Available online at www.sciencedirect.com
ScienceDirect
Applied Numerical Mathematics 59 (2009) 877–890
APPLIED NUMERICAL MATHEMATICS
www.elsevier.com/locate/apnum

Model reduction in state identification problems with an application to determination of thermal parameters

Janne M.J. Huttunen *, Jari P. Kaipio

Department of Physics, University of Kuopio, P.O. Box 1627, 70211 Kuopio, Finland

Available online 29 March 2008

Abstract

Large-dimensional parameter estimation problems are often computationally unstable and are therefore characterized as ill-posed inverse problems. Inverse problems tolerate measurement and modelling errors poorly which usually calls for accurate computational implementations of the underlying models. These implementations often turn out to be computationally too demanding for a specific application, especially in case of time-varying problems. The so-called approximation error approach has recently been developed to cope with both modelling and numerical discretization errors. This approach has been applied to both stationary (time-invariant) and nonstationary problems. Given a fixed available computational capacity, the employment of the approximation error approach usually yields significantly better estimates than with a conventional error model. In addition, the error estimates are more feasible than with a conventional error model. In this paper we extend the previous results and provide computationally efficient forms for the extended Kalman filters for large-dimensional state identification problems. We apply the approach to the determination of distributed thermal parameters of tissue. In the measurement setting the tissue is heated with focused ultrasound and the temperature evolution is observed through magnetic resonance imaging.

Keywords: Parameter identification; Approximation error theory; Adaptive estimation; Model reduction; Thermal tissue parameters

1. Introduction

The state-space estimation problems are problems in which the unknown state X_k of the system and the measurements Y_k are modelled using the state-space model comprising of the evolution and observation models:

$$\begin{cases} X_{k+1} = F_{k+1}(X_k, W_{k+1}), & k = 1, \ldots, N-1, \\ Y_k = G_k(X_k) + V_k, & k = 1, \ldots, N, \end{cases}$$

where F_k and G_k are functions, W_k is the state noise and V_k is the measurement noise. The goal is to estimate the state X_k based on the measurements $Y_1, \ldots, Y_k$. In the linear Gaussian case, optimal estimates in the mean square sense can be computed using the Kalman filter (KF) recursion [30]. For nonlinear state-space models, suboptimal estimates can be computed using the extended Kalman filter (EKF) [1].

* Corresponding author.
E-mail address: jmhuttun@physics.uku.fi (J.M.J. Huttunen).

0168-9274/$30.00
doi:10.1016/j.apnum.2008.03.029

In this paper we consider problems in which a state-space model contains unknown parameters which are of interest. These problems are called as state identification problems. In particular, we address parameters estimation problems in which the state dimension can be very large.

This kind of a situation occurs when the evolution model is derived from, say, a parabolic partial differential equation and the distributed parameters are to be determined. As an example, later in this paper we study the estimation of the thermal parameters of human tissue, in our case, the thermal conductivity and perfusion coefficients.

Large-dimensional state estimation and identification problems often have the instability characteristics of ill-posed inverse problems. Within the inverse problem community these are referred to as nonstationary inverse problems, in distinction with stationary inverse problems in which all unknowns are time-independent [28]. In many applications which lead to inverse problems, measurements are obtained sequentially and therefore the measurements can correspond to different targets. This situation leads to nonstationary inverse problems.

When solving inverse problems, the available computational resources are often limited. This is especially the case with control related problems and other applications with continuous data streams. These limitations may lead to computational realizations that render too poor, or even meaningless estimates. On the other hand, in many situations one can use significantly larger computational resources prior to the onset of the data stream or the primary application. In such cases the approximation and modelling error approach can be employed [28,29,22,21].

In this approach the statistical discrepancy between the predictions of an accurate model and an approximate model is computed. This discrepancy is called the approximation error and is then interpreted as a (state-dependent) additional error term which may find itself in the observation and/or state evolution models. The approach is Bayesian in that the computation of the approximation error statistics is carried out over the modelled prior distribution of the unknowns.

In stationary inverse problems the approach has been successfully applied to several applications, mainly in tomography-related inverse problems such as electrical impedance tomography and optical tomography. The sources of approximation errors include conventional (mesh) model reduction in partial differential equations [28,5], using isotropic model in case of anisotropy [19], and unknown boundary data in case of truncation of the computational domain [32].

In state-space identification problems there are two categories of approaches. If the model parameters are inherently time varying, one needs to set up a first order Markov model for their evolution and then augment the state evolution model. For example, in heat equation related problems, in which the standard state variable would be temperature, the ambient temperature can also be time-varying unknown variable and therefore the respective evolution model would be added to the evolution model for the temperature itself.

If the model parameters are time-invariant, iterative block algorithms can be employed to compute, for example, the maximum likelihood or some Bayesian estimates. However, the identification of time-invariant parameters can also be tackled with the time-varying parameter approach by adjusting the properties of the evolution model.

In this paper we consider the approximation error methods for state-space identification problems. We use a Markov model for the model parameters and apply the approximation error analysis for the augmented problem. We also derive a version for the extended Kalman filter specifically for problems in which the number of measurements is high. The filter is a hybrid of the innovation and information filter forms for the Kalman filter [1]. This hybrid filter is better suited for computationally efficient implementation of the approximation error for the EKF. However, the filter is also applicable to general nonlinear state-space estimation problems.

Furthermore, we apply the hybrid filter to a relatively large-dimensional problem for which the approach is meant for in the first place, rather than class room examples that have been studied earlier.

This paper is organized as follows. In Section 2 we describe the nonstationary parameter estimation problem which is considered in this paper and formulate the problem as a state-space identification problem. In Section 3 we analyze approximation errors in the associated state-space models and derive the modified state-space model. In Section 4 we derive recursive estimator formulas, the hybrid filter, for the parameter estimation. In Section 5 we apply the proposed approach to the estimation of the thermal parameters of human tissue when the temperature observations are derived from diffuse tensor mode using magnetic resonance imaging. A discussion is given in Section 6.

2. Nonstationary parameter identification

We consider the following nonlinear state-space model

$$\begin{cases} X_{k+1} = F_{k+1}(X_k, \theta, W_{k+1}), & k = 1, \ldots, N-1, \\ Y_k = G_k(X_k, \theta) + V_k, & k = 1, \ldots, N, \end{cases} \tag{1}$$

where X_k is the state of the system, Y_k are measurements, W_k is the state noise, V_k is the observation noise, θ is a vector containing unknown state-space model parameters and F_k and G_k are known mappings. The problem is to determine the unknown parameters θ based on measurements $Y_1, \ldots, Y_N$.

The unknown parameters θ can be determined with several different approaches. For example, when the model parameters are time-invariant, we can try to construct the likelihood function $\ell(Y_1, \ldots, Y_N|\theta)$ and compute the maximum likelihood estimate for θ. For other approaches, see [17,1]. In this paper we use the adaptive estimation approach, that is, we augment the state to include the unknown parameters and use state-space estimation techniques for the augmented model.

For this end, we treat the unknown parameters θ as a stochastic process and set $X_k^{\mathrm{A}} = (X_k, \theta_k)$. When estimating time-invariant parameters with this approach, it is customary to write a random walk model for θ_k, although more complicated models could also be used. Ideally, we would set $\theta_{k+1} = \theta_k + \omega_{k+1}$ with vanishing state noise ω_k accompanied with a feasible initial mean and covariance. Depending on the structure of the problem, however, we would use a nonvanishing covariance for the state noise for stability reasons. Thus an evolution model for X_k^{A} can be written as

$$X_{k+1}^{\mathrm{A}} = \begin{pmatrix} X_{k+1} \\ \theta_{k+1} \end{pmatrix} = \begin{pmatrix} F_{k+1}(X_k, \theta_k, W_{k+1}) \\ \theta_k + \omega_{k+1} \end{pmatrix} =: F_{k+1}^{\mathrm{A}}\big(X_k^{\mathrm{A}}, W_{k+1}^{\mathrm{A}}\big), \tag{2}$$

where $W_k^{\mathrm{A}} = (W_k, \omega_k)$. The estimates for the augmented state X_k^{A} can then be computed using extended Kalman filtering, see [1]. It is to be noted, however, that EKF estimates are not strictly speaking optimal in any sense. The estimates for the model parameters θ are obtained directly from the estimate of X_N^{A}.

3. Model reduction errors

In this section we give a brief account of the approximation error approach for state-space problems. We consider here only the case of pure model reduction, see [21]. In this paper we do not focus on how model reduction is carried out per se, but rather on how the model reduction related errors can be treated. For model reduction methods and approaches, see [3].

The approximation error approach for stationary cases was introduced in [28,29]. Some aspects of modelling errors that are related to a specific nonstationary inverse problem, a process tomography application with electrical impedance tomographic measurements and convection–diffusion state evolution models with unknown boundary data, were already treated in [43,42] and their semigroup theory in [40,41]. The numerical verification related to [40,41] is carried out in [23].

The more general nonstationary linear extension that considers state reduction with respect to both state and evolution equations and the modified Kalman recursions was given in [22]. The nonlinear extension with the extended Kalman filter (EKF) recursions was presented [21].

Assume that we have obtained the following (discretized) state-space model that is adequately accurate for our purposes but computationally too heavy:

$$\begin{cases} X_{k+1}^{\mathrm{D}} = F_{k+1}^{\mathrm{D}}(X_k^{\mathrm{D}}, \theta_k, W_{k+1}^{\mathrm{D}}), & k = 0, \ldots, N-1, \\ Y_k = G_k^{\mathrm{D}}(X_k^{\mathrm{D}}, \theta_k) + V_k^{\mathrm{D}}, & k = 1, \ldots, N, \end{cases} \tag{3}$$

where X_k^{D} is the state of the system, F_k^{D} and G_k^{D} are known mappings, W_{k+1}^{D} is the state noise and V_k^{D} is the measurement noise, which are generally not the same process as in (1). This evolution model is assumed to be accurate in the sense that the uncertainties caused by discretization are negligible with respect to the state and observation noise processes. When, for example, finite element schemes are used for numerical discretization, this model can be constructed with fine spatial and temporal discretization. However, this model often turns out to be a very high-dimensional one and computationally heavy and therefore we avoid to use this model in the identification process.

We assume that a low-dimensional state X_k can be obtained from X_k^{D} by a linear operator (typically a projection) P, that is, $X_k = PX_k^{\mathrm{D}}$. We write

$$\begin{aligned} X_{k+1} &= PX_k^{\mathrm{D}} = PF_{k+1}^{\mathrm{D}}\big(X_k^{\mathrm{D}}, \theta_k, W_{k+1}^{\mathrm{D}}\big) \\ &= F_{k+1}(X_k, \theta_k, W_{k+1}) + PF_{k+1}^{\mathrm{D}}\big(X_k^{\mathrm{D}}, \theta_k W_{k+1}^{\mathrm{D}}\big) - F_{k+1}(X_k, \theta_k, W_{k+1}) \\ &= F_{k+1}(X_k, \theta_k, W_{k+1}) + \epsilon_{k+1} \end{aligned} \tag{4}$$

where the stochastic process

$$\epsilon_k = PF_k^{\mathrm{D}}\big(X_{k-1}^{\mathrm{D}}, \theta_{k-1}, W_k^{\mathrm{D}}\big) - F_k(X_{k-1}, \theta_{k-1}, W_k), \quad k = 1, \ldots, N, \tag{5}$$

represents the errors in the evolution model caused by model reduction. Similarly, the observation model becomes

$$\begin{aligned} Y_k &= G_k(X_k, \theta_k) + G_k^{\mathrm{D}}\big(X_k^{\mathrm{D}}, \theta_k\big) - G_k(X_k, \theta_k) + V_k \\ &= G_k(X_k, \theta_k) + \nu_k + V_k \end{aligned} \tag{6}$$

where the process

$$\nu_k = G_k^{\mathrm{D}}\big(X_k^{\mathrm{D}}, \theta_k\big) - G_k(X_k, \theta_k), \quad k = 1, \ldots, N, \tag{7}$$

represents model reduction error in the observation model.

By using equations (4) and (6), the augmented state-space model for the system can be written as

$$\begin{cases} X_{k+1}^{\mathrm{A}} = F_{k+1}^{\mathrm{A}}(X_k^{\mathrm{A}}, W_{k+1}^{\mathrm{A}}) + \epsilon_{k+1}^{\mathrm{A}}, & k = 0, \ldots, N-1, \\ Y_k = G_k(X_k^{\mathrm{A}}) + \nu_k + V_k, & k = 1, \ldots, N, \end{cases} \tag{8}$$

where $\epsilon_k^{\mathrm{A}} = (\epsilon_k, 0)$. This model can be used in the estimation if the statistics of the processes ϵ_k and ν_k is known.

3.1. Temporal discretization

In physical problems, the state evolution equations are often derived from parabolic partial differential equations. In these problems, spatial discretization leads to stiff differential equations and accurate modelling requires that the temporal integration should be relatively fine. In this section, the aim is to use longer time stepping than would usually be accepted, see also [21].

The evolution model corresponding to temporal integration is written as

$$\bar{X}_{j+1} = \bar{F}_{j+1}(\bar{X}_j, \theta_j, \overline{W}_{j+1}), \quad j = 0, \ldots, \bar{N} - 1, \tag{9}$$

where $\bar{X}_j$ is the state of the system, $\bar{F}_j$ are known mappings and $\overline{W}_j$ is the state noise. We form a mapping F_k and a process W_k for a state-space model of the form (1) corresponding to measurement times.

For simplicity, we assume that the time between for measurements is a multiple of the time step used in temporal integration, that is, $X_k = \bar{X}_{kN_\tau}$ for some integer N_τ. Furthermore, again for simplicity, we assume that θ_j does not change between measurements, i.e., $\theta_j = \theta_k$ if $j = kN_\tau, \ldots, (k+1)N_\tau - 1$.

We use Eq. (9) to compute X_{k+1} from X_k recursively, i.e., we start from $\bar{X}_{kN_\tau} = X_k$, we compute $\bar{X}_{kN_\tau+j}$ for all $j = 1, \ldots, N_\tau$ by using model (9) recursively, and finally we set $F_{k+1}(X_k, W_{k+1}) := \bar{X}_{kN_\tau+N_\tau}$ where

$$W_{k+1} = (\overline{W}_{kN_\tau+1}, \ldots, \overline{W}_{kN_\tau+N_\tau}). \tag{10}$$

For details, see [21].

For the recursive estimator presented in the next section, we need to compute Jacobian matrices of F_{k+1}. These Jacobian matrices can be computed using the chain rule, see [21] for details.

4. Recursive estimator formulas

The extended Kalman filter (EKF) in its usual form is not directly applicable with the model (8) due to the fact that the error processes ϵ_k and ν_k depend on the state, see [21]. In this section we present an extension of EKF to accommodate the model (8). Furthermore, we describe how the statistics of the error processes ϵ_k and ν_k can be

computed, and finally we introduce a modification to the extension for cases in which the number of measurements is high compared to the number of state variables.

Denote by $\hat{X}_k^{\mathrm{A}}$ and $\hat{\Gamma}_k^{\mathrm{A}}$ the filter estimate and the error covariance for the augmented state X_k^{A} based on data $Y_1, \ldots, Y_k$, respectively. Furthermore, denote the predictor (i.e., the estimate based on data $Y_1, \ldots, Y_{k-1}$) and its error covariance by $\tilde{X}_k^{\mathrm{A}}$ and $\tilde{\Gamma}_k^{\mathrm{A}}$, respectively. The extension of EKF for the error model (8) is given by the formulas (see [21])

$$\tilde{X}_{k+1}^{\mathrm{A}} = F_{k+1}^{\mathrm{A}}\big(\hat{X}_k^{\mathrm{A}}\big) + \bar{\epsilon}_{k+1}^{\mathrm{A}}, \tag{11}$$

$$\tilde{\Gamma}_{k+1}^{\mathrm{A}} = \partial F_{k+1}^{\mathrm{A}} \hat{\Gamma}_k^{\mathrm{A}} \partial F_{k+1}^{\mathrm{A}}{}^{\mathrm{T}} + \partial_w F_{k+1}^{\mathrm{A}} \Gamma_{W_{k+1}^{\mathrm{A}}} \partial_w F_{k+1}^{\mathrm{A}}{}^{\mathrm{T}} + \Gamma_{\epsilon_{k+1}^{\mathrm{A}}}, \tag{12}$$

$$K_{k+1} = \tilde{\Gamma}_{k+1}^{\mathrm{A}} \partial G_{k+1}^{\mathrm{T}} \big[\partial G_{k+1} \tilde{\Gamma}_{k+1}^{\mathrm{A}} \partial G_{k+1}^{\mathrm{T}} + \Gamma_{\nu_{k+1}} + \Gamma_{V_{k+1}}\big]^{-1}, \tag{13}$$

$$\hat{X}_{k+1}^{\mathrm{A}} = \tilde{X}_{k+1}^{\mathrm{A}} + K_{k+1}\big(Y_{k+1} - G_{k+1}\big(\tilde{X}_{k+1}^{\mathrm{A}}\big) - \bar{\nu}_{k+1}\big), \tag{14}$$

$$\hat{\Gamma}_{k+1}^{\mathrm{A}} = (I - K_{k+1} \partial G_{k+1}) \tilde{\Gamma}_{k+1}^{\mathrm{A}} \tag{15}$$

where $\partial F_{k+1}^{\mathrm{A}}$ and $\partial_w F_{k+1}^{\mathrm{A}}$ are the Jacobian matrices of F_{k+1}^{A} with respect to X^{A} and w, respectively, evaluated at $(\hat{X}_k^{\mathrm{A}}, 0)$, ∂G_{k+1} is the Jacobian matrix of G_{k+1} evaluated at $\tilde{X}_{k+1}^{\mathrm{A}}$, $\bar{\epsilon}_{k+1}^{\mathrm{A}}$ and $\bar{\nu}_{k+1}$ are the expectations of $\epsilon_{k+1}^{\mathrm{A}}$ and ν_{k+1}, respectively, and $\Gamma_{W_{k+1}^{\mathrm{A}}}$, $\Gamma_{\epsilon_{k+1}^{\mathrm{A}}}$, $\Gamma_{V_{k+1}}$ and $\Gamma_{\nu_{k+1}}$ are the covariances of W_{k+1}^{A}, $\epsilon_{k+1}^{\mathrm{A}}$, V_{k+1} and ν_{k+1}, respectively.

The matrix K_{k+1} given by Eq. (13) is typically called the *Kalman gain* and its usual interpretation is to be a "gain" matrix which, multiplied with the difference between the actual and predicted measurements, gives an update for the state variable (cf. Eq. (14)).

For the initial state, we set $\tilde{X}_0^{\mathrm{A}}$ and $\tilde{\Gamma}_0^{\mathrm{A}}$ based on our prior uncertainty of the initial state and the parameters. The next task is to compute the expectation and the covariance of ϵ_k and ν_k.

4.1. The expectation and the covariance of ϵ_k^{A} and ν_k

In the case of linear state-space models, the expectation and the covariance of ϵ_k and ν_k can be computed explicitly. However, in the present case the evolution model is nonlinear and similar explicit equations do not generally exists. In this paper the expectations and the covariances are determined by averaging over the prior model.

We use the accurate model (3) and the initial state model (prior model) for $(X_0^{\mathrm{D}}, \theta)$ to generate samples from the evolution. In some cases we can assume that the evolution model for X_k^{D} which is conditioned on the initial state is such that the evolution has vanishing covariance. In other words, the conditional evolution model can be treated as a deterministic model and therefore we can set $W_k^{\mathrm{D}} = 0$ for the state noise. This is warranted if we can assume that the model per se is the correct one. In many practical cases, however, the models are only approximate and also the numerically accurate evolution model has to be equipped with a nonvanishing state noise term. In such cases, the determination of the state noise may be a tedious task, see [43]. We draw N_{s} samples θ^j from this distribution as well as N_{s} samples $X_0^{\mathrm{D},j}$ from the distribution of the initial state. In addition, we have to form a realization of the state noise $W_{k+1}^{\mathrm{D},j}$. Then the following steps are carried out for each time step $k = 1, \ldots, N$:

(i) update the state:

$$X_{k+1}^{\mathrm{D},j} = F_{k+1}^{\mathrm{D}}\big(X_k^{\mathrm{D},j}, \theta^j, W_{k+1}^{\mathrm{D},j}\big) \quad \text{and set} \quad X_{k+1}^j = P X_{k+1}^{\mathrm{D},j};$$

(ii) compute samples for ϵ_k and ν_k:

$$\epsilon_k^j = P F_k^{\mathrm{D}}\big(X_{k-1}^{\mathrm{D},j}, \theta^j, W_{k+1}^{\mathrm{D},j}\big) - F_k\big(X_{k-1}^j, \theta^j, W_{k+1}^{\mathrm{D},j}\big),$$
$$\nu_k^j = G_k^{\mathrm{D}}\big(X_k^{\mathrm{D},j}, \theta^j\big) - G_k\big(X_k^j, \theta^j\big);$$

(iii) compute sample mean and sample covariance for ϵ_k and ν_k:

$$\bar{\epsilon}_k \approx \frac{1}{N_{\mathrm{s}}} \sum_{j=1}^{N_{\mathrm{s}}} \epsilon_k^j, \qquad \Gamma_{\epsilon_k} \approx \frac{1}{N_{\mathrm{s}} - 1} \sum_{j=1}^{N_{\mathrm{s}}} \big(\epsilon_k^j - \bar{\epsilon}_k\big)\big(\epsilon_k^j - \bar{\epsilon}_k\big)^{\mathrm{T}},$$

$$\bar{\nu}_k \approx \frac{1}{N_\mathrm{s}} \sum_{j=1}^{N_\mathrm{s}} \nu_k^j, \qquad \Gamma_{\nu_k} \approx \frac{1}{N_\mathrm{s}-1} \sum_{j=1}^{N_\mathrm{s}} \big(\nu_k^j - \bar{\nu}_k\big)\big(\nu_k^j - \bar{\nu}_k\big)^\mathrm{T}.$$

For ϵ_k^A we have $\bar{\epsilon}_k^\mathrm{A} = (\bar{\epsilon}_k, 0)$ and $\Gamma_{\epsilon_k^\mathrm{A}} = \mathrm{diag}(\Gamma_{\epsilon_k}, 0)$. When N_s is increased, the quality of the estimates of the expectations and covariances is typically improved. The convergence is guaranteed (by the law of large numbers) when the generated samples are independent and the error processes ϵ_k and ν_k are square integrable. The latter condition depends on the nature of the mappings F_k and G_k as well as the distributions from which the samples are generated.

Contrary to general statistical principles, it has turned out that in practice the number of samples N_s does not have to be high when compared to the number of (in this case) observations and state variables [28,32]. There is not a proof of this property, however. Technically this means that the covariance estimates are rank-deficient, but this turns out not to be a problem.

It is to be noted that the expectations $\bar{\epsilon}_k$ and $\bar{\nu}_k$ and the covariance matrices Γ_{ϵ_k} and Γ_{ν_k} should be computed and stored for all $k = 1, \ldots, N$ before the presented estimation algorithm can be used. However, if the number of measurements is large, Γ_{ν_k} can be very large dense matrix and therefore storage requirements can be enormous. Furthermore, the computation of the inverse matrix in the Kalman gain equation (13) can be an excessively complex task. In the following we present a modification to the filter which allows us to save storage space and reduce computational complexity.

4.2. Hybrid filter

The form (13) is not computationally efficient if the number of measurements is high compared to the dimension of the state-space. With standard Kalman filters we would use the information filter that uses a modified state variable and propagates the inverses of the covariances rather than the covariances themselves [1]. For example, in the situation described in Section 5, the measurements are MR thermal images which are often at least of the size 128×128 so that we have of the order 10^4–10^6 measurements. Therefore, for example, $\partial G_{k+1} \tilde{\Gamma}_{k+1}^\mathrm{A} \partial G_{k+1}{}^\mathrm{T}$ and $\Gamma_{\nu_{k+1}}$ as full matrices would be too large to be stored and too heavy to compute.

For the EKF equipped with the approximation errors, the information filter as such is not feasible. We can, however, make use of the Sherman–Morrison–Woodbury (SMW) formula [16] to derive more efficient forms for Eqs. (13)–(15). The result is a hybrid since it does not use the modified state, which would also turn out to be computationally infeasible. We note that the SMW formula can be used to show the equivalence of the innovation and information filters.

The Sherman–Morrison–Woodbury formula can be written as

$$\big(ARA^\mathrm{T} + P\big)^{-1} = P^{-1} - P^{-1}A\big(A^\mathrm{T}P^{-1}A + R^{-1}\big)^{-1}A^\mathrm{T}P^{-1}$$

where A, P and R are matrices such that R and P are invertible. If we choose $P = \Gamma_\nu + \Gamma_V$, $A = \partial G$ and $R = \tilde{\Gamma}$ (we momentarily omit the subscripts $k+1$ and the superscript A to simplify notations), the SMW formula gives for the Kalman gain

$$\begin{aligned}
K &= \tilde{\Gamma}\partial G^\mathrm{T}\big[\partial G\tilde{\Gamma}\partial G^\mathrm{T} + \Gamma_\nu + \Gamma_V\big]^{-1} \\
&= \tilde{\Gamma}\partial G^\mathrm{T}P^{-1} - \tilde{\Gamma}\partial G^\mathrm{T}P^{-1}\partial G\big(\partial G^\mathrm{T}P^{-1}\partial G + \tilde{\Gamma}^{-1}\big)^{-1}\partial G^\mathrm{T}P^{-1} \\
&= \tilde{\Gamma}\partial G^\mathrm{T}P^{-1} - \tilde{\Gamma}\big(\partial G^\mathrm{T}P^{-1}\partial G + \tilde{\Gamma}^{-1}\big)\big(\partial G^\mathrm{T}P^{-1}\partial G + \tilde{\Gamma}^{-1}\big)^{-1}\partial G^\mathrm{T}P^{-1} \\
&\quad + \tilde{\Gamma}\tilde{\Gamma}^{-1}\big(\partial G^\mathrm{T}P^{-1}\partial G + \tilde{\Gamma}^{-1}\big)^{-1}\partial G^\mathrm{T}P^{-1} \\
&= \tilde{\Gamma}\partial G^\mathrm{T}P^{-1} - \tilde{\Gamma}\partial G^\mathrm{T}P^{-1} + \big(\partial G^\mathrm{T}P^{-1}\partial G + \tilde{\Gamma}^{-1}\big)^{-1}\partial G^\mathrm{T}P^{-1} \\
&= \big[\partial G^\mathrm{T}(\Gamma_\nu + \Gamma_V)^{-1}\partial G + \tilde{\Gamma}^{-1}\big]^{-1}\partial G^\mathrm{T}(\Gamma_\nu + \Gamma_V)^{-1}.
\end{aligned} \tag{16}$$

The covariances Γ_V are often diagonal and can be inverted easily. However, the covariance Γ_ν is not generally sparse and therefore $\Gamma_\nu + \Gamma_V$, as large full matrices, may be heavy to be inverted.

To overcome this potential problem, we note that the covariance of ν_k can be written in the form $\Gamma_\nu = \Xi\Xi^\mathrm{T}$ without any further computational load. This is due to the fact that the covariance estimate can be written as an outer

product in the first place, that is, $\Gamma_\nu = (N_\mathrm{s}-1)^{-1}\mathrm{X_s}\mathrm{X_s^T}$ where $\mathrm{X_s}$ is a matrix which columns are vectors $\nu^j - \bar{\nu}$ (i.e., the samples of ν subtracted by the means). Thus we can set $\Xi = (N_\mathrm{s}-1)^{-1/2}\mathrm{X_s}$. The rank-deficiency of Γ_ν is usually not a problem since Γ_V is in most cases positive definite. An approximate decomposition of this type can also be obtained by using eigenvalue decomposition.

The SMW formula with choices $A = \Xi$, $P = \Gamma_V$ and $R = I$ gives

$$(\Gamma_\nu + \Gamma_V)^{-1} = \Gamma_V^{-1} - \Gamma_V^{-1}\Xi\big(\Xi^\mathrm{T}\Gamma_V^{-1}\Xi + I\big)^{-1}\Xi^\mathrm{T}\Gamma_V^{-1}.$$

Denote $\Lambda = (\Xi^\mathrm{T}\Gamma_V^{-1}\Xi + I)^{-1}$. Then

$$\begin{aligned} K &= \big[\partial G^\mathrm{T}\big(\Gamma_V^{-1} - \Gamma_V^{-1}\Xi\Lambda\Xi^\mathrm{T}\Gamma_V^{-1}\big)\partial G + \tilde{\Gamma}^{-1}\big]^{-1}\partial G^\mathrm{T}\big(\Gamma_V^{-1} - \Gamma_V^{-1}\Xi\Lambda\Xi^\mathrm{T}\Gamma_V^{-1}\big) \\ &= \big[\tilde{\Gamma}^{-1} + \partial G^\mathrm{T}\Gamma_V^{-1}\partial G - \partial G^\mathrm{T}\Gamma_V^{-1}\Xi\Lambda\Xi^\mathrm{T}\Gamma_V^{-1}\partial G\big]^{-1}\big(\partial G^\mathrm{T}\Gamma_V^{-1} - \partial G^\mathrm{T}\Gamma_V^{-1}\Xi\Lambda\Xi^\mathrm{T}\Gamma_V^{-1}\big). \end{aligned}$$

Furthermore, let $\Phi = \Xi^\mathrm{T}\Gamma_V^{-1}\partial G$ and $\Psi = \tilde{\Gamma}^{-1} + \partial G^\mathrm{T}\Gamma_V^{-1}\partial G - \Phi^\mathrm{T}\Lambda\Phi$. Thus the Kalman gain become

$$K = \Psi^{-1}\big(\partial G^\mathrm{T} - \Phi^\mathrm{T}\Lambda\Xi^\mathrm{T}\big)\Gamma_V^{-1}. \tag{17}$$

The Kalman gain matrix for the recursive estimator can be computed using this formula. The computationally most demanding tasks are to compute $(\tilde{\Gamma}^\mathrm{A})^{-1}$ for the approximate model and invert the matrix Ψ.[1]

Consider a situation in which the mapping G_{k+1} is linear. This kind of a situation occurs in the simulation presented in Section 5. Substitution of (17) to Eqs. (14) and (15) gives

$$\begin{aligned} \hat{X}_{k+1}^\mathrm{A} &= \tilde{X}_{k+1}^\mathrm{A} + \Psi_{k+1}^{-1}G_{k+1}^\mathrm{T}\Gamma_{V_{k+1}}^{-1}\bar{Y}_{k+1} - \Psi_{k+1}^{-1}G_{k+1}^\mathrm{T}\Gamma_{V_{k+1}}^{-1}G_{k+1}\tilde{X}_{k+1}^\mathrm{A} \\ &\quad - \Psi_{k+1}^{-1}\Phi_{k+1}^\mathrm{T}\Lambda_{k+1}\Xi_{k+1}^\mathrm{T}\Gamma_{V_{k+1}}^{-1}\bar{Y}_{k+1} + \Psi_{k+1}^{-1}\Phi_{k+1}^\mathrm{T}\Lambda_{k+1}\Phi_{k+1}\tilde{X}_{k+1}^\mathrm{A}, \end{aligned} \tag{18}$$

$$\hat{\Gamma}_{k+1}^\mathrm{A} = \tilde{\Gamma}_{k+1}^\mathrm{A} - \Psi_{k+1}^{-1}G_{k+1}^\mathrm{T}\Gamma_{V_{k+1}}^{-1}G_{k+1}\tilde{\Gamma}_{k+1}^\mathrm{A} + \Psi_{k+1}^{-1}\Phi_{k+1}^\mathrm{T}\Lambda_{k+1}\Phi_{k+1}\tilde{\Gamma}_{k+1}^\mathrm{A}, \tag{19}$$

where $\bar{Y}_{k+1} = Y_{k+1} - \bar{\nu}_{k+1}$. This form is computationally more efficient compared to the form consisting of Eqs. (17), (14) and (15). However, the multiplication order should be chosen carefully.

Finally, we note that with this hybrid filter, one needs to store the matrices $\mathrm{X_s}$ (for all $k = 1, \ldots, N$) instead of the covariances matrices $\Gamma_{\nu_1}, \ldots, \Gamma_{\nu_N}$. This is a huge benefit with respect to storage requirements. Namely, the number of samples N_s can be much smaller than the number of measurements m and therefore the matrices $\mathrm{X_s}$ containing $m \times N_\mathrm{s}$ elements require significantly less space than the covariances Γ_{ν_k} that are dense $m \times m$-matrix.

5. Estimation of the thermal parameters of tissue

As an example we consider the estimation of the thermal properties of tissue, in our case, the heterogeneous thermal conductivity and perfusion coefficients. Approaches for this task are described in [10,49,20], for similar problems, see also [4,12,11,48,47]. In [10,49,20] this problem was attacked by using ultrasound induced heating of the tissue and magnetic resonance (MR) thermal imaging [37,27] to obtain temperature measurements.

The motivation in [10,49,20] was related to model-based optimal control for ultrasound surgery, in which task the thermal properties are assumed to be known when treatment is planned [14,35,49]. In ultrasound surgery MR temperature imaging can used to ensure correct location of a focal spot of the applied ultrasound field [36]. Also, the temperature data can be used as the system feedback for the model-based treatment planning methods, see [34]. Furthermore, MR images can be used to ensure that the whole target tissue (tumor) is ablated (clinically destroyed) after the treatment [36,50]. Whereas in ultrasound surgery the tissue temperature is raised to 60°–95 °C, in the approach [10,49,20] the tissue is warmed only by a couple of degrees so that the tissue is not permanently damaged.

The overview of the approach is as follows. A number of ultrasound transducers are placed on the skin and are driven typically by 10^4–10^6 Hz scale frequencies. The amplitude and relative phase of each transducer is controlled so as to guide the ultrasound energy and absorption distribution within the tissue. The absorption of energy induces

[1] In practice, the matrix K is, however, computed by solving the equation $\Psi K = (\partial G^\mathrm{T} - \Phi^\mathrm{T}\Lambda\Xi^\mathrm{T})\Gamma_V^{-1}$, for example, using the Cholesky decomposition.

small temperature changes in the tissue and these are observed through MR thermal imaging. The tissue properties are estimated from these measurements, for example, using likelihood estimation scheme.

The approach in this section is similar to [20], with the exception that here we treat the problem as a state-space identification problem and use the approximation error approach to deal with the heavily approximate run time evolution and observation models.

5.1. Heat equation with distributed ultrasound absorption source

Let $\Omega \subset \mathbb{R}^{2,3}$ be a bounded domain corresponding to a part of the human body. The temperature evolution in tissue is most commonly modelled using the Pennes' bioheat equation [38]

$$\rho C_{\mathrm{T}} \frac{\partial T}{\partial t} = \nabla \cdot (\kappa \nabla T) - w_{\mathrm{B}} C_{\mathrm{B}} (T - T_{\mathrm{A}}) + q, \tag{20}$$

where T is the temperature, ρ is the density of the medium, κ is the thermal conductivity, C_{T} and C_{B} are the heat capacity of the tissue and blood, respectively, w_{B} is the blood flow rate, T_{A} is the temperature of arterial blood and q is the (distributed) heat source. In the case of ultrasound absorption, the source term is of the form

$$q = \frac{\alpha |P|^2}{\rho c}, \tag{21}$$

where P is the ultrasound pressure field, α is the absorption coefficient, ρ is the density of the medium and c is the speed of sound [13,39]. In this paper, these coefficients are taken to be known. We denote $\beta = w_{\mathrm{B}} C_{\mathrm{B}}$ which parameter is called the perfusion coefficient.

We endow the problem with the initial and boundary conditions

$$\begin{aligned} T(0,\cdot) &= T_{\mathrm{A}} \quad \text{in } \Omega, \\ T(t,\cdot) &= T_{\mathrm{A}} \quad \text{on } \partial\Omega, \end{aligned}$$

where T_{A} is the ambient (body) temperature which is very slowly varying and can be measured accurately.

All the above coefficients are taken to be distributed parameters. However, within each tissue type they are assumed to be constants. This segmentation into tissue types is a routine task and is carried out, for example, based on MRI images. Furthermore, the parameters are assumed to be constant with respect to temperature, which is warranted since only small temperature changes are inflicted [51,2,52].

The ultrasound field P is assumed to be time-harmonic, which is warranted since the period of vibration is typically five orders of magnitude shorter than the time for observable temperature changes in tissue. Thus we can solve the spatially dependent part p of the wave equation separately. The pressure field p is the solution of the Helmholtz equation

$$\nabla \cdot \left(\frac{1}{\rho} \nabla p \right) + \frac{k^2}{\rho} p = 0, \tag{22}$$

where k is the wave number and ρ is the density of the medium.

In an absorbing medium, the wave number is of the form $k = 2\pi f/c + \mathrm{i}\alpha$, where f is the frequency of the wave field, c is the speed of sound and α is the absorption coefficient [6]. The solution of the Helmholtz equation with high wave number can easily turn out to be a computationally prohibitive task. In this paper, the solution is computed using a full wave approach that is called the ultraweak variational formulation (UWVF) [8,9]. For details of a computationally feasible implementation, see [25,24,20].

5.2. Evolution and observation models

Eq. (20) is discretized spatially using the finite element method with linear basis functions while the temporal integration is carried out by the implicit Euler method. The numerical form for the evolution over the time steps can be written in our case as [20]

$$\begin{aligned} &\left[M + (S + \bar{M}) \Delta\tau \right] \bar{X}_{j+1} = M \bar{X}_j + (S + \bar{M}) \bar{T}_a \Delta\tau + \tilde{M} d_{j+1} \Delta\tau, \\ &\bar{X}_0 = \bar{T}_{\mathrm{A}}, \end{aligned} \tag{23}$$

where $\bar{X}_j$ and d_j are vectors consisting of tissue temperature and intensity $|P|^2$ in nodes, respectively, $\bar{T}_\mathrm{A}$ is a (constant) vector of the arterial blood temperature T_A, the matrices M, S, $\bar{M}$, $\tilde{M}$ are related to the finite element and implicit Euler implementation and $\Delta\tau$ is the time step. It is to be noted that the matrix S depends on κ and $\bar{M}$ depends on β.

Denote by θ the vector of unknown parameters, that is, $\theta = (\kappa, \beta)$, and define a function $\bar{F}$ by

$$\bar{F}(x, \theta, w) = A(\theta)^{-1}\big[Mx + \big(S(\kappa) + \bar{M}(\beta)\big)\bar{T}_a\Delta\tau + \tilde{M}d_{j+1}\Delta\tau\big] + w \tag{24}$$

where $A(\theta) = M + (S(\kappa) + \bar{M}(\beta))\Delta\tau$. Then the evolution model for $\bar{X}_j$ can be written as

$$\bar{X}_j = \bar{F}_{j+1}(\bar{X}_j, \theta_j, \overline{W}_j), \quad j = 1, \ldots, \bar{N} - 1, \tag{25}$$

where $\overline{W}_j$ is the state noise.

The evolution model (25) corresponds to integration over time intervals that are determined by the implicit Euler method. However, as discussed above, these intervals could be longer than usually required. The numerical evolution model corresponding to time stepping over the measurement intervals is carried out as described in Section 3.1. This model is written as

$$X_{k+1} = F_{k+1}(X_k, \theta_k, W_{k+1}), \quad k = 0, \ldots, N - 1,$$

where F_k and W_k are determined as described in Section 3.1. The Jacobian matrices of F_{k+1} can be computed using the chain rule as described in [21].

The measurements are temperature images given by MR thermal imaging. Hence the observation model can be written as

$$Y_k = \mathcal{M}X_k + V_k = \underbrace{(\mathcal{M}\ 0)X_k^\mathrm{A}}_{:=G_k(X_k,\theta_k)} + V_k, \tag{26}$$

where Y_k are measurements at times t_k, $\mathcal{M}$ is an interpolation matrix from the computational mesh to the measurement grid and V_k is measurement noise.

5.3. Computation of the Jacobian matrices

For the computation of the state covariances, we need the Jacobian matrices of F_k and G_k. From Eq. (24), the Jacobian matrices of $\bar{F}_j$ with respect to x and w are $\partial_x\bar{F}_j = A^{-1}M$ and $\partial_w\bar{F}_j = I$, respectively. By multiplying Eq. (24) with A and then differentiating the both sides of the resulting equations, we obtain

$$\big[M + (S + \bar{M})\Delta\tau\big]\frac{\partial\bar{F}_{j+1}}{\partial\kappa^{(r)}} + \frac{\partial S}{\partial\kappa^{(r)}}\bar{F}_{j+1}\Delta\tau = \frac{\partial S}{\partial\kappa^{(r)}}\bar{T}_a\Delta\tau,$$
$$\big[M + (S + \bar{M})\Delta\tau\big]\frac{\partial\bar{F}_{j+1}}{\partial\beta^{(r)}} + \frac{\partial\bar{M}}{\partial\beta^{(r)}}\bar{F}_{j+1}\Delta\tau = \frac{\partial\bar{M}}{\partial\beta^{(r)}}\bar{T}_a\Delta\tau.$$

Consequently, the components of the Jacobian matrix $\partial_\theta\bar{F}_k$ can be constructed iteratively as

$$\frac{\partial\bar{F}_{j+1}}{\partial\kappa^{(r)}} = -A^{-1}\frac{\partial S}{\partial\kappa^{(r)}}(\bar{F}_{j+1} - T_a)\Delta\tau,$$
$$\frac{\partial\bar{F}_{j+1}}{\partial\beta^{(r)}} = -A^{-1}\frac{\partial\bar{M}}{\partial\beta^{(r)}}(\bar{F}_{j+1} - \bar{T}_a)\Delta\tau.$$

The Jacobian matrices for F_{k+1} can be constructed using the chain rule as described in [21]. The Jacobian matrix of the linear mapping G_k is $\partial G_k = (\mathcal{M}\ 0)$.

5.4. Numerical results

In the simulation, we considered a two-dimensional spatial domain based on an X-ray image from a cross section of a breast with a tumor that has been identified and segmented by a medical doctor. The domain was divided into

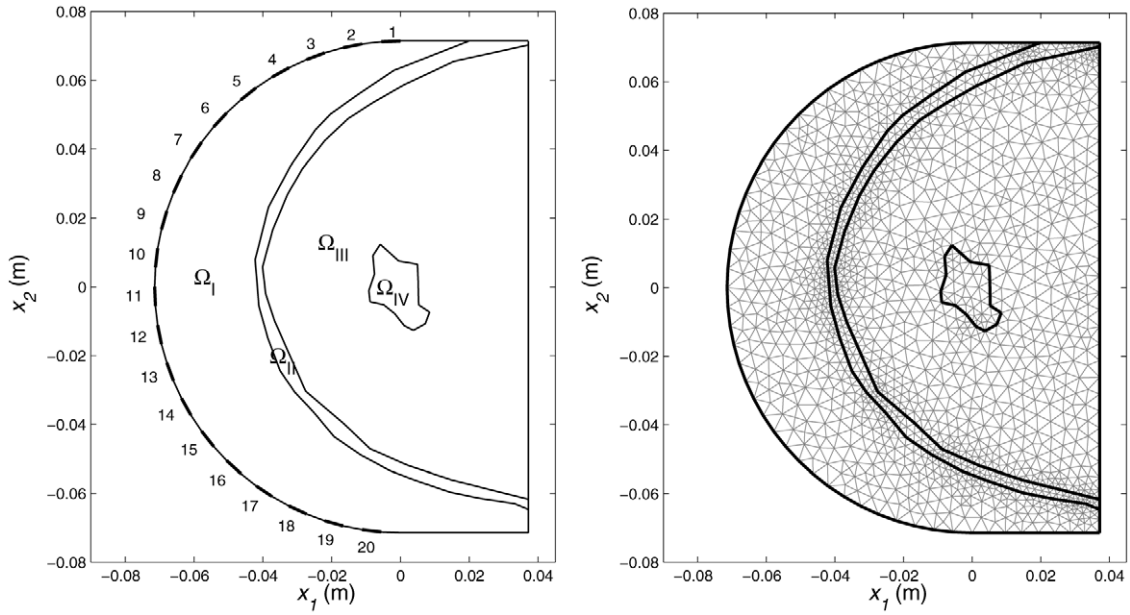

Fig. 1. Left: the computational domain. The domain is divided into four subdomains Ω_{I}–Ω_{IV} in which physical coefficients are assumed to be constants. The numbered black segments on the left boundary represent the 20 ultrasound transducers. Right: the computational mesh consisting of 2965 elements and 1533 nodes.

Table 1
The acoustic parameters used in the simulation

Subdomain	α [Nep/m]	c [m/s]	ρ [kg/m^3]
Ω_{I}	0[a,d]	1500[a]	1000
Ω_{II}	12[a]	1610[a]	1200[b,d]
Ω_{III}	5[a,c]	1485[a]	1020[e]
Ω_{IV}	5[c]	1547[a]	1050[b]

[a] [13]. [b] [15]. [c] [18]. [d][33]. [e][45].

four subdomains: Ω_{I} is a water bed that is needed for the practical setup, Ω_{II} is skin, Ω_{III} is the breast tissue and Ω_{IV} is the tumor. The ultrasound transducer elements were placed on the left side of the domain so that the geometrical focus of the transducer system was located in the center of the subdomain Ω_{IV}. In the treatment phase, the focal spot of ultrasound field can be made to scan over the tumor by adjusting amplitudes and phases of transducers. In the parameter estimation phase the ultrasound field cannot be described as a focused field, see [20]. The computational geometry is shown in Fig. 1.

The acoustic parameters of the tissues were chosen to correspond to the values reported in [13,15,18,33,45] and are shown in Table 1. In conflicting cases we have used the averages of the reported values. The employed thermal parameters are summarized in Table 2.

We set the heat capacity of blood as $C_{\mathrm{B}} = 3770\ \mathrm{J/kg\,K}$ [33] and the ambient temperature as $T_{\mathrm{A}} = 37\,^{\circ}\mathrm{C}$. The conductivity and the perfusion coefficient of water were taken as known. Hence, the unknown parameter vector was $\theta = (\kappa^{(\mathrm{II})}, \kappa^{(\mathrm{III})}, \kappa^{(\mathrm{IV})}, \beta^{(\mathrm{II})}, \beta^{(\mathrm{III})}, \beta^{(\mathrm{IV})})^{\mathrm{T}}$ where the superscripts refer to the subdomains.

The frequency of ultrasound was 500 kHz that corresponds to a wavelength of about 3 mm depending on the tissue type. The Helmholtz equation was solved using the UWVF with a computational mesh consisting of 7113 elements and 3664 nodes. We note that with standard finite element type schemes the number of unknowns would be 10–100 times larger [25].

It is clear that the estimate quality depends, among other things, on the evolution of the distributed heat source. We used optimized transducer excitations that were computed using the method presented in [20]. Loosely speaking, the aim is to cause maximal temperature elevations and local temperature gradients without going over 42 °C anywhere in

Table 2
The thermal parameters used in the simulation. Published values were used to the initial state so that the means were set to correspond to the average values reported in [33,44,31]. The perfusion of tumor varies enormously depending on tumor [46]. In this case the perfusion was chosen to have a higher value than in the surrounding healthy tissue

Subdomain	C_{T} [J/kg K]	Means of initial distribution		True values	
		κ [W/K m]	w_{B} [kg/m^3 s]	κ [W/Km]	w_{B} [kg/m^3 s]
Ω_{I}	4190[a]	0.60[a]	0	0.60	0
Ω_{II}	3770[a]	0.50[a]	1.0[a]	0.65	1.2
Ω_{III}	3550[b,c]	0.50[b]	0.7[b]	0.40	0.9
Ω_{IV}	3770[c]	0.65[c]	2.3	0.80	1.7

[a] [33]. [b] [45]. [c] [31].

tissue domain. The maximum temperature rise caused by the employed excitations was 3.24 °C and the total excitation and measurement time was 228 s.

The synthetic data was generated using a computational mesh consisting of 65 072 elements and 32 831 nodes and the time step was 0.01 s. The computed temperature evolution was interpolated to a 128×128 grid for the computation of the synthetic MRI temperature measurement data. Additive zero mean Gaussian noise with standard deviation of 1.0 °C was added to form the synthetic data. The MRI temperature estimate noise depends on the used MRI sequence, pixel/voxel size and the rate at which the estimates are produced. The used value is typical for the employed setting.

The acquisitions times for MRI temperature maps have been reported to be between 2.3 and 7.2 seconds [26,7]. In this simulation, time between measurements is chosen to be 4.5 s so that the number of measurement frames is 50.

The high-dimensional reference model to compute X_k^{D} was constructed using a computational mesh containing 51 248 elements and 25 823 nodes with time steps of $\Delta t = 0.03$ s. The estimates were computed in a mesh consisting of 2965 elements and 1533 nodes shown in Fig. 1. The time step in temporal integration was set as $\Delta t = 0.15$ s. The required statistics of the approximation errors were computed as described in Section 4.1 (the state noise was ignored). The number of samples was $N_{\mathrm{s}} = 1000$. It is to be noted that the accurate model is only needed for the simulation of the evolution – no state estimates have to be computed with this model.

For the extended Kalman filter recursion, $\hat{X}_0$, $\hat{\Gamma}_0$ and the state noise covariance Γ_{W_k} were chosen as follows. The initial joint distribution of the unknown coefficients and the initial temperature was taken to be Gaussian. The mean of the initial temperature was set to a constant 37 °C. The means of the parameters κ and β are shown in Table 2. The initial temperature and the coefficients were chosen to be mutually independent so that the standard deviations of the initial temperature is 0.1 °C and the standard deviations of the parameters κ and β were set to be 50% of their respective parameter values. The latter standard deviations are so large, that they would permit even nonphysical (negative) coefficient estimates easily. Thus we can claim that the estimates are determined by data only and that the initial state has little or no effect on the estimates. The state noise covariance is a diagonal matrix such that the variance of the components is $0.001^2 \Delta t$. The variance was chosen empirically such that it was the smallest possible level without estimator became unstable.

We computed two estimates: an estimate computed using the approximation error model (8), and an estimate computed using the conventional augmented model (2). We note that the covariances Γ_{V_k} and Γ_{v_k} are $16\,384 \times 16\,384$-matrices. The matrix in the Kalman gain equation (13) which should be inverted is also of the same size and therefore the inversion would be an excessively complex task. Thus the former estimate is computed using the hybrid filter presented in Section 4.2 (Eqs. (18)–(19)) and, for the latter estimate, the Kalman gain matrix is computed with the aid of the SMW equation (Eq. (16) with $\Gamma_{v_k} = 0$). The covariance Γ_{V_k} is diagonal and can be inverted easily. The computational load for computing both estimates during the estimation process is almost the same.

The coefficient estimates are components of the state estimate at final time, that is, $\hat{X}_N^{\mathrm{A}}$, corresponding to the coefficients κ and β. Respectively, we use the square roots $\hat{\sigma}$ of the diagonal elements of the corresponding covariance $\hat{\Gamma}_N$ as error estimates. For the error estimates to be meaningful, the actual estimate errors should seldom be larger than $\pm 2\hat{\sigma}$.

The computed estimates are shown in Table 3. The results indicate that, with this example, the use of the approximation error model gives substantial improvement to the accuracy of the estimates. This is also the case with other examples studied in papers cited elsewhere in this paper. The estimates for subdomain Ω_{II} are clearly worse than for the other subdomains. This is due to the fact that the subdomain Ω_{II} corresponds to the very thin skin layer from which little information is actually obtained.

Table 3
The estimates $\hat{X}$ and the error estimates $\hat{\sigma}$ for the approximation error models (8) and the conventional augmented model (2)

	True	Approximation error model			Conventional model		
		$\hat{X}$	$\hat{\sigma}$	True error	$\hat{X}$	$\hat{\sigma}$	True error
$\kappa^{(\mathrm{II})}$	0.650	0.597	0.025	0.053 ($2.1 \times \hat{\sigma}$)	0.567	0.020	0.083 ($4.2 \times \hat{\sigma}$)
$\kappa^{(\mathrm{III})}$	0.400	0.421	0.014	0.021 ($1.5 \times \hat{\sigma}$)	0.523	0.024	0.123 ($5.2 \times \hat{\sigma}$)
$\kappa^{(\mathrm{IV})}$	0.800	0.732	0.093	0.069 ($0.7 \times \hat{\sigma}$)	1.367	0.142	0.567 ($4.0 \times \hat{\sigma}$)
$\beta^{(\mathrm{II})}$	4524	6777	634	2253 ($3.6 \times \hat{\sigma}$)	2104	485	2420 ($5.0 \times \hat{\sigma}$)
$\beta^{(\mathrm{III})}$	3393	3283	82	110 ($1.3 \times \hat{\sigma}$)	4695	85	1302 ($15.3 \times \hat{\sigma}$)
$\beta^{(\mathrm{IV})}$	6409	6241	318	168 ($0.5 \times \hat{\sigma}$)	8951	360	2542 ($7.1 \times \hat{\sigma}$)

As for the error estimates, the approximation error model can also be said to yield feasible error estimates, whereas the conventional model gives too optimistic error estimates. It is interesting to note the effect of the nonlinearity of the overall model. For linear problem the approximation error model always gives – in addition to better estimates as such – larger error estimates than the conventional model. In the present case the approximation error model actually gives smaller error estimates than the conventional model.

6. Discussion

In this paper we carried out the approximation error analysis for state-space identification problems. We treated the identification problem as an adaptive estimation problem. By computing the second order statistics for the model reduction related errors in the associated evolution and observation models, we derived a modification to the extended Kalman filter. Furthermore, we derived a hybrid filter which is computationally more efficient in a situation in which the number of measurements is high when compared to the dimension of the state vector. As an application we studied estimation of the distributed thermal parameters of tissue.

Given a specified numerical accuracy, the approach permits one to use a computationally more efficient implementation. Conversely, with a specified computational load, more accurate estimates can be obtained. Moreover, the predicted estimation errors are better in accordance with the actual errors.

In the studied example, the approximation error method was able to handle the model reduction well. It is impossible, however, to predict how well the method works with an application without trying out. On the other hand, if an accurate estimation scheme has been set up, the model reduction is easy and straightforward to implement.

Acknowledgements

This work was supported by the Academy of Finland (application number 213476, Finnish Programme for Centres of Excellence in Research 2006-2011) and Finnish Funding Agency for Technology and Innovation (TEKES, project 40285/05). We wish also acknowledge PhD Tomi Huttunen for providing ultrasound fields and PhD Matti Malinen for providing computational meshes.

References

[1] B.D.O. Anderson, J.B. Moore, Optimal Filtering, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1979.
[2] D.P. Anhalt, K. Hynynen, R.B. Roemer, Patterns of changes of tumour temperatures during clinical hyperthermia: implications for treatment plannin, evaluation and control, International Journal of Hyperthermia 11 (3) (1995) 425–436.
[3] A.C. Antoulas, Approximation of Large-Scale Dynamical Systems, Advances in Design and Control, vol. 6, Society for Industrial and Applied Mathematic, Philadelphia, PA, 2005.
[4] H. Arkin, K.R. Holmes, M.M. Chen, W.G. Bottje, Thermal pulse decay method for simultaneous measurement of local thermal conductivity and blood perfusion: a theoretical analysis, Journal of Biomechanical Engineering 108 (3) (1986) 208–214.
[5] S.R. Arridge, J.P. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, M. Vauhkonen, Approximation errors and model reduction with an application in optical diffusion tomography, Inverse Problems 22 (1) (2006) 175–196.
[6] A.B. Bhatia, Ultrasound Absorption: An Introduction to the Theory of Sound Absorption and Dispersion in Gases, Liquids, and Solids, Oxford University Press, 1967.
[7] C. Bohris, J.W. Jenne, R. Rastert, I. Simiantonakis, G. Brix, J. Spoo, M. Hlavac, R. Nemeth, P.E. Huber, J. Debus, MR monitoring of focused ultrasound surgery in a breast tissue model in vivo, Magnetic Resonance Imaging 19 (2001) 167–175.

[8] O. Cessenat, Application d'une nouvelle formulation variationelle des equations d'ondes harmonigues, problemes de Helmholtz 2D et de Maxwell 3D, Ph.D. thesis, Paris IX Dauphine, 1996.
[9] O. Cessenat, B. Despres, Application of an ultra-weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem, SIAM Journal of Numerical Analysis 35 (1) (1998) 255–299.
[10] H.L. Cheng, D.B. Plewes, Tissue thermal conductivity by magnetic resonance thermometry and focused ultrasound heating, Journal of Magnetic Resonance Imaging 16 (5) (2002) 598–609.
[11] S.T. Clegg, R.B. Roemer, A comparative evaluation of unconstrained optimization methods applied to the thermal tomography problem, Transactions of the ASME 107 (1985) 228–233.
[12] A.M. Divrik, R.B. Roemer, T.C. Cetas, Inference of complete tissue temperature fields from a few measured temperatures: An unconstrained optimization method, IEEE Trans. BME 31 (1).
[13] F.A. Duck, A.C. Baker, H.C. Starritt, Ultrasound in Medicine, Institute of Physics Publishing, Bristol, 1998.
[14] B. Erdmann, J. Lang, M. Seebass, Optimization of temperature distributions for regional hyperthermia based on a nonlinear heat transfer model, in: K. Diller (Ed.), Biotransport: Heat and Mass Transfer in Living Systems, Annals of the New York Academy of Sciences 858 (1998) 36–46.
[15] M. Gautherie (Ed.), Methods of External Hyperthermic Heating, Springer-Verlag, New York, 1990.
[16] G.H. Golub, C.F. van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, 1989.
[17] G. Goodwin, R. Payne, Dynamic System Identification, Academic Press, New York, 1977.
[18] S.A. Goss, R.L. Johnston, F. Dunn, Compilation of empirical ultrasonic properties of mammalian tissues. II, The Journal of the Acoustical Society of America 68 (1) (1980) 93–108.
[19] J. Heino, E. Somersalo, J.P. Kaipio, Compensation for geometric mismodelling by anisotropies in optical tomography, Optics Express 13 (2005) 296–308.
[20] J.M.J. Huttunen, T. Huttunen, M. Malinen, J.P. Kaipio, Determination of heterogeneous thermal parameters using ultrasound induced heating and mr thermal mapping, Physics in Medicine and Biology 52 (2006) 1011–1032.
[21] J.M.J. Huttunen, J.P. Kaipio, Approximation error analysis in nonlinear state estimation with an application to state-space identification, Inverse Problems 23 (2007) 2141–2157.
[22] J.M.J. Huttunen, J.P. Kaipio, Approximation errors in nonstationary inverse problems, Inverse Problems and Imaging 1 (1) (2007) 77–93.
[23] J.M.J. Huttunen, H.-K. Pikkarainen, Discretization error in dynamical inverse problems: one-dimensional model case, Journal of Inverse and Ill-Posed Problems 15 (2007) 365–386.
[24] T. Huttunen, The ultraweak variational formulation for ultrasound transmission problems, Ph.D. thesis, University of Kuopio, February 2004.
[25] T. Huttunen, P. Monk, J.P. Kaipio, Computational aspects of the ultraweak variational formulation, Journal of Computational Physics 182 (2002) 27–46.
[26] K. Hynynen, O. Pomeroy, D.N. Smith, P.E. Huber, N.J. McDannold, J. Kettenbach, J. Baum, S. Singer, F.A. Jolesz, Mr imaging-guided focused ultrasound surgery of fibroadenomas in the breast: A feasibility study, Radiology 219 (1) (2001) 176–185.
[27] Y. Ishihara, A. Calderon, H. Watanabe, K. Okamoto, Y. Suzuki, K. Kuroda, Y. Suzuki, A precise and fast temperature mapping using water proton chemical shift, Magnetic Resonance in Medicine 34 (1995) 814–823.
[28] J.P. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Applied Mathematical Sciences, vol. 160, Springer-Verlag, New York, 2004.
[29] J.P. Kaipio, E. Somersalo, Statistical inverse problems: Discretization, model reduction, and inverse crimes, Journal of Computational and Applied Mathematics 198 (2007) 493–504.
[30] R.E. Kalman, A new approach to linear filtering and prediction problems, Journal of Basic Engineering Transactions of the ASME (Series D) 82 (1960) 35–45.
[31] J. Lang, B. Erdmann, M. Seebass, Impact of nonlinear heat transfer on temperature control in regional hyperthermia, IEEE Transaction on Biomedical Engineering 46 (9) (1999) 1129–1138.
[32] A. Lehikoinen, S. Finsterle, A. Voutilainen, L.M. Heikkinen, M. Vauhkonen, J.P. Kaipio, Approximation errors and truncation of computational domains with application to geophysical tomography, Inverse Problems and Imaging 1 (2) (2007) 371–389.
[33] K. Mahoney, T. Fjield, N. McDannold, G. Clement, K. Hynynen, Comparison of modeled and observed *in vivo* temperature elevations induced by focused ultrasound: Implications for treatment planning, Physics in Medicine and Biology 46 (2001) 1785–1798.
[34] M. Malinen, S.R. Duncan, T. Huttunen, J.P. Kaipio, Feedforward and feedback control of ultrasound surgery, Applied Numerical Mathematics 56 (2006) 55–79.
[35] M. Malinen, T. Huttunen, J.P. Kaipio, Thermal dose optimization method for ultrasound surgery, Physics in Medicine and Biology 48 (2003) 745–762.
[36] N. McDannold, K. Hynynen, D. Wolf, G. Wolf, F.A. Jolesz, MRI evaluation of thermal ablation of tumors with focused ultrasound, Journal of Magnetic Resonance Imaging 8 (1998) 91–100.
[37] D.L. Parker, V. Smith, P. Sheldon, L.E. Crooks, L. Fussell, Temperature distribution measurements in two-dimensional NMR imaging, Medical Physics 10 (1983) 321–325.
[38] H.H. Pennes, Analysis of tissue and arterial blood temperatures in the resting human forearm, Journal of Applied Physiology 1 (2) (1948) 93–122.
[39] A.D. Pierce, Acoustics: An Introduction to its Physical Principles and Applications, Acoustical Society of America, New York, 1991.
[40] H.-K. Pikkarainen, A mathematical model for electrical impedance process tomography, Doctoral dissertation, Helsinki University of Technology, 2005.
[41] H.-K. Pikkarainen, Stochastic approach to electric process tomography, in: M.J. Colaco, H.R.B. Orlande, G.S. Dulikravich (Eds.), Proceedings of the Inverse Problems, Design and Optimization Symposium, vol. II, E-papers Publishing House, Ltd., Rio de Janeiro, Brazil, 2005.

[42] A. Seppänen, M. Vauhkonen, E. Somersalo, J.P. Kaipio, State space models in process tomography – approximation of state noise covariance, Inverse Problems and Engineering 9 (2001) 561–585.
[43] A. Seppänen, M. Vauhkonen, P.J. Vauhkonen, E. Somersalo, J.P. Kaipio, State estimation with fluid dynamical evolution models in process tomography – an application with impedance tomography, Inverse Problems 17 (2001) 467–484.
[44] M.G. Skinner, M.N. Iizuka, M.C. Kolios, M.D. Sherar, A theoretical comparison of energy sources – microwave, ultrasound and laser – for interstitial thermal therapy, Physics in Medicine and Biology 43 (1998) 3535–3547.
[45] M.G. Skinner, M.N. Iizuka, M.C. Kolios, M.D. Sherar, A theoretical comparison of energy sources – microwave, ultrasound and laser – for interstitial thermal therapy, Physics in Medicine and Biology 43 (1998) 3535–3547.
[46] C.W. Song, A. Lokshina, J.G. Rhee, M. Pattenm, S.H. Levitt, Implications of blood blow in hyperthermic treatment of tumors, IEEE Transactions on Biomedical Engineering 31 (1) (1984) 9–16.
[47] C. Sumi, J. Kuwabara, Determination of thermal conductivity distribution from internal temperature distribution measurements, Review of Scientific Instruments 77 (2006) 064904.
[48] S.A. Telenkov, J. Youn, D.M. Goodman, A.J. Welch, T.E. Milner, Non-contact measurement of thermal diffusivity in tissue, Physics in Medicine and Biology 46 (2001) 551–558.
[49] A. Vanne, K. Hynynen, MRI feedback temperature control for focused ultrasound surgery, Physics in Medicine and Biology 48 (2003) 31–43.
[50] N. Vykhodtseva, V. Sorrentino, F.A. Jolesz, R.T. Bronson, K. Hynynen, MRI detection of the thermal effects of focused ultrasound in the brain, Ultrasound in Medicine and Biology 26 (5) (2000) 871–880.
[51] F.M. Waterman, L. Tupchong, R.E. Nerlinger, J. Matthews, Blood flow in human tumors during local hyperthermia, International Journal of Radiation Oncology*Biology*Physics 20 (6) (1991) 1255–1262.
[52] A.E. Worthington, J. Trachtenberg, M.D. Sherar, Ultrasound properties of human prostate tissue during heating, Ultrasound in Medicine & Biology 28 (10) (2002) 1311–1318.