

An asymptotically optimal indirect approach to continuous-time system identification

Rodrigo A. González, Cristian R. Rojas and James S. Welsh

Abstract—The indirect approach to continuous-time system identification consists in estimating continuous-time models by first determining an appropriate discrete-time model. For a zero-order hold sampling mechanism, this approach usually leads to a transfer function estimate with relative degree 1, independent of the relative degree of the strictly proper real system. In this paper, a refinement of these methods is developed. Inspired by the indirect prediction error method, we propose an estimator that enforces a fixed relative degree in the continuous-time transfer function estimate, and show that the estimator is consistent and asymptotically efficient. Extensive numerical simulations are put forward to show the performance of this estimator when contrasted with other indirect and direct methods for continuous-time system identification.

Index Terms—System identification; Continuous-time systems; Parameter estimation; Sampled data.

I. INTRODUCTION

System identification deals with the problem of estimating adequate models of dynamical systems from input-output data. The methods developed over the years in this field have seen applications in many areas of science and engineering, and comprehensive literature has been written on the subject [1]–[3].

When postulating a mathematical model for describing a dynamical system based on sampled data, one must decide between obtaining a discrete-time (DT), or a continuous-time (CT) model. In a predominantly digital era, DT system identification has been studied thoroughly (see, e.g., [1], [2]). Nevertheless, interest in CT models still persists due to its advantages over discrete-time. For example, grey-box modelling [4], which is commonly based on physical principles and conservation laws, is naturally suited for continuous time, as the parameters can usually be better interpreted in this domain. Also, CT models are known to have more intuitive dynamics, and they do not depend on a sampling period.

In CT system identification there are two main approaches, namely the direct and indirect approaches. For direct CT system identification, a CT model is obtained directly from the sampled data. The main difficulty present in the direct methods is the handling of derivatives, as they are not immediately available from discrete data points without amplifying

noise [5]. To effectively deal with this issue, many well known methods have been proposed [6], with success in real applications [7]. On the other hand, indirect methods for CT modelling first determine a suitable DT model via DT system identification methods like the Prediction Error Methods (PEM) or Maximum Likelihood (ML) [2], and then transform this model into a CT equivalent model. Evidence has been shown regarding the advantages of direct over indirect CT model identification [8], although with a precise initialisation of PEM, the approaches seem comparable for certain sampling periods [9].

Even though the indirect approach seems easy to implement, as there is much theory and literature concerning DT system identification, there are reasons why this approach is not always recommended. First, it may suffer from numerical inaccuracies at fast sampling, and requires a precise initialisation. In addition, it is not possible to select the desired numerator order of the CT model, as the estimated DT model will generally lead to a CT model with relative degree 1 in the case of sampling by a zero-order hold mechanism. Hence, an unnecessarily complex model structure is being estimated, which leads to a loss in accuracy according to the parsimony principle [1].

In this paper, we introduce a method that optimally imposes a desired relative degree in the indirect approach to continuous-time system identification. Based on Indirect PEM [10], we prove that the proposed estimator is a consistent and asymptotically efficient estimator of the system's true parameter vector. Extensive numerical simulations show that the new method imposes the correct relative degree, while improving the statistical properties of the transfer function estimate, and achieves a performance that compares favourably against both standard direct and indirect approaches.

The remainder of this paper is organised as follows. In Section II the problem is formulated. Section III provides an introduction to the indirect approach for CT system identification. In Section IV we derive an estimator that optimally enforces the desired relative degree for the indirect approach, and determine its properties. Section V illustrates the method with extensive numerical examples. Finally, conclusions are drawn in Section VI.

This work was supported by the Swedish Research Council under contract number 2016-06079 (NewLEADS).

R. A. González and C. R. Rojas are with the Department of Automatic Control and ACCESS Linnaeus Centre, KTH Royal Institute of Technology, 10044 Stockholm, Sweden (e-mails: grodrigo@kth.se; crro@kth.se).

James S. Welsh is with the School of Electrical Engineering and Computer Science, University of Newcastle, Australia (e-mail: james.welsh@newcastle.edu.au).

II. PROBLEM FORMULATION

Consider a linear time-invariant, causal, stable, single input single output, CT system

$$y(t) = G_0(\rho)u(t) \quad (1)$$

$$= \frac{\beta_{n-r}\rho^{n-r} + \beta_{n-r-1}\rho^{n-r-1} + \cdots + \beta_1\rho + \beta_0}{\rho^n + \alpha_{n-1}\rho^{n-1} + \cdots + \alpha_1\rho + \alpha_0}u(t),$$

where ρ is the Heaviside operator, i.e., $\rho g(t) := dg(t)/dt$, and r is the relative degree of the system. We denote $\theta_c^0 := [\beta_{n-r} \ \cdots \ \beta_0 \ \alpha_{n-1} \ \cdots \ \alpha_0]^\top$ as the true CT system parameter vector.

Suppose that the input-output signals are sampled with period h and the resulting output is contaminated by an additive zero-mean white noise sequence $\{e(kh)\}_{k \in \mathbb{N}}$ of variance σ^2 . That is,

$$y_m(kh) = y(kh) + e(kh), \quad k \in \mathbb{N}. \quad (2)$$

The goal of CT system identification using sampled data is to obtain a CT transfer function estimate for $G_0(\rho)$, given N discrete input-output data measurements $\{u(kh), y_m(kh)\}_{k=1}^N$ and knowledge about the physical characteristics of the system, or the intersample behaviour. In this paper, we assume that the input is a piecewise constant signal between samples (i.e., the input has a zero-order hold behaviour [11]).

For obtaining a model of $G_0(\rho)$, a simple way to proceed is to identify the zero-order hold equivalent model given the input and output data measurements by using standard PEM in the DT domain, and then return to the CT domain via zero-order hold equivalences. Although this procedure has good statistical properties, it does not impose relative degree constraints in the CT domain. Our goal is to optimally impose this constraint, which should lead to a statistically improved estimate of $G_0(\rho)$.

III. THE INDIRECT APPROACH TO CONTINUOUS-TIME SYSTEM IDENTIFICATION

One approach to identifying a CT system is to first estimate the DT model given the input and output data samples, and then translate this model into continuous time. This is called the indirect approach, since it relies on DT system identification theory instead of obtaining immediately a CT model using CT system identification methods.

Much literature has been written regarding the first step of the indirect approach [1], [2]. The theoretically optimal solution is to apply the maximum likelihood method. This method is known to give consistent and asymptotically efficient estimates under very general conditions. Under the assumption that the additive white noise is Gaussian, the ML method is equivalent to PEM, which is one of the most celebrated parametric methods, and is available in the MATLAB System Identification Toolbox [12].

If a CT model for (1) is required, we should propose a DT model structure of the form

$$H(z) = \frac{b_{n-1}z^{n-1} + b_{n-2}z^{n-2} + \cdots + b_1z + b_0}{z^n + a_{n-1}z^{n-1} + \cdots + a_1z + a_0}. \quad (3)$$

If we define $\theta_d = [b_{n-1} \ \cdots \ b_0 \ a_{n-1} \ \cdots \ a_0]^\top$ and denote \hat{y} as the model's output, then the ML estimate for θ_d is

$$\hat{\theta}_d = \arg \min_{\theta_d} \frac{1}{\sigma^2} \sum_{k=1}^N \|y_m(kh) - \hat{y}(kh, \theta_d)\|^2.$$

The next step is to transform this DT transfer function into an adequate CT model. This can be done in several ways. For example, the well-known Tustin transformation can be applied on the DT transfer function estimate by letting

$$z = \frac{sh + 2}{2 - sh},$$

as reported in [13]. If it is assumed that the CT input signal is piecewise constant, the most natural mapping (used in e.g. [9], [14]), is the zero-order hold sampling equivalence

$$G_0(s) = s\mathcal{L} \left\{ \mathcal{Z}^{-1} \left\{ \frac{H_0(z)}{1 - z^{-1}} \right\} \Big|_{kh=t} \right\}, \quad (4)$$

where \mathcal{L} and \mathcal{Z} denote the Laplace and Z transforms, respectively. This mapping is known to be a troublesome part of the indirect approach, as it can be ill conditioned and its uniqueness depends on a correct choice of the sampling period.

In both of these mappings, the resulting CT model can have numerator parameters that exceed the desired numerator degrees. This is generally the case when the relative degree of $G_0(s)$ is greater than 1. This problem contributes to poor accuracy and high standard deviations at high frequencies. One very simple way of treating this issue is setting to zero the numerator coefficients which should be zero, but this is not the best way of taking care of the information [9].

IV. OPTIMAL ENFORCEMENT OF RELATIVE DEGREE

In this section we develop an indirect estimator for the CT parameter vector θ_c^0 that renders a CT transfer function estimate $\hat{G}(s)$ of a desired relative degree r . For this matter, we first focus on the PEM estimator of the zero-order hold equivalent model of $G_0(s)$.

For simplicity, we assume that the correct model order has been found. A model with structure (3) is obtained by PEM, and the covariance matrix of $\hat{\theta}_d$ is also estimated. We know that the CT zero-order hold equivalent of this model is in general given by

$$\hat{G}(s) = \frac{\hat{\beta}_{n-1}s^{n-1} + \hat{\beta}_{n-2}s^{n-2} + \cdots + \hat{\beta}_1s + \hat{\beta}_0}{s^n + \hat{\alpha}_{n-1}s^{n-1} + \cdots + \hat{\alpha}_1s + \hat{\alpha}_0}.$$

Define $\hat{\theta}_c = [\hat{\beta}_{n-1} \ \cdots \ \hat{\beta}_0 \ \hat{\alpha}_{n-1} \ \cdots \ \hat{\alpha}_0]^\top$, and denote by θ_d^0 the true DT parameter vector. The parameters in $\hat{\theta}_c$ are related to $\hat{\theta}_d$ by the zero-order hold equivalence equations that can be derived by using (4) and comparing coefficients. This relation is a nonlinear mapping $f : \hat{\theta}_c \rightarrow f(\hat{\theta}_c) = \hat{\theta}_d$, which is differentiable almost everywhere. Hence, the following asymptotic relationship is valid for the covariance matrices of $\hat{\theta}_d$ and $\hat{\theta}_c$:

$$\begin{aligned} \Sigma_{\hat{\theta}_d} &= \mathbb{E}\{(\hat{\theta}_d - \theta_d^0)(\hat{\theta}_d - \theta_d^0)^\top\} \\ &\approx \mathbb{E}\{J(\hat{\theta}_c - \theta_c^0)(\hat{\theta}_c - \theta_c^0)^\top J^\top\} \\ &= J\Sigma_{\hat{\theta}_c}J^\top, \end{aligned} \quad (5)$$

where J is the Jacobian matrix of f evaluated at the naive estimation of θ_c^0 , that is, throwing away the high order coefficients of the numerator of \hat{G} which should be zero¹.

Now, we propose to find an appropriate projection of $\hat{\theta}_c$ into a proper subspace of the parameter space that yields the desired relative degree. This subspace is simply the one formed by all vectors with the first $r - 1$ elements set to zero. Hence, we decide to study the following problem:

$$\begin{aligned} \tilde{\theta}_c &= \arg \min_{\theta} \frac{1}{2} (\hat{\theta}_c - \theta)^\top \Sigma_{\hat{\theta}_c}^{-1} (\hat{\theta}_c - \theta) \\ \text{s.t. } & [\mathbf{I}_{r-1} \quad 0] \theta = 0, \end{aligned} \quad (6)$$

where \mathbf{I}_{r-1} is the identity matrix of dimension $r - 1$, 0 is the null matrix of appropriate dimensions, and $\Sigma_{\hat{\theta}_c}^{-1} = J^\top \Sigma_{\hat{\theta}_d}^{-1} J$. The optimisation problem in this context can be interpreted as an application of the Indirect PEM [10].

By Lagrange multiplier theory, the optimization problem in (6) is equivalent to calculating, for a suitable λ ,

$$\tilde{\theta}_c = \arg \min_{\theta} \frac{1}{2} (\hat{\theta}_c - \theta)^\top \Sigma_{\hat{\theta}_c}^{-1} (\hat{\theta}_c - \theta) + \lambda^\top [\mathbf{I}_{r-1} \quad 0] \theta.$$

Partitioning $\Sigma_{\hat{\theta}_c}$ appropriately (dropping the subindex for simplicity), and differentiating with respect to θ we obtain

$$\tilde{\theta}_c = \hat{\theta}_c - \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{r-1} \\ 0 \end{bmatrix} \lambda, \quad (7)$$

Imposing the constraints from (6) we obtain $\lambda = \Sigma_{11}^{-1} [\mathbf{I}_{r-1} \quad 0] \hat{\theta}_c$. If we denote by C the Cholesky factorization matrix of $\Sigma_{\hat{\theta}_c}$ [15] (i.e., a lower triangular matrix with positive diagonal entries such that $\Sigma_{\hat{\theta}_c} = CC^\top$) we can write (7) as

$$\begin{aligned} \tilde{\theta}_c &= \hat{\theta}_c - \Sigma \begin{bmatrix} \mathbf{I}_{r-1} \\ 0 \end{bmatrix} \Sigma_{11}^{-1} [\mathbf{I}_{r-1} \quad 0] \hat{\theta}_c \\ &= \hat{\theta}_c - \begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{r-1} \\ 0 \end{bmatrix} C_{11}^{-1} [\mathbf{I}_{r-1} \quad 0] \hat{\theta}_c \\ &= C \begin{bmatrix} 0_{r-1} & 0^\top \\ 0 & \mathbf{I}_{2n-r+1} \end{bmatrix} C^{-1} \hat{\theta}_c. \end{aligned} \quad (8)$$

That is, the estimator (8) can be seen as an \mathcal{L}^2 best approximation to the PEM CT estimate $\hat{\theta}_c$ that imposes the desired relative degree. Note that this estimator relies on PEM giving a good initial estimate of the continuous-time model parameters.

A. Properties

We briefly present the most important properties of estimator (8) in the following theorems.

Theorem 4.1: Consider the system described by (1) and (2), where $\{e(kh)\}_{k=1}^N$ is a Gaussian white noise sequence. Assume that the sampling frequency $2\pi/h$ is larger than twice the largest imaginary part of the s -domain poles and there is no delay in the real system². Then, the estimator (8) is a consistent and asymptotically efficient estimator of the

¹Note that the standard PEM estimate could have also been used for this evaluation, and the asymptotic relation still holds.

²These conditions can be relaxed, as long as the sampling frequency is such that the $z \rightarrow s$ transformation is well defined.

real vector parameter θ_c^0 , provided the DT model set (with the chosen relative degree) contains the real system.

Proof: Under the Gaussian noise assumption, the DT PEM estimate can be interpreted as the ML estimate. Under the proposed sampling frequency, the $z \rightarrow s$ transformation is unique for sufficiently large N [16]. Hence, by the invariance principle of ML estimators [17], the CT equivalent of the system's parameters is also an ML estimate.

To prove the theorem, we only require that the assumptions in [10] are satisfied in this scenario, and then directly apply the results obtained in the cited contribution. First, note that the model structure given by this procedure contains the models with the desired relative degree (and the contention is proper if $r > 1$), with a linear mapping between parameter vectors given by the matrix

$$T := [0_{(2n-r+1) \times (r-1)} \quad \mathbf{I}_{2n-r+1}].$$

Furthermore, provided that the DT model set contains the real system, both structures give parameter identifiability. Also note that $\Sigma_{\hat{\theta}_c}$, obtained via (5), is a consistent estimate of the covariance matrix of $\hat{\theta}_c$. Hence, the results in [10] follow. Namely, the normalized estimation errors³ $\sqrt{N}(\hat{\theta}_c - \theta_c^0)/\sigma$ and $\sqrt{N}(\tilde{\theta}_c - \theta_c^0)/\sigma$ are asymptotically normally distributed with zero means and their asymptotic covariance matrices satisfy the relation

$$[\text{AsCov}(\tilde{\theta}_c - \theta_c^0)]^{-1} = T[\text{AsCov}(\hat{\theta}_c - \theta_c^0)]^{-1}T^\top. \quad (9)$$

Moreover, following the steps in [10, Section 3], the improved PEM estimate is \sqrt{N} -consistent and has the same asymptotic distribution as $\hat{\theta}_c$, thereby proving its asymptotic efficiency. ■

Remark 4.1: Note that by (8) and (9), the asymptotic covariances can be shown to satisfy the following properties:

- $\text{AsCov}(\tilde{\theta}_c - \theta_c^0, \hat{\theta}_c - \tilde{\theta}_c) = 0$,
- $\text{AsCov}(\tilde{\theta}_c - \theta_c^0) = \text{AsCov}(\hat{\theta}_c - \theta_c^0) - \text{AsCov}(\hat{\theta}_c - \tilde{\theta}_c)$.

Both of these claims follow by applying properties 10.5 and 10.6 from [18, Chapter 10] to this context. These properties imply that for sufficiently large N , the proposed estimator can only decrease the covariance of the estimated parameters compared to standard PEM. The asymptotic covariances satisfy a Pythagorean relation, as the PEM estimate is projected orthogonally on to the proper subspace where θ_c^0 lies.

Next, we establish that imposing a larger relative degree improves accuracy of the estimates, provided that the highest relative degree model structure contains the real system.

Theorem 4.2: Given $G_0(s)$ of order n and relative degree $r > 1$, consider CT candidate models of relative degree r_1 and r_2 and their improved PEM parameter vector estimates $\hat{\theta}_c^{r_1}$ and $\hat{\theta}_c^{r_2}$ respectively. If $r_1 < r_2 \leq r$, then their asymptotic covariance matrices satisfy $\text{AsCov}(\hat{\theta}_c^{r_2}) \preceq \text{AsCov}(\hat{\theta}_c^{r_1})$.

Proof: The proof follows by applying [19, Theorem 2]. Details are omitted due to length restrictions. ■

³For simplicity, we discard the enforced zero elements of $\tilde{\theta}_c$ in this analysis. We assume that the vector θ_c^0 has the appropriate dimension, where its zero elements have been discarded in the expression $\sqrt{N}(\theta_c - \theta_c^0)/\sigma$.

Remark 4.2: The relative degree of the CT system is not always known. In some practical applications, physical knowledge about the system can give intuition. In addition, statistical measures can be used such as the coefficient of determination [7], or the Young Information Criterion [20].

V. MONTE CARLO SIMULATION STUDIES

We now study the performance of the proposed estimator under a series of experiments.

The system considered is the Rao-Garnier system [8], which is a linear fourth-order non-minimum phase system with complex poles that has been tested in many publications (see e.g. [7], [21], [22]) of CT system identification:

$$G_0(s) = \frac{-6400s + 1600}{s^4 + 5s^3 + 408s^2 + 416s + 1600}.$$

This system is interesting since it has two damped oscillatory modes at 2 and 20 rad/sec with damping of 0.1 and 0.25 respectively, and has a non-minimum phase zero at $s = 0.25$. It is known that this is a particularly difficult system to estimate by PEM/ML, since these methods may converge to a local minimum if they are not well initialised [21].

Three methods were compared: PEM, PEM with relative degree enforcement (labeled PEMrd (8)), and the simplified refined IV method for continuous-time systems (SRIVC), which is one of the most successful direct methods available, and has been suggested for general use in one of the most recent surveys on CT system identification [7]. Each method was tested in M different Monte Carlo runs, and they have been evaluated according to the average normalized square error of the system estimate

$$\text{MSE } \hat{G} = \frac{1}{M} \sum_{i=1}^M \frac{\|\hat{G}_i - G_0\|_2^2}{\|G_0\|_2^2},$$

the average normalized square error of the parameter vectors

$$\text{MSE } \hat{\theta} = \frac{1}{M} \sum_{i=1}^M \frac{\|\hat{\theta}_c^i - \theta_c^0\|_2^2}{\|\theta_c^0\|_2^2},$$

and the fit measure

$$\text{Fit} = \frac{100}{M} \sum_{i=1}^M \left[1 - \frac{\|\hat{y}^i - y\|_2}{\|y - \bar{y}\|_2} \right],$$

where y is the noise-free output sequence, \hat{y}^i is the simulated output sequence of the i -th estimated model, and \bar{y} is the average value of $\{y(kh)\}_{k=1}^N$.

We have run PEM using the MATLAB System Identification Toolbox [12] with the `oe` command, and assumed that the correct order of the system is known. The search algorithm for PEM has been initialised with the estimate given by the Null Space Fitting method⁴ [23]. We based PEMrd on the PEM estimate previously obtained. The required Jacobian matrix has been numerically calculated via finite differences, and the correct relative degree has been imposed. We have

⁴PEM initialised with the estimate from SRIVC (as in [9]) has also been tested with similar results.

used the command `d2c` of MATLAB in both cases. SRIVC has been implemented with the CONTSID toolbox version 7.2 for MATLAB [6, Chapter 9] with default initialisation, and has been set to estimate the model

$$\hat{G}(s) = \frac{\beta_1 s + \beta_0}{s^4 + \alpha_3 s^3 + \alpha_2 s^2 + \alpha_1 s + \alpha_0}.$$

A. Effect of the number of data points and sampling rate

In this study, we designed the input as a pseudorandom binary sequence (PRBS) of amplitude switches between 0 and 2. For the first input sequence, the number of stages of the shift register is $n = 10$ and the data length of the shortest interval is $p = 7$. Hence, a sequence of $N = 7161$ data points has been obtained. The noise is a zero-mean white Gaussian noise signal, where the variance has been set such that the signal-to-noise ratio (SNR) between the noiseless output sequence and the noise equals 10 dB.

Three Monte Carlo studies were performed with $M = 500$ runs of different noise realisations each one, for $h = 0.01, 0.05, 0.1$. The results are shown in Table I.

TABLE I
MONTE CARLO RESULTS FOR PRBS INPUT OF LENGTH $N = 7161$.

h	Method	MSE \hat{G}	MSE $\hat{\theta}$	Fit
0.01	PEM	$1.11 \cdot 10^{-4}$	$6.76 \cdot 10^{-5}$	98.97
	PEMrd	$0.73 \cdot 10^{-4}$	$4.28 \cdot 10^{-5}$	99.12
	SRIVC	$0.73 \cdot 10^{-4}$	$4.29 \cdot 10^{-5}$	99.12
0.05	PEM	$2.00 \cdot 10^{-3}$	$5.42 \cdot 10^{-4}$	98.97
	PEMrd	$1.41 \cdot 10^{-3}$	$3.93 \cdot 10^{-4}$	99.11
	SRIVC	$1.40 \cdot 10^{-3}$	$3.93 \cdot 10^{-4}$	99.11
0.1	PEM	$3.28 \cdot 10^{-3}$	$7.94 \cdot 10^{-4}$	98.94
	PEMrd	$1.91 \cdot 10^{-3}$	$4.80 \cdot 10^{-4}$	99.09
	SRIVC	$1.92 \cdot 10^{-3}$	$4.75 \cdot 10^{-4}$	99.09

In order to analyse the performance of each estimator under less data, we have set the number of stages to $n = 9$ and the data length of the shortest interval to $p = 3$, resulting in an input of $N = 1533$ data points. With the same SNR as the test above, the results for a Monte Carlo simulation of 500 runs for each sampling period are in Table II.

TABLE II
MONTE CARLO RESULTS FOR PRBS INPUT OF LENGTH $N = 1533$.

h	Method	MSE \hat{G}	MSE $\hat{\theta}$	Fit
0.01	PEM	$5.88 \cdot 10^{-4}$	$3.65 \cdot 10^{-4}$	97.78
	PEMrd	$4.03 \cdot 10^{-4}$	$2.06 \cdot 10^{-4}$	98.07
	SRIVC	$4.02 \cdot 10^{-4}$	$2.06 \cdot 10^{-4}$	98.07
0.05	PEM	$7.43 \cdot 10^{-4}$	$4.54 \cdot 10^{-4}$	97.79
	PEMrd	$4.32 \cdot 10^{-4}$	$2.79 \cdot 10^{-4}$	98.12
	SRIVC	$4.32 \cdot 10^{-4}$	$2.79 \cdot 10^{-4}$	98.12
0.1	PEM	$1.96 \cdot 10^{-2}$	$2.92 \cdot 10^{-3}$	97.74
	PEMrd	$1.08 \cdot 10^{-2}$	$2.18 \cdot 10^{-3}$	98.03
	SRIVC	$9.99 \cdot 10^{-3}$	$3.88 \cdot 10^{-3}$	97.98

Both Tables I and II show that the refined PEM estimator statistically improves the estimates given by PEM, and is a very competitive method against SRIVC, even for high frequency sampling. Note that under less data points, PEMrd still outperforms PEM for every sampling period, which

indicates that the asymptotic properties studied in Section IV can be observed in practical finite data cases as well.

Remark 5.1: In Tables I and II we have discarded cases where PEM gave estimates with one pole in the negative real axis. Fortunately this scenario is very uncommon (9 cases seen in 3000 runs). A similar phenomenon was observed in Table II for $h = 0.1$ in SRIVC, where 2 estimates gave negative fits. These simulations were not considered either.

B. Multisine input

A different input was also tested. We have taken 2000 data points of a multisine input given by the sum of sine waves of angular frequencies $\omega = 0.5, 1, 5, 8, 10, 12, 15, 20, 25, 30$, which was sampled by a zero order hold. The first 200 data points have been discarded. The standard deviation of the noise has been set to 0.1, and the median of 500 Monte Carlo runs of the normalized model error, parameter error, and fit have been obtained. The results are in Table III.

TABLE III

MONTE CARLO RESULTS FOR MULTISINE INPUT OF LENGTH $N = 2000$.

h	Method	$\frac{\ \hat{G} - G_0\ ^2}{\ G_0\ ^2}$	$\frac{\ \hat{\theta} - \theta_0\ ^2}{\ \theta_0\ ^2}$	Fit
0.01	PEM	$8.80 \cdot 10^{-5}$	$7.27 \cdot 10^{-5}$	99.43
	PEMrd	$1.35 \cdot 10^{-5}$	$4.44 \cdot 10^{-6}$	99.82
	SRIVC	$1.44 \cdot 10^{-1}$	$6.21 \cdot 10^{-2}$	80.85
0.02	PEM	$1.42 \cdot 10^{-4}$	$8.37 \cdot 10^{-5}$	99.27
	PEMrd	$2.29 \cdot 10^{-5}$	$1.01 \cdot 10^{-5}$	99.73
	SRIVC	$2.14 \cdot 10^{-2}$	$4.60 \cdot 10^{-3}$	94.31

In this example, SRIVC performs poorly, while PEM and PEMrd normally reach the global optimum. The standard implementation of SRIVC in CONTSID considers initial conditions equal to zero, which leads to inaccurate estimates⁵. Even though the fit is very near the optimal, PEMrd is consistently better than standard PEM at all metrics.

C. Mean and covariance of the estimated parameters

As established in Theorem 4.1, improved PEM should reduce (at least asymptotically) the covariance of the parameter vector. To test this, we obtained the mean and standard deviation of each parameter given by the Monte Carlo study under the setup of Section V-A with $h = 0.05$, $N = 7161$, and under the same stopping rule stated in Remark 5.1. The results in Table IV show that the mean values are similar for all methods (except standard PEM, which does not estimate the correct model structure), and PEMrd provides the lowest standard deviation for every parameter.

D. Direct comparison with standard PEM

In this subsection we analyse the improvements of the novel method over the standard PEM estimates. To show the impact of enforcing the correct relative degree, we focused on the Bode diagrams of the CT models obtained by PEM and PEMrd. In Figure 1 we have plotted the frequency

⁵The command `tfest` from MATLAB SYSID Toolbox has also been tested. In this implementation of SRIVC, estimation of initial conditions is taken into account, with results similar to PEMrd in this test.

response estimates of 100 Monte Carlo simulations under the setup in Section V-A, for $h = 0.05$, $N = 7161$.

It is clear by Fig. 1 that the improvement over PEM is mainly at high frequencies. This is intuitive, since the relative degree determines the asymptotic slope of the Bode diagram of magnitude. The proposed method enforces the true asymptotic slope, leading to an important gain in accuracy in both magnitude and phase.

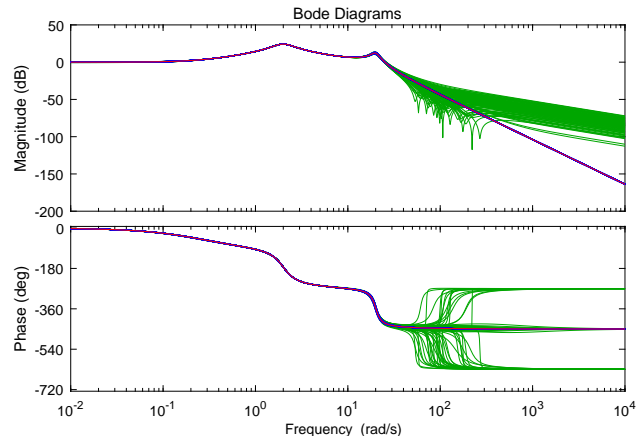


Fig. 1. Bode Diagrams of 100 estimates of $G_0(s)$ by PEM (green), 100 estimates of $G_0(s)$ by PEMrd (blue), and the real system (red).

Also, direct comparison plots between PEM and PEMrd are shown in Figure 2. These plots compare the normalized model and parameter error, and the fit of PEM and PEMrd for each Monte Carlo experiment (500 in total). PEMrd outperforms standard PEM in most Monte Carlo simulations, specially in the fit comparison, where 496 out of 500 experiments have lead to an increase in fit under PEM refinement.

E. Random systems

To obtain the average performance of each estimator, we tested them on a data set created with 500 random systems of order 3 and relative degree 2, by using the `rss` command in MATLAB. The slowest pole has been set to have real part less than -0.1. The input was a unit variance Gaussian white noise, and the zero-mean additive white noise had standard deviation equal to 5% of the maximum value of the noiseless output. The sampling period has been chosen as 10 times faster than the fastest pole or zero of the real system. Again, a zero-order hold sampling was considered.

We have computed the median of the metrics used above, as in Section V-B, for the 500 random systems. We also have counted failures of the estimators, which are the cases when the estimates produced a negative fit, when the algorithm crashed, or when it was not possible to correctly initialise the PEM estimate having tried reducing by a factor of 2 the sampling rate for initialisation estimation.

The results can be seen in Table V. Although all estimators report failures, the PEMrd again shows promising results in all metrics. As pointed out in [9] and seen in these tests, initialisation aspects are in fact a major issue concerning the reliability of these algorithms.

TABLE IV

ESTIMATED PARAMETER VALUE MEANS AND STANDARD DEVIATIONS FOR EACH METHOD, CONSIDERING $h = 0.05$.

Method	Parameter True value	b_1	b_2	b_3	b_4	a_1	a_2	a_3	a_4
		0	0	-6400	1600	5	408	416	1600
PEM	Mean	-0.04	0.30	-6403.01	1601.33	5.01	408.16	416.25	1600.49
	Std. Dev.	0.96	11.41	147.13	47.85	0.40	7.98	9.27	33.29
PEMrd	Mean	0	0	-6397.25	1599.79	4.99	407.88	415.84	1599.27
	Std. Dev.	0	0	122.39	42.39	0.32	7.11	8.33	28.59
SRIVC	Mean	0	0	-6399.19	1600.49	5.01	407.99	416.61	1599.71
	Std. Dev.	0	0	132.05	44.21	0.34	7.75	8.98	31.1

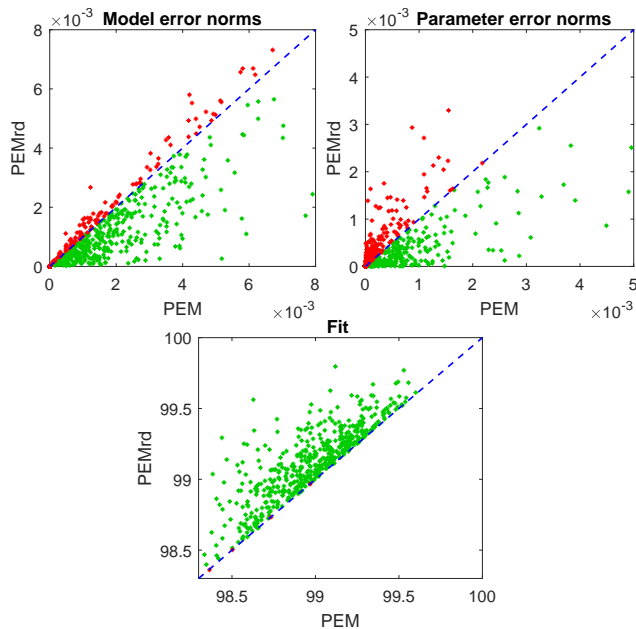


Fig. 2. Direct comparison plots between PEM and PEMrd for 500 Monte Carlo simulations. The green dots correspond to Monte Carlo simulations where PEMrd outperforms PEM. Red dots represent the opposite, and the dashed blue line is the separatrix.

TABLE V

MONTE CARLO SIMULATION RESULTS FOR RANDOM SYSTEMS.

Method	$\frac{\ \hat{G} - G_0\ ^2}{\ G_0\ ^2}$	$\frac{\ \hat{\theta} - \theta_0\ ^2}{\ \theta_0\ ^2}$	Fit	N ^o Failures
PEM	$1.45 \cdot 10^{-4}$	$3.33 \cdot 10^{-3}$	98.87	12
PEMrd	$1.05 \cdot 10^{-4}$	$1.87 \cdot 10^{-3}$	99.00	16
SRIVC	$1.11 \cdot 10^{-4}$	$2.13 \cdot 10^{-3}$	98.97	30

VI. CONCLUSIONS

We have proposed a refinement to the standard PEM estimator for indirect continuous-time system identification that achieves an asymptotically optimal desired relative degree enforcement. An explicit expression for this estimator has been found, and its statistical properties have been analysed. Extensive simulations using both standard benchmarks and random systems have been put forward with promising results. These have shown that the refinement to standard PEM leads to an important improvement in all the statistical metrics studied, and its performance is comparable, if not better, to SRIVC for all sampling periods in this study provided PEM is initialised correctly.

REFERENCES

- [1] T. Söderström and P. Stoica, *System Identification*. Prentice-Hall, 1989.
- [2] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice-Hall, 1999.
- [3] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*. John Wiley & Sons, 2012.
- [4] T. P. Bohlin, *Practical Grey-box Process Identification: Theory and Applications*. Springer, 2006.
- [5] G. Rao and H. Unbehauen, "Identification of continuous-time systems," *IEE Proceedings-Control theory and applications*, vol. 153, no. 2, pp. 185–220, 2006.
- [6] H. Garnier and L. Wang (Eds.), *Identification of Continuous-time Models from Sampled Data*. Springer, 2008.
- [7] 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.
- [8] G. Rao and H. Garnier, "Numerical illustrations of the relevance of direct continuous-time model identification," in *15th Triennial IFAC World Congress on Automatic Control*, Barcelona, Spain, vol. 35, pp. 133–138, 2002.
- [9] L. Ljung, "Experiments with identification of continuous time models," in *15th IFAC Symposium on System Identification*, Saint Malo, France, vol. 42, pp. 1175–1180, Elsevier, 2009.
- [10] T. Söderström, P. Stoica, and B. Friedlander, "An Indirect Prediction Error Method for System Identification," *Automatica*, vol. 27, no. 1, pp. 183–188, 1991.
- [11] K. J. Åström and B. Wittenmark, "Computer Controlled Systems: Theory and Design," *Prentice-Hall*, 1984.
- [12] L. Ljung, *System Identification Toolbox 7: Getting Started Guide*.
- [13] H. Unbehauen and G. Rao, "Continuous-time Approaches to System Identification - A Survey," *Automatica*, vol. 26, no. 1, pp. 23–35, 1990.
- [14] N. K. Sinha, "Identification of continuous-time systems from samples of input-output data: An introduction," *Sadhana*, vol. 25, no. 2, pp. 75–83, 2000.
- [15] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2012.
- [16] I. Kollar, G. Franklin, and R. Pintelon, "On the Equivalence of z-domain and s-domain Models in System Identification," in *Instrumentation and Measurement Technology Conference*, Brussels, Belgium, vol. 1, pp. 14–19, IEEE, 1996.
- [17] P. W. Zehna, "Invariance of maximum likelihood estimators," *Annals of Mathematical Statistics*, vol. 37, no. 3, p. 744, 1966.
- [18] C. Gourieroux and A. Monfort, *Statistics and econometric models*, vol. 1. Cambridge University Press, 1995.
- [19] R. A. González, P. E. Valenzuela, C. R. Rojas, and R. A. Rojas, "Optimal enforcement of causality in non-parametric transfer function estimation," *IEEE Control Systems Letters*, vol. 1, no. 2, pp. 268–273, 2017.
- [20] P. Young, "Recursive estimation, forecasting and adaptive control," *Control and Dynamic systems*, vol. 30, no. 3, pp. 119–165, 1989.
- [21] L. Ljung, "Initialisation aspects for subspace and Output-Error identification methods," in *European Control Conference (ECC)*, Cambridge, UK, pp. 773–778, 2003.
- [22] J. S. Welsh, J. C. Agüero, and M. Alamir, "Continuous-time System Identification using Indirect Inference," in *15th IFAC Symposium on System Identification*, Saint Malo, France, vol. 42, pp. 1169–1174, 2009.
- [23] M. Galrinho, C. R. Rojas, and H. Hjalmarsson, "The weighted null-space fitting method," *arXiv preprint arXiv:1708.03946*, 2017.