IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Coherence-Based Input Design for Nonlinear Systems

Javad Parsa, Cristian R. Rojas, and Håkan Hjalmarsson

***Abstract*— Many off-the-shelf generic non-linear model structures have inherent sparse parametrizations. Volterra series and non-linear Auto-Regressive with eXogeneous inputs (NARX) models are examples of this. It is well known that sparse estimation requires low mutual coherence, which translates into input sequences with certain low correlation properties. This paper highlights that standard optimal input design methods do not account for this requirement which may lead to designs unsuitable for this type of model structure. To tackle this problem, the paper proposes incorporating a coherence constraint to standard input design problems. The coherence constraint is defined as the ratio between the diagonal and non-diagonal entries of the Fisher information matrix (FIM) and can be easily added to any input design problem for nonlinear systems, while the resulting problem remains convex. The paper provides a theoretical analysis of how the range of the optimal objective function of the original problem is affected by the coherence constraint. Additionally, the paper presents numerical evaluations of the proposed approach's performance on a Volterra series model in comparison to state-of-the-art algorithms.**

I. Introduction

A. Sparse System Identification

The science of constructing mathematical models for dynamic systems by measuring their inputs and outputs is known as system identification. Mathematical models are an approximation of a system's actual behavior, and various assumptions can lead to distinct models. This paper centers on nonlinear models that can be expressed as a linear regression, which is defined as

$$\mathbf{y} = \mathbf{\Phi}\boldsymbol{\theta} + \mathbf{e}, \tag{1}$$

in which $\mathbf{\Phi} \in \mathbb{R}^{N\times n_\theta}$, $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta\times 1}$, $\mathbf{y} \in \mathbb{R}^{N\times 1}$ and $\mathbf{e} \in \mathbb{R}^{N\times 1}$ indicate a regressor matrix that consists of nonlinear terms of input values, a parameter vector, an observation vector, and white Gaussian noise with zero mean and covariance matrix $\sigma^2\mathbf{I}$, i.e., $\mathbf{e} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, respectively. Nonlinear models that are linear in parameters can be formulated as in (1).

The Polynomial NARX models are widely recognized nonlinear models used for system identification [1]. The model is represented by (1), where the regressor $\mathbf{\Phi}$ includes polynomials of past input and output signals. The Volterra model is another nonlinear model used in system identification, which is also expressed by (1) [1]. The number of parameters in a Volterra model is determined by the number of kernels P and the model memory L, and is given by $n_\theta = \binom{L+P}{P}$.

It is evident from the polynomial NARX and Volterra models that there can be an enormous number of parameters, with only a small number of them being non-zero. Furthermore, not estimating more parameters than necessary is beneficial for a low variance error in parameter estimation [2]. Consequently, sparse parameter estimation can be an effective approach to estimate this type of models.

In this case, we can estimate the parameter vector $\boldsymbol{\theta}$ in (1) by minimizing the error $\|\mathbf{y} - \mathbf{\Phi}\boldsymbol{\theta}\|_2^2$ with respect $\boldsymbol{\theta}$ while ℓ_0 pseudo norm of $\boldsymbol{\theta}$, i.e., $\|\boldsymbol{\theta}\|_0$, which denotes the number of nonzero elements, is smaller than s; this method is called P_0^ϵ. Since P_0^ϵ is a NP-hard problem [3], the ℓ_0 can be approximated with ℓ_1 which results in the so called P_1^ϵ method, also known as LASSO (least absolute shrinkage and selection operator) [4].

B. Mutual Coherence

Several papers have been published to analyze the requirements for an accurate sparse estimation [3], [5]–[7]. A significant number of these studies underscore the crucial role of mutual coherence, which denotes the highest possible absolute correlation between distinct columns of a regressor, i.e.,

$$\mu_{\mathbf{\Phi}} = \max_{i\neq j} \frac{|\langle \boldsymbol{\phi}_i, \boldsymbol{\phi}_j\rangle|}{\|\boldsymbol{\phi}_i\|_2\|\boldsymbol{\phi}_j\|_2}, \tag{2}$$

in which $\mu_{\mathbf{\Phi}}$ and $\boldsymbol{\phi}_i$ denote the mutual coherence of $\mathbf{\Phi}$ and the i-th column of $\mathbf{\Phi}$, respectively.

Specifically, the mutual coherence is essential in sparse estimation algorithms such as P_0^ϵ, which is evident from its ability to limit the estimation error, denoted as $\|\boldsymbol{\theta}^\epsilon - \boldsymbol{\theta}_0\|_2$, where $\boldsymbol{\theta}^\epsilon$ and $\boldsymbol{\theta}_0$ represent the solutions obtained by P_0^ϵ and the true parameters, respectively. A lower mutual coherence leads to a smaller upper bound on the estimation error, with the minimum bound being achieved when $\mu_{\mathbf{\Phi}} = 0$ [8]. Similarly, the importance of mutual coherence in P_1^ϵ has been established in [3]. In a nutshell, sparse estimation requires a low mutual coherence in order to be accurate.

C. Input Design for Nonlinear Systems

In this paper, we aim to design the input sequence $\mathcal{U}_{n_{seq}} := \{u_{n_{seq}}, \cdots, u_1\}$ that maximizes the accuracy of a given model, expressed as

$$y_t = G(\mathcal{U}_t)^T\boldsymbol{\theta}_0 + e_t. \tag{3}$$

Here, $\mathcal{U}_t = (u_t, \cdots, u_{t-k})$ in which k is the maximum time lag, and G is a nonlinear function. This type of model structure is very common in black-box identification of nonlinear systems, where a range of candidate non-linearities are included in the model but where many of these can be expected to be zero. Volterra and NARX models are two examples of this. The maximum accuracy is measured by a scalar function of the FIM, such as $\lambda_{\min}(\mathbf{I}_F)$ or $\log\det(\mathbf{I}_F)$, where $\lambda_{\min}$, det, and $\mathbf{I}_F$ represent the minimum eigenvalue, determinant, and FIM, respectively. To begin with, the FIM in the linear regression model (1) when $\mathbf{e} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ is given by

$$\mathbf{I}_F = \frac{1}{\sigma^2}\mathbf{\Phi}^T\mathbf{\Phi}. \tag{4}$$

Instead of directly designing the input such that the corresponding $\mathbf{\Phi}$ results in desired properties of the FIM, which is a generally acknowledged difficult problem, we will follow the classical approach and design the per sample FIM given by

$$\bar{\mathbf{I}}_F = \frac{1}{\sigma^2}\,\mathbb{E}\left\{\sum_{t=1}^{n_{seq}} \boldsymbol{\omega}_t \boldsymbol{\omega}_t^T\right\} \tag{5}$$

where $\boldsymbol{\omega}_t = G(\mathcal{U}_t)$, where in turn $\mathcal{U}_t$ is a stationary stochastic process. Due to stationarity and the law of large numbers

$$\lim_{N\to\infty} \frac{1}{N}\mathbf{I}_F = \bar{\mathbf{I}}_F, \quad \text{w.p. } 1$$

This leads to a two-step procedure: 1) design the per sample FIM. 2) Generate a realization of the input from the obtained stationary stochastic process. The motivation for this approach is that the first step can often be solved using convex optimization. The method has a long history, see e.g. [9]–[12].

In our setting, the per sample FIM is given by [12], [13]

$$\bar{\mathbf{I}}_F = \frac{1}{\sigma^2}\,\mathbb{E}\left\{\sum_{t=1}^{n_{seq}} \boldsymbol{\omega}_t \boldsymbol{\omega}_t^T\right\} = \frac{1}{\sigma^2}\int_{\mathbb{R}^{n_{seq}}} \sum_{t=1}^{n_{seq}} \boldsymbol{\omega}_t \boldsymbol{\omega}_t^T p(\mathcal{U}_{n_{seq}}) d\mathcal{U}_{n_{seq}}, \tag{6}$$

where $p(\mathcal{U}_{n_{seq}})$ is the probability density function of $\mathcal{U}_{n_{seq}}$. The per sample FIM exhibits different behaviors for linear and nonlinear systems. In nonlinear systems, the per sample FIM is dependent on the probability function, while in linear systems, the per sample FIM is an affine function of the input spectrum [10], [11], [14], [15]. In linear systems, a finite representation for the input spectrum is considered as a design variable to reduce complexity. In nonlinear models, the input signal u_t can be restricted to take a finite number d_{seq} of values, to decrease computational complexity [12]. We will use $\mathcal{D}$ to show the set of these values. Since the input signal u_t can only take a finite number of values, $p(\mathcal{U}_{n_{seq}})$ then becomes a probability mass function and corresponds to a stationary process. Moreover, assuming the input signal u_t to be a stationary process with finite memory can further reduce computational complexity [16]. Therefore the optimal pmf is selected from the set $\mathcal{P}_{\mathcal{D}}$ which is characterized as

$$\mathcal{P}_{\mathcal{D}} := \Big\{ p : \mathcal{D}^{n_{seq}} \to \mathbb{R} \Big| \sum_{\mathbf{x}\in\mathcal{D}^{n_{seq}}} p(\mathbf{x}) = 1; p(\mathbf{x}) \geq 0,$$
$$\forall \mathbf{x} \in \mathcal{D}^{n_{seq}}, \sum_{u_t\in\mathcal{D}} p(u_t, \mathbf{w}) = \sum_{u_t\in\mathcal{D}} p(\mathbf{w}, u_t), \forall \mathbf{w} \in \mathcal{D}^{n_{seq}-1} \Big\}.$$

By restricting the input values as per the above conditions, the expression for $\bar{\mathbf{I}}_F$ in (6) becomes

$$\bar{\mathbf{I}}_F = \frac{1}{\sigma^2} \sum_{\mathcal{U}_{n_{seq}}\in\mathcal{D}^{n_{seq}}} \sum_{t=1}^{n_{seq}} \boldsymbol{\omega}_t \boldsymbol{\omega}_t^T p(\mathcal{U}_{n_{seq}}). \tag{7}$$

In [12], a convex optimization problem is proposed to design the optimal FIM for nonlinear models. The optimization aims to minimize the estimation error with respect to the pmf $p(\mathcal{U}_{n_{seq}})$ while $p(\mathcal{U}_{n_{seq}}) \in \mathcal{P}_{\mathcal{D}}$. It is of the form

$$\min_{p(\mathcal{U}_{n_{seq}})\in\mathcal{P}_{\mathcal{D}}} \quad \text{Estimation error}$$
$$\text{st} \qquad \bar{\mathbf{I}}_F = \frac{1}{\sigma^2} \sum_{\mathcal{U}_{n_{seq}}\in\mathcal{D}^{n_{seq}}} \sum_{t=1}^{n_{seq}} \boldsymbol{\omega}_t \boldsymbol{\omega}_t^T p(\mathcal{U}_{n_{seq}}), \tag{8}$$

in which the estimation error can be defined by different criteria such as A-optimality, D-optimality, E-optimality, or L-optimality [9]. The number of free variables in the above problem is d_{seq}^{k+1}.

In [13], an alternative approach to designing inputs in nonlinear models was proposed, which is based on graph theory. The approach involves using the fact that each element in the convex polyhedron $\mathcal{P}_{\mathcal{D}}$ can be represented by a convex combination of its extreme points. Therefore, for all $z \in \mathcal{P}_{\mathcal{D}}$, we have

$$z = \sum_{i=1}^{n_{\mathcal{A}}} \beta_i a_i, a_i \in \mathcal{A}_{\mathcal{P}} \tag{9}$$

in which set $\mathcal{A}_{\mathcal{P}}$ includes all extreme points of $\mathcal{P}_{\mathcal{D}}$, $\beta_i \geq 0$ and $\sum_{i=1}^{n_{\mathcal{A}}} \beta_i = 1$. The usage of graph theory is to find the set $\mathcal{A}_{\mathcal{P}}$. According to [17, Theorem 6], each a_i can be described by a prime cycle of a graph.

As a result, in [13] the per sample FIM which is associated with the $i-th$ prime cycle is given by

$$\bar{\mathbf{I}}_F^{(i)} \approx \frac{1}{\sigma^2 N_m} \sum_{t=1}^{N_m} \boldsymbol{\omega}_t \boldsymbol{\omega}_t^T, \tag{10}$$

in which N_m is sufficiently large and $\boldsymbol{\omega}_t$ depends on $\{u_t^i\}_{t=0}^{t=N_m}$ which is obtained from a_i. It is possible to reduce the sum (7) from dimension n_{seq} to dimension 1 by using the approximation of $\bar{\mathbf{I}}_F^{(i)}$ in (10).

The final optimization problem to find the optimal FIM in [13] is given by

$$\min_{\beta_1,\beta_2,\cdots,\beta_{n_{\mathcal{A}}}} \quad \text{Estimation error}$$
$$\text{st} \qquad \bar{\mathbf{I}}_F^{app} = \sum_{i=1}^{n_{\mathcal{A}}} \beta_i \bar{\mathbf{I}}_F^{(i)}$$
$$\sum_{i=1}^{n_v} \beta_i = 1$$
$$\beta_i \geq 0, \ \forall i \in \{1, \cdots, n_{\mathcal{A}}\}. \tag{11}$$

Methods (8) and (11) are equivalent. For large values of n_{seq}, method (11) is preferred in terms of computational complexity [16].

The optimal pmf and per sample FIM are the results of both methods, which are given by equations (8) and (11). In the second step of the method we need to generate a realization from the corresponding distribution. To this end, a method based on a Markov chain is proposed in [16]. First, the optimal pmf is used to design a transition probability matrix $\mathbf{A}$ using Algorithm 1 in [16]. Then, a Markov chain is run using a random initial state and the transition matrix $\mathbf{A}$ to generate the input sequence $\mathcal{U}_{n_{seq}}$. Once the input sequence is obtained, the regressor $\mathbf{\Phi}$ is constructed based on this sequence. This regressor matrix is now an approximation of the optimal regressor which would have been obtained if the FIM as given by (4) had been used. It is worth noting

that when the resulting Markov chain exhibits ergodicity, although the input sequence is generated as a realization of a stationary stochastic process, it gives approximately the desired average properties of the estimator of $\boldsymbol{\theta}$ for different noise realizations [16].

Remark 1: In some cases, it is possible for the optimal solution to be non-ergodic, indicating the presence of disconnected subgraphs. To address this issue, a potential solution is to add a small perturbation to the resulting Markov chain of the input. By doing so, the perturbed Markov chain becomes ergodic [16].

D. Contribution

Accurate sparse estimation requires a low mutual coherence, but the input design problem for nonlinear models lacks coherence constraints, as demonstrated in (11) and (8). This often results in high correlation between input sequences, leading to a regressor in the model (1) with a high mutual coherence. Although [18] proposes a method for limiting coherence in linear systems, it cannot be applied to nonlinear systems. The importance of low mutual coherence in nonlinear models, such as polynomial NARX and Volterra series, is greater than in linear models because nonlinear models tend to be intrinsically sparse. Therefore, this paper proposes a novel approach for designing inputs that enables the production of accurate sparse estimates in nonlinear models. The contributions of this paper are as follows:

1) An innovative method is introduced to design input sequences that provide precise sparse estimation. This is achieved by adding coherence constraints to the original approach described in equations (11) and (8), while still maintaining convexity of the optimization problem.
2) Theorem 2 establishes an upper bound on the mutual coherence that guarantees correct support estimation of $\boldsymbol{\theta}$ in (1) using the P_1^ϵ method, as stated in [19]. The ranges of the optimal objective function of both methods (the proposed approach and the original method outlined in (11)) are theoretically derived for comparison (Theorem 3).
3) The proposed method is numerically evaluated on a Volterra series model and compared to state-of-the-art algorithms.

II. Proposed algorithm

In this section, a new input design method for nonlinear sparse models is formulated.

According to the definition of mutual coherence in (2) and using the above per sample FIM, it is straightforward to show that

$$\mu_{\boldsymbol{\Phi}} = \max_{k\neq j} \frac{|\bar{\mathbf{I}}_F(k,j)|}{\sqrt{\bar{\mathbf{I}}_F(k,k)\bar{\mathbf{I}}_F(j,j)}}. \tag{12}$$

To account for the accuracy of sparse estimation, we propose incorporating a constraint of the form $\{\max_{k\neq j} \frac{|\bar{\mathbf{I}}_F(k,j)|}{\sqrt{\bar{\mathbf{I}}_F(k,k)\bar{\mathbf{I}}_F(j,j)}} \leq \eta\}$ into the optimization problems (8) and (11). Imposing this constraint on (8) results in the following convex optimization problem

$$\begin{aligned} \min_{p(\mathcal{U}_{n_{seq}})\in\mathcal{P}_{\mathcal{D}}} \quad & \text{Estimation error} \\ \text{st} \quad & \bar{\mathbf{I}}_F = \frac{1}{\sigma^2} \sum_{\mathcal{U}_{n_{seq}}\in\mathcal{D}^{n_{seq}}} \sum_{t=1}^{n_{seq}} \boldsymbol{\omega}_t \boldsymbol{\omega}_t^T p(\mathcal{U}_{n_{seq}}) \\ & \max_{k\neq j} \frac{|\bar{\mathbf{I}}_F(k,j)|}{\sqrt{\bar{\mathbf{I}}_F(k,k)\bar{\mathbf{I}}_F(j,j)}} \leq \eta \end{aligned} \tag{13}$$

To solve the problem (13), we can use a convex optimization toolbox such as CVX [20]. To incorporate the coherence constraint $|\bar{\mathbf{I}}_F(k,j)| \leq \eta\sqrt{\bar{\mathbf{I}}_F(k,k)\bar{\mathbf{I}}_F(j,j)}, k\neq j$, in CVX we can use the geometric mean by $\sqrt{\bar{\mathbf{I}}_F(k,k)\bar{\mathbf{I}}_F(j,j)} = geo_{mean}([\bar{\mathbf{I}}_F(k,k), \bar{\mathbf{I}}_F(j,j)])$. Furthermore, in order to constrain the mutual coherence in the aforementioned problem, a tuning parameter, η, is employed. The paper [7] presents a theoretical bound for η that guarantees the correct estimation of the support vector of $\boldsymbol{\theta}$ through P_1^ϵ. The following theorem describes this bound.

Theorem 2 ([7]): Consider the model $\mathbf{y} = \boldsymbol{\Phi}\boldsymbol{\theta}_0 + \mathbf{e}$ where $\mathbf{e} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, and where $\boldsymbol{\theta}_0$ denotes a sparse parameter vector with support Λ_0 and $\|\boldsymbol{\theta}_0\|_0 = s$. Assume that the mutual coherence of $\boldsymbol{\Phi}$ satisfies the bound $0 \leq \mu_{\boldsymbol{\Phi}} < \frac{1}{3s}$. Then the support of the solution $\hat{\boldsymbol{\theta}}$ obtained through P_1^ϵ is Λ_0 with probability at least

$$\Big(1 - \frac{(n_\theta - s)}{n_\theta^{1+\nu}}\Big)\Big(1 - e^{-\frac{s}{7}}\Big), \quad \forall \nu > 0. \tag{14}$$

According to this theorem, sparse parameter vectors can be accurately estimated using P_1^ϵ if there is a sufficiently small mutual coherence. While the above theorem has been established for P_1^ϵ, similar results can also be obtained for other sparse estimation methods, emphasizing the importance of coherence constraints in input design.

Moreover, adding the coherence constraint affects the estimation error compared to the main problem. Therefore, one of the questions that arises is how much the range of objective function changes by adding the coherence constraint. The next theorem answers this question.

Theorem 3: Consider the following optimization problem:

$$\begin{aligned} \max_{\beta_1,\beta_2,\cdots,\beta_{n_v}} \quad & \lambda_{\min}(\bar{\mathbf{I}}_F^{app}) \\ \text{st} \quad & \bar{\mathbf{I}}_F^{app} = \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)} \\ & \sum_{i=1}^{n_v} \beta_i = 1 \\ & \beta_i \geq 0, \ \forall i \in \{1,\cdots,n_v\}, \end{aligned} \tag{15}$$

in which $\bar{\mathbf{I}}(i)_F \geq 0$. Let the corresponding optimal objective function be denoted by $\lambda^*_{\min}$. Then it holds that $\lambda^*_{\min} \leq n_\theta I_F^{\max}$, in which $I_F^{\max} = \max_{1\leq i\leq n_v, 1\leq k\leq n_\theta} \bar{\mathbf{I}}_F^{(i)}(k,k)$. Adding the coherence constraint $\{\max_{k\neq j} \frac{|\bar{\mathbf{I}}_F^{app}(k,j)|}{\sqrt{\bar{\mathbf{I}}_F^{app}(k,k)\bar{\mathbf{I}}_F^{app}(j,j)}} \leq \eta\}$ to the above problem yields a new input design problem, for which we denote the optimal objective function by $\lambda^{\eta^*}_{\min}$. It then holds that $\lambda^{\eta^*}_{\min} \leq \lambda^*_{\min}$ and $\lambda^{\eta^*}_{\min} \leq (\eta(n_\theta - 1) + 1) I_F^{\max}$.

Proof: See Appendix I. ∎

Based on the above theorem, the upper bound of the optimal objective function in the new input design problem

is impacted by two factors: the number of parameters and the value of the tuning parameter η. When the value of η decreases, the range of the optimal objective function becomes smaller. However, when $\eta = 1$, the range is not affected. As the number of columns in matrix $\mathbf{\Phi}$ increases with n_θ, there are more correlations between these columns that must be restricted to obtain feasible solutions. This results in a larger reduction of the feasible set, which in turn causes an increase in the difference between the upper bounds of $\lambda^*_{\min}$ and $\lambda^{\eta^*}_{\min}$. However, even if the estimation accuracy (as measured by $\lambda^*_{\min}$, or another model quality measure) may seem to be reduced by including the coherence constraint, the estimation accuracy when a sparse estimation method is used may improve. This is in fact the reason for including the constraint. We will see an example of this in Section III.

The algorithm is summarized in Algorithm 1.

Algorithm 1 Coherence-based Input Design for Nonlinear Systems (CIDNS)

Require: input levels, η, $\boldsymbol{\omega}_t$ and n_{seq}.

1) Design the optimal per sample FIM by solving (13).
2) Design the input sequences through the method proposed in [16].
3) Construct the optimal regressor $\mathbf{\Phi}$ by the obtained input sequence from the previous step.
4) Apply a sparse estimation algorithm to model (1).

Output: $\hat{\boldsymbol{\theta}}$.

III. Numerical Study

This section assesses the effectiveness of the new proposed approach for designing inputs in nonlinear systems. As previously stated, the Volterra model is a commonly used nonlinear model, and its parameter vectors typically exhibit sparsity. For this numerical study, we examine the Volterra model presented in [21], which is given by

$$\begin{aligned} y(t) &= 0.5 + 0.8u(t) + 0.12u(t)u(t-1) \\ &- 0.43u(t-2)u(t-3) + 0.21u(t-2)u(t-4) + e(t) \end{aligned} \tag{16}$$

in which the observation was corrupted by additive white Gaussian noise $\mathbf{e} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, which was selected to obtain SNR (Signal-to-Noise Ratio) between 9 and 30. The SNR is given by

$$\text{SNR} = 10 \log \frac{\|\Phi\boldsymbol{\theta}\|_2^2}{\mathbb{E}(\mathbf{e}^T\mathbf{e})}. \tag{17}$$

As the model given in (16) has $P = 2$ kernels and $L = 5$, there are 21 parameters in total. However, out of these 21 parameters, only five of them possess non-zero values. Two ways of measuring the estimation performance are used: NRMSE, which quantifies the error in sparse estimation, and is defined as $\text{NRMSE} = \frac{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2}{\|\boldsymbol{\theta}_0\|_2}$, and POD, which measures the proportion of correctly identified non-zero coefficients in the parameter vector, and is given by

$$\text{POD} = \frac{\|\boldsymbol{\theta} \odot \boldsymbol{\theta}_0\|_0}{s}, \tag{18}$$

in which $\odot$ denotes element-wise product. Before computing the POD, coefficients with magnitudes lower than 10^{-3} are truncated to zero in the simulations. To determine the appropriate threshold value, we initially estimate the parameter vector resulting in $\boldsymbol{\theta}$ and apply 200 different thresholds spaced logarithmically within the range of $[10^{-5}, 10^{-2}]$. This yields parameter vectors $\boldsymbol{\theta}^t_i$, where $1 \le i \le 200$. Subsequently, we measure the error $err_i = \frac{|\|\mathbf{y}-\mathbf{\Phi}\boldsymbol{\theta}\|_2 - \|\mathbf{y}-\mathbf{\Phi}\boldsymbol{\theta}^t_i\|_2|}{\|\mathbf{y}-\mathbf{\Phi}\boldsymbol{\theta}\|_2}$. To fine-tune the threshold value, we identify the maximum threshold that satisfies the condition $err_i \le \epsilon$, where ϵ represents a small positive constant. The study considers the input levels in $\mathcal{D} \in \{-0.1, 0, 0.4, 0.8, 1.3, 1.6, 2.3\}$. The optimization problem used for designing the optimal input involves using $-\log\det(\mathbf{I}_F)$ as the estimation error. Consequently,

$$\max_{p(\mathcal{U}_{n_{seq}}) \in \mathcal{P}_\mathcal{D}} \quad \log\det(\bar{\mathbf{I}}_F) \tag{19a}$$

$$\text{st} \qquad \bar{\mathbf{I}}_F = \sum_{\mathcal{U}_{n_{seq}} \in \mathcal{D}^{n_{seq}}} \sum_{t=1}^{n_{seq}} \boldsymbol{\omega}_t \boldsymbol{\omega}_t^T p(\mathcal{U}_{n_{seq}}) \tag{19b}$$

$$|\bar{\mathbf{I}}_F(k,j)| \le \eta \sqrt{\bar{\mathbf{I}}_F(k,k)\bar{\mathbf{I}}_F(j,j)}, k \ne j. \tag{19c}$$

The number of free variables in the above optimization problem is $7^5 = 16807$. The CVX toolbox is employed to solve (19). For comparison, we use standard input design where the mutual coherence constraint (19c) has been removed, called IDNS (Input Design for Nonlinear Systems). The proposed method in [16] is utilized to generate the input sequences. To estimate the parameter vector $\boldsymbol{\theta}$, we employ the proposed method CIDNS in conjunction with sparse estimation algorithms LADMM (LASSO via Alternating Direction Method of Multipliers) [22] and Sparse Bayesian Approach to System Identification (referred to as SPBSI) [23], resulting in CIDNS-LADMM and CIDNS-SPBSI. Furthermore, we combine the standard method IDNS with ordinary Least Squares (LS) and LADMM, and SPBSI. This results in IDNS-LADMM, IDNS-SPBSI, and IDNS-LS. The simulations were conducted using MATLAB R2022a on a computer with a 4.00 GHz I7 CPU and 16 GB RAM, operating on the Microsoft Windows 10 operating system.

Two hyperparameters are required for the algorithms: λ_{SPBSI} (SPBSI as in [23]), and λ_{LADMM} (LADMM). We chose both to be 0.25 as this resulted in the lowest error $\|\mathbf{y}-\mathbf{\Phi}\boldsymbol{\theta}\|_2$ during experimentation. The number of observations is 1000.

To select an appropriate value for the parameter η, the coherence constraint can restrict the feasible set of IDNS substantially because of the nonlinear terms in the regressor. If a small value of η is chosen, this can lead to an infeasible problem. However, if η is too large, it can result in a high sparse estimation error. To address this, in this study, we have picked the value of $\eta = 0.9$ after conducting some preliminary experiments. In order to investigate the performance of the proposed method for different values of η, we also evaluate CIDNS-LADMM for $\eta = 0.95$. Our simulation results indicate that $\eta = 0.9$ is reasonable for the considered problem as it ensures problem feasibility while achieving accurate sparse estimation.

The determinant of the Fisher information matrix, $\det(\mathbf{I}_F)$, for IDNS and CIDNS with $\eta = 0.9$ were 0.18,

and 0.025, respectively. This suggests that the estimation accuracy should be worse for CIDNS. However, as remarked after Theorem 3, this holds for non-sparse estimation. To evaluate the performance when sparse estimation is used we first note that the mutual coherence obtained through IDNS is 0.99, while it has dropped to $\mu_{\mathbf{\Phi}} = 0.9$ for CIDNS with $\eta = 0.9$, suggesting improved behavior of CIDNS for sparse estimation. To further examine this we have conducted 300 experiments for each SNR level $\{9, 12, 15, 18, 21, 24, 27, 30\}$, reported in Figures 1a and 1b, respectively. From these figures, we see a significant improvement when using the coherence constraint, both with respect to the NRMSE and with respect to POD. Based on these numerical results, it can be concluded that using a coherence constraint, as in CIDNS, in conjunction with sparse estimation methods such as LADMM and SPBSI, significantly reduces the NRMSE and leads to higher POD, especially at low SNRs. This implies that reducing the mutual coherence can lead to improved performance in sparse estimation.

To examine how the coherence constraint affects problem (19), we plot the histogram of the resulting input sequences from CIDNS with $\eta = 0.9$ and IDNS in Figure 1c. The figure reveals that IDNS prefers to use input levels of $\{-0.1, 1.3, 2.3\}$, whereas CIDNS uses $\{-0.1, 0.8, 2.3\}$. Additionally, Figure 2a depicts the resulting optimal pmf from IDNS and CIDNS. This figure shows that reducing the coherence leads to greater sparsity in the pmf, with the optimal pmf of CIDNS being more sparse than that of IDNS. Furthermore, the maximum probabilities in CIDNS occur when it designs input sequences using input levels of -0.1 and 0.8. This indicates that CIDNS attempts to decrease coherence by focusing more on the input levels of -0.1 and 0.8 among the seven input levels.

Figures 2b and 2c depict the input signal variations over time at input levels that yield the maximum resulting probability from CIDNS with $\eta = 0.9$ and IDNS, respectively, as shown in Figure 2a. In Figure 2c, the input signals 2, 3, 4, and 5 alternate between input levels of -0.1 and 2.3 over time, while input signal 1 remains constant, resulting in high mutual coherence in IDNS. Conversely, when a coherence constraint is present, as illustrated in Figure 2b, the input signals fluctuate between -0.1 and 0.8, resulting in non-constant input signals over time.

IV. Conclusion

This paper introduces a novel technique to enhance sparse estimation accuracy in nonlinear systems by restricting the mutual coherence during input design. The method's objective is to generate a regressor with low coherence, which is critical for accurate sparse estimation. The maximum ratio between the diagonal and non-diagonal elements of the Fisher information matrix determines the mutual coherence, making it easy to incorporate into standard input design problems, while the resulting problems remain convex. It is theoretically analyzed how the proposed idea affects the optimal objective function of the original approach. Based on numerical results, the proposed approach has a lower normalized root mean square error and a higher accuracy for identifying non-zero elements in the parameter vector than state-of-the-art input design algorithms.

Appendix I
Proof of Theorem 3

To prove Theorem 3, we first find the range of the optimal cost of the original problem (15). Using the first constraint of (15) results in

$$\bar{\mathbf{I}}_F = \begin{bmatrix} \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(1,1) & \cdots & \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(1,n_\theta) \\ \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(2,1) & \cdots & \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(2,n_\theta) \\ \vdots & \cdots & \vdots \\ \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(n_\theta,1) & \cdots & \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(n_\theta,n_\theta) \end{bmatrix}. \quad (20)$$

According to the Gershgorin circle theorem [24], we have

$$\Big|\lambda_m - \sum_{k\neq j, j=1}^{n_\theta} \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(k,j)\Big| \leq \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(k,k), \quad (21)$$

in which λ_m denotes the m-th eigenvalue of $\bar{\mathbf{I}}_F$. Then, we can write

$$\lambda_m \leq \sum_{k\neq j, j=1}^{n_\theta} \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(k,j) + \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(k,k). \quad (22)$$

Since $I_F^{\max} = \max_{1\leq i\leq n_v, 1\leq k\leq n_\theta} \bar{\mathbf{I}}_F^{(i)}(k,k)$ and $\sum_{i=1}^{n_v} \beta_i = 1$, we can write

$$\begin{aligned} \lambda_m &\leq \sum_{k\neq j, j=1}^{n_\theta} \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)}(k,j) + \sum_{i=1}^{n_v} \beta_i I_F^{\max} \\ &\leq \sum_{k\neq j, j=1}^{n_\theta} \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{nd}(k,j) + I_F^{\max} = \sum_{k\neq j, j=1}^{n_\theta} \bar{\mathbf{I}}_F^{nd}(k,j) + I_F^{\max} \end{aligned} \quad (23)$$

where $\bar{\mathbf{I}}_F^{nd}(k,j) = \max_{1\leq i\leq n_v} (\bar{\mathbf{I}}_F^{(i)}(k,j))$. Since $\bar{\mathbf{I}}_F^{nd}(k,j) \leq I_F^{\max}$, we have,

$$\lambda_m \leq \sum_{k\neq j, j=1}^{n_\theta} \bar{\mathbf{I}}_F^{nd}(k,j) + 1 \leq (n_\theta - 1) I_F^{\max} + I_F^{\max} = n_\theta I_F^{\max}.$$

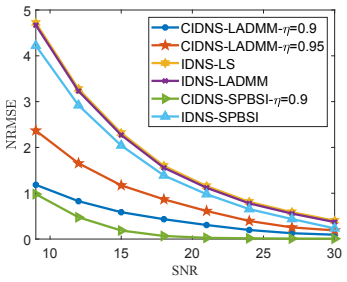
On the other hand, since the FIM is a positive definite matrix, $\lambda_m > 0$. Using above inequality, we have

$$0 < \lambda_m \leq n_\theta I_F^{\max}. \quad (24)$$

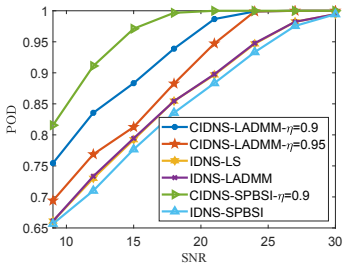
Finally, according to the above inequality, the optimal cost of the convex problem (15) satisfies $\lambda_{\min}^* \in (0, n_\theta I_F^{\max}]$. We add the coherence constraint to the original problem which results in

$$\begin{aligned} \max_{\beta_1, \beta_2, \cdots, \beta_{n_v}} \quad & \lambda_{\min}(\bar{\mathbf{I}}_F^{app}) \\ \text{st} \quad & \bar{\mathbf{I}}_F^{app} = \sum_{i=1}^{n_v} \beta_i \bar{\mathbf{I}}_F^{(i)} \\ & \max_{k\neq j} \frac{|\bar{\mathbf{I}}_F^{app}(k,j)|}{\sqrt{\bar{\mathbf{I}}_F^{app}(k,k)\bar{\mathbf{I}}_F^{app}(j,j)}} \leq \eta \\ & \sum_{i=1}^{n_v} \beta_i = 1 \\ & \beta_i \geq 0, \ \forall i \in \{1, \cdots, n_v\}. \end{aligned} \quad (25)$$

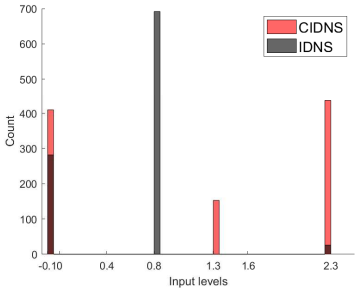
Similarly, the optimal cost of the above problem is derived. Since $I_F^{\max} = \max_{1\leq i\leq n_v, 1\leq k\leq n_\theta} \bar{\mathbf{I}}_F^{(i)}(k,k)$, it is straightforward to show that $\bar{\mathbf{I}}_F^{app}(j,j) \leq I_F^{\max}$. Using (23) results in



(a) NRMSE

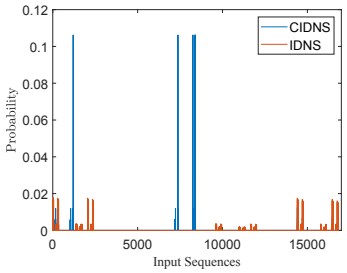


(b) POD

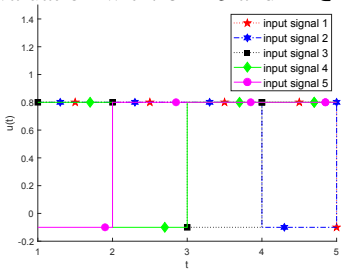


(c) Histogram of optimal input sequences

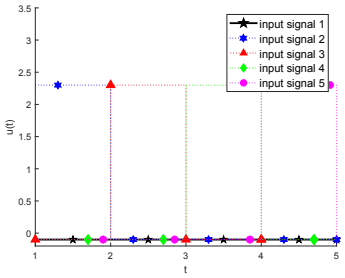
Fig. 1: Evaluation with: $s = 5$ and $\mathbf{\Phi} \in \mathbb{R}^{1000\times 21}$.



(a) Probability mass function



(b) CIDNS



(c) IDNS

Fig. 2: Evaluation of input design problems.

$$\lambda_m^{\eta} \leq \sum_{k\neq j, j=1}^{n_\theta} \bar{\mathbf{I}}_F^{app}(k,j) + I_F^{\max},$$

where λ_m^{η} denotes the m-th eigenvalue of $\bar{\mathbf{I}}_F^{app}$. According to the coherence constraint, we have $\max_{k\neq j} |\bar{\mathbf{I}}_F^{app}(k,j)| \leq \eta I_F^{\max}$.

Using above inequality we can write

$$\lambda_m^{\eta} \leq \sum_{k\neq j, j=1}^{n_\theta} \bar{\mathbf{I}}_F(k,j) + I_F^{\max} \leq (\eta(n_\theta - 1) + 1) I_F^{\max}.$$

Thus, the optimal cost of the problem (25) satisfies $\lambda_{\min}^{\eta^*} \in (0, (\eta(n_\theta - 1) + 1) I_F^{\max}]$. Moreover, since there is an additional constraint in (25) compared to (15), it is straightforward to show that $\lambda_{\min}^{\eta^*} \leq \lambda_{\min}^{*}$. This concludes the proof of Theorem 3.

REFERENCES

[1] Johan Schoukens and Lennart Ljung, "Nonlinear system identification: A user-oriented road map," *IEEE Control Systems Magazine*, vol. 39, no. 6, pp. 28–99, 2019.

[2] L. Ljung, *System Identification: Theory for the User*, Prentice-Hal, 1999.

[3] M. Elad, *Sparse and Redundant Representations*, Springer, 2010.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society (Series B)*, vol. 58, pp. 267–288, 1996.

[5] J. Parsa and H Hjalmarsson, "On reducing the coherence in sparse system identification," *59th IEEE Conference on Decision and Control (CDC)*, 2020.

[6] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[7] Zvika Ben-Haim, Yonina C. Eldar, and Michael Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5030–5043, 2010.

[8] D. L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Info. Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[9] G. Goodwin and R. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*, Academic Press, New York, 1977.

[10] X. Bombois, G. Scorletti, M. Gevers, P. Van den Hof, and R. Hildebrand, "Least costly identification experiment for control," *Automatica*, pp. 1651–1662, 2006.

[11] H. Hjalmarsson, "System identification of complex and structured systems," *European Journal of Control*, pp. 275–310, 2009.

[12] Christian A. Larsson, Hkan Hjalmarsson, and Cristian R. Rojas, "On optimal input design for nonlinear FIR-type systems," in *49th IEEE Conference on Decision and Control (CDC)*, 2010, pp. 7220–7225.

[13] Patricio E. Valenzuela, Cristian R. Rojas, and Hkan Hjalmarsson, "Optimal input design for non-linear dynamic systems: A graph theory approach," in *52nd IEEE Conference on Decision and Control*, 2013, pp. 5740–5745.

[14] J. Parsa and H. Hjalmarsson, "Optimal input design through infinity norm minimization using proximal mapping," in *60th IEEE Conference on Decision and Control*, 2021.

[15] J. Parsa, C. R. Rojas, and H. Hjalmarsson, "Low coherence regressor design in sparse system identification," *IEEE Transaction of Automatic Control (Submitted)*, 2021.

[16] Patricio E. Valenzuela, Cristian R. Rojas, and Hkan Hjalmarsson, "A graph theoretical approach to input design for identification of nonlinear dynamical models," *Automatica*, vol. 51, pp. 233–242, 2015.

[17] Arif Zaman, "Stationarity on Finite Strings and Shift Register Sequences," *The Annals of Probability*, vol. 11, no. 3, pp. 678 – 684, 1983.

[18] Javad Parsa, Cristian R. Rojas, and Hkan Hjalmarsson, "Application-oriented input design with low coherence constraint," *IEEE Control Systems Letters*, vol. 7, pp. 193–198, 2023.

[19] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *In Proc. Asilomar Conf. Signal Syst. Comput.*, 1993.

[20] M. Grant and S. Boyd., "Cvx: Matlab software for disciplined convex programming (web page and software)," Tech. Rep., Stanford.

[21] Sayed Mohammad Reza Loghmanian, Rubiyah Yusof, and Marzuki Khalid, "Nonlinear dynamic system identification using volterra series: Multi-objective optimization approach," in *2011 Fourth International Conference on Modeling, Simulation and Applied Optimization*, 2011, pp. 1–5.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[23] W. Pan, Y. Yuan, J. Gonalves, and GB. Stan, "A sparse bayesian approach to the identification of nonlinear state-space systems," *IEEE Trans. on Automatic Control*, vol. 61, no. 1, 2016.

[24] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, Cambridge University Press, 2 edition, 2012.