On the Relation between Discrete and Continuous-time Refined Instrumental Variable Methods

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

Abstract— The Refined Instrumental Variable method for discrete-time systems (RIV) and its variant for continuous-time systems (RIVC) are popular methods for the identification of linear systems in open-loop. The continuous-time equivalent of the transfer function estimate given by the RIV method is commonly used as an initialization point for the RIVC estimator. In this paper, we prove that these estimators share the same converging points for finite sample size when the continuous-time model has relative degree zero or one. This relation does not hold for higher relative degrees. Then, we propose a modification of the RIV method whose continuous-time equivalent is equal to the RIVC estimator for any non-negative relative degree. The implications of the theoretical results are illustrated via a simulation example.

I. Introduction

System identification deals with the problem of obtaining mathematical models of dynamical systems from data [1]. A distinction is made between discrete-time (DT) and continuous-time (CT) models. In DT system identification, it is assumed that a complete description of the system can be made by only observing its behavior at specific time instants. On the contrary, CT system identification seeks models that reflect the properties of the system for any moment in time.

There are two main approaches to CT system identification: indirect and direct [2]. The indirect approach consists in estimating a DT model with the data, and then converting this model into continuous-time. On the other hand, the direct approach does not use intermediate DT models. It is known that the standard indirect approach leads to models with relative degree one, independent of that of the strictly proper CT system [3]; such issue does not arise in the direct approach, since any number of zeros can be accommodated without the need of an additional optimization step.

For either the direct or indirect approach, refined instrumental variable methods can be applied. The Refined Instrumental Variable (RIV) method for DT systems and its simplified embodiment (SRIV, [4]) are used for estimating Box-Jenkins and output error models respectively, and the CT equivalents of their estimated models are commonly used for initializing the direct approach [5], [6]. Some of the most celebrated direct methods are the CT variants of these estimators, called RIVC and SRIVC [7]. The RIVC and SRIVC

algorithms compute iterative instrumental variable steps by prefiltering the data through CT filters, and are initialized by the CT equivalent of the RIV or SRIV estimators in the MATLAB Contsid Toolbox [8].

The RIV methods have been used for modeling climate dynamics [9] as well as mechanical systems [10], and extensions of these estimators have been proposed for LPV [11], Hammerstein-Wiener [12] and unstable systems [13]. A comprehensive overview of these methods can be found in [14]; however, such work overlooks a relation between the converging points of the DT and CT RIV variants.

In this paper, we show that the indirect approach with the RIV estimator provides the same estimate as RIVC at convergence of its iterations (considering *finite* sample size) for CT systems with relative degree one or zero. Although this result no longer holds for higher relative degrees due to parsimony issues, we propose a modification to the RIV estimator, termed Adapted RIV (ARIV), that is shown to extend this equivalence for any non-negative relative degree. As a byproduct, we show that the ARIV estimator can impose relative degree constraints in its CT equivalent without additional optimization steps. The theoretical results include the relationship between the simplified versions of these estimators (SRIV and SRIVC) as a special case.

The remainder of the paper is organized as follows. In Section II, we describe the system and model. Section III presents the notation of the unified RIV estimator; this estimator is later analyzed in Section IV, which contains the main contributions. In Section V we provide a simulation example, and we conclude the paper in Section VI.

II. System and model setup

Consider the single-input single-output, linear and time-invariant (LTI), asymptotically stable, CT system

$$x(t) = \frac{B_c^*(p)}{A_c^*(p)} u(t),$$

where p is the Heaviside operator (i.e., $pu(t) = \mathrm{d}u(t)/\mathrm{d}t$), and $u(t)$ is the input. The numerator and denominator polynomials $B_c^*(p)$ and $A_c^*(p)$ are coprime and given by

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

with $a_n^* \neq 0$, and $n \geq m$. These polynomials are jointly described by the parameter vector

$$\boldsymbol{\theta}_c^* = \begin{bmatrix} a_1^*, & a_2^*, & \ldots, & a_n^*, & b_0^*, & b_1^*, & \ldots, & b_m^* \end{bmatrix}^\top. \quad (1)$$

A noisy output measurement is retrieved every h[s], i.e.,

$$y(kh) = x(kh) + \frac{C^*(q)}{D^*(q)}e(kh),$$

where q denotes the forward shift operator, $e(kh)$ describes a zero-mean white noise stochastic process of finite variance that is uncorrelated with the input sequence $u(kh)$, and

$$C^*(q) = 1 + c_1^* q^{-1} + c_2^* q^{-2} + \cdots + c_{m_c}^* q^{-m_c},$$
$$D^*(q) = 1 + d_1^* q^{-1} + d_2^* q^{-2} + \cdots + d_{n_d}^* q^{-n_d},$$

with the polynomial degrees satisfying $n_d \geq m_c$. The coefficients c_i^* and d_j^* are combined in a parameter vector $\boldsymbol{\eta}^*$ of the same form as $\boldsymbol{\theta}_c^*$ in (1).

Based on N input and output data samples, $\{u(kh), y(kh)\}_{k=1}^N$, the goal is to determine a model for $G_c^*(p) := B_c^*(p)/A_c^*(p)$ in the CT or DT domain, and possibly of the noise filter $H^*(q) := C^*(q)/D^*(q)$ also.

A. Discrete-time equivalent and inverse of sampling

Throughout this paper we assume that the input signal is constant between samples, i.e., it has a zero-order hold (ZOH) intersample behavior. Therefore, the system we intend to model can be exactly described at the sampling instants by its DT ZOH equivalent, which has the form

$$y(kh) = G_d^*(q)u(kh) + H^*(q)e(kh), \qquad (2)$$

with $G_d^*(q) = B_d^*(q)/A_d^*(q)$, where

$$A_d^*(q) = \alpha_n^* q^n + \alpha_{n-1}^* q^{n-1} + \cdots + \alpha_1^* q + 1,$$
$$B_d^*(q) = \beta_n^* q^n + \beta_{n-1}^* q^{n-1} + \cdots + \beta_1^* q + \beta_0^*.$$

The parameter vector associated with $G_d^*(q)$ is denoted as $\boldsymbol{\theta}_d^*$ and has the same structure as $\boldsymbol{\theta}_c^*$ but is formed by the DT system parameters instead of the CT parameters. If the CT system is strictly proper, then $\beta_n^* = 0$ and the last element of $\boldsymbol{\theta}_d^*$ is omitted. Note that, however, almost any strictly proper CT system leads to $\beta_{n-1}^* \neq 0$. Although this fact is well known [15], explicit results on the resulting DT relative degree are difficult to find in the literature. For completeness, the formal statement with its proof is presented next.

Proposition 2.1: Consider a strictly proper, LTI, CT system $G_c^*(p)$. The relative degree of the DT ZOH equivalent of $G_c^*(p)$ is $r \geq 1$ if and only if $y_c^*(h) = y_c^*(2h) = \cdots = y_c^*([r-1]h) = 0$ and $y_c^*(rh) \neq 0$, where $y_c^*(t)$ is the step response of $G_c^*(p)$ and h is the sampling period.

Proof: The DT ZOH equivalent of $G_c^*(p)$ is computed by

$$G_d^*(z) = (1 - z^{-1})\mathcal{Z}\{y_c^*(kh)\}$$
$$= \frac{y_c^*(h)}{z} + \frac{y_c^*(2h) - y_c^*(h)}{z^2} + \frac{y_c^*(3h) - y_c^*(2h)}{z^3} + \cdots.$$

Therefore, the transfer function $G_d^*(z)$ has relative degree r if and only if $y_c^*(rh) - y_c^*([r-1]h) \neq 0$ and

$$y_c^*(h) = y_c^*(2h) - y_c^*(h) = \cdots = y_c^*([r-1]h) - y_c^*([r-2]h) = 0,$$

from which the statement follows. ∎

The indirect approach to CT system identification requires a link between the DT parameter vector $\boldsymbol{\theta}_d$ and its CT equivalent $\boldsymbol{\theta}_c$. This link is presented in Definition 2.1, which is used for analyzing the relationship between discrete and continuous-time refined instrumental variable methods.

Definition 2.1 (Inverse ZOH transformation): Given any DT transfer function described by $\boldsymbol{\theta}_d$, we define the *inverse ZOH transformation* by $\mathbf{f}: \mathbb{R}^{2n+1} \to \mathbb{R}^{2n+1}; \boldsymbol{\theta}_d \mapsto \boldsymbol{\theta}_c = \mathbf{f}(\boldsymbol{\theta}_d)$. The vector $\mathbf{f}(\boldsymbol{\theta}_d)$ is of the form (1) with $m = n$ that describes the transfer function $\mathbf{C}_c(p\mathbf{I} - \mathbf{A}_c)^{-1}\mathbf{B}_c + D_c$, where $\mathbf{C}_c = \big[\beta_0 - \frac{\beta_n}{\alpha_n}, \beta_1 - \frac{\beta_n\alpha_1}{\alpha_n}, \ldots, \beta_{n-1} - \frac{\beta_n\alpha_{n-1}}{\alpha_n}\big]$, $D_c = \beta_n/\alpha_n$, and where $\mathbf{A}_c \in \mathbb{R}^{n \times n}$ with $\mathbf{B}_c \in \mathbb{R}^{n \times 1}$ are computed by

$$\begin{bmatrix} \mathbf{A}_c & \mathbf{B}_c \\ \mathbf{0} & 0 \end{bmatrix} = \frac{1}{h}\log\left(\begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ \frac{-1}{\alpha_n} & \frac{-\alpha_1}{\alpha_n} & \frac{-\alpha_2}{\alpha_n} & \ldots & \frac{-\alpha_{n-1}}{\alpha_n} & \frac{1}{\alpha_n} \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix}\right).$$

In case $\beta_n = 0$ (i.e., the DT model is strictly proper), we know that $b_n = 0$ and we thus consider $\mathbf{f}$ to be the mapping between the strictly proper models only.

The function $\mathbf{f}$ is known to be well-defined and bijective when 1) there is no negative real pole in the DT transfer function associated with $\boldsymbol{\theta}_d$, and 2) the sampling radian frequency π/h is larger than twice the largest imaginary part of the poles related to $\mathbf{f}(\boldsymbol{\theta}_d)$ [16]. The functions $\mathbf{f}$ and $\mathbf{f}^{-1}$ are well-defined and differentiable in the domain where these two conditions hold, thus leading to its Jacobian matrix $\partial \mathbf{f}/\partial \boldsymbol{\theta}_d$ being non-singular at any point of that domain [17].

III. REFINED INSTRUMENTAL VARIABLES: A UNIFIED NOTATION

The motivation behind refined instrumental variables is to solve the non-convex optimization problem of maximum likelihood estimation by iterations stemming from a pseudo-linear regression. For brevity, similar to [14], here we jointly introduce the RIV (for DT identification) and RIVC (for CT identification) estimators via a unified notation.

We define ξ as the unified operator (i.e., q for the RIV estimator and p for RIVC), and the unified subscript $(\cdot)_\nu$, where $\nu \in \{d, c\}$ depending on whether the RIV or RIVC estimator is considered. Assuming that the jth iteration of the system parameters estimate, $\boldsymbol{\theta}_{\nu,j}$ (with model $B_{\nu,j}(\xi)/A_{\nu,j}(\xi)$), is given, the jth noise model iteration of the RIV and RIVC estimators is obtained by fitting an ARMA model to the estimated noise sequence $\{y(kh) - [B_{\nu,j}(\xi)/A_{\nu,j}(\xi)]u(kh)\}_{k=1}^N$. This procedure gives the noise model $C_{j+1}(q)/D_{j+1}(q)$, represented by the noise model parameter vector $\boldsymbol{\eta}_{j+1}$.

There are several ways to fit an ARMA model to obtain $\boldsymbol{\eta}_{j+1}$. In this work we consider the prediction error method [1]. Formally, we define the function that links $\boldsymbol{\theta}_{\nu,j}$ and $\boldsymbol{\eta}_{j+1}$ as $\mathbf{g}_\nu: \mathbb{R}^{n+m+1} \to \mathbb{R}^{m_c+n_d}; \boldsymbol{\theta}_{\nu,j} \mapsto \boldsymbol{\eta}_{j+1} = \mathbf{g}_\nu(\boldsymbol{\theta}_{\nu,j})$, where

$$\mathbf{g}_\nu(\boldsymbol{\theta}_{\nu,j}) = \arg\min_{\boldsymbol{\eta}} \sum_{k=1}^N \left[\frac{D(q,\boldsymbol{\eta})}{C(q,\boldsymbol{\eta})}\left(y(kh) - \frac{B_{\nu,j}(\xi)}{A_{\nu,j}(\xi)}u(kh)\right)\right]^2.$$

Remark 3.1: The notation $C(q,\boldsymbol{\eta})$ and $D(q,\boldsymbol{\eta})$ is used to stress the dependence of these polynomials on the parameter

vector $\boldsymbol{\eta}$. Similarly, in the sequel we use $A_\nu(\xi, \boldsymbol{\theta}_{\nu,j})$ and $B_\nu(\xi, \boldsymbol{\theta}_{\nu,j})$ interchangeably for $A_{\nu,j}(\xi)$ and $B_{\nu,j}(\xi)$.

After the noise model iteration is computed, an RIV step is proposed for minimizing $\mathbf{g}_\nu(\boldsymbol{\theta}_{\nu,j})$ upon convergence [14]. The jth system model iteration is given by

$$\boldsymbol{\theta}_{\nu,j+1} = \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\nu,f}(kh, \boldsymbol{\theta}_{\nu,j}) \boldsymbol{\varphi}_{\nu,f}^\top(kh, \boldsymbol{\theta}_{\nu,j}) \right]^{-1}$$
$$\times \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\nu,f}(kh, \boldsymbol{\theta}_{\nu,j}) y_{\nu,f}(kh, \boldsymbol{\theta}_{\nu,j}) \right], \quad (3)$$

where the matrix being inverted is called the *modified normal matrix*. The filtered regressor $\boldsymbol{\varphi}_{\nu,f}(kh, \boldsymbol{\theta}_{\nu,j})$, filtered instrument $\hat{\boldsymbol{\varphi}}_{\nu,f}(kh, \boldsymbol{\theta}_{\nu,j})$ and filtered output $y_{\nu,f}(kh, \boldsymbol{\theta}_{\nu,j})$ are respectively[1] given by

$$\boldsymbol{\varphi}_{\nu,f}(kh, \boldsymbol{\theta}_{\nu,j}) = \frac{D_{j+1}(q)}{C_{j+1}(q)} \left[\frac{-\xi}{A_{\nu,j}(\xi)} y(kh), \ldots, \frac{-\xi^n}{A_{\nu,j}(\xi)} y(kh), \right.$$
$$\left. \frac{1}{A_{\nu,j}(\xi)} u(kh), \ldots, \frac{\xi^m}{A_{\nu,j}(\xi)} u(kh) \right]^\top,$$

$$\hat{\boldsymbol{\varphi}}_{\nu,f}(kh, \boldsymbol{\theta}_{\nu,j}) = \frac{D_{j+1}(q)}{C_{j+1}(q)} \left[\frac{-\xi B_{\nu,j}(\xi)}{A_{\nu,j}^2(\xi)} u(kh), \ldots, \right.$$
$$\left. \frac{-\xi^n B_{\nu,j}(\xi)}{A_{\nu,j}^2(\xi)} u(kh), \frac{1}{A_{\nu,j}(\xi)} u(kh), \ldots, \frac{\xi^m}{A_{\nu,j}(\xi)} u(kh) \right]^\top,$$

$$y_{\nu,f}(kh, \boldsymbol{\theta}_{\nu,j}) = \frac{D_{j+1}(q)}{C_{j+1}(q)} \frac{1}{A_{\nu,j}(\xi)} y(kh). \quad (4)$$

In case the noise model is not estimated and thus set to unity (i.e., $C_j(q) = D_j(q) = 1$) the iterations in (3) describe the SRIV and SRIVC estimators [4], [7].

Remark 3.2: In the expressions above we have introduced a mixed notation of CT filters with sampled signals for the case $\nu = \mathrm{c}, \xi = p$ (i.e., for the RIVC and SRIVC estimators). If $G(p)$ is a CT filter and $x(kh)$ is a sampled signal, then $G(p)x(kh)$ implies that the signal $x(kh)$ is being interpolated using a ZOH, and the resulting output of the filter is sampled at $t = kh$.

Given the parameter estimate vector $\bar{\boldsymbol{\theta}}_\mathrm{d}$ obtained from the RIV iterations (3) at convergence as j tends to infinity for a fixed N, the associated CT parameter vector (i.e., the indirect approach estimate) is simply given by $\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d})$.

IV. RELATION BETWEEN DISCRETE-TIME AND CONTINUOUS-TIME REFINED INSTRUMENTAL VARIABLE METHODS

Before stating the main theoretical results of this paper, we present two assumptions:

Assumption 4.1: The system $G_\mathrm{c}^*(p)$ is proper ($n \geq m$) and asymptotically stable, with $A_\mathrm{c}^*(p)$ and $B_\mathrm{c}^*(p)$ being coprime.

Assumption 4.2: For all j sufficiently large, $A_{\mathrm{d},j}(q)$ has no negative real zero in the RIV algorithm, and the sampling

frequency π/h is larger than twice the largest imaginary part of the zeros of $A_{\mathrm{c},j}(p)$ in the RIVC algorithm.

While both assumptions are standard in the analysis of RIV algorithms (see, e.g., [18], [19]), the asymptotic stability of $G_\mathrm{c}^*(p)$ may not be needed in case the identification is performed in closed-loop, although a more involved filtering technique must be implemented [13]. Furthermore, Assumption 4.2 is equivalent to requiring the inverse ZOH transformation $\mathbf{f}$ to be well defined along the iterative process. This will typically be the case if the sampling period is chosen adequately and the model iterations do not diverge.

Theorem 4.1 shows that for models with relative degree zero or one, the RIV method with inverse ZOH transformation $\mathbf{f}$ provides the same model as the RIVC method upon convergence as the number of iterations j tends to infinity for a fixed and finite sample size N.

Theorem 4.1: Consider the RIV and RIVC estimators described in Section III, and assume that upon convergence in iterations ($j \to \infty$) for a fixed sample size, their modified normal matrices are non-singular. Furthermore, assume that Assumptions 4.1 and 4.2 are satisfied, and that $m = n - 1$ or $m = n$. Then, the RIVC and RIV methods have the same number of limiting points and they are linked by $\bar{\boldsymbol{\theta}}_\mathrm{c} = \mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d})$, where $\bar{\boldsymbol{\theta}}_\mathrm{c}$ and $\bar{\boldsymbol{\theta}}_\mathrm{d}$ are the limiting point(s) of the RIVC and RIV methods, respectively.

Proof: Any limiting point $\bar{\boldsymbol{\theta}}_\mathrm{d} = \lim_{j \to \infty} \boldsymbol{\theta}_{\mathrm{d},j}$ of the RIV iterations must satisfy

$$\bar{\boldsymbol{\theta}}_\mathrm{d} = \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\mathrm{d},f}(kh, \bar{\boldsymbol{\theta}}_\mathrm{d}) \boldsymbol{\varphi}_{\mathrm{d},f}^\top(kh, \bar{\boldsymbol{\theta}}_\mathrm{d}) \right]^{-1}$$
$$\times \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\mathrm{d},f}(kh, \bar{\boldsymbol{\theta}}_\mathrm{d}) y_{\mathrm{d},f}(kh, \bar{\boldsymbol{\theta}}_\mathrm{d}) \right].$$

Provided the modified normal matrix is non-singular, the condition above reduces to

$$\sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\mathrm{d},f}(kh, \bar{\boldsymbol{\theta}}_\mathrm{d}) \left[y_{\mathrm{d},f}(kh, \bar{\boldsymbol{\theta}}_\mathrm{d}) - \boldsymbol{\varphi}_{\mathrm{d},f}^\top(kh, \bar{\boldsymbol{\theta}}_\mathrm{d}) \bar{\boldsymbol{\theta}}_\mathrm{d} \right] = \mathbf{0}. \quad (5)$$

After some algebraic manipulations, we can see that

$$\hat{\boldsymbol{\varphi}}_{\mathrm{d},f}(kh, \bar{\boldsymbol{\theta}}_\mathrm{d}) = \frac{D(q, \mathbf{g}_\mathrm{d}(\bar{\boldsymbol{\theta}}_\mathrm{d}))}{C(q, \mathbf{g}_\mathrm{d}(\bar{\boldsymbol{\theta}}_\mathrm{d}))} \frac{\partial}{\partial \boldsymbol{\theta}_\mathrm{d}} \frac{B_\mathrm{d}(q, \boldsymbol{\theta}_\mathrm{d})}{A_\mathrm{d}(q, \boldsymbol{\theta}_\mathrm{d})} u(kh) \bigg|_{\boldsymbol{\theta}_\mathrm{d} = \bar{\boldsymbol{\theta}}_\mathrm{d}}, \quad (6)$$

and

$$y_{\mathrm{d},f}(kh, \bar{\boldsymbol{\theta}}_\mathrm{d}) - \boldsymbol{\varphi}_{\mathrm{d},f}^\top(kh, \bar{\boldsymbol{\theta}}_\mathrm{d}) \bar{\boldsymbol{\theta}}_\mathrm{d} =$$
$$\frac{D(q, \mathbf{g}_\mathrm{d}(\bar{\boldsymbol{\theta}}_\mathrm{d}))}{C(q, \mathbf{g}_\mathrm{d}(\bar{\boldsymbol{\theta}}_\mathrm{d}))} \left(y(kh) - \frac{B_\mathrm{d}(q, \bar{\boldsymbol{\theta}}_\mathrm{d})}{A_\mathrm{d}(q, \bar{\boldsymbol{\theta}}_\mathrm{d})} u(kh) \right). \quad (7)$$

Replacing (6) and (7) in (5) leads to

$$\sum_{k=1}^{N} \frac{D(q, \mathbf{g}_\mathrm{d}(\bar{\boldsymbol{\theta}}_\mathrm{d}))}{C(q, \mathbf{g}_\mathrm{d}(\bar{\boldsymbol{\theta}}_\mathrm{d}))} \frac{\partial}{\partial \boldsymbol{\theta}_\mathrm{d}} \frac{B_\mathrm{d}(q, \boldsymbol{\theta}_\mathrm{d})}{A_\mathrm{d}(q, \boldsymbol{\theta}_\mathrm{d})} u(kh) \bigg|_{\boldsymbol{\theta}_\mathrm{d} = \bar{\boldsymbol{\theta}}_\mathrm{d}} \times$$
$$\frac{D(q, \mathbf{g}_\mathrm{d}(\bar{\boldsymbol{\theta}}_\mathrm{d}))}{C(q, \mathbf{g}_\mathrm{d}(\bar{\boldsymbol{\theta}}_\mathrm{d}))} \left(y(kh) - \frac{B_\mathrm{d}(q, \bar{\boldsymbol{\theta}}_\mathrm{d})}{A_\mathrm{d}(q, \bar{\boldsymbol{\theta}}_\mathrm{d})} u(kh) \right) = \mathbf{0}. \quad (8)$$

The existence of the inverse ZOH transformation for $\bar{\boldsymbol{\theta}}_\mathrm{d}$ is ensured by Assumption 4.2. Since the

---

[1] Note that $\boldsymbol{\varphi}_{\nu,f}, \hat{\boldsymbol{\varphi}}_{\nu,f}$ and $y_{\nu,f}$ have a subscript f to stress the prefiltering process that must be performed. Also, note that for simplicity we have not written the dependence of these signals on the noise model.

ZOH equivalence relation is exact at the sampling instants [20], the equality $[B_\mathrm{d}(q,\bar{\boldsymbol{\theta}}_\mathrm{d})/A_\mathrm{d}(q,\bar{\boldsymbol{\theta}}_\mathrm{d})]u(kh) = [B_\mathrm{c}(p,\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d}))/A_\mathrm{c}(p,\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d}))]u(kh)$ holds for all $k \in \mathbb{N}$. By the same argument, we find that $\mathbf{g}_\mathrm{d}(\bar{\boldsymbol{\theta}}_\mathrm{d}) = \mathbf{g}_\mathrm{c}(\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d}))$. Moreover, the chain rule for gradients yields

$$
\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\theta}_\mathrm{d}} \frac{B_\mathrm{d}(q,\boldsymbol{\theta}_\mathrm{d})}{A_\mathrm{d}(q,\boldsymbol{\theta}_\mathrm{d})}u(kh)\bigg|_{\boldsymbol{\theta}_\mathrm{d}=\bar{\boldsymbol{\theta}}_\mathrm{d}} &= \frac{\partial}{\partial\boldsymbol{\theta}_\mathrm{d}} \frac{B_\mathrm{c}(p,\mathbf{f}(\boldsymbol{\theta}_\mathrm{d}))}{A_\mathrm{c}(p,\mathbf{f}(\boldsymbol{\theta}_\mathrm{d}))}u(kh)\bigg|_{\boldsymbol{\theta}_\mathrm{d}=\bar{\boldsymbol{\theta}}_\mathrm{d}} \\
&= \frac{\partial\mathbf{f}}{\partial\boldsymbol{\theta}_\mathrm{d}}\bigg|_{\boldsymbol{\theta}_\mathrm{d}=\bar{\boldsymbol{\theta}}_\mathrm{d}} \frac{\partial}{\partial\boldsymbol{\theta}_\mathrm{c}} \frac{B_\mathrm{c}(p,\boldsymbol{\theta}_\mathrm{c})}{A_\mathrm{c}(p,\boldsymbol{\theta}_\mathrm{c})}u(kh)\bigg|_{\boldsymbol{\theta}_\mathrm{c}=\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d})}. \quad (9)
\end{aligned}
$$

Note that $\partial\mathbf{f}/\partial\boldsymbol{\theta}_\mathrm{d}$ is non-singular thanks to $\mathbf{f}$ and $\mathbf{f}^{-1}$ being continuous and differentiable in the domain of interest due to Assumption 4.2. Such matrix does not depend on N and can therefore be factored out of the sum in (8) thanks to the linearity of the transfer functions. Thus, (8) is equivalent to

$$
\begin{aligned}
&\sum_{k=1}^{N} \frac{D(q,\mathbf{g}_\mathrm{c}(\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d})))}{C(q,\mathbf{g}_\mathrm{c}(\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d})))} \frac{\partial}{\partial\boldsymbol{\theta}_\mathrm{c}} \frac{B_\mathrm{c}(p,\boldsymbol{\theta}_\mathrm{c})}{A_\mathrm{c}(p,\boldsymbol{\theta}_\mathrm{c})}u(kh)\bigg|_{\boldsymbol{\theta}_\mathrm{c}=\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d})} \times \\
&\frac{D(q,\mathbf{g}_\mathrm{c}(\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d})))}{C(q,\mathbf{g}_\mathrm{c}(\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d})))} \left(y(kh) - \frac{B_\mathrm{c}(p,\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d}))}{A_\mathrm{c}(p,\mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d}))}u(kh) \right) = \mathbf{0}.
\end{aligned}
$$

On the other hand, by following the same arguments to derive (8) but for the RIVC method, any limiting point $\bar{\boldsymbol{\theta}}_\mathrm{c}$ satisfies

$$
\begin{aligned}
&\sum_{k=1}^{N} \frac{D(q,\mathbf{g}_\mathrm{c}(\bar{\boldsymbol{\theta}}_\mathrm{c}))}{C(q,\mathbf{g}_\mathrm{c}(\bar{\boldsymbol{\theta}}_\mathrm{c}))} \frac{\partial}{\partial\boldsymbol{\theta}_\mathrm{c}} \frac{B_\mathrm{c}(p,\boldsymbol{\theta}_\mathrm{c})}{A_\mathrm{c}(p,\boldsymbol{\theta}_\mathrm{c})}u(kh)\bigg|_{\boldsymbol{\theta}_\mathrm{c}=\bar{\boldsymbol{\theta}}_\mathrm{c}} \times \\
&\frac{D(q,\mathbf{g}_\mathrm{c}(\bar{\boldsymbol{\theta}}_\mathrm{c}))}{C(q,\mathbf{g}_\mathrm{c}(\bar{\boldsymbol{\theta}}_\mathrm{c}))} \left(y(kh) - \frac{B_\mathrm{c}(p,\bar{\boldsymbol{\theta}}_\mathrm{c})}{A_\mathrm{c}(p,\bar{\boldsymbol{\theta}}_\mathrm{c})}u(kh) \right) = \mathbf{0}.
\end{aligned}
$$

By comparing the characterizations of $\bar{\boldsymbol{\theta}}_\mathrm{d}$ and $\bar{\boldsymbol{\theta}}_\mathrm{c}$, we conclude that the limiting points of the RIV and RIVC estimators are linked by $\bar{\boldsymbol{\theta}}_\mathrm{c} = \mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d})$. ∎

Corollary 4.1: Under the same assumptions as in Theorem 4.1, the SRIVC and SRIV methods have the same number of limiting points and they are linked by $\bar{\boldsymbol{\theta}}_\mathrm{c} = \mathbf{f}(\bar{\boldsymbol{\theta}}_\mathrm{d})$, where $\bar{\boldsymbol{\theta}}_\mathrm{c}$ and $\bar{\boldsymbol{\theta}}_\mathrm{d}$ are the limiting point(s) of the SRIVC and SRIV methods, respectively.

Proof: Direct by fixing $C_j(q) = D_j(q) = 1$ in the proof for Theorem 4.1. ∎

Remark 4.1: The non-singularity of the modified normal matrices of the refined instrumental variable methods depends on the persistence of excitation order of the input, as well as the amount of over-parametrization, if any [19]. General conditions for the non-singularity of this matrix have not been addressed in the literature, although sufficient conditions for the generic non-singularity of such matrix for the SRIVC method can be found in [18]. Similar conditions can be derived for the SRIV estimator, although they are outside the scope of this work.

Theorem 4.1 shows that, under similar initialization conditions, the indirect and direct approaches with refined instrumental variables approach the *same* estimate at convergence in iterations when $m = n - 1$ or $m = n$. This theorem only applies for such values of m since only in these two cases there is a bijection between the CT and DT models (corresponding to $\mathbf{f}$ or its reduced version). In such scenarios, one could argue that it is useless to perform the RIVC

iterations with the RIV estimate as an initialization point, as it is done in some applications of the RIVC method [6]. However, it has been noted in [21] that the CT equivalent of the RIV estimator can fail to deliver reliable models, usually in the presence of fast sampling or stiff systems. Possible discrepancies between both methods when $m = n - 1$ or $m = n$ can be explained by

- Ill-conditioning of the Jacobian $\partial\mathbf{f}/\partial\boldsymbol{\theta}_\mathrm{d}$: for stiff systems this matrix can be severely ill-conditioned, which can affect the convergence of the iterative procedure of the RIV estimator. Also, instability of the estimates can arise more frequently within the iterations, thus requiring ad-hoc stabilization steps [13].

- Misspecification of the intersample behavior in the RIVC method: If the intersample behavior used for prefiltering the input in the RIVC method does not match with the nature of the DT equivalents in the RIV estimator, then both methods will deliver different results in general.

- Choice of initialization: Convergence is not guaranteed for finite samples, and the estimators may converge to (different) local minima if poorly initialized.

A. The Adapted RIV estimator

The indirect and direct approaches no longer produce the same estimates for $m < n - 1$, since they propose different model structures. It has been noted in [6] that the indirect approach will typically lead to worse results when $m < n-1$. However, it is possible to establish a link between the indirect and direct procedures if the filtered instrument and regressor vectors of the RIV method are appropriately modified. In the sequel, we consider the system parametrization

$$
G_\mathrm{d}^*(q) = \frac{\sum_{i=0}^{m} N_{\mathrm{d},i}^*(q)\gamma_i^*}{A_\mathrm{d}^*(q)}
$$

with parameter vector $\boldsymbol{\rho}_\mathrm{d}^* = [\alpha_1^*, \ldots, \alpha_n^*, \gamma_0^*, \ldots, \gamma_m^*]^\top$, and where $N_{\mathrm{d},i}^*(q)$ are the numerator polynomials of the DT equivalents of $p^i/A_\mathrm{c}^*(p)$. This novel parametrization is such that its numerator coefficients γ_i^* correspond exactly to those of the numerator polynomial of its CT equivalent. By leveraging the state-space description of $p^i/A_\mathrm{c}^*(p)$, it is possible to show that

$$
N_{\mathrm{d},i}^*(q) = \alpha_n^*\mathbf{e}_{i+1}^\top\mathrm{adj}(q\mathbf{I} - e^{\mathbf{A}_\mathrm{c}^*h})(\mathbf{I} - e^{\mathbf{A}_\mathrm{c}^*h})\mathbf{e}_1,
$$

where $\mathrm{adj}(\cdot)$ denotes the adjugate matrix, $\mathbf{e}_j$ is the jth column of the identity matrix of appropriate size, and $\mathbf{A}_\mathrm{c}^*$ is the state matrix of the CT equivalent of $G_\mathrm{d}^*(q)$, written as

$$
\mathbf{A}_\mathrm{c}^* = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ \frac{-1}{a_n^*} & \frac{-a_1^*}{a_n^*} & \frac{-a_2^*}{a_n^*} & \ldots & \frac{-a_{n-1}^*}{a_n^*} \end{bmatrix}.
$$

Note that the $N_{\mathrm{d},i}^*(q)$ polynomials are in general of order $n-1$ due to Proposition 2.1, and they satisfy $N_{\mathrm{d},i}^*(1) = 0$ for $i > 0$. The vector $\boldsymbol{\rho}_\mathrm{d}^*$ is related to $\boldsymbol{\theta}_\mathrm{c}^*$ via $\mathbf{f}$ for the denominator

coefficients, and via the identity map for the numerator coefficients. This transformation will be denoted as $\tilde{\mathbf{f}}$ and it is bijective in the same domain as $\mathbf{f}$ in Definition 2.1.

Definition 4.1 (Adapted RIV and SRIV estimators): We define the Adapted RIV (ARIV) estimator with noise model iterations given by $\boldsymbol{\eta}_{j+1} = \tilde{\mathbf{g}}_{\mathrm{d}}(\boldsymbol{\rho}_{\mathrm{d},j})$, with

$$
\tilde{\mathbf{g}}_{\mathrm{d}}(\boldsymbol{\rho}_{\mathrm{d},j}) = \arg\min_{\boldsymbol{\eta}} \sum_{k=1}^{N} \left[\frac{D(q,\boldsymbol{\eta})}{C(q,\boldsymbol{\eta})} \right.
$$
$$
\left. \times \left(y(kh) - \frac{\sum_{i=0}^{m} N_{\mathrm{d},i}(q,\boldsymbol{\rho}_{\mathrm{d},j})\gamma_{i,j}}{A_{\mathrm{d}}(q,\boldsymbol{\rho}_{\mathrm{d},j})} u(kh) \right) \right]^2, \quad (10)
$$

and system model iterations (3) with filtered output (4), but where the filtered regressor and instrument vectors are, respectively, given by

$$
\boldsymbol{\varphi}_{\mathrm{d},f}(kh,\boldsymbol{\rho}_{\mathrm{d},j}) = \frac{D_{j+1}(q)}{C_{j+1}(q)} \left[\frac{-q}{A_{\mathrm{d},j}(q)} y(kh), \dots, \frac{-q^n}{A_{\mathrm{d},j}(q)} y(kh), \right.
$$
$$
\left. \frac{N_{\mathrm{d},0,j}(q)}{A_{\mathrm{d},j}(q)} u(kh), \dots, \frac{N_{\mathrm{d},m,j}(q)}{A_{\mathrm{d},j}(q)} u(kh) \right]^\top, \quad (11)
$$
$$
\hat{\boldsymbol{\varphi}}_{\mathrm{d},f}(kh,\boldsymbol{\rho}_{\mathrm{d},j}) = \frac{D_{j+1}(q)}{C_{j+1}(q)} \left[\frac{-M_{\mathrm{d},1,j}(q)}{A_{\mathrm{d},j}^2(q)} u(kh), \dots, \right.
$$
$$
\left. \frac{-M_{\mathrm{d},n,j}(q)}{A_{\mathrm{d},j}^2(q)} u(kh), \frac{N_{\mathrm{d},0,j}(q)}{A_{\mathrm{d},j}(q)} u(kh), \dots, \frac{N_{\mathrm{d},m,j}(q)}{A_{\mathrm{d},j}(q)} u(kh) \right]^\top,
$$
$$
(12)
$$

with $N_{\mathrm{d},l,j}(q)$ and $M_{\mathrm{d},r,j}(q)$ ($l = 0, \dots, m$, $r = 1, \dots, n$) being, respectively, the numerator polynomials of the DT equivalents of $p^l / A_{\mathrm{c}}(p, \tilde{\mathbf{f}}(\boldsymbol{\rho}_{\mathrm{d},j}))$ and

$$
[p, \dots, p^n, 0, \dots, 0] \frac{\partial \tilde{\mathbf{f}}(\boldsymbol{\rho}_{\mathrm{d}})}{\partial \alpha_r} \bigg|_{\boldsymbol{\rho}_{\mathrm{d}} = \boldsymbol{\rho}_{\mathrm{d},j}} \frac{B_{\mathrm{c}}(p, \tilde{\mathbf{f}}(\boldsymbol{\rho}_{\mathrm{d},j}))}{A_{\mathrm{c}}^2(p, \tilde{\mathbf{f}}(\boldsymbol{\rho}_{\mathrm{d},j}))}.
$$

Upon convergence in iterations, the resulting DT model is given by $\left(\sum_{i=0}^{m} N_{\mathrm{d},i}(q,\bar{\boldsymbol{\rho}}_{\mathrm{d}})\bar{\gamma}_i \right) / A_{\mathrm{d}}(q,\bar{\boldsymbol{\rho}}_{\mathrm{d}})$, where $\bar{\boldsymbol{\rho}}_{\mathrm{d}} = \lim_{j\to\infty} \boldsymbol{\rho}_{\mathrm{d},j}$ and $\bar{\gamma}_i = \lim_{j\to\infty} \gamma_{i,j}$. The Adapted SRIV (ASRIV) estimator is defined as the ARIV estimator but with a fixed noise model $C_j(q) = D_j(q) = 1$.

Theorem 4.2: Consider the ARIV algorithm in Definition 4.1 and the RIVC estimator in Section III, and assume that upon convergence in iterations for a fixed sample size, their modified normal matrices are non-singular. If Assumptions 4.1 and 4.2 are satisfied, then the ARIV and RIVC estimators have the same number of limiting points and they are linked by $\bar{\boldsymbol{\theta}}_{\mathrm{c}} = \tilde{\mathbf{f}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}})$, where $\bar{\boldsymbol{\theta}}_{\mathrm{c}}$ and $\bar{\boldsymbol{\rho}}_{\mathrm{d}}$ are the limiting point(s) of the RIVC and ARIV methods, respectively.

Proof: At convergence, a limiting point $\bar{\boldsymbol{\rho}}_{\mathrm{d}}$ of the ARIV iterations must satisfy (5), however now

$$
y_{\mathrm{d},f}(kh,\bar{\boldsymbol{\rho}}_{\mathrm{d}}) - \boldsymbol{\varphi}_{\mathrm{d},f}^\top(kh,\bar{\boldsymbol{\rho}}_{\mathrm{d}})\bar{\boldsymbol{\rho}}_{\mathrm{d}}
$$
$$
= \frac{D(q,\tilde{\mathbf{g}}_{\mathrm{d}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}}))}{C(q,\tilde{\mathbf{g}}_{\mathrm{d}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}}))} \left(y(kh) - \frac{\sum_{i=0}^{m} N_{\mathrm{d},i}(q,\bar{\boldsymbol{\rho}}_{\mathrm{d}})\bar{\gamma}_i}{A_{\mathrm{d}}(q,\bar{\boldsymbol{\rho}}_{\mathrm{d}})} u(kh) \right)
$$
$$
= \frac{D(q,\mathbf{g}_{\mathrm{c}}(\tilde{\mathbf{f}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}})))}{C(q,\mathbf{g}_{\mathrm{c}}(\tilde{\mathbf{f}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}})))} \left(y(kh) - \frac{B_{\mathrm{c}}(p,\tilde{\mathbf{f}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}}))}{A_{\mathrm{c}}(p,\tilde{\mathbf{f}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}}))} u(kh) \right).
$$

After some algebraic manipulations, it can be noted that the entries of $\hat{\boldsymbol{\varphi}}_{\mathrm{d},f}(kh,\bar{\boldsymbol{\rho}}_{\mathrm{d}})$ satisfy

$$
\frac{-M_{\mathrm{d},r}(q,\bar{\boldsymbol{\rho}}_{\mathrm{d}})}{A_{\mathrm{d}}^2(q,\bar{\boldsymbol{\rho}}_{\mathrm{d}})} u(kh) = \frac{\partial}{\partial \alpha_r} \frac{B_{\mathrm{c}}(p,\tilde{\mathbf{f}}(\boldsymbol{\rho}_{\mathrm{d}}))}{A_{\mathrm{c}}(p,\tilde{\mathbf{f}}(\boldsymbol{\rho}_{\mathrm{d}}))} \bigg|_{\boldsymbol{\rho}_{\mathrm{d}} = \bar{\boldsymbol{\rho}}_{\mathrm{d}}}
$$

for $r = 1, \dots, n$, and

$$
\frac{N_{\mathrm{d},l}(q,\bar{\boldsymbol{\rho}}_{\mathrm{d}})}{A_{\mathrm{d}}(q,\bar{\boldsymbol{\rho}}_{\mathrm{d}})} u(kh) = \frac{\partial}{\partial \gamma_l} \frac{\sum_{i=0}^{m} N_{\mathrm{d},i}(q,\boldsymbol{\rho}_{\mathrm{d}})\gamma_i}{A_{\mathrm{d}}(q,\boldsymbol{\rho}_{\mathrm{d}})} u(kh) \bigg|_{\boldsymbol{\rho}_{\mathrm{d}} = \bar{\boldsymbol{\rho}}_{\mathrm{d}}}
$$
$$
= \frac{\partial}{\partial \gamma_l} \frac{B_{\mathrm{c}}(p,\tilde{\mathbf{f}}(\boldsymbol{\rho}_{\mathrm{d}}))}{A_{\mathrm{c}}(p,\tilde{\mathbf{f}}(\boldsymbol{\rho}_{\mathrm{d}}))} \bigg|_{\boldsymbol{\rho}_{\mathrm{d}} = \bar{\boldsymbol{\rho}}_{\mathrm{d}}}
$$

for $l = 1, \dots, m$. Therefore, by the chain rule for gradients (cf. (9)),

$$
\hat{\boldsymbol{\varphi}}_{\mathrm{d},f}(kh,\bar{\boldsymbol{\rho}}_{\mathrm{d}}) =
$$
$$
\frac{D(q,\mathbf{g}_{\mathrm{c}}(\tilde{\mathbf{f}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}})))}{C(q,\mathbf{g}_{\mathrm{c}}(\tilde{\mathbf{f}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}})))} \frac{\partial \tilde{\mathbf{f}}}{\partial \boldsymbol{\rho}_{\mathrm{d}}} \bigg|_{\boldsymbol{\rho}_{\mathrm{d}} = \bar{\boldsymbol{\rho}}_{\mathrm{d}}} \frac{\partial}{\partial \boldsymbol{\theta}_{\mathrm{c}}} \frac{B_{\mathrm{c}}(p,\boldsymbol{\theta}_{\mathrm{c}})}{A_{\mathrm{c}}(p,\boldsymbol{\theta}_{\mathrm{c}})} u(kh) \bigg|_{\boldsymbol{\theta}_{\mathrm{c}} = \tilde{\mathbf{f}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}})}.
$$

The rest of the proof follows the same lines as the proof of Theorem 4.1 after (9) and is therefore omitted. ∎

Corollary 4.2: Under the same assumptions as in Theorem 4.2, the ASRIV and the SRIVC estimators have the same number of limiting points, and they are linked by $\bar{\boldsymbol{\theta}}_{\mathrm{c}} = \tilde{\mathbf{f}}(\bar{\boldsymbol{\rho}}_{\mathrm{d}})$, where $\bar{\boldsymbol{\theta}}_{\mathrm{c}}$ and $\bar{\boldsymbol{\rho}}_{\mathrm{d}}$ are the limiting point(s) of the SRIVC and ASRIV methods, respectively.

Proof: Direct from fixing $C_j(q) = D_j(q) = 1$ in the proof for Theorem 4.2. ∎

In summary, the ARIV method in Definition 4.1 provides a DT estimator whose CT equivalent model has a fixed relative degree $n - m$. This implementation has the advantage of enforcing smoothness properties of the CT system directly into the DT estimate, without the need of additional optimization steps such as in [3]. The proposed method closes the gap between DT and CT refined instrumental variable methods, since the CT equivalent of the ARIV estimator corresponds exactly to the RIVC estimator for a finite sample size.

B. The Adapted RIVC estimator

The analog problem of the previous subsection for CT system identification consists of directly identifying a CT system whose DT equivalent has relative degree $n - m > 1$. This problem arises if the system is known to have a fixed number of sample time-delays. Instead of computing a time-delayed CT system, which typically requires iterative non-convex optimization steps, a computationally-cheap approach that directly imposes sample time-delays in the estimated transfer function can be obtained by an adapted form of the RIVC estimator, similar to Definition 4.1. In this case, we must parameterize the CT system as

$$
G_{\mathrm{c}}^*(p) = \frac{\sum_{i=0}^{m} N_{\mathrm{c},i}^*(p)\nu_i^*}{A_{\mathrm{c}}^*(p)}
$$

with parameter vector $\boldsymbol{\rho}_{\mathrm{c}}^* = [a_1^*, \dots, a_n^*, \nu_0^*, \dots, \nu_m^*]^\top$, and where $N_{\mathrm{c},i}^*(p)$ are the numerator polynomials of the CT equivalent of $q^i / A_{\mathrm{d}}^*(q)$. The noise model is computed from an ARMA estimation step similar to (10), while the system parameter vector $\boldsymbol{\rho}_{\mathrm{c}}^*$ can be estimated from a refined

Method	α_1	α_2	α_3	α_4	β_0	β_1	β_2	β_3	d_1	c_1
SRIV	4.06e-5	3.05e-4	3.16e-4	4.81e-5	9.61e-4	4.25e-3	3.40e-3	5.59e-4	-	-
ASRIV	2.46e-5	2.50e-4	3.64e-4	6.71e-5	9.69e-6	1.05e-4	8.79e-5	1.56e-5	-	-
RIV	3.97e-5	3.00e-4	3.15e-4	4.81e-5	9.58e-4	4.25e-3	3.42e-3	5.60e-4	6.83e-5	1.13e-4
ARIV	2.23e-5	2.02e-4	2.65e-4	4.55e-5	4.84e-6	5.32e-5	4.39e-5	7.91e-6	6.82e-5	1.13e-4

instrumental variable procedure with filtered regressor and instrument vectors given by CT variants of (11) and (12). We omit the full expressions due to space constraints.

V. SIMULATION EXAMPLE

To provide an example of the theoretical results, we consider the Rao-Garnier benchmark system [22]

$$G_\text{c}^*(p) = \frac{-4p + 1}{0.000625p^4 + 0.003125p^3 + 0.255p^2 + 0.26p + 1},$$

with noise model $H^*(q) = (1 + 0.4q^{-1})/(1 - 0.7q^{-1})$. Considering a sampling period $h = 0.05$[s], the DT equivalent of $G_\text{c}^*(p)$ has a parameter vector given by $\boldsymbol{\theta}_\text{d}^* = [-1.069, 0.546, -1.979, 1.65, 0.991, 2.665, -2.241, -1.268]^\top$. The input is a ZOH-interpolated multisine with angular frequencies $\omega = 1, 1.9, 2.1, 18, 22$ [rad/s], and the white noise $e(kh)$ filtered by $H^*(q)$ has variance equal to 6, which gives a signal-to-noise ratio of approximately 26 [dB]. If a DT model is sought, then one can use the RIV and SRIV estimators described in Section III. However, DT identification methods that offer more flexibility than these two are the ASRIV and ARIV methods, since these algorithms can adjust for the smoothness of the CT step response via the relative degree enforcement of the CT equivalent transfer function estimate.

Table I shows the mean square error (MSE) of each parameter when performing 500 Monte Carlo runs of $N = 10^4$ samples each, with varying noise realizations. While being competitive or marginally better in terms of the MSE of the denominator parameters, the ASRIV and ARIV methods can identify the numerator parameters with at least one order of magnitude less of MSE compared to their non-adapted counterparts. We note that this improvement has been achieved using only discrete-time tools; the ASRIV and ARIV methods are in fact equal at convergence to the DT equivalents of the estimates obtained from the SRIVC and RIVC methods, respectively.

VI. CONCLUSIONS

In this paper, we proved that the CT equivalent of the RIV estimator shares the same limiting points with the RIVC estimator for relative degrees one and zero. Such relation fails for $m < n-1$ since the indirect approach generally does not produce an estimate with the desired model structure. We also introduced the Adapted RIV estimator, which provides the correct relative degree in its CT equivalent. Under a similar logic, it is shown that it is also possible to adapt the RIVC estimator such that its DT equivalent has relative degree greater than one by construction. These adapted estimators provide the missing link between the refined instrumental variable methods in the DT and CT domain.

REFERENCES

[1] L. Ljung, *System Identification: Theory for the User,* 2nd ed. Prentice-Hall, 1999.

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

[3] R. A. González, C. R. Rojas, and J. S. Welsh, "An asymptotically optimal indirect approach to continuous-time system identification," in *57th IEEE Conference on Decision and Control (CDC)*, pp. 638–643, 2018.

[4] P. C. Young, "Some observations on instrumental variable methods of time-series analysis," *International Journal of Control*, vol. 23, no. 5, pp. 593–612, 1976.

[5] H. Garnier, "Direct continuous-time approaches to system identification. Overview and benefits for practical applications," *European Journal of Control*, vol. 24, pp. 50–62, 2015.

[6] L. Ljung, "Experiments with identification of continuous time models," in *15th IFAC Symposium on System Identification,* Saint Malo, France, vol. 42, pp. 1175–1180, 2009.

[7] P. C. Young and A. J. Jakeman, "Refined instrumental variable methods of recursive time-series analysis. Part III, Extensions," *International Journal of Control*, vol. 31, no. 4, pp. 741–764, 1980.

[8] H. Garnier, M. Gilson, H. Muller, and F. Chen, "A new graphical user interface for the CONTSID toolbox for Matlab," *IFAC-PapersOnLine*, vol. 54, no. 7, pp. 397–402, 2021.

[9] P. C. Young, *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner,* 2nd Edition. Springer, 2012.

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

[11] V. Laurain, M. Gilson, R. Tóth, and H. Garnier, "Refined instrumental variable methods for identification of LPV Box–Jenkins models," *Automatica*, vol. 46, no. 6, pp. 959–967, 2010.

[12] B. Ni, M. Gilson, and H. Garnier, "Refined instrumental variable method for Hammerstein–Wiener continuous-time model identification," *IET Control Theory & Applications*, vol. 7, no. 9, pp. 1276–1286, 2013.

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

[14] P. C. Young, "Refined instrumental variable estimation: Maximum Likelihood optimization of a unified Box–Jenkins model," *Automatica*, vol. 52, pp. 35–46, 2015.

[15] J. I. Yuz and G. C. Goodwin, *Sampled-Data Models for Linear and Nonlinear Systems*. Springer, 2014.

[16] I. Kollar, G. Franklin, and R. Pintelon, "On the equivalence of z-domain and s-domain models in system identification," in *IEEE Instrumentation and Measurement Technology Conference*, vol. 1, pp. 14–19, 1996.

[17] W. Rudin, *Principles of Mathematical Analysis,* 3rd Edition. McGraw-Hill, 1976.

[18] S. Pan, R. A. González, J. S. Welsh, and C. R. Rojas, "Consistency analysis of the Simplified Refined Instrumental Variable method for Continuous-time systems," *Automatica*, vol. 113, Art. 108767, 2020.

[19] R. A. González, C. R. Rojas, S. Pan, and J. S. Welsh, "Theoretical and practical aspects of the convergence of the SRIVC estimator for over-parameterized models," *Automatica*, vol. 142, Art. 110355, 2022.

[20] K. J. Åström and B. Wittenmark, "Computer Controlled Systems: Theory and Design," *Prentice-Hall*, 1984.

[21] H. Garnier and P. C. Young, "The advantages of directly identifying continuous-time transfer function models in practical applications," *International Journal of Control*, vol. 87, no. 7, pp. 1319–1338, 2014.

[22] G. P. Rao and H. Garnier, "Numerical illustrations of the relevance of direct continuous-time model identification," in *15th IFAC World Congress,* Barcelona, Spain, vol. 35, pp. 133–138, 2002.