Geophysical Journal International

Geophys. J. Int. (2025) **00,** 1–27
GJI Seismology

https://doi.org/10.1093/gji/ggaf131

Bridging the gap between SOLA and deterministic linear inferences in the context of seismic tomography

Adrian M. Mag [1], Christophe Zaroli [2] and Paula Koelemeijer [1]

[1] *Department of Earth Sciences, University of Oxford, Oxford, United Kingdom. E-mail: marin.mag@stx.ox.ac.uk*
[2] *Institut Terre et Environnement de Strasbourg, Université de Strasbourg, EOST, CNRS, UMR 7063, 5 rue Descartes, Strasbourg F-67084, France*

Accepted 2025 March 28. Received 2025 March 26; in original form 2024 July 12

SUMMARY

Seismic tomography is routinely used to image the Earth's interior using seismic data. However, in practice, data limitations lead to discretized inversions or the use of regularizations, which complicates tomographic model interpretations. In contrast, Backus–Gilbert inference methods make it possible to infer properties of the true Earth, providing useful insights into the internal structure of our planet. Two related branches of inference methods have been developed–the Subtractive Optimally Localized Averages (SOLA) method and Deterministic Linear Inference (DLI) approaches—each with their own advantages and limitations. In this contribution, we show how the two branches can be combined to derive a new framework for inference, which we refer to as SOLA-DLI. SOLA-DLI retains the advantages of both branches: it enables us to interpret results through the target kernels, rather than the imperfect resolving kernels, while also using the resolving kernels to inform us on trade-offs between physical parameters. We therefore highlight the importance and benefits of a more careful consideration of the target kernels. This also allows us to build families of models, rather than just constraining properties, using these inference methods. We illustrate the advantages of SOLA-DLI using three case studies, assuming error-free data at present. In the first, we illustrate how properties such as different local averages and gradients can be obtained, including associated bounds on these properties and resolution information. Our second case study shows how resolution analysis and trade-offs between physical parameters can be analysed using SOLA-DLI, even when no data values or errors are available. Using our final case study, we demonstrate that SOLA-DLI can be utilized to obtain bounds on the coefficients of basis function expansions, which leads to discretized models with specific advantages compared to classical least-squares solutions. Future work will focus on including data errors in the same framework. This publication is accompanied by a SOLA-DLI software package that allows the interested reader to reproduce our results and to utilize the method for their own research.

Key words: Inverse theory; Seismology; Seismic tomography; Surface waves and free oscillations..

1 INTRODUCTION

Seismic tomography relies on mathematical inversions (Nolet 2008; Rawlinson *et al.* 2010) to model Earth's interior from collected data. Resulting tomography models highlight persistent features, such as subducted plates, rising plumes and large scale velocity anomalies, believed to mirror real Earth characteristics (Ritsema & Lekić 2020). Improving these models often involves the development of new models with different data or methods to enhance the resolution of certain features or to reduce uncertainties. However, seismic inversions encounter a major challenge: data scarcity. This leads to non-uniqueness in solutions (e.g. Tarantola 1987), which often is mitigated using regularization. Yet, such prior information might inadvertently impose unrealistic constraints or introduce artefacts in the models (e.g. Nolet 2008; Zaroli *et al.* 2017). While incorporating such new information is not inherently wrong, it must be accurate and well-understood to avoid misinterpretations of the resulting seismic tomography models.

In contrast, inference methods aim to constrain some specific properties of the unknown model. In geophysics, these methods can be traced back to the seminal papers of Backus and Gilbert (Backus & Gilbert 1967a, b, 1970), where they attempted to obtain the highest resolution local averages of a continuous unknown model using just the data as constraints. The methodology introduced in Backus & Gilbert (1970) has since been used in various branches of geophysics, for example, deconvolution (Oldenburg

1981), geomagnetism (Backus 1988a), seismology (Zaroli 2016). Furthermore, this foundational work has inspired the development of two branches of linear inference methods (see Fig. A1).

The first branch started with three contributions from Backus (Backus 1970a, b, c) where the goal was to find a specific linear property of the unknown model, rather than just the highest resolution local average. Backus showed that data alone cannot provide any information on nearly all linear properties, and introduced an additional prior constraint on the model space in the form of a model norm bound (Backus 1970a). The use of prior model information (e.g. a norm bound) is what distinguishes this branch, which we refer to as the DLI (Deterministic Linear Inference) branch due to the deterministic nature of the norm bound prior information (as opposed to a probabilistic prior information). Backus (1970b) and Backus (1970c) further investigated how to deal with properties that cannot be naturally expressed on Hilbert spaces, and how to approach situations where we have bounds on the norm of a truncated expansion of the model rather than the model itself. Parker (1977) re-derived the findings of Backus (1970a) in a modified framework, defining a finite-dimensional 'property–data' space separate from the infinite–dimensional model space. This approach was then used to constrain the coefficients of a basis expansion for an unknown model by applying data constraints along with a model norm bound that differed from the one used by Backus (1970a). More recently, Al-Attar (2021) has placed the method of Backus (1970a) and Parker (1977) in a more general mathematical framework.

The second branch has mainly been represented by the SOLA method (Subtractive Optimally Localized Averages), developed by Pijpers & Thompson (1992, 1994), although similar methods had previously been used in deconvolution theory (Oldenburg 1981). The SOLA branch was introduced into the seismic tomography community by Zaroli (2016) and has received increasing attention in the past decade (e.g. Zaroli *et al.* 2017; Lau & Romanowicz 2021; Latallerie *et al.* 2022; Amiri *et al.* 2023; Restelli *et al.* 2024). Fundamentally, SOLA resembles the method by Backus (1970a), with the distinction that it lacks any prior model information such as the model norm bound used in the DLI branch. In the absence of additional prior constraints, it yields only an approximate local average–precisely defined by a resolving kernel R. Given an unknown model $\bar{m}$, the Backus (1970a) method finds a set of possible values for $\int T\bar{m}$ (the desired property), where T represents a predefined weight function, known as the target kernel. In contrast, SOLA provides a single value, $\int R\bar{m}$ (the approximate property), under error-free conditions, where R is similar to T. Consequently, results obtained with approaches from the DLI branch can be interpreted in terms of the target kernels, whereas results from the SOLA branch are interpreted through the resolving kernels (see Table A1).

The primary advantage of linear inference methods over inversions lies in their ability to provide detailed uncertainty and resolution analyses. However, this benefit comes at the cost of linearity; these methods are only applicable to linear problems or weakly nonlinear ones through linearization. While Snieder (1991) extended the method of Backus & Gilbert (1970) to address weakly nonlinear problems, the SOLA and DLI branches lack such generalizations. Furthermore, compared to other linear methods, linear inference techniques are not ideal candidates for iterative solvers, as they focus on extracting properties of the model rather than constructing models (though, as we will show later, it is possible to build discretized models as well). In nonlinear problems, typically a nonlinear method (such as Bayesian inversion) or an iterative solver is used to arrive at a model that is considered relatively close to the true model. The employed methods must have convergent properties, but they do not necessarily have the ability to provide resolution and uncertainty information. Linear inferences could then be used as a final step to provide comprehensive uncertainty and resolution analysis.

We propose that the combination of the two methodological branches offers a more comprehensive base framework for geophysical inferences. By framing the interpretation in terms of target kernels, as is implicitly done in the DLI branch, we ensure the results are easily and consistently interpretable, which is particularly important for specific applications, such as determining relationships between seismic velocities. If the interpretation is to be placed on the target kernels, then we argue that more care should be taken when designing the target kernels. However, the impact of target kernel selection has not been directly studied in the SOLA branch; simple target kernels, such as boxcar and Gaussian functions, have typically been chosen for their ease of use (e.g. Zaroli *et al.* 2017; Restelli *et al.* 2024). A more careful consideration of the target kernels not only ensures that the advantage of easier interpretability is not lost, but it can also lead to tighter property bounds, as we will demonstrate.

Although the DLI branch of methods does not require the explicit use of resolving kernels, these kernels are central in the SOLA branch. We will demonstrate that, with a slight modification of the SOLA approach, the resolving kernels can also be seen as an implicit component of approaches in the DLI branch. Even within DLI methods, we can thus use these resolving kernels to obtain additional insights into spatial trade-offs and contamination from other physical parameters. In addition, if the interpretation is placed on the target kernels, it is possible to use inference methods to obtain discretized models, rather than just properties of models.

By combining the two branches, we obtain in essence a deterministic linear inference method, similar to Al-Attar (2021), but with modified property bounds and a direct incorporation of resolving kernels (an idea stemming from SOLA) into the analysis. Therefore, this combination should be regarded as a 'SOLA-infused deterministic linear inference' method, which we will refer to as 'SOLA-DLI'.

In this contribution, we do not consider noise in the data. However, this does not mean that the data are perfect. Even a noise-free data set is not 'perfect' if it lacks enough information to fully constrain the model space to a single solution, that is, it is incomplete. As Backus & Gilbert (1967a) demonstrated, an infinite-dimensional model space requires an infinite number of independent data to provide complete constraints. Both branches of linear inference methods discussed earlier can address data noise and incompleteness, but they take fundamentally different approaches to handling incompleteness. Approaches from the DLI branch integrate incompleteness errors into the property bounds, while the SOLA branch captures these errors in the resolving kernels. Notably, the treatment of data noise varies even within each branch (e.g. Backus 1970a outlines two distinct approaches for handling them).

As the two methods we aim to combine differ fundamentally in how they address incompleteness in the data, we focus in this work exclusively on errors arising from data incompleteness rather than data noise. As a result, the framework we present is not immediately ready for most real-world applications, but it forms a foundation for future developments where data noise will be incorporated. In the mean time, the theory already has potential practical applications even without data noise considerations. For example, in design optimization problems, where data have yet to be measured, it is the 'geometry of the data set' that drives the optimization problem.

This is exactly an element that can be addressed with the theory presented here.

The remainder of this paper is organized as follows: Section 2 first explores the relationship between the DLI and SOLA branches, before combining them into the joint SOLA-DLI framework. In addition, we discuss the use of target and resolving kernels in SOLA-DLI, specifically, showing how the choice of target kernels influences the types of properties that can be constrained and how some may be better constrained than others. It further develops the theoretical framework for cases involving multiple physical parameters, explaining the roles and interpretations of resolving and contaminant kernels within this context. Additionally, it demonstrates how families of models can be derived using DLI-based inference methods. Section 3 presents three practical examples with synthetic, noise-free data to illustrate the theoretical concepts introduced in Section 2. Finally, Sections 4 and 5 provide a discussion and conclusion, respectively. Appendix A provides a general perspective on inference methods, while appendices B, C, D provide supplementary mathematical derivations that offer additional detail to support Section 2.

2 THEORY

In this section, we will mathematically describe the link between the two branches of linear inferences, taking the DLI branch as starting point and subsequently introducing elements from the SOLA branch to establish the SOLA-DLI framework. We also examine how the choice of target kernels influences the inference outcomes and discuss key considerations necessary for ensuring the correct interpretation of results. Additionally, we demonstrate how resolving kernels can be employed to analyse trade-offs between physical parameters and how the model norm bound can serve to estimate these trade-offs, offering a potential alternative to the 3-D noise method employed so far (Masters & Gubbins 2003; Restelli *et al.* 2024). Finally, we evaluate the strengths and limitations of linear inference methods for deriving discretized models, highlighting why 'models' obtained with SOLA should strictly speaking be considered proxies to a model, rather than an actual model.

Throughout the paper, we will adopt a modern mathematical notation similar to Al-Attar (2021), which is applicable to both Banach and Hilbert spaces. This operator-based formalism is particularly well-suited for comparing and combining the SOLA and DLI methods, as it more readily clarifies the connections between these approaches.

2.1 Combining the DLI and SOLA branches into SOLA-DLI

2.1.1 Deterministic linear inferences

Let d be some error-free data, m a model and G a linear forward operator. We can express the model-data relationship as follows:

$$G(m) = d. \tag{1}$$

We refer to such model-data relationships as 'deterministic data constraints', assuming the data are known exactly (no data noise). The model belongs to a model space $\mathcal{M}$, while the data reside in a data space $\mathcal{D}$. In inversions, we aim to find the model solution from the data by inverting the forward relation (eq. 1). However, in most cases, the forward relation cannot be inverted due to insufficient or inadequate data. For continuous models, this scenario can result in either no solutions or infinitely many solutions (Backus & Gilbert 1967b). In the absence of data noise, no solutions occur only when the data lie outside the range of the forward operator, making them incompatible with the physical laws governing the system. Typically, in such situations, we would employ a different forward relation. Throughout this paper, we will assume that for any data d, there exists at least one model such that $d = G(m)$ (in other words, G is surjective), leading to an infinite set of solutions, denoted by S (see Fig. 1a). Inversions can then be conducted by imposing constraints (regularizations) on the model space $\mathcal{M}$ until a single model, $\tilde{m}$, is 'selected'. For instance, one might choose the model with the smallest average gradient (the flattest model), or with the smallest norm. However, if the implicit assumptions of the chosen regularization are incorrect, the resulting model may not accurately represent reality.

We often seek specific properties of the true model $\bar{m}$ rather than the entire model itself. These properties, for example the average structure over some volume within the Earth or the depths of discontinuities, belong to a distinct space known as the property space $\mathcal{P}$, following the work of Al-Attar (2021). Therefore, we can define a different (inference) problem as:

Given that:

$$G(\bar{m}) = d \tag{2}$$

Find:

$$\mathcal{T}(\bar{m}) = \bar{p} \tag{3}$$

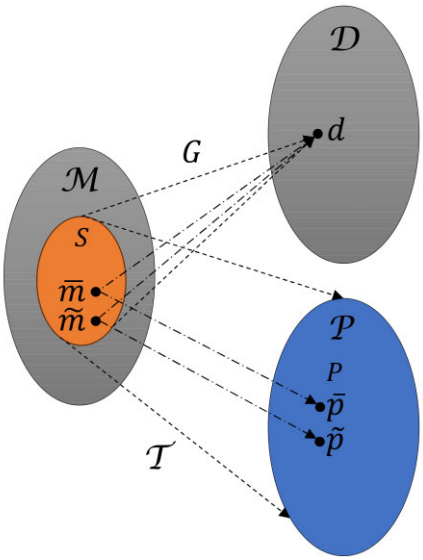
where $\mathcal{T}$ (the property mapping) is a linear relation that extracts a property of any model, and $\bar{p} \in \mathcal{P}$ represents the value extracted by $\mathcal{T}$ when applied to the true model $\bar{m}$. It can be shown that in most practical situations, the desired property $\bar{p}$ can take any value given a finite number of deterministic data constraints (Backus 1970a; Al-Attar 2021). In other words, given the data constraints, $\bar{p}$ may take any value from the property space $\mathcal{P}$ (see Fig. 1a), leaving us unable to definitively determine the property of the true model $\bar{m}$. Backus (1970a) demonstrated that this issue can be overcome by introducing a norm bound M in the model space:

$$\|m\|_{\mathcal{M}} \leq M. \tag{4}$$

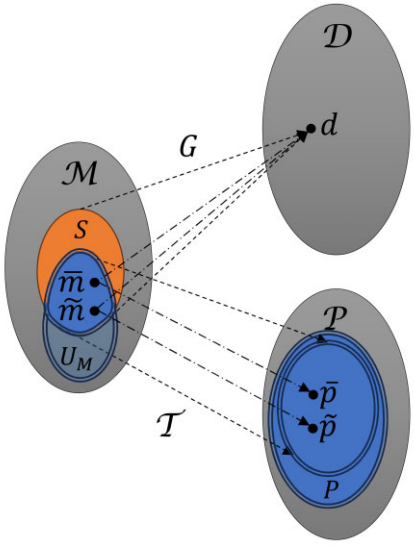
This constraint on the model space differs from constraints typically imposed during regularization because it does not aim to isolate a single model. Instead, the model norm bound restricts solutions to a bounded subset of $\mathcal{M}$. If the set of models satisfying the norm bound is denoted by U_M (eq. 4), then the set of solutions respecting both the norm bound and the data constraint is $U_M \cap S$, which is a bounded subset (Al-Attar 2021). Al-Attar (2021) further showed that this constraint results in the true property $\bar{p}$ being confined within a bounded subset $P \subset \mathcal{P}$, provided that the norm of the true model is less than the chosen norm bound (see Fig. 1b for a visual representation of these concepts). Note that the subset P is not a sharp bound on the values of the true property, meaning that $\mathcal{T}(U_M \bigcap S) \subseteq P$ (theoretically, better approximations are possible). Without additional prior information, all properties in P are equally likely to represent the true property $\bar{p}$.

If the model space $\mathcal{M}$, data space $\mathcal{D}$, and property space P are Hilbert spaces, and the forward and property mappings $G, \mathcal{T}$ are continuous linear mappings with G being surjective, then the solution to eq. (3) with data and model norm bound constraints (eqs 1 and 4) is given by (Al-Attar 2021):

$$\bar{p} \in \{p \in \mathcal{P} | \left\langle \mathcal{H}^{-1}(p - \tilde{p}), p - \tilde{p} \right\rangle \leq M^2 - \|\tilde{m}\|^2_{\mathcal{M}}\}, \tag{5}$$



(a) No norm bound applied on the model space.



(b) With a norm bound applied on the model space.

Figure 1. Schematic of general linear inference problems, illustrating the effect of bounding the model space (see Table A2 for symbol definitions). Sets with thin lines for margins represent unbounded sets, while sets with double line margins represent bounded sets. The true model is denoted by $\bar{m}$ and the least norm model solution is denoted by $\tilde{m}$. (a) When no bounds are imposed, the property of a model that respects the data constraint may take any value in the property space, which is an unbounded set (see Al-Attar 2021, Theorem 2.2). In other words, $\mathcal{T}(S) = P = \mathcal{P}$. (b) Applying the norm bound on the model space leads to the intersection between S and U_M to be bounded, which gets mapped under $\mathcal{T}$ to a bounded subset of $\mathcal{P}$. We thus have the following relation: $\mathcal{T}(U_M \bigcap S) \subseteq P \subset \mathcal{P}$.

where $\tilde{m}$ is the least norm solution to the data constraint (eq. 1), and $\tilde{p} = \mathcal{T}(\tilde{m})$ is the desired property of the least norm solution. This result implies that the true property $\bar{p}$ lies within a hyperellipsoid defined by the inequality in eq. (5). The shape of this hyperellipsoid is determined by the operator $\mathcal{H}$, which is given by Al-Attar (2021):

$$\mathcal{H} = \mathcal{T}\mathcal{T}^* - \mathcal{T}G^*(GG^*)^{-1}G\mathcal{T}^*. \tag{6}$$

It can further be shown that GG^* is invertible (Al-Attar 2021), if the model and data space are Hilbert spaces, and if G is a continuous linear and surjective mapping (we provide proof of its surjectivity in Appendix B).

2.1.2 *Considerations on the model norm bound*

The true model properties lie within the property bounds only if the norm of the true model is smaller than or equal to the norm bound. Therefore, it is crucial to choose a conservative norm bound in order to minimize the risk of inferring incorrect information about the true model. Typically, we select the norm bound to be greater than the norm of the least norm solution (which we will simply refer to as the least norm from here on).

Higher norm bounds result in larger property bounds, which in turn reduces the inference power. If it is not possible to justify a sufficiently small norm bound that results in meaningful property bounds, then a least norm regularization should not be used either. At first glance, the norm bound might appear to be a stringent constraint that is difficult to justify physically. In contrast, least norm regularization, commonly used in inversion problems, simply assumes that 'the true model should have a small norm'. However, such regularization essentially selects a single model, which is actually more stringent than the inequality constraint used in inference methods.

Deriving a norm bound directly from physical arguments is often challenging, but it is possible in specific cases. For example, when modelling the Earth's magnetic field using spherical harmonic coefficients, physical constraints like power dissipation can provide a norm bound. In seismology, point-wise upper bounds on properties are often more accessible. From these bounds, a model norm bound can be derived by constructing a piecewise function $\bar{m}(x) \leq b(x)$, leading to:

$$\int_\Omega \bar{m}^2 \, \mathrm{d}\Omega \leq \int_\Omega b^2 \, \mathrm{d}\Omega = M^2.$$

While this approach transforms point-wise bounds into an L_2 norm bound, it often overestimates the bounds, resulting in overly large property bounds. A more precise alternative would be the supremum norm:

$$\|m\|_\infty = \sup_{x \in \Omega} |m(x)|.$$

Alternatively, bounds based on model regularity (e.g. smoothness constraints) may be adopted, but they involve Sobolev spaces that account for derivatives. Although such approaches are more rigorous, they are mathematically complex and not fully developed, as discussed in Al-Attar (2021). Here, we use the L_2 norm for simplicity as the exploration of alternative norms is beyond the scope of our work, but we acknowledge that other norms may be better suited.

2.1.3 SOLA and resolving kernels

We can specificize the DLI inference problem to obtain the theory of (Backus 1970a) by assuming the following form for the data and property mapping:

$$[G(m)]_i = \langle K_i, m\rangle_{\mathcal{M}} \tag{7}$$

$$[\mathcal{T}(m)]_k = \langle T^{(k)}, m\rangle_{\mathcal{M}}, \tag{8}$$

where $\langle\cdot,\cdot\rangle_{\mathcal{M}}$ denotes the model space inner product, $K_i \in \mathcal{M}$ are data sensitivity kernels, and $T^{(k)} \in \mathcal{M}$ are target kernels. Then, eqs (5) and (6) correspond to the solution of Backus (1970a, eq. 4). We note that only the target kernels $T^{(k)}$ and sensitivity kernels K_i appear in this formulation; resolving kernels are neither required nor used.

Resolving kernels do play a role in SOLA-type linear inferences, where the problem described by eqs (2) and (3) is 'solved' without incorporating prior information on the model norm bound. However, the SOLA framework usually assumes an even more specific form for the data and property mapping than that shown in eqs (7) and (8), namely:

$$[G(m)]_i = \textstyle\int_\Omega K_i m \mathrm{d}\Omega, \tag{9}$$

$$[\mathcal{T}(m)]_k = \textstyle\int_\Omega T^{(k)} m \mathrm{d}\Omega, \tag{10}$$

which corresponds to choosing the model space inner product to be

$$\langle f, g\rangle_{\mathcal{M}} = \int_\Omega fg \mathrm{d}\Omega.$$

and the model space to be some corresponding function space, such as $L^2[\Omega]$, where Ω is some spatial domain. This choice of inner product, which is normally implicitly assumed in SOLA-type inference problems, has implications for the norm bound and its effectiveness. In particular, using this inner product leads to the following norm:

$$\|m\|_{\mathcal{M}} = \int_\Omega f^2 \mathrm{d}\Omega. \tag{11}$$

The goal of SOLA is to find some real weights $x_i^{(k)}$ such that (Zaroli 2019):

$$\int_\Omega T^{(k)} \bar{m} \mathrm{d}\Omega \approx \int_\Omega \sum_i^{N_d} x_i^{(k)} K_i \bar{m} \mathrm{d}\Omega \tag{12}$$

where N_d is the number of sensitivity kernels and the dimension of the data space $\mathcal{D} = \mathbb{R}^{N_d}$. The resolving kernels are then defined to be:

$$R^{(k)} = \sum_i^{N_d} x_i^{(k)} K_i \tag{13}$$

The resulting SOLA solution is given by:

$$\operatorname{argmin}_{x_i^{(k)}} \left[\int_\Omega (T^{(k)} - \sum_i^{N_d} x_i^{(k)} K_i)^2 \mathrm{d}\Omega \right] \tag{14}$$

$$\text{s.t.} \int_\Omega R^{(k)} \mathrm{d}\Omega = 1. \tag{15}$$

This constrained optimization problem will provide the weights that will produce unimodular resolving kernels similar to the target kernels. Subsequently, the true properties can be approximated by:

$$[\mathcal{T}(\bar{m})]_k \approx [\mathcal{R}(\bar{m})]_k = \int_\Omega R^{(k)} \bar{m} \mathrm{d}\Omega. \tag{16}$$

If we drop the unimodularity condition (eq. 15), then it can be shown (Appendix C) that the solution to the optimization problem in eq. (14) is the same as the solution to:

$$\int_\Omega T^{(k)} K_j \mathrm{d}\Omega - \sum_i^{N_d} \left(\int_\Omega K_j K_i \mathrm{d}\Omega \right) x_i^{(k)} = 0. \tag{17}$$

It is obvious from eqs (9) and (10) that:

$$\int_\Omega T^{(k)} K_j \mathrm{d}\Omega = [\mathcal{T} G^*]_{kj}, \tag{18}$$

$$\int_\Omega K_j K_i \mathrm{d}\Omega = [G G^*]_{ji}. \tag{19}$$

If we further define $X_{ki} = x_i^{(k)}$, then we find:

$$X = \mathcal{T} G^* (G G^*)^{-1}. \tag{20}$$

Therefore, $X : \mathcal{D} \to \mathcal{P}$ is a linear mapping that maps from the data space to the property space. In the absence of the unimodularity condition (eq. 15), $X(d)$ will be the solution in the SOLA framework, obtained here without the need for a model norm bound. It is also obvious that:

$$\mathcal{R} = XG = \mathcal{T} G^* (G G^*)^{-1} G \tag{21}$$

This is the same operator that can be recognized within the definition of $\mathcal{H}$ in eq. (6). This shows that resolving kernels are implicitly present in the solution of the norm-bound (DLI) branch of inference methods. In fact, we have the relation:

$$\mathcal{H} = (\mathcal{T} - \mathcal{R})(\mathcal{T} - \mathcal{R})^*, \tag{22}$$

which shows that the matrix $\mathcal{H}$ encodes the difference between resolving kernels and target kernels. This difference arises due to data incompleteness and can only be quantified in the property space when norm prior bounds on the model space are incorporated.

2.1.4 The combined SOLA-DLI framework

After establishing the link between the DLI and SOLA branches, combining the two primarily involves a practical step. This step entails calculating both the property bounds (using eq. 5) and the resolving kernels (using eq. 20), and interpreting the final result using both.

In addition, we also introduce a small modification to eq. (5). It can be shown that if eq. (5) holds, the following is also true (see Appendix D3):

$$\tilde{p}^{(k)} \in \left[\tilde{p}^{(k)} - \epsilon^{(k)}, \tilde{p}^{(k)} + \epsilon^{(k)} \right], \tag{23}$$

where

$$\epsilon^{(k)} = \sqrt{\left(M^2 - \|\tilde{m}\|_{\mathcal{M}}^2\right) \mathcal{H}_{kk}}. \tag{24}$$

$$\tilde{p} = \mathcal{T}(\tilde{m}) = \mathcal{R}(\tilde{m}). \tag{25}$$

Eq. (23) represents a hyperparallelepiped in $\mathcal{P}$ that encloses the hyperellipsoid defined by eq. (5). This provides more conservative error bounds. The rationale for this modification is twofold. First, eq. (5) requires the computation of the full matrix $\mathcal{H}$ and solving a system of linear equations involving $\mathcal{H}^{-1}$. In contrast, eq. (23) only needs the diagonal terms of $\mathcal{H}$, thus decreasing the computational cost significantly. Secondly, while eq. (5) incorporates information about the trade-offs between error bounds of different properties, this information is often difficult to visualize and interpret in practice. Conversely, eq. (23) can be easily plotted (visual explanation as

to why this is the case is given in Fig. A1), making it more practical for applications.

Historically, the DLI branch has been applied primarily to models involving a single physical parameter, whereas the SOLA approach has also been employed for cases where the data depend on multiple physical parameters simultaneously (e.g. Masters & Gubbins 2003; Restelli *et al.* 2024). This historical distinction may partly explain why resolving kernels have traditionally not been used in the DLI branch, as their utility becomes more significant when dealing with multiple physical parameters. We will further explore the increased significance of resolving kernels in the context of multiple physical parameters in Section 2.3, where we discuss the concept of contaminant kernels. However, given the importance of the target kernels, we first revisit these in the next section.

2.2 Choice of target kernels

Different information about the unknown model can be extracted by choosing appropriate target kernels, which need to be carefully designed if we interpret the results through them. To illustrate this, we introduce a simplified setup, where we assume the model to be a triplet of piece-wise continuous and bounded functions $m = (m^1, m^2, m^3)$ defined on the interval [0,1]. This leads to a 1-D inference problem, however, the results can be easily generalized. The true model is assumed to be known and is plotted in Fig. 2. This model is arbitrary and has no physical significance.

2.2.1 Local average targets

Previous studies have primarily used the box car function as a target kernel for its simplicity and ease of interpretation—it gives a uniform local average (Masters & Gubbins 2003; Restelli *et al.* 2024). However, many other types of target kernels could be used to obtain local averages. Here, we introduce three different averaging target kernels:

Uniform Local Average (for reference):

$$T_U^{(k)}(r) := \begin{cases} C & r \in V_k \\ 0 & \text{else} \end{cases} \tag{26}$$

Gaussian Local Average:

$$T_G^{(k)}(r) := C \exp\left[-\frac{\left\| r - r^k \right\|_2^2}{2\sigma^2} \right] \quad r \in \Omega \tag{27}$$

Bump Function Average:

$$T_B^{(k)}(r) := \begin{cases} C \exp\left[\frac{w^2}{2(r - r^{(k)})^2 - w^2} \right] & r \in V_k \\ 0 & \text{else} \end{cases} \tag{28}$$

where

$$V_k = \left[r^{(k)} - \frac{w}{2}, r^{(k)} + \frac{w}{2} \right] \tag{29}$$

is the compact support of the boxcar and bump function with width w, σ is the standard deviation of the Gaussian, and C is in each case an appropriate normalization constant that ensures the unimodularity of each target kernel. Examples of these averaging kernels are plotted in Fig. 3. Using these target kernels we can define, for example, the following property mappings for physical parameter m^1:

$$\tilde{p}^{1,(k)}_{U/G/B} = \mathcal{T}_{U/G/B}(m) = \int_0^1 T^{1,(k)}_{U/G/B}(r) m^1(r) \mathrm{d}r. \tag{30}$$

The property vector extracted by each such property mapping (uniform/Gaussian/bump) is a vector of local averages centred at a set of points $\{r^{(k)}\}$ that we call 'enquiry points'. In Fig. 3, we also plot the property vectors $\tilde{p}^{1,(k)}_{U/G/B}$ for 1000 evenly spaced enquiry points (right column). The grey hatched regions are parts of the domain where the target kernels $T^{1,(k)}_{U/G/B}$ are clipped and therefore their associated results uninterpretable.

Each of these target kernels has advantages and disadvantages. The boxcar function (Restelli *et al.* 2024) is simple and has compact support, providing a clear interpretation of the resolution of these kernels. However, most sensitivity kernels used in seismology are smooth, typically giving rise to poor resolving kernels that do not resemble boxcar functions. Consequently, the property error bounds are large.

Gaussian targets are often better reconstructed and thus lead to better constrained property values. They are, however, not defined on a compact domain and restricting a Gaussian to a compact domain leads to clipping. A clipped Gaussian is no longer a Gaussian, and different centring of the Gaussian leads to different clipping and therefore a 'non-uniform' interpretation of the property values. If most of the Gaussian is located well within the bounds of the model domain, then the errors introduced by clipping can be negligible, but not readily quantifiable. For such target kernels we define a 'width' that contains some large and arbitrary percentage of the function's weight (such as 90 per cent) and pretend as if the entire weight of the function is concentrated in this region.

Bump functions are both smooth and defined on a compact support, therefore offering some of the advantages of both boxcars and Gaussian targets. The one shown here is just one example of a family of functions with similar characteristics.

2.2.2 Local gradient targets

If we want to obtain some local estimate of the gradient of a 1-D physical parameter such as:

$$\tilde{p}^{l,(k)} = \int_0^1 T^{l,(k)} \frac{\mathrm{d}m^l}{\mathrm{d}r} \mathrm{d}r, \tag{31}$$

we can use integration by parts to obtain:

$$\tilde{p}^{l,(k)} = -\int_0^1 \frac{\mathrm{d}T^{l,(k)}}{\mathrm{d}r} m^l \mathrm{d}r + \left[T^{l,(k)} m^l \right]_0^1. \tag{32}$$

For target kernels $T^{l,(k)}$ with compact support eq. (32) will reduce to:

$$\tilde{p}^{l,(k)} = -\int_0^1 \frac{\mathrm{d}T^{l,(k)}}{\mathrm{d}r} m^l \mathrm{d}r \tag{33}$$

in the interval where the target kernels are not clipped. For any other target kernel, the second term will not necessarily be 0. However, it will be very close to zero if the target kernel is centred well within the bounds of the domain. Therefore, we can use eq. (33) to define new target kernels that extract local gradients of the true model.

In the bottom row of Fig. 3 we plot the derivative of a Gaussian, Bump and a Triangular function. The derivative of a Boxcar gives a sum of Dirac delta distributions. These cannot be used in our framework as they do not belong to a useful Hilbert space. Instead, we have opted for the derivative of a triangular function, which yields a Haar function. It is clear from Fig. 3(d) that the true property values obtained using the different gradient target kernels pick out the discontinuities of the true model.

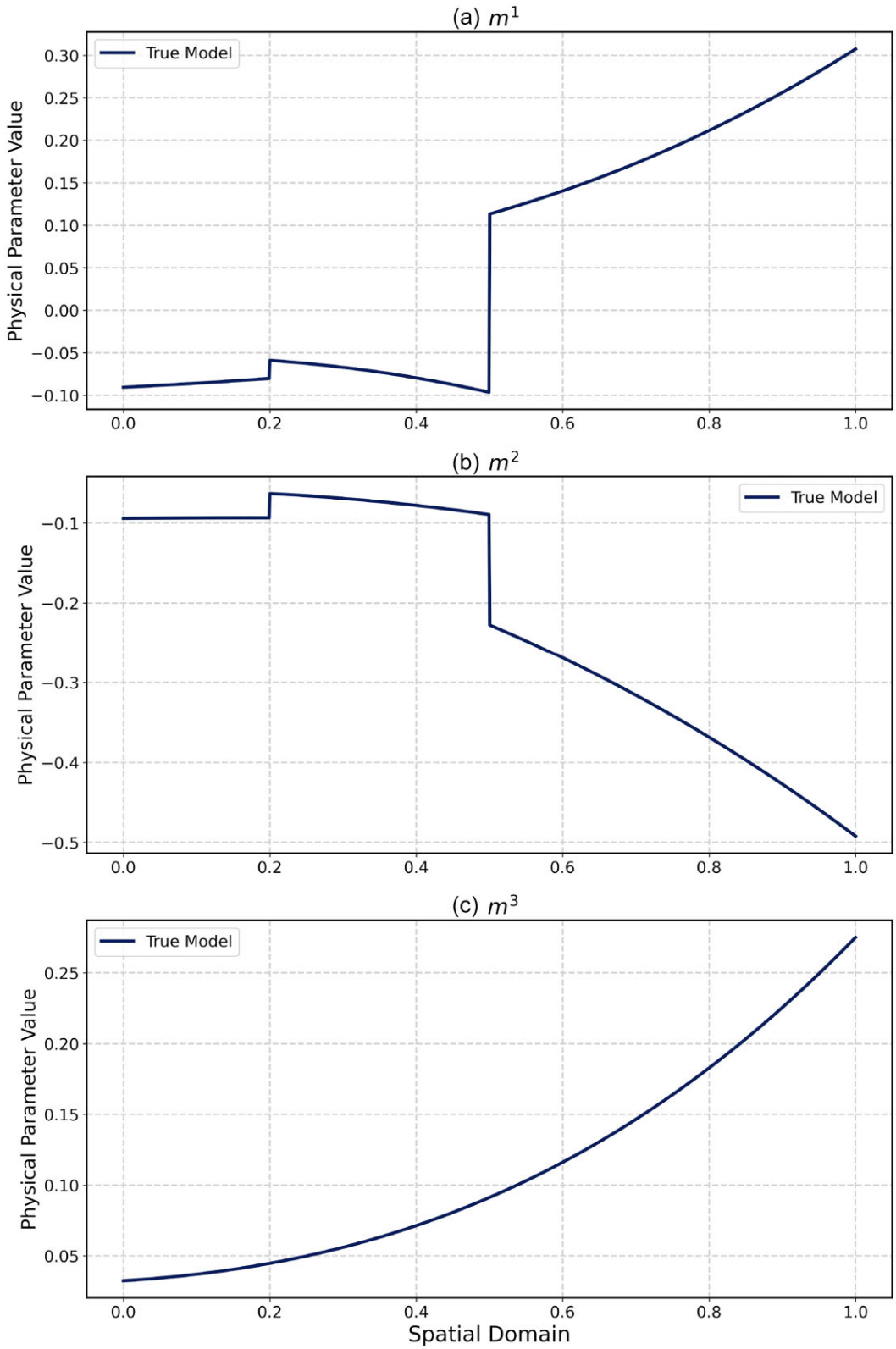


Figure 2. An arbitrary synthetic 'true model'. m^j denotes the physical parameters of the model.

The idea of using different target kernels to extract different types of information about the unknown model has been applied previously in helioseismology by Pijpers & Thompson (1994). They used Gaussian target kernels for extracting average information, and derivatives of the Gaussian to extract first and higher order derivatives of the model. However, as far as we are aware, this approach has not yet been used in seismic tomography. More importantly, the work of Pijpers & Thompson (1994) regard this approach

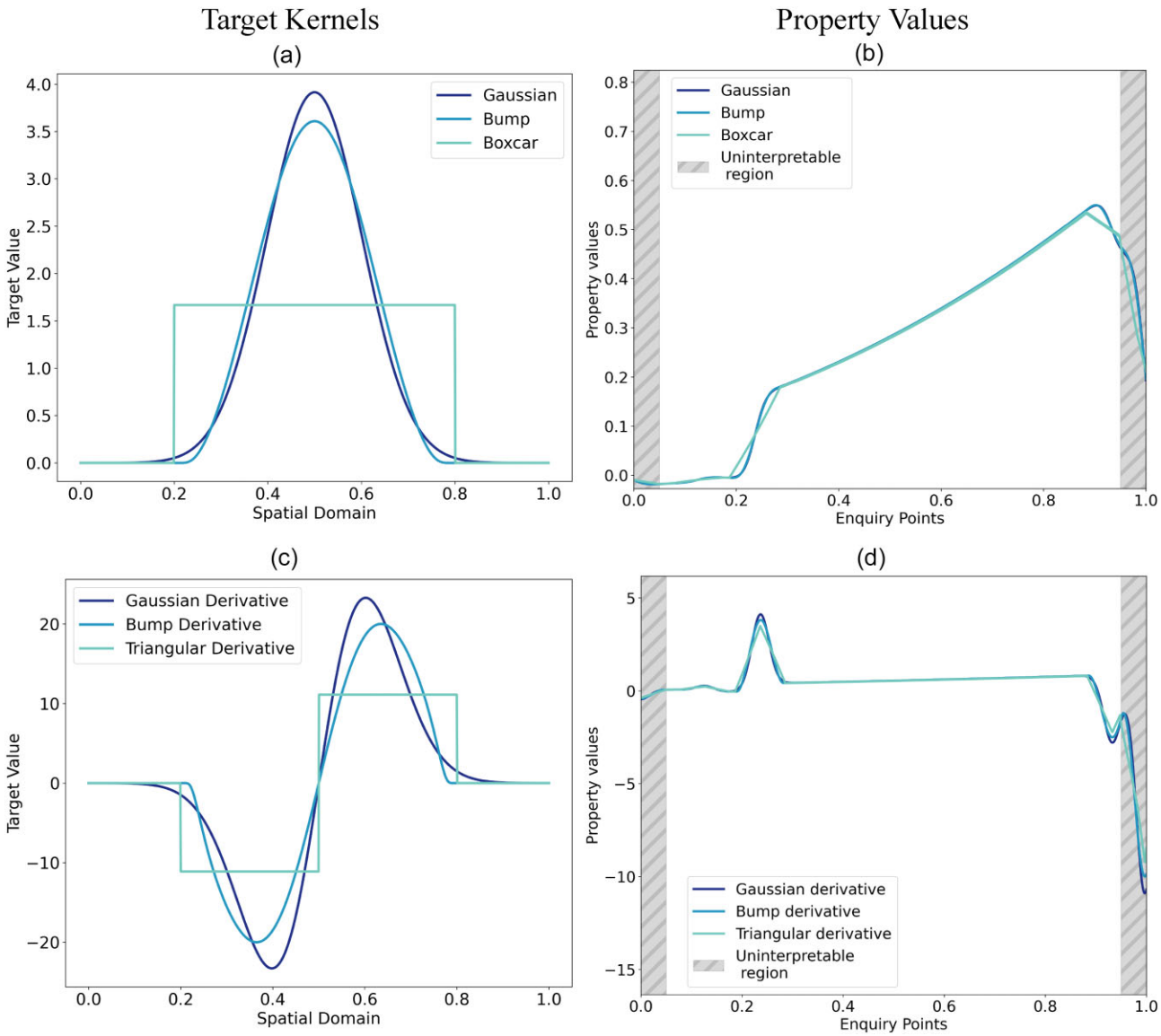


Figure 3. Examples of averaging (a) and gradient (c) target kernels and corresponding properties of model parameter m^1 (b and d) obtained using these target kernels. First column: averaging and gradient target kernels with width 0.6. Second column: property values as a function of 1000 enquiry points for different local averages and gradients of the true model using target kernels with width 0.1. The grey hatched regions represent parts of the domain where the target kernels are clipped (half-width of the target kernels). The target kernels for physical parameters m^2, m^3 are 0 since we are not interested in these parameters.

as an inversion, whereas we believe it should be considered as an inference problem instead. While Lau & Romanowicz (2021) investigated discontinuities inside the Earth using a SOLA approach, they used scalar value targets for the change across discontinuity and half-Gaussian target kernels to determine volumetric trade-offs.

2.3 Resolving and contaminant kernels

In seismology, we often analyse data that depend on multiple physical parameters, for example, compressional wave speed (v_p), shear wave speed (v_s), and density (ρ). In general, this dependence can be expressed as:

$$[G(m)]_i = \sum_j^{N_m} \int_\Omega K_i^j m^j \, \mathrm{d}\Omega, \tag{34}$$

where N_m represents the number of physical parameters. The property mapping can then be described by:

$$[\mathcal{T}(m)]_k = \sum_j^{N_m} \int_\Omega T^{j,(k)} m^j \, \mathrm{d}\Omega. \tag{35}$$

If we consider each physical parameter as residing in its own Hilbert space ($m^j \in \mathcal{M}_j$), the model space can be defined as the direct sum of these individual spaces. Consequently, a model is represented as a tuple:

$$m = (m^1, m^2, \ldots, m^{N_m}). \tag{36}$$

Furthermore, a norm bound for this composite model space can be derived from independent norm bounds applied to each physical

parameter. This is given by:

$$\|m\|_{\mathcal{M}} = \sum_{j}^{N_m} \|m^j\|_{\mathcal{M}_j}. \tag{37}$$

The solution for the property bounds is then provided by eq. (5), while the corresponding resolving kernels are expressed as:

$$R^{j,(k)} = \sum_{i}^{N_d} x_i^{(k)} K_i^j. \tag{38}$$

It is important to note that every target kernel has an associated resolving kernel. When we are interested in a specific property of the l-th physical parameter, we typically set all the target kernels associated with other physical parameters to zero. Ideally, their associated resolving kernels would then also be zero or close to zero. However, in practice, these resolving kernels are rarely zero. This discrepancy increases the property bounds, making it more difficult to constrain the desired property. This issue effectively highlights the trade-offs that exist between physical parameters, thus providing useful information when interpreting the results. Notably, if a property of the l-th parameter strongly trades off with the l'-th parameter, this will be visible in the l'-th resolving kernel.

Any resolving kernel that is non-zero when it should ideally be zero is referred to as a contaminant kernel. Lau & Romanowicz (2021) used such contaminant kernels to quantify errors arising from trade-offs within a SOLA context. By using the model norm bound in our SOLA-DLI approach, we effectively and automatically account for these trade-offs, as they are integrated into the property bounds.

2.4 Obtaining discretized models through target kernels

One perceivable downside of linear inferences, such as SOLA-DLI, is the seeming impossibility of obtaining models that cover the full spatial domain (Valentine & Sambridge 2023). We illustrate here how SOLA-DLI can in fact be used to obtain discretized models by choosing appropriate target kernels, and discuss some advantages compared to simpler, classic inversions.

Consider a model $m \in \mathcal{M}$ related to some data $d \in \mathcal{D}$ by:

$$d_i = [G(m)]_i = \langle K_i, m\rangle_{\mathcal{M}}. \tag{39}$$

A common method to remove non-uniqueness, besides regularization, is discretization. Typically, a set of orthonormal basis functions $\{B_l\} \in \mathcal{M}$ is chosen and any model in $\mathcal{M}$ is projected on the subspace formed by the span of this set, leading to a parallel $m^{\parallel}$ and perpendicular $m^{\perp}$ component of the model (i.e. $m^{\parallel}$ is the component that can be expressed with $\{B_l\}$ and $m^{\perp}$ is the residual term):

$$m = m^{\parallel} + m^{\perp}, \tag{40}$$

$$m^{\parallel} = \sum_{l} p_l B_l \;\rightarrow\; \text{projection}, \tag{41}$$

$$m^{\perp} = m - m^{\parallel}, \tag{42}$$

where p_l are the coefficients given by the projection of m onto the basis functions:

$$p_l = \langle B_l, m\rangle_{\mathcal{M}}. \tag{43}$$

We can then reformulate the initial inverse problem as:

$$\text{Find } \{p_l\} \text{s.t.}$$

$$G\left(\sum_{l} p_l B_l\right) = d_i - G(\bar{m}^{\perp}). \tag{44}$$

The data correction term $G(\bar{m}^{\perp})$ subtracts from the original data the component corresponding to the part of the true model that is not within the span of the basis functions. In real applications, this term can never be computed since we do not know the true model $\bar{m}$, nor how much of it is outside the span of $\{B_l\}$. This term is therefore typically omitted and the equation solved in practice is just given by:

$$G\left(\sum_{l} p_l B_l\right) = d_i \tag{45}$$

which, combined with eq. (39), leads to the discretized inverse problem:

$$d_i = \sum_{l} \langle K_i, B_l\rangle\, p_l \tag{46}$$

In the seismic tomography literature, the matrix $\langle K_i, B_l\rangle$ is often denoted by G. However, we will not use that notation here since we already have a distinct (but related) use of the letter G.

When the number of coefficients p_l is chosen to be smaller than the number of data, such that eq. (46) is overdetermined, it is often solved in a least square (or regularized least square) manner to produce the coefficients $\{\hat{p}_l\}$. These are systematically different from the true coefficients $\{\bar{p}_l\}$, because the correction term $G(\bar{m}^{\perp})$ is ignored. Including more data while keeping the same basis functions $\{B_l\}$ will not eliminate the systematic error caused by omitting the correction term. In order to converge to the true solution $\{\bar{p}_l\}$, one has to increase both the number of data and the number of basis functions in the expansion. Increasing the number of basis functions shrinks the space in which $\bar{m}^{\perp}$ resides and thus decreases the size of the correction term $G(\bar{m}^{\perp})$. Some methods exist to mitigate or eliminate the systematic error introduced by ignoring the data correction. Trampert & Snieder (1996), for example, refer to the effect of the uncorrected data as leakage, and offer a method of suppressing it based on soft priors. A different method of overcoming this issue is using quadratic bounds on the model space (Backus 1988a), which we will discuss in the SOLA-DLI framework.

The inverse problem discussed above (eq. 46) can be turned into an inference problem by defining the property mapping as:

$$[\mathcal{T}(m)]_l = \langle B_l, m\rangle_{\mathcal{M}} = p_l. \tag{47}$$

By also providing a model norm bound, the basis coefficients can be solved for using eq. (5).

As the number of data increases, the bounds on the coefficients decrease (assuming error-free data), and the mid value of these bounds approaches the true property. This contrasts with the behaviour of a simple least-norm solution that–without the addition of more basis functions–will converge towards a systematically incorrect answer. Therefore, framing the problem as an inference problem, and using norm bounds, we can avoid the 'leakage problem'. This is basically the same idea as that of Backus (1988a) since a norm bound is just a specific case of a quadratic prior.

This approach and other similar ones have previously been explored by several authors (e.g. Parker 1977; Al-Attar 2021), especially in the context of geomagnetic modelling problems (e.g. Backus 1988a, 1989). However, these studies used mostly spherical harmonics expansions of the model. In the spirit of Section 2.2, we note that the choice of target kernels (implicitly the basis function expansion in this case) has an impact on the size of the property bounds (here the expansion coefficients), which mirrors the idea that not all basis function expansions are created equal, some of

them being naturally better constrained by the data geometry than others.

Once the property bounds have been found, they can be sampled and mapped back to the model space using the adjoint of the property mapping, thereby producing actual models. Since this method generates a family of models rather than a single model, it is unclear which particular model should be selected if one intends to run an iterative inversion based on this method.

3 APPLICATIONS

We use three case studies to showcase the advantages and capabilities of the SOLA-DLI method introduced in the previous Section. In Case 1 (Subsection 3.1), we show the effect of the prior model norm bound and choice of target kernels on the solution, illustrating how different types of properties can be constrained. In Case 2 (Subsection 3.2), we illustrate SOLA-DLI can be utilized to perform a simple resolution and trade-off analysis, even without data errors. Finally, in Case 3 (Subsection 3.3), we demonstrate how discretized model solutions can be obtained using SOLA-DLI, comparing the results with a least-squares inversion solution.

3.1 Case 1: effect of different target kernels

In this completely synthetic case study, we show how the choice of target kernels influences the inference results. We also illustrate how the prior information and the desired resolution change the local property estimates.

3.1.1 Setup

We consider a 1-D model space containing three physical parameters m^1, m^2, m^3, all of which are piece-wise continuous functions defined on the interval [0,1]. The synthetic true model (Fig. 4) is generated quasi-randomly and has no physical meaning.

The model-data relationship for d_i with $i \in \{1, 2, .., N\}$ is given by:

$$d_i = [G(m^1, m^2, m^3)]_i = \int_0^1 K_i^1(r)m^1(r)\mathrm{d}r + \int_0^1 K_i^2(r)m^2(r)dr + \int_0^1 K_i^3(r)m^3(r)\mathrm{d}r. \quad (48)$$

For each physical parameter, the sensitivity kernels are produced quasi-randomly using the equation:

$$K_i^j(r) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(r-\mu_{i,j})^2}{2\sigma^2}\right) \sin(\omega r) \sum_q c_q (r - r_q)^2 \quad (49)$$

where $\mu_{i,j}, c_q, r_q, \omega$ are randomly generated (see Fig. 5). We choose to use $N = 150$ (e.g. 150 observations) with the sensitivity kernels computed for each physical parameter. To simulate the lack of data sensitivity to a particular region (e.g. no *S*-wave sensitivity in the Earth's outer core), we manually set the sensitivity kernels for m^2 to zero in the interval [0.5, 0.75]. The synthetic (error-free) data are then produced using eq. (48) combined with the synthetic sensitivity kernels and the synthetic true model. As target kernels we use those defined in eqs (26), (27) and (28) choosing a width of 0.2.

The least norm solution to this problem (eq. 48) is given by the Moore–Penrose right-inverse:

$$\tilde{m} = G^*(GG^*)^{-1}d \quad (50)$$

and shown in Fig. 6. This is a regularized inverse solution obtained by selecting the solution with the least norm from the set of all possible solutions. We note that in this case the least norm solution approximates the true model reasonably well, except in the regions with no sensitivity (where the solution is set equal to zero), indicating that the true model norm is very close to the least norm.

To solve the SOLA-DLI inference problem, upper bound functions b^i are chosen arbitrarily (Fig. 4) such that:

$$\left|m^j(r)\right| \leq b^j(r) \quad \forall r \in [0, 1], \quad (51)$$

which leads to the following upper bound on the model norm:

$$\left\|m^j\right\|_{\mathcal{M}^j} = \sqrt{\int_0^1 (m^j)^2\mathrm{d}r} \leq \sqrt{\int_0^1 (b^j)^2\mathrm{d}r} = M^j, \quad (52)$$

$$\|m\|_{\mathcal{M}} \leq M = M^1 + M^2 + M^3. \quad (53)$$

In real applications, the upper bound functions b^j should be chosen carefully based on physical arguments, for example using constraints from mineral physics, as discussed already in Section 2.1.2.

3.1.2 Local averages and gradients

As introduced in Section 2.2, we consider three types of local averages (uniform local averages, Gaussian averages and bump averages) and three types of locally averaged gradients (triangular averaged gradients, bump averaged gradients and Gaussian averaged gradients). In this case study, we are specifically interested in obtaining these properties for parameter m^2 given its region of no sensitivity. We evaluate the properties at 100 equally spaced inquiry points in the spatial domain, with the results plotted in Figs 7 and 8.

For each type of property, at each of the 100 enquiry points, the solution (eq. 23) provides both an upper and a lower bound. Fig. 7 shows that the uniform local average is the least constrained property, while the Gaussian average is the best constrained. This result is not surprising, given that the sensitivity kernels are Gaussians modulated by polynomial and sinusoidal functions. If the sensitivity kernels were more similar to boxcar functions, we should expect the uniform local averages to be better constrained.

Regions with no sensitivity are poorly constrained, as here the only constraint comes from the model norm bound. While it is unsurprising that some properties are better constrained than others, it is particularly notable that the property bounds for the local averages are so large that they provide little information about the true property values, even in regions with data sensitivity. Conversely, the Gaussian averages yield such tight bounds that we can be highly confident in the actual property values, assuming that the prior information is correct. These distinct differences in the ability to constrain the property are significant, as we are often interested in extracting meaningful information about the true model, rather than obtaining a specific type of average. Thus, it is important to recognize that the choice of averaging type can significantly impact how much information is obtained.

When we aim to obtain locally averaged gradients as properties, smoother target kernels again lead to significantly better property bounds (Fig. 8), similar as noted for local averages. Notice however, that while the resolving kernels look similar, the property bounds of the gradients are typically larger than for the averages (compare Figs 7 and 8).

In the absence of unimodularity conditions, the SOLA solution ('approximate property') is obtained by mapping the true model through the approximate mapping $\mathcal{R}$. Because the true model is close to the least norm model, a comparison between the true and

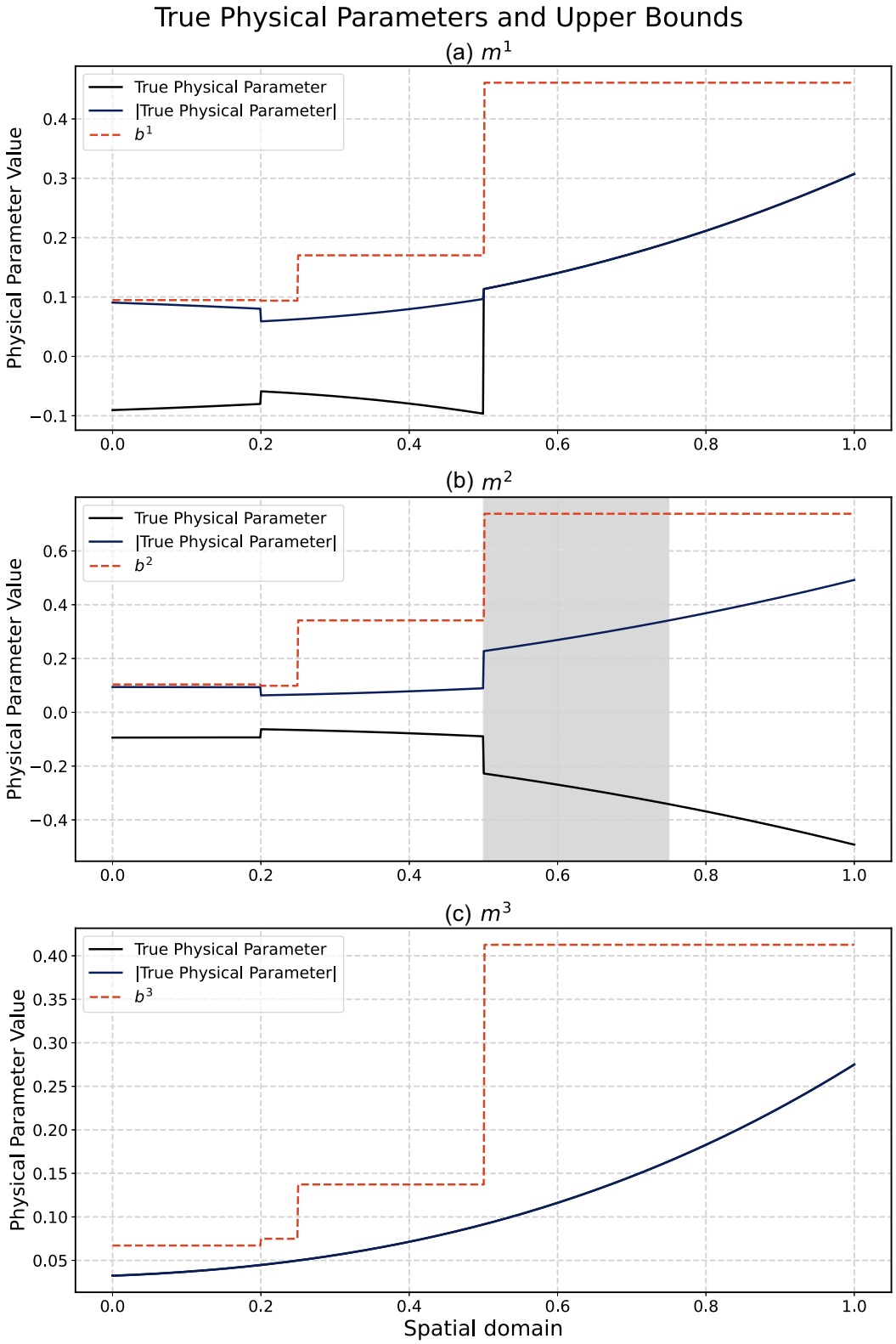


Figure 4. Case 1: True model and model norm bounds. Panels (a)–(c) show the synthetic quasi-randomly generated true model (comprised of three physical parameters) and some arbitrary piece-wise upper bound functions (b_i) used for computing the norm bound. In each panel, we present both the physical parameter (black), and the absolute value of the physical parameter (blue).

approximate property values will give the false impression that it outperforms the DLI method. However, we must remember that approximate property values do not provide us the desired information

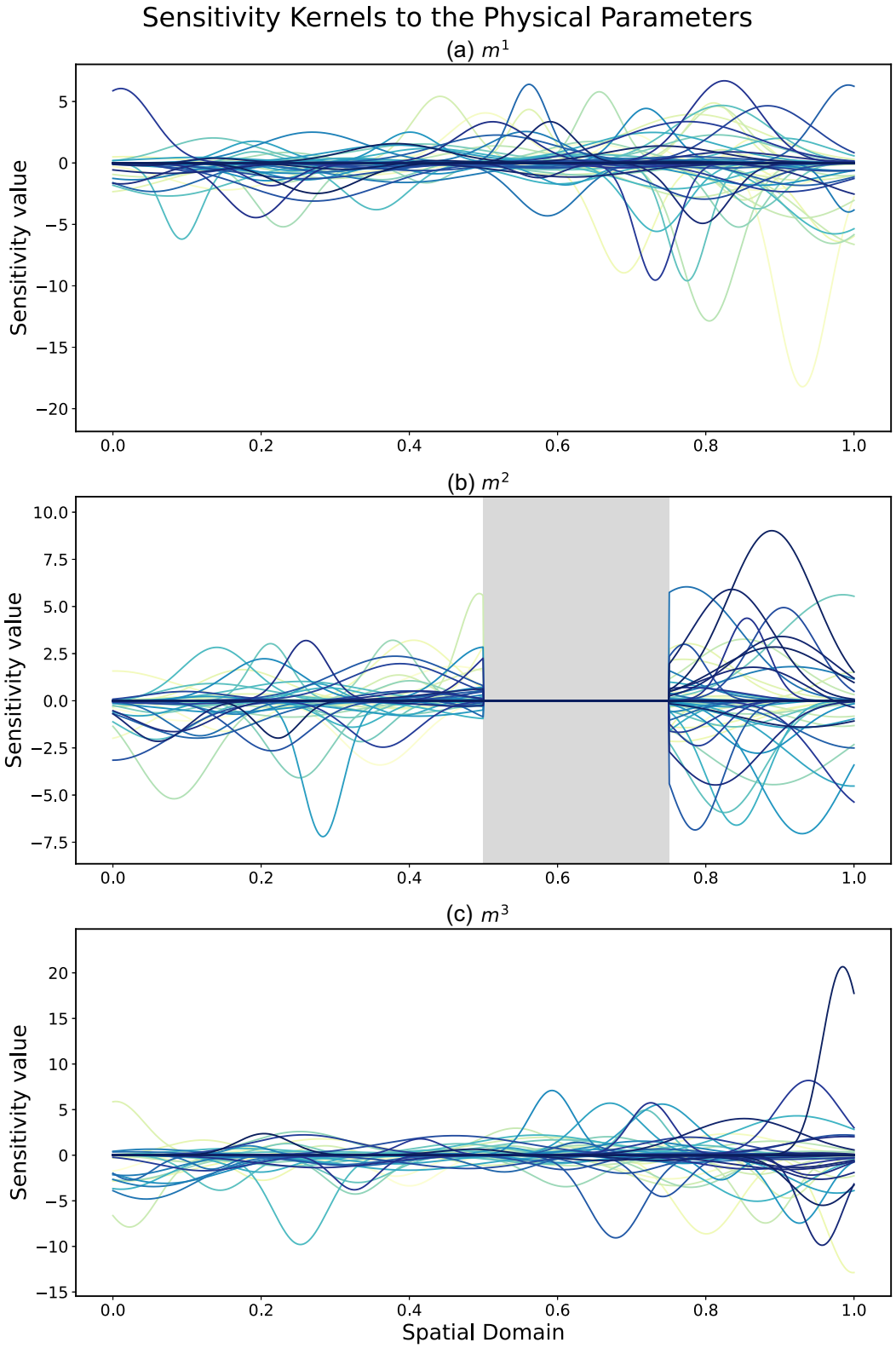


Figure 5. Case 1: Sensitivity kernels. Panels (a)–(c) show the synthetic quasi-randomly generated sensitivity kernels for physical parameters m^1, m^2, m^3. The region with no sensitivity to m^2 is shaded in grey.

about the true property values. In addition, the approximate property values must be interpreted through the resolving kernels, which can be rather different from the target kernels, and also vary in shape from one enquiry point to another. Furthermore, we believe that the SOLA method also benefits from better designed target kernels, as

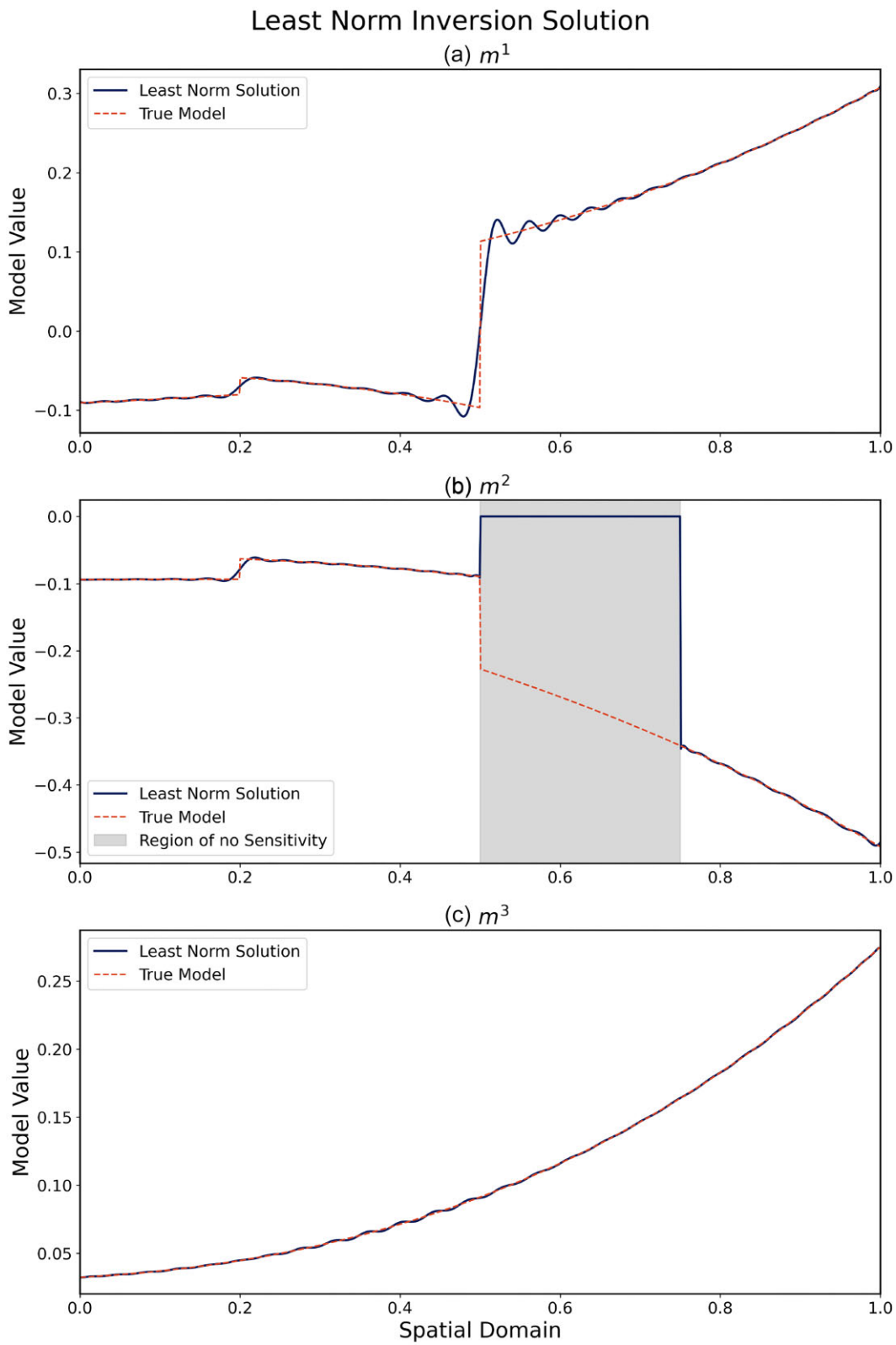


Figure 6. Case 1: least norm solution for (m^1, m^2, m^3) obtained using eq. (48).

they can lead to better resolving kernels and an easier interpretation of the results.

3.1.3 Effect of the prior model norm bound

When we change the norm bound prior information, only the property error bounds are affected (see eq. 24). This is illustrated in Fig. 9, where we show results for three different upper bounds

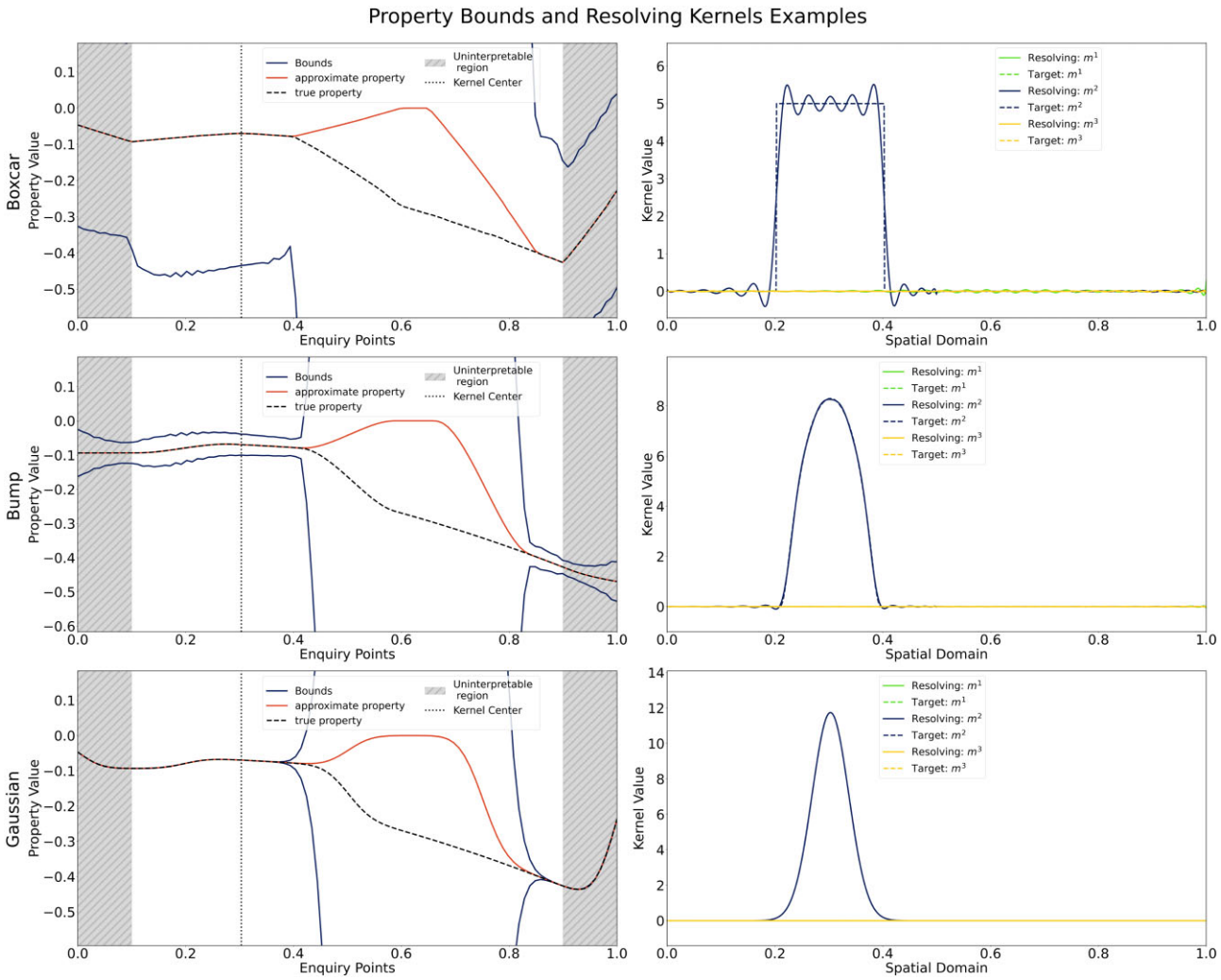


Figure 7. Case 1: SOLA-DLI solutions for three different types of local average properties. First column: solution bounds for three types of local averages of physical parameter m^2 evaluated at 100 evenly spaced enquiry points. Second column: target and resolving kernels for each type of property at the enquiry point located at $r^k = 0.3$ with width 0.2. The approximate property represents the SOLA solution in the absence of the unimodularity condition, which is mathematically just the true model mapped through the approximate mapping $\mathcal{R}$.

on the true model. Bound 3 is the most conservative, assuming a constant function three times larger than the maximum of the true physical parameter. Bounds 1 and 2 are tighter and therefore assume more prior knowledge. The bottom panel of the same figure illustrates that, as expected, tighter norm bounds lead to tighter property bounds. In all cases, the range remains centred on the approximate property. It is interesting to note that restricting the bounding function b^j in some local region does not lead to a tighter property bounds at an enquiry point in the same region, but rather it will lead to a uniform decrease of the property bounds at all enquiry points.

3.1.4 Effect of target kernel width

Changing the width of the target kernels can be interpreted as changing the resolution of the property evaluated at a given enquiry point. To investigate this, we have varied the target width between 1 and 100 per cent of the domain width and computed the relative error bounds for all the enquiry points and widths. The results are plotted in Fig. 10. The relative error bound shown in the first column is defined as:

$$e^{(k)} = \frac{\epsilon^{(k)}}{\max(\tilde{p}) - \min(\tilde{p})} \tag{54}$$

where $\tilde{p}$ is the property of the least norm model solution $\tilde{m}$. This metric has been chosen as the absolute error ϵ is not a good metric for determining whether a property is well constrained, while the classic relative error defined as $\epsilon / \bar{p}$ cannot be computed without knowing the true property $\bar{p}$. While there is no quantitative rule for what constitutes an unacceptable high relative property error bound, we believe any relative error higher than 100 per cent is 'certainly too high', and relative errors less than 10 per cent are 'generally good'.

In general, we find that for all target kernel types the relative error bounds increase when we decrease the width of the target kernel (i.e. increase resolution). In addition, regions with no sensitivity always lead to large relative errors. As expected, the width of the uninterpretable regions at the edge of the domain increases with larger target kernel width (decreasing resolution) as the half-width of the kernel increases. Finally, in this setup, we find that particular properties, e.g. Gaussian averages, are constrained better (i.e. lower

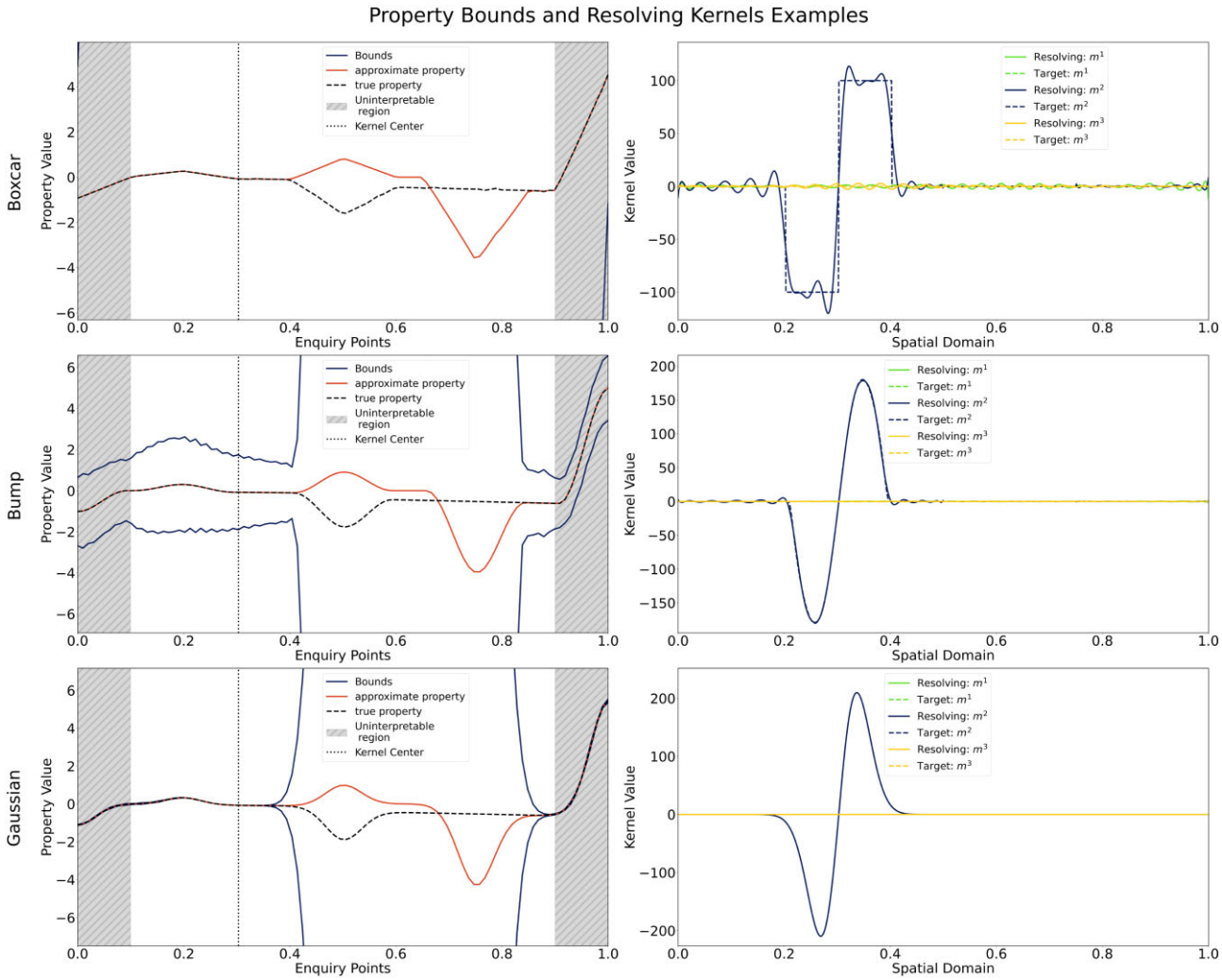


Figure 8. Case 1: SOLA-DLI solutions for three different types of local averaged gradient properties, similar to Fig. 7.

relative error bounds) than uniform or bump local averages, for all enquiry points and for target kernel widths (i.e. all resolutions), likely due to the fact that Gaussian-like sensitivity kernels were used.

This case study illustrates the general notion that we typically use inference methods to answer specific questions about a true model rather than finding the entire model itself. In SOLA-DLI, these questions are encoded in our chosen target kernels, which should be carefully designed to improve the property bounds and facilitate straightforward interpretations. The differing extent to which we are able to retrieve different target kernels effectively shows that our data can answer some questions better than others.

3.2 Case 2: quasi synthetic normal mode application

In this quasi-synthetic case study, we illustrate how to use SOLA-DLI to conduct a simple resolution analysis without real data or model values nor any prior information, based solely on the sensitivity kernels of the data set. We also illustrate how the results of such a resolution analysis can be linked to trade-offs between physical parameters.

3.2.1 Setup

We consider a model formed by the triplet $m = (\delta \ln(v_s), \delta \ln(v_p), \delta \ln(\rho))$, where v_s is shear-wave speed, v_p is compressional-wave speed, and ρ is density (Fig. 11). Each physical parameter is assumed to be a piece-wise continuous function defined over the interval $[0, R_E]$ where R_E is Earth's radius (approximately 6371 km). We aim to constrain Gaussian averages and gradients of this synthetic true model using realistic normal mode sensitivity kernels (Woodhouse & Dahlen 1978). Specifically, we select the same modes as in the SP12RTS data set (Koelemeijer *et al.* 2016; Restelli *et al.* 2024), that is, 143 modes with their sensitivity to $\delta \ln(v_s)$, $\delta \ln(v_p)$ and $\delta \ln(\rho)$ concentrated mostly in the mantle (see Fig. 12).

3.2.2 Resolution analysis

Before introducing any data or model values, we are able to perform a simple resolution analysis to investigate where and on what spatial scale our data contain information regarding the Earth model. While the SOLA-DLI solution itself depends on the model norm bound via M (see eq. 24), indirectly on the data via $\|\tilde{m}\|_{\mathcal{M}}$ (see eq. 25), and on the relationships between the target kernels and sensitivity kernels

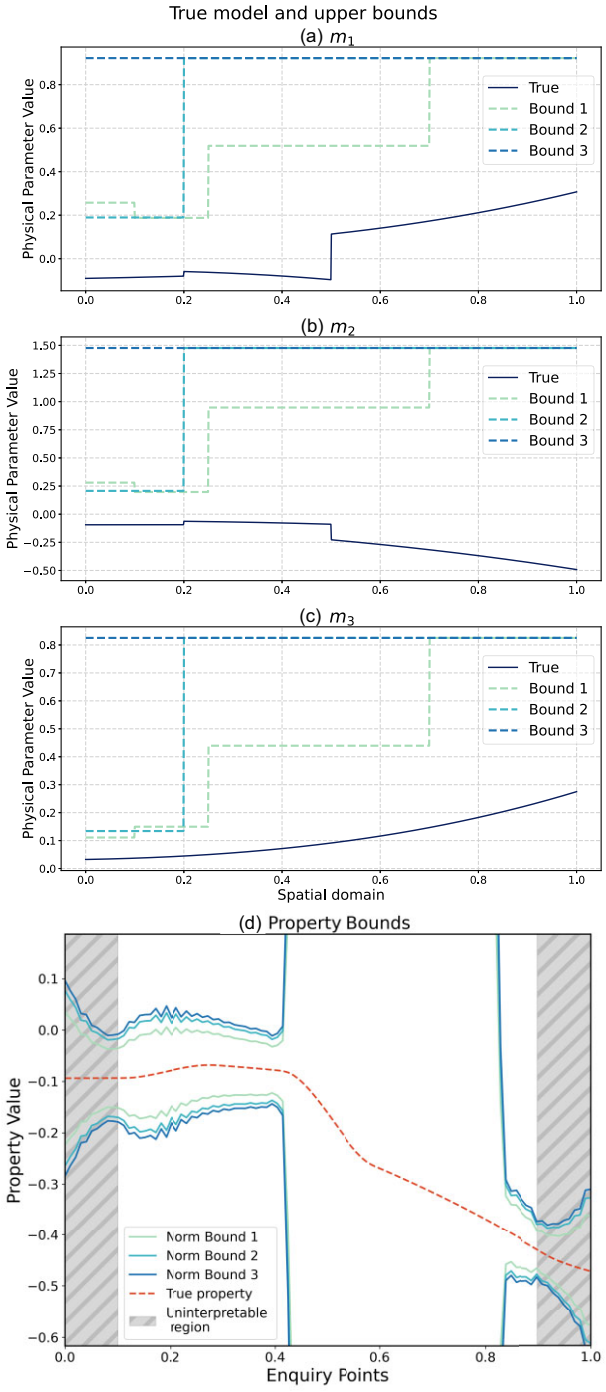


Figure 9. Case 1: effect of prior norm bound on the property bounds. Panels (a)–(c) indicate the levels of three different upper bounds on all three model parameters of the true model. Our choice of norm bound functions results in the following prior norm bounds (M^i): 2.44 (Bound 1), 3.89 (Bound 2) and 4.34 (Bound 3), which are all larger than the true model norm of 0.32. (d) Solutions corresponding to the three different model upper bounds, using a bump average for m^2 as example. Tighter norm bounds lead to tighter constraints on the desired properties.

via $\mathcal{H}$ (see eqs D14–D23), the resolving kernels only depend on the data geometry, that is, the data sensitivity kernels.

The diagonal elements of the matrix $\mathcal{H}$ can be shown to equal:

$$\mathcal{H}_{kk} = \sum_{j}^{N_m} \left\| T^{j,(k)} - A^{j,(k)} \right\|^2_{\mathcal{M}_j} \tag{55}$$

which essentially quantifies the cumulative difference between our target and resolving kernels. Using $\mathcal{H}$ we can also define the resolving misfit as a more useful metric:

$$R_k = \frac{\sqrt{\mathcal{H}_{kk}}}{\sum_j^{N_m} \left\| T^{j,(k)} \right\|_{\mathcal{M}_j}} = \frac{\sqrt{\sum_j^{N_m} \left\| T^{j,(k)} - A^{j,(k)} \right\|^2_{\mathcal{M}_j}}}{\sum_j^{N_m} \left\| T^{j,(k)} \right\|_{\mathcal{M}_j}} \tag{56}$$

which is a generalization of the 'resolution misfit' defined in Restelli *et al.* (2024). The resolving misfit is 0 when all the resolving kernels associated with some property evaluated at $r^{(k)}$ are equal to the corresponding target kernels. This would mean that our data contain exact information about the desired property and the property error bounds are 0. On the other hand, the resolving misfit is equal to 1 when our resolving kernels are zero, which would correspond to a complete lack of sensitivity of our data to the desired property. It is important to note that the computation of the resolving misfit does not use the data vector d nor any prior model information, it only uses the 'geometry of the data set' (Latallerie *et al.* 2024). Fig. 13 illustrates the information that is provided by the resolving misfit (left column). As indicated by a low resolving misfit (darker shades of blue), our data mostly contain information in the mantle, as expected from this selection of sensitivity kernels. The resolving misfit is also typically low for wide target kernels (> 18 per cent domain width, or more than 1000 km). Wide gradient kernels can be better recovered in the mantle, while wide averaging kernels can be better recovered in the lower outer core, again indicating that our choice of target (i.e. property) is important.

3.2.3 *Trade-offs between physical parameters*

When our data are sensitive to two or more physical parameters, it may become difficult or impossible to obtain properties of a single parameter in isolation from the others. These trade-offs between physical parameters pose problems for interpretations, particularly in regions such as the lower mantle where the sensitivity of normal modes to seismic velocities and density is similar.

Our setup with SOLA-DLI, where we explicitly set the target kernels for parameters not of interest to zero, enables us to easily visualize and consider model parameter trade-offs. Suppose we are interested in some local property of $\delta \ln(\rho)$, for example the Gaussian local average density in the deep mantle or the density jump across the 660 discontinuity as characterized by a Gaussian gradient (Lau & Romanowicz 2021). If we choose low resolution (wide) target kernels (middle column), we find that the resolving kernels for $\delta \ln(\rho)$ match the target kernels well (Fig. 13). Furthermore, the resolving kernels for $\delta \ln(v_s)$ and $\delta \ln(v_p)$ also match their respective target kernels, which are just zero. Such zero or near zero resolving kernels indicate that the trade-off between the physical parameter of interest and the other physical parameters is small. However, if we choose higher resolution (thin) target kernels (right column), we notice that the resolving kernels are struggling to match their respective target kernels. The resolving kernels for $\delta \ln(v_p)$ and particularly $\delta \ln(v_s)$ are far from zero, indicating significant trade-offs with the desired property of density, which are regarded as *contaminants*. Such trade-offs between physical parameters are naturally

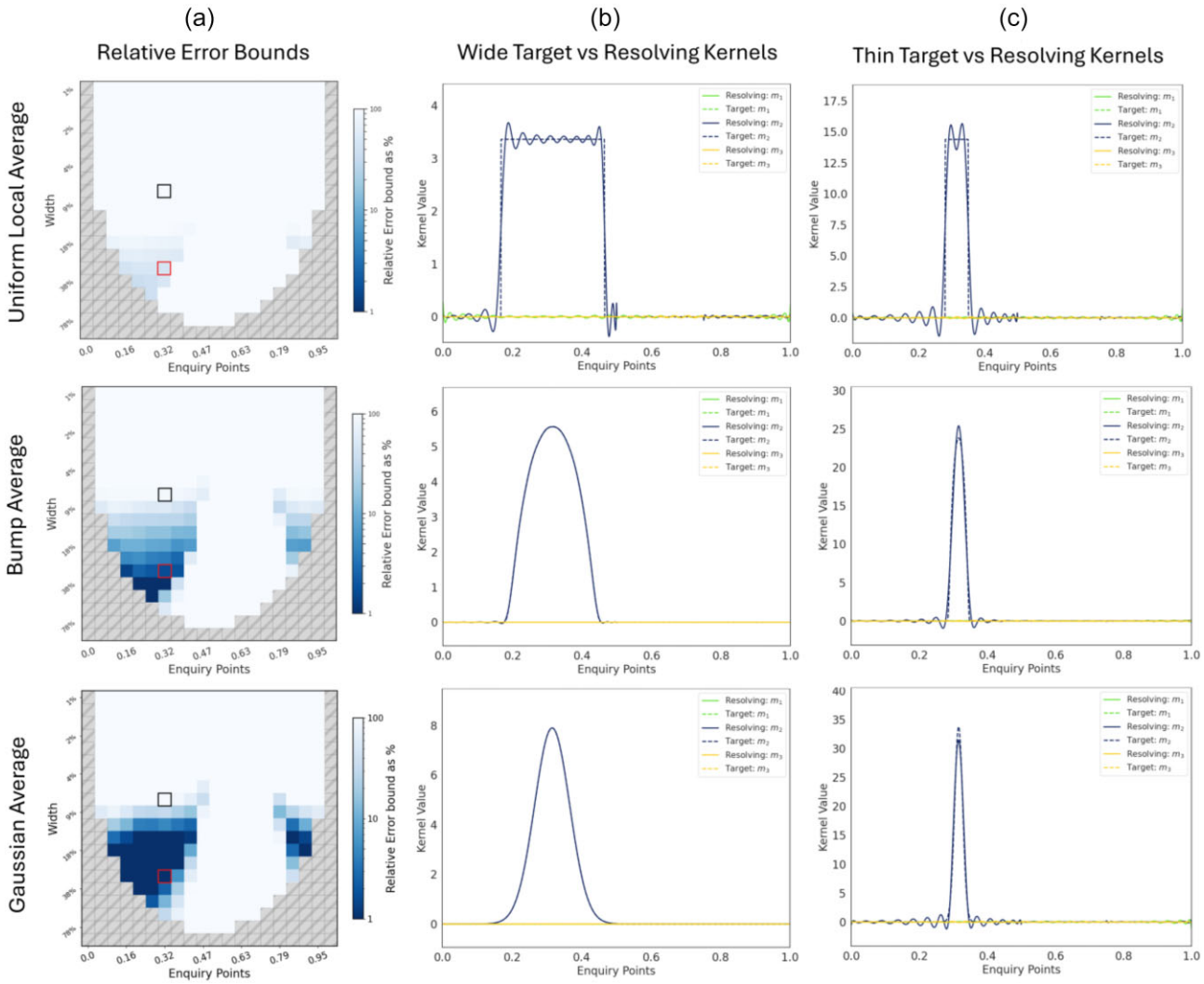


Figure 10. Case 1: relative error bounds with examples of resolving kernels compared to their target kernels. The different rows correspond to different types of target kernels, for example, uniform local average (top), bump average (middle) and Gaussian average (bottom). In the three columns, we show (a) the relative error bounds e; (b) examples for wide target and resolving kernels, corresponding to the red squares in panels (a) and (c) examples for narrow target and resolving kernels, corresponding to the black squares in (a).

taken into account by SOLA-DLI and typically result in higher error bounds on the property. If instead we would account for the sensitivity to $\delta \ln(v_p)$ and $\delta \ln(v_s)$ by scaling the sensitivity kernels, we would obtain tighter bounds, at the expense of assuming more prior information, similar to the results of Restelli *et al.* (2024) using the '3-D noise' approach in their SOLA inversions.

3.3 Case 3: discretized inversions using continuous SOLA-DLI

This final case study serves to illustrate how we can obtain a family of discretized model solutions using SOLA-DLI, and how this approach compares to a typical least-squares inversion model solution.

3.3.1 Setup

Here, we consider a model m with only one physical parameter, denoted also m (see the true model in Fig. 14a). Our model space $\mathcal{M}$ is $PCb[0, 1]$ and the data are given by:

$$d_i = \langle K_i, m \rangle_{\mathcal{M}}\,, \tag{57}$$

where K_i are some quasi-randomly functions, generated again using eq. (49) (see Fig. 14b).

In this setup, we choose to discretize the model using a Fourier expansion. The resulting basis functions are (see Fig. 14c):

$$B_l(r) = \begin{cases} 1, & l = 0 \\ \sqrt{2}\sin\left(2\pi \frac{l+1}{2} r\right), & l \text{ odd} \\ \sqrt{2}\cos\left(2\pi \frac{l}{2} r\right), & l \text{ even} \end{cases} \tag{58}$$

and a possible model expansion with Fourier coefficients p_l is given by:

$$m(r) \approx \sum_l p_l B_l(r). \tag{59}$$

The discretized model–data relation used for the least-squares inversion is:

$$d_i = \sum_l \langle K_i, B_l \rangle_{\mathcal{M}}\, p_l = \sum_l \Gamma^*_{il} p_l, \tag{60}$$

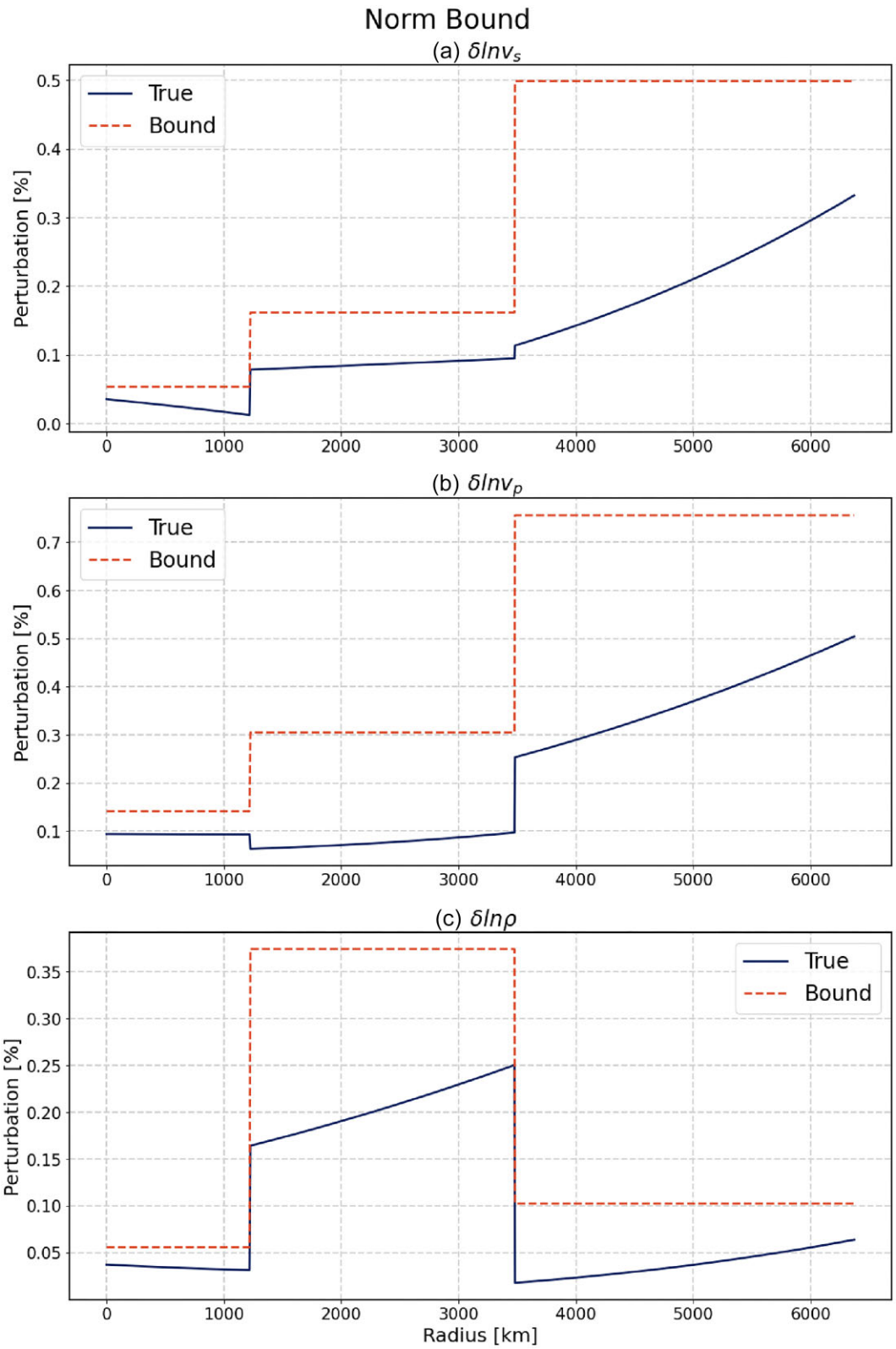


Figure 11. Case 2: arbitrary quasi-random synthetic true model and the upper bound functions used to compute the prior upper bound norm.

where (see also Appendix D2):

$$\Gamma^*_{ij} = \langle K_i, B_l \rangle_{\mathcal{M}} \,. \tag{61}$$

This leads to the following least-squares solution for p_l:

$$\hat{p} = (\Gamma\Gamma^*)^{-1}\Gamma d. \tag{62}$$

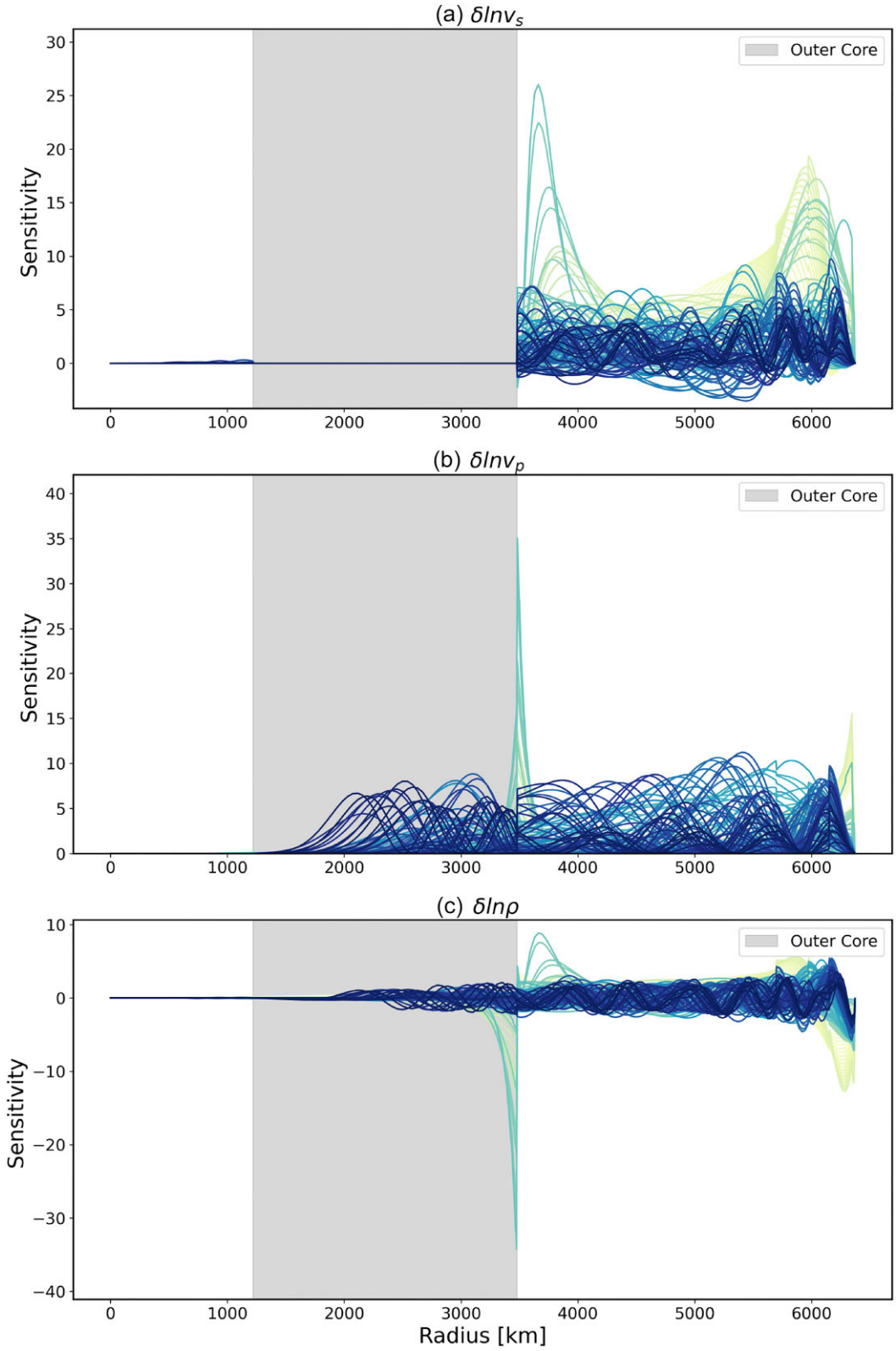


Figure 12. Case 2: normal mode sensitivity kernels for (a) $\delta \ln(v_s)$, (b) $\delta \ln(v_p)$ and (c) $\delta \ln(\rho)$, obtained using a modified version of OBANI based on the work of Woodhouse & Dahlen (1978). The shaded region indicates the depth range of the outer core, where the sensitivity to v_s is zero.

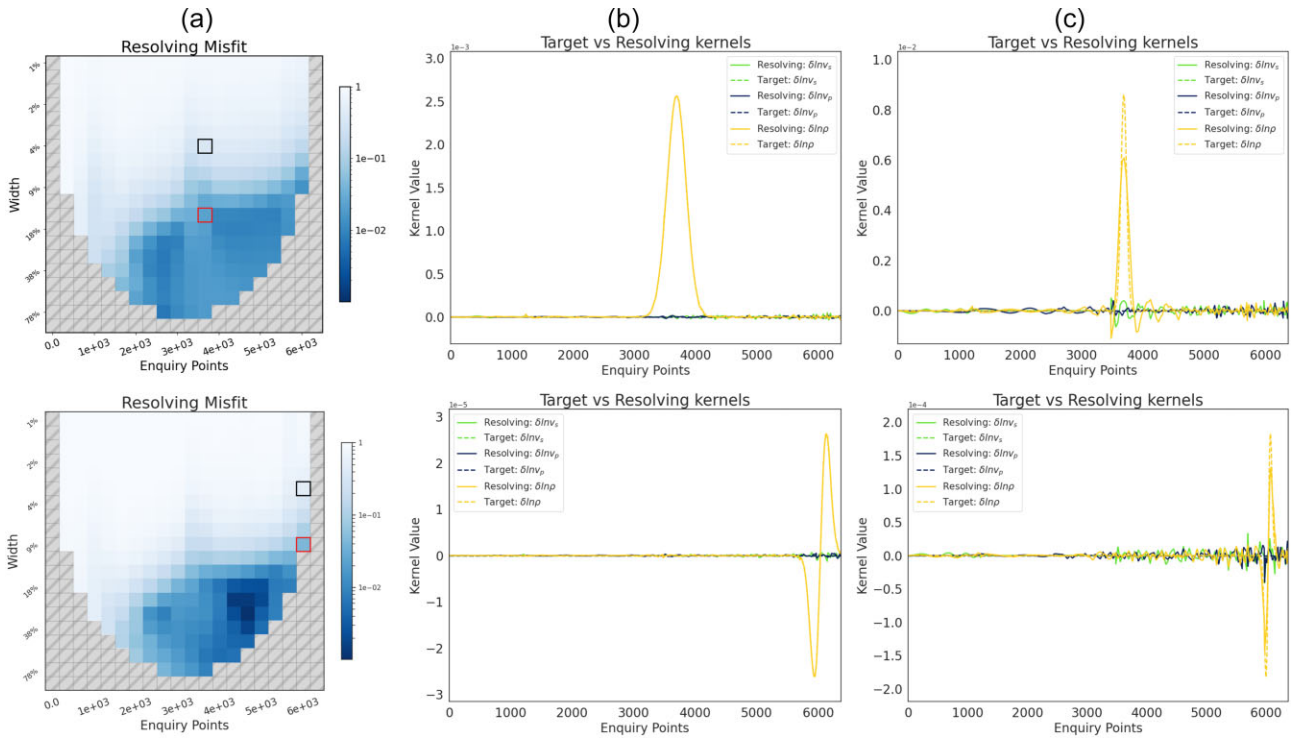


Figure 13. Case 2: resolution analysis for a Gaussian average (top) and gradient (bottom) target for $\delta \ln(rho)$ using realistic mode sensitivity kernels. The resolving misfit (left) and kernels (middle and right) can be computed without the need for data or any prior norm bound information. The middle and right panels illustrate the target and resolving kernels for a wide and thin target, including the contaminant kernels that indicate trade-offs between physical parameters.

Using the least-squares solution $\{\hat{p}_l\}$, we can thus find the corresponding model solution by using the Fourier expansion:

$$\hat{m} = \sum_l \hat{p}_l B_l. \tag{63}$$

To obtain the SOLA-DLI solution, we consider the Fourier coefficients p_l to be elements of a property vector obtained from the property mapping:

$$p_l = [\mathcal{T}(m)]_l = \langle B_l, m \rangle_{\mathcal{M}}. \tag{64}$$

We also introduce a prior model norm bound (see Fig. 14a). This leads to the following SOLA-DLI problem:

Given

$$d_i = \langle K_i, m \rangle_{\mathcal{M}} \tag{65}$$

Find

$$p_l = \langle B_l, m \rangle_{\mathcal{M}}. \tag{66}$$

This problem is readily solved using eq. (23) to obtain upper and lower bounds for the possible values of the Fourier coefficients:

$$\bar{p}_l \in [\tilde{p}_l - \epsilon_l, \tilde{p}_l + \epsilon_l]. \tag{67}$$

In this case $\tilde{p}_l$ are the Fourier coefficients of the least norm solution to eq. (57), which are not to be confused with the single least-squares solution $\hat{p}_l$. In contrast, the property bounds obtained from SOLA-DLI (eq. 67) offer a family of solutions that can be sampled.

3.3.2 *Discretized least-squares versus SOLA-DLI solution*

We compute both the least-squares and discretized SOLA-DLI solution using different number of data points (50, 70 or 100), solving for 29 Fourier coefficients. SOLA-DLI initially provides property bounds on the Fourier coefficients, and we therefore have to draw samples from these distribution for each Fourier coefficient distribution to obtain a possible model solution, illustrated in Fig. 15.

When using few data (Fig. 15, the least-squares inversion generally struggles to retrieve Fourier coefficients close to the true ones, while the bounds of the SOLA-DLI solution always encompass the true coefficients (true properties). That said, for certain Fourier coefficients and in certain parts of the model, the least-squares solution does appear to approach the true property (Fourier coefficients) and the true model better than the SOLA-DLI solution. Increasing the number of data to 70 leads to a better least-squares solution, especially for the first 10 Fourier coefficients, and tighter bounds of the SOLA-DLI solution. However, it now becomes clear that the SOLA-DLI bounds offer more accurate information, always encompassing the true Fourier coefficients and better resembling the true model compared to the least-squares solution. When we further increase the number of data points to 100 (see Figs 15c and f), we note that the SOLA-DLI solution converges closely to the true Fourier coefficients and model, while the least-squares inversion systematically deviates.

In our synthetic setup, it is possible to explicitly compute the data correction term, which captures the components of the true model that are not within the span of the basis functions (see Section 2.4, eq. 44). When we correct the data, using our knowledge of the true model, we find that the least-squares inversion solution converges to the true Fourier coefficients, even for few data (Fig. 16). This demonstrates the equivalence of the discretized least-squares and SOLA-DLI solutions. However, in real world applications, when the true model is unknown, this data correction term cannot be computed. Consequently, the SOLA-DLI solution should be preferred

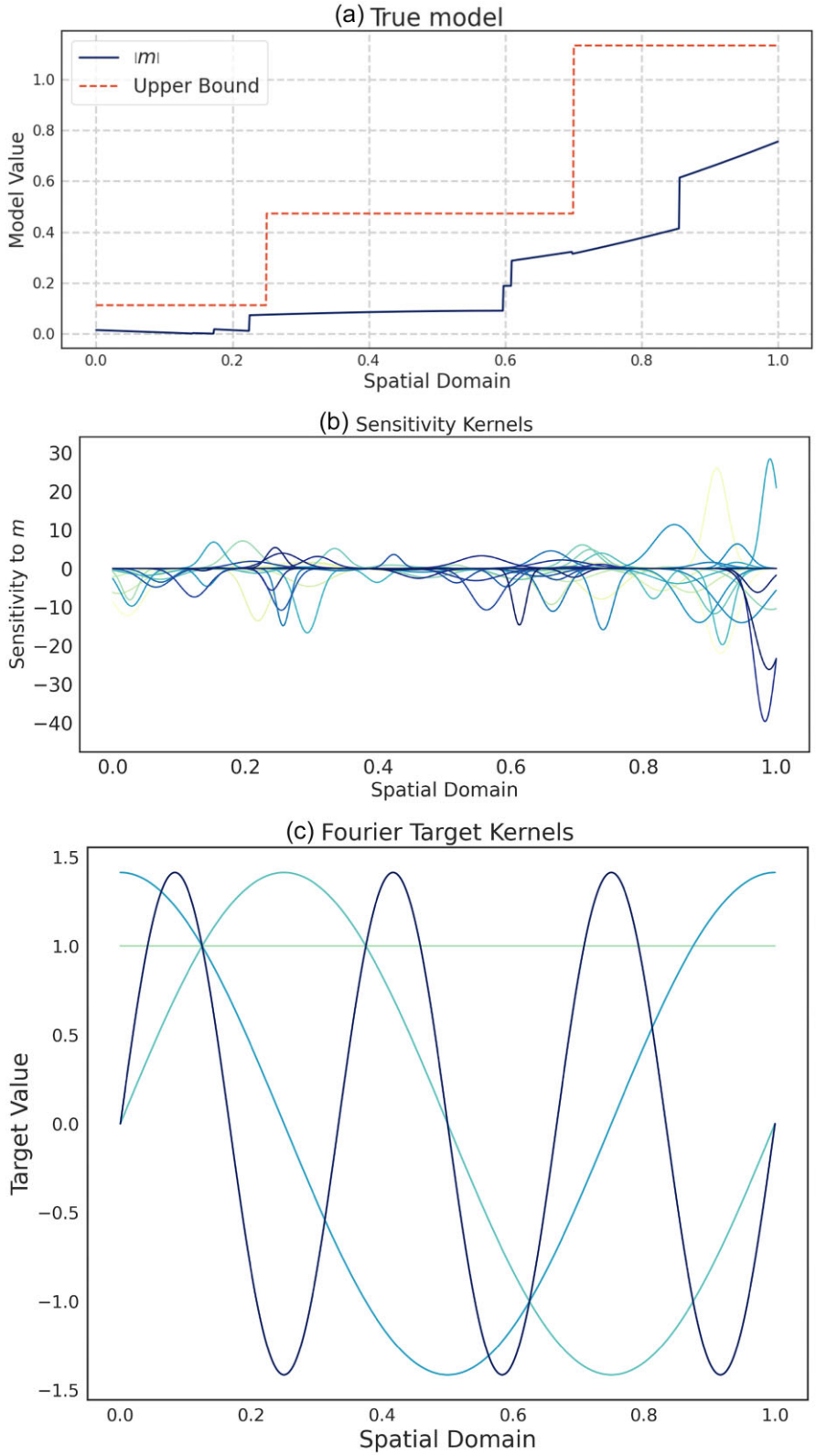


Figure 14. Case 3: Model setup and kernels. (a) True model m with the upper bound function used to compute the prior model norm bound. (b) Synthetic quasi-random sensitivity kernels. (c) Examples of four Fourier basis functions.

over the discretized least-squares inversion method. As mentioned before, there are other methods (e.g. Trampert & Snieder 1996) for bypassing or approximating the effect of the data correction term, which should also be preferred over a simple least norm inversion.

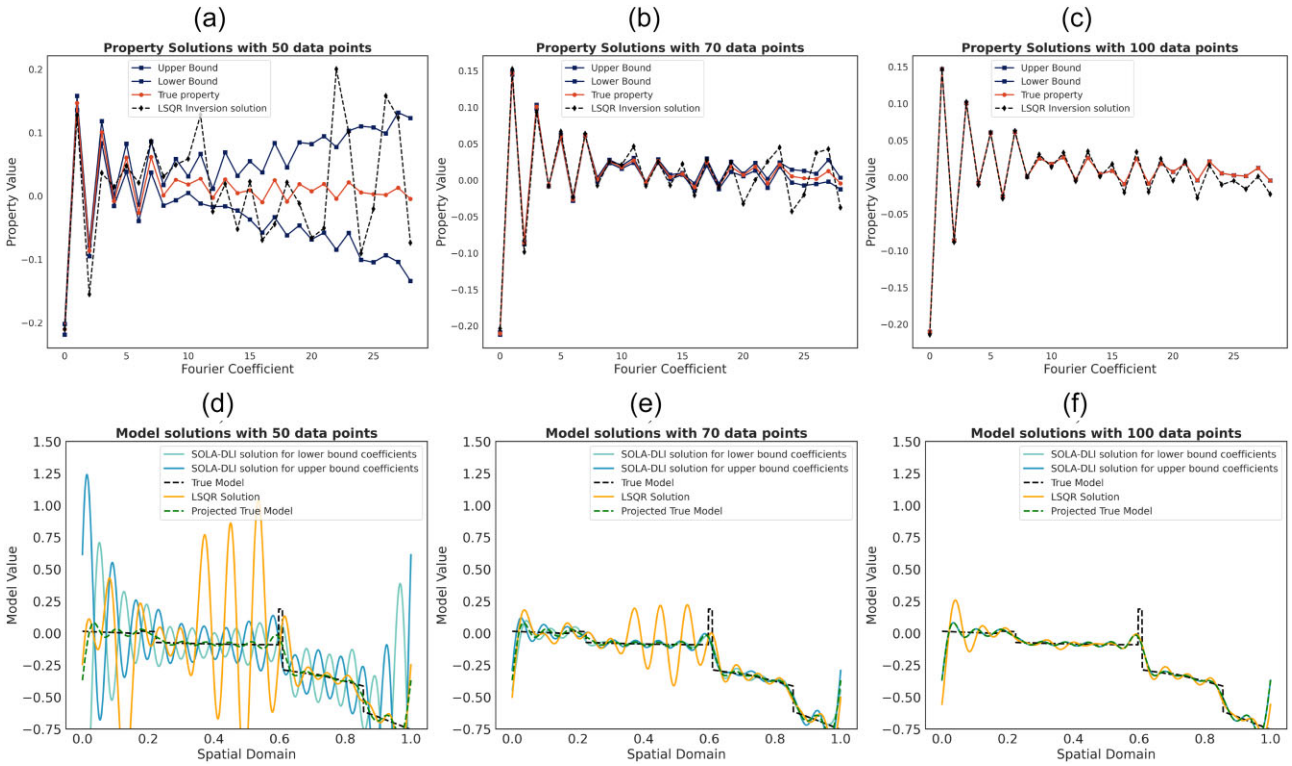


Figure 15. Case 3: comparison between least-squares and SOLA-DLI solution for a model discretized using Fourier basis functions. (a–c) Fourier coefficients from discretized least-squares inversion and bounds of the SOLA-DLI solution using (a) 50 data points, (b) 70 data points and (c) 100 data points. (d–f) Discretized model solution from discretized least-squares and two samples from the SOLA-DLI property bounds using (d) 50 data points, (e) 70 data points and (f) 100 data points.

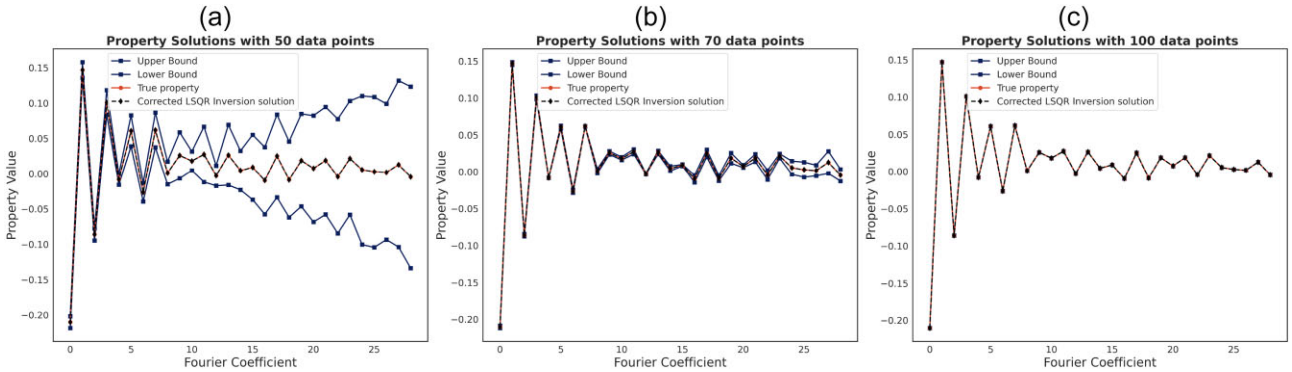


Figure 16. Case 3: comparison between Fourier coefficients obtained using SOLA-DLI and the discretized least-squares method with an additional data correction term, using (a) 50 data points, (b) 70 data points and (c) 100 data points. A comparison with the least-squares solutions in Fig. 15 indicates that the data correction leads to the systematic error in the Fourier coefficients.

4 DISCUSSION

In this contribution, we have introduced the SOLA-DLI framework, which combines the advantages of both DLI and SOLA branches of inferences. At present, we have focused on error-free data, as the fundamental distinction between the two branches lies in their treatment of uncertainties arising from incomplete data, not from how data noise is incorporated. However, for any real-world application, it is essential to address data noise. Al-Attar (2021), Parker (1977) and Backus (1970a) have each proposed methods for incorporating noise into DLI-based approaches. Since the SOLA-DLI framework integrates both methodologies, we can draw upon these existing approaches and adapt them as necessary to introduce data noise into the framework. There are numerous ways to achieve this, but it is not yet clear which approach would best balance computational efficiency with the need to produce property bounds that are not excessively large. Depending on the chosen approach, the matrix X may be affected by noise, which would alter the final form of the resolving kernels. However, this does not render them unusable or uncomputable, and crucially, it does not alter their interpretation. We believe that the selection of target kernels also remains important, but in the presence of data noise, a particular set of target kernels A may perform better than another set B, whereas in the absence of noise, set B might outperform set A. This variability does not undermine the points that we make in the present contribution, which is that some target kernels are more effective than others. A potential direction for future research could be target optimization,

where, given a set of data, the goal is to identify those target kernels that produce the most effective constraints from a family of, for example, averaging weight functions. We anticipate that the methods behind such an optimization algorithm would need to account for data noise.

We expect that the careful treatment of target kernels will become more involved when going to 2-D or 3-D cases (e.g. Zaroli 2016; Latallerie *et al.* 2022; Freissler *et al.* 2024). However, it also opens up more possibilities. In higher dimensions, we can design target kernels sensitive to directional gradients or local curvature using kernels that represent smoothed Laplacian operators. Such target kernels would for example amplify the presence of a gap between two peaks rather than smoothing the peaks into one, which could be useful for studying plumes (similar to the idea of point-spread functions of Fichtner & Trampert 2011). Another possible extension of our work would be to replace the deterministic prior information with probabilistic information by placing a prior measure on the model space, which can be updated using noisy data measurements and propagated into the property space. Backus (1970a) already mentioned such a modification and Al-Attar (2021) added to this discussion. Such a modification would lead to yet another possible mechanism for dealing with data noise.

The introduction of prior model information via the model norm bound is of great importance in the SOLA-DLI method. The model norm bound (L_2 norm) chosen here is the most common due to its mathematical simplicity, but as pointed out by Al-Attar (2021) and discussed in Section 2.1.2, there might be better prior constraints. Other model norms may allow to place bounds on the maximum point-value of the true model, or its gradients (smoothness) (Stark & Hengartner 1993). Such modifications may necessitate the use of more general spaces than Hilbert spaces, which adds significant theoretical complications. We refer the interested reader to Al-Attar (2021) for the required theoretical modifications.

The computational cost of the methods presented arises from multiple sources. Computing the matrices Λ ($N_d \times N_d$) and Γ ($N_p \times N_d$) requires at most $N_m(N_d^2/2 + N_p N_d)$ integrations for N_m model parameters, with the cost depending on the sensitivity and target kernels used. Sensitivity kernels, especially in the case of finite-frequency adjoint methods, can be expensive to compute. If sensitivity kernels already exist, the integration cost for SOL-DLI depends only on the number of kernels and the integration scheme. Since Λ^{-1} is rarely computed explicitly, applying it involves solving $N_p + 1$ linear systems, similar to the cost of obtaining a SOLA or DLI solution without data noise (Al-Attar 2021), and much lower than the classic Backus–Gilbert method (Backus & Gilbert 1970; Pijpers & Thompson 1992). For SOLA-DLI, these $N_p + 1$ solves yield the final solution, while normal DLI requires additional computations involving $\mathcal{H}^{-1}$ to assess the hyperellipsoid constraints. When accounting for data noise, the cost depends on the method used to incorporate it, which will likely be adapted from existing SOLA or DLI approaches. Thus, the total cost of SOLA-DLI is expected to be comparable to DLI or SOLA, making it computationally attractive for inference problems.

5 CONCLUSION

In this contribution, we have presented the theory and possible applications of the SOLA-DLI framework, which combines the Backus–Gilbert based SOLA method with DLI. To derive this framework, we have first demonstrate the links between these two branches of inference methods, before showing how the combined framework is capable of providing a more comprehensive analysis. We have particularly emphasized the distinction between interpreting results through target kernels versus resolving kernels. As a result, target kernel design is significantly more important in SOLA-DLI. In addition, the framework is capable of incorporating multiple physical parameters, with trade-offs captured by contaminant kernels. Furthermore, we have demonstrated how discretized models can be obtained using these linear inference methods, highlighting the advantages and disadvantages associated with different approaches. All of these theoretical aspects are practically demonstrated through three synthetic, noise-free case studies, with software provided to enable the reader to explore these further themselves.

ACKNOWLEDGMENTS

We thank the Editor (Carl Tape), Malcolm Sambridge and an anonymous reviewer for their comments, which have substantially improved the manuscript. AMM was funded by a NERC DTP studentship NE/S007474/1 and gratefully acknowledges their support. AMM also received funding from Royal Society grant RF\ERE\210182 awarded to PK. CZ acknowledges financial support from ITES (Institut Terre et Environnement de Strasbourg, UMR 7063) for a research visit to Oxford. PK acknowledges financial support from a Royal Society University Research Fellowship (URF\1\180377). Normal mode sensitivity kernels were computed using a modified version of OBANI, which is based on the work of Woodhouse & Dahlen (1978). The authors thank Sam Scivier, Franck Latallerie and David Al-Attar for fruitful theoretical discussions. The following Python packages were used extensively for producing and plotting the synthetic data: Scipy (Virtanen *et al.* 2020), Numpy (Harris *et al.* 2020) and Matplotlib (Hunter 2007). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

DATA AVAILABILITY

The sensitivity kernels and all the codes used to produce the figures in this paper can be found at https://github.com/Adrian-Mag/SOLA_DLI (Mag *et al.* 2024).

REFERENCES

Al-Attar, D., 2021. Linear inference problems with deterministic constraints, preprintarXiv:2104.12256.

Amiri, S., Maggi, A., Tatar, M., Zigone, D. & Zaroli, C., 2023. Rayleigh wave group velocities in north-west iran: Sola Backus–Gilbert vs. fast marching tomographic methods, *Seismica,* **2**(2).

Backus, G., 1970a. Inference from inadequate and inaccurate data, i, *Proc. Natl. Acad. Sci. USA,* **65**(1), 1–7.

Backus, G., 1970b. Inference from inadequate and inaccurate data, iii, *Proc. Natl. Acad. Sci. USA,* **67**(1), 282–289.

Backus, G., 1970c. Inference from inadequate and inaccurate data, ii∗, *Proc. Natl. Acad. Sci. USA,* **65**(2), 281–287, .

Backus, G. & Gilbert, F., 1968a. The resolving power of gross earth data, *Geophys. J. Int.,* **16**(2), 169–205.

Backus, G. & Gilbert, F., 1968b. The resolving power of gross earth data, *Geophys. J. Int.,* **16**(2), 169–205.

Backus, G. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross earth data, *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Sci.,* **266**(1173), 123–192.

Backus, G.E., 1988a. Bayesian inference in geomagnetism, *Geophys. J. Int.,* **92**(1), 125–142.

Backus, G.E., 1988b. Comparing hard and soft prior bounds in geophysical inverse problems, *Geophys. J. Int.,* **94**(2), 249–261.
Backus, G.E., 1989. Confidence set inference with a prior quadratic bound, *Geophys. J. Int.,* **97**(1), 119–150.
Backus, G.E. & Gilbert, J., 1967a. Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. Int.,* **13**(1–3), 247–276.
Backus, G.E. & Gilbert, J.F., 1967b. Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. Int.,* **13**(1–3), 247–276.
Fichtner, A. & Trampert, J., 2011. Resolution analysis in full waveform inversion, *Geophys. J. Int.,* **187**(3), 1604–1624.
Freissler, R., Schuberth, B. S.A. & Zaroli, C., 2024. A concept for the global assessment of tomographic resolution and uncertainty, *Geophys. J. Int.,* ggae178, doi:10.1093/gji/ggae178, .
Harris, C.R. *et al.*, 2020. Array programming with NumPy, *Nature,* **585**(7825), 357–362, .
Hunter, J.D., 2007. Matplotlib: A 2-D graphics environment, *Comput. Sci. Eng.,* **9**(3), 90–95, .
Koelemeijer, P., Ritsema, J., Deuss, A. & Van Heijst, H.-J., 2016. SP12RTS: a degree-12 model of shear-and compressional-wave velocity for Earth's mantle, *Geophys. J. Int.,* **204**(2), 1024–1039.
Latallerie, F., Zaroli, C., Lambotte, S. & Maggi, A., 2022. Analysis of tomographic models using resolution and uncertainties: a surface wave example from the pacific, *Geophys. J. Int.,* **230**(2), 893–907.
Latallerie, F., Zaroli, C., Lambotte, S., Maggi, A., Walker, A. & Koelemeijer, P., 2024. Surface-wave tomography with 3D resolution and uncertainty, *Seismica*.
Lau, H.C. & Romanowicz, B., 2021. Constraining jumps in density and elastic properties at the 660 km discontinuity using normal mode data via the Backus–Gilbert method, *Geophys. Res. Lett.,* **48**(9), e2020GL092217.
Mag, A., Zaroli, C. & Koelemeijer, P., 2024. Adrian-Mag/SOLA_DLI: SOLA-DLI.
Masters, G. & Gubbins, D., 2003. On the resolution of density within the Earth, *Phys. Earth planet. Inter.,* **140**(1–3), 159–167.
Nolet, G., 2008. *A Breviary of Seismic Tomography,* Cambridge University Press, Cambridge, UK.
Oldenburg, D., 1981. A comprehensive solution to the linear deconvolution problem, *Geophys. J. Int.,* **65**(2), 331–357.
Parker, R.L., 1977. Linear inference and underparametrized models, *Rev. Geophys.,* **15**(4), 446–456.
Pijpers, F. & Thompson, M., 1992. Faster formulations of the optimally localized averages method for helioseismic inversions, *Astron. Astrophys.,* **262**(2), L33–L36.
Pijpers, F. & Thompson, M., 1994. The SOLA method for helioseismic inversion, *Astron. Astrophys.,* **281**(1), 231–240.
Rawlinson, N., Pozgay, S. & Fishwick, S., 2010. Seismic tomography: a window into deep earth, *Phys. Earth planet. Inter.,* **178**(3–4), 101–135.
Restelli, F., Zaroli, C. & Koelemeijer, P., 2024. Robust estimates of the ratio between s- and p-wave velocity anomalies in the earth's mantle using normal modes, *Phys. Earth planet. Inter.,* **347,** 107 135.
Ritsema, J. & Lekić, V., 2020. Heterogeneity of seismic wave velocity in earth's mantle, *Annu. Rev. Earth Planet. Sci.,* **48,** 377–401.
Snieder, R., 1991. An extension of Backus–Gilbert theory to nonlinear inverse problems, *Inverse Problems,* **7**(3), 409.
Stark, P.B., 2008. Generalizing resolution, *Inverse Problems,* **24**(3), 034 014.
Stark, P.B. & Hengartner, N.W., 1993. Reproducing Earth's kernel: uncertainty of the shape of the core-mantle boundary from PKP and PcP traveltimes, *J. Geophys. Res. Solid Earth,* **98**(B2), 1957–1971.
Tarantola, A., 1987. *Inverse Problem Theory,* Elsevier, Amsterdam.
Trampert, J. & Snieder, R., 1996. Model estimations biased by truncated expansions: possible artifacts in seismic tomography, *Science,* **271**(5253), 1257–1260.
Valentine, A.P. & Sambridge, M., 2023. Emerging directions in geophysical inversion, *Applications of Data Assimilation and Inverse Problems in the Earth Sciences, Vol. 5*
Virtanen, P. *et al.*, 2020. SciPy 1.0: fundamental algorithms for scientific computing in python, *Nat. Methods,* **17,** 261–272, .
Woodhouse, J.H. & Dahlen, F.A., 1978. The effect of a general aspherical perturbation on the free oscillations of the earth, *Geophys. J. Int.,* **53**(2), 335–354.
Zaroli, C., 2016. Global seismic tomography using Backus–Gilbert inversion, *Geophys. J. Int.,* **207**(2), 876–888, .
Zaroli, C., 2019. Seismic tomography using parameter-free Backus–Gilbert inversion, *Geophys. J. Int.,* **218**(1), 619–630.
Zaroli, C., Koelemeijer, P. & Lambotte, S., 2017. Toward seeing the Earth's interior through unbiased tomographic lenses, *Geophys. Res. Lett.,* **44**(22), 11–399.

APPENDIX A: OVERVIEW OF INFERENCE METHODS

In this apppendix, we present an informal overview of inference methods in the absence of data noise. We consider the most general form of an inference problem to be (see also Fig. 1b):

$$\begin{aligned} &\text{Given:}\\ &G(\bar{m}) = d\\ &\text{Find:}\\ &\mathcal{T}(\bar{m}) = \bar{p}\\ &\text{where}\\ &G : \mathcal{M} \to \mathcal{D}\\ &\mathcal{T} : \mathcal{M} \to \mathcal{P} \end{aligned}$$

The true model $\bar{m}$ is unknown, and we only have data constraints at our disposal to find some properties $\bar{p}$ of that true model. In general, there are six choices we have to make before attempting to solve this problem: we must decide what $G, \mathcal{T}, \mathcal{M}, \mathcal{D}, \mathcal{P}$ are, and whether we want to introduce prior information or not, which we will discuss below.

A1 Choice of $\mathcal{D}$, $\mathcal{P}$, $\mathcal{M}$ and $\mathcal{T}$

For most applications, we are only able to measure a finite number of data and we are typically interested in a finite number of properties. In addition, the data and properties are usually real. Therefore, in most cases there is only one option for the data space $\mathcal{D}$ and property space $\mathcal{P}$: they are $\mathbb{R}^{N_d}$ and $\mathbb{R}^{N_p}$ where N_d, N_p are the number of data, and the number of properties, respectively.

The model space $\mathcal{M}$ is most often a function space or a finite dimensional real vector space. On a more fundamental level, we are interested in whether the space is simply a Banach space or if it possesses an inner product structure, making it a Hilbert space. Some authors have proposed solutions to inference problems in Banach spaces (e.g. Stark 2008; Al-Attar 2021), while most others have focused on the more structured Hilbert space (e.g. Backus & Gilbert 1967b; Backus 1970a; Pijpers & Thompson 1994; Zaroli 2016; Al-Attar 2021), as this simplifies the mathematics.

The property mapping $\mathcal{T}$ is typically chosen to be a linear functional, as most inference problems focus on point evaluation, basis coefficients, or local averages, all of which are linear functionals. In this paper, we argue that a more careful consideration of these functionals can lead to improved results in inference problems, particularly in the context of SOLA/DLI-type inference problems (as discussed in Section 2.2).

Table A1. Comparison of different Backus–Gilbert based inference methods with noise-free data. Note that the original papers use A instead of R for resolving (or averaging) kernels, but we prefer the general R as we consider a range of targets.

	Original Backus–Gilbert (BG)	SOLA	Deterministic linear inferences (DLI)	SOLA-DLI
Solution	$\bar{p} \approx \int_\Omega R\bar{m}d\Omega$ where $R = \sum_i x_i K_i$ and x_i minimize $\int_\Omega J^2 R^2 d\Omega$ s.t. $\int_\Omega R d\Omega = 1$	$\bar{p} \approx \int_\Omega R\bar{m}d\Omega$ where $R = \sum_i x_i K_i$ and x_i minimize $\int_\Omega (R-T)^2 d\Omega$ s.t. $\int_\Omega R d\Omega = 1$	$\langle \mathcal{H}^{-1}(p-\tilde{p}), p-\tilde{p}\rangle \leq M^2 - \|\tilde{m}\|^2_{\mathcal{M}}$	$\tilde{p} \in \tilde{p} + [-\epsilon, +\epsilon]$ where $\epsilon = \sqrt{(M^2 - \|\tilde{m}\|^2_{\mathcal{M}}) diag(\mathcal{H})}$
Prior information	None	None	$\|\bar{m}\| \leq M$	$\|\bar{m}\| \leq M$
Interpretation of results	Resolving kernel	Resolving kernel	Target kernel	Target kernel (+ Resolving kernels)
Contaminant kernels used?	No	Yes	No	Yes
Can produce model/model Proxy	Model proxy	Model proxy	Model	Model
References	Backus & Gilbert (1967b, 1968a, 1968b, 1970)	Oldenburg (1981), Pijpers & Thompson (1994), Masters & Gubbins (2003), Zaroli (2016, 2019)	Backus (1970a, b, c), Parker (1977), Al-Attar (2021)	This contribution

Table A2. Table summarizing the main mathematical symbols used in the manuscript. Elements are grouped on columns and rows depending on the relationships between them. For example, $\bar{m}$ is part of the model space $\mathcal{M}$ and is related to $\bar{d}$ through G, which is determined by K. Similarly, $\tilde{p}$ is related to $\bar{m}$ through $\mathcal{R}$ and to $\tilde{m}$ through $\mathcal{T}$.

$\mathcal{M}$ Model space	$\mathcal{D}$ Data space	$\mathcal{P}$ Property space	Mapping	Kernels
$\bar{m}$: TRUE model	$\bar{d}$: TRUE data		G: forward mapping	K: sensitivity kernels
$\bar{m}$: TRUE model		$\bar{p}$: TRUE property	$\mathcal{T}$: property mapping	T: target kernels
$\bar{m}$: TRUE model		$\tilde{p}$: approximate property	$\mathcal{R}$: approximate mapping	R: resolving kernels
$\tilde{m}$: least norm model		$\tilde{p}$: approximate property	$\mathcal{T}$: property mapping	T: target kernels
		ϵ: property error	$\mathcal{H}$: hyperellipsoid matrix	
j: physical parameter index	i: data index	k: property index		
N_m: number of physical parameters	N_d: number of data			N_p : number of properties

A2 Choice of forward mapping G

We are now left with making the two most important decisions. First, we need to decide whether G is a linear or non-linear mapping. Most often, the forward problem is non-linear, which leads to complicated inference problems. While there is some work on this front in the inference field (Snieder 1991), the problem is generally too difficult to tackle analytically or requires vast computational resources. For this reason, inference problems usually assume G to be linear and bounded (and therefore continuous when dealing with normed spaces), resulting in linear inferences.

For linear inferences, we can delve deeper into the structure of G. If $\mathcal{M}$ is a Hilbert space, G is often defined as a vector of inner products with some known members of $\mathcal{M}$ (commonly referred to as sensitivity kernels in seismology) (e.g. Backus 1970a). SOLA methods, for instance, specifically use the L_2 inner product on the model space $L_2[\Omega]$ with Ω some spatial domain, leading to the well-known form of the forward operator seen in eq. (9).

More complicated linear forward mappings can be used if we consider that data frequently depend on multiple physical parameters, such as shear and compressional wave speeds, as well as density. These problems can be addressed by considering $\mathcal{M}$ as a direct sum of Hilbert spaces, which itself forms a Hilbert space (e.g. Lau & Romanowicz 2021), and a forward mapping of the form given in eq. (34). In this paper, we argue that under such choices, the analysis provided by SOLA methods offers insights that are not readily accessible through DLI methods alone.

A3 Without prior information

The last, and arguably the most important decision, concerns prior information. We note that the choices of $\mathcal{M}$, $\mathcal{D}$ and G already encode some level of prior information. However, when referring to additional prior information, we assume that these choices have already been fixed. If we decide not to use any additional prior information, we would follow along the route of MOLA/SOLA methods (e.g. Backus & Gilbert 1970; Oldenburg 1981; Pijpers & Thompson 1994; Zaroli 2016). For these methods, it can be shown that we typically cannot directly infer $\mathcal{T}(\bar{m}) = \bar{p}$. Instead of obtaining the properties of interest, we must settle for approximate properties $\mathcal{R}(\bar{m})$ (see Fig. A1). Given only the data values and geometry, the goal is then to construct an approximate mapping $\mathcal{R}$ such that:

$$\mathcal{R} = XG, \quad \text{where } \mathcal{R} : \mathcal{M} \to \mathcal{P} \text{ and } X : \mathcal{D} \to \mathcal{P}.$$

In essence, this involves determining the $N_d \times N_p$ elements of the X mapping. The original method of Backus and Gilbert (Backus & Gilbert 1967a, b, 1968b, a, 1970) proposed an approximate mapping designed to obtain the highest-resolution local averages at N_p points of the unknown model. This mapping is obtained by minimizing the cost functions for each point one by one (see Fig. A1):

$$\int_\Omega (J^{(k)} A^{(k)})^2 \mathrm{d}\Omega, \quad s.t. \int_\Omega A^{(k)} = 1$$

where $J^{(k)}$ is a weight function with increasing weight further away from the k-th point where maximum resolution is desired. In contrast, for SOLA Pijpers & Thompson (1994) constructed an approximate unimodular mapping that resembles the predefined $\mathcal{T}$ by minimizing the cost function:

$$\mathrm{Tr}[(\mathcal{T}-\mathcal{R})(\mathcal{T}-\mathcal{R})^*], \quad \text{subject to } \mathcal{R}(1)=1,$$

where 1 represents the constant function on the left-hand side and the N_p-dimensional vector of ones on the right-hand side. This is the most generic formulism, a more specific form is given in Fig. A1. It turns out that solving the SOLA optimization problem is computationally more efficient than solving the optimization problem needed for the original Backus–Gilbert method (Pijpers & Thompson 1992; Al-Attar 2021). This is due to the fact that the matrix to be inverted for SOLA depends only on the sensitivity kernels, while for the Backus–Gilbert method it depends on both the sensitivity kernels and the spatially dependent functions J^k, and thus needs to be inverted again for each k-th point.

For linear inferences on a Hilbert model space, where the forward mapping is defined via projections onto sensitivity kernels, the approximate mapping $\mathcal{R}$ is associated with resolving kernels. These resolving kernels offer valuable insights into the interpretation of the approximate properties $\mathcal{R}(\bar{m})$. Even if data noise is ignored, the resolving kernels will be imperfect due to data incompleteness and trade-offs between physical parameters. The shape of these resolving kernels thus provides information about data limitations and parameter trade-offs.

A4 With additional prior information

If we introduce additional prior information, we have the choose between soft and hard prior information Backus (1988b). Hard priors are those where we assume that the true model must lie with 100 per cent certainty within a subset of the model space.

The norm bound used in this paper is an example of a hard quadratic bound (see more here, Backus 1989). In this paper, we refer to linear inferences with hard prior information as DLI methods (Deterministic Linear Inferences) due to the deterministic nature of the prior information (Al-Attar 2021). However, in the literature, these kind of problems are also referred to as Confidence Interval Sets as the solutions are intervals in which the property is found (Backus 1988a).

Soft bounds can be imposed, for example, by introducing a soft prior via a regularization term that penalizes some undesired feature of the model (for example, penalizing large norms). Soft priors via penalty terms may be considered 'softer' from the perspective that we prefer some models to be penalized more than others, based on our belief that they are less likely to be true. However, in practice, an optimization process is carried out, resulting typically in a single solution. This solution strikes a balance between fitting the data and minimizing the penalty. However, this is a single solution, and the act of optimizing a penalized cost function effectively collapses the model space to a single point (thus solving the problem of non-uniqueness) and will reject any other model. In contrast, a hard prior will immediately remove some models that are deemed unacceptable, but it will usually keep many others that are deemed acceptable, without discriminating between them. Therefore, a hard prior will be more inclusive and less stringent than a penalty-based soft prior.

Soft prior assumptions and similar regularizations are essential in inversion methods, as they help address non-uniqueness, which cannot be resolved without such constraints. This is why inference methods like DLI tend to have lower precision–their assumptions are too weak to break the non-uniqueness. Essentially, these methods trade precision for accuracy, as weaker assumptions reduce the likelihood of introducing bias into the solution.

Another way to impose a soft bound is by making the model space a probabilistic space with a measure to describe our prior knowledge. This would eliminate some sets of models that have zero probability, but it will usually keep many models, giving higher probability to some compared to others. Overall, we believe that the hard priors used in this paper are less stringent than penalty-based soft priors, but more stringent than probabilistic soft priors.

APPENDIX B: SURJECTIVITY of G

An inverse problem requires three components: model space $\mathcal{M}$, data space $\mathcal{D}$ and a forward relation $G : \mathcal{M} \to \mathcal{D}$. In our case, let $\mathcal{M} = L_2(\Omega)$, a Hilbert space defined on a compact domain $\Omega \subset \mathbb{R}$, and $\mathcal{D} = \mathbb{R}^{N_d}$ for some $N_d \in \mathbb{N}$. As discussed in Section 2.1.3 (eq. 9), the forward relation is defined as:

$$[G(m)]_i = \int_\Omega K_i(x) m(x)\,\mathrm{d}x. \tag{B1}$$

with $x \in \Omega$. To demonstrate that G is surjective, we utilize its dual G'. The dual space $\mathcal{M}'$ consists of linear forms $m' \in \mathcal{M}'$ defined on $\mathcal{M}$ that map elements $m \in \mathcal{M}$ to $\mathbb{R}$:

$$m' : \mathcal{M} \to \mathbb{R}$$

Since $\mathcal{M}$ is a Hilbert space, the Riesz Representation Theorem establishes an isomorphism $\mathcal{L}_\mathcal{M} : \mathcal{M}' \to \mathcal{M}$. This means that for each $m \in \mathcal{M}$, there exists a unique $m' \in \mathcal{M}'$ such that

$$\mathcal{L}_\mathcal{M}(m') = m.$$

Similarly, the dual space of the data space $\mathcal{D}$ is $\mathcal{D}'$ with an isomorphism $\mathcal{L}_\mathcal{D} : \mathcal{D}' \to \mathcal{D}$. If $G : \mathcal{M} \to \mathcal{D}$, then the dual mapping is defined by

$$d'(G(m)) = (G'(d'))(m), \forall m \in \mathcal{M}, \forall d \in \mathcal{D}$$

Rearranging gives:

$$d'(d) = m'(m).$$

We can express the relationship between the spaces as:

$$m' = \mathcal{L}_\mathcal{M} \circ G' \circ \mathcal{L}_\mathcal{D}^{-1}(d).$$

We can define G^*, the adjoint of G, as:

$$G^* = \mathcal{L}_\mathcal{M} \circ G' \circ \mathcal{L}_\mathcal{D}^{-1}. \tag{B2}$$

An equivalent definition states:

$$\langle G(m), d\rangle_\mathcal{D} = \langle m, G^*(d)\rangle_\mathcal{M}, \forall d \in \mathcal{D}, m \in \mathcal{M} \tag{B3}$$

where $\langle\cdot,\cdot\rangle_\mathcal{M}$ denotes the inner product on $\mathcal{M}$ and similarly $\langle\cdot,\cdot\rangle_\mathcal{D}$ is the inner product on $\mathcal{D}$. This implies

$$G^*(d) = \sum_i^{N_d} d_i K_i(x).$$

where $x \in \Omega$. We can now prove the surjectivity of G.

Proof. According to Proposition 2.1 from Al-Attar (2021), G is surjective if and only if $\ker(G') = \{0\}$, where

$$\ker(G') = \{d' \in \mathcal{D}' \mid G'(d') = 0\}.$$

Using the relation between the dual and the adjoint of G (eq. B2), we find that:

$$G' = \mathcal{L}_{\mathcal{M}}^{-1} G^* \mathcal{L}_{\mathcal{D}}$$

where we have omitted the composition symbol '$\circ$'. Therefore, we have to show that:

$$\ker(\mathcal{L}_{\mathcal{M}}^{-1} G^* \mathcal{L}_{\mathcal{D}}) = \{0\}.$$

We will show this by assuming the contrary and showing that it leads to a contradiction. Let us assume that there exists a $d' \in \mathcal{D}'$, $d' \neq 0$ such that $\mathcal{L}_{\mathcal{M}}^{-1} G^* \mathcal{L}_{\mathcal{D}}(d') = 0 \in \mathcal{M}'$. We know that $\mathcal{L}_{\mathcal{D}}$ is an isomorphism, therefore $\mathcal{L}_{\mathcal{D}}(d') \in \mathcal{D}$ and $\mathcal{L}_{\mathcal{D}}(d') \neq 0$. Applying the adjoint of G, we have:

$$G^*(\mathcal{L}_{\mathcal{D}}(d')) = \sum_i^{N_d} [\mathcal{L}_{\mathcal{D}}(d')]_i K_i(x)$$

However, $\{K_i\}$ are linearly independent, therefore

$$G^*(\mathcal{L}_{\mathcal{D}}(d')) = \sum_i^{N_d} [\mathcal{L}_{\mathcal{D}}(d')]_i K_i(x) = 0 \in \mathcal{M} \text{ iff } \mathcal{L}_{\mathcal{D}}(d') = 0$$

which we already know is not the case. This means that $G^*(\mathcal{L}_{\mathcal{D}}(d')) \neq 0 \in \mathcal{M}$. Since $\mathcal{L}_{\mathcal{M}}^{-1}$ is bijective, the non-zero element $G^*(\mathcal{L}_{\mathcal{D}}(d'))$ gets mapped onto a non-zero element of $\mathcal{M}'$, which contradicts the initial assumption. □

As a final note, we want to emphasize that the approximate nature of the theory does not imply that G is not surjective (as we have proven above).

APPENDIX C: X FROM SOLA

For SOLA inferences, in the absence of unimodularity conditions, we want to solve:

$$\text{argmin}_{x_i^{(k)}} \left[\int_\Omega (T^{(k)} - \sum_i^{N_d} x_i^{(k)} K_i)^2 \mathrm{d}\Omega \right]. \tag{C1}$$

Mathematically, this is a multi-objective minimization problem, because we want to find $x_i^{(k)}$ that minimize concomitantly all squared differences between the targets and their corresponding averaging kernels. Since we give the same importance to each target-resolving kernel error we want to minimize, we can use the classic Pareto method. In other words, we try to minimize:

$$\text{argmin}_{x_i^{(k)}} \left[\sum_k^{N_p} \int_\Omega (T^{(k)} - \sum_i^{N_d} x_i^{(k)} K_i)^2 \mathrm{d}\Omega \right]. \tag{C2}$$

In practice, the minimization problem for each property can be solved independently for all other properties (mathematically equivalent to eq. C2). This leads to an embarrassingly parallel algorithm (Zaroli *et al.* 2017).

Using matrix calculus, the solution can readily be found. We first take the gradient of eq. (C2) and set it to zero:

$$\frac{\partial}{\partial x_l^{(q)}} \sum_k^{N_p} \left[\int_\Omega (T^{(k)} - \sum_i^{N_d} x_i^{(k)} K_i)^2 \mathrm{d}\Omega \right] = 0. \tag{C3}$$

Because

$$\frac{\partial}{\partial x_l^{(q)}} \left(T^{(k)} - \sum_i^{N_d} x_i^{(k)} K_i \right)^2 = -2 \left(T^{(k)} - \sum_i^{N_d} x_i^{(k)} K_i \right) \times \left(\sum_j^{N_m} \frac{\partial x_j^{(k)}}{\partial x_l^{(q)}} K_j \right)$$

Eq. (C3) becomes:

$$-2 \sum_k^{N_p} \int_\Omega (T^{(k)} - \sum_i^{N_d} x_i^{(k)} K_i) \left(\sum_j^{N_m} \frac{\partial x_j^{(k)}}{\partial x_l^{(q)}} K_j \right) \mathrm{d}\Omega = 0 \tag{C4}$$

It is obviously that:

$$\frac{\partial x_j^{(k)}}{\partial x_l^{(q)}} = \delta_{lj} \delta_{qk},$$

which can be substituted in eq. (C4) to obtain:

$$\int_\Omega (T^{(q)} - \sum_i^{N_d} x_i^{(q)} K_i) K_l \mathrm{d}\Omega = 0. \tag{C5}$$

Separating the integral in eq. (C5), we now get:

$$\sum_i^{N_d} x_i^{(q)} \int_\Omega K_i K_l \mathrm{d}\Omega = \int_\Omega T^{(q)} K_l \mathrm{d}\Omega.$$

This can be written in matrix form as:

$$X\Lambda = \Gamma,$$

where

$$\Lambda_{il} = \int_\Omega K_i K_l \mathrm{d}\Omega$$

$$\Gamma_{ql} = \int_\Omega T^{(q)} K_l \mathrm{d}\Omega$$

$$X_{qi} = x_i^{(q)}.$$

This finally gives us:

$$X = \Gamma \Lambda^{-1} \tag{C6}$$

which is equivalent to the eqs (18), (19) and (20).

APPENDIX D: ELEMENTS NEEDED FOR THE DERIVATION OF THE SOLA-DLI SOLUTION

In this appendix, we derive some further equations and equalities related to the material presented in Section 2.1.

D1 Data-model relationships

Given a model $m = (m^1, m^2, ...)$, the relationship between the model and data is defined by:

$$d_i = [G(m)]_i = \sum_j^{N_m} \left\langle K_i^j, m^j \right\rangle_{\mathcal{M}_j}. \tag{D1}$$

It is useful to define:

$$G^j(m) = \left\langle K_i^j, m^j \right\rangle_{\mathcal{M}_j} \tag{D2}$$

$$G = \sum_j^{N_m} G^j. \tag{D3}$$

The adjoint of G is defined in eq. (B3) and repeated for convenience:

$$\left\langle G(m), d' \right\rangle_{\mathcal{D}} = \left\langle m, G^*(d') \right\rangle_{\mathcal{M}} \tag{D4}$$

for all $m \in \mathcal{M}$ and $d' \in \mathcal{D}$. We expand the LHS:

$$\sum_i^{N_d} \sum_j^{N_m} \left\langle K_i^j, m^j \right\rangle_{\mathcal{M}_j} d'_i = \left\langle m, G^*(d') \right\rangle_{\mathcal{M}} . \tag{D5}$$

For the RHS, we use the formula for the inner product in the direct sum space $\mathcal{M}$:

$$\langle a, b \rangle_{\mathcal{M}} = \sum_j^{N_m} \left\langle a^j, b^j \right\rangle_{\mathcal{M}_j} , \tag{D6}$$

where a, b are some members of $\mathcal{M}$. Therefore, we write

$$\sum_i^{N_d} \sum_j^{N_m} \left\langle K_i^j, m^j \right\rangle_{\mathcal{M}_j} d'_i = \sum_j^{N_m} \left\langle m^j, G^{j*}(d') \right\rangle_{\mathcal{M}_j} . \tag{D7}$$

Taking the sum over i inside, we can also write:

$$\sum_j^{N_m} \left\langle \sum_i^{N_d} d'_i K_i^j, m^j \right\rangle_{\mathcal{M}_j} = \sum_j^{N_m} \left\langle m^j, G^{j*}(d') \right\rangle_{\mathcal{M}_j} \tag{D8}$$

and we identify:

$$G^{j*}(d') = \sum_i^{N_d} d'_i K_i^j , \tag{D9}$$

$$G^* = (G^{1*}, G^{2*}, ...). \tag{D10}$$

G^* maps elements from the data space to elements (tuples) in the model space. A similar approach shows that the adjoint of the property mapping $\mathcal{T}$ is given by:

$$\mathcal{T}^{j*}(p) = \sum_k^{N_p} p^{(k)} T^{j,(k)}, \tag{D11}$$

$$\mathcal{T}^* = (\mathcal{T}^{1*}, \mathcal{T}^{2*}, ...). \tag{D12}$$

D2 H matrix

The $\mathcal{H}$ matrix introduced in Section 2.1 quantifies the difference between the target and resolving kernels. It is defined by Al-Attar (2021, see eq. 2.84) as:

$$\mathcal{H} = HH^*, \tag{D13}$$

$$H = \mathcal{T} - \mathcal{R}, \tag{D14}$$

where $\mathcal{R}$ is the 'approximate mapping', given by:

$$\mathcal{R} = \mathcal{T} G^* (GG^*)^{-1} G. \tag{D15}$$

This mapping takes any model $m \in \mathcal{M}$ into the data space $d \in \mathcal{D}$, then finds the least norm solution to $G(m) = d$ and maps this least norm solution into the property space. When applied to one of the possible model solutions $U_M \bigcap S$ (see Fig. 1b), it gives the property of the model solution that has the smallest norm. Combining (D13), (D14) and (D15) we obtain:

$$\mathcal{H} = (\mathcal{T} - \mathcal{R})(\mathcal{T} - \mathcal{R})^*, \tag{D16}$$

$$\mathcal{H} = \mathcal{T}\mathcal{T}^* - \mathcal{T} G^* (GG^*)^{-1} G \mathcal{T}^*. \tag{D17}$$

Let us denote

$$\Lambda := GG^*. \tag{D18}$$

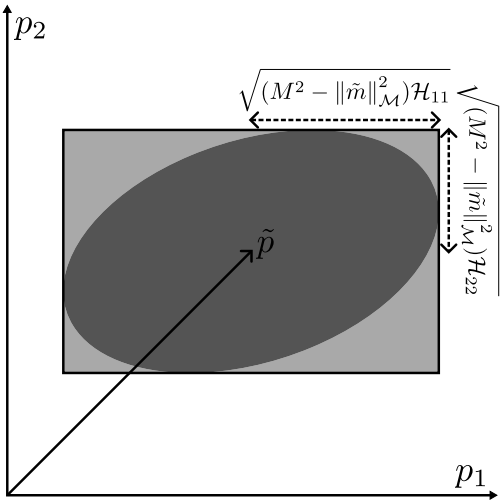


Figure A1. Illustration of the relationship between the hyperellipsoid defined in eq. (D25) and the hyperparallelepiped defined in eq. (D27) for the case when the property space is 2-D. p_1 and p_2 could represent, for example, two local averages at two different spatial locations. The dark shaded ellipse contains all the possible combinations of these two properties given by the tighter inequality of eq. (D25). The lighter grey shaded rectangle contains all the possible combinations of these two properties under the simplified inequality in eq. (D27).

Using a simple application of eqs (D1) and (D10), we can then easily obtain:

$$\Lambda_{iq} = \sum_j^{N_m} \left\langle K_i^j, K_q^j \right\rangle_{\mathcal{M}_j} . \tag{D19}$$

Similarly, we denote:

$$\chi := \mathcal{T}\mathcal{T}^*, \tag{D20}$$

$$\chi_{k,l} = \sum_j^{N_m} \left\langle T^{j,(k)}, T^{j,(l)} \right\rangle_{\mathcal{M}_j} \tag{D21}$$

and

$$\Gamma := \mathcal{T} G^*, \tag{D22}$$

$$\Gamma_{ki} = \sum_j^{N_m} \left\langle T^{j,(k)}, K_i^j \right\rangle_{\mathcal{M}_j} . \tag{D23}$$

Using the definitions of Λ, χ, Γ we can write (D17) as:

$$\mathcal{H} = \chi - \Gamma \Lambda^{-1} \Gamma^T. \tag{D24}$$

Fig. A1 provides a visualization of the ellipse in the property space as determined by $\mathcal{H}$ when only two properties are considered.

D3 Error bounds

The error bounds defined in eqs (23) and (24) are derived from the property bounds defined by Al-Attar (2021, see eq. 2.84) as:

$$\left\langle \mathcal{H}^{-1}(p - \tilde{p}), p - \tilde{p} \right\rangle_{\mathcal{P}} \leq M^2 - \|\tilde{m}\|^2_{\mathcal{M}} \tag{D25}$$

Eq. (D25) describes a hyperellipsoid centred on $\tilde{p}$ with major axes given by the eigenvalues of $\mathcal{H}^{-1}$ scaled by $\sqrt{M^2 - \|\tilde{m}\|^2_{\mathcal{M}}}$. If the matrix $\mathcal{H}$ is diagonal, then the inverse is trivial to find and the hyperellipsoid has its major axes aligned with the coordinate axes

of the property space. In all other cases, the hyperellipsoid will have some arbitrary orientation and $\mathcal{H}$ will be difficult to invert numerically.

To avoid numerical complications, we use here a different, more relaxed approximation for the error bounds given in the form:

$$\|p - \tilde{p}\|_{\mathcal{P}}^2 \leq (M^2 - \|\tilde{m}\|_{\mathcal{M}}^2)\text{diag}(\mathcal{H}) \tag{D26}$$

where $\text{diag}(\mathcal{H})$ is the diagonal of $\mathcal{H}$. We can also write this in component form:

$$\left\|p^{(k)} - \tilde{p}^{(k)}\right\|_{\mathcal{P}}^2 \leq (M^2 - \|\tilde{m}\|_{\mathcal{M}}^2)\mathcal{H}_{kk} \tag{D27}$$

Inequality (D27) describes a hyperparallelepiped that contains the error bounds of (D25) with sides parallel to the coordinate axes of $\mathcal{P}$ (see Fig. A1). As this approximation does not require the inversion of the $\mathcal{H}$ matrix, it is computationally advantageous. Visually, the hyperellipsoid fits 'perfectly' inside the hyperparallelepiped (Fig. A1), but the error bounds of the hyperparallelepiped are easier to visualize in a static plot (see for example first column of Fig. 7). The hyperellipsoid encodes the correlations between the error bounds of the various components of the property vector (such as the correlation between the error bounds of two different local averages). Plotting the bounds for each component of the property vector simultaneously would therefore be very difficult, since the error bounds of each property component would depend on the values of the bounds on all other property components. The hyperparallelepiped ignores these correlations, simplifying thus the plotting. However, it overestimates the property bounds, which will likely make it more difficult to interpret the property values.

To show how (D27) arises from (D25), we need to prove the following:

Given that

$$x^T A^{-1} x \leq b \tag{D28}$$

Show that

$$x_k^2 \leq b A_{kk} \tag{D29}$$

where $x = \tilde{p} - \epsilon$, $A = \mathcal{H}$ and $b = M^2 - \|\tilde{m}\|_{\mathcal{M}}^2$. To prove this, we start by finding the maximum extent of the hyperellipsoid (D28) along the k^{th} coordinate axis, which can be described mathematically as:

$$\text{Find } \max(c^T x) \tag{D30}$$

$$\text{Given that } x^T A^{-1} x \leq b. \tag{D31}$$

where c^T will be chosen later to be a vector with all entries 0 except the k^{th} one. We shall use the Lagrangian approach to solve this problem. We introduce the slack constant s and use it to transform the inequality (D31) into an equality (slack constraint):

$$x^T A^{-1} x - b + s^2 = 0. \tag{D32}$$

Let λ be a Lagrange multiplier. The problem then becomes finding the extremum points of the Lagrangian:

$$f(x, \lambda) = c^T x + \lambda(b - x^T A^{-1} x - s^2). \tag{D33}$$

Differentiating f with respect to x and setting the result to zero leads to:

$$c - 2\lambda A^{-1} x = 0. \tag{D34}$$

Note that $\lambda = 0$ leads to $c = 0$, which is a contradiction. Therefore we must have $\lambda \neq 0$ and $s^2 = 0$, which means that our constraint is active. Assuming A^{-1} to be invertible, we obtain:

$$x = \frac{Ac}{2\lambda}. \tag{D35}$$

We next differentiate f with respect to λ (using $s^2 = 0$ since we have shown the constraint to be active) to obtain the second Lagrange equation. Setting the result equal to zero leads to:

$$b - x^T A^{-1} x = 0. \tag{D36}$$

Substituting (D35) into (D36) and rearranging for b, we obtain:

$$b = \left(\frac{Ac}{2\lambda}\right)^T A^{-1} \frac{Ac}{2\lambda}. \tag{D37}$$

Since A is symmetric this leads to:

$$\lambda^2 = \frac{c^T Ac}{4b}. \tag{D38}$$

Assuming that A is positive definite (its eigenvalues give the lengths of the hyperellipsoids' major axes), we must have

$$\lambda = \frac{\sqrt{c^T Ac}}{2\sqrt{b}}. \tag{D39}$$

Finally, using (D38) and (D35) the optimal vector solution x can be expressed for any vector c as:

$$x = \frac{Ac}{2\lambda} = \frac{Ac}{\frac{\sqrt{c^T Ac}}{\sqrt{b}}} = \frac{\sqrt{b}Ac}{\sqrt{c^T Ac}}, \tag{D40}$$

and the maximal value of $c^T x$ is thus:

$$\sqrt{bc^T Ac}. \tag{D41}$$

Now, we consider a fixed index k between 1 and N, and we define c to be the following vector:

$$c := (\delta_{ik})_{1 \leq i \leq N}, \tag{D42}$$

where δ_{ik} is 1 if $i = k$, and 0 if $i \neq k$. Substituting this for c in (D41), we obtain:

$$\max(x_k) = \sqrt{bA_{kk}} \tag{D43}$$

or equivalently:

$$x_k \leq \sqrt{bA_{kk}} \tag{D44}$$

If instead we choose $c = -\delta_{ik}$, then we have:

$$x_k \geq -\sqrt{bA_{kk}} \tag{D45}$$

These two inequalities can be summarized in the final answer:

$$(x_k)^2 \leq bA_{kk} \tag{D46}$$