Adaptive Input Design for LTI Systems

László Gerencsér, Håkan Hjalmarsson, *Fellow, IEEE,* and Lirong Huang

Abstract—**Optimal input design for parameter estimation has obtained extensive coverage in the past. A key problem here 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 this problem, there are few results that address dynamical systems. This paper presents sufficient conditions for convergence/consistency and asymptotic optimality for a class of adaptive systems consisting of a recursive prediction error estimator and an input generator depending on the time-varying parameter estimates. The results apply to a general family of single input single output linear time-invariant systems. An important application is adaptive input design for which the results imply that, asymptotically in the sample size, the adaptive scheme recovers the same accuracy as the off-line prediction error method that uses data from an experiment where perfect knowledge of the system has been used to design an optimal input spectrum.**

Index Terms—**Linear time-invariant (LTI), recursive prediction error (RPE), single-input single-output (SISO).**

I. INTRODUCTION

WITH the rapid developments in model-based engineering, compare with the petrochemical industry where it is reported that all plants employ model predictive control, the high cost of modeling is coming more and more into focus as a limiting factor [1]. Often the only practical means to modeling is data-driven modeling, i.e., system identification. For this type of modeling, the major part of the cost is associated with performing experiments on the plant in question. A key variable here is the duration of the experiments since it strongly couples to costs in terms of personnel, energy, material and production losses.

For dynamical systems, it has been shown that careful design of the experiment can lead to quite drastic reduction in the required experimental time as compared to standard white noise excitation or step testing [2], [3]. It has also been stressed that the experimental conditions are essential for making system identification robust with respect to many of the design variables that are involved, e.g., model structure and orders, and with respect to the resulting end performance [4].

The aforementioned observations have prompted renewed interest in optimal experiment design—a topic that has been studied extensively over the past decades, see, e.g., [5]–[12] and references therein. Recent advances include novel computationally tractable algorithms [13], least-costly and application oriented frameworks [14], [4], closed-loop methods [15]–[19], and extensions to non-linear models [20]–[22].

A key problem in optimal experiment design is that the optimal experiment typically depends on the system parameters that are to be identified. One of the fundamental routes to cope with this problem is to employ adaptive schemes, meaning that as information from the system is gathered the experimental conditions are changed. Adaptive design is usually called sequential design in the statistics literature, where there exist a rich collection of results and applications (see, e.g., [23] and the references therein).

When only the input excitation is considered part of the experiment design, we will use the terminology input design. Adaptive input design has been studied in many works in engineering literature (see, e.g., [24]–[28] and [29]). However, as pointed out in [30] and [28], there are few results that address this problem for dynamical systems. Given the increasing practical relevance of input design, it is becoming urgent to provide a solid theoretical foundation for such methods.

When the system is linear time-invariant and belongs to the model set, and the input is (quasi-)stationary, it is only the second-order properties of the input that asymptotically (in the sample size) influence the model quality. Thus, in this case, it is the spectrum, or equivalently the autocorrelation sequence, of the input that is the design variable in optimal input design. The actual input sequence can be generated by filtering white noise through an input spectrum shaping filter corresponding to a stable spectral factor of the optimal input spectrum [13]. Building on this, an obvious approach to adaptive input design is to combine a recursive identification scheme with a time-varying input spectrum shaping filter, computed from the solution of the optimal input design problem using the most recent model estimate as a substitute for the true system.

Such a certainty equivalence approach leads to an adaptive feedback system where, similar to adaptive control, the input properties change over time depending on the response of the system. From a performance perspective, there are several issues that are non-trivial to analyze:

    (i) Under which conditions will the parameter estimates of such a procedure converge?

(ii) If the algorithm converges, will it be consistent, i.e. will the model parameters correspond to the true system parameters?

(iii) If the algorithm converges to a correct system description, how does the resulting (large sample) accuracy compare to the accuracy an oracle, having access to the unknown true parameters for the experiment design already at the beginning of the experiment, could achieve?

In regards to (iii), notice that even if the parameters converge to the true values so that, as the experiment time progresses towards infinity, the input behaves closer and closer to a stationary signal having the optimal spectrum, suboptimal experimental conditions prevail in the meantime and it is not evident that the algorithm is able to catch up with the accumulated loss of accuracy this causes—this strongly depends on the rate of convergence of the algorithm.

An early version of the above concept was presented in [24]. A severe limitation was that the parameter estimation was not recursive, requiring re-identification using all past data for each new measurement. Furthermore, no statistical analysis was provided and even if, for this off-line algorithm, (i) and (ii) can be dealt with rather straightforwardly using results from [31], (iii) is non-trivial to analyze since the input signal is non-stationary. Subsequently the recursive certainty equivalence approach, adopted in this contribution, was outlined in [27], but without details of formal treatment of (i)–(iii), the technical foundations of which were given in the paper [32] that appeared a few months later. Recently, [28] takes 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 [24]. The advantage of using ARX-models is that the analysis of the recursive least-squares method can be carried out with a powerful result in [33].

There exists an extensive body of literature on general recursive stochastic algorithms, e.g. [34]–[40]. However, the available results, representing variations of what is loosely called the ODE method, do not seem to be applicable to prove a certainity equivalence principle for adaptive input design in general. The key technical advance needed for the proof of such results has been developed in [32] by proving a very strong form of an ODE principle for general recursive estimation schemes introduced in [34] and discussed in much details in [36].

Building on this work, the objective of this paper is to provide the complete theoretical foundations of the adaptive input design framework outlined in [27], providing results for (i)–(iii), hereby validating current practice in input design.

While we will cover (i) and (ii), our primary objective will be to deal with (iii). In particular, with θ^*, $\hat{\theta}_n$ and $\hat{\hat{\theta}}_n^*$ denoting the true parameter vector, the parameter estimate in the adaptive algorithm, and the off-line parameter estimate obtained from an experiment using the optimal input, respectively, we will be interested in establishing conditions for when adaptive input design asymptotically yields the same asymptotic accuracy as the optimal non-adaptive design in the sense that

$$\sqrt{n}(\hat{\theta}_n - \theta^*)$$

and

$$\sqrt{n}(\hat{\hat{\theta}}_n^* - \theta^*)$$

have the same asymptotic covariance matrix. A pre-requisite for this is (of course) that the recursive estimation algorithm is able to achieve this when the optimal input is used. In proving these kind of results, we formulate and prove a novel, significantly stronger form of certainty equivalence: namely, we show that $\hat{\theta}_n - \hat{\hat{\theta}}_n^* = o(n^{-1/2})$ in a sense explained below, see Theorem 4.1. Another ambition has been to cover the general class of single-input single-output (SISO) linear time-invariant (LTI) systems and associated model structures considered in [41]. The recursive prediction error (RPE) approach [34], [36] fulfills these objectives. However, in its initial form, this algorithm requires a so-called boundedness condition ensuring some kind of stability of the state process associated with the RPE algorithm and a projection mechanism to ensure that the estimators stay in a compact domain, and one generally cannot exclude the possibility that the sequence of estimates gets trapped at the boundary where the projection takes place. A variety of ideas have emerged to rectify these deficiences. In this paper we follow the method put forward in [32], [39], where stability of the state-process is ensured assuming a specific form of joint stability of the state-transition matrices, and the projection is replaced by a resetting mechanism which allows almost sure convergence to the true parameter vector to be established together with a rate of convergence for the moments of the estimation error. A restrictive assumption here is that the asymptotic prediction error criterion is only allowed to have the true parameter vector as stationary point. This is a more severe condititon than identifiability. However, for a method that, as in the case of RPE, is based on gradient based non-linear search the best one can hope for is that convergence takes place to the set of stationary points. Notice that the corresponding off-line result [31], which proves convergence to the global minimum, makes the assumption that the global minimum can be found-something that is not easy to guarantee in practice using gradient based methods, on-line as well as off-line. However, as our focus is (iii), which has convergence to the true system parameters as a pre-requisite, we have chosen to base our algorithm and analysis on the work [39], [32], thus avoiding the issue of clustering at the boundary. Recently, a novel recursive algorithm for ARMAX models has been proposed in [40] using expanding truncations for which an almost sure convergence result has been established. Unfortunately, for our considerations, this convergence result applies only when the input is white and, furthermore, the asymptotic accuracy of this algorithm is not known, and hence, at least at present, this algorithm is not suited to our purpose.

The paper starts off in Section II by introducing the system and model assumptions, together with the input signal generation mechanism that will be employed. The latter depends on the estimated parameter vector. Off-line prediction error identification is discussed in Section III. The complete adaptive algorithm, comprising the true system, the recursive estimation algorithm and the input generator, is presented in Section IV. Formal results on convergence/consistency and asymptotic distribution are provided at the end of the section. These results are quite general in that they make no specific use of the functional

relationship between the parameter estimate and the input generator, other than that this is a sufficiently smooth map. These results are then placed in the context of adaptive input design in the following Section V, where a complete adaptive input design algorithm is presented, together with the result that this algorithm achieves the same asymptotic accuracy as an oracle. The algorithm is illustrated on a numerical example in Section VI. Conclusions are provided in Section VII. Most of the proofs are provided in appendices.

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 Ω is the sample space, $\mathcal{F}$ is the σ-algebra that defines events E in Ω which are measurable, i.e., for which the probability $\mathbb{P}(E)$ is defined. Let $\mathbb{E}[\cdot]$ 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 the square matrix P is nonsingular, its inverse is denoted by P^{-1}. I_m stands for the identity matrix of dimension m, $0_{m \times n}$ stands for the zero matrix of dimensions $m \times n$, $0_m = 0_{m \times 1}$ stands for the zero vector of dimension m, and 0 denotes the zero matrix of appropriate dimensions. Denote by $\rho(\cdot)$ the maximum eigenvalue, minimum eigenvalue, and spectral radius of a matrix, respectively. For a vector, let $\| \cdot \|$ denote the Euclidean norm and for a matrix the norm induced by the Euclidean norm. Unless explicitly stated, matrices are assumed to have real entries and compatible dimensions.

II. LTI SYSTEM, INPUT SIGNAL AND MODEL

For the sake of clarity we present our results for single-input-single-output (SISO) systems. Extension to multi-input-multi-output systems is straightforward along the lines of [27].

Let us consider a SISO, finite-dimensional linear stochastic control-system described compactly by the state-space equations

$$\xi_{n+1}^y = A^y \xi_n^y + B^y u_n + K^y e_n,$$
$$y_n = C^y \xi_n^y + e_n. \tag{1}$$

Here, $u = (u_n)$ is the control-signal, $e = (e_n)$ is the noise-signal, $y = (y_n)$ is the observed output signal, and $\xi^y = (\xi_n^y)$ is the state-variable, all of them real-valued scalars, except possibly the real-valued state-vector ξ_n^y. For the sake of simplicity we assume that the system is at rest prior to time $n = 0$, in particular $u_n = 0, e_n = 0$ and $\xi_n^y = 0$ for $n < 0$. This will be no limitation in terms of the asymptotic analysis we will consider.

As the system will be operating in open loop, stability will be required.

Condition 2.1: The system matrix A^y has all its eigenvalues strictly inside the unit circle.

The noise signal is subject to the following condition.

Condition 2.2: The noise process is $e = (e_n), n \geq 0$ is an independent and identically distributed (i.i.d.) sequence with zero-mean and a positive finite variance, say σ_e^2, i.e.

$$\mathbb{E}[e_n] = 0, \qquad 0 < \mathbb{E}[e_n^2] = \sigma_e^2 < \infty.$$

Moreover, for some $c > 0$ we have

$$\mathbb{E}[e^{c\, e_n^2}] < \infty. \tag{2}$$

The input signal u is the output of the following state-space equations:

$$\bar{\xi}_{n+1}^u(\eta) = A^u(\eta)\bar{\xi}_n^u + B^u(\eta)w_n,$$
$$\bar{u}_n(\eta) = C^u(\eta)\bar{\xi}_n^u(\eta) + D^u(\eta)w_n. \tag{3}$$

Here, η is a parameter vector that influences the state-space matrices and hence can be used to control the spectrum of the input signal; it will be tuned over time by the adaptive algorithm. However, for the time being η will be considered fix, taking its values from an open set $D_\eta \subset \mathbb{R}^{n_\eta}$. To indicate signals that are generated by time-invariant filters we use overline notation, e.g. $\bar{u}_n(\eta)$ as in (3) and $\bar{y}_n(\eta)$ for y_n in (1) when u is generated by (3) with η fix.

The signal w driving u in (3) is subject to the following condition.

Condition 2.3: The random process $w = (w_n)$, $n \geq 0$ is an i.i.d. sequence with zero-mean and variance 1, i.e., for all $n \geq 0$, we have

$$\mathbb{E}[w_n] = 0, \qquad \mathbb{E}[w_n^2] = 1.$$

Moreover, the processes $w = (w_n)$ and $e = (e_n), n \geq 0$ are independent. Finally, for some $c > 0$, we have

$$\mathbb{E}[e^{c\, w_n^2}] < \infty. \tag{4}$$

The objective is to identify a model of the transfer functions from the input u and the noise e to the output y of the system (1). However, despite that we only are interested in the input-output relationships, we will use a model of state-space type as it will be used in a time-varying fashion in our adaptive algorithm. Its state-space matrices $A(\theta), B(\theta), C(\theta)$, and $K(\theta)$ are parametrized by a, to be identified, parameter vector θ belonging to some open domain $D_\theta \in \mathbb{R}^{n_\theta}$. Thus, the model is given by

$$\bar{\xi}_{n+1}(\theta; \eta) = A(\theta)\bar{\xi}_n(\theta; \eta) + B(\theta)\bar{u}_n(\eta) + K(\theta)e_n,$$
$$\bar{y}_n(\theta; \eta) = C(\theta)\bar{\xi}_n(\theta; \eta) + e_n. \tag{5}$$

Such a model can be expressed in transfer function form as

$$y_n = G(q, \theta)u_n + H(q, \theta)e_n$$

where $G(q, \theta)$ and $H(q, \theta)$ are rational functions in the backwards shift operator q^{-1}. This model class encompass a wide variety of standard black-box structures such as output-error, Box-Jenkins and ARMAX, as well as parametrized state-space models. It also includes tailor-made transfer function and state-space models, e.g. where the parameters have physical interpretations. For details, we refer to [41].

III. OFF-LINE PREDICTION ERROR IDENTIFICATION

To identify the system parameter θ we proceed mostly along standard lines of arguments. We provide a summary for the sake of defining our notations and conditions.

A. The Prediction Error

Notice that the noise e in (1) can be computed as

$$\xi^y_{n+1} = A^y \xi^y_n + B^y u_n + K^y (y_n - C^y \xi^y_n)$$

$$= (A^y - K^y C^y) \xi^y_n + B^y u_n + K^y y_n,$$

$$e_n = -C^y \xi^y_n + y_n.$$

For a pair of tentative values $\theta \in D_\theta$ and $\eta \in D_\eta$, let us therefore define the estimated noise process $\bar{\varepsilon}_n(\theta; \eta)$ for $n \geq 0$ by

$$\bar{\xi}_{n+1}(\theta; \eta) = (A(\theta) - K(\theta)C(\theta))\bar{\xi}_n(\theta; \eta)$$

$$+ B(\theta)\bar{u}_n(\eta) + K(\theta)\bar{y}_n(\eta),$$

$$\bar{\varepsilon}_n(\theta; \eta) = -C(\theta)\bar{\xi}_n(\theta; \eta) + \bar{y}_n(\eta), \qquad (6)$$

with 0 initial conditions. The cost function based on N observations $y_0, \cdots, y_{N-1}$, obtained as the negative conditional likelihood function in the case when e is Gaussian, modulo constants, is defined by

$$V_N(\theta; \eta) = \sum_{n=0}^{N-1} \frac{1}{2} \bar{\varepsilon}_n^2(\theta; \eta). \qquad (7)$$

The off-line estimator $\hat{\theta}_N(\eta)$ is defined as the global minimizer of $V_N(\theta; \eta)$ w.r.t. θ in D_θ.

B. The Prediction Error Gradient

To minimize the cost function (7) in practice we need to compute its gradient w.r.t. θ. Then we get for the partial derivatives w.r.t. θ, the column vector

$$V_{\theta,N}(\theta; \eta) := \frac{\partial}{\partial \theta} V_N(\theta; \eta) = \sum_{n=0}^{N-1} \bar{\varepsilon}_{\theta,n}(\theta; \eta)\bar{\varepsilon}_n(\theta; \eta) \qquad (8)$$

where $\bar{\varepsilon}_{\theta,n}(\theta; \eta) = \partial \bar{\varepsilon}_n(\theta; \eta)/\partial \theta$. In standard procedures, the estimate is obtained by setting the partial derivatives w.r.t. θ equal to 0, i.e., solving the equation

$$V_{\theta,N}(\theta; \eta) = 0. \qquad (9)$$

A rigorous definition of $\hat{\theta}_N(\eta)$ in this context, taking account the possibility of multiple solutions or no solutions, was given in [42].

For actual computations, we need the gradient process $\bar{\varepsilon}_{\theta,n}(\theta; \eta)$. Denote by subscript θ_k the partial differential with respect to an arbitrary element of θ. Then differentiation of (6) yields that the complete gradient of $\bar{\varepsilon}_n(\theta; \eta)$ with respect to θ can be generated by a set of state-space systems, indexed by $k = 1, \ldots, n_\theta$, coupled in parallel

$$\bar{\xi}_{\theta_k,n+1}(\theta; \eta) = (A(\theta) - K(\theta)C(\theta))\bar{\xi}_{\theta_k,n}(\theta; \eta)$$

$$+ (A(\theta) - K(\theta)C(\theta))_{\theta_k}\bar{\xi}_n(\theta; \eta)$$

$$+ B_{\theta_k}(\theta)\bar{u}_n(\eta) + K_{\theta_k}(\theta)\bar{y}_n(\eta),$$

$$\bar{\varepsilon}_{\theta_k,n}(\theta; \eta) = -C(\theta)\bar{\xi}_{\theta_k,n}(\theta; \eta) - C_{\theta_k}(\theta)\bar{\xi}_n(\theta; \eta). \qquad (10)$$

Here, we notice that $\bar{\varepsilon}_{\theta_k,n}(\theta; \eta)$ is the output of a state-space system with inputs $\bar{\xi}_n(\theta; \eta), \bar{u}_n(\eta)$ and $\bar{y}_n(\eta)$, and $A(\theta) - K(\theta)C(\theta)$ as state-transition matrix.

Note that the state transition matrix of the overall system above is block-diagonal, with diagonal blocks equal to $A(\theta) - K(\theta)C(\theta)$. A time-varying version of the above system will be incorporated into our adaptive algorithm, see (24).

C. Asymptotic Properties

In the process of system identification, the search domain for the systems parameters will be denoted by $D_{\theta 0}$, with $D_{\theta 0}$ being a compact domain contained in D_θ. In formulating the technical details we need the following definitions.

Definition 3.1: 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 $M_k(\bar{s}) := \sup_{n \geq 0} \mathbb{E}^{1/k}[|\bar{s}_n|^k] < \infty$ for all $1 \leq k < \infty$.

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

Definition 3.2: A stochastic process $z = (z_n(\omega)), n \geq 0$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called asymptotically wide sense stationary with exponentially decaying error if it can be approximated by a wide sense stationary (w.s.st.) process $z = (z_n^{(s)}(\omega)), -\infty < n < \infty$ so that $z_n - z_n^{(s)} = O_M(\alpha^n)$ for $n \geq 0$ with some $0 < \alpha < 1$.

The signals generating the off-line estimate are given by (1) (the system), (3) (the input generator), (6) (the prediction error), and (10) (the prediction error gradient). Under Condition 2.1, with D_θ being such that $A(\theta) - K(\theta)C(\theta)$ is stable for all $\theta \in D_\theta$, and with the input generator (3) being stable for $\eta \in D_\eta$, it is easily seen that the joint process $(\bar{\varepsilon}(\theta; \eta), \bar{\varepsilon}_\theta(\theta; \eta))$ is asymptotically w.s.st. with exponentially decaying error locally uniformly for $\theta \in D_\theta$ and for $\eta \in D_\eta$. It follows that the following asymptotic cost function is well-defined:

$$W(\theta; \eta) = \lim_{n \to \infty} \frac{1}{2} \mathbb{E}[\bar{\varepsilon}_n^2(\theta; \eta)]. \qquad (11)$$

Smoothness of $W(\theta; \eta)$ w.r.t. $(\theta; \eta)$ in the open domain $D_\theta \times D_\eta$ will follow along standard lines under appropriate conditions, see Condition 3.2 below. Now we introduce the notion that the true system is in the model set.

Condition 3.1: For any $\eta \in D_\eta$ there is a unique $\theta^* \in \text{int} D_{\theta 0}$ such that $\bar{\varepsilon}_n(\theta^*; \eta) = e_n$.

Condition 3.1 is often decomposed into two separate conditions: i) identifiability, i.e., each θ corresponds to a unique prediction error transfer function from u and y to $\bar{\varepsilon}(\theta)$, and ii) a persistence of excitation condition on the input u [41].

We will require the following smoothness condition.

Condition 3.2: The matrices $A(\theta), B(\theta), C(\theta)$ and $K(\theta)$ are four-times continuously differentiable functions of θ in D_θ.

With Conditions 3.1 and 3.2 in force it follows that for any $\eta \in D_\eta$, θ^* is the unique global minimizer of $W(\theta; \eta)$ and, under the stability assumption we introduced above for D_θ, it follows from [31] that the off-line estimator converges to θ^* with probability 1 for any feasible η. Furthermore, the estimator will have the asymptotic covariance matrix

$$\lim_{N \to \infty} N \cdot \mathbb{E}[(\hat{\theta}_N(\eta) - \theta^*)(\hat{\theta}_N(\eta) - \theta^*)^T] \qquad (12)$$

equal to the Cramér-Rao lower bound for Gaussian noise, regardless of the distribution of the actual noise see [43], [41].

The asymptotic covariance matrix is given by

$$\sigma_e^2 \tilde{R}^{-1}(\theta^*; \eta) \tag{13}$$

where

$$\tilde{R}(\theta; \eta) := \lim_{n \to \infty} \mathrm{E}[\varepsilon_{\theta,n}(\theta; \eta)\varepsilon_{\theta,n}^T(\theta; \eta)]. \tag{14}$$

IV. RECURSIVE ESTIMATION WITH AN ADAPTIVE INPUT

A. Introduction

A Gauss-Newton type of recursive estimator is given by

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \frac{1}{n}\hat{R}_n^{-1}\varepsilon_{\theta,n}\varepsilon_n,$$

$$\hat{R}_n = \hat{R}_{n-1} + \frac{1}{n}(\varepsilon_{\theta,n}\varepsilon_{\theta,n}^T - \hat{R}_n) \tag{15}$$

where ε_n and $\varepsilon_{\theta,n}$ are on-line estimators of $\bar{\varepsilon}_n(\hat{\theta}_{n-1})$ and $\bar{\varepsilon}_{\theta,n}(\hat{\theta}_{n-1})$, respectively. The algorithm is started from some initial conditions $\hat{\theta}_0$ and $\hat{R}_0$.

In (15), $\varepsilon_{\theta,n}\varepsilon_n$ can be considered an estimate of the gradient

$$W_\theta(\theta; \eta) = \lim_{n \to \infty} \mathbb{E}[\bar{\varepsilon}_{\theta,n}(\theta; \eta)\bar{\varepsilon}_n(\theta; \eta)] \tag{16}$$

and $\hat{R}_n$ a Gauss-Newton estimate of the Hessian $W_{\theta\theta}(\theta; \eta)$ at the current estimate of θ^*, i.e., at $\theta = \hat{\theta}_n$. In order to ensure that the estimator does not leave its domain of definition D_θ, and even stays in a compact domain, say $D_{\theta 0}$, recursive estimation schemes such as (15) typically need to be complemented with either a stopping mechanism or a resetting mechanism, see [44] and [39].

In this work, we consider the recursive estimation algorithm (15) modified by a resetting mechanism. For this purpose, we first consider two compact truncation domains $D_{\theta 0} \subset D_\theta$ and $D_{R0} \subset D_R = \mathbb{R}^{n_\theta \times n_\theta +}$, the latter denoting the set of n_θ by n_θ symmetric positive definite matrices. For now, we impose only the minimal condition that the initial values satisfy $\hat{\theta}_0 \in \mathrm{int}D_{\theta 0}$ and $\hat{R}_0 \in \mathrm{int}D_{R0}$. Then we first define a temporary value for the next estimate as

$$\hat{\theta}_{n+1-} = \hat{\theta}_n - \frac{1}{n+1}R_n^{-1}\varepsilon_{\theta,n+1}\varepsilon_{n+1}, \tag{17}$$

$$R_{n+1-} = R_n + \frac{1}{n+1}(\varepsilon_{\theta,n+1}\varepsilon_{\theta,n+1}^T - R_n), \tag{18}$$

Then define $(\hat{\theta}_{n+1}, R_{n+1})$ to be equal to $\tag{19}$

$$\begin{cases} (\hat{\theta}_{n+1-}, R_{n+1-}), & \hat{\theta}_{n+1-} \in D_{\theta 0}, R_{n+1-} \in D_{R0} \\ (\hat{\theta}_0, R_0), & \text{otherwise.} \end{cases} \tag{20}$$

In regards to ε_n and $\varepsilon_{\theta,n}$, here we employ the standard procedure [36] of making the off-line expressions (6) (the prediction error), and (10) (the prediction error gradient) time-varying by replacing θ with the current estimate. *A major feature of our algorithm is that we will also allow the input generator to be model dependent, i.e. we will allow η to be a function of θ, $\eta = \eta(\theta)$, and use the current estimate as θ in this relationship, making also (3) time-varying.* The motivation for this is that the desired experimental conditions may depend on the true system. At this time we assume the minimal condition that $\eta(\cdot)$, defined on D_θ,
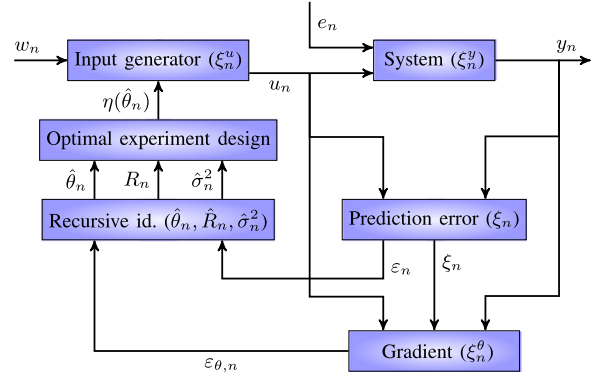

Fig. 1. Schematic of the adaptive system. Variables within parentheses are state variables.

is a continuous function of θ. We will return to the topic of adaptive input design in the next section. This leads to the following time-varying adaptive system:

$$\xi_{n+1}^y = A^y \xi_n^y + B^y u_n + K^y e_n,$$

$$y_n = C^y \xi_n^y + e_n, \tag{21}$$

$$\xi_{n+1}^u = A^u(\eta(\hat{\theta}_n))\xi_n^u + B^u(\eta(\hat{\theta}_n))w_n,$$

$$u_n = C^u(\eta(\hat{\theta}_n))\xi_n^u + D^u(\eta(\hat{\theta}_n))w_n, \tag{22}$$

$$\xi_{n+1} = (A(\hat{\theta}_n) - K(\hat{\theta}_n)C(\hat{\theta}_n))\xi_n$$
$$+ B(\hat{\theta}_n)u_n + K(\hat{\theta}_n)y_n, \tag{23}$$

$$\varepsilon_n = -C(\hat{\theta}_n)\xi_n + y_n,$$

$$\xi_{\theta_k,n+1} = (A(\hat{\theta}_n) - K(\hat{\theta}_n)C(\hat{\theta}_n))\xi_{\theta_k,n}$$
$$+ (A(\hat{\theta}_n) - K(\hat{\theta}_n)C(\hat{\theta}_n))_{\theta_k}\xi_n$$
$$+ B_{\theta_k}(\hat{\theta}_n)u_n + K_{\theta_k}(\hat{\theta}_n)y_n, \tag{24}$$

$$\varepsilon_{\theta_k,n} = -C(\hat{\theta}_n)\xi_{\theta_k,n} - C_{\theta_k}(\hat{\theta}_n)\xi_n, \quad k = 1 \dots n_\theta.$$

The schematic of these equations is given in Fig. 1.

To compactify the notation, we will collect the state vectors for the system comprising (3) (the input generator), (1) (the system), (23) (the prediction error), and (24) (the prediction error gradient), into one single state vector

$$\Phi_n := \begin{bmatrix} (\xi_n^u)^T & (\xi_n^y)^T & (\xi_n)^T & (\xi_{\theta,n})^T \end{bmatrix}^T$$

with $(\xi_{\theta,n})$ denoting the concatenation of the column vectors $(\xi_{\theta_k,n})$.

We shall also extend the above adaptive algorithm with an estimator of σ_e^2, see (28) below, as the noise variance often is required in experiment design. Then for some matrices $A^\Phi(\theta), B^\Phi(\theta), C^\Phi(\theta)$ and $D^\Phi(\theta)$, we can write the entire adaptive system as follows.

Adaptive system

$$\hat{\theta}_0 \in \text{int} D_{\theta 0}, \quad \hat{R}_0 \in \text{int} D_{R0}, \quad \hat{\sigma}_0^2 = 0, \tag{25}$$

$$\Phi_{n+1} = A^\Phi(\hat{\theta}_n)\Phi_n + B^\Phi(\hat{\theta}_n)\begin{bmatrix} w_n \\ e_n \end{bmatrix}, \tag{26}$$

$$\begin{bmatrix} \varepsilon_n \\ \varepsilon_{\theta,n} \end{bmatrix} = C^\Phi(\hat{\theta}_n)\Phi_n + D^\Phi(\hat{\theta}_n)\begin{bmatrix} w_n \\ e_n \end{bmatrix}, \tag{27}$$

$$\hat{\sigma}_{n+1}^2 = \hat{\sigma}_n^2 + \frac{1}{n+1}(\varepsilon_n^2 - \hat{\sigma}_n^2), \tag{28}$$

$$\hat{\theta}_{n+1-} = \hat{\theta}_n - \frac{1}{n+1}\hat{R}_n^{-1}\varepsilon_{\theta,n+1}\varepsilon_{n+1}, \tag{29}$$

$$\hat{R}_{n+1-} = \hat{R}_n + \frac{1}{n+1}(\varepsilon_{\theta,n+1}\varepsilon_{\theta,n+1}^T - \hat{R}_n), \tag{30}$$

$$(\hat{\theta}_{n+1}, \hat{R}_{n+1}) =$$
$$\begin{cases} (\hat{\theta}_{n+1-}, \hat{R}_{n+1-}), & \hat{\theta}_{n+1-} \in D_{\theta 0}, \hat{R}_{n+1-} \in D_{R0} \\ (\hat{\theta}_0, \hat{R}_0), & \text{otherwise.} \end{cases} \tag{31}$$

From Fig. 1, we see that the subsystems generating the input, the system, the prediction error and the prediction error gradient are connected in cascade. This implies that the state-transition matrix $A^\Phi(\theta)$ in (26) is lower block triangular. Furthermore, each block on the diagonal of $A^\Phi(\theta)$ is either $A^u(\eta(\theta))$, A^y or $A(\theta) - K(\theta)C(\theta)$.

B. Asymptotic Properties

In this section, we will analyze the adaptive system (25)–(31) as n grows towards infinity. Our main interest is the behavior of $\hat{\theta}_n$. Our analysis is based on the results of [39] and extended in [32].

Let $\eta^* = \eta(\theta^*)$ denote the optimal experimental setting assuming that the system parameter θ^* is known, and let $\hat{\theta}_N(\eta^*)$ denote the off-line estimator of θ^* under the optimal experimental conditions. Then our ultimate goal is to establish the asymptotic equivalence of the recursive estimator $\hat{\theta}_N$ generated by our adaptive algorithm and the optimal off-line estimator $\hat{\theta}_N(\eta^*)$ in a technical sense to be specified in Theorems 4.1 and 4.1 below.

A pre-requisite for this analysis is to establish the overall stability of the time-varying system. We will base our analysis on the joint spectral radius.

Definition 4.1: For a given set of fixed size square matrices, say $\mathcal{A}$, the joint spectral radius is defined as

$$\lambda(\mathcal{A}) = \sup_I \limsup_n ||A(n)\cdots A(1)||^{1/n}, \quad A(k) \in \mathcal{A} \tag{32}$$

with I denoting the set of all possible infinite selection of sequences $A(k)$. The set $\mathcal{A}$, is said to be jointly stable if $\lambda(\mathcal{A}) < 1$.

We will need the following assumption.

Condition 4.1: Let $D_{\eta 0} \subset \mathbb{R}^{n_\eta}$ be a compact set. The joint spectral radius of $\Sigma_u = \{A^u(\eta)) : \eta \in D_{\eta 0}\}$ is less than one.

Since the input generator (3) is in the hands of the user, Condition 4.1 can be ensured using techniques from the theory on stability of linear time-varying systems. As we will see in Section V, in adaptive input design $A^u(\cdot)$ does typically not depend on η, in which case the condition on the joint spectral radius of Σ_u is trivially satisfied.

Condition 4.2: The joint spectral radius of $\Sigma_\theta = \{A(\theta) - K(\theta)C(\theta) : \theta \in D_{\theta 0}\}$ is less than one.

In transfer function notation, the stability condition on $A(\theta) - K(\theta)C(\theta)$ corresponds to that the transfer functions $H^{-1}(q, \theta)$ and $H^{-1}(q, \theta)G(q, \theta)$ are stable. This is a standard condition [41], however joint stability as required in Condition 4.2 is at first sight certainly restrictive. However, joint stability represents the state-of-the art in recursive parameter estimation and in fact it is often tacitly assumed in the literature. Moreover, this assumption was justified, under appropriate technical modifications of standard procedures, in [32]. The arguments presented there applies in our case.

Remark 4.1: Observe that Condition 4.2 is trivially satisfied for ARARX models, i.e., models of the type

$$A(q)y_t = B(q)u_t + \frac{1}{D(q)}e_t,$$

where $A(q), B(q)$ and $D(q)$ are parametrized polynomials in q^{-1}. This follows since $H^{-1}(q, \theta) = A(q)D(q)$ and $H^{-1}(q, \theta, \theta)G(q, \theta) = B(q)D(q)$. The ARARX model is a stochastic model commonly used in economics, engineering, health and medical science literature (see, e.g., [45]–[52] and the references therein).

For ARMAX models, i.e., models of the type

$$A(q)y_t = B(q)u_t + C(q)e_t,$$

Condition 4.2 can be relaxed when the input is not adaptively updated, e.g., the method in [40] applies to white inputs.

We now have the following stability result.

Lemma 4.1: Suppose that $\eta(\cdot)$, defined on D_θ, is a continuous function of θ, and the image of $D_{\theta 0}$ is a subset of $D_{\eta 0}$. Then Conditions 2.1, 4.1 and 4.2 imply that the set of state-transition matrices $A^\Phi(\theta)$ with $\theta \in D_{\theta 0}$ is jointly stable.

Proof: The result is a direct consequence of Lemma 1.1 of Appendix A. ∎

It follows that, under the conditions of the lemma above, the map from u and e to Φ as defined by (25)–(31) is BIBO stable in L_p for any $1 \le p \le \infty$.

The following smoothness assumptions for the input generator mechanism will be required.

Condition 4.3: The matrices $A^u(\eta), B^u(\eta), C^u(\eta)$ and $D^u(\eta)$ are four-times continuously differentiable on an open domain D_η, satisfying $D_{\eta 0} \subset D_\eta \subset \mathbb{R}^{n_\eta}$.

Condition 4.4: The map $\eta(\theta)$ is four-times continuously differentiable on D_θ and its image of $D_{\theta 0}$ is a subset of $D_{\eta 0}$, defined in Condition 4.1.

For recursive algorithms of the type (15) one can at best find a stationary point of $W(\theta; \eta)$. A critical assumption in the ensuing developments is the following condition, which is a strengthening of Condition 3.1.

Condition 4.5: With D_η as in Condition 4.1, the equation

$$W_\theta(\theta; \eta) = 0 \tag{33}$$

has the unique solution $\theta = \theta^*$, which is assumed to belong to $\text{int} D_{\theta 0}$, for any $\eta \in D_\eta$. Moreover it is assumed that θ^* is locally identifiable, i.e., the Hessian-matrix $W_{\theta\theta}(\theta^*; \eta)$ is positive definite for any $\eta \in D_\eta$.

The verification of the above condition is far from trivial. In the remarkable paper [53], Åström and Söderström prove that in the case of an ARMA process $Dy = Ce$, using standard parametrization, the above condition holds under minimal conditions on C and D.

It is well known that the algorithm (25)–(31) can be viewed as finite-difference equations which has a natural connection with a set of ordinary differential equations (ODEs) (see [34], [36], [54] and [35]). The ODE associated with the algorithm is given as follows (see, e.g., [34], [39] and [32])

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_t = -\frac{1}{t}R_t^{-1}W_\theta(\theta_t, \eta(\theta_t)), \tag{34a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}R_t = \frac{1}{t}\left(\tilde{R}(\theta_t, \eta(\theta_t)) - R_t\right) \tag{34b}$$

for $t \geq 1$ with initial condition for $t = 1$ equal to $(\theta_0, R_0) = (\hat{\theta}_0, \hat{R}_0)$.

The adaptive system (25)–(31) has a particular structure that we will explore. In order to express these properties, we introduce the joint vector-variable $z = \begin{bmatrix} \theta^T & (\mathrm{vec}\,R)^T \end{bmatrix}^T$, and define the corresponding domains of definitions and search domains as $D = \{z : \theta \in D_\theta, R \in D_R\}$ and $D_0 = \{z : \theta \in D_{\theta 0}, R \in D_{R0}\}$. We also let

$$\frac{\mathrm{d}}{\mathrm{d}t}z_t = \frac{1}{t}h(z_t) \tag{35}$$

denote the vectorized version of the ODE (34).

Lemma 4.2: Let Conditions 3.1, 3.2, 4.3, 4.4, and 4.5 be in force. Then:

i) It holds that

$$W_{\theta\eta}(\theta^*; \eta) = 0 \tag{36}$$

for any $\eta \in D_\eta$.

ii) The equation

$$W_\theta(\theta; \eta(\theta)) = 0 \tag{37}$$

has the unique solution $\theta = \theta^*$ in D_θ.

iii) It holds that

$$W_{\theta\theta}(\theta^*; \eta) = \tilde{R}(\theta^*; \eta) > 0 \tag{38}$$

for any $\eta \in D_\eta$.

iv) The ODE (34) has unique equilibrium $\theta_t = \theta^*$ and $\hat{R}_t = R^* := \tilde{R}(\theta^*; \eta(\theta^*))$

v) The Jacobian of h in (35), evaluated at $z^* = \begin{bmatrix} (\theta^*)^T & (\mathrm{vec}\,R^*)^T \end{bmatrix}^T$, has the structure

$$\begin{bmatrix} -I_{n_\theta} & 0 \\ * & -I_{n_\theta^2} \end{bmatrix}. \tag{39}$$

Proof: Consider (33) as an implicit function to be solved for θ. Differentiating w.r.t. η, we immediately get $W_{\theta\eta}(\theta^*; \eta) = 0$ for any $\eta \in D_\eta$, implying i). ii) follows from i) and Condition 4.5. For iii), we notice that from (16) it follows that $W_{\theta\theta}(\theta^*; \eta)$ consists of two terms, $\tilde{R}(\theta^*; \eta)$ and

$$\lim_{n\to\infty}\mathbb{E}[\bar{\varepsilon}_{\theta\theta n}(\theta^*; \eta)\bar{\varepsilon}_n(\theta^*; \eta)].$$

By Condition 3.1, $\bar{\varepsilon}_n(\theta^*; \eta) = e_n$ which is independent of $\bar{\varepsilon}_{\theta\theta n}(\theta^*; \eta)$, since the latter quantity is a function of $\{e_k\}, k < n$

only. Hence, the term above is zero and iv) follows. Finally, v) follows from i).																■

Condition 4.6: Let the unique solution to (35) with $z_s = \xi$ for $s \geq 1$ be denoted $z(t, s, \xi)$. Also let $z_0 = \begin{bmatrix} (\theta_0)^T & (\mathrm{vec}\,R_0)^T \end{bmatrix}^T$. Let $D_0 \subset D$ be a compact truncation domain such that $z^* \in \mathrm{int}\,D_0$.

The following conditions hold:

i) There exist compact convex sets $D'_{\theta 0} \subset D_\theta$ and $D'_{R0} \subset D_R$ such that, with $D'_0 = \{z : \theta \in D'_{\theta 0}, R \in D'_{R0}\}$

$$z(t, s, \xi) \in D'_0 \ for\ \xi \in D_0 \ and\ z(t, s, \xi) \in D \ for\ \xi \in D'_0 \tag{40}$$

for all $t \geq s \geq 1$. In addition $\lim_{t\to\infty} z(t, s, \xi) = z^*$ for $\xi \in D'_0$,

ii) We have an initial estimate $z_0 = \xi_0$ such that for all $t \geq s \geq 1$ we have $z(t, s, \xi_0) \in \mathrm{int}\,D_0$.

Applying [32], we get the remarkable result that the recursive estimator $\hat{\theta}_n$ generated by our adaptive algorithm is asymptotically equivalent to the optimal off-line estimator, in a sense specified in the theorem below.

Theorem 4.1 Consider the adaptive system (25)–(31). Let the following conditions be in force:

1) *Stochastic processes:* Conditions 2.2 and 2.3.
2) *Smoothness:* Conditions 3.2, 4.3 and 4.4.
3) *Identifiability:* Conditions 3.1 and 4.5.
4) *Stability:* Conditions 2.1, 4.1, and 4.2.
5) *ODE:* Conditions 4.6.

Then we have

$$\hat{\theta}_n - \hat{\hat{\theta}}_n(\eta^*) = O_M(n^{-1/2-\delta}), \tag{41}$$

for some constant $\delta > 0$.

Proof: See Appendix B.																■

For the interpretation of the result, recall that it can be shown, following the arguments of [42] step by step, that we have

$$\hat{\hat{\theta}}_n(\eta^*) - \theta^* = -\frac{1}{n}\sigma_e^2\,(R^*)^{-1}\sum_{n=0}^{N-1}\bar{\varepsilon}_{\theta,n}(\theta^*; \eta^*)e_n + O_M(n^{-1}).$$

It follows that $\hat{\hat{\theta}}_n(\eta^*) - \theta^* = O_M(n^{-1/2})$. Since the difference between $\hat{\theta}_n$ and $\hat{\hat{\theta}}_n(\eta^*)$ is order of magnitude $O_M(n^{-1/2-\delta})$, we conclude that statistically important properties $\hat{\hat{\theta}}_n(\eta^*)$ will be automatically inherited by $\hat{\theta}_n$. In particular, consider the covariance matrix of $\hat{\hat{\theta}}_n(\eta^*) - \theta^*$, which is readily seen to be

$$\mathbb{E}[\hat{\hat{\theta}}_n(\eta^*) - \theta^*)(\hat{\hat{\theta}}_n(\eta^*) - \theta^*)^T]$$

$$= \frac{1}{n}\sigma_e^2\,(R^*)^{-1} + O(n^{-1}). \tag{42}$$

As a direct consequence of the above we have the following corollary.

Corollary 4.1: Consider the adaptive system (25)–(31) and let the assumptions in Theorem 4.1 be in force. Then the covariance matrix of the error process $(\hat{\theta}_n - \theta^*)$ satisfies

$$\mathbb{E}[(\hat{\theta}_n - \theta^*)(\hat{\theta}_n - \theta^*)^T]$$

$$= \frac{1}{n}\sigma_e^2\,(R^*)^{-1} + O(n^{-1-\delta}), \tag{43}$$

for some $\delta > 0$.

V. ADAPTIVE INPUT DESIGN

We will now apply the results presented in the previous s ection to the case where the input generator (3) corresponds to the solution of an optimal input design problem. We will tailor our results to the general frameworks in [13], [14], [4], which consists of the two steps: (i) Design of the input spectrum by way of a semidefinite program (SDP), and (ii) spectral factorization of the corresponding spectrum, yielding the input generator.

The main objective is to establish conditions under which Theorem 4.1 holds, as this will then establish that adaptive input design asymptotically achieves the same accuracy as optimal input design in the sense that the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta^*)$ for the adaptive scheme is the same as for the off-line case using the optimal input. The condition related to the input generator are Conditions 4.1, 4.3, and 4.4. This means that our main tasks are to establish stability of the time-varying linear system (3) and that the map from the model parameters to the state space matrices in (3) is sufficiently smooth.

For these considerations, the essential characteristic of the optimal input design problems in the aforementioned references is that they can be formulated as

$$\min_{\eta \in \mathbb{R}^{n_\eta}, \gamma \in \mathbb{R}^{n_\gamma}} \gamma_1 \tag{44}$$

$$\text{s.t. } M(\eta, \gamma, \theta) \geq 0, \tag{45}$$

where (45) is a Linear Matrix Inequality (LMI), i.e. $M(\eta, \gamma, \theta)$ is affine in the decision variables η and γ, so that the problem is an SDP.

The decision variable $\eta = \begin{bmatrix} \eta_1 & \cdots & \eta_{n_\eta} \end{bmatrix}^T \in \mathbb{R}^{n_\eta}$ contains the coefficients in a finite expansion of the input spectrum

$$\Phi_u(e^{i\omega}, \eta) = \sum_{k=1}^{n_\eta} \eta_k (\mathcal{B}_k(e^{i\omega}) + \mathcal{B}_k^*(e^{i\omega})), \tag{46}$$

where $\{\mathcal{B}_k\}_{k=1}^{n_\eta}$ are stable rational basis functions. It is common to use an input generator of FIR type. This corresponds to the basis

$$\mathcal{B}_k(z) = \begin{cases} 1/2 & k = 1, \\ z^{-(k-1)} & k > 1. \end{cases} \tag{47}$$

The matrix $M(\eta, \gamma, \theta)$ is block diagonal where each block captures, e.g., signal constraints/criteria and model quality constraints/criteria, see below. The formulation (44)–(45) covers both the case where a model quality measure is optimized subject to constraints on the used signals, or the opposite formulation (known as least-costly design [14]).

The auxiliary variable $\gamma = \begin{bmatrix} \gamma_1 & \cdots & \gamma_{n_\gamma} \end{bmatrix}^T \in \mathbb{R}^{n_\gamma}$ is (partly) used to incorporate a condition that ensures that Φ_u, defined in (46), is non-negative. The latter can be ensured by the positive real lemma (see, e.g., [13, Lemma 2.1]) and corresponds to an LMI. For the case of an input generator of FIR type, i.e.,

the basis(47), it takes the form

$$\mathbf{K}(Q; \{A^u, B^u, C^u, D^u\}) :=$$

$$\begin{bmatrix} Q - (A^u)^T Q A^u & -(A^u)^T Q B^u \\ -(B^u)^T Q A^u & -(B^u)^T Q B^u \end{bmatrix} + \begin{bmatrix} 0 & (C^u)^T \\ C^u & 2D^u \end{bmatrix} \geq 0 \tag{48}$$

where

$$A^u = \begin{bmatrix} 0_{n_\eta - 2} & I_{n_\eta - 2} \\ 0 & 0_{n_\eta - 2}^T \end{bmatrix}, \quad B^u = \begin{bmatrix} 0_{n_\eta - 2} \\ 1 \end{bmatrix},$$

$$C^u = C^u(\eta) = [\eta_{n_\eta} \quad \cdots \quad \eta_2], \quad D^u = D^u(\eta) = \frac{1}{2}\eta_1. \tag{49}$$

The unique elements of $Q = Q^T \geq 0$ are elements of γ. $\mathbf{K}(Q; \{A^u, B^u, C^u, D^u\})$ is thus one of the blocks of $M(\eta, \gamma, \theta)$.

Signal constraints are in terms of constraints on signal spectra, either energy constraints or frequency-by-frequency constraints. To illustrate the expressions involved, with the input spectrum given by (46), the input energy for an experiment of length N can, using Parseval's theorem, be expressed as

$$N\mathbb{E}[u_t^2(\eta)] = \frac{N}{2\pi} \int_{-\pi}^{\pi} \Phi_u(e^{i\omega}, \eta) \, d\omega = \sum_{k=1}^{n_\eta} \beta_k \, \eta_k \tag{50}$$

where $\beta_k = \frac{N}{2\pi} \int_{-\pi}^{\pi} (\mathcal{B}_k(e^{i\omega}) + \mathcal{B}_k^*(e^{i\omega}))d\omega$. Similarly, the noise-free output energy of a model can be expressed as

$$N\mathbb{E}\left[\left(C(\theta)(qI - A(\theta))^{-1} B(\theta)u_t(\eta)\right)^2\right] = \sum_{k=1}^{n_\eta} \alpha_k(\theta) \, \eta_k$$

where

$$\alpha_k(\theta) = \frac{N}{2\pi} \int_{-\pi}^{\pi} \left| C(\theta)(e^{i\omega}I - A(\theta))^{-1} B(\theta) \right|^2$$

$$\times (\mathcal{B}_k(e^{i\omega}) + \mathcal{B}_k^*(e^{i\omega}))d\omega. \tag{51}$$

The blocks of (45) that correspond to model quality measures are affine functions of the information matrix. Modulo a normalization constant, the information matrix corresponds to $\tilde{R}(\theta; \eta)$ defined in (14). Employing Parseval's formula and (46), we can write

$$\tilde{R}(\theta; \eta) = \sum_{k=1}^{n_\eta} \eta_k \, \tilde{R}_k(\theta) + \tilde{R}_e(\theta), \tag{52}$$

where

$$\tilde{R}_k(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_u(e^{i\omega}, \theta)\Gamma_u^*(e^{i\omega}, \theta)(\mathcal{B}_k(e^{i\omega}) + \mathcal{B}_k^*(e^{i\omega}))d\omega, \tag{53}$$

$$\tilde{R}_e(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_e(e^{i\omega}, \theta)\Gamma_e^*(e^{i\omega}, \theta)d\omega \tag{54}$$

where $\Gamma_u(z, \theta)$ and $\Gamma_e(z, \theta)$ are stable rational vector-valued functions for $\theta \in D_\theta$, see [13], [41]. The term $\tilde{R}_e$ is due to the noise excitation. Thus the information matrix is an affine function of η, and hence the blocks of (45) that correspond to model quality measures are affine functions of η as well.

The expressions (51), (53)–(54), are representative for the type of dependence between M and θ. In summary, the optimal input design frameworks of [13], [14], [4] lead to SDPs that can be written as (44)–(45), with

$$M(\eta, \gamma, \theta)$$

$$= \sum_{k=1}^{n_\eta} \eta_k \, M_k(\theta) + \sum_{k=1}^{p_\gamma} \gamma_k \, M_{n_\eta + k}(\theta) + M_{n_\eta + p_\gamma + 1}(\theta) \quad (55)$$

where

$$M_k(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{\Gamma}_k(e^{i\omega}, \theta) \bar{\Gamma}_k^*(e^{i\omega}, \theta) \quad (56)$$

where in turn $\{\tilde{\Gamma}_k(z, \theta)\}$ and $\{\bar{\Gamma}_k(z, \theta)\}$ are vector-valued rational transfer functions in z, with coefficients possibly depending on θ, stable on $D_{\theta 0}$.

It is clear from (55) that (44)–(45) is an SDP in η and γ. Spectral factorization of the resulting spectrum (46) yields a stable filter which we denote $G_u(z, \eta(\theta))$. Realizing this filter in state-space form gives the input generator (3). The filter will share poles with the basis functions $\{\mathcal{B}_k(q)\}_{k=1}^{n_\eta}$. Thus it is only the numerator coefficients that depend on η so G_u can be written

$$G_u(z, \eta) = \frac{\sum_{k=0}^{m} g_k(\eta) z^{-k}}{d(z)}, \quad (57)$$

for some fix denominator polynomial $d(z)$. We can thus realize the filter in state-space form (3) using a controllable form [55], where A^u and B^u are fixed matrices and where C^u and D^u depend linearly on the filter coefficients $\{g_k(\eta)\}_{k=1}^{m}$.

Remark 5.1: It may not be possible to generate all possible information matrices using a finite expansion (46) of the input spectrum. This can be guaranteed using the so-called *partial correlation parametrization* which is a linear infinite dimensional parametrization of the input spectrum, chosen such that (52) still has a finite number of nonzero terms. Using partial correlation also leads to optimal input design problems of the type (44)–(45), but the computation of the spectral factor G_u differs. We refer to [13] for details.

We now summarize the adaptive input design algorithm that we will analyze.

For the above algorithm, we have the following result.

Theorem 5.1: Suppose that
  (i) $M(\eta, \gamma, \theta)$ in (45) is given by (55)–(56), where $\{\tilde{\Gamma}_k(z, \theta)\}$ and $\{\bar{\Gamma}_k(z, \theta)\}$ are vector-valued rational transfer functions in z, with coefficients possibly depending on θ, stable on D_θ.
 (ii) Problem (44)–(45) is well posed in the sense that for each $\theta \in D_\theta$, the minimum is bounded from below. Assume also that (45) is strictly feasible for any θ in D_θ.
(iii) Problem (44)–(45) has a unique solution for every $\theta \in D_\theta$.
(iv) Conditions 2.1, 2.2, 2.3, 3.2, 4.2, 4.5, and 4.6 hold.
Then $\{\hat{\theta}_n\}$ generated by Algorithm 5.1 possesses the properties stated in Theorems 4.1 and Corollary 4.1.

Algorithm V.1:

1) *Parametrization:* Choose stable rational basis functions $\{\mathcal{B}_k(z)\}_{k=1}^{n_\eta}$ in the input spectrum expansion (46).
2) *Initial estimate:* Define compact sets $D_{\theta 0} \subset \mathbb{R}^{n_\theta}$ and $D_{R0} \subset \mathbb{R}^{n_\theta \times n_\theta +}$ and take $\hat{\theta}_0 \in \text{int} D_{\theta 0}, R_0 \in \text{int} D_{R0}$ and $n = 0$.
3) *Generate input process:* Take $\{w_n\}$, to be used as the signal generating the input (see Fig. 1), to be a sequence of independent random variables satisfying Condition 2.3.
4) *Input spectrum update:* Compute the optimal solution $\eta(\hat{\theta}_n)$ to (44)–(45).
5) *Input filter update:* Compute the corresponding stable minimum phase input filter $G_u(z, \eta(\hat{\theta}_n))$, with struc ture (57), by spectral factorization of the correspo nding input spectrum $\Phi_u(e^{i\omega}, \eta(\hat{\theta}_n))$.
6) *Input generator update:* Compute the controllable state-space realization of transfer function $G_u(z, \eta(\hat{\theta}_n))$, giving $A^u(\eta(\hat{\theta}_n)), B^u(\eta(\hat{\theta}_n))$, $C^u(\eta(\hat{\theta}_n))$ and $D^u(\eta(\hat{\theta}_n))$.
7) *Measurement update:* Compute and apply the input signal u_{n+1}, generated by (3) with $\eta = \eta(\hat{\theta}_n)$, to the true system and collect a new measurement y_{n+1} from the true system.
8) *Parameter estimate update:* The updated recursive estimate $\hat{\theta}_{n+1}$ is computed by (29)–(31).
9) *Iterate:* Replace n by $n + 1$ and go to step 4).

Corollary 5.1: Let

$$z_n = F(q) \begin{bmatrix} \Phi_n \\ \varepsilon_n \\ \varepsilon_{\theta, n} \end{bmatrix},$$

where Φ_n, ε_n and $\varepsilon_{\theta, n}$ are defined in (26)–(27), and where $F(q)$ is a stable transfer function matrix. Then

$$\frac{1}{n} \sum_{k=1}^{n} z_k z_{k-\tau}^T \quad (58)$$

exists almost surely for any integer τ. The limit equals the corresponding correlation for the same signals when the input is generated by (3) with η taken as the solution to (44)–(45) with $\theta = \theta^*$, the parameter vector corresponding to the true system.

Proof: See Appendix C. ∎

Remark 5.2: By Corollary 4.1, it follows that Theorem 5.1 shows that the adaptive Algorithm 5.1 asymptotically recovers the same accuracy as using the optimal input during the experiment together with the off-line prediction error method.

Remark 5.3: It follows from (58) that the sample input power

$$\bar{u}_n^2 := \frac{1}{n} \sum_{k=1}^{n} u_k^2 \quad (59)$$

converges almost surely to the power of the optimal input signal. Furthermore, for $\hat{\sigma}_n^2$ defined by (28), it follows that $\hat{\sigma}_n^2 \to \sigma_e^2$ almost surely.

Remark 5.4: The condition on well-posedness is not restrictive. For example, it is trivially satisfied for the common objective of minimizing some measure of the experimental effort, e.g., the input energy.

Remark 5.5: Condition 4.5 implies that the solution to (44)–(45) has to correspond to a nonzero input spectrum for any θ in D_θ.

Theorem 5.1 requires strict feasibility of the SDP (44)–(45). In the next lemma we establish that this holds generally for the constraints used in [4], [13], [14]. We state the results for the commonly used FIR basis, but the results are straightforward to extend to a general stable rational set of basis functions.

Lemma 5.1: Let Z be a positive (semi-)definite matrix. Then the LMI (48) associated with the positivity condition $\Phi_u(\omega) \geq 0, \forall \omega$ and the quality constraint

$$\tilde{R}(\theta; \eta) \geq Z \qquad (60)$$

are strictly feasible.

Proof: See Appendix D. ∎

Remark 5.6: Not all quality constraints in [4], [13], [14] are of the type (60). For example, [13] employ quality constraints of the type

$$\mu - \mathrm{Tr}\, Z \geq 0$$
$$\begin{bmatrix} Z & V^* \\ V & \tilde{R}(\theta; \eta) \end{bmatrix} \geq 0 \qquad (61)$$

where μ and V are fix quantities, and where $Z = Z^T \in \mathbb{R}^{p_z \times p_z}$ is an auxiliary variable. If we take $Z = \mu/(2p_z)I$, Schur complement gives that the condition

$$\tilde{R}(\theta; \eta) - V Z^{-1} V^* = \tilde{R}(\theta; \eta) - \frac{2p_z}{\mu} V V^* > 0 \qquad (62)$$

implies (61) with strict inequalities, i.e. strict feasibility. The condition (62) is of the type (60) and hence Lemma 5.1 applies also to constraints of the type (61).

Remark 5.7: It is straightforward to extend Theorem 5.1 to the case where the system operates in closed loop with a fix stabilizing LTI controller, and the experiment design problem concerns designing the optimal reference signal. The expressions for signal spectra and the information matrix become more involved, but retain the structure (55)–(56) that we rely on for the theorem.

Remark 5.8: For input design problems where some signal size measure is the objective function, the first phase of Algorithm 5.1 may generate excessive excitation if the trajectory of the parameter estimate $\{\theta_n\}$ passes through models that correspond to systems that are difficult to identify, i.e. require large signal sizes in order to achieve the quality specified by (44)–(45). A practical way to avoid this is to limit the signal size in an initial phase.

VI. APPLICATION TO MINIMUM VARIANCE CONTROL

A. Optimal Input Design for Minimum Variance Control

In this Section we will study the problem of optimal open loop input design when the model is to be used for (approximate) minimum variance control of an ARMAX system

$$A(q, \theta)y_n = B(q, \theta)u_n + C(q, \theta)e_n, \qquad (63)$$

where $A(q, \theta) = 1 + a_1 z^{-1} + \ldots + a_{n_a} z^{-n_a}$, $B(q, \theta) = b_1 z^{-1} + \ldots + b_{n_b} z^{-n_b}$ and $C(q, \theta) = 1 + c_1 z^{-1} + \ldots + c_{n_c} z^{-n_c}$, and where $\theta = [a_1 \ldots a_{n_a}\ b_1\ \ldots b_{n_b}\ c_1 \ldots c_{n_c}]^T$. We will assume that the system is minimum phase, i.e., $z^{n_b} B(z, \theta) = \sum_{k=1}^{n_b} b_k z^{n_b - k}$ has all its roots strictly inside the unit circle.

The minimum variance controller for (63) under the minimum phase assumption is given by [56]

$$u_n = -K^*(q, \theta)y_n, \ K^*(q, \theta) := \frac{C(q, \theta) - A(q, \theta)}{B(q, \theta)}. \qquad (64)$$

Using an estimate $\hat{\theta}_N$ of the model parameters θ, and the certainty equivalence principle, i.e. replacing the unknown parameter by its estimate, leads to use of the controller $K^*(q, \hat{\theta}_N)$ which gives the closed-loop output

$$y_n(K^*(\hat{\theta}_N), \theta) := \frac{1}{1 + \frac{B(q,\theta)}{A(q,\theta)} K(q, \hat{\theta}_N)} \frac{C(q, \theta)}{A(q, \theta)} e_n.$$

Following [57]–[59] we will use the degradation from the ideal performance as performance measure, i.e. we take

$$J(\hat{\theta}_N, \theta) = \mathbb{E}[y_n^2(K^*(\hat{\theta}_N)|\hat{\theta}_N] - \mathbb{E}[y_n^2(K^*(\theta)))^2]$$
$$= \mathbb{E}[(y_n^2(K^*(\hat{\theta}_N))|\hat{\theta}_N] - \sigma_e^2$$
$$= \mathbb{E}\left[\left(\frac{1}{1 + \frac{B(q,\theta)}{A(q,\theta)} K(q, \hat{\theta}_N)} \frac{C(q, \theta)}{A(q, \theta)} e_n\right)^2 |\hat{\theta}_N\right]$$
$$- \sigma_e^2.$$

Here, expectation is taken with respect to the noise sequence $\{e_n\}$, while keeping the estimate $\hat{\theta}_N$ fix. This means that $J(\hat{\theta}_N, \theta)$ measures the increase in output variance that occurs when the parameter estimate $\hat{\theta}_N$ is used instead of the true θ in the minimum variance controller. To account for the randomness of the identification experiment we will also take expectation over the distribution of the identified model parameters $\hat{\theta}_N$; we denote this expectation by $\mathbb{E}_{\hat{\theta}_N}[\cdot]$. We will thus use $\mathbb{E}_{\hat{\theta}_N}[J(\hat{\theta}_N, \theta)]$ as measure of the performance degradation. This is thus the performance degradation, as compared to the minimum variance controller, that can be expected by i) performing an identification experiment, ii) computing an estimate $\hat{\theta}_N$, and iii) applying the controller K^*, with the unknown θ substituted by $\hat{\theta}_N$, to the true system and computing the increase in output variance.

The performance degradation $\mathbb{E}_{\hat{\theta}_N}[J(\hat{\theta}_N, \theta)]$ depends on the experimental conditions and the objective of is to choose these such that the performance degradation is minimized. For computational tractability we will, again following [57]–[59], use the asymptotic performance degradation

$$\bar{V}(\theta) := \lim_{N \to \infty} N\mathbb{E}_{\hat{\theta}_N}[J(\hat{\theta}_N, \theta)] \qquad (65)$$

as our performance degradation measure. Experience [57]–[59] shows that $\bar{V}(\theta)$ is an accurate approximation of $N\mathbb{E}_{\hat{\theta}_N}[J(\hat{\theta}_N, \theta)]$ even for fairly small sample sizes N. It is

straightforward to show that [32]

$$\bar{V}(\theta) = \sigma_e^2 \, \mathrm{Tr}\{W(\theta)\tilde{R}^{-1}(\theta;\eta)\}, \tag{66}$$

where

$$W(\theta) = \mathbb{E}\left[\left(\left.\frac{\partial y_n(K^*(\eta))}{\partial \eta}\right|_{\eta=\theta}\right)^T \left.\frac{\partial y_n(K^*(\eta))}{\partial \eta}\right|_{\eta=\theta}\right]. \tag{67}$$

For large model orders, it was shown in [57] that a closed loop identification experiment using the minimum variance controller (64) minimizes $\bar{V}$ over all possible experimental conditions. This was later shown to hold also for models of finite order [58], [59]. Here, we will restrict the experiment to be open loop and consider the case when the input generator is of FIR type, i.e. we will use the basis (47).

Simple computations then give that Γ_u in (53) and Γ_e in (54), respectively, can be expressed as

$$\Gamma_u(q,\theta) = \begin{bmatrix} \frac{B(q,\theta)}{C(q,\theta)A(q,\theta)}\,\Gamma_{n_a}(q) \\[2ex] \frac{-1}{C(q,\theta)}\,\Gamma_{n_b}(q) \\[2ex] 0_{n_c \times 1} \end{bmatrix},$$

$$F_e(q,\theta) = \begin{bmatrix} \frac{1}{A(q,\theta)}\,\Gamma_{n_a}(q) \\[2ex] 0_{n_b \times 1} \\[2ex] \frac{-1}{C(q,\theta)}\,\Gamma_{n_c}(q) \end{bmatrix}, \quad \Gamma_n(q) = \begin{bmatrix} q^{-1} \\ \vdots \\ q^{-n} \end{bmatrix}.$$

To emphasize that $\bar{V}(\theta)$ is a function of the spectrum parameters we write $\bar{V}(\theta,\eta)$. We will consider the optimal input design problem consisting of minimizing the performance degradation subject to a constraint p_u on the input power, i.e.

$$\begin{aligned} \min_\eta \quad & \bar{V}(\theta,\eta) \\ \text{s.t.} \quad & \mathbb{E}[u_n^2] \le p_u, \quad \Phi_u(e^{i\omega},\eta) \ge 0, \; \forall \omega, \end{aligned} \tag{68}$$

where $\bar{V}(\theta,\eta)$ can be expressed as (66). Taking $W^{1/2}(\theta)$ to be the matrix square root of $W(\theta)$, this weighted trace problem can, using Schur complemnt, be expressed as the semi-definite program

$$\begin{aligned} \min_{\eta,Z,Q} \quad & \mathrm{Tr}\{Z\} \\ \text{s.t.} \quad & \begin{bmatrix} Z & W^{1/2}(\theta) \\ W^{1/2}(\theta) & \sum_{k=1}^{n_\eta} \tilde{R}_k(\theta)\,\eta_k + \tilde{R}_e(\theta) \end{bmatrix} \ge 0, \\ & \eta_1 \le p_u, \; K(Q,\{A^u(\eta),B^u(\eta),C^u(\eta),D^u(\eta)\}) \ge 0 \end{aligned} \tag{69}$$

where $K(Q,\{A^u(\eta),B^u(\eta),C^u(\eta),D^u(\eta)\})$ is given in (48). For details, see [13]. Observes that (69) conforms to the structure of (44)–(45). The unique elements of Z and Q in (69) correspond to the vector γ in (44)–(45), the left-hand sides of the three inequalities in (69) correspond to three diagonal blocks of $M(\eta,\gamma,\theta)$ in (45). In adaptive input design, θ in (69) is replaced by the most recent estimate.
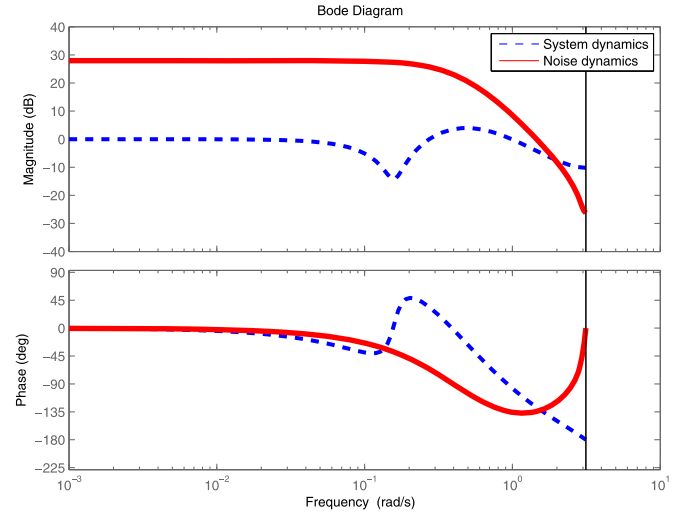

Fig. 2.   Bode diagram of system dynamics and noise dynamics.

B. Numerical Illustration

We will now consider the case where

$$A(q,\theta) = (1 - 0.6q^{-1})^2(1 - 0.7q^{-1})^2,$$
$$B(q,\theta) = c\,(1 - 0.98\,e^{i\pi/20}q^{-1})(1 - 0.98\,e^{-i\pi/20}q^{-1}),$$
$$C(q,\theta) = (1 - 0.8q^{-1})(1 + 0.8q^{-1})$$

where c is adjusted so that the static gain $B(1,\theta)/A(1,\theta) = 1$. The noise variance $\sigma_e^2 = 0.001$. The Bode diagrams of the system dynamics $B(q,\theta)/A(q,\theta)$ and noise dynamics $C(q,\theta)/A(q,\theta)$ are shown in Fig. 2.

In the experiment design, the input power bound $p_u = 1$ and $n_\eta = 10$. The model uses $n_a = 4, n_b = 3$ and $n_c = 2$.

We will compare 4 different methods in a Monte Carlo study consisting of $M = 48$ experiments, where each experiment consists of $N = 10.000$ input/output samples. For each method, the sample mean of the performance degradation is computed for sampling time n as

$$\tilde{J}_n := \frac{n}{M} \sum_{k=1}^{M} J(\hat{\theta}_n^k, \theta)$$

where $\hat{\theta}_n^k$ is the parameter estimate at time n, $1 \le n \le N$, in experiment k, $1 \le k \le M$. We have normalized with n since then, at least according to the asymptotic theory, $\tilde{J}_n$ is independent of n.

The methods we compare are:

 i) The input is white noise with power p_u and $\hat{\theta}_n^k$ is computed using the MATLAB System Identification Toolbox (SITB) off-line prediction error method as implemented in the function `armax`.

 ii) The input is taken as white noise filtered through a filter which is the stable minimum phase spectral factor corresponding to the optimal spectrum of problem (69). $\hat{\theta}_n^k$ is computed off-line as in i).

 iii) Algorithm 5.1 is used. The recursive estimator uses the SITB algorithm `rarmax` and the optimal input design problem (69) is solved using YALMIP [60] and SeDuMi
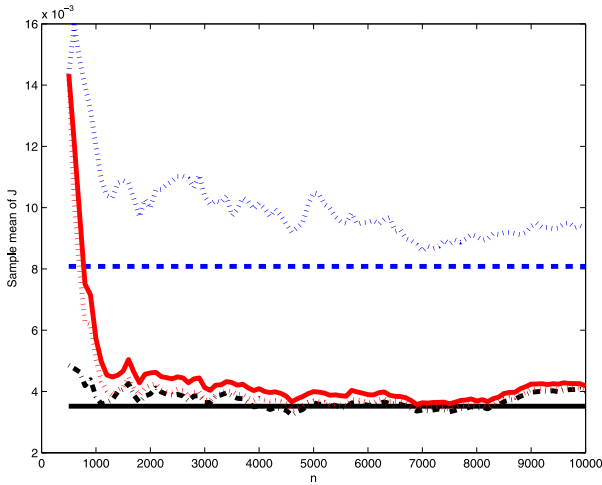
Fig. 3. Sample mean $\tilde{J}_n$ of the performance degradations over the different Monte Carlo experiments. Horizontal dashed blue and solid black lines correspond to the asymptotic theory when the input is white and optimal, respectively. Blue dotted: Method i)—White input/off-line. Black dashed: Method ii)—Optimal input/off-line. Red solid: Method iii)—Adaptive input/on-line. Red dotted: Method iv)—Adaptive input/off-line.

[61]. The first $n = n_{init} := 500$ samples uses the same white input as Method i). The algorithm is then initialized at $n = n_{init} := 500$ with the off-line estimate from Method i) (which uses a white input). The initial $\hat{R}_0$ is taken as the inverse of the estimated covariance matrix for the off-line estimate, normalized with the estimated noise variance. To obtain initial conditions for the states of the recursive estimator that correspond to this initial estimate and uncertainty, we use the fact that when recursive PEM is applied repeatedly to a data set, its estimate approaches that of the off-line estimate [36]. Thus `rarmax` is applied 51 times to the data set consisting of the first n_{init} samples and the final states from `rarmax` are taken as initial states when computing $\hat{\theta}_n^k$ for $n = n_{init} + 1$. For this start-up procedure, the off-line estimate is used as initial estimate and $\hat{R}_0$ is as described above. The initial states are taken to be zero. For $n = n_{init} + 1$ and onwards, $\hat{\theta}_n^k$ is computed using recursive PEM using the SITB function `rarmax`.

iv) The adaptive input generated in iii) is used but the parameter estimates are computed off-line as in i).

Thus, three different data generation mechanisms have been used: a white input, the optimal input and an adaptively generated input. The same white noise sequence has been used as noise $\{e_n\}$ in all cases. Also, the same white noise sequence has been used as source for the input in all three cases.

Fig. 3 shows the sample mean $\tilde{J}_n$ of the performance degradations over the different Monte Carlo experiments. Here we see that using a white input (Method i)) gives significantly worse performance than the other methods. As the methods using the adaptive input use the same input as Method i) prior to $n = 500$, they have the same performance as this method at $n = 500$ but then start to catch up with the performance of Method ii) (which uses the optimal input). The on-line estimate takes somewhat longer time to catch up than the off-line estimate (both use the same data).

VII. CONCLUSION

In this contribution, we have shown that an on-line estimator of system parameters controlling a parameter-dependent input generator has the same asymptotic behavior in a strong sense as an off-line estimator using experimental conditions set up under the hypothesis that the system is known. The result holds for stable SISO LTI systems and for a wide class of model structures.

As an application, we have proposed an optimal adaptive input design method for stable LTI systems based on the certainty equivalence principle. This is a formal development of the scheme outlined in [27], which establishes convergence and asymptotic efficiency, based on the results of [32]. The asymptotic theory is backed-up by a finite-sample simulation study.

In terms of contributions to recursive estimation per se, the condition on joint stability of the state transition matrices in question, phrased in [32], has been refined, and a relevant auxiliary result for block-triangular matrices has been given. In addition a clarification, needed for the proof of [32, Theorem 4.3] has been added.

APPENDIX

A. Proof of Lemma 4.1

The proof follows directly from the following general result:

Lemma A.1: Let $\mathcal{A}$ be a bounded family of 2×2 block-triangular matrices with fixed sized, square diagonal blocks, denoted by A_{11} and by A_{22}. Then the joint stability of the set of diagonal blocks A_{11} and by A_{22}, say $\mathcal{A}_{11}$ and $\mathcal{A}_{22}$, implies that $\mathcal{A}$ itself is jointly stable.

Proof: The conditions of the lemma trivially imply that the the time-varying linear system

$$x_{n+1}^1 = A_{11,n} x_n^1 + u_n^1, \tag{70}$$

$$x_{n+1}^2 = A_{21,n} x_n^1 + A_{22,n} x_n^2 + u_n^2 \tag{71}$$

is BIBO stable in L_∞. Hence, by [62], it follows that the joint spectral radius of $\mathcal{A}$ is less than 1.

B. Proof of Theorem 4.1

We start by observing that the adaptive system (25)–(31) has the format

$$\phi_{n+1} = A(x_n)\phi_n + B(x_n)\tilde{e}_n, \tag{72}$$

$$x_{n+1-} = x_n + \frac{1}{n+1} Q(\phi_{n+1}), \tag{73}$$

$$x_{n+1} = x_{n+1-} \quad \text{if} \quad x_{n+1-} \in D_0, \tag{74}$$

$$x_{n+1} = x_0 \quad \text{if} \quad x_{n+1-} \notin D_0, \tag{75}$$

where $\phi_n = \Phi_n, A = A^\Phi, B = B^\Phi, x_n = \hat{\theta}_n$, and $\tilde{e}_n = \begin{bmatrix} w_n & e_n \end{bmatrix}^T$, and where Q is a quadratic function. To avoid divergence, the estimates x_n are enforced to stay in a compact truncation domain D_0. In [32], this scheme was coined DFL scheme with resetting after [63], [34].

If x_n converges, the behavior of (72) will approach that of what is called the frozen-system, a parameter-dependent $\mathbb{R}^r$-valued vector-valued stochastic process $(\bar{\phi}_n(x))$, with $n \geq 0$

and $x \in D \subset \mathbb{R}^p$ denoting the parameter, where D is an open domain, defined by the state-space equation

$$\overline{\phi}_{n+1}(x) = A(x)\overline{\phi}_n(x) + B(x)e_n \qquad (76)$$

with some non-random initial condition $\overline{\phi}_0(x)$. The mean-field, defined as

$$G(x) = \lim_{n \to \infty} \mathrm{E}Q(\overline{\phi}_n(x)) \qquad (77)$$

plays a central role in the analysis. More precisely, the behavior of the *associated ODE*

$$\dot{z}_t = \frac{1}{t}G(z_t), \qquad z_s = \xi, s \geq 1 \qquad (78)$$

determine the asymptotic properties of the DFL-scheme. For the adaptive system (25)–(31), G corresponds to h in (35), and under the conditions of the theorem it follows that $G(x)$ is well defined. From v) in Lemma 4.2 it follows that the adaptive system we consider is a partially *stochastic Newton method* [32]. Splitting x as $x = \begin{bmatrix} (x^1)^T & (x^2)^T \end{bmatrix}^T$ where x^1 corresponds to θ, and Q accordingly, Theorem 4.4 in [32] gives

$$x_N^1 - x^{1*} = \frac{1}{N}\sum_{n=1}^{N} Q^1(\overline{\phi}_n(x^*)) + O_M(N^{-1/2-\delta}) \qquad (79)$$

with some $\delta > 0$, We will set aside the verification of the conditions for this theorem for the moment and proceed by noting that (79) yields a particularly useful insight into the relationship between recursive and off-line estimators. Namely, define the off-line estimator $\hat{x}_N$ of x^* as "the solution" of

$$U_N(x) = \sum_{n=1}^{N} Q^1(\overline{\phi}_n(x)) = 0 \qquad (80)$$

with respect to x. Then it is easy to see that the result given in [42] can be extended at no cost. Noting that the Jacobian matrix of the right-hand side of the associated ODE at $x = x^*$ is of the form given in (39), we get for the first component of $\hat{x}_N$

$$\hat{x}_N^1 - x^{1*} = \frac{1}{N}\sum_{n=1}^{N} Q^1(\overline{\phi}_n(x^*)) + O_M(N^{-1}). \qquad (81)$$

Note that the dominant term of the r.h.s is identical with the dominant term on the r.h.s. of (79). Let us write $\widehat{\widehat{x}}_N = x_N$ for the recursive estimator of x^*. Then we arrive at the following result, see [32, (4.7)]:

$$\widehat{\widehat{x}}_N^1 - \hat{x}_N^1 = O_M(N^{-1/2-\delta}). \qquad (82)$$

It follows, among others, that the asymptotic covariance matrix of the off-line estimator $\hat{x}_N^1$, if it exists, equals the asymptotic covariance matrix of the recursive estimator $\widehat{\widehat{x}}_N^1$.

The result (82) is exactly (41) that we would like to prove. What remains is thus to verify that the conditions of Theorem 4.4 in [32] are satisfied. Before this, however, a comment is due: It is claimed in [32] that Theorem 4.4 follows directly from Theorem 4.3 which follows directly from Theorem 4.2, which in turn directly follows from Theorem 4.1, for which a complete proof is given. To justify these claims we give a hint showing the interconnection between discrete time (Theorems

4.2–4.4) and continuous time results (Theorems 4.1). For this we note that the statement of Theorem 4.2 of [32] implies that Theorems 4.3 is equivalent to saying that, under the conditions of the theorem, we have

$$x_N - x^* = \int_{t=1}^{N} \frac{\partial y}{\partial \xi}(N, t, x^*)\frac{1}{t}Q(\overline{\phi}_t^c(x^*))dt + O_M(N^{-1/2-\delta}),$$

where $Q(\overline{\phi}_t^c(x^*))$ is a piecewise constant extension of $Q(\overline{\phi}_n(x^*))$ to real values t.

To see the equivalence of the two representations of $x_N - x^*$ we have to note only that $||\frac{\partial y}{\partial \xi}(N, t, x^*) - \frac{\partial y}{\partial \xi}(N, n, x^*)|| \leq C(\frac{n}{N})^{\alpha-1}\frac{1}{n}$, and then apply the moment inequality given as Theorem 1.1 in [64].

We are now ready to verify the conditions of [32, Theorem 4.3]. Starting with Condition 3.7, notice first that the elements of A^Φ and B^Φ are multinomials in the elements of the state-space matrices of the model and the input generator (3), and the derivatives of $A(\theta), B(\theta), C(\theta)$, and $D(\theta)$. This and Conditions 3.2, 4.3 and 4.4 imply that A^Φ and B^Φ are three times continuously differentiable. Furthermore, the joint stability condition in Condition 3.7 is implied by Lemma 4.1. Conditions 3.8 and 3.9 on the driving noise $\tilde{e}_n$ are trivially satisfied by the i.i.d. Conditions 2.2 and 2.3. In regards to [32, Condition 3.4], we notice that part (iii) is a trivial corollary of part (i), and hence can be removed. Parts (i) and (ii), save for (3.21), follow from Condition 4.6, where a minor typo is also corrected, relaxing the original formulation: convergence of $y(t, s, \xi)$ to x^* is required for $\xi \in D_0'$ rather than for $\xi \in D$.

For (3.21), it can be shown, using a compactness argument, that if the Jacobian matrix

$$A^* = \frac{\partial G(x)}{\partial x}\Big|_{x=x^*} \qquad (83)$$

is stable (having all its eigenvalues on left half of the complex plane), then

$$\left\|\frac{\partial}{\partial \xi}y(t, s, \xi)\right\| \leq C_0(s/t)^\alpha \qquad (84)$$

with some $C_0 \geq 1, \alpha > 0$ for all $\xi \in D_0'$ and $t \geq s \geq 1$ follows from the remaining components of Condition 3.4. This situation holds in this case as v) in Lemma 4.2 gives that all eigenvalues of A^* are -1.

Finally, we notice that the validity of Theorem 4.3 also requires a correction of the theorem. The correction is that we explicitly require that $G(x) = 0$ has a unique solution x^* in D, in accordance with Condition 3.3. In our setting, this follows from iv) in Lemma 4.2.

Now, in the case of adaptive experiment design the off-line estimator defined in the context of the general estimation method under (80) would be defined, according to (8) and (9), by

$$V_{\theta,N}(\theta; \eta(\theta)) = \sum_{n=0}^{N-1} \overline{\varepsilon}_{\theta,n}(\theta; \eta(\theta))\overline{\varepsilon}_n(\theta; \eta(\theta)) = 0. \qquad (85)$$

Let this (uncomputable) off-line estimator be denoted by $\hat{\overline{\theta}}_N(\eta)$. Then the arguments given in [42] can be applied step by step to

yield that

$$\hat{\hat{\theta}}_N(\eta) - \theta^* = -\frac{1}{N}(R^*)^{-1}\sum_{n=1}^{N}\bar{\varepsilon}_{\theta,n}(\theta^*;\eta(\theta^*))e_n + O_M(N^{-1}) \tag{86}$$

since $\bar{\varepsilon}_n(\theta^*;\eta(\theta^*)) = e_n$. Then Theorem 4.1 implies that $\hat{\hat{\theta}}_N(\eta) - \hat{\theta}_N = O_M(N^{-1/2-\delta})$.

Now if we consider the off-line estimator $\hat{\theta}_N(\eta^*)$, obtained under *optimal* experimental conditions defined by

$$V_{\theta,N}(\theta;\eta^*) = \sum_{n=0}^{N-1}\bar{\varepsilon}_{\theta,n}(\theta;\eta^*)\bar{\varepsilon}_n(\theta;\eta^*) = 0 \tag{87}$$

then, applying the arguments given in [42] once again, we get that $\hat{\theta}_N(\eta^*) - \theta^*$ can be written exactly as the right hand side of (86). Therefore, we have $\hat{\theta}_N(\eta) - \hat{\theta}_N(\eta^*) = O_M(N^{-1})$. It follows that $\hat{\theta}_N(\eta^*) - \hat{\theta}_N = O_M(N^{-1/2-\delta})$, as stated.

C. Proof of Theorem 5.1

The results follow from Theorem 4.1 and Theorem 4.1 if we can verify Conditions 4.3, 4.4 and 4.1. As noted before Theorem 5.1, A_z can be kept fix and since the basis functions $\{\mathcal{B}_k(q)\}_{k=1}^{n_\eta}$ are stable its spectral radius is less than one. Thus Condition 4.1 holds trivially.

Let us now examine the map from θ to η. Firstly, (55)–(56) imply that $M(\eta,\gamma,\theta)$ is continuously differentiable of any order with respect to θ on D_θ. Secondly, if the (primal) problem (44)–(45) is strictly feasible and bounded from below, and the solution is unique, [65, Theorem 1] gives that the solution is differentiable with respect to perturbations of M. The essence of the proof is that the equations (9) in [65] have a non-singular Jacobian permitting the implicit function theorem to be applied. The result of Theorem 1 in [65] can be extended by noting that the equations in (9) are continuously differentiable of any order, and hence the implicit function theorem gives that the solution is continuously differentiable of any order with respect to perturbations of M [66]. In summary, the map from θ to η, as defined by (44)–(45), is continuously differentiable of any order under the assumptions of the theorem. Thus Condition 4.4 has been verified.

Next, we study the map from η to the filter coefficients of $G_u(q,\eta)$, i.e. the spectral factorization step. For simplicity of exposition, we restrict our analysis to the FIR case where $\mathcal{B}_1(z) = 1/2$ and $\mathcal{B}_k(z) = z^{-k+1}$ for $k > 1$. The case of general rational stable basis functions can be handled along the same lines but is more involved. Consider

$$G_u(z) = \sum_{k=1}^{n_\eta} g_k z^{-(k-1)},$$

$$\Phi(z) = G_u(z)G_u(z^{-1}) = \sum_{k=1}^{n_\eta}\eta_k\left(z^{-(k-1)} + z^{(k-1)}\right).$$

We will use the implicit function theorem [66] to prove that the map from $\eta = \begin{bmatrix} \eta_1, & \dots & ,\eta_{n_\eta} \end{bmatrix}^T$ to $g = \begin{bmatrix} g_1, & \dots & ,g_{n_\eta} \end{bmatrix}^T$, defined by $\Phi(z) = G_u(z)G_u(z^{-1})$ is

continuously differentiable of any order. Firstly, the map from g to η is given by

$$\eta(g) := \oint_{|z|=1} G_u(z)G_u(z^{-1})\Gamma(z)\frac{dz}{z}$$

where here $\Gamma(z) = \begin{bmatrix} 1 & z^{-1} & \dots & z^{-n_\eta+1} \end{bmatrix}^T$. Differentiation under the integral sign, the Jacobian of this map is

$$J(g) : = \oint_{|z|=1}\left(G_u(z)\Gamma(z)\Gamma^T(z^{-1})\right.$$
$$\left. + G_u(z^{-1})\Gamma(z)\Gamma^T(z)\right)\frac{dz}{z}.$$

Let $g \neq 0$, let $\alpha = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{n_\eta} \end{bmatrix}^T$ and let $\alpha(z) = \alpha^T\Gamma(z)$ be the associated polynomial (in z^{-1}) of degree $n_\eta - 1$. Suppose that $J(g)\alpha = 0$ for an $g \neq 0$. This can be expressed

$$J(g)\alpha := \oint_{|z|=1}(G_u(z)\alpha(z^{-1}) + G_u(z^{-1})\alpha(z))\Gamma(z)\frac{dz}{z} = 0. \tag{88}$$

Here, $G_u(z)\alpha(z^{-1}) + G_u(z^{-1})\alpha(z)$ is a symmetric polynomial in $z^{-(n_\eta-1)},\dots,z^{n_\eta-1}$. Hence, expression (88) implies that this polynomial is identically zero. If $\alpha \neq 0$, it must hold that $\alpha(z) = G_u(z)$ since $G_u(z)$ and $G_u(z^{-1})$ are coprime. But then $G_u(z)\alpha(z^{-1}) + G_u(z^{-1})\alpha(z) = 2G_u(z)G_u(z^{-1})$ is nonzero, contradicting our assumption that $g \neq 0$. Hence, $\alpha = 0$ is the only solution to $J(g)\alpha = 0$ and $J(g)$ is non-singular. Using this and that $\eta(g)$ is continuously differentiable of any order, it follows from the implicit function theorem that the map from η to g is continuously differentiable of any order. This implies Condition 4.3.

The convergence of sample correlations of the type (58) follows in exactly the same way as (B.2) in [28]. This concludes the proof.

D. Proof of Lemma 5.1

We start with the positivity condition. With $\tilde{r} = \begin{bmatrix} \eta_{n_\eta} & \dots & \eta_2 \end{bmatrix}^T$, we can write the matrix in (48) as

$$\begin{bmatrix} Q & (C^u)^T \\ C^u & 2D^u \end{bmatrix} - \begin{bmatrix} A^u & B^u \end{bmatrix}^T Q \begin{bmatrix} A^u & B^u \end{bmatrix}$$

$$= \begin{bmatrix} Q & \tilde{r} \\ \tilde{r}^T & \eta_1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix}.$$

This is a positive definite matrix if we take $\tilde{r} = 0$ and Q to be diagonal with strictly monotonically increasing elements along the diagonal, and take η_1 to be greater than the maximal value of Q.

Maintaining $\tilde{r} = 0$, (52) gives

$$\tilde{R}(\theta;\eta) = \eta_1 \tilde{R}_1(\theta). \tag{89}$$

Take $\alpha \in \mathbb{R}^{p_\theta}$ to have unit norm. Then

$$\alpha^T\tilde{R}_1(\theta)\alpha = \frac{1}{\pi}\int_{-\pi}^{\pi}|\alpha(e^{i\omega},\theta)|^2 d\omega > 0 \tag{90}$$

where $\alpha(z,\theta) = \alpha^T\Lambda(z,\theta)$ is a stable rational function. The inequality follows since $|\alpha(e^{i\omega},\theta)|^2$ is positive and has at most

a finite number of zeros on the unit circle. Combining (89)–(90) gives that the minimum eigenvalue of $\tilde{R}(\theta; \eta)$ can be made as large as desired by picking η_1 large enough.

In summary, η_1 large enough and $\eta_2 = \ldots = \eta_{n_\eta} = 0$ ensures feasibility of the constraints in Lemma 5.1, and the lemma has been proven.

REFERENCES

[1] Y. Zhu, "System identification for process control: Recent expertise and outlook," *Int. J. Model., Identif. and Control*, vol. 6, no. 2, pp. 20–32, 2009.

[2] M. Barenthin, H. Jansson, and H. Hjalmarsson, "Applications of mixed $\mathcal{H}_\infty$ and $\mathcal{H}_2$ input design in identification," in *Proc. 16th World Congr. Autom. Control*. Prague, Czech Republik: IFAC, 2005, paper Tu-A13-TO/1.

[3] Y. Zhu, R. Patwardhanb, S. Wagnerb, and J. Zhaoa.

[4] H. Hjalmarsson, "System identification of complex and structured systems," *Eur. J. Control*, vol. 15, no. 4, pp. 275–310, 2009, plenary address, Eur. Control Conf.

[5] A. Atkinson and R. Bailey, "One hundred years of the design of experiments on and off the pages of Biometrika," *Biometrika*, vol. 88, pp. 53–97, 2001.

[6] 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 Comput. Biol.*, vol. 5, pp. 1–12, 2009.

[7] K. Bernaerts, K. P. M. Gysemans, T. N. Minh, and J. F. Van Impe, "Optimal experiment design for cardinal values estimation: Guidelines for data collection," *Int. J. Food Microbiol.*, vol. 100, pp. 153–165, 2005.

[8] H. A. Dror and D. M. Steinberg, "Robust experimental design for multivariate generalized linear models," *Technometrics*, vol. 48, pp. 520–529, 2006.

[9] J. Hahn, K. Hirano, and D. Karlan, "Adaptive experimental design using propensity score," *J. Bus. & Econom. Statist.*, vol. 29, pp. 96–108, 2011.

[10] T. Lohmann, H. G. Bock, and J. P. Schloeder, "Numerical methods for parameter estimation and optimal experiment design in chemical reaction systems," *Indust. & Eng. Chem. Res.*, vol. 31, pp. 54–57, 1992.

[11] L. Pronzato, "Optimal experimental design and some related control problems," *Automatica*, vol. 44, no. 2, pp. 303–325, Feb. 2008.

[12] H. Wynn, "The sequential generation of d-optimum experimental designs," *Ann. Math. Stat.*, vol. 41, pp. 1655–1664, 1970.

[13] H. Jansson and H. Hjalmarsson, "Input design via LMIs admitting frequency-wise model specifications in confidence regions," *IEEE Trans. Autom. Control*, vol. 50, no. 10, pp. 1534–1549, 2005.

[14] X. Bombois, G. Scorletti, M. Gevers, P. M. J. Van den Hof, and R. Hildebrand, "Least costly identification experiment for control," *Automatica*, vol. 42, no. 10, pp. 1651–1662, 2006.

[15] H. Hjalmarsson and H. Jansson, "Closed loop experiment design for linear time invariant dynamical systems via LMIs," *Automatica*, vol. 44, no. 3, pp. 623–636, 2008.

[16] R. Hildebrand, M. Gevers, and G. Solari, "Closed-loop optimal experiment design: Solution via moment extension," *IEEE Trans. Autom. Control*, vol. 60, no. 7, pp. 1731–1744, 2015.

[17] J. Rathouský and V. Havlena, "MPC-based approximate dual controller by information matrix maximization," *Int. J. Adapt. Control and Signal Process.*, vol. 27, no. 11, pp. 974–999, 2013.

[18] G. Marafioti, R. Bitmead, and M. Hovd, "Persistently exciting model predictive control," *Int. J. Adapt. Control and Signal Process.*, vol. 28, no. 6, pp. 536–552, 2013.

[19] C. Larsson, C. Rojas, X. Bombois, and H. Hjalmarsson, "Experimental evaluation of model predictive control with excitation (MPC-X) on an industrial depropanizer," *J. Process Control*, vol. 31, pp. 1–16, Jul. 2015.

[20] P. Valenzuela, C. Rojas, and H. Hjalmarsson, "Optimal input design for non-linear dynamic systems: A graph theory approach," in *Proc. 52st IEEE Conf. Decision and Control*, Florence, Italy, 2013.

[21] M. Forgione, X. Bombois, P. Van den Hof, and H. Hjalmarsson, "Experiment design for parameter estimation in nonlinear systems based on multilevel excitation," in *Eur. Control Conf.*, Strasbourg, France, 2014.

[22] A. De Cock, M. Gevers, and J. Schoukens, "A preliminary study on optimal input design for nonlinear systems," in *Proc. 52nd IEEE Conf. Decision and Control*, Florence, Italy, 2013.

[23] T. L. Lai, "Sequential analysis: Some classical problems and new challenges," *Statistica Sinica*, vol. 11, pp. 303–408, 2001.

[24] K. Lindqvist and H. Hjalmarsson, "Identification for control: Adaptive input design using convex optimization," in *Proc. IEEE Conf. Decision and Control*, Orlando, FL, Dec. 2001, pp. 4326–4331.

[25] V. M. Pérez, J. Renaud, and L. T. Watson, "Adaptive experimental design for construction of response surface approximations," in *Proc. 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conf.*, 2001, pp. AIAA-2001–1622.

[26] M. Sasena, M. Parkinson, P. Goovaerts, P. Papalambros, and M. Reed, "Adaptive experimental design applied to an ergonomics testing procedure," in *Proc. DETC'02, ASME 2002 Design Engineering Tech. Conf. and Computers and Information in Engineering Conf.*, 2002, p. DETC2002/DAC34091.

[27] L. Gerencsér and H. Hjalmarsson, "Adaptive input design in system identification," in *Proc. 44th IEEE Conf. Decision and Control and European Control Conf.*, Seville, Spain, Dec. 12–15 2005, pp. 4988–4993.

[28] 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, no. 3, pp. 623–633, Mar. 2009.

[29] L. Huang, H. Hjalmarsson, and L. Gerencsér, "Adaptive experiment design for ARMAX systems," in *Proc. 51st IEEE Conf. Decision and Control*, Maui, HI, 2012.

[30] H. Hjalmarsson, "From experiment design to closed loop control," *Automatica*, vol. 41, no. 3, pp. 393–438, Mar. 2005.

[31] L. Ljung, "Convergence analysis of parametric identification methods," *IEEE Trans. Autom. Control*, vol. AC-23, pp. 770–783, May 1978.

[32] L. Gerencsér, "A representation theorem for recursive estimators," *SIAM J. Control and Optimiz.*, vol. 44, no. 6, pp. 2123–2188, 2006.

[33] T. Lai and C. Wei, "Least-squares estimates in stochastic regression-models with applications to identification and control of dynamic systems," *Ann. Statist.*, vol. 10, no. 1, pp. 154–166, 1982.

[34] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Autom. Control*, vol. AC-22, no. 4, pp. 551–575, Aug. 1977.

[35] H. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer-Verlag, 2003.

[36] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.

[37] H.-F. Chen, *Stochastic Approximation and Its Applications*. Kluwer/Cambridge University Press, 2002.

[38] H.-F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*. Boston, MA: Birkhäuser, 1991.

[39] L. Gerencsér, "Rate of convergence of recursive estimators," *SIAM J. Control and Optimiz.*, vol. 30, no. 5, pp. 1200–1227, 1992.

[40] H.-F. Chen, "New approach to recursive identification for ARMAX systems," *IEEE Trans. Autom. Control*, vol. 55, pp. 868–879, 2010.

[41] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[42] L. Gerencsér, "On the martingale approximation of the estimation error of ARMA parameters," *Syst. Control Lett.*, vol. 15, pp. 417–423, 1990.

[43] L. Ljung and P. Caines, "Asymptotic normality of prediction error estimation for approximate system models," *Stochastics*, vol. 3, no. 1, pp. 29–46, 1979.

[44] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Berlin, Germany: Springer-Verlag, 1987.

[45] R. Almeida, E. Pueyo, J. Martínez, A. Rocha, and P. Laguna, "Quantification of the QT variability related to HRV: Robustness study facing automatic delineation and noise on the ECG," *Comput. Cardiol.*, pp. 769–772, 2004.

[46] R. Diversi, R. Guidorzi, and U. Soverini, "Identification of arx and ararx models in the presence of input and output noises," *Eur. J. Control*, vol. 16, p. 2010, 242–255.

[47] M. Haest, G. Bastin, M. Gevers, and V. Wertz, "ESPION: An expert system for system identification," *Automatica*, vol. 26, pp. 85–95, 1990.

[48] R. Kohn, "Asymptotic estimation and hypothesis testing results for vector linear time series models," *Econometrica*, vol. 47, pp. 1005–1030, 1979.

[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 Circul. Physiol.*, 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] A. Porta, G. Baselli, E. Caiani, A. Malliani, F. Lombardi, and S. Cerutti, "Quantifying electrocardiogram rt-rr variability interactions," *Med. & Biol. Eng. & Comput.*, vol. 36, pp. 27–34, 1998.

[52] F. Tjärnström and L. Ljung, "l_2 model reduction and variance reduction," *Automatica*, vol. 38, pp. 1517–1530, 2002.

[53] K. J. Åström and T. Söderström, "Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model," *IEEE Trans. Autom. Control*, vol. AC-19, pp. 769–773, 1974.

[54] H. Kushner, "Stochastic approximation: A survey," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, pp. 87–96, 2010.

[55] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[56] K. Åström and B. Wittenmark, *Computer Controlled Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[57] M. Gevers and L. Ljung, "Optimal experiment designs with respect to the intended model application," *Automatica*, vol. 22, no. 5, pp. 543–554, 1986.

[58] R. Hildebrand and G. Solari, "Identification for control: Optimal input intended to identify a minimum variance controller," *Automatica*, vol. 43, no. 5, pp. 758–767, 2007.

[59] J. Mårtensson, C. Rojas, and H. Hjalmarsson, "Conditions when minimum variance control is the optimal experiment for identifying a minimum variance controller," *Automatica*, vol. 47, no. 3, pp. 578–583, Mar. 2011.

[60] J. Löfberg, "Yalmip: A toolbox for modeling and optimization in MAT-LAB," in *Proc. CACSD Conf.*, Taipei, Taiwan, 2004. [Online]. Available: http://control.ee.ethz.ch/joloef/yalmip.php.

[61] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimiz. Methods and Softw.*, vol. 11, pp. 625–653, 1999.

[62] G. Michaletzky and L. Gerencsér, "BIBO stability of linear switching systems," *IEEE Trans. Autom. Control*, vol. 47, no. 11, pp. 1895–1898, 2002.

[63] D. P. Derevitskii and A. L. Fradkov, *Applied Theory of Discrete Adaptive Control Systems*. Moscow, Russia: Nauka, 1981.

[64] L. Gerencsér, "On a class of mixing processes," *Stochastics*, vol. 26, pp. 165–191, 1989.

[65] R. W. Freund and F. Jarre, "A sensitivity result for semidefinite programs," *Oper. Res. Lett.*, vol. 32, pp. 126–132, 2004.

[66] T. Saaty and J. Bram, *Nonlinear Mathematics*. New York: McGraw-Hill, 1964.

Håkan Hjalmarsson (M'98–SM'11–F'13) was born in 1962. He received the M.S. degree in electrical engineering in 1988, and the Licentiate and Ph.D. degrees in automatic control in 1990 and 1993, respectively, all from Linköping University, Linköping, Sweden.

He is a Professor at the School of Electrical Engineering, KTH, Stockholm, Sweden. He has held visiting research positions at California Institute of Technology, Louvain University, and at the University of Newcastle, Australia. His research interests include system identification, signal processing, control and estimation in communication networks and automated tuning of controllers.

Dr. Hjalmarsson has served as an Associate Editor for *Automatica* (1996–2001) and for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL (2005–2007) and been Guest Editor for the *European Journal of Control and Control Engineering Practice*. He is a Chair of the IFAC Coordinating Committee CC1 Systems and Signals. In 2001, he received the KTH Award for Outstanding Contribution to Undergraduate Education.

Lirong Huang received the Ph.D. degree from the Department of Mathematics and Statistics, University of Strathclyde, Glasgow, U.K.

As postdoctorate research associate, he worked with the Automatic Control Lab at KTH Royal Institute of Technology, Stockholm, Sweden, and then the Automatic Control Lab and Institute of Molecular Systems Biology, ETH Zurich, Switzerland. His research interests includes stochastic dynamics, modelling and algorithms.

László Gerencsér received the M.Sc. and Ph.D. degrees in mathematics from Eötvös Loránd University (ELTE), Budapest, Hungary, in 1969 and 1970, respectively.

Since 1970, he has been with MTA SZTAKI (Institute for Computer Science and Control of the Hungarian Academy of Sciences). He established and headed the Stochastic Systems Research Group from 1998 to 2012, after which he joined the Systems and Control Laboratory. Previously, he held a one-year visiting position at the Department of Mathematics, Chalmers University of Technology, Göteborg, Sweden, in 1986. From 1998 to 1991, he was a Visiting Professor with the Department of Electrical Engineering, McGill University, Montreal, QC, Canada. From 1997 to 2001, he had held a Széchenyi Professorship with the Eötvös Loránd University, Budapest, Hungary. His research interests include stochastic systems, recursive identification, stochastic adaptive control, input design, randomization methods, hidden Markov models, financial time series, change-detection, empirical characteristic functions (ECF), and Lévy processes.

Dr. Gerencsér was an Associate Editor for the *SIAM Journal of Control and Optimization* from 1998 to 2004, and from 2001 to 2004 he was an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He was awarded the title of Doctor of the Hungarian Academy of Sciences (in mathematics) in 2000 for recognition of his research activity the title. For his work in simultaneous perturbation stochastic approximation (SPSA), he received the Hart Prize for Independent Research from The Johns Hopkins University in 2002. He was the IPC Chair of the 19th International Symposium on Mathematical Theory of Networks and Systems (MTNS2010) Budapest in 2010.