Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Adaptive input design in system identification

László Gerencsér and Håkan Hjalmarsson[a]

Abstract— **Recently there have been significant developments in re-casting experiment design problems in system identification as convex optimization problems. The practical implementation of these methods is hampered by the fact that typically the "data" in the optimization problem depend on the to be identified system. In this contribution we propose an adaptive certainty equivalence solution based on a recursively identified model. The input design is adapted by taking one Newton step using data from the last identified model.**

I. Introduction

Experiment design in system identification has a long history. The classical approach is to use the input spectrum to shape the covariance matrix of the parameter estimates such that the objectives are met [7], [15].

When the high-order variance expression for identified frequency functions [27] appeared this allowed input design in the frequency domain which lead to a new flurry of activities [14], [18], [28], [8].

More recently the pendulum has swung back and several new results based on the classical approach have appeared [25], [6], [26], [16], [21], [3]. These recent contributions explore advances in convex optimization.

This paper contributes to this development. We will consider identification of time-discrete linear time-invariant systems within the prediction error framework, [28], [30]. Based on N observed input-output data points this framework delivers a frequency response estimate $G(e^{j\omega}, \hat{\theta}_N)$, where $\hat{\theta}_N$ is the prediction error estimate of a vector $\theta \in \mathbb{R}^n$ that parametrizes a set of transfer functions, $G(e^{j\omega}, \theta)$, together with an uncertainty region. The estimated model $G(e^{j\omega}, \hat{\theta}_N)$ will typically deviate from the system being estimated. In this contribution we will only consider the case where the true system belongs to the model class and hence only variance errors occurs. The variance errors are influenced by the second order statistics of the input used for the identification. No new input design problem will be considered. Instead we will focus on how to implement these methods in practice – the key problem being that the data required to solve the convex optimization problem associated with the input design problem in general requires the true system to be known. Here we propose an adaptive certainty equivalence solution based on a recursively identified model. In this approach the input design is changed in each iteration by taking one Newton step in a barrier method using data from the last identified model.

An adaptive method for this type of input design problem has previously been suggested in [26]. However, there the optimization problem is solved in each iteration. A two-step procedure has been proposed in [1] where in the first step data are collected over a short time interval with a fix input design. A model is identified from this data set and replaces the true system in an optimal input design. In the second step the input resulting from this design is used to generate a second data set on which the final model is estimated.

II. Multi-variable linear stochastic control systems: technical conditions

The set of real numbers will be denoted by $\mathbb{R}$, the p-dimensional Euclidean space will be denoted by $\mathbb{R}^p$. The Euclidean-norm of $x \in \mathbb{R}^p$ will be denoted by $|x|$. We shall often use subscripts to indicate partial derivatives. Let $(y_n), \quad 0 \leq n < \infty$ be a vector-valued, wide-sense stationary stochastic process defined by a finite-dimensional linear stochastic control-system:

$$y = G(\theta^*)u + H(\theta^*)e. \tag{1}$$

Condition 2.1: $G(\theta)$ and $H(\theta)$ are causal, rational transfer functions of the backward shift operator q^{-1}, moreover $H(\theta)$ is square, and inverse stable for $\theta \in D$ where $D \subset \mathbb{R}^p$ is an open domain.

Condition 2.2: There exists a minimal state-space realization of $H(\theta)$, say

$$H(\theta) = I + C(\theta)(q^{-1}I - A(\theta))^{-1}B(\theta), \tag{2}$$

such that the matrices $A(\theta), B(\theta)$ and $C(\theta)$ are three-times continuously differentiable functions of θ, and similarly for $G(\theta)$.

To characterize the input and the noise process we need the concept of L-mixing processes, elaborated in [9], which is a generalization of what is called "exponentially stable processes" in the system-identification literature, see Definition 3.1 in 8.3 of [5] or [29]. A slightly more restrictive class of processes, called L^+-mixing processes has been introduced in [10], and will be used in formulating our conditions. We will also need the following definition: a discrete-time $\mathbb{R}^p$-valued stochastic process (u_n) is *M-bounded* if for all $1 \leq q < \infty$

$$M_q(u) := \sup_{n \geq 0} \mathrm{E}^{1/q}\left[|u_n|^q\right] < \infty. \tag{3}$$

In this case we write $u_n = O_M(1)$. For a stochastic process $(z_n), n \geq 0$ and a positive sequence (c_n) we write $z_n = O_M(c_n)$ if $u_n = z_n/c_n = O_M(1)$. Finally, we say that (u_n) is *M^*-bounded* if for some $\varepsilon > 0$ we have $\sup_n \mathrm{E}\, \exp \varepsilon |u_n|^2 < \infty$.

Condition 2.3: The input process $(u_n), \quad n \geq 0$ and the noise process is $(e_n), \quad 0 \leq n \geq 0$ are both second-order stationary stochastic processes such that they are both in class M^*, and they are both L^+-mixing. In addition (e_n)

0-7803-9568-9/05/$20.00 ©2005 IEEE

is a martingale-difference process with respect to $(\mathcal{F}_n)$ with constant conditional covariance:

$$\mathrm{E}\,[e_n|\mathcal{F}_{n-1}] = 0, \qquad \mathrm{E}\,[e_n e_n^T|\mathcal{F}_{n-1}] = \Lambda^*$$

almost surely, with $\Lambda^* > 0$.

Let $\theta \in D_\theta$ where D_θ is an open set and let Λ be a symmetric positive definite matrix. Define the second order stationary process $\overline{\varepsilon}(\theta) = H^{-1}(\theta)(y - G(\theta)u)$. The overline indicates that we deal with frozen-parameter processes. Define the asymptotic cost function by

$$W(\theta, \Lambda) = \lim_{n\to\infty} \frac{1}{2}\mathrm{E}\,[\overline{\varepsilon}_n^T(\theta)\Lambda^{-1}\overline{\varepsilon}_n(\theta)] + \frac{1}{2}\log\ \det\Lambda. \quad (4)$$

The Hessian of W with respect to θ at (θ^*, Λ^*) is

$$R^* = W_{\theta\theta}(\theta^*, \Lambda^*) = \lim_{n\to\infty} \mathrm{E}\,[\overline{\varepsilon}_{\theta,n}^T(\theta^*)(\Lambda^*)^{-1}\overline{\varepsilon}_{\theta,n}(\theta^*)]. \quad (5)$$

Condition 2.4: The equation

$$W_\theta(\theta, \Lambda) = \lim_{n\to\infty} \mathrm{E}\,[\overline{\varepsilon}_{\theta,n}^T(\theta)(\Lambda)^{-1}\overline{\varepsilon}_n(\theta)] = 0 \quad (6)$$

has a unique solution $\theta = \theta^*$ in D_θ for any symmetric, positive definite Λ and the Hessian-matrix $R^* = W_{\theta\theta}(\theta^*, \Lambda^*)$ is positive definite.

III. Input design

In this section we will outline the basic features of the input design formulations used in, e.g., [21], [3]. For simplicity of presentation we will consider single-input/single-output systems only. However, the main results of the paper will apply to multivariable systems.

Consider that we are interested in some scalar *performance criterion* J which depends on the system, i.e. $J = J(\theta^*)$. J may for example represent a weighted sum of the frequency responses at some given frequencies or a pole or zero location.

The accuracy of the corresponding estimate $\hat{J}_N = J(\hat{\theta}_n)$ is asymptotically (in n) given by Gauss' approximation formula

$$Var(\hat{J}_n) \approx \frac{1}{n}\,[J'(\theta^*)]^T\,P(\theta^*)J'(\theta^*) \quad (7)$$

where the asymptotic covariance matrix P is defined as

$$P(\theta^*) = \lim_{n\to\infty}\ n\mathrm{E}\left[(\hat{\theta}_n - \theta^*)(\hat{\theta}_n - \theta^*)^T\right] \quad (8)$$

From (7) it is clear that $P(\theta^*)$ plays a crucial role in experiment design. It is through this quantity that we can influence the accuracy of $\hat{J}_n$.

The way the input influences $P(\theta^*)$ is through its spectrum. In fact the inverse of P depends affinely on the input spectrum under the assumption that the data is collected in open-loop [21], [28]. Hence, in open loop operation the input power spectrum $\Phi_u(\omega)$ is our design variable.

A typical *input design* problem could be to, for a given experiment length N, find the minimal input power required to guarantee a certain accuracy γ_2 of $\hat{J}_N$, i.e.

$$\begin{array}{ll} \underset{\Phi_u}{\text{minimize}} & \gamma_1 \\ \text{subject to} & \mathrm{E}[u_n^2] \le \gamma_1 \\ & Var(\hat{J}_N) \le \gamma_2 \\ & \Phi_u(\omega) \ge 0 \end{array} \quad (9)$$

As stated, the above problem is untractable from an optimization point of view since the free variable Φ_u is infinite dimensional and since $Var(\hat{J}_N)$ is non-convex with respect to Φ_u.

There are several ways to handle the first issue, see [21]. Here we will employ a *finite dimensional* parametrization of Φ_u

$$\Phi_u = \sum_{k=-(M-1)}^{M-1} \tilde{c}_{|k|}\ \mathcal{B}_k \quad (10)$$

where M is a finite positive integer. Here the $\mathcal{B}_k$-s are known basis functions and $\tilde{c} = [\tilde{c}_0, \tilde{c}_1, \ldots, \tilde{c}_M]^T$ is the free variable. For the particular choice $\mathcal{B}_k(e^{j\omega}) = e^{-j\omega k}$, the coefficients $\tilde{c}_k$ have the interpretation as auto-correlations.

The positiveness constraint on the input spectrum that appear in (9) translates into a linear matrix inequality (LMI) in $\tilde{c}$ and a new $M \times M$ matrix valued variable $Q = Q^T$ by way of the Kalman-Yakubovich-Popov (KYP) lemma, see e.g. [31]. We express this constraint as

$$K(Q, \tilde{c}) \ge 0 \quad (11)$$

Notice also that the input power constraint that appear in (9) is linear in $\tilde{c}$ since

$$\mathrm{E}[u_n^2] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_u(\omega)d\omega$$

We can thus write $\mathrm{E}[u_n^2] = b^T\tilde{c}$ where $b \in \mathbb{R}^M$ is a vector depending on the basis functions.

The *convexification* of $Var(\hat{J}_N)$ can be carried out using the Schur-complement which leads to that the two input design problems discussed above are equivalent to the following program

$$\begin{array}{ll} \underset{\gamma_1,\tilde{c},Q}{\text{minimize}} & \gamma_1 \\ \text{subject to} & b^T\tilde{c} \le \gamma_1 \\ & \begin{bmatrix} \gamma_2 & [J'(\theta^*)]^T \\ J'(\theta^*) & P^{-1}(\theta^*, \tilde{c}) \end{bmatrix} \ge 0 \\ & K(Q, \tilde{c}) \ge 0 \end{array} \quad (12)$$

where the argument $\tilde{c}$ has been added to P to indicate the dependence on the input spectrum. Notice that the first matrix inequality is an LMI in $\tilde{c}$ since $P^{-1}(\theta^*, \tilde{c})$ depends affinely on $\tilde{c}$. Thus (12) is a convex program. It is possible to extend and modify this basic problem formulation in a number of ways, see [21] for details. From an optimization point of view, however, the resulting problems have structures similar to (12) so for the considerations in this paper it is sufficient to study the above problem. Also closed loop experiment design formulations lead to the same type of convex optimization problems [22].

Let $\eta = [\tilde{c}^T,\ \mathrm{Vec}(Q)^T,\ \gamma_i]^T$. The actual input is computed by the state-space equation

$$\begin{aligned} z_{n+1} &= A_z(\eta)z_n + B_z(\eta)w_n \\ u_n &= C_z(\hat{\eta})z_n + D_z(\eta)w_n \end{aligned} \quad (13)$$

where the matrices A_z, B_z, C_z and D_z are obtained by spectral factorization of (10), and w_n is white noise.

For fixed θ^* finding a solution to (12) amounts to solving for a point which satisfies the Karush-Kuhn-Tucker conditions, which can be written as

$$F(\theta^*, \eta) = 0. \tag{14}$$

for some function F (see [4] for details). The latter can be solved by Newton's iteration. We assume that F includes the weight matrix that is needed for a Newton's iteration, so that we can use the recursion:

$$\hat{\eta}_n = \hat{\eta}_{n-1} + \lambda_n F(\theta^*, \hat{\eta}_{n-1}). \tag{15}$$

The *optimal* η will be denoted by η^*.

IV. Adaptive input design

The key problem with (12) is that θ^* is in fact unknown. However, if a parameter estimate is available it is natural to use the certainty equivalence solution to (12). For a *fixed* input, given, say by η, a Newton-type recursive prediction error estimate of θ^* is defined as follows:

$$\begin{aligned} R_n &= R_{n-1} + \frac{1}{n}(\varepsilon_{\theta,n}\Lambda_n^{-1}\varepsilon_{\theta,n}^T - R_n^{-1}) \\ \Lambda_n &= \Lambda_{n-1} + \frac{1}{n}(\varepsilon_n\varepsilon_n^T - \Lambda_{n-1}) \\ \hat{\theta}_n &= \hat{\theta}_{n-1} - \frac{1}{n}R_n^{-1}\varepsilon_{\theta,n}\Lambda_n^{-1}\varepsilon_n, \end{aligned} \tag{16}$$

Here ε_n and $\varepsilon_{\theta,n}$ are on-line estimates of $\overline{\varepsilon}_n(\hat{\theta}_{n-1})$ and $\overline{\varepsilon}_{\theta,n}(\hat{\theta}_{n-1})$, respectively. They can be generated by a finite dimensional linear system with state matrices depending on the time-varying parameters $\hat{\theta}_k$. Setting

$$\psi = (\theta, \mathrm{Vec}\Lambda, \mathrm{Vec}R)$$

a shorthand notation for algorithm (16) will be

$$\hat{\psi}_n = \hat{\psi}_{n-1} - \frac{1}{n}H(\hat{\psi}_{n-1}, \eta;\ \varepsilon_{\theta,n}, \varepsilon_n). \tag{17}$$

A simple *adaptive algorithm* would be to solve in each time-step n the equation $F(\hat{\theta}_n, \eta) = 0$ for η. However, for short sampling rates, this may be computationally prohibitive. Instead, one may take a single Newton-step towards the solution to (14) for each time-step. Thus we get the following *adaptive input design* algorithm, in which the frozen parameters η in (17) and θ in the recursion for η are replaced by their on-line estimates:

$$\hat{\psi}_n = \hat{\psi}_{n-1} - \frac{1}{n}H(\hat{\psi}_{n-1}, \eta_{n-1};\ \varepsilon_{\theta,n}, \varepsilon_n), \tag{18}$$

$$\hat{\eta}_n = \hat{\eta}_{n-1} + \lambda_n F(\hat{\psi}_n, \hat{\eta}_{n-1}). \tag{19}$$

Here λ_n is the step-size for the algorithm. Since no noise is explicitly present in this recursion, it may seem reasonable to use a fixed step size $\lambda_n = \lambda$. On the other hand the *analysis* of the joint algorithm is much simpler using existing results if we use the standard stepsize $1/n$. It will be argued below, that the asymptotic covariance matrix of $\hat{\theta}_n$ is the same for both choices.

The actual input is computed by the time-varying version of the filter (13). Equations (18), (19) together with the time-varying state space systems generating $\varepsilon_n, \varepsilon_{\theta,n}$ and u_n can be seen as a stochastic approximation procedure, which can be analyzed via the methods of [10] and [11].

The search domains or truncation domains for the systems parameters, the noise covariance and the input parameters are denoted by $D_{\theta 0}, D_{\Lambda 0}$ and $D_{\eta 0}$, respectively. They are compact domains inside the respective feasible open sets, and containing the optimal values in their interiors. In choosing the truncation domain $D_{\eta 0}$ it is critical for the algorithm that $\tilde{c}_n$ corresponds to a spectrum.

Boundedness of the estimates will be enforced by a *resetting* mechanism: if say $\hat{\theta}_n$ would leave $D_{\theta 0}$, then we redefine $\hat{\theta}_n$ to be $\hat{\theta}_0$, see (3.34) of [11]. Conditions for the position of $\hat{\theta}_0$ relative to $D_{\theta 0}$ are given in Condition 3.4 of [11].

We need to strengthen the persistent excitation condition $R^* > 0$. To emphasize that the input, and thus the estimated innovation depends on η write $\overline{\varepsilon}_n(\theta) = \overline{\varepsilon}_n(\theta;\eta) = \overline{\varepsilon}_n(\theta;\eta,\theta^*)$ and define, assuming that the true system parameter is θ,

$$R^*(\theta,\eta;\Lambda) = \lim_{n\to\infty} \mathrm{E}\ [\overline{\varepsilon}_{\theta,n}^T(\theta;\eta,\theta)\Lambda^{-1}\overline{\varepsilon}_{\theta n}(\theta;\eta,\theta)].$$

Condition 4.1: The search domains $D_{\theta 0}, D_{\Lambda 0}$ and $D_{\eta 0}$ are such that for some $0 < \kappa_1 < \kappa_2$ we have for all set of feasible parameters

$$\kappa_1 I \le R^*(\theta,\eta;\Lambda) \le \kappa_2 I.$$

To ensure the stability of the time-varying filters generating the processes ε_n, $\varepsilon_{\theta,n}$ and u_n we need the following condition, which can be often satisfied by an appropriate state-space realization, see the discussion in [11]:

Condition 4.2: The family of matrices $A(\theta),\ \theta \in D_{\theta 0}$ is jointly stable, in the sense that there exist a single symmetric positive definite $r \times r$ matrix V and $0 < \lambda < 1$ such that for all $\theta \in D_{\theta,0}$

$$A^T(\theta)VA(\theta) \le \lambda V.$$

Similarly, the family of matrices $A_z(\eta),\ \eta \in D_{\eta 0}$ is jointly stable.

V. The asymptotic covariance matrix

First we consider the problem of determining the *asymptotic covariance matrix* of recursive estimators, a problem that has been discussed in a number of works in a partially rigorous manner. A rigorous result has been derived in Theorem 13, Chapter 4.5,Part II of [2] in a *series* model, where the initial time tends to infinity, and thus the probability of exiting the truncation domain tends to 0, assuming a Markovian state dynamics. For results in the context of weak convergence theory see [24]. A recent general result for recursive estimation processes driven by a mixing random field and equipped with a resetting mechanism has been given in [11]. The main advance of this result relative to the cited result of [2] is that the asymptotic covariance matrix is obtained for a *single* process. We also get a rate of convergence for the covariance-matrix sequence, which is useful in applications such as the analysis of performance

degradation to statistical parametric uncertainty. Set

$$R(\theta,\Lambda) = \lim_{n\to\infty} \mathrm{E}\,[\overline{\varepsilon}^T_{\theta,n}(\theta)\Lambda^{-1}\overline{\varepsilon}_{\theta n}(\theta)]$$
$$\Lambda(\theta) = \lim_{n\to\infty} \mathrm{E}\,[\overline{\varepsilon}_n(\theta)\overline{\varepsilon}^T_n(\theta)].$$

We will use the notation $\Lambda^* = \Lambda(\theta^*)$ and $R^* = R(\theta^*,\Lambda^*)$. It is well known [28] that $P(\theta^*) = (R^*)^{-1}$. For the sake of clarity we consider first the recursive estimation method (16) or (17) with a fixed input design η, with the technical conditions stated exactly. The following result follows from [11]:

Theorem 5.1: Consider the full RPE estimator (16) modified with the resetting mechanism described above. Assume that Conditions 2.1, 2.2, 2.3, 2.4, 4.2 are satisfied, and the initial conditions are appropriately positioned relative to truncation domains (cf. Condition 3.4 of [11]). Then the asymptotic covariance-matrix of the error process $(\hat{\theta}_t - \theta^*)$, defined by

$$S^* = \lim_{n\to\infty} n\mathrm{E}[(\hat{\theta}_n - \theta^*)(\hat{\theta}_n - \theta^*)^T],$$

exists, and we have $S^* = (R^*)^{-1}$. More exactly we have with some $\varepsilon_2 > 0$

$$\mathrm{E}[(\hat{\theta}_n - \theta^*)(\hat{\theta}_n - \theta^*)^T] = \frac{1}{n}(R^*)^{-1} + O(n^{-1-\varepsilon_2}).$$

An analogous result holds for the extended parameter vector $\psi = (\theta, \mathrm{Vec}\Lambda, \mathrm{Vec}R)$.

The results of [11] are also applicable for the adaptive input design method given by (18) and (19). First, fix η, and write $\overline{\varepsilon}_n(\theta) = \overline{\varepsilon}_n(\theta;\eta) = \overline{\varepsilon}_n(\theta;\eta,\theta^*)$. Then the asymptotic covariance matrix of $\hat{\theta}_n$ will be written as

$$S^*(\eta) = S^*(\eta;\theta^*,\Lambda^*) = R^*(\theta^*,\eta;\Lambda^*)^{-1}.$$

It should be noted that the function $S^*(\eta;\theta^*,\Lambda^*)$ is *explicitly* computable. The associated ODE has the following structure:

$$\begin{aligned}\dot{y}_t &= \frac{1}{t}G(y_t,v_t)\\ \dot{v}_t &= \frac{1}{t}F(y_t,v_t),\end{aligned} \tag{20}$$

where y corresponds to ψ and v corresponds to η. Here G stands for the expectation of the correction term on right hand side of (16), while F has been introduced in (14).

Since for each fixed v we have $G(\psi^*,v) = 0$, the Jacobian of the right hand side is a block-triangular 2×2 matrix with stable diagonal blocks at the equilibrium point. Thus the local asymptotic stability of the joint ODE follows. The asymptotic covariance matrix of the sample means of the correction terms in (18) and (19) evaluated at the optimal parameter values, i.e. $H(\psi^*,\eta^*;\ \overline{\varepsilon}_{\theta,n}(\theta^*),\mathrm{e}_n)$ and $F(\psi^*,\eta^*)$ is a 2×2 block matrix with the only non-zero block at position $(1,1)$. Applying the results of [11] we get that the asymptotic covariance matrix of $(\hat{\theta}_n,\hat{\eta}_n)$ satisfies a Lyapunov-equation the $(1,1)$ block of which is identical with the Lyapunov-equation given for the asymptotic covariance matrix of $\hat{\psi}_n$ when $\eta = \eta^*$ is fixed. Thus we get the following result:

Proposition: The adaptive input design algorithm (18) and (19) solves the optimal input design problem. In particular we have

$$\lim_{n\to\infty} n\mathrm{E}[(\hat{\theta}_n - \theta^*)(\hat{\theta}_n - \theta^*)^T] = S(\eta^*), \tag{21}$$

In addition we get that $\hat{\psi}_n$ and $\hat{\eta}_n$ are asymptotically uncorrelated.

Remark: A natural alternative would be to use a fixed step-size in the deterministic recursion. Recursive estimation methods which are a mixture of a stochastic and a deterministic procedure have been discussed heuristically back in [12]. Since then a number of techniques have been developed, and a rigorous analysis have become feasible. The advantage of using a mixed algorithm is that the asymptotic covariance matrix of the first component $\hat{\psi}_n$ would be generally smaller than for the standard stochastic approximation method using stepsize $1/n$ for both components. However, if the two components of the joint parameter estimator are asymptotically uncorrelated, then it is easily seen that no reduction is achieved.

VI. Misspecified models

When the true system is not in the model class we face a completely new situation. First, the optimal θ^*, defined as the solution of an asymptotic ML equation, does depend on the input. Let it be denoted by $\theta^*(\eta)$. An important aspect of the input design problem is then to determine η such that $\theta^*(\eta)$ models the system properties of interest. In *identification for control* [17] criteria of the type

$$J(\theta) = \int_0^{2\pi} ||(G(e^{i\omega},\theta) - G^*(e^{i\omega}))U(e^{i\omega})||^2 W(e^{i\omega})d\omega, \tag{22}$$

where $G^*(e^{i\omega})$ is the true system, and where W is a suitably chosen frequency weighting filter which also may depend on the true system, are important. In this situation, the performance index $J(\theta)$ can not be expressed explicitly in terms of the assumed system parameters and input spectrum, rather we need to resort to experiments. After some algebra and simplifications we arrive at the following non-standard stochastic approximation problem: solve

$$F(\eta) = f(\mathrm{E}Q(\overline{\varepsilon}(\eta)) = 0, \tag{23}$$

where f, Q are explicitly known smooth functions, and $\overline{\varepsilon}(\eta)$ is a computable process. We call this problem a *non-linear stochastic approximation* problem.

As a simple benchmark example consider the following problem: let f be a scalar-valued function defined on R^p, and let $a \in R^p$ be an unknown vector. Assume that for any scalar η we can take measurements

$$y_n(\eta) = a\eta + e_n, \tag{24}$$

where e_n is a zero mean i.i.d. sequence. Solve for η the equation

$$F(\eta) = f(a\eta) = 0 \tag{25}$$

using the noisy measurements y_n.

A possible solution to the above general problem can be obtained by evaluating $F(\eta)$ with increasing accuracy by taking an increasing number of samples to approximate the

expectation which is the argument of f. A mathematically more attractive solution is obtained if we approximate the required expectation by an on-line computable average called z_n. Then we get a recursive method given by the pair of equations:

$$\begin{aligned}\eta_{n+1} =& \eta_n + \frac{1}{r} f(z_{n+1}) \\ z_{n+1} =& \frac{1}{n+1} \sum_{r=1}^{n+1} Q_r(\overline{\varepsilon}(\eta_{r-1})).\end{aligned} \tag{26}$$

Using a recursive form for computing z_{n+1} a stochastic approximation process the associated ODE of which is

$$\begin{aligned}\dot{\eta}_t &= f(z_t) \\ \dot{z}_t &= g(\eta_t) - z_t\end{aligned} \tag{27}$$

with $g(\eta) = \mathrm{E}Q(\overline{\varepsilon}(\eta))$.

Theorem 6.1: Assume that the differential equation

$$\dot{\eta}_t = F(\eta_t) = f(g(\eta_t)) \tag{28}$$

is exponentially asymptotically stable at η^* and the eigenvalues of $(\partial/\partial\eta)F(\eta)|_{\eta=\eta^*}$ are real. Then the extended ODE (27) is also exponentially asymptotically stable at $(\eta^*, g(\eta^*))$.

Proof: The Jacobian of (27) is

$$\begin{pmatrix} 0 & f_z(z^*) \\ g_\eta(\eta^*) & -I \end{pmatrix},$$

where subscripts denote partial derivatives. Let

$$B = f_z(z^*) \quad \text{and} \quad C = g_\eta(\eta^*).$$

The the assumption is that BC is stable, i.e. it has all its eigenvalues on the left half plane. The determinant of the above matrix is $\det BC \neq 0$, thus 0 is not an eigenvalue.

Consider the Schur-component

$$\det(-I - \lambda I + C\lambda^{-1}B).$$

Put this equal to zero and multiply by $-\lambda$. Then we have to solve

$$\det(\lambda^2 I + \lambda I - CB) = 0.$$

Now take a coordinate transformation that takes CB into its Jordan-form. The the above determinant can be factored according to the Jordan-blocks. Let $J(\mu)$ be a Jordan-block of CB which has μ in its diagonal, and μ is an eigenvalue of CB. Then its determinant is μ^k where k is the size of the block. Thus the corresponding determinant will become $(\lambda^2 + \lambda - \mu)^k$. Since the set of non-zero eigenvalues of BC and CB coincide, we have that μ is either 0 or negative by assumption. The case $\mu = 0$ would imply $\lambda = 0$, which has already been excluded. Thus the roots of $\lambda^2 + \lambda - \mu$ lie in the left half plane, as stated.

VII. Numerical illustration

In this section we present results when the algorithm (16), (19) and (13) was applied to the problem of estimating the static gain of the system

$$G = \frac{2q^{-1} + q^{-2}}{1 - 1.0607q^{-1} + 0.5625q^{-2}}$$

The objective is to estimate the static gain to within an accuracy of 0.01 using minimum input power over an experiment of length $N = 50$. The update mechanism (19) for the input design problem corresponds to one iteration in SDPT3 [23] (Version 3.0) which is a primal-dual interior point algorithm which uses the path following paradigm.

First an initial experiment of length $N = 25$ using a white noise input with unit variance was conducted. The corresponding parameter estimate, together with the input design corresponding to this parameter estimate were used as initial values in the algorithm.

The algorithm (16), (19) and (13) generated static gain estimates with a variance of 0.0173 using a mean input power of 1.38 over 50 Monte-Carlo runs. This can be compared with the optimal design for which the variance of static gain estimates was 0.0231 using a mean input power of 1.18. The reason why neither design reaches the desired performance specification is due to finite sample effects not covered by the asymptotic theory in Section IV. In Figure 1 the input for one simulation is compared to the optimal input. Clearly the adaptively generated input converges to the optimal one.

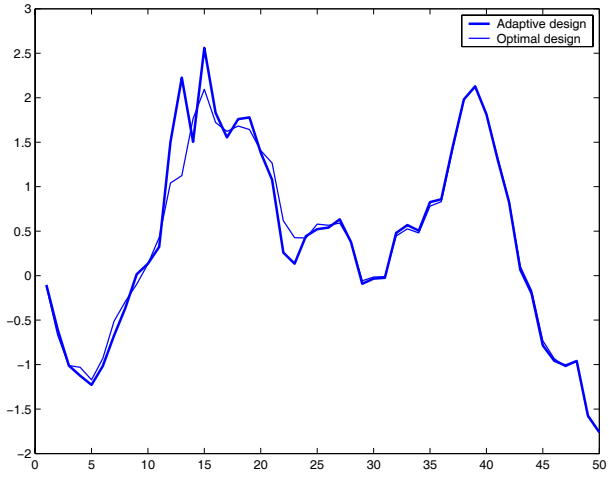

Fig. 1. Adaptive input compared to the optimal input.

VIII. Conclusions

A certainty equivalence based adaptive input design method has been presented and its asymptotical statistical properties have been analyzed under the assumption that the true system is in the model set. Performance degradation due to statistical uncertainty, also called regret, is of great interest in adaptive prediction, adaptive input design and adaptive control of stochastic systems. These results can be also applied in the context of identification for control, see [13], [19], [20].

We have also presented a novel approach to adaptive input design when the true system is not in the model class. The idea is to rely on experiments to compute necessary quantities and results in a non-standard SA problem.

REFERENCES

[1] M. Barenthin, H. Jansson, and H. Hjalmarsson. Applications of mixed $\mathcal{H}_\infty$ and $\mathcal{H}_2$ input design in identification. In *16th World Congress on Automatic Control*, Praha, 2005. IFAC.

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

[3] X. Bombois, G. Scorletti, M. Gevers, R. Hildebrand, and P. Van den Hof. Cheapest open-loop identification for control. In *Conference on Decision and Control*, Paradise Island, Bahamas, December 2004. IEEE.

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2003.

[5] P.E. Caines. *Linear Stochastic Systems*. Wiley, 1988.

[6] B. L. Cooley and J. H. Lee. Control-relevant experiment design for multivariable systems described by expansions in orthonormal bases. *Automatica*, 37:273–281, 2001.

[7] V. V. Fedorov. *Theory of Optimal Experiments, volume 12 of* Probability and Mathematical Statistics. Academic Press, 1972.

[8] U. Forssell and L. Ljung. Some results on optimal experiment design. *Automatica*, 36(5):749–756, May 2000.

[9] L. Gerencsér. On a class of mixing processes. *Stochastics*, 26:165–191, 1989.

[10] L. Gerencsér. Rate of convergence of recursive estimators. *SIAM J. Control and Optimization*, 30(5):1200–1227, 1992.

[11] L. Gerencsér. A representation theorem for the error of recursive estimators. *SIAM J. Control and Optimization*, 2005, to appear.

[12] L. Gerencsér, Zs. Lipcsey, and Zs. Vágó. Convergence of mixed type (stochastic-deterministic) recursive estimation methods. In L. Ljung and K.J. Åström, editors, *Bridge Between Control Science and Technology. Proc. of the 9th Triennal World Congress of IFAC, Budapest*, pages 1035–1038. Pergamon Press, Oxford, 1984.

[13] M. Gevers. Towards a joint design of identification and control ? In H.L. Trentelman and J.C. Willems, editors, *Essays on Control : Perspectives in the Theory and its Applications*, pages 111–151. Birkhäuser, 1993.

[14] M. Gevers and L. Ljung. Optimal experiment designs with respect to the intended model application. *Automatica*, 22:543–554, 1986.

[15] G.C. Goodwin and R.L. Payne. *Dynamic System Identification: Experiment Design and Data Analysis, volume 136 of* Mathematics in Science and Engineering. Academic Press, 1977.

[16] R. Hildebrand and M. Gevers. Identification for control: Optimal input design with respect to a worst case ν-gap cost functionhow to decompose semidefinite discrete-time algebraic riccati equations. *SIAM Journal on Control and Optimization*, 41(5):1586–1608, 2003.

[17] H. Hjalmarsson. From experiment design to closed loop control. *Automatica*, 41(3):393–438, March 2005.

[18] H. Hjalmarsson, M. Gevers, and F. De Bruyne. For model-based control design, closed loop identification gives better performance. *Automatica*, 32(12):1659–1673, 1996.

[19] H. Hjalmarsson, M. Gevers, and F. De Bruyne. For model based control design criteria, closed loop identification gives better performance. *Automatica*, 32:1659–1673, 1996.

[20] H. Hjalmarsson and K. Lindqvist. Identification for control: L_2 and L_∞ methods. In *Conference on Decision and Control*, Orlando, Florida, USA, December 2001. IEEE.

[21] H. Jansson and H. Hjalmarsson. A general framework for mixed $\mathcal{H}_\infty$ and $\mathcal{H}_2$ input design. *IEEE Transactions for Automatic Control*, 2005. To appear.

[22] Henrik Jansson. *Experiment Design with Applications in Identification for Control*. PhD thesis, S3–Automatic Control, Royal Institute of Technology, Stockholm, Sweden, 2004.

[23] R. H. Tütüncü K. C. Toh, M. J. Todd. Sdpt: a matlab software package for semidefinite programming. *Optimization Methods and Software*, pages 545–581, 1999.

[24] H.J. Kushner and G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer Verlag. New York, 1997.

[25] K. Lindqvist and H. Hjalmarsson. Optimal input design using linear matrix inequalities. In *SYSID 2000, IFAC Symp. on System Identification*, 2000.

[26] K. Lindqvist and H. Hjalmarsson. Identification for Control: Adaptive input design using convex optimization. In *Proc. 40th IEEE Conf. on Decision and Control*, Orlando, FL, 2001.

[27] L. Ljung. Asymptotic variance expressions for identified black-box transfer function models. *IEEE Transactions on Automatic Control*, AC-30:834–844, 1985.

[28] L. Ljung. *System Identification - Theory For the User, 2nd ed*. PTR Prentice Hall, Upper Saddle River, N.J, 1999.

[29] L. Ljung and T. Söderström. *Theory and practice of recursive identification*. The MIT Press, 1983.

[30] T. Söderström and P. Stoica. *System Identification*. Prentice-Hall International, Hemel Hempstead, UK, 1989.

[31] V. A. Yakubovich. Solution of certain matrix inequalities occuring in the theory of automatic control. *Docl. Acad. Nauk. SSSR*, pages 1304–1307, 1962.