# Masked Multiple State Space Model Identification Using FRD and Evolutionary Optimization

Mehmet Önder Efe, Burak Kürkçü, Coşku Kasnakoğlu, Zaharuddin Mohamed, and Zhijie Liu, *Member, IEEE*

***Abstract*—Identification of dynamical systems from frequency response data (FRD) has extensively been studied and effective techniques have been developed. Given different FRD sets obtained from different systems and a fixed state space model structure, is it possible to find a constant parameter vector containing $(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})$ quadruple's numerical content and a FRD-associated mask vector set that approximates the spectral information available in each FRD set? This article proposes a genetic algorithm based optimization approach to determine the real parameter vector $(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})$ and the binary mask vector through a sequential optimization scheme. We study state space models for matching FRD from multiple systems. Results show that the proposed optimization approach solves the problem and compresses multiple dynamical models into a single masked one.**

***Index Terms*—Genetic algorithms (GAs), identification, masked models, optimization, state space models.**

## NOMENCLATURE

| | |
|---|---|
| $\alpha_b$ | Random vector containing zeros and ones. |
| $\gamma_c$ | Small positive value for mutation operator in real parameter optimization subroutine. |
| $\lambda_m$ | Relative importance weight of the magnitude term. |
| $\lambda_p$ | Relative importance weight of the phase term. |
| $\lambda_s$ | Relative importance weight of the stability term. |
| $\mathbb{B}$ | Real set containing only 0 and 1. |
| $\mathbb{R}$ | Set of real numbers. |
| $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}$ | System matrices. |
| $\mathbf{M}$ | Concatenated form of mask matrices. |
| $\mathbf{m}$ | Concatenated (row) mask vector. |
| $\mathbf{m}^*$ | Best value of the concatenated mask vector observed so far. |
| $\mathbf{m}_l$ | Adjustable real parameter vector for the $l$th FRD. |
| $\mathbf{W}$ | Concatenated form of system matrices. |
| $\mathbf{w}$ | Adjustable real parameter vector. |
| $\mathbf{w}^*$ | Best value of the parameter vector observed so far. |
| $\mathbf{x}$ | State vector. |
| $\mathbf{x}_s$ | State vector of the subspace method. |
| $\mathcal{S}^b\{\cdot\}$ | Random initializer of binary vector sets. |
| $\mathcal{S}^c\{\cdot\}$ | Random initializer of real vector sets. |
| $\mu_b$ | Threshold value for mutation operator in binary parameter optimization subroutine. |
| $\mu_c$ | Threshold value for mutation operator in real parameter optimization subroutine. |
| $\omega_{lp}$ | $p$th frequency point in the $l$th FRD. |
| $\sigma_c$ | Constant for mutation operator in real parameter optimization subroutine. |
| $\Upsilon$ | Row generating operator. |
| $\{\mathbf{m}\}_1^{n_b}$ | Mask vector population. |
| $\{\mathbf{w}\}_1^{n_c}$ | Parameter vector population. |
| $d_{lp}$ | $p$th measured value in $l$th FRD. |
| $g_{\max}$ | Maximum iteration (generation) number. |
| $J$ | Cost (loss) function. |
| $L$ | Number of different systems to be identified. |
| $n_b$ | Population size for binary search. |
| $n_c$ | Population size for real search. |
| $n_{Cb}$ | Number of binary offsprings generated at each generation. |
| $n_{Cc}$ | Number of real offsprings generated at each generation. |
| $n_c$ | Population size for real parameter search. |
| $P_l$ | Number of rows in $l$th FRD. |
| $r_c$ | Normal distributed random variable ($\mathcal{N}(0,1)$). |
| $r_m$ | Uniformly distributed random variable from (0,1). |
| $u$ | Input of the state space system. |
| $y$ | Output of the state space system. |

Manuscript received 22 November 2023; revised 12 February 2024 and 30 March 2024; accepted 9 April 2024. Date of publication 26 April 2024; date of current version 2 July 2024. Paper no. TII-23-4658. *(Corresponding author: Mehmet Önder Efe.)*

Mehmet Önder Efe is with the Department of Computer Engineering, Hacettepe University, Ankara 06800, Türkiye (e-mail: onderefe@hacettepe.edu.tr).

Burak Kürkçü is with the Department of Mechanical Engineering, University of California, Berkeley, CA 94720 USA, on leave from the Department of Computer Engineering, Hacettepe University, Ankara 06800, Türkiye (e-mail: bkurkcu@berkeley.edu).

Coşku Kasnakoğlu is with the Electrical and Electronics Engineering Department, TOBB University of Economics and Technology, Ankara 06510, Türkiye (e-mail: kasnakoglu@etu.edu.tr).

Zaharuddin Mohamed is with the Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia (e-mail: zahar@fke.utm.my).

Zhijie Liu is with the School of Intelligence Science and Technology, University of Science and Technology Beijing, Beijing 100083, China (e-mail: liuzhijie2012@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2024.3388605.

Digital Object Identifier 10.1109/TII.2024.3388605

$y_s$ Output of the subspace identification method based model.

## I. Introduction

SYSTEM identification has been a core issue for many decades and as the tools of data acquisition become easily accessible, the identification algorithms constitute the crux of the control system design process and effective algorithms emerge. Many successful approaches have been proposed in the past and the ones based on frequency response data (FRD) had numerous industrial and scientific consequences.

Given a frequency response dataset obtained over a set of frequency points, a parameterized model can be tuned to match the magnitude and phase information available in the FRD, [1], [2], [3], [4]. The tuned model is accompanied by a performance level that is tightly dependent upon the order of the model. In the literature, techniques yielding the transfer function (TF) models have become popular, and commercial software adopt such techniques for practicing engineers, [5], [6]. For single-input-single-output (SISO) TF models of order $n$ have at most $2n$ adjustable parameters and depending on the complexity of the given FRD, large values of $n$ may be inevitable, [7]. The same SISO system identification problem can be cast into state space models as well. In such a setting, the matrices embodying the system require tuning of $(n+1)^2$ parameters. Obviously, the search space for state space models have more flexibility than TF models and in this article, we adopt state space models.

Toward this goal, many aspects of the system identification problem has been scrutinized in the past. Identification of linear/nonlinear systems under the presence of noise has been considered in [8], [9], [10], [11], linear and parameter varying (LPV) model identification using neural networks is reported by [12], [13], [14], using block structured architecture [15] and numerical algorithms for subspace identification (N4SID) have been studied in [16] and [17], subspace identification based on FRD is reported in [18]. Regarding the subspace identification approaches, identification from time domain data using combined invariant subspace and subspace identification method has been proposed by [19]. Modeling of the latent dynamics in state space for high dimensional time series in [20] and use of conditional maximum likelihood identification in the context of identifying one general state space system in [21] are the remarkable works reporting successful results. Identification of Hammerstein–Wiener systems using a combination of gravitational search algorithm and particle swarm optimization is presented in [22]. Closed-loop delta-operator-based subspace identification method for continuous-time systems is proposed in [23], where instrumental variable method and principal component analysis are utilized to solve the problems of biased results as the system operates in closed loop. Identification of convergent continuous-time Lur'e-type state space systems is studied in [24] and use of machine learning has been introduced as a remedy to the data driven identification of dynamical systems [25], [26].

The works cited so far considered single FRD, i.e., a set of frequency points and a set of associated complex numbers, describing the state space system numerically and indirectly. The optimization process finds a set of real parameters, which are the numerical content of $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ matrices for a state space system. After the modeling phase, those parameters are frozen for the available datasets.

The importance of the problem addressed here is that multiple state space systems can be obtained by switching from a single mother system via mask vectors. Especially when hardware applications are considered, the fact that many spectral features can be implemented on a single hardware will lead to future electronic systems being reduced in size and being designed more effectively. Another field of application is to represent the approximate dynamical behavior of irrational TFs, like heat flow generating an infinite dimensional TF or processes governed by partial differential equations, [27]. Further to these, the method can be used to store a family of responses generated by varying a particular design parameter of a unique system. For example, a family of proportional integral and derivative (PID) controllers working at multiple operating points [28] or filters [29] can be contained within a single mother model. Last but not the least, the approach presented here is extendable to discrete time systems.

Now consider a set of frequency points and more than one FRD sets (frequency responses) associated to different systems, choose a state space model and simultaneously tune the $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ matrices in such a way that different switchings reproduce the corresponding system's FRD. The approach proposed in this work makes use of the redundant capacity in a given state space model in such a way that maximal overlaps among the given datasets are discovered and their combinations are activated via appropriate on/off switchings of the relevant parameters to reproduce the FRD sets via a single, yet switched, system also called a *subsystem* in the sequel.

The stage set for the problem involves optimizing a parameter set that contains real numbers and another parameter set composed of binary numbers i.e., mask vectors. The former establishes the system's numerical content, while the latter activates/deactivates the associated variables. This obviously requires an iterative optimization and we adopt a genetic optimization approach for parameter tuning, which best suits the needs of the current work. Our contribution to the subject area explains the aforementioned optimization scheme for storing multiple spectral information into a single mother state space model.

This rest of this article is organized as follows: Section II describes the problem analytically. Section III explains the evolutionary approach adopted to solve the problem and the algorithmic flow. Section IV is devoted the numerical studies. Section V provides the comparison with subspace identification method. Finally, Section VI concludes this article.

## II. Problem Formulation

Systems and control theory provides many useful alternatives to reach a dynamic model that is based on a given FRD. These approaches require choosing a model, then a fitting algorithm

optimizes the model parameters in such a way that a chosen loss function is minimized. Loss function may quantify the similarity of spectral results or time domain results. The available algorithms typically generate a TF that matches the given FRD, which might have been obtained from a continuous time or discrete time system, with what is generated by the developed model, [1], [2], [3], [5], [6].

In the literature, we have very mature algorithms that are based on single dataset based identification of systems, i.e., a set of frequencies are selected and the corresponding complex numbers are associated to each particular frequency, then a model is devised. In this study, we propose an approach that switches ONor switches OFF the values inside a state space model such that each particular switching of the dynamic model generates one of the given FRD sets. The question here is *how we could achieve such a real and binary variable set*.

Let $\mathbf{x} \in \mathbb{R}^{n\times 1}$ be the state vector, $u \in \mathbb{R}$ be the scalar input signal, $y \in \mathbb{R}$ be the output signal and $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times 1}$, $\mathbf{C} \in \mathbb{R}^{1\times n}$ and $\mathbf{D} \in \mathbb{R}$ be the system matrices. A continuous time dynamic system in state space can be given by (1)–(2) or in (3), which we will call the mother system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{1}$$

$$y = \mathbf{C}\mathbf{x} + \mathbf{D}u \tag{2}$$

$$\begin{pmatrix}\dot{\mathbf{x}}\\ y\end{pmatrix} = \begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ u\end{pmatrix}. \tag{3}$$

Let there be $L$ different FRD sets. For the $l$th FRD, let $\mathbf{M}_l := \left(\begin{smallmatrix}\mathbf{m}_{l\mathbf{A}} & \mathbf{m}_{l\mathbf{B}}\\ \mathbf{m}_{l\mathbf{C}} & \mathbf{m}_{l\mathbf{D}}\end{smallmatrix}\right) \in \mathbb{B}^{(n+1)\times(n+1)}$ be the mask matrix and $\mathbf{W} := \left(\begin{smallmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{smallmatrix}\right) \in \mathbb{R}^{(n+1)\times(n+1)}$ be the parameter matrix that is to be masked by $\mathbf{M}_l$. The $l$th masked system that approximates the $l$th FRD can be given as

$$\dot{\mathbf{x}}_l = (\mathbf{m}_{l\mathbf{A}} \odot \mathbf{A})\mathbf{x}_l + (\mathbf{m}_{l\mathbf{B}} \odot \mathbf{B})u \tag{4}$$

$$y_l = (\mathbf{m}_{l\mathbf{C}} \odot \mathbf{C})\mathbf{x}_l + (\mathbf{m}_{l\mathbf{D}} \odot \mathbf{D})u \tag{5}$$

where the operator $\odot$ performs an element-wise multiplication of its arguments and $l = 1, 2, \ldots, L$. Let $\mathbf{a}_i$ be the $i$th column of $\mathbf{A}$, then we have $\mathbf{A} = (\mathbf{a}_1\mathbf{a}_2\ldots\mathbf{a}_n)$. Define the row generating operator $\Upsilon$ as $\Upsilon\{\mathbf{A}\} := (\mathbf{a}_1^T\mathbf{a}_2^T\ldots\mathbf{a}_n^T)$, which is an $1 \times n^2$ vector. Now, define $\mathbf{w} := \Upsilon\{\mathbf{W}\}$, $\mathbf{m}_l := \Upsilon\{\mathbf{M}_l\}$, $\mathbf{m} := (\mathbf{m}_1\mathbf{m}_2\ldots\mathbf{m}_L)$ and minimize the cost in the following:

$$J := \frac{\lambda_m}{P_l}\sum_{l=1}^{L}\sum_{p=1}^{P_l}\Big(\left(20\log|d_{lp}| - 20\log|y_l(\mathbf{w},\mathbf{m}_l,\omega_{lp})|\right)^2 \frac{\lambda_p}{P_l}\angle\left(d_{lp} - y_l(\mathbf{w},\mathbf{m}_l,\omega_{lp})\right)^2\Big) + \lambda_s\sum_{l=1}^{L}\frac{\operatorname{sgn}\left(|\lambda_{\max}\{\mathbf{m}_{l\mathbf{A}} \odot \mathbf{A}\}|\right)+1}{2} \tag{6}$$

where $\lambda_m > 0$, $\lambda_p > 0$, and $\lambda_s > 0$ are weight parameters determining the relative importance of the magnitude, phase, and the stability information, respectively. In above, $\angle(.)$ is the angle operator and $\lambda_{\max}(\cdot)$ stands for the maximum eigenvalue. Choosing $\lambda_s > 0$ ensures performing search among solutions having all left half plane (LHP) poles, i.e., stable models. For the state space multiple model identification problem, the complex number $d_{lp}$ is the scalar target value for the $l$th frequency response dataset's $p$th row and $\omega_{lp}$ is the frequency value in the $l$th dataset's $p$th row, i.e.,

$$y_l(\mathbf{w},\mathbf{m}_l,\omega_{lp}) = (\mathbf{m}_{l\mathbf{C}} \odot \mathbf{C})(j\omega_{lp}\mathbf{I}_{n\times n} - \mathbf{m}_{l\mathbf{A}} \odot \mathbf{A})^{-1}\times (\mathbf{m}_{l\mathbf{B}} \odot \mathbf{B}) + (\mathbf{m}_{l\mathbf{D}} \odot \mathbf{D}) \tag{7}$$

where

$$\mathbf{m}_l = \Upsilon\left\{\begin{pmatrix}\mathbf{m}_{l\mathbf{A}} & \mathbf{m}_{l\mathbf{B}}\\ \mathbf{m}_{l\mathbf{C}} & \mathbf{m}_{l\mathbf{D}}\end{pmatrix}\right\}, \mathbf{w} = \Upsilon\left\{\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{pmatrix}\right\}. \tag{8}$$

Based on the above formulation, the inherent connection between the mask and the mother system is defined in (4) and (5), where Binary mask values and real parameter values are to be optimized by minimizing the cost function given in (6). For a frozen set of real parameters, every different mask vector defines a new dynamical system, i.e., a new FRD. The crux of the presented approach is to find out a good mask vector and a good real parameter vector that fits all FRDs with an acceptable cost value.

Denoting the concatenated set of frequencies by $\mathcal{F}$ and the concatenated set of FRD sets by $\mathcal{D}$, the optimization problem for the multiple FRD identification problem can be given as follows:

$$\begin{aligned}&\text{minimize } J(\mathbf{w},\mathbf{m},\mathcal{F},\mathcal{D})\\ &\text{s.t. } \mathbf{w} \in \mathbb{R}^{1\times(n+1)^2}, \mathbf{m} \in \mathbb{B}^{1\times L(n+1)^2}\end{aligned} \tag{9}$$

where the order of the system, $n$, is chosen by the designer. The value of $\lambda_p$ in (6) determines the relative importance of phase information. If $\lambda_p = 0$, the optimization approach matches solely the magnitudes and the resulting system may be a nonminimum phase system although the FRD generating counterpart was a minimum phase one, or vice versa. This is further strengthened by using the third term in (6), where the stability of the masked models are ensured by choosing an appropriate value for $\lambda_s$. One has to note that models obtained via minimizing the cost in (9) are valid only over the chosen frequency spectrum. At this point, one could wonder why genetic algorithms (GAs) are chosen instead of deterministic approaches. The answer to this question is threefold: 1) Deterministic methods often struggle in such scenarios as they attempt to evaluate the entire solution space in a single step. In such cases, GAs can be more effective due to their ability to explore potential solutions across a wide search space through parallel computation and population-based approaches. 2) Certain optimization problems tend to get stuck in local minima with deterministic methods. Especially in complex objective functions with multiple local minima, GAs may perform better as they have the ability to explore potential solutions across a broad search space. 3) If the optimization problem exhibits a certain structure or characteristic as we have here, GAs are more suitable for leveraging or adapting to such structures. Remembering the search space containing real and binary variables, it becomes comprehensible to advice using GAs. In the next section, we explain how an evolutionary algorithm based iterative approach can be implemented to solve the problem.

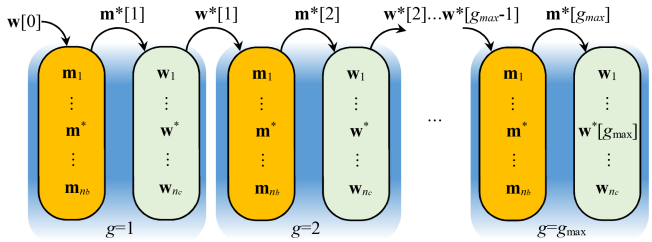


Fig. 1. Temporal flow of the optimization algorithm.

## III. Genetic Algorithms

GAs have demonstrated very successful results in various optimization problems. Their independence from the need for derivatives in the loss function grants them significant preference. This attribute becomes notably advantageous in certain contexts, [30], [31]. In this section, we focus our discussion on binary and real search algorithms tailored for the loss function in (6). The challenge lies in establishing the hierarchy of binary and real search subroutines.

The initial collection of mask vectors is chosen randomly from $\mathbb{B}^{n_b \times L(n+1)^2}$, where $n_b$ represents the number of candidate solutions, corresponding to the population's size. To identify the best performing mask vector within this set, an initial vector $\mathbf{w}$ is required. $\mathbf{w}$ is chosen randomly from $\mathbb{R}^{1\times(n+1)^2}$. The optimal mask vector is determined by evaluating the cost $J(\mathbf{w}, \mathbf{m}, \mathcal{F}, \mathcal{D})$.

The process involves a loop over generations denoted by $g$. In the first phase of the loop, the binary subroutine manages parent selection, crossover, mutation, and evaluates the cost for the subsequent mask generation ($\mathbf{m}$). The best performing mask vector for the $g$th generation is established, with $\mathbf{w}$ remaining unchanged.

For the real variable optimization, the first generation is specially configured. The previously used $\mathbf{w}$ is included in the first generation, derived from $\mathbb{R}^{n_c\times(n+1)^2}$. Here, $n_c$ denotes the population size for the GA that optimizes the real variables $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ of the optimization problem. The other entries, excluding the initial $\mathbf{w}$ vector, are randomly assigned in the first generation. The real variable optimization employs the best mask vector from the preceding subroutine, generating the next iteration through parent selection, crossover, mutation operations, and cost evaluation using (6). Subsequently, a new population of $\mathbf{w}$ vectors is obtained, and the loop progresses by returning to the binary (mask) optimization stage utilizing the best $\mathbf{w}$ vector found. This process is summarized in Algorithm 1 and depicted in Fig. 1. In Fig. 1, $g = 1$ stage, a random real vector ($\mathbf{w}$) enters and an optimized mask ($\mathbf{m}$) is generated for the next generation.

In Fig. 1, each generation comprises two consecutive subroutines. The sequence starts with $\mathbf{w}[0]$ and, after acquiring the optimal mask ($\mathbf{m}^\star$), the algorithm seeks the best real vector ($\mathbf{w}^\star$), and so forth. The essence of this optimization process lies in the fact that the binary subroutine uses the locally best parameter vector ($\mathbf{w}^\star$), while the real optimization process employs the locally best mask vector ($\mathbf{m}^\star$) from consecutive

**Algorithm 1:** Evolutionary Search for $\mathbf{m}$ and $\mathbf{w}$.

1: Initial mask generation $\mathcal{S}^b\{\{\mathbf{m}\}_1^{n_b}\}$
2: Initial parameter vector $\mathcal{S}^c\{\mathbf{w}\}$ ($\mathbf{w}^*$ initially)
3: **for** $g = 1$ to $g_{\max}$ **do**
4: `%Binary optimization subroutine`
5: Parent selection from $\{\mathbf{m}\}_1^{n_b}$
6: Crossover among selected (binary) parents
7: Mutation in $\{\mathbf{m}\}_1^{n_b}$
8: Cost evaluation with $\{\mathbf{m}\}_1^{n_b}$ and $\mathbf{w}^*$
9: Sort and select
10: Choose current best $\mathbf{m}^* \leftarrow \text{argmin} J(\{\mathbf{m}\}_1^{n_b}, \mathbf{w}^*, \mathcal{F}, \mathcal{D})$
11: `% Real optimization subroutine`
12: **if** $g == 1$ **then**
13: Initial parameter generation $\mathcal{S}^c\{\{\mathbf{w}\}_1^{n_c-1}\}$
14: Append $\mathbf{w}^*$ (used initially) as $n_c^{\text{th}}$ candidate
15: **end if**
16: Parent selection from $\{\mathbf{w}\}_1^{n_c}$
17: Crossover among selected (real) parents
18: Mutation in $\{\mathbf{w}\}_1^{n_c}$
19: Cost evaluation with $\{\mathbf{w}\}_1^{n_c}$ and $\mathbf{m}^*$
20: Sort and select
21: Choose current best $\mathbf{w}^* \leftarrow \text{argmin}\, J(\mathbf{m}^*, \{\mathbf{w}\}_1^{n_b}, \mathcal{F}, \mathcal{D})$
22: **if** $J <= tolerance$ **then**
23: Exit
24: **end if**
25: **end for**

generations. This iterative procedure gradually minimizes the cost denoted by $J$ in (6), ultimately leading to the discovery of an optimal or near-optimal solution. Depending on the accuracy expectations, the designer is allowed to seek for a solution in larger dimensional state space models, which may offer better accuracies.

In Algorithm 1, the crossover operation (line 6) functions as a two-input, two-output operator. It operates based on a roulette wheel selection strategy, accepting two vectors (parents) and executing multiple exchanges to generate two offspring vectors. In our algorithm, a random number determines the strategy for the current parents. Among three equiprobable outcomes, the first outcome employs a single-point crossover, the second uses a double-point crossover, and the last outcome generates a random vector $\alpha_b \in \mathbb{B}^{1\times(n+1)^2}$ comprising zeros and ones. The two offspring vectors are derived using the uniform crossover approach, defined as follows:

$$\text{offspring}_1 = \alpha_b \odot \text{parent}_1 + (\mathbf{1} - \alpha_b) \odot \text{parent}_2$$
$$\text{offspring}_2 = (\mathbf{1} - \alpha_b) \odot \text{parent}_1 + \alpha_b \odot \text{parent}_2$$

where $\mathbf{1}$ is a $1 \times (n+1)^2$ vector composed of all ones. The strategy defined above eliminates the drawbacks of using single type crossover and we generate $n_{Cb}$ offsprings, where $n_{Cb}$ is an even number as we generate two offsprings at each trial and it is equal to $n_b$ in this study.

Mutation operator in line 7 of Algorithm 1 is a single input single output operator that enhances the exploratory capability of the search mechanism. At the $g$th generation, the $i$th entry of $\mathbf{m}[g]$, denoted by $\mathbf{m}_i[g]$, is modified using a uniformly distributed random number, $r_m$, which varies in between 0 and 1. For a given threshold, say $\mu_b$, if $r_m < \mu_b$, the value of $\mathbf{m}_i[g]$ is flipped, i.e., $\mathbf{m}_i[g] \leftarrow 1 - \mathbf{m}_i[g]$. Since the distribution of $r_m$ is uniform, the variable $\mu_b$ is the flipping probability of the relevant variable.

At every generation, the population together with the offsprings are evaluated using the loss function in (6), the individuals are sorted according to their performances. The poorest ones are then deleted to maintain the population size constant, i.e., $n_b$. This ensures that the $(g+1)$th population will not contain poorer individuals than the poorest one of the $g$th generation.

For the real optimization subroutine of Algorithm 1, roulette wheel selection is used to determine parents (line 16) and uniform crossover approach defined below is adopted (line 17)

$$\text{offspring}_1 = \alpha_c \odot \text{parent}_1 + (\mathbf{1} - \alpha_c) \odot \text{parent}_2$$
$$\text{offspring}_2 = (\mathbf{1} - \alpha_c) \odot \text{parent}_1 + \alpha_c \odot \text{parent}_2$$

where $\alpha_c$ is a uniform random number that varies in between $(-\gamma_c, 1+\gamma_c)$ with $0 < \gamma_c \ll 1$ being a positive value improving the exploration capability yielding offsprings that are slightly different than ordinary mixture (i.e., $\gamma_c = 0$) of their parents. For the $i$th entry of the $\mathbf{w}$ vector, which contains the real parameters, mutation operator taking place in line 19 of Algorithm 1 implements $\mathbf{w}_i[g] \leftarrow \mathbf{w}_i[g] + \sigma_c r_{ci}$ if $r_{ci} > \mu_c$ with $\mu_c$ being a small positive threshold, $0 < \sigma_c < 1$ being a constant design parameter and $r_{ci}$ being a normally distributed random variable. The evaluation, sorting and selection procedure is the same as we discussed for binary case, and the result in $g$th generation is $\mathbf{w}^\star$, which is to be used by the binary optimization subroutine in $(g+1)$th generation.

The algorithm defined above generates well performing binary mask population for the current best parameter vector ($\mathbf{w}^\star[g]$) and a well performing parameter vector population for the best mask vector ($\mathbf{m}^\star[g]$) until either a predefined number of generations ($g_{\max}$) is achieved or $J \leq \text{tolerance}$ is reached. Since the optimization problem given in (9) involves finding real vector $\mathbf{w} \in \mathbb{R}^{1\times(n+1)^2}$ and the Binary vector $\mathbf{m} \in \mathbb{R}^{1\times L(n+1)^2}$ classical optimization schemes does not help working in such a mixed search space. This was our motivation to use genetic optimization as given under the Algorithm 1's structure.

As seen from the given discussion, the optimization process considering only the binary or only the real case adopts an evolutionary optimization process to distinguish well performing solutions. Our paper integrates them in a way that a suitable mask vector accompanied by a parameter vector can be reached to fit multiple FRDs under different switchings, the outcomes of which will be discussed next.

Last but not the least, one can choose to enhance crossover and mutation strategies to obtain a better population that displays better convergence characteristics. The goal in this study is to demonstrate that the proposed scheme generates useful results under standard GA settings.

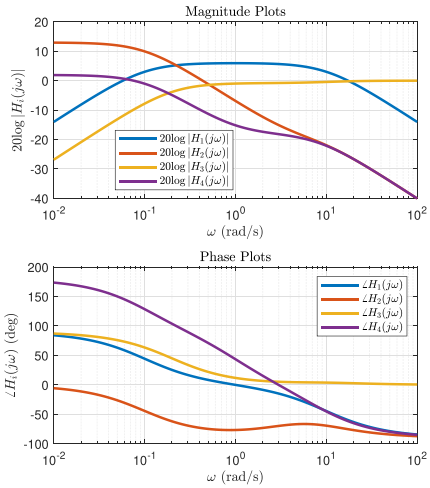


Fig. 2. Bode magnitude and phase plots of the four systems that are represented in a single masked model.

## IV. Illustrative Example

To provide an illustrative example, four systems characterized by equations (10)–(13) have been taken into account. The magnitude and phase data for each system were computed over the frequency range $\omega \in [0.01, 100]$ rad/s, yielding the frequency domain specifics depicted in Fig. 2. We considered a logarithmically spaced $P_l = 100$ points for $l = 1, 2, 3, 4$. The figure demonstrates that these four systems exhibit distinct behaviors, encompassing characteristics of low-pass, high-pass, and band-pass filters with gain values below or above 0 dB level at different frequencies. Such characteristics are frequently encountered in the industrial practice

$$H_1(s) = \frac{20s}{(s+0.1)(s+10)} \tag{10}$$

$$H_2(s) = \frac{s+4}{(s+0.1)(s+9)} \tag{11}$$

$$H_3(s) = \frac{s(s+10)}{(s+0.2)(s+11)} \tag{12}$$

$$H_4(s) = \frac{s-1}{(s+0.1)(s+8)}. \tag{13}$$

The poles of the four systems that generate the target spectral data seem in two clusters, which are around 0.1 and 10. This is deliberate to see the spectral changes occur within the chosen frequency spectrum. If the FRD generating models have spectral pictures, the meaningful change of which take place at mutually exclusive spectral regions, then packing those into a single mother model would require larger $n$.

One may choose the order of the mother model by considering the orders of the given TFs in (10)–(13). The sum of the orders is eight, therefore, a mother model of order less than this value could be considered as an economic representation. However, in a practical setting, one may have spectral information rather than the TFs and in those cases, the value of $n$ can be set by trial and error, which may introduce some redundancy in the final mother model.

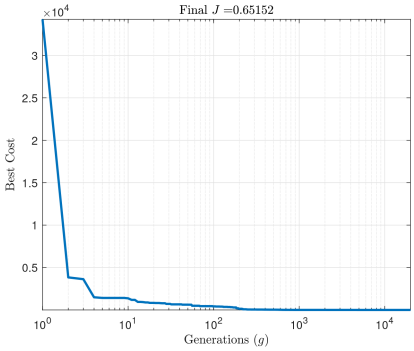


Fig. 3. Evolution of the best cost over the generations, where $g_{\max} = 20\,000$.

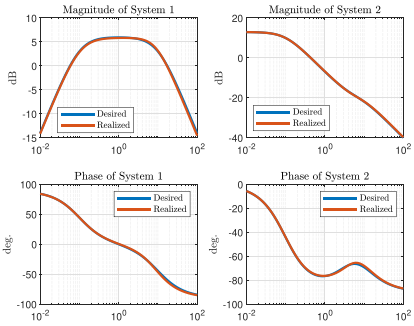


Fig. 4. Obtained results for the first and the second subsystems after applying the synthesized masks to the mother state space model in (14) and (15).

Optimization was applied to fine-tune the parameters of a fifth-order state-space system ($n = 5$) as per the proposed methodology. The real parameters are constrained to the interval $[-10, 10]$. The evolution of the cost function for the composite problem involving these four systems is illustrated in Fig. 3. The optimal cost value resulting from this optimization was noted as $J = 0.65152$ and the mother system given in (14) and (15) is obtained after $g_{\max} = 20\,000$ generations taking almost an hour on an Intel I7 computer having 2.4 GHz CPU. All simulations are conducted in MATLAB environment

$$\dot{\mathbf{x}} = \begin{pmatrix} -8.24 & 2.73 & -1.04 & -1.38 & 2.00 \\ -2.29 & -7.85 & 0.09 & 0.16 & 0.25 \\ -0.70 & 0.17 & -4.98 & 0.40 & -0.63 \\ 0.69 & -0.12 & -0.60 & -1.37 & 0.11 \\ 0.80 & 0.16 & -1.37 & -0.02 & -0.29 \end{pmatrix} \mathbf{x} + \begin{pmatrix} -1.81 \\ 9.02 \\ 2.12 \\ 2.13 \\ -2.14 \end{pmatrix} u \tag{14}$$

$$y = \begin{pmatrix} -0.53 & 2.08 & -0.55 & 2.29 & -0.57 \end{pmatrix} \mathbf{x} + 1.07u. \tag{15}$$

Fig. 4 displays the alignment of the obtained subsystems with the targeted spectral data for $l = 1, 2$. For the last two subsystems, i.e., $l = 3, 4$, the results are shown in Fig. 5. This visualization confirms that both magnitude and phase curves can

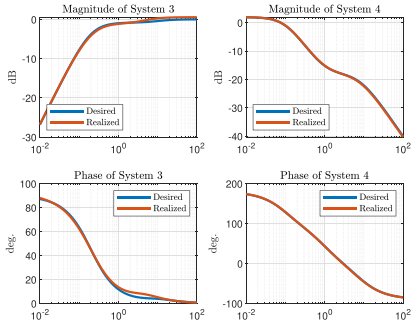


Fig. 5. Obtained results for the third and the fourth subsystems after applying the synthesized masks to the mother state space model in (14) and (15).

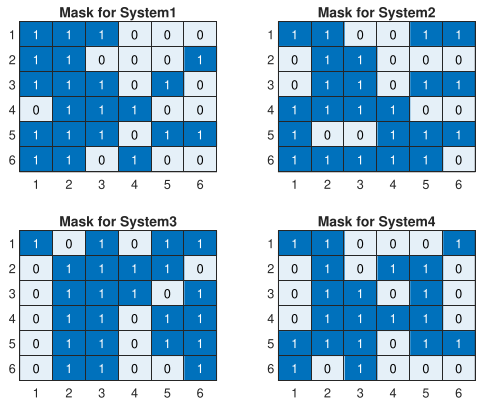


Fig. 6. Obtained masks that are applied to the mother system's $\left(\begin{smallmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{smallmatrix}\right)$ quadruple given in (14) and (15).

be accurately approximated. One might say that some frequency values are not well approximated. First of all, it is necessary to emphasize the success of the proposed algorithm in expressing four different systems with a 5th order masked model. If each system were implemented separately, this problem could be solved with an 8th order model. Here, dimensional simplification should not be overlooked when representing spectral information with a smaller dimensional system. In a more practical scenario, one might capture a resonant peak available in the Bode magnitude plot. In such cases, use of frequency weighted cost function may be considered as a remedy, which is not implemented here.

Fig. 6 depicts the masks utilized to create these subsystems. It is evident that a variety of masks were generated, enabling the production of subsystems derived from the mother system given in (14) and (15). These outcomes validate the effectiveness of the suggested approach.

Selecting the relative importance parameters $\lambda_m$, $\lambda_p$, and $\lambda_s$ within the study holds significant importance. The third term in (6) is zero for stable masked models. If the stability is not a critical issue, the designer may choose $\lambda_s = 0$. In our study, we would like to observe similarity in the time domain responses, then we set $\lambda_s = 10\,000$, which tells that we do not accept unstable approximations to the systems in (10)–(13). Confining ourselves to the stable region, a $\lambda_p$ value that is excessively larger

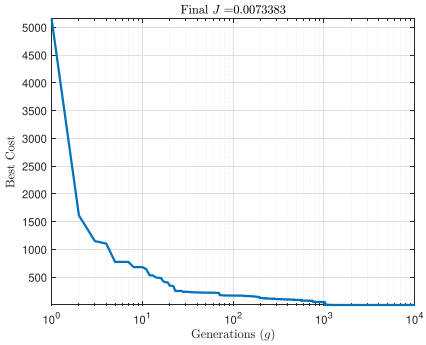


Fig. 7. Evolution of the best cost over the generations where $g_{\max} = 10\,000$.

than $\lambda_m$ would predominantly align with phase information, whereas a value that is excessively small would predominantly align with magnitude information. Through our discussed example, it was noted that a suitable $\lambda_p$ value, specifically $\lambda_p = 0.1$ and $\lambda_m = 1$, was identified, where the alteration range of phase information in degrees did not excessively influence the cost function. Consequently, these values were implemented in the shown results.

In the literature, subspace identification methods are known to solve problems if there is only one FRD or time domain data, [32], [33], [34], which discuss the theoretical aspects of the subspace identification scheme. To the best of our knowledge, no other alternative provides a means to obtain multiple subsystems after applying a proper mask to a mother system to fit the desired FRD.

In the presented simulation work, we chose $n_b = n_c = 20$. The population size is an important parameter, if it is large, the optimization takes a long time, alternatively, for small population sizes, the algorithm is likely to get trapped to a local minima. In addition to this, we set $\mu_b = \mu_c = 0.02$, which suggests 2% change in each mutation operation, and $\sigma_c = 0.1$, $\gamma_c = 0.2$ are chosen.

## V. Comparison With Subspace Identification Method

In order to compare our approach with subspace identification method, which was applied successfully to industrial problems [35], [36], we chose the system in (10) and sought for four models ($L = 4$) that have the same Bode magnitude and phase data. The reason behind the choice of $L$ is to see possible different solutions generating the same spectral information. During the optimization process, we chose $n_b = n_c = 50$ individuals and $n = 3$, which is the order of the masked models to be developed. All other optimization parameters were kept the same. In Fig. 7, the evolution of the cost ($J$) is shown. Since the four FRDs contain the same information, the cost decreases to a very low value after $g_{\max} = 10\,000$ generations. The optimization process yields the mother model given in (16) and (17), and the associated masks are shown in Fig. 8.

$$\dot{\mathbf{x}} = \begin{pmatrix} -1.30 & -1.85 & 0.22 \\ -6.58 & -0.37 & 0.49 \\ -4.48 & -2.22 & -10.00 \end{pmatrix} \mathbf{x} + \begin{pmatrix} -2.69 \\ -0.90 \\ -5.41 \end{pmatrix} u \quad (16)$$

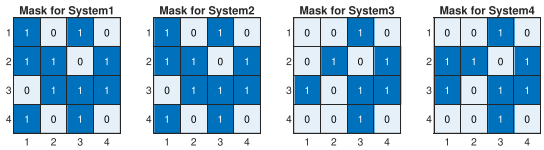


Fig. 8. Obtained masks that are applied to the mother system's $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$ quadruple given in (16) and (17).

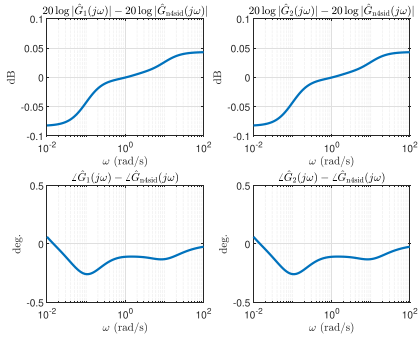


Fig. 9. Difference between the magnitude and phase curves between the first two subsystems and that generated by subspace identification algorithm. One should note that the first two masks are identical and the masked subsystems are the same.

$$y = \begin{pmatrix} -5.21 & 5.48 & -3.68 \end{pmatrix} \mathbf{x}. \quad (17)$$

According to Fig. 8, first two systems have the same mask thereby leading to the same system. In other words, we see three different systems generated by our approach and we will compare the Bode magnitude and phase curves obtained via subspace identification method. Let $\hat{G}_l(j\omega)$ be the TF of the $l$th subsystem generated by the proposed algorithm and let $G_{\mathrm{n4sid}}(j\omega)$ be the TF obtained from the subspace identification algorithm, which generates the following state space model:

$$\dot{\mathbf{x}}_s = \begin{pmatrix} -0.87 & 15.26 & -14.29 \\ 0.46 & -9.23 & -10.82 \\ 0.00 & 0.00 & -109.75 \end{pmatrix} \mathbf{x}_s + \begin{pmatrix} 3.77 \\ -2.32 \\ 0.00 \end{pmatrix} u \quad (18)$$

$$y_s = \begin{pmatrix} 2.28 & -4.93 & 0.99 \end{pmatrix} \mathbf{x}_s. \quad (19)$$

When the Bode magnitude and phase data are plotted for (10), for our approach and for subspace identification method, the obtained curves turn out to be totally indistinguishable. Therefore, we show the differences in Figs. 9 and 10, where we see the difference in $\pm 0.1$ dB interval in magnitudes, $\pm 0.3$ degrees in phases.

According to the presented comparison, we conclude that the proposed algorithm is able to generate a model, which can compete with subspace identification method, further, the proposed scheme is able to store different systems' spectral content within a single masked model, which cannot be achieved via subspace identification method. The disadvantage of the proposed approach is the time required to reach an acceptable $J$ value. This is mainly because the proposed technique is based on a search based optimization, which gradually improves the performance.

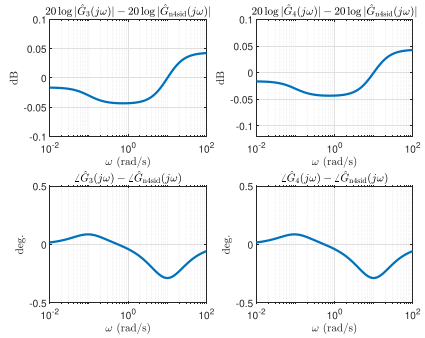


Fig. 10. Difference between the magnitude and phase curves between the last two subsystems and that generated by subspace identification algorithm.

Based on the described algorithm and the studied numerical examples, the contribution of the current work is to define an algorithm that encodes multiple spectral information into a single switchable mother model. The work differentiates from the existing body of literature in terms of mask update which is carried out simultaneously with the real parameter update.

## VI. Conclusion

This article describes an effective switching and optimization approach to store multiple FRDs within a mother state space model. The problem is cast into an optimization problem that has binary and continuous variables and an effective solution is proposed. The proposed optimization strategy finds out the best performing subnetwork. The proposed technique seeks for the optimum value of a cost function, which is defined over all subsystems obtained via available masks. The evolutionary process finds out best matching numerical content and associated mask values. The proposed approach is justified via a simulations covering spectrally dissimilar system behaviors. The current study advances the subject area toward multifunctional state space models that may contain dissimilar spectral information in an optimized dynamical structure accompanied by an optimized mask.

## References


[1] H. Garnier, M. Mensler, and A. Richard, "Continuous-time model identification from sampled data: Implementation issues and performance evaluation," *Int. J. Control*, vol. 76, no. 13, pp. 1337–1357, Jan. 2003.

[2] H. Garnier, "Direct continuous-time approaches to system identification. overview and benefits for practical applications," *Eur. J. Control*, vol. 24, pp. 50–62, 2015.

[3] L. Ljung, "Experiments with identification of continuous time models," *IFAC Proc.*, vol. 42, no. 10, pp. 1175–1180, 2009.

[4] P. Young and A. Jakeman, "Refined instrumental variable methods of recursive time-series analysis part III. extensions," *Int. J. Control*, vol. 31, no. 4, pp. 741–764, Apr. 1980.

[5] A. A. Özdemir and S. Gümüşsoy, "Transfer function estimation in system identification toolbox via vector fitting," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6232–6237, Jul. 2017.

[6] Z. Drmač, S. Gügercin, and C. Beattie, "Quadrature-based vector fitting for discretized H2 approximation," *SIAM J. Sci. Comput.*, vol. 37, no. 2, pp. A625–652, Jan. 2015.

[7] H. Vold, J. Crowley, and G. T. Rocklin, "New ways of estimating frequency response functions," *Sound Vib.*, vol. 18, pp. 34–38, Nov. 1984.

[8] F. Zhu and X. Wang, "Recursive identification of state space systems with coloured process noise and measurement noise," *Int. J. Comput. Appl. Technol.*, vol. 67, no. 2-3, pp. 129–140, 2021.

[9] X. Liu and X. Yang, "Identification of nonlinear state-space systems with skewed measurement noises," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 69, no. 11, pp. 4654–4662, Nov. 2022.

[10] X. Liu and X. Yang, "Variational identification of linearly parameterized nonlinear state-space systems," *IEEE Trans. Control Syst. Technol.*, vol. 31, no. 4, pp. 1844–1854, Jul. 2023.

[11] Y. Gu, W. Dai, Q. Zhu, and H. Nouri, "Hierarchical multi-innovation stochastic gradient identification algorithm for estimating a bilinear state-space model with moving average noise," *J. Comput. Appl. Math.*, vol. 420, 2023, Art. no. 114794.

[12] Y. Bao, J. M. Velni, and M. Shahbakhti, "Epistemic uncertainty quantification in state-space LPV model identification using Bayesian neural networks," *IEEE Contr. Syst. Lett.*, vol. 5, no. 2, pp. 719–724, Apr. 2021.

[13] Y. Bao and J. M. Velni, "Data-driven linear parameter-varying model identification using transfer learning," *IEEE Contr. Syst. Lett.*, vol. 5, no. 5, pp. 1579–1584, Nov. 2021.

[14] M. Forgione and D. Piga, "Continuous-time system identification with neural networks: Model structures and fitting criteria," *Eur. J. Control*, vol. 59, pp. 69–81, May 2021.

[15] M. Mejari, B. Mavkov, M. Forgione, and D. Piga, "Direct identification of continuous-time LPV state-space models via an integral architecture," *Automatica*, vol. 142, Aug. 2022, Art. no. 110407.

[16] C. M. Pappalardo, F. Califano, S. I. Lok, and D. Guida, "A systematic computational and experimental study of the principal data-driven identification procedures. Part I: Analytical methods and computational algorithms," *J. Appl. Comput. Mechanics*, vol. 9, no. 2, pp. 529–549, Apr. 2023.

[17] H. Bagua, A. Hafaifa, A. Iratni, and M. Guemana, "Model variables identification of a gas turbine using a subspace approach based on input/output data measurements," *Control Theory Technol.*, vol. 19, no. 2, pp. 183–196, May 2021.

[18] M. Mondal and J. A. Ramos, "Estimation of humidity sensor equivalent circuit parameters utilizing frequency response measurements combined with subspace identification and similarity transformations," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 2006909.

[19] C. Huang, "A combined invariant-subspace and subspace identification method for continuous-time state-space models using slowly sampled multi-sine-wave data," *Automatica*, vol. 140, Jun. 2022, Art. no. 110261.

[20] J. Yu and S. J. Qin, "Latent state space modeling of high-dimensional time series with a canonical correlation objective," *IEEE Contr. Syst. Lett.*, vol. 6, pp. 3469–3474, 2022.

[21] L. Xiao, H. Ogai, W. Jianhong, and R. A. Ramirez Mendoza, "Conditional maximum likelihood identification for state space system," *Mechatronic Syst. Control*, vol. 49, no. 1, pp. 1–8, 2021.

[22] T. Zong, J. Li, and G. Lu, "Identification of Hammerstein-Wiener systems with state-space subsystems based on the improved PSO and GSA algorithm," *Circuits Syst. Signal Process.*, vol. 42, no. 5, pp. 2755–2781, May 2023.

[23] M. Yu, G. Guo, and J. Liu, "Closed-loop delta-operator-based subspace identification for continuous-time systems utilising the parity space," *Int. J. Syst. Sci.*, vol. 52, no. 15, pp. 3323–3334, Nov. 2021.

[24] M. F. Shakib, A. Y. Pogromsky, A. Pavlov, and N. van de Wouw, "Computationally efficient identification of continuous-time Lur'e-type systems with stability guarantees," *Automatica*, vol. 136, Feb. 2022, Art. no. 110012.

[25] H. Mania, M. I. Jordan, and B. Recht, "Active learning for nonlinear system identification with guarantees," *J. Mach. Learn. Res.*, vol. 23, pp. 1–30, 2022.

[26] C. M. Pappalardo, F. Califano, S. I. Lok, and D. Guida, "A systematic computational and experimental study of the principal data-driven identification procedures. Part II numerical analysis and experimental testing," *J. Appl. Comput. Mechanics*, vol. 9, no. 2, pp. 550–589, Apr. 2023.

[27] M. O. Efe and H. Özbay, "Proper orthogonal decomposition for reduced order modeling: 2D heat flow," in *Proc. IEEE Int. Conf. Control Appl.*, 2003, pp. 1273–1278.

[28] D. E. Seborg, T. F. Edgar, and D. A. Mellichamp, *Process Dynamics and Control*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2004.

[29] S. Palani, *Discrete Time Systems and Signal Processing*. Berlin, Germany: Springer, 2023.

[30] S. N. Sivanandam and S. N. Deepa, *Introduction to Genetic Algorithms*. Berlin, Germany: Springer, 2008.

[31] A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms: A tutorial," *Rel. Eng. Syst. Saf.*, vol. 91, no. 9, pp. 992–1007, 2006.
[32] P. van Overschee and B. De Moor, *Subspace Identification of Linear Systems: Theory, Implementation, Applications*. Berlin, Germany: Springer, 1996.
[33] M. Verhaegen, "Identification of the deterministic part of MIMO state space models," *Automatica*, vol. 30, pp. 61–74, 1994.
[34] T. McKelvey, H. Akcay, and L. Ljung, "Subspace-based multivariable system identification from frequency response data," *IEEE Trans. Autom. Control*, vol. 41, no. 7, pp. 960–979, Jul. 1996.
[35] Z. Masoumi, B. Moaveni, and S. M. M. Gazafrudi, "Signal-model-based fault diagnosis in windings of synchronous generator," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 2942–2951, Mar. 2023.
[36] Y. C. Jiang, S. Yin, and O. Kaynak, "Optimized design of parity relation-based residual generator for fault detection: Data-driven approaches," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 1449–1458, Feb. 2021.

**Mehmet Önder Efe** received the Ph.D. degree from Electrical Electronics Engineering Department, Bogazici University, Istanbul, Türkiye.

Having spent a year with Carnegie Mellon and another with The Ohio State University, he worked at several private universities in Türkiye in between 2003 and 2013. Since 2013, he has been with the Department of Computer Engineering, Hacettepe University. His research interests include control systems, neural networks, and autonomous vehicles.

**Burak Kürkçü** received the B.Sc. degree from Istanbul Technical University, Istanbul, Türkiye, in 2010, and the M.Sc. and Ph.D. degrees from the Department of Electrical and Electronics Engineering, TOBB University of Economics and Technology, Ankara, Türkiye, in 2015 and 2019, respectively.

He is currently a Research Scholar with the Department of Mechanical Engineering, University of California, Berkeley, Berkeley, CA USA. He is also an Assistant Professor with the Department of Computer Engineering at Hacettepe University, Ankara. Before joining the Computer Engineering Department, Hacettepe University, he was a Control System Design Engineer with ASELSAN, Inc., for over a decade. He is the author or coauthor of several publications focusing on the robust control theory and applications.

Dr. Kürkçü was the recipient of the IEEE Turkey Ph.D. Thesis Award in 2020. He is an Associate Editor for *Transactions of the Institute of Measurement and Control, Measurement and Control*, and the Guest Editor of the *Turkish Journal of Electrical Engineering and Computer Science*.

**Coşku Kasnakoğlu** received the B.S. degrees from the Department of Electrical and Electronics Engineering and the Department of Computer Engineering, Middle East Technical University, Ankara, Türkiye, in 2000, and the M.S. and Ph.D. degree from the Department of Electrical and Computer Engineering, Ohio State University, Columbus, OH, USA, in 2003 and 2007, respectively.

He is currently a Professor with the Department of Electrical and Electronics Engineering, TOBB University of Economics and Technology, Ankara.

**Zaharuddin Mohamed** received the B.Eng. degree from the Department of Electronics and Electrical Engineering from Universiti Kebangsaan Malaysia, Malaysia, in 1993, and the M.Sc. and Ph.D. degrees in control systems engineering from the University of Sheffield, Sheffield, U.K., in 1995 and 2003, respectively.

He is currently with the Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Malaysia. His research interest includes control of underactuated systems, robotics, and mechatronics.

**Zhijie Liu** (Member, IEEE) received the B.Sc. degree in automation from the China University of Mining and Technology Beijing, Beijing, China, in 2014, and the Ph.D. degree in control science and engineering from Beihang University, Beijing, in 2019.

In 2017, he was an RA with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA, for 12 months. He is currently a Professor with the School of Intelligence Science and Technology, University of Science and Technology Beijing, Beijing. His research interests include adaptive control, modeling, and vibration control for flexible structures, and distributed parameter system.