# AVA Inversion of PP Reflection in a VTI Medium using Proximal Splitting Algorithm

**A. Zidan [1*], Y. E. Li. [1], A. Cheng [1]**

**Key Points:**

- To alleviate the ill-posedness of the seismic amplitude inversion, and obtain a unique and stable solution.

*Department of Civil and Environmental Engineering,
National University of Singapore,
21 Lower Kent Ridge Rd,
Singapore, 119077

Corresponding author: Ahmed Zidan, azidan89@hotmail.com

## Abstract

Amplitude variation with offset (AVO) inversion, particularly for more than two model parameters, is a highly ill-posed problem and, hence, regularization is indispensable. Here, we propose a regularized inverse problem to mitigate the ill-posedness of the amplitude inversion. The regularization is added to measure the difference in information between the a priori probability density function and the predicted probability density of the inverted parameters. Information theory provides a collection of contrast functions which quantify the divergence from one probability distribution to another, such as the relative entropy. The a priori density is approximated by a Gaussian mixture model, obtained from well logs and rock physics model. The mixture model is a density estimator, providing the statistical properties of the model parameters of interest. The likelihood of the data and the divergence are combined in an augmented Lagrangian scheme, the alternating direction method of multipliers (ADMM), to obtain a unique solution that best generate the recorded seismic data and satisfy the geological constraints conveyed by the a priori probability density function. The proposed inversion scheme is then applied to the anisotropy AVO inversion, for estimating the elastic and seismic anisotropy parameters of shale formations. Compared to the unconstrained minimization, the P- and S-wave velocity, and $\varepsilon$ are better recovered, moreover, density and Thomsen's $\delta$ are well-constrained.



## Plain Language Summary

Geophysical inverse problems are highly ill-posed, due to variable sensitivities of different model parameters, measurement errors, among other reason. To alleviate ill-posedness of the seismic amplitude inversion, a regularization function that measures the distance to a priori mixture model is proposed. The a priori mixture model is first obtained from well log and rock physics data, independently of the seismic data. The a priori model conveys the statistical properties of the parameters of interest. Then, the functional and regularization are combined via a proximal splitting scheme, to obtain a unique solution that is close to the a priori mixture model and best generate the observed seismic data.


## 1 Introduction

Seismic amplitude inversion has been successfully applied to predict the elastic properties of the subsurface (Hampson et al., 2005; Russell & Hampson, 2006); jointly with rock physics inversion, models for reservoir lithology and fluid properties are built (Bosch et al., 2010; Grana & Della Rossa, 2010). In a strong vertical transverse isotropy (VTI) medium, i.e., shale, isotropic approximations of the Zoeppritz equations fail to predict the correct AVO responses, which leads to incorrect lithology and fluid predictions, and fail to accurately describe the geomechanical behavior of the shale rocks (Sayers, 2013b). The anisotropy of shale is attributed to preferential orientation of clay minerals, micro-cracks, organic matter, amongst others, hence, anisotropy exists at all scale lengths (Sayers, 2013a). Since there is no direct method to measure anisotropy in vertical wells, rock physics modeling (RPM) is used to estimate the low-to-medium frequency anisotropy parameters necessary to initialize the seismic amplitude inversion for the lithology and fluid predictions. Bandyopadhyay (2009) showed that, increasing the volume fraction of kerogen increases the rock stiffnesses, and subsequently the elastic anisotropy. Consequently, the estimated seismic anisotropy parameters are of great relevance for the organic shale exploration and production.

Daley & Hron (1977) gave the exact solution for reflection and transmission coefficients in a transversely isotropic medium. Many linear approximations have been subsequently introduced assuming weak elastic contrasts (Thomsen et al., 1993; Rüger, 1997). Despite a vertical transverse medium (VTI) is described by the Thomsen's parameters $\varepsilon, \ \delta,$ and $\gamma$ (Thomsen, 1986), the P-P reflection coefficients in a VTI medium depends only on $\varepsilon$ and $\delta$ (Rüger, 1997). The effect of the NMO anisotropy $\delta$ appears at the small-to-middle angles of incidence (Banik, 1987), while P-wave anisotropy $\varepsilon$ influences larger angles of incidence greater than 30° (Kim et al., 1993). Plessix & Bork (2001) studied the AVO response in a VTI medium based on the least-squares function;

showing that, it is difficult to obtain all five parameters for small-to-medium angles of incidence due to the existence of local minimum solutions. F. Zhang et al. (2019) used implicit constraints to derive an approximation of the P-P reflection coefficients in a VTI medium, consisting of only three parameters. In addition, joint P-P and P-S amplitude inversion of a VTI medium has been proposed to better constrain the density and seismic anisotropy parameters (Nadri & Hartley, 2007; Lu et al., 2018; Luo et al., 2020; Zhou et al., 2020).

Seismic amplitude inversion is highly ill-posed due to variable-sensitivities among model parameters; moreover, a small amount of noise in the observed seismic data results in a very large change in the estimates (Tarantola, 2005). Furthermore, the density and Thomsen's anisotropy inversions are very challenging due to less information in the near-angle traces, whereas the far-angle traces are typically distorted (non-flatten) (L. Liu et al., 2013). To quote Treitel & Lines (2001), "A good match between the observed and theoretical geophysical responses provides us with a necessary, but by no means sufficient condition for the calculation to converge to the *ground truth* below us.", a solution to ill-posed problems is to add regularization that enhances the stability of the inversion, such as $L_2-$norm promoting smoothness in the solution (Velis, 2008), and $L_1-$norm to promote sparsity in the solution (Y. Wang, 2010) and as an anti-noise functional (C. Liu et al., 2015). Total variation $(TV-\text{norm})$ has been successfully used to promote sparsity of the reflection coefficients (F. Zhang et al., 2014) and acoustic impedance (Wu, He, et al., 2019). Zand et al. (2020) used the $TV-$ norm regularization to obtain a stable solution of the least-square reverse-time migration (LSRTM). Despite the norm regularization functions to provide a desired shape of the solution, it can not infer the statistical information about the model parameters of interest. Different geoscience data follow different distributions, such as the P- and S-wave velocity follow a Gaussian distribution (Hernlund & Houser, 2008) and a Lévy-stable distribution (Painter & Paterson, 1994), the P- and S-wave reflection coefficients follow a Lévy-stable distribution (Painter et al., 1995) and a Gaussian mixture distribution (Mukerji et al., 2009), and the porosity follows a log-normal distribution (Berezin, 1982). Accordingly, the Bayesian inference uses a prior probability distribution that infer the statistical properties of the model parameters to obtain the posterior distribution of the unknown model parameters, using deterministic approaches (Downton & Lines, 2001; Buland & Omre, 2003; Downton, 2005; Grana, 2020), or stochastic methods (D. Zhu & Gibson, 2018; Wu, Li, et al., 2019; K. Li et al., 2020).

Zidan (2022) proposed a generalized framework for Bayesian inversion using a lower-bound estimate of the statistical information of the model parameters of interest, along with a Bregman divergence to regularize the amplitude inversion by measuring the distance to the a priori model. However, different rock facies and fluids saturations result in different elastic and petrophysical properties, and subsequently lead to multimodal behavior of the marginal and joint distributions of the model parameters of interest. Examples of such problems are petrophysical inversion of seismic data (Connolly & Hughes, 2016; de Figueiredo et al., 2017, 2018a; Kolbjørnsen et al., 2020), and anisotropy AVO inversion (Lu et al., 2018; Zhou et al., 2020), where more parameters must be estimated. To this end, a prior mixture model is necessary to infer the correct statistical information about the parameters of interest. Additionally, the mixture models provide good fit to heavy-tailed data (Mukerji et al., 2009; Grana & Bronston, 2015). Grana & Della Rossa (2010) proposed a Gaussian mixture model for the litho-fluid classes. Grana et al. (2017) obtained an analytical solution of the posterior distribution using a prior Gaussian mixture model. Fjeldstad & Grana (2018) used Markov chain Monte Carlo (McMC) to sample from a Gaussian mixture posterior density for predicting lithology-fluid classes. de Figueiredo et al. (2019) jointly inverted for facies and elastic properties, using a Gaussian mixture prior density. Consequently, a single-component density function is incompetent to represent the model space, and subsequently infer the wrong statistical information about the model parameters.

To quote Goodfellow et al. (2016), "The basic intuition behind information theory is that learning that an unlikely event has occurred is more informative than learning that a likely event has occurred.", we can calculate the amount of information conveyed by the inverted parameters using the relative entropy (Kullback & Leibler, 1951). Here, we propose a regularized inverse problem that holds for mixture probability density functions. The regularization is based on $f$-divergences, which measure the difference between two probability distributions, i.e., the a priori distribution and predicted distribution. Firstly, an inverse rock physics (IRPM) problem is solved using well log and seismic data for estimating the P-wave anisotropy $\varepsilon$ and normal-moveout (NMO) anisotropy $\delta$ parameters, in order to fully cover the low-to-medium frequency gap of the limited-bandwidth seismic data (Zidan et al., 2021). Then, a priori joint distribution is obtained using all five parameters $\left[V_p,\ V_s,\ \rho,\ \delta,\ \varepsilon\right]$. The mixture model contains the statistical information about the model parameters of different litho-facies. Next, the anisotropy AVO modeling based on Rüger's P-P reflection coefficients in a VTI medium is performed. A regularized inverse problem that measures the residual between the amplitudes of observed and synthetic seismic angle gathers, and the divergence of the predicted probability density from the a priori probability density is constructed. Subsequently, the alternating direction of multipliers method (ADMM) is used to minimize the functional and regularization alternately. The performance of the proposed approach is subsequently tested on synthetic and real seismic angel gathers. The regularized optimization successfully constrains the low-sensitivity density and $\delta$ parameters, and a better recovery of the elastic velocities and $\varepsilon$.

# Methodology

## 1.1 Anisotropic AVO modeling

Reflection amplitudes are parameterized by the angles of incidence and elastic properties of rock layers across the interface. Assuming incident plane waves at an interface between two VTI media, the energy partitioning is described by Daley & Hron (1977). Various weak-contrast approximations have been introduced, including Ursin & Haugen (1996); Vavryčuk (1999); Shaw & Sen (2004). Thomsen et al. (1993) derived a linear approximation of the PP reflection coefficients for weak anisotropy VTI media:

$$\begin{aligned} R_{PP}^{VTI}(\theta) &= \frac{1}{2}\left[\frac{\Delta Z_0}{\bar{Z}_0}\right] \\ &+ \frac{1}{2}\left[\frac{\Delta V_{p0}}{\bar{V}_{p0}} - \left(\frac{2\bar{V}_{s0}}{\bar{V}_{p0}}\right)\frac{\Delta G_0}{\bar{G}_0} + (\delta_2 - \delta_1)\right]\sin^2\theta \\ &+ \frac{1}{2}\left[\frac{\Delta V_{p0}}{\bar{V}_{p0}} - (\delta_2 - \delta_1 - \varepsilon_2 + \varepsilon_1)\right]\tan^2\theta\sin^2\theta, \end{aligned} \tag{1}$$

where $Z_0 = \rho V_{p0}$ and $G_0 = \rho V_{s0}^2$ are the vertical P-wave impedance and vertical S-wave modulus (Castagna & Backus, 1993). The $\Delta\delta$ appears on both $\sin^2\theta$ and $\tan^2\theta\sin^2\theta$ terms, resulting in inaccurate reflection coefficients for large angles of incidence ($> 20°$). Moreover, for angles larger than $45°$, the $\tan^2\theta\sin^2\theta$ dominates the AVO-gradient term, and equation 1 breaks down (Rüger, 2002). Rüger (1997) decomposed the reflection coefficients into isotropic and anisotropic terms:

$$R_{PP}^{VTI}(\theta) = R_{PP}^{iso}(\theta) + R_{PP}^{ani}(\theta), \tag{2}$$

where,

$$\begin{aligned} R_{PP}^{iso}(\theta) &= \frac{1}{2}\left[\frac{\Delta Z_0}{\bar{Z}_0}\right] \\ &+ \frac{1}{2}\left[\frac{\Delta V_{p0}}{\bar{V}_{p0}} - \left(\frac{2\bar{V}_{s0}}{\bar{V}_{p0}}\right)\frac{\Delta G_0}{\bar{G}_0}\right]\sin^2\theta \\ &+ \frac{1}{2}\left[\frac{\Delta V_{p0}}{\bar{V}_{p0}}\right]\tan^2\theta\sin^2\theta, \end{aligned} \tag{3}$$

$$R_{PP}^{ani}(\theta) = \frac{1}{2}\Delta\delta\sin^2\theta + \frac{1}{2}\Delta\varepsilon\tan^2\theta\sin^2\theta. \tag{4}$$

In the above equation, $\Delta\delta$ controls the small-angle reflection coefficients through the $\sin^2\theta$ term, whereas $\Delta\varepsilon$ controls the large-angle reflection coefficients with the $\tan^2\theta\sin^2\theta$ term. Consequently, equation 2 is more stable for large angles of incidence. Figure 1a shows comparison of the exact isotropic and exact anisotropic solutions in a VTI medium, the difference between the two solutions particularly for the mid-to-far angle of incidence may lead to the wrong lithology and fluid content. Equation 2 is extended to a time-continuous reflectivity function reads per Stolt & Weglein (1985):

$$R_{PP}^{VTI}(\theta) = a_{V_p}(\theta)\frac{\partial \ln V_p}{\partial t} + a_{V_s}(\theta)\frac{\partial \ln V_s}{\partial t} + a_\rho(\theta)\frac{\partial \ln \rho}{\partial t} + a_\delta(\theta)\frac{\partial \delta}{\partial t} + a_\varepsilon(\theta)\frac{\partial \varepsilon}{\partial t}, \tag{5}$$

where,

$$a_{V_p}(\theta) = \frac{1}{2}(1 + \sin^2\theta + \sin^2\theta\tan^2\theta), \tag{6}$$

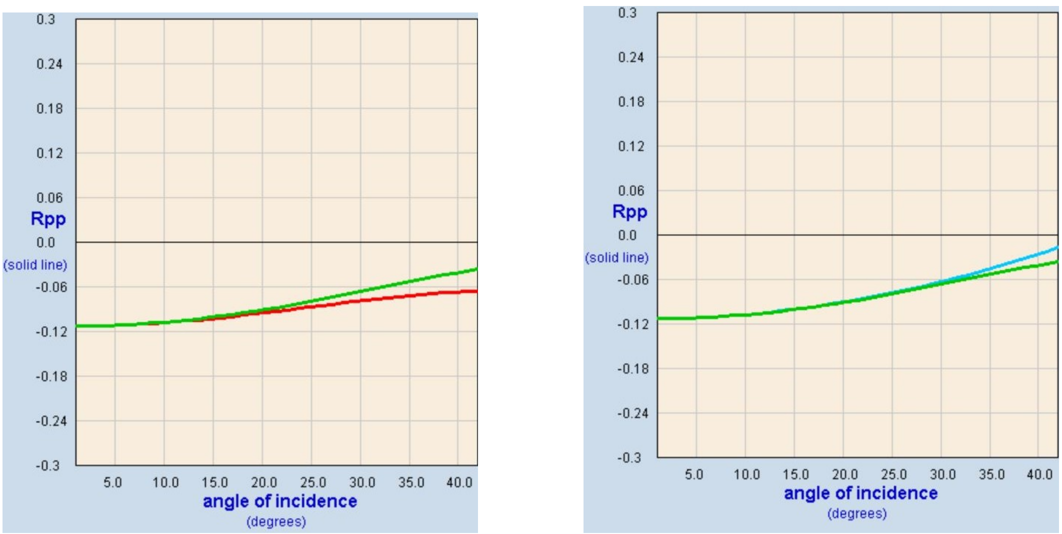
$$a_{V_s}(\theta) = -\left(\frac{2\bar{V}_s(t)}{\bar{V}_p(t)}\right)^2 \sin^2\theta, \tag{7}$$

$$a_\rho(\theta) = \frac{1}{2}\left(1 - \left(\frac{2\bar{V}_s(t)}{\bar{V}_p(t)}\right)^2 \sin^2\theta\right), \tag{8}$$

$$a_\delta(\theta) = \frac{1}{2}\sin^2\theta, \tag{9}$$

$$a_\varepsilon(\theta) = \frac{1}{2}\sin^2\theta\tan^2\theta. \tag{10}$$

In the above equation, accurate estimation of the background $\bar{V}_s/\bar{V}_p$ is important, particularly for relatively strong elastic contrasts (Alemie, 2010); in practice, it is usually taken as a constant calculated from well logs. Figure 1b shows comparison between the Daley & Hron (1977) and the Rüger (1997) solutions in a VTI medium. In spite of the good fit between the two solutions, equation 5 is highly ill-posed and rarely inverted directly.




(a) (b)

Figure 1: (a) Comparison between the exact isotropic (red) and exact anisotropic (green) solutions in a VTI medium. (b) Comparison between the exact anisotropic (green) and Rüger anisotropic (cyan) solutions in a VTI medium. We have used the CREWES VTI Explorer 1.1 to perform the above simulation.

### 1.2 Alternating direction method of multipliers (ADMM)

In constrained optimization, the goal is to solve:

$$\hat{x} = \underset{x}{\operatorname{argmin}}\, f(x) + g(x), \tag{11}$$

where $f(x)$ and $g(x)$ are convex functions. The two functions can be separate using variable splitting:

$$\begin{aligned} \hat{x} = \quad & \underset{x,z}{\operatorname{argmin}}\, f(x) + g(z), \\ \text{s.t.} \quad & x = z. \end{aligned} \tag{12}$$

The consensus constraint $(x = z)$ can be presented as a penalty term using the augmented Lagrangian (Parikh & Boyd, 2014):

$$L_\rho(x, z, y) = f(x) + g(z) + y^T (x - z) + \frac{\rho}{2} \|x - z\|_2^2, \tag{13}$$

where $y$ is the Lagrange multiplier vector, and $\rho > 0$. Equation 13 can be solved iteratively using the scaled dual variable $u^k = \left({}^1\!/\!_\rho\right) y^k$ at $k^{th}$ iteration and $\lambda = \left({}^1\!/\!_\rho\right)$ (Boyd et al., 2011; Parikh & Boyd, 2014):

$$x^{k+1} = \mathbf{prox}_{\lambda f}\left(z^k - u^k\right), \tag{14}$$

$$z^{k+1} = \mathbf{prox}_{\lambda g}\left(x^{k+1} + u^k\right), \tag{15}$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1}. \tag{16}$$

The above equation is the proximal form of the ADMM, where $\mathbf{prox}_{\lambda f}$, $\mathbf{prox}_{\lambda g}$ are the proximity operators for $f$ and $g$ with penalty $\lambda$, respectively. The proximity operator of a function, e.g., $g$, is defined in equation 17, where we try to minimize $g(x)$, yet stay close to a given value $y$.

$$\mathbf{prox}_{\lambda g}(y) = \underset{x}{\operatorname{argmin}}\, g(x) + \frac{1}{2\lambda} \|x - y\|_2^2. \tag{17}$$

For $f$ and $g$ convex, and $k^{th}$ iteration, the ADMM has $\mathcal{O}({}^1\!/\!_k)$ rate of convergence (He & Yuan, 2012; Hong et al., 2016). Many variants of the ADMM have been developed, such as linearized (Goldfarb et al., 2013), online (H. Wang & Banerjee, 2013), stochastic (Ouyang et al., 2013; Huang et al., 2019), and accelerated (J. Zhang et al., 2019).

The function $f(x)$ can be set as the mean-square error (MSE) of the amplitude differences between synthetic and observed seismic data:

$$MSE\left(\hat{y_\theta},\ y_\theta\right) = \frac{1}{n} \sum_i \left(\hat{y}_{i,\theta} - y_{i,\theta}\right)^2, \tag{18}$$

where $\hat{y_\theta},\ y_\theta$ are the synthetic and observed amplitudes at angle of incidence $\theta$, respectively. The $L_2$ loss function is strictly convex, with a positive-definite Hessian. The $L_2$ loss function is more sensitive to noise, as it is related to the short-tailed Gaussian density (Guitton & Symes, 2003; Tarantola, 2005). In general, seismic data contains different types of noises, and therefore noise-insensitive loss functions are favorable, such as the $L_1$ and Huber misfit functions. However, the mean-squares and Huber errors are not normally distributed, and therefore can lead to erroneous results (Kosheleva & Kreinovich, 2017) and, hence, a regularization $g(z)$ may help in stabilizing the inversion.

### 1.3 $f$-divergences

$f$-divergences (a.k.a. Csiszár $f$-divergences) are contrast functions that measure the dissimilarity between two probability distributions (Csiszár, 2008). Let $p$ a prior probability density defined from the available well logs and rock physics model, and $q$ is the predicted probability distribution from seismic amplitude inversion, then for a convex function $f$ with $f(1) = 0$, the $f$-divergence of $p$ from $q$ is:

$$D_f(p, q) = \mathbb{E}_q \left[ f(u(x)) \right], \tag{19}$$

where $u$ is the density ratio $u(x) = {p(x)}/{q(x)}$. The divergence measure depends on the choice of the function $f$, i.e., the KL divergence is a special case of the $f$-divergences for $f(x) = x \log x$ (Polyanskiy & Wu, 2014). Examples of probability metrics $f$ include (Liese & Vajda, 2006):

$$\text{(Forward KL-div):} \quad D_{KL}(p\|q) = u(x) \log u(x) - (u(x) - 1), \tag{20}$$

$$\text{(Reverse KL-div):} \quad D_{KL}(q\|p) = -\log u(x) + (u(x) - 1), \tag{21}$$

$$\text{(Jensen-Shannon):} \quad D_{JS}(p\|q) = \frac{1}{2}\left[ (u(x)+1) \log\left(\frac{2}{u(x)+1}\right) + u(x) \log u(x) \right]. \tag{22}$$

The $f$-divergences are non-negative, and convex:

$$D_f \left[ \lambda p_1 + (1-\lambda) p_2 \| \lambda q_1 + (1-\lambda) q_2 \right] \le \lambda D_f \left( p_1 \| q_1 \right) + (1-\lambda) D_f \left( p_2 \| q_2 \right). \tag{23}$$

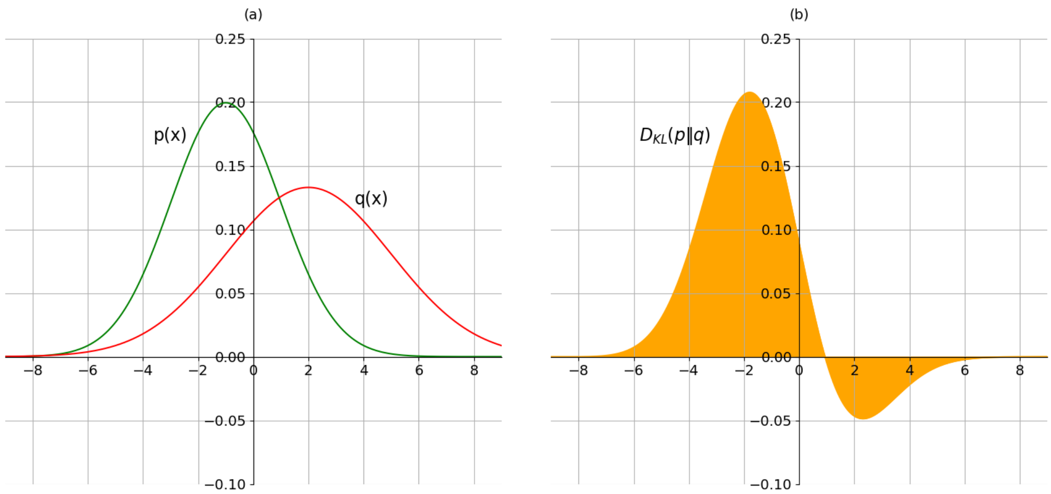



Figure 2: (a) Illustration of the difference in information between the true distribution $p(x)$ and observed distribution $q(x)$. (b) The Kullback–Leibler metric is calculated as the area under the curve $D_{KL}(p\|q) = \int p(x) \ln \frac{p(x)}{q(x)}$.

$f$-divergences appear in many machine learning applications and related fields (Nowozin et al., 2016; Ke et al., 2020; Ghasemipour et al., 2020; Yu et al., 2020; Gimenez & Zou, 2019). Figure 2 shows how to estimate the similarity between two Gaussian probability distributions via information. The Kullback–Leibler metric (area under the curve) increases much faster with increasing the difference in the mean value as compared to the difference in the variance value between the two distributions.

To solve equation 19, $i.i.d.$ (independent and identically distributed) samples are generated $\{x_i\}_{i=1}^{M}$ from $q$, and then the Monte Carlo estimation is used as following (Mnih & Rezende, 2016):

$$D_f(p, q) = \frac{1}{M} \sum_{i=1}^{M} f(u(x_q^{(i)})), \quad x_q^{(i)} \sim q(x). \tag{24}$$

In figure 3a, despite staring with a normal Gaussian distribution $q^0(x)$, we iteratively fit a mixture model of same number of components as $p(x)$; nonetheless, we can always choose to fit a simpler model as shown in figure 3b. In equation 19, $q(x)$ is calculated using the inverted model parameters $\{x\}$, and thereby would always have a significant mass that leads to zero-forcing of $q$, i.e., $p(x) = 0 \Rightarrow q(x) = 0$, hence, minimization of the reverse KL divergence is prompted as it steers clear of regions where $q(x)$ is high and $p(x)$ is small (a.k.a. mode-seeking). On the other hand, the forward KL divergence approximates $p$ distribution across all its modes, by seeking the mean (a.k.a. mean-seeking). For our proposal, due to variable-sensitivities among model parameters, noises in the observed data, and bad starting model, the inverted parameters are incorrect leading to a variational distribution $q^0(x)$ (orange) that is far away from the a priori distribution (red) and, hence, minimizing the divergence between the two distributions can mitigate ill-posedness of the inverse problem.

Intuitively, we use $f$-divergence to measure how far is the predicted parameters from the a priori probability density function, and consequently rejects local minima that have large divergence values. Figure 4 illustrates how a solution with larger divergence value from the a priori density is identified as a local minimum solution and, hence, inversion is proceeded until reaching a solution with a small divergence value, and best generate the observed seismic data. In a proximal splitting framework, both the likelihood of data and regularization are minimized, alternately. For $f$-divergence, minimization is done with respect to the model parameters $\{x\}$ that used to approximate the variational distribution $q$, and without needing to sample from $q$. The seismic amplitude inversion is an underdetermined, hence, a direct method might not be accurate for the $x$-minimization step. Moreover, there is no closed form solution of the $f$-divergence proximity operator. Assuming $f$ and $g$ are smooth functions, we use the L-BFGS-B iterative solver to carry out the two primal variables' minimization, the $x$-minimization and $z$-minimization (Boyd et al., 2011). The L-BFGS-B algorithm is useful to handle bound constraints on the variables, where the Hessian is updated at each iteration (Byrd et al., 1995; C. Zhu et al., 1997).

### 1.4 The a priori model

Because seismic anisotropy can not be measured directly in vertical wells, a rock physics model is built to predict the elastic anisotropy parameters $\delta$ and $\varepsilon$. A constrained inverse rock physics problem is constructed, in which the Hudson-Cheng crack model is combined with the anisotropy AVO convolution model to predict the Thomsen's anisotropy parameters (Zidan et al., 2021).

A prior probability density function of the elastic and seismic anisotropy parameters is then defined, using a Gaussian mixture model and a multivariate Gaussian model. The a priori distribution provides the statistical information that helps in stabilizing the amplitude inversion. A Gaussian mixture model is defined as a weighted sum of various Gaussian probability density functions (Reynolds, 2009):

$$p(x|\lambda) = \sum_{i=1}^{M} w_i \; g(x|\mu_i, \Sigma_i), \tag{25}$$

with,

$$\lambda = [w_i, \; \mu_i, \; \Sigma_i], \tag{26}$$

$$\sum_{i=1}^{M} w_i = 1, \tag{27}$$

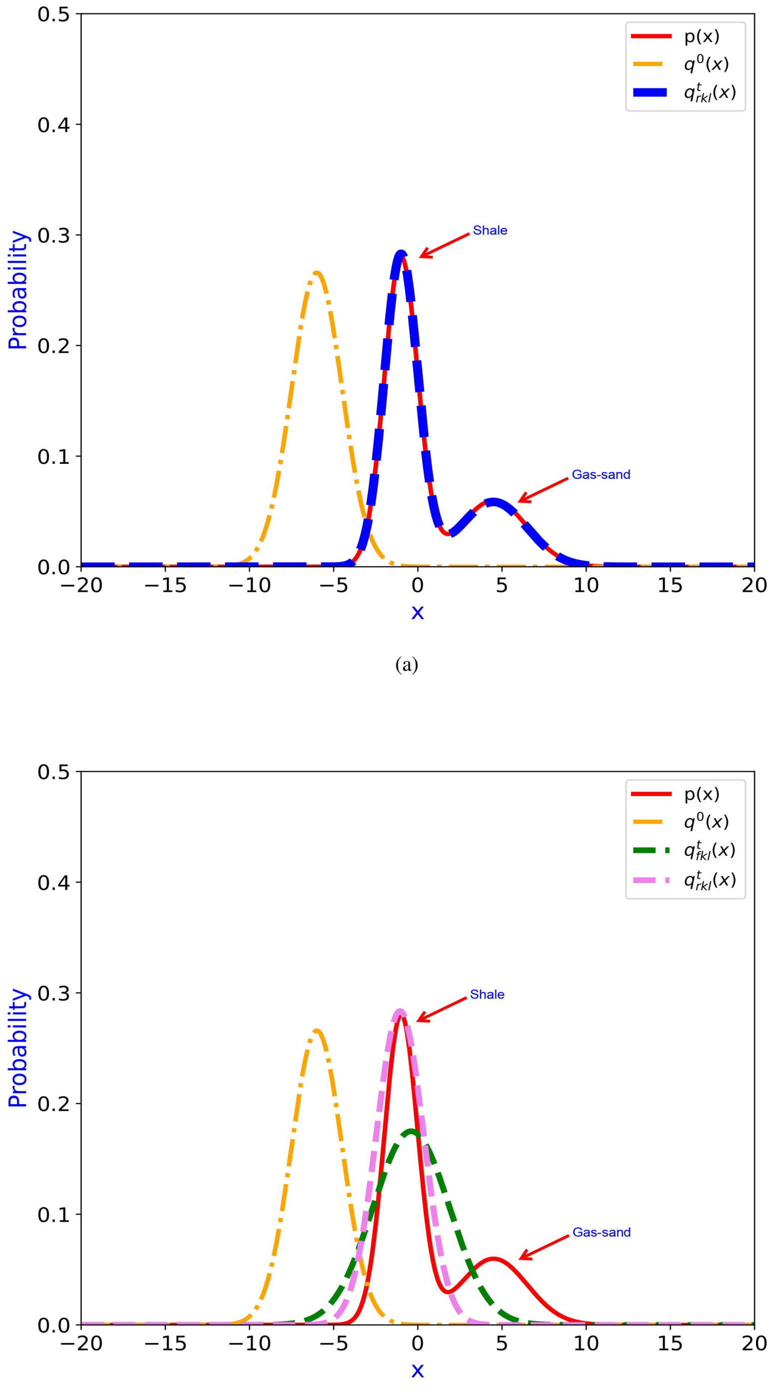



Figure 3: Minimization of *f*-divergence between (a) two mixture models using the reverse KL divergence, and (b) a single and mixture models using the forward and reverse KL divergences.

where $x$ is an N-dimensional data vector; $g(x|\mu_i, \sigma_i)$ are the Gaussian densities components, each defined by a mean vector $\mu_i$, and a covariance matrix $\Sigma_i$. $w_i$ is a weight assigned to each component:

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{\left[-\frac{1}{2}(x-\mu_i)' \Sigma_i^{-1} (x-\mu_i)\right]\right\}. \tag{28}$$

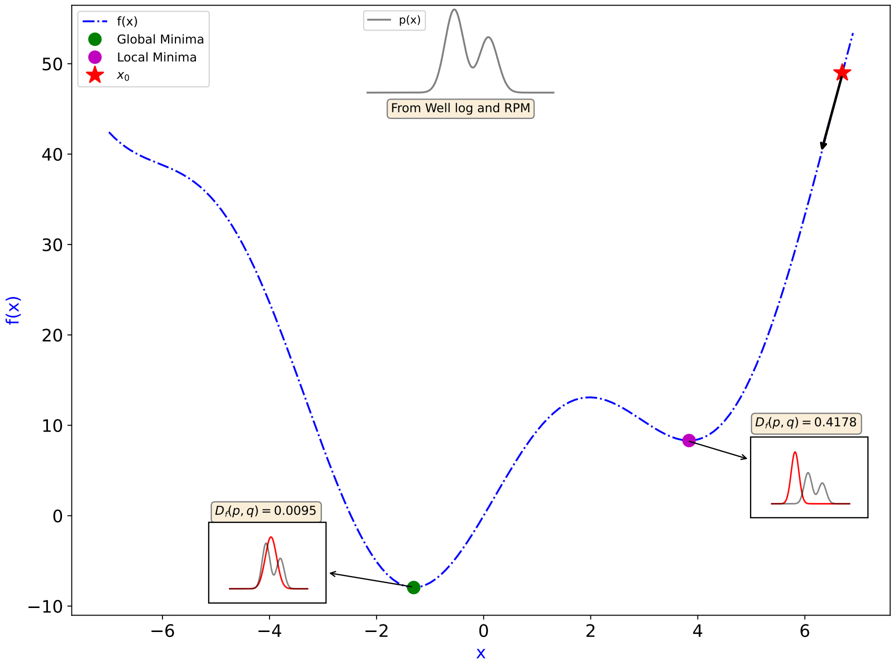



Figure 4: Illustration of the proposed constraint optimization, which aims to skip local minima solutions with large $f$-divergence values from the a priori model. The probability density in (red) obatined from the amplitude inversion, is compared with the probability density obtained from well logs and rock physics data (gray).

Here, the Gaussian mixture model is emphasized as a density estimator that best describe the input parameters and fully represents the model space. To achieve the best density estimation of the GMM, a *full*-type covariance matrix is used. In addition, the optimal number of components is estimated by adjusting the model likelihood, via minimizing the Bayesian information criterion (BIC) and the Akaike information criterion (AIC) (VanderPlas, 2016). Smooth Gaussian components, each with a mean vector and a covariance matrix, are fitted using Expectation–Maximization approach (E-M) (Reynolds, 2009). The E-M algorithm is an iterative approach, where the mixture parameters $\lambda$ are updated to increase the likelihood of the model (VanderPlas, 2016):

$$\hat{w}_i = \frac{1}{T} \sum_{t=1}^{T} Pr\left(i|x_t, \lambda\right), \tag{29}$$

$$\hat{\mu}_i = \frac{\sum_{t=1}^{T} Pr\left(i|x_t, \lambda\right)\ x_t}{\sum_{t=1}^{T} Pr\left(i|x_t, \lambda\right)}, \tag{30}$$

$$\hat{\Sigma}_i = \frac{\sum_{t=1}^{T} Pr\left(i|x_t, \lambda\right)\ x_t^2}{\sum_{t=1}^{T} Pr\left(i|x_t, \lambda\right)} - \hat{\mu}_i^2, \tag{31}$$

with,

$$Pr\left(i|x_t,\lambda\right) = \frac{w_i\ g(x_t|\mu_i,\Sigma_i)}{\sum_{k=1}^{M} w_k\ g(x_t|\mu_k,\Sigma_k)}, \tag{32}$$

where $\lambda = [w_i,\ \mu_i,\ \Sigma_i]$ are the initial mixture parameters. $\hat{w}_i,\ \hat{\mu_i},\ \hat{\Sigma}_i$ are the updated mixture parameters at $i^{th}$ iteration. $Pr\left(i|x_t,\lambda\right)$ is the posterior probability for $i^{th}$ component. Figure 5 shows the Bayesian information criterion and Akaike information criterion as a function of the number of the GMM components. The number of Gaussian components that enhances density estimation of the GMM is around $25-30$ for the AIC, whereas BIC suggest a simpler model of $9$ components; the simple model of $9$ components is selected. The inversion results would therefore depend on the choice of the a priori density function, i.e., number of the mixture model components. The Gaussian mixture and multivariate Gaussian models are shown in figures 6a and 6b, respectively. All the 2D projections of the negative log-likelihood of the a priori distributions and the data points are plotted. The GMM preferably addresses the multimodal behavior of the model parameters. Nevertheless, there would be a covariance matrix for each component, and subsequently the Mahalanobis distance is not applicable (Zidan, 2022). Furthermore, the analytical solution of the posterior distribution for a GMM is not tractable and, hence, requires a stochastic sampling algorithm to explore the mixture prior density function, e.g., Markov chain Monte Carlo. However, the stochastic sampling algorithms are computationally expensive, and converges slowly in high-dimensional models. Moreover, the AVO models depend on the boundary properties rather than layer properties, which necessitates a longer McMC chain. Alternatively, we propose a deterministic approach to regularize the seismic amplitude inversion using a divergence measure from the a priori probability density function.

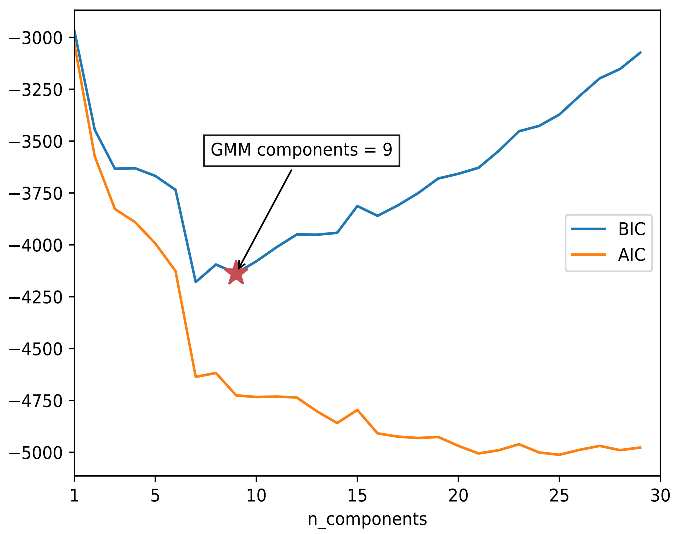
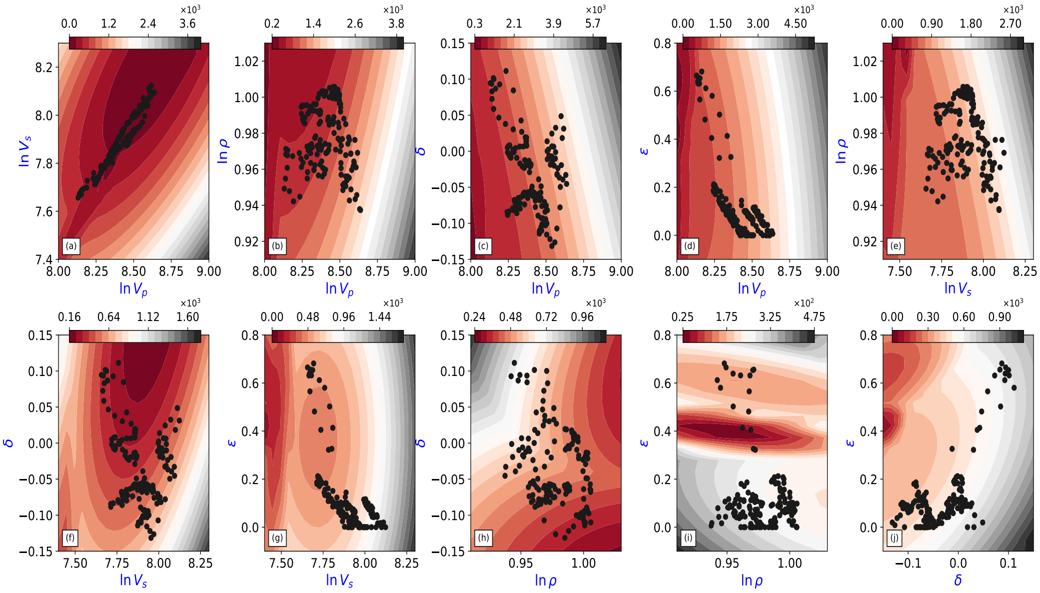


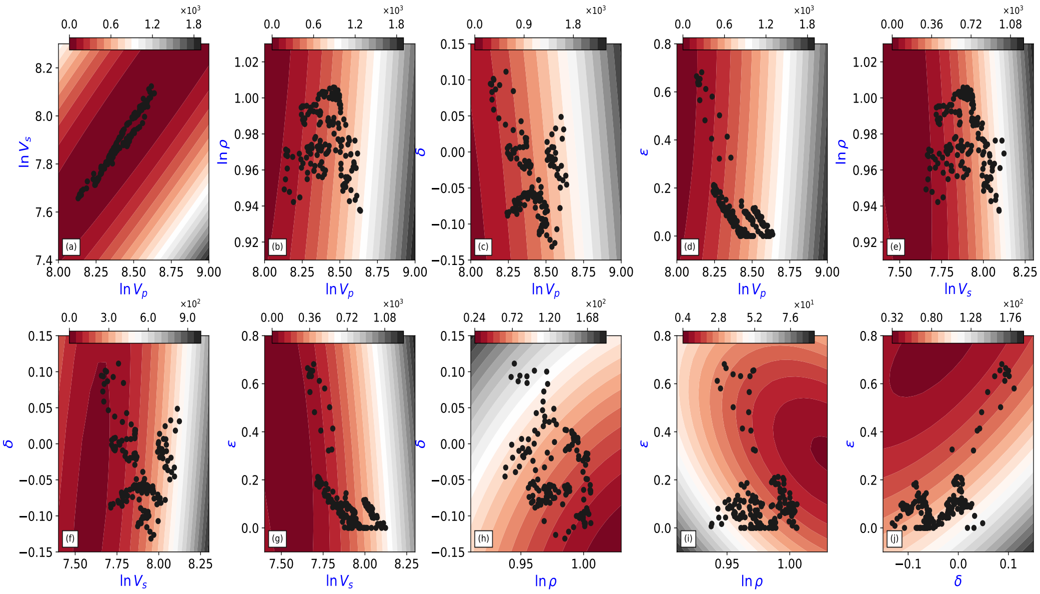

Figure 5: The optimal number of Gaussian components based on the AIC and BIC criterion.

## Results

Figure 7 shows cross-plot of the AVO intercept-and-gradient attributes, calculated from the pre-stack seismic angle gathers using the two term Aki-Richard's equation. The AVO response of the organic-rich shale is of class IV (negative intercept and positive gradient), where the absolute values of the intercept and gradient increase with the kerogen content and porosity (Y. Li et al.,




(a)




(b)

Figure 6: The negative log-likelihood of the a priori probability distribution using a (a) Gaussian mixture model, and (b) multivariate Gaussian model. The GMM components is set to 9. Black dots are the input data samples from well logs and rock physics modelling.

2015). As shown, it is hard to distinct class IV AVO response of the organic-rich shales from the surrounding rocks. Despite conventional AVO attributes provide information about interfaces (top and base of the shale reservoir), they cannot infer the effective moduli of the shale reservoirs. Consequently, it is necessary to estimate supplementary parameters that are sensitive to the organic matters and correlated with the gas content of the sweetest intervals. The feasibility of the proposed regularization is demonstrated using synthetic seismic angle gathers. The model parameters consist of the P- and S-wave velocity, density, and Thomsen's parameters $\delta$ and $\varepsilon$. The synthetic angle gathers are obtained by convolving the P-P reflection coefficients calculated using equation 5 with a statistical wavelet. Figure 8 shows the synthetic angle gather consisting of $41$ traces of range $1 - 41°$.

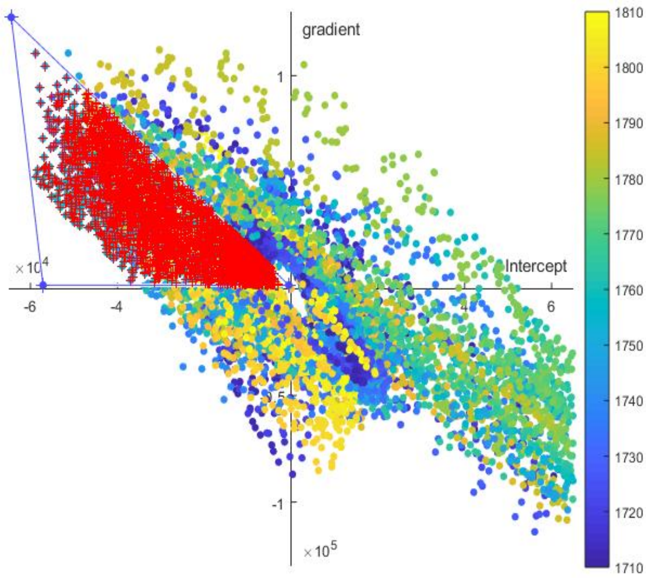



Figure 7: Cross-plot of the AVO intercept and gradient attributes calculated from pre-stack seismic angle gathers, using the two term Aki-Richard's equation. The red dots represent the class IV AVO responses of the organic-rich shales.

The information contained in the a priori mixture density and the difference in information of the a priori distribution from the variational (predicted) distribution is measured via $f$-divergence. Subsequently, the amplitude inversion can be regularized by measuring how much information lost when substituting the a priori probability density $p$ with the variational density $q$. In an augmented Lagrangian scheme, i.e., ADMM, the data residual between synthetic and real angle gathers is measured, next regularization is imposed by measuring the information lost in the model space, and then the functional and regularization are coordinated globally via constraints, i.e., the dual variable for the ADMM.

To assess the proposed approach, results are compared with the unconstrained minimization of the L-BFGS-B algorithm. We use the same low-frequency model and boundaries as for the proposed approach. The initial model is necessary to fully cover the low-to-medium frequency gap of the seismic data. The proposed method is first tested on synthetic data with high signal-to-noise ratio of $S/N = \infty$. For this test, the mean-square error misfit and the Kullback-Leibler divergence regularization are used. Figure 9 shows the inversion results of the unconstrained L-BFGS-B and constrained ADMM methods. Despite the high signal-to-noise ratio, the unconstrained L-BFGS-B algorithm converges to a local minimum and couldn't update all the five parameters properly due to the variable-sensitivities among the model parameters and missing information in the observed data (limited aperture and S-wave), particularly the density and Thomsen's $\delta$ as shown in figure 9a. Using ADMM with the KL divergence regularization of the GMM a priori model, the model parameters $V_p,\ V_s$ and $\varepsilon$ have been recovered successfully, and $\rho$ and $\delta$ are preferably constrained, as shown in figure 9b. The constrained minimization is next run using the multivariate Gaussian a priori model, as shown in figure 9c. Similarly, all five parameters have been recovered fairly well, yet the a priori density of the multivariate Gaussian could not properly model particular litho-facies,

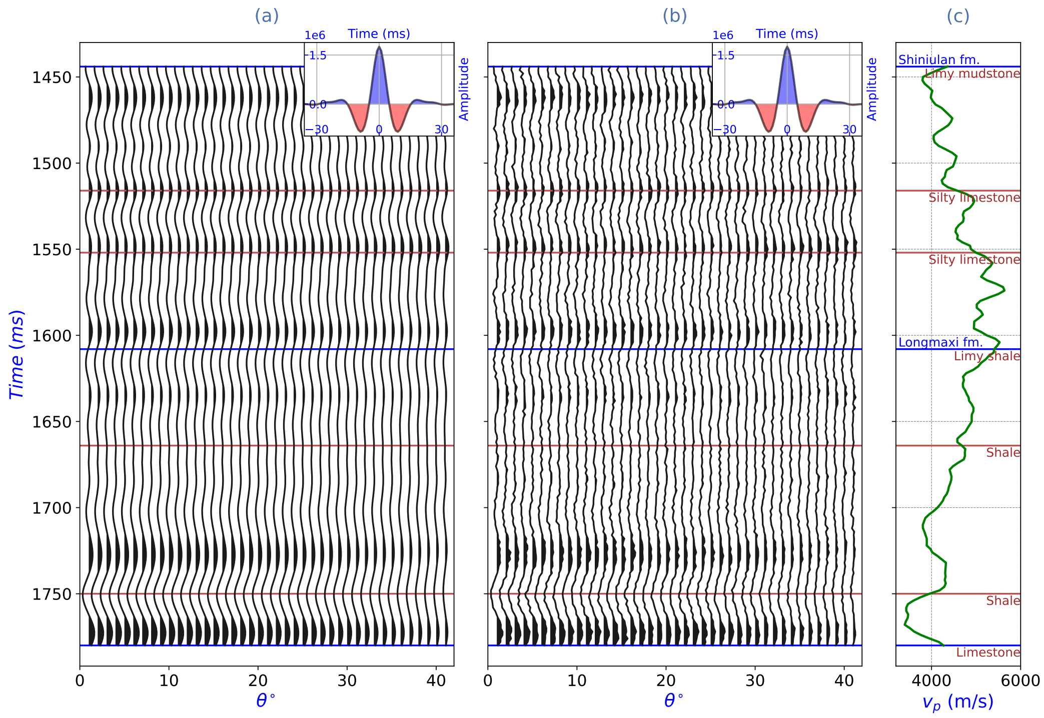



Figure 8: Synthetic pre-stack angle gathers with (a) $S/N = \infty$, (b) $S/N = 9$, along with (c) the P-wave velocity from well logs, and the statistical wavelet.

e.g., the shale gas zone as compared to the GMM. Notably, the log interval is consisting of alternating shale and mudstone, and thereby there are no significant variations in the lithofacies.

Additionally, a simpler (single-component) variational distributions of a multivariate Gaussian and a multivariate student's t- are used to model the inverted parameters, and to approximate the a priori Gaussian mixture model. Figures 10a and 10b show the inversion results of the multivariate Gaussian and multivariate student's t-distributions. Despite using simpler models of the same (Gaussian) and different (student's t-) density functions to fit the 9-components GMM, the $f$-divergence minimization successfully regularizes the parameters' updates corresponding to the a priori Gaussian mixture model.

Next, the proposed regularization is tested against noise by adding a random noise to the pre-stack angle gather corresponding to signal-to-noise ratio of $S/N = 9$ as shown in figure 8b. The Huber misfit and KL divergence are used. The unconstrained optimization fails to obtain the correct updates of model parameters, and the Huber misfit function fails to handle the low signal-to-noise ratio of the seismic data, as shown in figure 11a. In the constrained optimization, all five model parameters are fairly recovered, by constraining the model updates using the statistical properties of the unknown parameters. When the information lost from the a priori probability distribution is large, the dual variable is updated and proceed to next iteration until a solution of lower divergence and residual values is reached, as shown in figure 11b.

Then, the regularization is tested with a poor starting model, represented by a constant mean value for each parameter. Because the initial model is far from the true solution, the unconstrained optimization fails to converge to the true solution, as illustrated in figure 12a. On the contrary, minimization step of the KL divergence imposes regularization that prompts solutions close to the a priori information, hence, fairly recover all the five model parameters as shown in figure 12b. Additionally, a single-component multivariate Gaussian and student's t-distributions are used as variational distributions to fit the a priori model of 9-components GMM. Figure 13 shows that the regularization functions, likewise, properly constrain the model parameters.

Finally, the proposed constrained optimization is applied to pre-stack seismic data acquired at the Zhaotong national shale gas demonstration area, Sichuan basin. The marine shales of the Silurian Longmaxi formation occur along synclinal belt, the sweetest intervals have an average thickness of $35.5m$ and rich in organic matter ($TOC \approx 3.5\%$). The estimated anisotropy parameters are therefore used to accurately map the occurrence of the sweetest intervals across the seismic array. The maximum angle of incidence at the reservoir interval is estimated as $42°$, and all the $(1–42°)$ are used in the inversion process. The seismic convolution model of the Rüger equation, and the statistical wavelets estimated from the seismic data are used to synthesize angle gathers, which subsequently compared with the real angle gathers at each CDP location. The low-frequency models for the elastic and Thomsen's anisotropy parameters are built based on the well log data and the structure and stratigraphy interpretation. A priori Gaussian mixture model is built, and a single-component Gaussian density function is used as the variational distribution. The sweetest shale intervals are well-identified with a lower P- and S-wave velocity, and density, and higher $\delta$ and $\varepsilon$ along the synclinal structure at about $1750ms$ as shown in figures 14 and 15. The near-angle seismic traces have less information about density, while most information is in the far-angle traces, hence, the density result is less stable as it is more sensitive to the noise level in the data, particularly misalignment of reflectors at far angles of incidence.

## Discussion

To alleviate ill-posedness of the amplitude inversion, the objective function is split into a loss, and regularization that addresses the statistical properties conveyed by a priori density function of the unknown parameters. The regularization is based on measuring the distance between two probability distributions. A priori mixture model, and a variational distribution of the same number of components or simpler are used as regularization. The proposed approach works as following, first the data residual between the observed and synthetic data is minimized, next the information loss in the model space is minimized, and then both the functional and regularization are coordinated globally via the dual variable. Despite the proposed regularization provides better constrains on the density and Thomsen's parameter $\delta$, P-S data and larger angles of incidence are required to fully constrain the inversion results. The proposed approach is very useful in solving geophysical inverse problems that involve different moduli and elastic attributes, such as the Young's modulus, Poisson's ratio, incompressibility, rigidity, Lame' parameters, and density (Goodway et al., 1997; Xu & Bancroft, 1998; Gray et al., 1999; Golalzadeh et al., 2008; Zong et al., 2013; Yin et al., 2015); moreover, the joint elastic and petrophysical models, such as the P- and S-wave velocity, density, effective porosity, clay volume, and water saturation (Bosch, 2004; Z. Li et al., 2016; M. Liu & Grana, 2018; de Figueiredo et al., 2018b; Guo et al., 2021; K. Li et al., 2021). The latter AVO models necessitate robust a priori information to better addresses the joint distributions of the different model parameters.

The performance of the proposed regularization is demonstrated using a high signal-to-noise seismic, low signal-to-noise seismic, and bad starting model. In the high signal-to-noise seismic, the proposed regularization successfully eases the ill-posedness due to variable-sensitivities among model parameters. With bad staring model, the regularization fairly mitigate the problem of frequency gap between the seismic data and prior model, by updating the dual variable of the functional and regularization. The regularization is also capable of mitigating the low signal-to-noise observed seismic data. Furthermore, the proposed regularization is suited for many probability distributions that can best describe the statistics of the well logs data, such as parametric distributions (e.g., log-normal distribution), and non-parametric distributions (e.g., kernel density estimation). However, the resulting solution from ADMM is not exact because of the partial updates for the dual variable. Using a Bregman divergence as a regularization function (Zidan, 2022), a stationary statistical correlation information matrix is used, which can be extended in time using a temporal correlation function (Buland & Omre, 2003), and in space using a spatial correlation model (de Figueiredo et al., 2018b). Nonetheless, $f$-divergence is considered non-stationary and can be used for each interface, without a temporal nor a spatial correlation functions (for sufficient well-control), hence, can be used for the entire well log interval and across the seismic array. Furthermore, the ADMM provides an adaptable framework to incorporate additional regularization functions, such as noise-reduction and sparsity-promoting functions. The ADMM can be extended for parallel implementation, to solve $N$ sub-problems in parallel at each iteration (Boyd et al., 2011; Deng et al., 2017).

Nevertheless, the divergence measure heavily depends on the choice of the probability metric; moreover, it is required to approximate a variational distribution from the inverted model parameters. Furthermore, optimization is done over the same data points (inverted model) that are used to approximate the variational distribution. This deterministic update might lead to biased estimates of the descent directions on the $f$-divergence. To enhance the estimates of the functional value and gradients, the Auxiliary $f$-divergence or the Fenchel-conjugate $f$-divergence can be used, which yield the upper and lower bounds on the divergence measure, respectively (M. Zhang et al., 2019). Furthermore, the tail-adaptive $f$-divergence can be used to achieve mass-covering of the target distribution (D. Wang et al., 2018). Other divergence measures such as the Alpha-, Beta- and Gamma-divergences can also be adapted for better convergence (Cichocki & Amari, 2010). Having said that, it is necessary to estimate the descent direction without the need to approximate a variational distribution. Zidan (2022) used the Stein's method to estimate the optimal descent direction that maximally decreases the KL divergence without needing to approximate a variational distribution. The estimated descent direction depends only on the a priori probability density through a score function.

## Conclusion

We proposed a constrained optimization scheme that can be used with a priori mixture models. We then applied the proposed regularization to anisotropy amplitude inversion in a VTI medium. We first combined the well logs and rock physics data to build a joint mixture prior probability distribution that conveys the statistical properties of the parameters of interest. We then set up a regularized inverse problem using the alternating direction of multipliers method (ADMM), in which the functional and regularization are solved separately. The goal is to obtain a single stable solution that minimizes the data residual, yet stay close to the a priori mixture probability distribution. Such constraints are necessary when the starting model is far away from the true solution. In comparison with the unconstrained optimization, the vertical velocities are better recovered, and the density and anisotropy parameters are well-constrained.

## 2 ACKNOWLEDGMENTS

The authors acknowledge BGP for the well log data and seismic data. The work is supported by the Singapore Economic Development Board Petroleum Engineering Professorship program. A. Zidan is also supported by SINGA scholarship. The data that support the finding of this study are available online at `https://doi.org/10.5281/zenodo.4304555`.

## References

Alemie, W. (2010). *Regularization of the avo inverse problem by means of a multivariate cauchy probability distribution* (MSc dissertation). University of Alberta.

Bandyopadhyay, K. (2009). *Seismic anisotropy: Geological causes and its implications to reservoir geophysics* (PhD dissertation). Stanford University.

Banik, N. (1987). An effective anisotropy parameter in transversely isotropic media. *Geophysics*, *52*(12), 1654–1664.

Berezin, V. (1982). Probability distribution of porosity in sandstones. *International Geology Review*, *24*(10), 1197–1200.

Bosch, M. (2004). The optimization approach to lithological tomography: Combining seismic data and petrophysics for porosity prediction. *Geophysics*, *69*(5), 1272–1282.

Bosch, M., Mukerji, T., & Gonzalez, E. F. (2010). Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. *Geophysics*, *75*(5), 75A165–75A176.

Boyd, S., Parikh, N., & Chu, E. (2011). *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc.

Buland, A., & Omre, H. (2003). Bayesian linearized avo inversion. *Geophysics*, *68*(1), 185–198.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, *16*(5), 1190–1208.

Castagna, J. P., & Backus, M. M. (1993). *Offset-dependent reflectivity—theory and practice of avo analysis*. Society of Exploration Geophysicists.

Cichocki, A., & Amari, S.-i. (2010). Families of alpha-beta-and gamma-divergences: Flexible and robust measures of similarities. *Entropy*, *12*(6), 1532–1568.

Connolly, P. A., & Hughes, M. J. (2016). Stochastic inversion by matching to large numbers of pseudo-wells. *Geophysics*, *81*(2), M7–M22.

Csiszár, I. (2008). Axiomatic characterizations of information measures. *Entropy*, *10*(3), 261–273.

Daley, P. F., & Hron, F. (1977). Reflection and transmission coefficients for transversely isotropic media. *Bulletin of the seismological society of America*, *67*(3), 661–675.

de Figueiredo, L. P., Grana, D., Bordignon, F. L., Santos, M., Roisenberg, M., & Rodrigues, B. B. (2018a). Joint bayesian inversion based on rock-physics prior modeling for the estimation of spatially correlated reservoir properties. *Geophysics*, *83*(5), M49–M61.

de Figueiredo, L. P., Grana, D., Bordignon, F. L., Santos, M., Roisenberg, M., & Rodrigues, B. B. (2018b). Joint bayesian inversion based on rock-physics prior modeling for the estimation of spatially correlated reservoir properties. *Geophysics*, *83*(5), M49–M61.

de Figueiredo, L. P., Grana, D., Roisenberg, M., & Rodrigues, B. B. (2019). Gaussian mixture markov chain monte carlo method for linear seismic inversion. *Geophysics*, *84*(3), R463–R476.

de Figueiredo, L. P., Grana, D., Santos, M., Figueiredo, W., Roisenberg, M., & Neto, G. S. (2017). Bayesian seismic inversion based on rock-physics prior modeling for the joint estimation of acoustic impedance, porosity and lithofacies. *Journal of Computational Physics*, *336*, 128–142.

Deng, W., Lai, M.-J., Peng, Z., & Yin, W. (2017). Parallel multi-block admm with o (1/k) convergence. *Journal of Scientific Computing*, *71*(2), 712–736.

Downton, J. E. (2005). *Seismic parameter estimation from avo inversion*. University of

Calgary, Department of Geology and Geophysics.

Downton, J. E., & Lines, L. R. (2001). Constrained three parameter avo inversion and uncertainty analysis. In *Seg technical program expanded abstracts 2001* (pp. 251–254). Society of Exploration Geophysicists.

Fjeldstad, T., & Grana, D. (2018). Joint probabilistic petrophysics-seismic inversion based on gaussian mixture and markov chain prior models probabilistic petroelastic prediction. *Geophysics*, *83*(1), R31–R42.

Ghasemipour, S. K. S., Zemel, R., & Gu, S. (2020). A divergence minimization perspective on imitation learning methods. In *Conference on robot learning* (pp. 1259–1277).

Gimenez, J. R., & Zou, J. (2019). *A unified f-divergence framework generalizing vae and gan* (Tech. Rep.). Stanford University.

Golalzadeh, A., Javaherian, A., & Nabi-Bidhendi, M. (2008). Estimation of lame's parameters from p-waves in a vti medium. *Journal of Geophysics and Engineering*, *5*(1), 37–45.

Goldfarb, D., Ma, S., & Scheinberg, K. (2013). Fast alternating linearization methods for minimizing the sum of two convex functions. *Mathematical Programming*, *141*(1), 349–382.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Goodway, B., Chen, T., & Downton, J. (1997). Improved avo fluid detection and lithology discrimination using lamé petrophysical parameters;"$\lambda\rho$","$\mu\rho$", & "$\lambda/\mu$ fluid stack", from p and s inversions. In *Seg technical program expanded abstracts 1997* (pp. 183–186). Society of Exploration Geophysicists.

Grana, D. (2020). Bayesian petroelastic inversion with multiple prior models. *Geophysics*, *85*(5), M57–M71.

Grana, D., & Bronston, M. (2015). Probabilistic formulation of avo modeling and avo-attribute-based facies classification using well logs. *Geophysics*, *80*(4), D343–D354.

Grana, D., & Della Rossa, E. (2010). Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion. *Geophysics*, *75*(3), O21–O37.

Grana, D., Fjeldstad, T., & Omre, H. (2017). Bayesian gaussian mixture linear inversion for geophysical inverse problems. *Mathematical Geosciences*, *49*(4), 493–515.

Gray, D., Goodway, B., & Chen, T. (1999). Bridging the gap: Using avo to detect changes in fundamental elastic constants. In *Seg technical program expanded abstracts 1999* (pp. 852–855). Society of Exploration Geophysicists.

Guitton, A., & Symes, W. W. (2003). Robust inversion of seismic data using the huber norm. *Geophysics*, *68*(4), 1310–1319.

Guo, Q., Ba, J., Fu, L.-Y., & Luo, C. (2021). Joint seismic and petrophysical nonlinear inversion with gaussian mixture-based adaptive regularization. *Geophysics*, *86*(6), 1–57.

Hampson, D. P., Russell, B. H., & Bankhead, B. (2005). Simultaneous inversion of pre-stack seismic data. In *Seg technical program expanded abstracts 2005* (pp. 1633–1637). Society of Exploration Geophysicists.

He, B., & Yuan, X. (2012). On the o(1/n) convergence rate of the douglas–rachford alternating direction method. *SIAM Journal on Numerical Analysis*, *50*(2), 700–709.

Hernlund, J. W., & Houser, C. (2008). On the statistical distribution of seismic velocities in earth's deep mantle. *Earth and Planetary Science Letters*, *265*(3-4), 423–437.

Hong, M., Luo, Z.-Q., & Razaviyayn, M. (2016). Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM Journal on Optimization*, *26*(1), 337–364.

Huang, F., Chen, S., & Huang, H. (2019). Faster stochastic alternating direction method of multipliers for nonconvex optimization. In *International conference on machine learning* (pp. 2839–2848).

Ke, L., Choudhury, S., Barnes, M., Sun, W., Lee, G., & Srinivasa, S. (2020). Imitation learning as f-divergence minimization. In *International workshop on the algorithmic foundations of robotics* (pp. 313–329).

Kim, K. Y., Wrolstad, K. H., & Aminzadeh, F. (1993). Effects of transverse isotropy on p-wave avo for gas sands. *Geophysics*, *58*(6), 883–888.

Kolbjørnsen, O., Buland, A., Hauge, R., Røe, P., Ndingwan, A. O., & Aker, E. (2020). Bayesian seismic inversion for stratigraphic horizon, lithology, and fluid prediction. *Geophysics*, *85*(3), R207–R221.

Kosheleva, O., & Kreinovich, V. (2017). Why deep learning methods use kl divergence instead of least squares: A possible pedagogical explanation..

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79–86.

Li, K., Yin, X., Zong, Z., & Grana, D. (2021). Estimation of porosity, fluid bulk modulus, and stiff-pore volume fraction using a multi-trace bayesian avo petrophysics inversion in multi-porosity reservoirs. *Geophysics*, *87*(1), 1–101.

Li, K., Yin, X.-Y., Zong, Z.-Y., & Lin, H.-K. (2020). Seismic avo statistical inversion incorporating poroelasticity. *Petroleum Science*, *17*(5), 1237–1258.

Li, Y., Guo, Z.-Q., Liu, C., Li, X.-Y., & Wang, G. (2015). A rock physics model for the characterization of organic-rich shale from elastic properties. *Petroleum Science*, *12*(2), 264–272.

Li, Z., Song, B., Zhang, J., & Hu, G. (2016). Joint elastic and petrophysical inversion using prestack seismic and well log data. *Exploration Geophysics*, *47*(4), 331–340.

Liese, F., & Vajda, I. (2006). On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, *52*(10), 4394–4412.

Liu, C., Song, C., Lu, Q., Liu, Y., Feng, X., & Gao, Y. (2015). Impedance inversion based on l1 norm regularization. *Journal of Applied Geophysics*, *120*, 7–13.

Liu, L., Sun, S. Z., Luo, C., Deng, X., Jing, B., Pan, Y., … Zhao, H. (2013). A noise-resistant density inversion algorithm and its application on high efficiency well selection for complex carbonate reservoir. In *Seg technical program expanded abstracts 2013* (pp. 3083–3087). Society of Exploration Geophysicists.

Liu, M., & Grana, D. (2018). Stochastic nonlinear inversion of seismic data for the estimation of petroelastic properties using the ensemble smoother and data reparameterization. *Geophysics*, *83*(3), M25–M39.

Lu, J., Wang, Y., Chen, J., & An, Y. (2018). Joint anisotropic amplitude variation with offset inversion of pp and ps seismic data. *Geophysics*, *83*(2), N31–N50.

Luo, C., Ba, J., Carcione, J. M., Huang, G., & Guo, Q. (2020). Joint pp and ps pre-stack ava inversion for vti medium based on the exact graebner equation. *Journal of Petroleum Science and Engineering*, *194*, 107416.

Mnih, A., & Rezende, D. (2016). Variational inference for monte carlo objectives. In *International conference on machine learning* (pp. 2188–2196).

Mukerji, T., Routh, P., & Ball, V. (2009). Statistical modeling of seismic reflectivities comparing lévy stable and gaussian mixture distributions. In *Seg technical program expanded abstracts 2009* (pp. 2085–2089). Society of Exploration Geophysicists.

Nadri, D., & Hartley, B. (2007). Non-linear joint avo inversion of pp and ps waves in a vti medium. *ASEG Extended Abstracts*, *2007*(1), 1–4.

Nowozin, S., Cseke, B., & Tomioka, R. (2016). f-gan: Training generative neural samplers using variational divergence minimization. In *Proceedings of the 30th international conference on neural information processing systems* (pp. 271–279).

Ouyang, H., He, N., Tran, L., & Gray, A. (2013). Stochastic alternating direction method of multipliers. In *International conference on machine learning* (pp. 80–88).

Painter, S., Beresford, G., & Paterson, L. (1995). On the distribution of seismic reflection coefficients and seismic amplitudes. *Geophysics*, *60*(4), 1187–1194.

Painter, S., & Paterson, L. (1994). Fractional lévy motion as a model for spatial variability in sedimentary rock. *Geophysical Research Letters*, *21*(25), 2857–2860.

Parikh, N., & Boyd, S. (2014). Proximal algorithms. *Foundations and Trends in* *optimization*, *1*(3), 127–239.

Plessix, R.-E., & Bork, J. (2001). Quantitative estimate of vti parameters from ava responses [link]. *Geophysical Prospecting*, *48*(1), 87–108.

Polyanskiy, Y., & Wu, Y. (2014). Lecture notes on information theory. *Lecture Notes for ECE563 (UIUC) and*, *6*(2012-2016), 7.

Reynolds, D. A. (2009). Gaussian mixture models. *Encyclopedia of biometrics*, *741*.

Rüger, A. (1997). P-wave reflection coefficients for transversely isotropic models with vertical and horizontal axis of symmetry. *Geophysics*, *62*(3), 713–722.

Rüger, A. (2002). *Reflection coefficients and azimuthal avo analysis in anisotropic media*. Society of Exploration Geophysicists.

Russell, B., & Hampson, D. (2006). The old and the new in seismic inversion. *CSEG Recorder*, *31*(10), 5–11.

Sayers, C. M. (2013a). The effect of kerogen on the avo response of organic-rich shales. *The Leading Edge*, *32*(12), 1514–1519.

Sayers, C. M. (2013b). The effect of kerogen on the elastic anisotropy of organic-rich shales. *Geophysics*, *78*(2), D65–D74.

Shaw, R. K., & Sen, M. K. (2004). Born integral, stationary phase and linearized reflection coefficients in weak anisotropic media. *Geophysical Journal International*, *158*(1), 225–238.

Stolt, R. H., & Weglein, A. B. (1985). Migration and inversion of seismic data. *Geophysics*, *50*(12), 2458–2472.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

Thomsen, L. (1986). Weak elastic anisotropy. *Geophysics*, *51*(10), 1954–1966.

Thomsen, L., Castagna, J., & Backus, M. (1993). Weak anisotropic reflections. *Offset-dependent reflectivity—Theory and practice of AVO analysis: Soc. Expl. Geophys*, 103–111.

Treitel, S., & Lines, L. (2001). Past, present, and future of geophysical inversion—a new millennium analysis. *Geophysics*, *66*(1), 21–24.

Ursin, B., & Haugen, G. U. (1996). Weak-contrast approximation of the elastic scattering matrix in anisotropic media. *Pure and Applied Geophysics*, *148*(3), 685–714.

VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. " O'Reilly Media, Inc.".

Vavryčuk, V. (1999). Weak-contrast reflection/transmission coefficients in weakly anisotropic elastic media: P-wave incidence. *Geophysical Journal International*, *138*(2), 553–562.

Velis, D. R. (2008). Stochastic sparse-spike deconvolution. *Geophysics*, *73*(1), R1–R9.

Wang, D., Liu, H., & Liu, Q. (2018). Variational inference with tail-adaptive f-divergence. *arXiv preprint arXiv:1810.11943*.

Wang, H., & Banerjee, A. (2013). Online alternating direction method (longer version). *arXiv preprint arXiv:1306.3721*.

Wang, Y. (2010). Seismic impedance inversion using l1-norm regularization and gradient descent methods. *Journal of Inverse and Ill-Posed Problems*, *18*(7), 823–838.

Wu, H., He, Y., Chen, Y., Li, S., & Peng, Z. (2019). Seismic acoustic impedance inversion using mixed second-order fractional atpv regularization. *IEEE Access*, *8*, 3442–3452.

Wu, H., Li, S., Chen, Y., & Peng, Z. (2019). Bayesian acoustic impedance inversion with gamma distribution. In *Journal of physics: Conference series* (Vol. 1237, p. 022034).

Xu, Y., & Bancroft, J. C. (1998). Avo case study: Extraction of lame's parameters from vertical component seismic data.

Yin, X.-Y., Liu, X.-J., & Zong, Z.-Y. (2015). Pre-stack basis pursuit seismic inversion for brittleness of shale. *Petroleum Science*, *12*(4), 618–627.

Yu, L., Song, Y., Song, J., & Ermon, S. (2020). Training deep energy-based models with f-divergence minimization. In *International conference on machine learning* (pp.

10957–10967).

Zand, T., Siahkoohi, H. R., Malcolm, A., Gholami, A., & Richardson, A. (2020). Consensus optimization of total variation–based reverse time migration. *COMPUTATIONAL GEOSCIENCES*.

Zhang, F., Dai, R., & Liu, H. (2014). Seismic inversion based on l1-norm misfit function and total variation regularization. *Journal of Applied Geophysics*, *109*, 111–118.

Zhang, F., Zhang, T., & Li, X.-Y. (2019). Seismic amplitude inversion for the transversely isotropic media with vertical axis of symmetry. *Geophysical Prospecting*, *67*(9), 2368–2385.

Zhang, J., Peng, Y., Ouyang, W., & Deng, B. (2019). Accelerating admm for efficient simulation and optimization. *ACM Transactions on Graphics (TOG)*, *38*(6), 1–21.

Zhang, M., Bird, T., Habib, R., Xu, T., & Barber, D. (2019). Variational f-divergence minimization. *arXiv preprint arXiv:1907.11891*.

Zhou, L., Chen, Z.-C., Li, J.-Y., Chen, X.-H., Liu, X.-Y., & Liao, J.-P. (2020). Nonlinear amplitude versus angle inversion for transversely isotropic media with vertical symmetry axis using new weak anisotropy approximation equations. *Petroleum Science*, *17*(3), 628–644.

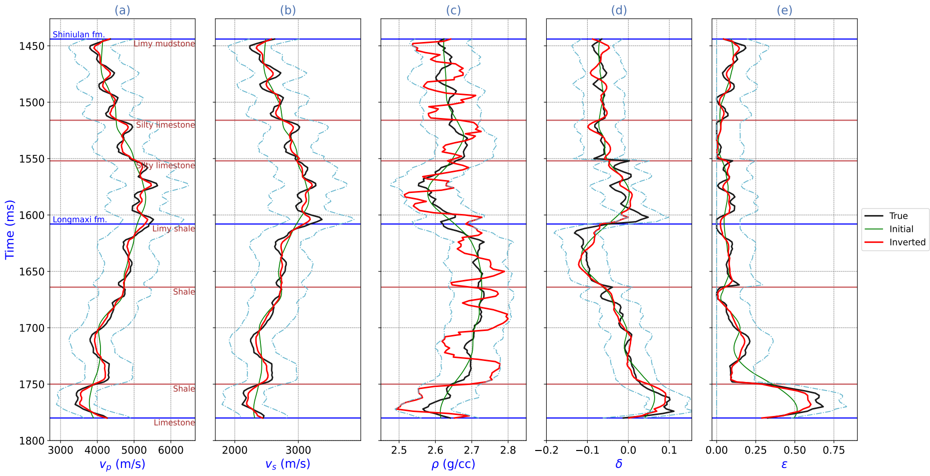
Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, *23*(4), 550–560.

Zhu, D., & Gibson, R. (2018). Seismic inversion and uncertainty quantification using transdimensional markov chain monte carlo method. *Geophysics*, *83*(4), R321–R334.

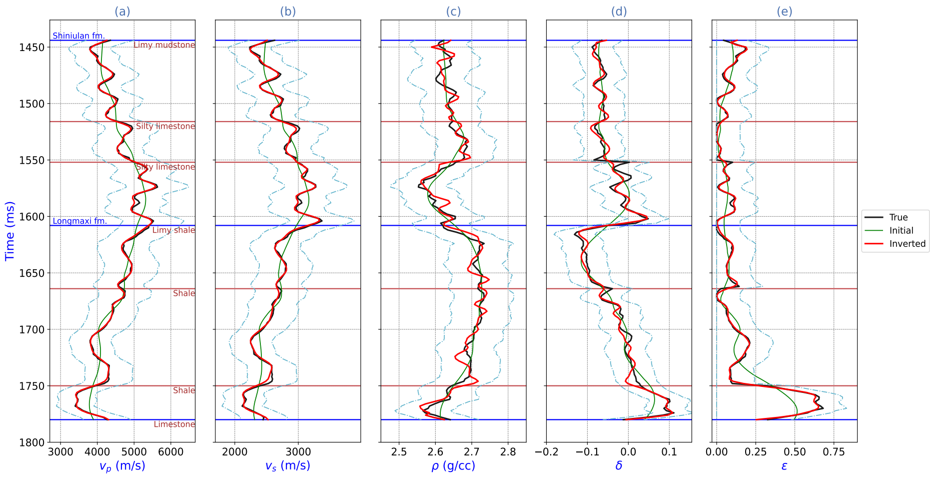
Zidan, A. (2022). *Shale-gas reservoir characterization and sweet spot prediction* (PhD dissertation). National University of Singapore.

Zidan, A., Li, Y., & Cheng, A. (2021). A pareto multi-objective optimization approach for anisotropic shale models. *Journal of Geophysical Research: Solid Earth*, e2020JB021476. doi: http://doi.org/10.1029/2020JB021476

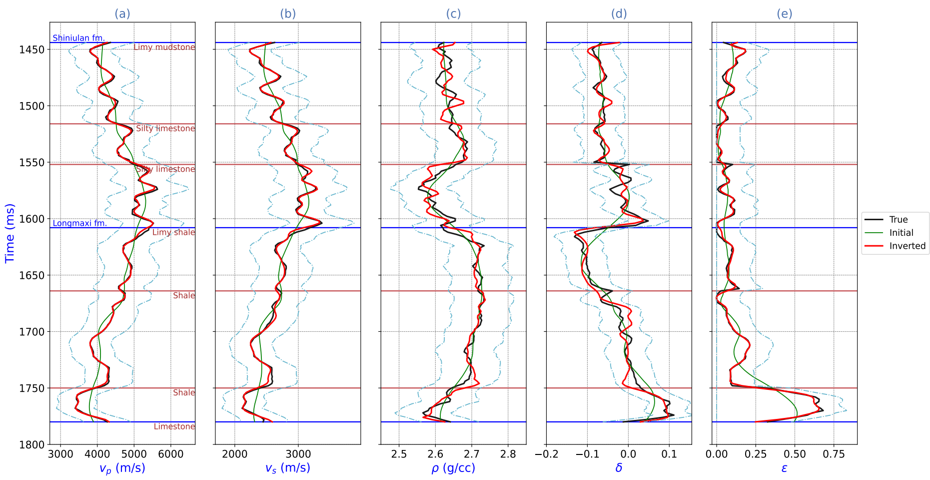
Zong, Z., Yin, X., & Wu, G. (2013). Elastic impedance parameterization and inversion with young's modulus and poisson's ratio. *Geophysics*, *78*(6), N35–N42.




(a)

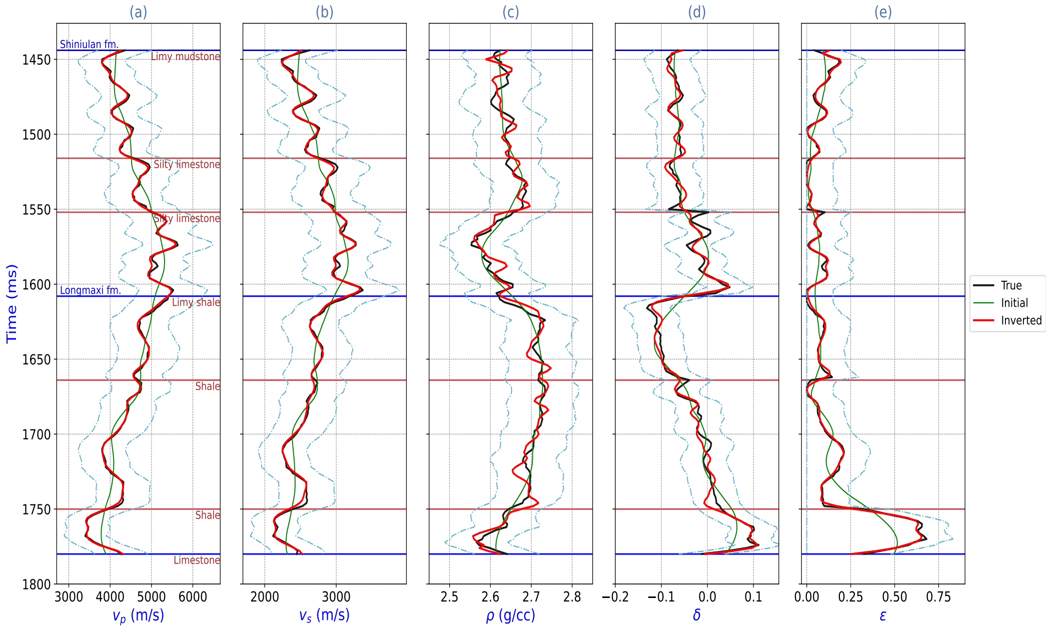



(b)

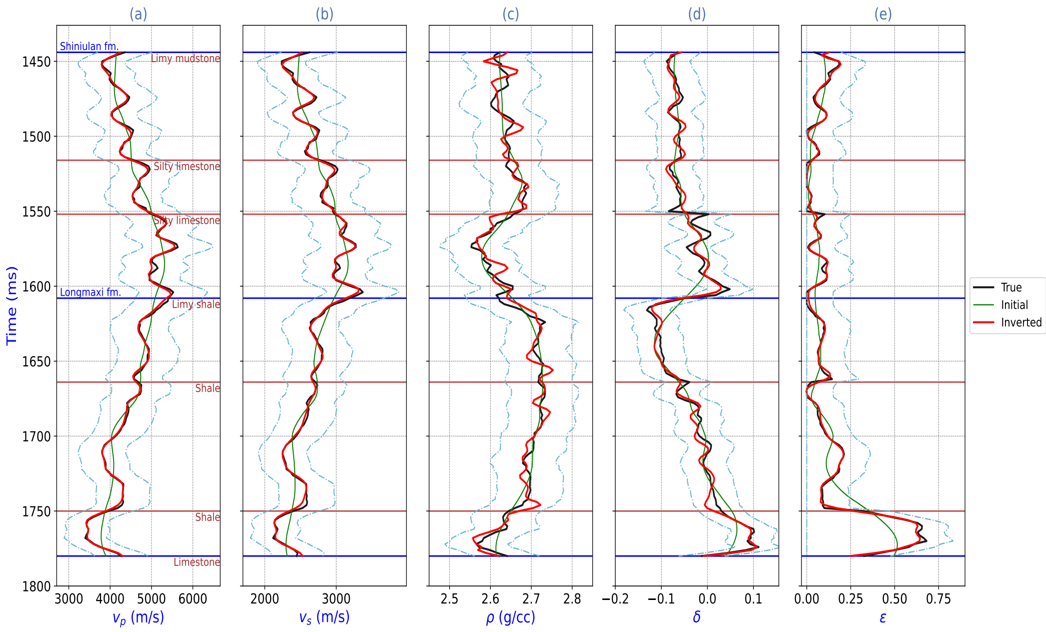



(c)

Figure 9: Inversion results of the (a) unconstrained L-BFGS-B, (b) ADMM of the 9-components-GMM, and (c) ADMM of the multivariate Gaussian. The signal-to-noise ratio is set as $S/N = \infty$. Cyan lines represent the upper and lower boundary constraints.

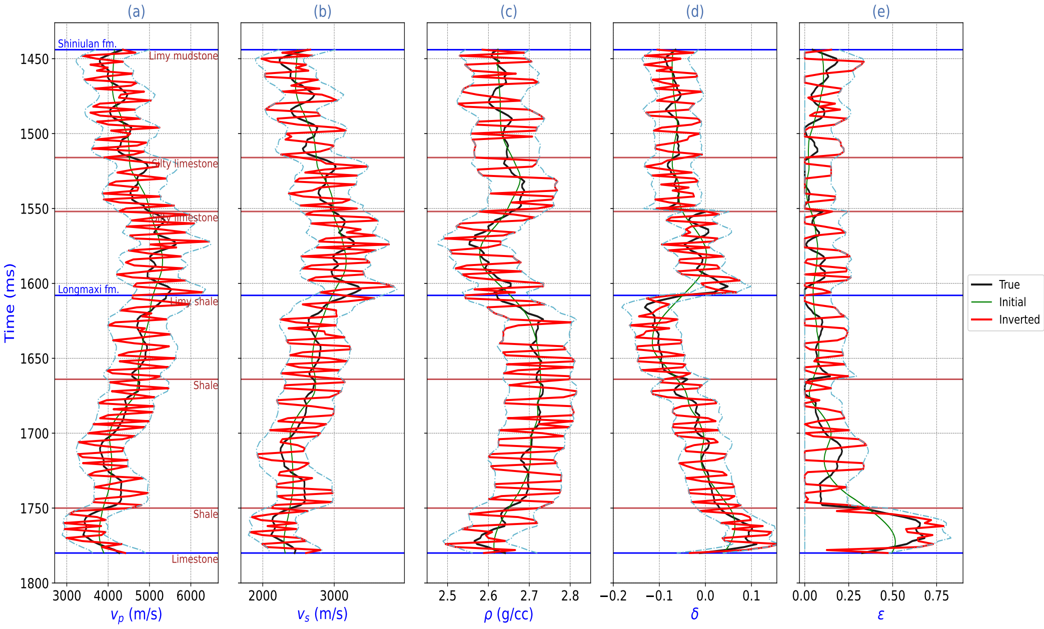



(a)

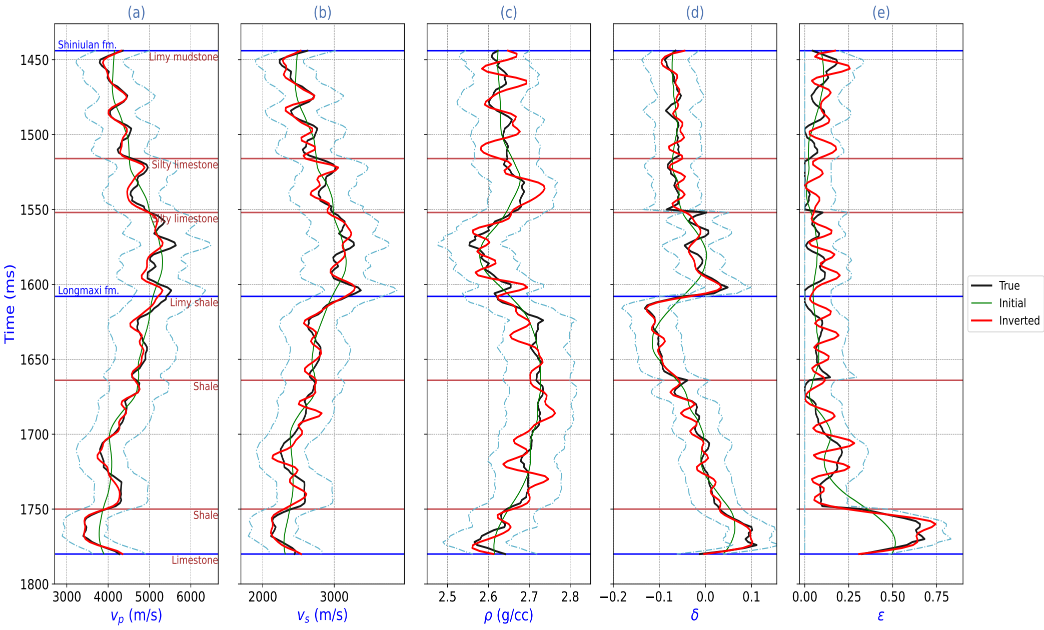



(b)

Figure 10: Inversion results of using a (a) single-component multivariate Gaussian, and (b) single-component multivariate student's t-distribution as the variational distribution to fit the 9-components Gaussian mixture a priori model. The signal-to-noise ratio is set as $S/N = \infty$. Cyan lines represent the upper and lower boundary constraints.

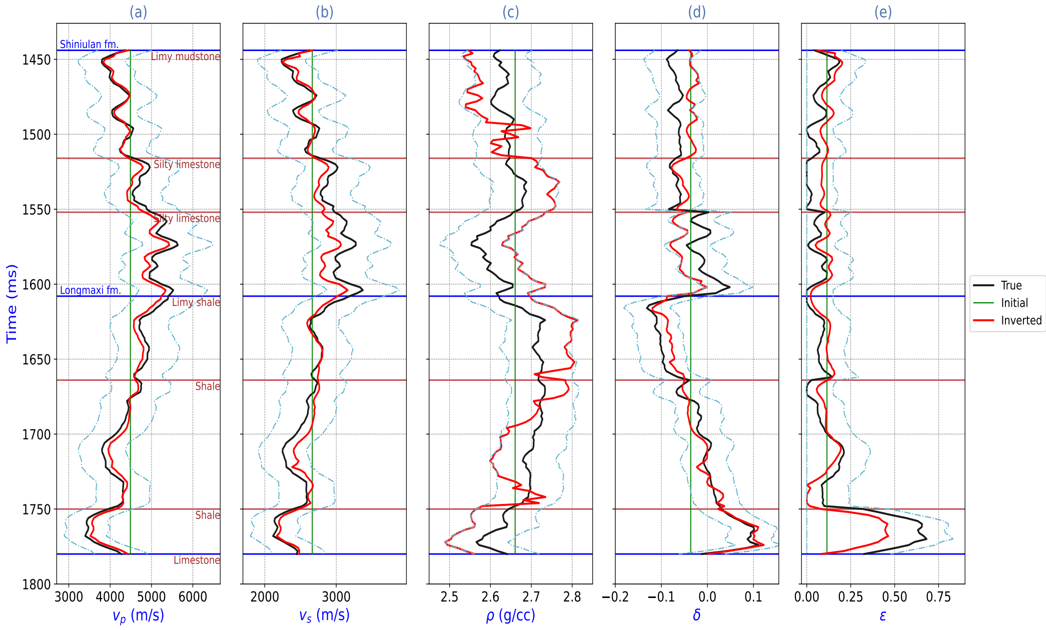



(a)

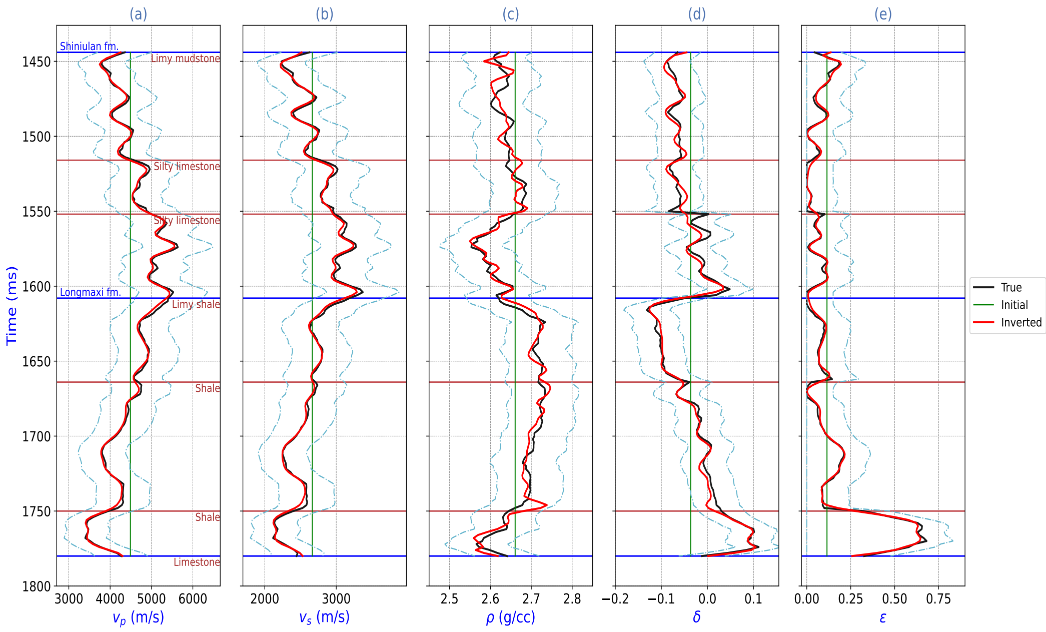



(b)

Figure 11: Inversion results of the (a) unconstrained optimization of the *L=BFGS-B*, and (b) ADMM of 9-components GMM. The signal-to-noise ratio is set as $S/N = 9$. Cyan lines represent the upper and lower boundary constraints.

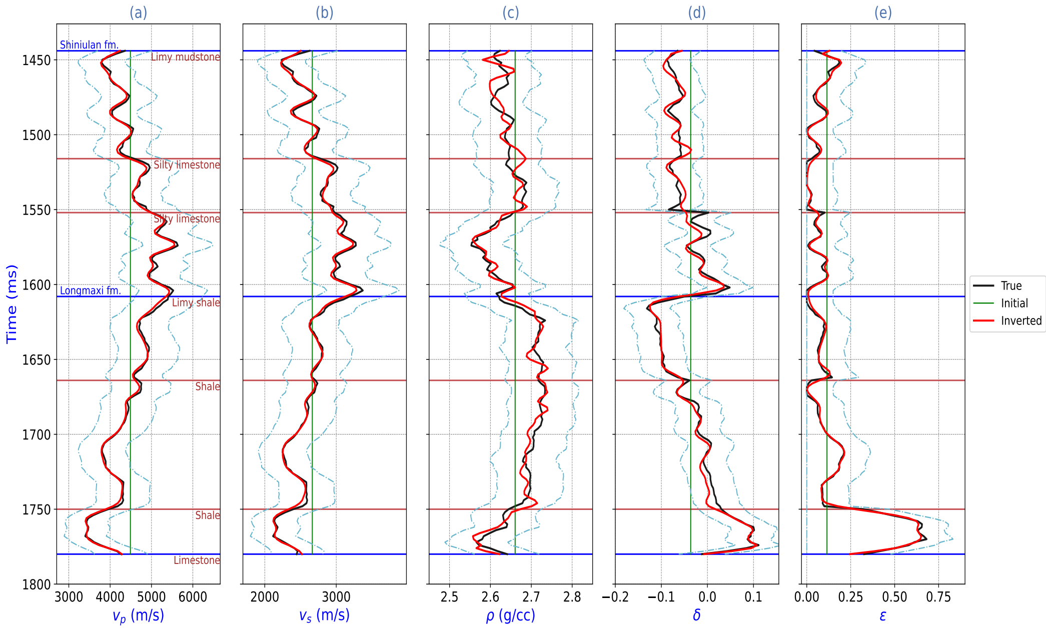



(a)

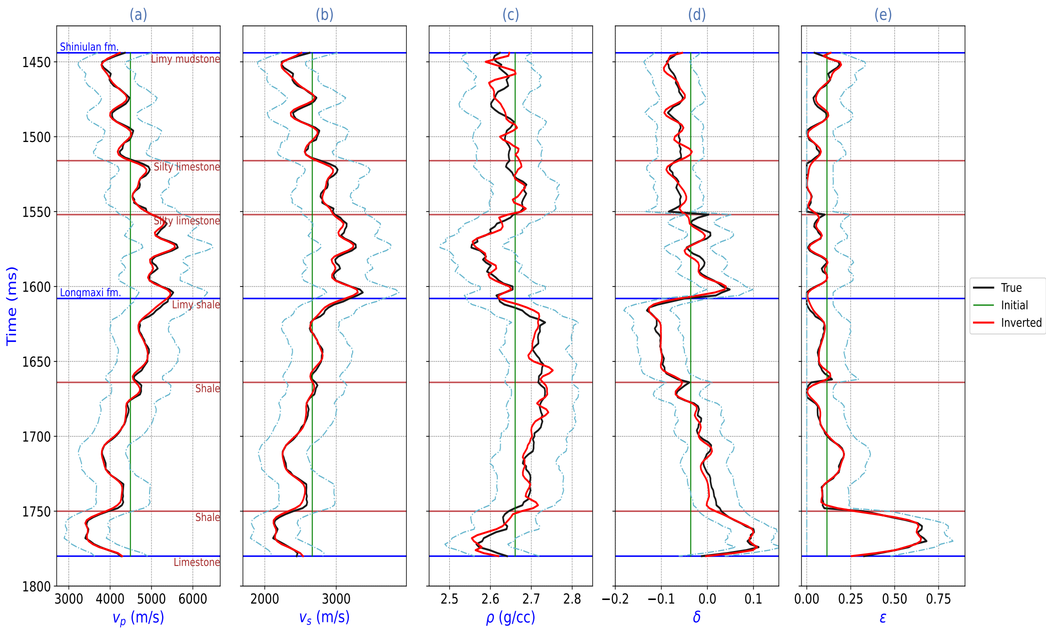



(b)

Figure 12: Inversion results of the (a) unconstrained L-BFGS-B, and (b) ADMM of 9-components GMM. The signal-to-noise ratio is set as $S/N = \infty$, and with a bad initial model (the mean value for each parameter along the time axis). Cyan lines represent the upper and lower boundary constraints.

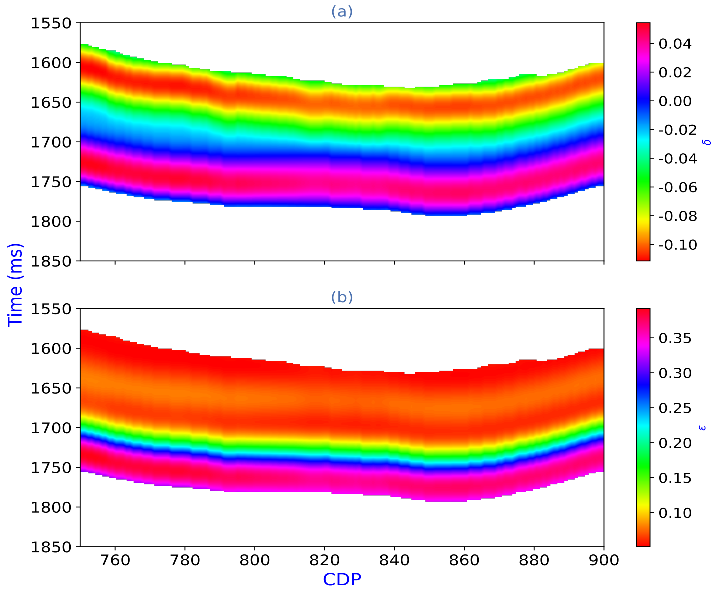



(a)

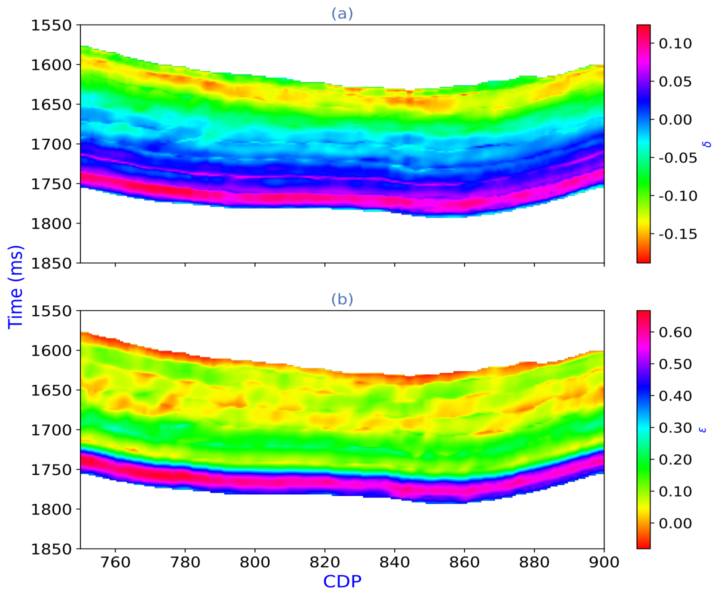



(b)

Figure 13: Inversion results of using a (a) single-component multivariate Gaussian, and (b) single-component multivariate student's t-distribution as the variational distribution to fit the 9-components Gaussian mixture a priori model. The signal-to-noise ratio is set as $S/N = \infty$, and with a bad initial model (the mean value for each parameter along the time axis). Cyan lines represent the upper and lower boundary constraints.

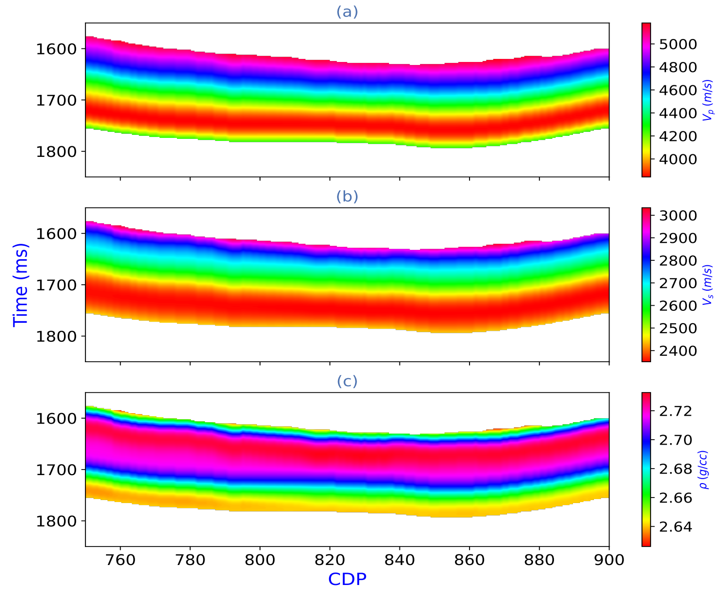



(a)

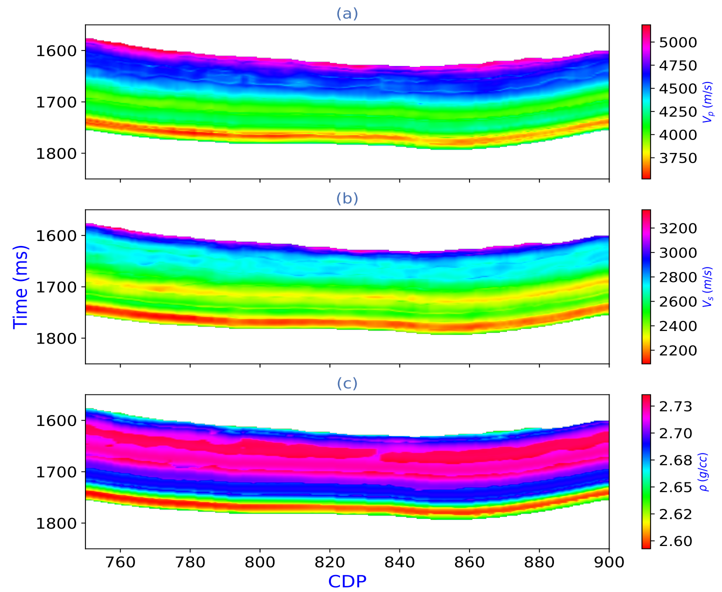



(b)

Figure 14: (a) Low-frequency anisotropy models of the (a) near-vertical anisotropy ($\delta$), and (b) P-wave anisotropy ($\varepsilon$), which used to initialize (b) the amplitude-versus-offset inversion using the proposed constrained optimization.




(a)




(b)

Figure 15: (a) Low-frequency elastic models of the (a) P-wave velocity ($V_p$), (b) S-wave velocity ($V_s$), and (c) density ($\rho$), which used to initialize (b) the amplitude-versus-offset inversion using the proposed constrained optimization.