Research Article

Effective Parameter Dimension via Bayesian Model Selection in the Inverse Acoustic Scattering Problem

Abel Palafox, Marcos A. Capistrán, and J. Andrés Christen

CIMAT A.C., Jalisco S/N, Valenciana, 36240 México, GTO, Mexico

Correspondence should be addressed to Abel Palafox; abel.palafox@cimat.mx

Received 8 April 2014; Revised 23 July 2014; Accepted 23 July 2014; Published 7 September 2014

Academic Editor: Fatih Yaman

Copyright © 2014 Abel Palafox et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We address a prototype inverse scattering problem in the interface of applied mathematics, statistics, and scientific computing. We pose the acoustic inverse scattering problem in a Bayesian inference perspective and simulate from the posterior distribution using MCMC. The PDE forward map is implemented using high performance computing methods. We implement a standard Bayesian model selection method to estimate an effective number of Fourier coefficients that may be retrieved from noisy data within a standard formulation.

1. Introduction

Inverse problems are typically ill-posed and analytical solutions are seldom available. Approaches to inverse problems in the interface of applied mathematics, statistics, and scientific computing represent a setting with a myriad of tools for robust solutions, for a number of reasons including its sound theoretical setting for uncertainty quantification [1, 2], prediction, and decision making. See [3] for a recent review. Although these interdisciplinary approaches are becoming prevalent [4–8], topics such as knowledge representation in complex systems [9] and effective dimension are not as mature yet. Sparsity promoting regularization and Bayesian methods have received attention recently [8, 10].

In this paper we consider a nonlinear acoustic scattering inverse problem as a case study. The rationale is that the well posedness of the direct problem and robust numerical methods has been comprehensively studied [11, 12]. This fact allows us to address the inverse problem with a reliable direct problem solver.

The propagation of acoustic waves in a homogeneous isotropic medium with constant speed of sound is governed by the linear wave equation:

$$U_{tt} = c^2 \Delta U \quad \text{in } \mathbb{R}^n, n = 2, 3 \quad (1)$$

for a velocity potential U and c the speed of sound in the medium. For monochromatic time-harmonic waves with frequency ω we have

$$U(x, t) = \text{Re} \left(e^{i\omega t} u(x) \right), \quad (2)$$

where the space dependent term $u(x)$ satisfies the Helmholtz equation. Consider the following:

$$\Delta u + k^2 u = 0, \quad (3)$$

where $k = \omega/c$ is called the *wave number*. Given an obstacle of compact support $D \subset \mathbb{R}^n$ ($n = 2, 3$), its forward scattering problem is governed by the Helmholtz equation in $\mathbb{R}^n - \overline{D}$. The total wave $u(x) = \exp(ikx \cdot d) + u^s(x)$ is a superposition of the incident wave $\exp(ikx \cdot d)$ and the scattered wave $u^s(x)$, and it is subject to a boundary condition on $\Gamma = \partial D$. The boundary condition may be of type $u = 0$ (Dirichlet), $\partial u / \partial \nu = 0$ (Neumann), or $\partial u / \partial \nu + i k \lambda u = 0$ (impedance). Also, the scattered waves u^s satisfy the Sommerfeld radiation condition

$$\lim_{r \rightarrow \infty} r^{(n-1)/2} \left(\frac{\partial u^s}{\partial r} - i k u^s \right) = 0, \quad r = \|x\|_2 \quad (4)$$

and are referred to as *radiating solutions* of the Helmholtz equation.

Given an incident wave $\exp(ikx \cdot d)$, the obstacle boundary Γ uniquely determines the scattered wave $u^s(x)$, and every scattered wave $u^s(x)$ which is a radiating solution of the Helmholtz equation has the asymptotic behavior of an outgoing spherical wave. Consider the following:

$$u(x) = \frac{e^{ik|x|}}{|x|} \left\{ u_\infty(\hat{x}) + O\left(\frac{1}{|x|}\right) \right\}, \quad (5)$$

where $\hat{x} = x/|x|$ and $u_\infty(\hat{x})$ are the scattering amplitude or *far field pattern*. We denote by F_Γ the boundary to far field mapping:

$$F_\Gamma(\hat{x}) = u_\infty(\hat{x}). \quad (6)$$

The mapping from the obstacle boundary Γ to the far field scattering amplitude u_∞ is injective [11, 12].

In this work we address the problem of estimating the boundary Γ given noisy measurements of the far field data u_∞ . This paper is organized as follows. For the sake of making this paper self-contained, in Section 2 we briefly review useful results used throughout the paper. The acoustic layer potential and integral equation method are presented in Section 2.1. We discuss the numerical solution of the forward mapping and its implementation in Section 2.2. In Section 2.3 we present in detail the Bayesian formulation of the inverse problem along with the probability distributions involved. The inverse problem results and the effective dimension analysis are presented in Section 3. We consider the effective dimension exploration as the main contribution of this work.

2. Materials and Methods

In this section we describe the forward mapping evaluation using the integral equations method and the layer acoustic potentials approach as in [11, 12].

2.1. Single Layer and Double Layer Potentials. The fundamental solution of the Helmholtz equation (3) is

$$\Phi(x, y) = \begin{cases} \frac{1}{4\pi} \frac{e^{ik|x-y|}}{|x-y|} & n = 3 \\ \frac{i}{4} H_0^{(1)}(k|x-y|) & n = 2, \end{cases} \quad (7)$$

for $x \neq y$, where $|\cdot|$ denotes the L_2 norm and $H_0^{(1)}$ is the Hankel function of first kind and order zero. Given an integrable function φ , the integral operators

$$(S\varphi)(x) = \int_\Gamma \varphi(y) \Phi(x, y) ds(y), \quad x \in \mathbb{R}^n \setminus \Gamma, \quad (8)$$

$$(K\varphi)(x) = \int_\Gamma \varphi(y) \frac{\partial \Phi(x, y)}{\partial \nu(y)} ds(y), \quad x \in \mathbb{R}^n \setminus \Gamma, \quad (9)$$

with ν being the unit normal vector to Γ directed to the exterior of D , are called acoustic *single layer* and acoustic *double layer* potentials for density φ , respectively. Both, the single and double layer potentials are radiant solutions of

the Helmholtz equation in $\mathbb{R}^n \setminus D$. The single layer potential is continuous in $\mathbb{R}^n \setminus D$ (including the boundary Γ) and consequently it is defined as in (8) also for $x \in \Gamma$. The double layer potential is only continuous in $\mathbb{R}^n \setminus \overline{D}$ (not including Γ); nonetheless $(K\varphi)(x)$ is defined in the boundary as its corresponding limit when x approaches Γ , namely,

$$(K\varphi)(x) = \frac{1}{2} \varphi(x) - \int_\Gamma \frac{\partial \Phi(x, y)}{\partial \nu(y)} \varphi(y) ds(y), \quad x \in \Gamma; \quad (10)$$

see [12] for details.

The direct problem is formulated in the form of a *combined potential* using the Dirichlet boundary condition:

$$\frac{1}{2} \varphi + K\varphi - i\eta S\varphi = -u^{\text{inc}}, \quad (11)$$

where $u^{\text{inc}} = \exp(ikx \cdot d)$ is the incident wave, η is a coupling factor (in our case we set $\eta = k$), and S and K are the acoustic potentials defined above.

The three potentials (single layer, double layer, and combined potential) reproduce the same far field pattern given a boundary Γ . However, the combined potential operator in (11) is the sum of the identity and a compact operator, and therefore its singular values are bounded away from zero. Consequently, the combined potential is more stable regarding numerical implementation and then we use the combined potential in the remainder.

The far field pattern for the combined potential is given by

$$u_\infty(\hat{x}) = \frac{e^{-i(\pi/4)}}{\sqrt{8\pi k}} \int_\Gamma \{k\nu(y) \cdot \hat{x} + \eta\} e^{-ik\hat{x} \cdot y} \varphi(y) ds(y). \quad (12)$$

The corresponding far field pattern for the single-layer potential is given by

$$u_\infty(\hat{x}) = \frac{e^{i(\pi/4)}}{\sqrt{8\pi k}} \int_\Gamma e^{-ik\hat{x} \cdot y} \varphi(y) ds(y). \quad (13)$$

The integral (12) can be evaluated numerically using the trapezoidal rule after solving the integral equation (11) for φ . It is known [13] that the trapezoidal rule has high accuracy when integrating over periodic functions and therefore is a sensible method to use in this case.

2.2. Numerical Solution of the Forward Map. As a test case we consider the two-dimensional kite-shaped domain shown in Figure 1. The parametric representation for the boundary is given by $\Gamma(t) = (\cos t + 0.65 \cos 2t - 0.65, 1.5 \sin t)$, $0 \leq t \leq 2\pi$.

The combined potential equation (11) gives rise to a linear system upon discretization:

$$A\varphi = g, \quad (14)$$

where φ is the density, $g = -2u^{\text{inc}}$, and the matrix A has the form

$$A = I - L - ikM, \quad (15)$$

FIGURE 1: Synthetic example: a kite-shaped object. Numerical results for forward mapping evaluations are obtained by generating synthetic far field pattern measurements for this smooth, nonconvex boundary.

FIGURE 2: Elapsed times of serial and parallel numerical implementations for a single forward mapping evaluation using the combined potential.

where I is the $N \times N$ identity matrix and N is the number of knots used to discretize the boundary Γ . L and M are the discrete kernels corresponding to single layer potential and double layer potential, respectively (for details see [12]).

Matrix A in (15) is square, nonsymmetric, dense, complex, and well conditioned. Mathematical software such as Matlab or Python can be used to evaluate the direct problem.

FIGURE 3: Relative errors of the forward mapping evaluation on direction $d = (1, 0)$, for the smooth and periodic kite-shaped boundary, varying the number of points N and wave number $k = 1$ fixed. This plot with logarithmic scale for y -axis illustrates the exponential convergence of Nystrom method for combined and single layer potentials.

However, it must be taken into account that a big number of evaluations of the forward mapping must be done in a Bayesian statistics approach.

Consequently, in order to have an efficient numerical machinery to evaluate the direct problem we have implemented a parallel version of the general conjugate residuals method (GCR) to solve the linear system (14). The GCR method was programmed in C on a graphics processing unit (GPU). Of note, a serial version of the GCR method programmed in C is slow, even compared with matlab and python solvers of the linear systems; see Figure 2. The implementation details for parallel computing scheme are presented on Appendix B. Although the evaluation of the far field pattern (12) is not computationally expensive, it can be done in parallel. Details are shown also in Appendix B.

The far field pattern value $u_\infty(d)$ obtained for the kite in the directions $d = (1, 0)$ closely resembles results previously reported in [12]. The convergence of Nystrom method, used to numerically evaluate the integral operators, is illustrated on Figure 3. We use as a reference solution the far field pattern value $u_\infty(1, 0)$ with wave number $k = 1$, obtained with the combined potential for $N = 1024$ points, and we plot in logarithmic scale for y -axis the relative errors for far field patterns obtained with combined potential (solid line in Figure 3) and single layer potential (dashed line in Figure 3) for number of points $N = 8, 16, 32, 64, 128, 256$.

2.3. Bayesian Formulation. First let us remember the Bayes rule:

$$\pi(\theta | u_\infty) = \frac{\pi(u_\infty | \theta) \pi(\theta)}{\pi(u_\infty)}. \quad (16)$$

The *posterior density* $\pi(\theta \mid u_\infty)$ quantifies the uncertainty of the parameter θ to be recovered from data u_∞ . The *prior density* $\pi(\theta)$ expresses the information regarding the unknown parameter θ . $\pi(u_\infty \mid \theta)$ is the *likelihood* function, that is, the measurement error distribution assumed, conditional on θ . The density $\pi(u_\infty)$ is a normalizing constant. Below we describe these density functions for our inverse problem.

We consider the inverse problem defined by the nonlinear forward mapping (6) from the point of view of Bayesian statistics. In Bayesian statistical inversion theory, the solution of the inverse problem is the conditional probability of parameters, given the data (posterior probability). We assume that the data are noisy measurements of far field pattern u_∞ and that the boundary Γ may be described by a vector of parameters θ as follows.

We construct a prior model based on a parameterization for Γ considering that it is a simple 2π -periodic curve and $\Gamma \in C^2(\mathbb{R}^2)$. We express Γ as follows:

$$\Gamma(t) = r(t) (\cos t, \sin t), \quad (17)$$

where $t \in [0, 2\pi)$ and

$$r(t) = \frac{a_0}{2} + \sum_{i=1}^{\infty} a_i \cos(it) + b_i \sin(it). \quad (18)$$

Thus the radius of Γ is expressed in terms of a Fourier series. The vector of parameters

$$\theta = (a_0, a_1, \dots, a_n, b_1, b_2, \dots, b_n, \dots) \quad (19)$$

contains the Fourier series coefficients in (18). For a boundary Γ that is defined by a vector of coefficients θ we denote the forward mapping by F_θ . We refer to

$$\theta_n = (a_0, a_1, \dots, a_n, b_1, b_2, \dots, b_n) \quad (20)$$

as the finite vector of $c = 2n + 1$ Fourier coefficients to represent a boundary Γ_n .

2.3.1. Posterior Distribution. The posterior distribution does not have to our knowledge a closed form. Consequently, we resort to *Markov chain Monte Carlo* (MCMC) simulation methods to analyze the posterior distribution. We use the t-walk [14], a general purpose sampling algorithm for continuous distributions, in order to sample satisfactorily from the posterior distribution. We describe briefly the t-walk algorithm and implementation details in Appendix A.

The t-walk algorithm evaluates the energy function (minus log of the nonnormalized posterior). Consider the following:

$$\varepsilon(\theta) = E - \log L(\theta, u_\infty) - \log \pi_\Theta(\theta), \quad (21)$$

for an arbitrary constant E (i.e., minus log of the nonnormalized posterior) on each iteration. In practice the constant E is not used and the energy is evaluated as minus log of the product likelihood times the prior one.

The output simulations of the t-walk are vectors of Fourier coefficients θ_n for n fixed, each one of which defines a boundary Γ_n through (18). We refer subsequently to the simulations θ_n , obtained with the t-walk, as MCMC simulations or posterior samples and we refer to a *simulated boundary* as the curve Γ_n which is defined by a simulated vector θ_n .

2.3.2. Prior Distribution. We establish the prior model as follows. Suppose that $\Gamma \in C^2(\mathbb{R}^2)$ corresponds to a star shaped domain. Then Γ admits a representation (17)-(18). Furthermore, the mean square error for approximating Γ by $\tilde{\Gamma}_n$ is given by

$$\|\Gamma - \tilde{\Gamma}_n\|^2 = \sum_{m>n} |(a_m, b_m)|^2. \quad (22)$$

Further details of this consequence of Parseval inequality are described in [15].

Due to smoothness of Γ , the amplitude of the Fourier coefficients decays quickly along the series. In fact

$$|(a_m, b_m)| = O(m^{-2}). \quad (23)$$

Note that m is the position of the coefficient in the Fourier series. Therefore the norm $|(a_m, b_m)|$ of the coefficients a_m, b_m is by definition of $O(m^{-2})$ bounded by

$$|(a_m, b_m)| < \frac{C}{m^2}, \quad (24)$$

for some positive constant $C \in \mathbb{R}$.

We pose a normal distribution for each pair a_m, b_m of coefficients as follows:

$$a_m, b_m \sim N\left(0, \frac{C}{4m^2}\right), \quad (25)$$

with a_m, b_m mutually independent. We note that the variance was set to $C/4m^2$. The rationale is that with the variance in this way, we guarantee that the pair a_m, b_m satisfies the inequality (24) with probability higher than 0.99. Moreover the variance decreases when increasing position m , that is, for high order coefficients along the Fourier series. Consequently, this modeling incorporates the smoothness of Γ on the prior distribution.

The coefficient a_0 in (18) is related to the size of the scattering object. We assume that

$$a_0 \sim N(2.5, 0.14). \quad (26)$$

It is not clear how to set the constant C in distribution (25). This constant can be seen as a scaling factor and the parameter m controls the spread of the distribution by $1/4m^2$. A single constant C can be chosen for all prior distributions (25) independently of the value of the index $m = 1, \dots, n$. For this aim we sample coefficients $\tilde{\theta}_n$ from the distributions (25) varying the value of C . We add the condition

$$r(t) > 0.75 \quad \forall t \in [0, 2\pi) \quad (27)$$

on the radius (18) to guarantee that the sample curves do not intersect themselves. The sampled curves are presented on Figure 4. As we observe, using the values of $C = 0.5$ and $C = 1$ (see Figures 4(a) and 4(b)) we obtain very smooth curves. On the other hand, Figures 4(c) and 4(d) show curves with more oscillations. For our purposes, a value of $C = 2$ is chosen.

FIGURE 4: Prior distribution samples for coefficients on our boundary representation. The sample is drawn from a Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = C/4m^2$, where m is the position of the coefficient on a Fourier series and C is constant. We vary the variance and set the value of C depending on the expected smoothness of the boundary to recover.

2.3.3. Likelihood. Assume that the measurements of the far field pattern have additive Gaussian noise

$$u_{\infty}(\hat{x}) = F^{\Gamma}(\theta) + \eta_{\hat{x}}, \quad (28)$$

with $\eta_{\hat{x}} \sim N(0, \sigma^2)$ for a boundary Γ . The likelihood is given by the noise distribution

$$L(\theta, u_{\infty}) = (2\pi\sigma)^{-N/2} \exp \left\{ -\frac{1}{2} \sum_{\hat{x}} \left\| \frac{u_{\infty}(\hat{x}) - F^{\Gamma}(\hat{x})}{\sigma} \right\|^2 \right\}, \quad (29)$$

assuming independent measurements. We assume that the measurements u_{∞} are taken on evenly spaced directions $\hat{x}$.

3. Results and Discussion

In order to avoid committing an “inverse crime” [5] in our numerical experiments we proceed as follows. First we generate synthetic noisy far field measurements from the kite-shaped object (see Figure 1). Second, this kite is not obtained from any finite Fourier series expansion and it is therefore not included in any of our models. This forces our

FIGURE 5: Marginal posterior distributions for a_i 's coefficients for $c = 11$. These unimodal distributions are used to estimate coefficients of the inverse problem.

methodology to cope with diverse shapes, being the finite Fourier series only an approximation. Third, synthetic data generation is done using the single-layer potential approach, whereas for MCMC steps we evaluate the forward mapping

using the combined potential. Of note, both the single and the combined potential approaches approximate the same far field pattern when N is large enough, however they are numerically different.

FIGURE 6: Continued.

FIGURE 6: Probability Regions from the MCMC simulations (gray). The boundaries defined from mean of marginal posterior distribution are shown in blue. The corresponding MAP boundaries are shown in green. The results are compared to the kite object in dashed red line.

We present results of the MCMC simulations, varying the number of Fourier coefficients in Section 3.1. Results representing the main contribution of this paper are presented on Section 3.2.

3.1. Inverse Problem Results. For our experiments we discretize the boundary of the kite-shaped object with 256 points. The far field pattern was computed on 16 evenly spaced points. We use 16 incident waves and we set the coupling parameter η to be equal to the wave number $k = 1$. The synthetic data are generated with additive Gaussian noise $N(0, \sigma^2)$ with $\sigma = 0.1$ (equivalently $\text{SNR} = 0.97$). We generate 100,000 MCMC simulations of the posterior distribution for each number of coefficients ($c = 3, 5, \dots, 31, 33$). The transient stage of the MCMC is taken into account and we remove the first 5,000 iterations in each case. We highlight that it takes about 7 hours to produce 100,000 MCMC simulations on a Python implementation. With our parallel implementation the execution time is reduced to about 1 hour.

Due to space constraints, we present only marginal posterior probability distributions of cosine terms for $c = 11$ coefficients in Figure 5. As we observe, for each coefficient, the corresponding distribution is unimodal with low variation. The same feature is presented for coefficients for cases $c \in \{3, 5, \dots, 31, 33\}$, $c \neq 11$ (not shown).

We present in Figure 6 the probability region (in gray) for the number of Fourier coefficients for cases $c \in \{3, 5, \dots, 31, 33\}$. For this aim we draw 1000 simulated boundaries subsampled with their corresponding IAT factor in each case. The true kite-shaped boundary is shown as a red dashed line. From Figure 6 it is apparent that as we increase the number of coefficients the samples have more oscillations and the probability region becomes imprecise. This fact exhibits the numerical instability of Fourier-based representation for the solutions of the inverse problem of interest.

Also, in Figure 6 we show what we refer to as MAP boundary (dashed line) and mean boundary (continuous line) for each value of c . The MAP boundary is the curve defined by the simulated vector θ_n that has the lowest energy value (maximum a posteriori). The mean boundary is the curve defined by the mean of each coefficient of simulated vectors θ_n . MAP boundaries and mean boundaries seem to approximate the true curve with approximately the same quality. In fact, we cannot qualitatively distinguish the best approximation from $c = 9$ coefficients to $c = 33$. This issue gives rise to the analysis presented on the next section.

3.2. Effective Dimension. We refer to *effective dimension* as the number of parameters that can be properly retrieved from a noisy data set. In the context of this paper, effective dimension

FIGURE 7: Minus logarithm of estimated normalizing constants for Bayesian model comparisons. Best model refers to the highest normalizing constant ($c = 11$ for this case). Continuous line shows the distance between the kite and the mean boundaries. Dashed line corresponds to the distance between the kite and the MAP. The boxplot corresponds to distances of the last 1000 simulations of the posterior to the kite. Red scale corresponds to minus logarithm of normalizing constants and blue scale is for the distances.

is related to the uncertainty quantification of parameters, that is, identifying the number of Fourier coefficients on the representation given by (17)–(20), required to approximate properly the boundary of the kite (see Figure 1) from synthetic far field measurements. As an aside comment we want to mention that effective dimension can be seen as an indicator for a parsimonious approximation regarding computational cost. However this feature is not used here due to the fact that effective dimension analysis is performed as a posterior procedure. In this section we have changed the notation for the distributions to be consistent with [14].

We define a collection of k models given by

$$\pi_{u^\infty|\Theta_n}^n(u^\infty | \theta_n), \quad (30)$$

with $n \in \{1, 2, \dots, k\}$. Θ_n is a vector of random variables. A realization of Θ_n is a vector θ_n defined in (20), and $\pi_{u^\infty|\Theta_n}^n$ is the corresponding posterior distribution for model n . For simplicity we use as index n to indicate the model corresponding to $c = 2n + 1$ Fourier coefficients.

We use the super-model approach; that is, a new model is defined

$$\pi_{u^\infty|\Theta, M}(u^\infty | \theta, n) = \pi_{u^\infty|\Theta_n}^n(u^\infty | \theta_n), \quad (31)$$

where $M = n$ is the indicator of the model and $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_k)$.

The posterior probability for model $M = n$ of explaining the observed far field pattern is given by

$$\pi(M = n | u_\infty) \propto \pi_M(n) \pi_{u_\infty}^n(u_\infty), \quad (32)$$

where $\pi_M(n)$ is the prior probability of model $M = n$ and $\pi_{u_\infty}^n(u_\infty)$ is the corresponding normalizing constant (also known as the marginal likelihood) for model n . A straightforward way to perform model selection is by computing (32) for all the models and choosing the model with highest probability. Indeed, the difficulty regarding the computation of (32) is evaluating the normalizing constant $\pi_{u_\infty}^n(u_\infty)$. To perform our MCMC only the nonnormalized posterior is needed and therefore $\pi_{u_\infty}^n(u_\infty)$ was not previously calculated (this is the usual case when doing MCMC in a fixed number of parameters in a Bayesian application). Nevertheless, the MCMC simulations may be used to estimate $\pi_{u_\infty}^n(u_\infty)$ and provide information to explore the effective dimension. This is explained below.

The ratio of two normalizing constants is defined as follows:

$$r = \frac{\pi_{u_\infty}^l(u_\infty)}{\pi_{u_\infty}^s(u_\infty)}, \quad (33)$$

where $s, l \in \{1, 2, \dots, k\}$, $s \neq l$. Considering a uniform prior distribution for the models, that is, $\pi_M(n) = 1 \setminus n$, model s is better than model l when $r > 1$, and they are indistinguishable if $r = 1$ and l is better than s when $r < 1$. Then, the best model has the highest normalizing constant value.

A consistent estimator of the ratio is

$$\hat{r} = \frac{(1/m_l) \sum_{i=1}^{m_l} \pi_{u^\infty|\Theta_l}^l(u^\infty | \theta_{l,i}) \pi_{\Theta_l}^l(\theta_{l,i}) / \pi_l^l(\theta_{l,i})}{(1/m_s) \sum_{i=1}^{m_s} \pi_{u^\infty|\Theta_s}^s(u^\infty | \theta_{s,i}) \pi_{\Theta_s}^s(\theta_{s,i}) / \pi_s^l(\theta_{s,i})}, \quad (34)$$

where π_l^l, π_s^l are completely known importance sampling densities for $\pi_{\Theta_l|u^\infty}^l$ and $\pi_{\Theta_s|u^\infty}^s$, respectively, and the sets $\{\theta_{l,1}, \theta_{l,2}, \dots, \theta_{l,m_l}\}$ and $\{\theta_{s,1}, \theta_{s,2}, \dots, \theta_{s,m_s}\}$ are samples of the posterior distribution for models l and s , respectively, (see [14] for details).

We observe from (34) that for a model s , the product

$$\pi_{u^\infty|\Theta_s}^s(u^\infty | \theta_{s,i}) \pi_{\Theta_s}^s(\theta_{s,i}) \quad (35)$$

for each $\theta_{s,i} \in \{\theta_{s,1}, \theta_{s,2}, \dots, \theta_{s,m_s}\}$,

is the corresponding likelihood times prior distribution. This product is in fact the energy (21) using a suitable constant E . Then the resulting sample from MCMC posterior simulations is used to compute (34), as a byproduct of the former with marginal computational cost added. Some details about approximating normalizing constant from MCMC simulations are discussed in [16].

The importance sampling density π_s^l must be comparable to the product (35). We propose to numerically estimate the density π_s^l for every model using a Kernel density estimation (KDE) [17] of the posterior distribution using the MCMC sample. A rough estimate may be used, for the sole purpose of estimating r .

The kernel density estimator has the form

$$\hat{\pi}_s(\theta_s^* | u_\infty) = \frac{1}{m_s h_{m_s}} \sum_{i=1}^{m_s} K_s\left(\frac{\theta_s^* - \theta_{s,i}}{h_{m_s}^i}\right), \quad (36)$$

where the kernel K is a bounded density on R^s and h_{m_s} is the bandwidth. For this estimator we define a Gaussian kernel

$$K_s(x) = (2\pi)^{-s/2} e^{-x^2/2} \quad (37)$$

and a bandwidth

$$h_{m_s} = \prod_{i=1}^{m_s} h_{m_s}^i, \quad (38)$$

where $h_{m_s}^i = 1.06\sigma_i m_s^{-1/5}$ and σ_i is the standard deviation for each coefficient on the model s . We compute

$$c_s = \left(\frac{1}{m_s} \right) \sum_{i=1}^{m_s} \frac{\pi_{u^\infty|\Theta_s}^s(u^\infty | \theta_{s,i}) \pi_{\Theta_s}^s(\theta_{s,i})}{\pi_s^I(\theta_{s,i})}, \quad (39)$$

for each model $s \in \{1, 2, \dots, k\}$ then all the models can be compared together. Then effective dimension corresponds to the model with the highest normalizing constant, equivalently the lowest value of minus logarithm of the normalizing constant.

We present in Figure 7 the minus logarithm of the estimated normalizing constants (39) for all models (red line). As we see in this figure, the lowest value is obtained with $n = 11$ coefficients. Also we present in the same Figure the distance (norm L_2 of the pointwise difference) between the radius of the kite and the radius of the simulated boundaries. The continuous line shows the difference with respect to the mean boundaries. The dashed line corresponds to the distances with respect to the map boundaries. The boxplot in the same figure corresponds to distances of the last 1000 simulations of the posterior. These last simulations are taken and subsampled with their corresponding IAT factor. These boundaries are in fact the probability region shown in Figure 6.

Based on the Bayes factors, the best model corresponds to $n = 11$ (we recall that $c = 2n + 1$). Of note, there are MAP boundaries (e.g., $c \in \{13, 19, 25, 29\}$) with lower mean squared error compared with the MAP for $c = 11$. Also the mean boundaries from $c = 15$ to $c = 33$ have lower distances to the MAP. However as we observe on the boxplot, the mean of the error decreases from 3 to 11 Fourier coefficients and then oscillations appear, increasing spread and increasing error from 13 to 33 coefficients. Despite the low error of MAP and mean boundaries for models with 13 to 33 coefficients, due to the spread shown in the boxplot, the best model is certainly given by Bayes' factor (i.e., for $c = 11$). A qualitative analysis that agrees with these results is also discussed in Section 3.

Fourier series is a classical representation for smooth periodic closed curves. Furthermore, we have used a well understood and high order numerical method for forward mapping evaluation. However, these two elements by themselves are not enough to have confidence regarding the solution of the inverse problem. The quality of the solution depends also on the number of coefficients of the Fourier series used for the representation.

Our effective dimension discussion relies on MCMC methods to provide a quantitative strategy to determine the number of parameters that can be retrieved given a data set.

4. Conclusions

In this work we address the acoustic inverse scattering problem with a classical Fourier-based representation of the solution. We pose the inverse problem as a Bayesian inference problem and use the output of a MCMC method (namely, the t-walk) for our effective dimension results. For the corresponding direct problem we have used the classical layer potential approach, which was solved in a fast and reliable manner with a robust numerical method and parallel computing. Using Fourier series to represent solutions allows for a straightforward formulation that incorporates the smoothness of the solutions into the prior distribution. On the other hand, the finite Fourier representation is numerically unstable. Although other approaches to represent the scattering obstacle are applicable (e.g., wavelet basis which correspond to Besov priors), a fundamental question remains: How much information can be retrieved, within the representation, from a noisy data set?

The main contribution of this paper is the effective dimension method, which is a quantitative method to estimate and quantify the uncertainty of the estimable parameters given a noisy data set. Given a parametric representation of the solution of the inverse problem, the effective dimension method is implemented via Bayesian model selection where the normalizing constant for each model is approximated using the MCMC output. Of note, the effective dimension method is applicable regardless of the parametric representation of the solution.

Appendices

A. The t-Walk

The t-walk (for "traverse" or "thoughtful" walk) is a MCMC sampler for arbitrary continuous distributions that require no tuning. The t-walk maintains two independent points in the sample space and all moves are based on four proposals (walk, traverse, hop, and blow) that are accepted with a standard Metropolis-Hastings acceptance probability on the product space. These moves produce an efficient sampling algorithm that is invariant to scale and approximately invariant to affine transformations of the state space.

For an objective function (e.g., posterior distribution) $\pi(x)$, $x \in \mathcal{X}$, and $\mathcal{X} \subset \mathbb{R}^n$, a new objective function is defined as $f(x, x') = \pi(x)\pi(x')$ in the corresponding product space $\mathcal{X} \times \mathcal{X}$. Then two proposals are considered as follows:

$$(y, y') = \begin{cases} (x, h(x, x')), & \text{with probability } 0.5 \\ (h(x, x'), x'), & \text{with probability } 0.5, \end{cases} \quad (\text{A.1})$$

where $h(\cdot, \cdot)$ is defined by one of the following four moves.

(i) *Walk Move*. The walk move is defined by the function

$$h_w(x, x')_j = \begin{cases} x_j + (x_j - x'_j)\alpha_j, & I_j = 1 \\ x_j, & I_j = 0, \end{cases} \quad (\text{A.2})$$

for $j = 1, 2, \dots, n$, where $\alpha_j \in \mathbb{R}$ are i.i.d. r.v. with density

$$\psi_w(\alpha_j) = \begin{cases} \frac{1}{k\sqrt{1+\alpha_j}}, & \alpha \in \left[\frac{-a_w}{1+a_w}, a_w\right] \\ 0, & \text{otherwise} \end{cases} \quad (\text{A.3})$$

with $a_w = 1.5$, $k = 2(\sqrt{1+a_w} - 1/\sqrt{1+a_w})$, and $\alpha_j = (a_w/(1+a_w))(-1+2u+a_w u^2)$ with $u \sim U(0, 1)$. In this case the Hastings ratio is

$$\frac{g(x|y, x')}{g(y|x, x')} = 1, \quad (\text{A.4})$$

for both cases on (A.1).

(ii) *Traverse Move*. The traverse move is defined by the function

$$h_t(x, x')_j = \begin{cases} x'_j + \beta(x'_j - x_j), & I_j = 1 \\ x_j, & I_j = 0, \end{cases} \quad (\text{A.5})$$

where $\beta \in \mathbb{R}^+$ is a r.v. with density

$$\begin{aligned} \psi_t(\beta) = & \frac{a_t - 1}{2a_t} \{(a_t + 1)\beta^{a_t} I_{(0,1]}(\beta)\} \\ & + \frac{a_t + 1}{2a_t} \{(a_t - 1)\beta^{-a_t} I_{(1,\infty)}(\beta)\} \end{aligned} \quad (\text{A.6})$$

which is a mixture of densities and can be easily sampled as follows:

$$\beta = \begin{cases} \frac{1}{u^{a_t+1}}, & \text{with probability } \frac{a_t - 1}{2a_t} \\ \frac{1}{u^{1-a_t}}, & \text{with probability } \frac{a_t + 1}{2a_t}, \end{cases} \quad (\text{A.7})$$

with $a_t = 6$ and $u \sim U(0, 1)$. The acceptance ratios are

$$\frac{\pi(y')}{\pi(x')} \beta^{n_t-2} \quad \text{or} \quad \frac{\pi(y)}{\pi(x)} \beta^{n_t-2}, \quad (\text{A.8})$$

for first and second cases on (A.1), respectively.

(iii) *Hop Move*. The hop move is defined by the function

$$h_h(x, x') = \begin{cases} x_j + \frac{\sigma(x, x')}{3} z_j, & I_j = 1 \\ x_j, & I_j = 0, \end{cases} \quad (\text{A.9})$$

with $z_j \sim N(0, 1)$ and $\sigma(x, x') = \max_{I_j=1} |x_j - x'_j|$. For this proposal

$$\begin{aligned} g_h(y|x, x') &= \frac{(2\pi)^{-n_t/2} 3^{n_t}}{\sigma(x, x')^{n_t}} \exp \left\{ -\frac{9}{2\sigma(x, x')^2} \sum_{I_j=1} (y_j - x_j)^2 \right\} \\ &\times \prod_{I_j=0} \delta_{x_j}(y_j). \end{aligned} \quad (\text{A.10})$$

(iv) *Blow Move*. The blow move is defined by the function

$$h_b(x, x') = \begin{cases} x'_j + \sigma(x, x') z_j, & I_j = 1 \\ x_j, & I_j = 0, \end{cases} \quad (\text{A.11})$$

with $z_j \sim N(0, 1)$ and $\sigma(x, x') = \max_{I_j=1} |x_j - x'_j|$. For this proposal

$$\begin{aligned} g_b(y|x, x') &= \frac{(2\pi)^{-n_t/2}}{\sigma(x, x')^{n_t}} \exp \left\{ -\frac{1}{2\sigma(x, x')^2} \sum_{I_j=1} (y_j - x_j)^2 \right\} \\ &\times \prod_{I_j=0} \delta_{x_j}(y_j). \end{aligned} \quad (\text{A.12})$$

The numerical implementation of the algorithm only requires the user to define three functions.

- (i) *Initialization*. A function to generate the two different initial guesses x and x' .
- (ii) *Support*. Defines the support of the target function. Points outside of the support are rejected.
- (iii) *Energy*. In this function the minus logarithm of the target density is evaluated.

The algorithm is available for download from Andres Christen's personal web page <http://www.cimat.mx/~jac/twalk> and it has been implemented on C, C++, Matlab, R, and Python languages.

B. Parallel Computing

Although MCMC methods are by definition serial procedures, the high computational cost can be reduced by performing each evaluation of the objective function in a parallel computing scheme when possible. In this appendix we present a way to solve the direct problem for combined potential and Nystrom method.

We recall the linear system matrix for combined potential and Nystrom method:

$$(I - L + ikM)\varphi = g, \quad (\text{B.1})$$

where L, M are kernels for double layer potential and single layer potential (for details see [12]). The matrix coefficient ij corresponds to

$$(I - L + ikM)_{ij} = (I_{ij} - L_{ij} + ikM_{ij}), \quad (\text{B.2})$$

where $I_{ij} = 0$ for $i \neq j$ and $I_{ij} = 1$ for $i = j$, $L_{ij} = L(t_i, t_j)$, and $M_{ij} = M(t_i, t_j)$, with $t_i = 2\pi i/N$, $t_j = 2\pi j/N$, and $i, j = 0, 1, \dots, N$.

The evaluation of each entry of the matrix involves the numerical evaluation of Bessel functions which is computationally costly. On the other hand, each entry of the matrix is independent. Then the matrix setup is recommended to be performed in a parallel scheme in order to reduce the

```

 $r_0 = -u_i - A\varphi_0$ 
 $p_0 = r_0$ 
for  $i < N$  do
   $a_i = \frac{(r_i, Ap_i)}{(Ap_i, Ap_i)}$ 
   $x_{i+1} = x_i + a_i p_i$ 
   $r_{i+1} = r_i - a_i Ap_i$ 
   $p_{i+1} = r_i + \sum_{j=0}^i b_j^{(i)} p_j$ , where  $b_j^{(i)} = -\frac{(Ar_{i+1}, Ap_j)}{(Ap_j, Ap_j)}$ 
end for

```

ALGORITHM 1: Conjugate residuals method (GCR).

computing time. That is, given a number of knots N we define a grid of threads in CUDA of size $N \times N$ and each entry ij is evaluated in a different thread.

In order to solve the linear system (B.1) we choose the GCR method which is described in Algorithm 1. The dot products involved are performed based on the cascading algorithm of the *optimizing parallel reduction in CUDA* sample from Nvidia CUDA toolkit documentation (see web page <http://docs.nvidia.com/cuda>).

The most demanding part is computing the vector p since a set of i coefficients b should be computed on each iteration. For this aim, we use i CUDA blocks and we compute the corresponding $b_j^{(i)}$ within each block by using cascading algorithm.

When using multiple incident waves, the term u_i becomes a matrix. This is equivalent to solve as many linear systems as incident waves. The same algorithm is performed using a CUDA block for each incident wave excepting the computing of $b_j^{(i)}$ which is done for a single incident wave at time. The computing of the far field pattern can be done in parallel by approximating the integral (12) for the multiple incident waves. However this latter evaluation is not expensive and it takes no advantage of the parallel computing.

In our experiments, this parallel computing implementation has a performance enhancement of about 7x speedup over the original. An implementation with a more effective method than this parallel version of GCR is left as future work.

Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank Editor Fatih Yaman and anonymous referees for their constructive and detailed remarks which have helped to improve this paper in a major way. This research was supported by Grant GTO-2011-C04-168676 Guanajuato-CONACyT.

References

- [1] T. Oden, R. Moser, and O. Ghattas, "Computer predictions with quantified uncertainty, part I," *SIAM News*, vol. 43, no. 9, 2010.
- [2] T. Oden, R. Moser, and O. Ghattas, "Computer predictions with quantified uncertainty, part II," *SIAM News*, vol. 43, no. 10, pp. 1–4, 2010.
- [3] C. Fox, H. Haario, and J. Christen, *Bayesian Theory and Applications*, Inverse Problems, 2013.
- [4] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [5] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, vol. 160 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 2005.
- [6] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, Pa, USA, 2005.
- [7] L. Biegler, G. Biros, O. Ghattas et al., Eds., *Large-Scale Inverse Problems and Quantification of Uncertainty*, Wiley Series in Computational Statistics, John Wiley & Sons, New York, NY, USA, 2011.
- [8] V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen, "Sparsity-promoting Bayesian inversion," *Inverse Problems*, vol. 28, no. 2, Article ID 025005, 2012.
- [9] U. Grenander and M. I. Miller, "Representations of knowledge in complex systems," *Journal of the Royal Statistical Society*, vol. 56, no. 4, pp. 549–603, 1994.
- [10] C. A. Zarzer, "On Tikhonov regularization with non-convex sparsity constraints," *Inverse Problems*, vol. 25, no. 2, Article ID 025006, 13 pages, 2009.
- [11] D. L. Colton and R. Kress, *Integral Equation Methods in Scattering Theory*, vol. 57, John Wiley & Sons, New York, NY, USA, 1983.
- [12] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, vol. 93 of *Applied Mathematical Sciences*, Springer, Berlin, Germany, 2nd edition, 1998.
- [13] G. Dalquist and A. Björck, *Numerical Methods in Scientific Computing*, vol. 102, SIAM, 2008.
- [14] M. Chen, Q. Shao, and J. G. Ibrahim, *Monte Carlo methods in Bayesian Computation*, Springer Series in Statistics, Springer, New York, NY, USA, 2000.
- [15] M. A. Pinsky, *Introduction to Fourier Analysis and Wavelets*, vol. 102 of *Graduate Studies in Mathematics*, American Mathematical Society, 2002.
- [16] P. de Valpine, "Improved estimation of normalizing constants from Markov chain MONte Carlo output," *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 333–351, 2008.
- [17] B. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability no. 26, Chapman & Hall, New York, NY, USA, 1986.