



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *IEEE Signal Processing Letters*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Hu, J., Zachariah, D., Wigren, T., Stoica, P. (2025)

Adaptive Experiment Design for Nonlinear System Identification with Operational Constraints

IEEE Signal Processing Letters, : 1-5

<https://doi.org/10.1109/lsp.2025.3639512>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-572530>

Adaptive Experiment Design for Nonlinear System Identification with Operational Constraints

Jingwei Hu, Dave Zachariah, Torbjörn Wigren, Petre Stoica

Abstract—We consider the joint problem of online experiment design and parameter estimation for identifying nonlinear system models, while adhering to system constraints. We utilize a receding horizon approach and propose a new adaptive input design criterion, which is tailored to continuously updated parameter estimates, along with a new sequential estimator. We demonstrate the ability of the method to design informative experiments online, while steering the system within operational constraints.

Index Terms—input design, adaptive design, constrained experiments

I. INTRODUCTION

We consider discrete-time nonlinear dynamical systems of the form

$$x_{t+1} = f_{\theta}(x_t, u_t), \quad (1)$$

where both states x_t and parameters θ are unknown. We aim to address the joint problem of (i) *estimating* θ from the inputs u_t and measurements of the system outputs y_t , and (ii) *designing* u_t so as to make the data informative about θ , while steering the system states within operational constraints.

There is a rich set of methods to tackle the estimation and design problems separately when the system is known to be *linear*. A classic design choice for u_t is a pseudorandom binary sequence (PRBS) [10] due to its excitation properties and ease of implementation. More efficient input designs are possible by leveraging a known structure of the system using the statistically motivated Fisher information matrix or asymptotic error covariance matrices, in the design criterion [1], [4], [6], [7], [12], [13], [17], [23], [31].

The high cost of conducting experiments motivates the development of efficient design schemes that continuously *adapt* u_t to the information available up to time t [2], [12], [16], [22], [33]. Informative experiments may, however, have adverse effects on the system in operation. (See [8], [9], [28] for a control perspective.) An open-loop input design method for *nonlinear* systems was introduced in [11], that penalizes designs in which the estimated states violate the constraints. This is achieved by linearizing the state-space model and designing u_t over a fixed horizon. More recently, a closed-loop (receding horizon) method was developed in [3] that also penalizes constraint violations. This penalized design approach requires a delicate balancing of competing objectives, where the information criterion is based on direct combination of dimensions of θ that typically have incommensurable units.

The contribution of this paper is:

- a statistically motivated experiment design method, adapted to the available information at each time step, that can steer the system within operational constraints; and
- an online parameter estimation method that is efficient and insensitive to initial estimates.

The proposed joint design and estimation method enables safer and more efficient experiments for nonlinear systems.

II. PROBLEM FORMULATION

The system output is assumed to follow the measurement model

$$y_t = Hx_t + e_t \quad (2)$$

where y_t is a $d_y \times 1$ vector and $e_t \sim \mathcal{N}(0, V)$ is a white noise process with a diagonal covariance matrix V with diagonal elements given by the vector v . Sensor specifications typically provide relatively accurate estimates of the variances v . The dimensions of u_t , θ and x_t are $d_u \times 1$, $d_{\theta} \times 1$ and $d_x \times 1$, respectively.

A sequence of input-output signals $\{(u_0, y_1), \dots, (u_{t-1}, y_t)\}$ provides information about θ in (1) via (2). At any t , our aim is to design subsequent inputs u_{t+i} so that the resulting observations provide the maximal *additional* information about θ , while steering the system within user-defined operational constraints:

$$x_{\min} \leq x_{t+i} \leq x_{\max}, \quad (3)$$

which can be time varying. This is readily extended to constraints on the outputs Hx_t via (2). A fundamental challenge here is that x_{t+i} is given by the forward recursion of (1):

$$\begin{aligned} x_{t+i} &= f_{\theta}(f_{\theta}(\dots f_{\theta}(x_t, u_t) \dots u_{t+i-2}), u_{t+i-1}) \\ &\equiv f_{\theta}^i(x_t, u_{t:t+i-1}), \quad \forall i > 0 \end{aligned} \quad (4)$$

where θ and x_{t+1} have to be estimated from data. Thus (4) renders the constraints (3) nonlinear in the designed inputs, which must also satisfy

$$u_{\min} \leq u_{t+i} \leq u_{\max}. \quad (5)$$

Let \mathcal{U} denote the set of input sequences that satisfy (5) and let \mathcal{X} be the set of resulting states that satisfy (3). It is assumed that a feasible sequence exists in \mathcal{U} . To tackle the computational challenges ensuing from (4), we will leverage advances in automatic differentiation methods.

Notation: $\|z\|_W = \sqrt{z^T W z}$ is a weighted norm using a positive definite matrix W . The element-wise product and division operator are denoted \odot and \oslash , while $(z)^+$ sets the negative elements of a vector z to 0, i.e., $[(z)^+]_i = \max(0, z_i)$.

III. METHOD

Suppose we are to design a k -step input sequence $U_t = [u_t, u_{t+1}, \dots, u_{t+k-1}]$. The horizon length k should be set sufficiently long to steer the system within operational constraints. The designed sequence yields corresponding outputs $\{y_{t+1}, \dots, y_{t+k}\}$. We begin by formulating a suitable adaptive design criterion for U_t that quantifies the additional information gained about θ relative to what was available at t .

A. Adaptive design criterion

Using (2) and (4), the likelihood of the k new outputs depends on θ, v and the unknown state x_t . Implementing U_t yields the following Fisher information matrix (FIM) of the unknown variables:

$$J_{\theta, x_t, v}(U_t) = \mathbb{E} \left[\begin{bmatrix} \frac{\partial \ell}{\partial \theta} \\ \frac{\partial \ell}{\partial x_t} \\ \frac{\partial \ell}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{\partial \ell}{\partial \theta} \\ \frac{\partial \ell}{\partial x_t} \\ \frac{\partial \ell}{\partial v} \end{bmatrix}^\top \right], \quad (6)$$

where ℓ denotes the (Gaussian) log-likelihood function of the data. It follows from the Cramér-Rao inequality [21], [29] that

$$C_{t+k} = [I_{d_\theta} \quad 0] J_{\theta, x_t, v}^{-1} [I_{d_\theta} \quad 0]^\top \quad (7)$$

is the minimum achievable error covariance matrix for any unbiased estimator $\hat{\theta}_{t+k}$ that incorporates the k new samples.

Let $\hat{\theta}_t$ denote a prior estimate at time t with error covariance matrix $C_t \succeq C_{t+k}$. Then the following bounded criterion

$$\mathcal{J}(U_t; \theta, x_t, v) = \text{tr} \{ C_{t+k} C_t^{-1} \} \leq d_\theta, \quad (8)$$

quantifies the *relative* error improvement gained by the design U_t . The upper bound d_θ is attained in the extreme case of *uninformative* experiments. The criterion (8) bounds the mean squared-error of $\hat{\theta}_{t+k}$, standardized to the errors at t :

$$\mathcal{J} \leq \mathbb{E} \left[\|\theta - \hat{\theta}_{t+k}\|_{C_t^{-1}}^2 \right], \quad (9)$$

via the Cramér-Rao inequality. Thus \mathcal{J} is an L-criterion [7, ch. 2.9] *adapted* to the information about θ available at t that renders the errors in the components of θ commensurable.

A receding horizon formulation for adaptive experiment design is then

$$\begin{aligned} U_t^* &= \arg \min_{U_t \in \mathcal{U}} \mathcal{J}(U_t; \theta, x_t, v) \\ &\text{subject to } x_{t+i} \in \mathcal{X}, \forall i = 1, \dots, k, \end{aligned} \quad (10)$$

which is evaluated at estimates $\hat{\theta}$, \hat{x}_t and \hat{v} given in Section III-C. However, both the constraints in (10) and the log-likelihood for all data render the problem rather intractable. Next, we address both issues.

B. Relaxation

The likelihood function for the past data $y_{1:t}$ does not depend on U_t and is parameterized using an initial state x_0 . To obtain a tractable FIM that uses all data, we consider summarizing the past data $y_{1:t}$ using maximum likelihood estimates (MLE) $\hat{\theta}$, \hat{x}_t and \hat{v} . The distribution of the estimates

is denoted $q(\hat{\theta}, \hat{x}_t, \hat{v} | \theta, x_t, v)$. Under standard regularity conditions [21, ch. 9] [32, app. B], q is asymptotically normal and the estimator efficient so that with a Gaussian data model (2), its asymptotic covariance matrix is block-diagonal:

$$\mathbb{V} \left[\begin{bmatrix} \hat{\theta} & \hat{x}_t & \hat{v} \end{bmatrix}^\top \right] = \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}, \quad (11)$$

where the lower block Q belongs to \hat{v} which is thus uncorrelated with $\hat{\theta}$ and \hat{x}_t . We will therefore take q to be a Gaussian distribution as summary of the past data. With a given covariance matrix (11), this is the most conservative distribution in the sense that it yields the smallest FIM [26], [30]. (When the MLE is asymptotically sufficient [15], q also corresponds to a factorization of the past likelihood of $y_{1:t}$.)

The likelihood function then becomes

$$\begin{aligned} \tilde{\ell} &= \ln q(\hat{\theta}, \hat{x}_t, \hat{v} | \theta, x_t, v) + \sum_{i=1}^k p(y_{t+i} | u_{t:t+i-1}, \theta, x_t, v) \\ &= -\frac{1}{2} \left\| \begin{bmatrix} \theta \\ x_t \end{bmatrix} - \begin{bmatrix} \hat{\theta} \\ \hat{x}_t \end{bmatrix} \right\|_{P^{-1}}^2 - \frac{1}{2} \|v - \hat{v}\|_{Q^{-1}}^2 \\ &\quad - \frac{1}{2} \sum_{i=1}^k \|\varepsilon_{t+i}\|_{V^{-1}}^2 - \frac{k}{2} \ln |V| + \text{const.}, \end{aligned} \quad (12)$$

where we used the prediction error $\varepsilon_{t+i} \equiv y_{t+i} - H f_\theta^i(x_t, u_{t:t+i-1})$ for compactness. The corresponding conservative FIM (6) becomes [29]:

$$\begin{aligned} \tilde{J}_{x_t, \theta, v} &= \begin{bmatrix} \tilde{J}_\theta & \tilde{J}_{\theta x_t} & 0 \\ \tilde{J}_{\theta x_t}^\top & \tilde{J}_{x_t} & 0 \\ 0 & 0 & \tilde{J}_v \end{bmatrix} \\ &= \begin{bmatrix} P^{-1} + \sum_{i=1}^k E_{t+i}^\top V^{-1} E_{t+i} & 0 \\ 0 & Q^{-1} + \frac{k}{2} V^{-2} \end{bmatrix}, \end{aligned} \quad (13)$$

where

$$E_{t+i} = H \begin{bmatrix} \frac{\partial f_\theta^i(x_t, u_{t:t+i-1})}{\partial \theta} & \frac{\partial f_\theta^i(x_t, u_{t:t+i-1})}{\partial x_t} \end{bmatrix}$$

is the Jacobian of the prediction error ε_{t+i} . Applying (13) to (7), we finally obtain

$$\tilde{C}_{t+k} \equiv (\tilde{J}_\theta - \tilde{J}_{\theta x_t} \tilde{J}_{x_t}^{-1} \tilde{J}_{\theta x_t}^\top)^{-1} \succeq C_{t+k}. \quad (14)$$

Inserting the matrix in (14) into the criterion (8) yields a tractable, conservative minimum error bound $\tilde{\mathcal{J}} \geq \mathcal{J}$. (The bound could be loose for small t , but can become tight when asymptotic sufficiency of the MLE holds). The dominant cost of evaluating $\tilde{\mathcal{J}}$ is therefore the inversion of \tilde{J}_{x_t} .

We now turn to relaxing the state constraints \mathcal{X} in (10). We aim for designs that continually steer the state to satisfy the operational constraints. This is achieved by introducing a penalty term to the design problem (10). Whereas \mathcal{J} measures a standardized squared deviation of $\hat{\theta}_{t+k}$ from the unknown θ , we let penalty $\mathcal{J}_\mathcal{X}$ measure a standardized squared exceedance of x_{t+i} from the state boundaries. Specifically,

$$\begin{aligned} \mathcal{J}_\mathcal{X}(U_t; \theta, x_t) &= \frac{1}{k} \times \\ &\sum_{i=1}^k \|(Sx_{\min} - Sx_{t+i})^+ + (Sx_{t+i} - Sx_{\max})^+\|^2, \end{aligned} \quad (15)$$

where S is a diagonal matrix that standardizes violations from the constraints. That is, $[S]_{i,i} = 1/(x_{\max}^{(i)} - x_{\min}^{(i)})$ or 0, depending on whether x_i is constrained. The relaxed form of (10) then becomes

$$U_t^* = \arg \min_{U_t \in \mathcal{U}} \tilde{\mathcal{J}}(U_t; \theta, x_t, v) + \gamma \mathcal{J}_{\mathcal{X}}(U_t; \theta, x_t), \quad (16)$$

where $\gamma \geq 0$ is a weight that controls the adherence to the state constraints. Note that when the state sequence x_{t+i} is in \mathcal{X} , the cost $\mathcal{J}_{\mathcal{X}}$ vanishes. By setting, for instance, $\gamma > d_\theta/(0.01)^2$ the cost of exceeding the state boundaries by 1%, is greater than the cost of using uninformative inputs ($\tilde{\mathcal{J}} = d_\theta$). This incentivizes designs that steer the system within the state constraint set \mathcal{X} . Note that $\tilde{\mathcal{J}}$ also provides a diagnostic for how informative the experiment is.

C. Online estimator

Joint estimation of (θ, x_t, v) for the design (16) involves $d_\theta + d_x + d_y$ parameters and is a challenging task for standard recursive estimators. To reduce the sensitivity to poor initial estimates, we update $\hat{\theta}$ periodically using blocks of $b \geq d_\theta + d_x + d_y$ samples. At the same time, we maintain a running estimation of x_t using the model in (1).

Suppose t is a multiple of b , so that $\tau = t - b$ indicates the end of the previous block of samples. Let $(\hat{\theta}_\tau, \hat{x}_{\tau|\tau}, \hat{v}_\tau)$ denote the prior estimates at τ . Using the likelihood function along with the b subsequent samples, new estimates $\hat{\theta}_t, \hat{v}_t$, and a *smoothed* estimate $\hat{x}_{\tau|t}$ is obtained by solving:

$$\begin{aligned} (\hat{\theta}_t, \hat{x}_{\tau|t}, \hat{v}_t) = \arg \min_{\theta, x_\tau, v} & \sum_{i=1}^b \|\varepsilon_{\tau+i}\|_{V^{-1}}^2 + b \ln |V| \\ & + \left\| \begin{bmatrix} \theta \\ x_\tau \end{bmatrix} - \begin{bmatrix} \hat{\theta}_\tau \\ \hat{x}_{\tau|\tau} \end{bmatrix} \right\|_{P_{\tau|\tau}^{-1}}^2 + \|v - \hat{v}_\tau\|_{Q_\tau^{-1}}^2, \end{aligned} \quad (17)$$

which is a maximum likelihood estimator. The covariance matrices are therefore approximated using the corresponding inverse FIM:

$$P_{\tau|t} = \begin{bmatrix} \tilde{J}_\theta & \tilde{J}_{\theta x_\tau} \\ \tilde{J}_{\theta x_\tau}^\top & \tilde{J}_{x_\tau} \end{bmatrix}^{-1} \quad \text{and} \quad Q_t = \tilde{J}_v^{-1} \quad (18)$$

based on the form in (13) with plug-in estimates. The state estimate in (17) and the joint covariance (18) are then sequentially updated to t :

$$(\hat{x}_{\tau|t}, P_{\tau|t}) \rightarrow (\hat{x}_{\tau+1|t}, P_{\tau+1|t}) \rightarrow \dots \rightarrow (\hat{x}_{t|t}, P_{t|t})$$

using the prediction step of the unscented Kalman filter [5], [14], [24]. The procedure is described as follows:

Let $\hat{z}_t = \begin{bmatrix} \hat{\theta} \\ \hat{x}_t \end{bmatrix}$ denote an initial estimate with an error covariance matrix P_t . The unscented transform is a deterministic sampling method that produces a one-step prediction, and its approximate error covariance, using $S = 2d_z + 1$ samples with $d_z = d_\theta + d_x$, is given by:

$$(\hat{z}_{t+1}, P_{t+1}) = T(\hat{z}_t, P_t; u_t). \quad (19)$$

The augmented state estimate is computed as

$$\hat{z}_{t+1} = \frac{1}{S} \sum_{s=0}^{S-1} w^{(s)} z_{t+1}^{(s)}, \quad \text{where } z_{t+1}^{(s)} = \begin{bmatrix} \hat{\theta}^{(s)} \\ f_{\hat{\theta}^{(s)}}(\hat{x}_t^s, u_t) \end{bmatrix}$$

and

$$z_t^{(s)} = \begin{cases} \hat{z}_t & \text{for } s = 0 \\ \hat{z}_t + \eta^{1/2} [P_t^{1/2}]_s & \text{for } s = 1, \dots, d_z \\ \hat{z}_t - \eta^{1/2} [P_t^{1/2}]_s & \text{for } s = d_z + 1, \dots, 2d_z \end{cases}. \quad (20)$$

The weights $w^{(s)}$ and η are given in [5], [14], [24]. The covariance matrix P_t is computed similarly via outer products of the samples. Eq. (19) is applied sequentially to produce running predictions of x_t for use in (16).

D. Adaptive design method

Using the joint estimate $(\hat{\theta}, \hat{x}_t, \hat{v})$ described above, we can finally evaluate the criterion in (16). This design problem can be converted into an unconstrained form by transforming the inputs $\mathcal{U} \rightarrow \mathbb{R}^{d_u}$ as follows:

$$\tilde{\mathcal{U}} = [\tilde{u}(u_t), \dots, \tilde{u}(u_{t+k-1})] \quad (21)$$

where $\tilde{u}(u_t) = g((u_t - u_{\min}) \odot (u_{\max} - u_{\min}))$ and where g is a differentiable and invertible function, e.g., the element-wise inverse sigmoid function $[g(z)]_i = \ln \frac{z_i}{1-z_i}$.

Thus (21) is unconstrained and computationally convenient for adaptive experiment design. The map $\tilde{\mathcal{U}}_t$ recovers the original inputs by:

$$U_t = [\tilde{u}^{-1}(\tilde{u}_t), \dots, \tilde{u}^{-1}(\tilde{u}_{t+k-1})] \quad (22)$$

where $\tilde{u}^{-1}(\tilde{u}_t) = g^{-1}(\tilde{u}_t) \odot (u_{\max} - u_{\min}) + u_{\min}$ and $[g^{-1}(z)]_i = \frac{1}{1+e^{-z_i}}$. Using the transformed inputs $\tilde{\mathcal{U}}_t$, by substituting (22) in (16), results in an unconstrained optimization problem. At time t , the input

$$u_t^* = \tilde{u}^{-1}(u_t^*)$$

is computed and applied in the subsequent time step.

The adaptive input design and online estimation method can now be summarized as follows:

- 1) Observe b new samples, solve (17) and compute (18).
- 2) Update estimate of x_t and P_t using (19).
- 3) Solve (16) in unconstrained form and apply input u_t^* .
- 4) Update $t := t + 1$
- 5) Go to 1) if t is a multiple of b , otherwise go to 2).

To solve the non-convex minimization problem in (17), we employ a numerical optimization strategy suitable for rugged cost surfaces. Specifically, we utilize the Basin Hopping algorithm [19], [25] to locate an approximate solution to (17), then using a quasi-Newton limited-memory BFGS algorithm [20] for local optimization. To solve the design problem (16), we leverage an available automatic differentiation method [27] alongside the Adam optimizer [18], using a constant stepsize of 10^{-3} . The code implementations are available online.¹ The runtime of the design method using a standard laptop with single thread CPU is in the range of 0.2 to 5 seconds, depending on k and model complexity.

¹<https://github.com/jingwei91hu/adaptive-input-design>

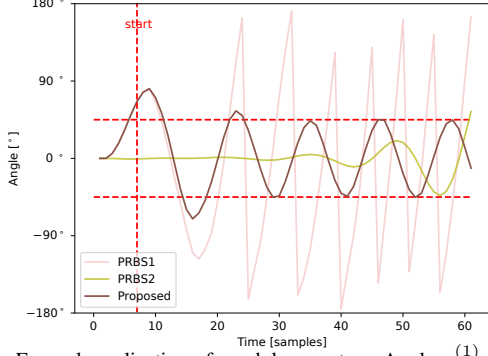


Fig. 1. Example realization of pendulum system. Angle $x_t^{(1)}$ when applying two different PRBS inputs and the adaptive design that steers the angle within the interval $[-45^\circ, 45^\circ]$.

IV. NUMERICAL EXPERIMENTS

In the experiments presented below, we use a nonlinear pendulum model as an example of (1):

$$\begin{aligned} x_{t+1}^{(1)} &= x_t^{(1)} + x_t^{(2)} \Delta T \\ x_{t+1}^{(2)} &= x_t^{(2)} + (\theta_1 \sin x_t^{(1)} + \theta_2 u_t) \Delta T, \end{aligned} \quad (23)$$

where $x_t^{(1)}$ represents the pendulum's angular displacement and $x_t^{(2)}$ denotes its angular velocity u_t . Additional examples are provided in the Supplementary Materials.

The experiments are initialized with a block of b samples generated from PRBS inputs. The initial prior estimates, $\hat{\theta}_0$ and $\hat{x}_{1|0}$, are randomly drawn from a Gaussian distribution with zero mean and chose to be uninformative by assigning a large covariance matrix $P_{1|0} = 10^4 \cdot I$. The initial noise estimates \hat{v}_0 are presumed to be accurate, and their covariance matrix Q_0 is configured to correspond to a 0.1% uncertainty in the noise standard deviations. For all the experiments, the weight in (16) was set to $\gamma = 400 \gg d_\theta$, where d_θ equals 2 or 3 below.

To provide a succinct evaluation of the parameter estimates, we use the normalized mean squared error

$$\text{MSE}_t = \sum_{k=1}^{d_\theta} \frac{\mathbb{E}[|\theta_k - \hat{\theta}_{k,t}|^2]}{\theta_k^2}$$

and the corresponding (approximate) Cramér-Rao bounds (CRB) obtained by evaluating the FIM in (14) at the true parameters. In addition, we quantify the (relative) violation of operational constraints (3) using:

$$\text{OCV}_t = \mathbb{E} [\|(Sx_{\min} - Sx_t)^+ + (Sx_t - Sx_{\max})^+ \|],$$

with S from (15).

We set the model parameters to $\theta_1 = -24$, determined by the gravitational acceleration, and $\theta_2 = 1$, related to the pendulum's length. The sampling period is $\Delta T = 0.1$ s. Only the angular displacement is observed, meaning $H = [1 \ 0]$ in (2), with standard deviation of noise $\sqrt{v} = 0.01$ rad. The system is initialized at $x_0 = [0 \ 0]^T$ with the goal of gathering informative samples while keeping the pendulum within $[x_{\min}^{(1)}, x_{\max}^{(1)}] = [-45^\circ, +45^\circ]$. The input constraint was

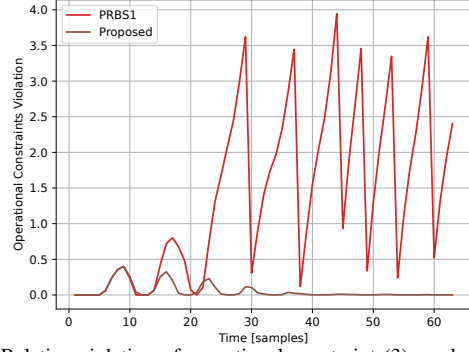


Fig. 2. Relative violation of operational constraint (3), evaluated using 100 Monte Carlo runs.

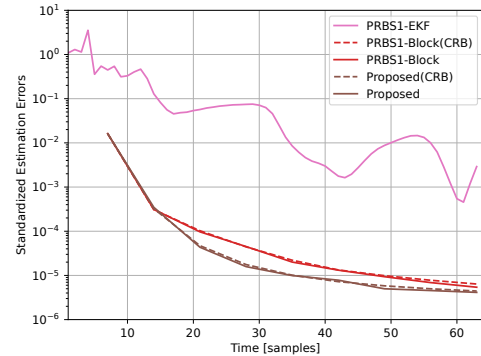


Fig. 3. Mean-square errors of parameter estimator along with (approximate) Cramér-Rao bounds. The estimation errors for PRBS2 were too large to be included in the plot. Results are based on 100 Monte Carlo runs.

$|u_t| \leq 10$. In the experiment, for the adaptive design, we use horizon length $k = 6$ and block size $b = 7$.

Figure 1 provides an illustration of three experiments: The first is PRBS1 which only satisfies input amplitude constraints, whereas the second one, PRBS2, is tuned to the best of our ability to satisfy the operational constraints during an experiment. In contrast, the adaptive design steers the angular state within its constraints. A more systematic comparison of constraint violations is provided in Fig. 2.

Finally, Figure 3 shows the estimation errors and CRBs. We see that the online estimator outperforms the Extended Kalman Filter (EKF), with parameter-augmented states, when using the PRBS1 design. We also note that the proposed design lowers the bound while steering within the state constraints. The resulting errors of the online errors are correspondingly lower.

V. CONCLUSION

We have proposed an adaptive experiment design method that steers the system within system constraints and an on-line estimator for nonlinear systems that provides parameter estimates and necessary state estimates. We demonstrate the ability of the method to perform safer experiments online, while steering the system within operational constraints, yet yield substantial gains in estimation accuracy and, therefore, reduced experiment time.

REFERENCES

- [1] M. Annergren, C. Larsson, H. Hjalmarsson, X. Bombois, and B. Wahlberg, "Application-oriented input design in system identification: Optimal input design for control [applications of control]," *IEEE Control Systems Magazine*, vol. 37, no. 2, pp. 31–56, 2017.
- [2] S. Arimoto and H. Kimura, "Optimum input test signals for system identification—an information-theoretical approach," *International Journal of Systems Science*, vol. 1, no. 3, pp. 279–290, 1971.
- [3] M. Babar and M. Baglietto, "Optimal feedback input design for dynamic nonlinear systems," *International Journal of Control*, vol. 94, no. 8, pp. 2264–2281, 2021.
- [4] X. Bombois, M. Gevers, R. Hildebrand, and G. Solari, "Optimal experiment design for open and closed-loop system identification," *Communications in Information and Systems*, vol. 11, no. 3, pp. 197–224, 2011.
- [5] J. V. Candy, *Bayesian signal processing: classical, modern, and particle filtering methods*. John Wiley & Sons, 2016.
- [6] D. Eckhard, A. Bazanella, C. Rojas, and H. Hjalmarsson, "Input design as a tool to improve the convergence of pem," *Automatica*, vol. 49, no. 11, pp. 3282–3291, 2013.
- [7] V. Federov, *Theory of optimal experiments*. Academic Press, 1972.
- [8] A. Feldbaum, *Optimal Control Systems*. Academic Press, 1965.
- [9] S. Fflila, P. Dufour, and H. Hammouri, "Optimal input design for online identification: a coupled observer-mpc approach," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 11 457–11 462, 2008.
- [10] K. Godfrey, "Introduction to binary signals used in system identification," in *International Conference on Control 1991. Control'91*. IET, 1991, pp. 161–166.
- [11] G. Goodwin, "Optimal input signals for nonlinear-system identification," vol. 118, no. 7, pp. 922–926, 1971.
- [12] —, "Dynamic system identification: experiment design and data analysis," *Mathematics in science and engineering*, vol. 136, 1977.
- [13] H. Jansson and H. Hjalmarsson, "Input design via lmis admitting frequency-wise model specifications in confidence regions," *IEEE transactions on Automatic Control*, vol. 50, no. 10, pp. 1534–1549, 2005.
- [14] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
- [15] S. Kaufman, "Asymptotic efficiency of the maximum likelihood estimator," *Annals of the Institute of Statistical Mathematics*, vol. 18, pp. 155–178, 1966.
- [16] L. Keviczky, *On Some Questions of Input Signal Synthesis*, ser. Research Reports TFRT-3047. Department of Automatic Control, Lund Institute of Technology, 1972.
- [17] J. Kiefer, "Optimum experimental designs," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 21, no. 2, pp. 272–304, 1959.
- [18] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [19] R. Leary, "Global optimization on funneling landscapes," *Journal of Global Optimization*, vol. 18, pp. 367–383, 2000.
- [20] D. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical programming*, vol. 45, no. 1, pp. 503–528, 1989.
- [21] L. Ljung, *System Identification: Theory for the User*. Prentice Hall, 1999.
- [22] H. Mania, M. I. Jordan, and B. Recht, "Active learning for nonlinear system identification with guarantees," *Journal of Machine Learning Research*, vol. 23, no. 32, pp. 1–30, 2022.
- [23] R. Mehra, "Optimal inputs for linear system identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 3, pp. 192–200, 1974.
- [24] H. Menegaz, J. Ishihara, G. Borges, and A. Vargas, "A systematization of the unscented kalman filter theory," *IEEE Transactions on automatic control*, vol. 60, no. 10, pp. 2583–2598, 2015.
- [25] B. Olson, I. Hashmi, K. Molloy, and A. Shehu, "Basin hopping as a general and versatile optimization framework for the characterization of biological macromolecules," *Advances in Artificial Intelligence*, vol. 2012, no. 1, p. 674832, 2012.
- [26] S. Park, E. Serpedin, and K. Qaraqe, "Gaussian assumption: The least favorable but the most useful [lecture notes]," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 183–186, 2013.
- [27] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS-W*, 2017.
- [28] J. Qian, M. Nadri, and P. Dufour, "Optimal input design for parameter estimation of nonlinear systems: case study of an unstable delta wing," *International Journal of Control*, vol. 90, no. 4, pp. 873–887, 2017.
- [29] T. Söderström and P. Stoica, *System Identification*. Prentice Hall, 1989.
- [30] P. Stoica and P. Babu, "The gaussian data assumption leads to the largest cramer-rao bound [lecture notes]," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 132–133, 2011.
- [31] P. Stoica and T. Söderström, "A useful input parameterization for optimal experiment design," *IEEE Transactions on Automatic Control*, vol. 27, no. 4, pp. 986–989, 1982.
- [32] P. Stoica, R. L. Moses, et al., *Spectral analysis of signals*. Prentice Hall, 2005.
- [33] V. Stojanovic and V. Filipovic, "Adaptive input design for identification of output error model with constrained output," *Circuits, Systems, and Signal Processing*, vol. 33, pp. 97–113, 2014.