

where  $\alpha > 0$ . The FWI problem in equation 7 ensures that the estimated model, while satisfying the data (and the wave equation), is smooth along the direction  $\theta$ .

Figure 1, top row, shows the anisotropic regularization balls for only one term of the regularizer equation 4 and for different values of  $[\sigma]_i = \sigma$  and  $[\theta]_i = \theta$ . We observe that for  $\sigma = 0.5$ , the anisotropic regularization becomes equivalent to the standard isotropic regularization. When the value of  $\sigma$  increases, reaching the maximum considered value of 0.9, the degree of regularization applied in the direction  $\theta$  and its normal is maximally different, resulting in a needle-shaped ellipse. This shape favors models with elongated features aligned with  $\theta$ , while allowing variations in the normal direction to it, as illustrated in the bottom row.

### 3 Algorithm

The augmented Lagrangian function of equation 7 is

$$\begin{aligned} \mathcal{L}(\theta, \sigma, u_s, m, \lambda, \nu) = & \frac{1}{2} \sum_{s=1}^{n_s} \|Pu_s - d_s\|_2^2 + \alpha \mathcal{R}(m, \theta, \sigma) \\ & + \frac{\mu}{2} \sum_{s=1}^{n_s} \|A(m)u_s - b_s\|_2^2 - \lambda_s^T (A(m)u_s - b_s) \\ & + \frac{\tau}{2} \|\theta - z\|_2^2 - \nu^T (\theta - z), \end{aligned}$$

where  $z(\theta) = \min(\max(\theta, -\frac{\pi}{2}), \frac{\pi}{2})$ ,  $\lambda_s$  and  $\nu$  are Lagrange multipliers associated to the constraints, and  $\mu > 0$  and  $\tau > 0$  are the penalty parameters. The regularization parameter  $\alpha > 0$  is given, being typically determined by the discrepancy principle.

This problem can be efficiently solved using the alternating direction method of multipliers (ADMM, [15]), resulting in the following iterative procedure starting from initial guess  $m$  (typically encoding some prior information), (spatially invariant)  $\sigma = 0.5$  and initial multipliers  $\lambda_s = \nu = 0$ :

$$\theta^+ = \arg \min_{\theta} \mathcal{L}(\theta, \sigma, u_s, m, \lambda, \nu) \quad (8a)$$

$$\sigma^+ = \arg \min_{\sigma} \mathcal{L}(\theta^+, \sigma, u_s, m, \lambda, \nu) \quad (8b)$$

$$u_s^+ = \arg \min_u \mathcal{L}(\theta^+, \sigma^+, u, m, \lambda, \nu) \quad (8c)$$

$$m^+ = \arg \min_m \mathcal{L}(\theta^+, \sigma^+, u_s^+, m, \lambda, \nu) \quad (8d)$$

$$\lambda_s^+ = \lambda_s - \mu(A(m^+)u_s^+ - b_s) \quad (8e)$$

$$\nu^+ = \nu - \tau(\theta^+ - z^+). \quad (8f)$$

Here, the updated variables at each iteration are denoted by a superscript “+”. Each step of the algorithm addresses a specific subproblem, as described more in details below.

Subproblem equation 8a involves minimizing the regularization function  $\mathcal{R}$ , as defined in equation 4, with respect to  $\theta$ , given the current values of  $m$  and  $\sigma$ . This can be efficiently achieved using a single iteration of the Gauss-Newton method. To ensure the stability of the  $\theta$  update, it is essential to incorporate a smoothing term, as suggested by [11].

Subproblem equation 8b involves minimizing  $\mathcal{R}$  with respect to  $\sigma$ , resulting in:

$$[\sigma]_i = \frac{[g_{z'}]_i^2}{[g_{x'}]_i^2 + [g_{z'}]_i^2}, \quad (9)$$

where  $([g_{x'}]_i, [g_{z'}]_i)^T = R([\theta]_i)[\nabla m]_i$  represent the rotated gradient of the current model at the  $i$ th pixel. This expression for equation 9 is justified by considering that, when  $[\theta]_i$  corresponds to the correct orientation angle, we typically expect  $|[g_{x'}]_i| \ll |[g_{z'}]_i|$ . In this case, directly minimizing  $[g_{x'}]_i^2 + [g_{z'}]_i^2$  would heavily penalize  $[g_{z'}]_i$ , which is undesirable because we aim to apply more smoothing along the direction defined by  $[\theta]_i$ . The weights  $[\sigma]_i$  in equation 9 balance the contributions of gradient components in the weighted sum  $[\sigma]_i^2 [g_{x'}]_i^2 + (1 - [\sigma]_i)^2 [g_{z'}]_i^2$ , ensuring that the magnitudes of both terms are approximately equal before penalization. This adaptive weighting mechanism effectively enhances smoothing along the structural direction  $[\theta]_i$  while allowing variations across it, eventually enabling the recovery of dominating structures and details in the model  $m$ . To maintain stability during the  $\sigma$  update, the computed weights are smoothed using a convolution with a  $5 \times 5$  averaging filter; alternatively, one could incorporate a smoothing term.

A comprehensive analysis of the methods for solving subproblems equation 8c and equation 8d can be found in [14].