Approximation errors and model reduction with an application in optical diffusion tomography

S R Arridge[1]**, J P Kaipio**[2]**, V Kolehmainen**[2]**, M Schweiger**[1]**, E Somersalo**[3]**, T Tarvainen**[2] **and M Vauhkonen**[2]

[1] Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK
[2] Department of Applied Physics, University of Kuopio, PO Box 1627, 70211 Kuopio, Finland
[3] Institute of Mathematics, Helsinki University of Technology, Otakaari 1, 02015 TKK, Finland

E-mail: Jari.Kaipio@uku.fi

Abstract

Model reduction is often required in several applications, typically due to limited available time, computer memory or other restrictions. In problems that are related to partial differential equations, 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. Optical diffusion tomography (ODT) is an example in which the typical required accuracy for the forward problem solver leads to computational times that may be unacceptable both in biomedical and industrial end applications. In this paper we review the approximation error theory and investigate the interplay between the mesh density and measurement accuracy in the case of optical diffusion tomography. 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.

1. Introduction

Consider the following inverse problem with additive noise model: estimate a parameter x from noisy data y,

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

where e denotes the noise and A is a known mapping. The data vector y is assumed to be finite dimensional, $y \in \mathbb{R}^m$, while the function x is a distributed parameter. The mapping A may be linear or non-linear, the focus in this paper being on non-linear models. For the sake of definiteness, let us specify the inverse problem as that of optical diffusion tomography

(ODT): a bounded body $\Omega \in \mathbb{R}^d, d = 2, 3$, with unknown optical properties is illuminated with known near-infrared sources and the scattered and transmitted light is measured on the body's surface. In this application, the estimated variable x is considered to be a pair of real functions (μ_a, μ_s) representing the optical absorption and scattering coefficients, and the forward model is a diffusion equation. For a general reference of this problem, see [1].

The forward model in optical tomography is defined by a partial differential equation. The distributed parameter that we are interested in appears as a non-constant coefficient of the 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 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.

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,

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

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 2 summarizes the Bayesian framework of inverse problems and the statistical approximation error theory. An overview of the application, the optical diffusion tomography, is given in section 3. Section 4 contains computed examples where the statistical theory is applied to the ODT inverse problem.

2. Methods

In this section, we give a brief review of the Bayesian framework for inverse problems and apply it to the approximation error analysis. For general accounts on statistical inversion theory, see for example the recent books [2, 3]. For the basics of approximation error theory see [2, 4].

2.1. 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 x_h and y are random variables in finite-dimensional spaces $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Assume that their joint probability distribution in $\mathbb{R}^n \times \mathbb{R}^m$ is absolutely continuous with respect to the Lebesgue measure, and hence expressible in terms of a joint probability density $\pi(x_h, y)$. The key identity relates the conditional probability densities to the joint probability density,

$$\pi(x_h, y) = \pi(x_h)\pi(y \mid x_h) = \pi(y)\pi(x_h \mid y), \tag{4}$$

where the marginal probability densities are

$$\pi(x_h) = \int_{\mathbb{R}^m} \pi(x_h, y)\,\mathrm{d}y, \qquad \pi(y) = \int_{\mathbb{R}^n} \pi(x_h, y)\,\mathrm{d}x_h.$$

If x_h is the variable to be estimated based on the observed values of y, we call $\pi(x_h)$ the prior probability density and $\pi(y \mid x_h)$ the likelihood density. In this framework, the solution of the inverse problem is the posterior probability density $\pi(x_h \mid y)$ that according to (4) is

$$\pi(x_h \mid y) = \frac{\pi(x_h)\pi(y \mid x_h)}{\pi(y)}, \qquad y = y_{\mathrm{measured}}, \tag{5}$$

which is the well-known Bayes' formula. 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

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

where π_{noise} is the probability distribution of the noise e. A particular but versatile special case is when x_h and e are mutually independent and Gaussian,

$$x_h \sim \mathcal{N}(x_{h*}, \Gamma_{x_h}), \qquad e \sim \mathcal{N}(e_*, \Gamma_e),$$

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. In this case, the posterior density (5) becomes

$$\pi(x_h \mid y) \propto \exp\left(-\tfrac{1}{2}(x_h - x_{h*})^{\mathrm{T}}\Gamma_{x_h}^{-1}(x_h - x_{h*})\right.$$
$$\left. - \tfrac{1}{2}(y - A_h(x_h) - e_*)^{\mathrm{T}}\Gamma_e^{-1}(y - A_h(x_h) - e_*)\right). \tag{6}$$

Observe that this equation applies only if x_h and e are mutually independent. If this is not the case, no closed formula for the posterior density can be written. A notable exception is when the mapping A_h is linear, i.e., we have the linear additive model

$$y = A_h x_h + e.$$

In this case, the joint probability density is the Gaussian, with mean and covariance

$$\mathrm{E}\left\{\begin{pmatrix} x_h \\ y \end{pmatrix}\right\} = \begin{pmatrix} \mathrm{E}\{x_h\} \\ \mathrm{E}\{y\} \end{pmatrix} = \begin{pmatrix} x_{h*} \\ A_h x_{h*} + e_* \end{pmatrix},$$
$$\mathrm{cov}\begin{pmatrix} x_h \\ y \end{pmatrix} = \begin{pmatrix} \Gamma_{x_h} & \Gamma_{x_h y} \\ \Gamma_{y x_h} & \Gamma_y \end{pmatrix} = \begin{pmatrix} \Gamma_{x_h} & \Gamma_{x_h}A_h^{\mathrm{T}} + \Gamma_{x_h e} \\ A_h\Gamma_{x_h} + \Gamma_{e x_h} & A_h\Gamma_{x_h}A_h^{\mathrm{T}} + \Gamma_{x_h e}A_h^{\mathrm{T}} + A_h\Gamma_{x_h e} + \Gamma_e \end{pmatrix},$$

where Γ_{x_h}, Γ_y and Γ_e denote the covariance matrices of the random vectors x_h, y and e, respectively, and the double subindex notation such as $\Gamma_{x_h y}$ means the cross covariance of x_h and y defined as

$$\Gamma_{x_h y} = \mathrm{E}\{(x_h - \mathrm{E}\{x_h\})(y - \mathrm{E}\{y\})^{\mathrm{T}}\} \in \mathbb{R}^{n \times m},$$

and the other cross correlations correspondingly. The posterior density in this case is also the Gaussian with mean and covariance

$$E\{x_h \mid y\} = x_{h*} + \Gamma_{x_h y} \Gamma_y^{-1} (y - E\{y\}) \tag{7}$$

$$\text{cov}(x_h \mid y) = \widetilde{\Gamma}_{x_h} \tag{8}$$

where $\widetilde{\Gamma}_{x_h} = \Gamma_{x_h} - \Gamma_{x_h y} \Gamma_y^{-1} \Gamma_{y x_h}$ is the Schur complement of Γ_{x_h}.

2.2. 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. There are very few general results concerning the statistical inversion theory in infinite-dimensional spaces. The notable exception is when the variables are Gaussian and the forward map is linear. For the Hilbert space theory, see, e.g., [5], and for more general distribution space theory, [6]. To avoid theoretical difficulties, we assume that the continuous model $x \mapsto A(x)$ can be approximated in a satisfactory manner by a densely discretized finite-dimensional model

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

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

$$y = A_\delta(x_\delta) + e, \qquad x_\delta \in \mathbb{R}^N, \qquad N = N_\delta. \tag{9}$$

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,

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

Furthermore, 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

$$y = A_h(x_h) + [A_\delta(x_\delta) - A_h(x_h)] + e = A_h(x_h) + \varepsilon(x_\delta) + e. \tag{10}$$

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 $\pi(x_h \mid y)$.*

In the literature, methods of treating errors in the forward model have been discussed. A classical problem, known as total least squares (TLS), considers linear models where the forward map is of the form $A = A_o + E : \mathbb{R}^n \to \mathbb{R}^m$, where A_o is a known matrix and E is unknown (see, e.g., [7, 8]). Such problems are encountered, e.g. in blind deconvolution. Since in the present case, the dimensions of x_h and x_δ are not necessarily the same, while the mappings A_h and A_δ are known, the TLS is not applicable, even in the framework of local linearization. However, the modelling error approach discussed here can be used as an alternative for resolving the TLS problems from the Bayesian viewpoint.

2.3. Implementation

In the Gaussian linear case, the approximation error problem can be solved exactly by using the results of subsection 2.1. These results have been presented in [4] and not repeated here. 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 (9). We can generate a sample of vectors in $\mathbb{R}^N$,

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

such that these vectors are distributed according to the prior density. If the prior is Gaussian or another 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

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

can be written as

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

and, using the sample S, we have the approximation as a sample average,

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

With a similar reasoning, one can easily 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 approximation for the noise covariance. The Gaussian approximation for the approximation error and the unknown 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

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

and

$$\mathrm{cov}(n) \approx \frac{1}{L} \sum_{\ell=1}^{L} \left[A_\delta\left(x_\delta^{(\ell)}\right) - A_h\left(Px_\delta^{(\ell)}\right) - \varepsilon_*\right]\left[A_\delta\left(x_\delta^{(\ell)}\right) - A_h\left(Px_\delta^{(\ell)}\right) - \varepsilon_*\right]^{\mathrm{T}} + \Gamma_e = \Gamma_n.$$

The Gaussian approximation for the noise covariance is thus

$$\begin{aligned}\pi(n) &\approx \pi_{\mathrm{G}}(n) \\ &\propto \exp\left(-\tfrac{1}{2}(n - \varepsilon_* - e_*)^{\mathrm{T}}\Gamma_n^{-1}(n - \varepsilon_* - e_*)\right).\end{aligned}$$

To simplify the analysis further, we write an approximation where the mutual dependence of x_h and the approximation error are ignored. By (6), this leads to an approximation that is referred to as *enhanced error model* [4, 2],

$$\begin{aligned}\pi(x_h \mid y) \propto \exp\big(&-\tfrac{1}{2}(x_h - x_{h*})^{\mathrm{T}}\Gamma_{x_h}^{-1}(x_h - x_{h*}) \\ &-\tfrac{1}{2}(y - A(x_h) - \varepsilon_* - e_*)^{\mathrm{T}}\Gamma_n^{-1}(y - A(x_h) - \varepsilon_* - e_*)\big),\end{aligned} \qquad (11)$$

where

$$x_{h*} = P\mathrm{E}\{x_\delta\} \in \mathbb{R}^n, \qquad \Gamma_{x_h} = P\Gamma_{x_\delta}P^\mathrm{T} \in \mathbb{R}^{n \times n}.$$

Let us write the factorizations (e.g., Cholesky factorizations),

$$L_{x_h}^\mathrm{T} L_{x_h} = \Gamma_{x_h}^{-1}, \qquad L_{e+\varepsilon}^\mathrm{T} L_{e+\varepsilon} = \Gamma_n^{-1},$$

we observe that the computation of the maximum *a posteriori* (MAP) estimate, within the validity of the approximations above, amounts to minimizing the functional

$$F(x_h) = \|L_{e+\varepsilon}(y - A_h(x_h) - \varepsilon_* - e_*)\|^2 + \|L_{x_h}(x_h - x_{h*})\|^2. \tag{12}$$

This computational problem is superficially of the same form as the computation of a Tikhonov regularized solution. The most common choice for the Tikhonov regularization would be to set $\varepsilon \equiv 0$, $\Gamma_{x_h} = \alpha^{-1}I$, $x_* = 0$ and $e \sim \mathcal{N}(0, \sigma^2 I)$ so that the problem (12) would reduce to the minimization of the functional

$$x_h \mapsto \frac{1}{\sigma^2}\|y - A_h(x_h)\|^2 + \alpha\|x_h\|^2.$$

Thus the enhanced error model (12) is appealing because we incorporate the model inaccuracy into the same framework as our general machinery for MAP estimation.

Previous studies [2, 4] show that in some cases, neglecting the stochastic dependence of the estimated signal x_h and the modelling error has a negligible effect and in others a significant effect. The effect was negligible in the electrical impedance tomography (EIT) studies in [2] which suggests that the situation might be the same with ODT.

Finally, we note that the approximation error theory also fits into the framework of iterative regularization methods, wherein the prior term $\|L_{x_h}(x_h - x_{h*})\|^2$ in (12) is not explicitly stated but instead an iterative scheme is used for the approximate minimization of the likelihood term $\|L_{e+\varepsilon}(y - A_h(x_h) - \varepsilon_* - e_*)\|^2$ and truncation of the iterations prior to convergence takes the role of regularization.

3. Optical diffusion tomography

Optical diffusion tomography (ODT) is a non-invasive imaging method in which images of the optical absorption and scattering functions within a turbid medium are derived based on measurements of near-infrared light on the surface of the object [1]. The main emphasis in the application of ODT has been in medicine, where applications include the detection and classification of tumours from breast tissue, monitoring of infant brain tissue oxygenation level and functional brain activation studies. For a recent review on the clinical applications, see [9]. In addition to medical applications, there has been recently a growing interest towards the use of optical methods in industrial process tomography [10].

In the experimental set-up, m_s optic fibres are placed on the source positions $\varepsilon_k \subset \partial\Omega$ on the boundary of the body Ω, and m_d optic fibres are placed in the detector positions $\zeta_i \subset \partial\Omega$.

In this paper we consider the so-called frequency domain measurements. Light from a sinusoidally modulated laser source is guided to the body via one of the source locations at ε_k and the amplitudes and phase shifts of the transmitted light are measured on all the detector locations $\zeta_i, i = 1, \ldots, m_d$ using the detector fibres and light sensitive detectors. This process is then repeated for all the m_s source locations. The objective is to reconstruct the absorption and scattering functions within the body Ω based on these transmission data. This reconstruction problem is a nonlinear and highly ill-posed inverse boundary value problem. In the following, we discuss briefly the models and methods that are used for the ODT problem in this paper.

3.1. Diffusion approximation model

The most commonly used forward model for optical diffusion tomography is the diffusion approximation (DA) to the radiative transfer equation (RTE). The diffusion approximation model is also used in this paper. In this paper we assume that the diffusion approximation is an accurate representation of the ODT measurements. The issue how well the diffusion model approximates the radiative transfer equation or the actual ODT measurements is not considered in this paper. For further details on derivation and properties of the forward models and boundary conditions, see [1, 2, 11, 12].

Let $\Omega \subset \mathbb{R}^p (p = 2, 3)$ denote the model domain. The frequency domain version of the diffusion approximation with the Robin-type boundary condition is of the form [12]:

$$-\nabla \cdot \kappa(r)\nabla \Phi_k(r, \omega) + \mu_a(r)\Phi_k(r, \omega) + \frac{i\omega}{c}\Phi_k(r, \omega) = q_{0,k}(r, \omega), \qquad r \in \Omega \qquad (13)$$

$$\Phi_k(r, \omega) + \frac{1}{2\gamma}\kappa(r)\frac{\partial \Phi_k(r, \omega)}{\partial \nu} = g_k(r, \omega) \qquad r \in \partial\Omega, \qquad (14)$$

where ω is the modulation frequency of the light source, $\Phi_k(r, \omega)$ is the photon density (for the source at $\varepsilon_k \subset \partial\Omega$), $\mu_a(r)$ is the absorption coefficient (mm^{-1}), $\kappa(r) = (p(\mu_a(r) + \mu_s(r)))^{-1}$ (mm) is the diffusion coefficient, $p = 2, 3$ is the dimension of the domain, $\mu_s(r)$ is the reduced scattering coefficient (mm^{-1}), c is the speed of light in the medium, $q_{0,k}(r, \omega)$ is the distribution of sources inside Ω, $g_k(r, \omega)$ models the boundary sources at the source sites ε_k, γ is a dimension-dependent constant ($\gamma_2 = 1/\pi$, $\gamma_3 = 1/4$) and ν is the outer normal at $\partial\Omega$ [12, 2]. As the model for the sources we use the diffuse boundary source model in which the source is modelled as an inward directed diffuse photon current f_s at the source sites:

$$g_k(r, \omega) = \begin{cases} \dfrac{f_s}{2\gamma} & r \in \varepsilon_k \\ 0 & r \in \partial\Omega \backslash \varepsilon_k. \end{cases} \qquad (15)$$

and the internal source term is $q_{0,k}(r, \omega) = 0$. Assume that r_i is a point at the detector site ζ_i. The measured flux $z_{i,k}(r_i, \omega)$ is obtained as:

$$z_{i,k}(r_i, \omega) = 2\gamma\Phi_k(r_i, \omega). \qquad (16)$$

3.2. FEM implementation of the diffusion model

The numerical solution of the diffusion approximation model (13)–(15) is based on the finite element method (FEM). The derivation of the finite element discretization has been presented in previous publications, see e.g. [1, 11–14]. Here we repeat only the resulting FEM equations.

In the FEM approximation, the domain Ω is divided into N_e elements Ω_ℓ which are joined by N_n node points. The photon density is approximated in a finite-dimensional basis

$$\Phi^h = \sum_{i=1}^{N_n} \phi_i \varphi_i(r), \qquad (17)$$

where φ_i are the nodal basis functions of the finite element mesh. In this paper, we use piecewise linear polynomials φ_i which have support over the elements that are joined by node i. By inserting approximation (17) into the variational formulation of (13)–(15) and using the basis functions φ_j as the test functions, we arrive at the matrix equation

$$(K(\kappa) + C(\mu_a) + R + i\omega Z)\Phi^h = G, \qquad (18)$$

where the elements of the system matrices are given by

$$K_{i,j} = \int_{\Omega} \kappa \nabla \varphi_i \cdot \nabla \varphi_j \, dr \tag{19}$$

$$C_{i,j} = \int_{\Omega} \mu_a \varphi_i \varphi_j \, dr \tag{20}$$

$$Z_{i,j} = \frac{1}{c} \int_{\Omega} \varphi_i \varphi_j \, dr \tag{21}$$

$$R_{i,j} = \int_{\partial\Omega} 2\gamma \varphi_i \varphi_j \, dS \tag{22}$$

and the source vector is of the form

$$G_j = \int_{\partial\Omega} 2 f_s \varphi_j \, dS. \tag{23}$$

The FEM approximation Φ_k^h for the kth source is obtained formally as

$$\Phi_k^h = (K(\kappa) + C(\mu_a) + R + i\omega Z)^{-1} G.$$

The complex measurement vector

$$z^{(k)} = \left(z_{1,k}, \ldots, z_{m_d,k} \right)^{\mathrm{T}} \in \mathbb{C}^{m_d}$$

containing the measured flux at the measurement sites $\{\zeta_i\}$ for the source at ε_k is obtained by

$$z^{(k)} = M \Phi_k^h,$$

where the elements of the measurement matrix M are defined by

$$M_{i,\ell} = \begin{cases} 2\gamma, & \text{if node } r_\ell \in \zeta_i \subset \partial\Omega \\ 0, & \text{otherwise} \end{cases}, \qquad i = 1, \ldots, m_d \quad \ell = 1, \ldots, N_n. \tag{24}$$

To complete the specification on the forward model of ODT, we give the notation that will be used for the data and the FEM-based forward solver in the subsequent sections. Let

$$z = (z^{(1)}, \ldots, z^{(m_s)})^{\mathrm{T}} \in \mathbb{C}^{m_d m_s}$$

denote the concatenated vector of complex data for all the m_s sources. The practical devices for frequency-domain optical tomography measure the amplitude and phase shift of the complex-valued flux. The forward solution can be represented in the corresponding form by

$$y = \begin{pmatrix} \mathrm{re}(\log(z)) \\ \mathrm{im}(\log(z)) \end{pmatrix} \in \mathbb{R}^m, \qquad m = 2m_s m_d. \tag{25}$$

At this stage, we specify the discretization for the unknowns (μ_a, μ_s) for the inverse problem. We use piecewise constant approximations

$$\mu_a(r) = \sum_{j=1}^{n_p} \mu_{a,j} \chi_j(r) \tag{26}$$

$$\mu_s(r) = \sum_{j=1}^{n_p} \mu_{s,j} \chi_j(r), \tag{27}$$

where χ_j denote the characteristic function of disjoint image pixels. Within the expression (26)–(27), we identify the coefficients (distributed parameters) $(\mu_a(r), \mu_s(r))$ with the vectors

$$\mu_{\mathrm{a}} = \left(\mu_{\mathrm{a},1}, \ldots, \mu_{\mathrm{a},n_p}\right)^{\mathrm{T}} \in \mathbb{R}^{n_p} \tag{28}$$

$$\mu_{\mathrm{s}} = \left(\mu_{\mathrm{s},1}, \ldots, \mu_{\mathrm{s},n_p}\right)^{\mathrm{T}} \in \mathbb{R}^{n_p} \tag{29}$$

and the parameter vector for the inverse problem becomes

$$x_h = \begin{pmatrix} \mu_{\mathrm{a}} \\ \mu_{\mathrm{s}} \end{pmatrix} \in \mathbb{R}^n, \qquad n = 2n_p. \tag{30}$$

For the FEM-based forward solver we use the notation

$$y = A_h(x_h), \tag{31}$$

where the subindex h refers to the finite element mesh that is used.

3.3. MAP estimate for ODT

In this section, we briefly describe the Bayesian MAP estimation problems for optical diffusion tomography with the conventional and enhanced error models. The Bayesian approach to optical diffusion tomography with the conventional likelihood (measurement) model has previously been considered in [12, 14–19].

We assume here that the prior distribution of the unknown target x_h is modelled by a proper Gaussian density:

$$x_h \sim \mathcal{N}\left(x_{h*}, \Gamma_{x_h}\right), \qquad \Gamma_{x_h}^{-1} = L_{x_h}^{\mathrm{T}} L_{x_h}. \tag{32}$$

Consider first the conventional likelihood model. A typical assumption in ODT is that the measurements are corrupted by additive zero-mean Gaussian noise e which is independent of the unknown x_h. This assumption is also made here, and thus, we write the observation model as

$$y = A_h(x_h) + e, \qquad e \sim \mathcal{N}(0, \Gamma_e), \qquad \Gamma_e^{-1} = L_e^{\mathrm{T}} L_e. \tag{33}$$

The computation of the MAP estimate for models (32)–(33) amounts to the minimization problem:

$$\|L_e(y - A_h(x_h))\|^2 + \left\|L_{x_h}(x_h - x_{h*})\right\|^2 \longrightarrow \min. \tag{34}$$

In the subsequent sections, we refer to the solution of (34) as the MAP estimate with the conventional likelihood (measurement) model.

Consider next the application of the enhanced error model. If we write an observation model of the form (10) with the assumption $e \sim \mathcal{N}(0, \Gamma_e)$ and proceed as in section 2.3, the computation of the MAP estimate amounts to minimizing

$$\|L_{e+\varepsilon}(y - A_h(x_h) - \varepsilon_*)\|^2 + \left\|L_{x_h}(x_h - x_{h*})\right\|^2 \longrightarrow \min, \tag{35}$$

where $L_{e+\varepsilon}^{\mathrm{T}} L_{e+\varepsilon} = (\Gamma_e + \Gamma_\varepsilon)^{-1} = \Gamma_n^{-1}$. In the following, we refer to the solution of (35) as the MAP estimate with the enhanced error model.

The minimization problems (34)–(35) can be solved, for example, by the Gauss–Newton algorithm. Note that when the number n of unknowns is large, the Gauss–Newton updates can be efficiently computed with the implicit Gauss–Newton method, see for example [20].

3.4. A proper smoothness prior model

The conventional smoothness prior for the image vector $x_h = (\mu_{\mathrm{a}}, \mu_{\mathrm{s}})^{\mathrm{T}} \in \mathbb{R}^n$ is usually defined as

$$\pi(x_h) \propto \exp\left\{-\tfrac{1}{2}\|Lx_h\|^2\right\} = \exp\left\{-\tfrac{1}{2}x_h^{\mathrm{T}} B x_h\right\}, \tag{36}$$

where $B = L^{\mathrm{T}} L$, L is a block matrix

$$L = \begin{pmatrix} \alpha_{\mu_{\mathrm{a}}} D & 0 \\ 0 & \alpha_{\mu_{\mathrm{s}}} D \end{pmatrix} \qquad (37)$$

and D is a first-order difference matrix between adjacent pixels in the image basis. The prior density (36) is improper (i.e., B is not invertible) because matrix D has null space $\mathcal{N}(D) = \mathrm{span}\{c\}$, where $c = -\sqrt{2/n}(1, 1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^{n/2}$. Therefore, the conventional smoothness prior cannot be directly applied in the approximation error theory. However, we can use (36) as a starting point in the construction of a proper smoothness prior model.

To get a proper density, we follow the procedure in [2, 21]. Assume that we have prior information on the actual values and uncertainty of the absorption and scattering functions at some of the image pixels. Let k be the number of these pixels. By possibly reordering the elements of x_h, we may write the partition $x_h = (x_{h,1}, x_{h,2})^{\mathrm{T}} \in \mathbb{R}^n$, where $x_{h,2} \in \mathbb{R}^{2k}$ contain the absorption and scattering parameters of the 'specified' pixels where we can specify the prior model for the actual absorption and scattering values, and $x_{h,1} \in \mathbb{R}^{n-2k}$ contain the absorption and scattering parameters of the remaining, 'unspecified' pixels. By partitioning accordingly the matrix

$$L^{\mathrm{T}} L = B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

in equation (36), we can write a proper conditional smoothness prior for $x_{h,1}$ conditioned on $x_{h,2}$ as [2]:

$$\pi(x_{h,1}|x_{h,2}) \propto \exp\left\{-\tfrac{1}{2}\left(x_{h,1} + B_{11}^{-1} B_{12} x_{h,2}\right)^{\mathrm{T}} B_{11}^{-1}\left(x_{h,1} + B_{11}^{-1} B_{12} x_{h,2}\right)\right\}. \quad (38)$$

Next, assume that the absorption and scattering in the specified pixels is modelled by a Gaussian prior

$$\pi(x_{h,2}) \propto \exp\left\{-\tfrac{1}{2}(x_{h,2} - x_{h*,2})^{\mathrm{T}} \Gamma_0^{-1}(x_{h,2} - x_{h*,2})\right\}. \qquad (39)$$

Using equations (38)–(39), we can now obtain a new, proper smoothness prior for x_h as [2]:

$$\pi(x_{h,1}, x_{h,2}) = \pi(x_{h,1}|x_{h,2})\pi(x_{h,2}) \propto \exp\left\{-\tfrac{1}{2}(x_h - x_{h*})^{\mathrm{T}} \Gamma_{x_h}^{-1}(x_h - x_{h*})\right\}, \qquad (40)$$

where

$$x_{h*} = \begin{pmatrix} -B_{11}^{-1} B_{12} x_{h*,2} \\ x_{h*,2} \end{pmatrix}, \qquad \Gamma_{x_h}^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{21} B_{11}^{-1} B_{12} + \Gamma_0^{-1} \end{pmatrix}. \qquad (41)$$

Furthermore, let us denote the Cholesky decompositions of B_{11} and Γ_0^{-1} by $B_{11} = L_1^{\mathrm{T}} L_1$ and $\Gamma_0^{-1} = L_0^{\mathrm{T}} L_0$. Then we can write the prior model (40)–(41) as

$$\pi(x_h) \propto \exp\left\{-\tfrac{1}{2}\left\|L_{x_h}(x_h - Q x_{h*,2})\right\|^2\right\}, \qquad (42)$$
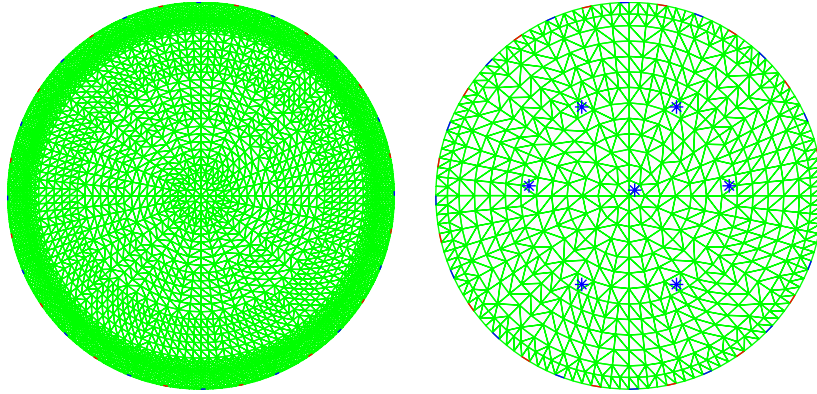
where

$$L_{x_h} = \begin{pmatrix} L_1 & L_1 B_{11}^{-1} B_{12} \\ 0 & L_0 \end{pmatrix}, \qquad Q = \begin{pmatrix} -B_{11}^{-1} B_{12} \\ I \end{pmatrix}. \qquad (43)$$

4. Numerical results

4.1. The ODT set-up

We evaluate the enhanced error model in the case $\Omega \subset \mathbb{R}^2$. The object domain is a circle with radius of 25 mm. The measurement set-up in the simulations consists of $m_s = 16$ sources and

Figure 1. Two finite element meshes used in the simulations. Left: h^1 ($N_n = 4217$, $N_e = 7920$) is the mesh for the accurate forward model, i.e. $A_{h^1}(x_h) = A_\delta(x_\delta)$. Right: h^5 ($N_n = 713$, $N_e = 1296$) is the mesh for a target model. The asterisks denote the locations of the 'specified' (conditioned) pixels in the construction of the proper smoothness prior model (see section 4.2). The elements of $x_{h,2} \in \mathbb{R}^{14}$ correspond to the absorption and scattering coefficients of the seven specified pixels.

Table 1. The finite element meshes for the simulations. Mesh h^0 is used for the simulation of the measurement data. h^1 is the mesh for the accurate forward model, i.e. $A_{h^1}(x_h) = A_\delta(x_\delta)$. h^k, $k = 2, \dots, 5$ are meshes for the target models with different accuracy. N_n is the number of nodes and N_e is the number of triangle elements in the mesh.

| Mesh	N_n	N_e
h^0	5625	10 736
h^1	4217	7 920
h^2	2057	3 856
h^3	1513	2 768
h^4	1001	1 744
h^5	713	1 296

$m_d = 16$ detectors, located at equispaced intervals on the boundary $\partial\Omega$. With this set-up, the vector of frequency domain ODT data (25) in the simulations is $y \in \mathbb{R}^{512}$.

For the finite-element discretization, we create a set of different meshes with triangular elements. The details of the meshes are given in table 1. The mesh h^0 is used for the simulation of the measurement data, h^1 is the mesh for the accurate forward model (i.e., $A_{h^1}(x_h) = A_\delta(x_\delta)$) and h^k, $k = 2, \dots, 5$ are meshes for the target models. Images of the meshes h^1 and h^5 are shown in figure 1.

For the representation of the functions (μ_a, μ_s), we divide the domain Ω into $n_p = 524$ quadrilateral pixels, and the functions (μ_a, μ_s) are approximated in piecewise constant bases of the form (26)–(27), leading to $x_h \in \mathbb{R}^{1048}$ unknowns in the inverse problem. The MAP estimation problems (34) and (35) are solved by a Gauss–Newton algorithm with an explicit line search algorithm. For further details see [14, 20].

4.2. The prior model and the Gaussian approximation for the approximation error

We construct a proper first-order smoothness prior distribution as explained in section 3.4. We aim to control the approximative correlation length of the coefficient distributions as follows.
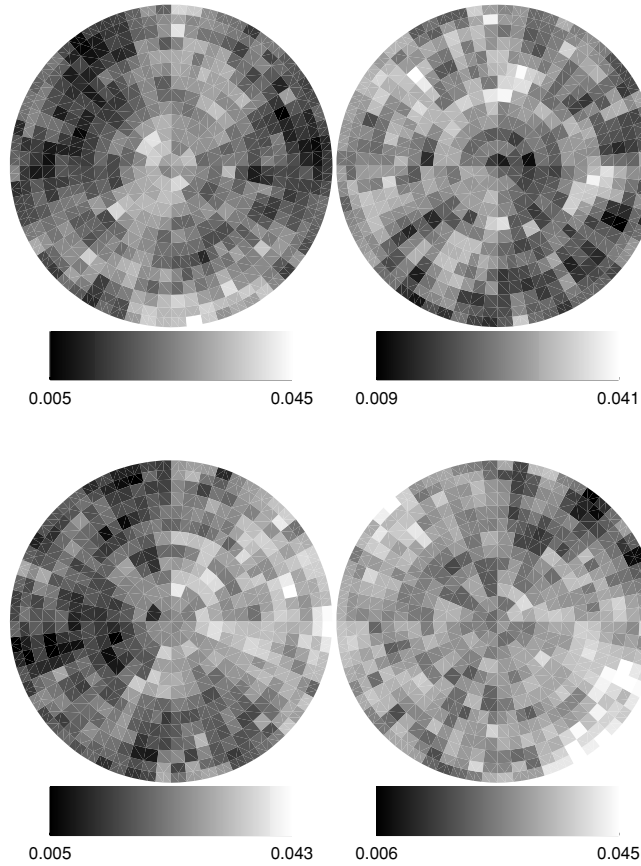
Figure 2. Four draws from the proper smoothness prior model (42). The correlation length is approximately one-quarter of the diameter of the disk.

We choose seven pixels as 'specified' pixels such that their distances from each other is approximately one-quarter of the diameter of the domain Ω. The locations of the specified pixels are shown with asterisks in figure 1. The absorption and scattering parameters of the specified pixels define elements $x_{h,2} \in \mathbb{R}^{14}$ of the unknown x_h. For the marginal distribution of the parameters $x_{h,2}$, we specify a Gaussian prior model of the form (39) with diagonal covariance matrix Γ_0 (i.e., elements of $x_{h,2}$ are assumed mutually independent). For the prior mean of $x_{h,2}$, we assume values $\mu_{a*} = 0.025$ mm^{-1} and $\mu_{s*} = 2$ mm^{-1}. The standard deviations $(\sigma_{\mu_a}, \sigma_{\mu_s})$ are chosen such that $[\mu_{a*} - 3\sigma_{\mu_a}, \mu_{a*} + 3\sigma_{\mu_a}] = [0.005, 0.045]$ mm^{-1} and $[\mu_{s*} - 3\sigma_{\mu_s}, \mu_{s*} + 3\sigma_{\mu_s}] = [0.5, 3.5]$ mm^{-1}, respectively.

These choices correspond to the assumption that absorption and scatter values at the specified pixels are with prior probability 0.99 in the given intervals. These intervals of parameter values are approximations for the range of absorption and scattering values found typically in medical applications of ODT. Once the marginal distribution for $x_{h,2}$ has been specified, the parameters α_{μ_a} and α_{μ_s} for the difference operator, see equation (37), are tuned such that the pixelwise variances in the proper smoothness prior model $\pi(x_h)$, equation (42), for the absorption and scattering images become approximately equal. That is, we choose α_{μ_a} and α_{μ_s} such that we have $\text{var}_{\pi(x_h)}(\mu_{a,k}) \approx c_1$ and $\text{var}_{\pi(x_h)}(\mu_{s,k}) \approx c_2$ for all k. For general

details on the equalization of the variances over subdomains, see [21]. Figure 2 shows four random draws from the resulting prior model $\pi(x_h)$.

In this study, we ignore the cross correlation of x_h and ε, that is, we employ the enhanced error model (12). The results given below indicate that this would be sufficient choice for ODT. To compute the enhanced error model between a target model (meshes h^k, $k = 2, \ldots 5$) and the accurate forward model (mesh h^1) we proceed as explained in section 2.3. We draw a set of i.i.d. samples $\{x_h^{(\ell)}, \ell = 1, \ldots, r\}$ from the prior distribution $\pi(x_h)$ and compute the corresponding approximation errors

$$\varepsilon_{h^k}^{(\ell)} = A_{h^1}\big(x_h^{(\ell)}\big) - A_{h^k}\big(x_h^{(\ell)}\big), \qquad k = 2, \ldots, 5. \tag{44}$$

Note that we use the same 524-pixel representation for the absorption and scattering images in the accurate forward model and the target models. Therefore, the vectors $\varepsilon_{h^k}^{(\ell)}$ do not contain error due to the projection of the unknown x_h from a higher-dimensional space to a lower-dimensional one (i.e., we have $P = I$). The distribution of the approximation error $\varepsilon_{h^k} \sim \mathcal{N}\big(\varepsilon_{h^k*}, \Gamma_{\varepsilon_{h^k}}\big)$ is approximated from the samples as

$$\varepsilon_{h^k*} = \frac{1}{r}\sum_{\ell=1}^{r}\varepsilon_{h^k}^{(\ell)}, \qquad \Gamma_{\varepsilon_{h^k}} = \frac{1}{r-1}\sum_{\ell=1}^{r}\varepsilon_{h^k}^{(\ell)}\varepsilon_{h^k}^{(\ell)\mathrm{T}} - \varepsilon_{h^k*}\varepsilon_{h^k*}^{\mathrm{T}}. \tag{45}$$

In this study, we use $r = 2500$ samples for the construction of the enhanced error model.

Figure 3 shows an example of the approximation errors. The thick lines in the top figures show the absolute value of the mean ε_{h^5*} which was computed using the accurate forward model A_{h^1} and the target model A_{h^5} (amplitude on left, phase shift on right). The thin line shows the standard deviations of the random noise e with the relative noise level $\delta_e = 0.5\%$, which is used for the evaluation of the enhanced error model for different meshes in the next section. As can be seen, the means of the channelwise approximation errors between the models A_{h^1} and A_{h^5} exceed clearly the typical levels of random noise for ODT. The bottom figures show the estimated standard deviations of ε_{h^5} with thick lines and the standard deviations of e with thin lines. As can be seen, with the relative noise level $\delta_e = 0.5\%$ the standard deviations in amplitude are of the same magnitude but in the phase data the standard deviations of ε_{h^5} are larger than the standard deviations of e. Based on these figures, we would expect to get poor reconstructions if we were to use target model A_{h^5} with the conventional likelihood (measurement) model.

4.3. Estimation results over the prior distribution

The effects of a fixed approximation error to reconstruction with different noise levels were studied in [2]. The investigated problems were full angle computerized tomography (CT) and electrical impedance tomography (EIT). In both problems the result was that once the additive noise level gets as small as the approximation error level, the estimation errors $\mathrm{E}\{\|x - \hat{x}\|^2\}$ tend to grow with decreasing noise levels. On the other hand, if the approximation error was modelled, the estimation errors decreased first and then, with very small noise levels, stagnated or increased very slowly. The latter behaviour can be conjectured to be due to (a) using the enhanced error model instead of the complete approximation error model, (b) approximate prior models and (c) using a Gaussian approximation error model while the actual approximation errors were non-Gaussian.

4.3.1. Fixed noise level, decreasing accuracy of the target model.
We compute the estimation errors

$$\varrho(\hat{x}_h) = \mathrm{E}\{\|x_h - \hat{x}_h\|^2\}\big(\mathrm{tr}\,\Gamma_{x_h} + \|x_{h*}\|^2\big)^{-1}$$
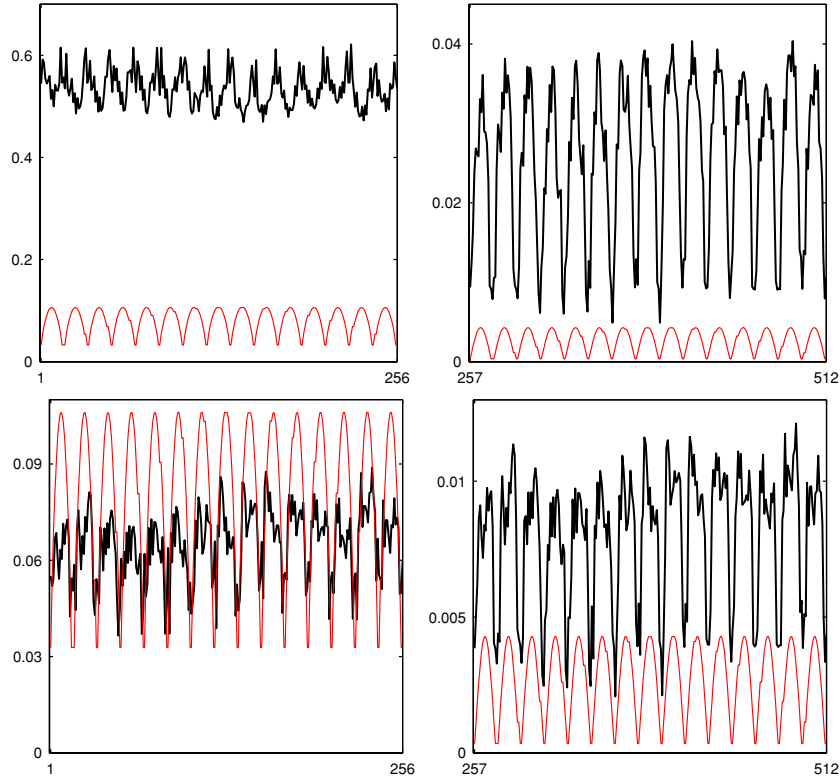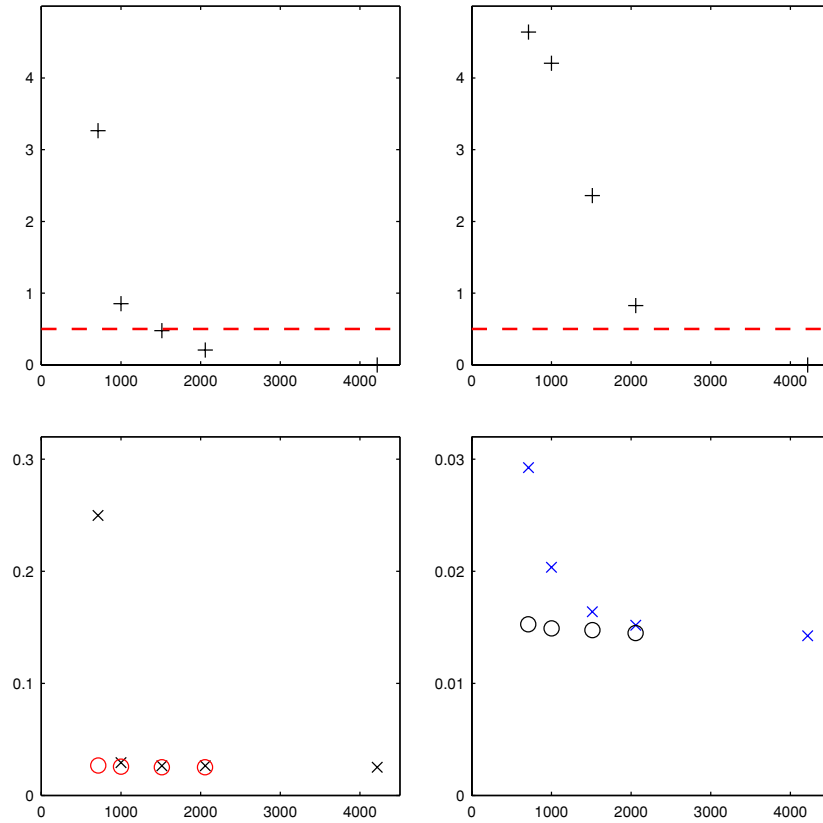
Figure 3. Example of the approximation error. Top: the absolute value of the mean of the approximation errors $|\varepsilon_{h^5*}|$ between the accurate model A_{h^1} ($N_n = 4217$, $N_e = 7920$) and the target model A_{h^5} ($N_n = 713$, $N_e = 1296$) is shown with thick line. The thin line shows the standard deviations $\sigma_{e,j}$ of the random noise e corresponding to the relative noise level $\delta_e = 0.5\%$ (i.e., $\sigma_{e,j} = \delta_e|y_{*,j}|/100$). Left image shows the amplitude part and right image the phase shift part of the approximation error vector. The abscissae denote the measurement index j. Bottom: thick line shows the standard deviations $\sigma_{\varepsilon_{h^5}}$ of the approximation errors ε_{h^5}. Thin line shows the standard deviations of e with $\delta_e = 0.5\%$.

separately for the absorption and scattering coefficients. There is a common conjecture that the absorption coefficient is more sensitive to the amplitude data whereas the scattering coefficient is more sensitive to the relative phase data.

The employed target models (meshes h^k, $k = 2, \ldots, 5$) have numbers of nodes between 2057 and 713 while the accurate forward model (h^1) has 4027 nodes. To compute the simulated data, we draw another set of 100 i.i.d. samples from the prior distribution and compute the respective forward solutions in the mesh h^0 with 5625 nodes and add random noise to the simulated data as $y = A_{h^0}(x) + e$, where realizations of e are drawn from the distribution $\mathcal{N}(0, \Gamma_e)$ with $\Gamma_e = \text{diag}(\sigma_{e,1}^2, \ldots, \sigma_{e,m}^2)$ and $\sigma_{e,j} = \delta_e|y_{*,j}|/100$ (i.e., we assume that the channelwise standard deviation of the noise is δ_e per cent of the absolute value of the computed data). Here we use relative noise level $\delta_e = 0.5\%$.

In figure 4, we show the approximation errors for both amplitude and relative phase data together with the estimation errors $\varrho(\mu_{a,\text{MAP}})$ and $\varrho(\mu_{s,\text{MAP}})$ for the absorption and scattering coefficients. The estimation errors are computed as the ensemble mean of the reconstructions from the 100 realizations of the data. The MAP estimates with the conventional likelihood

Figure 4. Top row: expected relative approximation errors $E\{\|A_{h^1}(x_h) - A_{h^k}(x_h)\|/\|A_{h^1}(x_h)\|\}$ with respect to the number of nodes in the FE-mesh. Left: amplitude data. Right: phase data. The dashed line indicates the relative level ($\delta_e = 0.5\%$) of the additive random noise that is used in the simulation. Bottom row: the expected errors $E\{\|\mu_{a,MAP} - \mu_a\|^2\}(\text{trace}(\Gamma_{\mu_a}) + \|\mu_{a*}\|^2)^{-1}$ (left) and $E\{\|\mu_{s,MAP} - \mu_s\|^2\}(\text{trace}(\Gamma_{\mu_a}) + \|\mu_{s*}\|^2)^{-1}$ (right) in the reconstructed images with respect to the number of nodes N_n in the FE-mesh. The errors with the enhanced error model are shown with circles (o) and the errors with the conventional likelihood model are shown with crosses (x). The expectation is computed as sample average over 100 realizations of data that were simulated in the mesh h^0.

model are computed by solving (34) and the respective estimates with the enhanced error model by solving (35). The estimates are computed with the Gauss–Newton algorithm with an explicit line search. For details, see [14, 20]. The results show the following. If the conventional likelihood model is used, the estimate errors start to increase when the approximation errors exceed the noise level. Note the similarity in the trends of amplitude data and μ_a, and phase data and μ_s. This similarity supports the common conception that the amplitude data are more sensitive to absorption and phase data more sensitive to scattering. On the other hand, if the enhanced error model is employed, the estimate errors increase only marginally even when the target model is reduced drastically.

It is to be noted that the estimation errors for both coefficients are not extremely large with a ~ 1000-node mesh even when the conventional likelihood model is used. This is in contrast with the common conception of how dense the meshes with ODT should be. The reason here is probably due to that the correct, albeit discretized, prior model was used. On
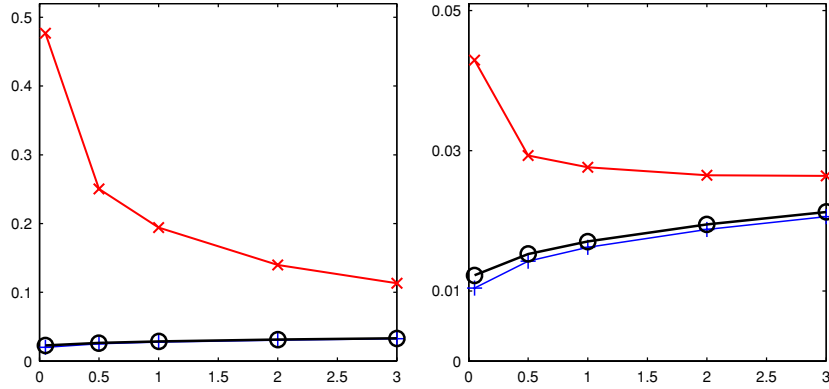
Figure 5. The expected errors $E\{\|\mu_{a,MAP} - \mu_a\|^2\}(\text{trace}(\Gamma_{\mu_a}) + \|\mu_{a*}\|^2)^{-1}$ (left) and $E\{\|\mu_{s,MAP} - \mu_s\|^2\}(\text{trace}(\Gamma_{\mu_a}) + \|\mu_{s*}\|^2)^{-1}$ (right) in the reconstructed images with respect the noise level δ_e (%). The error with conventional 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^5 ($N_n = 713$) is shown with (x), and the error with the enhanced error model in the sparse mesh h^5 ($N_n = 713$) is shown with (o). The expectations are computed as sample averages over 100 realizations of the data.

the other hand, if this is the sole reason, it can be argued that the previously required mesh densities may be mostly due to poor prior models or infeasible regularization methods.

4.3.2. Fixed mesh, decreasing noise level. We employ the target model A_{h^5} with the 713 node mesh here. With this mesh the approximation errors are larger than the noise levels. The estimation errors with different relative noise levels δ_e between 3% and 0.05% are shown in figure 5. We also show estimation errors when the accurate forward model $A_{h^1}(x_h)$ ($N_n = 4217$) with conventional likelihood model is used. The estimation errors $\varrho(\mu_{a,MAP})$ and $\varrho(\mu_{s,MAP})$ are again estimated based on 100 reconstructed data sets. The behaviour of the estimation errors has the same characteristics as with previous studies with different problems. When the target model $A_{h^5}(x_h)$ and conventional likelihood model is used, the estimation errors increase although the noise level decreases. On the other hand, the estimate errors decrease monotonically with both the accurate forward model $A_{h^1}(x_h)$ using the conventional likelihood model and the target model $A_{h^5}(x_h)$ with the enhanced error model. The fact that the estimate errors when using the enhanced and target models are not much worse than when using an accurate forward model, supports the simpler choice of enhanced error model over the more accurate approximation error model which takes into account the cross correlation between x_h and ε.

4.3.3. Reconstruction of a test phantom. As explained above and shown in figure 2, the prior probability distribution is concentrated on coefficients with smooth spatial distribution, and moreover, the absorption 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 very 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.
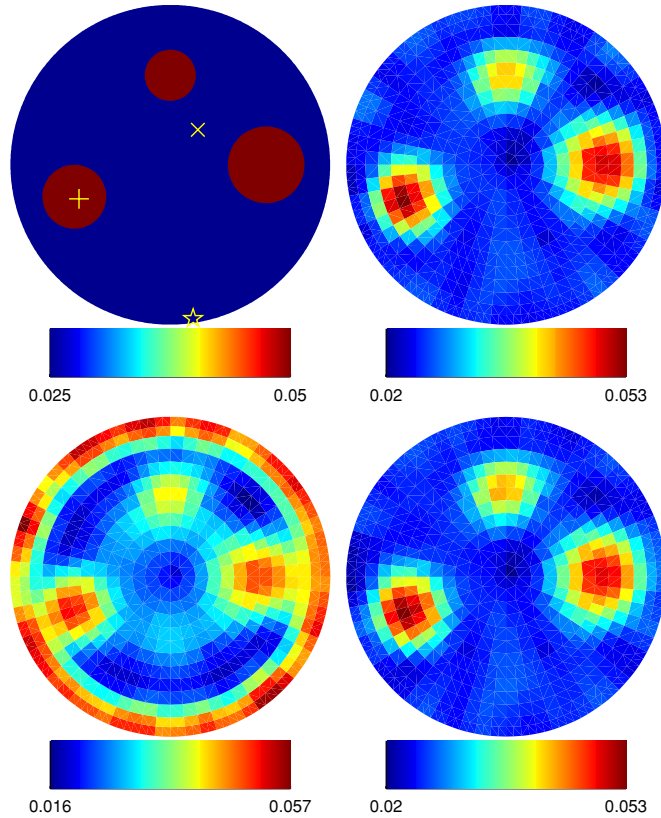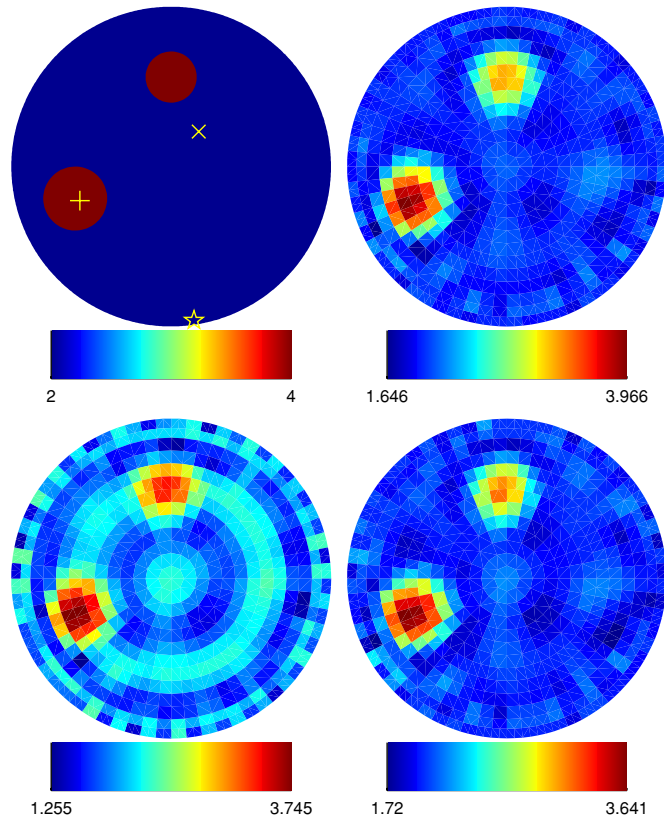
Figure 6. Reconstructed absorption images μ_a. Top left: the target. Top right: reconstruction with conventional likelihood model using the accurate forward model $A_{h^1}(x_h)$ ($N_n = 4217$). Bottom left: reconstruction with the conventional likelihood model using the target model $A_{h^5}(x_h)$ ($N_n = 713$). Bottom right: reconstruction with the enhanced error model using the target model $A_{h^5}(x_h)$ ($N_n = 713$). The relative noise level was $\delta_e = 0.5\%$.

To avoid too optimistic results and to assess the robustness of the modelling error theory developed here, 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 $A_{h^5}(x_h)$ with enhanced error model and the conventional likelihood model. The actual spatial coefficient distributions and the estimates are shown in figures 6 and 7. The relative noise level δ_e was 0.5%.

The results with the accurate model and the target model with enhanced error model are essentially equivalent while the target model with conventional likelihood model exhibits severe annular estimation errors. The errors are largest near the boundaries and have a similar overcompensation–undercompensation structure as with the EIT results in [2], where the results concerned ensemble characteristics over the correct prior model.

Note that there is little or no crosstalk between the estimates of the absorption and scattering coefficients. This suggests that the often encountered crosstalk may be due to the implicit regularization properties of the employed inversion method. With many regularization methods it can be difficult to approximate the implicit prior model behind the computational scheme.

Figure 7. Reconstructed scatter images μ_s. 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 $A_{h^5}(x_h)$ ($N_n = 713$). Bottom right: reconstruction with the enhanced error model using the target model $A_{h^5}(x_h)$ ($N_n = 713$). The relative noise level was $\delta_e = 0.5\%$.

4.3.4. Approximate estimate errors. One of the most appealing topics in the statistical inversion paradigm is that it is possible to provide reliable statistics for the estimation errors. The actual reliability has to be assessed with respect to the accuracy of the underlying likelihood and prior models: if they are correct or at least feasible, the error estimates are reliable; if they are not, very little can be said. On the other hand, since with inverse problems the prior variances are usually much larger than the likelihood variances, the accuracy of the prior model is not as stringent as with the likelihood. Thus, when significant approximation errors are present, the reliability of the estimate error can be very low.

In some applications the reliability of the estimates can be fundamental. Such a case is the screening of patients for tumours, to decide who should be subjected to more extensive diagnoses. Making a false positive screening decision, that is, claim no tumour when a tumour is present, is a grave one. More generally, statistical decision limits are always computed based on the distributions themselves rather than the point estimates [22].

In figure 8, we show the Gaussian approximations for marginal densities for both coefficients in the three pixels which are marked in figure 7. The computation of the actual marginal densities requires the implementation of a MCMC algorithm for the problem. In the
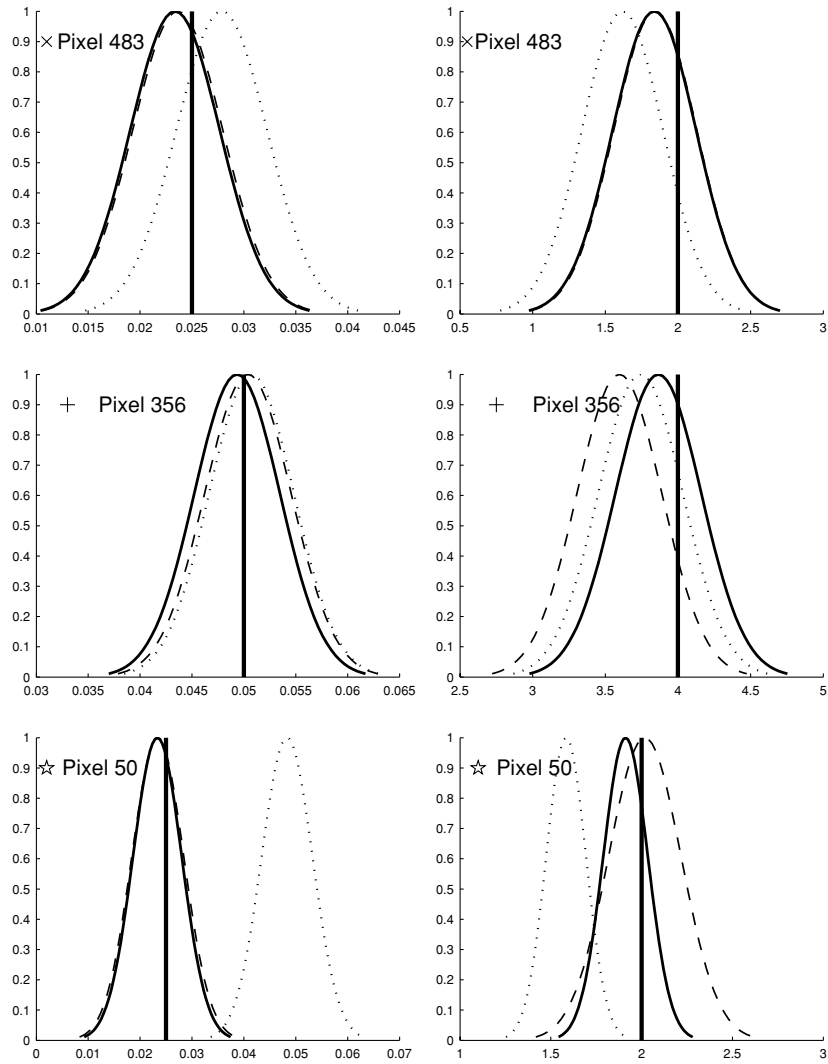
Figure 8. The (approximate) marginal densities of the absorption (left) and scattering coefficients (right) in the three pixels shown in figure 7. The vertical lines denote the actual values, the solid lines the marginal densities with the accurate forward model $A_{h^1}(x_h)$ ($N_n = 4217$) with conventional likelihood model. The dashed lines denote the marginal densities with target model $A_{h^5}(x_h)$ ($N_n = 713$) with enhanced error model and the dotted lines denote the marginals with the target model $A_{h^5}(x_h)$ with the conventional likelihood model. The densities are nonnormalized. The relative noise level was $\delta_e = 0.5\%$.

case of most pixels, the target model with enhanced error model gives reliable error estimates which are comparable or very near to those which were obtained when the accurate forward model with conventional likelihood model is employed. On the other hand, in the case of pixel number 50 when using the target model with conventional likelihood model, the actual coefficient values have essentially zero probability with respect to the approximate marginal density. In other words, we would claim that given the measurements, the correct value is impossible. Thus, at least in the present case, the approximation error model yields also more accurate (approximate) estimates for the accuracy of the (point) estimates.

5. 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 with 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 very 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. Furthermore, the reliability of the estimation errors was improved over the ones given when using the conventional error model.

In practice, optical diffusion tomography should be modelled with three-dimensional models. The extension of the enhanced error model to 3D is straightforward. We also expect that the relative reduction in the model size is more significant in 3D than in 2D. In this study, we assumed that the diffusion approximation to the radiative transfer equation (RTE) would be an accurate model for the actual ODT measurements, and we considered the compensation of reduction in the discretization accuracy by the approximation error theory. The future topics for the research include investigation of how well the approximation error method could be used to compensate the modelling errors between the diffusion approximation and arguably more accurate but computationally tedious RTE.

Acknowledgments

References

[1] Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* **15** R41–93
[2] Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems* (New York: Springer)
[3] Tarantola A 2004 *Inverse Problem Theory and Methods for Model Parameter Estimation* (Philadelphia, PA: SIAM)
[4] Kaipio J P and Somersalo E 2005 Statistical inverse problems: discretization, model reduction and inverse crimes *J. Comput. Appl. Math.* in press
[5] Mandelbaum A 1984 Linear estimators and measurable linear transformations on a Hilbert space *Z. Wahrscheinlichkeitstheor.* **3** 385–98
[6] Lehtinen M S, Päivärinta L and Somersalo E 1989 Linear inverse problems for generalized random variables *Inverse Problems* **5** 599–612
[7] van Huffel S and Vandeville J 1991 *The Total Least Squares Problem: Computational Aspects and Analyis* (SIAM)
[8] Nievergilt Y 1994 Total least squares: state of the art regression in numerical analysis *SIAM Rev.* **36** 258–64
[9] Gibson A P, Hebden J C and Arridge S R 2005 Recent advances in diffuse optical imaging *Phys. Med. Biol.* **50** R1–R43
[10] Williams R A and Beck M S (ed) 1995 *Process Tomography, Principles, Techniques and Applications* (Oxford: Butterworth-Heinemann)
[11] Schweiger M, Arridge S R, Hiraoka M and Delpy D T 1995 The finite element model for the propagation of light in scattering media: boundary and source conditions *Med. Phys.* **22** 1779–92
[12] Heino J and Somersalo E 2002 Estimation of optical absorption in anisotropic background *Inverse Problems* **18** 559–73

[13] Tarvainen T, Vauhkonen M, Kolehmainen V and Kaipio J P 2005 Hybrid radiative-transfer-diffusion model for optical tomography *Appl. Opt.* **44** 876–86

[14] Kolehmainen V 2001 Novel approaches to image reconstruction in diffusion tomography *PhD thesis* University of Kuopio, Kuopio, Finland

[15] Ye J C, Webb K J, Bouman C A and Millane R P 1999 Optical diffusion tomography by iterative coordinate-descent optimization in a Bayesian framework *J. Opt. Soc. Am.* A **16** 2400–12

[16] Ye J C, Bouman C A, Webb K J and Millane R P 2001 Nonlinear multigrid algorithms for Bayesian optical diffusion tomography *IEEE Trans. Image Process.* **10** 909–22

[17] Milstein A B, Oh S, Reynolds J S, Webb K J, Bouman C A and Millane R P 2002 Three dimensional Bayesian optical diffusion tomography with experimental data *Opt. Lett.* **27** 95–7

[18] Eppstein M J, Dougherty D E, Hawysz D J and Sevick-Muraca E M 1999 Three-dimensional Bayesian optical image reconstruction with domain decomposition *IEEE Trans. Med. Imaging* **23** 147–63

[19] Prince S, Kolehmainen V, Kaipio J P, Franceschini M A, Boas D and Arridge S R 2003 Time-series estimation of biological factors in optical diffusion tomography *Phys. Med. Biol.* **48** 1491–504

[20] Schweiger M, Arridge S R and Nissilä I 2005 Gauss–Newton method for image reconstruction in diffuse optical tomography *Phys. Med. Biol.* **50** 2365–86

[21] Calvetti D, Kaipio J P and Somersalo E 2005 Aristotelian prior boundary conditions *Int. J. Math. Comput. Sci.* at press

[22] Melsa J L and Cohn D L 1978 *Decision and Estimation Theory* (New York: McGraw-Hill)