Full-waveform inversion with Mumford-Shah regularization

Ajinkya Kadu, Rajiv Kumar†, Tristan van Leeuwen**
* *Mathematical Institute, Utrecht University*
† *Georgia Institute of Technology*

SUMMARY

Full-waveform inversion (FWI) is a non-linear procedure to estimate subsurface rock parameters from surface measurements of induced seismic waves. This procedure is ill-posed in nature and hence, requires regularization to enhance some structure depending on the prior information. Recently, Total-Variation (TV) regularization has gained popularity due to its ability to produce blocky structures. Contrary to this, the earth behaves more like a piecewise smooth function. TV regularization fails to enforce this prior information into FWI. We propose a Mumford-Shah functional to incorporate the piecewise smooth spatial structure in the FWI procedure. The resulting optimization problem is solved by a splitting method. We show the improvement in results against TV regularization on two synthetic camembert examples.

INTRODUCTION

Full waveform inversion (FWI) is a non-linear data-fitting procedure where we iteratively estimate high-resolution velocity models of the subsurface by minimizing the difference between the synthetic and recorded data. These high-resolution velocity models are used to perform reservoir characterization, time-lapse monitoring, as well as aid in identifying potential geo-hazards. However, FWI often suffers from so-called cycle-skipping (Beydoun and Tarantola, 1988), which is a common source of local-minima. Moreover, the observed seismic data often lacks low frequencies and long offets, and are contaminated by noise (see Virieux and Operto (2009) for a recent overview on FWI).

One way to overcome the non-uniqueness of FWI is to add regularization to the data-fitting terms, which results in stable solutions. Various strategies have been proposed to impose the regularization such as Tikhonov (Tikhonov, 1963; Asnaashari et al., 2013) and sparsity-promotion based regularization (Li et al., 2012). Recently, the Total Variation (TV) regularization method has been proposed, which resolves the sharp interfaces via preserving the edges and discontinuities (Rudin et al., 1992; Lin and Huang, 2014). The central idea of TV regularization is to impose sparsity on the gradient of the model parameters. Esser et al. (2018) further showed the advantages of successively relaxed asymmetric total-variation constraints to perform the automatic salt flooding.

Although TV regularization can circumvent the local-minima issue, it has a tendency to reduce the contrast at edges and over-smooth the flat regions, resulting in staircase effects in the velocity model. To further shed light on this effect, we run a small experiment on vertical trace from a complex velocity model as shown in Figure 1. We can clearly see that TV regularization approximates the smooth dipping part of the velocity model with a constant flat velocity model.

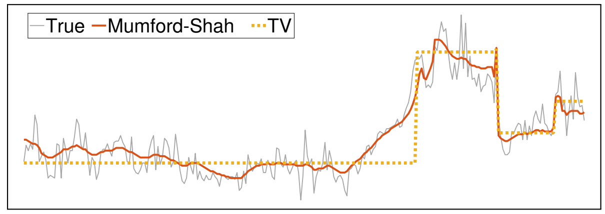

Figure 1: Comparison of the Mumford-Shah (red color) with TV method (yellow color) on 1D example.

To preserve the piecewise smooth behavior of the velocity models, we borrow the ideas from the image analysis literature and propose an FWI framework, which uses the Mumford-Shah functional as a regularizer (Mumford and Shah, 1989). The Mumford-Shah functional provides a prototypical form of all regularizers, which aims at combining the smoothing of the homogeneous region with the enhancement of edges (see Figure 1). The proposed FWI algorithm is based on splitting the problem using a technique described in Zheng and Aravkin (2018) and we employ an alternating minimization strategy to solve the problem.

The paper is organized as follows: we begin with the regularized FWI and discuss the drawbacks of the current regularization techniques. Next we introduce the Mumford-Shah segmentation procedure and a corresponding regularization term which induces a piecewise smooth model. We discuss the integration of such regularization in the FWI framework and propose an alternate minimization strategy to solve the resulting problem efficiently. Finally, we demonstrate the method on two camembert models, and compare the results with the TV.

THEORY

The regularized FWI problem in its least-squares formulation (Tarantola, 1984) reads

$$\min_{m} \quad \frac{1}{2}\|F(m) - d\|_2^2 + \mathcal{R}(m),$$

where F is a forward modeling operator, m defines the subsurface model, for instance, P-wave velocity or density or both, and d represents the seismic data acquired at number of receivers. $\|\cdot\|_2$ represents the Euclidean norm. $\mathcal{R}(m)$ is the regularization function which incorporates the prior information about the model.

The most popular regularization strategies are Tikhonov regularization and TV. Tikhonov regularization, defined as $\mathcal{R}(m) = \|\nabla m\|_2^2$, promotes smoothness in the model parameters by penalizing its spatial gradient. On the contrary, TV, defined as $\mathcal{R}(m) = \|\nabla m\|_1$, promotes jumps in the model leading to a piecewise constant image (Unser et al., 2017). The sparsity in the gradient is promoted through the ℓ_1-norm. Each of these regularizations has its own benefits, but these methods fail when the model we're interested in is piecewise smooth. Hence, we resort to methods in the image segmentation literature to reconstruct a piecewise smooth model.

Mumford-Shah functional

For the image segmentation problem, Mumford and Shah (1989) have proposed the following formulation to segment image function $f : \Omega \to \mathbb{R}$ defined on open bounded set Ω:

$$\min_{u,\Gamma} \left\{ \int_{\Omega} |u(x) - f(x)|^2 \,\mathrm{d}x + \alpha \int_{\Omega \setminus \Gamma} |\nabla u(x)|^2 \,\mathrm{d}x + \lambda \int_{\Gamma} \mathrm{d}x \right\},$$

where u is the segmented image and $\Gamma \subset \Omega$ is a set of boundaries. The first term represents the mismatch between the true image and segmented image over domain Ω. The second term penalizes the gradient of the segmented image outside the region Γ and the last term approximates the length of the boundary Γ. In summary, the Mumford-Shah functional creates a piecewise smooth image u by penalizing the smoothness over the region $\Omega \setminus \Gamma$ and length of the boundary Γ simultaneously. The parameter α controls the smoothness of the region $\Omega \setminus \Gamma$ and λ controls the length of the boundary. The smoothness increases with increasing α and similarly, the length of boundary decreases with increase in λ producing the edges in an image.

Although this formulation has gained great popularity in the image segmentation community a decade back, solving the minimization problem in two variables u and Γ remains hard. It is also important to start with a right boundary Γ for an optimization method to converge (Vese and Chan, 2002). Ambrosio and Tortorelli proposed a simpler version which approximates the Mumford-Shah functional but heavily depends on an extra parameter ϵ (Ambrosio and Tortorelli, 1990).

Relaxation of Mumford-Shah functional

To overcome these challenges of solving the Mumford-Shah functional, Strekalovskiy and Cremers (2014) proposed to relaxed the Mumford-Shah regularization function using the following formulation:

$$\min_{u} \int_{\Omega} \left[|u(x) - f(x)|^2 + \mathcal{R}_{\mathrm{MS}}\left(\nabla u\left(x\right)\right) \right] \mathrm{d}x, \qquad (1)$$

$$\text{where} \quad \mathcal{R}_{\mathrm{MS}}\left(g\right) = \min\left(\alpha |g|^2, \lambda\right).$$

This regularization function, denoted by $\mathcal{R}_{\mathrm{MS}}(g)$, is also known as *truncated* quadratic regularizer. It penalizes the gradient till a certain threshold is reached. After the threshold, the regularizer is constant and any extra changes are not penalized. This regularizer indeed separates the region Ω into two parts: a smooth part and the boundary Γ. The boundaries are defined by

$$\Gamma = \left\{ x \in \Omega \mid |\nabla u(x)| > \sqrt{\lambda/\alpha} \right\}.$$

The proposed regularization term is non-convex in nature. See Figure 2 for comparison of the proposed regularization with ℓ_1- and ℓ_2-norm regularization. Before we delve into solving the problem, we define a proximal operator of a general functional $h : \mathbb{X} \to \mathbb{R}$ as

$$\mathbf{prox}_{\tau,h}(\hat{x}) = \operatorname*{argmin}_{x \in \mathbb{X}} \left\{ \tfrac{1}{2\tau} \|x - \hat{x}\|_2^2 + h(x) \right\}$$

for parameter $\tau > 0$ and argument $\hat{x} \in \mathbb{X}$. The proximal operators for data misfit $D := \int_{\Omega} |u(x) - f(x)|^2 \,\mathrm{d}x$ and the proposed Mumford-Shah regularizer is given below:

$$\mathbf{prox}_{\tau,D}(\tilde{u}) = \frac{\tilde{u} + 2\tau f}{1 + 2\tau},$$

$$\mathbf{prox}_{\tau,\mathcal{R}_{MS}}(\tilde{g}) = \begin{cases} \frac{1}{1+2\tau\alpha}\tilde{g} & \text{if} \quad |\tilde{g}| \le \sqrt{\frac{\lambda}{\alpha}(1 + 2\tau\alpha)} \\ \tilde{g} & \text{else} \end{cases}.$$

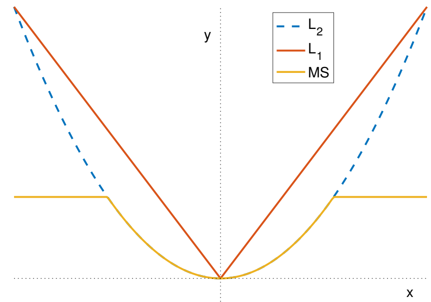

Figure 2: Comparison of Mumford-Shah regularizer (referred to as MS) with other regularization (ℓ_1 and ℓ_2).

Due to simplicity of the proximal operators, the relaxed problem (1) can be solved quickly using primal-dual algorithm (Chambolle and Pock, 2011). The algorithm is described in Algorithm 1. This algorithm has only two segmentation parameters α and λ. As mentioned earlier, the α induces the smoothness and λ controls the length of boundaries. This algorithm produces segmented image u from a given image f in less than a second with $N = 10000$ iterations.

Algorithm 1 Fast Mumford-Shah segmentation

Require: Image f, parameters α, λ, tolerance ϵ
  Initialize: $u^0 = f, \bar{u}^0 = u^0, p^0 = 0, \tau_0 = 1/4, \sigma_0 = 1/2$
Ensure: u^N
for $n = 0$ to $N - 1$ **do**
  $p^{n+1}(x) = \mathbf{prox}_{\sigma,g}\left(p^n(x) + \sigma_n \nabla \bar{u}^n(x)\right)$
  $u^{n+1}(x) = \mathbf{prox}_{\tau,f}\left(u^n(x) + \tau_n \operatorname{div} p^{n+1}(x)\right)$
  $\theta_n = 1/\sqrt{1 + 4\tau_n}$
  $\tau_{n+1} = \theta_n \tau_n, \sigma_{n+1} = \sigma_n/\theta_n$
  $\bar{u}^{n+1} = u^{n+1} + \theta_n\left(u^{n+1} - u^n\right)$
  if $\|u^{n+1} - u^n\|_2 < \epsilon$ **then**
    break
  end if
end for

ALGORITHM

In this section, we describe the integration of Mumford-Shah functional into FWI framework which results in a Mumford-Shah regularized full-waveform inversion (MS-FWI). We propose an alternating minimizing strategy to solve the problem efficiently.

MS-FWI problem

To simplify the notations, we work with the misfit function defined by f and regularization function defined by g. We discretize the model on a grid of total size n to make the problem computationally tractable. The full-waveform inversion regularized by Mumford-Shah functional takes the following form:

$$\min_{\mathbf{m}\in\mathbb{R}^n} \quad f(\mathbf{m}) + g(\mathbf{A}\mathbf{m}), \tag{2}$$

$$\text{where} \quad f(\mathbf{m}) = \tfrac{1}{2}\|F(\mathbf{m}) - \mathbf{d}\|_2^2, \qquad g(\mathbf{z}) = \mathcal{R}_{\text{MS}}(\mathbf{z}).$$

The matrix A denotes the discretization of the gradient using finite-difference. This problem can be solved using various splitting methods, including primal-dual method (Chambolle and Pock, 2011), alternating direction method of multipliers (Boyd et al., 2011), and split-bregman method (Goldstein and Osher, 2009). All these methods require solving the minimization with respect to $\mathbf{m}$ at each step completely, which becomes expensive. To avoid full solve at each step, we discuss a simple strategy to solve such problem in the next subsection.

Alternating minimization

Recently Zheng and Aravkin (2018) proposed a strategy to solve a problem of form (2). The authors split the variables and add a penalty term on the difference and then use the alternating minimization strategy. We resort to a similar version of this method and re-write equation (2) as

$$\min_{\mathbf{m}\,,\,\mathbf{z}} \quad f(\mathbf{m}) + g(A\mathbf{z}) + \rho\|\mathbf{z} - \mathbf{m}\|^2, \tag{3}$$

where $\rho > 0$ is a parameter chosen appropriately. We make use of an alternating minimization strategy to solve the above problem:

$$\mathbf{m}^{k+1} := \operatorname*{argmin}_{\mathbf{m}} \left\{ f(\mathbf{m}) + \rho\|\mathbf{m} - \mathbf{z}^k\|^2 \right\}, \tag{4}$$

$$\mathbf{z}^{k+1} := \operatorname*{argmin}_{\mathbf{z}} \left\{ g(\mathbf{A}\mathbf{z}) + \rho\|\mathbf{z} - \mathbf{m}^{k+1}\|^2 \right\}. \tag{5}$$

By splitting the problem as shown in (3), we are able to decouple the FWI minimization (4) and segmentation step (5). Moreover, the minimization in $\mathbf{m}$ can be restricted to a single update at each iteration. This property of the method improves computational time sufficiently. It is also important to note that the minimization in variable $\mathbf{z}$ is a segmentation of an image $\mathbf{m}^{k+1}$, which can be evaluated within a fraction of a second. The MS-FWI framework is described in Algorithm 2. In step 2, t_k is a step length obtained from either line-search method or a Lipschitz constant of the gradient. Step 3 refers to fast Mumford-Shah segmentation introduced in algorithm 1.

The convergence of our algorithm depends mildly on the parameter ρ. In our experiments, we have taken $\rho = 100$ and observed a linear convergence with iterations.

Algorithm 2 MS-FWI Algorithm

Require: segmentation parameters (α, λ), optimization parameter ρ, initial model $\mathbf{m}^0$, K , tolerance ϵ
Initialize: $\mathbf{z}^0 = \mathbf{m}^0$

Ensure: $\mathbf{m}^K$ - final model and $\mathbf{z}^K$ - segmented model

1: **for** $k = 0$ to $K - 1$ **do**
2: $\quad \mathbf{m}^{k+1} = \mathbf{m}^k - t_k \left(\nabla f(\mathbf{m}^k) + 2\rho\left(\mathbf{m}^k - \mathbf{z}^k\right)\right)$
3: $\quad \mathbf{z}^{k+1} = \text{fastMumfordShah}(\mathbf{m}^{k+1})$
4: $\quad$ **if** $\|\mathbf{m}^{k+1} - \mathbf{m}^k\|_2 < \epsilon$ **then**
5: $\quad\quad$ break
6: $\quad$ **end if**
7: **end for**

EXAMPLES

To demonstrate the capabilities of the proposed method, we present numerical experiments on two synthetic camembert model with acoustic data. The first camembert model is the classic model with a blob of constant velocity. The second camembert model consists of a circular disk with linear gradient velocity model in the oblique direction.

Experimental setup

We work with the models of size 1 km in each z and x direction, discretized with a 10-m grid spacing. For simplicity, we perform transmission experiment with sources placed at $x = 20$ m and receivers at $x = 980$ m. A total of 21 sources were placed 50 m apart. The source is a Ricker wavelet with a 10-Hz peak frequency and zero time lag. The data were acquired with receivers placed 10 m apart. To avoid a full inverse crime, we add a Gaussian noise of 30 dB SNR to the data. We model waves using scalar Helmholtz equation with perfectly matched layer boundary conditions on all sides (Da Silva and Herrmann, 2017). For inversion, we consider data of frequencies 5, 6, 7, 8 Hz. The purpose of these experiments are to show that the Mumford-Shah functional preserves the piecewise behavior of the velocity model, whereas TV introduces staircase effects during FWI.

We use $\rho = 100$ and $\epsilon = 10^{-6}$ for MS-FWI reconstructions and perform a total of $K = 500$ iterations. A step length t_k has been estimated using backtracking line-search method. To compare the MS-FWI reconstructions, we consider three methods: 1) classical FWI without any regularization, 2) classical non-regularized FWI followed by segmentation step (referred to as FWI-s henceforth), and 3) FWI with total-variation regularization (TV-FWI). In FWI and FWI-s, we perform a total of 500 L-BFGS iterations. In FWI-s, we first invert the velocity model using the classical FWI without any regularization, and then apply a single step of segmentation using Mumford-Shah algorithm mentioned in Algorithm 1. We make use of the formulation proposed in Peters and Herrmann (2017) for TV-FWI and perform a total of 500 constrained iterations.

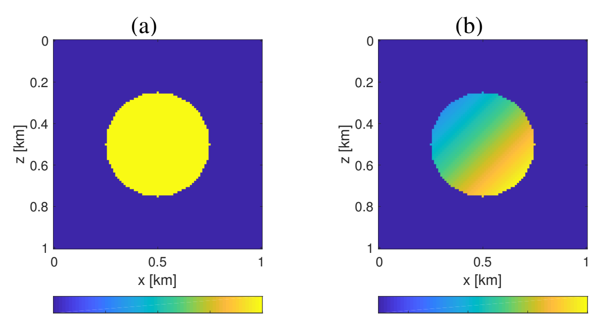

Figure 3: True velocity (in m/s) of synthetic (a) simple camembert model, and (b) gradient camembert model.

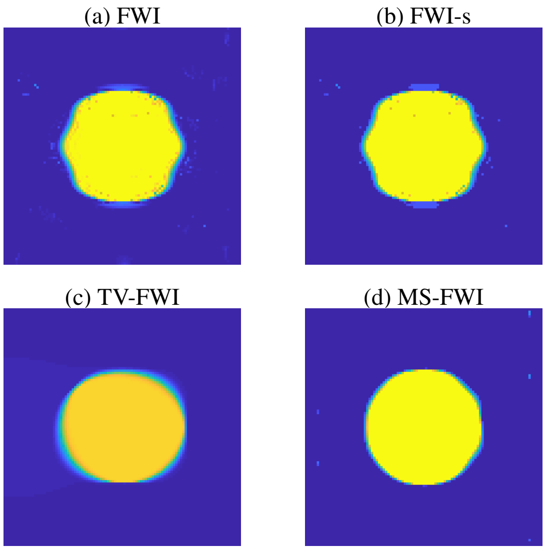

Figure 4: Reconstruction of simple camembert model with various methods.

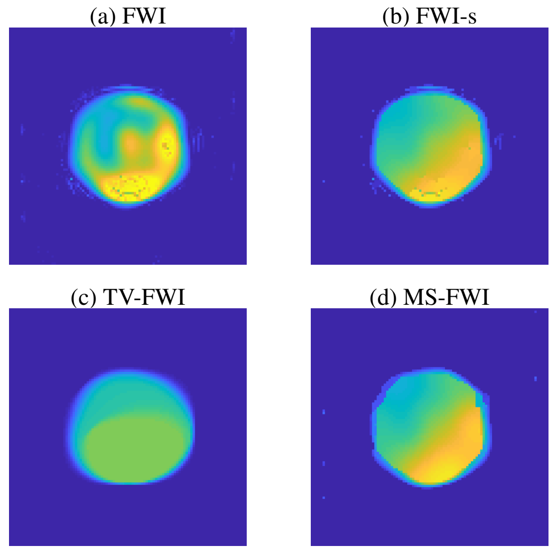

Figure 5: Reconstruction of gradient camembert model with various methods.

Method	Simple camembert	Gradient camembert
FWI	0.2006	0.2514
FWI-s	0.1997	0.2242
TV-FWI	0.2588	0.2948
MS-FWI	0.1227	0.2078

Table 1: Normalized model misfit (NMM) for various reconstruction methods. NMM = $\|\mathbf{m}^{\text{rec}} - \mathbf{m}^{\text{true}}\|_2 / \|\mathbf{m}^{\text{init}} - \mathbf{m}^{\text{true}}\|_2$.

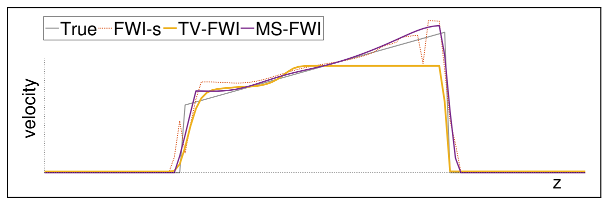

Figure 6: Reconstruction of vertical trace of gradient camembert model at $x = 0.5$ km.

Simple camembert model

Figure 3(a) shows a model with background velocity of 2500 m/s. It contains a circular disk of 250 m radius with velocity of 2900 m/s. We take $\alpha = 10^3$ and $\lambda = 1$ as segmentation parameter in both FWI-s and MS-FWI. A constraint parameter $\tau = 0.6$ has been chosen for TV-FWI. The results of the reconstructions from these methods have been presented in Figure 4. The FWI and consequently FWI-s fail to capture the shape of the blob, while TV-FWI incorrectly predicts the blob velocity. It is evident from Table 1 that the MS-FWI is the winner in this experiment.

Gradient camembert model

Figure 3(b) shows the model with background velocity of 2500 m/s and a circular disk with gradient in the velocity from 2700 m/s to 3200 m/s. We take $\alpha = 50$ and $\lambda = 1$ as segmentation parameter in both FWI-s and MS-FWI. Similarly, constraint parameter $\tau = 0.7$ has been chosen for TV-FWI. The results of the reconstructions from these methods have been presented in Figure 5.

Figure 6 shows the vertical tarce extracted from the inverted velocity models with linear gradient inside the circular disk. It is quite evident in both the examples that MS-FWI is able to capture and preserve the piecewise smoothness of the velocity model compared to the TV regularizer.

CONCLUSIONS

In this paper, we introduce a Mumford-Shah segmentation approach to include better prior information about the model in the seismic inversion. The Mumford-Shah regularizer make use of a non-convex penalty which has a simple proximal operator. This regularizer is integrated into the FWI procedure through penalization. The resulting formulation is solved alternatively to obtain a piecewise smooth model. We have shown that the method outperforms Total-Variation regularization to capture the trend on two different variants of camembert model. Our formulation is very general and has the potential to be applied to wide range of inverse problems.

ACKNOWLEDGMENT

This work is part of the Industrial Partnership Programme (IPP) 'Computational sciences for energy research' of the Netherlands Organisation for Scientific Research (NWO). This research programme is co-financed by Shell Global Solutions International B.V. T.v.L. is financially supported by NWO as part of research programme 613.009.032. R.K. would like to thank the member organizations of the SINBAD II project and the SINBAD Consortium for supporting this work. We thank Nick Luiken for helpful discussions and reviewing the manuscript.

REFERENCES

Ambrosio, L., and V. M. Tortorelli, 1990, Approximation of functional depending on jumps by elliptic functional via t-convergence: Communications on Pure and Applied Mathematics, **43**, 999–1036.

Asnaashari, A., R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, 2013, Regularized seismic full waveform inversion with prior model information: Geophysics, **78**, R25–R36.

Beydoun, W. B., and A. Tarantola, 1988, First born and rytov approximations: Modeling and inversion conditions in a canonical example: The Journal of the Acoustical Society of America, **83**, 1045–1055.

Boyd, S., N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al., 2011, Distributed optimization and statistical learning via the alternating direction method of multipliers: Foundations and Trends® in Machine learning, **3**, 1–122.

Chambolle, A., and T. Pock, 2011, A first-order primal-dual algorithm for convex problems with applications to imaging: Journal of mathematical imaging and vision, **40**, 120–145.

Da Silva, C., and F. J. Herrmann, 2017, A unified 2d/3d large scale software environment for nonlinear inverse problems: arXiv preprint arXiv:1703.09268.

Esser, E., L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, 2018, Total variation regularization strategies in full-waveform inversion: SIAM Journal on Imaging Sciences, **11**, 376–406.

Goldstein, T., and S. Osher, 2009, The split bregman method for l1-regularized problems: SIAM journal on imaging sciences, **2**, 323–343.

Li, X., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: Geophysics, **77**, A13–A17.

Lin, Y., and L. Huang, 2014, Acoustic-and elastic-waveform inversion using a modified total-variation regularization scheme: Geophysical Journal International, **200**, 489–502.

Mumford, D., and J. Shah, 1989, Optimal approximations by piecewise smooth functions and associated variational problems: Communications on pure and applied mathematics, **42**, 577–685.

Peters, B., and F. J. Herrmann, 2017, Constraints versus penalties for edge-preserving full-waveform inversion: The Leading Edge, **36**, 94–100.

Rudin, L. I., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: Physica D: nonlinear phenomena, **60**, 259–268.

Strekalovskiy, E., and D. Cremers, 2014, Real-time minimization of the piecewise smooth mumford-shah functional: European Conference on Computer Vision, Springer, 127–141.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

Tikhonov, A. N., 1963, On the solution of ill-posed problems and the method of regularization: Doklady Akademii Nauk, Russian Academy of Sciences, 501–504.

Unser, M., J. Fageot, and J. P. Ward, 2017, Splines are universal solutions of linear inverse problems with generalized tv regularization: SIAM Review, **59**, 769–793.

Vese, L. A., and T. F. Chan, 2002, A multiphase level set framework for image segmentation using the mumford and shah model: International journal of computer vision, **50**, 271–293.

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

Zheng, P., and A. Aravkin, 2018, Fast methods for nonsmooth nonconvex minimization: arXiv preprint arXiv:1802.02654.