

PAPER

# Bayesian approach to inverse scattering with topological priors

To cite this article: Ana Carpio *et al* 2020 *Inverse Problems* **36** 105001

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

# Bayesian approach to inverse scattering with topological priors

Ana Carpio<sup>1,2,4</sup> , Sergei Iakunin<sup>1,3</sup>  and Georg Stadler<sup>2</sup> 

<sup>1</sup> Universidad Complutense de Madrid, 28040 Madrid, Spain

<sup>2</sup> Courant Institute, New York University, NY 10012, United States of America

<sup>3</sup> Basque Center for Applied Mathematics-BCAM, 48009 Bilbao, Spain

E-mail: [ana\\_carpio@mat.ucm.es](mailto:ana_carpio@mat.ucm.es), [siakunin@bcamath.org](mailto:siakunin@bcamath.org) and [stadler@cims.nyu.edu](mailto:stadler@cims.nyu.edu)

Received 18 March 2020, revised 10 July 2020

Accepted for publication 28 July 2020

Published 24 September 2020



CrossMark

## Abstract

We propose a Bayesian inference framework to estimate uncertainties in inverse scattering problems. Given the observed data, the forward model and their uncertainties, we find the posterior distribution over a finite parameter field representing the objects. To construct the prior distribution we use a topological sensitivity analysis. We demonstrate the approach on the Bayesian solution of 2D inverse problems in light and acoustic holography with synthetic data. Statistical information on objects such as their center location, diameter size, orientation, as well as material properties, are extracted by sampling the posterior distribution. Assuming the number of objects known, comparison of the results obtained by Markov Chain Monte Carlo (MCMC) sampling and by sampling a Gaussian distribution found by linearization about the maximum *a posteriori* estimate show reasonable agreement. The latter procedure has low computational cost, which makes it an interesting tool for uncertainty studies in 3D. However, MCMC sampling provides a more complete picture of the posterior distribution and yields multi-modal posterior distributions for problems with larger measurement noise. When the number of objects is unknown, we devise a stochastic model selection framework.

Keywords: inverse scattering, Bayesian inference, topological prior, PDE-constrained optimization, MCMC sampling

(Some figures may appear in colour only in the online journal)

## 1. Introduction

Inverse scattering techniques are a common tool to detect objects in areas such as medicine, geophysics, or public security. The basic structure of the underlying mathematical problem

<sup>4</sup> Author to whom any correspondence should be addressed.

is as follows. An incident wave field illuminates a set of objects integrated in an ambient medium. The resulting wave field is measured at a set of detectors. Given the measured data, the goal is to reconstruct the unknown objects and their material properties. In practice, the process is affected by different sources of errors and uncertainty, such as external noise in the recorded data and errors in the measurement systems and governing mathematical equations. Some deterministic approaches are able to provide reasonable reconstructions of objects under specific conditions. However, these reconstructions depend on the choice of tuning parameters, such as thresholds in direct methods and regularization or stopping criteria in iterative procedures. Moreover, deterministic approaches do not provide information on the confidence we are allowed to have in the results and do not shine light on correlations between the inferred parameters. This gap is addressed by a Bayesian probabilistic formulation which provides a more complete picture of the reconstructed parameters and their uncertainties.

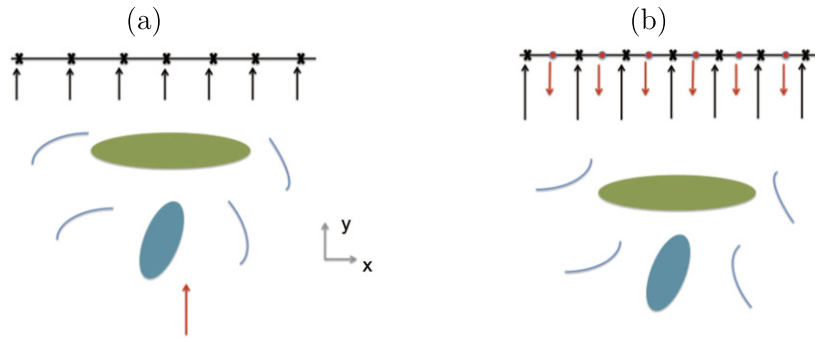
Details of the imaging process depend on the type of waves employed. To fix ideas, we focus here on situations where the physical process is modelled by a wave equation. This is the case, for instance, in inverse acoustic scattering [7, 20, 30, 58], and inverse electromagnetic scattering when using polarized radiation [2, 13, 23, 49]. When the incident wave is time harmonic,  $U_{\text{inc}}(\mathbf{x}, t) = e^{-i\omega t} u_{\text{inc}}(\mathbf{x})$ , the total wave field is time harmonic too, i.e.,  $U(\mathbf{x}, t) = e^{-i\omega t} u(\mathbf{x})$ . Its amplitude  $u(\mathbf{x})$  obeys a Helmholtz transmission problem, which, in two dimensions, is

$$\begin{cases} \Delta u + \kappa_e^2 u = 0 & \text{in } \Omega_e, \\ \Delta u + \kappa_i^2 u = 0 & \text{in } \Omega_i, \\ u^- - u^+ = 0, \quad \beta \frac{\partial u^-}{\partial \mathbf{n}} - \frac{\partial u^+}{\partial \mathbf{n}} = 0 & \text{on } \partial\Omega_i, \\ \lim_{|\mathbf{x}| \rightarrow 0} |\mathbf{x}|^{1/2} \left( \frac{\partial}{\partial |\mathbf{x}|} (u - u_{\text{inc}}) - i k_e (u - u_{\text{inc}}) \right) = 0, \end{cases} \quad (1)$$

where  $\Omega_i$  is an inclusion,  $\Omega_e = \mathbb{R}^2 \setminus \overline{\Omega_i}$ , and  $\mathbf{n}$  is the unit outer normal vector for  $\Omega_i$ . The symbols  $-$  and  $+$  denote values from inside and outside  $\Omega_i$ , respectively. The Sommerfeld radiation condition on the propagation of the scattered field  $u_{\text{sc}} = u - u_{\text{inc}}$  at infinity implies that only outgoing waves are allowed. The parameters  $\beta, \kappa_i, \kappa_e \geq 0$  depend on the frequency  $\omega$  and the material properties. Moreover,  $\kappa_e \sim k_e$  is assumed to be constant outside of a ball containing the objects and the detectors  $\mathbf{x}_j, j = 1, \dots, N$ .

To model the measurement process, we generate data  $\mathbf{d}$  by solving the forward problem (1) with inclusions we consider as the ‘truth’ and evaluate the resulting wave field at detectors located at  $\mathbf{x}_j, j = 1, \dots, N$ . We then add independent additive white noise of a magnitude specified in each problem to obtain ‘synthetic’ measurement. Depending on the application, the measured data are complex-valued amplitude fields  $u(\mathbf{x}_j)$  (in microwave imaging [2, 49] or acoustic holography [30], for instance) or real-valued intensities  $|u(\mathbf{x}_j)|^2$  (in light microscopy [19, 47]). Given the measured data and the ambient properties  $\kappa_e$ , our goal is to find the inclusions  $\Omega_i = \cup_{\ell=1}^L \Omega_i^\ell$ . Different experimental imaging set-ups correspond to different arrangements of emitted incident waves and detector distributions, see figure 1. Here, we focus on the configuration displayed in figure 1(a) for our numerical tests, though the methods extend to other arrangements. We will consider both types of data, complex valued fields and real intensities, which is the case in acoustic and light holography [19, 30, 47], respectively. Bayesian methods are particularly interesting for acoustic imaging, since the magnitude of noise in the recorded data is usually larger.

If we assume that each object can be approximated by star-shaped parameterizations, that is, the boundary of each  $\Omega_i^\ell$  is defined by ‘rays’ emerging from a center, the inverse scattering problem amounts to finding a set of parameters  $\boldsymbol{\nu}$ : the number of components, their centers



**Figure 1.** Schematic arrangements of incident waves (red beams) and detector grids (black crosses) in different imaging set-ups. (a) In holography and microscopy, one incident planar wave interacts with the objects and is recorded at a fixed grid of detectors behind the objects [30, 47]. (b) In acoustic imaging of materials, waves emitted from a grid of sources interact with the medium and the reflected waves are recorded at a grid of receivers in the same region [4, 27, 33]. Blue lines represent waves interacting with the objects and black arrows indicate the reflections arriving at the detector grid. Microwave imaging uses set-ups similar to (a) for multiple incident wave frequencies, while the relative positions of objects, emitters and receptors are rotated to increase the number of independent observations [2, 49].

and the coefficients of the trigonometric expansions defining their boundaries, as well as their material constants  $\kappa_i^\ell$ . The coefficients  $\kappa_i^\ell$  may be spatially dependent, which increases the complexity of the problem. Here, we take them to be constant and usually known. We represent star-shaped objects using expansions in trigonometric polynomials for the radius [11, 13, 38] instead of general functions varying with the angle [7, 25] to reduce the parameter dimension.

There is a broad literature on deterministic inverse scattering approaches, which basically fall in two categories. Non-iterative methods such as linear sampling, factorization, multiple signal classification, orthogonality sampling, direct sampling and topological derivative analysis [8, 26, 39, 42] provide direct approximations to the objects from the data. Iterative techniques may refine this information at a higher computational cost, and vary widely with the specific application. In most set-ups, one can resort to regularized nonlinear least-squares optimization formulations, where the governing equation is treated as an equality constraint. The idea is to seek shapes that fit the measurements within the noise level, and to use regularization or early termination of iterative algorithms to prevent overfitting. In principle, the existence of several local minima is a possibility—if local minima occur depends on the distribution of detectors, the incident wave directions and the local wave speed. To obtain the best estimate for the true configurations, one may follow different strategies: produce sharp initial guesses of the objects or select wide ranges of detectors and incident waves or frequencies [1, 11, 14, 34, 35, 37], when possible, and include additional regularizing terms in the cost functional, see [13, 38] for instance.

Such optimization-based methods are naturally related to Bayesian approaches, which consider all variables in the inverse problem as random variables. Assuming the variables are a finite-dimensional vector  $\nu$ , the densities of the random variables are related using Bayes' formula [40, 56]:

$$p_{\text{pt}}(\nu) := p(\nu|\mathbf{d}) = \frac{p(\mathbf{d}|\nu)}{p(\mathbf{d})} p_{\text{pr}}(\nu). \quad (2)$$

Here,  $p_{\text{pr}}(\cdot)$  is the prior density of the variables, which incorporates prior (expert) knowledge available about the variables;  $p(\mathbf{d}|\boldsymbol{\nu})$  is the likelihood or conditional probability of the observations  $\mathbf{d}$  given the variables  $\boldsymbol{\nu}$ ; and  $p_{\text{pt}}(\mathbf{d}|\boldsymbol{\nu})$  is the posterior density of the parameters given the data, which is the solution of the Bayesian inverse problem. The density  $p(\mathbf{d})$  is a scaling that does not depend on the parameters. Full characterization of the posterior density is an extremely challenging probability problem for moderate and high-dimensional parameters  $\boldsymbol{\nu}$ , and often one has to rely on approximations of the posterior distribution. One approximation is finding the maximum *a posteriori* (MAP) point, i.e., the parameters that maximize the posterior density. This amounts to an optimization problem that, making assumptions on the prior and likelihood and taking the negative logarithm of the densities, becomes a nonlinear least-squares problem of the form used in deterministic inversion; see section 2. In this optimization problem, the regularization is implied by the prior density and the likelihood corresponds to the data misfit term. Linearization about the MAP point, also called Laplace approximation, results in an approximation of the posterior density by a multivariate Gaussian, allowing for computationally efficient manipulations even for high- or infinite-dimensional parameters [6, 58]. Alternatively, to avoid a Gaussian approximation of the posterior, we can sample the posterior distribution using, for instance, MCMC techniques [25, 32, 54] at a higher computational cost.

The results obtained with Bayesian methods depend on the choice of the prior [19, 25]. Prior distributions can involve additional parameters, so-called ‘hyperparameters’ [43]. There are different approaches to handling hyperparameters. A straightforward strategy consists in fixing subjective guesses for them. The results depend on how good such subjective guesses are. A second possibility is to select empirical guesses, i.e., one solves the problem using different hyperparameters and then chooses the most appropriate one [9]. Finally, one might introduce additional probability densities for the hyperparameters, and work in an hierarchical Bayes framework [5]. In this work, we construct subjective guesses for the prior distributions from a topological sensitivity analysis of the cost functionals. This allows us to handle imaging set-ups with rather limited data, i.e., in which detectors are located in a narrow area, only one incident wave is used, multiple penetrable objects may be present, and the recorded data may not be the complex amplitude field, but just its real modulus. We will show that both the Bayesian linearized approach and MCMC sampling provide reasonable descriptions of statistical properties of the objects for various noise levels in the observations, when the number of object components is known. Previous work in 3D light holographic settings also applied MCMC sampling to infer the location, size and refractive index of single spherical particles [19]. In [50–52], scattering from single 2D sound-soft objects is considered. Here, the objects are placed at known locations and inference is based on far-field, small-noise, complex data. Similar to this work, MCMC sampling is used to infer Fourier modes in a starshaped parametrization, amongst other parametrizations. Bayesian approaches relying on more complex parameterizations varying with the angle are presented in [7, 25] for 2D acoustic and tomography set-ups.

When the number of objects components is not known, this number becomes an additional unknown. Compared to the other parameters, it is of a rather different nature: it is discrete, and it controls the presence of other parameter blocks. In this case, we reformulate our problem within the frameworks of hierarchical Bayesian modeling and model selection [15]. The number of objects can then be selected by empirical arguments or by MCMC methods with variable selection [18], see also [24] for a hierarchical level set approach. [50] applies model selection to estimate the number of modes in the radii of single starshaped objects instead.

The next sections are organized as follows. In section 2, we devise a general Bayesian approach for object reconstruction, which we test in the physical set-ups described in section 3.

Section 4 uses topological sensitivity analysis to estimate the number of objects and define priors. Assuming the number of objects and their material properties known, we compute MAP points for their shape and then sample the Gaussian posterior approximation obtained by linearization at the MAP point in section 5. The results for different geometries are compared to MCMC sampling of the full posterior distribution in section 6. Section 7 discusses cases where we allow for variations in the material constants of the inclusions. In section 8, the number of objects is considered as unknown, and thus we implement a Bayesian strategy for model selection. Finally, section 9 presents our conclusions. For completeness, a final appendix A discusses formulas for different derivatives relevant in sections 5 and 7.

## 2. Bayesian approach for a given number of objects

In this section we develop a framework for object detection. We assume that we can approximate the objects using star-shaped parameterizations, but more general parameterizations could be used instead. The parameters for an inclusion consisting of  $L \geq 1$  object components are collected in a vector  $\boldsymbol{\nu} := (\boldsymbol{\nu}^1, \dots, \boldsymbol{\nu}^L) \in \mathbb{R}^{L(2M+3)}$ , where

$$\boldsymbol{\nu}^\ell = (c_x^\ell, c_y^\ell, a_0^\ell, a_1^\ell, \dots, a_M^\ell, b_1^\ell, \dots, b_M^\ell), \quad \ell = 1, \dots, L. \quad (3)$$

Here,  $(c_x^\ell, c_y^\ell)$  are the centers and  $r^\ell(t)$  the radii of the object components, associated to the parameterization

$$\mathbf{q}^\ell(t) = (c_x^\ell, c_y^\ell) + r^\ell(t)(\cos(2\pi t), \sin(2\pi t)), \quad t \in [0, 1], \quad (4)$$

$$r^\ell(t) = a_0^\ell + 2 \sum_{m=1}^M a_m^\ell \cos(2\pi m t) + 2 \sum_{m=1}^M b_m^\ell \sin(2\pi m t), \quad (5)$$

for  $\ell = 1, \dots, L$ . Note that similar parametrizations are available in three dimensions, where they could rely for instance on spherical harmonics instead of Fourier series [13]. The number of modes  $M$ , which we fix, controls the complexity of the boundary. The Bayesian approach to inverse problems requires to define a prior distribution for the parameters  $\boldsymbol{\nu}$ . We choose  $p(\boldsymbol{\nu})$  as a multivariate Gaussian

$$p(\boldsymbol{\nu}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{|\boldsymbol{\Gamma}_{\text{pr}}|}} \exp\left(-\frac{1}{2}(\boldsymbol{\nu} - \boldsymbol{\nu}_0)' \boldsymbol{\Gamma}_{\text{pr}}^{-1} (\boldsymbol{\nu} - \boldsymbol{\nu}_0)\right) \quad (6)$$

with covariance matrix  $\boldsymbol{\Gamma}_{\text{pr}}$  and  $n := L(2M + 3)$ . Notice that the radii (5) belong to the space of trigonometric polynomials  $T_{2M+1}$ , and are expanded in an orthonormal basis  $\phi_1, \dots, \phi_{2M+1}$ , so that the mass matrix associated to this basis is the identity. Otherwise, the mass matrix would enter (6), see [6]. Modeling of the prior distribution, i.e., the choices for the covariance matrix  $\boldsymbol{\Gamma}_{\text{pr}}$  and the mean  $\boldsymbol{\nu}_0$ , are discussed in section 4.

We illuminate the  $L$  objects with an incident plane wave of amplitude  $u_{\text{inc}}(\mathbf{x})$ , generating data  $\mathbf{d}$  at detectors  $\mathbf{x}_j$ ,  $j = 1, \dots, N$ . We denote by  $\mathbf{f}: \mathbb{R}^{L(2M+3)} \rightarrow \mathbb{R}^N$  the parameter-to-observable map, i.e., the mathematical description of this process. To be more precise, for parameters  $\boldsymbol{\nu}$ , we denote by  $u_{\Omega_\nu}$  the solution of the wave equation (1) with object  $\Omega_i = \Omega_\nu$  defined by (3)–(5). Then  $\mathbf{f}(\boldsymbol{\nu}) = (f(u_{\Omega_\nu}(\mathbf{x}_j)))_{j=1}^N$ , where  $f$  is the measurement operator (which may be real  $f(u) = |u|^2$  or complex valued  $f(u) = u$ ).

We assume additive Gaussian measurement noise, i.e., the observations and parameters are related by

$$\mathbf{d} = \mathbf{f}(\boldsymbol{\nu}) + \boldsymbol{\varepsilon}. \quad (7)$$

Here, the measurement noise  $\varepsilon$  is distributed as a multivariate Gaussian  $\mathcal{N}(0, \mathbf{\Gamma}_n)$  with mean zero and covariance matrix  $\mathbf{\Gamma}_n$ . We consider the noise level for each sensor to be equal and uncorrelated, so that  $\mathbf{\Gamma}_n$  is a real diagonal matrix. For complex valued data, the additive noise is represented by a standard complex Gaussian variable whose real and imaginary parts are both real Gaussians of the form  $\mathcal{N}(0, \mathbf{\Gamma}_n/2)$  [46].

Due to these assumptions, on the measurement noise, the conditional probability density  $p(\mathbf{d}|\boldsymbol{\nu})$  takes the form

$$p(\mathbf{d}|\boldsymbol{\nu}) = \frac{1}{(2\pi)^{N/2} \sqrt{|\mathbf{\Gamma}_n|}} \exp\left(-\frac{1}{2}\|\mathbf{f}(\boldsymbol{\nu}) - \mathbf{d}\|_{\mathbf{\Gamma}_n^{-1}}^2\right), \quad (8)$$

where  $\|\mathbf{v}\|_{\mathbf{\Gamma}_n^{-1}}^2 = \mathbf{v}^T \mathbf{\Gamma}_n^{-1} \mathbf{v}$ . Combining (2), (8), (6) and neglecting normalization constants, the posterior density becomes, up to multiplicative constants,

$$p_{\text{pt}}(\boldsymbol{\nu}) \propto \exp\left(-\frac{1}{2}\|\mathbf{f}(\boldsymbol{\nu}) - \mathbf{d}\|_{\mathbf{\Gamma}_n^{-1}}^2 - \frac{1}{2}\|\boldsymbol{\nu} - \boldsymbol{\nu}_0\|_{\mathbf{\Gamma}_{\text{pr}}^{-1}}^2\right). \quad (9)$$

Taking logarithms, the problem of maximizing the posterior probability of the parameter set  $\boldsymbol{\nu}$  given the data  $\mathbf{d}$  is identical to minimizing the regularized cost objective [3]:

$$J(\boldsymbol{\nu}) := \frac{1}{2}\|\mathbf{f}(\boldsymbol{\nu}) - \mathbf{d}\|_{\mathbf{\Gamma}_n^{-1}}^2 + \frac{1}{2}\|\boldsymbol{\nu} - \boldsymbol{\nu}_0\|_{\mathbf{\Gamma}_{\text{pr}}^{-1}}^2, \quad (10)$$

where we neglect additional terms involving only the covariances. The first part of functional (10) is related to the standard cost used in deterministic inverse problems, whereas the second part originating from the prior takes the role of the regularization, which prevents ill-posedness and overfitting of the observation data.

In our experiments, we choose a diagonal covariance matrix  $\mathbf{\Gamma}_n = \text{diag}(\sigma_1^2, \dots, \sigma_N^2)$ , and set all the variances equal to a constant  $\sigma_{\text{noise}}^2$ . Thus,  $\sqrt{|\mathbf{\Gamma}_n|} = \prod_{j=1}^N \sigma_j = \sigma_{\text{noise}}^N$ . The mean  $\boldsymbol{\nu}_0$  in the prior multivariate Gaussian density and the elements of the covariance matrix  $\mathbf{\Gamma}_{\text{pr}}$  are considered hyperparameters to be selected as discussed in the next section. This selection reflects our uncertainty in the prior information available, and affects the resulting parameter inference. We introduce a strategy to generate guesses for all the hyperparameters based on the topological fields of the objective functional (12). For that purpose, the available data are split in two parts: a fraction  $\mathbf{d}^{(1)}$  is used to generate the prior, whereas the remaining values  $\mathbf{d}^{(2)} =: \mathbf{d}$  are used in (9) to define the posterior distribution. We use intersperse grids as detailed in section 5.2. We generate synthetic data  $\mathbf{d}^{(1)}$  and  $\mathbf{d}$  for our tests solving (1) by BEM methods using ‘true’ object inclusions, and then add noise (as detailed in our numerical tests) to these observations.

### 3. Physical set-ups

We will study the behavior of the methods presented in this paper in light and acoustic holography set-ups, adjusted to figure 1(a). In this section we briefly summarize the physics background and the parameter choices we make for the remainder of this paper.

Let us fix a reference length scale  $L$ . Typical object sizes may range from  $L/10$  to  $2L$ , for instance, while the distance to the detectors is about  $5L$ . In an acoustic setting, the evolution of the total wave field  $U$  is governed by the wave equation  $\rho(\mathbf{x}) U_{tt}(\mathbf{x}, t) - \text{div}(\alpha(\mathbf{x}) \nabla U(\mathbf{x}, t)) = 0$ , where  $\rho$  is the density and  $\alpha$  represents a relevant ‘modulus’. In liquids,  $\alpha$  represents the bulk modulus and in solids,  $\alpha$  is Young’s modulus. In gases,  $\alpha = \gamma p$ , where  $p$  is the pressure and  $\gamma$  is related to the specific heat. The speed of sound in the medium is then  $c = (\alpha/\rho)^{1/2}$ .

When the applied field is time-harmonic, then  $U(\mathbf{x}, t) = e^{-i\omega t} u(\mathbf{x})$ , where the amplitude  $u(\mathbf{x})$  obeys  $\operatorname{div}(\alpha(\mathbf{x}) \nabla u(\mathbf{x})) + \rho(\mathbf{x}) \omega^2 u(\mathbf{x}) = 0$ .

We nondimensionalize the problem according to  $x = x'L$ ,  $y = y'L$ ,  $\Omega = \Omega'L$ ,  $u_{\text{inc}} = u_0 u'_{\text{inc}}$ ,  $u = u_0 u'$  and then drop the  $'$  for ease of notation. Here,  $L$  is the reference length and  $u_0$  the modulus of the amplitude of the incident wave. The total amplitude field is then governed by (1), where the parameters are the dimensionless wavenumbers,

$$\kappa_e = \frac{2\pi\nu}{c_e} L, \quad \kappa_i = \frac{2\pi\nu}{c_i} L, \quad (11)$$

as well as the ratio  $\beta = \frac{\alpha_i}{\alpha_e}$ . We denote by  $\nu = \frac{\omega}{2\pi}$  the frequency of the emitted sound wave and by  $c_e$  and  $c_i$  the sound speed in the ambient medium and inside the objects, respectively. The incident amplitude for the plane wave is  $u_{\text{inc}} = e^{i\kappa_e y}$ , where  $y$  points in the direction of the detectors.

Frequencies for sound lie in the range 20 Hz–20 kHz. The speed of sound in air is  $343 \text{ m s}^{-1}$ . If for instance  $L \sim 6 \text{ cm}$ ,  $\kappa_e \sim 12$ –20 for 11–18 kHz. Depending on the material the object is made of, it can be sound-soft (it absorbs sound), sound-hard (it reflects sound), or penetrable. We assume penetrable objects. For the other types, the governing equations and formulas should be adjusted following [10].

In a light holography set-up, the general framework is similar [12, 13], assuming we use polarized light in the presence of few well separated objects. Then, the equations governing the amplitude field can be approximated by (1) and the wavenumbers are again given by (11),  $c$  and  $\nu$  representing light wavespeed and frequencies. In this case,  $\beta = \frac{\mu_e}{\mu_i}$ ,  $\mu_e$  and  $\mu_i$  are material permeabilities. For the biological applications we target,  $\beta \sim 1$ . Visible light wavelengths lie the range of 400–700 nm and result in wavenumbers  $\kappa_e \sim 12$ –20, setting  $L \sim 1 \mu\text{m}$ , for instance. While in classical light holography, the measured data are real intensities  $|u|^2$  [12], it has become possible to record complex-valued data  $u$  in acoustic holography set-ups [30]. The latter are affected by larger noise magnitudes.

#### 4. Topological selection of priors

Topological derivative methods generate first guesses of objects without *a priori* information, other than the measured noisy data, the ambient medium properties and the incident wave. In deterministic frameworks, such guesses are then improved by level set, shape derivative, topological derivative or Gauss–Newton iterations [11, 13, 14, 35]. Following a Bayesian approach, we propose the following procedure to generate prior densities from topological sensitivity studies of the underlying unregularized cost functional (12). Such prior knowledge can significantly influence Bayesian inference results, as, e.g., observed in [19, 25].

##### 4.1. Topological derivative of the cost functional

Given  $\mathcal{R} \subset \mathbb{R}^2$ , the topological derivative of

$$J_c(\mathbb{R}^2 \setminus \overline{\Omega_\nu}) := \frac{1}{2} \sum_{j=1}^N |d_j - f(u_{\Omega_\nu}(\mathbf{x}_j))|^2, \quad (12)$$

is a scalar field  $D_T(\mathbf{x}, \mathcal{R})$  satisfying [57]:

$$J_c(\mathcal{R} \setminus \overline{B(\mathbf{x}, \varepsilon)}) = J_c(\mathcal{R}) + D_T(\mathbf{x}, \mathcal{R}) \operatorname{meas}(B(\mathbf{x}, \varepsilon)) + o(\varepsilon^2)$$

for any  $\mathbf{x} \in \mathcal{R}$  and any small radius  $\varepsilon > 0$ . Here,  $\text{meas}(B(\mathbf{x}, \varepsilon))$  is the volume of the ball centered at  $\mathbf{x}$  with radius  $\varepsilon$ . When  $D_T(\mathbf{x}, \mathcal{R}) < 0$ , the cost functional decreases by removing small balls centered at  $\mathbf{x}$ . This suggests that placing objects  $\Omega$  in regions where the topological derivative is negative, the cost functional should decrease. In this way, we obtain guesses for object component locations. When  $\mathcal{R} = \mathbb{R}^2$ , the topological derivative admits an explicit expression in terms of auxiliary forward and adjoint fields [11, 33]. For  $\mathbf{x} \in \mathbb{R}^2$

$$D_T(\mathbf{x}, \mathbb{R}^2) = \text{Re} \left[ \frac{2(1 - \beta)}{1 + \beta} \nabla u_{\text{inc}}(\mathbf{x}) \nabla \bar{p}_{\text{inc}}(\mathbf{x}) + (\beta \kappa_i^2 - \kappa_e^2) u_{\text{inc}}(\mathbf{x}) \bar{p}_{\text{inc}}(\mathbf{x}) \right], \quad (13)$$

where  $\kappa_e$  is the wavenumber for the outer medium,  $\kappa_i$  the wavenumber for the inclusion whose shape we are seeking and  $u_{\text{inc}}$  is the incident wave. When  $\kappa_e$  is constant, the conjugate adjoint field is

$$\bar{p}_{\text{inc}}(\mathbf{x}) = \frac{i}{4} \sum_{j=1}^N H_0^{(1)}(\kappa_e |\mathbf{x} - \mathbf{x}_j|) \chi(\mathbf{x}_j), \quad (14)$$

with  $H_0^{(1)}$  the Hankel function and  $\chi(\mathbf{x}_j) = \overline{(d_j - f(u_{\text{inc}}(\mathbf{x}_j)))} f'(u_{\text{inc}}(\mathbf{x}_j))$ . Standard choices for the measurement operator are  $f(u) = u$  with  $f'(u) = 1$  and  $f(u) = |u|^2$  with  $f'(u) = 2\bar{u}$ . When  $\kappa_e$  varies spatially inside a bounded region,  $\bar{p}_{\text{inc}}$  is a solution of:

$$\Delta \bar{p} + \kappa_e^2 \bar{p} = \sum_{j=1}^N \chi(\mathbf{x}_j) \delta_{\mathbf{x}_j} \text{ in } \mathbb{R}^2, \quad \lim_{|\mathbf{x}| \rightarrow \infty} |\mathbf{x}|^{1/2} \left( \frac{\partial \bar{p}}{\partial |\mathbf{x}|} - i k_e \bar{p} \right) = 0,$$

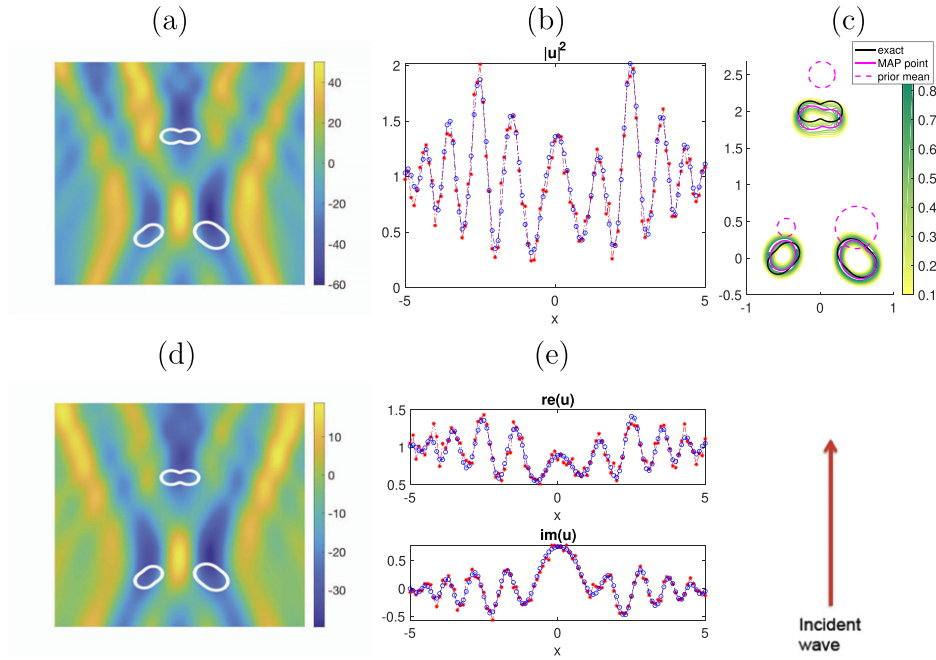
$k_e$  being its constant value at infinity.

#### 4.2. Prior selection

Figure 2 displays the topological derivative field (13) for three objects when  $\kappa_i = 15.12$  and  $\kappa_e = 12.56$ . Regions where large negative values are attained provide guesses of the object locations, which we use as prior knowledge as follows. We choose a constant  $C_0 \in (0, 1)$  and define the set  $\Omega_0 = \{\mathbf{x} \in R_{\text{obs}} | D_T(\mathbf{x}, \mathbb{R}^2) < (1 - C_0) \min_{\mathbf{y} \in R_{\text{obs}}} D_T(\mathbf{y}, \mathbb{R}^2)\}$  in the region where we look for objects, the so-called observation region  $R_{\text{obs}}$ . Then, we fit circles to each connected component to construct our initial guess  $\Omega_{\nu_0}$ .

For single objects, we test 50 values between 0.01 and 0.3 for  $C_0$ , choosing the one which yields the smallest value for the cost functional. Here, depending on the value selected for  $C_0$ , we will capture one, two or three dominant negative regions. Since we assume we know the number of objects, we set  $C_0$  to capture  $L = 3$  objects. Star-shaped objects are a deformation of circles. We fit circles to each component  $\Omega_{\nu_0}^\ell$  as follows. For  $\ell = 1, \dots, L$ ,  $(c_{x,0}^\ell, c_{y,0}^\ell)$  is the center of mass of the component and  $a_{0,0}^\ell$  is the minimum distance from the center to the boundary (or the average of minimum and maximum distances). We set  $a_{m,0}^\ell = 0$  and  $b_{m,0}^\ell = 0$ ,  $m = 1, \dots, M$  to avoid inserting unnecessary bias in the hyperparameters. Finally, we check that  $J_c(\mathbb{R}^2 \setminus \Omega_{\nu_0}) \leq J_c(\mathbb{R}^2)$  for the cost functional (12). Otherwise, we divide  $a_{0,0}^\ell$  by 2 until this requirement is fulfilled.

These values define the prior mean  $\nu_0$  in (9) and (10), which are also the initial guesses shown in the figures discussed in sections 5.2 and 6. We select covariances that render non-positive radii in (5) unlikely. For that purpose, we choose  $\Gamma_{\text{pr}}$  as a block diagonal matrix. In each block, the components decrease with increasing mode number. Additionally, in our imaging set-up, the uncertainty in the direction of the incident wave  $c_{y,0}^\ell$  is larger than the uncertainty in  $c_{x,0}^\ell$ . This is due to the increased uncertainty of the topological derivative in the incident wave,



**Figure 2.** Topological derivative (13) for a configuration with three objects whose contours are superimposed: (a) with data  $|u(\mathbf{x}_j)|^2$ ,  $j = 1, \dots, 101$ , (d) with data  $u(\mathbf{x}_j)$ ,  $j = 1, \dots, 101$ , corrupted by noise of magnitude 5% and 10%, respectively. The noisy data, depicted in panels (b) and (e), are recorded in the set-ups represented in Figure 1(a). Red asterisks represent the noisy data, whereas blue circles are the ‘true’ synthetic values generated solving the forward problem by BEM. Panel (c) represents the information on the objects inferred for (a) and (b) by the Bayesian linearized approach developed in section 5 with  $M = 5$ . Parameter values:  $\kappa_i = 15.12$ ,  $\kappa_e = 12.56$ ,  $\beta = 1$ .

i.e., the  $y$ -direction. Specifically, we choose the variances  $(\sigma_x^\ell)^2 = 0.1$  and  $(\sigma_y^\ell)^2 = 0.2$ . The variances for the radius mode coefficients  $a_{m,0}$  and  $b_{m,0}$  are inspired by convergence results for deterministic approaches which use  $H^s$ -norms for the radius, that is, weighted  $L^2$ -norms forcing decay [38]. We set  $(\sigma_0^\ell)^2 = 0.1$  and  $(\sigma_m^\ell)^2 = 0.1/(1 + m^2)^s$ ,  $s$  large,  $1 \leq m \leq M$ . In our tests, we usually set  $s = 3$ . In this way, the prior favors regular shapes, i.e., shapes with  $r(t) > 0$ .

Once the prior distribution is defined, we resort to different techniques to explore the posterior distribution. Methods for doing this as well as numerical results are presented in the next two sections. The simplest one computes a maximum *a posteriori* (MAP) estimate and samples the linearized posterior distribution to infer properties of the objects which generated the data, as illustrated in figure 2(c). Green contours represent probabilities of belonging to a boundary (built from the samples), whereas the magenta curve represents the MAP point.

## 5. Sampling from a Bayesian linearized formulation

An approximation of the posterior density (9), which builds on tools often available for optimization in deterministic inverse problems, is the Laplace approximation obtained by linearization at the maximum a posteriori (MAP) point. The approach first computes the MAP parameter

vector  $\boldsymbol{\nu}_{\text{MAP}}$ , which minimizes the negative log likelihood (10), and then approximates the posterior distribution by a multivariate Gaussian  $\mathcal{N}(\boldsymbol{\nu}_{\text{MAP}}, \boldsymbol{\Gamma}_{\text{pt}})$  with posterior covariance matrix  $\boldsymbol{\Gamma}_{\text{pt}} = \mathbf{H}^{-1}$ , where  $\mathbf{H}$  is the Hessian of (10) evaluated at  $\boldsymbol{\nu}_{\text{MAP}}$  [40, 56]. If the parameter-to-observable map  $\mathbf{f}(\cdot)$  were linear (and with the additive Gaussian noise and Gaussian prior assumptions made in section 2), this approximation of the posterior distribution would be exact. In general, the accuracy of this Gaussian posterior approximation depends on the degree of nonlinearity of  $\mathbf{f}$ .

Efficient computation of the MAP point is discussed in section 5.1. Once  $\boldsymbol{\nu}_{\text{MAP}}$  is available, the Hessian can either be computed explicitly or using low-rank approximations for large or infinite-dimensional parameters [6, 54]. For a parameter-to-observable map  $\mathbf{f}(\boldsymbol{\nu}) = (f(u_{\Omega_\nu}(\mathbf{x}_j)))_{j=1}^N$  with measurement operator  $f : \mathbb{C} \rightarrow \mathbb{C}$ ,  $f(u) = u$  one obtains

$$\boldsymbol{\Gamma}_{\text{pt}} = (\text{Re}[\mathbf{F}^{\text{ad}} \boldsymbol{\Gamma}_{\text{n}}^{-1} \mathbf{F}] + \boldsymbol{\Gamma}_{\text{pr}}^{-1})^{-1}. \quad (15)$$

Here,  $\mathbf{F}$  is the Jacobian matrix of  $\mathbf{f}(\cdot)$  evaluated at  $\boldsymbol{\nu}_{\text{MAP}}$  and its adjoint  $\mathbf{F}^{\text{ad}} = \overline{\mathbf{F}}^T$  is the conjugate transpose of  $\mathbf{F}$ . If the measurement operator is  $f : \mathbb{C} \rightarrow \mathbb{R}$ ,  $f(u) = |u|^2$ , the first part of the Hessian, which represents the amount of information learned from the data, is  $\text{Re}[\mathbf{F}^{\text{ad}} \mathbf{M}_h \boldsymbol{\Gamma}_{\text{n}}^{-1} \mathbf{F}]$ , where  $\mathbf{M}_h$  is a real diagonal matrix, see section 5.1. Samples from this posterior distribution approximation  $\mathcal{N}(\boldsymbol{\nu}_{\text{MAP}}, \boldsymbol{\Gamma}_{\text{pt}})$  can be drawn as

$$\boldsymbol{\nu}^{\text{pt}} = \boldsymbol{\nu}_{\text{MAP}} + \boldsymbol{\Gamma}_{\text{pt}}^{1/2} \mathbf{n}, \quad (16)$$

where  $\mathbf{n}$  is a vector of independent and identically distributed (i.i.d.) standard normal random values and  $\boldsymbol{\Gamma}_{\text{pt}}^{1/2}$  is a square root of the positive posterior covariance matrix [58]. Analogously, samples from the prior are drawn using  $\boldsymbol{\nu}^{\text{pr}} = \boldsymbol{\nu}_0 + \boldsymbol{\Gamma}_{\text{pr}}^{1/2} \mathbf{n}$ .

To compute the MAP point, we compared different strategies to solve the nonlinear least-squares problem (10): gradient descent, Newton methods, Gauss–Newton (GN) and Levenberg–Marquardt (LM) variants. We detail the procedure finally used next.

### 5.1. Computing the MAP point

Newton methods to minimize a functional  $J(\boldsymbol{\nu})$  implement the iteration  $\boldsymbol{\nu}^{k+1} = \boldsymbol{\nu}^k - \mathbf{H}^{-1} \mathbf{g}$ , where  $\mathbf{H}$  is the Hessian and  $\mathbf{g}$  is the gradient of  $J$ . This is equivalent to solving systems of the form  $\mathbf{H} \boldsymbol{\xi}^{k+1} = -\mathbf{g}$ , followed by the update step  $\boldsymbol{\nu}^{k+1} := \boldsymbol{\nu}^k + \boldsymbol{\xi}^{k+1}$ . Levenberg–Marquardt approaches add a ‘damping’ term  $\mu \mathbf{I}$ ,  $\mu > 0$ , to the system matrix. The value of  $\mu$  is adjusted at each iteration. If the objective yields a strong decrease from one iteration to the next, small values of  $\mu$  are used and the method resembles a Newton-type scheme. When the decrease is slow, large values of  $\mu$  are selected and the method becomes closer to gradient descent. We use a variant of this method proposed by Fletcher [28], which scales this ‘damping’ term to allow for larger steps in directions along which the gradient is smaller using  $\mu \text{diag}(\mathbf{H})$  instead of  $\mu \mathbf{I}$ .

For the objective (10) and the two forms of the parameter-to-observable map  $\mathbf{f}$ , we compute gradients and Hessians as follows, where for the latter we neglect second-order derivatives of measurement operators. The resulting Hessian approximation is sometimes called Gauss–Newton Hessian. Given a parameter-to-observable map  $\mathbf{f}(\boldsymbol{\nu}) = (f(u_{\Omega_\nu}(\mathbf{x}_j)))_{j=1}^N$  with measurement operator  $f : \mathbb{C} \rightarrow \mathbb{C}$ ,  $f(u) = u$ , we have

$$\begin{aligned} \mathbf{g}(\boldsymbol{\nu}) &:= \text{Re}[\mathbf{F}^{\text{ad}}(\boldsymbol{\nu}) \boldsymbol{\Gamma}_{\text{n}}^{-1} (\mathbf{f}(\boldsymbol{\nu}) - \mathbf{d})] + \boldsymbol{\Gamma}_{\text{pr}}^{-1} (\boldsymbol{\nu} - \boldsymbol{\nu}_0), \\ \mathbf{H}^{\text{GN}}(\boldsymbol{\nu}) &:= \text{Re}[\mathbf{F}^{\text{ad}}(\boldsymbol{\nu}) \boldsymbol{\Gamma}_{\text{n}}^{-1} \mathbf{F}(\boldsymbol{\nu})] + \boldsymbol{\Gamma}_{\text{pr}}^{-1}, \end{aligned}$$

where here and in the following,  $\mathbf{F}$  denotes the Fréchet derivative of the parameter-to-observable map  $\boldsymbol{\nu} \rightarrow (u_{\Omega_{\boldsymbol{\nu}}}(\mathbf{x}_j))_{j=1}^N$ ,  $u_{\Omega_{\boldsymbol{\nu}}}$  being the solution of (1). The characterization of this operator in terms of the solutions of auxiliary boundary value problems, as well as the way to calculate  $\mathbf{F}$ , are discussed in appendix A. For a measurement operator  $f : \mathbb{C} \rightarrow \mathbb{R}$ ,  $f(u) = |u|^2$ , and diagonal constant  $\Gamma_n$ , gradient and Hessian are

$$\mathbf{g}(\boldsymbol{\nu}) := \frac{1}{\sigma_{\text{noise}}^2} \text{Re}[\mathbf{F}^{\text{ad}}(\boldsymbol{\nu}) \mathbf{M}_g (\mathbf{f}(\boldsymbol{\nu}) - \mathbf{d})] + \Gamma_{\text{pr}}^{-1} (\boldsymbol{\nu} - \boldsymbol{\nu}_0),$$

$$\mathbf{H}^{\text{GN}}(\boldsymbol{\nu}) := \frac{1}{\sigma_{\text{noise}}^2} \text{Re}[\mathbf{F}^{\text{ad}}(\boldsymbol{\nu}) \mathbf{M}_h \mathbf{F}(\boldsymbol{\nu})] + \Gamma_{\text{pr}}^{-1},$$

where  $\mathbf{M}_g$  and  $\mathbf{M}_h$  are diagonal matrices defined as follows:

$$\mathbf{M}_g = \text{diag} [(2u_{\Omega_{\boldsymbol{\nu}}}(\mathbf{x}_j))_{j=1}^N], \quad \mathbf{M}_h = \text{diag} [(6|u_{\Omega_{\boldsymbol{\nu}}}(\mathbf{x}_j)|^2 - 2d_j)_{j=1}^N].$$

To summarize, we use the Levenberg–Marquardt method with scaled diagonal matrix and the Gauss–Newton Hessian to compute the MAP point. That is, starting from  $\boldsymbol{\nu}^0 = \boldsymbol{\nu}_0$ , we use the iteration  $\boldsymbol{\nu}^{k+1} = \boldsymbol{\nu}^k + \boldsymbol{\xi}^{k+1}$ , where  $\boldsymbol{\xi}^{k+1}$  is the solution of

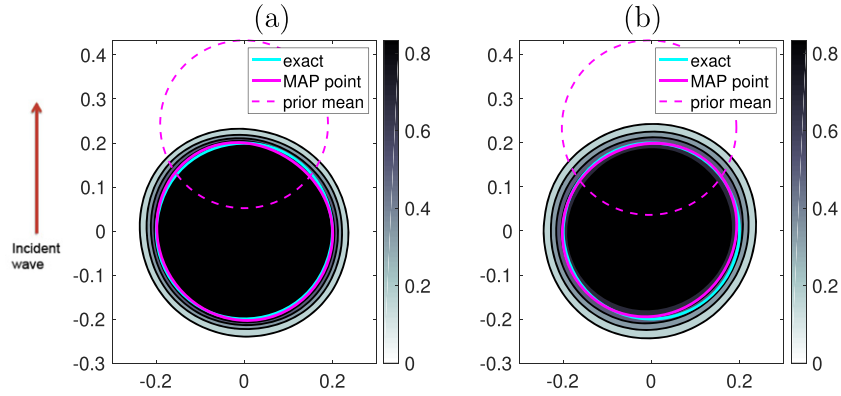
$$(\mathbf{H}_{\lambda_k}^{\text{GN}}(\boldsymbol{\nu}^k) + \mu_k \text{diag}(\mathbf{H}_{\lambda_k}^{\text{GN}}(\boldsymbol{\nu}^k))) \boldsymbol{\xi}^{k+1} = -\mathbf{g}_{\lambda_k}(\boldsymbol{\nu}^k).$$

Here, the subscript  $\lambda_k$  indicates that we multiply  $\Gamma_{\text{pr}}^{-1}$  by a factor  $\lambda_k$  in the initial steps to balance the different orders of magnitude of the two terms defining the cost  $J$  in (10). The initial value  $\lambda_0 = 0.1 \sigma_{\text{noise}}^{-2}$  decreases each iteration by a factor  $2/3$  until it reaches the value  $\lambda = 1$  corresponding to (10). We set  $\mu_k = 10^{-3}$  for each iteration, and check if the functional decreases sufficiently. If it does not, this value is increased by a fixed factor. This usually only happens for a few steps, and  $\mu$  is bounded from above by 1. The iteration stops when the difference between the new value of the cost and the previous one is smaller than  $\tau \sigma_{\text{noise}}^{-2}$ , where  $\tau$  is a tolerance (usually  $\tau = 10^{-5}$ ). The final parametrization  $\boldsymbol{\nu}_{\text{MAP}}$  is considered the MAP point. Computing this MAP point typically requires 15–25 iterations for problems with single objects.

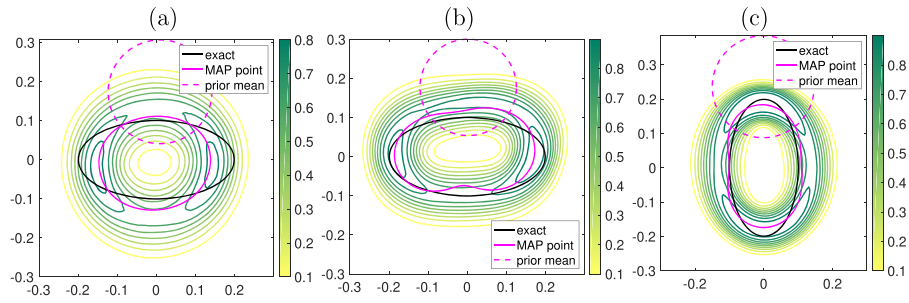
Notice that a star-shaped parameterization of an object is uniquely defined when its center is fixed. However, star-shaped objects (circles or ellipses, for instance) may be parameterized with different centers, at least when infinitely many coefficients are used in the expansion. If we consider only the cost functional (12), we may encounter different minima defining the same or very similar objects. Some optimization strategies [13] overcome this issue by enforcing that the center of the object must be the center of mass. In our Bayesian approach, the regularization (which results from our choice of prior) in (10) results in a preference for one parametrization for the same object.

## 5.2. Object detection with quantified uncertainty

Once we have computed the MAP point, we can apply the Bayesian linearized framework to object detection. We have performed our tests in the set-up shown in figure 1(a), adapted to either acoustic or light holography; these settings are described in detail in section 3. We place detectors at  $\mathbf{x}_j = -5 + 0.05j$ ,  $j = 0, \dots, 200$ , on a uniform grid of step size 0.05. The subgrid with step size 0.1 is used to compute the MAP point, whereas the data at intermediate detectors is used to generate the topological priors. The same splitting is used for our MCMC sampling experiments in the next section.



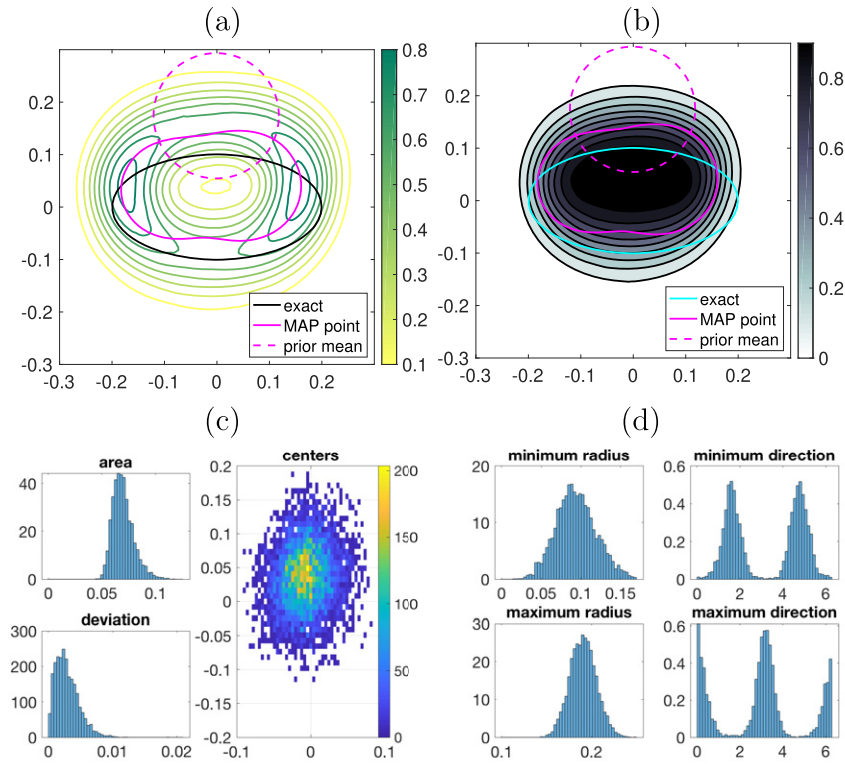
**Figure 3.** Linearized Bayesian solution for synthetic measurements of complex amplitudes using a sphere centered at  $(0, 0)$  with radius 0.2 with noise (a) 1% and (b) 5%. Cyan curves show the true object used to generate the synthetic data, whereas the obtained MAP points are shown in magenta. Dashed curves are the prior means constructed by topological methods, which are also used as initial guesses in the optimization. Black contours show the probability of points to belong to the object. Parameter values are  $\kappa_i = 15.12$ ,  $\kappa_e = 12.56$ ,  $\beta = 1$ , and  $M = 5$ . The arrow indicates the incidence direction. Detectors are placed at a distance 5 in that direction, as in figure 1(a).



**Figure 4.** (a) and (b) Inferred shapes for an ellipse centered at  $(0, 0)$  with semi-axes of lengths 0.2 and 0.1 in the  $x$  and  $y$  directions, respectively, when (a)  $\kappa_i = 15.12$ ,  $\kappa_e = 12.56$ , and (b)  $\kappa_i = 25.12$ ,  $\kappa_e = 20.56$ , with  $\beta = 1$ ,  $M = 5$ . (c) Displays results for the same parameters as in (b) orienting the ellipse in the  $y$  direction. Green contours show the pointwise marginal for each point on the curve. Data: Complex amplitudes at detectors placed at a distance 5 in the incidence direction, with 5% noise.

Figures 2–5 show the results for two sets of wavenumbers, namely  $\kappa_i = 15.12$ ,  $\kappa_e = 12.56$ , and  $\kappa_i = 25.12$ ,  $\kappa_e = 20.56$ . The incident wave is  $u_{\text{inc}} = e^{i\kappa_e y}$ , where the  $y$ -coordinate is oriented as in figure 1(a).

The topological approach described in section 4 is used to define the priors, and the prior mean is used as initialization to compute the MAP point. The functional (10) is minimized as explained in section 5.1 to obtain the MAP point. Linearizing about it and approximating the posterior probability by a Gaussian, we generate samples by means of (16). Based on them, we compute probabilities for points to lie inside the object (e.g., figure 4), the pointwise marginals of the contour (e.g., figure 5), as well as the marginal distributions of the centers of mass, the maximum and minimum distances from the center to the object border and their orientation.



**Figure 5.** Inferred shapes from complex data with 10% noise for the physical parameters  $\kappa_i = 25.12$ ,  $\kappa_e = 20.56$ ,  $\beta = 1$ , and  $M = 5$ . The true shape is an ellipse centered at  $(0, 0)$  with semi-axes of length 0.2 and 0.1 in the  $x$  and  $y$  directions, respectively. Contours represent the contours for the pointwise marginal for points on the curve in (a) or the probability of being inside the object in (b). The histograms in (c) and (d) are discrete approximations of the densities for the distribution of the area, the deviation from a circular shape, the center of mass, the minimal and maximal radius and the angle of the direction of the minimal and maximal extension of the inferred objects.

A small number of samples have negative radius as defined by (4) and (5) resulting in shapes with loops, which we discard for these calculations.

Figure 3 compares the results for a circle-shaped object as the noise magnitude increases from 1% to 5%, for the lower wavenumbers. The position, size and shape are inferred with small uncertainty. Switching to ellipsoidal shapes, the values of  $\kappa_i$ ,  $\kappa_e$  need to be increased for a more precise description of the dimensions and the orientation of the object, see figure 4. However, smaller values locate the true center more accurately. A similar phenomenon is observed in 3D [13]. Improved resolution has also been achieved increasing the wavenumber in acoustic 2D settings [4] with full aperture far field measurements, however, the number of incident directions and sampling points is increased too. Here, we keep the same single incident direction and limited aperture sampling points. The inferred shapes are still reasonable for higher noise magnitudes. Figure 5 illustrates the results for an ellipse, including statistics for its center of mass, size and orientation, with 10% noise. Similar results are obtained when halving the number of detectors (these results are not shown here). If we reduce the number of detectors further, the distance between them becomes larger than the object size. As

a consequence, the MAP point loses the elliptical shape, but its location and size are still adequate. In figure 5, we additionally study the uncertainty in the inferred objects in terms of quantities that do not depend on the parametrization, namely the area, the center of mass and the minimal and maximal radii and their directions. The area for a closed curve  $\mathcal{C}$  is computed as  $\int_{\mathcal{C}} r(s) ds$ . The center of mass of a curve with the parameterization  $\mathbf{q}(t) = (x(t), y(t))$ ,  $t \in I$ , is  $\int_I \mathbf{q}(t) |\mathbf{q}'(t)| dt / \int_I |\mathbf{q}'(t)| dt$ , where  $|\mathbf{q}'(t)| dt = \sqrt{x'(t)^2 + y'(t)^2} dt = ds$  is the differential of arch length. The deviation from a circular shape is  $\int_{\mathcal{C}} |r(s) - r_{av}| ds$ , where  $r(s)$  is the distance to the center of mass and  $r_{av}$  the average radius. The maximum and minimum radii are the longest and shortest distances of the curve to the center of mass, respectively. The angle between the  $x$ -axis and the direction of longest and shortest radius is denoted as maximum and minimum direction in figure 5. Figures 3–5 use as data complex amplitudes measured at 101 detectors placed at a distance 5 of the object in the incidence direction. As commented earlier, figure 2(c) uses real intensities at the detectors as measured data. Unlike ellipsoidal shapes, these three objects can be exactly parametrized for a small number  $M = 5$  of modes to represent the object geometry. The reconstruction of the configuration is reasonable.

The computational cost of this procedure is moderate. The MAP points are usually obtained in about 15–25 iterations, and sampling with the expression (16) requires no additional forward solves besides computing the Hessian at the MAP point. The figures discussed here use 10,000 samples. However, we have introduced some approximations in the Bayesian inference process: first, when approximating the posterior by a Gaussian, second, when approximating the full Hessian with the Gauss–Newton Hessian at the MAP point. To assess the validity of the procedure, we will compare to results obtained by MCMC sampling of the whole posterior distribution.

## 6. Sampling with Markov Chain Monte Carlo

MCMC techniques have the potential to fully explore and statistically characterize the posterior distributions without linearizing the parameter-to-observable map. A Markov chain is a sequential stochastic process, which moves from one state to another within an allowed set of states:  $X^0 \rightarrow X^1 \rightarrow \dots \rightarrow X^k \rightarrow \dots$ . To define a Markov chain we need three elements: (1) the state space, that is, the set of states  $X$  the chain is allowed to reach, (2) the transition operator  $p(X^{k+1} | X^k)$  which establishes the probability of transitioning from state  $X^k$  to  $X^{k+1}$ , and (3) the initial distribution  $\pi_0$  which defines the initial probability of being in any of the possible states. To generate a Markov chain, one moves from one state to another guided by the transition operator  $p(X^{k+1} | X^k)$ .

In our context, we wish to sample posterior distributions  $\pi$  by means of Markov chains, and a natural choice for  $\pi_0$  is the prior distribution. There are many MCMC variants for sampling posterior distributions [3], which are based on different transition operators. MCMC algorithms often suffer from ‘slow mixing’, i.e., the underlying Markov Chain takes (too) many samples to explore the parameter space and thus provide a good characterization of the distribution we aim at sampling. For large scale problems near a continuous limit, preconditioned Crank–Nicholson and stochastic Newton variants have been shown successful [17, 25, 54]. However, when the target distribution is multimodal, these samplers may fail to jump from one mode to another. Thus, we resort to affine invariant MCMC samplers working with several chains [32] because their strategy to generate new proposals reduces the occurrence of samples with negative radii during the sampling process and the use of many chains allows us to handle multimodal distributions.

In our imaging set-up, the initial distribution  $\pi_0$  is the truncated prior distribution, where we neglect normalization factors

$$\pi_0(\boldsymbol{\nu}) = \begin{cases} 0, & \text{if intersections,} \\ \exp(-\frac{1}{2}(\boldsymbol{\nu} - \boldsymbol{\nu}_0)' \mathbf{T}_{\text{pr}}^{-1}(\boldsymbol{\nu} - \boldsymbol{\nu}_0)), & \text{otherwise.} \end{cases} \quad (17)$$

By intersection we mean that the radius of the star-shaped curves with coefficients  $\boldsymbol{\nu}$  vanishes at at least one point and thus the curve is degenerate, and, for multiple objects, that we have intersecting or nested boundaries. Disregarding the normalization constants, which do not play a role in MCMC methods, the posterior to be sampled is

$$\pi(\boldsymbol{\nu}) = \pi_0(\boldsymbol{\nu}) \exp\left(-\frac{1}{2}(\mathbf{d} - \mathbf{f}(\boldsymbol{\nu}))' \mathbf{T}_n^{-1}(\mathbf{d} - \mathbf{f}(\boldsymbol{\nu}))\right).$$

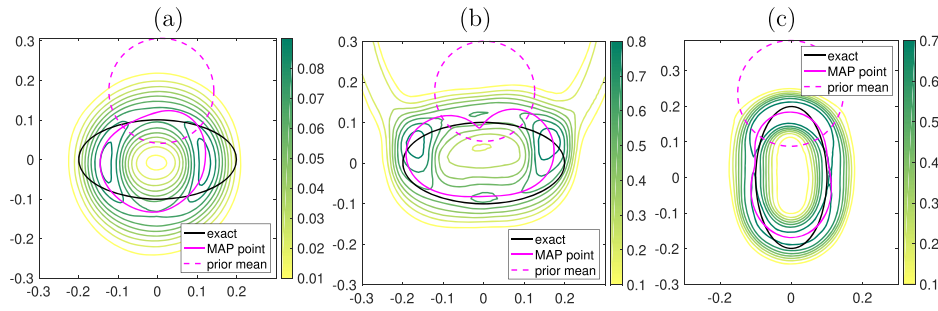
For the algorithm to sample the posterior distribution we follow [32], which we summarize for completeness:

- Initialization: generate the initial positions of the walkers  $X_0^w \in \mathbb{R}^d$ ,  $w = 1, \dots, W$  by sampling from the prior distribution  $\pi_0$ . Choose the acceptance parameter  $a$  (typically 2) and the number of steps  $K$ .
- for each step  $k = 0, \dots, K - 1$ , evolve the walkers  $w = 1, \dots, W$  as follows:
  - \* Draw a walker  $X_k^q$  at random from the set of walkers  $\{X_k^j\}_{j \neq w}$ .
  - \* Choose a random value  $z_w$  from the distribution  $g(z) = \frac{1}{\sqrt{z}}$  when  $z \in [1/a, a]$ , zero otherwise.
  - \* Calculate proposition  $X_{\text{prop}}^w = X_k^q + z_w(X_k^w - X_k^q)$ .
  - \* Calculate  $s = z_w^{d-1} \frac{\pi(X_{\text{prop}}^w)}{\pi(X_k^w)}$ . Calculate  $s = \text{Min}(1, s)$ .
  - \* Draw  $r$  with probability  $\mathcal{U}(0, 1)$ . If  $r \leq s$  set  $X_{k+1}^w = X_{\text{prop}}^w$ , otherwise set  $X_{k+1}^w = X_k^w$ .
- Final result: the Markov chains  $\{X_0^w, \dots, X_K^w\}$ , for all the walkers  $w = 1, \dots, W$ .

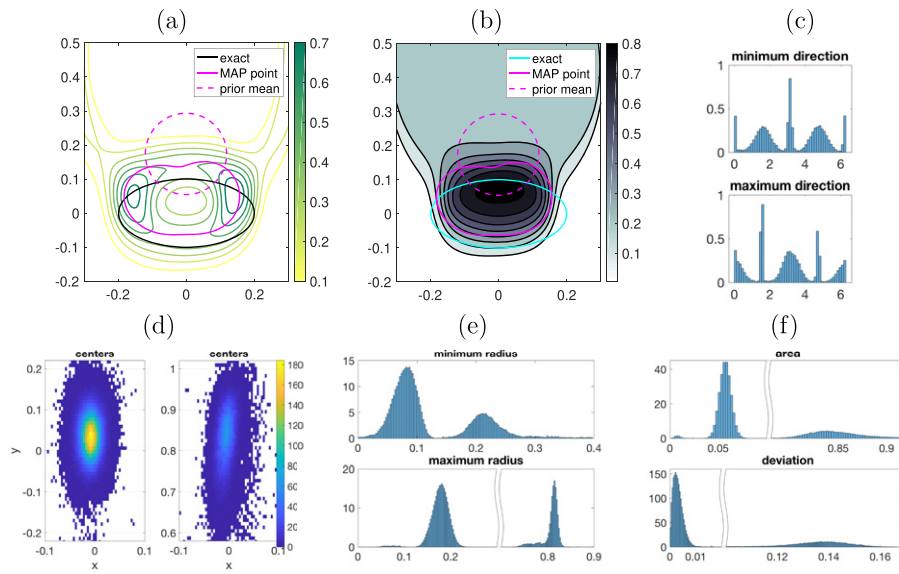
The number of walkers is chosen as at least  $W > 2d$  to increase mixing and to take advantage of parallelization [29].

With this algorithm, we revisit the configurations studied by Bayesian linearized techniques. Figures 6 and 7 are the counterparts of figures 4 and 5, generated by MCMC sampling. We have used Gelman–Rubin tests [31] to check that the sample distributions under study are properly converged. Notice that in both cases we have rejected samples with negative radius. The results are similar but the computational cost is higher, several tens of thousand samples have been used, depending on the shapes and wavenumbers. Each sample requires solving a forward problem to evaluate the likelihood. Magenta curves in this case represent the sample with maximum probability, which is an approximation to the MAP point. The approximation improves increasing the number of samples.

Comparing figures 4 and 6, we see that the approximations to the MAP point are similar. However, there are differences in the lower probability contours around the MAP point for the ellipse oriented orthogonally to the incidence direction and the larger values of  $\kappa$ , see panel (b). These low probability features are converged as when increasing the number of Monte Carlo samples, they remain mostly unchanged. We analyze this phenomenon in more detail in figures 7 and 8 for larger noise. Comparing figures 5 and 7, the MAP and expected curves show again reasonable agreement. However, the Bayesian posterior becomes bimodal, as seen in the



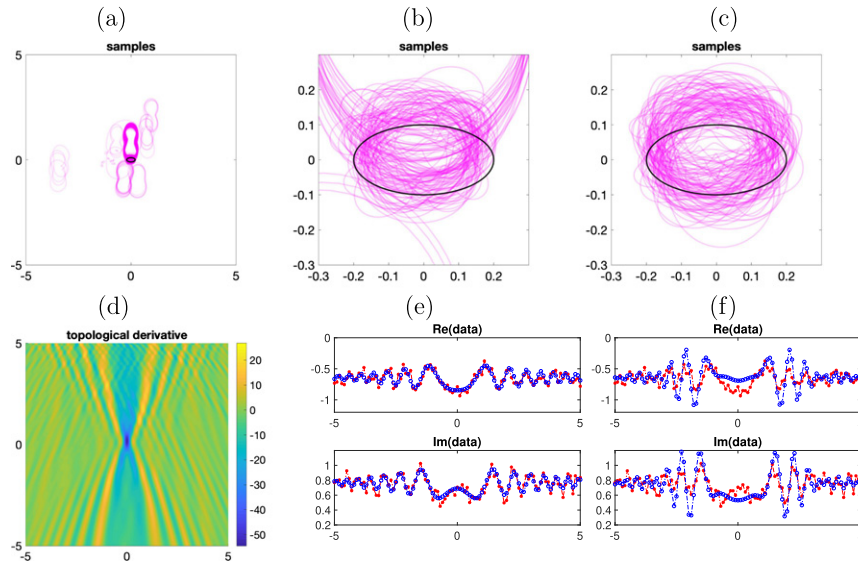
**Figure 6.** Same as figure 4 but with MCMC sampling. Sampling parameters are  $W = 200$ ,  $K = 200$ ,  $B = 35\,000$  and  $a = 2$  in (a), and are increased to  $K = 500$  and  $B = 55\,000$  in (b) and (c).



**Figure 7.** Same as figure 5 but with MCMC sampling. Sampling parameters are  $a = 2$ ,  $W = 800$ ,  $K = 1500$ ,  $B = 80\,000$ . Due to the bi-model nature of the posterior distribution, only parts of the figures are shown. See figure 8 for a study explaining the bimodality of the posterior.

histograms for centers, radii, orientation and area. These observations persist varying the number of walkers  $W = 200, 400, 800$ , steps  $K = 500, 1000, 3000, 7000, 14000$ , and acceptance rates  $a = 2, 1.1, 2.5$ .

To further study this, figure 8(a) displays a collection of samples. Most of them wrap around the true object, as shown in panel (b). However, a significant number of samples is oriented in the incidence direction of the waves, along the  $y$  axis, orthogonal to the true orientation of the object. Tracking the evolution of the initial walkers, we observe that when they are above a certain size, the proposed curves evolve toward this second family, whereas smaller shapes get closer to the true object. Comparing figure 8(d) with Figure 8(a), we notice that samples concentrate in regions of large negative values for the topological derivative. Some concentrate



**Figure 8.** Samples from the problems discussed in figures 5 and 7. Shown in (a) are the last 800 walker states generated by the MCMC algorithm in the imaging region  $[-5, 5] \times [-5, 5]$ , in (b) zoom-ins showing 100 of such samples around the true object, and in (c) 100 samples generated by Bayesian linearized sampling. Using the linearized approach, all samples wrap around the true object (black curve), while samples from the posterior obtained using MCMC show that other shapes are consistent with the data. To study this further, in (d) the topological derivative of the cost functional (12) is shown, and in (e) and (f) a comparison of the data (asterisks) with observations generated from a sample wrapped around the object (circles in (e)) and by one of the large, elongated samples (circles in (f)).

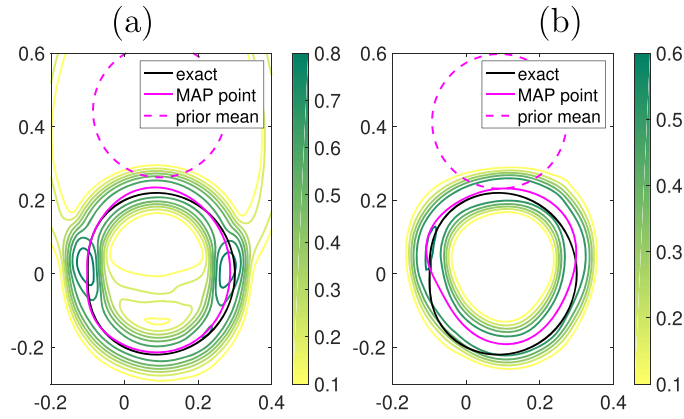
around the spot where largest negative values are attained, marking the true object location, while the other elongate along stripes of less negative values. As commented in section 4.1, we expect the physical cost functional (12) to decrease by removing regions of negative values of the topological derivative, that is, placing objects in them. Panels (d) and (e) compare the synthetic data used in the simulations to the measurements corresponding to samples wrapped around the true object and to elongated samples.

Elongation and loss of axial resolution are aberrations present in traditional holographic reconstruction techniques based on numerical backpropagation [59] due to the use of only one incident direction. They indicate the potential presence of additional local minima in the original cost functional and highlight ambiguity due to ill-posedness, in particular if the data contains larger noise levels.

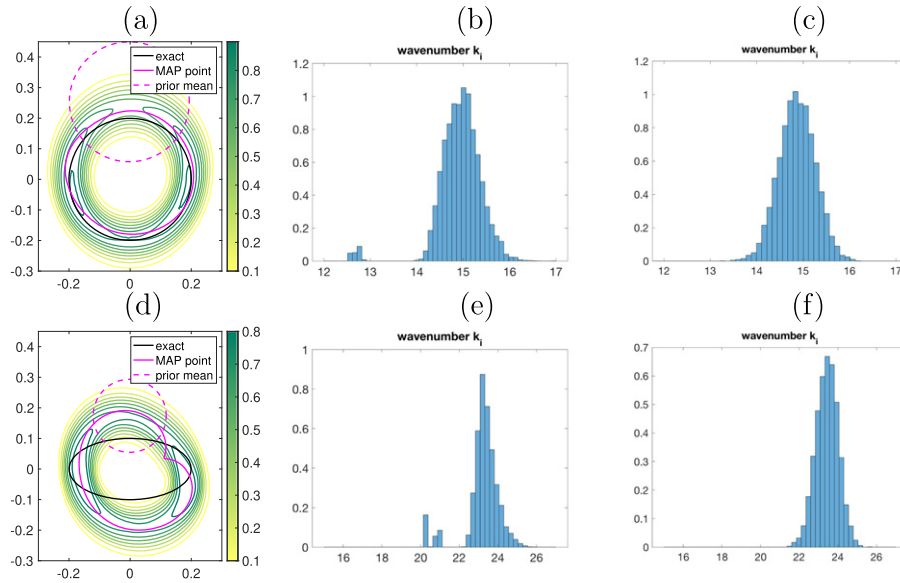
Figure 9 compares the results for a asymmetric egg-like shapes when we replace complex valued data by just intensities. As can be seen, when only using intensities, the uncertainty in the shape increases.

## 7. Varying the material properties

In the previous sections, we fix the object properties and aim at inferring their geometry. In practice, one may have to infer also material constants entering the equations such as  $\kappa_i$ . In



**Figure 9.** Results for an egg-like object inferred from complex data  $u(\mathbf{x}_j)$  (a), and from the modulus data  $|u(\mathbf{x}_j)|^2$  (b),  $j = 1, \dots, 101$ , with 5% noise. Parameters are  $\kappa_i = 24.79$ ,  $\kappa_e = 20.6$ ,  $\beta = 1$  and  $M = 5$ . Sampling parameters are  $a = 2$ ,  $W = 200$  and  $K = 3000$ .



**Figure 10.** (a) Same as figure 3(b) but with MCMC sampling for unknown  $\kappa_i$ , shown in (b). Shown in (d) is the counterpart of figure 4(b) using intensity data with MCMC sampling for unknown  $\kappa_i$ , shown in (e). (c) and (f) are the equivalent of (b) and (e) obtained by linearizing about the MAP point. We use the same MCMC sampling parameters as for the previous figures. The true values for  $\kappa_i$  are  $\kappa_i = 15.12$  in (a)–(c) and  $\kappa_i = 25.12$  in (d)–(f). Expected values are  $\kappa_i = 15.04$  and  $\kappa_i = 23.18$ , respectively.

the MCMC framework, this can be handled by adding an additional element to the prior. The forward problem has one parameter in addition to those defining the shapes, which enters the equation. Hence, we sample with respect to one more parameter. The results are shown in figure 10.

In the Bayesian linearized framework, we also include a new parameter in the prior. When optimizing, we simply compute derivatives with respect to one parameter more. The Fréchet derivative with respect to  $\kappa_i$  is then the solution of a boundary value problem obtained differentiating the Helmholtz equations with respect to  $\kappa_i$ , see appendix A2. Once this is done, a similar iteration to that proposed in section 5.1 provides a MAP estimate. Sampling by means of (16) we obtain figures 10(c) and (f).

## 8. Model selection when the number of objects is unknown

In the previous sections, we inferred object shapes from observational data assuming that the number of objects is known. Here, we are considering the number of objects  $L$  as an additional variable to be inferred. The number of objects is substantially different from other variables that characterize the object geometry because it is discrete and changing it amounts to the addition or removal of blocks of parameters defining object components. Thus, the number of objects defines different models with different numbers of parameters [15]. The probability of the data given the model, that is, the number of objects  $m$ , is computed by integrating over the model parameters:

$$p(\mathbf{d}|m) = \int_{\nu} p(\mathbf{d}|\nu)p(\nu|m)d\nu, \quad (18)$$

where  $p(\mathbf{d}|\nu)$  is the likelihood (8) and  $p(\nu|m)$  the prior (17) given the model. It can be evaluated by sampling one of them,  $p(\nu|m)$  for instance, and computing a Monte Carlo estimate based on these samples  $\nu_i$ ,  $i = 1, \dots, S$ , amounting to  $\frac{1}{S} \sum_{i=1}^S p(\mathbf{d}|\nu_i)$ , [3]. The resulting quantity is called the evidence for model  $m$  [53]. We have implemented this procedure for the synthetic observations coming from the configuration with 3 objects shown in figure 2. Computing the evidence for models with  $m = 1, 2, 3, 4$ , we find a clear maximum for  $m = 3$ , i.e., the true number of objects. Once we have selected a number of objects, we sample the posterior distribution (9) using the MCMC techniques described in section 6 to infer the expected objects, which resemble the MAP points in figure 2(c).

## 9. Summary and conclusions

We have developed a Bayesian framework for object detection which uses topological methods to generate priors. In this approach, objects are represented by star-shaped parameterizations. Assuming the number of objects and their material properties are known, we compute the ‘maximum a posteriori’ (MAP) estimate for the parameters defining centers and radii through minimization of the proper cost functional. Linearizing the parameter-to-observable map about the MAP point, one can generate samples of the Laplace approximation to the posterior distribution to quantify the uncertainty in the object location and its shape at a low computational cost. We test the scheme in 2D holography imaging set-ups, for both acoustic and light waves. The former uses complex fields measured at detectors as data, whereas in the latter only real intensities are available. In these set-ups, wave fields are governed by transmission boundary problems for Helmholtz equations and the incident waves reduce to a single beam. For small noise magnitudes, many shapes can be inferred with moderate uncertainty for a wide range of wavenumbers. As the magnitude of the noise increases, larger wavenumbers (i.e., smaller frequencies) provide better shape descriptions whereas smaller wavenumbers approximate their location more accurately. Even for noise magnitudes of 10% and higher, topological derivatives

are useful to generate priors, and the Bayesian approach with complex data results in parameter inference with moderate uncertainties. This is particularly relevant for acoustic imaging set-ups, in which the magnitude of the noise is larger.

To assess the performance of Bayesian linearized methods, i.e., Laplace posterior approximations, we compared the results with those obtained by direct MCMC sampling of the full posterior distribution, finding reasonable agreement. However, MCMC sampling provides more complete insight into the structure of the posterior distribution, which become multimodal for larger noise levels. Comparisons are made under the assumption of piecewise constant material properties, which allows us to use fast boundary element schemes to solve forward and Fréchet problems. Also, working with 2D star-shaped objects represented in a trigonometric basis, we only need to infer 14–50 parameters for sets of 1–3 objects. When the number of objects is also unknown, we used a Bayesian model selection strategy to obtain information on the number of objects, as well as on their location and shapes.

Our methods extend to 3D set-ups, involving a larger number of parameters. Whereas Bayesian linearized techniques may yet be efficient once the derivatives needed to implement optimization techniques are characterized as solutions of specific boundary problems, the computational cost of direct MCMC sampling increases. Here, computing a MAP point requires 15–25 Newton-type descent steps. Each iteration implies solving a small linear system, as well as boundary value problems for the derivatives. Instead, MCMC requires solving a large amount of boundary value problems. The problem becomes even more difficult if we allow for spatial variations in the material parameters. Tempering approaches [19] may help to reduce that cost.

Topological sensitivity is a flexible tool to obtain prior information. It uses an explicit formula which accommodates limited aperture data, only one incident wave, large noise, and measurements of the full complex field, or functions of it, such as intensities. Moreover, it has an interpretation as a measure of the decrease of the physical cost functional which provides insight into the multimodality of the posterior. However, the proposed Bayesian framework is independent of the choice of the prior. Other initialization procedures, such as linear or direct sampling [8, 39], may be used to construct priors when the combination of receivers and incident directions provides complex amplitude data widely distributed around the object.

Here, we mostly consider tests in which the material properties are considered known. Assuming material properties to be characterized by constant parameters, the proposed framework can be adapted by including a few additional parameters. To consider more general spatially variable material properties, one could combine these methods with those developed in [6, 58] and implement coupled BEM–FEM or spectral-FEM solvers as in [12, 13]. The methods would extend to wave fields governed by systems different from Helmholtz equations, provided characterizations for the derivatives and adequate solvers for the boundary value problems involved are available. Finally, note that the objects in our tests are assumed to be stationary. Time-dependent Bayesian methods to track moving contours are discussed in [55].

## Acknowledgments

A Carpio and S Iakunin acknowledge partial support from the FEDER/MICINN-AEI grant MTM2017-84446-C2-1-R. G Stadler acknowledges partial support from the US National Science Foundation Grant #1723211 and from KAUST under Award #OSR-2018-CARF-3666. A Carpio thanks RE Caflisch for hospitality during a sabbatical stay at the Courant Institute, NYU, and DG Grier for an introduction to acoustic holography.

## Appendix A. Characterization of Fréchet derivatives

In this appendix, we study Fréchet derivatives for transmission problems of the form (1), where the total wavefield  $u = u_{\text{inc}} + u_{\text{sc}}$  in the exterior region  $\Omega_e = \mathbb{R}^2 \setminus \bar{\Omega}_i$  and the transmitted wave field  $u = u_{\text{tr}}$  in a bounded smooth inclusion  $\Omega_i$ . We require Fréchet derivatives of the solutions with respect to the parameters defining  $\Omega_i$  and with respect to  $\kappa_i$ . Fréchet derivatives with respect to the domain are calculated using integral equations in [36, 38]. We recall these expressions here and give a different proof by variational methods. Similar, but simpler, variational arguments yield expressions for the Fréchet derivatives with respect to problem coefficients, for similar arguments see [11, 58]. Our summary also includes remarks on the relation to shape derivatives, as well as computational details.

To simplify, the parameters  $\kappa_e, \kappa_i, \beta$  are taken to be constant, real and positive. General conditions on the parameters  $\kappa_e, \kappa_i, \beta$  guaranteeing existence and uniqueness of a solution  $u \in H_{\text{loc}}^1(\mathbb{R}^2)$  for this problem can be found in [16, 44, 45]. When  $\partial\Omega_i \in C^2$ ,  $u$  is in  $H^2(\Omega_i) \cup H_{\text{loc}}^2(\Omega_e)$ , which ensures continuity away from the interface.

Let us recall that given two Banach spaces  $X, Y$  and a function  $\mathcal{F}: D(\mathcal{F}) \subset X \rightarrow Y$ , its Fréchet derivative  $\mathcal{F}': X \rightarrow Y$  is a linear bounded operator satisfying  $\mathcal{F}(x + \xi) = \mathcal{F}(x) + \mathcal{F}'(x)\xi + o(\xi)$  for  $\xi \in X$  as  $\|\xi\|_X \rightarrow 0$ , for any  $x \in X$ . In terms of the directional Gateaux derivative  $\mathcal{F}'(x)\xi = D_\xi \mathcal{F}(x)$  with  $D_\xi \mathcal{F}(x) = \lim_{\tau \rightarrow 0} \frac{\mathcal{F}(x + \tau\xi) - \mathcal{F}(x)}{\tau}$ . In our context, Fréchet derivatives can be characterized as solutions of adequate boundary value problems.

### A1. Fréchet derivative with respect to the domain

We consider variable domains  $\Omega_i = \Omega^t$ , whose boundaries  $\Gamma^t$  are generated from a smooth curve  $\Gamma^0 \in C^2$  (twice differentiable) following a family of deformations  $\Gamma^t = \{\mathbf{x} + t\mathbf{V}(\mathbf{x}) \mid \mathbf{x} \in \Gamma^0\}$ , along a smooth vector field  $\mathbf{V} \in C^2(\Gamma^0)$ . The solutions of (1) with  $\Omega_i = \Omega^t$  are denoted by  $u^t$ . For small  $t > 0$ ,  $\Gamma^t \in C^2$  is a perturbation of  $\Gamma^0$ . The deformation  $\mathbf{x}^t = \phi^t(\mathbf{x}) = \mathbf{x} + t\mathbf{V}(\mathbf{x})$  maps  $\Omega^0$  to  $\Omega^t$ . For small  $t$ ,  $\phi^t$  is a diffeomorphism and its inverse  $\eta^t$  maps  $\Omega^t$  to  $\Omega^0$ . We extend  $\mathbf{V}$  to  $\mathbb{R}^2$  in such a way that it decays fast away from  $\Gamma^0$ , while preserving the same regularity.

The operator  $\mathcal{F}$  that assigns to  $\mathbf{V} \in C^1(\partial\Omega_i)$  the far field values  $(u(\mathbf{x}_j))_{j=1}^N$  of the solution of the forward problem (1) is Fréchet differentiable with derivative  $\langle \mathcal{F}'(\partial\Omega_i), \mathbf{V} \rangle = (v(\mathbf{x}_j))_{j=1}^N$ , where  $v$  is the solution of

$$\begin{aligned} \Delta v + \kappa_e^2 v &= 0 \quad \text{in } \Omega_e, \quad \Delta v + \kappa_i^2 v = 0 \quad \text{in } \Omega_i, \\ v^- - v^+ &= -(\mathbf{V} \cdot \mathbf{n}) \left( \frac{\partial u^-}{\partial \mathbf{n}} - \frac{\partial u^+}{\partial \mathbf{n}} \right) \quad \text{on } \partial\Omega_i, \\ \beta \frac{\partial v^-}{\partial \mathbf{n}} - \frac{\partial v^+}{\partial \mathbf{n}} &= \frac{d}{ds} \left[ (\mathbf{V} \cdot \mathbf{n}) \frac{d}{ds} (\beta u^- - u^+) \right] + (\mathbf{V} \cdot \mathbf{n}) (\beta \kappa_i^2 u^- - \kappa_e^2 u^+) \quad \text{on } \partial\Omega_i, \\ \lim_{|\mathbf{x}| \rightarrow 0} |\mathbf{x}|^{1/2} \left( \frac{\partial}{\partial |\mathbf{x}|} v - \kappa_e v \right) &= 0, \end{aligned} \tag{A.1}$$

and  $\frac{d}{ds}$  is the derivative with respect to the arclength, as proven in [36, 38].

In practice, the forward (1) and Fréchet (A.1) problems are discretized and the solution operators they induce are represented by matrices. We use the boundary element formulations introduced in [21, 22] to approximate the solutions. To do so, we recast (1) as a boundary value problem for the scattered and the transmitted fields with transmission conditions  $u_{\text{tr}} - u_{\text{sc}} = u_{\text{inc}}$  and  $\beta \partial_{\mathbf{n}} u_{\text{tr}} - \partial_{\mathbf{n}} u_{\text{sc}} = \partial_{\mathbf{n}} u_{\text{inc}}$ .

Note that the difference between two close domains  $\Omega_\nu$  and  $\Omega_{\nu+\xi}$  parameterized by (4) and (5) defines a vector perturbation  $\mathbf{V}$  of the boundary in terms of the difference of their parameterizations, which can be written in the form with (4) and (5) with parameters  $\xi$  [14]. Thus, the action of the Fréchet derivative is identified with the action of a matrix on a vector, that is,  $\langle \mathcal{F}'(\partial\Omega_\nu), \mathbf{V} \rangle$  becomes  $\mathbf{F}(\nu)\xi \sim (v(\mathbf{x}_j))_{j=1}^N$ . In this way, we compute the matrices  $\mathbf{F}(\nu)$  employed in section 5.1.

The characterization (A.1) is proven using integral equations in [36, 38] for  $\mathbf{V} \in C^1$ . We give an alternative proof here using a variational approach inspired by the work in [48] for 2D exterior elasticity problems with zero Dirichlet boundary conditions on a moving boundary, which clarifies the role of the transmission conditions.

**Theorem 1.** *Keeping the previous notations and assumptions, the ‘Fréchet derivative’ of the far field of the solution  $u$  of (1) with respect to the domain is given by the far field of the solution of the boundary value problem (A.1).*

**Proof.** The proof proceeds in the following steps.

Step 1. *Variational formulation.* Firstly, we reformulate the transmission problem (1) as an equivalent boundary value problem posed in a bounded domain. Let  $\Gamma_R$  be a large circle which encloses the objects  $\Omega_i$ . The Dirichlet-to-Neumann operator [27, 41] associates to any Dirichlet data on  $\Gamma_R$  the normal derivative of the solution of the exterior Dirichlet problem:  $L : H^{1/2}(\Gamma_R) \rightarrow H^{-1/2}(\Gamma_R)$ ,  $f \mapsto L(f) = \partial_n v$ , where  $v \in H_{\text{loc}}^1(\mathbb{R}^2 \setminus \overline{B_R})$ ,  $B_R := B(\mathbf{0}, R)$ , is the unique solution of

$$\begin{cases} \Delta v + \kappa_e^2 v = 0 & \text{in } \mathbb{R}^2 \setminus \overline{B_R}, \quad v = f \quad \text{on } \Gamma_R, \\ \lim_{r \rightarrow \infty} r^{1/2}(\partial_r v - i\kappa_e v) = 0. \end{cases}$$

In the sequel,  $\partial_n$  stands for the normal derivative at the interface and  $\partial_r$  denotes the radial derivative.  $H_{\text{loc}}^1(\mathbb{R}^2 \setminus \overline{B_R})$  is the usual Sobolev space and  $H^{1/2}(\Gamma_R)$  and  $H^{-1/2}(\Gamma_R)$  denotes the standard trace spaces. We replace (1) by an equivalent boundary value problem set in  $B_R$  with a non-reflecting boundary condition on  $\Gamma_R$  defined by the Dirichlet-to-Neuman operator:

$$\begin{cases} \Delta u + \kappa_e^2 u = 0 & \text{in } \Omega'_e := B_R \setminus \overline{\Omega_i}, \quad \beta \Delta u + \beta \kappa_i^2 u = 0 & \text{in } \Omega_i, \\ u^- - u^+ = 0 & \text{on } \partial\Omega_i, \quad \beta \partial_n u^- - \partial_n u^+ = 0 & \text{on } \partial\Omega_i, \\ \partial_n(u - u_{\text{inc}}) = L(u - u_{\text{inc}}) & \text{on } \Gamma_R. \end{cases} \quad (\text{A.2})$$

Unit normals are exterior to  $\partial\Omega_i$  and to  $\Gamma_R$ . The solution  $u$  of (A.2) also solves the variational problem

$$\begin{aligned} u &\in H^1(B_R), \quad b(\Omega_i; u, w) = \ell(w), \quad \forall w \in H^1(B_R), \\ b(\Omega_i; u, w) &= \int_{\Omega'_e} (\nabla u \nabla \overline{w} - \kappa_e^2 u \overline{w}) d\mathbf{x} + \int_{\Omega_i} (\beta \nabla u \nabla \overline{w} - \beta \kappa_i^2 u \overline{w}) d\mathbf{x} \\ &\quad - \int_{\Gamma_R} L u \overline{w} dS_{\mathbf{x}}, \quad \forall u, w \in H^1(B_R), \\ \ell(w) &= \int_{\Gamma_R} (\partial_n u_{\text{inc}} - L u_{\text{inc}}) \overline{w} dS_{\mathbf{x}}, \quad \forall w \in H^1(B_R). \end{aligned} \quad (\text{A.3})$$

Step 2. *Deformed problems.* Since  $\mathbf{V}$  decreases rapidly to zero away from  $\Gamma^0$ ,  $\phi^t(\Gamma_R) = \Gamma_R$  and  $\phi^t(\mathbf{x}_j) = \mathbf{x}_j$ ,  $j = 1, \dots, N$ . Setting in (A.3)  $\Omega_i = \phi^t(\Omega^0) = \Omega^t$  and  $\Omega'_e = \phi^t(\Omega'_e) = B_R \setminus \overline{\Omega^t}$ ,

the variational reformulations in the deformed domains take the form: Find  $u^t \in H^1(B_R)$  such that

$$\begin{aligned} b^t(\Omega^t; u^t, w) &= \ell(w), \quad \forall w \in H^1(B_R), \\ b^t(\Omega^t; u, w) &= \int_{B_R \setminus \bar{\Omega}^t} (\nabla_{\mathbf{x}^t} u \nabla_{\mathbf{x}^t} \bar{w} - \kappa_e^2 u \bar{w}) d\mathbf{x}^t - \int_{\Gamma_R} L u \bar{w} dS_{\mathbf{x}} \\ &\quad + \int_{\Omega^t} (\beta \nabla_{\mathbf{x}^t} u \nabla_{\mathbf{x}^t} \bar{w} - \beta \kappa_i^2 u \bar{w}) d\mathbf{x}^t, \quad \forall u, w \in H^1(B_R). \end{aligned} \quad (\text{A.4})$$

Step 3. *Change variables to initial configuration.* We now transform all the quantities appearing in (A.4) back to the initial configuration  $\Omega^0$ . The process is similar to transforming deformed configurations back to a reference configuration in continuum mechanics. The deformation gradient is the Jacobian of the change of variables

$$\mathbf{J}^t(\mathbf{x}) = \nabla_{\mathbf{x}} \phi^t(\mathbf{x}) = \left( \frac{\partial x_i^t}{\partial x_j}(\mathbf{x}) \right) = \mathbf{I} + t \nabla \mathbf{V}(\mathbf{x}), \quad (\text{A.5})$$

and its inverse  $(\mathbf{J}^t)^{-1} = \left( \frac{\partial x_i}{\partial x_j^t} \right)$  is the Jacobian of the inverse change of variables. Then, volume and surface elements are related by

$$d\mathbf{x}^t = \det \mathbf{J}^t(\mathbf{x}) d\mathbf{x}, \quad dS_{\mathbf{x}^t} = \det \mathbf{J}^t(\mathbf{x}) \|(\mathbf{J}^t(\mathbf{x}))^{-T} \mathbf{n}\| dS_{\mathbf{x}}, \quad (\text{A.6})$$

and the chain rule for derivatives reads  $\nabla_{\mathbf{x}} u(\mathbf{x}^t(\mathbf{x})) = (\mathbf{J}^t(\mathbf{x}))^T \nabla_{\mathbf{x}^t} u(\mathbf{x}^t(\mathbf{x}))$ , that is,  $\nabla_{\mathbf{x}^t} u = (\mathbf{J}^t)^{-T} \nabla_{\mathbf{x}} u$ . For each component we have

$$\frac{\partial u}{\partial x_{\alpha}^t}(\mathbf{x}^t(\mathbf{x})) = \frac{\partial u}{\partial x_k}(\mathbf{x}^t(\mathbf{x})) (J^t)^{-1}_{k\beta}(\mathbf{x}). \quad (\text{A.7})$$

Defining  $\hat{u}(\mathbf{x}) = u^t \circ \phi^t(\mathbf{x}) = u^t(\mathbf{x}^t(\mathbf{x}))$ , changing variables and using (A.6) and (A.7) yields:

$$\begin{aligned} b_i^t(\Omega^t; u^t, w) &= \int_{\Omega^t} \left[ \beta \frac{\partial u^t}{\partial x_{\alpha}^t}(\mathbf{x}^t) \frac{\partial \bar{w}}{\partial x_{\alpha}^t}(\mathbf{x}^t) - \beta \kappa_i^2 u^t(\mathbf{x}^t) \bar{w}(\mathbf{x}^t) \right] d\mathbf{x}^t \\ &= \int_{\Omega^0} \beta \left[ \frac{\partial \hat{u}}{\partial x_p}(\mathbf{x}) (J^t)^{-1}_{p\alpha}(\mathbf{x}) \frac{\partial \hat{\bar{w}}}{\partial x_q}(\mathbf{x}) (J^t)^{-1}_{q\alpha}(\mathbf{x}) - \beta \kappa_i^2 \hat{u}(\mathbf{x}) \hat{\bar{w}}(\mathbf{x}) \right] \\ &\quad \times \det \mathbf{J}^t(\mathbf{x}) d\mathbf{x} = \hat{b}_i^t(\Omega^0; \hat{u}, \hat{w}). \end{aligned} \quad (\text{A.8})$$

A similar relation holds on  $B_R \setminus \bar{\Omega}^t$  defining  $b_e^t(B_R \setminus \bar{\Omega}^t; u^t, w) = \hat{b}_e^t(B_R \setminus \bar{\Omega}^0; \hat{u}, \hat{w})$ . For  $w \in H^1(B_R)$ , we have  $\hat{w} \in H^1(B_R)$ . Therefore, we obtain the equivalent variational formulation: Find  $\hat{u} \in H^1(B_R)$  such that

$$\hat{b}^t(\Omega^0; \hat{u}, w) = \hat{b}_i^t(\Omega^0; \hat{u}, w) + \hat{b}_e^t(B_R \setminus \bar{\Omega}^0; \hat{u}, w) - \int_{\Gamma_R} L \hat{u} \bar{w} dS_{\mathbf{x}} = \ell(w),$$

for  $w \in H^1(B_R)$ . Let us analyze the dependence on  $t$  of the terms appearing in  $\hat{b}_i^t$  and  $\hat{b}_e^t$ . From the definitions of the Jacobian matrices (A.5) we obtain [27, 48]

$$\det \mathbf{J}^t(\mathbf{x}) = 1 + t \operatorname{div}(\mathbf{V}(\mathbf{x})) + O(t^2), \quad (\mathbf{J}^t)^{-1}(\mathbf{x}) = \mathbf{I} - t \nabla \mathbf{V}(\mathbf{x}) + O(t^2). \quad (\text{A.9})$$

Inserting (A.9) in (A.8) we find

$$\ell(w) = \hat{b}'(\Omega^0; \hat{u}, w) = b^0(\Omega^0; \hat{u}, w) + t[I_1(\hat{u}) + I_2(\hat{u}) + I_3(\hat{u})] + O(t^2), \quad (\text{A.10})$$

where

$$\begin{aligned} b^0(\Omega^0; \hat{u}, w) &= \int_{\Omega^0} \left[ \beta \frac{\partial \hat{u}}{\partial x_\alpha} \frac{\partial \bar{w}}{\partial x_\alpha} - \beta \kappa_i^2 \hat{u} \bar{w} \right] \mathrm{d}\mathbf{x} + \int_{B_R \setminus \bar{\Omega}^0} \left[ \frac{\partial \hat{u}}{\partial x_\alpha} \frac{\partial \bar{w}}{\partial x_\alpha} - \kappa_e^2 \hat{u} \bar{w} \right] \mathrm{d}\mathbf{x} \\ &\quad - \int_{\Gamma_R} L \hat{u} \bar{w} \mathrm{d}S_{\mathbf{x}}, \end{aligned} \quad (\text{A.11})$$

$$\begin{aligned} I_1(\hat{u}) &= \int_{\Omega^0} \left[ \beta \frac{\partial \hat{u}}{\partial x_\alpha} \frac{\partial \bar{w}}{\partial x_\alpha} - \beta \kappa_i^2 \hat{u} \bar{w} \right] \frac{\partial V_p}{\partial x_p} \mathrm{d}\mathbf{x} + \int_{B_R \setminus \bar{\Omega}^0} \left[ \frac{\partial \hat{u}}{\partial x_\alpha} \frac{\partial \bar{w}}{\partial x_\alpha} - \kappa_e^2 \hat{u} \bar{w} \right] \frac{\partial V_p}{\partial x_p} \mathrm{d}\mathbf{x}, \\ I_2(\hat{u}) &= - \int_{\Omega^0} \beta \frac{\partial \hat{u}}{\partial x_p} \frac{\partial \bar{w}}{\partial x_\alpha} \frac{\partial V_p}{\partial x_\alpha} \mathrm{d}\mathbf{x} - \int_{B_R \setminus \bar{\Omega}^0} \frac{\partial \hat{u}}{\partial x_p} \frac{\partial \bar{w}}{\partial x_\alpha} \frac{\partial V_p}{\partial x_\alpha} \mathrm{d}\mathbf{x}, \\ I_3(\hat{u}) &= - \int_{\Omega^0} \beta \frac{\partial \hat{u}}{\partial x_\alpha} \frac{\partial \bar{w}}{\partial x_q} \frac{\partial V_q}{\partial x_\alpha} \mathrm{d}\mathbf{x} - \int_{B_R \setminus \bar{\Omega}^0} \frac{\partial \hat{u}}{\partial x_\alpha} \frac{\partial \bar{w}}{\partial x_q} \frac{\partial V_q}{\partial x_\alpha} \mathrm{d}\mathbf{x}. \end{aligned} \quad (\text{A.12})$$

Step 4. *Variational problem for the domain derivative  $u'$ .* Let us compare the transformed function  $\hat{u}$  and the solution  $u^0$  of  $b^0(\Omega^0; u^0, w) = \ell(w)$ . Thanks to (A.10), for any  $w \in H^1(B_R)$  we have

$$b^0(\Omega^0; \hat{u} - u^0, w) = -t[I_1(\hat{u}) + I_2(\hat{u}) + I_3(\hat{u})]. \quad (\text{A.13})$$

Well posedness of the variational problems (A.2) with respect to changes in domains  $\Omega^t$  implies uniform bounds on the solutions for  $t \in [0, T]$ :  $\|u^t\|_{H^1(B_R)} \leq C(T)$ ,  $\|\hat{u}\|_{H^1(B_R)} \leq C(T)$ . The right hand side in (A.13) tends to zero as  $t \rightarrow 0$ . Well posedness of the variational problem again implies  $\hat{u} \rightarrow u^0$  in  $H^1(B_R)$ .

Dividing (A.13) by  $t$ , we find  $b^0(\Omega^0; \frac{\hat{u} - u^0}{t}, w) = -[I_1(\hat{u}) + I_2(\hat{u}) + I_3(\hat{u})]$ . Then, the limit  $\dot{u} = \lim_{t \rightarrow 0} \frac{\hat{u} - u^0}{t}$  satisfies

$$b^0(\Omega^0; \dot{u}, w) = -[I_1(u^0) + I_2(u^0) + I_3(u^0)]. \quad (\text{A.14})$$

The function  $\dot{u}$  is the so called ‘material derivative’, that is,  $\dot{u} = \frac{\partial u}{\partial t} + \mathbf{V} \cdot \nabla u^0$ . The domain derivative is then  $u' = \dot{u} - \mathbf{V} \cdot \nabla u^0$ . Then,

$$b^0(\Omega^0; u', w) = b^0(\Omega^0; \dot{u}, w) - b^0(\Omega^0; \mathbf{V} \cdot \nabla u^0, w), \quad (\text{A.15})$$

where  $b^0(\Omega^0; \mathbf{V} \cdot \nabla u^0, w)$  is obtained from (A.11) replacing  $\hat{u}$  by  $\frac{\partial u^0}{\partial x_p} V_p$ .

Step 5. *Differential equation for  $u'$ .* We evaluate the terms in the right hand side of (A.14) to calculate the right hand side in (A.15) using (A.12). Recalling the equations for  $u^0$  and that  $\mathbf{V}$  vanishes at  $\Gamma_R$ , (A.15) becomes after integrating by parts

$$\begin{aligned} b^0(\Omega^0; u', w) &= \int_{\Gamma^0} \left( \beta \frac{\partial (u^0)^-}{\partial x_\alpha} - \frac{\partial (u^0)^+}{\partial x_\alpha} \right) \frac{\partial \bar{w}}{\partial x_q} V_q n_\alpha \mathrm{d}S_{\mathbf{x}} \\ &\quad - \int_{\Gamma^0} \left( \beta \frac{\partial (u^0)^-}{\partial x_\alpha} - \frac{\partial (u^0)^+}{\partial x_\alpha} \right) \frac{\partial \bar{w}}{\partial x_\alpha} V_p n_p \mathrm{d}\mathbf{x} + \int_{\Gamma^0} (\beta \kappa_i^2 (u^0)^- - \kappa_e^2 (u^0)^+) \bar{w} V_q n_q \mathrm{d}\mathbf{x}. \end{aligned} \quad (\text{A.16})$$

Integrating by parts in  $b^0(\Omega^0; u', w)$  for  $w$  vanishing on  $\Gamma^0$  this identity yields the equation (A.1) for  $u'$ :  $\Delta u' + \kappa_e^2 u' = 0$  in  $B_R \setminus \overline{\Omega^0}$ ,  $\beta \Delta u' + \beta \kappa_i^2 u' = 0$  in  $\Omega^0$ , as well as the radiating boundary condition at infinity.

Step 6. *Transmission conditions for  $u'$ .* When  $\beta = 1$ ,  $u^0$  and its derivatives  $\nabla u^0$  are continuous across  $\Gamma^0$ . As a result,  $u'$  is continuous across  $\Gamma^0$  and it belongs to  $H_{\text{loc}}^1(\mathbb{R}^2)$ . Integrating by parts in (A.16) we get

$$\beta \frac{\partial(u')^-}{\partial \mathbf{n}} - \frac{\partial(u')^+}{\partial \mathbf{n}} = (\beta \kappa_i^2 (u^0)^- - \kappa_e^2 (u^0)^+) (\mathbf{V} \cdot \mathbf{n}) \quad \text{on } \Gamma^0.$$

When  $\beta \neq 1$ ,  $\nabla u^0$  is not longer continuous across  $\Gamma^0$ . Denoting by  $\mathbf{t}$  and  $\mathbf{n}$  the unit tangent and normal vectors on  $\Gamma^0$ , we have  $\nabla(u^0)^\pm = (\nabla(u^0)^\pm \cdot \mathbf{t})\mathbf{t} + (\nabla(u^0)^\pm \cdot \mathbf{n})\mathbf{n}$ . The relation  $u' = \dot{u} - \mathbf{V} \cdot \nabla u^0$  gives the jump at the interface:

$$(u')^- - (u')^+ = -(\mathbf{V} \cdot \mathbf{n}) \left( \frac{\partial(u^0)^-}{\partial \mathbf{n}} - \frac{\partial(u^0)^+}{\partial \mathbf{n}} \right).$$

Notice that  $u^0$  being continuous across  $\Gamma^0$ , the tangent derivatives  $\frac{\partial u^0}{\partial \mathbf{t}} = \nabla u^0 \cdot \mathbf{t}$  too. To obtain a transmission condition for the derivatives of  $u'$  at the interface  $\Gamma^0$  we revisit (A.16). The first term vanishes due to the transmission boundary conditions satisfied by  $u^0$  at  $\Gamma^0$ . The second term can be rewritten as

$$-\int_{\Gamma^0} \left( \beta \frac{\partial(u^0)^-}{\partial \mathbf{n}} - \frac{\partial(u^0)^+}{\partial \mathbf{n}} \right) \frac{\partial \bar{w}}{\partial \mathbf{n}} V_p n_p d\mathbf{x} - \int_{\Gamma^0} \left( \beta \frac{\partial(u^0)^-}{\partial \mathbf{t}} - \frac{\partial(u^0)^+}{\partial \mathbf{t}} \right) \frac{\partial \bar{w}}{\partial \mathbf{t}} V_p n_p d\mathbf{x}.$$

The first integral vanished again due to the transmission boundary conditions satisfied by  $u^0$ , whereas the second one, together with the third integral on the right hand side of (A.16) provides the transmission boundary condition in (A.1).  $\square$

Given the characterization of the Fréchet derivative (A.1), we find an expression for the shape derivative of the cost (12) along a smooth vector field  $\mathbf{V}$ , defined as  $DJ_c(\Omega_i) \cdot \mathbf{V} = \frac{d}{dt} J_c(\phi^t(\Omega_i))|_{t=0}$ , keeping the previous notations and following [12, 13]:

$$\begin{aligned} \langle DJ_c(\mathbb{R}^2 \setminus \overline{\Omega_i}), \mathbf{V} \rangle &= \text{Re} \left[ \int_{\partial \Omega_i} ((1 - \beta) (\beta \partial_{\mathbf{n}} u^- \partial_{\mathbf{n}} \bar{p}^- + \partial_{\mathbf{t}} u^- \partial_{\mathbf{t}} \bar{p}^-) \right. \\ &\quad \left. + (\beta \kappa_i^2 - \kappa_e^2) u \bar{p}) \mathbf{V} \cdot \mathbf{n} dS_{\mathbf{x}} \right], \end{aligned} \quad (\text{A.17})$$

where  $u$  and  $\bar{p}$  solve the forward and adjoint problems (1) and (A.18)

$$\begin{cases} \Delta \bar{p} + \kappa_e^2 \bar{p} = \sum_{j=1}^N \chi(\mathbf{x}_j) \delta_{\mathbf{x}_j} & \text{in } \mathbb{R}^2 \setminus \overline{\Omega_i}, \quad \Delta \bar{p} + \kappa_i^2 \bar{p} = 0 & \text{in } \Omega_i, \\ \bar{p}^- - \bar{p}^+ = 0 & \text{on } \partial \Omega_i, \quad \beta \partial_{\mathbf{n}} \bar{p}^- - \partial_{\mathbf{n}} \bar{p}^+ = 0 & \text{on } \partial \Omega_i, \\ \lim_{r \rightarrow \infty} r^{1/2} (\partial_r \bar{p} - i k_e \bar{p}) = 0, \end{cases} \quad (\text{A.18})$$

$$\chi(u(\mathbf{x}_j)) = \begin{cases} \overline{d_j - u(\mathbf{x}_j)}, & \text{for } f(u) = u, \\ 2(d_j - |u(\mathbf{x}_j)|^2) \overline{u(\mathbf{x}_j)}, & \text{for } f(u) = |u|^2, \end{cases} \quad (\text{A.19})$$

with  $r = |\mathbf{x}|$ . This expression agrees with that established in [10, 12] when  $\mathbf{V} = V_n \mathbf{n}$ . The same proofs hold for general fields  $\mathbf{V}$  keeping track of the terms involving tangential components

and the transmission boundary conditions. Writing  $\bar{p} = \bar{p}' + \bar{p}_{\text{inc}}$ , where  $\bar{p}_{\text{inc}}$  is given by (14),  $\bar{p}'$  is a solution of a problem of the form (1) with incident wave  $\bar{p}_{\text{inc}}$ , solvable by boundary elements too.

For the functional (10), the derivatives with respect to  $\nu$  are given in section 5.1 in terms of Fréchet derivatives. In terms of shape derivatives, the gradient method reads  $\nu_{n+1,m}^\ell = \nu_{n,m}^\ell - \tau_{n,m}^\ell [\sigma_{\text{noise}}^{-2} \langle DJ_{\text{c}}(\mathbb{R}^2 \setminus \bar{\Omega}_{\nu_n}), \mathbf{V}_m \rangle + \sigma_m^{-2} (\nu_{n,m} - \nu_{0,m})]$ , for  $\ell = 1, \dots, L$ ,  $m = 1, \dots, 2M + 3$ ,  $\tau_{n,m}^\ell > 0$  small, with directions  $\mathbf{V}_1 = (1, 0)$ ,  $\mathbf{V}_2 = (0, 1)$ ,  $\mathbf{V}_{2+m+1} = \cos(2\pi mt)(\cos(2\pi t), \sin(2\pi t))$ , for  $m = 0, \dots, M$  and  $\mathbf{V}_{3+M+m} = \sin(2\pi mt)(\cos(2\pi t), \sin(2\pi t))$ , for  $m = 1, \dots, M$ . We have used the above formulas to check codes and iterative procedures.

## A2. Fréchet derivative with respect to the coefficients

An analogous but much simpler procedure to that followed the previous section, provides an expression for Fréchet derivatives when  $\kappa_i$  is a constant.

**Theorem 2.** *Keeping the notations and assumptions of the previous section, the ‘Fréchet derivative’ of the far field of the solution  $u$  of (1) with respect to  $\kappa_i$  is given by the far field of the solution of the boundary value problem (A.20):*

$$\begin{aligned} \Delta v + \kappa_e^2 v &= 0 \quad \text{in } \Omega_e, \quad \Delta v + \kappa_i^2 v = -2\kappa_i u \quad \text{in } \Omega_i, \\ v^- - v^+ &= 0 \quad \text{on } \partial\Omega_i, \quad \beta \frac{\partial v^-}{\partial \mathbf{n}} - \frac{\partial v^+}{\partial \mathbf{n}} = 0 \quad \text{on } \partial\Omega_i, \\ \lim_{|\mathbf{x}| \rightarrow 0} |\mathbf{x}|^{1/2} \left( \frac{\partial}{\partial |\mathbf{x}|} v - \imath k_e v \right) &= 0. \end{aligned} \quad (\text{A.20})$$

**Proof.** Let  $u$  be the solution of (1),  $\kappa_i$  constant, and  $u^t$  the solution of (1) with coefficient  $\kappa_i^t = \kappa_i + t$ . Then  $u^t - u$  is a solution of (1) with right hand side  $f^t = -2\kappa_i t u^t - t^2 u^t$  and zero incident wave. As argued in step 1 of the proof of theorem 1, the forward problem (1) admits the variational reformulation (A.2). Well-posedness of the variational problems (A.2) with respect to changes in the coefficients implies uniform bounds on the solutions for  $t \in [0, T]$ :  $\|u^t\|_{H^1(B_R)} \leq C(T)$ . Thus, the right hand sides  $f^t$  tend to zero as  $t \rightarrow 0$ . Well-posedness of the variational problem again implies that  $u^t \rightarrow u^0 = u$  in  $H^1(B_R)$ . The quotients  $\frac{u^t - u^0}{t}$  are solutions of (1) with right-hand side  $\frac{f^t - f^0}{t} = -2\kappa_i u^t - t u^t$ , which tends to  $-2\kappa_i u$  in  $H^1(B_R)$ . Well-posedness of the variational problem again with respect to the right hand side implies that the limit  $\dot{u} = \lim_{t \rightarrow 0} \frac{u^t - u^0}{t}$  is a solution of (A.20).  $\square$

This boundary problem has a non zero right-hand side. We may solve it by finite elements. However, in an iterative optimization procedure that requires solving problems of the form (A.20) for different  $\Omega_i$  in each iteration, we have to use fine meshes and remesh each time we change the proposed objects, which is expensive. Alternatively, we can solve for the right hand side in the whole space by convolution with the Green function of the exterior Helmholtz problem and correct it solving a transmission problem with zero source by BEM. The convolution makes this option equally expensive. From the computational point of view it is more convenient to use the definition  $\frac{\partial u(\kappa_i)}{\partial t} = \frac{u(\kappa_i + t) - u(\kappa_i)}{t}$  for small  $t > 0$  to approximate it. In this way we only have to solve an additional forward problem with parameter  $\kappa_i + t$  by BEM per iteration. In this way we complete the Fréchet matrices  $\mathbf{F}(\nu)$  with an additional column to obtain the Fréchet matrices  $\mathbf{F}(\nu, \kappa_i)$ .

## ORCID iDs

Ana Carpio  <https://orcid.org/0000-0001-8623-6521>

Sergei Iakunin  <https://orcid.org/0000-0002-6355-9907>

Georg Stadler  <https://orcid.org/0000-0001-7762-6544>

## References

- [1] Ahn C Y, Jeon K, Ma Y K and Park W K 2014 A study on the topological derivative-based imaging of thin electromagnetic inhomogeneities in limited-aperture problems *Inverse Problems* **30** 105004
- [2] Belkebir K and Saillard M 2001 Testing inversion algorithms against experimental data *Inverse Problems* **17** 1565–71
- [3] Bishop C M 2006 *Pattern Recognition and Machine Learning* (Berlin: Springer)
- [4] Borges C, Gillman A and Greengard L 2017 High resolution inverse scattering in two dimensions using recursive linearization *SIAM J. Imag. Sci.* **10** 641–64
- [5] Box G E P and Tiao G C 1965 Multiparameter problem from a Bayesian point of view *Ann. Math. Stat.* **36** 1468–82
- [6] Bui-Thanh T, Ghattas O, Martin J and Stadler G 2013 A computational framework for infinite-dimensional Bayesian inverse problems: Part I. The linearized case, with application to global seismic inversion *SIAM J. Sci. Comput.* **35** A2494–523
- [7] Bui-Thanh T and Ghattas O 2014 An analysis of infinite dimensional Bayesian inverse shape acoustic scattering and its numerical approximation *SIAM/ASA J. Uncertain. Quantification* **2** 203–22
- [8] Cakoni F, Colton D and Monk P 2011 *The Linear Sampling Method in Inverse Electromagnetic Scattering* (Philadelphia, PA: SIAM)
- [9] Carlin B P and Louis T L 2000 *Bayes and Empirical Bayes Methods for Data Analysis* 2nd edn (London: Chapman and Hall)
- [10] Carpio A and Rapún M L 2008 Topological derivatives for shape reconstruction *Lect. Notes Math.* **1943** 85–133
- [11] Carpio A and Rapún M L 2008 Solving inverse inhomogeneous problems by topological derivative methods *Inverse Problems* **24** 045014
- [12] Carpio A, Dimiduk T G, Selgas V and Vidal P 2018 Optimization methods for in-line holography *SIAM J. Imag. Sci.* **11** 923–56
- [13] Carpio A, Dimiduk T G, Le Louër F and Rapún M L 2019 When topological derivatives met regularized Gauss–Newton iterations in holographic 3D imaging *J. Comp. Phys.* **388** 224–51
- [14] Caubet F, Godoy M and Conca C 2016 On the detection of several obstacles in 2D Stokes flow: topological sensitivity and combination with shape derivatives *Inverse Probl. Imag.* **10** 327–67
- [15] Chipman H, George E I and McCulloch R E 2001 The practical implementation of Bayesian model selection *Model Selection (IMS Lecture Notes-Monograph Series vol 38)* (Beachwood, OH: Institute of Mathematical Statistics) pp 65–116
- [16] Costabel M and Stephan E 1985 A direct boundary integral equation method for transmission problems *J. Math. Anal. Appl.* **106** 367–413
- [17] Cotter S L, Roberts G, Stuart A M and White D 2013 MCMC methods for functions: modifying old algorithms to make them faster *Stat. Sci.* **28** 424–46
- [18] Dellaportas P, Forster J J and Ntzoufras I 2002 On Bayesian model and variable selection using MCMC *Stat. Comput.* **12** 27–36
- [19] Dimiduk T G and Manoharan V N 2016 Bayesian approach to analyzing holograms of colloidal particles *Opt. Express* **24** 24045–60
- [20] Dominguez N and Gibiat V 2010 Non-destructive imaging using the time domain topological energy method *Ultrasonics* **50** 367–72
- [21] Domínguez V, Lu S and Sayas F J 2014 A Nyström flavored Calderón calculus of order three for two dimensional waves, time-harmonic and transient *Comput. Math. Appl.* **67** 217–36
- [22] Domínguez V, Lu S and Sayas F J 2014 A fully discrete Calderon calculus for two dimensional time harmonic waves *Int. J. Numer. Anal. Model.* **11** 332–45
- [23] Dorn O and Lesselier D 2006 Level set methods for inverse scattering *Inverse Problems* **22** R67–131

- [24] Dunlop M M, Iglesias M A and Stuart A M 2017 Hierarchical Bayesian level set inversion *Stat. Comput.* **27** 1555–84
- [25] Dunlop M M 2016 Analysis and computation for Bayesian inverse problems *PhD Thesis* Warwick
- [26] Feijoo G R 2004 A new method in inverse scattering based on the topological derivative *Inverse Problems* **20** 1819–40
- [27] Feijoo G R, Oberai A A and Pinsky P M 2004 An application of shape optimization in the solution of inverse acoustic scattering problems *Inverse Problems* **20** 199–228
- [28] Fletcher R 1971 Modified Marquardt subroutine for non-linear least squares *Tech. Rep.* 197213
- [29] Foreman-Mackey D, Hogg D W, Lang D and Goodman J 2013 emcee: The MCMC Hammer *Publ. Astron. Soc. Pac.* **125** 306–12
- [30] Gao H W, Mishra K I, Winters A, Wolin S and Grier D G 2018 Flexible wide-field high-resolution scanning camera for continuous-wave acoustic holography *Rev. Sci. Instrum.* **89** 114901
- [31] Gelman A and Rubin D B 1992 Inference from iterative simulation using multiple sequences *Stat. Sci.* **7** 457–72
- [32] Goodman J and Weare J 2010 Ensemble samplers with affine invariance *Commun. Appl. Math. Comput. Sci.* **5** 65–80
- [33] Guzina B and Bonnet M 2006 Small-inclusion asymptotic of misfit functionals for inverse problems in acoustics *Inverse Problems* **22** 1761–85
- [34] Guzina B and Pourhamadian F 2015 Why the high-frequency inverse scattering by topological sensitivity may work *Proc. R. Soc. A* **471** 2179
- [35] He L, Kao C Y and Osher S 2007 Incorporating topological derivatives into shape derivatives based level set methods *J. Comp. Phys.* **225** 891–909
- [36] Hettlich F 1995 Fréchet derivatives in inverse obstacle scattering *Inverse Problems* **11** 371–82
- [37] Hintermüller M and Laurain A 2008 Electrical impedance tomography: from topology to shape *Control and Cybernetics* **37** 913–33
- [38] Hohage T and Schormann C 1998 A Newton-type method for a transmission problem in inverse scattering *Inverse Problems* **14** 1207–27
- [39] Ito K, Jin B and Zou J 2013 A direct sampling method for inverse electromagnetic medium scattering *Inverse Problems* **29** 095018
- [40] Kaipio J and Somersalo E 2006 *Statistical and Computational Inverse Problems* vol 160 (Berlin: Springer)
- [41] Keller J B and Givoli D 1989 Exact non-reflecting boundary conditions *J. Comp. Phys.* **82** 172–92
- [42] Kirsch A 2002 The MUSIC-algorithm and the factorization method in inverse scattering theory for inhomogeneous media *Inverse Problems* **18** 1025–40
- [43] Klein A, Falkner S, Bartels S, Hennig P and Hutter F 2017 Fast Bayesian optimization of machine learning hyperparameters on large datasets *Proc. 20th Int. Conf. on Artificial Intelligence and Statistics (AISTATS) 2017* Fort Lauderdale, Florida, USA vol 54 JMLR: W & CP
- [44] Kleinman R E and Martin P 1988 On single integral equations for the transmission problem of acoustics *SIAM J. Appl. Math.* **48** 307–25
- [45] Kress R and Roach G F 1978 Transmission problems for the Helmholtz equation *J. Math. Phys.* **19** 1433–7
- [46] Lapidot A 2009 *A Foundation in Digital Communication* (Cambridge: Cambridge University Press)
- [47] Lee S H, Roichman Y, Yi G R, Kim S H, Yang S M, Blaaderen A, van Oostrum P and Grier D G 2007 Characterizing and tracking single colloidal particles with video holographic microscopy *Opt. Express* **15** 18275–82
- [48] Li P, Wang Y, Wang Z and Zhao Y 2016 Inverse obstacle scattering for elastic waves *Inverse Problems* **32** 115018
- [49] Litman A and Crocco L 2009 Testing inversion algorithms against experimental data: 3D targets *Inverse Problems* **25** 020201
- [50] Palafox A, Capistrán M A and Christen J A 2014 Effective parameter dimension via Bayesian model selection in the inverse acoustic scattering problem *Math. Probl. Eng.* **2014** 427203
- [51] Palafox A 2016 Uncertainty quantification on the inverse scattering problem *PhD Thesis* CIMAT
- [52] Palafox A, Capistrán M A and Christen J A 2017 Point cloud-based scatterer approximation and affine invariant sampling in the inverse scattering problem *Math. Methods Appl. Sci.* **40** 3393–403
- [53] Penny W, Mattout J and Trujillo-Barreto N 2006 Bayesian model selection and averaging *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (Amsterdam: Elsevier)

- [54] Petra N, Martin J, Stadler G and Ghattas O 2014 A computational framework for infinite-dimensional Bayesian inverse problems: Part II. Stochastic Newton MCMC with application to ice sheet flow inverse problems *SIAM J. Sci. Comput.* **36** A1525–55
- [55] Rathin Y, Vaswani N, Tannenbaum A and Yezzi A 2007 Tracking deforming objects using particle filtering for geometric active contours *IEEE Trans. Pattern Anal. Mach. Intell.* **29** 1470–5
- [56] Tarantola A 2005 *Inverse Problem Theory and Methods for Model Parameter Estimation* (Philadelphia, PA: SIAM)
- [57] Sokolowski J and Zochowski A 1999 On the topological derivative in shape optimization *SIAM J. Control Optim.* **37** 1251–72
- [58] Zhu H, Li S, Fomel S, Stadler G and Ghattas O 2016 A Bayesian approach to estimate uncertainty for full-waveform inversion using *a priori* information from depth migration *Geophysics* **81** R307–23
- [59] Vincent T 2012 *Introduction to Holography* (Boca Raton, FL: CRC Press)