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 behave 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 [CITE HERE], 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[CITE HERE].

One way to overcome the non-uniqueness of FWI is to add regularization to the data-fitting terms, which results in stable solutions [CITE HERE]. Various strategies have been proposed to impose the regularization such as Tikhonov [CITE HERE], sparsity-promotion based regularization [CITE HERE] and box-constraints [CITE HERE]. Recently, the Total Variation (TV) regularization method has been proposed, which resolves the sharp interfaces via preserving the edges and discontinuities [CITE HERE]. The central idea of TV regularization is to impose sparsity on the gradient of the model parameters. [CITE HERE] 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 plot a vertical trace extracted from the piece-wise smooth camembert model in Figure 1a. The corresponding projected trace on the TV-norm ball is shown in Figure 1b. We can clearly see that TV regularization approximates the smooth dipping part of the velocity model with a constant flat velocity model.

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 [CITE HERE]. 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 [CITE HERE]. The proposed FWI algorithm is based on splitting the problem using technique described in [CITE HERE] 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 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 (**?**) 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 Total-variation method. Tikhonov regularization defined as $\mathcal{R}(m) = \|\nabla m\|_2^2$, promotes smoothness in the model parameters by penalizing its spatial gradient. Total variation defined as $\mathcal{R}(m) = \|\nabla m\|_1$, on other hands, promotes jumps in the model. 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 [CITE HERE] 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 still remains hard. It is also important to start with a right boundary Γ for an optimization method to converge. Ambrosio and Tortorelli proposed a simpler version which approximates the Mumford-Shah functional but heavily depends on an extra parameter ϵ [CITE HERE]. We refer to the paper [CITE HERE] which discusses the effect of this parameter on image segmentation.

Relaxation of Mumford-Shah functional

As proposed in [CITE HERE], the Mumford-Shah regularization function can be relaxed using the following formulation:

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

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

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 [FIG HERE] 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

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

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:

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

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

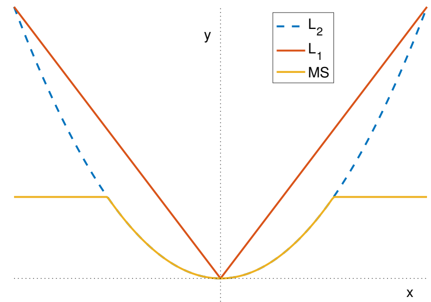

Figure 1: Comparison of Mumford-Shah regularizer (referred to as MS) with other regularization (ℓ_1 and ℓ_2).

Due to simplicity of the proximal operators, the relaxed problem (1) can be solved quickly using primal-dual algorithm [CITE HERE]. The algorithm is described in Algorithm 1. A segmented image is obtained in less than a second with $N = 10000$ iterations using this algorithm.

Algorithm 1 Fast Mumford-Shah segmentation

Require: Image f, α, λ
Initialize: $u^0 = f, \bar{u}^0 = u^0, p^0 = 0, \tau_0 = 1/4, \sigma_0 = 1/2$
Ensure: u^N
1: **for** $n = 0$ to $N - 1$ **do**
2: $p^{n+1}(x) = \mathbf{prox}_{\sigma,g}\left(p^n(x) + \sigma_n \nabla \bar{u}^n(x)\right)$
3: $u^{n+1}(x) = \mathbf{prox}_{\tau,f}\left(u^n(x) - \tau_n \operatorname{div} p^{n+1}(x)\right)$
4: $\theta_n = \frac{1}{\sqrt{1+4\tau_n}}$
5: $\tau_{n+1} = \theta_n \tau_n, \sigma_{n+1} = \frac{\sigma_n}{\theta_n}$
6: $\bar{u}^{n+1} = u^{n+1} + \theta_n \left(u^{n+1} - u^n\right)$
7: **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:

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

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

The matrix A denotes the discretization of the gradient using finite-difference. This problem can be solved using many methods including, primal-dual method [CITE HERE], alternating direction method of multipliers [CITE HERE], split-bregman

[CITE HERE]. All these methods require solving the minimization with respect to $\mathbf{m}$ at each step, 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 Sasha [CITE HERE] proposed a strategy to solve problem of form (2). The authors split the variables and add a penalty term on the difference and then use alternating minimization strategy. We resort to a similar version of this method:

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

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. Also 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 algorithm is described in Algorithm 2. In step 2, t_k refers to a step length which is obtained from line search method or Lipschitz constant of the gradient. Step 3 refers to fast Mumford-Shah segmentation introduced in algorithm 1.

Algorithm 2 MS-FWI Algorithm

Require: α, λ, ϵ, ρ, $\mathbf{m}^0$, K
Initialize: $\mathbf{z}^0 = \mathbf{m}^0$
Ensure: $\mathbf{m}$ - final model and $\mathbf{z}$ - segmented model
1: **for** $k = 0$ to K **do**
2: $\mathbf{m}^{k+1} = \mathbf{m}^k - t_k \left(\nabla f(\mathbf{m}^k) + 2\rho \left(\mathbf{m}^k - \mathbf{z}^k \right) \right)$
3: $\mathbf{z}^{k+1}$ = fastMumfordShah($\mathbf{m}^{k+1}$)
4: **end for**

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

EXAMPLES

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

Simple Camembert model

Figure 2 shows a model 1 km long and 1 km deep, discretized with a 10-m grid spacing. The sediment is a staircase in depth below a 300-m water layer with velocities between 1500 and 4000 m/s. The salt body, embedded in the sediment, has its top at 200 m below the water bottom with a constant velocity of 4500 m/s. Sources were placed 10 m deep and 200 m apart. The source is a Ricker wavelet with a 15-Hz peak frequency and zero time lag. The data were acquired with receivers placed 50 m apart starting at a smallest offset of 100 m up to a largest of 4 km. To avoid a full inverse crime, a different finite-difference code generated the data for a model discretized with a much finer grid spacing of 50/8 m. The amplitude of the source wavelet is estimated for each frequency at every step (**?**).

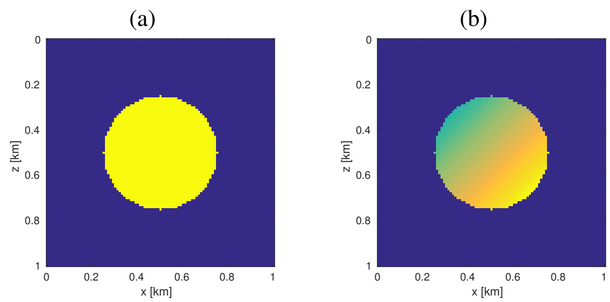

Figure 2: True velocity (in m/s) of synthetic (a) simple camembert model, and (b) gradient camembert model.

For the classic full-waveform inversion, we apply a spectral projected gradient method with bounds constraints (**?**) on the velocity. For the initial model, we take a linear velocity profile with depth. The inversion is performed in a multi-scale fashion over the frequency range 2.5–4.5 Hz with 200 iterations for each frequency batch and a total of 3 passes (**?**) over the frequency range. Figure **??** displays the results. The top of the salt near the water bottom has been reconstructed reasonably well, but not the salt body below. The sediment structure at larger depths is also lost.

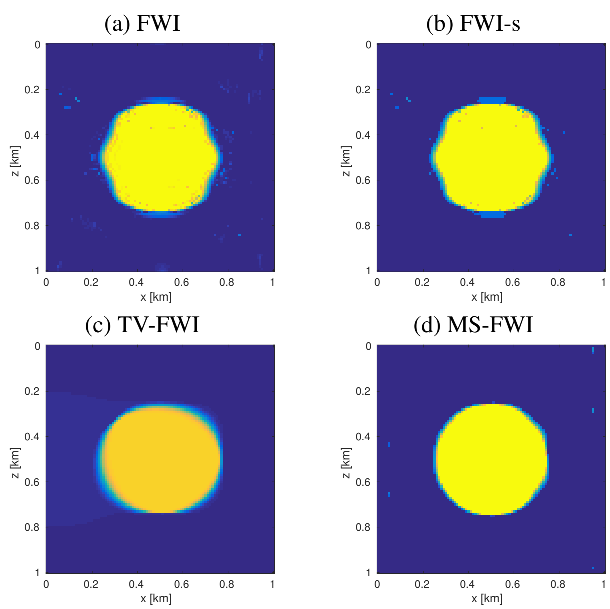

Figure 3: Reconstruction of simple camembert model with various methods

Method	Simple camembert	Gradient camembert
FWI	3553.63	3200.41
TV-FWI	4583.71	3633.27
MS-FWI	2173.29	2712.41

Table 1: Model misfit for various reconstruction methods. Model misfit is defined as ℓ_2-norm of the difference of the reconstructed model with the true model: $\|\mathbf{m}^{\text{rec}} - \mathbf{m}^{\text{true}}\|_2$

The parametric level-set parameters $\boldsymbol{\alpha}$ are initialized as shown in Figure **??**. All negative RBFs have a value of -1 and all positive of $+1$. We place $25(z) \times 20(x)$ nodes over the model grid for the background. The background parameters $\boldsymbol{\beta}$ are initialized with a smooth linear trend in depth. We incorporate prior information about the water bottom, at $z = 350$ m, in the initial model.

The optimization over β is performed with the interior-point method (`fmincon` in MATLAB®). We restrict ourselves to $J = 10$ iterations in algorithm **??**. We apply the `minFunc` (**?**) code in MATLAB® to optimize over α, limited to $K = 20$ iterations. In both these steps, the step direction is calculated by at most 10 iterations of the conjugate gradient method. A total of 4 passes are made over the frequency range, along with 2 inner iterations for each frequency batch. In total, we perform about 960 iterations. The Heaviside width parameter (κ) is initialized with 0.05 and reduced by 20% after each frequency pass to produce sharp boundaries for the salt.

Gradient Camembert model

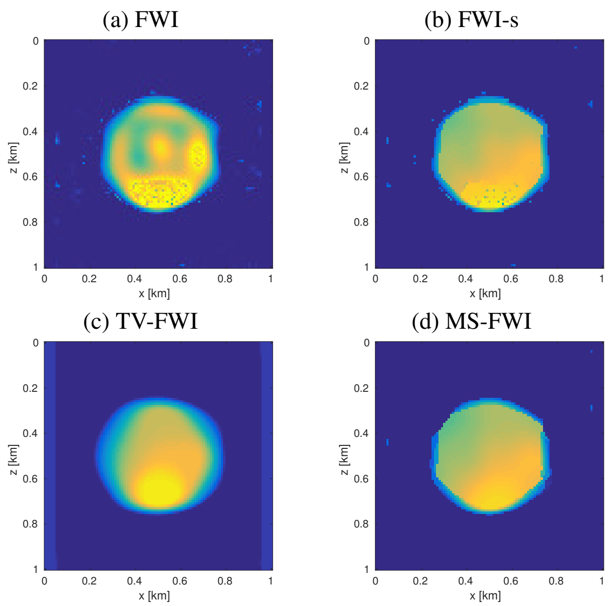

Figure 4: Reconstruction of gradient camembert model with various methods

Figure **??** shows the model obtained with the proposed method. The salt is reconstructed accurately at its top and sides. The sediment structure is reconstructed well down to a depth of 1.5 km as shown in Figure **??**. Figure **??** illustrates the need for multiple passes over the frequency range. Figure **??** shows that the method manages to fit the data for the lower frequencies but not for the higher.

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. The regularization is integrated into the FWI procedure by a simple penalization trick. The resulting formulation is solved alternatively to obtain a piecewise smooth model. We have shown that the method effectively captures 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.