Blind Nonparametric Estimation of SISO Continuous-time Systems

Augustus Elton * **Rodrigo A. González** ** **James S. Welsh** *
Tom Oomen **,*** **Cristian R. Rojas** ****

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

Abstract: Blind system identification is aimed at finding parameters of a system model when the input is inaccessible. In this paper, we propose a blind system identification method that delivers a single-input single-output, continuous-time model in a nonparametric kernel form. We take advantage of the representer theorem to form a joint maximum *a posteriori* estimator of the input and system impulse response. The identified system model and input are optimised in sequence to overcome the blind problem with generalised cross validation used to select appropriate hyperparameters given some fixed input sequence. We demonstrate via Monte Carlo simulations the accuracy of the method in terms of estimating the input.

Keywords: Continuous-time system identification, Nonparametric methods, Identifiability

1. INTRODUCTION

Blind system identification (BSI) is exercised in lieu of standard system identification (Ljung, 1999) when the user only has access to output measurements. These problems generally are ill-posed and require some prior knowledge about the input for the system to be identifiable up to a scalar constant of the true system (Bai and Fu, 1999). Similarly, blind equalisation is derived from the same BSI problem where some prior knowledge about the system must be provided for the input to be identified. These procedures have been applied in the areas of biomedicine, image reconstruction and most commonly in communications (Abed-Meraim et al., 1997).

Techniques are generally developed around particular applications, which has resulted in a broad range of blind methods appropriate to single-input single-output (SISO), single-input multiple-output (SIMO) and multiple-input multiple-output systems. These techniques cannot in general be directly applied to any arbitrary output to form a unique solution, as often strict requirements about the system or the input must be met, such as signal stationarity or multiple channels. For example, the cross-relation method (Xu et al., 1995) is a second order statistic-based method that is limited to SIMO systems where there are multiple correlated outputs.

The blind system identification of SISO models may not be tractable and generally, conditions about the input or the system are imposed to make the problem feasible. Bai and Fu (1999) developed an oversampling technique for SISO, linear BSI that uses an infinite impulse response, which has been further developed for Hammerstein and Wiener systems (Bai, 2002). This approach identifies a discrete-time system blindly by reconstructing the output information into a SIMO problem with respect to sets of the output samples, which is similar in construction to the cross-relation method. A strict requirement is that the output is over-sampled by a factor greater than the input, where the input is assumed to be constant during the over-sampling period. Moreover, subspace methods have been developed, where the input is assumed to belong to a known subspace. For instance Ohlsson et al. (2014) used lifting to perform the blind identification of an ARX system. In relation to our proposal, Bottegal et al. (2015) constructed a blind nonparametric estimate using a kernel-based method.

Although progress has been made on BSI, at present, they 1) cannot be employed directly to admit irregular output sampling, and 2) generically suffer from ill-conditioning when converting them to continuous-time (Garnier and Young, 2014), which is physically relevant for some biological applications (Friedrich et al., 2017). This paper addresses these issues. In particular,

(C1) We develop a joint maximum *a posteriori* input and continuous-time impulse response estimator that promotes sparsity in the input changes and regularisation in the impulse response. For this the following aspects are addressed.
  (a) We propose that the input and impulse response estimate can be updated sequentially, within a kernel-regularised identification framework.
  (b) We show that the kernel matrix may be decomposed by assuming that the input is held constant between samples for a stable spline kernel.
  (c) We provide an expression for the transfer function related to the impulse response estimate.

(C2) We verify the algorithm via Monte Carlo simulations, where the effects of the hyperparameters and noise are explored.

⋆ This work was supported in part by the Australia government Research Training Program (RTP) scholarship and in part by the Swedish Research Council under contract number 2016-06079 (NewLEADS), by the Digital Futures project EXTREMUM, and by the research program VIDI with project number 15698, which is (partly) financed by the Netherlands Organization for Scientific Research (NWO).

The paper is organised as follows. In Section 2 the BSI problem is presented. Section 3 describes the blind joint estimation scheme, which is the main contribution of this work. An analysis on the blind estimator is performed in Section 4 via Monte Carlo simulations and Section 5 draws conclusions. Proofs of the main theoretical results can be found in the Appendix.

2. PROBLEM FORMULATION

Consider the linear time-invariant, asymptotically stable, continuous-time system

$$\tilde{y}(t) = \int_{-\infty}^{t} g(\tau)u(t-\tau)\mathrm{d}\tau, \tag{1}$$

where $u(t)$ is the input signal and $g(t)$ is the impulse response of the system. Equivalently, we can write $\tilde{y}(t) = (g * u)(t)$, where the symbol $*$ denotes convolution. The output signal is given by N noisy measurements of $\tilde{y}(t)$ at time instants $t_i > 0$,

$$y(t_i) = \tilde{y}(t_i) + v(t_i), \text{ for } t_i = [t_0, t_1, \ldots, t_N], \tag{2}$$

where $v(t_i)$ is a zero-mean white Gaussian noise with a variance of σ^2. Our aim is to estimate the input sequence $\{u(t_i)\}_{i=0}^{N-1}$, and a nonparametric transfer function for the impulse response $g(t)$, directly in continuous-time, using knowledge of the output measurements $\{y(t_i)\}_{i=1}^{N}$. Without implying some knowledge on the input or system the problem is ill-posed and the system is not identifiable. In this work, we assume that the input intersample behaviour is a zero-order hold and that the changes in the input amplitude may occur sparsely over time.

3. JOINTLY REGULARISED INPUT AND IMPULSE RESPONSE ESTIMATION

In a blind setting, where $\{y(t_i)\}_{i=1}^{N}$ is the only information available, identifying $g(t)$ and $u(t)$ is ill-posed; there is an infinite amount of solutions available to reconstruct the output. To restrict the solutions, we will solve the BSI problem in a Bayesian setting by introducing adequate priors for g and the input sequence $\mathbf{u} := [u(t_0), u(t_1), \ldots, u(t_{N-1})]^\top$, with the goal of solving a maximum *a posteriori* estimation problem.

To this end, we define the output measurement vector as $\mathbf{y} = [y(t_1), y(t_2), \ldots, y(t_N)]^\top$, and assume that $g(t)$ can be modelled as a zero-mean Gaussian process. Our proposal, outlining the first contribution (C1), is that we compute a joint input and impulse response estimate

$$\begin{aligned}\begin{bmatrix}\hat{g}\\ \hat{\mathbf{u}}\end{bmatrix} &= \arg\max_{g,\mathbf{u}} \mathrm{p}(g, \mathbf{u}|\mathbf{y})\\ &= \arg\max_{g,\mathbf{u}} \frac{\mathrm{p}(g, \mathbf{u}, \mathbf{y})}{\mathrm{p}(\mathbf{y})}\\ &= \arg\max_{g,\mathbf{u}} \log(\mathrm{p}(\mathbf{y}|g, \mathbf{u})) + \log(\mathrm{p}(g)) + \log(\mathrm{p}(\mathbf{u})), \end{aligned} \tag{3}$$

where in the last line we have exploited the prior independence of g and $\mathbf{u}$, and the monotonicity of the logarithm function. The three terms in (3) will be analysed next. First, note that

$$\mathrm{p}\left(y(t_i)|g, \mathbf{u}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(y(t_i) - (g * u)(t_i)\right)^2\right),$$

where we have used the fact that $v(t_i)$ in (2) is Gaussian-distributed. Therefore, the first term in the cost (3) can be expressed as

$$\log(\mathrm{p}(\mathbf{y}|g, \mathbf{u})) = \frac{-N}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{N}\frac{\left(y(t_i) - (g*u)(t_i)\right)^2}{\sigma^2}. \tag{4}$$

Note that $-(N/2)\log(2\pi\sigma^2)$ is constant with respect to the impulse response and input, and can be excluded from the maximisation in (3). With regards to the second term in (3), the prior $\mathrm{p}(g)$ is written with an abuse of notation since the concept of probability density function is not well defined in infinite-dimensional function spaces (Aravkin et al., 2014). Based on the relationship between kernel methods and Gaussian processes (Pillonetto et al., 2022), we can restrict the optimisation problem (3) to impulse responses belonging to a reproducing kernel Hilbert space (RKHS) $\mathcal{G}$ (with norm $\|\cdot\|_{\mathcal{G}}$), and pick

$$\log(\mathrm{p}(g)) = -\gamma\|g\|_{\mathcal{G}}^2 + C, \tag{5}$$

where C is a normalising constant, and γ is a regularisation parameter. Such choice of prior is formally justified under certain conditions; we refer to Aravkin et al. (2014) for the details. Finally, the input prior $\mathrm{p}(\mathbf{u})$ is designed to take into account the sparsity of the changes in the input. This can be achieved by modelling the input differences by a Laplace distribution

$$\log\left(\mathrm{p}(\mathbf{u})\right) = -\lambda\sum_{i=1}^{N-1}|u(t_i) - u(t_{i-1})|, \tag{6}$$

where λ is a penalty parameter which affects the number of changes forced into the input sequence, similar to that used in fused-lasso, l_1-mean filtering and total variation denoising estimation (Rojas and Wahlberg, 2014).

Substituting terms from (4), (5) and (6) into (3), yields the reduced joint impulse response and input optimisation problem

$$\begin{aligned}\begin{bmatrix}\hat{g}\\ \hat{\mathbf{u}}\end{bmatrix} = \arg\min_{g\in\mathcal{G},\mathbf{u}\in\mathbb{R}^N}\Bigg(&\sum_{i=1}^{N}\left(y(t_i) - (g*u)(t_i)\right)^2\\ &+ \tilde{\gamma}\|g\|_{\mathcal{G}}^2 + \tilde{\lambda}\sum_{i=1}^{N-1}|u(t_i) - u(t_{i-1})|\Bigg),\end{aligned} \tag{7}$$

where we have denoted $\sigma^2\gamma$ and $\sigma^2\lambda$ as $\tilde{\gamma}$ and $\tilde{\lambda}$, respectively. These parameters, together with the kernel, are unknown and must be tuned with the data at hand.

3.1 Kernel Selection and Tuning

In the context of kernel methods for system identification, the impulse response $g(t)$ is modelled by a Gaussian process where the covariance $\mathbb{E}(g(t)g(\tau))$ is described by a positive semi-definite kernel function, $k(t,\tau)\colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$. The selection of the kernel function permits the incorporation of prior knowledge to the estimator, such as BIBO stability. Furthermore, there exists a variety of well-established kernel descriptions that can be used to parameterise kernel functions (Pillonetto et al., 2022). In this paper we consider, while not being limited to, the stable spline kernel of order q with the form

$$k(t,\tau) = s_q(e^{-\beta t}, e^{-\beta\tau}), \tag{8}$$

where β is a strictly positive hyperparameter, and s_q is the regular spline kernel of order q. In general, s_q can be expressed as (Scandella et al., 2022, Prop. 2.1)

$$s_q(e^{-\beta t}, e^{-\beta\tau}) = \sum_{r=0}^{q-1}\gamma_{q,r}\begin{cases}e^{-\beta(2q-r-1)t}e^{-r\beta\tau} & \text{if } t \geq \tau,\\ e^{-\beta(2q-r-1)\tau}e^{-r\beta t} & \text{if } t < \tau,\end{cases}$$

where

$$\gamma_{q,r} = \frac{(-1)^{q+r-1}}{r!(2q-r-1)!}. \tag{9}$$

For example, for $q = 1$ we have $s_1(e^{-\beta t}, e^{-\beta\tau}) = e^{-\beta\max(t,\tau)}$.

Now that we have a model for the impulse response we must select the kernel hyperparameters $\boldsymbol{\rho} := [\beta, \tilde{\gamma}]$ and choose the penalty parameter $\tilde{\lambda}$. Standard continuous-time approaches can be used to tune these hyperparameters for a given input sequence. Due to the connections between regularised least squares methods and the regularised function estimation in RKHSs (Pillonetto et al., 2022), we use generalised cross validation, which finds the hyperparameters, $\hat{\boldsymbol{\rho}} := [\hat{\beta}, \hat{\gamma}]$, that minimise the predicted residual sum of squares

$$\hat{\boldsymbol{\rho}}(\hat{\mathbf{u}}) = \underset{\boldsymbol{\rho}\in\Gamma}{\arg\min} \sum_{k=1}^{N} \left(\frac{y(t_k) - \hat{y}(t_k)}{1 - \operatorname{tr}(\hat{\mathbf{H}}(\boldsymbol{\rho}))/N} \right)^2, \tag{10}$$

where $\hat{\mathbf{H}}(\boldsymbol{\rho}) = \hat{\mathbf{K}}(\boldsymbol{\rho})(\hat{\mathbf{K}}(\boldsymbol{\rho})+\tilde{\gamma}\mathbf{I}_N)^{-1}$. The predicted output $\hat{y}(t_k)$ evaluated at $\hat{\boldsymbol{\rho}}(\hat{\mathbf{u}})$ is found by

$$\hat{\mathbf{y}}(\hat{\boldsymbol{\rho}}) = \hat{\mathbf{K}}(\hat{\boldsymbol{\rho}})(\hat{\mathbf{K}}(\hat{\boldsymbol{\rho}}) + \tilde{\gamma}\mathbf{I}_N)^{-1}\mathbf{y}.$$

The kernel matrix $\mathbf{K}$ has entries given by

$$\mathbf{K}_{ij} = \int_0^\infty \int_0^\infty k(t,\tau)u(t_i - t)u(t_i - \tau)\mathrm{d}t\mathrm{d}\tau, \tag{11}$$

and $\hat{\mathbf{K}}(\boldsymbol{\rho})$ is the kernel matrix $\mathbf{K}$ defined by (11) evaluated at $\hat{\mathbf{u}}$ from (7). The penalty parameter $\tilde{\lambda}$ is often set prior to the optimisation and can be varied until a desired sparsity in the input changes is obtained.

With the formation of the problem in (7) where the parameter selection is based off the criterion in (10), the problem is still not feasible. The following section optimises the input and impulse response estimates in sequence rather then computing a joint estimate, which constitutes contribution (C1.a) of the paper.

3.2 Sequential Joint Solution

The following theorem provides alternative expressions for $\hat{g}(t)$ and $\hat{\mathbf{u}}$ in (7) that are later used for deriving a continuous-time frequency response function estimate based solely on output data.

Theorem 1. Consider the cost in (7), where $\mathcal{G}$ is an RHKS with an arbitrary positive semi-definite kernel function $k\colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$. Then, the optimal input sequence $\hat{\mathbf{u}}$, with elements $\hat{u}(t_i)$, can be computed from

$$\hat{\mathbf{u}} = \underset{\mathbf{u}\in\mathbb{R}^N}{\arg\min} \left(\tilde{\gamma}\mathbf{y}^\top(\mathbf{K} + \tilde{\gamma}\mathbf{I}_N)^{-1}\mathbf{y} + \tilde{\lambda}\|\mathbf{D}\mathbf{u}\|_1 \right), \tag{12}$$

where the kernel matrix $\mathbf{K}$ is given in (11) and the penalty matrix $\mathbf{D} \in \mathbb{R}^{(N-1)\times N}$ is given by

$$\mathbf{D} = \begin{bmatrix} 1 & -1 & & \mathbf{0} \\ & 1 & -1 & \\ & & \ddots & \\ \mathbf{0} & & 1 & -1 \end{bmatrix}. \tag{13}$$

Moreover, the impulse response $\hat{g}$ that minimises the cost in (7) is given by

$$\hat{g}(t) = \sum_{i=1}^{N} \hat{c}_i \hat{g}_i(t), \tag{14}$$

where the representer functions are

$$\hat{g}_i(t) = \int_0^\infty k(t,\tau)\hat{u}(t_i - \tau)\mathrm{d}\tau, \tag{15}$$

and $\hat{c}_i$ is the ith entry of the optimal weighting vector

$$\hat{\mathbf{c}} = (\hat{\mathbf{K}} + \tilde{\gamma}\mathbf{I}_N)^{-1}\mathbf{y}, \tag{16}$$

with $\hat{\mathbf{K}}$ being the kernel matrix $\mathbf{K}$ defined in (11) evaluated at the optimal input sequence $\hat{\mathbf{u}}$.

Proof. See Appendix A. □

Once the optimisation problem in (12) is solved, Theorem 1 gives a closed-form expression for the impulse response estimate, which holds for equidistant and non-equidistant sampling. This input optimisation can be difficult to solve, especially if many different input values are present, due to the computation of the matrix inverse of dimension $N \times N$. In this paper, we decompose the kernel matrix $\mathbf{K}$ to make the input cost evaluation easier by assuming that the sampling times are equally spaced by some known period h. This adds to contribution (C1.b).

Lemma 2. Consider the kernel matrix $\mathbf{K}$ with entries described in (11), and the stable spline kernels in (8) which are parameterised by β. If $u(t)$ is constant between the instants $t = 0, h, \ldots, Nh$, then $\mathbf{K}$ admits the decomposition

$$\mathbf{K} = \boldsymbol{\Phi}\mathcal{O}_\beta\boldsymbol{\Phi}^\top, \tag{17}$$

where $\boldsymbol{\Phi}$ is given by

$$\boldsymbol{\Phi} = \begin{bmatrix} u(0) & & & 0 \\ u(h) & u(0) & & \\ \vdots & & \ddots & \\ u([N-1]h) & u([N-2]h) & \cdots & u(0) \end{bmatrix}, \tag{18}$$

and the matrix $\mathcal{O}_\beta \in \mathbb{R}^{N\times N}$ has entries

$$\mathcal{O}_{\beta,ij} = \sum_{r=0}^{q-1} \frac{\gamma_{q,r} e^{-\beta h(2q-1)\max\{i,j\}}}{\beta^2 r(2q-r-1)} \begin{cases} a(\beta) & \text{if } i=j, \\ b_{i-j}(\beta) & \text{if } i\neq j, \end{cases} \tag{19}$$

where

$$a(\beta) = 2[(2q-r-1) + re^{\beta h(2q-1)} - (2q-1)e^{\beta hr}]/(2q-1),$$
$$b_{i-j}(\beta) = e^{-\beta hr(1-|i-j|)}(e^{\beta rh}-1)(e^{\beta h(2q-1)} - e^{\beta rh}),$$

with $q \geq 1$ being the order of the stable spline kernel.

Proof. See Appendix B. □

Given the decomposition in Lemma 2, we consider the Cholesky factorisation $\mathcal{O}_\beta/\tilde{\gamma} = \mathbf{L}\mathbf{L}^\top$ (i.e., $\mathbf{L}$ is an upper triangular matrix with positive diagonal), and the thin QR factorisation (Chen and Ljung, 2013; González et al., 2021)

$$\begin{bmatrix} \boldsymbol{\Phi}\mathbf{L} & \mathbf{y} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} = \mathbf{Q}\begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & r \end{bmatrix}, \tag{20}$$

where $\mathbf{Q}$ is an orthogonal matrix, $\mathbf{R}_1$ is an upper triangular matrix with positive diagonal entries of dimension $N \times N$, and $r > 0$. Note that the following identities are satisfied:

$$\mathbf{R}_1^\top\mathbf{R}_1 = \mathbf{L}^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\mathbf{L} + \mathbf{I},$$
$$\mathbf{R}_1^\top\mathbf{R}_2 = \mathbf{L}^\top\boldsymbol{\Phi}^\top\mathbf{y},$$
$$\mathbf{R}_2^\top\mathbf{R}_2 + r^2 = \|\mathbf{y}\|^2.$$

By leveraging the matrix inversion lemma (Horn and Johnson, 2012) and the identities above, the first summand in (12) can be written as

$$\begin{aligned} \tilde{\gamma}\mathbf{y}^\top(\mathbf{K}+\tilde{\gamma}\mathbf{I})^{-1}\mathbf{y} &= \mathbf{y}^\top\left[\mathbf{I} - \boldsymbol{\Phi}\mathbf{L}(\mathbf{L}^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\mathbf{L}+\mathbf{I})^{-1}\mathbf{L}^\top\boldsymbol{\Phi}^\top\right]\mathbf{y} \\ &= \mathbf{R}_2^\top\mathbf{R}_2 + r^2 - \mathbf{R}_2^\top\mathbf{R}_1(\mathbf{R}_1^\top\mathbf{R}_1)^{-1}\mathbf{R}_1^\top\mathbf{R}_2 \\ &= r^2. \end{aligned}$$

Therefore, we have proven the following corollary.

Corollary 3. If $u(t)$ is constant between the time instants $t = 0, h, \ldots, (N-1)h$, then $\hat{\mathbf{u}}$ in (12) can be computed from

$$\hat{\mathbf{u}} = \underset{\mathbf{u}\in\mathbb{R}^N}{\arg\min} \left([r(\mathbf{u})]^2 + \tilde{\lambda}\|\mathbf{D}\mathbf{u}\|_1 \right), \tag{21}$$

where $r(\mathbf{u})$ is given by the thin QR factorisation in (20).

In summary, we have obtained an inverse-free cost function for finding the optimal input sequence, whose solution can

be used to directly obtain the continuous-time impulse response estimate using Theorem 1.

3.3 Nonparametric Transfer Function Estimation

In many practical scenarios, a frequency-domain estimate of the system g is sought after. Contrary to the impulse response estimation procedure, the transfer function estimate associated with the estimated kernel hyperparameters and input sequence does not require the computation of the integrals in (15). Next, we provide a closed form expression for the transfer function estimate of g, which corresponds to contribution (C1.c) of this paper. Note this only holds for equidistant sampling.

Theorem 4. Consider the joint optimisation problem in (7). The transfer function estimate associated with the minimiser of (7) can be formed as

$$\hat{G}(s;\hat{\mathbf{u}}) = \mathcal{K}^\top(s)(\hat{\mathbf{\Phi}}^\top\hat{\mathbf{\Phi}}\mathcal{O}_\beta + \tilde{\gamma}\mathbf{I})^{-1}\hat{\mathbf{\Phi}}^\top\mathbf{y}, \tag{22}$$

where $\hat{\mathbf{\Phi}}$ is the Toeplitz matrix in (18) evaluated at $\hat{\mathbf{u}}$, and $\mathcal{K}(s)$ is a vector of size N with entries

$$\mathcal{K}_l(s) = \sum_{r=0}^{q-1} \frac{\gamma_{q,r} e^{-l\beta h(2q-r-1)}(e^{\beta h(2q-r-1)} - 1)}{\beta(s+r\beta)(2q-r-1)} + \frac{(-1)^q \beta^{2q-1} e^{-lh(s+\beta[2q-1])}(e^{h(s+\beta[2q-1])} - 1)}{(s+\beta[2q-1])\prod_{k=0}^{2q-1}(s+k\beta)}, \tag{23}$$

where q is the order of the stable spline kernel.

Proof. See Appendix C. □

Remark 5. The blind system identification problem can only estimate the system or input up to a scalar multiple of the true system. This can be seen in (22) if the hyperparameter $\tilde{\gamma}$ is also gain-dependent. Indeed, if we consider the notation $\hat{G}(s;\hat{\mathbf{u}},\tilde{\gamma})$ to stress the dependence of $\tilde{\gamma}$ in the transfer function estimate, by setting $\bar{\tilde{\gamma}} = \alpha^2\tilde{\gamma}$ and $\bar{\hat{\mathbf{u}}} = \alpha\hat{\mathbf{u}}$ with $\alpha \neq 0$, we find that $\hat{G}(s;\bar{\hat{\mathbf{u}}},\bar{\tilde{\gamma}}) = \hat{G}(s;\hat{\mathbf{u}},\tilde{\gamma})/\alpha$, as expected. This is not a limitation of the proposal, rather it is inherent to the blind identification problem.

3.4 Complete Algorithm

We estimate the input sequence in (21) and the transfer function estimate from (22) by forming the kernel hyperparameters via the optimisation in (10). The function *init_input* in Algorithm 1 initialises the first n components of the input by applying a grid search for the first n input samples to the input cost objective function in (12). It is evident from the construction of the regressor matrix in (18) that the first inputs have a larger effect on the cost. Thus for computational efficiency the first three values of the input are initialised and the rest are set to zero.

Algorithm 1 Blind nonparametric continuous-time system identification for fixed sparsity factor $\tilde{\lambda}$

1: Input: data $\{y(t_i)\}_{i=1}^N$, initial estimates $[\beta,\tilde{\gamma},\tilde{\lambda}] \leftarrow [1, 0.1, 0.1]$, number of input samples to initialise $n \leftarrow 3$
2: Construct $\mathbf{y} \leftarrow [y(t_1), ..., y(t_N)]^\top$
3: Form $\hat{\boldsymbol{\rho}}^{(0)} \leftarrow [\tilde{\gamma}, \beta]$
4: Initialise $\hat{\mathbf{u}}^{(0)} \leftarrow$ init_input$(n, \{y(t_i)\}_{i=1}^n, \hat{\boldsymbol{\rho}}^{(0)})$
5: $\hat{\mathbf{u}} \leftarrow \arg\min_{\mathbf{u}\in\mathbb{R}^N} \left([r(\mathbf{u})]^2 + \tilde{\lambda}\|\mathbf{D}\mathbf{u}\|_1\right)$ ▷ Given $\hat{\mathbf{u}}^{(0)}$, $\hat{\boldsymbol{\rho}}^{(0)}$
6: $\hat{\boldsymbol{\rho}}(\hat{\mathbf{u}}) \leftarrow \arg\min_{\boldsymbol{\rho}\in\Gamma} \sum_{k=1}^N \left(\frac{\mathbf{y}-\hat{\mathbf{y}}}{1-\mathrm{tr}(\mathbf{H}(\boldsymbol{\rho},\hat{\mathbf{u}}))/N}\right)^2$
7: Output: $\hat{\mathbf{u}}$ and compute $\hat{G}(s;\hat{\mathbf{u}})$ from (22)

4. SIMULATIONS

The simulation results are outlined as follows. First we explain the experimental setup in Section 4.1. We then examine the effects of varying the input penalty hyperparameter, λ, in Section 4.2, and assess input fit scores from differing noise levels in 4.3. This provides for our final contribution (C2). The parameter λ is set prior to the optimisation in Step 5 of Algorithm 1.

4.1 Simulation setup

We consider the first order system with time constant $\tau > 0$

$$G(s) = \frac{1}{\tau s + 1}, \tag{24}$$

which is simulated using a known persistently exciting input $\tilde{u}(t)$, where the output is contaminated with Gaussian i.i.d noise and sampled at a rate of $T_s = \tau/5$ seconds. The estimated input, $\hat{\mathbf{u}}$ and transfer function estimate, $\hat{G}(s,\hat{\mathbf{u}})$ are obtained using Algorithm 1, where a stable spline kernel of order 1 is selected. The results of a Monte Carlo analysis, in Section 4.3, are evaluated using an input fit score

$$\mathrm{F} := \left(1 - \|\alpha(\hat{\mathbf{u}} - \tilde{\mathbf{u}})\|_2 / \|\tilde{\mathbf{u}} - \bar{\tilde{\mathbf{u}}}\|_2\right) \times 100\%, \tag{25}$$

where $\bar{\tilde{\mathbf{u}}}$ is the mean of the true input vector $\tilde{\mathbf{u}}$ and the scaling constant α is the median of the vector $\tilde{\mathbf{u}} \cdot (1/\hat{\mathbf{u}})$.

4.2 Case Study I: Effect of λ and promoting sparsity

The blind form of the estimator is investigated for a range of λ_i values, where the ith input sequence estimate, $\hat{\mathbf{u}}_i$ corresponds to $[\lambda_1 = 0.001, \lambda_2 = 0.009, \lambda_3 = 0.086, \lambda_4 = 0.794]$. Here, the system is excited by a piecewise constant signal with a period of $T_i = 5\tau$ seconds, where the input, $\tilde{\mathbf{u}}$, and output, $\tilde{\mathbf{y}}$, are sampled at an interval of $T_s = \tau/5$ seconds to produce $N = 100$ samples, as shown in Fig. 1. The noisy output, $\mathbf{y}$, is formed to have an SNR of 20 dB.

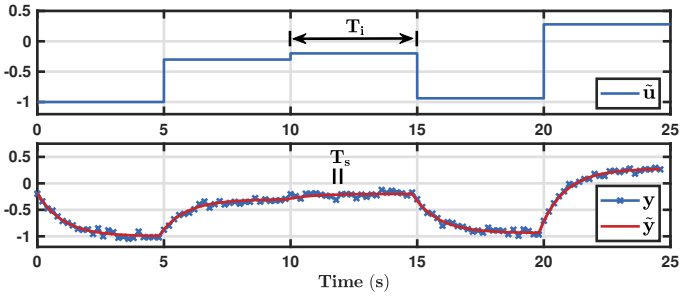

Fig. 1. Simulation of the system in (24) given the input $u(t)$, including $\tilde{y}(t)$ from (1) and $y(t)$ from (2).

Fig. 2 shows the results when no knowledge of the input period, T_i, is used. This is achieved by only considering that the input is constant between sampling intervals T_s. It is evident that for increasing values of λ, the changes in the input become more sparse and smooth. In this example an accurate estimation of the true input is particularly difficult since the number of input values to be estimated is equal to the number of output data points available.

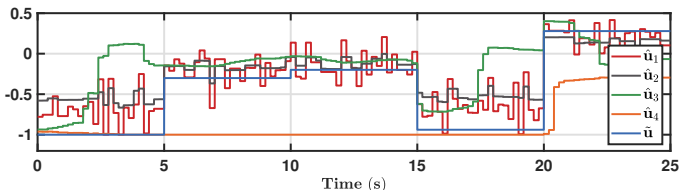

Fig. 2. Effects of varying λ with no input information.

To increase the performance of the input estimation, we include some prior input information. Here, we assume the input is held for τ seconds and optimise for $N/5$ unique values in the input optimisation. This improves the estimation of the input due to reduction of the parameters in the optimisation, as shown in Fig. 3.

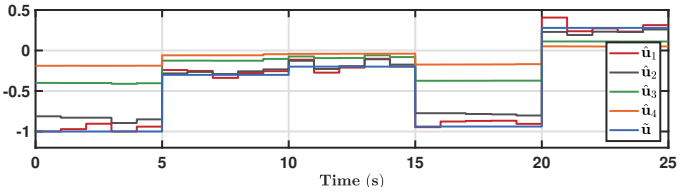

Fig. 3. Effects of varying λ with the assumption that the input signal is constant over τ seconds.

4.3 Case Study II: Effect of Noise

Lastly, the input optimisation is performed for different levels of SNR between $y(t_i)$ and $v(t_i)$ in (2). The Monte Carlo runs involve simulating the system, in (24), with different sequences of inputs. In Algorithm (1) we set λ to zero and use the known constant period of the input. Here, we simulate each input as a piecewise constant signal with $T_i = 5\tau$, $T_s = \tau/5$ and 10 input changes to produce a total of 250 samples. Also, the constant input period T_i is known *priori*. As expected, the input fit score improves for larger signal-to-noise ratios, shown in Fig. 4. Overall, there were $[13, 8, 6, 7, 6]$ failed runs at a SNR of $[6, 12, 20, 40, 60]$ dB, respectively, due to poor input initialisations.

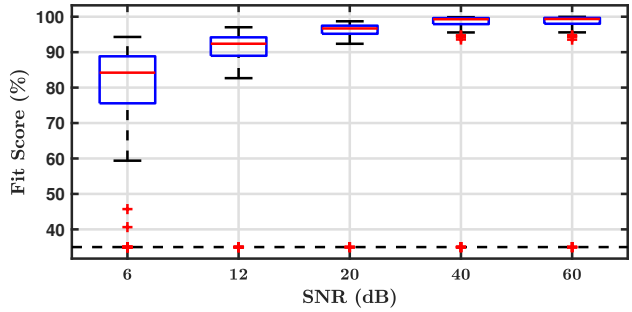

Fig. 4. Effects of varying SNR levels on the input fit score for 100 Monte Carlo runs per noise level.

5. CONCLUSIONS

This paper has presented a blind nonparametric estimator for continuous-time systems. A joint input and impulse estimation scheme is proposed where input and impulse response estimates are optimised in sequence. This approach incorporates regularisation in the impulse response estimate, while also promoting sparsity in the input amplitude changes. Many simulations were conducted to highlight its performance. As expected, the blind estimation improved significantly when the constant input period is known, the amount of samples was increased and when there were reduced levels of noise in the measured output.

REFERENCES

Abed-Meraim, K., Qiu, W., and Hua, Y. (1997). Blind system identification. *Proceedings of the IEEE*, 85(8), 1310–1322.

Aravkin, A.Y., Bell, B.M., Burke, J.V., and Pillonetto, G. (2014). The connection between Bayesian estimation of a Gaussian random field and RKHS. *IEEE Trans. on Neural Networks and Learning Systems*, 26(7), 1518–1524.

Bai, E.W. (2002). A blind approach to the Hammerstein–Wiener model identification. *Automatica*, 38(6), 967–979.

Bai, E.W. and Fu, M. (1999). Blind system identification and channel equalization of IIR systems without statistical information. *IEEE Trans. Signal Processing*, 47(7), 1910–1921.

Bottegal, G., Risuleo, R.S., and Hjalmarsson, H. (2015). Blind system identification using kernel-based methods. *IFAC-PapersOnLine*, 48(28), 466–471.

Chen, T. and Ljung, L. (2013). Implementation of algorithms for tuning parameters in regularized least squares problems in system identification. *Automatica*, 49(7), 2213–2220.

Friedrich, J., Zhou, P., and Paninski, L. (2017). Fast online deconvolution of calcium imaging data. *PLoS Computational Biology*, 13(3), 1–26.

Garnier, H. and Young, P.C. (2014). The advantages of directly identifying continuous-time transfer function models in practical applications. *Int. Journal of Control*, 87(7), 1319–1338.

González, R.A., Rojas, C.R., and Hjalmarsson, H. (2021). Non-causal regularized least-squares for continuous-time system identification with band-limited input excitations. In *Proceedings of the 60th IEEE Conference on Decision and Control*, 114–119.

Horn, R.A. and Johnson, C.R. (2012). *Matrix Analysis,* 2nd ed. Cambridge University Press.

Ljung, L. (1999). *System Identification,* 2nd ed. Prentice-Hall.

Ohlsson, H., Ratliff, L., Dong, R., and Sastry, S.S. (2014). Blind identification via lifting. *IFAC Proceedings Volumes*, 47(3), 10367–10372.

Pillonetto, G., Chen, T., Chiuso, A., De Nicolao, G., and Ljung, L. (2022). *Regularized System Identification.* Springer.

Rojas, C.R. and Wahlberg, B. (2014). On change point detection using the fused lasso method. *arXiv preprint arXiv:1401.5408.*

Scandella, M., Mazzoleni, M., Formentin, S., and Previdi, F. (2022). Kernel-based identification of asymptotically stable continuous-time linear dynamical systems. *Int. Journal of Control*, 95(6), 1668–1681.

Schölkopf, B., Herbrich, R., and Smola, A.J. (2001). A generalized representer theorem. In *Int. Conference on Computational Learning Theory*, 416–426.

Xu, G., Liu, H., Tong, L., and Kailath, T. (1995). A least-squares approach to blind channel identification. *IEEE Trans. Signal Processing*, 43(12), 2982–2993.

Appendix A. PROOF OF THEOREM 1

Proof. Denote $V(g, \mathbf{u})$ as the cost in (7). For a fixed input, let $\hat{g}_{\mathbf{u}} = \arg\min_{g\in\mathcal{G}} V(g, \mathbf{u})$, and let $\hat{\mathbf{u}}$ be the minimiser of $V(\hat{g}_{\mathbf{u}}, \mathbf{u})$. We then have for any $g \in \mathcal{G}$ and $\mathbf{u} \in \mathbb{R}^N$ that

$$V(g, \mathbf{u}) \geq V(\hat{g}_{\mathbf{u}}, \mathbf{u}) \geq V(\hat{g}_{\hat{\mathbf{u}}}, \hat{\mathbf{u}}),$$

indicating that we can minimise $V(g, \mathbf{u})$ for a fixed input sequence and then minimise $V(\hat{g}_{\mathbf{u}}, \mathbf{u})$ with respect to $\mathbf{u}$ to obtain the optimal point $(\hat{g}_{\hat{\mathbf{u}}}, \hat{\mathbf{u}})$. For a fixed input sequence $\mathbf{u}$, the optimisation problem in (7) is simply the standard continuous-time impulse response estimation problem in an RKHS framework (see, e.g., Scandella et al. (2022))

$$\hat{g}_{\mathbf{u}} = \arg\min_{g\in\mathcal{G}} \left(\sum_{i=1}^{N} (y(t_i) - (g * u)(t_i))^2 + \tilde{\gamma}\|g\|_{\mathcal{G}}^2 \right). \quad \text{(A.1)}$$

Thus, the representer theorem (Schölkopf et al., 2001) applied to (A.1) directly leads to $\hat{g}_{\mathbf{u}}$ being given by (14),

with representers $\hat{g}_i$ and weighting vector $\hat{\mathbf{c}}$ of the form (15) and (16), respectively, but evaluated at $\mathbf{u}$ instead of $\hat{\mathbf{u}}$. Evaluating these expressions at the optimal input sequence leads to the impulse response estimate of the theorem.

For the input sequence optimisation, we note that the representer theorem provides the convolution representation $(\hat{g}_{\mathbf{u}} * u)(t_i) = \mathbf{K}_i^\top (\mathbf{K} + \tilde{\gamma}\mathbf{I}_N)^{-1}\mathbf{y}$ for any bounded $\mathbf{u}$. Thus,

$$\begin{aligned}&\sum_{i=1}^{N}\big(y(t_i)-(\hat{g}_{\mathbf{u}}*u)(t_i)\big)^2+\tilde{\gamma}\|\hat{g}_{\mathbf{u}}\|_{\mathcal{G}}^2\\ &=\|\mathbf{y}-\mathbf{K}(\mathbf{K}+\tilde{\gamma}\mathbf{I}_N)^{-1}\mathbf{y}\|^2+\tilde{\gamma}\mathbf{y}^\top(\mathbf{K}+\tilde{\gamma}\mathbf{I}_N)^{-1}\mathbf{K}(\mathbf{K}+\tilde{\gamma}\mathbf{I}_N)^{-1}\mathbf{y}\\ &=\tilde{\gamma}\mathbf{y}^\top(\mathbf{K}+\tilde{\gamma}\mathbf{I}_N)^{-1}\mathbf{y}. \qquad \text{(A.2)}\end{aligned}$$

On the other hand, by rearranging terms, the Laplacian prior term in (7) can be written as

$$\tilde{\lambda}\sum_{i=1}^{N-1}|u(t_i)-u(t_{i-1})| = \tilde{\lambda}\|\mathbf{D}\mathbf{u}\|_1, \qquad \text{(A.3)}$$

where $\mathbf{D}$ is defined in (13). Replacing (A.2) and (A.3) in the cost (7) leads to the optimisation problem (12), which concludes the proof. □

Appendix B. PROOF OF LEMMA 2

Proof. By separating the double integral in (11) at the axis $\tau = \xi$, we can decompose $\mathbf{K}_{ij}$ $(i, j = 1, 2, \ldots, N)$ into $\mathbf{A}_{ij} + \mathbf{A}_{ji}$, where

$$\begin{aligned}\mathbf{A}_{ij} &= \int_0^\infty\int_0^\xi u(ih-\xi)u(jh-\tau)k(\xi,\tau)\mathrm{d}\tau\mathrm{d}\xi\\ &=\sum_{r=0}^{q-1}\gamma_{q,r}\int_0^\infty u(ih-\xi)e^{-\beta(2q-r-1)\xi}\int_0^\xi u(jh-\tau)e^{-\beta r\tau}\mathrm{d}\tau\mathrm{d}\xi. \qquad \text{(B.1)}\end{aligned}$$

We first compute the inner integral. By exploiting the zero-order hold representation (valid for $t \in (0, Nh)$)

$$u(t)=\sum_{k=0}^{N-1}u(kh)\big(\mu(t-hk)-\mu(t-h[k+1])\big) \qquad \text{(B.2)}$$

with $\mu(\cdot)$ being the Heaviside function, we can write

$$\int_0^\xi u(jh-\tau)e^{-\beta r\tau}\mathrm{d}\tau = \sum_{k=0}^{N-1}u(kh)g_{j-k}(\xi),$$

where

$$g_{j-k}(\xi)=\begin{cases}(r\beta)^{-1}\big(e^{-r\beta h[j-k-1]} & \text{if } j>k \text{ and } \xi\geq h[j-k-1],\\ \quad -e^{-r\beta\min\{h[j-k],\xi\}}\big) & \\ 0 & \text{otherwise.}\end{cases}$$

Note that the case $r = 0$ can be easily extracted from $\lim_{r\to 0} g_{jk}(\xi)$. Therefore,

$$u(ih-\xi)\int_0^\xi u(jh-\tau)e^{-\beta r\tau}\mathrm{d}\tau=\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}u(lh)u(kh)f_{i-l,j-k}(\xi),$$

where

$$\begin{aligned}f_{i-l,j-k}(\xi)&=g_{j-k}(\xi)\big(\mu(h[i-l]-\xi)-\mu(h[i-l-1]-\xi)\big)\\ &=\begin{cases}(r\beta)^{-1}(e^{-r\beta h[i-l-1]}-e^{-r\beta\xi}) & \text{if } i-l=j-k>0,\\ \times\big(\mu(\xi-h[i-l-1])-\mu(\xi-h[i-l])\big) & \\ (r\beta)^{-1}(e^{-r\beta h[j-k-1]}-e^{-r\beta h[j-k]}) & \text{if } i-l>j-k>0,\\ \times\big(\mu(\xi-h[i-l-1])-\mu(\xi-h[i-l])\big) & \\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Inserting this result into (B.1) and integrating, we obtain

$$\mathbf{A}_{ij} =\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\sum_{r=0}^{q-1}u(kh)u(lh)\gamma_{q,r}F_{i-l,j-k}(\beta)$$

where

$$F_{i-l,j-k}(\beta) = \int_{h[i-l-1]}^{h[i-l]} f_{i-l,j-k}(\xi)e^{-\beta(2q-r-1)\xi}\mathrm{d}\xi. \qquad \text{(B.3)}$$

By writing $\mathbf{A}_{ji}$ in similar fashion, we reach the following description for $\mathbf{K}_{ij}$:

$$\mathbf{K}_{ij}=\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\sum_{r=0}^{q-1}u(kh)u(lh)\gamma_{q,r}[F_{i-l,j-k}(\beta)+F_{j-k,i-l}(\beta)].$$

Next, note that $F_{i-l,j-k}(\beta) + F_{j-k,i-l}(\beta) = 0$ if $l \geq i$, and also if $k \geq j$. Furthermore, for any integers $k \in \{0, 1, \ldots, j-1\}$ and $l \in \{0, 1, \ldots, i-1\}$ that satisfy $i - l \neq j - k$, we have $F_{i-l,j-k}(\beta) \neq 0$ if and only if $F_{j-k,i-l}(\beta) = 0$. Thus, we may write $\mathbf{K}_{ij}$ as

$$\mathbf{K}_{ij} =\sum_{k=0}^{j-1}\sum_{l=0}^{i-1}u(kh)u(lh)\tilde{F}_{i-l,j-k}(\beta),$$

with $\tilde{F}_{i-l,j-k}(\beta)$ being given by

$$\tilde{F}_{i-l,j-k}(\beta)=\sum_{r=0}^{q-1}\gamma_{q,r}\begin{cases}2F_{i-l,j-k}(\beta) & \text{if } i-l=j-k,\\ F_{\max\{i-l,j-k\},\min\{i-l,j-k\}}(\beta) & \text{if } i-l\neq j-k.\end{cases}$$

Alternatively, we can write this entry of the kernel matrix as $\mathbf{U}_j^\top \mathcal{O}_\beta \mathbf{U}_i$, where $\mathbf{U}_j$, $\mathbf{U}_i$ are the jth and ith columns of $\mathbf{\Phi}^\top$ respectively, and $\mathcal{O}_\beta$ has entries that are given by $\tilde{F}_{i,j}(\beta)$. Expanding $\tilde{F}_{i,j}(\beta)$ by computing (B.3) returns the conditional expression in (19). Since $\mathcal{O}_\beta$ does not depend on i nor j, we can describe the matrix $\mathbf{K}$ by stacking the vectors $\mathbf{U}_j$ and $\mathbf{U}_i$, leading to (17). □

Appendix C. PROOF OF THEOREM 4

Proof. To obtain the transfer function description we can apply the Laplace transform directly to both sides of (14):

$$\hat{G}(s,\hat{\mathbf{u}}) = \sum_{l=1}^{N}\hat{G}_l(s,\hat{\mathbf{u}})\hat{c}_l, \qquad \text{(C.1)}$$

where $\hat{c}_l$ is the lth element of the optimal vector from (16). The transfer function $\hat{G}_l(s, \hat{\mathbf{u}})$ is computed by

$$\hat{G}_l(s,\hat{\mathbf{u}}) = \int_0^\infty K(s,\tau)\hat{u}(lh-\tau)\mathrm{d}\tau.$$

If we take into consideration that the input is under a zero-order-hold assumption (recall Eq. (B.2)), we then have

$$\hat{G}_l(s,\hat{\mathbf{u}}) = \sum_{k=0}^{l-1}\left(\hat{u}(kh)\int_{h[l-k-1]}^{h[l-k]}K(s,\tau)\mathrm{d}\tau\right).$$

The Laplace transform of the stable spline kernel of order q can be computed (see, e.g., Scandella et al. (2022)) as

$$K(s;\tau)=\sum_{r=0}^{q-1}\frac{\gamma_{q,r}e^{-\beta(2q-r-1)\tau}}{s+r\beta}+\frac{(-1)^q\beta^{2q-1}e^{-(s+\beta[2q-1])\tau}}{\prod_{k=0}^{2q-1}(s+k\beta)},$$

where $\gamma_{q,r}$ is given in (9). By integrating this expression with respect to τ we obtain $\hat{\mathbf{G}}_l(s,\hat{\mathbf{u}}) = \sum_{k=0}^{l-1}\hat{u}(kh)\mathcal{K}_{l-k}(s)$, where $\mathcal{K}_{l-k}(s)$ is defined in (23). In other words, we have shown that $\hat{\mathbf{G}}_l(s,\hat{\mathbf{u}})$ is simply the lth row of $\hat{\mathbf{\Phi}}$ multiplied by $\mathcal{K}$. Replacing this expression in (C.1) and later expanding $\hat{\mathbf{c}}$ in (16) and rearranging, we reach (22). □