Parsimonious Identification of Continuous-Time Systems: A Block-Coordinate Descent Approach ⋆

Rodrigo A. González * **Cristian R. Rojas** ** **Siqi Pan** ***
James S. Welsh ***

* Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
** Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden
*** School of Engineering, University of Newcastle, Callaghan, NSW, Australia

Abstract: The identification of electrical, mechanical, and biological systems using data can benefit greatly from prior knowledge extracted from physical modeling. Parametric continuous-time identification methods can naturally incorporate this knowledge, which leads to interpretable and parsimonious models. However, some applications lead to model structures that lack parsimonious descriptions using unfactored transfer functions, which are commonly used in standard direct approaches for continuous-time system identification. In this paper we characterize this parsimony problem, and develop a block-coordinate descent algorithm that delivers parsimonious models by sequentially estimating an additive decomposition of the transfer function of interest. Numerical simulations show the efficacy of the proposed approach.

Keywords: Continuous-time system identification; block-coordinate descent; parsimony.

1. INTRODUCTION

Continuous-time system identification methods (Garnier and Wang, 2008) are popular and successful in a wide range of practical applications thanks to several advantages they enjoy compared to the discrete-time algorithms (Söderström and Stoica, 1989). One of these traits is that some continuous-time methods allow the direct incorporation of the *a priori* knowledge of the relative degree of the physical systems they model. This leads to more parsimonious representations, which means that simpler models (in terms of number of parameters) can be used to accurately describe the phenomenon at hand.

With regards to model flexibility, most linear continuous-time identification methods parameterize the model structure as an unfactored transfer function with a user-defined number of poles and zeros. This is the case for the Prediction Error Method for continuous-time systems (PEM), as well as the Simplified Refined Instrumental Variable method for Continuous-time systems (SRIVC, Young and Jakeman (1980)) and the Least-Squares State-Variable Filter method (LSSVF, Young (1965)). However, some practical applications related to, e.g., motion systems and vibration analysis, consider systems that are more easily interpreted as a sum of transfer functions with distinct denominators, typically corresponding to different resonant modes. The SRIVC method cannot handle the estimation

of such additive continuous-time systems, since such model structure does not yield a pseudolinear regression suitable for constructing the filtered instrument and regressor vectors (Garnier et al., 2007). In addition, it is known that numerical conditioning issues may arise when estimating high-order or highly-resonant systems, which are typically parameterized as the sum of second-order continuous-time systems (Gilson et al., 2017).

In this paper we propose an algorithm that delivers parsimonious models for additive continuous-time systems. This methods performs a block coordinate descent using SRIVC as a tool for decreasing the output error cost at each iteration. More precisely,

C1 We obtain an explicit condition under which it is preferable to estimate additive models instead of unfactored transfer functions, which is relevant when deciding what model structure should be considered for continuous-time system identification;

C2 We prove the global convergence (for large sample size) of a coordinate-block descent algorithm for the identification of additive continuous-time systems;

C3 We provide closed-form expressions of the SRIVC iterations that, at convergence for finite sample size and under mild conditions, are proven to deliver a critical point of the cost function being minimized at each block-coordinate descent step.

The paper is structured as follows. In Section 2, we introduce the problem setup and discuss different model parameterizations. Section 3 introduces the parsimony problem for continuous-time models, while the block-

coordinate descent method that solves this problem is presented in Section 4. A numerical study can be found in Section 5, and the paper is concluded in Section 6. Proofs of the main results can be found in the Appendix.

2. SYSTEM AND MODEL SETUP

Consider the single-input single-output (SISO), linear and time-invariant, continuous-time system

$$x(t) = \frac{B^*(p)}{A^*(p)} u(t), \tag{1}$$

where p is the Heaviside operator, and $u(t)$ is the input signal. The numerator and denominator polynomials $B^*(p)$ and $A^*(p)$ are assumed coprime and given by

$$A^*(p) = a_n^* p^n + a_{n-1}^* p^{n-1} + \cdots + a_1^* p + 1,$$
$$B^*(p) = b_m^* p^m + b_{m-1}^* p^{m-1} + \cdots + b_1^* p + b_0^*,$$

with $a_n^* \neq 0$ and $n \geq m$. The polynomials $A^*(p)$ and $B^*(p)$ are jointly described by the parameter vector

$$\boldsymbol{\theta}^* = [a_1^*, \ a_2^*, \ \ldots, \ a_n^*, \ b_0^*, \ b_1^*, \ \ldots, \ b_m^*]^\top. \tag{2}$$

The system $G^*(p) = B^*(p)/A^*(p)$ can also be described in its modal (or *additive*) form. This modal form, which describes the system as a finite sum of transfer functions (Pan et al., 2021), leads to the alternative description

$$x(t) = \sum_{i=1}^{K} G_i^*(p) u(t), \tag{3}$$

where $G_i^*(p) = B_i^*(p)/A_i^*(p)$, and the transfer function polynomials $A_i^*(p)$ and $B_i^*(p)$ have degrees n_i and m_i, respectively ($m_i \leq n_i$). We assume without loss of generality that the $A_i^*(p)$ polynomials are anti-monic (i.e., their constant coefficient is fixed to 1), and they are jointly coprime. For this representation we write $\boldsymbol{\theta}_i^*$ as the parameter vector that describes the transfer function $G_i^*(p)$, similarly to (2). In addition, for each separate submodel to be identifiable, we assume that at most one subsystem $G_i^*(p)$ is biproper.

Given the system in (3), a noisy output measurement is retrieved every h[s]. That is,

$$y(kh) = x(kh) + v(kh),$$

where $v(kh)$ is assumed to be a zero-mean stationary random process of variance σ^2 that is uncorrelated with the input. We assume that the input has a zero-order hold (ZOH) intersample behavior, although our results can also be extended to arbitrary inputs (González et al., 2020).

This paper studies how to determine a model for the additive decomposition of $G^*(p)$ in (3), with known model structure, based on N input and output data samples $\{u(kh), y(kh)\}_{k=1}^N$. In addition, we are interested in comparing both standard (1) and additive (3) forms in terms of their parsimony when implementing continuous-time system identification methods. As the models depend on the parameter vector $\boldsymbol{\theta}$ that is being estimated, we write them as $G(p, \boldsymbol{\theta})$, or $G(p)$, if the dependence in $\boldsymbol{\theta}$ is obvious.

3. PARSIMONY IN CONTINUOUS-TIME SYSTEM IDENTIFICATION

Direct continuous-time methods such as the SRIVC estimator may suffer from a lack of parsimony when we identify the sum of transfer functions of particular relative degrees. The following proposition, which constitutes Contribution C1 of this paper, indicates the number of additional

parameters that are being estimated if the user decides to estimate (3) with a model structure $G(p) = B(p)/A(p)$ of relative degree r instead of estimating the parameters of each transfer function $B_i^*(p)/A_i^*(p)$ separately.

Proposition 1. Consider the system in (3), and the model structure $G(p) = \sum_{i=1}^K B_i(p)/A_i(p)$. If one instead decides to use the model structure $G(p) = B(p)/A(p)$ for identification, with minimal relative degree r that contains the true system, then the latter model structure incurs in a lack of parsimony if and only if

$$\sum_{i=1}^{K} r_i - r > K - 1, \tag{4}$$

where $r_i = n_i - m_i$ is the relative degree of $B_i(p)/A_i(p)$. The excess in (4), i.e., the difference between the left and right hand sides, is the number of additional parameters that the model structure $G(p) = B(p)/A(p)$ considers.

Proof. The number of parameters to be estimated in each model $B_i(p)/A_i(p)$ is $n_i + m_i + 1$, which leads to the need of estimating $\sum_{i=1}^K (n_i + m_i) + K$ parameters if one considers the model structure $G(p) = \sum_{i=1}^K B_i(p)/A_i(p)$. On the other hand, the model structure $G(p) = B(p)/A(p)$ requires estimating $2\sum_{i=1}^K n_i - r + 1$ parameters. Subtracting both of these expressions leads to an excess of parameters given by $\sum_{i=1}^K r_i - r - K + 1$. If such quantity is greater than zero, then we reach the condition in (4). □

A consequence of this result is that if $r = 0$ or $r = 1$, then the model structure $G(p) = B(p)/A(p)$ suffers from a lack of parsimony if there exists a transfer function $G_i^*(p)$ in the expansion (3) with relative degree greater than one.

Example 3.1. Consider the system

$$G^*(p) = \frac{3}{0.25p^2 + 0.25p + 1} + \frac{1}{0.0025p^2 + 0.01p + 1}. \tag{5}$$

This system corresponds to a truncated modal description of a flexible structure, such as a piezoelectric laminate beam (Moheimani et al., 2003). Only 6 parameters must be estimated if the following model structure is used:

$$G(p) = \frac{b_{01}}{a_{21}p^2 + a_{11}p + 1} + \frac{b_{02}}{a_{22}p^2 + a_{12}p + 1}. \tag{6}$$

On the other hand, if this modal decomposition is not taken into account and one decides to estimate the model

$$G(p) = \frac{b_2 p^2 + b_1 p + b_0}{a_4 p^4 + a_3 p^3 + a_2 p^2 + a_1 p + 1} \tag{7}$$

with no constraints on the parameter values, then 7 parameters must be estimated. This model structure leads to a lack of parsimony compared to (6).

4. BLOCK-COORDINATE DESCENT METHOD FOR CONTINUOUS-TIME SYSTEM IDENTIFICATION

In this section we present a method that solves the parsimony problem described in Section 3 for the identification of linear continuous-time systems. The goal is to estimate the parameters of additive models of the form (3) by solving the following minimization problem:

$$\min_{\substack{\boldsymbol{\theta}_i \in \Omega_i, \\ i=1,\ldots,K}} \frac{1}{N} \sum_{k=1}^{N} \left[y(kh) - \sum_{i=1}^{K} G_i(p, \boldsymbol{\theta}_i) u(kh) \right]^2, \tag{8}$$

with $\Omega_i \subset \mathbb{R}^{n_i+m_i+1}$ being a compact set where the parameters of the ith subsystem are assumed to lie. Note that to solve the optimization problem in (8) one cannot directly apply refined instrumental variable methods (i.e., the SRIVC method (Young and Jakeman, 1980)), since the denominator polynomials of each submodel $G_i(p, \boldsymbol{\theta}_i)$ are distinct. Instead, we propose a *block-coordinate descent* algorithm, in which the cost function is iteratively minimized with respect to $\boldsymbol{\theta}_i$ while leaving the other decision variables fixed. To this end, we define the total cost function

$$V_N(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_K):=\frac{1}{N}\sum_{k=1}^{N}\left[y(kh)-\sum_{i=1}^{K}G_i(p,\boldsymbol{\theta}_i)u(kh)\right]^2,$$

where $\boldsymbol{\theta}_i \in \Omega_i$ for $i = 1, \ldots, K$. Algorithm 1 describes the general proposed procedure.

Algorithm 1 Block-coordinate descent algorithm

1: Input: initial parameter vector $\boldsymbol{\theta}_i^1$ for each i
2: **for** $l = 1, 2, \ldots$ **do**
3:     **for** $i = 1, \ldots, K$ **do**
4:        $\boldsymbol{\theta}_i^{l+1} \leftarrow \underset{\boldsymbol{\theta}_i \in \Omega_i}{\arg\min}\, V(\boldsymbol{\theta}_1^{l+1}, \ldots, \boldsymbol{\theta}_{i-1}^{l+1}, \boldsymbol{\theta}_i, \boldsymbol{\theta}_{i+1}^l, \ldots, \boldsymbol{\theta}_K^l)$
5:     **end for**
6: **end for**
7: Output: parameter vectors $\lim_{l\to\infty} \boldsymbol{\theta}_i^l$, $i = 1, 2, \ldots, K$.

Let $\mathbf{x}^l := [(\boldsymbol{\theta}_1^l)^\top, \ldots, (\boldsymbol{\theta}_K^l)^\top]^\top \in \Omega \subset \mathbb{R}^P$, with $\Omega = \prod_{i=1}^K \Omega_i$ being the parameter space, and $P = \sum_{i=1}^K (n_i + m_i) + K$ being the total number of parameters to estimate. Each iteration of Algorithm 1 (in l) can therefore be written as $\mathbf{x}^{l+1} = \mathcal{A}(\mathbf{x}^l)$, where $\mathcal{A}: \mathbb{R}^P \to \mathbb{R}^P$ is a mapping that can be described as a composition of functions:

$$\mathcal{A} = \mathbf{S} \circ \mathbf{C}^K \circ \mathbf{S} \circ \mathbf{C}^{K-1} \circ \cdots \circ \mathbf{S} \circ \mathbf{C}^1, \qquad (9)$$

where we denote the choice function $\mathbf{C}^i(\mathbf{x}) := (\mathbf{x}, i)$, the joint parameter vector $\mathbf{x} := [\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_K^\top]^\top$, and the optimization step $\mathbf{S}(\mathbf{x}, i) := (\boldsymbol{\theta}_1, \ldots, \overline{\boldsymbol{\theta}}_i, \ldots, \boldsymbol{\theta}_K)$, with $\overline{\boldsymbol{\theta}}_i = \arg\min_{\boldsymbol{\theta}_i \in \Omega_i} V_N(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_i, \ldots, \boldsymbol{\theta}_K)$. The following result concerns the convergence of Algorithm 1 to a stationary point of the cost (8), and constitutes Contribution C2.

Theorem 2. (Global convergence of Algorithm 1). Consider the set of fixed points $\Gamma_N = \{\mathbf{x} \in \Omega : \nabla V_N(\mathbf{x}) = \mathbf{0}\}$. For a sufficiently large N, the limit of any convergent subsequence of $\{\mathbf{x}^l\}$ obtained from the iterations $\mathbf{x}^{l+1} = \mathcal{A}(\mathbf{x}^l)$ almost surely belongs to Γ_N.

Proof. See Appendix A. □

A crucial part of Theorem 2 is the characterization of the solution set Γ_N. One can establish that, in addition to the uniform convergence of V_N as N tends to infinity, the gradient ∇V_N also converges uniformly to $\nabla \bar{V}$, where $\bar{V}$ is defined in Eq. (A.1). This implies that the set of stationary points of V_N, Γ_N, converges to those of $\bar{V}$, say, $\bar{\Gamma} = \{\mathbf{x} \in \Omega : \nabla \bar{V}(\mathbf{x}) = \mathbf{0}\}$, in the sense that for every $\epsilon > 0$ and almost every realization ξ, there exists a sample size $N(\xi, \epsilon) \in \mathbb{N}$ such that for all $N \geq N(\xi, \epsilon)$ and every $\mathbf{x} \in \Gamma_N$ there is a $\mathbf{y} \in \bar{\Gamma}$ such that $\|\mathbf{x} - \mathbf{y}\| < \epsilon$, and also for every $\mathbf{y} \in \bar{\Gamma}$ there is a $\mathbf{x} \in \Gamma_N$ such that $\|\mathbf{x} - \mathbf{y}\| < \epsilon$.

Unfortunately, $\bar{\Gamma}$ may contain not only the global minima of $\bar{V}$ (corresponding to all permutations $\boldsymbol{\theta}_i$, $\boldsymbol{\theta}_j$, $i \neq j$, that share the same model structure), but also other local minima and saddle points (Regalia, 1995, Section 7.1). Some

references give sufficient conditions under which $\bar{\Gamma}$ contains only the global minima. For example, if $\{u(kh)\}_{k\in\mathbb{Z}}$ is white noise and the model structure exactly contains the true system, then according to Lemma 4.1 of Söderström (1975b), every stationary point which is not a global minimizer of $\bar{f}$ must give pole-zero cancellations, thereby leading to the same minimal realization. This result, however, is not valid in general for arbitrary inputs.

Remark 1. For Theorem 2 to hold, it is only required for V_N to be decreasing at each step. This means that it is sufficient to find a parameter vector $\overline{\boldsymbol{\theta}}_i$ that reduces the cost V_N instead of minimizing it.

4.1 SRIVC for computing the descent step

The block-coordinate descent algorithm described in Algorithm 1 requires a way to compute $\mathbf{S}(\mathbf{x}, i)$ at each iteration, for each $i = 1, 2, \ldots, K$. That is, we need to compute

$$\boldsymbol{\theta}_i^{l+1} = \underset{\boldsymbol{\theta}_i \in \Omega_i}{\arg\min}\, \frac{1}{N}\sum_{k=1}^{N}\left[y(kh) - \sum_{j=1}^{i-1}G_j(p, \boldsymbol{\theta}_j^{l+1})u(kh)\right.$$
$$\left. - \sum_{j=i+1}^{K}G_j(p, \boldsymbol{\theta}_j^l)u(kh) - G_i(p, \boldsymbol{\theta}_i)u(kh)\right]^2 \qquad (10)$$

for $i = 1, 2, \ldots, K$. The key insight is that, for $\Omega_i = \mathbb{R}^{n_i+m_i+1}$ and fixed values of $\{\boldsymbol{\theta}_j^{l+1}\}_{j=1}^{i-1}$, $\{\boldsymbol{\theta}_j^l\}_{j=i+1}^{K}$, the optimization problem in (10) reduces to a nonlinear least-squares problem that can be solved via SRIVC iterations. Indeed, if we define the residual output of each submodel

$$\tilde{y}(kh) := y(kh) - \sum_{j=1}^{i-1}G_j(p, \boldsymbol{\theta}_j^{l+1})u(kh) - \sum_{j=i+1}^{K}G_j(p, \boldsymbol{\theta}_j^l)u(kh),$$

then $\boldsymbol{\theta}_i^{l+1}$ must satisfy the first-order optimality condition

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(kh, \boldsymbol{\theta}_i^{l+1})e(kh, \boldsymbol{\theta}_i^{l+1}) = \mathbf{0}, \qquad (11)$$

where the gradient and total residual are, respectively,

$$\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(kh, \boldsymbol{\theta}_i^{l+1}) = \left[\frac{-pB_i^{l+1}(p)}{[A_i^{l+1}(p)]^2}u(kh), \ldots, \frac{-p^nB_i^{l+1}(p)}{[A_i^{l+1}(p)]^2}u(kh),\right.$$
$$\left.\frac{1}{A_i^{l+1}(p)}u(kh), \ldots, \frac{p^m}{A_i^{l+1}(p)}u(kh)\right]^\top,$$
$$e(kh, \boldsymbol{\theta}_i^{l+1}) = \tilde{y}(kh) - \frac{B_i^{l+1}(p)}{A_i^{l+1}(p)}u(kh),$$

with B_i^{l+1} and A_i^{l+1} being the numerator and denominator polynomials of the ith submodel evaluated at $\boldsymbol{\theta}_i^{l+1}$.

Lemma 4.1, which constitutes Contribution C3 of this paper, provides the SRIVC iterations that are shown to deliver stationary points of the cost in (10) at convergence in iterations under mild conditions.

Lemma 4.1. (SRIVC iterations). For an initial model parameter estimate $\boldsymbol{\theta}_{i,0}^{l+1}$ and $s = 0, 1, 2, \ldots$, consider the following SRIVC iterations

$$\boldsymbol{\theta}_{i,s+1}^{l+1} = \left[\frac{1}{N}\sum_{k=1}^{N}\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(kh, \boldsymbol{\theta}_{i,s}^{l+1})\boldsymbol{\varphi}_{\mathrm{f}}^\top(kh, \boldsymbol{\theta}_{i,s}^{l+1})\right]^{-1} \qquad (12\mathrm{a})$$
$$\times \left[\frac{1}{N}\sum_{k=1}^{N}\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(kh, \boldsymbol{\theta}_{i,s}^{l+1})\tilde{y}_{\mathrm{f}}(kh, \boldsymbol{\theta}_{i,s}^{l+1})\right],$$

where the filtered regressor $\boldsymbol{\varphi}_\text{f}$ and filtered residual output $\tilde{y}_\text{f}$ are given by

$$\boldsymbol{\varphi}_\text{f}(kh, \boldsymbol{\theta}_{i,s}^{l+1}) = \left[\frac{-p}{A_{i,s}^{l+1}(p)} \tilde{y}(kh), \ldots, \frac{-p^n}{A_{i,s}^{l+1}(p)} \tilde{y}(kh), \right.$$
$$\left. \frac{1}{A_{i,s}^{l+1}(p)} u(kh), \ldots, \frac{p^m}{A_{i,s}^{l+1}(p)} u(kh) \right]^\top, \quad (13)$$

$$\tilde{y}_\text{f}(kh, \boldsymbol{\theta}_{i,s}^{l+1}) = \frac{1}{A_{i,s}^{l+1}(p)} \tilde{y}(kh). \quad (14)$$

If the matrix being inverted in (12a) is non-singular for all integers s large enough, then any converging point (when s tends to infinity), if they exist, satisfies the first-order optimality condition (11).

Proof. See Appendix B. $\qquad\square$

In practice, one can terminate the SRIVC procedure for each submodel whenever V_N has strictly decreased from its initial value. This termination rule is in agreement with Remark 1. Also, note that a decrease in the cost function requires to initialize the methods close to the global optimum, since the SRIVC method does not guarantee global convergence for finite-sample size. This can be done by applying the standard SRIVC estimator to find a model for (1), finding the partial fraction expansion, and then deleting the unwanted high-order numerator terms.

Remark 2. The non-singularity of the matrix in (12a) depends on the persistence of excitation of the input, as well as on the interpolation error when constructing the filtered output in the regressor vector. The generic non-singularity result in Theorem 1 of Pan et al. (2020) can be extended to the case in (12a) by including the model parameters of the other submodels as part of the genericity statement. Such proof, however, is outside of the scope of the current paper.

The method we propose for identifying additive continuous-time systems is detailed in Algorithm 2. Apart from what has been discussed, additional techniques can be fit to the algorithm to robustify it. These techniques include a) including a non-fixed step size (Söderström and Stoica, 1982) in the incremental form of these iterations, b) admitting unstable models by ad-hoc prefiltering (González et al., 2022), and c) introducing randomization when choosing the next submodel to be updated (i.e., the i index). The details of these extensions are left out of our exposition for simplicity only.

5. SIMULATIONS

We now verify the applicability of the proposed method through two numerical experiments.

5.1 Case Study 1: 4th order system

We consider $G^*(p)$ as in (5), which can also be written as

$$G^*(p) = \frac{0.2575p^2 + 0.28p + 4}{0.000625p^4 + 0.003125p^3 + 0.255p^2 + 0.26p + 1}. \quad (15)$$

Note that the denominator polynomial coincides with that of the Rao-Garnier system, which is a benchmark for linear continuous-time system identification (Rao and Garnier, 2002). We compare the proposed block-coordinate descent (BCD) method with the SRIVC estimator and the indirect

---

Algorithm 2: Block-coordinate descent method for additive continuous-time system identification

1: Input: $\{u(kh), y(kh)\}_{k=1}^N$, initialization $\mathbf{x}_1 = [\boldsymbol{\theta}_1^{1\top}, \ldots, \boldsymbol{\theta}_K^{1\top}]^\top$, tolerance factor ϵ, maximum number of $\mathcal{A}$-iterations M, and maximum number of SRIVC iterations M_s
2: $l \leftarrow 1$, $s \leftarrow 1$,
3: **while** $l \leq M$ **do**
4:     **for** $i = 1, \ldots, K$ **do**
5:         $\boldsymbol{\theta}_{i,1}^{l+1} \leftarrow \boldsymbol{\theta}_i^l$, $s \leftarrow 1$
6:         **while** $s \leq M_\text{s}$ **do**
7:             Compute $\boldsymbol{\theta}_{i,s+1}^{l+1}$ using SRIVC (12)
8:             **if** $V(\boldsymbol{\theta}_1^{l+1}, \ldots, \boldsymbol{\theta}_{i-1}^{l+1}, \boldsymbol{\theta}_{i,s+1}^{l+1}, \boldsymbol{\theta}_{i+1}^l, \ldots, \boldsymbol{\theta}_K^l) < V(\boldsymbol{\theta}_1^{l+1},$
            $\ldots, \boldsymbol{\theta}_{i-1}^{l+1}, \boldsymbol{\theta}_{i,1}^{l+1}, \boldsymbol{\theta}_{i+1}^l, \ldots, \boldsymbol{\theta}_K^l)$ **then**
9:                 $\boldsymbol{\theta}_i^{l+1} \leftarrow \boldsymbol{\theta}_{i,s+1}^{l+1}$, $s \leftarrow M_\text{s}$
10:             **end if**
11:             $s \leftarrow s + 1$
12:         **end while**
13:     **end for**
14:     $\mathbf{x}_{l+1} \leftarrow [\boldsymbol{\theta}_1^{l+1\top}, \ldots, \boldsymbol{\theta}_K^{l+1\top}]^\top$
15:     **if** $\dfrac{\|\mathbf{x}_{l+1} - \mathbf{x}_l\|_2}{\|\mathbf{x}_l\|_2} < \epsilon$ **then**
16:         $\hat{\mathbf{x}} \leftarrow \mathbf{x}_{l+1}$, $l \leftarrow M$
17:     **end if**
18:     $l \leftarrow l + 1$
19: **end while**
20: Output: $\hat{\mathbf{x}}$ and its associated models $\{\hat{G}_i(p)\}_{i=1}^K$.

---

approach using the SRIV estimator (Young, 1976) converted to continuous-time. The standard SRIVC estimator uses $n = 4$ and $m = 2$ as the model polynomial degrees, in agreement with the model structure (7). A Monte Carlo simulation is performed in order to test the fit and the mean square error (MSE) of the estimated parameters of all methods. The input is a Gaussian white noise of unit variance that is interpolated with a ZOH, and the measurement noise is also white and Gaussian, of unitary variance. Five hundred Monte Carlo runs are computed with $N = 10000$ and $h = 0.005$[s]. In each run, all methods are initialized with an additive model whose parameters deviate at most 10% from the true parameters.

Once the model is obtained from the data in each Monte Carlo run, the fit metric is computed by

$$\text{fit} = 100 \left(1 - \|\hat{\mathbf{x}}^i - \mathbf{x}\|_2 / \|\mathbf{x} - \bar{x}\mathbf{1}\|_2 \right),$$

where $\mathbf{x}$ is the noiseless output sequence, $\hat{\mathbf{x}}^i$ is the simulated output sequence of the ith estimated model, and $\bar{x}$ is the average value of $\{x(kh)\}_{k=1}^N$.

In Figure 1 we observe the benefit of obtaining parsimonious models by plotting the boxplot for the fit metric of all methods. We have also computed the MSE of each parameter of the equivalent sum model of the form (15) in Table 1. The indirect approach method (SRIV) provides models with relative degree equal to one almost always, hence inducing over-parametrization. The SRIVC estimator avoids this over-parametrization issue yet still cannot provide parsimonious models due to Proposition 1. The BCD method is the only one that provides the correct model structure, which leads to the best performance in the fit metric and MSEs of each estimated parameter.

Table 1. MSEs of the estimated parameters for each method, Case Study 1.

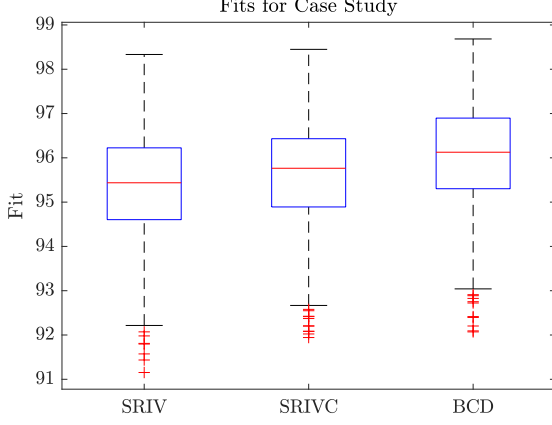| Method	Parameter True value	a_1^* 0.26	a_2^* 0.255	a_3^* $3.125 \cdot 10^{-3}$	a_4^* $6.25 \cdot 10^{-4}$	b_0^* 4	b_1^* 0.28	b_2^* 0.2575	b_3^* 0
SRIV	MSE	$3.51 \cdot 10^{-4}$	$8.53 \cdot 10^{-5}$	$1.74 \cdot 10^{-8}$	$5.48 \cdot 10^{-10}$	$2.44 \cdot 10^{-2}$	$6.03 \cdot 10^{-3}$	$1.07 \cdot 10^{-4}$	$1.55 \cdot 10^{-7}$
SRIVC	MSE	$3.52 \cdot 10^{-4}$	$7.59 \cdot 10^{-5}$	$1.71 \cdot 10^{-8}$	$4.28 \cdot 10^{-10}$	$2.44 \cdot 10^{-2}$	$4.56 \cdot 10^{-3}$	$1.05 \cdot 10^{-4}$	0
BCD	MSE	$3.41 \cdot 10^{-4}$	$3.45 \cdot 10^{-5}$	$1.62 \cdot 10^{-8}$	$2.16 \cdot 10^{-10}$	$2.41 \cdot 10^{-2}$	$4.45 \cdot 10^{-3}$	$9.04 \cdot 10^{-5}$	0

Fig. 1. Boxplot of the fit metric for Case Study 1.

5.2 Case Study 2: high-order, highly-resonant system

We test the proposed method on a 16th order system:

$$G(p) = \sum_{i=1}^{8} \frac{c_i}{(p/\omega_i)^2 + 2(\xi_i/\omega_i)p + 1},$$

where the natural frequencies ω_i are located between 6 and 470 [rad/s], and the damping ratios are values between 0.001 and 0.0017. The input is a multisine formed by 16 sinusoids of random phase, and two hundred Monte Carlo runs are recorded with $N = 3000$, $h = 0.001$[s], and initial conditions equal to zero. Each run includes a zero-mean additive noise of variance equal to 2.25, which corresponds to approximately a signal-to-noise ratio of 20[dB]. Both SRIVC and BCD methods are initialized at the estimator given by the LSSVF method (Young, 1965). The tolerance factor of each method is set to $\epsilon = 10^{-16}$, and a maximum of 10 $\mathcal{A}$-iterations with $M_s = 200$ is used for BCD.

Figure 2 shows the fit of the SRIVC estimator compared to the proposed BCD method in a boxplot format, and also in a direct comparison plot. In this study, the BCD method returned better fit compared to the SRIVC estimator in 71.5% of the runs, and no run led to a failure in producing a result. In addition, Table 2 shows the MSEs of the parameters c_i for each approach. The MSEs related to the BCD method are lower or equal to the MSEs of the SRIVC estimates for all the numerator coefficients, which can be explained by the fact that a more parsimonious model is being fit to the data.

6. CONCLUSIONS

In this paper we have derived a method for estimating continuous-time models in an additive form. First, we explored the fact that certain systems might benefit from a more parsimonious model description if a modal form is considered. We proposed a block-coordinate descent method to identify models with this structure, and we proved its convergence. The numerical simulations show
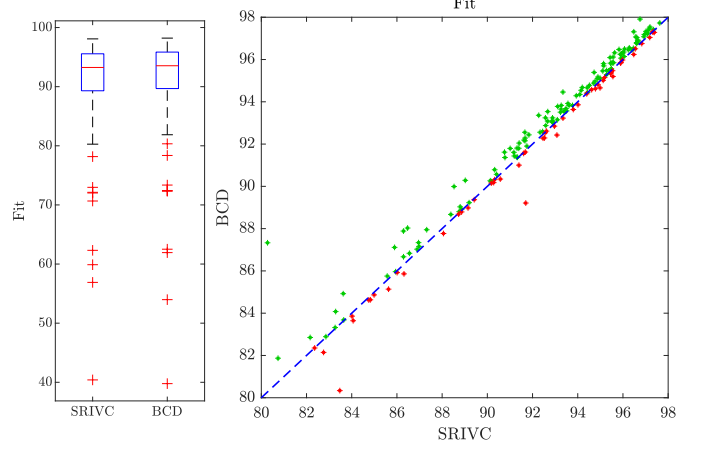

Fig. 2. Left figure: Boxplot of the fit metric for the SRIVC and BCD methods, Case Study 2. Right figure: Comparison plot between BCD and SRIVC, Case Study 2. Green dots correspond to runs where BCD outperforms SRIVC. Red dots represent the opposite, and the dashed blue line is the separatrix.

that the method delivers more accurate models than the standard indirect and direct methods, and its potential for the estimation of large-order and highly-resonant systems has been studied.

REFERENCES

Garnier, H., Gilson, M., Young, P.C., and Huselstein, E. (2007). An optimal IV technique for identifying continuous-time transfer function model of multiple input systems. *Control Engineering Practice*, 15(4), 471–486.

Garnier, H. and Wang, L. (2008). *Identification of Continuous-time Models from Sampled Data.* Springer.

Gilson, M., Welsh, J.S., and Garnier, H. (2017). A frequency localizing basis function-based IV method for wideband system identification. *IEEE Transactions on Control Systems Technology*, 26(1), 329–335.

González, R.A., Rojas, C.R., Pan, S., and Welsh, J.S. (2020). Consistent identification of continuous-time systems under multisine input signal excitation. *Automatica*, 133, Article 109859.

González, R.A., Rojas, C.R., Pan, S., and Welsh, J.S. (2022). Refined instrumental variable methods for unstable continuous-time systems in closed-loop. *International Journal of Control*, 1–15.

Luenberger, D.G. and Ye, Y. (2008). *Linear and Nonlinear Programming,* 3rd Edition. Springer.

Moheimani, S.O.R., Halim, D., and Fleming, A.J. (2003). *Spatial Control of Vibration: Theory and Experiments.* World Scientific.

Pan, S., González, R.A., Welsh, J.S., and Rojas, C.R. (2020). Consistency analysis of the Simplified Refined Instrumental Variable method for Continuous-time systems. *Automatica*, 113, Article 108767.

Table 2. MSEs of the estimated numerator parameters c_i for SRIVC and BCD, Case Study 2.

Method	Parameter True value	c_1 0.66	c_2 0.24	c_3 0.48	c_4 0.15	c_5 0.09	c_6 0.15	c_7 0.09	c_8 0.06
SRIVC	MSE	$6.52 \cdot 10^{-4}$	$6.74 \cdot 10^{-5}$	$1.61 \cdot 10^{-5}$	$1.13 \cdot 10^{-5}$	$2.19 \cdot 10^{-5}$	$1.04 \cdot 10^{-5}$	$6.06 \cdot 10^{-6}$	$3.87 \cdot 10^{-7}$
BCD	MSE	$6.33 \cdot 10^{-4}$	$6.42 \cdot 10^{-5}$	$1.61 \cdot 10^{-5}$	$1.09 \cdot 10^{-5}$	$2.14 \cdot 10^{-5}$	$1.04 \cdot 10^{-5}$	$6.00 \cdot 10^{-6}$	$3.84 \cdot 10^{-7}$

Pan, S., Nguyen, Q.C., Nguyen, V.T., and Welsh, J.S. (2021). Continuous-time system identification of a flexible cantilever beam. In *2021 IEEE Conference on Control Technology and Applications (CCTA)*, 868–873.

Rao, G.P. and Garnier, H. (2002). Numerical illustrations of the relevance of direct continuous-time model identification. In *15th IFAC World Congress on Automatic Control,* Barcelona, Spain, volume 35, 133–138.

Regalia, P.A. (1995). *Adaptive IIR Filtering in Signal Processing and Control.* Marcel Dekker, New York.

Söderström, T. (1975a). Ergodicity results for sample covariances. *Problems of Control and Information Theory*, 4(2), 131–138.

Söderström, T. and Stoica, P. (1982). Some properties of the output error method. *Automatica*, 18(1), 93–99.

Söderström, T. and Stoica, P. (1989). *System Identification.* Prentice-Hall.

Söderström, T. (1975b). On the uniqueness of maximum likelihood identification. *Automatica*, 11(2), 193–197.

Young, P.C. (1965). The determination of the parameters of a dynamic process. *IERE Journal of Radio and Electronic Engineering*, 29, 345–361.

Young, P.C. (1976). Some observations on instrumental variable methods of time-series analysis. *International Journal of Control*, 23(5), 593–612.

Young, P.C. and Jakeman, A.J. (1980). Refined instrumental variable methods of recursive time-series analysis. Part III, Extensions. *International Journal of Control*, 31(4), 741–764.

Appendix A. PROOF OF THEOREM 2

Proof. We adapt the theory in Section 7.7 of Luenberger and Ye (2008) to first analyze the global convergence of Algorithm 1 to the solution set $\tilde{\Gamma}_N = \{\mathbf{x} \in \Omega \colon \mathcal{A}(\mathbf{x}) = \mathbf{x}\}$. Afterwards, we prove that $\tilde{\Gamma}_N \subseteq \Gamma_N$. By Lemma B.1 of Söderström and Stoica (1989) (see also Söderström (1975a)) $V_N(\mathbf{x})$ converges almost surely (as $N \to \infty$) to

$$\bar{V}(\mathbf{x}) = \mathbb{E}\left\{ \left[y(kh) - \sum_{i=1}^{K} G_i(p, \boldsymbol{\theta}_i) u(kh) \right]^2 \right\} \quad (A.1)$$

$$= \frac{1}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \left| \sum_{i=1}^{K} \left[\tilde{G}_i^{\prime *}(e^{i\omega h}) - \tilde{G}_i(e^{i\omega h}, \boldsymbol{\theta}_i) \right] \right|^2 \Phi_u(\omega) d\omega + \sigma^2,$$

where Φ_u is the spectrum of the sampled input signal, and $\tilde{G}_i^*$, $\tilde{G}_i$ are the discrete-time ZOH equivalents of the system G_i^* and model G_i, respectively. Since V_N is continuous and Ω is compact, this convergence is uniform. In the following, we prove the necessary ingredients for applying the Global Convergence Theorem of Luenberger and Ye (2008).

- *Well-posedness of* $\mathbf{S}$: Due to the uniform convergence of V_N to $\bar{V}$, for almost every realization ξ, there exists a sample size $N'(\xi) \in \mathbb{N}$ such that for all $\mathbf{x} = [\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_K^\top]^\top \in \Omega$ and $N \geq N'(\xi)$, the minimizer of $V(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}, \boldsymbol{\theta}_{i+1}, \ldots, \boldsymbol{\theta}_K)$ with respect to $\boldsymbol{\theta}$

is unique. Thus, for all $N \geq N'(\xi)$, $\mathbf{S}$ is a well-defined point-to-point mapping.

- *Descent of algorithm (with respect to V_N)*: From the definition of the algorithm $\mathcal{A}$ in (9) and the solution set $\tilde{\Gamma}_N$, it follows that if $\mathbf{x} \in \tilde{\Gamma}_N$, then $V_N(\mathcal{A}(\mathbf{x})) \leq V_N(\mathbf{x})$. Otherwise, if $\mathbf{x} \notin \tilde{\Gamma}_N$, then $\mathbf{x} \neq \mathcal{A}(\mathbf{x})$ and therefore we must have $V_N(\mathcal{A}(\mathbf{x})) < V_N(\mathbf{x})$. Thus, V_N is a continuous *descent function* for $\tilde{\Gamma}_N$ and $\mathcal{A}$.

- *Closedness of $\mathcal{A}$*: From the theorem in Section 8.4 of Luenberger and Ye (2008), the map $\mathbf{S}$ is closed. In addition, the maps $\mathbf{C}^i$ $(i = 1, \ldots, n)$ are continuous, and thus closed. Therefore, by Corollary 1 in Section 7.7 of Luenberger and Ye (2008), $\mathcal{A}$ is closed.

From the previous points, it follows that all the conditions for the Global Convergence Theorem in Section 7.7 of Luenberger and Ye (2008) hold. This implies that, for almost every realization and for sufficiently large N, the limit of every converging subsequence of $(\mathbf{x}^l)_{l \in \mathbb{N}}$ belongs to the solution set $\tilde{\Gamma}_N$.

Now, to prove that $\tilde{\Gamma}_N \subseteq \Gamma_N$, take $\bar{\mathbf{x}} = [\bar{\boldsymbol{\theta}}_1^\top, \ldots, \bar{\boldsymbol{\theta}}_K^\top]^\top \in \tilde{\Gamma}_N$. Since $\bar{\boldsymbol{\theta}}_1$ is an critical point of the function $V_N^1(\boldsymbol{\theta}_1) := V_N(\boldsymbol{\theta}_1, \bar{\boldsymbol{\theta}}_2, \ldots, \bar{\boldsymbol{\theta}}_K)$ for fixed $\bar{\boldsymbol{\theta}}_2, \ldots, \bar{\boldsymbol{\theta}}_K$, then V_N^1 must satisfy the first-order optimality condition $\frac{\partial V_N^1}{\partial \boldsymbol{\theta}_1}|_{\boldsymbol{\theta}_1 = \bar{\boldsymbol{\theta}}_1} = \mathbf{0}$. By definition, this means that $\frac{\partial V_N}{\partial \boldsymbol{\theta}_1}|_{\boldsymbol{\theta}_1 = \bar{\boldsymbol{\theta}}_1} = \mathbf{0}$. Repeating this argument for $\bar{\boldsymbol{\theta}}_2, \ldots, \bar{\boldsymbol{\theta}}_K$ leads to the desired conclusion, namely, that $\bar{\mathbf{x}} \in \Gamma_N$. □

Appendix B. PROOF OF LEMMA 4.1

Proof. As $s \to \infty$, any converging point $\boldsymbol{\theta}_i^{l+1} := \lim_{s \to \infty} \boldsymbol{\theta}_{i,s}^{l+1}$ of the SRIVC iterations in (12) must satisfy

$$\boldsymbol{\theta}_i^{l+1} = \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\mathrm{f}}(kh, \boldsymbol{\theta}_i^{l+1}) \boldsymbol{\varphi}_{\mathrm{f}}^\top(kh, \boldsymbol{\theta}_i^{l+1}) \right]^{-1}$$
$$\times \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\mathrm{f}}(kh, \boldsymbol{\theta}_i^{l+1}) \tilde{y}_{\mathrm{f}}(kh, \boldsymbol{\theta}_i^{l+1}) \right],$$

which, given the non-singularity of the normal matrix above, is equivalent to

$$\frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\mathrm{f}}(kh, \boldsymbol{\theta}_i^{l+1}) \left[\tilde{y}_{\mathrm{f}}(kh, \boldsymbol{\theta}_i^{l+1}) - \boldsymbol{\varphi}_{\mathrm{f}}^\top(kh, \boldsymbol{\theta}_i^{l+1}) \boldsymbol{\theta}_i^{l+1} \right] = \mathbf{0}. \quad (B.1)$$

However, by leveraging the expressions for $\boldsymbol{\varphi}_{\mathrm{f}}$ and $\tilde{y}_{\mathrm{f}}$ in (13) and (14), we find that

$$\tilde{y}_{\mathrm{f}}(kh, \boldsymbol{\theta}_i^{l+1}) - \boldsymbol{\varphi}_{\mathrm{f}}^\top(kh, \boldsymbol{\theta}_i^{l+1}) \boldsymbol{\theta}_i^{l+1}$$
$$= \frac{1}{A_i^{l+1}(p)} \tilde{y}(kh) - \frac{(1 - A_i^{l+1}(p))}{A_i^{l+1}(p)} \tilde{y}(kh) - \frac{B_i^{l+1}(p)}{A_i^{l+1}(p)} u(kh)$$
$$= e(kh, \boldsymbol{\theta}_i^{l+1}).$$

Thus, replacing this result into (B.1) also leads to (11), concluding the proof. □