Identification of piecewise constant parameters in nonlinear models

Identification of Piecewise Constant Parameters in Nonlinear Models

Giuseppe Carlo Calafiore[a], Giulia Fracastoro[a], Lorenzo Zino[a,1]

[a]*Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, 10129, Italy*

Abstract

Many real-world dynamical systems are characterized by different temporal phases, with sudden changes in the values of the system's parameters in correspondence to variations from one phase to another. Identifying the system's parameters and these switching instants from potentially noisy measurements of the system's states is a relevant problem in several applications. We here propose a novel approach for estimating the time-varying parameters of a broad class of nonlinear dynamical systems from noisy state measurements. We formulate the problem as a mixed-integer quadratic program (MIQP) including a sparsity constraint to enforce the piecewise constant nature of the parameters. Then, we develop a convex relaxation of the problem in the form of a quadratic program (QP). The solution of the relaxed convex QP and/or the sub-optimal solutions of the MIQP returned by the MIQP solvers provide us with computationally-efficient approximations that can be used effectively in those large-dimensional cases in which the solution of the original MIQP is difficult to obtain. After validating our approach in a controlled experiment, we demonstrate its potential on two real-world case studies regarding marketing and epidemiological applications.

Keywords: System identification, Nonlinear dynamics, Mixed-integer optimization, Convex programming.

*Corresponding author L. Zino (email: `lorenzo.zino@polito.it`)

Email addresses: `giuseppe.calafiore@polito.it` (Giuseppe Carlo Calafiore), `giulia.fracastoro@polito.it` (Giulia Fracastoro), `lorenzo.zino@polito.it` (Lorenzo Zino)

1. Introduction

Identifying the parameters of a dynamical system is a problem of paramount importance in many real-world applicative fields, spanning from engineering to physics, from economics and social science to epidemiology, [1]. The field of system identification is mature and well-developed, especially in the context of linear systems, see, e.g., classical references such as [2, 3, 4, 5], which developed upon earlier approaches on least-squares methods for identifying the parameters of linear time-invariant systems [6], and on design of techniques based on maximum-likelihood estimation for stochastic linear systems [7]. In more recent times, optimization methods —in particular convex optimization— emerged as powerful tools for effective parameter identification, see, e.g., [8].

These classical approaches focus primarily on the scenario in which the dynamics involved are linear and the parameters do not change with time. However, many important applications do not fit in such context, and hence classical identification methods have limited applicability in these cases. In fact, the majority of real-world phenomena are characterized by inherent nonlinearities and non-stationary parameters' behavior. In this paper we focus in particular on the case where the parameters change abruptly, rather than smoothly with time. This happens often in systems that exhibit different temporal phases, with sudden changes in the values of the parameters between phases, due to sudden changes in the environment. An example that, sadly, became very popular in the past few years comes from the mathematical modeling of epidemics, whereby the contagion mechanism is typically modeled by nonlinear dynamics [9], and the implementation of nonpharmaceutical interventions, vaccination campaigns, as well as the appearance of new variants of the disease may yield abrupt changes in the value of the model parameters, see, e.g., [10, 11]. This example illustrates the importance of developing algorithms that are able to identify piece-wise constant parameters for nonlinear dynamical systems from

available noisy measurements, and thus to detect the key phases of the phenomenon under investigation.

Several methods have been developed for nonlinear system identification, including the use of maximum-likelihood methods [12], particle swarm optimization [13], machine learning [14], and neural networks [15, 16]; for more details, see, e.g., the surveys in [17, 18]. Also the problem of identifying time-varying parameters for linear dynamical systems has been addressed in several studies, which include approaches that are applicable to the piece-wise constant case. Efforts have been made to extend existing methods for constant parameters, including the use of least-squares and wavelet functions [19, 20, 21, 22], maximum-likelihood [23], generalization of the Extended Kalman filters [24], and semidefinite programming methods [25]. However, these methods do not enforce bounds on the number of changing points (switching instants) of the parameters, and typically yield a smooth continuously-changing estimation of the parameters. Other methods that enforce piecewise-constant variations often rely on the critical assumption that the switching instants of the parameters are known a priori [23]; this assumption can be rather unrealistic and it is a key hypothesis that we lift in our approach. Approaches for linear regression with time-varying parameters include [26, 27]. Of particular interest in this context is the use of hybrid systems theory to address the problem of estimating piecewise-constant parameters of linear (or affine) dynamical systems, by framing it as a parameter identification problem for piecewise affine switched systems, assuming an underlying autoregressive model [28, 29].

Quite surprisingly, little material is available for the case of nonlinear dynamical systems with piecewise constant parameters and noisy measurements, which is relevant for many applications. Assuming underlying nonlinear autoregressive exogenous models, parameter identification algorithms for piecewise nonlinear switched systems have been developed by means of different approaches, including maximum likelihood frameworks, iterative filter-based techniques, and and residual-based methods [30, 31, 32, 33]. Most of the existing methods used to identify parameters in the presence of such abrupt changes assume an underly-

ing Markov model that captures the changes in the parameters, and rely on the use of iterative filtering or bayesian inference [34, 35]. However, these methods based on hybrid systems are designed for systems that switch repeatedly between different modes according to purely stochastic rules, each one characterized by a set of parameters. Hence, these methods are unfit for the scenario of interest in this paper, in which parameters switch asynchronously and modes are in general not repeated and regulated by rules that are not purely stochastic, and for which one cannot assume absence of memory as in classical Markov chains. Moreover, these methods do not typically allow one to set a priori the total number of switches of the parameters. This limits the possibility of using them to identify and characterize a predetermined number of key switches, which could be useful in many real-world applications, including social and economic systems. Finally, ad-hoc methods have been developed to deal with specific real-world problems. For instance, in [36] the authors deal with the problem of identifying epidemic parameters from publicly available data, and they propose the use of an ℓ_1-type penalty function to enforce sparsity in the discrete derivatives of the time-varying parameters. However, a general and flexible formalization of an optimization framework to address this important problem is still missing.

In this paper, we aim at filling this gap by presenting a novel methodological framework for the identification of piecewise constant parameters from noisy observations for a broad class of nonlinear dynamical systems. Technically, our goal is to identify a sequence of parameters that: i) minimizes the error between the predicted states and the observed states, ii) controls the impact of noise in the measurements, while iii) keeping small the number of times the parameters vary during the entire time-window, thus enforcing a piecewise-constant nature for such parameters. To this aim, we formulate the identification problem as an optimization problem with a quadratic objective function that encapsulates the prediction error and the impact of the noise, while the third goal is introduced as a sparsity constraint on the parameters' discrete-time derivatives. Using suitable algebraic developments, we demonstrate that the latter can be conveniently re-cast as a linear constraint for a set of binary variables.

Hence, the proposed parameter identification algorithm ultimately reduces to a mixed-integer quadratic program (MIQP), whose solution can be computed or approximated via state-of-the-art optimization solvers in small-to-medium size scenarios, see, e.g., [37]. The computational complexity of the proposed method, however, remains NP-complete [38], which may hinder its applicability in large-scale systems and/or when the parameters change several times in the observation time-window. To address these limitations, we further develop a convex relaxation of the proposed MIQP into a computationally efficient quadratic program (QP), which is then demonstrated to be a practically useful proxy of the original MIQP.

After a validation on a synthetic dataset, we test the proposed framework on two real-world case studies. First, we consider a marketing scenario in which the temporal evolution of the market shares of three competing brands is modeled using a generalized Lotka–Volterra model, based on [39], and we calibrate it on the Japanese beer market, using data from [40]. This case study demonstrates the potentiality of our optimization-based approach, which is able to correctly identify the different phases of the market and the key switching instants discussed in the literature [41]. Moreover, the moderate size of the dataset allows us to perform a comparison between the MIQP and the relaxed QP. Such a comparison shows that sub-optimal solutions of the MIQP are good proxies of the exact solutions, but their computation becomes computationally challenging when the parameters have several changes over the time-window. In this scenarios, the QP is fast to be solved and its solution is a quite reliable proxy of the solution of the MIQP and can thus be used when solving the MIQP —or even computing good approximations of the solution— becomes practically infeasible. Second, we discuss a case study based on the spread of COVID-19 in Italy, using a susceptible–infected–removed–deceased (SIRD) model, similar to [10]. Here, we utilize a longer dataset spanning more than 2 years, which makes the solution of the MIQP computationally unfeasible (and also the computation of reliable sub-optimal solutions is challenging), highlighting the importance and effectiveness of the computationally-efficient QP relaxation. Our results allow

us to gain interesting insight into the different phases that characterized the temporal evolution of the COVID-19 pandemic in Italy.

The rest of the paper is organized as follows. In Section 2, we present the setting and the considered model class. In Section 3, we introduce our optimization-based approach for identification, and in Section 4 we discuss in depth two application case studies. Section 5 concludes the paper.

2. Problem Setup

2.1. Notation

We denote the set of real, nonnegative real, nonnegative integer, and strictly positive integer numbers as $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{N}$, and $\mathbb{N}_+$, respectively. Given a positive integer $n \in \mathbb{N}_+$, a (column) vector $\boldsymbol{x} \in \mathbb{R}^n$ is denoted with bold lowercase font, with ith entry x_i. Given two positive integers $n, m \in \mathbb{N}_+$, a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ is denoted with bold capital font, with A_{ij} denoting the generic jth entry of the ith row and $\boldsymbol{a_j}$ denoting its jth column vector. Given a vector $\boldsymbol{x}$ or a matrix $\boldsymbol{A}$, $\boldsymbol{x}^\top$ and $\boldsymbol{A}^\top$ are the transpose vector and matrix, respectively. Given a vector $\boldsymbol{x}$, we denote by $\mathrm{diag}(\boldsymbol{x})$ the square matrix with the elements of $\boldsymbol{x}$ on the diagonal and 0 on off-diagonal entries. The vector $\boldsymbol{e}_i$ denotes a vector of all zeros, except for a 1 in the ith entry. Given a matrix $\boldsymbol{A}$, we denote by $\mathrm{tr}(\boldsymbol{A})$ its trace. The matrix of all ones is denoted by $\mathbf{1}$, the matrix of all zeros is $\mathbf{0}$, and the identity matrix is $\boldsymbol{I}$. When needed, dimensions are denoted as one positive integer index (for vectors) and two positive integer indices (for matrices), which are the number of rows and columns, respectively. The set of n-dimensional symmetric positive semi-definite matrices is denoted by $\mathcal{S}^n$. The sign operator sgn() and inequalities applied to a vector or a matrix are intended entrywise (i.e., $\boldsymbol{x} \geq 0 \iff x_i \geq 0$, for all i).

2.2. Model class

We consider a class of discrete-time nonlinear dynamical systems of the form

$$\boldsymbol{x}(t+1) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t); \boldsymbol{\theta}(t)), \tag{1}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the n-dimensional state variable of the system at time $t \in \mathbb{N}$, $\boldsymbol{u}(t) \in \mathbb{R}^p$ is the p-dimensional input vector, $\boldsymbol{\theta}(t) \in \mathbb{R}^m$ is a vector containing the m possibly time-varying parameters of the model, and $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^n$ is the state-transition map, on which we make the following assumption.

Assumption 1. *The map $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}; \boldsymbol{\theta}) : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^n$ is differentiable with respect to $\boldsymbol{x}$ and affine in the parameter vector $\boldsymbol{\theta}$, i.e., we can write*

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}; \boldsymbol{\theta}) = \boldsymbol{f_0}(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u})\boldsymbol{\theta}, \tag{2}$$

with vector $\boldsymbol{f_0}(\boldsymbol{x}, \boldsymbol{u}) \in \mathbb{R}^n$ and matrix $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{u}) \in \mathbb{R}^{n \times m}$, for all $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{u} \in \mathbb{R}^p$.

At each time $t \in \mathbb{N}$, there is available an observation of the state of the system $\boldsymbol{x}(t)$, denoted as $\bar{\boldsymbol{x}}(t) \in \mathbb{R}^n$, which may be subject to noise and errors in the measurement. Specifically, we assume that the measurement is in the form

$$\bar{\boldsymbol{x}}(t) = \boldsymbol{x}(t) + \boldsymbol{\xi}(t), \tag{3}$$

where $\boldsymbol{\xi}(t) \in \mathbb{R}^n$ is a vector that accounts for errors in the measurement of $\boldsymbol{x}(t)$. We observe that we can write the evolution of the measured state $\bar{\boldsymbol{x}}(t)$ via the following recursion:

$$\begin{aligned} \bar{\boldsymbol{x}}(t+1) &= \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t); \boldsymbol{\theta}(t)) + \boldsymbol{\xi}(t+1) \\ &= \boldsymbol{f}(\bar{\boldsymbol{x}}(t) - \boldsymbol{\xi}(t), \boldsymbol{u}(t); \boldsymbol{\theta}(t)) + \boldsymbol{\xi}(t+1) \\ &= \boldsymbol{f}(\bar{\boldsymbol{x}}(t), \boldsymbol{u}(t); \boldsymbol{\theta}(t)) - \boldsymbol{J_f}(\bar{\boldsymbol{x}}(t))\,\boldsymbol{\xi}(t) + \boldsymbol{\xi}(t+1) + o(\|\boldsymbol{\xi}(t)\|_2), \end{aligned} \tag{4}$$

where $\boldsymbol{J_f}(\bar{\boldsymbol{x}}(\boldsymbol{t}))$ denotes the Jacobian matrix of $\boldsymbol{f}$ with respect to $\boldsymbol{x}$, evaluated in $\boldsymbol{x} = \bar{\boldsymbol{x}}(t)$. Hence, assuming that the error size $\|\boldsymbol{\xi}(t)\|_2$ is sufficiently small, we can approximate $\bar{\boldsymbol{x}}(t+1)$ by neglecting the higher-order terms, obtaining the approximate recursion:

$$\bar{\boldsymbol{x}}(t+1) \simeq \boldsymbol{f}(\bar{\boldsymbol{x}}(t), \boldsymbol{u}(t); \boldsymbol{\theta}(t)) - \boldsymbol{J_f}(\bar{\boldsymbol{x}}(t))\,\boldsymbol{\xi}(t) + \boldsymbol{\xi}(t+1). \tag{5}$$

As a consequence of Assumption 1, the Jacobian matrix $\boldsymbol{J_f}(\bar{\boldsymbol{x}}(t))$ takes the form

$$\boldsymbol{J_f}(\bar{\boldsymbol{x}}(t)) = \boldsymbol{J_{f_0}}(\bar{\boldsymbol{x}}(t)) + \sum_{i=1}^{m} \theta_i \boldsymbol{J_{f_i}}(\bar{\boldsymbol{x}}(t)), \tag{6}$$

where $\boldsymbol{J_{f_0}}$ is the Jacobian matrix of $\boldsymbol{f_0}$ and $\boldsymbol{J_{f_i}}$ is the Jacobian matrix of the ith column of matrix $\boldsymbol{F}(\boldsymbol{x},\boldsymbol{u})$, denoted as $\boldsymbol{f_i}(\boldsymbol{x},\boldsymbol{u})$. Therefore, the approximate recursion Eq. (5), neglecting the higher-order terms, becomes

$$\begin{aligned}\bar{\boldsymbol{x}}(t+1) \;\simeq\; & \boldsymbol{f_0}(\bar{\boldsymbol{x}}(t),\boldsymbol{u}(t)) + \boldsymbol{F}(\bar{\boldsymbol{x}}(\boldsymbol{t}),\boldsymbol{u}(\boldsymbol{t}))\boldsymbol{\theta}(t) - \boldsymbol{J_{f_0}}(\bar{\boldsymbol{x}}(t))\boldsymbol{\xi}(\boldsymbol{t}) \\ & - \sum_{i=1}^{m} \theta_i(t)\boldsymbol{J_{f_i}}(\bar{\boldsymbol{x}}(t))\boldsymbol{\xi}(t) + \boldsymbol{\xi}(t+1).\end{aligned} \tag{7}$$

Letting

$$\boldsymbol{G}(\boldsymbol{\theta}(t)) = \boldsymbol{G}(\boldsymbol{\theta}(t);\bar{\boldsymbol{x}}(t),\boldsymbol{u}(t)) := \boldsymbol{J_{f_0}}(\bar{\boldsymbol{x}}(t)) + \sum_{i=1}^{m} \theta_i(t)\boldsymbol{J_{f_i}}(\bar{\boldsymbol{x}}(t)), \tag{8}$$

we rewrite Eq. (7) more compactly as

$$\bar{\boldsymbol{x}}(t+1) \simeq \boldsymbol{f_0}(t) + \boldsymbol{F}(t)\boldsymbol{\theta}(t) - \boldsymbol{G}(\boldsymbol{\theta}(t))\boldsymbol{\xi}(t) + \boldsymbol{\xi}(t+1), \tag{9}$$

where $\boldsymbol{f_0}(t) := \boldsymbol{f_0}(\bar{\boldsymbol{x}}(t),\boldsymbol{u}(t))$ and $\boldsymbol{F}(t) := \boldsymbol{F}(\bar{\boldsymbol{x}}(t),\boldsymbol{u}(t))$.

2.3. Objectives

Our objective is to estimate the time-varying parameter vector $\boldsymbol{\theta}(t)$ from a given time sequence of noisy observations of the state variables $\bar{\boldsymbol{x}}(0),\ldots,\bar{\boldsymbol{x}}(T-1)$, over a time-window of duration $T \in \mathbb{N}_+$. In particular, in our parameter identification, we aim at achieving the following three objectives:

1. Minimize the error between the predicted state of the system and the observations through the entire time-window;
2. Keep the impact of the measurement noise under control; and
3. Limit the number of times the parameters vary during the entire time-window.

3. A mixed-integer optimization approach

We next illustrate how to synthesize the identification objectives stated in Section 2.3 via a suitable mixed-integer optimization problem. Preliminarily, we make the following further assumptions on the noise sequence and on the parameter vector.

Assumption 2. *The noise sequence* $\boldsymbol{\xi}(t) \in \mathbb{R}^n$, $t = 0, 1, \ldots$, *is a sequence of independent random variables with mean equal to* $\mathbf{0}$ *and given sequence of covariance matrices* $\boldsymbol{Q}(t) \in \mathcal{S}^n$, $t = 0, 1, \ldots$, *known a priori.*

Assumption 3. *Bounds on the amplitude of* $\boldsymbol{\theta}(t)$ *are known a priori. That is, there exists a given constant* $C > 0$ *such that* $\|\boldsymbol{\theta}(t)\|_\infty \leq C$, *for all* $t \in \mathbb{N}$.

We consider a given time horizon $T \in \mathbb{N}_+$, for which we have collected measured data $\bar{\boldsymbol{x}}(0), \ldots, \bar{\boldsymbol{x}}(T-1)$. At each t, we recognize in Eq. (9) two independent error contributions to $\bar{\boldsymbol{x}}(t+1)$: the direct measurement error $\boldsymbol{\xi}(t+1)$, and the indirect term $\boldsymbol{G}(\boldsymbol{\theta}(t))\boldsymbol{\xi}(t)$ which quantifies the dynamic effect of the error in $\bar{\boldsymbol{x}}(t)$ on $\bar{\boldsymbol{x}}(t+1)$, up to the first order approximation. From Assumption 2, the covariance matrix of this error term is $\boldsymbol{G}(\boldsymbol{\theta}(t))\boldsymbol{Q}(t)\boldsymbol{G}(\boldsymbol{\theta}(t))^\top$.

As discussed in Section 2.3, the goal of this paper is to determine an estimate of the parameter vector $\boldsymbol{\theta}(t)$ from the noisy sequence of data $\bar{\boldsymbol{x}}(1), \ldots, \bar{\boldsymbol{x}}(T)$ with the objective of minimizing the sum of two terms. The first objective is the total (weighted) squared prediction error

$$\mathcal{E} := \sum_{t=0}^{T-1} \|\boldsymbol{W}(t)(\bar{\boldsymbol{x}}(t+1) - \boldsymbol{f_0}(t) - \boldsymbol{F}(t)\boldsymbol{\theta}(t))\|_2^2, \tag{10}$$

where $\boldsymbol{W}(t) = \mathrm{diag}(\boldsymbol{w}(t))$, with $\boldsymbol{w}(t) > 0$, $t = 0, \ldots, T-1$ is a (possibly time-varying) diagonal weight matrix. The weights in $\boldsymbol{w}(t)$ re-scale the contribution of each component of the state vector $\boldsymbol{x}$ to the total prediction error, giving flexibility to the identification approach for dealing with several issues, including i) different orders of magnitude between the variables, ii) different reliabilities of different state measurements, and iii) focus of the identification process on fitting the parameters better on a limited time-interval.

The second objective is to control the total indirect measurement errors, by keeping small the sum of the traces of their covariance matrices:

$$\mathcal{V} := \sum_{t=0}^{T-1} \mathrm{tr}(\boldsymbol{G}(\boldsymbol{\theta}(t))\boldsymbol{Q}(t)\boldsymbol{G}^\top(\boldsymbol{\theta}(t))) = \|G(\boldsymbol{\theta}(t))\boldsymbol{Q}^{1/2}(t)\|_F^2. \tag{11}$$

Observe that the weighted prediction error in Eq. (10) is convex in $\boldsymbol{\theta}(t)$, and also the error cost in Eq. (11) is convex, since $\boldsymbol{G}$ is affine in $\boldsymbol{\theta}(t)$. The parameter

identification problem is next formalized as a minimization problem with respect to the parameter vector $\boldsymbol{\theta}(t)$, with a cost function equal to a positive linear combination of $\mathcal{E}$ and $\mathcal{V}$, which is convex, and a constraint which encodes the third objective in our identification approach, that is the requirement that the time evolution of the parameter is piece-wise constant.

To this end, we let $\Delta\boldsymbol{\theta}(t)$ denote the first-order difference operator, i.e., $\Delta\boldsymbol{\theta}(t) = \boldsymbol{\theta}(t)-\boldsymbol{\theta}(t-1)$, for $t = 1, \ldots, T-1$. Stacking all the T parameter vectors $\boldsymbol{\theta}(0), \ldots, \boldsymbol{\theta}(T-1)$ in the Tm-dimensional vector $\boldsymbol{\Theta} := [\boldsymbol{\theta}(0)^\top, \ldots, \boldsymbol{\theta}(T-1)^\top]^\top$ and all the $T-1$ first-order differences $\Delta\boldsymbol{\theta}(1), \ldots, \Delta\boldsymbol{\theta}(T-1)$ in the $(T-1)m$-dimensional vector

$$\Delta\boldsymbol{\Theta} := \begin{bmatrix} \Delta\boldsymbol{\theta}(1) \\ \Delta\boldsymbol{\theta}(2) \\ \vdots \\ \Delta\boldsymbol{\theta}(T-1) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}(1) - \boldsymbol{\theta}(0) \\ \boldsymbol{\theta}(2) - \boldsymbol{\theta}(1) \\ \vdots \\ \boldsymbol{\theta}(T-1) - \boldsymbol{\theta}(T-2) \end{bmatrix}$$

we have that $\Delta\boldsymbol{\Theta} = \boldsymbol{D}\boldsymbol{\Theta}$, where the matrix

$$\boldsymbol{D} := \begin{bmatrix} -\boldsymbol{I}_{m,m} & \boldsymbol{I}_{m,m} & \boldsymbol{0}_{m,m} & \boldsymbol{0}_{m,m} & \cdots & \boldsymbol{0}_{m,m} \\ \boldsymbol{0}_{m,m} & -\boldsymbol{I}_{m,m} & \boldsymbol{I}_{m,m} & \boldsymbol{0}_{m,m} & \cdots & \boldsymbol{0}_{m,m} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0}_{m,m} & \boldsymbol{0}_{m,m} & \cdots & \boldsymbol{0}_{m,m} & -\boldsymbol{I}_{m,m} & \boldsymbol{I}_{m,m} \end{bmatrix}$$

is a block-circulant matrix of dimension $(T-1)m \times Tm$. It is clear that the number of jumps, or switching instants, in the time evolution of the parameters in $\boldsymbol{\theta(t)}$ is given by the number of nonzero entries in vector $\Delta\boldsymbol{\Theta}$. The number of nonzero entries of a vector $\boldsymbol{x}$ is called the *cardinality*, denoted as $\|\boldsymbol{x}\|_0$, and can be given a "variational" characterization, as specified in the following.

Lemma 1. *For any given vector $\boldsymbol{x} \in \mathbb{R}^n$, the cardinality (i.e., the number of nonzero entries) of $\boldsymbol{x}$, denoted by the 0-norm of $\boldsymbol{x}$, $\|\boldsymbol{x}\|_0$, can be characterized through the following identities:*

1. $\|\boldsymbol{x}\|_0 = \min_{\boldsymbol{z}\in[0,1]^n} \sum_{i=1}^n z_i$, subject to $\boldsymbol{x} = \mathrm{diag}(\boldsymbol{z})\boldsymbol{x}$.
2. $\|\boldsymbol{x}\|_0 = \min_{\boldsymbol{z}\in[0,1]^n} \sum_{i=1}^n z_i$, subject to $\boldsymbol{x}^\top\boldsymbol{x} \leq \boldsymbol{x}^\top\mathrm{diag}(\boldsymbol{z})\boldsymbol{x}$.

3. $\|\boldsymbol{x}\|_0 = \min_{\boldsymbol{z}\in[-1,1]^n} \|\boldsymbol{z}\|_1$, subject to $\|\boldsymbol{x}\|_1 \leq \boldsymbol{z}^\top \boldsymbol{x}$.

4. *If* $\|\boldsymbol{x}\|_\infty \leq C$, *it further holds that*

$$\|\boldsymbol{x}\|_0 = \min_{\boldsymbol{z}\in\{0,1\}^n} \sum_{i=1}^{n} z_i, \text{ subject to } -C\boldsymbol{z} \leq \boldsymbol{x} \leq C\boldsymbol{z}. \tag{12}$$

Proof. First, we observe that, for any $\boldsymbol{z} \in [0,1]^n$ such that $\boldsymbol{x} = \text{diag}(\boldsymbol{z})\boldsymbol{x}$ it necessarily holds true that $x_i \neq 0 \implies z_i = 1$. Hence, we can immediately conclude that $\|\boldsymbol{x}\|_0 \leq \sum_{i=1}^n z_i$. Then, we observe that the equality is attained for $\boldsymbol{z}^*$ defined entrywise as $z_i^* = 1$ if $x_i \neq 0$ and $z_i^* = 0$ otherwise. This yields the first identity. Identities 2. and 3. are obtained following a similar argument, with the unique difference that, in case 3., the equality is attained for $\boldsymbol{z}^* = \text{sgn}(\boldsymbol{x})$. For 4., observe that for any $\boldsymbol{z} \in \{0,1\}^n$ such that $-C\boldsymbol{z} \leq \boldsymbol{x} \leq C\boldsymbol{z}$ it necessarily holds true that $x_i \neq 0 \implies z_i = 1$. Hence, we can immediately conclude that $\|\boldsymbol{x}\|_0 \leq \sum_{i=1}^n z_i$. Then, we observe again that the equality is attained for $\boldsymbol{z}^*$ defined entry-wise as $z_i^* = 1$ if $x_i \neq 0$ and $z_i^* = 0$ otherwise, where the constraint $-Cz_i \leq x_i \leq Cz_i$ holds true, since $-C \leq x_i \leq C$, yielding the claim. □

From item 4. of Lemma 1, we also obtain the following equivalent conditions

$$\|\boldsymbol{x}\|_0 \leq k \quad \iff \quad \exists\, \boldsymbol{z} \in \{0,1\}^n : \sum_{i=1}^{n} z_i \leq k, \;\; -C\boldsymbol{z} \leq \boldsymbol{x} \leq C\boldsymbol{z}. \tag{13}$$

Let now $\boldsymbol{e_i} \in \mathbb{R}^m$ denote the ith standard basis vector of $\mathbb{R}^m$, and let us define the following block-circulant matrix:

$$\boldsymbol{E_i} := \begin{bmatrix} \boldsymbol{e_i}^\top & \boldsymbol{0}_m^\top & \boldsymbol{0}_m^\top & \dots & \boldsymbol{0}_m^\top \\ \boldsymbol{0}_m^\top & \boldsymbol{e_i}^\top & \boldsymbol{0}_m^\top & \dots & \boldsymbol{0}_m^\top \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0}_m^\top & \boldsymbol{0}_m^\top & \boldsymbol{0}_m^\top & \dots & \boldsymbol{e_i}^\top \end{bmatrix}$$

which is a block-diagonal matrix with $T-1$ rows and $(T-1)m$ columns. Then, we say that the ith parameter sequence $[\theta_i(0), \dots, \theta_i(T-1)]$ is k-piecewise constant if the vector $[\Delta\theta_i(1), \dots \Delta\theta_i(T-1)]$ is k-sparse, i.e., if its cardinality is k. Using the matrix $\boldsymbol{E_i}$, we have that the ith parameter sequence $[\theta_i(0), \dots, \theta_i(T-1)]$ is k-piecewise constant if and only if $\|\boldsymbol{E_i}\boldsymbol{D}\boldsymbol{\Theta}\|_0 \leq k$.

Under Assumptions 1–3, we can finally formulate our parameter identification problem as an optimization problem where we aim to minimize a linear combination of the two terms $\mathcal{E}$ and $\mathcal{V}$, weighted by a nonnegative coefficient $\gamma \geq 0$, while requiring the parameter sequences to be piecewise constant.

Problem 1. *We formulate the following mixed-integer quadratic program (MIQP).*

$$\begin{aligned} \min_{\boldsymbol{\Theta}\in\mathbb{R}^{Tm},\boldsymbol{Z}\in\{0,1\}^{(T-1)\times m}} \quad & \sum_{t=0}^{T-1} \|\boldsymbol{W}(t)(\bar{\boldsymbol{x}}(t+1) - \boldsymbol{f_0}(t) - \boldsymbol{F}(t)\boldsymbol{\theta}(t))\|_2^2 \\ & +\gamma \sum_{t=0}^{T-1} \|\boldsymbol{G}(\boldsymbol{\theta}(t))\boldsymbol{Q}^{1/2}(t)\|_F^2 \\ \textit{s.t.:} \quad & \mathbf{1}^\top \boldsymbol{z}_i \leq k_i, \quad i = 1,\dots,m, \\ & -C\boldsymbol{z}_i \leq \boldsymbol{E}_i\boldsymbol{D}\boldsymbol{\Theta} \leq C\boldsymbol{z}_i, \quad i = 1,\dots,m, \\ & \|\boldsymbol{\Theta}\|_\infty \leq C. \end{aligned} \tag{14}$$

where k_i, $i = 1,\dots,m$, are given upper bounds on the sparsity of $[\Delta\theta_i(1),\dots,\Delta\theta_i(T-1)]$, and $\gamma \geq 0$ is a tradeoff parameter controlling the penalty on the induced error term.

The cost function to be minimized in Problem 1 comes from the explicit expressions for $\mathcal{E}$ and $\mathcal{V}$ derived in Eq. (10) and Eq. (11). The constraint $\|\boldsymbol{\Theta}\|_\infty \leq C$ comes from Assumption 3, while the other constraints in Problem 1 are derived from the k-sparsity constraints $\|\boldsymbol{E_i}\boldsymbol{D}\boldsymbol{\Theta}\|_0 \leq k_i$, for all $i = 1,\dots,m$ by using Eq. (13). The $m(T-1)$ additional binary variables are collected in the matrix $\boldsymbol{Z} \in \{0,1\}^{(T-1)\times m}$, with columns $\boldsymbol{z}_i$, $i = 1,\dots,m$, thus obtaining the MIQP in Eq. (14).

3.1. Validation

Before further discussing our parameter identification method, presenting a convex relaxation, and illustrating their practical use through some case study, we report a validation on a simple example. Here, we assume that the ground truth is known a priori, and we show that the proposed method is able to correctly identify piecewise constant parameters of a nonlinear dynamical system from noisy measurements.

To this aim, we generate a time-series from a bi-dimensional generalized Lotka–Volterra model with no spontaneous growth [42], which is characterized by the dynamical system

$$\begin{aligned} x_1(t+1) &= x_1(t) + \theta_1(t) x_1(t) x_2(t) \\ x_2(t+1) &= x_2(t) + \theta_2(t) x_1(t) x_2(t), \end{aligned} \tag{15}$$

with initial condition $\boldsymbol{x}(0) = [0.2, 0.2]$ and time-varying parameters equal to

$$\theta_1(t) = \begin{cases} 0.2, & 0 \leq t \leq 11, \\ -0.1, & 12 \leq t \leq 19, \\ 0.05, & 20 \leq t \leq 29, \end{cases} \qquad \theta_2(t) = \begin{cases} -0.12, & 0 \leq t \leq 6, \\ 0.07, & 7 \leq t \leq 21, \\ 0.15, & 22 \leq t \leq 29. \end{cases} \tag{16}$$

We can check that the map in Eq. (15) satisfies Assumption 1. In fact, we can write it as $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{f_0}(\boldsymbol{x}) + \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\theta}$, with

$$\boldsymbol{f_0}(\boldsymbol{x}) = \boldsymbol{x}, \qquad \boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} x_1 x_2 & 0 \\ 0 & x_1 x_2 \end{bmatrix}, \tag{17}$$

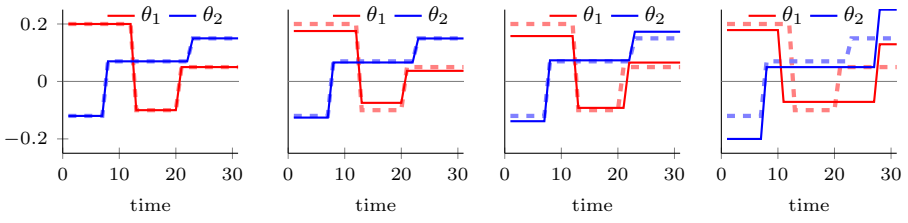
where we have omitted the time-dependencies of the variables and of the parameters for brevity. We can write the Jacobian of the map in Eq. (15) as

$$\boldsymbol{J}_f(\boldsymbol{x}, \boldsymbol{\theta}) = \begin{bmatrix} 1 + \theta_1 x_2 & \theta_1 x_1 \\ \theta_2 x_2 & 1 + \theta_2 x_1 \end{bmatrix} \tag{18}$$

and, following our approach, we define

$$\boldsymbol{G}(\boldsymbol{\theta}) = \boldsymbol{I}_{2,2} + \begin{bmatrix} x_2 & x_1 \\ 0 & 0 \end{bmatrix} \theta_1 + \begin{bmatrix} 0 & 0 \\ x_2 & x_1 \end{bmatrix} \theta_2. \tag{19}$$

Using Eq. (15), we generate a sample path $\boldsymbol{x}(0), \ldots, \boldsymbol{x}(T)$, and we generate the measurements $\bar{\boldsymbol{x}}(0), \ldots, \bar{\boldsymbol{x}}(T)$, with $T = 30$, by adding to each entry $\boldsymbol{x}(t)$ a noise with covariance matrix $\boldsymbol{Q(t)} = \sigma I$, for all t. In our experiments, we test different values of σ, and we set the weight matrix equal to the identity, i.e., $\boldsymbol{W}(t) = \boldsymbol{I}_{2,2}$ for all $t = 0, \ldots, T-1$, since the two variables have the same order of magnitude. Then, we solved the MIQP on a 4.7 GHz 10-Core Intel Core i7 using Mosek [37] implemented in MatLab R2022b with CVX [43, 44], which took just few seconds to reach a solution.



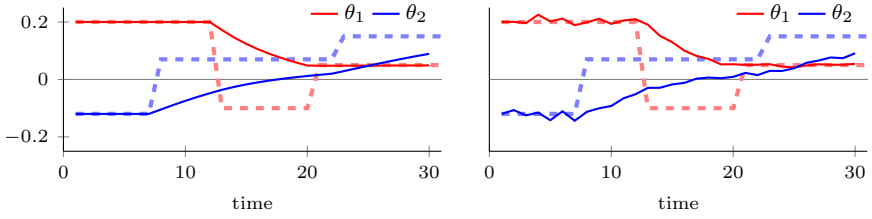
(a) $\sigma = 0$ (b) $\sigma = 0.005$ (c) $\sigma = 0.01$ (d) $\sigma = 0.02$

Figure 1: Validation on synthetic data with different levels of noise σ, with $k_1 = k_2 = 2$, $C = 0.25$, and $\gamma = 1$. The dashed curves illustrate the real profile of the two parameters.

The results, reported in Fig. 1, validate our method. In fact, in the absence of noise (i.e., $\sigma = 0$), our method is able to identify the exact value of the parameters in Eq. (16); see Fig. 1a. For small to moderate values of the noise (signal-to-noise ratios[1] of approximately 60 and 15 in Figs. 1b and 1c, respectively), we observe that the MIQP is still very effective in detecting the correct switching instants of the parameters and provides very accurate estimates of their values, as one can see in Figs. 1b and 1c, with a relative ℓ_2-norm error averaged over 100 realization of the noise equal to 3.15% and 9.74%, respectively. Predictably, the accuracy decreases as the noise reaches an order of magnitude comparable with the measurements (signal-to-noise ratio of approximately 4), as in Fig. 1d, where the mean relative ℓ_2-norm error is larger than 40%.

Finally, we compare our novel technique with a classical method for estimating time-varying parameters. Specifically, we replicate the scenarios with no noise and small noise (i.e., $\sigma = 0$ and $\sigma = 0.005$), and we estimate the time-varying parameters using an extended Kalman filter [24]. The results, illustrated in Fig. 2, suggest that the constraint on the number of parameter switches, and the ensuing mixed-integer approach, is key to obtain piecewise-linear estimates for the parameters. In fact, even in the absence of noise, methods based on extended Kalman filters tend to enforce a smooth evolution of the parameters

[1]The signal-to-noise ratio is the ratio between the mean square value of the state and the mean square value of the noise.



(a) $\sigma = 0$ (b) $\sigma = 0.005$

Figure 2: Parameter estimation replicating scenarios in Figs. 1a and 1b, using an extended Kalman filter. The dashed curves illustrate the real profile of the two parameters.

and hence generally fail in correctly detecting abrupt discontinuous changes.

3.2. Convex relaxation of Problem 1

While Problem 1 can be solved numerically in moderate-sized instances using state-of-the art solvers for MIQP [37], it is known that MIQPs are computational challenging in general, due to the presence of binary variables which make them NP-Complete [38]. It is therefore of interest to consider, along with the original MIQP formulation, also a tractable convex relaxation of Problem 1. While different methods can be used to achieve this goal [45], one such relaxation is readily obtained from the characterization in Eq. (12) by relaxing the domain of $\boldsymbol{z} \in \{0,1\}^n$ to the (convex) domain $\boldsymbol{z} \in [0,1]^n$. The following lemma holds.

Lemma 2. *For any $\boldsymbol{x} \in \mathbb{R}^n$ with $\|\boldsymbol{x}\|_\infty \leq C$ it holds that*

$$\|\boldsymbol{x}\|_0 \geq \frac{1}{C}\|\boldsymbol{x}\|_1 = \min_{\boldsymbol{z}\in[0,1]^n} \sum_{i=1}^{n} z_i, \;\; \text{subject to} \;\; -C\boldsymbol{z} \leq \boldsymbol{x} \leq C\boldsymbol{z}. \tag{20}$$

Proof. Since enlarging the feasible set can only improve the optimization objective, we have from Eq. (12) that

$$\|\boldsymbol{x}\|_0 \geq \min_{\boldsymbol{z}\in[0,1]^n} \sum_{i=1}^{n} z_i, \;\; \text{subject to} - C\boldsymbol{z} \leq \boldsymbol{x} \leq C\boldsymbol{z}. \tag{21}$$

Further, it is straightforward to check that the optimal z for the optimization problem in Eq. (21) is $z_i^* = |x_i|/C$, for all i, from which it follows that the optimal value of this problem is $\|\boldsymbol{x}\|_1/C$. □

From Lemma 2 we easily obtain the implications

$$\|\boldsymbol{x}\|_0 \leq k \Longrightarrow \frac{1}{C}\|\boldsymbol{x}\|_1 \leq k \iff \exists\, \boldsymbol{z} \in [0,1]^n : \sum_{i=1}^{n} z_i \leq k,\ -C\boldsymbol{z} \leq \boldsymbol{x} \leq C\boldsymbol{z}, \tag{22}$$

which are used to obtain the following convex relaxed version of Problem 1.

Problem 2. *A convex relaxation of the MIQP in (1) is given by the QP:*

$$\begin{aligned} \min_{\boldsymbol{\Theta}\in\mathbb{R}^{Tm},\boldsymbol{Z}\in[0,1]^{(T-1)\times m}} \quad & \sum_{t=0}^{T-1} \|\boldsymbol{W}(t)(\bar{\boldsymbol{x}}(t+1) - \boldsymbol{f_0}(t) - \boldsymbol{F}(t)\boldsymbol{\theta}(t))\|_2^2 \\ & +\gamma \sum_{t=0}^{T-1} \|\boldsymbol{G}(\boldsymbol{\theta}(t))\boldsymbol{Q}^{1/2}(t)\|_F^2 \\ s.t.: \quad & \mathbf{1}^\top \boldsymbol{z_i} \leq k_i, \quad i = 1,\dots,m, \\ & -C\boldsymbol{z_i} \leq \boldsymbol{E_i}\boldsymbol{D}\boldsymbol{\Theta} \leq C\boldsymbol{z_i}, \quad i = 1,\dots,m, \\ & \|\boldsymbol{\Theta}\|_\infty \leq C. \end{aligned} \tag{23}$$

The QP in Problem 2 can be easily solved in polynomial time using standard and efficient tools, see, e.g., Mosek [37]. However, it is worth noticing that this relaxation comes at a cost. In fact, the solution of the QP in Problem 2 may, in general, not satisfy the bound imposed on the number of switch instants k_i.

Remark 1 (Use of the estimated model)**.** It is worth to briefly discuss about the possible uses of a model identified via the proposed approach. Once we solved Problem 1, we find ourselves with the estimated time profiles of the model parameters $\boldsymbol{\theta}(t)$ over the horizon $[0, T-1]$. It is important to observe that these estimated parameters are determined so to "explain" the observations over the considered horizon, and shall not be used, in general, for extrapolating in time (i.e., for forward predicting) the system behavior. In fact, no hypotheses are made on the dynamic behavior of the parameter $\boldsymbol{\theta}(t)$ itself, so that we do not know how the parameters will evolve from T onward. A forward prediction can only be made by making further assumptions on the *future* behavior of $\boldsymbol{\theta}(t)$; e.g., one may assume that $\boldsymbol{\theta}(t)$ will stay constant at its last estimated value $\boldsymbol{\theta}(T-1)$ for a certain number of forward periods. Besides this specific case, the main use of the estimated model is for analyzing the observed dynamic phenomenon

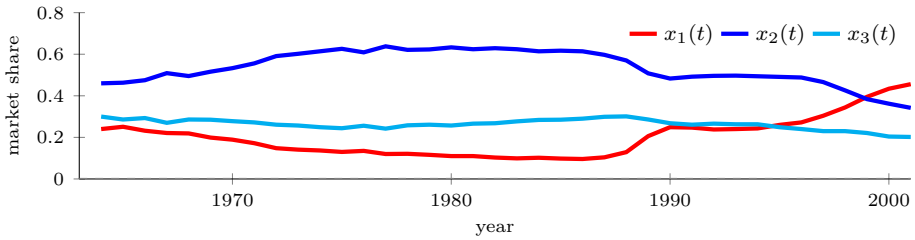


Figure 3: Temporal evolution of the market shares of *Asahi* (red), *Kirin* (blue), and *Sapporo* (cyan) from 1963 to 2001. Data from [40].

inside the estimation horizon $[0, T-1]$, and extracting useful information from the values of the parameters and their switching instants.

4. Case Studies

We next present two case studies in which we use the proposed approach for calibrating nonlinear models based on real-world data. The first case study, discussed in Section 4.1, presents the competition among different brands in the Japanese beer market and is focused on unveiling their interactions; the second one, reported in Section 4.2, is an epidemiological case study, which focused on the estimation of the parameters of a COVID-19 SIRD model based on Italy epidemiological data covering the entire duration of the pandemic.

4.1. Japanese beer market

In the first case study, we consider the competition among the three main brands of beer in the Japanese market —namely, *Asahi*, *Kirin*, and *Sapporo*— for which data are available with a one-year temporal granularity from 1963 to 2000. The dataset, publicly available in [40], comprises thus 38 data points consisting in the relative shares of the three brands in the market, as illustrated in Fig. 3.

We focus our analysis on unveiling the relations among the three brands. To this aim, inspired by [39], we represent the dynamics through a generalized Lotka–Volterra model, neglecting spontaneous growths and focusing on

the interaction effects. This yields the following three-dimensional nonlinear dynamical system:

$$\begin{aligned} x_1(t+1) &= x_1(t) + \theta_1(t)x_1(t)x_2(t) + \theta_2(t)x_1(t)x_3(t) \\ x_2(t+1) &= x_2(t) + \theta_3(t)x_1(t)x_2(t) + \theta_4(t)x_2(t)x_3(t) \\ x_3(t+1) &= x_3(t) + \theta_5(t)x_1(t)x_3(t) + \theta_6(t)x_2(t)x_3(t), \end{aligned} \tag{24}$$

where $\boldsymbol{x}(t) = [x_1(t), x_2(t), x_3(t)]^\top$ are the shares of *Asahi*, *Kirin*, and *Sapporo* in the Japanese beer market at time t, respectively. The system is characterized by a six-dimensional parameter vector $\boldsymbol{\theta}(t) = [\theta_1(t), \theta_2(t), \theta_3(t), \theta_4(t), \theta_5(t), \theta_6(t)]^\top$, where each entry captures the interaction effect between two brands. Specifically, $\theta_1(t)$ and $\theta_2(t)$ capture the effect on the temporal evolution of the shares of *Asahi* caused by its interaction with *Kirin* and *Sapporo*, respectively. Similarly, $\theta_3(t)$ and $\theta_4(t)$ capture the effect on *Kirin* of its interaction with *Asahi* and *Sapporo*, respectively; and $\theta_5(t)$ and $\theta_6(t)$ the effect on *Sapporo* of its interaction with *Asahi* and *Kirin*, respectively. Similar to Eq. (15), we can check that the map in Eq. (24) satisfies Assumption 1 and we can write the matrices $\boldsymbol{F}$ and $\boldsymbol{G}$ following similar arguments, obtaining

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} x_1x_2 & x_1x_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & x_1x_2 & x_2x_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_1x_3 & x_2x_3 \end{bmatrix}, \tag{25}$$

and

$$\begin{aligned} \boldsymbol{G}(\boldsymbol{\theta}) = \boldsymbol{I}_{3,3} &+ \begin{bmatrix} x_2 & x_1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\theta_1 + \begin{bmatrix} x_3 & 0 & x_1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\theta_2 + \begin{bmatrix} 0 & 0 & 0 \\ x_2 & x_1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\theta_3 \\ &+ \begin{bmatrix} 0 & 0 & 0 \\ 0 & x_3 & x_2 \\ 0 & 0 & 0 \end{bmatrix}\theta_4 + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ x_3 & 0 & x_1 \end{bmatrix}\theta_5 + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & x_3 & x_2 \end{bmatrix}\theta_6. \end{aligned} \tag{26}$$

We performed the parameter identification using both the original MIQP in Problem 1 and the QP in Problem 2, in order to compare their outcome and the computational complexity. After re-formulating some constraints through

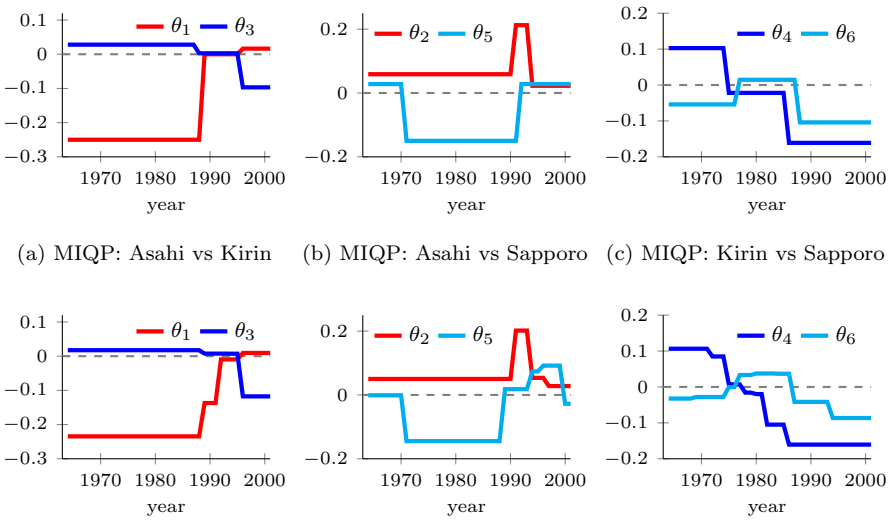
the addition of slack variables, the MIQP involves 1875 variables (222 of which are binary) and 1122 scalar constraints. Again, we solved both problems using Mosek implemented in MatLab R2022b with CVX.

In our numerical experiments, we set the weight matrix equal to the identity, i.e., $\boldsymbol{W}(t) = \boldsymbol{I}_{3,3}$ for all $t = 0, \ldots, T-1$, since the three variables have the same order of magnitude and comparable data quality. Moreover, since we had no information on the measurement error, but the three variables are bounded and slowly-changing over the entire time-window, we estimated the covariance matrix $\boldsymbol{Q}(t)$ using the data $\boldsymbol{x}(0), \ldots, \boldsymbol{x}(T)$, by setting $\boldsymbol{Q}(t) = \text{diag}([\sigma_1, \sigma_2, \sigma_3])$ as a time-invariant diagonal matrix with ith diagonal entry equal to the sample variance of the ith variable, i.e., $\sigma_i := \frac{1}{T-1} \sum_{t=1}^{T} (x_i(t) - \frac{1}{T} \sum_{t=1}^{T} x_i(t))^2$. In this scenario, the objective function of our optimization problems reduces to

$$\sum_{t=0}^{T-1} \|\bar{\boldsymbol{x}}(t+1) - \boldsymbol{x}(t) - \boldsymbol{F}(\boldsymbol{x}(t))\boldsymbol{\theta}(t)\|_2^2 + \gamma ||\text{diag}([\sigma_1, \sigma_2, \sigma_3]) G(\boldsymbol{\theta}(t))||_F^2, \quad (27)$$

with $\boldsymbol{F}(\boldsymbol{x})$ from Eq. (25) and $\boldsymbol{G}(\boldsymbol{\theta})$ from Eq. (26).

First, we used the MIQP problem formulation to identify the parameters of Eq. (24), setting $k_i = 2$, for all $i = 1, \ldots, 6$, $C = 0.25$, and $\gamma = 2$. The output of our parameter identification algorithms are reported in panels (a–c) of Fig. 4. The results provide interesting insight into the system analyzed. In fact, they allow not only to shed light on the interactions between the three different brands, but also to identify the key switching instants in the Japanese beer market. In fact, following [39], the signs of the parameters determine the (competitive or mutualistic) role between two brands: if both negative they are in pure competition; if both positive they are mutualistic; while if one is positive and the other negative, they have a prey-predator type of interaction. For instance, Fig. 4(c) shows two interesting turning points in the interactions between *Kirin* and *Sapporo*. Until the second part of the 1970s, the two brands display a pray-predator type of interaction beneficial to Kirin (in fact, θ_4 is positive and θ_6 is negative). Then, the sign of such interaction is reverted in 1977 in favor of *Sapporo*. This coincides with the year in which *Sapporo* introduced its first draft beer, named "Black Label" [41]. Then, this period



(a) MIQP: Asahi vs Kirin (b) MIQP: Asahi vs Sapporo (c) MIQP: Kirin vs Sapporo

(d) QP: Asahi vs Kirin (e) QP: Asahi vs Sapporo (f) QP: Kirin vs Sapporo

Figure 4: Comparison of the six time-varying parameters estimated using the MIQP (a-c) and using the QP relaxation (d-f) for the Japanese beer market case study, with $k_i = 2$, for all $i = 1, \ldots, 6$, $C = 0.25$, and $\gamma = 2$.

lasts until the end of the 1980s, when the interactions started becoming purely competitive. Indeed, it is known that this last period was characterized by an intense competition between the brands over dry beers, which goes under the name of "dry war" [41]. In fact, in 1987 Asahi launched *Super Dry* gaining momentum over the other brands, which started a strong competition, as we observed in panel (c). This is reflected in the parameters in Fig. 4. In fact, we observe an increase in both red curves at the end of the 1980s in panels (a) and (b), while the interaction between Kirin and Sapporo becomes purely competitive, as observed in the above. Hence, this simple case study elucidates how our algorithm can be used to understand the impact of historical events on the market and detect key phases and turing points.

Second, we performed the same parameter identification using the QP relaxation in Problem 2. The results are reported in panels (d–f) of Fig. 4. A comparison of these panels with the previous ones (a–c) suggests that the out-

Table 1: Time needed (measured in seconds) to compute a optimal and sub-optimal solutions of the MIQP for the Japanese beer market case study, for increasing values of k and different relative optimality tolerance employed by the mixed-integer optimizer (5%, 1%, or 0.5%), compared to the time needed to solve the QP (last row).

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
optimal	43.67	19,645.12	>1 day	>1 day	>1 day	>1 day
0.5%	1.28	23.66	445.00	2,261.47	6,784.81	16,467.30
1%	0.23	1.33	2.73	11.75	22.58	41.72
5%	0.11	0.13	0.17	0.25	0.45	1.28
QP	0.81	0.78	0.95	0.82	1.09	1.04

come of the QP relaxation was a good proxy for the MIQP in the case of interest. In fact, the solution of the QP approximates with sufficiently good accuracy not only the value of the majority of the parameters (the sign of the parameter, which determines whether the brands have a positive or negative interaction, coincides in more that 96% of the instances), but also their main switching instants (relative ℓ_2-norm error of about 10%). Only minor differences are observed. In particular, since the solution of the QP is allowed to change value more than k_i times, some abrupt changes of parameters in the solutions of the MIQP are split into two smaller steps —see, for instance, the evolution of the parameter θ_4 (blue curve) in panels (c) and (f). Even for θ_4, however, the signs coincide in all instants, except for one, in which they are both close to 0. It is worth noticing that, while the QP took less than one second to be solved, the computation of the exact solution of the MIQP required approximately 5.5 hours.

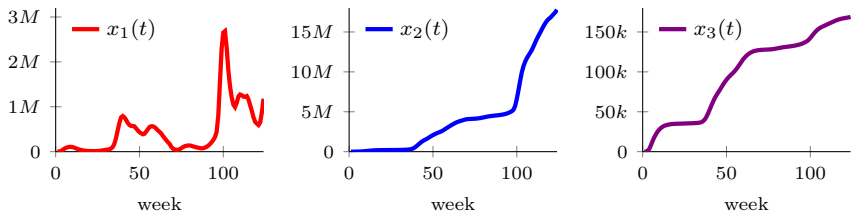
Based on this observation, we performed an additional study to investigate the computational effort required to solve the MIQP. Specifically, we computed the time needed to compute sub-optimal solutions[2] of the MIQP for increasing

[2]A solution with 1% tolerance is obtained by allowing the objective function to be 1% larger than its optimal value.

values of k_i, by setting $k_i = k$, for all $i = 1, \ldots, 6$, and testing values of k from 1 to 6. The results are reported in Table 1, and they also show that sub-optimal solutions of the MIQP with low relative optimality tolerance can be efficiently computed. Importantly, while sub-optimal solutions with larger tolerance show noticeable differences with respect to the exact solution of the MIQP reported in Fig. 4 (relative ℓ_2-norm error greater than 25% when tolerance is 1%), the one obtained with 0.5% tolerance for $k = 2$ (computed in less than 24 seconds) is a good proxy of the optimal one (relative ℓ_2-norm error smaller than 10%). To summarize, the analysis of this case study suggests that the MIQP approach developed in this paper could be useful to identify the parameters of low-dimensional systems, possibly using sub-optimal solutions. Furthermore, the solution of the convex relaxation in Problem 2 is in general a good proxy of MIQP solution and its use is suggested when dealing with higher dimensional systems, as we shall see in the next case study. In fact, our convex relaxation has performances that are comparable with a sub-optimal solutions of the MIQP with 0.5% tolerance (no more than 10% relative ℓ_2-norm error), but its computational time does not grow with k_i. This suggests that, among the two approximations, the convex relaxation could be preferred when the parameters have many changing points, while for $k_i \leq 3$, sub-optimal solutions of the MIQP with at most 0.5% tolerance can be computed in a reasonable time and may have better performances.

4.2. COVID-19 in Italy

In the second case study, we calibrate an epidemic model using data on the COVID-19 pandemic in Italy, from its inception (February 24, 2020) till July 12, 2022. Specifically, we use publicly available data reported by the Civil Protection Department of the Italian Government on the number of reported cases, recoveries, and deaths [46]. We aggregated data on a weekly basis to account for the natural oscillations in the number of tests performed during the different days of the week. Hence, our dataset comprises 124 weekly data points for reported cases, total recovered people, and total deaths, as shown in Fig. 5.



(a) Reported cases (b) Recovered (c) Total deaths

Figure 5: COVID-19 weekly figures from February 24, 2020 to July 12, 2022 in Italy, as reported in the official dataset available in [46].

We formulate the epidemic model as a discrete-time SIRD model with time-varying parameters, similar to the one proposed in [10], with time unit equal to one week. Such a temporal granularity allows for neglecting the delay between contagion and infectiousness, since the duration of the latency period is shorter than one week in the large majority of the cases [47]. We observe that the SIRD model can be written in the form of Eq. (1), with affine mapping with respect to the parameters. Let us define the state variable vector $\boldsymbol{x}(t) := [x_1(t), x_2(t), x_3(t)]^\top$, whose entries are respectively the total number of infected, recovered, and deceased individuals in the population at time t, with the number of susceptible individuals equal to $n - x_i(t) - x_2(t) - x_3(t)$, where n is the total population size. Let us further define the time-varying parameter vector $\boldsymbol{\theta}(t) = [\theta_1(t), \theta_2(t), \theta_3(t)]^\top$ to contain the infection rate, recovery rate, and death rate at time t, respectively. Then, we can write the following dynamical system:

$$\begin{aligned} x_1(t+1) &= x_1(t) + \tfrac{1}{n}\theta_1(t)(n - x_1(t) - x_2(t) - x_3(t))x_1(t) - (\theta_2(t) + \theta_3(t))x_1(t) \\ x_2(t+1) &= x_2(t) + \theta_2(t)x_1(t) \\ x_3(t+1) &= x_3(t) + \theta_3(t)x_1(t), \end{aligned} \tag{28}$$

where we note that the system is autonomous, i.e., $\boldsymbol{u}(t) = \mathbf{0}$, for all $t \in \mathbb{N}$. Note that the second and third equations are linear and of straightforward interpretation: each week, a fraction $\theta_2(t)$ of the individuals who were infected the previous week have recovered and a fraction $\theta_3(t)$ of them have died. The

first equation, instead, besides having a negative contribution equal to the two terms described in the above, has a positive nonlinear term that accounts for new infections. In particular, the number of newly infected individuals is equal to the product between susceptible individuals, $n - x_1(t) - x_2(t) - x_3(t)$, the fraction of infected individuals in the population $\frac{1}{n}x_1(t)$, and the infection rate $\theta_1(t)$. This nonlinear term is typical from the epidemic literature, see [42] for more details.

We can check that the map in Eq. (28) satisfies Assumption 1. In fact, we can write it as $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{f_0}(\boldsymbol{x}) + \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\theta}$, with

$$\boldsymbol{f_0}(\boldsymbol{x}) = \boldsymbol{x}, \qquad \boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} \frac{1}{n}x_1(n - x_1 - x_2 - x_3) & -x_1 & -x_1 \\ 0 & x_1 & 0 \\ 0 & 0 & x_1 \end{bmatrix}, \tag{29}$$

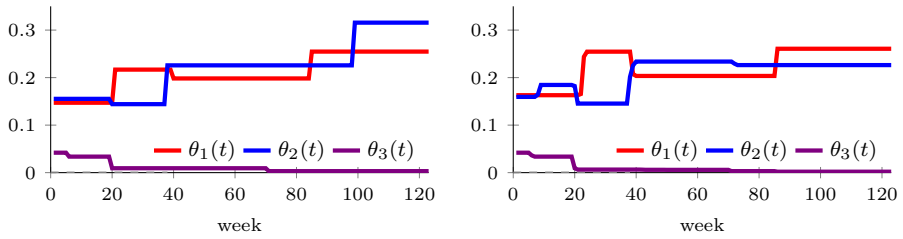
where we have omitted the time-dependencies of the variables and parameters for brevity. Note that we can write the Jacobian of the map in Eq. (29) as

$$\boldsymbol{J}_f(\boldsymbol{x}, \boldsymbol{\theta}) = \begin{bmatrix} 1 + \frac{1}{n}\theta_1(n - 2x_1 - x_2 - x_3) - \theta_2 - \theta_3 & 0 & 0 \\ \theta_2 & 1 & 0 \\ \theta_3 & 0 & 1 \end{bmatrix} \tag{30}$$

and

$$\boldsymbol{G}(\boldsymbol{\theta}) = \boldsymbol{I}_{3,3} + \begin{bmatrix} \frac{1}{n}(n - 2x_1 - x_2 - x_3) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\theta_1 + \begin{bmatrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\theta_2 + \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\theta_3 \tag{31}$$

Before formalizing the parameter identification problem, we observe from Fig. 5 that the three components of the state variable vector $\boldsymbol{x}(t)$ have different orders of magnitude across them and large variations in time, due to different testing policies and accumulation of recovered and total deaths. This calls for the definition of a non-trivial sequence of weight matrices $\boldsymbol{W}(t)$ to re-scale the components of $\boldsymbol{x}(t)$ to comparable quantities, and thus avoid the objective function to be dominated by just the few entries with larger order of magnitude. In particular, we set $\boldsymbol{W}(t) = \text{diag}(\bar{\boldsymbol{x}}(t)^{-1})$, so that each the term in the first summand



(a) MIQP (b) QP

Figure 6: Parameters estimated for the COVID-19 case study obtained as a) a sub-optimal solution of the MIQP (40% tolerance) and b) the solution of the QP relaxation for $k_1 = k_2 = k_3 = 3$ with $C = 0.5$ and $\gamma = 0.2$.

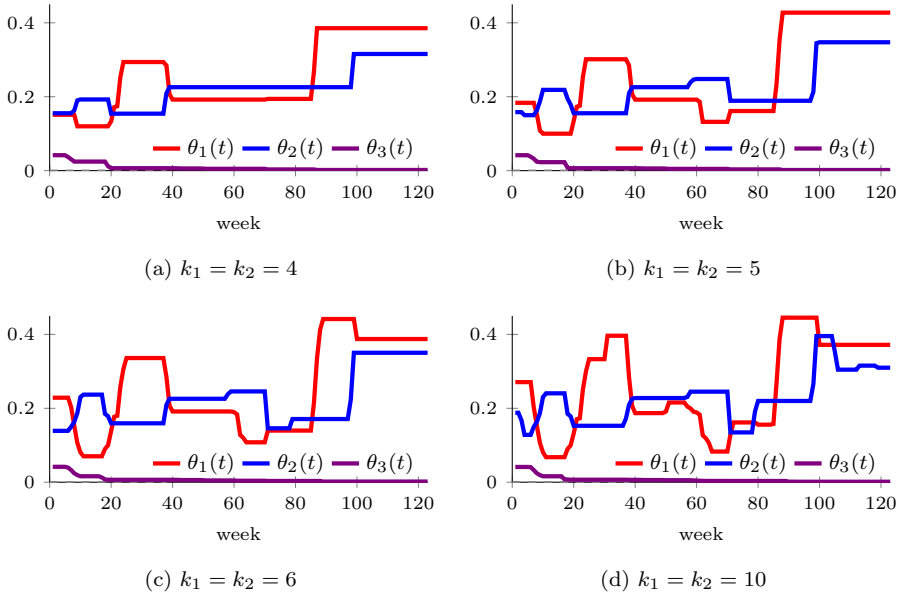
of the objective function measures the relative error of the predicted state with respect to the measured one. Then, similar to the previous case study, we set $\boldsymbol{Q}(t)$ as a time-invariant diagonal matrix with ith diagonal entry equal to the sample variance of the ith variable.

In the 124 weeks considered in the dataset, several events occurred, including two major lockdown periods (Spring 2020 and Winter-Spring 2021), the vaccination and booster campaigns (Spring-Summer 2021 and Fall 2021-Winter 2022, respectively), and several epidemic waves due to the diffusion of new variants. Hence, we expect the presence of several switching instants in the values of the piecewise constant parameters θ_1, θ_2, and θ_3. We verify this conjecture by showing in Fig. 6 that the parameters estimated by setting too small values of k_i are not particularly informative on the different phases of the pandemic, since they cannot capture, e.g., the key changes that occurred in the infection rate when lockdown measures are imposed at mid March 2020. However, increasing k_i makes the solution of the MIQP hard to compute in practice. Indeed, computing a sub-optimal solution of the MIQP with $k_1 = k_2 = k_3 = 3$ at 20% sub-optimality tolerance takes almost 4 hours, and for $k_1 = k_2 = k_3 = 6$ after 1 day of optimization the sub-optimality tolerance is still larger than 40%. This example highlights the importance of our computationally-efficient QP relaxation, which runs in few seconds even in this more complex case study, and provides a good proxy of the MIQP solution (see the results in Section. 4.1, and

compare the two panels in Fig. 6). Hence, in the following we focus on solving the QP, for which we use again Mosek implemented in MatLab R2022b with CVX. After the addition of suitable slack variables for obtaining a standard QP formulation, we get 4009 scalar variables and 1338 constraints.

We solved the QP in Problem 2 with different values of k_i and we compared the results obtained. After a preliminary analysis, we observe that the parameter $\theta_3(t)$ associated with the mortality of the disease tends to have few changes with respect the other parameters with a three-step decrease, corresponding with i) the start of the decrease of the first wave (mid April 2020), ii) the diffusion of antigenic tests, which increased the number of detected mild cases (end of summer 2020), and iii) the peak of the second-dose vaccination campaign (summer 2021). Differently, the value of the parameter associated to the contagion and recovery rates show more switching instants. This may be due to the fact that these parameters are not only affected by the characteristics of the disease and on the state of the vaccination campaign, but are also strongly impacted by human behavior. In view of these preliminary observations, we decided to fix the value of $k_3 = 3$, and vary the value of $k_1 = k_2$ in the range $k_1 = k_2 \in \{4, 5, 6, 10\}$.

The results are reported in Fig. 7. Unsurprisingly, we observe that, as we increase $k_1 = k_2$, the parameter identification method becomes able to detect an increasing number of key switching instants for the model parameters. More interestingly, we also observe that the effect of increasing k seems to be incremental: first, the algorithm detects the most crucial changes; then, as k grows large, smaller changes are added to those already detected. For $k_1 = k_2 \geq 5$, we observe that the algorithm identifies the key phases of the first pandemic wave, with a steep decrease of the infection rate and increase of the recovery rate in correspondence of the first lock down, which is then reverted at the end of the summer, slowly leading to the second wave in fall-winter 2020. Then, the infection rate $\theta_1(t)$ show a rapid decrease in correspondence to the start of the second lockdown, followed by another decrease, during the peak of the vaccination campaign; while the recovery rate shows an opposite, but less pronounced,



(a) $k_1 = k_2 = 4$ (b) $k_1 = k_2 = 5$

(c) $k_1 = k_2 = 6$ (d) $k_1 = k_2 = 10$

Figure 7: Parameters estimated for the COVID-19 case study using the QP relaxation for different values of k_1 and k_2. Common parameters are $k_3 = 3$, $C = 0.5$ and $\gamma = 0.2$.

behavior. Besides capturing parameter changes due to human behavior, our algorithm identifies also key natural events, such as the diffusion of new variants. For instance, the diffusion of the Omicron variant can be seen in the steep increase in θ_1 around week 90 (mid November 2021).

Finally, we would like to remark that, in this case study, we used a simple SIRD model to describe the evolution of COVID-19. Such a choice has been done for the sake of simplicity. More complex and realistic epidemic models that include, e.g., further health states [48] or spatially-distributed populations [49, 50], may be embedded within our framework, and our parameter identification method can be used to shed light on further important aspects, such as estimating the fraction of non-detected infected individuals or evaluating the impact of spatial heterogeneity.

5. Conclusion

In this paper we presented an optimization-based system identification approach that can be effectively used for estimating time-varying but piecewise constant parameters in a broad class of nonlinear dynamical models. The proposed method, which is based on casting the estimation problem as a MIQP, has several advantages, including the possibility to set the number of changes in the model parameters and incorporating additional information available on the data quality, without relaying on assumptions on the switching rule or on knowledge of the switching instants. The proposed optimization technique can be conveniently approximated via sub-optimal solutions of the MIQP as well as relaxed into a computationally-efficient QP, whose solution is shown to be in practice a good proxy of the original one. After validation on a synthetic case study, the practical use of our method and its approximations was finally demonstrated on two case studies.

The results presented in this paper, besides demonstrating the effectiveness of the proposed method, arise a few points for discussion, which pave the way for directions of future research. First, our algorithm requires the knowledge (or, at least, an estimate) of the covariance matrix of the measurement errors. In our case studies, we have assumed it to be constant and we have estimated it using the covariance of the measured data. However, more refined methods to estimate such a matrix should be designed in order to adopt our method in real-world scenarios, especially when measurement errors change over time. Second, our case studies suggest that the parameter k_i that determines the number of times the parameter varies during the entire time-window is crucial. In order to tune such a parameter, one should use the information available on the system, and combine it with heuristics. A simple heuristic could be to re-iterate the procedure, starting with k_i large, and then decrease it until the solution remains quite stable. Future work should focus on testing such heuristic and design more refined ones. Third, despite the empirical observation that the solution of the QP is often a sufficiently good proxy of the MIQP

solution, more refined approximations methods should be designed, possibly leveraging finer regularization techniques for the cardinality bound [45], and rigorous theoretical guarantees on their performance must be derived. Fourth, an extensive comparison of the performance of different methods to estimate time-varying parameters with abrupt changes (e.g., comparing our method with algorithms based on recursive least-squares with forgetting factors [20, 22] or switching systems [33]) is an important direction of future investigation in order to determine key advantages and disadvantages of each of these methods.

Despite its generality and flexibility, our parameter identification approach is not exempt from limitations. First, we assumed that the entire state of the system is observable, up to some level of uncertainty. In many realistic scenarios, however, only some of the variables of the systems can be observed (e.g., the number of detected infected individuals in an epidemiological application), while it is not possible to have information on other variables (e.g., the undetected infected individuals). Future extensions of the approach shall be developed to account for cases in which only part of the state is observable. Second, our method relies on a first-order expansion of the system's dynamics. While this is a standard practice and it is admissible in many real-world applications such as those considered in our case studies, further analysis in needed to deal with highly nonlinear dynamics, where high-order terms may not be neglected.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L. Ljung, Perspectives on system identification, Annual Reviews in Control 34 (1) (2010) 1–12. doi:10.1016/j.arcontrol.2009.12.001.

[2] K. Åström, P. Eykhoff, System identification—a survey, Automatica 7 (2) (1971) 123–162. doi:10.1016/0005-1098(71)90059-8.

[3] T. Söderström, P. Stoica, System Identification, Prentice Hall, Hoboken NJ, US, 1989.

[4] L. Ljung, System Identification: Theory for the User, 2nd Edition, Pearson Education, London, UK, 1998.

[5] L. Ljung, H. Hjalmarsson, H. Ohlsson, Four encounters with system identification, European Journal of Control 17 (5) (2011) 449–471. doi:10.3166/ejc.17.449-471.

[6] V. Strejc, Least squares parameter estimation, Automatica 16 (5) (1980) 535–550. doi:10.1016/0005-1098(80)90077-1.

[7] R. Kashyap, Maximum likelihood identification of stochastic linear systems, IEEE Transactions on Automatic Control 15 (1) (1970) 25–34. doi:10.1109/TAC.1970.1099344.

[8] L. Vandenberghe, Convex optimization techniques in system identification, in: Proceedings of the 16th IFAC Symposium on System Identification, Vol. 45, 2012, pp. 71–76. doi:10.3182/20120711-3-BE-2027.00244.

[9] P. E. Paré, C. L. Beck, T. Başar, Modeling, estimation, and analysis of epidemics over networks: An overview, Annual Reviews in Control 50 (2020) 345–360. doi:10.1016/j.arcontrol.2020.09.003.

[10] G. C. Calafiore, C. Novara, C. Possieri, A time-varying SIRD model for the COVID-19 contagion in Italy, Annual Reviews in Control 50 (2020) 361–372. doi:10.1016/j.arcontrol.2020.10.005.

[11] R. Alisic, P. E. Paré, H. Sandberg, Change time estimation uncertainty in nonlinear dynamical systems with applications to COVID-19, International Journal of Robust and Nonlinear Control 33 (9) (2023) 4732–4760. doi:10.1002/rnc.5974.

[12] H. U. Voss, J. Timmer, J. Kurths, Nonlinear dynamical system identification from uncertain and indirect measurements, International Journal of Bifurcation and Chaos 14 (06) (2004) 1905–1933. doi:10.1142/s0218127404010345.

[13] M. Schwaab, E. C. Biscaia, Jr., J. L. Monteiro, J. C. Pinto, Nonlinear parameter estimation through particle swarm optimization, Chemical Engineering Science 63 (6) (2008) 1542–1552. doi:10.1016/j.ces.2007.11.024.

[14] A. Chiuso, G. Pillonetto, System identification: A machine learning perspective, Annual Review of Control, Robotics, and Autonomous Systems 2 (1) (2019) 281–304. doi:10.1146/annurev-control-053018-023744.

[15] M. Norgaard, O. Ravn, N. Poulsen, L. Hansen, Neural Networks for Modelling and Control of Dynamic Systems, Springer, London, UK, 2000.

[16] M. Forgione, D. Piga, Continuous-time system identification with neural networks: Model structures and fitting criteria, European Journal of Control 59 (2021) 69–81. doi:10.1016/j.ejcon.2021.01.008.

[17] J. Noël, G. Kerschen, Nonlinear system identification in structural dynamics: 10 more years of progress, Mechanical Systems and Signal Processing 83 (2017) 2–35. doi:10.1016/j.ymssp.2016.07.020.

[18] J. Schoukens, L. Ljung, Nonlinear system identification: A user-oriented road map, IEEE Control Systems Magazine 39 (6) (2019) 28–99. doi:10.1109/MCS.2019.2938121.

[19] M. Tsatsanis, G. Giannakis, Time-varying system identification and model validation using wavelets, IEEE Transactions on Signal Processing 41 (12) (1993) 3512–3523. doi:10.1109/78.258089.

[20] C. Paleologu, J. Benesty, S. Ciochina, A robust variable forgetting factor recursive least-squares algorithm for system identification, IEEE Signal Processing Letters 15 (2008) 597–600. doi:10.1109/LSP.2008.2001559.

[21] Y. Li, H.-l. Wei, S. A. Billings, Identification of time-varying systems using multi-wavelet basis functions, IEEE Transactions on Control Systems Technology 19 (3) (2011) 656–663. doi:10.1109/TCST.2010.2052257.

[22] A. Goel, D. S. Bernstein, A targeted forgetting factor for recursive least squares, in: 2018 IEEE Conference on Decision and Control (CDC), IEEE, 2018, p. 3899–3903. doi:10.1109/cdc.2018.8619181.

[23] K. Esfandiari, K. S. Narendra, Identification and control of linear systems with piece-wise constant parameters, in: 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2022, pp. 3006–3011. doi:10.1109/SMC53654.2022.9945441.

[24] E. Wan, R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation, in: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, 2000, pp. 153–158. doi:10.1109/ASSPCC.2000.882463.

[25] Z. Liu, L. Vandenberghe, Semidefinite programming methods for system realization and identification, in: Proceedings of the 48h IEEE Conference on Decision and Control held jointly with 2009 28th Chinese Control Conference, 2009, pp. 4676–4681. doi:10.1109/CDC.2009.5400177.

[26] N. Beck, Time-varying parameter regression models, American Journal of Political Science 27 (3) (1983) 557. doi:10.2307/2110985.

[27] A. Glushchenko, K. Lastochkin, Unknown piecewise constant parameters identification with exponential rate of convergence, International Journal of Adaptive Control and Signal Processing 37 (1) (2023) 315–346. doi:10.1002/acs.3533.

[28] J. Roll, A. Bemporad, L. Ljung, Identification of piecewise affine systems via mixed-integer programming, Automatica 40 (1) (2004) 37–50. doi:10.1016/j.automatica.2003.08.006.

[29] A. Bemporad, A. Garulli, S. Paoletti, A. Vicino, A bounded-error approach to piecewise affine system identification, IEEE Transactions on Automatic Control 50 (2005) 1567–1580. doi:10.1109/TAC.2005.856667.

[30] A. S. Willsky, Detection of abrupt changes in dynamic systems, Springer-Verlag, 2005, p. 27–49. doi:10.1007/bfb0006388.

[31] L. Bako, K. Boukharouba, S. Lecoeuche, An $\ell_0 - \ell_1$ norm based optimization procedure for the identification of switched nonlinear systems, in: Proceedings of the 49th IEEE Conference on Decision and Control, 2010, pp. 4467–4472. doi:10.1109/CDC.2010.5717199.

[32] F. Lauer, G. Bloch, R. Vidal, A continuous optimization framework for hybrid system identification, Automatica 47 (3) (2011) 608–613. doi:10.1016/j.automatica.2011.01.020.

[33] F. Bianchi, M. Prandini, L. Piroddi, A randomized two-stage iterative method for switched nonlinear systems identification, Nonlinear Analysis: Hybrid Systems 35 (2020) 100818. doi:10.1016/j.nahs.2019.100818.

[34] L. Ljung, S. Gunnarsson, Adaptation and tracking in system identification—a survey, Automatica 26 (1) (1990) 7–21. doi:10.1016/0005-1098(90)90154-a.

[35] P. Fearnhead, Exact and efficient bayesian inference for multiple changepoint problems, Statistics and Computing 16 (2) (2006) 203–213. doi:10.1007/s11222-006-8450-8.

[36] G. C. Calafiore, G. Fracastoro, Age structure in SIRD models for the COVID-19 pandemic—a case study on Italy data and effects on mortality, PLOS ONE 17 (2) (2022) 1–17. doi:10.1371/journal.pone.0264324.

[37] E. D. Andersen, C. Roos, T. Terlaky, On implementing a primal-dual interior-point method for conic quadratic optimization, Mathematical Programming 95 (2) (2003) 249–277. doi:10.1007/s10107-002-0349-3.

[38] A. D. Pia, S. S. Dey, M. Molinaro, Mixed-integer quadratic programming is in NP, Mathematical Programming 162 (1-2) (2016) 225–240. doi:10.1007/s10107-016-1036-0.

[39] A. Marasco, A. Picucci, A. Romano, Market share dynamics using Lotka–Volterra models, Technological Forecasting and Social Change 105 (2016) 49–62. doi:10.1016/j.techfore.2016.01.017.

[40] A. Marasco, A. Picucci, A. Romano, Determining firms' utility functions and competitive roles from data on market shares using Lotka–Volterra models, Data in Brief 7 (2016) 709–713. doi:10.1016/j.dib.2016.03.020.

[41] J. W. Alexander, Brewed in Japan: The Evolution of the Japanese Beer Industry, University of Hawaii Press, Honolulu HI, US, 2014.

[42] F. Brauer, C. Castillo-Chavez, Mathematical models in population biology and epidemiology, 2nd Edition, Springer, New York NY, USA, 2012. doi:10.1007/978-1-4614-1686-9.

[43] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, `http://cvxr.com/cvx` (Mar. 2014).

[44] M. Grant, S. Boyd, Graph implementations for nonsmooth convex programs, in: V. Blondel, S. Boyd, H. Kimura (Eds.), Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, 2008, pp. 95–110.

[45] A. M. Tillmann, D. Bienstock, A. Lodi, A. Schwartz, Cardinality minimization, constraints, and regularization: A survey, arXiv preprint 2106.09606 (2022). doi:10.48550/arXiv.2106.09606.

[46] Dipartimento della Protezione Civile, COVID-19 Italia - Monitoraggio situazione, `https://github.com/pcm-dpc/COVID-19` (2020).

[47] H. Xin, Y. Li, P. Wu, Z. Li, E. H. Y. Lau, Y. Qin, L. Wang, B. J. Cowling, T. K. Tsang, Z. Li, Estimating the latent period of coronavirus disease

2019 (COVID-19), Clinical Infectious Diseases 74 (9) (2021) 1678–1681. doi:10.1093/cid/ciab746.

[48] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, Nature Medicine 26 (6) (2020) 855–860. doi:10.1038/s41591-020-0883-7.

[49] F. Della Rossa, D. Salzano, A. Di Meglio, F. De Lellis, M. Coraggio, C. Calabrese, A. Guarino, R. Cardona-Rivera, P. De Lellis, D. Liuzza, F. Lo Iudice, G. Russo, M. di Bernardo, A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic, Nature Communications 11 (1) (2020) 5106. doi:10.1038/s41467-020-18827-5.

[50] F. Parino, L. Zino, M. Porfiri, A. Rizzo, Modelling and predicting the effect of social distancing and travel restrictions on COVID-19 spreading, Journal of the Royal Society Interface 18 (2021) 20200875. doi:10.1098/rsif.2020.0875.