International Journal of Control

ISSN: 0020-7179 (Print) 1366-5820 (Online) Journal homepage: https://www.tandfonline.com/loi/tcon20

Continuous-time model identification: application on a behavioural (miLife) study

Korkut Bekiroglu, Michael A. Russell, Constantino Lagoa, Rong Su, Mario Sznaier, Stephanie T. Lanza & Candice L. Odgers

To cite this article: Korkut Bekiroglu, Michael A. Russell, Constantino Lagoa, Rong Su, Mario Sznaier, Stephanie T. Lanza & Candice L. Odgers (2020): Continuous-time model identification: application on a behavioural (miLife) study, International Journal of Control, DOI: 10.1080/00207179.2019.1706101

To link to this article: https://doi.org/10.1080/00207179.2019.1706101

Accepted author version posted online: 17 Dec 2019.
Published online: 06 Jan 2020.

INTERNATIONAL JOURNAL OF CONTROL
https://doi.org/10.1080/00207179.2019.1706101

Taylor & Francis
Taylor & Francis Group

Continuous-time model identification: application on a behavioural (miLife) study

Korkut Bekiroglu [a], Michael A. Russell[b], Constantino Lagoa[c], Rong Su[d], Mario Sznaier[e], Stephanie T. Lanza[f] and Candice L. Odgers[g]

[a]Electrical Engineering Technology, SUNY Polytechnic Institute, Utica, NY, USA; [b]Department of Biobehavioral Health and The Methodology Center, Penn State University, University Park, PA, USA; [c]Department of Electrical Engineering, Penn State University, University Park, PA, USA; [d]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore; [e]Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA; [f]Department of Biobehavioral Health and The Methodology Center to Department of Biobehavioral Health and the Edna Bennett Pierce Prevention Research Center, The Pennsylvania State University, University Park, PA, USA; [g]Sanford School of Public Policy, Duke University, Durham, NC, USA

ABSTRACT

To develop efficient just-in-time personalised treatments, dynamical models are needed that provide a description of how an individual responds to treatment. However, available system identification approaches cannot effectively be applied to most behavioural datasets since, usually, the data collected is subjected to a large amount of noise and time sampling is not uniform. To be able to circumvent these issues, in this paper a new method is proposed for parsimonious system identification of continuous-time systems that does not require specially structured data. The developed algorithm provides an effective way to leverage these 'non-standard' datasets to identify continuous time dynamical models that are compatible with a-priori information available on the process. The algorithm developed is tested on data obtained from a behavioural study on adolescents and violence. The objective is to model the temporal dynamics of the association between violence exposure and mental health symptoms (depression and anxiety) in day-to-day life among a sample of adolescents at heightened risk for both substance use exposure and problem behaviour. The information extracted from individual models of behaviour such as the maximum burden and the time of fading away of depression/anxiety does differ substantially from person to person. This information has the potential to be useful to design personalised interventions that would have a better chance of succeeding.

ARTICLE HISTORY
Received 2 April 2019
Accepted 13 December 2019

KEYWORDS
Parsimonious continuous time system estimation; non-uniformly sampled data; dynamical modelling for behavioural problems; atomic norm minimisation

1. Introduction

1.1 Motivation

Evaluating and modelling instantaneous changes in human behaviour is a long-standing problem which, if successfully addressed, can lead to more effective treatments for social and behavioural problems. This problem is a difficult one because of the complex and nonlinear structure of human behaviour and the characteristics of the data collected. Human behaviour data that is usually collected is not, in our view, sufficient to determine a complete and exact model. This being said, 'simple' linear models obtained from such data can give important clues about individual behaviour and can inform on how to design an effective personalised treatment regime. Obviously, a behavioural treatment based on these models is not optimal. However, give the 'coarse nature' of most available treatments (meaning not many options for intervention), such models can in most cases be 'good enough,' in the sense that a precise model would not result in a much different treatment (Deshpande, Nandola, Rivera, & Younger, 2014; Rivera, 2012). At least, it is possible to design behavioural treatments that are better than the current state of the art. Furthermore, traditional approaches to the analysis of such datasets in the social and behavioural sciences are limited to the use of 'off the shelf' statistical tools – such as multilevel modelling (MLM) or generalised estimating equations (GEE). With the assumption that the study group is a representative sample of the population, the findings are assumed to be applicable to a large set of individuals, leading to a 'one size fits all' approach to treatment development. However, more recent methods are able to collect large amounts of data from individuals. Hence, it is now possible to develop *personalised models* of behaviour which can be leveraged by clinicians to improve the way treatment decisions are made. For instance, if the model of behaviour for an alcohol user is known, then the association of exposures (e.g. treatment, violence) with alcohol use behaviour can be accurately evaluated, and the best treatment for this specific person can be chosen among the set of all possible treatments. However, to be able to determine dynamical models of social and behavioural problems, one needs to be able to extract information from datasets that do not conform with the needs of most available system identification procedures. To illustrate this concept, we used the following behavioural problem: the association of violence and depression/anxiety in young adolescents.

1.1.1 Real data implementation

The proposed impulse response estimation method is used to investigate the association of violence and depression/anxiety in

CONTACT Korkut Bekiroglu korkutbekiroglu@gmail.com Electrical Engineering Technology, SUNY Polytechnic Institute, Utica, NY, USA

young adolescents (miLife study, Odgers & Russell, 2017; Russell, Wang, & Odgers, 2016). Knowing the level and fading time of depression and anxiety that is caused by violent events can give crucial information for personalised intervention design. In particular, such information can be exploited to develop efficient individually tailored treatment regimes for adolescents that are exposed to violence.

As one might expect, the results in this paper reveal that the different adolescents have different behavioural responses when exposed to such traumatic events. Therefore individual response, as modelled by an impulse response to violent events, can be used to classify the adolescents. In Section 6, the associations of violence and depression/anxiety for 45 adolescents are investigated. In this study, the patients were contacted via telephone, a method of sampling the individual response that results in having non-uniformly spaced samples of the levels of violence and depression. Such non-uniformly sampled data is not suitable for discrete time system identification methods. One could use available interpolation based methods to approximate missing data. However, as the results in this paper show, this type of methods often fail if the data is not sampled sufficiently fast. Therefore, to be able to identify meaningful dynamical models that are consistent with non-uniformly sampled data, one needs a new approach that directly uses this fragmented data to determine the models.

1.2 Literature review on continuous-time system identification

Having a temporally dense, uniformly sampled dataset for dynamical modelling is not an issue if one has a sensor able to collect evenly sampled data. However, in spite of recent technological developments, collecting uniformly sampled medical/behavioural data for some processes is not always possible. For instance even though collecting uniformly sampled physical activity data with a sensor such as a smartphone, fit-bit is feasible (Ashour et al., 2016), collecting smoking behaviour data such as the smoked number of cigarettes per day is still not attainable with a sensor (e.g.Bekiroglu, Lagoa, Murphy, & Lanza, 2016; Lagoa, Bekiroglu, Lanza, & Murphy, 2014). The challenges that come from the nature of such behavioural data can be addressed, however, through the use of continuous time system identification methods. Motivated by these challenges, a computationally efficient algorithm to the problem of low order continuous time system identification is proposed.

Classical linear continuous-time system identification algorithms fall into two general categories (Rao & Unbehauen, 2006): (i) *The indirect approach*: Initially the discrete time model is identified with uniformly sampled input/output data (Yoshimura, Matsubayashi, & Inoue, 2019) and then the identified model is transformed (e.g. bilinear transformation) to the continuous time model, see, e.g.(Ohta & Kawai, 2004). Even though transforming a discrete time model to a continuous time model is theoretically possible, such process causes a numerical problem when the sampling rate is high. Moreover, this type of algorithms requires uniformly sampled data. (ii) *The direct approach*: The derivative is approximated to construct an input-output estimation of a continuous time system from discrete datasets. In other words, a state variable filter or a causal, stable, realisable linear operator is employed to approximate the higher order derivatives, see, e.g.Garnier (2011); Johansson (2009); Mercère, Ouvrard, Gilson, Garnier (2007); Young (1981). In Garnier (2011), it is shown that this method could handle non-uniformly sampled data if filtered data can be sampled uniformly. In contrast to the direct approach, our method only uses the available non-uniformly sampled data without any filtering or derivative estimation. This data is directly used to identify the impulse response of the continuous-time system. In addition to these works, ARMA model estimation from irregular sampled data is presented in Chen, Agüero, Gilson, Garnier, and Liu (2017). In this approach, initial state is assumed to have a normal distribution to estimate continuous time system models. Our method does not need this type of assumption. We directly estimate the impulse response of continuous time system from irregularly sampled data. Also, the methods proposed in Yuz, Alfaro, Agüero, and Goodwin (2011) and Banbura, Giannone, Modugno, and Reichlin (2010) require fast-sampled data that is not the case for the proposed method. This is also not the case in most of the behavioural problems.

Furthermore, most of the results available on parsimonious model identification are for discrete time models (see, e.g. Bühlmann & Van De Geer, 2011; Godoy, Agüero, Carvajal, Goodwin, & Yuz, 2014; Yilmaz, Bekiroglu, Lagoa, & Sznaier, 2017 and references therein). Recently published book (Bühlmann & Van De Geer, 2011) also present many lasso-based statistical modelling methods where the model coefficients are sparsified. In the proposed work, the number of exponentials used in the impulse response (i.e. order of the model) are sparsified. We are directly enforcing parsimonious dynamical model. Moreover, *low order continuous time model identification* research is still limited in the literature. Recently a study addressed the same problem for continuous time system if the data is sampled uniformly (Yue, Thunberg, Ljung, & Goncalves, 2016). In contrast to Yue et al. (2016), this article introduces an algorithm to address this problem without having limitations on uniform sampling. The preliminary versions of some of the ideas in this paper have appeared in the conference paper (Bekiroglu, Lagoa, Lanza, & Sznaier, 2017). Additional results in here include an analysis of the optimal sampling rate of the proposed algorithm, the extension of the basic algorithm to handle initial conditions, not trivial because of the non-uniformly sampled data. Finally a real data-set miLife study is analysed with the proposed method (Odgers & Russell, 2017; Russell et al., 2016).

1.3 Contributions

The main contributions of the paper can be summarised as

(i) Relaxation of an NP-hard low-order continuous time system identification algorithm using atomic norm concept is introduced.
(ii) Randomised Frank–Wolfe algorithm for the discrete-time system in Yilmaz et al. (2017) is modified to solve the relaxed continuous time system identification problem.
(iii) The method is tested on an miLife study to evaluate how adolescents are affected by violence. Based on this 45

patients' data, it is shown that each person might have different level of depression or anxiety for violence. However, we should note that this method should be tested on a richer data set to make a more concrete result.

1.4 The sequel

The model specification and system identification problem for zero and non-zero initial condition cases, corresponding assumptions and optimisation problems are introduced in Section 2. Then the convex relaxation of continuous time parsimonious system identification problem is given in Section 3. What follows is the randomised Frank–Wolfe algorithm to solve this relaxed problem in Section 4. Finally, paper ends with examples in Sections 5 and 6, and concluding remarks in Section 7.

2. Methods

2.1 Model specification

Assuming the initial condition is equal to zero and the observations are corrupted by bounded noise, output of a single input single output (SISO) linear time invariant (LTI) continuous time system Ω can be represented as

$$\mathbf{y}(t_i) = \int_0^{t_i} \mathbf{C}e^{\mathbf{A}(t_i-\tau)}\mathbf{B}\mathbf{u}(\tau)\,\mathrm{d}\tau + \varepsilon(t_i), \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times 1}$ and $\mathbf{C} \in \mathbb{R}^{1\times n}$ represent state, input and output matrices respectively. Also $\mathbf{u}$ represents the input signal, and $\mathbf{y}(t_i)$ is output at time t_i corrupted by some bounded noise ε (e.g. $\|\varepsilon\|_2 \leq \Xi_{\max}$). The non-uniform sampling instants t_i are collected in the following vector:

$$\mathbf{t} = \{t_0, t_1, t_2, \ldots, t_m\}, \quad \text{where } t_0 < t_1 < t_2 < \cdots < t_m.$$

Remark 1: The sequence of times t_i above can be any strictly increasing sequence. No assumptions on the sampling times, like t_i commensurate, are made.

2.2 System identification problem

In this section, we start by assuming that the initial conditions are zero and define the system identification problem for strictly proper SISO LTI systems. Note that this problem is later extended the non-zero initial conditions case. We start by stating the assumptions that are made on the system and signals.

(A.1) Convolution of any exponential with input signal $\mathbf{u}$ (continuous time) can be computed in closed form and/or numerically.

Example 1: If the transfer function (closed-form description) of input signal $U(s)$ is known, then the output $\mathbf{y}(t_i)$ could be calculated with the inverse Laplace transform of this multiplication $\mathbf{y}(t_i) = \mathscr{L}^{-1}(G(s)U(s))$. For the step response of $U(s) = 1/s$ and transfer function $G(s) = 1/(s-p)$, time response of output is simply

$$\mathbf{y}(t_i) = \mathscr{L}^{-1}\left(\frac{1}{s}\frac{1}{s-p}\right) = \frac{e^{pt_i}-1}{p},$$

where p is a pole and $\mathscr{L}^{-1}$ is the inverse Laplace transform.

Note that this assumption is a reasonable one for medical/behavioural processes since most of the inputs in these applications are modelled as combinations of impulses or steps which satisfy the assumption above.

(A.2) The system to be identified does not have repeated poles.[1]

(A.3) The initial conditions are assumed to be zero.

Remark 2 ((Discussion on Input Assumption)): Given the fact that the main objective is to identify the model of the system from sparse non-uniformly sampled noisy data, we need (i) to have complete knowledge of the input applied and (ii) to be able to do elementary calculations with this input signal. The assumption above allows for fast and precise computation of the quantities involved in the proposed algorithm.

Problem 1: Considering an SISO LTI continuous-time system Ω and assume that

- the poles of the system to be identified belong to a compact set $\mathbb{S}_\rho = \{p \in \mathbb{C} \text{ or } p \in \mathbb{R} : -\rho < Re(p) < 0 \text{ and } -\pi/T_s < Im(p) < \pi/T_s\}$ (a rectangular region in left-half plane) (Figure 1),
- input signal $\mathbf{u}(t)$ and output noisy measurements $\mathbf{y}(t_i)$ are available, and
- the measurement noise ε is bounded (e.g.$\|\varepsilon\|_2 \leq \Xi$).

Then the problem is to identify the lowest order model $\tilde{\Omega}$ that is the estimation of true plant Ω while explaining available input–output data within the given bound on estimation error.

Remark 3 ((On the set $\mathbb{S}_\rho$)): The poles of the system Ω are assumed to belong to the compact set $\mathbb{S}_\rho$ that contains only stable poles but not necessarily asymptotically stable poles. This set can be defined based on a-priori information on bounds on the time constant of the system. We should remark that this set is only one possible choice. Other types of a-priori information on the system (such as bounds on overshoot, rise-time, etc .) can be used to determine the 'right' set $\mathbb{S}_\rho$ for the specific system identification problem to be solved. The only restriction is that it should be compact. In principle, the proposed approach can also be used for the identification of unstable plants but we have found that exponential growth of the impulse response can lead to numerical problems.

2.2.1 Optimisation for Problem thm1

The impulse response of any finite dimensional proper LTI system (1) with non-repeated poles in the compact set $\mathbb{S}_\rho$ can be represented as linear combination of the impulse responses of first-order systems (sum of exponentials) by using partial fraction expansion. Therefore, the system identification Problem 1

can be formulated as follows:

$$\min_{c_p\in\mathbb{C},p\in\mathbb{S}_\rho} \quad \text{cardinality}\{c_p : c_p \neq 0\} \tag{2a}$$

$$\text{s.t.} \quad \sum_{i=0}^{m}\left[\sum_{p\in\mathbb{S}_\rho}[c_p e^{pt} * \mathbf{u}(t)]\Big|_{t=t_i} - \mathbf{y}(t_i)\right]^2 \leq \Xi_{\max}, \tag{2b}$$

$$c_{p^*} = c_p^*, \tag{2c}$$

where while the objective function minimises the number of non-zero elements of vector c_p (minimum number of poles or exponentials), the constraint imposes fidelity to the data collected. The constraint in (2b) enforces a real valued impulse response. Note that the number of exponential used to estimate the impulse response is equal to the system order. Therefore minimising the vector c_p is equal to minimising the order of linear dynamical systems.

2.2.2 Extension of Problem thm1 to non-zero initial condition

In this section, Problem 1 and its optimisation problem formulation is extended for non-zero initial condition case.

Problem 2: Considering the scenario in Problem 1, the problem is again to identify the lowest order model $\tilde{\Omega}$ when initial conditions are not necessarily zero.

In this case, output of the system can be represented as

$$\mathbf{y}(t_i) = \mathbf{C}e^{\mathbf{A}t_i}\mathbf{x}_0 + \int_0^{t_i}\mathbf{C}e^{\mathbf{A}(t_i-\tau)}\mathbf{B}\mathbf{u}(\tau)\mathrm{d}\tau + \varepsilon(t_i), \tag{3}$$

where $\mathbf{x}_0 \in \mathbb{R}^{n\times 1}$ is the vector of initial conditions. As before, the response of the system can be represented as a finite sum of exponentials, i.e.

$$\mathbf{C}e^{\mathbf{A}t_i}\mathbf{x}_0 = \sum_{j=1}^{n} c_{p_j}^{ic} e^{p_j t_i} : p_j \in \mathbb{S}_\rho, \quad c_{p_j}^{ic} \in \mathbb{C},$$

$$\text{where } c_{p_j}^{ic} = \mathbf{C}(j)\mathbf{x}_0(j)$$

$$\mathbf{C}e^{\mathbf{A}t_i}\mathbf{B} = \sum_{j=1}^{n} c_{p_j}^{u} e^{p_j t_i} : p_j \in \mathbb{S}_\rho, \quad c_{p_j}^{u} \in \mathbb{C},$$

where $c_{p_j}^{u} = \mathbf{C}(j)\mathbf{B}(j)$ and $\mathbf{C}(j)$ is the jth element of the row vector $\mathbf{C}$ and $\mathbf{B}(j)$ is the jth element of the column vector $\mathbf{B}$.

Remark 4: Note that the same sparsity pattern must be enforced in $\mathbf{C}$ and $\mathbf{CB}$ in order to have the same poles in both initial condition and input responses. This is later implemented using the concept of block sparsity.

2.2.3 Optimisation for Problem thm2

Non-zero initial conditions case can be formulated as

$$\min_{c_p^{ic},\, c_p^{u}\in\mathbb{C},\, c_p\in\mathbb{R},\, p\in\mathbb{S}_\rho} \quad \text{cardinality}\{c_p : c_p \neq 0\} \tag{4a}$$

$$\text{s.t.} \quad \sum_{i=0}^{m}\left[\sum_{p\in\mathbb{S}_\rho}\left\{c_p^{ic}e^{pt} + \left[c_p^{u}e^{pt} * \mathbf{u}(t)\right]\right\}\Big|_{t=t_i} - \mathbf{y}(t_i)\right]^2 \leq \Xi_{\max} \tag{4b}$$

$$\max\{|c_p^{ic}|,\ |c_p^{u}|\} \leq c_p \text{ for all } p \in \mathbb{S}_\rho \tag{4c}$$

$$c_{p^*}^{u} = c_p^{*u} \quad \text{and} \quad c_{p^*}^{ic} = c_p^{*ic}. \tag{4d}$$

Similar to Problem (2a), while the objective is to minimise the cardinality of vector $\mathbf{c}_p$, the constraint (4b) imposes a certain fidelity for a given bound on noise and the constraint in (4c) enforces a block sparsity on vectors $c_{p_j}^{ic}$, $c_{p_j}^{u}$ (having same sparsity pattern – Remark 4) that imposes the use of the same poles for initial condition and impulse responses of the system.

This optimisation problem gives the most parsimonious model under the given constraints. However, there is no available method in the literature to solve this non-convex optimisation problem since cardinality minimisation is an NP-hard problem (Donoho, 2006). Furthermore there are infinitely many poles in the admissible pole set $\mathbb{S}_\rho$. Therefore, we introduce a convex relaxation of the optimisation problems (2a) and (4a) that can be efficiently solved. Before closing the section, one should note that the response of e^{pt} is complex value if $p \notin \mathbb{R}$. This might result in a complex valued c_p. To easily address the fact that impulse responses are real and eliminate the complex part from the calculations, in the next sections we propose a new set of atomic impulse responses to be used in the identification algorithm. These include elementary impulse responses from both first- and second-order systems.

3. Convex relaxation

To relax the NP-hard optimisation problems (2a) and (4a), a suitable atomic norm minimisation problem is introduced. As a first step of this atomic norm minimisation problem, atomic sets of elementary impulse and input responses for zero and nonzero initial conditions are defined.

3.1 Set of elementary impulse responses

As a first step towards a convex relaxation of the parsimonious system identification problem defined in the previous section, we introduce a set of elementary impulse responses that will later be used to define an atomic norm minimisation problem. Even though the impulse response of first-order systems utilised in (2a) can in principle be used, the impulse response of these first-order systems has an imaginary part when $Im(p) \neq 0$. This results in numerical problems when trying to approximate the solution of the optimisation problems (2a) or (4a). To address this issue, we propose the following set of real valued elementary

impulse responses

$$\begin{aligned}\mathcal{I}_1 &= \{\pm\alpha_p^1(e^{pt_i} + e^{p^*t_i}) : p \in \mathbb{S}_\rho \text{ and } p \in \mathbb{C}\}\\ \mathcal{I}_2 &= \{\pm\alpha_p^2(-je^{pt_i} + je^{p^*t_i}) : p \in \mathbb{S}_\rho \text{ and } p \in \mathbb{C}\}\\ \mathcal{I}_3 &= \{\pm\alpha_p^3(e^{pt_i}) : p \in \mathbb{S}_\rho \text{ and } p \in \mathbb{R}\},\end{aligned} \tag{5}$$

where real numbers α_p's are scaling factors. Selecting the right scaling is extremely important when working with convex relaxations of cardinality functions, as it has been widely mentioned in the literature on sparsity. However, choosing the 'right' scaling factors is still an open question. Since the scaling factors for discrete time systems introduced in Shah, Bhaskar, Tang, and Recht (2012) and Yilmaz et al. (2017) have shown to generically perform well, we have adapted them for the continuous time problem addressed in this paper. More precisely, we use

$$\begin{aligned}\Gamma &= \frac{Re(\beta_p) - \beta_a - Re\left(e^{2p}e^{2t_m p^*}\beta_p\right) + |e^p|^{2N+2}\beta_a}{1 - |e^p|^2}\\ \alpha_p^1 &= \sqrt{2\left(Re(\beta_p^2) + \beta_a^2\right) + 2\sqrt{2\Gamma\left(|\beta_p|^2 - \beta_a^2\right)}}^{-1}\\ \alpha_p^2 &= \sqrt{2\left(\beta_a^2 - Re(\beta_p^2)\right) + 2\sqrt{2\Gamma\left(|\beta_p|^2 - \beta_a^2\right)}}^{-1}\\ \alpha_p &= (1 - e^{2p})/(1 - e^{p(2t_m+2)}),\end{aligned} \tag{6}$$

where

$$\beta_p = \frac{1 - e^{2t_m p}}{1 - e^{2p}} \quad \text{and} \quad \beta_a = \frac{1 - |e^p|^{2t_m}}{1 - |e^p|^2}$$

and t_m is the time that the final sample is obtained. Furthermore in order to deal with the numerical problems, the terms e^{2p} and $|e^p|^2$ are slightly perturbed when they are equal to 1 to improve numerical stability. Finally, the set of elementary impulse responses is defined as

$$\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3.$$

3.2 Atom set

3.2.1 Atom set for zero initial conditions

The atomic sets for proposed method can now be defined. The set of atoms is actually a collection of the responses of elementary systems, defined in (5), to the input at the sample time t_i. Linear combination of the responses of these elementary systems will form the system's output as seen in (1). More precisely, the atom set for zero initial conditions is defined as

$$\begin{aligned}\mathcal{A} \doteq \{a \in \mathbb{R}^{m+1} :\\ a = [(\mathbf{u} * \mathbf{h})(t_0)\ (\mathbf{u} * \mathbf{h})(t_1) \cdots (\mathbf{u} * \mathbf{h})(t_m)]^T, \mathbf{h} \in \mathcal{I}\},\end{aligned} \tag{7}$$

where $(\mathbf{u} * \mathbf{h})(t_i)$ is the value of the convolution of $\mathbf{u}$ and $\mathbf{h}$ at time t_i. Furthermore, given Assumption A.1, the elements of the set $\mathcal{A}$ are easily computable. Therefore, we can define a map between the elements of $\mathcal{A}$ and $\mathcal{I}$ in closed form.

Definition 1: $\mathcal{K} : \mathcal{A} \to \mathcal{I}$ and satisfies

$$\mathbf{h} = \mathcal{K}(a) \text{ if } a = (\mathbf{h} * \mathbf{u})(t_{i-1}) \text{ for all } i = 1, 2, \ldots, m+1. \tag{8}$$

3.2.2 Atom set for non-zero initial condition

In the case of non-zero initial conditions, the response of the system is the sum of the initial condition and input responses. As it is explained in Section 2.2.3 and (4a), the initial condition and input responses share the same poles. This can be enforced by using block sparsity as explained in Remark 4. However, even though the initial condition and input responses are the response of same poles, the sign and magnitude can be different. This issue can be addressed by having different sign combinations of pole response. Therefore the set of atoms for non-zero initial condition case is slightly different from the set in (7). More precisely, it is of the form

$$\begin{aligned}\bar{\mathcal{A}} \doteq \{a \in \mathbb{R}^{m+1} :\\ a_{ic} = [\check{\mathbf{h}}_{ic}(t_0)\check{\mathbf{h}}_{ic}(t_1) \cdots \check{\mathbf{h}}_{ic}(t_m)]^T,\\ a_u = [(\mathbf{u} * \check{\mathbf{h}}_u)(t_0)(\mathbf{u} * \check{\mathbf{h}}_u)(t_1) \cdots (\mathbf{u} * \check{\mathbf{h}}_u)(t_m)]^T\\ \text{where } \check{\mathbf{h}}_{ic} \in \pm\mathbf{h},\ \check{\mathbf{h}}_u \in \pm\mathbf{h} \text{ and } \mathbf{h} \in \mathcal{I}\\ a = [a_{ic}\ a_u]^T\}.\end{aligned} \tag{9}$$

Again one can define a mapping between the elements of $\bar{\mathcal{A}}$ and $\mathcal{I}$ in closed form.

Definition 2: $\bar{\mathcal{K}} : \bar{\mathcal{A}} \to \mathcal{I}$ and satisfies

$$\mathbf{h} = \bar{\mathcal{K}}(a) \text{ if } a = [\check{\mathbf{h}}_{ic}(t_{i-1})(\mathbf{u} * \check{\mathbf{h}}_u)(t_{i-1})]^T \text{ for all } i = 1, 2, \ldots, m+1.$$

3.3 Convex relaxation of optimisation problem in (eqn2a)

Given the discussion in Section 2.2.1 and the definition of the atom set $\mathcal{A}$, the time domain response of any LTI system whose poles are in $\mathcal{S}_\rho$ can be approximated as

$$\mathbf{y} = [\mathbf{y}(t_0)\ \mathbf{y}(t_1) \cdots \mathbf{y}(t_m)]^T \approx \sum_{a \in \mathcal{A}} c_a a,$$

where equality holds in noise free case and if the system does not have repeated poles. Hence the problem of low order system identification from non-uniformly sampled time domain data can be formulated as finding estimated system input response

$$\tilde{\mathbf{y}} = \sum_{a \in \mathcal{A}} c_a a \tag{10}$$

and corresponding impulse response

$$\mathbf{g} = \sum_{a \in \mathcal{A}} c_a \mathcal{K}(a)$$

such that distance between the estimated $\tilde{\mathbf{y}}$ and the collected output data $\mathbf{y}$ is less than a given bound (e.g. $\|\mathbf{y} - \tilde{\mathbf{y}}\|_2 \leq \Xi_{\max}$), and the number of elements of $\mathcal{A}$ that are used in the expressions (10) is the lowest possible one. Finally to obtain a convex relaxation of (2a), the following atomic norm is considered

$$\|\tilde{\mathbf{y}}\|_{\mathcal{A}} \doteq \left\{\inf \sum_{a \in \mathcal{A}} |c_a| : \tilde{\mathbf{y}} = \sum_{a \in \mathcal{A}} c_a a,\ c_a \in \mathbb{R}\right\}. \tag{11}$$

Given this atomic norm definition and non-uniformly sampled outputs $\mathbf{y} = [\mathbf{y}(t_0)\mathbf{y}(t_1) \cdots \mathbf{y}(t_m)]^T$, a convex relaxation

of (2a) is

$$\begin{aligned} \min_{\tilde{\mathbf{y}}} \quad & \frac{1}{2}\|\mathbf{y}-\tilde{\mathbf{y}}\|_2^2 \\ \text{s.t.} \quad & \|\tilde{\mathbf{y}}\|_{\mathcal{A}} \le \tau \\ & \tilde{\mathbf{y}} = \sum_{a\in\mathcal{A}} c_a a, \end{aligned} \tag{12}$$

where τ is a bound on the atomic norm of output of the system.[2]

3.4 Convex relaxation of (eqn4a)

In this section, the atomic norm minimisation problem is extended for non-zero initial condition case with the help of set $\bar{\mathcal{A}}$ in (9). Given the following vector:

$$\mathbf{c}_a = [c^{ic}\ c^{u}],$$

the output of an LTI system for nonzero initial condition case can be represented as

$$\tilde{\mathbf{y}} = \sum_{a\in\bar{\mathcal{A}}} \mathbf{c}_a a = \sum_{a=[a_{ic}\,a_{\mathbf{u}}]^T\in\bar{\mathcal{A}}} (c^{ic}a_{ic} + c^{u}a_{u})$$

and corresponding impulse response is

$$\mathbf{g} = \sum_{a\in\bar{\mathcal{A}}} c_a \bar{\mathcal{K}}(a).$$

Similar to the problem (4a), this atomic norm definition enforces a block sparsity on vector $\mathbf{c}_a$. Given this definition, the following atomic norm is defined.

$$\begin{aligned} \|\tilde{\mathbf{y}}\|_{\bar{\mathcal{A}}} \doteq \Bigg\{ & \inf \sum_{a\in\bar{\mathcal{A}}} \|\mathbf{c}_a\|_\infty : \tilde{\mathbf{y}} = \sum_{a\in\bar{\mathcal{A}}} c_a a \\ & = \sum_{a=[a_{ic}\ a_{\mathbf{u}}]^T\in\bar{\mathcal{A}}} (c^{ic}a_{ic} + c^{u}a_{u}),\ c^{ic},\ c^{u}\in\mathbb{R} \Bigg\}. \end{aligned}$$

Now a convex relaxation of (4a) is

$$\begin{aligned} \min_{\tilde{\mathbf{y}}} \quad & \frac{1}{2}\|\mathbf{y}-\tilde{\mathbf{y}}\|_2^2 \\ \text{s.t.} \quad & \|\tilde{\mathbf{y}}\|_{\bar{\mathcal{A}}} \le \tau \\ & \tilde{\mathbf{y}} = \sum_{a\in\bar{\mathcal{A}}} \mathbf{c}_a a. \end{aligned} \tag{13}$$

In the next section, an algorithm to solve these relaxed problems is given. Although the problem above is a convex problem, it is a challenging one since, to the best of our knowledge, there is no closed-form expression for the atomic norm used. This is, in part, a consequence of the fact that the atomic set considered is uncountable. There are two main approaches in the literature to this class of problems: (i) grid the atom set and use this as a way to approximate the atomic norm (Shah et al., 2012) or (ii) use a randomised version of the classical Frank–Wolfe algorithm (Jaggi, 2013; Yilmaz et al., 2017). The main drawback of the first approach is that, to have good approximations, the grid has to be 'very dense' leading to complex high dimensional optimisation problems. Hence, in the next section, we provide a modification of the randomised Frank–Wolfe algorithm especially suited for the continuous time parsimonious system identification problem considered in this paper.

4. System identification algorithm

Optimisation Problems (12) and (13) are particular cases of a convex problem with an atomic norm constraint and a Frank–Wolfe type algorithm can be used to solve it (Jaggi, 2013). However, the classical Frank–Wolfe algorithm cannot be directly applied to the problems described in this paper because of the following reasons: (i) the atomic norm used in our formulation, to the best of our knowledge, does not have a computable closed form and (ii) optimisation Problem (12) or (13) provides a non-uniformly sampled estimate of the response of the system which is not suitable for continuous time system identification.

Note that the available measurements are sampled non-uniformly, and the optimisation problem directly uses these measurements to solve the proposed optimisation problem given in (12) or (13). If a classical version of the Frank–Wolfe algorithm is used, the estimated response obtained would be a non-uniformly sampled one that provides a trade-off between sparsity and fitting error. However, this non-uniformly sampled response cannot be directly used in available algorithms to calculate the continuous-time system parameters. These algorithms need uniformly sampled data.

To address these limitations, we propose a modified version of the randomised Frank–Wolfe algorithm. The first concern is already addressed by introducing a randomised search that has a convergent feasible estimate without having to compute the value of the atomic norm at each iteration. In addition, to address the second concern that is explained in previous paragraph, two variables are simultaneously updated at each iteration k. These two variables are

- Estimation of system's output at the (non-uniform) sample times $t_0, t_1, \ldots, t_m$ which is denoted by $\tilde{\mathbf{y}}_k$.
- Uniformly sampled system's impulse response estimation is denoted by $\tilde{\mathbf{g}}_k \doteq [\tilde{\mathbf{g}}_k(0)\ \tilde{\mathbf{g}}_k(T_s)\ \tilde{\mathbf{g}}_k(2T_s)\cdots\tilde{\mathbf{g}}_k(NT_s)]$, where T_s is sampling period.

As it is explained in the previous section, the atoms used for computing estimate $\tilde{\mathbf{y}}_k$ can be easily used to calculate a vector of uniform samples of the impulse response of the system, suitable for model identification. Even though the sample times of the data collected can be arbitrary, the sampling interval T_s used in the definition of $\tilde{\mathbf{g}}_k$ should satisfy some criteria to avoid aliasing, i.e. to uniquely recover the impulse response estimation $\tilde{\mathbf{g}}_k(t)$ from its samples $\tilde{\mathbf{g}}_k(nTs)$, T_s needs to be small enough. However, since the definition of the set $\mathbb{S}_p$ implies that all the poles of the p system to be identified satisfy

$$-\pi/T_s \le Im(p) \le \pi/T_s,$$

and by using classical results on sampling, it can be shown that a sample period of T_s is enough to recover the impulse response from its samples.

Finally, once one has an estimation of the impulse response at evenly sampled times, $[\tilde{\mathbf{g}}_k(0)\ \tilde{\mathbf{g}}_k(T_s)\ \bar{\mathbf{g}}_k(2T_s)\cdots\tilde{\mathbf{g}}_k(NT_s)]$, the transfer function of the continuous time system can be effortlessly identified by using standard algorithms, provided that one has enough number of samples N.

Remark 5 ((On $\tilde{\mathbf{g}}_k(t)$)): Instead of introducing the additional variable $\tilde{\mathbf{g}}_k(t)$, we could alternatively keep track of the atoms being used to approximate the response $\tilde{\mathbf{y}}_k$ and periodically eliminate those with negligible coefficients. We opted to use the extra variable $\tilde{\mathbf{g}}_k(t)$ since it leads to a simple algorithm which has been shown to have very good performance in the experiments that we have conducted.

Algorithm 1 Modified Randomised Frank- -Wolfe Algorithm

Input:

→ Bound on atomic norm τ

→ $\tilde{\mathbf{y}}^{initial} \leftarrow \tau a_0$ for arbitrary $a_0 \in \mathcal{A}$ and $\tilde{\mathbf{g}}^{initial} \leftarrow \tau\tilde{\mathbf{g}}_0$ where $\tilde{\mathbf{g}}_0 = \tau[\mathbf{h}_0(0)\ \mathbf{h}_0(T_s)\cdots\mathbf{h}_0(NTs)]^T$, $\mathbf{h}_0 = \mathcal{K}(a_0)$

→ Measurement y and its sampled time.

1: **for** k = 0,1,2,3,... **do**
2: Randomly pick uniformly $p_k \in \mathbb{S}_\rho$.
3: Let $\mathcal{A}(p_k)$ be set of all atoms with a pole at p_k.
4: $a_k \leftarrow \text{argmin}_{a\in\mathcal{A}(p_k)}\langle\nabla f(\tilde{\mathbf{y}}_k), \pm a\rangle$ ▷ $a_k \in \mathcal{A}$
5: $\mathbf{h}_k \leftarrow \mathcal{K}(a_k)$ ▷ $\mathbf{h}_k \in \mathcal{I}$
6: $\alpha_k \leftarrow \text{argmin}_{\alpha\in[0,1]} f(\tilde{\mathbf{y}}_k + \alpha[\tau a_k - \tilde{\mathbf{y}}_k])$
7: $\tilde{\mathbf{y}}_{k+1} \leftarrow \tilde{\mathbf{y}}_k + \alpha_k[\tau a_k - \tilde{\mathbf{y}}_k]$
8: $\tilde{\mathbf{g}}_{k+1} \leftarrow \tilde{\mathbf{g}}_k + \alpha_k\{\tau[\mathbf{h}_k(0)\ \mathbf{h}_k(T_s)\cdots\mathbf{h}_k(NTs)]^T - \bar{\mathbf{g}}_k\}$
9: **end for**

Output:

→ Uniformly sampled $\tilde{\mathbf{g}}$ and non-uniformly sampled system's output estimation $\tilde{\mathbf{y}}$

Next the proposed modification version of Frank–Wolfe algorithm is presented. Note that the convergence rate of the randomised Frank of algorithm is already proven in Yilmaz et al. (2017). Since the randomisation step in the provided algorithm is same with the one in Yilmaz et al. (2017), Algorithm 1 has same convergence rate.

Remark 6: In this algorithm, atomic set $\mathcal{A}$ and mapping $\mathcal{K}$ in Definition 1 are used since zero initial conditions were assumed. To address the case with arbitrary initial conditions, just replace the atomic set $\mathcal{A}$ and mapping $\mathcal{K}$ by $\bar{\mathcal{A}}$ and $\bar{\mathcal{K}}$ respectively.

The randomised Frank–Wolfe Algorithm 1 is computationally very efficient since:

- the iterations in the algorithm only entail inner products with $\tilde{\mathbf{y}}$ and a and the computation of the gradient with respect to $\tilde{\mathbf{y}}$, a computational burden proportional to the number of sampling times m,
- the estimate of the impulse response $\bar{\mathbf{g}}_k$ does not play any role in any of the computations except by being updated once the atom a_k is chosen,
- the computation of the value of the output of the map $\mathcal{K}$ in step 5 is trivial since we know the pole(s) of a_k.

Remark 7: To identify the continuous time system model, length of the uniformly sampled impulse response signal needs only to be long enough to contain all needed information.[3] Thus only a sufficient number of samples for this impulse response signal need to be stored in the memory for system identification, not the whole time interval spanned by available data.

Next, details on each of the steps of the algorithm are given.

Inputs: A randomly picked pole in $\mathbb{S}_\rho$ is scaled by τ and then used to initialise an output response $\tilde{\mathbf{y}}^{initial}$ and impulse response $\tilde{\mathbf{g}}_0$.

Step 2 A pole is randomly picked in given set $\mathbb{S}_\rho$.

Step 3 The corresponding atom for sampled pole p_k is calculated using the atomic set (7).

Step 4 Then the steepest descent direction is searched for calculated atom. In other words, inner product of the gradient vector $(\nabla f(\tilde{\mathbf{y}}_k) = (\tilde{\mathbf{y}}_k - \mathbf{y})$ for (12)) with randomly chosen $\pm a$ is calculated, and the optimum atom a_k is identified. In here we need to highlight the following fact to show why this algorithm is computationally very efficient.

Remark 8: The algorithm only uses the non-uniformly sampled output $\mathbf{y}$ and its approximation $\tilde{\mathbf{y}}_k$ in these steps. The uniformly sampled atom a_k does not play a role in the calculations.

Step 5 Using mapping $\mathcal{K}$, the $\mathbf{h}_k \in \mathcal{I}$ at iteration k is calculated based on the pole of a_k in step 5.

Step 6 Calculating the optimum step size $\alpha \in [0, 1]$ needs a second-order polynomial minimisation, where the closed-form formula for the optimum α^* is

$$\alpha^* = \max(0, \min(\alpha_u, 1)) \text{ where}$$
$$\alpha_u = \frac{(\tilde{\mathbf{y}}_k - \mathbf{y})^T(\tau a_k - \tilde{\mathbf{y}}_k)}{(\tau a_k - \tilde{\mathbf{y}}_k)^T(\tau a_k - \tilde{\mathbf{y}}_k)}. \tag{14}$$

Step 7 Current non-uniformly sampled $\tilde{\mathbf{y}}$ is updated.

Step 8 Uniformly sampled impulse response $\tilde{\mathbf{g}}$ is updated.

As a final step for proposed method, one of the available Ho's algorithm (or subspace methods) can be employed to obtain a state space realisation of the system from estimated uniformly sampled impulse response $\tilde{\mathbf{g}}$.

5. Academic examples

In this section, academic examples are provided to illustrate the performance of the proposed method and benchmark comparison. An artificially generated example is first studied in detail. Then, the interpolation ability of the method on a sparsely and non-uniformly sampled data is compared with the benchmarks. Finally estimation and sparsification performance of the method is evaluated based on 200 different simulations for step and impulse responses.

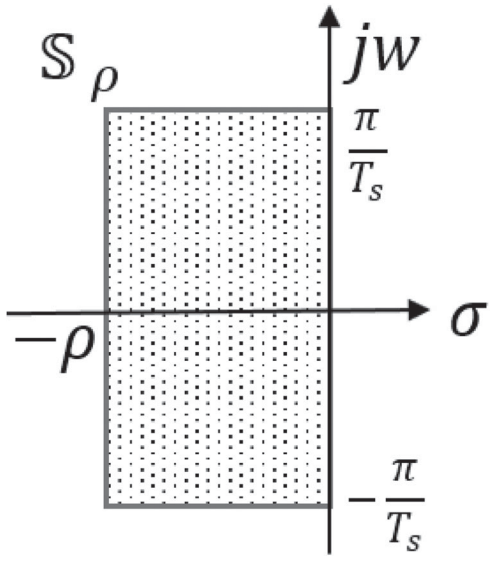

Figure 1. Set $\mathbb{S}_\rho$.

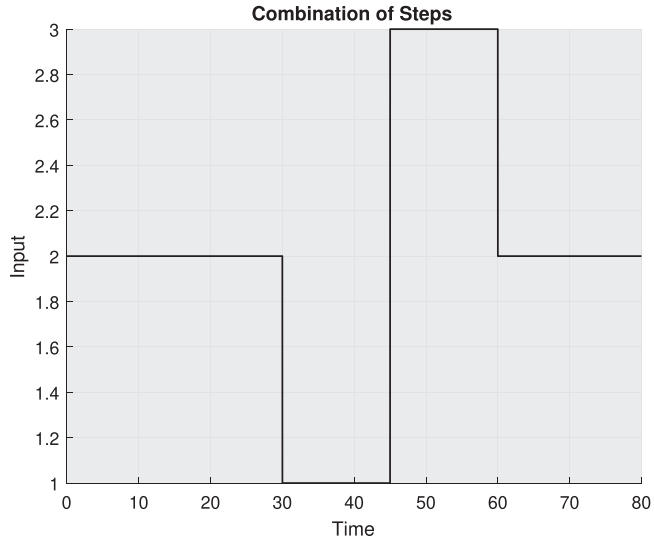

Figure 2. Combination of Steps – Treatment.

5.1 Combination of step response example

The performance of the proposed method is illustrated with an academic example. In this example, a randomly generated system $(0.4802s + 0.09276)/(s^2 + 0.3863s + 7.6628)$ is excited with the input given in Figure 2. Then the output is contaminated with Gaussian noise $\mathcal{N}(0, 0.035 \times \max(\mathbf{y}))$ and the measurements used for impulse response estimation are random non-uniform samples of this noisy output. The signal to noise ratio is calculated by using Matlab SNR function and it is 7.86. To evaluate the performance of the method, the true, measurement and estimated output of the system and its impulse response estimation for the given input is depicted in Figure 3. In addition, the variance accounted for (VAF) is calculated as a metric for both estimated output and estimated impulse response. The formula for the VAF is

$$VAF = \left(1 - \frac{var(y_{true} - y_{est})}{var(y_{true})}\right) * 100,$$

where *var* represent variance. VAF of output is 98.7 and VAF of impulse response is 99.5 for this specific example. The running time of Algorithm 1 is 2.78 s. The sampling rate for impulse response is set at 0.01. In this example, we found the bound on atomic norm $\tau = 2.4$ by trial and error. Because of the atomic norm definition, there is a trade of between noise and parsimony of the system. If one decreases the value of τ, the resulting fitting error is larger (objective of the optimisation problem in (13)) but a lower dimensional system is obtained (a system with a fewer number of poles). If one increases the value of τ, more complex systems are allowed resulting in a better fitting error with a system of higher order. There is no known systematic way for selecting τ, therefore this needs further research. Finally, we choose $\rho = 4$ for this simulation and $T_s = 0.01$ s.

5.1.1 Effect of τ selection

The effect of τ is investigated under the same conditions that we used for combination of step response example and the problem was solved for $\tau \in [0.1, 0.2, 0.3, \ldots 5]$. The error between true

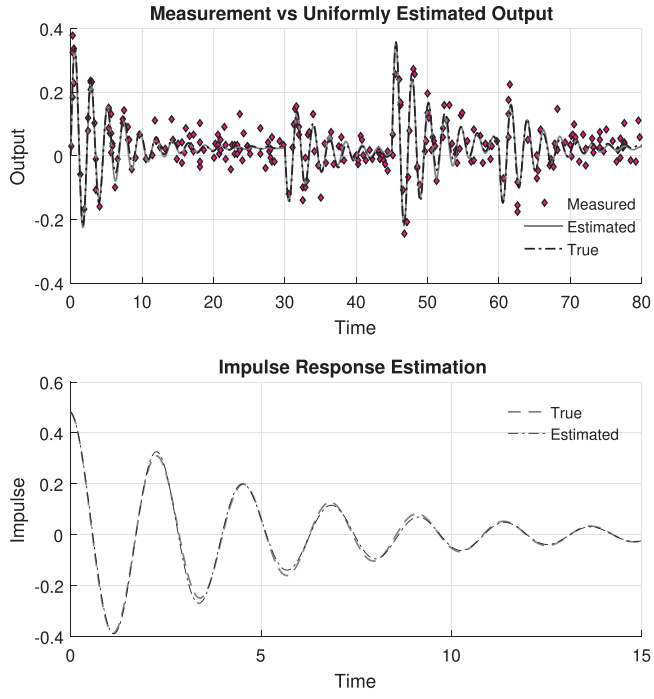

Figure 3. Measurement versus uniformly sampled estimation of output and Impulse response estimation.

impulse response and estimated impulse response is measured with $\|\mathbf{g}_{true} - \tilde{\mathbf{g}}\|_2 / \|\mathbf{g}_{true}\|_2$. Also we used $1/(\sum_{i>n} \sigma(i)/\sigma(1))$ as a sparsity measure where $\sigma(i)$ is the normalised singular values vector of the Hankel matrix of the estimated impulse response. Figure 4 depicts the influence of τ in the results obtained.

5.2 Benchmark comparison performance on interpolation

5.2.1 Interpolation

Non-uniformly sampled step response of $(0.5867s + 0.14116)/(s^2 + 0.4829s + 11.92)$ for 20 s with 25 sampled data is contaminated with Gaussian noise $\mathcal{N}(0, 0.008 \times \max(\mathbf{y}))$. The τ

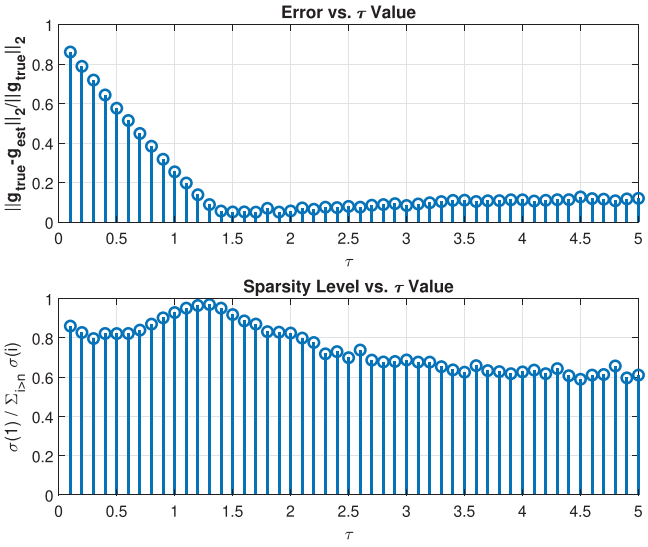

Figure 4. Effect of τ value on impulse response estimation and sparsity.

is again chosen as $\tau = 2.93$ by trial and error. Different types of interpolation methods are available in MATLAB toolboxes. However, most of the methods in toolboxes such as *misdata* function assume commensurate measurement times and the missing measurements are treated as *NaN*. This is not the case in problems involving behavioural data since variables are not sampled uniformly. MATLAB has some functions to deal with such types of data. In this example, the functions *resample* and *interp*1 are employed. In these functions, even though the user can specify interpolation method such as spline, the function itself can select best method for given data. Therefore we leave functions *resample* and *interp*1 to choose best method to interpolate given data. Note that if the interpolation works well, one of the indirect method introduced in introduction can be employed for continuous time system identification. The results of *resample* and *interp*1, and proposed method are given in Figure 5.

5.2.2 Direct approach – CONSTID

In this section, same example is tested with a mature continuous time system identification toolbox CONSTID (Garnier, Wang, & Young, n.d., Chapter 9) that can handle non-uniformly data directly. In this toolbox, there are five different functions which are LSSVF (LS-based state-variable filter (SVF) method for CT ARX models), IVSVF (IV-based SVF method for CT ARX models), COE (non-linear optimisation method for CT hybrid OE models), SRIVC (optimal instrumental variable method for CT hybrid OE models) and SIDGPMF (subspace-based generalised Poisson moment functionals method for CT). Even though some of the functions has some limitations, four of them are tested in this comparison. For instance, as it is stated in Garnier et al. (n.d., Chapter 9) lssvf is very vulnerable to noisy data. Therefore, ivsvf is advised for noisy data. However, ivsvf and coe require the cut-off frequency of the SVF. Finally, the srivc is suggested to overcome these drawbacks. For this specific example, we start with 25 sampled data, and CONSTID fails because of inadequate data. Then we gradually increased the amount of sampled data from 25 to 50. Even though it identify a system

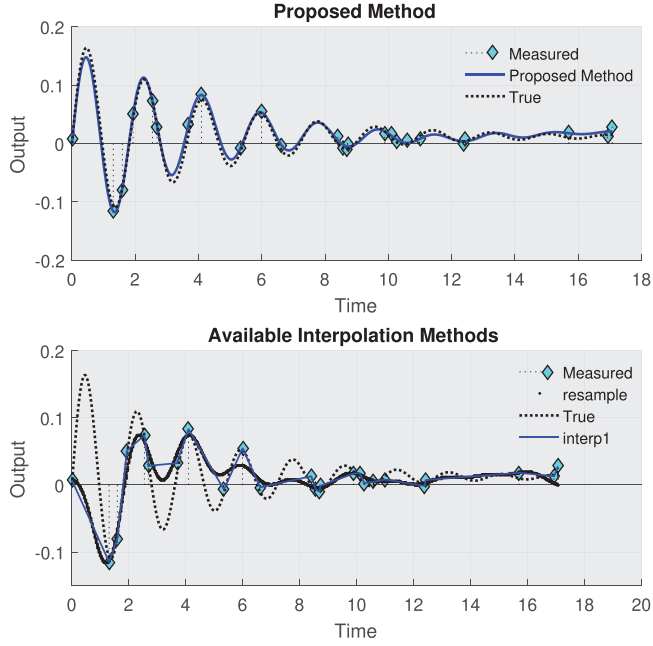

Figure 5. Interpolation from *resample* and *interp*1 and proposed method.

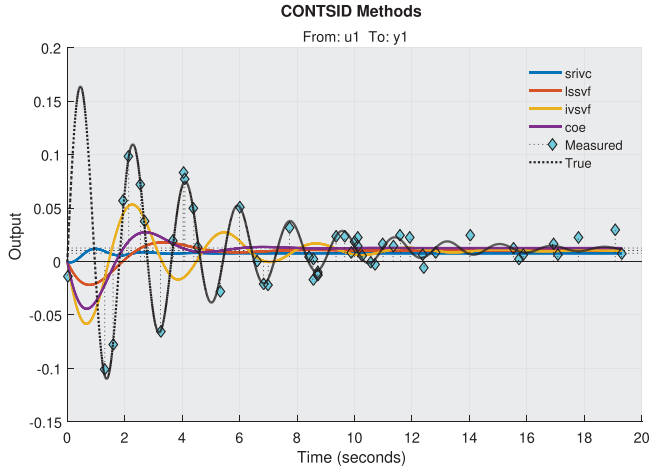

Figure 6. CONSTID identification result.

when we have 50 data, it far from the true impulse response (see Figure 6). Note that after trying a few cut-off frequency, the cut-off frequency for lssvf, ivsvf and coe is chosen as 0.4.

Even though the available methods fail to identify a system that interpolates the given data, the proposed method successfully identified the impulse response of the system from a noisy and very sparsely sampled dataset. In our method, the poles are chosen in the set $\mathbb{S}_2$.

5.3 Average performance

In this section, the performance of the method for non-zero initial condition is analysed based on 200 different experiments for both step response and impulse response. In these experiments, the random stable systems are picked in the set $\mathbb{S}_\rho$ while ρ ranging from 1 to 6. Additionally the order of these random

Table 1. Experiment statistics.

	Estimation error			Sparsity measure		
	Min	Avr	Max	Min	Avr	Max
Step response	0.015	0.503	1.648	0.000	0.389	2.786
Impulse response	0.0089	1.383	32.882	0.000	0.384	3.284

systems is also randomly picked between 2 and 8. In each simulation, the data length was picked between 10 and 60 randomly. Finally the output of this system is corrupted by additive noise $\mathcal{N}(0, 0.005 \times \max(\mathbf{y}))$. Then findings are outlined in Table 1. In this table, estimation error is defined as

$$\|\mathbf{g}_{true} - \tilde{\mathbf{g}}\|_2 / \|\mathbf{g}_{true}\|_2.$$

Normalised singular values vector of Hankel matrix of the estimated impulse response, i.e. $\sum_{i>n} \sigma(i)/\sigma(1)$, is used as sparsity measure, where n is the LTI system order. Note that this measure is identically 0 for true impulse response. We have chosen τ equal to the true atomic norm of the system to be identified, for consistency across all simulations. Note that if the true system is known, then τ can be easily calculated by using partial fraction expansion. Therefore, the true bound on the atomic norm for the proposed method for Algorithm 1 is provided for each simulation.

6. miLife study results

In this section, the performance of the proposed method is evaluated based on a real adolescents depression/anxiety data.

6.1 Study and measurements

Ecological momentary assessment (EMA) data from the miLife Study is analysed in this paper. The miLife Study used EMA via mobile phones to track daily experiences, symptoms and behaviours of young adolescents ($N = 151$) at heightened risk for both exposure to violence and mental health problems. Adolescents were, on average, 13 years of age (with ages ranging from 11 to 15 years, SD = 0.91). The sample was 48% female and ethnically diverse (57.3% Caucasian, 23.3% Hispanic, 4.0% African-American, 4.7% Native American, 4.0% Asian and 6.7% Other). One in three families in the sample 'occasionally' or 'often' had difficulty paying for food or other necessities, 40% reported difficulties paying for bills and 8% reported that they were currently receiving government services or assistance. Parental reports were collected for 93% of the adolescents in the sample ($n = 141$). Adolescents from low socioeconomic status neighbourhoods were recruited via a brief telephone screen (full details regarding recruitment are provided elsewhere: Odgers & Russell, 2017; Russell et al., 2016). Adolescents with three or more risk factors reported by the parent (i.e. behavioural difficulties, inattention or hyperactivity or early initiation of substances, or a parent with a substance use problem) were invited to participate in the study. Parents provided consent and the adolescents provided assent.

Depression and anxiety symptoms were measured three times daily using EMA. Adolescents reported on the presence (yes/no) of five depressive (e.g. I feel hopeless, like nothing matters) and four anxiety symptoms (e.g. I am worried). The sum of these symptoms was taken at each assessment. Violence exposure was measured in the evening each day. Adolescents were asked whether they witnessed people fighting: (a) at home, (b) in school, (c) in their neighbourhoods or (d) somewhere else (e.g. 'Did you see people fighting in your home today?' Yes/No). If exposure was reported in any context, violence exposure was coded 1 for the day, 0 if not. Violence exposure occurred on 9.7% of the over 4300 study days.

To use the proposed approach in a behavioural treatment setting, we first need to determine which variables should be treated as inputs and which ones should be considered as outputs. Since the behavioural problems are more abstract than the mechanical/electrical systems, we use some of the results available in the literature to guide us in this process. The literature on adolescent stress studies shows that adolescents exposed to violence from different sources such as in their families, communities and schools are at increased risk for depression and anxiety (Fowler, Tompsett, Braciszewski, Jacques-Tiura, & Baltes, 2009; Guerra, Rowell Huesmann, & Spindler, 2003; Odgers & Russell, 2017). Therefore, to investigate the association between violence and adolescent depression or anxiety, depression or anxiety is assumed to be an output for our LTI, SISO system while violence is assumed to be input of the system.

6.2 Data processing

Total 151 adolescents were tracked for 38 days using EMA. First 8 days is used for baseline analysis. In our analysis, we only include adolescents who reported witnessing violence at least twice during the 30-day EMA period, because adolescents who were exposed to only one violent incident may not provide enough information for us to determine, with any reliability, how much the system is perturbed by violence. As a result of this simple filtering, we only modelled 45 adolescents out of 151 adolescents in the study. In addition, some of the people eliminated did not have enough information (number of depression/anxiety measurements) to derive a meaningful dynamical model. Among adolescents with sufficient data, we modelled the output variable (i.e.depression or anxiety, $\mathbf{y}$ in Equation (1)) for which we had the largest number of measurements. Therefore, two groups of adolescents are analysed: those for whom we modelled depression (referred to hereafter as the **depression group**, 20 adolescents) and those for whom we modelled anxiety (**anxiety group**, 25 adolescents). Violence measurements were assumed to be a binary input ($\mathbf{u}$ in Equation (1)) for our model such that the measurement is 1 if the violence occurred and 0 otherwise. In system theory language, we are randomly perturbing system with an impulse while the initial conditions are not equal to zero. In other words, output is a response of series of impulses while the initial conditions are not zero at the time of input applied.

6.3 Adolescents classification

The model characteristics obtained for each of the 45 adolescents are summarised in Tables 2 and 3. The information in Tables 2 and 3 can be used to classify the adolescents based on the time course of affective reactivity to violence exposure.

Table 2. Patients model specifications of depression group.

Adolescent	Peak value	Settling time (hours)	# Violence	# Depression (anxiety)
1	0.285	0.66	3	16
2	0.4055	0.77	2	14
3	1.4872	1.14	3	11
4	0.5792	0.93	4	35
5	1.4069	1.21	20	27
6	0.3002	2.24	2	32
7	0.8926	1.07	20	31
8	1.2451	1.07	3	28
9	0.7902	1	3	15
10	0.3476	0.79	2	31
11	1.3287	1.2	8	34
12	1.905	1.23	2	35
13	0.2005	0.51	8	29
14	1.5631	1.94	11	41
15	0.2898	0.7	3	21
16	0.2242	1.54	9	36
17	1.9026	1.18	2	30
18	0.496	0.8	2	52
19	0.4037	0.79	10	35
20	1.7531	1.31	3	31

Table 3. Patients model specifications of anxiety group.

Adolescent	Peak value	Settling time (hours)	# Violence	# Depression (anxiety)
1	0.3789	0.53	3	15
2	0.9896	1	12	32
3	0.7137	1.1	13	44
4	0.7585	1.07	9	31
5	0.8220	1.05	5	31
6	0.1185	0.75	9	37
7	1.4548	1.11	3	37
8	0.1677	0.56	3	33
9	0.6643	1	6	33
10	1.7162	1.44	12	28
11	1.9484	1.24	2	14
12	0.5342	0.86	3	30
13	1.2803	1.61	2	18
14	1.256	11.11	15	35
15	0.1372	0.45	5	31
16	0.1903	0.77	3	33
17	0.6378	0.88	3	35
18	0.4625	0.78	2	22
19	0.4676	0.78	2	35
20	0.3031	1.24	7	31
21	0.7016	0.87	7	27
22	0.1924	0.66	4	28
23	0.0606	0.68	7	30
24	1.4270	1.5	13	29
25	1.2581	0.65	15	33

In Table 2, for instance, we highlight the one (total 8 adolescents) who has a settling time (final value of impulse response +0.05) less than 1 h. Also 13 adolescents in anxiety group show similar characteristics (see highlighted adolescents in Table 3). This shows that the coupling of violence exposure with anxiety for this group of adolescents is fading faster than the others. This information could be used by clinicians to manipulate the intensity of the treatment. Alternatively one can use to classify these adolescents based on the impulse response peak value. For instance, eight adolescents in depression group and seven adolescents in anxiety group show much higher initial depression or anxiety against the violence. This information suggests the possibility that these adolescents might need a more aggressive treatment strategy to minimise the impact of violence exposures, although causal inferences cannot be drawn due to the observational nature of the data collection.

6.4 Personal treatment

In addition to this classification approach, clinicians might find the identified models useful for the development of personalised treatments (e.g. using model predictive control, as in Bekiroglu et al., 2016). However, the data set used only concerned 45 subjects. Therefore, although promising preliminary results were obtained, the proposed work should be tested on larger data sets and see if these preliminary results can be extended to larger populations. We should note that an approach based on individual behavioural models is a paradigm shift from how data analysis and treatment design is done is most of current behavioural studies. The fact that, in many situations, we can now collect large amounts of individual data, means that a personalised approach to analysis and treatment design is feasible. This contrasts with 'classical approaches' where the lack of intensive longitudinal data meant that it was only possible to obtain population models and treatments could only be determined based on that.

7. Conclusion

A new parsimonious continuous time system identification algorithm for non-uniformly sampled datasets is developed for miLife study to analyse the association between violence and an adolescent's depression and anxiety. We show that some of the application areas such as behavioural treatment design problem requires a model despite non-uniformly sampled data and nonlinearity to improve the performance of processes. For that reason, we propose a method that can use non-uniformly sampled data directly to provide a parsimonious estimate of the impulse response of a continuous time system. To develop this method, an atomic norm approach is utilised. We show that the proposed method provides a systematic way of estimating a person based model for medical/behavioural problem. The performance of the proposed method is tested on miLife study and our findings show that reactions to the violence can differ from person to person. These findings reveal new insights for behavioural/social sciences to develop new micro-scale treatment design for each patient. In terms of system theory, future work should investigate an on-line sparse continuous time system identification algorithm to improve the performance of adaptive treatment algorithms. Furthermore, a systemic way needs to be investigated for true τ search.

Notation

Scalars are denoted by lower case letters (e.g. x), vector-valued signals by lowercase boldface letters (e.g. $\mathbf{x} \in \mathcal{L}^p$), where $\mathcal{L}^p$, $0 \leq p \leq \infty$ denotes the usual $\mathcal{L}^p$ space with the norm $\|.\|_p$, transpose of the vector is denoted as $\mathbf{x}^T$, the jth element of vector is denoted as $\mathbf{x}(j)$, the real numbers by $\mathbb{R}$, the complex numbers by $\mathbb{C}$, the integers by $\mathbb{Z}$ and the exponential for each element of vector $\mathbf{x}$ by e^x. For a signal $\mathbf{x}$, $\mathbf{x}(t_i)$ denotes the value of the signal at time t_i. The matrices are presented by bold capital letter (e.g.

X). For a complex number $p \in \mathbb{C}$, $Re(p)$ denotes the real part of p and $Im(p)$ denotes the imaginary part of p. Also p^* represents the complex conjugate of complex number p. e^p represents the exponential of complex number p. For a closed rectangle in left-half plane in $\mathbb{C}$ is denoted by $\mathbb{S}_\rho = \{p \in \mathbb{C} \text{ or } p \in \mathbb{R} : -\rho < Re(p) < 0 \text{ and } -\pi/T_s < Im(p) < \pi/T_s\}$ where T_s is sampling interval. conv(S) represents the convex hull of the set S. Furthermore the convolution of the signals **g** and **u** is denoted by $\mathbf{g} * \mathbf{u}$. The inverse Laplace transform is denoted by $\mathscr{L}^{-1}$.

Notes

1. Note that repeated poles might be approximated with non-repeated poles with an arbitrary perturbation (Van Dooren, Gallivan, & Absil, 2010).
2. If the true system is known, then τ can be easily calculated by using partial fraction expansion. Since the transfer function of an LTI system can be represented as a linear sum of the element of the set in (5) then absolute sum ($\|.\|_{\ell_1}$) of the weight of each element in this partial expansion will be the bound on the inf $\sum_{a\in\mathcal{A}} |c_a|$ in equation (11).
3. Since the set $\mathbb{S}_\rho$ contains only the stable poles but not necessarily asymptotically stable poles, the stable poles converge to zero. Therefore instead of calculating the tail of impulse response with bunch of zeros on the tail of impulse response, one can cut the tail to identify the coefficients of the model.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was partially supported by National Institutes of Health (NIH) grant P50 DA039838 and National Science Foundation (NSF) Grants CNS-1329422, EECS 1808381 and the Building and Construction Authority through the NRF GBIC Program with the project reference NRF2015ENC-GBICRD001-057. The work described is the sole responsibility of the authors and does not necessarily represent the official views of NIH or NSF.

ORCID

Korkut Bekiroglu http://orcid.org/0000-0003-4582-0216

References

Ashour, M., Bekiroglu, K., Yang, C.-h. H., Lagoa, C., Conroy, D., Smyth, J., & Lanza, S (2016). On the mathematical modeling of the effect of treatment on human physical activity. In *2016 IEEE conference on control applications (CCA)*(pp. 1084–1091).

Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2010). Now-casting and the real-time data flow. In *Handbook of economic forecasting* (Vol. 2, pp. 195–237). New York, NY: Elsevier.

Bekiroglu, K., Lagoa, C., Lanza, S. T., & Sznaier, M. (2017). System identification algorithm for non-uniformly sampled data. *IFAC-PapersOnLine*, *50*(1), 7296–7301.

Bekiroglu, K., Lagoa, C., Murphy, S. A., & Lanza, S. T. (2016). Control engineering methods for the design of robust behavioral treatments. *IEEE Transactions on Control Systems Technology*, *PP*(99), 1–12.

Bühlmann, P., & Van De Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Springer Science & Business Media. https://books.google.com/books?hl = en&lr = &id = S6jYXmh988UC&oi = fnd&pg = PR3&ots = VM8c_5vo4s&sig = szdw4EIFYxx-PGhIA_qfW89EF3Y

Chen, F., Agüero, J. C., Gilson, M., Garnier, H., & Liu, T. (2017). EM-based identification of continuous-time ARMA models from irregularly sampled data. *Automatica*, *77*, 293–301.

Deshpande, S., Nandola, N. N., Rivera, D. E., & Younger, J. W. (2014). Optimized treatment of fibromyalgia using system identification and hybrid model predictive control. *Control Engineering Practice*, *33*, 161–173.

Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, *52*, 1289–1306.

Fowler, P. J., Tompsett, C. J., Braciszewski, J. M., Jacques-Tiura, A. J., & Baltes, B. B. (2009). Community violence: A meta-analysis on the effect of exposure and mental health outcomes of children and adolescents. *Development and Psychopathology*, *21*(1), 227–259.

Garnier, H. (2011). *Data-based continuous-time modelling of dynamic systems*. 2011 International symposium on advanced control of industrial processes (ADCONIP) (Ml), Hangzhou (pp. 146–153).

Garnier, H., Wang, L., & Young, P. C. (n.d.). *Direct identification of continuous-time models from sampled data: Issues, basic solutions and relevance*. London: Springer.

Godoy, B. I., Agüero, J. C., Carvajal, R., Goodwin, G. C., & Yuz, J. I. (2014). Identification of sparse FIR systems using a general quantisation scheme. *International Journal of Control*, *87*(4), 874–886.

Guerra, N. G., Rowell Huesmann, L., & Spindler, A. (2003). Community violence exposure, social cognition, and aggression among urban elementary school children. *Child Development*, *74*(5), 1561–1576.

Jaggi, M. (2013). *Revisiting {frank-wolfe}: Projection-free sparse convex optimization*. Proceedings of the 30th international conference on machine learning, Atlanta, GA, USA (pp. 427–435).

Johansson, R. (2009). *Continuous-time model identification and state estimation using non-uniformly sampled data*. 15th IFAC symposium on system identification, Budapest, Hungary.

Lagoa, C. M., Bekiroglu, K., Lanza, S. T., & Murphy, S. A.. (2014). Designing adaptive intensive interventions using methods from engineering. *Journal of Consulting and Clinical Psychology*, *82*(5), 868–878.

Mercère, G., Ouvrard, R., Gilson, M., & Garnier, H. (2007). *Subspace based methods for continuous-time model identification of MIMO systems from filtered sampled data*. 2007 European control conference (ECC), Kos, Greece (pp. 4057–4064).

Odgers, C. L., & Russell, M. A. (2017). Violence exposure is associated with adolescents' same- and next-day mental health symptoms. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*(12), 1310–1318.

Ohta, Y., & Kawai, T (2004). Continuous-time subspace system identification using generalized orthonormal basis functions. In *16th International symposium on mathematical theory of networks and systems*, Leuven, Belgium.

Rao, G. P., & Unbehauen, H. (2006). Identification of continuous-time systems. *IEE Proceedings-Control Theory and Applications*, *153*(2), 185–220.

Rivera, D. E. (2012). Optimized behavioral interventions: What does system identification and control engineering have to offer? *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *45*(16), 882–893.

Russell, M. A., Wang, L., & Odgers, C. L. (2016). Witnessing substance use increases same-day antisocial behavior among at-risk adolescents: Gene-environment interaction in a 30-day ecological momentary assessment study. *Development and Psychopathology*, *28*(4), 1441–1456.

Shah, P., Bhaskar, B. N., Tang, G., & Recht, B (2012). Linear system identification via atomic norm regularization. In *2012 IEEE 51st IEEE conference on decision and control (CDC)* (pp. 6265–6270).

Van Dooren, P., Gallivan, K., & Absil, P. A. (2010). H2-optimal model reduction with higher-order poles. *SIAM Journal on Matrix Analysis and Applications*, *31*, 2738–2753.

Yilmaz, B., Bekiroglu, K., Lagoa, C., & Sznaier, M. (2017). A randomized algorithm for parsimonious model identification. *IEEE Transactions on Automatic Control*, *PP*(99), 1.

Yoshimura, S., Matsubayashi, A., & Inoue, M. (2019). System identification method inheriting steady-state characteristics of existing model. *International Journal of Control*, *92*(11), 2701–2711.

Young, P. (1981). Parameter estimation for continuous-time models – a survey. *Automatica*, *17*(1), 23–39.

Yue, Z., Thunberg, J., Ljung, L., & Goncalves, J (2016). Identification of sparse continuous-time linear systems with low sampling rate: Exploring matrix logarithms. arXiv preprint arXiv:1605.08590.

Yuz, J., Alfaro, J., Agüero, J., & Goodwin, G. (2011). Identification of continuous-time state-space models from non-uniform fast-sampled data. *IET Control Theory & Applications*, *5*(7), 842–855.