process. The objective function to be minimized takes the form:

$$\min_{\mathbf{m}} \phi(m) = \|\mathbf{G}\,\rho - \mathbf{d}^{obs}\|_2^2 + \beta \sum_{r=s,x,y,z} \alpha_r \|\mathbf{W}_r\,\mathbf{G}_r\,\rho\|_2^2 \tag{3.11}$$

$$\text{subject to } \phi_d \leq \phi_d^*$$

where I set the reference density ρ^{ref} to zero. Since (3.11) is linear with respect to the density contrast model ρ, I can solve it uniquely for a fixed trade-off parameter β. I repeat this process for variable β values until a find a solution that satisfies $\phi_d \approx N$. Figure 3.2(a) presents a vertical section through the recovered density model. The density anomaly is imaged at roughly the right position, but the edges of the block are poorly defined. As normally obtained with $\ell_2 - norm$ penalties, the solution is smooth and density values remain small (near zero reference model). Hence the need to explore different regularization functions that can better resolve compact objects.

3.1 General ℓ_p-norm regularization

Alternatively, researchers have explored the use of non-ℓ_2 measures to promote the recovery of compact anomalies. Approximations to ℓ_1-norm such as the Huber norm (Huber, 1964)

$$\sum_i |m_i|^p \approx \sum_i \left\{ \begin{array}{ll} m_i^2, & |m_i| \leq \varepsilon, \\ 2\varepsilon|m_i| - \varepsilon^2, & |m_i| > \varepsilon, \end{array} \right\}$$

and the Ekblom norm (Ekblom, 1973):

$$\sum_i |m_i|^p \approx \sum_i (m_i^2 + \varepsilon^2)^{p/2} \tag{3.12}$$

have received considerable attention in geophysical inversion and signal processing (Daubechies et al., 2010; Farquharson and Oldenburg, 1998; Gorodnitsky and Rao, 1997; Li, 1993; Sun and Li, 2014). Likewise, the Lawson's measure (Lawson,
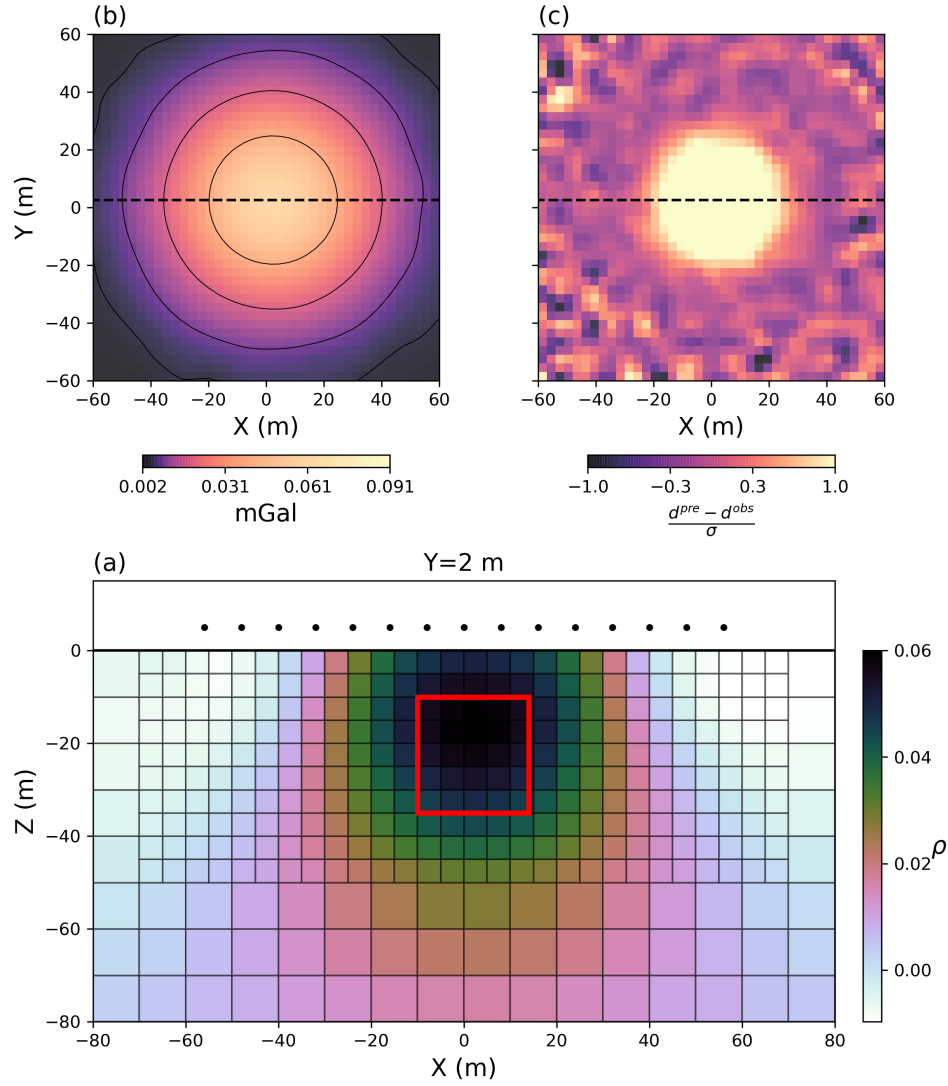
Figure 3.2: (a) Vertical section through the inverted density model using the conventional ℓ_2-norm regularization, (b) predicted and (c) normalized data residual. Outline of the true model (red) is shown for reference.

1961)

$$\sum_i |m_i|^p \approx \sum_i \frac{m_i{}^2}{(m_i{}^2 + \varepsilon^2)^{1-p/2}} \, , \tag{3.13}$$

has been proposed to approximate ℓ_0-norm and it has proven useful in generating minimum support models. This formulation has received considerable attention in the literature. (Ajo-Franklin et al., 2007; Barbosa and Silva, 1994; Last and Kubik, 1983; Portniaguine, 1999). Figure 3.3 compares the ℓ_p-norms with the Lawson approximation over a range of model values. As $\varepsilon \rightarrow 0$, the approximation approaches the ℓ_p-norm on the complete interval $p \in [0,2]$. While (3.13) would in theory permit us to explore a wide range of solutions for $0 \le p \le 2$, its numerical implementation remains challenging. Most algorithms have been limited to the ℓ_0, ℓ_1, and ℓ_2-norm measure applied evenly to all components of the model objective function.

Recent efforts by Sun and Li (2014) has shown promise in further exploring the model space by varying ℓ_p-norm measures locally. They divided the inversion domain into regions reacting favourably to either the ℓ_1 or ℓ_2-norm regularization. The automated process could adapt to complex geological scenarios where both smooth and blocky anomalies are present. Building upon the work I introduced in my M.Sc. thesis (Fournier, 2015), I want to extend the work of Sun and Li (2014) and further generalized the mixed norm inversion for $p \in [0\ 2]$.

3.1.1 Synthetic 1D problem

To develop my methodology it suffices to work with a simple test example. In Figure 3.4(a) I present a synthetic 1D model made up of a boxcar anomaly. The region is divided into 50 uniform cells distributed along the interval $[0 \le x \le 1]$. I define a synthetic geophysical experiment such that the data ($\mathbf{d}^{obs}$) are

$$\mathbf{d}^{obs} = \mathbf{F}\,\mathbf{m}^{true} + \mathbf{e} \, , \tag{3.14}$$

The kernel coefficients F_{ij} are sampled from a standard normal distribution of positive values multiplied by the discretization intervals. Choosing a stochastic kernel function for a linear inverse problem is unusual. Smooth functions are usually em-
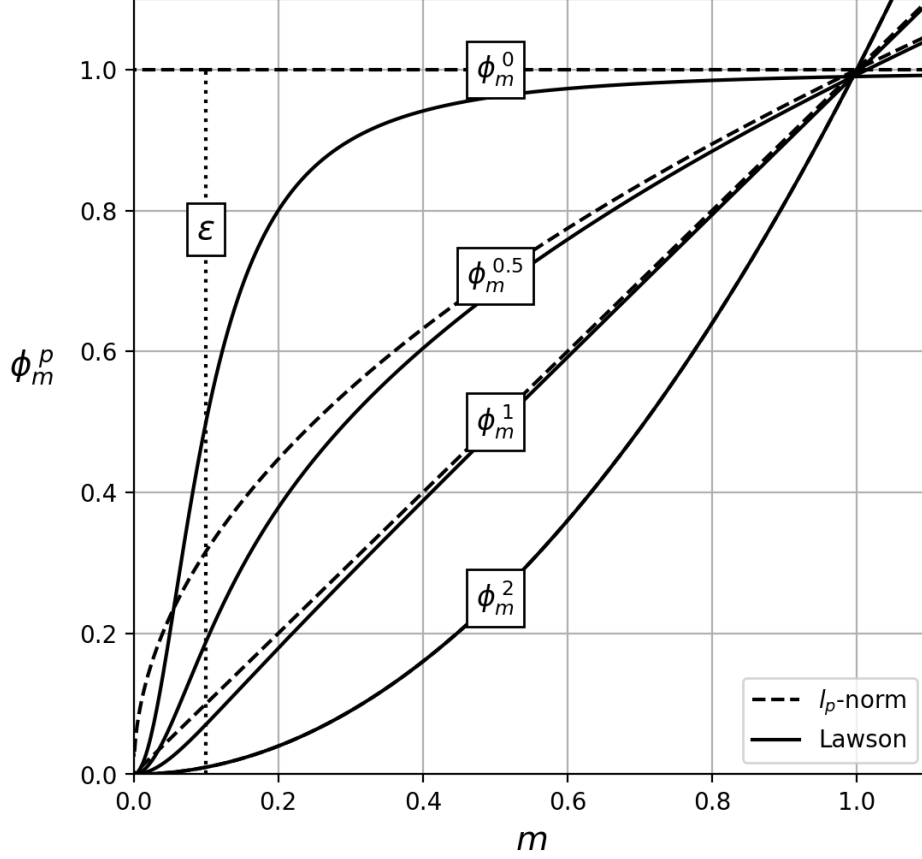
Figure 3.3: Approximated ℓ_p-norm using the Lawson measure (Lawson, 1961) over a range of p-values and for a fixed threshold parameter $\varepsilon = 10^{-1}$.

ployed (polynomials, decaying exponentials, sinusoidals), but my choice will serve to highlight the effects of various regularization functions. I generate 10 data, so $\mathbf{F} \in \mathbb{R}^{N \times M}$ where $M = 50$ and $N = 10$. Random Gaussian error $\mathbf{e}$ (σ=0.025) is added to simulate noise (Fig. 3.4(c)).

To begin my analysis, I invert my synthetic dataset with two simple regularization functions. Figure 3.5(a) presents the recovered model after reaching the target misfit ($\phi_d^* = N$) using the smallness term alone ($\alpha_x = 0$). The solution exhibits high variability similar to the stochastic kernel function, but model parameters remain near the implied zero reference value. Next, I invert the data using the model gradi-
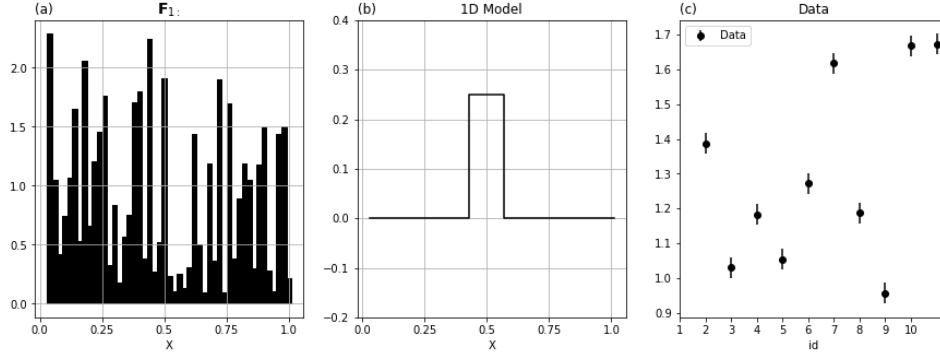
Figure 3.4: Linear forward problem made up of: (a) an example kernel function ; (b) model; (c) observed data with assigned standard errors.

ent term ($\alpha_s = 0$); this yields the smoother model presented in 3.5(b). The solution shows less spatial variability and the horizontal position of the boxcar anomaly is better located.

Next, I combine both regularization functions so that the solution remains close to the reference value and smooth. I need to determine the length scale weighting proposed in (3.6). For my problem $h = 0.02$ and hence I set $\alpha_x = 1$ and $\alpha_s = 2500$. Inverting with these parameter values yields the model in 3.5(c). Convergence curves presented in Figure 3.5(d) show the evolution of ϕ_d and ϕ_m as a function of iteration. As β decreases (shown in Figure 3.5(e)), the misfit, ϕ_d progressively decreases while model complexity, indicated by ϕ_m, progressively increases.

Visually, the solution 3.5(c) exhibits characteristics of remaining near the zero reference value while also attempting to be smooth. Numerical evaluation of the two components of the regularization function, presented in Table 3.1, show that $\phi_s = 3.31$ and $\phi_x = 1.13$. This might suggest that both ϕ_s and ϕ_x are roughly equal in importance.

Rather than work with global norms, in this study, I propose to quantify the relative importance of the terms in the regularization function based on their partial derivatives, or gradients. From (3.10) I expect to find an optimal solution where the sum of the gradients vanishes, either because all components are equal to zero, or because multiple gradients have opposite signs. To quantify the size of the
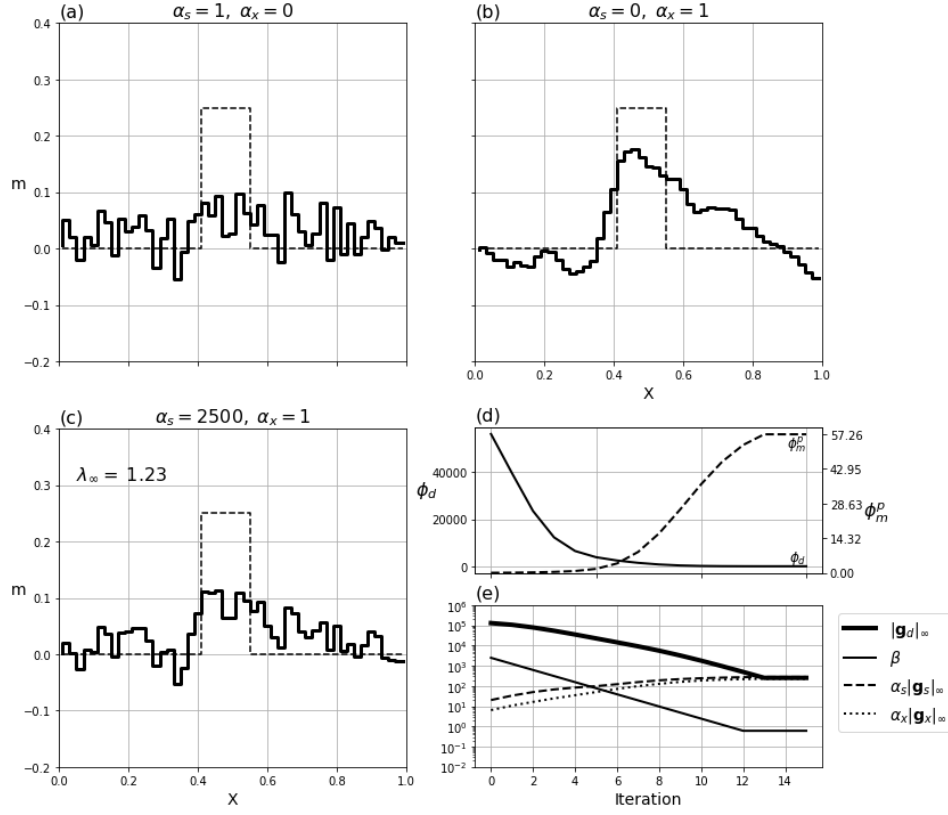
36

Figure 3.5: Solution to the 1D inverse problem using (a) an ℓ_2-norm on the model ($\alpha_x = 0$), (b) the ℓ_2-norm on model gradients ($\alpha_s = 0$) and (c) combined regularization function ($\alpha_s = 2500$, $\alpha_x = 1$). (d) Convergence curve comparing the misfit (ϕ_d) and the regularization (ϕ_m) as a function of iterations. (e) Comparative plot for the relative contribution of the different components of the objective function measured in terms of maximum absolute gradient ($\|\mathbf{g}_i\|_\infty$)

gradients I use the infinity-norm

$$\|\mathbf{g}_r\|_\infty = \|\nabla_m \phi_t\|_\infty \qquad (3.15)$$

The $\|\mathbf{g}_r\|_\infty$ metric is appealing for a few reasons: (a) it is directly linked to the minimization process because I use gradient descent methods, (b) it does not depend on the dimension M of the parameter space as do other measures that involve a sum of components of the vector, (c) the theoretical maximum can be calculated

α_s	α_x	$\alpha_s\,\phi_s$	$\alpha_x\,\phi_x$	λ_∞
2500	1	3.31	1.13	1.23

Table 3.1: Norm values and proportionality ratio obtained for the 1D solution presented in Figure 3.5(c). A proportionality ratio of $\lambda_\infty \approx 1$ indicates that the components of the regularization function are both contributing significantly to the final solution.

analytically for any given ℓ_p-norm function. These properties will become useful in the following section when I attempt to balance different norm penalties applied on a cell-by-cell basis.

Figure 3.5(e) compares $\|\mathbf{g}_d\|_\infty$, $\alpha_s\|\mathbf{g}_s\|_\infty$ and $\alpha_x\|\mathbf{g}_x\|_\infty$ over the iterative process. I note that, under the current α-scaling strategy proposed in (3.6), the individual partial derivatives for ϕ_s and ϕ_x also appear to be proportional in magnitude. To quantify this I define a proportionality ratio:

$$\lambda_\infty = \frac{\alpha_s\,\|\mathbf{g}_s\|_\infty}{\alpha_x\,\|\mathbf{g}_x\|_\infty} \tag{3.16}$$

I shall use λ_∞ as an indicator to evaluate the relative influence of (any) two terms in the regularization function. For my example $\lambda_\infty = 1.23$, from which I infer that ϕ_s and ϕ_x are contributing nearly equally to the solution (Table 3.1). As I further generalize the regularization function for arbitrary ℓ_p-norm measures, I will attempt to maintain this proportionality ratio ($\lambda_\infty \approx 1$) between competing functions so that my modeling objectives are preserved throughout the inversion process.

3.1.2 Iterative Re-weighted Least Squares algorithm

Solutions obtained with ℓ_2-norm regularization functions provided some insight about the sought model but better representations can be obtained by employing general ℓ_p-norms:

$$\phi_s^p = \sum_i |m_i|^p \tag{3.17}$$

My main focus is in the regularization function in (3.3) which I approximate with the Lawson norm such that

$$\phi_m = \sum_{r=s,x} \int_V \frac{f_r(m)^2}{\left(f_r(m)^2 + \varepsilon^2\right)^{1-p_r/2}} \, dV \, , \tag{3.18}$$

This measure makes the inverse problem non-linear with respect to the model. The common strategy is to linearize the function through the Iterative Reweighted Least-Squares (IRLS) approach where the denominator is replaced by model parameters from the most recent iteration. The smallest model component can be written as:

$$\phi_s^{p_s} = \sum_{i=1}^{M} \frac{m_i^2}{\left((m_i^{(k-1)})^2 + \varepsilon^2\right)^{1-p_s/2}} V_i \tag{3.19}$$

where $m_i^{(k-1)}$ are model parameters obtained at a previous iteration and V_i are volume terms connected with the discretization. In (3.19) I have explicitly written the objective function as $\phi_s^{p_s}$ to indicate that I am evaluating a smallest model component with an ℓ_p-norm with $p = p_s$. The approximated norm can be expressed in linear form as:

$$\phi_s^{p_s} = \|\mathbf{V}_s \, \mathbf{R}_s \, \mathbf{m}\|_2^2 \, . \tag{3.20}$$

where the IRLS weights $\mathbf{R}_s$ are defined as

$$\mathbf{R}_s = \text{diag}\left[\mathbf{r}_s\right]^{1/2}$$
$$r_{s_i} = \left(\left(m_i^{(k-1)}\right)^2 + \varepsilon^2\right)^{p_s/2-1} . \tag{3.21}$$

Carrying out the same procedure on the measure of model derivatives yields

$$\phi_x^{p_x} = \sum_{i=1}^{M-1} \frac{\left(\frac{m_i - m_{i+1}}{h_i}\right)^2}{\left[\left(\frac{m_i^{(k-1)} - m_{i+1}^{(k-1)}}{h_i}\right)^2 + \varepsilon^2\right]^{1-p_s/2}} V_i \tag{3.22}$$

where h_i defines the cell-center distance between neighboring model parameters. This can be expressed in linear form as

$$\phi_x^{p_x} = \|\mathbf{V}_x \mathbf{R}_x \mathbf{G}_x \mathbf{m}\|_2^2 \, , \tag{3.23}$$

where the gradient operator and the corresponding IRLS weights are calculated by

$$\mathbf{G}_x = \begin{bmatrix} -h_1^{-1} & h_1^{-1} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & h_{M-1}^{-1} & h_{M-1}^{-1} \end{bmatrix} . \tag{3.24}$$

and

$$\mathbf{R}_x = \operatorname{diag}\left[\mathbf{r}_x\right]^{1/2}$$
$$r_{x_i} = \left[\left(\frac{m_i^{(k-1)} - m_{i+1}^{(k-1)}}{h_i} \right)^2 + \varepsilon^2 \right]^{p_x/2-1} . \tag{3.25}$$

respectively. The final regularization function is thus

$$\phi_m^p = \alpha_s \|\mathbf{V}_s \mathbf{R}_s \mathbf{m}\|_2^2 + \alpha_x \|\mathbf{V}_x \mathbf{R}_x \mathbf{G}_x \mathbf{m}\|_2^2 \, . \tag{3.26}$$

The core IRLS procedure described in Table 3.2 involves two main stages:

1. Stage 1 solves the inverse problem using ℓ_2-norms presented in (3.5). The assumption is made that the globally convex ℓ_2-norm regularized inversion is a good approximation of the true solution and it is used to form the initial IRLS weights defined in (3.21). The β parameter is controlled by a cooling schedule that starts with a high value and is successively decreased until $\phi_d \approx \phi_d^*$.

2. Stage 2 starts from the solution obtained in Stage 1 and solves the inverse problem iteratively using the regularization in (3.26) and a standard Gauss-Newton procedure. A gradient descent direction $\delta\mathbf{m}$ is found by solving

$$\mathbf{H} \, \delta\mathbf{m} = \mathbf{g} \tag{3.27}$$

where **H** is the approximate Hessian and **g** is the gradient of the objective function. I use the Conjugate Gradient method (Hestenes and Stiefel, 1952) to solve this system.

The model update at the kth iteration is

$$\mathbf{m} = \mathbf{m}^{(k-1)} + \alpha \delta \mathbf{m} \tag{3.28}$$

where the step length α is found by a line-search back-stepping method (Nocedal and Wright, 1999). Gradient steps are only performed if the data misfit remains within the user-defined tolerance η_{ϕ_d}.

$$\frac{|\phi_d - \phi_d^*|}{\phi_d^*} \leq \eta_{\phi_d} \tag{3.29}$$

If outside the tolerance, the algorithm repeats the Gauss-Newton calculation with the previous $\mathbf{m}^{(k-1)}$ and a different β-value, either lower or higher depending on the achieved ϕ_d. This β-search step is an important component in the workflow when the minimization switches between an l_2 to an l_p objective function because ϕ_m^p can vary markedly. This can force a change of β by a few orders of magnitude in some cases. Once an appropriate β has been found such that (3.29) is respected, the model update $\mathbf{m}^{(k)}$ is accepted and used for the next iteration cycle. The IRLS process continues until the change in regularization falls below some pre-defined tolerance η_{ϕ_m}

$$\frac{|\phi_m^{(k-1)} - \phi_m^{(k)}|}{\phi_m^{(k)}} < \eta_{\phi_m} \tag{3.30}$$

I set to $\eta_{\phi_m} = 10^{-5}$ (0.01% change) in all my experiments. Using the above algorithm I now explore specific inversions for a fixed $\varepsilon = 10^{-3}$ and uniform norms, with $p = 1$ and $p = 0$, applied on the model and model gradients.

3.1.3 Case 1: ℓ_1-norm ($p_s = p_x = 1$)

I first address the convex case for $p_s = p_x = 1$ for which optimality can be guaranteed (Daubechies et al., 2010; Osborne, 1985). Using the procedure prescribed in Table (3.2), I invert the 1D problem with three different regularization functions:

Stage 1: Initialization (ϕ_m^2)	Stage 2: IRLS (ϕ_m^p)
$\min\limits_{m}\ \phi_d + \beta\phi_m^2$ s.t. $\phi_d = \phi_d^*$ $\beta^{(0)}, \mathbf{m}^{(0)}, \mathbf{R}^{(0)}, \phi_m^{(0)}$	**while** $\dfrac{\lvert \phi_m^{(k-1)} - \phi_m^{(k)} \rvert}{\phi_m^{(k)}} > \eta_{\phi_m}$ do β-**Search** $k := k+1$ $\beta^{(k)}, \mathbf{m}^{(k)}, \mathbf{R}^{(k)}$

β-Search

Solve $\mathbf{H}\,\delta\mathbf{m} = \mathbf{g}$

$\mathbf{m} = \mathbf{m}^{(k-1)} + \alpha\,\delta\mathbf{m}$

if $\dfrac{\lvert \phi_d - \phi_d^* \rvert}{\phi_d^*} > \eta_{\phi_d}$

adjust β, re-do

else

continue

Table 3.2: IRLS algorithm in pseudo-code made of two stages: Stage 1 Initialization with convex least-squares inversion, Stage 2 IRLS updates with inner β-search steps.

(a) l_1-norm measure of the model ($\alpha_x = 0$), (b) l_1-norm measure of the model gradients ($\alpha_s = 0$) and (c) for the combined penalties using $\alpha_s = 2500$, $\alpha_x = 1$, which I previously used for the l_2-norm inversion.

As shown in Figure 3.6(a), the first inversion is successful in recovering a sparse solution. From Linear Programming (LP) theory, the expected optimal solution would have as many non-zero parameters as there are linearly independent constraints or 10 values in this case. For comparison, I solve the LP problem by the Simplex routine from the open-source library `Scipy.Optimization.linprog` (Jones et al., 2001). Figure 3.6(a) compares both solutions and shows that my implementation of IRLS for l_1-norm yields a solution in close agreement with the Simplex routine. A better approximation could be obtained (not shown here) by lowering the threshold parameter ε. I will examine this aspect of the algorithm in the following section.

Figure 3.6(b) presents the solution for the l_1-norm applied to the model gradients. The final solution is *blockier* and the general shape of the boxcar model has
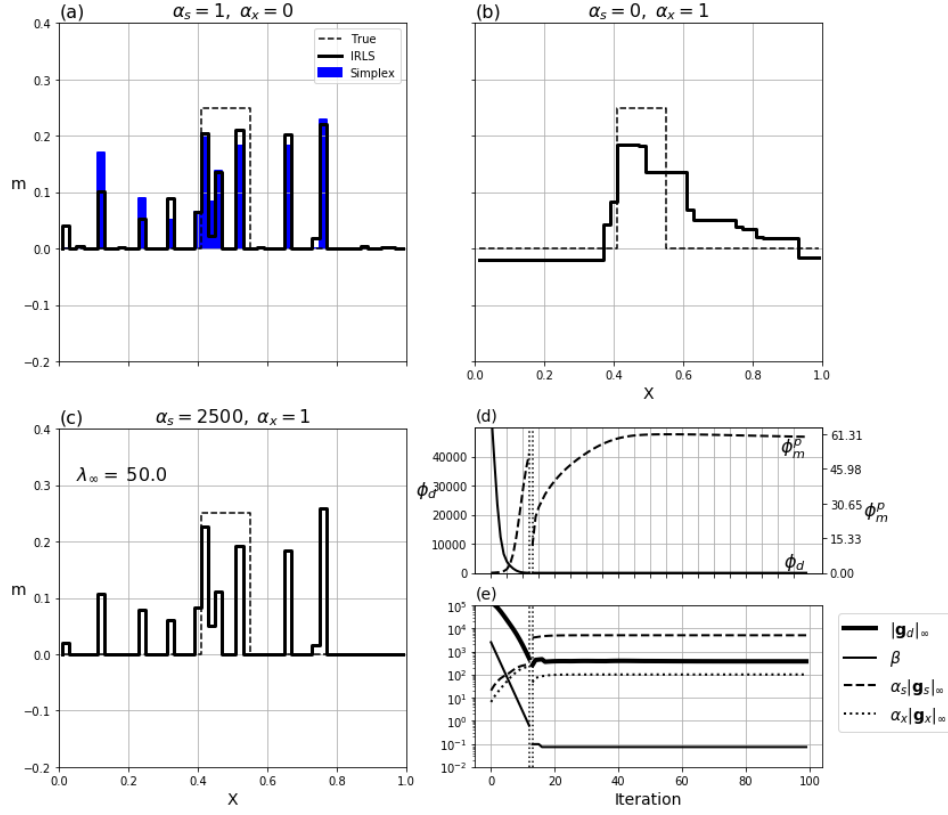
Figure 3.6: (a) Two solutions using an ℓ_1-norm on the model: (blue) Simplex, and (black) IRLS method. (b) Solution obtained with the approximated ℓ_1-norm (IRLS) penalty on model gradients alone and (c) with the combined penalty functions ($\alpha_s = 2500$, $\alpha_x = 1$). The calculated proportionality ratio λ_∞ indicates that the combined penalties is dominated by the ϕ_s^1 term. (d) Convergence curve and (e) maximum partial derivatives associated with the components of the objective function as a function of iterations for the inversion in (c). The vertical dotted lines indicate the change in regularization from an ℓ_2-norm to ℓ_1-norm measure.

been improved. Lastly, the solution obtained with the combined l_1-norm regularization on the model and gradients is shown in Figure 3.6(c); it is similar to that in Figure 3.6(a). This shows that the smallest model component has dominated the solution. This is quantified by the evaluated proportionality ratio $\lambda_\infty = 50$; setting $\alpha_s = 2500$ is too large.

To understand the previous result, I can factor out the cell centered distance h_i
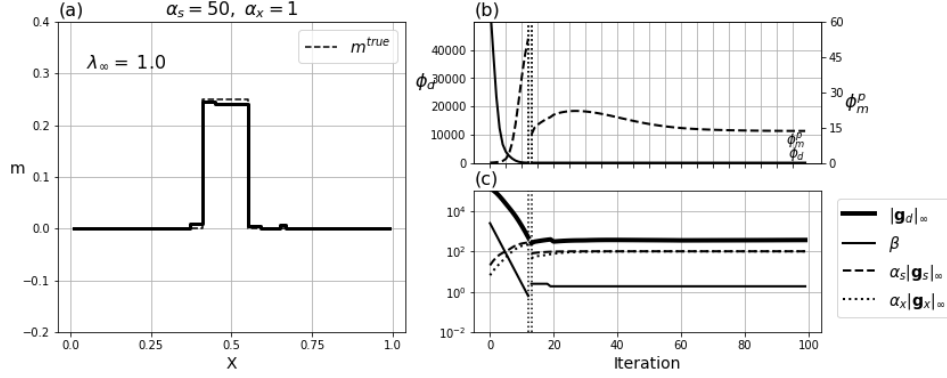
Figure 3.7: (a) Solution obtained with the combined penalty functions $\alpha_s \phi_s^1 + \alpha_x \phi_x^1$ after re-adjustment of $\alpha_s = 50$, $\alpha_x = 1$. (b) Convergence curve and (c) maximum partial derivatives associated with the components of the objective function as a function of iteration.

from (3.22) such that

$$\phi_x^{p_x} = \sum_{i=1}^{M-1} \frac{h_i^2 \left(m_i - m_{i+1}\right)^2}{h_i^{p-2} \left[\left(m_i^{(k-1)} - m_{i+1}^{(k-1)}\right)^2 + h_i^2 \varepsilon^2\right]^{1-p_s/2}} V_i \qquad (3.31)$$

This expression clearly shows a differences in scales between ϕ_s^p and ϕ_x^p, previously fixed in (3.6), that now depends on the p-value such that

$$\frac{\left[\phi_x^p\right]}{\left[\phi_s^p\right]} = [h]^{-p} . \qquad (3.32)$$

In accordance to this new relationship, I can re-adjust the importance of ϕ_s^p by setting $\alpha_s = 50$. After applying this change I recover the model presented in Figure 3.7(a). The combined assumption of a piece-wise continuous and sparse model yields a solution that closely resembles the true boxcar model. The recovery of $\mathbf{m}^{true}$ has remarkably improved compared to the ℓ_2-norm solutions (Fig. 3.5), and this demonstrates the power of customizable objective functions. It is important to notice that the re-adjustment of α_s has brought the partial derivatives of $\phi_s^{p_s}$ and $\phi_x^{p_x}$ to a comparable level, with a final proportionality ratio $\lambda_\infty = 1.01$. Even though

44

I have changed the norms during the inversion, the contribution of both penalty functions has remained at a comparable level during the transition between Stage 1 and 2 of the algorithm (Fig 3.7(c)).

3.1.4 Case 2: ℓ_0-norm ($p_s = p_x = 0$)

The main advantage of the IRLS formulation is that it permits, in theory, approximating any norm including the non-linear approximation for $p < 1$. The goal is to potentially recover a model with even fewer non-zero parameters than that obtained by solving the problem with $p = 1$. The IRLS formulation for $p = 0$ has been implemented for various geophysical problems under different names: such as the *compact* inversion (Last and Kubik, 1983), *minimum support* functional (Portniaguine and Zhdanov, 2002), and others (Ajo-Franklin et al., 2007; Barbosa and Silva, 1994; Blaschek et al., 2008; Chartrand, 2007; Stocco et al., 2009).

Following the same IRLS methodology as described in Table 3.2, I invert the synthetic 1D problem with three assumptions: (a) ℓ_0-norm applied on the model ($\alpha_x = 0$), (b) ℓ_0 on model gradients ($\alpha_s = 0$) and combined penalties ($\alpha_s = 1$, $\alpha_x = 1$). Figure 3.8 presents the solutions for all three cases. I note that in the first case, (a), the approximate ℓ_0-norm inversion recovers a sparser solution than obtained with the ℓ_1-norm; there are only eight non-zeros parameters. Similarly for case (b), I recover a model with fewer changes in model values. Finally in case (c) the solution obtained with the combined ℓ_0-norm penalties matches almost perfectly the true boxcar anomaly. The final proportionality ratio as calculated from (3.16) indicates once again a good balance between the penalty functions ($\lambda_\infty = 1.13$).

To summarize this section, I have now recovered nine models using different ℓ_p-norm penalties applied on the model and model gradient. All solutions presented in Figure 3.5, 3.6 and 3.8 can reproduce the data within the predefined data tolerance ($\phi_d^{(k)} \approx N$). Without prior knowledge about the true signal, all these solutions would be valid candidates to explain the observed geophysical data.

3.2 Mixed norm regularization

While I was successful in recovering a solution that closely resembles the boxcar model, the same penalty functions might not be appropriate for other models, such
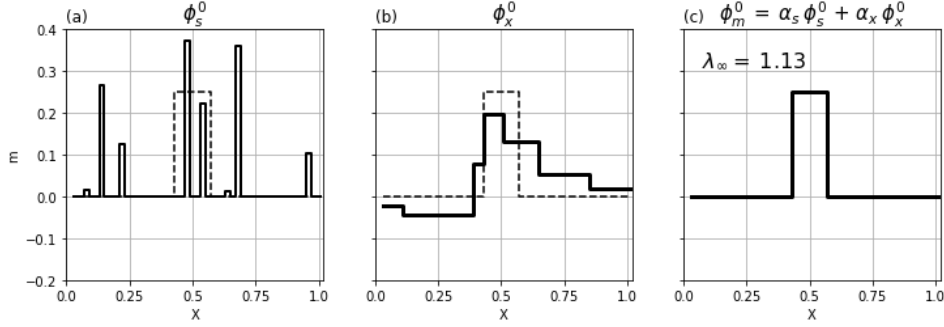
Figure 3.8: Solution to the 1D inverse problem using an approximate ℓ_0-norm (a) on the model, (b) on model gradients and (c) combined penalty functions using the IRLS algorithm ($\alpha_s = 1$, $\alpha_x = 1$). All three solutions honor the data within the target misfit ϕ_d^*.

as compact targets with smooth edges. I thus explore a broader range of solutions by using the Lawson approximation in (3.19) for any combination of norms on the range $0 \leq p \leq 2$.

The idea of combining different norm measures for the simultaneous recovery of smooth and compact features has partially been explored by Sun and Li (2014) on a 2D seismic tomography problem. They demonstrated the benefits of dividing model space into regions with different ℓ_p-norm penalties. The choice of norms was limited to be either l_1 or l_2. Little has been published however on the independent mixing of model and gradient norms on the range $p \in [0, 2]$, although this problem was initially addressed in (Fournier, 2015). I now apply my algorithm to minimize

$$\phi_m^p = \alpha_s \phi_s^0 + \alpha_x \phi_x^2 . \tag{3.33}$$

where once again the superscript indicates the ℓ_p-norm measure used in each function. Based upon my previous work, I expect the solution to be sparse, in terms of non-zero model parameters, while smooth with respect to the model gradients. Unfortunately, following the current IRLS strategy, I recover the model presented in Figure 3.9(a). The anomaly is concentrated near the boxcar but appears to be dominated by ϕ_s^0. There seems to be only marginal influence from ϕ_x^2. Comparing the partial derivatives of the objective function confirms this. After convergence
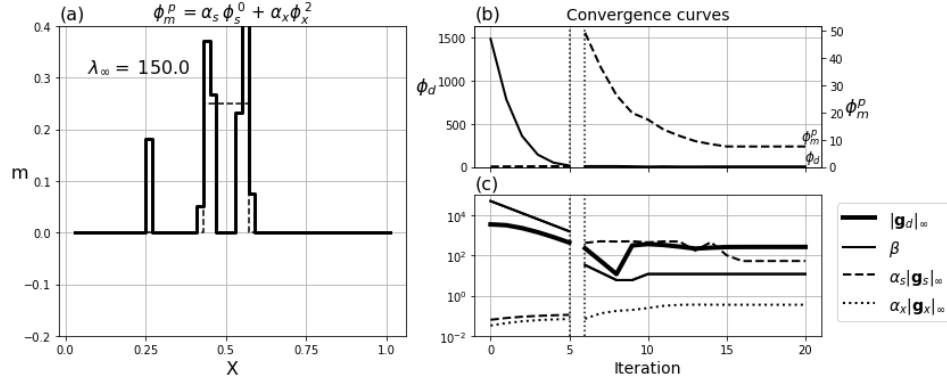
Figure 3.9: (a) Recovered model and (b) convergence curves using the conventional IRLS method for $p_s = 0$, $p_x = 2$ and a fixed threshold parameter $\varepsilon = 10^{-3}$ ($\alpha_s = \alpha_x = 1$). (c) Trade-off parameter and maximum gradients for the different components of the objective function. At the start of Stage 2 (iteration 6), the sudden increase in $\|\mathbf{g}_s\|_\infty$ is matched with a decrease in β. Throughout the inversion, $\|\mathbf{g}_x\|_\infty$ remains small in magnitude.

the calculated proportionality ratio is $\lambda_\infty = 159$. This is a significant change from the end of Stage 1 where $\lambda_\infty \approx 1$. Clearly, iteration 6, at Stage 2 of the IRLS, took the solution away from the proportionality condition (Fig 3.9(c)). I hypothesize that a more desirable solution could be obtained if proportionality was preserved among the components of the objective function throughout the IRLS process. In the following sections, I provide an important modification to the standard IRLS algorithm to achieve this goal.

3.2.1 Scaled-IRLS steps

Since the inverse problem is solved using gradients of the composite objective function $\phi(\mathbf{m})$, the relative magnitude of the individual gradients is a driving force in controlling the iteration step in (3.27). I want to ensure that each penalty term in the objective function is playing a significant role. Taking the partial derivatives of the linearized Lawson norm as prescribed in (3.19) yields:

$$g^p = \frac{\partial \phi^p}{\partial m} = \frac{f(m)}{\left(f(m^{(k-1)})^2 + \varepsilon^2\right)^{1-p/2}} V \,. \tag{3.34}$$

47

From Figure 3.10(a), I note that the magnitude of the derivatives increases rapidly for small p values as $m_i \to 0$. This trend is accentuated for small ε values as demonstrated in Figure 3.10(b) for $p = 0$. The magnitude of derivatives for $p < 2$ increase rapidly as $m \to 0$ and $\varepsilon \to 0$. This results in gradient steps in equation (3.10) that are dominated by sparse norms. This property of the Lawson approximation is important because, when attempting to combine different norm penalties within the same objective function, there will be a systematic bias towards small ℓ_p-norm penalties. To circumvent this bias I define the following gradient-based scaling

$$\gamma = \left[\frac{\|g^2\|_\infty}{\|g^p\|_\infty} \right]^{1/2} \tag{3.35}$$

By using this scaling I can equalize the size of the gradients associated with any ℓ_p-norm. I can easily compute $\|g^p\|_\infty$ for any function $f(m)$ by taking the partial derivative of (3.34) and setting it to zero. This maximum gradient of the Lawson approximation occurs at $f(m)^*$

$$f(m)^* = \begin{cases} \infty \text{ or } f(m)_{\max}, & p \geq 1 \\ \dfrac{\varepsilon}{\sqrt{1-p}}, & p < 1 , \end{cases} \tag{3.36}$$

from which I can calculate $\|g^p\|_\infty$ by substituting $f(m)^*$ into (3.34). Figure 3.10(c) presents the derivatives for different approximated ℓ_p-norms after applying the corresponding γ-scale. Note that the largest derivative of any norm is at most as large as the ℓ_2-norm penalty for $m \in \mathbb{R}$. The scaling of the norm derivatives guarantees that two penalties can co-exist and impact the solution at every step of the IRLS, regardless of the chosen $\{p, \varepsilon\}$-values.

I therefore define Scaled-IRLS weights such that (3.21) become:

$$\mathbf{R}_s = \gamma_s \, \mathrm{diag} \left[\left((\mathbf{m}^{(k-1)})^2 + \varepsilon^2 \right)^{p_s/2 - 1} \right]^{1/2} \tag{3.37}$$

where the scaling parameter γ_s is:

$$\gamma_s = \left[\frac{\|\mathbf{g}_s^2\|_\infty}{\|\mathbf{g}_s^{p_s}\|_\infty} \right]^{1/2} \tag{3.38}$$
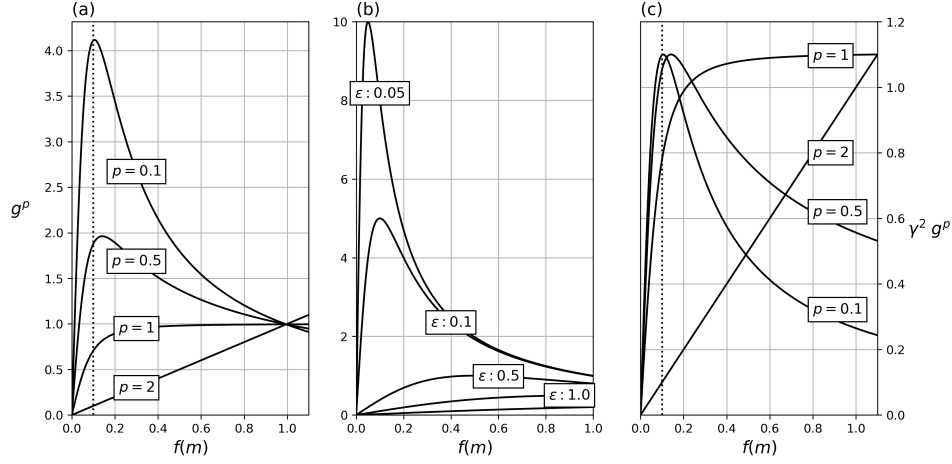
48

Figure 3.10: Derivatives of the Lawson approximation over a range of model values for (a) a fixed threshold parameter $\varepsilon = 10^{-1}$ over a range of p values and for (b) a fixed $p = 0$ over a range of ε values. (c) Applying the γ-scaling to the gradients brings all maximums to be equal irrespective of p and ε.

The role γ_s is to reference the partial derivatives of the approximated ℓ_p-norms to the derivatives of its ℓ_2-norm measures. This re-scaling is done for two reasons. First, at the transition between Stage 1 and 2, it preserves the balance between the misfit and regularization terms and thus no large adjustment in the trade-off parameter β is needed. Secondly, this ensures that proportionality between ϕ_s^0 and ϕ_x^2 is preserved during the ℓ_p-norm inversion.

Two options are possible to compute the γ-scalings: (a) take the maximum absolute gradient directly from the gradient values in (3.34), or (b) calculate $\left\| \mathbf{g}_j^p \right\|_\infty$ analytically from (3.36). I have found that Option 2 is more stable since it is based upon a theoretical maximum of the gradient and not on a particular realization of that maximum that arises from the distribution of values in the current model $\mathbf{m}^k$.

The outcome of the re-scaling strategy is shown in Figure 3.11(a). The solution seems to have my desired properties of being sparse in terms of the number of non-zero model values and the model has smooth edges. The maximum partial derivatives, shown in Figure 3.11(c), confirm that the scaling strategy was successful in balancing the impact of the two components of the regularization. This is quantified by the calculated proportionality ratio $\lambda_\infty = 0.7$. It is an improvement
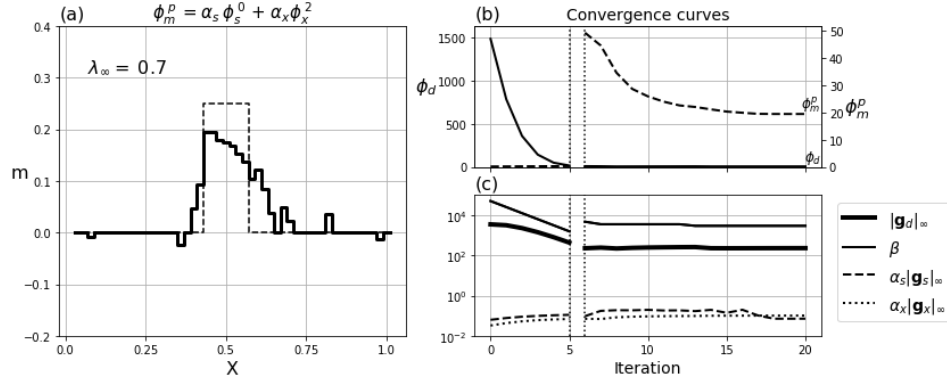
Figure 3.11: (a) Recovered model and (b) convergence curves using the Scaled-IRLS approach for $p_s = 0$, $p_x = 2$ and a fixed threshold parameter $\varepsilon = 1e - 3$ ($\alpha_s = \alpha_x = 1$). (c) Trade-off parameter and maximum gradients for the different components of the objective function. The scaling procedures preserves the proportionality between $\left\| \mathbf{g}_s^{p_s} \right\|_\infty$ and $\left\| \mathbf{g}_x^{p_x} \right\|_\infty$ throughout the iteration process. The trade-off β-parameter needed only to be adjusted slightly at the beginning of Stage 2.

over the previous solution with a ratio of 150 (Figure 3.9), but it appears that the algorithm has reached a steady state solution with slightly more influenced from ϕ_s. In the following section I provide a strategy to better preserve proportionality between each model update through a cooling strategy.

Scaled model derivatives

Applying the same scaling strategy to the model derivative requires additional care. Directly implementing the γ-scaling as in (3.38) to ϕ_x^p I would obtain that

$$\gamma_x \approx \left[h^{p-2} \frac{\left\| \hat{\mathbf{g}}_x^2 \right\|_\infty}{\left\| \hat{\mathbf{g}}_x^{p_s} \right\|_\infty} \right]^{1/2}. \tag{3.39}$$

where $\hat{\mathbf{g}}_x^p$ denotes the normalized model differences that would arise from taking the gradient of (3.31). This multiplying factor h^{p-2} is closely related to the α_s scaling strategy specified in (3.32). Both have the goal of scaling the components of the objective function in terms of the discretized length. As I further generalize

the IRLS inversion to allow for mixed-norm measured that vary on a cell by cell and in a 3D grid, this multiple level of scaling can rapidly becomes cumbersome. In order to simplify the implementation, I am proposing to measure the model gradients on a normalized grid such that

$$\mathbf{D}_x = \begin{bmatrix} \hat{h}_1^{-1} & -\hat{h}_1^{-1} & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & \hat{h}_{M-1}^{-1} & -\hat{h}_{M-1}^{-1} \end{bmatrix}, \qquad (3.40)$$

where $\hat{\mathbf{h}} = \mathbf{h}/min(\mathbf{h})$. Normalizing the length scales to unity brings both $\phi_s^{p_s}$ and $\phi_x^{p_x}$ to be dimensionally equivalent. The scaled IRLS weights become

$$\mathbf{R}_x = \gamma_x \, \text{diag}\left[\left(\left(\mathbf{D}_x \mathbf{m}^{(k-1)}\right)^2 + \varepsilon^2\right)^{p_x/2-1}\right]^{1/2}, \qquad (3.41)$$

where

$$\gamma_x = \left[\frac{\|\hat{\mathbf{g}}_x^2\|_\infty}{\|\hat{\mathbf{g}}_x^{p_x}\|_\infty}\right]^{1/2}. \qquad (3.42)$$

In order to demonstrate that this change in scale does not change the overall objective function, I proceed with two inversions. I repeat the experiment presented in Section 3.1.3 for $p_s = p_x = 1$, also shown in Figure 3.12(a) for comparison. First, I use the normalized gradient measure with uniform scaling ($\alpha_s = \alpha_x = 1$). The recovered model shown in Figure 3.12(b) is almost identical to the previous solution. Small discrepancies between the two models can be attributed to slight differences in the iterative process. As presented in Table 3.3, the global objective function $\phi(m)$ remains unchanged with both approaches. Changes in scale between individual components of the objective function (ϕ_d, ϕ_s^p, ϕ_x^p) are absorbed by their respective hyper-parameter (β, α_s, α_x).

The second experiment tests the case of a non-uniform discretization. Using the same noisy data, I refine the mesh on the right-half of the domain by dividing the cell size by a factor 2 ($h_i = 1/100$). I also re-adjust the kernel functions F_{ij} within the refined region such that each random coefficients are sampled twice but over a smaller cell length. The final inversion mesh contains 75 parameters, compared
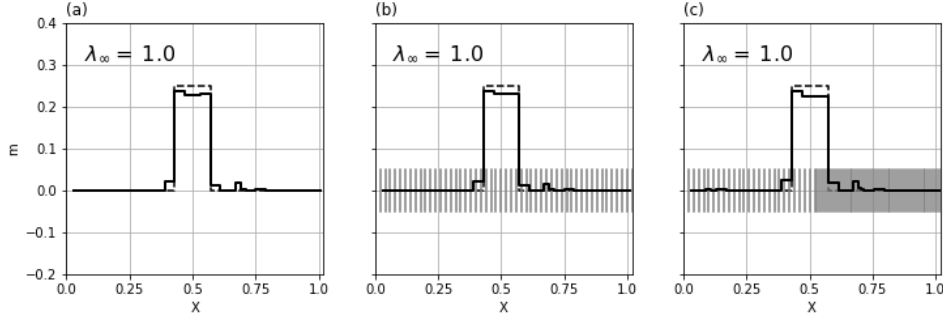
Figure 3.12: Recovered 1D models for $p_s = p_x = 1$ using different scaling strategies and parameterization. (a) Solution previously shown in Figure 3.7 that uses the standard gradient measure (α_s=50). (b) Solution obtained with the normalized gradient measure ($\alpha_s = \alpha_x = 1$). (c) Model recovered with a different parameterization such that the right half of the domain has cells with half the size. Small discrepancies between the three solutions can be attributed to slight differences in the iterative process.

ϕ_d	β	α_s	α_x	$\beta\alpha_s\phi_s$	$\beta\alpha_x\phi_x$
4.87	65.13	50.0	1.0	56.18	8.60
5.13	3381.87	1.0	1.0	58.33	8.81
4.93	3236.60	1.0	1.0	55.81	8.59

Table 3.3: Components of the objective function corresponding to the inversion results presented in Figure 3.12 for $p_s = p_x = 1$.

to 50 parameters in the previous experiment. After convergence of the algorithm I recover the model presented in Figure 3.12(c). Once again, the final model and the calculated components of the objective function (Table 3.3) are almost identical to the previous experiments. This demonstrates that the normalization of the model derivatives does not change the global objective function, and that the solution remains independent on the choice of discretization.

3.2.2 Threshold ε-parameter

While I have improved the flexibility of the IRLS algorithm, I have yet to address the threshold ε-parameter which has been held fixed. The choice of threshold parameters remains a subject of disagreement among researchers. In the early