysics because it uses the full information contained in the waveforms, considers all wave phenomena and obtains sharper and more accurate images of the subsurface than traditional seismic imaging techniques, such as seismic travel-time tomography [6,7] and inversion based on surface wave dispersion [8].

FWI is a non-linear problem, and therefore the subsurface models are iteratively updated using the full information of the propagated seismic wave travelling through the modelled geological structure [9] using a gradient-based optimisation

* Corresponding author.
E-mail address: sergioluiz.pesquisa@gmail.com (S.L.E.F. da Silva).

https://doi.org/10.1016/j.physa.2020.124473
0378-4371/

method that leads to the best fit between simulated and observed data [2,3]. In each iteration, the wavefields generated by the seismic sources are simulated by the wave equation, and the simulated data are compared with the observed. The subsurface properties model of the previous iteration is updated to decrease the residuals, which are saved to update the model in the next iteration.

However, despite its potential, it is still a technique that faces some challenges. The main challenge for FWI is that it is inherently an ill-posed inverse problem [10] because many models fit the same data set (non-unity of solution). In addition, the presence of noise decreases the accuracy of the measurements and makes the FWI unstable and difficult to solve, increasing the ill-posedness of the problem. Moreover, in the 1960s several researchers noted that to solve ill-posed problems it is necessary to provide additional information. Therefore, several regularisation techniques for producing adequate models were proposed [11–13]. However, all these techniques impose characteristics on the solution, but we do not have *a priori* information about the subsurface model or the type of noise present in the observed data.

In the literature, the most widely used objective function in the optimisation process consists in the minimisation of the least-squares norm (L_2-norm) of the residuals [14]. Moreover, the minimisation of the least-squares distance is very popular, and from a theoretical perspective, it assumes a Gaussian distribution of residuals [15]. However, in cases where the residual probability distribution is quite different from the Gaussian probability distribution, the image FWI produces are not satisfactory. Indeed, the least-squares distance suffers from sensitivity to noise, especially to data outliers.

In this work, we go beyond the Gaussian distribution to solve the FWI problem and introduce an alternative objective function which assumes a more general and flexible probability distribution for the residuals. This new class of distribution has statistical properties that are closer to the computed residual distribution. This new probability distribution is obtained from the maximisation of Tsallis entropy [16] under appropriated constraints [17] and is known as q-Gaussian probability distribution [18,19]. The q-Gaussian distribution is not rare, and also arise as a function of two random gamma variables [20] as well as solutions of generalised Fokker–Planck equations [21]. A wide class of systems obeys q-Gaussian statistics [22], such as in weak chaos detection [23], a generalisation of Gauss' law of error [24], rotational velocity of stars [25], as well as random walk model to describe animal cell migration tracks [26].

This paper is organised as follows: In Section 2 we briefly review the usual formulation of the FWI. In Section 3 we use the Tsallis entropy and apply the concepts of maximum entropy [27,28] and maximum likelihood [29] to obtain the q-objective function. Also, we explicitly derive the gradient function relative to the new objective function. In Section 4 we apply our method to a classical model used in exploration geophysics and compare the traditional FWI results with the q-FWI, which produces a stable and robust objective function. Finally, in Section 5 we discuss the advantages of our approach and show the superiority of our methodology to overcome noise with strong outliers in the data.

2. Full-waveform inversion

Full-waveform inversion is formulated as a nonlinear inverse problem. The forward problem consists of modelling the propagation of the seismic waves using the wave equation. The FWI problem generally uses the acoustic approximation [9]. In this wave equation version, the model parameters are the velocities of the acoustic waves (P-wave) of the medium. The inverse problem consists in determining the velocity model parameters iteratively using information extracted from the full waveform recorded in the receivers (observed seismic data).

The acoustic wave equation in the time domain for a medium with velocity $c(\mathbf{x})$ is given by:

$$\nabla^2 u(\mathbf{x}, t) - \frac{1}{c^2(\mathbf{x})}\frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} = f(t)\delta(\mathbf{x} - \mathbf{x}_s) \tag{1}$$

where $\mathbf{x}$ is the position (spatial coordinates), the time is denoted by t, ∇^2 is the Laplacian operator, u is the seismic wavefield, and the term denoted by $f(t)\delta(\mathbf{x} - \mathbf{x}_s)$ represents the seismic source at the position $\mathbf{x}_s$ where δ is the Dirac Delta function. Applying the Fourier transform to Eq. (1), we have the wave equation in the frequency-space domain:

$$\nabla^2 \hat{u}(\mathbf{x}, \omega) + \frac{\omega^2}{c^2(\mathbf{x})}\hat{u}(\mathbf{x}, \omega) = \hat{f}(\omega)\delta(\mathbf{x} - \mathbf{x}_s) \tag{2}$$

where $\hat{u}$ is the seismic wavefield in the frequency domain, ω is the angular frequency, and $\hat{f}$ is the Fourier transform of the source term.

Eq. (2) can be written as a system of linear equations of the form:

$$A(\mathbf{m}, \omega)\hat{\mathbf{u}}(\omega) = \hat{\mathbf{f}}(\omega) \tag{3}$$

where the parameter of interest, $\mathbf{m} \in \mathbb{R}^M$, is defined as the squared slowness ($\mathbf{m} = 1/c^2(\mathbf{x})$) and $\hat{\mathbf{f}} \in \mathbb{C}^{M\times 1}$. In addition, $\mathbf{A}(\mathbf{m}, \omega) \in \mathbb{C}^{M\times M}$ is the discretised Helmholtz operator (impedance matrix) [30,31] for the frequency-domain acoustic wave equation given by $\mathbf{A}(\mathbf{m}, \omega) = \mathbf{L} + \omega^2 diag(\mathbf{m})$ with $\mathbf{L} \in \mathbb{C}^{M\times M}$ the discretised Laplacian operator. Finally, M is the number of model parameters (discretised points).

In the traditional approach, FWI is formulated as a least-squares-constrained optimisation problem [14]. The frequency-domain FWI is given by the optimisation problem below:

$$\min_{\mathbf{m},\hat{u}} \ \frac{1}{2}\sum_{s,r} \left\| S_{s,r}\hat{u}_s - d_{s,r} \right\|_2^2 \qquad subject \quad to \qquad A(\mathbf{m})\hat{u}_s = \hat{f}_s \tag{4}$$

where $S_{s,r} \in \mathbb{R}^{N_r \times M}$ represents the sampling operator that extracts the seismic wavefield $\hat{u}_s \in \mathbb{C}^{M\times 1}$ at the receiver (r index) positions for the ith source (s index) [32], $d_{s,r} \in \mathbb{C}^{N_r \times 1}$ is the observed seismic data and $\|.\|_2$ denotes the least-squares distance or the L_2-norm. N_r is the number of receivers. In Eq. (4) the angular frequency (ω) and the spatial coordinate ($\mathbf{x}$) are both implicit.

As the wave equation is always satisfied, the constraint in Eq. (4) can be eliminated, i.e., $\hat{u}_s = A^{-1}(\mathbf{m})\hat{f}_s$ where $\hat{u}_s$ is the solution of the wave equation for the source $\hat{f}_s$. Thus, the modelled data is obtained by: $d^{mod}_{s,r}(\mathbf{m}) = S_{s,r}\hat{u}(\mathbf{m}) = S_{s,r}A^{-1}(\mathbf{m})\hat{f}_s$.

Let the residuals be given by the misfit between the modelled data and the observed data: $\Delta d_{s,r}(\mathbf{m}) = d^{mod}_{s,r}(\mathbf{m}) - d_{s,r}$; the model parameters are recovered by solving the following unconstrained optimisation problem:

$$\min_{\mathbf{m}} \ \phi_{L_2}(\mathbf{m}) \coloneqq \frac{1}{2}\sum_{s,r} \Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m}) \tag{5}$$

where the superscript † refer to the transpose conjugate (adjoint).

Another way to study the objective function shown in Eq. (5) is using the probabilistic maximum likelihood method [15,29,33]. In this approach, the traditional FWI determines the *maximum a posteriori* solution to estimate the parameters of the model. In this context, to minimise $\phi_{L_2}(\mathbf{m})$ is equivalent to maximising the Gaussian likelihood given by:

$$\mathcal{L}_G(\mathbf{m}) \propto \exp\left(-\frac{1}{2}\sum_{s,r} \Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m})\right) \tag{6}$$

where $\mathcal{L}_G(\mathbf{m})$ is the *posteriori probability*. Note that the *maximum a posteriori* value of $\mathbf{m}$ in Eq. (6) is obtained by minimising the negative of $\ln(\mathcal{L}_G(\mathbf{m}))$, which is equivalent to minimising Eq. (5). Furthermore, we notice that we are dealing with a standard normal distribution, that is, a distribution with mean zero ($\mu = 0$) and variance one ($\Omega^2 = 1$). We notice also that, in order to assume an *a priori* Gaussian distribution for the residuals, we have to suppose that the residuals are independent and normally distributed [15].

3. FWI and the q-entropy

3.1. Tsallis entropy and probability distribution estimation

Entropy is an important concept in statistical physics and has been applied in diverse contexts, from communication theory [34] to biological systems [35]. For the continuous case, the Boltzmann–Gibbs–Shannon (BGS) entropy [36] is expressed as:

$$E_s\Big(p(x)\Big) = -k\int_{\mathcal{X}} p(x)\ \ln\big(p(x)\big)dx \tag{7}$$

where k is a conventional positive constant, and $p(x)$ is the probability density function over a random variable $x \in \mathcal{X}$.

Furthermore, $p(x)$ is normalised:

$$\int_{\mathcal{X}} p(x)dx = 1 \tag{8}$$

Inspired by multifractal concepts, Tsallis postulated a generalisation of the BGS entropy [16] which is defined by:

$$E_q\Big(p(x)\Big) \equiv \frac{k}{q-1}\left(1 - \int_{-\infty}^{\infty} p^q(x)dx\right) \tag{9}$$

where q is a real parameter. The connection between the Tsallis framework and BGS framework is attained at the limit $q \to 1$ [16].

The q-probability distribution can be derived by applying the maximum entropy principle [27] in the Tsallis entropy (Eq. (9)) under appropriate constraints [17,37]. From this method, we obtain the q-likelihood function ($\mathcal{L}_q$) that best represents the stochastic behaviour associated with the q-parameter. In this way, there are several Tsallis distributions which depend on the constraints. In this work, we considered the constraint given in Eq. (8) and also

$$\langle x^2\rangle_q \equiv \frac{\int_{-\infty}^{\infty} p^q(x)x^2dx}{\int_{-\infty}^{\infty} p^q(x)dx} = \sigma^2 \tag{10}$$

where σ^2 represents the normalised q-expectation [17,19].

Following along Gibbs–Jaynes' path, the maximisation of the Tsallis entropy under constraints (8) and (10) yields to the q-Gaussian distribution [16,19]:

$$p_q(x) = \frac{1}{\sigma\mathcal{K}_q}\left[1 + \frac{q-1}{3-q}\frac{x^2}{\sigma^2}\right]^{\frac{1}{1-q}} \tag{11}$$

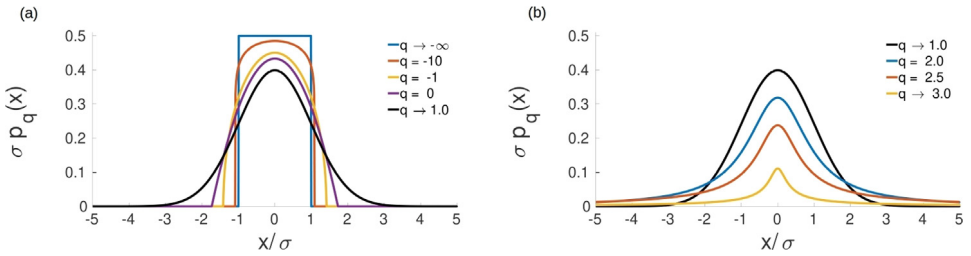

Fig. 1. Probability plots of q-Gaussian distributions for typical values of q. The black curve represents the Gaussian distribution ($q \to 1$) with zero mean. (a) In the $q < 1$ case, the curve has a cut-off (bounded random variable) and tends to a uniform distribution in the special case: $q \to -\infty$. (b) In the $1 < q < 3$ case, the q-Gaussian looks like a Gaussian distribution, being symmetrical and bell-shaped, but with a long tail.

where $\mathcal{K}_q$ is the normalising constant that satisfy condition (8). According to [19]:

$$\mathcal{K}_q = \begin{cases} \sqrt{\frac{\pi(3-q)}{q-1}}\,\frac{\Gamma\left(\frac{2-q}{1-q}\right)}{\Gamma\left(\frac{5-3q}{2(1-q)}\right)} & \text{for} \quad -\infty < q < 1 \\ \\ \sqrt{\frac{\pi(3-q)}{1-q}}\,\frac{\Gamma\left(\frac{3-q}{2(q-1)}\right)}{\Gamma\left(\frac{1}{q-1}\right)} & \text{for} \quad 1 < q < 3 \end{cases} \tag{12}$$

where $\Gamma(z)$ is the Gamma Function [38]. In the $q < 1$ case, the distribution is light-tailed [39], as can be seen in Fig. 1(a). The particular case $q \to -\infty$ corresponds to the uniform distribution (blue curve in Fig. 1(a)). For $q \to 1$ (both right and left limits) we recover the Gaussian distribution (black curve in Fig. 1). In the $1 < q < 3$ case, the q-Gaussian resembles the Gaussian distribution, being symmetrical and bell-shaped, but with long tails that decrease as a power law: $p_q(|x|) \propto |x|^{-2/(q-1)}$ [22]. For $q = 2$ the q-Gaussian distribution becomes the Cauchy distribution. In the $q \geq 3$ case, no q-Gaussian distribution exists because it does not satisfy the normalisation condition in Eq. (8). Moreover, the $p_q(x)$ probability distribution is symmetric, and has mean zero. Therefore, the mean and the variance are the same as the Gaussian distribution. We note that the quality of our results is not influenced by the normalised q-expectation, Eq. (10), so we consider $\sigma^2 = 1$.

3.2. q-Gaussian objective function

Following the FWI notation shown in the previous section, for $x = \Delta d(\mathbf{m})$ and $\sigma^2 = 1$, Eq. (11) assumes the form:

$$p_q(\mathbf{m}) = \frac{1}{\mathcal{K}_q}\left[1 + \left(\frac{q-1}{3-q}\right)\Delta d^{\dagger}(\mathbf{m})\Delta d(\mathbf{m})\right]^{\frac{1}{1-q}}. \tag{13}$$

To obtain a good estimate of the model parameters ($\mathbf{m}$), we must obtain the parameters of the model that maximise the likelihood function of the observed data ($d_{s,r}$). In this way, the q-Gaussian likelihood function is derived using Eq. (13) from the maximum likelihood method:

$$\mathcal{L}_q(\mathbf{m}) = \prod_{s,r} p_q\big(\Delta d_{s,r}(\mathbf{m})\big) \tag{14}$$

Taking the natural logarithm of the Eq. (14):

$$ln\big(\mathcal{L}_q(\mathbf{m})\big) = ln\left\{\prod_{s,r}\frac{1}{\mathcal{K}_q}\left[1 + \left(\frac{q-1}{3-q}\right)\Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m})\right]^{\frac{1}{1-q}}\right\} \tag{15}$$

That is:

$$ln\big(\mathcal{L}_q(\mathbf{m})\big) = ln\left(\frac{1}{\mathcal{K}_q}\right) + \frac{1}{1-q}\sum_{s,r} ln\left[1 + \left(\frac{q-1}{3-q}\right)\Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m})\right] \tag{16}$$

We notice that to minimise the negative of Eq. (16) is equivalent to minimising the following function:

$$\phi_q(\mathbf{m}) = \frac{1}{q-1}\sum_{s,r} ln\left[1 + \left(\frac{q-1}{3-q}\right)\Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m})\right] \tag{17}$$

This is the q-Gaussian objective function, and we call q-FWI the optimisation problem that uses this objective function for seismic inversion.

The conventional objective function is recovered from Eq. (17) at the limit $q \to 1$ using the L'Hôpital's rule :

$$\lim_{q\to 1}\phi_q(\mathbf{m}) = \lim_{q\to 1}\sum_{s,r}\frac{\frac{\partial}{\partial q}\left\{ln\left[1+\left(\frac{q-1}{3-q}\right)\Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m})\right]\right\}}{\frac{\partial(q-1)}{\partial q}}$$

$$\lim_{q\to 1}\phi_q(\mathbf{m}) = \lim_{q\to 1}\sum_{s,r}\frac{2(3-q)^{-2}\Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m})}{1+\left(\frac{q-1}{3-q}\right)\Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m})}$$

$$\lim_{q\to 1}\phi_q(\mathbf{m}) = \frac{1}{2}\sum_{s,r}\Delta d_{s,r}(\mathbf{m})^{\dagger}\Delta d_{s,r}(\mathbf{m}) = \phi_{L_2}(\mathbf{m}) \tag{18}$$

quod erat demonstrandum.

3.3. Gradient of the q-Gaussian objective function

FWI is usually performed using a gradient-based method using a second-order Taylor–Lagrange approximation [14]. Starting with an initial model $\mathbf{m}_0$, the iterative process to obtain an optimal minimum consists in updating the model according to:

$$\mathbf{m}_{j+1} = \mathbf{m}_j - \alpha_j \mathbf{H}^{-1}\nabla_{\mathbf{m}}\phi(\mathbf{m}_j) \tag{19}$$

where α_j is the steplength computed through a search strategy [40], $\mathbf{H}$ is the Hessian matrix and $\nabla_{\mathbf{m}}\phi(\mathbf{m}_j) = \partial\phi(\mathbf{m}_j)/\partial\mathbf{m}$ is the gradient of the objective function $\phi(\mathbf{m}_j)$ with respect to model parameters in the jth iteration. The Hessian is a square matrix composed of second-order partial derivatives of the objective function.

The derivative of the objective function $\phi_q(\mathbf{m})$ with respect to each model element m_l is given by:

$$\frac{\partial\phi_q(\mathbf{m})}{\partial m_l} = \frac{1}{q-1}\sum_{s,r}\frac{\left(\frac{q-1}{3-q}\right)\frac{\partial(\Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m}))}{\partial m_l}}{1+\left(\frac{q-1}{3-q}\right)\Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m})} \tag{20}$$

or:

$$\frac{\partial\phi_q(\mathbf{m})}{\partial m_l} = \sum_{s,r}\mathbb{R}\left[\left(\frac{\partial\Delta d_{s,r}(\mathbf{m})}{\partial m_l}\right)^{\dagger}\frac{2\Delta d_{s,r}(\mathbf{m})}{3-q+(q-1)\Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m})}\right] \tag{21}$$

where $\mathbb{R}$ denotes the real part and:

$$\left(\frac{\partial\Delta d_{s,r}(\mathbf{m})}{\partial m_l}\right)^{\dagger} = -\omega^2 S^{\dagger}_{r,s} diag(\mathbf{u}_s)^{\dagger}A^{-1}(\mathbf{m}) \tag{22}$$

is known as the Fréchet derivative [14].

For comparison, the gradient of the objective function of the traditional FWI is given by:

$$\frac{\partial\phi_{L_2}(\mathbf{m})}{\partial m_l} = \sum_{s,r}\mathbb{R}\left[\left(\frac{\partial\Delta d_{s,r}(\mathbf{m})}{\partial m_l}\right)^{\dagger}\Delta d_{s,r}(\mathbf{m})\right] \tag{23}$$

Note that when $q \to 1$ the q-FWI objective function becomes the traditional FWI objective function, which gradient is given by Eq. (23) [14]. Comparing Eqs. (21) and (23), one can see that the gradient of the proposed objective function is weighted by the factor $2/[3-q+(q-1)\Delta d^{\dagger}_{s,r}(\mathbf{m})\Delta d_{s,r}(\mathbf{m})]$. In other words, the ith term of Eq. (21) is down-weighted by the magnitude of the ith residual. That means the gradient of the q-Gaussian objective function becomes less sensitive to outliers (large errors) compared to the gradient of the traditional objective function.

4. Numerical results

To demonstrate the potentialities of the proposed objective function, we use the Marmousi velocity model [41] as the target P-wave velocity model (true model), which is shown in Fig. 2(a). The model is largely used to test seismic imaging reconstruction techniques. Reconstructing this model is a challenge because it presents a geometry with many reflectors, sloping geological strata, and strong medium velocity variations, from 1.5 km/s to 4.7 km/s. Fig. 2(b) shows the initial model used to start the optimisation process, which is obtained from the true velocity model using a Gaussian smoother with a standard deviation of 400 m. The models consist of 700 and 315 grid cells in the horizontal and vertical directions, respectively (220,500 total grid points).

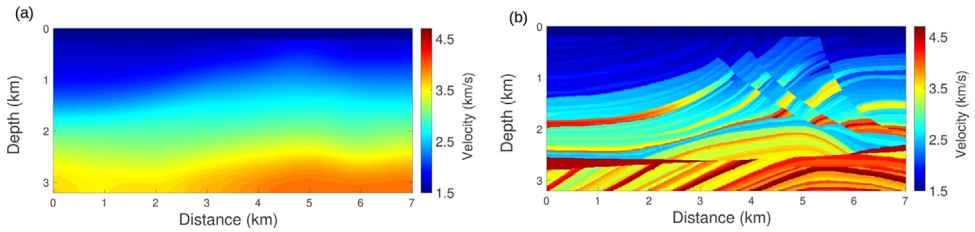

Fig. 2. (a) True P-wave Marmousi velocity model. (b) Initial P-wave velocity model obtained from smoothing the true model by a Gaussian smoother with standard deviation of 400 m.

The seismic wavefield modelling is performed using a 2-D frequency-domain acoustic wave equation with a classical 9-point finite-difference stencil with perfectly matched layer (PML) absorbing boundaries [42] for the spatial discretisation with Dirichlet boundary condition to simulate an infinite medium.

The water layer, on top of the model, is assumed to be known and is kept at a constant with velocity of 1.5 km/s during the optimisation process. A Ricker wavelet [43] centred at 8 Hz is used as source term. Because of the fast convergence, the optimisation process is performed with the quasi-Newton method *l*-BFGS (*Limited-memory* Broyden–Fletcher–Goldfarb–Shanno) [40].

The data was generated using an in-line fixed acquisition at a depth of 50 m, involving $N_s = 35$ equally spaced sources located every 200 m, from 10 m to 6,810 m. For each source, $N_r = 175$ receivers were located every 40 m, from 10 m to 6,970 m, at a depth of 50 m. The impedance matrix (Helmholtz operator) was discretised on a grid with a spacing of 10 m, to guarantee at least 7 wavelengths per discretisation point. In the inversion process, we employed sequentially four frequency groups, from the lowest to the highest frequencies, to mitigate the non-linearity of the problem [44]: {3, 4, 5}, {6, 7, 8}, {9, 10, 11} and {12, 13, 14} Hz. For each frequency group, we computed 20 *l*-BFGS iterations.

We considered two scenarios to illustrate the robustness and stability of our proposal. In the first one, we simulate the FWI with Gaussian noise added to the original data. In the second case, we added Gaussian noise and spiky noise (outliers) to the original data. We emphasise that the initial model and the acquisition geometry are the same in all numerical experiments. Tests have been performed on the data without added noise, but in this case the results of inversion with FWI or q-FWI methods are similar.

We performed five inversions for each scenario: in the first, we employ the traditional approach (Eq. (5)), and in the others we use the q-FWI with $q = 1.1, 1.5, 2.0, 2.5$ and 2.9 (Eq. (17)). We chosen the values in the interval $1 < q < 3$ because the q-Gaussian distribution is favoured for its long tails in comparison to the Gaussian, making it less sensitive to outliers. Diversely, in the $q < 1$ case the q-distribution is light-tailed, and the random variable is bounded. The long tail is interesting because the outlier residuals will not be neglected during the data inversion process.

4.1. Addition to Gaussian noise

In the first scenario, to demonstrate the robustness of the proposed method, we use a data contaminated by Gaussian white noise with a signal-to-noise ratio (SNR) of 10 dB. The inversion results using traditional FWI and q-FWI with $q = 1.1, 1.5, 2.0, 2.5$ and 2.9 are illustrated in Fig. 3(a)–(f), respectively. Both traditional FWI and q-FWI yield estimations close to the true model in the shallowest geological structures (less than 2 km). However, layers with depths greater than 2 km are not well reconstructed by the traditional FWI. Note that the velocity structure is significantly improved at depths greater than 2.0 km, where the q-FWI method reconstructs layers with high resolution, mainly for $q = 2.5$.

We compare the reconstruction of the traditional FWI and q-FWI cases with the true model. The velocity model relative errors corresponding to Fig. 3(a)–(f) are 5.7, 5.6, 5.2, 4.9, 4.6, and 5.0%, respectively.

4.2. Addition to Gaussian noise and outliers

In this second scenario, we repeat the previous experiment except that we have contaminated 12% of the recorded seismic traces in the receivers with Gaussian noise of the same energy as the observed data, to simulate the presence of spikes (outliers), i.e., doubling its values. The $21 = 0.12 \times N_r$ receivers that had the signals disturbed with spiky noise were chosen randomly, using a uniform distribution. This simulates a situation where powerful data pre-processing is not needed. The data collected are, in many cases, contaminated by outliers.

Fig. 4(a) shows the model obtained using the traditional FWI. Fig. 4(b)–(f) show the q-FWI inversion results for $q = 1.1, 1.5, 2.0, 2.5$ and 2.9, respectively. Note that the inversion with the traditional FWI fails to achieve a good reconstruction, but the proposed method provides images close to the true model. Traditional FWI generated many artefacts from the velocity model estimation obtained with the q-FWI.

In the presence of outliers, the traditional FWI fails to recover most of the velocity model, while our method succeeds to recover the main structures of the velocity model. The performance of the q-FWI with outliers is comparable to the traditional FWI with only Gaussian noise as shown in Section 4.1. The velocity model relative errors corresponding to Fig. 4(a)–(f) are 14.0, 6.0, 4.9, 4.6, 4.5, and 5.1%, respectively.

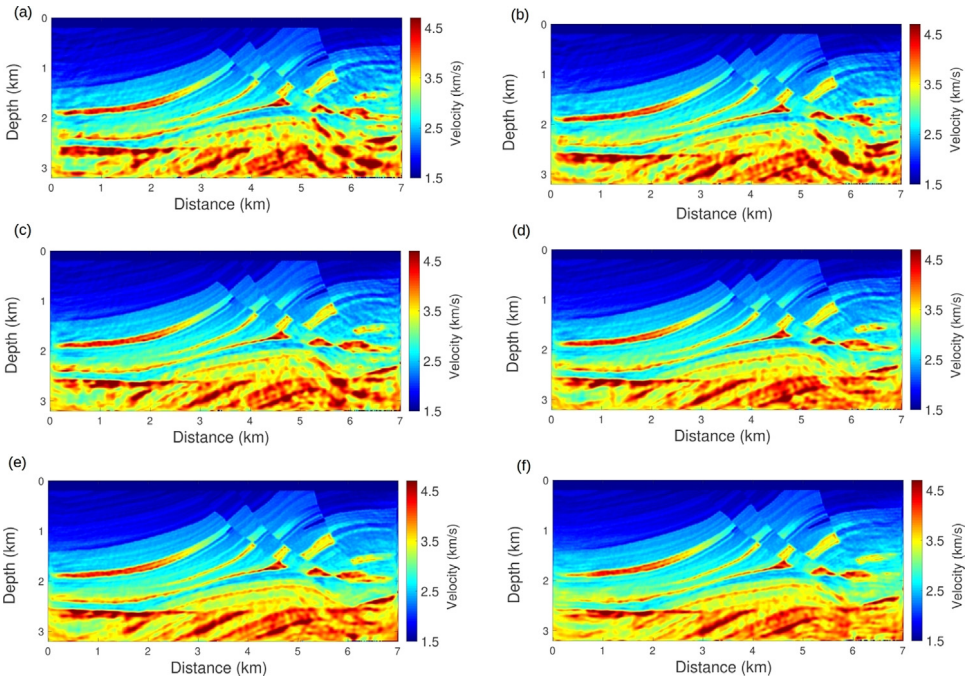

Fig. 3. Inversion results for Gaussian-noise-added case. (a) Velocity estimation with traditional FWI; velocity estimation with q-FWI for: (b) $q = 1.1$, (c) $q = 1.5$, (d) $q = 2.0$, (e) $q = 2.5$, (f) $q = 2.9$.

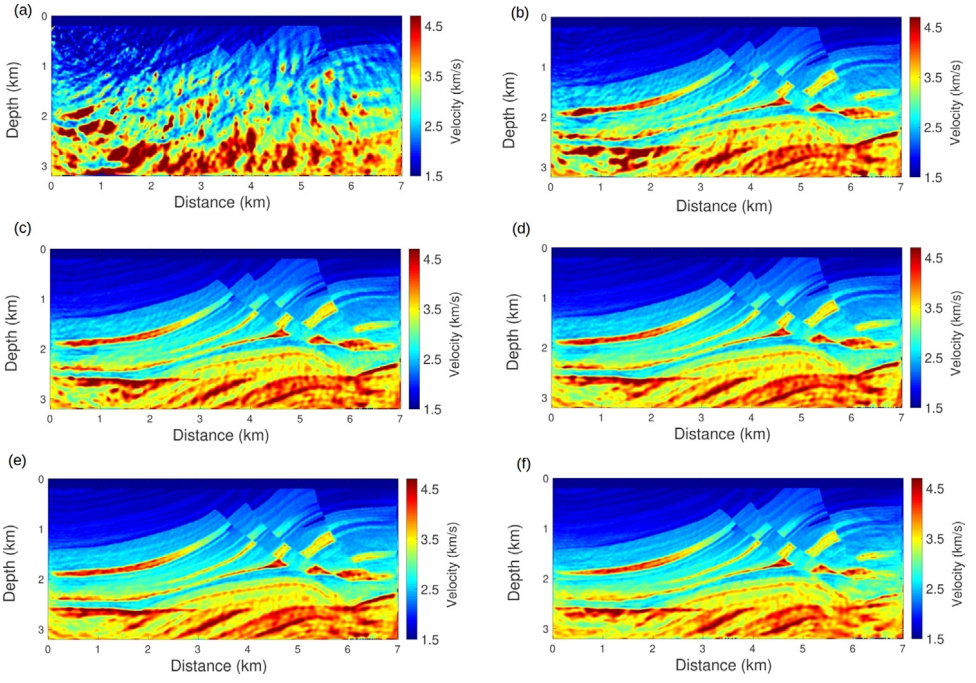

Fig. 4. Inversion results with outliers in the data set. (a) Velocity estimation with traditional FWI; velocity estimation with q-FWI for: (b) $q = 1.1$, (c) $q = 1.5$, (d) $q = 2.0$, (e) $q = 2.5$, (f) $q = 2.9$.

5. Conclusion

To mitigate the influence of noise in the reconstruction of the velocity models, we proposed a novel objective function based on the q-Gaussian distribution. We call the proposed method by the abbreviation q-FWI, in reference to the

q-statistics. A numerical study with a high-resolution complex velocity model demonstrates the effectiveness of the proposed inversion methodology. In this complex model, we tested the robustness and stability of the q-statistics for two scenarios with noise: with and without outliers.

The results show that the FWI based on q-statistics, and by extension on q-entropy, is a powerful methodology in noisy environments. The proposed objective function of the q-FWI is less sensitive to the presence of noise, especially to outliers. In addition, the q-FWI provides a better image reconstruction, for no additional computational cost, than the traditional approach. In our tests, the best q-parameter for the reconstruction model was $q = 2.5$. The long tail of the q-distribution incorporated the information of the outliers and played an optimal role in the reconstruction of the image. In conclusion, the FWI based on q-statistics provides better results than traditional FWI based on the Gaussian distribution in noisy data environments and may become a valuable tool in exploration geophysics, especially when the data quality is poor and it is desirable to bypass the heavy data pre-processing work that would otherwise be necessary.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Sérgio Luiz Eduardo Ferreira da Silva: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Carlos A.N. da Costa:** Conceptualization, Methodology, Validation, Formal analysis, Writing - original draft, Visualization. **Pedro Tiago C. Carvalho:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **João Medeiros de Araújo:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Liacir dos Santos Lucena:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Visualization. **Gilberto Corso:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Supervision.

Acknowledgements

The authors gratefully acknowledge support from Shell Brazil through the *Methods for Full Waveform Inversion* project at Federal University of Rio Grande do Norte and the strategic importance of the support given by Brazilian National Agency for Petroleum, Natural Gas and Biofuels (ANP) through the R&D levy regulation. The authors also gratefully acknowledge support from the Brazilian National Council for Scientific and Technological Development (CNPq). We also thank the two anonymous reviewers and the editor of Physica A for their comments that helped to improve the paper throughout the revision.

References

[1] O. Yilmaz, Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data, Society of exploration geophysicists, 2001.
[2] A. Tarantola, Inversion of seismic reflection data in the acoustic approximation, Geophysics 49 (8) (1984) 1259–1266.
[3] P. Lailly, The seismic inversion problem as a sequence of before stack migrations, in: Conference on Inverse Scattering, Theory and Application. Conference on Inverse Scattering: Theory and Application, in: Proceedings in Applied Mathematics Series, SIAM, 1983, pp. 206–220.
[4] L. Sirgue, R.G. Pratt, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies, Geophysics 69 (1) (2004) 231–248.
[5] J. Virieux, A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, W. Zhou, 6. an introduction to full waveform inversion, in: Encyclopedia of Exploration Geophysics, 2017, pp. R1–1–R1–40.
[6] Seismic traveltime Tomography, in: Fundamentals of Geophysical Interpretation, Society of Exploration Geophysicists, 2004.
[7] R.P. Bording, A. Gersztenkorn, L.R. Lines, J.A. Scales, S. Treitel, Applications of seismic travel-time tomography, Geophys. J. Int. 90 (2) (1987) 285–303.
[8] Z. Liu, J. Li, G. Schuster, 3d wave-equation dispersion inversion of surface waves, in: SEG 2017 Workshop: Full-Waveform Inversion and beyond, Beijing, China, 20-22 November 2017, 2017, pp. 22–26.
[9] A. Fichtner, Full Seismic Waveform Modelling and Inversion, Springer Verlag, 2010.
[10] J. Hadamard, Sur les problèmes aux dérivés partielles et leur signification physique, Princet. Univ. Bull. 13 (1902) 49–52.
[11] A. N. Tikhonov, V. Ya. Arsenin, Solution of ill-posed problems, Math. Comput. - Math. Comput. 32 (1978).
[12] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1) (1992) 259–268.
[13] M. Benning, M. Burger, Modern regularization methods for inverse problems, Acta Numer. 27 (2018) 1–111.
[14] J. Virieux, S. Operto, An overview of full-waveform inversion in exploration geophysics, Geophysics 74 (2009) WCC1–WCC26.
[15] A. Tarantola, Inverse Problem Theory - and Methods for Model Parameter Estimation, SIAM, 2005.
[16] C. Tsallis, Possible generalization of boltzmann-gibbs statistics, J. Stat. Phys. 52 (1) (1988) 479–487.
[17] C. Tsallis, R. Mendes, A. Plastino, The role of constraints within generalized nonextensive statistics, Physica A 261 (3) (1998) 534–554.
[18] H.J. Hilhorst, G. Schehr, A note onq-gaussians and non-gaussians in statistical mechanics, J. Stat. Mech. Theory Exp. 2007 (06) (2007) P06003.
[19] D. Prato, C. Tsallis, Nonextensive foundation of levy distributions, Phys. Rev. E 60 (1999) 2398–2401.
[20] A.A. Budini, Extended q-gaussian and q-exponential distributions from gamma random variables, Phys. Rev. E 91 (2015) 052113, http://dx.doi.org/10.1103/PhysRevE.91.052113.
[21] A. Plastino, A. Plastino, Non-extensive statistical mechanics and generalized fokker-planck equation, Physica A 222 (1) (1995) 347–354.
[22] S. Picoli Jr., R.S. Mendes, L.C. Malacarne, R.P.B. Santos, Q-distributions in complex systems: a brief review, Braz. J. Phys. 39 (2009) 468–474.

[23] C.G. Antonopoulos, H. Christodoulidi, Weak chaos detection in the fermi–pasta–ulam- system using q-gaussian statistics, Int. J. Bifurcation Chaos 21 (08) (2011) 2285–2296.
[24] H. Suyari, M. Tsukada, Law of error in tsallis statistics, IEEE Trans. Inform. Theory 51 (2) (2005) 753–757.
[25] J.C. Carvalho, R. Silva, J.D. do Nascimento jr., B.B. Soares, J.R.D. Medeiros, Observational measurement of open stellar clusters: A test of kaniadakis and tsallis statistics, EPL (Europhys. Lett.) 91 (6) (2010) 69002, http://dx.doi.org/10.1209/0295-5075/91/69002.
[26] P.C.A. da Silva, T.V. Rosembach, A.A. Santos, M.S. Rocha, M.L. Martins, Normal and tumoral melanocytes exhibit q-gaussian random search patterns, Plos One 9 (9) (2014) 1–13.
[27] E.T. Jaynes, Information theory and statistical mechanics, Phys. Rev. 106 (1957) 620–630.
[28] E.T. Jaynes, Information theory and statistical mechanics. ii, Phys. Rev. 108 (1957) 171–190.
[29] D. Ferrari, Y. Yang, Maximum l q -likelihood estimation, Ann. Statist. 38 (2) (2010) 753–783.
[30] Z. Chen, D. Cheng, W. Feng, T. Wu, An optimal 9-point finite difference scheme for the helmholtz equation with pml, Int. J. Numer. Anal. Model. 10 (2013) 389–410.
[31] K.J. Marfurt, Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations, Geophysics 49 (5) (1984) 533–549.
[32] T. van Leeuwen, F.J. Herrmann, A penalty method for PDE-constrained optimization in inverse problems, Inverse Problems 32 (1) (2015) 015007.
[33] W. Menke, Geophysical Data Analysis: Discrete Inverse Theory, Academic Press, 1984, pp. 79–99.
[34] M. Kulisiewicz, P. Kazienko, B.K. Szymanski, R. Michalski, Entropy Measures of Human Communication Dynamics, Scientific Reports, 2018, pp. 2045–2322.
[35] M. Costa, A.L. Goldberger, C.-K. Peng, Multiscale entropy analysis of biological signals, Phys. Rev. E 71 (2005) 021906, http://dx.doi.org/10.1103/PhysRevE.71.021906.
[36] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.
[37] A.-H. Sato, Q-gaussian distributions and multiplicative stochastic processes for analysis of multiple financial time series, J. Phys. Conf. Ser. 201 (2010) 012008.
[38] E. Artin, M. Butler, The Gamma Function, in: Dover Books on Mathematics, Dover Publications, 2015.
[39] M.C. Bryson, Heavy-tailed distributions: Properties and tests, Technometrics 16 (1) (1974) 61–68.
[40] J. Nocedal, S.J. Wright, Numerical Optimization, second ed., Springer, New York, NY, USA, 2006.
[41] G.S. Martin, R. Wiley, K.J. Marfurt, Marmousi2: An elastic upgrade for marmousi, Lead. Edge 25 (2) (2006) 156–166.
[42] J. P. Berenger, A perfect matched layer for the absorption of electromagnetic waves, J. Comput. Phys. 114 (1994) 185–200.
[43] N. Ricker, Further developments in the wavelet theory of seismogram structure, Bull. Seismol. Soc. Am. 33 (3) (1943) 197–228.
[44] R.G. Pratt, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model, Geophysics 64 (3) (1999) 888–901.