Recursive Model-Based Virtual In-Cylinder Pressure Sensing for Internal Combustion Engines

Doctoral Thesis

to be awarded the degree
Doctor of Engineering (Dr.-Ing.)

submitted by
Runzhe Han
from China

approved by the
Department of Mathematics/Computer Science and Mechanical Engineering
Clausthal University of Technology

Date of oral examination
December 18, 2020

Dean
Prof. Dr.-Ing. Volker Wesling

Chairperson of the Board of Examiners
Prof. Dr.-Ing. Christian Rembe

Supervising Tutor
Prof. Dr.-Ing. Christian Bohn

Reviewer
Prof. Dr.-Ing. Clemens Gühmann

Contents

Acknowledgements

This thesis is the summary of my research activities from 2016 to 2020 at the Institute of Electrical Information, Clausthal University of Technology, under the supervision and guidance of Prof. Christian Bohn.

I express my deep gratitude to Prof. Christian Bohn for giving me the opportunity to work with him and for providing me with continuous support and advice in scientific as well as personal matters. Without his continuous encouragement, this work would not have materialized. I was usually inspired by the deep insight of Prof. Christian Bohn in control and system identification.

My thanks also goes to Prof. Clemens Gühmann, Technical University of Berlin, who gave me critical remarks and constructive suggestions contributing significantly to my view on this work. I appreciate his willingness to act as a referee for this thesis.

Over the years in Clausthal, each one of my colleagues has contributed to this work in his or her unique way. Particularly, I would like to thank Dr. Georg Bauer, Dr. Stephan Beitler, and Mr. Tobias Jäger. I have learnt a lot from Dr. Georg Bauer in digital signal processing and system theory, and Dr. Stephan Beitler has taught me many things about internal-combustion engines. With the technical help from Mr. Tobias Jäger and Dr. Stephan Beitler, the engine test bench for my research was successfully built, and I highly appreciate them.

Finally, I want to thank my parents for their love and sustained support for my study abroad.

Claustahl-Zellerfeld, December 2020 Runzhe Han

Symbols

$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	the set of real-valued n-dimensional vectors
$\mathbb{C}$	the set of complex numbers
$\mathbb{C}^{m\times n}$	the set of complex-valued m by n matrices
$\mathbb{N}$	the set of natural numbers
$\in$	belong to
$\boldsymbol{I}_n$	$n \times n$ identity matrix
$\boldsymbol{I}$	identity matrix without the dimension index
$\tilde{\boldsymbol{I}}_N$	unit vector with N rows
$\mathbf{0}$	zero matrix or zero vector
$\hat{}$	reconstructed or estimated value
$^{\mathrm{T}}$	real-valued matrix transposition
$^{\mathrm{H}}$	complex-valued matrix transposition
$\cdots$	the transient process between two stationary operating conditions
$\mathrm{E}[\cdot]$	statistical expected value
$\delta(\cdot)$	unit impulse
δ_{ij}	Kronecker delta function
$\|\cdot\|_2$	2-norm of a vector
$\mathrm{rank}(\cdot)$	the rank of a matrix
$\delta(k)$	Dirac delta function
$\mathrm{E}[\cdot \mid \cdot]$	conditional expectation
$\otimes$	Kronecker product
$O(\cdot)$	" ordo " in Taylor's formula with remainder
†	Moore-Penrose inverse
$\mathrm{Re}(\cdot)$	real part value of a complex data
$\mathrm{Im}(\cdot)$	imaginary part value of a complex data
$\mathrm{vec}(\cdot)$	a column vector formed by stacking the columns of a matrix on top of each other
$\mathrm{average}(\cdot)$	the average value of all the elements in a vector
$\approx$	approximation
$\mathrm{range}(\cdot)$	the column space of a matrix
$*$	convolution
$\frac{\partial y(x)}{\partial x}$	first order partial derivative of the function $y(x)$ with respect to the variable x
$\frac{d}{dt}$	first order differential operator with respect to time
$\frac{d^2}{dt^2}$	second order differential operator with respect to time
$\mu(\cdot)$	mean value of a random sequence
$\sigma(\cdot)$	standard deviation of a random sequence

Abbreviations

IC	Internal Combustion
FRF	Frequency Response Function
IIR	Infinite Impulse Response
FIR	Finite Impulse Response
RBF	Radial Basis Function
ANN	Artificial Neural Network
SISO	Single-Input Single-Output
IAS	Instantaneous Angular Speed
LTI	Linear Time-Invariant
LTV	Linear Time-Variant
IO	Input-Output
MIMO	Multiple-Input Multiple-Output
SVD	Singular Value Decomposition
DC	Direct Current
FFT	Fast Fourier Transform
DFT	Discrete Fourier Transform
LPM	Local Polynomial Method
LM	Levenberg-Marquardt
TVDB	Time-Varying Delay Block

List of Figures

List of Tables

Recursive Model-Based Virtual In-Cylinder Pressure Sensing for Internal Combustion Engines

by **Runzhe Han**
Department of Mathematics/Computer Science and Mechanical Engineering,
Clausthal University of Technology

Abstract

The in-cylinder pressure signal is a very useful indicator for modern high-performance internal combustion engines. Unfortunately, direct measurements of the in-cylinder pressure are impractical because installing cylinder pressure sensors is difficult and conditions in internal combustion engine cylinders are adverse. Numerous methods (such as virtual sensing methods) have been investigated to reconstruct the in-cylinder pressure from externally measured signals, such as the engine block structural vibration signal and the engine crank angular speed.

Many of the proposed methodologies have shown promising results. However, there still exist some drawbacks, such as ill-conditioned inversion and the need of large number of data to derive an inverse model by artificial neural networks. In this thesis, considering current in-cylinder pressure reconstruction problems, linear model-based, nonlinear model-based, and inverse model-based in-cylinder pressure reconstruction methods, which are alternative to existing cylinder pressure reconstruction methods, are proposed. All the proposed methods are based on the recursive state reconstruction by using the Kalman filter or observer such that a direct inversion can be avoided. Moreover, all the proposed methods are recursively conducted in time domain, so they are suitable for real-time implementations and they also do not have frequency-domain problems such as spectral leakage. Additionally, all the proposed methods are model-based methods, and the models are identified by using system identification techniques excluding artificial neural networks, so the need of a large number of data is not necessary.

For system identification and the validation of the proposed methods, the datasets under different engine operating conditions were acquired from a four-cylinder diesel engine. Data acquired is from the operating condition 1200 rpm, 60 Nm to the operating condition 3000 rpm, 180 Nm. The reconstructed cylinder pressure curves and two combustion metrics cylinder pressure peak and peak location were used for validating the proposed cylinder pressure reconstruction methods. According to the cylinder pressure reconstruction results obtained based on using the proposed methods in this thesis, it can be found that all the proposed methods can be used under both stationary and non-stationary operating conditions, and the reconstructed cylinder pressure results are comparable among

existing cylinder pressure reconstruction methods. Furthermore, it can also be found that there exist several factors affecting the pressure reconstruction accuracy, such as the quality of the identified models, delay block and instantaneous engine cycle frequency.

Chapter 1

Introduction

1.1 Importance of Cylinder Pressure

Nowadays, internal-combustion (IC) engines are still the main power source for automobiles despite a growing number of greener alternatives, such as electrical motors. In order to fulfil rigorous ecological standards regarding the air pollution, more and more leading automobile manufactures are investigating advanced IC engines by involving the newest combustion control strategies. In-cylinder pressure-based engine combustion control is one of the technologies to enhance the combustion efficiency of the engine, and simultaneously reduce harmful emissions [1]. In addition, knowledge of cylinder pressure traces can also provide information for both misfire detection and knock detection. Thus obtaining the instantaneous information of the cylinder pressure signal is of importance for engine optimization and control. Cylinder pressure sensors are available for direct measurements and are fitted routinely during engine development and also on some production motorsport engines such as in Formula One. In the following of this chapter, the methods of cylinder pressure measurement and reconstruction are first reviewed, afterwards the motivation, the objectives, and a brief introduction of the results, and outline of this thesis about cylinder pressure reconstruction are given. Without confusion, the cylinder pressure mentioned in this thesis is referred to the engine in-cylinder pressure.

1.2 State of the Art of Obtaining Cylinder Pressure

In this section, several direct methods of the cylinder pressure measurement are first discussed, focusing on the particular installation arrangement including flush mounted sensors and those integrated with the spark or glow plug. Then indirect methodologies for pressure reconstruction are described.

1.2.1 Direct Measurement

Piezoelectric pressure transducers are widely used to measure the cylinder pressure for research and development purposes because of their small size, light weight, fast response time, and low sensitivity to environmental conditions. However laboratory-grade transducers are expensive, and require a high-cost charge amplifier to convert the transducer output signal to a measurable voltage signal. Additionally, both transducers and amplifiers need calibration. There are several kinds of installation forms for these transducers:

- *Flush mounted*
  The pressure transducer is flush-mounted to the cylinder head such that a specially machined bore is required to carry the transducer. However, on most engines this kind of mounting can make positioning and sealing of the transducer difficult and costly to achieve. In terms of advantages, because the transducer is flush with the cylinder head, turbulence and cavitation effect can be avoided.

- *Integrated into a spark plug or a glow plug*
  This kind pressure transducer is integrated into a spark plug (for gasoline engines) or a glow plug (for diesel engines), so the installation is straightforward. However, this kind of pressure transducer has a limited frequency response compared with the flush-mounted pressure transducer [2]. Furthermore, for spark plugs with an integrated pressure transducer require a small eccentricity of the electrode due to the limited available space, so the eccentricity should be always chosen as small as possible.

Additionally, optical measurement devices of the cylinder pressure have also been reported. While they still present the general design difficulties of a straightforward measurement sensor and have not been widely used even in the laboratory [2].

1.2.2 Indirect Cylinder Pressure Reconstruction

The reconstruction of the cylinder pressure using information available from other sensors fitted to engines has been investigated by many researchers since the mid 1980s. The indirect methodologies for the cylinder pressure reconstruction can be called virtual sensing [3]. Three main approaches of virtual sensing of the cylinder pressure have been investigated.

1.2.2.1 Engine structure vibrations/engine block acoustic emissions-based reconstruction

The fast pressure change in a cylinder during combustion leads to both engine structure vibrations and engine block acoustic emissions. The potential of engine block vibration signals and engine block acoustic emissions to recover the cylinder pressure has been explored [4–7]. The results showed that engine block vibration signals and engine acoustic emissions contain information related to the combustion process, and can easily be measured, normally on the cylinder head (or the engine block) and by using a microphone, respectively. However, these signals also contain other non-combustion information because the cylinder pressure is not the unique source causing vibrations and acoustic emissions. Based on the frequency response function (FRF) between the cylinder pressure signal and the vibration signal (or the acoustic signal), various methods have been investigated, such as inverse filtering applying cepstrum techniques [4, 8, 9] or time domain smoothing techniques [10], but also nonlinear methods like radial basis function (RBF) networks [11].

1.2.2.2 Crank angular speed-based reconstruction

The fluctuating waveform of engine speed versus crank angle also contains information about the cylinder-by-cylinder combustion pressure [12]. How the speed fluctuation varies

with the cylinder pressure changes has been explored by many researchers. The paper [13] used the FRF between the cylinder pressure and the crank angular speed signal and applied an FRF mapping to improve the cylinder pressure reconstruction accuracy under time-varying working points. The paper [14] modeled the cylinder pressure via the crank angular speed from a statistical point of view. The papers [15] and [2] used the extended sliding observer and the Kalman filter respectively based on an engine energy model whereas artificial neural networks (ANNs) approaches have been investigated by [16–19].

1.2.2.3 Combination of engine structure vibrations and crank angular speed-based reconstruction

It has been shown that both vibration and crank angular speed signal contain information about the cylinder pressure but mainly in different frequency regions [12]. In [12], a complex RBF network was proposed for the reconstruction of the cylinder pressure using both vibration and crank angular speed signal.

More cylinder pressure reconstruction methods can be found in [20].

1.3 Motivation, Thesis Objectives, and Results

1.3.1 Motivation

According to the descriptions of the direct measurement methods, it can be known that the use of direct sensing is however still a major problem for volume-produced engines owing to high sensor costs, limited durability, and design difficulties, which potentially creates very expensive service costs. Moreover, positioning can also be problematic, namely the process of finding sensor access on a crowed cylinder head, which is a particular problem for smaller and compact engines. Thus indirect cylinder pressure reconstruction is preferred.

Many of indirect cylinder pressure reconstruction methods have shown promising results. However, there still exist a number of drawbacks in existing methods. For example, for inverse FRF-based methods (with or without cepstrum techniques), the problems such as spectrum leakage, ill-conditioned inversion (nonminimum-phase zeros), or FRF variations may occur. In addition, inverse FRF-based methods may encounter a difficulty in multivariable systems. ANNs normally need large amounts of acquired data to train the network. For engine energy model-based methods, the engine energy model has to be known and needs to be calibrated first, which can be expensive and time-consuming in practice especially when the model is used for different types of engines. Besides, a majority of the researches of the cylinder pressure reconstruction have been performed on one-cylinder or two-cylinder engines which are relatively small in size with low power output.

Thus the motivation of this thesis is to handle the above-mentioned problems existing in current reconstruction methods.

1.3.2 Thesis Objectives

Based on the motivation, the main objectives of this thesis are formulated as follows:

- Use a four-cylinder diesel engine with a typical engine speed range and a typical engine torque range to validate the proposed cylinder pressure reconstruction methods in this thesis.

- Identify the model between four cylinder pressure signals and one vibration signal by using state-space model identification methods such that large amounts of data for training is not necessary, and first-principle model, i.e., energy model, is either not needed.

- Based on the identified models, derive recursive time-domain cylinder pressure indirect reconstruction methods such that they can be implemented recursively in digital computers and avoid the spectrum leakage problem. Additionally, direct inversion is not needed in the proposed methods such that the ill-conditioned inversion can be avoided.

1.3.3 Results

On the basis of the thesis objectives, the results obtained in the thesis are briefly introduced below:

- Both linear and nonlinear models between the cylinder pressure and the vibration were identified for cylinder pressure reconstruction, and finally for linear model identification, the final time-domain identified model was chosen as a model with 9th order, and the final frequency-domain identified model was chosen as a model with 8th order. While for nonlinear model identification, a Wiener model structure was involved, and finally a Wiener model with 9 poles and 7 zeros was chosen.

- There are totally four cylinder pressure reconstruction methods proposed in the thesis, and among these methods there are three methods formulated based on using a delay block bank, while the remained one method is based on inverse model identification. The cylinder pressure reconstruction results can show that in some engine operating ranges, the curves of the cylinder pressure were reconstructed well besides the bottom part by using delay block-based methods, while for the inverse model identification-based method, the curves of the cylinder pressure were not well reconstructed, but the total method without delay blocks is more easily formulated.

1.4 Thesis Outline

On the basis of the thesis objectives, the thesis is organized as follows:

In Chapter 2, basic theories used in subsequent chapters are illustrated. Because all the cylinder pressure reconstruction methods proposed in this thesis are formulated as state reconstruction problem, the theories of observer and Kalman fitler are illustrated. Moreover, the cylinder pressure signal is modeled based on the theory of order tracking, so order tracking is discussed in detail. System identification techniques for both linear models and nonlinear models are used for the identification of the model between the cylinder pressure and vibration. To make this thesis fairly self-contained in the sense

that all steps have to be carried out to implement identification algorithms, detailed descriptions on the system identification algorithms used in this thesis are included.

In Chapter 3, a four-cylinder diesel engine with its control and measurement system used for validating the proposed methods is first described. Afterwards, a systematic design process of the signal processing of the collect signals (including four cylinder pressure signals, one engine structural vibration signal, and one crank angular speed signal) from the engine test bench is given.

In Chapter 4, according to the system identification techniques introduced in Chapter 3, different models between four cylinder pressure signals and one vibration signal are identified. The types of the identified models are time-domain identified linear model, frequency-domain identified linear model, and nonlinear model. The models identified in this chapter are used in the proposed cylinder pressure reconstruction methods in Chapter 5, Chapter 6, and Chapter 7.

In Chapter 5, based on using both engine structural vibration signal and engine speed signal, a linear framework is proposed for the cylinder pressure reconstruction. Specifically, three delay blocks are involved to transform the identified linear models in into a single-input single-output (SISO) linear model, afterwards an augmented model can be obtained by augmenting the SISO model with the state of the cylinder No. 1 pressure signal model, and finally the Kalman filter for the augmented model can be implemented to sequentially estimate the cylinder No. 1 pressure signal, and simultaneously other three cylinder pressure signals can also be reconstructed based on using the estimated cylinder No. 1 pressure signal and delay blocks.

In Chapter 6, the instantaneous engine cycle frequency, which is necessary to formulate the delay blocks, is estimated through the vibration signal rather than being calculated in the proposed method introduced in Chapter 2. Thus the cylinder pressure estimation method proposed in this chapter is able to merely use the vibration signal to estimate four cylinder pressure signals.

In Chapter 7, a nonlinear framework, which can be seen as an extension of the linear approach, is developed. The objective of nonlinear framework is to assess whether cylinder pressure reconstruction accuracy can be improved by using a nonlinear model between cylinder pressure and vibration. The only difference between linear framework and nonlinear framework is that in the nonlinear method the identified nonlinear model between four cylinder pressure signals and one vibration signal is used, and the corresponding nonlinear Kalman filter is then implemented.

In Chapter 8, a cylinder pressure reconstruction method without delay blocks, which is called inverse model-based method, is proposed.

Finally, the thesis closes in Chapter 9 with discussions and conclusions.

It should be noted that in this thesis all the proposed methods are implemented in discrete-time domain, so all the basics introduced in Chapter 2 correspond to the discrete-time domain.

Chapter 2

Basic Theory

2.1 Introduction

This chapter reviews basic theories of linear state observer, linear and nonlinear Kalman filtering, order tracking, and system identification which are used in the sequent chapters. The outline of this chapter is described as follows. The theories of the linear observer and linear Kalman filter are introduced in Section 2.2, followed by the introduction of the theory of nonlinear Kalman filtering which is described in Section 2.3. Based on these state reconstruction theories, in Section 2.4, the specific modeling and reconstruction of the frequency-modulated signal are investigated, followed by three sections (from Section 2.5 to Section 2.7) describing system identification techniques. Finally, a summary is given in Section 2.8.

2.2 State Reconstruction for Linear Systems

In this section, state reconstruction is introduced for linear discrete-time systems. Both observer and Kalman filter for state reconstruction are depicted. In the introduction of the Kalman filter, as basics the definitions of white noise process and innovation process are first given. The stability of the Kalman filter and several practical problems of the Kalman filter implementation are also discussed in the end of this section.

2.2.1 The Asymptotic Observer

Consider a linear time-invariant (LTI) model

$$\begin{cases} \boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{u}(k), \\ \boldsymbol{y}(k) = \boldsymbol{C}\boldsymbol{x}(k) + \boldsymbol{D}\boldsymbol{u}(k), \end{cases} \tag{2.1}$$

where $\boldsymbol{u}(k) \in \mathbb{R}^m$, $\boldsymbol{y}(k) \in \mathbb{R}^p$, and $\boldsymbol{x}(k) \in \mathbb{R}^n$ represent the model input, output, and state variable vector, respectively. The matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ are constant.

An observer for the reconstruction of the state $\boldsymbol{x}(k)$ of the model (2.1) can be denoted as

$$\hat{\boldsymbol{x}}(k+1) = \boldsymbol{A}\hat{\boldsymbol{x}}(k) + \boldsymbol{B}\boldsymbol{u}(k) + \boldsymbol{L}(\boldsymbol{y}(k) - \boldsymbol{C}\hat{\boldsymbol{x}}(k) - \boldsymbol{D}\boldsymbol{u}(k)), \tag{2.2}$$

where the symbol "ˆ" denotes the reconstructed or estimated value, and $\boldsymbol{L}$ is a gain matrix.

The difference $\tilde{\boldsymbol{x}}(k)$ between the reconstructed state $\hat{\boldsymbol{x}}(k)$ and the real state $\boldsymbol{x}(k)$ satisfies

$$\tilde{\boldsymbol{x}}(k+1) = (\boldsymbol{A} - \boldsymbol{L}\boldsymbol{C})\tilde{\boldsymbol{x}}(k), \tag{2.3}$$

and therefore, if $\boldsymbol{L}$ is chosen such that the state equation (2.3) is asymptotically stable, the difference $\boldsymbol{x}_{\mathrm{e}}(k)$ goes to zero for $k \to +\infty$, i.e.,

$$\lim_{k\to+\infty} \tilde{\boldsymbol{x}}(k) = \lim_{k\to+\infty} (\boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)) = \boldsymbol{0}. \tag{2.4}$$

According to the above analysis, the observer (2.2) can be called an asymptotic observer when the value of gain matrix $\boldsymbol{L}$ is chosen suitably [21, 22]. In this thesis, the observer for LTI systems is only considered.

2.2.2 The Kalman Filter

At the start of this section, the definitions of both white noise process and innovation process are introduced.

Definition 2.1. [23] The scalar random process $\{X(k)\}$ which satisfies two conditions:

(i) $\mathrm{E}[X(k)] = 0$,

(ii) $\mathrm{E}[X(k)X(j)] = Q(k)\delta_{kj}$,

where $\mathrm{E}[\cdot]$ denotes the statistical expected value, and δ_{kj} is the Kronecker Delta function, and

$$\delta_{kj} = \begin{cases} 0, & \text{if} \quad k \neq j, \\ 1, & \text{if} \quad k = j, \end{cases} \tag{2.5}$$

is said to be a white noise process. □

It is easy to generalize the above definition to the corresponding vector case, e.g., a vector white noise process satisfies: (i) $\mathrm{E}[\boldsymbol{X}(k)] = \boldsymbol{0}$ and (ii) $\mathrm{E}\left[\boldsymbol{X}(k_1)\boldsymbol{X}^{\mathrm{T}}(k_2)\right] = \boldsymbol{Q}(k_1)\delta_{k_1k_2}$, where the matrix $\boldsymbol{Q}(k_1) \geq \boldsymbol{0}$ and the symbol " $^{\mathrm{T}}$ " denotes the real-valued matrix transpose. With the above definition, the introduction of the Kalman filter theory is conducted as follows.

Given a linear time-variant (LTV) model with additive error:

$$\begin{cases} \boldsymbol{x}(k+1) = \boldsymbol{A}(k)\boldsymbol{x}(k) + \boldsymbol{B}(k)\boldsymbol{u}(k) + \boldsymbol{E}(k)\boldsymbol{w}(k), \\ \boldsymbol{y}(k) = \boldsymbol{C}(k)\boldsymbol{x}(k) + \boldsymbol{D}(k)\boldsymbol{u}(k) + \boldsymbol{v}(k), \end{cases} \tag{2.6}$$

where $\boldsymbol{u}(k) \in \mathbb{R}^m$, $\boldsymbol{y}(k) \in \mathbb{R}^p$, and $\boldsymbol{x}(k) \in \mathbb{R}^n$ represent the model input, output, and state variable vector, respectively. $\boldsymbol{w}(k) \in \mathbb{R}^q$ and $\boldsymbol{v}(k) \in \mathbb{R}^p$, and $\{\boldsymbol{w}(k)\}$ and $\{\boldsymbol{v}(k)\}$ are white noise processes, and $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ are uncorrelated with $\boldsymbol{x}(0)$ (a random variable with constant mean $\boldsymbol{x}_0$ and constant variance $\boldsymbol{P}_0$) and with each other. The matrices $\boldsymbol{A}(k)$, $\boldsymbol{B}(k)$, $\boldsymbol{C}(k)$, $\boldsymbol{D}(k)$, and $\boldsymbol{E}(k)$ are time-variant. The covariance matrices of $\{\boldsymbol{w}(k)\}$ and $\{\boldsymbol{v}(k)\}$ are $\boldsymbol{Q}(k)\delta_{kj}$ and $\boldsymbol{R}(k)\delta_{kj}$, respectively, i.e.,

$$\mathrm{E}\left[\boldsymbol{w}(k)\boldsymbol{w}^{\mathrm{T}}(j)\right] = \boldsymbol{Q}(k)\delta_{kj} \quad \text{and} \quad \mathrm{E}\left[\boldsymbol{v}(k)\boldsymbol{v}^{\mathrm{T}}(j)\right] = \boldsymbol{R}(k)\delta_{kj}.$$

Prior to introducing the Kalman filter algorithm for the model (2.6), define at each time k the a priori (before including the measurement $\boldsymbol{y}(k)$) estimate

$$\hat{\boldsymbol{x}}^-(k) = \mathrm{E}\left[\boldsymbol{x}(k)|\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots, \boldsymbol{y}(k-1)\right],$$

where $\mathrm{E}\left[\cdot \mid \cdot\right]$ denotes the conditional expectation, and error variance

$$\boldsymbol{P}^-(k) = \mathrm{E}\left[\left(\boldsymbol{x}(k) - \hat{\boldsymbol{x}}^-(k)\right)\left(\boldsymbol{x}(k) - \hat{\boldsymbol{x}}^-(k)\right)^{\mathrm{T}}\right],$$

and the a posteriori (after including the measurement $\boldsymbol{y}(k)$) estimate

$$\hat{\boldsymbol{x}}(k) = \mathrm{E}\left[\boldsymbol{x}(k)|\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots, \boldsymbol{y}(k)\right],$$

and error variance

$$\boldsymbol{P}(k) = \mathrm{E}\left[\left(\boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)\right)\left(\boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)\right)^{\mathrm{T}}\right],$$

and $\hat{\boldsymbol{x}}^-(k)$ is also called one-step-ahead prediction.

Based on the above descriptions, conventionally the Kalman filter algorithm for the model (2.6) can be summarized as follows if the start time is set to zero [24]:

- *Initialization*

$$\boldsymbol{P}(0) = \boldsymbol{P}_0 \quad \text{and} \quad \hat{\boldsymbol{x}}(0) = \boldsymbol{x}_0.$$

- *Time update (effect of system dynamics)*
  Error variance

$$\boldsymbol{P}^-(k+1) = \boldsymbol{A}(k)\boldsymbol{P}(k)\boldsymbol{A}^{\mathrm{T}}(k) + \boldsymbol{E}(k)\boldsymbol{Q}(k)\boldsymbol{E}^{\mathrm{T}}(k). \tag{2.7}$$

  Estimate

$$\hat{\boldsymbol{x}}^-(k+1) = \boldsymbol{A}(k)\hat{\boldsymbol{x}}(k) + \boldsymbol{B}(k)\boldsymbol{u}(k). \tag{2.8}$$

- *Measurement update (effect of measurement $\boldsymbol{y}(k)$)*
  Gain calculation

$$\boldsymbol{K}(k+1) = \boldsymbol{P}^-(k+1)\boldsymbol{C}^{\mathrm{T}}(k+1)\left(\boldsymbol{C}(k+1)\boldsymbol{P}^-(k+1)\boldsymbol{C}^{\mathrm{T}}(k+1) + \boldsymbol{R}(k+1)\right)^{-1}, \tag{2.9}$$

  where $\boldsymbol{K}(k)$ is called the Kalman gain.

  Error variance update

$$\boldsymbol{P}(k+1) = \left(\boldsymbol{I}_n - \boldsymbol{K}(k+1)\boldsymbol{C}(k+1)\right)\boldsymbol{P}^-(k+1), \tag{2.10}$$

  where $\boldsymbol{I}_n$ denotes $n \times n$ identity matrix.

  Estimate update

$$\begin{aligned}\hat{\boldsymbol{x}}(k+1) = \hat{\boldsymbol{x}}^-(k+1) + \boldsymbol{K}(k+1)\Big(&\boldsymbol{y}(k+1) - \boldsymbol{C}(k+1)\hat{\boldsymbol{x}}^-(k+1) \\ &-\boldsymbol{D}(k+1)\boldsymbol{u}(k+1)\Big).\end{aligned} \tag{2.11}$$

If Equation (2.11) in Equation (2.8) is either used, there results an alternative group

of Kalman filter recursive equations (the a priori filter formulation) [24], i.e.,

$$\boldsymbol{P}^{-}(k+1) = \boldsymbol{A}(k)\Big(\boldsymbol{P}^{-}(k) - \boldsymbol{P}^{-}(k)\boldsymbol{C}^{\mathrm{T}}(k)\Big(\boldsymbol{C}(k)\boldsymbol{P}^{-}(k)\boldsymbol{C}^{\mathrm{T}}(k) + \boldsymbol{R}(k)\Big)^{-1}\boldsymbol{C}(k)\boldsymbol{P}^{-}(k)\Big)\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{E}(k)\boldsymbol{Q}(k)\boldsymbol{E}^{\mathrm{T}}(k), \tag{2.12}$$

$$\boldsymbol{K}(k) = \boldsymbol{P}^{-}(k)\boldsymbol{C}^{\mathrm{T}}(k)\Big(\boldsymbol{C}(k)\boldsymbol{P}^{-}(k)\boldsymbol{C}^{\mathrm{T}}(k) + \boldsymbol{R}(k)\Big)^{-1}, \tag{2.13}$$

and

$$\hat{\boldsymbol{x}}^{-}(k+1) = \boldsymbol{A}(k)\left(\boldsymbol{I}_n - \boldsymbol{K}(k)\boldsymbol{C}(k)\right)\hat{\boldsymbol{x}}^{-}(k) + \boldsymbol{B}(k)\boldsymbol{u}(k) + \boldsymbol{A}(k)\boldsymbol{K}(k)\left(\boldsymbol{y}(k) - \boldsymbol{D}(k)\boldsymbol{u}(k)\right), \tag{2.14}$$

where the initial values can be chosen as

$$\boldsymbol{P}^{-}(0) = \boldsymbol{P}_0^{-} \quad \text{and} \quad \hat{\boldsymbol{x}}^{-}(0) = \boldsymbol{x}_0^{-},$$

where the random variable $\boldsymbol{x}^{-}(0)$ is with constant mean $\boldsymbol{x}_0^{-}$ and constant variance $\boldsymbol{P}_0^{-}$.

Remark 2.1. Equation (2.14) can be transformed into

$$\hat{\boldsymbol{x}}^{-}(k+1) = \boldsymbol{A}(k)\hat{\boldsymbol{x}}^{-}(k) + \boldsymbol{B}(k)\boldsymbol{u}(k) + \boldsymbol{A}(k)\boldsymbol{K}(k)\Big(\boldsymbol{y}(k) - \boldsymbol{C}(k)\hat{\boldsymbol{x}}^{-}(k) - \boldsymbol{D}(k)\boldsymbol{u}(k)\Big), \tag{2.15}$$

and by observation, Equation (2.15) is identical in structure to Equation (2.2), so Equation (2.15), which can be used to design the deterministic state observer, is just a time-variant version of Equation (2.2). □

2.2.3 Stability of the Kalman Filter

For the Kalman filter for the model (2.6), according to [25, 26], if $(\boldsymbol{A}(k), \boldsymbol{C}(k))$ is uniformly detectable, the Kalman filter error variance will be bounded.

When in the model (2.6) the matrices $\boldsymbol{A}(k) \equiv \boldsymbol{A}$, $\boldsymbol{B}(k) \equiv \boldsymbol{B}$, $\boldsymbol{C}(k) \equiv \boldsymbol{C}$, $\boldsymbol{D}(k) \equiv \boldsymbol{D}$, $\boldsymbol{E}(k) \equiv \boldsymbol{E}$, $\boldsymbol{Q}(k) \equiv \boldsymbol{Q}$, and $\boldsymbol{R}(k) \equiv \boldsymbol{R}$, where the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{E}$, $\boldsymbol{Q}$, and $\boldsymbol{R} > \boldsymbol{0}$ are constant, the above conclusion about the Kalman filter stability can either be hold. Furthermore, if $(\boldsymbol{A}, \boldsymbol{C})$ is detectable and $\left(\boldsymbol{A}, \boldsymbol{E}\sqrt{\boldsymbol{Q}}\right)$ is reachable, then there will be a steady-state Kalman gain $\boldsymbol{K} = \boldsymbol{P}^{-}\boldsymbol{C}^{\mathrm{T}}\left(\boldsymbol{C}\boldsymbol{P}^{-}\boldsymbol{C}^{\mathrm{T}} + \boldsymbol{R}\right)^{-1}$, where $\boldsymbol{P}^{-}$ denotes the unique positive definite limiting solution of Equation (2.12), which is independent of $\boldsymbol{P}_0$, and the Kalman filter will be asymptotically stable, which can result in a steady-state Kalman filter [24]. The existence of the steady-state Kalman gain can indicate how to find a stabilized gain $\boldsymbol{L}$ in the asymptotic observer introduced in Section 2.2.1, i.e., $\boldsymbol{L}$ can be chosen as $\boldsymbol{A}\boldsymbol{K}$ based on the structure of Equation (2.15).

The definitions of controllability (for continuous-time systems), reachability (for discrete-time systems), stabilizability, observability, and detectability can be referred to [27]. Additionally, all of them correspond to all the states in the state space, e.g., if a system is observable, it means that the system is completely observable.

2.2.4 Implementation of the Kalman Filter

As for the linear Kalman filter implementation, a number of problems may occur, such as the problem of mismodeling, the numerical precision problem, and the correlated noise problem. So the standard Kalman filter as described above should be modified. Several modified Kalman filters are illustrated below to solve the above problems.

2.2.4.1 Kalman filter with fading memory

The Kalman filter with fading memory is identical to the standard Kalman filter [28], with the exception that the time-update equation for the computation of the a priori error variance has a λ factor in its first term. This serves to handel the mismodeling problem, which results in the filter giving more credence to the measurement. This is equivalent to increasing $\boldsymbol{Q}(k)$, which also results in the filter giving relatively more credence to the measurement. The fading-memory filter for the model (2.6) can be summarized as follows:

$$\boldsymbol{P}^{-}(k+1) = \frac{1}{\lambda}\boldsymbol{A}(k)\boldsymbol{P}(k)\boldsymbol{A}^{\mathrm{T}}(k) + \boldsymbol{E}(k)\boldsymbol{Q}(k)\boldsymbol{E}^{\mathrm{T}}(k), \tag{2.16}$$

$$\hat{\boldsymbol{x}}^{-}(k+1) = \boldsymbol{A}(k)\hat{\boldsymbol{x}}(k) + \boldsymbol{B}(k)\boldsymbol{u}(k), \tag{2.17}$$

$$\boldsymbol{K}(k+1) = \boldsymbol{P}^{-}(k+1)\boldsymbol{C}^{\mathrm{T}}(k+1)\left(\boldsymbol{C}(k+1)\boldsymbol{P}^{-}(k+1)\boldsymbol{C}^{\mathrm{T}}(k+1) + \boldsymbol{R}(k+1)\right)^{-1}, \tag{2.18}$$

$$\boldsymbol{P}(k+1) = \left(\boldsymbol{I}_n - \boldsymbol{K}(k+1)\boldsymbol{C}(k+1)\right)\boldsymbol{P}^{-}(k+1), \tag{2.19}$$

$$\hat{\boldsymbol{x}}(k+1) = \hat{\boldsymbol{x}}^{-}(k+1) + \boldsymbol{K}(k+1)\left(\boldsymbol{y}(k+1) - \boldsymbol{C}(k+1)\hat{\boldsymbol{x}}^{-}(k+1) - \boldsymbol{D}(k+1)\boldsymbol{u}(k+1)\right), \tag{2.20}$$

where the forgetting factor $\lambda \in \mathbb{R}$ and $0 < \lambda < 1$, and the initial values $\boldsymbol{P}(0) = \boldsymbol{P}_0$ and $\hat{\boldsymbol{x}}(0) = \boldsymbol{x}_0$.

If $\lambda = 1$ in Equation (2.16) then the fading-memory filter is equivalent to the standard Kalman filter. In most applications, $\frac{1}{\lambda}$ is only slightly greater than 1 (e.g., $\frac{1}{\lambda} \approx 1.01$).

2.2.4.2 Square-root Kalman filter

The square-root Kalman filter can mathematically increase the precision of the Kalman filter when hardware precision is not available.

There exist several kinds of square-root Kalman filters [28], while this thesis focuses on introducing the Potter's algorithm [29]. But in Potter's algorithm, $\boldsymbol{w}(k)$ should be zero and the measurements should be scalar. So in the model (2.6), $\boldsymbol{Q}(k)$ should be set to $\boldsymbol{0}$, and the output vector signal $\boldsymbol{y}(k)$ and the vector signal $\boldsymbol{v}(k)$ should be replaced with $y(k) \in \mathbb{R}$ and $v(k) \in \mathbb{R}$, respectively. The corresponding covariance function of the white noise process $\{v(k)\}$ is $R(k)\delta_{kj}$. Suppose at each time k the a priori error variance $\boldsymbol{P}^{-}(k) = \boldsymbol{S}^{-}(k)(\boldsymbol{S}^{-}(k))^{\mathrm{T}}$. Also, suppose that the a posteriori error variance $\boldsymbol{P}(k) = \boldsymbol{S}(k)\boldsymbol{S}^{\mathrm{T}}(k)$. With the above descriptions, the specific square-root filtering algorithm is as follows:

- *Initialization*

  $$\boldsymbol{S}(0) = \boldsymbol{S}_0 \quad \text{and} \quad \hat{\boldsymbol{x}}(0) = \boldsymbol{x}_0,$$

  where the matrices $\boldsymbol{S}_0$ and $\boldsymbol{x}_0$ are constant.

- *Time update (effect of system dynamics)*
  Error variance
$$\boldsymbol{S}^{-}(k+1) = \boldsymbol{A}(k)\boldsymbol{S}(k), \tag{2.21}$$
  Estimate
$$\hat{\boldsymbol{x}}^{-}(k+1) = \boldsymbol{A}(k)\hat{\boldsymbol{x}}(k) + \boldsymbol{B}(k)\boldsymbol{u}(k). \tag{2.22}$$

- *Measurement update (effect of measurement $y(k)$)*
  Gain calculation
$$\boldsymbol{F}(k+1) = \boldsymbol{S}^{-}(k+1)\boldsymbol{C}^{\mathrm{T}}(k+1), \tag{2.23}$$
$$\alpha(k+1) = \frac{1}{\boldsymbol{F}^{\mathrm{T}}(k+1)\boldsymbol{F}(k+1) + R(k+1)}, \tag{2.24}$$
$$\boldsymbol{K}(k+1) = \alpha(k+1)\boldsymbol{S}^{-}(k+1)\boldsymbol{F}(k+1), \tag{2.25}$$
  where $\boldsymbol{F}(k)$ and $\alpha(k)$ are the intermediate variables.
  Error variance update
$$\gamma(k+1) = \frac{1}{1 + \sqrt{R(k+1)}\sqrt{\alpha(k+1)}}, \tag{2.26}$$
$$\boldsymbol{S}(k+1) = \boldsymbol{S}^{-}(k+1) - \gamma(k)\boldsymbol{K}(k+1)\boldsymbol{F}^{\mathrm{T}}(k+1), \tag{2.27}$$
  where $\gamma(k)$ is an intermediate variable.
  Estimate update
$$\hat{\boldsymbol{x}}(k+1) = \hat{\boldsymbol{x}}^{-}(k+1) + \boldsymbol{K}(k+1)\Big(y(k+1) - \boldsymbol{C}(k+1)\hat{\boldsymbol{x}}^{-}(k+1) - \boldsymbol{D}(k+1)\boldsymbol{u}(k+1)\Big). \tag{2.28}$$

If necessary, the above square-root Kalman filter can also involve an forgetting factor λ, and Equation (2.21) should be changed into

$$\boldsymbol{S}^{-}(k+1) = \frac{1}{\lambda}\boldsymbol{A}(k)\boldsymbol{S}(k). \tag{2.29}$$

2.2.4.3 Kalman filter for correlated noise systems

In the model (2.6), if the white noise processes $\{\boldsymbol{w}(k)\}$ and $\{\boldsymbol{v}(k)\}$ are also correlated with each other, i.e.,

$$\mathrm{E}\left[\begin{pmatrix}\boldsymbol{w}(k)\\ \boldsymbol{v}(k)\end{pmatrix}\begin{pmatrix}\boldsymbol{w}^{\mathrm{T}}(j) & \boldsymbol{v}^{\mathrm{T}}(j)\end{pmatrix}\right] = \begin{pmatrix}\boldsymbol{Q}(k) & \boldsymbol{M}^{\mathrm{T}}(k)\\ \boldsymbol{M}(k) & \boldsymbol{R}(k)\end{pmatrix}\delta_{kj},$$

where $\boldsymbol{M}(k) \not\equiv \boldsymbol{0}$, a modified Kalman filter for the model (2.6) should be implemented [24, 28].

By modifying Equations (2.12) to (2.14), the Kalman filter for correlated noise systems can be illustrated as follows:

$$\boldsymbol{P}^{-}(k+1) = \boldsymbol{A}(k)\boldsymbol{P}^{-}(k)\boldsymbol{A}^{\mathrm{T}}(k)$$

$$
\begin{aligned}
&-\boldsymbol{K}(k)\left(\boldsymbol{C}(k)\boldsymbol{P}^{-}(k)\boldsymbol{C}^{\mathrm{T}}(k)+\boldsymbol{R}(k)\right)\boldsymbol{K}^{\mathrm{T}}(k)\\
&+\boldsymbol{E}(k)\boldsymbol{Q}(k)\boldsymbol{E}^{\mathrm{T}}(k),
\end{aligned}
\tag{2.30}
$$

$$
\boldsymbol{K}(k)=\left(\boldsymbol{A}\boldsymbol{P}^{-}(k)\boldsymbol{C}^{\mathrm{T}}(k)+\boldsymbol{E}(k)\boldsymbol{M}^{\mathrm{T}}(k)\right)\left(\boldsymbol{C}\boldsymbol{P}^{-}(k)\boldsymbol{C}^{\mathrm{T}}(k)+\boldsymbol{R}(k)\right)^{-1},
\tag{2.31}
$$

and

$$
\hat{\boldsymbol{x}}^{-}(k+1)=\boldsymbol{A}(k)\hat{\boldsymbol{x}}^{-}(k)+\boldsymbol{B}(k)\boldsymbol{u}(k)+\boldsymbol{K}(k)\left(\boldsymbol{y}(k)-\boldsymbol{C}(k)\hat{\boldsymbol{x}}^{-}(k)-\boldsymbol{D}(k)\boldsymbol{u}(k)\right).
\tag{2.32}
$$

When $\boldsymbol{M}(k)\equiv\boldsymbol{0}$ in Equation (2.31), the Kalman filter for correlated noise systems can immediately be changed into the standard Kalman filter that is introduced in Section 2.2.2.

Remark 2.2. For the Kalman filter for correlated noise systems, in the model (2.6) if the matrices $\boldsymbol{A}(k)\equiv\boldsymbol{A}$, $\boldsymbol{B}(k)\equiv\boldsymbol{B}$, $\boldsymbol{C}(k)\equiv\boldsymbol{C}$, $\boldsymbol{D}(k)\equiv\boldsymbol{D}$, $\boldsymbol{E}(k)\equiv\boldsymbol{E}$, $\boldsymbol{Q}(k)\equiv\boldsymbol{Q}$, $\boldsymbol{R}(k)\equiv\boldsymbol{R}$, and $\boldsymbol{M}(k)\equiv\boldsymbol{M}$, where the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{E}$, $\boldsymbol{Q}$, $\boldsymbol{R}>\boldsymbol{0}$, and $\boldsymbol{M}$ are constant, and if $(\boldsymbol{A},\boldsymbol{C})$ is observable and $\left(\boldsymbol{A},\boldsymbol{E}\sqrt{\boldsymbol{Q}}\right)$ is reachable, then there will be a steady-state Kalman gain $\boldsymbol{K}=\left(\boldsymbol{M}+\boldsymbol{A}\boldsymbol{P}^{-}\boldsymbol{C}^{\mathrm{T}}\right)\left(\boldsymbol{C}\boldsymbol{P}^{-}\boldsymbol{C}^{\mathrm{T}}+\boldsymbol{R}\right)^{-1}$, where $\boldsymbol{P}^{-}$ denotes the unique positive definite limiting solution of Equation (2.30), which is independent of $\boldsymbol{P}_0$, and the Kalman filter will be asymptotically stable, which can result in a steady-state Kalman filter [22, 30]. □

Remark 2.3. For the correlated case, even if $\boldsymbol{M}\neq\boldsymbol{0}$, the conventional Kalman filter can still be implemented, but the corresponding Kalman filter would be suboptimal in this case [24]. □

2.3 Nonlinear Kalman Filtering for Nonlinear Systems

As a counterpart of the linear Kalman filter, nonlinear Kalman filters are designed for nonlinear systems. There are several kinds nonlinear Kalman filters such as the linearized Kalman filter, extended Kalman filter (EKF), and second-order EKF [28, 31]. Because in this thesis only EKF is of interest if nonlinear systems are encountered in the sequent chapters, in this section only the specific introduction of EKF is illustrated.

Consider a nonlinear state-space model with additive error:

$$
\begin{cases}
\boldsymbol{x}(k+1)=\boldsymbol{g}(\boldsymbol{x}(k),\boldsymbol{u}(k))+\boldsymbol{w}(k),\\
\boldsymbol{y}(k)=\boldsymbol{h}(\boldsymbol{x}(k),\boldsymbol{u}(k))+\boldsymbol{v}(k),
\end{cases}
\tag{2.33}
$$

where $\boldsymbol{u}(k)\in\mathbb{R}^m$, $\boldsymbol{y}(k)\in\mathbb{R}^p$, and $\boldsymbol{x}(k)\in\mathbb{R}^n$ represent the model input, output, and state variable vector, respectively. $\boldsymbol{w}(k)\in\mathbb{R}^n$ and $\boldsymbol{v}(k)\in\mathbb{R}^p$, and $\{\boldsymbol{w}(k)\}$ and $\{\boldsymbol{v}(k)\}$ are white noise processes, and $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ are uncorrelated with $\boldsymbol{x}(0)$ (a random variable with constant mean $\boldsymbol{x}_0$ and constant variance $\boldsymbol{P}_0$) and with each other. The functions $\boldsymbol{g}(\cdot)$ and $\boldsymbol{h}(\cdot)$ are nonlinear vector-valued functions which are smooth. The covariance matrices of $\{\boldsymbol{w}(k)\}$ and $\{\boldsymbol{v}(k)\}$ are $\boldsymbol{Q}(k)\delta_{kj}$ and $\boldsymbol{R}(k)\delta_{kj}$, respectively, i.e.,

$$
\mathrm{E}\left[\boldsymbol{w}(k)\boldsymbol{w}^{\mathrm{T}}(j)\right]=\boldsymbol{Q}(k)\delta_{kj}\quad\text{and}\quad\mathrm{E}\left[\boldsymbol{v}(k)\boldsymbol{v}^{\mathrm{T}}(j)\right]=\boldsymbol{R}(k)\delta_{kj}.
$$

The discrete-time EKF algorithm for the nonlinear model (2.33) can be summarized as follows if the start time is set to zero [31]:

- *Initialization*

$$\boldsymbol{P}(0) = \boldsymbol{P}_0 \quad \text{and} \quad \hat{\boldsymbol{x}}(0) = \boldsymbol{x}_0.$$

- *Time update (effect of system dynamics)*
  Error variance

$$\boldsymbol{P}^-(k+1) = \boldsymbol{A}_{\mathrm{L}}(k)\boldsymbol{P}(k)\boldsymbol{A}_{\mathrm{L}}^{\mathrm{T}}(k) + \boldsymbol{Q}(k). \tag{2.34}$$

  Estimate

$$\hat{\boldsymbol{x}}^-(k+1) = \boldsymbol{g}(\hat{\boldsymbol{x}}(k), \boldsymbol{u}(k)). \tag{2.35}$$

- *Measurement update (effect of measurement $\boldsymbol{y}(k)$)*
  Gain update

$$\boldsymbol{K}(k+1) = \boldsymbol{P}^-(k+1)\boldsymbol{C}_{\mathrm{L}}^{\mathrm{T}}(k+1)\left(\boldsymbol{C}_{\mathrm{L}}(k+1)\boldsymbol{P}^-(k+1)\boldsymbol{C}_{\mathrm{L}}^{\mathrm{T}}(k+1) + \boldsymbol{R}(k+1)\right)^{-1}, \tag{2.36}$$

  where $\boldsymbol{K}(k)$ is called the Kalman gain.

  Error variance

$$\boldsymbol{P}(k+1) = \left(\boldsymbol{I}_n - \boldsymbol{K}(k+1)\boldsymbol{C}_{\mathrm{L}}(k+1)\right)\boldsymbol{P}^-(k+1). \tag{2.37}$$

  Estimate

$$\hat{\boldsymbol{x}}(k+1) = \hat{\boldsymbol{x}}^-(k+1) + \boldsymbol{K}(k+1)\left(\boldsymbol{y}(k+1) - \boldsymbol{h}(\hat{\boldsymbol{x}}^-(k+1), \boldsymbol{u}(k+1))\right). \tag{2.38}$$

- *Jacobians*

$$\boldsymbol{A}_{\mathrm{L}}(k) \approx \left.\frac{\partial \boldsymbol{g}(\boldsymbol{x}(k), \boldsymbol{u}(k))}{\partial \boldsymbol{x}^{\mathrm{T}}(k)}\right|_{\boldsymbol{x}(k)=\hat{\boldsymbol{x}}(k), \boldsymbol{u}(k)=\boldsymbol{u}(k)}, \tag{2.39}$$

$$\boldsymbol{C}_{\mathrm{L}}(k+1) \approx \left.\frac{\partial \boldsymbol{h}(\boldsymbol{x}(k), \boldsymbol{u}(k))}{\partial \boldsymbol{x}^{\mathrm{T}}(k)}\right|_{\boldsymbol{x}(k)=\hat{\boldsymbol{x}}^-(k+1), \boldsymbol{u}(k)=\boldsymbol{u}(k+1)}, \tag{2.40}$$

where $\frac{\partial \boldsymbol{g}(\boldsymbol{x}(k), \boldsymbol{u}(k))}{\partial \boldsymbol{x}(k)}$ denotes the partial derivative of the function $\boldsymbol{g}(\boldsymbol{x}(k), \boldsymbol{u}(k))$ with respect to the variable $\boldsymbol{x}(k)$.

Remark 2.4. By observing the EKF equations, it can be known that EKF can also be transformed into a square-root form or a square-root form with a forgetting factor. □

2.4 Recursive Order Tracking

The problem of order tracking is usually encountered in rotating machines [32]. In rotating machines, a single order $u_i(k) \in \mathbb{R}$ may be mathematically defined by the time-variant phasor described as [32]

$$u_i(k) = A_i(k) \sin\left(\sum_{j=0}^{k} 2\pi f_i(j) T_{\mathrm{s}} + \phi_i(0)\right), \tag{2.41}$$

where T_s denotes the sampling period, $A_i(k)$ denotes the amplitude of the order $u_i(k)$, $\phi_i(0)$ represents the initial phase of order $u_i(k)$ at the step $k = 0$, and $f_i(k)$ denotes instantaneous frequency of order $u_i(k)$, and

$$f_i(k) = \begin{cases} l_i f_{\mathrm{ref}}(k), & \text{if} \quad i \neq 0, \\ 0, & \text{if} \quad i = 0, \end{cases} \tag{2.42}$$

where l_i can be a fraction or an integer, and $f_{\mathrm{ref}}(k)$, which is called the reference frequency, is the instantaneous frequency of a reference shaft in rotating machines.

Multiple orders are normally present in a dataset acquired from an operating machine. The combination of orders may be described mathematically by a summation of time-variant phasors. The combination with an error is expressed as

$$u(k) = \sum_{i=0}^{n_\mathrm{u}} u_i(k) + e(k), \tag{2.43}$$

where $u(k) \in \mathbb{R}$, and the number n_u stands for the total order number in the collected signal $u(k)$, especially, $u_0(k)$ denotes the DC offset corresponding to the frequency $f_0(k) = 0$ Hz.

The objective of the order tracking problem is to reconstruct each order $u_i(k)$ based on using the summed signal $u(k)$ which is known. Below, three subsections are used to formulate the order tracking problem. In Section 2.4.1, the order tracking with time-invariant amplitude is introduced while in Section 2.4.2, the order tracking with time-variant amplitude is illustrated. In Section 2.4.3, the order tracking with unknown instantaneous frequency is depicted.

2.4.1 Order Tracking with Time-Invariant Amplitude

Based on the fact that

$$\begin{aligned} 2u_i(k+1) &= A_i(k+1)\sin(\varphi_i(k+1) - \varphi_i(k) + \varphi_i(k)) \\ &= A_i(k+1)\left(\sin(\varphi_i(k+1) - \varphi_i(k))\cos(\varphi_i(k))\right. \\ &\quad \left. + \cos(\varphi_i(k+1) - \varphi_i(k))\sin(\varphi_i(k))\right) \\ &= A_i(k+1)\left(\sin\left(2\pi f_i(k)T_\mathrm{s}\right)\cos(\varphi_i(k)) + \cos\left(2\pi f_i(k)T_\mathrm{s}\right)\sin(\varphi_i(k))\right), \end{aligned} \tag{2.44}$$

where $\varphi_i(k+1) = \sum_{j=0}^{k+1} 2\pi f_i(j)T_\mathrm{s} + \phi_i(0)$, and when the amplitude of the order $u_i(k)$ is time-invariant, i.e., $A_i(k+1) = A_i(k)$, it can be obtained that

$$\begin{pmatrix} u_i(k+1) \\ A_i(k+1)\cos(\varphi_i(k+1)) \end{pmatrix} = \begin{pmatrix} \cos\left(2\pi f_i(k)T_\mathrm{s}\right) & \sin\left(2\pi f_i(k)T_\mathrm{s}\right) \\ -\sin\left(2\pi f_i(k)T_\mathrm{s}\right) & \cos\left(2\pi f_i(k)T_\mathrm{s}\right) \end{pmatrix} \begin{pmatrix} u_i(k) \\ A_i(k)\cos(\varphi_i(k)) \end{pmatrix}, \tag{2.45}$$

based on Equation (2.45), it is obvious that when the amplitude $A_i(k)$ is time-invariant and the reference frequency $f_{\mathrm{ref}}(k)$ is constant or time-variant, the signal $u(k)$ in Equation (2.43) can be expressed as the output of a linear state-space model in time domain [33, 34], i.e.,

$$\begin{cases} \boldsymbol{x}(k+1) = \boldsymbol{A}(f_{\mathrm{ref}}(k))\boldsymbol{x}(k), \\ u(k) = \boldsymbol{C}\boldsymbol{x}(k) + e(k). \end{cases} \tag{2.46}$$

The matrices $\boldsymbol{A}(f_{\mathrm{ref}}(k))$ and $\boldsymbol{C}$ are given as

$$\boldsymbol{A}(f_{\mathrm{ref}}(k)) = \left(\begin{array}{c|ccc} 1 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{A}_1(f_{\mathrm{ref}}(k)) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{A}_{n_{\mathrm{u}}}(f_{\mathrm{ref}}(k)) \end{array}\right) \tag{2.47}$$

and

$$\boldsymbol{C} = \left(1 \,\vdots\, \boldsymbol{C}_1 \quad \cdots \quad \boldsymbol{C}_{n_{\mathrm{u}}}\right), \tag{2.48}$$

respectively.

The individual block entries in the matrices $\boldsymbol{A}(k)$ and $\boldsymbol{C}$ are

$$\boldsymbol{A}_i(f_{\mathrm{ref}}(k)) = \begin{pmatrix} \cos\left(2\pi l_i f_{\mathrm{ref}}(k) T_{\mathrm{s}}\right) & \sin\left(2\pi l_i f_{\mathrm{ref}}(k) T_{\mathrm{s}}\right) \\ -\sin\left(2\pi l_i f_{\mathrm{ref}}(k) T_{\mathrm{s}}\right) & \cos\left(2\pi l_i f_{\mathrm{ref}}(k) T_{\mathrm{s}}\right) \end{pmatrix} \tag{2.49}$$

and

$$\boldsymbol{C}_i = \begin{pmatrix} 1 & 0 \end{pmatrix}, \tag{2.50}$$

respectively.

The state vector of the model (2.46) is expressed as

$$\boldsymbol{x}(k) = \left(x_0^{\mathrm{T}}(k) \quad \boldsymbol{x}_1^{\mathrm{T}}(k) \quad \cdots \quad \boldsymbol{x}_{n_{\mathrm{u}}}^{\mathrm{T}}(k)\right)^{\mathrm{T}}. \tag{2.51}$$

Based on the model (2.46) and the known signal $u(k)$, the order $u_i(k)$ can be reconstructed under different assumptions on the term $e(k)$:

(i) When $\{e(k)\}$ is a white noise process, and the state $\boldsymbol{x}(0)$ is a random variable with constant mean and constant variance, and $e(k)$ is uncorrelated with $\boldsymbol{x}(0)$, Equations (2.7) to (2.11) (Kalman filter recursion equations) can be used directly for the model (2.46), then with the estimated state $\hat{\boldsymbol{x}}(k)$, the reconstructed orders $\hat{u}_i(k) = \boldsymbol{C}_i \hat{\boldsymbol{x}}_i(k)$ for $i = 1, 2, \ldots, n_{\mathrm{u}}$.

(ii) When $e(k) \equiv 0$ and the frequency $f_{\mathrm{ref}}(k)$ is constant, the conventional observer introduced in Section 2.2.1 can be implemented such that the orders $u_i(k)$, $i = 1, 2, \ldots, n_{\mathrm{u}}$ can be reconstructed.

2.4.2 Order Tracking with Time-Variant Amplitude

According to the model (2.46), when the amplitude $A_0(k)$ is time-variant, the model (2.46) can still hold. However, besides the amplitude $A_0(k)$ if the amplitude $A_i(k)$ is time-variant, it is not suitable for us to use the state-space model (2.46) to represent the signal $u(k)$, and the reason is that for the modeling of the signal $u(k)$, the transition equation $\boldsymbol{x}(k+1) = \boldsymbol{A}(f_{\mathrm{ref}}(k))\boldsymbol{x}(k)$ will not be accurate enough if $A_i(k)$ is time-variant, and the corresponding reason can be found according to Equations (2.44) and (2.45). Furthermore, if $\{e(k)\}$ is a white noise process, and $\boldsymbol{x}(0)$ is a random variable with constant mean and constant variance, and $e(k)$ is uncorrelated with $\boldsymbol{x}(0)$, the Kalman filer with fading memory introduced in Section 2.2.4.1 can be used to handle the equivalent mismodeling problem. Furthermore, if the numerical precision problem simultaneously

occurs, the square-root Kalman filter with fading memory as depicted in Section 2.2.4.2 should be implemented.

2.4.3 Order Tracking with Unknown Instantaneous Frequency

The above cases of order tracking are based on the assumption that the instantaneous frequency $f_i(k)$ is known, however, sometimes the information of the instantaneous frequency is not available, thus an approach to handle the order tracking problem with unknown instantaneous frequency is necessary. Based on the frequency tracker investigated in [33], the order tracking problem without knowing the information of the instantaneous frequency can still be handled. The idea behind the order tracking without known instantaneous frequency is depicted below.

Regard the unknown reference frequency $f_{\mathrm{ref}}(k)$ as a state variable $x_{\mathrm{f}}(k)$, and augment it with the state of the model (2.46) in which $\{e(k)\}$ is still assumed to be a white noise process, then

$$\begin{cases} \begin{pmatrix} \boldsymbol{x}(k+1) \\ x_{\mathrm{f}}(k+1) \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}(f_{\mathrm{ref}}(k)) & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}(k) \\ x_{\mathrm{f}}(k) \end{pmatrix}, \\ u(k) = \begin{pmatrix} \boldsymbol{C} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}(k) \\ x_{\mathrm{f}}(k) \end{pmatrix} + e(k). \end{cases} \tag{2.52}$$

According to Equation (2.47) and Equation (2.49), it can be known that the matrix $\boldsymbol{A}(f_{\mathrm{ref}}(k))$ is a function of the frequency $f_{\mathrm{ref}}(k)$, as a result the model (2.52) is a nonlinear model, the EKF introduced in Section 2.3 can be used to track both order $u_i(k)$ and frequency $f_{\mathrm{ref}}(k)$. If necessary, for implementation the form of EKF should be modified, e.g., when the amplitude $A_i(k)$ (except of $A_0(k)$) is time-variant. The stability analysis of the EKF used for the order tracking can be referred to [35].

2.5 Linear System Identification in Time Domain

State-space models are good models for modern control and estimation techniques, moreover many industrial processes can be described very accurately by this type of models [36]. Subspace identification techniques are effective for identifying state-space models. Based on the points mentioned above, in this section, time-domain subspace identification methods [22, 36–40] for discrete-time LTI state-space models are reviewed.

Given the following discrete-time causal state-space model with unknown parameters:

$$\begin{cases} \boldsymbol{x}(k+1, \boldsymbol{\theta}_{\mathrm{s}}^*) = \boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{s}}^*)\boldsymbol{x}(k, \boldsymbol{\theta}_{\mathrm{s}}^*) + \boldsymbol{B}(\boldsymbol{\theta}_{\mathrm{s}}^*)\boldsymbol{u}(k) + \boldsymbol{w}(k), \\ \boldsymbol{y}(k) = \boldsymbol{C}(\boldsymbol{\theta}_{\mathrm{s}}^*)\boldsymbol{x}(k, \boldsymbol{\theta}_{\mathrm{s}}^*) + \boldsymbol{D}(\boldsymbol{\theta}_{\mathrm{s}}^*)\boldsymbol{u}(k) + \boldsymbol{v}(k), \end{cases} \tag{2.53}$$

where $\boldsymbol{x}(k, \boldsymbol{\theta}_{\mathrm{s}}^*) \in \mathbb{R}^{n_{\mathrm{m}}}$, $\boldsymbol{u}(k) \in \mathbb{R}^m$, and $\boldsymbol{y}(k) \in \mathbb{R}^p$ represent the state vector, input signal, and output signal, respectively. The parameter vector $\boldsymbol{\theta}_{\mathrm{s}}^*$ represents unknown parameters in the state-space model, and the parameter vector $\boldsymbol{\theta}_{\mathrm{s}}^* \in \boldsymbol{\Omega}_{\mathrm{s}}$ is formed by stacking the unknown parameters which represent the total elements in the state-space matrices of the model (2.53), and $\boldsymbol{\Omega}_{\mathrm{s}}$ denotes the specified parameter set which constraints the parameter vector. $\{\boldsymbol{w}(k)\}$ and $\{\boldsymbol{v}(k)\}$ are white noise processes, and $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ are uncorrelated with $\boldsymbol{x}(0, \boldsymbol{\theta}_{\mathrm{s}}^*)$ (a random variable with constant mean and constant variance) and with each

other. $\boldsymbol{u}(k)$ is uncorrelated with $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$. The covariance matrix of $\{\boldsymbol{w}(k)\}$ and the covariance function of $\{\boldsymbol{v}(k)\}$ are $\boldsymbol{Q}\delta_{kj}$ and $\boldsymbol{R}\delta_{kj}$, respectively. $\mathrm{E}\left[\boldsymbol{w}(k)\boldsymbol{v}^{\mathrm{T}}(j)\right] = \boldsymbol{M}\delta_{kj}$. $\boldsymbol{Q}$, $\boldsymbol{R}$, and $\boldsymbol{M}$ are constant and unknown. The matrix pair $(\boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{s}}^{*}), \boldsymbol{C}(\boldsymbol{\theta}_{\mathrm{s}}^{*}))$ is observable, and the matrix pair $(\boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{s}}^{*}), \boldsymbol{B}(\boldsymbol{\theta}_{\mathrm{s}}^{*}))$ and the matrix pair $\left(\boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{s}}^{*}), \sqrt{\boldsymbol{Q}}\right)$ are reachable. The state equation is asymptotically stable.

Suppose that the initial time is in the infinitely remote past, and the model (2.53) satisfies the condition of obtaining the steady-state Kalman filter introduced in Section 2.2.4.3. So the optimal one-step-ahead predictor of $\boldsymbol{y}(k)$ in the model (2.53) can be given by

$$\begin{cases} \hat{\boldsymbol{x}}^{-}(k+1, \boldsymbol{\theta}_{\mathrm{s}}) = \boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{s}})\hat{\boldsymbol{x}}^{-}(k, \boldsymbol{\theta}_{\mathrm{s}}) + \boldsymbol{B}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{u}(k) + \boldsymbol{K}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{\varepsilon}(k, \boldsymbol{\theta}_{\mathrm{s}}), \\ \hat{\boldsymbol{y}}^{-}((k, \boldsymbol{\theta}_{\mathrm{s}}) = \boldsymbol{C}(\boldsymbol{\theta}_{\mathrm{s}})\hat{\boldsymbol{x}}^{-}(k, \boldsymbol{\theta}_{\mathrm{s}}) + \boldsymbol{D}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{u}(k), \end{cases} \tag{2.54}$$

where the parameter vector $\boldsymbol{\theta}_{\mathrm{s}}$ is formed by stacking the unknown parameters which represent the total elements in the state-space matrices of the model (2.54), and

$$\boldsymbol{\varepsilon}(k, \boldsymbol{\theta}_{\mathrm{s}}) = \boldsymbol{y}(k) - \boldsymbol{C}(\boldsymbol{\theta}_{\mathrm{s}})\hat{\boldsymbol{x}}^{-}(k, \boldsymbol{\theta}_{\mathrm{s}}) - \boldsymbol{D}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{u}(k), \tag{2.55}$$

and the steady-state Kalman gain $\boldsymbol{K}(\boldsymbol{\theta}_{\mathrm{s}})$ is given by

$$\boldsymbol{K}(\boldsymbol{\theta}_{\mathrm{s}}) = \left(\boldsymbol{M} + \boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{P}^{-}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{C}^{\mathrm{T}}(\boldsymbol{\theta}_{\mathrm{s}})\right)\left(\boldsymbol{C}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{P}^{-}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{C}^{\mathrm{T}}(\boldsymbol{\theta}_{\mathrm{s}}) + \boldsymbol{R}\right)^{-1}, \tag{2.56}$$

where $\boldsymbol{P}^{-}(\boldsymbol{\theta}_{\mathrm{s}})$ is the solution of the algebraic Riccati equation

$$\begin{aligned} \boldsymbol{P}^{-}(\boldsymbol{\theta}_{\mathrm{s}}) = {} & \boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{P}^{-}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{\theta}_{\mathrm{s}}) \\ & -\boldsymbol{K}(\boldsymbol{\theta}_{\mathrm{s}})\left(\boldsymbol{C}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{P}^{-}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{C}^{\mathrm{T}}(\boldsymbol{\theta}_{\mathrm{s}}) + \boldsymbol{R}\right)\boldsymbol{K}^{\mathrm{T}}(\boldsymbol{\theta}_{\mathrm{s}}) \\ & +\boldsymbol{E}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{Q}\boldsymbol{E}^{\mathrm{T}}(\boldsymbol{\theta}_{\mathrm{s}}). \end{aligned} \tag{2.57}$$

Based on the predictor (2.54), an alternative form can be used to represent the state-space model (2.53), so our identification problem can be transform into the problem of the identification of the model

$$\begin{cases} \hat{\boldsymbol{x}}^{-}(k+1, \boldsymbol{\theta}_{\mathrm{s}}) = \boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{s}})\hat{\boldsymbol{x}}^{-}(k, \boldsymbol{\theta}_{\mathrm{s}}) + \boldsymbol{B}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{u}(k) + \boldsymbol{K}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{\varepsilon}(k, \boldsymbol{\theta}_{\mathrm{s}}), \\ \boldsymbol{y}(k) = \boldsymbol{C}(\boldsymbol{\theta}_{\mathrm{s}})\hat{\boldsymbol{x}}^{-}(k, \boldsymbol{\theta}_{\mathrm{s}}) + \boldsymbol{D}(\boldsymbol{\theta}_{\mathrm{s}})\boldsymbol{u}(k) + \boldsymbol{\varepsilon}(k, \boldsymbol{\theta}_{\mathrm{s}}), \end{cases} \tag{2.58}$$

which is called the innovation form of the model (2.53) [22, 37].

From the model (2.53) (without displaying the symbol " $\boldsymbol{\theta}_{\mathrm{s}}$ "), the following equation can be obtained:

$$\boldsymbol{Y}_r(k) = \boldsymbol{O}_r\boldsymbol{x}(k) + \boldsymbol{S}_r\boldsymbol{U}_r(k) + \boldsymbol{V}_r(k), \tag{2.59}$$

where

$$\boldsymbol{Y}_r(k) = \begin{pmatrix} \boldsymbol{y}(k) \\ \boldsymbol{y}(k+1) \\ \vdots \\ \boldsymbol{y}(k+r-1) \end{pmatrix}, \tag{2.60}$$

$$\boldsymbol{U}_r(k) = \begin{pmatrix} \boldsymbol{u}(k) \\ \boldsymbol{u}(k+1) \\ \vdots \\ \boldsymbol{u}(k+r-1) \end{pmatrix}, \tag{2.61}$$

$$\boldsymbol{O}_r = \begin{pmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \vdots \\ \boldsymbol{CA}^{r-1} \end{pmatrix}, \tag{2.62}$$

and

$$\boldsymbol{S}_r = \begin{pmatrix} \boldsymbol{D} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{CB} & \boldsymbol{D} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{CA}^{r-2}\boldsymbol{B} & \boldsymbol{CA}^{r-3}\boldsymbol{B} & \cdots & \boldsymbol{CB} & \boldsymbol{D} \end{pmatrix}, \tag{2.63}$$

and the ith block component of $\boldsymbol{V}_r(k)$

$$\boldsymbol{V}_r^{(i)}(k) = \boldsymbol{CA}^{i-2}\boldsymbol{w}(k) + \boldsymbol{CA}^{i-3}\boldsymbol{w}(k+1) + \ldots + \boldsymbol{Cw}(k+i-2) + \boldsymbol{v}(k+i-1), \tag{2.64}$$

and $r \geq n_{\mathrm{m}}$.

Based on the input-output (IO) data of the model (2.53), i.e., $(\boldsymbol{u}(k))_{k=0}^{N-1}$ and $(\boldsymbol{y}(k))_{k=0}^{N-1}$, introduce the following equations

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_r(s_{\mathrm{m}}) & \boldsymbol{Y}_r(s_{\mathrm{m}}+1) & \cdots & \boldsymbol{Y}_r(s_{\mathrm{m}}+M-1) \end{pmatrix}, \tag{2.65}$$

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}(s_{\mathrm{m}}) & \boldsymbol{x}(s_{\mathrm{m}}+1) & \cdots & \boldsymbol{x}(s_{\mathrm{m}}+M-1) \end{pmatrix}, \tag{2.66}$$

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_r(s_{\mathrm{m}}) & \boldsymbol{U}_r(s_{\mathrm{m}}+1) & \cdots & \boldsymbol{U}_r(s_{\mathrm{m}}+M-1) \end{pmatrix}, \tag{2.67}$$

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_r(s_{\mathrm{m}}) & \boldsymbol{V}_r(s_{\mathrm{m}}+1) & \cdots & \boldsymbol{V}_r(s_{\mathrm{m}}+M-1) \end{pmatrix}, \tag{2.68}$$

where s_{m} is chosen as the largest value of s_1 and s_2 which are involved in the instrumental-variable matrix in Equation (2.73).

With Equations (2.65) to (2.68), Equation (2.59) can be rewritten as

$$\boldsymbol{Y} = \boldsymbol{O}_r\boldsymbol{X} + \boldsymbol{S}_r\boldsymbol{U} + \boldsymbol{V}. \tag{2.69}$$

According to the condition that the matrix pair $(\boldsymbol{A}, \boldsymbol{C})$ is observable, an analysis of the rank of $\boldsymbol{Y} - \boldsymbol{S}_r\boldsymbol{U} + \boldsymbol{V}$ would directly provide the rank of $\boldsymbol{X}$, because $\mathrm{rank}\,(\boldsymbol{O}_r\boldsymbol{X}) = \mathrm{rank}\,(\boldsymbol{Y} - \boldsymbol{S}_r\boldsymbol{U} + \boldsymbol{V})$, provided that $\boldsymbol{O}_r$ has full (column) rank [41], where $\mathrm{rank}\,(\cdot)$ denotes the rank of a matrix. Therefore the order of the model (2.53) to be identified can be determined when $M \geq r \geq n_{\mathrm{m}}$. The value of $-\boldsymbol{S}_r\boldsymbol{U} + \boldsymbol{V}$ is unknown, but if the effect of $\boldsymbol{U}$ and $\boldsymbol{V}$ can be eliminated, the model order can still be determined.

Eliminating the influence of $\boldsymbol{U}$ and $\boldsymbol{V}$ can be done by the orthogonal projection technique and the instrumental variable method, respectively [37]. By involving an orthogonal matrix and an instrumental-variable matrix in Equation (2.69), the following equation can

be obtained:

$$\frac{1}{M}\boldsymbol{O}_r\boldsymbol{X}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}+\frac{1}{M}\boldsymbol{V}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}=\boldsymbol{O}_r\frac{1}{M}\boldsymbol{X}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}+\frac{1}{M}\boldsymbol{V}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}=\frac{1}{M}\boldsymbol{Y}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}},\tag{2.70}$$

where

$$\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}=\boldsymbol{I}_M-\boldsymbol{U}^{\mathrm{T}}\left(\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}\right)^{-1}\boldsymbol{U},\tag{2.71}$$

if $\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}$ is nonsingular, and the matrix $\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}$ performs projection, orthogonal to the matrix $\boldsymbol{U}$ (see the books [22, 36] for the specific definition and properties (e.g., orthogonality) of orthogonal projection), i.e.,

$$\boldsymbol{U}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}=\boldsymbol{U}-\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}\left(\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}\right)^{-1}\boldsymbol{U}=\boldsymbol{0}.\tag{2.72}$$

The matrix $\boldsymbol{Z}\in\mathbb{R}^{s\times M}$ is denoted as

$$\boldsymbol{Z}=\begin{pmatrix}\boldsymbol{Z}^{(0)} & \boldsymbol{Z}^{(1)} & \cdots & \boldsymbol{Z}^{(M-1)}\end{pmatrix},\tag{2.73}$$

and the ith element in $\boldsymbol{Z}$ can be represented as

$$\boldsymbol{Z}^{(i)}=\begin{pmatrix}\boldsymbol{u}(i)\\ \vdots\\ \boldsymbol{u}(i+s_2-1)\\ \boldsymbol{y}(i)\\ \vdots\\ \boldsymbol{y}(i+s_1-1)\end{pmatrix}.\tag{2.74}$$

Remark 2.5. $\boldsymbol{Z}$ is called an instrumental-variable matrix [22, 37]. □

According to the properties of the instrumental-variable matrix $\boldsymbol{Z}$ [22], the following equations can be derived:

$$\lim_{M\to+\infty}\frac{1}{M}\boldsymbol{V}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}=\boldsymbol{0},\tag{2.75}$$

$$\operatorname{rank}\left(\lim_{M\to+\infty}\frac{1}{M}\boldsymbol{X}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}\right)=n_{\mathrm{m}},\tag{2.76}$$

and furthermore with the assumption that the matrix $\boldsymbol{O}_r$ has full rank, it can be obtained that

$$\operatorname{rank}\left(\lim_{M\to+\infty}\frac{1}{M}\boldsymbol{Y}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}\right)=n_{\mathrm{m}}\tag{2.77}$$

of which the derivation can be found in [22].

As mentioned above, $\operatorname{rank}(\boldsymbol{O}_r\boldsymbol{X})=\operatorname{rank}(\boldsymbol{Y}-\boldsymbol{S}_r\boldsymbol{U}+\boldsymbol{V})$ when $\boldsymbol{O}_r$ has full rank, thus in Equation (2.70) it can be known that based on checking the rank of the matrix $\lim_{M\to+\infty}\frac{1}{M}\boldsymbol{Y}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}$ or the rank of the matrix $\boldsymbol{O}_r$ (if the matrix $\boldsymbol{O}_r$ can be calculated), then the number n (i.e., the order of the model (2.53)) can immediately be determined.

Furthermore, based on the matrix $\boldsymbol{O}_r$ in Equation (2.62), the following two equations can be used for deriving the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$:

$$\hat{\boldsymbol{C}}=\boldsymbol{O}_r(1:p,1:n_{\mathrm{m}}),\tag{2.78}$$

$$\boldsymbol{O}_r(p+1:pr,1:n_{\mathrm{m}}) = \boldsymbol{O}_r(1:p(r-1),1:n_{\mathrm{m}})\hat{\boldsymbol{A}}, \tag{2.79}$$

where the notation $\boldsymbol{O}_r(1:p,1:n_{\mathrm{m}})$ denotes the matrix obtained by extracting the rows $1,2,\ldots,p$ and the columns $1,2,\ldots,n_{\mathrm{m}}$ from the matrix $\boldsymbol{O}_r$.

Therefore the value of the matrix $\boldsymbol{O}_r$ is necessary to be known for both model order determination and the calculation of the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$. The strategy for calculating the matrix $\boldsymbol{O}_r$ is illustrated as follow.

According to Equations (2.70) and (2.75), the following equation can be obtained:

$$\begin{aligned}\lim_{M\to+\infty}\left(\boldsymbol{O}_r\frac{1}{M}\boldsymbol{X}\boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}^{\mathrm{T}}}\boldsymbol{Z}^{\mathrm{T}}+\frac{1}{M}\boldsymbol{V}\boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}^{\mathrm{T}}}\boldsymbol{Z}^{\mathrm{T}}\right) &= \lim_{M\to+\infty}\left(\boldsymbol{O}_r\frac{1}{M}\boldsymbol{X}\boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}^{\mathrm{T}}}\boldsymbol{Z}^{\mathrm{T}}\right)\\ &= \lim_{M\to+\infty}\left(\frac{1}{M}\boldsymbol{Y}\boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}^{\mathrm{T}}}\boldsymbol{Z}^{\mathrm{T}}\right)\\ &= \boldsymbol{\Psi}.\end{aligned} \tag{2.80}$$

Remark 2.6. It should be mentioned that for the calculation of $\boldsymbol{\Psi}$ in Equation (2.80), one efficient numerical calculation is to apply the RQ factorization [42] of the data matrix

$$\begin{pmatrix}\boldsymbol{U}^{\mathrm{T}} & \boldsymbol{Z}^{\mathrm{T}} & \boldsymbol{Y}^{\mathrm{T}}\end{pmatrix}^{\mathrm{T}} = \boldsymbol{R}_{\mathrm{d}}\boldsymbol{Q}_{\mathrm{d}}, \tag{2.81}$$

where $\boldsymbol{R}_{\mathrm{d}}$ is here a lower triangular, $pr+mr+s$ square matrix, while $\boldsymbol{Q}_{\mathrm{d}}$ is an orthogonal $(pr+mr+s)\times M$ matrix. Specifically, first perform the RQ factorization on the matrix

$$\begin{pmatrix}\boldsymbol{U}\\ \boldsymbol{Z}\\ \boldsymbol{Y}\end{pmatrix} = \begin{pmatrix}\boldsymbol{R}_{11} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{R}_{31} & \boldsymbol{R}_{32} & \boldsymbol{R}_{33} & \boldsymbol{0}\end{pmatrix}\begin{pmatrix}\boldsymbol{Q}_1\\ \boldsymbol{Q}_2\\ \boldsymbol{Q}_3\\ \boldsymbol{Q}_4\end{pmatrix} \tag{2.82}$$

with the orthogonality of the matrix $\begin{pmatrix}\boldsymbol{Q}_1^{\mathrm{T}} & \boldsymbol{Q}_2^{\mathrm{T}} & \boldsymbol{Q}_3^{\mathrm{T}} & \boldsymbol{Q}_4^{\mathrm{T}}\end{pmatrix}^{\mathrm{T}}$ [22], i.e.,

$$\begin{cases}\boldsymbol{Q}_i\boldsymbol{Q}_j^{\mathrm{T}} = \boldsymbol{0}, & \text{if} \quad i\neq j,\\ \boldsymbol{Q}_i\boldsymbol{Q}_i^{\mathrm{T}} = \boldsymbol{I},\end{cases} \tag{2.83}$$

where $\boldsymbol{I}$ denotes the identity matrix without the dimension index, then it can be obtained that

$$\boldsymbol{Y}\boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}^{\mathrm{T}}}\boldsymbol{Z}^{\mathrm{T}} = \boldsymbol{R}_{32}\boldsymbol{R}_{22}^{\mathrm{T}}. \tag{2.84}$$

□

By performing the singular value decomposition (SVD) [42] of the matrix $\boldsymbol{\Psi}$ which is equal to $\lim_{M\to+\infty}\left(\frac{1}{M}\boldsymbol{R}_{32}\boldsymbol{R}_{22}^{\mathrm{T}}\right)$, it can be obtained that

$$\begin{aligned}\boldsymbol{\Psi} &= \boldsymbol{U}_{\mathrm{s}}\boldsymbol{S}_{\mathrm{s}}\boldsymbol{V}_{\mathrm{s}}^{\mathrm{T}}\\ &= \begin{pmatrix}\boldsymbol{U}_1 & \boldsymbol{U}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{S}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\begin{pmatrix}\boldsymbol{V}_1^{\mathrm{T}}\\ \boldsymbol{V}_2^{\mathrm{T}}\end{pmatrix}\\ &= \boldsymbol{U}_1\boldsymbol{S}_1\boldsymbol{V}_1^{\mathrm{T}},\end{aligned} \tag{2.85}$$

where the rank of the square matrix $\boldsymbol{S}_1$ is n_{m} (which is the model order).

By observing Equations (2.80) and (2.85), the matrix $\boldsymbol{O}_r$ can be calculated as

$$\boldsymbol{O}_r = \boldsymbol{U}_1 \boldsymbol{T}, \tag{2.86}$$

where a similarity transformation $\boldsymbol{T}$, is introduced to make the recovered observability matrix $\boldsymbol{O}_r$, e.g., the similarity transformation $\boldsymbol{T}$ can be chosen as an identity matrix or be set to $\boldsymbol{S}_1^{\frac{1}{2}}$, because it is not necessary to recover number-wise the original matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ from which the IO data was generated [36, 37].

In practice, the value of M is finite, but the calculation of the matrix $\boldsymbol{O}_r$ can still be conducted according to the above equations. Based on calculation of the matrix $\boldsymbol{O}_r$ in Equation (2.86), the model order can be determined, and the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ can be calculated based on Equations (2.78) and (2.79). The stability of the calculated matrix $\hat{\boldsymbol{A}}$ can be enforced by reflecting unstable poles into the unit disk [43].

Remark 2.7. In practice, the model order should be chosen according to the obvious gap in the singular value spectrum [44], specifically, the n most significant values of the singular values in the matrix $\boldsymbol{S}_{\mathrm{s}}$ are kept and the remaining ones are set to zeros. □

The calculation of the matrix $\boldsymbol{K}$ can be based on Equation (2.56) [36], however, as seen in Equation (2.56), besides the values of the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$, the values of the matrices $\boldsymbol{Q}$, $\boldsymbol{R}$, and $\boldsymbol{M}$ are also necessary. The values of $\boldsymbol{Q}$, $\boldsymbol{R}$, and $\boldsymbol{M}$ can be obtained by implementing a robust identification algorithm (see the chapter six in [36] or the paper [45]), thus the calculated value of the matrix $\boldsymbol{K}$ can be obtained based on Equation (2.56) and the estimates $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{C}}$, $\hat{\boldsymbol{Q}}$, $\hat{\boldsymbol{R}}$, and $\hat{\boldsymbol{M}}$. It should be noted that in the implementation of the robust identification algorithm, the RQ factorization technique is involved.

The above steps are based on an infinite-length IO data, however for implementation of the above steps in practice, the finite-length IO data should be substituted into Equation (2.65) and Equation (2.67), respectively, and then the value of matrix $\boldsymbol{O}_r$ can be calculated (see Equation (2.86)), afterwards the calculated $\boldsymbol{A}$, $\boldsymbol{C}$, and $\boldsymbol{K}$ can be obtained.

Based on the IO data $(\boldsymbol{u}(k))_{k=0}^{N-1}$ and $(\boldsymbol{y}(k))_{k=0}^{N-1}$ and the calculated matrices $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{C}}$, and $\hat{\boldsymbol{K}}$ from the IO data, the matrices $\boldsymbol{B}$ and $\boldsymbol{D}$ in the innovation model (2.58) can be estimated below, and there are two different kinds of cost functions for the estimation [37, 46]:

- *Prediction focus*
  For the prediction focus, the matrices $\boldsymbol{B}$ and $\boldsymbol{D}$ can be estimated from the following linear regression problem [36]

$$\begin{aligned}\min_{\boldsymbol{B},\boldsymbol{D},\boldsymbol{x}^-(0)} \frac{1}{N}\sum_{k=0}^{N-1} \| \boldsymbol{y}(k) - \hat{\boldsymbol{C}}\left(q\boldsymbol{I}_{n_{\mathrm{m}}} - \hat{\boldsymbol{A}} + \hat{\boldsymbol{K}}\boldsymbol{C}\right)^{-1}\left(\boldsymbol{B} - \hat{\boldsymbol{K}}\boldsymbol{D}\right)\boldsymbol{u}(k) - \boldsymbol{D}\boldsymbol{u}(k) \\ -\hat{\boldsymbol{C}}\left(q\boldsymbol{I}_{n_{\mathrm{m}}} - \hat{\boldsymbol{A}} + \hat{\boldsymbol{K}}\boldsymbol{C}\right)^{-1}\hat{\boldsymbol{K}}\boldsymbol{y}(k) - \hat{\boldsymbol{C}}\left(q\boldsymbol{I}_{n_{\mathrm{m}}} - \hat{\boldsymbol{A}} + \hat{\boldsymbol{K}}\boldsymbol{C}\right)^{-1}\boldsymbol{x}^-(0)\delta(k) \|_2^2,\end{aligned} \tag{2.87}$$

  where $\|\cdot\|_2$ denotes the 2-norm of a vector, q denotes the one-step forward-shift operator [37], and the matrix $\boldsymbol{D}$ and the initial state $\boldsymbol{x}^-(0)$ can easily be excluded or included if necessary, e.g., they can be set to zero or regarded as an independent parameter to be estimated.

- *Simulation focus*
  As a counterpart of prediction focus, without the weighting (i.e., the noise model $H(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})$), for the simulation focus, the matrices $\boldsymbol{B}$ and $\boldsymbol{D}$ can be estimated from the linear regression problem [36]

$$\min_{\boldsymbol{B},\boldsymbol{D},\boldsymbol{x}^{-}(0)} \frac{1}{N} \sum_{k=0}^{N-1} \| \boldsymbol{y}(k) - \hat{\boldsymbol{C}} \left(q\boldsymbol{I}_{n_{\mathrm{m}}} - \hat{\boldsymbol{A}}\right)^{-1} \boldsymbol{B}\boldsymbol{u}(k) - \boldsymbol{D}\boldsymbol{u}(k) - \hat{\boldsymbol{C}} \left(q\boldsymbol{I}_{n_{\mathrm{m}}} - \hat{\boldsymbol{A}}\right)^{-1} \boldsymbol{x}^{-}(0)\delta(k) \|_2^2, \tag{2.88}$$

  where the matrix $\boldsymbol{D}$ and the initial state $\boldsymbol{x}^{-}(0)$ can easily be excluded or included if necessary.

For implementation purpose, the optimization problems (2.87) and (2.88) can be rewritten as the forms without shift operator q, which can be referred to [36]. It should also be noted that the optimization problems (2.87) and (2.88) are obviously convex optimization problems.

Remark 2.8. As introduced in [37, 40], by adding two weighting matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ in Equation (2.85), a weighted matrix $\boldsymbol{\Psi}_{\mathrm{W}}$ can be obtained, i.e.,

$$\boldsymbol{\Psi}_{\mathrm{W}} = \boldsymbol{W}_1 \boldsymbol{\Psi} \boldsymbol{W}_2, \tag{2.89}$$

then different weightings can lead to different subspace identification methods, which can be illustrated as follows:

- Equation (2.85): $\boldsymbol{W}_1 = \boldsymbol{I}_r$, $\boldsymbol{W}_2 = \boldsymbol{I}_r$.
- MOESP [22, 47]: $\boldsymbol{W}_1 = \boldsymbol{I}_r$, $\boldsymbol{W}_2 = \left(\frac{1}{M}\boldsymbol{Z}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}\right)^{-1} \boldsymbol{Z}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}$.
- N4SID [36]: $\boldsymbol{W}_1 = \boldsymbol{I}_r$, $\boldsymbol{W}_2 = \left(\frac{1}{M}\boldsymbol{Z}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}\right)^{-1} \boldsymbol{Z}$.
- CVA [48]: $\boldsymbol{W}_1 = \left(\frac{1}{M}\boldsymbol{Y}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Y}\right)^{\frac{-1}{2}}$, $\boldsymbol{W}_2 = \left(\frac{1}{M}\boldsymbol{Z}\boldsymbol{\Pi}_{\boldsymbol{U}^{\mathrm{T}}}^{\perp}\boldsymbol{Z}^{\mathrm{T}}\right)^{\frac{-1}{2}}$.

As a result different subspace identification methods can be cast in one unified framework based on using the weighting matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$. □

Subspace identification methods (e.g., N4SID, MOESP, and CVA) have proven to be very successful for identification of multivariable, linear black-box state-space models. These methods are associated with a number of design variables [37], or user choices. These include prediction horizons (i.e., the values of r, s_1, and s_2), weighting matrices (i.e., $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$), and identification focuses (i.e., simulation focus and prediction focus). It is known that these choices may have a big influence on the model quality while there is no comprehensive theory for rational decision making [46].

Finally, a procedure of the time-domain subspace identification for estimating the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$ based on using the IO data $(\boldsymbol{u}(k))_{k=0}^{N-1}$ and $(\boldsymbol{y}(k))_{k=0}^{N-1}$ is illustrated in Algorithm 2.1.

Remark 2.9. In Algorithm 2.1, there is also an additional option of enforcing stability, as aforementioned. Algorithm 2.1 has been implemented in the Matlab function n4sid.m in [49]. □

Algorithm 2.1: Time-domain subspace identification.

1 Choose the values of r, s_1, and s_2 (i.e., prediction horizons), and then formulate the matrices $\boldsymbol{U}$ and $\boldsymbol{Y}$, and the instrumental matrix $\boldsymbol{Z}$ using Equation (2.65), Equation (2.67), and Equation (2.73), respectively.
2 With the matrices $\boldsymbol{U}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$, use RQ factorization to calculate the matrix $\boldsymbol{\Psi}$.
3 Choose weighting matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, and then calculate $\boldsymbol{\Psi}_{\mathrm{W}}$ in Equation (2.89).
4 Replace $\boldsymbol{\Psi}$ with $\boldsymbol{\Psi}_{\mathrm{W}}$ in Equation (2.85), and obtain $\boldsymbol{U}_1$, and then calculate the extended observability matrix $\boldsymbol{O}_r$ using Equation (2.86).
5 Determine the model order according to the matrix $\boldsymbol{S}_1$ in Equation (2.85), and compute the estimates of $\boldsymbol{A}$ and $\boldsymbol{C}$ using Equation (2.78) and Equation (2.79), respectively.
6 Estimate the matrices $\boldsymbol{Q}$, $\boldsymbol{R}$, and $\boldsymbol{M}$, and then calculate the steady-state Kalman gain $\boldsymbol{K}$.
7 Estimate $\boldsymbol{B}$, $\boldsymbol{D}$, and $\boldsymbol{x}^{-}(0)$ using either the prediction option (2.87) or simulation option (2.88) as per requirements.

2.6 Linear System Identification in Frequency Domain

In this section, it is assumed that the input signal and output signal of the model (2.53) are periodic signals, thus it is natural to use frequency-domain system identification methods because the spectrum leakage can be avoided for the fast Fourier transform (FFT) of periodic signals with integer number of periods [22, 50, 51]. Based on using the discrete Fourier transform (DFT) of a discrete-time signal with finite length, the DFT of both input sequence $(\boldsymbol{u}(k))_{k=0}^{N-1}$ and output sequence $(\boldsymbol{y}(k))_{k=0}^{N-1}$ of the model (2.53) can be obtained, and the DFT values of the input and output sequences are denoted as $\boldsymbol{U}(n)$ and $\boldsymbol{Y}(n)$ at the frequency nf_{s}/N (in Hz), respectively. The definition of discrete Fourier transform (DFT) is introduced as follows:

Definition 2.2. [52] The DFT of a finite discrete-time sequence $(u(k))_{k=0}^{N-1}$ can be defined as

$$U(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} u(k) e^{-j2\pi nk/N}, \tag{2.90}$$

where $j^2 = -1$, the index $n \in \mathbb{N}$ points to the frequency nf_{s}/N, and $u(k) \in \mathbb{R}$ satisfies the condition for DFT [51]. Some pitfalls such as the spectrum leakage in the computation of DFT can be referred to [51]. □

In frequency-domain system identification, with the DFT values of both input sequence $(\boldsymbol{u}(k))_{k=0}^{N-1}$ and output sequence $(\boldsymbol{y}(k))_{k=0}^{N-1}$ at the frequencies of interest, the model between $\boldsymbol{u}(k)$ and $\boldsymbol{y}(k)$ in either transfer function form [53] or state-space model form [50, 54] can be identified. In this section, a frequency-domain system identification method is introduced to identify the model in state-space form, specifically, the so-called local polynomial method (LPM) [55, 56] is first used to calculate the FRF values at the frequencies of interest, and then based on the FRF points the frequency-domain subspace identification method proposed in [57] is implemented, but the noise variance is not included here. For

more about subspace identification in frequency domain, refer to [50, 58]. Based on the above simplified introduction, below two steps are used to illustrate the total frequency-domain system identification method.

2.6.1 Step I: LPM

The model (2.53) is equivalent to the model (2.58), and convert the model (2.58) to the following the transfer function form:

$$\boldsymbol{y}(k) = \boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{u}(k) + \boldsymbol{H}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{e}(k), \tag{2.91}$$

where the parameter vector $\boldsymbol{\theta}_{\mathrm{p}}$ is formed by stacking the unknown parameters which represent the coefficients of the transfer functions in the model (2.91), and q^{-1} is the one-step backwards-shift operator, and

$$\boldsymbol{e}(k) = \boldsymbol{\varepsilon}(k, \boldsymbol{\theta}_{\mathrm{s}}), \tag{2.92}$$

$$\boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}}) = \boldsymbol{D}(\boldsymbol{\theta}_{\mathrm{s}}) + \boldsymbol{C}(\boldsymbol{\theta}_{\mathrm{s}}) \left(q\boldsymbol{I}_n - \boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{s}})\right)^{-1} \boldsymbol{B}(\boldsymbol{\theta}_{\mathrm{s}}), \tag{2.93}$$

$$\boldsymbol{H}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}}) = \boldsymbol{I}_p + \boldsymbol{C}(\boldsymbol{\theta}_{\mathrm{s}}) \left(q\boldsymbol{I}_n - \boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{s}})\right)^{-1} \boldsymbol{K}(\boldsymbol{\theta}_{\mathrm{s}}). \tag{2.94}$$

If the model uncertainty which is denoted as $\boldsymbol{y}_{\mathrm{s}}(k)$ is involved into the model (2.91), the following model can be obtained:

$$\boldsymbol{y}(k) = \boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{u}(k) + \boldsymbol{y}_{\mathrm{s}}(k) + \boldsymbol{H}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{e}(k). \tag{2.95}$$

Furthermore, for practical identification purpose, because the length of the IO data is finite for identification, i.e., $(\boldsymbol{u}(k))_{k=0}^{N-1}$ and $(\boldsymbol{y}(k))_{k=0}^{N-1}$, the model (2.95) must be extended with the initial conditions, i.e., the transient terms $\boldsymbol{t}_{\mathrm{G}}(k)$ and $\boldsymbol{t}_{\mathrm{H}}(k)$ caused by the dynamics of the model $\boldsymbol{G}$ (corresponding to $\boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})$) and the model $\boldsymbol{H}$ (corresponding to $\boldsymbol{H}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})$), respectively, should be involved [55], i.e.,

$$\begin{aligned}\boldsymbol{y}(k) &= \boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{u}(k) + \boldsymbol{t}_{\mathrm{G}}(k) + \boldsymbol{t}_{\mathrm{H}}(k) + \boldsymbol{y}_{\mathrm{s}}(k) + \boldsymbol{H}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{e}(k) \\ &= \boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{u}(k) + \boldsymbol{t}_{\mathrm{GH}}(k) + \boldsymbol{y}_{\mathrm{s}}(k) + \boldsymbol{H}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{e}(k),\end{aligned} \tag{2.96}$$

where $\boldsymbol{t}_{\mathrm{GH}}(k) = \boldsymbol{t}_{\mathrm{G}}(k) + \boldsymbol{t}_{\mathrm{H}}(k)$.

By conducting DFT (2.90) on the two sides of Equation (2.96) at the frequency Ω_n, the following equation without displaying the parameter vector $\boldsymbol{\theta}_{\mathrm{p}}$ can be obtained:

$$\begin{aligned}\boldsymbol{Y}(n) &= \boldsymbol{G}(\Omega_n)\boldsymbol{U}(n) + \boldsymbol{T}_{\mathrm{G}}(\Omega_n) + \boldsymbol{T}_{\mathrm{H}}(\Omega_n) + \boldsymbol{Y}_{\mathrm{s}}(n) + \boldsymbol{H}(\Omega_n)\boldsymbol{E}(n) \\ &= \boldsymbol{G}(\Omega_n)\boldsymbol{U}(n) + \boldsymbol{T}_{\mathrm{GH}}(\Omega_n) + \boldsymbol{Y}_{\mathrm{s}}(n) + \boldsymbol{H}(\Omega_n)\boldsymbol{E}(n),\end{aligned} \tag{2.97}$$

where $\Omega_n = e^{jw_nT_{\mathrm{s}}}$, and $w_n = 2\pi f_n$, $\boldsymbol{G}(\Omega_n)$ and $\boldsymbol{H}(\Omega_n)$ represent the FRF values of the model $\boldsymbol{G}$ and the model $\boldsymbol{H}$ at the specified frequency f_n, the terms $\boldsymbol{T}_{\mathrm{G}}(\Omega_n)$, $\boldsymbol{T}_{\mathrm{H}}(\Omega_n)$, and $\boldsymbol{T}_{\mathrm{GH}}(\Omega_n)$, which are DFTs of $\boldsymbol{t}_{\mathrm{G}}(k)$, $\boldsymbol{t}_{\mathrm{H}}(k)$, and $\boldsymbol{t}_{\mathrm{GH}}(k)$ at the frequency Ω_n, stand for the leakage error in the frequency domain, and $\boldsymbol{T}_{\mathrm{GH}}(\Omega_n) = \boldsymbol{T}_{\mathrm{G}}(\Omega_n) + \boldsymbol{T}_{\mathrm{H}}(\Omega_n)$, and the terms $\boldsymbol{E}(n)$ and $\boldsymbol{Y}_{\mathrm{s}}(n)$ are DFTs of $\boldsymbol{e}(k)$ and $\boldsymbol{y}_{\mathrm{s}}(k)$ at the frequency Ω_n, respectively.

Remark 2.10. If an integer number of periods of the steady-state response to a periodic excitation is measured, then the leakage error $\boldsymbol{T}_{\mathrm{G}}(\Omega_n)$ in Equation (2.97) is zero. Note,

however, that the leakage term $\boldsymbol{T}_{\mathrm{H}}(\Omega_n)$ corresponding to the noise $\boldsymbol{e}(k)$ (which is non-periodic) still remains different from zero [50]. □

The estimation of the value of $\boldsymbol{G}(\Omega_n)$ using LPM is based on the assumption that the FRF $\boldsymbol{G}(w)$ and the leakage $\boldsymbol{T}_{\mathrm{GH}}(w)$ are smooth functions of frequency $w \in \mathbb{R}$ in rad/s. Thus, they can be approximated by complex polynomials within a narrow frequency band. The polynomial approximations at frequencies w_{n+l} ($l = 0, \pm 1, \pm 2, \ldots$) of $\boldsymbol{G}(w)$ and $\boldsymbol{T}_{\mathrm{GH}}(w)$ of the order η and centered around frequency w_n are respectively given as

$$\boldsymbol{G}(\Omega_{n+l}) = \boldsymbol{G}(\Omega_n) + \sum_{s=1}^{\eta} \boldsymbol{g}_s(n) l^s + \boldsymbol{O}\left(\left(\frac{l}{N}\right)^{(\eta+1)}\right) \tag{2.98}$$

and

$$\boldsymbol{T}_{\mathrm{GH}}(\Omega_{n+l}) = \boldsymbol{T}_{\mathrm{GH}}(\Omega_n) + \sum_{s=1}^{\eta} \boldsymbol{t}_s(n) l^s + \frac{1}{\sqrt{N}}\boldsymbol{O}\left(\left(\frac{l}{N}\right)^{(\eta+1)}\right), \tag{2.99}$$

where $\boldsymbol{g}_s(n)$ and $\boldsymbol{t}_s(n)$ are the Taylor coefficients of $\boldsymbol{G}(w)$ and $\boldsymbol{T}_{\mathrm{GH}}(w)$, respectively. The value of M can be chosen by users, but usually η is set to 2 in the literature. The notation $\boldsymbol{O}(\cdot)$ is called " ordo " in Taylor's formula with remainder [59].

Substituting Equations (2.98) and (2.99) into Equation (2.97) at frequency w_{n+l}, it can be obtained that

$$\begin{aligned}\boldsymbol{Y}(n+l) &= (\boldsymbol{G}(\Omega_n) + \sum_{s=1}^{\eta} \boldsymbol{g}_s(n) l^s)\boldsymbol{U}(n+l) \\ &\quad + (\boldsymbol{T}_{\mathrm{GH}}(\Omega_n) + \sum_{s=1}^{\eta} \boldsymbol{t}_s(n) l^s) + \boldsymbol{V}(n+l) \\ &= \boldsymbol{\Theta}\boldsymbol{K}(n+l) + \boldsymbol{V}(n+l),\end{aligned} \tag{2.100}$$

where the term $\boldsymbol{V}(n+l) = \boldsymbol{Y}_{\mathrm{s}}(n+l) + \boldsymbol{H}(w_{n+l})\boldsymbol{E}(n+l)$, $\boldsymbol{\Theta}$ is the matrix of the unknown complex parameters, namely the Taylor coefficients of $\boldsymbol{G}(w)$ and $\boldsymbol{T}_{\mathrm{GH}}(w)$ at the frequency w_n, i.e.,

$$\boldsymbol{\Theta} = \begin{pmatrix}\boldsymbol{G}(\Omega_n) & \boldsymbol{g}_1(n) & \boldsymbol{g}_2(n) & \cdots & \boldsymbol{g}_\eta(n) & \boldsymbol{T}_{\mathrm{GH}}(\Omega_n) & t_1(n) & t_2(n) & \cdots & t_\eta(n)\end{pmatrix}, \tag{2.101}$$

and $\boldsymbol{K}(n+l)$ contains the input data

$$\boldsymbol{K}(n+l) = \begin{pmatrix}\boldsymbol{K}_1(l) \otimes \boldsymbol{U}(n+l) \\ \boldsymbol{K}_1(l)\end{pmatrix} \tag{2.102}$$

with $\boldsymbol{K}_1(l) = \begin{pmatrix}1 & l & \cdots & l^\eta\end{pmatrix}^{\mathrm{T}}$ and " $\otimes$ " denoting the Kronecker product [42].

Collecting Equation (2.100) for the $2q+1$ neighbours of frequency w_n ($l = -q, -q+1, \ldots, 0, \ldots, q-1, q$) gives the following equation

$$\boldsymbol{Y}_q = \boldsymbol{\Theta}\boldsymbol{K}_q + \boldsymbol{V}_q, \tag{2.103}$$

where $\boldsymbol{Y}_q$, $\boldsymbol{K}_q$, and $\boldsymbol{V}_q$ are matrices with the forms

$$\boldsymbol{Y}_q = \begin{pmatrix}\boldsymbol{Y}(n-q) & \boldsymbol{Y}(n-q+1) & \cdots & \boldsymbol{Y}(q) & \cdots & \boldsymbol{Y}(n+q)\end{pmatrix}, \tag{2.104}$$

$$\boldsymbol{K}_q = \begin{pmatrix} \boldsymbol{K}(n-q) & \boldsymbol{K}(n-q+1) & \cdots & \boldsymbol{K}(q) & \cdots & \boldsymbol{K}(n+q) \end{pmatrix}, \tag{2.105}$$

and

$$\boldsymbol{V}_q = \begin{pmatrix} \boldsymbol{V}(n-q) & \boldsymbol{V}(n-q+1) & \cdots & \boldsymbol{V}(q) & \cdots & \boldsymbol{V}(n+q) \end{pmatrix}, \tag{2.106}$$

respectively.

It should be noted that q is a parameter of LPM that has to be chosen by users under the constraint $2q+1 \geq (\eta+1)(m+1)$, with m being the number of the input signals of the model $\boldsymbol{G}$. The constraint ensures that Equation (2.103) is an over-determined set of equations for the unknown $\boldsymbol{\Theta}$ that is to be solved in a least-squares sense

$$\min_{\boldsymbol{\Theta}} \sum_{i=1}^{2q+1} \|\boldsymbol{Y}_q(1:p,i) - \boldsymbol{\Theta}\boldsymbol{K}_q(1:p,i)\|_2^2, \tag{2.107}$$

where p denotes the number of the output signals of the model $\boldsymbol{G}$. The notation $\boldsymbol{Y}_q(1:p,i)$ denotes the elements in the ith column of the matrix Y_q.

Remark 2.11. For the left and right borders of the frequency band of interest, the $2q+1$ neighbouring frequencies cannot be centered around the frequency w_n, but have to be shifted to the left or to the left as mentioned in [55]. Additionally, the choice of q is a trade-off between an effective noise reduction (big q) and a low interpolation error (small q) [55]. □

Actually the solution of Equation (2.107) can be directly obtained as

$$\hat{\boldsymbol{\Theta}} = \boldsymbol{Y}_q \boldsymbol{K}_q^{\dagger}, \tag{2.108}$$

where the symbol " † " denotes the Moore-Penrose inverse [42].

According to Equation (2.101) and the estimated value $\hat{\boldsymbol{\Theta}}$ which is the solution of the optimization problem (2.107), the estimated FRF at the frequency w_n can be obtained using

$$\hat{\boldsymbol{G}}(\Omega_n) = \hat{\boldsymbol{\Theta}} \begin{pmatrix} \boldsymbol{I}_m \\ \boldsymbol{0} \end{pmatrix}. \tag{2.109}$$

The steps above are repeated for every frequency of interest, and finally an estimated sequence $\left(\hat{\boldsymbol{G}}(\Omega_n)\right)_{n\in \mathrm{T}}$ can be obtained, where T denotes the set of the frequencies of interest.

2.6.2 Step II: Frequency-Domain Subspace Identification

According to the model (2.53), the model $\boldsymbol{G}$ can be represented by using the following state-space model without explicitly displaying the unknown parameter vector $\boldsymbol{\theta}_{\mathrm{s}}^*$ in the state-space matrices:

$$\begin{cases} \boldsymbol{x}_{\mathrm{G}}(k+1) = \boldsymbol{A}\boldsymbol{x}_{\mathrm{G}}(k) + \boldsymbol{B}\boldsymbol{u}_{\mathrm{G}}(k), \\ \boldsymbol{y}_{\mathrm{G}}(k) = \boldsymbol{C}\boldsymbol{x}_{\mathrm{G}}(k) + \boldsymbol{D}\boldsymbol{u}_{\mathrm{G}}(k), \end{cases} \tag{2.110}$$

where the state vector $\boldsymbol{x}_{\mathrm{G}}(k)$, the input $\boldsymbol{u}_{\mathrm{G}}(k)$, and the output $\boldsymbol{y}_{\mathrm{G}}(k)$ correspond to the model $\boldsymbol{G}$. $\boldsymbol{x}_{\mathrm{G}}(k)$, $\boldsymbol{u}_{\mathrm{G}}(k)$, and $\boldsymbol{y}_{\mathrm{G}}(k)$ are different with $\boldsymbol{x}(k)$, $\boldsymbol{u}(k)$, and $\boldsymbol{y}(k)$ in the model (2.53), respectively.

With the model (2.110) and the estimated FRF points $\left(\hat{\boldsymbol{G}}(\Omega_n)\right)_{n\in\mathrm{T}}$, in the following the specific frequency-domain subspace identification method proposed in [57] is introduced, and for more details it can also be referred to [50, 58].

First recall the definition of the one-sided z-transform:

Definition 2.3. [52] The one-sided z-transform (or say unilateral z-transform), $X(z)$, of the sequence $(x(k))_{k=0}^{+\infty}$ is given by

$$X(z)=\sum_{k=0}^{+\infty} z^{-k}x(k). \tag{2.111}$$

for all z such that the result of the summation is well defined, denoted by the region of convergence (ROC). □

The following shift property of the one-sided z-transform is critical, and not hard to establish. Suppose that $x(k) \overset{z}{\leftrightarrow} X(z)$ and $g(k) \overset{z}{\leftrightarrow} G(z)$. Then if $g(k) = x(k+1)$, it can be obtained that $G(z) = z\left(X(z) - x(0)\right)$. The z-transform can also be used for multidimensional functions.

With the above definition of the one-sided z-transform and the corresponding shift property, perform the one-sided z-transform on both sides of the state-space model (2.110) [60]

$$\begin{cases} z\boldsymbol{X}_{\mathrm{G}}(z) - z\boldsymbol{x}_{\mathrm{G}}(0) = \boldsymbol{A}\boldsymbol{X}_{\mathrm{G}}(z) + \boldsymbol{B}\boldsymbol{U}_{\mathrm{G}}(z), \\ \boldsymbol{Y}_{\mathrm{G}}(z) = \boldsymbol{C}\boldsymbol{X}_{\mathrm{G}}(z) + \boldsymbol{D}\boldsymbol{U}_{\mathrm{G}}(z), \end{cases} \tag{2.112}$$

where $z = e^{(\sigma_{\mathrm{re}}+jw)T_{\mathrm{s}}}$, and $\sigma_{\mathrm{re}} + jw \in \mathbb{C}$ is a complex value with the real part σ_{re} and the imaginary part w.

By assuming the initial value $\boldsymbol{x}_{\mathrm{G}}(0) = \boldsymbol{0}$ and setting $\sigma_{\mathrm{re}} = 0$, it can be obtained that

$$\begin{cases} \Omega_n\boldsymbol{X}_{\mathrm{G}}(n) = \boldsymbol{A}\boldsymbol{X}_{\mathrm{G}}(n) + \boldsymbol{B}\boldsymbol{U}_{\mathrm{G}}(n), \\ \boldsymbol{Y}_{\mathrm{G}}(n) = \boldsymbol{C}\boldsymbol{X}_{\mathrm{G}}(n) + \boldsymbol{D}\boldsymbol{U}_{\mathrm{G}}(n), \end{cases} \tag{2.113}$$

at the frequency w_n, and $\boldsymbol{U}_{\mathrm{G}}(n)$, $\boldsymbol{Y}_{\mathrm{G}}(n)$, and $\boldsymbol{X}_{\mathrm{G}}(n)$ denote the DFTs of $(\boldsymbol{u}_{\mathrm{G}}(k))_{k=0}^{N-1}$, $(\boldsymbol{y}_{\mathrm{G}}(k))_{k=0}^{N-1}$, and $(\boldsymbol{x}_{\mathrm{G}}(k))_{k=0}^{N-1}$ in the model (2.110) at the frequency w_n.

Because $\boldsymbol{G}(\Omega_n)\boldsymbol{I}_m = \boldsymbol{G}(\Omega_n)$, in the model (2.113) $\boldsymbol{U}_{\mathrm{G}}(n)$, $\boldsymbol{Y}_{\mathrm{G}}(n)$, and $\boldsymbol{X}_{\mathrm{G}}(n)$ can be replaced with $\boldsymbol{I}_m$, $\boldsymbol{G}(\Omega_n)$, and $\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n)$, respectively, where $\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) \in \mathbb{C}^{n_{\mathrm{m}}\times m}$ (which is a DFT value at the frequency w_n) denotes the state of the following state-space model

$$\begin{cases} \Omega_n\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) = \boldsymbol{A}\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \boldsymbol{B}\boldsymbol{I}_m, \\ \boldsymbol{G}(\Omega_n) = \boldsymbol{C}\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \boldsymbol{D}\boldsymbol{I}_m, \end{cases} \tag{2.114}$$

i.e.,

$$\begin{cases} \Omega_n\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) = \boldsymbol{A}\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \boldsymbol{B}, \\ \boldsymbol{G}(\Omega_n) = \boldsymbol{C}\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \boldsymbol{D}. \end{cases} \tag{2.115}$$

Multiply the second equation of (2.115) by Ω_n^l, and elaborate it by repeatedly substituting $\Omega_n^l\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n)$ with the first equation of (2.115), i.e.,

$$\Omega_n^l\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) = \Omega_n^{l-1}\left(\boldsymbol{C}\Omega_n\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \Omega_n\boldsymbol{D}\right)$$

$$
\begin{aligned}
&= \Omega_n^{l-1} \left(\boldsymbol{C}\boldsymbol{A}\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \boldsymbol{C}\boldsymbol{B} + \Omega_n \boldsymbol{D} \right) \\
&= \Omega_n^{l-2} \left(\boldsymbol{C}\boldsymbol{A}^2\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \boldsymbol{C}\boldsymbol{A}\boldsymbol{B} + \Omega_n \boldsymbol{C}\boldsymbol{B} + \Omega_n^2 \boldsymbol{D} \right) \\
&= \dots \\
&= \boldsymbol{C}\boldsymbol{A}^l\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \left(\boldsymbol{C}\boldsymbol{A}^{l-1}\boldsymbol{B} + \Omega_n \boldsymbol{C}\boldsymbol{A}^{l-2}\boldsymbol{B} + \dots + \Omega_n^{l-1}\boldsymbol{C}\boldsymbol{B} + \Omega_n^l \boldsymbol{D} \right). \quad (2.116)
\end{aligned}
$$

Write down Equation (2.116) for $l = 0, \dots, r-1$ with $r \geq n_{\mathrm{m}}$:

$$
\begin{aligned}
&\boldsymbol{G}(\Omega_n) = \boldsymbol{C}\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \boldsymbol{D}, && (2.117) \\
&\Omega_n \boldsymbol{G}(\Omega_n) = \boldsymbol{C}\boldsymbol{A}\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \boldsymbol{C}\boldsymbol{B} + \Omega_n \boldsymbol{D}, && (2.118) \\
&\dots \\
&\Omega_n^{r-1} \boldsymbol{G}(\Omega_n) = \boldsymbol{C}\boldsymbol{A}^{r-1}\boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(n) + \boldsymbol{C}\boldsymbol{A}^{r-2}\boldsymbol{B} + \dots + \Omega_n^{r-2}\boldsymbol{C}\boldsymbol{B} + \Omega_n^{r-1}\boldsymbol{D}. && (2.119)
\end{aligned}
$$

Define

$$
\boldsymbol{W}_r(n) = \begin{pmatrix} 1 & \Omega_n & \cdots & \Omega_n^{r-1} \end{pmatrix}^{\mathrm{T}}, \quad (2.120)
$$

and then apply Equations (2.62), (2.63) and (2.120) to the r equations from Equation (2.117) to Equation (2.119), and do this for every element in the sequence $(\boldsymbol{G}(\Omega_n))_{n=0}^{M-1}$ ($n \in \mathrm{T}$, and here the set T contains all the natural numbers in the range $[0, M-1]$), afterwards the following relation can be obtained:

$$
\boldsymbol{G}_M = \boldsymbol{O}_r \boldsymbol{X}_M + \boldsymbol{S}_r \boldsymbol{I}_M, \quad (2.121)
$$

where $\boldsymbol{O}_r$ is defined in Equation (2.62), the matrices $\boldsymbol{G}_M$, $\boldsymbol{X}_M$, and $\boldsymbol{W}_M$ are respectively defined as

$$
\boldsymbol{G}_M = \begin{pmatrix} \boldsymbol{W}_r(1) \otimes \boldsymbol{G}(w_1) & \boldsymbol{W}_r(2) \otimes \boldsymbol{G}(w_2) & \cdots & \boldsymbol{W}_r(M) \otimes \boldsymbol{G}(w_M) \end{pmatrix}, \quad (2.122)
$$

$$
\boldsymbol{X}_M = \begin{pmatrix} \boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(1) & \boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(2) & \cdots & \boldsymbol{X}_{\mathrm{G}}^{\mathrm{a}}(M) \end{pmatrix}, \quad (2.123)
$$

and

$$
\boldsymbol{W}_M = \begin{pmatrix} \boldsymbol{W}_r(1) \otimes \boldsymbol{I}_m & \boldsymbol{W}_r(2) \otimes \boldsymbol{I}_m & \cdots & \boldsymbol{W}_r(M) \otimes \boldsymbol{I}_m \end{pmatrix}. \quad (2.124)
$$

The complex data equation in now converted into a real equation

$$
\boldsymbol{G}_M^{\mathrm{re}} = \boldsymbol{O}_r \boldsymbol{X}_M^{\mathrm{re}} + \boldsymbol{S}_r \boldsymbol{W}_M^{\mathrm{re}}. \quad (2.125)
$$

Remark 2.12. As seen in Equation (2.125), there is no noise-related term, thus it is not necessary to make the value of M be infinite in order to realize some asymptotic properties as discussed in [50], which is different with the frequency-domain system identification method introduced in [58]. □

In Equation (2.125), $\boldsymbol{G}_M^{\mathrm{re}} = \begin{pmatrix} \mathrm{Re}\,(\boldsymbol{G}_M) & \mathrm{Im}\,(\boldsymbol{G}_M) \end{pmatrix}$, $\boldsymbol{X}_M^{\mathrm{re}} = \begin{pmatrix} \mathrm{Re}\,(\boldsymbol{X}_M) & \mathrm{Im}\,(\boldsymbol{X}_M) \end{pmatrix}$, and $\boldsymbol{W}_N^{\mathrm{re}} = \begin{pmatrix} \mathrm{Re}\,(\boldsymbol{W}_M) & \mathrm{Im}\,(\boldsymbol{W}_M) \end{pmatrix}$, and $\mathrm{Re}\,(\cdot)$ and $\mathrm{Im}\,(\cdot)$ respectively represent the real part value and the imaginary part value of a complex data [61].

Equation (2.125) is similar to Equation (2.69), i.e., the orthogonal projection technique can be used to eliminate the effect from the term $\boldsymbol{S}_r \boldsymbol{W}_N^{\mathrm{re}}$, then the value of $\boldsymbol{O}_r$ can be calculated in Equation (2.125). Here the RQ factorization technique as introduced in

Remark 2.6 is implemented to calculate the matrix $\boldsymbol{O}_r$.

Two steps, which are used for describing the calculation of the matrix $\boldsymbol{O}_r$, are illustrated as follows:

- *Step A: RQ factorization*
  Perform the RQ factorization method:

$$\begin{pmatrix} \boldsymbol{W}_M^{\mathrm{re}} \\ \boldsymbol{G}_M^{\mathrm{re}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{Q}_1 \\ \boldsymbol{Q}_2 \\ \boldsymbol{Q}_3 \end{pmatrix}, \tag{2.126}$$

  then

$$\boldsymbol{G}_M^{\mathrm{re}} = \boldsymbol{R}_{21}\boldsymbol{Q}_1 + \boldsymbol{R}_{22}\boldsymbol{Q}_2, \tag{2.127}$$

  furthermore, according to the definition of the orthogonal projection matrix (see Equation (2.71)) and the orthogonality of the matrix $\begin{pmatrix} \boldsymbol{Q}_1^{\mathrm{T}} & \boldsymbol{Q}_2^{\mathrm{T}} & \boldsymbol{Q}_3^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}}$ [22], the orthogonal projection matrix $\boldsymbol{\Pi}_{\boldsymbol{W}_M^{\mathrm{re}}}^{\perp}$ of the matrix $\boldsymbol{W}_M^{\mathrm{re}}$ (which is equal to $\boldsymbol{R}_{11}\boldsymbol{Q}_1$) can be obtained:

$$\begin{aligned} \boldsymbol{\Pi}_{\boldsymbol{W}_M^{\mathrm{re}}}^{\perp} &= \boldsymbol{I} - (\boldsymbol{W}_M^{\mathrm{re}})^{\mathrm{T}} \left(\boldsymbol{W}_M^{\mathrm{re}} (\boldsymbol{W}_M^{\mathrm{re}})^{\mathrm{T}} \right)^{-1} \boldsymbol{W}_M^{\mathrm{re}} \\ &= \boldsymbol{I} - \boldsymbol{Q}_1^{\mathrm{T}} \boldsymbol{R}_{11}^{\mathrm{T}} \left(\boldsymbol{R}_{11} \boldsymbol{Q}_1 \boldsymbol{Q}_1^{\mathrm{T}} \boldsymbol{R}_{11} \right)^{-1} \boldsymbol{R}_{11} \boldsymbol{Q}_1 \\ &= \boldsymbol{I} - \boldsymbol{Q}_1^{\mathrm{T}} \boldsymbol{Q}_1 \\ &= \boldsymbol{Q}_2^{\mathrm{T}} \boldsymbol{Q}_2 + \boldsymbol{Q}_3^{\mathrm{T}} \boldsymbol{Q}_3, \end{aligned} \tag{2.128}$$

  and therefore

$$\begin{aligned} \boldsymbol{G}_M^{\mathrm{re}} \boldsymbol{\Pi}_{\boldsymbol{W}_M^{\mathrm{re}}}^{\perp} &= \boldsymbol{O}_r \boldsymbol{X}_M^{\mathrm{re}} \boldsymbol{\Pi}_{\boldsymbol{W}_M^{\mathrm{re}}}^{\perp} \\ &= \boldsymbol{R}_{21}\boldsymbol{Q}_1\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{Q}_2 + \boldsymbol{R}_{22}\boldsymbol{Q}_2\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{Q}_2 + \boldsymbol{R}_{21}\boldsymbol{Q}_1\boldsymbol{Q}_3^{\mathrm{T}}\boldsymbol{Q}_3 + \boldsymbol{R}_{22}\boldsymbol{Q}_2\boldsymbol{Q}_3^{\mathrm{T}}\boldsymbol{Q}_3 \\ &= \boldsymbol{R}_{22}\boldsymbol{Q}_2 \\ &\triangleq \boldsymbol{R}_{\boldsymbol{Q}}. \end{aligned} \tag{2.129}$$

  If the estimated FRF points $\left(\hat{\boldsymbol{G}}(\Omega_n) \right)_{n=0}^{M-1}$ are substituted into the above equations, the estimate $\hat{\boldsymbol{R}}_{\boldsymbol{Q}}$ can be obained.

- *Step B: Calculation of $\boldsymbol{O}_r$*
  Similar to Equation (2.80), perform SVD of the matrix $\hat{\boldsymbol{R}}_{\boldsymbol{Q}}$ in Equation (2.129), i.e.,

$$\hat{\boldsymbol{R}}_{\boldsymbol{Q}} = \boldsymbol{U}_{\mathrm{f}} \boldsymbol{S}_{\mathrm{f}} \boldsymbol{V}_{\mathrm{f}}^{\mathrm{T}}, \tag{2.130}$$

  where as illustrated in Remark 2.7 the matrix $\boldsymbol{S}_{\mathrm{f}}$ can also be used to determine the model order, i.e., the number of the most significant values of the singular values in the matrix $\boldsymbol{S}_{\mathrm{f}}$ can approximately be seen as the model order n_m.

  Here it should be noted that because $\mathrm{range}\,(\boldsymbol{R}_{22}) = \mathrm{range}\left(\hat{\boldsymbol{R}}_{\boldsymbol{Q}} \right)$ [47], in Equation (2.130) the matrix $\hat{\boldsymbol{R}}_{\boldsymbol{Q}}$ can be replaced with the matrix $\hat{\boldsymbol{R}}_{22}$.

Then by observing Equation (2.130), set

$$\hat{\boldsymbol{O}}_r = \boldsymbol{U}_{\mathrm{f}}(1:r, 1:n_{\mathrm{m}})\boldsymbol{T}_{\mathrm{f}}, \tag{2.131}$$

where $\hat{\boldsymbol{O}}_r$ denotes the estimate of $\boldsymbol{O}_r$, and $\boldsymbol{T}_{\mathrm{f}}$ represents a similarity transformation, and the reason why it is included in Equation (2.131) can be referred to the involved matrix $\boldsymbol{T}$ in Equation (2.86).

If $\boldsymbol{T}_{\mathrm{f}}$ is set to $\boldsymbol{I}$, it can be obtained that

$$\hat{\boldsymbol{O}}_r = \boldsymbol{U}_{\mathrm{f}}(1:r, 1:n_{\mathrm{m}}), \tag{2.132}$$

then the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ of the model $\boldsymbol{G}$ (i.e., the model (2.110)) can be derived using Equations (2.78) and (2.79). If necessary, the stability of $\boldsymbol{A}$ can be guaranteed using the methods (e.g., reflection) mentioned in the introduction of time-domain system identification.

With the estimated matrices $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{C}}$, the estimation of the matrices $\boldsymbol{B}$ and $\boldsymbol{D}$ of the model $\boldsymbol{G}$ can be derived by minimizing the following cost function:

$$\boldsymbol{J}_M(\boldsymbol{\theta}_{\mathrm{s}}) = \frac{1}{M}\sum_{n=0}^{M-1} \boldsymbol{\varepsilon}_{\mathrm{f}}^{\mathrm{H}}(\Omega_n, \hat{\boldsymbol{A}}, \boldsymbol{B}, \hat{\boldsymbol{C}}, \boldsymbol{D})\boldsymbol{\varepsilon}_{\mathrm{f}}(\Omega_n, \hat{\boldsymbol{A}}, \boldsymbol{B}, \hat{\boldsymbol{C}}, \boldsymbol{D}), \tag{2.133}$$

where the symbol " $^{\mathrm{H}}$ " denotes the complex-valued matrix transposition, and

$$\boldsymbol{\varepsilon}_{\mathrm{f}}(\Omega_n, \hat{\boldsymbol{A}}, \boldsymbol{B}, \hat{\boldsymbol{C}}, \boldsymbol{D}) = \mathrm{vec}\left(\boldsymbol{G}(\Omega_n, \hat{\boldsymbol{A}}, \boldsymbol{B}, \hat{\boldsymbol{C}}, \boldsymbol{D}) - \hat{\boldsymbol{G}}(\Omega_n)\right), \tag{2.134}$$

where the symbol $\mathrm{vec}\,(\cdot)$ denotes a column vector formed by stacking the columns of a matrix on top of each other, and

$$\boldsymbol{G}(\Omega_n, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}) = \boldsymbol{C}\left(\Omega_n \boldsymbol{I}_{n_m} - \boldsymbol{A}\right)^{-1}\boldsymbol{B} + \boldsymbol{D}. \tag{2.135}$$

The minimization problem of the cost function corresponding to the matrices $\boldsymbol{B}$ and $\boldsymbol{D}$ is a linear regression problem. With the above steps, the values of the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ of the model $\boldsymbol{G}$ can be calculated, i.e., the identified model $\hat{\boldsymbol{G}}$ in state-space form of $\boldsymbol{G}$ can be obtained.

With the above two steps (i.e., Step I: LPM and Step II: Frequency-domain subspace identification), the introduction for frequency-domain system identification is thus finished. Additionally, $\boldsymbol{J}_M(\boldsymbol{\theta}_{\mathrm{s}})$ can be minimized with respect to all the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$, and the specific method for dealing with complex data can be referred to [58]. Additionally, the nonlinear problem (2.133) can be solved using the Levenberg-Marquardt (LM) algorithm [58]. Thus if necessary, the LM algorithm can then be implemented for refining with the initialization of the frequency-domain subspace identification introduced above.

According to the above introduction of the frequency-domain system identification, it can be found that there are several tunable parameters which are η (see Equation (2.98)), q (see Equation (2.103)) and r (see $\boldsymbol{O}_r$). A summary of the frequency-domain system identification for the model $\boldsymbol{G}$ based on using the IO data $(\boldsymbol{u}(k))_{k=0}^{N-1}$ and $(\boldsymbol{y}(k))_{k=0}^{N-1}$ is

made, and the summary is illustrated in Algorithm 2.2.

Algorithm 2.2: Frequency-domain system identification for the model $\boldsymbol{G}$.

1 Use LPM in Section 2.6.2 to estimate the FRF points $\left(\hat{\boldsymbol{G}}(\Omega_n)\right)_{n\in\mathrm{T}}$.
2 Calculate the extended observability matrix $\boldsymbol{O}_r$ using Equation (2.131).
3 Compute estimates of $\boldsymbol{A}$ and $\boldsymbol{C}$ using Equation (2.78) and Equation (2.79), respectively.
4 Estimate $\boldsymbol{B}$ and $\boldsymbol{D}$ by solving the optimization problem (2.133).

Remark 2.13. In Algorithm 2.2, there is also an additional option of enforcing stability by using reflection as mentioned in Algorithm 2.1. Additionally, if necessary, LM algorithm can be used for refining the identified model from Algorithm 2.2. □

2.7 Nonlinear System Identification in Time Domain

Wiener and Hammerstein models the most known and the most widely implemented members of the class of block-oriented nonlinear models which can include complex models [62]. Because Wiener models have the ability to approximate almost any nonlinear systems [63]. In this thesis, only discrete-time Wiener model identification is used, and in this section the Wiener model identification is introduced. The Wiener model structure with output error is illustrated in Figure 2.1.

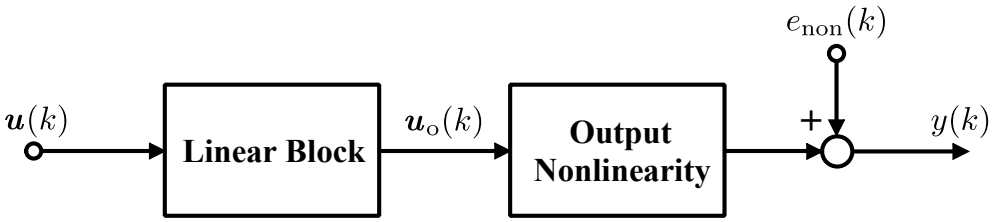

Figure 2.1: Wiener model with additive error.

As displayed in Figure 2.1, the nonlinear model contains both nonlinear part and linear part. The linear block may be parametric or nonparametric while the nonlinear elements may be memory or memoryless. The discrete-time nonlinear model considered in this section is in the form of

$$\begin{cases} \boldsymbol{u}_{\mathrm{o}}(k) = \boldsymbol{f}_{\mathrm{L}}(k) * \boldsymbol{u}(k), \\ \boldsymbol{y}(k) = \boldsymbol{f}_{\mathrm{N}}(\boldsymbol{u}_{\mathrm{o}}(k)) + \boldsymbol{e}(k), \end{cases} \tag{2.136}$$

where the symbol " $*$ " denotes the convolution, $\boldsymbol{f}_{\mathrm{L}}(k)$ represents the impulse response function of the linear block denoted as $\boldsymbol{F}_{\mathrm{L}}$, and $\boldsymbol{f}_{\mathrm{N}}(\cdot)$ denotes the output nonlinearity. As aforementioned, $\boldsymbol{u}(k) \in \mathbb{R}^m$ denotes the input signal, and $\boldsymbol{y}(k) \in \mathbb{R}^p$ denotes the output signal. The term $\boldsymbol{e}(k)$ denotes the output error.

Several assumptions on the nonlinear model (2.136) are made as follows:

(i) $\{\boldsymbol{e}(k)\}$ denotes a white noise process, and $\mathrm{E}\left[\boldsymbol{e}(k)\boldsymbol{e}^{\mathrm{T}}(j)\right] = \boldsymbol{R}\delta_{kj}$.

(ii) The input signal $\boldsymbol{u}(k)$ is uncorrelated with $\boldsymbol{e}(k)$.

(iii) Perform one-sided z-transform on $\boldsymbol{f}_{\mathrm{L}}(k)$, i.e., $\boldsymbol{f}_{\mathrm{L}}(k) \overset{z}{\hookleftarrow} \boldsymbol{F}_{\mathrm{L}}(z^{-1})$, where

$$\boldsymbol{F}_{\mathrm{L}}(z^{-1}) = \boldsymbol{C}_{\mathrm{L}}\left(z\boldsymbol{I} - \boldsymbol{A}_{\mathrm{L}}\right)^{-1}\boldsymbol{B}_{\mathrm{L}} + \boldsymbol{D}_{\mathrm{L}} \tag{2.137}$$

represents the discrete-time transfer function of the linear block, and the transfer function $\boldsymbol{F}_{\mathrm{L}}(z^{-1})$ is rational, asymptotically stable, and causal. The element in the ith row and jth column in the $p \times m$ matrix $\boldsymbol{F}_{\mathrm{L}}(z^{-1})$ is denoted as

$$F_{\mathrm{L}}^{ij}(z^{-1}) = \frac{B_{ij}(z^{-1})}{A_{ij}(z^{-1})} = \frac{b_1^{ij} + b_2^{ij}z^{-1} + b_3^{ij}z^{-2} + \ldots + b_{n_{\mathrm{b}}}^{ij}z^{-n_{\mathrm{b}}+1}}{1 + a_1^{ij}z^{-1} + a_2^{ij}z^{-2} + \ldots + a_{n_{\mathrm{a}}}^{ij}z^{-n_{\mathrm{a}}}}. \tag{2.138}$$

Based on $\boldsymbol{F}_{\mathrm{L}}(z^{-1})$, the following relation can be obtained:

$$\boldsymbol{u}_{\mathrm{o}}(k) = \boldsymbol{F}_{\mathrm{L}}(q^{-1})\boldsymbol{u}(k), \tag{2.139}$$

where $\boldsymbol{F}_{\mathrm{L}}(q^{-1})$ denotes the transfer operator of $\boldsymbol{F}_{\mathrm{L}}$.

(iv) The element in the ith row of the nonlinear function $\boldsymbol{f}_{\mathrm{N}}(\cdot)$ is a single-variable polynomial function, i.e., $f_{\mathrm{N}}^{i}(x) = c_1^i x^n + c_2^i x^{n-1} + \ldots + c_n^i x + c_{n+1}^i$, where x denotes the argument of the function $f_{\mathrm{N}}^{i}(\cdot)$, and c_j^i, $j = 1, 2, \ldots, n+1$, and n denote the coefficients and degree of the polynomial function, respectively.

Based on Equations (2.136) and (2.139), the simulated output can be obtained given the input $\boldsymbol{u}(k)$:

$$\hat{\boldsymbol{y}}(k, \boldsymbol{\theta}_{\mathrm{p}}) = \boldsymbol{f}_{\mathrm{N}}\left(\boldsymbol{F}_{\mathrm{L}}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{u}(k), \boldsymbol{\theta}_{\mathrm{p}}\right) + \boldsymbol{y}_{\mathrm{t}}(k, \boldsymbol{\theta}_{\mathrm{p}}), \tag{2.140}$$

where the parameter vector $\boldsymbol{\theta}_{\mathrm{p}}$ contains all the unknown parameters c_j^i, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, n+1$, $b_1^{ij}, b_2^{ij}, b_3^{ij}, \ldots, b_{n_{\mathrm{b}}}^{ij}$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, m$, and $a_1^{ij}, a_2^{ij}, \ldots, a_{n_{\mathrm{a}}}^{ij}$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, m$ in the model (2.136), and $\boldsymbol{y}_{\mathrm{t}}(k, \boldsymbol{\theta}_{\mathrm{p}})$ denotes represents a transient term caused by the dynamics of the model $\boldsymbol{F}_{\mathrm{L}}$.

As seen in Equation (2.140), the transient term $\boldsymbol{y}_{\mathrm{t}}(k, \boldsymbol{\theta}_{\mathrm{p}})$ is included, however, in the process of nonlinear system identification it is neglected, i.e.,

$$\hat{\boldsymbol{y}}(k, \boldsymbol{\theta}_{\mathrm{p}}) = \boldsymbol{f}_{\mathrm{N}}\left(\boldsymbol{F}_{\mathrm{L}}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{u}(k), \boldsymbol{\theta}_{\mathrm{p}}\right). \tag{2.141}$$

Given the IO data $(\boldsymbol{u}(k))_{k=0}^{N-1}$ and $(\boldsymbol{y}(k))_{k=0}^{N-1}$, the identification of the linear model $\boldsymbol{F}_{\mathrm{L}}$ and the nonlinear function $\boldsymbol{f}_{\mathrm{N}}(\cdot)$ can be summarized in Algorithm 2.3.

Remark 2.14. The optimization problem (2.142) should also be reformulated as a difference equation for implementation, and iterative optimization methods such as Gauss-Newton algorithm can be used for solving the optimization problem [64, 65]. The identification algorithm can be realized by the Matlab function nlhw.m in [49]. □

Algorithm 2.3: Nonlinear model identification algorithm.

1 Specify the number of poles and zeros of the linear model $\boldsymbol{F}_{\mathrm{L}}$.
2 Specify the number of the coefficients of the nonlinear function $\boldsymbol{f}_{\mathrm{N}}(\cdot)$.
3 Minimize the quadratic error criterion

$$J_N(\boldsymbol{\theta}_{\mathrm{p}}) = \frac{1}{N}\sum_{k=0}^{N-1} \parallel \boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k, \boldsymbol{\theta}_{\mathrm{p}}) \parallel_2^2 \tag{2.142}$$

to estimate the parameter vector $\boldsymbol{\theta}_{\mathrm{p}}$.

2.8 Summary

In this chapter, the techniques of the linear observer, linear Kalman filter, nonlinear Kalman filtering, the recursive model-based order tracking problem, and system identification are briefly introduced. To see more details about the observer, refer to [66], while more aspects of the linear Kalman filter such as the initialization (i.e., choice of initial values of $\boldsymbol{P}_0$ and $\boldsymbol{x}_0$, and their forgetting effects on the Kalman filter under certain conditions), stability, statistical properties, Kalman filter derivations, and practical considerations when using the Kalman filter can be referred to [24, 26, 26, 28, 31, 67–69]. For more details about nonlinear Kalman filters (e.g., the stability and optimality of nonlinear Kalman filtering), refer to [24, 26, 26, 28, 31, 67, 68], and it should be pointed out that nonlinear Kalman filtering is a complex problem, and it still needs more investigations. In the order tracking problem, under different cases such as the case of time-variant amplitude $A_i(k)$, several suggestions on how to implement the observer or Kalman filtering are given. Several kinds of system identification techniques are depicted. Some other aspects such as asymptotic analysis (i.e., asymptotic distribution of parameter estimates) under some assumptions (e.g., some assumptions on noise statistics), iterative optimization methods in identification (e.g., Gauss-Newton algorithm) are not depicted in detail. Additionally, some open problems in system identification are also not mentioned. These aspects and more details about system identification can be found in [37, 70] (time-series model identification), [22, 36–40, 71] (state-space model identification and subspace identification in time domain), [50] (subspace identification in frequency domain), [72, 73] (recursive system identification), [37, 70, 74] (non-parametric model identification, such as correlation analysis and spectral analysis), and [75] (errors-in-variables system identification). While for nonlinear system identification, refer to [37, 62, 63, 73, 76].

Chapter 3

Test Bench and Signal Processing

3.1 Introduction

This chapter introduces a diesel test bench and illustrates how the collected signals from the test bench are processed. After signal processing, the collected signals can then be used for system identification and validating the proposed cylinder pressure reconstruction methods. Below, in Section 3.2, a full description on the engine test bench is given including some key parameters of the engine. Signal processing of the collected data is depicted in Section 3.3. Finally, a summary is made in Section 3.4.

3.2 Test Bench and Data Acquisition

Figure 3.1 shows an image of the test bench[1], which consists of the control and measurement systems and the four-stroke diesel engine. Three necessary parameters for validating the proposed methods can be acquired from the test bench. Specifically, over different operating ranges, the piezoelectric transducer (integrated in the glow plug), the accelerometer, and the flywheel angular velocity sensor can be respectively used for synchronously measuring the cylinder pressure signal, the vibration signal of the engine block, and the engine crank angular speed signal. The accelerometer was mounted on the engine surface. The sampling frequency of the measurement system was chosen as 100 kHz. In this paper, the raw instantaneous engine speed, the engine vibration, and the cylinder pressure signal are called the measured signals. The main characteristics of the engine are displayed in the following Table 3.1.

Table 3.1: Engine main characteristics.

Parameter	Value
Combustion	Compression ignited
Stroke (S)	95.5 mm
Bore (D)	81 mm
Length of the connecting rod	144.11 mm
Crank radius (S/2)	47.75 mm
Number of cylinders	4

[1]The engine test bench was developed by Clausthal University of Technology.

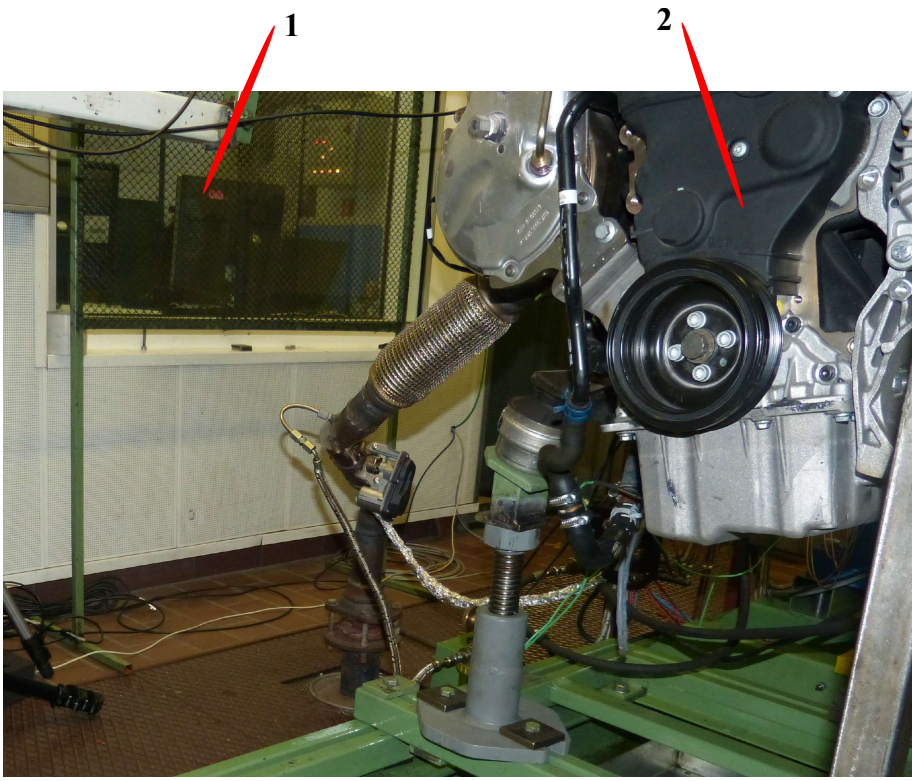

Figure 3.1: Test bench instrumentation (1: control and measurement systems; 2: Volkswagen 2.0 TDI diesel engine with four cylinders).

It is necessary to mention that the acceleration sensor was exactly mounted on the outer wall of the IC engine of the test bench.

3.3 Signal Processing

The collected signals from the test bench need to be processed for the purposes of system identification and the implementation of the proposed methods. Prior to introducing the signal processing for measurements, without confusion, the definition of the operating condition should be clarified. Engine operating condition means that the engine is running under a group of parameter settings, such as the instantaneous engine speed and load settings. Operating condition can be categorized into stationary operating condition and non-stationary operating condition. In this thesis, for ease of description, only the instantaneous engine speed (in rpm) and load (in Nm) are used to specify a stationary operating condition without including other parameters.

The collected measurements from the test bench over various operating conditions were used for identifying both linear models and nonlinear models (between four cylinder pressure signals and one engine vibration signal) and validating the proposed methods. To be more specific, for time-domain system identification, two datasets, A and B were respectively given for model validation and model identification. Both dataset A and dataset B contain the collected cylinder pressure, engine vibration, and instantaneous engine speed signal which are synchronized. The operating condition ranges of the datasets A and B are described in Table 3.2.

Table 3.2: Datasets for identification and validation.

Dataset	Content
Dataset A	$Z_1 \cdots Z_2 \cdots Z_3$
Dataset B	$Z_4 \cdots Z_5 \cdots Z_6 \cdots Z_7 \cdots Z_8 \cdots Z_9$

Specifically, in Table 3.2, the numbers Z_1, Z_2, and Z_3 represent the stationary operating conditions 1200 rpm, 60 Nm, 1200 rpm, 120 Nm, and 1200 rpm, 180 Nm, respectively. While Z_4, Z_5, and Z_6 represent the stationary operating conditions 2100 rpm, 60 Nm, 2100 rpm, 120 Nm, and 2100 rpm, 180 Nm, respectively. The numbers Z_7, Z_8, and Z_9 represent the stationary operating conditions 3000 rpm, 60 Nm, 3000 rpm, 120 Nm, and 3000 rpm, 180 Nm, respectively. The symbol " $\cdots$ " denotes the transient process between two stationary operating conditions. Between the operating condition Z_3 and the operating condition Z_4, there is either a transient process, which is not illustrated in the table.

While for frequency-domain system identification, three stationary operating conditions Z_2, Z_5, and Z_7 are used for model identification, and Z_t (which is a part of the transient process between the stationary operating condition Z_3 and the stationary operating condition Z_4, i.e., Z_3 and Z_4 are not included) is used for model validation.

The datasets displayed in Table 3.2 are also used for validating the effectiveness of the proposed cylinder pressure reconstruction methods. Specifically, nine stationary operating conditions Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8, and Z_9, and three non-stationary operating conditions Z_t, $\tilde{Z}_\mathrm{t}$ (which is a part of the transient process between the stationary operating condition Z_6 and the stationary operating condition Z_7, i.e., Z_6 and Z_7 are not included), and $\breve{Z}_\mathrm{t}$ (which is from Z_6 to Z_7, i.e., Z_6 and Z_7 are included) are used for validating the reconstruction methods.

In the following of this section, signal processing for the measurements (cylinder pressure signal, vibration signal, and instantaneous engine speed signal) is executed which includes a downsampling process (i.e., first choose an anti-aliasing filter, and then decrease the sampling frequency of the measurements) in Section 3.3.1, a cylinder pressure calibration process in Section 3.3.2, and an instantaneous engine speed calculation process in Section 3.3.3. The instantaneous engine speed calculation is necessary for the proposed cylinder pressure reconstruction methods in Chapter 5 and Chapter 7, therefore the accuracy of the calculation of the instantaneous engine speed should be guaranteed.

3.3.1 Low-Pass Filters and Downsampling

The aim of downsampling is to decrease the sampling frequency in the cylinder pressure reconstruction process. In the downsampling process, an anti-aliasing filter is necessary, and because the high frequencies of the cylinder pressure are not of interest, a low-pass filter is chosen as the anti-aliasing filter. An introduction of low-pass filters is first given before downsampling. According to the length of the impulse response of a filter, impulse response (IIR) filters and finite impulse response (FIR) filters are illustrated below.

3.3.1.1 IIR filters

A SISO discrete-time IIR system can be described by

$$\sum_{k=0}^{N} a_k y(n-k) = \sum_{r=0}^{M} b_r u(n-r), \tag{3.1}$$

where $u(k) \in \mathbb{R}$ and $y(k) \in \mathbb{R}$ denote the input and output of the system (3.1). $a_k \in \mathbb{R}$, $k = 0, 1, \ldots, N$ and $b_r \in \mathbb{R}$, $r = 0, 1, \ldots, M$ are constant coefficients.

By suitably choosing the values of a_k, $k = 0, 1, \ldots, N$ and b_r, $r = 0, 1, \ldots, M$, four kinds of discrete-time IIR filters, which are IIR low-pass filters, IIR high-pass filters, IIR band-pass filters, and IIR band-stop filters, can be obtained. Other three kinds of IIR filters can be derived based on low-pass filters, and an IIR low-pass filter can be derived based on the bilinear transformation of prototype analog filters such as Butterworth, Chebyshev, and elliptic function filters [77].

3.3.1.2 FIR filters

Modify Equation (3.1) by normalizing the coefficient of $y(n)$ to be unity (i.e., $a_0 = 1$), and the rearranged SISO discrete-time IIR system is of the form

$$y(n) + \sum_{k=1}^{N} a_k y(n-k) = \sum_{r=0}^{M} b_r u(n-r). \tag{3.2}$$

If all a_k coefficients are set to zero, then Equation (3.2) reduces to

$$y(n) = \sum_{r=0}^{M} b_r u(n-r). \tag{3.3}$$

Thus Equation (3.3) can be used to describe an FIR system, i.e., the impulse response of FIR system can be represented as

$$h(n) = \sum_{r=0}^{M} b_r \delta(n-r), \tag{3.4}$$

where $\delta(n)$ denotes the unit impulse [52], i.e.,

$$\delta(n) = \begin{cases} 0, & \text{if} \quad n \neq 0, \\ 1, & \text{if} \quad n = 0. \end{cases} \tag{3.5}$$

There are mainly three methods for the design of FIR filters [51]: a) impulse response truncation (windowing method), b) frequency sampling, and c) optimal method.

Compared with FIR filters, the major advantages of IIR filters is that for the same frequency characteristics, they require fewer coefficients than FIR filters. This leads to fewer operations, thus being able to achieve higher throughput. Meanwhile, IIR filters require the smallest storage requirement, since they need a least number of coefficients for achieving specified characteristics. In addition, the filtering of the collected data from the engine test bench does not need to result in a linear-phase characteristic, therefore finally an IIR low-pass filter (a Butterworth filter) was used for processing the collected

data.

3.3.1.3 Downsampling

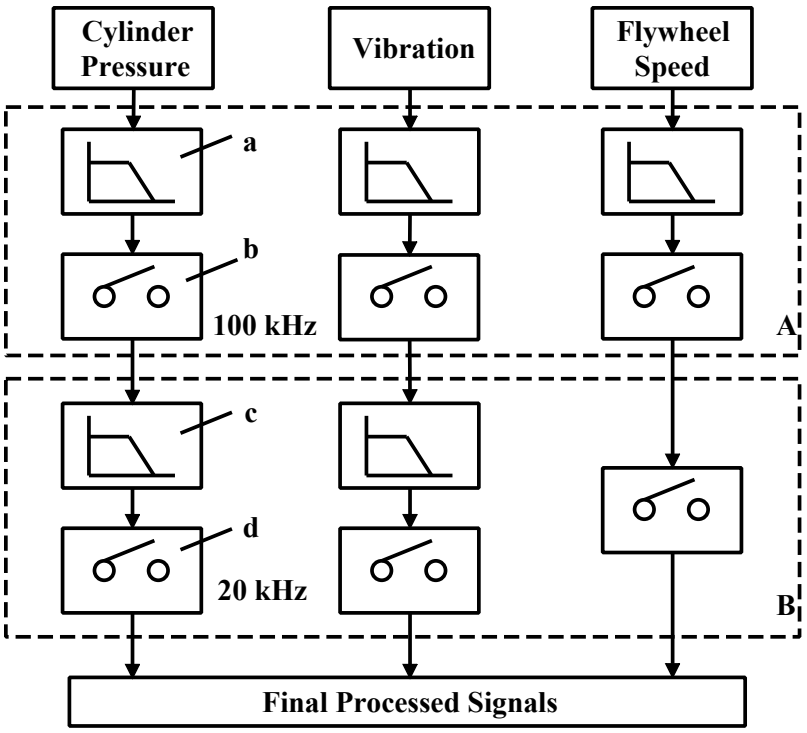

Figure 3.2: Collection and downsampling (a: low-pass filter; b: sampling; c: 7-th order Butterworth low-pass filter with cutoff frequency 1 kHz; d: sampling; A: data collection in the hardware (ETAS unit); B: downsampling).

All the data were collected from ETAS (a measurement device) illustrated in the part I of Figure 3.2, and the sampling frequency was 100 kHz. This preliminary sampling frequency is enough, i.e., the effective maximum frequency of the cylinder pressure signal can be kept without spectrum aliasing [52]. The downsampling process, which consists of a low-pass filter and a resampling process, is depicted in the part II of Figure 3.2. With the downsampling process shown in the part II of Figure 3.2, the sampling frequency for all the measurements were changed into 20 kHz. The sampling frequency after downsampling shown in Figure 3.2 is denoted as f_s which is equal to 20 kHz.

As aforementioned, the low-pass filtering is necessary before the downsampling process (from 100 kHz to 20 kHz), several 7th-order SISO Butterworth low-pass filters with different cutoff frequencies were used to determine a suitable cutoff frequency for the cylinder pressure signal under the stationary operating condition 3000 rpm, 120 Nm. Obviously, as shown in Figure 3.3[2], The frequency 1 kHz can be specified as the cutoff frequency. After the cutoff frequency checking procedure like above for several other stationary operating conditions, the cutoff frequencies of both cylinder pressure signal and engine vibration

[2]The unit of the horizontal axis is degree which is transformed from the unit seconds.

signal were finally chosen as 1 kHz. While the low-pass filter is not used for the instantaneous engine speed signal collected from ETAS, and the reason is that for the calculation of the instantaneous engine speed signal as illustrated in Section 3.3.3, just the pulse number is counted.

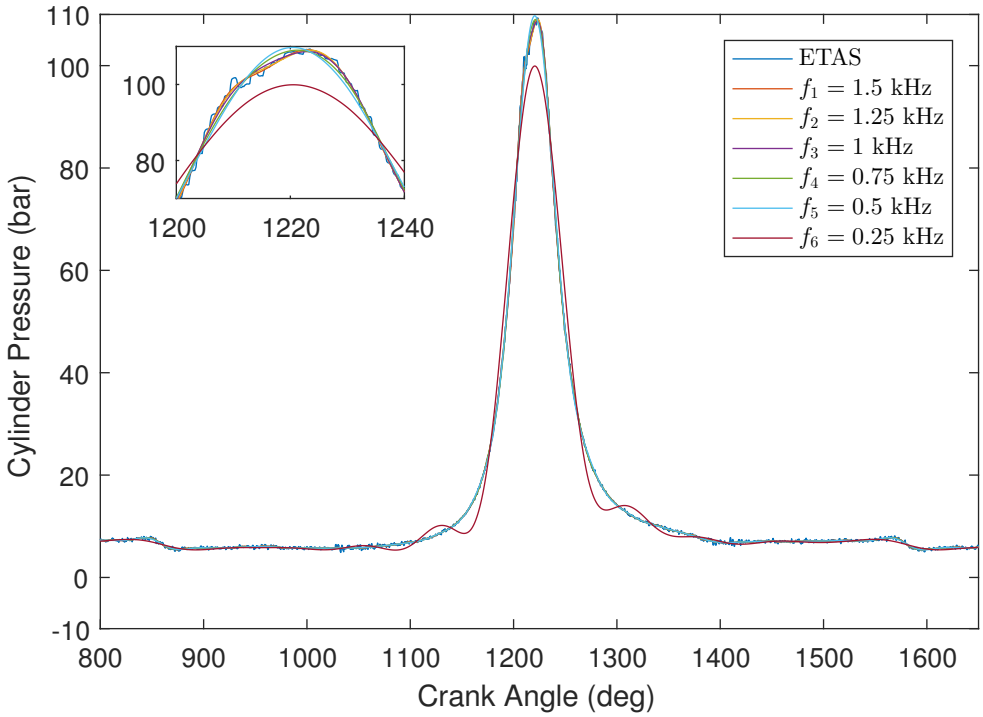

Figure 3.3: Butterworth low-pass filter for the cylinder No. 1 pressure signal under the stationary operating condition 3000 rpm, 120 Nm. ($f_1, f_2, \ldots, f_6$ denote six different cutoff frequencies.)

3.3.2 Cylinder Pressure Calibration

The second part of signal processing is to calibrate the cylinder pressure signal after downsampling. The reason why the calibration is necessary is that rapid temperature changes in both piezoelectric transducer housing and the quartz sensing element could change the direct current (DC) voltage of the transducer output. Based on the engine parameters given in Table 3.1, the least-squares method studied in [78] was used for the cylinder pressure calibration. One point should be noted that the pressure samples under one stationary operating condition used for the calibration method must be collected between intake valve closing and the start of injection [78]. After calibration, the sensor DC offset can be improved while the original sensor gain was still used. As an example, the calibration result for the stationary operating condition 3000 rpm, 120 Nm is displayed in Figure 3.4.

3.3.3 Engine Speed Calculation

The final part in signal processing is about the instantaneous engine speed calculation. Because the collected flywheel angular speed signal is a square pulse signal, the pulse signal should be transformed into the instantaneous angular speed (IAS) in rad/s. Accurately speaking, IAS denotes the instantaneous engine crank angular speed.

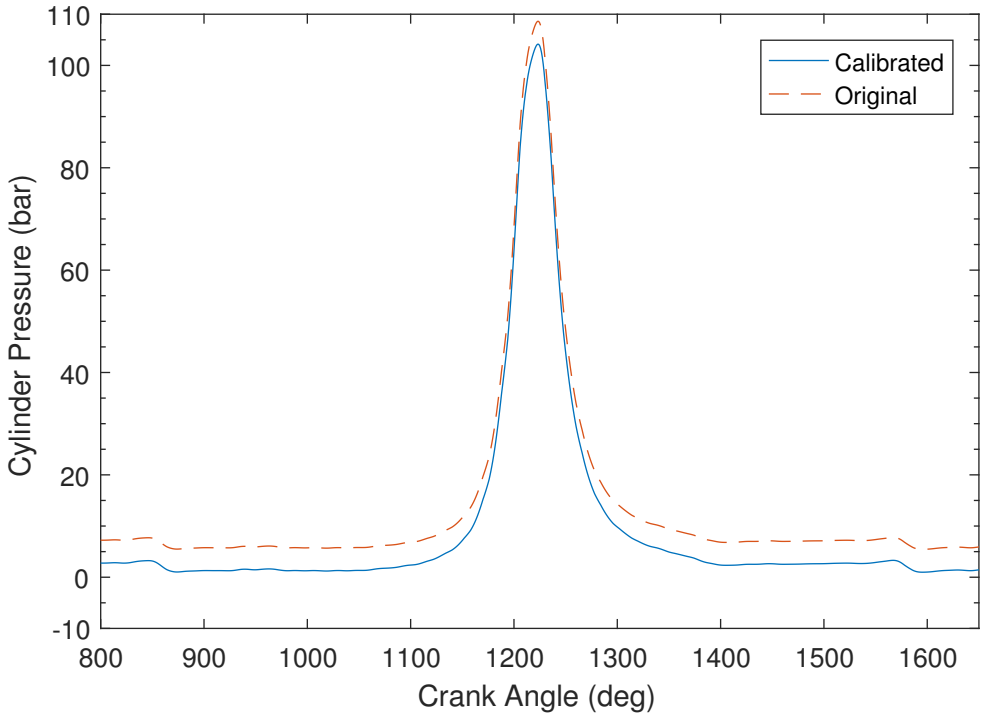

Figure 3.4: Cylinder pressure calibration.

The principle the calculation for the IAS $\tilde{\omega}$ of the engine, which is shown in Figure 3.5. Specifically, use ith $\Delta\theta^{\mathrm{p}}$ (i.e., $\Delta\theta_i^{\mathrm{p}}$) as an example, by using the flywheel angular position signal from the Hall-effect sensor, then the IAS signal $\tilde{\omega}(\theta_i^{\mathrm{p}})$ can approximately be computed as

$$\tilde{\omega}(\theta_i^{\mathrm{p}}) \approx \frac{\Delta\theta_i^{\mathrm{p}}}{\tilde{\tau}(\theta_i^{\mathrm{p}})}, \tag{3.6}$$

where θ_i^{p} denotes the angle obtained by the flywheel angular position sensor, $\Delta\theta_i^{\mathrm{p}}$ denotes the angle between two consecutive falling edges, and $\Delta\theta_i^{\mathrm{p}}$ is 6° (and 12° for the reference marker), and $\tilde{\tau}(\theta_i^{\mathrm{p}})$ denotes the value to approximate the elapsed time $\tau(\theta_i^{\mathrm{p}})$ between two consecutive falling edges. Additionally, it can be referred to [79] for more details about several factors affecting the precision of the calculation of IAS.

The relationship used in this paper between $\tilde{\omega}(k)$ and $\tilde{\omega}(\theta_i^{\mathrm{p}})$ is

$$\tilde{\omega}(k) = \tilde{\omega}(\theta_i^{\mathrm{p}}) \quad \text{for} \quad k = \frac{\theta_{i-1}^{\mathrm{p}}}{T_{\mathrm{s}}^{\mathrm{A}}}, \ldots, \frac{\theta_i^{\mathrm{p}}}{T_{\mathrm{s}}^{\mathrm{A}}}, \tag{3.7}$$

where θ_i^{p} an integer multiple of $T_{\mathrm{s}}^{\mathrm{A}}$.

In this thesis, the instantaneous engine cycle frequency $f(k)$ in Hz (here one engine cycle consists of four engine strokes, i.e., 720° per cycle), which is half of the instantaneous engine angular speed frequency, is needed. In order to simulate the real-time situations, the frequency $f(k)$ can approximately be calculated as

$$f(k) \approx \frac{\tilde{\omega}_{\mathrm{d}}(k)}{4\pi}, \tag{3.8}$$

where $\tilde{\omega}_{\mathrm{d}}(k)$ denotes the delayed version of the IAS $\tilde{\omega}(k)$. The sampling frequency for

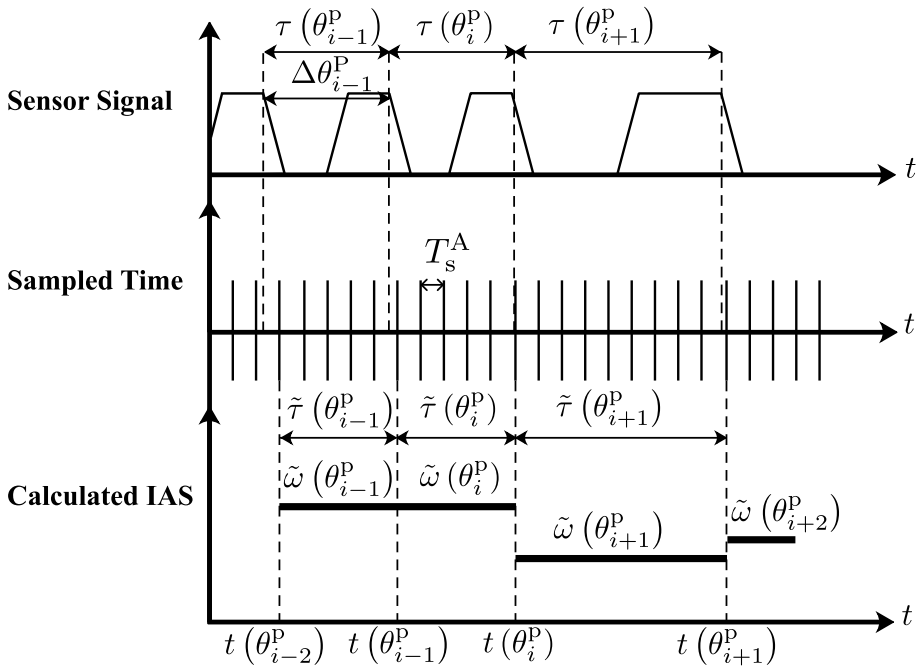

Figure 3.5: IAS calculation (t: time; θ: crank angle; i: integral number; The wider pulse of the sensor signal: the reference marker of the engine flywheel; The sampling period $T_{\mathrm{s}}^{\mathrm{A}}$: 100 kHz).

both $f(k)$ and $\tilde{\omega}_{\mathrm{d}}(k)$ is 20 kHz.

The speed $\tilde{\omega}_{\mathrm{d}}(k)$ can be calculated as

$$\tilde{\omega}_{\mathrm{d}}(k) = \tilde{\omega}(\theta_i^{\mathrm{p}}) \quad \text{for} \quad k = t_{\mathrm{s}}(k), \ldots, t_{\mathrm{e}}(k), \tag{3.9}$$

where $t_{\mathrm{s}}(k) = \mathrm{find}(t(k) \geq t(\theta_{i-1}^{\mathrm{p}}), 1) + 1$, and $t_{\mathrm{e}}(k) = \mathrm{find}(t(k) \geq t(\theta_i^{\mathrm{p}}), 1)$. " find " denotes the function in Matlab. The sampling frequency for $t(k)$ is 20 kHz.

It should be noted that in the calculation process of IAS, there are two sampling frequencies, i.e., 100 kHz (corresponding to the sampling period $T_{\mathrm{s}}^{\mathrm{A}}$) and 20 kHz (i.e., f_{s}), the frequency 100 kHz is used for increasing the precision of the calculation of the instantaneous engine speed, and the lower frequency 20 kHz is used for the purpose of implementability.

Remark 3.1. In Section 3.3.1, the low-pass filtering process can be regarded as prefiltering process in system identification [37] which will be discussed in Chapter 4. □

3.4 Summary

In this chapter, signal processing for the collected data from the engine test bench is conducted. In the downsampling process, a 7th-order Butterworth low-pass filter with the cutoff frequency 1 kHz was used as the anti-filter. Afterwards, a process for calibrating the cylinder pressure signal is described, and the difference between the calibrated cylinder pressure signal and the cylinder pressure signal without calibration is obviously shown in Figure 3.4. The instantaneous engine cycle frequency $f(k)$ in Hz is also calculated, and in

this thesis, the method of calculating the frequency $f(k)$ is used in several cylinder pressure reconstruction methods where the instantaneous engine speed is needed. All the processed data discussed in Section 3.3 is ready for the purposes of system identification (in both time domain and frequency domain) and validating the cylinder pressure reconstruction in the sequent chapters.

Chapter 4

Identification of the Model Between Cylinder Pressure and Engine Structural Vibration

4.1 Introduction

In this chapter, the objective is to identify an open-loop black-box model between four inputs (cylinder pressure signal) and one output (engine structural vibration signal) based on using the system identification methods introduced in Chapter 2. Because of the complex relationship between the cylinder pressure and the vibration, besides linear models, nonlinear models are identified either in this chapter. All the identified models are discrete-time models. Below, the system-identification cycle in practice is first introduced in Section 4.1. In Section 4.3, both time-domain identified model and frequency-domain identified model are derived. Then in Section 4.4, a nonlinear relationship between four cylinder pressure signals and one vibration signal is built. In the end of this chapter, a summary is given.

4.2 The System-Identification Cycle

In this section, a summary for the use of the identification methods introduced in Chapter 2 is made. Specifically, Figure 4.1 is first used to illustrate the system-identification cycle and then some related descriptions are made for Figure 4.1.

As shown in Figure 4.1, there are several key elements in the system-identification cycle:

- *Experiment*
  The identification of the model between four cylinder pressure signals and one vibration signal needs the collected cylinder pressure signal and vibration signal from the test bench. In order to make sure that the identified model can be used in practice, when do experiments the IO data should be collected for identification in typical engine operating ranges. Additionally, to be able to estimate a model (linear or nonlinear) from the IO data, the data should contain enough information, i.e., the condition of persistent excitation should be satisfied [22, 37]. As an example of the persistent excitation in time-domain subspace identification, at least the collected input sequence should satisfy the persistent excitation condition such that

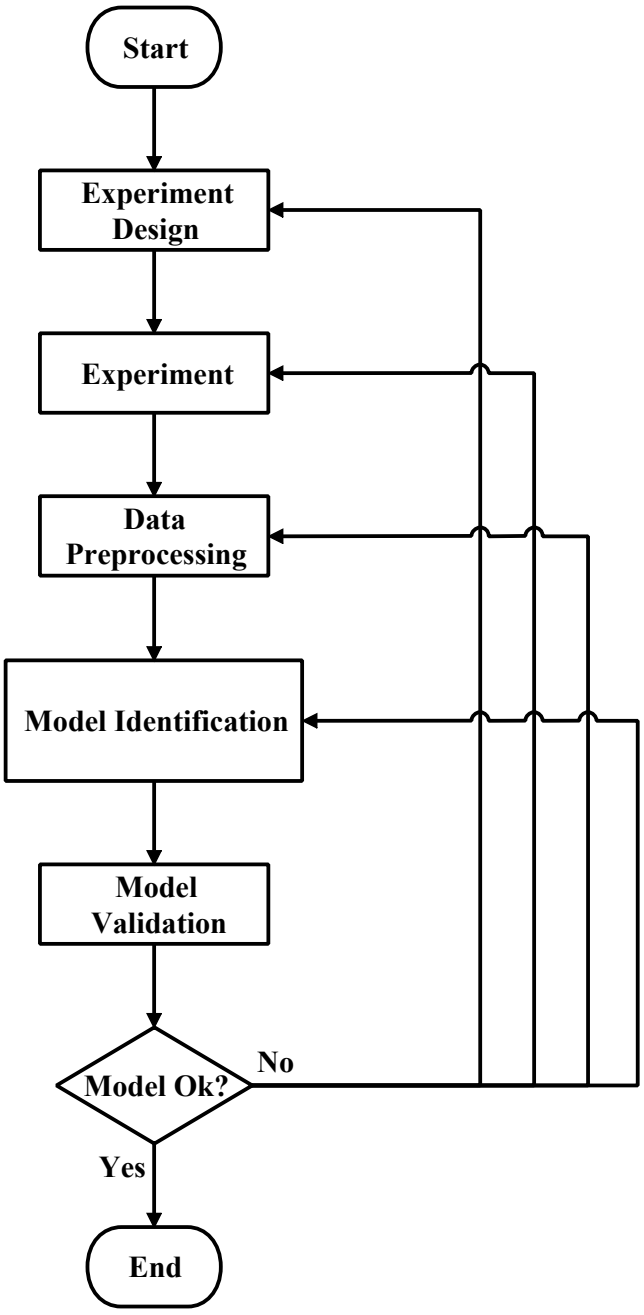

Figure 4.1: A schematic view of the key elements in the system-identification cycle.

the matrix $\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}$ is nonsingular in Equation (2.71), but it should be noted that this persistency of excitation of the input signal may be not sufficient. The reader can refer to see more about persistent excitation in subspace identification in the book [22]. Moreover, it is deserved to say that different identification methods may need different persistent excitation conditions.

- *Data preprocessing*
  In Chapter 3, signal processing for the collected IO data is illustrated in Figure 3.2, and both downsampling and low-pass filtering are involved. For linear system identification, the functions and effects of the prefiltering is also illustrated in Remark 4.1, while for nonlinear system identification, the functions and effects of the prefiltering are more complex, and they are not considered in this thesis.

- *Model identification*
  In Section 2.5 and Section 2.6, time-domain system identification (Algorithm 2.1) and frequency-domain system identification (Algorithm 2.2) are introduced. In Section 2.7, nonlinear system identification is described (Algorithm 2.3). For all the identification methods introduced in the above, the parameters related to the noise are regarded as tunable parameters in the cylinder pressure reconstruction methods in the thesis, and the tuning is specifically illustrated in the proposed reconstruction methods.

- *Model validation*
  As shown in Table 3.2, two datasets are used for identifying the model between four cylinder pressure signals and one vibration signal. One dataset is used for identification and self model validation, and the other dataset is used for cross model validation. Both self model validation and cross model validation use the following fitness criterion to measure the quality of identified models:

$$\text{fitness} = 100\left(1 - \frac{\parallel \boldsymbol{Y}(k) - \hat{\boldsymbol{Y}}(k) \parallel_2}{\parallel \boldsymbol{Y}(k) - \text{average}\left(\boldsymbol{Y}(k)\right)\tilde{\boldsymbol{I}}_N \parallel_2}\right), \tag{4.1}$$

  where $\boldsymbol{Y}(k) = \begin{pmatrix}\boldsymbol{y}^{\mathrm{T}}(0) & \boldsymbol{y}^{\mathrm{T}}(1) & \cdots & \boldsymbol{y}^{\mathrm{T}}(N-1)\end{pmatrix}^{\mathrm{T}}$, and $\hat{\boldsymbol{Y}}(k)$ represents vector of which the elements are the simulated output of the identified model, and average $(\cdot)$ represent the average value of all the elements in a vector, $\tilde{\boldsymbol{I}}_N$ denotes the unit vector with N rows, and the fraction denotes the normalized root-mean-square error (NRMSE).

In Figure 4.1, several feedback loops which can be used to improve the identified results can be found. For example, when the identified model is not acceptable, one possible reason is that the model identification method is not chosen suitably, so in this case, the model identification method should be improved or changed.

4.3 Results of Linear System Identification

Based on the system-identification cycle illustrated in Section 4.2, the identification of the model between the cylinder pressure and the vibration is conducted in this section, and the corresponding results are obtained either.

Suppose there exists a relationship between four cylinder pressure signals and one vibration signal, and it can be represented as

$$y(k) = \boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{u}(k) + H(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})e(k), \tag{4.2}$$

where $\boldsymbol{u}(k) = \begin{pmatrix} P_1(k) & P_3(k) & P_4(k) & P_2(k) \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^4$, $P_i(k)$ denotes the cylinder No. i pressure signal after signal processing in Chapter 3. $y(k) \in \mathbb{R}$ denotes the engine vibration signal after signal processing. $\{e(k)\}$ denotes a scalar white noise process, of which the covariance function is $\sigma^2\delta_{kj}$, i.e., $\mathrm{E}\left[e(k)e^{\mathrm{T}}(j)\right] = \sigma^2\delta_{kj}$, and the value of σ^2 is positive, constant and unknown. The parameter vector $\boldsymbol{\theta}_{\mathrm{p}} \in \boldsymbol{\Omega}_{\mathrm{p}}$ is formed by stacking all the scalar parameters to be identified in $\boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})$ and $H(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})$, and $\boldsymbol{\Omega}_{\mathrm{p}}$ denotes the specified parameter set which constraints the parameter vector [22], e.g., a stability constraint. $\boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})$ and $H(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})$ denote the transfer functions (or operators) of two LTI models (one deterministic sub-model $\boldsymbol{G}$ and one stochastic sub-model H, respectively). $\boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})$ and $H^{-1}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})$ are rational, asymptotically stable, and causal. The product $H(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})e(k)$ represent a result induced by the disturbance and measurement noise. The input vector signal $\boldsymbol{u}(k)$ is assumed to be noiseless.

Remark 4.1. As mentioned in Remark 3.1 in Chapter 3, the input-output (IO) data and the noise term in the system (4.2) are prefiltered by using the same Butterworth low-pass filter with 7th order (which is denoted as $L_{\mathrm{f}}(q^{-1})$ here), and prefiltering does not change the IO relation for the linear system (4.2), which can be illustrated as follows:

$$L_{\mathrm{f}}(q^{-1})y_{\mathrm{o}}(k) = \boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})\boldsymbol{L}_{\mathrm{f}}(q^{-1})\boldsymbol{u}_{\mathrm{o}}(k) + L_{\mathrm{f}}(q^{-1})H(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})e_{\mathrm{o}}(k), \tag{4.3}$$

where $y_{\mathrm{o}}(k)$ and $\boldsymbol{u}_{\mathrm{o}}(k)$ denote the original signals with the sampling frequency 100 kHz collected from ETAS. $\{e_{\mathrm{o}}(k)\}$ represents the assumed white noise process in the original system without low-pass filtering. The output $L_{\mathrm{f}}(q^{-1})y_{\mathrm{o}}(k) = y(k)$, the input $\boldsymbol{L}_{\mathrm{f}}(q^{-1})\boldsymbol{u}_{\mathrm{o}}(k) = \boldsymbol{u}(k)$, and the noise part $L_{\mathrm{f}}(q^{-1})H(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})e_{\mathrm{o}}(k) = H(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})e(k)$. $\boldsymbol{L}_{\mathrm{f}}(q^{-1})$ is defined as $L_{\mathrm{f}}(q^{-1})\boldsymbol{I}_4$.

The prefilter $L_{\mathrm{f}}(q^{-1})$, which is a low-pass filter, can affect the bias distribution of the resulting model and can also be to remove disturbances of high frequencies not of interest in the modeling [22, 37]. As a conclusion, the prefilter $L_{\mathrm{f}}(q^{-1})$ is good for the modeling of the system (4.2) where low frequencies (smaller than 1 kHz) are of interest. □

Without considering the stochastic sub-model, in this section linear system identification techniques introduced in Chapter 2 are implemented to identify the deterministic sub-model of the model (4.2), and the objective of system identification is to parameterize the transfer functions $\boldsymbol{G}(q^{-1}, \boldsymbol{\theta}_{\mathrm{p}})$, which can be denoted by the mapping $\mathcal{M}_1 : \boldsymbol{\Omega}_{\mathrm{p}} \mapsto \mathcal{R}_{n_{\mathrm{m}}}^{1\times 4}$, where $\mathcal{R}_{n_{\mathrm{m}}}^{1\times 4}$ denotes the set of all 1×4 (four-input one-output) asymptotically stable and causal rational transfer functions with n_{m}th order. According to the properties of the mapping such as surjective, injective, and bijective, it can be determined whether a unique model can be identified or not. However, in practice, often the main objective is to find a model that describes IO data, and uniqueness is not needed [22]. The above problem of uniqueness refers to the concept identifiability which is introduced in [37, 40, 70].

4.3.1 Results of System Identification in Time Domain

Based on two datasets (dataset A and dataset B) introduced in Chapter 3, Algorithm 2.1 was used for the identification of the model $\boldsymbol{G}$. In the implementation process of Algorithm 2.1, the stability of the identified models was forced by reflecting unstable poles into the unit disk [43], and the feedthrough term $\boldsymbol{D}$ in the model $\boldsymbol{G}$ was not involved, and the prediction horizons were determined by using the Akaike information criterion (AIC) [46]. The identification results are shown in Table 4.1. The different models correspond to subspace estimates with CVA and MOESP weightings, with prediction focus or simulation focus (CVA(s) and MOESP(s) for simulation focus). The measure (4.1) is used in the table.

Table 4.1: Identification results of the model $\boldsymbol{G}$ using Algorithm 2.1.

Model order	CVA	MOESP	CVA(s)	MOESP(s)
4	49.10 (33.82)	23.58 (12.75)	66.80 (62.92)	53.14 (41.34)
5	59.96 (48.80)	40.17 (37.36)	61.20 (62.71)	56.33 (55.96)
6	61.92 (46.81)	37.17 (18.70)	65.56 (68.15)	59.53 (44.58)
7	61.52 (51.72)	52.01 (52.38)	66.10 (56.18)	72.31 (63.55)
8	69.94 (60.27)	62.39 (47.89)	71.17 (66.09)	70.92 (60.39)
9	67.49 (59.62)	66.72 (60.38)	73.42 (63.25)	73.69 (65.52)
10	69.22 (62.74)	64.43 (25.18)	70.39 (67.02)	71.30 (66.12)

Note: The fitness in bracket denotes the cross-validation fitness while the fitness without bracket denotes the self-validation fitness.

By observing the identification results in Table 4.1 the model with 9th order and MOESP(s) was finally chosen as the model $\boldsymbol{G}$ used for the cylinder pressure reconstruction. The reasons why the above model was chosen are as follows:

(i) In this thesis, the derivation and formulation of the methods of cylinder pressure reconstruction are emphasized, while the best model is not necessary.

(ii) In practice the best model among the identified models is difficult to find.

In the sequent sections of this chapter, the model selection criterion will be also based on the above reasons. Besides, according to the identification results, it can also be found that in general the identification results with simulation focus are better than the results with prediction focus.

The Bode plot, pole-zero map, and impulse response of the chosen model are illustrated in Figure 4.2, Figure 4.3, and Figure 4.4, respectively. In these figures, G_i, $i = 1, 2, \ldots, 4$, denote four sub-models of the model $\boldsymbol{G}$, i.e., the transfer function $\boldsymbol{G}(z) = \begin{pmatrix} G_1(z) & G_2(z) & G_3(z) & G_4(z) \end{pmatrix}$.

As seen in Figure 4.2, the Bode magnitude curves of the four sub-models are similar while the Bode phase curves of the four sub-models are totally different with each other. Figure 4.3 shows that all the sub-models are stable, and simultaneously shows the poles of the sub-models have similar coordinates. Figure 4.4 illustrates the decays of the impulse responses of the sub-models.

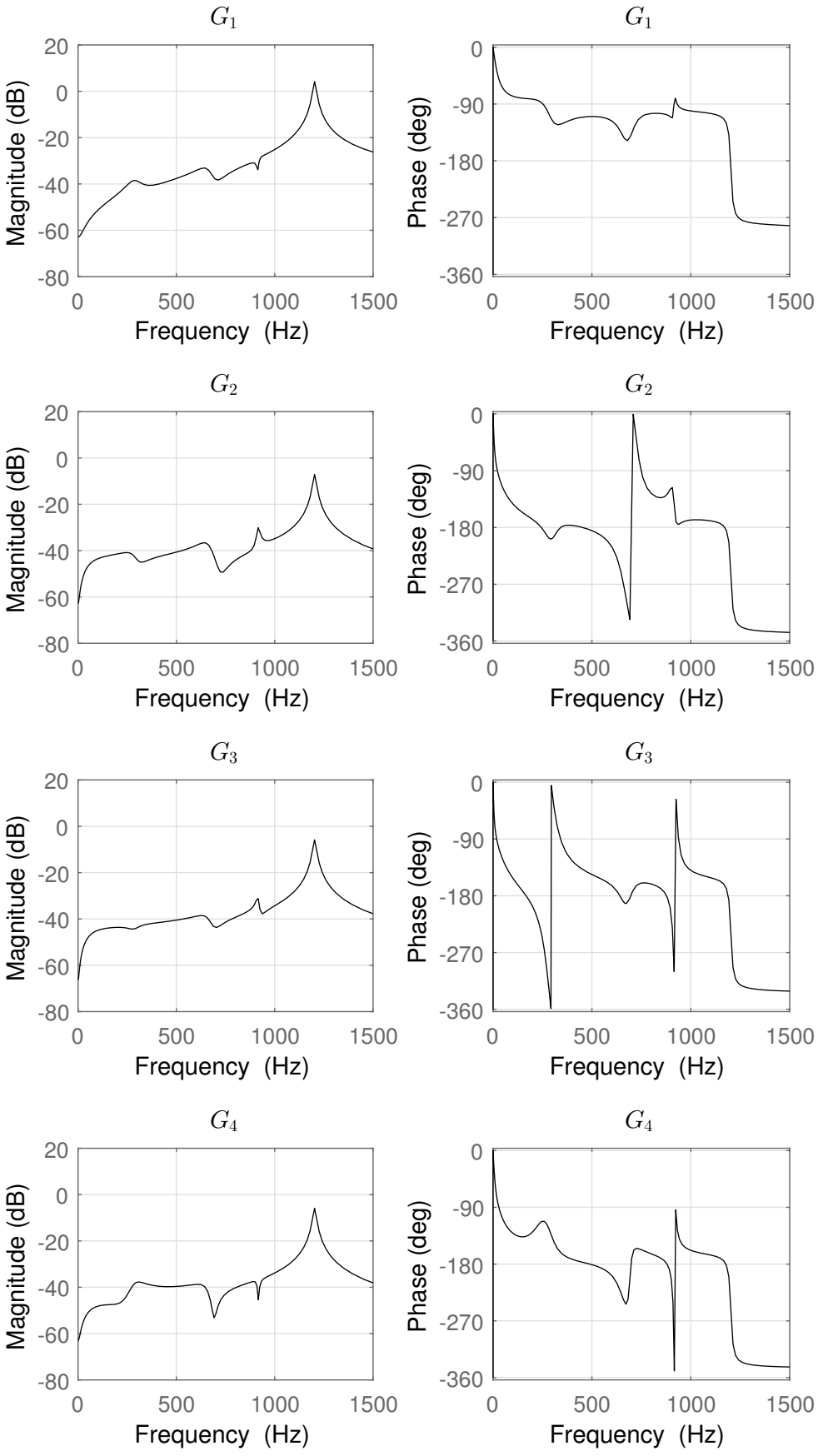

Figure 4.2: Bode plot with phase wrapping.

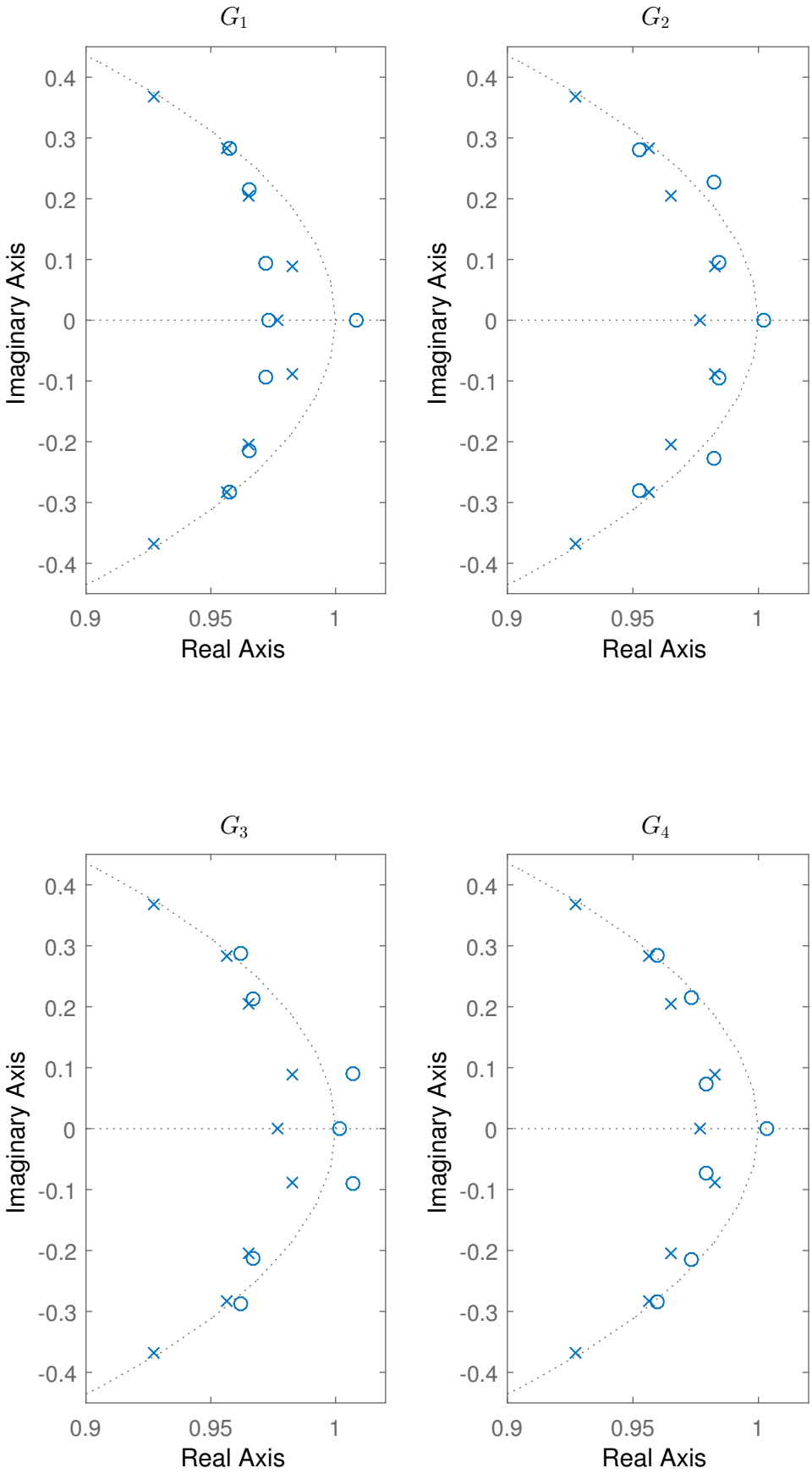

Figure 4.3: Pole-zero map.

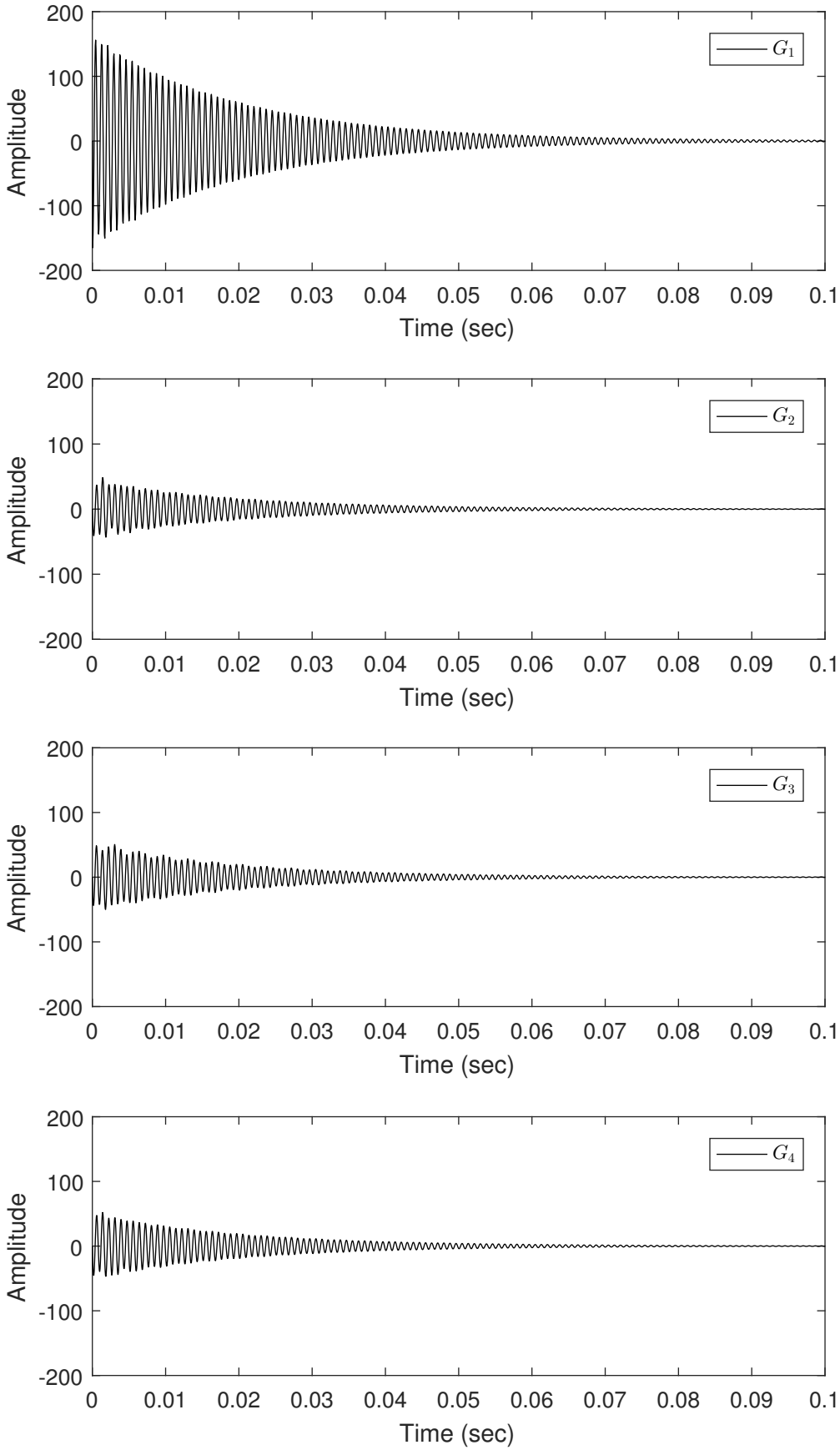

Figure 4.4: Impulse response.

4.3.2 Results of System Identification in Frequency Domain

In the first step of frequency-domain system identification, LPM was used to estimate the FRF points of the model $\boldsymbol{G}$ at the specified frequencies. As introduced in Chapter 3, the stationary operating conditions Z_2 (1200 rpm, 120 Nm), Z_5 (2100 rpm, 120 Nm), and Z_8 (3000 rpm, 120 Nm) are used for model identification. But prior to system identification, the frequencies of interest need to be specified. Figure 4.5, Figure 4.6, and Figure 4.7 (three semilog plots) display the FFT results of the cylinder No. 1 pressure signal and the vibration signal under the above three stationary operating conditions. These figures indicate that under stationary operating conditions both cylinder pressure signal and vibration signal can approximately be seen as a periodic signal. Therefore the frequencies for identification can be specified by finding the corresponding peaks in the spectrums. Based on the spectrums, finally 40 consecutive peaks (i.e., the red circles in the figures) were selected for each stationary operating condition. It should also be noted that the selected frequencies in the spectrums of the cylinder No. 1 pressure signal are the same as the ones of the other three pressure signals. The specific peak number selection procedure is illustrated in Algorithm 4.1.

Algorithm 4.1: Peak number selection procedure.

1 Choose the cutoff frequency f_c in Hz for the collected signals from ETAS (see Section 3.3.1.3 and Figure 3.3).

2 Calculate the selected peak number n_p as the greatest integer less than or equal to $\frac{f_c}{f_b}$, where f_b denotes the basic frequency of the cylinder pressure under the engine speed 3000 rpm (i.e., $f_b = 25$ Hz).

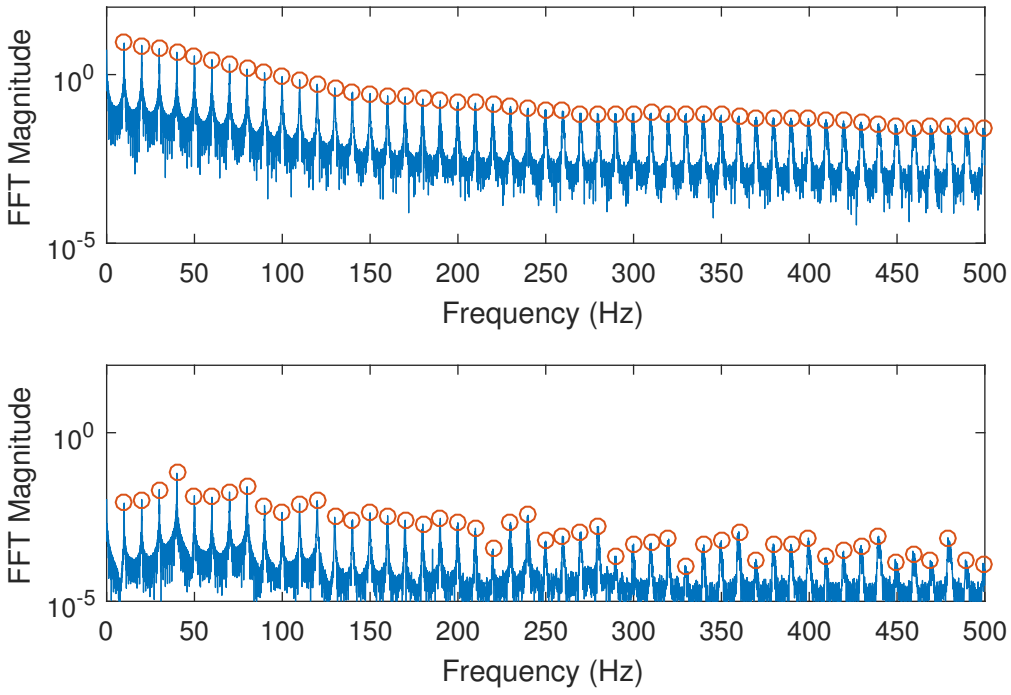

Figure 4.5: FFT magnitudes of the cylinder No. 1 pressure signal (upper figure) and vibration signal (lower figure) under the stationary operating condition 1200 rpm, 120 Nm.

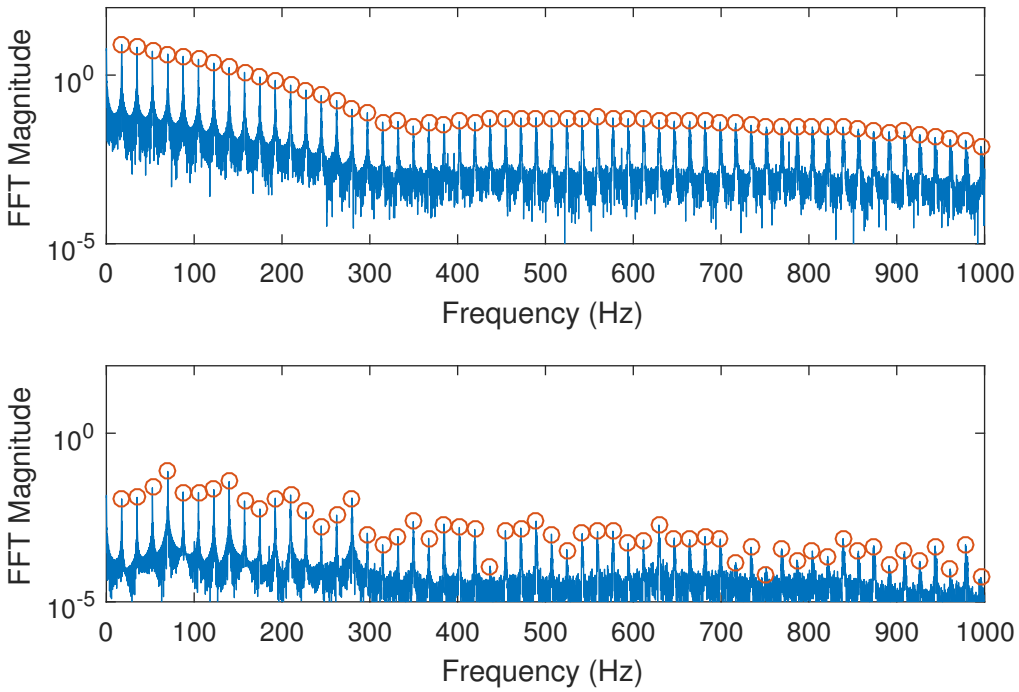

Figure 4.6: FFT magnitudes of the cylinder No. 1 pressure signal (upper figure) and vibration signal (lower figure) under the stationary operating condition 2100 rpm, 120 Nm.

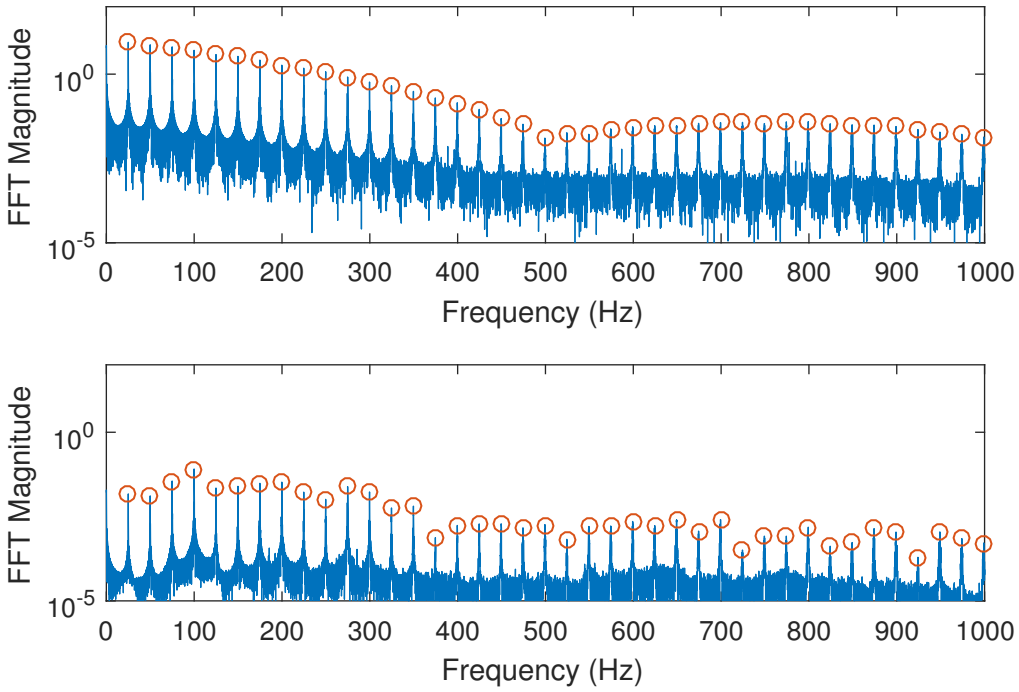

Figure 4.7: FFT magnitudes of the cylinder No. 1 pressure signal (upper figure) and vibration signal (lower figure) under the stationary operating condition 3000 rpm, 120 Nm.

With the total specified 120 frequencies and corresponding FFT magnitudes derived from Z_1, Z_2, and Z_3, Algorithm 2.2 was implemented to identify the model $\boldsymbol{G}$. In the first step of Algorithm 2.2, η was set to 2 in Equation (2.98), and in Equation (2.103) q was set to 15. The subspace algorithm was looped over different model orders (n_{m} was chosen from 4 to 10) and different block rows (r was chosen from $n_{\mathrm{m}}+1$ to 30) of the extended observability matrix. The stability of the identified models was forced by reflecting unstable poles into the unit disk [43]. The LM algorithm was carried out to optimize the model parameters obtained with the subspace algorithm.

The final identification results of frequency-domain system identification are displayed in Table 4.2. The measure (4.1) is used in the table. It can be seen that in the table, except the model with 8th order, other models are identified as well as validated based on detrended IO data. The reason why the detrended IO data is used is that in the frequency-domain system identification process, the frequency components corresponding to the DC offsets of both cylinder pressure and vibration were not used. However, for the cylinder pressure reconstruction, if the model identified based on the detrended IO data is used, the input (i.e., the vibration signal) of the cylinder pressure reconstructor should be first detrended, so it is not convenient to reconstruct the cylinder pressure. As mentioned above, the model with 8th order was also validated based on the IO data without detrending, and the corresponding identification results are illustrated in Table 4.2 (the model order with rectangular box). While other models validated by using the IO data without detrending have bad results which are not illustrated in the table.

Table 4.2: Identification results of the model $\boldsymbol{G}$ using Algorithm 2.2.

Model order	Self validation			Cross validation
4	$13.96(Z_2^{\mathrm{d}})$	$33.22(Z_5^{\mathrm{d}})$	$40.90(Z_8^{\mathrm{d}})$	$25.35(Z_{\mathrm{t}}^{\mathrm{d}})$
5	$-4.57(Z_2^{\mathrm{d}})$	$23.19(Z_5^{\mathrm{d}})$	$23.47(Z_8^{\mathrm{d}})$	$14.59(Z_{\mathrm{t}}^{\mathrm{d}})$
6	$-4.22(Z_2^{\mathrm{d}})$	$11.16(Z_5^{\mathrm{d}})$	$16.76(Z_8^{\mathrm{d}})$	$6.17(Z_{\mathrm{t}}^{\mathrm{d}})$
7	$12.18(Z_2^{\mathrm{d}})$	$30.50(Z_5^{\mathrm{d}})$	$30.30(Z_8^{\mathrm{d}})$	$24.99(Z_{\mathrm{t}}^{\mathrm{d}})$
8	$53.42(Z_2^{\mathrm{d}})$	$61.46(Z_5^{\mathrm{d}})$	$63.53(Z_8^{\mathrm{d}})$	$57.35(Z_{\mathrm{t}}^{\mathrm{d}})$
8	$52.96(Z_2)$	$60.78(Z_5)$	$62.21(Z_8)$	$57.05(Z_{\mathrm{t}})$
9	$17.44(Z_2^{\mathrm{d}})$	$36.13(Z_5^{\mathrm{d}})$	$38.44(Z_8^{\mathrm{d}})$	$32.07(Z_{\mathrm{t}}^{\mathrm{d}})$
10	$-0.89(Z_2^{\mathrm{d}})$	$11.89(Z_5^{\mathrm{d}})$	$17.84(Z_8^{\mathrm{d}})$	$7.65(Z_{\mathrm{t}}^{\mathrm{d}})$

Note: The operating conditions in the bracket are used for model validation. Z_2^{d} denotes the operating condition Z_2 with detrended cylinder pressure and vibration signal.

By observing the identification results in Table 4.2, and based on the above discussions, the model with 8th order was finally chosen as the model $\boldsymbol{G}$ which is used for the cylinder pressure reconstruction. Additionally, the model with 8th order can be directly used for cylinder pressure reconstruction without detrending.

The Bode plot, pole-zero map, and impulse response of the chosen model are illustrated in Figure 4.8, Figure 4.9, and Figure 4.10, respectively. As displayed in Figure 4.8, both Bode magnitude curves and Bode phase curves of the four sub-models are obviously different with each other. The stability is shown in Figure 4.9, and the decays of the impulse responses of the sub-models are also depicted in Figure 4.10.

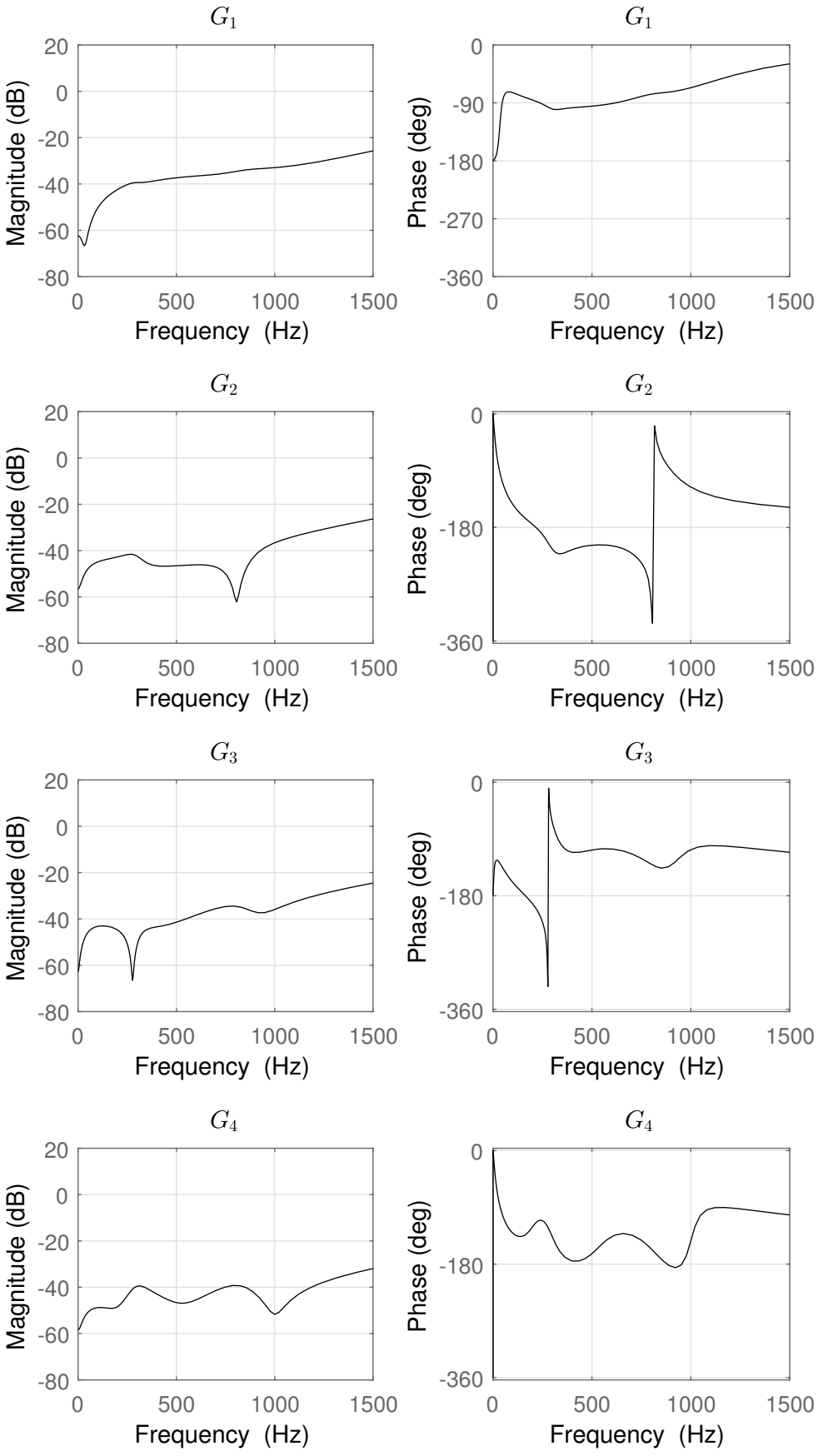

Figure 4.8: Bode plot with phase wrapping.

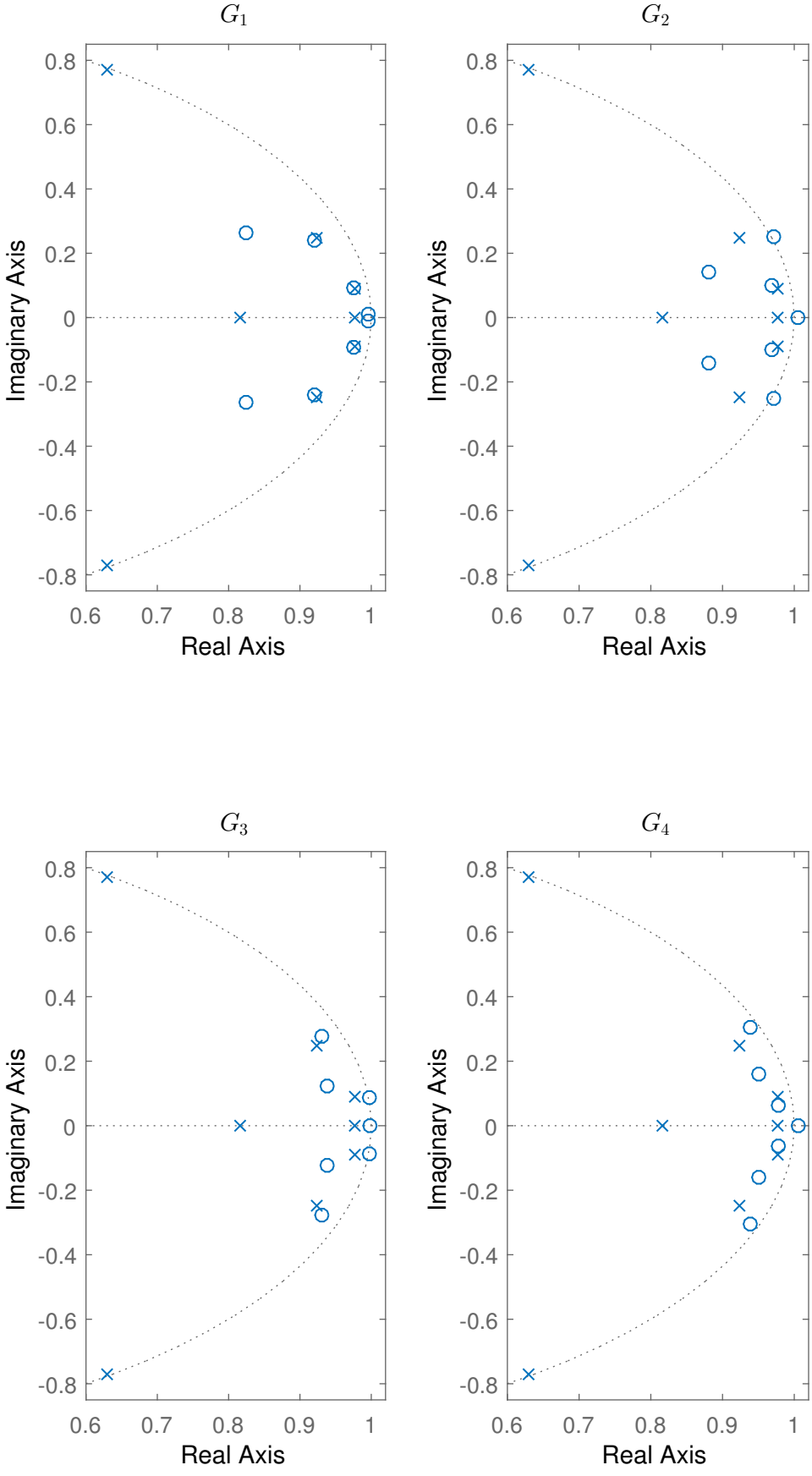

Figure 4.9: Pole-zero map.

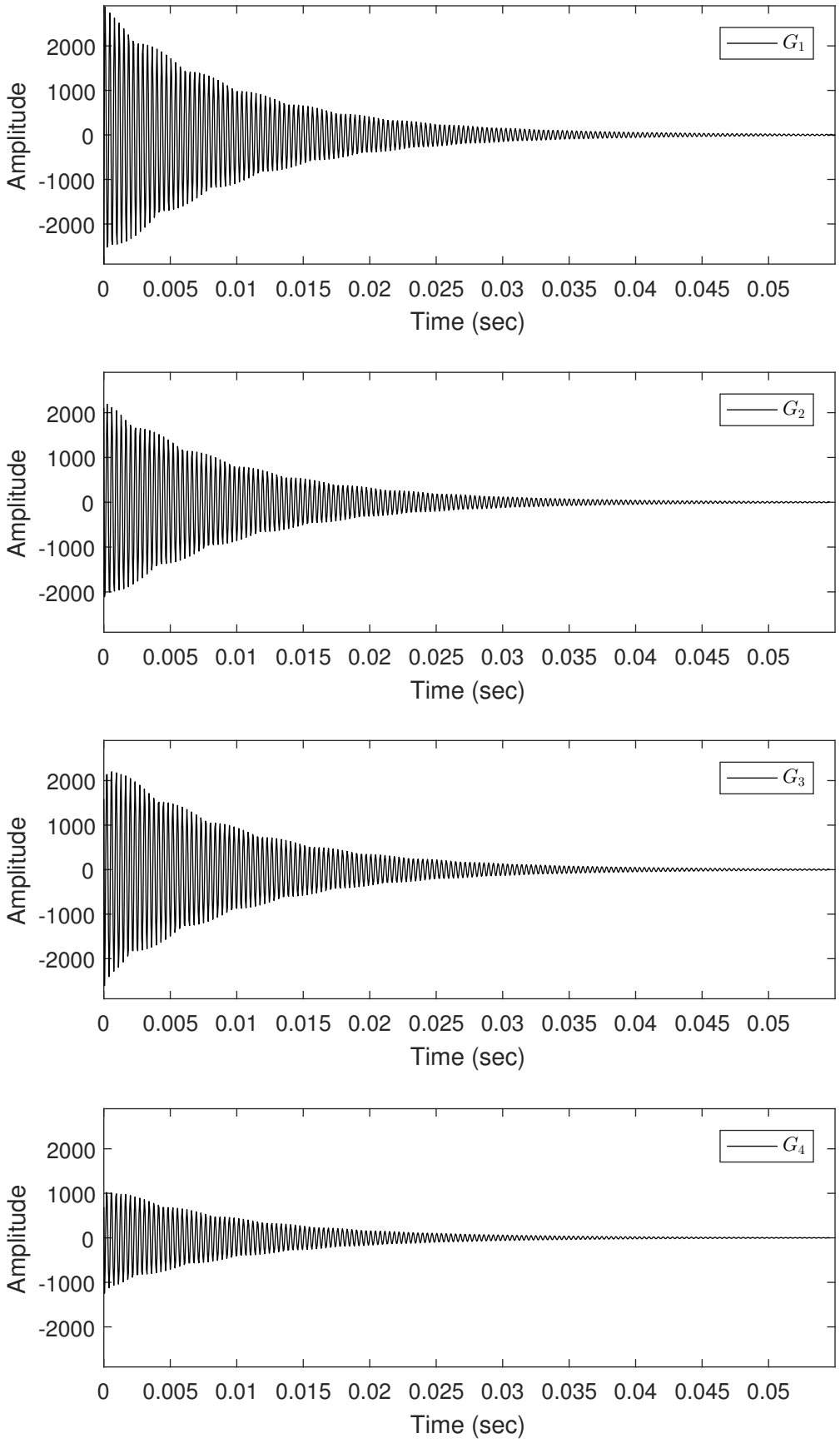

Figure 4.10: Impulse response.

4.4 Results of Nonlinear System Identification

Suppose there exists a nonlinear model between four cylinder pressure signals and one vibration signal, as displayed in Figure 2.1, the nonlinear model (denoted as $\boldsymbol{G}_{\text{non}}$) contains both nonlinear deterministic model and output error. Specifically, the nonlinear model $\boldsymbol{G}_{\text{non}}$ considered in this section is denoted as

$$\begin{cases} \boldsymbol{u}_{\text{o}}(k) = \boldsymbol{f}_{\text{L}}(k) * \boldsymbol{u}(k), \\ y(k) = f_{\text{N}}(\boldsymbol{u}_{\text{o}}(k)) + e_{\text{non}}(k), \end{cases} \tag{4.4}$$

where the symbol " $*$ " denotes the convolution, $\boldsymbol{f}_{\text{L}}(k)$ represents the impulse response function of the linear block denoted as $\boldsymbol{F}_{\text{L}}$, and $f_{\text{N}}(\cdot)$ denotes the output nonlinearity. The linear block and the output nonlinearity are concatenated to form the nonlinear deterministic model which is denoted as $\boldsymbol{G}_{\text{non}}^{\text{d}}$. As aforementioned, $\boldsymbol{u}(k) \in \mathbb{R}^4$ denotes a vector consist of four cylinder pressure signals, and $y(k) \in \mathbb{R}$ denotes the engine vibration signal. The term $e_{\text{non}}(k)$ denotes the output error.

Based on dataset B introduced in Chapter 3, Algorithm 2.3 was used for identifying the nonlinear model $\boldsymbol{G}_{\text{non}}^{\text{d}}$. Dataset A was used for model validation. The degree of the polynomial function $f_{\text{N}}(\cdot)$ was set to 2 to capture complexity. The identification results are illustrated in Table 4.3. The specified poles and zeros number are also involved. The measure (4.1) is used.

Table 4.3: Identification results of the model $\boldsymbol{G}_{\text{non}}^{\text{d}}$ using Algorithm 2.3.

Poles number	Zeros number	Self validation	Cross validation
4	2	3.42	2.13
5	3	37.52	33.43
6	4	27.64	25.75
7	5	63.70	56.55
8	6	75.53	57.16
9	7	76.70	70.38
10	8	78.10	64.24

Finally, by observing the identification results in Table 4.3 the model with 9th order was chosen as the model $\boldsymbol{G}_{\text{non}}^{\text{d}}$ used for the cylinder pressure reconstruction.

4.5 Summary

In this chapter, based on the identification algorithms introduced in Chapter 2, three kinds of models are finally identified, i.e., the time-domain identified model, frequency-domain identified model, and nonlinear model. All of them will be used in cylinder pressure reconstruction methods in the sequent chapters. It should be mentioned again that in this thesis the derivation and illustration of cylinder pressure reconstruction methods are the main points, so the best identified model is not pursued.

Chapter 5

Linear Model-Based Cylinder Pressure Estimation Using Both Engine Structural Vibration and Engine Speed Signal

5.1 Introduction

In this chapter, a discrete-time method for the linear model-based cylinder pressure estimation using both engine structural vibration and instantaneous engine speed signal is explored [80, 81]. The specific algorithm is illustrated in Section 5.2, followed by the cylinder pressure estimation results shown in Section 5.3 and Section 5.4, more specifically, in Section 5.3, the cylinder pressure estimation results are based on the time-domain identified model, while in Section 5.4, the estimation results are based on the frequency-domain identified model. In Section 5.5, based on a so-called ideal model, the proposed cylinder pressure estimation method is further validated. Finally, a summary is given.

5.2 Cylinder Pressure Estimation Algorithm

In this section, a method of cylinder pressure estimation by using both engine structural vibration and instantaneous engine speed signal is proposed. As illustrated in Figure 5.1, the total framework contains two parts, i.e., part A and part B.

Part A can be seen as an offline design procedure, which contains three components:

(i) Identify the model (4.2) between four (processed) cylinder pressure signals and one (processed) vibration signal. As aforementioned, the deterministic sub-model in the model (4.2) is denoted as $\boldsymbol{G}$, of which the transfer function is $\boldsymbol{G}(z) = \boldsymbol{C}\left(z\boldsymbol{I}_{n_\mathrm{m}} - \boldsymbol{A}\right)^{-1}\boldsymbol{B} + \boldsymbol{D}$ (see Equation (2.110) without showing $\boldsymbol{\theta}_\mathrm{p}$). $e_\mathrm{l}(k)$ denotes the scalar output error.

(ii) Regard other three cylinder pressure signals as the delay versions of the cylinder No. 1 pressure signal. The reason why the delay block model $\tilde{G}_\mathrm{d}^\mathrm{a}$ can be involved is that according to the mechanism of the engine the phase difference is 180° between every two consecutive engine strokes and the pressure signal shapes corresponding to different cylinders are nearly the same in every engine cycle, and the delay in angle

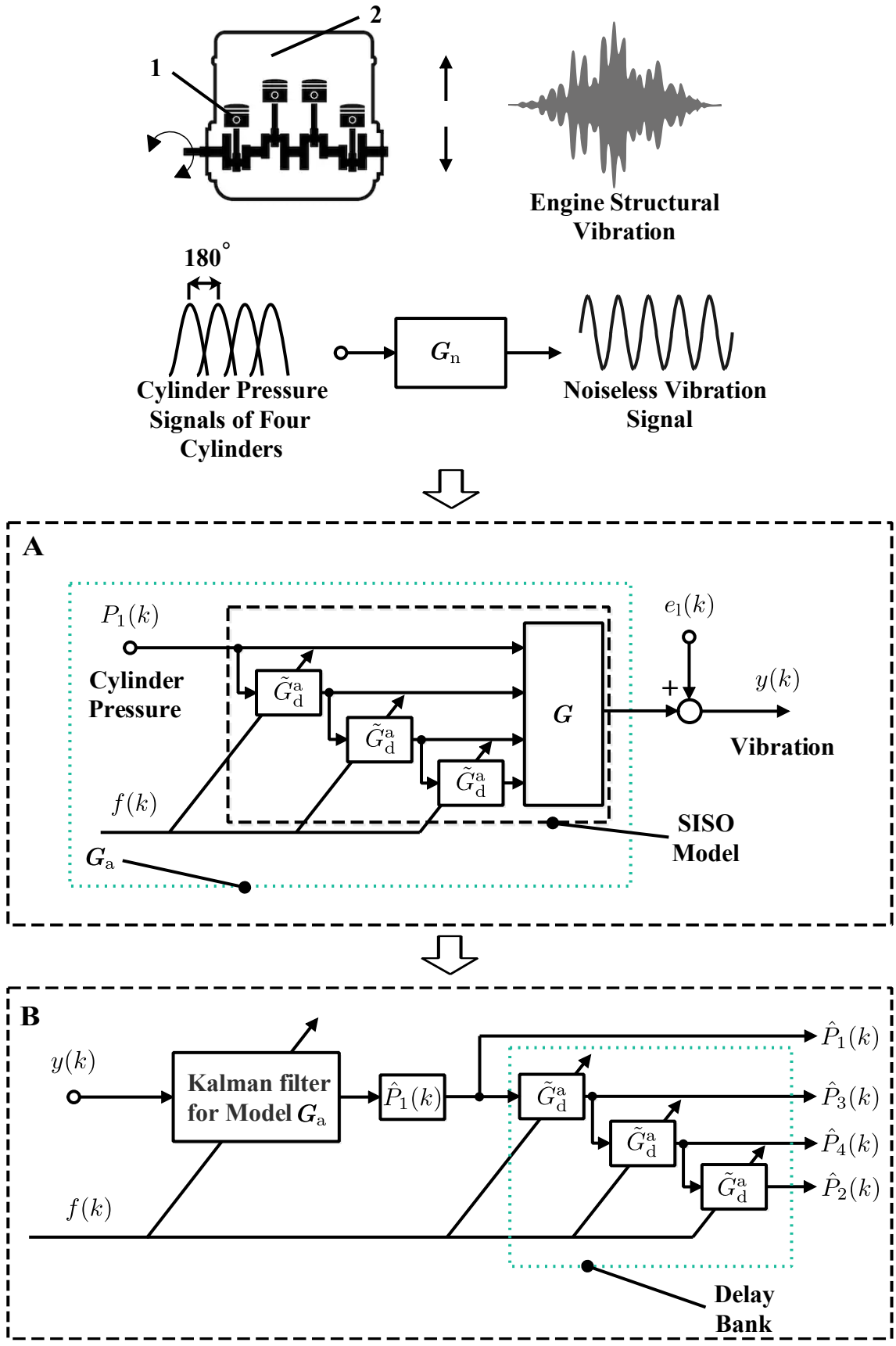

Figure 5.1: Cylinder pressure estimation using both engine vibration and speed signal (1: the cylinder No. 1; 2: the engine block of the engine test bench; $\boldsymbol{G}_n$: the model between four noiseless cylinder pressure signals and one noiseless vibration signal).

domain can be transformed to the delay in time domain. As a result, by combing three delay systems with the model $\boldsymbol{G}$, a SISO model can be obtained as shown in part A of Figure 5.1. Furthermore, the reason why a SISO model is required is that the observability the augmented model described below should be guaranteed.

(iii) Augment the cylinder pressure signal model with the model $\boldsymbol{G}$ and three delay systems (i.e., state augmentation), then the augmented model $\boldsymbol{G}_{\mathrm{a}}$ can be obtained. The cylinder pressure signal modeling can be referred to the order modeling as introduced in Section 2.4.

Part B can be regarded as an implementation procedure. With the augmented model $\boldsymbol{G}_{\mathrm{a}}$, a linear Kalman filter for it can be obtained, and then by using the vibration signal as the input of the Kalman filter, the state estimate of the augmented model can be sequentially obtained, and based on the estimated state of the augmented model, the estimated value of the cylinder No. 1 pressure signal can be recursively obtained, and simultaneously other three cylinder pressure signals can also be obtained recursively based on the delay block bank which contains three delay blocks.

Below, both part A and part B are described in detail, i.e., the following aspects such as cylinder pressure signal modeling, delay block, model augmentation, and cylinder pressure estimation are illustrated.

5.2.1 Cylinder Pressure Signal Modeling

In a four-cylinder engine, each cylinder fires once every one engine cycle (i.e., 720° crank angle). Thus under stationary operating conditions, the cylinder pressure signal of each cylinder is approximately periodic with a period 4π rad such that in discrete-angle domain the cylinder pressure signal can be approximately written as a Fourier series [52] with constant amplitudes $a(0)$ and $a(i)$ and phases $\alpha(i)$, i.e.,

$$P_m(\varphi(k)) \approx a(0) + \sum_{i=1}^{n_{\mathrm{p}}} a(i) \sin\left(\frac{1}{2} i\varphi(k) + \alpha(i)\right), \tag{5.1}$$

where $P_m(\varphi(k))$ denotes the cylinder No. m pressure signal, the variable n_{p} stands for the considered harmonic number, and $\varphi(k)$ denotes the engine crank angle in rad, and

$$\varphi(k) = \sum_{0}^{k} 2\pi f(k) T_{\mathrm{s}}, \tag{5.2}$$

where $f(k)$ denotes the instantaneous engine cycle frequency, the sampling period $T_{\mathrm{s}} = \frac{1}{f_{\mathrm{s}}}$, and the sampling frequency f_{s} is equal to 20 kHz.

According to Equation (5.1) and the order tracking introduced in Section 2.4, the cylinder pressure signal $P_m(\varphi(k))$ can be expressed as the output of a state-space model in time domain, i.e.,

$$\begin{cases} \boldsymbol{x}_{P_m}(k+1) = \boldsymbol{A}_{P_m}(f(k))\boldsymbol{x}_{P_m}(k), \\ P_m(k) = \boldsymbol{C}_{P_m}\boldsymbol{x}_{P_m}(k) + e_{P_m}(k), \end{cases} \tag{5.3}$$

where the state vector $\boldsymbol{x}_{P_m}(k) \in \mathbb{R}^{2n_{\mathrm{p}}+1}$, and $e_{P_m}(k)$ denotes the error term which could be induced by the modeling error when using a state-space model to represent the cylinder

pressure signal.

The occurrence of the modeling error is also the reason why the approximation symbol "$\approx$" should be involved in Equation (5.1). But in the model (4.2), the input $\boldsymbol{u}(k)$ is assumed to be noiseless for simplifying system identification.

The matrices $\boldsymbol{A}_{P_m}(f(k))$ and $\boldsymbol{C}_{P_m}$ are respectively given as

$$\boldsymbol{A}_{P_m}(f(k)) = \left(\begin{array}{c|ccc} 1 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{A}_1(f(k)) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{A}_{n_\mathrm{p}}(f(k)) \end{array}\right) \tag{5.4}$$

and

$$\boldsymbol{C}_{P_m} = \left(\begin{array}{c|ccc} 1 & \boldsymbol{C}_1 & \cdots & \boldsymbol{C}_{n_\mathrm{p}} \end{array}\right). \tag{5.5}$$

The individual block entries in these block matrices follow from the state-space representation of a frequency-modulated sine wave as

$$\boldsymbol{A}_i(f(k)) = \begin{pmatrix} \cos(2\pi i f(k) T_\mathrm{s}) & \sin(2\pi i f(k) T_\mathrm{s}) \\ -\sin(2\pi i f(k) T_\mathrm{s}) & \cos(2\pi i f(k) T_\mathrm{s}) \end{pmatrix} \tag{5.6}$$

and

$$\boldsymbol{C}_i = \begin{pmatrix} 1 & 0 \end{pmatrix}. \tag{5.7}$$

Remark 5.1. The number "1" in both (5.4) and (5.5) corresponds to the DC offset of $P_m(k)$. It should also be noted that the form of the state-space model (5.3) is hold if the amplitude $a(i)$ is constant. While for time-variant amplitude $a(i)$ (i.e., under non-stationary conditions), a forgetting factor introduced in Section 2.2.4.1 can be involved in the Kalman filter for the model $\boldsymbol{G}_\mathrm{a}$ shown in Figure 5.1. □

5.2.2 Delay Block

For the four-cylinder engine test bench shown in Figure 3.1, as aforementioned, every two consecutive cylinder pressure signals have nearly the same curve and have a fixed phase difference 180° (π rad), i.e.,

$$\begin{cases} P_3(\varphi(k)) \approx P_1(\varphi(k) + \pi), \\ P_4(\varphi(k)) \approx P_3(\varphi(k) + \pi), \\ P_2(\varphi(k)) \approx P_4(\varphi(k) + \pi). \end{cases} \tag{5.8}$$

If $\frac{f_\mathrm{s}}{4f(k)}$ is an integer at each step k, in (5.8) ($f(k) \neq 0$ when the engine is in working mode), the phase delay π in discrete-time domain can be equivalently expressed as $\frac{f_\mathrm{s}}{4f(k)}$, i.e.,

$$\begin{cases} P_3(k) \approx P_1\left(k + \dfrac{f_\mathrm{s}}{4f(k)}\right), \\ P_4(k) \approx P_3\left(k + \dfrac{f_\mathrm{s}}{4f(k)}\right), \\ P_2(k) \approx P_4\left(k + \dfrac{f_\mathrm{s}}{4f(k)}\right). \end{cases} \tag{5.9}$$

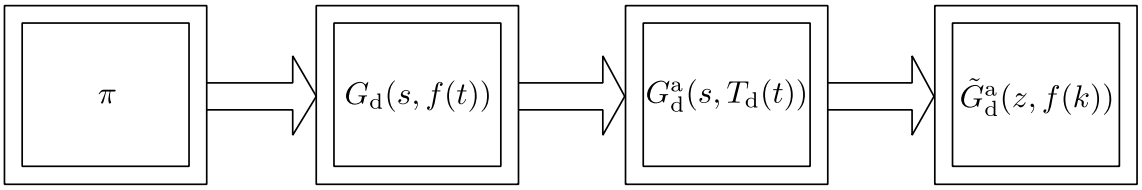

Figure 5.2: Delay block evolution flow.

However, if the sampling frequency f_{s} is too large, $\frac{f_{\mathrm{s}}}{4f(k)}$, which denotes the discrete steps corresponding to the conceptual time-varying delay block transfer function $z^{-N(k)}$ with $N(k) = \frac{f_{\mathrm{s}}}{4f(k)}$, could then be too large for practical implementations. The above analysis is just a rough evaluation because $N(k)$ may be a non-integer number, but it can give us a sense how the number discrete steps can change. In order to avoid the large number of the steps, in the remainder of this section, a strategy is proposed, and the rough idea behind the strategy is that first equivalently transform the phase delay π into the delay $\frac{1}{4f(t)}$ ($f(t) \neq 0$) in continuous-time domain corresponding to the conceptual time-varying delay block transfer function $G_{\mathrm{d}}(s, f(t))$, afterwards discretize $G_{\mathrm{d}}(s, f(t))$ such that the above large step problem can be avoided and simultaneously a discrete-time model, which can be directly involved in cylinder pressure reconstruction, can be obtained.

Below, three parts (from Section 5.2.2.1 to Section 5.2.2.3) are used to illustrate the specific strategy, and the total process is shown in Figure 5.2.

5.2.2.1 Conceptual time-varying delay block transfer function

As mentioned above, the function $G_{\mathrm{d}}(s, f(t))$ is called the conceptual time-varying delay block (TVDB) transfer function, because the transfer function is only defined for LTI systems. When the frequency $f(t)$ is constant, i.e., $f(t) = c$ with c a constant, the function $G_{\mathrm{d}}(s, f(t))$ turns to be a normal transfer function. The conceptual TVDB transfer function can be expressed as

$$G_{\mathrm{d}}(s, f(t)) = e^{-\frac{1}{4c}s}, \tag{5.10}$$

however, when the frequency $f(t)$ is time-variant, Equation 5.10 will not be hold, which will be illustrated in the next section, in which an approximation of $G_{\mathrm{d}}(s, f(t))$ with time-variant frequency $f(t)$ will be obtained.

5.2.2.2 Delay block transfer function approximation

The time-varying dead time $\frac{1}{4f(t)}$ can bring some additional terms to the conceptual TVDB transfer function [82], which can lead to a more complex form of the conceptual TVDB transfer function. Fortunately, Proposition 5.1 can help us to avoid the effects of the additional terms as well as get a compact form of the conceptual TVDB transfer function by using an approximation method.

Proposition 5.1. The conceptual TVDB transfer function $G_{\mathrm{d}}(s, f(t))$ can be approximately denoted as

$$G_{\mathrm{d}}(s, f(t)) \approx e^{-\frac{1}{4f_{\mathrm{a}}(t)}s}, \tag{5.11}$$

where $f_{\mathrm{a}}(t)$ denotes the zero-order hold [60] for the discrete-time frequency $f(k)$ with the sampling period T_{s}. □

Proof. See Appendix A. □

Remark 5.2. $f_{\mathrm{a}}(t)$ can approximate $f(t)$ better if the sampling period T_{s} is smaller. □

With the conclusion in Proposition 5.1, the Padé approximation technique can be involved to transform the approximation of $G_{\mathrm{d}}(s, f(t))$ into a stable and causal rational transfer function [83] under the time-varying frequency $f(t)$, i.e.,

$$\begin{aligned} G_{\mathrm{d}}(s, f(t)) &\approx e^{-T_{\mathrm{d}}(t)s} \\ &\approx \sum_{i=0}^{\omega+\nu}(-1)^i \frac{(T_{\mathrm{d}}(t)s)^i}{i!} \\ &= \frac{\sum_{i=0}^{\omega} g_i s^i}{\sum_{i=0}^{\nu} h_i s^i} \\ &= G_{\mathrm{d}}^{\mathrm{a}}(s, T_{\mathrm{d}}(t)), \end{aligned} \tag{5.12}$$

where $T_{\mathrm{d}}(t)$ denotes the approximated dead time, and $T_{\mathrm{d}}(t) = \frac{1}{4f_{\mathrm{a}}(t)}$. g_i and h_i are respectively the coefficients of the numerator and denominator polynomials [83], and they are respectively given as

$$g_i = (-1)^i \frac{(\omega+\nu-i)!\,\omega!}{(\omega+\nu)!\,i!\,(\omega-i)!}(T_{\mathrm{d}}(t))^i \tag{5.13}$$

and

$$h_i = \frac{(\omega+\nu-i)!\,\nu!}{(\omega+\nu)!\,i!\,(\nu-i)!}(T_{\mathrm{d}}(t))^i, \tag{5.14}$$

where m and n are integers.

Remark 5.3. According to the Padé approximation in [83] the author recommended the approximation $G_{\mathrm{d}}^{\mathrm{a}}(s, T_{\mathrm{d}}(t))$ where the numerator's degree is one less than that of the denominator. Thus here, set $\omega = \nu - 1$, i.e., $G_{\mathrm{d}}^{\mathrm{a}}(s, T_{\mathrm{d}}(t))$ is strictly proper. □

A controllable realization [27] of the approximated TVDB transfer function $G_{\mathrm{d}}^{\mathrm{a}}(s, T_{\mathrm{d}}(t))$ in Equation (5.12)) can be represented as

$$\begin{cases} \dot{\boldsymbol{x}}_{\mathrm{d}}(t) = \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))\boldsymbol{x}_{\mathrm{d}}(t) + \boldsymbol{B}_{\mathrm{d}}u_{\mathrm{d}}(t), \\ y_{\mathrm{d}}(t) = \boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))\boldsymbol{x}_{\mathrm{d}}(t), \end{cases} \tag{5.15}$$

where $\boldsymbol{x}_{\mathrm{d}}(t) \in \mathbb{R}^v$, $u_{\mathrm{d}}(t) \in \mathbb{R}$, and $y_{\mathrm{d}}(t) \in \mathbb{R}$ denote the state vector, the input, and the output, respectively.

The matrices $\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))$, $\boldsymbol{B}_{\mathrm{d}}$, and $\boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ are respectively expressed as

$$\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t)) = \begin{pmatrix} -\frac{h_{\nu-1}}{h_\nu} & -\frac{h_{\nu-2}}{h_\nu} & \cdots & -\frac{h_1}{h_\nu} & -\frac{h_0}{h_\nu} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \tag{5.16}$$

$$\boldsymbol{B}_{\mathrm{d}} = \begin{pmatrix} 1 \\ \boldsymbol{0} \end{pmatrix}, \tag{5.17}$$

and

$$\boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t)) = \begin{pmatrix} \frac{g_\omega}{h_\nu} & \frac{g_{\omega-1}}{h_\nu} & \cdots & \frac{g_1}{h_\nu} & \frac{g_0}{h_\nu} \end{pmatrix}, \tag{5.18}$$

where $h_\nu \neq 0$.

5.2.2.3 Delay block discretization

In this section, the discretization process of the approximated conceptual TVDB transfer function $G_{\mathrm{d}}^{\mathrm{a}}(s, T_{\mathrm{d}}(t))$ (i.e., the model (5.15)) is illustrated. As shown in Figure 5.2, the discretization of the model (5.15) is conducted after obtaining the function $G_{\mathrm{d}}^{\mathrm{a}}(s, T_{\mathrm{d}}(t))$, i.e., the process from $G_{\mathrm{d}}^{\mathrm{a}}(s, T_{\mathrm{d}}(t))$ to $\tilde{G}_{\mathrm{d}}^{\mathrm{a}}(z, f(k))$. However, without any strategy, directly discretizing the model (5.15) is non-trivial because of the complex form of the time-varying matrices $\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ and $\boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ of which the argument is the dead time $T_{\mathrm{d}}(t)$ illustrated in the above. Below, a method for handling the non-trivial discretization problem is derived. The main idea behind the method contains three parts:

(i) Substitute the value $T_{\mathrm{d}}(t) = 1$ into the model (5.15) which is a LTI model.

(ii) Parameterize the matrices $\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ and $\boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ based on the LTI model obtained in (i) such that a model depending on $T_{\mathrm{d}}(t)$ can be obtained (see Figure 5.3).

(iii) Discretize the model obtained in (ii).

Below, the above three parts are described in more detail one by one.

By substituting $T_{\mathrm{d}}(t) = 1$ into the model (5.15), a LTI state-space model can be obtained:

$$\begin{cases} \dot{\boldsymbol{x}}_{\mathrm{d}}(t) = \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}\, \boldsymbol{x}_{\mathrm{d}}(t) + \boldsymbol{B}_{\mathrm{d}} u_{\mathrm{d}}(t), \\ y_{\mathrm{d}}(t) = \boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}\, \boldsymbol{x}_{\mathrm{d}}(t), \end{cases} \tag{5.19}$$

where the matrices $\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}$ and $\boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}$ are time-invariant.

Based on the model (5.19), the matrices $\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ and $\boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ can be parameterized. Proposition 5.2 can explain how to get these parameterized matrices.

Proposition 5.2. Based on the transfer function of the model (5.19), the conceptual time-varying transfer function of the model (5.15) can be obtained as

$$G_{\mathrm{d}}^{\mathrm{a}}(s, T_{\mathrm{d}}(t)) = \frac{\boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}}{T_{\mathrm{d}}(t)} \left(s\boldsymbol{I}_\nu - \frac{\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}}{T_{\mathrm{d}}(t)} \right)^{-1} \boldsymbol{B}_{\mathrm{d}}. \tag{5.20}$$

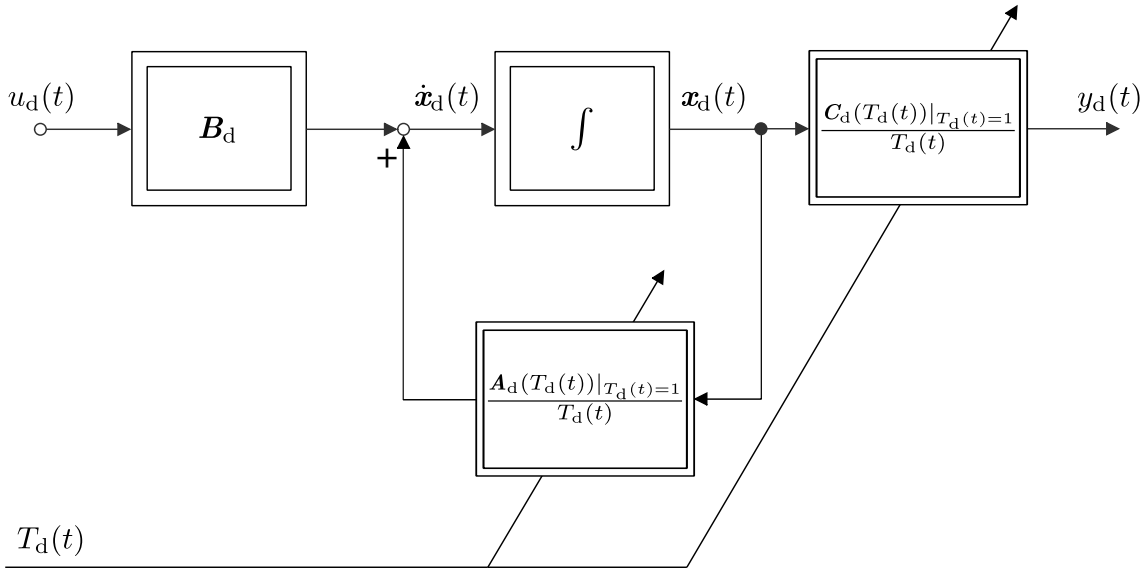

Figure 5.3: Model depending on $T_\mathrm{d}(t)$.

□

Proof. See Appendix B. □

On the basis of Proposition 5.2 the matrices $\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))$ and $\boldsymbol{C}_\mathrm{d}(T_\mathrm{d}(t))$ of the model (5.15), which are functions of the variable $T_\mathrm{d}(t)$, can be respectively represented as

$$\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t)) = \frac{\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1}}{T_\mathrm{d}(t)} \tag{5.21}$$

and

$$\boldsymbol{C}_\mathrm{d}(T_\mathrm{d}(t)) = \frac{\boldsymbol{C}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1}}{T_\mathrm{d}(t)}. \tag{5.22}$$

Based on Equations (5.21) and (5.22), the zero-order-hold discretization of the model (5.15) can be represented as

$$\begin{cases} \boldsymbol{x}_\mathrm{d}(k+1) = \boldsymbol{G}_\mathrm{d}(T_\mathrm{d}(k))\boldsymbol{x}_\mathrm{d}(t) + \boldsymbol{H}_\mathrm{d}(T_\mathrm{d}(k))u_\mathrm{d}(k), \\ y_\mathrm{d}(k) = \boldsymbol{C}_\mathrm{d}(T_\mathrm{d}(k))\boldsymbol{x}_\mathrm{d}(k), \end{cases} \tag{5.23}$$

where $T_\mathrm{d}(k) = \frac{1}{4f_\mathrm{a}(k)}$.

The matrices $\boldsymbol{G}_\mathrm{d}(T_\mathrm{d}(k))$ and $\boldsymbol{H}_\mathrm{d}(T_\mathrm{d}(k))$ are respectively given as

$$\boldsymbol{G}_\mathrm{d}(T_\mathrm{d}(k)) = \boldsymbol{\Phi}\left((k+1)T_\mathrm{s}, kT_\mathrm{s}\right) \tag{5.24}$$

and

$$\boldsymbol{H}_\mathrm{d}(T_\mathrm{d}(k)) = \int_{kT_\mathrm{s}}^{(k+1)T_\mathrm{s}} \boldsymbol{\Phi}\left((k+1)T_\mathrm{s}, \beta\right) \boldsymbol{B}_\mathrm{d}\, d\beta, \tag{5.25}$$

where $\boldsymbol{\Phi}(\cdot,\cdot)$ denotes the state transition matrix of the model (5.23).

However, it is difficult to analytically calculate the values of $\boldsymbol{G}_\mathrm{d}(T_\mathrm{d}(k))$ and $\boldsymbol{H}_\mathrm{d}(T_\mathrm{d}(k))$ in Equation (5.24) and Equation (5.25), respectively. Thus the numerical calculation is

necessary. Here, the same method as derived in the paper [84] is used when numerically calculating the matrix $\boldsymbol{G}_{\mathrm{d}}(T_{\mathrm{d}}(k))$, and an alternative method is used to numerically calculate the matrix $\boldsymbol{H}_{\mathrm{d}}(T_{\mathrm{d}}(k))$. The numerical results for $\boldsymbol{G}_{\mathrm{d}}(T_{\mathrm{d}}(k))$ and $\boldsymbol{H}_{\mathrm{d}}(T_{\mathrm{d}}(k))$ are respectively given as

$$\begin{aligned}\boldsymbol{G}_{\mathrm{d}}(T_{\mathrm{d}}(k)) \approx\ & \boldsymbol{I}_{\nu} + 4T_{\mathrm{s}}f_{\mathrm{a}}(k)\ \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1} + \frac{1}{2!}\left(4T_{\mathrm{s}}f_{\mathrm{a}}(k)\ \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}\right)^{2} \\ & + \ldots + \frac{1}{l_{\mathrm{a}}!}\left(4T_{\mathrm{s}}f_{\mathrm{a}}(k)\ \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}\right)^{l_{\mathrm{a}}} \\ =\ & \boldsymbol{G}_{\mathrm{d}}^{\mathrm{a}}(f(k)),\end{aligned} \tag{5.26}$$

and

$$\begin{aligned}\boldsymbol{H}_{\mathrm{d}}(T_{\mathrm{d}}(k)) \approx\ & T_{\mathrm{s}}\left(\boldsymbol{I}_{\nu} + \frac{1}{2!}\left(4T_{\mathrm{s}}f_{\mathrm{a}}(k)\ \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}\right)\right. \\ & \left. + \ldots + \frac{1}{l_{\mathrm{a}}!}\left(4T_{\mathrm{s}}f_{\mathrm{a}}(k)\ \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}\right)^{l_{\mathrm{a}}-1}\right)\boldsymbol{B}_{\mathrm{d}} \\ =\ & \boldsymbol{H}_{\mathrm{d}}^{\mathrm{a}}(f(k)),\end{aligned} \tag{5.27}$$

where $f_{\mathrm{a}}(k) = f(k)$ (see Proposition 5.1), and l_{a} and $l_{\mathrm{a}} - 1$ respectively denote the largest exponent number for numerical approximations of $\boldsymbol{G}_{\mathrm{d}}(T_{\mathrm{d}}(k))$ and $\boldsymbol{H}_{\mathrm{d}}(T_{\mathrm{d}}(k))$.

The specific derivations for the above numerical results can be found in Appendix C. Below, an example is given to illustrate the effectiveness of the above two numerical approximations.

Based on the above numerical approximations for the matrix $\boldsymbol{G}_{\mathrm{d}}(T_{\mathrm{d}}(k))$ and the matrix $\boldsymbol{H}_{\mathrm{d}}(T_{\mathrm{d}}(k))$, the numerically discretized version of the approximated TVDB model (5.15) can be obtained, and the discretized model can be represented as

$$\begin{cases}\boldsymbol{x}_{\mathrm{d}}^{\mathrm{a}}(k+1) = \boldsymbol{G}_{\mathrm{d}}^{\mathrm{a}}(f(k))\boldsymbol{x}_{\mathrm{d}}^{\mathrm{a}}(t) + \boldsymbol{H}_{\mathrm{d}}^{\mathrm{a}}(f(k))u_{\mathrm{d}}(k), \\ y_{\mathrm{d}}^{\mathrm{a}}(k) = \boldsymbol{C}_{\mathrm{d}}(f(k))\boldsymbol{x}_{\mathrm{d}}^{\mathrm{a}}(k),\end{cases} \tag{5.28}$$

where $\boldsymbol{x}_{\mathrm{d}}^{\mathrm{a}}(k) \in \mathbb{R}^{v}$, $u_{\mathrm{d}}(k) \in \mathbb{R}$, and $y_{\mathrm{d}}^{\mathrm{a}}(k) \in \mathbb{R}$ denote the state vector, the input, and the output, respectively, and $\boldsymbol{C}_{\mathrm{d}}(f(k)) = 4f_{\mathrm{a}}(k)\ \boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}$ (see Equation (5.22)).

The conceptual time-varying transfer function of the model (5.28) is represented as $\tilde{G}_{\mathrm{d}}^{\mathrm{a}}(z, f(k))$, and based on the above process which is from (5.8) to (5.28) and which is also called delay block evolution illustrated in Figure 5.2, the phase delay π can be transformed from $G_{\mathrm{d}}(s, f(t))$ to $\tilde{G}_{\mathrm{d}}^{\mathrm{a}}(z, f(k))$. The obtained model $\tilde{G}_{\mathrm{d}}^{\mathrm{a}}$ (i.e., the model (5.28)) is finally is used for the cylinder pressure reconstrusction as illustrated in Figure 5.1.

It should be pointed out that there still exist two problems in the above delay block discretization process:

(i) As aforementioned, the model (5.15) is in controllable canonical form, however, the observability cannot be guaranteed.

(ii) As seen in Equation (5.16), the expression form of the matrix $\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ is prone to numerical problems.

A balanced realization of the model (5.15), which can handle the above two problems simultaneously, can be involved in the discretization process. The specific realization and

corresponding results can be referred to Appendix D. In this thesis, balanced realization is involved, and based on one conclusion from Appendix D that the balanced realization problem of the model (5.15) can be transformed into the balanced realization problem of the model (5.19), and the total discretization procedure of the model (5.15) as introduced at the start of this section should be modified as follows:

(i) Substitute the value $T_{\mathrm{d}}(t) = 1$ into the model (5.15) which is a LTI model.

(ii) Make a balanced realization of the LTI model obtained in (i) such that a minimally-realized LTI model can be obtained.

(iii) Parameterize the matrices $\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ and $\boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ based on the minimally-realized LTI model obtained in (ii) such that a model depending on $T_{\mathrm{d}}(t)$ can be obtained (see Figure 5.3).

(iv) Discretize the model obtained in (iii).

Because the final discretized model (5.28) will be used for cylinder pressure reconstruction, the reachability and observability of the model (5.28) should be still verified after numerical approximation even though the model (5.15) is minimally realized after balanced realization. The corresponding verifications can be found in Appendix E.

In the remainder of this section, two examples related to the discretized model (5.28) will be given, and based on the examples some conclusions will be obtained either.

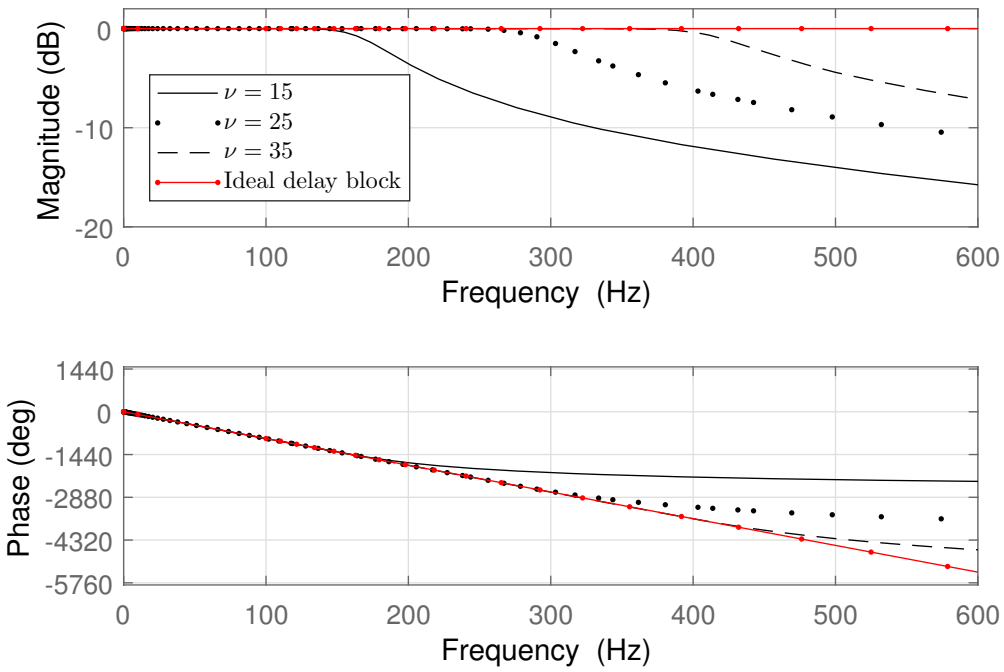

Figure 5.4: Bode plot of the discretized TVDB models with different model orders.

Example 5.1. In the model (5.28), fix the value of $f(k)$ to 10 Hz, and set l_{a} to 10 and set T_{s} to 1×10^{-4} seconds for the numerical approximations of $\boldsymbol{G}_{\mathrm{d}}(T_{\mathrm{d}}(k))$ and $\boldsymbol{H}_{\mathrm{d}}(T_{\mathrm{d}}(k))$ (see Equations (5.26) and (5.27)), then three discretized TVDB models can be obtained by selecting three different model orders ($\nu = 15$, $\nu = 25$, and $\nu = 35$) for the model (5.28). The Bode plots of three discretized TVDB models with different model orders are

illustrated in Figure 5.4. By observing Figure 5.4, it can be known that under fixed l_{a}, fixed sampling period, and fixed dead time, the passband cutoff frequencies [51] of the discretized TVDB becomes smaller as ν decreases. Moreover, by comparing with the ideal delay block (i.e., $z^{-\frac{1}{4f(k)T_{\mathrm{s}}}}$), the effectiveness of the discretization of TVDB is verified. □

Example 5.2. Given two angle-domain periodic signals, one signal $u(k) = \sin(\varphi(k)) + \sin(2\varphi(k)) + \sin(3\varphi(k))$ and the other signal $u_{\mathrm{d}}(k) = \sin\left(\varphi(k) - \frac{\pi}{2}\right) + \sin\left(2\left(\varphi(k) - \frac{\pi}{2}\right)\right) + \sin\left(3\left(\varphi(k) - \frac{\pi}{2}\right)\right)$, where the signal $u_{\mathrm{d}}(k)$ is the delay of the signal $u(k)$, $\varphi(k) - \varphi(k-1) = 2\pi f^*(k)T_{\mathrm{s}}^*$, and the initial angle value $\varphi(0) = 2\pi f^*(0)T_{\mathrm{s}}^*$. The curve of the instantaneous frequency $f^*(k)$ is shown in Figure 5.5, and the sampling period T_{s}^* is equal to 1×10^{-4} seconds.

According to the above information it can be known that $u(k)$ and $u_{\mathrm{d}}(k)$ have the same curve but different phases. The delay between these two signals in angle domain is $\frac{\pi}{2}$, thus in continuous-time domain, the delay becomes $\frac{1}{4f^*(t)}$, and in discrete-time domain, the corresponding delay changes to $\frac{1}{4f^*(k)}$.

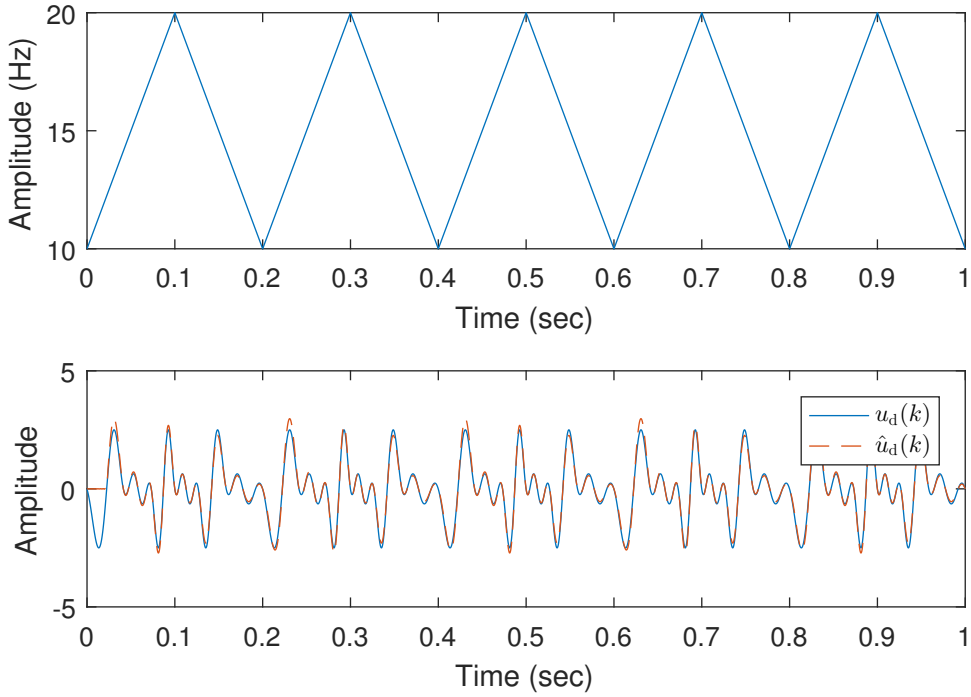

Figure 5.5: Curve of instantaneous frequency $f^*(k)$ (upper figure) and simulated output of the discretized TVDB model (lower figure).

Use the signal $u(k)$ as the input of the model (5.28) with $\nu = 35$ and $l_{\mathrm{a}} = 10$, as shown in Figure 5.5 the simulated output $\hat{u}_{\mathrm{d}}(k)$ is nearly the same as the signal $u_{\mathrm{d}}(k)$, which can illustrate the effectiveness of the delay block discretization method proposed in this section. □

5.2.3 Model Augmentation

As shown in Figure 5.1, the SISO model with an additive error can be represented as

$$\begin{cases} \boldsymbol{x}_{\mathrm{s}}(k+1) = \boldsymbol{A}_{\mathrm{s}}(f(k))\boldsymbol{x}_{\mathrm{s}}(k) + \boldsymbol{B}_{\mathrm{s}}(f(k))P_1(k), \\ y(k) = \boldsymbol{C}_{\mathrm{s}}(f(k))\boldsymbol{x}_{\mathrm{s}}(k) + \boldsymbol{D}_{\mathrm{s}}P_1(k) + e_{\mathrm{l}}(k), \end{cases} \tag{5.29}$$

where $\boldsymbol{x}_{\mathrm{s}}(k) \in \mathbb{R}^{n_{\mathrm{s}}}$, and the output error $e_{\mathrm{l}}(k)$ is induced by the measurement noise, the disturbance on the model $\boldsymbol{G}$, the cylinder-to-cylinder difference, and the error in the formulation process of the TVDB model (5.28).

The matrices $\boldsymbol{A}_{\mathrm{s}}(f(k))$, $\boldsymbol{B}_{\mathrm{s}}(f(k))$, $\boldsymbol{C}_{\mathrm{s}}(f(k))$, $\boldsymbol{D}_{\mathrm{s}}$, and the state vector $\boldsymbol{x}_{\mathrm{s}}(k)$ are respectively given as

$$\boldsymbol{A}_{\mathrm{s}}(f(k)) = \left(\begin{array}{c|ccc} \boldsymbol{A} & \boldsymbol{B}\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{C}_{\mathrm{d}}(f(k)) \end{pmatrix} \\ \hline \boldsymbol{0} & \begin{pmatrix} \boldsymbol{G}_{\mathrm{d}}^{\mathrm{a}}(f(k)) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{H}_{\mathrm{d}}^{\mathrm{a}}(f(k))\boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{G}_{\mathrm{d}}^{\mathrm{a}}(f(k)) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_{\mathrm{d}}^{\mathrm{a}}(f(k))\boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{G}_{\mathrm{d}}^{\mathrm{a}}(f(k)) \end{pmatrix} \end{array}\right), \tag{5.30}$$

$$\boldsymbol{B}_{\mathrm{s}}(f(k)) = \begin{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \boldsymbol{B}^{\mathrm{T}} & \left(\boldsymbol{H}_{\mathrm{d}}^{\mathrm{a}}(f(k))\right)^{\mathrm{T}} & \boldsymbol{0} \end{pmatrix}^{\mathrm{T}}, \tag{5.31}$$

$$\boldsymbol{C}_{\mathrm{s}}(f(k)) = \left(\begin{array}{c|c} \boldsymbol{C} & \boldsymbol{D}\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{C}_{\mathrm{d}}(f(k)) \end{pmatrix} \end{array}\right), \tag{5.32}$$

$$\boldsymbol{D}_{\mathrm{s}} = \boldsymbol{D}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{5.33}$$

and

$$\boldsymbol{x}_{\mathrm{s}}(k) = \begin{pmatrix} \boldsymbol{x}^{\mathrm{T}}(k) & \boldsymbol{x}_1^{\mathrm{T}}(k) & \boldsymbol{x}_2^{\mathrm{T}}(k) & \boldsymbol{x}_3^{\mathrm{T}}(k) \end{pmatrix}^{\mathrm{T}}, \tag{5.34}$$

where $\boldsymbol{x}_1(k)$, $\boldsymbol{x}_2(k)$, and $\boldsymbol{x}_3(k)$ denote the state vectors of three delay block models shown in Figure 5.1, respectively.

By augmenting the model (5.29) with the state of the model of the cylinder No. 1 pressure signal (i.e., the model (5.3)), an augmented model $\boldsymbol{G}_{\mathrm{a}}$ can be obtained

$$\begin{cases} \boldsymbol{x}_{\mathrm{a}}(k+1) = \boldsymbol{A}_{\mathrm{a}}(f(k))\boldsymbol{x}_{\mathrm{a}}(k), \\ y(k) = \boldsymbol{C}_{\mathrm{a}}(f(k))\boldsymbol{x}_{\mathrm{a}}(k)) + e_{\mathrm{a}}(k), \end{cases} \tag{5.35}$$

where the state vector $\boldsymbol{x}_{\mathrm{a}}(k) \in \mathbb{R}^{n_{\mathrm{a}}}$, and the term $e_{\mathrm{a}}(k)$ denotes the output error which is induced by $e_{\mathrm{l}}(k)$ and $e_{P_{\mathrm{l}}}(k)$.

The matrices $\boldsymbol{A}_{\mathrm{a}}(f(k))$, $\boldsymbol{C}_{\mathrm{a}}(f(k))$, and the state vector $\boldsymbol{x}_{\mathrm{a}}(k)$ are respectively given as

$$\boldsymbol{A}_{\mathrm{a}}(f(k)) = \begin{pmatrix} \boldsymbol{A}_{\mathrm{s}}(f(k)) & \boldsymbol{B}_{\mathrm{s}}(f(k))\boldsymbol{C}_{P_{\mathrm{l}}} \\ \boldsymbol{0} & \boldsymbol{A}_{P_{\mathrm{l}}}(f(k)) \end{pmatrix}, \tag{5.36}$$

$$\boldsymbol{C}_{\mathrm{a}}(f(k)) = \begin{pmatrix} \boldsymbol{C}_{\mathrm{s}}(f(k)) & \boldsymbol{D}_{\mathrm{s}}\boldsymbol{C}_{P_{\mathrm{l}}} \end{pmatrix}, \tag{5.37}$$

and

$$\boldsymbol{x}_{\mathrm{a}}(k) = \begin{pmatrix} \boldsymbol{x}_{\mathrm{s}}^{\mathrm{T}}(k) & \boldsymbol{x}_{P_{\mathrm{l}}}^{\mathrm{T}}(k) \end{pmatrix}^{\mathrm{T}}. \tag{5.38}$$

5.2.4 Cylinder Pressure Estimation

Based on the cylinder pressure estimation framework displayed in Figure 5.1 and the above analysis, briefly speaking, by using the Kalman filter for the augmented model (5.35), the state of the augmented model can be estimated, and then the cylinder No. 1 pressure signal can be estimated based on the model (5.3), simultaneously other three cylinder pressure signals can be obtained either by using delay blocks.

Algorithm 5.1: Cylinder pressure estimation algorithm using the square-root Kalman filter with a forgetting factor.

Initialization: $\boldsymbol{S}(0) = 100\boldsymbol{I}_{n_\mathrm{a}}$, $\hat{\boldsymbol{x}}_\mathrm{a}(0) = \boldsymbol{0}$

for $k = 1 \to N-1$ **do**

`/* Square-root Kalman filter with a forgetting factor */`

$\boldsymbol{S}^-(k) = \frac{1}{\lambda}\boldsymbol{A}_\mathrm{a}(f(k-1))\boldsymbol{S}(k-1)$ `// λ denotes the forgetting factor`

$\hat{\boldsymbol{x}}_\mathrm{a}^-(k) = \boldsymbol{A}_\mathrm{a}(f(k-1))\hat{\boldsymbol{x}}_\mathrm{a}(k-1)$

$\boldsymbol{F}(k) = \boldsymbol{S}^-(k)\boldsymbol{C}_\mathrm{a}^\mathrm{T}(f(k))$

$\alpha(k) = \frac{1}{\boldsymbol{F}^\mathrm{T}(k)\boldsymbol{F}(k)+\sigma_\mathrm{a}^2}$

$\boldsymbol{K}(k) = \alpha(k)\boldsymbol{S}^-(k)\boldsymbol{F}(k)$

$\gamma(k) = \frac{1}{1+\sqrt{\sigma_\mathrm{a}^2}\sqrt{\alpha(k)}}$

$\boldsymbol{S}(k) = \boldsymbol{S}^-(k) - \gamma(k)\boldsymbol{K}(k)\boldsymbol{F}^\mathrm{T}(k)$

$\hat{\boldsymbol{x}}_\mathrm{a}(k) = \hat{\boldsymbol{x}}_\mathrm{a}^-(k) + \boldsymbol{K}(k)\left(y(k) - \boldsymbol{C}_\mathrm{a}(f(k))\hat{\boldsymbol{x}}_\mathrm{a}^-(k)\right)$

`/* Cylinder pressure estimation */`

$\hat{P}_1(k) = \begin{pmatrix}\boldsymbol{0} & \boldsymbol{C}_{P_1}\end{pmatrix}\hat{\boldsymbol{x}}_\mathrm{a}(k)$

$\hat{P}_3(k) = \tilde{G}_\mathrm{d}^\mathrm{a}(q^{-1}, f(k))\hat{P}_1(k)$ `// ` $\tilde{G}_\mathrm{d}^\mathrm{a}(q^{-1}, f(k))$ `denotes the conceptual time-varying transfer operator of the model` (5.28) `(i.e.,` $\tilde{G}_\mathrm{d}^\mathrm{a}$`)`

$\hat{P}_4(k) = \tilde{G}_\mathrm{d}^\mathrm{a}(q^{-1}, f(k))\hat{P}_3(k)$

$\hat{P}_2(k) = \tilde{G}_\mathrm{d}^\mathrm{a}(q^{-1}, f(k))\hat{P}_4(k)$

In the model (5.35), $\{e_\mathrm{a}(k)\}$ is assumed to be a scalar white noise process, of which the covariance function is $\sigma_\mathrm{a}^2\delta_{kj}$, and the value of σ_a^2 is tunable. Additionally, $e_\mathrm{a}(k)$ is also assumed to be uncorrelated with $\boldsymbol{x}_\mathrm{a}(0)$ which is assumed to be a random variable with constant mean and variance. Thus the specific cylinder pressure estimation algorithm can be summarized in Algorithm 5.1. In the algorithm, the square-root Kalman filter with a forgetting factor is used for handling the numerical and non-stationary problem (see Section 2.4.2) and the mismodeling problems of TVDB and $\boldsymbol{G}$.

In order to make sure that the square-root Kalman filter used in Algorithm 5.1 is convergent, the observability of the augmented model (5.35) should be guaranteed. In Appendix F, the specific verification of the observability of the augmented model (5.35) is given. According to the verification, it can be known that the augmented model (5.35) is uniformly observable under certain conditions, and as a result the square-root Kalman filter with a forgetting factor for the augmented model is convergent. The reason why the stability of the square-root Kalman filter can be guaranteed under the observability condition is that the square-root Kalman filter with a forgetting factor can be directly derived from the normal Kalman filter [28], and the normal Kalman filter for the model (5.35) is stable under the condition that the model (5.35) uniformly observable (see Section 2.2.3), as a result the stability of the square-root Kalman filter with a forgetting for the

model (5.35) can also be guaranteed.

5.3 Cylinder Pressure Estimation Results Based on Time-Domain Identified Model

For the cylinder pressure estimation in Algorithm 5.1, there are a number of tunable parameters, and the parameters used is displayed in Table 5.1.

Table 5.1: Tunable parameters.

Parameter	Value
ν	35
l_{a}	10
n_{p}	40
σ_{a}^2	2.3×10^{-5}
λ	0.9995
$\hat{\boldsymbol{G}}$	9th order and MOESP(s)

In Table 5.1, the values of σ_{a}^2 and λ were chosen according to empirical analysis, while the model $\hat{\boldsymbol{G}}$ was chosen based on the identification results in Section 4.3.1. The specific selection methods of the values of ν, l_{a}, and $n_{\mathrm{p}} = 40$ are described as follows:

(i) The values of ν and l_{a} were chosen based on the performance of the discretized TVDB model (5.28) under three stationary operating conditions Z_2 (1200 rpm, 120 Nm), Z_5 (2100 rpm, 120 Nm), and Z_8 (3000 rpm, 120 Nm).

(ii) The value n_{p} was selected based on Algorithm 4.1.

As for the first item mentioned above, the TVDB performance for the cylinder pressure estimation is validated based on the strategy depicted in Figure 5.6. Specifically, first use the cylinder No. 1 pressure as the input of the delay bank which consists of three tunable TVDBs, and then simulated cylinder pressure outputs can be obtained, and finally a specified TVDB can be selected based on comparing the outputs of the tunable TVDBs.

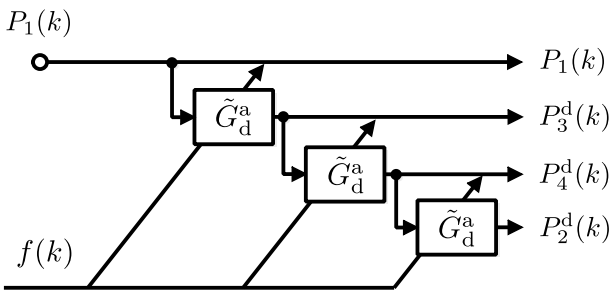

Figure 5.6: TVDB validation strategy.

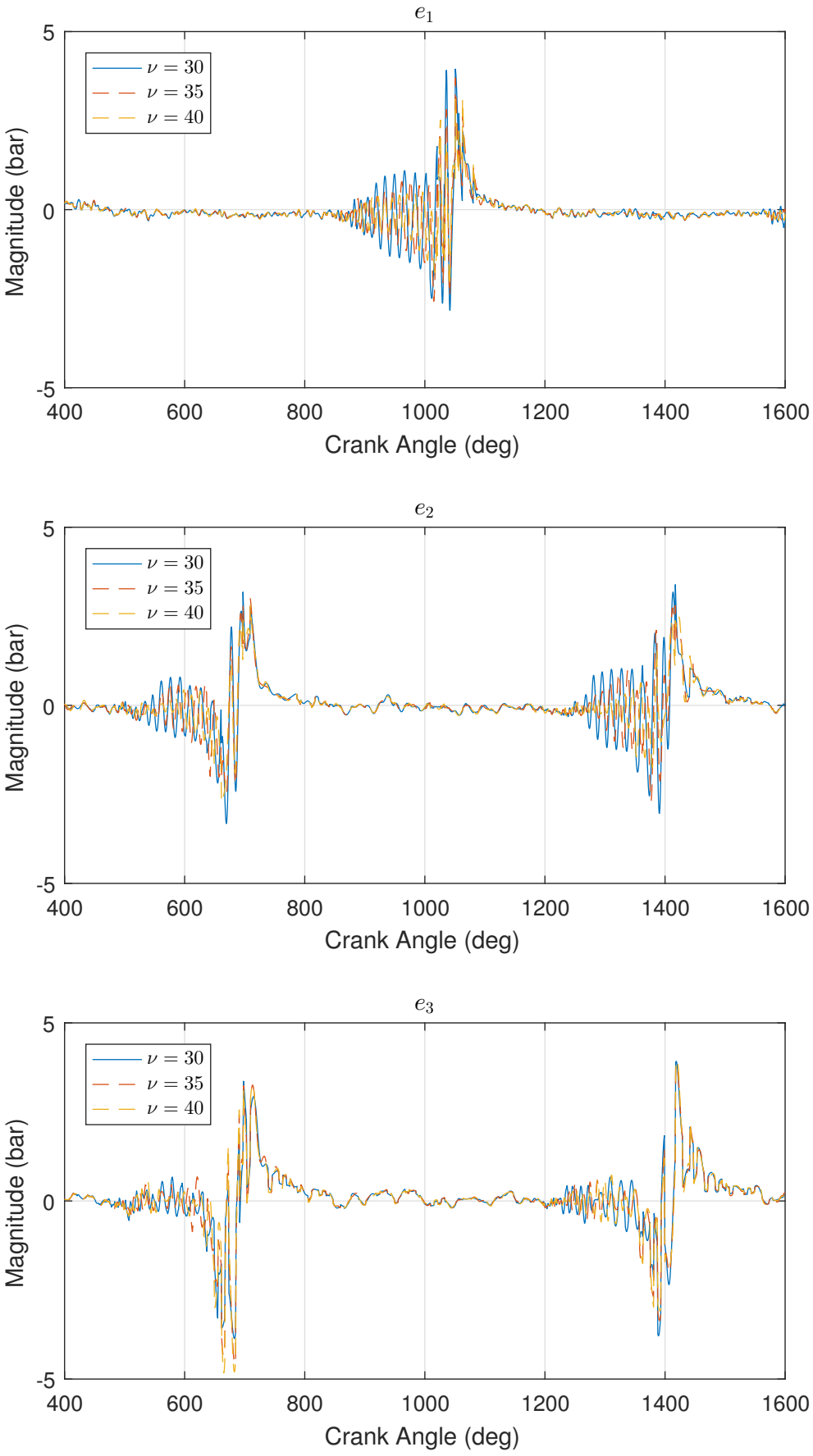

Figure 5.7: The difference between P_3 and P_3^{d} under Z_2 (upper figure), Z_5 (middle figure), and Z_8 (lower figure).

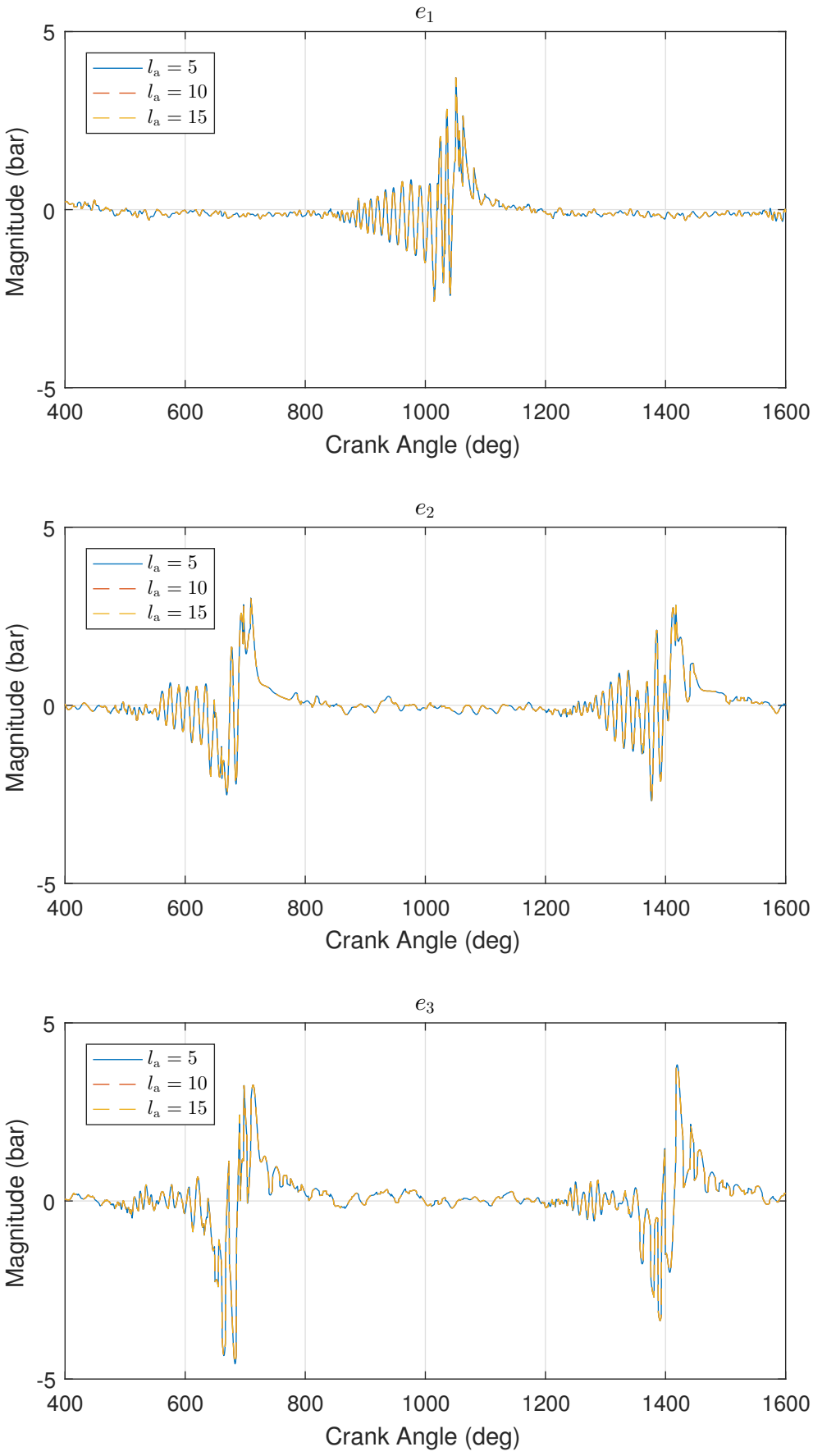

Figure 5.8: The difference between P_3 and P_3^{d} under Z_2 (upper figure), Z_5 (middle figure), and Z_8 (lower figure).

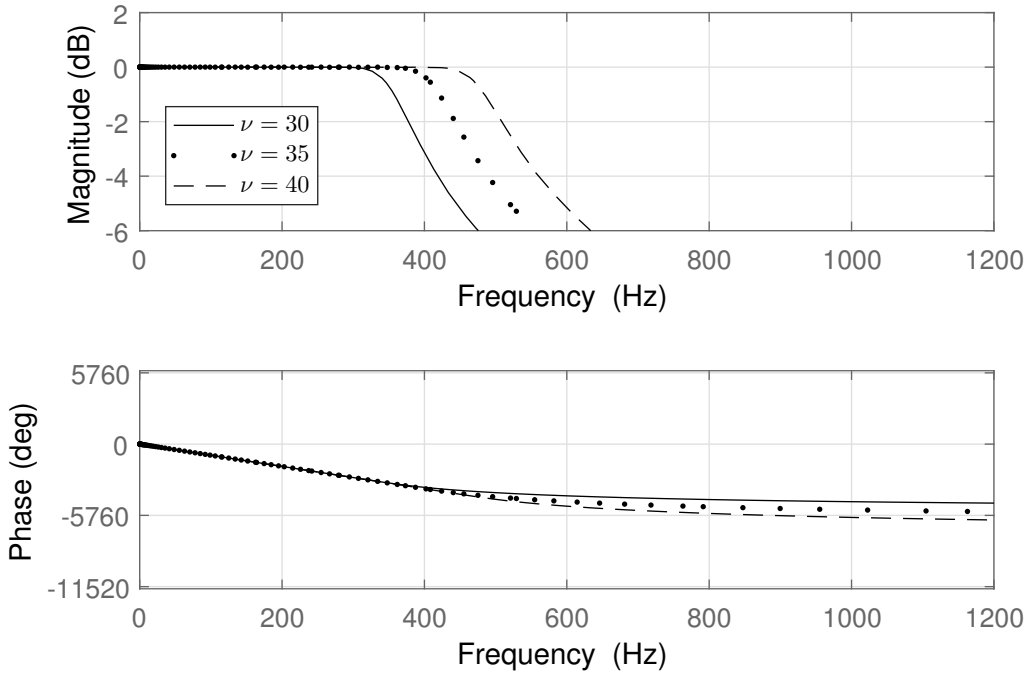

Figure 5.9: Bode plots of TVDB with three different values of ν, the same frequency $f(k)$ which is 10 Hz (i.e., exact 1200 rpm), and the same l_{a} which is 10.

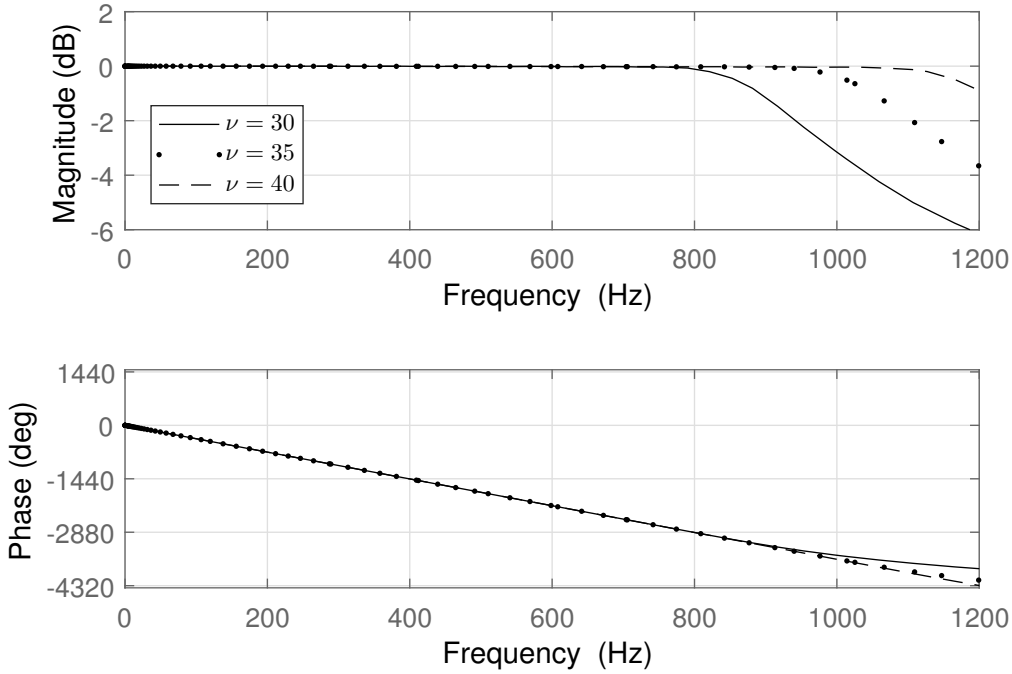

Figure 5.10: Bode plots of TVDB with three different values of ν, the same frequency $f(k)$ which is 25 Hz (i.e., exact 3000 rpm), and the same l_{a} which is 10.

The validation results are shown in Figure 5.7 and Figure 5.8. In the figures, e_1, e_2, and e_3 denote the differences between P_3 (the cylinder No. 3 pressure signal) and P_3^{d} (the delayed version of the cylinder No. 3 pressure signal) under the operating conditions Z_2, Z_5, and Z_8, respectively, are illustrated. Specifically, in Figure 5.7, the value of l_{a} is set to 10, and three different values of ν are specified, and according to the results it can be known that the performance of the TVDB with $\nu = 35$ and the performance of the TVDB with $\nu = 40$ are nearly the same, and both of them are better than the performance of the TVDB with $\nu = 30$. While in Figure 5.8, the value of ν is set to 35, and three different values of l_{a} are specified, and the corresponding results show that the values of l_{a} do not affect much on the performance of TVDB. In both figures, the absolute values of the errors e_1, e_2, and e_3 do not exceed 2 bar in general. Additionally, in the figures, only the cylinder No. 3 pressure signal is considered to validate the performance of TVDB.

Finally, the reasons why ν and l_{a} were respectively chosen as 35 and 10 are as follows:

(i) The TVDB performance under $\nu = 35$ is almost the same as the TVDB performance under $\nu = 40$.

(ii) The performance of the TVDB with three different values of l_{a} is almost the same.

(iii) As shown in Figure 5.9 and Figure 5.10, the passband cutoff frequency of the TVDB with $\nu = 35$ is enough even though it is smaller than the passband cutoff frequency of the TVDB with $\nu = 40$.

After the validation of the TVDB performance for the cylinder pressure estimation, Algorithm 5.1 was implemented to estimate four cylinder pressure signals. From Figure 5.11 to Figure 5.19, the estimated curves of four cylinder pressure signals under nine stationary operating conditions are displayed.

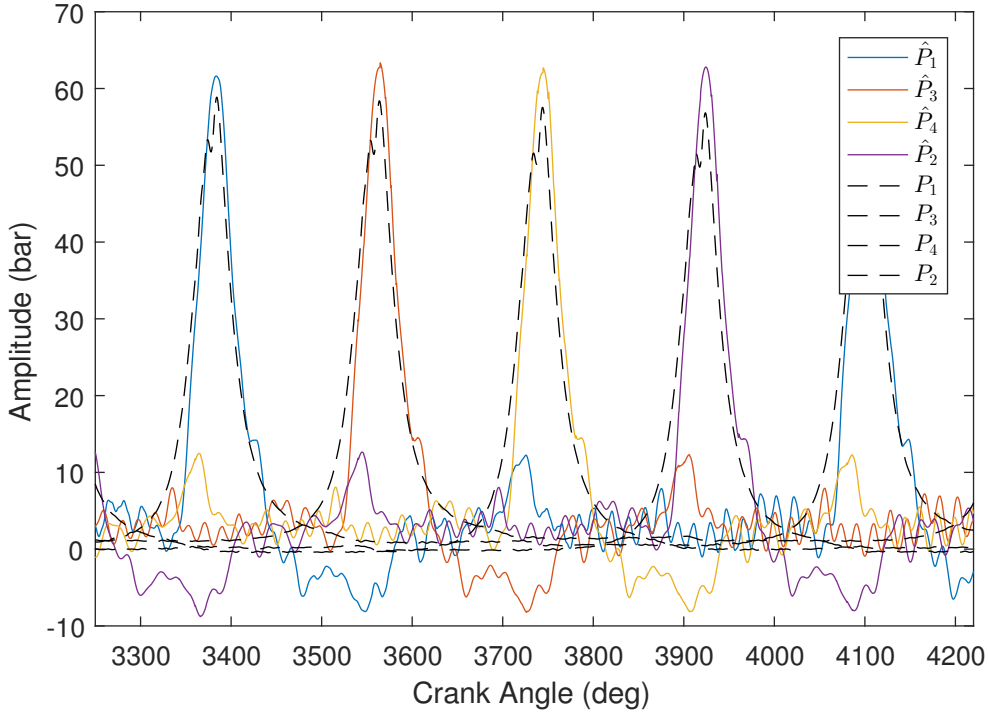

Figure 5.11: Cylinder pressure estimation under the stationary operating condition Z_1 (1200 rpm, 60 Nm) using Algorithm 5.1.

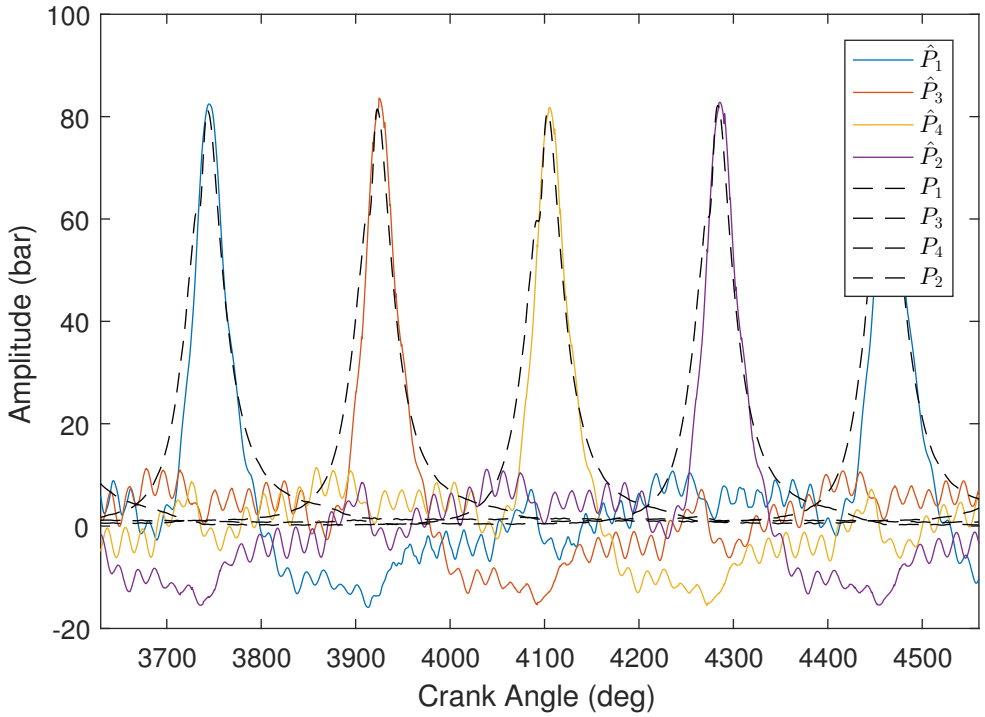

Figure 5.12: Cylinder pressure estimation under the stationary operating condition Z_2 (1200 rpm, 120 Nm) using Algorithm 5.1.

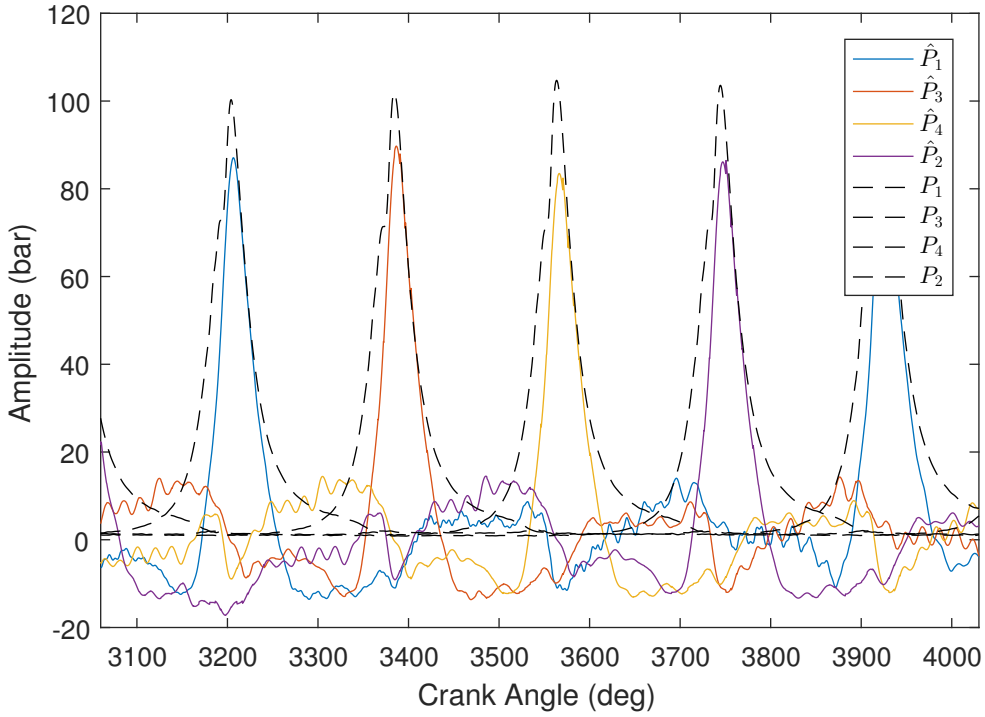

Figure 5.13: Cylinder pressure estimation under the stationary operating condition Z_3 (1200 rpm, 180 Nm) using Algorithm 5.1.

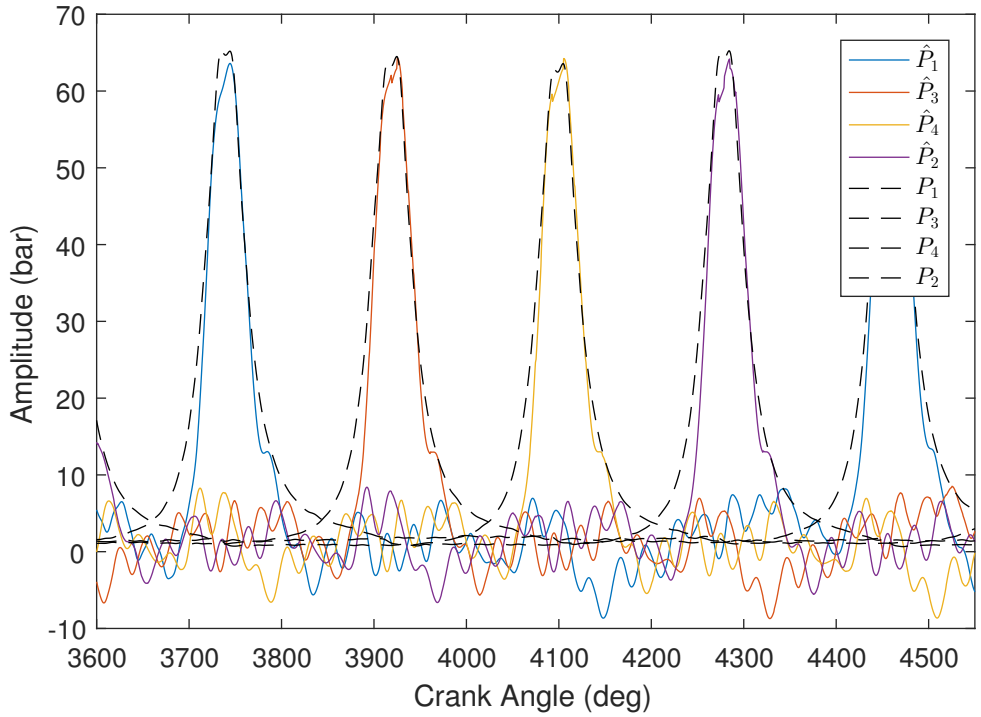

Figure 5.14: Cylinder pressure estimation under the stationary operating condition Z_4 (2100 rpm, 60 Nm) using Algorithm 5.1.

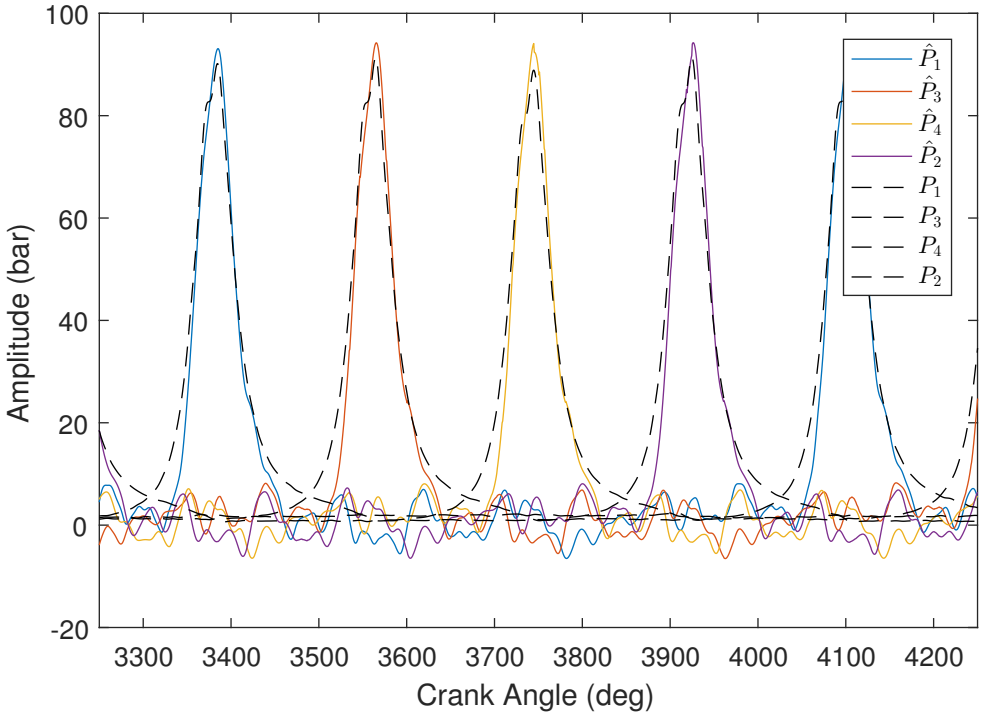

Figure 5.15: Cylinder pressure estimation under the stationary operating condition Z_5 (2100 rpm, 120 Nm) using Algorithm 5.1.

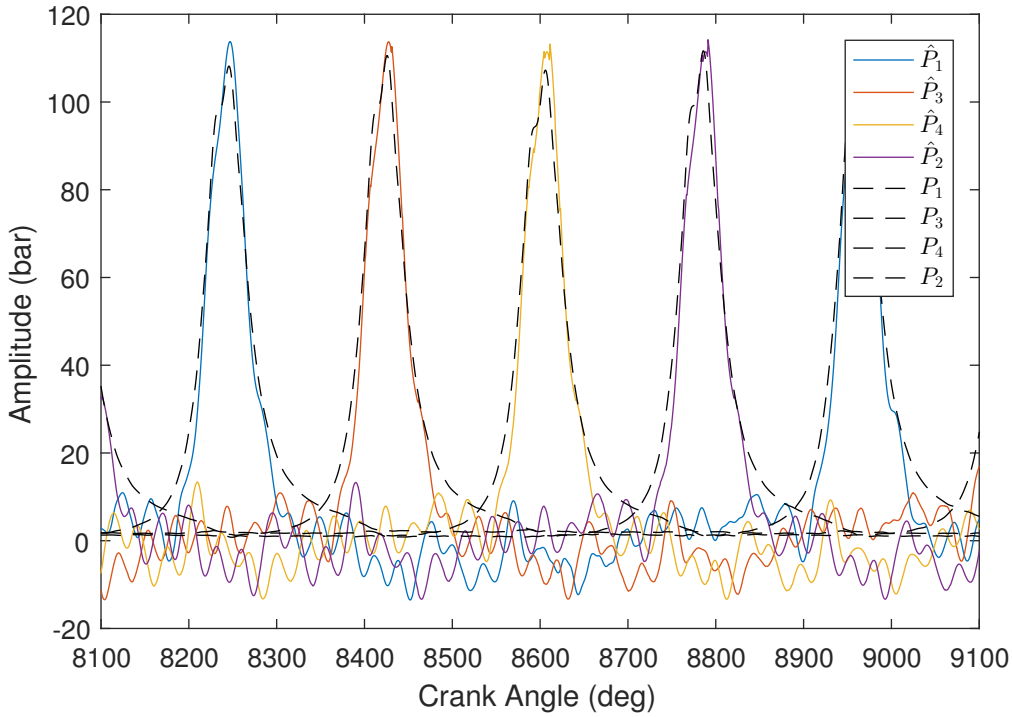

Figure 5.16: Cylinder pressure estimation under the stationary operating condition Z_6 (2100 rpm, 180 Nm) using Algorithm 5.1.

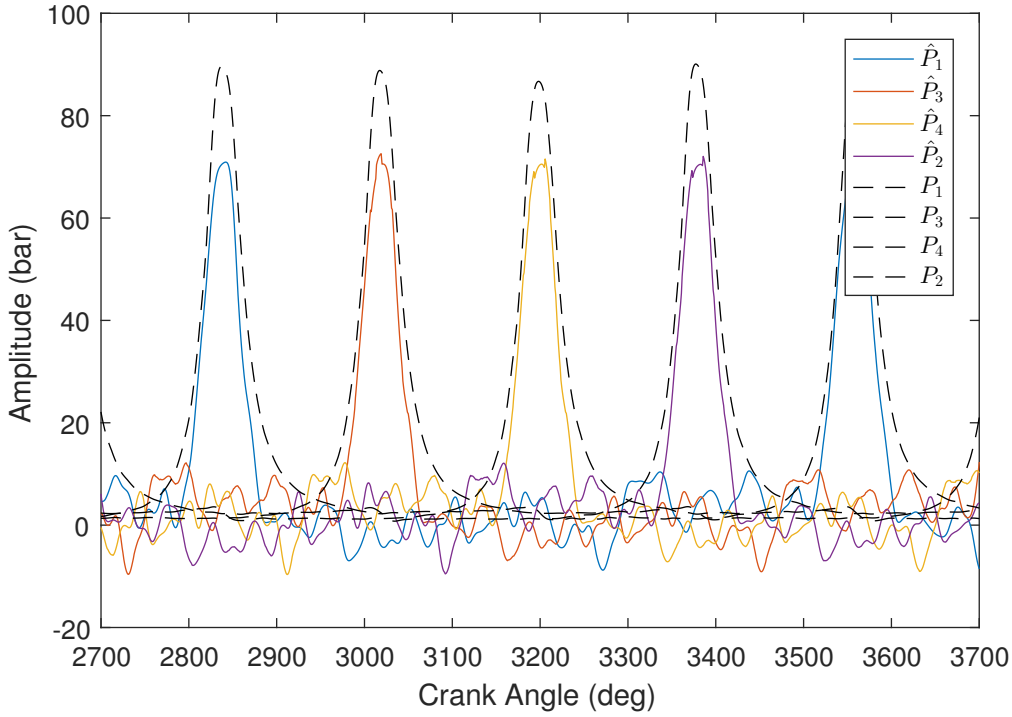

Figure 5.17: Cylinder pressure estimation under the stationary operating condition Z_7 (3000 rpm, 60 Nm) using Algorithm 5.1.

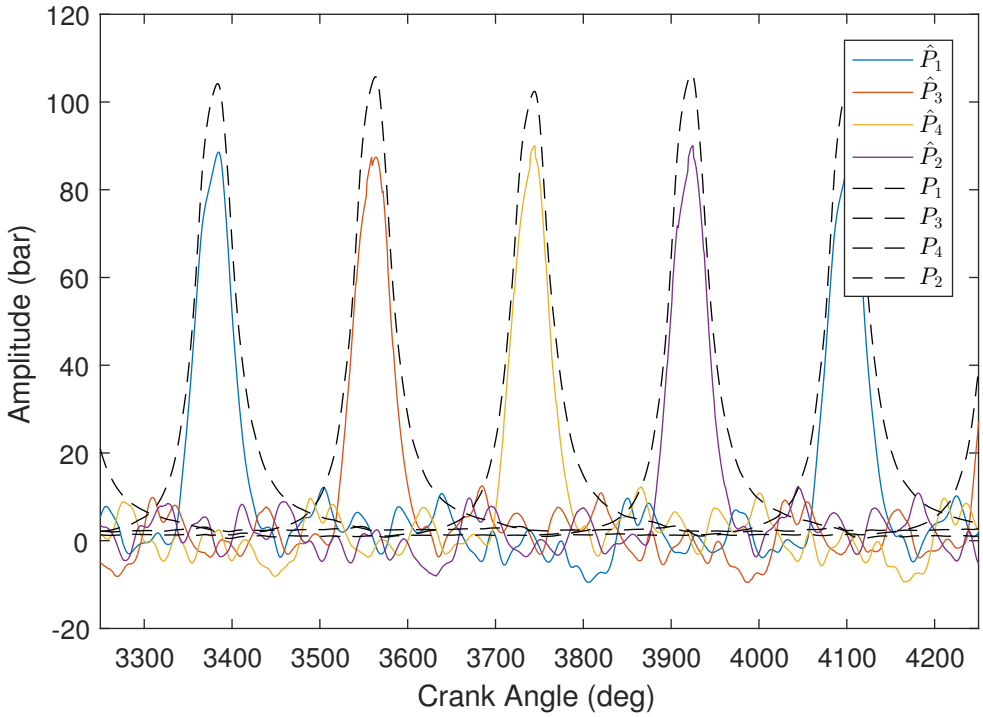

Figure 5.18: Cylinder pressure estimation under the stationary operating condition Z_8 (3000 rpm, 120 Nm) using Algorithm 5.1.

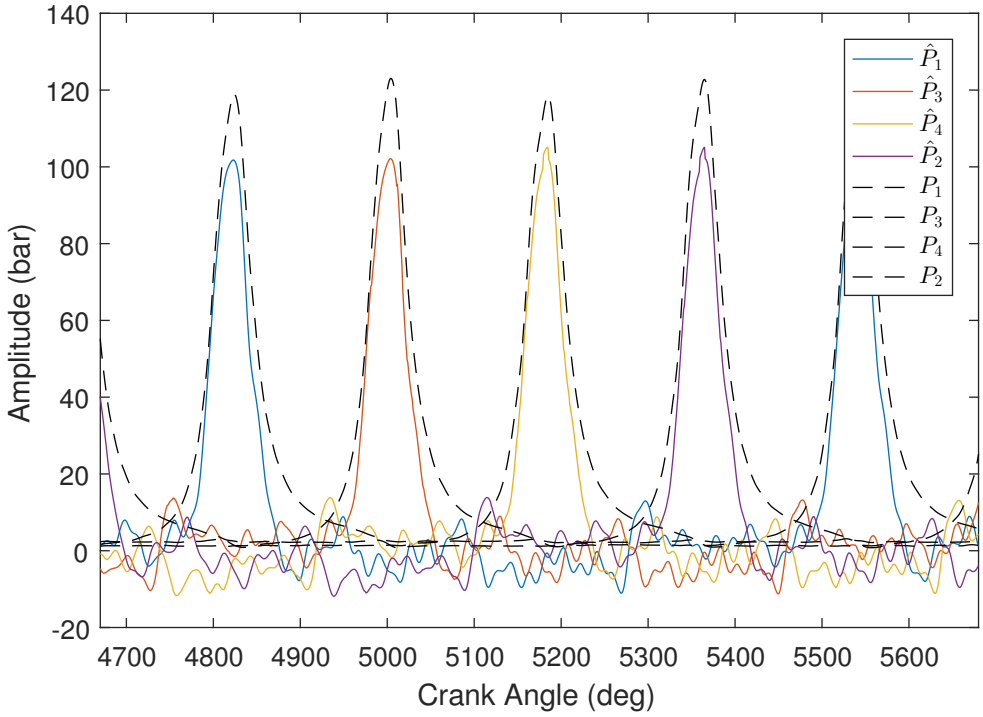

Figure 5.19: Cylinder pressure estimation under the stationary operating condition Z_9 (3000 rpm, 180 Nm) using Algorithm 5.1.

Table 5.2: Cylinder No. 1 pressure estimation results evaluation based on Algorithm 5.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	-4.29%	2.09%	13.40%	2.02%	-3.30%	-4.30%	21.89%	13.25%	13.63%
$\sigma(e_{P_{\max}})$	1.19%	1.31%	3.52%	1.65%	1.65%	1.31%	2.08%	2.62%	1.71%
$\mu(e_{P_{\mathrm{loc}}})$	$-0.62°$	$-1.91°$	$-1.83°$	$-5.13°$	$-0.90°$	$-1.22°$	$-3.13°$	$-0.03°$	$1.01°$
$\sigma(e_{P_{\mathrm{loc}}})$	$1.56°$	$0.40°$	$0.44°$	$4.11°$	$0.44°$	$0.50°$	$1.86°$	$0.91°$	$0.63°$

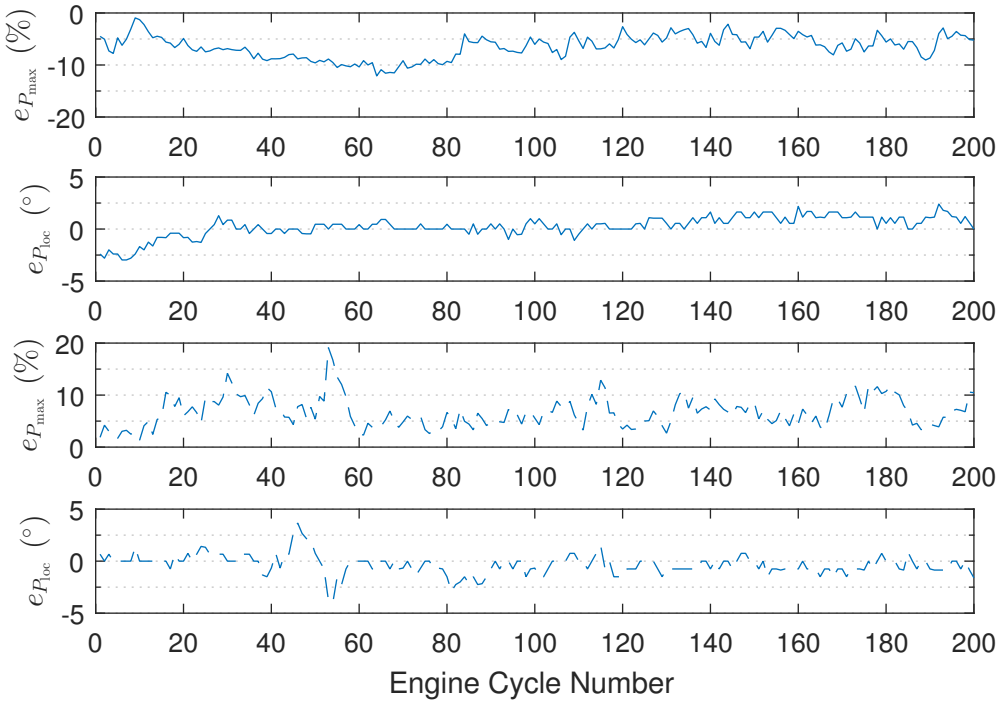

Figure 5.20: Cylinder No. 1 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\mathrm{t}}$ (lower two subfigures) using Algorithm 5.1.

To validate the effectiveness of the proposed cylinder pressure estimation method, two characteristic parameters, which are the relative error of pressure peak $P_{\max}$ and the absolute error of peak location P_{loc}, were used to evaluate estimation results. At mth peak, the two characteristic parameters can be represented as

$$e^m_{P_{\max}} = \frac{P^m_{\max} - \hat{P}^m_{\max}}{P^m_{\max}} \tag{5.39}$$

and

$$e^m_{P_{\mathrm{loc}}} = P^m_{\mathrm{loc}} - \hat{P}^m_{\mathrm{loc}}, \tag{5.40}$$

respectively.

Under nine stationary operating conditions, the average and standard deviation of the characteristic parameters are illustrated in Table 5.2. Specifically, $\mu(e_{P_{\max}})$ and $\sigma(e_{P_{\max}})$ represent the average and the standard deviation of the relative error of pressure peak corresponding to 50 engine cycles, respectively. While $\mu(e_{P_{\mathrm{loc}}})$ and $\sigma(e_{P_{\mathrm{loc}}})$ respectively denote the average and standard deviation of absolute error of peak location corresponding

to 50 engine cycles. But Table 5.2 is only for the cylinder No. 1, for other cylinders, see Table G.1, Table G.2, and Table G.3 in Appendix G.

According to the results summarized in the above tables, the value of $\mu(e_{P_{\max}})$ varies from -9.40% to 21.89%, but in general the absolute value of $\mu(e_{P_{\max}})$ is less than 14%. Additionally, in general the value of $\sigma(e_{P_{\max}})$ is less than 2%. For the peak location estimation, most of the absolute values of $\mu(e_{P_{\mathrm{loc}}})$ do not exceed $2.5°$, and almost all the values of $\sigma(e_{P_{\mathrm{loc}}})$ are less than $2°$.

While for non-stationary operating conditions (i.e., the transient process Z_{t}[1] and the transient process $\tilde{Z}_{\mathrm{t}}$[2]), the pressure peak and peak location estimation results for cylinder No. 1 is illustrated in Figure 5.20, and for other cylinders, see the figures from Figure G.1 to Figure G.3 in Appendix G. According to the results in the figures, it can be found that for lower instantaneous engine speeds (i.e., under the condition Z_{t}), in general, the absolute values of $e_{P_{\max}}$ do not exceed 10%, and the absolute values of $e_{P_{\mathrm{loc}}}$ are smaller than $5°$. As a comparison, for higher instantaneous engine speeds (i.e., under the condition $\tilde{Z}_{\mathrm{t}}$), generally the absolute values of $e_{P_{\max}}$ are less than 15%, while the absolute values of $e_{P_{\mathrm{loc}}}$ do not exceed $5°$.

5.4 Cylinder Pressure Estimation Results Based on Frequency-Domain Identified Model

In Section 5.3, the cylinder pressure estimation results based on the time-domain identified model are displayed. While, in this section, the estimation results based on the frequency-domain identified model are shown. Algorithm 5.1 was still used, however, the time-domain identified model was replaced with the frequency-domain identified model. The frequency-domain identified model is with 8th order and identified by using the IO data without detrending (see Section 4.3.2). Thus, for this case, the tunable parameters in Algorithm 5.1 are depicted in Table 5.3.

Table 5.3: Tunable parameters.

Parameter	Value
ν	35
l_{a}	10
n_{p}	40
σ_{a}^2	2.3×10^{-5}
λ	0.9995
$\hat{\boldsymbol{G}}$	frequency-domain identified model with 8th order

From Figure 5.21 to Figure 5.29, the curves of the estimated cylinder pressure are shown under nine different stationary operating conditions, i.e., from the operating condition Z_1 to the operating condition Z_9.

[1] Z_{t} is a part of the transient process between the stationary operating condition Z_3 and the stationary operating condition Z_4, i.e., Z_3 and Z_4 are not included.

[2] $\tilde{Z}_{\mathrm{t}}$ is a part of the transient process between the stationary operating condition Z_6 and the stationary operating condition Z_7, i.e., Z_6 and Z_7 are not included.

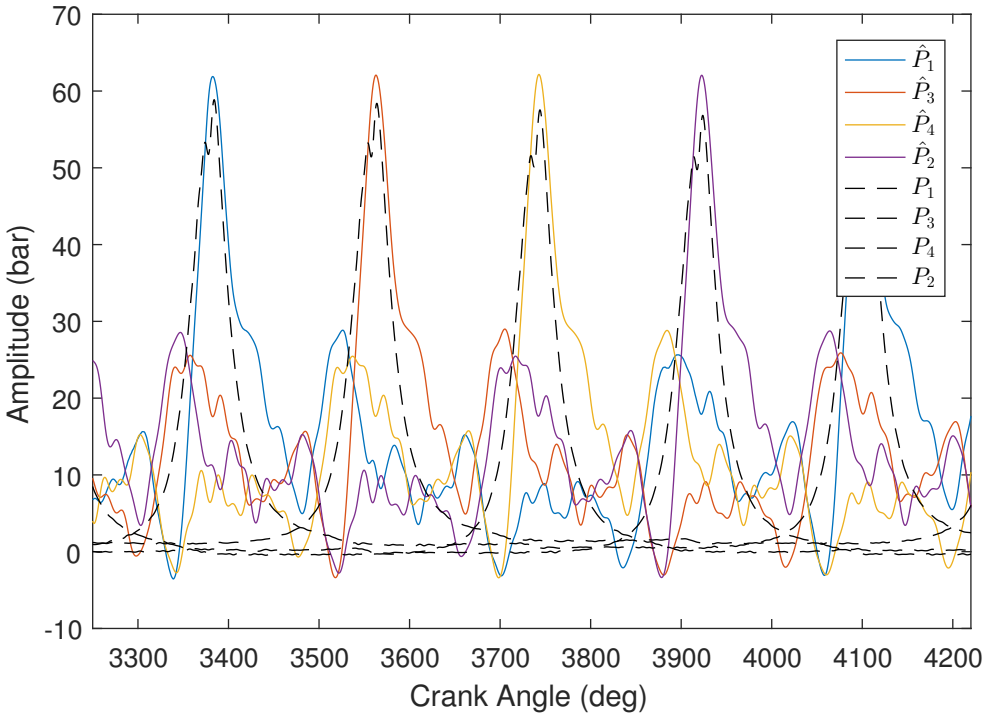

Figure 5.21: Cylinder pressure estimation under the stationary operating condition Z_1 (1200 rpm, 60 Nm) using Algorithm 5.1.

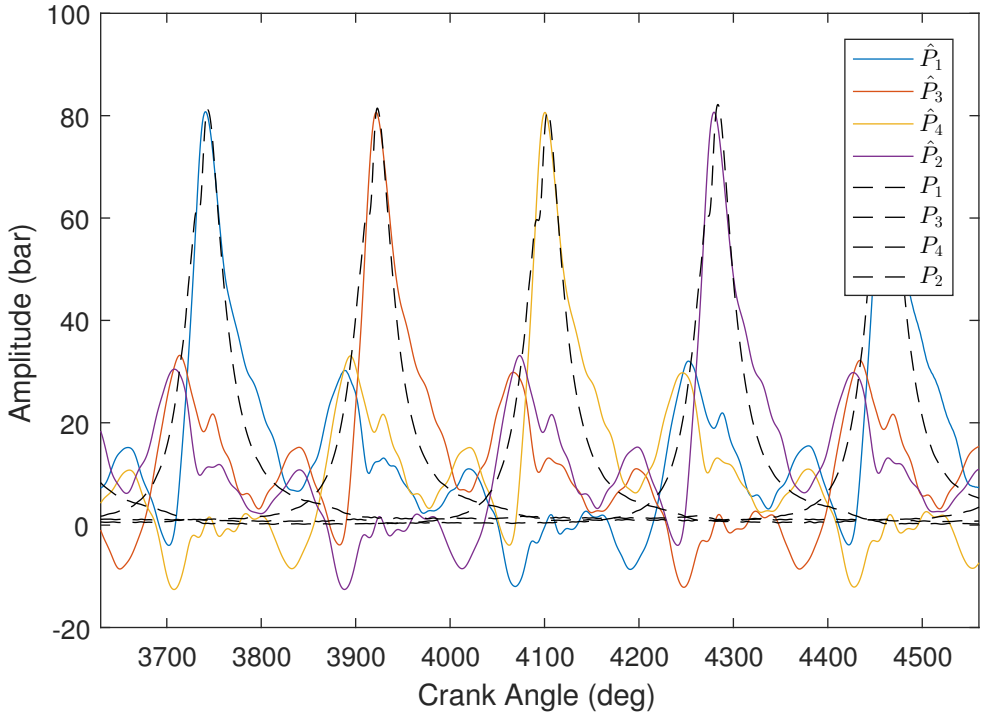

Figure 5.22: Cylinder pressure estimation under the stationary operating condition Z_2 (1200 rpm, 120 Nm) using Algorithm 5.1.

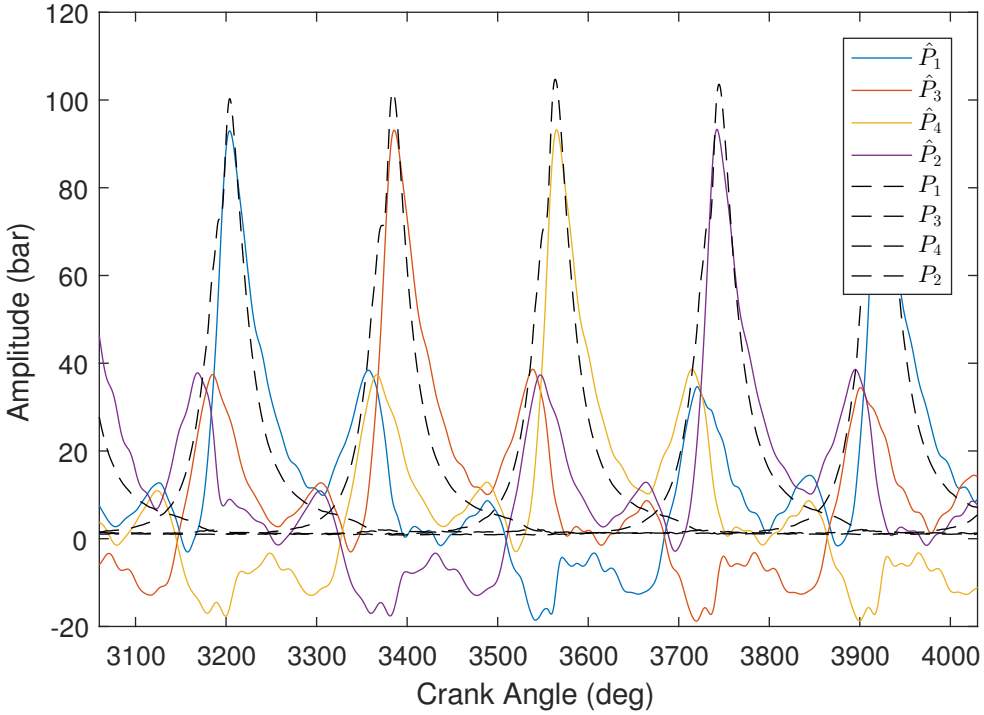

Figure 5.23: Cylinder pressure estimation under the stationary operating condition Z_3 (1200 rpm, 180 Nm) using Algorithm 5.1.

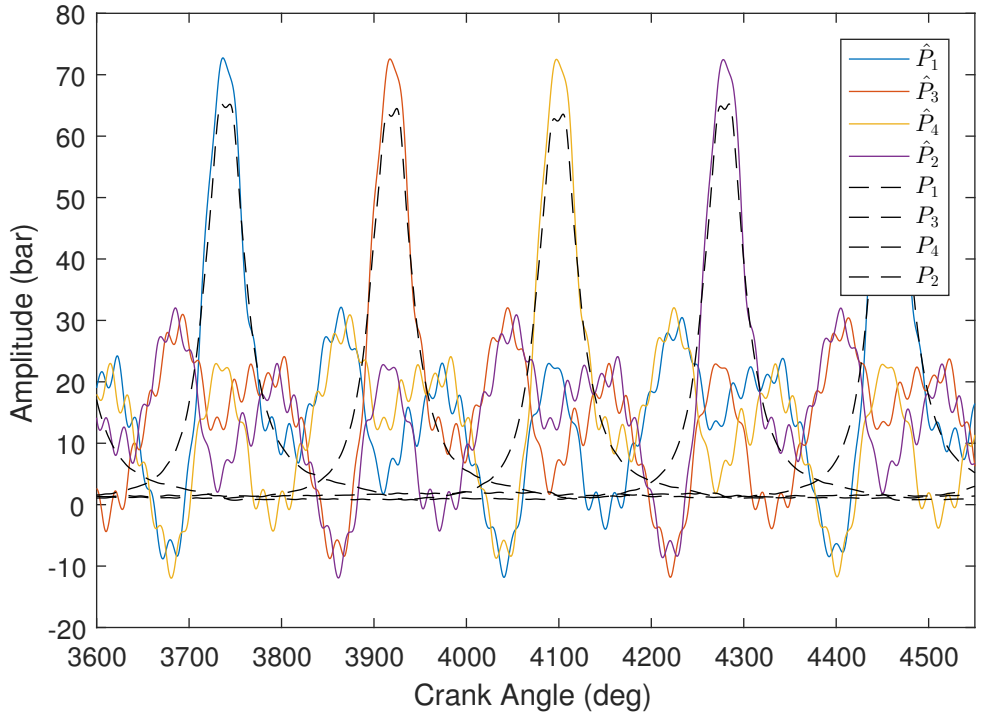

Figure 5.24: Cylinder pressure estimation under the stationary operating condition Z_4 (2100 rpm, 60 Nm) using Algorithm 5.1.

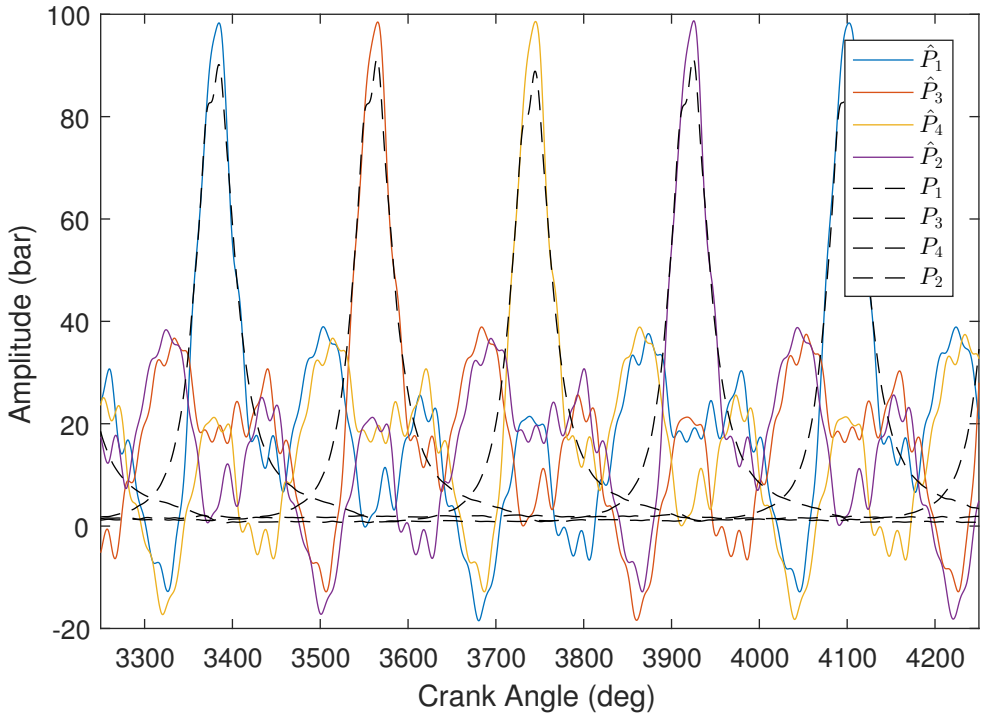

Figure 5.25: Cylinder pressure estimation under the stationary operating condition Z_5 (2100 rpm, 120 Nm) using Algorithm 5.1.

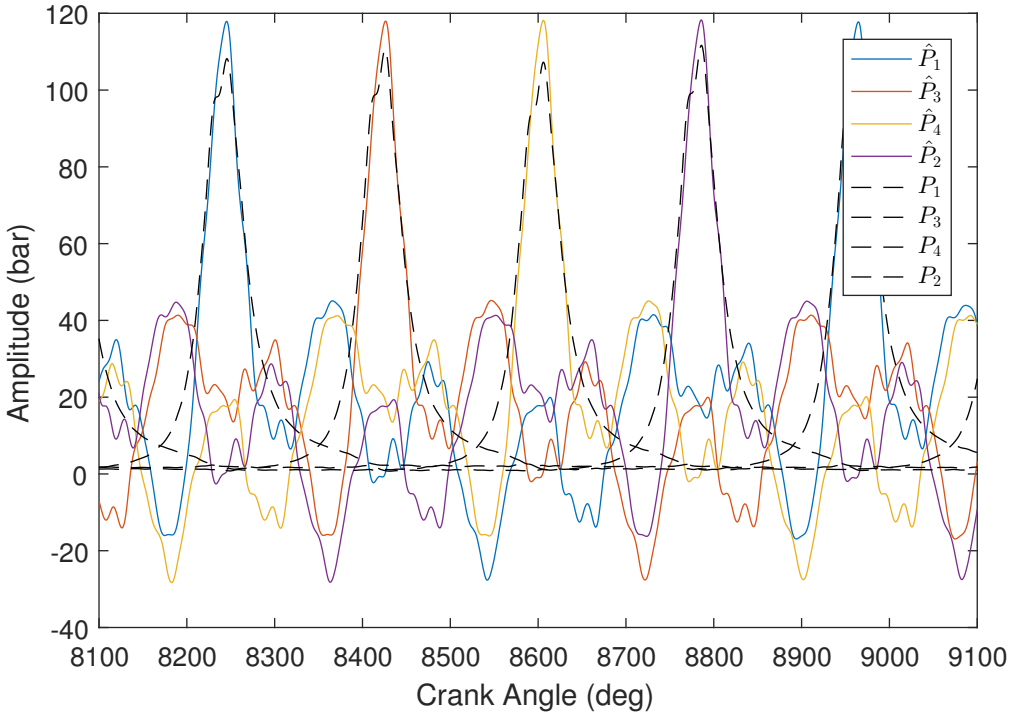

Figure 5.26: Cylinder pressure estimation under the stationary operating condition Z_6 (2100 rpm, 180 Nm) using Algorithm 5.1.

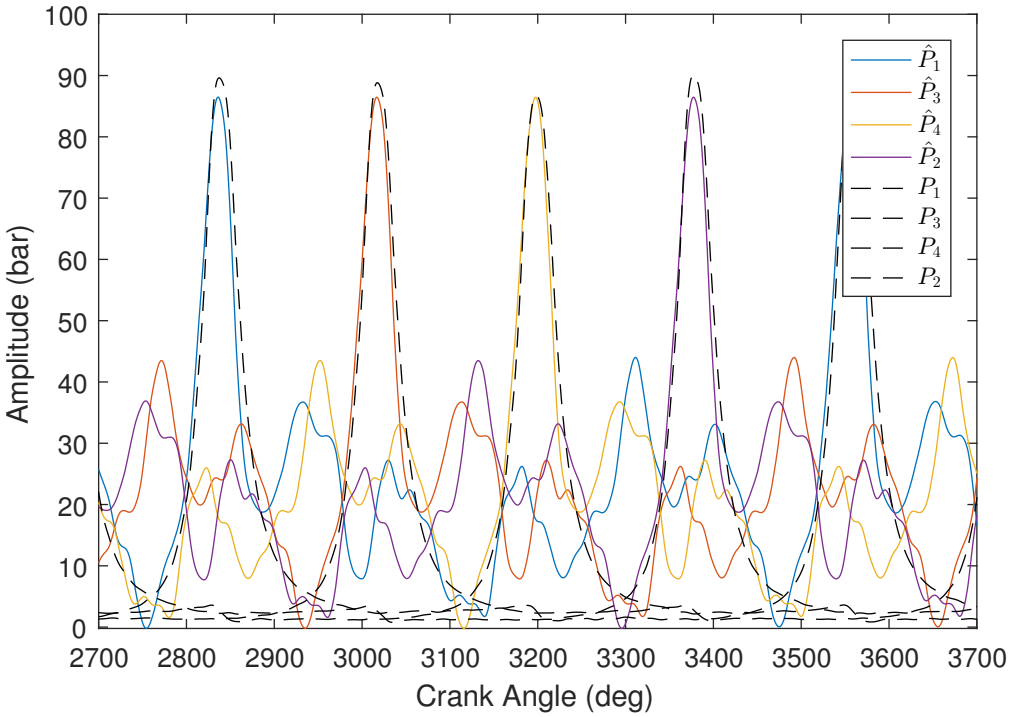

Figure 5.27: Cylinder pressure estimation under the stationary operating condition Z_7 (3000 rpm, 60 Nm) using Algorithm 5.1.

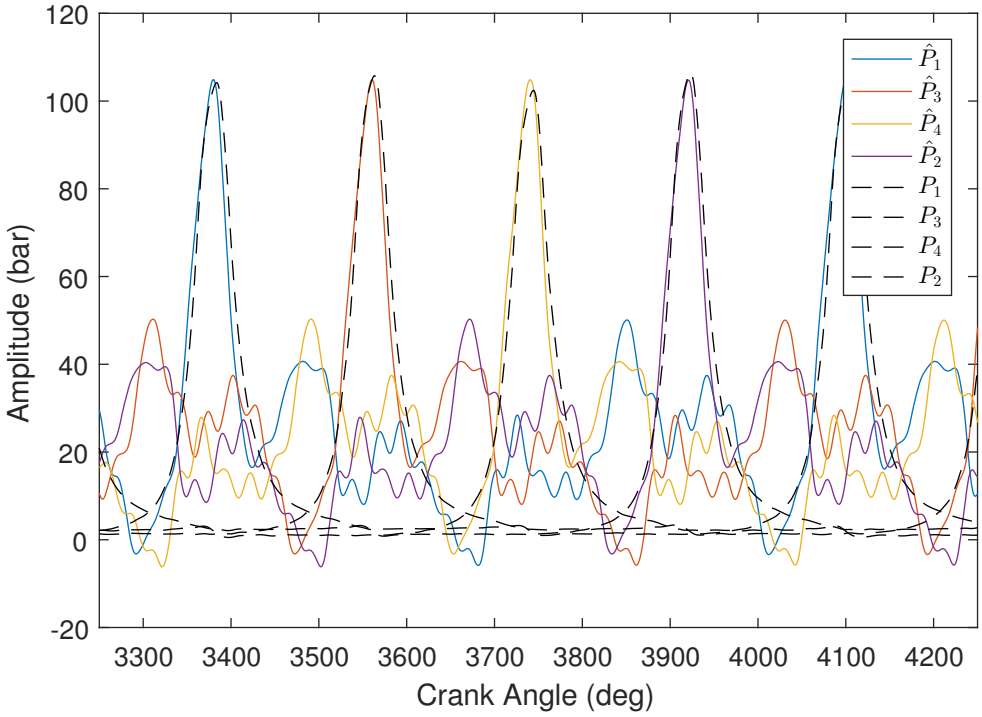

Figure 5.28: Cylinder pressure estimation under the stationary operating condition Z_8 (3000 rpm, 120 Nm) using Algorithm 5.1.

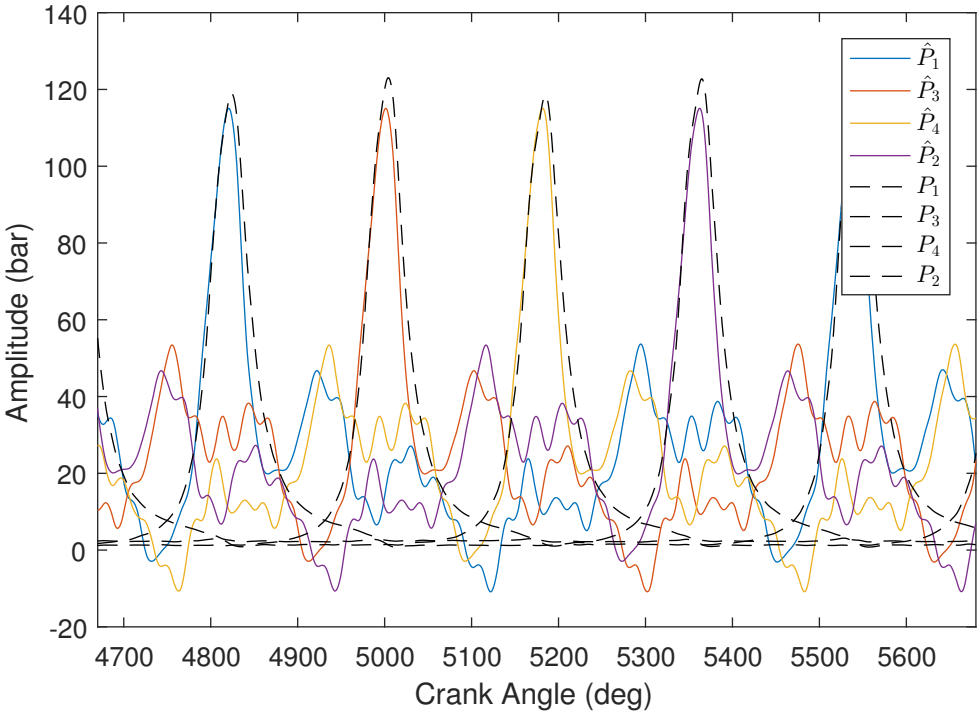

Figure 5.29: Cylinder pressure estimation under the stationary operating condition Z_9 (3000 rpm, 180 Nm) using Algorithm 5.1.

Table 5.4: Cylinder No. 1 pressure estimation results evaluation based on Algorithm 5.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	-6.32%	1.61%	8.15%	-10.73%	-9.19%	-8.53%	2.88%	-0.76%	3.47%
$\sigma(e_{P_{\max}})$	0.77%	1.04%	2.55%	0.41%	0.56%	0.55%	0.38%	0.39%	0.40%
$\mu(e_{P_{\mathrm{loc}}})$	$1.23°$	$-0.41°$	$-1.36°$	$-0.60°$	$0.98°$	$1.66°$	$1.15°$	$3.06°$	$3.34°$
$\sigma(e_{P_{\mathrm{loc}}})$	$1.37°$	$1.63°$	$2.03°$	$4.20°$	$0.79°$	$0.96°$	$0.60°$	$0.59°$	$0.56°$

Note: mean($e_{P_{\max}}$) denotes the average value of 50 cylinder pressure peak values.

Under nine stationary operating conditions, the mean value and standard deviation of the characteristic parameters are illustrated in Table 5.4 for the cylinder No. 1 pressure signal, and for other cylinders, see the tables from Table G.4 to Table G.6 in Appendix G. In the tables, generally, the absolute values of $\mu(e_{P_{\max}})$ do not exceed 10%, while most of the absolute values of $\mu(e_{P_{\mathrm{loc}}})$ are smaller than 2°. Furthermore, most of the standard deviation values of $e_{P_{\max}}$ are smaller than 1%, while for peak location, almost all the values of $e_{P_{\mathrm{loc}}}$ do not exceed 1.5°.

Under non-stationary operating conditions (i.e., the transient processes Z_{t} and $\tilde{Z}_{\mathrm{t}}$), as shown in Figure 5.30 (for the cylinder No. 1 pressure signal) and the figures from Figure G.4 to Figure G.6 in Appendix G (for other cylinders), at lower instantaneous engine speeds (i.e., under the condition Z_{t}), in general the absolute value of $e_{P_{\max}}$ varies from 0% to 10%, while for higher instantaneous engine speeds (i.e., under the condition $\tilde{Z}_{\mathrm{t}}$, generally most of the absolute values of $e_{P_{\max}}$ are smaller than 5%. For peak location estimation, generally most of the absolute values of $e_{P_{\mathrm{loc}}}$ at both lower instantaneous engine speeds and higher instantaneous engine speeds are smaller than 5°.

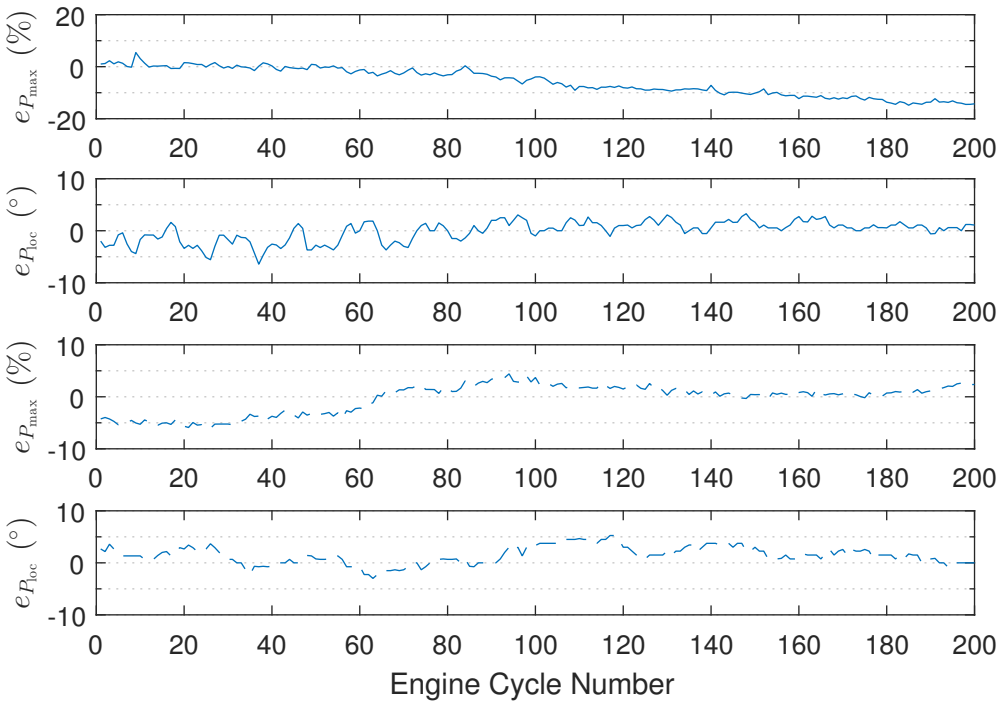

Figure 5.30: Cylinder No. 1 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\mathrm{t}}$ (lower two subfigures) using Algorithm 5.1.

By comparing the cylinder pressure estimation results in this section with the cylinder pressure estimation results in Section 5.3, it can be known that based on using the frequency-domain identified model, the combustion metrics pressure peak and peak location are better estimated in general, but the estimated cylinder pressure curves become worse, i.e., the estimated cylinder pressure curves based on using the frequency-domain identified model have bigger fluctuations at the bottom.

5.5 An Ideal Model-Based Cylinder Pressure Estimation

According to the cylinder pressure estimation results illustrated in Section 5.3 and Section 5.4, it can be known that the model quality can affect the estimation precision of the cylinder pressure signal, even though it is obvious that the Kalman filter performance can be affected by the model quality. So in this section, the performance of the cylinder pressure estimation method proposed in Section 5.2 is explored when the model $\boldsymbol{G}$ is ideal. The idea is depicted in Figure 5.31. Specifically, first use the collected input signal $\boldsymbol{u}(k) = \begin{pmatrix} P_1(k) & P_3(k) & P_4(k) & P_2(k) \end{pmatrix}^{\mathrm{T}}$ as the input of the time-domain identified model $\hat{\boldsymbol{G}}$, and then the corresponding simulated output signal $y_{\mathrm{s}}(k)$ can be obtained. Additionally, in Figure 5.31, the output noise sequence $\{e_{\mathrm{s}}(k)\}$ denotes a scalar white noise process, of which the covariance function is $\sigma_{\mathrm{s}}^2 \delta_{kj}$ with σ_{s}^2 known. Finally, with the simulated output $y_{\mathrm{s}}(k)$, four cylinder pressure signal can be recursively estimated. According to the description of the above idea, in the total process illustrated in Figure 5.31 the time-domain identified model $\hat{\boldsymbol{G}}$ can be seen as an ideal model, i.e., no model identification error.

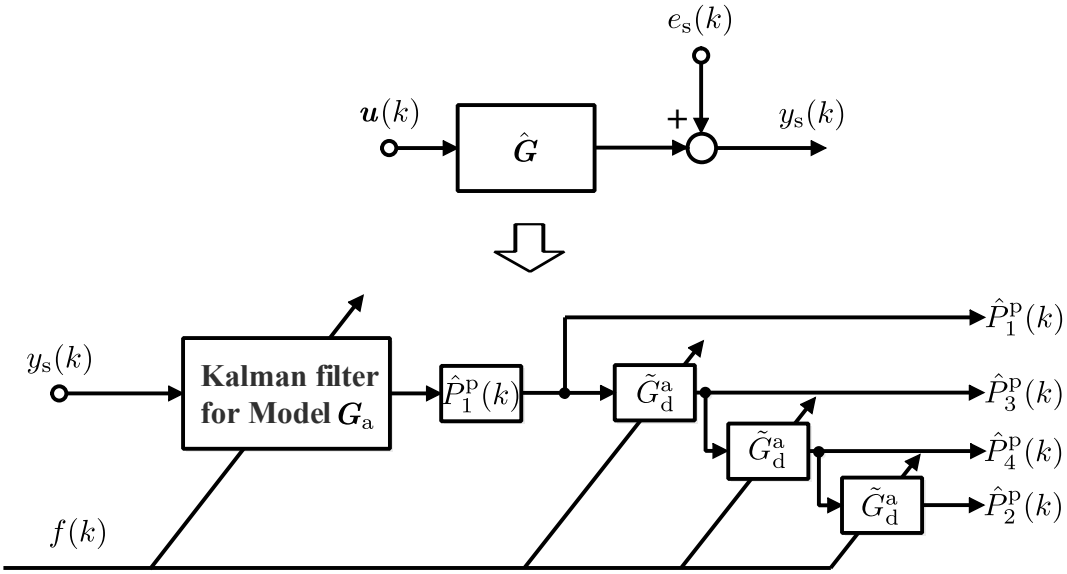

Figure 5.31: An ideal model-based method validation.

Based on the above process, the simulated output signals can be obtained under three stationary operating conditions Z_2 (1200 rpm, 120 Nm), Z_5 (2100 rpm, 120 Nm), and Z_8 (3000 rpm, 120 Nm), then with the simulated outputs, by implementing Algorithm 5.1 the estimated cylinder pressure signals can be obtained. The tunable parameters in Algorithm 5.1 for this case is given in Table 5.5.

Table 5.5: Tunable parameters.

Parameter	Value
ν	35
l_a	10
n_p	40
σ_a^2	1×10^{-5}
λ	0.9995
$\hat{\boldsymbol{G}}$	9th order and MOESP(s)

The final cylinder pressure estimation results under three stationary operating conditions are shown in Figure 5.32 and Table 5.6. According to the results, it can be known that under the three stationary operating conditions, the values of $\mu(e_{P_{\max}})$ are around 10%, and the values of $\mu(e_{P_{\text{loc}}})$ are around 2°. The standard deviation values $\sigma(e_{P_{\max}})$ and $\sigma(e_{P_{\text{loc}}})$ are smaller than 2.6% and 1.6°, respectively. Generally, the cylinder pressure curves estimated based on the ideal model are better than the curves estimated based on the time-domain identified model, especially at the bottom. The results in Table 5.6 can indicate that the quality of the model $\boldsymbol{G}$ is not the unique factor which can affect the cylinder pressure estimation precision.

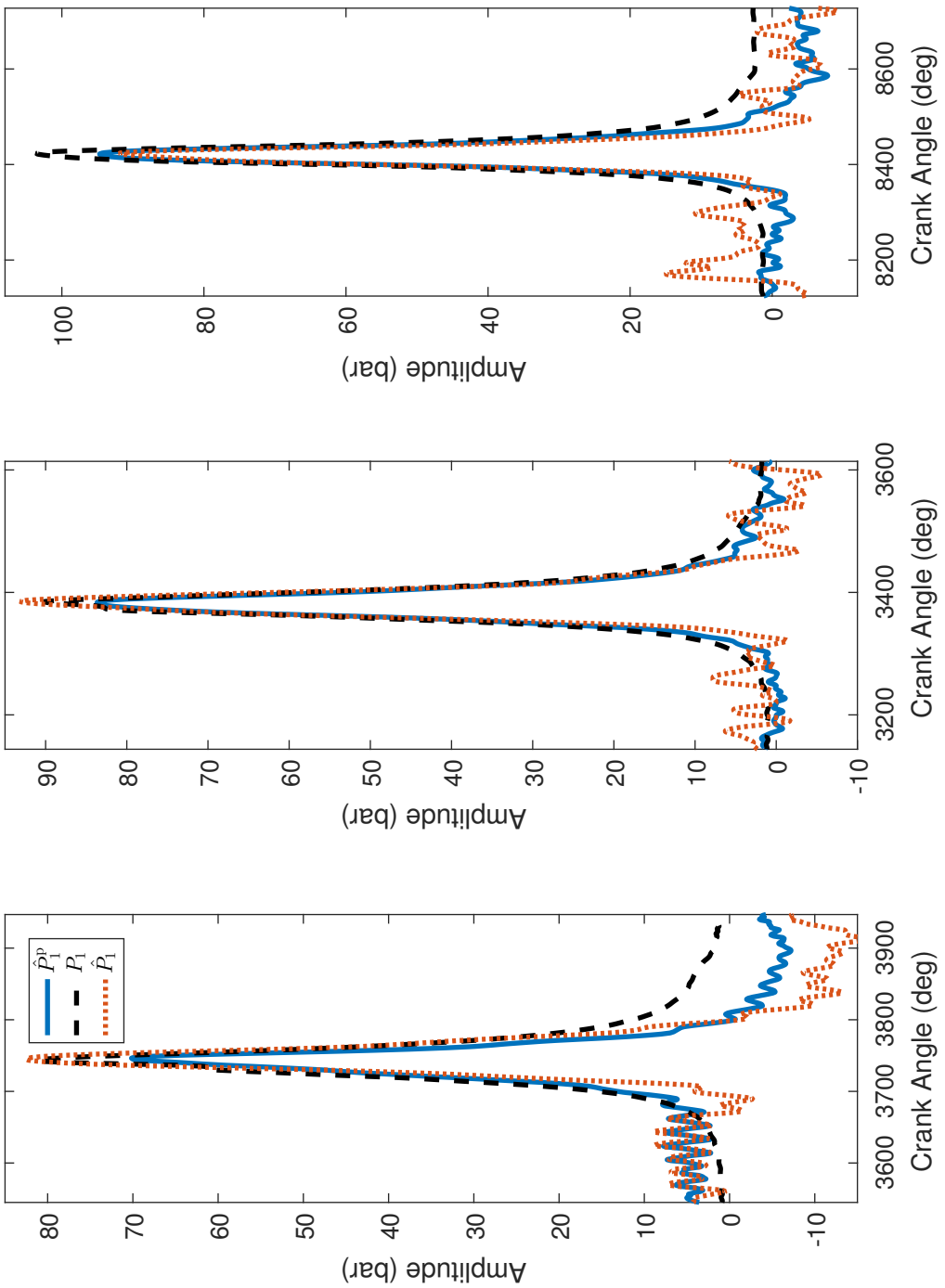

Figure 5.32: Cylinder No. 1 pressure estimation results under the conditions Z_2 (1200 rpm, 120 Nm), Z_5 (2100 rpm, 120 Nm), and Z_8 (3000 rpm, 120 Nm) using Algorithm 5.1 with an ideal model (the doted lines correspond to estimated results based on the time-domain identified model which are illustrated in Section 5.3).

Table 5.6: Cylinder No. 1 pressure estimation results evaluation based on Algorithm 5.1 (50 engine cycles for each operating condition).

	Z_2	Z_5	Z_8
$\mu(e_{P_{\max}})$	12.91% (2.09%)	7.01% (−3.30%)	9.44% (13.25%)
$\sigma(e_{P_{\max}})$	2.17% (1.31%)	1.48% (1.65%)	2.52% (2.62%)
$\mu(e_{P_{\mathrm{loc}}})$	−2.77° (−1.91°)	0.44° (−0.90°)	1.75° (−0.03°)
$\sigma(e_{P_{\mathrm{loc}}})$	0.33° (0.40°)	1.02° (0.44°)	1.58° (0.91°)

5.6 Summary

In this chapter, the cylinder pressure estimation results based on the time-domain identified model, the frequency-domain identified model, and the so-called ideal model are illustrated. It indicates that the model quality can affect the cylinder pressure estimation results. Generally, the estimated combustion metrics (pressure peak and peak location) based on the frequency-domain identified model are better than the ones estimated based on the time-domain identified model. But the estimated curves of the cylinder pressure signals are worse based on the frequency-domain identified model.

As mentioned in Section 5.5, there should be several kinds factors that affect the cylinder pressure estimation algorithm proposed in this chapter. The main factors are introduced as follows:

- *Modeling*
  Obviously, according to the results shown in the above sections, the identification accuracy of the model $\boldsymbol{G}$ has a big effect on the cylinder pressure estimation. Several factors can determine the model quality:

  (i) Inaccurate measurements, e.g., cylinder pressure transducers can be affected by the high temperature in cylinders.

  (ii) Inaccurate input or output number, i.e., the cylinder pressure is not the only one source causing engine structural vibration.

  (iii) Inaccurate model structure assumption, e.g., the linear assumption on $\boldsymbol{G}$ may lead to some errors if nonlinearity is severe.

- *TVDB*
  The use of the conceptual TVDB is based on the assumption that the shapes of the pressure signals of the four cylinders are the same in every engine cycle which is not strictly true because of some factors such as cylinder-to-cylinder injection and air-EGR dispersions, so this assumption can lead to some errors in estimation. Furthermore, the Padé approximation and numerical discretization of the conceptual continuous-time TVDB can also bring some additional errors (see Equation (5.28)).

- *Instantaneous engine cycle frequency calculation*
  The instantaneous engine cycle frequency is used to calculate the time-variant dead time of TVDB, so the calculation of the instantaneous engine cycle frequency has

an effect on the cylinder pressure estimation. It should be noted that in the cylinder pressure estimation process based on the frequency-domain identified model, the engine cycle frequency $f(k)$ was assumed to be constant in a big angle range (around 540°, which was chosen based on practical situation), the reason for using the big angle range rather than using 6° (and 12° for the reference marker) in the estimation process based on the time-domain identified model is that in LPM, for each stationary operating condition, only one fundamental frequency is used, which is not true in practice, so a more smooth curve of calculated instantaneous engine speed is needed. In this thesis, only for the cylinder pressure estimation based on the frequency-domain identified model a big angle range was used, for other cases 6° (and 12° for the reference marker) are used in the calculation process of the frequency $f(k)$.

Chapter 6

Linear Model-Based Cylinder Pressure Estimation Merely Using Engine Structural Vibration Signal

6.1 Introduction

In Chapter 5, the cylinder pressure estimation is based on the known IAS, however, sometimes the engine crank angular speed information maybe unknown. Considering the above problem, in this chapter, the problem when only vibration signal is available for the cylinder pressure estimation is handled [85]. Thus different with the cylinder pressure estimation method proposed in the previous chapter, the cylinder pressure reconstruction by merely using vibration signal is proposed in this chapter. This chapter is organized as follows. In Section 6.2, the proposed cylinder pressure estimation algorithm is illustrated, and then the corresponding cylinder pressure estimation results are shown in Section 6.3, and finally, in Section 6.4, a summary is given.

6.2 Cylinder Pressure Estimation Algorithm

In this section, the proposed method of cylinder pressure estimation by merely using engine structural vibration is illustrated, and in Figure 6.1 the idea behind the proposed method is depicted. Compared with the cylinder pressure estimation framework in Figure 5.1, the vibration signal and an estimator E_{f} are first used to estimate the instantaneous engine cycle frequency $f(k)$ instead of using a crank angular speed sensor to determine the instantaneous engine cycle frequency. Then with the vibration signal and the estimated instantaneous engine cycle frequency $\hat{f}(k)$, the same cylinder pressure estimation method proposed in Chapter 5 can be used to reconstruct the cylinder pressure signals. Below, the problem of the estimation of the instantaneous engine cycle frequency $f(k)$ by using the vibration signal is firstly formulated, and afterwards the derivation of the cylinder pressure estimation algorithm using merely engine vibration signal is given.

6.2.1 Estimation of Instantaneous Engine Cycle Frequency

The vibration signal $y(k)$ is mainly induced by the cylinder pressure which can be considered as a combination of the DC offset $A_0(k)$, a certain number of order (or say spectral)

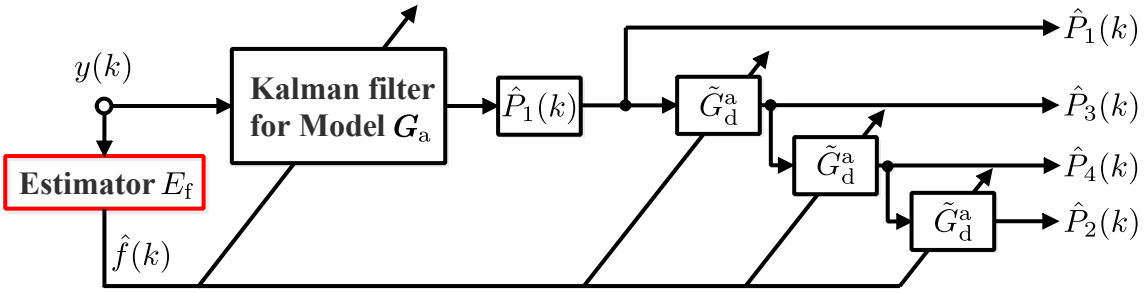

Figure 6.1: Cylinder pressure estimation merely using engine vibration signal.

components $y_i(k), i = 1, 2, \cdots, n_\mathrm{y}$ (see [52]) and an error term $e_\mathrm{y}(k)$, i.e.,

$$\begin{aligned} y(k) &= A_0(k) + \sum_{i=1}^{n_\mathrm{y}} y_i(k) + e_\mathrm{y}(k), \\ &= A_0(k) + \sum_{i=1}^{n_\mathrm{y}} A_i(k)\,\Theta_i(k) + e_\mathrm{y}(k), \end{aligned} \tag{6.1}$$

where n_y is the number of the order components of interest. $A_i(k)$ refers to the amplitude envelope of the l_ith order component, and $\Theta_i(k)$ denotes the carrier signal [33], which can be represented as

$$\Theta_i(k) = \sin(\varphi_i(k)) = \sin(l_i \sum_{j=0}^{k} 2\pi f(j) T_\mathrm{s} + \varphi_i(0)), \tag{6.2}$$

where $\varphi_i(k)$ denotes the instantaneous phase and $\varphi_i(0)$ is the initial phase of the l_ith order component.

Similar to the cylinder pressure signal modeling as illustrated in Equation (5.3), the vibration signal $y(k)$ in Equation (6.1) can also be denoted as the output of a state-space model when $A_i(k)$ is constant, i.e.,

$$\begin{cases} \boldsymbol{x}_\mathrm{y}(k+1) = \boldsymbol{A}_\mathrm{y}(f(k))\boldsymbol{x}_\mathrm{y}(k), \\ y(k) = \boldsymbol{C}_\mathrm{y}\boldsymbol{x}_\mathrm{y}(k) + e_\mathrm{y}(k), \end{cases} \tag{6.3}$$

where the state vector $\boldsymbol{x}_\mathrm{y}(k) \in \mathbb{R}^{2n_\mathrm{y}+1}$.

The matrices $\boldsymbol{A}_\mathrm{y}(k)$ and $\boldsymbol{C}_\mathrm{y}$ are given as

$$\boldsymbol{A}_\mathrm{y}(f(k)) = \left(\begin{array}{c|ccc} 1 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{A}_1(f(k)) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{A}_{n_\mathrm{y}}(f(k)) \end{array}\right), \tag{6.4}$$

and

$$\boldsymbol{C}_\mathrm{y} = \left(1 \;\vdots\; \boldsymbol{C}_1 \quad \cdots \quad \boldsymbol{C}_{n_\mathrm{y}}\right), \tag{6.5}$$

respectively.

The individual block entries in these block matrices follow from the state-space representation of a frequency-modulated sine wave as

$$\boldsymbol{A}_i(f(k)) = \begin{pmatrix} \cos(2\pi i f(k) T_\mathrm{s}) & \sin(2\pi i f(k) T_\mathrm{s}) \\ -\sin(2\pi i f(k) T_\mathrm{s}) & \cos(2\pi i f(k) T_\mathrm{s}) \end{pmatrix} \tag{6.6}$$

and

$$\boldsymbol{C}_i = \begin{pmatrix} 1 & 0 \end{pmatrix}. \tag{6.7}$$

When the frequency $f(k)$ is unknown, the remainder of this section is used to illustrated how to estimate the instantaneous engine cycle frequency $f(k)$. But firstly both the amplitudes A_i, $i = 1, 2, \cdots, n_\mathrm{y}$ and the frequency $f(k)$ are assumed to be time-invariant.

According to the frequency tracker introduced in [33], the unknown frequency $f(k)$ can be regarded as a state variable $x_\mathrm{f}(k)$, and then it can be augmented with the state of the model (6.3), afterwards the following model can be derived:

$$\begin{cases} \begin{pmatrix} \boldsymbol{x}_\mathrm{y}(k+1) \\ x_\mathrm{f}(k+1) \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_\mathrm{y}(x_\mathrm{f}(k)) & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_\mathrm{y}(k) \\ x_\mathrm{f}(k) \end{pmatrix}, \\ y(k) = \begin{pmatrix} \boldsymbol{C}_\mathrm{y} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_\mathrm{y}(k) \\ x_\mathrm{f}(k) \end{pmatrix} + e_\mathrm{y}(k). \end{cases} \tag{6.8}$$

The model (6.8) is a nonlinear model, so nonlinear Kalman filters can be used for the estimation of the frequency $f(k)$ (i.e., $x_\mathrm{f}(k)$). But for the frequency estimation purpose, only knowing the first order of the signal $y(k)$ is enough for us to track the frequency. Thus a band-pass filter can be used for filtering out the orders not of interest in the signal $y(k)$, and the signal $y(k)$ after filtering is denoted as $y_\mathrm{b}(k)$. For the design of the band-pass filter, the linear phase is not necessary because the objective is to merely estimate the frequency $f(k)$, therefore IIR filters can be used instead of using FIR filters to avoid big dimensional problem. An IIR filter such as the Butterworth bandpass filter can be included. The lowest engine speed and the highest engine speed used for the cylinder pressure reconstruction study in this thesis are around 1200 rpm and around 3000 rpm, respectively, so the lower cutoff frequency and the higher cutoff frequency of the Butterworth bandpass filter were respectively set to 10 Hz and 25 Hz. As can be seen, the highest cutoff frequency is more than two times as large as the lowest cutoff frequency, so the filtered signal $y_\mathrm{b}(k)$ can be represented as the output of the following state-space model with an output error:

$$\begin{cases} \boldsymbol{x}_\mathrm{b}(k+1) = \boldsymbol{A}_\mathrm{b}(f(k))\boldsymbol{x}_\mathrm{b}(k), \\ y_\mathrm{b}(k) = \boldsymbol{C}_\mathrm{b}\boldsymbol{x}_\mathrm{b}(k) + e_\mathrm{b}(k), \end{cases} \tag{6.9}$$

where $\boldsymbol{A}_\mathrm{b}(f(k))$ and $\boldsymbol{C}_\mathrm{b}$ are respectively denoted as

$$\boldsymbol{A}_\mathrm{b}(f(k)) = \begin{pmatrix} \boldsymbol{A}_1(f(k)) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2(f(k)) \end{pmatrix}, \tag{6.10}$$

$$\boldsymbol{C}_\mathrm{b} = \begin{pmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 \end{pmatrix}, \tag{6.11}$$

and $e_\mathrm{b}(k)$ represents the output error after band-pass filtering.

Similar to the model (6.8), augmenting the state variable $x_{\mathrm{f}}(k)$ with the state of the model (6.9) and replacing $f(k)$ with the state variable $x_{\mathrm{f}}(k)$ result in the following augmented state-space model:

$$\begin{cases} \boldsymbol{x}_{\mathrm{e}}(k+1) = f_{\mathrm{e}}(\boldsymbol{x}_{\mathrm{e}}(k)), \\ y_{\mathrm{b}}(k) = \boldsymbol{C}_{\mathrm{e}}\boldsymbol{x}_{\mathrm{e}}(k) + e_{\mathrm{b}}(k). \end{cases} \tag{6.12}$$

In the model (6.12), the expressions of $\boldsymbol{x}_{\mathrm{e}}(k)$, $f_{\mathrm{e}}(\boldsymbol{x}_{\mathrm{e}}(k))$, and $\boldsymbol{C}_{\mathrm{e}}$ are

$$\boldsymbol{x}_{\mathrm{e}}(k) = \begin{pmatrix} \boldsymbol{x}_{\mathrm{b}}(k) \\ x_{\mathrm{f}}(k) \end{pmatrix}, \tag{6.13}$$

$$f_{\mathrm{e}}(\boldsymbol{x}_{\mathrm{e}}(k)) = \begin{pmatrix} \boldsymbol{A}_{\mathrm{b}}(x_{\mathrm{f}}(k)) & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{pmatrix}, \tag{6.14}$$

and

$$\boldsymbol{C}_{\mathrm{e}} = \begin{pmatrix} \boldsymbol{C}_{\mathrm{b}} & 0 \end{pmatrix}, \tag{6.15}$$

respectively.

The model (6.12) is a nonlinear model, so the EKF introduced in Section 2.4.3 can be used for the estimation of the frequency $f(k)$ (i.e., the state variable $x_{\mathrm{f}}(k)$).

However, the formulation of the model (6.12) is based on the assumption that the amplitude A_1 of the first order in the signal $y_{\mathrm{b}}(k)$ and the frequency $f(k)$ are time-invariant. When the amplitude $A_1(k)$ and the frequency $f(k)$ are time-variant, there will be a mismodeling problem for the modeling of the model (6.12), thus the term $\boldsymbol{Q}(k)$ should be included in EKF, and increasing $\boldsymbol{Q}(k)$ can handle the problem of mismodeling, which is illustrated in Section 2.2.4.1. Here, the term $\boldsymbol{Q}(k)$ is involved and denote it as $\boldsymbol{Q}_{\mathrm{b}}$ in the EKF for the model (6.12), and the value of $\boldsymbol{Q}_{\mathrm{b}}$ is tunable. Furthermore, in the model (6.12), $\{e_{\mathrm{b}}(k)\}$ is assumed to be a scalar white noise process, of which the covariance function is $R_{\mathrm{b}}\delta_{kj}$, and the value of R_{b} is tunable. $e_{\mathrm{b}}(k)$ is also assumed to be uncorrelated with $\boldsymbol{x}_{\mathrm{e}}(0)$ which is assumed to be a random variable with constant mean and variance.

The above frequency tracking process can be represented by a frequency estimator E_{f} which consists of EKF and a band-pass filter. The frequency estimator E_{f} is depicted in Figure 6.2.

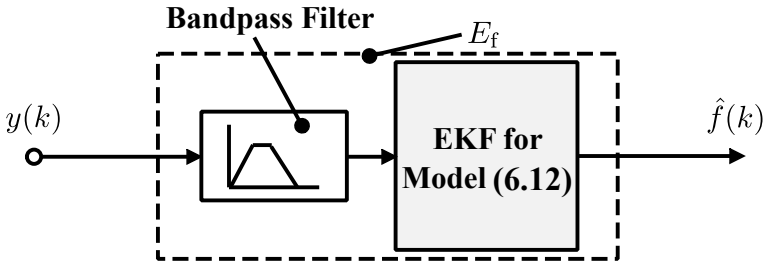

Figure 6.2: Proposed estimator E_{f} for frequency tracking.

There is an initialization problem for the EKF for the instantaneous engine cycle frequency estimation, because in the matrix $\boldsymbol{A}_{\mathrm{b}}(x_{\mathrm{f}}(k))$ in (6.14) there are two blocks

$\boldsymbol{A}_1(x_\mathrm{f}(k))$ and $\boldsymbol{A}_2(x_\mathrm{f}(k))$ corresponding to $f(k)$ and $2f(k)$, respectively, so the initial value of the frequency $f(k)$ in EKF should be near the tracked frequency, if not there will be an initialization problem. As an example, if the instantaneous engine cycle frequency to be estimated is 25 Hz, and if the initial value of the frequency is set to 10 Hz, EKF will finally obtain the value around 12.5 Hz as the estimated instantaneous engine cycle frequency. So in the instantaneous engine cycle frequency estimation algorithm a strategy for solving the initialization problem should be included.

Above all, the instantaneous engine cycle frequency estimation algorithm is summarized in Algorithm 6.1.

Algorithm 6.1: Instantaneous engine cycle frequency estimation.

Initialization: $\boldsymbol{P}_\mathrm{e}(0) = 10\boldsymbol{I}_5$, $\hat{\boldsymbol{x}}_\mathrm{e}(0) = 10\begin{pmatrix}\boldsymbol{0}\\1\end{pmatrix}$

for $k = 1 \to N-1$ **do**

  /* Extended Kalman filter */

  $\boldsymbol{P}_\mathrm{e}^-(k) = \boldsymbol{A}_\mathrm{e}(k-1)\boldsymbol{P}_\mathrm{e}(k-1)\boldsymbol{A}_\mathrm{e}^\mathrm{T}(k-1) + \boldsymbol{Q}_\mathrm{b}$ // $\boldsymbol{A}_\mathrm{e}(k) \approx \left.\frac{\partial f_\mathrm{e}(\boldsymbol{x}_\mathrm{e}(k))}{\partial \boldsymbol{x}_\mathrm{e}(k)}\right|_{\boldsymbol{x}_\mathrm{e}(k)=\hat{\boldsymbol{x}}_\mathrm{e}(k)}$

  $\hat{\boldsymbol{x}}_\mathrm{e}^-(k) = f_\mathrm{e}(\hat{\boldsymbol{x}}_\mathrm{e}(k-1))$

  $\boldsymbol{K}_\mathrm{e}(k) = \boldsymbol{P}_\mathrm{e}^-(k)\boldsymbol{C}_\mathrm{e}^\mathrm{T}\left(\boldsymbol{C}_\mathrm{e}\boldsymbol{P}_\mathrm{e}^-(k)\boldsymbol{C}_\mathrm{e}^\mathrm{T} + R_\mathrm{b}\right)^{-1}$

  $\boldsymbol{P}_\mathrm{e}(k) = \left(\boldsymbol{I}_5 - \boldsymbol{K}_\mathrm{e}(k)\boldsymbol{C}_\mathrm{e}\right)\boldsymbol{P}_\mathrm{e}^-(k)$

  $\hat{\boldsymbol{x}}_\mathrm{e}(k) = \hat{\boldsymbol{x}}_\mathrm{e}^-(k) + \boldsymbol{K}_\mathrm{e}(k)\left(y_\mathrm{b}(k) - \boldsymbol{C}_\mathrm{e}\hat{\boldsymbol{x}}_\mathrm{e}^-(k)\right)$

  /* Instantaneous engine cycle frequency estimation */

  $R(k) = y_\mathrm{b}(k) - \begin{pmatrix}0 & 0 & 1 & 0 & 0\end{pmatrix}\boldsymbol{x}_\mathrm{e}(k)$

  if $k \geq k_1$ **then**

    if $\sigma(R) \leq h$ // $\sigma(R)$ denotes the standard deviation of the sequence $(R(k))_{k=1}^{k_1-1}$, and h represent a threshold

    then

      $\hat{f}(k) = 2\hat{x}_\mathrm{f}(k)$

    else

      $\hat{f}(k) = \hat{x}_\mathrm{f}(k)$

    end

  end

end

As seen in the algorithm, the threshold h can be set based on observing the FFT spectrums of the vibration signal under different stationary operating conditions and the value of $R(k)$. For example, if the instantaneous engine cycle frequency to be estimated is 25 Hz, and if the initial value of the frequency is set to 10 Hz, without any strategy EKF will obtain the value around 12.5 Hz as the estimated instantaneous engine cycle frequency, and in this case $R(k)$ should be very small because the main part of $y_\mathrm{b}(k)$ corresponds to the frequency 25 Hz. Given another example, if the instantaneous engine cycle frequency to be estimated is 10 Hz, and if the initial value of the frequency is set to 10 Hz, without any strategy EKF will obtain the value around 10 Hz as the estimated instantaneous engine cycle frequency, and for this case, the main part of $y_\mathrm{b}(k)$ corresponds to 10 Hz such that the value of $R(k)$ is larger. So based on setting the threshold h, the initialization problem of the EKF in Algorithm 6.1 can be handled.

6.2.2 Cylinder Pressure Estimation

Replacing the estimated instantaneous engine cycle frequency $\hat{f}(k)$ with the calculated frequency $f(k)$ in the model (5.35) results in the following model:

$$\begin{cases} \boldsymbol{x}_{\mathrm{a}}(k+1) = \boldsymbol{A}_{\mathrm{a}}(\hat{f}(k))\boldsymbol{x}_{\mathrm{a}}(k), \\ y(k) = \boldsymbol{C}_{\mathrm{a}}(\hat{f}(k))\boldsymbol{x}_{\mathrm{a}}(k)) + e_{\mathrm{e}}(k), \end{cases} \tag{6.16}$$

where the output error $e_{\mathrm{e}}(k)$ is induced by $e_{\mathrm{a}}(k)$ and the difference between $f(k)$ and $\hat{f}(k)$.

In the model (6.16), $\{e_{\mathrm{e}}(k)\}$ is assumed to be a scalar white noise process, of which the covariance function is $\sigma_{\mathrm{a}}^2\delta_{kj}$, and the value of σ_{a}^2 is tunable. Additionally, $e_{\mathrm{e}}(k)$ is also assumed to be uncorrelated with $\boldsymbol{x}_{\mathrm{a}}(0)$ which is assumed to be a random variable with constant mean and variance. Then the square-root Kalman filter with a forgetting factor for the model (6.16) can be used to estimate the cylinder pressure signals as illustrated in Algorithm 5.1. As a summary, for the cylinder pressure estimation by merely using the engine vibration signal, Algorithm 6.1 is firstly used to estimate the frequency $f(k)$, and then Algorithm 5.1 is implemented. The cylinder pressure reconstruction method by merely using the engine structural vibration signal is summarized in Algorithm 6.2.

Algorithm 6.2: Cylinder pressure estimation algorithm merely using engine structural vibration signal.

Initialization: $\boldsymbol{S}(0) = 100\boldsymbol{I}_{n_{\mathrm{a}}}$, $\hat{\boldsymbol{x}}_{\mathrm{a}}(0) = \boldsymbol{0}$

for $k = 1 \to N-1$ **do**

`/* Instantaneous engine cycle frequency estimation */`

$\hat{x}_{\mathrm{f}}(k) \to \hat{f}(k)$ `// See Algorithm 6.1 for the estimation of` $x_{\mathrm{f}}(k)$

`/* Square-root Kalman filter with a forgetting factor */`

$\boldsymbol{S}^{-}(k) = \frac{1}{\lambda}\boldsymbol{A}_{\mathrm{a}}(\hat{f}(k-1))\boldsymbol{S}(k-1)$ `// λ denotes the forgetting factor`

$\hat{\boldsymbol{x}}_{\mathrm{a}}^{-}(k) = \boldsymbol{A}_{\mathrm{a}}(\hat{f}(k-1))\hat{\boldsymbol{x}}_{\mathrm{a}}(k-1)$

$\boldsymbol{F}(k) = \boldsymbol{S}^{-}(k)\boldsymbol{C}_{\mathrm{a}}^{\mathrm{T}}(\hat{f}(k))$

$\alpha(k) = \frac{1}{\boldsymbol{F}^{\mathrm{T}}(k)\boldsymbol{F}(k)+\sigma_{\mathrm{e}}^2}$

$\boldsymbol{K}(k) = \alpha(k)\boldsymbol{S}^{-}(k)\boldsymbol{F}(k)$

$\gamma(k) = \frac{1}{1+\sqrt{\sigma_{\mathrm{e}}^2}\sqrt{\alpha(k)}}$

$\boldsymbol{S}(k) = \boldsymbol{S}^{-}(k) - \gamma(k)\boldsymbol{K}(k)\boldsymbol{F}^{\mathrm{T}}(k)$

$\hat{\boldsymbol{x}}_{\mathrm{a}}(k) = \hat{\boldsymbol{x}}_{\mathrm{a}}^{-}(k) + \boldsymbol{K}(k)\left(y(k) - \boldsymbol{C}_{\mathrm{a}}(\hat{f}(k))\hat{\boldsymbol{x}}_{\mathrm{a}}^{-}(k)\right)$

`/* Cylinder pressure estimation */`

$\hat{P}_1(k) = \begin{pmatrix}\boldsymbol{0} & \boldsymbol{C}_{P_1}\end{pmatrix}\hat{\boldsymbol{x}}_{\mathrm{a}}(k)$

$\hat{P}_3(k) = \tilde{G}_{\mathrm{d}}^{\mathrm{a}}(q^{-1}, \hat{f}(k))\hat{P}_1(k)$

$\hat{P}_4(k) = \tilde{G}_{\mathrm{d}}^{\mathrm{a}}(q^{-1}, \hat{f}(k))\hat{P}_3(k)$

$\hat{P}_2(k) = \tilde{G}_{\mathrm{d}}^{\mathrm{a}}(q^{-1}, \hat{f}(k))\hat{P}_4(k)$

end

The stability analysis of the square-root Kalman filter in Algorithm 6.2 is the same as the stability analysis in Algorithm 5.1.

6.3 Cylinder Pressure Estimation Results

In this section, the cylinder pressure estimation is conducted based on using Algorithm 6.2. As seen in Algorithm 6.2, the instantaneous engine cycle frequency $f(k)$ is estimated rather than being calculated. Below, the effectiveness of the estimation of the instantaneous engine cycle frequency $f(k)$ is first illustrated, and then the cylinder pressure estimation results related to pressure peak and peak location are given under nine stationary operating condition. Additionally, the results related to pressure peak and peak location are also given under two non-stationary operating conditions.

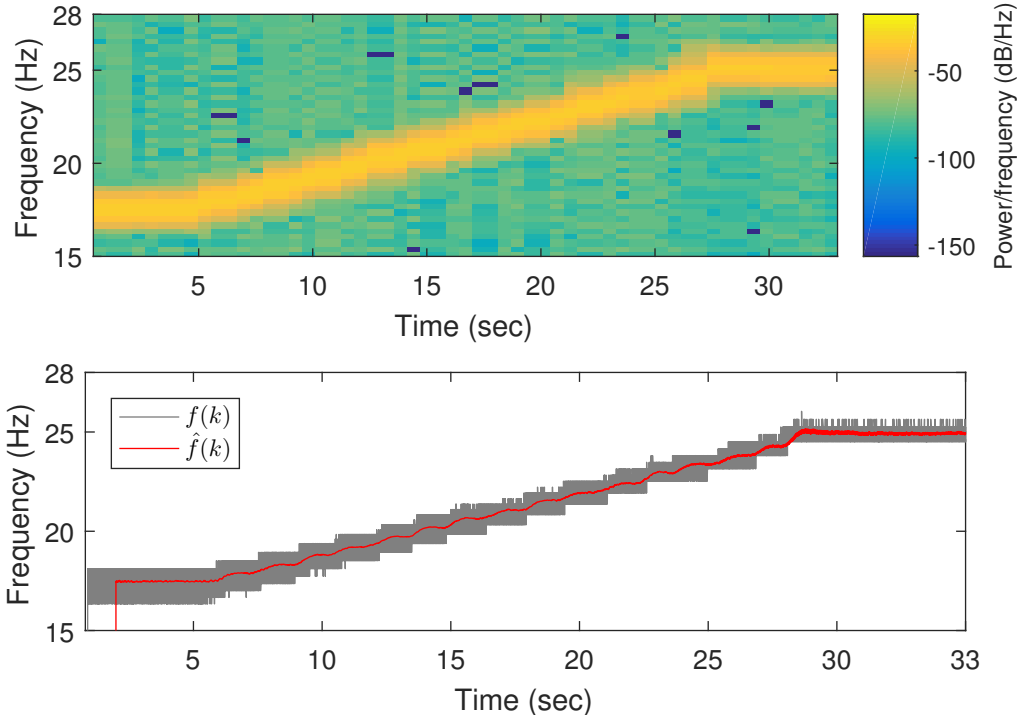

Figure 6.3: Estimated instantaneous engine cycle frequency.

In Figure 6.3, the effectiveness of the proposed frequency estimator E_{f} shown in Figure 6.2 can be validated, the estimated frequency has the same tendency of the corresponding calculated frequency $f(k)$ under the operating condition $\breve{Z}_{\mathrm{t}}$. Furthermore, the estimated frequency $\hat{f}(k)$ is more smooth than the calculated frequency $f(k)$. As a comparison, the spectrogram of the vibration signal $y(k)$ is also given for displaying the variations of the first order based on the short-time Fourier transform [51]. The results indicate that the estimator E_{f} can track time-variant frequency well. The parameters used in the frequency estimator are displayed in Table 6.1.

Table 6.1: Tunable parameters.

Parameter	Value
lower cutoff frequency	10 Hz
higher cutoff frequency	25 Hz
$\boldsymbol{Q}_{\mathrm{b}}$	$\boldsymbol{I}_5$
R_{b}	50
h	0.002

In Table 6.1, the lower cutoff frequency and the higher cutoff frequency are the frequencies specified for the band-pass filter in Figure 6.2.

With the estimated frequency $\hat{f}(k)$, based on using Algorithm 6.2 the cylinder pressure estimation results are given below. There are a number of tunable parameters in Algorithm 6.2 which are displayed in Table 6.2.

Table 6.2: Tunable parameters.

Parameter	Value
ν	35
l_{a}	10
n_{p}	40
σ_{a}^2	2.3×10^{-5}
λ	0.9995
$\hat{\boldsymbol{G}}$	9th order and MOESP(s) (time-domain identified model)

From Figure 6.4 to Figure 6.12, the curves of the estimated cylinder pressure are shown under nine different stationary operating conditions. Under nine stationary operating conditions, the mean value and standard deviation of the characteristic parameters are illustrated in Table 6.3 for the cylinder No. 1. For other cylinders, see tables from Table G.7 to Table G.9. In the tables, generally, the absolute values of $\mu(e_{P_{\max}})$ do not exceed 20%, while most of the absolute values of $\mu(e_{P_{\mathrm{loc}}})$ are smaller than 3°. Furthermore, most of the standard deviation values of $\mu(e_{P_{\max}})$ are smaller than 1%, while for peak location, almost all the values of $\sigma(e_{P_{\mathrm{loc}}})$ do not exceed 1.5°.

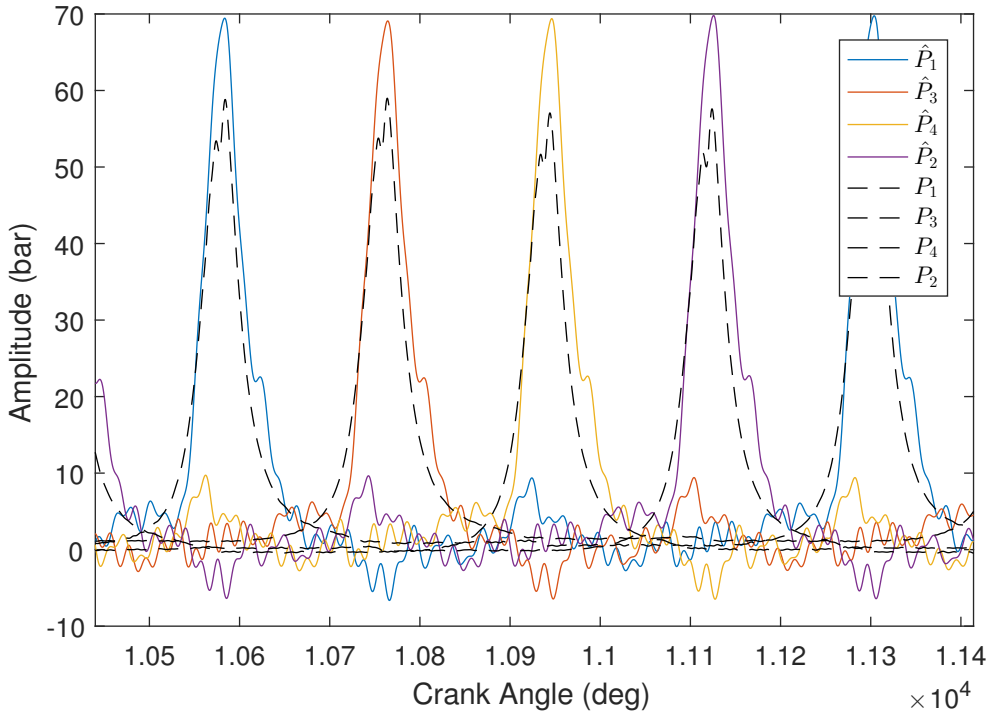

Figure 6.4: Cylinder pressure estimation under the stationary operating condition Z_1 (1200 rpm, 60 Nm) using Algorithm 6.2.

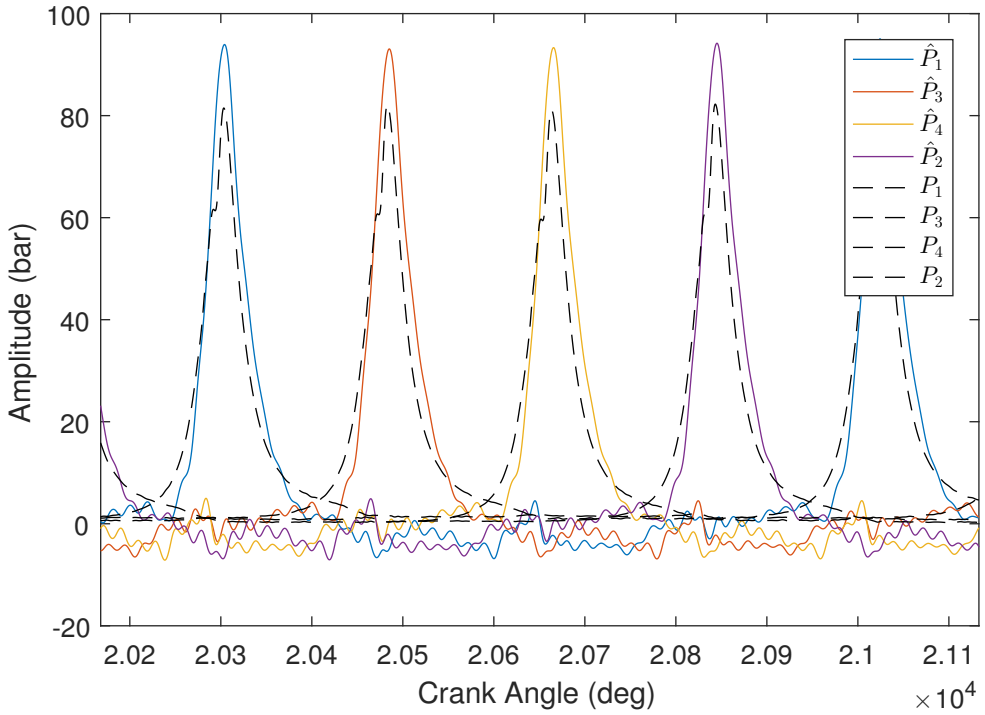

Figure 6.5: Cylinder pressure estimation under the stationary operating condition Z_2 (1200 rpm, 120 Nm) using Algorithm 6.2.

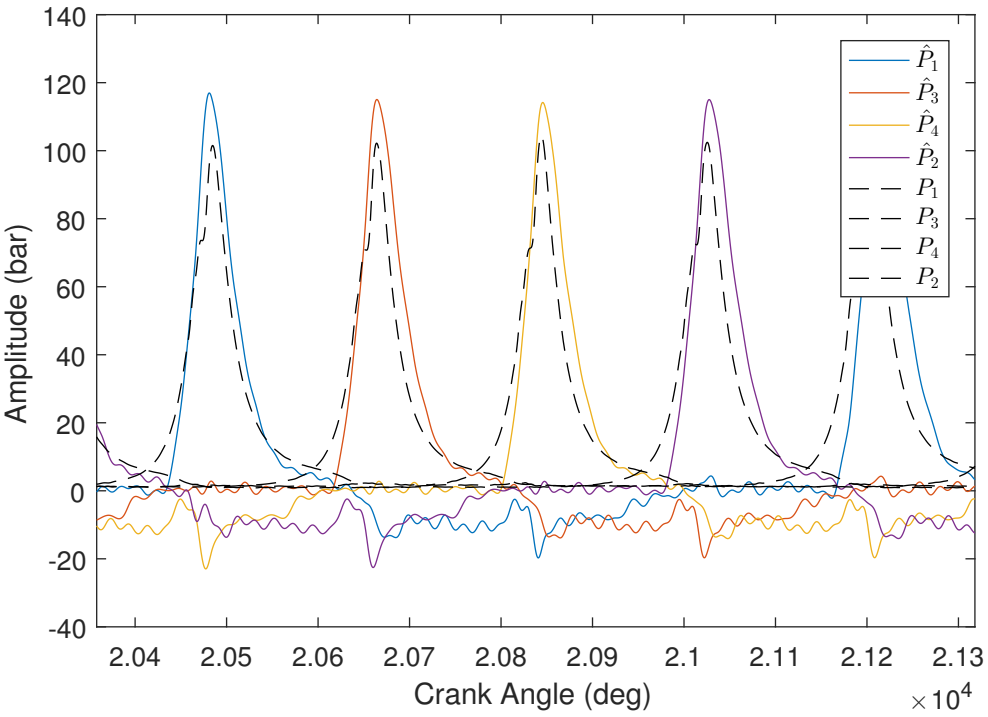

Figure 6.6: Cylinder pressure estimation under the stationary operating condition Z_3 (1200 rpm, 180 Nm) using Algorithm 6.2.

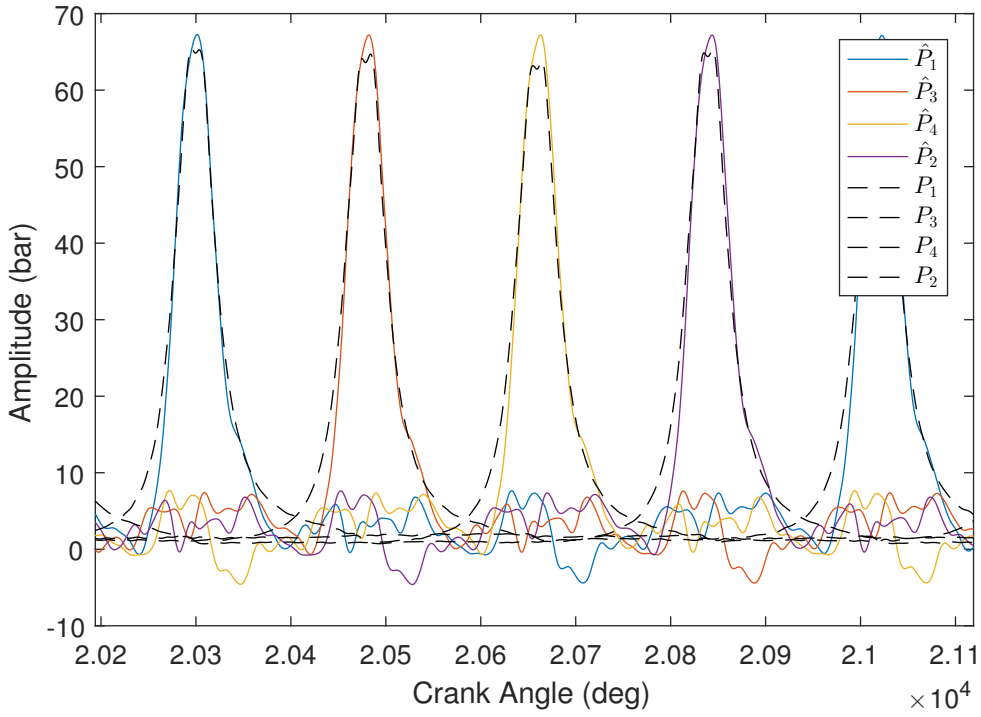

Figure 6.7: Cylinder pressure estimation under the stationary operating condition Z_4 (2100 rpm, 60 Nm) using Algorithm 6.2.

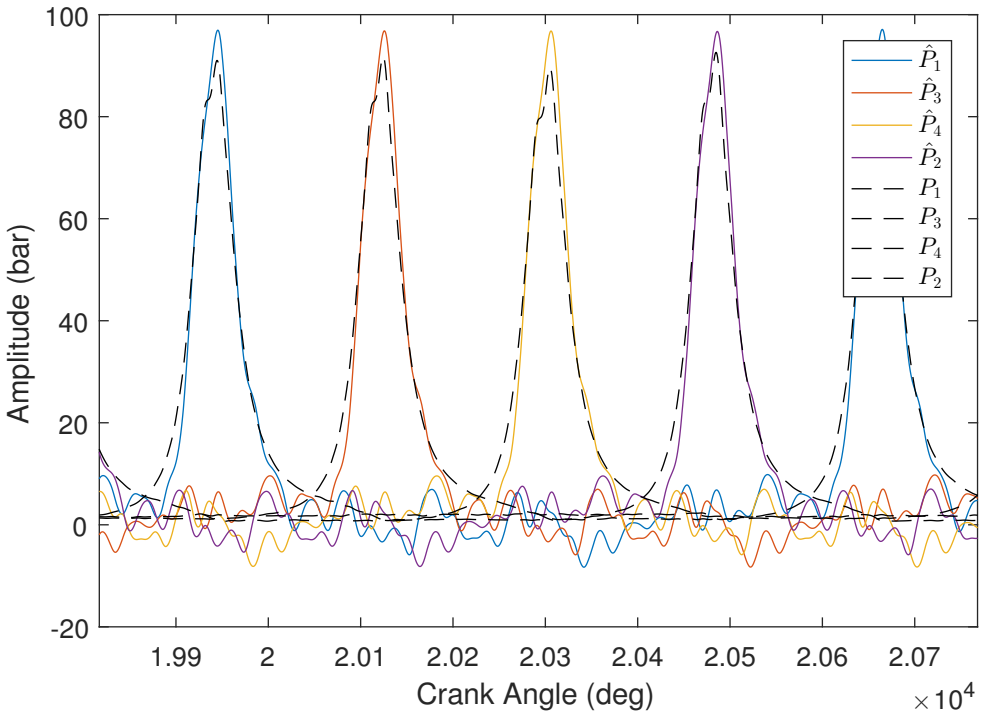

Figure 6.8: Cylinder pressure estimation under the stationary operating condition Z_5 (2100 rpm, 120 Nm) using Algorithm 6.2.

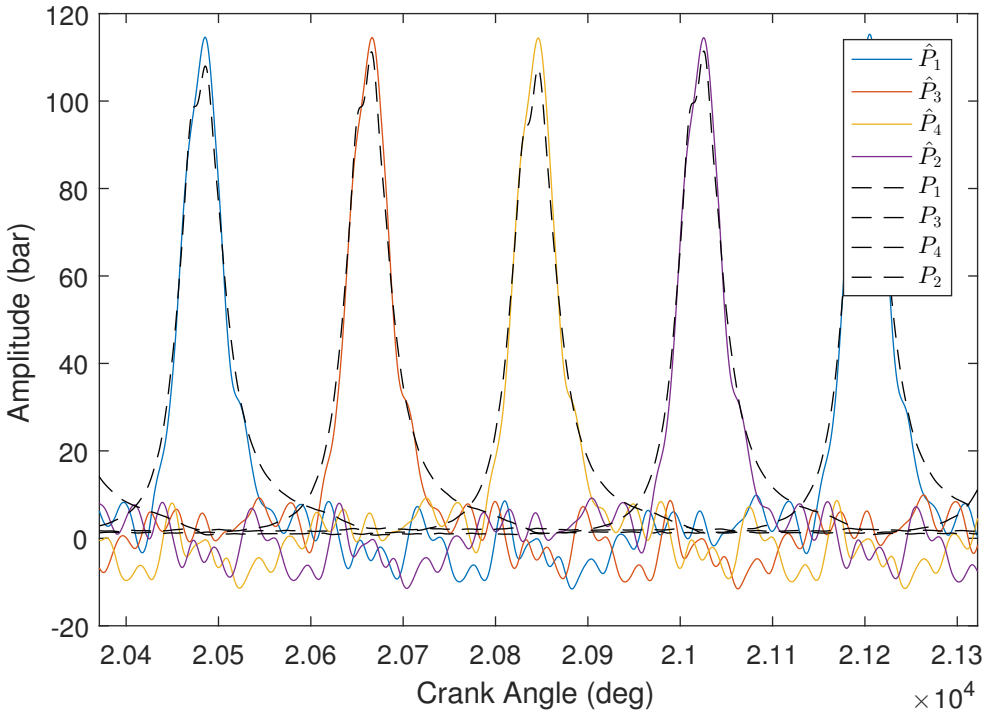

Figure 6.9: Cylinder pressure estimation under the stationary operating condition Z_6 (2100 rpm, 180 Nm) using Algorithm 6.2.

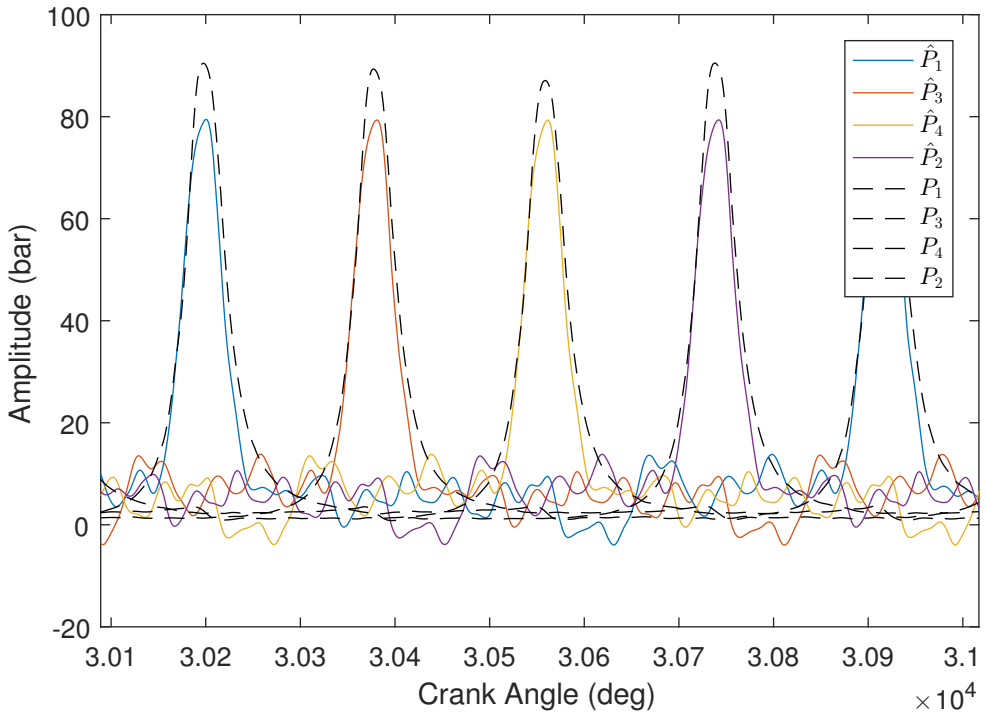

Figure 6.10: Cylinder pressure estimation under the stationary operating condition Z_7 (3000 rpm, 60 Nm) using Algorithm 6.2.

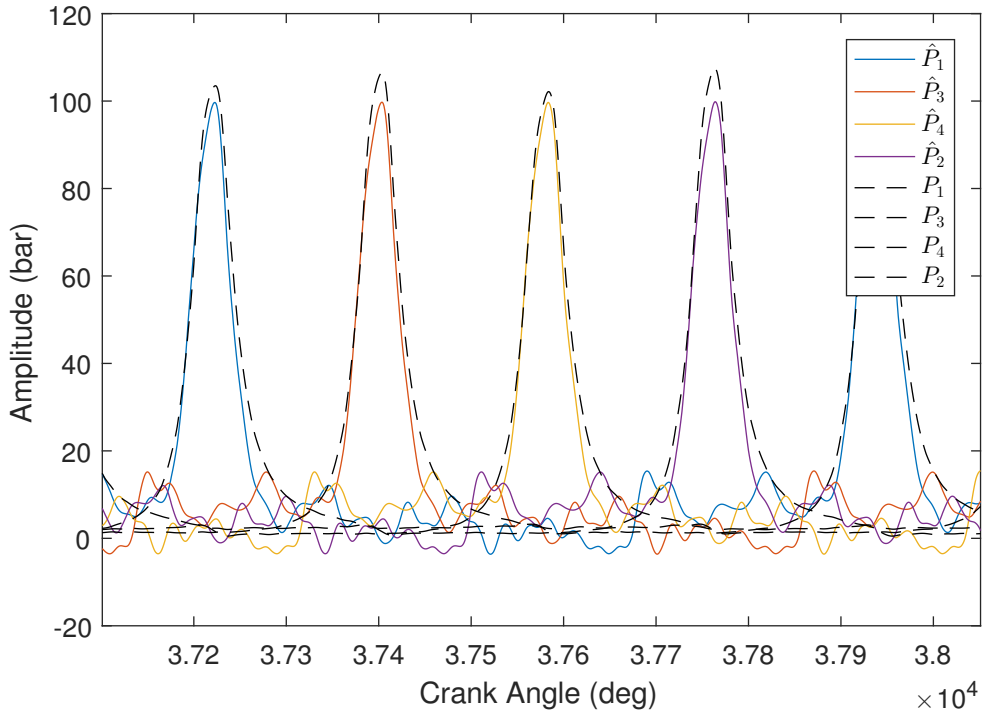

Figure 6.11: Cylinder pressure estimation under the stationary operating condition Z_8 (3000 rpm, 120 Nm) using Algorithm 6.2.

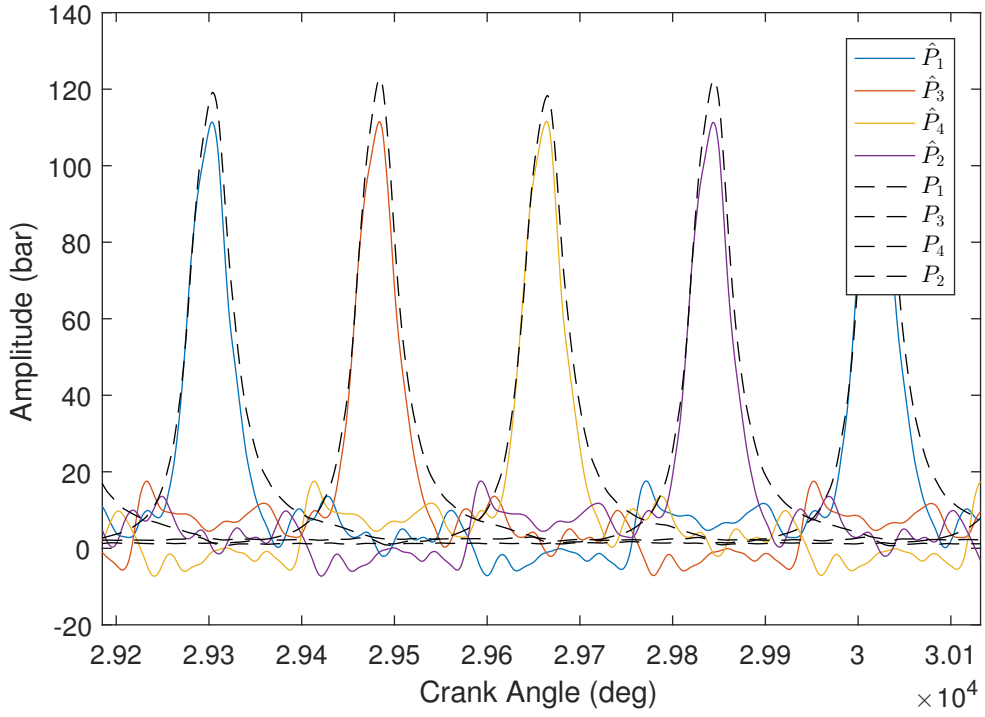

Figure 6.12: Cylinder pressure estimation under the stationary operating condition Z_9 (3000 rpm, 180 Nm) using Algorithm 6.2.

Table 6.3: Cylinder No. 1 pressure estimation results evaluation based on Algorithm 6.2 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	-18.10%	15.45%	-10.93%	-2.62%	-6.68%	-5.93%	11.82%	4.25%	6.63%
$\sigma(e_{P_{\max}})$	1.12%	0.68%	3.52%	0.33%	0.50%	0.50%	0.33%	0.43%	0.41%
$\mu(e_{P_{\mathrm{loc}}})$	$-0.64°$	$-1.48°$	$-1.46°$	$-4.47°$	$0.29°$	$-0.05°$	$-3.75°$	$0.90°$	$1.02°$
$\sigma(e_{P_{\mathrm{loc}}})$	$0.74°$	$0.49°$	$4.66°$	$4.17°$	$0.71°$	$0.49°$	$1.11°$	$1.05°$	$1.47°$

Note: mean($e_{P_{\max}}$) denotes the average value of 50 cylinder pressure peak values.

For the case of cylinder pressure estimation based on Algorithm 6.2, in general the estimation results (including the estimated curves and the estimated combustion metrics $P_{\max}$ and P_{loc}) under the stationary operating conditions Z_7, Z_8, and Z_9 are better than the estimation results under the stationary operating conditions Z_1, Z_2, and Z_3. Compared with the cylinder pressure estimation results in Section 5.3, generally the estimation results in this section become better under the stationary operating conditions Z_7, Z_8, and Z_9.

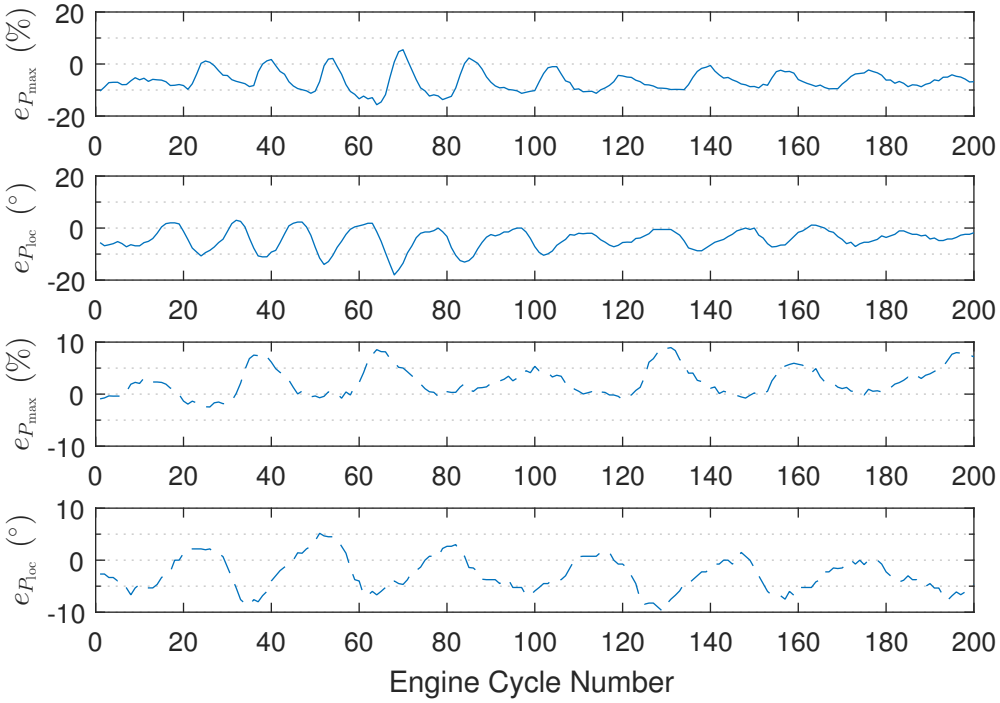

Figure 6.13: Cylinder No. 1 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\mathrm{t}}$ (lower two subfigures) using Algorithm 6.2.

Under non-stationary operating conditions, the characteristic parameters $e_{P_{\max}}$ and $e_{P_{\max}}$ are illustrated in the figures from Figure 6.13 and the figures from Figure G.7 to Figure G.9. According to the results, it can be found that in general, the absolute values of $e_{P_{\max}}$ do not exceed 20%, and the absolute values of $e_{P_{\mathrm{loc}}}$ are smaller than 10°.

The estimated cylinder pressure peak $P_{\max}$ and peak location P_{loc} under the operating condition $\tilde{Z}_{\mathrm{t}}$ are better than the results under the operating condition Z_{t}. Additionally, by a comparison with the values of $e_{P_{\max}}$ and $e_{P_{\mathrm{loc}}}$ under the non-stationary operating conditions in Section 5.3, the curves of the $e_{P_{\max}}$ and $e_{P_{\mathrm{loc}}}$ displayed in Figure 6.13 and the figures from Figure G.7 to Figure G.9 fluctuate more greatly.

6.4 Summary

In this chapter, the cylinder pressure is estimated based on the estimation method which merely uses the engine structural vibration signal. According to the estimated results under both stationary operating conditions and non-stationary operating conditions, they can show that even though both Algorithm 5.1 and Algorithm 6.2 are conducted based on the same model $\hat{\boldsymbol{G}}$ (i.e., 9th order and MOESP(s)), the cylinder pressure estimation results in this section become different with the estimation results in Section 5.3. Thus the conclusion of the calculation of the instantaneous engine cycle frequency can affect the cylinder pressure estimation in Chapter 5 is validated.

Chapter 7

Nonlinear Model-Based Cylinder Pressure Estimation

7.1 Introduction

In Chapter 5 and Chapter 6, cylinder pressure estimation is conducted in a linear framework. However, as aforementioned, in practice the relationship between four cylinder pressure signals and one vibration signal is complex, and the model accuracy can affect cylinder pressure estimation results. Therefore a nonlinear model between four cylinder pressure signals and one vibration signal has an advantage to describe the complexity. Additionally, as investigated in Section 5.5 even though based on using an ideal model, the cylinder pressure reconstruction results cannot be improved much, it is still meaningful to check the cylinder pressure estimation results based on using a nonlinear model because the ideal model cannot be obtained in practice. In this chapter, based on the nonlinear model $\hat{\boldsymbol{G}}_{\text{non}}$ identified in Chapter 4, a nonlinear model-based cylinder pressure estimation framework is developed. This chapter is organized as follows. In Section 7.2, the specific nonlinear cylinder pressure estimation algorithm is first illustrated, followed by Section 7.3, in which the corresponding cylinder pressure estimation results are given. Finally, in Section 7.3, a summary of this chapter is given.

7.2 Nonlinear Cylinder Pressure Estimation Algorithm

The total nonlinear cylinder pressure estimation algorithm is illustrated in Figure 7.1. The total framework contains two parts, i.e., part A and part B.

Part A can be seen as an offline design procedure, which contains three components, and they are described as follows:

(i) Identify the nonlinear model (2.136) between four (processed) cylinder pressure signals and one (processed) vibration signal. As aforementioned, the deterministic sub-model in the model (2.136) is denoted as $\boldsymbol{G}_{\text{non}}^{\text{d}}$. $e_{\text{n}}(k)$ denotes the scalar output error.

(ii) Regard other three cylinder pressure signals as the delay versions of the cylinder No. 1 pressure signal. As a result, by combing three delay systems with the identified

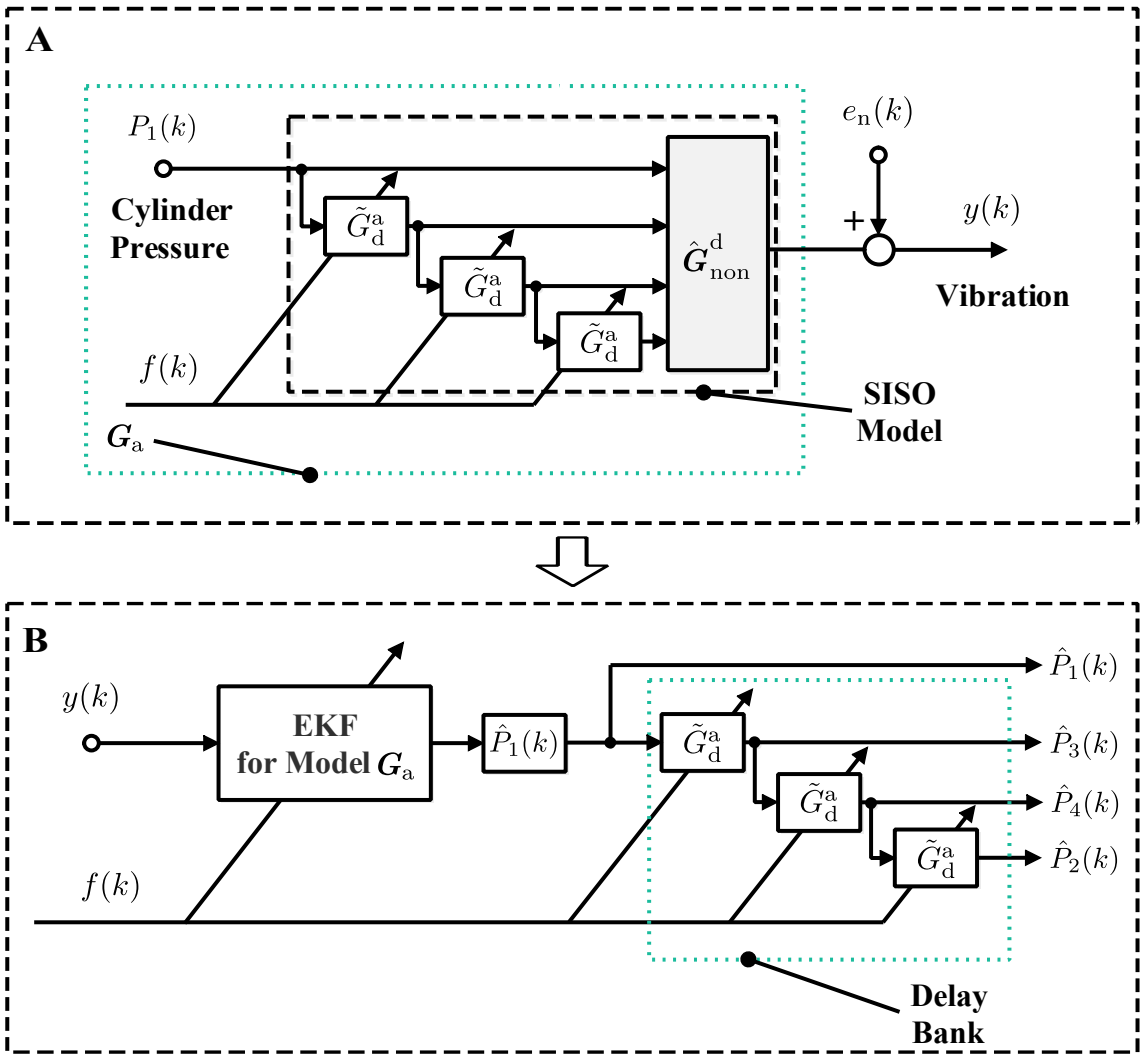

Figure 7.1: Nonlinear cylinder pressure estimation framework.

model $\hat{\boldsymbol{G}}_{\mathrm{non}}^{\mathrm{d}}$, a SISO model can be finally obtained as displayed in part A of Figure 7.1.

(iii) Augment the cylinder pressure signal model with the identified model $\hat{\boldsymbol{G}}_{\mathrm{non}}^{\mathrm{d}}$ and three delay systems (i.e., state augmentation), then the augmented model $\boldsymbol{G}_{\mathrm{a}}$ can be obtained. The cylinder pressure signal modeling can be referred to the order modeling as introduced in Section 5.2.1.

Part B can be regarded as an implementation procedure. With the nonlinear augmented model $\boldsymbol{G}_{\mathrm{a}}$, a nonlinear Kalman filter for it can be derived, and then by using the vibration signal as the input of the nonlinear Kalman filter (e.g., EKF), the state of the augmented model can be recursively estimated, and with the estimated state, the estimated value of the cylinder No. 1 pressure signal can be recursively obtained, and simultaneously other three cylinder pressure signals can also be obtained based on the delay block bank which contains three delay blocks.

Below, the model augmentation and nonlinear cylinder pressure estimation algorithm are described in Section 7.2.1 and Section 7.2.2, respectively.

7.2.1 Model Augmentation

With the TVDB model (5.28), the SISO model with additive error shown in Figure 7.1 can be represented as

$$\begin{cases} \boldsymbol{x}_{\mathrm{s}}(k+1) = \boldsymbol{A}_{\mathrm{s}}(f(k))\boldsymbol{x}_{\mathrm{s}}(k) + \boldsymbol{B}_{\mathrm{s}}(f(k))P_1(k), \\ y(k) = f_{\mathrm{N}}\left(\boldsymbol{C}_{\mathrm{s}}(f(k))\boldsymbol{x}_{\mathrm{s}}(k)\right) + e_{\mathrm{n}}(k), \end{cases} \tag{7.1}$$

where the state vector $\boldsymbol{x}_{\mathrm{s}}(k) \in \mathbb{R}^{n_{\mathrm{s}}}$, and the output error $e_{\mathrm{n}}(k)$ is induced by the measurement noise, the disturbance on the model $\boldsymbol{G}^{\mathrm{d}}_{\mathrm{non}}$, the cylinder-to-cylinder difference, and the error in the formulation process of TVDB model (5.28).

The matrices $\boldsymbol{A}_{\mathrm{s}}(f(k))$, $\boldsymbol{B}_{\mathrm{s}}(f(k))$, $\boldsymbol{C}_{\mathrm{s}}(f(k))$, $\boldsymbol{D}_{\mathrm{s}}$, and the state vector $\boldsymbol{x}_{\mathrm{s}}(k)$ are respectively given as

$$\boldsymbol{A}_{\mathrm{s}}(f(k)) = \left(\begin{array}{c:ccc} \boldsymbol{A}_{\mathrm{L}} & \boldsymbol{B}_{\mathrm{L}} \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{C}_{\mathrm{d}}(f(k)) \end{pmatrix} \\ \hdashline \boldsymbol{0} & \begin{pmatrix} \boldsymbol{G}^{\mathrm{a}}_{\mathrm{d}}(f(k)) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{H}^{\mathrm{a}}_{\mathrm{d}}(f(k))\boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{G}^{\mathrm{a}}_{\mathrm{d}}(f(k)) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}^{\mathrm{a}}_{\mathrm{d}}(f(k))\boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{G}^{\mathrm{a}}_{\mathrm{d}}(f(k)) \end{pmatrix} \end{array}\right), \tag{7.2}$$

$$\boldsymbol{B}_{\mathrm{s}}(f(k)) = \left(\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \boldsymbol{B}_{\mathrm{L}}^{\mathrm{T}} \quad \left(\boldsymbol{H}^{\mathrm{a}}_{\mathrm{d}}(f(k))\right)^{\mathrm{T}} \quad \boldsymbol{0}\right)^{\mathrm{T}}, \tag{7.3}$$

$$\boldsymbol{C}_{\mathrm{s}}(f(k)) = \left(\begin{array}{c:c} \boldsymbol{C}_{\mathrm{L}} & \boldsymbol{D}_{\mathrm{L}} \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{\mathrm{d}}(f(k)) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{C}_{\mathrm{d}}(f(k)) \end{pmatrix} \end{array}\right), \tag{7.4}$$

$$\boldsymbol{D}_{\mathrm{s}} = \boldsymbol{D}_{\mathrm{L}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{7.5}$$

and

$$\boldsymbol{x}_{\mathrm{s}}(k) = \left(\boldsymbol{x}^{\mathrm{T}}(k) \quad \boldsymbol{x}_1^{\mathrm{T}}(k) \quad \boldsymbol{x}_2^{\mathrm{T}}(k) \quad \boldsymbol{x}_3^{\mathrm{T}}(k)\right)^{\mathrm{T}}, \tag{7.6}$$

where $\boldsymbol{x}_1(k)$, $\boldsymbol{x}_2(k)$, and $\boldsymbol{x}_3(k)$ denote the state vectors of three delay block models shown in Figure 7.1, respectively.

By augmenting the model (7.1) with the state of the model of the cylinder No. 1 pressure signal (see the model (5.3)), the following augmented model $\boldsymbol{G}_{\mathrm{a}}$ can be obtained:

$$\begin{cases} \boldsymbol{x}_{\mathrm{a}}(k+1) = \boldsymbol{A}_{\mathrm{a}}(f(k))\boldsymbol{x}_{\mathrm{a}}(k), \\ y(k) = f_{\mathrm{N}}\left(\boldsymbol{C}_{\mathrm{s}}(f(k))\boldsymbol{C}_{\mathrm{a}}\boldsymbol{x}_{\mathrm{a}}(k)\right) + e_{\mathrm{a}}(k), \end{cases} \tag{7.7}$$

where the state vector $\boldsymbol{x}_{\mathrm{a}}(k) \in \mathbb{R}^{n_{\mathrm{a}}}$, and the term $e_{\mathrm{a}}(k)$ denotes the output error which is induced by $e_{\mathrm{n}}(k)$ and $e_{P_1}(k)$.

The matrices $\boldsymbol{A}_{\mathrm{a}}(f(k))$ and $\boldsymbol{C}_{\mathrm{a}}(f(k))$, and the state vector $\boldsymbol{x}_{\mathrm{a}}(k)$ are given as

$$\boldsymbol{A}_{\mathrm{a}}(f(k)) = \begin{pmatrix} \boldsymbol{A}_{\mathrm{s}}(f(k)) & \boldsymbol{B}_{\mathrm{s}}(f(k))\boldsymbol{C}_{P_1} \\ \boldsymbol{0} & \boldsymbol{A}_{P_1}(f(k)) \end{pmatrix}, \tag{7.8}$$

$$\boldsymbol{C}_{\mathrm{a}} = \begin{pmatrix} \boldsymbol{I}_{n_{\mathrm{s}}} & \boldsymbol{0} \end{pmatrix}, \tag{7.9}$$

and

$$\boldsymbol{x}_{\mathrm{a}}(k) = \begin{pmatrix} \boldsymbol{x}_{\mathrm{s}}^{\mathrm{T}}(k) & \boldsymbol{x}_{P_1}^{\mathrm{T}}(k) \end{pmatrix}^{\mathrm{T}}, \tag{7.10}$$

respectively.

7.2.2 Nonlinear Cylinder Pressure Estimation

Based on the nonlinear cylinder pressure estimation framework displayed in Figure 7.1 and the above analysis, four cylinder pressure signals can be recursively estimated. In the implementation process, the nonlinear square-root Kalman filter with a forgetting factor is used for handling the numerical problem, the non-stationary problem, and the mismodling problem (see Section 5.3). Prior to introducing the specific reconstruction method, the sequence $\{e_{\mathrm{a}}(k)\}$ is assumed to be a scalar white noise process, of which the covariance function is $\sigma_{\mathrm{a}}^2\delta_{kj}$, i.e., $\mathrm{E}\left[e_{\mathrm{a}}(k)e_{\mathrm{a}}^{\mathrm{T}}(j)\right] = \sigma_{\mathrm{a}}^2\delta_{kj}$, and the value of σ_{a}^2 is tunable. $e_{\mathrm{a}}(k)$ is also assumed to be uncorrelated with $\boldsymbol{x}_{\mathrm{a}}(0)$ which is assumed to be a random variable with constant mean and variance. The specific nonlinear cylinder pressure estimation procedure is given in Algorithm 7.1.

Algorithm 7.1: Nonlinear cylinder pressure estimation algorithm using the square-root Kalman filter with a forgetting factor.

Initialization: $\boldsymbol{S}(0) = 10\boldsymbol{I}_{n_{\mathrm{a}}}$, $\hat{\boldsymbol{x}}_{\mathrm{a}}(0) = \boldsymbol{I}_{n_{\mathrm{a}}}$

for $k = 1 \to N-1$ **do**

/* Nonlinear square-root Kalman filter with a forgetting factor */

$\boldsymbol{S}^{-}(k) = \frac{1}{\lambda}\boldsymbol{A}_{\mathrm{a}}(f(k-1))\boldsymbol{S}(k-1)$ // λ denotes the forgetting factor, see Remark 2.4

$\hat{\boldsymbol{x}}_{\mathrm{a}}^{-}(k) = \boldsymbol{A}_{\mathrm{a}}(f(k-1))\hat{\boldsymbol{x}}_{\mathrm{a}}(k-1)$

$\boldsymbol{F}(k) = \boldsymbol{S}^{-}(k)\boldsymbol{C}_{\mathrm{aL}}^{\mathrm{T}}(k)$ // $\boldsymbol{C}_{\mathrm{aL}}(k) \approx \left.\frac{\partial f_{\mathrm{N}}(\boldsymbol{C}_{\mathrm{s}}(f(k))\boldsymbol{C}_{\mathrm{a}}\boldsymbol{x}_{\mathrm{a}}(k))}{\partial \boldsymbol{x}_{\mathrm{a}}^{\mathrm{T}}(k)}\right|_{\boldsymbol{x}_{\mathrm{a}}(k)=\hat{\boldsymbol{x}}_{\mathrm{a}}^{-}(k)}$

$\alpha(k) = \frac{1}{\boldsymbol{F}^{\mathrm{T}}(k)\boldsymbol{F}(k)+\sigma_{\mathrm{a}}^2}$

$\boldsymbol{K}(k) = \alpha(k)\boldsymbol{S}^{-}(k)\boldsymbol{F}(k)$

$\gamma(k) = \frac{1}{1+\sqrt{\sigma_{\mathrm{a}}^2}\sqrt{\alpha(k)}}$

$\boldsymbol{S}(k) = \boldsymbol{S}^{-}(k) - \gamma(k)\boldsymbol{K}(k)\boldsymbol{F}^{\mathrm{T}}(k)$

$\hat{\boldsymbol{x}}_{\mathrm{a}}(k) = \hat{\boldsymbol{x}}_{\mathrm{a}}^{-}(k) + \boldsymbol{K}(k)\left(y(k) - \boldsymbol{C}_{\mathrm{aL}}(k)\hat{\boldsymbol{x}}_{\mathrm{a}}^{-}(k)\right)$

/* Cylinder pressure estimation */

$\hat{P}_1(k) = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{C}_{P_1} \end{pmatrix}\hat{\boldsymbol{x}}_{\mathrm{a}}(k)$

$\hat{P}_3(k) = \tilde{G}_{\mathrm{d}}^{\mathrm{a}}(q^{-1}, f(k))\hat{P}_1(k)$ // $\tilde{G}_{\mathrm{d}}^{\mathrm{a}}(q^{-1}, f(k))$ denotes the conceptual time-varying transfer operator of the model (5.28) (i.e., $\tilde{G}_{\mathrm{d}}^{\mathrm{a}}$)

$\hat{P}_4(k) = \tilde{G}_{\mathrm{d}}^{\mathrm{a}}(q^{-1}, f(k))\hat{P}_3(k)$

$\hat{P}_2(k) = \tilde{G}_{\mathrm{d}}^{\mathrm{a}}(q^{-1}, f(k))\hat{P}_4(k)$

The stability of the nonlinear square-root Kalman filter in Algorithm 7.1 is not emphasized in this thesis, and it can be referred to [26].

7.3 Nonlinear Cylinder Pressure Estimation Results

Algorithm 7.1 was implemented to estimate four cylinder pressure signals. From figure 7.2 to Figure 7.10, the estimated curves of four cylinder pressure signals under nine stationary operating conditions are displayed.

For cylinder pressure estimation in Algorithm 7.1, there are a number of tunable parameters, and the tunable parameters used here are displayed in Table 7.1. Specifically, besides the identified model, other parameters in Algorithm 7.1 are chosen as the same as the parameters used in the linear cylinder pressure estimation framework in Section 5.3.

Table 7.1: Tunable parameters.

Parameter	Value
ν	35
l_{a}	10
n_{p}	40
σ_{a}^2	2.3×10^{-5}
λ	0.9995
$\hat{\boldsymbol{G}}_{\mathrm{non}}^{\mathrm{d}}$	9th order (nonlinear model)

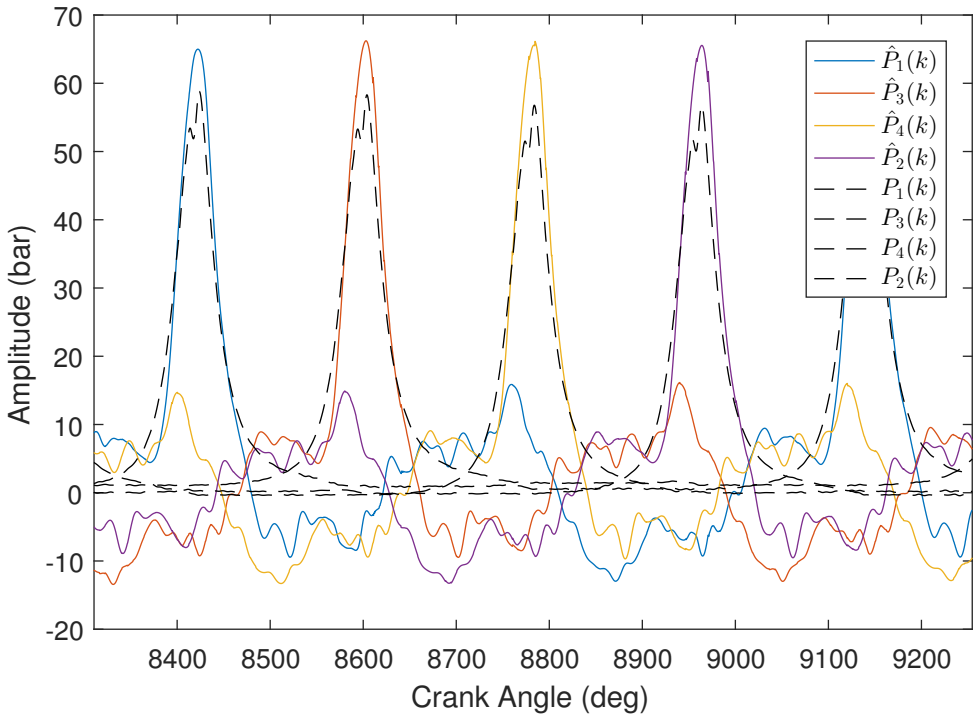

Figure 7.2: Cylinder pressure estimation under the stationary operating condition Z_1 (1200 rpm, 60 Nm) using Algorithm 7.1.

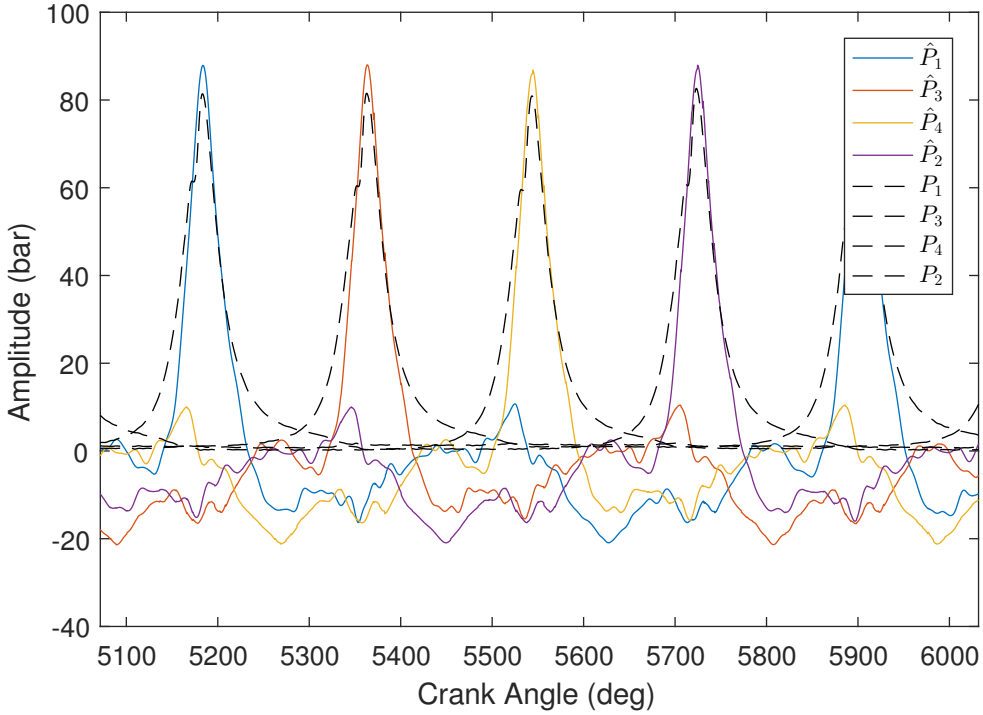

Figure 7.3: Cylinder pressure estimation under the stationary operating condition Z_2 (1200 rpm, 120 Nm) using Algorithm 7.1.

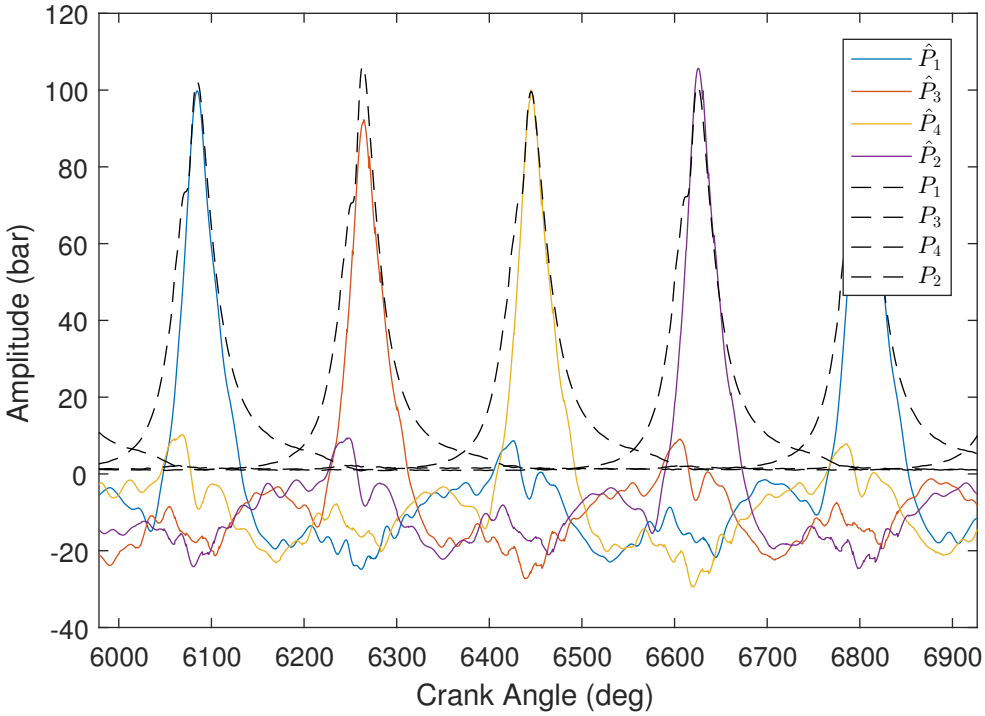

Figure 7.4: Cylinder pressure estimation under the stationary operating condition Z_3 (1200 rpm, 180 Nm) using Algorithm 7.1.

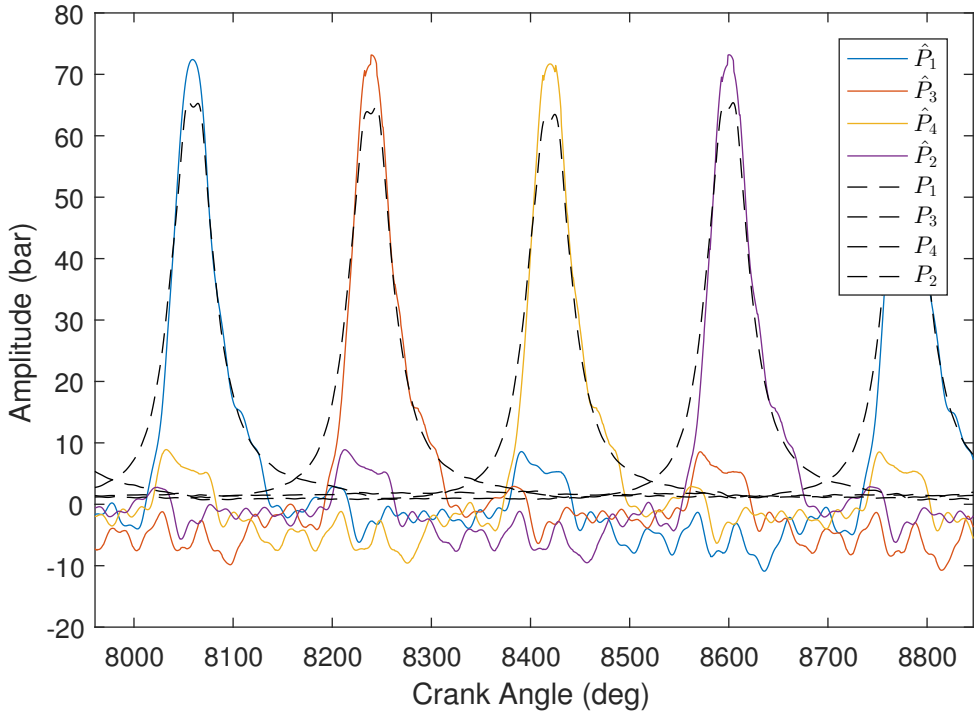

Figure 7.5: Cylinder pressure estimation under the stationary operating condition Z_4 (2100 rpm, 60 Nm) using Algorithm 6.2.

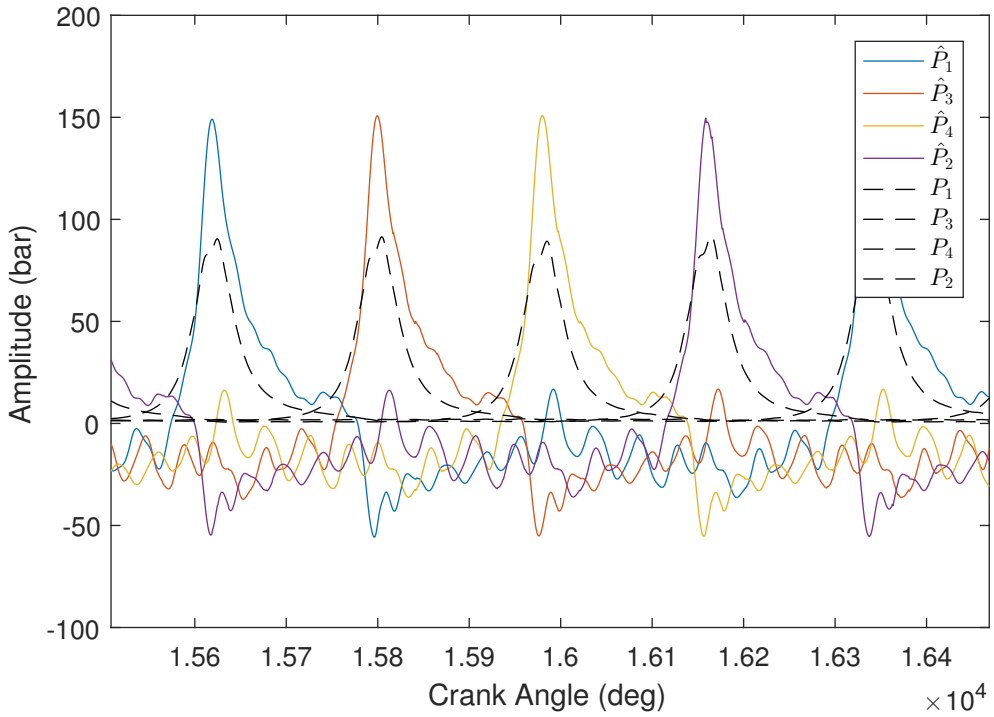

Figure 7.6: Cylinder pressure estimation under the stationary operating condition Z_5 (2100 rpm, 120 Nm) using Algorithm 7.1.

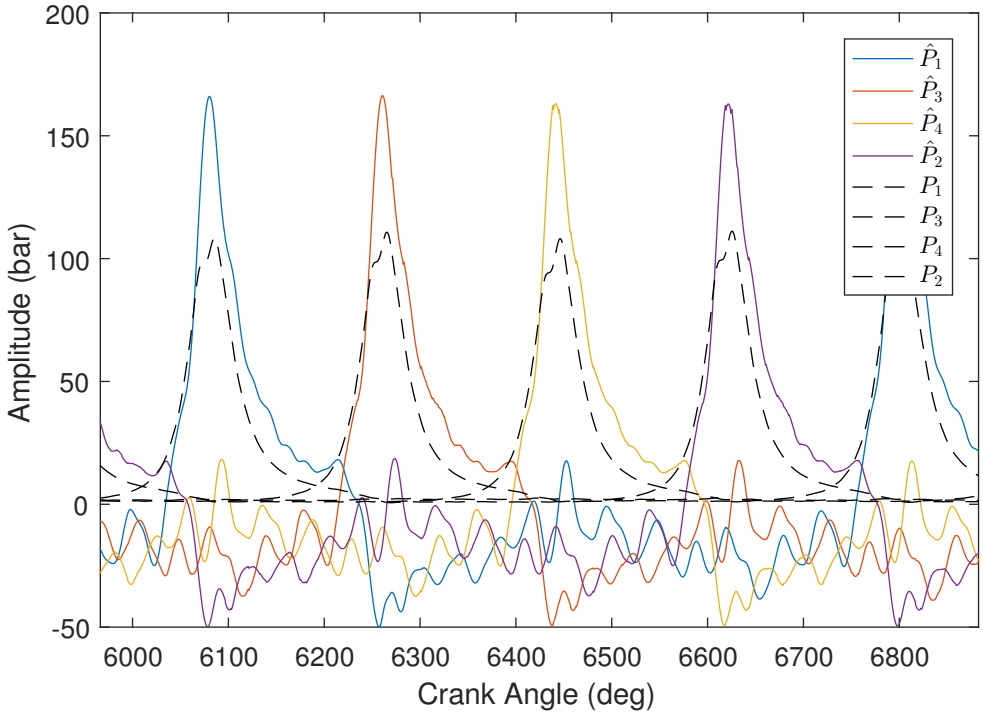

Figure 7.7: Cylinder pressure estimation under the stationary operating condition Z_6 (2100 rpm, 180 Nm) using Algorithm 7.1.

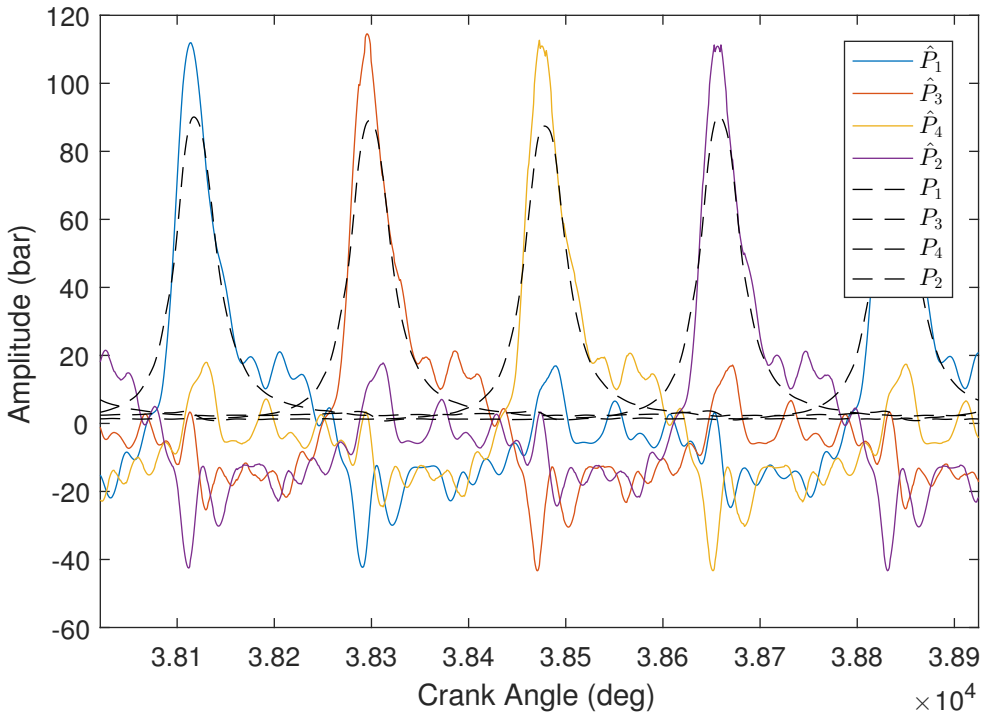

Figure 7.8: Cylinder pressure estimation under the stationary operating condition Z_7 (3000 rpm, 60 Nm) using Algorithm 7.1.

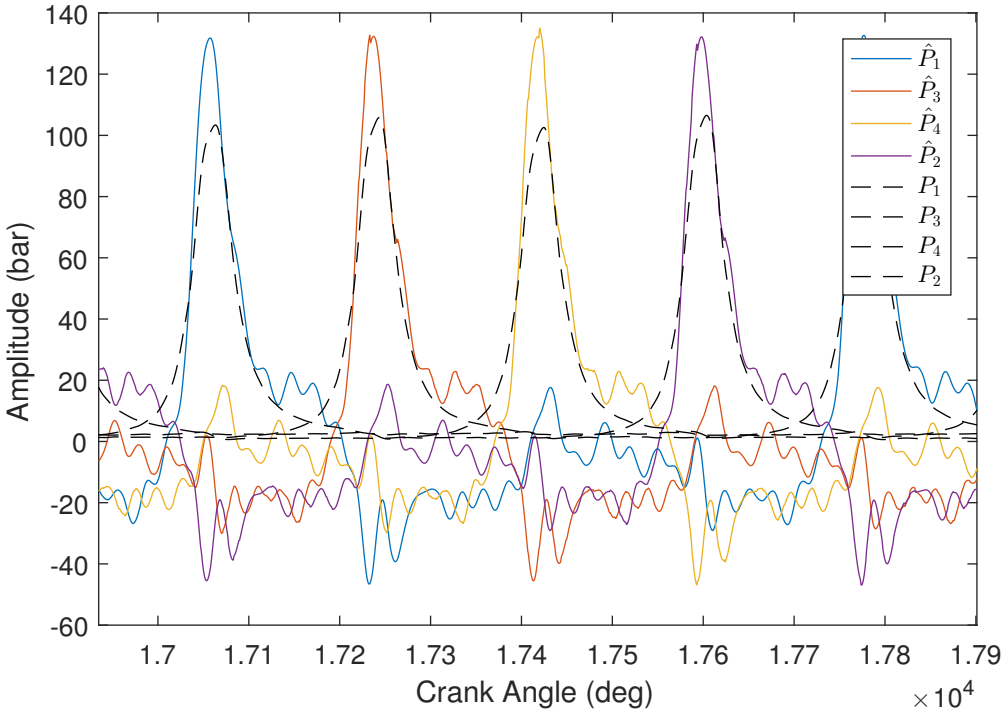

Figure 7.9: Cylinder pressure estimation under the stationary operating condition Z_8 (3000 rpm, 120 Nm) using Algorithm 7.1.

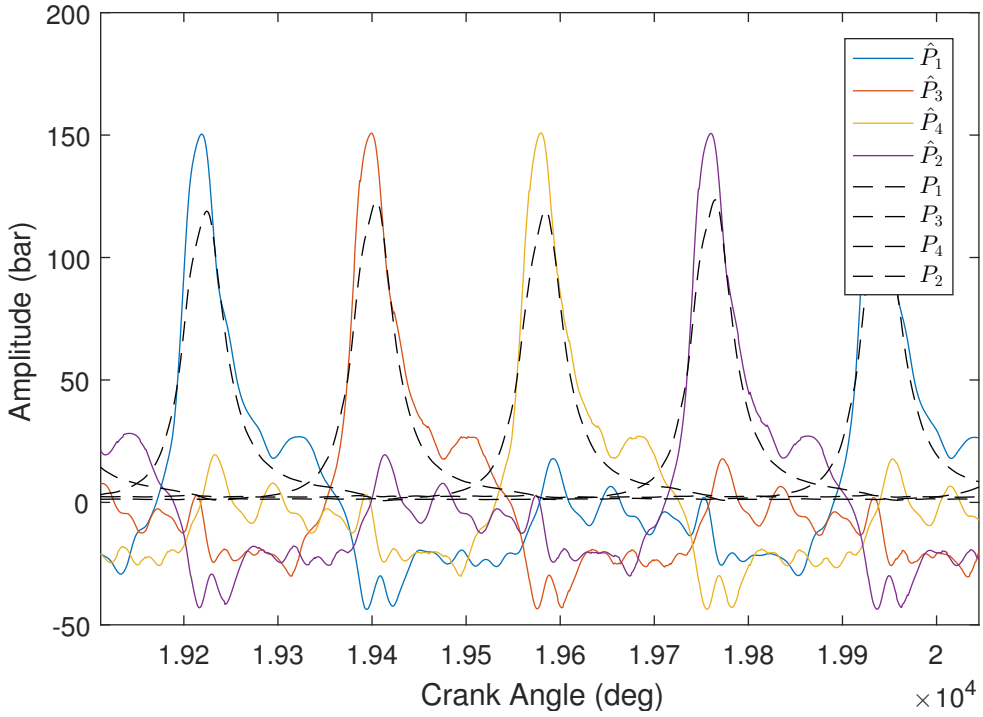

Figure 7.10: Cylinder pressure estimation under the stationary operating condition Z_9 (3000 rpm, 180 Nm) using Algorithm 7.1.

Corresponding to the figures from Figure 7.2 to Figure 7.10, Table 7.2 and the tables from Table G.10 to Table G.12 display the cylinder pressure estimation results (i.e., the average values $\mu(e_{P_{\max}})$ and $\mu(e_{P_{\mathrm{loc}}})$ and standard deviation values $\sigma(e_{P_{\max}})$ and $\sigma(e_{P_{\mathrm{loc}}})$ introduced in Chapter 5) under nine stationary operating conditions. The values of $\mu(e_{P_{\max}})$ vary from -65.14% to 8.44%, but most of the absolute values of $\mu(e_{P_{\max}})$ are less than 30%. Additionally, most of values of $\sigma(e_{P_{\max}})$ are less than 2.5%. For the peak location estimation, most of the absolute values of $\mu(e_{P_{\mathrm{loc}}})$ do not exceed $5°$, while most of values of $\sigma(e_{P_{\mathrm{loc}}})$ are less than $1°$.

For the case of cylinder pressure estimation based on the nonlinear model $\hat{\boldsymbol{G}}^{\mathrm{d}}_{\mathrm{non}}$, in general the estimation results (including the estimated curves and the estimated combustion metrics $P_{\max}$ and P_{loc}) under the stationary operating conditions Z_1, Z_2, and Z_3 are better than the estimation results under other stationary operating conditions. Compared with the cylinder pressure estimation results in Section 5.3, the estimation results based on the nonlinear model are worse in general.

Table 7.2: Cylinder No. 1 pressure estimation results evaluation based on Algorithm 7.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	-11.30%	7.81%	1.03%	-11.44%	-61.24%	-51.50%	-23.64%	-28.41%	-25.54%
$\sigma(e_{P_{\max}})$	1.45%	0.97%	1.64%	0.95%	2.03%	1.27%	0.97%	0.95%	1.10%
$\mu(e_{P_{\mathrm{loc}}})$	$1.67°$	$-0.77°$	$-0.51°$	$-0.68°$	$6.09°$	$5.45°$	$4.28°$	$5.74°$	$4.59°$
$\sigma(e_{P_{\mathrm{loc}}})$	$0.55°$	$0.24°$	$0.45°$	$4.17°$	$0.38°$	$0.38°$	$0.71°$	$0.77°$	$0.55°$

Note: mean$(e_{P_{\max}})$ denotes the average value of 50 cylinder pressure peak values.

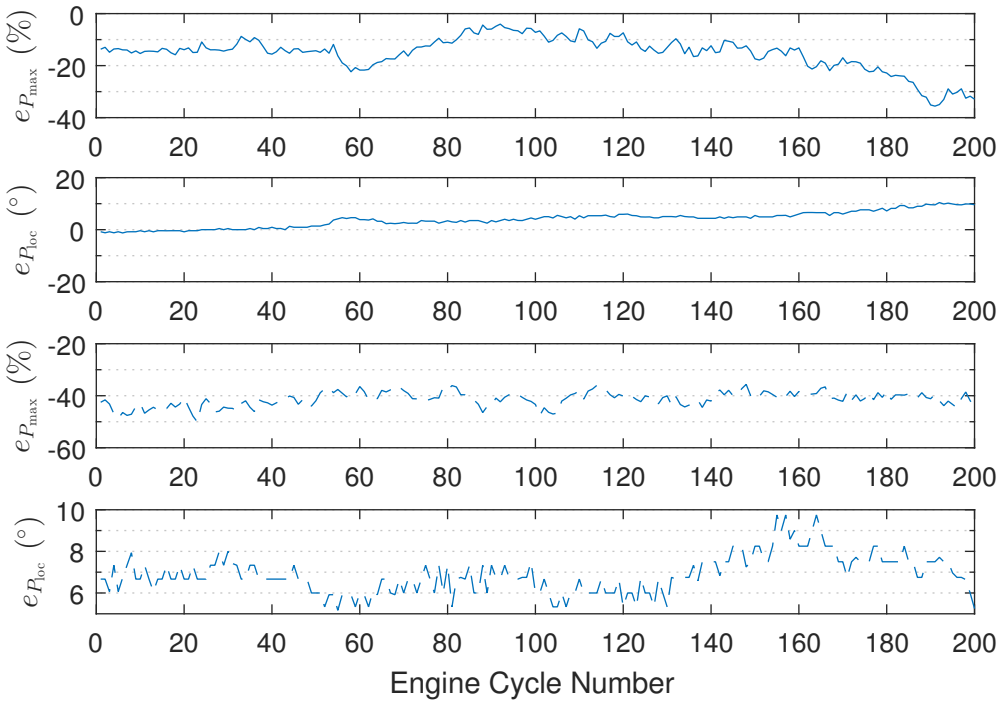

Figure 7.11: Cylinder No. 1 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\mathrm{t}}$ (lower two subfigures) using Algorithm 7.1.

For non-stationary operating conditions (i.e., the transient process Z_t and the transient process $\tilde{Z}_\mathrm{t}$, the pressure peak and peak location estimation results are illustrated in Figure 7.11 and the figures from Figure G.10 to Figure G.12. According to the results, it can be found that for lower engine speeds (i.e., under the condition Z_t), in general, the absolute values of e_{P_max} do not exceed 20%, and the absolute values of e_{P_loc} are smaller than 10°. As a comparison, for higher engine speeds (i.e., under the condition $\tilde{Z}_\mathrm{t}$), generally the absolute values of e_{P_max} are less than 50%, while the absolute values of e_{P_loc} do not exceed 10°.

The cylinder pressure estimation results (i.e., the pressure peak and peak location) under the operating condition Z_t are better than the results under the operating condition $\tilde{Z}_\mathrm{t}$. With the estimation results under both stationary operating conditions and non-stationary operating conditions, it can be known that the cylinder pressure estimation is especially effective at low engine speeds. Additionally, by a comparison with the cylinder pressure estimation results under non-stationary operating conditions in Section 5.3 and Section 5.4, the estimation results become worse generally.

7.4 Summary

In this chapter, the cylinder pressure estimation is conducted by using the nonlinear framework as illustrated in Figure 7.1. The nonlinear estimation framework was validated under both stationary operating conditions and non-stationary operating conditions. Compared with the estimation results in the linear framework proposed in Chapter 5, under the nonlinear framework the estimation results do not become better even though the nonlinear model $\hat{\boldsymbol{G}}^\mathrm{d}_\mathrm{non}$ has higher fitness, which also indicates that the model quality is not the unique factor which can affect the cylinder pressure estimation.

Chapter 8

Inverse Model-Based Cylinder Pressure Reconstruction

8.1 Introduction

Given an LTI system model, the inversion of the model has numerous practical implementations including but not limited to input reconstruction [86, 87] and inverse model-based feedforward control [88, 89]. For nonminimum-phase systems[1] or strictly proper systems, such inversion is not always straightforward. In addition, multivariable systems and unstable systems impose additional complications. System inversion has received significant attention, also from a theoretical perspective [90]. Comparing with the methods proposed in Chapter 5, Chapter 6, and Chapter 7, in this chapter, an inverse identification-based cylinder pressure reconstruction method, which can avoid the inverse problems which are encountered in nonminimum-phase systems or strictly proper systems, is proposed. The proposed method does not need delay blocks and the calculation of the engine speed. The outline of the subsequent sections of this chapter is depicted as follows. In Section 8.2, the main algorithm of inverse model-based cylinder pressure reconstruction is first illustrated, and then in Section 8.3 and Section 8.4, the corresponding reconstruction results and a numerical example are shown. At last, a summary of this chapter is given.

8.2 Cylinder Pressure Reconstruction by Using Inverse Model Identification

Write the model (4.2) a little more explicitly, and then the following equation can be obtained:

$$y(k) = \begin{pmatrix} G_1(q^{-1},\boldsymbol{\theta}_\mathrm{p}) & G_3(q^{-1},\boldsymbol{\theta}_\mathrm{p}) & G_4(q^{-1},\boldsymbol{\theta}_\mathrm{p}) & G_2(q^{-1},\boldsymbol{\theta}_\mathrm{p}) \end{pmatrix} \boldsymbol{u}(k) + H(q^{-1},\boldsymbol{\theta}_\mathrm{p})e(k), \tag{8.1}$$

where $\begin{pmatrix} G_1(q^{-1},\boldsymbol{\theta}_\mathrm{p}) & G_3(q^{-1},\boldsymbol{\theta}_\mathrm{p}) & G_4(q^{-1},\boldsymbol{\theta}_\mathrm{p}) & G_2(q^{-1},\boldsymbol{\theta}_\mathrm{p}) \end{pmatrix} = \boldsymbol{G}(q^{-1},\boldsymbol{\theta}_\mathrm{p})$, and then if four transfer functions in $\boldsymbol{G}(q^{-1},\boldsymbol{\theta}_\mathrm{p})$ are the same and equal to $G_\mathrm{s}(q^{-1},\boldsymbol{\theta}_\mathrm{p})$, a SISO model with additive error can be obtained:

$$y(k) = G_\mathrm{s}(q^{-1},\boldsymbol{\theta}_\mathrm{p})u(k) + H(q^{-1},\boldsymbol{\theta}_\mathrm{p})e(k), \tag{8.2}$$

[1] A causal LTI system is referred to as stable (minimum-phase) if and only if all poles (zeros) are inside the unit circle, otherwise the system is referred to as unstable (nonminimum-phase).

where $u(k) = \sum_{i=1}^{4} P_i(k) \in \mathbb{R}$ which denotes the summation of four cylinder pressure signals, and the number of the poles in $G_{\rm s}(q^{-1}, \boldsymbol{\theta}_{\rm p})$ is $n_{\rm s}$.

The object of this chapter is to reconstruct the curve of the sum of four cylinder pressure signals by merely using the vibration signal $y(k)$.

Denote the transfer function of the discrete-time model $G_{\rm s}$ (corresponding to $G_{\rm s}(q^{-1}, \boldsymbol{\theta}_{\rm p})$) as

$$G_{\rm s}(z) = \boldsymbol{C}_{\rm s} \left(z\boldsymbol{I}_{n_{\rm s}} - \boldsymbol{A}_{\rm s}\right)^{-1} \boldsymbol{B}_{\rm s} + D_{\rm s} \triangleq \left(\begin{array}{c|c} \boldsymbol{A}_{\rm s} & \boldsymbol{B}_{\rm s} \\ \hline \boldsymbol{C}_{\rm s} & D_{\rm s} \end{array}\right), \tag{8.3}$$

and also denote the transfer operator of the model $G_{\rm s}$ as $G_{\rm s}(q^{-1})$ without displaying the parameter vector $\boldsymbol{\theta}_{\rm p}$.

The problem of cylinder pressure reconstruction can be transformed into the problem of the input reconstruction of the model $G_{\rm s}$. The input reconstruction problem is illustrated in Figure 8.1. In the figure, $v(k) = H(q^{-1}, \boldsymbol{\theta}_{\rm p})e(k)$.

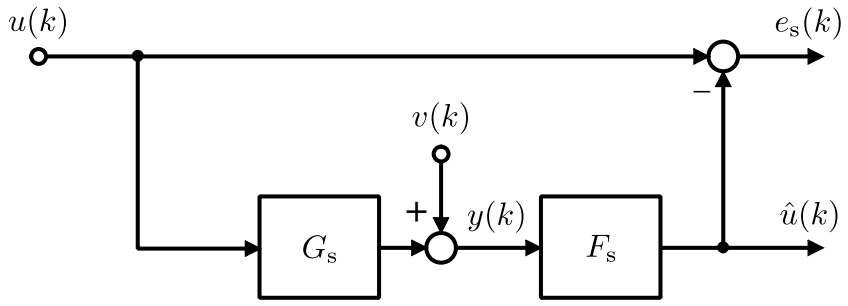

Figure 8.1: Input reconstruction problem.

The difference between $u(k)$ and its reconstructed value $\hat{u}(k)$ is represented as

$$\begin{aligned} e_{\rm s}(k) &= u(k) - \hat{u}(k) \\ &= u(k) - F_{\rm s}(q^{-1}) \left(G_{\rm s}(q^{-1})u(k) + v(k)\right) \\ &= \left(1 - F_{\rm s}(q^{-1})G_{\rm s}(q^{-1})\right) u(k) - F_{\rm s}(q^{-1})v(k), \end{aligned} \tag{8.4}$$

where $F_{\rm s}(q^{-1})$ denotes the transfer operator of the model $F_{\rm s}$.

According to Equation (8.4), it can be found that when

$$F_{\rm s}(z) = G_{\rm s}^{\dagger}(z), \tag{8.5}$$

where $F_{\rm s}(z)$ is the transfer function of the model $F_{\rm s}$, the contribution of $e_{\rm s}(k)$ induced by the input $u(k)$ can be eliminated, and the remained error $-F_{\rm s}(q^{-1})v(k)$ can be suppressed by enhancing both model quality of the model (8.2) and measurement accuracy of the signal $y(k)$. In this chapter, given the nominal model (8.2), our objective is to realize high-accuracy reconstruction of the sum $u(k)$ of four cylinder pressure signals by eliminating the error in $e_{\rm s}(k)$ induced by the input signal $u(k)$.

Based on [91], the inverse $G_{\rm s}^{\dagger}(z)$ can be represented as

$$G_{\rm s}^{\dagger}(z) \triangleq \left(\begin{array}{c|c} \boldsymbol{A}_{\rm s} - \boldsymbol{B}_{\rm s}D_{\rm s}^{\dagger}\boldsymbol{C} & -\boldsymbol{B}_{\rm s}D_{\rm s}^{\dagger} \\ \hline D_{\rm s}^{\dagger}\boldsymbol{C}_{\rm s} & D_{\rm s}^{\dagger} \end{array}\right). \tag{8.6}$$

At least two challenges are associated with the direct use of (8.6):

(i) $D_{\mathrm{s}}^{\dagger}$ does not exist.

(ii) Nonminimum-phase zeros in $G_{\mathrm{s}}(z)$.

Below, with the objective of this chapter mentioned above, a method is proposed to find the inverse model F_{s} such that $F_{\mathrm{s}}(z) = G_{\mathrm{s}}^{\dagger}(z)$ and the above two challenges can be handled. As a result, based on the obtained inverse model and the signal $y(k)$, the sum $u(k)$ of four cylinder pressure signals can be reconstructed. The specific cylinder pressure reconstruction algorithm is described using three parts:

(i) Derive an observer-based multisine signal reconstructor.

(ii) Based on the observer-based reconstructor, use the frequency-domain subspace identification method to identify the inverse model F_{s}.

(iii) Use the vibration signal $y(k)$ as the input of the identified inverse model $\hat{F}_{\mathrm{s}}$ to obtain the corresponding response, i.e., the reconstructed sum $\hat{u}(k)$ of four cylinder pressure signals.

The above three parts are depicted one by one in the sequent sections.

8.2.1 Observer-Based Multisine Signal Reconstructor

In this section, the observer-based multisine signal reconstructor is first derived, and then several properties of the reconstructor is illustrated.

8.2.1.1 Formulation of observer-based multisine signal reconstructor

As illustrated in Chapter 5, under stationary operating conditions the cylinder pressure signal can be modeled as the output of a state-space model (see the model (5.3)), so the sum $u(k)$ of four cylinder pressure signals can also be modeled as the output of a state-space model which is denoted as G_{u}, i.e.,

$$\begin{cases} \boldsymbol{x}_{\mathrm{u}}(k+1) = \boldsymbol{A}_{\mathrm{u}}(f(k))\boldsymbol{x}_{\mathrm{u}}(k), \\ u(k) = \boldsymbol{C}_{\mathrm{u}}\boldsymbol{x}_{\mathrm{u}}(k) + e_{\mathrm{u}}(k), \end{cases} \tag{8.7}$$

where the state vector $\boldsymbol{x}_{\mathrm{u}}(k) \in \mathbb{R}^{2n_{\mathrm{u}}+1}$, and $e_{\mathrm{u}}(k)$ denotes the error term.

The matrices $\boldsymbol{A}_{\mathrm{u}}(f(k))$ and $\boldsymbol{C}_{\mathrm{u}}$ are respectively given as

$$\boldsymbol{A}_{\mathrm{u}}(f(k)) = \left(\begin{array}{c|ccc} 1 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{A}_1(f(k)) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{A}_{n_{\mathrm{u}}}(f(k)) \end{array}\right) \tag{8.8}$$

and

$$\boldsymbol{C}_{\mathrm{u}} = \left(1 \;\middle|\; \boldsymbol{C}_1 \;\; \cdots \;\; \boldsymbol{C}_{n_{\mathrm{u}}}\right). \tag{8.9}$$

The individual block entries in these block matrices follow from the state-space representation of a frequency-modulated sine wave as

$$\boldsymbol{A}_i(f(k)) = \begin{pmatrix} \cos(2\pi i f(k) T_\mathrm{s}) & \sin(2\pi i f(k) T_\mathrm{s}) \\ -\sin(2\pi i f(k) T_\mathrm{s}) & \cos(2\pi i f(k) T_\mathrm{s}) \end{pmatrix} \tag{8.10}$$

and

$$\boldsymbol{C}_i = \begin{pmatrix} 1 & 0 \end{pmatrix}. \tag{8.11}$$

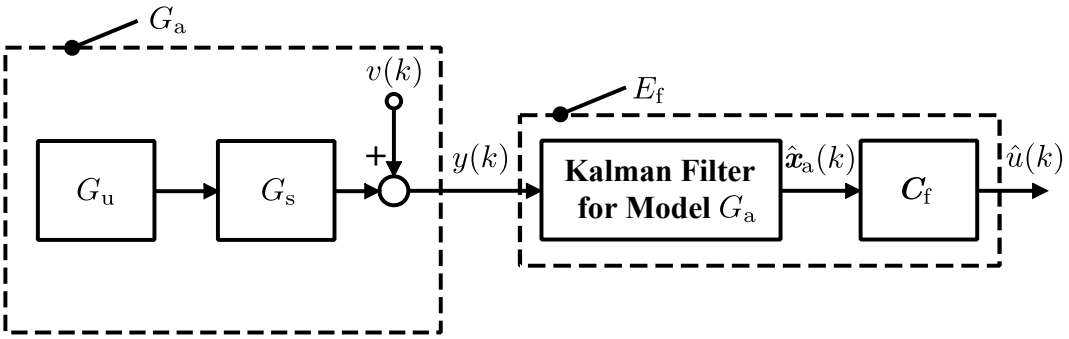

Figure 8.2: Kalman filter-based estimator E_f.

As seen in Figure 8.2, by augmenting the model (8.2) with the state of the model (8.7), an augmented model G_a can be finally obtained:

$$\begin{cases} \boldsymbol{x}_\mathrm{a}(k+1) = \boldsymbol{A}_\mathrm{a}(f(k))\boldsymbol{x}_\mathrm{a}(k), \\ y(k) = \boldsymbol{C}_\mathrm{a}\boldsymbol{x}_\mathrm{a}(k)) + e_\mathrm{a}(k), \end{cases} \tag{8.12}$$

where the state vector $\boldsymbol{x}_\mathrm{a}(k) \in \mathbb{R}^{n_\mathrm{a}}$, and the term $e_\mathrm{a}(k)$ denotes the output error which is induced by $v(k)$ and $e_\mathrm{u}(k)$.

The matrices $\boldsymbol{A}_\mathrm{a}(f(k))$ and $\boldsymbol{C}_\mathrm{a}$ are given as

$$\boldsymbol{A}_\mathrm{a}(f(k)) = \begin{pmatrix} \boldsymbol{A}_\mathrm{s} & \boldsymbol{B}_\mathrm{s}\boldsymbol{C}_\mathrm{u} \\ \boldsymbol{0} & \boldsymbol{A}_\mathrm{u}(f(k)) \end{pmatrix} \tag{8.13}$$

and

$$\boldsymbol{C}_\mathrm{a} = \begin{pmatrix} \boldsymbol{C}_\mathrm{s} & \boldsymbol{D}_\mathrm{s}\boldsymbol{C}_\mathrm{u} \end{pmatrix}, \tag{8.14}$$

respectively.

The state vector of the model (8.12) is given as

$$\boldsymbol{x}_\mathrm{a}(k) = \begin{pmatrix} \boldsymbol{x}_\mathrm{s}^\mathrm{T}(k) & \boldsymbol{x}_\mathrm{u}^\mathrm{T}(k) \end{pmatrix}^\mathrm{T}. \tag{8.15}$$

where $\boldsymbol{x}_\mathrm{s}(k) \in \mathbb{R}^{n_\mathrm{s}}$ denotes the state vector of the model G_s.

Under stationary operating conditions, with knowing the signal $y(k)$, the linear Kalman filter for the augmented model (8.12) is first implemented, then the state estimate $\hat{\boldsymbol{x}}_\mathrm{a}(k)$ can be obtained, and finally $\hat{u}(k) = \boldsymbol{C}_\mathrm{f}\hat{\boldsymbol{x}}_\mathrm{a}(k)$, where $\boldsymbol{C}_\mathrm{f} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{C}_\mathrm{u} \end{pmatrix}$. The above estimation process for the signal $u(k)$ is illustrated in Figure 8.2, and the Kalman filter-based estimator is denoted as E_f.

If only a number of frequency components in the signal $u(k)$ are of interest, e.g., at constant frequencies $0, f_\mathrm{b}, 2f_\mathrm{b}, \ldots, (M-1)f_\mathrm{b}$, and meanwhile if the sum of these frequency components are denoted as $u_\mathrm{b}(k)$, similar to the signal $u(k)$, the signal $u_\mathrm{b}(k)$ can be represented as the output of the following model G_b:

$$\begin{cases} \boldsymbol{x}_\mathrm{b}(k+1) = \boldsymbol{A}_\mathrm{b}\boldsymbol{x}_\mathrm{b}(k), \\ u_\mathrm{b}(k) = \boldsymbol{C}_\mathrm{b}\boldsymbol{x}_\mathrm{b}(k)), \end{cases} \tag{8.16}$$

where the state vector $\boldsymbol{x}_\mathrm{b}(k) \in \mathbb{R}^{2(M-1)+1}$.

The matrices $\boldsymbol{A}_\mathrm{b}$ and $\boldsymbol{C}_\mathrm{b}$ are respectively given as

$$\boldsymbol{A}_\mathrm{b} = \left(\begin{array}{c|ccc} 1 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{A}_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{A}_{M-1} \end{array}\right) \tag{8.17}$$

and

$$\boldsymbol{C}_\mathrm{b} = \left(\begin{array}{c|ccc} 1 & \boldsymbol{C}_1 & \cdots & \boldsymbol{C}_{M-1} \end{array}\right). \tag{8.18}$$

The individual block entries in these block matrices follow from the state-space representation of a frequency-modulated sine wave as

$$\boldsymbol{A}_i = \begin{pmatrix} \cos(2\pi i f_\mathrm{b} T_\mathrm{s}) & \sin(2\pi i f_\mathrm{b} T_\mathrm{s}) \\ -\sin(2\pi i f_\mathrm{b} T_\mathrm{s}) & \cos(2\pi i f_\mathrm{b} T_\mathrm{s}) \end{pmatrix} \tag{8.19}$$

and

$$\boldsymbol{C}_i = \begin{pmatrix} 1 & 0 \end{pmatrix}. \tag{8.20}$$

With the signal model (8.16), the signal u_b can be reconstructed by using the reconstructor E_o displayed in Figure 8.3.

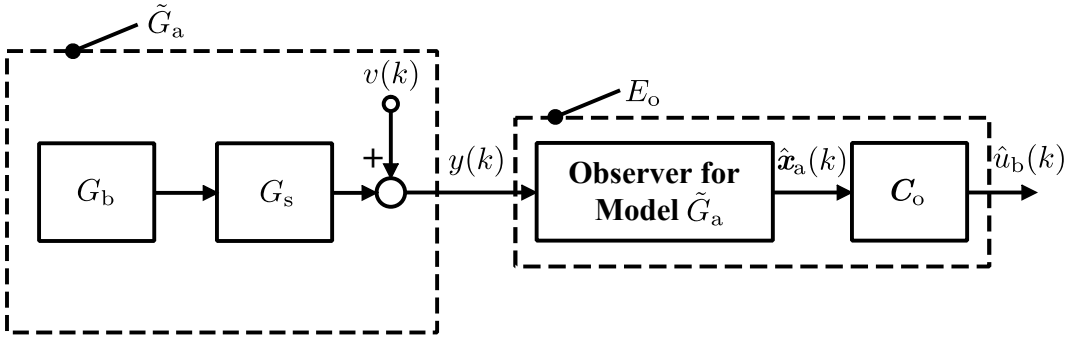

Figure 8.3: Observer-based reconstructor E_o.

As seen in Figure 8.3, by augmenting the model (8.2) with the state of the model (8.16), an augmented model $\tilde{G}_\mathrm{a}$ can be finally obtained:

$$\begin{cases} \tilde{\boldsymbol{x}}_\mathrm{a}(k+1) = \tilde{\boldsymbol{A}}_\mathrm{a}\tilde{\boldsymbol{x}}_\mathrm{a}(k), \\ y(k) = \tilde{\boldsymbol{C}}_\mathrm{a}\tilde{\boldsymbol{x}}_\mathrm{a}(k) + v(k), \end{cases} \tag{8.21}$$

where the state vector $\tilde{\boldsymbol{x}}_{\mathrm{a}}(k) \in \mathbb{R}^{\tilde{n}_{\mathrm{a}}}$.

Suppose that the error term $v(k)$ can be represented as

$$\begin{cases} \boldsymbol{x}_{\mathrm{v}}(k+1) = \tilde{\boldsymbol{A}}_{\mathrm{a}} \boldsymbol{x}_{\mathrm{v}}(k) + \boldsymbol{m}(k), \\ v(k) = \tilde{\boldsymbol{C}}_{\mathrm{a}} \boldsymbol{x}_{\mathrm{v}}(k) + n(k) \end{cases} \tag{8.22}$$

where the state vector $\boldsymbol{x}_{\mathrm{v}}(k) \in \mathbb{R}^{\tilde{n}_{\mathrm{a}}}$, and $\{\boldsymbol{m}(k)\}$ and $\{n(k)\}$ are white noise processes uncorrelated with $\tilde{\boldsymbol{x}}_{\mathrm{a}}(0)$ (a random variable with constant mean $\tilde{\boldsymbol{x}}_0$ and constant variance $\tilde{\boldsymbol{P}}_0$). The covariance matrices of $\{\boldsymbol{m}(k)\}$ and $\{n(k)\}$ are $\boldsymbol{Q}\delta_{kj}$ and $R\delta_{kj}$, respectively. Furthermore the values of both $\boldsymbol{Q}$ and R are assumed to be tunable.

The matrices $\tilde{\boldsymbol{A}}_{\mathrm{a}}$ and $\tilde{\boldsymbol{C}}_{\mathrm{a}}$ are respectively given as

$$\tilde{\boldsymbol{A}}_{\mathrm{a}} = \begin{pmatrix} \boldsymbol{A}_{\mathrm{s}} & \boldsymbol{B}_{\mathrm{s}}\boldsymbol{C}_{\mathrm{b}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathrm{b}} \end{pmatrix} \tag{8.23}$$

and

$$\tilde{\boldsymbol{C}}_{\mathrm{a}} = \begin{pmatrix} \boldsymbol{C}_{\mathrm{s}} & \boldsymbol{D}_{\mathrm{s}}\boldsymbol{C}_{\mathrm{b}} \end{pmatrix}. \tag{8.24}$$

The state vector of the model (8.21) is given as

$$\tilde{\boldsymbol{x}}_{\mathrm{a}}(k) = \begin{pmatrix} \boldsymbol{x}_{\mathrm{s}}^{\mathrm{T}}(k) & \boldsymbol{x}_{\mathrm{b}}^{\mathrm{T}}(k) \end{pmatrix}^{\mathrm{T}}. \tag{8.25}$$

In Figure 8.3, an observer-based reconstructor E_{o} is used to reconstruct the signal $u_{\mathrm{b}}(k)$. In the reconstructor, the observer (which is actually a steady-state Kalman filter) is used for the stochastic model $\tilde{G}_{\mathrm{a}}$. The reason why the observer is used rather than using the Kalman filter with time-variant Kalman gain is that a time-invariant reconstructor E_{o} is necessary in the proposed cylinder pressure reconstruction method.

The observer-based reconstructor E_{o} can be represented as

$$\begin{cases} \hat{\tilde{\boldsymbol{x}}}_{\mathrm{a}}(k+1) = \left(\tilde{\boldsymbol{A}}_{\mathrm{a}} - \boldsymbol{L}_{\mathrm{o}}\tilde{\boldsymbol{C}}_{\mathrm{a}}\right)\hat{\tilde{\boldsymbol{x}}}_{\mathrm{a}}(k) + \boldsymbol{L}_{\mathrm{o}} y(k), \\ \hat{u}_{\mathrm{b}}(k) = \boldsymbol{C}_{\mathrm{o}}\hat{\tilde{\boldsymbol{x}}}_{\mathrm{a}}(k), \end{cases} \tag{8.26}$$

where $\boldsymbol{C}_{\mathrm{o}} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{C}_{\mathrm{b}} \end{pmatrix}$.

As illustrated in Section 2.2.2 (see Remark 2.1), the stabilized gain $\boldsymbol{L}_{\mathrm{o}}$ can be derived based on using the steady-state Kalman gain of the steady-state Kalman filter for the augmented model $\tilde{G}_{\mathrm{a}}$. Specifically, choose the observer gain $\boldsymbol{L}_{\mathrm{o}}$ such that $\tilde{\boldsymbol{A}}_{\mathrm{a}} - \boldsymbol{L}_{\mathrm{o}}\tilde{\boldsymbol{C}}_{\mathrm{a}}$ can be guaranteed to be asymptotically stable, which requires specifying $\boldsymbol{Q}$ and R and solving an algebraic Riccati equation associated with the optimal filtering problem [68]. The existence of a unique stabilized solution of the algebraic Riccati equation can be realized under certain conditions which are illustrated in Section 2.2.3:

(i) $\boldsymbol{Q} \geq 0$ and $R > 0$.

(ii) $(\tilde{\boldsymbol{A}}_{\mathrm{a}}, \sqrt{\boldsymbol{Q}})$ is reachable.

(iii) $(\tilde{\boldsymbol{A}}_{\mathrm{a}}, \tilde{\boldsymbol{C}}_{\mathrm{a}})$ is detectable.

By tuning the values of $\boldsymbol{Q}$ and R, the first two conditions can easily be satisfied, and for the third condition, $(\tilde{\boldsymbol{A}}_{\mathrm{a}}, \tilde{\boldsymbol{C}}_{\mathrm{a}})$ can be guaranteed to be observable if the signal model is observable (see Appendix 5E).

Because the signal $u_{\mathrm{b}}(k)$ is a multisine signal, the above formulated reconstructor E_{o} is called observer-based multisine signal reconstructor.

8.2.1.2 Properties of the observer-based reconstructor

The existence of the FRF of the reconstructor E_{o} can be guaranteed because the reconstructor E_{o} can be guaranteed to be asymptotically stable by choosing a stabilized gain $\boldsymbol{L}_{\mathrm{o}}$.

Under the condition of the existence of the FRF of the reconstructor E_{o}, several properties of the reconstructor are illustrated as follows:

(i) The reconstructor is asymptotically optimal (i.e., as k tends to $+\infty$, the reconstructor is optimal) for the reconstruction of the signal u_{b} if the assumptions on the model $\tilde{G}_{\mathrm{a}}$ are satisfied, the model G_{s} is identified perfectly, the values of $\boldsymbol{Q}$ and R are true, and the reconstructor E_{o} is asymptotically stable.

(ii) The FRF values of the reconstructor are only observable at the frequencies specified in the matrix $\boldsymbol{A}_{\mathrm{b}}$.

The above properties are obvious. In the following example (i.e., Example 8.1), the second property of the reconstructor is illustrated.

Example 8.1. Given a SISO model G shown in Figure 8.4, the input signal $u(k) = \sum_{i=1}^{8} \sin(2\pi i f k T_{\mathrm{s}})$, and the transfer function $G(z) = \frac{z+0.5}{z^2-1.5z+0.7}$, where $T_{\mathrm{s}} = 0.001$ seconds and $f = 10$ Hz.

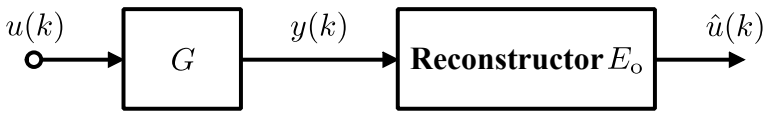

Figure 8.4: Reconstruction of $u(k)$.

Formulate the observer-based reconstructor E_{o} shown in Figure 8.4 for the signal $u(k)$ according to the reconstructor formulation steps described in Section 8.2.1.2. For the formulation of the reconstructor, set $\boldsymbol{Q} = 0.25\boldsymbol{I}_{18}$ and $R = 0.25$ such that an asymptotically stable reconstructor E_{o} can be obtained, and the Bode plot of the reconstructor E_{o} is illustrated in Figure 8.5.

With the response of the SISO model G and the formulated reconstructor E_{o}, the reconstruction result of the signal $u(k)$ can be obtained. Specifically, as shown in Figure 8.5, use the response $y(k)$ as the input of the reconstructor E_{o}, and then the output is the reconstructed value of $u(k)$. The reconstruction result of $u(k)$ is displayed in Figure 8.6.

According to the reconstruction result of $u(k)$ in Figure 8.6, it can be known that the proposed reconstructor of the input signal $u(k)$ of the SISO model G is effective. It can also be seen that a direct inversion of the model G is avodied for the input reconstruction purpose.

Based on the above formulated reconstructor E_{o}, in the remained part of this example, the second property of the reconstructor is explained and validated.

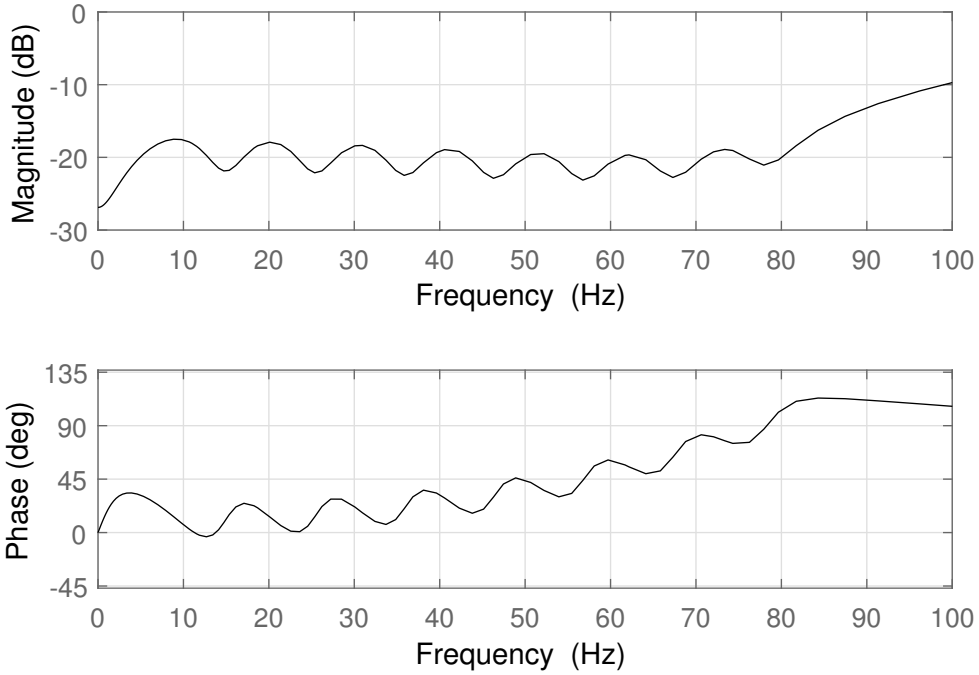

Figure 8.5: Bode plot the reconstructor E_o.

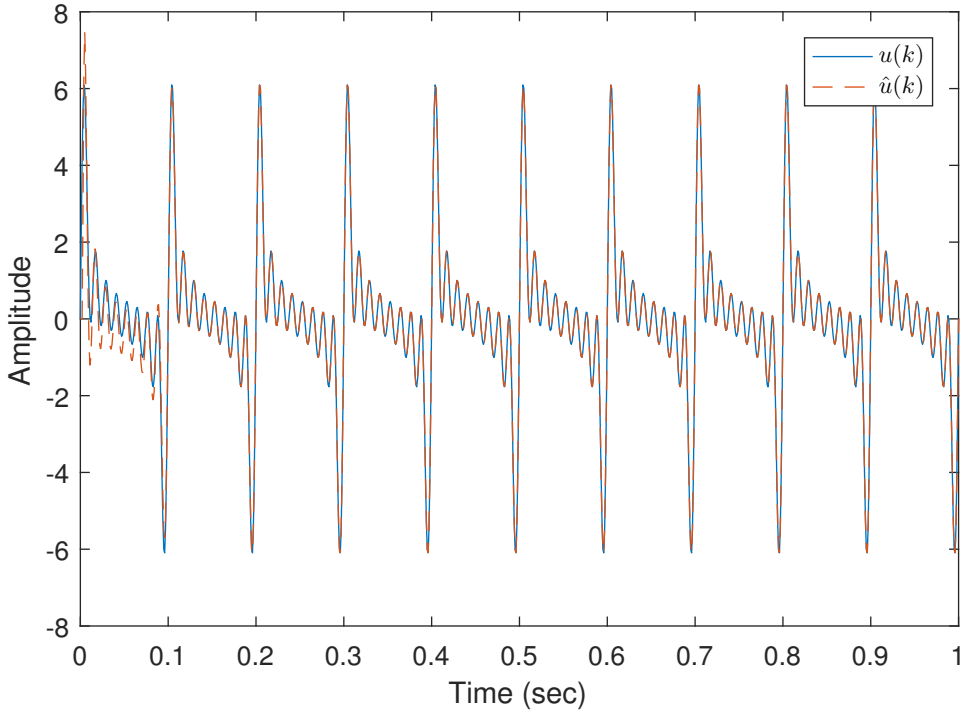

Figure 8.6: Reconstruction of $u(k)$.

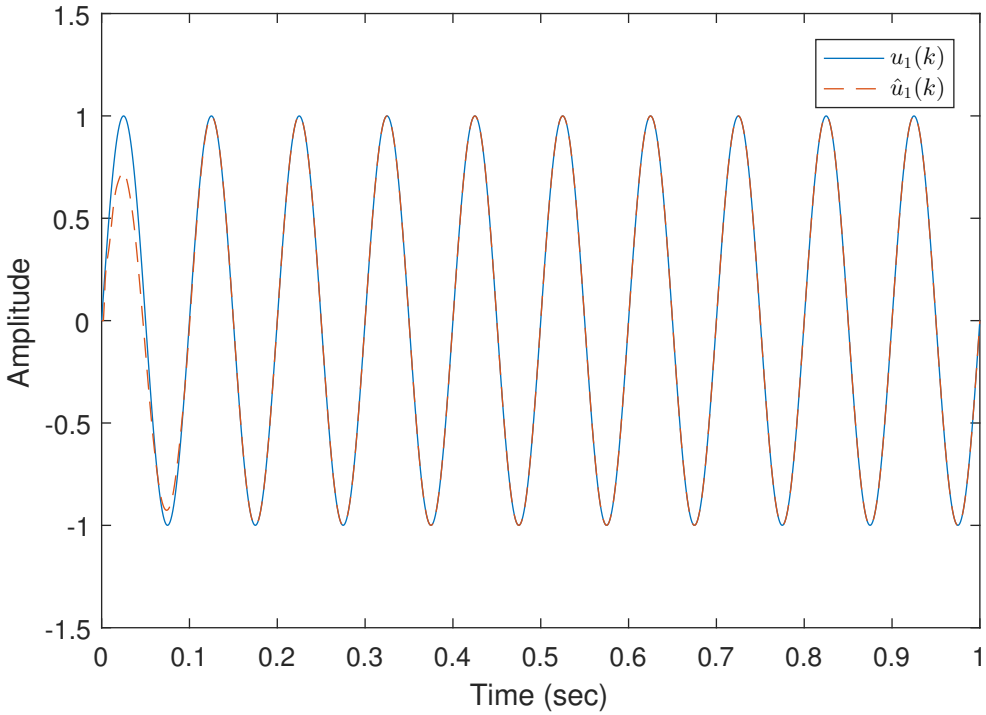

Figure 8.7: Reconstruction of $\sin(2\pi f_1 k T_\mathrm{s})$.

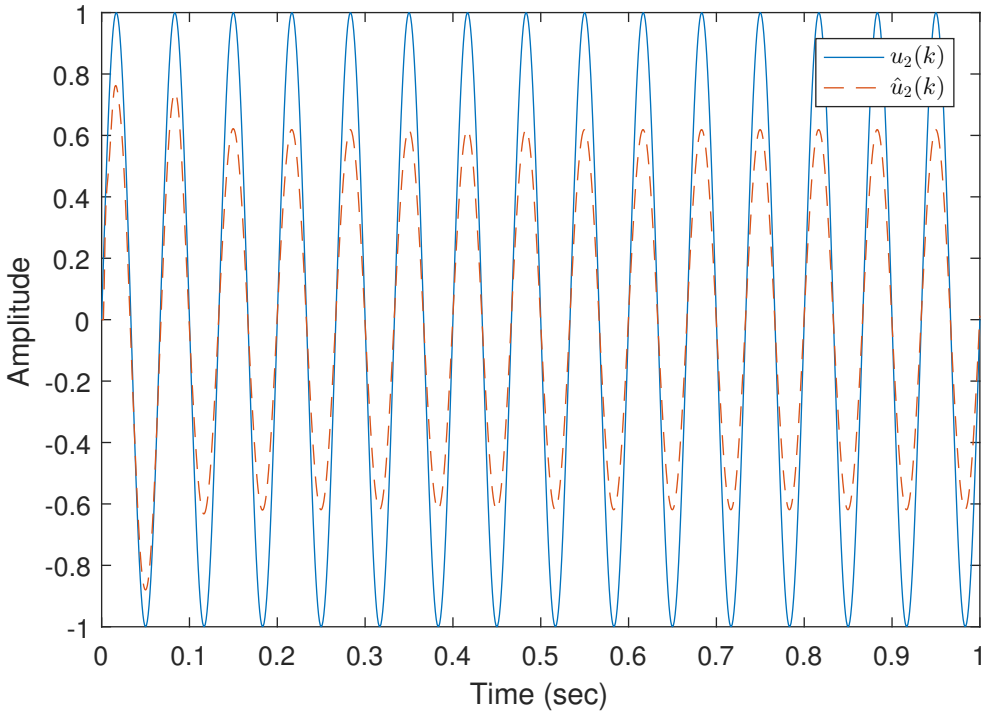

Figure 8.8: Reconstruction of $\sin(2\pi f_2 k T_\mathrm{s})$.

Given two signals $u_1(k) = \sin(2\pi f_1 k T_\mathrm{s})$ and $u_2(k) = \sin(2\pi f_2 k T_\mathrm{s})$, where $f_1 = 10$ Hz and $f_2 = 15$ Hz, use the two signals as the input of the SISO model G, respectively, then two output responses $y_1(k)$ and $y_2(k)$ can be obtained, respectively. Based on $y_1(k)$, $y_2(k)$, and the reconstructor E_o, the reconstructed values of $u_1(k)$ and $u_2(k)$ can be obtained. As seen in Figure 8.7 and Figure 8.8, the reconstruction of $u_2(k)$ is worse than the reconstruction of $u_1(k)$, which can give a validation of the second property of the reconstructor E_o. The frequency 10 Hz is included in the model of the signal $u(k)$ while the frequency 15 Hz is not included. The magnitudes and phases of the FRF values at frequencies 10 Hz and 15 Hz are shown in Figure 8.5. In summary, the unobservability of the FRF value at 15 Hz is caused by the unobservability of the state of signal model of $u(k)$ corresponding to the frequency 15 Hz not specified in the signal $u(k)$. Thus the concept of the so-called observability of the FRF value is explained and simultaneously the observability property is validated. □

8.2.2 Inverse Model Identification

As shown in Section 8.2.1.1, the formulation of the reconstructor E_o only depends on the model G_s, the frequency components in the signal u_b, $\boldsymbol{Q}$, and R, among which the frequency components are manually specified. The formulation indicates that the IO data of the model G_s is not used to formulate the reconstructor, therefore if only a number of frequency components of the sum $u(k)$ of four cylinder pressure signals are of interest, just specify the frequencies of these frequency components, and then involve these frequencies in the model of the reconstructor E_o. The FRF points of the E_o at the specified frequencies can be involved into a sequence which can be denoted as

$$(E_\mathrm{o}(\Omega_n))_{n=0}^{M-1}, \tag{8.27}$$

where $\Omega_n = e^{jw_n T_\mathrm{s}}$ with $j^2 = -1$ and $w_n = 2\pi n f_\mathrm{b}$.

Actually, for the cylinder pressure reconstruction, an infinite number of frequency components need to be involved, i.e., a frequency range (or say an engine speed range) is of interest. However, including an infinite number of frequency components is not possible. But a finite number of frequencies from the frequency range of interest can be specified, then a finite number of observable FRF points of the reconstructor E_o can be obtained, finally a frequency-domain curve fitting can be implemented such that the identified model can be used as an inverse model in the frequency range of interest. The above process is called inverse model identification. Frequency-domain subspace identification introduced in Section 2.6.2 can be used for the curving fitting purpose given the sequence $(E_\mathrm{o}(\Omega_n))_{n=0}^{M-1}$.

8.2.3 Cylinder Pressure Reconstruction Algorithm

Based on the identified model from inverse model identification method illustrated in Section 8.2.2, the inverse model can be identified, and then given the engine structural vibration signal as the input of the inverse model, the sum $u(k)$ of four cylinder pressure signals can be recursively reconstructed. Below, the specific cylinder pressure reconstruction algorithm is summarized in Algorithm 8.1.

Algorithm 8.1: Inverse model-based cylinder pressure reconstruction.

1 Identify the SISO model G_{s} use the time-domain subspace identification method illustrated in Chapter 4.
2 Specify a finite number of frequencies in the frequency range of interest, so we can obtain the sequence $(nf_{\mathrm{b}})_{n=0}^{M-1}$.
3 Formulate the reconstructor E_{o} depicted in Section 8.2.1.1.
4 Collect the FRF points of E_{o} at specified frequencies such that a sequence $(E_{\mathrm{o}}(\Omega_n))_{n=0}^{M-1}$ can be obtained.
5 Based on the sequence $(E_{\mathrm{o}}(\Omega_n))_{n=0}^{M-1}$, implement the frequency-domain subspace identification method, and finally we can obtain an identified model $\tilde{E}_{\mathrm{o}}$ with model order n_{E}.
6 $\hat{u}(k) = \tilde{E}_{\mathrm{o}}(q^{-1})y(k)$. // $\tilde{E}_{\mathrm{o}}(q^{-1})$ denotes the transfer operator of $\tilde{E}_{\mathrm{o}}$

8.3 Cylinder Pressure Reconstruction Results

According to the cylinder pressure reconstruction algorithm introduced in Section 8.2.3, the SISO model G_{s} was first identified by using the subspace identification method illustrated in Chapter 4, and two datasets (dataset A and dataset B) introduced in Chapter 3 were used for the identification and validation purposes, and the identification results are shown in Table 8.1. The measure (4.1) is used in the table.

Table 8.1: Identification results of the model G_{s}.

Model order	CVA	MOESP	CVA(s)	MOESP(s)
4	20.17 (9.41)	−2.67 (−9.69)	38.93 (42.97)	39.12 (40.45)
5	30.64 (26.26)	−26.00 (−89.66)	39.69 (42.15)	39.50 (42.38)
6	30.97 (20.96)	−1.25 (−3.40)	39.74 (43.15)	39.63 (42.84)
7	32.15 (26.90)	25.34 (25.23)	39.78 (42.75)	39.63 (41.78)
8	37.66 (35.17)	32.90 (24.15)	39.77 (43.35)	39.87 (25.83)
9	36.32 (34.58)	34.62 (33.98)	38.27 (35.61)	38.37 (36.36)
10	36.21 (36.89)	32.20 (9.79)	40.04 (40.28)	39.88 (42.65)

Note: The meanings of the abbreviations in the first row of this table can be referred to Section 4.3.1. The fitness in bracket denotes the cross-validation fitness while the fitness without bracket denotes the self-validation fitness.

According to the system identification results displayed in Table 8.1, the model with 6th order and CVA(s) was finally chosen as the model G_{s} used for the cylinder pressure reconstruction.

According to the cutoff frequency mentioned in Chapter 3, a finite number of frequencies in the frequency range of interest were specified, as a result the sequence $(nf_{\mathrm{b}})_{n=1}^{150}$ was obtained, where $f_{\mathrm{b}} = 6$ Hz. Then based on the identified SISO model, the specified frequencies, and the specified $\boldsymbol{Q}$ (equal to $0.25\boldsymbol{I}_{306}$) and R (equal to 0.25), the corresponding reconstructor E_{o} depicted in Section 8.2.1.1 was derived. The Bode plot of the obtained reconstructor E_{o} is shown in Figure 8.9.

Based on the derived reconstructor E_{o}, the sequence $(E_{\mathrm{o}}(\Omega_n))_{n=1}^{150}$ was obtained, where $\Omega_n = e^{j2\pi n f_{\mathrm{b}} T_{\mathrm{s}}}$. Based on the sequence $(E_{\mathrm{o}}(\Omega_n))_{n=1}^{150}$, the frequency-domain subspace identification method introduced in Section 4.3.2 was implemented. Specifically, the subspace algorithm was looped over different model orders (n_{E} was chosen from 4 to 10) and different block rows (r was chosen from $n_{\mathrm{E}}+1$ to 30) of the extended observability matrix. The stability of the identified models was forced by reflecting unstable poles into the unit disk [43]. The LM algorithm was carried out to optimize the model parameters obtained with the subspace algorithm. Finally, according to the cylinder pressure reconstruction results, the identified model with 8th order was used as the model $\tilde{E}_{\mathrm{o}}$. The Bode plot of $\tilde{E}_{\mathrm{o}}$ is illustrated in Figure 8.10.

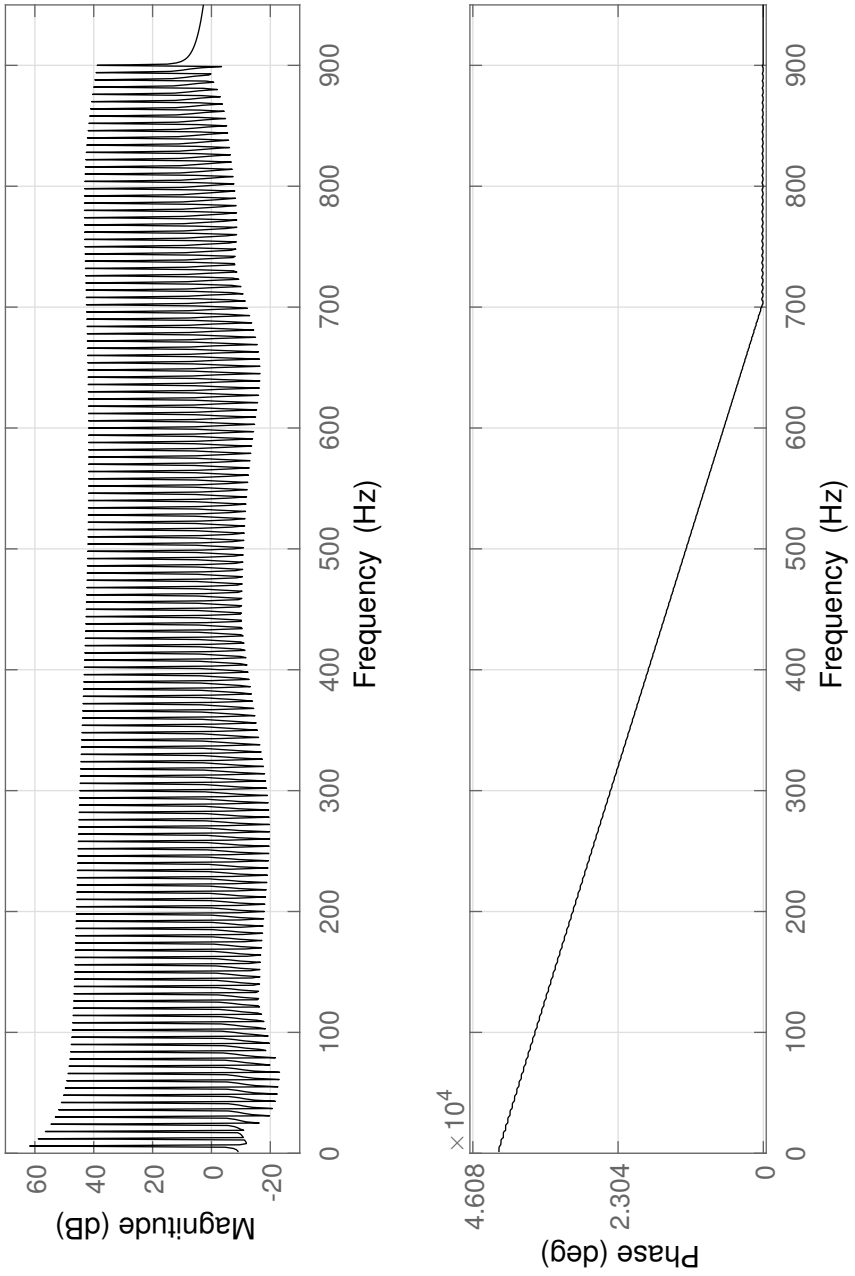

Figure 8.9: Bode plot of the obtained reconstructor E_{o}.

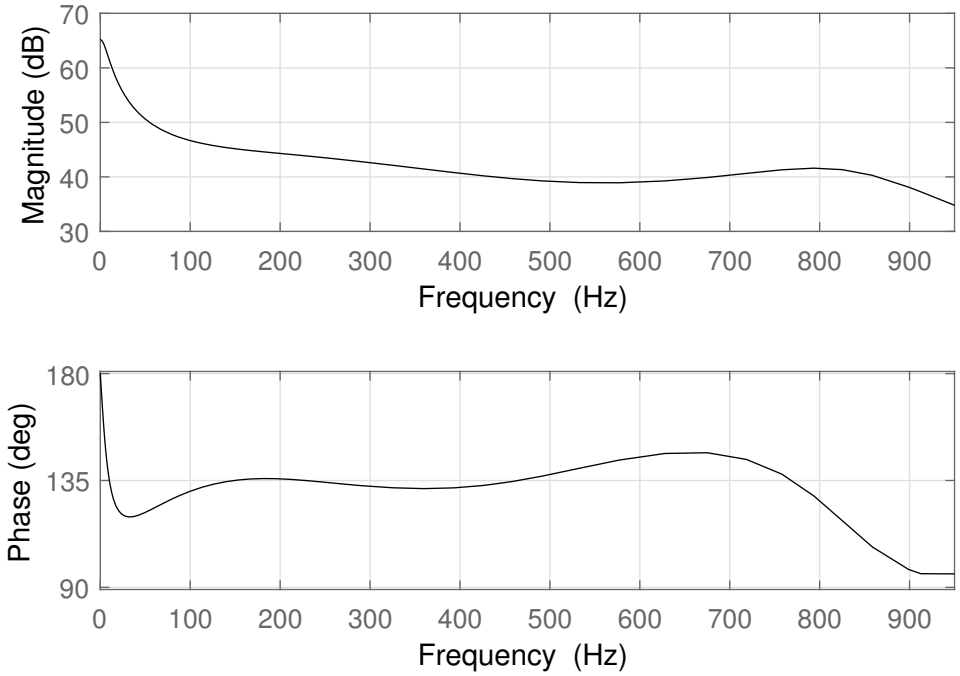

Figure 8.10: Bode plot of the model $\tilde{E}_{\mathrm{o}}$.

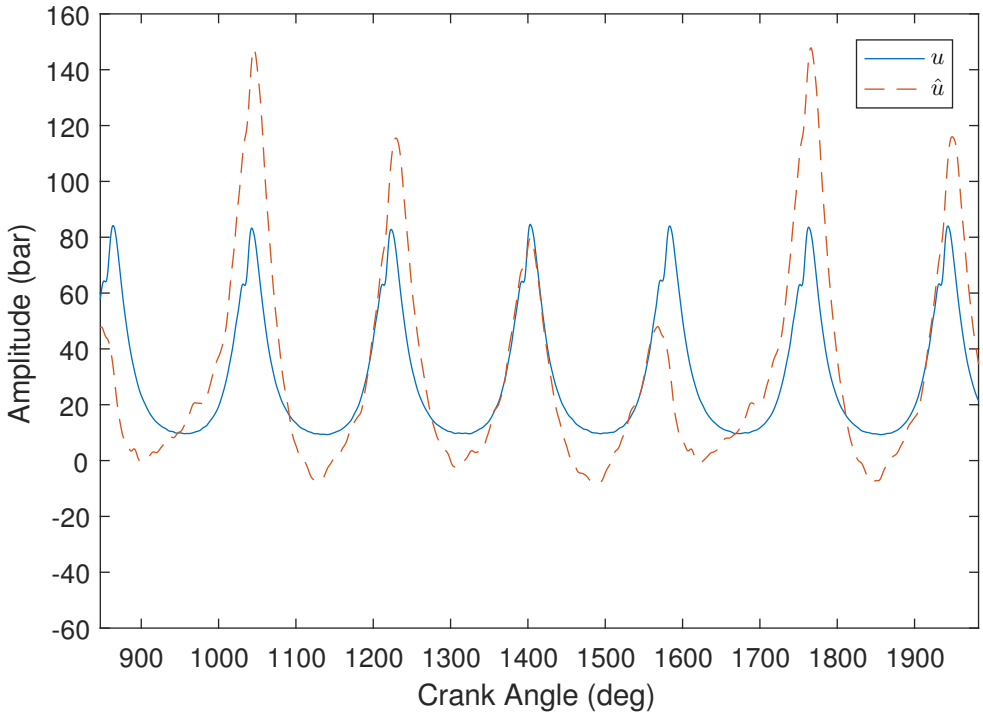

Figure 8.11: Cylinder pressure reconstruction under the stationary operating condition Z_2 (1200 rpm, 120 Nm) using Algorithm 8.1.

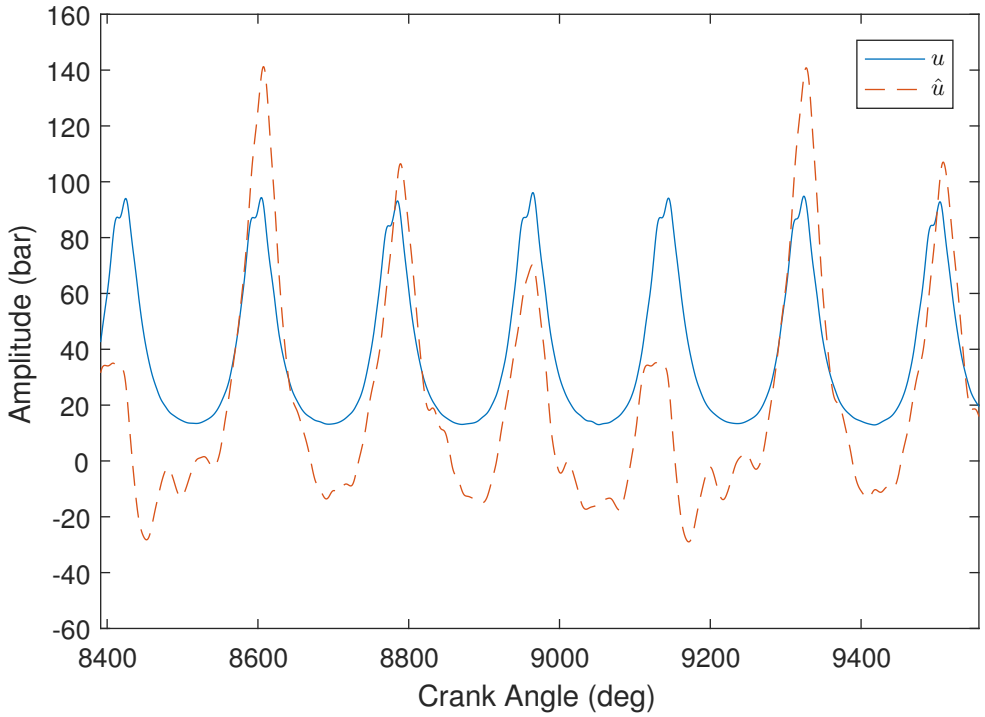

Figure 8.12: Cylinder pressure reconstruction under the stationary operating condition Z_5 (2100 rpm, 120 Nm) using Algorithm 8.1.

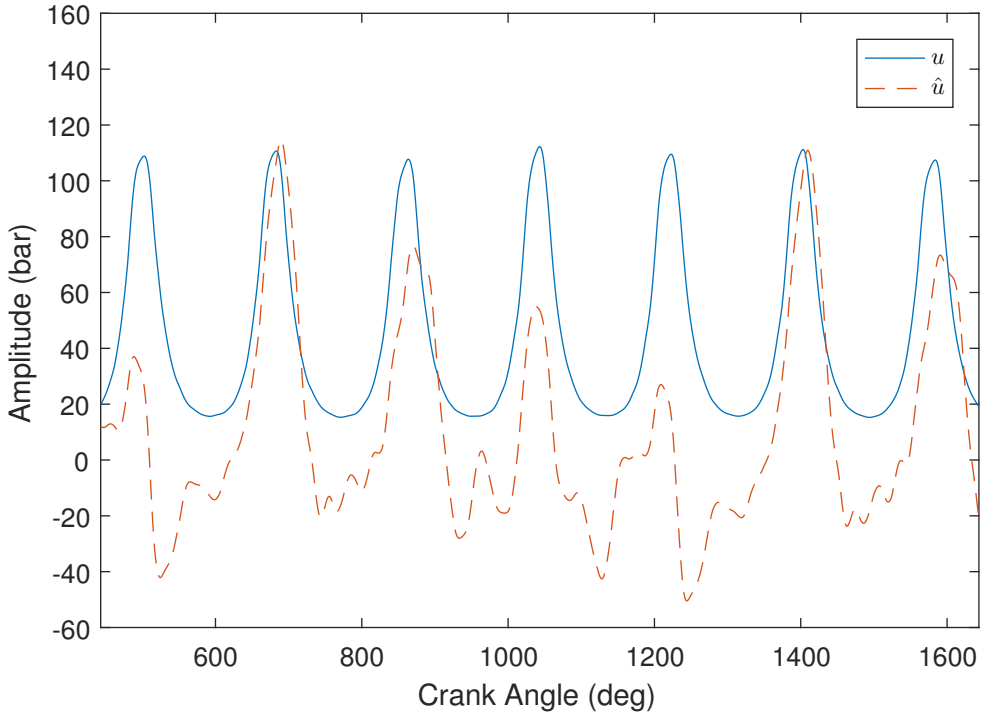

Figure 8.13: Cylinder pressure reconstruction under the stationary operating condition Z_8 (3000 rpm, 120 Nm) using Algorithm 8.1.

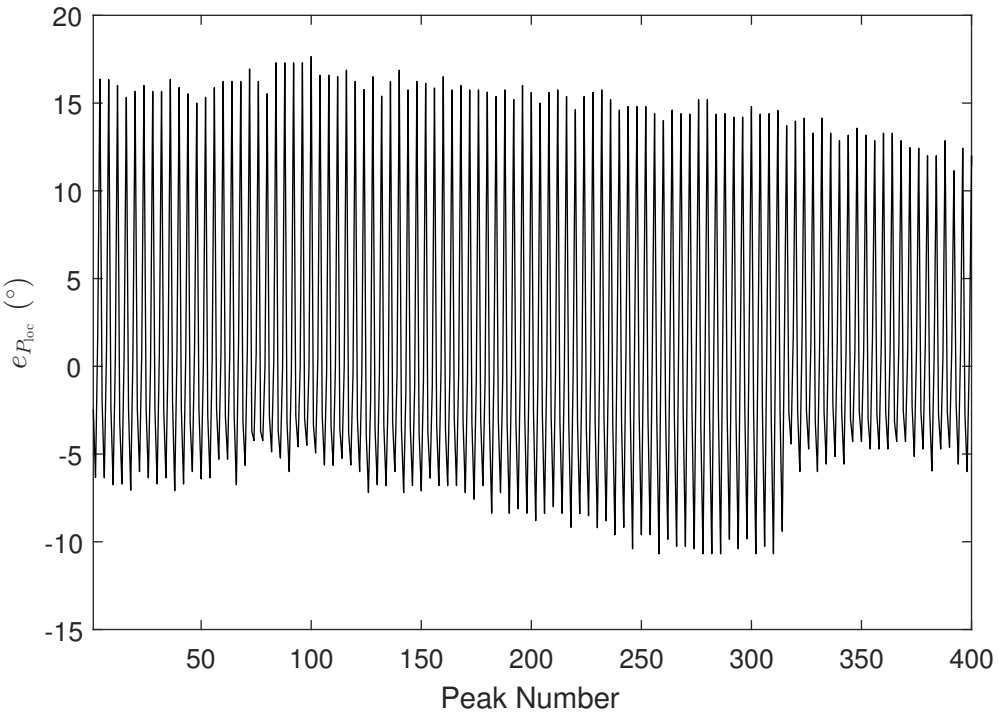

Figure 8.14: Cylinder pressure peak location reconstruction under the operating condition Z_{t} (just show 400 peak locations) using Algorithm 8.1.

With the identified model $\tilde{E}_{\mathrm{o}}$ and corresponding output vibration signal, the cylinder pressure signal reconstruction results can be obtained under three different stationary operating conditions. The results are displayed in Figure 8.11, Figure 8.12, and Figure 8.13. According to the cylinder pressure reconstruction results, it can be known that the reconstruction quality is not good in general, especially the pressure peak ($P_{\max}$) is not reconstructed well, while the reconstruction of the peak location (P_{loc}) is better. From the results, it can also be known that the proposed inverse model-based cylinder pressure reconstruction method at least can work even though the reconstruction precision is not good.

As an example, merely reconstructing the peak location under the operating condition Z_{t} is displayed in Figure 8.14. The reconstruction results in Figure 8.14 indicate that in practical applications, if the peak location is of interest only, and if the requirement of the accuracy of the reconstructed peak location is not high, the inverse model-based cylinder pressure reconstruction method proposed in this chapter can be chosen because this method can easily be realized.

In the end of this section, it should be noted that the cylinder pressure reconstruction results shown in the figures are not original results because 60 bar were added into them, i.e., the original DC offset is not accurate. Two possible reasons why the DC offset is not accurate are as follows:

(i) In the frequency-domain subspace identification method (see Algorithm 2.2), the FRF point value at the frequency 0 Hz was not well estimated.

(ii) The SISO model G_{s} was not identified well.

8.4 A Numerical Example for Validation

Because the identified SISO model $\hat{G}_{\mathrm{s}}$ does not have a good quality. The practical example in the above cannot sufficiently show that effectiveness of the inverse model-based input reconstruction. Here a numerical example is used to illustrate the effectiveness of the proposed input reconstruction method without the effect from the inaccurate SISO model.

Example 8.2. Given a transfer function model described by $G(z) = \frac{z}{z^2-1.5z+0.7}$. The input signal to be reconstructed is $u(k) = \sin(2\pi f_1 k T_{\mathrm{s}}) + \sin(2\pi f_2 k T_{\mathrm{s}}) + \sin(2\pi f_3 k T_{\mathrm{s}}) + \sin(2\pi f_4 k T_{\mathrm{s}})$, where $f_1 = 100$ Hz, $f_2 = 200$ Hz, $f_3 = 300$ Hz, and $f_4 = 400$ Hz, and the sampling period T_{s} is 1×10^{-5} seconds.

By observing the input signal $u(k)$, a finite number of frequencies were specified in the frequency range of interest, as a result the sequence $(nf_{\mathrm{b}})_{n=1}^{80}$ was obtained, where $f_{\mathrm{b}} = 6$ Hz. Set $\boldsymbol{Q} = 0.25\boldsymbol{I}_{162}$ and $R = 0.25$. Then the estimator E_{o} can be constructed, and the Bode plot the reconstructor E_{o} is illustrated in Figure 8.15. Afterwards, the FRF points of E_{o} at the specified frequencies can be obtained.

Based on the sequence $(E_{\mathrm{o}}(\Omega_n))_{n=1}^{80}$, the frequency-domain subspace identification method introduced in Section 4.3.2 was implemented to obtain the model $\tilde{E}_{\mathrm{o}}$. Specifically, the subspace algorithm was looped over different model orders (n_{E} was chosen from 2 to 8) and different block rows (r was chosen from $n_{\mathrm{E}} + 1$ to 30) of the extended observability matrix. The stability of the identified models was forced by reflecting unstable poles into the unit disk [43]. The LM algorithm was carried out to optimize the model parameters obtained with the subspace algorithm. The Bode plot the identified model $\tilde{E}_{\mathrm{o}}$ with 5th order is shown in Figure 8.16.

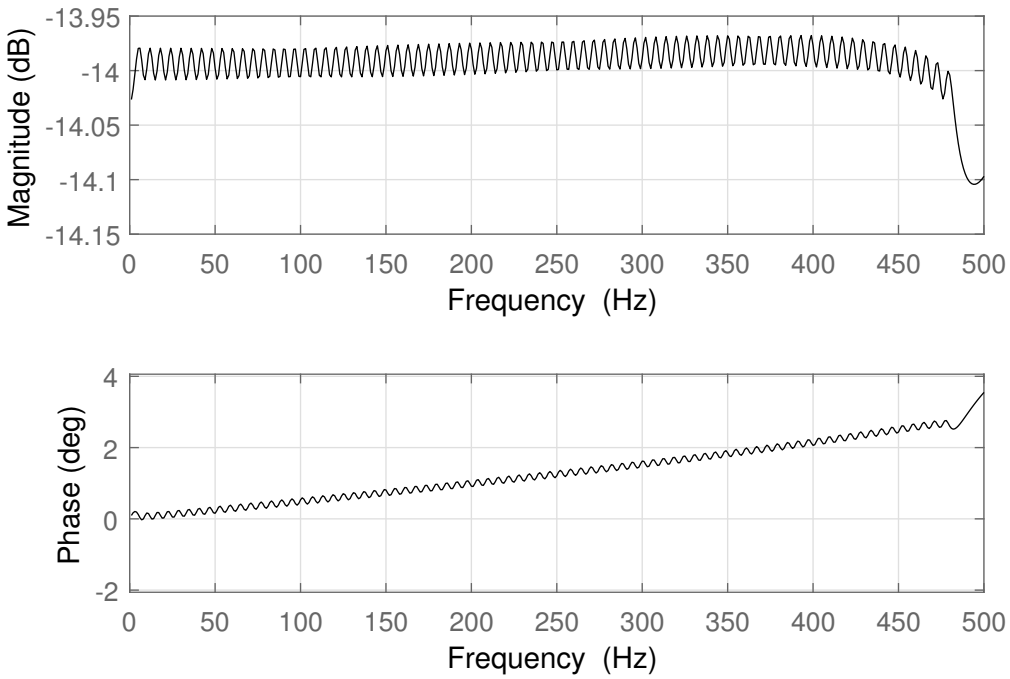

Figure 8.15: Bode plot of the model E_{o}.

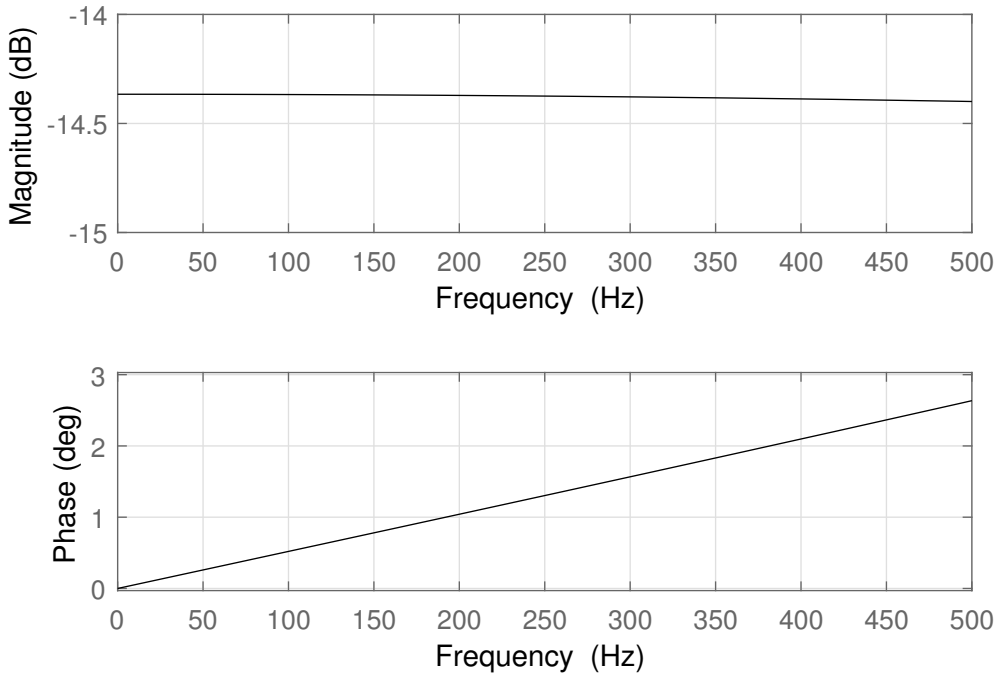

Figure 8.16: Bode plot of the model $\tilde{E}_{\mathrm{o}}$ with 5th order.

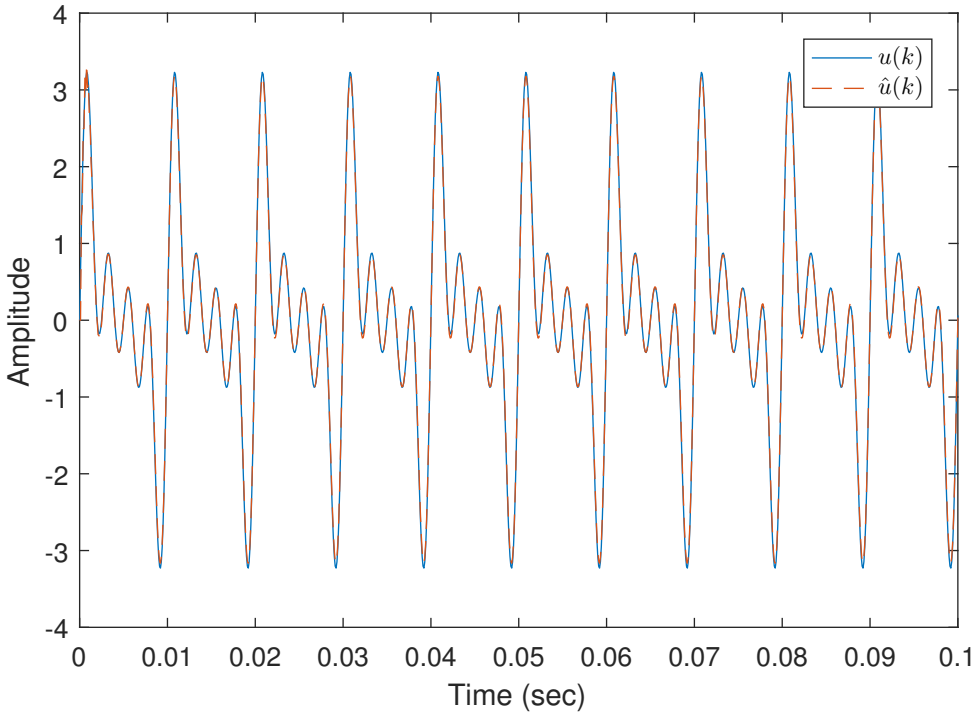

Figure 8.17: Input reconstruction.

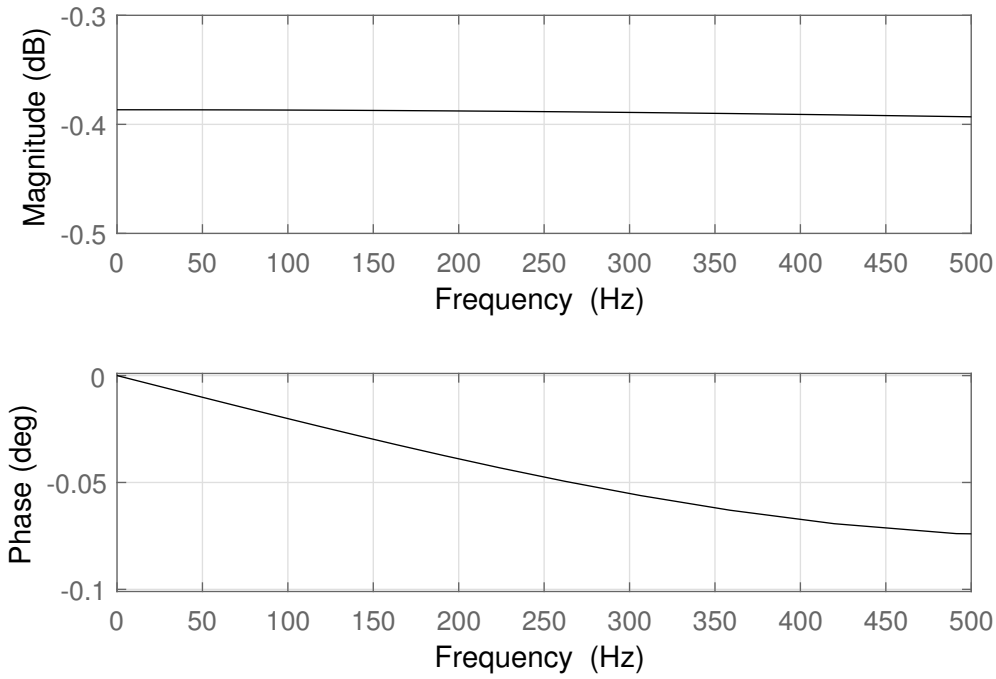

Figure 8.18: Bode plot of the model $\tilde{G}$.

By using the input signal $u(k)$ and the model $G(z)$, the simulated output was obtained, afterwards with the simulated output and the so-called inverse model $\tilde{E}_{\mathrm{o}}$, the reconstruction result of the input signal $u(k)$ is illustrated in Figure 8.17.

Furthermore, as shown in Figure 8.18, the bode plot of the model $\tilde{G}$ ($\tilde{G}(q^{-1}) = G(q^{-1})\tilde{E}_{\mathrm{o}}(q^{-1})$) can indicate that the inverse model $\tilde{E}_{\mathrm{o}}$ is effective in the frequency range from 0 Hz to 500 Hz. □

Example 8.2 demonstrates the applicability of the proposed inversion method for nonminimum-phase systems. Moreover, it also indicates that if the model $G(z)$ is accurate, the final input reconstruction result can be good.

8.5 Summary

The inverse model-based input reconstruction method proposed in this chapter can handle the problems (as depicted in Section 8.2) brought by direct inversion methods. From the formulation of the reconstructor E_{o} to the inverse model identification, the properties of the reconstructor E_{o} are illustrated while the asymptotic property of the inverse model identification is not analyzed. As seen in Section 8.2, actually the results of the identification of the inverse model can be validated by observing the quality of the input reconstruction results. According to the cylinder pressure reconstruction results, it can be known that at least the reconstructed pressure peak location can be used for some applications even though the reconstructed pressure peak is bad. Furthermore, according to the cylinder pressure reconstruction results in Section 8.3 and the numerical example illustrated in Section 8.4, the model quality can largely affect the effectiveness of the observer-based reconstructor.

Chapter 9

Discussion, Conclusions, and Future Research

9.1 Discussion and Conclusions

IC engine fuel economy, CO_2, and harmful emissions can all benefit from the use of closed-loop combustion control [1]. Additionally, misfire/knock detection systems can also monitor the health and performance of IC engines. But to be successful, both engine combustion control and misfire/knock detection systems require sensing of the instantaneous pressure signals on all cylinders. Cylinder pressure sensors are available for direct measurement and are fitted routinely during engine development and also on some production engines. However, the use of direct sensing is still a major problem for volume-produced engines owing to high sensor cost and limited lifetime. In addition, installing issues can either be problematic.

In order to avoid the drawbacks of the direct sensing of the cylinder pressure, numerous attempts have been made to find indirect ways to reconstruct cylinder pressure using information from other sensors fitted to engines. This process is also called virtual sensing process [3] and involves construction or identification of an inverse mathematical model linking the alternative sensor signal to cylinder pressure. The main types of alternative sensor signals have been the use of the engine speed signal or the engine structural vibration signal. The aim being that when a well-constructed or well-identified inverse model is supplied with the alternative sensor signal as an input, it gives an instantaneous reconstruction of the cylinder pressure as the output which can then be used for feedback combustion control or misfire detection or knock detection. However, there exist a number of difficulties in current cylinder pressure reconstruction methods, such as ill-conditioned inverse problems (brought from inverse model-based methods), spectrum leakage problems (caused by FRF-based methods), and the requirement of large number of IO data (when using ANNs).

With the objective which is to handle the existing problems in current cylinder pressure reconstruction methods, in this thesis, based on the identified models between four cylinder pressure signals and one engine structural vibration signal, the linear model-based, nonlinear model-based, and inverse model-based cylinder pressure reconstruction method are proposed. These methods are time-domain methods and are alternative to current existing cylinder pressure reconstruction methods. All the proposed methods can be cast in one unified framework which can be called input reconstruction. Below the reason why they can be cast in one unified framework is illustrated.

Prior to giving the reason, a summary of the system identification in Chapter 4 is briefly made. Both linear models and nonlinear models between four cylinder pressure signals and one vibration signal were identified based on using the collected data from the engine test bench shown in Figure 3.1. Linear models were obtained based on using both time-domain system identification and frequency-domain system identification. For the model identification, just the deterministic sub-models are of interest, and the second-order statistics of the noise process are regarded as tunable parameters.

With the identified models, the proposed cylinder pressure reconstruction methods can be conducted. Both linear model-based and nonlinear model-based cylinder pressure reconstruction methods are based on the conceptual time-varying transfer functions derived based on linear and nonlinear Kalman filtering. The calculated instantaneous engine cycle frequency is the time-variant parameter of the corresponding time-variant models. Actually, these time-variant models can be seen as inverse models of which the input is the engine structural vibration signal. One difficulty in the formulations of these time-variant models is the derivation of the approximated TVDB. While the inverse model-based cylinder pressure reconstruction method is based on a time-invariant transfer function which is derived based on the theory of observer and frequency-domain system identification. This method can build the relation between four cylinder pressure signals and one vibration signal and can avoid using the calculated instantaneous engine cycle frequency and the approximated TVDB. Generally, with the above analysis all the proposed cylinder pressure reconstruction methods can be cast in one unified framework which uses the output vibration signal to reconstruct the input cylinder pressure signals. The unified framework is illustrated in Figure 9.1.

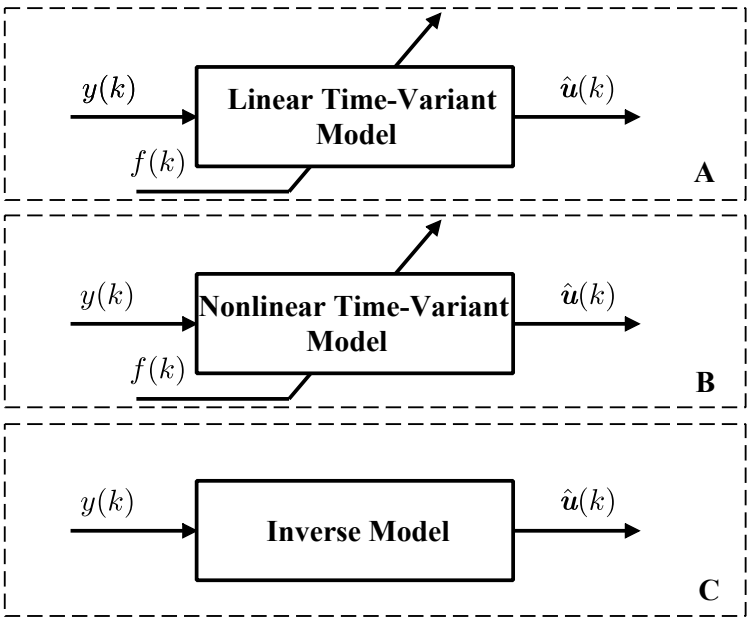

Figure 9.1: The framework of input reconstruction (A: cylinder pressure estimation methods in Chapter 5 and Chapter 6; B: cylinder pressure estimation method in Chapter 7; C: cylinder pressure reconstruction method in Chapter 8).

There are a number of factors affecting the cylinder pressure reconstruction precision, such as the approximated TVDB, the precision of the calculation of the instantaneous engine cycle frequency, and the quality of the identified model between four cylinder

pressure signals and one engine structural vibration signal. Most other researches have obtained $e_{P_{\max}}$ within the range of $5\% \sim 10\%$ and $e_{P_{\text{loc}}}$ within the range of $-5^\circ \sim +5^\circ$ [19], thus the estimated combustion metrics cylinder pressure peak and peak location in Chapter 5, Chapter 6, and Chapter 7 are comparable at some engine operating ranges. It should be noted that even though among all the proposed cylinder pressure reconstruction methods the estimated cylinder pressure curves are the worst based on the inverse model-based method, it is still possible for us to use inverse model-based cylinder pressure reconstruction method when high accuracy of the cylinder pressure reconstruction is not required because the inverse model-based method proposed in Chapter 8 can be easily realized.

Finally, a clarification should be made that in this thesis deriving alternative cylinder pressure reconstruction algorithms which can avoid the drawbacks in existing methods is the main objective, and the comparisons between the proposed methods in this thesis and other methods have not been made, and the enhancement of the accuracy of the reconstructed cylinder pressure using a specific cylinder pressure reconstruction method is not investigated.

9.2 Future Research

In the future, enhancing the precision of the proposed cylinder pressure reconstruction methods can be seen as the unified objective of all possible extensions of this thesis. Several possible extensions for the objective are illustrated as follows:

(i) As seen from the cylinder pressure reconstruction results, it can be known that the quality of the model between four cylinder pressure signals and one vibration signal can affect the cylinder pressure reconstruction accuracy. Other kinds of nonlinear model structures can be explored.

(ii) The effects of the tunable parameters on the cylinder pressure reconstruction in the chapters from Chapter 5 to Chapter 7 can be investigated. Additionally, the effects of the calculated/estimated instantaneous engine cycle frequency on the cylinder pressure reconstruction should also be explored.

(iii) Strategies for decreasing the fluctuations at the bottom of the reconstructed cylinder pressure curves displayed in the chapters from Chapter 5 to Chapter 7 should be studied.

(iv) For the nonlinear framework in Chapter 7, the coefficients of the nonlinearity could be included as states in the model used for EKF. Then, the coefficients could be estimated online so that the linear part of the model would remain constant whereas the nonlinear part would be adapted. Alternatively, because the Wiener model consists of a linear model and an output nonlinearity, the output of the linear model could be obtained by entering the vibration signal through the inverse nonlinearity. This would remove the nonlinearity and a linear Kalman filter could be used.

(v) The proposed inverse model-based cylinder pressure reconstruction method in Chapter 8 can only be used for SISO models. However, according to the system identification results, it is obvious that the identified SISO models do not have good fitness.

As a result the potential of the proposed inverse model-based cylinder pressure reconstruction method cannot be sufficiently reached. Therefore it is necessary to generalize the inverse model-based method so that it can be used for multiple-input single-output models.

(vi) According to the paper [5], the installing position of the accelerometer can affect the correlation between the cylinder pressure and vibration signal. Thus the cylinder pressure reconstruction with different installed positions of the accelerometer can be studied.

Appendix

A Proof of Proposition 5.1

Prior to giving the proof of Proposition 5.1, the definition of one-sided Laplace transform is given below.

Definition .1. [52] The one-sided Laplace transform (or say unilateral Laplace transform) of the signal $l(t)$ is given by

$$L(l(t)) = \int_{0^-}^{+\infty} l(t)e^{-st}\,dt. \tag{A.1}$$

for all s such that the result of the integration is well defined, denoted by the region of convergence (ROC). □

By using one-sided Laplace transform and its inverse, the constant time delay operation can be represented as

$$\begin{aligned} g(t-\tau) &= L^{-1}(e^{-\tau s}L(g(t))) \\ &= L^{-1}\left(\left(1+(-\tau s)+\frac{1}{2}(-\tau s)^2+\ldots\right)L(g(t))\right) \\ &= e^{-\tau\left(\frac{d}{dt}\right)}g(t), \end{aligned} \tag{A.2}$$

where $\frac{d}{dt}$ denotes the first order differential operator with respect to time t. τ is a fixed dead time. $g(t-\tau)$ denotes the delay of $g(t)$, and $g(t)=0$ for $t<0$.

If the dead time τ is time-varying, i.e., a time-dependent dead time $\tau(t)$, according to [82] and Equation (A.2), the following equation can be obtained:

$$g(t-\tau(t)) = \frac{1}{1-\frac{d}{dt}\tau(t)}\left(g(t)-\frac{d}{dt}(T(t)g(t))+\frac{1}{2}\frac{d^2}{dt^2}(T(t)g(t))\mp\ldots\right), \tag{A.3}$$

where $\frac{d^2}{dt^2}$ denotes the second order differential operator with respect to the time t, and $T(t)$ represents an intermediate variable.

$f_{\mathrm{a}}(t)$ can be seen as an approximation of $f(t)$ because $f_{\mathrm{a}}(t)$ is the zero-order hold of $f(k)$. Additionally, it is obvious that $f_{\mathrm{a}}(t)$ is a staircase signal, of which the function is regarded as a right-continuous function. Then come back to our problem, as aforementioned, the delay between every two consecutive cylinder pressure signals in time domain can be represented as $\tau(t)=\frac{1}{4f(t)}$, thus $\tau(t)\approx\frac{1}{4f_{\mathrm{a}}(t)}$. The dead time $\frac{1}{4f_{\mathrm{a}}(t)}$ is approximately used as the delay between every two consecutive cylinder pressure signals.

Finally, based on the above analysis and Equation (A.3), if $g(t)$ denotes the cylinder

No. 1 cylinder pressure signal and $g(t-\tau(t))$ denotes the cylinder No. 3 cylinder pressure signal, $g(t-\tau(t)) \approx e^{-\frac{1}{4f_{\mathrm{a}}(t)}\left(\frac{d}{dt}\right)} g(t)$ without the additional terms $\frac{d}{dt}\tau(t)$ and $T(t)$. So the TVDB transfer function can be approximately denoted as $e^{-\frac{1}{4f_{\mathrm{a}}(t)}s}$.

B Proof of Proposition 5.2

The transfer function of the model (5.19) can be represented as

$$H_{\mathrm{d}}^{\mathrm{a}}(s) = \boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1} \left(s\boldsymbol{I}_{\nu} - \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}\right)^{-1} \boldsymbol{B}_{\mathrm{d}}, \tag{B.1}$$

and by substituting $T_{\mathrm{d}}(t)s$ into s in Equation (B.1), the transfer function of the model (5.15) can be obtained as

$$G_{\mathrm{d}}^{\mathrm{a}}(s, T_{\mathrm{d}}(t)) = \frac{\boldsymbol{C}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}}{T_{\mathrm{d}}(t)} \left(s\boldsymbol{I}_{\nu} - \frac{\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}}{T_{\mathrm{d}}(t)}\right)^{-1} \boldsymbol{B}_{\mathrm{d}}. \tag{B.2}$$

C Numerical Approximations

The state transition matrix of the model (5.15) can be represented as

$$\boldsymbol{\Phi}(t, t_0) = \boldsymbol{I}_{\nu} + \int_{t_0}^{t} \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(h_1))\, dh_1 + \int_{t_0}^{t}\int_{t_0}^{h_1} \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(h_1))\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(h_2))\, dh_2 dh_1 + \ldots, \tag{C.1}$$

furthermore, $\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))$ and $\int_{t_0}^{t} \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(h_1))\, dh_1$ commute (see Appendix 5B), i.e.,

$$\boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t)) \int_{t_0}^{t} \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(h_1))\, dh_1 = \int_{t_0}^{t} \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(h_1))\, dh_1 \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t)), \tag{C.2}$$

then based on Equation (C.2), the series in Equation (C.1) can be rewritten as

$$\boldsymbol{\Phi}(t, t_0) = \sum_{i=0}^{+\infty} \frac{1}{i!} \left(\int_{t_0}^{t} \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(h_1))\, dh_1\right)^{i}. \tag{C.3}$$

Based on Equation (C.3), it is obvious that $\boldsymbol{\Phi}(t, t_0) = e^{\int_{t_0}^{t} \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(h_1))\, dh_1}$. Perform the rectangle integration [28] on the right side of Equation (C.3) with $t_0 = kT_{\mathrm{s}}$, $t = (k+1)T_{\mathrm{s}}$, and the maximum value of i chosen as l_{a}, the following formula can be obtained:

$$\begin{aligned}
\boldsymbol{G}_{\mathrm{d}}(T_{\mathrm{d}}(k)) &= \boldsymbol{\Phi}\left((k+1)T_{\mathrm{s}}, kT_{\mathrm{s}}\right) \\
&\approx \boldsymbol{I}_{\nu} + 4T_{\mathrm{s}}f_{\mathrm{a}}(k)\ \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1} + \frac{1}{2!}\left(4T_{\mathrm{s}}f_{\mathrm{a}}(k)\ \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}\right)^{2} \\
&\quad + \ldots + \frac{1}{l_{\mathrm{a}}!}\left(4T_{\mathrm{s}}f_{\mathrm{a}}(k)\ \boldsymbol{A}_{\mathrm{d}}(T_{\mathrm{d}}(t))|_{T_{\mathrm{d}}(t)=1}\right)^{l_{\mathrm{a}}}.
\end{aligned} \tag{C.4}$$

The numerical calculation for the matrix $\boldsymbol{H}_{\mathrm{d}}(T_{\mathrm{d}}(k))$ is illustrated as follows:

$$\boldsymbol{H}_{\mathrm{d}}(T_{\mathrm{d}}(k)) = \int_{kT_{\mathrm{s}}}^{(k+1)T_{\mathrm{s}}} \boldsymbol{\Phi}\left((k+1)T_{\mathrm{s}}, \beta\right) \boldsymbol{B}_{\mathrm{d}}\, d\beta$$

$$
\begin{aligned}
&\approx \int_{kT_\mathrm{s}}^{(k+1)T_\mathrm{s}} \left(\boldsymbol{I}_\nu + \left(\int_\beta^{(k+1)T_\mathrm{s}} 4f(t)\,dt \right) \boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right. \\
&\quad + \frac{1}{2!} \left(\left(\int_\beta^{(k+1)T_\mathrm{s}} 4f(t)\,dt \right) \boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right)^2 \\
&\quad \left. + \ldots + \frac{1}{l_\mathrm{a}!} \left(\left(\int_\beta^{(k+1)T_\mathrm{s}} 4f(t)\,dt \right) \boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right)^{l_\mathrm{a}} \right) \boldsymbol{B}_\mathrm{d}\,d\beta \\
&\approx \int_{kT_\mathrm{s}}^{(k+1)T_\mathrm{s}} \left(\boldsymbol{I}_\nu + 4f_\mathrm{a}(\beta)((k+1)T_\mathrm{s}-\beta)\,\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right. \\
&\quad + \frac{1}{2!} \left(4f_\mathrm{a}(\beta)\left((k+1)T_\mathrm{s}-\beta\right)\,\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right)^2 \\
&\quad \left. + \ldots + \frac{1}{l_\mathrm{a}!} \left(4f_\mathrm{a}(\beta)\left((k+1)T_\mathrm{s}-\beta\right)\,\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right)^{l_\mathrm{a}} \right) \boldsymbol{B}_\mathrm{d}\,d\beta \\
&= \left(T_\mathrm{s}\boldsymbol{I}_\nu + \left(\frac{1}{2!} 4f_\mathrm{a}(\beta)((k+1)T_\mathrm{s}-\beta)^2\,\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right) \right|_{kT_\mathrm{s}}^{(k+1)T_\mathrm{s}} \\
&\quad + \left(\frac{1}{3!} 4f_\mathrm{a}(\beta)\left((k+1)T_\mathrm{s}-\beta\right)^3\,\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right) \Big|_{kT_\mathrm{s}}^{(k+1)T_\mathrm{s}} \\
&\quad + \ldots + \left(\frac{1}{l_\mathrm{a}!} 4f_\mathrm{a}(\beta)((k+1)T_\mathrm{s}-\beta)^{l_\mathrm{a}}\,\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right) \Big|_{kT_\mathrm{s}}^{(k+1)T_\mathrm{s}} \Bigg) \boldsymbol{B}_\mathrm{d} \\
&= T_\mathrm{s} \left(\boldsymbol{I}_\nu + \frac{1}{2!} \left(4T_\mathrm{s} f_\mathrm{a}(k)\,\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right) \right. \\
&\quad \left. + \ldots + \frac{1}{l_\mathrm{a}!} \left(4T_\mathrm{s} f_\mathrm{a}(k)\,\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1} \right)^{l_\mathrm{a}-1} \right) \boldsymbol{B}_\mathrm{d}. \qquad \text{(C.5)}
\end{aligned}
$$

It should be noted that for the second-time approximation in Equation (C.5), the rectangle integration is also implemented.

D Balanced Realization

As introduced in Proposition 5.1, $f_\mathrm{a}(t)$ is right-continuous, so according to $T_\mathrm{d}(t) = \frac{1}{4f_\mathrm{a}(t)}$, it can be found that $T_\mathrm{d}(t)$ is also right-continuous, i.e., in every sampling period T_s, the value of $T_\mathrm{d}(t)$ is constant. So in the model (5.15) the matrices $\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))$ and $\boldsymbol{C}_\mathrm{d}(T_\mathrm{d}(t))$, which depend on $T_\mathrm{d}(t)$, are constant in every sampling period T_s.

For the balanced realization of an LTV model, the balanced transformation matrix is time-variant [92]. Denote the balanced transformation matrix of the model (5.15) as $\boldsymbol{T}_\mathrm{b}(t)$, because the matrices $\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))$, $\boldsymbol{B}_\mathrm{d}$, and $\boldsymbol{C}_\mathrm{d}(T_\mathrm{d}(t))$ are constant in every sampling period, the matrix $\boldsymbol{T}_\mathrm{b}(t)$ is constant in every sampling period either. The balanced realization of the model (5.15) can be represented as

$$
\begin{cases}
\dot{\boldsymbol{x}}_\mathrm{d}(t) = \boldsymbol{T}_\mathrm{b}^{-1}(t)\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))\boldsymbol{T}_\mathrm{b}(t)\boldsymbol{x}_\mathrm{d}(t) - \boldsymbol{T}_\mathrm{b}^{-1}(t)\dot{\boldsymbol{T}}_\mathrm{b}(t)\boldsymbol{x}_\mathrm{d}(t) + \boldsymbol{T}_\mathrm{b}^{-1}(t)\boldsymbol{B}_\mathrm{d}u_\mathrm{d}(t), \\
y_\mathrm{d}(t) = \boldsymbol{C}_\mathrm{d}(T_\mathrm{d}(t))\boldsymbol{T}_\mathrm{b}(t)\boldsymbol{x}_\mathrm{d}(t).
\end{cases} \qquad \text{(D.1)}
$$

The functions of $\boldsymbol{T}_\mathrm{b}(t)$, $\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))$, and $\boldsymbol{C}_\mathrm{d}(T_\mathrm{d}(t))$ can be regarded as right-continuous functions. Define the break point between one sampling period and its consecutive period

as t_b, then

$$\begin{aligned}\dot{\boldsymbol{x}}_\mathrm{d}(t_\mathrm{b}) &= \lim_{t\to t_\mathrm{b}^+}\dot{\boldsymbol{x}}_\mathrm{d}(t),\\ &= \boldsymbol{T}_\mathrm{b}^{-1}(t_\mathrm{b}^+)\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t_\mathrm{b}^+))\boldsymbol{T}_\mathrm{b}(t_\mathrm{b}^+)\boldsymbol{x}_\mathrm{d}(t_\mathrm{b}^+) - \boldsymbol{T}_\mathrm{b}^{-1}(t_\mathrm{b}^+)\dot{\boldsymbol{T}}_\mathrm{b}(t_\mathrm{b}^+)\boldsymbol{x}_\mathrm{d}(t_\mathrm{b}^+) + \boldsymbol{T}_\mathrm{b}^{-1}(t_\mathrm{b}^+)\boldsymbol{B}_\mathrm{d}u_\mathrm{d}(t_\mathrm{b}^+),\\ &= \boldsymbol{T}_\mathrm{b}^{-1}(t_\mathrm{b}^+)\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t_\mathrm{b}^+))\boldsymbol{T}_\mathrm{b}(t_\mathrm{b}^+)\boldsymbol{x}_\mathrm{d}(t_\mathrm{b}^+) + \boldsymbol{T}_\mathrm{b}^{-1}(t_\mathrm{b}^+)\boldsymbol{B}_\mathrm{d}u_\mathrm{d}(t_\mathrm{b}^+), \qquad \text{(D.2)}\end{aligned}$$

where t_b^+ denotes the time t tends to t_b from right, and $\dot{\boldsymbol{T}}_\mathrm{b}(t_\mathrm{b}^+) = \boldsymbol{0}$, and

$$\begin{aligned}y_\mathrm{d}(t_\mathrm{b}) &= \lim_{t\to t_\mathrm{b}^+} y_\mathrm{d}(t),\\ &= \boldsymbol{C}_\mathrm{d}(T_\mathrm{d}(t_\mathrm{b}^+))\boldsymbol{T}_\mathrm{b}(t_\mathrm{b}^+)\boldsymbol{x}_\mathrm{d}(t_\mathrm{b}^+). \qquad \text{(D.3)}\end{aligned}$$

According to Equations (D.2) and (D.3), it can be known that in every sampling period T_s the balanced realization of the model (5.15) is the same as the balanced realization of LTI models.

Based on Equations (5.21) and (5.22), Equations (D.2) and (D.3) can be transformed into

$$\begin{cases}\dot{\boldsymbol{x}}_\mathrm{d}(t_\mathrm{b}) = \boldsymbol{T}_\mathrm{b}^{-1}(t_\mathrm{b}^+)\dfrac{\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1}}{T_\mathrm{d}(t_\mathrm{b}^+)}\boldsymbol{T}_\mathrm{b}(t_\mathrm{b}^+)\boldsymbol{x}_\mathrm{d}(t_\mathrm{b}^+) + \boldsymbol{T}_\mathrm{b}^{-1}(t_\mathrm{b}^+)\boldsymbol{B}_\mathrm{d}u_\mathrm{d}(t_\mathrm{b}^+),\\ y_\mathrm{d}(t_\mathrm{b}) = \dfrac{\boldsymbol{C}_\mathrm{d}(T_\mathrm{d}(t))|_{T_\mathrm{d}(t)=1}}{T_\mathrm{d}(t_\mathrm{b}^+)}\boldsymbol{T}_\mathrm{b}(t_\mathrm{b}^+)\boldsymbol{x}_\mathrm{d}(t_\mathrm{b}^+).\end{cases} \qquad \text{(D.4)}$$

By observing (D.4), the problem of the balanced realization of the model (5.15) can be transformed into the problem of the balanced realization of the model (5.19).

E Reachability and Observability of the TVDB Model

According to the calculation of the instantaneous engine cycle frequency $f(k)$ in Figure 3.5, in each $\Delta\theta^\mathrm{p}$, the model (5.15) is a LTI model, and then it is minimally-realized in every $\Delta\theta^\mathrm{p}$. According to the preservation of reachability and observability under sampling with a zero-order hold [27], it can be known that in every $\Delta\theta^\mathrm{p}$, a controllable, observable, continuous-time, time-invariant linear model (5.15) with input that passes through a period-T_s sampler and zero-order hold yields a reachable, observable, discrete-time, time-invariant linear model (5.23) if

$$\lambda_k - \lambda_i \neq \frac{2\pi q j}{T_\mathrm{s}}, \quad q = \pm 1, \pm 2, \ldots, \qquad \text{(E.1)}$$

where $j^2 = -1$, for every pair of eigenvalues of λ_k and λ_j of the matrix $\boldsymbol{A}_\mathrm{d}(T_\mathrm{d}(t))$.

Based on Equation (E.1), it can be known that in every $\Delta\theta^\mathrm{p}$, by choosing a suitable sampling period T_s, both reachability and observability can be preserved, i.e., the model (5.23) is both reachable and observable, thus the reachability and observability of the model (5.28) can be hold if a suitable l_a is selected. Finally, based on the above analysis, the LTV model (5.28) is uniformly reachable and uniformly observable. One possible way

to choose the values of T_s and l_a is that make the sampling period T_s smaller while make the value of l_a larger.

F Observability of the Augmented Model

The observability of the time-variant augmented model (5.35) (i.e., the model $\boldsymbol{G}_\mathrm{a}$) can be verified by observing the observability of the augmented model in every angle $\Delta\theta^\mathrm{p}$, and the reason is that in every angle $\Delta\theta^\mathrm{p}$ the augmented model is a LTI model, and then if the observability of the augmented model can be guaranteed in every angle $\Delta\theta^\mathrm{p}$, the time-variant augmented model (5.35) will be uniformly observable. Below, the specific observability verification of the augmented model in every angle $\Delta\theta^\mathrm{p}$ is illustrated.

From the paper [93], it can be known that the stochastic augmented model (5.35) is observable if and only if its deterministic sub-model is observable. Therefore, the observability of $(\boldsymbol{A}_\mathrm{a}(f(k)), \boldsymbol{C}_\mathrm{a}(f(k)))$ can be verify just in every angle $\Delta\theta^\mathrm{p}$. Use ith $\Delta\theta^\mathrm{p}$ (i.e., $\Delta\theta_i^\mathrm{p}$) as an example, and in this angle duration, according to the SISO model (5.29), suppose the following equations can be obtained:

$$\boldsymbol{x}_\mathrm{s}(1) = \boldsymbol{A}_\mathrm{s}(f(k))\boldsymbol{x}_\mathrm{s}(0) + \boldsymbol{B}_\mathrm{s}(f(k))\boldsymbol{C}_{P_1}\boldsymbol{x}_{P_1}(0), \tag{F.1}$$

$$\boldsymbol{x}_\mathrm{s}(2) = \boldsymbol{A}_\mathrm{s}(f(k))\boldsymbol{x}_\mathrm{s}(1) + \boldsymbol{B}_\mathrm{s}(f(k))\boldsymbol{C}_{P_1}\boldsymbol{A}_{P_1}(f(k))\boldsymbol{x}_{P_1}(0), \tag{F.2}$$

$$\dots$$

$$\boldsymbol{x}_\mathrm{s}(n_\mathrm{u}) = \boldsymbol{A}_\mathrm{s}(f(k))\boldsymbol{x}_\mathrm{s}(2n_\mathrm{p}) + \boldsymbol{B}_\mathrm{s}(f(k))\boldsymbol{C}_{P_1}\boldsymbol{A}_{P_1}^{2n_\mathrm{p}}(f(k))\boldsymbol{x}_{P_1}(0), \tag{F.3}$$

then the following equation can be obtained:

$$\boldsymbol{O}\boldsymbol{x}_{P_1}(0) = \boldsymbol{J} \tag{F.4}$$

with

$$\boldsymbol{O} = \begin{pmatrix} \boldsymbol{B}\boldsymbol{C}_{P_1} \\ \boldsymbol{B}\boldsymbol{C}_{P_1}\boldsymbol{A}_{P_1}(f(k)) \\ \vdots \\ \boldsymbol{B}\boldsymbol{C}_{P_1}\boldsymbol{A}_{P_1}^{2n_\mathrm{p}}(f(k)) \end{pmatrix} \tag{F.5}$$

and

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{x}_\mathrm{s}(1) - \boldsymbol{A}_\mathrm{s}(f(k))\boldsymbol{x}_\mathrm{s}(0) \\ \boldsymbol{x}_\mathrm{s}(2) - \boldsymbol{A}_\mathrm{s}(f(k))\boldsymbol{x}_\mathrm{s}(1) \\ \vdots \\ \boldsymbol{x}_\mathrm{s}(2n_\mathrm{p}+1) - \boldsymbol{A}_\mathrm{s}(f(k))\boldsymbol{x}_\mathrm{s}(2n_\mathrm{p}) \end{pmatrix}. \tag{F.6}$$

By choosing suitable values of T_s and l_a, as illustrated in Appendix 5D the SISO model (5.29) can be guaranteed to be observable in every $\Delta\theta^\mathrm{p}$, therefore the initial state vector $\boldsymbol{x}_\mathrm{s}(0)$ of the SISO model (5.29) can always be determined by the response sequence $(y(k))_{k=0}^{2n_\mathrm{p}}$, as a result the state $\boldsymbol{x}_\mathrm{s}(k)$ for $k \geq 1$ can either be determined by using the state transition equation, as a result, the value matrix $\boldsymbol{J}$ can be obtained by using the response sequence $(y(k))_{k=0}^{2n_\mathrm{p}}$. Additionally, if the signal model (5.3) is observable, it can be found that the matrix $\boldsymbol{O}$ has a Moore-Penrose inverse, as a result,

$$\boldsymbol{x}_{P_1}(0) = \boldsymbol{O}^\dagger\boldsymbol{J}. \tag{F.7}$$

Based on the above analysis, in every $\Delta\theta^{\mathrm{p}}$ the observability problem of the augmented model (5.35) can be transformed into the one of the signal model (5.3). If the matrix $\boldsymbol{A}_{P_1}(f(k))$ has different eigenvalues, the observability of the signal model (5.3) can be guaranteed. So in order to guarantee the observability, different frequency components can be included in the input signal $P_1(k)$ to make sure that $\boldsymbol{A}_m(f(k)) \neq \boldsymbol{A}_n(f(k))$ in $\boldsymbol{A}_{P_1}(f(k))$, where $m \neq n$. Thus in every angle $\Delta\theta^{\mathrm{p}}$, $(\boldsymbol{A}_{\mathrm{a}}(f(k)), \boldsymbol{C}_{\mathrm{a}}(f(k)))$ can be guaranteed to be observable under certain conditions (i.e., suitable values of T_{s} and l_{a}, and different frequency components involved in $P_1(k)$). As a result, under certain conditions the time-variant augmented model (5.35) can be guaranteed to be uniformly observable.

G Cylinder Pressure Reconstruction Results

In this section, cylinder pressure reconstruction results obtained in Chapter 5, Chapter 6, and Chapter 7 are illustrated. There are two types of results to be shown in this section. The first type of the results displays the mean and standard deviation of the two characteristic parameters $e_{P_{\max}}$ and $e_{P_{\mathrm{loc}}}$ under stationary operating conditions, while the second type of the results shows the curves of the two characteristic parameters $e_{P_{\max}}$ and $e_{P_{\mathrm{loc}}}$ under non-stationary operating conditions. It should be mentioned that the tables from Table G.1 to Table G.3 correspond to the cylinder pressure estimation results based on the time-domain identified model while the tables from Table G.4 to Table G.6 correspond to the cylinder pressure estimation results based on the frequency-domain identified model. Furthermore, the figures from Figure G.1 to Figure G.3 correspond to the cylinder pressure estimation results based on the time-domain identified model while the figures from Figure G.4 to Figure G.6 correspond to the cylinder pressure estimation results based on the frequency-domain identified model.

Table G.1: Cylinder No. 3 pressure estimation results evaluation based on Algorithm 5.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	-7.10%	2.46%	10.29%	-0.39%	-3.02%	-2.06%	20.10%	14.29%	16.13%
$\sigma(e_{P_{\max}})$	1.39%	1.38%	10.45%	1.89%	1.84%	1.70%	2.27%	2.65%	1.76%
$\mu(e_{P_{\mathrm{loc}}})$	$-0.66°$	$-2.14°$	$-1.51°$	$0.47°$	$-1.13°$	$-1.53°$	$-2.59°$	$-0.65°$	$0.68°$
$\sigma(e_{P_{\mathrm{loc}}})$	$0.77°$	$0.27°$	$0.88°$	$2.35°$	$0.61°$	$0.63°$	$3.21°$	$1.88°$	$0.72°$

Table G.2: Cylinder No. 4 pressure estimation results evaluation based on Algorithm 5.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	-8.62%	2.23%	9.24%	-1.53%	-5.62%	-4.79%	18.46%	10.99%	12.96%
$\sigma(e_{P_{\max}})$	1.39%	1.32%	12.04%	1.98%	1.81%	1.62%	2.37%	2.79%	2.07%
$\mu(e_{P_{\mathrm{loc}}})$	$-0.90°$	$-2.22°$	$-1.61°$	$0.38°$	$-1.07°$	$-2.04°$	$-3.45°$	$-0.50°$	$0.46°$
$\sigma(e_{P_{\mathrm{loc}}})$	$0.65°$	$0.31°$	$0.70°$	$1.45°$	$0.75°$	$1.93°$	$3.64°$	$1.26°$	$0.81°$

Table G.3: Cylinder No. 2 pressure estimation results evaluation based on Algorithm 5.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	−9.40%	2.08%	12.33%	0.98%	−2.05%	−1.61%	21.17%	14.88%	15.85%
$\sigma(e_{P_{\max}})$	1.21%	1.54%	5.52%	1.86%	1.71%	1.50%	2.11%	2.80%	1.95%
$\mu(e_{P_{\text{loc}}})$	−1.34°	−2.41°	−2.33°	−0.72°	−1.72°	−3.31°	−3.22°	−1.24°	0.10°
$\sigma(e_{P_{\text{loc}}})$	0.68°	0.36°	1.02°	3.26°	0.72°	1.82°	2.71°	1.25°	1.44°

Table G.4: Cylinder No. 3 pressure estimation results evaluation based on Algorithm 5.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	−7.25%	1.58%	7.85%	−12.20%	−8.42%	−6.21%	2.02%	1.52%	6.48%
$\sigma(e_{P_{\max}})$	0.83%	1.01%	4.69%	0.68%	0.52%	0.60%	0.36%	0.45%	0.44%
$\mu(e_{P_{\text{loc}}})$	0.43°	−1.82°	−1.01°	5.12°	0.07°	0.67°	1.44°	2.50°	2.98°
$\sigma(e_{P_{\text{loc}}})$	1.31°	1.79°	3.76°	1.62°	0.78°	0.96°	0.47°	0.45°	0.61°

Table G.5: Cylinder No. 4 pressure estimation results evaluation based on Algorithm 5.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	−9.29%	0.95%	7.04%	−13.47%	−11.33%	−9.44%	−0.18%	−2.10%	3.32%
$\sigma(e_{P_{\max}})$	0.89%	0.74%	4.39%	0.58%	0.53%	0.59%	0.38%	0.37%	0.36%
$\mu(e_{P_{\text{loc}}})$	−0.05°	−0.65°	−0.85°	5.32°	0.22°	1.06°	1.23°	2.98°	3.18°
$\sigma(e_{P_{\text{loc}}})$	1.37°	2.12°	4.23°	1.06°	0.82°	1.03°	0.59°	0.50°	0.56°

Table G.6: Cylinder No. 2 pressure estimation results evaluation based on Algorithm 5.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	−10.08%	2.20%	9.47%	−10.59%	−7.66%	−6.00%	3.45%	2.26%	6.56%
$\sigma(e_{P_{\max}})$	0.87%	1.18%	2.51%	0.59%	0.46%	0.60%	0.42%	0.43%	0.48%
$\mu(e_{P_{\text{loc}}})$	−0.68°	0.14°	−1.90°	4.15°	−0.08°	1.17°	0.67°	2.03°	2.51°
$\sigma(e_{P_{\text{loc}}})$	1.46°	2.18°	3.82°	3.03°	0.84°	1.05°	0.52°	0.51°	0.57°

Table G.7: Cylinder No. 3 pressure estimation results evaluation based on Algorithm 6.2 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	−18.42%	14.75%	−10.73%	−4.24%	−5.70%	−3.55%	10.80%	6.47%	9.54%
$\sigma(e_{P_{\max}})$	0.98%	0.71%	4.32%	0.45%	0.42%	0.50%	0.33%	0.46%	0.44%
$\mu(e_{P_{\text{loc}}})$	−1.68°	−3.09°	−1.76°	1.38°	−0.61°	−0.97°	−3.63°	0.45°	0.82°
$\sigma(e_{P_{\text{loc}}})$	0.73°	0.59°	5.15°	0.72°	0.79°	0.58°	1.04°	1.13°	1.48°

Table G.8: Cylinder No. 4 pressure estimation results evaluation based on Algorithm 6.2 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	−20.85%	16.13%	−11.93%	−5.60%	−8.54%	−6.62%	8.94%	3.05%	6.42%
$\sigma(e_{P_{\max}})$	1.12%	0.80%	5.33%	0.35%	0.47%	0.56%	0.35%	0.50%	0.45%
$\mu(e_{P_{\text{loc}}})$	−2.67°	−2.93°	−2.48°	1.50°	−0.41°	−0.62°	−3.63°	0.87°	0.92°
$\sigma(e_{P_{\text{loc}}})$	0.73°	0.58°	4.48°	0.66°	0.81°	0.61°	1.19°	1.16°	1.62°

Table G.9: Cylinder No. 2 pressure estimation results evaluation based on Algorithm 6.2 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	−22.43%	14.93%	−10.80%	−1.26%	−4.73%	−3.20%	12.19%	7.20%	9.65%
$\sigma(e_{P_{\max}})$	0.82%	0.91%	5.28%	4.00%	0.51%	0.53%	0.31%	0.49%	0.49%
$\mu(e_{P_{\text{loc}}})$	−2.88°	−2.01°	−3.62°	−1.66°	−0.55°	−0.47°	−4.26°	0.02°	0.29°
$\sigma(e_{P_{\text{loc}}})$	0.85°	0.67°	6.44°	6.92°	0.94°	0.63°	1.25°	1.25°	1.67°

Table G.10: Cylinder No. 3 pressure estimation results evaluation based on Algorithm 7.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	−14.29%	8.44%	−3.16%	−14.55%	−60.97%	−49.12%	−26.94%	−27.03%	−21.91%
$\sigma(e_{P_{\max}})$	1.43%	1.15%	11.22%	1.20%	2.13%	1.47%	1.32%	2.25%	1.80%
$\mu(e_{P_{\text{loc}}})$	0.33°	−1.35°	−0.57°	5.25°	5.44°	4.97°	4.53°	5.17°	4.41°
$\sigma(e_{P_{\text{loc}}})$	0.80°	0.42°	0.76°	0.83°	0.46°	0.44°	0.79°	1.79°	1.33°

Table G.11: Cylinder No. 4 pressure estimation results evaluation based on Algorithm 7.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	−16.10%	7.91%	−4.17%	−15.36%	−65.14%	−52.66%	−29.04%	−31.27%	−26.20%
$\sigma(e_{P_{\max}})$	1.78%	1.14%	12.36%	1.40%	2.32%	1.75%	2.08%	2.80%	1.77%
$\mu(e_{P_{\text{loc}}})$	−0.37°	−1.35°	−0.44°	5.42°	5.60°	5.52°	4.45°	5.60°	4.16°
$\sigma(e_{P_{\text{loc}}})$	0.54°	0.27°	0.63°	0.73°	0.60°	1.27°	0.82°	1.90°	0.77°

Table G.12: Cylinder No. 2 pressure estimation results evaluation based on Algorithm 7.1 (50 engine cycles for each operating condition).

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
$\mu(e_{P_{\max}})$	−16.88%	7.05%	−0.49%	−12.69%	−59.28%	−46.32%	−24.90%	−25.93%	−22.47%
$\sigma(e_{P_{\max}})$	1.47%	1.32%	3.75%	1.40%	2.22%	8.41%	1.17%	2.42%	2.18%
$\mu(e_{P_{\text{loc}}})$	−0.49°	−1.32°	−1.08°	3.33°	5.44°	7.25°	4.21°	4.49°	3.38°
$\sigma(e_{P_{\text{loc}}})$	0.43°	0.20°	0.49°	3.45°	0.71°	3.41°	1.42°	0.63°	1.13°

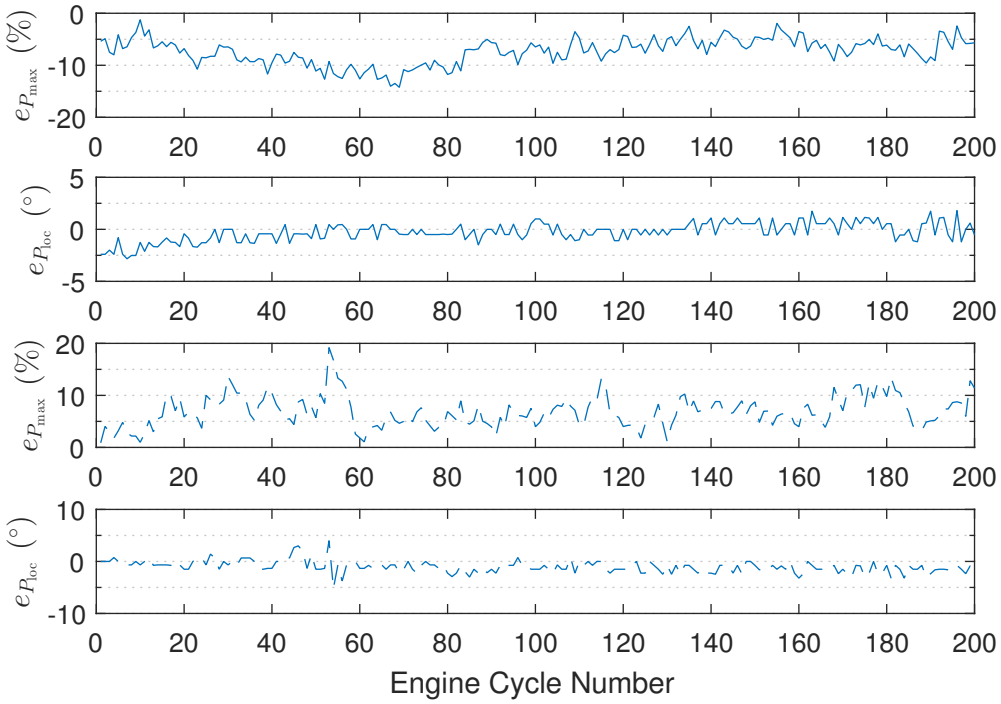

Figure G.1: Cylinder No. 3 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\text{t}}$ (lower two subfigures) using Algorithm 5.1.

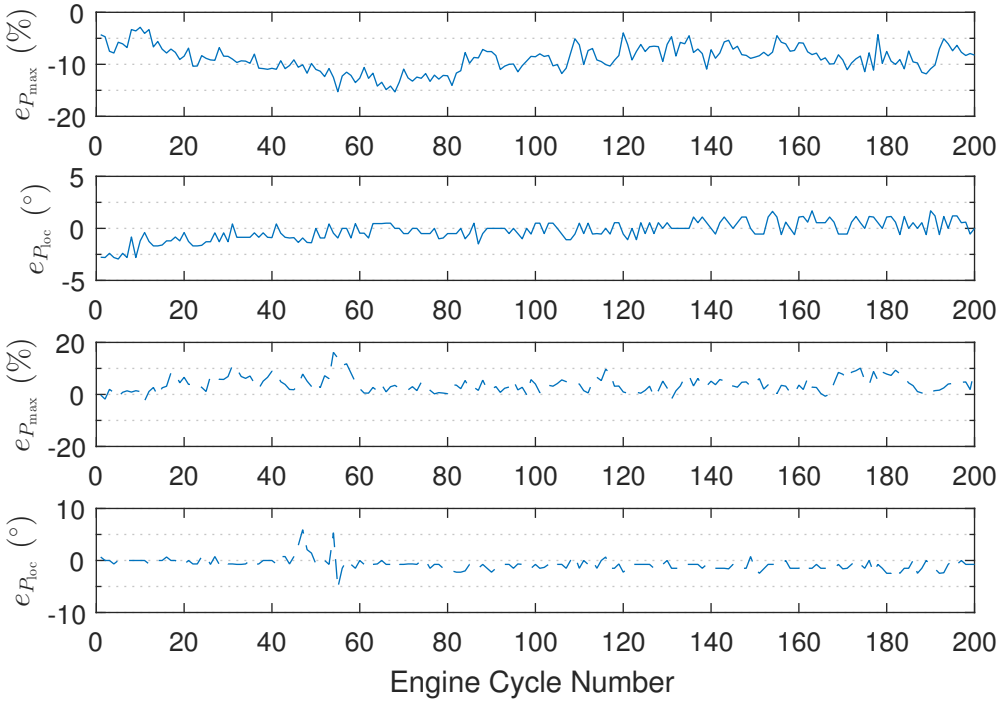

Figure G.2: Cylinder No. 4 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\text{t}}$ (lower two subfigures) using Algorithm 5.1.

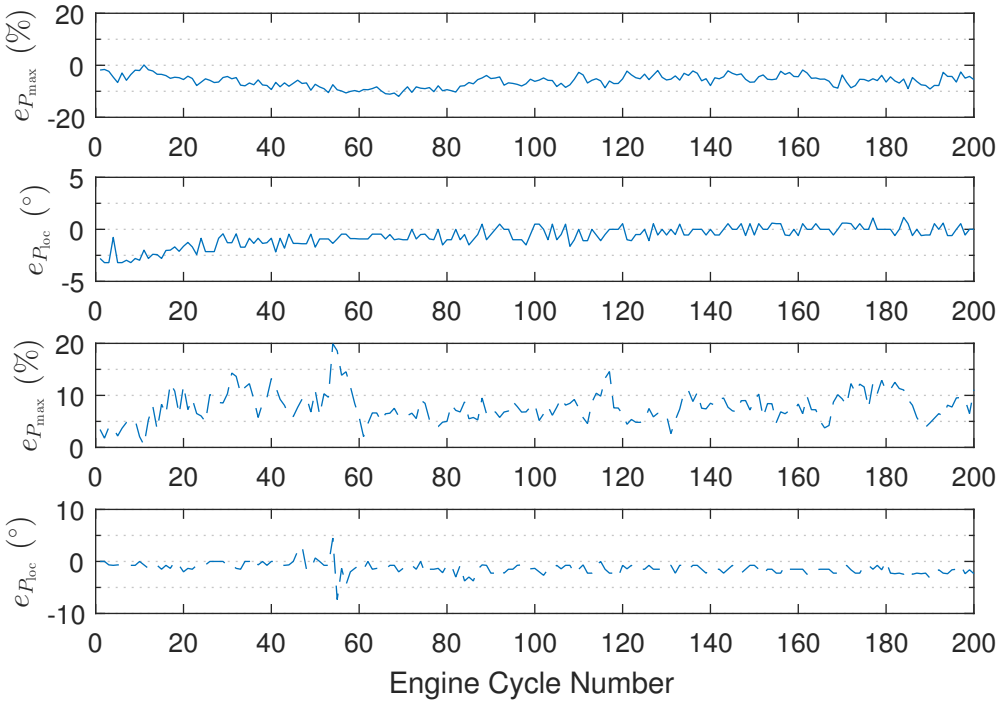

Figure G.3: Cylinder No. 2 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\text{t}}$ (lower two subfigures) using Algorithm 5.1.

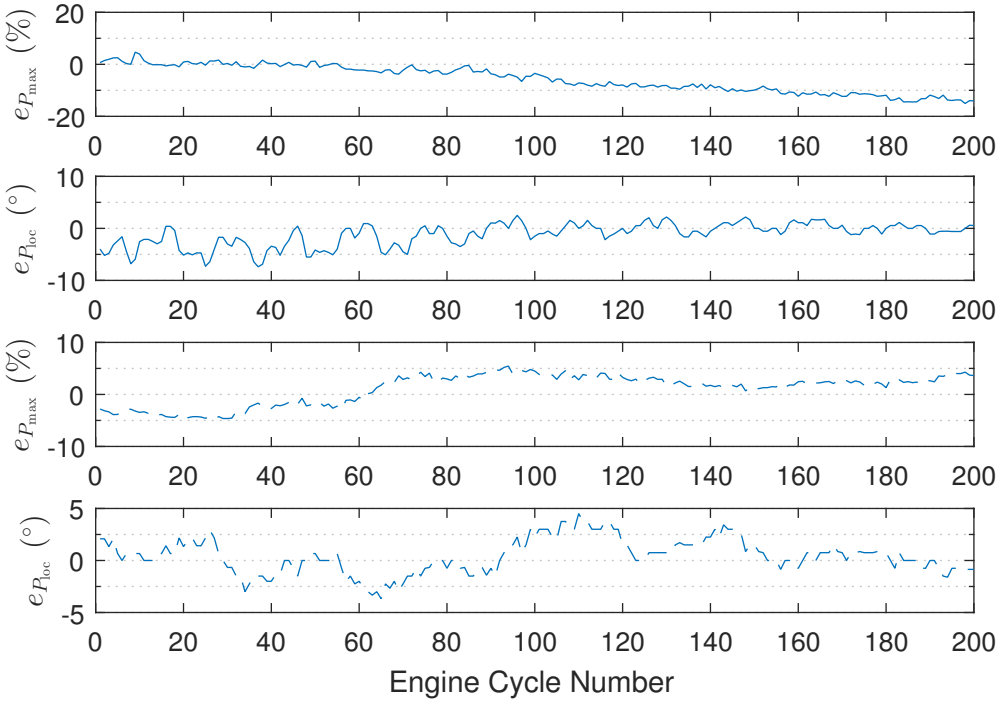

Figure G.4: Cylinder No. 3 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\text{t}}$ (lower two subfigures) using Algorithm 5.1.

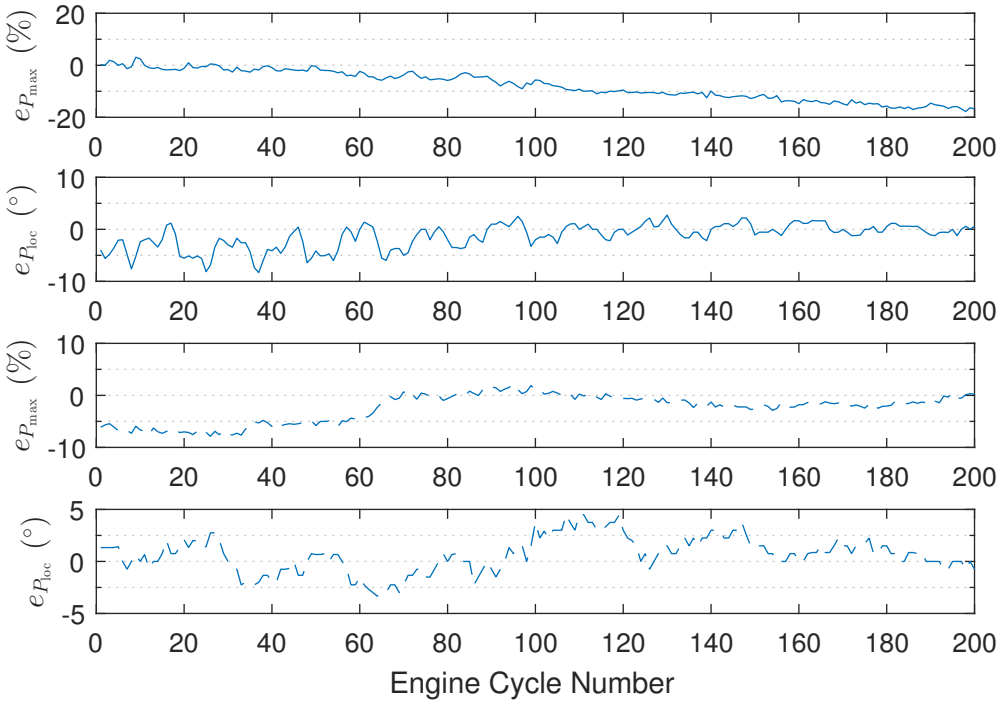

Figure G.5: Cylinder No. 4 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\text{t}}$ (lower two subfigures) using Algorithm 5.1.

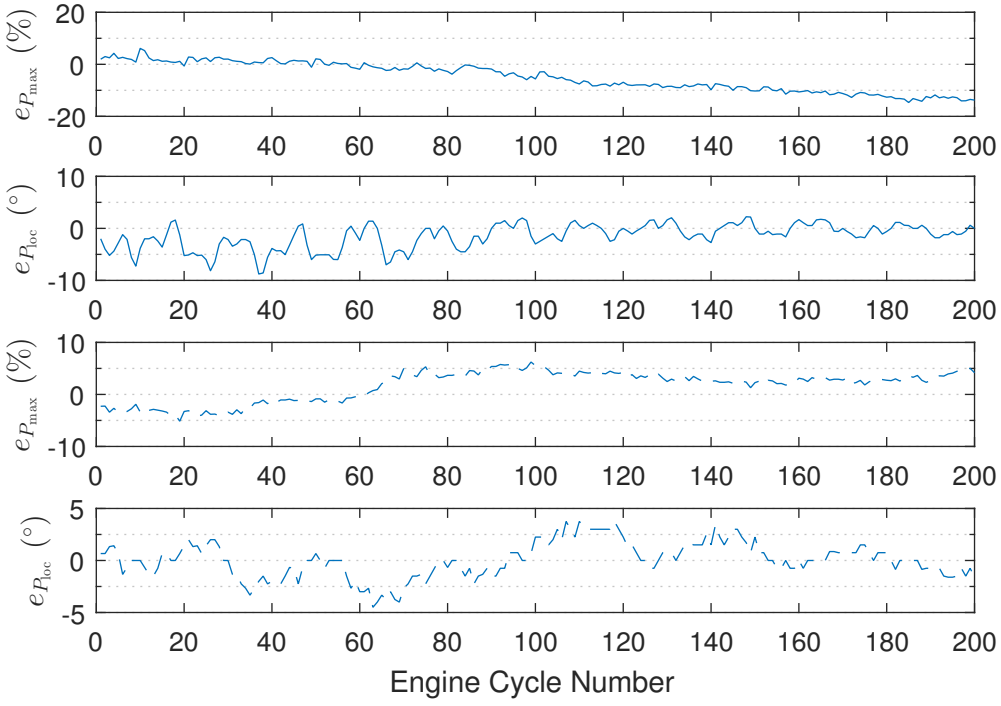

Figure G.6: Cylinder No. 2 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\text{t}}$ (lower two subfigures) using Algorithm 5.1.

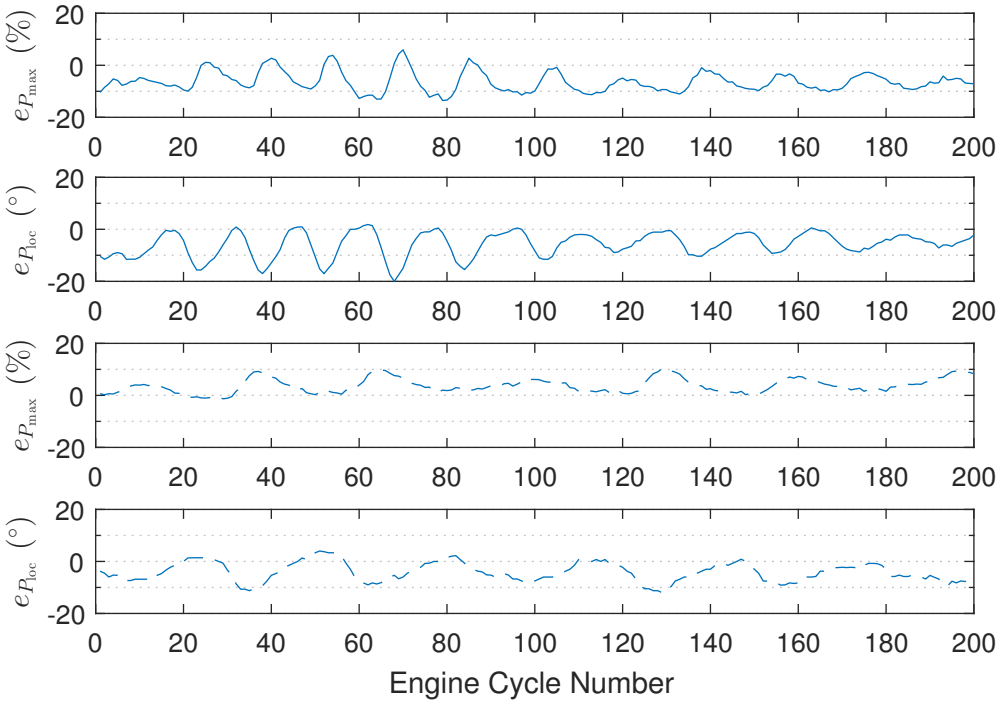

Figure G.7: Cylinder No. 3 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\text{t}}$ (lower two subfigures) using Algorithm 6.2.

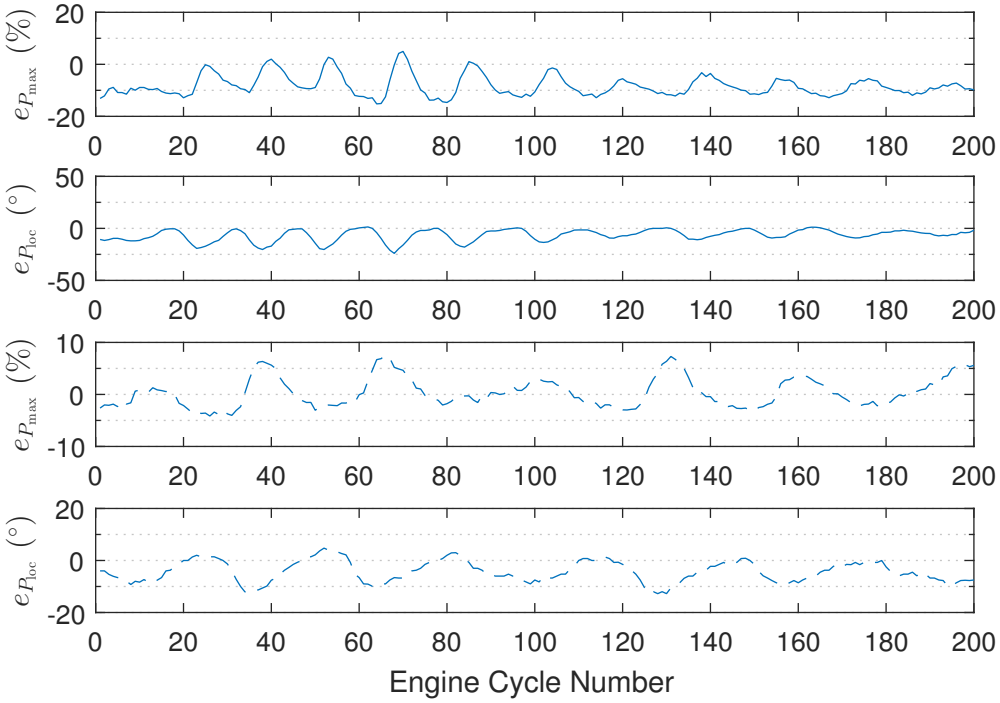

Figure G.8: Cylinder No. 4 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\text{t}}$ (lower two subfigures) using Algorithm 6.2.

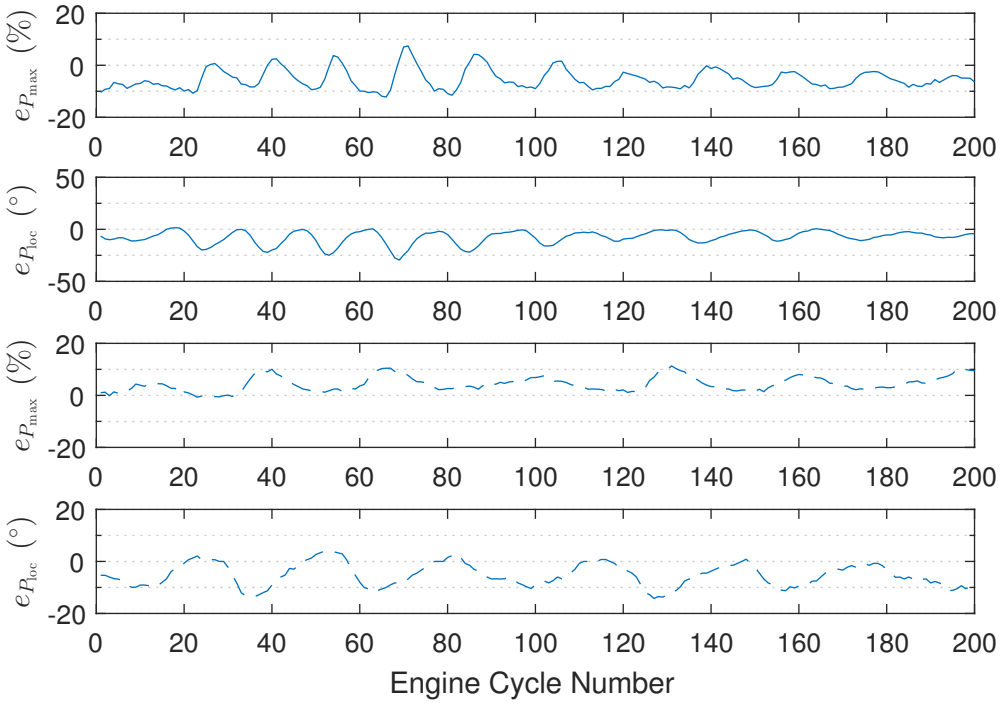

Figure G.9: Cylinder No. 2 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\mathrm{t}}$ (lower two subfigures) using Algorithm 6.2.

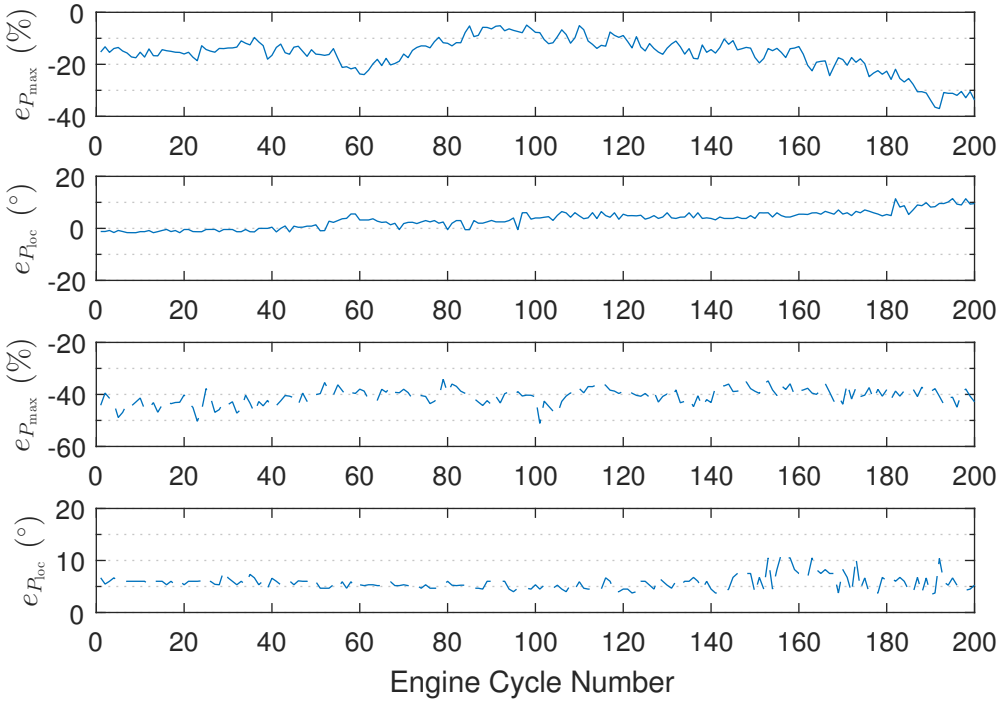

Figure G.10: Cylinder No. 3 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\mathrm{t}}$ (lower two subfigures) using Algorithm 7.1.

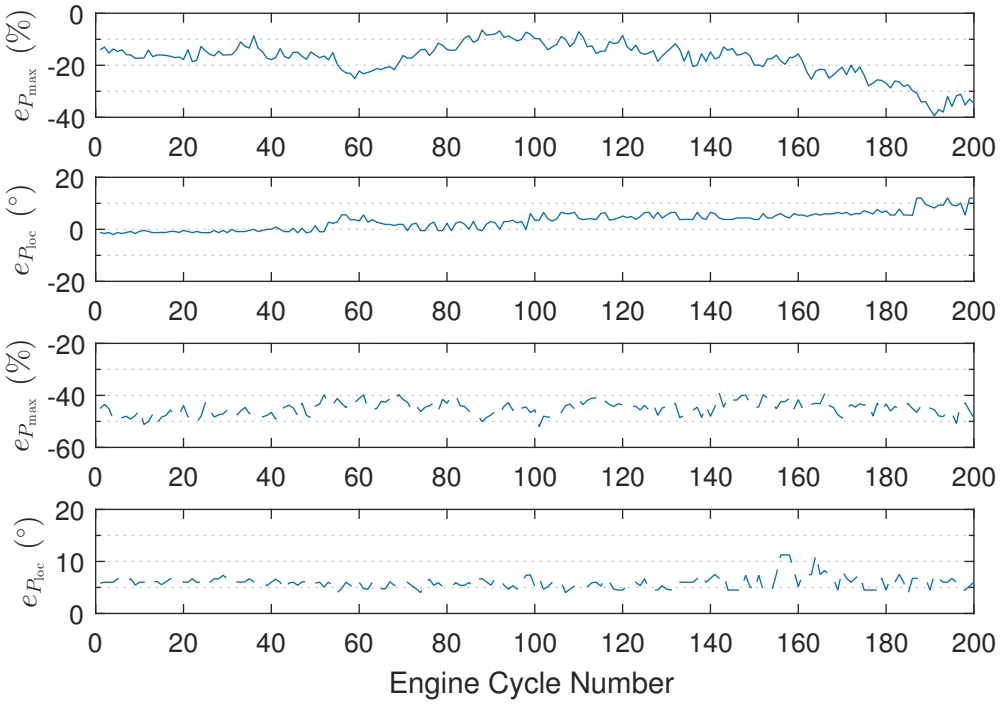

Figure G.11: Cylinder No. 4 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\mathrm{t}}$ (lower two subfigures) using Algorithm 7.1.

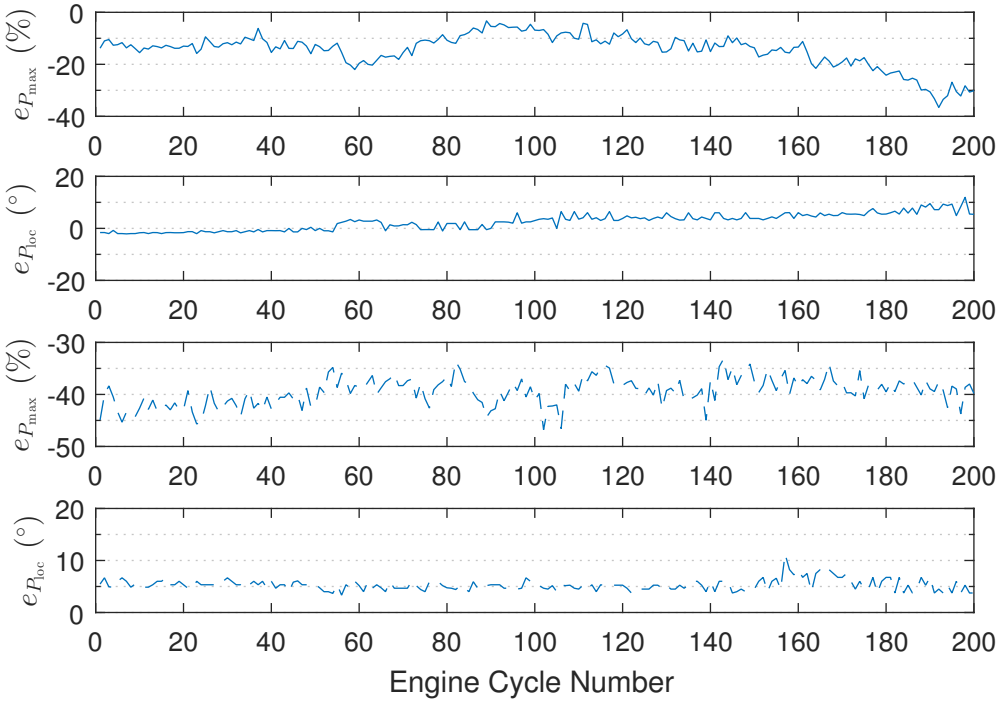

Figure G.12: Cylinder No. 2 pressure estimation results evaluation under the operating conditions Z_{t} (upper two subfigures) and $\tilde{Z}_{\mathrm{t}}$ (lower two subfigures) using Algorithm 7.1.

References

[1] I. Haskara and Y. Wang, "Cylinder pressure-based combustion controls for advanced diesel combustion with multiple-pulse fuel injection," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 6, pp. 2143–2155, 2013.

[2] A. Al-Durra, M. Canova, and S. Yurkovich, "A model-based methodology for real-time estimation of diesel engine cylinder pressure," *Journal of Dynamic Systems, Measurements, and Control*, vol. 133, no. 3, pp. 031 005–1–031 005–9, 2011.

[3] L. Fortuna, S. Graziani, A. Rizzo, and M. G. Xibilia, *Soft Sensors for Monitoring and Control of Industrial Processes*. London, UK: Springer, 2007.

[4] M. El-Ghamry, J. A. Steel, R. L. Reuben, and T. L. Fog, "Indirect measurement of cylinder pressure from diesel engines using acoustic emission," *Mechanical Systems and Signal Processing*, vol. 19, no. 4, pp. 751–765, 2005.

[5] C. J. Polonowksi, V. K. Mathur, J. D. Naber, and J. R. Blough, "Accelerometer based sensing of combustion in a high speed HPCR diesel engine," in *Proceedings of the SAE World Congress & Exhibition*, Detorit, MI, 16–19 April 2007.

[6] L. Arnone, S. Manelli, G. Chiatti, and O. Chiavola, "In-cylinder pressure analysis through accelerometer signal processing for diesel engine combustion optimization," in *Proceedings of the SAE 2009 Noise and Vibration Conference and Exhibition*, Saint Charles, CA, 19–21 May 2009.

[7] F. Taglialatela, N. Cesario, M. Porto, S. S. Merola, P. Sementa, and B. M. Vaglieco, "Use of accelerometers for spark advance control of SI engines," *SAE International Journal of Engine*, vol. 2, no. 1, pp. 971–981, 2009.

[8] P. Azzoni and A. Paniero, "Cylinder pressure reconstruction by cepstrum analysis," in *Proceedings of the 22nd International Symposium on Automotive Technology and Automation*, Florence, Italy, 14–18 May 1990, pp. 629–637.

[9] Y. Ren, *Detection of knocking Combination in Diesel Engines by Inverse Filtering of Structural Vibration Signals*. Sydney, Austrilia: University of New South Wales, 1999.

[10] Y. Gao and R. B. Randall, "Reconstruction of diesel engine cylinder pressure using a time domain smoothing technique," *Mechanical Systems and Signal Processing*, vol. 13, no. 5, pp. 709–722, 1999.

[11] H. Du, L. Zhang, and X. Shi, "Reconstructing cylinder pressure from vibration signals based on radial basis function networks," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 215, no. D6, pp. 761–767, 2001.

[12] R. Johnsson, "Cylinder pressure reconstruction based on complex radial basis function networks from vibration and speed signals," *Mechanical Systems and Signal Processing*, vol. 20, no. 8, pp. 1923–1940, 2006.

[13] D. Moro, N. Cavina, and F. Ponti, "In-cylinder pressure reconstruction based on instantaneous engine speed signal," *Journal of Engineering of Gas Turbines and Power*, vol. 124, no. 1, pp. 220–225, 2002.

[14] F. T. Connolly and A. E. Yagle, "Modeling and identification of combustion pressure process in internal combustion engines," *Mechanical Systems and Signal Processing*, vol. 8, no. 1, pp. 1–19, 1994.

[15] Y. Shiao and J. J. Moskwa, "Cylinder pressure and combustion heat release estimation for si engine diagnostics using nonlinear sliding observers," *IEEE Transations on Control Systems Technology*, vol. 3, no. 1, pp. 70–78, 1995.

[16] P. J. Jacob, F. Gu, and A. D. Ball, "Non-parametric models in the monitoring of engine performance and condition: Part 1: Modeling of non-linear engine processes," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 213, no. D1, pp. 73–81, 1999.

[17] F. Gu, P. J. Jacob, and A. D. Ball, "Non-parametric models in the monitoring of engine performance and condition: Part 2: Non-instrusive estimation of diesel engine cylinder pressure and its use in fault detection," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 213, no. D1, pp. 135–143, 1999.

[18] R. Potenza, J. F. Dunne, and D. Richardson, "Multicylinder engine pressure reconstruction using narx neural networks and crank kinematics," *International Journal of Engine Research*, vol. 8, no. 6, pp. 499–518, 2007.

[19] C. Bennett, *Reconstruction of Gasoline Engine In-Cylinder Pressures Using Recurrent Neural Networks*. Brighton, UK: University of Sussex, 2014.

[20] L. Jia, J. D. Naber, and J. R. Blough, "Review of sensing methodologies for estimation of combustion metrics," *Journal of Combustion*, vol. 2016, no. 8593523, pp. 1–9, 2016.

[21] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ: Prentice-Hall, 2000.

[22] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Sqaures Approach*. Cambridge, UK: Cambridge University Press, 2007.

[23] A. Spanos, *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*. Cambridge, UK: Cambridge University Press, 1999.

[24] F. L. Lewis, L. Xie, and D. Popa, *Optimal and Robust Estimation: With an Introduction to Stochastic Control Theory.* Boca Raton, FL: CRC Press, 2008.

[25] B. D. O. Anderson and J. B. Moore, "Detectability and stabilizability of time-varying discrete-time linear systems," *SIAM Journal on Control and Optimization*, vol. 19, no. 1, pp. 20–32, 1981.

[26] T. Karvonen, *Stability of Linear and Non-Linear Kalman Filters.* Helsinki, Finland: University of Helsinki, 2014.

[27] W. J. Rugh, *Linear System Theory.* Upper Saddle River, NJ: Prentice Hall, 1996.

[28] D. Simon, *Optimal State Estimation: Kalman, H infinity, and Nonlinear Approaches.* Hoboken, NJ: John Wiley & Sons, 2006.

[29] P. Kaminski, A. Bryson, and S. Schmidt, "Discrete square root filtering: A survey of current techniques," *IEEE Transations on Automatic Control*, vol. 16, no. 6, pp. 727–736, 1971.

[30] B. D. O. Anderson and J. B. Moore, *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall, 1979.

[31] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB.* New York, NY: John Wiley & Sons, 2001.

[32] J. R. Blough, "A survey of DSP methods for rotating machinery analysis, what is needed, what is available," *Journal of Sound and Vibration*, vol. 262, no. 3, pp. 707–720, 2003.

[33] S. Bittanti and S. M. Savaresi, "On the parameterization and design of an extended kalman filter frequency tracker," *IEEE Transactons on Automatic Control*, vol. 45, no. 9, pp. 1718–1724, 2000.

[34] C. Bohn, O. Magnor, and M. Schultalbers, "State observer based analysis of crankshaft speed measurements with application to misfire detection," in *Proceedings of the 2005 International Conference on Control and Automation*, Budapest, Hungary, 26–29 June 2005, pp. 239–244.

[35] B. L. Scala, R. Bitmead, and M. R. James, "Conditions for stability of the extended kalman filter and their application to the frequency racking problem," *Mathematics of Control, Signals and Systems*, vol. 8, pp. 1–26, 1995.

[36] P. van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory-Implementation-Applications.* Boston, MA: Kluwer Academic Publishers, 1996.

[37] L. Ljung, *System Identification: Theory for the User.* Upper Saddle River, NJ: Prentice Hall PTR, 1999.

[38] T. Katayama, *Subspace Methods for System Identification.* London, UK: Springer, 2005.

[39] S. J. Qin, "An overview of subspace identification," *Computers and Chemical Engineering*, vol. 30, no. 10–12, pp. 1502–1513, 2006.

[40] A. K. Tangirala, *Principles of System Identification: Theory and Practice*. Boca Raton, FL: CRC Press, 2015.

[41] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton, NJ: Princeton University Press, 2009.

[42] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins University Press, 1996.

[43] L. Balogh, *Model Fitting in Frequency Domain Imposing Stability of the Model*. Brussels, Belgium: Uitgeverij VUBPRESS Brussels University Press, 2009.

[44] H. S. Ulusoy, M. Q. Feng, and P. J. Fanning, "System identification of a building from multiple seismic records," *Earthquake Engineering and Structural Dynamics*, vol. 40, no. 6, pp. 661–674, 2011.

[45] P. van Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.

[46] L. Ljung, "Aspects and experiences of user choices in subspace identification methods," *IFAC-PapersOnLine*, vol. 36, no. 16, pp. 1765–1770, 2003.

[47] M. Verhaegen, "Identification of the deterministic part of MIMO state space models given in innovations form from input-output data," *Automatica*, vol. 30, no. 1, pp. 61–74, 1994.

[48] W. E. Larimore, "Canonical variate analysis in identification, filtering and adaptive control," in *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, HI, 5–7 December 1990, pp. 596–604.

[49] L. Ljung, *System Identification Toolbox™ User's Guide*. Natick, MA: MathWorks, 2016.

[50] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*. New York, NY: IEEE Press, 2001.

[51] K. M. M. Prabhu, *Window Functions and Their Applications in Signal Processing*. Boca Raton, FL: CRC Press, 2014.

[52] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*. Upper Saddle River, NJ: Prentice Hall, 1997.

[53] C. Bohn and H. Unbehauen, "Minmax and least squares multivariable transfer function curving fitting: Error criteria, algorithms and comparisons," in *Proceedings of the American Control Conference*, Philadelphia, PA, 24–26 June 1998, pp. 3189–3193.

[54] C.-W. Chen, J.-N. Juang, and G. Lee, "Frequency domain state-space system identification," *Journal of Vibration and Acoustics*, vol. 116, no. 4, pp. 523–528, 1994.

[55] R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbé, "Estimation of nonparametric noise and FRF models for multivariable systems: Part I: Theory," *Mechanical Systems and Signal Processing*, vol. 24, no. 3, pp. 573–595, 2010.

[56] ——, "Estimation of nonparametric noise and FRF models for multivariable systems: Part II: Extensions, applications," *Mechanical Systems and Signal Processing*, vol. 24, no. 3, pp. 596–616, 2010.

[57] T. McKelvey, H. Akçay, and L. Ljung, "Subspace-based multivariable system identification from frequency response data," *IEEE Transactons on Automatic Control*, vol. 41, no. 7, pp. 960–979, 1996.

[58] J. Paduart, *Identification of Nonlinear Systems using Polynomial Nonlinear State Space Models*. Brussels, Belgium: Uitgeverij VUBPRESS Brussels University Press, 2008.

[59] A. I. Khuri, *Advanced Calculus with Applications in Statisticsä*. Hoboken, NJ: John Wiley & Sons, 2003.

[60] L. A. Zadeh and C. A. Desoer, *Linear System Theory: The State Space Approach*. New York, NY: McGraw-Hill Book, 1963.

[61] R. Remmert and G. Schumacher, *Funktionentheorie I*. Berlin, Germany: Springer, 1989.

[62] A. Janczak, *Identification of Nonlinear Systems Using Neural Networks and Polynomial Models*. Berlin, Germany: Springer, 2005.

[63] F. Giri and E.-W. Bai, Eds., *Block-Oriented Nonlinear System Identification*. Berlin, Germany: Springer, 2010.

[64] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. London, UK: Academic Press, 1981.

[65] S. Boyed and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[66] J. O'Reilly, *Observers for Linear Systems*. London, UK: Academic Press, 1983.

[67] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York, NY: Academic Press, 1970.

[68] A. Gelb, Ed., *Applied Optimal Estimation*. Cambridge, MA: The MIT Press, 1974.

[69] J. Humpherys and J. West, "Kalman filtering with newton's method," *IEEE Control Systems Magazine*, vol. 30, no. 6, pp. 523–528, 2010.

[70] T. Söderström and P. Stoica, *System Identification*. Hemel Hempstead, UK: Prentice Hall International, 1989.

[71] J.-N. Juang, *Applied System Identification*. Upper Saddle River, NJ: Prentice Hall PTR, 1994.

[72] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: The MIT Press, 1983.

[73] C. Bohn and H. Unbehauen, *Identifikation dynamischer Systeme*. Wiesbaden, Germany: Springer Vieweg, 2016.

[74] R. Isermann and M. Münchhof, *Identification of Dynamic Systems: An Introduction with Applications*. Berlin, Germany: Springer, 2011.

[75] T. Söderström, *Errors-in-Variables Methods in System Identification*. Cham, Switzerland: Springer International Publishing, 2018.

[76] S. A. Billings, *Nonlinear System Identification: NARMAX Methods in Time, Frequency, and Spatio-Temporal Domains*. West Sussex, UK: John Wiley & Sons, 2013.

[77] J. O. Smith III, *Introduction to Digital Filters with Audio Applications*. http://w3k.org/books/filters1p3/index.html: W3K Publishing, 2007.

[78] K. Lee, M. Yoon, and M. Sunwoo, "A study on pegging methods for noisy cylinder pressure signal," *Control Engineering Practice*, vol. 16, no. 8, pp. 922–929, 2008.

[79] H. André, F. Girardin, A. Bourdon, J. Antoni, and R. Rémond, "Precision of the ias monitoring system based on the elapsed time method in the spectral domain," *Mechanical Systems and Signal Processing*, vol. 44, no. 1–2, pp. 14–30, 2014.

[80] R. Han, C. Bohn, and G. Bauer, "Recursive engine in-cylinder pressure estimation using Kalman filter and structural vibration signal," *IFAC-PapersOnLine*, vol. 51, no. 31, pp. 700–705, 2018.

[81] ——, "Virtual engine in-cylinder pressure sensor for automobiles and agricultural tractors," *IFAC-PapersOnLine*, vol. 53, no. 1, pp. 543–548, 2020.

[82] G. A. Korn, "Network approximations of variable time delays require caution," *IRE Transactions on Electronic Computers*, vol. EC-11, no. 1, p. 82, 1962.

[83] M. Vajta, "Some remarks on padé-approximations," in *Proceedings of the 3rd TEMPUS-INTCOM Symposium on Intelligent Systems in Control and Measurements*, Veszprém, Hungary, 9–14 September 2000, pp. 1–6.

[84] R. Han, C. Bohn, and G. Bauer, "A novel inverse filtering method for systems with multiple input signals," in *Proceedings of the 18th International Conference on Control, Automation and Systems*, PyeongChang, South Korea, 17–20 October 2018, pp. 369–374.

[85] ——, "Recursive engine in-cylinder pressure estimation merely using structural vibration signal," in *Proceedings of the 2020 European Control Conference*, Saint Petersburg, Russia, 12–15 May 2020, pp. 240–245.

[86] T. Uhl, "The inverse identification problem and its technical application," *Archive of Applied Mechanics*, vol. 77, no. 5, pp. 325–337, 2007.

[87] Y. Jung and M. Enqvist, "Estimating models of inverse systems," in *Proceedings of the 52nd IEEE Conference on Decision and Control*, Florence, Italy, 10–13 December 2013, pp. 7143–7148.

[88] Y. Li and T. Tomizuka, "Two-degree-of-freedom control with robust feedback control for hard disk servo systems," *IEEE/ASME Transactions on Mechatronics*, vol. 4, no. 1, pp. 17–24, 1999.

[89] F. Boeren, T. Oomen, and M. Steinbuch, "Iterative motion feedforward tuning: A data-driven approach based on instrumental variable identification," *Control Engieering Practice*, vol. 37, pp. 11–19, 2015.

[90] L. M. Silverman, "Inversion of multivatiable linear systems," *IEEE Transactions on Automatic Control*, vol. 14, no. 3, pp. 270–276, 1969.

[91] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Englewood Cliffs, NJ: Prentice Hall, 1996.

[92] C. Batlle and N. Roqueiro, "Balanced model order reduction method for systems depending on a parameter," *IFAC-PapersOnLine*, vol. 52, no. 1, pp. 412–417, 2019.

[93] H.-F. Chen, "On stochastic observability and controllability," *Automatica*, vol. 16, pp. 179–190, 1980.