Contributions to the identification of continuous-time models from irregularly sampled data

Fengwei Chen

École doctorale IAEM Lorraine

Contributions à l'identification de modèles à temps continu à partir de données échantillonnées à pas variable

Thèse présentée et soutenue publiquement le 21 Novembre 2014 pour l'obtention du

Doctorat de l'Université de Lorraine

Mention Automatique, Traitement du Signal et des Images, Génie Informatique

par

Fengwei CHEN

Composition du jury

Président:	Didier MAQUIN	Professeur Université de Lorraine
Rapporteurs:	Martine OLIVI	Chargée de Recherche, HDR INRIA
	Xavier BOMBOIS	Maître de Conférences, HDR Delft University of Technology
Examinateurs:	Jonas SJÖBERG	Professeur Chalmers University of Technology
Co-directeur:	Marion GILSON	Professeur Université de Lorraine
Directeur:	Hugues GARNIER	Professeur Université de Lorraine

Centre de Recherche en Automatique de Nancy
UMR 7039 Université de Lorraine - CNRS
2 rue Jean Lamour, F-54519 Vandoeuvre-les-Nancy, France

Remerciements

A Monsieur le Professeur Hugues GARNIER
Merci d'avoir accepté de diriger cette thèse. Merci pour votre disponibilité, votre écoute, vos conseils et votre patience. Ces années à vos côtés m'ont permis d'acquérir de nombreuses connaissances scientifiques. Soyez assurée de mon profond respect et de ma grande reconnaissance.

A Madame le Professeur Marion GILSON
Merci d'avoir accepté de co-diriger cette thèse. Merci pour votre sympathie,votre disponibilité, vos idées et conseils, ainsi que pour votre aide précieuse. Soyez assuré de mon profond respect et de ma grande reconnaissance.

A Madame le docteur Martine OLIVI et Monsieur le docteur Xavier BOMBOIS
Merci d' avoir accepté de juger ce travail en qualité de rapporteur.

A Monsieur le professeur Jonas SJÖBERG et Monsieur le professeur Didier MAQUIN
Merci d'avoir accepté de faire partie de ce jury. Veuillez trouver ici le témoignage de ma reconnaissance et de mon plus profond respect.

A tous mes collègues et amis du Laboratoire
Un grand Merci 'a Monsieur le professeur Dominique SAUTER pour son accueil lors de mon arrivée à la gare de Nancy.

Je suis très reconnaissant à Madame le Docteur Floriane pour sa gentillesse et sa sympathie, pour m'avoir aider dans la vie quotidienne notamment dans les démarches administratives.

Je tiens à remercier Monsieur le Docteur Vincent pour le partage de ses connaissances.

Je tiens à remercier chaleureusement *Julien* pour son sympathie et son accompagnement notamment pendant des congrès; *Boyi* pour la discussion intéressante et les bons moments passés ensembles; *Meriem*, *Tatiana*, *Brandon* et *Jean-Baptiste* pour leur accompagnment en trouvant le logement et en faisant l'inscription; *Arturo* et *Yusuf* pour son lecture de mon manuscrit.

A mes amis chinois
Je vous remercie pour vos présences, vos amitiés et pour tout ce que nous avons partagé. plus particulièrement à *Xiaoyan, Yanan, Yong XU, Jing, Jingwen, Xin, Yongxin, Xiaoyun, Kun, Yong HU, Song, Chongsheng, Xuanxuan, Yuanzhao, Jun, Yue, Quan*.

A mes parents pour leur soutien et leur amour inestimable.

Un grand merci à China Scholarship Council (CSC) pour ce financement de thèse.

Résumé et contributions

Les modèles mathématiques des systèmes dynamiques sont requis dans la plupart des domaines de recherche scientifique et ils se présentent sous formes diverses, équations différentielles, équations aux différences, représentation d'état et ou encore sous forme fonctions de transfert. L'approche la plus largement utilisée pour la modélisation mathématique se fonde sur la construction d'équations mathématiques issues des lois de la Physique, connues pour décrire de façon précise le comportement du système. Cette approche présente plusieurs inconvénients, ainsi les modèles obtenus sont souvent difficiles à identifier du fait de leur complexité. Cette dernière rend également difficile l'utilisation de ces modèles dans des applications telles que la conception du système de commande.

Une alternative à cette modélisation purement physique est l'identification des systèmes, fondée sur l'estimation d'un modèle à partir de données expérimentales. Un des avantages de cette méthode est qu'elle peut être appliquée à tout système dont on peut recueillir des données. Par ailleurs, les modèles obtenus sont parcimonieux et décrivent au mieux le comportement du systèmes autour de son point de fonctionnement. Ces modèles peuvent être soit de type boîte noire, sous forme d'état ou de fonction de transfert, soit de type boîte grise avec un lien directe avec les équations de la Physique.

La plupart des systèmes dynamiques évoluent naturellement à temps continu (TC), et sont décrits par des équations différentielles, telles que les équations de conservation, de masse, de diffusion etc. Cependant, paradoxalement, les approches d'identification de systèmes les plus connues sont fondées sur des modèles à temps discret (TD), sans lien direct avec les équations de la Physique, essentiellement du fait du développement du "tout numérique" de ces vingt dernières années.

Ainsi, beaucoup moins d'attention a été consacrée à la modélisation de modèles à TC à partir de données échantillonnées et par conséquent, ces méthodes sont encore peu connues alors qu'elles peuvent parfois être mieux adaptées à certains problèmes d'identification que celles fondées sur les modèles à TD. L'identification de modèles à temps continu est en effet un problème d'une importance considérable qui trouve ses applications dans quasiment toutes les disciplines des sciences et de l'ingénierie.

Problème et méthodologie générale de l'identification des systèmes

Un système linaire à temps continu, invariant dans le temps est décrit par

$$y_t = G(p)u_t + \xi_t \tag{1}$$

où u représente le signal d'entrée, y le signal de sortie, G la fonction de transfert, p l'opérateur de différentiation et ξ_t regroupe les perturbations de toutes natures. Ces perturbations ξ_t regroupent les erreurs de mesure, les perturbations non mesurées, les erreurs de modèle ou encore l'association de ces derniers.

Les modèles peuvent être classés suivant trois grandes catégories :

- Modèle boîte grise, où le modèle est construit à temps continu sur la base des équations de la Physique ; les paramètres du modèle représentent alors des valeurs inconnues du système et ont une interprétation physique directe ;
- Modèle boîte noire, flexible, pouvant être utilisés dans tout type d'application. Ces modèles peuvent être à temps continu ou discret, et les paramètres n'ont pas nécessairement d'interprétation directe physique, même si les modèles à TC sont plus proches du modèle physique que les modèles à TD ;
- Modèle mécanistique issu sur des données. Ces modèles sont de type boîte noire mais ne sont considérés crédibles que s'ils ont une interprétation physique significative, valide.

Dans cette thèse, nous nous intéresserons aux modèles de type boîte noire. Le lecteur se reportera, par exemple, à [Boh06] (et les références inclues) pour les modèles de type boîte grise; ou à [YG06] (et les références inclues), pour le modèle mécanistique.

Les éléments fondamentaux de la procédure d'identification de système sont:

- le choix du signal d'excitation ;
- la détermination de la structure de modèle ;
- le critère d'identification ;
- la validation du modèle identifié.

En résumé, l'identification de système consiste donc à estimer un modèle dynamique, à temps continu ou discret, à partir de données mesurées. Cette procédure peut être résumée par la figure 1.

L'identification de modèles de type boîte noire, à temps continu, à partir de données mesurées, peut être réalisée suivant deux approches distinctes :

- l'approche indirecte, qui se décompose en deux étapes. Tout d'abord, un modèle à temps discret est obtenu en appliquant les méthodes usuelles d'identification de modèles à TD à partir de données échantillonnées; puis le modèle TD est transformé en un modèle à temps continu. Cette approche indirecte présente l'avantage d'utiliser les méthodes bien établies d'identification à TD [Lju99, SS89, You11]. Les méthodes les plus connues permettant d'obtenir des estimées convergentes sont les méthodes de minimisation de l'erreur de prédiction ainsi que les techniques de variable instrumentale;

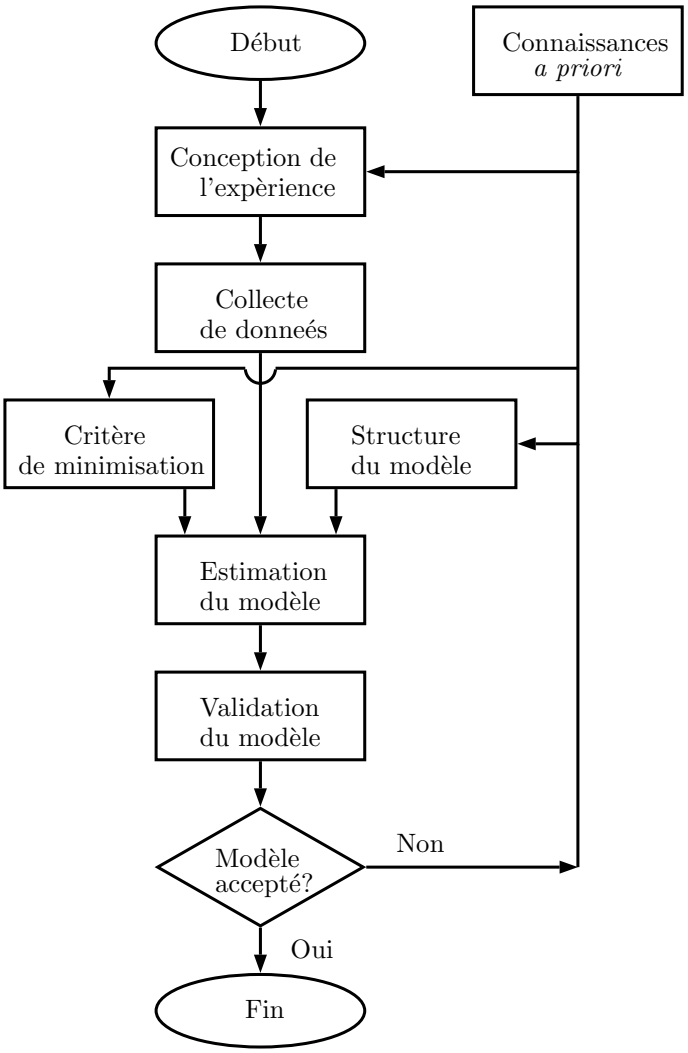

Figure 1: Procédure d'estimation (voir [Lju99]).

- l'approche directe, où un modèle à TC est directement estimé à partir des données mesurées ; ces méthodes font l'objet de cette thèse. Le grand intérêt de ces approches est que le modèle conserve sa forme originale à temps continu et qu'elles peuvent naturellement identifier les systèmes à partir de données échantillonnées à pas variable.

Données échantillonnées à pas variable

Dans un certain nombre de situations, l'identification du système doit être réalisée à partir de données échantillonnées à pas variable. Cette situation est souvent rencontrée dans les domaines biomédical, environnemental, dans le cas des systèmes mécaniques où un échantillonnage angulaire est réalisé ou lorsque les données transitent sur un

réseau. L'identification directe de modèles à temps continu est l'approche à privilégier ; les paramètres des modèles à temps discret étant dépendants

de la période d'échantillonnage, leur estimation devient plus complexe alors que les approches directes peut s'appliquer moyennant une adaptation au niveau de l'implantation des algorithmes pour prendre en compte le caractère irrégulier des données disponibles. Un exemple de données d'entrée/sortie échantillonnées à pas variable est représenté sur la Figure 1.2.

Quelques approches d'identification de modèles à temps continu ont été récemment développées [EG08, SCB92, LS02, LMS07, Mos08a, You11], et des résultats d'applications sont présentés dans [Eng07]. Les travaux présentés dans ce mémoire ont pour objectifs de proposer des solutions à des problèmes ouverts pour élargir le champ d'applications des méthodes disponibles pour identifier les systèmes à partir de données échantillonnées à pas variable.

Organisation du mémoire de thèse

Chapitre 2

L'échantillonnage d'un processus stochastique à temps continu est étudié dans le chapitre 2. Certains concepts de base concernant les processus stochastiques à temps continu (bruit blanc, modèle à temps discret équivalent) sont rappelés. Plusieurs exemples sont également décrits pour illustrer ces notions qui serviront dans les développements qui suivront.

Chapitre 3

Dans le Chapitre 3, la méthode espérance-maximisation EM est exploitée pour estimer les paramètres de modèles de bruit à temps continu CAR et CARMA à partir des données échantillonnées à pas variable. Cet algorithme, qui allie simplicité de mise en oeuvre et efficacité, a été utilisé avec succès pour estimer les paramètres de modèles à temps discret. La méthode EM proposée estime les paramètres d'un modèle CAR décrit à l'aide de l'opérateur δ et converge bien lorsque l'écart entre les instants d'échantillonnages sont faibles. Cependant, les estimées sont biaisées lorsque les intervalles d'échantillonnage sont plus grands. Une méthode d'interpolation est alors proposée et intégrée dans la version initiale de la méthode pour réduire le biais de l'estimateur. Les exemples de simulation montrent que la technique d'interpolation permet d'obtenir un compromis entre le taux de la convergence et la précision de l'estimation. L'algorithme EM ne peut pas estimer directement les paramètres de la partie MA du modèle CARMA. L'estimateur final du modèle CARMA combine un algorithme numérique de type gradient pour estimer la partie MA à l'algorithme EM pour la partie CAR.

Chapitre 4

Dans le chapitre 4, un estimateur de type variable instrumentale optimale pour identifier les paramètres de modèles de type Box-Jenkins à

temps continu est développé. Ce

dernier est itératif et exploite les techniques de régression linéaire. Il présente l'avantage de fournir des

estimées non biaisées lorsque le bruit de mesure est coloré et sa convergence est peu sensible au choix

du vecteur de paramètres initial. Une difficulté majeure dans le cas où les données sont

échantillonnées à pas variable concerne néanmoins l'estimation du modèle ARMA à temps continu (CARMA).

Les algorithmes présentés dans le chapitre 4 sont intégrés dans l'estimateur complet de variable instrumentale optimale pour les modèles Box-Jenkins à temps continu.

Une version étendue au cas de l'identification en boucle fermée est également développée.

Chapitre 5

Dans le chapitre 5, un estimateur robuste pour l'identification de systèmes à retard est proposé. Cette

classe de systèmes est très largement rencontrée en pratique et les méthodes disponibles ne peuvent pas traiter

le cas de données échantillonnées à pas variable. Le retard n'est pas contraint à être un multiple de la période

d'échantillonnage, contrairement à l'hypothèse traditionnelle dans le cas de modèles à temps discret. L'estimateur

développé est de type bootstrap et combine la méthode de variable instrumentale itérative pour les paramètres de

la fonction de transfert du système avec un algorithme numérique de type gradient pour estimer le retard. Un filtrage de type

passe-bas est introduit pour élargir la région de convergence pour l'estimation du retard. Les performances de l'estimateur proposé sont illustrées à l'aide de simulation

de Monte-Carlo.

Travaux publiés

Revues internationales

F. Chen, H. Garnier, and M. Gilson (2014). Robust identification of continuous-time models with arbitrary time-delay from irregularly sampled data. *Journal of Process Control*, 25(1):19–27, 2015.

F. Chen, J.C. Agüero, M. Gilson, and H. Garnier (2014). EM-based estimation of continuous-time AR models from irregularly sampled data. *Submitted to Automatica*.

Congrès internationaux avec comité de lecture et actes

F. Chen, H. Garnier, and M. Gilson (2013). Refined instrumental variable identification of continuous-time OE and BJ models from irregularly sampled data. *11th IFAC International Workshop on Adaptation and Learning in Control and Signal Processing*, Caen, France.

F. Chen, M. Gilson, H. Garnier, J. Agüero, and J. Schorsch (2014). Closed-loop identification of continuous-time systems from irregularly sampled data. *13th European Control Conference*, Strasbourg, France.

F. Chen, H. Garnier, M. Gilson, J. C. Agüero, and B. I. Godoy (2014). Identification of continuous-time transfer function models from irregularly sampled data in presence of colored noise. *19th IFAC World Congress*, Cape Town, South Africa.

F. Chen, H. Garnier, and M. Gilson (2014). Parameter and time-delay identification of continuous-time models from irregularly sampled data. *53rd IEEE Conference on Decision and Control*, Los Angeles, CA, USA.

Contents

List of Figures

List of Tables

List of abbreviations

ARX	AutoRegressive model with eXternal inputs
ARMA	AutoRegressive Moving Average
BJ	Box-Jenkins
CAR	Continuous-time AutoRegressive model
CARX	Continuous-time AutoRegressive model with eXternal inputs
CBJ	Continuous-time Box-Jenkins model
CLRIVC	Refined Instrumental Variable method for Closed-Loop Continuous-time systems
CLRIVC-EM	combined Refined Instrumental Variable and Expectation-Maximization method for Closed-Loop Continuous-time systems
CLSRIVC	Simplified Refined Instrumental Variable for Closed-Loop Continuous-time systems
COE	Continuous-time Output Error model
CT	Continuous-Time
DT	Discrete-Time
DBM	Data-Based Mechanism
EM	Expectation-Maximizaton
FOH	First-Order-Hold
Gard-EM	combined Gradient and Expectation-Maximization method
HBJ	Hybrid Box-Jenkins model
IV	Instrumental Variable
LPV	Linear parameter varying
LSSVF	State Variable Filter based Least-Squares method
MCS	Monte-Carlos Simulation
MISO	Multi-Input Single-Output
ML	Maximum-Likelihood
ODE	Ordinary Differential Equation
OE	Output Error
PEM	Prediction error minimization
PSD	Power Spectral Density
RIV	Refined Instrumental Variable
RIVC	Refined Instrumental Variable method for Continuous-time systems
RIVC-EM	combined Refined Instrumental Variable and Expectation-Maximization method
RTS	Rauch-Tung-Striebel
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio

SRIVC	Simplified Refined Instrumental Variable method for Continuous-time systems
SVF	State Variable Filter
TDSRIVC	Simplified Refined Instrumental Variable method for Continuous-time systems with Time-Delay
ZOH	Zero-Order-Hold

List of Symbols

$\boldsymbol{\eta}$	Vector of unknown parameters in the noise transfer function $H(p)$
$C_c(p)$	Controller
δ_τ	Dirac's delta function
δ_ℓ	Kronecker's delta function
e	Output error
$\mathscr{F}$	Fourier transform
$\mathscr{F}^{-1}$	Inverse Fourier transform
$F(p)$	State variable filter
$\boldsymbol{F}_c, \boldsymbol{G}_c, \boldsymbol{H}_c$	State-space matrices in continuous-time
$\boldsymbol{F}_{h_k}, \boldsymbol{G}_{h_k}, \boldsymbol{H}_{h_k}$	State-space matrices in discrete-time
$\boldsymbol{F}^\delta_{h_k}, \boldsymbol{G}^\delta_{h_k}, \boldsymbol{H}^\delta_{h_k}$	State-space matrices using δ-operator
$G(p)$	Plant part of a system
$H(p)$	Noise part of a system
h_k	k-the Sampling interval
$\bar{h}$	Upper bound of the sampling interval
$\underline{h}$	Lower bound of the sampling interval
$\boldsymbol{I}_n$	Identity matrix of dimension n
$\boldsymbol{J}$	Smoother gain
$\boldsymbol{K}$	Kalman gain
$l(\boldsymbol{\theta})$	Likelihood function
N	Number of observations
$\boldsymbol{P}$	Error covariance of the estimated states
p	Differential operator
P_e	Power of $e(t)$
$\boldsymbol{\phi}$	Regressor
$\boldsymbol{\psi}$	Instrumental variable
$\Phi_e(j\omega)$	Spectrum of $e(t)$
$\Phi^d_\xi(e^{j\omega h})$	Spectrum of the sampled sequence $\xi(kh)$
$\mathscr{Q}$	The auxiliary function used in the EM algorithm
$r_e(\tau)$	Autocorrelation function of $e(t)$
σ^2	Intensity of the CT white noise e_t
t_k	k-th sampling instants
τ	Time-delay
$\boldsymbol{\rho}$	Vector of the parameters in the plant transfer function $G(p)$
u	Input signal

$\boldsymbol{w}_t$ Process noise of the state-space model
v_t Measurement noise of the state-space model
x Noise-free response
ξ Colored measurement noise
y Output signal
$\boldsymbol{z}$ State variable

Chapter 1

Introduction

Mathematical models of dynamic systems are required in most areas of scientific enquiry and take various forms, such as differential equations, difference equations, state-space equations and transfer functions. The most widely used approach for mathematical modelling involves the construction of mathematical equations based on physical laws that are known to govern the behaviour of the system. Amongst the drawbacks of this approach are that the resulting models are often complex and not easily estimated directly from the available data because of the identifiability problems caused by over-parameterisation. This complexity also makes them difficult to use in applications such as control system design.

If sufficient experimental or operational data are available, an alternative to physically-based mathematical modelling is data-based 'system identification', which can be applied to virtually any system and typically yields relatively simple models that can well describe the system's behaviour within a defined operational regime. Such models can be either in a 'black-box' form, which describes only the input–output behaviour, or in some other, internally descriptive form, such as state-space equations, that can be interpreted in physically meaningful terms.

Dynamic systems in the physical world are naturally described in continuous-time (CT), differential equation terms because the physical laws, such as conservation equations, have been evolved mainly in this form. Paradoxically, however, the best known system identification schemes have been based on discrete-time (DT) models (sometimes referred to as sampled-data models), without much concern for the merits of natural continuous-time model descriptions and their associated identification methods. In fact, the development of CT system identification techniques occurred in the the last century, before the development of the DT techniques, but was overshadowed by the more extensive DT developments. This was mainly due to the 'go completely discrete-time' trend that was spurred by parallel developments in digital computers.

Much less attention has been devoted to CT modelling from DT data and many practitioners appear unaware that such alternative methods not only exist but may be better suited to their modelling problems. The identification of continuous-time models is indeed a problem of considerable importance that has applications in virtually all disciplines of science and engineering.

This chapter is organised as follows: in the first section, the general procedure for system identification is reviewed; thereafter, the motivation for using irregularly sampled data is presented

in Section 1.2; subsequently, basic solutions to CT model identification are presented in Section 1.3; after that, in Section 1.4, we present the special topics in system identification that are considered in this thesis; finally, Section 1.5 gives the outline of this thesis and Section 1.6 lists the related publications.

1.1 System Identification Problem and Procedure

A linear time-invariant continuous-time system with input u and output y can always be described by

$$y_t = G(p)u_t + \xi_t \tag{1.1}$$

where G is the transfer function, p the time-domain differential operator and the additive term ξ_t represents errors and disturbances of all natures. The source of ξ_t could be measurement errors, unmeasured process disturbances, model inadequacy, or combinations of these. It is assumed that the input signal $\{u_t, t_1 < t < t_N\}$ is applied to the system, with u_t and the output y_t both sampled at discrete times $t_1, \cdots, t_N$. The sampled signals are denoted by $\{u_{t_k}, y_{t_k}\}$. The identification problem can be stated as follows: determine a continuous-time model for the original CT system from N sampled measurements of the input and output $Z^N = \{u_{t_k}, y_{t_k}\}_{k=1}^N$.

There are three different kinds of parameterised models:

- grey-box models, where the model is constructed in continuous-time from basic physical principles and the parameters represent unknown values of the system coefficients that, at least in principle, have a direct physical interpretation. Such models are also known as physically parameterised or tailor-made models;
- black-box models, which are families of flexible models of general applicability. The parameters in such models, which can be continuous time or discrete time, have no direct physical interpretation (even though the CT version is closer to the physically parameterised model than the DT version), but are used as vehicles to describe the properties of the input– output relationships of the system. Such models are also known as readymade models;
- data-based mechanistic (DBM) models, which are effectively models identified initially in a black-box, generic model form but only considered credible if they can be interpreted in physically meaningful terms. In this thesis, we restrict our attention to black-box model identification. The reader is referred, for instance, to [Boh06] and the references therein, for grey-box model identification; and [YG06] and the references therein, for DBM model identification.

The basic ingredients for the system identification problem are as follows

- the data set;
- a model description class (the model structure);
- a criterion of fit between data and models;
- a way to evaluate the resulting models.

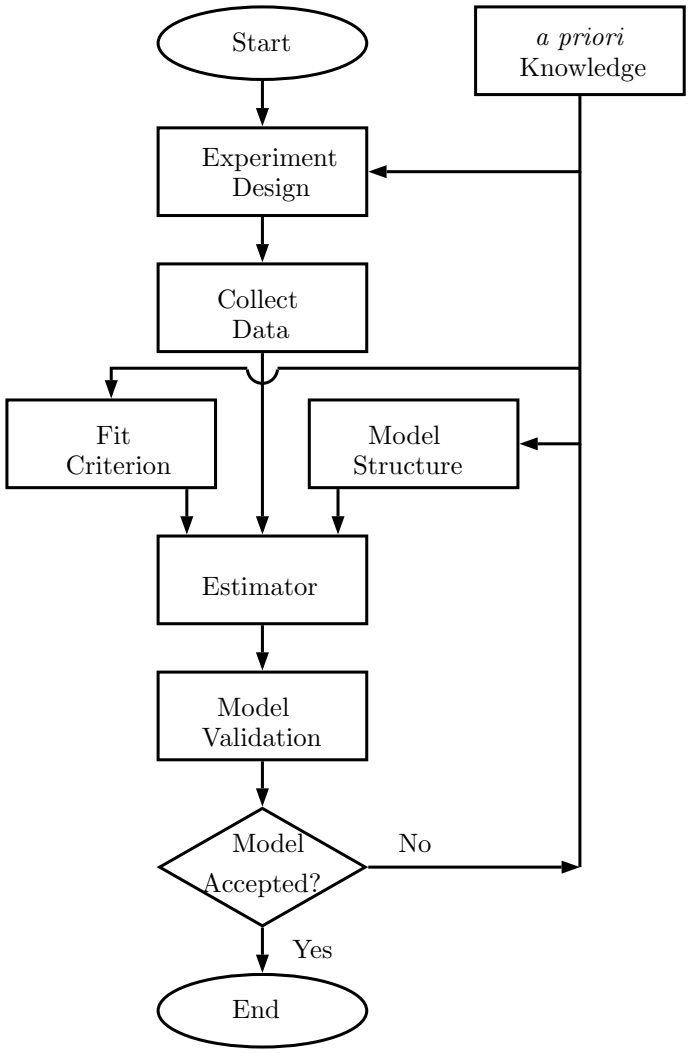

Figure 1.1: A flow chart of the identification procedure (see [Lju99]).

System identification deals with the problem of determining mathematical models of dynamical, continuous-time systems using measured input–output data. Basically this means that a set of candidate models is chosen and then a criterion of fit between model and data is developed. Finally, the model that best describes the data according to the criterion, within the model set, is computed using some suitable algorithm. The flow chart of the estimation procedure is presented in Figure 1.1.

There are two fundamentally different time-domain approaches to the problem of obtaining a black-box CT model of a natural CT system from its sampled input–output data:

- the indirect approach, which involves two steps. First, a DT model for the original CT system is obtained by applying DT model estimation methods to the available sampled data; and then the DT model is transformed into the required CT form. This indirect approach has the advantage that it uses well-established DT model identification methods [Lju99, SS89, You11]. Examples of such methods, which are known to give consistent and statistically efficient estimates under very general conditions, are prediction error methods and optimal instrumental variable techniques;

- the direct approach, where a CT model is obtained immediately using CT model identification methods, such as those discussed in this thesis. Without relying any longer on analogue computers, the present techniques exploit the power of the digital tools. In this direct approach, the model remains in its original CT form.

Independent of how the identification problem is approached, a model parametrization will lead to the definition of a predictor

$$\hat{x}_{t_k}(\boldsymbol{\rho}) = g(\boldsymbol{\rho}, Z^{k-1}) \tag{1.2}$$

that depends on the unknown parameter vector $\boldsymbol{\rho}$, and past data Z^{k-1}. The general procedure for estimating a parameterized model from sampled data, regardless of whether it is a CT or DT model, is as follows:

1. from observed data and the predictor $\hat{x}_{t_k}(\boldsymbol{\rho})$, form the sequence of prediction errors

$$\varepsilon_{t_k}(\boldsymbol{\rho}) = y_{t_k} - \hat{x}_{t_k}(\boldsymbol{\rho}) \quad k = 1, \cdots, N \tag{1.3}$$

2. filter the prediction errors through a linear filter $F(\bullet)$ to enhance or attenuate interesting or unimportant frequency bands in the signals

$$\varepsilon_{F,t_k}(\boldsymbol{\rho}) = F(\bullet)\varepsilon_{t_k}(\boldsymbol{\rho}) \tag{1.4}$$

   where $\bullet$ can be the shift operator if the filter is in discrete time or the differential operator when the filter is in continuous time;

3. choose a scalar-valued, positive function $l(\cdot)$ to measure the size or norm of the filtered prediction error

$$l\left(\varepsilon_{F,t_k}(\boldsymbol{\rho})\right) \tag{1.5}$$

4. minimise the sum of these norms

$$\hat{\boldsymbol{\rho}} = \arg \min_{\boldsymbol{\rho}} J_N(\boldsymbol{\rho}) \tag{1.6}$$

   where

$$J_N(\boldsymbol{\rho}) = \frac{1}{N} \sum_{k=1}^{N} l\left(\varepsilon_{F,t_k}(\boldsymbol{\rho})\right) \tag{1.7}$$

This procedure is general and pragmatic, in the sense that it is independent of the particular CT or DT model parametrization being used, although this parametrization will affect the minimization procedure.

1.2 Irregularly Sampled Data

In some situations, it is difficult to obtain equidistant sampled data. This problem arises in medicine, environmental science, transport and traffic systems, astrophysics and other areas,

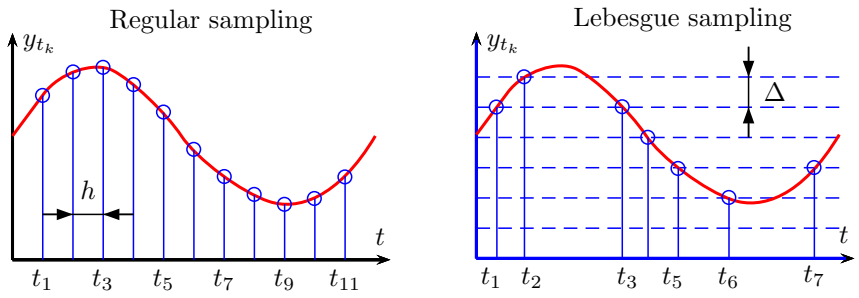

Figure 1.2: Regular sampling and Lebesgue sampling.

where measurement is not under the control of the experimenter or where uniform sampling is practically impossible. For these irregularly sampled data systems, the standard DT linear, time-invariant models will not be applicable because the assumption of a uniformly sampled environment, as required for the existence of such discrete-time models, is violated. On the other hand, the coefficients of CT models are assumed to be independent of the sampling period and so they have a built-in capability to cope with the irregularly sampled data situation. With a small modification of the data handling procedure, the measurements are considered as points on a continuous line, which do not need to be equidistantly spaced.

Here two applications of irregular sampling are listed in the follows.

- Lebesgue sampling (see *e.g.* [ÅB03]): this sampling scheme is illustrated in Figure 1.2, where a sequence of regularly sampled data is plotted int the left figure. The Lebesgue sampling uses another mechanism to generate the measurements, it firstly monitors the CT output signal, when a certain amount (Δ) of the output signal is detected, the sampler generates an observation. The advantage of this sampling scheme is that, in some cases, it produces less number of observations and saves the storage space.

- Networked control system (see *e.g.* [WZC09]): the block diagram of a networked control system is given in Figure 1.3 and the sampled data is plotted in Figure 1.4. The data exchange between the computer and the process is carried out by the network. Assuming that the CT output of the plant is regularly sampled and the measurement is denoted by y_{t_k}, the computer receives the data at time-instant t_i. Because of the random delay and data losses of the network, even though the original signal is regularly sampled, the received data, which denoted by $y^c_{t_i}$, at the computer side is irregularly distributed (see Figure 1.4). As soon as the computer receives the measurement, it generates the control signal $u^c_{t_i}$ and transmits it to the actuator, the received data is denoted by u_{t_j}. It can be observed from Figure 1.4 that the sampled input-output signals are non-synchronized and irregularly sampled.

There are many papers, where the problem of identification from irregularly sampled data has been investigated, see *e.g.* [EG08, SCB92, LS02, LMS07, Mos08b, YAAG11], and some applications are presented in [Eng07]. Since this thesis aims at developing some identification algorithms that can handle irregular sampled data, other issues, such as how the irregular data is generated, are omitted in this thesis. A few basic identification schemes which can handle irregularly sampled data are briefly reviewed in the next section.

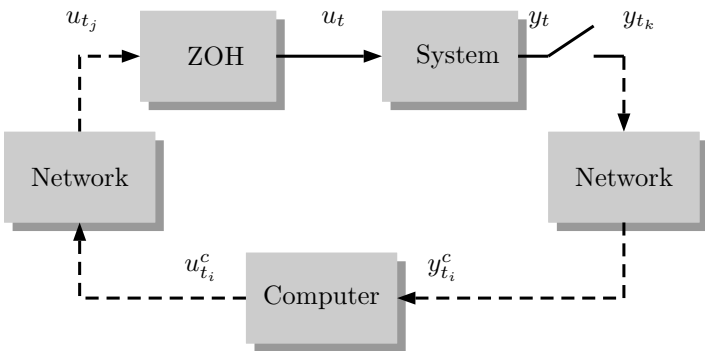

Figure 1.3: The block diagram of a networked control system.

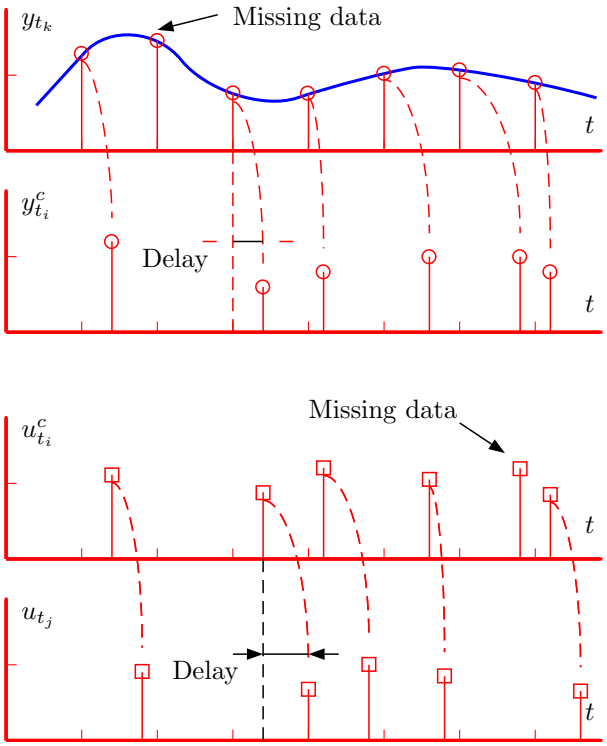

Figure 1.4: The sampled data for a networked control system.

1.3 Direct Continuous-time Model Identification

The objective of this thesis is to estimate linear single-input single-output (SISO) transfer function models, thus we assume the following dynamical system

$$x_t = G(p)u_t = \frac{B(p)}{A(p)}u_t \tag{1.8}$$

where u_t and x_t denote the excitation signal and the noise-free response at time-instant t, respectively, p denotes the differential operator, *i.e.* $p = \frac{d}{dt}$. $B(p)$ and $A(p)$ of degrees of n_b and n_a ($n_b \leq n_a$) are assumed to be coprime. Moreover, $A(p)$ is assumed to have roots located in the left half plane.

$$B(p) = b_0 p^{n_b} + b_1 p^{n_b-1} + \cdots + b_{n_b}$$

$$A(p) = p^{n_a} + a_1 p + \cdots + a_{n_a}$$

The unknown parameters are stacked column-wise in

$$\boldsymbol{\rho} = \begin{bmatrix} a_1 & \cdots & a_{n_a} & b_0 & \cdots & b_{n_b} \end{bmatrix}^T \tag{1.9}$$

The system is assumed to be sampled at time-instant t_k, for $k = 1, 2, \cdots, N$. The time-varying sampling interval is denoted as

$$h_k = t_{k+1} - t_k, \quad k = 1, 2, \cdots, N-1 \tag{1.10}$$

The noise free signal x_t is always contaminated by an additive noise ξ_t in practical situations, thus the measurement equation is then given by

$$y_{t_k} = x_{t_k} + \xi_{t_k} \tag{1.11}$$

Assuming that the model structure and the degrees n_b and n_a are known, the identification objective is to estimate the unknown parameters in $\boldsymbol{\rho}$ from irregularly sampled data $Z^N = \{u_{t_k}, y_{t_k}\}_{k=1}^N$.

1.3.1 The Derivative Approximation Method for CARX Model Estimation

The derivative approximation method for CT autoregressive with exogenous input (CARX) model identification was presented in *e.g.* [TFCB97, LMS07]. The basic idea of this method is to estimate the unknown parameters by the least-squares method, where the involved time-derivatives are approximated by finite difference.

Assuming that the measurement noise ξ_t is given as

$$\xi_t = \frac{1}{A(p)} e_t \tag{1.12}$$

where e_t is a CT white noise. Then, the model can be rewritten into the following equivalent CARX form

$$A(p) y_t = B(p) u_t + e_t \tag{1.13}$$

Since the unmeasured time-derivatives of the input-output signals are needed in parameter estimation, the following approach was used (see *e.g.* [LMS07]) to approximate them from the irregularly sampled data

$$p^i y_{t_k} \approx D^i y_{t_k} \tag{1.14}$$

where D is the difference operator. Given approximated time-derivatives, a regression model can be obtained

$$D^{n_a} y_{t_k} = \boldsymbol{\phi}_{t_k}^T \boldsymbol{\rho} + \varepsilon_{t_k} \tag{1.15}$$

where ε_{t_k} is the equation error and

$$\boldsymbol{\phi}_{t_k}^T = \begin{bmatrix} -D^{n_a-1} y_{t_k} & \cdots & -y_{t_k} & D^{n_b} u_{t_k} & \cdots & u_{t_k} \end{bmatrix}$$

The least-squares (LS) estimate of $\boldsymbol{\rho}$ is given by

$$\hat{\boldsymbol{\rho}} = \left(\sum_{k=1}^N \boldsymbol{\phi}_{t_k} \boldsymbol{\phi}_{t_k}^T \right)^{-1} \sum_{k=1}^N \boldsymbol{\phi}_{t_k} D^n y_{t_k} \tag{1.16}$$

1.3.2 SVF-based Methods for CT Model Estimation

The use of a state variable filter (SVF) in CT system identification was suggested in *e.g.* [You11, TB95]. This SVF-based method has proven popular since it provides a very convenient way to generate the prefiltered derivatives from the sampled input and output data.

Consider a SVF $F(p)$ taking the following structure

$$F(p) = \frac{1}{(p + \omega_c^{\text{SVF}})^{n_a}} \tag{1.17}$$

where ω_c^{SVF} is the breakpoint frequency. By applying $F(p)$ on both sides of equation (1.13), we can obtain a linear regression model built from the filtered signals and time-derivatives

$$p^{n_a} y_{F,t_k} = \boldsymbol{\phi}_{F,t_k}^T \boldsymbol{\rho} + \varepsilon_{F,t_k} \tag{1.18}$$

where $[\cdot]_F$ denotes the filtering operation $[\cdot]_F = F(p)[\cdot]$ (this issue will be detailed Chapter 4) and

$$\boldsymbol{\phi}_{F,t_k}^T = \begin{bmatrix} -p^{n_a-1} y_{F,t_k} & \cdots & -y_{F,t_k} & p^{n_b} u_{F,t_k} & \cdots & u_{F,t_k} \end{bmatrix}$$

Subsequently, the SVF-based least-squares (LSSVF) estimator is given by

$$\hat{\boldsymbol{\rho}}_{ls} = \left(\sum_{k=1}^{N} \boldsymbol{\phi}_{F,t_k} \boldsymbol{\phi}_{F,t_k}^T \right)^{-1} \sum_{k=1}^{N} \boldsymbol{\phi}_{F,t_k} p^{n_a} y_{F,t_k} \tag{1.19}$$

Since ε_{F,t_k} is always colored, the LS estimator (1.19) is biased. Several methods, such as the prediction error minimization (PEM) and instrumental variable (IV) methods, have been proposed to remove the bias. Amongst these methods, this thesis focuses more specially on the IV techniques which present the great advantage of being able to consistently identify CT models while relying on simple linear (regression-like) algorithms. The IV estimate can be obtained by solving the following equation

$$\hat{\boldsymbol{\rho}}_{iv} = \text{sol} \left\{ \sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k} \left(y_{F,t_k} - \boldsymbol{\phi}_{F,t_k}^T \boldsymbol{\rho} \right) = 0 \right\} \tag{1.20}$$

where $\boldsymbol{\psi}_{t_k}$ is called the instrumental variable and $\boldsymbol{\psi}_{F,t_k}$ is its filtered version using $F(p)$. One intuitive choice of $\boldsymbol{\psi}_{t_k}$ can be

$$\mathring{\boldsymbol{\psi}}_{t_k}^T = \begin{bmatrix} -p^{n_a-1} x_{t_k} & \cdots & -x_{t_k} & p^{n_b} u_{t_k} & \cdots & u_{t_k} \end{bmatrix} \tag{1.21}$$

where x_{t_k} is the noise-free response of the true system. However, this value cannot be obtained since the noise-free signal x_{t_k} is not measured. A practical way is to approximate it by using the estimated signal $\hat{x}_{t_k}$

$$x_{t_k}(\hat{\boldsymbol{\rho}}_{ls}) = G(p, \hat{\boldsymbol{\rho}}_{ls}) u_{t_k} \tag{1.22}$$

where $G(p, \hat{\boldsymbol{\rho}}_{ls})$ represents the auxiliary model built from the LSSVF estimate. Then, we obtain the SVF-based instrumental variable estimator (IVSVF) given by

$$\hat{\boldsymbol{\rho}}_{iv} = \left(\sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k}(\hat{\boldsymbol{\rho}}_{ls}) \boldsymbol{\phi}_{F,t_k}^T \right)^{-1} \sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k}(\hat{\boldsymbol{\rho}}_{ls}) p^{n_a} y_{F,t_k} \tag{1.23}$$

where

$$\boldsymbol{\psi}_{t_k}^T(\hat{\boldsymbol{\rho}}_{ls}) = \begin{bmatrix} -p^{n_a-1} x_{t_k}(\hat{\boldsymbol{\rho}}_{ls}) & \cdots & -x_{t_k}(\hat{\boldsymbol{\rho}}_{ls}) & p^{n_b} u_{t_k} & \cdots & u_{t_k} \end{bmatrix} \tag{1.24}$$

1.3.3 The Optimal IV Method for COE Model Estimation

Let us consider the following CT output-error (COE) model

$$y_{t_k} = \frac{B(p)}{A(p)} u_{t_k} + e_{t_k} \tag{1.25}$$

where e_{t_k} is a DT white measurement noise. It has been shown that the unbiased and minimum variance conditions for the IV estimator are (see *e.g.* [YJ80, SS83])

$$\begin{cases} F^{\text{opt}}(p) = \dfrac{1}{A(p)} \\ \boldsymbol{\psi}_{F,t_k}^{\text{opt}} = F^{\text{opt}}(p)\mathring{\boldsymbol{\psi}}_{t_k} \end{cases} \tag{1.26}$$

The following comments can be made from the IV optimal conditions:

1. Filtering operation is a distinguishing feature of optimal IV solution.
2. The optimal IV estimate is achieved when the instrument variable is equal to the filtered noise-free version of the regression vector.
3. One particularly successful implementation of the optimal IV method is known as the simplified refined IV method for CT system identification (simplified RIVC, SRIVC), where an iterative procedure is used, in which, at each iteration, the auxiliary model is used to generate the instrumental variables and prefilter based on the parameters obtained at the previous iteration.

The SRIVC algorithm was first presented in [YJ80] and can be summarized as follows

Algorithm 1.1 (The SRIVC algorithm)**.**

---

1. *Initialization. Specify a ω_c^{SVF} and use the LSSVF method to compute an initial estimate $\hat{\boldsymbol{\rho}}^0$*
2. *Iteration.*
   for j=0:convergence
   (a) *Generate the noise-free response $\hat{x}_{t_k}(\hat{\boldsymbol{\rho}}^j)$ using the estimated model built from $\hat{\boldsymbol{\rho}}^j$.*

$$\hat{x}_{t_k}\left(\hat{\boldsymbol{\rho}}^j\right) = \frac{B\left(p, \hat{\boldsymbol{\rho}}^j\right)}{A\left(p, \hat{\boldsymbol{\rho}}^j\right)} u_{t_k}$$

   (b) *Update the prefilter*

$$F\left(p, \hat{\boldsymbol{\rho}}^j\right) = \frac{1}{A\left(p, \hat{\boldsymbol{\rho}}^j\right)}$$

   and compute the filtered signals in $\hat{\boldsymbol{\psi}}_{F,t_k}\left(\hat{\boldsymbol{\rho}}^j\right)$, $\boldsymbol{\phi}_{F,t_k}\left(\hat{\boldsymbol{\rho}}^j\right)$, and $y_{F,t_k}\left(\hat{\boldsymbol{\rho}}^j\right)$.

   (c) *Compute a new estimate*

$$\hat{\boldsymbol{\rho}}^{j+1} = \left[\boldsymbol{\psi}_{F,t_k}\left(\hat{\boldsymbol{\rho}}^j\right)\boldsymbol{\phi}^T_{F,t_k}\left(\hat{\boldsymbol{\rho}}^j\right)\right]^{-1}\boldsymbol{\psi}_{F,t_k}\left(\hat{\boldsymbol{\rho}}^j\right)y_{F,t_k}\left(\hat{\boldsymbol{\rho}}^j\right)$$

end

The following example is used to illustrate the use of the SRIVC method.

Example 1.1. *Consider the a second order COE system*

$$\begin{cases} x_t = \dfrac{b_0^o}{p^2 + a_1^o p + a_2^o}u_t \\ y_{t_k} = x_{t_k} + e_{t_k} \end{cases} \tag{1.27}$$

where $a_1^o = 2.8, a_2^o = 4$ *and* $b_0^o = 5$, e_{t_k} *is a DT white measurement noise. The simulation is carried out under the following assumptions*

1. *The input signal* u_t *takes the form of a multisine signal*

$$u_{t_k} = sin(0.714t_k) + sin(1.428t_k) + sin(2.142t_k) \tag{1.28}$$

We assume that the intersample behavior of u_{t_k} *is zero-order-hold (ZOH).*

2. *The sampling interval* h_k *is uniformly distributed over the following interval*

$$h_k \sim U[0.01, \bar{h}] \tag{1.29}$$

where $\bar{h}$ *increases from* $0.01s$ *to* $1s$.

The choice of the sampling interval is usually coupled to the time constant of the system. A rule of thumb for DT model identification is to choose the uniformly sampling frequency about 10 times the bandwidth of the system. This corresponds about to placing $4 - 10$ *sampling points over the rise time of the step response. It may be observed that the system has a rise time of about* $2s$. *A basic choice for the uniform sampling period would be between* $0.2s < T_s < 0.5s$. *Since* h_k *is uniformly distributed,* $\bar{h}$ *will be twice this value.*

3. *The Signal-to-Noise Ratio (SNR) is set to* 10 *dB*

$$\textit{SNR} = 10\ \ log_{10}\left(\frac{P_x}{P_e}\right)$$

where P_x *denotes the averaged power of* x_t.

4. *The number of observations is* $N = 1000$. *A Monte Carlo simulation (MCS) of* 200 *realizations are performed, the mean and standard deviation are computed from these realizations. The stopping rule of the iterative algorithm is*

$$max\left\{\left|\frac{\hat{\boldsymbol{\rho}}^{i+1} - \hat{\boldsymbol{\rho}}^i}{\hat{\boldsymbol{\rho}}^i}\right|\right\} \leq 10^{-4} \tag{1.30}$$

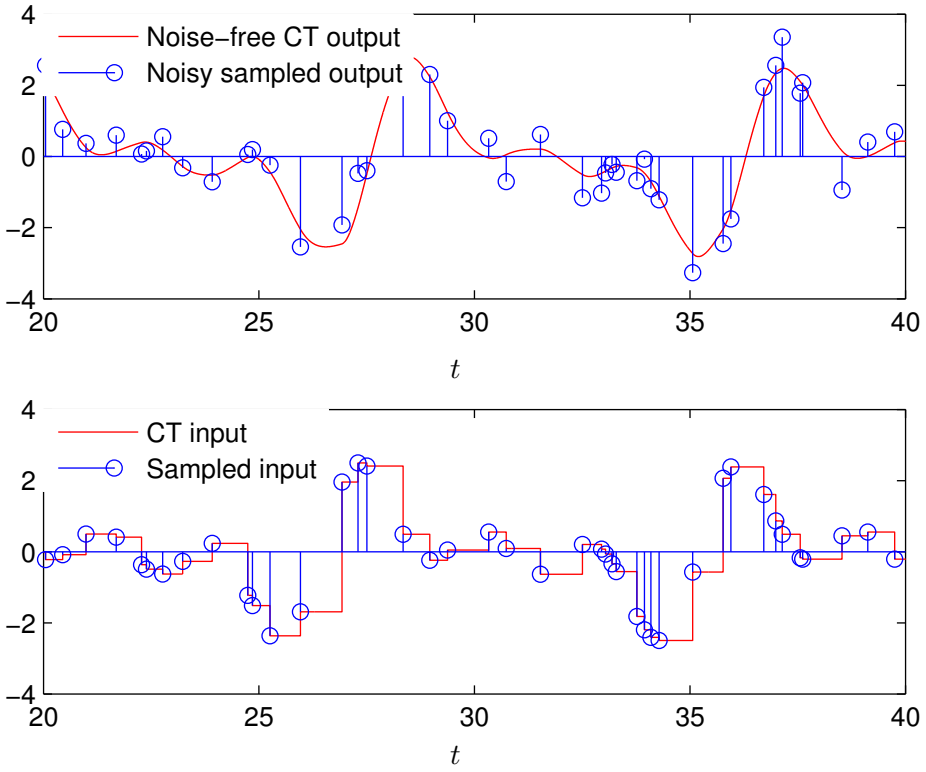

Figure 1.5: A portion of irregularly sampled data ($\bar{h} = 1s$).

A portion of the observed input-output signals is plotted in Figure 1.5, where the maximum sampling interval $\bar{h}$ is set to $1s$. The SRIVC parameter estimates are presented in Figure 1.6. It shows that the results remain very accurate for all $\bar{h}$ over a large considered range. It may be noticed that the mean and standard deviation of the estimates are almost identical over the range of $\bar{h}$.

Table 1.1 lists the averaged number of iterations for convergence. It shows that the SRIVC converges quite quickly for the different values of $\bar{h}$.

$\bar{h}$	0.01	0.05	0.1	0.2	0.4	0.6	0.8	1
N_{iter}	3.7	3.8	4.0	4.2	4.6	5.4	6.4	7.4

Table 1.1: The averaged number of iterations for convergence.

The SRIVC algorithm is therefore a very useful and computationally efficient estimator in this non-uniformly sampling situation where the additive noise is purely white. This algorithm is optimal under these conditions and it functions even if the sampling period varies over a quite large range.

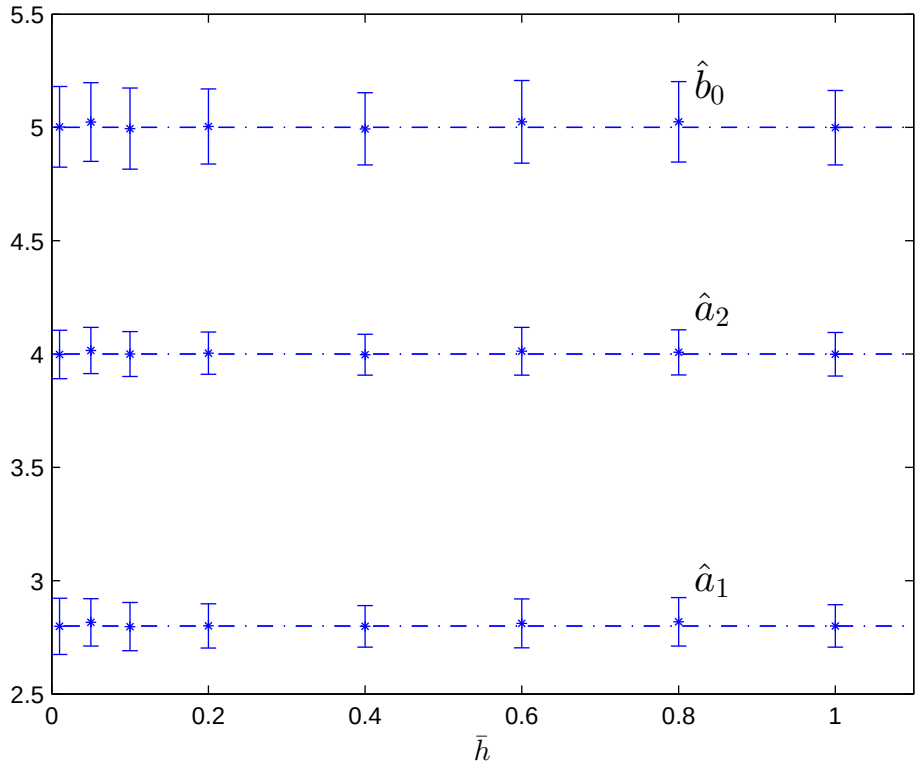

Figure 1.6: Parameter estimates by using the SRIVC method. $(*)$ - the mean value. Vertical bars - the standard deviation.

1.4 Main Contributions of This Thesis

This SRIVC algorithm is known to be one of the most reliable algorithm in practice (see *e.g.* [GY14]). It provides optimal estimates when the time-delay is *a priori* known and the additive noise is white. However, if the noise is colored with a rational power spectral density (PSD) function, the estimates still remain consistent, because of the iterative IV mechanism, but they do not have minimum variance. To account for the colored noise process, the RIVC algorithm for hybrid Box-Jenkins (HBJ) model estimation from regularly sampled data has been recently developed. Here the model of the dynamics is estimated in CT form while the additive noise model is estimated as a DT autoregressive moving average (ARMA) model (see [YGG08]). However, when the data are irregularly sampled, this RIVC algorithm cannot be used anymore and the RIVC algorithm requires to use a CT model for the additive noise process. These facts motivate the following contributions of this thesis

1. CT autoregressive (AR)/ARMA model estimation from irregularly sampled data.
2. CT Box-Jenkins model estimation from irregularly sampled data, where the noise is represented by a CT AR/ARMA model.
3. CT OE model estimation with a unknown time-delay from irregularly sampled data. The difficulty lies in this situation is that the time-delay is a non-linear parameter in the model, it cannot be estimated in regression form along with other linear parameters.

1.4.1 CT Box-Jenkins Model Estimation

Assuming that a colored noise process ξ_t is given by

$$\xi_t = H(p) = \frac{C(p)}{D(p)} e_t \tag{1.31}$$

where e_t is a CT white noise. Then, the CBJ model can be given as

$$\begin{cases} x_t = G(p)u_t = \dfrac{B(p)}{A(p)} u_t \\ \xi_t = H(p)e_t = \dfrac{C(p)}{D(p)} e_t \\ y_t = x_t + \xi_t \end{cases} \tag{1.32}$$

$D(p)$ and $C(p)$ are polynomials of degrees n_d and n_c, respectively

$$\begin{aligned} C(p) &= p^{n_c} + c_1 p^{n_c - 1} + \cdots + c_{n_c} \\ D(p) &= p^{n_d} + d_1 p^{n_d - 1} + \cdots + d_{n_d} \quad (n_d \geq n_c) \end{aligned}$$

Then unknown parameters are stacked in a column vector

$$\boldsymbol{\eta} = \begin{bmatrix} d_1 & \cdots & d_{n_d} & c_1 & \cdots & c_{n_c} \end{bmatrix}^T$$

The RIVC method (see [YGG08]) can be used to estimate a CBJ model with a slight modification. We propose to estimate the plant model by IV method and the noise model by ML method. This modified RIVC algorithm will be discussed in detail in Chapter 4.

1.4.2 COE System Identification with Unknown Time-delay

A COE model with time-delay can be as follows

$$y_t = G(p)u_{t-\tau} + e_t = \frac{B(p)}{A(p)} u_t(\tau) + e_t \tag{1.33}$$

The existence of a unknown time-delay greatly complicates system identification, since it is a non-linear parameter in the model and cannot be easily estimated along with the other linear parameters in the same step. Time-delay estimation has been extensively studied in the literature, see *e.g.* [Bjö03, AHS06b, LG08, YG06, NXS10]. In this thesis, a new method is proposed to estimate CT plant parameters and time-delay from irregularly sampled data. The proposed method uses the SRIVC method to estimate the linear plant parameters and an adaptive numerical method to estimate the time-delay. In order to improve the performance of the numerical search, a low-pass filtering technique is introduced to widen the convergence region for time-delay estimation.

1.5 Outline of This Thesis

Chapter 2

Since the assumed colored additive noise is a stochastic process, sampling of a CT stochastic system is firstly recalled in Chapter 2. Although this topic is not new and has been extensively studied in the last two decades, for the sake of clarity, some basic concepts and results (such as the concept of CT white noise, CT stochastic process, and the related sampling issues) are presented here. Several examples are also included to give the reader an illustrative understanding.

Chapter 3

In chapter 3, the expectation-maximization-based (EM) method is developed to estimate CT AR/ARMA models from irregularly sampled data. The EM algorithm has been successfully used in DT model estimation for several years and shows attractive convergence properties. The proposed EM-based method uses a δ-operator model to approximate a CT model, and it works quite well when the sampling intervals are small. However, the approximation error is non-negligible when the sampling interval becomes larger, therefore a state interpolation technique is then applied to reduce the bias of the parameter estimates. The simulated examples show that the interpolation technique provides a trade-off between the convergence rate and the estimation accuracy. Since the EM algorithm cannot be used to estimate MA parameters, an additional numerical search of the MA parameters is combined to the EM algorithm. This leads to the Grad-EM algorithm for CT ARMA model estimation.

Chapter 4

In chapter 4, the problem of CT model identification from non-uniformly sampled input-output data in presence of colored noise is considered. A CT Box-Jenkins model structure is used to describe the system, which provides independent parameterizations for the plant and noise. The CT Box-Jenkins model is assumed to be estimated in a separable way. More precisely, the plant is estimated by the IV technique, while the noise is estimated by the EM algorithm (see Chapter 3). In estimating each of them, the other is assumed to be fixed. Both the open-loop and closed-loop situations are considered. The performance of the proposed methods is evaluated by numerical examples.

Chapter 5

This chapter presents a new approach to estimate CT OE models with arbitrary time-delay from irregularly sampled input-output data. It is based on the separable nonlinear least-squares method which combines in a bootstrap manner the iterative optimal IV method for transfer function model estimation with an adaptive gradient-based technique that searches for the optimal time-delay. Since the objective function may have several local minima, the initialization requires special attention. Here, a low-pass filtering strategy is used to widen the convergence

region around the global minimum. Simulation results are included to illustrate the performance of the proposed method.

1.6 Publications

This thesis is based on the following papers.

Journals

F. Chen, H. Garnier, and M. Gilson (2014). Robust identification of continuous-time models with arbitrary time-delay from irregularly sampled data. *Journal of Process Control*, 25(1):19–27, 2015.

F. Chen, J.C. Agüero, M. Gilson, and H. Garnier (2014). EM-based estimation of continuous-time AR models from irregularly sampled data. *Submitted to a Journal*.

Conferences

F. Chen, H. Garnier, and M. Gilson (2013). Refined instrumental variable identification of continuous-time OE and BJ models from irregularly sampled data. *11th IFAC International Workshop on Adaptation and Learning in Control and Signal Processing*, Caen, France.

F. Chen, M. Gilson, H. Garnier, J. Agüero, and J. Schorsch (2014). Closed-loop identification of continuous-time systems from irregularly sampled data. *13th European Control Conference*, Strasbourg, France.

F. Chen, H. Garnier, M. Gilson, J. C. Agüero, and B. I. Godoy (2014). Identification of continuous-time transfer function models from irregularly sampled data in presence of colored noise. *19th IFAC World Congress*, Cape Town, South Africa.

F. Chen, H. Garnier, and M. Gilson (2014). Parameter and time-delay identification of continuous-time models from irregularly sampled data. *53rd IEEE Conference on Decision and Control*, Los Angeles, CA, USA.

Chapter 2

Sampling of Continuous-time Stochastic Systems

This chapter reviews the known results of CT stochastic system sampling, where the word 'stochastic' means that this process cannot be described by an explicit mathematical function. The theory of stochastic processes is quite involved, the newcomer to this field should observe that no one can be competent in a field as complicated as the theory of stochastic processes with a single reading of a short chapter of this type, good references are *e.g.* [Åst70, BH97, Jaz70]. For deterministic systems, sampling means that the obtained DT systems exactly describe the relationship between u_t and y_t at the sampling instant t_k. However, the sampling of a stochastic system is not defined in this sense, because only the output of a stochastic process can be measured (the input is a white noise thus cannot be measured). In this case, sampling means that the second order property (such as power spectral density function, or spectrum) of the DT system output keeps the same as the CT one at the sampling instant t_k.

Many papers assume that CT systems are sampled at regular frequency. Then, the equivalent DT systems have a closed form, see *e.g.* [Söd02, WLS93, Åst70, FG96]. However, when data are collected unevenly in time, the variation of the sampling interval gives rise to a time-varying DT system. Amongst the available model structures, we recommend the use of time-varying DT state-space models (or sampled data models). Compared with other choices, such as the ARMA model used in [Söd90], a time-varying DT state-space model is particularly useful in the irregular sampling framework, since it explicitly includes the sampling interval h_k. The DT matrices are functions uniquely dependent on h_k when the underlying CT system is invariant. By contrast, if an ARMA model is adopted, the zero positions of the ARMA model will be very hard to determine (see [Söd90]). The sampled data model will be extensively used in Chapter 3 for state estimation and CT system identification.

This chapter is organized as follows: the concept of CT white noise is reviewed in Section 2.1; then, the CT system to be sampled is given in Section 2.2; subsequently, three sampling schemes, namely the instantaneous, integrated and simplistic sampling schemes are introduced in Section 2.3, 2.4 and 2.5, respectively; finally we draw some conclusions in Section 2.6 and list the relevant Matlab routines in Section 2.7.

2.1 The Concept of Continuous-time White Noise

A CT white noise is a special kind of stochastic process. An important descriptor for stationary stochastic processes is the *power spectral density* (PSD) function, or *spectrum*, which can be computed by using the well-known *Wiener–Khinchin* relation

$$\Phi_x(j\omega) = \mathscr{F}\left\{r_x(\tau)\right\} = \int_{-\infty}^{\infty} r_x(\tau)e^{-j\omega\tau}d\tau \tag{2.1}$$

where $\mathscr{F}\left\{\cdot\right\}$ indicates the Fourier transform, $\Phi_x(j\omega)$ is called the PSD function of x_t. $r_x(\tau)$ is the *autocorrelation* function

$$r_x(\tau) = \mathbb{E}\left\{x_t x_{t+\tau}\right\} \tag{2.2}$$

where $\mathbb{E}\left\{\cdot\right\}$ denotes the expectation operator. The autocorrelation function is also an important descriptor for stochastic processes. If $r_x(\tau)$ decreases rapidly with respect to τ, it is equivalent to say that x_t changes rapidly with respect to time, and vice versa. A CT white noise e_t is defined to have zero mean and a constant PSD function

$$\mathbb{E}\left\{e_t\right\} = 0 \tag{2.3}$$
$$\Phi_e(j\omega) = \sigma^2 \tag{2.4}$$

The autocorrelation function of e_t can be obtained by applying the inverse Fourier transform to (2.4)

$$r_e(\tau) = \sigma^2\delta_\tau \tag{2.5}$$

where δ_τ is the Dirac's delta function. Equation (2.5) shows that a CT white noise is uncorrelated to time. However, when analyzing its variance, we found ourselves with a dilemma that a CT white noise has infinite variance! It jumps extremely fast and infinitely far around zero. The reason can be explained as: a CT white noise is assumed to have a constant PSD function, indicating that it contains harmonics for all frequencies, and each has the same power, therefore a CT white noise has infinite power.

Obviously, a CT noise is only a mathematical conjecture. When returning to the physical world, a more reasonable alternative is to consider a band-limited white noise, which has only a constant PSD function over a finite range of frequencies

$$\Phi_{ebl}(j\omega) = \begin{cases} \sigma^2 & |\omega| < \omega_b \\ 0 & |\omega| \geq \omega_b \end{cases} \tag{2.6}$$

where ω_b is the bandwidth. Then, it turns out that the autocorrelation function of a band-limited white noise can be as follows

$$r_{ebl}(\tau) = \frac{1}{2\pi}\int_{-\omega_b}^{\omega_b} \Phi_{ebl}(j\omega)e^{j\omega\tau}d\omega = \frac{\omega_b\sigma^2}{\pi} sinc\left(\omega_b\tau\right) \tag{2.7}$$

where $sinc$ is the cardinal sine function. As it can be seen from Figure 2.1, a band-limited noise is correlated at neighboring points, which greatly complicates the analysis. By contrast, a CT white noise is quite easy to analyze since it is uncorrelated at distinct time-instants. Therefore, the CT white noise concept will be used in the following chapters to facilitate the theoretical analysis.

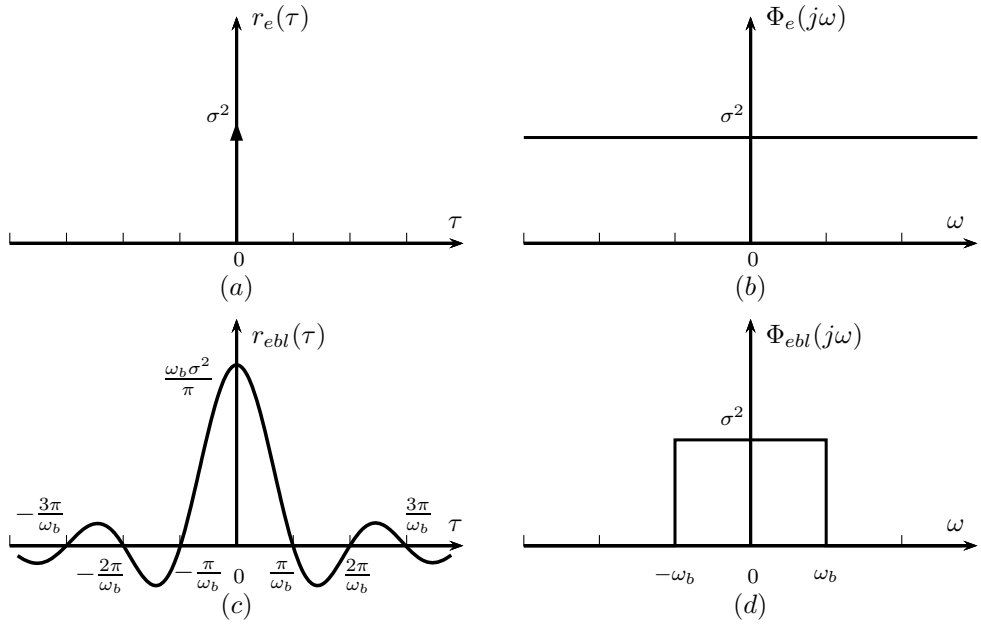

Figure 2.1: The autocorrelation function $(a), (c)$ and power spectral density $(b), (d)$ of the CT white and band-limited white noises.

2.2 Continuous-time Stochastic System

As it is stated in Chapter 1, the additive measurement noise ξ_t takes the following form

$$\xi_t = H(p)e_t = \frac{C(p)}{D(p)}e_t \tag{2.8}$$

where p denotes the differential operator. $D(p)$ and $C(p)$ are polynomials of the structures

$$C(p) = p^{n_c} + c_1 p^{n_c-1} + \cdots + c_{n_c}$$
$$D(p) = p^{n_d} + d_1 p^{n_d-1} + \cdots + d_{n_d}$$

Moreover, $D(p)$ and $C(p)$ are assumed to be coprime and have all zeros located in the left half plane. The degrees of the denominator and the numerator are denoted by n_d and n_c $(n_d \geq n_c)$, respectively. e_t has a constant PSD function of σ^2

$$\mathbb{E}\left\{e_t e_{t+\tau}\right\} = \sigma^2\delta_\tau \tag{2.9}$$

The unknown parameters are stacked in a column vector

$$\boldsymbol{\eta} = [d_1 \ \cdots \ d_{n_d} \ c_1 \ \cdots \ c_{n_c}]^T \tag{2.10}$$

The PSD function of ξ_t is given as (see [Söd02, Lar05])

$$\Phi_\xi(j\omega) = \sigma^2 \frac{C(-j\omega)C(j\omega)}{D(-j\omega)D(j\omega)} \tag{2.11}$$

Note that $n_d > n_c$ is assumed in [Lar05] to guarantee that ξ_t has a finite variance. However, this assumption is not necessary in this thesis. When ξ_t has infinite variance $(n_d = n_c)$, the integrated sampling scheme (see Section 2.4) can then be applied.

2.3 Instantaneous Sampling

2.3.1 Sampled Data Model

When $n_d > n_c$, ξ_t has a finite variance and the instantaneous sampling scheme can be applied. Write System (2.8) into its equivalent state-space form

$$\begin{cases} \dot{\boldsymbol{z}}_t = \boldsymbol{F}_c \boldsymbol{z}_t + \boldsymbol{w}_t \\ \xi_t = \boldsymbol{H}_c \boldsymbol{z}_t \end{cases} \tag{2.12}$$

where $\boldsymbol{z}_t$ is a n_d-dimensional state variable vector and $\dot{\boldsymbol{z}}_t$ denotes its time-derivative. A possible choice for $\boldsymbol{F}_c$ and $\boldsymbol{H}_c$ can be the following controllable canonical form

$$\boldsymbol{F}_c = \begin{bmatrix} -d_1 & \cdots & -d_{n_d} \\ \boldsymbol{I}_{n_d-1} & & \boldsymbol{0} \end{bmatrix}, \ \boldsymbol{H}_c = \begin{bmatrix} \boldsymbol{0} & c_1 & \cdots & c_{n_c} \end{bmatrix}$$

$\boldsymbol{w}_t$ is white and has the following covariance matrix

$$\mathbb{E}\{\boldsymbol{w}_t \boldsymbol{w}_{t+\tau}^T\} = \boldsymbol{Q}_c \delta_\tau = \begin{bmatrix} \sigma^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \delta_\tau \tag{2.13}$$

Equation (2.12) takes the form of a first order ordinary differential equation, the solution of the process equation is given as

$$\boldsymbol{z}_t = e^{\boldsymbol{F}_c t} \boldsymbol{z}_{t_0} + \int_0^t e^{\boldsymbol{F}_c (t-s)} \boldsymbol{w}_s ds \tag{2.14}$$

where $\boldsymbol{z}_{t_0}$ is the initial state at time $t = 0$. Consequently, the *sampled data model* for the noise process is given as

$$\begin{cases} \boldsymbol{z}_{t_{k+1}} = \boldsymbol{F}_{h_k} \boldsymbol{z}_{t_k} + \tilde{\boldsymbol{w}}_{t_k} \\ \xi_{t_k} = \boldsymbol{H}_c \boldsymbol{z}_{t_k} \end{cases} \tag{2.15}$$

where

$$\boldsymbol{F}_{h_k} = e^{\boldsymbol{F}_c h_k}, \ \tilde{\boldsymbol{w}}_{t_k} = \int_0^{h_k} e^{\boldsymbol{F}_c t} \boldsymbol{w}_{t_{k+1}-t} dt$$

It can be readily shown that $\tilde{\boldsymbol{w}}_{t_k}$ is a DT white noise having the covariance matrix $\boldsymbol{Q}_{h_k}$

$$\boldsymbol{Q}_{h_k} = \mathbb{E}\{\tilde{\boldsymbol{w}}_{t_k} \tilde{\boldsymbol{w}}_{t_k}^T\} = \int_0^{h_k} e^{\boldsymbol{F}_c t} \boldsymbol{Q}_c e^{\boldsymbol{F}_c^T t} dt \tag{2.16}$$

When the CT system is time-invariant, $\boldsymbol{F}_{h_k}$ and $\boldsymbol{Q}_{h_k}$ are uniquely defined by h_k, details to compute these terms are given in Appendix A.

Remark 2.1. *Note that even though $\boldsymbol{Q}_c$ is singular, $\boldsymbol{Q}_{h_k}$ is still full rank and positive definite (see e.g. [GAG+13]), therefore model (2.15) needs n_d channels of uncorrelated DT white noise input. The sampled data model (2.15) behaves as a multi-input single-output model. The full rank covariance also gives much convenience in using the expectation-maximization algorithm (see Chapter 3).*

2.3.2 Aliasing

To illustrate the aliasing phenomenon, we assume in this subsection that the system is regularly sampled. By using the matrices given in (2.15), the PSD function of the sampled sequence ξ_{kh} is (see [Söd02, WLS93])

$$\Phi_\xi^d\left(e^{j\omega h}\right) = h\boldsymbol{H}_c\left(e^{j\omega h}\boldsymbol{I} - \boldsymbol{F}_h\right)^{-1}\boldsymbol{Q}_h\left(e^{-j\omega h}\boldsymbol{I} - \boldsymbol{F}_h^T\right)^{-1}\boldsymbol{H}_c^T \tag{2.17}$$

where h is the sampling period. It has been shown that the folding of high frequency component to low frequency component (aliasing) occurs in sampling, see *e.g.* [Söd02, Söd90, Wah88, WLS93, Lar05, Yuz05]. The relationship between $\Phi_\xi^d\left(e^{j\omega h}\right)$ and $\Phi_\xi(j\omega)$ is represented by the following lemma

Lemma 2.2 (Aliasing)**.** *Assuming that the stochastic process ξ_t is periodically, the PSD function of the sampled sequence ξ_{kh} can also be formulated in terms of*

$$\Phi_\xi^d\left(e^{j\omega h}\right) = \sum_{\ell=-\infty}^{+\infty}\Phi_\xi\left(j\omega + j\frac{2\pi}{h}\ell\right) \tag{2.18}$$

where h denotes the sampling period and $\Phi_\xi\left(j\omega\right)$ is the PSD function of ξ_t defined in (2.11).

Proof. See [FG96]. The PSD function of ξ_{kh} is defined to be the discrete Fourier transform of its autocorrelation function

$$r_\xi(\ell h) = \mathbb{E}\left\{\xi_{kh}\xi_{kh+\ell h}\right\} \tag{2.19}$$

$$\Phi_\xi^d\left(e^{j\omega h}\right) = h\sum_{\ell=-\infty}^{+\infty} r_\xi(\ell h)e^{-j\omega\ell h} \tag{2.20}$$

where ℓ is an integer. $r_\xi(\ell h)$ can be obtained by applying the inverse Fourier transform to $\Phi_\xi(j\omega)$

$$r_\xi(\ell h) = r_{\xi_t}|_{t=\ell h} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\Phi_\xi(js)e^{js\ell h}ds \tag{2.21}$$

Substituting (2.21) into (2.20) leads to

$$\begin{aligned}\Phi_\xi^d\left(e^{j\omega h}\right) &= \frac{h}{2\pi}\int_{-\infty}^{+\infty}\Phi_\xi(js)\sum_{\ell=-\infty}^{+\infty}e^{j(s-\omega)\ell h}ds = \int_{-\infty}^{+\infty}\Phi_\xi(js)\sum_{\ell=-\infty}^{+\infty}\delta\left(s-\omega-\frac{2\pi}{h}\ell\right)ds \\ &= \sum_{\ell=-\infty}^{+\infty}\Phi_\xi\left(j\omega + j\frac{2\pi}{h}\ell\right)\end{aligned} \tag{2.22}$$

where $\delta(\cdot)$ is Dirac's delta function. □

The following example gives a more illustrative explanation of the aliasing phenomenon.

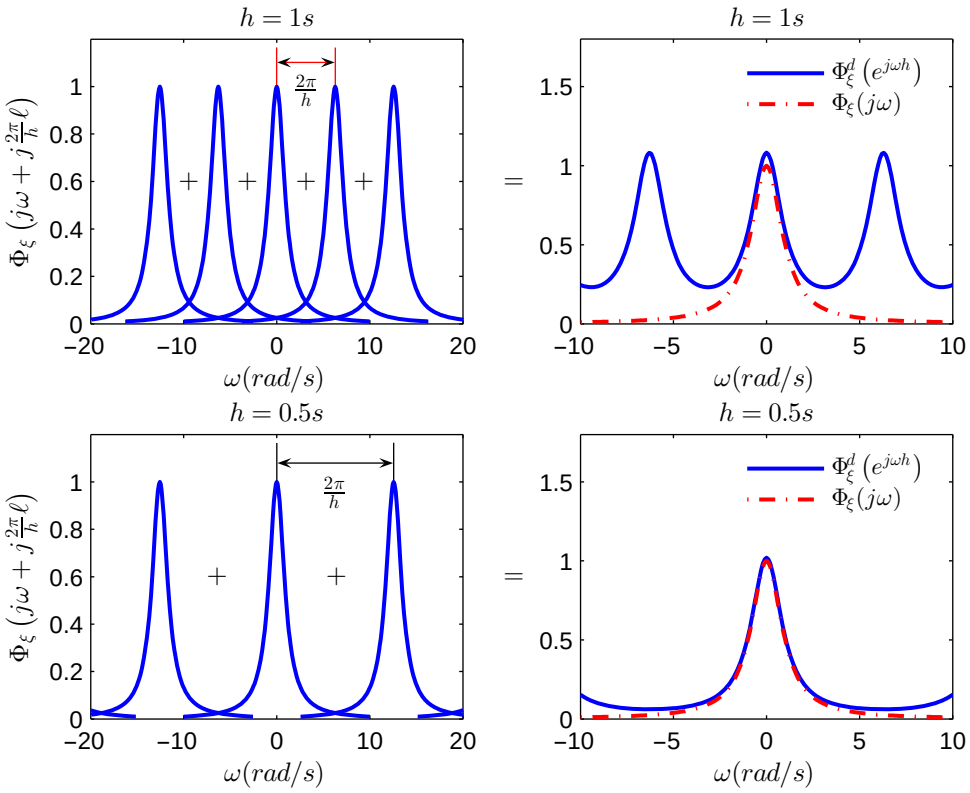

Figure 2.2: The PSD function of a CAR process and its sampled version by using the simplistic sampling scheme.

Example 2.1 (Aliasing)**.** *Consider a first order CT AR (CAR) process*

$$\xi_t = \frac{1}{p+1} e_t \tag{2.23}$$

where e_t is a CT white noise having a constant PSD function $\Phi_e(j\omega) = 1$. According to (2.11), the PSD function of ξ_t can be readily given as

$$\Phi_\xi(j\omega) = \frac{1}{\omega^2 + 1} \tag{2.24}$$

By applying the instantaneous sampling scheme, the PSD function of the sampled sequence ξ_{kh} takes the following form

$$\Phi^d_\xi\left(e^{j\omega h}\right) = \sum_{\ell=-\infty}^{+\infty} \Phi_\xi\left(j\omega + j\frac{2\pi}{h}\ell\right) = \frac{h\left(1 - e^{-2h}\right)}{2\left(e^{j\omega h} - e^{-h}\right)\left(e^{-j\omega h} - e^{-h}\right)} \tag{2.25}$$

where the second equation in (2.25) is obtained by using (2.17). Two sampling periods $h = 0.5s$ and $h = 1s$ are tried in this example. The results are presented in Figure 2.2, where $\left\{\Phi_\xi\left(j\omega + j\frac{2\pi}{h}\ell\right), \ell = -\infty, \cdots, \infty\right\}$ are plotted in the left column. It can be observed that the distance between two successive elements of $\left\{\Phi_\xi\left(j\omega + j\frac{2\pi}{h}\ell\right), \ell = -\infty, \cdots, \infty\right\}$ is larger when the sampling period h is smaller, this indicates that the folding of high frequency component to low frequency component is reduced, therefore the PSD function of the sampled process becomes closer to the original CT process. We can infer that $\Phi^d_\xi\left(e^{j\omega h}\right)$ converges to $\Phi_\xi(j\omega)$ when the sampling period h tends to zero.

2.4 Integrated Sampling

When $n_d = n_c = m$, System (2.8) can be rearranged as

$$\xi_t = \frac{C(p)}{D(p)} e_t = e_t + \frac{\bar{C}(p)}{D(p)} e_t \tag{2.26}$$

where

$$\bar{C}(p) = \bar{c}_1 p^{m-1} + \bar{c}_2 p^{m-2} + \cdots + \bar{c}_m, \quad \bar{c}_i = c_i - d_i$$

Similar to (2.12), the equivalent CT state-space representation can be given as

$$\begin{cases} \dot{\boldsymbol{z}}_t = \boldsymbol{F}_c \boldsymbol{z}_t + \boldsymbol{w}_t \\ \xi_t = \boldsymbol{H}_c \boldsymbol{z}_t + v_t \end{cases} \tag{2.27}$$

where

$$\boldsymbol{F}_c = \begin{bmatrix} -d_1 & \cdots & -d_m \\ \boldsymbol{I}_{m-1} & & \boldsymbol{0} \end{bmatrix}, \ \boldsymbol{H}_c = \begin{bmatrix} \bar{c}_1 & \cdots & \bar{c}_m \end{bmatrix}$$

$\boldsymbol{w}_t$ and v_t are white and have the following covariance matrix

$$\mathbb{E}\left\{ \begin{bmatrix} \boldsymbol{w}_t \\ v_t \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_{t+\tau} \\ v_{t+\tau} \end{bmatrix}^T \right\} = \begin{bmatrix} \boldsymbol{Q}_c & \boldsymbol{S}_c \\ \boldsymbol{S}_c^T & \boldsymbol{R}_c \end{bmatrix} \delta_\tau$$

where

$$\boldsymbol{Q}_c = \begin{bmatrix} \sigma^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \ \boldsymbol{S}_c = \begin{bmatrix} \sigma^2 \\ \boldsymbol{0} \end{bmatrix}, \ \boldsymbol{R}_c = \sigma^2$$

Since

$$\xi_t = \boldsymbol{H}_c \left(j\omega \boldsymbol{I} - \boldsymbol{F}_c\right)^{-1} \boldsymbol{w}_t + v_t = \begin{bmatrix} \boldsymbol{H}_c \left(j\omega \boldsymbol{I} - \boldsymbol{F}_c\right)^{-1} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_t \\ v_t \end{bmatrix} \tag{2.28}$$

The PSD function of ξ_t can be alternatively given as

$$\Phi_\xi(j\omega) = \begin{bmatrix} \boldsymbol{H}_c \left(j\omega \boldsymbol{I} - \boldsymbol{F}_c\right)^{-1} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_c & \boldsymbol{S}_c \\ \boldsymbol{S}_c^T & \boldsymbol{R}_c \end{bmatrix} \begin{bmatrix} \left(-j\omega \boldsymbol{I} - \boldsymbol{F}_c^T\right)^{-1} \boldsymbol{H}_c^T \\ \boldsymbol{I} \end{bmatrix} \tag{2.29}$$

As it can be seen from (2.28), v_t appears as a direct term in ξ_t, therefore ξ_t has infinite variance. Direct sampling of ξ_t leads to an impractical signal having infinite variance. To circumvent this problem, it is suggested to pass ξ_t through a low-pass filter to attenuate the high frequency components before sampling, this kind of scheme is called integrated sampling. The *integrated sampler* is given as (see Figure 2.3 for the block diagram)

$$\xi_{t_k} = \frac{1}{h_k} \int_{t_k}^{t_k + \delta_k} \xi_s ds \tag{2.30}$$

The low-pass filter has the following frequency response function (see Figure 2.4 for the im-

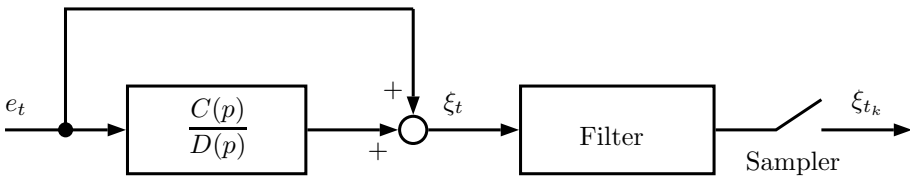

Figure 2.3: The block diagram of the integrated sampling scheme.

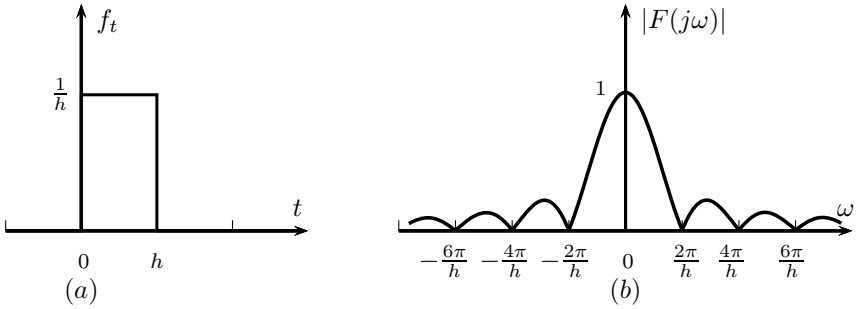

Figure 2.4: The impulse response (a) and amplitude characteristics (b) of the anti-aliasing filter.

pulse and frequency responses of the low-pass filter)

$$F(j\omega) = \frac{1 - e^{-j\omega h}}{j\omega h} = sinc\left(\frac{\omega h}{2}\right) e^{\frac{-j\omega h}{2}} \tag{2.31}$$

The integrated sampling scheme has been discussed in many papers, see *e.g.* [LW10, GAG+13, WLS93, FG96, Åst70]. By applying (2.30) to the ξ_t, the measurement equation becomes

$$\xi_{t_k} = \frac{1}{h_k}\left(\boldsymbol{H}_c \int_0^{h_k} e^{\boldsymbol{F}_c t} dt \; \boldsymbol{z}_{t_k} + \boldsymbol{H}_c \int_0^{h_k} \int_0^t e^{\boldsymbol{F}_c s} ds \; \boldsymbol{w}_{t_{k+1}-t} dt + \int_0^{h_k} v_{t_{k+1}-t} dt\right) \tag{2.32}$$

Then, the integrated sampled data model can be as follows

$$\begin{cases} \boldsymbol{z}_{t_{k+1}} = \boldsymbol{F}_{h_k} \boldsymbol{z}_{t_k} + \tilde{\boldsymbol{w}}_{t_k} \\ \xi_{t_k} = \boldsymbol{H}_{h_k} \boldsymbol{z}_{t_k} + \tilde{v}_{t_k} \end{cases} \tag{2.33}$$

where

$$\begin{aligned} \boldsymbol{F}_{h_k} &= e^{\boldsymbol{F}_c h_k} \\ \boldsymbol{H}_{h_k} &= \frac{1}{h_k} \boldsymbol{H}_c \int_0^{h_k} e^{\boldsymbol{F}_c t} dt \\ \tilde{\boldsymbol{w}}_{t_k} &= \int_0^{h_k} e^{\boldsymbol{F}_c t} \boldsymbol{w}_{t_{k+1}-t} dt \\ \tilde{v}_{t_k} &= \frac{1}{h_k} \boldsymbol{H}_c \int_0^{h_k} \int_0^t e^{\boldsymbol{F}_c s} ds \; \boldsymbol{w}_{t_{k+1}-t} dt + \frac{1}{h_k} \int_0^{h_k} v_{t_{k+1}-t} dt \end{aligned}$$

$\tilde{\boldsymbol{w}}_{t_k}$ and $\tilde{v}_{t_k}$ are white and have following covariance matrix

$$\mathbb{E}\left\{ \begin{bmatrix} \tilde{\boldsymbol{w}}_{t_k} \\ \tilde{v}_{t_k} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{w}}_{t_{k+\ell}} \\ \tilde{v}_{t_{k+\ell}} \end{bmatrix}^T \right\} = \begin{bmatrix} \boldsymbol{Q}_{h_k} & \boldsymbol{S}_{h_k} \\ \boldsymbol{S}_{h_k}^T & \boldsymbol{R}_{h_k} \end{bmatrix} \delta_\ell$$

To compute $\boldsymbol{Q}_{h_k}$, $\boldsymbol{R}_{h_k}$ and $\boldsymbol{S}_{h_k}$, let us firstly define the following notations

$$\boldsymbol{F}_t = e^{\boldsymbol{F}_c t},\ \bar{\boldsymbol{F}}_t = \int_0^t \boldsymbol{F}_s ds,\ \bar{\bar{\boldsymbol{F}}}_t = \int_0^t \bar{\boldsymbol{F}}_s ds \tag{2.34}$$

Based on the above notations, we have

$$\boldsymbol{Q}_{h_k} = \mathbb{E}\{\tilde{\boldsymbol{w}}_{t_k}\tilde{\boldsymbol{w}}_{t_k}^T\} = \int_0^{h_k} \boldsymbol{F}_t \boldsymbol{Q}_c \boldsymbol{F}_t^T dt \tag{2.35}$$

$$\begin{aligned}
\boldsymbol{R}_{h_k} =& \mathbb{E}\{\tilde{v}_{t_k}\tilde{v}_{t_k}^T\} = \frac{1}{h_k^2}\boldsymbol{H}_c \int_0^{h_k}\int_0^{h_k} \bar{\boldsymbol{F}}_t \underbrace{\mathbb{E}\{\boldsymbol{w}_{t_{k+1}-t}\boldsymbol{w}_{t_{k+1}-s}^T\}}_{\boldsymbol{Q}_c\delta_{t-s}} \bar{\boldsymbol{F}}_s^T ds\, dt\, \boldsymbol{H}_c^T \\
&+ \frac{1}{h_k^2}\boldsymbol{H}_c \int_0^{h_k}\int_0^{h_k} \bar{\boldsymbol{F}}_t \underbrace{\mathbb{E}\{\boldsymbol{w}_{t_{k+1}-t}v_{t_{k+1}-s}^T\}}_{\boldsymbol{S}_c\delta_{t-s}} ds\, dt \\
&+ \frac{1}{h_k^2} \int_0^{h_k}\int_0^{h_k} \underbrace{\mathbb{E}\{v_{t_{k+1}-s}\boldsymbol{w}_{t_{k+1}-t}^T\}}_{\boldsymbol{S}_c^T\delta_{t-s}} \bar{\boldsymbol{F}}_t^T ds\, dt\, \boldsymbol{H}_c^T \\
&+ \frac{1}{h_k^2} \int_0^{h_k}\int_0^{h_k} \underbrace{\mathbb{E}\{v_{t_{k+1}-t}v_{t_{k+1}-s}^T\}}_{\boldsymbol{R}_c\delta_{t-s}} ds\, dt \\
=& \frac{1}{h_k^2}\boldsymbol{H}_c \int_0^{h_k} \bar{\boldsymbol{F}}_t \boldsymbol{Q}_c \bar{\boldsymbol{F}}_t^T dt\, \boldsymbol{H}_c^T + \frac{1}{h_k^2}\boldsymbol{H}_c\bar{\bar{\boldsymbol{F}}}_{h_k}\boldsymbol{S}_c + \frac{1}{h_k^2}\boldsymbol{S}_c^T\bar{\bar{\boldsymbol{F}}}_{h_k}^T\boldsymbol{H}_c^T + \frac{1}{h_k}\boldsymbol{R}_c
\end{aligned} \tag{2.36}$$

$$\boldsymbol{S}_{h_k} = \mathbb{E}\{\tilde{\boldsymbol{w}}_{t_k}\tilde{v}_{t_k}^T\} = \frac{1}{h_k}\int_0^{h_k} \boldsymbol{F}_t\boldsymbol{Q}_c\bar{\boldsymbol{F}}_t^T dt\, \boldsymbol{H}_c^T + \frac{1}{h_k}\bar{\boldsymbol{F}}_{h_k}\boldsymbol{S}_c \tag{2.37}$$

An efficient method to compute $\boldsymbol{F}_{h_k}, \boldsymbol{H}_{h_k}, \boldsymbol{Q}_{h_k}, \boldsymbol{S}_{h_k}$ and $\boldsymbol{R}_{h_k}$ is presented in Appendix A. If ξ_{t_k} is uniformly sampled with the period h, then the PSD of the sampled sequence has a PSD function of

$$\Phi_\xi^d(e^{j\omega h}) = h\begin{bmatrix}\boldsymbol{H}_h\left(e^{j\omega h}\boldsymbol{I}-\boldsymbol{F}_h\right)^{-1} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{Q}_h & \boldsymbol{S}_h\\ \boldsymbol{S}_h^T & \boldsymbol{R}_h\end{bmatrix}\begin{bmatrix}\left(e^{-j\omega h}\boldsymbol{I}-\boldsymbol{F}_h^T\right)^{-1}\boldsymbol{H}_h^T \\ \boldsymbol{I}\end{bmatrix} \tag{2.38}$$

The following example is used to illustrate how the integrated sampling scheme and the sampling period h affect the PSD function of the sampled sequence.

Example 2.2. *Consider the following CT stochastic system*

$$\xi_t = \frac{C(p)}{D(p)}e_t = \frac{p^2+3p+5}{p^2+2p+2}e_t \tag{2.39}$$

where e_t is a CT white noise having a constant PSD function $\Phi_e(j\omega) = 1$. System (2.39) has the following equivalent CT state-space representation

$$\begin{cases}\dot{\boldsymbol{z}}_t = \boldsymbol{F}_c\boldsymbol{z}_t + \boldsymbol{w}_t \\ \xi_t = \boldsymbol{H}_c\boldsymbol{z}_t + v_t\end{cases} \tag{2.40}$$

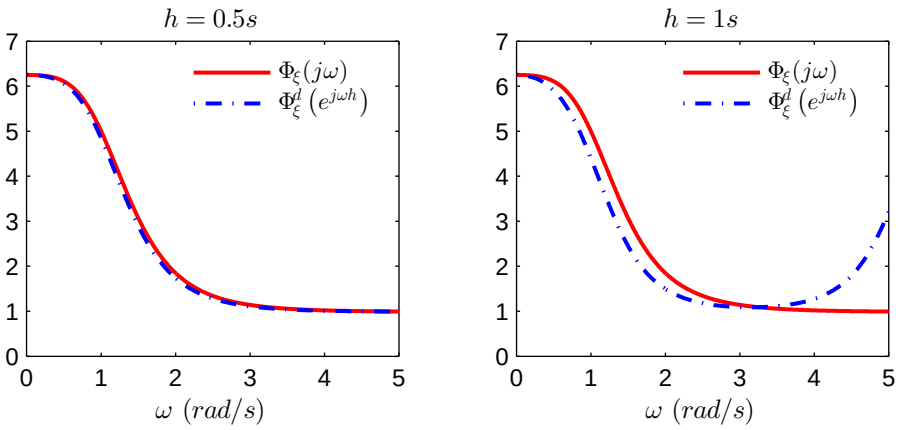

Figure 2.5: The PSD functions for two sampling periods by using the integrated sampling scheme.

where

$$\boldsymbol{F}_c = \begin{bmatrix} -2 & -2 \\ 1 & 0 \end{bmatrix}, \boldsymbol{H}_c = \begin{bmatrix} 1 \\ 3 \end{bmatrix}^T, \ \boldsymbol{Q}_c = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ \boldsymbol{S}_c = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ \boldsymbol{R}_c = 1$$

According to (A.8) – (A.21), an integrated sampled DT state-space model is obtained as

$$\begin{cases} \boldsymbol{z}_{kh+h} = \boldsymbol{F}_h \boldsymbol{z}_{kh} + \tilde{w}_{kh} \\ \xi_{kh} = \boldsymbol{H}_h \boldsymbol{z}_{kh} + \tilde{v}_{kh} \end{cases} \tag{2.41}$$

For $h = 0.5s$

$$\boldsymbol{F}_{0.5} = \begin{bmatrix} 0.2415 & -0.5816 \\ 0.2908 & 0.8231 \end{bmatrix}, \boldsymbol{H}_{0.5} = \begin{bmatrix} 1.1124 \\ 2.4524 \end{bmatrix}^T, \boldsymbol{Q}_{0.5} = \begin{bmatrix} 0.1931 & 0.0423 \\ 0.0423 & 0.0192 \end{bmatrix}, \boldsymbol{S}_{0.5} = \begin{bmatrix} 0.7054 \\ 0.2388 \end{bmatrix}$$
$$\boldsymbol{R}_{0.5} = 3.2976$$

For $h = 1s$

$$\boldsymbol{F}_1 = \begin{bmatrix} -0.1108 & -0.6191 \\ 0.3096 & 0.5083 \end{bmatrix}, \boldsymbol{H}_1 = \begin{bmatrix} 1.0471 \\ 1.9120 \end{bmatrix}^T, \boldsymbol{Q}_1 = \begin{bmatrix} 0.1990 & 0.0479 \\ 0.0479 & 0.0687 \end{bmatrix}, \boldsymbol{S}_1 = \begin{bmatrix} 0.3795 \\ 0.4052 \end{bmatrix}$$
$$\boldsymbol{R}_1 = 2.4802$$

The PSD functions for the sampled sequence ξ_{kh} *and the CT signal* ξ_t *are plotted in Figure 2.5. When the sampling period is small* ($h = 0.5s$)*, the sampled sequence has nearly the same PSD function as the CT one within* $\omega \in [0, 5] rad/s$*. However, it becomes more biased when the sampling period turns larger.*

2.5 Simplistic Sampling

The computation load of the matrices of the integrated sampled data model (2.33) is heavy. Thus, the following simplistic sampling scheme is considered ([BH97, LW10])

$$\xi_{t_k} = \frac{1}{h_k} \int_{t_k}^{t_k+\delta_k} \xi_s ds = \frac{1}{\delta_k} \int_{t_k}^{t_k+\delta_k} (\boldsymbol{H}_c \boldsymbol{z}_s + v_s)\, ds$$

$$= \boldsymbol{H}_c \boldsymbol{z}_{t_k} + \frac{1}{h_k} \int_{t_k}^{t_k+\delta_k} v_s ds \tag{2.42}$$

where it is assumed that $\boldsymbol{z}_t$ is constant during the sampling interval, the averaging on $\boldsymbol{z}_t$ is thus ignored. Under this assumption, the sampled data model by using simplistic sampling scheme can be as follows

$$\begin{cases} \boldsymbol{z}_{t_{k+1}} = \boldsymbol{F}_{h_k} \boldsymbol{z}_{t_k} + \tilde{\boldsymbol{w}}_{t_k} \\ \xi_{t_k} = \boldsymbol{H}_c \boldsymbol{z}_{t_k} + \tilde{v}_{t_k} \end{cases} \tag{2.43}$$

where

$$\boldsymbol{F}_{h_k} = e^{\boldsymbol{F}_c h_k}$$

$$\mathbb{E}\left\{ \begin{bmatrix} \tilde{\boldsymbol{w}}_{t_k} \\ \tilde{v}_{t_k} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{w}}_{t_k} \\ \tilde{v}_{t_k} \end{bmatrix}^T \right\} = \begin{bmatrix} \boldsymbol{Q}_{h_k} & \boldsymbol{S}_{h_k} \\ \boldsymbol{S}_{h_k}^T & \boldsymbol{R}_{h_k} \end{bmatrix}$$

where

$$\boldsymbol{Q}_{h_k} = \int_0^{h_k} \boldsymbol{F}_t \boldsymbol{Q}_c \boldsymbol{F}_t^T dt, \ \boldsymbol{R}_{h_k} = \frac{1}{h_k} \boldsymbol{R}_c, \ \boldsymbol{S}_{h_k} = \frac{1}{h_k} \int_0^{h_k} e^{\boldsymbol{F}_c t} dt \ \boldsymbol{S}_c$$

The following example is used to illustrate the characteristics of the simplistic sampling scheme

Example 2.3. *Consider the same system as it is given in Example 2.2*

$$\xi_t = \frac{C(p)}{D(p)} e_t = \frac{p^2 + 3p + 5}{p^2 + 2p + 2} e_t \tag{2.44}$$

where e_t is a CT white noise having a constant PSD function $\Phi_e(j\omega) = 1$. According to (2.43), a DT state-space model can be obtained by using the simplistic sampling scheme

$$\begin{cases} \boldsymbol{z}_{kh+h} = \boldsymbol{F}_h \boldsymbol{z}_{kh} + \tilde{w}_{kh} \\ \xi_{kh} = \boldsymbol{H}_h \boldsymbol{z}_{kh} + \tilde{v}_{kh} \end{cases} \tag{2.45}$$

Four different sampling periods are tried $h = \{0.01, 0.1, 0.5, 1\}s$. The PSD functions for the simplistic and integrated sampling schemes are plotted in Figure 2.6. It illustrates that when the sampling period is small ($h = 0.01s$ and $h = 0.1s$), both the two sampling schemes given accurate results. When the sampling period becomes larger ($h = 0.5s$ and $h = 0.1s$), the simplistic scheme becomes more biased. By contrast, the integrated scheme remains relatively better.

2.6 Conclusion

In this chapter, sampling issues of CT stochastic systems have been reviewed. The concept of CT white noise has been recalled in the beginning. Three sampling schemes have been reviewed: 1) the instantaneous sampling scheme, which is used to sample a system when the system output has finite variance; 2) the integrated sampling scheme, which is proposed to sample a stochastic process having infinite variance; 3) the simplistic scheme, which is the simplified version of the integrated sampling scheme that can reduce the computational load of the DT model matrices. Other issues such as aliasing, the computation of the DT model matrices have also been discussed.

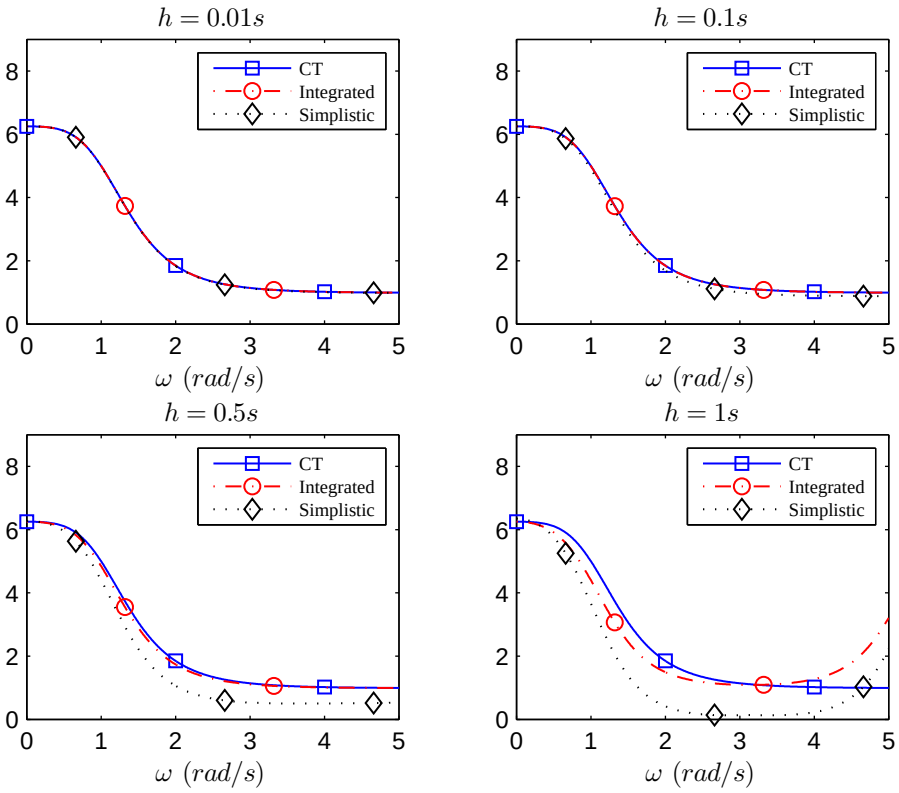

Figure 2.6: The PSD functions for different sampling periods by using the integrated and simplistic sampling schemes.

2.7 Relevant Matlab Routines

The relevant Matlab routines developed for this chapter are given below

1. `c2d_int`/`c2d_simp`: integrated/simplistic sampling of the given stochastic system.

   Syntax: `[Fh,Hh,Qh,Sh,Rh]=c2d_int/c2d_simp(Fc,Hc,Qc,Sc,Rc,h)`

   - `Fc,Hc,Qc,Sc,Rc` are the CT model matrices, h is the sampling interval.
   - `Fh,Hh,Qh,Sh,Rh` are the returned DT model matrices.

2. `c2d_ins`: instantaneous sampling of the given stochastic system.

   Syntax: `[Fh,Qh]=c2d_ins(Fc,Qc,h)`

3. `simcs_ins`/`simc_int`/`simc_simp`: one random realization of the given stochastic system using the instantaneous/integrated/simplistic sampling scheme, respectively.

   $$y_t = \frac{C(p)}{D(p)} e_t$$

   where e_t is a CT white noise having a constant PSD function $\Phi_e(j\omega) = 1$.

   Syntax: `y=simcs_xxx(den,num,t)`

- t is a column vector containing the sampling instants.

Chapter 3

Identification of Continuous-time Noise Models from Irregularly Sampled Data

Noise exists everywhere in the physical world, therefore the data collected from the real application is inevitably noise-contaminated. In order to obtain high-quality statistical estimation results from noisy data, this chapter investigates the problem of noise modeling. As it was stated in Section 1.4, recent identification schemes are quite dedicated to the estimation of Hybrid Box-Jenkins models, where the plant is modeled in CT while the noise is modeled in DT (see [YGG08, Joh94]). However, this hybrid model cannot be used when data are irregularly sampled because of the DT noise model assumption. In this thesis, we try our first attempt to estimate CT AR/ARMA (CAR/CARMA) noise models from irregularly sampled data, which is a part of the work for estimating CT Box-Jenkins models from irregularly sampled data presented in the next chapter.

Estimation of CAR/CARMA models has been discussed in several recent papers. For example, in [Söd91, KMU11], indirect approaches were suggested to estimate CARMA models from regularly sampled data, where a DT model is firstly estimated and then it is converted to a CT model. A frequency domain approach was developed in [GL09] to estimate directly the CARMA model. In [SCB92, LS02, Mos08b], derivative approximation methods were presented to estimate CAR parameters from irregularly sampled data by linear regression. The same idea was used in [FSM$^+$99, Pha00], where some compensating techniques were developed to reduce the bias. The criteria for uniquely determination the CAR parameters from sampled data was discussed in [KUW14].

In this thesis, the EM-based approach to estimate CAR/CARMA models from irregularly sampled data is suggested. The proposed method employs a separable way to estimate the CARMA parameters. Precisely, the EM algorithm is applied to estimate the CAR parameters, while a numerical method is used to estimate the MA parameters.

For the CAR parameter estimation, the goal is to develop a maximum-likelihood (ML) estimator by using the EM algorithm, which was firstly proposed in [DLR77] and becomes a well-accepted approach to solve ML problems. Compared with the traditional ML method, which uses numerical methods to maximize the non-linear log-likelihood function, the EM algorithm is quite easy to implement and show several advantages. As pointed out in [Shu82], the traditional ML estimators have some unattractive features. First, they involve the calcula-

tion of the inverse of Hessian matrix, which can be very large or may not exist if we have a large number of parameters to estimate. Secondly, the numerical methods may not necessarily increase the log-likelihood function in the successive iterations. Fortunately, these unattractive features can be avoided by using the EM algorithm. The EM algorithm is an iterative procedure. Each iteration consists of an expectation step followed by a maximization step, each step is easier to solve than the original problem. The M-step even has closed-form solutions in some cases. The convergence analysis was discussed in [Wu83]. The EM algorithm was first adapted to estimate DT state-space models in [Shu82], later developments can be found in *e.g.* [Isa93, GA05, GN05, SWN11, YAAG11, ATY$^+$12, WSLN13, CGG14, AYG07, KA08].

This chapter is organized as follows. In Section 3.1, the derivative approximation method for CAR model estimation is reviewed. The basic theory of the EM algorithm is recalled in Section 3.2. Subsequently, the basic EM algorithm for CAR model estimation is detailed in Section 3.3. Then Section 3.4 presents the improved EM algorithm for CAR model estimation by using the state interpolation technique. The problem of CARMA model estimation is detailed in Section 3.6. Finally, we present some examples in Section 3.7 and draw some conclusions in Section 3.8.

3.1 Review of the Derivative Approximation Method for CAR Model Estimation

This derivative approximation method for CT stochastic process identification was developed by Söderström and his colleagues in early 1990s, who claimed the following advantages

- It shows good computational efficacy for both uniform and non-uniformly sampled data.
- It has good numerical properties for fast sampled data.

3.1.1 Problem Statement

Consider the following CAR process

$$D(p, \boldsymbol{\eta}^o)\xi_t = e_t \tag{3.1}$$

where *p* is the differential operator. $D(p, \boldsymbol{\eta}^o)$ is a polynomial of degree n_d, which is assumed to be *a priori* known. e_t is a Gaussian CT white noise with the intensity σ_o^2

$$D(p, \boldsymbol{\eta}^o) = p^{n_d} + d_1^o p^{n_d-1} + \cdots + d_{n_d}^o$$
$$\mathbb{E}\left\{e_t e_{t+\tau}^T\right\} = \sigma_o^2 \delta_\tau$$

It is further assumed that all the zeros of $D(p, \boldsymbol{\eta}^o)$ are located in the left half plane. The unknown parameters are stacked in a column vector $\boldsymbol{\theta}^o$

$$\boldsymbol{\eta}^o = \begin{bmatrix} d_1^o & \cdots & d_{n_d}^o \end{bmatrix}^T \tag{3.2}$$
$$\boldsymbol{\theta}^o = \begin{bmatrix} \boldsymbol{\eta}^o; & \sigma_o^2 \end{bmatrix} \tag{3.3}$$

Without loss of generality, System (3.1) is assumed to be sampled at time-instants t_k, for $k = 1, 2, \cdots, N$. The time-varying sampling period is denoted by h_k

$$h_k = t_{k+1} - t_k, \;\; k = 1, 2, \cdots, N-1 \tag{3.4}$$

Then the identification objective is to estimate the unknown parameters in $\boldsymbol{\theta}^o$ from non-uniformly sampled data $\{\xi_{t_{1:N}}\}$ when n_d is known.

3.1.2 Parameter Estimators

Equation (3.1) has the following regression model

$$\xi_{t_k}^{(n_d)} = \boldsymbol{\phi}_{t_k}^T \boldsymbol{\eta} + \varepsilon_{t_k} \tag{3.5}$$

where

$$\boldsymbol{\phi}_{t_k}^T = \left[\xi_{t_k}^{(n_d-1)}, \xi_{t_k}^{(n_d-2)}, \cdots, \xi_{t_k}\right] \tag{3.6}$$

When the time-derivatives of ξ_t are obtained, $\boldsymbol{\eta}$ can be estimated by means of least-squares. However, these time-derivatives are usually not measured. Thus, an alternative is suggested in [SCB92] to approximate the differential operator by the difference operator. One solution is briefly recalled and used here. Substituting p^i by a difference operator D^i leads to

$$p^i f_{t_k} \approx D^i f_{t_k} = \sum_{\mu=0}^{i} \beta_k(i,\mu) f_{t_{k+\mu}} \tag{3.7}$$

To concrete the approximation, it is natural that $\beta_k(i,\mu)$ satisfies the following condition

$$\sum_{\mu=0}^{i} \beta_k(i,\mu)(t_{k+\mu} - t_k)^\nu = \begin{cases} i! & \nu = i \\ 0 & \nu = 0, ..., i-1 \end{cases} \tag{3.8}$$

where the solution of $\beta_k(i,\mu)$ is given by (see *e.g.* [LS02])

$$\beta_k(i,\mu) = \frac{i!}{\prod_{\substack{s=0 \\ s\neq\mu}}^{i}(t_{k+\mu} - t_{k+s})}, \;\; \mu = 0, \cdots, i \tag{3.9}$$

By using the approximated derivatives, we have the following regression model

$$D^{n_d}\xi_{t_k} = \boldsymbol{\phi}_{t_k}^T \boldsymbol{\eta} + \varepsilon_{t_k} \tag{3.10}$$

where

$$\boldsymbol{\phi}_{t_k} = \left[D^{n_d-1}\xi_{t_k}, D^{n_d-2}\xi_{t_k}, \cdots, \xi_{t_k}\right] \tag{3.11}$$

The unknown parameters can be estimated by the following two schemes: the *shifted least-squares* and *instrumental variable methods*.

- The shifted least-squares method (see [LS02]):

$$\hat{\boldsymbol{\eta}} = \left[\sum_{k=1}^{N} \boldsymbol{\phi}_{t_k}\boldsymbol{\phi}_{t_k}^T\right]^{-1} \sum_{k=1}^{N} \boldsymbol{\phi}_{t_k} D^{n_d}\xi_{t_{k+s}} \tag{3.12}$$

  where s is a non-negative integer. This leads to the forward shifting of the n_d-th derivative, thus it is named shifted least-squares. It was suggested in [LS02] that $s = n_d - 1$.

- The IV method (see [SCB92]):

$$\hat{\boldsymbol{\eta}} = \left[\sum_{k=1}^{N} \boldsymbol{\psi}_{t_k} \boldsymbol{\phi}_{t_k}^T\right]^{-1} \sum_{k=1}^{N} \boldsymbol{\psi}_{t_k} D^{n_d} \xi_{t_{k+s}} \tag{3.13}$$

where

$$\boldsymbol{\psi}_{t_k}^T = \left[\xi_{t_{k-L-1}}, \cdots, \xi_{t_{k-L-n_d}}\right] \tag{3.14}$$

where $\boldsymbol{\psi}_{t_k}$ is the instrumental variable built up from a delayed version of the measured output. L is a nonnegative integer. It was shown in [SCB92] that $L = 0$ is the best choice.

Example 3.1. *Consider a second order CAR process (see [SCB92])*

$$\left(p^2 + d_1^o p + d_2^o\right) \xi_t = e_t \tag{3.15}$$

where $d_1^o = 2, d_2^o = 2$, e_t *has the intensity* $\sigma_o^2 = 1$. *The simulation is performed under the following conditions*

- ξ_t *is instantaneously observed for* $N = 10,000$ *times, the sampling period* h_k *is uniformly distributed in the following interval*

$$h_k \sim U\left[0.01, \bar{h}\right] s$$

where $\bar{h}$ *increases gradually from* $0.01s$ *to* $0.1s$ *with step size* $0.01s$.

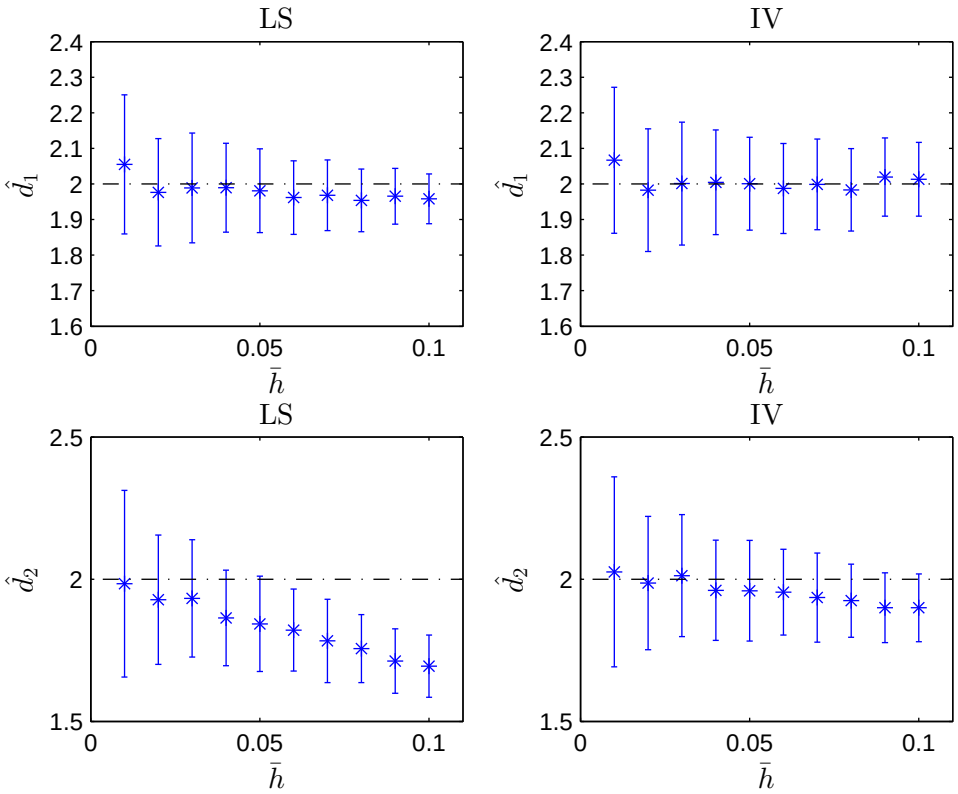

Figure 3.1: The mean ($*$) and standard deviation (vertical bar) of the parameter estimates. Top – Estimated by the shifted least-squares method. Bottom – Estimated by the IV method.

- *A Monte-Carlo simulation with 100 realizations is run, from which the mean and variance of the estimated parameters are computed.*

The parameter estimates are presented in Figure 3.1. It is clear that the bias of d_2 increases monotonically with respect to $\bar{h}$, but the IV-based approach behaves much better. Another interesting phenomenon is that when $\bar{h}$ increases, the standard deviation slightly reduces. This is benefited from the increase of the time-span of the sampled data, which means the sampled data gathers more information.

3.2 Review of the EM Algorithm

3.2.1 Maximum Likelihood Method

The maximum likelihood (ML) estimation method has been developed for a century, good references are *e.g.* [SS89, Lju99]. The basic idea of the ML method is to select a set of model parameters that maximize the joint probability density (PDF) of the observations. More specifically, let $\{\xi_{t_{1:N}}\}$ denote the observations, the PDF function is given by

$$p_{\boldsymbol{\theta}}\left(\xi_{t_{1:N}}\right) = p\left(\xi_{t_{1:N}}|\boldsymbol{\theta}\right) = p_{\boldsymbol{\theta}}\left(\xi_{t_1}\right)\prod_{k=2}^{N} p_{\boldsymbol{\theta}}\left(\xi_{t_k}|\xi_{t_{1:k-1}}\right) \tag{3.16}$$

where the second equation in (3.16) is obtained by repeating the well-known Bayes' rule. The subscript $[\cdot]_{\boldsymbol{\theta}}$ means that $[\cdot]$ is parameterized by $\boldsymbol{\theta}$. With this in mind, the likelihood function is defined to be the logarithm of the joint PDF

$$l(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}\left(\xi_{t_{1:N}}\right) = \log p_{\boldsymbol{\theta}}\left(\xi_{t_1}\right) + \sum_{k=2}^{N} p_{\boldsymbol{\theta}}\left(\xi_{t_k}|\xi_{t_{1:k-1}}\right) \tag{3.17}$$

Then the ML estimate of $\boldsymbol{\theta}$ is as follows

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} l(\boldsymbol{\theta}) \tag{3.18}$$

Numerical methods are typical ways to solve (3.18). As it was recalled in the beginning of this chapter, $l(\boldsymbol{\theta})$ is always a non-linear function with respect to $\boldsymbol{\theta}$, the numerical methods present several disadvantages. For this reason, the EM algorithm is chosen since it offers a more robust alternative to the traditional approach.

3.2.2 Expectation-Maximization Algorithm

Instead of direct maximizing $l(\boldsymbol{\theta})$, the EM algorithm uses an iterative way to implement the ML method. Each iteration consists of an expectation step followed by a maximization step, and each step is much easier to solve than the original problem. One flexibility of the EM algorithm is the choice of the *hidden variables*, which helps to simplify the complexity of

parameter estimation. Let $\{\boldsymbol{\zeta}_{t_{1:N}}\}$ denote the hidden variable, then, $l(\boldsymbol{\theta})$ can be reformulated as (by using the Bayes' rule)

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log p_{\boldsymbol{\theta}}\left(\xi_{t_{1:N}}\right) = \log \frac{p_{\boldsymbol{\theta}}\left(\xi_{t_{1:N}}, \boldsymbol{\zeta}_{t_{1:N}}\right)}{p_{\boldsymbol{\theta}}\left(\boldsymbol{\zeta}_{t_{1:N}} | \boldsymbol{\zeta}_{t_{1:N}}\right)} \\ &= \log p_{\boldsymbol{\theta}}\left(\xi_{t_{1:N}}, \boldsymbol{\zeta}_{t_{1:N}}\right) - \log p_{\boldsymbol{\theta}}\left(\xi_{t_{1:N}} | \boldsymbol{\zeta}_{t_{1:N}}\right) \end{aligned} \tag{3.19}$$

where $\{\xi_{t_{1:N}}, \boldsymbol{\zeta}_{t_{1:N}}\}$ is called the *complete data*. Since $l(\boldsymbol{\theta})$ depends on the unknown parameters in $\boldsymbol{\theta}$, the EM algorithm considers the following conditioned PDF given $\xi_{t_{1:N}}$ and $\hat{\boldsymbol{\theta}}^0$

$$\begin{aligned} l\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0\right) &= \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{l(\boldsymbol{\theta}) | \xi_{t_{1:N}}\right\} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\log p_{\boldsymbol{\theta}}\left(\xi_{t_{1:N}}, \boldsymbol{\zeta}_{t_{1:N}}\right) | \xi_{t_{1:N}}\right\} - \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\log p_{\boldsymbol{\theta}}\left(\boldsymbol{\zeta}_{t_{1:N}} | \xi_{t_{1:N}}\right) | \xi_{t_{1:N}}\right\} \\ &= \mathscr{Q}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0\right) - \mathscr{H}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0\right) \end{aligned} \tag{3.20}$$

According to Jensen's equality (see *e.g.* [Rao65]), it can be proven that $\mathscr{H}\left(\hat{\boldsymbol{\theta}}^0, \hat{\boldsymbol{\theta}}^0\right) \geq \mathscr{H}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0\right)$. If we can find a $\boldsymbol{\theta}$ leading to $\mathscr{Q}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0\right) \geq \mathscr{Q}\left(\hat{\boldsymbol{\theta}}^0, \hat{\boldsymbol{\theta}}^0\right)$, it is equivalent to say that the likelihood increases $l\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0\right) \geq l\left(\hat{\boldsymbol{\theta}}^0, \hat{\boldsymbol{\theta}}^0\right)$. Consequently, the main works of the EM iteration becomes: compute the expectation $\mathscr{Q}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0\right)$ given $\hat{\boldsymbol{\theta}}^0, \xi_{t_{1:N}}$ and then maximize $\mathscr{Q}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0\right)$ with respect to $\boldsymbol{\theta}$ to obtain a new estimate. This iterative procedure will be terminated until a stationary point of $\mathscr{Q}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0\right)$ is found.

The EM algorithm is summarized below

Algorithm 3.1 (The EM algorithm)**.**

---

1. *Initialization, give the initial value* $\hat{\boldsymbol{\theta}}^0$.
2. *Iterative estimation.*
   for j=0:convergence
   (a) *Construct the auxiliary function*
   $$\mathscr{Q}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^j\right) = \mathbb{E}_{\hat{\boldsymbol{\theta}}^j}\left\{\log p_{\boldsymbol{\theta}}\left(\xi_{t_{1:N}}, \boldsymbol{\zeta}_{t_{1:N}}\right) | \xi_{t_{1:N}}\right\} \tag{3.21}$$
   (b) *Maximize the auxiliary function with respect to* $\boldsymbol{\theta}$
   $$\hat{\boldsymbol{\theta}}^{j+1} = \arg\max_{\boldsymbol{\theta}} \mathscr{Q}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^j\right) \tag{3.22}$$
   end

3.3 Basic EM Algorithm for CAR Model Estimation

3.3.1 δ-operator Model

The CAR model can be written into its equivalent state space form

$$\begin{cases} \dot{\boldsymbol{z}}_t = \boldsymbol{F}_c \boldsymbol{z}_t + \boldsymbol{w}_t \\ \xi_t = \boldsymbol{H}_c \boldsymbol{z}_t \end{cases} \tag{3.23}$$

$\boldsymbol{F}_c, \boldsymbol{H}_c$ take the controllable canonical form and $\boldsymbol{w}_t$ is a Gaussian CT noise

$$\boldsymbol{F}_c = \begin{bmatrix} -d_1^o & \cdots & -d_n^o \\ \boldsymbol{I}_{n-1} & & \boldsymbol{0} \end{bmatrix},\ \boldsymbol{H}_c = \begin{bmatrix} \boldsymbol{0} \\ 1 \end{bmatrix},\ \mathbb{E}\left\{\boldsymbol{w}_t \boldsymbol{w}_{t+\tau}^T\right\} = \boldsymbol{Q}_c \delta_\tau = \begin{bmatrix} \sigma_o^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \delta_\tau$$

By applying the instantaneous sampling scheme (see Chapter 2), the equivalent DT model can be given by

$$\begin{cases} \boldsymbol{z}_{t_{k+1}} = \boldsymbol{F}_{h_k} \boldsymbol{z}_{t_k} + \boldsymbol{w}_{t_k} \\ \xi_{t_k} = \boldsymbol{H}_c \boldsymbol{z}_{t_k} \end{cases} \tag{3.24}$$

where

$$\boldsymbol{F}_{h_k} = e^{\boldsymbol{F}_c h_k},\ \mathbb{E}\left\{\boldsymbol{w}_{t_k} \boldsymbol{w}_{t_{k+\ell}}^T\right\} = \boldsymbol{Q}_{h_k} \delta_\ell = \int_0^{h_k} e^{\boldsymbol{F}_c t} \boldsymbol{Q}_c e^{\boldsymbol{F}_c^T t} dt\ \delta_\ell \tag{3.25}$$

Let us define that

$$\boldsymbol{z}_{t_k}^\delta = h_k^{-1} \left(\boldsymbol{z}_{t_{k+1}} - \boldsymbol{z}_{t_k}\right) \tag{3.26}$$

$$\boldsymbol{F}_{h_k}^\delta = h_k^{-1} \left(\boldsymbol{F}_{h_k} - \boldsymbol{I}\right), \tag{3.27}$$

$$\boldsymbol{w}_{t_k}^\delta = h_k^{-1} \boldsymbol{w}_{t_k} \tag{3.28}$$

Then, according to (3.24), we have the following δ-operator model (see [FG96])

$$\begin{cases} \boldsymbol{z}_{t_k}^\delta = \boldsymbol{F}_{h_k}^\delta \boldsymbol{z}_{t_k} + \boldsymbol{w}_{t_k}^\delta \\ \xi_{t_k} = \boldsymbol{H}_c \boldsymbol{z}_{t_k} \end{cases} \tag{3.29}$$

Let

$$\boldsymbol{Q}_{h_k}^\delta = h_k \mathbb{E}\left\{\boldsymbol{w}_{t_k}^\delta \left(\boldsymbol{w}_{t_k}^\delta\right)^T\right\} = h_k^{-1} \boldsymbol{Q}_{h_k} \tag{3.30}$$

It is obvious that when the sampling interval h_k tends to zero, the following relationships hold

$$\lim_{h_k \to 0} \boldsymbol{F}_{h_k}^\delta = \boldsymbol{F}_c,\ \lim_{h_k \to 0} \boldsymbol{Q}_{h_k}^\delta = \boldsymbol{Q}_c \tag{3.31}$$

The above equations imply that the δ-operator model has some connections with the CT model. In particularly, it can be used to approximate the CT model in fast sampling situation and will be considered in the following sections ([YAAG11, MG90, GAG+13]).

3.3.2 The EM Algorithm

The key point of the EM algorithm is to choose the hidden variable. In [Shu82], it was suggested to choose $\boldsymbol{z}_t$. Then, the complete data thus becomes $\{\xi_{t_{1:N}}, \boldsymbol{z}_{t_{1:N}}\}$. At this moment, we find ourselves in a trouble: the relationship between ξ_t and $\boldsymbol{z}_t$ is deterministic since no measurement noise is assumed. In this case, we consider the following PDF function

$$l(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}} \left(\boldsymbol{z}_{t_{1:N}}\right) \tag{3.32}$$

It can be inferred that the same results can be drawn from $l'(\boldsymbol{\theta})$ and $l(\boldsymbol{\theta})$, therefore $l'(\boldsymbol{\theta})$ is used for further study. The EM algorithm constructs the following auxiliary function for parameter estimation

$$\mathscr{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0) = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0} \left\{ l(\boldsymbol{\theta}) | \xi_{t_{1:N}} \right\} \tag{3.33}$$

The E-step and M-step of the EM algorithm are detailed in the following two lemmas

Lemma 3.1 (E-step)**.** *Consider the model (3.24). Choose $\boldsymbol{z}_{t_{1:N}}$ as the complete data. Given the initial estimate $\hat{\boldsymbol{\theta}}^0$ and the initial state*

$$\boldsymbol{z}_{t_1} \sim \mathscr{N}\left(\boldsymbol{\mu}_{t_1}, \boldsymbol{P}_{t_1}\right)$$

By using the following relationships

$$\begin{aligned}
\boldsymbol{z}_{t_k}^{\delta} &= h_k^{-1}\left(\boldsymbol{z}_{t_{k+1}} - \boldsymbol{z}_{t_k}\right) \\
\boldsymbol{F}_{h_k}^{\delta} &= h_k^{-1}\left(\boldsymbol{F}_{h_k} - \boldsymbol{I}\right) \\
\boldsymbol{Q}_{h_k}^{\delta} &= h_k \boldsymbol{Q}_{h_k}
\end{aligned}$$

We have

$$\begin{aligned}
&-2\mathscr{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0) \\
=&\log\det \boldsymbol{P}_{t_1} + \sum_{k=1}^{N-1} \log\det\left(h_k \boldsymbol{Q}_{h_k}^{\delta}\right) + \mathrm{Tr}\left\{\boldsymbol{P}_{t_1}^{-1}\left(\boldsymbol{z}_{t_1} - \boldsymbol{\mu}_{t_1}\right)\left(\boldsymbol{z}_{t_1} - \boldsymbol{\mu}_{t_1}\right)^T\right\} \\
&+ \sum_{k=1}^{N-1} \mathrm{Tr}\left\{h_k\left(\boldsymbol{Q}_{h_k}^{\delta}\right)^{-1}\left[\boldsymbol{\Phi}_{t_k} - \boldsymbol{\Psi}_{t_k}\left(\boldsymbol{F}_{h_k}^{\delta}\right)^T - \boldsymbol{F}_{h_k}^{\delta}\boldsymbol{\Psi}_{t_k}^T + \boldsymbol{F}_{h_k}^{\delta}\boldsymbol{\Upsilon}_{t_k}\left(\boldsymbol{F}_{h_k}^{\delta}\right)^T\right]\right\}
\end{aligned} \tag{3.34}$$

where

$$\boldsymbol{\Phi}_{t_k} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_k}^{\delta}\left(\boldsymbol{z}_{t_k}^{\delta}\right)^T \middle| \xi_{t_{1:N}}\right\} \tag{3.35}$$

$$\boldsymbol{\Psi}_{t_k} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_k}^{\delta}\boldsymbol{z}_{t_k}^T \middle| \xi_{t_{1:N}}\right\} \tag{3.36}$$

$$\boldsymbol{\Upsilon}_{t_k} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_k}\boldsymbol{z}_{t_k}^T | \xi_{t_{1:N}}\right\} \tag{3.37}$$

Proof. See Appendix C.1 □

Lemma 3.2 (M-step)**.** *Consider the auxiliary function (3.68), the CT matrices that maximize $\mathscr{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0)$ are given by*

$$\hat{\boldsymbol{\eta}}^T = -\sum_{k=1}^{N-1} h_k \boldsymbol{W}^T \boldsymbol{\Psi}_{t_k}\left(\sum_{k=1}^{N-1} h_k \boldsymbol{\Upsilon}_{t_k}\right)^{-1} \tag{3.38}$$

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{k=1}^{N-1} h_k\left(\boldsymbol{W}^T \boldsymbol{\Phi}_{t_k} \boldsymbol{W} - \hat{\boldsymbol{\eta}}^T \boldsymbol{\Upsilon}_{t_k}\hat{\boldsymbol{\eta}}\right) \tag{3.39}$$

where $\boldsymbol{W} = [1 \;\; \boldsymbol{0}]^T$. The new estimate $\hat{\boldsymbol{\theta}}$ is given by $\hat{\boldsymbol{\theta}} = \left[\hat{\boldsymbol{\eta}};\; \hat{\sigma}^2\right]$.

Proof. See Appendix C.2. □

The construction of $\boldsymbol{\Phi}_{t_k}$, $\boldsymbol{\Psi}_{t_k}$ and $\boldsymbol{\Upsilon}_{t_k}$ in the E-step can be classified as state estimation. It shows that only three elementary terms need to be evaluated

$$\mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_k}\xi_{t_k}^T\middle|\xi_{t_{1:N}}\right\} = \hat{\boldsymbol{z}}_{t_k}^{t_N}\xi_{t_k}^T \tag{3.40}$$

$$\mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_k}\boldsymbol{z}_{t_k}^T\middle|\xi_{t_{1:N}}\right\} = \hat{\boldsymbol{z}}_{t_k}^{t_N}(\hat{\boldsymbol{z}}_{t_k}^{t_N})^T + \boldsymbol{P}_{t_k}^{t_N} \tag{3.41}$$

$$\mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_{k+1}}\boldsymbol{z}_{t_k}^T\middle|\xi_{t_{1:N}}\right\} = \hat{\boldsymbol{z}}_{t_{k+1}}^{t_N}(\hat{\boldsymbol{z}}_{t_k}^{t_N})^T + \boldsymbol{P}_{t_{k+1},t_k}^{t_N} \tag{3.42}$$

where the computation of $\hat{\boldsymbol{z}}_{t_k}^{t_N}, \boldsymbol{P}_{t_k}^{t_N}, \boldsymbol{P}_{t_{k+1},t_k}^{t_N}$ are given in the next section.

3.3.3 State Estimation

Let us declare the following definitions

- *a priori* estimate: $\hat{\boldsymbol{z}}_{t_k}^{t_{k-1}}$

  $$\hat{\boldsymbol{z}}_{t_k}^{t_{k-1}} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_k}|\xi_{t_{1:k-1}}\right\} \tag{3.43}$$

  $\hat{\boldsymbol{z}}_{t_k}^{t_{k-1}}$ is conditioned on all of the measurements up to the time-instant t_{k-1}, it is a prediction of $\boldsymbol{z}_{t_k}$ using the previous information.

- *a posteriori* estimate: $\hat{\boldsymbol{z}}_{t_k}^{t_k}$:

  $$\hat{\boldsymbol{z}}_{t_k}^{t_k} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_k}|\xi_{t_{1:k}}\right\} \tag{3.44}$$

  $\hat{\boldsymbol{z}}_{t_k}^{t_k}$ is conditioned on all of the measurements up to t_k.

- Estimation error covariance: $\boldsymbol{P}_{t_k}^{t_{k-1}}$ and $\boldsymbol{P}_{t_k}^{t_k}$

  $$\begin{aligned}\boldsymbol{P}_{t_k}^{t_{k-1}} &= \operatorname{cov}\left\{\hat{\boldsymbol{z}}_{t_k}^{t_{k-1}}, \hat{\boldsymbol{z}}_{t_k}^{t_{k-1}}\right\} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\left(\boldsymbol{z}_{t_k} - \hat{\boldsymbol{z}}_{t_k}^{t_{k-1}}\right)\left(\boldsymbol{z}_{t_k} - \hat{\boldsymbol{z}}_{t_k}^{t_{k-1}}\right)^T\right\} \\ &= \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\tilde{\boldsymbol{z}}_{t_k}^{t_{k-1}}\left(\tilde{\boldsymbol{z}}_{t_k}^{t_{k-1}}\right)^T\right\} \end{aligned} \tag{3.45}$$

  $$\boldsymbol{P}_{t_k}^{t_k} = \operatorname{cov}\left\{\hat{\boldsymbol{z}}_{t_k}^{t_k}, \hat{\boldsymbol{z}}_{t_k}^{t_k}\right\} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\tilde{\boldsymbol{z}}_{t_k}^{t_k}\left(\tilde{\boldsymbol{z}}_{t_k}^{t_k}\right)^T\right\} \tag{3.46}$$

  where $\tilde{\boldsymbol{z}}_{t_k}^{t_{k-1}}$ and $\tilde{\boldsymbol{z}}_{t_k}^{t_k}$ are called estimation error.

- Innovation: $\tilde{\xi}_{t_k}$

  $$\tilde{\xi}_{t_k} = \xi_{t_k} - \hat{\xi}_{t_k}^{t_{k-1}} \tag{3.47}$$

- Smoothed estimate, error covariance and lag-one covariance: $\hat{\boldsymbol{z}}_{t_k}^{t_N}, \boldsymbol{P}_{t_k}^{t_N}, \boldsymbol{P}_{t_{k+1},t_k}^{t_N}$

  $$\hat{\boldsymbol{z}}_{t_k}^{t_N} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_k}|\xi_{t_{1:N}}\right\} \tag{3.48}$$

  $$\boldsymbol{P}_{t_k}^{t_N} = \operatorname{cov}\left\{\hat{\boldsymbol{z}}_{t_k}^{t_N}, \hat{\boldsymbol{z}}_{t_k}^{t_N}\right\} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\tilde{\boldsymbol{z}}_{t_k}^{t_N}\left(\tilde{\boldsymbol{z}}_{t_k}^{t_N}\right)^T\right\} \tag{3.49}$$

  $$\boldsymbol{P}_{t_{k+1},t_k}^{t_N} = \operatorname{cov}\left\{\hat{\boldsymbol{z}}_{t_{k+1}}^{t_N}, \hat{\boldsymbol{z}}_{t_k}^{t_N}\right\} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\tilde{\boldsymbol{z}}_{t_{k+1}}^{t_N}\left(\tilde{\boldsymbol{z}}_{t_k}^{t_N}\right)^T\right\} \tag{3.50}$$

Lemma 3.3 (The Kalman filter)**.** *Assume the state-space model is given by (3.24), given the observations* $\{\xi_{t_{1:N}}\}$ *and the initial state*

$$\boldsymbol{z}_{t_1}^{t_1} \sim \mathcal{N}\left(\boldsymbol{\mu}_{t_1}, \boldsymbol{P}_{t_1}\right)$$

The equations of the Kalman filter are listed below

- *Time update:*

$$\hat{\boldsymbol{z}}_{t_{k+1}}^{t_k} = \boldsymbol{F}_{h_k} \hat{\boldsymbol{z}}_{t_k}^{t_k} \tag{3.51}$$
$$\boldsymbol{P}_{t_{k+1}}^{t_k} = \boldsymbol{F}_{h_k} \boldsymbol{P}_{t_k}^{t_k} \boldsymbol{F}_{h_k} + \boldsymbol{Q}_{h_k} \tag{3.52}$$

- *Measurement update:*

$$\hat{\boldsymbol{z}}_{t_k}^{t_k} = \hat{\boldsymbol{z}}_{t_k}^{t_{k-1}} + \boldsymbol{K}_{t_k} \tilde{\xi}_{t_k} \tag{3.53}$$
$$\boldsymbol{P}_{t_k}^{t_k} = \boldsymbol{P}_{t_k}^{t_{k-1}} - \boldsymbol{K}_{t_k} \boldsymbol{\Omega}_{t_k} \boldsymbol{K}_{t_k}^T \tag{3.54}$$

  where

$$\tilde{\xi}_{t_k} = \xi_{t_k} - \boldsymbol{H}_{h_k} \hat{\boldsymbol{z}}_{t_k}^{t_{k-1}} \tag{3.55}$$
$$\boldsymbol{K}_{t_k} = \boldsymbol{P}_{t_k}^{t_{k-1}} \boldsymbol{H}_{h_k}^T \boldsymbol{\Omega}_{t_k}^{-1} \tag{3.56}$$
$$\boldsymbol{\Omega}_{t_k} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\tilde{\xi}_{t_k} \tilde{\xi}_{t_k}^T\right\} = \boldsymbol{H}_{h_k} \boldsymbol{P}_{t_k}^{t_{k-1}} \boldsymbol{H}_{h_k}^T \tag{3.57}$$

$\boldsymbol{K}_{t_k}$ *is called Kalman gain.*

Proof. See *e.g.* [Sim06]. □

Lemma 3.4 (The Rauch–Tung–Striebel (RTS) smoother)**.** *Assume the state-space model is given by (3.24), given the observations* $\{\xi_{t_{1:N}}\}$ *and initial conditions* $\hat{\boldsymbol{z}}_{t_N}^{t_N}$, $P_{t_N}^{t_N}$, $\boldsymbol{P}_{t_{k_N}, t_{N-1}}^{t_N} = (\boldsymbol{I} - \boldsymbol{K}_{t_N} \boldsymbol{H}_c) \boldsymbol{F}_{h_{N-1}} \boldsymbol{P}_{t_{N-1}}^{t_{N-1}}$*, the RTS smoother is summarized as*

$$\hat{\boldsymbol{z}}_{t_k}^{t_N} = \hat{\boldsymbol{z}}_{t_k}^{t_k} + \boldsymbol{J}_{t_k}\left(\hat{\boldsymbol{z}}_{t_{k+1}}^{t_N} - \hat{\boldsymbol{z}}_{t_{k+1}}^{t_k}\right) \tag{3.58}$$
$$\boldsymbol{P}_{t_k}^{t_N} = \boldsymbol{P}_{t_k}^{t_k} + \boldsymbol{J}_{t_k}\left(\boldsymbol{P}_{t_{k+1}}^{t_N} - \boldsymbol{P}_{t_{k+1}}^{t_k}\right) \boldsymbol{J}_{t_k}^T \tag{3.59}$$
$$\boldsymbol{P}_{t_{k+1}, t_k}^{t_N} = \boldsymbol{P}_{t_{k+1}}^{t_{k+1}} \boldsymbol{J}_{t_k}^T + \boldsymbol{J}_{t_{k+1}}\left(\boldsymbol{P}_{t_{k+2}, t_{k+1}}^{t_N} - \boldsymbol{F}_{h_{k+1}} \boldsymbol{P}_{t_{k+1}}^{t_{k+1}}\right) \boldsymbol{J}_{t_k}^T \tag{3.60}$$

where

$$\boldsymbol{J}_{t_k} = \boldsymbol{P}_{t_k}^{t_k} \boldsymbol{F}_{h_k}^T \left(\boldsymbol{P}_{t_{k+1}}^{t_k}\right)^{-1} \tag{3.61}$$

Proof. See *e.g.* [RTS65, SS11]. □

3.3.4 Consistency

Amongst the properties of the estimated parameters, an interesting one is consistency, which indicates the estimates converge to the true values as the number of observations tends to infinity. In our context, a δ-operator model is used to approximate the CT model, so the concept of

asymptotic consistency is redefined to include that the maximum sampling interval $\bar{h}$ tends to zero (see *e.g.* [SCB92]).

Assuming that the E-step and the M-step of Algorithm 3.1 are given by Lemma 3.1 and Lemma 3.2, respectively. Then, the converged estimate by Algorithm 3.1 is asymptotically consistent

$$\lim_{N\to\infty}\lim_{\bar{h}\to 0}\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^o \tag{3.62}$$

3.4 Improved EM Algorithm for CAR Model Estimation

The EM algorithm presented in Section 3.3 is accurate for fast sampled data, but sometimes the sampling period might be not so small and the parameter estimates will be biased. Since the bias of the parameter estimates is determined by h, a straightforward way to reduce the bias is reducing h. However, the sampling interval is determined by the experimenter and cannot be changed afterward. In this thesis, we propose to interpolate the state to reduce the sampling interval. It seems that the main difficulty related to this idea is the state estimation at a time-instant with missing measurement.

This section employs an state interpolation technique to improve the accuracy of CT parameter estimates. Thus, we assume that $t_k < t_{k,1} < t_{k,2} < \cdots < t_{k,n_k} < t_{k+1}$ are n_k user-specified time-instants within the interval (t_k, t_{k+1}). Note that there are no measurements available at these interpolated time-instants. For the sake of consistency, let $t_{k,0} = t_k$, $t_{k,n_k+1} = t_{k+1}$, the interpolation interval is denoted as

$$h_{k,i} = t_{k,i+1} - t_{k,i}, \quad 0 \le i \le n_k$$

A new q-operator model can be obtained as

$$\begin{cases} \boldsymbol{z}_{t_{k,i+1}} = \boldsymbol{F}_{h_{k,i}}\boldsymbol{z}_{t_{k,i}} + \boldsymbol{w}_{h_{k,i}} \\ y_k = \boldsymbol{H}_c\boldsymbol{z}_{t_{k,0}} \end{cases} \tag{3.63}$$

where $\boldsymbol{z}_{t_{k,i}}$ denotes the state at time instant $t_{k,i}$, $\boldsymbol{w}_{h_{k,i}}$ is the new DT process noise having the covariance matrix $\boldsymbol{Q}_{h_{k,i}}$. Model (3.63) can be written into the δ-operator form (see [MG90])

$$\begin{cases} \boldsymbol{z}^{\delta}_{t_{k,i}} = \boldsymbol{F}^{\delta}_{h_{k,i}}\boldsymbol{z}_{t_{k,i}} + \boldsymbol{w}^{\delta}_{h_{k,i}} \\ y_k = \boldsymbol{H}_c\boldsymbol{z}_{t_{k,0}} \end{cases} \tag{3.64}$$

where

$$\begin{aligned} &\boldsymbol{z}^{\delta}_{t_{k,i}} = h^{-1}_{k,i}\left(\boldsymbol{z}_{t_{k,i+1}} - \boldsymbol{z}_{t_{k,i}}\right) \\ &\mathbb{E}\left\{\boldsymbol{w}^{\delta}_{h_{k,i}}(\boldsymbol{w}^{\delta}_{h_{k,i+\ell}})^T\right\} = h^{-1}_{k,i}\boldsymbol{Q}^{\delta}_{h_{k,i}}\delta_\ell \end{aligned}$$

The matrices $\boldsymbol{F}^{\delta}_{h_{k,i}}$, $\boldsymbol{Q}^{\delta}_{h_{k,i}}$ take the form of

$$\boldsymbol{F}^{\delta}_{h_{k,i}} = h^{-1}_{k,i}(\boldsymbol{F}_{h_{k,i}} - \boldsymbol{I}), \;\; \boldsymbol{Q}^{\delta}_{h_{k,i}} = h^{-1}_{k,i}\boldsymbol{Q}_{h_{k,i}} \tag{3.65}$$

From (3.65), it is easy to derive that the bias order is $O(h_{k,i})$ when $\boldsymbol{F}^{\delta}_{h_{k,i}}$, $\boldsymbol{Q}^{\delta}_{h_{k,i}}$ are used to approximate $\boldsymbol{F}_c$, $\boldsymbol{Q}_c$, it also implies that the bias can be reduced by choosing a smaller interpolation interval $h_{k,i}$.

3.4.1 The EM algorithm

The flexibility of using the EM algorithm is that the user can choose the *hidden variable*. In this subsection, by choosing $\boldsymbol{z}_{t_{1:N,1:n_k}}$ as another hidden variable, $l'(\boldsymbol{\theta})$ can be decomposed in the following way

$$\begin{aligned} l'(\boldsymbol{\theta}) =& \log p_{\boldsymbol{\theta}}\left(\boldsymbol{z}_{t_{1:N}}\right) = \log \frac{p_{\boldsymbol{\theta}}\left(\boldsymbol{z}_{t_{1:N}}, \boldsymbol{z}_{t_{1:N,1:n_k}}\right)}{p_{\boldsymbol{\theta}}\left(\boldsymbol{z}_{t_{1:N,1:n_k}} \middle| \boldsymbol{z}_{t_{1:N}}\right)} \\ =& \log p_{\boldsymbol{\theta}}\left(\boldsymbol{z}_{t_{1:N}}, \boldsymbol{z}_{t_{1:N,1:n_k}}\right) - \log p_{\boldsymbol{\theta}}\left(\boldsymbol{z}_{t_{1:N,1:n_k}} \middle| \boldsymbol{z}_{t_{1:N}}\right) \end{aligned} \tag{3.66}$$

Then

$$\mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{l'(\boldsymbol{\theta}) | \xi_{t_{1:N}}\right\} = \mathscr{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0) - \mathscr{H}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0) \tag{3.67}$$

where

$$\begin{aligned} \mathscr{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0) &= \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\log p_{\boldsymbol{\theta}}\left(\boldsymbol{z}_{t_{1:N}}, \boldsymbol{z}_{t_{1:N,1:n_k}}\right) \middle| \xi_{t_{1:N}}\right\} \\ \mathscr{H}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0) &= \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\log p_{\boldsymbol{\theta}}\left(\boldsymbol{z}_{t_{1:N,1:n_k}} \middle| \boldsymbol{z}_{t_{1:N}}\right) \middle| \xi_{t_{1:N}}\right\} \end{aligned}$$

The E-step and M-step are given below

- E-step: consider the model structures (3.63) and (3.64). Given the measurement $\xi_{t_{1:N}}$, the initial parameter vector $\hat{\boldsymbol{\theta}}^0$ and the initial state

  $$\boldsymbol{z}_{t_1} \sim \mathscr{N}\left(\boldsymbol{\mu}_{t_1}, P_{t_1}\right)$$

  The $\mathscr{Q}$-function containing $\boldsymbol{F}^{\delta}_{h_{k,i}}$ and $\boldsymbol{Q}^{\delta}_{h_{k,i}}$ is given by

  $$\begin{aligned} & -2\mathscr{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0) \\ & = \log\left|\boldsymbol{P}_{t_1}\right| + \operatorname{Tr}\left\{\boldsymbol{P}_{t_1}^{-1}\left(\boldsymbol{z}_{t_{t_1}} - \boldsymbol{\mu}_{t_1}\right)\left(\boldsymbol{z}_{t_{t_1}} - \boldsymbol{\mu}_{t_1}\right)^T\right\} + \sum_{k=1}^{N-1}\sum_{i=0}^{n_k} \log\left|h_{k,i}\boldsymbol{Q}^{\delta}_{h_{k,i}}\right| \\ & \quad + \sum_{k=1}^{N-1}\sum_{i=0}^{n_k} \operatorname{Tr}\left\{h_{k,i}\left(\boldsymbol{Q}^{\delta}_{h_{k,i}}\right)^{-1}\Big[\boldsymbol{\Phi}_{t_{k,i}} - \boldsymbol{\Psi}_{t_{k,i}}\left(\boldsymbol{F}^{\delta}_{h_{k,i}}\right)^T - \boldsymbol{F}^{\delta}_{h_{k,i}}\boldsymbol{\Psi}^T_{k,i} \right. \right. \\ & \quad \left. \left. + \boldsymbol{F}^{\delta}_{h_{k,i}}\boldsymbol{\Upsilon}_{t_{k,i}}\left(\boldsymbol{F}^{\delta}_{h_{k,i}}\right)^T\Big]\right\} \end{aligned} \tag{3.68}$$

  where

  $$\begin{aligned} \boldsymbol{\Phi}_{t_{k,i}} &= \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}^{\delta}_{t_{k,i}}(\boldsymbol{z}^{\delta}_{t_{k,i}})^T \middle| \xi_{t_{1:N}}\right\} \\ \boldsymbol{\Psi}_{t_{k,i}} &= \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}^{\delta}_{t_{k,i}}\boldsymbol{z}^T_{t_{k,i}} \middle| \xi_{t_{1:N}}\right\} \\ \boldsymbol{\Upsilon}_{t_{k,i}} &= \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_{k,i}}\boldsymbol{z}^T_{t_{k,i}} | \xi_{t_{1:N}}\right\} \end{aligned}$$

- M-step: the parameter estimate $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\eta}}; \hat{\sigma}^2]$ that maximize $\mathscr{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0)$ are given by

  $$\hat{\boldsymbol{\eta}}^T = -\sum_{k=1}^{N-1}\sum_{i=0}^{n_k} \boldsymbol{W}\boldsymbol{\Psi}_{t_{k,i}}\left(\sum_{k=1}^{N-1}\sum_{i=0}^{n_k}\boldsymbol{\Upsilon}_{t_{k,i}}\right)^{-1} \tag{3.69}$$

$$\hat{\sigma}^2 = \frac{1}{m}\sum_{k=1}^{N-1}\sum_{i=0}^{n_k} h_{k,i}\left[\boldsymbol{W}\boldsymbol{\Phi}_{t_{k,i}}\boldsymbol{W}^T - \hat{\boldsymbol{\eta}}^T\boldsymbol{\Upsilon}_{t_{k,i}}\hat{\boldsymbol{\eta}}\right] \tag{3.70}$$

where $m = \sum_{k=1}^{N-1} n_k$.

3.5 State estimation

$\boldsymbol{\Phi}_{t_{k,i}}$, $\boldsymbol{\Psi}_{t_{k,i}}$ and $\boldsymbol{\Upsilon}_{t_{k,i}}$ involved in $\mathscr{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0)$ contain the following two terms

$$\mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_{k,i}}\boldsymbol{z}_{t_{k,i}}^T\middle|\xi_{t_{1:N}}\right\} = \hat{\boldsymbol{z}}_{t_{k,i}}^{t_N}(\hat{\boldsymbol{z}}_{t_{k,i}}^{t_N})^T + \boldsymbol{P}_{t_{k,i}}^{t_N} \tag{3.71}$$

$$\mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_{k,i+1}}\boldsymbol{z}_{t_{k,i}}^T\middle|\xi_{t_{1:N}}\right\} = \hat{\boldsymbol{z}}_{t_{k,i+1}}^{t_N}(\hat{\boldsymbol{z}}_{t_{k,i}}^{t_N})^T + \boldsymbol{M}_{t_{k,i}}^{t_N} \tag{3.72}$$

where the smoothed state $\hat{\boldsymbol{z}}_{t_{k,i}}^{t_N}$, the error covariance matrix $\boldsymbol{P}_{t_{k,i}}^{t_N}$, and the lag-one error covariance $\boldsymbol{M}_{t_{k,i}}^{t_N}$ are defined below

$$\begin{aligned}\hat{\boldsymbol{z}}_{t_{k,i}}^{t_N} &= \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_{k,i}}|\xi_{t_{1:N}}\right\},\ \boldsymbol{P}_{t_{k,i}}^{t_N} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\tilde{\boldsymbol{z}}_{t_{k,i}}^{t_N}(\tilde{\boldsymbol{z}}_{t_{k,i}}^{t_N})^T\right\}\\ \boldsymbol{M}_{t_{k,i}}^{t_N} &= \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\tilde{\boldsymbol{z}}_{t_{k,i+1}}^{t_N}(\tilde{\boldsymbol{z}}_{t_{k,i}}^{t_N})^T\right\}\end{aligned}$$

where $\tilde{\boldsymbol{z}}_{t_{k,i}}^{t_N} = \boldsymbol{z}_{t_{k,i}} - \hat{\boldsymbol{z}}_{t_{k,i}}^{t_N}$.

At the sampling instant $\{t_k\}_{k=1}^{t_N}$ where the observations are available, $\hat{\boldsymbol{z}}_{t_k}^{t_N}$, $\boldsymbol{P}_{t_k}^{t_N}$ (or $\hat{\boldsymbol{z}}_{t_{k,0}}^{t_N}$, $\boldsymbol{P}_{k,0}^{t_N}$) can be estimated by using the well-known RTS smoother from the sampled output data. However, the estimation of $\hat{\boldsymbol{z}}_{t_{k,i}}^{t_N}$, $\boldsymbol{P}_{t_{k,i}}^{t_N}$ and $\boldsymbol{M}_{t_{k,i}}^{t_N}$ $(0 < i \leq n_k)$ are not considered in the standard RTS smoother, thus the RTS smoother is modified as follows

Proposition 3.5. *Assuming that the filtered states*

$$\hat{\boldsymbol{z}}_{t_k}^{t_k} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_k}|\xi_{t_{1:k}}\right\},\ \hat{\boldsymbol{z}}_{t_{k+1}}^{t_k} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_{k+1}}|\xi_{t_{1:k}}\right\} \tag{3.73}$$

and their error covariance matrices $\boldsymbol{P}_{t_k}^{t_k}, \boldsymbol{P}_{t_{k+1}}^{t_k}$ are obtained by using the Kalman filter. The smoothed state $\hat{\boldsymbol{z}}_{t_k}^{t_N}$ and its error covariance matrix $\boldsymbol{P}_{t_k}^{t_N}$ are obtained by using the RTS smoother. The smoothed state and its error covariance matrix at the time instant $t_{k,i}$ $(i \leq n_k)$ can be computed in the following way

1. *Prediction*
   The prediction $\hat{\boldsymbol{z}}_{t_{k,i}}^{t_k}$ and its error covariance matrix $\boldsymbol{P}_{t_{k,i}}^{t_k}$ are

$$\hat{\boldsymbol{z}}_{t_{k,i}}^{t_k} = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0}\left\{\boldsymbol{z}_{t_{k,i}}|\xi_{t_{1:k}}\right\} = \boldsymbol{F}_{\tau_{k,i}}\hat{\boldsymbol{z}}_{t_k}^{t_k} \tag{3.74}$$

$$\boldsymbol{P}_{t_{k,i}}^{t_k} = \boldsymbol{F}_{\tau_{k,i}}\boldsymbol{P}_{t_k}^{t_k}\boldsymbol{F}_{\tau_{k,i}}^T + \boldsymbol{Q}_{\tau_{k,i}} \tag{3.75}$$

   where $\tau_{k,i} = t_{k,i} - t_k$. $\boldsymbol{F}_{\tau_{k,i}}$, $\boldsymbol{Q}_{\tau_{k,i}}$ are parameterized by $\hat{\boldsymbol{\theta}}^0$ and $\tau_{k,i}$.

2. *Smoothing*
   The smoothed state $\hat{\boldsymbol{z}}_{t_{k,i}}^{t_N}$ and its error covariance matrix $\boldsymbol{P}_{t_{k,i}}^{t_N}$ are

$$\hat{\boldsymbol{z}}_{t_{k,i}}^{t_N} = \hat{\boldsymbol{z}}_{t_{k,i}}^{t_k} + \boldsymbol{J}_{t_{k,i}}\left(\hat{\boldsymbol{z}}_{t_{k+1}}^{t_N} - \hat{\boldsymbol{z}}_{t_{k+1}}^{t_k}\right) \tag{3.76}$$

$$\boldsymbol{P}_{t_{k,i}}^{t_N} = \boldsymbol{P}_{t_{k,i}}^{t_k} + \boldsymbol{J}_{t_{k,i}}\big(\boldsymbol{P}_{t_{k+1}}^{t_N} - \boldsymbol{P}_{t_{k+1}}^{t_k}\big)\boldsymbol{J}_{t_{k,i}}^T \tag{3.77}$$

where

$$\boldsymbol{J}_{t_{k,i}} = \big(\boldsymbol{F}_{\tau_{k,i}}\boldsymbol{P}_{t_k}^{t_k}\boldsymbol{F}_{h_k}^T + \boldsymbol{Q}_{\tau_{k,i}}\boldsymbol{F}_{h_k-\tau_{k,i}}^T\big)\big(\boldsymbol{P}_{t_{k+1}}^{t_k}\big)^{-1} \tag{3.78}$$

3. *Lag-one covariance*
   When $i < n_k$, the lag-one covariance $\boldsymbol{M}_{t_{k,i}}^{t_N}$ is

$$\begin{aligned}\boldsymbol{M}_{t_{k,i}}^{t_N} =& \boldsymbol{F}_{\tau_{k,i+1}}\boldsymbol{P}_{t_k}^{t_k}\boldsymbol{F}_{\tau_{k,i}}^T + \boldsymbol{F}_{h_{k,i}}\boldsymbol{Q}_{\tau_{k,i}} \\ &+ \boldsymbol{J}_{t_{k,i+1}}\big(\boldsymbol{P}_{t_{k+1}}^{t_N} - \boldsymbol{P}_{t_{k+1}}^{t_k}\big)\boldsymbol{J}_{t_{k,i}}^T\end{aligned} \tag{3.79}$$

   When $i = n_k$

$$\begin{aligned}\boldsymbol{M}_{t_{k,i}}^{t_N} =& \boldsymbol{F}_{h_k}\boldsymbol{P}_{t_k}^{t_k}\boldsymbol{F}_{\tau_{k,i}}^T + \boldsymbol{F}_{h_{k,i}}\boldsymbol{Q}_{\tau_{k,i}} \\ &+ \big(\boldsymbol{P}_{t_{k+1}}^{t_N} - \boldsymbol{P}_{t_{k+1}}^{t_k}\big)\boldsymbol{J}_{t_{k,i}}^T\end{aligned} \tag{3.80}$$

Proof. See Appendix C.3. □ □

There is a last question that how to build, for example $\boldsymbol{F}_h$ and $\boldsymbol{Q}_h$ from the previous estimate $\hat{\boldsymbol{\theta}} = \big[\hat{\boldsymbol{\eta}}; \hat{\sigma}^2\big] = \big[\hat{d}_1, \cdots, \hat{d}_n, \hat{\sigma}^2\big]^T$. Let

$$\boldsymbol{F}_h = \big[\boldsymbol{f}_h^1 \ \cdots \ \boldsymbol{f}_h^n\big]^T, \ \boldsymbol{Q}_h = [q_h^{ij}]_{i,j=1}^n$$

According to the relationships between the q-operator and δ-operator model matrices, $\boldsymbol{f}_h^1$ and q_h^{11} can be approximated as

$$\boldsymbol{f}_h^1 = \big[1 - \hat{d}_1 h, -\hat{d}_2 h, \cdots, -\hat{d}_n h\big]^T, \ q_h^{11} = \hat{\sigma}^2 h \tag{3.81}$$

Let

$$\bar{\boldsymbol{F}}_c = \begin{bmatrix} -\hat{d}_1 & \cdots & -\hat{d}_n \\ \boldsymbol{I}_{n-1} & & \boldsymbol{0} \end{bmatrix}, \ \bar{\boldsymbol{Q}}_c = \begin{bmatrix} \hat{\sigma}^2 & \\ & \boldsymbol{0} \end{bmatrix} \tag{3.82}$$

Compute

$$\bar{\boldsymbol{F}}_h = e^{\bar{\boldsymbol{F}}_c h}, \ \bar{\boldsymbol{Q}}_h = \int_0^h e^{\bar{\boldsymbol{F}}_c t}\boldsymbol{Q}_c e^{\bar{\boldsymbol{F}}_c^T t} dt \tag{3.83}$$

The remaining elements in $\boldsymbol{F}_h$ and $\boldsymbol{Q}_h$ are approximated by

$$\boldsymbol{f}_h^i = \bar{\boldsymbol{f}}_h^i, \ q_h^{ij} = \bar{q}_h^{ij} \tag{3.84}$$

3.6 Extension to CARMA Model Estimation

In this section the estimation of a CARMA process is considered

$$\left(p^{n_d} + d_1 p^{n_d-1} + \cdots + d_{n_d}\right) \xi_t = \left(p^{n_c} + c_1 p^{n_c-1} + \cdots + d_{n_c}\right) e_t \tag{3.85}$$

where e_t is a CT white noise with the intensity of σ^2. It is further assumed that $n_c < n_d$, thus ξ_t has finite variance and the instantaneous sampling scheme can be applied. The corresponding sampled data model for (3.85) is then given as

$$\begin{cases} \boldsymbol{z}_{t_{k+1}} = \boldsymbol{F}_{h_k} \boldsymbol{z}_{t_k} + \tilde{\boldsymbol{w}}_{t_k} \\ \xi_{t_k} = \boldsymbol{H}_c \boldsymbol{z}_{t_k} \end{cases} \tag{3.86}$$

where

$$\boldsymbol{F}_{h_k} = e^{\boldsymbol{F}_c h_k}, \ \boldsymbol{Q}_{h_k} = \mathbb{E}\left\{\tilde{\boldsymbol{w}}_{t_k} \tilde{\boldsymbol{w}}_{t_k}^T\right\} = \int_0^{h_k} e^{\boldsymbol{F}_c t} \boldsymbol{Q}_c e^{\boldsymbol{F}_c^T t} dt$$

$$\boldsymbol{F}_c = \begin{bmatrix} -d_1 & \cdots & \cdots & -d_{n_d} \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & 1 & 0 \end{bmatrix}, \boldsymbol{H}_c^T = \begin{bmatrix} \boldsymbol{0} \\ 1 \\ c_1 \\ \vdots \\ c_{n_c} \end{bmatrix}, \ \boldsymbol{Q}_c = \begin{bmatrix} \sigma^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

Since the measurement noise of a CARMA model is missing, it is impossible to estimate $\boldsymbol{H}_c$ by the EM algorithm. To solve this problem, a combined gradient and EM algorithm, termed as Grad-EM, is proposed. An additional gradient search for the unknown parameters in $\boldsymbol{H}_c$ is added to the EM iterations.

Model (3.86) has the following innovation form

$$\begin{cases} \hat{\boldsymbol{z}}_{t_{k+1}}^{t_k} = \boldsymbol{F}_{h_k} \hat{\boldsymbol{z}}_{t_k}^{t_{k-1}} + \boldsymbol{K}_{t_k} \epsilon_{t_k} \\ \xi_{t_k} = \boldsymbol{H}_c \hat{\boldsymbol{z}}_{t_k}^{t_{k-1}} + \epsilon_{t_k} \end{cases} \tag{3.87}$$

where ϵ_{t_k} is a white innovation sequence

$$\begin{aligned} &\epsilon_{t_k} \sim \mathcal{N}\left(0, \boldsymbol{\Omega}_{t_k}\right) \\ &\boldsymbol{\Omega}_{t_k} = \boldsymbol{H}_c \boldsymbol{P}_{t_k}^{t_{k-1}} \boldsymbol{H}_c^T \\ &\boldsymbol{\Lambda}_{t_k} = \boldsymbol{F}_{h_k} \boldsymbol{P}_{t_k}^{t_{k-1}} \boldsymbol{H}_c^T \\ &\boldsymbol{K}_{t_k} = \boldsymbol{\Lambda}_{t_k} \boldsymbol{\Omega}_{t_k}^{-1} \\ &\boldsymbol{P}_{t_{k+1}}^{t_k} = \boldsymbol{F}_{h_k} \boldsymbol{P}_{t_k}^{t_{k-1}} \boldsymbol{F}_{h_k}^T + \boldsymbol{Q}_{h_k} - \boldsymbol{K}_{t_k} \left(\boldsymbol{F}_{h_k} \boldsymbol{P}_{t_k}^{t_{k-1}} \boldsymbol{H}_c^T\right)^T \end{aligned}$$

Then the log-likelihood function defined in (3.17) can be equivalently formulated as (see *e.g.* [ATY+12, Lju99])

$$l(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{k=1}^{N} \left(\epsilon_{t_k}^T \boldsymbol{\Omega}_k^{-1} \epsilon_{t_k} + \log \det \boldsymbol{\Omega}_{t_k}\right) + \text{constant} \tag{3.88}$$

The gradient method for parameter estimation can be given by

$$\hat{\boldsymbol{\theta}}^{j+1} = \hat{\boldsymbol{\theta}}^{j} + \mu^{j} \left[\nabla^2 l\big(\hat{\boldsymbol{\theta}}^{j}\big)\right]^{-1} \nabla l\big(\hat{\boldsymbol{\theta}}^{j}\big) \tag{3.89}$$

where μ^j is the step length, $\nabla l\left(\boldsymbol{\theta}^j\right)$ and $\nabla^2 l\left(\boldsymbol{\theta}^j\right)$ are the gradient and the approximated Hessian matrix, respectively.

Provided that the initial value $\boldsymbol{z}_{t_1}$ is independent of $\boldsymbol{\theta}$. The derivatives with respect to $\boldsymbol{\theta}$ in the above equation can be computed recursively by

$$\frac{\partial \epsilon_{t_k}}{\partial \boldsymbol{\theta}_i} = -\frac{\partial \boldsymbol{H}_c}{\partial \boldsymbol{\theta}_i}\hat{\boldsymbol{z}}_{t_k}^{t_{k-1}} - \boldsymbol{H}_c \frac{\partial \hat{\boldsymbol{z}}_{t_k}^{t_{k-1}}}{\partial \boldsymbol{\theta}_i} \tag{3.90}$$

$$\frac{\partial \hat{\boldsymbol{z}}_{t_{k+1}}^{t_k}}{\partial \boldsymbol{\theta}_i} = \frac{\partial \boldsymbol{F}_{h_k}}{\partial \boldsymbol{\theta}_i}\hat{\boldsymbol{z}}_{t_k}^{t_{k-1}} + \boldsymbol{F}_{h_k}\frac{\partial \hat{\boldsymbol{z}}_{t_k}^{t_{k-1}}}{\partial \boldsymbol{\theta}_i} + \frac{\partial \boldsymbol{K}_{t_k}}{\partial \boldsymbol{\theta}_i}\epsilon_{t_k} + \boldsymbol{K}_{t_k}\frac{\partial \epsilon_{t_k}}{\partial \boldsymbol{\theta}_i} \tag{3.91}$$

$$\frac{\partial \boldsymbol{\Omega}_{t_k}}{\partial \boldsymbol{\theta}_i} = \frac{\partial \boldsymbol{H}_c}{\partial \boldsymbol{\theta}_i}\boldsymbol{P}_{t_k}^{t_{k-1}}\boldsymbol{H}_c^T + \boldsymbol{H}_c\frac{\partial \boldsymbol{P}_{t_k}^{t_{k-1}}}{\partial \boldsymbol{\theta}_i}\boldsymbol{H}_c^T + \boldsymbol{H}_c\boldsymbol{P}_{t_k}^{t_{k-1}}\frac{\partial \boldsymbol{H}_c^T}{\partial \boldsymbol{\theta}_i} \tag{3.92}$$

$$\frac{\partial \boldsymbol{\Lambda}_{t_k}}{\partial \boldsymbol{\theta}_i} = \frac{\partial \boldsymbol{F}_{h_k}}{\partial \boldsymbol{\theta}_i}\boldsymbol{P}_{t_k}^{t_{k-1}}\boldsymbol{H}_c^T + \boldsymbol{F}_{h_k}\frac{\partial \boldsymbol{P}_{t_k}^{t_{k-1}}}{\partial \boldsymbol{\theta}_i}\boldsymbol{H}_c^T + \boldsymbol{F}_{h_k}\boldsymbol{P}_{t_k}^{t_{k-1}}\frac{\partial \boldsymbol{H}_c^T}{\partial \boldsymbol{\theta}_i} \tag{3.93}$$

$$\frac{\partial \boldsymbol{K}_{t_k}}{\partial \boldsymbol{\theta}_i} = \frac{\partial \boldsymbol{\Lambda}_{t_k}}{\partial \boldsymbol{\theta}_i}\boldsymbol{\Omega}_{t_k}^{-1} + \boldsymbol{\Lambda}_{t_k}\boldsymbol{\Omega}_{t_k}^{-1}\frac{\partial \boldsymbol{\Omega}_{t_k}}{\partial \boldsymbol{\theta}_i}\boldsymbol{\Omega}_{t_k}^{-1} \tag{3.94}$$

$$\begin{aligned}\frac{\partial \boldsymbol{P}_{t_{k+1}}^{t_k}}{\partial \boldsymbol{\theta}_i} =& \frac{\partial \boldsymbol{F}_{h_k}}{\partial \boldsymbol{\theta}_i}\boldsymbol{P}_{t_k}^{t_{k-1}}\boldsymbol{F}_{h_k}^T + \boldsymbol{F}_{h_k}\frac{\partial \boldsymbol{P}_{t_k}^{t_{k-1}}}{\partial \boldsymbol{\theta}_i}\boldsymbol{F}_{h_k}^T + \boldsymbol{F}_{h_k}\boldsymbol{P}_{t_k}^{t_{k-1}}\frac{\partial \boldsymbol{F}_{h_k}^T}{\partial \boldsymbol{\theta}_i} \\ &- \frac{\partial \boldsymbol{K}_{t_k}}{\partial \boldsymbol{\theta}_i}\boldsymbol{\Lambda}_{t_k}^T - \boldsymbol{K}_{t_k}\frac{\partial \boldsymbol{\Lambda}_{t_k}^T}{\partial \boldsymbol{\theta}_i} + \frac{\partial \boldsymbol{Q}_{h_k}}{\partial \boldsymbol{\theta}_i}\end{aligned} \tag{3.95}$$

Let $\boldsymbol{\theta}_F$ and $\boldsymbol{\theta}_H$ denote the unknown parameters in $\boldsymbol{F}_c$ and $\boldsymbol{H}_c$, respectively. The Grad-EM algorithm consists of two steps

1. We estimate $\boldsymbol{\theta}_H$ by the traditional gradient algorithm with frozen $\boldsymbol{\theta}_F$, and choosing the step size μ^j such that the log-likelihood function increases,
2. We estimate $\boldsymbol{\theta}_F$ using the EM algorithm with frozen $\boldsymbol{\theta}_H$, this step is the same as estimating a CAR model.

Algorithm 3.2 (The Grad-EM algorithm)**.**

---

1. *Initialization: Set the initial guess* $\hat{\boldsymbol{\theta}}^0 = \left\{\hat{\boldsymbol{\theta}}_F^0, \hat{\sigma}_0^2, \hat{\boldsymbol{\theta}}_H^0\right\}$
2. *Iteration:*
   for j=0:convergence
   (a) Estimate $\boldsymbol{\theta}_H$ *using numerical search method*

$$\hat{\boldsymbol{\theta}}_H^{j+1} = \hat{\boldsymbol{\theta}}_H^{j} + \mu^{j}\left[\nabla^2 l\big(\hat{\boldsymbol{\theta}}_H^{j}\big)\right]^{-1}\nabla l\big(\hat{\boldsymbol{\theta}}_H^{j}\big) \tag{3.96}$$

   (b) Estimate $\boldsymbol{\theta}_F$ *and* σ^2 *using the EM algorithm*

- *E-step*

$$\mathscr{Q}(\boldsymbol{\theta}_F, \sigma^2, \hat{\boldsymbol{\theta}}_F^j, \hat{\sigma}_j^2, \hat{\boldsymbol{\theta}}_H^{j+1}) = \mathbb{E}\big\{l(\boldsymbol{\theta}_F, \sigma^2)|\xi_{t_{1:N}}, \hat{\boldsymbol{\theta}}_F^j, \hat{\boldsymbol{\theta}}_H^{j+1}, \hat{\sigma}_j^2\big\}$$

- *M-step*

$$\big\{\hat{\boldsymbol{\theta}}_F^{j+1}, \hat{\sigma}_{j+1}^2\big\} = \arg \max_{\boldsymbol{\theta}_F, \sigma^2} \mathscr{Q}(\boldsymbol{\theta}_F, \sigma^2, \hat{\boldsymbol{\theta}}_F^j, \hat{\sigma}_j^2, \hat{\boldsymbol{\theta}}_H^{j+1})$$

(c) Form $\hat{\boldsymbol{\theta}}^{j+1} = \big\{\hat{\boldsymbol{\theta}}_F^{j+1}, \hat{\sigma}_{j+1}^2, \hat{\boldsymbol{\theta}}_C^{j+1}\big\}$

end

3.7 Numerical Examples

3.7.1 CAR Model Estimation from Irregularly Sampled Data

Consider the following CAR process

$$\left(p^2 + d_1^o p + d_2^o\right) \xi_t = e_t \tag{3.97}$$

where $d_1^o = 2, d_2^o = 2$, e_t has the intensity of $\sigma_o^2 = 1$. This model was also used in [SCB92, FSM+99, Pha00, LS02].

In the beginning, a simple example is given to learn the difficulty to identify the given system. It would be straightforward that the CAR process parameters are easy or hard to estimate if the relationships between $l(\boldsymbol{\theta})$ and $\boldsymbol{\theta}$ can be obtained. Let us assume that

- ξ_t is instantaneously sampled for $N = 5000$ times, h_k is uniformly distributed in the following interval

$$h_k \sim U\left[0.01, 0.09\right] s$$

- d_1, d_2 pick the following values sequentially

$$d_1 = \{0.1, 0.2, \cdots, 3.9\}\,,\ d_2 = \{0.1, 0.2, \cdots, 3.9\}$$

  σ^2 is fixed to 1.

The surface plot of $l(\boldsymbol{\theta})$ as a function of d_1, d_2 is shown in Fig. 3.2. It is obvious from this figure that $l(\boldsymbol{\theta})$ is convex over the given area, any initialization $\left(\hat{d}_1^0, \hat{d}_2^0\right) \in [0.1, 3.9]$ has a large possibility to converge to the global maximum.

Based on the above results, the CAR estimation problems for different sampling situations are considered in the following.

In this section, when the state interpolation technique is applied, the interval is chosen that $h_{k,i} = \lambda$ $(i < n_k)$, $h_{k,n_k} = h_k - n_k\lambda < \lambda$, here, $\lambda = 0.01$ is chosen. The state interpolation will not be performed by setting $n_k = 0$. The simulation is assumed to be carried out under the following assumptions

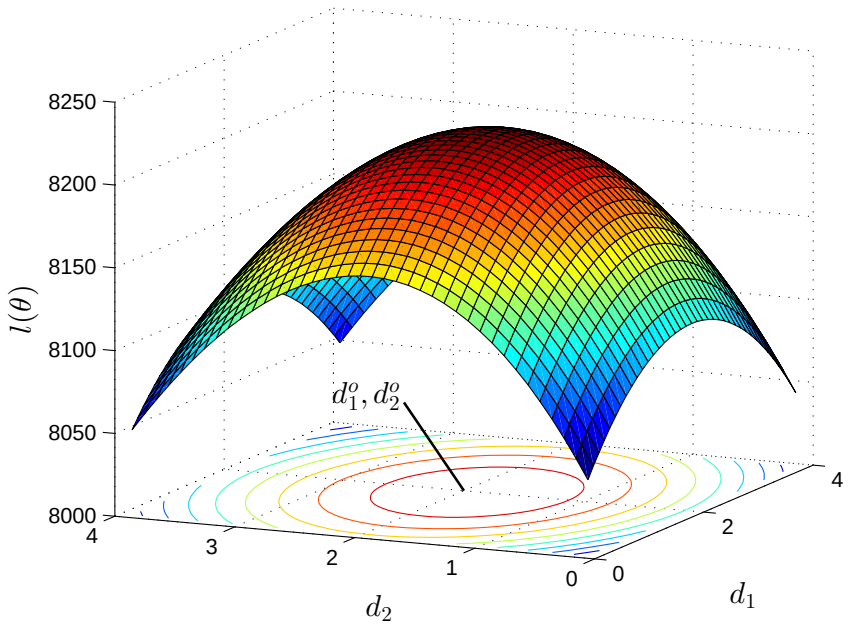

Figure 3.2: $l(\boldsymbol{\theta})$ as a function of the two parameters of the CAR process.

- h_k is uniformly distributed in the following interval

  $$h_k \sim U\left[\underline{h}, \bar{h}\right]$$

  where $\underline{h} = 0.01$, $\bar{h}$ increases from 0.01 to 0.09 with the increment of 0.01.

- Monte-Carlo simulations with 100 realizations are performed. Each realization has $N = 10000$ observations.

- For each run, $\hat{d}_1^0, \hat{d}_2^0$ are initialized as random numbers, $\hat{\sigma}_0^2$ is initialized by a fixed value.

  $$\hat{d}_1^0 \sim U\left[0.1, 3.9\right],\ \hat{d}_2^0 \sim U\left[0.1, 3.9\right],\ \hat{\sigma}_0^2 = 0.8$$

- The EM algorithm will be terminated once the maximum relative change of $\hat{\boldsymbol{\theta}}$ is smaller that 0.01%.

Two groups of results are obtained and plotted in Fig. 3.3. One of them uses the state interpolation technique to reduce the bias, whose results are plotted in sub-figures $(b), (d), (f), (h)$, the other does not use state interpolation. It is shown in this figure that the refinement brought by interpolation is significant. Several conclusions can also be drawn below:

1. Without interpolation, the bias of the parameter estimates increases as $\bar{h}$ increases. This phenomenon is shown in $(a), (c), (g)$.
2. With state interpolation, the bias of the parameter estimates keeps constant as $\bar{h}$ increases. This phenomenon is shown in $(b), (d), (f)$.
3. The standard deviation of the parameter estimates decreases as $\bar{h}$ increases. It is true that the time span of the sampled data increases when $\bar{h}$ becomes larger, which implies

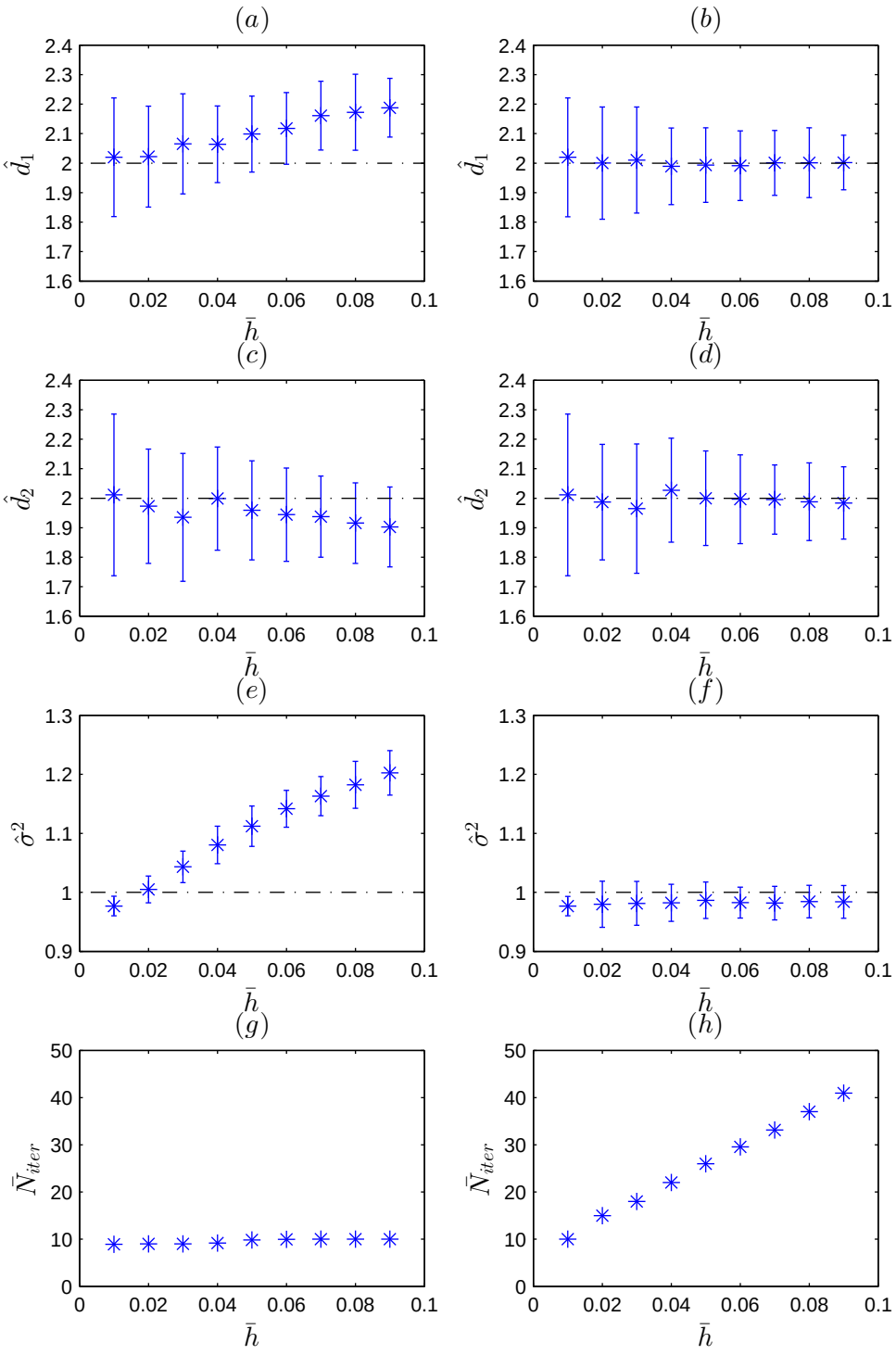

Figure 3.3: Parameter estimates by using the EM algorithm from irregularly sampled data. $(a), (c), (e), (g)$ – Estimation without state interpolation. $(b), (d), (f), (h)$ – Estimation with state interpolation. $\bar{N}_{iter}$ – Averaged number of iterations for convergence. $(*)$ – The mean values of the estimates. Vertical bar – The standard deviations of the estimates.

that the sampled data contains more information of the system dynamics over a longer period. The decrease in standard deviation is thus reasonable.

The use of state interpolation greatly improves the accuracy of the parameter estimates. However, this improvement does not come for free, the trade-off is the slowdown of the convergence rate. This is clear by comparing sub-figures (g) and (h) of Fig. 3.3, the number of iterations $\bar{N}_{iter}$ increases linearly as $\bar{h}$ becomes larger when state interpolation is applied.

3.7.2 The impact of the interpolation interval

In the previous section, one question remains still unclear, namely how does λ affect the performance of the proposed estimator? The previous simulation examples may give us some hints, that is smaller values of λ produce better approximation at the price of reducing the convergence rate, and vice versa. This fact will be made clearer in this section. Consider the same system in (3.97) and make the following assumptions

- Monte-Carlo simulation with 100 runs is performed. Each realization has $N = 10000$ observations, with a constant sampling period $h = 0.2$ (here the regular sampling assumption is used since we want to study how the interpolation interval affects the convergence region).
- Initial values $\hat{d}_1^0, \hat{d}_2^0$ and $\hat{\sigma}_0^2$ are chosen as

$$\hat{d}_1^0 = \hat{d}_2^0 = 1,\ \hat{\sigma}_0^2 = 0.5$$

- λ picks the following values sequentially

$$\lambda = \{0.05, 0.11, 0.15, 0.19\}$$

- For each λ, the averaged likelihood function value at i-th iteration, denoted by $\bar{l}(\boldsymbol{\theta}^i)$, is computed and stored.

The averaged log-likelihood function $\bar{l}(\boldsymbol{\theta}^i)$ for each λ is plotted in Figure 3.4. When λ has a larger value, $\bar{l}(\boldsymbol{\theta})$ increases more quickly during the first few iterations (see the dotted-dark and dashed-magenta curves). However, the convergence rate is reduced when λ turns smaller (see the solid-red and dashed-dot blue curves). The benefit of reducing λ is the increase of $\bar{l}(\boldsymbol{\theta})$ (or equivalently, the increase of the estimation accuracy). This fact can be justified by comparing the steady value of $\bar{l}(\boldsymbol{\theta})$. In Figure 3.4, we can see that the solid-red line ($\lambda = 0.05$) has the largest likelihood after 8-th iteration.

Another question is how small λ can reach, can the proposed method work for a arbitrary small value of λ? We can answer yes theoretically and in practice we must say no. Because of the computational error, the interpolation does not make sense when λ is too small.

3.7.3 CARMA Model Estimation from Irregularly Sampled Data

Consider the following CARMA process

$$\left(p^2 + d_1 p + d_2\right) \xi_t = \left(p + c_1\right) e_t \tag{3.98}$$

with $d_1 = 1, d_2 = 2, c_1 = 5$, e_t has the intensity $\sigma^2 = 1$.

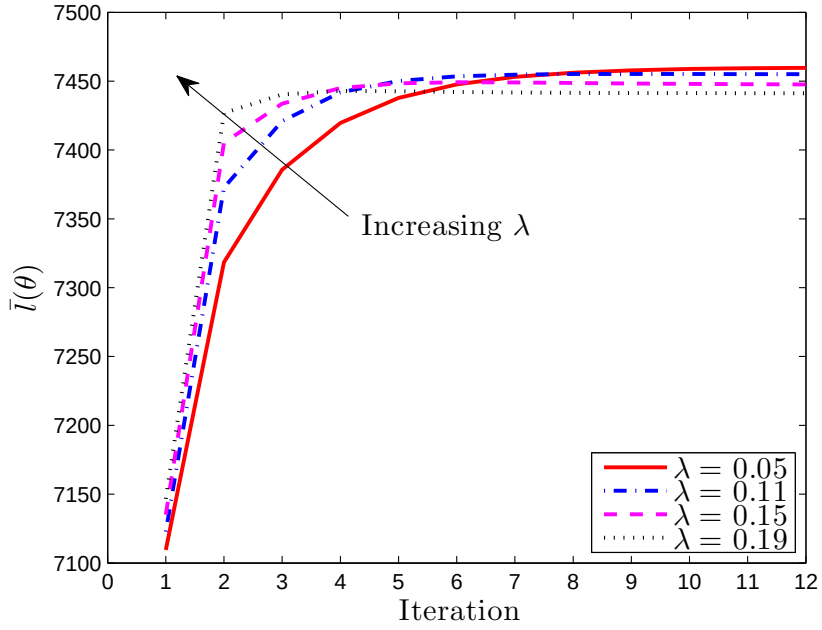

Figure 3.4: The influence of change λ on the performance of the proposed method.

- ξ_t is instantaneously observed for $N = 10,000$ times, the sampling period h_k uniformly distributes in the following interval

  $$h_k \sim U\left[0.01, \bar{h}\right] s$$

  where $\bar{h}$ increases gradually from $0.01s$ to $0.1s$ with step size $0.01s$.

- Monte-Carlo simulation with 100 realizations is performed, the mean and variance of estimated parameters are computed from these realizations.

- The initial conditions for the EM algorithm are

  $$\hat{d}_1^0 \sim (0,3],\ \hat{d}_2^0 \sim U(0,3],\ \hat{c}_1^0 \sim U[0,10],\ \hat{\sigma}^0 = 0.5$$

  The stopping rule for the EM algorithm is

  $$\max\left\{\left|\frac{\hat{\boldsymbol{\theta}}^{j+1} - \hat{\boldsymbol{\theta}}^j}{\hat{\boldsymbol{\theta}}^j}\right|\right\} < 10^{-4}$$

The Monte-Carlos simulation results are plotted in Figure 3.5.

3.8 Conclusion

This chapter has addressed the problem of CAR/CARMA model estimation from non-uniformly sampled data. The proposed method is a EM-based method, which uses a δ-operator model to approximate the CT model. The proposed method works quite well for fast sampled

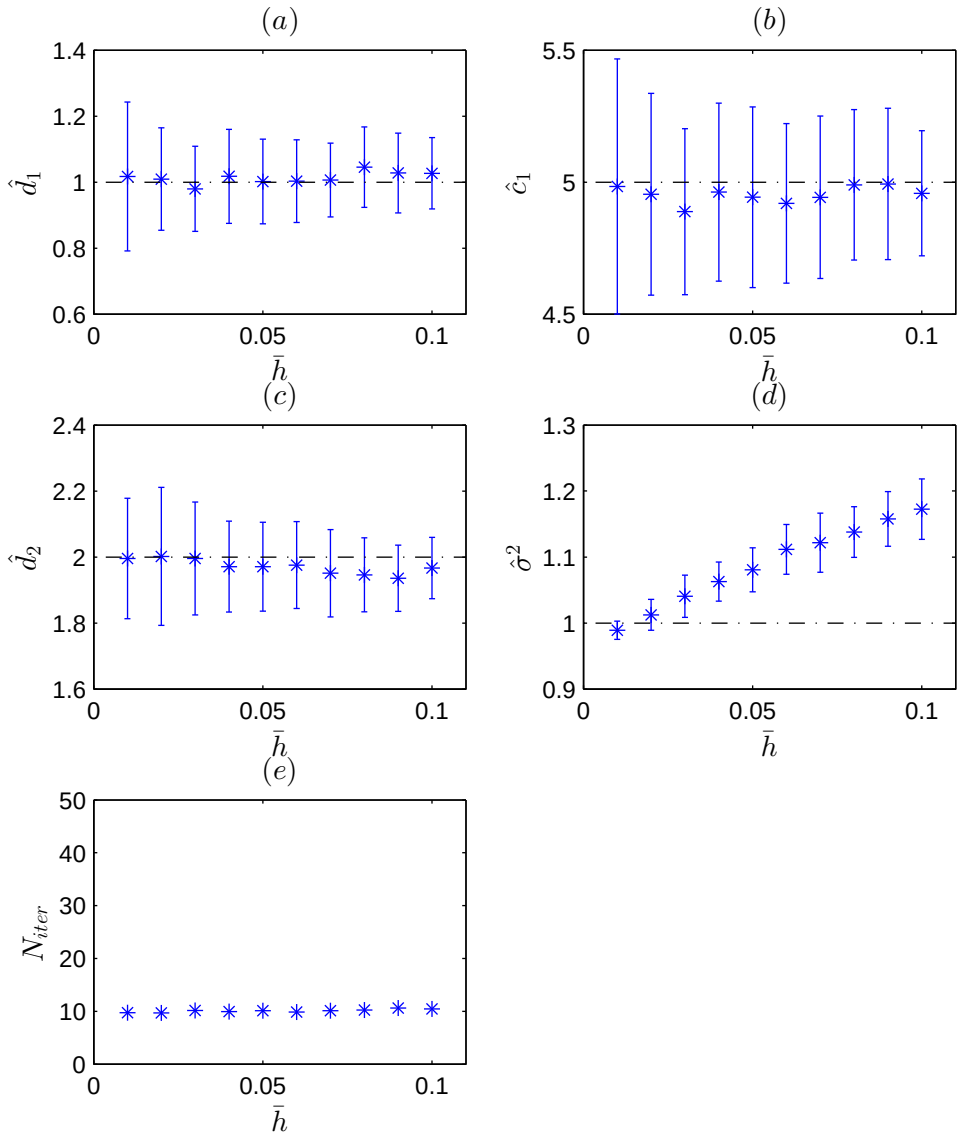

Figure 3.5: Monte-Carlos simulation results for the Grad-EM algorithm.

data. In the case of moderate sampling intervals, a state interpolation technique has been introduced to improve the accuracy of the estimator. This state interpolation technique requires to evaluate the state at a time-instant with missing measurement. Therefore the Kalman filter and Rauch-Tung-Striebel smoother have been re-derived for this purpose. It was shown that the choice of interpolation interval λ provides a freedom to choose between convergence rate and estimation accuracy. More precisely, the accuracy can be improved by slowing the convergence rate, and vice versa. Due to the fact that the proposed EM algorithm cannot estimate MA parameters, an additional numerical search was appended to the the EM algorithm to estimate the MA parameters. This yields the Grad-EM algorithm for CARMA model estimation.

3.9 Relevant Matlab Routines

1. `rts`: the Rauch-Tung-Striebel smoother for CAR/CARMA model.

   Syntax: `[xs,Ps,Ms]=rts(Fc,Hc,Qc,y,t)`

   - `Fc,Hc,Qc` are the CT state-space model, y and t are the sampled output and the sampling instants.
   - `xs,Ps,Ms` are the smoothed state, error covariance and the lag-one error covariance (3D matrices).

2. `rts_intpl`: RTS smoother with state interpolation

   Syntax: `[xs,Ps,Ms,ts]=rts_intpl(Fd,Hc,Qd,y,t,lambda)`

   - `Fd,Hd,Qd,Sd,Rd`: the δ-operator model $\boldsymbol{F}_\lambda^\delta$, $\boldsymbol{H}_c$, $\boldsymbol{Q}_\lambda^\delta$, `lambda`: the interpolation interval.
   - `ts`: the interpolated sampling-instants.

3. `car_em`: the EM algorithm for CAR model estimation without state interpolation.

   Syntax: `[theta,niter]=car_em(y,t,nd)`

   - `y,t,nd`: the sampled data, the sampling instants and the degree of the denominator.
   - `theta,niter`: estimated parameters and the iteration number.

4. `car_em_intpl`: the EM algorithm for CAR model estimation with state interpolation.

   Syntax: `[theta,niter]=car_em_intpl(y,t,nd,lambda)`

5. `carma_gradem`: the GRAD-EM algorithm for CARMA model estimation without state interpolation.

   Syntax: `[theta,niter]=carma_gradem(y,t,[nd,nc])`

Chapter 4

Identification of CT Box-Jenkins Models from Irregularly Sampled Data

Identification of COE models from irregularly sampled data was reviewed in Section 1.3, where the additive measurement noise is assumed to be white and the iterative SRIVC method is used to compute the optimal estimates. However, most practical systems do not conform the white measurement noise assumption, therefore, in this chapter, we consider a more general case when the measured output signal is corrupted by an additive colored measurement noise. CT model identification from irregularly sampled data has been recently considered from several different perspectives. In [TFCB97], derivative approximation methods were proposed to identify CARX models. In [Joh09], identification of continuous-time models, which are composed of an input-output model and a stochastic innovations model, from irregularly sampled input-output data sequences was studied. More recently, in [GC12], the minimum distortion filtering was applied to the problem of system identification when the data is collected with a irregular sampling period. The EM algorithm was used in [YAAG11] to estimate a CT state space model from irregular fast sampled data, a so called incremental model was used to approximate a CT model when the sampling period is small.

This chapter focuses on the use of the optimal IV methods for transfer function model identification, thus we use the CT Box-Jenkins (CBJ) models to represent dynamical systems. One main advantage of using the Box-Jenkins model is that the plant and noise models are independently parameterized (or equivalently, the denominators of the plant and noise models have no common roots), therefore the separable procedure can be applied in parameter estimation.

Another interesting problem considered here is closed-loop system identification from irregularly sampled data. The existence of the feedback, which induces the correlation between the disturbances and the control signal, raises a difficulty in system identification. Although the RIVC method has already been extended to handle closed-loop system identification (see *e.g.* [GGYdH08]) and shows some advantages, it cannot handle irregularly sampled data because of the DT noise model assumption. Therefore the second goal of this chapter is to extend the closed-loop RIVC (CLRIVC) method developed in [GGYdH08] to handle irregularly sampled data.

This chapter is composed of two sections: the first section considers the problem of open-loop system identification, the RIVC-based method (called RIVC-EM) for CBJ model estimation is

presented; the second section considers closed-loop system identification, the CLRIVC method is extended to handle irregular sampled data in presence of colored measurement noise.

4.1 Open-loop System Identification

4.1.1 Problem Statement

Consider the following CT system (see also Section 1.4.1)

$$\begin{cases} x_t = G(p, \boldsymbol{\rho}^o)u_t = \dfrac{B(p, \boldsymbol{\rho}^o)}{A(p, \boldsymbol{\rho}^o)}u_t \\ \xi_t = H(p, \boldsymbol{\eta}^o)e_t = \dfrac{C(p, \boldsymbol{\eta}^o)}{D(p, \boldsymbol{\eta}^o)}e_t \\ y_t = x_t + \xi_t \end{cases} \tag{4.1}$$

where u_t, x_t are the excitation signal and the noise-free response, respectively. e_t is a Gaussian CT white noise having the following covariance function

$$\mathbb{E}\left\{e_t e_{t+\tau}\right\} = \sigma_o^2 \delta_\tau$$

where δ_τ is the Dirac's delta function. $[\cdot]^o$ denotes the true parameter. The plant $G(p, \boldsymbol{\rho}^o)$ and the noise $H(p, \boldsymbol{\eta}^o)$ models are stable and have no common factors. It is further assumed that $H(p, \boldsymbol{\eta}^o)$ is invertibly stable. The polynomials $B(p, \boldsymbol{\rho}^o), A(p, \boldsymbol{\rho}^o), C(p, \boldsymbol{\eta}^o)$ and $D(p, \boldsymbol{\eta}^o)$ take the following structure

$$B(p, \boldsymbol{\eta}^o) = \sum_{i=0}^{n_b} b_i p^{n_b - i},\ A(p, \boldsymbol{\rho}^o) = p^{n_a} + \sum_{i=1}^{n_a} a_i p^{n_a - i}$$

$$C(p, \boldsymbol{\eta}^o) = p^{n_c} + \sum_{i=1}^{n_c} c_i p^{n_c - i},\ D(p, \boldsymbol{\rho}^o) = p^{n_d} + \sum_{i=1}^{n_d} d_i p^{n_d - i}$$

where $n_a \geq n_b$. For the sake of simplicity, it is assumed that ξ_t has finite variance so that the instantaneous sampling scheme (see Section 2.3) can be applied, therefore the degrees of $D(p, \boldsymbol{\rho}^o)$ and $C(p, \boldsymbol{\rho})$ are assumed that $n_d > n_c$. The unknown parameters are stacked in the column vectors

$$\boldsymbol{\theta}^o = \left[\boldsymbol{\rho}^o;\ \boldsymbol{\eta}^o;\ \sigma_o^2\right] \tag{4.2}$$

$$\boldsymbol{\rho}^o = \left[a_1^o\ \cdots\ a_{n_a}^o\ b_0^o\ \cdots\ b_{n_b}^o\right]^T \in \mathbb{R}^{n_a + n_b + 1} \tag{4.3}$$

$$\boldsymbol{\eta}^o = \left[d_1^o\ \cdots\ d_{n_d}^o\ c_1^o\ \cdots\ c_{n_c}^o\right]^T \in \mathbb{R}^{n_d + n_c} \tag{4.4}$$

The input and output signals u_t, y_t are instantaneously sampled at irregular time-instant t_k, for $k = 1, 2, \cdots, N$, and this gives rise to u_{t_k}, y_{t_k}. The time varying sampling interval is denoted as

$$h_k = t_{k+1} - t_k \tag{4.5}$$

The identification objective is to estimate the unknown parameters in $\{\boldsymbol{\rho}^o, \boldsymbol{\eta}^o, \sigma_o^2\}$ from the irregularly sampled input and output data $\{u_{t_k}, y_{t_k}\}_{k=1}^N$.

4.1.2 The Optimal IV Estimator

In this section the main conditions for obtaining optimal (consistent and minimum variance) IV parameter estimates are recalled. Similar to the SRIVC method, let $F(p)$ be a stable filter, a CBJ model has the following alternative regression form

$$y_{F,t_k} = \boldsymbol{\phi}_{F,t_k}^T \boldsymbol{\rho} + \varepsilon_{F,t_k} \tag{4.6}$$

with

$$\boldsymbol{\phi}_{F,t_k}^T = \left[-y_{F,t_k}^{(n_a-1)} \ \cdots \ -y_{F,t_k} \ u_{F,t_k}^{(n_b)} \ \cdots \ u_{F,t_k} \right]$$

where $[\cdot]_F$ denotes the filtering operation, *i.e.* $[\cdot]_F = F(p)[\cdot]$. The IV estimator is given by

$$\hat{\boldsymbol{\rho}} = \left(\sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k} \boldsymbol{\phi}_{F,t_k}^T \right)^{-1} \sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k} y_{F,t_k}^{(n_a)} \tag{4.7}$$

where $\boldsymbol{\psi}_{F,t_k}$ is the instrumental variable. It has been shown that a minimum variance estimator is achieved under the following conditions [YJ80] (see also [SS83]):

$$\begin{cases} \boldsymbol{\psi}_{F,t_k}^{\text{opt}} = F^{\text{opt}}(p) \mathring{\boldsymbol{\phi}}_{t_k} \\ F^{\text{opt}}(p) = \dfrac{1}{H(p,\boldsymbol{\eta}^o) A(p,\boldsymbol{\rho}^o)} \end{cases} \tag{4.8}$$

where $\mathring{\boldsymbol{\phi}}_{t_k}$ is the noise-free version of $\boldsymbol{\phi}_{t_k}$

$$\mathring{\boldsymbol{\phi}}_{t_k}^T = \left[-x_{t_k}^{(n_a-1)} \ \cdots \ -x_{t_k} \ u_{t_k}^{(n_b)} \ \cdots \ u_{t_k} \right] \tag{4.9}$$

Since the optimal conditions need the true parameters in $\boldsymbol{\rho}^o, \boldsymbol{\eta}^o$, the RIVC algorithm makes use of an iterative way to compute the optimal estimate. The combined RIVC and EM (RIVC-EM) algorithm is summarized below.

Algorithm 4.1 (The RIVC-EM algorithm)**.**

Step 1. Initialization*: Specify the initial guesses of the noise model $\hat{\boldsymbol{\eta}}^0, \hat{\sigma}^0$, the cutoff frequency ω_c^{SVF}. Apply the LSSVF algorithm to compute $\hat{\boldsymbol{\rho}}^0$.*

Step 2. Iterative IV estimation with prefilters*:*

$for\ j = 1 : convergence$

(1) If the estimated plant model of $(j-1)$th iteration is unstable, reflect the unstable zeros of the estimated polynomial $A(p, \hat{\boldsymbol{\rho}}^{j-1})$ into the stable region of the complex plane. Generate the noise response by using the 'auxiliary model'

$$x_{t_k}(\hat{\boldsymbol{\rho}}^{j-1}) = G(p, \hat{\boldsymbol{\rho}}^{j-1}) u_{t_k}$$

(2) Obtain the latest estimate $\{\hat{\boldsymbol{\eta}}^j, \hat{\sigma}^j\}$ *of the noise model parameters from the estimated noise sequence* $\xi_{t_k}(\hat{\boldsymbol{\rho}}^{j-1})$ *by the EM algorithm introduced in Chapter 3*

$$\xi_{t_k}(\hat{\boldsymbol{\rho}}^{j-1}) = y_{t_k} - x_{t_k}(\hat{\boldsymbol{\rho}}^{j-1}) \tag{4.10}$$

(3) Prefilter u_{t_k}, y_{t_k}, $x_{t_k}(\hat{\boldsymbol{\rho}}^{j-1})$ *and their time-derivatives by the filter*

$$F(p, \hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j) = \frac{1}{A(p, \hat{\boldsymbol{\rho}}^{j-1})H(p, \hat{\boldsymbol{\eta}}^j)}$$

(4) Based on these prefiltered data, compute a new estimate $\hat{\boldsymbol{\rho}}^j$ *of the plant model parameter vector*

$$\hat{\boldsymbol{\rho}}^j = \left[\sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k}(\hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j)\boldsymbol{\phi}^T_{F,t_k}(\hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j)\right]^{-1} \sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k}(\hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j) y^{(n_a)}_{F,t_k}(\hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j) \tag{4.11}$$

where $\boldsymbol{\phi}_{F,t_k}(\hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j)$, $\boldsymbol{\psi}_{F,t_k}(\hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j)$ *are defined in* (4.6) *and* (4.8) *but where a dependency to the noise model parameter estimate* $\hat{\boldsymbol{\eta}}^j$ *should be made clear.*

end

The block diagram of the RIVC-EM algorithm is given in Figure 4.1.

4.1.3 Implementation of the CT Filtering Operation

In the filtering operation, the applied input of the filter is a reconstructed 'CT signal' (see Figure 4.2), which is built from the measurements $\{u_{t_k}, y_{t_k}\}_{k=1}^N$ and the assumed inter-sample

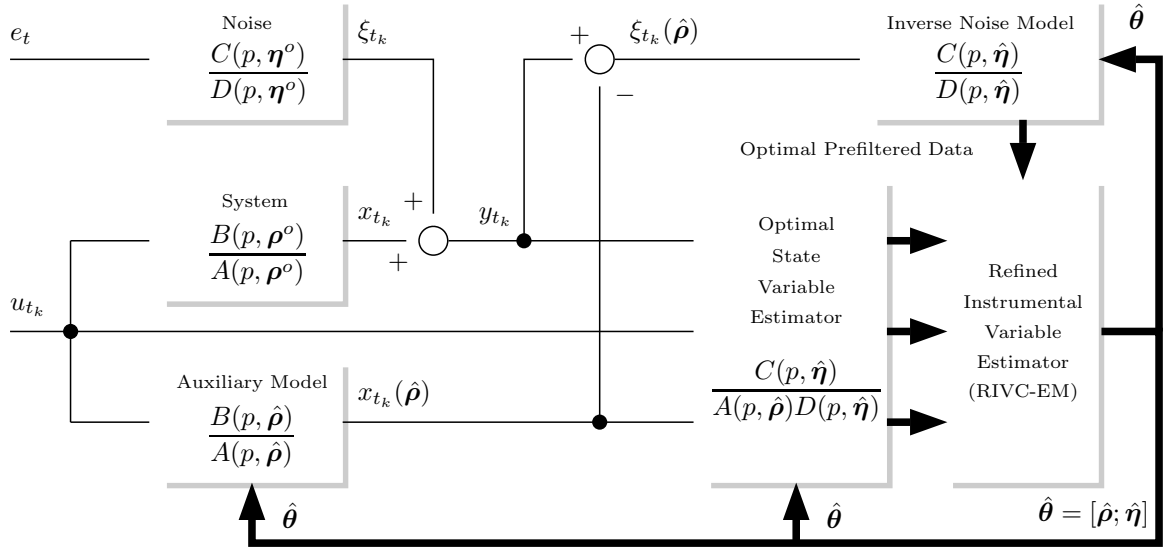

Figure 4.1: The Block Diagram of the RIVC-EM algorithm (see [You11]).

behavior. In many applications, especially in computer controlled systems, the inter-sample behavior of u_t is always known, such as zero-order-hold (ZOH) or first-order-hold (FOH), therefore u_t can always be exactly reconstructed. Unfortunately, the inter-sample behavior of y_t is not exactly known. In this case, it is reasonable to assume that y_t is FOH. More details on the digital implementation to compute filtered derivatives from sampled data are given in Appendix D.

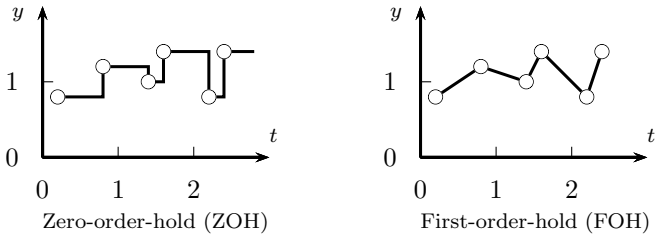

Figure 4.2: Reconstruction of CT signals from the sampled data for two intersample behavior assumptions.

4.1.4 Numerical Examples

Consider the following CT system

$$\mathscr{S}\begin{cases} x_t = \dfrac{b_0^o p + b_1^o}{p^2 + a_1^o p + a_2^o} u_t \\ \xi_t = \dfrac{p + c_1^o}{p^2 + d_1^o p + d_2^o} e_t \\ y_{t_k} = x_{t_k} + \xi_{t_k} \end{cases} \tag{4.12}$$

where $a_1^o = 2.8, a_2^o = 4, b_0^o = 1, b_1^o = 3, d_1^o = 1, d_2^o = 2, c_1^o = 5, \sigma_o^2 = 1.8 \times 10^{-3}$.

The Simulation is carried out under the following conditions:

1. u_t is chosen to be a pseudo-random binary sequence (PRBS) that is generated from a shift register with 9 stages. The synchronizing clock period is $1s$.
2. The sampling period h_k is uniformly distributed over the following interval

$$h_k \sim U[0.01, 0.09]s \tag{4.13}$$

3. Monte Carlo simulations (MCS) with 100 realizations are performed. Each realization has $N = 10,000$ observations. A portion of the measured input-output data is plotted in Figure 4.3, where the uniform distribution of the sampling interval is shown in the top sub-figure.
4. The cutoff frequency ω_c^{SVF} of the initial state variable filter is set to $\omega_c^{\text{SVF}} = 0.6\pi$.
5. The initial values of d_1, d_2, c_1 are random numbers uniformly distributed in the following intervals, the initial guess of $\hat{\sigma}^0$ is assumed to be a fixed value.

$$\hat{d}_1^0, \hat{d}_2^0 \sim U\left[0, 3\right],\ \hat{c}_1^0 \sim U\left[0, 10\right],\ \hat{\sigma}_0^2 = 1 \tag{4.14}$$

6. The stopping rule is chosen to be

$$\max\left\{\left|\frac{\hat{\boldsymbol{\rho}}^{i+1}-\hat{\boldsymbol{\rho}}^{i}}{\hat{\boldsymbol{\rho}}^{i}}\right|\right\}\leq 10^{-4} \tag{4.15}$$

	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_0$	$\hat{b}_1$	$\hat{d}_1$	$\hat{d}_2$	$\hat{c}_1$	$\hat{\sigma}^2$	$\bar{N}_{iter}$
True	2.8	4	1	3	1	2	5	1.8×10^{-3}	
SRIVC	2.8006 $\pm$0.1588	3.9849 $\pm$0.3405	1.0017 $\pm$0.0319	2.9961 $\pm$0.2644	——	——	——	——	3.7
RIVC-EM	2.8032 $\pm$0.0918	4.0075 $\pm$0.1472	1.0001 $\pm$0.0083	3.0079 $\pm$0.1277	1.0632 $\pm$0.1319	1.9657 $\pm$0.1061	5.0888 $\pm$0.3253	2.1×10^{-3} $\pm7.5\times10^{-5}$	12.1

Table 4.1: Mean and standard deviation of the parameter estimates. $\hat{\sigma}^2$ - Estimated intensity of the CT white noise. $\bar{N}_{iter}$ - Averaged number of iterations for convergence.

The estimated parameters are presented in Table 4.1. The SRIVC method from the CONTSID toolbox[1] is run here to give a comparison. As it can be seen from this table, both the SRIVC and the RIVC-EM methods are unbiased. However, because of the mis-specification of the noise model, the variance of the SRIVC estimates is not optimal. The proposed RIVC-EM algorithm takes the merits of the noise modeling, so it reduces the variance of the estimates. However, this improvement is at the expense of increasing computational load. As it can be observed from this table, the RIVC-EM method needs on average 12 iterations to converge. The Bode plots of the 100 RIVC-EM model estimates are also plotted in Figure 4.4.

[1] see http://www.cran.uhp-nancy.fr/contsid/

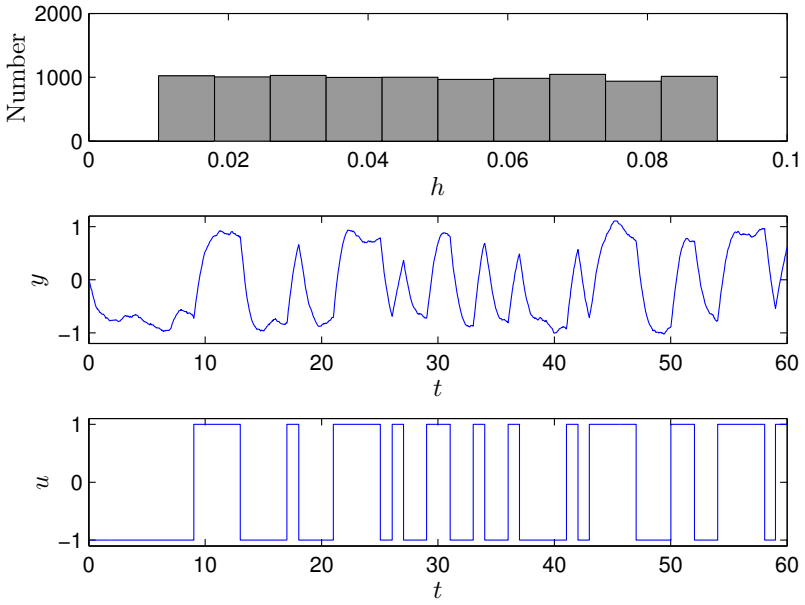

Figure 4.3: A portion of the sampled data. Top - The distribution of the sampling interval. Middle - The sampled output. Bottom - The sampled input.

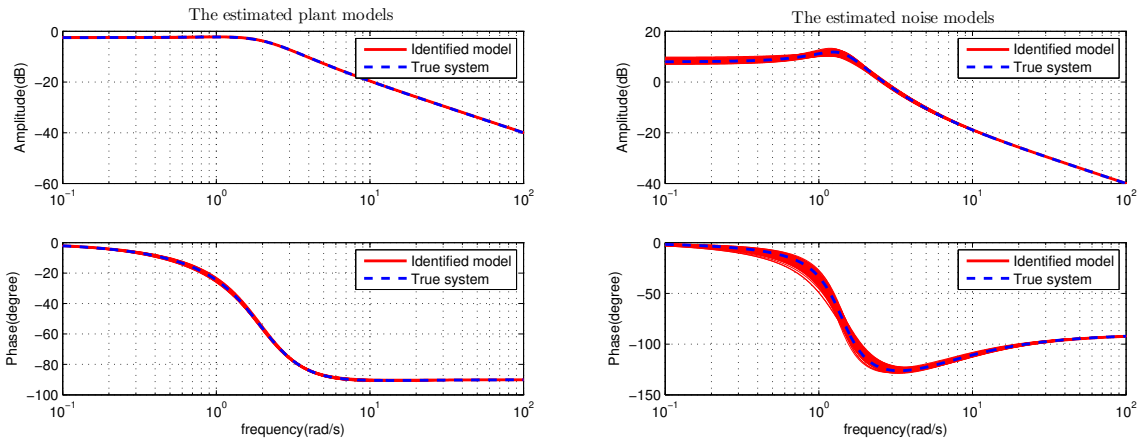

Figure 4.4: Bode plots of the RIVC-EM estimates with the true system

4.2 Closed-loop System Identification

4.2.1 Problem Statement

The block diagram of a linear SISO closed-loop system is shown in Figure 4.5, where the process, the noise and the controller are denoted by $G(p, \boldsymbol{\rho}^o)$, $H(p, \boldsymbol{\eta}^o)$ and $C_c(p)$, respectively. u_t and y_t are the input and the output signals,

$$r_t = r_{1,t} + C_c(p) r_{2,t} \tag{4.16}$$

is the external signal. ξ_t is the additive colored noise. e_t is assumed to be a Gaussian CT white noise and has the intensity σ_o^2

$$\mathbb{E}\{e_t e_\tau\} = \sigma_o^2 \delta_{t-\tau} \tag{4.17}$$

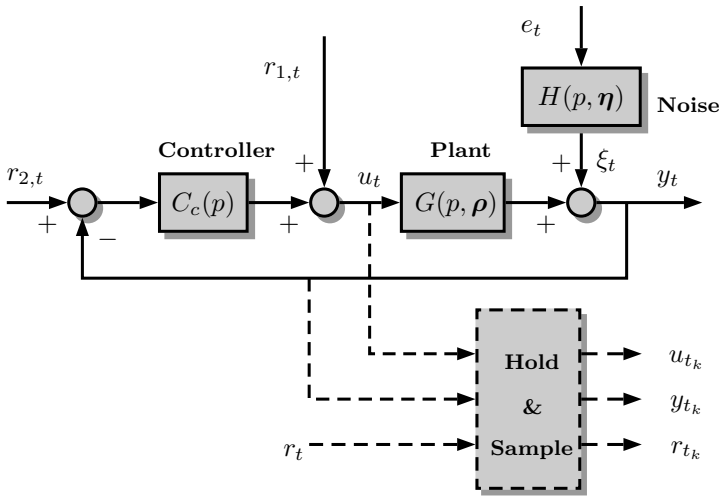

Figure 4.5: The block diagram of a closed-loop system

where δ_t is the Dirac's delta function. With all the notations defined above, the CT system can be given as (see [GGYdH08])

$$\mathscr{S}: \begin{cases} y_t = G(p, \boldsymbol{\rho}^o)u_t + H(p, \boldsymbol{\eta}^o)e_t \\ u_t = r_t - C_c(p)y_t \end{cases} \tag{4.18}$$

where $G(p, \boldsymbol{\rho}^o), H(p, \boldsymbol{\eta}^o), \boldsymbol{\rho}^o, \boldsymbol{\eta}^o$ are defined in the same way as those in the open-loop system (4.1). $C_c(p)$ is assumed to be *a priori* known. Then, system (4.18) can be rewritten into the following equivalent form with respect to r_t and e_t

$$\begin{cases} y_t = \dfrac{G(p, \boldsymbol{\rho}^o)}{1 + G(p, \boldsymbol{\rho}^o)C_c(p)}r_t + \dfrac{H(p, \boldsymbol{\eta}^o)}{1 + G(p, \boldsymbol{\rho}^o)C_c(p)}e_t \\ u_t = \dfrac{1}{1 + G(p, \boldsymbol{\rho}^o)C_c(p)}r_t - \dfrac{C_c(p)H(p, \boldsymbol{\eta}^o)}{1 + G(p, \boldsymbol{\rho}^o)C_c(p)}e_t \end{cases} \tag{4.19}$$

The noise-free input-output terms of (4.19) are given by

$$\begin{cases} x_t = \dfrac{G(p, \boldsymbol{\rho}^o)}{1 + G(p, \boldsymbol{\rho}^o)C_c(p)}r_t \\ v_t = \dfrac{1}{1 + G(p, \boldsymbol{\rho}^o)C_c(p)}r_t \end{cases} \tag{4.20}$$

y_t, u_t and r_t are assumed to be sampled at irregular time-instant t_k, for $k = 1, 2, \cdots, N$. The sampled data are denoted by y_{t_k}, u_{t_k} and r_{t_k}, respectively. The identification objective is to estimate $\{\boldsymbol{\rho}^o, \boldsymbol{\eta}^o, \sigma^o\}$ from the sampled data set $\{y_{t_{1:N}}, u_{t_{1:N}}, r_{t_{1:N}}\}$.

4.2.2 Instrumental Variable Estimator

A closed-loop system can be estimated in the same way as an open-loop system if u_{t_k} and y_{t_k} are available. However, because of the feedback, u_{t_k} is also corrupted by a colored noise, even though the assumed additive noise is white. Therefore the open-loop approach cannot produce the optimal estimates. The following IV-based method is presented to yield consistent estimates. From (4.18), the closed-loop system has also the regression form

$$y_{F,t_k} = \boldsymbol{\phi}_{F,t_k}^T \boldsymbol{\rho} + \varepsilon_{F,t_k} \tag{4.21}$$

with

$$\boldsymbol{\phi}_{F,t_k}^T = \left[-y_{F,t_k}^{(n_a-1)} \ \cdots \ -y_{F,t_k} \ u_{F,t_k}^{(n_b)} \ \cdots \ u_{F,t_k}\right]$$

where $[\cdot]_F$ denotes the filtering operation. Subsequently, the IV estimator is given by

$$\hat{\boldsymbol{\rho}} = \left(\sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k}\boldsymbol{\phi}_{F,t_k}^T\right)^{-1} \sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k} y_{F,t_k}^{(n_a)} \tag{4.22}$$

where $\boldsymbol{\psi}_{F,t_k}$ is the instrument.

It has been shown that a minimum variance estimator is achieved under the following conditions (see *e.g.* [GGYV08])

$$\begin{cases} \boldsymbol{\psi}_{F,t_k}^{\text{opt}} = F^{\text{opt}}(p)\mathring{\boldsymbol{\phi}}_{t_k} \\ F^{\text{opt}}(p) = \dfrac{1}{H(p,\boldsymbol{\eta}^o)A(p,\boldsymbol{\rho}^o)} \end{cases} \tag{4.23}$$

where $\mathring{\boldsymbol{\phi}}_{t_k}$ is the noise-free version of $\boldsymbol{\phi}_{t_k}$

$$\mathring{\boldsymbol{\phi}}_{t_k}^T = \big[-x_{t_k}^{(n_a-1)} \ \cdots \ -x_{t_k} \ v_{t_k}^{(n_b)} \ \cdots \ v_{t_k} \big] \tag{4.24}$$

where x_{t_k}, v_{t_k} are given in (4.20).

The combined CLRIVC and EM (CLRIVC-EM) algorithm for closed-loop system identification is summarized below

Algorithm 4.2 (The CLRIVC-EM algorithm)**.**

---

1. **Initialization***: Initialize $F(p)$ with the cutoff frequency of ω_c^{SVF}. Specify the initial estimates $\{\hat{\boldsymbol{\eta}}^0, \sigma^0\}$ using the open-loop data $\{u_{t_k}, y_{t_k}\}_{k=1}^N$ and compute $\hat{\boldsymbol{\rho}}^0$.*

2. **Iterative IV estimation with prefilters***:*

   $for\ j = 1 : convergence$

   (1) If the estimated plant model is unstable, reflect the unstable zeros of $A(p, \hat{\boldsymbol{\rho}}^{j-1})$ into the stable region of the complex plane. Generate the noise-free signals $x_{t_k}\big(\hat{\boldsymbol{\rho}}^{j-1}\big)$ and $v_{t_k}\big(\hat{\boldsymbol{\rho}}^{j-1}\big)$ with the 'auxiliary model' $G\big(p, \hat{\boldsymbol{\rho}}^{j-1}\big)$ built from the previously estimated parameter vector $\hat{\boldsymbol{\rho}}^{j-1}$

$$x_{t_k}\big(\hat{\boldsymbol{\rho}}^{j-1}\big) = \frac{G\big(p, \hat{\boldsymbol{\rho}}^{j-1}\big)}{1 + G\big(p, \hat{\boldsymbol{\rho}}^{j-1}\big)C_c(p)} r_{t_k}$$
$$v_{t_k}\big(\hat{\boldsymbol{\rho}}^{j-1}\big) = \frac{1}{1 + G\big(p, \hat{\boldsymbol{\rho}}^{j-1}\big)C_c(p)} r_{t_k}$$

   (2) Obtain the latest estimate $\{\hat{\boldsymbol{\eta}}^j, \hat{\sigma}^j\}$ based on the estimated noise sequence using the proposed EM algorithm.

$$\hat{\xi}_{t_k} = \big(1 + G\big(p, \hat{\boldsymbol{\rho}}^{j-1}\big)C_c(p)\big)\ \big(y_{t_k} - x_{t_k}\big(\hat{\boldsymbol{\rho}}^{j-1}\big)\big)$$

   (3) Prefilter the input u_{t_k}, the output y_{t_k} and the estimated noise-free response $x_{t_k}(\hat{\boldsymbol{\rho}}^{j-1})$, $v_{t_k}(\hat{\boldsymbol{\rho}}^{j-1})$ by

$$F(p, \hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j) = \frac{1}{A(p, \hat{\boldsymbol{\rho}}^{j-1})H(p, \hat{\boldsymbol{\eta}}^j)} \tag{4.25}$$

   with the estimated parameter vector $\hat{\boldsymbol{\rho}}^{j-1}$ obtained at the previous iteration and $\hat{\boldsymbol{\eta}}^j$ obtained in Step (2).

(4) Based on these prefiltered signals, compute a new estimate $\hat{\boldsymbol{\rho}}^j$

$$\hat{\boldsymbol{\rho}}^j = \left[\sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k}(\hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j)\boldsymbol{\phi}^T_{F,t_k}(\hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j)\right]^{-1} \times \left[\sum_{k=1}^{N} \boldsymbol{\psi}_{F,t_k}(\hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j) y^{(n_a)}_{F,t_k}(\hat{\boldsymbol{\rho}}^{j-1}, \hat{\boldsymbol{\eta}}^j)\right]$$

end

4.2.3 Numerical Example

Consider the following two systems

$$\mathscr{S}_1 : \begin{cases} G(p, \boldsymbol{\rho}^o) = \dfrac{3}{p+2} \\ H(p, \boldsymbol{\eta}^o) = \dfrac{1}{p+0.5} \\ C_c(p) = 1 \\ \sigma^2 = 5.7 \times 10^{-3} \end{cases}, \mathscr{S}_2 : \begin{cases} G(p, \boldsymbol{\rho}^o) = \dfrac{p+3}{p^2+2.8p+4} \\ H(p, \boldsymbol{\eta}^o) = \dfrac{1}{p+1} \\ C_c(p) = 1 \\ \sigma^2 = 2.4 \times 10^{-3} \end{cases} \tag{4.26}$$

The external signal r_t is chosen to be a PRBS, which is generated from a shift register with 9 stages with the synchronizing clock of $0.2s$. Moreover, the signal-to-noise ratio (SNR) is approximately 20 dB for both examples, where the SNR is defined as

$$\text{SNR} = 10 \log_{10}\left(\frac{P_x}{P_e}\right)$$

where P_x denotes the power of the noise-free signal x_t.

The following sampling scheme is considered

- First, the true system is regularly sampled with the sampling period of $\Delta = 0.0001s$, the resultant sampled output is denoted by Z^M.
- Then, we draw a irregular sequence from the regularly sampled sequence Z^M, the down sampling period is given by

  $$h_k = t_{k+1} - t_k = n_k \Delta$$

  where n_k is a regularly distributed random integer

  $$n_k \in [10, 50]$$

  Consequently, the sampling period h_k satisfies that

  $$h_k \sim U[0.001, 0.005]s$$

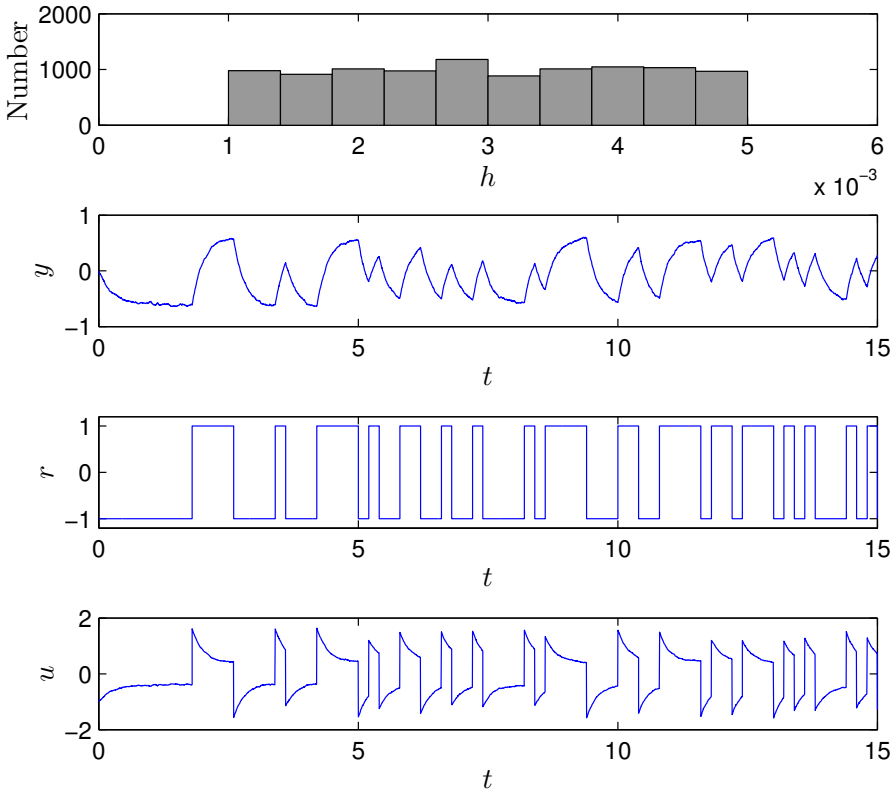

Figure 4.6: A portion of the input-output data ($\mathscr{S}_1$).

- Finally, the irregularly sampled data is collected in Z^N

The estimation procedure is carried out under the following conditions

- The length of the sampled data is $N = 10,000$.
- Monte-Carlo simulation (MCS) of 100 runs is performed, the mean value and standard deviation of the parameter estimates are computed from these realizations.
- The stopping rule for the CLRIVC-EM algorithm is

$$\max\left\{\left|\frac{\hat{\boldsymbol{\theta}}^{i+1}-\hat{\boldsymbol{\theta}}^{i}}{\hat{\boldsymbol{\theta}}^{i}}\right|\right\} \leq 10^{-4} \tag{4.27}$$

- The cutoff frequency is chosen as $\omega_c^{\text{SVF}} = 0.6\pi$. The initial guess of σ is $\hat{\sigma}_0^2 = 1$.

A portion of the sampled input-output data of $\mathscr{S}_1$ is plotted in Figure 4.6. The distribution of h_k is shown in the top sub-figure.

In order to make this example more persuasive, we applied two different estimation scheme: the first is the CLSRIVC algorithm (see [GGYdH08]), the second is the proposed CLRIVC-EM algorithm. In the case of $\mathscr{S}_2$, the Bode plots of the estimated models (CLRIVC-EM algorithm) along with the true system are displayed in Fig. 4.7. The mean values and standard deviations of the parameter estimates are gathered in Table 4.2.

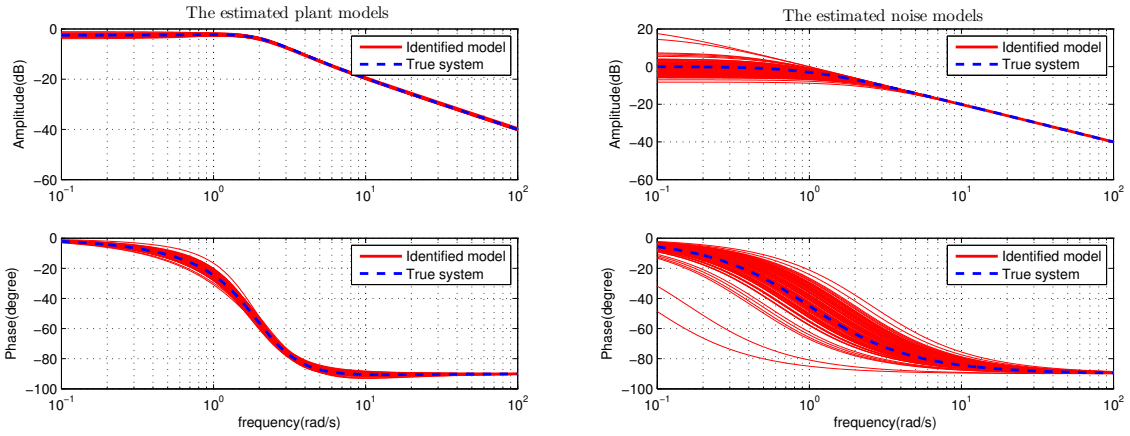

Figure 4.7: The Bode plots of the 100 estimated plant and noise models ($\mathscr{S}_2$) by CLRIVC-EM algorithm with the true system.

		$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_0$	$\hat{b}_1$	$\hat{d}_1$	$\hat{\sigma}^2$	$\bar{N}_{iter}$
System	True value	2	—	3	—	0.5	5.7332×10^{-3}	
$\mathscr{S}_1$	CLSRIVC	1.9649 $\pm$0.1092	—	2.9781 $\pm$0.0437	—	—	—	3.5
	CLRIVC-EM	1.9943 $\pm$0.0576	—	2.9770 $\pm$0.0207	—	0.5634 $\pm$0.3047	7.6268×10^{-3} $\pm2.1757\times10^{-3}$	3.1
		$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_0$	$\hat{b}_1$	$\hat{d}_1$	$\hat{\sigma}^2$	N_{iter}
System	True value	2.8	4	1	3	1	2.3810×10^{-3}	
$\mathscr{S}_2$	CLSRIVC	2.8492 $\pm$0.3636	4.0941 $\pm$0.5617	0.9902 $\pm$0.0310	3.0937 $\pm$0.5185	—	—	6.2
	CLRIVC-EM	2.7616 $\pm$0.2724	3.9541 $\pm$0.3747	0.9934 $\pm$0.0141	2.9413 $\pm$0.3540	1.1367 $\pm$0.4275	2.9444×10^{-3} $\pm5.0692\times10^{-4}$	5.3

Table 4.2: Mean and standard deviation of the parameter estimates. $\bar{N}_{\text{iter}}$ – Number of iterations for convergence.

From Table 4.2, we can see that the mean values of the estimated parameters by using the CLRIVC-EM and CLSRIVC algorithms are very close. This is normal, because of the consistency of the IV-based methods. However, since the noise is mis-specified, the variance of CLSRIVC estimates is not minimized. It can be seen from Table 4.2 that the CLRIVC-EM algorithm delivers unbiased estimates with smaller variance than the usual CLSRIVC algorithm.

4.3 Conclusion

The issue of CT Box-Jenkins model identification from irregularly sampled data has been considered in this chapter. Refined instrumental variable-based method has been developed to solve this problem. This approach was soon extended to handle CT closed-loop system identification. The performance of the proposed scheme has been investigated by means of simulation examples.

4.4 Relevant Matlab Routines

1. `rivc_em`: RIVC-EM algorithm for CBJ model estimation.

   Syntax: `[theta,eta,niter,sigma2]=rivc_em(data,nn,fcsvf,eta0)`

   - `data` is an `iddata` object containing $\{y_{t_k}, u_{t_k}\}_{k=1}^{N}$.
   - `nn`=$[n_a, n_b, n_c, n_d]$ specifies the degrees of the polynomials.
   - `fcsvf`=f_c^{SVF}, `eta0`=$\hat{\boldsymbol{\eta}}^0$ are user provided initial values.
   - The returned values are $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}, N_{iter}, \hat{\sigma}^2$

2. `simc_fw`: digital simulation of the filtering operation.

   `y=simc_fw(A,B,u,t,'intersample','zoh')`

   $$y_t = \frac{B(p)}{A(p)} u_t$$

   - u, t provides the sampled data and sampling instants.
   - Intersample behavior can also be chosen as `'foh'`.

3. `clrivc_em`: CLRIVC-EM algorithm for closed-loop system identification.

   Syntax: `[theta,eta,niter,sigma2]=clrivc_em(dataol,datacl,nn,Con,fcsvf)`

   - `dataol`, `datacl` are the open-loop and closed-loop data, `Con` is the controller $C_c(p)$
   - Other variables are defined the same as those in `rivc_em`.

Chapter 5

Identification of CT Systems with Time-delay from Irregularly Sampled Data

Many practical systems, such as thermal and chemical processes, biological systems, have inherent time-delay. There has been a continuing interest in methods of identifying CT systems with time-delay. However, there is still no clear agreement on which method is best. A very nice survey of various methods is presented in [Bjö03]. A linear filter based method was recently introduced in [AHS06b], where the time-delay along with the transfer function model parameters are estimated simultaneously through a simple linear regression. In [YIKW07], iterative global non-linear least-squares and instrumental variable methods were proposed, in which the plant parameters and time-delay are estimated in a separable way. In [NXS10], a combined continuous wavelet transform (CWT) and cross correlation method is used to estimate the time-delay for MIMO dynamical systems, the unbiased time-delay was found out by calculating and handling the cross correlation between the CWT coefficients of the input and output signal. There are also many other methods dedicated to consider step response data, see *e.g.* [WZ01, LG08, LG10, AHS06a].

This chapter presents a new method to identify CT delay systems when the measured data are irregularly sampled. The proposed method is based on the separable nonlinear least-squares method and estimates in a combined way the plant parameters by the reliable SRIVC method and the time-delay by an adaptive gradient search. To robustify the time-delay search against its initialization, the ideas presented in [FMS96] are adopted. A low-pass CT filtering is introduced to widen the convergence region for parameter estimation. Digital implementation issues of the CT filtering operations require special attention in this irregularly sampled data situation and appropriate methods are suggested. For a properly chosen filter, the proposed method converges to the global minimum for a larger range of initial values. Compared with a fully gradient-based method, the SRIVC-based method reduces the computational load, since the dimension of the unknown parameters estimated by the gradient search is minimized.

This chapter is organized as follows. The identification problem is stated in Section 5.1. Then, independent estimations of the plant parameters and the time-delay are detailed in Section 5.2 and 5.3, respectively. Subsequently, the bootstrap method for time-delay system identification

is described in Section 5.4. After that, the low-pass filtering technique to enlarge the global convergence region is introduced in Section 5.5. Simulation examples are presented in Section 5.6 to illustrate the effectiveness of the proposed method. Finally, Section 5.7 presents the main conclusions.

5.1 Problem Statement

Consider the following COE system

$$\begin{cases} x(t) = G(p, \boldsymbol{\rho}^o)u(t - \tau^o) = \dfrac{B(p, \boldsymbol{\rho}^o)}{A(p, \boldsymbol{\rho}^o)}u(t - \tau^o) \\ y(t_k) = x(t_k) + e(t_k) \end{cases} \tag{5.1}$$

where $u(t), x(t)$ and τ^o are the excitation signal, the noise-free response and the pure time-delay, respectively. p denotes the differentiation operator. $B(p, \boldsymbol{\rho}^o)$ and $A(p, \boldsymbol{\rho}^o)$ are polynomials assumed to be coprime which have the following structure

$$\begin{aligned} B(p, \boldsymbol{\rho}^o) &= b_0^o p^m + b_1^o p^{m-1} + \cdots + b_m^o \\ A(p, \boldsymbol{\rho}^o) &= a_0^o p^n + a_1^o p^{n-1} + \cdots + 1, \quad (n \geq m) \end{aligned}$$

The true parameters are stacked in a column vector $\boldsymbol{\theta}^o$

$$\boldsymbol{\theta}^o = [\boldsymbol{\rho}^o;\ \tau^o] = [a_0^o\ \cdots\ a_{n-1}^o\ b_0^o\ \cdots\ b_m^o\ \tau^o]^T \in \mathbb{R}^{n+m+2} \tag{5.2}$$

$u(t)$ and $y(t)$ are assumed to be sampled at time-instants $\{t_k\}_{k=1}^N$, the sampled signals $u(t_k)$ and $y(t_k)$ are assumed to be recorded at irregular time-instants, the time-varying sampling interval is denoted as

$$h_k = t_{k+1} - t_k, \quad k = 1, 2, \cdots, N-1 \tag{5.3}$$

Thus, the identification problem can be stated as: assume that the degrees n and m are known, the identification objective is to estimate the unknown plant parameters and time-delay from the irregularly sampled data $Z^N = \{u(t_k), y(t_k)\}_{k=1}^N$.

5.2 Estimation of the Plant Parameters when the Time-delay is Known

In this section, estimation of the plant parameters with the known time-delay is detailed. When the time-delay is assumed to be known, the estimation problem is simplified as a traditional CT model estimation problem. Amongst the available methods, the SRIVC algorithm has proven to be one of the most efficient algorithms in practical applications (see, *e.g.*, some recent examples in [GGYH07, GY14]) and is used in the proposed approach.

5.2.1 The SRIVC Method

Prefiltering is an implicit or explicit component in CT transfer function model estimation. In the case of the SRIVC algorithm, a COE model (which takes the same form as system (5.1), but parameterized by $\boldsymbol{\rho}$ and τ) to be estimated can be expressed in the *pseudo-linear regression* manner. Let $F(p)$ be a stable filter taking the following form

$$F(p, \boldsymbol{\rho}) = \frac{1}{A(p, \boldsymbol{\rho})} \tag{5.4}$$

Then, the COE model can be equivalently reformulated in the following regression form (it is assumed here that $\tau = \tau^o$)

$$y_F(t_k, \boldsymbol{\rho}) = \boldsymbol{\phi}_F^T(t_k, \boldsymbol{\rho}, \tau^o)\boldsymbol{\rho} + \varepsilon(t_k, \boldsymbol{\rho}, \tau^o) \tag{5.5}$$

with

$$\boldsymbol{\phi}_F^T(t_k, \boldsymbol{\rho}, \tau^o) = \big[-y_F^{(n)}(t_k, \boldsymbol{\rho}) \ \cdots \ -y_F^{(1)}(t_k, \boldsymbol{\rho}) \ u_F^{(m)}(t_k - \tau^o, \boldsymbol{\rho}) \ \cdots \ u_F(t_k - \tau^o, \boldsymbol{\rho}) \big] \tag{5.6}$$

where $\varepsilon(t_k, \boldsymbol{\rho}, \tau^o)$ is the equation error, $[\cdot]_F$ denotes the filtering operation, *i.e.* $[\cdot]_F = F(p, \boldsymbol{\rho})[\cdot]$. In the SRIVC algorithm, the IV vector $\boldsymbol{\psi}_F^T(t_k, \boldsymbol{\rho}, \tau^o)$ is chosen in the following way

$$\begin{aligned} \boldsymbol{\psi}_F^T(t_k, \boldsymbol{\rho}, \tau^o) = \big[&-x_F^{(n)}(t_k, \boldsymbol{\rho}, \tau^o) \ \cdots \ -x_F^{(1)}(t_k, \boldsymbol{\rho}, \tau^o) \\ &u_F^{(m)}(t_k - \tau^o, \boldsymbol{\rho}) \ \cdots \ u_F(t_k - \tau^o, \boldsymbol{\rho}) \big] \end{aligned} \tag{5.7}$$

where $x(t_k, \boldsymbol{\rho}, \tau^o)$ is the noise-free response of the following auxiliary model

$$x(t_k, \boldsymbol{\rho}, \tau^o) = \frac{B(p, \boldsymbol{\rho})}{A(p, \boldsymbol{\rho})} u(t_k - \tau^o) \tag{5.8}$$

From N measured input-output data, the IV estimator is given as

$$\hat{\boldsymbol{\rho}} = \big[\boldsymbol{\Psi}_F(\boldsymbol{\rho}, \tau^o) \boldsymbol{\Phi}_F^T(\boldsymbol{\rho}, \tau^o) \big]^{-1} \boldsymbol{\Psi}_F(\boldsymbol{\rho}, \tau^o) \boldsymbol{y}_F(\boldsymbol{\rho}) \tag{5.9}$$

with

$$\begin{aligned} \boldsymbol{y}_F(\boldsymbol{\rho}) &= [y_F(t_s, \boldsymbol{\rho}) \ \cdots \ y_F(t_N, \boldsymbol{\rho})]^T \\ \boldsymbol{\Phi}_F(\boldsymbol{\rho}, \tau^o) &= [\boldsymbol{\phi}_F(t_s, \boldsymbol{\rho}, \tau^o) \ \cdots \ \boldsymbol{\phi}(t_N, \boldsymbol{\rho}, \tau^o)] \\ \boldsymbol{\Psi}_F(\boldsymbol{\rho}, \tau^o) &= \big[\boldsymbol{\psi}_F(t_s, \boldsymbol{\rho}, \tau^o) \ \cdots \ \boldsymbol{\psi}_F(t_N, \boldsymbol{\rho}, \tau^o) \big] \end{aligned}$$

where $t_s \geq \tau^o$.

Equation (3.39) contains unknown parameters on the right-hand side, therefore the SRIVC algorithm employs an iterative way to find out the optimal estimate, where at each iteration, the auxiliary model is used to generate the instrument and the prefilter based on the parameters obtained at the previous iteration, see, *e.g.* [YGG08, You11], for more details.

5.2.2 Implementation Issues of the CT Filtering Operation in the Presence of Arbitrary Time-delay

It is worth noticing that the computation of the SRIVC parameter estimates requires the value of prefiltered signals at the irregular time-instants t_k in both regression and instrument vectors, expressed under their developed forms in (5.6) and (5.7). The digital implementation issues of the CT filtering operations are well-known in CT model identification but they should be treated in an appropriate way here. Indeed, some components of the both vectors include an arbitrary time-delay and the latter requires special attention. Consider $u_F(t_k-\tau, \boldsymbol{\rho})$ for example

$$u_F(t_k - \tau, \boldsymbol{\rho}) = \frac{1}{A(p, \boldsymbol{\rho})} u(t_k - \tau) \tag{5.10}$$

The difficulty of computing $u_F(t_k - \tau, \boldsymbol{\rho})$ from irregularly sampled data is that $u(t)$ is unobserved at time-instant $t_k - \tau$. However, if the inter-sample behavior of $u(t_k)$ is *a priori* known, $u(t_k - \tau)$ can be exactly reconstructed from its neighboring data. This idea can be illustrated more clearly in Figure 5.1 where the zero-order-hold (ZOH) assumption is used. Of course, other higher order assumptions, *e.g.* first-order-hold (FOH), could also be used.

The digital simulation of $u_F(t_k - \tau)$ is carried out in state-space form. Let the equivalent CT state-space model of (D.1) denoted as

$$\begin{cases} \dot{\boldsymbol{z}}(t - \tau) = \boldsymbol{F}_c \boldsymbol{z}(t) + \boldsymbol{G}_c u(t - \tau) \\ u_F(t) = \boldsymbol{H}_c \boldsymbol{z}(t) \end{cases} \tag{5.11}$$

where $\dot{\boldsymbol{z}}(t)$ denotes the time-derivative of $\boldsymbol{z}(t)$. Equation (D.2) has the following sampled data model

$$\begin{cases} \boldsymbol{z}(t_{k+1} - \tau) = \boldsymbol{F}(h_k) \boldsymbol{z}(t_k - \tau) + \boldsymbol{M}(h_k) \\ u_F(t_k - \tau) = \boldsymbol{H}_c \boldsymbol{z}(t_k - \tau) \end{cases} \tag{5.12}$$

where

$$\boldsymbol{F}(h_k) = e^{\boldsymbol{F}_c h_k}, \ \boldsymbol{M}(h_k) = \int_{t_k - \tau}^{t_{k+1} - \tau} e^{\boldsymbol{F}_c (t_{k+1} - \tau - t)} \boldsymbol{G}_c u(t) dt$$

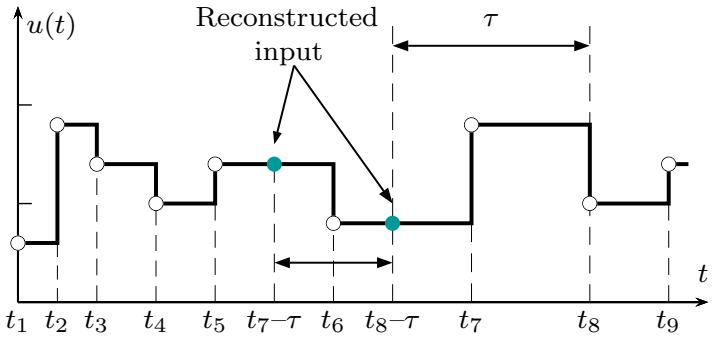

Figure 5.1: Reconstruction of the signals at unobserved time-instants using ZOH assumption.

where $h_k = t_{k+1} - t_k$. The computation of the integral $\boldsymbol{M}(h_k)$ needs more attention, consider the case that $u(t)$ is ZOH, $u(t)$ may change at some time-instants during the interval $[t_k - \tau, t_{k+1} - \tau]$, for example, $u(t)$ changes at time t_6 during $[t_7 - \tau, t_8 - \tau]$ in Figure 5.1. Let $t_{k,1}, t_{k,n_k}$, where $n_k \geq 2$ is a positive integer, denote $t_k - \tau, t_{k+1} - \tau$, respectively, and $\{t_{k,i} | i = 2, \cdots, n_k - 1\}$ the transition time-instants of $u(t)$, $\boldsymbol{M}(h_k)$ can be expanded as

$$\begin{aligned}
\boldsymbol{M}(h_k) &= \sum_{i=1}^{n_k-1} \int_{t_{k,i}}^{t_{k,i+1}} e^{\boldsymbol{F}_c(t_{k,n_k}-t)} dt \ \boldsymbol{G}_c u(t_{k,i}) \\
&= \sum_{i=1}^{n_k-1} e^{\boldsymbol{F}_c(t_{k,n_k}-t_{k,i+1})} \int_0^{h_{k,i}} e^{\boldsymbol{F}_c t} dt \ \boldsymbol{G}_c u(t_{k,i}) \\
&= \sum_{i=1}^{n_k-1} \prod_{j=i+1}^{n_k-1} \boldsymbol{F}(h_{k,j}) \boldsymbol{G}(h_{k,i}) u(t_{k,i})
\end{aligned} \tag{5.13}$$

where $h_{k,i} = t_{k,i+1} - t_{k,i}$ and

$$\boldsymbol{G}(h_{k,i}) = \int_0^{h_{k,i}} e^{\boldsymbol{F}_c t} dt \ \boldsymbol{G}_c \tag{5.14}$$

The exponentials involved in $\boldsymbol{F}(h_{k,i})$ and $\boldsymbol{G}(h_{k,i})$ could be computed by some standard methods, such as the `expm` routine in Matlab. However, the computation load in using `expm` will be very heavy when $h_{k,i}$ is time-varying, therefore, the following approach is derived to provides an fast approximation of $\boldsymbol{F}(h_{k,i})$ and $\boldsymbol{G}(h_{k,i})$

1. Make the following decomposition

   $$h_{k,i} = m_{k,i}\Delta + \delta_{k,i} < (m_{k,i} + 1)\Delta$$

   where Δ is the constant sampling period, $m_{k,i}$ is a positive integer, $\delta_{k,i} \geq 0$.

2. Compute $\boldsymbol{F}(\delta_{k,i})$, $\boldsymbol{G}(\delta_{k,i})$ by a fast approximated approach. Here we adopt the method presented in [Van78], and use the 4th-order Runge-Kutta (RK4) method to approximate the solutions.

3. Compute $\boldsymbol{F}(h_{k,i})$, $\boldsymbol{G}(h_{k,i})$ using the following relationships

   $$\begin{cases} \boldsymbol{F}(h_{k,i}) = \boldsymbol{F}(m_{k,i}\Delta)\boldsymbol{F}(\delta_{k,i}) \\ \boldsymbol{G}(h_{k,i}) = \boldsymbol{F}(m_{k,i}\Delta)\boldsymbol{G}(\delta_{k,i}) + \boldsymbol{G}(m_{k,i}\Delta) \end{cases}$$

Remark 5.1. *Note that $\{\boldsymbol{F}(j\Delta), \boldsymbol{G}(j\Delta) | j = 1, 2, \cdots\}$ involved in Step 3 can be precomputed and stored in a numerical array. The main computation load is then reduced to the approximation of $\boldsymbol{F}(\delta_{k,i})$, $\boldsymbol{G}(\delta_{k,i})$ using the RK4 method. The bias order of approximated $\boldsymbol{F}(\delta_{k,i})$, $\boldsymbol{G}(\delta_{k,i})$ satisfies that $\boldsymbol{O}(\delta_{k,i}^5) \leq \boldsymbol{O}(\Delta^5)$, which depends on the chosen Δ.*

Similar expressions can be obtained when the input of the filter is linear between the time-instants, as it is the case when the input is driven by a first-order hold (FOH). The latter assumption is used to compute the filtered version of the time-derivatives of $y(t_k)$.

5.3 Estimation of the Time-delay when the Plant is Known

In this section, the plant-independent estimation of the time-delay is investigated, therefore it is assumed that $\boldsymbol{\rho} = \boldsymbol{\rho}^o$. τ is assumed to be a 'pure delay' which can be expressed as an explicit parameter in the CT model. The prediction error as a function of τ is given as

$$\boldsymbol{\epsilon}(\boldsymbol{\rho}^o, \tau) = \boldsymbol{y} - G\left(p, \boldsymbol{\rho}^o\right) \boldsymbol{u}(\tau) \tag{5.15}$$

with

$$\boldsymbol{u}(\tau) = \left[u(t_s - \tau) \ \cdots \ u(t_N - \tau)\right]^T$$
$$\boldsymbol{\epsilon}(\boldsymbol{\rho}^o, \tau) = \left[\epsilon(t_s, \boldsymbol{\rho}^o, \tau) \ \cdots \ \epsilon(t_N, \boldsymbol{\rho}^o, \tau)\right]^T$$

where $t_s \geq \tau$. The estimate $\hat{\tau}$ is obtained by minimizing the following cost function

$$\hat{\tau} = \arg\min_{\tau} J_N(\boldsymbol{\rho}^o, \tau) \tag{5.16}$$

$$J_N(\boldsymbol{\rho}^o, \tau) = \frac{1}{2(N - s + 1)} \boldsymbol{\epsilon}^T(\boldsymbol{\rho}^o, \tau) \boldsymbol{\epsilon}(\boldsymbol{\rho}^o, \tau) \tag{5.17}$$

The Gauss-Newton method can be used to solve (5.16), thus $\hat{\tau}$ can be found by the following iterative search

$$\hat{\tau}^{j+1} = \hat{\tau}^j - \mu^j \left[\nabla^2 J_N(\boldsymbol{\rho}^o, \hat{\tau}^j)\right]^{-1} \nabla J_N(\boldsymbol{\rho}^o, \hat{\tau}^j) \tag{5.18}$$

where μ is the step length, $\nabla J_N(\boldsymbol{\rho}^o, \hat{\tau}^j)$ and $\nabla^2 J_N(\boldsymbol{\rho}^o, \hat{\tau}^j)$ denote the gradient and the approximated Hessian matrix, respectively

$$\nabla J_N(\boldsymbol{\rho}^o, \tau) = \boldsymbol{\epsilon}_\tau^T \boldsymbol{\epsilon}(\boldsymbol{\rho}^o, \tau)$$
$$\nabla^2 J_N(\boldsymbol{\rho}^o, \tau) = \boldsymbol{\epsilon}_\tau^T \boldsymbol{\epsilon}_\tau$$
$$\boldsymbol{\epsilon}_\tau = \frac{\partial \boldsymbol{\epsilon}(\boldsymbol{\rho}^o, \tau)}{\partial \tau}$$

5.4 Separable Method for Continuous-time System Identification with Time-delay

Estimation of $\boldsymbol{\rho}$ and τ has been considered independently in the previous sections. The two methods are now combined together in a bootstrap manner to estimate simultaneously the time-delay and transfer function parameters.

5.4.1 Algorithm

The prediction error as a function of $\boldsymbol{\rho}$ and τ can be given as

$$\boldsymbol{\epsilon}(\boldsymbol{\theta}) = \boldsymbol{y} - G(p, \boldsymbol{\rho}) \boldsymbol{u}(\tau)$$

Let us consider the following separable scheme for estimations of $\boldsymbol{\rho}$ and τ

Proposition 5.2 (Separable method)**.** *Let the estimate $\hat{\boldsymbol{\rho}}$ be a function of τ*

$$\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{\rho}}(\tau) = \left[\boldsymbol{\Psi}_F(\boldsymbol{\theta})\boldsymbol{\Phi}_F^T(\boldsymbol{\theta})\right]^{-1}\boldsymbol{\Psi}_F(\boldsymbol{\theta})\boldsymbol{y}_F(\boldsymbol{\rho}) \tag{5.19}$$

Then $\hat{\tau}$ can be obtained as

$$\arg\min_{\boldsymbol{\theta}} J_N(\boldsymbol{\theta}) = \arg\min_{\tau} \tilde{J}_N(\tau) \tag{5.20}$$

$$\hat{\tau} = \arg\min_{\tau} \tilde{J}_N(\tau) \tag{5.21}$$

where

$$\tilde{J}_N(\tau) = J_N(\boldsymbol{\theta})\big|_{\boldsymbol{\rho}=\hat{\boldsymbol{\rho}}}$$
$$J_N(\boldsymbol{\theta}) = \frac{1}{2(N-s+1)}\boldsymbol{\epsilon}^T(\boldsymbol{\theta})\boldsymbol{\epsilon}(\boldsymbol{\theta})$$

where s is chosen that $t_s \geq t_0$. The iterative estimation of $\hat{\boldsymbol{\rho}}$ and $\hat{\tau}$ are given as

$$\hat{\tau}^{j+1} = \hat{\tau}^j - \mu^j\left[\nabla^2\tilde{J}_N\left(\hat{\tau}^j\right)\right]^{-1}\nabla\tilde{J}_N\left(\hat{\tau}^j\right) \tag{5.22}$$

$$\hat{\boldsymbol{\rho}}^{j+1} = \left[\boldsymbol{\Psi}_F\left(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\right)\boldsymbol{\Phi}_F^T\left(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\right)\right]^{-1}\boldsymbol{\Psi}_F\left(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\right)\boldsymbol{y}_F\left(\hat{\boldsymbol{\rho}}^j\right) \tag{5.23}$$

where

$$\nabla\tilde{J}_N(\hat{\tau}^j) = \boldsymbol{\epsilon}_\tau^T\boldsymbol{\epsilon}(\hat{\boldsymbol{\theta}}^j) \tag{5.24}$$

$$\nabla^2\tilde{J}_N(\hat{\tau}) \approx \boldsymbol{\epsilon}_\tau^T\boldsymbol{\epsilon}_\tau - \boldsymbol{\epsilon}_\tau^T\boldsymbol{\epsilon}_{\boldsymbol{\rho}}\left(\boldsymbol{\epsilon}_{\boldsymbol{\rho}}^T\boldsymbol{\epsilon}_{\boldsymbol{\rho}}\right)^{-1}\boldsymbol{\epsilon}_{\boldsymbol{\rho}}^T\boldsymbol{\epsilon}_\tau \tag{5.25}$$

$$\boldsymbol{\epsilon}_\tau = \frac{\partial\boldsymbol{\epsilon}(\boldsymbol{\theta})}{\partial\tau}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^j} = pG(p, \hat{\boldsymbol{\rho}}^j)\boldsymbol{u}(\hat{\tau}^j) \tag{5.26}$$

$$\boldsymbol{\epsilon}_{\boldsymbol{\rho}} = \frac{\partial\boldsymbol{\epsilon}(\boldsymbol{\theta})}{\partial\boldsymbol{\rho}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^j} = -\boldsymbol{\Psi}_F^T(\hat{\boldsymbol{\theta}}^j) \tag{5.27}$$

Proof. See [Ngi01] and the references therein for the details of the separable non-linear least squares method. □

□

The proposed algorithm (TDSRIVC) for CT model identification with time-delay can be summarized as

Algorithm 5.1 (The basic TDSRIVC algorithm)**.**

---

1. *Initialization: set the tolerances $\Delta\tau_{min}$, $\Delta\tilde{J}_N(\tau)_{min}$ along with the boundaries τ_{min}, τ_{max}. Based on the chosen initial value $\hat{\tau}^0$ and the initial filter $F(p)$, use the SRIVC method to compute $\hat{\boldsymbol{\rho}}^0$*

   $$F(p) = \frac{1}{(p/\omega_F + 1)^n} \tag{5.28}$$

   where ω_F is a user provided value, see [YGG08].

2. *Iteration:*
   for j=0:convergence

   a. *Compute the increment* $\Delta\hat{\tau}^j$

$$\Delta\hat{\tau}^j = -\left[\nabla^2 \tilde{J}_N\left(\hat{\tau}^j\right)\right]^{-1} \nabla \tilde{J}_N\left(\hat{\tau}^j\right)$$

   b. *Perform the following*

      i. *Compute* $\hat{\tau}^{j+1} = \hat{\tau}^j + \Delta\hat{\tau}^j$.
      If $\hat{\tau}^{j+1} \notin [\tau_{min}, \tau_{max}]$, *let* $\Delta\hat{\tau}^j = \Delta\hat{\tau}^j/2$ *and repeat this step.*
      If $|\Delta\hat{\tau}^j| < \Delta\tau_{min}$, *break.*

      ii. *Estimate* $\boldsymbol{\rho}$ *using the SRIVC estimator*
      Update auxiliary model and prefilter

$$G\big(p, \hat{\boldsymbol{\rho}}^j\big) = B\big(p, \hat{\boldsymbol{\rho}}^j\big)/A\big(p, \hat{\boldsymbol{\rho}}^j\big)$$
$$F\big(p, \hat{\boldsymbol{\rho}}^j\big) = 1/A\big(p, \hat{\boldsymbol{\rho}}^j\big)$$

      Compute the filtered signals in

$$\boldsymbol{\Psi}_F\big(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\big),\ \boldsymbol{\Phi}_F\big(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\big),\ \boldsymbol{y}_F\big(\hat{\boldsymbol{\rho}}^j\big)$$

      Estimate $\boldsymbol{\rho}$

$$\hat{\boldsymbol{\rho}}^{j+1} = \Big[\boldsymbol{\Psi}_F\big(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\big)\boldsymbol{\Phi}_F^T\big(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\big)\Big]^{-1} \boldsymbol{\Psi}_F\big(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\big)\boldsymbol{y}_F\big(\hat{\boldsymbol{\rho}}^j\big)$$

      If $\tilde{J}_N\left(\hat{\tau}^{j+1}\right) \geq \tilde{J}_N(\hat{\tau}^j)$, *let* $\Delta\hat{\tau}^j = \Delta\hat{\tau}^j/2$ *and go to (i).*

   (a) *Check the stopping condition. Go to step (a) if the following condition is satisfied, else break.*

$$\tilde{J}_N(\hat{\tau}^j) - \tilde{J}_N\left(\hat{\tau}^{j+1}\right) \geq \Delta\tilde{J}_N(\tau)_{min}$$

end

The proposed bootstrap algorithm provides very good estimates when the initial time-delay value is closed to the true one. However, the gradient-based method might be trapped in a local minimum since the cost function $J(\boldsymbol{\rho}, \tau)$ is a nonlinear function of the time-delay τ. To circumvent this problem, a low-pass filtering strategy is introduced in the next section to improve the convergence performance.

5.5 Expansion of the Convergence Region by Low-pass Filtering

5.5.1 The Case without Low-pass Filtering

Consider the objective function with respect to the CT error

$$J(\boldsymbol{\rho}, \tau) = \frac{1}{2}\int_{-\infty}^{\infty} \epsilon^2(t, \boldsymbol{\rho}, \tau)dt \tag{5.29}$$

From Parseval's theorem, $\bar{J}(\boldsymbol{\rho}^o, \tau)$ has the following frequency-domain interpretation

$$J(\boldsymbol{\rho}, \tau) = \frac{1}{4\pi}\int_{-\infty}^{+\infty} \left[\left|G(i\omega, \boldsymbol{\rho}^o)e^{-i\omega\tau^o} - G(i\omega, \boldsymbol{\rho})e^{-i\omega\tau}\right|^2 \Phi_u(\omega) + \Phi_e(\omega)\right] d\omega \tag{5.30}$$

where $\Phi_u(\omega)$ and $\Phi_e(\omega)$ are the power spectral density (PSD) functions of $u(t)$ and $e(t)$, respectively. According to [FMS96], $G(i\omega, \boldsymbol{\rho}^o)$ can be formulated in terms of $G(i\omega, \boldsymbol{\rho})$ and the model error $M(\omega)e^{i\varphi(\omega)}$

$$G(i\omega, \boldsymbol{\rho}^o) = G(i\omega, \boldsymbol{\rho})\left[1 + M(\omega)e^{i\varphi(\omega)}\right] \tag{5.31}$$

Let the distance between τ and τ^o by $\delta\tau$

$$\delta\tau = \tau - \tau^o \tag{5.32}$$

substituting (5.31) and (5.32) into $J(\boldsymbol{\rho}, \tau)$ and removing the constant term yields the following equation

$$V(\boldsymbol{\rho}, \delta\tau) = \frac{1}{4\pi}\int_{-\infty}^{+\infty} |G(i\omega, \boldsymbol{\rho})|^2 \left|1 - e^{i\omega\delta\tau}\left[1 + M(\omega)e^{i\varphi(\omega)}\right]\right|^2 \Phi_u(\omega)d\omega \tag{5.33}$$

Since the same results can be drawn from both cost functions $J(\boldsymbol{\rho}, \tau)$ and $V(\boldsymbol{\rho}, \delta\tau)$, $V(\boldsymbol{\rho}, \delta\tau)$ will be used in the following sections for further analysis.

To quantify the width of the convergence region, the concept of *admissible region* is given below

Definition 5.3 (Admissible Region)**.** *If $\delta\tau_{max}$ is the maximum value which guarantees that when $\tau \in [\tau^o - \delta\tau_{max}, \tau^o + \delta\tau_{max}]$, $J(\boldsymbol{\rho}, \tau)$ is a convex function with respect to τ, then, $\delta\tau_{max}$ is termed as the admissible region.*

However, when $\boldsymbol{\rho} \neq \boldsymbol{\rho}^o$, the admissible region for $V(\boldsymbol{\rho}, \delta\tau)$ is quite hard to determine because $M(\omega) \neq 0$ (see [FMS96] for more discussions). In this chapter, we only analyze the simple case when $\boldsymbol{\rho} = \boldsymbol{\rho}^o$. The following theorem is used to compute $V(\boldsymbol{\rho}^o, \delta\tau)$ and the corresponding $\delta\tau_{max}$ based on the output signal.

Theorem 5.4. *Assuming that the PSD function of $u(t)$ is denoted by $\Phi_u(\omega)$. Let $x(t)$ denote the response of the true system*

$$x(t) = G(p, \boldsymbol{\rho}^o)u(t - \tau^o)$$

Then $V(\boldsymbol{\rho}^o, \delta\tau)$ can be computed by

$$V(\boldsymbol{\rho}^o, \delta\tau) = R_x(0) - R_x(\delta\tau) \tag{5.34}$$

where R_x is the autocorrelation function of $x(t)$

$$R_x(\delta\tau) = \mathbb{E}\left\{x(t)x(t-\delta\tau)\right\}$$

Proof. This theorem is quite similar to Theorem 1 in [FMS96], the difference is that we do not apply a low-pass filter here. By setting $M(\omega)$ to zero, $V(\boldsymbol{\rho}^o, \delta\tau)$ can be given as

$$\begin{aligned} V(\boldsymbol{\rho}^o, \delta\tau) =& \frac{1}{4\pi}\int_{-\infty}^{+\infty} \left|G(i\omega, \boldsymbol{\rho}^o)\right|^2 \left|1 - e^{i\omega\delta\tau}\right|^2 \Phi_u(\omega) d\omega \\ =& \frac{1}{4\pi}\int_{-\infty}^{+\infty} \Phi_x(\omega)\left(2 - e^{-i\omega\delta\tau} - e^{i\omega\delta\tau}\right) d\omega \end{aligned} \tag{5.35}$$

where $\Phi_x(\omega) = \left|G(i\omega, \boldsymbol{\rho}^o)\right|^2 \Phi_u(\omega)$ is the PSD of $x(t)$. By using the well-known *Wiener-Khinchin* theorem, (5.35) can be reformulated as

$$V(\boldsymbol{\rho}^o, \delta\tau) = R_x(0) - \frac{1}{2}R_x(-\delta\tau) - \frac{1}{2}R_x(\delta\tau) = R_x(0) - R_x(\delta\tau) \tag{5.36}$$

The second equation in (5.36) uses the fact that $R_x(\delta\tau)$ is an even function. □ □

The following example gives a graphical illustration of the relationship between $V(\boldsymbol{\rho}, \delta\tau)$ and the unknown parameters.

Example 5.1. *Consider the following noise-free system*

$$x(t_k) = \frac{1}{a_0^o p + 1} u(t_k - \tau^o) \tag{5.37}$$

where $a_0^o = 1, \tau^o = 12s$. The system is excited by $u(t)$ with $\Phi_u(\omega) = 1$. The number of observations is $N = 1000$, the sampling interval h_k is uniformly distributed in the following interval

$$h_k \sim U[0.01, 0.09]s$$

The surface plot of $V(\boldsymbol{\rho}, \delta\tau)$ as a function of a_0 and $\delta\tau$ is shown in the Figure 5.2. When $a_0 = a_0^o = 1$, the admissible region is $\delta\tau_{max} \approx 2.6$, it indicates that the time-delay search will converge to the true value when it is initialized within the region $[9.4, 14.6]s$. Note that the intersection line $V(\boldsymbol{\rho}^o, \delta\tau)$ of the surface plot and the transparent plane can be obtained by Theorem 5.4, thus $\delta\tau_{max} \approx 2.6$ can be obtained without any knowledge of the true parameters.

5.5.2 Expansion of the Convergence Region by Low-pass Filtering

To increase the chance of converging to the global minimum, the use of low-pass filter was initially suggested in [FMS96]. It was shown that the use of filtered signals in estimation can widen the global convergence region in searching the time-delay.

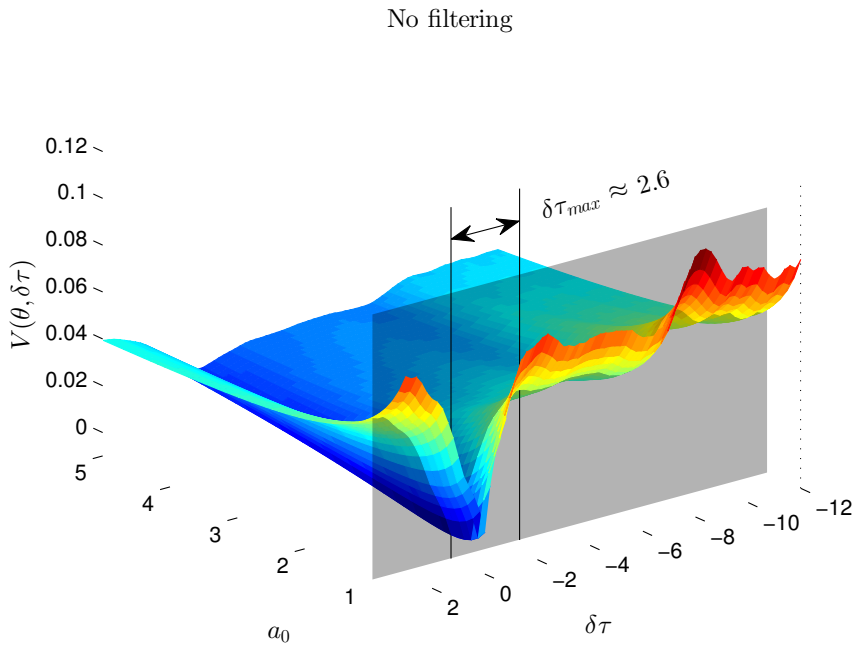

Figure 5.2: The surface plot of $V(\boldsymbol{\rho}, \delta\tau)$ as a function of a_0 and $\delta\tau$. $\delta\tau_{max}$ is computed when $a_0 = 1$. The intersection line of the transparent plane and the surface plot is $V(\boldsymbol{\rho}^o, \delta\tau)$.

Let the filtered error and cost function be denoted by

$$\bar{\epsilon}(t, \boldsymbol{\rho}, \tau) = L(p)\epsilon(t, \boldsymbol{\rho}, \tau) \tag{5.38}$$

$$\bar{J}(\boldsymbol{\rho}, \tau) = \frac{1}{2}\int_{-\infty}^{\infty} \bar{\epsilon}^2(t, \boldsymbol{\rho}, \tau)dt \tag{5.39}$$

where $L(p)$ is a CT low-pass filter. Even if the theory is attractive, from a practical side, there is an important issue to be discussed which is the choice of $L(i\omega)$. Amongst the possible options, an ideal low-pass filter is easier to specify and will be used in the proposed approach. In the situation of ideal low-pass filtering, Theorem 5.4 should be modified as follows

Proposition 5.5. *Assume $L(i\omega)$ is an ideal low-pass filter with the cutoff frequency of ω_L*

$$L(i\omega) = \begin{cases} 1 & |\omega| \leq \omega_L \\ 0 & \textit{otherwise} \end{cases} \tag{5.40}$$

Let

$$\bar{x}(t) = L(p)x(t) \tag{5.41}$$

where $x(t)$ is defined in Theorem 5.4. $\bar{[\cdot]}$ denotes the filtering operation. Based on this filtered signal, equation (5.34) in Theorem 5.4 is modified as

$$\bar{V}(\boldsymbol{\rho}^o, \delta\tau) = R_{\bar{x}}(0) - R_{\bar{x}}(\delta\tau) \tag{5.42}$$

The following example is used to illustrate the impact of the low-pass filter.

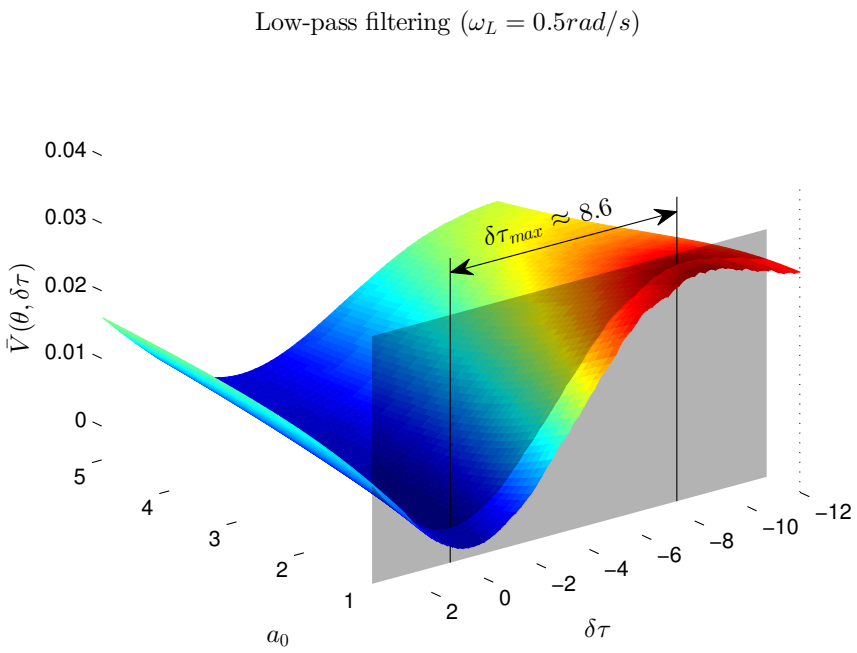

Figure 5.3: The filtered index term $\bar{V}(\boldsymbol{\rho}, \delta\tau)$ as a function of a_0 and $\delta\tau$ ($\omega_L = 0.5rad/s$). $\delta\tau_{max}$ is computed when $a_0 = 1$. The intersection line of the two the transparent plane and the surface plot is $V(\boldsymbol{\rho}^o, \delta\tau)$

Example 5.2. *Consider the same system as Example 5.1. An ideal low-pass filter $L(i\omega)$ having the cutoff frequency of $\omega_L = 0.5rad/s$ is applied to expand the convergence region. The filtered index term $\bar{V}(\boldsymbol{\rho}, \tau)$ as a function of τ and a_0 is plotted in Figure 5.3. When $a_0 = a_0^o = 1$, the admissible region is widened to $\delta\tau_{max} \approx 8.6$, the corresponding convergence region is $[3.4, 20.6]s$. This figure also illustrates that the sharp peaks of the cost function are 'spread out' by the low-pass filter, which makes the global minimum easier to reach.*

5.5.3 The Robust TDSRIVC Algorithm

Since the low-pass filter can widen the convergence region, the TDSRIVC algorithm is then modified to take this advantage. We suggest to run the TDSRIVC algorithm in the first m iterations by minimizing the filtered cost function $\bar{J}(\boldsymbol{\rho}, \tau)$, which could yield a relatively accurate estimate from a coarse initialization, then we move to minimize the unfiltered cost function $J(\boldsymbol{\rho}, \tau)$ until the convergence occurs. The modified algorithm is summarized below

Algorithm 5.2 (The robust TDSRIVC algorithm)**.**

---

1. *Initialization.*
2. *Iteration:*

- *The first stage:*
  for j=0:m
  Perform the estimation by minimizing the filtered cost function

$$\hat{\rho} = \arg\min_{\boldsymbol{\theta}} \bar{J}_N(\boldsymbol{\theta}) \tag{5.43}$$

  where $\bar{[\cdot]}$ *denotes the additional filtering operation by* $L(p)$*. Equations (5.22) and (5.23) are modified as*

$$\hat{\tau}^{j+1} = \hat{\tau}^j - \mu^j \left[\nabla^2 \bar{\tilde{J}}_N\left(\hat{\tau}^j\right)\right]^{-1} \nabla \bar{\tilde{J}}_N\left(\hat{\tau}^j\right) \tag{5.44}$$

$$\hat{\boldsymbol{\rho}}^{j+1} = \left[\bar{\boldsymbol{\Psi}}_F\left(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\right) \bar{\boldsymbol{\Phi}}_F^T\left(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\right)\right]^{-1} \bar{\boldsymbol{\Psi}}_F\left(\hat{\boldsymbol{\rho}}^j, \hat{\tau}^{j+1}\right) \bar{\boldsymbol{y}}_F\left(\hat{\boldsymbol{\rho}}^j\right) \tag{5.45}$$

  end

- *The second stage:*
  for j=m:convergence
  Perform the estimation by minimizing the unfiltered cost function

$$\hat{\rho} = \arg\min_{\boldsymbol{\theta}} J_N(\boldsymbol{\theta}) \tag{5.46}$$

  Equations (5.22) and (5.23) are used to estimate the delay and the plant parameters.
  end

Remark 5.6. *The question remains here to know how* ω_L *should be chosen. The previous statements conclude that the global minimum is easier to access for smaller values of* ω_L*. Then, one may ask, can* ω_L *be arbitrarily small? Our answer is no, the reason can be given as: 1) The NUFFT is used to implement the ideal low-pass filtering, which has a minimum resolution. So it does not make sense when* ω_L *is lower than this resolution. 2) Smaller values for* ω_L *mean that more information of the sampled data is removed, which can deteriorate the plant parameter estimation in Algorithm 5.2. A rule of thumb can be to choose the cut-off frequency of the ideal low-pass filter as* $\frac{1}{10}$ *to* $\frac{1}{2}$ *of the system bandwidth. Another problem is the choice of* m*, a rule of thumb is to chose a value between* 5 *and* 10*, the convergence region can always be extended.*

5.5.4 Implementation of the Ideal Low-pass Filtering from Irregularly Sampled Data

Once the cut-off frequency of the ideal filter is chosen, there is a last issue to be investigated: how to implement the ideal low-pass filtering operations in the present irregularly sampled data situation? One efficient way is to carry out the latter by using the non-uniform fast Fourier transform (NUFFT) and its inverse, by adding a rectangular window to the frequency data.

Different approximated methods for fast Fourier transform (FFT) from irregularly sampled data were reviewed in [War98]. The basic idea is to reconstruct a sequence of *regularly sampled*

data, then the traditional fast Fourier transform can be applied. Here we use the so called *polynomial interpolation through neighboring grid points* suggested in [War98] where the nearest 2 points are used to construct the regularly resampled data.

Assume the sampled data is $\{y(t_k)\}_{k=1}^{N}$, where t_k is irregularly spaced. Let $\{y'_p\}_{p=1}^{N'}$ denote the regularly resampled data, with constant sampling period Δ

$$N' = 2^M \geq N > 2^{M-1}, \quad M = 1, 2, 3, \cdots \tag{5.47}$$

$$y'_p = \frac{y(t_{k+1}) - y(t_k)}{t_{k+1} - t_k}(p\Delta - t_k) + y(t_k), \;\; t_k \leq p\Delta < t_{k+1} \tag{5.48}$$

Based on y'_p, apply fast Fourier transform

$$F_q = \sum_{p=1}^{N'} y'_p e^{-i2\pi \frac{(p-1)(q-1)}{N'}}, \;\; q = 1, \cdots, N' \tag{5.49}$$

Assume that the ideal low-pass filter $L(i\omega)$ has the cutoff frequency of ω_L, N_c is an integer $(1 \leq N_c < N'/2)$

$$\frac{N_c}{N'} \cdot \frac{2\pi}{\Delta} \leq \omega_L < \frac{N_c + 1}{N'} \cdot \frac{2\pi}{\Delta} \tag{5.50}$$

We define the following signal

$$G_q = \begin{cases} 0, & N_c + 1 < q \leq N' - N_c \\ 1, & \text{otherwise} \end{cases}$$

Then the filtered version of y'_p by $L(i\omega)$ can be obtained by the inverse Fourier transform

$$\bar{y}'_p = \frac{1}{N'} \sum_{q=1}^{N'} F_q G_q e^{i2\pi \frac{(p-1)(q-1)}{N'}}, \;\; p = 1, \cdots, N' \tag{5.51}$$

where $[\cdot]$ denotes the filtering operation by $L(i\omega)$. Finally $\bar{y}(t_k)$ can be obtained by the following interpolation

$$\bar{y}(t_k) = \frac{\bar{y}'_{p+1} - \bar{y}'_p}{\Delta}(t_k - p\Delta) + \bar{y}'_p, \quad p\Delta \leq t_k < (p+1)\Delta \tag{5.52}$$

5.6 Numerical Examples

5.6.1 Identification from Irregularly Sampled Data

In this section, the identification problem for systems of different orders is considered. In the simulation, the following conditions are assumed

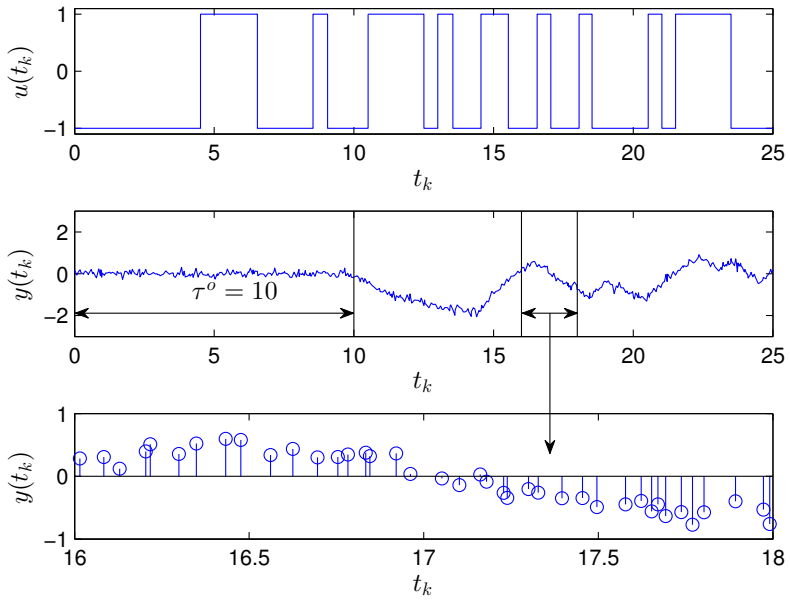

Figure 5.4: Sampled input-output data for System 1. Top-The input signal. Middle-The output signal where the presence of the delayed response can be observed and bottom-Zoom part of the sampled output where the irregular sampling can be observed.

1. The systems to be identified take the general form of (5.1), with true parameters given in Table 5.1. The input signal is chosen to be a pseudo random binary sequence (PRBS) switching between ± 1, which is generated from a 9-stage shift register with the clock period of $0.5s$. The input-output signals are observed at $N = 1500$ time-instants, the sampling interval h_k is uniformly distributed in the following interval

$$h_k \sim U[0.01, 0.09]s \tag{5.53}$$

The variance of $e(t_k)$ is adjusted to obtain a signal-to-noise ratio (SNR) of 15dB. The SNR is defined below

$$\text{SNR} = 10\log \frac{P_x}{P_e}$$

where P_x and P_e represent the average power of the noise-free signal $x(t_k)$ and the disturbance $e(t_k)$. A portion of the irregularly sampled data for System 1 is plotted in Figure 5.4.

2. Monte-Carlo (MC) simulation with 100 runs is performed to obtain the statistical estimation results. Each run has random realizations of h_k and $e(t_k)$. To initiate the TDSRIVC algorithm, one needs to provide ω_F and $\hat{\tau}^0$ (see Algorithm 5.1). It has been declared in [You11] (p250) that the choice of ω_F is not critical for the SRIVC algorithm since convergence occurs over a fairly wide range of values. In this section, ω_F, ω_L and $\hat{\tau}^0$ are chosen as

$$\omega_F = 1rad/s, \omega_L = 0.1rad/s,\ \hat{\tau}^0 = \{1, 3, 5, 7, 9\}\, s \tag{5.54}$$

No.	True system	P_{glob} without filtering					P_{glob} with filtering				
	$\hat{\tau}^0$	$1s$	$3s$	$5s$	$7s$	$9s$	$1s$	$3s$	$5s$	$7s$	$9s$
1.	$\frac{2e^{-8s}}{2s+1}$	5%	3%	9%	49%	97%	100%	99%	99%	100%	98%
2.	$\frac{3e^{-8s}}{0.25s^2+s+1}$	0%	1%	15%	100%	100%	100%	100%	100%	99%	93%
3.	$\frac{2e^{-8s}}{0.25s^2+0.7s+1}$	12%	10%	13%	98%	61%	98%	100%	99%	98%	62%
4.	$\frac{(-4s+1)e^{-8s}}{9s^2+2.4s+1}$	0%	8%	49%	97%	20%	98%	98%	99%	90%	76%

Table 5.1: Monte-Carlo simulation results for the proposed method with and without applying a low-pass filter from irregularly sampled data. P_{glob}-the ratio of global convergence.

3. The normalized root-mean-square error (NRMSE) is used as the benchmark for performance evaluation, it is computed by the following formula

$$\text{NRMSE} = 100\left(1 - \frac{\|y(t) - \hat{x}(t)\|_2}{\|y(t) - \mathbb{E}\{y(t)\}\|_2}\right) \tag{5.55}$$

where $\hat{x}(t)$ is the noise-free response of the estimated model. When SNR $= 15$dB, NRMSE satisfies the following inequality

$$\begin{aligned}\text{NRMSE} &\leq 100\left(1 - \frac{\sqrt{P_e}}{\sqrt{P_x + P_e}}\right) \\ &= 100\left[1 - \left(1 + 10^{\text{SNR}/10}\right)^{-1/2}\right] \approx 82.5\end{aligned}$$

In this chapter, the criterion for global convergence is chosen as NRMSE > 81.

4. The following estimation strategies are considered:

   A. Identification without low-pass filtering (Algorithm 5.1). τ_{min} and τ_{max} are set to 0 and 10, respectively. The stopping rule algorithm is that the relative change in $J_N\left(\hat{\boldsymbol{\theta}}\right)$ is lower than 0.01%. The basic TDSRIVC algorithm is allowed to have 50 iterations at most.

   B. Identification with low-pass filtering (Algorithm 5.2). τ_{min}, τ_{min} keep the same. The maximum number of iterations of the first estimation stage (With low-pass filtering) is set to 10. During these 10 iterations, the algorithm will be terminated if the relative change of $\bar{J}_N(\boldsymbol{\theta})$ is smaller than 0.01%. Subsequently, the algorithm proceeds to the second estimation stage (No filtering), the maximum number of iterations in this stage is set to 50.

MC simulations (each has 100 realizations) are performed by using different initializations, more specifically, ω_F and ω_L are fixed and $\hat{\tau}^0$ picks up a value from $\{1, 3, 5, 7, 9\}$ sequentially. The global convergence ratios of the MC simulations are gathered in Table 5.1. It can be observed that Strategy A (No filtering) is more probable to converge when good initial guesses of the time-delay are provided (Around $\tau^o = 8s$), while Strategy B (Low-pass filtering) can converge from a coarse initialization.

No.	True system	Method	$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{\tau}$	$\bar{N}_{iter}$	P_{glob}
1.	$\frac{2e^{-8s}}{2s+1}$	TDSRIVC	1.9999 $\pm$0.0207	——	2.0005 $\pm$ 0.0129	——	7.9996 $\pm$ 0.0038	11.7	100%
		PROCEST	1.9996 $\pm$0.0205	——	2.0003 $\pm$ 0.0127	——	7.9996 $\pm$ 0.0038	15.2	100%
2.	$\frac{3e^{-8s}}{0.25s^2+s+1}$	TDSRIVC	0.2488 $\pm$0.0077	0.9984 $\pm$ 0.0086	3.0007 $\pm$ 0.0180	——	8.0011 $\pm$ 0.0068	11.3	100%
		PROCEST	0.2489 $\pm$0.0042	0.9984 $\pm$0.0080	3.0007 $\pm$0.0168	——	8.0010 $\pm$0.0049	21.2	96%
3.	$\frac{2e^{-8s}}{0.25s^2+0.7s+1}$	TDSRIVC	0.2492 $\pm$0.0049	0.6992 $\pm$0.0058	2.0011 $\pm$0.0116	——	8.0009 $\pm$0.0057	10.7	100%
		PROCEST	0.2489 $\pm$0.0044	0.6987 $\pm$0.0057	2.0005 $\pm$0.0113	——	8.0013 $\pm$0.0053	21.5	73%
4.	$\frac{(-4s+1)e^{-8s}}{9s^2+2.4s+1}$	TDSRIVC	8.9988 $\pm$0.0563	2.3975 $\pm$0.0293	-3.9947 $\pm$0.0351	0.9985 $\pm$ 0.0158	7.9994 $\pm$0.0062	12.0	93%
		PROCEST	8.9993 $\pm$0.0557	2.3954 $\pm$0.0295	-3.9921 $\pm$0.0356	0.9981 $\pm$0.0161	7.9992 $\pm$ 0.0062	16.0	81%

Table 5.2: Estimated parameters from regularly sampled data with low-pass filtering (Strategy B). $\bar{N}_{iter}$-The averaged number of iterations for global convergence, P_{glob}-the ratio of global convergence. The mean value, the standard deviation are computed from the estimated models that converge to the global minimum.

5.6.2 Identification from Regularly Sampled Data: A Comparison with PROCEST

The proposed method can also handle uniformly sampled data by assuming that the sampling interval h_k is constant. In this case, the PROCEST routine available in the Matlab System Identification (SID) Toolbox (Version 2014a) is also used to give a comparison. PROCEST is a variation of the prediction error minimization (PEM) method, which can estimate up to a 3rd order simple process model plus delay from regularly sampled data only (see [Lju12]).

Compared with the previous section, the simulation conditions keep the same except the following two

- The constant sampling interval is chosen as $h_k = 0.05s$.
- The initial value of the time-delay is generated from a uniform distribution $\hat{\tau}^0 \sim U[0, 10]s$.

The settings for the PROCEST routine are given here: the minimum and maximum values of the time-delay τ are chosen as 0 and 10, respectively. `'SearchMethod'='gna'`, `'Tolerance'`=0.01%. When Strategy B is used in estimation, a focus filter (low-pass filter) is performed in the first stage by setting `'Focus'`=$[0, \omega_L]$, the maximum number of iterations is set to `'MaxIter'`=10. Then in the second stage, `'Focus'`, `'MaxIter'` are changed to `'Prediction'` and 50, respectively. The same filter $F(p)$ defined in (5.28) and time-delay $\hat{\tau}^0$ are used to initialize the PROCEST routine in identifying System $1 \sim 3$. In estimating System 4, the initial model is $(-4p+1)F(p)$. The number of iterations is the total iterations in the both estimation stages.

The results of strategy B are presented in Table 5.2. For the first order system, TDSRIVC and PROCEST routines provide identical and very good global convergence ratios. When the system has more parameters to estimate (System 3 and 4), the global convergence ratio slightly drops for the PROCEST routine while the proposed TDSRIVC method is still able to provide fairly good results. The proposed TDSRIVC seems however to converge more quickly. The robustness of TDSRIVC comes from the fact that it has only one dimension (the time-delay) to be estimated by numerical search, the other parameters are estimated by the robust SRIVC method. By contrast, all the parameters are estimated by numerical search in PROCEST.

5.7 Conclusion

In this chapter, an approach to continuous-time system identification with arbitrary time-delay from irregularly sampled input-output data has been presented. The proposed approach combines the SRIVC method in a bootstrap manner with an adaptive gradient-based search algorithm. The implementation issues of CT filtering from irregularly sampled data have been addressed. In order to improve the convergence property, a low-pass filter strategy has been introduced to widen the convergence region. Simulation results have shown that the proposed method is of low complexity and converges to the true parameters from a relatively wide range of time-delay initial values. Numerical examples have shown that the proposed method has similar or superior convergence property to the explicit method available in the System Identification toolbox for simple process plus delay estimation which can handle regularly sampled data only. This new algorithm provides therefore more choices for CT delayed system identification and constitutes a very useful addition to the CONTSID toolbox.

5.8 Relevant Matlab Routines

1. `simc_fw`: simulate the CT system with input delay.

$$y_t = \frac{B(p)}{A(p)} u_{t-\tau} \tag{5.56}$$

   Syntax: `y=simc_fw(A,B,u,t,'timedelay',td)`

   - `td` is optional.

2. `lf_nufft`: the ideal low-pass filter for non-uniformly sampled data.

   Syntax: `y=lf_nufft(y,t,fc)`

   - `fc`: the cutoff frequency of the filter.

3. `tdsrivc`: the SRIVC method for time-delay system identification.

   Syntax: `[theta,tau,niter]=tdsrivc(data,[na,nb],fcsvf,'cutoff',fc)`

   - `fcsvf`: the cutoff frequency of the SVF, `fc`(optional): the cutoff frequency of the ideal low-pass filter.

Chapter 6

Conclusion and Perspectives

6.1 Conclusion

The use of CT models to handle irregularly sampled data is a common view in the field of system identification. This thesis reviewed the general procedure for CT system identification, and recalled several available methods for simple CT model estimation from irregularly sampled data. However these simple models are not enough to explain most of the real systems. Therefore, this thesis was carried out by concerning the following two identification problems

1. How to obtain optimal estimates in presence of colored measurement noise? In regular sampling case, the hybrid Box-Jenkins models have been used successfully to model the colored noise in DT. However, when the regular sampling assumption is violated, the hybrid Box-Jenkins models cannot be used any more due to the DT noise model assumption. Therefore, this thesis focuses on the use of CT Box-Jenkins models, where the noise is also modeled in CT.

2. How to identify the CT system robustly when it has an arbitrary input time-delay? A delayed CT model contains both linear (the plant) and non-linear (the delay) parameters. Therefore the traditional non-linear least-squares method is an option for parameter estimation. However, the convergence performance of the non-linear least-squares method is quite weak, which requires always an accurate initialization of the time-delay to produce consistent estimates. Therefore, this thesis focuses on the development of new methods that are able to consistently estimate the time-delay even in the case of bad initializations.

In Chapter 2, the sampling issues of CT stochastic systems have been reviewed. The aim of sampling is to find out a DT time-varying state-space model (sampled data model) that has the same second order properties (*e.g.* the PSD function) as the CT system at the sampling instants. Three sampling schemes are reviewed in this chapter: $1)$ the instantaneous sampling scheme, which is used to sample a stochastic system when the output has finite variance; $2)$ the integrated sampling scheme, which is proposed to sample a stochastic process having infinite variance; $3)$ the simplistic sampling scheme, which is the simplified version of the integrated sampling scheme that can reduce the computational load of the DT model matrices. Other issues such as aliasing, the computation of the DT model matrices are also discussed in this chapter.

In Chapter 3, the estimation of CAR/CARMA noise models has been investigated. We firstly considered a simpler case: CAR model estimation. The proposed EM-based method uses a δ-operator model to approximate the CT model for fast sampled data. In order to improve the performance of the proposed method in handling slowly sampled data, a state interpolation technique is developed to increase the precision of the parameter estimates. It is shown that more accurate estimates can be obtained by choosing a smaller interpolation interval λ. Due to the fact that the proposed EM-based method cannot estimate MA parameters, an additional numerical search to estimate the MA parameters is combined to the EM algorithm. This leads to the Grad-EM algorithm for CARMA model estimation.

In Chapter 4, identification of CT Box-Jenkins models from irregularly sampled data has been studied. The proposed EM-based RIVC (RIVC-EM) algorithm estimates the plant model by the IV method and the CAR/CARMA noise model by the EM algorithm. This approach has been also extended to handle CT closed-loop system identification. The performance of the proposed methods has been investigated by means of simulation.

In Chapter 5, a new approach to estimate CT models with arbitrary time-delay from irregularly sampled data has been presented. The proposed method is based on the non-linear least-squares method which combines the SRIVC method for the plant parameter estimation with an adaptive gradient-based search algorithm for the time-delay estimation. Since the object function has many local minima with respect to the unknown parameters, a low-pass filter strategy was applied to widen the convergence region. Simulation results showed that the proposed method is of low complexity and converges to the true parameters from a relatively wide range of time-delay initial values.

Therefore, the contributions of this thesis can be summed up below

- The proposition and development of a EM-based method for continuous-time autoregressive/autoregressive moving average noise model identification.
- The proposition of a refined instrumental variable method for full continuous-time Box-Jenkins model identification in both open-loop and closed-loop situations.
- The proposition and development of simplified refined instrumental variable method for continuous-time model identification with arbitrary time-delay.

All the developed algorithms assume irregularly sampled data. The developed routines in this thesis will be included in the upcoming version of the CONTSID toolbox.

6.2 Perspectives

The presented work constitutes a basis for the development of identification methods in the following contexts

- The identification of a CT ARMA model by using the EM algorithm only. In this way, the AR and MA parameters are estimated in the same step, the complexity of the algorithm can be reduced.

- The identification of CT AR/ARMA models using integrated measurements. It is more practical to use integrated measurements since we always prefilter the signal before sampling.
- The estimation of MISO or MIMO models with multiple time-delay. The industrial process has always multiple input and multiple output, therefore the estimation of a MIMO system is appealing but more challenging.
- The estimation of CT models with time-delay from non-uniformly sampled data in presence of colored measurement noise.

Appendix A

Computation of the Sampled Data Model for Stochastic Systems

A.1 Computation of Matrix Exponentials

The sampled data model (2.33) contains several integrals involving matrix exponentials, which seem hard to compute. Fortunately, [Van78] and [LW10] provided an efficient and compact way to compute these terms. The formulae are given in the following theorem

Theorem A.1 (See [Van78])**.** *Let n_1, n_2, n_3, and n_4 be positive integers, and set m to be their sum, if the $m \times m$ block triangular matrix $\boldsymbol{W}$ is defined by*

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{B}_1 & \boldsymbol{C}_1 & \boldsymbol{D}_1 \\ 0 & \boldsymbol{A}_2 & \boldsymbol{B}_2 & \boldsymbol{C}_2 \\ 0 & 0 & \boldsymbol{A}_3 & \boldsymbol{B}_3 \\ 0 & 0 & 0 & \boldsymbol{A}_4 \end{bmatrix} \begin{matrix} \}n_1 \\ \}n_2 \\ \}n_3 \\ \}n_4 \end{matrix} \tag{A.1}$$

$$\underbrace{}_{n_1} \underbrace{}_{n_2} \underbrace{}_{n_3} \underbrace{}_{n_4} \tag{A.2}$$

then for $t \geq 0$

$$e^{\boldsymbol{W}t} = \begin{bmatrix} \boldsymbol{F}_{1,t} & \boldsymbol{G}_{1,t} & \boldsymbol{H}_{1,t} & \boldsymbol{K}_{1,t} \\ 0 & \boldsymbol{F}_{2,t} & \boldsymbol{G}_{2,t} & \boldsymbol{H}_{2,t} \\ 0 & 0 & \boldsymbol{F}_{3,t} & \boldsymbol{G}_{3,t} \\ 0 & 0 & 0 & \boldsymbol{F}_{4,t} \end{bmatrix} \tag{A.3}$$

where

$$\boldsymbol{F}_{j,t} = e^{\boldsymbol{A}_j t} \tag{A.4}$$

$$\boldsymbol{G}_{j,t} = \int_0^t e^{\boldsymbol{A}_j(t-s)} \boldsymbol{B}_j e^{\boldsymbol{A}_{j+1}s} ds \tag{A.5}$$

$$\boldsymbol{H}_{j,t} = \int_0^t e^{\boldsymbol{A}_j(t-s)} \boldsymbol{C}_j e^{\boldsymbol{A}_{j+2}s} ds + \int_0^t \int_0^s e^{\boldsymbol{A}_j(t-s)} \boldsymbol{B}_j e^{\boldsymbol{A}_{j+1}(s-r)} \boldsymbol{B}_{j+1} e^{\boldsymbol{A}_{j+2}r} dr\, ds \tag{A.6}$$

$$\boldsymbol{K}_{j,t} = \int_0^t e^{\boldsymbol{A}_j(t-s)} \boldsymbol{D}_j e^{\boldsymbol{A}_{j+3}s} ds$$

$$
\begin{aligned}
&+ \int_0^t \int_0^s e^{\boldsymbol{A}_j(t-s)} \boldsymbol{C}_j e^{\boldsymbol{A}_{j+2}(s-r)} \boldsymbol{B}_{j+2} e^{\boldsymbol{A}_{j+3} r} dr\ ds \\
&+ \int_0^t \int_0^s e^{\boldsymbol{A}_j(t-s)} \boldsymbol{B}_j e^{\boldsymbol{A}_{j+1}(s-r)} \boldsymbol{C}_{j+1} e^{\boldsymbol{A}_{j+3} r} dr\ ds \\
&+ \int_0^t \int_0^s \int_0^r e^{\boldsymbol{A}_j(t-s)} \boldsymbol{B}_j e^{\boldsymbol{A}_{j+1}(s-r)} \boldsymbol{B}_{j+1} e^{\boldsymbol{A}_{j+2}(r-w)} \boldsymbol{B}_{j+2} e^{\boldsymbol{A}_{j+3} w} dw\ dr\ ds
\end{aligned} \tag{A.7}
$$

Proof. See [Van78].

□

A.2 Computation of Sampled Data Models

Subsequently, we apply Theorem A.1 to the following matrix

$$
\boldsymbol{W} = \begin{pmatrix} -\boldsymbol{F}_c & \boldsymbol{I} & 0 & 0 & 0 \\ 0 & -\boldsymbol{F}_c & \boldsymbol{Q}_c & 0 & 0 \\ 0 & 0 & \boldsymbol{F}_c^T & \boldsymbol{I} & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{I} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{A.8}
$$

$$
e^{\boldsymbol{W} h_k} = \begin{pmatrix} \boldsymbol{F}_1(h_k) & \boldsymbol{G}_1(h_k) & \boldsymbol{H}_1(h_k) & \boldsymbol{K}_1(h_k) & \boldsymbol{M}_1(h_k) \\ 0 & \boldsymbol{F}_2(h_k) & \boldsymbol{G}_2(h_k) & \boldsymbol{H}_2(h_k) & \boldsymbol{K}_2(h_k) \\ 0 & 0 & \boldsymbol{F}_3(h_k) & \boldsymbol{G}_3(h_k) & \boldsymbol{H}_3(h_k) \\ 0 & 0 & 0 & \boldsymbol{F}_4(h_k) & \boldsymbol{G}_4(h_k) \\ 0 & 0 & 0 & 0 & \boldsymbol{F}_5(h_k) \end{pmatrix} \tag{A.9}
$$

Similar to (A.1), we rename the integrals involved in $\boldsymbol{Q}_{h_k}$, $\boldsymbol{R}_{h_k}$ and $\boldsymbol{S}_{h_k}$ as

$$
\boldsymbol{Q}_t = \int_0^t \boldsymbol{F}_s \boldsymbol{Q}_c \boldsymbol{F}_s^T ds \tag{A.10}
$$

$$
\bar{\boldsymbol{Q}}_t = \int_0^t \boldsymbol{F}_s \boldsymbol{Q}_c \bar{\boldsymbol{F}}_s^T ds \tag{A.11}
$$

$$
\bar{\bar{\boldsymbol{Q}}}_t = \int_0^t \bar{\boldsymbol{F}}_s \boldsymbol{Q}_c \bar{\boldsymbol{F}}_s^T ds \tag{A.12}
$$

The matrices can then be given as

$$
\boldsymbol{F}_{h_k} = \boldsymbol{F}_3^T(h_k) \tag{A.13}
$$

$$
\bar{\boldsymbol{F}}_{h_k} = \boldsymbol{G}_3^T(h_k) \tag{A.14}
$$

$$
\bar{\bar{\boldsymbol{F}}}_{h_k} = \boldsymbol{H}_3^T(h_k) \tag{A.15}
$$

$$
\boldsymbol{Q}_{h_k} = \boldsymbol{F}_{h_k} \boldsymbol{G}_2(h_k) \tag{A.16}
$$

$$
\bar{\boldsymbol{Q}}_{h_k} = \boldsymbol{F}_{h_k} \boldsymbol{H}_2(h_k) \tag{A.17}
$$

$$
\bar{\bar{\boldsymbol{Q}}}_{h_k} = \boldsymbol{F}_{h_k} \boldsymbol{K}_1(h_k) + \boldsymbol{K}_1^T(h_k) \boldsymbol{F}_{h_k}^T \tag{A.18}
$$

Based on the above results, the remainder matrices of (2.33) are given as

$$\boldsymbol{H}_{h_k} = \frac{1}{h_k}\boldsymbol{H}_c\bar{\boldsymbol{F}}_{h_k} \tag{A.19}$$

$$\boldsymbol{R}_{h_k} = \frac{1}{h_k^2}\left(\boldsymbol{H}_c\bar{\bar{\boldsymbol{Q}}}_{h_k}\boldsymbol{H}_c^T + \boldsymbol{H}_c\bar{\bar{\boldsymbol{F}}}_{h_k}\boldsymbol{S}_c + \boldsymbol{S}_c^T\bar{\bar{\boldsymbol{F}}}_{h_k}^T\boldsymbol{H}_c^T\right) + \frac{1}{h_k}\boldsymbol{R}_c \tag{A.20}$$

$$\boldsymbol{S}_{h_k} = \frac{1}{h_k}\left(\bar{\boldsymbol{Q}}_{h_k}\boldsymbol{H}_c^T + \bar{\boldsymbol{F}}_{h_k}\boldsymbol{S}_c\right) \tag{A.21}$$

Appendix B

Fast approximation of time-varying matrices

In Kalman filtering, computations of $\boldsymbol{F}_{h_k}$, $\boldsymbol{Q}_{h_k}$ are needed. This issue has been addressed in [Van78, LW10], a computational efficient method was proposed, and it is used in this paper.

By defining the following matrix and taking its exponential

$$\boldsymbol{M} = \begin{bmatrix} -\boldsymbol{F}_c & \boldsymbol{Q}_c \\ \boldsymbol{0} & \boldsymbol{F}_c^T \end{bmatrix} \tag{B.1}$$

$$\boldsymbol{M}_h = e^{\boldsymbol{M}h} = \begin{bmatrix} \boldsymbol{F}_1 & \boldsymbol{G}_1 \\ \boldsymbol{0} & \boldsymbol{F}_2 \end{bmatrix} \tag{B.2}$$

$\boldsymbol{F}_h$, $\boldsymbol{Q}_h$ are immediately obtained as follows (see [Van78, LW10])

$$\boldsymbol{F}_h = \boldsymbol{F}_2^T, \ \boldsymbol{Q}_h = \boldsymbol{F}_{h_k}\boldsymbol{G}_1 \tag{B.3}$$

The main work of the method is to solve (B.2), this can be easily done by using the `expm` routine in Matlab. However, in non-uniformly sampling case, the computation of $\boldsymbol{F}_{h_k}$, $\boldsymbol{Q}_{h_k}$ become a heavy burden in using `expm`. In this paper, the following method is used to approximate the time varying matrices $\boldsymbol{F}_{h_k}$, $\boldsymbol{Q}_{h_k}$ efficiently

1. First the following decomposition of h_k is performed

   $$h_k = jh_s + \Delta_k < (j+1)h_s + \Delta_k \tag{B.4}$$

   where j is a positive integer

2. Then $\boldsymbol{F}_{h_k}$, $\boldsymbol{Q}_{h_k}$ can be obtained by

   $$\boldsymbol{F}_{h_k} = \boldsymbol{F}_{jh_s}\boldsymbol{F}_{\Delta_k} \tag{B.5}$$

   $$\boldsymbol{Q}_{h_k} = \boldsymbol{Q}_{jh_s} + \boldsymbol{F}_{jh_s}\boldsymbol{Q}_{\Delta_k}\boldsymbol{F}_{jh_s}^T \tag{B.6}$$

3. $\boldsymbol{F}_{\Delta_k}$, $\boldsymbol{Q}_{\Delta_k}$ can be obtained by solving (B.2) and (B.3) using some a fast approximation method. In this paper, the 4th order Runge-Kutta (RK4) method is applied.

4. $\boldsymbol{F}_{jh_s}, \boldsymbol{Q}_{jh_s}$ can be precomputed and stored in a numerical array using the following recursion

$$\begin{aligned}\boldsymbol{F}_{jh_s+h_s} &= \boldsymbol{F}_{jh_s}\boldsymbol{F}_{hs} \\ \boldsymbol{Q}_{jh_s+h_s} &= \boldsymbol{Q}_{jh_s} + \boldsymbol{F}_{jh_s}\boldsymbol{Q}_{h_s}\boldsymbol{F}_{jh_s}^T\end{aligned}$$

Note that in above procedure, $\boldsymbol{F}_{h_s}, \boldsymbol{Q}_{h_s}$ can be precisely computed in advance. The main computational burden reduces to approximate $\boldsymbol{F}_{\Delta_k}, \boldsymbol{Q}_{\Delta_k}$ using the RK4 method.

Appendix C

Proofs

C.1 Proof of Lemma 3.1

Proof. From (3.67)

$$\mathscr{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^0) = \mathbb{E}_{\hat{\boldsymbol{\theta}}^0} \left\{ \log p_{\boldsymbol{\theta}} \left(\boldsymbol{z}_{t_{1:N}} \right) | \xi_{t_{1:N}} \right\} \tag{C.1}$$

Since $\boldsymbol{z}_t$ is a Markovian process, the current state $\boldsymbol{z}_{t_k}$ contains all the information up to time instant t_k, therefore the prediction of the future state is only dependent on $\boldsymbol{z}_{t_k}$

$$p_{\boldsymbol{\theta}} \left(\boldsymbol{z}_{t_{k+1}} | \boldsymbol{z}_{t_{1:k}} \right) = p_{\boldsymbol{\theta}} \left(\boldsymbol{z}_{t_{k+1}} | \boldsymbol{z}_{t_k} \right) \tag{C.2}$$

Given the fact that

$$\boldsymbol{z}_{t_1} \sim \mathscr{N} \left(\boldsymbol{\mu}_{t_1}, \boldsymbol{P}_{t_1} \right) \tag{C.3}$$

$$\boldsymbol{z}_{t_{k+1}} \sim \mathscr{N} \left(\boldsymbol{F}_{h_k} \boldsymbol{z}_{t_k}, \boldsymbol{Q}_{h_k} \right) \tag{C.4}$$

$$\boldsymbol{z}_{t_{k+1}} = h_k \boldsymbol{z}_{t_k}^{\delta} + \boldsymbol{z}_{t_k} \tag{C.5}$$

$$\boldsymbol{F}_{h_k} = \boldsymbol{I} + h_k \boldsymbol{F}_{h_k}^{\delta} \tag{C.6}$$

$$\boldsymbol{Q}_{h_k} = h_k \boldsymbol{Q}_{h_k}^{\delta} \tag{C.7}$$

Then

$$\begin{aligned}
&-2\log p_{\boldsymbol{\theta}} \left(\boldsymbol{z}_{t_{1:N}} \right) \\
&= -2\log \left[p_{\boldsymbol{\theta}} \left(\boldsymbol{z}_{t_1} \right) \prod_{k=1}^{N-1} p_{\boldsymbol{\theta}} \left(\boldsymbol{z}_{t_{k+1}} | \boldsymbol{z}_{t_k} \right) \right] = \log p_{\boldsymbol{\theta}} \left(\boldsymbol{z}_{t_1} \right) + \sum_{k=1}^{N-1} \log p_{\boldsymbol{\theta}} \left(\boldsymbol{z}_{t_{k+1}} | \boldsymbol{z}_{t_k} \right) \\
&= \log \det \boldsymbol{P}_{t_1} + \sum_{k=1}^{N-1} \log \det \boldsymbol{Q}_{h_k} + \left(\boldsymbol{z}_{t_1} - \boldsymbol{\mu}_{t_1} \right)^T \boldsymbol{P}_{t_1}^{-1} \left(\boldsymbol{z}_{t_1} - \boldsymbol{\mu}_{t_1} \right) \\
&\quad + \sum_{k=1}^{N-1} \left(\boldsymbol{z}_{t_{k+1}} - \boldsymbol{F}_{h_k} \boldsymbol{z}_{t_k} \right)^T \boldsymbol{Q}_{h_k}^{-1} \left(\boldsymbol{z}_{t_{k+1}} - \boldsymbol{F}_{h_k} \boldsymbol{z}_{t_k} \right) \\
&= \log \det \boldsymbol{P}_{t_1} + \sum_{k=1}^{N-1} \log \det \left(h_k \boldsymbol{Q}_{h_k}^{\delta} \right) + \left(\boldsymbol{z}_{t_1} - \boldsymbol{\mu}_{t_1} \right)^T \boldsymbol{P}_{t_1}^{-1} \left(\boldsymbol{z}_{t_1} - \boldsymbol{\mu}_{t_1} \right)
\end{aligned}$$

$$+\sum_{k=1}^{N-1}\left(\boldsymbol{z}_{t_k}^{\delta}-\boldsymbol{F}_{h_k}^{\delta}\boldsymbol{z}_{t_k}\right)^T\left(h_k^{-1}\boldsymbol{Q}_{h_k}^{\delta}\right)^{-1}\left(\boldsymbol{z}_{t_k}^{\delta}-\boldsymbol{F}_{h_k}^{\delta}\boldsymbol{z}_{t_k}\right) \tag{C.8}$$

By taking expectations on both sides of (C.8) and using the rule of trace, we obtain the result presented in Lemma 3.1. □

C.2 Proof of Lemma 3.2

Proof. Replacing $\boldsymbol{F}_{h_k}^{\delta}$ and $\boldsymbol{Q}_{h_k}^{\delta}$ in $\mathscr{Q}(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}^0)$ by $\boldsymbol{F}_c$ and $\boldsymbol{Q}_c$ leads to

$$\begin{aligned}-2\mathscr{Q}(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}^0)=&\sum_{k=1}^{N-1}\mathrm{Tr}\,\left\{h_k\boldsymbol{Q}_c^{-1}\left(\boldsymbol{\Phi}_{t_k}-\boldsymbol{\Psi}_{t_k}\boldsymbol{F}_c^T-\boldsymbol{F}_c\boldsymbol{\Psi}_{t_k}^T+\boldsymbol{F}_c\boldsymbol{\Upsilon}_{t_k}\boldsymbol{F}_c^T\right)\right\}\\&-\sum_{k=1}^{N-1}\log\det\left(h_k\boldsymbol{Q}_c^{-1}\right)+\cdots\end{aligned} \tag{C.9}$$

Then differentiate $-2\mathscr{Q}(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}^0)$ with respect to $\boldsymbol{F}_c$ and set the derivative to zero, then we have

$$-2\cdot\frac{\partial\mathscr{Q}(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}^0)}{\partial\boldsymbol{F}_c}=2\boldsymbol{Q}_c^{-1}\sum_{k=1}^{N-1}\left(-h_k\boldsymbol{\Psi}_{t_k}+h_k\boldsymbol{F}_c\boldsymbol{\Upsilon}_{t_k}\right)=0 \tag{C.10}$$

Since the unknown parameters are stacked in the first row of $\boldsymbol{F}_c$. We consider the first row of (C.10), let $\boldsymbol{W}=[1\ \ \boldsymbol{0}]^T$, the estimate of $\boldsymbol{\eta}$ can be given by

$$\hat{\boldsymbol{\eta}}=\sum_{k=1}^{N-1}h_k\boldsymbol{W}^T\boldsymbol{\Psi}_{t_k}\left(\sum_{k=1}^{N-1}h_k\boldsymbol{\Upsilon}_{t_k}\right)^{-1} \tag{C.11}$$

Differentiate $-2\mathscr{Q}(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}^0)$ with respect to $\boldsymbol{Q}_c^{-1}$ and set the derivative to zero, it follows that

$$-2\cdot\frac{\partial\mathscr{Q}(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}^0)}{\partial\boldsymbol{Q}_c^{-1}}=-\sum_{k=1}^{N-1}\boldsymbol{Q}_c+\sum_{k=1}^{N-1}h_k\left(\boldsymbol{\Phi}_{t_k}-\boldsymbol{\Psi}_{t_k}\boldsymbol{F}_c^T-\boldsymbol{F}_c\boldsymbol{\Psi}_{t_k}^T+\boldsymbol{F}_c\boldsymbol{\Upsilon}_{t_k}\boldsymbol{F}_c^T\right)=0 \tag{C.12}$$

According to the relationship between $\boldsymbol{Q}_c$ and σ^2, the estimate of $\hat{\sigma}^2$ can be given by

$$\hat{\sigma}^2=\frac{1}{N-1}\sum_{k=1}^{N-1}h_k\left(\boldsymbol{W}^T\boldsymbol{\Phi}_{t_k}\boldsymbol{W}-\hat{\boldsymbol{\eta}}^T\boldsymbol{\Upsilon}_{t_k}\hat{\boldsymbol{\eta}}\right) \tag{C.13}$$

□

C.3 Proof of Proposition 3.5

Proof. Prediction: (3.74)∼(3.75) are straightforward from the Kalman filter.

Smoothing: The geometric way (see *i.e.* [AK82]) is used to derive (3.76)∼(3.78). Since there is no measurement available at $t_{k,i}$, the predictions of $\boldsymbol{z}_{k+1}$ from $\hat{\boldsymbol{z}}^k_{k,i}$ and $\hat{\boldsymbol{z}}^k_k$ will be completely the same. For (3.78), according to equation (4) in [AK82], we have

$$\boldsymbol{J}_{k,i} = \text{cov}\big\{\boldsymbol{z}_{k,i}, \boldsymbol{z}_{k+1} - \hat{\boldsymbol{z}}^k_{k+1}\big\}\big(\boldsymbol{P}^k_{k+1}\big)^{-1} \tag{C.14}$$

Subsequently (C.14) becomes

$$\begin{aligned}\boldsymbol{J}_{k,i} =& \mathbb{E}\big\{\left(\boldsymbol{F}_{\tau_{k,i}}\boldsymbol{z}_k + \boldsymbol{w}_{\tau_{k,i}}\right)\left(\boldsymbol{F}_{h_k}\tilde{\boldsymbol{z}}^k_k + \boldsymbol{w}_{h_k}\right)^T\big\}\big(\boldsymbol{P}^k_{k+1}\big)^{-1} \\ =& \left[\boldsymbol{F}_{\tau_{k,i}}\boldsymbol{P}^k_k\boldsymbol{F}^T_{h_k} + \mathbb{E}\left\{\boldsymbol{w}_{\tau_{k,i}}\boldsymbol{w}^T_{h_k}\right\}\right]\left(\boldsymbol{P}^k_{k+1}\right)^{-1}\end{aligned} \tag{C.15}$$

where

$$\begin{aligned}\boldsymbol{w}_{\tau_{k,i}} &= \int_{t_k}^{t_{k,i}} e^{\boldsymbol{F}_c(t_{k,i}-t)}\boldsymbol{w}_t dt \\ \boldsymbol{w}_{h_k} &= \int_{t_k}^{t_{k+1}} e^{\boldsymbol{F}_c(t_{k+1}-t)}\boldsymbol{w}_t dt\end{aligned}$$

The expectation term involved in (C.15) becomes

$$\begin{aligned}\mathbb{E}\left\{\boldsymbol{w}_{\tau_{k,i}}\boldsymbol{w}^T_{h_k}\right\} =& \int_{t_k}^{t_{k,i}} e^{\boldsymbol{F}_c(t_{k,i}-t)}\boldsymbol{Q}_c e^{\boldsymbol{F}^T_c(t_{k+1}-t)}dt \\ =& \boldsymbol{Q}_{\tau_{k,i}}\boldsymbol{F}^T_{h_k-\tau_{k,i}}\end{aligned} \tag{C.16}$$

By combining (C.15) and (C.16), the smoother gain can be given as

$$\boldsymbol{J}_{k,i} = \big(\boldsymbol{F}_{\tau_{k,i}}\boldsymbol{P}^k_k\boldsymbol{F}^T_{h_k} + \boldsymbol{Q}_{\tau_{k,i}}\boldsymbol{F}^T_{h_k-\tau_{k,i}}\big)\big(\boldsymbol{P}^k_{k+1}\big)^{-1} \tag{C.17}$$

For (3.77), Property 6.2 in [SS11] can be used to derive this equation.

Lag-one covariance: Similar to Property 6.3 in [SS11], by subtracting $\boldsymbol{z}_{k,i+1}$ $(i < n_k)$ on both side of (3.76) and rearranging the terms, we have

$$\tilde{\boldsymbol{z}}^N_{k,i+1} + \boldsymbol{J}_{k,i+1}\hat{\boldsymbol{z}}^N_{k+1} = \tilde{\boldsymbol{z}}^k_{k,i+1} + \boldsymbol{J}_{k,i+1}\hat{\boldsymbol{z}}^k_{k+1} \tag{C.18}$$

Similarly

$$\tilde{\boldsymbol{z}}^N_{k,i} + \boldsymbol{J}_{k,i}\hat{\boldsymbol{z}}^N_{k+1} = \tilde{\boldsymbol{z}}^k_{k,i} + \boldsymbol{J}_{k,i}\hat{\boldsymbol{z}}^k_{k+1} \tag{C.19}$$

Multiplying each side of (C.18) by transpose of (C.19) and take expectation, we have (the cross terms are zero)

$$\begin{aligned}&\boldsymbol{M}^N_{k,i} + \boldsymbol{J}_{k,i+1}\mathbb{E}\big\{\hat{\boldsymbol{z}}^N_{k+1}(\hat{\boldsymbol{z}}^N_{k+1})^T\big\}\boldsymbol{J}^T_{k,i} \\ =& \boldsymbol{F}_{\tau_{k,i}}\boldsymbol{P}^k_k + \boldsymbol{J}_{k,i+1}\mathbb{E}\big\{\hat{\boldsymbol{z}}^k_{k+1}(\hat{\boldsymbol{z}}^k_{k+1})^T\big\}\boldsymbol{J}^T_{k,i}\end{aligned} \tag{C.20}$$

Provided that

$$\mathbb{E}\big\{\hat{\boldsymbol{z}}^N_{k+1}(\hat{\boldsymbol{z}}^N_{k+1})^T\big\} = \mathbb{E}\big\{\boldsymbol{z}_{k+1}\boldsymbol{z}^T_{k+1}\big\} - \boldsymbol{P}^N_{k+1} \tag{C.21}$$

$$\mathbb{E}\big\{\hat{\boldsymbol{z}}^k_{k+1}(\hat{\boldsymbol{z}}^k_{k+1})^T\big\} = \mathbb{E}\big\{\boldsymbol{z}_{k+1}\boldsymbol{z}^T_{k+1}\big\} - \boldsymbol{P}^k_{k+1} \tag{C.22}$$

Combining (C.20)∼(C.22) leads to (3.79)

$$\boldsymbol{M}^N_{k,i} = \boldsymbol{F}_{\tau_{k,i}}\boldsymbol{P}^k_k + \boldsymbol{J}_{k,i+1}(\boldsymbol{P}^N_{k+1} - \boldsymbol{P}^k_{k+1})\boldsymbol{J}^T_{k,i} \tag{C.23}$$

When $i = n_k$, Equation (C.18) becomes

$$\tilde{\boldsymbol{z}}^N_{k+1} + \hat{\boldsymbol{z}}^N_{k+1} = \tilde{\boldsymbol{z}}^k_{k+1} + \hat{\boldsymbol{z}}^k_{k+1} \tag{C.24}$$

Then, Equation (3.80) can be proved in the same way. □

Appendix D

Digital Implementation of the Continuous-time Filtering Operation

The filtering operation is carried out in state-space form. The equivalent state-space model is a first order differential equation and many mature algorithms are available to solve this equation.

Consider

$$u_{f,t_k} = \frac{1}{A(p,\boldsymbol{\rho})} u_{t_k} \tag{D.1}$$

First write (D.1) into following state-space representation

$$\begin{cases} \dot{\boldsymbol{z}}_{t_k} = \boldsymbol{F}_c \boldsymbol{z}_{t_k} + \boldsymbol{G}_c u_{t_k} \\ u_{f,t_k} = \boldsymbol{H}_c \boldsymbol{z}_{t_k} \end{cases} \tag{D.2}$$

where $\dot{z}_{t_k}$ denotes $\dfrac{dz_{t_k}}{dt}$ and

$$\boldsymbol{F}_c = \begin{bmatrix} -a_1 & \cdots & \cdots & -a_{n_a} \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & 1 & 0 \end{bmatrix}, \boldsymbol{G}_c = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{H}_c^T = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

Model (D.2) has an DT counterpart (assume u_{t_k} is 'ZOH')

$$\begin{cases} \boldsymbol{z}_{t_{k+1}} = \boldsymbol{F}_{h_k} \boldsymbol{z}_{t_k} + \boldsymbol{G}_{h_k} u_{t_k} \\ u_{f,t_{k+1}} = \boldsymbol{H}_c z_{t_{k+1}} \end{cases} \tag{D.3}$$

where

$$\boldsymbol{F}_{h_k} = e^{\boldsymbol{F}_c h_k}, \ \boldsymbol{G}_{h_k} = \int_0^{h_k} e^{\boldsymbol{F}_c t} dt \ \boldsymbol{G}_c$$

In irregular sampling case, (D.3) is time-varying, the computation loads of F_{h_k} and $\boldsymbol{G}_{h_k}$ are quite heavy when using some standard method, such as the `expm` routine in Matlab. Therefore

the following fourth-order Runge-Kutta (RK4) method is used to approximate them. Let us construct a matrix $\boldsymbol{W}$ according to Theorem A.1, $\boldsymbol{F}_{h_k}$, $\boldsymbol{G}_{h_k}$ can be obtained by computing

$$\boldsymbol{M}_{h_k} = e^{\boldsymbol{W} h_k} \tag{D.4}$$

which is the solution of the following differential equation

$$\dot{\boldsymbol{M}}_t = \boldsymbol{W}\boldsymbol{M}_t \tag{D.5}$$

Let $\boldsymbol{M}_0 = \boldsymbol{I}$, $\boldsymbol{M}_{h_k}$ can be approximated by the RK4 method

$$\boldsymbol{M}_{h_k} = \boldsymbol{I} + \frac{h_k}{6}(\boldsymbol{R}^1_{h_k} + 2\boldsymbol{R}^2_{h_k} + 2\boldsymbol{R}^3_{h_k} + \boldsymbol{R}^4_{h_k}) \tag{D.6}$$

where

$$\begin{aligned}
\boldsymbol{R}^1_{h_k} &= \boldsymbol{W}\boldsymbol{M}_0 \\
\boldsymbol{R}^2_{h_k} &= \boldsymbol{W}\left(\boldsymbol{M}_0 + 0.5\boldsymbol{R}^1_{h_k} h_k\right) \\
\boldsymbol{R}^3_{h_k} &= \boldsymbol{W}\left(\boldsymbol{M}_0 + 0.5\boldsymbol{R}^2_{h_k} h_k\right) \\
\boldsymbol{R}^4_{h_k} &= \boldsymbol{W}\left(\boldsymbol{M}_0 + \boldsymbol{R}^3_{h_k} h_k\right)
\end{aligned}$$

The truncation error of the RK4 method is $O(h^5)$. Sometimes this is not enough, the following way is used to reduce the approximation error

1. First specify the maximum step length h_s for the RK4 method. h_k can be decomposed as

   $$h_k = jh_s + \delta_k$$

   where $j \geq 0$ is a no-negative integer.

2. Evaluate $\boldsymbol{F}_{\delta_k}$, $\boldsymbol{G}_{\delta_k}$, using the RK4 method.

3. Compute $\boldsymbol{F}_{h_k}$, $\boldsymbol{G}_{h_k}$

   $$\begin{aligned}
   \boldsymbol{F}_{h_k} &= \boldsymbol{F}_{jh_s}\boldsymbol{F}_{\delta_k} \\
   \boldsymbol{G}_{h_k} &= \boldsymbol{G}_{jh_s} + \boldsymbol{F}_{jh_s}\boldsymbol{G}_{\delta_k}
   \end{aligned}$$

4. Compute $\boldsymbol{z}_{t_{k+1}}$

   $$\boldsymbol{z}_{t_{k+1}} = \boldsymbol{F}_{h_k}\boldsymbol{z}_{t_k} + \boldsymbol{G}_{h_k} u_{t_k}$$

Note that $\{\boldsymbol{F}_{jh_s}, \boldsymbol{G}_{jh_s},\ j = 1, 2, \cdots\}$ involved in Step 2 can be precomputed and stored in a numerical array by the following recursion

$$\begin{aligned}
\boldsymbol{F}_{jh_s} &= \boldsymbol{F}_{h_s}\boldsymbol{F}_{jh_s - h_s} \\
\boldsymbol{G}_{jh_s} &= \boldsymbol{F}_{h_s}\boldsymbol{G}_{jh_s - h_s} + \boldsymbol{G}_{h_s}
\end{aligned}$$

Bibliography

[ÅB03] K. J. Åström and B. Bernhardsson. Systems with Lebesgue sampling. In *Directions in Mathematical Systems Theory and Optimization (A. Rantzer, C.I. Byrnes (Eds.))*, pages 1–13, Berlin Heidelberg, 2003. Springer-Verlag.

[AHS06a] S. Ahmed, B. Huang, and S. L. Shah. Novel identification method from step response. *Control Engineering Practice*, 15(5):545–556, 2006.

[AHS06b] S. Ahmed, B. Huang, and S. L. Shah. Parameter and delay estimation of continuous-time models using a linear filter. *Journal of Process Control*, 16(4):323–331, 2006.

[AK82] C. F. Ansley and R. Kohn. A geometrical derivation of the fixed interval smoothing algorithm. *Biometrika*, 69(2):486–487, 1982.

[Åst70] K. J. Åström. *Introduction to Stochastic Control Theory*. Academic Press, New York, 1970.

[ATY+12] J. C. Agüero, W. Tang, J. I. Yuz, R. Delgado, and G. C. Goodwin. Dual time-frequency domain system identification. *Automatica*, 48:3031–3041, 2012.

[AYG07] J. C. Aguëro, J. I. Yuz, and G. C. Goodwin. Frequency domain identification of MIMO state space models using the EM algorithm. In *European Control Conference 2007*, Kos, Greece, 2–5 July 2007.

[BH97] R. G. Brown and P. Y. C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering: with MATLAB Exercises and Solutions*. John Wiley & Sons, New York, third edition, 1997.

[Bjö03] S. Björklund. *A survey and comparison of time-delay estimation methods in linear systems*. Ph.D. thesis, Linköping University, Sweden, 2003.

[Boh06] T. Bohlin. *Practical Grey-box Process Identification. Theory and Applications*. Advanced in Industrial Control. Springer, 2006.

[CGG14] F. Chen, H. Garnier, and M. Gilson. Robust identification of continuous-time models with arbitrary time-delay from irregularly sampled data. *Journal of Process Control*, 2014.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[EG08] F. Eng and F. Gustafsson. Identification with stochastic sampling time jitter. *Automatica*, 44(3):637–646, 2008.

[Eng07] F. Eng. *Non-Uniform Sampling in Statistical Signal Processing*. Phd thesis, Department of Electrical Engineering, Linköping university, Sweden, 2007.

[FG96] A. Feuer and G. C. Goodwin. *Sampling in Digital Signal Processing and Control*. Birkhäuser, Boston, 1996.

[FMS96] G. Ferretti, C. Maffezzoni, and R. Scattolini. On the identifiability of the time delay with least-squares methods. *Automatica*, 32(3):449–453, 1996.

[FSM+99] H. Fan, T. Söderström, M. Mossberg, B. Carlsson, and Y. Zhou. Estimation of continuous-time AR process parameters from discrete-time data. *IEEE Transactions on Signal Processing*, 47(5):1232–1244, 1999.

[GA05] G. C. Goodwin and J. C. Agüero. Approximate EM algorithms for parameter and state estimation in nonlinear stochastic models. In *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, Seville, Spain, December 12-15 2005.

[GAG+13] G. C. Goodwin, J. C. Agüero, M. E. Cea Garrido, M. E. Salgado, and J. I. Yuz. Sampling and sampled-data models. *IEEE Control Systems Magazine*, pages 34–53, October 2013.

[GC12] G. C. Goodwin and M.G. Cea. Application of minimum distortion filtering to identification of linear systems having non-uniform sampling period. In *System Identification, Environmental Modelling, and Control System Design (L. Wang and H. Garnier (Eds.))*, pages 197–114, London, 2012. Springer-Verlag.

[GGYdH08] M. Gilson, H. Garnier, P. C. Young, and P. Van den Hof. Instrumental variable methods for closed-loop continuous-time model identification. In *Identification of Continuous-time Models from Sampled Data, H. Garnier and L. Wang (Eds.)*, page 133–160, London, 2008. Springer-Verlag.

[GGYH07] H. Garnier, M. Gilson, P. C. Young, and E. Huselstein. An optimal IV technique for identifying continuous-time transfer function model of multiple input systems. *Control Engineering Practice*, 15(4):471–486, 2007.

[GGYV08] M. Gilson, H. Garnier, P. C. Young, and P. Van den Hof. Instrumental variable methods for closed-loop continuous-time model identification. In *Identification of Continuous-time Models from Sampled Data (H. Garnier and L. Wang (Eds.))*, pages 133–160, London, 2008. Springer.

[GL09] J. Gillberg and L. Ljung. Frequency-domain identification of continuous-time ARMA models from sampled data. *Automatica*, 45:1371–1378, 2009.

[GN05] S. Gibson and B. Ninness. Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica*, 41:1667–1682, 2005.

[GY14] H. Garnier and P. C. Young. The advantages of directly identifying continuous-time transfer function models in practical applications. *International Journal of Control*, 87(7):1319–1338, 2014.

[Isa93] A. Isaksson. Identification of ARX models subject to missing data. *IEEE Transactions on Automatic Control*, 38(5):813–819, 1993.

[Jaz70] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York and London, 1970.

[Joh94] R. Johansson. Identification of continuous-time models. *IEEE Transactions on Signal Processing*, 42(4):887–897, 1994.

[Joh09] R. Johansson. Continuous-time model identification and state estimation using In *15th IFAC Symposium on System Identification*, Saint-Malo, France, 6-8 July 2009.

[KA08] V. Kadirkamanathan and S. Anderson. Maximum–likelihood estimation of delta-domain model parameters from noisy output signals. *IEEE Transactions on Signal Processing*, 56(8):3765–3770, 2008.

[KMU11] H. Kirshner, S. Maggio, and M. Unser. A sampling theory approach for continuous ARMA identification. *IEEE Transactions on Signal Processing*, 59(10):4620–4634, 2011.

[KUW14] H. Kirshner, M. Unser, and J. P. Ward. On the unique identification of continuous-time uutoregressive models from sampled data. *IEEE Transactions on Signal Processing*, 62(6):1361–1376, 2014.

[Lar05] E. K. Larsson. Limiting sampling results for continuous-time arma systems. *International Journal of Control*, 78(7):461–473, 2005.

[LG08] T. Liu and F. Gao. Robust step-like identification of low order process model under nonzero initial conditions and disturbance. *IEEE Transactions on Automatic Control*, 53(11):2690–2695, 2008.

[LG10] T. Liu and F. Gao. A frequency domain step response identification method for continuous-time processes with time delay. *Journal of Process Control*, 20(7):800–809, 2010.

[Lju99] L. Ljung. *System Identification – Theory for the User*. Prentice-Hall, Upper Saddle River, 2nd edition edition, 1999.

[Lju12] L. Ljung. Version 8 of the system identification toolbox. In *16th IFAC Symposium on System Identification*, Brussels, Belgium, 11-13 July 2012.

[LMS07] E. K. Larsson, M. Mossberg, and T. Söderström. Identification of continuous-time ARX models from irregularly sampled data. *IEEE Transactions on Automatic Control*, 52(3):417–427, 2007.

[LS02] E. K. Larsson and T. Söderström. Identifcation of continuous-time AR processes from unevenly sampled data. *Automatica*, 38:709–718, 2002.

[LW10] L. Ljung and A. Wills. Issues in sampling and estimating continuous-time models with stochastic disturbances. *Automatica*, 46:925–931, 2010.

[MG90] R. H. Middleton and G. C. Goodwin. *Digital Control and Estimation*. Prentice-Hall, Englewood Cliffs, NJ, 1990.

[Mos08a] M. Mossberg. Estimation of continuous-time stochastic signals from sample covariances. *IEEE Transactions on Signal Processing*, 56(2):821–825, 2008.

[Mos08b] M. Mossberg. High-accuracy instrumental variable identification of continuous-time autoregressive processes from irregularly sampled noisy data. *IEEE Transactions on Signal Processing*, 56(8):4087–4091, 2008.

[Ngi01] L. S. H. Ngia. Separable nonlinear least-squares methods for efficient off-line and on-line modeling of systems using Kautz and Laguerre filters. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 48(6):562–579, 2001.

[NXS10] B. Ni, D. Xiao, and S. L. Shah. Time delay estimation for MIMO dynamical systems - with time - frequency domain analysis. *Journal of Process Control*, 20(1):83–94, 2010.

[Pha00] D. T. Pham. Estimation of continuous-time autoregressive model from finely sampled data. *IEEE Transactions on Signal Processing*, 48(9):2576–2584, 2000.

[Rao65] C. R. Rao. *Linear Statistical Inference and its Applications*. Wiley, New York, 1965.

[RTS65] H. E. Rauch, F. Tung, and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 8(3):1445–1450, 1965.

[SCB92] T. Söderström, B. Carlsson, and S. Bigi. On estimating continuous-time stochastic models from discrete-time data. UPTEC 92104R, Department of Technology, Uppsala University, July 1992.

[Shu82] R. H. Shumway. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253–264, 1982.

[Sim06] D. Simon. *Optimal State Estimation. Kalman, H_∞ and Nonlinear Approaches*. Wiley–Interscience, Hoboken, New Jersey, 2006.

[Söd90] T. Söderström. On zero locations for sampled stochastic systems. *IEEE Transactions on Automatic Control*, 35(11):1249–1253, 1990.

[Söd91] T. Söderström. Computing stochastic continuous-time models from ARMA models. *International Journal of Control*, 53(6):1311–1326, 1991.

[Söd02] T. Söderström. *Discrete-time Stochastic Systems*. Springer Verlag, London, 2002.

[SS83] T. Söderström and P. Stoica. *Instrumental variable methods for system identification*. Springer Verlag, New York, 1983.

[SS89] T. Söderström and P. Stoica. *System Identification*. Series in Systems and Control Engineering. Prentice Hall, Englewood Cliffs, 1989.

[SS11] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications*. Springer Science+Business Media, New York, third edition, 2011.

[SWN11] T. B. Schön, A. Wills, and B. Ninness. System identification of nonlinear state-space models. *Automatica*, 47(1):39–49, 2011.

[TB95] K. M. Tsang and S. A. Billings. Identification of systems from non-uniformly sampled data. *International Journal of Systems Science*, 10(26):1823–1837, 1995.

[TFCB97] T.Söderström, H. Fan, B. Carlsson, and S. Bigi. Least squares parameter estimation of continuous-time ARX models from discrete-time data. *IEEE Transactions on Automatic Control*, 42(5):659–672, 1997.

[Van78] C. F. Van Loan. Computing integrals involving the matrix exponentials. *IEEE Transactions on Automatic Control*, 23(3):395–404, 1978.

[Wah88] B. Wahlberg. Limit results for sampled systems. *International Journal of Control*, 48:1267–1283, 1988.

[War98] A. F. Ware. Fast approximate Fourier transforms for irregularly spaced data. *SIAM Review*, 40(4):838–856, 1998.

[WLS93] B. Wahlberg, L. Ljung, and T. Söderström. On sampling of continuous-time stochastic processes. *Control Theory and Advanced Technology*, 9(1):99–112, 1993.

[WSLN13] A. Wills, T. B. Schön, L. Ljung, and B. Ninness. Identification of Hammerstein–Wiener models. *Automatica*, 49:70–81, 2013.

[Wu83] C. F. J Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.

[WZ01] Q. Wang and Y. Zhang. Robust identification of continuous systems with dead-time from step responses. *Automatica*, 37(3):377–390, 2001.

[WZC09] J. Wang, W.X. Zheng, and T. Chen. Identification of linear dynamic systems operating in a networked environment. *Automatica*, 45(12):2763–2772, 2009.

[YAAG11] J. I. Yuz, J. Alfaro, J. C. Agüero, and G. C. Goodwin. Identification of continuous-time state-space models from non-uniform fast-sampled data. *IET Control Theory and Applications*, 5(7):842–85, 2011.

[YG06] P. C. Young and H. Garnier. Identification and estimation of continuous-time data-based mechanistic (DBM) models for environmental systems. *Environmental Modelling and Software*, 21(8):1055–1072, 2006.

[YGG08] P. C. Young, H. Garnier, and M. Gilson. Refined instrumental variable identification of continuous-time hybrid Box-Jenkins models. In *Identification of Continuous-time Models from Sampled Data (H. Garnier and L. Wang (Eds.))*, pages 91–132, London, 2008. Springer-Verlag.

[YIKW07] Z. Yang, H. Iemura, S. Kanae, and K. Wada. Identification of continuous-time systems with multiple unknown time delays by global nonlinear least-squares and instrumental variable methods. *Automatica*, 43(7):1257–1264, 2007.

[YJ80] P. C. Young and A. J. Jakeman. Refined instrumental variable methods of time-series analysis: Part III, extensions. *International Journal of Control*, 31:741–764, 1980.

[You11] P. C. Young. *Recursive Estimation and Time-series Analysis: An Introduction for the Student and Practitioner*. Springer-Verlag, Berlin, 2011.

[Yuz05] J. I. Yuz. *Sampled-data Models for Linear and Nonlinear Systems*. Phd thesis, School of Electrical Engineering and Computer Science, University of Newcastle, Australia, 2005.

Résumé

Cette thèse traite de l'identification de systèmes dynamiques à partir de données échantillonnées à pas variable. Ce type de données est souvent rencontré dans les domaines biomédical, environnemental, dans le cas des systèmes mécaniques où un échantillonnage angulaire est réalisé ou lorsque les données transitent sur un réseau. L'identification directe de modèles à temps continu est l'approche à privilégier lorsque les données disponibles sont échantillonnées à pas variable ; les paramètres des modèles à temps discret étant dépendants de la période d'échantillonnage. Dans une première partie, un estimateur optimal de type variable instrumentale est développé pour estimer les paramètres d'un modèle Box-Jenkins à temps continu. Ce dernier est itératif et présente l'avantage de fournir des estimées non biaisées lorsque le bruit de mesure est coloré et sa convergence est peu sensible au choix du vecteur de paramètres initial. Une difficulté majeure dans le cas où les données sont échantillonnées à pas variable concerne l'estimation de modèles de bruit de type AR et ARMA à temps continu (CAR et CARMA). Plusieurs estimateurs pour les modèles CAR et CARMA s'appuyant sur l'algorithme Espérance-Maximisation (EM) sont développés puis inclus dans l'estimateur complet de variable instrumentale optimale. Une version étendue au cas de l'identification en boucle fermée est également développée. Dans la deuxième partie de la thèse, un estimateur robuste pour l'identification de systèmes à retard est proposé. Cette classe de systèmes est très largement rencontrée en pratique et les méthodes disponibles ne peuvent pas traiter le cas de données échantillonnées à pas variable. Le retard n'est pas contraint à être un multiple de la période d'échantillonnage, contrairement à l'hypothèse traditionnelle dans le cas de modèles à temps discret. L'estimateur développé est de type bootstrap et combine la méthode de variable instrumentale itérative pour les paramètres de la fonction de transfert avec un algorithme numérique de type gradient pour estimer le retard. Un filtrage de type passe-bas est introduit pour élargir la région de convergence pour l'estimation du retard. Tous les estimateurs proposés sont inclus dans la boîte à outils logicielle CONTSID pour Matlab et sont évalués à l'aide de simulation de Monte-Carlo.

Mots-clés: modèles à temps continu, données échantillonnées à pas variable, estimateur de la variable instrumentale, modèle de Box-Jenkins, processus stochastique à temps continu, algorithme espérance-maximisation, systèmes à retard

Abstract

The output of a system is always corrupted by additive noise, therefore it is more practical to develop estimation algorithms that are capable of handling noisy data. The effect of white additive noise has been widely studied, while a colored additive noise attracts less attention, especially for a continuous-time (CT) noise. Sampling issues of CT stochastic processes are reviewed in this thesis, several sampling schemes are presented. Estimation of a CT stochastic process is studied. An expectation-maximization-based (EM) method to CT autoregressive/autoregressive moving average model is developed, which gives accurate estimation over a large range of sampling interval. Estimation of CT Box-Jenkins models is also considered in this thesis, in which the noise part is modeled to improve the performance of plant model estimation. The proposed method for CT Box-Jenkins model identification is in a two-step and iterative framework. Two-step means the plant and noise models are estimated in a separate and alternate way, where in estimating each of them, the other is assumed to be fixed. More specifically, the plant is estimated by refined instrumental variable (RIV) method while the noise is estimated by EM algorithm. Iterative means that the proposed method repeats the estimation procedure several times until a optimal estimate is found. Many practical systems have inherent time-delay. The problem of identifying delayed systems are of great importance for analysis, prediction or control design. The presence of a unknown time-delay greatly complicates the parameter estimation problem, essentially because the model are not linear with respect to the time-delay. An approach to continuous-time model identification of time-delay systems, combining a numerical search algorithm for the delay with the RIV method for the dynamic has been developed in this thesis. In the proposed algorithm, the system parameters and time-delay are estimated reciprocally in a bootstrap manner. The time-delay is estimated by an adaptive gradient-based method, whereas the system parameters are estimated by the RIV method. Since numerical method is used in this algorithm, the bootstrap method is likely to converge to local optima, therefore a low-pass filter has been used to enlarge the convergence region for the time-delay. The performance of the proposed algorithms are evaluated by numerical examples.

Keywords: Continuous-time models, irregularly sampled data, instrumental variable, Box-Jenkins models, continuous-time stochastic process, expectation-maximization algorithm, time-delay