Adaptive experiment design for LTI systems

Lirong Huang, Håkan Hjalmarsson, László Gerencsér

Abstract

Optimal experiment design for parameter estimation is a research topic that has been in the interest of various studies. A key problem in optimal input design is that the optimal input depends on some unknown system parameters that are to be identified. Adaptive design is one of the fundamental routes to handle this problem. Although there exist a rich collection of results on adaptive experiment design, there are few results that address these issues for dynamic systems. This paper proposes an adaptive input design method for general single-input single-output linear-time-invariant systems.

I. INTRODUCTION

Optimal experiment design for parameter estimation is a research topic that has been in the interest of various studies over the past century (see, e.g., [4], [5], [9], [10], [17], [25], [44], [54], [61] and references therein). The existing results include many analytical solutions (see [27], [30], [23] and the references therein). Typically, a key problem in optimal input design is that the optimal input depends on some unknown system properties, e.g., system parameters, that are to be identified. One of the fundamental routes to cope with this problem is to employ adaptive input schemes, i.e., as information from the system is gathered the input properties are changed. Adaptive design is usually called sequential design in statistics literature, where there exist a rich collection of results and applications (see, e.g., [35] and the references therein). Adaptive

* This work was supported by the European Research Council under the advanced grant LEARN, contract 26738, and by the Swedish Research Council under contract 621-2009-4017.

Lirong Huang and Håkan Hjalmarsson are with ACCESS Linnaeus Center and Automatic Control Lab, School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden (Email addresses: lirong.huang@ee.kth.se (L. Huang), hakan.hjalmarsson@ee.kth.se (H. Hjalmarsson)).

László Gerencsér is with Computer and Automation Institute of the Hungarian Academy of Science, Budapest, Hungary (Email address: gerencser@sztaki.hu).
experiment design has also been studied in many works in engineering literature (see, e.g., [39], [52], [56], [21], [22] and [29]). However, as pointed out in [27] and [22], there are few results that address these issues for dynamic systems.

An iterative method is outlined in [39], where, in each iteration, an optimal input design problem is solved with the unknown true system parameter replaced by the most recent estimate of the parameter. It is assumed that the optimization is restricted to inputs where a finite number of auto-correlations can be nonzero and the minimum phase spectral factor of the input spectrum that solves the optimization problem is used as a filter whose input is fed with a realization of a white noise process. However, in this work, no formal analysis of the statistical properties of the sequence of parameter estimates is provided and, moreover, all past data are used in the proposed scheme when updating the parameter estimate and therefore the procedure is not recursive. In [21], a general stochastic framework for optimal adaptive input design has been outlined. But it is noticed that some of the required technical results have not been completely clarified in the work. Recently, [22] take a different approach and focus on a smaller class of problems, namely, identification of ARX systems with input filter of finite impulse response (FIR) type as in [39]. The advantage of using ARX-models is that the analysis of the recursive least-squares method can be carried out with a simple but powerful result in [36].

In this paper, following the framework proposed in [21], we study the adaptive input design method for single-input single-output (SISO) linear time invariant (LTI) systems based on the certainty equivalence principle. This contribution is a formal development of the scheme outlined in [21] for general SISO LTI systems.

**Notation:** Throughout the paper, unless otherwise specified, we will employ the following notation. Our problem will be embedded in an underlying complete probability space \((\Omega, \mathcal{F}, \mathbb{P})\), where \(\Omega\) is the sample space, \(\mathcal{F}\) is the \(\sigma\)-algebra that defines events \(E\) in \(\Omega\) which are measurable, i.e., for which the probability \(\mathbb{P}(E)\) is defined. Let \(E[\cdot]\) be the expectation operator with respect to the probability measure. If \(A\) is a vector or matrix, its transpose is denoted by \(A^T\). If \(P\) is a square matrix, \(P > 0\) (\(P < 0\)) means that \(P\) is a symmetric positive (negative) definite matrix of appropriate dimensions while \(P \geq 0\) (\(P \leq 0\)) is a symmetric positive (negative) semidefinite matrix. If square matrix \(P\) is nonsingular, its inverse is denoted by \(P^{-1}\). \(I_m\) stands for the identity matrix of order \(m\), \(0_{m \times n}\) the zero matrix of dimensions \(m \times n\) and \(0_m = 0_{m \times 1}\). Denote
by $\lambda_M(\cdot)$, $\lambda_m(\cdot)$ and $\rho(\cdot)$ the maximum eigenvalue, minimum eigenvalue and spectral radius of a matrix, respectively. Let $|\cdot|$ denote the Euclidean norm of a vector and their induced norms of a matrix, respectively. Unless explicitly stated, matrices are assumed to have real entries and compatible dimensions.

II. LTI SYSTEM AND INPUT SIGNAL

Let us consider a general form of SISO LTI models (see, e.g., [42])

$$A^*(q)y_n = \frac{B^*(q)}{F^*(q)}u_n + \frac{C^*(q)}{D^*(q)}e_n$$  \hspace{1cm} (1)

where $A^*(q)$, $B^*(q)$, $F^*(q)$, $C^*(q)$ and $D^*(q)$ are polynomials in the backward shift operator $q^{-1}$ of degrees $p_a$, $p_b$, $p_f$, $p_c$ and $p_d$, respectively,

$$A^*(q) = 1 + \sum_{j=1}^{p_a} a_j^* q^{-j}, \quad B^*(q) = \sum_{j=1}^{p_b} b_j^* q^{-j}, \quad F^*(q) = 1 + \sum_{j=1}^{p_f} f_j^* q^{-j},$$

$$C^*(q) = 1 + \sum_{j=1}^{p_c} c_j^* q^{-j}, \quad D^*(q) = 1 + \sum_{j=1}^{p_d} d_j^* q^{-j},$$

which satisfies

**Assumption 1:** $A^*(z) \neq 0$, $F^*(z) \neq 0$, $C^*(z) \neq 0$ and $D^*(z) \neq 0$ for all $|z| \geq 1$;

**Assumption 2:** the model structure (5) below is globally identifiable at $\theta^*$ (see [42] Theorem 4.1, p116), i.e.,

i) there is no common factor to all $z^{p_a}A^*(z)$, $z^{p_b}B^*(z)$ and $z^{p_c}C^*(z)$,

ii) there is no common factor to $z^{p_b}B^*(z)$ and $z^{p_f}F^*(z)$,

iii) there is no common factor to $z^{p_c}C^*(z)$ an $z^{p_d}D^*(z)$,

iv) if $p_a \geq 1$, then there must be no common factor to $z^{p_f}F^*(z)$ an $z^{p_d}D^*(z)$,

v) if $p_d \geq 1$, then there must be no common factor to $z^{p_a}A^*(z)$ an $z^{p_b}B^*(z)$,

vi) if $p_f \geq 1$, then there must be no common factor to $z^{p_a}A^*(z)$ an $z^{p_c}C^*(z)$;

**Assumption 3:** the noise process $\{e_n\}$ is a sequence of independent random variables such that

$$\mathbb{E}[e_n] = 0, \quad \mathbb{E}[e_n^2] = \sigma_e^2, \quad \sup_n \mathbb{E}[\exp(\alpha e_n^2)] < \infty$$  \hspace{1cm} (2)

for some $\alpha > 0$, where $\sigma_e^2 > 0$ is unknown.
The system (1) has a state-space representation of the form
\[
\begin{align*}
    x_{n+1} &= A_\xi x_n + B_\xi u_n + K_\xi e_n \\
    y_n &= C_\xi x_n + e_n
\end{align*}
\] (3)
where \(x_n \in \mathbb{R}^n\), \(u_n \in \mathbb{R}\) and \(y_n \in \mathbb{R}\) represent the states, input and output of the system, respectively; the transition matrix \(A_\xi\) has all its eigenvalues strictly inside the unit circle, i.e., the system (3) is internally stable; and the matrix \(A_\xi - K_\xi C_\xi\) has all its eigenvalues strictly inside the unit circle, i.e., the system (3) is inversely stable from \(\{y_n\}\) to \(\{e_n\}\).

**Remark 2.1:** The condition in Assumption 1 on \(C^*(z)\) and \(D^*(z)\) is not a restriction while that on \(A^*(z)\) and \(F^*(z)\) is required for the stability of the system. Assumption 3 on the noise is certainly satisfied for wide-sense stationary Gaussian sequences (see [13], [19], [20]).

In many ways, system (1) is a natural model from a physical point of view: the dynamics from input \(u\) to output \(y\) is modeled separately from the measurement noise. For simplicity, assume that the system is at rest prior to time \(n = 0\), i.e., \(y_n = u_n = e_n = 0\) for \(n < 0\). Suppose that the true system can be expressed by (1) and write
\[
\theta^* = \begin{bmatrix} \theta_{A}^* & \theta_{B}^* & \theta_{C}^* & \theta_{D}^* \end{bmatrix}^T \in \text{int} D_\theta,
\] (4)
with \(\theta_{A}^* = (a_1^* \ a_2^* \ \cdots \ a_{p_a}^*)^T \in \mathbb{R}^{p_a}\), \(\theta_{B}^* = (b_1^* \ b_2^* \ \cdots \ b_{p_b}^*)^T \in \mathbb{R}^{p_b}\), \(\theta_{F}^* = (f_1^* \ f_2^* \ \cdots \ f_{p_f}^*)^T \in \mathbb{R}^{p_f}\), \(\theta_{C}^* = (c_1^* \ c_2^* \ \cdots \ c_{p_c}^*)^T \in \mathbb{R}^{p_c}\), \(\theta_{D}^* = (d_1^* \ d_2^* \ \cdots \ d_{p_d}^*)^T \in \mathbb{R}^{p_d}\)
for the true system parameters and denote by \(\theta\) an arbitrary vector of the same structure, \(\theta = \begin{bmatrix} \theta_{A}^T & \theta_{B}^T & \theta_{F}^T & \theta_{C}^T & \theta_{D}^T \end{bmatrix}^T\) with \(\theta_{A} = (a_1 \ a_2 \ \cdots \ a_{p_a})^T\), \(\theta_{B} = (b_1 \ b_2 \ \cdots \ b_{p_b})^T\), \(\theta_{F} = (f_1 \ f_2 \ \cdots \ f_{p_f})^T\), \(\theta_{C} = (c_1 \ c_2 \ \cdots \ c_{p_c})^T\) and \(\theta_{D} = (d_1 \ d_2 \ \cdots \ d_{p_d})^T\), where \(D_\theta \subset \mathbb{R}^{p_\theta}\) with \(p_\theta = p_a + p_b + p_f + p_c + p_d \geq 1\) is a compact set. We also write
\[
\begin{align*}
    A(q, \theta_A) &= 1 + \sum_{j=1}^{p_a} a_j q^{-j}, \quad B(q, \theta_B) = \sum_{j=1}^{p_b} b_j q^{-j}, \quad F(q, \theta_F) = 1 + \sum_{j=1}^{p_f} f_j q^{-j}, \\
    C(q, \theta_C) &= 1 + \sum_{j=1}^{p_c} c_j q^{-j}, \quad D(q, \theta_D) = 1 + \sum_{j=1}^{p_d} d_j q^{-j}
\end{align*}
\] (5)
and hence \(A(q, \theta_A) = A^*(q), \ B(q, \theta_B^*) = B^*(q), \ F(q, \theta_F^*) = F^*(q), \ C(q, \theta_C^*) = C^*(q)\) and \(D(q, \theta_D^*) = D^*(q)\). In this paper, \(D_\theta\) is defined as (32) (also denoted by (34)) below such that \(C(z, \theta_C) \neq 0\) and \(F(z, \theta_F) \neq 0\) for all \(|z| \geq 1\) on \(D_\theta\).
Remark 2.2: By continuity of the model structure (5) (see also Appendix A), there exists a subset $D_{\theta_0} \subseteq D_{\theta}$ with $\theta^* \in \text{int}D_{\theta_0}$ such that Assumption 2 holds for all $\theta \in D_{\theta_0}$, or say, the model structure (5) is globally identifiable at all $\theta \in D_{\theta_0}$, where $D_{\theta}$ is given by (32) below.

The generation of the input signal is given as follows.

Assumption 4: The input signal $\{u_n\}$ is defined in terms of an external source represented by a state-space system that is at rest prior to time $n = 0$,

\[
\begin{align*}
    z_{n+1} &= A_z(r(\theta_n))z_n + B_z(r(\theta_n))s_n \\
    u_n &= C_z(r(\theta_n))z_n + D_z(r(\theta_n))s_n
\end{align*}
\]  

(6)

where $\theta_n$ is the estimate of $\theta^*$, $r = r(\cdot)$ is a continuous function $D_{\theta}$ and $A_z \in \mathbb{R}^{m \times m}$, $B_z \in \mathbb{R}^m$, $C_z^T \in \mathbb{R}^m$, $D_z \in \mathbb{R}$ are continuous functions of $r$; the noise process $\{s_n\}$ is an independent sequence of random variables independent of $\{e_n\}$ such that

\[
E[s_n] = 0, \quad E[s_n^2] = 1, \quad \sup_n E[\exp(\alpha_s s_n^2)] < \infty
\]  

(7)

for some $\alpha_s > 0$.

According to [31, Corollary 1.1, p21] (see also [14] and [8]), the time-varying system (6) is bounded-input bounded-output (BIBO) stable if the following condition holds.

Assumption 5: The joint spectral radius of the bounded set of matrices $\Sigma_z = \{A_z(\cdot) : \theta \in D_{\theta}\}$ is less than one, i.e.,

\[
\rho(\Sigma_z) = \limsup_{n \to \infty} \rho_n(\Sigma_z) < 1,
\]  

(8)

where $\rho_n(\Sigma_z) = \sup\{[\rho(A)]^{1/n} : A \in \Sigma_z^n\}$ and $\Sigma_z^n = \{A_n \cdots A_2 A_1 : A_k \in \Sigma_z, k = 1, 2, \cdots, n\}$.

In practice, the input generator (6) should be designed such that this assumption is satisfied, for which an example will be given in the application of our proposed method.

Remark 2.3: Obviously, Assumption 5 implies that system (6) with $\theta_n$ fixed to any $\theta \in D_{\theta}$ is BIBO stable, which means that $G_s(q, \theta) = C_z(r(\theta)) [qI_m - A_z(r(\theta))]^{-1} B_z(r(\theta)) + D_z(r(\theta))$ is stable for any $\theta \in D_{\theta}$. Let us consider the case with fixed $\theta_n = \theta \in D_{\theta}$ and write $G_s(q) = G_s(q, \theta)$, which is a stable filter. So the input sequence $\{u_n\}$ generated by $u_n = G_s(q)n$ is a stationary signal with spectrum (see, e.g., [42, Theorem 2.2, p40])

\[
\Psi_u(e^{i\omega}, \hat{r}) = \sum_{\tau = -\infty}^{\infty} \hat{r}_\tau e^{i\omega\tau} = |G_s(e^{i\omega})|^2 \sigma_e^2,
\]
where \( \hat{r} = [\hat{r}_0 \ \hat{r}_1 \ \cdots \ \hat{r}_m \ \cdots]^T \) with \( \hat{r}_\tau = \hat{r}_{-\tau} = \mathbb{E}[u_n u_{n-\tau}] \) is the auto-correlations of the input signal \( u \). Clearly, this yields
\[
\Psi_u(e^{i\omega}, \hat{r}) \geq 0
\]
for all \( \omega \) and, moreover,
\[
\Psi_u(e^{i\omega}, \hat{r}) > 0
\]
for almost all \( \omega \), which means that the input signal \( \{u_n\} \) is persistently exciting (see, e.g., [42, Definition 13.2, p414]).

## III. Prediction Error Estimation

For any fixed \( \theta \in \mathcal{D}_\theta \), define the prediction error process by
\[
\varepsilon_n(\theta, r) = y_n(r) - \hat{y}_n(\theta, r)
\]
for all \( n \geq 0 \), where \( y_n(r) \) is the output of system (1) with input signal \( u_n(r) \) generated by (6) with \( \theta_n \) fixed to \( \theta_n = \theta \), and the one-step predictor \( \hat{y}_n(\theta, r) \) for LTI model (1) is
\[
\hat{y}_n(\theta, r) = \frac{D(q, \theta_D)B(q, \theta_B)}{C(q, \theta_C)F(q, \theta_B)}u_n(r) + \left[ 1 - \frac{D(q, \theta_D)A(q, \theta_A)}{C(q, \theta_C)} \right] y_n(r),
\]
which can also be written as a recursion
\[
C(q, \theta_C)F(q, \theta_B)\hat{y}_n(\theta, r) = F(q, \theta_B) \left[ C(q, \theta_C) - D(q, \theta_D)A(q, \theta_A) \right] y_n(r) + D(q, \theta_D)B(q, \theta_B)u_n(r).
\]

Introducing the auxiliary variables
\[
w_n(\theta, r) = \frac{B(q, \theta_B)}{F(q, \theta_F)}u_n(r) \quad \Rightarrow \quad w_n(\theta, r) = \sum_{j=1}^{p_b} b_j u_{n-j}(r) - \sum_{j=1}^{p_f} f_j u_{n-j}(\theta, r)
\]
and
\[
v_n(\theta, r) = A(q, \theta_A)y_n(r) - w_n(\theta, r) \quad \Rightarrow \quad v_n(\theta, r) = y_n(r) + \sum_{j=1}^{p_a} a_j y_{n-j}(r) - w_n(\theta, r),
\]
we have (see, e.g., [43] and [42])
\[
\varepsilon_n(\theta, r) = y_n(r) - \hat{y}_n(\theta, r) = y_n(r) - \theta^T \varphi_n(\theta, r)
\]
for all \( n \), where
\[
\varphi_n(\theta, r) = \begin{bmatrix} -\hat{y}_{n-1}(r) \ 
\hat{u}_{n-1}(r) \ 
-\hat{u}_{n-1}(\theta, r) \ 
\hat{\varepsilon}_{n-1}(\theta, r) \ 
-\hat{\varepsilon}_{n-1}(\theta, r) \end{bmatrix}^T
\]
with \( \tilde{y}_{n-1}(r) = [y_{n-1}(r) \cdots y_{n-p_n}(r)]^T \in \mathbb{R}^{p_n} \), \( \tilde{u}_{n-1}(r) = [u_{n-1}(r) \cdots u_{n-p_n}(r)]^T \in \mathbb{R}^{p_n} \),\n\n\( \tilde{w}_{n-1}(\theta, r) = [w_{n-1}(\theta, r) \cdots w_{n-p_n}(\theta, r)]^T \in \mathbb{R}^{p_n} \), \( \tilde{e}_{n-1}(\theta, r) = [\varepsilon_{n-1}(\theta, r) \cdots \varepsilon_{n-p_n}(\theta, r)]^T \in \mathbb{R}^{p_n} \). In this case, the input signal \( \{u_n\} \) is generated by system (6) with fixed \( \theta \in D_\theta \) and \( r = r(\theta) \). Note that \( y_n(r) = 0 \), \( u_n(r) = 0 \) and \( \varepsilon_n(\theta, r) = 0 \) for all \( n < 0 \) since the system is at rest prior to time \( n = 0 \). The overline indicates that (11) is defined as a frozen-parameter process (for fixed \( \theta \in D_\theta \)). For simplicity, we write \( u_n = u_n(\theta), e_n = e_n(\theta, r), e_{\theta, n} = e_{\theta, n}(\theta, r) \), etc. where there is no ambiguity, and, to emphasize the dependence on \( \theta \), we also write \( e_n(\theta) = e_n(\theta, r), e_{\theta, n}(\theta) = e_{\theta, n}(\theta, r) \) and so on.

The asymptotic cost function is defined by (see [41] and [19])

\[
W(\theta) = \lim_{n \to \infty} \frac{1}{2} \mathbb{E}[\varepsilon_n^2(\theta)].
\] (18)

Then the gradient and the Hessian of \( W \) are given by

\[
W_\theta(\theta) = \frac{\partial}{\partial \theta} W(\theta) = \lim_{n \to \infty} \mathbb{E}[\varepsilon_{\theta, n}(\theta)\varepsilon_n(\theta)]
\] (19)

and

\[
W_{\theta\theta}(\theta) = \frac{\partial^2}{\partial \theta^2} W(\theta) = \frac{\partial^2}{\partial \theta \partial \theta^T} W(\theta)
\] (20)

respectively, where \( \varepsilon_{\theta, n}(\theta) = \varepsilon_{\theta, n}(\theta, r) = \frac{\partial}{\partial \theta} \varepsilon_n(\theta, r) = [\frac{\partial}{\partial \theta_1} \varepsilon_n(\theta, r) \cdots \frac{\partial}{\partial \theta_p} \varepsilon_n(\theta, r)]^T \). We define

\[
G(\theta) = \lim_{n \to \infty} \mathbb{E}[\varepsilon_{\theta, n}(\theta)\varepsilon_{\theta, n}^T(\theta)]
\] (21)

for all \( \theta \in D_\theta \) and then have \( W_{\theta\theta}(\theta^*) = G(\theta^*) \).

Moreover, the model (11) gives

\[
\varepsilon_{\theta, n}(\theta, r) = -\varphi_n(\theta, r) = \left[ \tilde{y}_{n-1}^T(r) \quad -\tilde{u}_{n-1}^T(r) \quad \tilde{w}_{n-1}^T(\theta, r) \quad -\tilde{\varepsilon}_{n-1}^T(\theta, r) \quad \tilde{\varepsilon}_{n-1}^T(\theta, r) \right]^T
\] (22)

and

\[
\varepsilon_n(\theta, r) = \tilde{y}_{n-1}^T(r)(\theta_A - \theta_A^*) - \tilde{u}_{n-1}^T(r)(\theta_B - \theta_B^*) + \tilde{w}_{n-1}^T(\theta, r)(\theta_C - \theta_C^*) - \tilde{\varepsilon}_{n-1}^T(\theta, r)\theta_C - \theta_C^* + \Delta \tilde{w}_{n-1}\theta_F - \Delta \tilde{\varepsilon}_{n-1}\theta_F + e_n,
\] (23)

where \( \Delta \tilde{w}_{n-1} = \Delta \tilde{w}_{n-1}(\theta, r) = \tilde{w}_{n-1}(\theta, r) - \tilde{w}_{n-1}(\theta^*, r), \Delta \tilde{\varepsilon}_{n-1} = \Delta \tilde{\varepsilon}_{n-1}(\theta, r) = \tilde{\varepsilon}_{n-1}(\theta, r) - \tilde{\varepsilon}_{n-1}(\theta^*, r) \) and \( \Delta \tilde{e}_{n-1} = \Delta \tilde{e}_{n-1}(\theta, r) = \varepsilon_{n-1}(\theta, r) - \tilde{e}_{n-1} \) with \( \tilde{e}_{n-1} = \tilde{e}_{n-1}(\theta^*, r) = [e_{n-1} \cdots e_{n-p_e}]^T \).
In the limit $n \to \infty$, the prediction error estimate of $\theta$ is defined by

\[
E[\varepsilon_{\theta,n}(\theta)\varepsilon_n(\theta)] = E[\varepsilon_{\theta,n}(\theta, r)\varepsilon_n(\theta, r)] \\
= E[\varepsilon_{\theta,n}(\theta, r)\varepsilon_{\theta,n}(\theta, r)](\theta - \theta^*) + E[\varepsilon_{\theta,n}(\theta, r)\Delta \tilde{w}_{n-1}^T]\theta_F^* - E[\varepsilon_{\theta,n}(\theta, r)\Delta \tilde{z}_{n-1}^T]\theta_C^* \\
+ E[\varepsilon_{\theta,n}(\theta, r)\Delta \tilde{v}_{n-1}^T]\theta_D^* + E[\varepsilon_{\theta,n}(\theta, r)e_n] \\
= E[\varepsilon_{\theta,n}(\theta, r)\varepsilon_{\theta,n}(\theta, r)](\theta - \theta^*) + E[\varepsilon_{\theta,n}(\theta, r)\Delta \tilde{w}_{n-1}^T]\theta_F^* - E[\varepsilon_{\theta,n}(\theta, r)\Delta \tilde{z}_{n-1}^T]\theta_C^* \\
+ E[\varepsilon_{\theta,n}(\theta, r)\Delta \tilde{v}_{n-1}^T]\theta_D^* \\
= 0.
\] (24)

Obviously, $\theta = \theta^*$ is a solution to equation (24). Suppose that

\textit{Assumption 6:} $\theta = \theta^*$ is the unique solution to the normal equation (24) on $D_0$.

This implies that $\theta^*$ is consistently estimated when the input is generated according to (6) with $\theta = \theta^*$.

The model (11) together with the gradient (22) immediately suggests a Newton-type recursive prediction error estimate of $\theta^*$ as follows (see, e.g., [21] and [43])

\[
\theta_{n+1} = \theta_n - \frac{1}{n + 1}R_n^{-1}\varepsilon_{\theta,n+1}\varepsilon_{n+1},
\] (25)

\[
R_{n+1} = R_n + \frac{1}{n + 1}(\varepsilon_{\theta,n+1}\varepsilon_{\theta,n+1}^T - R_n),
\] (26)

where $\varepsilon_{n+1}$ and $\varepsilon_{\theta,n+1}$ are the online estimates of $\varepsilon_{n+1}(\theta_n)$ and $\varepsilon_{\theta,n+1}(\theta_n)$ given by (11) and (22), respectively.

In order to keep the estimates in some known domain, recursive estimation schemes such as (25)-(26) typically need to be complemented with either a projection or a resetting mechanism (see [12], [19], [21], [20], [22], [28], [34] and [40]). In this work, we consider the recursive estimation algorithm (25)-(26) with a resetting mechanism. Assume that $D_C$ and $D_F$ are convex compact sets with $\theta^*_C \in \text{int}D_C$ and $\theta^*_F \in \text{int}D_F$ such that there are constants $\lambda_c \in (0,1)$, $\lambda_f \in (0,1)$ and symmetric positive definite matrices $V_c \in \mathbb{R}^{p_c \times p_c}$, $V_f \in \mathbb{R}^{p_f \times p_f}$ satisfying

\[
\tilde{C}^T(\theta_C)V_c\tilde{C}(\theta_C) \leq \lambda_c V_c, \quad \forall \theta_C \in D_C \\
\tilde{F}^T(\theta_F)V_f\tilde{F}(\theta_F) \leq \lambda_f V_f, \quad \forall \theta_F \in D_F
\] (27) (28)

October 16, 2013 DRAFT
where \( \tilde{C}(\theta_C) \in \mathbb{R}^{p_c \times p_c} \) and \( \tilde{F}(\theta_F) \in \mathbb{R}^{p_f \times p_f} \) are the companion matrices of \( C(q, \theta_C) \) and \( F(q, \theta_F) \), respectively, that is,

\[
\tilde{C}(\theta_C) = \begin{bmatrix}
-c_1 & -c_2 & \cdots & -c_{p_c-1} & -c_{p_c} \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0
\end{bmatrix}
\quad \text{and} \quad
\tilde{F}(\theta_F) = \begin{bmatrix}
-f_1 & -f_2 & \cdots & -f_{p_f-1} & -f_{p_f} \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0
\end{bmatrix}.
\]

For convenience, we also write \( \tilde{C} = \tilde{C}(\theta) = \tilde{C}(\theta_C) \) and \( \tilde{F} = \tilde{F}(\theta) = \tilde{F}(\theta_F) \) where there is no ambiguity. Clearly, the existence of common Lyapunov functions \( V_c \) and \( V_f \) in (27) and (28) implies that the corresponding switching systems with transition matrices \( \{\tilde{C}(\theta_C) : \theta_C \in D_C\} \) are (uniformly asymptotically) stable and hence the joint spectral radii of bounded sets of matrices \( \Sigma_C = \{\tilde{C}(\theta_C) : \theta_C \in D_C\} \) and \( \Sigma_F = \{\tilde{F}(\theta_F) : \theta_F \in D_F\} \) are less than one, i.e.,

\[
\rho(\Sigma_C) = \limsup_{n \to \infty} \rho_n(\Sigma_C) < 1 \quad \text{and} \quad \rho(\Sigma_F) = \limsup_{n \to \infty} \rho_n(\Sigma_F) < 1,
\]

respectively, where \( \rho_n(\Sigma_C) = \sup\{[\rho(C)]^{1/n} : C \in \Sigma_C^n\} \) with \( \Sigma_C^n = \{C_n \cdots C_2 C_1 : C_k \in \Sigma_C, k = 1, 2, \cdots, n\} \) and \( \rho_n(\Sigma_F) = \sup\{[\rho(F)]^{1/n} : F \in \Sigma_F^n\} \) with \( \Sigma_F^n = \{F_n \cdots F_2 F_1 : F_j \in \Sigma_F, j = 1, 2, \cdots, n\} \) (see [14], [8], [31]).

**Remark 3.1:** Suppose that \( D_C \) and \( D_F \) are convex polyhedra with vertices \( \theta_{C,k}, k = 1, \cdots, n_c, \) and \( \theta_{F,j}, j = 1, \cdots, n_f, \) respectively. It is observed that symmetric positive definite matrices \( V_c \) and \( V_f \) satisfy (27) and (28) for all \( \theta_C \in D_C \) and \( \theta_F \in D_F \) if the following linear matrix inequalities (LMIs)

\[
\tilde{C}^T(\theta_{C,k}) V_c \tilde{C}(\theta_{C,k}) < V_c, \quad k = 1, \cdots, n_c
\]

\[
\tilde{F}^T(\theta_{F,j}) V_f \tilde{F}(\theta_{F,j}) < V_f, \quad j = 1, \cdots, n_f
\]

hold, respectively.

**Remark 3.2:** Conditions (27)-(28) are certainly restrictive (see the remark below Condition 4.5 in [19]). But it is noticed that these restrictive conditions vanish for ARARX systems, i.e., system (1) with \( p_c = p_f = 0 \). The ARARX is a stochastic model commonly used in economics, engineering, health and medical science literature (see, e.g., [3], [16], [24], [32], [49], [50], [51], [53], [59] and the references therein). As an application example of our proposed method, a problem of adaptive input design for a class of ARARX will be considered in Section VI.
Let $D_{\theta}$ be a compact set defined as follows
\[ D_{\theta} = \{ \theta : g(\theta) \leq K_{\theta}, \theta_C \in D_C, \theta_F \in D_F \}, \quad (32) \]
also denoted by (34) below, where $g : \mathbb{R}^{p_{\theta}} \rightarrow \mathbb{R}_+$ is a continuous function, $K_{\theta}$ is a positive constant, $D_C$ and $D_F$ are given by (27) and (28), respectively. Obviously, we have $G(\theta) = \lim_{n \to \infty} \mathbb{E}[\varepsilon_{\theta,n}(\theta)\varepsilon_{\theta,n}^T(\theta)] \geq 0$ for all $\theta \in D_{\theta}$. Under the above assumptions, we have the following important result:

**Lemma 3.1:** There exists a subset $D_{\theta^*} \subseteq D_{\theta}$ such that $\theta^* \in \text{int}D_{\theta^*}$ and $G(\theta) > 0$ for all $\theta \in D_{\theta^*}$.

**Proof:** See Appendix A.

This implies that
\[ R^* = G(\theta^*) = W_{\theta^*}(\theta^*) = \lim_{n \to \infty} \mathbb{E}[\varepsilon_{\theta,n}(\theta^*)\varepsilon_{\theta,n}^T(\theta^*)] > 0. \quad (33) \]

Let $D_R$ be a compact set of symmetric positive definite matrices defined as $D_R = \{ P \in \mathbb{R}^{p_{\theta} \times p_{\theta}} : \kappa_1 I_{p_{\theta}} \leq P \leq \kappa_2 I_{p_{\theta}} \}$, denoted by (35) below, where $\kappa_1$ and $\kappa_2$ are sufficiently small and large positive constants (that will be given by (88) in Appendix D), respectively. In summary, we present the entire adaptive system by (34)-(40) to be analyzed, where the notations are given in Appendix B.

### Adaptive system

\[ D_{\theta} = \{ \theta : g(\theta) \leq K_{\theta}, \theta_C \in D_C, \theta_F \in D_F \}, \quad (34) \]
\[ D_R = \{ P \in \mathbb{R}^{p_{\theta} \times p_{\theta}} : \kappa_1 I_{p_{\theta}} \leq P \leq \kappa_2 I_{p_{\theta}} \}, \quad (35) \]
\[ \theta_0 \in \text{int}D_{\theta}, \quad R_0 \in \text{int}D_R, \quad (36) \]
\[ \Phi_{n+1} = A_{\Phi}(\theta_n)\Phi_n + B_{\Phi}(\theta_n)\eta_n, \quad (37) \]
\[ \theta_{n+1} = \theta_n - \frac{1}{n+1} R_n^{-1} \varepsilon_{\theta,n+1}\varepsilon_{\theta,n+1}^T, \quad (38) \]
\[ R_{n+1} = R_n + \frac{1}{n+1} (\varepsilon_{\theta,n+1}\varepsilon_{\theta,n+1}^T - R_n), \quad (39) \]
\[ (\theta_{n+1}, R_{n+1}) = \begin{cases} (\theta_{n+1-}, R_{n+1-}), & \theta_{n+1-} \in D_{\theta}, R_{n+1-} \in D_R \\ (\theta_0, R_0), & \text{otherwise} \end{cases} \quad (40) \]
IV. CONVERGENCE AND ACCURACY ANALYSIS

In this section, we consider the convergence of the recursive estimation algorithm (34)-(40). It is well known that the algorithm (34)-(40) can be viewed as finite-difference equations, which has a natural connection with ordinary differential equations (ODEs) (see [40], [43], [33] and [34]). The ODE associated with the algorithm is given as follows (see, e.g., [40], [19] and [20])

\[
\frac{d}{dt}\theta_t = -R_t^{-1}W_\theta(\theta_t) \tag{41a}
\]

\[
\frac{d}{dt}R_t = G(\theta_t) - R_t \tag{41b}
\]

for \(t \geq 0\) with initial condition \((\theta_0, R_0)\), where \(W_\theta(\theta_t)\) and \(G(\theta_t)\) are defined by (19) and (21), respectively. Assume the following (see, e.g., [19])

**Assumption 7:** \(\theta_0 \in D_0\), where \(D_0 \subset \text{int}D_\theta\) is a compact set such that

\[
\{\theta_t : t \geq t_0, \theta_{t_0} \in D_0\} \subset \text{int}D_\theta. \tag{42}
\]

Moreover, another condition is imposed on the generator (6) of the input signal

**Assumption 8:** Functions \(r(\cdot), A_z(\cdot), B_z(\cdot), C_z(\cdot)\) and \(D_z(\cdot)\) are twice continuously differentiable with bounded partial derivatives up to second order on \(D_\theta\).

Let \((\mathcal{F}_n, \mathcal{F}_n^+)\), \(n \geq 0\), be a pair of families of \(\sigma\)-algebras such that (i) \(\mathcal{F}_n \subset \mathcal{F}\) is monotone increasing, (ii) \(\mathcal{F}_n^+ \subset \mathcal{F}\) is monotone decreasing, and (iii) \(\mathcal{F}_n\) and \(\mathcal{F}_n^+\) are independent for all \(n \geq 0\). In this paper, we set \(\mathcal{F}_n = \sigma\{e_t, s_t : 0 \leq t \leq n\}\) and \(\mathcal{F}_n^+ = \sigma\{e_t, s_t : t \geq n + 1\}\). According to (1) and (6), \(\{\eta_n\}\) with \(\eta_n = [e_{n+1} s_n]^T\) is wide-sense stationary and \(\{\eta_n^2\}\) is in class \(M^*\) (see [19]). It is also noticed that \(\{\eta_n\}\) is \(L\)-mixing with respect to the \(\sigma\)-algebras \((\mathcal{F}_n, \mathcal{F}_n^+)\) (see Appendix C). We establish the following theorem on convergence by applying the main results in [28] and [19] (see also [20]), which are listed in Appendix C.

**Theorem 4.1:** If Assumptions [1,7] hold, then \(\{(\theta_n, R_n)\}\) computed by the recursive algorithm (34)-(40) converges to \((\theta^*, R^*)\) a.s. as \(n \to \infty\), where \(R^* = G(\theta^*)\) is defined by (33). Moreover, if Assumption [8] also holds, then \(\{(\theta_n, R_n)\}\) satisfies

\[
\theta_n - \theta^* = O_M\left(n^{-1/2}\right) \quad \text{and} \quad R_n - R^* = O_M\left(n^{-1/2}\right) \quad \text{a.s. as } n \to \infty. \tag{43}
\]
Proof: See Appendix D.

Let the input signal denoted by \( \{ u_n^* \} \) be generated by (6) with \( \theta_n = \theta^* \) for all \( n \). Note that \( H(n, \theta^*) = -(R^*)^{-1} \varepsilon_{\theta,n}(\theta^*)e_n \) is a wide-sense stationary process with zero mean and hence [20, Condition 6.1] is satisfied. Under Assumptions [11,8] (see Theorem 4.1), [20, Theorem 6.2] implies that \( S^* = \lim_{n \to \infty} n \mathbb{E} \left[ (\theta_n - \theta^*)(\theta_n - \theta^*)^T \right] \) exists and it satisfies the Lyapunov equation

\[
(A^* + I_p/2)S^* + S^*(A^* + I_p/2)^T + P^* = 0
\]

with \( A^* = \frac{\partial}{\partial \theta} \left[ -R^{-1}(\theta)W_\theta(\theta) \right] \bigg|_{\theta = \theta^*} = -I_p \) (see, e.g., [46, (12), p175]) and therefore \( S^* = P^* \), where

\[
P^* = \sum_{n=-\infty}^{\infty} \mathbb{E} \left[ H(n, \theta^*)H^T(0, \theta^*) \right] = \sigma_e^2(R^*)^{-1}
\]

is the covariance matrix of \( \sqrt{n}(\theta_n - \theta^*) \) as \( n \to \infty \) when the input signal is generated by (6) with \( \theta_n = \theta^* \) for all \( n \).

In an adaptive input design context, the generator (6) arise from the use of the certainty equivalence principle, where the function \( r(\cdot) \) is obtained by solving some optimal input design problem at each step using the model parameter \( \theta \). The following theorem considers the asymptotic normality for the case where the adaptive input (6) is used instead.

**Theorem 4.2:** Suppose that Assumptions [11,8] hold. Then \( \{ (\theta_n, R_n) \} \) computed by the recursive algorithm (34)-(40) satisfies

\[
\sqrt{n}(\theta_n - \theta^*) \xrightarrow{L} \mathcal{N}(0_{p_\theta}, P^*) \quad \text{as} \quad n \to \infty,
\]

where \( P^* \) is the covariance matrix given by (44).

Proof: See Appendix E.

It is observed that Assumptions [11,3] are descriptions of the nature, i.e., LTI system (1) while, in practice, Assumptions [4,8] and [8] should be ensured by the input generator (6) that is designed by the user, which will be illustrated with an application example of our proposed method in the next section.

**V. ADAPTIVE INPUT DESIGN**

We will use the results presented in the previous section to design the adaptive input signal (6) for identification of LTI (1) based on the recursive algorithm (34)-(40). For open loop input
design where the input spectrum is restricted to be of the type \( \Psi_u(e^{j\omega}, \hat{r}) = \sum_{\tau=-m}^{m} \hat{r}_\tau e^{j\omega \tau} \) with \( \hat{r}_\tau = \hat{r}_{-\tau} = \mathbb{E}[u_n u_{n-\tau}] \), it has been shown in [30] (see also [21] and [22]) that a wide range of optimal input design problems can be formulated as the following problem

\[
\begin{aligned}
\min_{\hat{r},X} & \quad J(\hat{r},X,\theta) \\
\text{s.t.} & \quad M(\hat{r},X,\theta) \geq 0
\end{aligned}
\] (46)

where the scalar function \( J \) is linear in \( \hat{r} \) and \( X \) with \( X \) being some matrix-valued auxiliary variable such as the matrix \( Q \) in the following (48), and the constraints (47) can be interpreted as a set of linear matrix inequalities (LMIs) in \( \hat{r} \) and \( X \), see, e.g., (48), (51) and (52) below.

The positivity condition (9), by the positive-real lemma (see, e.g., [30, Lemma 2.1]), can be included in the constraints (47) as

\[
K(Q; \{A_u, B_u, C_u, D_u\}) = \begin{bmatrix}
Q - A_u^T Q A_u & -A_u^T Q B_u \\
B_u^T Q A_u & -B_u^T Q B_u
\end{bmatrix} + \begin{bmatrix}
0 & C_u^T \\
C_u & 2D_u
\end{bmatrix} \geq 0
\] (48)

with \( Q = Q^T \geq 0 \), where

\[
A_u = \begin{bmatrix}
0_{T-1} & 0 \\
I_{m-1} & 0_{m-1}
\end{bmatrix}, \quad B_u = [1 0_{T-1}]^T,
\]

\[
C_u = C_u(\hat{r}) = [\hat{r}_1 \cdots \hat{r}_m], \quad D_u = D_u(\hat{r}) = \frac{1}{2} \hat{r}_0
\] (49)

and hence \( \{A_u, B_u, C_u, D_u\} \) is a controllable state-space realization of the FIR system

\[
\Psi_u^+(e^{j\omega}) = \frac{1}{2} \hat{r}_0 + \sum_{\tau=1}^{m} \hat{r}_\tau e^{-j\omega \tau}.
\] (50)

Moreover, the property of persistent excitation (10) implies the following positivity condition of Toeplitz matrix (see [2], [42] and [57])

\[
R_u(\hat{r}) = \begin{bmatrix}
\hat{r}_0 & \cdots & \hat{r}_m \\
\vdots & \ddots & \vdots \\
\hat{r}_m & \cdots & \hat{r}_0
\end{bmatrix} > 0.
\] (51)

In practice, there is also a power constraint

\[
0 < \hat{r}_{\min} \leq \hat{r}_0 \leq \hat{r}_{\max} < +\infty.
\] (52)

Given any \( \theta \in D_\theta \), one can solve the convex optimization (46)-(47) numerically to any desired accuracy and obtain the auto-correlation sequence \( \hat{r}(\theta) = [\hat{r}_0(\theta) \hat{r}_1(\theta) \cdots \hat{r}_m(\theta)]^T \). Let
the optimal auto-correlations $\hat{r}(\theta^*) = [\hat{r}_0(\theta^*) \; \hat{r}_1(\theta^*) \; \ldots \; \hat{r}_m(\theta^*)]^T$ with $m^*_r = m_r(\theta^*)$, which yields a realization of the optimal input

$$u_n = \hat{F}(q, \theta^*)s_n = \sum_{\tau=0}^{m} \hat{f}_\tau(\theta^*) s_{n-\tau} = \sum_{\tau=0}^{m^*_r} \hat{f}_\tau(\theta^*) s_{n-\tau}, \quad (53)$$

where $\hat{F}(e^{i\omega}, \theta^*) = \sum_{\tau=0}^{m} \hat{f}_\tau(\theta^*) e^{i\omega\tau} = \sum_{\tau=0}^{m^*_r} \hat{f}_\tau(\theta^*) e^{i\omega\tau}$ is the minimum phase spectral factor given by spectral factorization (see [42]). Obviously, if $m^*_r = 0$, $u_n = \hat{f}_0(\theta^*)s_n$ and the transfer function $\hat{F}(q, \theta^*) = \hat{f}_0(\theta^*)$ gives a state-space realization of the form (6) with

$$A_z(\hat{r}(\theta^*)) = 0_{m \times m}, \quad B_z(\hat{r}(\theta^*)) = 0_m, \quad C_z(\hat{r}(\theta^*)) = 0^T_m, \quad D_z(\hat{r}(\theta^*)) = \hat{f}_0(\theta^*); \quad (54)$$

if $m^*_r \geq 1$, then, according to a state-space realization theorem (see, e.g., [55, 26.8 Theorem, p481]), the transfer function $\hat{F}(q, \theta^*)$ admits a state-space realization of the form (6) with

$$A_z(\hat{r}(\theta^*)) = \begin{bmatrix} 0^T_{m^*_r-1} & 0 & 0^T_{m-m^*_r} \\ I_{m^*_r-1} & 0_{m^*_r-1} & 0_{(m-m^*_r) \times (m-m^*_r)} \\ 0_{(m-m^*_r) \times (m^*_r-1)} & 0_{m-m^*_r} & 0_{(m-m^*_r) \times (m-m^*_r)} \end{bmatrix}, \quad B_z(\hat{r}(\theta^*)) = \begin{bmatrix} 1 \\ 0_{m-1} \end{bmatrix}, \quad (55)$$

$$C_z(\hat{r}(\theta^*)) = [\hat{f}_1(\theta^*) \; \ldots \; \hat{f}_{m^*_r}(\theta^*) \; 0^T_{m-m^*_r}], \quad D_z(\hat{r}(\theta^*)) = \hat{f}_0(\theta^*).$$

In practice, a natural approach to circumvent the problem that the optimal filter $\hat{F}(q, \theta^*)$ and hence the optimal generator (6) with $\{A_z(\hat{r}(\theta^*)), \; B_z(\hat{r}(\theta^*)), \; C_z(\hat{r}(\theta^*)), \; D_z(\hat{r}(\theta^*))\}$ depend on the true system parameters $\theta^*$ is to recursively estimate $\theta^*$ and use the certainty equivalence principle, which yields an adaptive filter

$$\hat{F}(q, \theta_n) = \sum_{\tau=0}^{m} \hat{f}_\tau(\theta_n) q^{-\tau} = \sum_{\tau=0}^{m^*_r(\theta_n)} \hat{f}_\tau(\theta_n) q^{-\tau}, \quad (56)$$

and its state-space realization (6) with $\{A_z(\hat{r}(\theta_n)), \; B_z(\hat{r}(\theta_n)), \; C_z(\hat{r}(\theta_n)), \; D_z(\hat{r}(\theta_n))\}$ of the form (54) or (55) when $m_r(\theta_n) = 0$ or $m_r(\theta_n) \geq 1$, respectively. It is noticed that the state-space realization $\{A_z(\hat{r}(\theta_n)), \; B_z(\hat{r}(\theta_n)), \; C_z(\hat{r}(\theta_n)), \; D_z(\hat{r}(\theta_n))\}$, of the form either (54) or (55) and with $\theta_n \in D_\theta$, satisfies the conditions imposed on system (6), i.e., Assumptions 4 and 8. This leads to an adaptive, or sequential, input design scheme as follows.

**Algorithm 5.1:** The details of the adaptive scheme are summarized as the following steps:

1) Fix the order $m$ of the linear time-varying system (6) that generates the input signals.
2) Initial estimate. Define $D_\theta$ and $D_R$ and set $\theta_0 \in D_\theta \subset \text{int}D_\theta$, $R_0 \in \text{int}D_R$ and $n = 0$.

3) Generate input process. Take $\{s_n\}$ to be a sequence of independent random variables satisfying (7).

4) Input spectrum update. Compute the optimal solution $\hat{r}(\theta_n)$ to (46)-(47) with $\theta = \theta_n$, where constraints (47) include LMIs (48), (51) and (52), if the optimal input design problem has at least one solution; otherwise, let $\hat{r}_0 = \hat{r}_{\text{max}}$ and $\hat{r}_\tau = 0$ for all $\tau > 0$.

5) Input filter update. Compute the corresponding stable minimum phase input filter (56) given by spectral factorization of the corresponding input spectrum $\Psi_u(e^{i\omega}, \hat{r}(\theta_n))$.

6) Input generator update. Compute the state-space realization of transfer function (56), which is represented by linear time-varying system (6) with $\{A_z, B_z, C_z, D_z\}$ of the form either (54) or (55).

7) Measurement update. Apply the input signal $u_{n+1}$ generated by (6) to the system (1) and collect a new measurement $y_{n+1}$ from the system (1).

8) Parameter estimate update. The update recursive estimate

$$
\theta_{n+1} = [\theta_{A,n+1}^{T} \theta_{B,n+1}^{T} \theta_{F,n+1}^{T} \theta_{C,n+1}^{T} \theta_{D,n+1}^{T}]^{T}
$$

is computed by (34)-(40).

9) Iterate. Replace $n$ by $n + 1$ and go to step 4).

By Theorems 4.1 and 4.2 we immediately have the following formal result for the convergence of the adaptive scheme Algorithm 5.1:

**Theorem 5.1:** Suppose that Assumptions 1-3 and 6-7 hold. Then $\{\theta_n\}$ generated by Algorithm 5.1 satisfies

$$
\theta_n - \theta^* \to 0 \quad \text{a.s.} \quad \text{and} \quad \sqrt{n} (\theta_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0_{p_a}, P^*)
$$

as $n \to \infty$, where $P^*$ is the covariance matrix given by (44).

Theorem 5.1 shows that the adaptive Algorithm 5.1 asymptotically recovers the same accuracy as if the optimal input were used.

VI. NUMERICAL ILLUSTRATION: $L_2$-GAIN ESTIMATION

The problem of $L_2$-gain estimate for FIR systems has been studied in [22], Section 6. In this section, we consider a class of ARARX systems with $p_a = 0$, $p_b \geq 2$ and $p_d \geq 1$ (see Remark
As in [22], the objective is to obtain a certain accuracy of an estimate of the squared $L_2$-gain

$$|G^*|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G^*(e^{i\omega})|^2 d\omega = \theta_B^* T \theta_B^*$$

of the system transfer function $G^*(q) = B^*(q)$ at the end of an experiment of length $N$, and at the same time use as little input power as possible. This problem can be formulated as follows (see [22])

$$\min_u \mathbb{E}[u_n^2]$$

s.t. $\text{Var}[\hat{G}_N]^2 \leq \gamma,$

(57)

where $\text{Var}[\cdot]$ is the variance operator with respect to the probability measure, $\hat{G}_N(q) = G(q, \theta_{B,N})$ represents the estimated transfer function with the truncated estimate of $\theta_B$ and the input signal is generated by linear time-varying system (6).

As in [22], we set the order $m = p_b - 1 > 0$ of the generator (6) of input signal. Note that (16) and (22) give $\xi_n = \xi_n(\theta, r) = y_n(r) + \xi_{\theta,n}^T(\theta, r) \theta, \xi_{\theta,n} = \xi_{\theta,n}(\theta, r) = [-\tilde{u}_{n-1}^T(\theta_B) \tilde{v}_{n-1}^T(\theta_B)]^T$ and $y_n(r) = \tilde{u}_{n-1}^T(\theta_B) \theta^* - \tilde{v}_{n-1}^T(\theta_B) \theta_D + e_n$, which with (15) yields

$$\tilde{v}_{n-1}(\theta_B) = \begin{bmatrix} y_{n-1}(r) - \tilde{u}_{n-1}^T(\theta_B) \\ \vdots \\ y_{n-p_d}(r) - \tilde{u}_{n-p_d}^T(\theta_B) \\ \end{bmatrix} = \begin{bmatrix} \tilde{u}_{n-2}^T(r) \\ \vdots \\ \tilde{u}_{n-p_d}(r) \\ \end{bmatrix} (\theta_B^* - \theta_B) - \begin{bmatrix} \tilde{v}_{n-1}^T(\theta_B) \\ \vdots \\ \tilde{v}_{n-p_d}^T(\theta_B) \end{bmatrix} \theta_D + \begin{bmatrix} e_{n-1} \\ \vdots \\ e_{n-p_d} \end{bmatrix}.$$

(58)

In the limit $n \to \infty$, the equation (24) is given as

$$\mathbb{E}[\xi_{\theta,n}^T(\theta, r) \xi_n(\theta, r)] = \mathbb{E}[\xi_{\theta,n}^T(\theta, r) \xi_{\theta,n}(\theta, r)] \theta + \mathbb{E}[\xi_{\theta,n}(\theta, r) y_n(r)]$$

$$= \begin{bmatrix} \mathbb{E}[\tilde{u}_{n-1}^T(\theta_B) \tilde{u}_{n-1}^T(\theta_B)] \theta_B^* - \theta_B \\ \vdots \\ \mathbb{E}[\tilde{v}_{n-1}^T(\theta_B) \tilde{v}_{n-1}^T(\theta_B)] \theta_B^* - \theta_B \\ \end{bmatrix} \theta_D + \begin{bmatrix} \mathbb{E}[\tilde{u}_{n-1}^T(\theta_B) \tilde{v}_{n-1}^T(\theta_B)] \theta_B^* - \theta_B \\ \vdots \\ \mathbb{E}[\tilde{v}_{n-1}^T(\theta_B) \tilde{v}_{n-1}^T(\theta_B)] \theta_B^* - \theta_B \\ \end{bmatrix} \theta_D$$

$$= 0,$$

(59)

that is,

$$R_u(\theta_B - \theta_B^*) + R_B \theta_D = [R_u(\theta_B + R_D)(\theta_B - \theta_B^*) = 0,$$

$$R_B^T(\theta_B - \theta_B^*) + \mathbb{E}[\tilde{v}_{n-1}(\theta_B) \tilde{v}_{n-1}^T(\theta_B)] \theta_D - \mathbb{E}[\tilde{v}_{n-1}(\theta_B) \tilde{v}_{n-1}^T(\theta_B)] \theta_D = 0,$$

(60a) (60b)
where $R_u(r) > 0$ is the Toeplitz matrix given by (51) with $r_j = r_{-j} = E[u_n u_{n-j}]$ and $r_j = 0$ for $|j| > m$,

$$R_B = \begin{bmatrix}
\sum_{j=1}^{p_b} r_j (b_j - b^*_j) & \sum_{j=1}^{p_b} r_{j+1}(b_j - b^*_j) & \cdots & \sum_{j=1}^{p_b} r_{j+p_d-1}(b_j - b^*_j) \\
\sum_{j=1}^{p_b} r_{j-1}(b_j - b^*_j) & \sum_{j=1}^{p_b} r_j (b_j - b^*_j) & \cdots & \sum_{j=1}^{p_b} r_{j+p_d-2}(b_j - b^*_j) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{j=1}^{p_b} r_{j-p_b+1}(b_j - b^*_j) & \sum_{j=1}^{p_b} r_{j-p_b+2}(b_j - b^*_j) & \cdots & \sum_{j=1}^{p_b} r_{j+p_d-p_b}(b_j - b^*_j)
\end{bmatrix}, \quad (61)$$

and

$$R_D = \begin{bmatrix}
\sum_{j=1}^{pd} d_j r_j & \sum_{j=1}^{pd} d_{j+1} r_{j+1} & \cdots & \sum_{j=1}^{pd} d_{j+p_d-1} r_{j+p_d-1} \\
\sum_{j=1}^{pd} d_{j-1} r_{j-1} & \sum_{j=1}^{pd} d_j r_j & \cdots & \sum_{j=1}^{pd} d_{j+p_d-2} r_{j+p_d-2} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{j=1}^{pd} d_{j-p_d+1} r_{j-p_d+1} & \sum_{j=1}^{pd} d_{j-p_d+2} r_{j-p_d+2} & \cdots & \sum_{j=1}^{pd} d_{j+p_d-p_d} r_j
\end{bmatrix}. \quad (62)$$

It is observed that (60a) has the unique solution $\theta_B = \theta^*_B$, i.e., $b_j = b^*_j$ for $1 \leq j \leq p_b$ when

$$R_u(r) + \frac{1}{2}(R_D^T + R_D) > 0. \quad (63)$$

Then, in this case, (60b) gives

$$R^*_u(\theta_D - \theta^*_D) = 0, \quad (64)$$

which has the unique solution $\theta_D = \theta^*_D$ since $R^*_u = E[\tilde{v}_{n-1}(\theta^*_B) \tilde{v}_{n-1}^T(\theta^*_B)] > 0$ (see (68) below). Therefore, $\theta = \theta^*$ is the unique solution to normal equation (59) on any compact set $D_\theta \subset \mathbb{R}^{p_0}$ with $\theta^* \in \text{int}D_\theta$. Let us consider $\theta_0 \in D_0 = \{\theta : |\theta_0| \leq K_{D_0}\}$ for some given $K_{D_0} \geq 0$ and $D_\theta = \{\theta : g(\theta) \leq K_\theta\}$ with continuous function $g(\theta) = |\theta \varepsilon(\theta)|^2$. Then there exists a constant $K_D = K_D(D_0) > 0$ such that $\theta^* \in \text{int}D_\theta$ and Assumption [21] is satisfied for all $K_\theta \geq K_D$.

We may choose $K_\theta$ sufficiently large and implement our proposed algorithm. Alternatively, in this case, where $\theta = \theta^*$ is the unique solution to (59) and hence the unique asymptotically stable equilibrium point of (41a) on $D_\theta$ for all $K_\theta > 0$, it is easy to verify that our proposed method (34)-(40) can work with an expanding truncation domain $D_\theta = \{\theta : g(\theta) \leq K_j\}$ (see [12] and [20]), in which $K_\theta$ will be increased from $K_j$ to $K_{j+1}$ if the estimate $(\theta_n, R_n)$ goes out of $D_\theta$ with $K_j$ and hence is reset to $(\theta_0, R_0)$, where $\{K_j\}$ is a sequence of positive real numbers increasingly diverging to infinity. Algorithm 5.1 with an adaptive truncation domain will be illustrated in the following numerical example.
A. The optimization problem

In the identification procedure, the new input of each step is determined by the solution of the optimization problem \(57\). Obviously, with the parametrization described above, the objective function \(E[\varepsilon_n^2]\) in \(57\) equals the first auto-correlation coefficient \(\hat{r}_0\). As suggested in [22], the variance constraint \(\text{Var}[|\tilde{G}_N|^2] \leq \gamma\) may be replaced by using a linear approximation of \(|G(q, \theta_N)|^2\) around the true value

\[
|\tilde{G}_N|^2 = |G^*|^2 + 2\theta_B^*T(\theta_B - \theta^*) (1 + \bar{\varepsilon}_N) \tag{65}
\]

where \(\bar{\varepsilon}_N = o(1)\) is the bounded error term such that all finite moments of \(\bar{\varepsilon}_N\) converge to 0 when \(\bar{\theta}_N - \theta^*\) tends to 0, which implies that the variance of the squared \(L_2\)-gain can be written as

\[
\text{Var}[|\tilde{G}_N|^2] = 4\theta_B^*T\text{Cov}[\theta_B, \theta^*] + \text{tr}\text{Cov}(\theta_B, \theta^*) \cdot o(1) \tag{66}
\]

with \(o(1) \rightarrow 0\) as \(N \rightarrow \infty\), where \(\text{Cov}[\cdot]\) is the covariance operator with respect to the probability measure. According to Theorem 5.1, the original variance constraint may be replaced by an approximation

\[
4\theta_B^*T\frac{\sigma_e^2}{N}(R_u^*)^{-1}\theta_B^* \leq \gamma, \tag{67}
\]

where \(R_u^* \in \mathbb{R}^{p_b \times p_b}\) is the principal submatrix of \(R^*\) and therefore

\[
P^* = \sigma_e^2(R^*)^{-1} = \sigma_e^2 \begin{bmatrix} R_u^* & 0 \\ 0 & R_v^* \end{bmatrix}^{-1} = \sigma_e^2 \begin{bmatrix} (R_u^*)^{-1} & 0 \\ 0 & (R_v^*)^{-1} \end{bmatrix}. \tag{68}
\]

The inequality (67), by Schur complements, can be expressed as

\[
\begin{bmatrix} R_u^* & 2\theta_B^* \\ 2\theta_B^{*T} & \frac{\gamma N}{\sigma_e^2} \end{bmatrix} \succeq 0. \tag{69}
\]

In the adaptive input design context, at each step we replace the true value \(\theta^*\) with the estimate \(\theta_n = [\theta_{B,n}^T, \theta_{D,n}^T]^T\). Therefore, the optimization problem that is solved at iteration step
n is formulated as

$$\min_{r, Q} \quad r_0$$

s.t. $$\begin{bmatrix} R_u(r) & 2\theta_{B,n} \\ 2\theta_{B,n}^T & \frac{\gamma N}{\sigma_e^2} \end{bmatrix} \geq 0$$

$$K(Q, \{A_u, B_u, C_u(r), D_u(r)\}) \geq \beta_K I_{p_b}$$

$$Q \geq 0$$

$$R_u(r) \geq \beta_R I_{p_b}$$

$$\hat{r}_{\max} \geq r_0 \geq \hat{r}_{\min}$$

$$R_u(r) + \frac{1}{2} [R^T_D(\theta_{D,n}) + R_D(\theta_{D,n})] \geq \beta_D I_{p_b},$$

where $\beta_K$, $\beta_R$ and $\beta_D$ are small positive numbers set to ensure the positivity condition (48), the persistent excitation condition (51) and the unique solution condition (63).

As in [22], the optimization is made with the MATLAB toolbox YALMIP ([45]) and the solver sdpt3 ([60]). Spectral factorization is used to compute the coefficients $\hat{f}_j(\theta_n)$, for $0 \leq j \leq m$, of input filter (56) from the solution $\hat{r}(\theta_n)$ of optimization problem (70). And then the input signal generator (6) of the form (54)-(55) is obtained as a state-space realization of the transfer function (56). The conditions of Theorem 5.1 are satisfied for the procedure described above, which implies that the parameter estimates will converge to the true value almost surely and the asymptotic accuracy for the adaptive design will be the same as for the optimal input.

B. Simulation results

The true parameters of the ARARX system with orders $p_a = 0$, $p_b = 4$ and $p_d = 3$ are $\theta^*_B = (0.9, 0.6, 0.2, 0.3)^T$, $\theta^*_D = (-1.2, 0.75, -0.2)^T$ and $\sigma_e^2 = 0.1$, which is derived from the FIR numerical example in [22]. As in [22], we set the order $m = 3$ for the linear time-varying system (6). In the following simulations, we employ the algorithm (34)-(40) with expending truncation domain and choose $D_\theta = \{\theta : |\theta\varepsilon(\theta)|^2 \leq K_j\}$ with $K_j = 10 + j$ and $j \geq 0$, $D_R$ with $\kappa_1 = 10^{-6}$ and $\kappa_2 = 10^{10}$, initial value $\theta_0 = [\theta_{B,0}^T, \theta_{D,0}^T]^T = 0_I^T$ and $R_0 = I_7$.

The total experiment length $N = 4 \times 10^3$, the required accuracy $\gamma = 10^{-4}$, $\beta_K = \beta_R = \hat{r}_{\min} = 10^{-3}$, $\hat{r}_{\max} = 5$ and $\beta_D = 10^{-6}$. Figs. 1 and 2 show a typical realization of Algorithm 5.1 for
estimates of $\theta_B$ and $X = [\theta_D^T \sigma_e^2]^T$, respectively, while Figs. 3 and 4 show a typical realization of algorithm (34)-(40) with optimal input signal that is generated by (6) with parameters obtained by solving optimization problem (70) with $\theta_n = \theta^*$. The realization of power $r_0$ corresponding to Figs 1-2 as well as $r_0$ of the optimal input (55), is shown in Fig. 5. And Fig. 6 shows the variance of the estimated $\mathcal{L}_2$-gain, $\text{Var}(|\tilde{G}_N|^2_2)$, estimated from 100 Monte Carlo simulations with the adaptive input and the optimal input, respectively.
Fig. 5. Solid line: the realization of auto-relation $r_0$ for Algorithm 5.1 corresponding to Figs. 1-2. Dotted line: auto-relation $r_0$ of the optimal input (55).

Fig. 6. Variance of the estimated $L_2$-gain, $\text{Var}(|\bar{G}_N|^2)$. Solid line: variance estimated from Monte Carlo simulations with the adaptive input. Dotted line: variance estimated from Monte Carlo simulations with the optimal input.

VII. CONCLUSION

This paper is a formal development of the scheme outlined in [21]. We have proposed an adaptive input design method for stable LTI systems based on the certainty equivalence principle. As an application example, we studied the adaptive input design problem of $L_2$-gain estimation for a class of ARARX systems. A numerical example was conducted to illustrate and verify the effectiveness of our proposed method.

APPENDIX A. PROOF OF LEMMA 3.1

Clearly, given any $\theta \in D_\theta$, $E[\bar{\varepsilon}_{\theta,n}(\theta, r)\bar{\varepsilon}_{\theta,n}^T(\theta, r)] \in \mathbb{R}^{p_{\theta} \times p_{\theta}}$ is a symmetric positive semidefinite matrix. Furthermore, $E[\bar{\varepsilon}_{\theta,n}(\theta, r)\bar{\varepsilon}_{\theta,n}^T(\theta, r)] = E[\varphi_n(\theta, r)\varphi_n^T(\theta, r)]$ is not symmetric positive definite if and only if there exists a nonzero vector $\nu \in \mathbb{R}^{p_{\theta}}$ such that $\nu^T E[\bar{\varepsilon}_{\theta,n}(\theta, r)\bar{\varepsilon}_{\theta,n}^T(\theta, r)]\nu = E[\nu^T \bar{\varepsilon}_{\theta,n}(\theta, r)\bar{\varepsilon}_{\theta,n}^T(\theta, r)\nu] = E[|\nu^T \bar{\varepsilon}_{\theta,n}(\theta, r)|^2] = E[|\nu^T \varphi_n(\theta, r)|^2] = 0$.

Let

$$G^*(q) = \frac{B^*(q)}{A^*(q)F^*(q)} \quad \text{and} \quad \bar{H}^*(q) = \frac{C^*(q)}{A^*(q)D^*(q)},$$

then $y_n(r) = \bar{G}^*(q)u_n(r) + \bar{H}^*(q)e_n$. By (22), we observe that

$$\bar{\varepsilon}_{\theta,n}(\theta, r) = \bar{\varepsilon}_{\theta,n}(\theta, r) = \begin{bmatrix} F_u(q, \theta) & F_e(q, \theta) \\ u_n(r) & e_n \end{bmatrix}^T = F_u(q, \theta)u_n(r) + F_e(q, \theta)e_n \quad (71)$$
and hence $\mathbb{E}[\varepsilon_{\theta,n}(\theta^*)\varepsilon_{\theta,n}^T(\theta^*)]$ are continuous on $D_N$ since both $F_u(q,\theta)$ and $F_e(q,\theta)$ are continuous on $D_N$, where

$$F_u(q,\theta) = \begin{bmatrix} F_{u\theta}(q) & -F_{u\theta}(q) & F_{uu}(q,\theta) & -F_{u\theta}(q,\theta) & F_{u\theta}(q,\theta) \end{bmatrix}^T,$$

$$F_e(q,\theta) = \begin{bmatrix} F_{e\theta}(q) & 0^T_{p_n} & 0^T_{p_f} & -F_{e\theta}(q,\theta) & F_{e\theta}(q,\theta) \end{bmatrix}^T,$$

$$F_{uu}(q,\theta) = \begin{bmatrix} q^{-1} & \cdots & q^{-p_a} \end{bmatrix} G(q), \quad F_{u\theta}(q) = \begin{bmatrix} q^{-1} & \cdots & q^{-p_a} \end{bmatrix},$$

$$F_{uu}(q,\theta) = \begin{bmatrix} q^{-1} & \cdots & q^{-p_a} \end{bmatrix} G(q), \quad F_{u\theta}(q) = \begin{bmatrix} q^{-1} & \cdots & q^{-p_a} \end{bmatrix}.$$

Particularly, we have

$$\varepsilon_{\theta,n}(\theta^*) = F_u(q,\theta^*)u_n(r) + F_e(q,\theta^*)\epsilon_n, \quad (72)$$

where $F_u(q,\theta^*) = \begin{bmatrix} F_{u\theta}(q) & -F_{u\theta}(q) & F_{uu}(q,\theta) & -F_{u\theta}(q,\theta) & F_{u\theta}(q,\theta) \end{bmatrix}$ and

$$F_e(q,\theta^*) = \begin{bmatrix} F_{e\theta}(q) & 0^T_{p_n} & 0^T_{p_f} & -F_{e\theta}(q,\theta^*) & F_{e\theta}(q,\theta^*) \end{bmatrix}^T.$$

Note that $\{u_n\}$ is generated by $\mathbf{6}$ with $\{s_n\}$ independent of $\{\epsilon_n\}$. For any nonzero vector $\nu \in \mathbb{R}^{p_0}$, we have

$$\nu^T \varepsilon_{\theta,n}(\theta,r) = \nu^T F_u(q,\theta^*)u_n(r) + \nu^T F_e(q,\theta^*)\epsilon_n, \quad (73)$$

and, by Parseval’s formula,

$$\mathbb{E}[|\nu^T \varepsilon_{\theta,n}(\theta^*)|^2] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \nu^T F_u(e^{i\omega},\theta^*) \nu^T F_e(e^{i\omega},\theta^*) \right| \begin{bmatrix} \Psi_u(\omega) & 0 \\ 0 & \sigma_e^2 \end{bmatrix} \begin{bmatrix} F_u^T(e^{i\omega},\theta^*) \nu \\ F_e^T(e^{i\omega},\theta^*) \nu \end{bmatrix} d\omega. \quad (74)$$

Since Assumption $[2]$ implies that there does not exist a vector $\nu \neq 0_{p_0}$ such that $\nu^T F_u(e^{i\omega},\theta^*) = 0$ for almost all $\omega$, (74) with $\mathbf{10}$ yields $\mathbb{E}[|\nu^T \varepsilon_{\theta,n}(\theta^*)|^2] > 0$ for any nonzero vector $\nu \in \mathbb{R}^{p_0}$, or say, $\mathbb{E}[\varepsilon_{\theta,n}(\theta^*)\varepsilon_{\theta,n}^T(\theta^*)] > 0$. Since Assumption $[2]$ holds on some neighborhood of $\theta^*$ (see Remark $[2.2]$), it follows the desired result.
APPENDIX B. NOTATIONS IN ADAPTIVE SYSTEM (34)-(40)

\[ \eta_n = [\varepsilon_{n+1} \ s_n]^T, \quad \Phi_n = [\Phi_{1,n}^T \ \Phi_{2,n}^T \ \Phi_{3,n}^T \ \Phi_{4,n}^T \ \Phi_{5,n}^T \ \Phi_{6,n}^T \ \Phi_{7,n}^T \ \Phi_{8,n}^T]^T \in \mathbb{R}^{m+p_0+n_x+2} \]

with

\[ \Phi_{1,n} = z_n, \ \Phi_{2,n} = \tilde{u}_{n-1}, \ \Phi_{3,n} = \varepsilon_n, \ \Phi_{4,n} = \xi_n, \ \Phi_{5,n} = \tilde{y}_{n-1}, \ \Phi_{6,n} = \tilde{w}_{n-1}, \ \Phi_{7,n} = \tilde{v}_{n-1}, \ \Phi_{8,n} = \tilde{v}_{n-1}, \]

\[
A_\Phi(\theta) = \begin{bmatrix}
A_z(r(\theta)) & 0 & 0 & 0 & 0 & 0 & 0 \\
\tilde{C}_z & \tilde{I}_u & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
B_\xi C_z(r(\theta)) & 0 & K_\xi & A_\xi & 0 & 0 & 0 \\
0 & \tilde{I}_y & \tilde{C}_y & 0 & 0 & 0 & 0 \\
0 & \tilde{B}_w^u & 0 & 0 & 0 & 0 & 0 \\
0 & \tilde{B}_v^u & \tilde{I}_v & \tilde{C}_v & 0 & \tilde{F}_v & \tilde{I}_v & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \tilde{D}_\xi & \tilde{C} \\
\end{bmatrix}, \quad B_\Phi(\theta) = \begin{bmatrix}
0 & B_z(r(\theta)) \\
0 & \tilde{D}_z \\
1 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
\end{bmatrix},
\]

\[ Q(\theta, \Phi_{n+1}) = -R_n^{-1}(\theta)\varepsilon_{\theta,n+1}(\theta) \left( C_\xi \Phi_{4,n+1} + \Phi_{4,n+1} + \theta^T \varepsilon_{\theta,n+1} \right), \]

\[ R_{n+1}(\theta) = \frac{n}{n+1} R_n(\theta) + \frac{1}{n+1} \varepsilon_{\theta,n+1}\varepsilon_{\theta,n+1}^T \]

with \( \varepsilon_{\theta,n}(\theta) = \begin{bmatrix} \Phi_{5,n} \\ -\Phi_{2,n} \\ \Phi_{6,n} \\ -\Phi_{8,n} \\ \Phi_{7,n} \end{bmatrix} \in \mathbb{R}^{p_0}, \quad \tilde{C}_z = \begin{bmatrix} C_z(r(\theta)) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{p_0 \times m}, \quad \tilde{I}_u = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \mathbb{R}^{p_0 \times p_0}, \]

\[
\tilde{D}_z = \begin{bmatrix} D_z(r(\theta)) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{p_0}, \quad \tilde{I}_y = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{p_0}, \quad \tilde{C}_y = \begin{bmatrix} C_\xi \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{p_0 \times n_x}, \quad \tilde{B}_w^u = \begin{bmatrix} \theta_B^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{p_f \times p_0},
\]

\[
\tilde{B}_v^u = \begin{bmatrix} -\theta_B^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{p_d \times p_0}, \quad \tilde{I}_v = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{p_d}, \quad \tilde{C}_v = \begin{bmatrix} C_\xi \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{p_d \times n_x}, \quad \tilde{F}_v = \begin{bmatrix} \theta_F^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{p_f \times p_d},
\]

October 16, 2013  
DRAFT
\[
\begin{pmatrix}
0 & 0 \\
I_{pd-1} & 0
\end{pmatrix} \in \mathbb{R}^{pd \times pd}, \quad
\begin{pmatrix}
-\theta_B^T \\
0 \\
\vdots \\
0
\end{pmatrix} \in \mathbb{R}^{pc \times pe}, \quad
\begin{pmatrix}
1 \\
0 \\
\vdots \\
0
\end{pmatrix} \in \mathbb{R}^{pc}, \quad
\begin{pmatrix}
C_\xi \\
0 \\
\vdots \\
0
\end{pmatrix} \in \mathbb{R}^{pc \times n_\xi},
\]

\[
\tilde{I}_v = \begin{pmatrix}
0 & 0 \\
I_{pd-1} & 0
\end{pmatrix} \in \mathbb{R}^{pd \times pd}, \quad
\tilde{B}_u = \begin{pmatrix}
-\theta_B^T \\
0 \\
\vdots \\
0
\end{pmatrix} \in \mathbb{R}^{pc \times pe}, \quad
\tilde{I}_e = \begin{pmatrix}
1 \\
0 \\
\vdots \\
0
\end{pmatrix} \in \mathbb{R}^{pc}, \quad
\tilde{C}_\xi = \begin{pmatrix}
C_\xi \\
0 \\
\vdots \\
0
\end{pmatrix} \in \mathbb{R}^{pc \times n_\xi}.
\]

\[
\tilde{F}_v = \begin{pmatrix}
\theta_F^T \\
0 \\
\vdots \\
0
\end{pmatrix} \in \mathbb{R}^{pc \times pf} \quad \text{and} \quad
\tilde{D}_v = \begin{pmatrix}
\theta_D^T \\
0 \\
\vdots \\
0
\end{pmatrix} \in \mathbb{R}^{pc \times pd}.
\]

\section*{Appendix C. Some Useful Results in Literature}

\textbf{Definition C.1:} A random process \( \{\bar{s}_n\}_{n \geq 0} \) is said to be M-bounded, which is denoted by \( \bar{s}_n = O_M(1) \), if \( M_q(\bar{s}) = \sup_{n \geq 0} E_1^{1/q}[|\bar{s}_n|^q] < \infty \) for all \( 1 \leq q < \infty \).

Suppose that \( \{t_n\} \) is a sequence of positive numbers. We also write \( \bar{s}_n = O_M(t_n) \) if \( \bar{s}_n/t_n = O_M(1) \).

\textbf{Definition C.2:} A random process \( \{\bar{s}_n\}_{n \geq 0} \) is \( L \)-mixing with respect to the \( \sigma \)-algebras \((F_n, F_n^+) \), \( n \geq 0 \), if the following conditions are satisfied:

\begin{enumerate}
  \item \( \bar{s}_n \) is \( F_n \) measurable,
  \item \( \bar{s}_n = O_M(1) \),
  \item \( \sum_{t=0}^{\infty} \gamma_q(t) < \infty \) for all \( 1 \leq q < \infty \), where
    \[
    \gamma_q(t) = \sup_{n \geq t} E_1^{1/q}[|\bar{s}_n - E[\bar{s}_n|F_n^+]|^q], \quad t \geq 0.
    \]
\end{enumerate}

Some useful theorems derived from the main results in \cite{19}, \cite{20} and \cite{28} are given as follows, which are applied to develop our results in this paper.

\textbf{Condition C.1:} The noise \( \{\eta_n\} \) in the system (37) is a sequence of independent random variables such that

\[
\sup_n E[\exp(\alpha_n|\eta_n|^2)] < \infty \quad (75)
\]

holds for some \( \alpha_n > 0 \).

\textbf{Condition C.2:} The time-varying system (37) is bounded input-bounded output (BIBO) stable.
Condition C.3: The ODE (41) has an asymptotically stable equilibrium point $X_\ast \in \text{int} D_X$ with $D_X \subset D_\ast$, where $D_X = D_\theta \times D_R$ and $D_\ast$ is the domain of attraction of $X_\ast$. The initial estimate $X_0$ is in the interior of $D_{X_0} = D_0 \times D_R$, where $D_0$ is a compact set defined by (42).

Condition C.4: The families of matrices $A_\Phi(\theta)$ and $B_\Phi(\theta)$, $\theta \in D_\theta$, are twice continuously differentiable with bounded partial derivatives up to second order in $D_\theta$.

Condition C.5: Denote by $X(t; \bar{t}, \bar{X})$ the solution to ODE (41) for $t \geq \bar{t} \geq 0$ with $X_{\bar{t}} = \bar{X}$. Assume that (41) has a unique equilibrium point $X_\ast \in \text{int} D_{X_0}$ on $D_X$ and $\bar{X} \in \text{int} D_{X_0}$, where $D_{X_0} \subset \text{int} D_X$ is a compact convex set that is invariant for (41) and $\{X(t; \bar{t}, \bar{X}) : t > \bar{t} \geq 0, \bar{X} \in D_{X_0}\} \subset \text{int} D_{X_0}$. Moreover, for every $\bar{X} \in D_{X_0}$, we have the Lyapunov exponent $-\alpha < -1/2$, i.e., there is a constant $\bar{C}_0 > 0$ such that
\[
\left| \frac{\partial}{\partial \bar{X}} X(t; \bar{t}, \bar{X}) \right| \leq \bar{C}_0 \exp \left( -\alpha(t - \bar{t}) \right)
\] (76) for all $t > \bar{t} \geq 0$.

The special case of [28, Theorem 3.1] with $M = 1$ is cited as

Theorem C.1: If Conditions C.1-C.3 hold, then process $\{X_n\}$ with $X_n = (\theta_n, R_n)$ computed by the recursive stochastic algorithm (34)-(40) converges to $X_\ast$ a.s. as $n \to \infty$.

A variant of [19, Theorem 4.1] (see also [20, Theorem 3.3]) is given as follows

Theorem C.2: Assume that Conditions C.1, C.2, C.4 hold and Condition C.5 also holds with $\bar{X} = X_0 = (\theta_0, R_0) \in \text{int} D_{X_0}$. Then $\{X_n\}$ with $X_n = (\theta_n, R_n)$ computed by the recursive stochastic algorithm (34)-(40) satisfies
\[
X_n - X_\ast = O_M(n^{-1/2}).
\] (77)

APPENDIX D. PROOF OF THEOREM 4.1

It is observed that, by (2) and (7), Condition C.1 is satisfied. Let us consider Condition C.2, i.e., the BIBO stability of the linear time-varying system (37). According to Theorem 2.1 [47], the switching system (37) is BIBO stable if and only if it is uniformly exponentially stable, or equivalently, uniformly asymptotically stable (see, e.g., [15]). Clearly, the BIBO stability is guaranteed by the joint stability of $A_\Phi(\theta)$ for all $\theta \in D_\theta$, that is, there exist a symmetric positive definite matrix $V_\Phi$ and a constant $\lambda_\Phi \in (0, 1)$ such that
\[
A_\Phi^T(\theta) V_\Phi A_\Phi(\theta) \leq \lambda_\Phi V_\Phi
\] (78)
for all $\theta \in D_\theta$ (see [19] Condition 4.1 and [20] Condition 3.7]). But the BIBO stability of system (37) is also ensured when the bounded set of matrices $\Sigma_\phi = \{A_\phi(\theta) : \theta \in D_\theta\}$ is LCP (left convergent products), i.e., every left-infinite product $\lim_{n \to \infty} A_n \cdots A_2 A_1$ converges, where $A_k \in \Sigma_\phi$ for all $k = 1, 2, \cdots, n$ (see [8] and [31]). In this work, we will show the BIBO stability of system (37) by applying an important result of the joint spectral radius (see [14], [8], [31] and the references therein). Let $\Sigma^n_\phi = \{A_n \cdots A_2 A_1 : A_k \in \Sigma_\phi, k = 1, 2, \cdots, n\}$. It is easy to observe that every product $\bar{A}_{\phi,n} \in \Sigma^n_\phi$ is a lower triangular matrix of the form

$$
\bar{A}_{\phi,n} = \begin{bmatrix}
\bar{A}_{z,n} & 0 & 0 & 0 & 0 & 0 & 0 \\
* & (\bar{I}_u)^n & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
* & * & * & (A_\xi)^n & 0 & 0 & 0 \\
* & * & * & 0 & 0 & 0 & 0 \\
* & * & * & * & F_n & 0 & 0 \\
* & * & * & * & * & (\bar{I}_v)^n & 0 \\
* & * & * & * & * & * & \bar{C}_n \\
\end{bmatrix}
$$

(79)

for all $n \geq 1$, where $\bar{A}_{z,n} \in \Sigma^n_z$, $\bar{C}_n \in \Sigma^n_C$, $\bar{F}_n \in \Sigma^n_F$ and the entries denoted by * can be zero or nonzero. Obviously, $(\bar{I}_u)^n$ and $(\bar{I}_v)^n$ are strictly lower triangular matrices (i.e., lower triangular matrices having zeros along their main diagonals) for all $n \geq 1$. In fact, there is a positive integer $n_0$ such that $(\bar{I}_u)^n = 0$ and $(\bar{I}_v)^n = 0$ for all $n \geq n_0$ since both $\bar{I}_u$ and $\bar{I}_v$ are nilpotent matrices. Note that the transition matrix $A_\xi$ has all its eigenvalues strictly inside the unit circle, that is, the spectral radius of $A_\xi$,

$$
\rho(A_\xi) = \max\{|\lambda_k| : \lambda_k \text{ is an eigenvalue of } A_\xi\} < 1,
$$

(80)

which yields $\rho((A_\xi)^n) = [\rho(A_\xi)]^n \leq \rho(A_\xi) < 1$ for all $n \geq 1$. Therefore, we have

$$
\rho_n(\Sigma_\phi) = \sup\{[\rho(\bar{A})]^{1/n} : \bar{A} \in \Sigma^n_\phi\} = \max\{\rho_n(\Sigma_z), \rho(A_\xi), \rho_n(\Sigma_C), \rho_n(\Sigma_F)\}
$$

(81)

for all $n \geq 1$, where $\rho_n(\Sigma_z)$, $\rho_n(\Sigma_C)$ and $\rho_n(\Sigma_F)$ are given in (8) and (29), respectively. But this combined with (8), (29) and (80) immediately implies

$$
\rho(\Sigma_\phi) = \limsup_{n \to \infty} \rho_n(\Sigma_\phi) < 1.
$$

(82)

By [31 Corollary 1.1, p21] (see also [8]), the switching system (37) is (uniformly asymptotically) stable and therefore BIBO stable. Therefore, Condition C.2 is satisfied.
Let us proceed to show the asymptotic stability of the associated ODE (41). Since $G(\theta)$ is continuous on the compact $D_\theta$ (see Appendix A), there exists $\kappa_0 > 0$ such that $0 \leq G(\theta) < \kappa_0 I_{p_0}$ for all $\theta \in D_\theta$. Note that ODE (41b) with initial value $R_0 > 0$ gives

$$R_t = e^{-\frac{1}{2}tI_{p_0}}R_0e^{-\frac{1}{2}tI_{p_0}^\tau} + e^{-\frac{1}{2}tI_{p_0}}\left[\int_0^t e^{\frac{1}{2}I_{p_0}^\tau G(\theta)\tau}e^{\frac{1}{2}I_{p_0}^\tau}d\tau\right] e^{-\frac{1}{2}tI_{p_0}}$$

(83)

for all $t \geq 0$, which yields

$$\kappa_\tau e^{-tI_{p_0}} < e^{-\frac{1}{2}tI_{p_0}^\tau}R_0 e^{-\frac{1}{2}tI_{p_0}^\tau} \leq R_t \leq R_0 + e^{-\frac{1}{2}tI_{p_0}^\tau}\left[\int_0^t e^{\frac{1}{2}I_{p_0}^\tau \kappa_0I_{p_0}e^{\frac{1}{2}I_{p_0}^\tau}d\tau}\right] e^{-\frac{1}{2}tI_{p_0}}$$

(84)

for all $t \in [0, \infty)$, where $\kappa_2 = \kappa_R + \kappa_0$ with $\kappa_R I_{p_0} > R_0 > \kappa_\tau I_{p_0} > 0$. This implies that

$$\kappa_2^{-1} I_{p_0} < R_t^{-1} < \kappa_\tau^{-1} e^{tI_{p_0}}, \quad \forall \ t \in [0, \infty).$$

(85)

Recall that the asymptotic cost function $W(\theta)$ defined by (18) has exactly one minimum $\theta^*$ on $D_\theta$ since (24) has the unique solution $\theta = \theta^*$. Obviously, $W(\theta) \geq W(\theta^*) > 0$ for all $\theta \in D_\theta$. By (41a) and Assumption 7, we observe

$$\frac{d}{dt}W(\theta) = -W^\tau_\theta(\theta)R_t^{-1}W_\theta(\theta) \leq -\lambda_m(R_t^{-1})|W_\theta(\theta)|^2 < -\kappa_2^{-1}|W_\theta(\theta)|^2$$

(86)

for all $t \in [0, \infty)$, and, particularly,

$$\frac{d}{dt}W(\theta) \leq -\kappa_2^{-1}w_{\theta^*}^2 < 0$$

for all $\theta \in D_\theta \setminus D_{\theta^*}$, where $w_{\theta^*} = \inf_{\theta \in D_\theta \setminus D_{\theta^*}}|W_\theta(\theta)|$. This implies that there is a finite positive constant $t_{\theta^*} \leq W(\theta_0)/(\kappa_2^{-1}w_{\theta^*}^2)$ such that $\theta_t \in D_{\theta^*}$ for all $t \geq t_{\theta^*}$. By Lemma 3.1 there is a positive constant $\kappa_{\theta^*}$ such that $G(\theta) \geq \kappa_{\theta^*} I_{p_0}$ for all $\theta \in D_{\theta^*}$. This combined with (83) and (84) gives

$$R_t \geq R_{t_{\theta^*}} + e^{-\frac{1}{2}tI_{p_0}}\left[\int_{t_{\theta^*}}^t e^{\frac{1}{2}I_{p_0}^\tau \kappa_{\theta^*}I_{p_0}e^{\frac{1}{2}I_{p_0}^\tau}d\tau}\right] e^{-\frac{1}{2}tI_{p_0}} > \kappa_\tau e^{-t_{\theta^*}I_{p_0}} + \kappa_{\theta^*} e^{-t_{\theta^*}I_{p_0}}$$

(87)

for all $t \geq t_{\theta^*}$. But (84) and (87) immediately yield $R_t > \kappa_1 I_{p_0}$ for all $t \geq 0$, where $\kappa_1 = \kappa_\tau e^{-t_{\theta^*}} > 0$. This combined with (84) and (85) gives

$$\kappa_1 I_{p_0} < R_t < \kappa_2 I_{p_0} \quad \text{and} \quad \kappa_2^{-1} I_{p_0} < R_t^{-1} < \kappa_1^{-1} I_{p_0}$$

(88)

for all $t \geq 0$, where positive constants $\kappa_1$ and $\kappa_2$ can be used to define $D_R$ in (35). So (86) holds for all $t \geq 0$. But, according to [38 VIII. Theorem, p66], this implies that the equilibrium $\theta^*$ of
(41a) is asymptotically stable, which also yields $R_t \to R^* = G(\theta^*)$ as $t \to \infty$. Therefore, the equilibrium $(\theta^*, R^*)$ of ODE (41) is asymptotically stable. But this with Assumption 7 implies that Condition C.3 holds. By Theorem C.1, $\{(\theta_n, R_n)\}$ computed by the recursive algorithm (34)-(40) converges to $(\theta^*, R^*)$ a.s. as $n \to \infty$.

Finally, we show (43) as follows. Note that Assumption 8 implies Condition C.4 and the Jacobian matrix of (41) at $(\theta^*, R^*)$ has the structure
\[
\begin{pmatrix}
-I_{p\theta} & 0 \\
* & -I_p
\end{pmatrix},
\]
all eigenvalues of which are equal to $-1$. It follows that Condition C.5 is satisfied with the Lyapunov exponent $-\alpha = -1 + c$ for any $c > 0$ in some invariant neighborhood of $(\theta^*, R^*)$ (see also proof of [19, Theorem 4.2]). Let $D_{\theta,R}$ be a compact convex invariant neighborhood such that $(\theta^*, R^*) \in \text{int} D_{\theta,R}$ and Condition C.5 is satisfied with the Lyapunov exponent $-\alpha < -1/2$. The proof of [28, Theorem 3.1] shows that there exists a sample dependent finite number $N_{\theta,R}$ such that $\{(\theta_n, R_n)\}_{n \geq N_{\theta,R}} \subset \text{int} D_{\theta,R}$ almost surely. Let us consider the sequence $\{(\theta_n, R_n)\}_{n \geq N_{\theta,R}}$.

But, by Theorem C.2, $\{(\theta_n, R_n)\}_{n \geq N_{\theta,R}}$ satisfies
\[
\theta_n - \theta^* = O_M((n - N_{\theta,R})^{-1/2}) \quad \text{and} \quad R_n - R^* = O_M((n - N_{\theta,R})^{-1/2})
\]
a.s. as $n \to \infty$. It is noticed that
\[
n^{1/2} = O_M((n - N_{\theta,R})^{1/2})
\]
a.s. as $n \to \infty$ since $\mathbb{P}\{N_{\theta,R} < \infty\} = 1$. So, by Cauchy-Schwarz inequality, (90) and (91) imply (43), which completes the proof.

**APPENDIX E. PROOF OF THEOREM 4.2**

Since, according to the proof of Theorem 4.1 (see Appendix D), the switching system (37) is BIBO stable and hence is uniformly exponentially stable, there are $C_\Phi > 0$ and $\lambda_\Phi \in (0, 1)$ such that (see [47])
\[
|\Phi_n| \leq C_\Phi \lambda_\Phi^n |\Phi_0| + \sum_{k=1}^{n} C_\Phi \lambda_\Phi^k |\eta_{n-k}| =: \hat{\Phi}_n.
\]
Therefore, $\{\hat{\Phi}_n\}$ and hence $\{\Phi_n\}$ are $L$-mixing processes since $\{e_n\}$, $\{w_n\}$ and hence $\{\eta_n\}$ are $L$-mixing processes (see [18]). It follows that the process
\[
\Delta \Phi_n = \Phi_n - \Phi_n^*
\]
is $L$-mixing, where $\{\Phi_n^\ast\}$ is generated by (37) with $\theta_n = \theta^\ast$ and $\varepsilon_{\theta,n} = \varepsilon_{\theta,n}(\theta^\ast) = -\varphi_n(\theta) = \left[\tilde{y}_{n-1}(\theta^\ast) - \tilde{u}_n T_{n-1}(\theta^\ast) - \tilde{\xi}_{n-1}(\theta^\ast) \tilde{v}_{n-1}(\theta^\ast)\right]^T$. Moreover, since system (37) is uniformly exponentially stable and, by Theorem 4.1, $\Delta \Phi_k \rightarrow 0$ a.s. as $n \rightarrow \infty$, we have $\Delta \Phi_n = o_M(1)$, i.e., $\Delta \Phi_n \rightarrow 0$ in $L_q$-norm for all $q \geq 1$. Clearly, this yields that

$$\Delta \varepsilon_{\theta,n} = \varepsilon_{\theta,n} - \varepsilon_{\theta,n}^\ast$$

(94)
is an $L$-mixing process and $\Delta \varepsilon_{\theta,n} = o_M(1)$ since

$$\Delta \varepsilon_{\theta,n} = [\Delta \Phi_{5,n}^T - \Delta \Phi_{2,n}^T \Delta \Phi_{6,n}^T - \Delta \Phi_{8,n}^T \Delta \Phi_{T,n}^T]^T,$$

where $\Delta \Phi_{k,n} = \Phi_{k,n} - \Phi_{k,n}^\ast$ for $k = 1, 2, \ldots, 8$.

Note that $\{\varepsilon_{\theta,n}\}$ is an $L$-mixing process and therefore

$$\frac{1}{n} \sum_{k=1}^{n} \varepsilon_{\theta,k}^\ast \varepsilon_{\theta,k}^T \rightarrow R^\ast \quad \text{a.s.}$$

(95)

and hence in law as $n \rightarrow \infty$. Moreover, by Cauchy-Schwarz inequality, we observe

$$\frac{1}{n} \sum_{k=1}^{n} \left[\varepsilon_{\theta,k}^\ast \Delta \varepsilon_{\theta,k}^T + \Delta \varepsilon_{\theta,k} \varepsilon_{\theta,k}^T + \Delta \varepsilon_{\theta,k} \Delta \varepsilon_{\theta,k}^T\right] \rightarrow 0,$$

(96)

$$\frac{1}{n} \sum_{k=1}^{n} \varepsilon_{\theta,k} \Delta \tilde{w}_{k-1}^T = \frac{1}{n} \sum_{k=1}^{n} \left(\varepsilon_{\theta,k}^\ast + \Delta \varepsilon_{\theta,k}\right) \Delta \tilde{w}_{k-1}^T \rightarrow 0,$$

(97)

$$\frac{1}{n} \sum_{k=1}^{n} \varepsilon_{\theta,k} \Delta \tilde{v}_{k-1}^T = \frac{1}{n} \sum_{k=1}^{n} \left(\varepsilon_{\theta,k}^\ast + \Delta \varepsilon_{\theta,k}\right) \Delta \tilde{v}_{k-1}^T \rightarrow 0,$$

(98)

$$\frac{1}{n} \sum_{k=1}^{n} \varepsilon_{\theta,k} \Delta \tilde{v}_{k-1}^T = \frac{1}{n} \sum_{k=1}^{n} \left(\varepsilon_{\theta,k}^\ast + \Delta \varepsilon_{\theta,k}\right) \Delta \tilde{v}_{k-1}^T \rightarrow 0,$$

(99)
in $L_q$ for any $q \geq 1$ and hence in law as $n \rightarrow \infty$. But, since both $\{\varepsilon_{\theta,n}\}$ and $\{\varepsilon_{\theta,n}^\ast\}$ are $L$-mixing processes, (95) and (96) give

$$\frac{1}{n} \sum_{k=1}^{n} \varepsilon_{\theta,k} \varepsilon_{\theta,k}^T \rightarrow R^\ast$$

(100)
in $L_q$ for any $q \geq 1$ and hence in law as $n \rightarrow \infty$. And the combination of (11), (24) and (97)-(99) yields

$$\left(\frac{1}{n} \sum_{k=1}^{n} \varepsilon_{\theta,k} \varepsilon_{\theta,k}^T\right) (\theta_n - \theta^\ast) = \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{\theta,k} (\theta^\ast - \Delta \tilde{w}_{k-1}^T \theta^\ast + \Delta \tilde{v}_{k-1}^T \theta^\ast - \Delta \tilde{v}_{k-1}^T \theta^\ast - e_k)$$

$$\rightarrow -\frac{1}{n} \sum_{k=1}^{n} \varepsilon_{\theta,k} e_k - \frac{1}{n} \sum_{k=1}^{n} \left(\varepsilon_{\theta,k}^\ast + \Delta \varepsilon_{\theta,k}\right) e_k - \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{\theta,k} e_k$$

(101)
in $L_q$ for any $q \geq 1$ and hence in law as $n \to \infty$. But, by a martingale central limit theorem (see, e.g., [26, Theorem 3.2, p58]), we have

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \varepsilon_{\theta,k} e_k \Rightarrow \mathcal{N}(0_p, \sigma^2 e^2 R^*) \quad \text{as} \quad n \to \infty. \quad (102)$$

Recall that the sequence $\{\varepsilon_{\theta,k}\}_{1 \leq k \leq n}$ is $\mathcal{F}_{n-1}$ measurable for all $n \geq 1$, where $\varepsilon_{\theta,k}$ is the online version of $\varepsilon_{\theta,k}$ defined by (22). So, by the martingale central limit theorem, the combination of (100), (101) and (102) yields the desired result (45). The proof is complete.

REFERENCES

[1] M. Agrawal, P. Stoica and P. Åhgren, Common factor estimation and two applications in signal processing, *Signal Processing*, vol.84, pp.421-429, 2004.

[2] B. Alkire and L Vandenberghe, Convex optimization problems involving finite autocorrelation sequences. *Mathematical Programming, Serial A*, vol.93 pp.331-359, 2002.

[3] R. Almeida, E. Pueyo, J.P. Martínez, A.P. Rocha and P. Laguna, Quantification of the QT variability related to HRV: robustness study facing automatic delineation and noise on the ECG. *Computers in Cardiology* (pp.769-772) 2004.

[4] A.C. Atkinson and R.A. Bailey, One hundred years of the design of experiments on and off the pages of Biometrika, *Biometrika*, vol.88, pp.45-97, 2001.

[5] S. Bandara, J.P. Schlöder, R. Eils, H.G. Bock and T. Meyer, Optimal experiment design for parameter estimation of a cell signaling model, *PLoS Computational Biology*, vol.5, pp.1-12, 2009.

[6] N. Barabanov, Lyapunov exponent and joint spectral radius: some known and new results. *In Proc. of the 44th IEEE Conference on Decision and Control and European Control Conference* (pp.2332-2337) 2005.

[7] A. Benveniste, M. Métivier and P. Priouret, *Adaptive algorithms and stochastic approximations*. Berlin, Germany: Springer-Verlag, 1990.

[8] M.A. Berger and Y. Wang. Bounded semigroups of matrices, *Linear Algebra and its Applications*, vol.166, pp.2127, 1992.

[9] K. Bernaerts, K.P.M. Gysemans, T. Nhan Minh and J.F. Van Impe, Optimal experiment design for cardinal values estimation: guidelines for data collection, *International Journal of Food Microbiology*, vol.100, pp.153-165, 2005.

[10] X. Bombois, G. Scorletti, M. Gevers, R. Hildebrand and P. Van den Hof, Cheapest open-loop identification for control. *In Proc. of the 43th IEEE conference on decision and control* (pp.382-387) 2004.

[11] V.S. Borkar, *Stochastic approximation: a dynamical systems viewpoint*, Cambridge, UK: Cambridge University Press, 2008.

[12] H.-F. Chen, *Stochastic approximation and its applications*, New York, USA: Kluwer Academic Publishers, 2002.

[13] H.-F. Chen and L. Guo, *Identification and stochastic adaptive control*. Boston, USA: Birkhäuser, 1991.

[14] I. Daubechies and J.C. Lagarias, Sets of matrices all infinite products of which converge, *Linear Algebra and its Applications*, vol.161, pp.227-263, 1992.

[15] W.P. Dayawansa and C.F. Martin, A converse Lyapunov theorem for a class of dynamical systems which undergo switching, *IEEE Trans. Automatic Control*, vol.44, pp.751-760, 1999.

[16] R. Diversi, R. Guidorzi and U. Soverini, Identification of ARX and ARARX models in the presence of input and output noises, *European J. Control*, vol.16, pp.242255, 2010.
[17] H.A. Dror and D.M. Steinberg, Robust experimental design for multivariate generalized linear models, *Technometrics*, vol.48, pp.520-529, 2006.
[18] L. Gerencsér, On a class of mixing processes, *Stochastics*, vol.26, pp.165-191, 1989.
[19] L. Gerencsér, Rate of convergence of recursive estimators, *SIAM J. Control and Optimization*, vol.30, pp.1200-1227.
[20] L. Gerencsér, A representation theorem for the error of recursive estimators, *SIAM J. Control and Optimization*, vol.44, pp.2123-2188, 2006.
[21] L. Gerencsér and H. Hjalmarsson, Adaptive input design in system identification. In *Proc. of the 44th IEEE Conference on Decision and Control and European Control Conference* (pp.4988-4993) 2005.
[22] L. Gerencsér, H. Hjalmarsson and J. Mårtensson, Identification of ARX systems with non-stationary inputs – asymptotic analysis with application to adaptive input design, *Automatica*, vol.45, pp.623-633, 2009.
[23] M. Gevers, X. Bombois, R. Hildebrand and G. Solari, Optimal experiment design for open and closed-loop system identification, *Communications in Information and Systems*, vol.11, pp.197-224, 2011.
[24] M. Haest, G. Bastin, M. Gevers and V. Wertz, ESPION: an expert system for system identification, *Automatica*, vol.26, pp.85-95, 1990.
[25] J. Hahn, K. Hirano and D. Karlan, Adaptive experimental design using propensity score, *Journal of Business & Economic Statistics*, vol.29, pp.96-108, 2011.
[26] P. Hall and C.C. Heyde, *Martingale limit theory and its application*, New York, USA: Academic Press, 1980.
[27] H. Hjalmarsson, From experiment design to closed-loop control, *Automatica*, vol.41, pp.393-438, 2005.
[28] L. Huang and H. Hjalmarsson, Recursive estimators with Markovian jumps, *Systems & Control Letters*, vol.61, pp.1009-1016, 2012.
[29] L. Huang, H. Hjalmarsson and L. Gerencsér, Adaptive experiment design for ARMAX systems. To appear in *Proc. of the 51st IEEE Conference on Decision and Control*.
[30] H. Jansson and H. Hjalmarsson, Input design via LMIs admitting frequency-wise model specifications in confidence regions, *IEEE Trans. Automatic Control*, vol.50, pp.1534-1549, 2005.
[31] R. Jungers, *The joint spectral radius: theory and application*, Berlin, Germany: Springer-Verlag, 2009.
[32] R. Kohn, Asymptotic estimation and hypothesis testing results for vector linear time series models, *Econometrica*, vol.47, pp.1005-1030, 1979.
[33] H. Kushner, Stochastic approximation: a survey, *Wiley Interdisciplinary Reviews: Computational Statistics*, vol.2, pp.87-96, 2010.
[34] H.J. Kushner and G. Yin, *Stochastic approximation and recursive algorithms and applications (2nd edition)*. New York, USA: Spring, 2003.
[35] T.L. Lai, Sequential analysis: some classical problems and new challenges, *Statistica Sinica*, vol.11, pp.303-408, 2001.
[36] T.L. Lai and C.-Z. Wei, Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *The Annals of Statistics*, vol.10, pp.154-166, 1982.
[37] T.L. Lai and C.-Z. Wei, Extended least squares and their applications to adaptive control and prediction in linear systems, *IEEE Trans. Automatic Control*, vol.31, pp.898-906, 1986.
[38] J. LaSalle and S. Lefschetz, *Stability by Liapunov’s direct method with applications*. New York, USA: Academic Press, 1961.
[39] K. Lindqvist and H. Hjalmarsson, Identification for control: adaptive input design using convex optimization. In *Proc. of the 40th IEEE Conference on Decision and Control* (pp.4326-4331) 2001.
[40] L. Ljung, Analysis of recursive stochastic algorithms, *IEEE Trans. Automatic Control*, vol.22, pp.551-575, 1977.
[41] L. Ljung and P.E. Caines, Asymptotic normality of prediction error estimators for approximate system models, *Stochastics*, vol.3, pp.29-46, 1979.
[42] L. Ljung, *System identification: Theory for the user (2nd edition)*. New Jersey, USA: Prentice Hall, 1999.
[43] L. Ljung and T. Söderström, *Theory and practice of recursive identification*. Massachusetts, USA: MIT Press, 1983.
[44] T. Lohmann, H.G. Bock and J.P. Schloeder, Numerical methods for parameter estimation and optimal experiment design in chemical reaction systems, *Industrial & Engineering Chemistry Research*, vol.31, pp.54-57, 1992.
[45] J. Löfberg, YALMIP: A toolbox for modeling and optimization in MATLAB, *Proceedings of IEEE International Symposium on Computer Aided Control System Design*, 2004.
[46] H. Lütkepohl, *Handbook of matrices*. Chichester, UK: John Wiley & Sons, 1996.
[47] G. Michaletzky and L. Gerencsér, BIBO stability of linear switching systems, *IEEE Trans. Automatic Control*, vol.47, pp.1895-1898, 2002.
[48] L.T. Muftuler and O. Nalcioglu, FMRI Signal Modeling Using System Identification Techniques. *Proc. Intl. Soc. Mag. Reson. Med.*, 8, 2000.
[49] G. Nollo, A. Porta, L. Faes, M.D. Greco, M. Disertori and F. Ravelli, Causal linear parametric model for baroreflex gain assessment in patients with recent myocardial infarction, *AJP-Heart and Circulatory Physiology*, vol.280, pp.H1830-H1839, 2001.
[50] A. Nooraii, J.A. Romagnoli and J. Figueroa, Process identification, uncertainty characterisation and robustness analysis of a pilot-scale distillation column, *J. Process Control*, vol.9, pp.247-264, 1999.
[51] M. Noriega, J.P. Martíne, P. Laguna, R. Bailón and R. Almeida, Respiration effect on Wavelet-based ECG T-wave end delineation strategies, *IEEE Trans. Biomedical Engineering*, to appear.
[52] V.M. Pérez, J.E. Renaud and L.T. Watson, Adaptive experimental design for construction of response surface approximations, *Proceedings of the 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, AIAA-2001-1622, 2001.
[53] A. Porta, G. Baselli, E. Caiani, A. Malliani, F. Lombardi and S. Cerutti, Quantifying electrocardiogram RT-RR variability interactions, *Medical & Biological Engineering & Computing*, vol.36, pp.27-34, 1998.
[54] L. Pronzato, Optimal experimental design and some related control problems, *Automatica*, vol.44, pp.303-325, 2008.
[55] W.J. Rugh, *Linear system theory*. New Jersey, USA: Prentice Hall, 1996.
[56] M. Sasena, M. Parkinson, P. Goovaerts, P. Papalambros and M. Reed, Adaptive experimental design applied to an ergonomics testing procedure. In *Proceedings of DETC’02, ASME 2002 Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, DETC2002/ DAC34091.
[57] T. Söderström and P. Stoica, *System identification*. New York, USA: Prentice Hall, 1989.
[58] P. Stoica, T. Mckelvey and J. Mari, MA estimation in polynomial time, *IEEE Trans. Signal Processing*, vol.48, pp.1999-2012, 2000.
[59] F. Tjärnström and L. Ljung, $L_2$ model reduction and variance reduction, *Automatica*, vol.38, pp.1517-1530, 2002.
[60] K. C. Toh, R.H. Tütüncü and M.J. Todd, On the implementation and usage of SDPT3a Matlab software package for semidefinite-quadratic-linear programming, version 4.0. (2006) Available from [http://www.math.nus.edu.sg/mattohkc/sdpt3.html](http://www.math.nus.edu.sg/mattohkc/sdpt3.html).
[61] H.P. Wynn, The sequential generation of $D$-optimum experimental designs, *The Annals of Mathematical Statistics*, vol.5, pp.1655-1664.

October 16, 2013 DRAFT