Backus Gilbert SOLA Targets as Filters

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

[1]Department of Earth Sciences, University of Oxford, Oxford, UK
[2]Institut Terre et Environnement de Strasbourg, Université de Strasbourg, EOST, CNRS, UMR 7063, 5 rue Descartes, Strasbourg F-67084, France

Abstract

Linear inferences based on Backus-Gilbert SOLA theory aim to extract information about properties of the unknown model rather than the model itself. In seismic tomography we may be interested in properties such as local averages or local gradients of the unknown tomographic model. The classic Backus Gilbert SOLA theory has been used so far to extract various kinds of local averages using target kernels that correspond to averaging filters. In this paper the target kernels are extended to more general integral kernels. This expands the ability of Backus Gilbert SOLA to extract information about general filters of the unknown model. The new theory, as well as the classic theory, is presented through the deterministic linear inference formalism developed by Al-Attar (2021). We apply the new theory to 1D, error-free, synthetic cases. We also demonstrate the ability of normal mode data to constrain the structure in the Earth.

1 Introduction

Seismic tomography relies on mathematical inversions (Rawlinson, Pozgay, and Fishwick (2010), Nolet (2008)) to model Earth's interior from collected data. These models highlight persistent features—like subducted plates, plumes, boundaries, or LLSVPs—believed to mirror real Earth characteristics (McNamara (2019), Koelemeijer (2021)). Improving these models often involves creating new ones with different data or methods to enhance feature resolution or reduce uncertainties.However, seismic inversions encounter a major challenge: data scarcity. This leads to non-uniqueness in solutions (G. Backus and F. Gilbert 1970) which can be mitigated using regularization. Yet, such information injections might inadvertently impose unrealistic constraints (Nolet (2008), Zaroli, Koelemeijer, and Lambotte (2017)). While incorporating new information isn't inherently wrong, it must be accurate and well-understood to avoid misinterpretations in the resulting models.

The data often provides insufficient constraints for the entire model but can offer useful constraints for certain parts or properties. Seeking specific properties rather than inverting the entire model falls under the domain of mathematical inference. If the complete model is known, all its properties are also known, so trying to invert for the entire model aims to constrain all its properties, leading to compromises. Directly constraining a set of properties ensures maximizing data use for the specific task. This is the core idea behind the Backus-Gilbert SOLA linear inference method.

The Backus-Gilbert SOLA method was originally used to compute local averages of the unknown model only from the data, instead of the model itself, without the need for regularization (G. E. Backus and J. F. Gilbert 1967, G. Backus and F. Gilbert 1968, G. Backus and F. Gilbert 1970). The idea was extended by Backus to compute more general "properties" of the unknown model by fitting linear combinations of the sensitivity kernels to some other predefined kernel (later named "Target kernel" by some authors) (G. Backus 1970a, G. Backus 1970b). Backus does not go deeper into what kind of properties one could analyze with this method, but he does realize that data alone cannot constrain sufficiently any such property and therefore uses a norm bound on the model space to obtain some useful information about the desired properties. These ideas have been popularized and practically applied by several authors for topics such as helioseismics (e.g. Pijpers and Thompson 1994, Pijpers and Thompson 1992, who also coined the name SOLA), deconvolutions (Oldenburg 1981), and seismic tomography (e.g. Zaroli 2016, Zaroli, Koelemeijer, and Lambotte 2017, Zaroli 2019, Amiri et al. 2023, Restelli, Zaroli, and Koelemijer 2023, Masters and Gubbins 2003), but mostly concentrated on extracting local average information about the unknown model. A notable exception is the paper of Pijpers and Thompson (1994) who used some Gaussian derivatives to extract information about gradients. Past studies also mostly overlooked the importance of placing a norm bound on the model space. Recently, Al-Attar (2021) generalized those methods and developed a rigorous mathematical framework where the model space, data space, and property space are Banach or Hilbert spaces where the importance of the model norm bound is highlighted.

This paper centers on the idea that different kinds of properties of some unknown model can be constrained by the data and some prior model norm bound information. We will concentrate, however, on how dif-

ferent types of information can be extracted by choosing appropriate target kernels for the linear combination of the sensitivity kernels. The next section will provide a brief overview of the mathematical setting for general deterministic linear inferences. The third section specializes to the error-free data case which depends on a single continuous model parameter. The most important equations are presented without derivation and minimal interpretation (the reader is encouraged to read the references for those equations). The idea of viewing the classic BG SOLA method as a filtering of the unknown model is introduced and two types of such filters are presented: local averages and gradients. Finally, we apply these ideas to normal mode data and show how different types of filters may be better suited for extracting certain information about the unknown model.

2 Linear Inferences

Let's assume that we have acquired some data d that are related to some model m via a linear relation G such that:

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

Most model-data relations in seismology are not linear, but they can be linearised using perturbation theory (Tromp, Tape, and Liu 2005). The model is part of the model space $m \in \mathcal{M}$, and the data is part of the data space $d \in \mathcal{D}$. In inversions we try to find the model from the data by inverting the forward relation (1). However, in most cases, the forward relation cannot be inverted due to insufficient or inadequate data. For continuous models, this leads to either no solutions, or infinitely many solutions (G. E. Backus and J. F. Gilbert 1967). In the absence of measurement errors, the only case when there are no solutions is when the data is outside the range of the forward operator. This means that the data is incompatible with the forward operator/physical law. If we find ourselves in such a situation we should use a different forward relation. Therefore we will assume that the data is compatible with the forward relation, and that we have infinitely many solutions (see Fig. (1). We will denote the set of solutions by $G^{-1}d$ (should not be confused with the inverse of G, which does not exist).

Inversions can be carried out by adding constraints (regularisations) to the model space $\mathcal{M}$ until a single model $\tilde{m}$ is "selected". For example, one may choose as the solution the model from $G^{-1}d$ that has the smallest average gradient, which corresponds in some sense to the smoothest model. If the assumptions built into the regularisations are not correct, then the model will not be a good representation of reality.

We often seek specific properties of the true model $\bar{m}$ rather than the entire model itself. These properties, such as average physical values in Earth regions or the presence of discontinuities at specific depths, belong to a distinct space known as the property space $\mathcal{P}$ (Al-Attar 2021).

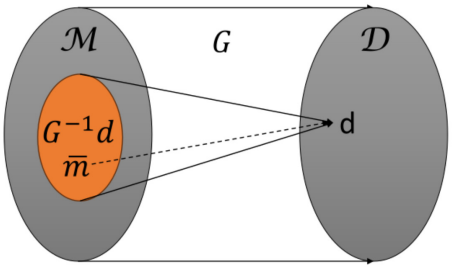
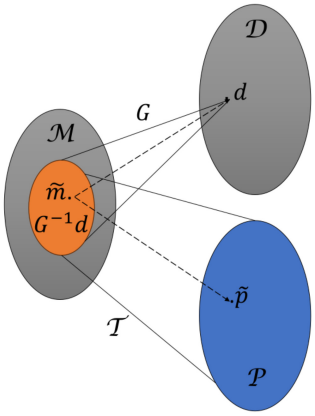
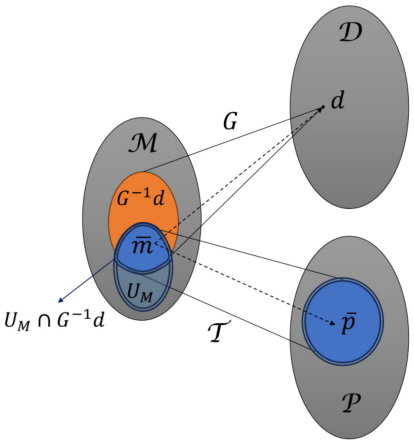


Figure 1: Schematic representation of the data space $\mathcal{D}$ and model space $\mathcal{M}$. G is surjective but not injective. This means that all of $\mathcal{M}$ gets mapped onto all of $\mathcal{D}$, however, there may be multiple elements of $\mathcal{M}$ that get mapped onto the same d. The set of all m such that $G(m) = d$ is called the pre-image of d and is denoted by $G^{-1}d$. The true model $\bar{m}$ is somewhere in $G^{-1}d$.



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



(b) Norm bound applied on the model space.

Figure 2: Representation of the effect of bounding the model space. Sets with thin lines for margins represent unbounded sets, while sets with double line margins represent bounded sets. When no bounds are imposed, the property of the true model may take any value in the property space, which is an unbounded set. Applying the norm bound on the model space leads to the intersection between $G^{-1}d$ and U_M to be bounded, which gets mapped under $\mathcal{T}$ to a bounded subset of $\mathcal{P}$.

Focusing on specific properties rather than the entire model aims to achieve more precise constraints.Therefore, our inference problem will be:

$$\begin{gathered}\text{Given that:}\\ G(m) = d\\ \text{Find:}\\ \mathcal{T}(\bar{m}) = \bar{p}\end{gathered} \tag{2}$$

where $\mathcal{T}$ is a linear map that represents a property of any model, and $\bar{p} \in \mathcal{P}$ represents the value of the property $\mathcal{T}$ when applied on the true model $\bar{m}$. It can be shown that in most practical situations the desired property $\bar{p}$ can be anything (Al-Attar 2021). In other words, given the data constraint, $\bar{p}$ may take any value from the property space $\mathcal{P}$ (see Fig. (2a)), and therefore we cannot say anything about the property $\mathcal{T}$ of the true model $\bar{m}$. This problem can be overcome by introducing a norm bound M in the model space (Al-Attar 2021):

$$\|m\|_{\mathcal{M}} \leq M \tag{3}$$

This is a constraint applied on the model space that differs from the constraints applied for regularisations because it does not aim to single out a model. This condition does not lead to an invertible system, but rather limits the possible solutions to a bounded subset of $\mathcal{M}$. If the set of models that respect the norm bound (3) is denoted U_M, then the set of solutions that respect the norm bound constraint and the data constraint is $U_M \cap G^{-1}d$, which is a bounded set. Al-Attar (2021) showed that such a constraint leads to the true property $\bar{p}$ being constrained in a bounded subset $P_M \subset \mathcal{P}$ obtained from mapping $U_M \cap G^{-1}d$ through $\mathcal{T}$ (see Fig. (2b)). All of the properties found in the bounded set P_M are equally likely to represent the true property $\bar{p}$ without any other prior information. If we choose the right property $\mathcal{T}$, we will be able to constrain the set of possible properties sufficiently for us to infer useful information about $\bar{p}$. Choosing the right $\mathcal{T}$ can be seen as finding the observable that can be best constrained by the data. Most importantly, this does not require us to introduce more information into the system, but rather puts us in the position of finding what the current information can constrain.

3 Classic SOLA Theory through Linear Inference Formalism

As mentioned before, many relations between the model and data in seismic tomography can be linearised and represented in the absence of measurement errors in the form:

$$d_i = G(\{m^j\}) = \sum_j \int K_i^j m^j \tag{4}$$

where d_j are the data, $\{m_i\}$ are the various model parameters that the data depend on (such as density, shear wave speed, compressional wave speed, locations of discontinuities, etc.), and $\{K_j^i\}$ are all the sensitivity kernels associated with each m_i (e.g. Tromp, Tape, and Liu 2005, Liu and Tromp 2008, Fichtner, Bunge, and Igel 2006). For the scope of this paper we will only deal with the error free case and data that only depend on one model parameter m. In this case, (4) simplifies to:

$$d_i = G(\bar{m}) = \int K_i \bar{m} \tag{5}$$

The sensitivity kernels and the model are both defined on some region Ω of the Earth, which may be 1D, 2D, or 3D. The data are always real, and the integral is computed over the entire spatial domain where the model is defined.

The Backus-Gilbert SOLA theory developed by Pijpers and Thompson (Pijpers and Thompson 1994), Zaroli (Zaroli 2019, Zaroli 2016, Zaroli, Koelemeijer, and Lambotte 2017), and others (Oldenburg 1981) has been used to find approximations of the true Earth $\bar{m}$ from linear combinations of the data constraints:

$$\tilde{p}^{(k)} = \int_\Omega A^{(k)}(\mathbf{r})\bar{m}(\mathbf{r})d\mathbf{r}^3 \tag{6}$$

where $\mathbf{r}$ is spatial position, and $A^{(k)}$ are called resolving kernels and are obtained from linear combinations of the sensitivity kernels:

$$A^{(k)}(\mathbf{r}) = \sum_i x_i^{(k)} K_i(\mathbf{r}) \tag{7}$$

If one finds the coefficients $x_i^{(k)}$ such that $A^{(k)}$ resemble some local average weight function $T^{(k)}$, then $\tilde{p}^{(k)}$ will represent approximate local averages of the true model. $T^{(k)}$ are called target kernels and represent the ideal resolving kernels we would want to obtain. In practice, it is impossible to obtain $A^{(k)} = T^{(K)}$, and in many cases $A^{(k)}$ may be nowhere close to resembling $T^{(k)}$, because the target kernels are independent of the sensitivity kernels. Therefore, in general:

$$\begin{gathered}\text{Approximate property of true model}\\ \tilde{p}^{(k)} = \mathcal{A}(\bar{m}) = \int_\Omega A^{(k)}(\mathbf{r})\bar{m}(\mathbf{r})d\mathbf{r}^3\end{gathered} \tag{8}$$

$$\begin{gathered}\text{True property of true model}\\ \bar{p}^{(k)} = \mathcal{T}(\bar{m}) = \int_\Omega T^{(k)}(\mathbf{r})\bar{m}(\mathbf{r})d\mathbf{r}^3\end{gathered} \tag{9}$$

$$\begin{gathered}\text{and}\\ \tilde{p}^{(k)} \neq \bar{p}^{(k)} \neq \bar{m}(\mathbf{r}^{(k)})\end{gathered} \tag{10}$$

where $\mathcal{A}$ is defined by (8) and is called the approximate property/averaging mapping. If we place a norm bound on the model space as discussed previously:

$$\|m\|_{\mathcal{M}} \leq M \tag{11}$$

then we find that the true property of the true model $\bar{p}$ must lie somewhere in the bounded property subset P_M defined by (Al-Attar 2021):

$$\begin{aligned}P_M = \{p \in \mathcal{M} |\, &\left\langle (HH^*)^{-1}(p-\tilde{p}), p-\tilde{p}\right\rangle_{\mathcal{P}}\\ &\leq M^2 - \|\tilde{m}\|^2_{\mathcal{M}}\}\end{aligned} \tag{12}$$

where $\langle\cdot,\cdot\rangle_{\mathcal{M}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{P}}$ are the inner products defined on the model space and property space respectively. $\tilde{m}$ is the least norm solution to (5), and

$$H = \mathcal{T} - \mathcal{A} \tag{13}$$

which we will call the deterministic error operator. H quantifies how different the true property/target mapping and the approximate property/averaging mapping are. Similarly, we will define $HH^* = \mathcal{H}$ and call it the deterministic hyperellipsoid matrix. The deterministic hyperellipsoid matrix is related to the resolution misfit RM used in other studies (Restelli, Zaroli, and Koelemijer 2023). The least norm solution of (5) is given by the Moore-Penrose inverse (Al-Attar 2021):

$$\tilde{m} = G^*(GG^*)^{-1}d \tag{14}$$

where * denotes operator adjoint. Equation (12) is a quadratic form in p and describes a hyperellipsoid in the property space centered on $\tilde{p}$ (Al-Attar 2021). The exact shape of the hyperellipsoid is encoded in the matrix $\mathcal{H}$ (the eigenvalues of $\mathcal{H}$ are the principal axes squared). The RHS acts as a scaling on the hyperellipsoid. Therefore, through the classic theory of SOLA we do not obtain a true average of the true model, but we can obtain a bounded subset that contains the true average of the true model. If the sensitivity kernels can be combined such that the resolving kernels resemble the target kernels well, then the bounded set P_M will be very small and we will be able to constrain the average of the true model very well. The property bound defined by the hyperellipsoid in (12) can be further bounded in a bigger hyperparallelepiped defined by

$$\left\{p \in \mathcal{P} | (p_i - \tilde{p}_i)^2 \leq \left(M^2 - \|\tilde{m}\|^2_{\mathcal{M}}\right)\mathcal{H}_{ii}\right\} \tag{15}$$

where $\mathcal{H}_{ii}$ denotes the diagonal elements of $\mathcal{H}$. The bounding hyperellipsoid of (12) is more difficult to visualize and work with numerically, so the bigger (more pessimistic) hyperparallelepiped (15) bound is usually preferred in practice. The model norm bound M is assumed known and the norm of the least square model $\|\tilde{m}\|_{\mathcal{M}}$ is independent of the target kernels, so only $\mathcal{H}$ depends on the target kernels, and can be computed using:

$$\mathcal{H} = \chi - U(\Lambda^{-1})U^T \tag{16}$$

$$\chi_{ij} = \int_\Omega T^{(i)}T^{(j)}d\mathbf{r}^3 \tag{17}$$

$$\Lambda_{ij} = \int_\Omega K_i K_j d\mathbf{r}^3 \tag{18}$$

$$U_{ij} = \int_\Omega T^{(i)}K_j d\mathbf{r}^3 \tag{19}$$

3.1 Target Kernels as Filters

A widely used target kernel is the boxcar function because its interpretation is very simple: it provides the local uniform average of the true model around some point of interest $\mathbf{r}^{(k)}$. This choice is very intuitive if one aims to use the BG SOLA theory as an inversion technique. However, the solution $\tilde{p}^{(k)}$ is not part of the model space $\mathcal{M}$ and therefore it should not be considered a model solution. Instead, $\tilde{p}^{(k)}$ is part of the property space $\mathcal{P}$, and can be interpreted as a filtered and discretized version of the true model. When the target kernels are boxcar functions, or Gaussian functions, what we obtain is a blurred version of the true model evaluated at some points $\mathbf{r}^{(k)}$ (see Fig. (4)). If, however, the target kernels are the derivative of a Gaussian, then $\tilde{p}^{(k)}$ will represent Gaussian smoothed local derivatives of the true model evaluated at some points $\mathbf{r}^{(k)}$. More generally, we can think of the target kernels as filtering kernels, similar to the kernels used in image processing. Some filters provide smoothing, others gradients, others pick up edges, and others may be used to emphasise other specific shapes and features. One may even try to obtain a Fourier or wavelet decomposition of the true model by using Fourier or wavelet base functions as target kernels. We can therefore generalise BG SOLA as a method for obtaining properties of the true model in the form of filters applied on the true model.

The idea of using different target kernels to extract different types of information about the unknown model has been applied before in helioseismics (Pijpers and Thompson 1994). They used Gaussian target kernels for extracting average information, and derivatives of the Gaussian for first and higher order derivatives. However, as far as we are aware this method has not been used yet in seismic tomography. More importantly, the work of Pijpers and Thompson uses the method as an inversion, whereas we believe it should be considered an inference method instead.

In the next two subsections I will present several types of resolving kernels that can be used for extracting local average information and several resolving kernels used for extracting gradient information.

3.1.1 Averaging filters

Model averages can be obtained from different target kernels such as:

Uniform Local Average:

$$T^{(k)} = \begin{cases} C & \mathbf{r}^{(k)} \in V_k \\ 0 & \text{else} \end{cases} \tag{20}$$

Gaussian Local Average:

$$T^{(k)} = C\exp\left[-\frac{\left\|\mathbf{r}-\mathbf{r}^k\right\|_2^2}{2\sigma^2}\right] \quad \mathbf{r}\in\Omega \tag{21}$$

Bump Function Average:

$$T^{(k)} = \begin{cases} C\exp\left[\frac{-1}{1-\sigma^2\left\|\boldsymbol{r}-\boldsymbol{r}^{(k)}\right\|_2^2}\right] & \mathbf{r}\in V_k' \\ 0 & else \end{cases} \tag{22}$$

where V_k is some predefined subset of the model domain Ω, $V_k' = \{\mathbf{r}\in\Omega|\ \left\|\boldsymbol{r}-\boldsymbol{r}^{(k)}\right\|_2 < \sigma^{-1}\}$ are the compact supports of the bump functions, and C is in each case an appropriate normalisation constant that ensures the unimodularity of each target kernel. 1D versions of these averaging kernels have been plotted in Fig. (3).

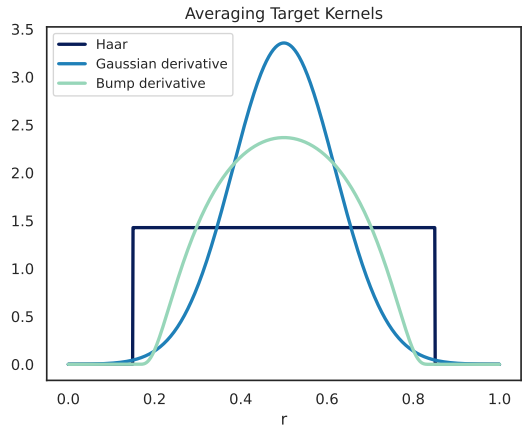


Figure 3: Examples of three averaging target kernels. The bump and boxcar kernels have a compact support width of 0.7.

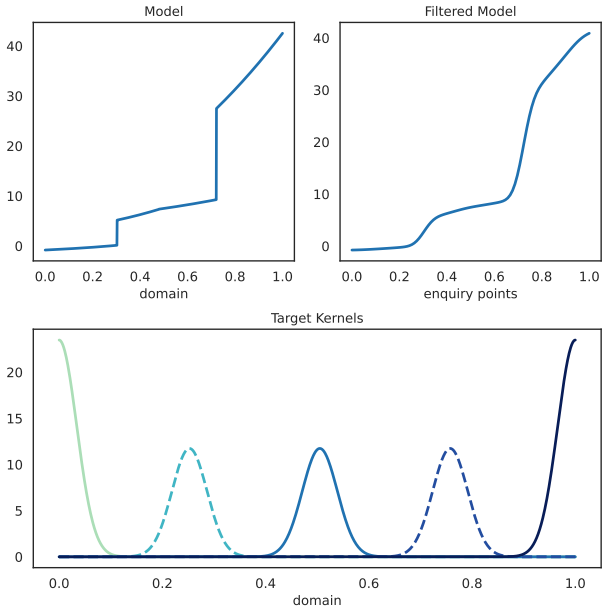


Figure 4: Example of Gaussian average filtering. Top left: original model, top right: filtered model, bottom: examples of the target kernels at different locations. Notice the clipped and scaled averaging kernels at the domain boundaries: such targets would not be accepted in our analysis. Alternating line styles have been used to distinguish between target kernels.

The uniform local average is the easiest to interpret and has a compact domain, however, it is difficult to obtain from linear combinations of sensitivity kernels due to the sharp corners (see Fig. (9)). Gaussian local averages are usually better recovered due to their smoothness that is often also characteristic of seismic sensitivity kernels. Unfortunately, Gaussian kernels do not have a compact support, so the definition of resolution associated with them is more subjective, leading to more difficult interpretations. Bump functions have compact support and are smooth (infinitely differentiable), thus offering some of the advantages of both Gaussian averages and uniform averages. Near the boundaries of the domain, the target kernels will be clipped, therefore changing their interpretation (a clipped Gaussian weight function will no longer provide a Gaussian average when applied on a function). For target kernels with a compact support of width w the only regions affected by the clipping will be one width away from the edges. Targets that are defined on the entire domain, such as Gaussians, will theoretically be affected everywhere as they will always be clipped. However, in practice the effect will only be noticeable near the edges because of the rapid fall-off of the function. We will abstain from constraining or trying to interpret the properties obtained from clipped target kernels in the case of targets with compact support. For targets defined over the entire domain, we will define a subjective width that contains a certain percentage of the total weight of the target kernel (such as 0.9 or 0.99 for example) and treat those target kernels as if they have compact support of that width. This introduces some numerical errors that must be acknowledged, therefore we would ideally work with targets that have compact support.

3.1.2 Gradient Filters

To obtain information about derivatives of the true model (also important for imaging discontinuities), we can use target kernels that are the negative derivatives of the averaging filters defined in the previous subsection (see Fig. (5)). However, this will not yield pure derivative information, but rather smoothed derivative information. For example, if we use the derivative of the bump function we will obtain the bump averaged derivative instead of the pure derivative.

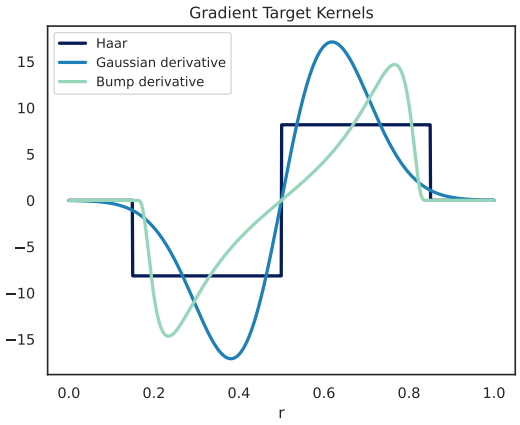


Figure 5: Examples of three derivative target kernels.

Similarly to the averaging target kernels, derivative target kernels have issues near the boundaries, and therefore such clipped targets will not be used (except for targets defined on the entire domain, where we will define a "fake" compact support). We have not included the derivative of the boxcar in Figure (5) since that leads to a δ distribution, which is unattainable in practice. Instead, we considered the derivative of a triangular function, which leads to a Haar-like target kernel. These gradient target kernels can also be

utilised to constrain the location of model discontinuities (see Fig. (6)).

Regardless of what filters we choose as the target kernels, the obtained filtered model is the generalised property defined in the previous two subsections and it can be bounded in the property space according to (12). The sensitivity kernels will dictate how well the properties can be constrained.

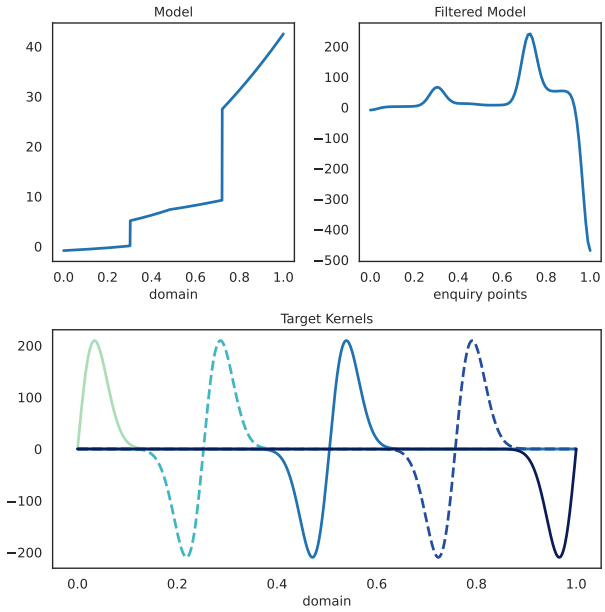


Figure 6: Example of derivative filtering. Top left: original model, top right: filtered model, bottom: examples of the target kernels at different locations. Notice the false derivative information near the edges in the filtered model.

4 Filters from Normal Mode Sensitivity Kernels in 1D

In this section we will study the ability of a set of normal mode sensitivity kernels to constrain three types of properties of a synthetic "true" model: local uniform averages, local Gaussian averages, and local Gaussian averages of the gradient.

Normal modes represent the fundamental vibrational patterns of the Earth in response to seismic disturbances. These oscillations, akin to the resonant frequencies of a musical instrument, arise from the interactions between seismic waves and the Earth's internal structure. Concurrently, sensitivity kernels serve as mathematical objects that quantify how variations in Earth's properties—such as density, elasticity, or geometry—affect observed seismic signals. While the full relation between the properties of the Earth and seismic data through normal mode theory is not linear, we can use sensitivity kernels to obtain an approximate linear relaitionship between the two. This forward relation takes the form of (4) (Woodhouse, Giardini, and Li 1986, Dahlen and Tromp 2020), but we will only focus on the simpler (although unrealistic case) where the data depends only on v_s according to (5). The normal modes used have been chosen to have sensitivity to the mantle, but some residual inner core sensitivity still remained. Some of the normal mode sensitivity kernels K_i are shown in Fig. (7). For an application of SOLA with boxcar target kernels for local average information see Restelli, Zaroli, and Koelemijer (2023).

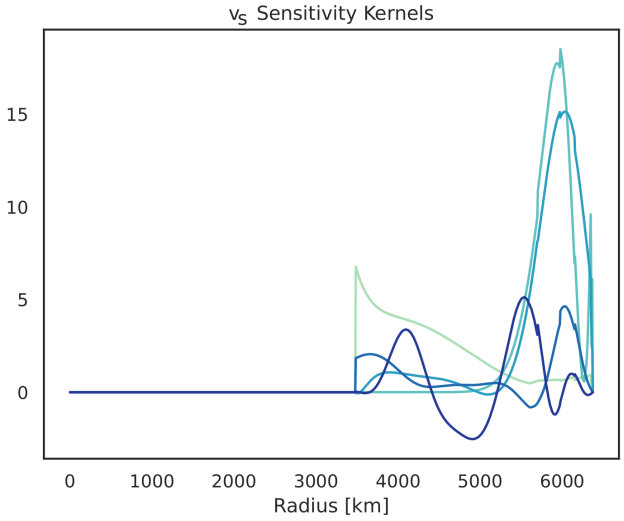


Figure 7: Examples of 1D normal mode sensitivity kernels to v_s computed in the PREM model (Dziewonski and Anderson 1981), plotted from the centre of the Earth (left) to the surface (right). The units of the sensitivity kernels will not be important since the toy problem does not contain real data or model. The units are assumed to be whatever they must be to have a uniform unit convention throughout the problem.

For extracting information about the unknown model we will use boxcar functions as defined in (20), the Gaussian function defined in (21), and the negative spatial derivative of the function defined in (21). For the purposes of this toy problem we will use a randomly generated "true" model $\bar{m}$ formed from piecewise continuous polynomials of unknown orders. An unknown number of discontinuities are also introduced and the number of discontinuities may be 0. We will assume that we know that the true model is bounded by some positive piece-wise function M:

$$|\bar{m}(r)| < M(r) \tag{23}$$

Normally M would be chosen based on physics related arguments. In this case, since the true model has no physical significance, we chose the upper bound by assuming that we know that the true model must not exceed some value $m_{max} = 5\max(\bar{m})$ anywhere in the model. Then, we can obtain the following upper bound:

$$\begin{aligned} \|\bar{m}\|_{\mathcal{M}} &= \sqrt{\langle \bar{m}, \bar{m} \rangle_{\mathcal{M}}} \\ \|\bar{m}\|_{\mathcal{M}} &= \sqrt{\int_0^R m^2(r) dr} \leq \sqrt{\int_0^R m_{max}^2 dr} \\ \|\bar{m}\|_{\mathcal{M}} &\leq m_{max} \sqrt{\int_0^R dr} = M \end{aligned} \tag{24}$$

where R is Earth's radius. The true (error-free) data $\{d_i\}_{i=1}^N$ are obtained from the true model via:

$$d_i = \int_\Omega K_i(r)\bar{m}(r)dr \tag{25}$$

From the true data we can immediately obtain the least norm solution $\tilde{m}$ (see Fig. (8)):

$$\tilde{m} = G^*(GG^*)^{-1}d = \sum_{i,j}(\Lambda^{-1})_{ij}d_jK_i(r) \tag{26}$$

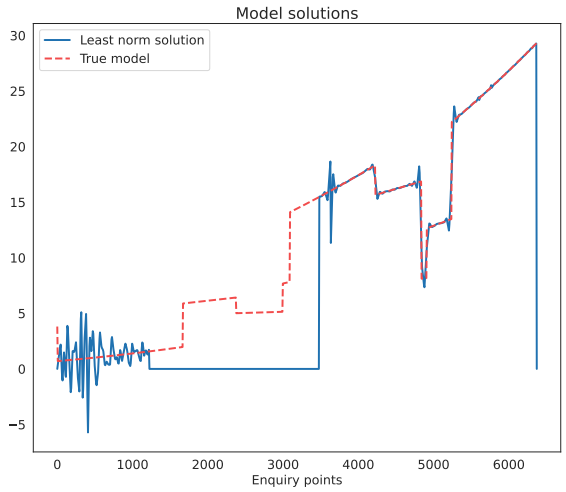


Figure 8: True model compared to the least norm solution.

For each property we will design target kernels that are spread uniformly over the spatial domain across 20 locations. For each spatial location, we will consider 20 different widths of the target kernels spaced logarithmically from 0.1% to 100% of the entire domain width ($6371km$). This will lead to 400 target kernels for each property type. For any target kernel T, the corresponding least norm solution property is:

$$\tilde{p} = \sum_{i,j}\int_\Omega T(r)K_i(r)(\Lambda^{-1})_{ij}d_jdr \tag{27}$$

However, the least norm solution is just one of infinitely many solutions and there is no reason to believe that it is the true solution. The other properties p that may equally represent the true property are found in the subspace P_M defined in (12). Therefore, for each individual target kernel we will compute the property bounds using:

$$\epsilon = \sqrt{\left(M^2 - \|\tilde{m}\|_{\mathcal{M}}^2\right)\mathcal{H}} \tag{28}$$
$$\mathcal{H} = \int_\Omega \left(T(r) - A(r)\right)^2 dr$$

where the resolving kernel is given by:

$$A(r) = \sum_{i,j}\int_\Omega T(r')K_i(r')dr'(\Lambda^{-1})_{ij}K_j(r) \tag{29}$$

and $\epsilon = \max_{p\in P_M}(p - \tilde{p})$. These offer the absolute property error bounds with respect to the least norm property solution $\tilde{p}$. However, we are more interested in the following error measure:

$$e = \frac{\epsilon}{\max(\tilde{p}) - \min(\tilde{p})} \tag{30}$$

In contrast to the relative error, which divides the absolute error by the least norm property pointwise, e divides all aboslute errors by the same value, which is the range of the least norm property. The reason for this choice is that the least norm property has 0 values in certain regions, which would lead to infinite relative errors at those location, even though the property bounds are both very close to zero. In Fig. (9) we have plotted e for all 400 target kernels for each of the three types of properties (type of target kernels). We have also selected two example target kernels and compare them with the corresponding resolving kernels to showcase how close the resolving kernel must be to the target kernel to produce the given relative error.

For the Gaussian resolving kernels we notice that the outer core has relative errors higher than 100% throughout, which comes as no surprise as v_s has no sensitivity in the outer core. We achieve excellent relative errors (less than 5%) in the mid-mantle at widths of $15\% - 40\%$ and in the outer inner core at widths of about 10%. Gaussian smoothed gradients showcase a very similar relative error distribution. It does show, however, lower errors in the mid-mantle at widths of about 17% and higher relative errors at widths smaller than 6%. The overall differences are not great, showing that our data can constrain quite well local gaussian averages and gaussian averaged gradients in the mid-mantle and some parts of the inner core. The most dramatic distribution of errors, however, we find in the case of uniform averages, where no region can produce error bounds less than 80%. Looking at the comparisons between target kernels and resolving kernels we see that Gaussian and Gaussian derivatives can be recovered very well in some regions, but boxcar functions cannot be recovered accurately for any width. This is not surprising given the smoothness of the normal mode sensitivity kernels. If the sensitivity kernels would have resembled boxcar functions, then we would naturally have expected better constraints on uniform averages.

In the first subfigure of Fig. (10) we selected a single width (resolution) of the Gaussian averaging targets of 25% of the domain width and looked how well constrained the associated property of the true model is. We find that the mid mantle ($r \in [4500km, 5500km]$) is the only region constrained well enough to warrant interpretation. The least norm property solution is plotted only for reference; it could very well be left out since it is not more likely to be true than any other imaginable property bounded by the property bounds.

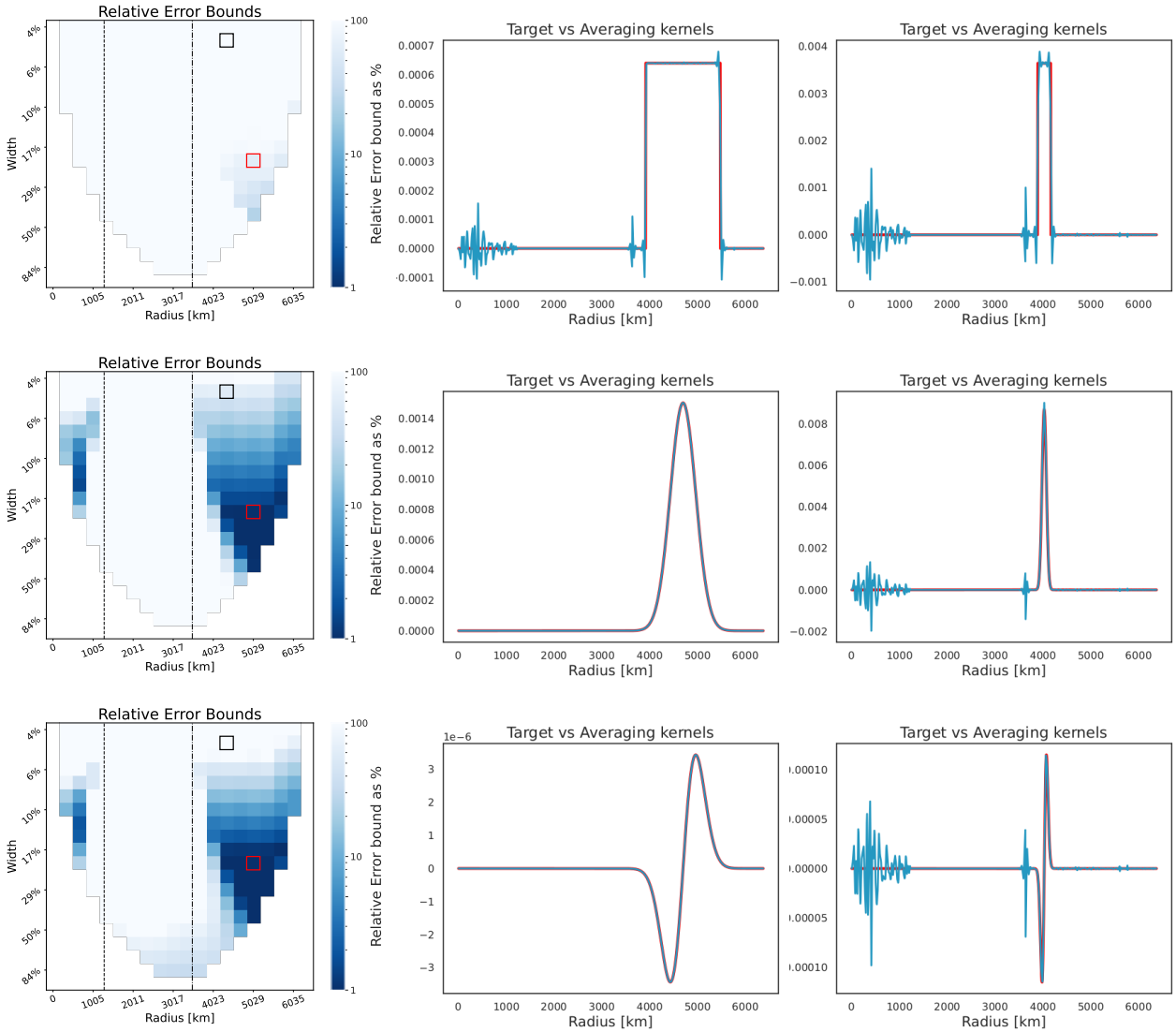


Figure 9: Comparison between different target kernel types and the resulting resolving kernels when applied to normal mode data. Relative error bounds (column 1) are given for uniform averages (row 1), Gaussian averages (row 2), and for Gaussian average of gradients (row 3). Column 2 shows how the target kernel compares to the resolving kernel at the red square (low error), and Column 3 shows how they compare at the black square (high error). The white regions in the relative error bounds plots are regions where the target kernels are clipped, and therefore are not interpretable. Regions with deep blue are regions where the property errors are very small. The dashed black line denotes the position of the ICB and the dot-dashed black line denotes the position of the CMB.

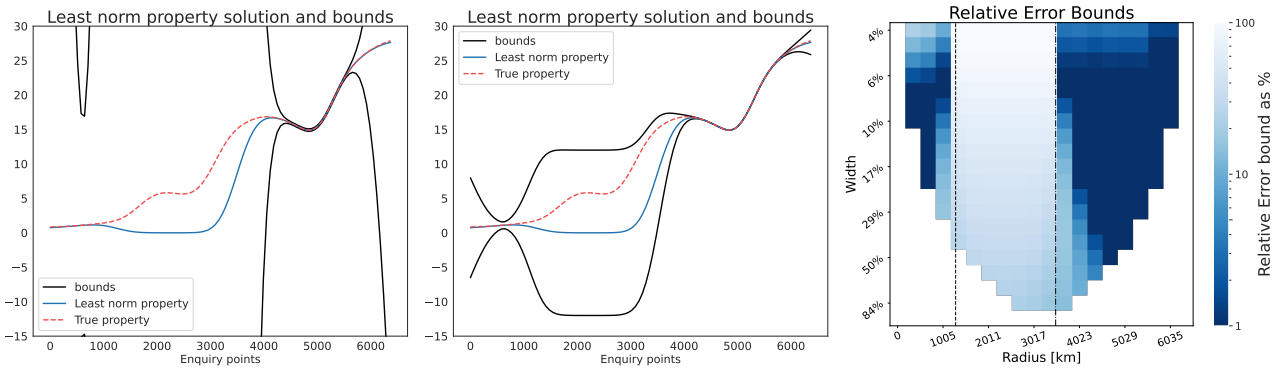


Figure 10: Effect of better model norm bound on the property constraints. (a) Property bounds on Gaussian averaging targets with width of 25% of domain width. Model norm bound used is $M \approx 10\|\bar{m}\|_{\mathcal{M}}$. The property bounds in some regions shoot outside of the y value range to values of roughly ± 200 and therefore they have not been plotted.(b) Property bounds on Gaussian averaging targets with width of 25% of domain width. Model norm bound used is $M = \|\bar{m}\|_{\mathcal{M}}$. (c) Property relative error bounds for $M = \|\bar{m}\|_{\mathcal{M}}$ model norm bound and Gaussian averaging targets.

One could note, however, that the least norm property approximates the true property better than our error bounds. This is because we choose loose constraints on the model norm. With tighter norm bounds the property bounds would be scaled down. For example, we choose $M = \|\tilde{m}\|_{\mathcal{M}}$ in Fig. (10), which is the best possible scenario, and also the least realistic. We now observe that the bounds in the mantle and the outer inner core are embracing the true property visually perfectly. If we look at the relative property bounds for all widths and spatial locations we see that most of the mantle and inner core can be well constrained. Even the outer core, where v_s sensitivity is 0, can be constrained slightly due to the model norm bound constraint.

Looking back at figure (8) we may notice how well the least norm solution agrees with the true solution. It also matches the locations of the discontinuities, except for the outer core where we have no sensitivity. In comparison, our Gaussian averages depicted in figure (10) seem to give us less information. First, we must note that a direct comparison between the least norm model solution $\tilde{m}$ and any possible property value $p \in P_M$ must be done with caution because those two mathematical objects live in different spaces. Second, the least norm solution assumes that the true solution is the solution with the least norm. In our case, computing the properties bounds was done assuming a norm bound that is 10 times higher than the true model norm, which is a rather conservative assumption compared to the least norm assumption. If one has very good reasons to believe that the true norm is very small, one can decrease the norm bound and obtain much tighter bounds on the desired properties, as seen in the second subfigure of figure (10) where we assumed that we know the true model norm. If the true model norm does actually match the least norm and we know that, then our property error bounds will drop to 0 everywhere. However if the true model norm is, in fact, not equal to the least norm, then the least norm solution will be systematically wrong and some features of that solution may not exist in reality. Our method on the other hand, will give deterministic bounds which are tight where the data can constrain the property and bounds that are very loose where the data cannot constrain the property.

5 Future Work

Extending the 1D applications presented here to 3D is trivial. We now need to think of target kernels as filters in 3D. Gradient targets will become directional, and Fourier decompositions will become spherical harmonic decompositions, etc. However, the introduction of model errors is more difficult. The deterministic property bounds will be replaced by equi-probabilistic bounds that are no longer easily computed (Al-Attar 2021). A further generalisation involves considering data that depends on multiple model parameters simultaneously. This situation could be tackled by forming a joint model space from the direct sums of the model spaces involved. When dealing with multiple model parameters, we will face the important and difficult problem of parameter trade-offs: do the data require a v_s perturbation, or a v_p perturbation, or both? To address this, we could try to constrain a single model parameter as well as possible while allowing the other to vary freely. By investigating the resulting resolving kernels that will arise for the unconstrained model parameters, we could shed light on the trade-offs between the constrained model parameter and all the other parameters.

While the ideas in this paper, and the possible generalisations mentioned above, are concerned with extending the BG SOLA method to deal with other observables than localised averages, one could also think of the reverse problem: given a certain property, what data should I use to constrain it as best as possible. By building a cost function dependent on the property bounds one could run an optimisation problem to find the best phases and/or frequencies for constraining the desired property.

6 Conclusion

We have expressed the classic BG SOLA theory in the more general framework of deterministic linear inferences developed by Al-Attar 2021. We then showed how different target kernels can be implemented for extracting different types of information about the unknown true model. This inference method only needs a norm bound as prior information, which leads to less subjectivity involved than in the case of regularized inversions. It is also more transparent about what the data can and cannot do since the main output of the method are some bounds rather than exact values: if the bounds are too large there is no point in attempting to draw conclusions about that piece of information. This "honesty" of the method will be useful when analysing the CMB topography, which is a region poorly covered by high quality data, and showcases many trade-offs with the surrounding structure. Instead of pushing more subjective information onto the model space we will target only those properties that can be found from the data and norm bound alone.

References

Amiri, Saman et al. (Dec. 2023). "Rayleigh wave group velocities in North-West Iran: SOLA Backus-Gilbert vs. Fast Marching tomographic methods". In: *Seismica* 2.2. DOI: 10.26443/seismica.v2i2.1011. URL: https://seismica.library.mcgill.ca/article/view/1011.

Al-Attar, David (2021). "Linear inference problems with deterministic constraints". In: *arXiv preprint arXiv:2104.12256*.

Backus, George (1970a). “Inference from inadequate and inaccurate data, I”. In: *Proceedings of the National Academy of Sciences* 65.1, pp. 1–7.
— (1970b). “Inference from inadequate and inaccurate data, III”. In: *Proceedings of the National Academy of Sciences* 67.1, pp. 282–289.
Backus, George and Freeman Gilbert (1968). “The resolving power of gross earth data”. In: *Geophysical Journal International* 16.2, pp. 169–205.
— (1970). “Uniqueness in the inversion of inaccurate gross earth data”. In: *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 266.1173, pp. 123–192.
Backus, George E and J Fo Gilbert (1967). “Numerical applications of a formalism for geophysical inverse problems”. In: *Geophysical Journal International* 13.1-3, pp. 247–276.
Dahlen, FAand and Jeroen Tromp (2020). “Theoretical global seismology”. In: *Theoretical Global Seismology*. Princeton university press.
Dziewonski, Adam M and Don L Anderson (1981). “Preliminary reference Earth model”. In: *Physics of the earth and planetary interiors* 25.4, pp. 297–356.
Fichtner, Andreas, H-P Bunge, and Heiner Igel (2006). “The adjoint method in seismology: I. Theory”. In: *Physics of the Earth and Planetary Interiors* 157.1-2, pp. 86–104.
Koelemeijer, Paula (2021). “Toward consistent seismological models of the core–mantle boundary landscape”. In: *Mantle convection and surface expressions*, pp. 229–255.
Liu, Qinya and Jeroen Tromp (2008). “Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods”. In: *Geophysical Journal International* 174.1, pp. 265–286.
Masters, Guy and David Gubbins (2003). “On the resolution of density within the Earth”. In: *Physics of the Earth and Planetary Interiors* 140.1-3, pp. 159–167.
McNamara, Allen K (2019). “A review of large low shear velocity provinces and ultra low velocity zones”. In: *Tectonophysics* 760, pp. 199–220.
Nolet, Guust (2008). “A breviary of seismic tomography”. In: *A Breviary of Seismic Tomography*.
Oldenburg, DW (1981). “A comprehensive solution to the linear deconvolution problem”. In: *Geophysical Journal International* 65.2, pp. 331–357.
Pijpers, FP and MJ Thompson (1992). “Faster formulations of the optimally localized averages method for helioseismic inversions”. In: *Astronomy and Astrophysics (ISSN 0004-6361), vol. 262, no. 2, p. L33-L36.* 262, pp. L33–L36.
— (1994). “The SOLA method for helioseismic inversion”. In: *Astronomy and Astrophysics (ISSN 0004-6361), vol. 281, no. 1, p. 231-240* 281, pp. 231–240.
Rawlinson, Nicholas, S Pozgay, and S Fishwick (2010). “Seismic tomography: a window into deep Earth”. In: *Physics of the Earth and Planetary Interiors* 178.3-4, pp. 101–135.
Restelli, Federica, Christophe Zaroli, and Paula Koelemijer (2023). “Robust estimates of the ratio between S-and P-wave velocity anomalies in the Earth’s mantle using normal modes”. In: *EarthArXiv eprints*, p. X5539D.
Tromp, Jeroen, Carl Tape, and Qinya Liu (2005). “Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels”. In: *Geophysical Journal International* 160.1, pp. 195–216.
Woodhouse, John H, Domenico Giardini, and Xiang-Dong Li (1986). “Evidence for inner core anisotropy from free oscillations”. In: *Geophysical Research Letters* 13.13, pp. 1549–1552.
Zaroli, Christophe (2016). “Global seismic tomography using Backus–Gilbert inversion”. In: *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society* 207.2, pp. 876–888.
— (2019). “Seismic tomography using parameter-free Backus–Gilbert inversion”. In: *Geophysical Journal International* 218.1, pp. 619–630.
Zaroli, Christophe, Paula Koelemeijer, and Sophie Lambotte (2017). “Toward seeing the Earth’s interior through unbiased tomographic lenses”. In: *Geophysical Research Letters* 44.22, pp. 11–399.