Adaptation in Stochastic Dynamic Systems—Survey and New Results IV: Seeking Minimum of API in Parameters of Data

Innokentiy V. Semushin, Julia V. Tsyganova
School of Mathematics and Information Technology, Ulyanovsk State University, Ulyanovsk, Russia
Email: kentvsem@gmail.com, tyanovaj@gmail.com

Received October 9, 2013; revised November 9, 2013; accepted November 16, 2013

Copyright © 2013 Innokentiy V. Semushin, Julia V. Tsyganova. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

This paper investigates the problem of seeking minimum of API (Auxiliary Performance Index) in parameters of Data Model instead of parameters of Adaptive Filter in order to avoid the phenomenon of over parameterization. This problem was stated by Semushin in [2]. The solution to the problem can be considered as the development of API approach to parameter identification in stochastic dynamic systems.

Keywords: Linear Stochastic Systems; Parameter Estimation; Model Identification; Identification for Control; Adaptive Control MSC (2010); 93E10; 93E12; 93E35

1. Introduction

The recent papers [1,2] gave a survey of the field of adaptation in stochastic systems as it has developed over the last four decades. The author’s research in this field was summarized and a novel solution for fitting an adaptive model in state space (instead of response space) was given.

In this paper, we further develop the Active Principle of Adaptation for linear time-invariant state-space stochastic MIMO filter systems included into the feedback or considered independently.

We solve the problem of seeking minimum of Auxiliary Performance Index (API) in parameters of Data Source Model (DSM) instead of parameters of Adaptive Filter (AF) in order to avoid difficulties known as Phenomenon of Over Parameterization (PhOP). The PhOP means that the number of parameters to be adjusted in AF is usually much greater than that in DSM. The solution of this problem will enable identification in the space of lower dimension and at the same time provide estimates of the given system state vector according to Original Performance Index (OPI). We verify the obtained theoretical results by two numerical simulation examples.

2. Parameterized Data Models \( \mathcal{D}(\theta) \)

Following the previous results of [1,2], we assume that all data models \( \mathcal{D}(\theta) \) forming a set \( \mathcal{D} \) are parameterized by an \( l \)-component vector \( \theta \). Each particular value of \( \theta \) (which does not depend on time) specifies a \( \mathcal{D}(\theta) \). Hence

\[
\mathcal{D} = \{ \mathcal{D}(\theta) | \theta \in \Theta \subset \mathbb{R}^l \}
\]

where \( \Theta \) is the compact subset of \( \mathbb{R}^l \). A given physical data model (PhDM) is described by the following equations:

\[
\begin{align*}
\mathcal{D}(\theta) : & \\
\nu_{t+1} = & \Phi(\theta)x_t + \Psi(\theta)u_t + \Gamma(\theta)w_t, t \in \mathbb{Z}_+ \\
y_t = & H(\theta)x_t + v_t, t \in \mathbb{Z}_i
\end{align*}
\]

where \( \mathbb{Z}_i \) denotes nonnegative integers, \( \mathbb{Z}_i \) strictly positive integers, and \( \mathbb{Z} \) all integers.

Every model \( \mathcal{D}(\theta) \) (2) is assumed to be acting between adjacent switches as long as it is sufficient for accepting as correct the basic theoretical statement (BTS) that all processes related to the \( \mathcal{D}(\theta) \) are wide-sense stationary. This statement amounts to the following assumptions. The random \( x_0 \) with \( E[\|x_0\|^2] < \infty \) is orthogonal [3] to \( w_t \) and \( v_t \), the zero-mean mutually orthogonal wide-sense stationary orthogonal sequences with \( E\{w_t w_t^T\} = Q(\theta) \geq 0 \) and \( E\{v_t v_t^T\} = R(\theta) > 0 \) for all \( t \in \mathbb{Z} \); \( w_t v_t^T \) is orthogonal to \( x_j \) and \( u_j \) for...
all \( f \geq t \); \( u_c \) is a given signal; it is an “external input” when considering the open-loop case or a control strategy function

\[
u_c = u_c \left( t, y', u^{-1} \right)
\]

when considering the closed-loop setup.

Stackable vectors of previous values

\[
X^t = (y^t, u^{-1} ) \quad \text{and} \quad u^{-1} = (u_{-1}, u_{-2}, \ldots, u_0)
\]

constitute the experimental condition \( X \) (refer to Ljung [4]).

By assumption, \( y^t \in \mathbb{R}^n \) is generated by the completely observable PhDM (2), so we can move from the physical state variables \( x \in \mathbb{R}^n \) in (2) to another \( x' \) through the similarity transformation \( x' = W_x x \). Such transformation uniquely determines a new state representation

\[
D'(\theta): \ x'^{t+1} = \Phi_x(\theta) x'^t + \Psi(\theta) u + \Gamma(\theta) w, t \in \mathbb{Z}_t
\]

of the standard observable data model (SODM) (cf. Semushin [2]).

For convenience in the below we shall omit the subscript \( \theta \) for all the matrices describing PhDM or SODM.

### 3. Parameterized Innovations

As before, the above data model of a time-invariant data source will be referred to as the conventional model, no matter whether it is PhDM (2) or SODM (5). Here we use another innovation model, that differs from the time-invariant (due to BTS) innovation model, presented in [2]:

\[
\mathcal{M}(\theta):\ x_{t+1} = \Phi x_{t} + \Psi u + G v, t \in \mathbb{Z}_t
\]

with \( t \in \mathbb{Z}_t \), the initial \( x_0 = x_0 \), and the \( E \) component vector \( \theta \) of unknown system parameters, which needs to be identified. This means that the entries of the matrices \( \Phi(\theta), \Psi(\theta), \Gamma(\theta), Q(\theta), H(\theta), R(\theta) \) are functions of \( \theta \in D(\theta) \). However, for the sake of simplicity we will suppress the corresponding notations below, i.e. instead of \( \Phi(\theta), \Psi(\theta), \Gamma(\theta), Q(\theta), H(\theta), R(\theta) \) we will write \( \Phi, \Psi, \Gamma, Q, H, R \). We make the same assumptions about the SODM.

### 4. Uncertainty Parameterization

In contrast to our previous work [2], we do not consider the four levels of uncertainty. Assume that system (2) (and also the SODM (5)) is parameterized by an \( l \)-component vector \( \theta \) of unknown system parameters, which needs to be identified. This means that the entries of the matrices \( \Phi(\theta), \Psi(\theta), \Gamma(\theta), Q(\theta), H(\theta), R(\theta) \) are functions of \( \theta \in D(\theta) \). However, for the sake of simplicity we will suppress the corresponding notations below, i.e. instead of \( \Phi(\theta), \Psi(\theta), \Gamma(\theta), Q(\theta), H(\theta), R(\theta) \) we will write \( \Phi, \Psi, \Gamma, Q, H, R \). We make the same assumptions about the SODM.

### 5. The Set \( \mathcal{A} \) of Adaptive Models \( \mathcal{M}(\hat{\theta}) \)

Let us consider the set of adaptive models

\[
\mathcal{A} = \{ \mathcal{M}(\hat{\theta}) \mid \hat{\theta} \in \Theta \subset \mathbb{R}^l \}
\]

Here we emphasize the fact that we construct adaptive models in the same class as \( \mathcal{M} \) belongs to with the only difference that the unknown parameter \( \theta \) in \( \mathcal{M}(\theta) \) is replaced by \( \hat{\theta} \) to obtain \( \mathcal{M}(\hat{\theta}) \). In so doing, each particular value of \( \hat{\theta} \), an estimate of \( \theta \), leads to a fixed model \( \mathcal{M}(\hat{\theta}) \). In accordance with The Active Principle of Adaptation (APA) [1], only when \( \hat{\theta} \) ranges over \( \Theta \) in search of \( \hat{\theta} \) for the goal \( \mathcal{M}(\hat{\theta}) \) or \( \mathcal{M}(\hat{\theta}) \) as governed by a smart, unsupervised helmsman equipped by a vision of the goal in state space and
able to pursue it, we obtain a model $\mathcal{M}(\hat{\theta})$ of active type within the set $\mathcal{A}$ (12). In this case, $\hat{\theta}$ will act as a self-tuned model parameter and so should be labeled by $\tau$, the time instant of model’s inner clock, in order to get thereby the emphasized notations $\hat{\theta}_t$ and $\mathcal{M}(\hat{\theta}_t)$ in describing parameter identification algorithms (PIAs) to be developed. From this point on $\mathcal{M}(\hat{\theta})$ becomes an adaptive estimator.

Remark 1 Note in passing that pace of $\tau$ may differ from that of $t$. We shall need to discriminate between $\tau$ and $t$ later when developing a PIA.

Remark 2 If we work in the context of SODM, the set

$$\mathcal{A} = \{ \mathcal{M}(\hat{\theta}) \mid \hat{\theta} \in \Theta \subset \mathbb{R}^1 \}$$

instead of (12) should be used.

At this junction, we identify the following tasks as pending:

1) Express $\mathcal{M}(\hat{\theta})$ or $\mathcal{M}(\hat{\theta})$ in an explicit form.
2) Build up APIs that could offer vision of the goal.
3) Examine APIs’ capacity to visualize the goal.
4) Develop a PIA that could help pursuing the goal.

Consider here the first three points consecutively.

5.1. Parameterized Adaptive Models

Reasoning from (6), (9), we set the adaptive model

$$\mathcal{M}(\hat{\theta}): \hat{g}_{t+1} = \Phi(\hat{\theta}) \hat{g}_{t+1} + \Psi(\hat{\theta}) \eta_{t+1} + \eta_{t+1}$$

or equivalently (due to $G_t = \Phi K_t$) the model

$$\hat{g}_{t+1} = \Phi(\hat{\theta}) \hat{g}_{t+1} + \Psi(\hat{\theta}) \eta_{t+1}$$

$$\mathcal{M}(\hat{\theta}): \hat{g}_{t+1} = \hat{g}_{t+1} + K_t(\hat{\theta}) \eta_{t+1}$$

(15)

$$y_t = H(\hat{\theta}) \hat{g}_{t+1} + \eta_{t+1}$$

as a member of $\mathcal{A}$ (12). Here $\hat{\theta}$ is the self-tuned parameter intended to estimate (in one-to-one correspondence) parameter $\theta$. In parallel, reasoning from $\mathcal{M}(\hat{\theta})$, we build the adaptive model

$$\mathcal{M}(\hat{\theta}): \hat{g}_{t+1} = \Phi_*(\hat{\theta}) \hat{g}_{t+1} + \Psi_*(\hat{\theta}) \eta_{t+1} + \eta_{t+1}$$

(16)

or equivalently (due to $G_t = \Phi_* K_*^t$) the model

$$\hat{g}_{t+1} = \Phi_*(\hat{\theta}) \hat{g}_{t+1} + \Psi_*(\hat{\theta}) \eta_{t+1}$$

$$\mathcal{M}(\hat{\theta}): \hat{g}_{t+1} = \hat{g}_{t+1} + K_*^t(\hat{\theta}) \eta_{t+1}$$

(17)

where $H_t$ does not depend on $\hat{\theta}$. Matrices $G_*(\hat{\theta})$ and $K_*^t(\hat{\theta})$ are evaluated according to (7), (8).
versus the estimates.

\[ \Phi(\beta) \]

and

\[ \Theta^* \]

according to OPI. \( \Phi \), thus bringing-

\[ \{ \Psi, \} \]

or

\[ \{ \psi \} \]

of parameter and state vector estimates.

\[ \text{Riccati equations, which positively affects the quality} \]

of unknown parameters. The minimization of API by

\[ \text{to get the optimal model} \]

or

\[ \text{identical to} (21), \]

that is

\[ \forall t \in \mathbb{Z}_+: \mathcal{M}^e(\hat{\Theta}^t) \equiv \mathcal{M}^e(\Theta^t), \]

in the following three setups:

\textbf{Setup 1} (Random Control Input) \( \{ u \} \) is a preassigned

zero-mean orthogonal wide-sense stationary process orthogonal to

\( \{ w, v \} \) but in contrast to \( \{ w \} \) and

\( \{ v \} \), known and serving as a testing signal;

\textbf{Setup 2} (Pure Filtering) \( \forall t \in \mathbb{Z}_+: u_t = 0, \) and

\textbf{Setup 3} (Close-loop Control) with \( \{ \psi \} \), which does not depend on

\( \Theta \).

\textbf{Proof} is similar to the proof of Theorem 2 in [2].

\textbf{Remark 3} Our main goal is to identify the vector \( \Theta \)

of unknown parameters. The minimization of API by

some PIA allows us to determine the optimal value \( \Theta^* \).

Then it must be substituted into (14)-(15) (or (16)-(17))

to get the optimal model \( \mathcal{M}(\hat{\Theta}) \) (or \( \mathcal{M}^e(\hat{\Theta}) \)). At the

same time, we obtain the optimal estimates \( \hat{x}_{i+1} \) (or \( \hat{x}_{i+1}^* \)) according to OPI.

5.3. Main Conceptual Novelty

Seeking minimum of API in parameters of Data Model

instead of parameters of Adaptive Filter is more profitable, as:

- It takes into account the dynamics of the discrete

Riccati equations, which positively affects the quality

of parameter and state vector estimates.
The number of unknown parameters can be substantially (an order of magnitude or more) reduced thus helping avoid the difficulties of PhOP.

API gradient is calculated easily—without the construction of sensitivity model of adaptive filter.

It can be implemented in the case of non-stationary systems, which is critical, for example, to handle the navigation data.

Thus, the proposed variant of API method is a new, thanks to the solution of its important tasks:

1) Numerical construction of API, which has the same minimizing argument that the OPI does;
2) The numerical minimization of the API by conventional optimization methods such as Newton-Raphson method, and
3) The combination of parameter identification of the system with the process of adaptive estimation of its states.

6. Simulation Examples

We take two examples to simulate:

**E1** Second order open-loop system with unknown parameters \((f_1, f_2, \beta, \alpha)\) is given by

\[
\begin{align*}
x_{r+1} &= \begin{bmatrix} 0 & 1 \\ f_1 & f_2 \end{bmatrix} x_r + \begin{bmatrix} 0 \\ \beta \end{bmatrix} u_r + \begin{bmatrix} 0 \\ \alpha \end{bmatrix} w_r \\
y_r &= \begin{bmatrix} 1 & 0 \end{bmatrix} x_r + v_r
\end{align*}
\]

Unknown parameters should be identified. Adaptive model parameter is the four-component vector

\[\hat{\theta} = (\hat{f}_1, \hat{f}_2, \hat{\beta}, \hat{\alpha}).\]

Its true value is

\[\theta^* = (-0.8, 0.1, 1.0, 0.4).\]

Covariances \(Q\) and \(R\) of the noises \(w_r\) and \(v_r\) are equal to 0.04 and 0.06, correspondingly.

**E2** The same (but closed-loop) system as in E1. The system is designed to operate with a minimum expected control cost

\[
J_e = \mathbb{E} \left\{ \frac{1}{500} \sum_{r=1}^{500} \left( x_r^2 + 5u_{r-1}^2 \right) \right\}.
\]

Unknown parameters \(\alpha, \beta, f_1\) and \(f_2\) should be identified. The true values of parameters are the same as in E1.

Simulation results of Figures 1-4 and 5-7 obtained from Julia Tsyganova’s MATLAB programs demonstrate equimodality (coincidence of the minimizing arguments) of the auxiliary performance index with the original performance index. It is seen that the minimums of OPI and API coincide near \(\theta^*\). Thus, the obtained results confirm applicability of the presented method.

---

**Figure 4. The values of \(J'_e\) and \(J'_f\) versus the estimates of parameter \(\alpha\) (Example E1).**

**Figure 5. The values of \(J'_e\) and \(J'_f\) versus the estimates of parameter \(f_1\) (Example E2).**

**Figure 6. The values of \(J'_e\) and \(J'_f\) versus the estimates of parameter \(f_2\) (Example E2).**
7. Conclusion

The present paper gives a comprehensive solution to the problem of seeking minimum of \( J^*(\hat{\theta}) \) in parameters \( \hat{\theta} \) of Data Model \( D(\hat{\theta}) \) or \( D'(\hat{\theta}) \) instead of parameters of Adaptive Filter \( M(\hat{\theta}) \) or \( M'(\hat{\theta}) \). The obtained results were verified by two numerical simulation examples.

Our further research is aimed at obtaining solutions to the following issues:

- Economic feasibility, numeric stability and convergence reliability of each proposed parameter identification algorithm.
- Numerical testing of the approach and determining the scope of its appropriate use in real life problems, for example, in Health Care field [6].

8. Acknowledgements

This work was partly supported by the RFBR Grant No. 13-01-97035.

REFERENCES