June 2019

# System Identification Based on Errors-In-Variables System Models

Hyundeok Kang
*Louisiana State University and Agricultural and Mechanical College*

# SYSTEM IDENTIFICATION BASED ON ERRORS-IN-VARIABLES SYSTEM MODELS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Division of Electrical and Computer Engineering

by
Hyundeok Kang
B.S., Inha University, 2007
M.S., Inha University, 2010
M.S., Louisiana State University, 2016
August 2019

# ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Dr. Guoxiang Gu for giving me the opportunity to do my Ph.D dissertation. Because of his continuous support, his kind, helpful advice and the interesting discussions, I was able to stay focused and not stray away from the essence of my work.

I also would like to thank my committee members, Dr. Morteza Naraghi-Pour, Dr. Xiangwei Zhou, Dr. Xiangyu Meng, Dr. Yuri Antipov and Dr. Kemin Zhou for taking time out of their busy schedules and consenting to be my committee, and for their valuable feedback.

In addition, I would like to thank the faculty, staff, and students in the Division of Electrical and Computer Engineering for all the help and support I have received.

Finally, I would like to thank my family and friends who put their faith in me and urged me to do better.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . . ii

ABSTRACT . . . . . iv

CHAPTER

1 INTRODUCTION . . . . . 1
- 1.1 Motivation . . . . . 1
- 1.2 Concept of EIV System Identification . . . . . 3
- 1.3 Existing Work . . . . . 4
- 1.4 Dissertation Work . . . . . 8

2 MATHEMATICAL PREPARATION . . . . . 12
- 2.1 Signals and Systems . . . . . 12
- 2.2 System Identification Problem . . . . . 15
- 2.3 EIV Estimation . . . . . 21

3 GRAPH SUBSPACE APPROACH . . . . . 27
- 3.1 Preliminaries . . . . . 27
- 3.2 Main Results . . . . . 31
- 3.3 Implementation . . . . . 42
- 3.4 Simulation Studies . . . . . 48

4 IDENTIFICATION WITH UNEQUAL NOISE VARIANCES . . . . . 57
- 4.1 Frisch Scheme . . . . . 57
- 4.2 GSA Combined with the Frisch Scheme . . . . . 67
- 4.3 New EIV Identification . . . . . 73

5 CONCLUSION AND FUTURE WORK . . . . . 92
- 5.1 Conclusion . . . . . 92
- 5.2 Future Work . . . . . 93

REFERENCES . . . . . 96

VITA . . . . . 103

# ABSTRACT

We study the identification problem for errors-in-variables (EIV) systems. Such an EIV model assumes that the measurement data at both input and output of the system involve corrupting noises. The least square (LS) algorithm has been widely used in this area. However, it results in biased estimates for the EIV-based system identification. In contrast, the total least squares (TLS) algorithm is unbiased, which is now well-known, and has been effective for estimating the system parameters in the EIV system identification.

In this dissertation, we first show that the TLS algorithm computes the approximate maximum likelihood estimate (MLE) of the system parameters and that the approximation error converges to zero asymptotically as the number of measurement data approaches infinity. Then we propose a graph subspace approach (GSA) to tackle the same EIV-based system identification problem and derive a new estimation algorithm that is more general than the TLS algorithm. Several numerical examples are worked out to illustrate our proposed estimation algorithm for the EIV-based system identification.

We also study the problem of the EIV system identification without assuming equal noise variances at the system input and output. Firstly, we review the Frisch scheme, which is a well-known method for estimating the noise variances. Then we propose a new method which is GSA in combination with the Frisch scheme (GSA-Frisch) algorithm via estimating the ratio of the noise variances and the system parameters iteratively. Finally, a new identification algorithm is proposed to estimate the system parameters based on the subspace interpretation without estimating noise variances or the ratio. This new algorithm is unbiased, and achieves the consistency of the parameter estimates. Moreover, it is low in complexity. The performance of the identification algorithm is examined by several numerical examples, and compared to the N4SID algorithm that has the Matlab codes available in Matlab toolboxes, and also to the GSA-Frisch algorithm.

# CHAPTER 1
# INTRODUCTION

This dissertation is on system identification, focusing on errors-in-variables (EIV) models. Such a model is extensively used in [7, 8, 57, 31, 77, 63, 39, 90, 91]. However, identification of such a model is full of challenges. We will propose a subspace based identification algorithm, and show its efficacy in estimation of the system parameters. Moreover, we will investigate various issues encountered in identifying the EIV systems. In this chapter, we will begin with the motivation of this dissertation work, followed by introducing the problem with the concept of EIV system identification. The existing work in this problem area will be reviewed and contributions of this dissertation will be presented, followed by organization and notations of the dissertation.

## 1.1 Motivation

System identification has been a core problem in model-based feedback control design. It is a methodology which uses statistical methods to build mathematical models of dynamical systems from measured data [65]. Its importance lies, especially for the control engineers, in the fact that it helps to not only figure out the unknown system but also predict the future behavior of the system so that we can design an appropriate controller for the system. The well-known books such as [41, 65] are standard texts providing treatise for identification of linear dynamic systems.

The initial work in the area of system identification has been focused on the case when output measurements are corrupted by noises, but the input signals are noise-free. The least-squares (LS) algorithm and its variants are well-known, and have been especially effective in identifying the linear system models. However, it becomes more interesting and harder when both the input and output measurements are corrupted by noises, which is referred to as errors-in-variables (EIV) system identification. It also has received great attention from this research community [7, 31, 37, 57, 64]. One of the traditional system identification methods

is LS estimation. However, the ordinary LS estimates are biased in general when employed for EIV system identification. Our motivation lies in tackling the bias problem of the LS estimates.

We are also motivated by the applications of EIV models in array signal processing for direction of arrival (DOA) estimation, time series modeling, blind channel estimation and equalization, astronomical data reduction, etc, in which the principle component analysis is the key ingredient embodied by the subspace approach [77, 78]. Such an approach may lead to the ML estimator under certain conditions. In fact even if the ML conditions do not hold, the subspace approach often leads to good estimation performance evidenced by the superresolution algorithm in DOA estimation [71] and by the popular second-order statistical method in blind channel estimation and equalization [75]. However, there lacks similar developments in the area of system identification due to probably the input and output relation of the measurement data that obscure the nature of the underlying estimation problem.

Another motivation is from the subspace based algorithms, such as the N4SID algorithm [14, 32, 80, 82, 81], because of its root in the principle component analysis (PCA) [33, 51] and canonical correlation analysis (CCA) [38, 36, 17], both being popular in statistical analysis and in various engineering applications. Indeed for finite order linear systems, if the noises in measuring the system input and output are Gauss white and have equal variances, then the EIV model will be resulted in, and the total least-squares (TLS) algorithm is applicable, leading to asymptotically unbiased estimates of the system parameters. Such a TLS algorithm admits a subspace interpretation [28] (Chapter 8) and works if noise variances are unequal but with the known ratio of the two variances. In fact, the maximum likelihood algorithm can be developed in [63, 69]. However, in the absence of knowledge on the noise variances, the system identification problem becomes a lot harder. Not only the maximum likelihood, but also the consistent algorithms are difficult to develop. This is evidenced by the BELS algorithm [31, 97, 93] and by the Frisch scheme [7, 57]. Both are viable methods, but fail to work if the signal-noise-ratio (SNR) is low.

## 1.2 Concept of EIV System Identification

The basic idea of system identification is the estimation of unknown system parameters by modeling the system based on the observed data. A simple diagram of a system with the input and output is illustrated in Figure 1.1. The system is assumed to be a finite dimensional linear time-invariant (FDLTI).

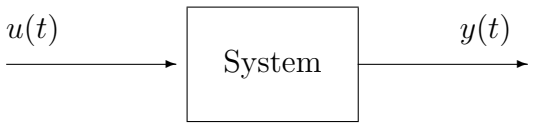


Figure 1.1. A system with input and output

If we know the collection of the input $\{u(t)\}$ and output $\{y(t)\}$ signals as an observed or measured data set in the noise free environment, we can easily estimate the system parameters by the well-known LS method. However, it is not easy to obtain the input and output data set, which has no corrupted noise. In most cases, some noises are involved in the measurement data. A diagram of the EIV signal and system model is depicted in Figure 1.2. Transfer function $P(z)$ represents the system model, $\tilde{u}(t)$ is the input measurement, $\tilde{y}(t)$

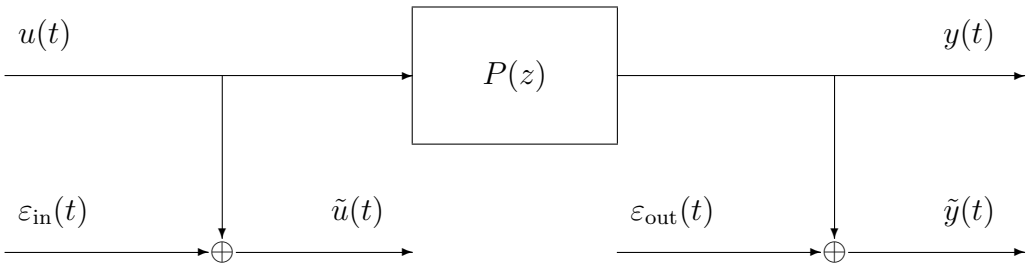


Figure 1.2. EIV system model

is the output measurement of the system, and $\varepsilon_{\text{in}}(t)$ and $\varepsilon_{\text{out}}(t)$ are the corrupting noises of the input and output in Figure 1.2. The EIV system identification is aimed at estimation of the parameters of $P(z)$ based on the given measurement data set $\{\tilde{u}(t)\}$ and $\{\tilde{y}(t)\}$.

## 1.3 Existing Work

In terms of the "errors-in-variables" problem as a research topic, not only the control engineers but also people from various fields have been interested in it. Söderström says "The area is therefore quite extensive and is also quickly growing with time. The vast majority of papers are written from an application perspective and can deal with biomedicine, chemistry, chemical engineering, earth sciences, econometrics, managements, mechanical engineering, finance, ecology, geosciences, image systems, time series analysis, etc." in his recent book [60]. It should be mentioned that the standard EIV estimation problem has long been studied in math literature, such as [26, 85, 23, 86, 48, 34, 10, 47, 35, 13, 20], due to the fact that it uses a lot of mathematical tools.

There have been many research papers that contributed to the EIV system identification problems historically. As an early work, a dynamic EIV model was developed in [40, 6], followed by many extended new methods and different solutions. The survey papers such as [58, 64, 59] give perspectives on the EIV system identification of the dynamic systems providing some popular methods and examples. We introduce some well-known methods for the EIV system identification next.

### Least Square Estimation

One of the traditional methods in system identification is the least square (LS) estimation. The pioneering work was done by the famous mathematician Carl Friedrich Gauss, followed by [2, 3]. LS estimation is aimed at finding the argument that minimizes the output equation error, which is the difference between the predicted output and the given measurement output. Let us consider the simple equation given by

$$x = Ab \tag{1.1}$$

where $x$ is a known column vector, $A$ is a known matrix and $b$ is an unknown column vector. Suppose that $x$ is known only approximately, i.e., we have only $\tilde{x} \neq x$. The LS solution to

estimate $b$ can be obtained by

$$\widehat{b}_{\mathrm{LS}} = (A'A)^{-1}A'\tilde{x}. \tag{1.2}$$

Similarly, in a dynamic system, the output equation can be written as

$$\tilde{y}(t) = \phi(t)'\theta + v_0(t), \tag{1.3}$$

where $\phi(t)$ is a regressor, which is a column vector consisting of the input and output measurements; $\theta$ is an estimated system parameter vector, which will be specifically defined in Chapter 2, and $v_0(t)$ is a noise term. By packing $\tilde{y}(t)$ together, $\phi(t)$ together, and $v_0(t)$ together from $t = 1$ to $t = N$, the LS solution can be computed by

$$\widehat{\theta}_{\mathrm{LS}} = \left[\frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi(t)'\right]^{-1}\left[\frac{1}{N}\sum_{t=1}^{N}\phi(t)\tilde{y}(t)\right]. \tag{1.4}$$

It is easily understood and widely used in this field. However, the drawbacks of the LS algorithm and its variants are quickly recognized, including the notable biased estimates in the presence of disturbances and noises at the input. In other words, the estimates do not converge to the true values as the number of sample measurement data goes to infinity. This issue will be treated in Chapter 2 with a detailed analysis. So, many new algorithms are proposed and developed to mitigate the weaknesses of the LS estimation.

**Instrumental Variable**

The instrumental variable (IV) has been studied by many researchers to overcome the limitations of the LS estimation focusing on the bias issue to achieve estimation consistency. The early work for this method was developed in [52, 88, 43, 87, 89, 21]. The modified algorithms are described in [56, 79, 70, 73]. IV methods can be seen as generalizations of the LS estimates [60]. To be specific, this method sets a correlation vector, of which the elements are called instrument variables, in a formula similar to the LS algorithm. The IV

estimation is computed by

$$\widehat{\theta}_{\mathrm{IV}} = \left[\frac{1}{N}\sum_{t=1}^{N} z(t)\phi(t)'\right]^{-1}\left[\frac{1}{N}\sum_{t=1}^{N} z(t)\tilde{y}(t)\right] \tag{1.5}$$

where $z(t)$ is a correlation vector consisting of the instrument variables chosen by the user under the specific conditions that the estimates tend to the true values for the large number of sample data. These conditions are

- $z(t)$ is strongly correlated with the regressor $\phi(t)$ so that the first term on the right-hand side in (1.5) is nonsingular,
- $z(t)$ is uncorrelated with $v_0(t)$.

The reasons those conditions are needed can be easily seen by substituting (1.3) with (1.5)

$$\begin{aligned}\widehat{\theta}_{\mathrm{IV}} &= \left[\frac{1}{N}\sum_{t=1}^{N} z(t)\phi(t)'\right]^{-1}\left[\frac{1}{N}\sum_{t=1}^{N} z(t)\{\phi(t)'\theta + v_0(t)\}\right] \\ &= \theta + \left[\frac{1}{N}\sum_{t=1}^{N} z(t)\phi(t)'\right]^{-1}\left[\frac{1}{N}\sum_{t=1}^{N} z(t)v_0(t)\right].\end{aligned} \tag{1.6}$$

If the conditions are true, the right-hand side in (1.6) is zero except $\theta$. Hence,

$$\widehat{\theta}_{\mathrm{IV}} = \theta \quad \text{as} \;\; N \longrightarrow \infty$$

unlike the LS estimation in general. In other words, IV estimation is a consistent estimator. Some examples were introduced and discussed regarding the choice of $z(t)$ and how to satisfy the aforementioned conditions in [66, 65]. However, it turns out the standard deviation of the estimates are huge, and it gives very low accuracy of the parameter estimates as disadvantages [60].

**Bias Compensation**

In order to avoid the bias issue, the bias compensation method also has been developed

by many researchers. There are two main estimation algorithms in this category, which are the bias-eliminating least square (BELS) and the Frisch scheme, focusing on the estimation of the noise variances.

- BELS: This algorithm was first developed in [97]. The basic concept of the algorithm is to remove the bias from the LS estimation, using the property that the bias is closely related to the noise variances. Basically, BELS first estimates the unknown noise variances and subtracts the noise variances from the approximate covariance matrix. Then, it estimates the system parameters using the LS algorithm iteratively with the updated estimates of the noise variances. Several approaches, especially regarding how to estimate the noise variances accurately or efficiently, were proposed in [92, 95, 93, 95, 72, 94, 96]. The disadvantage of BELS is the estimation of the noise variances is not accurate when the measurement noise level is high, which affects the accuracy of the system parameters [96].

- Frisch scheme: The early work of this algorithm was by the economist Ragnar Frisch for the static problem in [22]. After his work, many studies have appeared to develop the algorithm for the dynamic problem such as in [7, 11, 29, 70, 64, 8, 19, 61, 55]. It was also developed in the frequency domain mainly by Söderström in [62, 68, 69]. Similar to BELS, the Frish scheme is focused on estimates of the noise variances. The difference between them is that they use different mathematical tools for estimating the noise variances. Loosely speaking, BELS computes the estimates of the input and output noise variances together in each iteration step. In contrast, the Frisch scheme estimates the input noise variance first and then estimates the output noise variance using the estimated input noise variance iteratively.

**Maximum Likelihood**

The maximum likelihood (ML) estimation is based on the probability theory. It considers the probability density function of the observed data as a function of the unknown parameters

[41, 65, 60, 67, 12, 18]. Then it seeks the parameters that maximizes the probability function. It is widely used in the system identification because of it is a consistent estimator. For example, it has been developed for state space models such as in [53, 44] and also for the identification in the frequency domain such as [63, 69]. However, the disadvantage lies in the computational complexity that grows exponentially with the system order.

**Total Least Squares**

The total least squares (TLS) method has been studied in a bunch of papers, such as [27, 76, 49, 50, 16, 15], and the proceedings, such as [77, 78]. There is also a survey paper [42]. The TLS method is strongly related to the EIV model in light of the fact that TLS considers the case when both the input and the output measurement data are perturbed by the noises unlike the LS estimation. Thus, the TLS solution effectively works for the EIV system identification. See [26] for lucid convergence analysis and derivation of the TLS solution as a ML estimator. The detailed description will be provided in the next chapter.

## 1.4 Dissertation Work

### 1.4.1 Contribution

An attempt is made in this study to apply the results in [26] to identifying EIV systems. Specifically, we introduce graph metric [25, 84] that is induced by the input and output graph of the system. By suitable parameterization of the plant model, the system graph is identified as a closed subspace, termed as graph subspace, of the second-order statistics of the sample measurements which inherit the EIV structure albeit different from the standard EIV model in the literature.

Different from the robust control, this study is aimed at estimating the system parameters based on the finite measurement samples of the graph subspace of a given system. By blocking the input and output measurement data into vectors of different lengths, finite samples of the system graph subspace can be obtained, and their second order statistics

can be calculated. If the input data are persistently exciting, then the signal and noise subspaces of the sampled graph space can be distinguished asymptotically via the eigenvalue decomposition, leading to a consistent estimator. For finite measurements of the input and output data, signal and noise subspaces can be approximately computed. The computation procedure is similar to that of the TLS, and will be referred to as graph subspace algorithm (GSA). In a special case of the minimum blocking size, our graph subspace algorithm reduces to the TLS algorithm, and is thus more general. The contribution includes the proof of the asymptotic MLE nature of the TLS algorithm, and the proposal of the graph subspace algorithm that often outperforms the TLS, especially when the SNR is high. In addition, an iterative procedure is developed in order to improve the estimation accuracy in the case of a low SNR.

Also, in Chapter 4, we propose two methods for the EIV system identification, which are aimed to estimate system parameters with consistency in the presence of the input and output measurement noises when the noise variances are not only unknown but also not equal. One of them is GSA combined with the Frish scheme. To be specific, we estimate the ratio of the noise variances using the Frisch scheme, which is one of the well-known methods to estimate noise variances. The estimated ratio is employed in the GSA algorithm that is developed in Chapter 3, for the system parameter estimation. The contribution of this method is that it can be employed in practical cases in general, considering the noise variances are not equal and unknown.

The other method we propose is via defining the past and future inputs and outputs following similar ideas found in [14, 32, 80, 82, 81]. In this method, we estimate the system parameters without first estimating noise variances at both the input and the output. We are able to successfully carry out the asymptotic analysis, and show that measurement noises at both the input and output can be removed from system parameter estimates to achieve the consistent estimation, if the system input is persistently exiting and the measurement data increase asymptotically. Thus, the contribution of this new algorithm lies in a simple

and efficient way to estimate the system parameters without assuming the information from the noise variances.

### 1.4.2 Organization

In Chapter 2, we will present the mathematical background for the EIV system identification. It begins with the introduction of the basic concepts of random processes and the dynamic system model, which will be used for the development of the EIV system identification algorithm. In this chapter, the notation of the system identification problem will also be presented, followed by a description of the LS algorithm, which is widely used but biased in general, and the TLS algorithm in comparison with the LS algorithm. The standard EIV estimation method will be described as well.

Chapter 3 focuses on developing the graph subspace algorithm for identifying the parameterized EIV system model. It begins with the introduction of the notion of the graph subspace for dynamic systems and formulates the EIV identification problem using the concept of the graph subspace. We will demonstrate the asymptotic MLE property for the TLS estimate, and develop the graph subspace algorithm, which is more general than the TLS algorithm, including an iterative procedure to improve the estimation performance. It also contains an implementation procedure in the case when the input and output measurement noises are unknown, but the ratio of the two noise variances is known. The performance of the graph subspace algorithm is illustrated by several examples, presented in the simulation study section.

In Chapter 4, we will study the EIV system identification when the noise variances at the input and the output are unknown and unequal. This is a practical problem in EIV system identification since they are not only unequal but also unknown in practice. The first section introduces the Frisch scheme which is a well-known method to estimate the two different and unknown noise variances, followed by the estimation of system parameters based on the estimated noise variances and the TLS algorithm. Then the use of the Frisch scheme in

combination with the GSA will be studied, which has better performance, compared to that of the TLS algorithm. Also, we will propose a new algorithm, which estimates the system parameters directly without having to estimate the noise variances or using the ratio. Lastly, we will demonstrate the performance of the new algorithm with some simulation results. Chapter 5 concludes this dissertation with a summary of the dissertation research and the proposed topics for future research.

### 1.4.3 Notations

The notation in this dissertation is more or less standard. Denote $\mathbb{R}$ as the collection of all real numbers. Similarly, $\mathbb{R}^n$ denotes the collection of all real column vectors of size $n$. A matrix in $\mathbb{R}^{n\times m}$ is a finite two-dimensional array with elements from $\mathbb{R}$. For random variables, $\mathbb{E}\{\cdot\}$ denotes the expectation operation. In addition, $'$ denotes the transpose operation. Other notations will be made clear as we proceed.

# CHAPTER 2
# MATHEMATICAL PREPARATION

This chapter presents the mathematical background for the EIV system identification. In Section 2.1, we introduce the basic concepts of random processes for the system identification and the dynamic system model that we use for the development of the EIV system identification algorithm. In Section 2.2, the system identification problem is described, plus the notation. Also we present the LS algorithm, widely used for the identification problem. In Section 2.3, we describe the TLS algorithm in comparison with the LS algorithm. In addition, the standard EIV estimation method is presented.

## 2.1 Signals and Systems

### 2.1.1 Random Processes

In light of [83], a deterministic signal is a signal about which there is no uncertainty with respect to its value at any time. In contrast, a random signal is a signal about which there is uncertainty before it occurs. A random variable is described as the possible numerical outcomes of experiments whose results cannot be exactly predicted beforehand [41].

**Probability density function**

Let $x$ be a scalar random variable. Define $p(x)$ as the probability density function (PDF) of $x$, which satisfies

$$p(x) \geq 0, \quad \int_{-\infty}^{\infty} p(x)dx = 1,$$

for all $x$. Define $\mu$ and $\sigma^2$ as the expected or mean value and the variance of the random variable $x$ respectively. Then the Gaussian distributed PDF of $x$ can be written as

$$p(x) = \frac{1}{\sqrt{(2\pi)\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

**Maximum likelihood estimator**

Suppose that there is a group of $N$ independent random variables $x_{1:N}$. That is,

$$x_{1:N} = \{x_1, x_2, \cdots, x_N\}.$$

Let $\mu_i$ and $\sigma_i^2$ be the expected value and variance of $x_i$, respectively. Then, the PDF for $x_i$ can be written as

$$p(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right\}. \tag{2.1}$$

Since $\{x_1, x_2, \cdots, x_N\}$ are independent each other, we can form a likelihood function $\mathcal{L}(x_{1:N})$ that is the joint PDF of $x_{1:N}$ by products of (2.1) for $1 \leq i \leq N$ as follows:

$$\mathcal{L}(x) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right\}. \tag{2.2}$$

If we set $\theta$ as an unknown parameter vector that describes the properties of the observed variable, the maximum likelihood estimator (MLE) seeks $\theta$ which maximizes the likelihood function as follows:

$$\widehat{\theta}_{\mathrm{MLE}} = \underset{\theta}{\operatorname{argmax}}\, \mathcal{L}(\theta; x).$$

**Covariance**

If $x$ is a $D$ dimensional vector-valued random vector, then its covariance matrix is defined as

$$\Sigma_x = \mathbb{E}\{(x - \mu_x)(x - \mu_x)'\},$$

where $\mu_x$ is a mean vector. If $D$ components of $x$ are independent and identically distributed (i.i.d.), then $\Sigma_x$ is a diagonal matrix, and the diagonal elements of $\Sigma_x$ are the variance of $\{x_i\}$ with $x_i$ the $i$th element of $x$. That is,

$$\Sigma_x = \sigma_x^2 I_D, \tag{2.3}$$

where $\sigma_x^2$ is the variance of each $x_i$ and $I_D$ denotes $D$ dimensional identity matrix.

**Power spectral density**

A discrete-time vector random process $\{s(t)\}$ is said to be quasi-stationary, if its auto-correlation matrix

$$R_s(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E}\{s(t)s(t+\tau)'\} \tag{2.4}$$

exists for each integer $\tau$. Let $\Phi_s(\omega)$ be the power spectral density (PSD) of $\{s(t)\}$ defined as the discrete time Fourier transform of $R_s(\tau)$:

$$\Phi_s(\omega) = \sum_{\tau=-\infty}^{\infty} R_s(\tau)\mathrm{e}^{-j\omega\tau} \Longrightarrow R_s(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_s(\omega)\mathrm{e}^{-j\omega\tau}\, d\omega.$$

Then the mean or average power is the expression

$$\mathcal{P}_s := \mathrm{Tr}\left\{R_s(0)\right\} = \mathrm{Tr}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_s(\omega)\, d\omega\right\}. \tag{2.5}$$

Define $\mathcal{S}_n$ as the set of all vector quasi-stationary $\{s(t)\}$ with zero mean and bounded mean power. Then as $N \to \infty$, there holds [41] (Theorem 2.3), w. p. 1,

$$\frac{1}{N}\sum_{t=1}^{N} s(t)s(t-\tau)' \;\to\; R_s(\tau). \tag{2.6}$$

### 2.1.2 Dynamic System Model

In general, in the noise-free case, the dynamics of the single input and single output (SISO) plant model are governed by the difference equation using auto regression with exogeneous variables (ARX) model

$$y(t) + a_1 y(t-1) + \cdots a_{n_a} y(t-n_a) = b_0 u(t) + b_1 u(t-1) + b_2 u(t-2) + \cdots + b_{n_b} u(t-n_b)$$

$$\Longrightarrow y(t) = -\sum_{i=1}^{n_a} a_i y(t-i) + \sum_{i=0}^{n_b} b_i u(t-i), \tag{2.7}$$

where $\{u(t)\}$ is the system input, $\{y(t)\}$ is the system output and $\{a_i\}_{i=1}^{n_a}$, $\{b_i\}_{i=0}^{b_a}$ are the system parameters with integers $n_a \geq 1$ and $n_b \geq 1$. Note that we call

$$s(t) = \begin{bmatrix} y(t) \\ u(t) \end{bmatrix}$$

as a system graph signal. This is the basis of the graph subspace algorithm which will be studied and developed in Chapter 3.

## 2.2 System Identification Problem

### 2.2.1 Notations

Equation (2.7) can be written as

$$y(t) = \phi(t)'\theta, \tag{2.8}$$

where

$$\phi(t) = \begin{bmatrix} -y(t-1) & \cdots & -y(t-n_a) & u(t) & u(t-1) & \cdots & u(t-n_b) \end{bmatrix}'$$

is the regressor and

$$\theta = \begin{bmatrix} \theta_a \\ \theta_b \end{bmatrix}, \quad \theta_a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_a} \end{bmatrix}, \quad \theta_b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{n_b} \end{bmatrix}.$$

The goal of system identification is the estimation of $\theta$ based on noisy input and output measurements. The system under our study described by (2.7) admits transfer function

$P(z)$, given by

$$P(z) = \frac{b(z)}{a(z)}, \quad a(z) = 1 + \sum_{i=1}^{n_a} a_i z^{-i}, \;\; b(z) = \sum_{i=0}^{n_b} b_i z^{-i}. \tag{2.9}$$

When the input and output data are corrupted by the noises, the input and output measurements are described as

$$\tilde{s}(t) = \begin{bmatrix} \tilde{u}(t) \\ \tilde{y}(t) \end{bmatrix} = \begin{bmatrix} u(t) \\ y(t) \end{bmatrix} + \varepsilon(t), \qquad \varepsilon(t) = \begin{bmatrix} \varepsilon_{\text{in}}(t) \\ \varepsilon_{\text{out}}(t) \end{bmatrix}. \tag{2.10}$$

A common assumption on measurement noises is that they are i.i.d. Gaussian processes with the same variance for input and output noises, denoted by $\sigma_\varepsilon^2$. Define

$$v(t) = \frac{u(t)}{a(q)},$$

where $q^{-1}$ is the unit delay operator. Then

$$\begin{bmatrix} u(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a(q) \\ b(q) \end{bmatrix} v(t) \quad \Longrightarrow \quad \tilde{s}(t) = \begin{bmatrix} a(q) \\ b(q) \end{bmatrix} v(t) + \varepsilon(t).$$

The above shows that the estimation of $\theta_a$ and $\theta_b$ is equivalent to blind system identification [74], due to the unknown nature of $\{v(t)\}$.

We have the linear regression model $y(t) = \phi(t)'\theta$ for the noiseless data, yielding

$$\psi(t)' \begin{bmatrix} 1 \\ \theta \end{bmatrix} = 0, \quad \psi(t) = \begin{bmatrix} -y(t) \\ \phi(t) \end{bmatrix}. \tag{2.11}$$

So,

$$\psi(t) = \begin{bmatrix} -y(t) & -y(t-1) & \cdots & -y(t-n_a) & u(t) & u(t-1) & \cdots & u(t-n_b) \end{bmatrix}'$$

can be regarded as the augmented regressor. We set

$$\Psi_N = \begin{bmatrix} \psi(1)' \\ \vdots \\ \psi(N)' \end{bmatrix} = \begin{bmatrix} -T_{Y_{(N)}} & T_{U_{(N)}} \end{bmatrix}. \tag{2.12}$$

The matrices $T_{Y_{(N)}} \in \mathbb{R}^{N\times(n_a+1)}$ and $T_{U_{(N)}} \in \mathbb{R}^{N\times(n_b+1)}$ are Toeplitz specified as

$$T_{Y_{(N)}} = \begin{bmatrix} y(1) & y(0) & \cdots & y(1-n_a) \\ \vdots & y(1) & \ddots & \vdots \\ y(N-n_a) & \ddots & \ddots & y(0) \\ \vdots & \ddots & \ddots & y(1) \\ \vdots & \ddots & \ddots & \vdots \\ y(N) & \cdots & \cdots & y(N-n_a) \end{bmatrix},$$

$$T_{U_{(N)}} = \begin{bmatrix} u(1) & u(0) & \cdots & u(1-n_b) \\ \vdots & u(1) & \ddots & \vdots \\ u(N-n_b) & \ddots & \ddots & u(0) \\ \vdots & \ddots & \ddots & u(1) \\ \vdots & \ddots & \ddots & \vdots \\ u(N) & \cdots & \cdots & u(N-n_b) \end{bmatrix}.$$

The measurement data $\{\tilde{y}(t), \tilde{u}(t)\}$ are noisy and thus

$$\tilde{\Psi}_N = \Psi_N + \mathcal{E}_N \Longleftrightarrow \begin{bmatrix} \tilde{T}_{Y_{(N)}} & \tilde{T}_{U_{(N)}} \end{bmatrix} = \begin{bmatrix} T_{Y_{(N)}} & T_{U_{(N)}} \end{bmatrix} + \begin{bmatrix} \mathcal{E}_{Y_{(N)}} & \mathcal{E}_{U_{(N)}} \end{bmatrix}.$$

The noise matrix $\mathcal{E}_N \in \mathbb{R}^{N\times(n_a+n_b+2)}$ is given by $\mathcal{E}_N = \begin{bmatrix} \mathcal{E}_{Y_{(N)}} & \mathcal{E}_{Y_{(N)}} \end{bmatrix}$ of which both $\mathcal{E}_{Y_{(N)}}$

and $\mathcal{E}_{Y_{(N)}}$ have the Toeplitz structure:

$$\mathcal{E}_{Y_{(N)}} = \begin{bmatrix} \varepsilon_{\text{out}}(1) & \varepsilon_{\text{out}}(0) & \cdots & \varepsilon_y(1-n_a) \\ \vdots & \varepsilon_{\text{out}}(1) & \ddots & \vdots \\ \varepsilon_{\text{out}}(N-n_a) & \ddots & \ddots & \varepsilon_{\text{out}}(0) \\ \vdots & \ddots & \ddots & \varepsilon_{\text{out}}(1) \\ \vdots & \ddots & \ddots & \vdots \\ \varepsilon_{\text{out}}(N) & \cdots & \cdots & \varepsilon_{\text{out}}(N-n_a) \end{bmatrix},$$

$$\mathcal{E}_{U_{(N)}} = \begin{bmatrix} \varepsilon_{\text{in}}(1) & \varepsilon_{\text{in}}(0) & \cdots & \varepsilon_{\text{in}}(1-n_b) \\ \vdots & \varepsilon_{\text{in}}(1) & \ddots & \vdots \\ \varepsilon_{\text{in}}(N-n_b) & \ddots & \ddots & \varepsilon_{\text{in}}(0) \\ \vdots & \ddots & \ddots & \varepsilon_u(1) \\ \vdots & \ddots & \ddots & \vdots \\ \varepsilon_{\text{in}}(N) & \cdots & \cdots & \varepsilon_{\text{in}}(N-n_b) \end{bmatrix}.$$

### 2.2.2 LS Algorithm

Given predictor model $\widehat{y}(t) = \tilde{\phi}(t)'\widehat{\theta}$ and measurements of $\{\tilde{\phi}(t)\}_{t=1}^{N}$, the least square (LS) algorithm is aimed at finding the solution $\widehat{\theta} = \widehat{\theta}_{\text{LS}}$ to minimize

$$J = \frac{1}{N}\sum_{t=1}^{N} |\tilde{y}(t) - \widehat{y}(t)|^2 = \frac{1}{N}\sum_{t=1}^{N} |\tilde{y}(t) - \tilde{\phi}'(t)\widehat{\theta}|^2.$$

Define

$$\tilde{\Phi}_N = \begin{bmatrix} \tilde{\phi}(1)' \\ \vdots \\ \tilde{\phi}(N)' \end{bmatrix}, \quad \tilde{Y}_N = \begin{bmatrix} \tilde{y}(1) \\ \vdots \\ \tilde{y}(N) \end{bmatrix}.$$

Then $\widehat{\theta}_{\text{LS}}$ can be computed by $\widehat{\theta}_{\text{LS}} = (\tilde{\Phi}'_N\tilde{\Phi}_N)^{-1}\tilde{\Phi}'_N\tilde{Y}_N$.

In practice, the output equation error model, described by

$$\tilde{y}(t) = \frac{b(q)}{a(q)}u(t) + e(t)$$

is widely used, which is different from the model shown in Figure 1.2 due to the assumption on the noiseless input $u(t)$. Regardless of the whiteness of $e(t)$, there holds

$$\tilde{y}(t) = [1 - a(q)]\tilde{y}(t) + b(q)u(t) + v_0(t), \quad v_0(t) = a(q)e(t) \Longrightarrow \tilde{Y} = \Phi_N\theta + v_0,$$

where $v_0(t)$ is the additive measurement error. The LS algorithm is motivated by this model, described by

$$\tilde{y}(t) = \phi(t)'\theta + v_0(t), \tag{2.13}$$

where $\phi(t)$ does not involve noises. For instance, if $n_a = 0$ then the model in (2.13) agrees with that in Figure 1.2 by taking

$$v_0(t) = \varepsilon_{\text{out}}(t). \tag{2.14}$$

In this case, $\widehat{\theta}_{\text{LS}}$ is given by

$$\begin{aligned}\widehat{\theta}_{\text{LS}} &:= \left[\frac{1}{N}\sum_{i=1}^{N}\phi(i)\phi(i)'\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^{N}\phi(i)\{\phi(i)'\theta + v_0(i)\}\right] \\ &= \theta + \left[\frac{1}{N}\sum_{i=1}^{N}\phi(i)\phi(i)'\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^{N}\phi(i)v_0(i)\right].\end{aligned}$$

**Definition 1**: The input is persistently exciting of order $n_0$, if

$$\text{rank}\{R_*\} = n_0,$$

where $n_0 = \text{length}(\phi)$ and

$$R_* = \lim_{N\to\infty} R_N, \quad R_N := \frac{1}{N}\sum_{i=1}^{N} \phi(i)\phi(i)'.$$

Suppose $R_*$ has rank $n_0$ and $v_0(t)$ has zero mean and is independent of $\phi(t)$. Then there holds

$$\lim_{N\to\infty} \widehat{\theta}_{\text{LS}}(N) = \theta.$$

**Definition 2**: An estimate $\widehat{\theta}$ is said to be unbiased, if

$$\text{E}\{\widehat{\theta}\} = \theta.$$

Suppose that the input is persistently exciting of order $n_0$. Then the LS estimate is unbiased, if $n_a = 0$, $v_0(t)$ has zero mean and independent of $\phi(t)$. While an unbiased estimate is often desired, it does not mean a biased estimate is always abandoned. In fact people sometimes prefer to use a biased estimate due to its smaller variance error. In general, there is a tradeoff between bias and variance [41]. That is, the expected error can be decomposed as follows:

$$\text{E}\{\| \theta_0 - \widehat{\theta} \|^2\} = \text{bias}^2 + \sigma^2_{\widehat{\theta}} + \sigma^2_{\text{e}},$$

where $\theta_0$ is the true value, $\text{bias}^2$ denotes squared expected bias, $\sigma^2_{\widehat{\theta}}$ is the variance of $\widehat{\theta}$, and $\sigma^2_{\text{e}}$ is the variance of the error.

**Definition 3**: An estimate is called a consistent estimate, if it is asymptotically unbiased. Clearly consistency is required in system identification.

If $\varepsilon_{\text{in}}(t) = 0 \; \forall \; t$, $n_a = 0$, and $v_0(t)$ is white, the LS solution is MLE because $\Phi_N$ does not involve noises, which is known exactly. We can regard $\theta_{\text{LS}}$ as a solution to the perturbed equation:

$$(\tilde{Y} - \Delta_N) = \Phi_N \widehat{\theta} \tag{2.15}$$

where $\Delta_N$ has the smallest Euclidean norm for which $\widehat{\theta}$ has a solution. If $n_a > 0$, then the regressor $\tilde{\phi}(t)$ also depends on $\{v_0(k)\}_{k<t}$, and thus $\mathrm{E}\{\phi(t)v_0(t)\} \neq 0$, leading to the conclusion that the LS estimate is biased. In fact, the LS estimate is not consistent either. Only if $n_a = 0$, which is not realistic, is the LS estimate unbiased.

## 2.3 EIV Estimation

### 2.3.1 TLS Algorithm

The total least square (TLS) algorithm considers a more general case when each entry of $\tilde{\Phi}_N$ involves noise, assuming i.i.d. noise process. In this case we need to perturb not only $\tilde{Y}_N$ but also $\tilde{\Phi}_N$ in hope to "remove" the noise terms. For this purpose we rewrite the approximate equation $\tilde{Y}_N \approx \tilde{\Phi}_N\theta$ into

$$[-\tilde{Y}_N \;\; \tilde{\Phi}_N] \begin{bmatrix} 1 \\ \theta \end{bmatrix} \approx 0. \tag{2.16}$$

Denote that $\tilde{\Psi}_N = [\tilde{Y}_N \;\; \tilde{\Phi}_N] \in \mathbb{R}^{N\times(n_0+1)}$ with $n_0$ is the length of $\theta$. We now would like to perturb $\tilde{\Psi}_N$ to

$$\widehat{\Psi}_N = \tilde{\Psi}_N - \Delta_\Psi,$$

where $\Delta_\Psi = [-\delta Y_N \;\; \delta\Phi_N]$, $\delta Y_N = \tilde{Y}_N - \widehat{Y}_N$ and $\delta\Phi_N = \tilde{\Psi}_N - \widehat{\Psi}_N$, which is having the smallest induced 2-norm, i.e., $\| \Delta_\Psi \|_2 := \overline{\sigma}(\Delta_\Psi)$ such that,

$$\tilde{Y}_N - \delta Y_N \in \mathscr{R}(\tilde{\Phi}_N - \delta\Phi_N).$$

The above is equivalent to rank $\{\tilde{\Psi}_N - \Delta_\Psi\} = n_0$, even though $\tilde{\Psi}_N - \Delta_\Psi \in \mathbb{R}^{N\times(n_0+1)}$. Indeed because rank $\{\tilde{\Psi}_N - \Delta_\Psi\} = n_0 < n_0 + 1$, the exact equality in (2.16) can be made true and thus there exists some nonzero solution in the form of $\begin{bmatrix} 1 & \widehat{\theta}' \end{bmatrix}'$.

In light of SVD, $\tilde{\Psi}_N = U\Sigma V'$ where

$$\Sigma'\Sigma = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_{n_0+1}^2), \;\; U = \begin{bmatrix} u_1 & \cdots & u_{n_0+1} \end{bmatrix}, \;\; V = \begin{bmatrix} v_1 & \cdots & v_{n_0+1} \end{bmatrix}.$$

The perturbation $\Delta_\Psi$ with the smallest induced 2-norm is given by $\Delta_\Psi = \sigma_{n_0+1} u_{n_0+1} v'_{n_0+1}$.

Based on the solution $\Delta_\Psi = \sigma_{n_0+1} u_{n_0+1} v'_{n_0+1}$, there holds

$$\left(\tilde{\Psi}_N - \Delta_\Psi\right) \begin{bmatrix} 1 \\ \widehat{\theta}_{\mathrm{TLS}} \end{bmatrix} = \left(\sum_{i=1}^{n} \sigma_i u_i v_i'\right) v_{n_0+1}/v_{1,n_0+1} = 0 \implies \begin{bmatrix} 1 \\ \widehat{\theta}_{\mathrm{TLS}} \end{bmatrix} = \frac{v_{n_0+1}}{v_{1,n_0+1}}, \quad (2.17)$$

by the orthogonality of $\{v_i\}_{i=1}^{n_0}$ to $v_{n+1}$ where $v_{1,n_0+1} \neq 0$ is the first element of $v_{n_0+1}$. In summary, $\widehat{\theta}_{\mathrm{TLS}}$ can be obtained from $v_{n_0+1} \in \mathbb{R}^{n_0+1}$ by taking the last $n_0$ elements of $v_{n_0+1}$, divided by its first element that is nonzero with probability 1 [41].

Alternatively, the TLS estimate can also be computed from eigenvalue decomposition (EVD) of $\tilde{\Psi}'_N \tilde{\Psi}_N$ by computing its minimum eigenvalue (equal to $\sigma_{n_0+1}^2$) and the corresponding eigenvector that is $v_{n_0+1}$ due to

$$\tilde{\Psi}'_N \tilde{\Psi}_N = V S^2 V' \;\; \Rightarrow V S^2 V' \begin{bmatrix} 1 \\ \widehat{\theta}_{\mathrm{TLS}} \end{bmatrix} = \sigma_{n+1}^2 \begin{bmatrix} 1 \\ \widehat{\theta}_{\mathrm{TLS}} \end{bmatrix}. \quad (2.18)$$

Partition $V \in \mathbb{R}^{(n_0+1)\times(n_0+1)}$ according to

$$V = \begin{bmatrix} v_{11} & v_{12} \\ V_{21} & v_{22} \end{bmatrix}, \;\; V_{21} \in \mathbb{R}^{n_0 \times n_0}, \;\; v_{22} \in \mathbb{R}^{n_0}, \;\; v'_{11} \in \mathbb{R}^{n_0}.$$

Equality (2.18) implies $\widehat{\theta}_{\mathrm{TLS}} = v_{22} v_{12}^{-1} = -(V'_{21})^{-1} v'_{11}$ in light of $v_{12} = v_{1,n+1}, v_{22} = v_{2:n_0+1,n_0+1}$ and

$$\begin{bmatrix} v'_{11} & V'_{21} \end{bmatrix} \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix} = v'_{11} v_{12} + V'_{21} v_{22} = 0.$$

It is shown in [41] that by persistently exciting input the TLS estimate is unbiased asymptotically. The following results are new.

**Theorem 1.** *Assume that the input is persistently exciting and* $\{\varepsilon(t)\}$ *are independent Gaussian distributed random vectors with zero mean and known covariance* $\sigma_\varepsilon^2 I_2$ *for all* $t$*. If* $n_a = n_b$*, then the TLS estimate* $\widehat{\theta}_{\rm TLS}$ *is MLE asymptotically.*

Proof: We follow the same procedure as in [26]. The hypothesis on the measurement noise implies that its PDF is given by

$$f_\varepsilon(\{\varepsilon(t)\}) = \prod_{t=1}^{N} \frac{1}{2\pi\sigma_\varepsilon^2} \exp\left\{-\frac{[\tilde{y}(t)-y(t)]^2+[\tilde{u}(t)-u(t)]^2}{2\sigma_\varepsilon^2}\right\}.$$

The MLE $\widehat{\theta}_{\rm ML}$ maximizes $f_\varepsilon(\cdot)$, which is equivalent to the minimization of

$$\begin{aligned}\widehat{\varepsilon}_N^2 &:= \frac{1}{N}\sum_{i=1}^{N}[\tilde{y}(t)-y(t;\widehat{\theta}_{\rm ML})]^2+[\tilde{u}(t)-u(t;\widehat{\theta}_{\rm ML})]^2 = \frac{1}{N}\sum_{i=1}^{N}[\widehat{\varepsilon}_{\rm out}(t)^2+\widehat{\varepsilon}_{\rm in}(t)^2] \\ &= \frac{1}{N(n_a+1)}\left(\mathrm{Tr}\left\{\widehat{\mathcal{E}}'_{Y_{(N)}}\widehat{\mathcal{E}}_{Y_{(N)}}\right\}-\sum_{k=1}^{n_a}k[\widehat{\varepsilon}_{\rm out}(k-n_a)^2+\widehat{\varepsilon}_{\rm out}(N-k+1)^2]\right) \\ &+ \frac{1}{N(n_b+1)}\left(\mathrm{Tr}\left\{\widehat{\mathcal{E}}'_{U_{(N)}}\widehat{\mathcal{E}}_{U_{(N)}}\right\}-\sum_{k=1}^{n_b}k[\widehat{\varepsilon}_{\rm in}(k-n_b)^2+\widehat{\varepsilon}_{\rm in}(N-k+1)^2]\right).\end{aligned}$$

The above implies that if $N >> 1$, then there holds approximately

$$\widehat{\varepsilon}_N^2 \approx \frac{1}{N(n_a+1)}\left(\mathrm{Tr}\left\{\widehat{\mathcal{E}}'_{Y_{(N)}}\widehat{\mathcal{E}}_{Y_{(N)}}\right\}\right)+\frac{1}{N(n_b+1)}\left(\mathrm{Tr}\left\{\widehat{\mathcal{E}}'_{U_{(N)}}\widehat{\mathcal{E}}_{U_{(N)}}\right\}\right).$$

In light of the Gaussian assumption and the TLS solution that minimizes $\widehat{\varepsilon}_N^2$, as $N\to\infty$, there hold asymptotically

$$\frac{1}{N}\widehat{\mathcal{E}}'_{U_{(N)}}\widehat{\mathcal{E}}_{U_{(N)}} \to \sigma_\varepsilon^2 I_{n_b+1}, \quad \frac{1}{N}\widehat{\mathcal{E}}'_{Y_{(N)}}\widehat{\mathcal{E}}_{Y_{(N)}} \to \sigma_\varepsilon^2 I_{n_a+1}.$$

The above implies that $\widehat{\varepsilon}_N^2$ approaches asymptotically to

$$\widehat{\varepsilon}_N^2 \to \frac{1}{n_a+1}\mathrm{Tr}\left\{\sigma_\varepsilon^2 I_{n_a+1}\right\} + \frac{1}{n_b+1}\mathrm{Tr}\{\sigma_\varepsilon^2 I_{n_b+1}\} = \frac{1}{n_0+1}\mathrm{Tr}\{\sigma_\varepsilon^2 I_{n_0+1}\} = \sigma_\varepsilon^2.$$

It follows from [26] that the TLS is indeed approximately MLE, and approaches MLE as $N \to \infty$. □

In the case where $n_b \neq n_a$, we can add dummy parameters into $\theta_a$ or $\theta_b$ so that $n_b = n_a$ is true. These dummy parameters can be set to zero after estimation is done. Hence, the case $n_b \neq n_a$ does not pose a serious issue.

### 2.3.2 Standard EIV Estimation

We present an overview on standard EIV estimation, and its ML estimator based chiefly on the results in [26]. Let $\Theta \in \mathbb{R}^{p\times n}$ be the parameter matrix of interest satisfying

$$\Theta X = Y, \qquad X \in \mathbb{R}^{n\times N}, \qquad Y \in \mathbb{R}^{p\times N}, \tag{2.19}$$

where $N > n + p$. This is a so-called multivariate problem ($p \geq 1$), dual to the conventional equation of $X\Theta = Y$, due to the need for system identification in the multi-input and multiple-output (MIMO) setting. The precise values of $X$ and $Y$ are unknown. Instead only their measurements, denoted by $\tilde{X}$ and $\tilde{Y}$, respectively, are available, giving rise to the EIV model

$$\begin{bmatrix} \tilde{Y} \\ \tilde{X} \end{bmatrix} = \begin{bmatrix} Y \\ X \end{bmatrix} + E, \tag{2.20}$$

where elements of $E$ are i.i.d. random variables. Denote

$$\tilde{R} = \frac{1}{N}\begin{bmatrix} \tilde{Y} \\ \tilde{X} \end{bmatrix}\begin{bmatrix} \tilde{Y}' & \tilde{X}' \end{bmatrix} \in \mathbb{R}^{(n+p)\times(n+p)}, \tag{2.21}$$

which is the second-order statistics of the measurement samples. The following result holds.

**Theorem 2.** *Suppose that elements of $E$ are normal* i.i.d. *with mean zero and variance $\sigma^2$. Let $\tilde{R} = K\Lambda K'$ be the eigenvalue decomposition (EVD) with*

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{n+p})$$

*arranged in* ascending *order, and $K$ of the same size as $\tilde{R}$ being orthogonal. Partition*

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}, \tag{2.22}$$

*where $K_{11}, \Lambda_1 \in \mathbb{R}^{p\times p}$. Then $K_{11}$ and $K_{22}$ are nonsingular with probability 1 (w. p. 1), and*

$$\widehat{\Theta} = K_{12}K_{22}^{-1} = -(K_{11}')^{-1}K_{21}', \quad \widehat{\sigma}^2 = \frac{\mathrm{Tr}\{\Lambda_1\}}{n+p}, \tag{2.23}$$

*are ML estimators for $\Theta$ and $\sigma^2$, respectively, with $\mathrm{Tr}\{\cdot\}$ denoting the trace operation.*

The above result is quoted from [26] (Lemma 2.2 and Theorem 2.3) with only minor changes to fit the EIV identification. It is noted that the ML estimate $\widehat{\Theta}$ is the TLS solution to $\Theta\tilde{X} \approx \tilde{Y}$ [77], different from the ordinary LS solution $\widehat{\Theta}_{\mathrm{LS}} = \tilde{Y}\tilde{X}'(\tilde{X}\tilde{X}')^{-1}$, assuming $\det(\tilde{X}\tilde{X}') \neq 0$.

**Remark 1.** (a) There holds a strong consistency or $\widehat{\Theta} \to \Theta$ w. p. 1 as $N \to \infty$, provided that

$$\lim_{N\to\infty} \frac{1}{N} XX' = R_X > 0. \tag{2.24}$$

The condition for the weak consistency is also established in [24] that weakens $\frac{1}{N}$ in (2.24) to $\frac{1}{\sqrt{N}}$. It is noted that the ML estimates in (2.23) require to compute only the $p$ smallest eigenvalues of $\tilde{R}$ and their respective eigenvectors. If $p = 1$, both $\Lambda_1$ and $K_{11}$ are reduced to scalars, and $\Lambda_1 = \lambda_{\min}$ is the minimum eigenvalue of $\tilde{R}$.

(b) The ML estimators in (2.23) are obtained based on the Gaussian assumption, but their strong consistency is not. The ML estimate, denoted by $\widehat{\Theta}$, is called the generalized LS

solution to $\Theta\tilde{X} \approx \tilde{Y}$ [26], which is independent of the ML estimate $\widehat{\sigma}$. Moreover, the ML estimate $\widehat{\Theta}$ is obtained by first computing the ML estimates $(\widehat{X}, \widehat{Y})$ to $(\tilde{X}, \tilde{Y})$ subject to the rank condition

$$\operatorname{rank}\left\{\begin{bmatrix} \widehat{Y} \\ \widehat{X} \end{bmatrix}\right\} = n, \tag{2.25}$$

and then computing the ML estimate $\widehat{\Theta}$ as the unique solution to linear equation $\widehat{\Theta}\widehat{X} = \widehat{Y}$ [26, 77]. □

# CHAPTER 3
# GRAPH SUBSPACE APPROACH

This chapter is focused on developing the graph subspace algorithm for identifying the parameterized EIV system model. The notion of the system graph has been employed in [25, 84] for robust control. Its use in system identification will be proved to be productive. In Section 3.1, we will introduce the notion of the graph subspace for dynamic systems, and formulate the EIV identification problem using the concept of graph subspace. In Section 3.2, we show the asymptotic MLE property for the TLS estimate and develop the graph subspace algorithm which is more general than the TLS algorithm. In addition, an iterative procedure is proposed to improve the estimation performance. In Section 3.3, we discuss how to modify the algorithm for implementation in the case when the input and output measurement noises are unknown, but the ratio of the two noise variances is known. In Section 3.4, we present the numerical studies with several examples to illustrate the performance of the graph subspace algorithm.

## 3.1 Preliminaries

### 3.1.1 Graph Subspace

For simplicity, we consider only single-input and single-output (SISO) systems, although the results are applicable to multivariable systems. Let

$$P(z) = \frac{b(z)}{a(z)}, \quad \begin{bmatrix} b(z) \\ a(z) \end{bmatrix} = \sum_{k=0}^{n} \begin{bmatrix} b_k \\ a_k \end{bmatrix} z^{n-k}, \tag{3.1}$$

with $a_0 = 1$, be the transfer function of a given discrete-time SISO system. Parameterized model representation in (3.1) has no loss of generality for finite dimensional systems. Assume that $a(z)$ and $b(z)$ are coprime. Suppose that $a(z)$ is a Schur polynomial. Denote $\mathcal{B}$ as the set of all bounded signals. Then for each input $\{u(t)\} \in \mathcal{B}$, $y(t) = P(q)u(t) \in \mathcal{B}$ as well.

That is, the SISO system represented by $P(z)$ maps $\mathcal{B}$ into $\mathcal{B}$. The graph associated with $P(z)$, denoted as $\mathcal{G}_P$, consists of all such pairs $\{u(t), y(t)\} = \{u(t), P(q)u(t)\}$ that induce the graph metric [84].

The graph associated with $P(z)$ can be represented in an alternative way. Define

$$G(z) := \begin{bmatrix} b(z) \\ a(z) \end{bmatrix} z^{-n} = \sum_{k=0}^{n} \begin{bmatrix} b_k \\ a_k \end{bmatrix} z^{-k}. \tag{3.2}$$

Then the graph of $P(z)$ is given by $\mathcal{G}_P = G(q)\mathcal{B}$ that is a closed subspace [25, 84], and is termed as graph subspace in this dissertation. Define $v(t) := a(q)^{-1}u(t)$. It is clear that by the Schur stability of $a(z)$,

$$u(t) \in \mathcal{B} \quad \Longleftrightarrow \quad v(t) = a(q)^{-1}u(t) \in \mathcal{B}. \tag{3.3}$$

It follows that $s(t) := G(q)v(t) \in \mathcal{B}_2 := \mathcal{B} \oplus \mathcal{B}$ for each $v(t) \in \mathcal{B}$. More importantly,

$$\mathcal{G}_P = \left\{ s(t) = \begin{bmatrix} y(t) \\ u(t) \end{bmatrix} = G(q)v(t) : \quad v(t) \in \mathcal{B} \right\} \tag{3.4}$$

constitutes the graph subspace of $P(z)$.

Alternatively, denote

$$T_G(z) := z^{-n} \begin{bmatrix} -a(z) & b(z) \end{bmatrix} = G(z)' \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{3.5}$$

Then the graph subspace of $P(z)$ can be defined as

$$\mathcal{G}_P = \{ s(t) \in \mathcal{B}_2 : \quad T_G(q)s(t) = 0 \}. \tag{3.6}$$

Let the true parameter vector $\theta$, and blocked $s(t)$ of $(\ell+1)$-fold be denoted respectively by

$$\begin{aligned} \theta &= \left[\begin{array}{cccccc} b_0 & -a_1 & b_1 & \cdots & -a_n & b_n \end{array}\right]', \\ \psi_{s,\ell}(t) &= \left[\begin{array}{cccc} s(t)' & s(t-1)' & \cdots & s(t-\ell)' \end{array}\right]', \end{aligned} \tag{3.7}$$

satisfying $\ell \geq n$. So $\psi_{s,\ell}(t) \in \mathbb{R}^{2(\ell+1)}$. In the case of $\ell = n$, the equality $T_G(q)s(t) = 0$ in (3.6) is equivalent to

$$\left[\begin{array}{cc} -1 & \theta' \end{array}\right] \psi_{s,n}(t) = 0 \iff y(t) = \phi(t)'\theta$$

with $\phi(t)$ a permutated regressor:

$$\phi(t) = \left[\begin{array}{cccccc} u(t) & y(t-1) & u(t-1) & \cdots & y(t-n) & u(t-n) \end{array}\right]'$$

As a result, there exists an invertible map from $\mathcal{G}_P$ to

$$\underline{\mathcal{G}}_P := \left\{ \psi_{s,n}(t) \in \mathcal{B}_{2(n+1)} : \quad \left[\begin{array}{cc} -1 & \theta' \end{array}\right] \psi_{s,n}(t) = 0 \right\}.$$

### 3.1.2 EIV Identification

The input and output measurements are described by

$$\tilde{s}(t) = \left[\begin{array}{c} \tilde{y}(t) \\ \tilde{u}(t) \end{array}\right] = s(t) + \varepsilon(t), \quad \varepsilon(t) = \left[\begin{array}{c} \varepsilon_{\text{out}}(t) \\ \varepsilon_{\text{in}}(t) \end{array}\right]. \tag{3.8}$$

A common assumption on measurement noises for EIV systems is that $\{\varepsilon_{\text{in}}(t), \varepsilon_{\text{out}}(t)\}$ are i.i.d. Gaussian distributed with common variance $\sigma_\varepsilon^2$ and independent of $\{u(t), y(t)\}$. Also, we assume that a total of $(N+\ell)$ pairs of noisy input-output measurements $\{\tilde{u}(t), \tilde{y}(t)\}_{t=t_0}^{t_f}$ are available where $\ell \geq n$ and $t_f - t_0 = N + \ell - 1$. The problem of EIV identification is aimed at estimating the true parameter vector $\theta$ based on noisy input-output measurements

$\{\tilde{u}(t), \tilde{y}(t)\}_{t=t_0}^{t_f}$. Let $t_{0,\ell} = t_0 + \ell$. An $N$-column matrix

$$\tilde{\Psi}_{\ell,N}^{(s)} = \begin{bmatrix} \tilde{\psi}_{s,\ell}(t_{0,\ell}) & \tilde{\psi}_{s,\ell}(t_{0,\ell}+1) & \cdots & \tilde{\psi}_{s,\ell}(t_f) \end{bmatrix} \tag{3.9}$$

can be formed. Clearly $\tilde{\psi}_{s,\ell}(t) \in \mathbb{R}^{2(\ell+1)}$ by

$$\tilde{\psi}_{s,\ell}(t) = \psi_{s,\ell}(t) + \psi_{\varepsilon,\ell}(t),$$

where $\psi_{\varepsilon,\ell}(t)$ is similar to $\psi_{s,\ell}(t)$ with a difference in $u(t)$ and $y(t)$ being replaced by $\varepsilon_{\text{in}}(t)$ and $\varepsilon_{\text{out}}(t)$, respectively. It follows that

$$\tilde{\Psi}_{\ell,N}^{(s)} = \Psi_{\ell,N}^{(s)} + \Psi_{\ell,N}^{(\varepsilon)} \tag{3.10}$$

with $\Psi_{\ell,N}^{(s)}$ consisting of noiseless input-output measurement data and $\Psi_{\ell,N}^{(\varepsilon)}$ the error component of $\tilde{\Psi}_{\ell,N}^{(s)}$. There holds

$$\begin{bmatrix} -1 & \theta' \end{bmatrix} \Psi_{n,N}^{(s)} = 0. \tag{3.11}$$

The input $\{u(t)\}$ is said to be persistently exciting (PE), if

$$\operatorname{rank} \left\{ \frac{1}{N} \Psi_{\ell,N}^{(s)} \Psi_{\ell,N}^{(s)'} \right\} = \ell + n + 1, \tag{3.12}$$

for sufficiently large $N$, in which case there exists an invertible map from the range space of $\Psi_{\ell,N}^{(s)}$ to the graph subspace of $P(z)$. Recall the previous subsection. The PE condition is assumed throughout the dissertation; otherwise the system parameters are not identifiable even if $N \to \infty$.

**Remark 2.** It is beneficial to compare the EIV system identification with the standard EIV

estimation. By taking

$$\begin{bmatrix} \tilde{Y} \\ \tilde{X} \end{bmatrix} = \tilde{\Psi}_{\ell,N}^{(s)}, \quad \begin{bmatrix} Y \\ X \end{bmatrix} = \Psi_{\ell,N}^{(s)}, \quad E = \Psi_{\ell,N}^{(\varepsilon)},$$

and $\Theta = \theta'$ with $\tilde{Y}$ and $Y$ one-dimensional row vectors, i.e., $p = 1$, both (2.19) and (2.20) hold true in the case of $\ell = n$ in light of (3.11). Moreover, the PE condition in (3.12) is equivalent to (2.24). More importantly, the procedure derived in [26] for computing the MLE agrees with the TLS. Consequently the estimator associated with the TLS is a consistent estimator. The difference lies in the Toeplitz structure for $\tilde{\Psi}_{\ell,N}^{(s)}$. That is, both $\Psi_{\ell,N}^{(s)}$ and $E = \Psi_{\ell,N}^{(\varepsilon)}$ are (block) Toeplitz matrices. For this reason, the ML property for the standard EIV estimation shown in [26] does not hold anymore in general for the EIV identification. Nevertheless, the above observation will be useful in the next section. □

## 3.2 Main Results

### 3.2.1 Asymptotic MLE Property

It is clear that the TLS algorithm is the same as the MLE algorithm in [26] in the case of $\ell = n$. Define

$$\tilde{R}_n = \frac{1}{N} \tilde{\Psi}_{n,N}^{(s)} \tilde{\Psi}_{n,N}^{(s)'}.$$

The TLS computes the EVD of $\tilde{R}_n$ and it estimates $\widehat{\theta}$ using the eigenvector corresponding to the smallest eigenvalue. Specifically, if the eigenvector corresponding to the smallest eigenvalue of $\tilde{R}_n$ is

$$\tilde{v}_\eta = \begin{bmatrix} \tilde{v}_{\eta,1} & \tilde{v}_{\eta,2} & \cdots & \tilde{v}_{\eta,2(n+1)} \end{bmatrix}' \in \mathbb{R}^{2(n+1)},$$

where $\eta$ is the index of the eigenvector corresponding to the smallest eigenvalue of $\tilde{R}_n$, for example, $\eta = 1$ if the eigenvalues are in ascending order. Then the TLS algorithm sets

$$\widehat{\theta} = -\begin{bmatrix} \tilde{v}_{\eta,2} & \cdots & \tilde{v}_{\eta,2(n+1)} \end{bmatrix}' \tilde{v}_{\eta,1}^{-1} \tag{3.13}$$

as the estimated parameter vector. Recall (2.23) in Theorem 2 and that $\tilde{v}_{\eta,1} \neq 0$ w.p.1. Because the eigen-matrix of $\tilde{R}_n$ is the same as the left singular matrix of $\frac{1}{\sqrt{N}}\tilde{\Psi}_{n,N}^{(s)}$, singular value decomposition (SVD) can be used to compute the parameter estimate $\widehat{\theta}$ as well. The next result shows that the TLS produces an asymptotic MLE.

**Theorem 3.** *Let $\{\tilde{y}(t), \tilde{u}(t)\}_{t=t_0}^{t=t_f}$ be noisy input-output measurements of the EIV system described in (3.1) and (3.8), illustrated in Figure 1.2 with $N = t_f - t_0 + 1 >> 1$. For $\ell = n$, the estimated parameter vector $\widehat{\theta}$ as obtained in (3.13) is an asymptotic MLE of the true system parameter vector $\theta$. Under the PE condition in (3.12) and $\ell \geq n$, there holds $\widehat{\theta} \to \theta$ as $N \to \infty$.*

Proof: We follow the procedure in [26]. Let $\{\widehat{y}(t), \widehat{u}(t)\}_{t=t_0}^{t=t_f}$ be respective estimates of $\{\tilde{y}(t), \tilde{u}(t)\}_{t=t_0}^{t=t_f}$ satisfying

$$\text{rank}\left\{\widehat{\Psi}_{n,N}^{(s)}\right\} = 2n+1, \tag{3.14}$$

where $\widehat{\Psi}_{n,N}^{(s)}$ is defined in the same way as $\tilde{\Psi}_{n,N}^{(s)}$ except that $\{\tilde{y}(t), \tilde{u}(t)\}_{t=t_0}^{t=t_f}$ are replaced by $\{\widehat{y}(t), \widehat{u}(t)\}_{t=t_0}^{t=t_f}$, respectively. It follows from the hypotheses on Gaussian distribution and i.i.d. of the EIV system model that the MLE seeks $\{\widehat{y}(t), \widehat{u}(t)\}_{t=t_0}^{t=t_f}$ to minimize the following functional (that is the negative log-likelihood function, plus a constant):

$$f(\cdot) = \frac{1}{2\sigma_\varepsilon^2}\left(\sum_{t=t_0}^{t_f}[\tilde{u}(t) - \widehat{u}(t)]^2 + [\tilde{y}(t) - \widehat{y}(t)]^2\right),$$

subject to rank condition (3.14). It can be verified that

$$f(\cdot) = \frac{1}{2n_+\sigma_\varepsilon^2}\left[\text{Tr}\left\{\left(\tilde{\Psi}_{n,N}^{(s)} - \widehat{\Psi}_{n,N}^{(s)}\right)\left(\tilde{\Psi}_{n,N}^{(s)} - \widehat{\Psi}_{n,N}^{(s)}\right)'\right\}\right.$$

$$+\sum_{k=1}^{n}(n-k)\left\{\widehat{\varepsilon}_{\text{in}}(t_0+k)^2+\widehat{\varepsilon}_{\text{in}}(t_f-k)^2\right\}$$
$$+\sum_{k=1}^{n}(n-k)\left\{\widehat{\varepsilon}_{\text{out}}(t_0+k)^2+\widehat{\varepsilon}_{\text{out}}(t_f-k)^2\right\}\Bigg],$$

where $\widehat{\varepsilon}_{\text{in}}(k)=\tilde{u}(k)-\widehat{u}(k)$, $\widehat{\varepsilon}_{\text{out}}(k)=\tilde{y}(k)-\widehat{y}(k)$, and $n_+=n+1$. As a result, the negative log-likelihood function is dominated by the first term, and therefore

$$\frac{2n_+\sigma_\varepsilon^2}{N}f(\cdot)\approx \text{Tr}\left\{\frac{1}{N}\left(\tilde{\Psi}_{n,N}^{(s)}-\widehat{\Psi}_{n,N}^{(s)}\right)\left(\tilde{\Psi}_{n,N}^{(s)}-\widehat{\Psi}_{n,N}^{(s)}\right)'\right\}$$

with an error in the order of $\mathcal{O}(\frac{n^2}{N})$ that approaches zero as $N$ approaches infinity. It follows that minimization of the above trace yields the TLS algorithm that minimizes the negative log-likelihood function asymptotically. Consistency follows from the PE condition and

$$\tilde{R}_{\ell,\sigma}:=\lim_{N\to\infty}\frac{1}{N}\tilde{\Psi}_{\ell,N}^{(s)}\tilde{\Psi}_{\ell,N}^{(s)'}=\text{E}\{\psi_{s,\ell}(t)\psi_{s,\ell}(t)'\}+\sigma_\epsilon^2 I.$$

Hence, noise covariance can be removed asymptotically, even if $\sigma_\epsilon^2$ is unknown. Indeed, the $i$th eigenvalue of $\tilde{R}_{\ell,\sigma}$ is the same as $\sigma_\epsilon^2$ for $0<i\leq \ell-n+1$ if the eigenvalues are arranged in ascending order, in light of (3.12). Therefore, $\sigma_\epsilon^2$ can be estimated consistently, and removed as $N\to\infty$. It follows that parameter vector $\theta$ can be recovered asymptotically, which concludes the proof. □

**Remark 3.** The proof of Theorem 3 shows that the TLS algorithm minimizes an approximation of $f(\cdot)$ that is the negative log-likelihood function plus a constant, with the approximation error in the order of $\mathcal{O}(\frac{n^2}{N})$. Therefore, the TLS algorithm can be regarded as an approximate MLE for estimating the system parameters. □

In the case of $\ell=n$, the range space of eigenvector $\tilde{v}_\eta$ corresponding to the smallest eigenvalue of $\tilde{R}_n$ can be regarded as a noise subspace, and the range space spanned by the other $(2n+1)$ eigenvectors of $\tilde{R}_n$ can be regarded as a signal subspace. The notions of these two subspaces are proposed and studied in [1, 46] in order to solve the problem of

blind channel estimation. The parameter vector can be estimated by making use of the orthogonality of the two subspaces.

In the case when $\ell \neq n$, let

$$\tilde{R}_\ell = \frac{1}{N}\tilde{\Psi}_{\ell,N}^{(s)}\tilde{\Psi}_{\ell,N}^{(s)\prime}, \quad \ell > n. \tag{3.15}$$

Under the hypotheses of the Gaussian distribution and i.i.d. of the EIV system model, the asymptotic MLE, such as the one in Theorem 3, seeks block Toeplitz matrices $\widehat{\Psi}_{\ell,N}^{(s)}$ to minimize

$$\mathrm{Tr}\left\{\frac{1}{N}\left(\tilde{\Psi}_{\ell,N}^{(s)} - \widehat{\Psi}_{\ell,N}^{(s)}\right)\left(\tilde{\Psi}_{\ell,N}^{(s)} - \widehat{\Psi}_{\ell,N}^{(s)}\right)'\right\}, \tag{3.16}$$

subject to the rank condition

$$\mathrm{rank}\left\{\widehat{\Psi}_{\ell,N}^{(s)}\right\} = \ell + n + 1, \tag{3.17}$$

in light of (3.12). The restriction on the Toeplitz structure (the same as that of $\tilde{\Psi}_{\ell,N}^{(s)}$) in the case of $\ell > n$ differs from the case of $\ell = n$ and renders the problem of approximate MLE much harder.

### 3.2.2 GSA Algorithm

In this subsection, we formally introduce the GSA, the generalization of the TLS for EIV identification. Let $\tilde{R}_\ell = \tilde{K}\tilde{\Lambda}\tilde{K}'$ be the EVD. Then $\tilde{K} = \begin{bmatrix} \tilde{K}_\eta & \tilde{K}_s \end{bmatrix}$ of which columns of $\tilde{K}_\eta$ span the (approximate) noise subspace, and columns of $\tilde{K}_s$ span the (approximate) signal subspace. In addition, the noise subspace has dimension $d_\eta = \ell - n + 1$, and the signal subspace has dimension $d_s = \ell + n + 1$ [1, 46]. Similarly, $\tilde{\Lambda} = \mathrm{diag}(\tilde{\Lambda}_\eta, \tilde{\Lambda}_s)$ with

$$\tilde{\Lambda}_\eta = \mathrm{diag}(\tilde{\lambda}_1, \cdots, \tilde{\lambda}_{d_\eta}), \quad \tilde{\Lambda}_s = \mathrm{diag}(\tilde{\lambda}_{d_\eta+1}, \cdots, \tilde{\lambda}_{2(\ell+1)}),$$

and $\{\tilde{\lambda}_i\}$ arranged in ascending order. Partition

$$\begin{aligned} \tilde{K}_\eta &= \left[ \begin{array}{ccc} \tilde{v}_{\eta_1} & \cdots & \tilde{v}_{\eta_{d_\eta}} \end{array} \right] \in \mathbb{R}^{2(\ell+1)\times d_\eta}, \\ \tilde{K}_s &= \left[ \begin{array}{ccc} \tilde{v}_{s_1} & \cdots & \tilde{v}_{s_{d_s}} \end{array} \right] \in \mathbb{R}^{2(\ell+1)\times d_s}. \end{aligned} \tag{3.18}$$

Recall the true system parameter vector $\theta$. Denote

$$\underline{\theta} = \left[ \begin{array}{c} -1 \\ \theta \\ 0 \\ \vdots \\ 0 \end{array} \right], \quad S = \left[ \begin{array}{ccccc} 0 & \cdots & \cdots & \cdots & 0 \\ 1 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{array} \right] \tag{3.19}$$

as the length increased parameter vector of dimension $2(\ell+1)$, and the shift matrix of dimension $2(\ell+1)\times 2(\ell+1)$, respectively. Define matrix function

$$M_\eta(\theta) := \left[ \begin{array}{cccc} \underline{\theta} & S^2\underline{\theta} & \cdots & S^{2(d_\eta-1)}\underline{\theta} \end{array} \right] \in \mathbb{R}^{2(\ell+1)\times d_\eta}.$$

Then in the noise-free case, i.e., $\sigma_\varepsilon = 0$, there holds

$$M_\eta(\theta)' K_s = 0, \quad K_s := \tilde{K}_s\Big|_{\sigma_\varepsilon=0}.$$

In the noisy case of $\sigma_\varepsilon > 0$, $\widehat{\theta}$ is sought such that equation

$$M_\eta(\widehat{\theta})' \tilde{K}_s = 0 \tag{3.20}$$

admits a unique solution $\widehat{\theta}$, which imposes the block Toeplitz structure on $\widehat{\Psi}^{(s)}_{\ell,N}$.

Because of the lack of the mathematical tools, we remove the restriction on the block Toeplitz structure for $\widehat{\Psi}^{(s)}_{\ell,N}$ in minimizing cost functional (3.16), subject to condition (3.17).

As such, there does not exist a unique solution to (3.20) in general, in contrast to the case of $\ell = n$. Consequently we search for $\widehat{\theta}$ such that some cost functional of

$$M_\eta(\widehat{\theta})' \tilde{K}_s \tilde{\Lambda}_s^{1/2}$$

is minimized. Inclusion of $\tilde{\Lambda}_s^{1/2}$ departs the subspace algorithm developed in [1, 46]. For this purpose, partition

$$\tilde{K}_s \tilde{\Lambda}_s^{1/2} = \begin{bmatrix} \tilde{S}_1 \\ \vdots \\ \tilde{S}_{\ell+1} \end{bmatrix}, \quad \tilde{S}_i \in \mathbb{R}^{2 \times d_s}. \tag{3.21}$$

It is easy to see that

$$M_\eta(\widehat{\theta})' \tilde{K}_s \tilde{\Lambda}_s^{1/2} \approx 0$$

is equivalent to

$$\begin{bmatrix} -1 & \widehat{\theta}' \end{bmatrix} T_{\tilde{S}} \approx 0 \tag{3.22}$$

where $T_{\tilde{S}}$ is a block Toeplitz matrix with $2(n+1)$ rows, specified by

$$T_{\tilde{S}} = \begin{bmatrix} \tilde{S}_{d_\eta} & \cdots & \tilde{S}_{n+1} & \cdots & \tilde{S}_1 \\ \vdots & \ddots & & \ddots & \vdots \\ \tilde{S}_{\ell+1} & \cdots & \tilde{S}_{d_\eta} & \cdots & \tilde{S}_{n+1} \end{bmatrix} \tag{3.23}$$

in the case of $d_\eta > n+1$. If $d_\eta = n+1$, then

$$T_{\tilde{S}} = \begin{bmatrix} \tilde{S}_{n+1} & \cdots & \tilde{S}_1 \\ \vdots & \ddots & \vdots \\ \tilde{S}_{d_\eta} & \cdots & \tilde{S}_{n+1} \end{bmatrix}. \tag{3.24}$$

If $d_\eta < n+1$, then

$$T_{\tilde{S}} = \begin{bmatrix} \tilde{S}_{d_\eta} & \cdots & \tilde{S}_1 \\ \vdots & \ddots & \vdots \\ \tilde{S}_{n+1} & \cdots & \tilde{S}_{d_\eta} \\ \vdots & \ddots & \vdots \\ \tilde{S}_{\ell+1} & \cdots & \tilde{S}_{n+1} \end{bmatrix}. \tag{3.25}$$

(3.21) can be written as

$$\begin{bmatrix} -1 & \widehat{\theta}' \end{bmatrix} \begin{bmatrix} \beta \\ A \end{bmatrix} \approx 0,$$

where $\beta$ is the first row of $T_{\tilde{S}}$ and $A$ is the remaining part of $T_{\tilde{S}}$. Then we use the LS solution for the estimated system parameter. That is,

$$\begin{bmatrix} -1 & \widehat{\theta}'_{\text{GSA}} \end{bmatrix} \begin{bmatrix} \beta \\ A \end{bmatrix} = 0 \implies \widehat{\theta}_{\text{GSA}} = (AA')^{-1}A\beta'. \tag{3.26}$$

Prior to presenting the GSA algorithm, we restate the underlying assumption as follows. **Assumption**: Input and output data in (3.8) satisfy the PE condition as specified in (3.12), the measurement noises $\{\varepsilon_{\text{in}}(t), \varepsilon_{\text{out}}(t)\}$ are i.i.d. with zero-mean, and the system transfer function is given by

$$P(z) = \frac{b(z)}{a(z)}$$

with coefficients shown in the parameter vector in (3.7).

The GSA is summarized next.

- Step 1: Choose the block size $\ell \geq n$, form the matrix $\tilde{\Psi}^{(s)}_{\ell,N}$ as in (3.9), and compute

$$\tilde{R}_\ell = \frac{1}{N}\tilde{\Psi}^{(s)}_{\ell,N}\tilde{\Psi}^{(s)'}_{\ell,N}.$$

- Step 2: Compute the EVD for $\tilde{R}_\ell$ with the eigenvalues arranged in ascending order,

and partition $\tilde{K}_s\tilde{\Lambda}_s^{1/2}$ in accordance with (3.21).

- Step 3: Form matrix $T_{\tilde{S}}$ as specified in either (3.23) or (3.24) according to the value of $d_\eta$. Set

$$\beta = T_{\tilde{S}}(1,:), \quad A = T_{\tilde{S}}(2:2(n+1),:),$$

using the Matlab notation. Compute the estimate

$$\widehat{\theta}_{\rm GSA} = (AA')^{-1}A\beta'.$$

The above identification procedure for $\ell > n$ is referred to as GSA because of its connection to the system graph space. It is important to point out that this algorithm reduces to the TLS algorithm, in light of

$$T_{\tilde{S}} = \tilde{S}_1\tilde{\Lambda}_s^{1/2} = \tilde{K}_s\tilde{\Lambda}_s^{1/2}, \quad M_\eta(\widehat{\theta}) = \begin{bmatrix} -1 & \widehat{\theta}' \end{bmatrix},$$

and relation (2.23) in the case of $\ell = n$. In Step 3 of the above GSA, $\widehat{\theta}$ is the LS solution to

$$\begin{bmatrix} -1 & \widehat{\theta}' \end{bmatrix} T_{\tilde{S}} \approx 0. \tag{3.27}$$

We do not suggest the TLS solution in Step 3, because there are little noticeable changes in the simulation studies between using the LS or TLS algorithm to solve for the estimate $\widehat{\theta}$ from (3.27). In fact $\ell > n$ does not need to be a very large number. See the simulation studies section.

**Remark 4.** The results are also applicable to the case when measurement noises are not i.i.d. but having the covariance matrix

$$R_\varepsilon = \mathrm{E}\{\psi_{\varepsilon,\ell}(t)\psi_{\varepsilon,\ell}(t)'\} = \sigma_\varepsilon^2\Omega\Omega' \tag{3.28}$$

where $\Omega$ is a known nonsingular matrix and $\sigma_\varepsilon > 0$ is an unknown scalar. In this case, we can add computing

$$\tilde{R}_{\ell\Omega} = \Omega^{-1}\tilde{R}_\ell\Omega'^{-1}$$

to the end of Step 1 of the GSA. Such a change transforms the noise vector $\psi_{\varepsilon,\ell}(t)$ into

$$\psi^{(\Omega)}_{\varepsilon,\ell}(t) := \Omega^{-1}\psi_{\varepsilon,\ell}(t)$$

that has mean-zero and covariance $\sigma^2_\varepsilon I$. If $\psi_{\varepsilon,\ell}(t)$ has the Gaussian distribution, so does $\psi^{(\Omega)}_{\varepsilon,\ell}(t)$. In Step 2, we can replace $\tilde{R}_\ell$ with $\tilde{R}_{\ell\Omega}$, and $\tilde{K}_s\tilde{\Lambda}_s^{1/2}$ with $\tilde{K}_{s\Omega}\tilde{\Lambda}_{s\Omega}^{1/2}$ to signify the change due to $\Omega$. At the end of Step 2, we add EVD for

$$\Omega\tilde{K}_{s\Omega}\tilde{\Lambda}_{s\Omega}\tilde{K}'_{s\Omega}\Omega' = \tilde{K}_s\tilde{\Lambda}_s\tilde{K}'_s,$$

from which $\tilde{K}_s\tilde{\Lambda}_s^{1/2}$ is obtained. Step 3 of the GSA yields the GSA estimate $\widehat{\theta}$. We comment that components of $\{\psi^{(\Omega)}_{\varepsilon,\ell}(t)\}$ become i.i.d. with mean-zero and variance $\sigma^2_\varepsilon$, which helps to recover the same EIV identification problem as studied. In particular, the asymptotic MLE property in Theorem 2 holds if $\ell = n$. □

### 3.2.3 Iterative Method

The GSA is motivated by the subspace algorithm proposed in [1, 46] that shows improvements of the estimation performance as the block size $\ell$ increases. However, the source signal for blind channel estimation is assumed to be white, whereas $v(t)$ as defined in (3.3) is not. In fact, the non-whiteness of $v(t)$ may degrade the estimation performance as $\ell$ increases in the case of low SNR, as observed in the simulation studies. We thus propose an iterative procedure to search for more accurate parameter estimates for the case of $\ell > n$.

To be specific, consider first the TLS algorithm in the case of $\ell = n$. It has the interpre-

tation of seeking $\widehat{\theta}$ to minimize the cost function

$$J_N(\widehat{\theta}) := \left\| \frac{\begin{bmatrix} -1 & \widehat{\theta}' \end{bmatrix}}{\sqrt{1+\|\widehat{\theta}\|^2}} \tilde{\Psi}_{n,N}^{(s)} \right\|^2 .$$

There holds the inequality

$$\begin{aligned} \mathbb{E}\left\{ \min_{\widehat{\theta}} J_N(\widehat{\theta}) \right\} &\le \mathbb{E}\left\{ \left\| \frac{\begin{bmatrix} -1 & \theta' \end{bmatrix}}{\sqrt{1+\|\theta\|^2}} \tilde{\Psi}_{n,N}^{(s)} \right\|^2 \right\} \\ &= \frac{\begin{bmatrix} -1 & \theta' \end{bmatrix}}{1+\|\theta\|^2} \mathbb{E}\left\{ \tilde{\Psi}_{n,N}^{(s)} \tilde{\Psi}_{n,N}^{(s)'} \right\} \begin{bmatrix} -1 \\ \theta \end{bmatrix} \\ &= \sigma_\varepsilon^2. \end{aligned}$$

Next we consider use of a weighting matrix $W_{\widehat{\theta}}$ such that $M_\eta(\widehat{\theta}) W_{\widehat{\theta}}'$ is an orthogonal matrix to minimize

$$J_N(\widehat{\theta}) = \left\| W_{\widehat{\theta}} M_\eta(\widehat{\theta})' \tilde{K}_s \tilde{\Lambda}_s^{1/2} \right\|^2 .$$

Similarly, it can be seen that

$$\mathbb{E}\left\{ \min_{\widehat{\theta}} J_N(\widehat{\theta}) \right\} \le \mathbb{E}\left\{ \min_{\widehat{\theta}} \left\| W_{\widehat{\theta}} M_\eta(\widehat{\theta})' \tilde{\Psi}_{\ell,N}^{(s)} \right\|^2 \right\} = \sigma_\varepsilon^2.$$

Since minimization of $J_N(\widehat{\theta})$ leads to nonlinear optimization that is difficult to solve, a simple iterative procedure is proposed:

- Step 1: Use the GSA as an initial value of the estimated system parameter vector

$$\widehat{\theta}_{\mathrm{GSA}} =: \widehat{\theta}_0$$

- Step 2: Compute $\widehat{\theta}_i$ to minimize

$$J_N(\widehat{\theta}_i) = \left\| W_{\widehat{\theta}_{i-1}} M_\eta(\widehat{\theta}_i)' \tilde{K}_s \tilde{\Lambda}_s^{1/2} \right\|^2,$$

where $i$ is the $i$th iteration for $i = 1, 2, \cdots$, until $\|\widehat{\theta}_i - \widehat{\theta}_{i-1}\|$ is suitably small.

Minimization of $J_N(\widehat{\theta}_i)$ is a linear problem. Let $w_{\mu,\nu}^{(i)}$ be the $(\mu,\nu)$th entry of $W_{\widehat{\theta}_{i-1}} \in \mathbb{R}^{d_\eta \times d_\eta}$, and

$$\tilde{S}_{W_i}(\mu+1,\nu) = \sum_{\nu=1}^{d_\eta} w_{\mu,\nu}^{(i)} \tilde{S}_{\mu+\nu} \tag{3.29}$$

for $\mu = 0, 1, \cdots, n$ and $\nu = 1, \cdots, d_\eta$. Denote

$$\tilde{S}_{W_i} = \begin{bmatrix} \tilde{S}_{W_i}(1,1) & \cdots & \tilde{S}_{W_i}(1,d_\eta) \\ \vdots & \cdots & \vdots \\ \tilde{S}_{W_i}(n+1,1) & \cdots & \tilde{S}_{W_i}(n+1,d_\eta) \end{bmatrix}. \tag{3.30}$$

Note that $\tilde{S}_{W_i}$ is not a Toeplitz matrix in general. Then minimization of $J_N(\widehat{\theta}_i)$ is equivalent to the minimization of

$$\| \begin{bmatrix} -1 & \widehat{\theta}_i' \end{bmatrix} \tilde{S}_{W_i} \|^2.$$

Again, the LS algorithm can be used to compute $\widehat{\theta}_i$. We summarize Step 2 of the iterative algorithm, which is aimed at minimization of $J_N(\widehat{\theta}_i)$ given $\widehat{\theta}_{i-1}$ for $i \geq 1$ with $\widehat{\theta}_0 = \widehat{\theta}_{\text{GSA}}$ being the GSA estimate:

- Step 2-i: Form matrix $M_\eta(\widehat{\theta}_{i-1})$, and compute

$$W_{\widehat{\theta}_{i-1}} = [M_\eta(\widehat{\theta}_{i-1})' M_\eta(\widehat{\theta}_{i-1})]^{-1/2}.$$

- Step 2-ii: Form matrix $\tilde{S}_{W_i}$ according to (3.30) with each sub-matrix specified in (3.29).

- Step 2-iii: Set

$$A_W = \tilde{S}_{W_i}(2:2(n+1),:), \quad \beta_W = \tilde{S}_{W_i}(1,:),$$

and compute

$$\widehat{\theta}_i = (A_W A_W')^{-1} A_W \beta_W'.$$

- Step 2-iv: If

$$\|\widehat{\theta}_i - \widehat{\theta}_{i-1}\| < \epsilon$$

for some pre-specified tolerance $\epsilon > 0$, then stop and choose $\widehat{\theta}_i$ as the output. Otherwise, set

$$i = i + 1, \quad \widehat{\theta}_{i-1} = \widehat{\theta}_i,$$

and go to Step 2-i.

In Step 2-i, $W_{\widehat{\theta}_{i-1}}$ can be chosen as a symmetric and positive definite matrix for each $i \geq 1$.

**Remark 5.** The GSA proposed in this dissertation differs from the subspace algorithm in [1, 46] in that the signal subspace power represented by $\tilde{\Lambda}_s^{1/2}$ is included in $T_{\tilde{S}}$ for computing the parameter estimate $\widehat{\theta}$ via solving

$$\begin{bmatrix} -1 & \widehat{\theta}' \end{bmatrix} T_{\tilde{S}} \approx 0.$$

Hence, in the case of high SNR, the identification performance for the GSA can be improved more than that of the subspace algorithm. In the case of low SNR, the iterative procedure developed in this section can be employed to improve the identification performance. The simulation studies in the next section show that only a few iterations are needed. For the above reason, the GSA as proposed is not only more general than the TLS algorithm but also has more advantages than that of the subspace algorithm developed in [1, 46] for blind channel estimation. □

### 3.3 Implementation

In this section, we consider the implementation issue. While the GSA can be programmed directly, the use of the graph signal $\{s(t)\}$ may complicate the identification problem when

the two noise variances are not equal. The basic idea is to group input and output data separately so that the input and output noise variances will appear separately in the second statistics. Such a way helps to estimate the noise variances when the two of them are non-equal and unknown, which will be studied in the next chapter.

### 3.3.1 Modification

We define modified system parameter vector:

$$\bar{\theta} = \begin{bmatrix} -\theta_a \\ \theta_b \end{bmatrix}, \quad \theta_a = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}, \quad \theta_b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}, \quad a_0 = 1.$$

For the noisy measurements, we form the data vector as follows:

$$\psi_\ell(t) = \begin{bmatrix} \psi_{y_\ell}(t) \\ \psi_{u_\ell}(t) \end{bmatrix}, \quad \psi_{y_\ell}(t) = \begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(t-\ell) \end{bmatrix}, \quad \psi_{u_\ell}(t) = \begin{bmatrix} u(t) \\ u(t-1) \\ \vdots \\ u(t-\ell) \end{bmatrix}.$$

Define the data matrix as

$$\tilde{\Psi}_\ell = \begin{bmatrix} \tilde{\psi}_\ell(t_0) & \tilde{\psi}_\ell(t_0+1) & \cdots & \tilde{\psi}_\ell(t_0+N-1) \end{bmatrix}$$

with $t_{0,\ell} = t_0 + \ell$, $t_f = t_{0,\ell} + N - 1$. Similarly, we modify equation (3.9) to the following:

$$\tilde{\Psi}_\ell = \Psi_\ell + \Psi_{\varepsilon_\ell},$$

specified by,

$$\begin{aligned}
\tilde{\Psi}_\ell &= \left[\begin{array}{cccc} \tilde{\psi}_\ell(t_{0,\ell}) & \tilde{\psi}_\ell(t_{0,\ell}+1) & \cdots & \tilde{\psi}_\ell(t_f) \end{array}\right] \\
&= \left[\begin{array}{cccc} \psi_\ell(t_{0,\ell}) & \psi_\ell(t_{0,\ell}+1) & \cdots & \psi_{s_\ell}(t_f) \end{array}\right] + \Psi_{\varepsilon_\ell}, \\
\Psi_{\varepsilon_\ell} &= \left[\begin{array}{cccc} \psi_{\varepsilon_\ell}(t_{0,\ell}) & \psi_{\varepsilon_\ell}(t_{0,\ell}+1) & \cdots & \psi_{\varepsilon_\ell}(t_f) \end{array}\right],
\end{aligned}$$

where $\Psi_{\varepsilon_\ell}$ consists of input-output measurement noises. It follows that $\bar{\theta}'\psi_\ell(t) = 0\ \forall t \geq 0$, if $\ell = n$. In the case of $\ell > n$, we pad zeros to form

$$\theta_{a_\ell} = \left[\begin{array}{c} \theta_a \\ 0_{\ell-n} \end{array}\right] \in \mathbb{R}^{\ell+1}, \quad \theta_{b_\ell} = \left[\begin{array}{c} \theta_b \\ 0_{\ell-n} \end{array}\right] \in \mathbb{R}^{\ell+1},$$

where $0_k$ is a $k$-dimensional vector of zeros. Define the matrix

$$\begin{aligned}
M_{\ell,\eta}(\theta_{a_\ell}, \theta_{b_\ell}) &= \left[\begin{array}{cccccccc} -\theta'_{a_\ell} & 0 & \cdots & 0 & \theta'_{b_\ell} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\theta'_{a_\ell} & 0 & \cdots & 0 & \theta'_{b_\ell} \end{array}\right]', \\
&= \left[\begin{array}{cccccccccccc} -a_0 & \cdots & -a_n & 0 & \cdots & 0 & b_0 & \cdots & b_n & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -a_0 & \cdots & a_n & 0 & \cdots & 0 & b_0 & \cdots & b_n \end{array}\right]' \in \mathbb{R}^{2(\ell+1)\times d_\eta},
\end{aligned}$$

where $d_\eta$ is the the dimension of noise subspace of the auto-covariance matrix of the input and output measurement data:

$$\tilde{R}_{\mathrm{m}\ell} = \frac{1}{N}\tilde{\Psi}_\ell\tilde{\Psi}'_\ell.$$

Note that the subscript "m" indicates the modified forms for the measurement data to distinguish it from the one which involved the unmodified forms. Accordingly, we need to

replace (3.21) with the following

$$\tilde{K}_{\mathrm{m}\ell}\tilde{\Lambda}_{\mathrm{m}\ell}^{1/2} = \begin{bmatrix} \tilde{S}_{y,1} \\ \vdots \\ \tilde{S}_{y,\ell+1} \\ \tilde{S}_{u,1} \\ \vdots \\ \tilde{S}_{u,\ell+1} \end{bmatrix}, \ \tilde{S}_{y,i} \in \mathbb{R}^{1\times d_s}, \ \tilde{S}_{u,i} \in \mathbb{R}^{1\times d_s}, \tag{3.31}$$

where $\widehat{K}_{\mathrm{m}s}$ and $\widehat{\Lambda}_{\mathrm{m}s}$ are the eigenvectors and eigenvalues of the signal space of $\tilde{R}_{\mathrm{m}\ell}$ respectively. In light of

$$\tilde{R}_{\mathrm{m}\ell} = \tilde{K}_{\mathrm{m}s}\tilde{\Lambda}_{\mathrm{m}s}\tilde{K}'_{\mathrm{m}s},$$

we can see that

$$M_{\ell,\eta}(\widehat{\theta}_{a\ell}, \widehat{\theta}_{b\ell})\tilde{K}_{\mathrm{m}s}\tilde{\Lambda}_{\mathrm{m}s}^{1/2} \approx 0.$$

The above is equivalent to

$$\widehat{\tilde{\theta}}' T_{\mathrm{m}\tilde{S}} \approx 0, \tag{3.32}$$

where $T_{\mathrm{m}\tilde{S}}$ is given by

$$T_{\mathrm{m}\tilde{S}} = \begin{bmatrix} \tilde{S}_{y,d_\eta} & \cdots & \tilde{S}_{y,n+1} & \cdots & \tilde{S}_{y,1} \\ \vdots & \ddots & & \ddots & \vdots \\ \tilde{S}_{y,\ell+1} & \cdots & \tilde{S}_{y,d_\eta} & \cdots & \tilde{S}_{y,n+1} \\ \tilde{S}_{u,d_\eta} & \cdots & \tilde{S}_{u,n+1} & \cdots & \tilde{S}_{u,1} \\ \vdots & \ddots & & \ddots & \vdots \\ \tilde{S}_{u,\ell+1} & \cdots & \tilde{S}_{u,d_\eta} & \cdots & \tilde{S}_{u,n+1} \end{bmatrix} \tag{3.33}$$

in the case of $d_\eta > n+1$. If $d_\eta = n+1$, then

$$T_{\mathrm{m}\tilde{S}} = \begin{bmatrix} \tilde{S}_{y,n+1} & \cdots & \tilde{S}_{y,1} \\ \vdots & \ddots & \vdots \\ \tilde{S}_{y,d_\eta} & \cdots & \tilde{S}_{y,n+1} \\ \tilde{S}_{u,n+1} & \cdots & \tilde{S}_{u,1} \\ \vdots & \ddots & \vdots \\ \tilde{S}_{u,d_\eta} & \cdots & \tilde{S}_{u,n+1} \end{bmatrix}. \tag{3.34}$$

If $d_\eta < n+1$, then

$$T_{\mathrm{m}\tilde{S}} = \begin{bmatrix} \tilde{S}_{y,d_\eta} & \cdots & \tilde{S}_{y,1} \\ \vdots & \ddots & \vdots \\ \tilde{S}_{y,n+1} & \cdots & \tilde{S}_{y,d_\eta} \\ \vdots & \ddots & \vdots \\ \tilde{S}_{y,\ell+1} & \cdots & \tilde{S}_{y,n+1} \\ \tilde{S}_{u,d_\eta} & \cdots & \tilde{S}_{u,1} \\ \vdots & \ddots & \vdots \\ \tilde{S}_{u,n+1} & \cdots & \tilde{S}_{u,d_\eta} \\ \vdots & \ddots & \vdots \\ \tilde{S}_{u,\ell+1} & \cdots & \tilde{S}_{u,n+1} \end{bmatrix}. \tag{3.35}$$

We can then solve (3.32) using LS and find the estimated system parameters. It will give us the same values as those described in Section 3.2.

In addition, we also modify (3.29) for the iterative procedure as follows:

$$\tilde{S}_{W_y,i}(\mu+1,\nu) = \sum_{\nu=1}^{d_\eta} w_{\mu,\nu}^{(i)} \tilde{S}_{y,\mu+\nu}, \tag{3.36}$$

$$\tilde{S}_{W_u,i}(\mu+1,\nu) = \sum_{\nu=1}^{d_\eta} w_{\mu,\nu}^{(i)} \tilde{S}_{u,\mu+\nu}, \tag{3.37}$$

for $\mu = 0, 1, \cdots, n$ and $\nu = 1, \cdots, d_\eta$. Also (3.30) needs to be replaced with

$$\tilde{S}_{\mathrm{m}W_i} = \begin{bmatrix} \tilde{S}_{W_y,i}(1,1) & \cdots & \tilde{S}_{W_y,i}(1,d_\eta) \\ \vdots & \cdots & \vdots \\ \tilde{S}_{W_y,i}(n+1,1) & \cdots & \tilde{S}_{W_y,i}(n+1,d_\eta) \\ \tilde{S}_{W_u,i}(1,1) & \cdots & \tilde{S}_{W_u,i}(1,d_\eta) \\ \vdots & \cdots & \vdots \\ \tilde{S}_{W_u,i}(n+1,1) & \cdots & \tilde{S}_{W_u,i}(n+1,d_\eta) \end{bmatrix}. \tag{3.38}$$

Using the iterative procedure in Section 3.2.3 with the modified matrices and vectors, we can obtain the estimated system parameters.

### 3.3.2 Use of the Ratio of Noise Variances

Under the assumptions of the unequal noise, there holds

$$\lim_{N\to\infty} \frac{1}{N} \tilde{\Psi}_\ell' \tilde{\Psi}_\ell = \lim_{N\to\infty} \frac{1}{N} \Psi_\ell' \Psi_\ell + \begin{bmatrix} \sigma_y^2 I_{\ell+1} & 0 \\ 0 & \sigma_u^2 I_{\ell+1} \end{bmatrix}.$$

Define the root-squared ratio of the input and output noise variances as

$$\rho = \sqrt{\frac{\sigma_u^2}{\sigma_y^2}} = \frac{\sigma_u}{\sigma_y}.$$

If the output measurements are multiplied by $\rho$, we can form

$$\tilde{\psi}_\ell^{(\rho)}(t) = \begin{bmatrix} \rho\tilde{\psi}_{y_\ell}(t) \\ \tilde{\psi}_{u_\ell}(t) \end{bmatrix},$$

and

$$\tilde{\Psi}_\ell^{(\rho)} = \begin{bmatrix} \tilde{\psi}_\ell^{(\rho)}(t_{0,\ell}) & \tilde{\psi}_\ell^{(\rho)}(t_{0,\ell}+1) & \cdots & \tilde{\psi}_\ell^{(\rho)}(t_f) \end{bmatrix}.$$

If we know the ratio $\rho$, then asymptotically,

$$\frac{1}{N}\tilde{\Psi}_{\ell}^{(\rho)}\tilde{\Psi}_{\ell}^{(\rho)'} \;\rightarrow\; \frac{1}{N}\Psi_{\ell}^{(\rho)}\Psi_{\ell}^{(\rho)'} + \sigma_u^2 I_{2(\ell+1)} \tag{3.39}$$

as $N \rightarrow \infty$. This is equivalent to the case when the input and output noise variances are equal. Hence, the GSA can be applied to compute the estimate for $\theta_a$ and $\theta_b$. Caution needs to be taken on the scaling of $\rho$. That is, the GSA produces the estimate for $\theta_a$ and $\rho\theta_b$.

### 3.4 Simulation Studies

In this section, simulation studies on two different input cases will be presented, the white input case and the colored input case. Each case has two different plant models. To evaluate the performance of the proposed GSA, the root-mean-square error (RMSE), defined by

$$\text{RMSE} := \sqrt{\frac{1}{T}\sum_{i=1}^{T}\|\widehat{\theta}_i - \theta\|^2} \tag{3.40}$$

is used where $T$ is the number of ensembles and $\widehat{\theta}_i$ is the estimated parameter vector from the $i$th simulation. For the iterative procedure, we set the tolerance $\epsilon = 10^{-8}$ which is

$$\|\widehat{\theta}_i - \widehat{\theta}_{i-1}\| < 10^{-8}.$$

A total of 2,000 ensembles are used in our simulation studies. Note that the character for the block size $\ell$ is replaced with L in all figures. The solid line indicates the Cramér-Rao lower bound (CRLB).

**White Input**

Two plant models from [69] are used to illustrate the performance of the GSA algorithm in the case when the inputs $\{u(t)\}$ are white signals.

**Example 3.1** : The plant model is described by

$$P(z) = \frac{0.8 \times (1 - 1.2488z^{-1} + 0.9604z^{-2})}{1 - 1.4491z^{-1} + 0.9604z^{-2}}.$$

The SNR at the input and output is 3.0126dB and 5.9155dB, respectively, with $\sigma_\varepsilon = 0.5$. Figure 3.1 and Figure 3.2 show the simulation results for Example 3.1 with the white input $u(t)$.

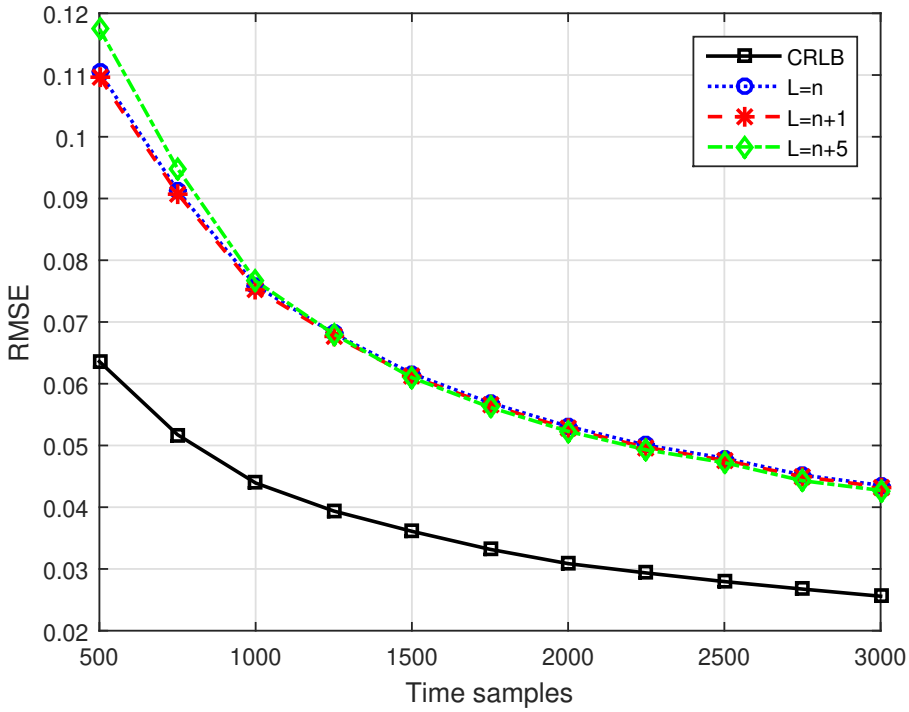


Figure 3.1. RMSEs for GSA in Example 3.1 without iteration

The RMSE curves for different L ($\ell$) values in Figure 3.1 show that the proposed GSA does not improve the estimation performance as L ($\ell$) increases, contrasting to the subspace algorithm for the blind channel estimation. The reason lies in that the corresponding $v(t)$ is not white for system identification based on the EIV model, whereas for blind channel estimation its source signal $v(t)$ is white. In addition, the SNR is relatively low. On the other hand, Figure 3.2 shows that the GSA with the iterative procedure improves the RMSE performance as L ($\ell$) increases. Table 3.1 shows the average number of iterations for the

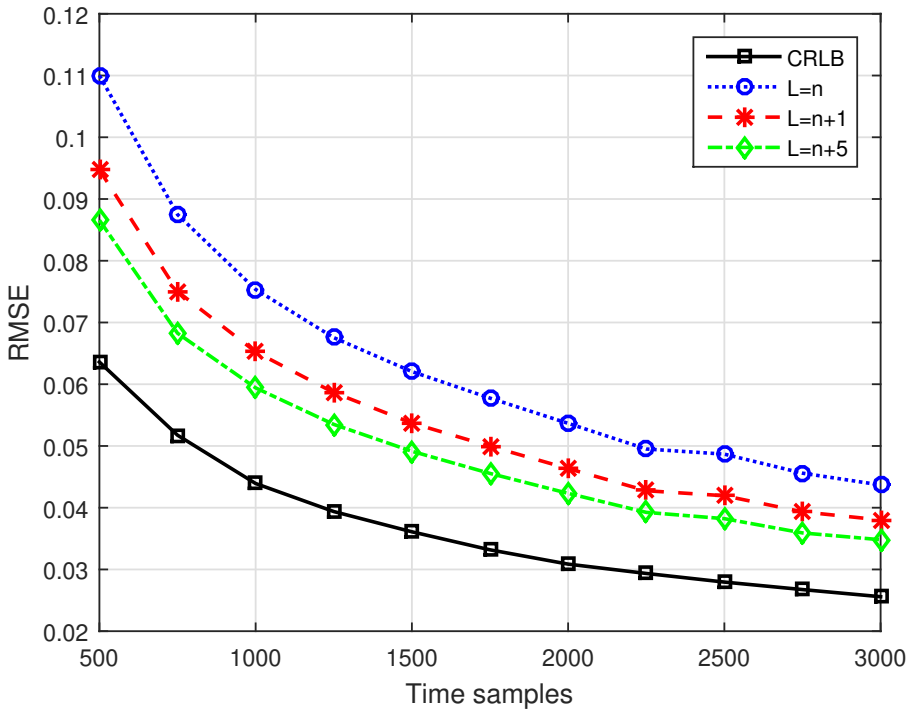


Figure 3.2. RMSEs for GSA in Example 3.1 with iteration

RMSE curves in Figure 3.2. In the case of $\ell = n$, we only need one instance of the iteration to reach the tolerance. As $\ell$ increases, more iterations are needed but not many, approximately three or four, as shown in the table.

Table 3.1. Average number of iterations in Example 3.1

| $N$ | 500 | 750 | 1000 | 1250 | 1500 | 1750 | 2000 | 2250 | 2500 | 2750 | 3000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\ell = n$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $\ell = n + 1$ | 4.38 | 4.23 | 4.11 | 4.02 | 3.95 | 3.91 | 3.87 | 3.84 | 3.78 | 3.79 | 3.75 |
| $\ell = n + 5$ | 5.09 | 4.87 | 4.70 | 4.62 | 4.52 | 4.45 | 4.40 | 4.35 | 4.33 | 4.32 | 4.26 |

**Example 3.2**: The plant model is described by

$$P(z) = \frac{1.5 - 0.9z^{-1} - 0.45z^{-2}}{1 - 0.5z^{-1} + 0.3z^{-2}}.$$

The SNR at the input and output is 2.9927dB and 8.4103dB, respectively, and $\sigma_\varepsilon = 0.5$.

The input $u(t)$ is again white. Figure 3.3 shows the RMSE in Example 3.2 without iteration. Figure 3.4 shows the simulation results in Example 3.2 with the iterative procedure.

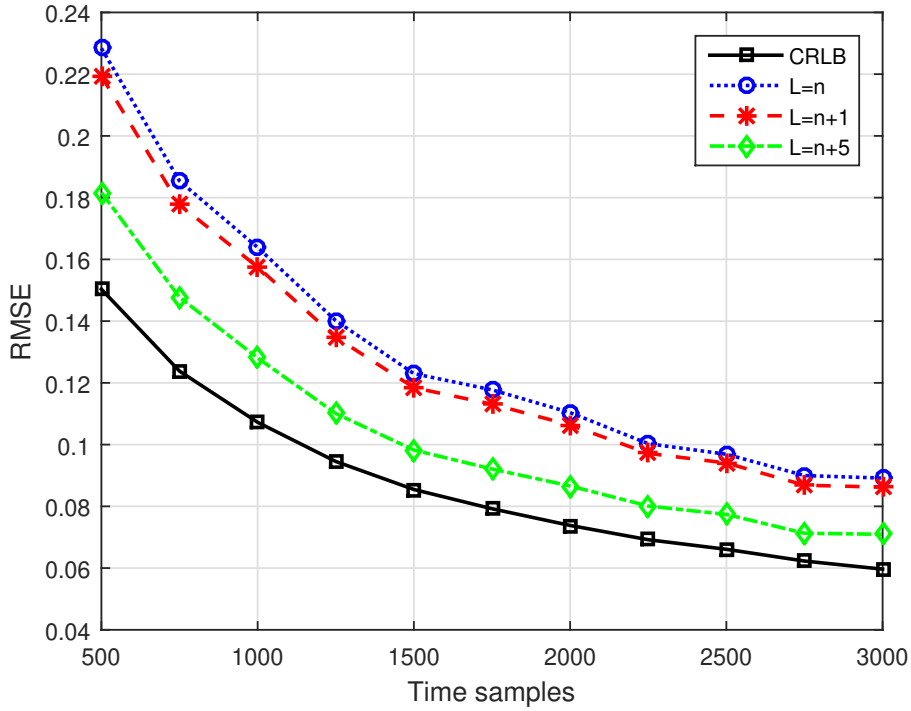


Figure 3.3. RMSEs for GSA in Example 3.2 without iteration

Figure 3.3 shows that the RMSE decreases as L ($\ell$) increases even though the iterative procedure is not used. Its reason lies in the greater SNR at the plant output than that in Example 1, which helps to improve the identification error without iteration. Figure 3.4 shows that the iterative procedure helps to improve the RMSE performance further, compared to those in Figure 3.3 especially when L $= n + 1$, i.e., $\ell = n + 1$. Table 3.2 shows the average number of iterations.

Table 3.2. Average number of iterations in Example 2

| $N$ | 500 | 750 | 1000 | 1250 | 1500 | 1750 | 2000 | 2250 | 2500 | 2750 | 3000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\ell = n$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $\ell = n + 1$ | 2.92 | 5.65 | 5.46 | 5.28 | 5.21 | 5.16 | 5.02 | 4.94 | 4.89 | 4.87 | 4.78 |
| $\ell = n + 5$ | 6.79 | 6.36 | 6.03 | 5.86 | 5.72 | 5.62 | 5.47 | 5.43 | 5.35 | 5.28 | 5.21 |

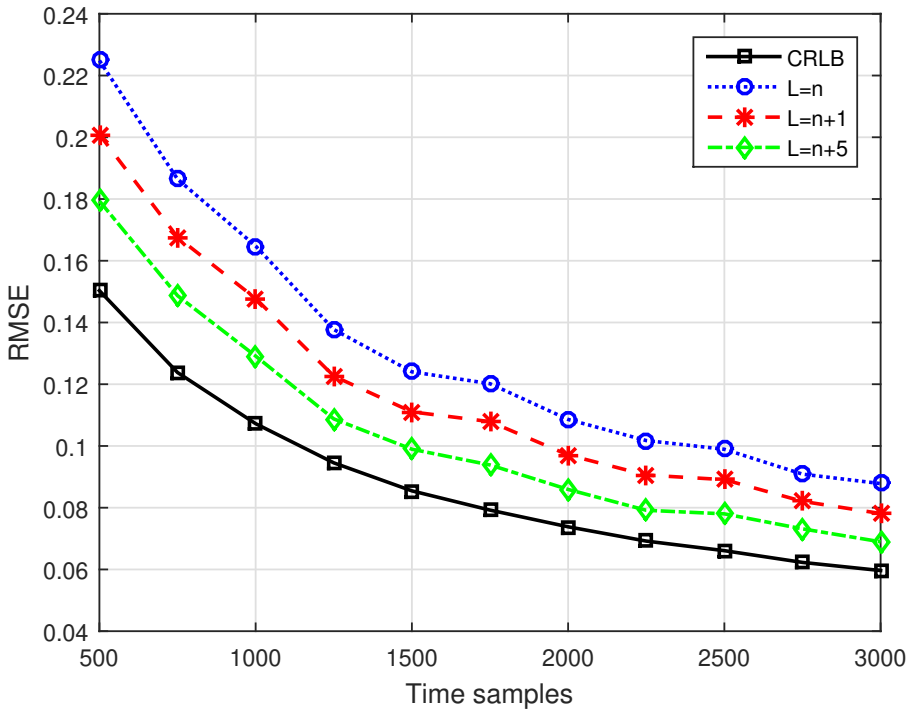


Figure 3.4. RMSEs for GSA in Example 2 with iteration

**Colored Input**

Two plant models modified from the literature [60, 69] are used to illustrate the performance of GSA algorithm with the colored input $u(t)$ signals.

**Example 3.3**: The plant model is from [60], given by

$$P(z) = \frac{2.6466z^{-1} + 1.3233z^{-2}}{1 - 1.5z^{-1} + 0.7z^{-2}}.$$

The input signal $u(t)$ is the filtered signal, described by

$$u(t) = \frac{1 - 1.684q^{-1} + 0.7812q^{-2}}{1 - 0.275q^{-1} - 0.2103q^{-2}}\ u_{\mathrm{n}}(t),$$

driven by the white Gaussian process $\{u_{\mathrm{n}}(t)\}$ with zero mean and variance one. The noise variance for $\varepsilon_{\mathrm{in}}(t)$ is 0.3, and $\varepsilon_{\mathrm{out}}(t)$ is 1. Assuming that the ratio of the two variances is known, we can multiply the output measurements by $\sqrt{0.3}$ to convert the EIV identification

problem into that of the equal variance 0.3, keeping in mind that the estimated numerator coefficients need to be scaled back by a factor of $1/\sqrt{0.3}$ before evaluating the estimation error. As a result, the SNR at the input and output is 10.56dB and 10.5dB, respectively.

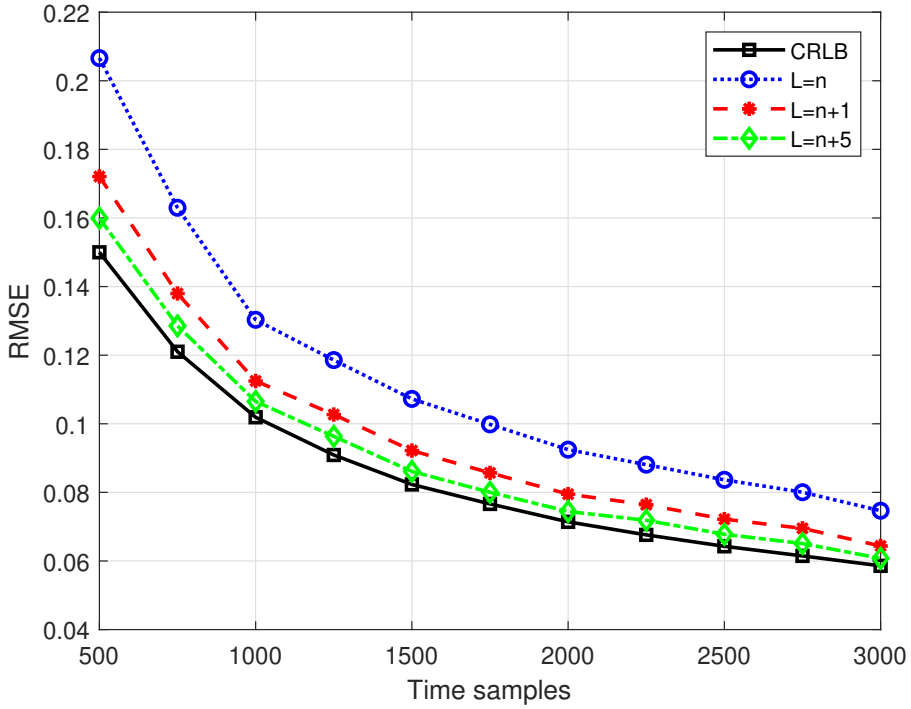


Figure 3.5. RMSEs for GSA in Example 3.3 without iteration

The RMSE curves for different L ($\ell$) values in Figure 3.5 show that the proposed GSA improves the estimation performance as L ($\ell$) increases, owing to the high SNR. On the other hand, Figure 3.6 shows that the GSA with the iterative procedure does not help to improve the RMSE, compared with Figure 3.5. Table 3.3 shows the average number of iterations.

Table 3.3. Average number of iterations in Example 3.3

| $N$ | 500 | 750 | 1000 | 1250 | 1500 | 1750 | 2000 | 2250 | 2500 | 2750 | 3000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\ell = n$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $\ell = n+1$ | 4.28 | 4.02 | 3.82 | 3.71 | 3.64 | 3.59 | 3.52 | 3.48 | 3.42 | 3.39 | 3.34 |
| $\ell = n+5$ | 5.65 | 5.31 | 5.09 | 4.96 | 4.85 | 4.79 | 4.72 | 4.66 | 4.69 | 4.54 | 4.49 |

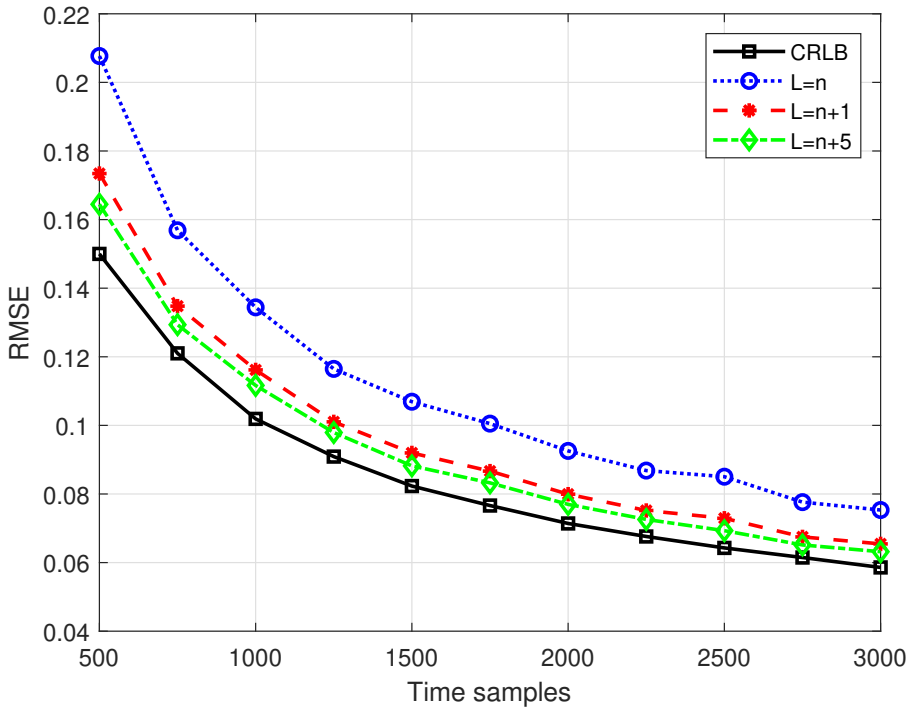


Figure 3.6. RMSEs for GSA in Example 3.3 with iteration

**Example 3.4**: The plant model is described by [69]

$$P(z) = \frac{0.8 \times (1 - 1.2488z^{-1} + 0.9604z^{-2})}{1 - 1.4491z^{-1} + 0.9604z^{-2}}.$$

Similar to Example 3.3, the input $u(t)$ is colored, generated by

$$u(t) = \frac{1 - 1.9837q^{-1} + 0.5852q^{-2}}{1 - 0.1975q^{-1} - 0.238q^{-2}}\ u_{\mathrm{n}}(t),$$

with $u_{\mathrm{n}}(t)$ white Gaussian of zero mean and variance 1. Figure 3.7 and Figure 3.8 show the simulation results for Example 3.4. The equal noise variances 1 are assumed. The SNRs at the plant input and output are 6.51dB and 5.09dB, respectively, which are considerably smaller compared with those in Example 3. This is probably the reason why the RMSE performance in Figure 3.7 does not show improvement as L ($\ell$) increases. However, the iterative procedure helps to reduce the RMSE as L ($\ell$) increases, as shown in Figure 3.8.

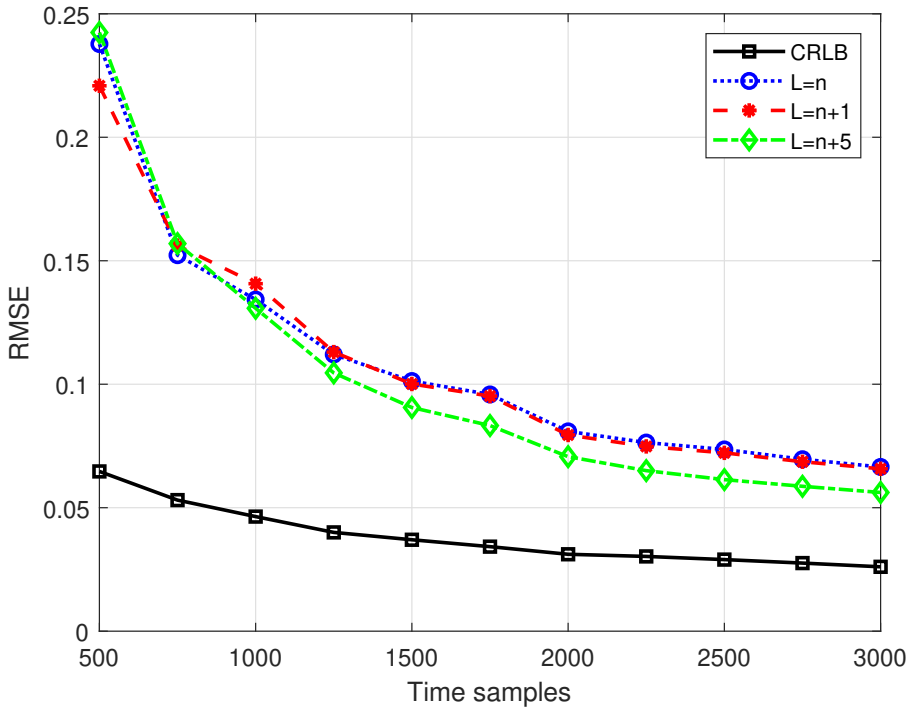


Figure 3.7. RMSEs for GSA in Example 3.4 without iteration

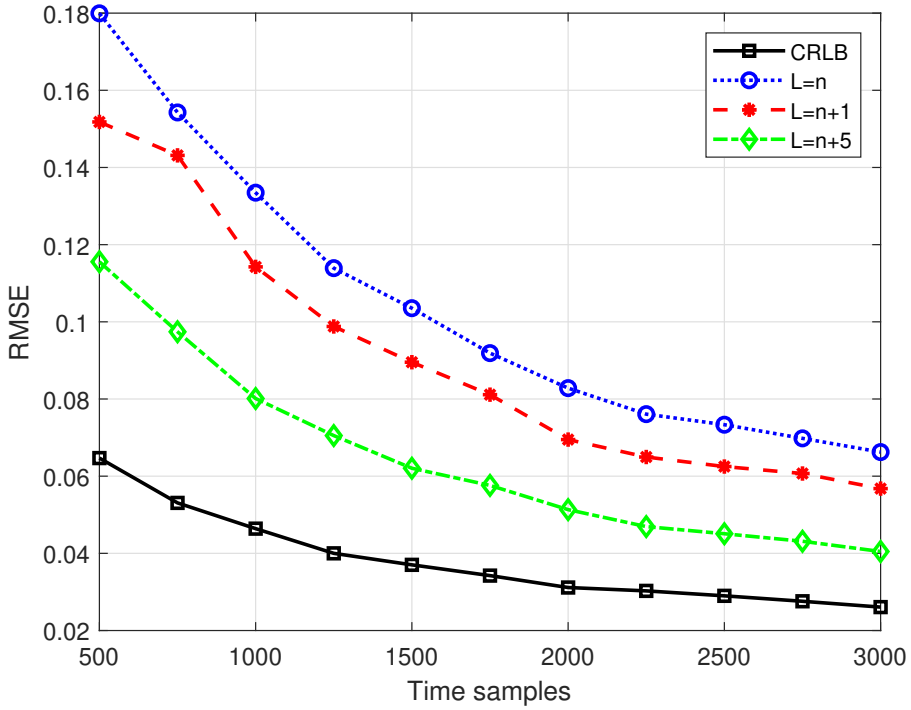


Figure 3.8. RMSEs for GSA in Example 3.4 with iteration

The SNR seems to be a crucial factor in determining the GSA's performance, highlighted in Remark 5. Table 3.4 shows the average number of iterations.

Table 3.4. Average number of iterations in Example 3.4

| $N$ | 500 | 750 | 1000 | 1250 | 1500 | 1750 | 2000 | 2250 | 2500 | 2750 | 3000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\ell = n$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $\ell = n + 1$ | 4.37 | 1.20 | 4.13 | 4.01 | 3.98 | 3.93 | 3.84 | 3.83 | 3.81 | 3.78 | 3.77 |
| $\ell = n + 5$ | 5.41 | 5.12 | 4.99 | 4.84 | 4.73 | 4.65 | 4.54 | 4.49 | 4.43 | 4.40 | 4.35 |

# CHAPTER 4
# IDENTIFICATION WITH UNEQUAL NOISE VARIANCES

In this chapter, we study the EIV system identification in the case where the noise variances $\sigma_u^2$ and $\sigma_y^2$ are unknown and unequal. In practice, often outputs of the system are more difficult to measure than the inputs. Therefore, this is a practical problem in EIV system identification. Two different methods will be studied. The first relies on the Frisch scheme, which estimates the two different noise variances, and then estimates the system parameters. In the literature [57], Frisch scheme in combination with the TLS algorithm has been investigated. We introduce the Frisch scheme in Section 4.1. The use of the GSA algorithm in combination with the Frisch scheme (GSA-Frisch) is studied in Section 4.2, which has better performance, compared with the TLS algorithm. Then we propose a new algorithm in Section 4.3 without estimating the two noise variances. In addition, the state-space formulation is presented for the N4SID algorithm, which is available in Matlab as a built-in function so that we can compare the performance of the new algorithm with that of the existing one in the simulation section. Finally, the results of this chapter are illustrated by the simulation results.

## 4.1 Frisch Scheme

In this section, we first provide an overview for the Frisch scheme in the EIV system identification based on the literatures [58], [7], [8], [57], focusing on how to estimate noise variances. We will also show some simulation examples to describe the algorithm.

### 4.1.1 Concept of Frisch Scheme

Let us consider the simple algebraic case first before extending it to the EIV system model. The equation is given by

$$\Psi_{n,N}^{(x)'}\alpha = 0, \tag{4.1}$$

where $\Psi_{n,N}^{(x)}$ is a Toeplitz matrix, and $\alpha$ is a column vector, specified by

$$\Psi_{n,N}^{(x)} = \begin{bmatrix} x_n & x_{n+1} & \cdots & \cdots & \cdots & x_{N+n-1} \\ x_{n-1} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ x_1 & \cdots & \cdots & x_n & \cdots & x_N \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \quad N > n.$$

Define $R_x$ as a covariance matrix

$$R_x = \lim_{N\to\infty} \frac{1}{N} \Psi_{n,N}^{(x)} \Psi_{n,N}^{(x)\prime}. \tag{4.2}$$

By (4.1), it is clearly seen that

$$R_x \alpha = 0. \tag{4.3}$$

The relation (4.3) plays an important role in the Frisch scheme. Suppose that noises are added into each element of $\Psi_{n,N}^{(x)}$. That is,

$$\tilde{\Psi}_{n,N}^{(x)} = \Psi_{n,N}^{(x)} + \Psi_{n,N}^{(e)}, \tag{4.4}$$

where $\Psi_{n,N}^{(e)}$ is the noise matrix that has a dimension of $(n \times N)$

$$\Psi_{n,N}^{(e)} = \begin{bmatrix} e_n & e_{n+1} & \cdots & \cdots & \cdots & e_{N+n-1} \\ e_{n-1} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ e_1 & \cdots & \cdots & e_n & \cdots & e_N \end{bmatrix}.$$

Define covariance matrices of $\tilde{\Psi}_{n,N}^{(x)}$ and $\Psi_{n,N}^{(e)}$ as follows:

$$\tilde{R_x} = \lim_{N\to\infty} \frac{1}{N} \tilde{\Psi}_{n,N}^{(x)} \tilde{\Psi}_{n,N}^{(x)\prime}, \quad R_\varepsilon = \lim_{N\to\infty} \frac{1}{N} \Psi_{n,N}^{(e)} \Psi_{n,N}^{(e)\prime}. \tag{4.5}$$

We assume that the given noise elements are Gaussian distributed with zero mean. The Frisch scheme focuses on finding all possible diagonal matrix $R_\varepsilon$ such that

$$R_x = \tilde{R}_x - R_\varepsilon \geq 0. \tag{4.6}$$

Define $\sigma_i^2$ as the variance of the $i$th column of ${\Psi_{n,N}^{(e)}}'$. If we consider the case when $R_\varepsilon = \text{diag}\left[\begin{array}{cccc} 0, & \cdots, & 0, & \widehat{\sigma}_n^2 \end{array}\right]$, (4.6) is written as

$$R_x = \tilde{R}_x - \text{diag}\left[\begin{array}{cccc} 0, & \cdots, & 0, & \widehat{\sigma}_n^2 \end{array}\right]. \tag{4.7}$$

Then the equation is rewritten as

$$R_x = \left[\begin{array}{cc} R_n & \tau_n \\ \tau_n' & \sigma_n^2 - \widehat{\sigma}_n^2 \end{array}\right], \tag{4.8}$$

where $R_n$ is defined as a matrix eliminating the $n$th row and $n$th column of $\tilde{R}_x$. We compute the determinant of the both sides of the equation (4.8) and by the property of a block matrix

$$\det(R_x) = \det(R_n)[\sigma_n^2 - \widehat{\sigma}_n^2 - \tau_n' R_n^{-1} \tau_n]. \tag{4.9}$$

Then we can get the maximum noise variance using the relation (4.6):

$$\widehat{\sigma}_{\max,n}^2 = \det[\tilde{R}_x]/\det[R_n]. \tag{4.10}$$

To obtain the above relation, we have used the fact that $\tilde{R}_x$ is singular, and thus $\sigma_n^2 = \tau_n' R_n^{-1} \tau_n$. Using the same procedure for the different cases of the diagonal matrix, as a result, we have a general expression

$$\widehat{\sigma}_{\max,i} = \det[\tilde{R}_x]/\det[R_i], \tag{4.11}$$

where $i = 1, \cdots, n$, $R_i$ is as a matrix eliminating the $i$th row and $i$th column of $\tilde{R}_x$. Thus we get the range of the estimated noise variance in the $i$ column of ${\Psi^{(e)}_{n,N}}'$ as follows

$$0 \leq \widehat{\sigma}_i^2 \leq \widehat{\sigma}_{\max,i}^2. \tag{4.12}$$

### 4.1.2 Frisch Scheme with TLS

In this section, we introduce the extended Frisch scheme for the EIV system model developed by [7] and define the Toeplitz matrices using the data for EIV model using the relation (3.8) and notation (3.9) as follows:

$$\tilde{\Psi}^{(u)}_{n,N} = \Psi^{(u)}_{n,N} + \Psi^{(\varepsilon_{\text{in}})}_{n,N},$$

$$\tilde{\Psi}^{(y)}_{n,N} = \Psi^{(y)}_{n,N} + \Psi^{(\varepsilon_{\text{out}})}_{n,N}.$$

Let $\tilde{\Psi} = \begin{bmatrix} \tilde{\Psi}^{(y)'}_{n,N} & \tilde{\Psi}^{(u)'}_{n,N} \end{bmatrix}$. Then its covariance matrix is

$$\tilde{R} = \frac{1}{N}\tilde{\Psi}'\tilde{\Psi}. \tag{4.13}$$

Then $\tilde{R}$ can be partitioned as follows:

$$\tilde{R} = \begin{bmatrix} \tilde{R}_{yy} & \tilde{R}_{yu} \\ \tilde{R}_{uy} & \tilde{R}_{uu} \end{bmatrix}, \tag{4.14}$$

where

$$\tilde{R}_{yy} = \frac{1}{N}\tilde{\Psi}^{(y)}_{n,N}\tilde{\Psi}^{(y)'}_{n,N}, \quad \tilde{R}_{yu} = \frac{1}{N}\tilde{\Psi}^{(y)}_{n,N}\tilde{\Psi}^{(u)'}_{n,N}, \quad \tilde{R}_{uy} = \frac{1}{N}\tilde{\Psi}^{(u)}_{n,N}\tilde{\Psi}^{(y)'}_{n,N}, \quad \tilde{R}_{uu} = \frac{1}{N}\tilde{\Psi}^{(u)}_{n,N}\tilde{\Psi}^{(u)'}_{n,N}.$$

The next relation holds by the assumption that the noise data are zero-mean, uncorrelated with each other, and $N$ approaches infinity.

$$\tilde{R} = R + R_\varepsilon, \quad R_\varepsilon = \begin{bmatrix} \sigma_y^2 I_{n+1} & 0 \\ 0 & \sigma_u^2 I_{n+1} \end{bmatrix}, \tag{4.15}$$

where

$$R = \frac{1}{N} \begin{bmatrix} \Psi_{n,N}^{(y)'} & \Psi_{n,N}^{(u)'} \end{bmatrix}' \begin{bmatrix} \Psi_{n,N}^{(y)'} & \Psi_{n,N}^{(u)'} \end{bmatrix},$$
$$R_\varepsilon = \frac{1}{N} \begin{bmatrix} \Psi_{n,N}^{(\varepsilon_{\text{out}})'} & \Psi_{n,N}^{(\varepsilon_{\text{in}})'} \end{bmatrix}' \begin{bmatrix} \Psi_{n,N}^{(\varepsilon_{\text{out}})'} & \Psi_{n,N}^{(\varepsilon_{\text{in}})'} \end{bmatrix},$$

and $\sigma_u^2$ and $\sigma_y^2$ are the noise variance of input and output, respectively. Similar to the algebraic case in the previous section, we focus on searching all possible $R_\varepsilon$ such that

$$R = \tilde{R} - \begin{bmatrix} \sigma_y^2 I_{n+1} & 0 \\ 0 & \sigma_u^2 I_{n+1} \end{bmatrix} \geq 0. \tag{4.16}$$

Let us consider the case when $\sigma_u^2 = 0$. By [54], $\tilde{R} - \text{diag}\left[\sigma_y^2 I_{n+1}, 0\right]$ is equivalent to

$$\text{diag}\left(\tilde{R}_{yy} - \tilde{R}_{uy}' \tilde{R}_{uu}^{-1} \tilde{R}_{uy} - \sigma_y^2 I_{n+1}, \tilde{R}_{uu}\right). \tag{4.17}$$

Using the fact that the covariance matrix $\tilde{R}_{uu}$ is a positive definite matrix and the condition (4.16),

$$\tilde{R}_{yy} - \tilde{R}_{uy}' \tilde{R}_{uu}^{-1} \tilde{R}_{uy} - \sigma_y^2 I_{n+1} \geq 0. \tag{4.18}$$

Hence, the maximum value of $\sigma_y^2$ is the smallest eigenvalue of the matrix

$$\tilde{R}_{yy} - \tilde{R}_{uy}' \tilde{R}_{uu}^{-1} \tilde{R}_{uy} \tag{4.19}$$

by the inequality of (4.18). Similarly, the maximum value of $\sigma_u^2$ can be determined by computing the smallest eigenvalue of the matrix

$$\tilde{R}_{uu} - \tilde{R}_{uy}\tilde{R}_{yy}^{-1}\tilde{R}_{uy}'. \tag{4.20}$$

In short,

$$\max(\sigma_y^2) = \min\Big(\mathrm{eig}\big(\tilde{R}_{yy} - \tilde{R}_{uy}'\tilde{R}_{uu}^{-1}\tilde{R}_{uy}\big)\Big), \tag{4.21}$$

$$\max(\sigma_u^2) = \min\Big(\mathrm{eig}\big(\tilde{R}_{uu} - \tilde{R}_{uy}\tilde{R}_{yy}^{-1}\tilde{R}_{uy}'\big)\Big) \tag{4.22}$$

using the Matlab commands. Indeed, we can find the relation between $\sigma_u^2$ and $\sigma_y^2$ from the above procedure, that is,

$$\sigma_y^2 = \min\Big(\mathrm{eig}\big(\tilde{R}_{yy} - \tilde{R}_{uy}'\left[\tilde{R}_{uu} - \sigma_u^2 I_{n+1}\right]^{-1}\tilde{R}_{uy}\big)\Big). \tag{4.23}$$

Figure 4.1 confirms equation (4.23). It shows the computed value of $\sigma_y^2$ as a function of $\sigma_u^2$, assuming $0 \le \sigma_u^2 \le 1$ for the test, using the system model in Example 3 in Chapter 3. In this example, the true input and output noise variances are $\sigma_u^2 = 0.3$ and $\sigma_y^2 = 1$ respectively. This figure illustrates that one of the pairs of $\{\sigma_u^2, \sigma_y^2\}$ crosses approximately the true value point [0.3, 1] in all the cases when $N = 1,000,\ 2,000$ and $3,000$.

### 4.1.3 Determining Noise Variances and System Parameters

In this section, we follow the searching procedure for noise variances $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_y^2$ suggested in [8] and [57], except that we use the TLS method to estimate the system parameters as initial values in the procedure.

The first step is to estimate the system parameter using each pair of $\{\sigma_u^2(i), \sigma_y^2(i)\}$ obtained in the previous section. We choose the TLS method to obtain the estimated values.

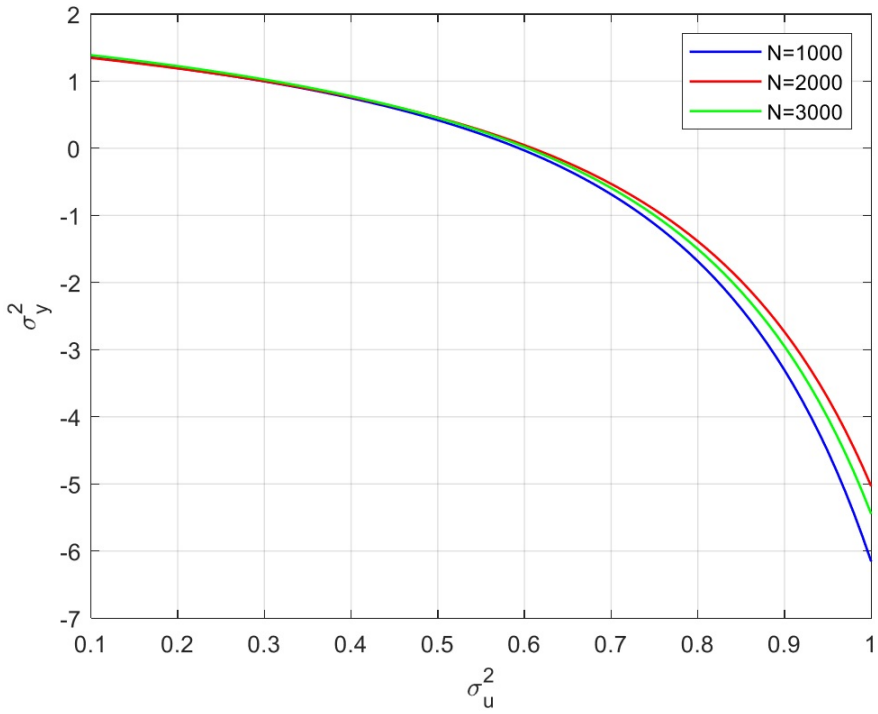


Figure 4.1. $\sigma_y^2$ computation corresponding to $\sigma_u^2$

According to Chapter 3, we need to compute the eigenvalue decomposition of the matrix

$$\tilde{R} - \begin{bmatrix} \sigma_y^2(i) I_{n+1} & 0 \\ 0 & \sigma_u^2(i) I_{n+1} \end{bmatrix} \tag{4.24}$$

where $\sigma_u^2(i)$ and $\sigma_y^2(i)$ are the $i$th pair of $\{\sigma_u^2(i), \sigma_y^2(i)\}$ and the range of the pair can be decided by the maximum value of $\sigma_u^2$ in the equation (4.22). If the eigenvector corresponding to the smallest eigenvalue is

$$v_{\tilde{R}} = \begin{bmatrix} v_{\tilde{R},1} & v_{\tilde{R},2} & \cdots & v_{\tilde{R},2(n+1)} \end{bmatrix}',$$

the estimated system parameter can be obtained by the TLS algorithm

$$\widehat{\theta}(i) = \begin{bmatrix} v_{\tilde{R},2} & \cdots & v_{\tilde{R},2(n+1)} \end{bmatrix}' v_{\tilde{R},1}^{-1} \tag{4.25}$$

which can be denoted as

$$\widehat{\theta}(i) = \begin{bmatrix} -1 & -\widehat{a}_1 & \cdots & -\widehat{a}_n & \widehat{b}_0 & \cdots & \widehat{b}_n \end{bmatrix}'.$$

The next step is to compute the residuals using the $\widehat{\theta}(i)$ for each $i$. Define $\{r_i(k)\}$ with

$$r_i(k) = \frac{1}{N}\sum_{t=1}^{N} \check{\phi}(t)\widehat{\theta}(i)\check{\phi}(t+k)\widehat{\theta}(i), \tag{4.26}$$

where

$$\check{\phi}(t) = \begin{bmatrix} \tilde{y}(t) & \tilde{y}(t-1) & \cdots & \tilde{y}(t-n) & \tilde{u}(t) & \tilde{u}(t-1) & \cdots & \tilde{u}(t-n) \end{bmatrix}.$$

Also define a theoretically computed noise covariance

$$r_{i,\mathrm{T}}(k) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon_{\mathrm{T}}(t)\widehat{\theta}(i)\varepsilon_{\mathrm{T}}(t+k)\widehat{\theta}(i), \tag{4.27}$$

where

$$\varepsilon_{\mathrm{T}}(t) = \begin{bmatrix} \check{\varepsilon}_{\mathrm{out}}(t) & \check{\varepsilon}_{\mathrm{out}}(t-1) & \cdots & \check{\varepsilon}_{\mathrm{out}}(t-n) & \check{\varepsilon}_{\mathrm{in}}(t) & \check{\varepsilon}_{\mathrm{in}}(t-1) & \cdots & \check{\varepsilon}_{\mathrm{in}}(t-n) \end{bmatrix}$$

is generated as i.i.d with the variances

$$\mathbb{E}[\check{\varepsilon}_{\mathrm{out}}(t)^2] = \sigma_y^2(i), \quad \mathbb{E}[\check{\varepsilon}_{\mathrm{in}}(t)^2] = \sigma_u^2(i).$$

Then we compare $r_i(k)$ with $r_{i,\mathrm{T}}(k)$ by defining $\{r_{i,\varepsilon}\}$ as

$$r_{i,\varepsilon} = \begin{bmatrix} r_i(0) - r_{i,\mathrm{T}}(0) \\ \vdots \\ r_i(m) - \rho_{i,\mathrm{T}}(m) \end{bmatrix}, \tag{4.28}$$

where $m$ is a chosen number by the user. The first element in $r_{i,\varepsilon}$ is zero as shown in [57] (Appendix I). Using a weighting matrix

$$W_r = \begin{bmatrix} m+1 & 0 & \cdots & \cdots & 0 \\ 0 & 2m & 0 & \cdots & 0 \\ \vdots & \ddots & 2(m-1) & 0 & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 2 \end{bmatrix},$$

define $V_r$ as

$$V_r(\sigma_u^2(i)) = r'_{i,\varepsilon} W_r r_{i,\varepsilon}, \tag{4.29}$$

which is the same form of the weighting matrix proposed in [57]. Then we search for the minimum value of $V_\rho$ and choose the corresponding pair from $\{\sigma_u^2(i), \sigma_y^2(i)\}$ as the estimate $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_y^2$. We made the $V_r$ curves in terms of $\sigma_u^2$ since it is regarded as a function of $\sigma_u^2$ in (4.29). Figure 4.2 shows the $V_r$ curves with Example 3. The curves illustrate that the minimum value of $V_r$ is approximately at $\sigma_u^2$=0.3 in all the cases when $N$=1,000, 2,000 and 3,000. That means we can estimate noise variances properly through the procedure we described.

### 4.1.4 Simulation Studies

This section's simulation studies are aimed at validating the Frisch scheme in combination with the TLS algorithm.

**Example 4.1**: We used the same plant model and the input signal $u(t)$, which is filtered in the same way, as those in Example 3.3. Table 4.1 presents the average values of $\widehat{\sigma}_u$, $\widehat{\sigma}_y$ and $\widehat{r}$ when the TLS algorithm is used for the noise variance estimation as described in the previous section. The number of ensembles is 1,000 for these simulation studies. It shows that the estimated values are close to their true values $\sigma_u^2 = 0.3$, $\sigma_y^2 = 1$ and $r = 0.5477$. It

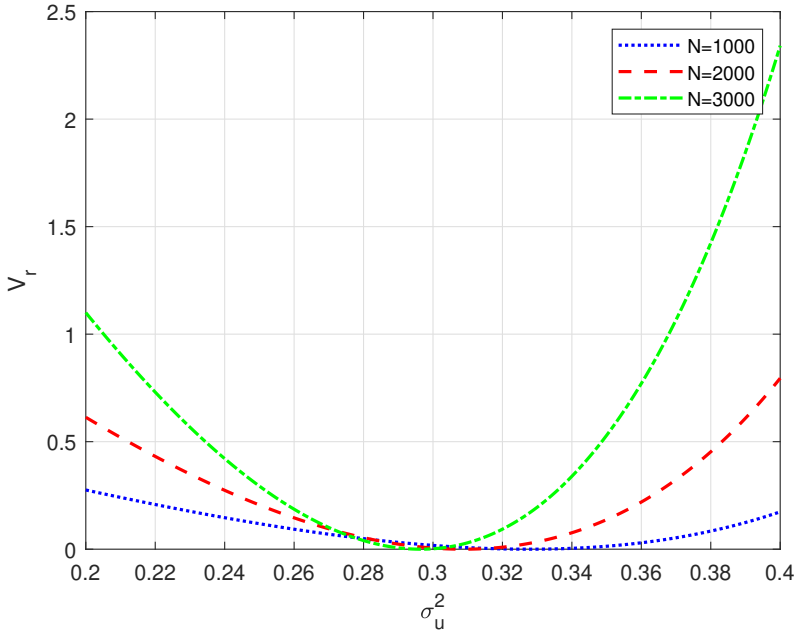


Figure 4.2. $V_r$ curves corresponding to $\sigma_u^2$

also tends to be more accurate as the number of sample data $N$ increases. Figure 4.3 shows

Table 4.1. Estimated values using the Frisch scheme with TLS

| $N$ | 500 | 1000 | 1500 | 2000 | 2500 | 3000 |
|---|---|---|---|---|---|---|
| $\widehat{\sigma}_u$ | 0.2870 $\pm$0.0502 | 0.2929 $\pm$0.0350 | 0.2961 $\pm$0.0281 | 0.2967 $\pm$0.0240 | 0.2962 $\pm$0.0232 | 0.2872 $\pm$0.0205 |
| $\widehat{\sigma}_y$ | 1.003 $\pm$0.1393 | 1.0061 $\pm$0.1000 | 1.0012 $\pm$0.0819 | 1.0014 $\pm$0.0707 | 1.0032 $\pm$0.0657 | 1.0054 $\pm$0.0623 |
| $\widehat{\rho}$ | 0.5389 $\pm$0.0772 | 0.5415 $\pm$0.0534 | 0.5453 $\pm$0.0437 | 0.5454 $\pm$0.0374 | 0.5443 $\pm$0.0357 | 0.5446 $\pm$0.0328 |

the RMSE for EIV system identification using the Frisch scheme in combination with the TLS algorithm. It shows that the RMSE curve is similar to the curve in the case of L $= n$ ($\ell = n$) in Figure 3.5. The GSA algorithm reduces to the TLS algorithm when $\ell = n$. Thus, this simulation's results show that the system parameter estimation works properly with the Frisch scheme in combination with the TLS algorithm.

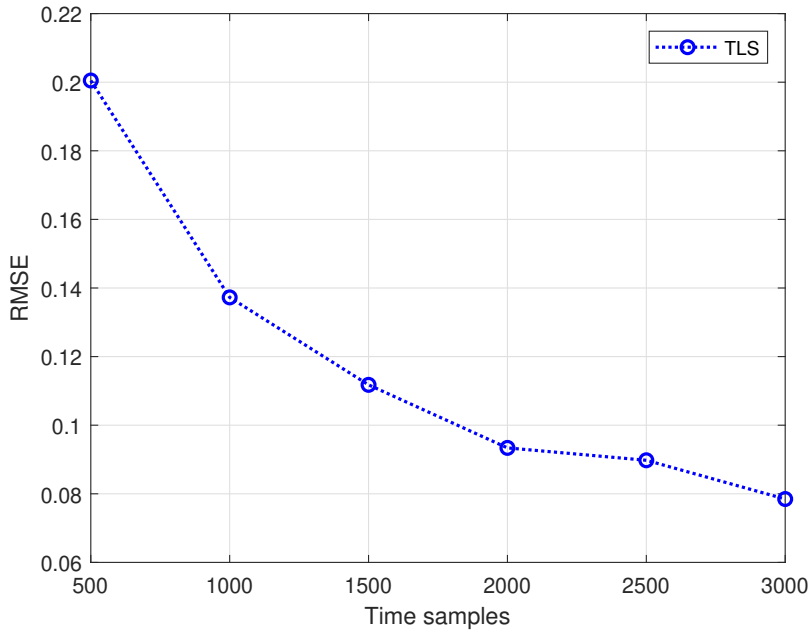


Figure 4.3. RMSE curve in Frisch scheme in combination with TLS

## 4.2 GSA Combined with the Frisch Scheme

In this section, we investigate how to combine the GSA and Frisch scheme. Specifically, we apply the GSA in the searching procedure for the ratio $\rho$. We also present how to estimate the system parameters by the GSA-Frisch method using numerical examples in the case when the noise variances are unknown and unequal.

### 4.2.1 Problem Formulation

For the given input and output measurement and the modified form of the data matrix in Section 3.3, the auto-covariance matrix is obtained and partitioned as

$$\tilde{R}_\ell = \frac{1}{N}\tilde{\Psi}_\ell\tilde{\Psi}'_\ell = \begin{bmatrix} \tilde{R}_{yy} & \tilde{R}_{yu} \\ \tilde{R}_{uy} & \tilde{R}_{uu} \end{bmatrix}.$$

Define the scaled auto-covariance matrix

$$\tilde{R}_{\rho_\ell} = \begin{bmatrix} \rho^2 \tilde{R}_{yy} & \rho \tilde{R}_{yu} \\ \rho \tilde{R}_{uy} & \tilde{R}_{uu} \end{bmatrix}, \quad \rho > 0.$$

Then (3.39) can be written in a different way:

$$\tilde{R}_{\rho_{0\ell}} \quad \rightarrow \quad \begin{bmatrix} \rho_0^2 R_{yy} & \rho_0 R_{yu} \\ \rho_0 R_{uy} & R_{uu} \end{bmatrix} + \sigma_u^2 I_{2(\ell+1)},$$

where $\rho_0$ is un unknown true ratio of the noise variances as $N \rightarrow \infty$. Hence, we notice again that GSA can be applied if the ratio of the noise variances is known. We propose a modified Frisch scheme to search iteratively for $\rho$, such that it converges to $\rho_0$. Without loss of generality, assume that $0 < \rho_0 \leq 1$ is the case in practice, because the output noise has a larger variance than the input noise does.

Even if $\rho \neq \rho_0$, $\tilde{R}_{\rho_{0L}} > 0$ generally. Hence we can apply GSA to estimate the system parameter vector. Define the Toeplitz matrix

$$\begin{aligned} T_{\bar{\theta},\rho} &= \begin{bmatrix} -\theta'_{a_\ell} & 0 & \cdots & 0 & \rho\theta'_{b_\ell} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\theta'_{a_\ell} & 0 & \cdots & 0 & \rho\theta_{b_\ell} \end{bmatrix}, \\ &= \begin{bmatrix} -a_0 & \cdots & -a_n & 0 & \cdots & 0 & \rho b_0 & \cdots & \rho b_n & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -a_0 & \cdots & -a_n & 0 & \cdots & 0 & \rho b_0 & \cdots & \rho b_n \end{bmatrix} \in \mathbb{R}^{d_\eta \times 2(\ell+1)}. \end{aligned}$$

Then there holds

$$T_{\bar{\theta},\rho} \Psi_\ell = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}$$

in the noise-free case.

Let $\widehat{\bar{\theta}}$ be the estimate obtained with the GSA. Then we have

$$T_{\widehat{\bar{\theta}},\rho}\tilde{\Psi}_\ell = \begin{bmatrix} \widehat{\varepsilon}_\rho(k_0) & \widehat{\varepsilon}_\rho(k_0+1) & \cdots & \cdots & \widehat{\varepsilon}_\rho(k_0+N-1) \\ \vdots & \ddots & \ddots & & \vdots \\ \widehat{\varepsilon}_\rho(k_0-\ell+n) & \cdots & \widehat{\varepsilon}_\rho(k_0) & \widehat{\varepsilon}_\rho(k_0+1) & \cdots & \widehat{\varepsilon}_\rho(k_0+N-\ell+n-1) \end{bmatrix}.$$

If $\widehat{\bar{\theta}} = \bar{\theta}$ and $\rho = \rho_0$, then $\widehat{\varepsilon}_\rho(k) = \varepsilon_{\rho_0}(k)$, that is, the model prediction error is induced by measurement errors. But if $\widehat{\bar{\theta}} \neq \bar{\theta}$ and $\rho \neq \rho_0$, then the error residue yields the following auto-covariance matrix:

$$\begin{aligned} \widehat{R}_{\varepsilon_\rho} = \frac{1}{N} T_{\widehat{\bar{\theta}},\rho}\tilde{\Psi}_\ell \tilde{\Psi}_\ell' T_{\widehat{\bar{\theta}},\rho}' &= \begin{bmatrix} \widehat{r}_\varepsilon(0,0) & \widehat{r}_\varepsilon(0,1) & \cdots & \widehat{r}_\varepsilon(0,m) \\ \widehat{r}_\varepsilon(1,0) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \widehat{r}_\varepsilon(m-1,m) \\ \widehat{r}_\varepsilon(m,0) & \cdots & \widehat{r}_\varepsilon(m,m-1) & \widehat{r}_\varepsilon(m,m) \end{bmatrix} \\ &= \begin{bmatrix} \widehat{r}_\varepsilon(0) & \widehat{r}_\varepsilon(1) & \cdots & \widehat{r}_\varepsilon(m) \\ \widehat{r}_\varepsilon(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \widehat{r}_\varepsilon(1) \\ \widehat{r}_\varepsilon(m) & \cdots & \widehat{r}_\varepsilon(1) & \widehat{r}_\varepsilon(0) \end{bmatrix}, \quad m = \ell - n. \end{aligned}$$

Because $N$ is finite, $\widehat{R}_{\varepsilon_\rho}$ is not exactly a Toeplitz matrix, i.e.,

$$\widehat{r}_\varepsilon(i-j) \neq \widehat{r}_\varepsilon(i,j).$$

One may average the diagonal elements to obtain an exactly Toeplitz matrix. That is, we can set

$$\widehat{r}_\varepsilon(0) = \frac{1}{m+1}\sum_{i=0}^{m}\widehat{r}_\varepsilon(i,i), \quad \widehat{r}_\varepsilon(k) = \frac{1}{2(m+1-k)}\sum_{k=|i-j|}\widehat{r}_\varepsilon(i,j), \quad k \neq 0. \tag{4.30}$$

The summation for $k = |i - j|$ means that the summation is over all possible integer pairs $(i, j)$ satisfying $i \neq j$ and $k = |i - j|$. There are a total of $2(m + 1 - k)$ terms in the case of $k \neq 0$.

Next we generate i.i.d. sequences $\{\delta_u(k), \delta_y(k)\}_{k=1}^{N+m}$ with mean zero and variance $\sigma^2$, equal to the average of the $m$ minimum eigenvalues of $\tilde{R}_{\rho\ell}$, and form the blocked Toeplitz matrix

$$\Psi_{\delta_\rho} = \begin{bmatrix} \Psi_{\delta_y} \\ \Psi_{\delta_u} \end{bmatrix}, \quad \Psi_{\delta_y} = \begin{bmatrix} \delta_y(m+1) & \cdots & \delta_y(N) & \delta_y(N+1) & \cdots & \delta_y(N+m) \\ \delta_y(m) & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \delta_y(N+1) \\ \delta_y(1) & \cdots & \delta_y(m) & \delta_y(m+1) & \cdots & \delta_y(N) \end{bmatrix},$$

$$\Psi_{\delta_u} = \begin{bmatrix} \delta_u(m+1) & \cdots & \delta_u(N) & \delta_u(N+1) & \cdots & \delta_u(N+m) \\ \delta_u(m) & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \delta_u(N+1) \\ \delta_u(1) & \cdots & \delta_u(m) & \delta_u(m+1) & \cdots & \delta_u(N) \end{bmatrix}.$$

The theoretical autocavariance matrix can then be obtained as follows:

$$\widehat{R}_\delta = \frac{1}{N} T_{\widehat{\theta},\rho} \tilde{\Psi}_{\delta_\rho} \tilde{\Psi}'_{\delta_\rho} T'_{\widehat{\theta},\rho} = \begin{bmatrix} \widehat{r}_\delta(0) & \widehat{r}_\delta(1) & \cdots & \widehat{r}_\delta(m) \\ \widehat{r}_\delta(-1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \widehat{r}_\delta(1) \\ \widehat{r}_\delta(-m) & \cdots & \widehat{r}_\delta(-1) & \widehat{r}_\delta(0) \end{bmatrix}.$$

Finally for each $\rho \in (0,\ 1]$, we compute

$$J(\rho) = \sum_{k=0}^{m} w_k [\widehat{r}_\delta(k) - \widehat{r}_\varepsilon(k)]^2$$

for some weights $w_k > 0$. Soderstrom's paper [57] chooses $w_0 = m+1$ and $w_k = 2(m+1-k)$ for $1 \leq i \leq m$, which are the same as those in the denominators of (4.30), respectively.

Hence, if this is what we choose for the weights, in light of (4.30), $J(\rho)$ can be written as

$$J(\rho) = \sum_{i=0}^{m} [\widehat{r}_{\varepsilon}(i,i) - \widehat{r}_{\delta}(i,i)]^2 + \sum_{k=1}^{m} \sum_{k=|i-j|} [\widehat{r}_{\varepsilon}(i,j) - \widehat{r}_{\delta}(i,j)]^2.$$

We compute $J(\rho)$ at each given $\rho$ with

$$\widehat{\theta}_{\rho} = \begin{bmatrix} -\widehat{\theta}_a' & \rho\widehat{\theta}_b' \end{bmatrix}'$$

which is the corresponding scaled estimated parameter vector. The estimate $\widehat{\rho}$ for $\rho_0$ is set to be the minimizer of $J(\rho)$. That is,

$$\widehat{\rho} = \arg\min_{0<\rho\leq 1} J(\rho).$$

### 4.2.2 Simulation Studies

In this section, we validate the estimate performance of the ratio using the search procedure described in the previous section. Also, we evaluate the performance of the system parameter estimate when the estimated ratio is used.

**Example 4.2**: We again used the same plant model and the input signal $u(t)$ as those in Example 3.3. In this example, we used 1,000 ensembles and $\rho \in [0.3,\ 0.8]$ with increments of 0.01 for the searching process. The true ratio of the noise variances is $\rho_0 = \sqrt{0.3} = 0.5477$. Table 4.2 shows the average values of the estimated ratios according to the number of sample data $N$ and block size $\ell$.

Table 4.2. Average values of the estimated ratio $\widehat{\rho}$

| $N$ | 500 | 1000 | 1500 | 2000 | 2500 | 3000 |
|---|---|---|---|---|---|---|
| $\ell = n$ | 0.5299 | 0.5382 | 0.5394 | 0.5430 | 0.5433 | 0.5431 |
| $\ell = n+1$ | 0.5328 | 0.5378 | 0.5416 | 0.5440 | 0.5430 | 0.5430 |
| $\ell = n+5$ | 0.4846 | 0.5170 | 0.5255 | 0.5341 | 0.5344 | 0.5364 |

Table 4.2 shows that the estimated values tend to close to the true value as $N$ increases. Figure 4.4 shows the RMSE between the true and estimated ratio. The RMSE is computed with

$$\text{RMSE} := \sqrt{\frac{1}{T}\sum_{i=1}^{T}(\widehat{\rho}_i - \rho_0)^2},$$

where $\widehat{\rho}_i$ is the estimated ratio from the $i$th simulation.

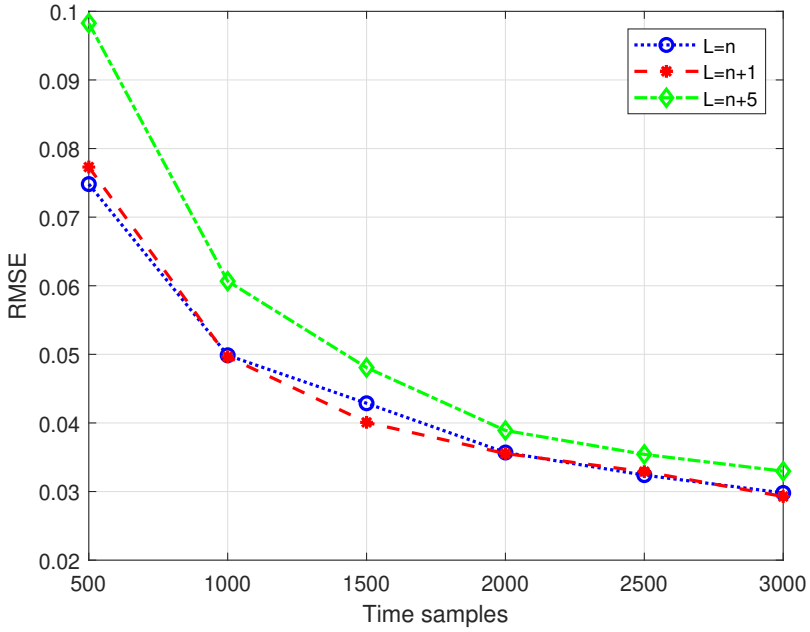


Figure 4.4. RMSE between the true ratio and the estimated ratio

Figure 4.4 shows that the estimate performance of the ratio degrades as $\ell$ increases, especially with small $N$. Nevertheless, Figure 4.5 shows that this method can be employed in the system identification. Figure 4.5 shows the RMSE curves between the true and estimated system parameter computed by (3.40) when the estimated ratios are applied for the GSA algorithm. It shows similar RMSE curves compared to Figure 3.3 which uses the true ratio value. It clearly shows that the performance improves as $\ell$ increases. Thus, it validates that the GSA-Frisch scheme effectively works for the system parameter estimation.

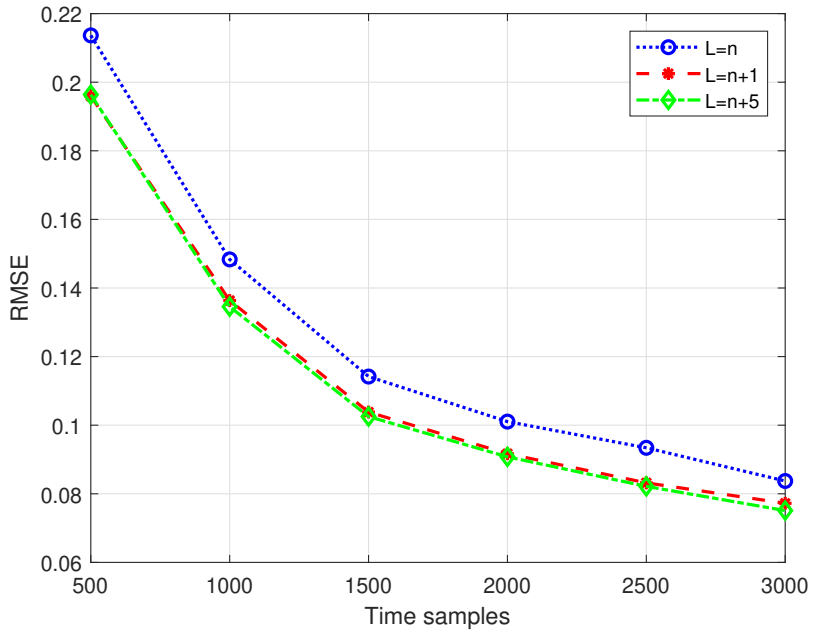


Figure 4.5. RMSE for EIV identification based on the estimated ratio

## 4.3 New EIV Identification

In this section, a new EIV system identification algorithm is presented. It is motivated by its success in blind channel estimation [1, 9, 46] and state-space model identification [14, 32, 80, 82, 81]. Its consistency will be analyzed and demonstrated by numerical examples. Also we compare the simulation results with those from GSA combined with the Frisch scheme.

### 4.3.1 Notations

For input signal $\{u(t)\}$ and $\ell \geq n$, define its $\ell$-blocked past and future signal vectors respectively as

$$u_{\rm p}(t) := \begin{bmatrix} u(t-1) & u(t-2) & \cdots & u(t-\ell-1) \end{bmatrix}',$$
$$u_{\rm f}(t) := \begin{bmatrix} u(t+\ell) & \cdots & u(t+1) & u(t) \end{bmatrix}'.$$

The subscripts p and f indicate the collection of past and future data, respectively. Next, we form the following past and future Toeplitz matrices of dimension $(\ell+1)\times N$ as

$$U_{\mathrm{p}_N} = \begin{bmatrix} u_{\mathrm{p}}(t) & u_{\mathrm{p}}(t+1) & \cdots & u_{\mathrm{p}}(t+N-1) \end{bmatrix},$$
$$U_{\mathrm{f}_N} = \begin{bmatrix} u_{\mathrm{f}}(t) & u_{\mathrm{f}}(t+1) & \cdots & u_{\mathrm{f}}(t+N-1) \end{bmatrix}.$$

Similarly, define $\ell$-blocked past and future signal vectors for the output signal $\{y(t)\}$ as

$$y_{\mathrm{p}}(t) := \begin{bmatrix} y(t-1) \\ y(t-2) \\ \vdots \\ y(t-\ell-1) \end{bmatrix}, \quad y_{\mathrm{f}}(t) := \begin{bmatrix} y(t+\ell) \\ \vdots \\ y(t+1) \\ y(t) \end{bmatrix},$$

and the Toeplitz matrices of dimension $(\ell+1)\times N$ as follows:

$$Y_{\mathrm{p}_N} = \begin{bmatrix} y_{\mathrm{p}}(t) & u_{\mathrm{p}}(t+1) & \cdots & y_{\mathrm{p}}(t+N-1) \end{bmatrix},$$
$$Y_{\mathrm{f}_N} = \begin{bmatrix} y_{\mathrm{f}}(t) & y_{\mathrm{f}}(t+1) & \cdots & y_{\mathrm{f}}(t+N-1) \end{bmatrix}.$$

In addition, past and future Toeplitz matrices for the noisy input and output measurements $\{\tilde{u}(k), \tilde{y}(k)\}$

$$\tilde{U}_{\mathrm{p}_N} = \begin{bmatrix} \tilde{u}_{\mathrm{p}}(t) & \tilde{u}_{\mathrm{p}}(t+1) & \cdots & \tilde{u}_{\mathrm{p}}(t+N-1) \end{bmatrix},$$
$$\tilde{U}_{\mathrm{f}_N} = \begin{bmatrix} \tilde{u}_{\mathrm{f}}(t) & \tilde{u}_{\mathrm{f}}(t+1) & \cdots & \tilde{u}_{\mathrm{f}}(t+N-1) \end{bmatrix},$$
$$\tilde{Y}_{\mathrm{p}_N} = \begin{bmatrix} \tilde{y}_{\mathrm{p}}(t) & \tilde{u}_{\mathrm{p}}(t+1) & \cdots & \tilde{y}_{\mathrm{p}}(t+N-1) \end{bmatrix},$$
$$\tilde{Y}_{\mathrm{f}_N} = \begin{bmatrix} \tilde{y}_{\mathrm{f}}(t) & \tilde{y}_{\mathrm{f}}(t+1) & \cdots & \tilde{y}_{\mathrm{f}}(t+N-1) \end{bmatrix},$$

where

$$\tilde{u}_{\mathrm{p}}(t) := \begin{bmatrix} \tilde{u}(t-1) \\ \tilde{u}(t-2) \\ \vdots \\ \tilde{u}(t-\ell-1) \end{bmatrix}, \tilde{u}_{\mathrm{f}}(t) := \begin{bmatrix} \tilde{u}(t+\ell) \\ \vdots \\ \tilde{u}(t+1) \\ \tilde{u}(t) \end{bmatrix},$$

$$\tilde{y}_{\mathrm{p}}(t) := \begin{bmatrix} \tilde{y}(t-1) \\ \tilde{y}(t-2) \\ \vdots \\ \tilde{y}(t-\ell-1) \end{bmatrix}, \tilde{y}_{\mathrm{f}}(t) := \begin{bmatrix} \tilde{y}(t+\ell) \\ \vdots \\ \tilde{y}(t+1) \\ \tilde{y}(t) \end{bmatrix}$$

can be defined.

Alternatively, we introduce different forms of the past and future data, which revealed better performance than the previous one from our simulation experiences. We form the following past and future Toeplitz matrices of dimension $(\ell+1) \times N$ for the noise-free input and output data as

$$\begin{aligned} U_{\mathrm{p}_N} &= \begin{bmatrix} u_{\mathrm{p}}(t-N+1) & \cdots & u_{\mathrm{p}}(t-1) & u_{\mathrm{p}}(t) \end{bmatrix}, \\ U_{\mathrm{f}_N} &= \begin{bmatrix} u_{\mathrm{f}}(t) & u_{\mathrm{f}}(t+1) & \cdots & u_{\mathrm{f}}(t+N-1) \end{bmatrix}, \\ Y_{\mathrm{p}_N} &= \begin{bmatrix} y_{\mathrm{p}}(t-N+1) & \cdots & y_{\mathrm{p}}(t-1) & y_{\mathrm{p}}(t) \end{bmatrix}, \\ Y_{\mathrm{f}_N} &= \begin{bmatrix} y_{\mathrm{f}}(t) & y_{\mathrm{f}}(t+1) & \cdots & y_{\mathrm{f}}(t+N-1) \end{bmatrix}. \end{aligned}$$

The Toeplitz matrices for the noisy input and output data, $\tilde{U}_{\mathrm{p}_N}, \tilde{U}_{\mathrm{f}_N}, \tilde{Y}_{\mathrm{p}_N}, \tilde{Y}_{\mathrm{f}_N}$, and for the measurement noises, $\Delta^{(u)}_{\mathrm{p}_N}, \Delta^{(u)}_{\mathrm{f}_N}, \Delta^{(y)}_{\mathrm{p}_N}, \Delta^{(y)}_{\mathrm{f}_N}$ are formed in the same way.

Then the past and future noise Toeplitz matrices

$$\Delta^{(u)}_{\mathrm{p}_N} = \tilde{U}_{\mathrm{p}_N} - U_{\mathrm{p}_N}, \qquad \Delta^{(u)}_{\mathrm{f}_N} = \tilde{U}_{\mathrm{f}_N} - U_{\mathrm{f}_N},$$

$$\Delta_{\mathrm{p}_N}^{(y)} = \tilde{Y}_{\mathrm{p}_N} - Y_{\mathrm{p}_N}, \qquad \Delta_{\mathrm{f}_N}^{(y)} = \tilde{Y}_{\mathrm{f}_N} - Y_{\mathrm{f}_N},$$

can also be defined, consistent with the EIV system and signal model in Figure 1.2, and those described in (3.1) and the following input and output relation:

$$a(q)\tilde{y}(t) - a(q)\delta_y(t) = b(q)\tilde{u}(t) - b(q)\delta_u(t). \tag{4.31}$$

For the given noisy and noise-free measurement data, define the following correlation matrices as

$$\tilde{R}_N := \frac{1}{N} \begin{bmatrix} \tilde{Y}_{\mathrm{f}_N} \\ \tilde{U}_{\mathrm{f}_N} \end{bmatrix} \begin{bmatrix} \tilde{Y}'_{\mathrm{p}_N} & \tilde{U}'_{\mathrm{p}_N} \end{bmatrix}, \tag{4.32}$$

$$R_N := \frac{1}{N} \begin{bmatrix} Y_{\mathrm{f}_N} \\ U_{\mathrm{f}_N} \end{bmatrix} \begin{bmatrix} Y'_{\mathrm{p}_N} & U'_{\mathrm{p}_N} \end{bmatrix}. \tag{4.33}$$

Similarly, for $\ell > n$, define two Toeplitz matrices of dimension $(\ell + 1 - n) \times (\ell + 1)$ for the system parameters as

$$T_a := \begin{bmatrix} a_0 & a_1 & \cdots & a_n & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_0 & a_1 & \cdots & a_n \end{bmatrix}, \tag{4.34}$$

$$T_b := \begin{bmatrix} b_0 & b_1 & \cdots & b_n & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_0 & b_1 & \cdots & b_n \end{bmatrix}. \tag{4.35}$$

### 4.3.2 Problem Formulation

In light of the EIV system and signal model in the previous chapter, there holds

$$\theta_a Y_{\mathrm{f}_N} = \theta_b U_{\mathrm{f}_N} \tag{4.36}$$

that is equivalent to

$$T_a[\tilde{Y}_{\mathrm{f}_N} - \Delta^{(y)}_{\mathrm{f}_N}] = T_b[\tilde{U}_{\mathrm{f}_N} - \Delta^{(u)}_{\mathrm{f}_N}]. \tag{4.37}$$

There also holds $\theta_a Y_{\mathrm{p}_N} = \theta_b U_{\mathrm{p}_N}$, which is equivalent to

$$T_a[\tilde{Y}_{\mathrm{p}_N} - \Delta^{(y)}_{\mathrm{p}_N}] = T_b[\tilde{U}_{\mathrm{p}_N} - \Delta^{(u)}_{\mathrm{p}_N}]. \tag{4.38}$$

Therefore, in the noise-free case, there hold

$$\begin{aligned} \begin{bmatrix} T_a & -T_b \end{bmatrix} \begin{bmatrix} Y_{\mathrm{f}_N} \\ U_{\mathrm{f}_N} \end{bmatrix} = 0, \\ \begin{bmatrix} T_a & -T_b \end{bmatrix} R_N = 0. \end{aligned} \tag{4.39}$$

Unfortunately, input and output measurement data are corrupted by noises. Hence, relations in (4.37)-(4.39) do not help much to solve the EIV identification problem. The following observation is instrumental.

**Lemma 1.** *Consider the EIV system and signal model in Figure 1.2 and described in (3.1) and (4.31). Suppose that measurement noises* $\{\varepsilon_{\mathrm{out}}(t), \varepsilon_{\mathrm{in}}(t)\}$ *are independent white processes, and uncorrelated to the uniformly bounded quasi-stationary input and output signals* $\{u(t), y(t)\}$. *There holds the following asymptotic equality:*

$$\lim_{N\to\infty} R_N = \lim_{N\to\infty} \tilde{R}_N. \tag{4.40}$$

Proof: Because $\{\varepsilon_{\text{out}}(t), \varepsilon_{\text{in}}(t)\}$ are independent white processes, and uncorrelated to the true input and output signal $\{u(t), y(t)\}$ that are quasi-stationary, there hold

$$\lim_{N\to\infty} \frac{1}{N} \Delta_{\text{p}_N}^{(y)} \Delta_{\text{f}_N}^{(u)'} = \lim_{N\to\infty} \frac{1}{N} \Delta_{\text{f}_N}^{(y)} \Delta_{\text{p}_N}^{(u)'} = 0, \tag{4.41}$$

$$\lim_{N\to\infty} \frac{1}{N} \Delta_{\text{p}_N}^{(y)} \Delta_{\text{f}_N}^{(y)'} = \lim_{N\to\infty} \frac{1}{N} \Delta_{\text{f}_N}^{(u)} \Delta_{\text{p}_N}^{(u)'} = 0. \tag{4.42}$$

By the definitions in (4.32) and (4.33), we have

$$\begin{aligned} \tilde{R}_N = R_N + \frac{1}{N} \begin{bmatrix} Y_{\text{f}_N} \\ U_{\text{f}_N} \end{bmatrix} \begin{bmatrix} \Delta_{\text{p}_N}^{(y)'} & \Delta_{\text{p}_N}^{(u)'} \end{bmatrix} + \frac{1}{N} \begin{bmatrix} \Delta_{\text{f}_N^{(y)}} \\ \Delta_{\text{f}_N}^{(u)} \end{bmatrix} \begin{bmatrix} Y'_{\text{p}_N} & U'_{\text{p}_N} \end{bmatrix} \\ + \frac{1}{N} \begin{bmatrix} \Delta_{\text{f}_N}^{(y)} \\ \Delta_{\text{f}_N}^{(u)} \end{bmatrix} \begin{bmatrix} \Delta_{\text{p}_N}^{(y)'} & \Delta_{\text{p}_N}^{(u)'} \end{bmatrix}. \end{aligned}$$

The last product in the above expression approaches zero by the asymptotic relations in (4.41) and (4.42). The hypotheses on the true input and output signals imply that the second and third products are also zero asymptotically as $N \to \infty$. Hence the limit in (4.40) holds true. □

The result in Lemma 1 suggests that

$$\begin{bmatrix} T_a & -T_b \end{bmatrix} \tilde{R}_N \approx 0 \tag{4.43}$$

for large $N$.

### 4.3.3 Applying TLS Algorithm

In the case when $\ell = n$, the TLS algorithm can be applied, considering that each element of $\tilde{R}_N$ involves measurement noises that are close to white and have similar variances. For $\ell > n$, a modified TLS algorithm can be developed. We present these two algorithms. The TLS algorithm is summarized prior to the modified TLS algorithm step by step as follows.

**TLS Algorithm:**

- Step 1: Given input and output data $\{\tilde{y}(t), \tilde{u}(t)\}$ as described in (3.1) and (3.8), form the past and future Toeplitz matrices $\{\tilde{Y}_{\mathrm{p},N}, \tilde{Y}_{\mathrm{f},N}\}$ and $\{\tilde{U}_{\mathrm{p},N}, \tilde{U}_{\mathrm{f},N}\}$ with $\ell = n$.

- Step 2: Form the correlation matrix $\tilde{R}_N$ as in (4.32) and compute its SVD as

$$\tilde{R}_N = \tilde{G}_N \tilde{\Sigma}_N \tilde{V}_N', \quad \tilde{G}_N = \begin{bmatrix} \tilde{g}_1 & \tilde{g}_2 & \cdots & \tilde{g}_{2(n+1)} \end{bmatrix}.$$

- Step 3: Set $\theta_{\mathrm{NEW}} = \begin{bmatrix} \theta_a & -\theta_b \end{bmatrix}$. Let $\tilde{g}_{2(n+1)}$ be the last column of $\tilde{G}_N$ and set the estimate in accordance with

$$\begin{aligned} \widehat{\theta}_{\mathrm{NEW}} &= \begin{bmatrix} \widehat{\theta}_a & -\widehat{\theta}_b \end{bmatrix} \\ &= \begin{bmatrix} 1 & \widehat{a}_1 & \cdots & \widehat{a}_n & -\widehat{b}_0 & -\widehat{b}_1 & \cdots & -\widehat{b}_n \end{bmatrix} \\ &= \tilde{g}'_{2(n+1)} / \tilde{g}_{1,2(n+1)}, \end{aligned} \tag{4.44}$$

  where $\tilde{g}_{1,2(n+1)}$ is the first element of $\tilde{g}_{2(n+1)}$. □

In the SVD step, $\tilde{\Sigma}_N$ is diagonal with $2(n+1)$ singular values of $\tilde{R}_N$ in descending order. In light of [26], $\tilde{g}_{1,2(n+1)} \neq 0$ with probability 1.

In the case of $\ell > n$, denote

$$\tilde{R}_{N,i} = [\tilde{R}_N(i : n+i, 1 : 2(\ell+1)); \tilde{R}_N(i+\ell+1 : i+\ell+n+1, 1 : 2(\ell+1))],$$

where $i = 1, \cdots, \ell - n + 1$, and $\tilde{R}_{N,i}$ consists of the $2(n+1)$ rows of $\tilde{R}_N$. Then the approximate relation (4.43) implies that

$$\theta_{\mathrm{NEW}} \begin{bmatrix} \tilde{R}_{N,1} & \tilde{R}_{N,2} & \cdots & \tilde{R}_{N,\ell-n+1} \end{bmatrix} \approx 0.$$

The previous TLS algorithm can be modified to compute the EVD for

$$\tilde{M}_N = \frac{1}{2(\ell+1)} \sum_{i=1}^{\ell-n+1} \tilde{R}_{N,i} \tilde{R}'_{N,i} = \tilde{G}_N \tilde{\Sigma}_N^2 \tilde{G}'_N. \tag{4.45}$$

The modified TLS algorithm is presented next.

**Modified TLS Algorithm:**

- Step 1: Given input and output data $\{\tilde{y}(t), \tilde{u}(t)\}$ as described in (3.1) and (4.31), form the past and future Toeplitz matrices $\{\tilde{Y}_{\mathrm{p},N}, \tilde{Y}_{\mathrm{f},N}\}$ and $\{\tilde{U}_{\mathrm{p},N}, \tilde{U}_{\mathrm{f},N}\}$ with $\ell > n$.

- Step 2: Form the correlation matrix $\tilde{R}_N$ as in (4.32), $\tilde{M}_N$ as in (4.45), and compute the EVD of $\tilde{M}_N$ as

$$\tilde{M}_N = \tilde{G}_N \tilde{\Sigma}_N^2 \tilde{G}'_N, \quad \tilde{G}_N = \left[ \begin{array}{ccc} \tilde{g}_1 & \cdots & \tilde{g}_{2(n+1)} \end{array} \right].$$

- Step 3: Let $\tilde{g}_{2(n+1)}$ be the last column of $\tilde{G}_N$ and set the estimate $\widehat{\theta}_{\mathrm{NEW}}$ in accordance with the same expression as in (4.44). □

We observe that if $\ell = n$ in the modified TLS algorithm, it reduces to the TLS algorithm. The consistency of the estimate $\widehat{\theta}_{\mathrm{NEW}}$ from these two algorithms is summarized next.

**Theorem 4.** *Under the same hypotheses as those of Lemma 1, plus the persistent excitation for the input, let $\widehat{\theta}_{\mathrm{NEW}}$ be obtained in (4.44) with either the TLS algorithm for the case of $\ell = n$, or the modified TLS algorithm for the case of $\ell > n$, there holds $\widehat{\theta}_{\mathrm{NEW}} \to \theta_{\mathrm{NEW}}$ as $N \to \infty$. Moreover, these two algorithms are unbiased in the sense that $\mathrm{E}\{\widehat{\theta}_{\mathrm{NEW}}\} = \theta_{\mathrm{NEW}}$.*

Proof: The unbiased property for the TLS and modified TLS algorithm follow from the asymptotic result in (4.40) of Lemma 1. The consistency of the algorithm follows the same reason and the persistent excitation assumption on the input signal. The detail is omitted. □

**Remark 6.** *The TLS and its modified algorithms have similar subspace interpretations as [1, 9, 46]. Indeed, in the case of $\ell = n$, we can regard $\mathcal{R}\{\tilde{g}_{2(n+1)}\}$ as the noise subspace, and $\mathcal{R}\{\begin{bmatrix} \tilde{g}_1 & \cdots & \tilde{g}_{2n+1} \end{bmatrix}\}$ as the signal subspace. Such an interpretation holds for the case of $\ell > n$.*

### 4.3.4 State-space Formulation

This section presents an overview on N4SID algorithm based on [80]. We observed that the EIV identification problem can be cast and solved by the subspace algorithm for the state-space model, which refers to the N4SID algorithm [14, 32, 80, 82, 81]. To be specific, assume that the system represented by $P(z)$ admits a minimal state-space realization, i.e.,

$$P(z) = \frac{b(z)}{a(z)} = D + C(zI - A)^{-1}B,$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^n$, $C \in \mathbb{R}^{1\times n}$, and $D \in \mathbb{R}$. Then by setting $w(t) = -B\varepsilon_{\text{in}}(t)$ and $v(t) = \varepsilon_{\text{out}}(t) - D\varepsilon_{\text{in}}(t)$, we obtain the following state-space description:

$$\begin{aligned} x(t+1) &= Ax(t) + B\tilde{u}(t) + w(t), \\ \tilde{y}(t) &= Cx(t) + D\tilde{u}(t) + v(t), \end{aligned}$$

identical to that of the N4SID algorithm in the literature. Moreover, the covariances of the process and measurement noises are given via the following expression:

$$\begin{aligned} \begin{bmatrix} Q & S \\ S' & R \end{bmatrix} &= \mathrm{E}\left\{ \begin{bmatrix} w(t) \\ v(t) \end{bmatrix} \begin{bmatrix} w(t) & v(t) \end{bmatrix} \right\} \\ &= \begin{bmatrix} \sigma_u^2 BB' & \sigma_u^2 BD' \\ \sigma_u^2 DB' & \sigma_y^2 + \sigma_u^2 DD' \end{bmatrix}. \end{aligned}$$

However, caution needs to be taken by noting that input $\{\tilde{u}(t)\}$ is correlated with $\{w(t), v(t)\}$, different from that assumed in the N4SID algorithm. Nevertheless, the N4SID algorithm seems to be the only algorithm that is available in Matlab and can be used to solve the EIV identification problem, compared with the proposed algorithm. In fact, the simulation results suggest that the N4SID algorithm is a viable algorithm for EIV identification. We will provide a brief review in this section.

In the N4SID algorithm, the past and future data of the input and output are defined a little differently. Specifically, for signal $\{s(t)\}$ and $\ell \geq n$, $s(t)$ is termed as the present, and

$$s_{\rm p}(t) := \begin{bmatrix} s(t-\ell) \\ \vdots \\ s(t-2) \\ s(t-1) \end{bmatrix}, \quad s_{\rm f}(t) := \begin{bmatrix} s(t+1) \\ s(t+2) \\ \vdots \\ s(t+\ell) \end{bmatrix},$$

as past and future, respectively, by an abuse of notation. As such, the Toeplitz matrices in the previous section now have Hankel structure:

$$\begin{aligned} S_{{\rm p}_N} &= \begin{bmatrix} s_{\rm p}(t) & s_{\rm p}(t+1) & \cdots & s_{\rm p}(t+N-1) \end{bmatrix}, \\ S_{t_N} &= \begin{bmatrix} s(t) & s(t+1) & \cdots & s(t+N-1) \end{bmatrix}, \\ S_{{\rm f}_N} &= \begin{bmatrix} s_{\rm f}(t+1) & s_{\rm f}(t+2) & \cdots & s_{\rm f}(t+N) \end{bmatrix}. \end{aligned}$$

With the above notations for $\{u(t), y(t)\}$ and $\{\tilde{u}(k), \tilde{y}(k)\}$, respectively, we form Hankel matrices of

$$\tilde{Y}_{{\rm p}_N}, \quad \tilde{Y}_{{\rm f}_N}, \quad \tilde{U}_{{\rm p}_N}, \quad \tilde{U}_{{\rm f}_N}.$$

Next define the extended observability matrix

$$\Gamma_\ell = \begin{bmatrix} C & CA & \cdots & CA^{\ell-1} \end{bmatrix}',$$

extended reversed controllability matrix as

$$\Delta_\ell = \begin{bmatrix} A^{\ell-1}B & A^{\ell-2}B & \cdots & AB & B \end{bmatrix},$$

and the lower block triangular Toeplitz matrix

$$H_\ell = \begin{bmatrix} D & 0 & \dots & 0 \\ CB & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{\ell-2}B & CA^{\ell-3}B & \dots & D \end{bmatrix}.$$

Finally, the state matrix is denoted by

$$X_\ell = \begin{bmatrix} x_\ell & x_{\ell+1} & x_{\ell+2} & \cdots & x_{\ell+N-1} \end{bmatrix}.$$

As the first step, N4SID algorithm carries out the RQ factorization of all the collection of the data as follows:

$$\begin{bmatrix} \tilde{U}_{\mathrm{p}_N} \\ \tilde{U}_{k_N} \\ \tilde{U}_{\mathrm{f}_N} \\ \tilde{Y}_{\mathrm{p}_N} \\ \tilde{Y}_{k_N} \\ \tilde{Y}_{\mathrm{f}_N} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ R_{61} & \cdots & \cdots & R_{66} \end{bmatrix} \begin{bmatrix} Q_1' \\ Q_2' \\ \vdots \\ Q_6' \end{bmatrix}.$$

For matrix $\Phi$ of the full row rank, denote

$$\Pi_\Phi = \Phi'(\Phi\Phi')^{-1}\Phi, \quad \Phi = \begin{bmatrix} \tilde{U}_{\mathrm{p}_N} \\ \tilde{U}_{k_N} \\ \tilde{U}_{\mathrm{f}_N} \\ \tilde{Y}_{\mathrm{p}_N} \end{bmatrix},$$

as the orthogonal projection along the range space of $\Phi'$. Then the following projection

$$\begin{aligned} Z_k &= \begin{bmatrix} \tilde{Y}_{k_N} \\ \tilde{Y}_{f_N} \end{bmatrix} \Pi_\Phi = R_{5:6,1:4} R^{-1}_{1:4,1:4} \Phi \\ &=: \begin{bmatrix} L_k^{(1)} & L_k^{(2)} & L_k^{(3)} \end{bmatrix} \begin{bmatrix} \tilde{U}_{\mathrm{p}_N} \\ \tilde{U}_{k_N} \\ \tilde{U}_{\mathrm{f}_N} \\ \tilde{Y}_{\mathrm{p}_N} \end{bmatrix} \end{aligned}$$

can be computed where $L_k^{(1)}, L_k^{(2)}$, and $L_k^{(3)}$ have dimension $\ell \times \ell$, and the subscript of $R$ indicates the indexes of the block row and column similar to the notation in Matlab. Similarly,

$$Z_{k+1} = \tilde{Y}_{\mathrm{f}_N}\Pi_\Phi = \begin{bmatrix} \tilde{U}_{\mathrm{p}_N} \\ \tilde{U}_{k_N} \\ \tilde{U}_{\mathrm{f}_N} \\ \tilde{Y}_{\mathrm{p}_N} \end{bmatrix} = R_{6:6,1:5} R^{-1}_{1:5,1:5} \begin{bmatrix} \tilde{U}_{\mathrm{p}_N} \\ \tilde{U}_{k_N} \\ \tilde{U}_{\mathrm{f}_N} \\ \tilde{Y}_{\mathrm{p}_N} \end{bmatrix} =: \begin{bmatrix} L_{k+1}^{(1)} & L_{k+1}^{(2)} & L_{k+1}^{(3)} \end{bmatrix} \begin{bmatrix} \tilde{U}_{\mathrm{p}_N} \\ \tilde{U}_{k_N} \\ \tilde{U}_{\mathrm{f}_N} \\ \tilde{Y}_{\mathrm{p}_N} \end{bmatrix},$$

where $L_{k+1}^{(1)}, L_{k+1}^{(2)}$ and $L_{k+1}^{(3)}$ have dimension $(\ell - 1) \times (\ell + 1)$.

In light of equation (46) in [80],

$$\Gamma_\ell \widehat{X}_\ell = Z_k - L_k^{(2)} \begin{bmatrix} \tilde{U}_{k_N} \\ \tilde{U}_{\mathrm{f}_N} \end{bmatrix} = M_\ell,$$

where $\widehat{X}_\ell$ is the estimated state vector. The matrix $\Gamma_\ell$ and the system order $n$ can be determined by the following SVD of $M$ with partition:

$$M_\ell = \begin{bmatrix} G_1 & G_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} V',$$

where $\Sigma_2$ consists of insignificant singular values, and

$$\Gamma_\ell \approx G_1 \Sigma_1^{1/2}.$$

The realization matrices $(A, C)$ can be estimated based on $\Gamma_\ell$. The other two realization matrices $(B, D)$ can be estimated via various ways, once the estimates of $(A, C)$ are available. We refer the readers to [80] for more details.

### 4.3.5 Simulation Studies

In this section, we present our simulation studies for the proposed EIV identification algorithm, applicable to unequal noise variances at the input and output. Because Matlab has the built-in function "n4sid", we compare the proposed algorithm using the TLS estimation with the N4SID algorithm in Example 4.3, 4.4, 4.6 and 4.7. We also compare the proposed algorithm using the modified TLS estimation with the GSA-Frisch algorithm in Example 4.5 and 4.8. The plant models in the literature [69] are used for Example 4.3 and 4.4 with the white input except at $b_0 = 0$ because physical systems are strictly causal, and the Matlab function "n4sid" defaults to $b_0$ as 0. Also, the plant model in [57] is used for Example 4.6 with the colored input. We assume that the past input and future inputs are well correlated not to make $\sum_{N=1}^{\infty} U_{fN} U_{pN}' = 0$. A total of 500 ensembles are used for the examples.

**White input**

**Example 4.3**: Consider transfer function

$$P(z) = \frac{0.8 \times (-1.2488z^{-1} + 0.9604z^{-2})}{1 - 1.4491z^{-1} + 0.9604z^{-2}}.$$

Identification of $P(z)$ is carried out with the TLS estimation in the proposed algorithm and with the N4SID algorithm. Figure 4.6 shows the RMSE plots with the solid line for the TLS with the proposed algorithm and dashed line for the N4SID algorithm. Noise variances

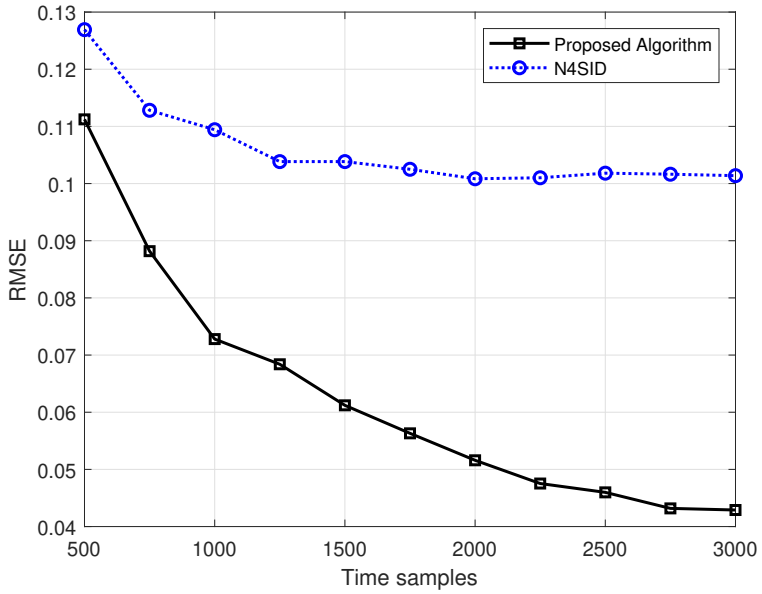


Figure 4.6. Proposed algorithm vs N4SID in Example 4.3

$\sigma_u^2 = 0.3^2$ and $\sigma_y^2 = 0.6^2$ are used. The SNR at the input and output is 10.2917dB and 15.4914dB, respectively. The RMSE curves in Figure 4.6 show that the TLS estimation in the proposed algorithm results in much smaller errors between the true and estimated system parameters.

**Example 4.4**: The plant model is described as

$$P(z) = \frac{-0.9z^{-1} - 0.45z^{-2}}{1 - 0.5z^{-1} + 0.3z^{-2}}.$$

Again, white inputs are used in this simulation study. The SNR at the plant input and output is 10.239dB and 6.7497dB, respectively, with the noise variances $\sigma_u^2 = 0.3^2$ and $\sigma_y^2 = 0.6^2$. Figure 4.7 shows the RMSE plots for the TLS and N4SID algorithm. In this example, the

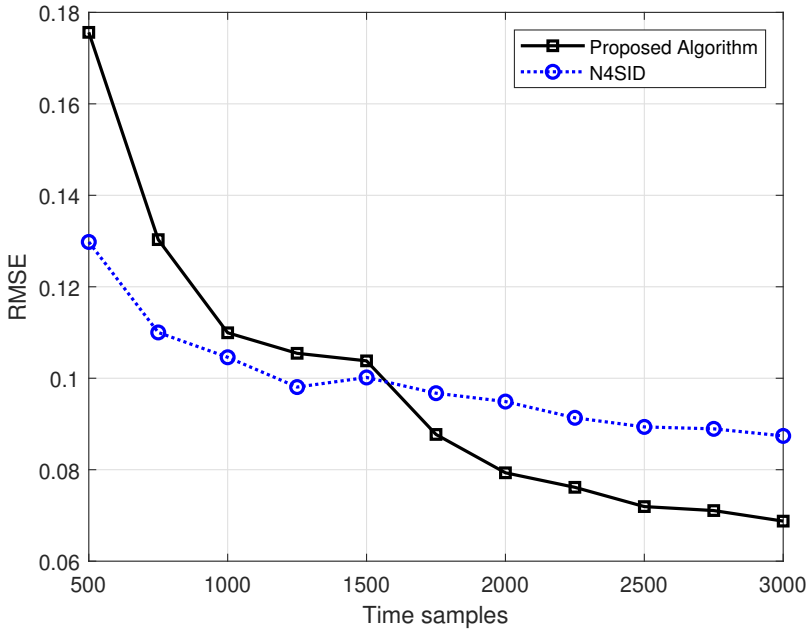


Figure 4.7. Proposed algorithm vs N4SID in Example 4.4

N4SID algorithm performs better than the TLS algorithm when $N < 1500$. However, the TLS algorithm performs better when $N > 1500$. Although the N4SID algorithm is not a consistent estimator for the EIV system and signal model, its performance for the small number of sample data is good for this example.

**Example 4.5**: The plant model is described as

$$P(z) = \frac{-0.9z^{-1} - 0.45z^{-2}}{1 - 0.5z^{-1} + 0.3z^{-2}}.$$

The input and output noise variances are $\sigma_u^2 = 0.3^2$ and $\sigma_y^2 = 0.6^2$, respectively. White inputs are used again in this simulation study. The SNR at the input is 10.239dB and 6.7497dB at the output. Figure 4.8 shows the RMSE plot with the new identification method in Example 5. Figure 4.9 shows the RMSE plot with the GSA-Frisch algorithm in Example 4.5. In the

case of $\ell = n$, the newly proposed identification algorithm performs worse than the GSA-Frisch algorithm. However, in both $\ell = n+1$ and $\ell = n+5$ cases, the new algorithm performs better than the GSA-Frisch algorithm. Moreover, the new algorithm is much simpler than the GSA-Frisch algorithm because it does not need the search procedure for estimating the ratio.

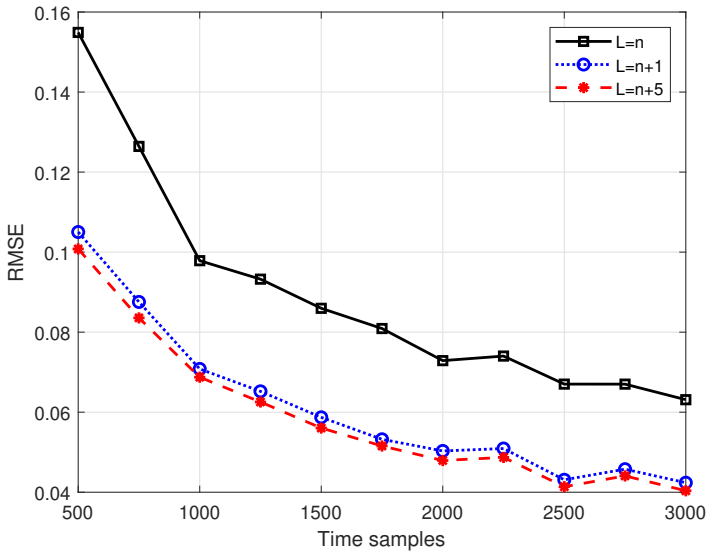


Figure 4.8. RMSE for EIV identification based on the modified TLS in Example 4.5

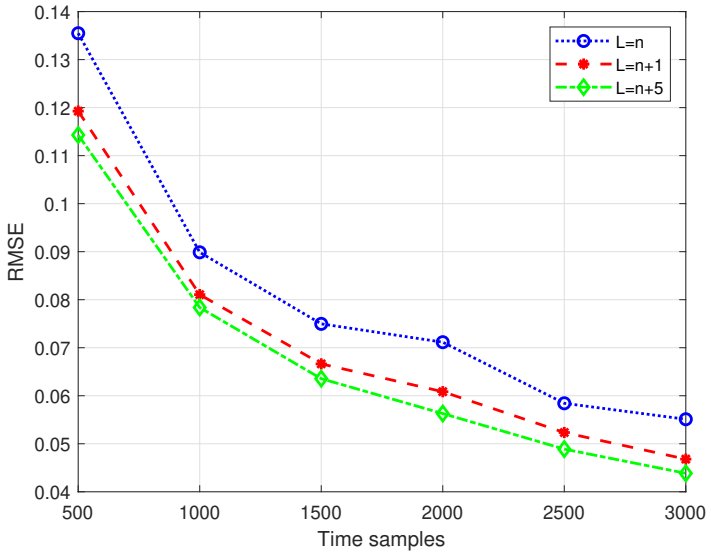


Figure 4.9. RMSE for EIV identification based on the GSA-Frisch algorithm in Example 4.5

**Colored input**

**Example 4.6**: The plant model is described as

$$P(z) = \frac{2z^{-1}}{1 - 0.8z^{-1}},$$
$$(1 - 0.5q^{-1})u(k) = (1 + 0.7q^{-1})u_o(k),$$

where $u_0(k)$ is the auxiliary input of the zero-mean white Gaussian process with variance 1. Hence $u(k)$ is not white. Figure 4.10 shows the RMSE when the equal noise variances of $\sigma_u^2 = 1$ and $\sigma_y^2 = 1$ are employed, which are the same as those in [57]. The SNR at the plant input is 4.4693dB and 19.4588dB at the output. It can clearly be seen that the proposed algorithm performs better than the N4SID algorithm.

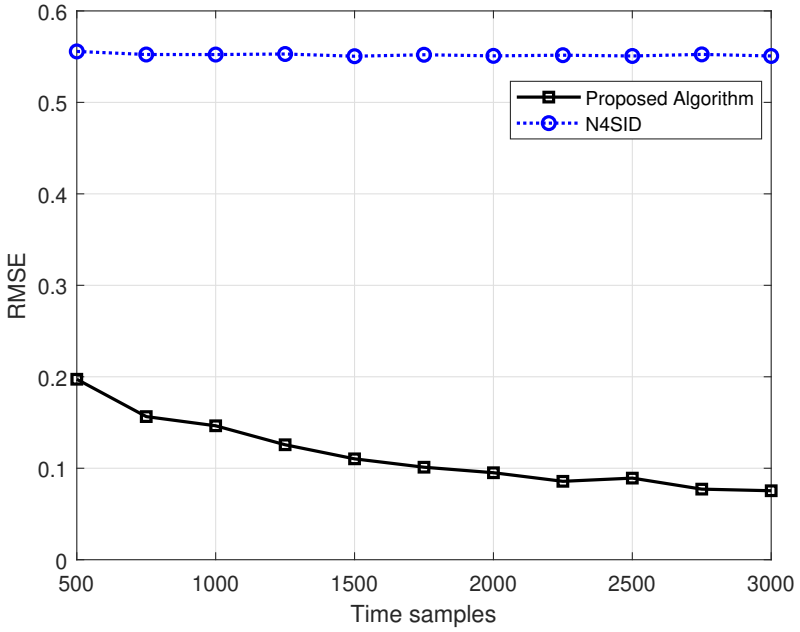


Figure 4.10. Proposed algorithm vs N4SID in Example 4.6

**Example 4.7**: In this example, we used the same plant model and the colored input as those in Example 4.6 but unequal noise variances $\sigma_u^2 = 0.3^2$ and $\sigma_y^2 = 0.6^2$. The SNR at the plant input and output is 14.9268dB and 23.8957dB, respectively. Figure 4.11 shows the RMSE in Example 4.7. Once again, the TLS estimation in the proposed algorithm performs

better than the N4SID algorithm.

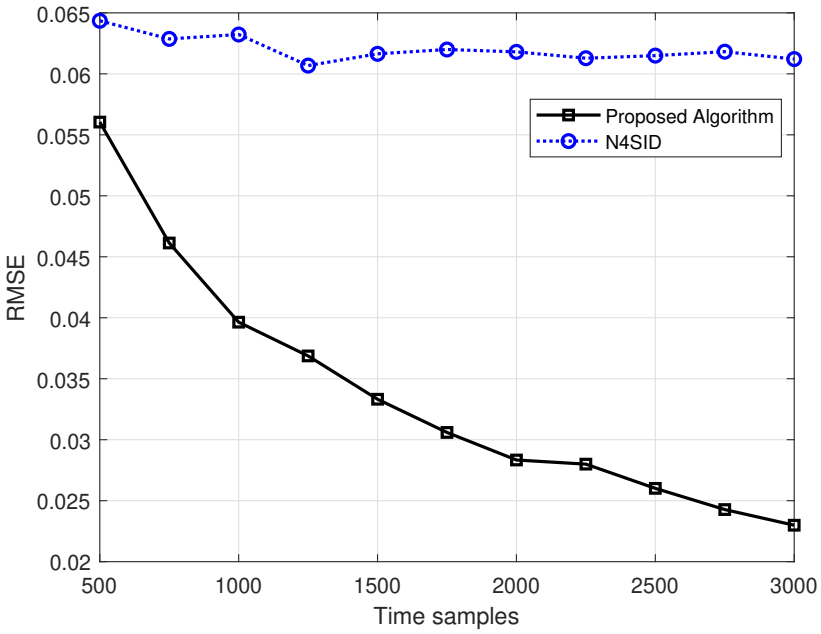


Figure 4.11. Proposed algorithm vs N4SID in Example 4.7

**Example 4.8**: In this example, the same plant model is used in Example 3.3. The colored input is again used. The noise variances at the plant input and output are $\sigma_u^2 = 0.3, \sigma_y^2 = 1$. The SNR at the input is 10.56dB and 10.5dB at the output. Figure 4.12 shows the RMSE plot with the modified TLS algorithm in the proposed identification method. When we compare it to the simulation results from the GSA-Frisch algorithm, which is Figure 4.5, the performance of the modified TLS algorithm in the proposed method overall is worse. The reason probably lies in the fact that the SNR in this example is higher than the SNR in Example 4.2. Although the modified TLS algorithm is not completely better than the GSA-Frisch algorithm in this example, it is still superior because of its low complexity, considering no estimating procedure is needed.

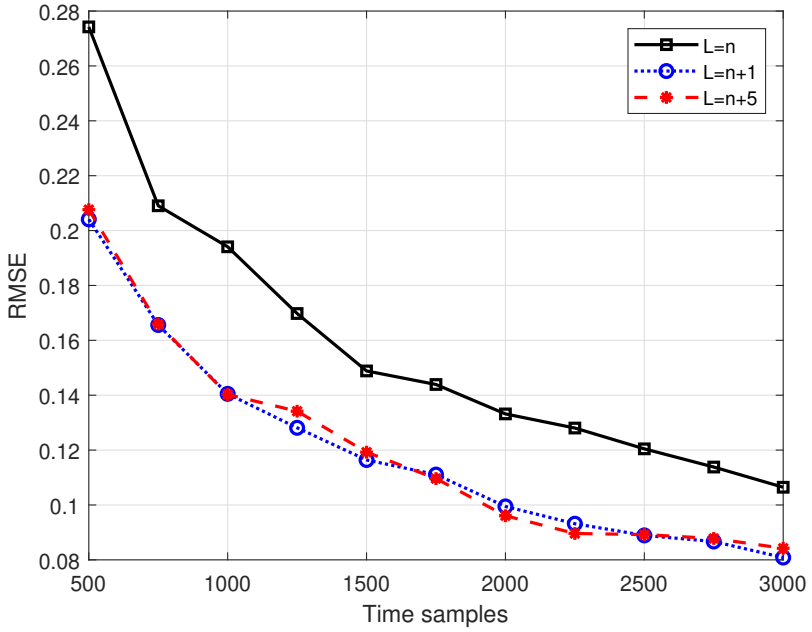


Figure 4.12. RMSE for EIV identification based on the modified TLS in Example 4.8

# CHAPTER 5
# CONCLUSION AND FUTURE WORK

## 5.1 Conclusion

In this dissertation, the EIV system identification has been studied. Chapter 2 provided some mathematical background and preparation for this study. A graph subspace method is developed in Chapter 3 to tackle the problem of EIV identification. Specifically, finite noisy samples of the system graph subspace are available by blocking the input and output measurement data into vectors of different length, and by computing the EVD of the respective second order statistics. Under the persistent excitation condition, the signal and noise subspaces of the sampled graph space can be obtained asymptotically. If the blocking size is minimum, then it recovers the TLS algorithm that is shown to produce an approximate MLE, and converges to the true MLE asymptotically. Compared to the TLS algorithm, the GSA improves the identification performance as the block size increases in the case of high SNR, and its iterative version improves the estimation accuracy in the case of low SNR. Compared to the MLE algorithm in [63], the GSA mitigates the NP hard and local minimum issues, while achieving the approximate and asymptotic MLE.

It is important to point out that the GSA assumes the same input and output noise variances or the known ratio of the two variances. The MLE algorithm developed in [63] has the same assumption. If such an assumption does not hold, then the Frisch scheme as studied in [57] can be employed to estimate first the noise variance at the input, then the noise variance at the output, and finally the system parameters iteratively. Alternatively, the iterative BELS algorithm in [97] can also be employed to estimate the noise variances and system parameters. While these two algorithms perform well in the case of high SNR, they do not work well in the low SNR environment.

So, in Chapter 4, we have studied the EIV identification problem for the case when the noise variance at the input differs from that at the output, and when the noise variances are

unknown. We proposed the GSA combined Frisch scheme for the unequal and unknown noise variances in Section 4.2. Specifically, a modified Frisch scheme was developed to estimate the ratio of the noise variances so that it can be applied to the GSA. Furthermore, we proposed identifying the EIV system directly without estimation of the noise variances or the ratio in Section 4.3. A TLS algorithm and a modified TLS algorithm are proposed, and shown to be asymptotically unbiased, and achieves the identifiability, if the input signal is persistently exciting, and the number of measurement data goes to infinity. Our identification result is validated with the simulation studies.

## 5.2 Future Work

In this dissertation, we proposed some efficient methods for the EIV system identification. However, there are still some problems we need to consider. In the following, we present three problems for future work.

1. Development of the EIV system identification without estimating the input and output variances

   • The use of Frisch scheme involves an iterative procedure and each iteration cycle involves high computational complexity that is the weakness of the Frisch scheme. On the other hand, Chapter 4 shows that there exist other approaches to deal with the unequal and unknown input-output noise variances. In fact, the identification performance for the algorithm developed in Section 4.3 exceeds the Frisch scheme for a class of input-output noises with much lower computational complexity. However, the estimation performance degenerates for other classes of input-output measurements. Hence, it becomes a challenge to developing effective estimation algorithms for EIV identification without estimating the two noise variances. A possible avenue in this research direction is to exploit the N4SID and modify it to solve the EIV system identification problem.

2. Development of the EIV system identification using the quantized input-output measurement data

   • Networked control systems have become popular for more than one decade now [4, 5]. For such control systems, the controller and plant situate in two different physical locations, and are connected via wireless communications. Because of the digital technology, the received input and output measurement data are likely to be quantized. How to estimate the system parameters using quantized measurement data becomes a very interesting and challenging problem. There are at least two theoretical issues here. The first one lies in the consistency of the TLS and GSA algorithms based on the quantized measurement data. That is, it remains unknown for the convergence of the estimated parameters to the true system parameters when the number of measurement data increases to infinity, due to the existence of the quantization error. The second issue is the model of the quantization error. In signal processing area, quantization errors are often treated as white noises with uniform distribution. If this were the case, then consistency would hold true, even if the input and output measurement data are quantized. However, there does not exist such a proof for the whiteness of the quantization error.

3. Development of the EIV system identification for Hammerstein-Wiener and Wiener-Hammerstein nonlinear systems

   • Almost all systems are nonlinear. How to identify nonlinear systems under the EIV framework is a very difficult problem. For instance, Hammerstein-Wiener and Wiener-Hammerstein systems are two of the most popular classes of parameterized nonlinear systems. Our simulation studies show that the identification errors associated with the TLS algorithm for these two classes of nonlinear systems indeed decrease as the number of measurement data increases. However, it remains unknown for the consistency of the TLS or GSA, when applied to these two classes of nonlinear systems. We suggest the

use of the nonlinear coprime factorization results [45, 30] to analyze the coprimeness of these two classes of parameterized nonlinear systems and to study the identifiability issue first, aided by numerical studies. Because the parameterizaton for these two classes of nonlinear systems is similar to that of the linear systems, we believe that the consistency holds.

# REFERENCES


[1] K. Abed-Meraim, P. Loubaton, and E. Moulines. A subspace algorithm for certain blind identification problems. *IEEE Trans. Inform. Theory*, 43:499–511, 1997.

[2] R.J. Adcock. Note on the methods of least squares. *The Analyst*, 4(6):183–184, 1877.

[3] R.J. Adcock. A problem in least squares. *The Analyst*, 5(2):53–54, 1878.

[4] P. Antsaklis and J. Baillieul. Special issue on networked control systems. *IEEE Transactions on Automatic Control*, 49, 2004.

[5] P. Antsaklis and J. Baillieul. Special issue on technology of networked control systems. *IEEE Transactions on Automatic Control*, 95, 2007.

[6] M. Aoki and PC. Yue. On a priori error estimate of some identification methods. *IEEE Transactions on Automatic Control*, AC-15:541–548, 1970.

[7] S. Beghelli, R.P. Guidorzi, and U. Soverini. The frisch scheme in dynamic system identification. *Automatica*, 26(1):171–176, 1990.

[8] S. Beghelli, R.P. Guidorzi, and U. Soverini. A new criterion in EIV identification and filtering applications. In *13th IFAC Symp. Syst. Identif.*, pages 1993–1998, Rotterdam, The Netherlands, Aug. 27-29, 2003.

[9] J. Cardoso and E. Moulines. Invariance of subspace based estimators. *IEEE Trans. Signal Processing*, 48:2495–2505, 2000.

[10] P. Castaldi and U. Soverini. Identification of errors-in-variables models and optimal output reconstruction. In *Proceedings of the 14th international symposium on mathematical theory of networks and systems (MTNS), Padova, Italy*, 1998.

[11] P. Castaldi, U. Soverini, and S. Beghelli. Identification of arx models in presence of input noise. In *Proceedings of the international conference on systems science*, Wroclaw, Poland, 1995.

[12] L.K. Chan and T.K. Mak. Maximum likelihood estimation in multivariate structural relationship. *Scandinavian Journal of Statistics*, 11(1):45–50, 1984.

[13] M. Chetoui, M. Thomassin, R. Malti, M. Aoun, S. Najar, M.N. Abdelkrim, and A. Oustaloup. New consistent methods for order and coefficient estimation of continuous-time errors-in-variables fractional models. *Computers and Mathematics with Applications*, 66(5):860–872, 2013.

[14] A.G. Costa Junior, J.B. Maldonado, F.A. Romero, J.C. Sanmartin, M. Valarezo, and H. Castillo. N4SID method applied to obtain a discrete-time linear state space system as a mathematical model of a jaw crusher prototype. In *CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies*, pages 1–6, October 18-20, 2017.

[15] B. De Moor. The singular value decompostion and long and short spaces of noisy matrices. *IEEE Trans Signal Process*, 41:2826–2838, 1993.

[16] B. De Moor and J. Vandewalle. Core problems in linear algebraic systems. *IEEE Trans Autom Control*, 35(5):563–566, 1990.

[17] U.B. Desai and S. Banerjee. Canonical correlation analysis, approximate covariance extension, and identification of stationary time series. *Automatica*, 22(5):593–597, 1986.

[18] R. Diversi, R. Guidorzi, and U. Soverini. Maximum likelihood identification of noisy input-output models. *Automatica*, 43:464–472, 2007.

[19] R. Diversi, R. Guidorzi, and U. Soverini. Frisch scheme–based algorithms for eiv identification. In *Proceedings of the 12th IEEE mediterranean conference on control and automation*, Kusadasi, Turkey, 2004.

[20] D. Fan and K. Lo. Recursive identification for dynamic linear systems from noisy input-output measurements. *Journal of Applied Mathematics*, 3(5):860–872, 2013.

[21] B.M. Finigan and I.H. Rowe. Strongly consistent parameter estimation by the introduction of strong instrumental variables. *IEEE Transactions on Automatic Control*, AC-19(6):825–830, 1974.

[22] R. Frisch. Statistical confluence analysis by means of complete regression systems. In *Technical Report 5, University of Oslo, Economics Institute, Oslo, Norway*, 1934.

[23] W.A. Fuller. Properties of some estimators for the errors-in-variables model. *The Annals of Statistics*, 8(2):407–422, 1980.

[24] P.P. Gallo. Consistency of regression estimates when some variables are subject to error. *Communications in Statistics: Theory and Methods*, 11:973–983, 1982.

[25] T.T. Georgiou and M.C. Smith. Optimal robustness in the gap metric. *IEEE Transactions on Automatic Control*, 35:673–686, 1990.

[26] L.J. Gleser. Estimation in a multivariate 'errors in variables' regression model: Large sample results. *The Annals of Statistics*, 9:24–44, 1981.

[27] G.H. Golub and C.F. Van Loan. An analysis of the total least squares problem. *SIAM J NUMER ANAL*, 17:883–893, 1980.

[28] G. Gu. *Discrete-Time Linear Systems – Theory and Design with Applications*. Springer, 2012.

[29] R. Guidorzi. Identification of multivariable processes in the frisch scheme context. In *Proceedings of the 12th international symposium on mathematical theory of networks and systems*, St Louis, USA, 1996.

[30] Z. Han and G. Chen. Dynamic right coprime factorization for nonlinear systems. In *Proceedings of the Second World Congress of Nonlinear Analysts*, volume 50, pages 3113–3120, December, 1997.

[31] M. Hong, T. Söderström, and W.X. Zheng. Accuracy analysis of bias-eliminating least squares estimates for errors-in-variables identification. *Automatica*, 43:1590–1596, 2007.

[32] L. Hongxia and C. Yuanli. N4SID based subspace identification method for normal temperature continuous transonic wind tunnel system. In *Control and Decision Conference*, pages 1394–1398, May 28-30, 2016.

[33] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, and 498–520, 1933.

[34] C. Hsiao, L. Wang, and Q. Wang. Estimation of nonlinear errors-in-variables models: an approximate solution. *Statistical Papers*, 38:1–25, 1997.

[35] L. Huwang, Y.H.S. Huang, and Y.H.T. Wang. Uniformly robust tests in errors-in-variables models. *Annals of the Institute of Statistical Mathematics*, 61:789–810, 2009.

[36] J. Kanti, T. Mardia, and J.M. Bibby. *Multivariate Analysis*. Academic Press, 1979.

[37] E. Karlsson, T. Söderström, and P. Stocia. Computing the cramer-rao lower bound for noisy input-output systems. *Signal Process.*, 80:2421–2447, 2000.

[38] T.R. Knapp. Canonical correlation analysis: A general parametric significance-testing system. *Psychological Bulletin*, 85(2):410–416, 1978.

[39] D. Kreiberg, T. Söderström, and F. Yang-Wallentin. Errors-in-variables system identification using structural equation modeling. *Automatica*, 66:218–230, 2016.

[40] MJ. Levin. Estimation of a system pulse transfer function in the presence of noise. *IEEE Transactions on Automatic Control*, AC-9:229–235, 1964.

[41] L. Ljung. *System Identification Theory for the User*. Prentice-Hall, 1999.

[42] I. Markovsky and S. Van Huffel. Overview of total least-squares methods. signal process. *Signal Process*, 87:2283– 2302, 2007.

[43] D.Q. Mayne. A method for estimating discrete time transfer functions. In *Advances in computer control*, second UKAC control convention, Bristol, UK, 1997.

[44] T. McKelvey. Identification of state-space models from time and frequency data. In *Ph.D. thesis, Linköping University*, 1995.

[45] J. B. Moore and L. Iricht. Coprime factorization over a class of nonlinear systems. In *1992 American Control Conference*, pages 3071–3075. IEEE, Chicago, IL, USA, 24-26 June 1992.

[46] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue. Subspace methods for the blind identification of multichannel fir filters. *IEEE Trans. Signal Processing*, 43:516–526, 1995.

[47] A. Musekiwa. Estimating the slope in the simple linear errors-in-variables model. In *Master thesis, Mathematical Statistics, Faculty of Science, University of Johannesburg*, 2005.

[48] E. Nowak. Identifiability in multivariate dynamic linear errors-in-variables models. *Journal of the American Statistical Association*, 87(419):713–723, 1992.

[49] C.C. Paige and Z. Strakoš. Scaled total least squares fundamentals. *Numer Math*, 91:117–146, 2002.

[50] C.C. Paige and Z. Strakoš. Core problems in linear algebraic systems. *SIAM J Matrix Anal Appl*, 27(3):861–875, 2006.

[51] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1930.

[52] O. Reiersøl. Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, 9:1–24, 1941.

[53] N.R. Sandell Jr and K.I. Yared. Maximum likelihood identification of state space models for linear systems. In *Technical report ESL-R-814, Electronic Systems Laboratory, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge*, 1978.

[54] I. Satake. *Linear Algebra*. Marcel Dekker, 1975.

[55] W. Scherrer and M. Deistler. A structure theory for linear dynamic errors-in-variables models. *SIAM Journal on Control and Optimization*, 36(6):2148–2175, 1998.

[56] T. Söderström. Identification of stochastic linear systems in presence of input noise. *Automatica*, 17:713–725, 1981.

[57] T. Söderström. Accuracy analysis of the Frisch scheme for identifying errors-in-variables systems. *IEEE Transactions on Automatic Control*, 52(6):985–997, 2007.

[58] T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43(6):939–958, 2007.

[59] T. Söderström. System identification for the errors-in-variables problem. *Trans Inst Meas Control*, 34(7):780–792, 2012.

[60] T. Söderström. *Errors-in-Variables Methods in System Identification*. Springer, 2018.

[61] T. Söderström. A frisch scheme for correlated output noise errors-in-variables identification. In *Proceedings of the 9th European control conference*, Kos, Greece, 2007.

[62] T. Söderström, K. Mahata, and U. Soverini. Identification of dynamic errors-in-variables model: approaches based on two-dimensional arma modelling of the data. *Automatica*, 39(5):929–935, 2003.

[63] T. Söderström and U. Soverini. Errors-in-variables identification using maximum likelihood estimation in the frequency domain. *Automatica*, 79(2):131–143, 2017.

[64] T. Söderström, U. Soverini, and K. Mahata. Perspectives on errors-invariables estimation for dynamic systems. *Signal Process.*, 82(8):1139–1154, 2002.

[65] T. Söderström and P. Stocia. *System Identification*. Prentice-Hall International, 1989.

[66] T. Söderström and P. Stoica. *Instrumental variable methods for system identification*. Springer, 1983.

[67] M.E. Solari. The "maximum likelihood solution" of the problem of estimating a linear functional relationship. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2):372–375, 1969.

[68] U. Soverini and T. Söderström. Frequency domain eiv identification: a frisch scheme approach. In *Proceedings of the 19th IFAC world congress*, Cape Town, South Africa, 2014.

[69] U. Soverini and T. Söderström. Frequency domain EIV identification combining the Frisch scheme and Yule-Walker equations. In *European Control Conference (Linz, Austria)*, pages 2038–2043, July 15-17, 2015.

[70] U. Soverini and T. Söderström. Identification methods of dynamic systems in presence of input noise. In *Proceedings of the 12th IFAC symposium on system identification*, Santa Barbara, California, 2000.

[71] P. Stocia and R. Moses. *Introduction to Spectral Analysis*. Prentice-Hall, 1997.

[72] P. Stoica, T. Söderström, and V. Simonyte. Study of a bias-free least squares parameter estimator. In *IEE Proc Control Theory Appl*, volume 142, page 1–6, September, 1995.

[73] S. Thil, M. Gilson, and H. Garnier. On instrumental variable-based methods for errors-in-variables model identification. In *Proceedings of the 17th IFAC world congress*, page 426–431, Seoul, Korea, 2008.

[74] L. Tong and S. Perreau. Multichannel blind identification: From subspace to maximum likelihood methods. *Proceedings of the IEEE*, 86(10):1951–1968, October 1998.

[75] L. Tong, G. Xu, and T. Kailath. Blind identification and equalization based on second-order statistics: A time domain approach. *IEEE Trans. Inform. Theory*, 40:340–349, 1994.

[76] J. Van Huffel, S. Vandewalle. *The total least squares problem: computation aspects and analysis*. SIAM, 1991.

[77] S. Van Huffel. Recent advances in total least squares techniques and error-in-variables modeling. In *SIAM*, 1997.

[78] S. Van Huffel and P.h. Lemmerling. Total least squares techniques and error-in-variables modeling. In *Analysis, Algorithms and Applications*, 2002.

[79] S. Van Huffel and J. Vandewalle. Comparison of total least squares and instrumental variable methods for parameter estimation of transfer function models. *Int J Control*, 50:1039–1056, 1989.

[80] P. Van Overschee and B. De Moor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1994.

[81] P. Van Overschee and B. De Moor. Closed-loop subspace system identification. In *Proceedings of the 36th Conference on Decision and Control*, pages 1848–1853, 1997.

[82] P. Van Overschee, B. De Moor, and J. Suykens. Subspace algorithms for system identification and stochastic realization. Technical report, ESAT-STSTA, 1990.

[83] B. Van Veen and S. Haykin. *Signals and Systems*. Wiley, 2003.

[84] M. Vidyasagar. Graph metric for unstable plants. *IEEE Transactions on Automatic Control*, 29:403–418, 1984.

[85] A. Wald. The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, 11(3):284–300, 1940.

[86] K.M. Wolter and W.A. Fuller. Estimation of nonlinear errors-in-variables models. *The Annals of Statistics*, 10(2):539–548, 1982.

[87] K.Y. Wong and E. Polak. Identification of linear discrete time systems using the instrumental variable approach. *IEEE Transactions on Automatic Control*, 12:707–718, 1967.

[88] P.C. Young. On aweighted steepest descentmethod of process parameter estimation. *Technical report, Engineering Laboratory, Cambridge University, United Kingdom*, 1965.

[89] P.C. Young. An instrumental variable method for real-time identification of a noisy process. *Automatica*, 6:271–287, 1970.

[90] E. Zhang and R. Pintelon. Identification of multivariable dynamic errors-in-variables system with arbitrary inputs. *Automatica*, 82:69–78, 2017.

[91] E. Zhang and R. Pintelon. Nonparametric identification of linear dynamic errors-in-variables systems. *Automatica*, 94:416–425, 2018.

[92] W.X. Zheng. Transfer function estimation form noisy input and output data. *Int J Adapt Control Signal Process*, 12:365–380, 1998.

[93] W.X. Zheng. A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, 47:1142–1147, 2002.

[94] W.X. Zheng. On least-squares identification of armax models. In *Proceedings of the 15th IFAC world congress*, volume 142, page 1–6, Barcelona, Spain, 2002.

[95] W.X. Zheng. An efficient algorithm for stochastic system identification with noisy input. In *Proceedings of the 38th IEEE conference on decision and control*, page 3657–3662, Phoenix, AZ, 1999.

[96] W.X. Zheng. Unbiased identification of stochastic linear systems from noisy input and output measurements. In *39th IEEE conference on decision and contrl*, Sydney, Australia, December, 2000.

[97] W.X. Zheng and C.B. Feng. Unbiased parameter estimation of linear systems in presence of input and output noise. *Int. J. Adapt. Control Signal Process*, 3:231–251, 1989.

# VITA

Hyundeok Kang was born in Seoul, South Korea. In 2007 he completed his Bachelor's in Electronic Engineering and in 2010 his Master's Degree in Information Engineering at Inha university, Incheon, South Korea. In the fall of 2014, he joined the Department of Electrical and Computer Engineering at Louisiana State University in the U.S.A. in pursuit of his graduate study. He obtained the Master's Degree of Science in Electrical Engineering in August 2016. He anticipates to be awarded the Doctor of Philosophy degree in Electrical Engineering in August 2019.