Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Approximation Errors and Model Reduction in Optical Tomography

V. Kolehmainen[§], S. R. Arridge[†], J.P. Kaipio[§], M. Schweiger[†], E. Somersalo[*], T. Tarvainen [§] and M. Vauhkonen[§]

***Abstract*— Model reduction is often required in optical diffusion tomography (ODT), typically due to limited available computation time or computer memory. In practice, this often means that we are bound to use sparse meshes in the model for the forward problem. Conversely, if we are given more and more accurate measurements, we have to employ increasingly accurate forward problem solvers in order to exploit the information in the measurements. In this paper we apply the approximation error theory to ODT. We show that if the approximation errors are estimated and employed, it is possible to use mesh densities that would be unacceptable with a conventional measurement model.**

I. Introduction

In optical diffusion tomography (ODT), a body $\Omega \in \mathbb{R}^d$ $(d = 2, 3)$ with unknown optical properties is illuminated with known near-infrared sources and a vector $y \in \mathbb{R}^m$ of measurements of the scattered and transmitted light is collected on the body's surface $\partial\Omega$. The inverse problem is to estimate the pair of functions $x = (\mu_{\rm a}, \mu_{\rm s})$ representing the optical absorption and scattering coefficients of the body based on the data y, given the measurement model

$$y = A(x) + e, \tag{1}$$

where e denotes the measurement noise and A is a known non-linear mapping, representing the solution of the diffusion (or transport) equation.

Since in general, no analytic solutions can be found, the forward model (1) is approximated by a discrete numerical scheme, e.g., finite element (FEM) approximation. Typically, the distributed parameter x is approximated by a representation in finite dimensional basis and the continuous model (1) is then replaced by an approximate equation,

$$y = A_h(x_h) + e, \tag{2}$$

where $x_h \in \mathbb{R}^n$ is a vector containing the degrees of freedom of the finite dimensional approximation of x, and $h > 0$ is a mesh parameter controlling the level of discretization. The theory of finite element method guarantees that for standard FEM discretizations, $A_h(x_h) \to A(x)$ as $h \to 0+$, i.e., the approximation (2) becomes exact within the measurement accuracy.

[§] Department of Physics, University of Kuopio, P.O.Box 1627, 70211 Kuopio, Finland Ville.Kolehmainen@uku.fi

[†] Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK

[*] Institute of Mathematics, Helsinki University of Technology, Otakaari 1, 02015 TKK, Finland

This work was supported by the Academy of Finland (projects 203985, 108299 and 206449)

From the point of view of inverse problems, the convergence of the forward model alone is not necessarily sufficient. As $h \to 0+$, the dimensionality of the approximation x_h usually increases, i.e., $n \to \infty$. This means that the complexity of the inverse problem of estimating x_h increases as the approximation improves. Hence, when the forward model is accurate, the inverse problem may be prohibitively large to be tackled by any computational scheme. On the other hand, if the forward model is inaccurate, the discretization error may become significant compared to the measurement error. Together with the fact that the inverse problem is ill-posed, the approximation error may destroy the quality of the estimate of x_h. This dichotomy is one of the bottlenecks of diffuse tomographic methods, in particular in three spatial dimensions, where the computational complexity is an issue even when relatively coarse meshes are used.

To overcome this dichotomy, we employ the Bayesian statistical inversion theory The key idea in this paper is to represent not just the measurement noise, but also computational model inaccuracy as a random variable. Hence, instead of the model (2), we write an *accurate* model,

$$y = A_h(x_h) + \big[A(x) - A_h(x_h)\big] + e = A_h(x_h) + \varepsilon(x) + e, \tag{3}$$

where the term $\varepsilon(x)$ is the modelling error. Since in the Bayesian paradigm, all variables, x included, are random variables, we can determine the probability distribution of the modelling error and thus treat it as noise.

The paper is organized as follows: Section II summarizes the Bayesian framework of inverse problems and the statistical approximation error theory. Section III contains computed examples where the theory is applied to the ODT inverse problem.

II. Methods

A. Inverse problems in Bayesian framework.

In the Bayesian approach, the inverse problems are viewed as problems of statistical inference. Hence, all variables are modelled as random variables, and measurements are used to infer on the probability distribution of the parameter of primary interest. We assume that the estimated parameter x_h and data y are random variables in finite dimensional spaces $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. In the Bayesian framework, the solution of the inverse problem is the posterior probability density $\pi(x_h \mid y)$:

$$\pi(x_h \mid y) = \frac{\pi(x_h)\pi(y \mid x_h)}{\pi(y)}, \quad y = y_{\rm measured}, \tag{4}$$

where $\pi(x_h)$ the prior probability density and $\pi(y \mid x_h)$ the likelihood density. In particular, if the noise is additive and the noise e and the unknown x_h are mutually independent, formula (2) leads to the likelihood density

$$\pi(y \mid x_h) = \pi_{\text{noise}}(y - A_h(x_h)),$$

where π_{noise} is the probability distribution of the noise e. Further, if x_h and e are mutually independent and Gaussian, $x_h \sim \mathcal{N}(x_{h*}, \Gamma_{x_h})$, $e \sim \mathcal{N}(e_*, \Gamma_e)$, the posterior density (4) becomes

$$\pi(x_h|y) \propto \exp\left(-\frac{1}{2}(x_h - x_{h*})^{\mathrm{T}}\Gamma_{x_h}^{-1}(x_h - x_{h*}) \right.$$
$$\left. -\frac{1}{2}(y - A_h(x_h) - e_*)^{\mathrm{T}}\Gamma_e^{-1}(y - A_h(x_h) - e_*)\right), \qquad (5)$$

where $x_{h*} \in \mathbb{R}^n$ and $e_* \in \mathbb{R}^m$ are the means and the symmetric positive definite matrices $\Gamma_{x_h} \in \mathbb{R}^{n\times n}$ and $\Gamma_e \in \mathbb{R}^{m\times m}$ are the covariance matrices, respectively.

B. Approximation errors

In this section we develop the central technique to treat the approximation errors. The starting point is the accurate model (1), where x is a distributed parameter. To avoid theoretical difficulties related to non-linear infinite dimensional problems, we assume that the model $x \mapsto A(x)$ can be approximated in a satisfactory manner by a densely discretized finite dimensional model

$$\mathbb{R}^N \to \mathbb{R}^m, \quad x_\delta \mapsto A_\delta(x_\delta), \quad \delta > 0 \text{ small.}$$

Hence, the discretized model that is exact within the measurement accuracy is

$$y = A_\delta(x_\delta) + e, \quad x_\delta \in \mathbb{R}^N, \quad N = N_\delta. \qquad (6)$$

Consider now a model reduction: choose a discretization parameter $h > \delta$, and let $x_h \in \mathbb{R}^n$, $n = n_h < N$, denote the representation of the distributed parameter in this coarse mesh. Assume further that there is a linear model reduction map,

$$P : \mathbb{R}^N \to \mathbb{R}^n, \quad x_\delta \mapsto x_h.$$

Further, let $A_h : \mathbb{R}^n \to \mathbb{R}^m$ denote the discretized forward map in the coarse mesh. We write the exact reduced model as

$$y = A_h(x_h) + \big[A_\delta(x_\delta) - A_h(x_h)\big] + e = A_h(x_h) + \varepsilon(x_\delta) + e. \qquad (7)$$

Within the deterministic inversion paradigm there are few means to estimate the approximation error $\varepsilon(x_\delta)$, except possibly for the upper bound of its norm. The same applies to the frequentist statistical paradigm as well. However, in the Bayesian paradigm, where x_δ is modelled as a random variable with the prior probability distribution, techniques for estimating the modelling error distribution are readily available.

We can define the approximation error problem as follows: *given the probability density of (x_δ, e), the model reduction operator P and the forward models A_δ and A_h, derive a computational model for the posterior density $\pi(x_h \mid y)$.*

C. Implementation

In the Gaussian linear case the approximation error problem can be solved exactly by using the results in [1]. Nonlinear forward models or non-Gaussian distributions usually require stochastic simulations.

Let $\pi(x_\delta)$ be the prior probability density in $\mathbb{R}^N$ corresponding to the accurate discrete model (6). We can generate a sample of vectors in $\mathbb{R}^N$,

$$S = \big\{x_\delta^{(1)}, x_\delta^{(2)}, \ldots, x_\delta^{(L)}\big\},$$

such that these vectors are distributed according to the prior density. If the prior is Gaussian or other standard parametric distribution, efficient random vector generators can be used. More generally, the sample can be generated by using Markov chain Monte Carlo (MCMC) techniques. Let $\pi_{\text{noise}}(e)$ denote the probability distribution of the additive noise vector $e \in \mathbb{R}^m$. The probability density of the random vector

$$n = \varepsilon(x_\delta) + e = \big[A_\delta(x_\delta) - A_h(Px_\delta)\big] + e$$

can be written as

$$\pi(n) = \int_{\mathbb{R}^N} \pi_{\text{noise}}\big(n - \big[A_\delta(x_\delta) - A_h(Px_\delta)\big]\big)\pi(x_\delta)\mathrm{d}x_\delta,$$

where we have used the relation $\pi(n) = \int_{\mathbb{R}^N} \pi(n \mid x_\delta)\pi(x_\delta)\mathrm{d}x_\delta$. Using the set S of samples, we have the approximation as a sample average,

$$\pi(n) \approx \frac{1}{L}\sum_{\ell=1}^{L} \pi_{\text{noise}}\big(n - \big[A_\delta(x_\delta^{(\ell)}) - A_h(Px_\delta^{(\ell)})\big]\big).$$

With a similar reasoning, one can find an expression for the sample based approximation for the joint probability distribution of (x_h, y) and thus for the posterior distribution.

The above expression is not very useful except for generating samples of the noise vector n. In practical calculations, we use a Gaussian prior and a Gaussian approximation for $\pi(n)$. The Gaussian approximation is both easy to construct and easy to exploit when compared to more complicated models. In most cases this is what we would first try since this model often leads to efficient computational implementation for the inverse solver and efficiency is what we are after in the first place.

To this end, we calculate first the sample based approximations for the mean and covariance of the noise vector $n = \varepsilon + e$. We have

$$\mathrm{E}\{n\} \approx \frac{1}{L}\sum_{\ell=1}^{L}\big[A_\delta(x_\delta^{(\ell)}) - A_h(Px_\delta^{(\ell)})\big] + e_* = \varepsilon_* + e_*,$$

and $\mathrm{cov}(n) = \Gamma_n = \Gamma_e + \Gamma_\varepsilon$, where Γ_ε is estimated as

$$\Gamma_\varepsilon \approx \frac{1}{L}\sum_{\ell=1}^{L} \xi_\varepsilon^{(\ell)}\xi_\varepsilon^{(\ell)\mathrm{T}}, \quad \xi_\varepsilon^{(\ell)} = \big[A_\delta(x_\delta^{(\ell)}) - A_h(Px_\delta^{(\ell)}) - \varepsilon_*\big]$$

The Gaussian approximation for the noise covariance is thus

$$\pi(n) \propto \exp\left(-\frac{1}{2}(n - \varepsilon_* - e_*)^{\mathrm{T}}\Gamma_n^{-1}(n - \varepsilon_* - e_*)\right). \quad (8)$$

To simplify the analysis further, we write an approximation where the mutual dependence of x_h and the approximation error is ignored. By (5), this leads to an approximation that is referred to as *enhanced error model* ([1]),

$$\pi(x_h \mid y) \propto \exp\left(-\frac{1}{2}(x_h - x_{h*})^{\mathrm{T}}\Gamma_{x_h}^{-1}(x_h - x_{h*})\right.$$
$$\left.-\frac{1}{2}(y - A(x_h) - \varepsilon_* - e_*)^{\mathrm{T}}\Gamma_n^{-1}(y - A(x_h) - \varepsilon_* - e_*)\right),$$

where $x_{h*} = P\mathrm{E}\{x_\delta\} \in \mathbb{R}^n$, $\Gamma_{x_h} = P\Gamma_{x_\delta}P^{\mathrm{T}} \in \mathbb{R}^{n\times n}$. If we write the Cholesky factorizations $L_{x_h}^{\mathrm{T}}L_{x_h} = \Gamma_{x_h}^{-1}$ and $L_{e+\varepsilon}^{\mathrm{T}}L_{e+\varepsilon} = \Gamma_n^{-1}$, the computation of the Maximum A Posteriori (MAP) estimate, within the validity of the approximations above, amounts to minimizing the functional

$$F(x_h) = \|L_{e+\varepsilon}(y - A_h(x_h) - \varepsilon_* - e_*)\|^2 + \|L_{x_h}(x_h - x_{h*})\|^2. \quad (9)$$

This computational problem is superficially of the same form as the computation of a MAP estimate with the conventional likelihood (measurement) model or a Tikhonov regularized solution (The MAP estimation problem with conventional likelihood model is obtained from (9) by replacing $L_{e+\varepsilon}$ by L_e and setting $\varepsilon_* = 0$). Thus the enhanced error model (9) is appealing because we incorporate the model inaccuracy into the same framework as our general machinery for MAP estimation.

III. Numerical results

We evaluate the enhanced error model in the case $\Omega \subset \mathbb{R}^2$. The object domain is a circle with radius of 25mm. The measurement setup in the simulations consists of 16 equivspaced suorces and detectors on the boundary $\partial\Omega$. With this setup, the vector of frequency domain ODT data is $y \in \mathbb{R}^{512}$, where the complex valued data has been splitted to the amplitude and phase parts.

In this study, the forward map A_h is based on FEM-solution of the diffusion equation. For the finite element discretization, we create three different meshes with triangular elements. The mesh h^0 (number of nodes $N_n = 5625$, number of elements $N_e = 10736$) is used for the simulation of the measurement data, h^1 ($N_n = 4217$, $N_e = 7920$) is the mesh for the accurate forward model (i.e., $A_{h^1}(x_h) = A_\delta(x_\delta)$) and h^2 ($N_n = 713$, $N_e = 1296$) is the mesh for the target model. For the representation of the functions $(\mu_\mathrm{a}, \mu_\mathrm{s})$ we divide the domain Ω into $n_p = 524$ pixels, leading to $x_\delta \in \mathbb{R}^{1048}$. In this study, we use the same 524-pixel representation in both, the accurate and target models, and therefore we have the projection operator $P = I$. The MAP estimation problem (9) is solved by a Gauss-Newton algorithm with an explicit line search algorithm.

As the prior model in the simulations we use a proper Gaussian smoothness prior

$$\pi(x_\delta) \propto \exp\{-\frac{1}{2}\|L_{x_\delta}(x_\delta - x_{\delta*})\|^2\}. \quad (10)$$

The details on the construction and tuning of the prior model can be found in [2]. Figure 1 shows one random draw (absorption left, scatter right) from the prior model

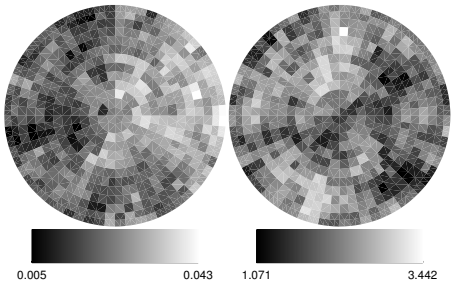

Fig. 1. One random draw from the proper smoothness prior model (10). Left:μ_a. Right:μ_s

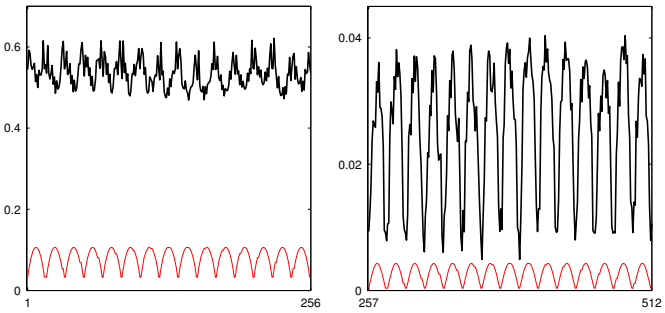

Fig. 2. Example of the approximation error. The absolute value of the mean of the approximation errors $|\varepsilon_*|$ between the accurate model A_{h^1} ($N_n = 4217$, $N_e = 7920$) and the target model A_{h^2} ($N_n = 713$, $N_e = 1296$) is shown with thick line. The thin line shows the standard deviations $\sigma_{e,j}$ of the random noise e corresponding to the relative noise level $\delta_e = 0.5\%$. Left image shows the amplitude part and right image phase shift part of the vector $|\varepsilon_*|$.

$\pi(x_\delta)$. The mean and covariance of the approximation error ε between the meshes h^1 and h^2 are estimated as sample averages over a ensemble of 2500 random draws from the prior (10) using the implementation explained in section II-C. Figure 2 shows the abolute value of the mean ε_* of the approximation errors. The thick lines in the figure show the $|\varepsilon_*|$ which was computed using the accurate forward model A_{h^1} and the target model A_{h^2} (amplitude on left, phase shift on right). The thin line shows the standard deviations of the random noise e with the relative noise level $\delta_e = 0.5\%$. As can be seen, the mean of the channelwise approximation errors exceed clearly the typical levels of random noise for ODT. Based on the figure, we would expect to get poor reconstructions if we were to use target model A_{h^2} with the conventional likelihood model.

Next, we consider the estimation errors

$$\varrho(\hat{x_h}) = \mathrm{E}\{\|x_h - \hat{x_h}\|^2\}(\mathrm{tr}\,\Gamma_{x_h} + \|x_{h*}\|^2)^{-1}$$

for the MAP-estimates with respect the level of random noise in the data. The estimation errors are computed as sample averages of 100 reconstructions computed from simulated data from set X of 100 i.i.d. samples from the prior $\pi(x_\delta)$. Note that the samples X were not included in the set S used for estimation of the statistics of ε. The simulated measurements are computed in the mesh h^0 ($N_n = 5625$) and random noise is added to the simulated data vectors as $y = A_{h^0}(x) + e$, where realisations of e are drawn from the

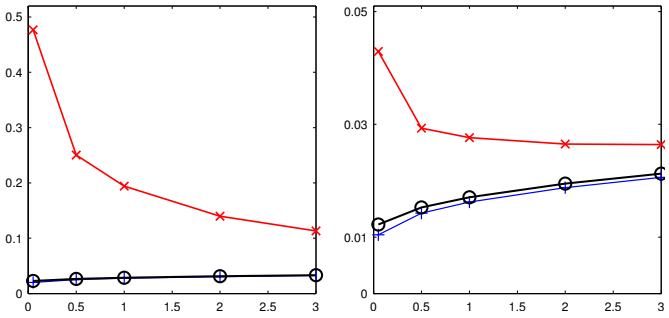

Fig. 3. The expected errors $E\{\|\mu_{\mathrm{a},MAP} - \mu_{\mathrm{a}}\|^2\}(\mathrm{trace}(\Gamma_{\mu_{\mathrm{a}}}) + \|\mu_{\mathrm{a}*}\|^2)^{-1}$ (left) and $E\{\|\mu_{\mathrm{s},\mathrm{MAP}} - \mu_{\mathrm{s}}\|^2\}(\mathrm{trace}(\Gamma_{\mu_{\mathrm{a}}}) + \|\mu_{\mathrm{s}*}\|^2)^{-1}$ (right) in the reconstructed images with respect the noise level δ_e (%). The error with conventional likelihood model in the dense computation mesh h^1 ($N_n = 4217$) is shown with (+), the error with the conventional likelihood model in the sparse computation mesh h^2 ($N_n = 713$) is shown with (x), and the error with the enhanced error model in the sparse mesh h^2 ($N_n = 713$) is shown with (○).

distribution $\mathcal{N}(0, \Gamma_e)$ with $\Gamma_e = \mathrm{diag}(\sigma_{e,1}^2, \ldots, \sigma_{e,m}^2)$ and $\sigma_{e,j} = \delta_e |y_{*,j}|/100$. (i.e, we assume that the channel wise standard deviation of the noise is δ_e percent of the absolute value of the computed data).

The estimation errors $\varrho(\hat{x_h})$ are computed for three different reconstructions: 1) MAP with the enhanced error model (9) using target model A_{h^2}, 2) MAP with conventional likelihood model using the accurate model A_{h^1} and 3) MAP with conventional likelihood model using the target model A_{h^2}. The estimation errors $\varrho(\mu_{\mathrm{a}})$ and $\varrho(\mu_{\mathrm{s}})$ with different noise levels δ_e between 3% and 0.05% are shown in Fig. 3. When the target model $A_{h^2}(x_h)$ and conventional likelihood model is used, the estimation errors increase although the noise level decreases. On the other hand, the estimate errors decrease monotonously with both the accurate forward model $A_{h^1}(x_h)$ using the conventional likelihood model and the target model $A_{h^2}(x_h)$ with the enhanced error model.

As explained above and shown in Fig. 1, the prior probability distribution is concentrated on coefficients with smooth spatial distribution, and moreover, the absorbtion and scattering coefficients are modelled as mutually independent random variables. We expect that for such coefficients, the prior model is reliable and the estimates are reasonable. When the estimation algorithm is applied to simulated or real data corresponding to coefficients that have a low probability with respect to the prior distribution, the estimates can be poor. This question is pronounced when the approximation error model is employed since the statistics of the modelling error is inferred from the prior distribution.

To avoid too optimistic results and to assess the robustness of the approximation error theory, we compute the MAP estimates when the actual spatial coefficient distributions are discontinuous and thus in conflict with the belief expressed by the prior density. The MAP estimates are again computed with the accurate model $A_{h^1}(x_h)$ ($N_n = 4217$) with conventional likelihood model and the 713-node target model $A_{h^2}(x_h)$ with enhanced error model and the conventional likelihood model. The actual spatial coefficient distribution

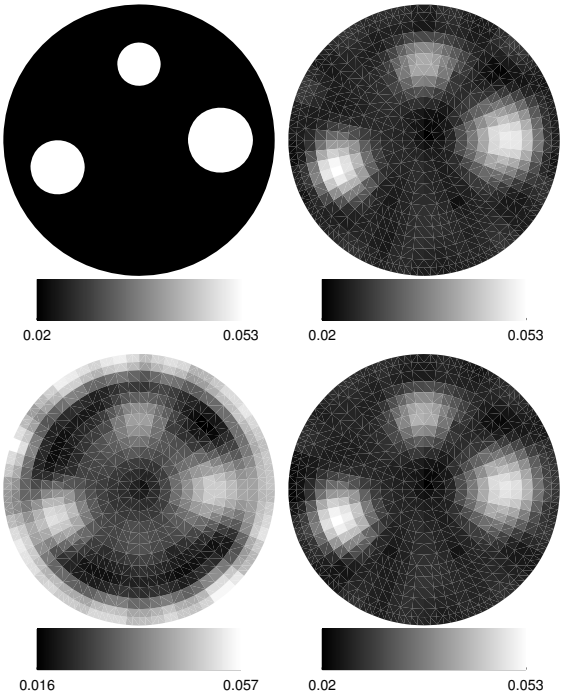

Fig. 4. Reconstructed absorption images μ_{a}. Top left: The target. Top right: Reconstruction with conventional likelihood model using the accurate forward model $A_{h^1}(x_h)$ ($N_n = 4217$). Bottom left: Reconstruction with the conventional likelihood model using the target model $A_{h^2}(x_h)$ ($N_n = 713$). Bottom right: Reconstruction with the enhanced error model using the target model $A_{h^2}(x_h)$ ($N_n = 713$). The relative noise level was $\delta_e = 0.5\%$.

and the estimates for the absorption coefficient are shown in Fig. 4. The results with the accurate model (top right) and the target model with enhanced error model (bottom right) are essentially equivalent while the target model with conventional likelihood model (bottom left) exhibits severe annular estimation errors, especially near the boundary.

IV. Conclusions

We have shown that the accuracy of the computational model for the forward problem can be relaxed if the approximation error model is used in optical diffusion tomography. The setting up of the approximation error model is a computationally intensive task while the use of the model is as with the conventional error model. The approach is especially attractive when the noise levels tend to be small, in which cases the approach allows for the exploitation of the actual accuracy of the measurements with computationally efficient forward models. It was shown that at least in the studied case, the approach is tolerant even to a qualitative misspecification of the prior model.

References

[1] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, New York, 2005.

[2] S.R. Arridge, J.P. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen and M. Vauhkonen. Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Problems*, 22:175–195, 2006.