Monograph:

Guo, L.Z., Billings, S.A. and Coca, D. (2008) Consistent parameter identification of partial differential equation models from noisy observations. Research Report. ACSE Research Report no. 978 . Automatic Control and Systems Engineering, University of Sheffield

Consistent parameter identification of partial differential equation models from noisy observations

Guo, L. Z., Billings, S. A., and Coca, D.

Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, S1 3JD
UK

Research Report No. 978
May 2008

Consistent parameter identification of partial differential equation models from noisy observations

Guo, L. Z., Billings, S. A., and Coca, D.

Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield S1 3JD, UK

Abstract

This paper introduces a new residual-based recursive parameter estimation algorithm for linear partial differential equations. The main idea is to replace unmeasurable noise variables by noise estimates and to compute recursively both the model parameter and noise estimates. It is proven that under some mild assumptions the estimated parameters converge to the true values with probability one. Numerical examples that demonstrate the effectiveness of the proposed approach are also provided.

1 Introduction

The identification problem for partial differential equation models which is often described under several different names including spatially extended systems, distributed parameter systems, and spatio-temporal dynamical systems, has been extensively studied for the past three decades. There are plenty of identification methods in the literature including statistical methods (Banks and Kunisch 1989, Fitzpatrick 1991), least squares methods (Yin and Fitzpatrick 1992, Coca and Billings 2000), finite dimensional approximation(Mao, Reich, Rosen 1994), singuler value decomposition (Gay and Ray 1995), neural networks (Gonzalez-Garcia, Rico-Martinez, and Kevrekidis 1998), orthogonal feedforward least squares method (Coca and Billings 2002, Guo and Billings 2006), maximal correlation method (Voss, Bunner, and Abel 1998), and some papers on practical issues (Point, Wouwer, and Remy 1996). One of the key issues in the identification of partial differential equations is the convergence and consistency of the estimator which has been studied by several authors (Banks and Kunisch 1989 and the references therein, Fitzpatrick 1991, Yin and Fitzpatrick 1992, Mao, Reich, Rosen 1994, Coca and Billings 2002). In this paper, a new residual based recursive identification algorithm for partial differential equation models is proposed, which can be implemented online easily. The basic idea behind the recursive algorithm is that the unmeasurable noises contained in the observations are replaced by their estimates or innovations at each time step in the algorithm. Unlike other methods, the proposed method considers the effects of both time and space sampling sizes on the convergence of the parameter estimates. Specifically, motivated by (Soderstrom, Fan, Carlsson, and Bigi 1997), the derivatives with respect to both time and space are approximated by a set of noisy samples rather than in a standard difference method such as Euler backword and forward schemes. It turns out that if the approximation parameters to the derivatives are selected with care under the constriants of natural conditions (Soderstrom, Fan, Carlsson,and Bigi 1997) and passive conditions, then the recursive algorithm will be consistently convergent under some mild conditions compared with the standard stationary and ergodic assumptions.

The paper is organised as follows. Section 2 defines the system identification problem which is studied in this paper. The residual based recursive algorithm is presented and consistency issues are discussed in section 3. Section 4 illustrates the proposed approach using some examples. Finally conclusions are drawn in section 5.

2 Problem statement

Consider the continuous spatio-temporal dynamical systems which can be described as the following linear partial differential equations

$$D^{(n,0)}u(t,x) + \sum_{i=0}^{n-1} \sum_{|j|\leq|m|} a_{i,j} D^{(i,j)}u(t,x) = f(t,x) \tag{1}$$

where $u(x,t) \in R^{n_u}$ is the dependent variable of the system, $t \in [0,\infty)$ denotes time and $x = (x_1, x_2, \cdots, x_{n_x}) \in \Omega \subset R^{n_x}$ denotes the spatial variable, $a_{i,j}, i = 0, 1, \cdots, n, |j| \leq |m|$ are the unknown constant parameters. $j = (j_1, j_2, \cdots, j_{n_x}) \in N^{n_x}$ is an n_x-dimensional multi-index with $|j| = \sum_{l=1}^{n_x} j_l$. $D^{(i,j)}u(t,x)$ is defined as $D_t^i D_x^j$ where $D_t^i = \partial^i/\partial t^i$ and $D_x^j = D_{x_1}^{j_1} D_{x_2}^{j_2} \cdots D_{x_{n_x}}^{j_{n_x}}$, $D_{x_l}^{j_l} = \partial^{j_l}/\partial x_l^{j_l}, l = 1, 2, \cdots, n_x,\ i = 0, 1, \cdots, n, |j| \leq |m|$. $f(t,x)$ is the external input to the system. The boundary and initial conditions are assumed to be

$$B(u(t,x)) = u_b(t,x), x \in \partial\Omega; T(u(0,x)) = u(x), x \in \Omega \tag{2}$$

where B is the differential operator which operates on the boundary $\partial\Omega$ of the spatial domain and T is a differential operator evaluated at time $t = 0$, providing the initial conditions of u and of time derivatives of u. For the sake of simplicity and without loss of generality, only the one dimensional case is considered in this paper, that is, $n_x = 1$ and $n_u = 1$.

For the purpose of identification, the system is observed in discrete-time at $t = 0, 1, 2, \cdots$ with a sampling interval h and in discrete-space at $x_k = x_{k-1} + H, k = 1, \cdots, K$ with a grid size H with the following observation function

$$y(t, x_k) = u(t, x_k) + e(t, x_k), k = 0, 1, \cdots, K \tag{3}$$

where the $e(t, x_k)$ is the measurement noise. This paper is concerned with the parameter identification problem from the sampled solution $y(t, x)$ and the external input $f(t, x)$, therefore here the solution of (1) is assumed to exist and to be unique, and the system is assumed to be input-output uniformly bounded.

The model (1) to (3) contributes a general representation for a large class of linear spatio-temporal systems such as wave, heat, and vibrating membrance systems. In order to identify a continuous model directly from nosiy data, the derivatives in (1) should be approximated with some difference operators. A general linear approximation of the differential operators by using the samples can be considered as

$$D^{(i,j)}u(t, x) \approx \hat{D}^{(i,j)}u(t, x) = \sum_{p,q} \beta_{i,j}(p, q)y(t + ph, x + qH) \tag{4}$$

where $\beta_{i,j}(p, q)$ are the weighting parameters. Assume that $u(t, x)$ is sufficiently differentiable so that a Taylor expansion of $u(t, x)$ yields

$$\begin{aligned}
\hat{D}^{(i,j)}u(t, x) &= \sum_{p,q} \beta_{i,j}(p, q)(u(t + ph, x + qH) + e(t + ph, x + qH)) \\
&= \sum_{p,q} \beta_{i,j}(p, q)u(t + ph, x + qH) + \sum_{p,q} \beta_{i,j}(p, q)e(t + ph, x + qH)) \\
&= \sum_{p,q} \beta_{i,j}(p, q)(\sum_{0 \le \alpha+\nu \le i+j} \frac{D^{(\alpha,\nu)}u(t, x)}{\alpha!\nu!}(ph)^{\alpha}(qH)^{\nu} + O(\bar{h})) \\
&\quad + \sum_{p,q} \beta_{i,j}(p, q)e(t + ph, x + qH))
\end{aligned} \tag{5}$$

with $\bar{h} = \max\{h, H\}$, then

$$\hat{D}^{(i,j)}u(t, x) = D^{(i,j)}u(t, x) + O(\bar{h}) + \sum_{p,q} \beta_{i,j}(p, q)e(t + ph, x + qH) \tag{6}$$

providing $\beta_{i,j}(p, q)$ satisfies the following conditions

$$\sum_{p,q} \beta_{i,j}(p, q)(ph)^{\alpha}(qH)^{\nu} = \begin{cases} 0, & 0 \le \alpha + \nu \le i + j \text{ but } \alpha \ne i, \nu \ne j \\ i!j!, & \alpha = i, \nu = j \end{cases} \tag{7}$$

which is referred to as the natural conditons by Soderstrom et al. (1997).

Remark 1 The minimal lengths of the indexes p and q in (β) are $i + 1$ and $j + 1$ respectively. For identification algorithms without iteration, the range of time index p can be set to be any integer whilst it must be negative for any iterative algorithm, that is only the time history data are available. In this paper, an iterative algorithm will be presented so that it is assumed that

p is negative. Generally, the range of the spatial index q will not be restricted. However it will be clear later on that the proposed algorithm in this paper requires the range of q to span to the boundary in order to estimate the noise around the spatial domain.

Remark 2 The introduction of the parameters $\beta_{i,j}(p,q)$ in (7) provides extra degrees of freedom which can be adjusted to produce desirable performance.

With these two remarks, for given sampling point (t, x_k) (4) is rewitten as

$$\begin{aligned} \hat{D}^{(n,0)}u(t,x_k) &= \beta_{n,0}(0,0)y(t,x_k) + \sum_{p=1}^{P}\sum_{q=-Q}^{Q}\beta_{n,0}(p,q)y(t-p,x_{k+q}) \\ \hat{D}^{(i,j)}u(t,x_k) &= \sum_{p=1}^{P}\sum_{q=-Q}^{Q}\beta_{i,j}(p,q)y(t-p,x_{k+q}) \text{ for } 0 \le i < n, 0 \le j \le m \end{aligned} \tag{8}$$

where $y(t-p, x_{k+q})$ denotes $y(t-ph, x_k+qH)$ for the sake of symbol simplicity. Note that the weights β have been chosen to be the same for all (t,x) in this paper. Actually they could be chosen differrently with different (t,x), which might be used to deal with uneven sampled data or multiscale data.

Combining (6), (8) and (1) yields

$$\begin{aligned} \hat{D}^{(n,0)}u(t,x_k) &= -\sum_{i=0}^{n-1}\sum_{j=0}^{m}a_{i,j}\hat{D}^{(i,j)}u(t,x_k) \\ &\quad +\beta_{n,0}(0,0)e(t,x_k) \\ &\quad +\sum_{p=1}^{P}b_{p,-Q}e(t-p,x_{k-Q}) + \cdots + \sum_{p=1}^{P}b_{p,Q}e(t-p,x_{k+Q}) \\ &\quad +f(t,x_k) + O(\bar{h}) \end{aligned} \tag{9}$$

where $b_{p,q} = \sum_{i=0}^{n}\sum_{j=0}^{m}a_{i,j}\beta_{i,j}(p,q)$.

Assume that $\beta_{n,0}(0,0) \neq 0$ and let $\varepsilon(t,x_k) = \beta_{n,0}(0,0)e(t,x_k)$, it then follows

$$\begin{aligned} \hat{D}^{(n,0)}u(t,x_k) &= -\sum_{i=0}^{n-1}\sum_{j=0}^{m}a_{i,j}\hat{D}^{(i,j)}u(t,x_k) \\ &\quad +\sum_{p=1}^{P}d_{p,-Q}\varepsilon(t-p,x_{k-Q}) + \cdots + \sum_{p=1}^{P}d_{p,Q}\varepsilon(t-p,x_{k+Q}) \\ &\quad +f(t,x_k) \\ &\quad +\varepsilon(t,x_k) \\ &\quad +O(\bar{h}) \end{aligned} \tag{10}$$

with $d_{p,q} = b_{p,q}/\beta_{n,0}(0,0)$. Then the following discrete-time and discrete-space linear regression

$$\mathbf{w}(t) = \Phi_0(t)\theta + \varepsilon(t) + O(\bar{h})\mathbf{1} \tag{11}$$

can then be constructed, where

$$\mathbf{w}(t) = \begin{bmatrix} w(t, x_0) \\ w(t, x_1) \\ \vdots \\ w(t, x_K) \end{bmatrix} \tag{12}$$

and

$$\Phi_0(t) = \begin{bmatrix} \phi_0^T(t, x_0) \\ \phi_0^T(t, x_1) \\ \vdots \\ \phi_0^T(t, x_K) \end{bmatrix} \tag{13}$$

and

$$\varepsilon(t) = \begin{bmatrix} \varepsilon(t, x_0) \\ \varepsilon(t, x_1) \\ \vdots \\ \varepsilon(t, x_K) \end{bmatrix} \tag{14}$$

where

$$\begin{aligned} w(t, x_k) &= \hat{D}^{(n,0)}u(t, x_k) \\ \phi_0^T(t, x_k) &= [-\hat{D}^{(0,0)}u(t,x) - \hat{D}^{(0,1)}u(t,x) \cdots - \hat{D}^{(0,m)}u(t,x); \\ &\quad -\hat{D}^{(1,0)}u(t,x) - \hat{D}^{(1,1)}u(t,x) \cdots - \hat{D}^{(1,m)}u(t,x); \cdots; \\ &\quad -\hat{D}^{(n-1,0)}u(t,x) - \hat{D}^{(n-1,1)}u(t,x) \cdots - \hat{D}^{(n-1,m)}u(t,x); \\ &\quad \varepsilon(t-1, x_{k-Q}), \varepsilon(t-2, x_{k-Q}), \cdots, \varepsilon(t-P, x_{k-Q}); \\ &\quad \vdots \\ &\quad \varepsilon(t-1, x_{k+Q}), \varepsilon(t-2, x_{k+Q}), \cdots, \varepsilon(t-P, x_{k+Q}); f(t,x)] \end{aligned} \tag{15}$$

and the parameters are

$$\begin{aligned} \theta^T &= [a_{0,0}, a_{0,1}, \cdots, a_{0,m}; a_{1,0}, a_{1,1}, \cdots, a_{1,m}; \cdots; \\ &\quad a_{n-1,0}, a_{n-1,1}, \cdots, a_{n-1,m}; d_{1,-Q}, \cdots, d_{P,-Q}; \cdots, d_{1,Q}, \cdots, d_{P,Q}; 1] \end{aligned} \tag{16}$$

Note that when x reaches or goes beyond the boundary of the spatial domain the corresonding samples will be replaced either by the boundary conditions or zeros. The objective of this paper is to develop an algorithm to estimate the parameter vector θ by using the noisy observations $y(t,x)$ and to show under what conditions the estimated parameters consistently converge to the true parameters when both time $t \to \infty$ and the spatial sampling size $\bar{h} \to 0$.

From (2) and (15) it can be observed that $w(t,x)$ is not only affected by the noise from the spatial location but also from other neighbouring spatial locations. These noise sequences are not measurable, instead, a recursive algorithm will be used and they will be estimated by the estimated residuals or the innovations.

3 The residual based recursive algorithm

Let $\hat{\theta}(t)$ be the estimate of θ at time instant t, the proposed residual based iterative algorithm can be stated as follows

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\Phi^T(t)(\mathbf{w}(t) - \Phi(t)\hat{\theta}(t-1)) \tag{17}$$

$$P^{-1}(t) = P^{-1}(t-1) + \Phi^T(t)\Phi(t), P(0) = p_0 I \tag{18}$$

$$\Phi(t) = \begin{bmatrix} \phi^T(t, x_0) \\ \phi^T(t, x_1) \\ \vdots \\ \phi^T(t, x_K) \end{bmatrix} \tag{19}$$

where

$$\begin{aligned}
\phi^T(t, x_k) = \; & [-\hat{D}^{(0,0)}u(t,x_k) - \hat{D}^{(0,1)}u(t,x_k) \cdots - \hat{D}^{(0,m)}u(t,x_k); \\
& -\hat{D}^{(1,0)}u(t,x_k) - \hat{D}^{(1,1)}u(t,x_k) \cdots - \hat{D}^{(1,m)}u(t,x_k); \cdots; \\
& -\hat{D}^{(n-1,0)}u(t,x_k) - \hat{D}^{(n-1,1)}u(t,x_k) \cdots - \hat{D}^{(n-1,m)}u(t,x_k); \\
& \hat{\varepsilon}(t-1, x_{k-Q}), \cdots, \hat{\varepsilon}(t-P, x_{k-Q}); \\
& \vdots \\
& \hat{\varepsilon}(t-1, x_{k+Q}), \cdots, \hat{\varepsilon}(t-P, x_{k+Q}); \\
& f(t, x_k)]
\end{aligned} \tag{20}$$

$$\hat{\varepsilon}(t) = \mathbf{w}(t) - \Phi(t)\hat{\theta}(t). \tag{21}$$

To initialise the algorithm, $\theta(0)$ will be set to be a small real vector, $\hat{\varepsilon}(\tau, x) = 0$ for $\tau \leq 0$.

To show the consistency of the algorithm, some assumptions have to be made.

- (A1) It is assumed that $\{\varepsilon(t), \mathscr{F}_t\}$ is a martingale difference sequence defined on a probability space with $\mathscr{F}_t$ is the σ algebra sequence generated by $\varepsilon(t)$ up to t.

- (A2) $E[\varepsilon(t)|\mathscr{F}_{t-1}] = 0, a.s.$
- (A3) $E[\varepsilon^T(t)\varepsilon(t-1)|\mathscr{F}_t] = \sigma^2(t) \leq \bar{\sigma}^2 < \infty, a.s.$
- (A4) $\eta_1 I \leq \frac{1}{t}\int_0^t \Phi_0(\tau)\Phi_0^T(\tau)d\tau \leq \eta_2 I, a.s.$.

The assumptions (A1) to (A3) indicate that the noise $\varepsilon(t,x)$ is zero-mean with bounded time and spatial-varying variances which shows the system concerned may not be stationary. Assumption (A4) is the persistent excitation condition. It is worth noting that with (A4) the trace $tr(P^{-1}(t))$ of $P^{-1}(t)$ is a strictly increasing function of t.

Now define the parameter estimation error $\tilde{\theta}(t)$ and the innovation $\tilde{\varepsilon}(t)$ as

$$\begin{aligned} \tilde{\theta}(t) &= \hat{\theta}(t) - \theta \\ \tilde{\varepsilon}(t) &= \mathbf{w}(t) - \Phi(t)\hat{\theta}(t-1) \end{aligned} \tag{22}$$

It follows that

$$\hat{\varepsilon}(t) = (I - \Phi(t)P(t)\Phi^T(t))\tilde{\varepsilon}(t) = (I + \Phi(t)P(t-1)\Phi^T(t))^{-1}\tilde{\varepsilon}(t) \tag{23}$$

Define

$$\begin{aligned} \tilde{\mathbf{w}}(t) &=: -\Phi(t)\tilde{\theta}(t) \\ &= -\Phi(t)(\hat{\theta}(t) - \theta) \\ &= \hat{\varepsilon}(t) - \mathbf{w}(t) + \Phi(t)\theta \\ &= D(z)(\hat{\varepsilon}(t) - \varepsilon(t)) + O(\bar{h})\mathbf{1} \end{aligned} \tag{24}$$

where

$$D(z) = \begin{bmatrix} 1 + D_0(z) & D_1(z) & \cdots & D_Q(z) & 0 & \cdots & 0 \\ D_0(z) & 1 + D_1(z) & \cdots & D_Q(z) & D_{Q+1}(z) & \cdots & 0 \\ & & \vdots & & & & \\ 0 & \cdots & D_{-Q+K}(z) & \cdots & \cdots & \cdots & 1 + D_K(z) \end{bmatrix} \tag{25}$$

with

$$D_l(z) = d_{1,l}z^{-1} + \cdots + d_{P,l}z^{-P}, l = -Q, \cdots, Q. \tag{26}$$

It is interesting to see that $\tilde{\mathbf{w}}(t)$ may be considered as the output of the multi-input multi-output linear systems $D(z)$ driven by the signals $(\hat{\varepsilon}(t) - \varepsilon(t))$, which are from different spatial locations. Consider the coefficients of $D_l(z)$ which are composed of a and β, it is assumed that the underlying system is such that the $\beta_{i,j}(p,q)$ can be chosen such that the above linear system (24) is passive in the sense given in the following assumption. Define

$$S(t) =: \sum_{i=1}^{t} 2(-\tilde{\mathbf{w}}^T(i)(\frac{1}{2}\tilde{\mathbf{w}}(i) + (\hat{\varepsilon}(i) - \varepsilon(i))) \tag{27}$$

then from (24) it is easy to see that for $\rho > 0$

$$S(t) = 2\sum_{i=1}^{t} \tilde{\mathbf{w}}^T(i)(D^{-1}(z) - \frac{1+\rho}{2}\mathbf{I})\tilde{\mathbf{w}}(i) + \rho\sum_{i=1}^{t} \tilde{\mathbf{w}}^T(i)\tilde{\mathbf{w}}(i) + O(\bar{h}). \tag{28}$$

Assumption (A5) The linear system (24) is finite gain stable and there exists a positive constant ρ such that

$$2\sum_{i=1}^{t} \tilde{\mathbf{w}}^T(i)(D^{-1}(z) - \frac{1+\rho}{2}\mathbf{I})\tilde{\mathbf{w}}(i) \geq 0, a.s. \tag{29}$$
$$S(t) \geq 0, a.s.$$

Now, the main result in this paper can be stated as the following theorem.

Theorem 1 For the given system (2) and the algorithm in (17) to (21), if the Assumptions (A1) to (A5) hold, then for any $\delta > 1$, we have

$$\|\hat{\theta}(t) - \theta\|^2 = O(\frac{(\ln tr(P_0^{-1}(t)))^\delta}{\lambda_{min}(P_0^{-1}(t))}) + O(\bar{h}), a.s. \tag{30}$$

where λ_{min} is the minimal eigenvalue of the matrix, and

$$\lim_{t\to\infty} \|\hat{\theta}(t) - \theta\|^2 = O(\bar{h}), a.s. \tag{31}$$

Proof. Let $V(t) = \tilde{\theta}^T(t)P^{-1}(t)\tilde{\theta}(t)$. From (18), it follows that

$$V(t) = \tilde{\theta}^T(t)P^{-1}(t-1)\tilde{\theta}(t) + \tilde{\theta}^T(t)\Phi^T(t)\Phi(t)\tilde{\theta}(t) \tag{32}$$

Then from (22), (23) and the matrix inversion lemma, it has

$$\begin{aligned}
V(t) &= \tilde{\theta}^T(t)P^{-1}(t-1)\tilde{\theta}(t) + \tilde{\theta}^T(t)\Phi^T(t)\Phi(t)\tilde{\theta}(t) \qquad (33)\\
&= V(t-1) + \tilde{\theta}^T(t-1)\Phi^T(t)\hat{\varepsilon}(t) + \tilde{\theta}^T(t)\Phi^T(t)\hat{\varepsilon}(t) + \tilde{\theta}^T(t)\Phi^T(t)\Phi(t)\tilde{\theta}(t)\\
&= V(t-1) + 2\tilde{\theta}^T(t)\Phi^T(t)\hat{\varepsilon}(t) - \tilde{\varepsilon}^T(t)\Phi(t)P(t)\Phi^T(t)\hat{\varepsilon}(t) + \tilde{\theta}^T(t)\Phi^T(t)\Phi(t)\tilde{\theta}(t)\\
&= V(t-1) + 2\tilde{\theta}^T(t)\Phi^T(t)\hat{\varepsilon}(t) - \tilde{\varepsilon}^T(t)\Phi(t)P(t)\Phi^T(t)(I - \Phi(t)P(t-1)\Phi^T(t))^{-1}\tilde{\varepsilon}(t)\\
&\quad + \tilde{\theta}^T(t)\Phi^T(t)\Phi(t)\tilde{\theta}(t)\\
&\leq V(t-1) + 2\tilde{\theta}^T(t)\Phi^T(t)\hat{\varepsilon}(t) + \tilde{\theta}^T(t)\Phi^T(t)\Phi(t)\tilde{\theta}(t)
\end{aligned}$$

$$
\begin{aligned}
&= V(t-1) - 2(\tilde{\mathbf{w}}^T(t))(-\frac{1}{2}\tilde{\mathbf{w}}(t) + (\hat{\varepsilon}(t) - \varepsilon(t))) + 2\tilde{\theta}^T(t-1)\Phi^T(t)\varepsilon(t) \\
&\quad +2(\varepsilon^T(t) - \hat{\varepsilon}^T(t))\Phi(t)P(t)\Phi^T(t)\varepsilon(t) + 2\varepsilon^T(t)\Phi(t)P(t)\Phi^T(t)\varepsilon(t)
\end{aligned}
$$

Since $S(t-1)$, $\tilde{\theta}^T(t-1)\Phi^T(t)$, and $\varepsilon(t)-\hat{\varepsilon}(t)$ are uncorrelated with $\varepsilon(t)$ and are $\mathscr{F}_{t-1}$ measurable, taking the conditional expectation on both side of $(V(t)+S(t))/(\ln tr(P^{-1}(t)))^\delta$ with respect to $\mathscr{F}_{t-1}$ and using (A1) to (A2) yields

$$
E[\frac{V(t)+S(t)}{(\ln tr(P^{-1}(t)))^\delta}|\mathscr{F}_{t-1}] \leq \frac{V(t-1)+S(t-1)}{(\ln tr(P^{-1}(t)))^\delta} + \frac{2\varepsilon^T(t)\Phi(t)P(t)\Phi^T(t)\varepsilon(t)}{(\ln tr(P^{-1}(t)))^\delta} \tag{34}
$$

To apply the martigale convergence theorem (Goodwin and Sin 1984) to the above equation, we need to show the sum of the last term on the right-hand side for t from 1 to ∞ is finite. In fact, noting that $P^{-1}(t)$ is a strictly increasing function of t,

$$
\begin{aligned}
\sum_{t=1}^{\infty}\frac{\varepsilon^T(t)\Phi(t)P(t)\Phi^T(t)\varepsilon(t)}{(\ln tr(P^{-1}(t)))^\delta} &\leq \sum_{t=1}^{\infty}\frac{\varepsilon^T(t)tr(\Phi(t)P(t)\Phi^T(t))\varepsilon(t)}{(\ln tr(P^{-1}(t)))^\delta} \qquad (35)\\
&\leq \bar{\sigma}^2\sum_{t=1}^{\infty}\frac{tr(\Phi(t)P(t)\Phi^T(t))}{(\ln tr(P^{-1}(t)))^\delta} \\
&\leq \bar{\sigma}^2\sum_{t=1}^{\infty}\frac{(K+1)\cdot tr(P^{-1}(t)) - tr(P^{-1}(t-1))}{tr(P^{-1}(t))((\ln tr(P^{-1}(t)))^\delta} \\
&= \bar{\sigma}^2\sum_{t=1}^{\infty}\frac{tr(P^{-1}(t)) - tr(P^{-1}(t-1))}{tr(P^{-1}(t))((\ln tr(P^{-1}(t)))^\delta} + K\bar{\sigma}^2\sum_{t=1}^{\infty}\frac{1}{(\ln tr(P^{-1}(t)))^\delta} \\
&= \bar{\sigma}^2\sum_{t=1}^{\infty}\int_{tr(P^{-1}(t-1))}^{tr(P^{-1}(t))}\frac{dx}{tr(P^{-1}(t))(\ln tr(P^{-1}(t)))^\delta} + K\bar{\sigma}^2\sum_{t=1}^{\infty}\frac{1}{(\ln tr(P^{-1}(t)))^\delta} \\
&\leq \bar{\sigma}^2\sum_{t=1}^{\infty}\int_{tr(P^{-1}(t-1))}^{tr(P^{-1}(t))}\frac{dx}{x(\ln x)^\beta} + K\bar{\sigma}^2\sum_{t=1}^{\infty}\frac{1}{(\ln tr(P^{-1}(t)))^\delta} \\
&= \bar{\sigma}^2\int_{tr(P^{-1}(0))}^{tr(P^{-1}(\infty))}\frac{dx}{x(\ln x)^\beta} + K\bar{\sigma}^2\sum_{t=1}^{\infty}\frac{1}{(\ln tr(P^{-1}(t)))^\delta} \\
&= \frac{\bar{\sigma}^2}{\delta-1}(\frac{1}{(\ln tr(P^{-1}(0)))^{\delta-1}} - \frac{1}{(\ln tr(P^{-1}(\infty)))^{\delta-1}}) \\
&\quad +K\bar{\sigma}^2\sum_{t=1}^{\infty}\frac{1}{(\ln tr(P^{-1}(t)))^\delta} \\
&< \infty
\end{aligned}
$$

It follows that $(V(t)+S(t))/(\ln tr(P^{-1}(t)))^\delta$ converges a.s. to a finite random variable, which means

$$
\begin{aligned}
V(t) &= O((\ln tr(P^{-1}(t)))^\delta), a.s., \qquad (36)\\
S(t) &= O((\ln tr(P^{-1}(t)))^\delta), a.s.
\end{aligned}
$$

and

$$\|\tilde{\theta}(t)\|^2 = \|\hat{\theta}(t) - \theta\|^2 = O(\frac{(\ln tr(P^{-1}(t)))^{\delta}}{\lambda_{min}(P^{-1}(t))}), a.s. \tag{37}$$

According to assumption (A5) and (24), one has

$$\sum_{i=1}^{t} \|\tilde{\mathbf{w}}(i)\|^2 = \sum_{i=1}^{t} \tilde{\mathbf{w}}^T(i)\tilde{\mathbf{w}}(i) \leq O((\ln tr(P^{-1}(t)))^{\delta}) + O(\bar{h}), a.s. \tag{38}$$

and there exist constants $k_1 > 0, k_2 > 0$ such that

$$\sum_{i=1}^{t} \|\hat{\varepsilon}(t) - \varepsilon(t)\|^2 \leq k_1 \sum_{i=1}^{t} \|\tilde{\mathbf{w}}(i)\|^2 + k_2 \leq O((\ln tr(P^{-1}(t)))^{\delta}) + O(\bar{h}), a.s. \tag{39}$$

Let $\tilde{\Phi}(t) =: \Phi(t) - \Phi_0(t)$, then it follows that

$$\begin{aligned} tr(\tilde{\Phi}(t)\tilde{\Phi}^T(t)) &= \sum_{k=0}^{K} \|\phi(t, x_k) - \phi_0(t, x_k)\|^2 \\ &= \sum_{k=0}^{K}\sum_{p=1}^{P}(\hat{\varepsilon}(t-p, x_k) - \varepsilon(t-p, x_k))^2 \\ &= O(\sum_{i=1}^{t} \|\hat{\varepsilon}(t) - \varepsilon(t)\|^2) \\ &\leq O((\ln tr(P^{-1}(t)))^{\delta}) + O(\bar{h}), a.s. \end{aligned} \tag{40}$$

so that

$$\begin{aligned} tr(\Phi^T(t)\Phi(t)) &\leq tr(\Phi_0^T(t)\Phi_0(t)) + tr(\tilde{\Phi}(t)\tilde{\Phi}^T(t)) \\ &\leq tr(\Phi_0^T(t)\Phi_0(t)) + O((\ln tr(P^{-1}(t)))^{\delta}) + O(\bar{h}), a.s. \end{aligned} \tag{41}$$

which yields

$$tr(P^{-1}(t)) = O(tr(P_0^{-1}(t))) + O(\bar{h}), a.s. \tag{42}$$

where $P_0^{-1}(t) = P_0^{-1}(t-1) + \Phi_0^T(t)\Phi_0(t)$. Similarly, it is easy to see

$$\lambda_{min}(P^{-1}(t)) = O(\lambda_{min}(P_0^{-1}(t)) + O(\bar{h}), a.s. \tag{43}$$

Combining (37) with (42) and (43) yields

$$\|\hat{\theta}(t) - \theta\|^2 = O(\frac{(\ln tr(P_0^{-1}(t)))^{\delta}}{\lambda_{min}(P_0^{-1}(t))}) + O(\bar{h}), a.s. \tag{44}$$

Then by the persistent excitation assumption (A4), one can conclude that

$$\lim_{t \to \infty} \|\hat{\theta}(t) - \theta\|^2 = O(\bar{h}), a.s. \tag{45}$$

Q.E.D

Remark 3 The passive conditions for the parameters β depends on the unknown model parameters which is difficult to validate. However such assumptions are often found in convengence and consistency analysis in parameter identification problems for conventional dynamic systems. How to determine these conditions is a problem under study. A possible way to deal with this problem is to include a method to estimate the β's following the estimated model parameters at each time step.

4 Numerical example - Shell-and-tube heat exchanger process

A fluid with constant density ρ and heat capacity C_p flows through the tube of a shell-and-tube heat exchanger with velocity v, as shown in Fig. (1). The fluid enters the tube at temperature u_0 and is heated from the shell side by condensing steam at temperature f. Assume that the tube has a uniform cross-section area S, length L and volumn $V = SL$. The surface area which are available for heat exchange is S_w, with a heat transfer coefficient δ. According to Ogunnaike and Ray(Ogunnaike and Ray 1994), the process composing of a convection term and a heat-exchange term can be model by the following PDE

$$\frac{\partial u(t,x)}{\partial t} = a_1 \frac{\partial u(t,x)}{\partial x} + a_2 u(t,x) + a_3 f(t,x) \tag{46}$$

where $u(t,x)$ denotes the fluid temperture at time t and position x, $a_1 = -v$, and $a_2 = -\delta S_w / \rho V C_p$. The boundary condition is specified at $x = 0$ since the inlet conditions can be assumed known

$$u(t,0) = u_0(t) \tag{47}$$

and the initial condition is some given initial temperature profile as

$$u(0,x) = u_i(x). \tag{48}$$

For the purpose of this numerical study, the values for the process parameters were chosen as $L = 1$, $a_1 = -1\mathrm{ms}^{-1}$, $a_2 = -2.92\mathrm{s}^{-1}$, and $a_3 = 2.92\mathrm{s}^{-1}$. The initial and boundary conditions were set to be $u_0(t) = 25°C$ and $u_i(x) = 25°C$. For the sake of simplicity, the control input $f(t,x) = f(t)$ was assumed to be independent of the spatial variable x and taken as the output of a second-order process

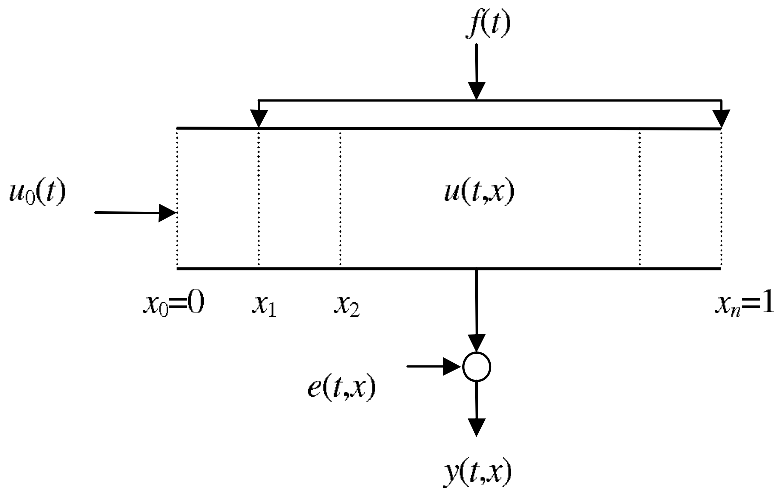

Figure 1: The shell-tube heat exchanger

$$p^2 f(t) + 3pf(t) + 3f(t) = 20v(t) \tag{49}$$

where $v(t)$ is a continuous-time white process with zero mean and unit variance. The above heat equation (46) was numerically solved with the above settings by a fourth-order Runge-Kutta method. For the purpose of simulation, the solution was sampled with a time-interval of $h = 0.01$ and a spatial-interval of $H = 0.01$. To apply the proposed recursive algorithm, those β's in the approximation to the derivatives $\partial u(t,x)/\partial t$ and $\partial u(t,x)/\partial x$ need to be determined according to the natural conditions (7) and passive assumptions (29). Let $y(t,x) = u(t,x) + e(t,x)$ be the measurements at the position x and time instant t and the approximations to the derivatives in this example be

$$\begin{aligned}
\frac{\partial u(t,x)}{\partial t} &= \beta_{1,0}(0,0)y(t,x) + \beta_{1,0}(1,0)y(t-1,x) + \beta_{1,0}(2,0)y(t-2,x) + \cdots + \beta_{1,0}(P_t,0)y(t-P_t,x) \qquad (50)\\
\frac{\partial u(t,x)}{\partial u} &= \beta_{0,1}(1,0)y(t-1,x) + \beta_{0,1}(2,0)y(t-2,x) + \cdots + \beta_{0,1}(P_x,0)y(t-P_x,x)\\
&+ \beta_{0,1}(1,1)y(t-1,x-1) + \beta_{0,1}(2,1)y(t-2,x-1) + \cdots + \beta_{0,1}(P_x,1)y(t-P_x,x-1)\\
&\;\;\vdots\\
&+ \beta_{0,1}(1,Q_x)y(t-1,x-Q_x) + \beta_{0,1}(2,Q_x)y(t-2,x-Q_x) + \cdots + \beta_{0,1}(P_x,Q_x)y(t-P_x,x-Q_x)
\end{aligned}$$

then the natural conditions are

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & P_t \end{bmatrix} \begin{bmatrix} \beta_{1,0}(0,0) \\ \beta_{1,0}(1,0) \\ \vdots \\ \beta_{1,0}(P_t,0) \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{h} \end{bmatrix} \tag{51}$$

for $\partial u(t,x)/\partial t$, and

	Data Size	$a_1 = -1$	$a_2 = -2.92$	$a_3 = 2.92$
	100	-3.35681	-8.07232	6.83338
	200	-2.35638	-6.66493	4.25968
SNR = 62	300	-1.71453	-5.22033	4.63909
	500	-1.39853	-3.96606	3.66518
	1000	-1.10333	-3.22524	2.90212
	100	-2.50930	-7.23002	5.00316
	200	-1.79061	-5.97266	3.97501
SNR = 48	300	-1.23798	-4.73361	4.26565
	500	-1.00949	-3.39527	3.18026
	1000	-0.97321	-2.56083	2.34338

Table 1: The parameter estimates using $\beta_{1,0}(0,0) = -\beta_{1,0}(1,0) = 1/h$ and $\beta_{0,1}(1,0) = -\beta_{0,1}(1,1) = 1/H$

$$\begin{bmatrix} 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & \cdots & 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & P_x & 1 & 2 & \cdots & P_x & \cdots & 1 & 2 & \cdots & P_x \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & \cdots & Q_x & Q_x & \cdots & Q_x \end{bmatrix} \begin{bmatrix} \beta_{0,1}(1,0) \\ \beta_{0,1}(2,0) \\ \vdots \\ \beta_{0,1}(P_x,0) \\ \beta_{0,1}(1,1) \\ \beta_{0,1}(2,1) \\ \vdots \\ \beta_{0,1}(P_x,1) \\ \vdots \\ \beta_{0,1}(1,Q_x) \\ \beta_{0,1}(2,Q_x) \\ \vdots \\ \beta_{0,1}(P_x,Q_x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{H} \end{bmatrix} \tag{52}$$

for $\partial u(t,x)/\partial u$. It is clear that for $P_t = 1$ and $P_x = 1, Q_x = 1$ this becomes the Euler method, that is $\beta_{1,0}(0,0) = -\beta_{1,0}(1,0) = 1/h$ and $\beta_{0,1}(1,0) = -\beta_{0,1}(1,1) = 1/H$ and for $P_t > 1$ or $P_x > 1, Q_x > 1$, these parameters can be determined according to the passive conditions (29) subject to the above natural conditions. In this paper, the simulation were conducted with two sets of β's: (1) $\beta_{1,0}(0,0) = -\beta_{1,0}(1,0) = 1/h$ and $\beta_{0,1}(1,1) = -\beta_{0,1}(1,2) = 1/H$ and (2) $\beta_{1,0}(0,0) = -\beta_{1,0}(1,0) = 1/h$ and $\beta_{0,1}(1,0) = \frac{3}{2H}, \beta_{0,1}(2,0) = -\frac{1}{2H}, \beta_{0,1}(1,1) = -\frac{3}{2H}, \beta_{0,1}(2,1) = \frac{1}{2H}$. The simulation results with different signal-to-noise ratios (SNR) and different data sizes are listed in Tables (1) and (2).

From Tables (1) and (2) it can be observed that the errors of parameter esimates generally decrease as the data size increases, and the proposed method can work well with different levels of measurement noises.

	Data Size	$a_1 = -1$	$a_2 = -2.92$	$a_3 = 2.92$
	100	-2.72233	-7.31312	5.52180
	200	-1.98843	-6.08481	4.04081
SNR = 62	300	-1.42064	-4.78006	4.29439
	500	-1.17389	-3.56129	3.31975
	1000	-0.94437	-2.86561	2.60291
	100	-1.77896	-5.72410	3.54736
	200	-1.23817	-4.60814	3.18878
SNR = 48	300	-0.83351	-3.67647	3.35002
	500	-0.64943	-2.49243	2.36227
	1000	-0.96299	-2.72609	2.60826

Table 2: The parameter estimates using $\beta_{1,0}(0,0) = -\beta_{1,0}(1,0) = 1/h$ and $\beta_{0,1}(1,0) = \frac{3}{2H}, \beta_{0,1}(2,0) = -\frac{1}{2H}, \beta_{0,1}(1,1) = -\frac{3}{2H}, \beta_{0,1}(2,1) = \frac{1}{2H}$

5 Conclusions

A recursive algorithm has been presented for the identification problem of continuous linear partial differential equation models. The analysis given in the paper has shown that the proposed method can produce a consistent parameter estimation. Because the selection of the design parameter β's is not very easy further studies are needed to develop new method to determine these parameters. Also it will be interesting to extend the proposed method to deal with the identification problem of nonlinear PDEs.

6 Acknowledgement

The authors gratefully acknowledge financial support from EPSRC (UK).

References

[1] Banks, H. T. and Kunisch, K., (1989) *Estimation Techniques for Distributed Parameter Systems*, Boston, Massachusetts: Birkhauser.

[2] Coca, D. and Billings, S. A., (2000) Direct parameter identification of distributed parameter systems, *Int. J. Systems Sci.*, Vol. 31, No.1, pp. 11-17.

[3] Coca, D. and Billings, S. A., (2002) Identification of finite dimensional models of infinite dimensional dynamical systems, *Automatica*, Vol. 38, pp. 1851-1865.

[4] Fitzpatrick, B. G., (1991) Bayesian analysis in inverse problems, *Inverse Problems*, Vol. 7, pp. 675-702.

[5] Gay, D. H. and Ray, W. H., (1995) Identification and control of distributed parameter systems by means of the singular value decomposition, *Chemical Engineering Science*, Vol. 50, No. 10, pp. 1519-1539.

[6] Gonzalez-Garcia, R., Rico-Martinez, R., and Kevrekidis, I. G., (1998) Identification of distributed parameter systems: a neural net approach, *Computer Chem. Eng.*, Vol. 22, Suppl. pp. S965-S968.

[7] Goodwin, G. C. and Sin, K. S., (1984), *Adaptive Filtering Prediction and Control*, New Jersey: Prentice-Hall, Inc.

[8] Guo, L. Z. and Billings, S. A., (2006) Identification of partial differential equation models for continuous spatio-temporal dynamical systems, *IEEE Trans. Circu. & Syst. II*, Vol. 53, No. 8, pp. 657-661.

[9] Mao, C., Reich, S., and Rosen, I. G., (1994) Approximation in the identification of nonlinear degenerate distributed parameter systems, *Nonlinear analysis, theory, methods, and applications*, Vol. 22, No. 1, pp. 91-120.

[10] Ogunnaike, B. A. and Ray W. H., (1994) *Process Dynamics, Modeling and Control*, New York: Oxford University Press.

[11] Point, N., Wouwer, A. V., and Remy, M., (1996) Practical issues in distributed parameter estimation: gradient computation and optimal experiment design, *Control Eng. Practice*, Vol.4, No. 11, pp. 1553-1562.

[12] Soderstrom, T., Fan, H., Carlsson, B., and Bigi, S., (1997) Least squares parameter estimation of continuous-time ARX models from discrete-time data, *IEEE Trans. Auto. Contr.*, Vol. 42, No. 5, pp. 659-673.

[13] Voss, H., Bunner, M. J., and Abel, M., (1998) Identification of continuous, spatiotemporal systems, *Physical Review E*, Vol. 57, No.3, pp. 2820-2823.

[14] Yin, G. and Fitzpatrick, B. G., (1992) Almost sure convegence in distributed parameter identification algorithms under correlated noise, *Appl. Math. Lett.*, Vol. 5, No. 4, pp. 41-44.