PAPER

On global normal linear approximations for nonlinear Bayesian inverse problems

IOP Publishing Inverse Problems

Inverse Problems **39** (2023) 054001 (18pp) https://doi.org/10.1088/1361-6420/acc129

On global normal linear approximations for nonlinear Bayesian inverse problems

Ruanui Nicholson[1], **Noémi Petra**[2,*], **Umberto Villa**[3] and **Jari P Kaipio**[4]

[1] Department of Engineering Science, University of Auckland, Auckland 1010, New Zealand
[2] Department of Applied Mathematics, University of California, Merced, CA 95343, United States of America
[3] Oden Institute for Computational Engineering & Sciences, The University of Texas at Austin, Austin, TX 78712, United States of America
[4] Department of Applied Physics, University of Eastern Finland, 70211 Kuopio, Finland

E-mail: npetra@ucmerced.edu

Received 3 August 2022; revised 20 January 2023
Accepted for publication 3 March 2023
Published 17 March 2023

Abstract

The replacement of a nonlinear parameter-to-observable mapping with a linear (affine) approximation is often carried out to reduce the computational costs associated with solving large-scale inverse problems governed by partial differential equations (PDEs). In the case of a linear parameter-to-observable mapping with normally distributed additive noise and a Gaussian prior measure on the parameters, the posterior is Gaussian. However, substituting an accurate model for a (possibly well justified) linear surrogate model can give misleading results if the induced model approximation error is not accounted for. To account for the errors, the Bayesian approximation error (BAE) approach can be utilised, in which the first and second order statistics of the errors are computed via sampling. The most common linear approximation is carried out via linear Taylor expansion, which requires the computation of (Fréchet) derivatives of the parameter-to-observable mapping with respect to the parameters of interest. In this paper, we prove that the (approximate) posterior measure obtained by replacing the nonlinear parameter-to-observable mapping with a linear approximation is in fact independent of the choice of the linear approximation when the BAE approach is employed. Thus, somewhat non-intuitively, employing the zero-model as the linear approximation gives the same approximate posterior as any other choice of linear approximations of the parameter-to-observable model. The independence of the linear approximation

* Author to whom any correspondence should be addressed.

is demonstrated mathematically and illustrated with two numerical PDE-based problems: an inverse scattering type problem and an inverse conductivity type problem.

Keywords: Bayesian inference, inverse problems governed by PDEs, linear approximation, model errors, Bayesian approximation errors, uncertainty quantification

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

1. Introduction

Solving large-scale inverse problems can become computationally challenging, particularly in the case of nonlinear parameter-to-observable maps. The computational challenges can be alleviated, however, by using an approximate forward model, see, for example, [1, 2]. A particularly obvious example is to replace the accurate (high-fidelity) nonlinear parameter-to-observable map with a linear (low-fidelity) approximation. Replacing an accurate parameter-to-observable mapping with an approximate one, even if seemingly well justified, will lead to model errors and uncertainties. It is well understood that if these uncertainties are not accounted for, the inversion results can be misleading, see for example [3, 4].

There are a variety of ways to account for model errors, e.g. machine learning-based approaches [5–7], employing Gaussian processes [3, 8, 9], and the Bayesian approximation error (BAE) approach [4, 10, 11]. In this paper we focus on the latter, which typically relies on posing the inverse problem in the Bayesian framework [10, 12]. Broadly speaking, the BAE approach propagates all modelling and measurement uncertainties into a single additive *total error* term which is then approximately (pre)marginalised over the prior model. The robustness and applicability of the approach has been demonstrated in a variety of settings, see for example [13–19].

Of specific relevance to the current study are the works [20–24], in which the accurate nonlinear parameter-to-observable map is replaced by an approximate linear counterpart while accounting for the model error using the BAE approach. Specifically, in [24] the inverse problem governed by the acoustic wave equation (modelling photoacoustic tomography) is formulated and solved for the initial pressure. An (approximate) linear parameter-to-observable map is introduced by ignoring the additional uncertainty in the speed of sound. The induced additional uncertainties are accounted for by the BAE approach. On the other hand, in [22, 23], the linear Born approximation is used to solve inverse scattering problems with high contrast materials, while in [20, 21] nonlinear parameter-to-observable maps related to diffuse optical tomography and electrical impedance tomography (EIT), respectively, are replaced with their respective Fréchet derivatives.

In the current paper we prove, and also illustrate with numerical examples, that when replacing a nonlinear parameter-to-observable map with a linear approximation, as long as the model errors are accounted for using the full Bayesian approximation error (full BAE) approach, the results are independent of the choice of linearisation. With the full BAE approach, we mean that (i) the measurement errors are modelled as additive and Gaussian, and that (ii) the primary unknown (parameter) is *not* approximated to be uncorrelated with the approximation errors.

How well the resulting approximate posterior represents the true posterior, is often a question of interest, and we refer to [12, 25, 26] for several results in this direction. However, in the

current paper we only concern ourselves with the independence of the approximate posterior to the choice of linearisation.

The remainder of the paper is organised as follows. In section 2 we briefly review the solution of linear(-ised) inverse problems within the Bayesian framework. In section 3 we provide a formal proof of the invariance of the resulting approximate posterior to the choice of linearisation. We illustrate this invariance through two numerical examples and investigate the convergence of the posterior models in sections 4.1 and 4.2. The implications of the results are discussed in section 5.

2. Global linear Gaussian approximations

After introducing the required notation in section 2.1, we outline the setup and solution of infinite-dimensional Bayesian linear inverse problems with additional auxiliary (random) variables. For an in-depth discussion of infinite-dimensional Bayesian inverse problems see [12], while for computational considerations, including discretisation details, see [27, 28].

2.1. Notation

The notation used in the current paper is based mostly on that provided in [12]. First, consider two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. Then for any $v_1 \in \mathcal{H}_1$ and $v_2 \in \mathcal{H}_2$, we define the (outer product) operator $v_1 \otimes v_2$ by $(v_1 \otimes v_2)u = \langle v_2, u\rangle_{\mathcal{H}_2} v_1$ for any $u \in \mathcal{H}_2$. We use μ to denote probability measures over infinite-dimensional Hilbert spaces and let $\mathbb{E}$ denote expectation. Taking $\mathcal{H}$ to be an infinite-dimensional Hilbert space, for $v \in \mathcal{H}$ with associated measure $\mu(v)$ we define the mean and covariance operator by

$$v_0 = \mathbb{E}v, \quad \mathcal{C}_{vv} = \mathbb{E}(v - v_0) \otimes (v - v_0),$$

respectively. We will assume throughout that all covariance operators are trace class (see, for example, [12, 29] for details on trace class covariance operators). This ensures that the squared L^2-norm of $v - v_0$ is bounded in expectation, that is $\mathbb{E}\|v - v_0\|^2 < +\infty$. Furthermore, we denote by $\mathcal{N}(v_0, \mathcal{C}_{vv})$ a Gaussian measure on v that has mean value $v_0 = \mathbb{E}v$ and covariance operator $\mathcal{C}_{vv}$. In the case of a joint measure $\mu(v_1, v_2)$ for $v_1 \in \mathcal{H}_1$ and $v_2 \in \mathcal{H}_2$ (both possibly infinite-dimensional), we denote the marginal measure of v_1 (resp. v_2) by μ_{v_1} (resp. μ_{v_2}).

For a finite dimensional random variable $w \in \mathbb{R}^p$ we denote by $\mathcal{N}(w_0, \Gamma_{ww})$ a Gaussian distribution on w with mean w_0 and covariance matrix Γ_{ww}. The cross-covariance operators between v and w are defined as

$$\mathcal{C}_{vw} = \mathbb{E}(v - v_0) \otimes (w - w_0), \quad \mathcal{C}_{wv} = \mathbb{E}(w - w_0) \otimes (v - v_0) = \mathcal{C}_{vw}^*.$$

Furthermore, for finite dimensional Hilbert spaces we use $\pi_w(w)$ to denote probability density function of w. Moreover, for $w_1, w_2 \in \mathbb{R}^p$, we denote by $\pi_{w_1}(w_2)$ the probability density function of w_1 evaluated at w_2. Finally, we use $\mathcal{L}(\mathcal{H}, \mathbb{R}^d)$ to denote the space of bounded linear operators from $\mathcal{H}$ to $\mathbb{R}^d$.

2.2. Linear(-ised) Bayesian inverse problems

We consider a set up where $d \in \mathbb{R}^q$ is the measurement, $e \in \mathbb{R}^q$ is an additive (measurement) error term and $m \in \mathcal{H}$ and $z \in \mathcal{Z}$ are the primary (interesting) and the auxiliary (uninteresting) unknowns, respectively, with $\mathcal{H}$ and $\mathcal{Z}$ denoting Hilbert spaces. These are related by the measurement (observation) model

$$d = \mathcal{G}(m, z) + e, \tag{2.1}$$

where $\mathcal{G} : \mathcal{H} \times \mathcal{Z} \to \mathbb{R}^q$ is the parameter-to-observable mapping.

One of the main motivation behind taking a (global) linear approximation is to avoid MCMC type algorithms to compute conditional mean estimates and/or iterative methods (successive linearisations) to compute maximum *a posteriori* estimates, and the related posterior error estimates.

Linearisation (affine approximation) involves writing

$$d = \mathcal{F}m + e + \varepsilon(m,z), \tag{2.2}$$

where $\varepsilon(m,z) = \mathcal{G}(m,z) - \mathcal{F}m$ is a random variable (the approximation/modelling error) induced by the joint distribution $\mu(m,z,e)$, and $\mathcal{F} \in \mathcal{L}(\mathcal{H},\mathbb{R}^q)$ is the linear map of the approximation. We also note that for nonadditive measurement errors, one can identify such errors as belonging to the auxiliary unknowns z [30]. Most commonly with linear approximations, $\varepsilon(m,z)$ is further approximated by a random variable independent of (m,z) or a fixed constant, and $\mathcal{F}$ is the derivative operator, i.e. $\mathcal{F} = \mathsf{D}_m\mathcal{G}$, evaluated at (m_0,z_0). However, as discussed below, other choices are also possible.

Below, we review the BAE approach in which ε is treated as a further additive error term. In particular, the full BAE approach considered in this work approximates the joint distribution of (m,e,ε) as Gaussian and premarginalise over (e,ε) [4]. The idea of premarginalisation can be understood by considering the 'ordinary additive error' measurement model: $d = \mathcal{G}(m) + e$, where m and e are mutually independent. Then, straightforward derivation gives the likelihood model $\pi_{d|m}(d\,|\,m) = \pi_e(d - \mathcal{G}(m))$. While the likelihood depends on the distribution of e, it does not include the random variable e itself: it has been premarginalised. Consider next the measurement model: $d = \mathcal{F}m + e + \varepsilon(m,z)$, where ε is clearly not mutually independent with m. This leads to

$$\pi_{d|m}(d\,|\,m) = \pi_{e+\varepsilon|m}(d - \mathcal{F}m\,|\,m).$$

In general, the density $\pi_{e+\varepsilon|m}(\cdot\,|\,m)$ does not have an analytical form and, furthermore, its implementation involves intractable integrals. The second order statistics (and thus the Gaussian approximation) for the joint distribution of (m,ε) are induced by $\mu(m,z)$ as well as $\mathcal{G}$ and $\mathcal{F}$, and can be estimated using draws from $\mu(m,z)$ and computing the related sample means and covariances.

Approximating $\pi(e)$ as Gaussian, taking $\mathcal{F}$ as the linear approximation of $\mathcal{G}$ and approximating the joint distribution $\mu(m,\varepsilon)$ as Gaussian, we obtain an analytic form for the likelihood

$$\pi_{d|m}(d\,|\,m) \propto \exp\left\{-\frac{1}{2}\left\|L_{\nu|m}(\tilde{d} - \tilde{\mathcal{F}}m)\right\|_2^2\right\}, \tag{2.3}$$

where (formally)

$$\nu = e + \varepsilon \tag{2.4}$$

$$\Gamma_{\nu|m} = \Gamma_{ee} + \Gamma_{\varepsilon\varepsilon} - \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1}\mathcal{C}_{m\varepsilon} \tag{2.5}$$

$$\Gamma_{\nu|m}^{-1} = L_{\nu|m}^{\mathrm{T}}L_{\nu|m} \tag{2.6}$$

$$\tilde{\mathcal{F}} = \mathcal{F} + \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1} \tag{2.7}$$

$$\tilde{d} = d - e_0 - \varepsilon_0 + \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1}m_0. \tag{2.8}$$

Above $\mathcal{C}_{\varepsilon m} = \mathcal{C}_{m\varepsilon}^*$ is the cross-covariance operator between ε and m (see section 2.1), $\Gamma_{\varepsilon\varepsilon}$ is the covariance of ε, and $\varepsilon_0 = \mathbb{E}\varepsilon$. With (a Gaussian approximation for) the prior model $\mu_m = \mathcal{N}(m_0,\mathcal{C}_{mm})$, (2.3)–(2.8) and using the Bayes' theorem, we get the following approximations

for the conditional mean estimate $\hat{m} \approx \mathbb{E}(m|d)$ and the posterior covariance operator $\hat{\mathcal{C}} \approx \mathcal{C}_{m|d}$,

$$\hat{\mathcal{C}} = \left(\tilde{\mathcal{F}}^* \Gamma_{\nu|m}^{-1} \tilde{\mathcal{F}} + \mathcal{C}_{mm}^{-1}\right)^{-1} \tag{2.9}$$

$$\begin{aligned} \hat{m} &= \hat{\mathcal{C}}\left(\tilde{\mathcal{F}}^* \Gamma_{\nu|m}^{-1} \tilde{d} + \mathcal{C}_{mm}^{-1} m_0\right) \\ &= \hat{\mathcal{C}}\tilde{\mathcal{F}}^* \Gamma_{\nu|m}^{-1} d + \hat{\mathcal{C}}\tilde{\mathcal{F}}^* \Gamma_{\nu|m}^{-1} (\mathcal{C}_{\varepsilon m} \mathcal{C}_{mm}^{-1} m_0 - e_0 - \varepsilon_0) + \hat{\mathcal{C}}\mathcal{C}_{mm}^{-1} m_0, \\ &= \mathcal{B}d + b, \end{aligned} \tag{2.10}$$

with $\mathcal{B} = \hat{\mathcal{C}}\tilde{\mathcal{F}}^* \Gamma_{\nu|m}^{-1}$ and $b = \hat{\mathcal{C}}\tilde{\mathcal{F}}^* \Gamma_{\nu|m}^{-1} (\mathcal{C}_{\varepsilon m} \mathcal{C}_{mm}^{-1} m_0 - e_0 - \varepsilon_0) + \hat{\mathcal{C}}\mathcal{C}_{mm}^{-1} m_0$. Note that while the linearisation $\mathcal{F}$ formally appears in (2.3), in section 3 we show that (2.3), and thus the triplet $(\hat{\mathcal{C}}, \mathcal{B}, b)$, is in fact independent of the choice of $\mathcal{F}$.

It is worth noting that the variational Bayesian approach (see for example [31–33] or the references therein) could also be used as an alternative approach to compute a Gaussian approximation to the posterior. However, when using the BAE approach the triplet $(\hat{\mathcal{C}}, \mathcal{B}, b)$ can be precomputed before the collection of data (see section 3.1), making computation of the posterior a linear algebra task only. On the other hand the variational Bayesian approach typically requires the measured data and relies on sample-based iterative methods.

3. Invariance of the Posterior to the linear approximations with BAE

Here we present our main result.

Theorem 1 (Invariance of the likelihood). *Let $\mathcal{H}$ and $\mathcal{Z}$ be Hilbert spaces, and assume $m \in \mathcal{H}$ and $z \in \mathcal{Z}$ have a joint prior measure $\mu(m,z)$. Furthermore, let m_0 and $\mathcal{C}_{mm}$ be the prior mean and trace class covariance operator of the random variable m. Suppose*

$$d = \mathcal{G}(m,z) + e, \tag{3.1}$$

where $\mathcal{G} : \mathcal{H} \times \mathcal{Z} \to \mathbb{R}^q$ is a nonlinear bounded parameter-to-observable map and $e \in \mathbb{R}^q$ has mean e_0 and covariance matrix Γ_{ee}. Then the approximate likelihood model involving $\mathcal{F} \in \mathcal{L}(\mathcal{H}, \mathbb{R}^q)$,

$$\pi_{d|m}(d|m) \propto \exp\left\{-\frac{1}{2}\|L_{\nu|m}(d - \mathcal{F}m - \bar{\varepsilon})\|_2^2\right\},$$

where $L_{\nu|m}^T L_{\nu|m} = \Gamma_{\nu|m}^{-1}$, and

$$\bar{\varepsilon} = e_0 + \varepsilon_0 + \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1}(m - m_0), \quad \Gamma_{\nu|m} = \Gamma_{ee} + \Gamma_{\varepsilon\varepsilon} - \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1}\mathcal{C}_{\varepsilon m}$$

is independent of the choice of $\mathcal{F}$.

Proof. First, we introduce the notation $g_0 = \mathbb{E}\mathcal{G}(m,z)$ and $\bar{g} = g_0 + \mathcal{C}_{\mathcal{G}m}\mathcal{C}_{mm}^{-1}(m - m_0)$ with $\mathcal{C}_{\mathcal{G}m} = \mathbb{E}(\mathcal{G}(m,z) - g_0) \otimes (m - m_0)$ and $\mathcal{C}_{m\mathcal{G}} = \mathcal{C}_{\mathcal{G}m}^*$. Then, letting $\mathcal{F} \in \mathcal{L}(\mathcal{H}, \mathbb{R}^d)$ be arbitrary with adjoint $\mathcal{F}^* \in \mathcal{L}(\mathbb{R}^d, \mathcal{H})$ and setting $f_0 = \mathbb{E}\mathcal{F}m = \mathcal{F}m_0$, we have

$$\mathcal{C}_{\mathcal{F}m} = \mathbb{E}(\mathcal{F}m - f_0) \otimes (m - m_0) = \mathcal{F}\mathbb{E}(m - m_0) \otimes (m - m_0) = \mathcal{F}\mathcal{C}_{mm}.$$

Then, noting that

$$\begin{aligned} \varepsilon_0 &= \mathbb{E}\mathcal{G}(m,z) - \mathbb{E}\mathcal{F}m = g_0 - \mathcal{F}m_0, \\ \mathcal{C}_{\varepsilon m} &= \mathbb{E}(\varepsilon - \varepsilon_0) \otimes (m - m_0) = \mathbb{E}(\mathcal{G}(m,z) - g_0 - \mathcal{F}m + \mathcal{F}m_0) \otimes (m - m_0) \\ &= \mathcal{C}_{\mathcal{G}m} - \mathcal{C}_{\mathcal{F}m}, \end{aligned}$$

we have

$$\begin{aligned} d - \mathcal{F}m - \bar{\varepsilon} &= d - \mathcal{F}m - \varepsilon_0 - \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1}(m - m_0) \\ &= d - \mathcal{F}m - g_0 + \mathcal{F}m_0 - \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1}(m - m_0) \\ &= d - g_0 - \mathcal{F}(m - m_0) - (\mathcal{C}_{\mathcal{G}m} - \mathcal{C}_{\mathcal{F}m})\mathcal{C}_{mm}^{-1}(m - m_0) \\ &= d - g_0 - \mathcal{F}(m - m_0) - \mathcal{C}_{\mathcal{G}m}\mathcal{C}_{mm}^{-1}(m - m_0) + \mathcal{F}(m - m_0) \\ &= d - g_0 - \mathcal{C}_{\mathcal{G}m}\mathcal{C}_{mm}^{-1}(m - m_0) \\ &= d - \bar{g}, \end{aligned}$$

which is independent of $\mathcal{F}$.

On the other hand, setting $\Gamma_{\mathcal{G}\mathcal{G}} = \mathbb{E}(\mathcal{G}(m,z) - g_0) \otimes (\mathcal{G}(m,z) - g_0)$, and $\Gamma_{\mathcal{G}\mathcal{F}} = \mathbb{E}(\mathcal{G}(m,z) - g_0) \otimes (\mathcal{F}m - f_0) = \Gamma_{\mathcal{F}\mathcal{G}}^*$ and noting that

$$\mathcal{C}_{\mathcal{F}\mathcal{F}} = \mathcal{F}\mathcal{C}_{mm}\mathcal{F}^*, \quad \mathcal{C}_{m\mathcal{F}} = \mathcal{C}_{mm}\mathcal{F}^*, \quad \Gamma_{\varepsilon\varepsilon} = \Gamma_{\mathcal{G}\mathcal{G}} - \Gamma_{\mathcal{G}\mathcal{F}} - \Gamma_{\mathcal{F}\mathcal{G}} + \Gamma_{\mathcal{F}\mathcal{F}},$$

we have

$$\begin{aligned} \Gamma_{\nu|m} &= \Gamma_{ee} + \Gamma_{\varepsilon\varepsilon} - \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1}\mathcal{C}_{m\varepsilon} \\ &= \Gamma_{ee} + \Gamma_{\mathcal{G}\mathcal{G}} - \Gamma_{\mathcal{G}\mathcal{F}} - \Gamma_{\mathcal{F}\mathcal{G}} + \Gamma_{\mathcal{F}\mathcal{F}} - \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1}\mathcal{C}_{m\varepsilon} \\ &= \Gamma_{ee} + \Gamma_{\mathcal{G}\mathcal{G}} - \Gamma_{\mathcal{G}\mathcal{F}} - \Gamma_{\mathcal{F}\mathcal{G}} + \Gamma_{\mathcal{F}\mathcal{F}} - (\mathcal{C}_{\mathcal{G}m} - \mathcal{C}_{\mathcal{F}m})\mathcal{C}_{mm}^{-1}(\mathcal{C}_{m\mathcal{G}} - \mathcal{C}_{m\mathcal{F}}) \\ &= \Gamma_{ee} + \Gamma_{\mathcal{G}\mathcal{G}} - \Gamma_{\mathcal{G}\mathcal{F}} - \Gamma_{\mathcal{F}\mathcal{G}} + \Gamma_{\mathcal{F}\mathcal{F}} - \mathcal{C}_{\mathcal{G}m}\mathcal{C}_{mm}^{-1}\mathcal{C}_{m\mathcal{G}} + \Gamma_{\mathcal{G}\mathcal{F}} + \Gamma_{\mathcal{F}\mathcal{G}} - \Gamma_{\mathcal{F}\mathcal{F}} \\ &= \Gamma_{ee} + \Gamma_{\mathcal{G}\mathcal{G}} - \mathcal{C}_{\mathcal{G}m}\mathcal{C}_{mm}^{-1}\mathcal{C}_{m\mathcal{G}}, \end{aligned}$$

which is also independent of $\mathcal{F}$. That is to say, the likelihood is invariant to the choice of the specific linearisation, $\mathcal{F}$.

□

Three remarks are now in order.

Remark 1 (Invariance of the posterior). In infinite dimensions, Bayes' rule (see [12, theorem 6.2]) is expressed using the Radon–Nikodym derivative of the posterior measure μ^d with respect to the prior measure μ_m,

$$\frac{\mathrm{d}\mu^d}{\mathrm{d}\mu_m}(m) \propto \pi_{\text{like}}(d|m). \tag{3.2}$$

Thus as the (approximate) likelihood $\pi_{\text{like}}(d|m)$ is independent of the choice of $\mathcal{F}$, so is the (approximate) posterior measure μ^d.

Remark 2. The zero operator $\mathcal{O}$ defined by $\mathcal{O}m = 0 \in \mathbb{R}^d$ for all $m \in \mathcal{H}$, is in $\mathcal{L}(\mathcal{H}, \mathbb{R}^d)$. Thus choosing $\mathcal{O}$ as the linearised model results in the same approximate likelihood (and thus posterior) as taking any other $\mathcal{F} \in \mathcal{L}(\mathcal{H}, \mathbb{R}^d)$ as the linearised model.

Remark 3. If the correlation between the approximation errors and the unknown are ignored (set to 0), i.e. the *enhanced error model* [4, 17], theorem 1 does not hold.

3.1. Convergence of the approximate joint second order statistics

Theorem 1 refers to the actual joint (second order) statistics (i.e. the actual means and covariances) and, thus, the result has to be taken as an asymptotic one with respect to the sample size N used to compute the sample statistics. In practice, all covariances in (2.3)–(2.8) except for Γ_{ee}, namely $\mathcal{C}_{mm}$, $\Gamma_{\varepsilon\varepsilon}$, $\mathcal{C}_{\varepsilon m}$, and $\mathcal{C}_{m\varepsilon}$, and the mean of the approximation error ε_0, are computed

Algorithm 3.1. Algorithm for computing the approximate likelihood (2.3). Note `steps 1--9` require no measured data, i.e., can be carried out *offline*.

Input: Joint prior distribution $\mu(m,z)$, covariance matrix of noise Γ_e, mean of noise e_0, data d, accurate forward model $\mathcal{G}(m,z)$, linear(-sed) model $\mathcal{F}(m,z)$

Output: Approximate likelihood $\pi_{d|m}(d\,|\,m)$.

1: **for** $i = 1$ **to** N **do**
2: Sample $(m^{(i)}, z^{(i)}) \sim \mu(m,z)$
3: Compute $\varepsilon^{(i)} = \mathcal{G}(m^{(i)}, z^{(i)}) - \mathcal{F}m^{(i)}$
} `parallelisable`

4: Compute $m_0 \approx \frac{1}{N}\sum_{i=1}^{N} m^{(i)}, \quad \mathcal{C}_{mm} \approx \frac{1}{N-1}\sum_{i=1}^{N}(m^{(i)} - m_0)(m^{(i)} - m_0)^*$

5: Compute $\varepsilon_0 \approx \frac{1}{N}\sum_{i=1}^{N} \varepsilon^{(i)}, \quad \Gamma_{\varepsilon\varepsilon} \approx \frac{1}{N-1}\sum_{i=1}^{N}(\varepsilon^{(i)} - \varepsilon_0)(\varepsilon^{(i)} - \varepsilon_0)^T$

6: Compute $\mathcal{C}_{m\varepsilon} = \mathcal{C}^*_{\varepsilon m} \approx \frac{1}{N-1}\sum_{i=1}^{N}(m^{(i)} - m_0)(\varepsilon^{(i)} - \varepsilon_0)^T$

7: Set $\Gamma_{\nu|m} = \Gamma_{ee} + \Gamma_{\varepsilon\varepsilon} - \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1}\mathcal{C}_{m\varepsilon} = (L^{\mathrm{T}}_{\nu|m} L_{\nu|m})^{-1}$

8: Set $\tilde{\mathcal{F}} = \mathcal{F} + \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1}$

9: Set $\tilde{d} = d - e_0 - \varepsilon_0 + \mathcal{C}_{\varepsilon m}\mathcal{C}_{mm}^{-1} m_0$

10: Set $\pi_{d|m}(d\,|\,m) \propto \exp\left\{-\frac{1}{2}\left\|L_{\nu|m}(\tilde{d} - \tilde{\mathcal{F}}m)\right\|_2^2\right\}$

as sample statistics based on draws from the joint prior $\mu(m,z)$ and the $\varepsilon(m,z)$ computed from these draws.

Specifically, for an ensemble of samples from the joint prior, $(m^{(i)}, z^{(i)}) \sim \mu(m,z)$ with $i = 1,2,\ldots,N$, each sample of the approximation error is computed as

$$\varepsilon^{(i)} = \varepsilon(m^{(i)}, z^{(i)}) = \mathcal{G}(m^{(i)}, z^{(i)}) - \mathcal{F}m^{(i)}.$$

The mean and covariance of ε are then approximated using the corresponding sample mean and sample covariance;

$$\mathbb{E}\varepsilon = \varepsilon_0 \approx \frac{1}{N}\sum_{i=1}^{N} \varepsilon^{(i)}, \quad \Gamma_{\varepsilon\varepsilon} \approx \frac{1}{N-1}\sum_{i=1}^{N}(\varepsilon^{(i)} - \varepsilon_0)(\varepsilon^{(i)} - \varepsilon_0)^T.$$

All means and (cross-)covariance matrices other than those of e are computed in a similar manner. We remark that all sampling can be carried out prior to the acquisition or consideration of any data, and is often referred to as being carried out *offline*.

For any $\mu(m,z)$ with finite variances, a bounded $\mathcal{G}$ and any bounded $\mathcal{F}$, each of the sample means and covariances converge as $N \to \infty$. The linear approximation does, however, change the statistics of the approximation error and, in particular, the small sample approximations. These (statistics) are a complicated (intractable) function of the fourth order statistics and are, in practice, estimated with Monte Carlo simulations. Thus, the choice of $\mathcal{F}$ can have an effect on the (small sample) performance of the global linear approximation. More specifically, the particular $\mathcal{F}$ chosen can be seen as a control variate [1, 34] for estimating the statistics of $\mathcal{G}$. As such, choosing an $\mathcal{F}$ that is (strongly) correlated with $\mathcal{G}$ can improve the performance (see section 4.2 and figure 8). However, it should be pointed out that asymptotically, the convergence rate of the Monte Carlo estimates of each of the means and covariance matrices remains of order $N^{-\frac{1}{2}}$ [34].

We summarise quantities required and the steps used for computing the approximate likelihood (2.3) in algorithm 3.1. Approximations for the conditional mean estimate $\hat{m}$

and the posterior covariance operator $\hat{C}$ can then be readily evaluated from the outputs of algorithm 3.1. using (2.9) and (2.10).

We also note that the typical sample sizes N are considerably less than the dimension of the discretised unknown, implying the sample covariance estimate of $\mathcal{C}_{mm}$ that occurs in the conditional covariance $\Gamma_{\nu|m}$ is not full rank and thus not invertible. While there are several approaches to handle this problem such as pseudoinverse and (Tikhonov type) ridge regression, we confine ourselves to the pseudoinverse in this paper. Naturally, it can be assumed that the point estimates get better as the covariance estimates get better. The catch here is, however, that the *posterior error estimates* are underestimated when the pseudoinverse is employed, and the underestimation is greatest with smallest sample sizes.

To show this, we compute statistics of the estimates as functions of the sample size N in sections 4.1 and 4.2. In particular, we compute the actual estimation errors as well as the posterior error estimates to illustrate this.

4. Numerical examples

In this section we consider two PDE-based inverse problems: an inverse scattering type problem and an inverse conductivity type problem. In both cases the forward problem is approximated computationally using the finite element method (FEM) with piece-wise linear continuous Lagrange basis functions. The parameter of interest m is also discretised using the finite element basis functions when we solve the forward problems. See [27, 28] for more details on discretisation of infinite dimensional Bayesian inverse problems using FEM.

We stress that the contribution of the present paper is that the choice of the approximative linear model does not play (asymptotically) a role when the full BAE approach is used. The focus *is not* on the performance of the full BAE versus neglecting modelling and approximation errors as such. For this reason, we do not compare results with those found when neglecting the approximation errors.

4.1. Example 1: inverse medium scattering

The first example we consider is a variation of an acoustic scattering problem that arises in a variety of applications, see, for example, [35, 36]. In particular, we consider the two-dimensional time harmonic acoustic scattering from a bounded penetrable obstacle. Almost the same problem was considered in [22] in which the linear Born (single scattering) approximation was used and the BAE was employed, although in the present paper we have near-field data instead of far-field data.

In this section, we do not have auxiliary unknowns z that would be premarginalized over. Thus, the only approximation/modelling errors are due to the linearisation. Furthermore, the idea is to investigate the use of a linearisation, as such, we do not consider different discretisations here (to avoid an inverse crime). In this section, we focus on the convergence of the estimates with the number of samples used to estimate the associated covariances. Section 4.2 considers a more elaborate example with auxiliary variables, and a comparison of the convergences of estimates with different linearisations, as well as the associated computational costs.

Specifically, let ω be the angular frequency, so that the wave number can be written as $k = \omega/c_0$, where c_0 is the background (i.e. outside any inhomogeneity) speed of sound. Then,

given an incident plane wave $u^i(x;v) = \exp(ikx\cdot v)$ propagating in direction v (with $\|v\| = 1$), and the speed of sound $c(x)$ inside the inhomogeneity, the total wave field, $u(x;v)$, and the scattered field, $u^s(x;v)$ satisfy

$$\begin{aligned} \Delta u + k^2 m u &= 0 \quad \text{in } \mathbb{R}^2, \\ u &= u^i + u^s, \\ \lim_{|r|\to\infty} r^{1/2}\left(\frac{\partial u^s}{\partial r} - iku^s\right) &= 0, \end{aligned} \tag{4.1}$$

where $m = c_0^2/c^2(x)$ denotes the refractive index. As is standard [35, 36] we assume the refractive index is positive with $m(x) = 1$ for $x \in \Omega \setminus D$. Defining $q(x) := 1 - m(x)$ (which is compactly supported in D), in the case of a penetrable inhomogeneous medium, the scattered field satisfies

$$\begin{aligned} \Delta u^s + k^2 m u^s &= k^2 q u^i \quad \text{in } \Omega \\ \lim_{|r|\to\infty}\left(\frac{\partial u^s}{\partial r} - iku^s\right) &= 0. \end{aligned} \tag{4.2}$$

We consider the so-called *multistatic* set up of the scattering problem in which the goal is to estimate m based on data consisting of point-wise measurements of (the real and imaginary parts of) the scattered field at fixed locations based on several different incident plane waves. That is, the (noiseless) data can be written as $d_{\ell j} = (\mathrm{Re}[u^s(x_\ell;v_j)], \mathrm{Im}[u^s(x_\ell;v_j)])$ for $\ell = 1,2\ldots,N_m$ and $j = 1,2\ldots,N_s$ with $\mathrm{Re}[\cdot]$ and $\mathrm{Im}[\cdot]$ denoting real and imaginary parts, and N_m, N_s denoting the number of measurement locations and number of different incident wave directions, respectively. Here, however, in contrast to [22], we consider near-field data at (finite) locations x_ℓ rather than in the respective directions.

The BAE approach was used to compensate for the linearisation of the forward problem in [22, 23], both of which employed the Born approximation

$$u^s(x) \approx -k^2 \int_D \frac{i}{4} H_0^{(1)}(k|x-s|) q(s) u^i(s)\, \mathrm{d}s = \mathcal{F}m, \tag{4.3}$$

where $H_0^{(1)}$ denotes the Hankel function of first kind and order zero. The Born approximation in (4.3) can be derived taking the first term in the Neumann series for solving the Lippmann-Schwinger equation, see, for example, [36, 37].

In this paper, we take the computational domain Ω to be a circle of radius 1.5, with D being the concentric unit circle of unit radius. We pose the Sommerfeld radiation condition on $\partial\Omega$ as an absorbing boundary condition [38, 39], i.e. we take

$$\frac{\partial u^s}{\partial r} - iku^s = 0, \quad \text{on } \partial\Omega \tag{4.4}$$

rather than at $|r| \to \infty$. We note that this is a mock-up of the standard scattering setup (that typically employs far-field data), and posing Sommerfeld conditions on $\partial\Omega$ is typically not justifiable as a physical model.

The finite element mesh used has a total of 7294 nodes with 3920 in D. We take 61 equally spaced measurements around the boundary of the domain for 61 uniformly spaced directions on the unit circle, that is, $N_m = N_s = 61$. We use this approach to provide the accurate nonlinear parameter-to-observable mapping $\mathcal{G}(m)$. The noise level is set at 3% of the peak value of the noiseless measurements, that is, $\delta_e = 3\times 10^{-2} d_{\max}$, where $d_{\max}$ is (the modulus of) the maximum noiseless measurement.

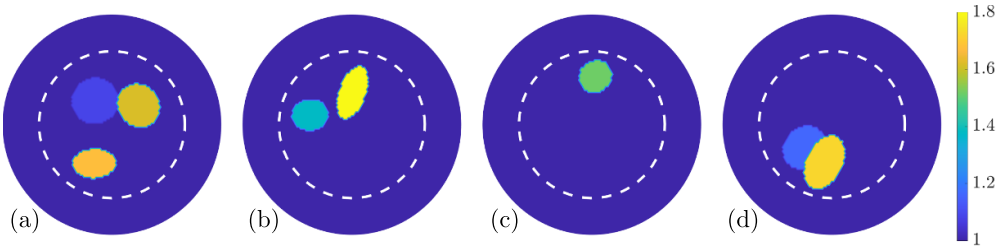

Figure 1. Four samples drawn form the prior distribution of refractive index. The white dashed line indicates the boundary of D.

The prior distribution on the refractive index used to compute the approximation error and the true resonator type refractive index are chosen to be similar to those used in [22]. Specifically, the sampled refractive indices consist of one to three high contrast ($m=1$ to $m=2$) ellipses with randomised shape and location. The number, the orientation, and the centre of the ellipses are each chosen uniformly, while the radius is chosen from a normal distribution of mean 0.25 and standard deviation of 0.07. In figure 1 we show four samples from the prior. Note that ellipses can partially overlap (occlude) each other.

Rather than employing the Born approximation, here, we employ the zero-map as the linear approximation, that is, we write $d=e+\varepsilon$ with $\mathcal{G}(m)-\mathcal{O}m$ having been absorbed in ε. Furthermore, we compute the associated joint covariances and approximated likelihood using algorithm 3.1 with $N=10$, $N=100$, $N=1000$, $N=5000$, $N=7500$, and $N=10\,000$. The (true) resonator and the estimated posterior means corresponding to different N are shown in figure 2. The reconstructions obtained with high sample numbers are similar to the resonator example in [22].

Naturally, both the estimate (reconstruction) as well as the error estimates can be assumed to increase in quality with increasing number of samples N. In figure 3, the true cross section of the target as well as the $\pm1\sigma$ and $\pm2\sigma$ error intervals are shown for different sample sizes N, where the posterior standard deviation is given by $\sigma(x)=(\mathcal{C}_{m|d}(x,x))^{\frac{1}{2}}$ for $x\in\Omega$. It is evident that using the pseudoinverse in the computation the conditional covariance $\Gamma_{\varepsilon|m}$ can yield severely underestimated posterior error estimates when small sample sized are used to approximate the covariances. This suggests to compute a large enough sample set or to use some other approach to estimate $\Gamma_{\varepsilon|m}$.

4.2. Example 2: inverse conductivity problem

We now consider an inverse conductivity type problem, which arises in a range of settings such as EIT [40, 41], and ground water flow [42, 43]. In particular, we consider a set up similar to the so-called continuum boundary condition version of the problem, see, for example, [41, 44, 45].

The data consists of (noisy) boundary measurements of the potentials u_j on $\partial\Omega$ for $j=1,2,\ldots,N_s$, which satisfy

$$\begin{aligned} -\nabla\cdot(\exp(m)\nabla u_j) &= 0 && \text{in } \Omega \\ \exp(m)\nabla u_j\cdot n &= g_j && \text{on } \partial\Omega\setminus S, \\ \exp(m)\nabla u_j\cdot n+u &= 0 && \text{on } S. \end{aligned} \tag{4.5}$$

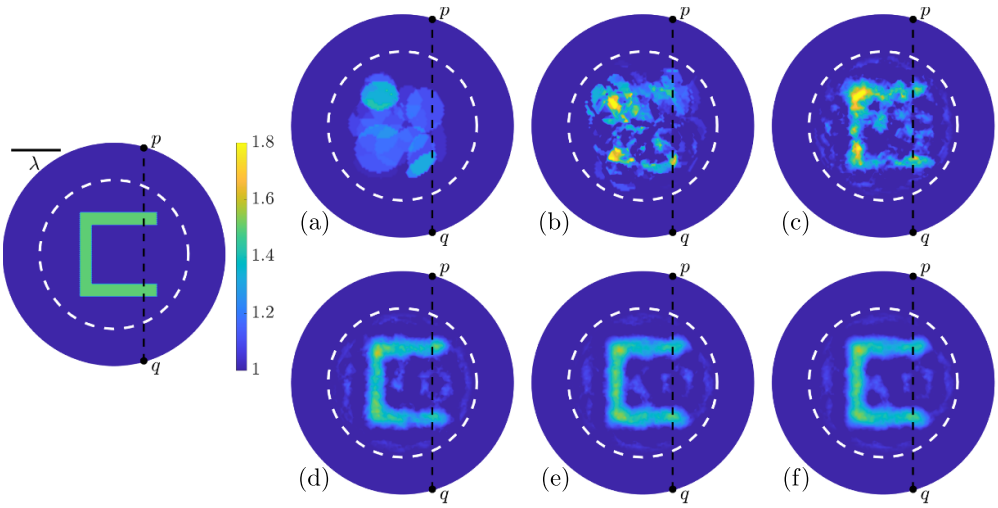

Figure 2. The true refractive index (far left) and the reconstructed (approximate conditional mean) refractive index using the zero-map as the linear approximation with $N = 10$ (a), $N = 100$ (b), $N = 1000$ (c), $N = 5000$ (d), $N = 7500$ (e), and $N = 10\,000$ (f). The wavelength λ is indicated at the top left corner of the true model.

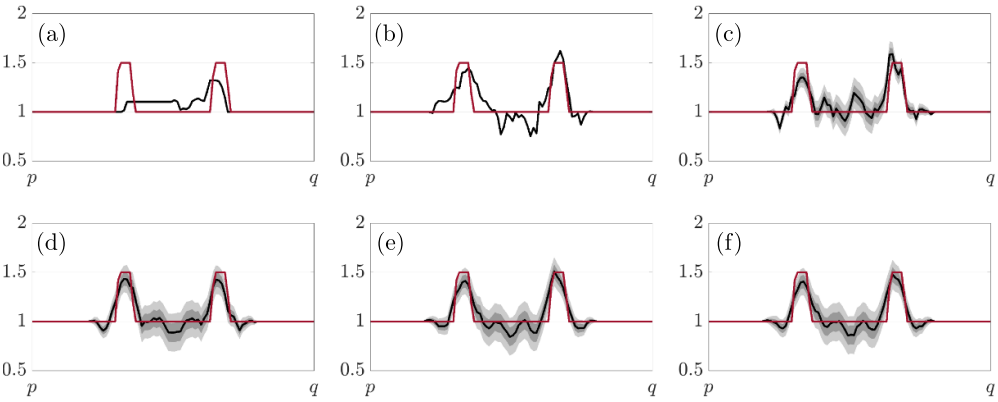

Figure 3. The approximate conditional mean along the cross-section $p - q$ (see figure 2). The red line is the true refractive index along the cross-section while the black lines show the conditional means found using the zero-map as the linear approximation. The darker grey and lighter grey show the (approximate) marginal posterior ± 1 and ± 2 standard deviations, respectively. The results are computed using $N = 10$ (a), $N = 100$ (b), $N = 1000$ (c), $N = 5000$ (d), $N = 7500$ (e), and $N = 10\,000$ (f).

In the context of EIT the above can be referred to as the *continuum model* [40, 41], where $\exp(m)$ represents the electrical conductivity, u_j the voltage, and g_j the applied boundary current. The (noiseless) data can be written as $d_{\ell j} = u_j(s_\ell)$ for $\ell = 1, 2 \ldots, N_m$ with N_m, N_s denoting the number of measurement points on the boundary and number of different applied currents, respectively. The domain Ω is modelled as starlike, meaning that its boundary can be

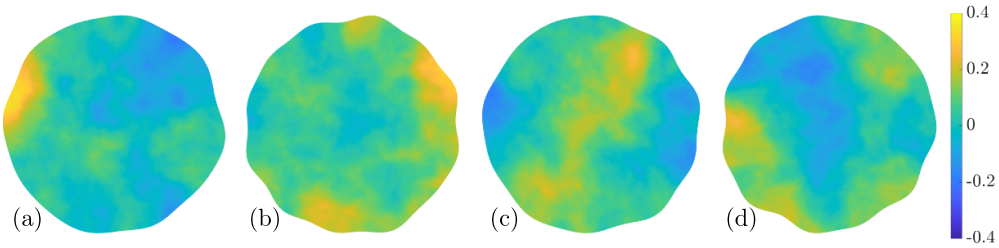

Figure 4. Four samples drawn form the joint prior distribution of the domain geometry and conductivity.

written using polar coordinates in the form $r = r(\theta)$ for $\theta \in [0, 2\pi]$. We set $S = \{(\theta, r(\theta)) | \theta \in [4.5, 4.9]\}$ and take $N_m = 16$ evenly spaced (in terms of θ) point measurements in $\partial\Omega \setminus S$. We apply $N_s = 15$ different currents of the form $g_i(\theta) = \exp(-\|\theta - i2\pi/N_s - \pi\|^2/16) - \exp(-\|\theta - i2\pi/N_s\|^2/16)$ for $j = 1, 2, \ldots N_s$. Equation (4.5) is solved using 4027 FEM nodes with linear basis functions for both the potential u and the (discretised) conductivity m.

In this example, we have an auxiliary parameter z that represents the (unknown, uninteresting) boundary geometry of the domain and write $\Omega(z)$ to underline the dependence of the domain shape on the parameterization z. As the domain is modelled as starlike it is represented numerically using a truncated Fourier series. The Fourier series is truncated after the first 25 Fourier basis functions, while the distribution of the Fourier coefficients is such that the coefficients are independent and identically distributed (iid) with normal distributions having mean 0 and standard deviation 0.005, other than the first coefficient, which has mean 1.[5] The parameter m of the conductivity distribution taken as (conditionally-)Gaussian so that for a given set of Fourier coefficients z (i.e. a given domain shape) the prior is of the form $m|z \sim \mathcal{N}(m_0, \mathcal{C}_{mm}(z))$. We take a prior mean of zero, i.e. $m_0 = 0$, and define the prior covariance operator using the inverse of an elliptic PDE operator [46, 47]. Specifically, we take

$$\mathcal{C}_{mm}(z) = \mathcal{A}(z)^{-2}, \quad \mathcal{A}(z) = \begin{cases} \gamma(\kappa I - \Delta), & \text{in } \Omega(z) \\ \nabla_n, & \text{on } \partial\Omega(z), \end{cases} \tag{4.6}$$

where Δ is the Laplacian operator, ∇_n is the directional derivative along the unit normal to the boundary, and $\gamma = 40$, $\kappa = 10$ control the marginal variance and correlation length of the random field respectively. The operator $\mathcal{A}(z)$ is discretised using FEM on the same mesh used for the forward problem. Four draws from the distribution of the geometry and conductivity are shown in figure 4. Solving (4.5) in the starlike domain (parametrised by the Fourier coefficients z) provides the accurate nonlinear parameter-to-observable mapping $\mathcal{G}(m, z)$.

We will only estimate the conductivity parameter m and premarginalise over *both* the use of the linearised model and the uncertain geometry (parametrized by the Fourier coefficients z). Here, we compute the approximate posterior mean and covariance estimates for both the zero map and the discretised Fréchet derivative linearisations. The Fréchet derivative is computed at $m = m_0$, with the boundary shape fixed to a unit circle which results from setting z

[5] Sampling from the joint prior is done by sampling the Fourier coefficients first and then the conductivity. Thus we can can ensure positivity of the radius of the domain before sampling for the conductivity or running the forward models. More specifically, any sample of the Fourier coefficients leading to a not strictly positive domain radius is resampled.

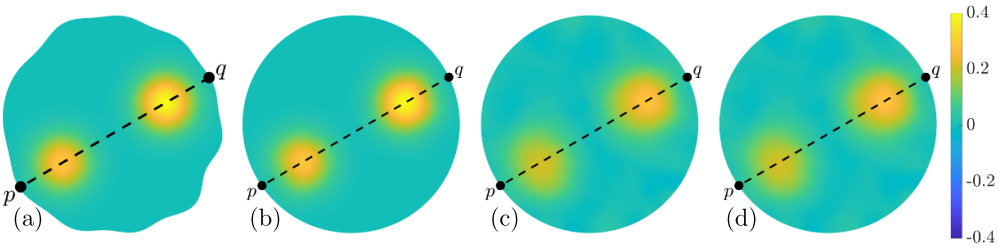

Figure 5. The actual conductivity inside the true geometry (a), the actual conductivity inside the geometry resulting for the (prior) expected value of all Fourier coefficients (b), the reconstructed (approximate conditional mean) conductivity using the (Fréchet) derivative linearisation with $N = 10\,000$ (c), and the reconstructed (approximate conditional mean) conductivity using the zero-map as the linear approximation with $N = 10\,000$ (d).

to its prior expected value. For details of the Fréchet derivative associated with (4.5) see, for example, [48]. Each linearisation induces different second order statistics for the approximation error ε. The true target and the estimates for both linearisations are shown in figure 5. With $N = 10\,000$ sample size, the estimates have essentially converged which is also suggested by the L_2 errors of the two estimates as shown below in figure 8. These computational results demonstrate the asymptotic convergence of the linearisations also in the case of auxiliary parameters.

The second main question of this example is whether the different linearisations yield significantly different convergence behaviour for the posterior error estimates. The convergence of the cross sections and the respective posterior error estimates for the zero and Fréchet linearisations (as N increases) are shown in figures 6 and 7, respectively. These figures show trends in agreement to those observed for the inverse scattering problem in section 4.1. For small numbers of samples ($N = 10$ and $N = 100$), the approximation posteriors (conditional means and posterior variances) found using either of the linearisations are essentially infeasible; the actual conductivity lies well outside the essential support of the approximate posteriors. It's worth noting, however, that the approximate posterior uncertainty for small numbers of samples appears significantly less underestimated when using the Fréchet linearisations than when using the zero map. For larger numbers of samples we see the posteriors begin to coincide.

To further investigate the convergence using different linearisations, in figure 8 we show the convergence (as N increases from 10 to 50 000) for both of the linearisations of the mean of the approximation errors ε_0, the L_2 error in the approximate conditional mean estimates, and the trace of the approximate posterior covariance matrices. The trace of the (approximate) posterior covariance matrix is one of several standard measure of the (approximate) posterior uncertainty [49]. These results were computed for five different *random seeds* and, furthermore, the reference mean of the approximation errors, denoted ε_0^*, and the reference approximate conditional mean estimate, denoted $\hat{m}^*$, are computed using 100 000 samples and the Fréchet derivative as the linearisation.

As seen in figure 8(a), for both choices of linearisation the mean of the approximation error essentially coincide and converge at the expected Monte Carlo rate of $N^{-\frac{1}{2}}$ [34]. The error in

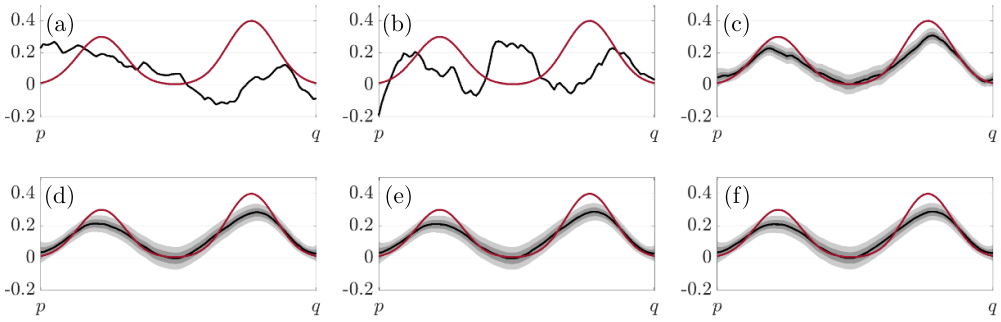

Figure 6. The approximate conditional mean along the cross-section $p-q$ (see figure 2). The red line is the true conductivity along the cross-section while the black lines show the conditional mean found using the zero-map as the linear approximation. The darker grey and lighter grey show the (approximate) marginal posterior ± 1 and ± 2 standard deviations, respectively. The results are computed using $N = 10$ (a), $N = 100$ (b), $N = 1000$ (c), $N = 5000$ (d), $N = 7500$ (e), and $N = 10\,000$ (f).

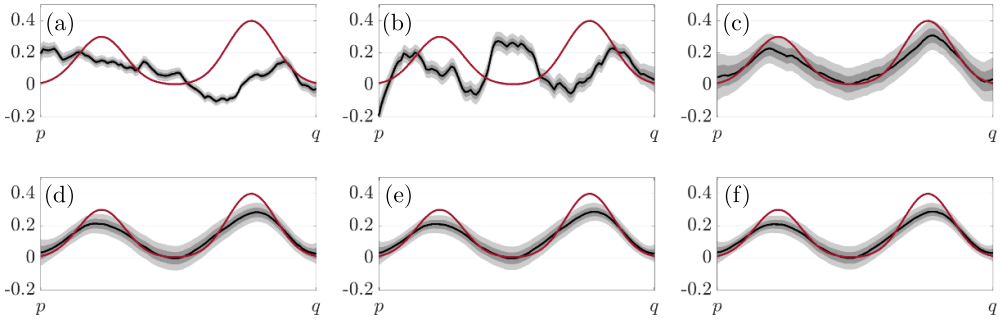

Figure 7. The approximate conditional mean along the cross-section $p-q$ (see figure 5). The red line is the true conductivity along the cross-section while the black lines show the conditional mean found using the Fréchet derivative as the linear approximation. The darker grey and lighter grey show the (approximate) marginal posterior ± 1 and ± 2 standard deviations, respectively. The results are computed using $N = 10$ (a), $N = 100$ (b), $N = 1000$ (c), $N = 5000$ (d), $N = 7500$ (e), and $N = 10\,000$ (f).

the approximate conditional mean estimates shown in figure 8(b) essentially coincide over the whole range of N, though there are some very small differences for $N = 10$ (these small differences for $N = 10$ can also be seen by comparing figures 6(a) and 7(a)). Lastly, in figure 8(c) we see that the trace of the approximate posterior covariance matrix using the zero map as the linearisation significantly under estimates the posterior uncertainty for smaller numbers of samples, while using the Fréchet derivative as the linearisation the posterior uncertainty is over estimated for smaller numbers of samples, though for $N \geqslant 10\,000$ we see the trace of the approximate posterior covariance matrices coincide for both choices of linearisation. We also note that across the five random seeds the qualitative behaviour is essentially the same.

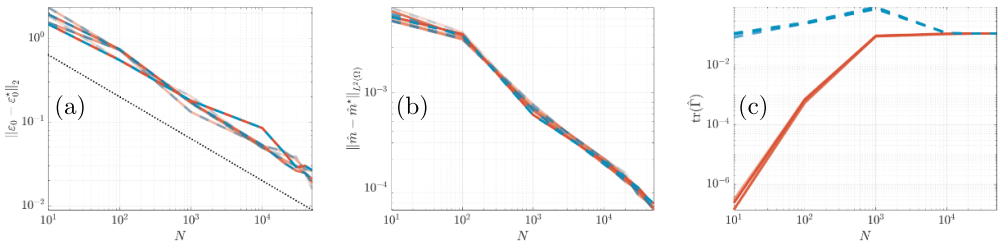

Figure 8. Convergence of the errors in the mean of the approximation error with the expected Monte Carlo convergence rate of $N^{-\frac{1}{2}}$ [34] shown as black dashed line (a). Convergence of the L_2 error in the MAP estimate (b). Convergence of the trace of the posterior covariance matrix (c). In all plots the blue lines show the convergence using the Fréchet derivative as the linear approximation while the red dashed lines show the convergence using the zero-map as the linear approximation. The results are computed using 5 different batches (with different random seeds) for the number of samples N increasing from 10 to 50 000. The estimates ε_0^* and $\hat{m}^*$ are computed using 100 000 samples.

5. Discussion and conclusions

In this paper, we considered inverse problems with nonlinear forward maps and the respective global linearisations (global affine approximations). In addition to the global linearisation, we assume that the measurement noise model and the prior model used in the inversion are Gaussian models. These approximations are common in resource-limited applications since they result in affine data-to-estimate maps as well as data-independent posterior covariance models.

We showed that when the full BAE approach is employed to compensate for the approximation error that arises due to the linearisation, measurement error and prior models, the approximate posterior model does asymptotically not depend on the choice of the linear approximation. We also illustrated numerically the result with inverse scattering and conductivity problems.

The result implies, in particular, that the asymptotic posterior models given by using the Fréchet derivative or the zero-map models as the linearisation are equal. While several works have used the full BAE with an approximate linearised solver, our result indicates that the Fréchet derivative (or any other linearisation) does not need to be constructed since the zero-map approximation yields the same posterior. That is to say, the approach could be used in a *black-box* manner, i.e. by simply running the accurate forward problem to compute the statistics of the approximation errors.

The result is, however, an asymptotic one since it relies on the actual second order joint statistics of the unknown and the approximation errors. We demonstrated that, when a large enough number of sample is used, the use of the zero-map approximation yields feasible estimates of the posterior mean and covariance. However, results also showed that the use of a too small sample size leads to both poor estimates and underestimated posterior error estimates when the pseudoinverse is used in the computation of the related conditional covariances. The latter can lead to severely misleading posterior estimates and interpretations.

The dependence of the convergence of the posterior models on the choice of linearisation was investigated for the inverse conductivity case. In this example, although the approximate

conditional mean estimates computed using the Fréchet derivative and the zero-map were fairly similar, the respective covariances were significantly different. Specifically, use of the Fréchet derivative resulted in a fairly accurate approximation of the trace of the posterior covariance matrix even for small numbers of samples, while use of the zero-map resulted in significantly under approximated posterior covariance matrix trace for small numbers of samples. For larger numbers of samples the two covariance matrices coincided as expected.

It should be pointed out, that as stated in the introduction, the results provided here do no give any immediate insight into how well the resulting approximate posterior compares to the true posterior. This of course is an interesting research question in itself.

Acknowledgment

The authors would like to thank Oliver Maclaren and Owen Dillion for helpful discussions about the manuscript as well as the anonymous reviewers for their valuable comments that contributed to improving the final version of this paper.

Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request. The data that support the findings of this study are available upon reasonable request from the authors.

Funding

This research was partially supported by US National Science Foundation DMS Grant 1654311.

ORCID iDs

Noémi Petra https://orcid.org/0000-0002-9491-0034
Umberto Villa https://orcid.org/0000-0002-5142-2559

References

[1] Peherstorfer B, Willcox K and Gunzburger M 2018 Survey of multifidelity methods in uncertainty propagation, inference and optimization *SIAM Rev.* **60** 550–91
[2] Frangos M, Marzouk Y, Willcox K and van Bloemen Waanders B 2010 *Surrogate and Reduced-Order Modeling: A Comparison of Approaches for Large-Scale Statistical Inverse Problems* (New York: Wiley) ch 7, pp 123–49
[3] Brynjarsdóttir J and O'Hagan A 2014 Learning about physical parameters: the importance of model discrepancy *Inverse Problems* **30** 114007
[4] Kaipio J P and Kolehmainen V 2013 *Approximate Marginalization Over Modeling Errors and Uncertainties in Inverse Problems* (Oxford: Oxford University Press) ch 32, pp 644–72
[5] Mozumder M, Hauptmann A, Nissila I, Arridge S R and Tarvainen T 2021 A model-based iterative learning approach for diffuse optical tomography *IEEE Trans. Med. Imaging* **41** 1289–99
[6] Sheriffdeen S, Ragusa J C, Morel J E, Adams M L and Bui-Thanh T 2019 Accelerating PDE-constrained inverse solutions with deep learning and reduced order models (arXiv:1912.08864)
[7] Lunz S, Hauptmann A, Tarvainen T, Schónlieb C-B and Arridge S 2021 On learned operator correction in inverse problems *SIAM J. Imaging Sci.* **14** 92–127

[8] Kennedy M C and O'Hagan A 2001 Bayesian calibration of computer models *J. R. Stat. Soc.* B **63** 425–64
[9] Higdon D, Kennedy M, Cavendish J C, Cafeo J A and Ryne R D 2004 Combining field data and computer simulations for calibration and prediction *SIAM J. Sci. Comput.* **26** 448–66
[10] Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems* (*Applied Mathematical Sciences* vol 160) (New York: Springer)
[11] Kaipio J and Somersalo E 2007 Statistical inverse problems: discretization, model reduction and inverse crimes *J. Comput. Appl. Math.* **198** 493–504
[12] Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559
[13] Babaniyi O, Nicholson R, Villa U and Petra N 2021 Inferring the basal sliding coefficient field for the Stokes ice sheet model under rheological uncertainty *Cryosphere* **15** 1731–50
[14] Nicholson R, Petra N and Kaipio J P 2018 Estimation of the Robin coefficient field in a Poisson problem with uncertain conductivity field *Inverse Problems* **34** 115005
[15] Brandt C and Seppänen A 2018 Recovery from errors due to domain truncation in magnetic particle imaging: approximation error modeling approach *J. Math. Imaging Vis.* **60** 1196–208
[16] Castello D A and Kaipio J P 2019 Modeling errors due to Timoshenko approximation in damage identification *Int. J. Numer. Methods Eng.* **120** 1148–62
[17] Arridge S R, Kaipio J P, Kolehmainen V, Schweiger M, Somersalo E, Tarvainen T and Vauhkonen M 2006 Approximation errors and model reduction with an application in optical diffusion tomography *Inverse Problems* **22** 175
[18] Liu D, Kolehmainen V, Siltanen S and Seppänen A 2015 A nonlinear approach to difference imaging in EIT; assessment of the robustness in the presence of modelling errors *Inverse Problems* **31** 035012
[19] Huttunen J M J and Kaipio J P 2007 Approximation errors in nonstationary inverse problems *Inverse Problems Imaging* **1** 77
[20] Tarvainen T, Kolehmainen V, Kaipio J P and Arridge S R 2010 Corrections to linear methods for diffuse optical tomography using approximation error modelling *Biomed. Opt. Express* **1** 209–22
[21] Calvetti D, Dunlop M, Somersalo E and Stuart A M 2018 Iterative updating of model error for Bayesian inversion *Inverse Problems* **34** 025008
[22] Kaipio J P, Huttunen T, Luostari T, Lähivaara T and Monk P B 2019 A Bayesian approach to improving the Born approximation for inverse scattering with high-contrast materials *Inverse Problems* **35** 084001
[23] Muhumuza K, Roininen L, Huttunen J M J and Lahivaara T 2020 A Bayesian-based approach to improving acoustic Born waveform inversion of seismic data for viscoelastic media *Inverse Problems* **36** 075010
[24] Tick J, Pulkkinen A and Tarvainen T 2019 Modelling of errors due to speed of sound variations in photoacoustic tomography using a Bayesian framework *Biomed. Phys. Eng. Express* **6** 015003
[25] Sprungk B 2020 On the local Lipschitz stability of Bayesian inverse problems *Inverse Problems* **36** 055015
[26] Goncharskii A V, Leonov A S and Yagola A G 1974 The regularization of incorrect problems with an approximately specified operator *USSR Comput. Math. Math. Phys.* **14** 195–201
[27] 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
[28] 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
[29] Da Prato G 2006 *An Introduction to Infinite-Dimensional Analysis* (Berlin: Springer Science & Business Media)
[30] Nicholson R and Kaipio J P 2020 An additive approximation to multiplicative noise *J. Math. Imaging Vis.* **62** 1–11
[31] Challis E and Barber D 2013 Gaussian Kullback–Leibler approximate inference *J. Mach. Learn. Res.* **14** 2239–86
[32] Blei D M, Kucukelbir A and McAuliffe J D 2017 Variational inference: a review for statisticians *J. Am. Stat. Assoc.* **112** 859–77
[33] Arridge S R, Ito K, Jin B and Zhang C 2018 Variational Gaussian approximation for poisson data *Inverse Problems* **34** 025005
[34] Robert C P and Casella G 1999 *Monte Carlo Statistical Methods* vol 2 (Berlin: Springer)

[35] Colton D, Coyle J and Monk P 2000 Recent developments in inverse acoustic scattering theory *SIAM Rev.* **42** 369–414
[36] Colton D and Kress R 2019 *Inverse Acoustic and Electromagnetic Scattering Theory* vol 93 (Berlin: Springer)
[37] Kirsch A 2011 *An Introduction to the Mathematical Theory of Inverse Problems* vol 120 (Berlin: Springer Science & Business Media)
[38] Feng K 1983 Finite element method and natural boundary reduction *Proc. Int. Congress of Mathematicians* pp 1439–53
[39] Ihlenburg F 2006 *Finite Element Analysis of Acoustic Scattering* vol 132 (Berlin: Springer Science & Business Media)
[40] Mueller J L and Siltanen S 2012 *Linear and Nonlinear Inverse Problems with Practical Applications* (Philadelphia, PA: SIAM)
[41] Borcea L 2002 Electrical impedance tomography *Inverse Problems* **18** R99
[42] Mattis S A, Butler T D, Dawson C N, Estep D and Vesselinov V V 2015 Parameter estimation and prediction for groundwater contamination based on measure theory *Water Resour. Res.* **51** 7608–29
[43] Bear J 2013 *Dynamics of Fluids in Porous Media* (Chelmsford, MA: Courier Corporation)
[44] Astala K and Päivärinta L 2006 Calderón's inverse conductivity problem in the plane *Ann. Math.* **163** 265–99
[45] Nachman A I 1996 Global uniqueness for a two-dimensional inverse boundary value problem *Ann. Math.* **163** 71–96
[46] Lindgren F, Rue H and Lindström J 2011 An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach *J. R. Stat. Soc.* B **73** 423–98
[47] Roininen L, Huttunen J M J and Lasanen S 2014 Whittle–Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography *Inverse Problems Imaging* **8** 561
[48] Lechleiter A and Rieder A 2008 Newton regularizations for impedance tomography: convergence by local injectivity *Inverse Problems* **24** 065009
[49] Alexanderian A 2021 Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: a review *Inverse Problems* **37** 043001