A matching pursuit approach to the geophysical inverse problem of seismic traveltime tomography under the ray theory approximation

N Schneider, V Michel, K Sigloch, E Totten

Geophysical Journal International

Geophys. J. Int. (2024) **238,** 1546–1581
Advance Access publication 2024 April 30
GJI Seismology

https://doi.org/10.1093/gji/ggae153

A matching pursuit approach to the geophysical inverse problem of seismic traveltime tomography under the ray theory approximation

N. Schneider [1], V. Michel [1], K. Sigloch [2] and E. J. Totten [3,*]

[1] *Department of Mathematics, Geomathematics Group Siegen, University of Siegen,* 57068 *Siegen, Germany.*
E-mail: naomi.schneider@mathematik.uni-siegen.de;michel@mathematik.uni-siegen.de
[2] *Université Côte d'Azur, Observatoire Côte d'Azur, CNRS, IRD, Géoazur, CS* 10269 - *F* 06905 *Sophia Antipolis, France*
[3] *Department of Earth Sciences, University of Oxford, Oxford OX*1 3*AN, UK*

Accepted 2024 April 23. Received 2024 March 29; in original form 2023 September 2

SUMMARY

Seismic traveltime tomography is a geophysical imaging method to infer the 3-D interior structure of the solid Earth. Most commonly formulated as a linearized inverse problem, it maps differences between observed and expected wave traveltimes to interior regions where waves propagate faster or slower than the expected average. The Earth's interior is typically parametrized by a single kind of localized basis function. Here we present an alternative approach that uses matching pursuits on large dictionaries of basis functions.Within the past decade the (Learning) Inverse Problem Matching Pursuits [(L)IPMPs] have been developed. They combine global and local trial functions. An approximation is built in a so-called best basis, chosen iteratively from an intentionally overcomplete set or dictionary. In each iteration, the choice for the next best basis element reduces the Tikhonov–Phillips functional. This is in contrast to classical methods that use either global or local basis functions. The LIPMPs have proven their applicability in inverse problems like the downward continuation of the gravitational potential as well as the MEG-/EEG-problem from medical imaging. Here, we remodel the Learning Regularized Functional Matching Pursuit (LRFMP), which is one of the LIPMPs, for traveltime tomography in a ray theoretical setting. In particular, we introduce the operator, some possible trial functions and the regularization. We show a numerical proof of concept for artificial traveltime delays obtained from a contrived model for velocity differences. The corresponding code is available online.

Key words: Body waves; Computational seismology; Eikonal equation; Seismic tomography; Waveform inversion; Seismology.

1 INTRODUCTION

Seismic traveltime tomography serves to infer the 3-D interior structure of the solid Earth. For this, it measures the characteristics of seismic waves that propagate through the Earth's deep interior, between sources (earthquakes) and receivers (seismometers) that are located at or near the surface. The most widely practised approach solves a linearized inverse problem, where measured traveltimes of waves propagating through the real, heterogeneous interior deviate moderately from forward-modelled traveltimes through a spherically symmetric reference Earth model. This reference velocity model is updated with moderate 3-D variations, in a linear inversion step that attributes traveltime anomalies (observed minus predicted) to discrete regions where wave velocities must be moderately faster or slower than in the reference model. The quantitative connection is made by means of efficiently computed sensitivity kernels (see Dahlen & Tromp 1998; Dahlen *et al.* 2000; Ben-Menahem & Singh 2000; Amirbekyan 2007; Nolet 2008; Aki & Richards 2009).

In particular, a Spherically symmetric, Non-Rotating, Elastic and Isotropic spherical shell (SNREI) model like the IASP91 (see Kennett & Engdahl 1991), is usually used. When comparing modelled traveltimes with true measurements, we expect a traveltime difference caused by anomalies within the Earth. Specifically, for total body-wave traveltimes of hundreds to 1000 s, the unexplained traveltime

*Now at: Geophysics Section, Dublin Institute for Advanced Studies, Dublin, Ireland.

delay is typically fractions of a second to several seconds. Traveltime tomography aims to approximate these deviations in the velocity field from the corresponding delays. Thus, traveltime tomography is a problem of inference, most often formulated as a linearized inverse problem. Unfortunately, not many theoretical results about it are known. In particular, a singular value decomposition of the associated mathematical forward operator (obtaining traveltime delays out of velocity anomalies), which would essentially accelerate the forward modelling, is unknown. The inverse problem is ill-posed, meaning that, for example the solution is non-unique. The latter causes, more practically speaking, that it is difficult to attribute a traveltime anomaly to a specific region along a waveform path (given the overall traveltime and its delay). For details on the theory, see e.g. Amirbekyan (2007). Traveltime tomography is most often approached with ray theory or with a more accurate, but also more demanding finite frequency strategy (Yomogida 1995; Dahlen & Tromp 1998; Marquering *et al.* 1998, 1999; Tian *et al.* 2007a, b; Nolet 2008; Sigloch 2008; Hosseini *et al.* 2020). Traveltime tomography is practised in two or three dimensions, and from length scales of millimetres (e.g. ultrasound waves in non-destructive testing of construction materials) to the planetary scale (in a spherical geometry). Our study focuses on this latter scale, commonly termed 'global traveltime tomography'. A broad range of global tomography models exists, and dozens can be inspected at the SubMachine webportal (https://www.earth.ox.ac.uk/~smachine/cgi/index.php Hosseini *et al.* 2018). These models differ in the types and amounts of data that entered the inverse problem, but another key difference concerns the parametrization of the sphere chosen by various workers. Basis functions include tetrahedra, voxels and spherical harmonics, which greatly differ in shape but have in common that they are largely chosen for computational convenience, rather than for some physical or statistical properties of Earth's mantle structure, the quantity to invert for. Such mantle properties are largely unknown, although one may reasonably postulate *a priori* notions, such as mantle structures being extended more horizontally than vertically. Hence, a better parametrization strategy could be to restrict basis functions minimally and instead let the data itself choose suitable functions.

This is especially interesting since some areas of the mantle are very well illuminated, others very poorly. However, we do not expect fundamental changes in convection style in different places due to the general viscosity of the mantle. Hence, poorly sampled regions might be filled and justified physically if the approximation method is free to choose types of basis functions taking into account the whole data distribution at once. This motivated us to apply a different type of approximation method for inverse problems to traveltime tomography which appears to be suitable to tackle these challenges of parametrization, underdeterminedness and regularization.

Here, we propose to use the Learning Regularized Functional Matching Pursuit (LRFMP) which is one of the (Learning) Inverse Problem Matching Pursuit [(L)IPMP] algorithms. The LRFMP is the realization of the RFMP with a learning add-on. The RFMP algorithm iteratively approximates the solution of an inverse problem by minimizing the corresponding Tikhonov–Phillips functional. The unique characteristic of the (L)IPMPs as well as the RFMP, is that the obtained approximation will be given in a so-called best basis of dictionary elements. The dictionary is an intentionally redundant set of diverse trial functions. Here, we consider orthogonal polynomials and linear tesseroid-based finite element hat functions (FEHFs). The best basis is usually also made of all types of trial functions given in the dictionary. The learning add-on enables the method to choose the best basis out of infinitely many trial functions. More details on the methods can be found in Berkel *et al.* (2011); Fischer (2011); Fischer & Michel (2012, 2013a, b); Michel & Telschow (2014, 2016); Telschow (2014); Michel (2015); Michel & Orzlowski (2017); Gutting *et al.* (2017); Telschow *et al.* (2018); Kontak (2018); Kontak & Michel (2018, 2019); Leweke (2018); Michel & Schneider (2020); Prakash *et al.* (2020); Schneider (2020) and Schneider & Michel (2022). These methods have already been successfully applied to inverse problems in geodesy and medical imaging as well as for normal mode tomography, and we briefly introduce them in Section 4, with an emphasis on adjustments made for traveltime tomography. Regarding the theory, a focus on the derivation of certain gradients and inner products needed for this adjustment is given in Section 3.3. These computations are provided as supplementary material in the appendix of this paper.

We demonstrate this new method on inversions for synthetic (i.e. invented) whole-mantle test structures, using realistic sampling geometries, that is actual earthquake-to-receiver paths that have yielded high-quality traveltime data for tomography in the ISC-EHB catalogue. For the latter, see for example Weston *et al.* (2018) and The International Seismological Centre (ISC) (2022).

In Section 2, we introduce the mathematical inverse problem of traveltime tomography. We present our choices of dictionary elements in Section 3, and we compute regularization terms related to a Sobolev space of these trial functions. We introduce the LRFMP in Section 4 and show a first proof of concept in Section 5. Section A details certain mathematical derivations regarding the trial functions and the methods. The corresponding code is available online (Schneider 2023).

1.1 Notation

The set of positive integers is denoted by $\mathbb{N}$. If we include 0, we use $\mathbb{N}_0$. For integers, we write $\mathbb{Z}$. For the set of real numbers, we use $\mathbb{R}$. The d-dimensional Euclidean space is called $\mathbb{R}^d$. Only positive real numbers are collected in $\mathbb{R}_+$. For $x \in \mathbb{R}^3$, we can use the common spherical coordinate transformation

$$x(r, \varphi, \theta) = \begin{pmatrix} r\sqrt{1-t^2}\cos(\varphi) \\ r\sqrt{1-t^2}\sin(\varphi) \\ rt \end{pmatrix} \tag{1}$$

for the radius $r \in \mathbb{R}_+$, the longitude $\varphi \in [0, 2\pi[$ and the latitude $\theta \in [0, \pi]$ which yields the polar distance $\cos(\theta) = t \in [-1, 1]$. A radius is denoted by $\mathcal{R} \in \mathbb{R}_+$ and the corresponding ball with radius $\mathcal{R}$ is defined by $\mathbb{B}_\mathcal{R} := \{x \in \mathbb{R}^3 : |x| \leq \mathcal{R}\}$. If $\mathcal{R} = 1$, we also abbreviate $\mathbb{B} := \mathbb{B}_1$.

Thus, a point $\xi(\varphi, t) \in \partial\mathbb{B} \subset \mathbb{R}^3$ is given by

$$\xi(\varphi, t) = \begin{pmatrix} \sqrt{1-t^2}\cos(\varphi) \\ \sqrt{1-t^2}\sin(\varphi) \\ t \end{pmatrix}. \tag{2}$$

Moreover, a well-known local orthonormal basis in $\mathbb{R}^3$ is defined by the vectors

$$\varepsilon^r(\varphi, t) = \begin{pmatrix} \sqrt{1-t^2}\cos(\varphi) \\ \sqrt{1-t^2}\sin(\varphi) \\ t \end{pmatrix}, \ \varepsilon^\varphi(\varphi, t) = \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \\ 0 \end{pmatrix}, \ \varepsilon^t(\varphi, t) = \begin{pmatrix} -t\cos(\varphi) \\ -t\sin(\varphi) \\ \sqrt{1-t^2} \end{pmatrix}, \tag{3}$$

see, for instance, Michel (2013).

2 SEISMIC TRAVELTIME TOMOGRAPHY

Seismic traveltime tomography estimates the propagation velocity of seismic waves as a function of spatial location inside the solid Earth. The sensitivity of traveltime measurements to Earth structures along the wave propagation path is volumetrically extended in practice, but can often be reasonably approximated as an (infinitely narrow) ray, in analogy to optical ray theory. Here we consider only (seismic) ray theory for P waves, although our approach could be extended to less approximative sensitivity modelling, such as Born/finite-frequency approximations, see for example Dahlen *et al.* (2000), and to S wave or other wave types. As we aim here for a proof of concept, we consider the accuracy of the classical ray theoretical infinite-frequency approach as adequate. For one seismic ray $\widetilde{X}$ between seismic source and receiver, the mathematical relationship between its traveltime ψ and the (P-)wave velocity c (and its slowness S, respectively), is given by the Eikonal equation (see e.g. Dahlen & Tromp 1998; Nolet 2008; Michel 2022),

$$\frac{1}{c(\widetilde{X}(s))} = S(\widetilde{X}(s)) = |\nabla_{\widetilde{X}}\psi(\widetilde{X}(s))|, \tag{4}$$

where $\widetilde{X} : [0, R] \to \mathbb{R}^3$ and the arc length s is chosen to parametrize $\widetilde{X}$. We reparametrize the ray here by a parameter $t \in [0, 1]$ such that $X : [0, 1] \to \mathbb{R}^3$ has the same curve, that is $X([0, 1]) = \widetilde{X}([0, R])$. The specific choice of the parameter transformation $s \leftrightarrow t$ depends on the implementation of the numerically calculated ray. The Eikonal equation yields the non-linear inverse problem

$$\int_{\widetilde{X}} S(\widetilde{X}(s))\, \mathrm{d}\sigma(\widetilde{X}(s)) = \psi(\widetilde{X}(R)) \ \Leftrightarrow \ \int_0^1 S(X(t))\left|X'(t)\right| \mathrm{d}t = \psi(X(1)) \tag{5}$$

with

$$s(t) = \int_0^t |X'(\tau)| d\tau. \tag{6}$$

Due to the difficulty of non-linear inverse problems, eq. (5) is usually linearized: instead of approximating the slowness S itself, we consider the deviation δS between reality and a reference model (see e.g. Amirbekyan 2007; Nolet 2008). This yields the seismic ray perturbation operator (SPO) given by

$$(\mathcal{T}\,\delta S)\left(\widetilde{X}_{\mathrm{ref}}\right) \coloneqq \int_0^R \delta S\left(\widetilde{X}_{\mathrm{ref}}(s)\right)\, \mathrm{d}s = \delta\psi\left(\widetilde{X}_{\mathrm{ref}}(R)\right) \tag{7}$$

where $\widetilde{X}_{\mathrm{ref}}$ stands for the ray obtained from a reference model and parametrized with the arc length s between 0 and its total arc length R. Moreover, $\delta\psi = \psi_{\mathrm{obs}} - \psi_{\mathrm{ref}}$ is the delay of the observed and (due to the reference model) expected traveltime. The more rays we consider, the better our approximation of δS can be. Thus, in practice, we use a family of rays $\{\widetilde{X}_{\mathrm{ref},i}\}_{i=1,\dots,\ell}$ with respect to a set of source–receiver pairs

$$\left\{\left(\widetilde{X}_{\mathrm{ref},i}(0), \widetilde{X}_{\mathrm{ref},i}(R)\right)\right\}_{i=1,\dots,\ell} \tag{8}$$

and consider the Discretized SPO (DSPO)

$$\left(\mathcal{T}_\urcorner^i\,\delta S\right)(X_{\mathrm{ref},i}) \coloneqq \int_0^R \delta S\left(\widetilde{X}_{\mathrm{ref},i}(s)\right) \mathrm{d}s = \delta\psi_i\left(\widetilde{X}_{\mathrm{ref},i}(R)\right) = \int_0^1 \delta S\left(X_{\mathrm{ref},i}(t)\right)\left|X'_{\mathrm{ref},i}(t)\right| \mathrm{d}t = \delta\psi_i\left(X_{\mathrm{ref},i}(1)\right), \tag{9}$$

$$\mathcal{T}_\urcorner\,\delta S = \delta\psi_\urcorner \tag{10}$$

with s as in eq. (6), $s(1) = R$, $\mathcal{T}_\urcorner \coloneqq (\mathcal{T}_\urcorner^i)_{i=1,\dots,\ell}$ and $\delta\psi_\urcorner \coloneqq (\delta\psi_i(X_{\mathrm{ref},i}(1)))_{i=1,\dots,\ell}$. Our aim is then to construct an approximation

$$\delta S \approx \sum_{i=1}^{I} \alpha_i d_i \tag{11}$$

for $I \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$ and some trial functions d_i. Note that, in seismology, instead of considering δS, the velocity anomaly (deviation from the reference velocity) is often considered. In our approach, this means, we would have to reformulate our approximation in the following way:

if we obtain δS as the linear combination, we approximate

$$\delta S = \frac{1}{c} - \frac{1}{c_{\text{ref}}} = -\frac{c - c_{\text{ref}}}{c c_{\text{ref}}}. \tag{12}$$

This reformulates to

$$\delta S + \frac{1}{c_{\text{ref}}} = \frac{1}{c}. \tag{13}$$

Thus, after approximating δS, we would transform it via

$$-\frac{\delta S}{\delta S + c_{\text{ref}}^{-1}} = -\left(-\frac{c - c_{\text{ref}}}{c c_{\text{ref}}}\right) c = \frac{c - c_{\text{ref}}}{c_{\text{ref}}} =: \frac{\mathrm{d}c}{c} \tag{14}$$

for a better comparison with the geophysics community. Since we are using a non-real, artificial deviation model (see Section 5.1), we abstain from this explicit reformulation but consider the numerator only in our resolution test.

3 TRIAL FUNCTIONS

We utilize two types of trial functions in our methods: global orthogonal polynomials and compactly supported linear finite element hat functions. We first present their definition and give examples of them.

3.1 Definition and examples

3.1.1 Tesseroid-based finite element hat functions

For a general introduction to finite elements, see e.g. Braess 2007; Grossmann *et al.* 2007; Johnson 2009; Schwarz & Köckler 2011. Regarding trial functions, we are interested in composite hat functions on tesseroids as the finite elements. Thus, in the sequel, we speak of finite element hat functions (FEHFs).

In seismology, using tetrahedra as finite elements is common (see e.g. Tian *et al.* 2007a, b; Sigloch 2008; Hosseini *et al.* 2020). We change the underlying finite element structure here because, from a seismological perspective, there is no physical need to use tetrahedra. Tesseroids, see Fukushima (2018), are rectangular parallelepipeds defined by upper and lower bounds with respect to the radius, the longitude and the latitude. Thus, a tesseroid may be even more reasonable than tetrahedra: such an underlying structure would be a realization of the interior of the Earth being structured or layered by gravitation. Moreover, the depth of the mantle (≈2900 km) is much smaller than the circumference of the Earth (≈40 000 km). Hence, the size of the heterogeneities in the mantle is smaller in the depth than it is latitudinally and longitudinally.

For the definition of a tesseroid, we obviously have six degrees of freedom. Due to the composite nature of the FEHFs, we move from the boundary-based view to a centre-based one. Then, the degrees of freedom are its radial centre R and its distance to each side ΔR as well as the longitudinal and latitudinal centres Φ and T and their respective distances to each side $\Delta\Phi$ and ΔT. The natural constraints of a tesseroid are

$$R \in [\rho\mathcal{R}, \mathcal{R}] = [R_{\min}, R_{\max}], \qquad 0 < \epsilon_R \leq \Delta R \leq \mathcal{R}/2, \tag{15}$$

$$\Phi \in [0, 2\pi] = [\Phi_{\min}, \Phi_{\max}], \qquad 0 < \epsilon_\Phi \leq \Delta\Phi \leq \pi, \tag{16}$$

$$T \in [-1 + \epsilon_T, 1 - \epsilon_T] = [T_{\min}, T_{\max}], \qquad 0 < \epsilon_T \leq \Delta T \leq 0.5. \tag{17}$$

Note that, for the derivative of an FEHF, we require that $\Delta R, \Delta\Phi, \Delta T \neq 0$ (see Section A1) and, thus, we introduce $\epsilon_R, \epsilon_\Phi$ and ϵ_T. Further note that, with $\rho \in [0, 1]$, we control the maximal possible depth of the tesseroids. For instance, if we set $\rho = 3482/6371 = 0.54654$, we allow the FEHF as deep as the core–mantle boundary. Furthermore, due to singularities at the poles (see Section A3) we need to stay away from the theoretically possible bounds ± 1 in the latitudinal case. We can identify (in polar coordinates) the ball with the domain

$$D := [0, \mathcal{R}] \times [P, P + 2\pi] \times [-1, 1] \quad \text{with} \quad P := \left\lfloor \frac{\Phi - \Delta\Phi}{\pi} \right\rfloor \pi, \tag{18}$$

the difference domain as

$$\Delta D := [\epsilon_R, \mathcal{R}/2] \times [\epsilon_\Phi, \pi] \times [\epsilon_T, 0.5] \tag{19}$$

$$= [\epsilon_R, R_{\max}/2] \times [\epsilon_\Phi, \Phi_{\max}/2] \times [\epsilon_T, (T_{\max} + \epsilon_T)/2] \tag{20}$$

and the tesseroid as

$$E := [\max(R_{\min}, R - \Delta R), \min(R_{\max}, R + \Delta R)] \times [\Phi - \Delta\Phi, \Phi + \Delta\Phi] \times [\max(T_{\min}, T - \Delta T), \min(T_{\max}, T + \Delta T)]. \tag{21}$$

Note that we need the definition of P as the starting point of a 2π-interval for technical reasons: in the appendix, we calculate the inner product of two FEHF which depends on the intersection of their respective tesseroids. For this intersection, it is important to know whether P is -1, 0 or 1 in practice.

For the sake of brevity, let $a \coloneqq (a_j)_{j=1,2,3} \coloneqq (r, \varphi, t)$, $A \coloneqq (A_j)_{j=1,2,3} \coloneqq (R, \Phi, T)$, $\Delta A \coloneqq (\Delta A_j)_{j=1,2,3} \coloneqq (\Delta R, \Delta\Phi, \Delta T)$,

$$A_{\min} \coloneqq (\rho\mathcal{R}, P, -1 + \epsilon_T) = (R_{\min}, \Phi_{\min}, T_{\min}), \tag{22}$$

$$A_{\max} \coloneqq (\mathcal{R}, P + 2\pi, 1 - \epsilon_T) = (R_{\max}, \Phi_{\max}, T_{\max}) \tag{23}$$

and the Cartesian product

$$\operatorname{supp}_{A,\Delta A} \coloneqq \prod_{j=1}^{3} [\max(A_{\min,j}, A_j - (\Delta A)_j), \min(A_{\max,j}, A_j + (\Delta A)_j)] = E. \tag{24}$$

With these definitions, we can consider the FEHFs. Commonly used in seismology are Lagrange finite element basis functions (see Tian *et al.* 2007a, b). Finite element basis functions for cuboids are given, for instance, in Maździarz (2010). We take the linear examples from there. Via translation to $\operatorname{supp}_{A,\Delta A}$ and the general notation just introduced, we then obtain the dictionary elements

$$N_{A,\Delta A}(x(a)) \coloneqq N_{(R,\Phi,T),(\Delta R,\Delta\Phi,\Delta T)}(r\xi(\varphi,t)) \coloneqq \chi_{\operatorname{supp}_{A,\Delta A}}(a) \prod_{j=1}^{3} \frac{\Delta A_j - |a_j - A_j|}{\Delta A_j}, \tag{25}$$

where χ denotes the characteristic function and $x(a) = r\xi(\varphi, t) \in \mathbb{B}_{\mathcal{R}}$. The FEHF $N_{A,\Delta A}$ attains its maximum in A, which is the centre of the respective volume element. It holds $N_{A,\Delta A}(A) = 1$. The function linearly decreases towards zero when moving towards $A \pm \Delta A$. It is constant zero outside of the volume element. Thus, it is only piecewise constant. An example is given in Fig. 1. Note that, in this example, we see that the hat is clearly visible. However, this shows that, although the theoretical support of an FEHF is a tesseroid, its visible support appears smaller.

3.1.2 *Polynomials*

For polynomials on a ball with radius $\mathcal{R}$, we consider the system

$$G^{\mathrm{I}}_{m,n,j}(r\xi(\varphi,t)) \coloneqq p_{m,n} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^n Y_{n,j}(\xi(\varphi,t)) \tag{26}$$

$$= p_{m,n} P_m^{(0,n+1/2)} \left(2\left(\frac{r}{\mathcal{R}}\right)^2 - 1\right) \left(\frac{r}{\mathcal{R}}\right)^n q_{n,j} P_{n,|j|}(t) \begin{cases} \sqrt{2}\cos(j\varphi), & j < 0 \\ 1, & j = 0 \\ \sqrt{2}\sin(j\varphi), & j > 0 \end{cases} \tag{27}$$

$$=: p_{m,n} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^n q_{n,j} P_{n,|j|}(t) \operatorname{Trig}(j\varphi) \tag{28}$$

with the normalization factors

$$p_{m,n} \coloneqq \sqrt{\frac{4m + 2n + 3}{\mathcal{R}^3}} \qquad \text{and} \qquad q_{n,j} \coloneqq q_{n,|j|} \coloneqq \sqrt{\frac{2n+1}{4\pi} \frac{(n - |j|)!}{(n + |j|)!}} \tag{29}$$

for $m,\ n \in \mathbb{N}_0,\ j \in \mathbb{Z},\ |j| \le n$ and where $P_m^{(\alpha,\beta)}$ denotes a Jacobi polynomial, $Y_{n,j}$ a spherical harmonic and $P_{n,|j|}$ an associated Legendre function. As can be seen here, we use fully normalized spherical harmonics in our implementation. For details on those composite functions, see for example Magnus *et al.* (1966); Müller (1966); Abramowitz & Stegun (1972); Szegö (1975); Freeden *et al.* (1998); Freeden & Schreiner (2009); Freeden & Gutting (2013) and Michel (2022). The functions G^{I} were investigated first by Dufour (1977); Ballani *et al.* (1993) and Michel (2013). For generalizations of this system, see Dunkl & Xu (2014) as well as Michel & Orzlowski (2016). Note that every G^{I} is an algebraic polynomial in $\mathbb{R}^3$ and is, therefore, well-defined on the whole ball. An example is given in Fig. 2.

3.2 A dictionary

With these two types of trial functions, we now build a so-called dictionary $\mathcal{D}$ which is an intentionally redundant set of functions. In the sequel, we work with the following dictionary

$$[\mathcal{G}_{M,N}]_{\mathrm{GI}} \coloneqq \left\{ G^{\mathrm{I}}_{m,n,j} \ :\ m,\ n \in \mathbb{N}_0,\ m \le M,\ n \le N,\ j \in \mathbb{Z},\ |j| \le n \right\}, \tag{30}$$

$$[\mathcal{A}]_{\mathrm{FEHF}} \coloneqq \left\{ N_{A,\Delta A} \ :\ A \in D,\ \Delta A \in \Delta D \right\}, \tag{31}$$

$$\mathcal{D} \coloneqq \mathcal{D}^{\mathrm{Inf}} \coloneqq [\mathcal{G}_{M,N}]_{\mathrm{GI}} \cup [\mathcal{A}]_{\mathrm{FEHF}}. \tag{32}$$

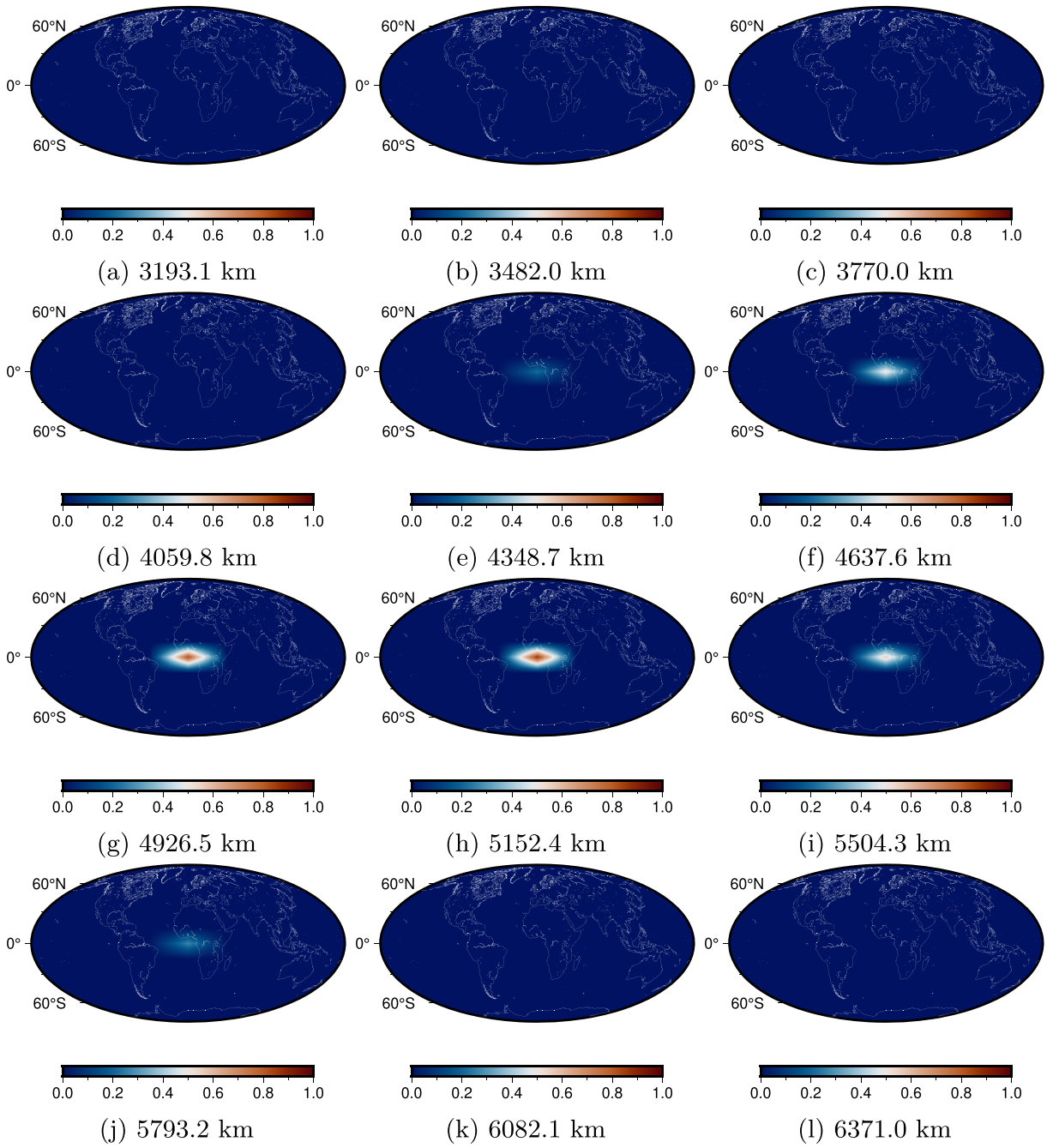

Figure 1. Example of an FEHF. We show $N_{(5096.8,\,0,\,0),\,(955.65,\,\pi/4,\,0.25)}$. The colour scales are adjusted for the purposes of comparison, that is the colourbar always has the same range to show where the influence of the function is larger.

We see that the trial function classes $[\mathcal{G}_{M,N}]_{\mathrm{GI}}$ and $[\mathcal{A}]_{\mathrm{FEHF}}$ are defined via the characteristic parameters of the trial functions. Note that the parameters of the FEHFs are continuous while those of the polynomials are discrete. We allow polynomials up to a maximum radial and angular degree. Theoretically, we could also allow $m,\ n \in \mathbb{N}_0$ and, thus, let $[\mathcal{G}_{M,N}]_{\mathrm{GI}}$ be infinite. In practice, however, this is not sensible as we will see later. Note, however, that $[\mathcal{A}]_{\mathrm{FEHF}}$ is truly infinite. We discuss the dictionary again in a larger context below when we introduce our approximation methods.

3.3 Regularization terms

As we are considering an ill-posed inverse problem, from a mathematical point of view, it is inevitable to include a regularization. For our approach, we use the Tikhonov–Phillips regularization. Thus, we need to determine a suitable function space for the penalty term. We decided to use the $\mathcal{H}_1(\mathbb{B})$-Sobolev space, see Yosida (1995); Adams & Fournier (2003); Heuser (2006); Braess (2007); Schwarz & Köckler (2011); Bhattacharyya (2012) and Werner (2018). The regularization terms are then obtained via

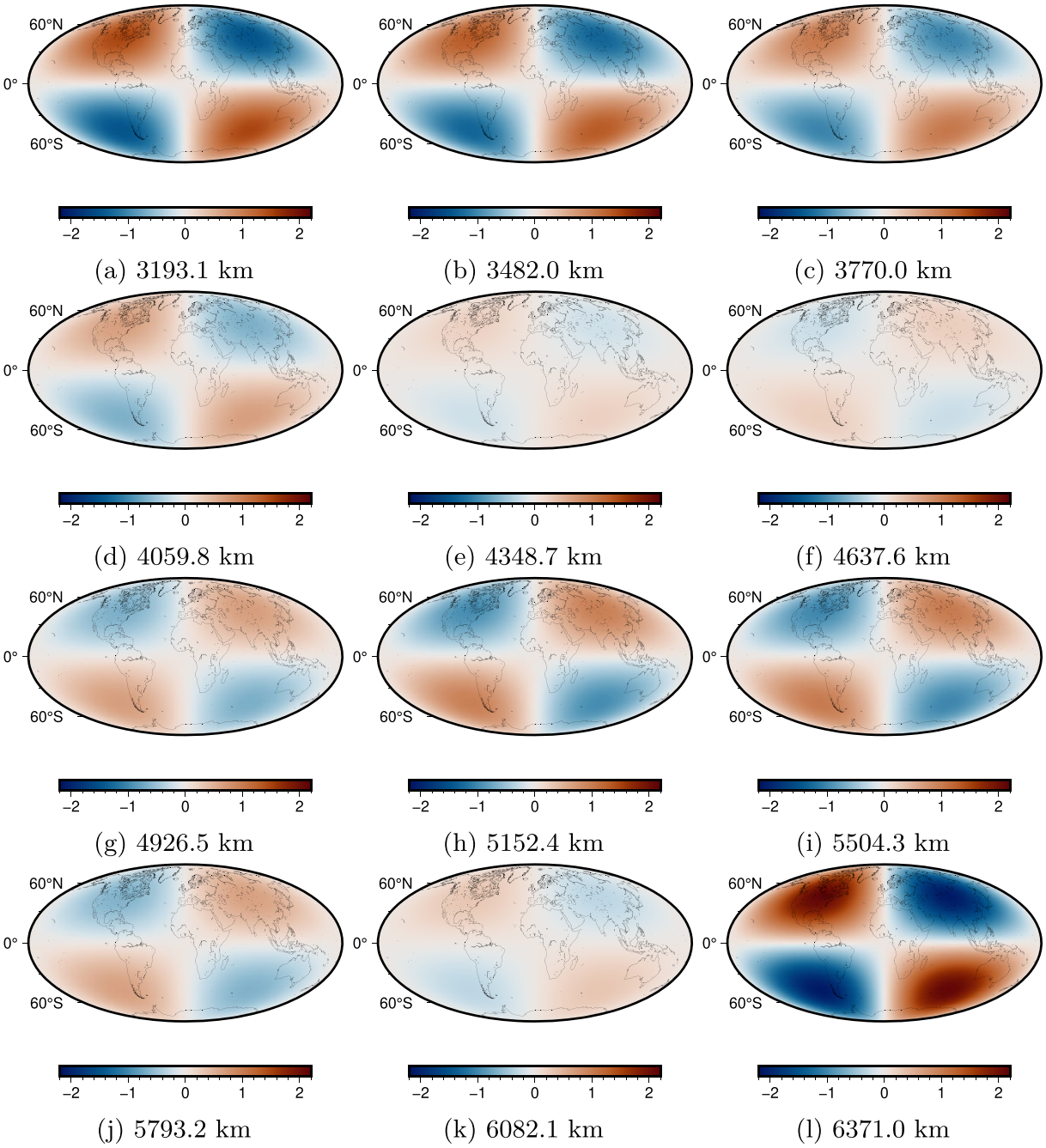

Figure 2. Example of a polynomial. We show $G^{\mathrm{I}}_{2,2,1}$. The colour scales are adjusted for the purposes of comparison, that is the colourbar always has the same range to show where the influence of the function is larger.

$$\langle d_i, d_j \rangle_{\mathcal{H}_1} = \left\langle \mathrm{D}^{(0)} d_i, \mathrm{D}^{(0)} d_j \right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R})} + \left\langle \mathrm{D}^{(1)} d_i, \mathrm{D}^{(1)} d_j \right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R}^3)} = \left\langle d_i, d_j \right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R})} + \left\langle \nabla_x d_i, \nabla_x d_j \right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R}^3)} \tag{33}$$

for two dictionary elements $d_i,\ d_j \in \mathcal{D}$. We have to determine these inner products for the different cases of dictionary elements in $\mathcal{D}$. We start with the determination of the gradients. For the FEHFs, we obtain

$$\nabla_{r\xi(\varphi,t)} N_{(R,\Phi,T),(\Delta R,\Delta\Phi,\Delta T)}(r,\varphi,t) = \chi_{\mathrm{supp}_{[(R,\Phi,T)-(\Delta R,\Delta\Phi,\Delta T),(R,\Phi,T)+(\Delta R,\Delta\Phi,\Delta T)]}}(r,\varphi,t) \tag{34}$$

$$\times \left(\varepsilon^r \frac{[-\operatorname{sgn}(r-R)]}{\Delta R} \frac{\Delta\Phi - |\varphi - \Phi|}{\Delta\Phi} \frac{\Delta T - |t-T|}{\Delta T} \right. \tag{35}$$

$$\left. + \frac{1}{r} \varepsilon^{\varphi} \frac{1}{\sqrt{1-t^2}} \frac{\Delta R - |r-R|}{\Delta R} \frac{[-\operatorname{sgn}(\varphi-\Phi)]}{\Delta\Phi} \frac{\Delta T - |t-T|}{\Delta T} \right. \tag{36}$$

$$+\frac{1}{r}\varepsilon^{t}\sqrt{1-t^2}\frac{\Delta R-|r-R|}{\Delta R}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\frac{[-\operatorname{sgn}(t-T)]}{\Delta T}\Bigg). \tag{37}$$

The derivation of this formula is done straightforwardly and can be found in Section A1. For the polynomials, we have

$$\nabla_{r\xi(\varphi,t)}\ G^{\mathrm{I}}_{m,n,j}(r\xi(\varphi,t)) = \left(\frac{\partial}{\partial x_k}G^{\mathrm{I}}_{m,n,j}(r\xi(\varphi,t))\right)_{k=1,2,3} \tag{38}$$

$$= p_{m,n}q_{n,j}\left(\left(P_m^{(0,n+1/2)}(I(r))\right)' I'(r)\left(\frac{r}{\mathcal{R}}\right)^n P_{n,|j|}(t)\mathrm{Trig}(j\varphi)\xi(\varphi,t)\right. \tag{39}$$

$$+\frac{n}{\mathcal{R}}P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1} P_{n,|j|}(t)\mathrm{Trig}(j\varphi)\xi(\varphi,t) \tag{40}$$

$$+\frac{j}{\mathcal{R}}P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1}\frac{1}{\sqrt{1-t^2}}P_{n,|j|}(t)\mathrm{Trig}(-j\varphi)\varepsilon^{\varphi}(\varphi,t) \tag{41}$$

$$\left.+\frac{1}{\mathcal{R}}P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1}\sqrt{1-t^2}P'_{n,|j|}(t)\mathrm{Trig}(j\varphi)\varepsilon^{t}(\varphi,t)\right) \tag{42}$$

$$=: p_{m,n}q_{n,j}\sum_{p=1}^{4}G^{\mathrm{I}}_{m,n,j;p}(r\xi(\varphi,t)), \tag{43}$$

with

$$I'(r) = \frac{4r}{\mathcal{R}^2}. \tag{44}$$

The computation of this gradient is shown in Section A2 in detail. In both cases, note that $\xi = \varepsilon^{r}, \varepsilon^{\varphi}$ and ε^{t} are vectors.

Then, we obtain the following values for the inner products. The detailed derivation is given in Section A3. For two polynomials, we have

$$\left\langle G^{\mathrm{I}}_{m,n,j}, G^{\mathrm{I}}_{m',n',j'}\right\rangle_{\mathcal{H}_1(\mathbb{B})} = \delta_{m,m'}\delta_{n,n'}\delta_{j,j'} + \delta_{n,n'}\delta_{j,j'}p_{m,n}p_{m',n} \tag{45}$$

$$\times\left[\frac{\mathcal{R}\sqrt{2}}{2^n}\int_{-1}^{1}\left(P_m^{(0,n+1/2)}(u)\right)'\left(P_{m'}^{(0,n+1/2)}(u)\right)'(1+u)^{n+3/2}\,\mathrm{d}u\right. \tag{46}$$

$$+\frac{\mathcal{R}n}{2^n\sqrt{2}}\int_{-1}^{1}\left[\left(P_m^{(0,n+1/2)}(u)\right)'P_{m'}^{(0,n+1/2)}(u)+P_m^{(0,n+1/2)}(u)\left(P_{m'}^{(0,n+1/2)}(u)\right)'\right](1+u)^{n+1/2}\,\mathrm{d}u \tag{47}$$

$$\left.+\frac{\mathcal{R}n(2n+1)}{2^{n+1}\sqrt{2}}\int_{-1}^{1}P_m^{(0,n+1/2)}(u)P_{m'}^{(0,n+1/2)}(u)(1+u)^{n-1/2}\,\mathrm{d}u\right], \tag{48}$$

where $\delta_{a,b}$ denotes the Kronecker Delta. Note that the remaining integrals must be computed numerically due to the exponents of $(1+u)$. For two FEHFs, we obtain

$$\langle N_{A,\Delta A}, N_{A',(\Delta A)'}\rangle_{\mathcal{H}_1(\mathbb{B})} \tag{49}$$

$$=\prod_{j=1}^{3}\int_{lb_{a_j}}^{ub_{a_j}}\frac{[\Delta A_j-|a_j-A_j|][(\Delta A_j)'-|a_j-A'_j|]}{\Delta A_j(\Delta A_j)'}\begin{Bmatrix}a_1^2, & j=1,\\ 1, & j=2,3\end{Bmatrix}\mathrm{d}a_j \tag{50}$$

$$+\sum_{k=1}^{3}\int_{lb_{a_k}}^{ub_{a_k}}\frac{\operatorname{sgn}(a_k-A_k)}{\Delta A_k}\frac{\operatorname{sgn}(a_k-A'_k)}{(\Delta A_k)'}\begin{Bmatrix}a_k^2, & k=1,\\ 1, & k=2,\\ 1-a_k^2, & k=3\end{Bmatrix}\mathrm{d}a_k \tag{51}$$

$$\times\prod_{j=1,\ j\neq k}^{3}\int_{lb_{a_j}}^{ub_{a_j}}\frac{\Delta A_j-|a_j-A_j|}{\Delta A_j}\frac{(\Delta A_j)'-|a_j-A'_j|}{(\Delta A_j)'}\begin{Bmatrix}\frac{1}{1-a_j^2}, & j=3,k=2,\\ 1, & \text{else}\end{Bmatrix}\mathrm{d}a_j, \tag{52}$$

where lb_{a_i} and ub_{a_i} are the lower and upper bound with respect to the dimension a_i of the intersection of the respective domains of the FEHFs $N_{A,\Delta A}$ and $N_{A',(\Delta A)'}$. Note that we need to determine these boundaries as well as the critical points in between them.

Finally, we consider the mixed case of a FEHF and a polynomial. This yields

$$\left\langle N_{A,\Delta A}, G^{\mathrm{I}}_{m,n,j}\right\rangle_{\mathcal{H}_1(\mathbb{B})}$$

$$= p_{m,n}q_{n,j}\int_{lb_r}^{ub_r}\frac{\Delta R-|r-R|}{\Delta R}P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^n r^2\mathrm{d}r\int_{lb_\varphi}^{ub_\varphi}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\mathrm{Trig}(j\varphi)\mathrm{d}\varphi\int_{lb_t}^{ub_t}\frac{\Delta T-|t-T|}{\Delta T}P_{n,|j|}(t)\mathrm{d}t \tag{53}$$

$$- p_{m,n} q_{n,j} \sum_{k=1}^{3} \int_{lb_{a_k}}^{ub_{a_k}} \frac{\operatorname{sgn}(a_k - A_k)}{\Delta A_k} \begin{cases} \left(P_m^{(0,n+1/2)}(I(a_k))\right)' I'(a_k) \left(\frac{a_k}{\mathcal{R}}\right)^n a_k^2 \\ + P_m^{(0,n+1/2)}(I(a_k)) \left(\frac{a_k}{\mathcal{R}}\right)^n n a_k, & k = 1 \\ j\,\mathrm{Trig}(-j a_k), & k = 2 \\ (1 - a_k^2) P'_{n,|j|}(a_k), & k = 3 \end{cases} \mathrm{d}a_k \tag{54}$$

$$\times \prod_{i=1, i\neq k}^{3} \int_{lb_{a_i}}^{ub_{a_i}} \frac{\Delta A_i - |a_i - A_i|}{\Delta A_i} \begin{cases} P_m^{(0,n+1/2)}(I(a_i)) \left(\frac{a_i}{\mathcal{R}}\right)^n, & i = 1 \\ \mathrm{Trig}(j a_i), & i = 2 \\ P_{n,|j|}(a_i), & i = 3, k = 1 \\ \frac{1}{1-a_i^2} P_{n,|j|}(a_i), & i = 3, k = 2 \end{cases} \mathrm{d}a_i \tag{55}$$

where $lb_r = lb_{a_1} = \max(R_{\min}, R - \Delta R)$, $ub_r = ub_{a_1} = \min(R_{\max}, R + \Delta R)$, $lb_\varphi = lb_{a_2} = \Phi - \Delta\Phi$, $ub_\varphi = ub_{a_2} = \Phi + \Delta\Phi$, $lb_t = lb_{a_3} = \max(T_{\min}, T - \Delta T)$ and $ub_t = ub_{a_3} = \min(T_{\max}, T + \Delta T)$. Note that we use the same notation as in eq. (43). The longitudinal integrals can be derived analytically (see Section A3). The radial and the latitudinal integrals must be computed numerically. Note that, in our experience, it is difficult but critical for our approach to determine the most efficient implementation for the latitudinal integrals.

4 The (LEARNING) INVERSE PROBLEM MATCHING PURSUITS

Next, we introduce our suggested algorithms for the approximation of ill-posed inverse problems: the (L)IPMPs. The IPMPs include: the Regularized Functional Matching Pursuit (RFMP) and the Regularized Orthogonal Functional Matching Pursuit (ROFMP). The LIPMPs are the respective counterparts that include a learning add-on: the LRFMP and the LROFMP. Note that there also exists the Regularized Weak Functional Matching Pursuit (RWFMP) whose idea can be included in the LRFMP and the LROFMP quite naturally as we explain below. To the best of our knowledge, the Geomathematics Group Siegen is among the first to adapt these algorithms for inverse problems. For more details and applications, see Berkel *et al.* (2011) Fischer (2011); Fischer & Michel (2012, 2013a, b); Michel & Telschow (2014, 2016); Telschow (2014); Michel (2015); Gutting *et al.* (2017); Michel & Orzlowski (2017); Leweke (2018); Kontak (2018); Kontak & Michel (2018, 2019); Telschow *et al.* (2018); Michel & Schneider (2020); Prakash *et al.* (2020); Schneider (2020) and Schneider & Michel (2022). We concentrate here on the LRFMP because it appears to be more suitable for traveltime tomography due to its efficiency. For this method, we introduce here the characteristics relevant for understanding and those that are newly adjusted to the particular problem at hand. Note, however, that the transfer of the latter to the LROFMP is straightforward and an implementation of it is included in the corresponding source code. We also direct the reader to the notations given in Section 2.

4.1 The (Learning) Regularized Functional Matching Pursuit

The RFMPs tackle inverse problems, such as the discretized, linearized traveltime tomography problem $T_\neg \delta S = \delta\psi_\neg$, which are often ill-posed by nature. The inevitable regularization for such kind of problems is implemented as a Tikhonov–Phillips regularization in these methods. This is an established and well-performing choice for many ill-posed inverse problems (see e.g. Louis 1989; Engl *et al.* 1996; Kirsch 1996; Hofmann 1999; Rieder 2003). An approximation f^* of the solution f is then found as the minimizer of the Tikhonov–Phillips functional. In particular, with the $\mathcal{H}_1(\mathbb{B})$ -Sobolev space introduced above, in the RFMP, we aim to determine f^* such that

$$\|\delta\psi_\neg - \mathcal{T}_\neg f^*\|_{\mathbb{R}^\ell}^2 + \lambda \|f^*\|_{\mathcal{H}_1(\mathbb{B})}^2, \qquad \lambda > 0, \tag{56}$$

is minimized. The first term is usually called the data misfit while the latter is the penalty term. Note that the corresponding Tikhonov–Phillips functional consists formally of only one penalty term instead of two (for smoothing and damping) as commonly used in seismology (see e.g. Charléty *et al.* 2013; Hosseini *et al.* 2020). We choose the $\mathcal{H}_1(\mathbb{B})$-Sobolev space for regularization here because we considered FEHFs for the approximation and they are tightly connected to (classical) Sobolev spaces. As we saw in eq. (33), the respective inner product is a sum of the $\mathrm{L}^2(\mathbb{B}, \mathbb{R})$- and the $\mathrm{L}^2(\mathbb{B}, \mathbb{R}^3)$-inner product due to which we re-enact the trade-off between smoothing and damping. Thus, this difference is only minor.

However, there is a more important difference between eq. (56) and the common seismological approach and it lies within the data misfit: due to uncertainties within the data, seismologists usually consider the (reduced) χ^2 in the computation process as well as for model selection (via the L-curve) instead of the pure residual $\delta\psi_\neg - \mathcal{T}_\neg f^*$ (see e.g. Hosseini *et al.* 2020). Mathematically, they are connected as follows:

$$\chi^2_{\mathrm{red}} = \frac{1}{\ell} \sum_{i=1}^{\ell} \left(\frac{(\delta\psi_\neg)_i - \mathcal{T}_\neg^i f^*}{\sigma_i} \right)^2 = \frac{1}{\ell} \left\| \frac{\delta\psi_\neg - \mathcal{T}_\neg f^*}{\sigma} \right\|_{\mathbb{R}^\ell}^2 \tag{57}$$

with

$$\frac{\delta\psi_\neg - \mathcal{T}_\neg f^*}{\sigma} := \left(\frac{(\delta\psi_\neg)_i - \mathcal{T}_\neg^i f^*}{\sigma_i} \right)_{i=1,\ldots,\ell}, \tag{58}$$

where $\sigma_i \in \mathbb{R}_+$ denotes the (known) uncertainty with respect to the ith measurement. As the uncertainty within the data cannot be circumvented, we, therefore, adjust the Tikhonov–Phillips functional considered in the RFMP for the case of traveltime tomography as well. Therefore, in the sequel, we consider the noise-cognizant Tikhonov–Phillips functional

$$\left\| \frac{\delta\psi_\neg - \mathcal{T}_\neg f^*}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \, \| f^* \|_{\mathcal{H}_1(\mathbb{B})}^2, \qquad \lambda > 0, \tag{59}$$

which shall be minimized in the RFMP. The minimizer f^* is then obtained iteratively as a linear combination of weighted dictionary elements $d_n \in \mathcal{D}$. Let N denote the current (or final) iteration. Then we have

$$f_N = f_0 + \sum_{n=1}^{N} \alpha_n d_n \tag{60}$$

in the case of the RFMP. Here, f_0 denotes a first approximation which is often the zero approximation in practice. As the dictionary is made of global polynomials and local FEHFs, the approximation consists—in all probability—of both types of trial functions.

The noise-cognizant Tikhonov–Phillips functional of the Nth step transfers then to

$$(\alpha, d) \mapsto \left\| \frac{R^N - \alpha \mathcal{T}_\neg d}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \, \| f_N + \alpha d \|_{\mathcal{H}_1(\mathbb{B})}^2, \qquad \lambda > 0, \tag{61}$$

$$R^{N+1} := R^N - \alpha_{N+1} \mathcal{T}_\neg d_{N+1} = \delta\psi - \mathcal{T}_\neg f_{N+1} \tag{62}$$

for the RFMP and with $R^0 = \delta\psi_\neg - \mathcal{T}_\neg f_0$ which yields $R^0 = y$ if $f_0 \equiv 0$. The main question is how to choose the dictionary element $d_{N+1} \in \mathcal{D}$ and the corresponding weight $\alpha_{N+1} \in \mathbb{R}$ such that the corresponding Tikhonov–Phillips functional is minimized. Similarly as in the literature on the (L)IPMPs, we can exchange the minimization of the noise-cognizant Tikhonov–Phillips functional by an equivalent maximization of the objective functions, see Section A4 for details:

$$\mathrm{RFMP}(d; N) := \frac{\left(\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{T}_\neg d}{\sigma} \right\rangle_{\mathbb{R}^\ell} - \lambda \, \langle f_N, d \rangle_{\mathcal{H}_1(\mathbb{B})} \right)^2}{\left\| \frac{\mathcal{T}_\neg d}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \, \| d \|_{\mathcal{H}_1(\mathbb{B})}^2} =: \frac{(A_N(d))^2}{B_N(d)}. \tag{63}$$

The weights are then easily obtained via

$$\alpha_{N+1} := \frac{A_N(d_{N+1})}{B_N(d_{N+1})}. \tag{64}$$

In practice, the IPMPs need termination and model selection criteria. As implemented for the Tikhonov–Phillips functional, we adjust them here as well. Usually, in seismological experiments, we would strive to let χ^2_{red} reach 1. This should yield the best trade-off between data (mis)fit, accuracy and smoothing. For the contrived data we use here, however, the uncertainty is assumed to be $\sigma \equiv 1$ s. This enables us to consider the relative data error $\| R^N \|_{\mathbb{R}^\ell} / \| R^0 \|_{\mathbb{R}^\ell}$ (as usually done in an IPMP) instead. If we used a varying uncertainty, then those data points with a high uncertainty, that is, a high value of σ_i, would be less important for our approximation as the first terms in the numerator and denominator of eq. (63) would be less significant. In other words, if we used a varying σ then we would trust in some data more than in some other. Then again, we should also rather consider χ_{red} than the RMSE. From the theoretical point-of-view, this does not appear to be problematic. However, it appears to be more reasonable to test this feature in the combination with the inversion of real data in the future.

In our experiments, we additionally perturb the delay vector with simulated noise. This again allows us to terminate the algorithm if the relative data error falls below this noise level. To avoid endless iterations for inappropriate parameters, we also set a maximum number of iterations. Among those models f_N obtained for diverse regularization parameters λ, we select the one (i.e. the regularization parameter) which yields the lowest relative root mean square error (RRMSE)

$$\left(\frac{\sum_{i=1}^{\kappa} \left(f\left(x^i\right) - f_N\left(x^i\right) \right)^2}{\sum_{i=1}^{\kappa} f\left(x^i\right)^2} \right)^{1/2} \tag{65}$$

for $\kappa \in \mathbb{N}$, where f is the (exact) solution which we use for our test and is given on the points x^i.

The IPMPs use by definition a finite dictionary $\mathcal{D}^{\mathrm{fin}} \subset \mathcal{D} = \mathcal{D}^{\mathrm{Inf}}$. In this case, the maximization of eq. (63) can be done by pairwise comparisons. However, this means that $\mathcal{D}^{\mathrm{fin}}$ must be chosen *a priori* either automatically or manually. The latter cannot be recommended in the case of the traveltime tomography because (i) we are inexperienced regarding which trial functions are needed as this is a novel application for the methods; (ii) the size of $\mathcal{D}^{\mathrm{fin}}$ grows tremendously due to the six characteristic parameters of the FEHFs and (iii) a possible bias by the choice of $\mathcal{D}^{\mathrm{fin}}$ cannot be quantified. In search of an automation of the *a priori* choice, the LIPMPs were developed [see Michel & Schneider (2020); Schneider (2020); Schneider & Michel (2022)]. They follow the same routine as the IPMPs but include an additional learning add-on. Then, the *a priori* dictionary choice is negligible. Though they do produce a learned dictionary which can be used as an automatically chosen

one in the IPMPs, they also proved to be useful as standalone approximation algorithms. Thus, in our experiment, we include the learning add-on, that is, we use the LRFMP.

4.2 The learning add-on

The idea is to allow all possible trial functions and, thus, use the infinite dictionary $\mathcal{D}$. As we saw before, $\mathcal{D}$ includes infinitely many trial functions of those types with continuous characteristic parameters, that is here the FEHFs. If the characteristic parameters are discrete as here with the polynomials, we still allow only a finite set. In each iteration, we determine optimized dictionary elements (or candidates) for each type of trial functions separately. Together they form again a very small, finite dictionary of candidates from which we obtain the overall most suitable function. Thus, the learning add-on is the determination of the finite dictionary of candidates in each iteration.

Recall that we allow polynomials up to a maximally possible radial and angular degree. The global trial functions in the dictionary are usually chosen to reconstruct global trends and, thus, high degrees and orders are often negligible. Hence, it suffices to consider some maximum radial and angular degree. Theoretically, these maximally possible degrees can be chosen to be very high such that the methods can learn a maximally needed radial and angular degree (see Schneider 2020; Schneider & Michel 2022). It remains to be seen whether this can be done in practice for traveltime tomography due to efficiency reasons [the curse of dimensionality occurs here: the set of all (Cartesian) polynomials $G^{\mathrm{I}}_{m,n,j}$ with degree $\leq N$ has a size of $\mathcal{O}(N^3)$, see Michel (1999)]. Note that, due to the finiteness, we can still obtain the maximizer of eq. (63) by pairwise comparisons. Furthermore, we can still use the common preprocessing routine of the IPMPs for the polynomials and improve efficiency in this way. Practically, this means we define a finite starting dictionary $\mathcal{D}^s$ which contains at least the chosen polynomials from $\mathcal{D}$.

In the case of the FEHFs, we use the truly infinite set of possible trial functions $[\mathcal{A}]_{\mathrm{FEHF}}$. The main challenge here is to maximize eq. (63) among all possible FEHFs. As a matter of fact, this is a non-linear constraint optimization problem with the objective function RFMP($N_{A,\,\Delta A}$; N) $\rightarrow$ max ! For maximizing RFMP($N_{A,\,\Delta A}$; N), recall the corresponding constraints with respect to A and ΔA given in eqs (15–17). We can solve this maximization with any kind of established optimization routine. For instance, methods from the NLOpt library [see Johnson (2019)], can be utilized. Note, however, that a gradient-based approach cannot be used here because it would necessitate the computation of the gradient of RFMP($N_{A,\,\Delta A}$; N) with respect to (A, ΔA). Unfortunately, this is not possible for practical purposes due to the absolute value in the definition of the function (see eq. 25). Further, from experience, we propose a 2-step-optimization procedure, that is, we run two optimization algorithms. First, we use a global method. Then we refine this solution by using a local counterpart starting from the former solution. This also enables us to soften the accuracy of the global optimization technique (i.e. its termination criteria) which decreases the runtime. Note that softening the termination criteria of the optimization is similar to including a weakness parameter as in the RWFMP. Also note that both solutions are inserted into the dictionary of candidates. Moreover, if the global method needs a starting point, we should insert a few FEHFs in the starting dictionary for this reason as well. Further note that this starting solution can also be included in the dictionary of candidates but should generally not be chosen, that is this starting solution should not have a major influence on the learnt approximation in general.

4.3 Additional divide-and-conquer strategy for traveltime tomography

We experienced that the optimization within the learning add-on is slowed down—amongst others—by the use of many rays at once. Thus, we considered how to sensibly use fewer rays at once while taking a numerically significant number of rays into account. This yields an additional divide-and-conquer strategy for our challenge at hand which proved to be helpful in practice. The main idea is to start with a low number of rays. In our experiments, we started with 1000 rays. If the relative data error falls below a certain threshold such as 50 per cent then we add the next package of 1000 rays to our consideration. We consider then the first 2000 rays in our algorithm with the exception of optimization of the FEHFs where we consider only the latest package of 1000 rays.

We are aware that this is a quite manual approach with certain seemingly arbitrary parameters. In the meantime, we also considered other aspects that concerned the efficiency such that it seems now to be possible to increase, for instance, the size of each ray package. This may form the basis of future research.

5 NUMERICAL IMPLEMENTATION AND TESTS

In this section, we show a numerical proof of concept for our TTIPMP code. We first introduce our experiment setting for reproducibility and afterwards our numerical results. Note that the corresponding code is available online (Schneider 2023).

The first-order objective of this paper is to describe a new ansatz for seismic traveltime tomography and to demonstrate that the method works in principle. For this reason, we consider it, first and foremost, to be sufficient to design a simple test example. Certainly, seismic tomography is always connected to challenges regarding the computational complexity and burden. Our approach is no different. In the future, further enhancements need to be done, which could, in combination with progress in computer technology, allow the use of larger data sets and more sophisticated test scenarios such as a chequerboard test. Since those aspects also mean that we would not yet claim to have a competitive method, it does not make sense to compare numerical results with those obtained with other methods.

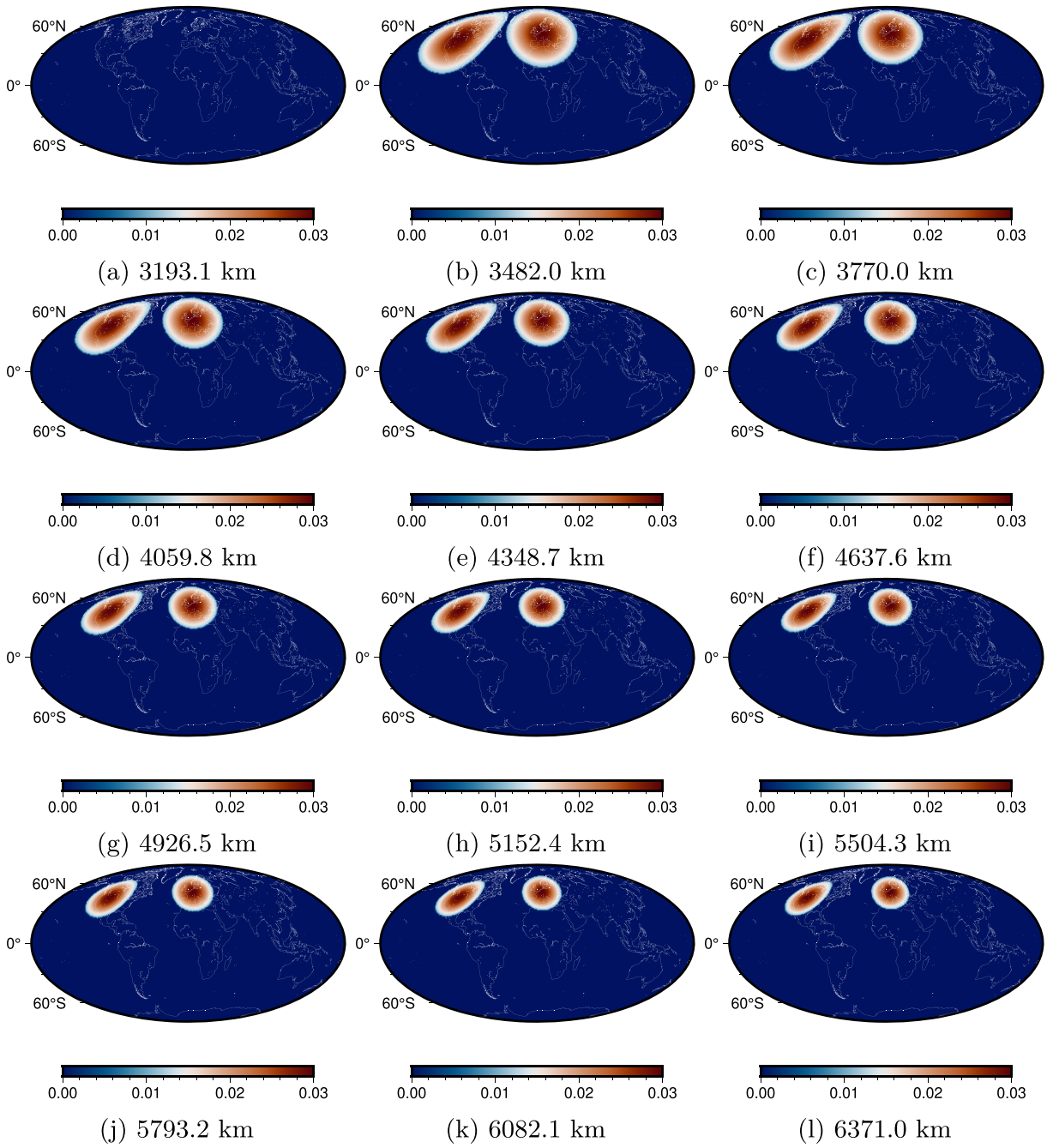

Figure 3. Resolution test input model for *P*-velocity consisting of two plumes below the Volcanic Eifel and Yellowstone volcano, shown at spheres with different radii. The values are in per cent.

5.1 Chosen earthquake data and specific experiment setting

We use the unit ball for our practical computations and show here the approximation of synthetic input structures modelled on mantle plumes, that is seismically slow, vertical, columnar features that extend from the core–mantle boundary to the surface, as a model for the deviation of the slowness in the interior of the Earth. That is, we show here a resolution test. The plumes are constructed to extend from the core–mantle boundary up to the Earth's surface. They lie below the Volcanic Eifel and the Yellowstone volcano. Note that vertically continuous plume conduits are among the structures suspected to occur in the real Earth (see e.g. Boschi *et al.* 2007; Tsekhmistrenko *et al.* 2021). For a visualization of this model (see Fig. 3). The diameter of the plumes are 1/40 of the circumference of the respective depth, that is 1718.30 km at the core–mantle boundary and 3143.96 km at the Earth's surface.

This enables us to compute a relative root mean square error as given in eq. (65) where we use a grid of $\kappa = 12 \cdot 65341 = 784\,092$ data points. This grid is given as 12 layers of an equi-angular grid, commonly also called a Driscoll–Healy grid (see e.g. Driscoll & Healy 1994;

Michel 2013), of 361 · 181 gridpoints. The data are perturbed with 5 per cent Gaussian noise, such that we have perturbed data y^δ given by

$$y_i^\delta = y_i \cdot (1 + 0.05 \cdot \varepsilon_i) \tag{66}$$

for the unperturbed data $y_i = (\mathcal{T} f)\left(\sigma^i \eta^i\right)$ and a unit normally distributed random number ε_i.

As we aim for a proof of concept, it suffices to consider a ray theoretical setting and the ISC-EHB seismic metadata [see Weston *et al.* 2018; The International Seismological Centre (ISC) 2022]. Note that these data were among others also used in Hosseini *et al.* (2020). In particular, we are using the ISC-EHB metadata from 1998 to 2016, respectively, but only the *P* waves. We use only the data up to 2016 which was thinned out a bit to be concentrated in Southeast Asia (western Pacific) as this was the working set of the seismological project partners (*cf.* paper by Totten *et al.* in preparation). Note, however, due to the divide-and-conquer strategy, we might not use all rays from these years in practice. At most we have used 318 542 rays. According to our strategy, we start with rays in 2016 and add further rays by moving back in time.

We correct these data with all necessary seismological corrections. The corrections are done in line with the tomography workflows exhibited in Hosseini *et al.* (2020), Mohammadzaheri *et al.* (2021) and Tsekhmistrenko *et al.* (2021). Moreover, with these metadata, we have a constant uncertainty $\sigma_i = 1\,\mathrm{s}$, $i = 1, ..., \ell$. In general, we cannot expect the rays to illuminate the Earth evenly, in particular also due to the use of thinned out data (see above). That is, we have to take into account that all of our metadata—the receiving seismological stations as well as the rays themselves—will be poorly distributed to a certain extent. In particular, the path coverage is sparse in shallow depths under the largely uninstrumented oceans as well as in the deepest parts of the mantle since we exclude core-diffracted body-wave paths. For a visualization of the ISC-EHB metadata rays, see Fig. 4. Despite the irregular distribution of the data, they still appear to be sufficiently concentrated near our structures. We consider the challenge posed by this distribution as a stress test for our method.

We solve the latitudinal integrals in the inner products of a polynomial and an FEHF with a Gauß–Legendre quadrature rule of 10^6 points and use an adaptive Gauß–Kronrod quadrature rule with an integration error of 10^{-4} anywhere else (i.e. for the DSPO as well as for other integrals in the inner products). In particular, the latter is necessary for efficiency reasons. We set $\epsilon_R = \epsilon_\Phi = \epsilon_T = 10^{-2}$ and $\rho = 3482/6371$. Recall that this sets the lower bound for the value of R of an FEHF to the core-mantle boundary in our setting. We use the GN_DIRECT_L and the LN_SBPLX algorithm for the global and local optimization in the learning add-on. We terminate the global algorithm if succeeding iterates (i.e. variation in the optimal point itself) vary less than 10^{-4} (i.e. xtol_rel $= 10^{-4}$) or succeeding function values (i.e. variation in the objective function of the current iterate) vary less than 10^0 (i.e. ftol_rel $= 10^0$). We terminate the local algorithm if successive iterates vary less than 10^{-8} or successive function values vary less than 10^{-4}. In analogy to Schneider (2020) and Schneider & Michel (2022), we terminate the optimization after 10 000 evaluations of the objective function or 600 s computation time. From experience, these termination criteria ensure that the optimization happens within a suitable time frame. Further, the loss in accuracy is negligible for our proof of concept. For more information on the termination criteria of the optimization methods, see Johnson 2019. We terminate the LRFMP either after 300 iterations, or if $\|R^N\|_{\mathbb{R}^\ell}/\|y\|_{\mathbb{R}^\ell}$ is less than the noise level or greater than 2 or if $|\chi^2_{\mathrm{red}} - 1| < 10^{-8}$. The regularization parameter is chosen such that the RRMSE is minimized after terminating successfully which means that either the iteration criterion or the residual criterion is fulfilled. We present here the result for $\lambda = 10^{-3}|y|$.

As a finite (starting) dictionary, we utilize

$$[\mathcal{G}_{5,5}]_{\mathrm{GI}} = \{(m, n, j) \,:\, m,\ n \in \mathbb{N}_0,\ m \le 5,\ n \le 5,\ j \in \mathbb{Z}_0,\ |j| \le n\}, \tag{67}$$

$$[\mathcal{A}]_{\mathrm{FEHF}} = \left\{(A, \Delta A) \,:\, A \in D_p,\ \Delta A = \Delta D_p\right\}, \tag{68}$$

$$D_p = \left\{\frac{3482}{6371} + \frac{2889i}{25484} \,:\, i = 0, ..., 4\right\} \times \left\{\frac{\pi i}{2} \,:\, i = 0, ..., 4\right\} \times \left\{-1 + \epsilon_T + \frac{(1-\epsilon_T)i}{2} \,:\, i = 0, ..., 4\right\}, \tag{69}$$

$$\Delta D_p = \left(\frac{2889}{25484}, \frac{\pi}{2}, \frac{1-\epsilon_T}{2}\right)^{\mathrm{T}}, \tag{70}$$

$$\mathcal{D} \coloneqq [\mathcal{G}]_{\mathrm{GI}} \cup [\mathcal{A}]_{\mathrm{FEHF}}. \tag{71}$$

Note that $\rho\mathcal{R} = 3482/6371$, $\mathcal{R} = 1.0$ and, thus, $\mathcal{R} - \rho\mathcal{R} = 2889/6371$ in our setting.

5.2 Synthetic inversion tests

We chose the regularization parameter to be $10^{-3}\|y\|_{\mathbb{R}^\ell}$ as this produced the lowest RRMSE among the tested values of parameters. The experiment terminated after 300 iterations with an RRMSE of 0.881173 and a relative data error of 0.154958. The absolute approximation error is shown in Fig. 6. The approximation is shown in Fig. 5. We scaled the figures to the size of the solution for better comparability (compare colour scales in Figs 3 and 6). We note that the errors are low for intermediate depths, where our teleseismic body wave paths sample the mantle most extensively, and higher for large and shallow depths. This is not surprising because of the irregular illumination of the Earth, which is common for teleseismic body waves and is intensified here due to the regional data set (see Fig. 4).

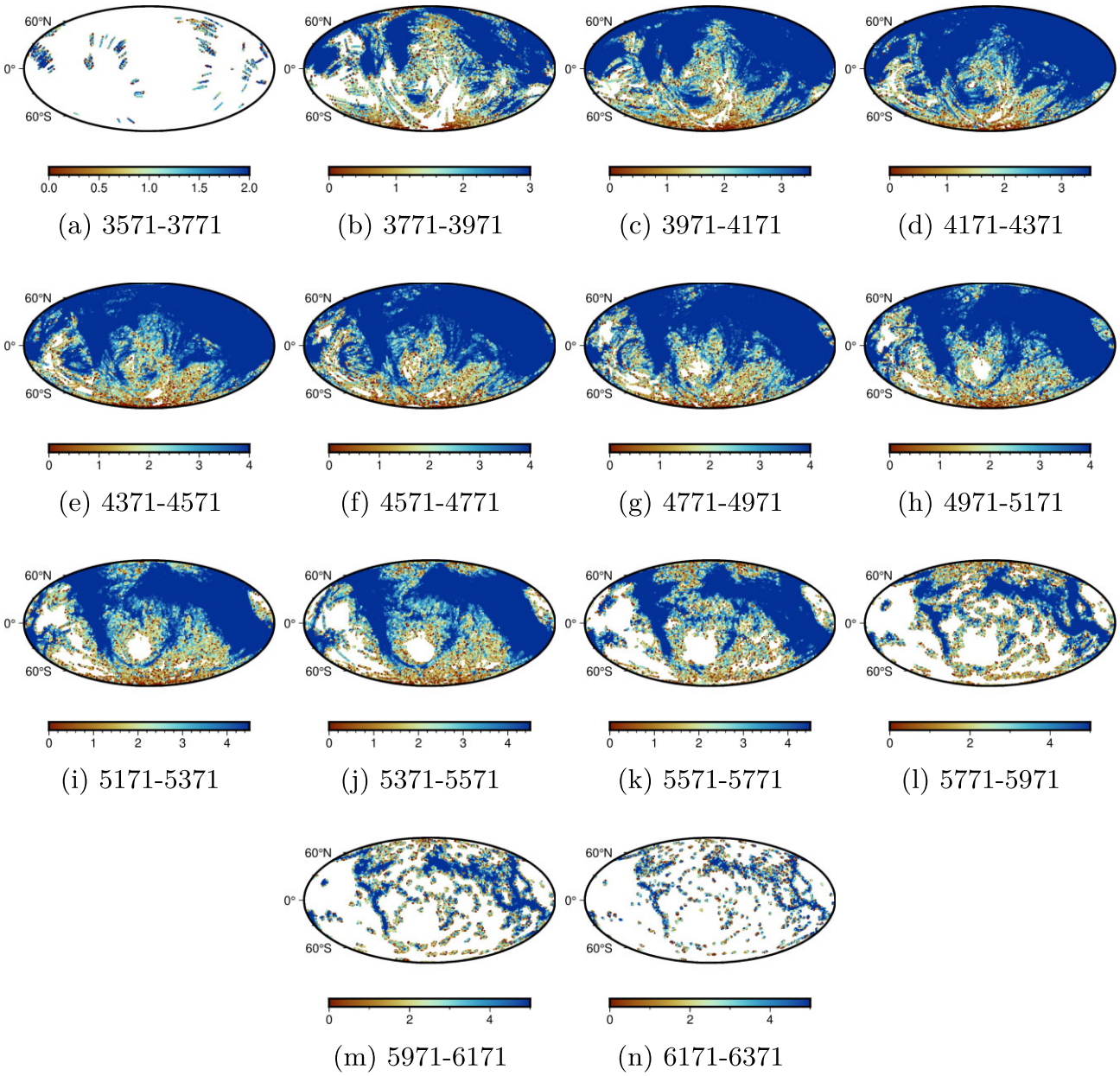

Figure 4. Distribution of ISC-EHB metadata rays at different spherical shells (radius in km given in subcaptions). The colour gives the number of obtained ray points within each $1^\circ \times 1^\circ$-degree voxel. White stands for no ray points in this voxel. The colourscale refers to the logarithm (to the base 10).

In particular, at the distances from 3770.9 to 4926.5 km to the centre, we see only very little remaining errors, and some of the errors are simply due to boundary effects since the method is unable to recover the precise geometry of the plumes. Moreover, those are close to the plumes, that is the region where our structure is given. Also in larger depths (radial range from 3193.1 to 3482.0 km), the main errors are situated in the Northern Hemisphere where our structure is located from the core–mantle boundary upwards. Since there is practically no data (see Fig. 4), in these depths, the method cannot register that the plumes are cut off at the core–mantle boundary and, further, it does not introduce random artefacts there but continues the structures. We assume that using additionally core-diffracted waves as well would erase the errors in these depths. In shallower depths (radii 5215.4–6082.1 km), we also see a similar continuing behaviour of the approximation. Comparing the approximation in these depths with the data distribution, we observe that the errors do not increase as rapidly as the data becomes sparse. Hence, also here we have a continuation of structure within sparser data regions without many artefacts. The errors increase in particular below the Southern Pacific at radii 5504.3 and 5793.2 km, which is typically not a very active seismic region and the data is not well-distributed there. Unfortunately, at the Earth's surface we have many artefacts. As the sparsity there is extremely high, we assume that a certain sparsity level can also be a limit to the method. However, note that the Earth's surface represents also the boundary of the considered regions (boundaries are known to possibly introduce additional challenges in approximation tasks).

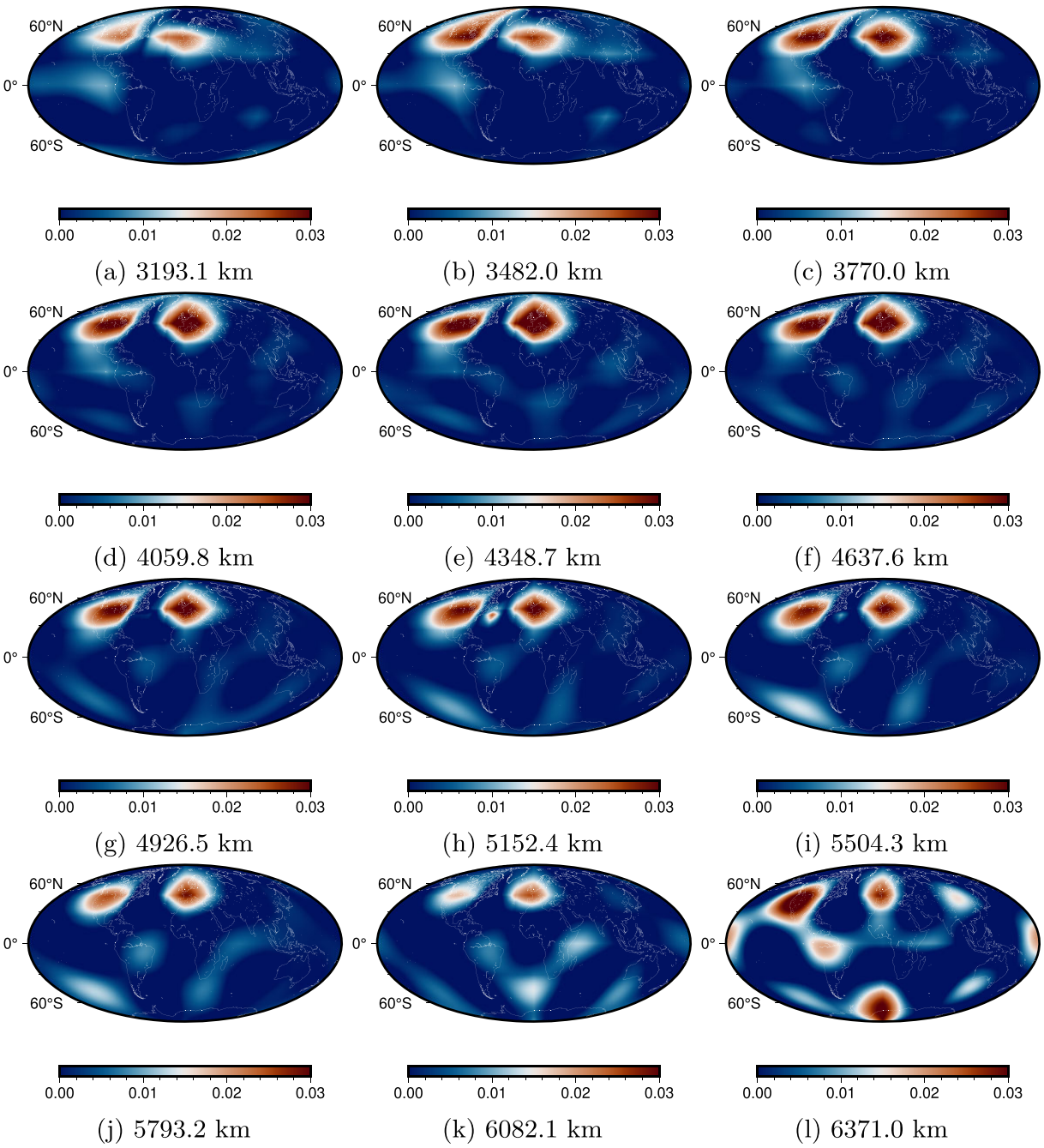

Figure 5. Approximation of the plumes model given in Fig. 3. The values are in per cent. The colour scales are adjusted for a better comparability, that is the colourbar always has the same range such that it can be easily compared with Fig. 3.

In Figs 7, 8 and 9, we show the chosen dictionary elements. Further, in Fig. 10, we provide a comparison to the FEHFs that were given in the starting dictionary. First of all, we note that the method chooses both local and global functions, much more polynomials than FEHFs to be precise. In particular, higher degrees and orders are preferred in general. This supports the idea that, due to the irregularly distributed data, the method uses global functions which fill empty blanks with similar structures as in regions with many data. This is at least the case, as long as there are enough data points nearby such as in medium depths. Thus, it would be interesting to investigate how the method would work with even higher degrees and orders. One possible hypothesis is that the method uses even fewer FEHFs in this instance. Regarding the chosen FEHFs, we see that the LRFMP obtains FEHFs as solutions of its optimization procedure (i.e. they are, in this respect, the best basis functions) in depths where more data is given. Nevertheless, also at these depths, the algorithm only selects a few of these basis functions. In regions where the data is sparse, the method falls back to the FEHFs given in the starting dictionary. In general this is not really desired. Note, however, that this also happens majorly near the Earth's surface which represents a specific challenge as discussed above. However,

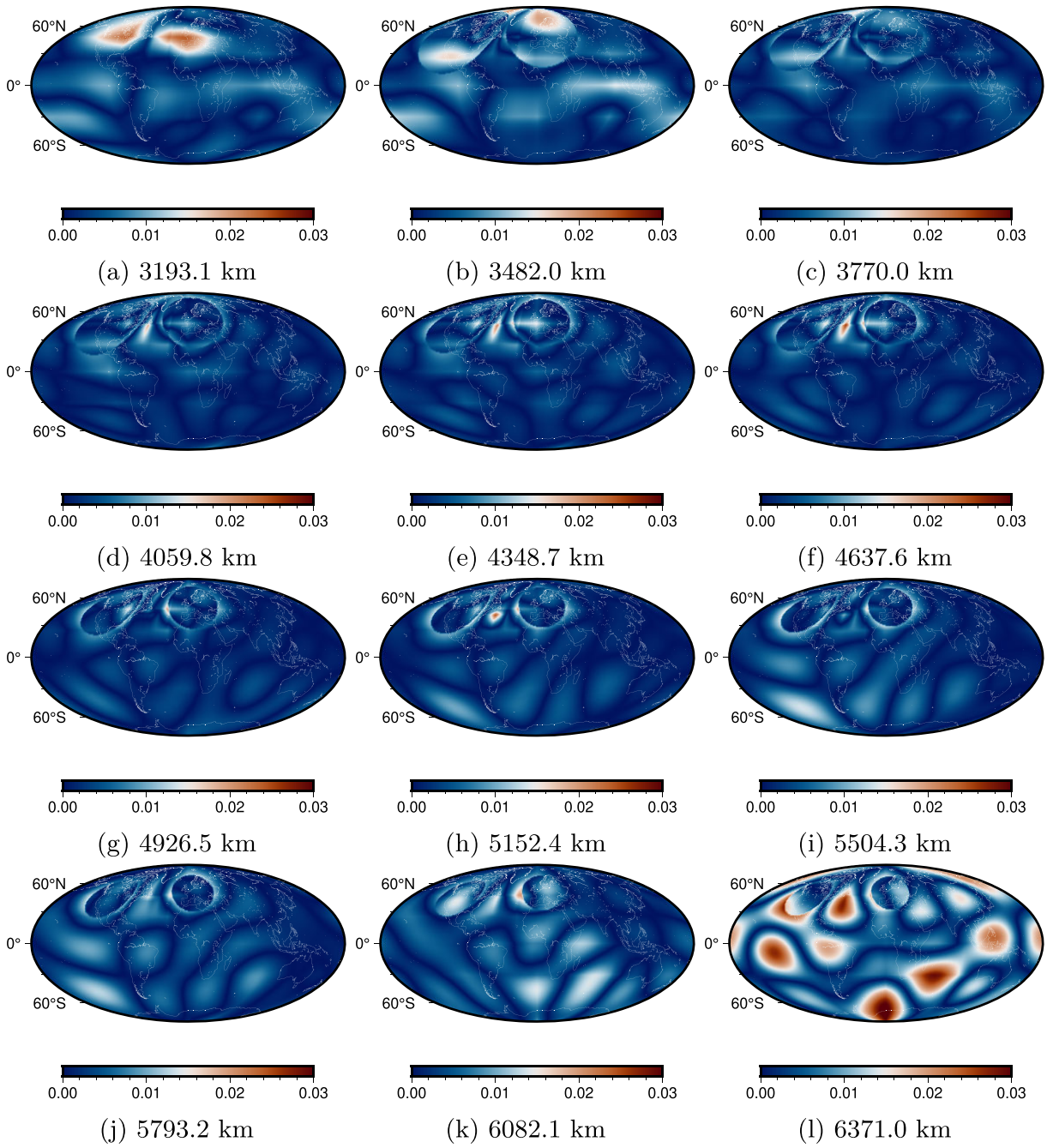

Figure 6. Absolute approximation error for the approximation of the plumes model given in Fig. 3. The values are in per cent. The colour scales are adjusted for a better comparability, that is the colourbar always has the same range to show where the errors are overall larger.

since the FEHFs are also generally less often chosen, this outcome suggests that FEHFs may not be the best choice within the LRFMP for traveltime tomography but with only sparse data the polynomials are also not suitable. It remains a question for future research to determine which alternative types of trial functions in the dictionary can yield better numerical results. For other applications such as gravitational field modelling, the LRFMP proved to perform well for the combination of orthogonal polynomials (spherical harmonics in this case) and localized scaling functions and wavelets constructed from reproducing kernels (see e.g. Schneider *et al.* 2023). However, the analogues of such kernels on balls are connected to essentially higher numerical costs, which is why we have not chosen them for our first experiments demonstrated here. In light of this, while this approach shows promise, more work needs to be done to improve the efficiency of their numerical calculation and integration.

Hence, we observe that the LRFMP can provide the possibility to reconstruct plumes within the Earth and constrain errors to their spatial locale, though the method certainly leaves the room for further improvements.

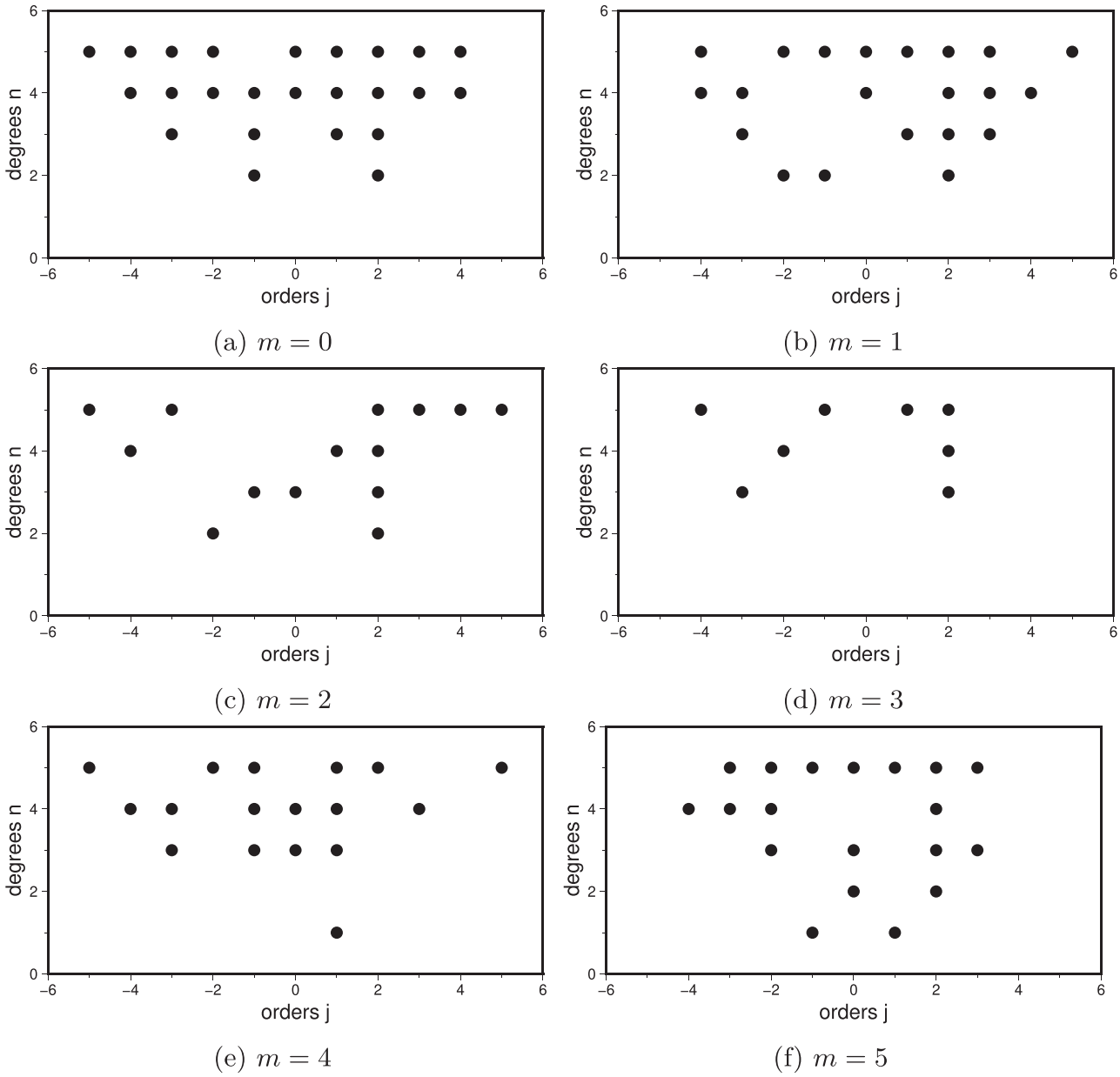

Figure 7. Chosen polynomials for the given radial degrees m.

6 DISCUSSION

The synthetic tests worked in the technical sense that our algorithm seems to have been correctly implemented and behaving broadly as expected, that is the test input structures were recovered fairly well. However, we hit computational limits even with this first, relatively 'easy' input pattern of two broad, slow plumes embedded into a background model. These limitations prevented us from proceeding to resolution tests of 'harder', more complex structures (e.g. chequerboards), or of more geologically realistic structure (many thin plumes, subducting slabs, etc.). This means we are not yet able to comment on hoped-for behaviours such as the algorithm selecting certain dictionary functions in well-sampled regions and reusing them in poorly sampled regions. This could provide an attractive strategy of acquiring knowledge in one region and carrying it *a priori* into less knowable regions of the mantle, which would however be expected to be physically analogous—an intelligent regularization technique. Extremely uneven sampling is a defining characteristic of global seismic tomography, and widely applied Tikhonov regularization strategies (norm damping, smoothing) are clearly simplistic. A judgement whether LRFMP can improve on those will have to await our technical ability to compute more complex test patterns.

The inverse problem of traveltime tomography is ill-posed. Unfortunately, we lack other, helpful theoretical insights such as a singular value decomposition of the corresponding forward operator, which makes the numerical computations rather time-consuming. Due to the lack of numerically exploitable properties of the respective operator, we have to take special care on how to use a sensible amount of rays in practice.

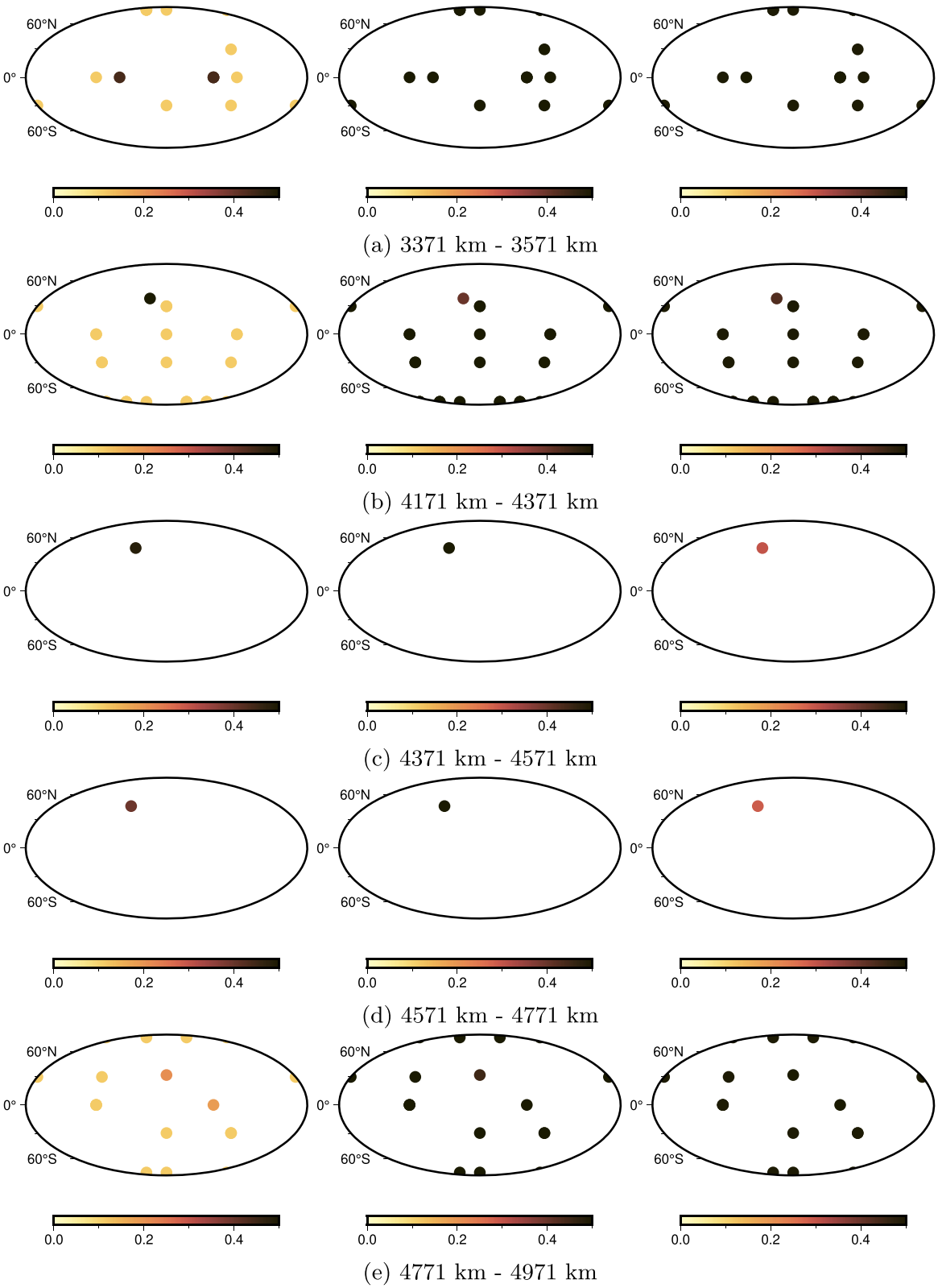

Figure 8. Chosen FEHFs within different spherical shells (*cf.* radii in Figure 4): the positions of the dots refer to the positions of the tesseroids' centre and the colours refer to their size in the three polar-coordinate directions, namely ΔR (left column), $\Delta\Phi$ (middle) and ΔT (right).

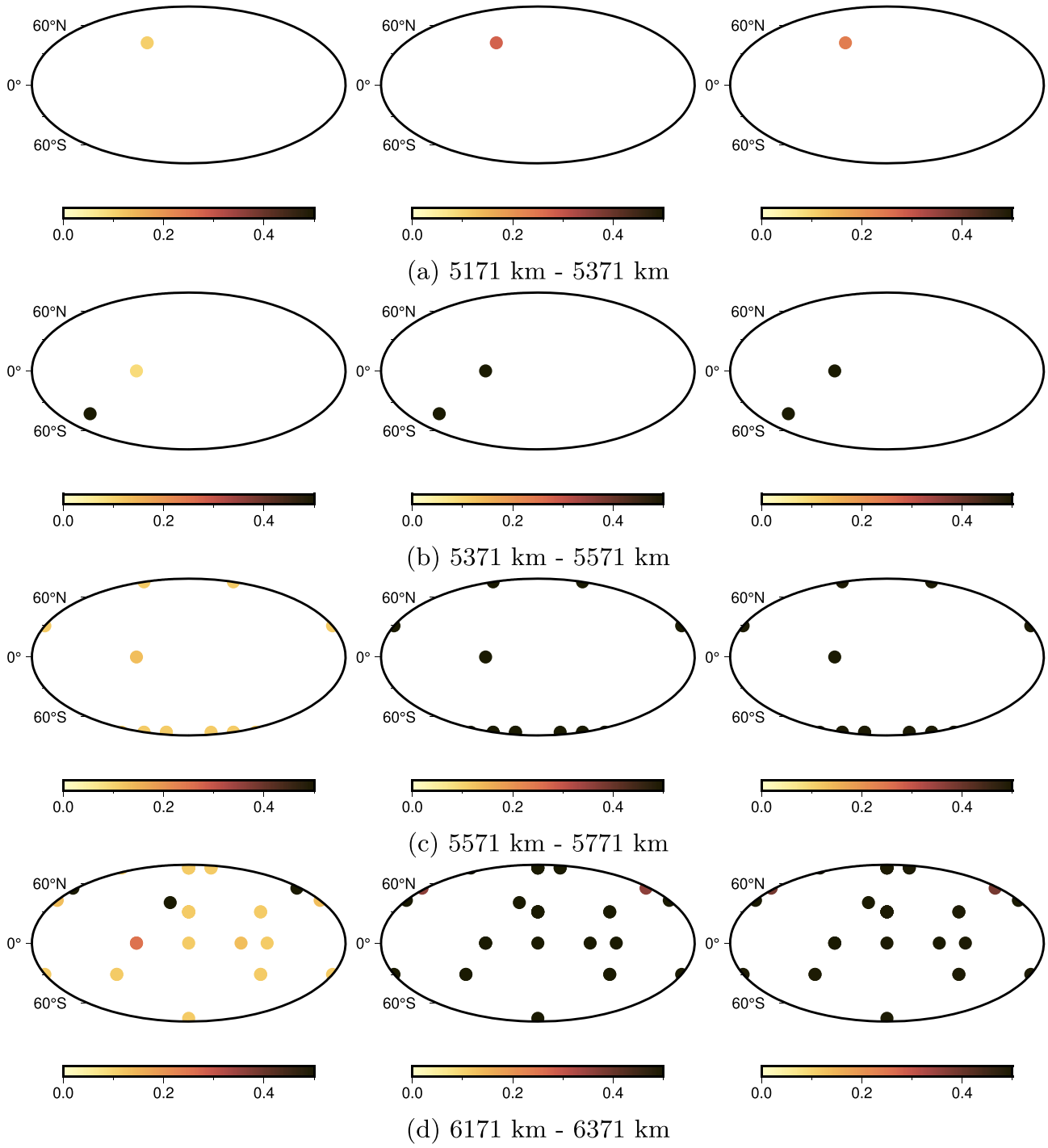

Figure 9. Illustration of the chosen FEHFs (see also Fig. 8).

In retrospect, it might have been preferable to start with a lower-dimensional test inverse problem, for example for mantle structure beneath one well-illuminated continent such as Europe or North America, though still in a spherical geometry. Ray theoretical or finite-frequency inversions have achieved very high resolution in such setups with only tens of thousands of wave paths (e.g. Sigloch 2011). This approach would have reduced the number of rays in our computations, compared to our >300 000 rays, which may still be on the low side for a global *P*-wave inversion. It would still have been possible to observe the selection of dictionary functions in under-constrained regions (around the edges of the targeted continent) and in the large non-constrained volumes that make up the rest of the mantle globally.

Besides these computational aspects, further work towards applicability in practice should include an efficient treatment of corrections for earthquake hypocentres ('source relocations'). Another question emerges from the result: which trial functions are best for this inverse

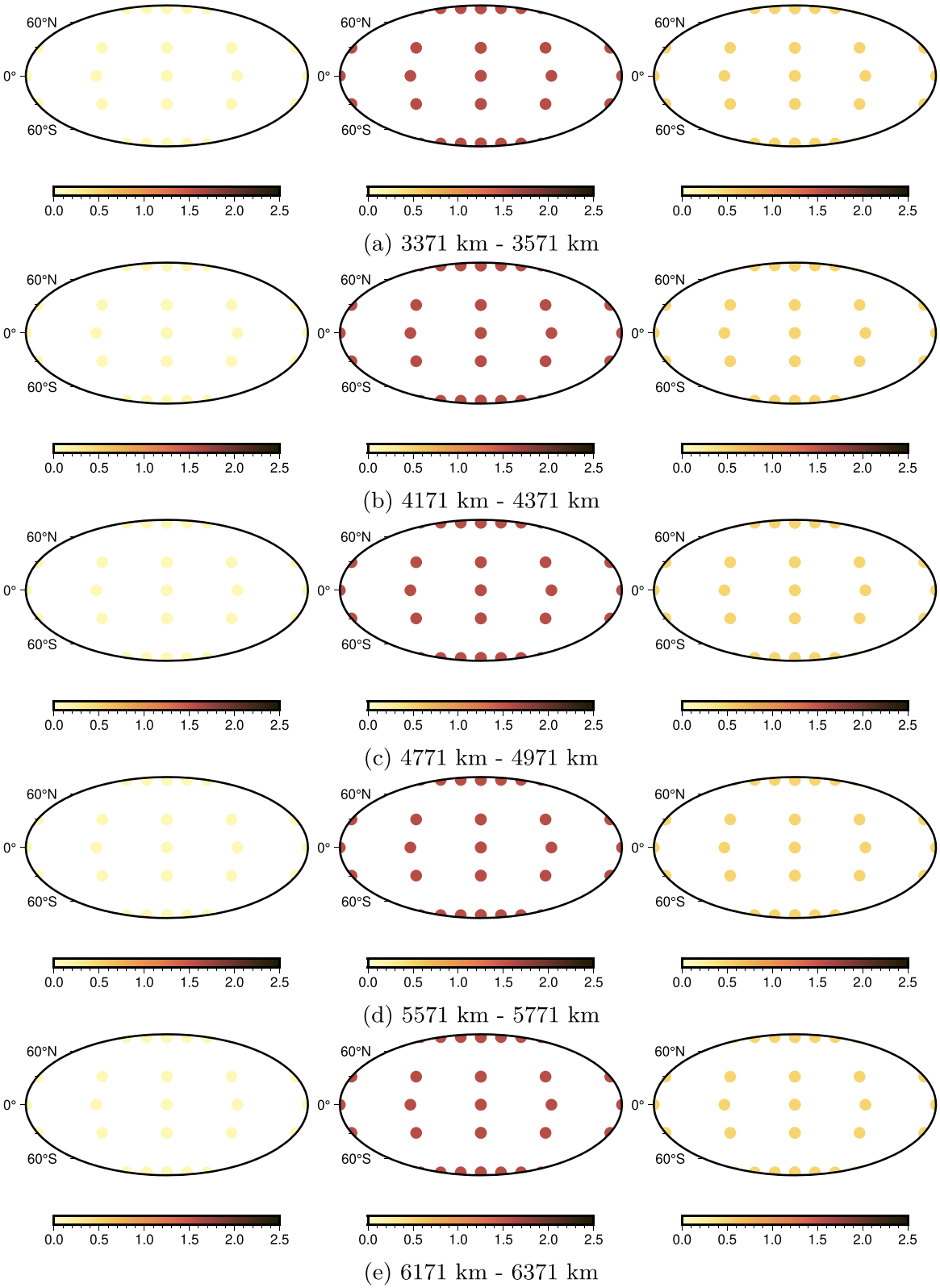

Figure 10. FEHFs of the starting dictionary within the given depth ranges (the illustration is analogous to Figs 8 and 9).

problem? Although we tried finite element hat functions in order to obtain a model that is comparable with other approaches, the results point towards a different dictionary for future research.

In general, we conclude that the LIPMPs may promise to provide an alternative numerical regularization method in comparison to other methods for traveltime tomography. In order to achieve competing status, the method still needs to be enhanced. Hence, this paper should be seen as a report about the intermediate results in the development of the LIPMPs for seismic tomography.

7 CONCLUSION AND OUTLOOK

Here, we proposed to use the LRFMP algorithm which yields an approximation in a chosen best basis of dictionary elements. For the latter, we allowed polynomials as well as tesseroid-based linear finite element hat functions in order to include global and local trial functions in the dictionary. As this is the first application of the LRFMP to traveltime tomography, we aimed for a proof of concept and followed the ray-theoretical approach.

Here we presented the method in general with an emphasis on aspects that have to be remodelled for traveltime tomography. These are—in addition to the choice of dictionary elements—the choice of a regularization space and the practical use of a necessary amount of data. For the former, we chose a Sobolev space due to the deep mathematical connections between finite elements and these spaces. For the latter, we introduced an additional divide-and-conquer strategy which allowed us to consider packages of rays iteratively and, in this way, overall consider a suitable amount of rays to obtain a proof of concept.

In our experiments, we considered a contrived Earth model consisting of two cone-shaped plumes between the core–mantle boundary and the Earth's surface. Our results showed that the LRFMP is able to reconstruct such structures. The approximation contains more errors where the data is relatively sparse.

Moreover, a numerical experiment with a varied data scenario showed that more sophisticated tests such as a chequerboard test would require further adaptations and enhancements of the method. For instance, the divide-and-conquer strategy (see Section 4.3) deals with an arbitrary choice of packages of 1000 rays. For complicated global structures, in contrast to our plume example, this might not be sufficient, since such few rays most likely cannot cover the anomalies sufficiently. In such cases, larger ray packages might make sense, but this requires a more scientific than arbitrary heuristic for choosing the package sizes. On the other hand, larger data packages will increase the computational complexity, which needs to be balanced as well.

In summary, future research could deal with a parameter study for the divide-and-conquer strategy in combination with other accuracy-lowering and efficiency-improving approaches as well as tackling more complicated Earth models.

ACKNOWLEDGMENTS

VM gratefully acknowledges the financial support by the German Research Foundation (DFG; Deutsche Forschungsgemeinschaft), project MI 655/14-1. KS was supported by the French government through the UCAJEDI Investments in the Future project, reference number ANR-15-IDEX-01. EJT was funded by the Clarendon scholarship of Oxford University Press (PhD) and supported by the UK National Environmental Research Council Doctoral Training Partnership in Environmental Research. We thank Maria Tsekhmistrenko, who maintains the tomographic inversion software on GitHub, and Afsaneh Mohammadzaheri, who assisted EJT with some data in the time of the COVID-19 pandemic. Last but not least, we are grateful for using the HPC Clusters Horus and Omni maintained by the ZIMT of the University of Siegen for our numerical results.

AUTHOR CONTRIBUTIONS

The research was carried out during NS's postdoc project. In this DFG-funded project, VM was the principal investigator. KS was the project partner from geosciences. Both supervised the project and assisted NS. EJT provided available data and software and assisted in the preparation of the tests.

CONFLICTS OF INTEREST

Not applicable.

AVAILABILITY OF DATA, MATERIAL AND CODE

The code is available at https://doi.org/10.5281/zenodo.8227888 under licence the CC-BY-NC-SA 3.0 DE. The solution shown here can be computed from it. The used and corrected rays accompany the code in a compressed format. Note that the ISC-EHB metadata is also freely available online.

REFERENCES

Abramowitz, M. & Stegun, I.A., 1972. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables,* 10th edn, Dover Publications, Inc.

Adams, R.A. & Fournier, J. J.F., 2003. *Sobolev Spaces,* 2nd edn, Elsevier Ltd.

Aki, K. & Richards, P.G., 2009. *Quantitative Seismology,* University Science Books.

Amirbekyan, A., 2007. The application of reproducing kernel based spline approximation to seismic surface and body wave tomography: theoretical aspects and numerical results, *PhD thesis*, University of Kaiserslautern, AG Geomathematik, http://kluedo.ub.uni-kl.de/volltexte/2007/2103/.

Ballani, L., Engels, J. & Grafarend, E.W., 1993. Global base functions for the mass density in the interior of a massive body (Earth), *Manuscripta Geodaetica,* **18,** 99–114.

Ben-Menahem, A. & Singh, S.J., 2000. *Seismic Waves And Sources,* 2nd edn, Dover Publication.

Berkel, P., Fischer, D. & Michel, V., 2011. Spline multiresolution and numerical results for joint gravitation and normal mode inversion with an outlook on sparse regularisation., *Int. J. Geomath.,* **1**(2), 167–204.

Bhattacharyya, P.K., 2012. *Distributions: Generalized Functions with Applications in Sobolev Spaces,* Walter de Gruyter GmbH & Co. KG.

Boschi, L., Becker, T.W. & Steinberger, B., 2007. Mantle plumes: dynamic models and seismic images, *Geochem. Geophys. Geosyst.,* **8,** doi:10.1029/2007GC001733.

Braess, D., 2007. *Finite Elemente – Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie,* 4th edn, Springer.

Charléty, J., Voronin, S., Nolet, G., Loris, I., Simons, F.J., Sigloch, K. & Daubechies, I.C., 2013. Global seismic tomography with sparsity constraints: comparison with smoothing and damping regularization, *J. Geophys. Res.,* **118**(9), 4887–4899.

Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology,* Princeton Univ. Press.

Dahlen, F.A., Hung, S.-H. & Nolet, G., 2000. Fréchet kernels for finite frequency traveltimes – I. Theory, *Geophys. J. Int.,* **141**(1), 157–174.

Driscoll, J.R. & Healy, D.M., 1994. Computing Fourier transforms and convolutions on the 2-sphere, *Adv. Appl. Math.,* **15**(2), 202–250.

Dufour, H.M., 1977. Fonctions orthogonales dans la sphère. Résolution théorique du problème potentiel terrestre, *Bull. Géod.,* **51,** 227–237.

Dunkl, C.F. & Xu, Y., 2014. *Orthogonal Polynomials of Several Variables,* 2nd edn, Cambridge Univ. Press.

Engl, H.W., Hanke, M. & Neubauer, A., 1996. *Regularization of Inverse Problems, Mathematics and Its Applications,* Kluwer Academic Publishers.

Fischer, D. & Michel, V., 2012. Sparse regularization of inverse gravimetry – case study: spatial and temporal mass variations in South America, *Inverse Problems,* **28**(6), doi:10.1088/0266-5611/28/6/065012.

Fischer, D. & Michel, V., 2013a. Automatic best-basis selection for geophysical tomographic inverse problems, *Geophys. J. Int.,* **193**(3), 1291–1299.

Fischer, D. & Michel, V., 2013b. Inverting GRACE gravity data for local climate effects, *J. Geod. Sci.,* **3**(3), 151–162.

Fischer, D., 2011. Sparse regularization of a joint inversion of gravitational data and normal mode anomalies, *PhD thesis*, University of Siegen, Geomathematics Group, Verlag Dr. Hut, Munich, http://dokumentix.ub.uni-siegen.de/opus/volltexte/2012/544/index.html.

Freeden, W. & Gutting, M., 2013. *Special Functions of Mathematical (Geo-)Physics,* Springer.

Freeden, W. & Schreiner, M., 2009. *Spherical Functions of Mathematical Geosciences – A Scalar, Vectorial, and Tensorial Setup,* Springer.

Freeden, W., Gervens, T. & Schreiner, M., 1998. *Constructive Approximation on the Sphere – with Applications to Geomathematics,* Oxford Univ. Press.

Fukushima, T., 2018. Accurate computation of gravitational field of a tesseroid, *J. Geod.,* **92**(12), 1371–1386.

Grossmann, C., Roos, H.-G. & Stynes, M., 2007. *Numerical Treatment of Partial Differential Equations,* Springer.

Gutting, M., Kretz, B., Michel, V. & Telschow, R., 2017. Study on parameter choice methods for the RFMP with respect to downward continuation, *Front. Appl. Math. Stat.,* **3,** doi:10.3389/fams.2017.00010.

Heuser, H., 2006. *Funktionalanalysis,* 4th edn, B. G. Teubner.

Hofmann, B., 1999. *Mathematik inverser Probleme,* B. G. Teubner.

Hosseini, K., Matthews, K.J., Sigloch, K., Shephard, G.E., Domeier, M. & Tsekhmistrenko, M., 2018. SubMachine: web-based tools for exploring seismic tomography and other models of Earth's deep interior, *Geochem. Geophys. Geosyst.,* **19**(5), 1464–1483.

Hosseini, K., Sigloch, K., Tsekhmistrenko, M., Zaheri, A., Nissen-Meyer, T. & Igel, H., 2020. Global mantle structure from multifrequency tomography using *P*, *PP* and *P*-diffracted waves, *Geophys. J. Int.,* **220**(1), 96–141.

Johnson, C., 2009. *Numerical Solution of Partial Differential Equations by the Finite Element Method,* Dover Publication, Dover ed., unabridged republ. of the work orig. publ. in 1987 by Cambridge Univ. Press.

Johnson, S.G., 2019. *The NLopt nonlinear-optimization package,* http://github.com/stevengj/nlopt and https://nlopt.readthedocs.io/en/latest/, both accessed 7 August 2023.

Kennett, B.L.N. & Engdahl, E.R., 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.,* **105**(2), 429–465.

Kirsch, A., 1996. *An Introduction to the Mathematical Theory of Inverse Problems,* Springer.

Kontak, M. & Michel, V., 2018. A greedy algorithm for nonlinear inverse problems with an application to nonlinear inverse gravimetry, *Int. J. Geomath.,* **9**(2), 167–198.

Kontak, M. & Michel, V., 2019. The regularized weak functional matching pursuit for linear inverse problems, *J. Inverse Ill-Posed Problems,* **27**(3), 317–340.

Kontak, M., 2018. Novel algorithms of greedy-type for probability density estimation as well as linear and nonlinear inverse problems, *PhD thesis*, University of Siegen, Geomathematics Group, Verlag Dr. Hut, Munich, http://dokumentix.ub.uni-siegen.de/opus/volltexte/2018/1316/index.html.

Leweke, S., 2018. The inverse magneto-electroencephalography problem for the spherical multiple-shell model: theoretical investigations and numerical aspects, *PhD thesis*, University of Siegen, Geomathematics Group, http://dokumentix.ub.uni-siegen.de/opus/volltexte/2019/1396/.

Louis, A.K., 1989. *Inverse und schlecht gestellte Probleme,* Teubner.

Magnus, W., Oberhettinger, F. & Soni, R., 1966. *Formulas and Theorems for the Special Functions of Mathematical Physics, Die Grundlehren der Mathematischen Wissenschaften 52,* 3rd edn, Springer-Verlag.

Marquering, H., Dahlen, F.A. & Nolet, G., 1999. Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.,* **137**(3), 805–815.

Marquering, H., Nolet, G. & Dahlen, F.A., 1998. Three-dimensional waveform sensitivity kernels, *Geophys. J. Int.,* **132**(3), 521–534.

Maździarz, M., 2010. Unified isoparametric 3D Lagrange finite elements, *Comp. Modell. Eng. Sci.,* **66**(1), 1–24.

Michel, V. & Orzlowski, S., 2016. On the null space of a class of Fredholm integral equations of the first kind, *J. Inverse Ill-Posed Problems,* **24**(6), 687–710.

Michel, V. & Orzlowski, S., 2017. On the convergence theorem for the Regularized Functional Matching Pursuit (RFMP) algorithm, *Int. J. Geomath.,* **8**(2), 183–190.

Michel, V. & Schneider, N., 2020. A first approach to learning a best basis for gravitational field modelling, *Int. J. Geomath.,* **11**(1), doi:10.1007/s13137-020-0143-5.

Michel, V. & Telschow, R., 2014. A non-linear approximation method on the sphere, *Int. J. Geomath.,* **5**(2), 195–224.

Michel, V. & Telschow, R., 2016. The regularized orthogonal functional matching pursuit for ill-posed inverse problems, *SIAM J. Numer. Anal.,* **54**(1), 262–287.

Michel, V., 1999. A multiscale method for the gravimetry problem: theoretical and numerical aspects of harmonic and anharmonic modelling, *PhD thesis*, University of Kaiserslautern, AG Geomathematik, Shaker Verlag, Aachen.

Michel, V., 2013. *Lectures on Constructive Approximation – Fourier, Spline, and Wavelet Methods on the Real Line, the Sphere, and the Ball,* Birkhäuser.
Michel, V., 2015. RFMP – an iterative best basis algorithm for inverse problems in the geosciences, in *Handbook of Geomathematics,* 2nd edn, pp. 2121–2147, eds Freeden, W., Nashed, M.Z. & Sonar, T., Springer.
Michel, V., 2022. *Geomathematics – Modelling and Solving Mathematical Problems in Geodesy and Geophysics,* Cambridge Univ. Press.
Mohammadzaheri, A., Sigloch, K., Hosseini, K. & Mihalynuk, M.G., 2021. Subducted lithosphere under South America from multifrequency P wave tomography, *J. Geophys. Res.,* **126**(6), e2020JB020704, doi:10.1029/2020JB020704.
Morse, P.M. & Feshbach, H., 1953a. *Methods of Theoretical Physics,* Vol. **1,** McGraw-Hill.
Morse, P.M. & Feshbach, H., 1953b. *Methods of Theoretical Physics,* Vol. **2,** McGraw-Hill.
Müller, C., 1966. *Spherical Harmonics,* Springer.
Nolet, G., 2008. *A Breviary of Seismic Tomography – Imaging the Interior of the Earth and Sun,* Cambridge Univ. Press.
Prakash, E., Radhamani, V. & Priyalatha, S.P.R., 2020. Determination of Earth's mass density distribution based on satellite data, *Adv. Math.: Sci. J.,* **9**(9), 7223–7233.
Rieder, A., 2003. *Keine Probleme mit Inversen Problemen. Eine Einführung in ihre stabile Lösung,* Vieweg.
Schneider, N. & Michel, V., 2022. A dictionary learning add-on for spherical downward continuation, *J. Geod.,* **96**(4), doi:10.1007/s00190-022-01598-w.
Schneider, N. 2023. (L)IPMP source code for travel time tomography, v3-tt-2023. *Zenodo*. https://doi.org/10.5281/zenodo.8227888.
Schneider, N., 2020. Learning dictionaries for inverse problems on the sphere, *PhD thesis*, University of Siegen, Geomathematics Group Siegen, doi:10.25819/ubsi/5431.
Schneider, N., Michel, V. & Sneeuw, N., 2023. High-dimensional experiments for the downward continuation using the LRFMP algorithm, preprint (arXiv:2308.04167).
Schwarz, H.R. & Köckler, N., 2011. *Numerische Mathematik,* 8th edn, Vieweg+Teubner.
Sigloch, K., 2008. Multiple-frequency body-wave tomography, *PhD thesis*, Princeton University, Princeton, NJ, USA.
Sigloch, K., 2011. Mantle provinces under North America from multifrequency P wave tomography, *Geochem. Geophys. Geosyst.,* **12,** doi:10.1029/2010GC003421.
Szegö, G., 1975. *Orthogonal Polynomials, American Mathematical Society Colloquium Publications Vol. XXIII,* American Mathematical Society.
Telschow, R., 2014. An orthogonal matching pursuit for the regularization of spherical inverse problems, *PhD thesis*, University of Siegen, Geomathematics Group, Verlag Dr. Hut, Munich.
Telschow, R., Gerhards, C. & Rother, M., 2018. On the approximation of spatial structures of global tidal magnetic field models, *Ann. Geophys.,* **36**(5), 1393–1402.
The International Seismological Centre (ISC), 2022. *ISC-EHB Bulletin,* http://www.isc.ac.uk/isc-ehb/, accessed 21 April 2022.
Tian, Y., Hung, S.-H., Nolet, G., Montelli, R. & Dahlen, F.A., 2007a. Dynamic ray tracing and traveltime corrections for global seismic tomography, *J. Comput. Phys.,* **226**(1), 672–687.
Tian, Y., Montelli, R., Nolet, G. & Dahlen, F.A., 2007b. Computing traveltime and amplitude sensitivity kernels in finite-frequency tomography, *J. Comput. Phys.,* **226**(2), 2271–2288.
Tsekhmistrenko, M., Sigloch, K., Hosseini, K. & Barruol, G., 2021. A tree of Indo-African mantle plumes imaged by seismic tomography, *Nat. Geosci.,* **14**(8), 612–619.
Werner, D., 2018. *Funktionalanalysis,* 8th edn, Springer.
Weston, J., Engdahl, E.R., Harris, J., Di Giacomo, D. & Storchak, D.A., 2018. ISC-EHB: reconstruction of a robust earthquake data set, *Geophys. J. Int.,* **214**(1), 474–484.
Yomogida, K., 1995. Fresnel zone inversion for lateral heterogeneities in the Earth, *Pure appl Geophys.,* **138**(3), 391–406.
Yosida, K., 1995. *Functional Analysis,* 6th edn, Springer, Reprint of the 1980 edition.

APPENDIX A: MATHEMATICAL DERIVATION OF SPECIFIC TERMS

A1 Derivation of $\nabla N_{A,\,\Delta A}$

As we are using the gradient in the $\mathrm{L}^2\left(\mathbb{B}, \mathbb{R}^3\right)$-integrals, we need the version of ∇ in spherical coordinates:

$$\nabla_{r\xi(\varphi,t)} = \varepsilon^r \frac{\partial}{\partial r} + \frac{1}{r}\nabla^* = \varepsilon^r \frac{\partial}{\partial r} + \frac{1}{r}\left(\varepsilon^\varphi \frac{1}{\sqrt{1-t^2}}\frac{\partial}{\partial \varphi} + \varepsilon^t \sqrt{1-t^2}\frac{\partial}{\partial t}\right) \tag{A1}$$

(see, e.g. Freeden *et al.* 1998; Freeden & Gutting 2013; Michel 2013). Then, for an FEHF, we obtain

$$\nabla_{r\xi(\varphi,t)} N_{(R,\Phi,T),(\Delta R,\Delta\Phi,\Delta T)}(r,\varphi,t) \tag{A2}$$

$$= \chi_{\mathrm{supp}[(R,\Phi,T)-(\Delta R,\Delta\Phi,\Delta T),(R,\Phi,T)+(\Delta R,\Delta\Phi,\Delta T)]}(r,\varphi,t)\left(\varepsilon^r \frac{\partial}{\partial r} N_{(R,\Phi,T),(\Delta R,\Delta\Phi,\Delta T)}(r,\varphi,t)\right. \tag{A3}$$

$$+ \frac{1}{r}\varepsilon^\varphi \frac{1}{\sqrt{1-t^2}}\frac{\partial}{\partial \varphi} N_{(R,\Phi,T),(\Delta R,\Delta\Phi,\Delta T)}(r,\varphi,t) \tag{A4}$$

$$\left. + \frac{1}{r}\varepsilon^t \sqrt{1-t^2}\frac{\partial}{\partial t} N_{(R,\Phi,T),(\Delta R,\Delta\Phi,\Delta T)}(r,\varphi,t)\right) \tag{A5}$$

$$= \chi_{\mathrm{supp}[(R,\Phi,T)-(\Delta R,\Delta\Phi,\Delta T),(R,\Phi,T)+(\Delta R,\Delta\Phi,\Delta T)]}(r,\varphi,t) \tag{A6}$$

$$\times \left(\varepsilon^r \frac{[-\operatorname{sgn}(r-R)]}{\Delta R}\frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi}\frac{\Delta T - |t-T|}{\Delta T}\right. \tag{A7}$$

$$+ \frac{1}{r}\varepsilon^\varphi \frac{1}{\sqrt{1-t^2}}\frac{\Delta R - |r-R|}{\Delta R}\frac{[-\operatorname{sgn}(\varphi-\Phi)]}{\Delta\Phi}\frac{\Delta T - |t-T|}{\Delta T} \tag{A8}$$

$$\left. + \frac{1}{r}\varepsilon^t \sqrt{1-t^2}\frac{\Delta R - |r-R|}{\Delta R}\frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi}\frac{[-\operatorname{sgn}(t-T)]}{\Delta T}\right) \tag{A9}$$

almost everywhere because

$$\frac{\partial}{\partial a_k}\frac{\Delta A_k - |a_k - A_k|}{\Delta A_k} = \frac{-\frac{\partial}{\partial a_k}|a_k - A_k|}{\Delta A_k} = \frac{-\operatorname{sgn}(a_k - A_k)}{\Delta A_k}. \tag{A10}$$

Note that this is only piecewise continuous with respect to the differentiated component but it is still continuous for the other components.

A2 Derivation of $\nabla G^{\mathrm{I}}_{m,n,j}$

Similarly, for the polynomials with $n \geq 1$, we obtain

$$\nabla_{r\xi(\varphi,t)} G^{\mathrm{I}}_{m,n,j}(r\xi(\varphi,t)) = p_{m,n}\left[\varepsilon^r \frac{\partial}{\partial r} + \frac{1}{r}\nabla^*\right]\left[P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^n Y_{n,j}(\xi(\varphi,t))\right] \tag{A11}$$

$$= p_{m,n}\varepsilon^r \frac{\partial}{\partial r}\left[P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^n\right] Y_{n,j}(\xi(\varphi,t)) + \frac{p_{m,n}}{r}\nabla^*\left[P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^n Y_{n,j}(\xi(\varphi,t))\right] \tag{A12}$$

$$= p_{m,n}\left[\left(P_m^{(0,n+1/2)}(I(r))\right)' I'(r)\left(\frac{r}{\mathcal{R}}\right)^n + \frac{n}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1}\right]\mu_n^{(1)} y_{n,j}^{(1)}(\xi(\varphi,t)) \tag{A13}$$

$$+ \frac{p_{m,n}}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1}\mu_n^{(2)} y_{n,j}^{(2)}(\xi(\varphi,t)) \tag{A14}$$

with

$$\mu_n^{(i)} := \begin{cases} 1, & i = 1 \\ \sqrt{n(n+1)}, & i = 2, 3, \end{cases} \tag{A15}$$

and the vector spherical harmonics $y_{n,j}^{(i)}$ (see e.g. Morse & Feshbach 1953a, b; Dahlen & Tromp 1998; Freeden & Schreiner 2009; Freeden & Gutting 2013; Michel 2022). In the case $n = 0$, the angular derivative as well as the derivative of $(r/R)^n$ does not exist:

$$\nabla_{r\xi(\varphi,t)} G^{\mathrm{I}}_{m,0,0}(r\xi(\varphi,t)) = p_{m,0}\left[\varepsilon^r \frac{\partial}{\partial r} + \frac{1}{r}\nabla^*\right] P_m^{(0,1/2)}(I(r))\, Y_{0,0}(\xi(\varphi,t)) \tag{A16}$$

$$= p_{m,0}\left[\varepsilon^r \frac{\partial}{\partial r} + \frac{1}{r}\nabla^*\right] P_m^{(0,1/2)}(I(r))\frac{1}{\sqrt{4\pi}} \tag{A17}$$

$$= p_{m,0}\left(P_m^{(0,1/2)}(I(r))\right)' I'(r)\left(\mu_0^{(1)}\right) y_{0,0}^{(1)}(\xi(\varphi,t)). \tag{A18}$$

Note that eq. (A18) can be written in the form eq. (A14) due to the multiplication with 0 (2nd term) and by defining $y_{0,0}^{(2)} := 0$ (3rd term). Hence, the gradient eq. (A14) is well-defined for all r and all possible m, n and j. For practical purposes, we need to specify the gradient in more detail:

$$\nabla_{r\xi(\varphi,t)} G^{\mathrm{I}}_{m,n,j}(r\xi(\varphi,t)) = p_{m,n}\left[\left(P_m^{(0,n+1/2)}(I(r))\right)' I'(r)\left(\frac{r}{\mathcal{R}}\right)^n + \frac{n}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1}\right] y_{n,j}^{(1)}(\xi(\varphi,t)) \tag{A19}$$

$$+ \frac{p_{m,n}}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1}\mu_n^{(2)} y_{n,j}^{(2)}(\xi(\varphi,t)) \tag{A20}$$

$$= p_{m,n}\left[\left(P_m^{(0,n+1/2)}(I(r))\right)' I'(r)\left(\frac{r}{\mathcal{R}}\right)^n q_{n,j} P_{n,|j|}(t)\mathrm{Trig}(j\varphi)\xi(\varphi,t)\right. \tag{A21}$$

$$+ \frac{n}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1} q_{n,j} P_{n,|j|}(t)\mathrm{Trig}(j\varphi)\xi(\varphi,t) \tag{A22}$$

$$\left. + \frac{1}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1} q_{n,j}\nabla^*\left(P_{n,|j|}(t)\mathrm{Trig}(j\varphi)\right)\right] \tag{A23}$$

$$= p_{m,n}q_{n,j}\left[\left(P_m^{(0,n+1/2)}(I(r))\right)' I'(r)\left(\frac{r}{\mathcal{R}}\right)^n P_{n,|j|}(t)\mathrm{Trig}(j\varphi)\xi(\varphi,t)\right. \tag{A24}$$

$$+ \frac{n}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1} P_{n,|j|}(t)\mathrm{Trig}(j\varphi)\xi(\varphi,t) \tag{A25}$$

$$+ \frac{1}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1}\frac{1}{\sqrt{1-t^2}} P_{n,|j|}(t)\left(\mathrm{Trig}(j\varphi)\right)' \varepsilon^{\varphi}(\varphi,t) \tag{A26}$$

$$\left. + \frac{1}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1}\sqrt{1-t^2} P'_{n,|j|}(t)\mathrm{Trig}(j\varphi)\varepsilon^t(\varphi,t)\right] \tag{A27}$$

$$= p_{m,n}q_{n,j}\left[\left(P_m^{(0,n+1/2)}(I(r))\right)' I'(r)\left(\frac{r}{\mathcal{R}}\right)^n P_{n,|j|}(t)\mathrm{Trig}(j\varphi)\xi(\varphi,t)\right. \tag{A28}$$

$$\left. + \frac{n}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1} P_{n,|j|}(t)\mathrm{Trig}(j\varphi)\xi(\varphi,t)\right] \tag{A29}$$

$$+\frac{j}{\mathcal{R}}P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1}\frac{1}{\sqrt{1-t^2}}P_{n,|j|}(t)\mathrm{Trig}(-j\varphi)\varepsilon^{\varphi}(\varphi,t) \tag{A30}$$

$$+\frac{1}{\mathcal{R}}P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^{n-1}\sqrt{1-t^2}P'_{n,|j|}(t)\mathrm{Trig}(j\varphi)\varepsilon^{t}(\varphi,t)\Bigg] \tag{A31}$$

$$=: p_{m,n}q_{n,j}\sum_{p=1}^{4}G^{\mathrm{I}}_{m,n,j;p}(r\xi(\varphi,t)). \tag{A32}$$

Note that the well-definedness of the terms

$$\frac{P_{n,|j|}(t)}{\sqrt{1-t^2}} \quad \text{and} \quad \sqrt{1-t^2}P'_{n,|j|}(t) \tag{A33}$$

used only for $n \geq 1$ was already discussed in Schneider (2020). The latter can also be computed with the respective algorithm given there. We discuss the former here in a bit more detail as this increases the efficiency in our implementation. Away from the poles, the term can be calculated straightforwardly. In a neighbourhood of the North and South Pole (i.e. for $t \to \pm 1$), we obtain

$$\lim_{t\to\pm1}\frac{P_{n,|j|}(t)}{\sqrt{1-t^2}} = \lim_{t\to\pm1}\left(1-t^2\right)^{(|j|-1)/2}P_n^{(|j|)}(t) = \begin{Bmatrix} \lim_{t\to\pm1}\left(1-t^2\right)^0 P'_n(t) = P'_n(\pm1), & |j|=1 \\ 0, & |j|>1, \end{Bmatrix} \tag{A34}$$

where the lower row holds because every derivative of a Legendre polynomial is bounded in $[-1, 1]$ for trivial reasons.

A3 Derivation of regularization terms

At first, we list certain trigonometric identities that we will need hereafter:

$$\sin(\alpha)+\sin(\beta) = 2\sin\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right), \quad \sin(\alpha)-\sin(\beta) = 2\cos\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right), \tag{A35}$$

$$\cos(\alpha)+\cos(\beta) = 2\cos\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right), \quad \cos(\alpha)-\cos(\beta) = -2\sin\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right), \tag{A36}$$

$$\sin(\alpha)\sin(\beta) = \frac{1}{2}\left(\cos\left(\alpha-\beta\right)-\cos\left(\alpha+\beta\right)\right), \quad \cos(\alpha)\cos(\beta) = \frac{1}{2}\left(\cos\left(\alpha-\beta\right)+\cos\left(\alpha+\beta\right)\right), \tag{A37}$$

$$\sin(\alpha)\cos(\beta) = \frac{1}{2}\left(\sin\left(\alpha-\beta\right)-\sin\left(\alpha+\beta\right)\right), \quad \int\sin(x)\cos(x)\mathrm{d}x = \frac{\sin^2(x)}{2}, \tag{A38}$$

$$\int x\sin(ax)\mathrm{d}x = \frac{\sin(ax)}{a^2}-\frac{x\cos(ax)}{a}, \qquad \int x\cos(ax)\mathrm{d}x = \frac{\cos(ax)}{a^2}+\frac{x\sin(ax)}{a}, \tag{A39}$$

$$\int\sin^2(x)\mathrm{d}x = \frac{x}{2}-\frac{\sin(2x)}{4}, \qquad \int\cos^2(x)\mathrm{d}x = \frac{x}{2}+\frac{\sin(2x)}{4}. \tag{A40}$$

We have to discuss the following inner products

$$\left\langle G^{\mathrm{I}}_{m,n,j}, G^{\mathrm{I}}_{m',n',j'}\right\rangle_{\mathcal{H}^1}, \qquad \left\langle N_{A,\Delta A}, N_{A',(\Delta A)'}\right\rangle_{\mathcal{H}^1}, \qquad \left\langle N_{A,\Delta A}, G^{\mathrm{I}}_{m,n,j}\right\rangle_{\mathcal{H}^1}, \tag{A41}$$

which practically means computing

$$\left\langle G^{\mathrm{I}}_{m,n,j}, G^{\mathrm{I}}_{m',n',j'}\right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R})}, \qquad \left\langle N_{A,\Delta A}, N_{A',(\Delta A)'}\right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R})}, \qquad \left\langle N_{A,\Delta A}, G^{\mathrm{I}}_{m,n,j}\right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R})}, \tag{A42}$$

$$\left\langle \nabla G^{\mathrm{I}}_{m,n,j}, \nabla G^{\mathrm{I}}_{m',n',j'}\right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R}^3)}, \quad \left\langle \nabla N_{A,\Delta A}, \nabla N_{A',(\Delta A)'}\right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R}^3)}, \quad \left\langle \nabla N_{A,\Delta A}, \nabla G^{\mathrm{I}}_{m,n,j}\right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R}^3)}. \tag{A43}$$

We start with two polynomials. We have

$$\left\langle G^{\mathrm{I}}_{m,n,j}, G^{\mathrm{I}}_{m',n',j'}\right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R})} = \delta_{m,m'}\delta_{n,n'}\delta_{j,j'} \tag{A44}$$

due to the fact that they constitute an orthonormal basis system in $\mathrm{L}^2(\mathbb{B}, \mathbb{R})$. Next, we consider the vectorial inner product of their gradients. Note that the vector spherical harmonics are orthonormal with respect to their degrees, order and type. With eqs (44), (A15) and the substitution $r = \mathcal{R}\sqrt{(1+u)/2}$, we obtain from eq. (A20)

$$\left\langle \nabla G^{\mathrm{I}}_{m,n,j}, \nabla G^{\mathrm{I}}_{m',n',j'} \right\rangle_{\mathrm{L}^2\left(\mathbb{B},\mathbb{R}^3\right)} = \delta_{n,n'}\delta_{j,j'} p_{m,n} p_{m',n} \tag{A45}$$

$$\times \left[\int_0^{\mathcal{R}} \left[\left(P_m^{(0,n+1/2)}(I(r))\right)' I'(r) \left(\frac{r}{\mathcal{R}}\right)^n + \frac{n}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{n-1} \right] \right. \tag{A46}$$

$$\left. \times \left[\left(P_{m'}^{(0,n+1/2)}(I(r))\right)' I'(r) \left(\frac{r}{\mathcal{R}}\right)^n + \frac{n}{\mathcal{R}} P_{m'}^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{n-1} \right] r^2 \mathrm{d}r \right. \tag{A47}$$

$$\left. + \left(\mu_n^{(2)}\right)^2 \int_0^{\mathcal{R}} \left[\frac{1}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{n-1} \right] \left[\frac{1}{\mathcal{R}} P_{m'}^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{n-1} \right] r^2 \mathrm{d}r \right] \tag{A48}$$

$$= \delta_{n,n'}\delta_{j,j'} p_{m,n} p_{m',n} \tag{A49}$$

$$\times \left[\left(\mu_n^{(1)}\right)^2 \int_0^{\mathcal{R}} \left(P_m^{(0,n+1/2)}(I(r))\right)' I'(r) \left(\frac{r}{\mathcal{R}}\right)^n \left(P_{m'}^{(0,n+1/2)}(I(r))\right)' I'(r) \left(\frac{r}{\mathcal{R}}\right)^n r^2 \mathrm{d}r \right. \tag{A50}$$

$$+ \left(\mu_n^{(1)}\right)^2 \int_0^{\mathcal{R}} \left(P_m^{(0,n+1/2)}(I(r))\right)' I'(r) \left(\frac{r}{\mathcal{R}}\right)^n \frac{n}{\mathcal{R}} P_{m'}^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{n-1} r^2 \mathrm{d}r \tag{A51}$$

$$+ \left(\mu_n^{(1)}\right)^2 \int_0^{\mathcal{R}} \frac{n}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{n-1} \left(P_{m'}^{(0,n+1/2)}(I(r))\right)' I'(r) \left(\frac{r}{\mathcal{R}}\right)^n r^2 \mathrm{d}r \tag{A52}$$

$$+ \left(\mu_n^{(1)}\right)^2 \int_0^{\mathcal{R}} \frac{n}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{n-1} \frac{n}{\mathcal{R}} P_{m'}^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{n-1} r^2 \mathrm{d}r \tag{A53}$$

$$\left. + \left(\mu_n^{(2)}\right)^2 \int_0^{\mathcal{R}} \frac{1}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{n-1} \frac{1}{\mathcal{R}} P_{m'}^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{n-1} r^2 \mathrm{d}r \right] \tag{A54}$$

$$= \delta_{n,n'}\delta_{j,j'} p_{m,n} p_{m',n} \tag{A55}$$

$$\times \left[\int_0^{\mathcal{R}} \left(P_m^{(0,n+1/2)}(I(r))\right)' \left(P_{m'}^{(0,n+1/2)}(I(r))\right)' \left(I'(r)\right)^2 \left(\frac{r}{\mathcal{R}}\right)^{2n} r^2 \mathrm{d}r \right. \tag{A56}$$

$$+ \frac{n}{\mathcal{R}} \int_0^{\mathcal{R}} \left[\left(P_m^{(0,n+1/2)}(I(r))\right)' P_{m'}^{(0,n+1/2)}(I(r)) + P_m^{(0,n+1/2)}(I(r)) \left(P_{m'}^{(0,n+1/2)}(I(r))\right)' \right] I'(r) \left(\frac{r}{\mathcal{R}}\right)^{2n-1} r^2 \mathrm{d}r \tag{A57}$$

$$\left. + \left(n^2 + n(n+1)\right) \int_0^{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) P_{m'}^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{2n} \mathrm{d}r \right] \tag{A58}$$

$$= \delta_{n,n'}\delta_{j,j'} p_{m,n} p_{m',n} \tag{A59}$$

$$\times \left[\int_0^{\mathcal{R}} \left(P_m^{(0,n+1/2)}(I(r))\right)' \left(P_{m'}^{(0,n+1/2)}(I(r))\right)' \frac{16r^2}{\mathcal{R}^4} \left(\frac{r}{\mathcal{R}}\right)^{2n} r^2 \mathrm{d}r \right. \tag{A60}$$

$$+ \frac{n}{\mathcal{R}} \int_0^{\mathcal{R}} \left[\left(P_m^{(0,n+1/2)}(I(r))\right)' P_{m'}^{(0,n+1/2)}(I(r)) + P_m^{(0,n+1/2)}(I(r)) \left(P_{m'}^{(0,n+1/2)}(I(r))\right)' \right] \frac{4r}{\mathcal{R}^2} \left(\frac{r}{\mathcal{R}}\right)^{2n-1} r^2 \mathrm{d}r \tag{A61}$$

$$\left. + \left(n^2 + n^2 + n\right) \int_0^{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) P_{m'}^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{2n} \mathrm{d}r \right] \tag{A62}$$

$$= \delta_{n,n'}\delta_{j,j'} p_{m,n} p_{m',n} \tag{A63}$$

$$\times \left[16 \int_0^{\mathcal{R}} \left(P_m^{(0,n+1/2)}(I(r))\right)' \left(P_{m'}^{(0,n+1/2)}(I(r))\right)' \left(\frac{r}{\mathcal{R}}\right)^{2n+4} \mathrm{d}r \right. \tag{A64}$$

$$+ 4n \int_0^{\mathcal{R}} \left[\left(P_m^{(0,n+1/2)}(I(r))\right)' P_{m'}^{(0,n+1/2)}(I(r)) + P_m^{(0,n+1/2)}(I(r)) \left(P_{m'}^{(0,n+1/2)}(I(r))\right)' \right] \left(\frac{r}{\mathcal{R}}\right)^{2n+2} \mathrm{d}r \tag{A65}$$

$$\left. + \left(2n^2 + n\right) \int_0^{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) P_{m'}^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}}\right)^{2n} \mathrm{d}r \right] \tag{A66}$$

$$= \delta_{n,n'}\delta_{j,j'} p_{m,n} p_{m',n} \tag{A67}$$

$$\times \left[16 \int_{-1}^{1} \left(P_m^{(0,n+1/2)}(u)\right)' \left(P_{m'}^{(0,n+1/2)}(u)\right)' \left(\frac{1+u}{2}\right)^{n+2} \frac{\mathcal{R}}{4} \sqrt{\frac{2}{1+u}} \mathrm{d}u \right. \tag{A68}$$

$$\left. + 4n \int_{-1}^{1} \left[\left(P_m^{(0,n+1/2)}(u)\right)' P_{m'}^{(0,n+1/2)}(u) + P_m^{(0,n+1/2)}(u) \left(P_{m'}^{(0,n+1/2)}(u)\right)' \right] \left(\frac{1+u}{2}\right)^{n+1} \frac{\mathcal{R}}{4} \sqrt{\frac{2}{1+u}} \mathrm{d}u \right. \tag{A69}$$

$$+ (n(2n+1)) \int_{-1}^{1} P_m^{(0,n+1/2)}(u) P_{m'}^{(0,n+1/2)}(u) \left(\frac{1+u}{2}\right)^n \frac{\mathcal{R}}{4} \sqrt{\frac{2}{1+u}} \mathrm{d}u \Bigg] \tag{A70}$$

$$= \delta_{n,n'} \delta_{j,j'} p_{m,n} p_{m',n} \tag{A71}$$

$$\times \Bigg[4\mathcal{R} \int_{-1}^{1} \left(P_m^{(0,n+1/2)}(u)\right)' \left(P_{m'}^{(0,n+1/2)}(u)\right)' \left(\frac{1+u}{2}\right)^{n+3/2} \mathrm{d}u \tag{A72}$$

$$+ \mathcal{R}n \int_{-1}^{1} \left[\left(P_m^{(0,n+1/2)}(u)\right)' P_{m'}^{(0,n+1/2)}(u) + P_m^{(0,n+1/2)}(u) \left(P_{m'}^{(0,n+1/2)}(u)\right)' \right] \left(\frac{1+u}{2}\right)^{n+1/2} \mathrm{d}u \tag{A73}$$

$$+ \frac{\mathcal{R}n(2n+1)}{4} \int_{-1}^{1} P_m^{(0,n+1/2)}(u) P_{m'}^{(0,n+1/2)}(u) \left(\frac{1+u}{2}\right)^{n-1/2} \mathrm{d}u \Bigg] \tag{A74}$$

$$= \delta_{n,n'} \delta_{j,j'} p_{m,n} p_{m',n} \tag{A75}$$

$$\times \Bigg[\frac{\mathcal{R}\sqrt{2}}{2^n} \int_{-1}^{1} \left(P_m^{(0,n+1/2)}(u)\right)' \left(P_{m'}^{(0,n+1/2)}(u)\right)' (1+u)^{n+3/2} \mathrm{d}u \tag{A76}$$

$$+ \frac{\mathcal{R}n}{2^n\sqrt{2}} \int_{-1}^{1} \left[\left(P_m^{(0,n+1/2)}(u)\right)' P_{m'}^{(0,n+1/2)}(u) + P_m^{(0,n+1/2)}(u) \left(P_{m'}^{(0,n+1/2)}(u)\right)' \right] (1+u)^{n+1/2} \mathrm{d}u \tag{A77}$$

$$+ \frac{\mathcal{R}n(2n+1)}{2^{n+1}\sqrt{2}} \int_{-1}^{1} P_m^{(0,n+1/2)}(u) P_{m'}^{(0,n+1/2)}(u) (1+u)^{n-1/2} \mathrm{d}u \Bigg]. \tag{A78}$$

Next, we discuss the inner products of two finite element hat functions. Generally, we have

$$\left\langle N_{A,\Delta A}, N_{A',(\Delta A)'} \right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R})} \tag{A79}$$

$$= \int_{\mathbb{B}} \chi_{\mathrm{supp}_{A,\Delta A}}(a) \chi_{\mathrm{supp}_{A',(\Delta A)'}}(a) \prod_{j=1}^{3} \frac{[\Delta A_j - |a_j - A_j|][(\Delta A_j)' - |a_j - A'_j|]}{\Delta A_j (\Delta A_j)'} \mathrm{d}a \tag{A80}$$

$$= \prod_{j=1}^{3} \int_{lb_{a_j}}^{ub_{a_j}} \frac{[\Delta A_j - |a_j - A_j|][(\Delta A_j)' - |a_j - A'_j|]}{\Delta A_j (\Delta A_j)'} \begin{Bmatrix} a_1^2, \, j = 1, \\ 1, \;\; j = 2, 3 \end{Bmatrix} \mathrm{d}a_j \tag{A81}$$

where lb_a denotes the lower and ub_a the upper bound of the respective a_j-integral. That is, principally, we have (informally speaking)

$$[lb, ub] = \mathrm{supp}_{A,\Delta A} \cap \mathrm{supp}_{A',(\Delta A)'} \tag{A82}$$

though we have to take a detailed look on the case a_2 if $P \neq P'$ (see below). As a generalization, we consider the integrals

$$\int_{lb_x}^{ub_x} \frac{[\Delta X - |x - X|][(\Delta X)' - |x - X'|]}{\Delta X (\Delta X)'} x^q \mathrm{d}x \tag{A83}$$

for $q \in \mathbb{N}_0$ (here, in particular, $q = 0$ and $q = 2$) in the sequel. Then, the $\mathrm{L}^2(\mathbb{B}, \mathbb{R})$-integral is obtained as the product of this integral for all variables $x = a_j, j = 1, 2, 3$. At first, we need to consider how the lower and upper bounds are defined. Furthermore, because the FEHFs are piecewise defined, we have to determine these critical points in between lb_x and ub_x as well. At last, we need to discuss the remaining integrals.

For $x = a_1 = r$ and $x = a_3 = t$, there are three cases depending on $X, \Delta X, X'$ and $(\Delta X)'$ that we have to consider:

(a) $\mathrm{supp}_{X,\Delta X} = [X - \Delta X, X + \Delta X]$

1. $\mathrm{supp}_{X',(\Delta X)'} = [X' - (\Delta X)', X + (\Delta X)']$
2. $\mathrm{supp}_{X',(\Delta X)'} = [X_{\min}, X' + (\Delta X)']$
3. $\mathrm{supp}_{X',(\Delta X)'} = [X' - (\Delta X)', X_{\max}]$

(b) $\mathrm{supp}_{X,\Delta X} = [X_{\min}, X + \Delta X]$

1. $\mathrm{supp}_{X',(\Delta X)'} = [X' - (\Delta X)', X' + (\Delta X)']$
2. $\mathrm{supp}_{X',(\Delta X)'} = [X_{\min}, X' + (\Delta X)']$
3. $\mathrm{supp}_{X',(\Delta X)'} = [X' - (\Delta X)', X_{\max}]$

(c) $\mathrm{supp}_{X,\Delta X} = [X - \Delta X, X_{\max}]$

1. $\mathrm{supp}_{X',(\Delta X)'} = [X' - (\Delta X)', X' + (\Delta X)']$
2. $\mathrm{supp}_{X',(\Delta X)'} = [X_{\min}, X' + (\Delta X)']$
3. $\mathrm{supp}_{X',(\Delta X)'} = [X' - (\Delta X)', X_{\max}]$

In any of these cases, we obtain the lower bound as

$$lb_x = \max[\max(X_{\min}, X - \Delta X), \max(X_{\min}, X' - (\Delta X)')] \tag{A84}$$

and the upper bound as

$$ub_x = \min[\min(X_{\max}, X + \Delta X), \min(X_{\max}, X' + (\Delta X)')] \tag{A85}$$

Note that if $lb_x \geq ub_x$ (componentwise), the intersection is the empty set and the integral vanishes. For $x = a_2 = \varphi$, we also have three relevant cases:

(a) $\operatorname{sgn}(P) = -1 \Rightarrow \operatorname{supp}_{X,\Delta X} = [\max(-\pi, X - \Delta X), \min(\pi, X + \Delta X)]$

1. $\operatorname{sgn}(P') = -1 \Rightarrow \operatorname{supp}_{X',(\Delta X)'} = [\max(-\pi, X' - (\Delta X)'), \min(\pi, X' + (\Delta X)')]$
2. $\operatorname{sgn}(P') = 0 \Rightarrow \operatorname{supp}_{X',(\Delta X)'} = [\max(0, X' - (\Delta X)'), \min(2\pi, X' + (\Delta X)')]$
3. $\operatorname{sgn}(P') = 1 \Rightarrow \operatorname{supp}_{X',(\Delta X)'} = [\max(\pi, X' - (\Delta X)'), \min(3\pi, X' + (\Delta X)')]$

(b) $\operatorname{sgn}(P) = 0 \Rightarrow \operatorname{supp}_{X,\Delta X} = [\max(0, X - \Delta X), \min(2\pi, X + \Delta X)]$

1. $\operatorname{sgn}(P') = -1 \Rightarrow \operatorname{supp}_{X',(\Delta X)'} = [\max(-\pi, X' - (\Delta X)'), \min(\pi, X' + (\Delta X)')]$
2. $\operatorname{sgn}(P') = 0 \Rightarrow \operatorname{supp}_{X',(\Delta X)'} = [\max(0, X' - (\Delta X)'), \min(2\pi, X' + (\Delta X)')]$
3. $\operatorname{sgn}(P') = 1 \Rightarrow \operatorname{supp}_{X',(\Delta X)'} = [\max(\pi, X' - (\Delta X)'), \min(3\pi, X' + (\Delta X)')]$

(c) $\operatorname{sgn}(P) = 1 \Rightarrow \operatorname{supp}_{X,\Delta X} = [\max(\pi, X - \Delta X), \min(3\pi, X + \Delta X)]$

1. $\operatorname{sgn}(P') = -1 \Rightarrow \operatorname{supp}_{X',(\Delta X)'} = [\max(-\pi, X' - (\Delta X)'), \min(\pi, X' + (\Delta X)')]$
2. $\operatorname{sgn}(P') = 0 \Rightarrow \operatorname{supp}_{X',(\Delta X)'} = [\max(0, X' - (\Delta X)'), \min(2\pi, X' + (\Delta X)')]$
3. $\operatorname{sgn}(P') = 1 \Rightarrow \operatorname{supp}_{X',(\Delta X)'} = [\max(\pi, X' - (\Delta X)'), \min(3\pi, X' + (\Delta X)')]$

If $\operatorname{sgn}(P) = \operatorname{sgn}(P')$ [i.e. the cases $(-1, -1), (0, 0)$ and $(1,1)$], we obtain the same lower and upper bound as in eqs (A84) and (A85), respectively, but with the respective $X_{\min} \in \{-\pi, 0, \pi\}$. If $\operatorname{sgn}(P) = -\operatorname{sgn}(P')$ while $\operatorname{sgn}(P) \neq \operatorname{sgn}(P')$ [i.e. the cases $(1, -1)$ and $(-1, 1)$], we shift one of them about 2π and obtain that $\operatorname{sgn}(P) = \operatorname{sgn}(P')$ (and $P = P'$). Last but not least, we have the cases $(0, 1), (1, 0), (0, -1)$ and $(-1, 0)$, that is where one, let it be $\operatorname{sgn}(P)$, equals zero and the other one, here $\operatorname{sgn}(P')$, is ± 1. If $\operatorname{sgn}(P') = -1$, then we can cut the support $\operatorname{supp}_{X',(\Delta X)'} \subset [-\pi, \pi]$ from the left-hand side at 0 and shift the part that is in $[-\pi, 0]$ into $[\pi, 2\pi]$. Then, we obtain $\operatorname{supp}^{\text{shifted}}_{X',(\Delta X)'} = [X' - (\Delta X)' + 2\pi, \min(X' + (\Delta X)' + 2\pi, 2\pi)] \cup [\max(0, X' - (\Delta X)'), X' + (\Delta X)']$. As a consequence, we obtain possibly two lower and upper bounds via

$$[X' - (\Delta X)' + 2\pi, \min(X' + (\Delta X)' + 2\pi, 2\pi)] \cap \operatorname{supp}_{X,\Delta X} \tag{A86}$$

and

$$[\max(0, X' - (\Delta X)'), X' + (\Delta X)'] \cap \operatorname{supp}_{X,\Delta X}. \tag{A87}$$

In particular, we have

$$lb_x = \begin{pmatrix} \max[\max(0, X - \Delta X), X' - (\Delta X)' + 2\pi] \\ \max[\max(0, X - \Delta X), \max(0, X' - (\Delta X)')] \end{pmatrix} \tag{A88}$$

and the upper bound as

$$ub_x = \begin{pmatrix} \min[\min(2\pi, X + \Delta X), \min(X' + (\Delta X)' + 2\pi, 2\pi)] \\ \min[\min(2\pi, X + \Delta X), X' + (\Delta X)'] \end{pmatrix}. \tag{A89}$$

If $\operatorname{sgn}(P') = 1$, we cut its domain at 2π and shift the part that is in $[2\pi, 3\pi]$ into $[0, \pi]$. Thus, analogously, we obtain $\operatorname{supp}^{\text{shifted}}_{X',(\Delta X)'} = [X' - (\Delta X)', \min(2\pi, X' + (\Delta X)')] \cup [\max(0, X' - (\Delta X)' - 2\pi), X' + (\Delta X)' - 2\pi]$ and

$$lb_x = \begin{pmatrix} \max[\max(0, X - \Delta X), X' - (\Delta X)'] \\ \max[\max(0, X - \Delta X), \max(0, X' - (\Delta X)' - 2\pi)] \end{pmatrix} \tag{A90}$$

and the upper bound as

$$ub_x = \begin{pmatrix} \min[\min(2\pi, X + \Delta X), \min(2\pi, X' + (\Delta X)')] \\ \min[\min(2\pi, X + \Delta X), X' + (\Delta X)' - 2\pi] \end{pmatrix}. \tag{A91}$$

For the cases where $\operatorname{sgn}(P') = 0$ and $\operatorname{sgn}(P) = \pm 1$, we obtain analogous solutions (exchange X with X', ΔX with $(\Delta X)'$ and vice versa). Note again that, in all cases, if $lb_x \geq ub_x$ (componentwise), the intersection is the empty set and the integral vanishes.

We explain the following steps of determining critical points and deriving the remaining integrals only for the case where we have one integration interval $[lb_x, ub_x]$. If we have two, we can execute these steps for both separately and add the integral values to obtain the value of (A83) with respect to $x = a_2 = \varphi$.

Critical points here are those points between the lower and upper bound(s) where one of the FEHF turn from increase to decrease of the hat (always with respect to a fixed dimension). Here is how to determine them. The critical points can obviously only be from $\{X, X \pm \Delta X, X', X' \pm (\Delta X)'\}$. Thus, we sort them for increasing order and then check each value whether it is in $[lb_x, ub_x]$. For practical purposes, it is sensible to count how many critical points—including lb_x and ub_x—we have. Let this count be $I \leq 6$. In the sequel, we set p_i for a critical point: $lb_x \leq p_i \leq ub_x, i = 1,..., I, I \leq 6$. Note that it cannot be more than 6 different critical points because of the definition of lb and ub. The integrals in eq. (A83) are, thus, equal to

$$\sum_{i=1}^{I-1} \int_{p_i}^{p_{i+1}} \frac{[\Delta X - |x - X|][(\Delta X)' - |x - X'|]}{\Delta X (\Delta X)'} x^q \mathrm{d}x. \tag{A92}$$

Note, that if two critical points p_i and p_{i+1} coincide, then the respective integral from p_i to p_{i+1} vanishes. Thus, we do not have to take care of this situation by ourselves. We first consider the following

$$\sum_{i=1}^{I-1} \int_{p_i}^{p_{i+1}} \frac{[\Delta X - |x - X|][(\Delta X)' - |x - X'|]}{\Delta X (\Delta X)'} x^q \mathrm{d}x \tag{A93}$$

$$= \sum_{i=1}^{I-1} \int_{p_i}^{p_{i+1}} \frac{[\Delta X - \mathrm{sgn}(x - X)(x - X)][(\Delta X)' - \mathrm{sgn}(x - X')(x - X')]}{\Delta X (\Delta X)'} x^q \mathrm{d}x \tag{A94}$$

$$= \sum_{i=1}^{I-1} \int_{p_i}^{p_{i+1}} \frac{\left[\Delta X \begin{bmatrix} + \\ - \\ - \\ + \end{bmatrix} (x - X)\right] \left[(\Delta X)' \begin{bmatrix} + \\ - \\ + \\ - \end{bmatrix} (x - X')\right]}{\Delta X (\Delta X)'} x^q \mathrm{d}x, \tag{A95}$$

where the last two cases must coincide. Note that we need to determine the values of $-\mathrm{sgn}(x - X)$ and $-\mathrm{sgn}(x - X')$ for this. For $x = a_1 = r$ and $x = a_3 = t$, the sign value can be obtained straightforwardly. Similarly, this holds for $x = a_2 = \varphi$, if $(P, P') \notin \{(0, 1), (1, 0), (0, -1), (-1, 0)\}$, that is, if we exclude the cases where we shifted the support of one FEHF in order to obtain the upper and lower bound. In the latter cases, we have to shift P or P' here accordingly to the shift done before to obtain the correct sign values. In practice, we can also shift P or P' once in the beginning of the computation of the inner product as well.

For readability, we only consider the integration of $x^q W(x; X, \Delta X, X', (\Delta X)')$ defined in

$$\int \frac{\left[\Delta X \begin{bmatrix} + \\ - \\ - \end{bmatrix} (x - X)\right] \left[(\Delta X)' \begin{bmatrix} + \\ - \\ + \end{bmatrix} (x - X')\right]}{\Delta X (\Delta X)'} x^q \mathrm{d}x =: \int x^q \frac{W(x; X, \Delta X, X', (\Delta X)')}{\Delta X (\Delta X)'} \mathrm{d}x \tag{A96}$$

$$= \frac{1}{\Delta X (\Delta X)'} \int x^q W\left(x; X, \Delta X, X', (\Delta X)'\right) \mathrm{d}x \tag{A97}$$

in the sequel. We obtain

$$\int x^q W\left(x; X, \Delta X, X', (\Delta X)'\right) \mathrm{d}x \tag{A98}$$

$$= \int \begin{bmatrix} + \\ + \\ - \end{bmatrix} x^{q+2} \begin{bmatrix} + \\ - \\ + \end{bmatrix} x^{q+1} \left(\begin{bmatrix} - \\ + \\ + \end{bmatrix} X \begin{bmatrix} - \\ + \\ + \end{bmatrix} X' \begin{bmatrix} + \\ + \\ + \end{bmatrix} \Delta X \begin{bmatrix} + \\ + \\ - \end{bmatrix} (\Delta X)' \right) \tag{A99}$$

$$\begin{bmatrix} + \\ + \\ - \end{bmatrix} x^q \left(X \begin{bmatrix} - \\ + \\ + \end{bmatrix} \Delta X \right) \left(X' \begin{bmatrix} - \\ + \\ - \end{bmatrix} (\Delta X') \right) \mathrm{d}x \tag{A100}$$

$$= \begin{bmatrix} + \\ + \\ - \end{bmatrix} \frac{x^{q+3}}{q+3} \begin{bmatrix} + \\ - \\ + \end{bmatrix} \frac{x^{q+2}}{q+2} \left(\begin{bmatrix} - \\ + \\ + \end{bmatrix} X \begin{bmatrix} - \\ + \\ + \end{bmatrix} X' \begin{bmatrix} + \\ + \\ + \end{bmatrix} \Delta X \begin{bmatrix} + \\ + \\ - \end{bmatrix} (\Delta X)' \right) \tag{A101}$$

$$\begin{bmatrix} + \\ + \\ - \end{bmatrix} \frac{x^{q+1}}{q+1} \left(X \begin{bmatrix} - \\ + \\ + \end{bmatrix} \Delta X \right) \left(X' \begin{bmatrix} - \\ + \\ - \end{bmatrix} (\Delta X') \right) \tag{A102}$$

$$= \left(\begin{bmatrix} + \\ + \\ - \end{bmatrix} (q+2)(q+1) x^{q+3} \begin{bmatrix} + \\ - \\ + \end{bmatrix} (q+3)(q+1) x^{q+2} \left(\begin{bmatrix} - \\ + \\ + \end{bmatrix} X \begin{bmatrix} - \\ + \\ + \end{bmatrix} X' \begin{bmatrix} + \\ + \\ + \end{bmatrix} \Delta X \begin{bmatrix} + \\ + \\ - \end{bmatrix} (\Delta X)' \right) \right. \tag{A103}$$

$$\begin{bmatrix}+\\+\\-\end{bmatrix}(q+3)(q+2)x^{q+1}\left(X\begin{bmatrix}-\\+\\+\end{bmatrix}\Delta X\right)\left(X'\begin{bmatrix}-\\+\\-\end{bmatrix}(\Delta X')\right)\frac{1}{(q+3)(q+2)(q+1)} \tag{A104}$$

$$=\left(\begin{bmatrix}+\\+\\-\end{bmatrix}(q+2)(q+1)x^2\begin{bmatrix}+\\-\\+\end{bmatrix}(q+3)(q+1)x\left(\begin{bmatrix}-\\+\\+\end{bmatrix}X\begin{bmatrix}-\\+\\+\end{bmatrix}X'\begin{bmatrix}+\\+\\+\end{bmatrix}\Delta X\begin{bmatrix}+\\+\\-\end{bmatrix}(\Delta X)'\right)\right. \tag{A105}$$

$$\left.\begin{bmatrix}+\\+\\-\end{bmatrix}(q+3)(q+2)\left(X\begin{bmatrix}-\\+\\+\end{bmatrix}\Delta X\right)\left(X'\begin{bmatrix}-\\+\\-\end{bmatrix}(\Delta X')\right)\right)\frac{x^{q+1}}{(q+3)(q+2)(q+1)}. \tag{A106}$$

With these values the $\mathrm{L}^2(\mathbb{B},\mathbb{R})$ inner product of two FEHFs is fully discussed. For the respective $\mathrm{L}^2\left(\mathbb{B},\mathbb{R}^3\right)$ inner product, we obtain

$$\left\langle\nabla_{r\xi(\varphi,t)}N_{(R,\Phi,T),(\Delta R,\Delta\Phi,\Delta T)},\nabla_{r\xi(\varphi,t)}N_{(R',\Phi',T'),((\Delta R)',(\Delta\Phi)',(\Delta T)')}\right\rangle_{\mathrm{L}^2\left(\mathbb{B},\mathbb{R}^3\right)} \tag{A107}$$

$$=\int_{\mathbb{B}}\chi_{\mathrm{supp}[(R,\Phi,T)-(\Delta R,\Delta\Phi,\Delta T),(R,\Phi,T)+(\Delta R,\Delta\Phi,\Delta T)]}(r,\varphi,t)\chi_{\mathrm{supp}[(R',\Phi',T')-((\Delta R)',(\Delta\Phi)',(\Delta T)'),(R',\Phi',T')+((\Delta R)',(\Delta\Phi)',(\Delta T)')]}(r,\varphi,t) \tag{A108}$$

$$\times\left(\frac{[-\operatorname{sgn}(r-R)]}{\Delta R}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\frac{\Delta T-|t-T|}{\Delta T}\frac{[-\operatorname{sgn}(r-R')]}{(\Delta R)'}\frac{(\Delta\Phi)'-|\varphi-\Phi'|}{(\Delta\Phi)'}\frac{(\Delta T)'-|t-T'|}{(\Delta T)'}\right. \tag{A109}$$

$$+\frac{1}{r^2}\frac{1}{1-t^2}\frac{\Delta R-|r-R|}{\Delta R}\frac{[-\operatorname{sgn}(\varphi-\Phi)]}{\Delta\Phi}\frac{\Delta T-|t-T|}{\Delta T}\frac{(\Delta R)'-|r-R'|}{(\Delta R)'}\frac{[-\operatorname{sgn}(\varphi-\Phi')]}{(\Delta\Phi)'}\frac{(\Delta T)'-|t-T'|}{(\Delta T)'} \tag{A110}$$

$$\left.+\frac{1}{r^2}(1-t^2)\frac{\Delta R-|r-R|}{\Delta R}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\frac{[-\operatorname{sgn}(t-T)]}{\Delta T}\frac{(\Delta R)'-|r-R'|}{(\Delta R)'}\frac{(\Delta\Phi)'-|\varphi-\Phi'|}{(\Delta\Phi)'}\frac{[-\operatorname{sgn}(t-T')]}{(\Delta T)'}\right)\mathrm{d}x(r,\varphi,t) \tag{A111}$$

$$=\int_{\mathbb{B}}\chi_{\mathrm{supp}[(R,\Phi,T)-(\Delta R,\Delta\Phi,\Delta T),(R,\Phi,T)+(\Delta R,\Delta\Phi,\Delta T)]}(r,\varphi,t)\chi_{\mathrm{supp}[(R',\Phi',T')-((\Delta R)',(\Delta\Phi)',(\Delta T)'),(R',\Phi',T')+((\Delta R)',(\Delta\Phi)',(\Delta T)')]}(r,\varphi,t) \tag{A112}$$

$$\times\frac{[-\operatorname{sgn}(r-R)]}{\Delta R}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\frac{\Delta T-|t-T|}{\Delta T}\frac{[-\operatorname{sgn}(r-R')]}{(\Delta R)'}\frac{(\Delta\Phi)'-|\varphi-\Phi'|}{(\Delta\Phi)'}\frac{(\Delta T)'-|t-T'|}{(\Delta T)'}\mathrm{d}x(r,\varphi,t) \tag{A113}$$

$$+\int_{\mathbb{B}}\chi_{\mathrm{supp}[(R,\Phi,T)-(\Delta R,\Delta\Phi,\Delta T),(R,\Phi,T)+(\Delta R,\Delta\Phi,\Delta T)]}(r,\varphi,t)\chi_{\mathrm{supp}[(R',\Phi',T')-((\Delta R)',(\Delta\Phi)',(\Delta T)'),(R',\Phi',T')+((\Delta R)',(\Delta\Phi)',(\Delta T)')]}(r,\varphi,t) \tag{A114}$$

$$\times\frac{1}{r^2}\frac{1}{1-t^2}\frac{\Delta R-|r-R|}{\Delta R}\frac{[-\operatorname{sgn}(\varphi-\Phi)]}{\Delta\Phi}\frac{\Delta T-|t-T|}{\Delta T}\frac{(\Delta R)'-|r-R'|}{(\Delta R)'}\frac{[-\operatorname{sgn}(\varphi-\Phi')]}{(\Delta\Phi)'}\frac{(\Delta T)'-|t-T'|}{(\Delta T)'}\mathrm{d}x(r,\varphi,t) \tag{A115}$$

$$+\int_{\mathbb{B}}\chi_{\mathrm{supp}[(R,\Phi,T)-(\Delta R,\Delta\Phi,\Delta T),(R,\Phi,T)+(\Delta R,\Delta\Phi,\Delta T)]}(r,\varphi,t)\chi_{\mathrm{supp}[(R',\Phi',T')-((\Delta R)',(\Delta\Phi)',(\Delta T)'),(R',\Phi',T')+((\Delta R)',(\Delta\Phi)',(\Delta T)')]}(r,\varphi,t) \tag{A116}$$

$$\times\frac{1}{r^2}(1-t^2)\frac{\Delta R-|r-R|}{\Delta R}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\frac{[-\operatorname{sgn}(t-T)]}{\Delta T}\frac{(\Delta R)'-|r-R'|}{(\Delta R)'}\frac{(\Delta\Phi)'-|\varphi-\Phi'|}{(\Delta\Phi)'}\frac{[-\operatorname{sgn}(t-T')]}{(\Delta T)'}\mathrm{d}x(r,\varphi,t) \tag{A117}$$

$$=\int_{lb_r}^{ub_r}\int_{lb_\varphi}^{ub_\varphi}\int_{lb_t}^{ub_t}\frac{[-\operatorname{sgn}(r-R)]}{\Delta R}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\frac{\Delta T-|t-T|}{\Delta T}\frac{[-\operatorname{sgn}(r-R')]}{(\Delta R)'}\frac{(\Delta\Phi)'-|\varphi-\Phi'|}{(\Delta\Phi)'}\frac{(\Delta T)'-|t-T'|}{(\Delta T)'}r^2\mathrm{d}r\,\mathrm{d}\varphi\,\mathrm{d}t \tag{A118}$$

$$+\int_{lb_r}^{ub_r}\int_{lb_\varphi}^{ub_\varphi}\int_{lb_t}^{ub_t}\frac{1}{r^2}\frac{1}{1-t^2}\frac{\Delta R-|r-R|}{\Delta R}\frac{[-\operatorname{sgn}(\varphi-\Phi)]}{\Delta\Phi}\frac{\Delta T-|t-T|}{\Delta T}\frac{(\Delta R)'-|r-R'|}{(\Delta R)'}\frac{[-\operatorname{sgn}(\varphi-\Phi')]}{(\Delta\Phi)'}\frac{(\Delta T)'-|t-T'|}{(\Delta T)'}r^2\mathrm{d}r\,\mathrm{d}\varphi\,\mathrm{d}t \tag{A119}$$

$$+\int_{lb_r}^{ub_r}\int_{lb_\varphi}^{ub_\varphi}\int_{lb_t}^{ub_t}\frac{1}{r^2}(1-t^2)\frac{\Delta R-|r-R|}{\Delta R}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\frac{[-\operatorname{sgn}(t-T)]}{\Delta T}\frac{(\Delta R)'-|r-R'|}{(\Delta R)'}\frac{(\Delta\Phi)'-|\varphi-\Phi'|}{(\Delta\Phi)'}\frac{[-\operatorname{sgn}(t-T')]}{(\Delta T)'}r^2\mathrm{d}r\,\mathrm{d}\varphi\,\mathrm{d}t \tag{A120}$$

$$=\int_{lb_r}^{ub_r}\frac{\operatorname{sgn}(r-R)}{\Delta R}\frac{\operatorname{sgn}(r-R')}{(\Delta R)'}r^2\mathrm{d}r\int_{lb_\varphi}^{ub_\varphi}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\frac{(\Delta\Phi)'-|\varphi-\Phi'|}{(\Delta\Phi)'}\mathrm{d}\varphi\int_{lb_t}^{ub_t}\frac{\Delta T-|t-T|}{\Delta T}\frac{(\Delta T)'-|t-T'|}{(\Delta T)'}\mathrm{d}t \tag{A121}$$

$$+\int_{lb_r}^{ub_r}\frac{\Delta R-|r-R|}{\Delta R}\frac{(\Delta R)'-|r-R'|}{(\Delta R)'}\mathrm{d}r\int_{lb_\varphi}^{ub_\varphi}\frac{\operatorname{sgn}(\varphi-\Phi)}{\Delta\Phi}\frac{\operatorname{sgn}(\varphi-\Phi')}{(\Delta\Phi)'}\mathrm{d}\varphi\int_{lb_t}^{ub_t}\frac{1}{1-t^2}\frac{\Delta T-|t-T|}{\Delta T}\frac{(\Delta T)'-|t-T'|}{(\Delta T)'}\mathrm{d}t \tag{A122}$$

$$+\int_{lb_r}^{ub_r}\frac{\Delta R-|r-R|}{\Delta R}\frac{(\Delta R)'-|r-R'|}{(\Delta R)'}\mathrm{d}r\int_{lb_\varphi}^{ub_\varphi}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\frac{(\Delta\Phi)'-|\varphi-\Phi'|}{(\Delta\Phi)'}\mathrm{d}\varphi\int_{lb_t}^{ub_t}(1-t^2)\frac{\operatorname{sgn}(t-T)}{\Delta T}\frac{\operatorname{sgn}(t-T')}{(\Delta T)'}\mathrm{d}t \tag{A123}$$

$$=\sum_{k=1}^{3}\int_{lb_{a_k}}^{ub_{a_k}}\frac{\operatorname{sgn}(a_k-A_k)}{\Delta A_k}\frac{\operatorname{sgn}(a_k-A_k')}{(\Delta A_k)'}\left\{\begin{array}{ll}a_k^2, & k=1,\\ 1, & k=2,\\ 1-a_k^2, & k=3\end{array}\right\}\mathrm{d}a_k \tag{A124}$$

$$\times\prod_{j=1,\ j\neq k}^{3}\int_{lb_{a_j}}^{ub_{a_j}}\frac{\Delta A_j-|a_j-A_j|}{\Delta A_j}\frac{(\Delta A_j)'-|a_j-A_j'|}{(\Delta A_j)'}\left\{\begin{array}{ll}\frac{1}{1-a_j^2}, & j=3,k=2,\\ 1, & \text{else}\end{array}\right\}\mathrm{d}a_j \tag{A125}$$

Note that $lb_r=lb_{a_1}, ub_r=ub_{a_1}, lb_\varphi=lb_{a_2}, ub_\varphi=ub_{a_2}, lb_t=lb_{a_3}$ and $ub_t=ub_{a_3}$. We see that also this integral fragments with respect to the variables a_j,$j=1, 2, 3$, and the only integrals left for a discussion are of the form

$$\int\frac{\operatorname{sgn}(x-X)\operatorname{sgn}(x-X')}{\Delta X(\Delta X)'}x^q\mathrm{d}x, \tag{A126}$$

again for $q\in\mathbb{N}_0$, here in particular $q=0$ and $q=2$, and

$$\int\frac{1}{1-t^2}\frac{\Delta T-|t-T|}{\Delta T}\frac{(\Delta T)'-|t-T'|}{(\Delta T)'}\mathrm{d}t. \tag{A127}$$

For each integral between two critical points, we obtain

$$\int\frac{\operatorname{sgn}(x-X)\operatorname{sgn}(x-X')}{\Delta X(\Delta X)'}x^q\mathrm{d}x=\pm\frac{1}{\Delta X(\Delta X)'}\int x^q\mathrm{d}x=\pm\frac{x^{q+1}}{(q+1)\Delta X(\Delta X)'} \tag{A128}$$

in the first case and

$$\int\frac{W(t;T,\Delta T,T',(\Delta T)')}{1-t^2}\mathrm{d}t \tag{A129}$$

$$=\int\begin{bmatrix}+\\+\\-\end{bmatrix}\frac{t^2}{1-t^2}\begin{bmatrix}+\\-\\+\end{bmatrix}\frac{t}{1-t^2}\left(\begin{bmatrix}-\\+\\+\end{bmatrix}T\begin{bmatrix}-\\+\\+\end{bmatrix}T'\begin{bmatrix}+\\+\\+\end{bmatrix}\Delta T\begin{bmatrix}+\\+\\-\end{bmatrix}(\Delta T)'\right) \tag{A130}$$

$$\begin{bmatrix}+\\+\\-\end{bmatrix}\frac{1}{1-t^2}\left(T\begin{bmatrix}-\\+\\+\end{bmatrix}\Delta T\right)\left(T'\begin{bmatrix}-\\+\\-\end{bmatrix}(\Delta T')\right)\mathrm{d}t \tag{A131}$$

$$=\begin{bmatrix}+\\+\\-\end{bmatrix}(\operatorname{atanh}(t)-t)\begin{bmatrix}-\\+\\-\end{bmatrix}\frac{\log\left(1-t^2\right)}{2}\left(\begin{bmatrix}-\\+\\+\end{bmatrix}T\begin{bmatrix}-\\+\\+\end{bmatrix}T'\begin{bmatrix}+\\+\\+\end{bmatrix}\Delta T\begin{bmatrix}+\\+\\-\end{bmatrix}(\Delta T)'\right) \tag{A132}$$

$$\begin{bmatrix}+\\+\\-\end{bmatrix}\operatorname{atanh}(t)\left(T\begin{bmatrix}-\\+\\+\end{bmatrix}\Delta T\right)\left(T'\begin{bmatrix}-\\+\\-\end{bmatrix}(\Delta T')\right) \tag{A133}$$

in the second case. Note that the latter is obviously not well-defined for $t=\pm1$ due to the logarithm which is why we have to diminish the domain for this variable in practice.

At last, we consider the mixed cases. We obtain

$$\left\langle N_{A,\Delta A},G^{\mathrm{I}}_{m,n,j}\right\rangle_{\mathrm{L}^2(\mathbb{B},\mathbb{R})} \tag{A134}$$

$$=\prod_{j=1}^{3}\int_{\max(A_{\min},A_j-\Delta A_j)}^{\min(A_{\max},A_j+\Delta A_j)}\frac{\Delta A_j-|a_j-A_j|}{\Delta A_j}G^{\mathrm{I}}_{m,n,j}(x(a_1,a_2,a_3))\left\{\begin{array}{ll}a_1^2, & j=1\\ 1, & j=2,3\end{array}\right\}\mathrm{d}a_j \tag{A135}$$

$$=p_{m,n}q_{n,j}\int_{\max(R_{\min},R-\Delta R)}^{\min(R_{\max},R+\Delta R)}\frac{\Delta R-|r-R|}{\Delta R}P_m^{(0,n+1/2)}(I(r))\left(\frac{r}{\mathcal{R}}\right)^n r^2\mathrm{d}r \tag{A136}$$

$$\times\int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi}\frac{\Delta\Phi-|\varphi-\Phi|}{\Delta\Phi}\operatorname{Trig}(j\varphi)\mathrm{d}\varphi\int_{\max(T_{\min},T-\Delta T)}^{\min(T_{\max},T+\Delta T)}\frac{\Delta T-|t-T|}{\Delta T}P_{n,|j|}(t)\mathrm{d}t. \tag{A137}$$

We immediately see that the integrals with respect to r and t cannot be calculated analytically. However, as they are 1-D integrals, they can easily be integrated numerically, for example with suitable software libraries. With respect to the φ-integral, we obtain

$$\frac{\Delta\Phi - |\varphi - \Phi|}{\Delta\Phi}\mathrm{Trig}(j\varphi) = \frac{\Delta\Phi - \mathrm{sgn}(\varphi - \Phi)(\varphi - \Phi)}{\Delta\Phi}\mathrm{Trig}(j\varphi) \tag{A138}$$

$$= \left(1 + \frac{\Phi\,\mathrm{sgn}(\varphi - \Phi)}{\Delta\Phi}\right)\mathrm{Trig}(j\varphi) - \frac{\mathrm{sgn}(\varphi - \Phi)}{\Delta\Phi}\,\varphi\mathrm{Trig}(j\varphi). \tag{A139}$$

Thus, the following cases remain:

$$I_1(j,\varphi) \coloneqq \int \varphi\mathrm{Trig}(j\varphi)\mathrm{d}\varphi = \int \left\{\begin{array}{c}\sqrt{2}\varphi\cos(j\varphi),\ j<0\\ \varphi,\ j=0\\ \sqrt{2}\varphi\sin(j\varphi),\ j>0\end{array}\right\}\mathrm{d}\varphi = \left\{\begin{array}{c}\sqrt{2}\left[\dfrac{\cos(j\varphi)}{j^2} + \dfrac{\varphi\sin(j\varphi)}{j}\right],\ j<0\\ \frac{1}{2}\varphi^2,\ j=0\\ \sqrt{2}\left[\dfrac{\sin(j\varphi)}{j^2} - \dfrac{\varphi\cos(j\varphi)}{j}\right],\ j>0\end{array}\right\} \tag{A140}$$

and

$$I_2(j,\varphi) \coloneqq \int \mathrm{Trig}(j\varphi)\mathrm{d}\varphi = \int \left\{\begin{array}{cc}\sqrt{2}\cos(j\varphi), & j<0\\ 1, & j=0\\ \sqrt{2}\sin(j\varphi), & j>0\end{array}\right\}\mathrm{d}\varphi = \left\{\begin{array}{cc}\frac{\sqrt{2}}{j}\sin(j\varphi), & j<0\\ \varphi, & j=0\\ -\frac{\sqrt{2}}{j}\cos(j\varphi), & j>0\end{array}\right\}. \tag{A141}$$

This yields

$$\int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \frac{\Delta\Phi - |\varphi - \Phi|}{\Delta\Phi}\mathrm{Trig}(j\varphi)\mathrm{d}\varphi \tag{A142}$$

$$= \int_{\Phi-\Delta\Phi}^{\Phi} \frac{\Delta\Phi - |\varphi - \Phi|}{\Delta\Phi}\mathrm{Trig}(j\varphi)\mathrm{d}\varphi + \int_{\Phi}^{\Phi+\Delta\Phi} \frac{\Delta\Phi - |\varphi - \Phi|}{\Delta\Phi}\mathrm{Trig}(j\varphi)\mathrm{d}\varphi \tag{A143}$$

$$= \int_{\Phi-\Delta\Phi}^{\Phi} \frac{\Delta\Phi + \varphi - \Phi}{\Delta\Phi}\mathrm{Trig}(j\varphi)\mathrm{d}\varphi + \int_{\Phi}^{\Phi+\Delta\Phi} \frac{\Delta\Phi - \varphi + \Phi}{\Delta\Phi}\mathrm{Trig}(j\varphi)\mathrm{d}\varphi \tag{A144}$$

$$= \int_{\Phi-\Delta\Phi}^{\Phi} \left(1 - \frac{\Phi}{\Delta\Phi} + \frac{\varphi}{\Delta\Phi}\right)\mathrm{Trig}(j\varphi)\mathrm{d}\varphi + \int_{\Phi}^{\Phi+\Delta\Phi} \left(1 + \frac{\Phi}{\Delta\Phi} - \frac{\varphi}{\Delta\Phi}\right)\mathrm{Trig}(j\varphi)\mathrm{d}\varphi \tag{A145}$$

$$= \frac{1}{\Delta\Phi}I_1(j,\varphi)|_{\Phi-\Delta\Phi}^{\Phi} + \left(1 - \frac{\Phi}{\Delta\Phi}\right)I_2(j,\varphi)|_{\Phi-\Delta\Phi}^{\Phi} - \frac{1}{\Delta\Phi}I_1(j,\varphi)|_{\Phi}^{\Phi+\Delta\Phi} + \left(1 + \frac{\Phi}{\Delta\Phi}\right)I_2(j,\varphi)|_{\Phi}^{\Phi+\Delta\Phi} \tag{A146}$$

$$= -\frac{1}{\Delta\Phi}I_1(j,\Phi-\Delta\Phi) - \left(1 - \frac{\Phi}{\Delta\Phi}\right)I_2(j,\Phi-\Delta\Phi) \tag{A147}$$

$$+ \frac{1}{\Delta\Phi}I_1(j,\Phi) + \left(1 - \frac{\Phi}{\Delta\Phi}\right)I_2(j,\Phi) + \frac{1}{\Delta\Phi}I_1(j,\Phi) - \left(1 + \frac{\Phi}{\Delta\Phi}\right)I_2(j,\Phi) \tag{A148}$$

$$- \frac{1}{\Delta\Phi}I_1(j,\Phi+\Delta\Phi) + \left(1 + \frac{\Phi}{\Delta\Phi}\right)I_2(j,\Phi+\Delta\Phi) \tag{A149}$$

$$= -\frac{1}{\Delta\Phi}I_1(j,\Phi-\Delta\Phi) - \left(1 - \frac{\Phi}{\Delta\Phi}\right)I_2(j,\Phi-\Delta\Phi) + \frac{2}{\Delta\Phi}I_1(j,\Phi) - \frac{2\Phi}{\Delta\Phi}I_2(j,\Phi) \tag{A150}$$

$$- \frac{1}{\Delta\Phi}I_1(j,\Phi+\Delta\Phi) + \left(1 + \frac{\Phi}{\Delta\Phi}\right)I_2(j,\Phi+\Delta\Phi). \tag{A151}$$

For the gradients, we have similarly

$$\left\langle \nabla_{r\xi(\varphi,t)}N_{(R,\Phi,T),(\Delta R,\Delta\Phi,\Delta T)},\ \nabla_{r\xi(\varphi,t)}G^{\mathrm{I}}_{m,n,j}\right\rangle_{\mathrm{L}^2\left(\mathbb{B},\mathbb{R}^3\right)} \tag{A152}$$

$$= \int_{\mathbb{B}} \chi_{\mathrm{supp}[(R,\Phi,T)-(\Delta R,\Delta\Phi,\Delta T),(R,\Phi,T)+(\Delta R,\Delta\Phi,\Delta T)]}(r,\varphi,t)\left(\varepsilon^r\frac{[-\mathrm{sgn}(r-R)]}{\Delta R}\frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi}\frac{\Delta T - |t-T|}{\Delta T}\right. \tag{A153}$$

$$\left. + \frac{1}{r}\varepsilon^{\varphi}\frac{1}{\sqrt{1-t^2}}\frac{\Delta R - |r-R|}{\Delta R}\frac{[-\mathrm{sgn}(\varphi-\Phi)]}{\Delta\Phi}\frac{\Delta T - |t-T|}{\Delta T} + \frac{1}{r}\varepsilon^{t}\sqrt{1-t^2}\frac{\Delta R - |r-R|}{\Delta R}\frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi}\frac{[-\mathrm{sgn}(t-T)]}{\Delta T}\right) \tag{A154}$$

$$\times \left(p_{m,n} q_{n,j} \sum_{p=1}^{4} G^{\mathrm{I}}_{m,n,j;p}(r\xi(\varphi,t)) \right) \mathrm{d}x(r,\varphi,t) \tag{A155}$$

$$= p_{m,n} q_{n,j} \int_{\mathbb{B}} \chi_{\mathrm{supp}[(R,\Phi,T)-(\Delta R,\Delta\Phi,\Delta T),(R,\Phi,T)+(\Delta R,\Delta\Phi,\Delta T)]}(r,\varphi,t) \tag{A156}$$

$$\times \left[\left(\varepsilon^r \frac{[-\operatorname{sgn}(r-R)]}{\Delta R} \frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi} \frac{\Delta T - |t-T|}{\Delta T} \right) \cdot \left(\sum_{p=1}^{2} G^{\mathrm{I}}_{m,n,j;p}(r\xi(\varphi,t)) \right) \right. \tag{A157}$$

$$+ \left(\frac{1}{r} \varepsilon^{\varphi} \frac{1}{\sqrt{1-t^2}} \frac{\Delta R - |r-R|}{\Delta R} \frac{[-\operatorname{sgn}(\varphi-\Phi)]}{\Delta\Phi} \frac{\Delta T - |t-T|}{\Delta T} \right) \cdot \left(G^{\mathrm{I}}_{m,n,j;3}(r\xi(\varphi,t)) \right) \tag{A158}$$

$$\left. + \left(\frac{1}{r} \varepsilon^{t} \sqrt{1-t^2} \frac{\Delta R - |r-R|}{\Delta R} \frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi} \frac{[-\operatorname{sgn}(t-T)]}{\Delta T} \right) \cdot \left(G^{\mathrm{I}}_{m,n,j;4}(r\xi(\varphi,t)) \right) \right] \mathrm{d}x(r,\varphi,t) \tag{A159}$$

$$= p_{m,n} q_{n,j} \int_{R-\Delta R}^{R+\Delta R} \int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \int_{T-\Delta T}^{T+\Delta T} \left(\varepsilon^r \frac{[-\operatorname{sgn}(r-R)]}{\Delta R} \frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi} \frac{\Delta T - |t-T|}{\Delta T} \right) \cdot \left(\sum_{p=1}^{2} G^{\mathrm{I}}_{m,n,j;p}(r\xi(\varphi,t)) \right) r^2 \mathrm{d}r\mathrm{d}\varphi\mathrm{d}t \tag{A160}$$

$$+ p_{m,n} q_{n,j} \int_{R-\Delta R}^{R+\Delta R} \int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \int_{T-\Delta T}^{T+\Delta T} \left(\frac{1}{r} \varepsilon^{\varphi} \frac{1}{\sqrt{1-t^2}} \frac{\Delta R - |r-R|}{\Delta R} \frac{[-\operatorname{sgn}(\varphi-\Phi)]}{\Delta\Phi} \frac{\Delta T - |t-T|}{\Delta T} \right) \cdot \left(G^{\mathrm{I}}_{m,n,j;3}(r\xi(\varphi,t)) \right) r^2 \mathrm{d}r\mathrm{d}\varphi\mathrm{d}t \tag{A161}$$

$$+ p_{m,n} q_{n,j} \int_{R-\Delta R}^{R+\Delta R} \int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \int_{T-\Delta T}^{T+\Delta T} \left(\frac{1}{r} \varepsilon^{t} \sqrt{1-t^2} \frac{\Delta R - |r-R|}{\Delta R} \frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi} \frac{[-\operatorname{sgn}(t-T)]}{\Delta T} \right) \cdot \left(G^{\mathrm{I}}_{m,n,j;4}(r\xi(\varphi,t)) \right) r^2 \mathrm{d}r\mathrm{d}\varphi\mathrm{d}t \tag{A162}$$

$$= p_{m,n} q_{n,j} \int_{R-\Delta R}^{R+\Delta R} \int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \int_{T-\Delta T}^{T+\Delta T} \left(\varepsilon^r \frac{[-\operatorname{sgn}(r-R)]}{\Delta R} \frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi} \frac{\Delta T - |t-T|}{\Delta T} \right) \tag{A163}$$

$$\times \left(\left(P_m^{(0,n+1/2)}(I(r)) \right)' I'(r) \left(\frac{r}{\mathcal{R}} \right)^n P_{n,|j|}(t) \mathrm{Trig}(j\varphi) \xi(\varphi,t) + \frac{n}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}} \right)^{n-1} P_{n,|j|}(t) \mathrm{Trig}(j\varphi) \xi(\varphi,t) \right) r^2 \mathrm{d}r\mathrm{d}\varphi\mathrm{d}t \tag{A164}$$

$$+ p_{m,n} q_{n,j} \int_{R-\Delta R}^{R+\Delta R} \int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \int_{T-\Delta T}^{T+\Delta T} \left(\frac{1}{r} \varepsilon^{\varphi} \frac{1}{\sqrt{1-t^2}} \frac{\Delta R - |r-R|}{\Delta R} \frac{[-\operatorname{sgn}(\varphi-\Phi)]}{\Delta\Phi} \frac{\Delta T - |t-T|}{\Delta T} \right) \tag{A165}$$

$$\times \left(\frac{j}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}} \right)^{n-1} \frac{1}{\sqrt{1-t^2}} P_{n,|j|}(t) \mathrm{Trig}(-j\varphi) \varepsilon^{\varphi}(\varphi,t) \right) r^2 \mathrm{d}r\mathrm{d}\varphi\mathrm{d}t \tag{A166}$$

$$+ p_{m,n} q_{n,j} \int_{R-\Delta R}^{R+\Delta R} \int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \int_{T-\Delta T}^{T+\Delta T} \left(\frac{1}{r} \varepsilon^{t} \sqrt{1-t^2} \frac{\Delta R - |r-R|}{\Delta R} \frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi} \frac{[-\operatorname{sgn}(t-T)]}{\Delta T} \right) \tag{A167}$$

$$\times \left(\frac{1}{\mathcal{R}} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}} \right)^{n-1} \sqrt{1-t^2} P'_{n,|j|}(t) \mathrm{Trig}(j\varphi) \varepsilon^{t}(\varphi,t) \right) r^2 \mathrm{d}r\mathrm{d}\varphi\mathrm{d}t \tag{A168}$$

$$= p_{m,n} q_{n,j} \int_{R-\Delta R}^{R+\Delta R} \frac{[-\operatorname{sgn}(r-R)]}{\Delta R} \left(P_m^{(0,n+1/2)}(I(r)) \right)' I'(r) \left(\frac{r}{\mathcal{R}} \right)^n r^2 \mathrm{d}r \int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi} \mathrm{Trig}(j\varphi)\mathrm{d}\varphi \int_{T-\Delta T}^{T+\Delta T} \frac{\Delta T - |t-T|}{\Delta T} P_{n,|j|}(t)\mathrm{d}t \tag{A169}$$

$$+ p_{m,n} q_{n,j} n \int_{R-\Delta R}^{R+\Delta R} \frac{[-\operatorname{sgn}(r-R)]}{\Delta R} r P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}} \right)^n \mathrm{d}r \int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi} \mathrm{Trig}(j\varphi)\mathrm{d}\varphi \int_{T-\Delta T}^{T+\Delta T} \frac{\Delta T - |t-T|}{\Delta T} P_{n,|j|}(t)\mathrm{d}t \tag{A170}$$

$$+ p_{m,n} q_{n,j} \int_{R-\Delta R}^{R+\Delta R} \frac{\Delta R - |r-R|}{\Delta R} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}} \right)^n \mathrm{d}r \int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \frac{[-\operatorname{sgn}(\varphi-\Phi)]}{\Delta\Phi} j\mathrm{Trig}(-j\varphi)\mathrm{d}\varphi \int_{T-\Delta T}^{T+\Delta T} \frac{\Delta T - |t-T|}{\Delta T} \frac{1}{1-t^2} P_{n,|j|}(t)\mathrm{d}t \tag{A171}$$

$$+ p_{m,n} q_{n,j} \int_{R-\Delta R}^{R+\Delta R} \frac{\Delta R - |r-R|}{\Delta R} P_m^{(0,n+1/2)}(I(r)) \left(\frac{r}{\mathcal{R}} \right)^n \mathrm{d}r \int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi} \mathrm{Trig}(j\varphi)\mathrm{d}\varphi \int_{T-\Delta T}^{T+\Delta T} \frac{[-\operatorname{sgn}(t-T)]}{\Delta T} (1-t^2) P'_{n,|j|}(t)\mathrm{d}t \tag{A172}$$

$$= p_{n,m} q_{n,j} \sum_{k=1}^{3} \int_{A_k-\Delta A_k}^{A_k+\Delta A_k} \frac{-\operatorname{sgn}(a_k - A_k)}{\Delta A_k} \begin{cases} \left(P_m^{(0,n+1/2)}(I(a_k))\right)' I'(a_k) \left(\frac{a_k}{\mathcal{R}}\right)^n a_k^2 + P_m^{(0,n+1/2)}(I(a_k)) \left(\frac{a_k}{\mathcal{R}}\right)^n n a_k, & k=1 \\ j\operatorname{Trig}(-j a_k), & k=2 \\ (1-a_k^2) P'_{n,|j|}(a_k), & k=3 \end{cases} \mathrm{d}a_k \tag{A173}$$

$$\times \prod_{i=1, i\neq k}^{3} \int_{A_i-\Delta A_i}^{A_i+\Delta A_i} \frac{\Delta A_i - |a_i - A_i|}{\Delta A_i} \begin{cases} P_m^{(0,n+1/2)}(I(a_i)) \left(\frac{a_i}{\mathcal{R}}\right)^n, & i=1 \\ \operatorname{Trig}(j a_i), & i=2 \\ P_{n,|j|}(a_i), & i=3, k=1 \\ \frac{1}{1-a_i^2} P_{n,|j|}(a_i), & i=3, k=2 \end{cases} \mathrm{d}a_i. \tag{A174}$$

Note that, for practical purposes, we still have to compute eight different integrals for this term. Moreover, the integrals with respect to $a_1 = r$ and $a_3 = t$ can still only be computed via numerical integration. For the case $a_2 = \varphi$, we obtain

$$\int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} -\frac{\operatorname{sgn}(\varphi-\Phi)}{\Delta\Phi} j\operatorname{Trig}(-j\varphi)\mathrm{d}\varphi = \int_{\Phi-\Delta\Phi}^{\Phi} \frac{1}{\Delta\Phi} j\operatorname{Trig}(-j\varphi)\mathrm{d}\varphi - \int_{\Phi}^{\Phi+\Delta\Phi} \frac{1}{\Delta\Phi} j\operatorname{Trig}(-j\varphi)\mathrm{d}\varphi \tag{A175}$$

$$= \frac{j}{\Delta\Phi} \left(\int_{\Phi-\Delta\Phi}^{\Phi} \operatorname{Trig}(-j\varphi)\mathrm{d}\varphi - \int_{\Phi}^{\Phi+\Delta\Phi} \operatorname{Trig}(-j\varphi)\mathrm{d}\varphi \right) \tag{A176}$$

$$= \frac{j}{\Delta\Phi} \left(2I_2(-j;\Phi) - I_2(-j;\Phi-\Delta\Phi) - I_2(-j;\Phi+\Delta\Phi)\right) \tag{A177}$$

and

$$\int_{\Phi-\Delta\Phi}^{\Phi+\Delta\Phi} \frac{\Delta\Phi - |\varphi-\Phi|}{\Delta\Phi} \operatorname{Trig}(j\varphi)\mathrm{d}\varphi, \tag{A178}$$

which we have already discussed above.

A4 Derivation of objective functions RFMP(· ; ·) and ROFMP(· ; ·) and related coefficients

We start with RFMP. The respective noise-cognizant Tikhonov–Phillips functional is given in eq. (63) by

$$\mathcal{J}^{\mathrm{SM}}(f_N + \alpha d; \mathcal{T}\neg, \lambda, \delta\psi, \sigma) \coloneqq \left\| \frac{R^N - \alpha \mathcal{T}\neg d}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| f_N + \alpha d \right\|_{\mathcal{H}_1(\mathbb{B})}^2, \qquad \lambda > 0, \tag{A179}$$

$$R^{N+1} \coloneqq R^N - \alpha_{N+1} \mathcal{T}\neg d_{N+1} = \delta\psi - \mathcal{T}\neg f_{N+1}. \tag{A180}$$

We now aim to determine $(\alpha, d),\ \alpha \in \mathbb{R},\ d \in \mathcal{D}$ which minimizes $\mathcal{J}^{\mathrm{SM}}$. We start as follows:

$$0 = \frac{\partial}{\partial\alpha} \mathcal{J}^{\mathrm{SM}}(f_N + \alpha d; \mathcal{T}\neg, \lambda, \delta\psi, \sigma) = \frac{\partial}{\partial\alpha} \left[\left\| \frac{R^N - \alpha \mathcal{T}\neg d}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| f_N + \alpha d \right\|_{\mathcal{H}_1(\mathbb{B})}^2 \right] \tag{A181}$$

$$= \frac{\partial}{\partial\alpha} \left[\left\| \frac{R^N}{\sigma} \right\|_{\mathbb{R}^\ell}^2 - 2\alpha \left\langle \frac{R^N}{\sigma}, \frac{\mathcal{T}\neg d}{\sigma} \right\rangle_{\mathbb{R}^\ell}^2 + \alpha^2 \left\| \frac{\mathcal{T}\neg d}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| f_N \right\|_{\mathcal{H}_1(\mathbb{B})}^2 + 2\alpha\lambda \left\langle f_N, d \right\rangle_{\mathcal{H}_1(\mathbb{B})}^2 + \alpha^2 \lambda \left\| d \right\|_{\mathcal{H}_1(\mathbb{B})}^2 \right] \tag{A182}$$

$$= -2 \left\langle \frac{R^N}{\sigma}, \frac{\mathcal{T}\neg d}{\sigma} \right\rangle_{\mathbb{R}^\ell}^2 + 2\alpha \left\| \frac{\mathcal{T}\neg d}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + 2\lambda \left\langle f_N, d \right\rangle_{\mathcal{H}_1(\mathbb{B})}^2 + 2\alpha\lambda \left\| d \right\|_{\mathcal{H}_1(\mathbb{B})}^2. \tag{A183}$$

This yields

$$\alpha_{N+1} = \frac{\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{T}\neg d}{\sigma} \right\rangle_{\mathbb{R}^\ell}^2 - \lambda \left\langle f_N, d \right\rangle_{\mathcal{H}_1(\mathbb{B})}^2}{\left\| \frac{\mathcal{T}\neg d}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| d \right\|_{\mathcal{H}_1(\mathbb{B})}^2}. \tag{A184}$$

Inserting this value into $\mathcal{J}^{\mathrm{SM}}(f_N + \alpha d; \mathcal{T}\neg, \lambda, \delta\psi, \sigma)$, we obtain

$$\mathcal{J}^{\mathrm{SM}}(f_N + \alpha_{N+1} d; \mathcal{T}\neg, \lambda, \delta\psi, \sigma) \tag{A185}$$

$$= \left\| \frac{R^N}{\sigma} \right\|_{\mathbb{R}^\ell}^2 - 2\alpha_{N+1} \left\langle \frac{R^N}{\sigma}, \frac{\mathcal{T}\neg d}{\sigma} \right\rangle_{\mathbb{R}^\ell}^2 + \alpha_{N+1}^2 \left\| \frac{\mathcal{T}\neg d}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| f_N \right\|_{\mathcal{H}_1(\mathbb{B})}^2 + 2\alpha_{N+1}\lambda \left\langle f_N, d \right\rangle_{\mathcal{H}_1(\mathbb{B})}^2 + \alpha_{N+1}^2 \lambda \left\| d \right\|_{\mathcal{H}_1(\mathbb{B})}^2 \tag{A186}$$

$$= \left\| \frac{R^N}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| f_N \right\|_{\mathcal{H}_1(\mathbb{B})}^2 - 2 \frac{\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{T}\neg d}{\sigma} \right\rangle_{\mathbb{R}^\ell}^2 - \lambda \left\langle f_N, d \right\rangle_{\mathcal{H}_1(\mathbb{B})}^2}{\left\| \frac{\mathcal{T}\neg d}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| d \right\|_{\mathcal{H}_1(\mathbb{B})}^2} \left[\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{T}\neg d}{\sigma} \right\rangle_{\mathbb{R}^\ell}^2 - \lambda \left\langle f_N, d \right\rangle_{\mathcal{H}_1(\mathbb{B})}^2 \right] \tag{A187}$$

$$+\left[\frac{\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{T}_\neg d}{\sigma}\right\rangle_{\mathbb{R}^\ell}^2 - \lambda \langle f_N, d\rangle_{\mathcal{H}_1(\mathbb{B})}^2}{\left\| \frac{\mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \|d\|_{\mathcal{H}_1(\mathbb{B})}^2}\right]^2 \left[\left\| \frac{\mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \|d\|_{\mathcal{H}_1(\mathbb{B})}^2\right] \tag{A188}$$

$$= \left\| \frac{R^N}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \|f_N\|_{\mathcal{H}_1(\mathbb{B})}^2 - \frac{\left(\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{T}_\neg d}{\sigma}\right\rangle_{\mathbb{R}^\ell}^2 - \lambda \langle f_N, d\rangle_{\mathcal{H}_1(\mathbb{B})}^2\right)^2}{\left\| \frac{\mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \|d\|_{\mathcal{H}_1(\mathbb{B})}^2} \tag{A189}$$

$$= \mathcal{J}^{\mathrm{SM}}(f_{N-1} + \alpha_N d_N; \mathcal{T}_\neg, \lambda, \delta\psi, \sigma) - \mathrm{RFMP}(d; N). \tag{A190}$$

Note that $\mathcal{J}^{\mathrm{SM}}(f_{N-1} + \alpha_N d_N; \mathcal{T}_\neg, \lambda, \delta\psi, \sigma)$ is fixed in the $(N+1)$th iteration. Thus, we see that, if we maximize RFMP$(\cdot\,;\cdot)$ as defined in eq. (63), we minimize the noise-cognizant Tikhonov–Phillips functional in the $(N+1)$th iteration.

Similarly, in the ROFMP, we consider the noise-cognizant Tikhonov–Phillips functional

$$\mathcal{J}_O^{\mathrm{SM}}\left(f_N^{(N)} + \alpha d; \mathcal{T}_\neg, \lambda, \delta\psi, \sigma\right) \coloneqq \left\| \frac{R^N - \alpha \mathcal{P}_{\mathcal{V}_N^\perp} \mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| f_N^{(N)} + \alpha\left(d - b_n^{(N)}(d)\right)\right\|_{\mathcal{H}_1(\mathbb{B})}^2, \tag{A191}$$

$$R^{N+1} \coloneqq R^N - \alpha_{N+1}^{(N+1)} \mathcal{P}_{\mathcal{V}_N^\perp} \mathcal{T}_\neg d_{N+1} = \delta\psi - \mathcal{T}_\neg f_{N+1}^{(N+1)}, \tag{A192}$$

for $\lambda > 0$. Confer Schneider, 2020 for a full definition. Again, we first consider the minimizer with respect to α:

$$0 = \frac{\partial}{\partial\alpha} \mathcal{J}_O^{\mathrm{SM}}\left(f_N^{(N)} + \alpha d; \mathcal{T}_\neg, \lambda, \delta\psi, \sigma\right) \tag{A193}$$

$$= \frac{\partial}{\partial\alpha}\left[\left\| \frac{R^N}{\sigma}\right\|_{\mathbb{R}^\ell}^2 - 2\alpha\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\rangle_{\mathbb{R}^\ell}^2 + \alpha^2 \left\| \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| f_N^{(N)}\right\|_{\mathcal{H}_1(\mathbb{B})}^2 + 2\alpha\lambda\left\langle f_N^{(N)}, d - b_n^{(N)}(d)\right\rangle_{\mathcal{H}_1(\mathbb{B})}^2 + \alpha^2\lambda \left\| d - b_n^{(N)}(d)\right\|_{\mathcal{H}_1(\mathbb{B})}^2\right] \tag{A194}$$

$$= -2\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\rangle_{\mathbb{R}^\ell}^2 + 2\alpha \left\| \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + 2\lambda\left\langle f_N^{(N)}, d - b_n^{(N)}(d)\right\rangle_{\mathcal{H}_1(\mathbb{B})}^2 + 2\alpha\lambda \left\| d - b_n^{(N)}(d)\right\|_{\mathcal{H}_1(\mathbb{B})}^2, \tag{A195}$$

which yields

$$\alpha_{N+1} = \alpha_{N+1}^{(N+1)} \coloneqq \frac{\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\rangle_{\mathbb{R}^\ell}^2 - \lambda\left\langle f_N^{(N)}, d - b_n^{(N)}(d)\right\rangle_{\mathcal{H}_1(\mathbb{B})}^2}{\left\| \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| d - b_n^{(N)}(d)\right\|_{\mathcal{H}_1(\mathbb{B})}^2}. \tag{A196}$$

Inserting $\alpha_{N+1}^{(N+1)}$ in $\mathcal{J}_O^{\mathrm{SM}}\left(f_N^{(N)} + \alpha d; \mathcal{T}_\neg, \lambda, \delta\psi, \sigma\right)$, we obtain

$$\mathcal{J}_O^{\mathrm{SM}}\left(f_N^{(N)} + \alpha_{N+1}^{(N+1)} d; \mathcal{T}_\neg, \lambda, \delta\psi, \sigma\right) \tag{A197}$$

$$= \left\| \frac{R^N}{\sigma}\right\|_{\mathbb{R}^\ell}^2 - 2\alpha_{N+1}^{(N+1)}\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\rangle_{\mathbb{R}^\ell}^2 + \left(\alpha_{N+1}^{(N+1)}\right)^2 \left\| \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 \tag{A198}$$

$$+ \lambda \left\| f_N^{(N)}\right\|_{\mathcal{H}_1(\mathbb{B})}^2 + 2\alpha_{N+1}^{(N+1)}\lambda\left\langle f_N^{(N)}, d - b_n^{(N)}(d)\right\rangle_{\mathcal{H}_1(\mathbb{B})}^2 + \left(\alpha_{N+1}^{(N+1)}\right)^2 \lambda \left\| d - b_n^{(N)}(d)\right\|_{\mathcal{H}_1(\mathbb{B})}^2 \tag{A199}$$

$$= \left\| \frac{R^N}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| f_N^{(N)}\right\|_{\mathcal{H}_1(\mathbb{B})}^2 \tag{A200}$$

$$- 2\frac{\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\rangle_{\mathbb{R}^\ell}^2 - \lambda\left\langle f_N^{(N)}, d - b_n^{(N)}(d)\right\rangle_{\mathcal{H}_1(\mathbb{B})}^2}{\left\| \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| d - b_n^{(N)}(d)\right\|_{\mathcal{H}_1(\mathbb{B})}^2}\left[\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\rangle_{\mathbb{R}^\ell}^2 + \lambda\left\langle f_N^{(N)}, d - b_n^{(N)}(d)\right\rangle_{\mathcal{H}_1(\mathbb{B})}^2\right] \tag{A201}$$

$$+ \left(\frac{\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\rangle_{\mathbb{R}^\ell}^2 - \lambda\left\langle f_N^{(N)}, d - b_n^{(N)}(d)\right\rangle_{\mathcal{H}_1(\mathbb{B})}^2}{\left\| \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| d - b_n^{(N)}(d)\right\|_{\mathcal{H}_1(\mathbb{B})}^2}\right)^2 \left[\left\| \frac{\mathcal{P}_{\mathcal{V}_N^\perp}\mathcal{T}_\neg d}{\sigma}\right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| d - b_n^{(N)}(d)\right\|_{\mathcal{H}_1(\mathbb{B})}^2\right] \tag{A202}$$

$$= \left\| \frac{R^N}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| f_N^{(N)} \right\|_{\mathcal{H}_1(\mathbb{B})}^2 - \frac{\left(\left\langle \frac{R^N}{\sigma}, \frac{\mathcal{P}_{\mathcal{V}_N^\perp} \mathcal{T}_\neg d}{\sigma} \right\rangle_{\mathbb{R}^\ell}^2 - \lambda \left\langle f_N^{(N)}, d - b_n^{(N)}(d) \right\rangle_{\mathcal{H}_1(\mathbb{B})}^2 \right)^2}{\left\| \frac{\mathcal{P}_{\mathcal{V}_N^\perp} \mathcal{T}_\neg d}{\sigma} \right\|_{\mathbb{R}^\ell}^2 + \lambda \left\| d - b_n^{(N)}(d) \right\|_{\mathcal{H}_1(\mathbb{B})}^2} \tag{A203}$$

$$= \mathcal{J}_O^{\mathrm{SM}} \left(f_{N-1}^{(N-1)} + \alpha_{N+1}^{(N+1)} d; \mathcal{T}_\neg, \lambda, \delta\psi, \sigma \right) - \mathrm{ROFMP}(d; N). \tag{A204}$$