Frequency Domain Identification of Continuous-Time Systems
Reconstruction and Robustness

Jonas Gillberg

Department of Electrical Engineering
Linköpings universitet, SE–581 83 Linköping, Sweden
Linköping 2006
Frequency Domain Identification of Continuous-Time Systems Reconstruction and Robustness

© 2006 Jonas Gillberg

gillberg@isy.liu.se
www.control.isy.liu.se
Division of Automatic Control
Department of Electrical Engineering
Linköpings universitet
SE–581 83 Linköping
Sweden

ISBN 91-85523-34-8      ISSN 0345-7524

Printed by LiU-Tryck, Linköping, Sweden 2006
To Karl-Eric, Gösta & Märtha who are no longer with us....
Abstract

Approaching parameter estimation from the discrete-time domain is the dominating paradigm in system identification. Identification of continuous-time models on the other hand is motivated by the fact that modelling of physical systems often take place in continuous-time. For many practical applications there is also a genuine interest in the parameters connected to these physical models.

The most important element of time- and frequency-domain identification from sampled data is the discrete-time system, which is connected to the parameters of the underlying continuous-time system. For deterministic systems it governs the frequency response from the sampled input to the sampled output. In case of time series, it models the spectrum of the sampled output.

As the rate of sampling increase, the relationship between the discrete- and continuous-time parameters can become more or less ill-conditioned. Mainly, because the gathering of the poles of the discrete-time system around the value 1 in the complex plane will produce numerical difficulties while mapping back to the continuous-time parameters. We will therefore investigate robust alternatives to using the exact discrete-time system, which are based on more direct use of the continuous-time system. Another, maybe more important, reason for studying such approximations is that they will provide insight into how one can deal with non-uniformly sampled data.

An equally important issue in system identification is the effect of model choice. The user might not know a lot about the system to begin with. Often, the model will only capture a particular aspect of the data which the user is interested in. Deviations can, for instance, be due to mis-readings while taking measurements or un-modelled dynamics in the case of dynamical systems. They can also be caused by misunderstandings about the continuous-time signal that underlies sampled data. From a user perspective, it is important to be able to control how and to what extent these un-modelled aspects influence the quality of the intended model.

The classical way of reducing the effect of modelling errors in statistics, signal processing and identification in the time-domain is to introduce a robust norm into the criterion function of the method. The thesis contains results which quantify the effect of broad-band disturbances on the quality of frequency-domain parameter estimates. It also contains methods to reduce the effect of narrow-band disturbances or frequency-domain outliers on frequency-domain parameter estimates by means of methods from robust statistics.
Acknowledgments

“No man is an island, entire of itself; every man is a piece of the continent, a part of the main.”

John Donne (1572 - 1631)

First of all I would like to thank my supervisors Prof. Lennart Ljung and Prof. Fredrik Gustafsson for their guidance and support throughout my research. I have learned to appreciate this combination since the personality, research and age profiles of Lennart and Fredrik seem to complement each other very well.

Several other persons but myself and my supervisors have contributed to the making of this thesis. I would like to mention Frida Eng with whom I have had many interesting discussions about problems connected to non-uniform sampling. I am also deeply in depth to the local $\LaTeX$guru at Automatic Control, Gustav Hendeby, who have answered even my most imbecilic questions on typesetting. Then, I would also like to thank Daniel Ankelhed, Janne Harju, Markus Gerdin, Johanna Wallén, Henrik Olsson, Daniel Axehill, David Törnqvist, Henrik Tidefeldt and Gustav Hendeby for reading parts of the manuscript.

Most of all I would like to express my most sincere gratitude to Anna who have shown an extraordinary patience with me during my work. I would also like to thank my father Anders and my mother Lisbeth who have been my most loyal supporters, cheering me up all the time. I am also very grateful for knowing Lars-Erik and Christina, parents of Anna, who have almost provided me with a second home in Norrköping and Arkösund. I would also like to mentioned the families of Ragnar Wallin and Fredrik Tjärnström who have enriched my and Annas social life during this brief period of time. Of course I should not forget to thank Erik Geijer Lundin for taking me pike fishing by Norra Finnö and Stockholm archipelago, and for keeping me company during Annas telephone conferences and Stockholm course work.

During the spring and summer of 2005 I had the opportunity to pay a three month visit to the Department of Fundamental Electricity and Instrumentation at the Vrije Universiteit Brussel. I would therefor like to thank Prof. Dr. ir. Rik Pintelon and Prof. Dr. ir. Johan Schoukens for making this productive stay possible.

This work was sponsored by the graduate school ECSEL, SSF and the VINNOVA ISIS competence center.
Contents

1 Introduction .............................................. 1
   1.1 Problem Specification ............................... 4
   1.2 Outline ........................................... 5
   1.3 Contributions .................................... 6

2 Basic Theory .......................................... 9
   2.1 Introduction ...................................... 9
   2.2 Outline .......................................... 9
   2.3 Stochastic Processes ............................... 10
   2.4 Linear Input-Output Models .......................... 11
      2.4.1 Sampled Input-Output Models .................. 12
   2.5 Linear Time Series Models .......................... 14
      2.5.1 Sampled Time Series Models .................. 16
   2.6 Spectral Analysis ................................ 18
      2.6.1 Spectral Analysis of Input Output Models ....... 19
      2.6.2 Spectral Analysis of Time Series Models ........ 20
   2.7 System Identification .............................. 21
      2.7.1 Maximum Likelihood Estimation of Linear Models .. 21
      2.7.2 Kalman Filter Based System Identification ....... 25
      2.7.3 Frequency Domain System Identification ........ 31
      2.7.4 Cramer-Rao Lower Bound ........................ 41
   2.8 Summary ........................................... 42

3 A Short Review of Direct Continuous-Time System Identification Methods 43
   3.1 Introduction ...................................... 43
   3.2 Outline .......................................... 43
   3.3 Direct Approach ................................... 44
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>Modulating Functions Methods</td>
<td>45</td>
</tr>
<tr>
<td>3.5</td>
<td>Linear Filter Methods</td>
<td>47</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Poisson Moment Functionals</td>
<td>48</td>
</tr>
<tr>
<td>3.6</td>
<td>Integration Methods</td>
<td>49</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Orthogonal Functions</td>
<td>50</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Numerical Integration Methods</td>
<td>51</td>
</tr>
<tr>
<td>3.6.3</td>
<td>Linear Integral Filter Methods</td>
<td>51</td>
</tr>
<tr>
<td>3.7</td>
<td>Simplified Refined Instrumental Variable Method</td>
<td>52</td>
</tr>
<tr>
<td>3.8</td>
<td>Summary</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>Identification of Input-Output Models from Uniformly Sampled Data</td>
<td>55</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>55</td>
</tr>
<tr>
<td>4.2</td>
<td>Outline</td>
<td>56</td>
</tr>
<tr>
<td>4.3</td>
<td>Pulse Transfer Function</td>
<td>56</td>
</tr>
<tr>
<td>4.4</td>
<td>Truncating the Pulse Transfer Function</td>
<td>59</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Numerical Illustration</td>
<td>60</td>
</tr>
<tr>
<td>4.5</td>
<td>Approximating the Continuous-Time System</td>
<td>63</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Sampling Zeros and Euler-Frobenius Polynomials</td>
<td>64</td>
</tr>
<tr>
<td>4.6</td>
<td>Approximating the Roll-Off Behavior</td>
<td>65</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Numerical Illustration</td>
<td>67</td>
</tr>
<tr>
<td>4.7</td>
<td>Estimating the Continuous-Time Fourier Transform</td>
<td>67</td>
</tr>
<tr>
<td>4.7.1</td>
<td>Numerical Illustration</td>
<td>71</td>
</tr>
<tr>
<td>4.8</td>
<td>Comparisons</td>
<td>71</td>
</tr>
<tr>
<td>4.9</td>
<td>Summary</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>Identification of Input-Output Models From Non-Uniformly Sampled Data</td>
<td>77</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>77</td>
</tr>
<tr>
<td>5.2</td>
<td>Outline</td>
<td>78</td>
</tr>
<tr>
<td>5.3</td>
<td>Polynomial Interpolation</td>
<td>79</td>
</tr>
<tr>
<td>5.4</td>
<td>Spline Interpolation</td>
<td>80</td>
</tr>
<tr>
<td>5.5</td>
<td>Linear Filtering and Spline Functions</td>
<td>83</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Z-transform of a B-Spline</td>
<td>84</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Fourier Transform of a Spline</td>
<td>86</td>
</tr>
<tr>
<td>5.6</td>
<td>Fundamental Polynomial B-Spline Function</td>
<td>86</td>
</tr>
<tr>
<td>5.7</td>
<td>Non-Uniform Sampling and Splines</td>
<td>87</td>
</tr>
<tr>
<td>5.8</td>
<td>Numerical Examples</td>
<td>89</td>
</tr>
<tr>
<td>5.9</td>
<td>Summary</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>Interpolation and the Estimation of the Continuous-Time Power Spectrum</td>
<td>95</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>95</td>
</tr>
<tr>
<td>6.2</td>
<td>Outline</td>
<td>96</td>
</tr>
<tr>
<td>6.3</td>
<td>Estimation of Power Spectrum</td>
<td>97</td>
</tr>
<tr>
<td>6.4</td>
<td>Interpolation and Spectral Bias</td>
<td>99</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Bias Due to Interpolation</td>
<td>101</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Bias Due to Leakage</td>
<td>102</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Periodogram Bias</td>
<td>103</td>
</tr>
<tr>
<td>6.5 Uniform Sampling</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>6.6 Summary</td>
<td>106</td>
<td></td>
</tr>
</tbody>
</table>

### 7 Properties of Bias and Variance
- **7.1 Introduction** | 109
- **7.2 Outline** | 110
- **7.3 Bias Expression** | 111
- **7.4 Variance Expression** | 112
- **7.5 Practical Considerations for Frequency Selection**
  - **7.5.1 Minimizing the Variance** | 113
  - **7.5.2 Minimizing the Bias** | 114
- **7.6 Summary** | 118
- **7.7 Appendix A** | 119

### 8 Application to Estimation of Tire Pressure
- **8.1 Introduction** | 121
- **8.2 Outline** | 122
- **8.3 Tire Pressure Modelling** | 122
- **8.4 Problem Specifics and Objectives** | 123
- **8.5 Time and Frequency Domain Approaches** | 124
- **8.6 Frequency Domain Estimation** | 125
- **8.7 Properties of Bias and Variance** | 127
- **8.8 Experimental Results** | 128
- **8.9 Summary** | 130

### 9 Robust Frequency Domain ARMA Modelling
- **9.1 Introduction** | 131
- **9.2 Outline** | 132
- **9.3 Infinitesimal Approach to Robust Estimation**
  - **9.3.1 Estimators** | 132
  - **9.3.2 The Influence Function** | 133
  - **9.3.3 M-estimators of Scale** | 134
  - **9.3.4 The Gross-Error Sensitivity** | 136
  - **9.3.5 Optimally B-Robust Estimators** | 137
  - **9.3.6 Optimally B-Robust Estimators of Scale** | 138
- **9.4 Robust Frequency Domain ARMA Estimation** | 139
  - **9.4.1 Multivariable Influence Function** | 140
  - **9.4.2 Multivariable Optimally B-Robust Estimator** | 141
- **9.5 Numerical Example** | 143
- **9.6 Summary** | 143
- **9.7 Appendix A** | 145
- **9.8 Appendix B Proof of Theorem 9.2** | 146
- **9.9 Appendix C Proof of Theorem 9.3** | 147
10 Identification of Continuous-Time ARMA Models 151
  10.1 Introduction ..................................................... 151
  10.2 Outline .......................................................... 152
  10.3 Model and Representations ...................................... 153
  10.4 Truncating the Poisson Summation Formula ................. 153
    10.4.1 Numerical Example ...................................... 154
  10.5 Direct Estimation .............................................. 155
  10.6 Estimating the Continuous-Time Spectrum .................. 156
    10.6.1 Numerical Illustration .................................. 160
  10.7 Interpretations in Terms of Splines ......................... 162
  10.8 Reconstruction by Smoothing .................................. 165
  10.9 Reconstruction by Factorization ............................. 168
    10.9.1 Reconstruction via Innovations ......................... 170
    10.9.2 Spectral Estimation .................................... 170
    10.9.3 Approximation ........................................... 170
    10.9.4 Numerical Experiments ................................. 171
  10.10 Summary ........................................................ 172

11 Conclusions and Further Research 175

Bibliography 177

Index 185
Symbols, Operators and Functions

- $p$: differentiation operator
- $\Phi_c(i\omega)$: continuous-time power spectrum
- $\Phi_d(e^{i\omega T_s})$: discrete-time spectrum
- $\hat{\Phi}_c(i\omega)$: continuous-time periodogram
- $\hat{\Phi}_d(e^{i\omega T_s})$: discrete-time periodogram
- $\theta$: parameter vector
- $\theta_c$: parameters of continuous-time model
- $\theta_d$: parameters of discrete-time model
- $\theta_0$: true parameters vector
- $\hat{\theta}$: estimated parameter vector
- $u(t)$: continuous-time input signal
- $e(t)$: continuous-time white noise
- $y(t)$: continuous-time output signal
- $U_T(i\omega)$: truncated Fourier transform of input signal
- $E_T(i\omega)$: truncated Fourier transform of input signal
- $Y_T(i\omega)$: truncated Fourier transform of output signal
- $Y_d(e^{i\omega T_s})$: discrete-time Fourier transform of output signal
- $\hat{Y}(i\omega)$: estimate of the continuous-time Fourier transform by $k$th order interpolation
- $\delta(t)$: Dirac delta distribution
- $\delta_i$: Kronecker delta function
- $\mathcal{N}(0,R)$: multivariable normal distribution with covariance matrix $R$
- $\text{AsN}(0,R)$: asymptotic multivariable normal distribution with covariance matrix $R$
- $\mathcal{U}(a,b)$: uniform distribution on the interval $[a,b]$
- $\text{Exp}(\lambda)$: exponential distribution with parameter $\lambda$
- $\text{AsExp}(\lambda)$: asymptotic exponential distribution with parameter $\lambda$

Abbreviations

- CARMA: continuous-time autoregressive moving-average
- CAR: continuous-time autoregressive
- COE: continuous-time output error
- ZOH: zero-order hold
- FOH: first-order hold
“The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification for such a mathematical construct is solely and precisely that it is expected to work.”

John von Neumann (1903 - 1957)

Complex industrial products such as automobiles, aircrafts etc. integrate a large number of components of different physical nature such as mechanics, electronics, hydraulics and fluids. Take for instance an automotive engine as the one in Figure 1.1. The number of components within these devices is growing and their desired behavior is getting more and more complicated. Apart from the products themselves, the process of product development is becoming more demanding. Greater functionality and better quality are to be implemented in less time, with less resources and less environmental impact. An effective product development process is therefore becoming more of a necessary condition for market success than the sufficient one it used to be. This is even more apparent when the market gets globalized and the competition gets tougher every single day.

In this context it is easy to understand why mathematical modelling and simulation have grown from a technical novelty fifty years ago, into a crucial component of product design today. Through the increased competition and the un-parallelled development of computer power, mathematical modelling has earned an important and respected role in modern engineering. Today many industrial companies are even reluctant to create systems with a behavior that cannot be modelled and simulated in advance. Mostly because of considerations revolving around cost.

The use of models has also added value to science and engineering by facilitating new and improved functionality. Good models can provide frameworks that can be used to interpret and make sense of measurements collected from a particular system. This
Introduction

Figure 1.1: A car engine.

Material can then be refined in order to form the base for more accurate decisions and/or more appropriate actions in order to control the state or output of the system. If properly designed, models can also make it possible to extract as much information as possible from available data, which can be important when gathering data is exceedingly difficult and/or expensive.

Figure 1.2: A saturated steam heat exchanger (Bittanti and Piroddi, 1997).

Now, if models are just that important, where are they used in practice? The following example might illuminate the issue. A typical device that can be found in a private house or an apartment building is a saturated steam heat exchanger such as the one in Figure 1.2. This piece of apparatus is connected to the local district heating system at one end, while the other end is connected to the building water heating system. As a home owner using the exchanger to heat your house, it is important that the hot water flowing through
radiators etc. has the correct temperature. In this setup, control of the temperature $T$ is achieved by varying the rate of the water flow $q$ through the exchanger. This causal relationship is illustrated in Figure 1.3. In order to control the temperature properly we need to know how changing the rate of flow affects the temperature. The objective of the model is to capture this relationship mathematically as accurately as possible. When a mathematical model is available, a device that automatically controls the water temperature can be readily manufactured.

How do we then produce the models we need? Producing models is more or less, as John von Neumann so eloquently put it in the quote at the beginning of this chapter, the work of scientists. Models are ultimately the product of observations and they can basically be devised in two different ways: by the so-called systems approach or the analytical approach (von Bertalanffy, 1968). In the first approach the object of study, the system, is decomposed into subsystems which are again decomposed into subsystems themselves. This process of reduction must of course end at some level of detail where some first principle rules. This could be an assumption or an empirically established fact. The analytical approach is the method for determining the cause and effect relationship of a system without decomposing it further. This is done from observed data with very little regard to the internal structure of the system. The analytical approach can be said to be the essence of system identification.

In system identification the user has retrieved a set of input and output data, such as the one in Figure 1.4, which consists of a sequence of measurements of flow rate and temperature readings. The classical approach, e.g., (Ljung, 1999) to constructing a model, is to describe the input-output relationship as a difference equation

$$T(k+1) + a_1 T(k) = b_1 q(k) + e(k)$$

which relates the different measurements to each other. The so called model parameters $\theta = (a_1, b_1)$, are then determined by minimizing the difference between the temperature predicted by the model, $\hat{T}$, and the measured temperature $T$ provided by the measured flow rates $q$ as below

$$\hat{\theta} = \arg\min_\theta \sum_{k=1}^{N} (T(k) - \hat{T}(\theta, k))^2.$$  

This is a very successful approach with numerous practical applications.
1.1 Problem Specification

The model class in (1.1) is labelled discrete-time because it relates the measurements $T(k), T(k+1)$ and $q(k)$ at discrete-time instances to each other. The model does however not capture what happens in between measurements. Differential equations such as

$$\frac{dT(t)}{dt} + a_1 T(t) = b_1 q(t)$$

(1.2)

is on the other hand a class of models which describes the input-output relationship continuously in time. Often these types of models are constructed from first principles and it is possible to attach a physical meaning to the parameters. The parameters can for instance, in the case of the heat exchanger, be a heat transfer coefficient relating the rate of transfer of heat from the saturated steam to the water. For some reason it might be important for the user to know this value. Such a continuous-time model would then be more appropriate than a discrete-time one where there is no or little direct physical information. Approaching parameter estimation from the discrete-time domain is the dominating paradigm in system identification. In the black-box discrete-time modelling framework however, the identified parameters often lack a physical interpretation. The main effort in this thesis is therefore directed towards the identification of these continuous-time models and their parameters in particular.

Uniform sampling has also been a standard assumption. A single sensor delivering measurements at a constant rate has been considered as the ideal situation. With the advent of networked asynchronous sensors the validity of this assumption has however changed. In fields such as economics and finance, uniform sampling might not be practically possible. This indicates a need for methods coping with non-uniform sampling.

The pivotal element of time- and frequency-domain identification from sampled data is the discrete-time system, which is connected to the parameters of the underlying continuous-time system. For input-output models, it governs the frequency response from the
sampled input to the sampled output. In case of time series, it models the spectrum of the sampled output.

In some cases, this discrete-time entity might be flawed. For instance, as the rate of sampling increase, the relationship between the discrete- and continuous-time parameters can become more or less ill-conditioned. Mainly, because the gathering of the poles of the discrete-time system around the value 1 in the complex plane will produce numerical difficulties while mapping back to the continuous-time parameters. The ultimate reason for the behavior is, that very limited change can occur between sampling instances at high sampling rates. It is therefore it is worthwhile to investigate robust alternatives to using the exact discrete-time system, which are based on a more direct use of the continuous-time system. Another, maybe more important, reason for studying such approximations, is that they will provide insight into how one can deal with non-uniformly sampled data in frequency domain identification.

Another important issue in system identification is the effect of model choice. The user might not know a lot about the system to begin with. However, for one thing he or she can be sure that in reality, the data will not behave exactly according to the initial model choice. Often, the model will only capture a particular aspect of the data which the user is interested in. Deviations can, for instance, be due to miss-readings while taking measurements or un-modelled dynamics in the case of dynamical systems. They can also be caused by misunderstandings about the continuous-time signal that underlies sampled data. From a user perspective, it is important to be able to control how and to what extent these un-modelled aspects influence the quality of the intended model.

The classical way of reducing the effect of modelling errors in statistics, signal processing and identification in the time-domain is to introduce a robust norm into the criterion function of the method. This thesis contains results which quantify the effect of broad-band disturbances on the quality of frequency-domain parameter estimates. It also contains methods designed to reduce the effect of narrow-band disturbances or frequency domain outliers on frequency-domain parameter estimates by means of methods from robust statistics.

1.2 Outline

The thesis is structured as illustrated in Figure 1.5. First, in Chapter 2, there is an introduction to stochastic processes, linear systems and time series models together with time and frequency domain system identification and estimation methods. Then, in Chapter 3, there will be a brief review of the classical direct identification methods for continuous-time systems.

In Chapter 4, where the actual contributions begin, a series of approximations of the discrete-time pulse transfer function are derived. One of these methods can be interpreted as a way of estimating the continuous-time Fourier transform from its discrete-time counterpart. This idea is continued in Chapter 5, where it is proved to be equivalent to interpolation in terms of polynomial spline functions. The concept of interpolation is then further extended to the estimation of a continuous-time system from non-uniformly sampled output data.

In Chapter 6, builds on the findings in Chapter 4, and Chapter 5, the effect of inter-
polation of non-uniformly sampled data in continuous time series models is investigated.

The conclusion is that this is equivalent to creating an estimate of the continuous-time covariance kernel, which in turn leads to an estimate of the continuous-time spectrum.

The spectral estimate in Chapter 6 will have a bias that decreases as the sampling interval $T_s$ approaches zero. At high frequencies, this spectral bias can however produce parameter bias if the spectral estimate is used for parameter identification. The parameter estimation method will be that of Whittle, which was introduced in Chapter 2, and methods for analyzing the relationship between spectral and parameter bias for this approach are presented in Chapter 7.

In Chapter 8, the methods found in Chapter 6 and Chapter 7 are used in an application aimed at estimating tire pressure from anti-lock break system (ABS) data. This basically means that one has to estimate the resonance peak of a second-order continuous-time autoregressive model from non-uniformly sampled data. An important issue in this chapter is how to deal with broad band disturbances which can be found around the resonance peak of interest.

In Chapter 9, the issue of narrow band frequency domain disturbances is confronted. These objects, which acts like statistical outliers in the frequency domain when estimating time series models, are dealt with using methods from robust statistics. This theory is applied to the Whittle likelihood estimator, where the user is responsible for balancing the trade-off between bias and variance.

In Chapter 10 the perspective moves back from non-uniform sampling to uniform sampling and estimation of continuous time series models. Conclusions from Chapter 6 and Chapter 4 lead us to consider a method to estimate the continuous-time spectrum from the discrete-time spectrum. In Chapter 10, methods for the estimation of continuous time series models are also presented and the idea of interpolation of the output sequence is questioned. Instead a method based on spectral factorization by the Kalman filter is used to reconstruct uniformly uniform samples with the correct statistical properties from non-uniformly sampled time series data.

### 1.3 Contributions

Most of the material in this thesis can be found in a number of earlier published reports. The parts on equidistantly sampled CARMA models in Chapter 10 are also treated in

J. Gillberg and L. Ljung. Frequency-domain identification of continuous-time ARMA models from sampled data. In *Proceedings of 16th IFAC World Congress*, Prague, Czech Republic, July 2005a

The ideas of Chapters 4 and 5 pertaining to the identification of COE models can be found in

J. Gillberg and L. Ljung. Frequency-domain identification of continuous-time OE models from sampled data. In *Proceedings of 16th IFAC World Congress*, Prague, Czech Republic, July 2005b

J. Gillberg and L. Ljung. Frequency-domain identification of continuous-time output-error models from non-uniformly sampled data. In *Proceedings
Material related to the tire pressure identification problem in Chapter 8 is also presented in


Finally, the issues in Chapter 6 and Chapter 7 regarding interpolation, bias and variance can be found in


J. Gillberg, F. Gustafsson, and R. Pintelon. Robust frequency domain ARMA modelling. In Proceedings of 14th IFAC Symposium on System Identification, Newcastle, Australia, March 2006
Publications on convex optimization for robust control which, for the sake of uniformity of subject in this thesis, have been left out are.


and


Journal Papers in preparation are

“Each problem that I solved became a rule which served afterwards to solve other problems.”

Rene Descartes (1596-1650), Discours de la Methode

2.1 Introduction

Linear systems and stationary stochastic processes are the bricks and mortar of automatic control and system identification. The purpose of this chapter is therefore to introduce these tools, and also to serve as a dictionary to which the reader can return. It contains a plethora of basic definitions that will be used again and again later on. If the reader feels completely familiar with all of these basic concepts, it is up to him or her to freely move forward to the following chapter.

2.2 Outline

First, in Section 2.3, there will be a brief introduction to stochastic processes. Here concepts such as weak stationarity, Gaussian processes, covariance kernels and covariance functions will be introduced. Then, in Section 2.4 and 2.5, the basic concepts of continuous and discrete-time input-output models and time series models, such as state space representations and their solutions are explained briefly. In Section 2.6 the spectral properties such as power spectral densities of stochastic processes and the Fourier transforms of signals emanating from linear systems are explained. Here, the classical transfer function formalism is also introduced, both in discrete and continuous-time. In Section 2.7, the basic methods for system identification of discrete and continuous-time models are described. The exact maximum likelihood approach to system identification is introduced
together with its Kalman filter formulation. Finally, Section 2.7.3 and Section 2.7.3, contains methods for the discrete- and continuous-time frequency domain identification of time series models.

### 2.3 Stochastic Processes

A **stochastic process** is a set of stochastic variables separated by an index \( t \) and defined on the same probability space \( \Omega \) (Papoulis, 1965; Cramér and Leadbetter, 1976). The process is denoted

\[
\{ x(t), \ t \in T \}.
\]  

(2.1)

The value \( t \) we call time, and common sets of time are

\[
t = \ldots, -1, 0, 1, 2, \ldots \quad \text{(Discrete-time)} \tag{2.2}
\]

\[
- \infty < t < \infty \quad \text{(Continuous-time).} \tag{2.3}
\]

Note that the process \( x \) is actually a function of two variables, the events \( \omega \in \Omega \) and the time \( t \), and it would be more correct to write

\[
\{ x(\omega, t), \ \omega \in \Omega, \ t \in T \}.
\]  

(2.4)

The material in this thesis we will be constrained to the subclass of **Gaussian processes**. A process \( \{ x(t), \ t \in T \} \) is defined as Gaussian if every vector

\[
(x(t_1), x(t_2), \ldots, x(t_{N_t}))
\]  

(2.5)

has an \( N_t \)-dimensional Normal distribution for every choice of \( N_t \) and time sequence \( \{ t_k \}_{k=1}^{N_t} \). For such a process it is sufficient to know the first and second order moments

\[
m(t) = E x(t) \tag{2.6}
\]

\[
k(t, u) = E (x(t) - m(t)) (x(u) - m(u)), \tag{2.7}
\]

where \( k(t, u) \) is known as the **covariance kernel**. If we define

\[
x = \begin{pmatrix} x(t_1) \\ \vdots \\ x(t_{N_t}) \end{pmatrix}, \quad \mu = \begin{pmatrix} m(t_1) \\ \vdots \\ m(t_{N_t}) \end{pmatrix}, \quad C = \begin{pmatrix} k(t_1, t_1) & \cdots & k(t_1, t_{N_t}) \\ \vdots & \ddots & \vdots \\ k(t_{N_t}, t_1) & \cdots & k(t_{N_t}, t_{N_t}) \end{pmatrix}
\]  

(2.8)

then the probability density function for this process will be

\[
f_{t_1, \ldots, t_{N_t}}(x) = \frac{1}{(2\pi)^{N_t/2} \det C^{1/2}} e^{-\frac{1}{2} (x-\mu)^T C^{-1} (x-\mu)}.
\]  

(2.9)

Despite being a subset of stochastic processes, Gaussian processes form a large class. In this text, we will almost entirely limit ourselves the subclass of **Gaussian stationary**
stochastic processes or stationary processes. A process such as the one in (2.1) is called stationary if every vector
\[(x(t_1 + \tau), x(t_2 + \tau), \ldots, x(t_N + \tau))\] (2.10)
has the same probability distribution as
\[(x(t_1), x(t_2), \ldots, x(t_n)).\] (2.11)
The property of stationarity seems reasonable. In the theory of stochastic processes one is interested in the statistical relationship between the values of the process at different time instances.

If this relation would also depend on the particular time instance, it would be much more complicated to use and we would also have much less information to our disposal while making predictions about the behavior of the process.

In the case of Gaussian processes a sufficient condition for stationarity is that
\[k(t + \tau, t) = r(\tau)\] (2.12)
for some function \(r(\tau)\). This property is known as weak stationarity for general stochastic processes, and \(r(\tau)\) is labelled the covariance function. In the following section we will in particular study the stationary processes that are the output of a linear dynamical systems whose input is a white Gaussian process.

2.4 Linear Input-Output Models

However, before we turn our attention to linear-dynamical stochastic systems, we will move our focus towards dynamical systems where the input is a deterministic mathematical function instead of a stochastic process. The reason for this is that parts of the theory of input-output systems then easily carries over to the somewhat more complicated time-series case.

There are basically two different ways of describing input-output models; by means of state variables or by input-output relations. In this section we will focus on the fist, leaving the input-output treatment to Section 2.6. In the state-space case, the basic form for linear systems is the linear state equations written as
\[
\begin{aligned}
\dot{x}(t) &= A(t)x(t) + B(t)u(t) \\
y(t) &= C(t)x(t) + D(t)u(t).
\end{aligned}
\] (2.13)
The \(n \times 1\) vector function of time \(x(t)\) is called the state vector, and its components are the state variables or simply states. The input signal is the \(m \times 1\) function \(u(t)\) and \(y(t)\) is the \(p \times 1\) function called the output signal (Rugh, 1996; Kailath, 1980). If the matrices \(A(t)\) \((n \times n)\), \(B(t)\) \((n \times m)\), \(C(t)\) \((p \times n)\) and \(D(t)\) \((p \times m)\) are constant, the system is called time-invariant, otherwise it is called time-varying.

The solution to (2.13) is well known and can be written as
\[
\begin{aligned}
x(t) &= \Phi(t, 0)x(0) + \int_0^t \Phi(t, \tau)u(\tau)d\tau \\
y(t) &= C(t)x(t) + D(t)u(t)
\end{aligned}
\] (2.14)
where \( \Phi(t, s) \) is the so-called transition matrix. Generally, the transition matrix is a complicated entity to compute analytically, but there are important special cases where an analytical expression will exist. Such a case is when the system is time-invariant, where the transition matrix is defined as

\[
\Phi(t, \tau) = \begin{cases} 
  e^{A(t-\tau)B} & t \leq \tau \\
  0 & \tau > t.
\end{cases}
\] (2.15)

This class of systems has been studied extensively and we therefore refer the reader to the books by Kailath (1980) and Rugh (1996) for a more in-depth exposition of the topic.

If we assume that the system in (2.13) is initially at rest, i.e., \( x(0) = 0 \), the expression in (2.14) can be reformulated into

\[
y(t) = \int_0^t g(t-\tau)u(\tau)d\tau,
\] (2.16)

where

\[
g(t) = Ce^{At}B
\] (2.17)

is known as the continuous-time impulse response. This name originates from the fact that if \( u(t) \) is a unit impulse, i.e., \( u(t) = \delta(t) \) and \( y(0) \) then

\[
y(t) = g(t).
\] (2.18)

All the equations above rely on the use of continuous functions. In practice, we cannot observe quantities such as \( y \) and \( u \) continuously, and we have to rely on a finite number of samples taken at different time instances. In the following subsection we will deal with sampled or discrete-time systems, which relate these finite measurements to each other using the properties of the underlying continuous-time system.

### 2.4.1 Sampled Input-Output Models

A common assumption in automatic control and signal processing is that the input of the continuous-time system is generated from the finite sequence \( u(t_k) \) using a zero-order hold (ZOH) circuit such that

\[
u(t) = \sum_{l=0}^{\infty} u(lT_s) \left( H(t-lT_s) - H(t-(l+1)T_s) \right)
\] (2.19)

where \( H \) is the Heaviside step function

\[
H(t) = \begin{cases} 
  0 & t \leq 0 \\
  1 & 0 < t.
\end{cases}
\] (2.20)
2.4 Linear Input-Output Models

This will yield an input signal $u(t)$ which is piecewise constant and the state-space representation of the deterministic sampled-data model for this input will be

$$
\begin{align*}
x(t_{k+1}) &= F(k)x(t_k) + \sum_{t_k \leq lT_s < t_{k+1}} G(t_k, lT_s) u(lT_s) \\
y(t_k) &= Cx(t_k) + Du(t_k)
\end{align*}
$$

(2.21)

where $F(t_k)$ and $G(t_k, lT_s)$ are

$$
F(t_k) = e^{A(t_{k+1} - t_k)} \min(t_{k+1}, (l+1)T_s) \\
G(t_k, lT_s) = \int_{lT_s}^{t_k} e^{A(t_{k+1} - \tau)} B d\tau.
$$

(2.22)

(2.23)

This relation can also be observed from an impulse response perspective, where the input and output expressions will be

$$
y(t_k) = \int_0^{t_k} g(t_k - \tau) \sum_{l=0}^{\infty} u(lT_s) (H(\tau - lT_s) - H(\tau - (l+1)T_s)) d\tau
$$

(2.24)

$$
= \sum_{l} \int_0^{lT_s} g(t_k - \tau) (H(\tau - lT_s) - H(\tau - (l+1)T_s)) d\tau u(lT_s)
$$

(2.25)

$$
= \sum_{l=0}^{\infty} \int_{t_k - lT_s}^{t_k} g(\tau) d\tau u(lT_s).
$$

(2.26)

In a more compact form these equations can then be written as

$$
y(t_k) = \sum_{l=0}^{\infty} g_d(k, l) u(lT_s)
$$

(2.27)

where

$$
g_d(k, l) = \int_{lT_s}^{\min(t_k, (l+1)T_s)} g(t_k - \tau) d\tau.
$$

(2.28)

The resulting system is therefore discrete-time and time-varying. If $t_k = kT_s$ on the other hand, the formulas above will simplify and the resulting systems will once again be time-invariant such that (2.27) will turn into

$$
y(kT_s) = \sum_{l=0}^{\infty} g_d((k - l)T_s) u(lT_s)
$$

(2.29)

where

$$
g_d(mT_s) = \int_0^{T_s} g(mT_s + \tau) d\tau.
$$

(2.30)

is known as the discrete-time impulse response.
2.5 Linear Time Series Models

In the case of input-output models both the input and the output are known functions. This means that we know what is the cause of our output. Sometimes, we might not know our input or we may not even know what it is. But, we can on the other hand be interested in how different measurements are related to each other in order to make educated predictions of the future output of this system. A model class which is capable of this in both continuous and discrete time are the *time series models* (Brockwell and Davis, 1987).

Linear continuous-time stochastic systems or time series are also known as linear *stochastic differential equation* (SDE) which in the time-invariant case are represented as

\[
\begin{align*}
\dot{x}(t) &= Ax(t)dt + BdW(t) \\
y(t) &= Cx(t).
\end{align*}
\]

(2.31)

Here \(A\) is a real valued \(n \times n\) matrix, \(B\) is an \(n \times m\) matrix and \(C\) is \(p \times n\). The \(n \times 1\) vector \(x(t)\) is the vector of states.

The representation in (2.31) might be unfamiliar to some readers and some may ask why we just don’t write

\[
\dot{x}(t) = Ax(t) + B\dot{w}(t).
\]

(2.32)

Based on many situations, for example in engineering, one is led to assume that \(\dot{w}(t)\) in (2.32) has approximately the following properties:

1. if \(t_1 \neq t_2\) then \(\dot{w}(t_1)\) and \(\dot{w}(t_2)\) are independent
2. \(\dot{w}(t)\) is stationary
3. \(\mathbb{E}\dot{w}(t) = 0\) \(\forall t.\)

The really disturbing problem with this, is that there does not exist any mathematically sensible stochastic process satisfying (1) and (2) above (Oksendal, 1998). Let us therefore consider a discrete form of (2.31)

\[
x(t_{k+1}) - x(t_k) = Ax(t_k)\Delta t_k + B\dot{w}(t_k)\Delta t_k
\]

(2.33)

where \(\Delta t_k = t_{k+1} - t_k\). If, we then define

\[
\dot{w}(t_k)\Delta t_k = W(t_{k+1}) - W(t_k) = \Delta W(t_k)
\]

(2.34)

where \(\dot{w}(t_k)\) would satisfy the conditions (1), (2) and (3), then the process \(W\) must be identical to the classical Wiener process

1. \(W(0) = 0\)
2. \(W(t_4) - W(t_3)\) and \(W(t_2) - W(t_1)\) are independent if \(0 \leq t_1 \leq t_2 \leq t_3 \leq t_4\)
3. \(W(t + s) - W(s) \sim \mathcal{N}(0, \sigma \sqrt{t})\) if \(0 \leq s < s + t.\)
From (2.33) one can then write

\[ x(t_k) = x(0) + \sum_{j=0}^{k-1} A x(t_l) \Delta t_l + \sum_{j=0}^{k-1} B \Delta W(t_k), \]  

(2.35)

and if the limit as \( \Delta t_l \to 0 \) would exist, the usual integration notation would give the integral equation

\[ x(t) = x(0) + \int_0^t A x(s) ds + \int_0^t B dW(s). \]  

(2.36)

The important issue is hence, the properties of integrals like

\[ \int_0^t f(s) dW(s) \]  

(2.37)

and this is the contribution of the fundamental work done by Itô (1951). We will not dwell too much over these technicalities in the rest of the text, but will instead refer the reader to the excellent books by Øksendal (1998) and Arnold (1974) on the topic. The solution to the equation in (2.31) is in this context defined by

\[
\begin{cases}
  x(t) &= e^{A t} x(0) + \int_0^t e^{A(t-\tau)} B dW(\tau) \\
  y(t) &= C x(t),
\end{cases}
\]  

(2.38)

and \( x(t) \) will be a zero mean Gaussian process with the first and second order moments \( m(t) = E x(t) \) and \( P_x(t) = E x(t) x^T(t) \) such that

\[
\begin{cases}
  \dot{m}(t) &= A m(t) \\
  \dot{P}_x(t) &= A P_x(t) + P_x(t) A^T + \sigma^2 B B^T.
\end{cases}
\]  

(2.39)

where \( \sigma \) is the variance of \( w \). The unique solutions to these equations are

\[
\begin{align*}
  m_x(t) &= e^{A t} m_x(0) \\
  P_x(t) &= e^{A t} P_x(0) e^{A^T t} + \sigma^2 \int_0^t e^{A(t-s)} B B^T e^{A^T(t-s)} ds,
\end{align*}
\]  

(2.40)

and the covariance between the stochastic states at time \( t + \tau \) and \( t \) is defined by

\[ k_x(t + \tau, t) = E x(t + \tau) x(t) = e^{A \tau} P_x(t) \quad \tau > 0. \]  

(2.41)

If the initial state covariance \( P_x(0) \) is equal to a \( P_x \) which is the solution of the Lyapunov equation

\[ 0 = A P_x + P_x A^T + \sigma^2 B B^T, \]  

(2.42)
then, the process \( x(t) \) will be stationary. Otherwise \( P_x(t) \rightarrow P_x \) as \( t \rightarrow \infty \) if the eigenvalues of \( A \) lie in the left hand part of the complex plane (Hannan, 1970; Doob, 1953). Stationarity then means that the covariance kernel in (2.41) will turn into the covariance function

\[
r_x(\tau) = e^{A\tau}P_x, \quad \tau > 0,
\]
and that the covariance function of the output process will be

\[
r_y(\tau) = Ce^{A\tau}P_xC^T, \quad \tau > 0.
\]

As was said previously in Section 2.4, continuous-time measurements are not quite realistic and therefore, in the following section, we will consider sampled versions the continuous time series models.

### 2.5.1 Sampled Time Series Models

The values of the state variables and the outputs of the stochastic differential equation in (2.31) at discrete time instances \( t_k \) will be related through the stochastic difference equations (Åström, 1970)

\[
\begin{align*}
x(t_{k+1}) &= F(\Delta t_k)x(t_k) + v(t_k) \\
y(t_k) &= Cx(t_k)
\end{align*}
\]

where the \( n \times n \) transition matrix \( F \) is defined by

\[
F(\Delta t_k) = e^{A\Delta t_k}
\]

and the \( n \times 1 \) Gaussian noise vector is defined by

\[
v(t_k) = \int_{t_k}^{t_{k+1}} e^{A(t_{k+1} - s)}BdW(s).
\]

This means that

\[
v(t_k) \sim \mathcal{N}(0, R(\Delta t_k))
\]

where

\[
R(\Delta t_k) = \sigma^2 \int_0^{\Delta t_k} e^{As}BB^T e^{A^Ts}ds
\]

and \( \Delta t_k = t_{k+1} - t_k \). It is surprising to see that there are in fact \( n \) input components of \( v \) present in the sampled discrete-time system in (2.45) while there is only one input present in the antecedent continuous-time system in (2.31). In the following example the sampling of a simple second order continuous-time stochastic system will illustrate this phenomenon.
Example 2.1
Assume that
\[
A = \begin{pmatrix} -a_1 & -a_2 \\ 1 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad C = \begin{pmatrix} 0 \\ 1 \end{pmatrix}^T.
\] (2.50)

Then
\[
F(\Delta t_k) = e^{A\Delta t_k} = \begin{pmatrix} F_{11}(\Delta t_k) & F_{12}(\Delta t_k) \\ F_{21}(\Delta t_k) & F_{22}(\Delta t_k) \end{pmatrix}
\]

where
\[
F_{11}(\Delta t_k) = \frac{r_1 e^{r_1 \Delta t_k} - r_2 e^{r_2 \Delta t_k}}{r_1 - r_2} \tag{2.51}
\]
\[
F_{12}(\Delta t_k) = \frac{-a_2 e^{r_1 \Delta t_k} + a_2 e^{r_2 \Delta t_k}}{r_1 - r_2} \tag{2.52}
\]
\[
F_{21}(\Delta t_k) = \frac{e^{r_1 \Delta t_k} - e^{r_2 \Delta t_k}}{r_1 - r_2} \tag{2.53}
\]
\[
F_{22}(\Delta t_k) = \frac{(r_1 + a_1) e^{r_1 \Delta t_k} - (r_2 + a_1) e^{r_2 \Delta t_k}}{r_1 - r_2}. \tag{2.54}
\]

We will also have
\[
R(\Delta t_k) = \begin{pmatrix} R_{11}(\Delta t_k) & R_{12}(\Delta t_k) \\ R_{21}(\Delta t_k) & R_{22}(\Delta t_k) \end{pmatrix}
\]

where
\[
R_{11}(\Delta t_k) = \sigma^2 \frac{\frac{1}{2} e^{2r_1 \Delta t_k} - 1 - \frac{r_1 r_2}{r_1 + r_2} \left( e^{(r_1 + r_2) \Delta t_k} - 1 \right) + \frac{r_1^2}{2r_2} \left( e^{2r_2 \Delta t_k} - 1 \right)}{(r_1 - r_2)^2} \tag{2.55}
\]
\[
R_{12}(\Delta t_k) = \sigma^2 \frac{\frac{1}{2} \left( e^{2r_1 \Delta t_k} - 1 \right) - \left( e^{(r_1 + r_2) \Delta t_k} - 1 \right) + \frac{1}{2} \left( e^{2r_2 \Delta t_k} - 1 \right)}{(r_1 - r_2)^2} \tag{2.56}
\]
\[
R_{21}(\Delta t_k) = \sigma^2 \frac{\frac{1}{2} \left( e^{2r_1 \Delta t_k} - 1 \right) - \left( e^{(r_1 + r_2) \Delta t_k} - 1 \right) + \frac{1}{2} \left( e^{2r_2 \Delta t_k} - 1 \right)}{(r_1 - r_2)^2} \tag{2.57}
\]
\[
R_{22}(\Delta t_k) = \sigma^2 \frac{\frac{1}{2} e^{2r_1 \Delta t_k} - 1 - \frac{2r_1 r_2}{r_1 + r_2} \left( e^{(r_1 + r_2) \Delta t_k} - 1 \right) + \frac{1}{2} \left( e^{2r_2 \Delta t_k} - 1 \right)}{(r_1 - r_2)^2} \tag{2.58}
\]

Everywhere
\[
r_1 = \frac{-a_1 + \sqrt{a_1^2 - 4a_2}}{2}
\]
\[
r_2 = \frac{-a_1 - \sqrt{a_1^2 - 4a_2}}{2}.
\]
2.6 Spectral Analysis

It is frequently a fruitful prospect to study the spectral properties of functions and stochastic processes. In this section, we will briefly introduce concepts such as the Laplace transform, the Fourier transform and the $z$-transform. These objects have been treated extensively in both new and old literature and are among the theoretical equivalents of musical evergreens. As such, they are “played” repeatedly and therefore, we only state a few properties. For the interested reader, we refer to the books by for instance Gasquet and Witomski (1998) or Bracewell (2000) for a surely deeper and richer treatment of the subject.

First, we start with the Laplace transformation. Assume that $f(t)$ is piecewise continuous, apart from finitely many impulses, then the Laplace transform is defined as

$$F(s) = \int_{-\infty}^{\infty} f(t) e^{-st} dt, \quad (2.55)$$

with the inversion formula

$$f(t) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} F(s) e^{st} ds, \quad a \geq \alpha, \quad (2.56)$$

if $f(t)$ does not contain any impulses and is differentiable at $t$. If we only consider points on the imaginary axis of the Laplace transform we will get the relations of the celebrated continuous-time Fourier Transform

$$F(i\omega) = F(s)|_{s=i\omega} = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt, \quad (2.57)$$

with the inversion formula

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\omega) e^{i\omega t} d\omega. \quad (2.58)$$

The common “denominator” for these transforms is that they decompose the function into a possibly infinite set of fundamental functions. Exponentially damped oscillations $e^{st}$ in the case of the Laplace transform and periodic oscillations $e^{i\omega t}$ in the case of the Fourier transform.

For discrete time sequences or functions $f(n)$, there exist decompositions similar to the continuous-time Laplace and Fourier transform. The discrete-time cousin to the Laplace transform is the $z$-transform which is defined by

$$F(z) = \sum_{n=0}^{\infty} f(n)z^{-n}, \quad (2.59)$$
with the inversion formula
\[
f(n) = \frac{1}{2\pi i} \int_{|z|=r} F(z)z^{n-1}dz. \tag{2.60}
\]
If the z-transform is evaluated along the unit circle \(z = e^{i\omega T_s}\) in the complex plane, the result will be the time-discrete Fourier transform (TDFT)
\[
F_{T_s}(e^{i\omega T_s}) = T_s F(z)|_{z=e^{i\omega T_s}} = T_s \sum_{k=-\infty}^{\infty} f(kT_s)e^{-i\omega T_sk}, \tag{2.61}
\]
which has the inversion formula
\[
f(nT_s) = \frac{1}{2\pi} \int_{-T_s}^{T_s} F_{T_s}(e^{i\omega T_s})e^{i\omega nT_s}d\omega. \tag{2.62}
\]
If the sequence \(f(nT_s)\) originates from a continuous-time signal, such that
\[
f(nT_s) = f(t)|_{t=nT_s}, \tag{2.63}
\]
then the relationship between the continuous and time-discrete Fourier transform is characterized by the well known Poisson summation formula, see e.g. Gasquet and Witomski (1998), which states that
\[
F_{T_s}(e^{i\omega T_s}) = \sum_{k=-\infty}^{\infty} F\left(i\omega + \frac{2\pi T_s}{k}\right). \tag{2.64}
\]
In the following subsection, these concepts will be applied to functions and stochastic processes connected to linear time-invariant systems.

### 2.6.1 Spectral Analysis of Input Output Models

From (2.16) it is known that the solution of a linear, time-invariant, deterministic and causal system can be described by its impulse response \(g(t)\) through the following convolution integral
\[
y(t) = \int_{0}^{t} g(t - \tau)u(t) = \int_{0}^{t} g(\tau)u(t - \tau)d\tau. \tag{2.65}
\]
Applying the continuous-time Fourier transform to both sides of the above equation, and using the convolution property of the transform will yield
\[
Y(i\omega) = \int_{-\infty}^{\infty} \int_{0}^{t} g(\tau)u(t - \tau)d\tau e^{-i\omega t}dt = G(i\omega)U(i\omega) \tag{2.66}
\]
where
\[ G(s) = \int_{-\infty}^{\infty} g(t) e^{-i\omega t} dt = \int_{0}^{\infty} C e^{At} B e^{-st} dt = C(i\omega I - A)^{-1} B \] (2.68)

is called the *continuous-time transfer function* of the continuous-time linear system.

The discrete-time setting is similar to the continuous-time one if the sampling is uniform. If one assumes that \( y(t) \) is observed at the sampling instants \( t_k = kT_s \), then
\[ y(kT_s) = \sum_{l=0}^{k} g_d((k-l)T_s) u(kT_s). \] (2.69)

If we apply the time-discrete Fourier transform in (2.61) to (2.69), this would yield
\[ Y_{T_s}(e^{i\omega T_s}) = G_d(e^{i\omega T_s}) U_{T_s}(e^{i\omega T_s}) \] (2.70)

where
\[ G_d(z) = T_s \sum_{k=-\infty}^{\infty} g_d(kT_s) z^{-k} \] (2.71)

is known as the *discrete-time transfer function*.

### 2.6.2 Spectral Analysis of Time Series Models

A number of spectral properties of stochastic processes are also well known. First of all, the covariance function in (2.12) has a form which implies that it will always be the Fourier transform
\[ r(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi(i\omega) e^{i\omega\tau} d\omega \] (2.72)

of the non-negative function \( \Phi(i\omega) \), called the *spectral density function* or simply *spectrum*. Moving from the covariance function \( r(\tau) \) to the spectrum \( \Phi(i\omega) \) can also be facilitated by the well-known transform formula
\[ \Phi(i\omega) = \int_{-\infty}^{\infty} r(\tau) e^{i\omega\tau} d\tau. \] (2.73)

If the process \( \{x(t), -\infty < t < \infty\} \) is only observed at the discrete-time instances \( t = kT_s \), the covariance function of the new discrete-time process
\[ \{x(t), t = \ldots, -2T_s, -T_s, 0, T_s, 2T_s, \ldots\} \] (2.74)

will be
\[ r_d(kT_s) = r(kT_s) \]
and the discrete-time power spectrum is
\[ \Phi_d(e^{j\omega T_s}) = T_s \sum_{k=-\infty}^{\infty} r_d(kT_s) e^{-j\omega T_s k}, \]  
\[ (2.75) \]
with the inverse relationship
\[ r_d(kT_s) = \frac{1}{2\pi} \int_{-\pi/T_s}^{\pi/T_s} \Phi_d(e^{j\omega T_s}, \theta) e^{j\omega kT_s} d\omega. \]
\[ (2.76) \]
Finally, the following relationship, similar to the Poisson summation formula e.g. Gasquet and Witomski (1998), exists between the discrete-time and continuous-time spectrums e.g. Wahlberg (1988)
\[ \Phi_d(e^{j\omega T_s}) = T_s \sum_{k=-\infty}^{\infty} \Phi_c(i\omega + \frac{j2\pi}{T_s} k). \]
\[ (2.77) \]

Until now, everything that have been said have dealt with explaining the deterministic or stochastic relationship between the input and the output in linear systems. In the rest of the thesis, the objective will instead be to identify this relationship from observation of the input and output, i.e. the inverse problem called system identification.

## 2.7 System Identification

As mentioned in the introduction in Chapter 1, system identification is concerned with estimating models which capture the desired deterministic or stochastic properties of data. Because it concerns estimation for stochastic data, the subject has strong connections to mathematical statistics. The following section will mainly review the fundamental methods of continuous-time system identification which are motivated by the Maximum Likelihood (ML) principle. First, in Subsection 2.7.1 there will be a discussion regarding the method for identifying parameters in a linear time-invariant stochastic system. Then, in Subsection 2.7.2 it will be shown how this approach is related to the use of a Kalman filter in order to perform a factorization of the covariance matrix of the stochastic process.

### 2.7.1 Maximum Likelihood Estimation of Linear Models

Denote the vector of possibly non-equidistant samples \{y(t_k)\}_{k=1}^{N_t} of the output of a linear system as
\[ Y_{N_t} = (y(t_1) \ y(t_2) \ y(t_3) \ldots y(t_{N_t}))^T. \]
\[ (2.78) \]
Assume that the probability density function of \(Y_{N_t}\) for these measurements is
\[ f(\theta; y(t_1), y(t_2), \ldots, y(t_{N_t})) = f(\theta; Y_{N_t}). \]
\[ (2.79) \]
Here $\theta \in \mathbb{R}^d$ contains the parameters that defines the model distribution. These are supposed to be unknown, and the purpose of the observations is in fact to estimate the vector $\theta$ using the information in $Y_{N_t}$. This is accomplished by an estimator

$$\hat{\theta}(Y_{N_t}),$$

(2.80)

which is an $\mathbb{R}^{N_t} \rightarrow \mathbb{R}^d$ function. If the observed value of $Y_{N_t}$ is $Y_{N_t}^*$, then the resulting estimate is denoted

$$\hat{\theta}_* = \hat{\theta}(Y_{N_t}^*).$$

(2.81)

A particular estimator that works under the principle that it maximizes the probability of the observed events, is the maximum likelihood estimator (MLE), introduced by Fisher (1912). The probability density function of the observations is given by (2.79) and the probability that the observations should take on a particular value $Y_{N_t}^*$ is therefore proportional to

$$f(\theta; Y_{N_t}^*).$$

(2.82)

This function is called the likelihood function, because it reflects the likelihood that the observed event would occur. A maximum likelihood estimator (ML) is one where we select $\theta$ such that the observations become as likely as possible. That is,

$$\hat{\theta}(Y_{N_t}^*) = \arg \max_{\theta} f_y(\theta; Y_{N_t}^*)$$

(2.83)

Since the logarithm is a strictly monotone function

$$\hat{\theta}_* = \arg \max_{\theta} f_y(\theta; Y_{N_t}^*)$$

(2.84)

will be equivalent to

$$\hat{\theta}_* = \arg \min_{\theta} L(\theta; Y_{N_t}^*)$$

(2.85)

where

$$L(\theta; Y_{N_t}^*) = - \log f(\theta; Y_{N_t}^*)$$

(2.86)

is known as the negative log-likelihood function.

**Input-Output Models**

Assume that the linear deterministic continuous-time system is written in the state space form

$$\begin{cases}
\dot{x}(t, \theta) = A(\theta)x(t, \theta) + B(\theta)u(t) \\
y(t, \theta) = C(\theta)x(t, \theta)
\end{cases}$$

(2.87)

and the input $u(t)$ is assumed to be zero-order hold as in (2.19). Suppose also that the disturbed output $y(t)$ has been measured at the discrete time instances $\{t_k\}_{k=1}^{N_t}$ such that

$$y_m(t_k) = y(t, \theta) + e(t_k),$$

(2.88)
where the elements of the sequence \( \{e(t_k)\}_{k=1}^{N_t} \) are independent and Gaussian with variance \( \lambda \). Then, the negative-log likelihood method for estimating the parameters would be (Söderström and Stoica, 1989)

\[
L(\theta) = -\log f(e(t_1, \theta), e(t_2, \theta) \ldots e(t_3, \theta)|\theta) \\
= \frac{1}{\lambda} \sum_{k=1}^{N_t} (y_m(t_k) - y(t_k, \theta))^2 + N_t \log \lambda. \tag{2.89}
\]

If we would minimize this expression analytically with respect to \( \lambda \) we get the following maximum-likelihood method for estimating the parameters in (2.87)

\[
\hat{\theta} = \operatorname{arg\min}_{\theta} \frac{1}{N_t} \sum_{k=1}^{N_t} (y_m(t_k) - y(t_k, \theta))^2 \tag{2.91}
\]

where the disturbance variance can be estimated as

\[
\hat{\lambda} = \frac{1}{N_t} \sum_{k=1}^{N_t} \left( y_m(t_k) - y(t_k, \hat{\theta}) \right)^2. \tag{2.92}
\]

**Indirect Method for Input-Output Models**

An alternative to the exact approach described previously is to go via discrete-time parameters \( \theta_d \). If discrete-time equidistantly sampled data is available a discrete-time model can be estimated using standard software (Mathworks, 2004). This model can then put into the state-space form

\[
\begin{cases}
x(kT_s + T_s) = F(\theta_d)x(kT_s) + G(\theta_d)u(kT_s) \\
y(kT_s) = C(\theta_d)x(kT_s) + D(\theta_d)u(kT_s).
\end{cases}
\]

where \( \theta_d \) represents the parameters of the discrete-time system. If the matrix \( F(\theta_d) \) has no eigenvalues on the negative real axis there exists a corresponding continuous-time system (Åström and Wittenmark, 1984)

\[
\begin{cases}
\dot{x}(t) = A(\theta)x(t) + B(\theta)u(t) \\
y(t) = C(\theta)x(t) + D(\theta)u(t)
\end{cases}
\]

where

\[
A(\theta) = \frac{\ln F(\theta_d)}{T_s} \\
B(\theta) = (F(\theta_d) - I)^{-1}A(\theta)G(\theta_d).
\]

The corresponding continuous-time transfer functions is then

\[G(s) = C(sI - A)^{-1}B + D\]

and continuous-time parameters can be acquired from this expression. In a similar fashion, continuous-time stochastic models can be identified from discrete time data.
Time Series Models

In the case of the process in (2.31), we will have a probability distribution

$$Y_{N_t} \sim \mathcal{N}(0, R(\theta))$$  \hfill (2.93)

where

$$R(\theta) = \begin{bmatrix}
    r_y(t_1 - t_1, \theta) & \ldots & r_y(t_1 - t_{N_t}, \theta) \\
    r_y(t_2 - t_1, \theta) & \ldots & r_y(t_2 - t_{N_t}, \theta) \\
    \vdots & \ddots & \vdots \\
    r_y(t_{N_t} - t_1, \theta) & \ldots & r_y(t_{N_t} - t_{N_t}, \theta)
\end{bmatrix}.$$  \hfill (2.94)

In the case of stationarity we also know from (2.44) that

$$r_y(\tau, \theta) = C(\theta) e^{A(\theta)\tau} P_x(\theta) C(\theta)^T, \quad \tau > 0$$

where

$$A(\theta) P_x(\theta) + P_x(\theta) A(\theta) + \sigma^2 B(\theta) B(\theta)^T = 0.$$  \hfill (2.95)

This paves the way for the Maximum-Likelihood estimation framework in (2.83) and the vector of samples will have the likelihood function

$$f(\theta; Y_{N_t}^*) = \frac{1}{(2\pi)^{N_t/2} \sqrt{\det R(\theta)}} e^{-\frac{1}{2} Y_{N_t}^{*T} R(\theta)^{-1} Y_{N_t}^*}.$$  \hfill (2.96)

and the log likelihood function will be

$$L(\theta) = -\log f(\theta; Y_{N_t}^*) = \frac{N}{2} \log 2\pi + \frac{1}{2} \log \det R(\theta) - \frac{1}{2} Y_{N_t}^{*T} R(\theta)^{-1} Y_{N_t}^*.$$  \hfill (2.97)

To sum things up we can therefore state that the “brute force” Maximum Likelihood (ML) method for estimating the parameters of the model in (2.31) would be

$$\hat{\theta}_* = \arg \min_\theta Y_{N_t}^{*T} R(\theta)^{-1} Y_{N_t}^* + \log \det R(\theta).$$  \hfill (2.98)

A monumental obstacle in the way of exploiting this approach fully, is that when $N_t$ grows large, the size of the $R(\theta)$ matrix will explode, and the inversion will be very costly. However, since the matrix $R(\theta)$ is symmetric and positive definite it is possible to perform an $LDLT^T$-factorization, a special form of the Cholesky factorization (Golub and Loan, 1996)

$$R(\theta) = L(\theta) D(\theta) L^T(\theta).$$  \hfill (2.99)

The objective function in the optimization can then be rewritten as

$$\hat{\theta} = \arg \min_\theta Y_{N_t}^{T} L^{-T}(\theta) D^{-1}(\theta) L^{-1}(\theta) Y_{N_t} + \sum_{k=1}^{N_t} \log D_{kk}(\theta)$$  \hfill (2.100)
where the $D_{kk}(\theta)$ are the eigenvalues $\lambda_k(R(\theta))$. The minimization can then be carried out using standard optimization method such as Gauss-Newton, Levenberg-Marquard or BFGS (Nocedal and Wright, 1999; Dennis and Schnabel, 1983).

This approach tends to work only if we have a medium or small number of measurements $N_t$ since the size of $R(\theta)$ is $N_t \times N_t$ and the Cholesky factorization would need approximately $O(N_t^3)$ operations.

It is interesting to note that the covariance function $r(\tau)$ will be close to zero for large $\tau$ since it decays as $e^{-\lambda \tau}$ for some $\lambda > 0$. This means that if the time of observation $t_{N_t}$ is large, most of the diagonals in the covariance matrix $R(\theta)$ in (2.94) will be more or less equal to zero. If one would set all but $M$ diagonals equal to zero where $M \ll N_t$ the matrix would become banded. The complexity of using a banded Cholesky factorization would now reduce to $O(M^2 N_t)$ operations. This property was utilized in a quite early paper by Jones (1977) on the topic of ML estimation from non-uniformly sampled data.

### 2.7.2 Kalman Filter Based System Identification

It has for a long time been known that doing the $LDL^T$-factorization in (2.99) and then taking the inverse as in (2.100), is equivalent to applying the Kalman filter (Kalman, 1960) to the output sequence $\{y(t_k)\}_{k=1}^{N_t}$. As we will see in the following section, this procedure is recursive and will circumvent the large matrices which comes with the direct use of the $LDL^T$ factorization. In this subsection we will explain the Kalman filter together with its use in system identification.

**Kalman Filter**

Discussing the Kalman filter usually begins with assuming that the data originates from the following state space model

$$
\begin{align*}
\begin{cases}
    x(t_{k+1}, \theta) = F(t_k, \theta) x(t_k, \theta) + u(t_k, \theta) \\
y(t_k) = C(\theta) x(t_k, \theta) + v(t_k, \theta),
\end{cases}
\end{align*}
$$

(2.101)

which can be of purely discrete form or originate from a continuous-time system. This entity, statistically describes the relationship between the output $\{y(t_k)\}_{k=1}^{N_t}$ and the so called process noise $\{u(t_k, \theta)\}_{k=1}^{N_t}$ and measurement noise $\{v(t_k, \theta)\}_{k=1}^{N_t}$. Most importantly, it models the covariance function $r(\tau, \theta)$. Here we assume that $u$ and $v$ are Gaussian stationary processes such that $E[u(t_k, \theta)] = 0$, $E[v(t_k, \theta)] = 0$ and

$$
E \begin{bmatrix} u(t_k, \theta) \\ v(t_k, \theta) \end{bmatrix} = \begin{bmatrix} Q(t_k, \theta) \\ S(t_k, \theta) \\ S(t_k, \theta)^T \\ R(t_k, \theta) \end{bmatrix}.
$$

(2.102)

An important consequence of the work of Kalman (Kalman, 1960) is that it is possible to find matrices $K_p(t_k, \theta)$ such that the state space model

$$
\begin{align*}
\begin{cases}
    \dot{x}(t_{k+1}, \theta) = F(t_k, \theta) \dot{x}(t_k, \theta) + K_p(t_k, \theta) e(t_k, \theta), \quad \dot{x}(0) = 0 \\
y(t_k) = C(\theta) x(t_k, \theta) + e(t_k, \theta)
\end{cases}
\end{align*}
$$

(2.103)

is statistically equivalent to the representation in (2.101). By this we mean that the covariance functions of $y(t_k)$ are the same for both representations. The form in (2.103) is
known as the innovations form where the innovations \( \{e(t_k, \theta)\}_{k=1}^{N_t} \) make up a Gaussian white noise process with the first and second order moments

\[
Ee(t_k, \theta) = 0 \quad (2.104)
\]

\[
Ee(t_k, \theta)e(t_l, \theta) = \begin{cases} 
0 & k \neq l \\
R_e(t_k, \theta) & k = l.
\end{cases} \quad (2.105)
\]

Here \( K_p(t_k, \theta) \) and \( R_e(t_k, \theta) \) are defined as

\[
K_p(t_k, \theta) = \left( F(t_k, \theta)P_p(t_k, \theta)C(\theta)^T + S(t_k, \theta) \right) R_e(t_k, \theta)^{-1} \quad (2.106)
\]

\[
R_e(t_k, \theta) = R(t_k, \theta) + C(\theta)P_p(t_k, \theta)C(\theta)^T, \quad (2.107)
\]

where the evolution of \( P_p(t_k, \theta) \) is dictated by the famous Riccati recursion

\[
P_p(t_{k+1}, \theta) = F(t_k, \theta)P_p(t_k, \theta)F(t_k, \theta)^T + Q(t_k, \theta) - K_p(t_k, \theta)R_e(t_k, \theta)K_p(t_k, \theta). \quad (2.108)
\]

The system (or filter) can also be decomposed into a time and a measurement update phase (Schmidt, 1966) where we, for the sake of simplicity, assume that \( S(t_k) = 0 \) (details see e.g. (Kailath et al., 2000)). The time update can then be written as

\[
\begin{align*}
\hat{x}_p(t_{k+1}, \theta) &= F(t_k, \theta)\hat{x}_f(t_{k+1}, \theta) \\
P_p(t_{k+1}, \theta) &= F(t_k, \theta)P_f(t_k, \theta)F(t_k, \theta)^T + Q(t_k, \theta),
\end{align*} \quad (2.109)
\]

and the measurement phase is defined as

\[
\begin{align*}
\hat{x}_f(t_k, \theta) &= \hat{x}_p(t_k, \theta) + K_f(t_k, \theta) (y(t_k) - C\hat{x}_p(t_k, \theta)) \\
K_f(t_k, \theta) &= P_p(t_k, \theta)C(t_k, \theta)^TR_e(t_k, \theta)^{-1} \\
P_f(t_k, \theta) &= P_p(t_k, \theta) - K_f(t_k, \theta)R_e(t_k, \theta)K_f(t_k, \theta)^T.
\end{align*} \quad (2.110)
\]

In the following example the filter setup is exemplified for a second order time series model with no measurement noise.

--- Example 2.2 ---

Assume that we have a stochastic model

\[
\begin{align*}
x(t_{k+1}, \theta) &= F(t_k, \theta)x(t_k, \theta) + u(t_k, \theta) \\
y(t_k) &= C(\theta)x(t_k, \theta)
\end{align*} \quad (2.111)
\]

where \( S(t_k, \theta) = 0, R(t_j, \theta) = 0 \) and

\[
F(t_k, \theta) = \begin{pmatrix} f_{11}(t_k, \theta) & f_{12}(t_k, \theta) \\ f_{21}(t_k, \theta) & f_{22}(t_k, \theta) \end{pmatrix} \quad (2.112)
\]

\[
Q(t_k, \theta) = \begin{pmatrix} q_{11}(t_k, \theta) & q_{12}(t_k, \theta) \\ q_{21}(t_k, \theta) & q_{22}(t_k, \theta) \end{pmatrix} \quad (2.113)
\]

\[
C(\theta) = \begin{pmatrix} 0 & 1 \end{pmatrix}. \quad (2.114)
\]
If we define the covariance matrices of the predicted and filtered state errors as

\[
P_p(t_k, \theta) = \begin{pmatrix} p^p_1(t_k, \theta) & p^p_2(t_k, \theta) \\ p^p_2(t_k, \theta) & p^p_3(t_k, \theta) \end{pmatrix},
\]

\[
P_f(t_k, \theta) = \begin{pmatrix} p^f_1(t_k, \theta) & p^f_2(t_k, \theta) \\ p^f_2(t_k, \theta) & p^f_3(t_k, \theta) \end{pmatrix}.
\]

(2.115) (2.116)

the Kalman filter measurement update will take on the form of

\[
P_f(t_k, \theta) = P_p(t_k, \theta) - K_f(t_k, \theta) R_e(t_k, \theta) K_f(t_k, \theta)
\]

\[
= P_p(t_k, \theta) + \frac{P_p(t_k, \theta) C^T C P_p(t_k, \theta)}{C P_p(t_k, \theta) C^T} \left( p^p_1(t_k, \theta) \quad p^p_2(t_k, \theta) \quad p^p_3(t_k, \theta) \right) \left( p^p_1(t_k, \theta) \quad p^p_2(t_k, \theta) \quad p^p_3(t_k, \theta) \right) = \left( \begin{array}{c} p^p_1(t_k, \theta) \\ p^p_2(t_k, \theta) \\ p^p_3(t_k, \theta) \end{array} \right) - \frac{p^p_1(t_k, \theta) p^p_2(t_k, \theta) p^p_3(t_k, \theta)}{p^p_3(t_k, \theta)} \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\]

and the estimate measurement update will be

\[
\hat{x}_f(t_k, \theta) = \hat{x}_p(t_k, \theta) + K_f(t_k, \theta) (y(t_k) - C \hat{x}_p(t_k, \theta))
\]

\[
= \left( \begin{array}{c} 1 - \frac{p^p_2(t_k, \theta)}{p^p_3(t_k, \theta)} \end{array} \right) \hat{x}_p(t_k, \theta) + \frac{p^p_2(t_k, \theta)}{p^p_3(t_k, \theta)} y(t_k)
\]

because

\[
K_f(t_k, \theta) = \frac{P_p(t_k, \theta) C^T}{C P_p(t_k, \theta) C^T} = \left( \begin{array}{c} \frac{p^p_2(t_k, \theta)}{p^p_3(t_k, \theta)} \\ 0 \end{array} \right).
\]

The time update phase will develop as

\[
P_p(t_{k+1}, \theta) = \begin{pmatrix} p^p_1(t_{k+1}, \theta) & p^p_2(t_{k+1}, \theta) \\ p^p_2(t_{k+1}, \theta) & p^p_3(t_{k+1}, \theta) \end{pmatrix}
\]

\[
= F(t_k, \theta) P_f(t_k, \theta) F(t_k, \theta) + Q(t_k, \theta)
\]

\[
= \begin{pmatrix} f_{11}(t_k, \theta) & f_{12}(t_k, \theta) \\ f_{21}(t_k, \theta) & f_{22}(t_k, \theta) \end{pmatrix} + \begin{pmatrix} q_{11}(t_k, \theta) & q_{12}(t_k, \theta) \\ q_{21}(t_k, \theta) & q_{22}(t_k, \theta) \end{pmatrix}.
\]

(2.117) (2.118) (2.119)

The connection between the \(L(\theta) D(\theta) L(\theta)^T\) factorization and the Kalman filter is now a matter of straightforward manipulations. If we define the vector of the innovations sequence

\[
E(\theta) = (e(t_1, \theta) \ e(t_2, \theta) \ e(t_3, \theta) \ \ldots \ e(t_N, \theta))^T
\]

(2.120)
the relationship between $Y_{N_t}$ and $E(\theta)$ will be defined by the equation

$$Y_{N_t} = L(\theta)E(\theta)$$

where the matrix $L(\theta)$ is a consequence of propagating the innovations form in (2.103). This yields an

$$L(\theta) = \begin{bmatrix}
I & 0 & \ldots & 0 \\
C(\theta)K_p(t_1, \theta) & I & \ldots & 0 \\
C(\theta)\Phi(t_2, t_1, \theta)K_p(t_1, \theta) & C(\theta)K_p(t_1, \theta) & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
C(\theta)\Phi(t_N, t_1, \theta)K_p(t_1, \theta) & C(\theta)\Phi(t_N, t_2, \theta)K_p(t_1, \theta) & \ldots & I
\end{bmatrix}$$

where the transition matrix $\Phi$ is defined as

$$\left\{ \begin{array}{l}
\Phi(t_k, t_l, \theta) = F(t_{i-1, \theta})F(t_{i-2, \theta}) \ldots F(t_j, \theta) \quad i > j \\
\Phi(t_k, t_k, \theta) = I.
\end{array} \right. \quad (2.121)$$

Thus, in order to compute the innovations, one would just need to invert $L(\theta)$ such that

$$E(\theta) = L(\theta)^{-1}Y_{N_t}(\theta). \quad (2.122)$$

Fortunately, this inverse will just be a consequence of inverting a lower triangular matrix, and the result will then be

$$L(\theta)^{-1} = \begin{bmatrix}
I & 0 & \ldots & 0 \\
-C(\theta)K_p(t_1, \theta) & I & \ldots & 0 \\
-C(\theta)\Phi_p(t_2, t_1, \theta)K_p(t_1, \theta) & -C(\theta)K_p(t_1, \theta) & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
-C(\theta)\Phi_p(t_N, t_1, \theta)K_p(t_1, \theta) & -C(\theta)\Phi_p(t_N, t_2, \theta)K_p(t_1, \theta) & \ldots & I
\end{bmatrix}$$

where a new transition matrix will be defined as

$$\left\{ \begin{array}{l}
\Phi_p(t_k, t_l, \theta) = F_p(t_{i-1, \theta})F(t_{i-2, \theta}) \ldots F(t_j, \theta) \quad , i > j \\
\Phi_p(t_k, t_k, \theta) = I
\end{array} \right. \quad (2.123)$$

with

$$F_p(t_k, \theta) = F(t_k, \theta) - K_p(t_k, \theta)C(\theta). \quad (2.124)$$

For the innovations form, the inversion of the state-space equations can be easily written as

$$\left\{ \begin{array}{l}
\hat{x}(t_{k+1}, \theta) = F_p(t_k, \theta)\hat{x}(t_k, \theta) + K_p(t_k, \theta)y(t_k), \quad \hat{x}(0) = 0 \\
e(t_k, \theta) = y(t_k, \theta) - C\hat{x}(t_k, \theta).
\end{array} \right. \quad (2.125)$$

This means that another way of estimating the parameters $\theta$ is to use the Kalman filter in (2.125) to compute the innovations $\{e(t_k, \theta)\}_{k=1}^{N_t}$. Since $e(t_k)$ are independent and for each $t_k$

$$e(t_k, \theta) \sim \mathcal{N}(0, R_e(t_k, \theta)) \quad (2.126)$$
2.7 System Identification

the maximum likelihood procedure using the likelihood function will be

\[
\begin{aligned}
\hat{\theta} &= \arg\min_{\theta} V(\theta) \\
V(\theta) &= \frac{1}{N} \sum_{k=1}^{N} e^{2(t_k, \theta)} + \log Re(t_k, \theta).
\end{aligned}
\] (2.127)

**Numerical Minimization**

Solving the minimization problem in (2.127) can be carried out using standard unconstrained optimization procedures (Nocedal and Wright, 1999; Dennis and Schnabel, 1983). These methods basically operate by creating a local quadratic model of the objective function around the current iterate \( \theta_k \), such that

\[
m_k(d_k) = V(\theta_k) + \nabla V(\theta_k)^T d_k + \frac{1}{2} d_k^T B_k d_k.
\] (2.128)

The minimizer of this optimization problem is

\[
p_k = -B_k^{-1} \nabla V(\theta_k)
\] (2.129)

which is used as a search direction. The new iterate will then be

\[
\theta_{k+1} = \theta_k + \alpha_k p_k,
\] (2.130)

where the step length \( \alpha_k \) is chosen through a line search (Ortega and Rheinboldt, 1970). In the classical Newton method the choice for \( B_k \) is

\[
B_k = \nabla^2 V(\theta_k).
\] (2.131)

This quantity can however be unnecessary complicated to compute. Modern quasi-Newton methods use the observed behavior of \( V(\theta_k) \) and \( \nabla V(\theta_k) \) in order to build curvature information and approximate the Hessian \( \nabla^2 V(\theta_k) \). One of the most well known and powerful such methods is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) in which \( B_k \) is updated through the formula

\[
B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}
\] (2.132)

where

\[
s_k = \theta_{k+1} - \theta_k
\] (2.133)

\[
y_k = \nabla V(\theta_{k+1}) - \nabla V(\theta_{k+1}).
\] (2.134)

Therefore the method only needs to be provided with the function value \( f(\theta_k) \) and the gradient \( \nabla f(\theta_k) \). This algorithm is implemented in the MATLAB Optimization Toolbox under the function name `fminunc`. In the following example we will show how the gradient can be computed recursively for the objective function in (2.127)
Gradient Computation

The elements $\frac{\partial}{\partial \theta} V(\theta)$ of the gradient $\nabla V(\theta)$ can be computed as

$$\frac{\partial}{\partial \theta} V(\theta) = \frac{1}{N} \sum_{k} 2 \frac{\partial}{\partial \theta} e(t_k, \theta) e(t_k, \theta) \frac{\partial}{\partial \theta} R_e(t_k, \theta) - \frac{\partial}{\partial \theta} R_e(t_k, \theta) e(t_k, \theta)^2 + \frac{\partial}{\partial \theta} C(t_k, \theta) R_e(t_k, \theta).$$

Since the derivatives of the prediction error are

$$\frac{\partial}{\partial \theta} e(t_k, \theta) = \frac{\partial}{\partial \theta} (y(t_k) - C(\theta) \hat{x}_p(t_k, \theta))$$

$$= - \frac{\partial}{\partial \theta} C(\theta) \hat{x}_p(t_k, \theta) - C(\theta) \frac{\partial}{\partial \theta} \hat{x}_p(t_k, \theta)$$

we need the quantities $\frac{\partial}{\partial \theta} \hat{x}_p(t_k, \theta)$ and $\frac{\partial}{\partial \theta} R_e(t_k, \theta)$. Fortunately, there is a straightforward way to compute these since

$$\frac{\partial}{\partial \theta} R_e(t_k, \theta) = \frac{\partial}{\partial \theta} R(t_k, \theta) + \frac{\partial}{\partial \theta} C(\theta) P_p(t_k, \theta) C(\theta)^T$$

$$+ C(\theta) \frac{\partial}{\partial \theta} P_p(t_k, \theta) C(\theta)^T$$

$$+ C(\theta) P_p(t_k, \theta) \frac{\partial}{\partial \theta} C(\theta)^T,$$

and

$$\frac{\partial}{\partial \theta} \hat{x}_p(t_k+1, \theta) = \frac{\partial}{\partial \theta} F(t_k, \theta) \hat{x}(t_k+1, \theta) + F(t_k, \theta) \frac{\partial}{\partial \theta} \hat{x}(t_k+1, \theta),$$

$$\frac{\partial}{\partial \theta} P_p(t_k+1, \theta) = \frac{\partial}{\partial \theta} F(t_k, \theta) P_f(t_k, \theta) F(t_k, \theta) + F(t_k, \theta) \frac{\partial}{\partial \theta} P_f(t_k, \theta) F(t_k, \theta)$$

$$+ F(t_k, \theta) P_f(t_k, \theta) \frac{\partial}{\partial \theta} F(t_k, \theta) + \frac{\partial}{\partial \theta} Q(t_k, \theta).$$

The derivatives of the measurement update equations are for the filtered estimate

$$\frac{\partial}{\partial \theta} \hat{x}_f(t_k, \theta) = \frac{\partial}{\partial \theta} \hat{x}_p(t_k, \theta) + \frac{\partial}{\partial \theta} K_f(t_k, \theta) e(t_k, \theta)$$

$$+ K_f(t_k, \theta) \frac{\partial}{\partial \theta} e(t_k, \theta),$$

for the filter gain

$$\frac{\partial}{\partial \theta} K_f(t_k, \theta) = \frac{\partial}{\partial \theta} P_p(t_k, \theta) C(\theta)^T R_e(t_k, \theta)^{-1}$$

$$+ P_p(t_k, \theta) \frac{\partial}{\partial \theta} C(\theta)^T R_e(t_k, \theta)^{-1}$$

$$- P_p(t_k, \theta) C(\theta)^T R_e(t_k, \theta)^{-1} \frac{\partial}{\partial \theta} R_e(t_k, \theta) R_e(t_k, \theta)^{-1}.$$
2.7 System Identification

and for the filter error covariance

\[
\frac{\partial}{\partial \theta_i} P_f(t_k, \theta) = \frac{\partial}{\partial \theta_i} P_p(t_k, \theta) - \frac{\partial}{\partial \theta_i} K_f(t_k, \theta) R_e(t_k, \theta) K_f(t_k, \theta) \\
- K_f(t_k, \theta) \frac{\partial}{\partial \theta_i} R_e(t_k, \theta) K_f(t_k, \theta) \\
- K_f(t_k, \theta) R_e(t_k, \theta) \frac{\partial}{\partial \theta_i} K_f(t_k, \theta).
\]

This means that what is basically needed in the computations are \( \frac{\partial}{\partial \theta_i} F(t_k, \theta) \), \( \frac{\partial}{\partial \theta_i} C(\theta) \) and \( \frac{\partial}{\partial \theta_i} Q(t_k, \theta) \). Every other quantity needed for \( \frac{\partial}{\partial \theta_i} V(\theta) \) can then be computed recursively.

An important special case in system identification is when the time between sampling instants is constant i.e. \( \Delta t_k = t_{k+1} - t_k = T_s \) for some \( T_s \) and for all \( k \). This simplifies many things. For instance, in the case of the Kalman filter formulation, the number of computations will be substantially reduced if the sampling interval is set to constant. This is a consequence of the fact that the matrices \( K_p(\theta) \) and \( R_e(\theta) \) will become constant

\[
K_p(\theta) = (F(\theta) P(\theta) C(\theta)^T + S(\theta)) R_e(\theta)^{-1}
\]

\[
R_e(\theta) = R(\theta) + C(\theta) P(\theta) C(\theta)^T,
\]

(2.135)

(2.136)

where \( P_p(\theta) \) is now defined by the Riccati equation

\[
P_p(\theta) = F(\theta) P_p(\theta) F(\theta)^T + Q(\theta) - K_p(\theta) R_e(\theta) K_p(\theta).
\]

(2.137)

By reparameterizing the problem such that \( R_e \) will be a parameter in itself it can be shown that the optimization with respect to \( R_e \) can be done analytically. The remaining parameters are then identified as

\[
\hat{\theta} = \arg \min_{\theta} \sum_{k=1}^{N_t} e^2(kT_s, \theta),
\]

(2.138)

and the estimate of \( \hat{R}_e \) is given by the formula

\[
\hat{R}_e = \frac{1}{N_t} \sum_{k=1}^{N_t} e^2(kT_s, \hat{\theta}).
\]

(2.139)

In this way a continuous-time model can be directly identified.

### 2.7.3 Frequency Domain System Identification

In the following section we will explain how to do system identification from input data that has undergone either a discrete- or continuous-time Fourier transform. First, there will be a short text explaining the classical Whittle likelihood estimator. Then, a more elaborate exposition on the continuous-time approach will follow.
Input-Output Models

Suppose that the input-output data of a deterministic system such as the one in (2.13) have been sampled with the constant sampling interval $T_s$. If the input is piecewise constant, the relationship between the discrete-time input and output can then be characterized by the discrete-time pulse transfer function $G_d$

$$G_d(e^{jωT_s}, θ) = C(θ)(e^{jωIT_s} - e^{A(θ)T_s})^{-1}Γ(θ) + D(θ) \tag{2.140}$$

$$Γ(θ) = \int_0^{T_s} e^{A(θ)t}B(θ)dt.$$

Now, define the discrete-time Fourier transform of the output $\{y(t_k)\}_{k=1}^{N_t}$ as

$$Y_d(e^{jωT_s}) = \frac{1}{\sqrt{N_t}} \sum_{k=1}^{N_t} y(kT_s)e^{-jωkT_s} \tag{2.141}$$

and the discrete Fourier transform at the frequencies $ω_1, \ldots, ω_{N_ω}$ as

$$Y_d = \{Y_d(e^{jω_1T_s}), \ldots, Y_d(e^{jω_{Nω}T_s})\}$$

where $ω_k = \frac{2π}{T_s}k$. Let, analogously $U_d$ denote the discrete-time Fourier transform of the input. Then, we may treat $Y_d$ and $U_d$ as observations, neglect transient (non-periodic) effects, and assume that the Fourier transforms are independent at different frequencies. All of this may be more or less good approximations – see pp 230 in (Ljung, 1999). Under these assumptions, the negative log-likelihood function for values of the discrete-time Fourier transform is (Pintelon and Schoukens, 2001)

$$L(θ) = -\log p(Y_d|θ)$$

$$= \frac{N_ω}{2} \log 2π + \sum_{k=1}^{N_ω} \frac{1}{2} |Y_d(e^{jω_kT_s}) - G_d(e^{jω_kT_s}, θ)U_d(e^{jω_kT_s})|^2$$

where $G_d$ is the so called pulse-transfer function (2.140) for the system $G_c$ in (2.13). The ML-procedure for estimating the parameters is then

$$\hat{θ} = \arg \min_θ V_d(θ)$$

where

$$V_d(θ) = \sum_{k=1}^{N_ω} |Y_d(e^{jω_kT_s}) - G_d(e^{jω_kT_s}, θ)U_d(e^{jω_kT_s})|^2. \tag{2.142}$$

This is a method that has long been used to estimate parameters of OE models (Mathworks, 2004). The transfer function $G_d$ can also be parameterized in terms of the discrete-time parameters. These can then be estimated and transformed to continuous-time using the matrix exponential.
Discrete Linear Time Series Models

Assume that we have sampled the output data of the system in (2.31) uniformly such that the vector of measurements is

\[ Y_{N_t} = \begin{pmatrix} y(0) & y(T_s) & y(2T_s) & \ldots & y(N_tT_s) \end{pmatrix}^T. \] (2.143)

Then the covariance matrix of \( Y_{N_t} \) in (2.94) is a Toeplitz matrix such that

\[
R(\theta) = \begin{pmatrix}
  r_y(0, \theta) & \ldots & r_y(-N_tT_s, \theta) \\
  r_y(T_s, \theta) & \ldots & r_y(-(N_t - 1)T_s, \theta) \\
  \vdots & \ddots & \vdots \\
  r_y(N_tT_s, \theta) & \ldots & r_y(0, \theta)
\end{pmatrix}
\] (2.144)

where the covariance function \( r_y(\tau, \theta) \) is defined by (2.12). This covariance function can, as indicated in (2.76), also be written as

\[
r_y(kT_s, \theta) = \frac{1}{2\pi} \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} \Phi_d(e^{i\omega T_s}, \theta) e^{i\omega kT_s} d\omega.
\] (2.145)

As pointed out in (2.97) the negative log-likelihood function for the parameters given the measurements \( Y_{N_t} \) will be

\[
L(\theta) = -\log f_y(Y_{N_t} | \theta) = \frac{N_t}{2} \log 2\pi + \frac{1}{2} \log \det R(\theta) + \frac{1}{2} Y_{N_t}^T R(\theta)^{-1} Y_{N_t}.
\] (2.146)

This expression is not new. However, by a famous result of Grenander and Szegö (1984) we can show that the second term in (2.146) will have the property that

\[
\frac{1}{N_t} \log \det R(\theta) \rightarrow \frac{1}{2\pi} \int_{-\frac{\pi}{T_s}}^{\frac{\pi}{T_s}} \log \Phi_d(e^{i\omega T_s}, \theta) d\omega
\] (2.147)
as $N_t \to \infty$. The third term in (2.146) can also be approximated as (Whittle, 1961)

$$\frac{1}{N_t} Y^T_{N_t} R^{-1}(\theta) Y_{N_t} \approx \sum_{k=1}^{N} \sum_{k=1}^{N} y(kT_s) \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\Phi_d(e^{i\omega T_s}, \theta)} e^{i\omega(k-l)T_s} d\omega \right\} y(lT_s)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\Phi_d(e^{i\omega T_s}, \theta)} e^{i\omega T_s} \sum_{k=1}^{N} \sum_{k=1}^{N} y(kT_s) y(lT_s) e^{i\omega(k-l)T_s} d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\sum_{k=1}^{N} y(kT_s) e^{-i\omega kT_s}}{\Phi_d(e^{i\omega T_s}, \theta)} d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\hat{\Phi}_d(e^{i\omega T_s})}{\Phi_d(e^{i\omega T_s}, \theta)} d\omega$$

where

$$\hat{\Phi}_d(e^{i\omega T_s}) = \left| \sum_{k=1}^{N_t} y(kT_s) e^{-i\omega kT_s} \right|^2$$

(2.148)

is the discrete-time periodogram of the sequence $y(kT_s), k = 1 \ldots N_t$. Altogether this means that, as $N_t \to \infty$

$$\frac{1}{N_t} L(\theta) \to \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\hat{\Phi}_d(e^{i\omega T_s})}{\Phi_d(e^{i\omega T_s}, \theta)} + \log \Phi_d(e^{i\omega T_s}, \theta) d\omega,$$

(2.149)

where the right hand side quantity is usually called the Whittle likelihood function (Whittle, 1961). This function requires the evaluation of an integral, which is normally approximated by its Riemann sum yielding the following approximate maximum likelihood method for estimating the parameters $\theta$

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N_t} \sum_{k=0}^{N_t} \hat{\Phi}_d(e^{i\frac{2\pi k}{N_t}}) + \log \Phi_d(e^{i\frac{2\pi k}{N_t}}, \theta).$$

(2.150)

This also simplifies the calculation of the periodogram

$$\hat{\Phi}_d(e^{i\omega T_s}) = \left| Y_d(e^{i\omega T_s}) \right|^2$$

(2.151)

$$Y_d(e^{i\omega T_s}) = \sum_{k=1}^{N_t} y(kT_s) e^{-i\omega kT_s},$$

(2.152)

which is only evaluated at the frequencies $\omega = \frac{2\pi k}{T_s N_t}, k = 1 \ldots N$ and hence $Y_d$ can be computed using the Fast Fourier Transform (FFT) at a cost of only $O(N_t \log N_t)$ operations.
Continuous Linear Time Series Models

This thesis mainly concerns the estimation of continuous-time parameters from sampled data. In the Whittle likelihood method, the identification takes place via the discrete-time system and its spectrum. The relationship between the discrete-time and continuous-time parameters can be quite complicated, and it would therefore be desirable to perform identification using the continuous-time parameters directly.

In this section we will establish how to estimate the parameters $\theta$, if someone would give us continuous-time measurements of the output $y(t)$ of (2.31) during a time interval $0 \leq t \leq T$. We would then compute the truncated continuous-time Fourier transform

$$Y_c(i\omega) = \frac{1}{\sqrt{T}} \int_0^T y(t)e^{-i\omega t} dt$$

(2.153)

and form the continuous-time periodogram

$$\hat{\Phi}_c(i\omega) = |Y_c(i\omega)|^2 = \left| \frac{1}{\sqrt{T}} \int_0^T y(t)e^{-i\omega t} dt \right|^2.$$  

(2.154)

The parameters would then be estimated by solving the following minimization problem

$$\hat{\theta} = \arg \min_{\theta} \sum_{k=1}^{N_\omega} \hat{\Phi}_c(i\omega_k) + \log \Phi_c(i\omega_k, \theta),$$

(2.155)

where the frequencies $\omega_k$, $k = 1, \ldots, N_\omega$ are chosen such that

$$\omega_k \in \left\{ \frac{2\pi}{T} l, l \in \mathbb{Z} \right\}.$$  

(2.156)

The expression in (2.155) is an approximate Maximum-Likelihood method where the model in (2.31) has been transformed into the frequency domain. The frequencies have been deliberately selected such that the Fourier transforms of the output at different frequencies are asymptotically uncorrelated. As in the Whittle likelihood approach the quadratic form in the time-domain criterion in (2.97) have been replaced by a summation in the frequency domain. Hence, the complexity has been greatly reduced and the method allows for the direct use of the continuous-time spectrum. The criterion here is the negative log-likelihood function for the real and imaginary parts of the Fourier transform, given the parameters $\theta$. The remaining part of this section will now be dedicated to motivating this approach. First a result on the distribution of the truncated Fourier transform in (2.153).

Lemma 2.1

Assume that one has a stochastic process $y(t)$ generated by the model in (2.31). Then, the truncated Fourier transform of the process from time $t = 0$ to $t = T$ will be

$$Y_c(i\omega) = \frac{B(i\omega)}{A(i\omega)} \sqrt{T} \int_0^T e^{-i\omega t} dW(t) + \frac{1}{\sqrt{T}} C(i\omega I - A)^{-1} (x(0) - e^{-i\omega T} x(T))$$

(2.157)
were \( x(0) \) and \( x(T) \) are stochastic variables denoting the states at the initial point and the end point.

**Proof:** The solution to the differential equation in (2.31) is given by (2.38) as

\[
y(t) = C e^{At} x(0) + \int_0^t C e^{A(t-\tau)} B dW(\tau).
\]  

The output \( y(t) \) is observed through a window

\[
W(t) = \frac{1}{\sqrt{T}} I_{[0,T]}
\]

where \( I \) is the indicator function. Then, we have the transform

\[
Y_c(i\omega) = \int_{-\infty}^{\infty} W(t) y(t) e^{-i\omega t} dt,
\]

which is a stochastic variable. The transform of the first term in (2.158) would then be

\[
\frac{1}{\sqrt{T}} \int_0^T C e^{(A-i\omega I)t} x(0) dt = \frac{1}{\sqrt{T}} C(i\omega I - A)^{-1} (I - e^{(A-i\omega I)t}) x(0)
\]

were \( x(0) \) is a stochastic variable. Integration by parts applied to the transform of the second part of (2.158) would yield

\[
\int_0^T f(t) B(t) dt = F(T) B(T) - \int_0^T F(t) B(t) dt
\]

(2.159)

where

\[
f(t) = C e^{(A-i\omega I)t}
\]

(2.161)

\[
F(t) = C (A - i\omega I)^{-1} e^{(A-i\omega I)t}
\]

(2.162)

\[
B(t) = \int_0^t e^{-A\tau} B dW(\tau).
\]

(2.163)
This relationship would then provide

\[ \int_0^T \int_0^t C e^{A(t-\tau)} B dW(\tau) e^{-i\omega t} d\tau = \]

\[ C(A - i\omega I)^{-1} e^{(A - i\omega I)T} \int_0^T e^{-At} B dW(t) - \int_0^T C(A - i\omega I)^{-1} e^{(A - i\omega I)T} e^{-At} B dW(t) \]

\[ = C(A - i\omega I)^{-1} e^{-i\omega T} \int_0^T e^{A(T-t)} B dW(t) - \int_0^T C(A - i\omega I)^{-1} e^{(A - i\omega I)T} e^{-At} B dW(t) \]

\[ = C(A - i\omega I)^{-1} e^{-i\omega T} \int_0^T e^{A(T-t)} B dW(t) - \int_0^T C(A - i\omega I)^{-1} B e^{-i\omega t} dW(t) \]

\[ = C(A - i\omega I)^{-1} e^{-i\omega T} \int_0^T e^{A(T-t)} B dW(t) + \frac{B(i\omega)}{A(i\omega)} \int_0^T e^{-i\omega t} dW(t). \]

One effect of the initial and end point conditions is that

\[ \int_0^T e^{A(T-t)} B dW(t) + e^{AT} x(0) = x(T), \]

and by this the transform will become

\[ Y_c(i\omega) = \frac{B(i\omega)}{A(i\omega)} \frac{1}{\sqrt{T}} \int_0^T e^{-i\omega t} dW(t) + \frac{1}{\sqrt{T}} C(i\omega I - A)^{-1}(x(0) - e^{-i\omega T} x(T)). \]

\[ \square \]

In order to determine the probability distribution of the Fourier transform, which will be normal, it is necessary to know the correlation between different frequency components. The following theorem illustrates this relationship.

**Theorem 2.1**

Assume that \( Y_c(i\omega) \) is defined as in (2.153) and \( \omega_k \) and \( \omega_l \) are defined as in (2.156). Further, assume that the process \( y(t) \) is stationary and that \( A \) in (2.31) has no eigenvalues on the imaginary axis yielding a stable system in (2.31). Then we have

\[ EY_c(i\omega_k)Y_c(i\omega_l) = \Phi_c(i\omega_k)\delta_{k+l} + \frac{1}{T} K(i\omega_k, i\omega_l). \]

where \( K \) is bounded and \( \Phi_c \) is defined as in (2.73). Observe that \( k, l \in \mathbb{Z} \) such that a negative \( l \) would represent the complex conjugate such that \( \bar{Y}_c(i\omega_l) = Y_c(-i\omega_l) \).
Proof: First, since $e^{i\omega T_s} = 1$ for the particular choice of $\omega$ in (2.156) we have

$$
Y_c(i\omega) = \frac{C(i\omega I - A)^{-1}}{\sqrt{T}} \left( B \int_0^T e^{-i\omega t}dW(t) + x(0) - x(T) \right)
$$

$$
= \frac{C(i\omega I - A)^{-1}}{\sqrt{T}} \left[ \int_0^T \left( e^{-i\omega t} - e^{A(T-t)} \right) BdW(t) + (I - e^{AT}) x(0) \right].
$$

This means that since $Ex_0e_x = 0$ and $Ede_tde_s = \delta(t - s)dt$ we have

$$
EY_c(i\omega_k)Y_c(i\omega_l) = \frac{C(i\omega_k I - A)^{-1}}{\sqrt{T}} E(T) \frac{(i\omega_l I - AT)^{-1}C^T}{\sqrt{T}}
$$

where

$$
E(T) = \int_0^T \left( e^{-i\omega_k t} - e^{A(T-t)} \right) \sigma^2 BB^T \left( e^{-i\omega_l t} - e^{A(T-T-t)} \right) dt
$$

(2.164)

$$
+ (I - e^{AT}) Ex(0)x(0)^T \left( I - e^{AT}T \right).
$$

(2.165)

The term inside the brackets will become

$$
E(T) = \sigma^2 BB^T \int_0^T e^{-i(\omega_k + \omega_l)t} dt - \sigma^2 BB^T \int_0^T e^{-i\omega_k t} e^{AT(T-t)} dt
$$

$$
- \int_0^T e^{-i\omega_l t} e^{A(T-t)} BB^T \sigma^2 dt + \int_0^T e^{A(T-t)} BB^T \sigma^2 e^{AT(T-t)} BB^T dt
$$

$$
+ Ex(0)x(0)^T - e^{AT} Ex(0)x(0)^T = Ex(0)x(0)^T e^{AT} + e^{AT} Ex(0)x(0)^T e^{AT}.
$$

Fortunately, since $x_t$ is a stationary continuous-time stochastic process (Davis, 1998)

$$
P_x(t) = Ex(t)x^T(t)
$$

$$
P_x(T) = e^{AT} P_x(0)e^{AT} + \int_0^T e^{A(T-t)} BB^T \sigma^2 e^{AT(T-t)} BB^T dt.
$$

In the stationary case $P_x(T) = P_x(0) = P_x$ where $P_x$ is the constants $n \times n$ matrix solution to the Lyapunov equation

$$
AP_x + P_x A^T + \sigma^2 BB^T = 0.
$$

Utilizing this expression, it is now possible to show that

$$
\int_0^T e^{A(T-t)} BB^T \sigma^2 e^{AT(T-t)} BB^T dt
$$

$$
+ Ex(0)x(0)^T - e^{AT} Ex(0)x(0)^T = Ex(0)x(0)^T e^{AT} + e^{AT} Ex(0)x(0)^T e^{AT}.
$$

$$
= (I - e^{AT})P_x + P_x(I - e^{AT})
$$
and
\[
\sigma^2 BB^T \int_0^T e^{-i\omega_k t} e^{A^T (T-t)} dt + \int_0^T e^{-i\omega_l t} e^{A (T-t)} dt BB^T \sigma^2
\]
\[
= \sigma^2 BB^T \left( e^{A^TT} - I \right) (i\omega_k I - A)^{-1} + (i\omega_l I - A)^{-1} \left( e^{AT} - I \right) BB^T \sigma^2.
\]

Finally, since \( \omega_k \) and \( \omega_l \) are defined on the grid (2.156) we get
\[
\int_0^T e^{-i(\omega_k + \omega_l) t} dt = T \delta_{k+l}.
\]

This means that
\[
EY_c(i\omega_k)Y_c(i\omega_l) = \Phi_c(i\omega_k)\delta_{k+l} + \frac{1}{T} K(i\omega_k, i\omega_l)
\]
where
\[
\Phi_c(i\omega_k) = C(i\omega_k I - A)^{-1} \sigma^2 BB^T (i\omega_k - A)^{-1} C^T
\]
and
\[
K(i\omega_k, i\omega_l) =
- C(i\omega_k I - A)^{-1} \left[ (i\omega_k I + A)^{-1} \sigma^2 BB^T + P_x \right] \frac{e^{A^TT} - I}{T}
+ \frac{e^{AT} - I}{T} \left[ (i\omega_l I + A)^{-1} BB^T \sigma^2 + P_x \right] \left( i\omega_k - A \right)^{-1} C^T
\]
since \( A \) is a stable matrix with no eigenvalues on the imaginary axis \( (i\omega I \pm A)^{-1} \) and \( e^{AT} \) will be bounded. Hence \( K \) will also be bounded.

From the previous result it is possible to derive an expression for the covariance matrix of the real and imaginary parts of the frequency components.

**Theorem 2.2**

Assume that \( Y_c(i\omega) \) is defined as in (2.153) and \( \omega_k > 0 \) and \( \omega_l > 0 \) are defined as in (2.156). Then
\[
E \begin{pmatrix} \text{Re } Y_c(i\omega_k) \\ \text{Im } Y_c(i\omega_k) \end{pmatrix} \begin{pmatrix} \text{Re } Y_c(i\omega_l) \\ \text{Im } Y_c(i\omega_l) \end{pmatrix}^T = \begin{pmatrix} \Phi_c(i\omega_l) \frac{2}{\Phi_c(i\omega_k)} \frac{0}{\Phi_c(i\omega_k)} \end{pmatrix} \delta_{k+l} + \frac{1}{T} K(i\omega_k, i\omega_l),
\]
where \( K \) is a bounded \( 2 \times 2 \) matrix.
Proof: Since

\[\text{Re}Y_c(i\omega) = \frac{Y_c(i\omega) + Y_c(-i\omega)}{2}\]
\[\text{Im}Y_c(i\omega) = \frac{Y_c(i\omega) - Y_c(-i\omega)}{2i}\]

the elements of the covariance matrix will be

\[\text{Re}Y_c(i\omega_k)\text{Re}Y_c(i\omega_l) = \left[\frac{Y_c(i\omega_k) + Y_c(-i\omega_k)}{2}\right] \left[\frac{Y_c(i\omega_l) + Y_c(-i\omega_l)}{2}\right]
= \frac{Y_c(i\omega_k)Y_c(i\omega_l) + Y_c(-i\omega_k)Y_c(-i\omega_l)}{4}
+ \frac{Y_c(-i\omega_k)Y_c(i\omega_l) + Y_c(-i\omega_l)Y_c(-i\omega_k)}{4}.
\]

The frequencies are positive \(\omega_k > 0\) and \(\omega_l > 0\) and therefore Theorem 2.1 states that

\[EY_c(i\omega_k)Y_c(i\omega_l) = \frac{1}{T}K(i\omega_k, i\omega_l)\]
\[EY_c(-i\omega_k)Y_c(-i\omega_l) = \frac{1}{T}K(-i\omega_k, -i\omega_l)\]
\[EY_c(i\omega_k)Y_c(-i\omega_l) = \Phi_c(i\omega_k)\delta_{k,l} + \frac{1}{T}K(i\omega_k, -i\omega_l)\]
\[EY_c(-i\omega_k)Y_c(i\omega_l) = \Phi_c(i\omega_k)\delta_{k,l} + \frac{1}{T}K(-i\omega_k, i\omega_l)\]

Hence

\[E\text{Re}Y_c(i\omega_k)\text{Re}Y_c(i\omega_l) = \frac{\Phi_c(i\omega_k)}{2}\delta_{k,l} + \frac{1}{T}K(i\omega_k, i\omega_l).\]

The other elements of the covariance matrix will follow analogously.

According to the theorem above

\[
E \begin{pmatrix}
\text{Re} Y_c(i\omega_1) \\
\text{Im} Y_c(i\omega_1) \\
\text{Re} Y_c(i\omega_2) \\
\text{Im} Y_c(i\omega_2)
\end{pmatrix} = 
\begin{pmatrix}
\Phi_c(i\omega_1) + \frac{1}{T}K & \frac{1}{T}K & \frac{1}{T}K & \frac{1}{T}K \\
\frac{1}{T}K & \Phi_c(i\omega_1) + \frac{1}{T}K & \frac{1}{T}K & \frac{1}{T}K \\
\frac{1}{T}K & \frac{1}{T}K & \Phi_c(i\omega_2) + \frac{1}{T}K & \frac{1}{T}K \\
\frac{1}{T}K & \frac{1}{T}K & \frac{1}{T}K & \Phi_c(i\omega_2) + \frac{1}{T}K
\end{pmatrix}.
\]

(2.166)

In the remaining part of the thesis it will be assumed that \(T\) is large enough that the effect of the \(K\) terms in expression (2.166) can be ignored. This means that the truncated Fourier transforms of \(y\) at the grid (2.156) are assumed to be independent and Gaussian with the distribution

\[Y_c(i\omega_k) \sim \mathcal{N}(0, \Phi_c(i\omega_k)).\]
Therefore the approximate likelihood function for the values of the truncated Fourier transform
\[ \{ Y_c(i\omega_1), Y_c(i\omega_2), \ldots, Y_c(i\omega_{N_\omega}) \} \]
on the frequency grid (2.156) is
\[ p(Y_c(i\omega_1), Y_c(i\omega_2), \ldots, Y_c(i\omega_{N_\omega}) | \theta) = \prod_{k=1}^{N_\omega} \frac{1}{2\pi \Phi_c(i\omega_k, \theta)} e^{-\frac{|Y_c(i\omega_k)|^2}{\Phi_c(i\omega_k, \theta)}}. \]
if the effects of the term \( \frac{1}{T} K \) is ignored. The negative log likelihood function will be
\[ L(\theta) = -\log p(Y_c(i\omega_1), Y_c(i\omega_2), \ldots, Y_c(i\omega_{N_\omega}) | \theta) \]
\[ = N_\omega \log 2\pi + \sum_{k=1}^{N_\omega} \frac{|Y_c(i\omega_k)|^2}{\Phi_c(i\omega_k, \theta)} + \log \Phi_c(i\omega_k, \theta). \]
Suppose now that a whole continuous time realization \( \{ y(t) : t \in [0, T] \} \) of the output of (2.31) is known. Define the \textit{continuous-time periodogram} of this output as
\[ \hat{\Phi}_c(i\omega) = |Y_c(i\omega)|^2. \] (2.167)
where \( Y_c \) is given by (2.153). The approximate Maximum Likelihood procedure (approximate since the terms proportional to \( \frac{1}{T} \) are ignored) for estimating the parameters is then
\[ \hat{\theta} = \arg \min_{\theta} L(\theta, \hat{\Phi}_c) \]
(2.168)
where
\[ L(\theta, \hat{\Phi}_c) = \sum_{k=1}^{N_\omega} \frac{\hat{\Phi}_c(i\omega_k)}{\Phi_c(i\omega_k, \theta)} + \log \Phi_c(i\omega_k, \theta), \]
(2.169)
which should be compared to (2.149). Before we close this section we would like to recommend the publications by Pintelon and Schoukens (2001), Brillinger (1981) and Dzhaparidze (1970) which also treat the issue of identification of discrete-time and continuous-time noise models.

2.7.4 Cramer-Rao Lower Bound

A property of the Maximum Likelihood estimator is that it is asymptotically unbiased (Cramér, 1946) under certain mild conditions
\[ \lim_{N_t \to \infty} \hat{\theta} \to \theta_0 \]
where \( \hat{\theta} \) is the estimate and \( \theta_0 \) are the true parameters. The quality of an estimator can be measured by its error covariance matrix
\[ P = E(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^T \]
and it is interesting to know how accurate these estimates can become. A lower limit for unbiased estimators is the Cramér-Rao lower bound (CRLB) (Cramér, 1946; Rao, 1946) which states that

\[ P \geq J^{-1}(\theta_0). \]

Here

\[ J(\theta_0) = E \left[ \frac{d}{d\theta} \log p(Y|\theta) \right] \left[ \frac{d}{d\theta} \log p(Y|\theta) \right]^T \Big|_{\theta=\theta_0} \]

is known as the Fisher information matrix and \( p(Y|\theta) \) is the likelihood function expressing the probability of the measurements \( Y \) provided that the true parameters are \( \theta_0 \).

An explicit expression for the Cramér-Rao lower bound for time series models known as the Slepian-Bangs formula (Slepian, 1954)

\[ \{J(\theta_0)\}_{ij} = \frac{1}{2} Tr \left( R(\theta)^{-1} \frac{\partial R(\theta)}{\partial \theta_i} R(\theta)^{-1} \frac{\partial R(\theta)}{\partial \theta_j} \right) \Big|_{\theta=\theta_0} \]

can be derived, where \( R(\theta) \) is the covariance matrix of the output sequence as defined in (2.94).

### 2.8 Summary

First, this chapter has contained a short introduction to the basics of linear systems theory and system identification. There was also a brief summary of the properties of stochastic processes, linear input-output models and time series models. The exposition concerned both discrete and continuous-time aspects, and also covered spectral properties.

The other half of the chapter was dedicated to system identification. Here, the classical time domain maximum likelihood approach based on the Kalman filter was explained to the reader. It was also shown that, under certain conditions, the Kalman filter can be used as an efficient way to factorize the covariance matrix of a stochastic process. This fact will be very useful in Chapter 10 while dealing with non-uniformly sampled stochastic processes.

Finally, frequency domain identification methods were considered. Here, the classical methods of Whittle (1961) and Dzhaparidze (1970) for discrete and continuous-time models, were dealt with in depth.
A Short Review of Direct Continuous-Time System Identification Methods

“Progress, far from consisting of change, depends on retentiveness. Those who cannot remember the past are condemned to repeat it.”

George Santayana (1863-1952)

3.1 Introduction

Parameter identification of continuous-time systems is by no means a new subject. In the days when computers were uncommon, the continuous-time perspective on identification was probably the dominating one. Research on the subject has been steadily going on since the early fifties, and today a number of excellent surveys of the area are available (Unbehauen and Rao, 1990), (Sinha and Rao, 1991) and (Mensler, 2003).

The previous chapters have mostly dealt with the more modern, so called indirect approach to continuous-time identification, where the sampled version of the system has played an important role. There are however also a large number of methods, which do not go via the sampled system, but instead attempts a direct identification of the continuous-time parameters using the sampled data. The present chapter will contain a short and not at all comprehensive review of this line of thought, with the purpose of making the reader familiar and aware of the work in this area.

3.2 Outline

First, in Section 3.3 there will be a short introduction to the two phase framework of the direct methods. Then, in Section 3.4 the approach based on modulating functions is explained. In Section 3.5 and 3.5.1 the linear filtering methods with the classical state
variable filters (SVF) and generalized Poisson moment functionals (GPMF) are introduced. In Section 3.6, 3.6.1 3.6.2 and 3.6.3 the focus is on integration methods such as orthogonal functions, numerical integration and linear integral filters. Finally, in Section 3.7 the simplified revised instrumental variable method (SRIVC) method is explained.

### 3.3 Direct Approach

Consider the linear continuous-time system

\[ \ddot{y}(t) + a_1 \dot{y}(t) + a_0 y(t) = b_1 \dot{u}(t) + b_0 u(t) + v(t) \]  

(3.1)

with the stochastic equation error \( v(t) \). Here, the parameters \( a_0, a_1, b_0 \) and \( b_1 \) are supposed to be estimated from a set of discrete-time measurements \( \{ y(t_k) \}_{k=1}^{N_t} \) and \( \{ u(t_k) \}_{k=1}^{N_t} \) of the deterministic input function and the stochastic output.

Assume for one moment that we do not only know the value of the input and the output at the sampling points, but at all points \( t \in [t_1, t_{N_t}] \). If we also know the value of all of the derivatives of the input and output, the expression in (3.1) will generate an infinite and over determined system of equations for the parameters.

Of course, the above method suffers from several drawbacks. Apart from the possible problems associated with the infinite dimensionality of \( y(t) \) and \( u(t) \), there is also a fundamental difficulty associated with the estimation of the time-derivatives. Consequently, all direct methods can be seen as ways of resolving these issues using an approach based on two steps. First, a primary stage, where the differential equation for the continuous-time system is transformed into a system of equations

\[ y_2(k) + a_1 y_1(k) + a_0 y_0(k) = b_1 u_1(k) + b_0 u_0(k) + v(k) \quad k = 1, \ldots, N_t \]  

(3.2)

by the application of some special functional to the equation in (3.1). Here \( y_0(k), y_1(k), y_2(k), u_0(k) \) and \( u_1(k) \) are the real or complex valued results from the transformation of the derivatives of the input and output. There are basically three different direct approaches for constructing this functional (Mensler, 2003; Unbehauen and Rao, 1987): the method of modulating functions, linear filtering methods and integration methods. These methods will be discussed further below.

The resulting equation in (3.2) is now algebraic (does not involve differentials) and can be put into a regression form, such that

\[ y_2(k) = \phi(k)^T \theta + v(k) \quad k = 1, \ldots, N_t \]  

(3.3)

where

\[ \phi(k)^T = (-y_1(k) - y_0(k) u_1(k) u_0(k)) + v(k) \]  

(3.4)

\[ \theta = (a_1 a_0 b_1 b_0)^T. \]  

(3.5)

In a secondary stage, the parameters are then identified by an adequate statistical estimation procedure from the system of regression equations in (3.3). A classical approach...
has been to use the least-squares (LS) method (Rao, 1973; Draper and Smith, 1981; Daniel and Wood, 1980), which in the case of equation (3.3) will become

$$\hat{\theta}_{Nt}^{LS} = \left( \frac{1}{N_t} \sum_{k=1}^{N_t} \phi(k) \phi^T(k) \right)^{-1} \frac{1}{N_t} \sum_{k=1}^{N_t} \phi(k) y_2(k). \quad (3.6)$$

Normally, this estimate will suffer from bias because the sequence \( \{v(k)\}_{k=1}^{N_t} \) will not be white. Instrumental variable (IV) methods have therefore become popular as a means for reducing this bias (Young, 1981). The main idea of the IV method is that the prediction error i.e.

$$\epsilon(n) = y_2(n) - \phi(n)^T \theta \quad (3.7)$$

should be independent of past data \( y(n - k), u(n - k) \) \( k = 1 \ldots n \). This would in turn mean that all nonlinear transformations of \( \epsilon(n) \) should be uncorrelated with past data. Such an approach is impossible for obvious reasons, and an alternative is to demand certain transformations of the data \( \zeta(k) \) and the prediction error \( \alpha(\epsilon(k, \theta)) \) to be uncorrelated. That is

$$\frac{1}{N_t} \sum_{k=1}^{N_t} \zeta(k) \alpha(\epsilon(k, \theta)) = 0. \quad (3.8)$$

These ideas applied to linear regression are known as instrumental variable methods and the elements of \( \zeta(k) \) are called instruments. This gives the estimates

$$\hat{\theta}_{Nt}^{IV} = \left( \frac{1}{N_t} \sum_{k=1}^{N_t} \phi(k) \zeta^T(k) \right)^{-1} \frac{1}{N_t} \sum_{k=1}^{N_t} \phi(k) y_f^{(n)}(k)^T \quad (3.9)$$

The choice of instruments is often particular to the problem at hand. For a more thorough treatment of the class of instrumental variable methods, please consult the excellent text by Söderström and Stoica (1983). In the sections below we will now address various ways of generating the algebraic equations used for estimation.

### 3.4 Modulating Functions Methods

An \( n \)th order modulating function \( \phi(t) \) is a smooth \( n - 1 \) times differentiable function such that

$$\phi^{(i)}(0) = 0$$

$$\phi^{(i)}(T) = 0$$

for \( i = 0, 1, \ldots, n - 1 \) on the interval \([0, T]\) (Shinbrot, 1957). Here \( \phi^{(i)}(t) \) denotes the \( i \)th derivative with respect to time.

The purpose of choosing the functions as above is to avoid the effect of initial conditions. If the differential equation in (3.1) is multiplied by such a function and is then
integrated over the interval it will yield
\[
\begin{align*}
\int_0^T \dddot{y}(t)\phi(t)\,dt + a_1 \int_0^T \ddot{y}(t)\phi(t)\,dt + a_0 \int_0^T \dot{y}(t)\phi(t)\,dt &= b_1 \int_0^T \ddot{u}(t)\phi(t)\,dt + b_0 \int_0^T \dot{u}(t)\phi(t)\,dt + \int_0^T v(t)\phi(t)\,dt.
\end{align*}
\]

However, because of the conditions put on the modulating function, the expression above can be rewritten by partial integration as
\[
\begin{align*}
\int_0^T \dddot{y}(t)\phi(t)\,dt - a_1 \int_0^T \ddot{y}(t)\phi(t)\,dt + a_0 \int_0^T \dot{y}(t)\phi(t)\,dt &= \gamma y_0 \gamma (\theta + \epsilon_n) \\
\epsilon_n &= \int_0^T v(t)\phi_n(t)\,dt
\end{align*}
\]

since the modulating functions are supposed to be \( n \) times differentiable. It is then possible to evaluate the integrals numerically using the fast Fourier transform to produce an algebraic equation (Pearson and Shen, 1993). If the integral expressions are defined as
\[
\begin{align*}
\gamma_{n-i} &= (-1)^i \int_0^T y(t)\phi_n^{(i)}(t)\,dt \\
\epsilon_n &= \int_0^T v(t)\phi_n(t)\,dt
\end{align*}
\]
the equation in (3.10) can be written as the regression equation
\[
\gamma y_0 = \gamma \theta + \epsilon_n
\]

where
\[
\gamma = [-\gamma_i \ldots - \gamma_n \gamma_{n-m} \ldots \gamma_{n-m}] \\
\theta = [a_n \ldots a_0 b_m \ldots b_0].
\]

Doing this over number of different modulating functions of the same class, a set of regression equations can be formed and parameters can then be estimated (Mensler, 2003). Two such classes will be described below.

One of two classes of functions that are usually mentioned in connection with the modulating functions method are the Fourier modulating functions
\[
\phi_{\mu,n}(t) = \frac{1}{T} e^{-i\mu\omega_0 t} (e^{-i\mu\omega_0 t} - 1)^n.
\]
The other class are the Hartley modulating functions

\[ \psi_{\mu,n}(t) = \sum_{k=0}^{n} (-1)^k \binom{k}{n} (\cos(n + \mu - k)\omega_0 t + \sin(n + \mu - k)\omega_0 t). \]

### 3.5 Linear Filter Methods

The linear filter approaches all have in common that they apply linear filters to the input and output signals in order to obtain a set of algebraic equations. This procedure is illustrated as

\[
F(p)\dddot{y}(t) + b_1 F(p)\dot{u}(t) + b_0 F(p)u(t) + F(p)v(t) = b_1 F(p)\dot{u}(t) + b_0 F(p)u(t) + F(p)v(t)
\]

(3.12)

where the filter \( F \) has been applied to (3.1). Here, \( F(p) \) is a continuous-time all-pole filter where \( p \) is the differentiation operator. It is usually defined as

\[
F(p) = \frac{1}{L(p)},
\]

(3.13)

where \( L \) is a polynomial with order greater than or equal to \( n \) to avoid direct differentiation. The parameter \( n \) is the maximum number of differentiations present in the differential equation. In this case \( n = 2 \).

The polynomial \( L \) can be easily compared to an observer polynomial, but the design uses a minimum of a priori information about the system. Basically, the approximate bandwidth of the system together with the order of the differential equation (Young, 1970). A particular choice is

\[
F(p) = \left( \frac{\lambda}{p + \lambda} \right)^n,
\]

which allows the user to tune the bandwidth of the filter through the parameter \( \lambda \) (Garnier et al., 2003). Here \( n \) is selected equal to the order of the system. For such a class of filters (3.12) can now be written as

\[
\frac{p^2}{L(p)} y(t) + a_1 \frac{p}{L(p)} y(t) + a_0 \frac{1}{L(p)} y(t) = b_1 \frac{p}{L(p)} u(t) + b_0 \frac{1}{L(p)} u(t) + \frac{1}{L(p)} v(t)
\]

where

\[
\frac{p^n}{L(p)}.
\]

are known as the state variable filters. If the outputs of these filters are denoted \( y_n(t), u_n(t) \) and \( v_n(t) \) and the effects of initial conditions are ignored, then (3.12)

\[
y_2(t) + a_1 y_1(t) + a_0 y_0(t) = b_1 u_1(t) + b_0 u_0(t) + v_0(t).
\]

(3.14)
This equation can be put in regression form

\[ y_2(t) = \phi(t)^T \theta + v_0(t) \]

where

\[ \phi(t)^T = [-y_1(t) - y_0(t) u_1(t) u_0(t)] \]

\[ \theta^T = [a_1 a_0 b_1 b_0] \].

### 3.5.1 Poisson Moment Functionals

A close relative to the state variable filter approach is the method of generalized Poisson moment functionals. Here a signal such as \( y \) is treated as a generalized function, a distribution, and can be expanded about instant \( t_0 \) as

\[ y(t') = \sum_{k=0}^{\infty} M_k[y(t)] e^{\lambda(t-t_0)} \delta^{(k)}(t-t_0), \]

where \( \delta^{(k)} \) are derivatives of the Dirac distribution (Saha and Rao, 1983). The coefficients of this representation are then computed as

\[ M_k[y] \triangleq \int_0^{t_0} y(t) m_k(t_0 - t) \, dt, \]

where

\[ m_k(t_0) \triangleq \beta^{k+1} t_0^k \frac{e^{-\lambda t}}{k!}. \]

and \( \beta \) and \( \lambda \) are tuning parameters. Here \( M_k \) is called the Poisson moment functional of order \( k \). An attractive feature of the method is that the integrals involved in the functionals can be obtained as the outputs of a cascade of filters as illustrated in Figure 3.1. For these

![Figure 3.1: Poisson Filter Chain](image)

\( y(t) \)

\( \frac{\beta}{s+\lambda} \)

\( \frac{\beta}{s+\lambda} \)

\( \frac{\beta}{s+\lambda} \)

\( M_0[y(t)] \)

\( M_1[y(t)] \)
filters there is a relationship between the derivatives of the signal $y$ such that

$$M_l[y^{(n)}(t)] = \sum_{j=0}^{i} (-1)^j \left( \begin{array}{c} j \\ i \end{array} \right) \beta^{i-j} \lambda^j M_{l-i+j}[y(t)]$$

$$- \sum_{q=1}^{i} \left[ y^{(q-1)}(0) \sum_{j=0}^{i-q} (-1)^j \left( \begin{array}{c} j \\ i-q \end{array} \right) \beta^{i-q-j} \lambda^j p_{l-i+j+q}(t) \right].$$

(3.15) (3.16)

Moment functionals $M_l$ of a derivative of $y$ or $u$ can hence be written in terms of lower order moment functionals of the original signals. Ignoring the effect if initial conditions we would for instance have (Mensler, 2003)

$$\begin{pmatrix} M_l[y^{(2)}(t)] \\ M_l[y^{(1)}(t)] \\ M_l[y(t)] \end{pmatrix} = \begin{pmatrix} \beta^2 & -2\lambda \beta & \lambda^2 \\ 0 & \beta & -\lambda \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} M_{l-2}[y(t)] \\ M_{l-1}[y(t)] \\ M_l[y(t)] \end{pmatrix}.$$ 

If the moment functional $M_n$ is applied to the test equation in (3.1) such that

$$M_l[\ddot{y}(t)] + a_1 M_l[\dot{y}(t)] + a_0 M_l[y(t)] = b_1 M_l[\dot{u}(t)] + b_0 M_l[u(t)] + M_l[v(t)],$$

it is by (3.15) possible to transform this system into a regression form

$$M_l[y^{(2)}(t)] = \phi(t)\theta + M_l[v(t)]$$

which can be used to produce a least squares estimate. The procedure is as mentioned earlier very similar to the state variable filter approach.

### 3.6 Integration Methods

In the integration approach the entire differential equation in (3.1) is integrated, say $n = 2$ times creating an expression such as

$$y(t) + a_1 y^{[1]}(t) + a_0 y^{[2]}(t) = b_1 u^{[1]}(t) + b_0 u^{[2]}(t) + c_1 \frac{t^1}{1!} + c_0 \frac{t^2}{2!} + v^{[2]}(t),$$

(3.17)

where

$$y^{[k]}(t) = \int \cdots \int y(t)dt$$

and $c_0$ and $c_1$ account for the effect of initial conditions (Whitfield and Messali, 1987). In order to produce algebraic equations that can be used to estimate parameters, the integrals have to be computed numerically. The different approaches below are basically different ways to perform this evaluation from samples of the input-output data.
3.6.1 Orthogonal Functions

In the orthogonal functions approach, the input $u(t)$ and output signals $y(t)$ are represented as

$$y(t) = \sum_{k=0}^{\infty} y_k \phi_k(t).$$

in a base of functions $\{\phi_k\}_{k=0}^{\infty}$ which is orthogonal with respect to a weighting function $w(x) \geq 0$ (Råde and Westergren, 1995). The coefficients $y_k$ are identified by projecting the function $y(t)$ onto the basis by evaluating scalar products like

$$y_k = \frac{\int_0^T w(t)y(t)\phi_k(t)dt}{\int_0^T w(t)\phi_k(t)\phi_k(t)dt}$$

numerically from samples of $y(t)$. The series representation is always truncated such that

$$y(t) \approx \sum_{k=0}^{N} y_k \phi_k(t) = y^T \phi(t)$$

with

$$y = (y_0 \ y_1 \ldots \ y_N)^T$$

$$\phi(t) = (\phi_0(t) \ \phi_1(t) \ldots \ \phi_N(t))^T.$$

This operation will automatically reduce the high frequency content of the representation and will therefore be an important design choice.

When the differential equation in (3.1) is in the integrated form (3.17)

$$y(t) + a_1 y^{[1]}(t) + a_0 y^{[2]}(t) = b_1 u^{[1]}(t) + b_0 u^{[2]}(t) + v^{[2]}(t)$$

the new representation allows it to be converted into an algebraic equation. The device which facilitates this is the so called operational matrix (Chen and Hsiao, 1975) for integration $P$, which has the property that

$$\int_0^t \phi(t)dt \approx P\phi(t).$$

This relation shows that integrals of the basis functions can be written as linear combinations of the basis functions themselves. If applied to the integrated differential equation above the operational matrix will produce

$$y^T \phi(t) + a_1 y^T P \phi(t) + a_0 y^T P^2 \phi(t) = b_1 y^T P \phi(t) + b_0 y^T P^2 \phi(t) + v^T P^2 \phi(t).$$

Equating the coefficients will then give a set of algebraic equations

$$y^T + a_1 y^T P + a_0 y^T P^2 = b_1 y^T P + b_0 y^T P^2 + v^T P^2,$$
which can be written in the form of linear regression

\[ y = \Psi w, \]

where \( w = [a_1 \ a_0 \ b_1 \ b_0] \) and \( \Psi \) depends on \( y \) and \( P \). That equation can then be used to estimate the parameters by least-squares or instrumental variable methods. Some of the orthogonal basis functions found in the literature are the Laguerre polynomials \( \text{(Hwang and Shih, 1982)} \), Fourier series \( \text{(Chung, 1987)} \), block pulse functions \( \text{(BPF)(Palanisamy and Bhattacharyya, 1981)} \) and Legendre polynomials \( \text{(Chang and Wang, 1982)} \).

### 3.6.2 Numerical Integration Methods

Since the continuous-time realizations of the input and output are not available, discrete-time signals are integrated numerically using piecewise constant or piecewise linear interpolation. These two types of interpolation also comes under the name of the block pulse functions \( \text{(BPF)} \) or the trapezoidal pulse functions \( \text{(TPF)} \). For the trapezoidal rule with equidistant sampling interval \( T_s \) the integration can be realized by the following expression

\[
y^{[k]}(t_i) = y^{[k]}(t_{i-1}) + T_s \frac{y^{[k-1]}(t_i) + y^{[k-1]}(t_{i-1})}{2}
\]

\[
u^{[k]}(t_i) = u^{[k]}(t_{i-1}) + T_s \frac{u^{[k-1]}(t_i) + u^{[k-1]}(t_{i-1})}{2}.
\]

A set of algebraic equations in a linear regression form can ultimately be derived by evaluating the integrals over different time intervals \( \text{(Whitfield and Messali, 1987)} \). When algebraic equations are present parameters can be readily estimated by for instance a least squares or instrumental variable approach \( \text{(Söderström and Stoica, 1983)} \).

### 3.6.3 Linear Integral Filter Methods

The linear integral filter approach \( \text{(LIF)} \) is closely related to the numerical integration methods \( \text{(Sagara and Zhao, 1990)} \). Numerically integrating a function over a small interval \([t, t - lT_s]\) can be written as

\[
y^{[1]} = \int_{t-lT_s}^t y(\tau) d\tau = \sum_{k=0}^l f_k y(t - kT_s),
\]

where \( f_i \) originate from a quadrature scheme such as the trapezoidal rule. Using the delay operator

\[ q^{-1} y(t) = y(t - T_s) \]

this relation can be represented as

\[ y^{[1]} = \sum_{k=0}^l f_k q^{-k} y(t). \quad (3.18) \]
The central result of the theory of linear integral filters is that if $y^{(j)}$ denotes the $j$:th derivative of $y(t)$ then $y^{(j)}(t)$ integrated $n$ times can be written as

$$I_n y^{(j)} \approx P_j y(t) = \sum_{k=0}^{nl} p_j^k q^{-k} y(t)$$

where $I_n$ is the operator for integrating $n$ times and

$$P_j = (1 - q^{-1})^j \sum_{k=0}^{l} f_k q^{-k} n^{-j}.$$ 

Here $p_j^i$ are real valued coefficients that have little to do with the differentiation operator $p$.

The method above avoids the problems of initial values encountered in the numerical integration approach. It can also be interpreted as linear filtering approach since multiple integration over the small interval is equivalent to pre-filtering with

$$L(s) = \frac{1 - e^{-sT_s l}}{s^n}.$$ 

This view supplies the user with $l$ as a tuning parameter for the filter. As for state variable filters and Poisson modulating functions it should be chosen in such a way that the bandwidth of the filter closely matches the bandwidth of the system.

By applying multiple integration to the system to be estimated a linear regression form can be derived. By varying the interval of integration $[t - T_s l]$ a set of algebraic equations can be produced and parameters can subsequently be estimated within a least squares or instrumental variable framework.

### 3.7 Simplified Refined Instrumental Variable Method

An algorithm that have proved itself both in practice and theory is the continuous-time simplified refined instrumental variable method (SRIVC)(Young and Jakeman, 1980). Assume for a minute that we have the model

$$y(t) = \frac{B(p, \theta)}{A(p, \theta)} u(t)$$

$$y_m(t_k) = y(t_k) + e(t_k)$$

where $e(t_k)$ is a white noise sequence. The the standard prediction error framework i.e. (Ljung, 1999) states that we should then estimate the parameters by minimizing the sum of the squares of the prediction errors

$$e(t_k) = y(t) - \left. \frac{B(p, \theta)}{A(p, \theta)} u(t) \right|_{t=t_k}.$$ 

Now, also assume that we know the polynomial $A(p, \theta)$. Then, we can rewrite (3.21) as

$$\epsilon(t) = \frac{1}{A(p)} (A(p)y(t) - B(p)u(t)).$$
By the application of the property that linear time-invariant systems commute, we can by further reformulation show that

\[ \epsilon(t) = A(p, \theta)y_f(t) - B(p, \theta)u_f(t) \]  

(3.23)

where

\[ y_f(t) = \frac{1}{A(p, \theta)}y(t) \]  

(3.24)

\[ u_f(t) = \frac{1}{A(p, \theta)}u(t). \]  

(3.25)

If we expand the expression in (3.23) further, we will then get

\[ \epsilon(t) = y_f^{(n)}(t) + a_{n-1}y_f^{(n-1)}(t) + \cdots + a_0y_f(t) \]

\[ - b_m u_f^{(m)}(t) - \cdots - b_0 u_f^{(0)}(t) \]  

(3.26)

(3.27)

where

\[ y_f^{(k)}(t) = p_k^k A(p, \theta)y(t), \quad k = 1 \ldots n. \]  

(3.28)

and the estimation error is now linear in the parameters. Hence we can rewrite (3.26) in linear regression form

\[ y_f^{(n)}(t_k) = \phi_f^T(t_k)\theta + \epsilon(t_k), \quad k = 1 \ldots N \]  

(3.29)

where

\[ \phi_f(t_k, \theta) = \left( -y_f^{(0)}(t_k) \ldots y_f^{(n-1)}(t_k) u_f^{(0)}(t_k) \ldots u_f^{(m)}(t_k) \right)^T \]  

(3.30)

and

\[ \theta = (a_0 \ldots a_{n-1} b_0 \ldots a_m)^T. \]  

(3.31)

The parameters can then be estimated by the least squares method such that

\[ \hat{\theta}_{LS}^N = \left( \frac{1}{N} \sum_{k=1}^N \phi_f(t_k) \right) \left( \frac{1}{N} \sum_{k=1}^N \phi_f(t_k) y_f^{(n)}(t_k)^T \right)^{-1} \]  

(3.32)

In practice \( \hat{\theta}_{LS}^N \) will however become biased. First of all, because we obviously do not know \( A(p, \theta) \) and instead we have pseudo-linear regression in the form of

\[ y_f^{(n)}(t_k, \theta) = \phi_f^T(t_k, \theta)\theta + \epsilon(t_k). \]  

(3.33)

The second reason is that the noise sequence \( \epsilon(t_k) \) can be far from white.

A popular way to resolve these issues is to start out with an initial guess for \( A(p, \theta) \). Then, compute \( \{u_f\}_{k=1}^{N_t} \) and \( \{y_f\}_{k=1}^{N_t} \), and estimate the parameters \( \theta_{LS} \) using the least
squares method. When these parameters are found, a prefiltered instrumental variable is generated by the model

\[ \hat{x}_f(t) = \frac{B(p, \theta_{LS}^N)}{A(p, \theta_{LS}^N)} u_f(t). \]  

(3.34)

where

\[ u_f(t) = \frac{1}{A(p, \theta_{LS}^N)} u(t). \]  

(3.35)

The associated IV vectors

\[ \zeta_f(t_k) = \left( -\hat{x}_f^{(0)}(t_k) \ldots - \hat{x}_f^{(n-1)}(t_k) u_f^{(0)}(t_k) \ldots u_f^{(m)}(t_k) \right)^T, \quad k = 1 \ldots N_t \]  

(3.36)

are then constructed, and an IV estimate found by

\[ \hat{\theta}_{IV}^N = \left( \frac{1}{N} \sum_{k=1}^{N} \phi_f(t_k) \zeta_f^T(t_k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \phi_f(t_k) y_f^{(n)}(t_k)^T. \]  

(3.37)

The parameters \( \theta_{IV}^N \) are the used iteratively in order to generate \( \{u_f\}_{k=1}^{N_t} \) and \( \{y_f\}_{k=1}^{N_t} \) and new pre-filtered instrumental variables

\[ \hat{x}_f(t) = \frac{B(p, \theta_{IV}^N)}{A(p, \theta_{IV}^N)} u_f(t). \]  

(3.38)

This procedure is repeated a couple of times, and parameter estimates are then retrieved.

3.8 Summary

In this chapter a variety of different approaches for identifying continuous-time systems have been illustrated. A common denominator for these time domain methods seems to be an effort to avoid estimating the continuous-time derivatives directly. Almost all approaches presented there include, directly or indirectly, some form of low pass filtering, mainly in order to attenuate high frequency noise that would otherwise be much amplified by the differentiation operation.
Identification of Input-Output Models from Uniformly Sampled Data

“If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.”

John von Neumann (1903 - 1957)

4.1 Introduction

In Section 2.7.1 a straightforward method for frequency domain identification of a continuous-time input output models was introduced. The steps in the process were the following:

First, equidistantly sampled input and output data \( \{ u(kT_s), y(kT_s) \}_{k=1}^{N_t} \) was acquired.

Then, the discrete-time Fourier transforms of the sampled data was computed as follows

\[
\begin{align*}
U(e^{i\omega_n T_s}) &= T_s \sum_{k=1}^{N_t} u(kT_s)e^{i\omega_n T_s}, \quad \omega_n = \frac{2\pi}{T_s} n, \ n = 0, \ldots, \left\lfloor \frac{N_t-1}{2} \right\rfloor \\
Y(e^{i\omega_n T_s}) &= T_s \sum_{k=1}^{N_t} y(kT_s)e^{i\omega_n T_s}.
\end{align*}
\] (4.1)

After that, the parameters would be identified by minimizing the sum of the square of the difference of the measured and expected frequency response as follows (Pintelon et al., 2004)

\[
\hat{\theta} = \arg \min_{\theta} \frac{1}{N_\omega} \sum_{k=1}^{N_\omega} \left| Y(e^{i\omega_k T_s}) - G_d(e^{i\omega_k T_s}, \theta)U(e^{i\omega_k T_s}) \right|^2
\] (4.2)

where \( N_\omega \) represented the number of frequency components. The pivotal element of this construction is the discrete-time transfer function \( G_d(z) \) in (2.71) which governs the frequency response \( G_d(e^{i\omega T_s}) \) of the system (therefore relating the frequency content of the
sampled input to that of the sampled output). In some cases, this so called pulse transfer function might be corrupted. As the rate of sampling increase, the relationship between the continuous- and discrete-time system can become more or less ill-conditioned. Mainly, this is due to the gathering of the poles of the discrete-time system around the value 1 in the complex plane. This will produce numerical difficulties while mapping back to the continuous-time parameters. We would therefore like to investigate robust alternatives to the exact $G_d(e^{i\omega T_s})$, which can circumvent such a problem by using the continuous-time frequency response $G_c(i\omega)$ directly. Another, maybe more important reason for studying such approximations is that they will provide insight into how one can deal with non-uniformly sampled data in frequency domain identification.

### 4.2 Outline

Most of the material in this chapter is based on the properties of the discrete-time transfer function $G_d(z)$. This entity, which will be introduced in Section 4.3, is also called the pulse transfer function, since it is the discrete-time transfer function for the magnitudes of individual pulses in the input pulse train.

The relationship between the pulse and continuous-time transfer function can be represented in terms of an infinite summation. In Section 4.4, the effect of truncating this sum during system identification from uniformly sampled data is investigated. The problems associated with the most basic form of truncation, a straight forward replacement of the pulse transfer function by the continuous-time transfer function are also illustrated.

While the method of approximation in Section 4.4 is the truncation of a sum, the means in Section 4.5, 4.6 and 4.7 is that of approximating the transfer function $G_c(s)$. Here, two different methods are derived, where the last one can be interpreted as a form of estimate of the continuous-time Fourier transform of the output $y(t)$.

### 4.3 Pulse Transfer Function

Consider a discrete-time system $G_d$ as in Figure 4.1, consisting of a zero-order hold circuit $H$ at the input, a continuous-time system $G_c$ and a sampling circuit at the output. Then,

![Figure 4.1: Input and sampling setup for a continuous-time system.](image)
The discrete-time pulse transfer function $G_d$ in (2.140) can also be written as

$$G_d(e^{i\omega T_s}, \theta) = \left(1 - e^{-i\omega T_s} \right) \sum_{k=-\infty}^{\infty} \frac{G_c(i\omega + i\frac{2\pi}{T_s} k, \theta)}{i\omega + i\frac{2\pi}{T_s} k}.$$  \hspace{1cm} (4.3)

This result can be found in the classical book by Åström and Wittenmark (1984) on computer controlled systems. However, since the result is central for this chapter, a brief summary of the thoughts leading to it will follow below.

First, define the Dirac “comb” function as

$$m(t) = \sum_{k=-\infty}^{\infty} \delta(t - T_s k)$$

where $\delta(t)$ is the Dirac delta function. Define the operator $\cdot^*$ such that for an arbitrary function $v$

$$v^*(t) = v(t)m(t)$$  \hspace{1cm} (4.4)

and for the Laplace transform

$$V^*(s) = \int_{0}^{\infty} v^*(t)e^{-st}dt.$$  \hspace{1cm} (4.5)

Then, one can denote

$$u^*(t) = u(t)m(t)$$

$$y^*(t) = y(t)m(t)$$

as the sampled versions of the input and output and their Laplace transforms. If the input is assumed to be zero-order hold its sampled version will pass through a hold circuit with transfer function

$$H(s) = \frac{1 - e^{-sT_s}}{s}$$

before entering the system $G$. Let the combination of the hold circuit and the system be defined as

$$F(s) = G(s)H(s).$$

Then, the relationship between the sampled versions of the output is

$$y^*(t) = m(t)F(s)u^*(t)$$

where $p$ is the differentiation operator. If $f$ is the impulse response of the system $F$, the sampled versions of the input and output will be related as

$$y^*(t) = (f(t) * u^*(t))^* = m(t) \int_{-\infty}^{\infty} f(t - \tau)m(\tau)u(\tau)d\tau.$$
At the same time

\[ (f^*(t) * u^*(t)) = \int_{-\infty}^{\infty} m(t - \tau)f(t - \tau)m(\tau)u(\tau)d\tau \]

\[ = \int_{-\infty}^{\infty} m(t)f(t - \tau)m(\tau)u(\tau)d\tau \]

since \( m(t - kT_s) = m(t) \) and \( m(\tau) \neq 0 \) only for \( \tau = kT_s \) (Åström and Wittenmark, 1984). Hence

\[ y^*(t) = (f(t) * u^*(t))^* = (f^*(t) * u^*(t)) \]

and an analogous expression holds for the Laplace transform

\[ Y^*(s) = [F(s) * U^*(s)]^* = F^*(s)U^*(s). \]

Since the transforms of the discrete input and output are trivial

\[ U^*(s) = \int_{0}^{\infty} u(t)m(t)e^{-st}dt = \sum_{k=0}^{\infty} u(kT_s)e^{-skT_s} \]

\[ Y^*(s) = \int_{0}^{\infty} y(t)m(t)e^{-st}dt = \sum_{k=0}^{\infty} y(kT_s)e^{-skT_s} \]

the only term of real interest is \( F^* \) for which

\[ F^*(s) = \int_{0}^{\infty} f(t)m(t)e^{-st}dt = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} F(s + \frac{2\pi}{T_s}k) \] (4.6)

because the product in the time domain can be expressed as a convolution in the frequency domain

\[ F^*(s) = \frac{1}{i2\pi} \int_{\gamma-i\infty}^{\gamma+i\infty} F(v)M(s - v)dv = \frac{1}{i2\pi} \int_{\gamma-i\infty}^{\gamma+i\infty} F(v) \frac{1}{1 - e^{-(s-v)T_s}} dv. \]

Placing the path of integration between the poles of \( F \) and \( M \) and completing it by a large semi-circle enclosing the poles of \( M \), residue calculus can be used under some mild conditions. The poles of \( M \) will be located at the zeros of \( e^{T_s(s-v)} = 1 \) which are

\[ v = s + i\frac{2\pi}{T_s}k, \; k = \ldots, -1, 0, 1, \ldots. \]
4.4 Truncating the Pulse Transfer Function

The residues can be proved to be

\[ \frac{1}{T_s} F(s + \frac{2\pi}{T_s}). \]

Since \( F(s) = G_c(s)H(s) \) and

\[ H(s + i\frac{2\pi}{T_s}k) = \frac{1 - e^{-(s+i\frac{2\pi}{T_s}k)}}{s + i\frac{2\pi}{T_s}k} \]
\[ = \frac{1 - e^{-sT_s e^{i2\pi k}}}{s + i\frac{2\pi}{T_s}k} \]
\[ = \frac{1 - e^{-sT_s}}{s + i\frac{2\pi}{T_s}k}. \]

Hence the expression

\[ G_d(e^{i\omega T_s}, \theta) = \left(1 - \frac{e^{-i\omega T_s}}{T_s}\right) \sum_{k=-\infty}^{\infty} G_c(i\omega + i\frac{2\pi}{T_s}k, \theta) \]

follows from (4.6). For a more detailed but similar discussion we refer the reader to Theorems 4.1 and 4.2 in the book by Åström and Wittenmark (1984).

4.4 Truncating the Pulse Transfer Function

A drawback connected with using the formula in (4.3) for estimation is of course the infinite sum. Good approximations can however be achieved with a limited number of terms, when the continuous-time system is strictly proper. Then, for each individual frequency \( \omega \), the contribution from higher order terms of the sum will approach zero as \( N_f \to \infty \).

In Figure 4.2 the second-order continuous-time OE-model

\[ G_c(s) = \frac{b_0s + b_1}{s^2 + a_1s + a_2} \]  

(4.7)

with true parameters \( a_1 = 2, a_2 = 3, b_0 = 1 \) and \( b_1 = 0.5 \) has been estimated using the method described in (4.2). The duration of the data set was \( T = 1000 \) s, the sampling time was \( T_s = 1 \) s and a random excitation signal was used. During estimation, the discrete-time counterpart \( G_d(e^{i\omega T_s}) \) to the continuous-time frequency response function \( G_c(i\omega) \) has been approximated as follows

\[ G_d(e^{i\omega T_s}, \theta) = \left(1 - \frac{e^{-i\omega T_s}}{T_s}\right) \sum_{k=-N_f}^{N_f} G_c(i\omega + i\frac{2\pi}{T_s}k, \theta) \]

(4.8)

In Figure 4.2, we have used \( N_f = 0 \) in (4.8), which means that we have assumed that the discrete- and continuous-time frequency response functions are almost equal. The figure illustrates the frequency-domain bias which could occur if the higher order terms in (4.8) are not taken into account. In Table 4.1 the parameter values for the same estimate as in Figure 4.2 are illustrated.
Figure 4.2: Bode diagram comparing a noise free estimate (dashed) of the true system \( G(s) = \frac{s^{0.5}}{s^2 + 2s + 3} \) (solid) using the method in (4.2) when \( N_f = 0 \) in (4.8). The figure illustrates the frequency-domain bias which could occur if the higher order terms in (4.8) are not taken into account.

4.4.1 Numerical Illustration

In Figure 4.3 the parameters of the model in (4.7) have been estimated using \( N_f = 0, \ldots, 10 \) in the expression for the frequency response of the pulse transfer function in (4.8). As can be seen, the parameter bias decreases as \( N_f \) increases. This is also illustrated numerically in Table 4.2. In Table 4.3 the effect of both the number \( N_f \) in (4.8) and the sampling times \( T_s \) are shown. As can be seen the bias also decrease with the sampling rate.

It should be noticed that the system in (4.7) is among the most difficult cases when using (4.8). Quite many high order terms may be needed at high frequencies since the gain there,

\[
|G_c(i\omega)| \approx \frac{b_0}{|\omega|},
\]

is decreasing slowly with increasing \( \omega \).
4.4 Truncating the Pulse Transfer Function

Table 4.1: Identified parameters for the transfer function $G_c(s) = \frac{s^{0.5}}{s^2+2s+3}$. Here $N_f = 0$ in (4.8) which means that one assumes that $G_d(e^{i\omega T_s}) = G_c(i\omega) \frac{H(i\omega)}{T_s}$. The sampling interval is $T_s = 1s$ and the parameters are biased.

<table>
<thead>
<tr>
<th>Method</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_0$</th>
<th>$b_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True System</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Estimated ($N_f = 0$)</td>
<td>1.3605</td>
<td>4.7870</td>
<td>0.5266</td>
<td>1.4300</td>
</tr>
</tbody>
</table>

Table 4.2: Identified parameters for the system $G_c(s) = \frac{s^{0.5}}{s^2+2s+3}$ versus the number $N_f$ of higher order terms in (4.8), $T_s = 1s$. The figure illustrates how the bias decreases as $N_f$ increases.

<table>
<thead>
<tr>
<th>$N_f$</th>
<th>$a_1 = 2$</th>
<th>$a_2 = 3$</th>
<th>$b_0 = 1$</th>
<th>$b_1 = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.3605</td>
<td>4.7870</td>
<td>0.5266</td>
<td>1.4300</td>
</tr>
<tr>
<td>1</td>
<td>1.7908</td>
<td>3.1375</td>
<td>0.9334</td>
<td>0.6408</td>
</tr>
<tr>
<td>2</td>
<td>1.9061</td>
<td>3.0930</td>
<td>0.9740</td>
<td>0.5906</td>
</tr>
<tr>
<td>3</td>
<td>1.9673</td>
<td>3.0706</td>
<td>1.0006</td>
<td>0.5640</td>
</tr>
<tr>
<td>4</td>
<td>1.9844</td>
<td>3.0028</td>
<td>1.0075</td>
<td>0.5333</td>
</tr>
<tr>
<td>5</td>
<td>1.9838</td>
<td>3.0063</td>
<td>1.0016</td>
<td>0.5289</td>
</tr>
<tr>
<td>6</td>
<td>1.9932</td>
<td>3.0207</td>
<td>1.0065</td>
<td>0.5280</td>
</tr>
<tr>
<td>7</td>
<td>1.9941</td>
<td>3.0324</td>
<td>1.0048</td>
<td>0.5307</td>
</tr>
<tr>
<td>8</td>
<td>1.9985</td>
<td>3.0263</td>
<td>1.0062</td>
<td>0.5252</td>
</tr>
<tr>
<td>9</td>
<td>2.0054</td>
<td>3.0263</td>
<td>1.0061</td>
<td>0.5259</td>
</tr>
<tr>
<td>10</td>
<td>2.0080</td>
<td>3.0189</td>
<td>1.0096</td>
<td>0.5161</td>
</tr>
</tbody>
</table>
Table 4.3: Identified parameters for the system $G_c(s) = \frac{s + 0.5}{s^2 + 2s + 3}$ versus the number $N_f$ of higher order terms in (4.8) and the sampling time $T_s$. The table illustrates how the bias decreases as $N_f$ increases and $T_s$ decreases.

<table>
<thead>
<tr>
<th>$N_f$</th>
<th>$T_s = 0.02$</th>
<th>$T_s = 0.1$</th>
<th>$T_s = 0.5$</th>
<th>$T_s = 1$</th>
<th>$T_s = 1.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0047</td>
<td>1.0003</td>
<td>1.0093</td>
<td>0.9263</td>
<td>0.3645</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4957</td>
<td>0.5057</td>
<td>0.5416</td>
<td>0.6516</td>
<td>0.5723</td>
</tr>
<tr>
<td>2</td>
<td>2.0288</td>
<td>2.0019</td>
<td>1.9882</td>
<td>1.7926</td>
<td>1.0915</td>
</tr>
<tr>
<td>3</td>
<td>2.9968</td>
<td>3.0109</td>
<td>3.0391</td>
<td>3.1560</td>
<td>2.7964</td>
</tr>
<tr>
<td>2</td>
<td>1.0047</td>
<td>1.0000</td>
<td>1.0072</td>
<td>0.9791</td>
<td>0.5087</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4957</td>
<td>0.5050</td>
<td>0.5245</td>
<td>0.5822</td>
<td>0.5119</td>
</tr>
<tr>
<td>2</td>
<td>2.0288</td>
<td>2.0018</td>
<td>1.9962</td>
<td>1.9114</td>
<td>1.3011</td>
</tr>
<tr>
<td>3</td>
<td>2.9968</td>
<td>3.0103</td>
<td>3.0215</td>
<td>3.0696</td>
<td>2.6539</td>
</tr>
<tr>
<td>5</td>
<td>1.0047</td>
<td>0.9997</td>
<td>1.0038</td>
<td>1.0036</td>
<td>0.6869</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4957</td>
<td>0.5045</td>
<td>0.5109</td>
<td>0.5341</td>
<td>0.4856</td>
</tr>
<tr>
<td>2</td>
<td>2.0288</td>
<td>2.0017</td>
<td>1.9998</td>
<td>1.9831</td>
<td>1.5824</td>
</tr>
<tr>
<td>3</td>
<td>2.9966</td>
<td>3.0098</td>
<td>3.0087</td>
<td>3.0260</td>
<td>2.6876</td>
</tr>
<tr>
<td>10</td>
<td>1.0047</td>
<td>0.9995</td>
<td>1.0021</td>
<td>1.0049</td>
<td>0.8129</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4957</td>
<td>0.5042</td>
<td>0.5056</td>
<td>0.5170</td>
<td>0.4878</td>
</tr>
<tr>
<td>2</td>
<td>2.0288</td>
<td>2.0016</td>
<td>2.0004</td>
<td>1.9969</td>
<td>1.7655</td>
</tr>
<tr>
<td>3</td>
<td>2.9967</td>
<td>3.0096</td>
<td>3.0042</td>
<td>3.0127</td>
<td>2.7935</td>
</tr>
</tbody>
</table>
4.5 Approximating the Continuous-Time System

In Section 4.4 the mode of approximation was the systematic truncation of the sum in (4.3). Another way to alter this relationship in a controlled manner would be to slightly change the properties of the continuous-time system $G_c$. Assume that $G_c$ has the following transfer function

$$G_c(s) = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_m}{s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_n}. \quad (4.10)$$

Then, for $\omega$ above the bandwidth of the system, we have

$$G_c(i\omega) \approx \frac{b_0}{(i\omega)^\ell}. \quad (4.11)$$

where $\ell = n - m$ is the so called pole excess or relative degree of the system. This means that, at high frequencies or over small time intervals the system approximately behaves as a set of $\ell$ integrators in series with a gain $b_0$. 

Figure 4.3: Identified parameters for the system $G_c(s) = \frac{s+0.5}{s^2+2s+3}$ versus the number $N_f$ of higher order terms in (4.8), $T_s = 1s$. The figure illustrates how the bias decreases as $N_f$ increases.
The high frequency part of the transfer function is usually called the roll-off and the relative degree basically tells us that the output \( y(t) \) of the system at high sampling rates almost behaves as if it would be \( \ell \) times continuously differentiable.

### 4.5.1 Sampling Zeros and Euler-Frobenius Polynomials

Assume that the system in (4.10) can be represented as

\[
G_c(s) = K \frac{(s - z_1)(s - z_2) \ldots (s - z_m)}{(s - p_1)(s - p_2) \ldots (s - p_n)}
\]

where \( K \) is the static gain, \( \{p_k\}_{k=1}^n \) are the continuous-time poles and \( \{z_k\}_{k=1}^m \) are the continuous-time zeros. Then, we know that the relative degree of the discrete-time pulse transfer function will always be 1 and the continuous-time poles will move to the discrete-time poles \( \{e^{p_k T_s}\}_{k=1}^n \) and the continuous-time zeros will move to the discrete-time zeros \( \{e^{z_k T_s}\}_{k=1}^m \).

The remaining \( \ell - 1 = n - m - 1 \) zeros, which are called sampling zeros will approach the roots of the well known Euler-Frobenius polynomials \( \Pi_\ell(z) \) (Åström et al., 1984; Weller et al., 2001). In fact, if the system is composed of only \( \ell \) integrators in series and a static gain \( b_0 \) such that

\[
G_c(s) = \frac{b_0}{s^\ell}
\]

then, the corresponding pulse transfer function will be

\[
G_d(z) = b_0 \frac{\Pi_\ell(z)}{\ell! \left( \frac{1}{T_s} \right)^\ell}.
\]

The Euler-Frobenius polynomials and their properties have a long pedigree dating back to the days of Euler in the 18th century, containing celebrated names such as Sobolev (1977) and Fröbenius (1910). In fact Fröbenius (1910) showed the interesting recursive relationship for these polynomials

\[
\Pi_{\ell+1} = (1 + \ell z) \Pi_\ell(z) + z(1 - z) \frac{d}{dz} \Pi_\ell(z) \quad \Pi_0(z) = 1.
\]

For practical purposes, the expansion

\[
\Pi_\ell(z) = b_1^\ell z^{\ell-1} + b_2^\ell z^{\ell-2} + \ldots + b_\ell^\ell, \quad \ell \geq 1
\]

where

\[
b_k^\ell = \sum_{l=1}^{k} (-1)^{k-l} l^\ell \binom{\ell + 1}{k - l}, \quad k = 1, \ldots, \ell
\]
Approximating the Roll-Off Behavior

In the previous section we used the summation formula (4.8) in order to estimate parameters. In this section, we will try to find a good approximation of the sum which is easier to calculate. For instance, define

$$F_{dc}^{(\ell)}(i\omega) = \frac{\Pi_{\ell}(e^{i\omega T_s})}{\ell! \left(\frac{e^{i\omega T_s} - 1}{T_s}\right)^\ell} - \frac{1}{(i\omega)^\ell} \frac{H(i\omega)}{T_s}$$

(4.23)

where $\Pi_{\ell}(z)$ are the Euler-Frobenius polynomials described in 4.5.1 and

$$H(i\omega) = 1 - e^{-i\omega T_s}$$

(4.24)

is the continuous-time transfer function of a zero order hold circuit. Assume, as before that the system $G_c$, defined in (4.7) has relative degree $\ell = n - m$. Then, we know that the high frequency gain will be

$$b_0 = \lim_{\omega \to \infty} G_c(i\omega)(i\omega)^\ell.$$

(4.25)

In the theorem below we use (4.25) and (4.23) in order to derive a new approximation of the discrete-time pulse transfer function $G_d(e^{i\omega T_s}, \theta)$

**Theorem 4.1**

Assume that $\ell \geq 1$, then as $T_s \to 0$

$$G_d(e^{i\omega T_s}, \theta) \to G_c(i\omega, \theta) \frac{H(i\omega)}{T_s} + b_0 F_{dc}^{(\ell)}(i\omega)$$

for each $\omega$ where $F_{dc}^{(\ell)}(i\omega)$ is defined as in (4.23) and $b_0$ as in (4.25). Also, $H$ and $F_{dc}^{(\ell)}$ are independent of the parameters in $\theta$. 
Proof: First of all we have

\[ \frac{b_0}{G_c(i\omega + i\frac{2\pi}{T_s} k)} \rightarrow 1 \]

as \( T_s \rightarrow 0 \) for \( k \neq 0 \). The explanation is that only the high frequency part of the transfer function will be visible in this frequency range. This, in turn, means that the discrete-time pulse transfer function in (4.3) can be approximated with

\[ G_d(e^{i\omega T_s}, \theta) \rightarrow G_c(i\omega, \theta) H(i\omega) + b_0 \sum_{k \neq 0} \frac{1 - e^{-i\omega T_s}}{(i\omega + i\frac{2\pi}{T_s} k)^\ell + 1} \]

as \( T_s \rightarrow 0 \). Here \( H \) is defined as in (4.24). Then, from Lemma 3.2 in (Wahlberg, 1988)

\[ \frac{e^{i\omega T_s} \Pi \ell(e^{i\omega T_s})}{\ell!(e^{i\omega T_s/T_s-1})^{\ell+1}} = \sum_{k=-\infty}^{\infty} \frac{1}{(i\omega + i\frac{2\pi}{T_s} k)^{\ell+1}} \]

and by multiplying by \( (1 - e^{-i\omega T_s})/T_s \) and then subtracting the central term where \( k = 0 \) we get

\[ F_{dc}^{(\ell)}(i\omega) = \frac{\Pi \ell(e^{i\omega T_s})}{\ell!(e^{i\omega T_s/T_s-1})^{\ell}} - \frac{1}{(i\omega)^\ell} \frac{H(i\omega)}{T_s}. \]

This completes the proof.

The terms on the right-hand side of expression (4.23) allow for a few interesting reflections. Let

\[ F_{dc}^{(\ell)}(i\omega) = F_1(e^{i\omega T_s}) - F_2(i\omega). \]

Then, the first term

\[ F_1(e^{i\omega T_s}) = \frac{\Pi \ell(e^{i\omega T_s})}{\ell!(e^{i\omega T_s/T_s-1})^{\ell}}, \]

can be interpreted as the frequency response of a system made up of a chain of \( \ell \) integrators where the input is subject to zero-order hold and the input and output are sampled. For more details, the reader is referred to Lemma 1 in the seminal paper on sampling zeros by Åström et al. (1984). The second term

\[ F_2(i\omega) = \frac{H(i\omega)}{(i\omega)^\ell} \]

represents the continuous-time transfer function of a hold circuit followed by the same chain of integrators.
4.6.1 Numerical Illustration

In Figure 4.4 the parameters of the continuous-time system in (4.7)

\[ G(s) = \frac{b_1 s + b_2}{s^2 + a_1 s + a_2} \]  
(4.26)

where \( b_1 = 1, b_2 = 0.5, a_1 = 2, \) and \( a_2 = 3 \) have been estimated using the method

\[
\hat{\theta} = \arg\min_{b_0, \theta} \frac{1}{N_\omega} \sum_{k=1}^{N_\omega} \left| Y_d(e^{i\omega_k T_s}) - \left( G_c(i\omega_k, \theta) H(i\omega) + b_0 F_{dc}^{(l)}(i\omega_k) \right) U_d(e^{i\omega_k T_s}) \right|^2 .
\]
(4.27)

with different sampling intervals \( T_s \). The process has in all cases been observed during \( T = N_t T_s = 1000 \) s and the excitation signal is random and binary. Frequency domain data up to the Nyquist frequency have been used such that

\[
\omega_k = \frac{2\pi}{T_s} k, \quad k = 1 \ldots N_\omega
\]
(4.28)

where \( N_\omega = \frac{N_t - 1}{2} \). The estimated parameter values for the system are also found in Table 4.4. As can be seen from both Figure 4.4 and Table 4.4 the approximation will be good at moderate sampling rates.

### Table 4.4: Identified parameters for the system \( G_c(s) = \frac{s + 0.5}{s^2 + 2s + 3} \) versus \( T_s \). The method used is that of (4.27).

<table>
<thead>
<tr>
<th>( T_s )</th>
<th>( a_1 = 2 )</th>
<th>( a_2 = 3 )</th>
<th>( b_0 = 1 )</th>
<th>( b_1 = 0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.9996</td>
<td>3.0055</td>
<td>0.9992</td>
<td>0.5020</td>
</tr>
<tr>
<td>0.1</td>
<td>1.9966</td>
<td>3.0075</td>
<td>0.9980</td>
<td>0.5023</td>
</tr>
<tr>
<td>0.2</td>
<td>1.9895</td>
<td>3.0116</td>
<td>0.9953</td>
<td>0.5045</td>
</tr>
<tr>
<td>0.3</td>
<td>1.9768</td>
<td>3.0161</td>
<td>0.9894</td>
<td>0.5068</td>
</tr>
<tr>
<td>0.4</td>
<td>1.9609</td>
<td>3.0309</td>
<td>0.9802</td>
<td>0.5117</td>
</tr>
<tr>
<td>0.5</td>
<td>1.9272</td>
<td>3.0418</td>
<td>0.9645</td>
<td>0.5192</td>
</tr>
<tr>
<td>0.6</td>
<td>1.8862</td>
<td>3.0356</td>
<td>0.9421</td>
<td>0.5225</td>
</tr>
<tr>
<td>0.7</td>
<td>1.8245</td>
<td>3.0183</td>
<td>0.9118</td>
<td>0.5234</td>
</tr>
<tr>
<td>0.8</td>
<td>1.7522</td>
<td>3.0135</td>
<td>0.8723</td>
<td>0.5331</td>
</tr>
<tr>
<td>0.9</td>
<td>1.6652</td>
<td>2.9433</td>
<td>0.8194</td>
<td>0.5225</td>
</tr>
</tbody>
</table>

4.7 Estimating the Continuous-Time Fourier Transform

Assume that there is no output noise in the continuous-time model in Figure 4.1. The frequency domain relationship between the sampled input and output will then be

\[
Y_d(e^{i\omega T_s}) = G_d(e^{i\omega T_s}, \theta) U_d(e^{i\omega T_s}) + \text{transients}
\]
where $G_d$ is defined by (4.3) if the input is assumed to be piecewise constant. In continuous-time, the corresponding relationship would be characterized by

$$Y_c(i\omega) = G_c(i\omega, \theta)U_c(i\omega) = G_c(i\omega, \theta)H(i\omega)U_d(e^{i\omega T_s}),$$

since the connection between the continuous- and discrete-time Fourier transforms of the input is

$$U_c(i\omega) = H(i\omega)U_d(e^{i\omega T_s})$$

due to the hold circuit $H$ in (4.24). This way of thinking will open for the estimation of the continuous-time Fourier transform of the output of the systems. Assume that we know the exact parameter values $\theta_0$. Then, we could compute the exact continuous-time Fourier transform of the output as

$$Y_c(i\omega) = \frac{G_c(i\omega, \theta_0)H(i\omega)}{G_d(e^{i\omega T_s}, \theta_0)}Y_d(e^{i\omega T_s}).$$

However, since we wish to identify the parameters, the knowledge of $\theta_0$ is quite unrealistic. Therefore, assume as before, that the system in (4.10) is strictly proper, stable and of
relative degree $\ell = n - m$. Further assume that the sampling time $T_s$ is such that the rate of sampling is above the system bandwidth. Define

$$F_{c,\ell+1,T_s}(i\omega) = T_s \left( \frac{e^{i\omega T_s} - 1}{i\omega T_s} \right)^{\ell+1} \Pi_{\ell}(e^{i\omega T_s})$$

(4.30)

where $\Pi_{\ell}(z)$ are the Euler-Frobenius polynomials in Section 4.5.1.

\[\begin{array}{c}
\text{Figure 4.5: Comparison of } G_c(i\omega)H(i\omega)/G_d(e^{i\omega T_s}) \text{ (solid) in (4.29) and } F_{c,\ell+1,T_s} \text{ (dotted) in (4.30) with } \ell = 2 \text{ for the continuous-time system } G_c(s) = \frac{1}{s^3 + 2s^2 + 3s + 4}. \\
\text{The sampling rate is } T_s = 0.5. \text{ For moderately high frequencies, the difference is quite small.}
\end{array}\]

In Figure 4.5 we can see that the frequency domain difference between

$$\frac{G_c(i\omega)H(i\omega)}{G_d(e^{i\omega T_s})}$$

(4.31)

in (4.29) and $F_{c,\ell+1,T_s}$ (4.30) for the transfer function

$$G(s) = \frac{1}{s^3 + 2s^2 + 3s + 4}$$

(4.32)
is quite small. The generality of this observation is also verified by the following theoretical result.

**Theorem 4.2**

Assume $F_{\ell+1}^{c,T_s}$ is defined as in (4.30) with $\ell \geq 1$. Then, for each $\omega$

$$\frac{G_c(i\omega, \theta_0) H(i\omega)}{G_d(e^{i\omega T_s}, \theta_0)} \rightarrow F_{\ell+1}^{c,T_s}(i\omega)$$

with the rate $O(T_s^{\ell+1})$ as $T_s \rightarrow 0$.

**Proof:** First of all we have

$$G_c\left(i\omega + \frac{i2\pi k}{T_s}\right) \rightarrow \frac{1}{(i\omega + \frac{i2\pi k}{T_s})^\ell}$$

as $T_s \rightarrow 0$ if $k \neq 0$ and $b_0$ is defined as in (4.25). This has the consequence that

$$\frac{G_c(i\omega) H(i\omega)}{G_d(e^{i\omega T_s})} \rightarrow \frac{G_c(i\omega)}{i\omega} + \sum_{k \neq 0} \frac{b_0}{(i\omega + \frac{i2\pi k}{T_s})^{\ell+1}}$$

as $T_s \rightarrow 0$ if we insert (4.33) in (4.3). From Lemma 3.2 in (Wahlberg, 1988) it is also clear that

$$F_c^\ell(i\omega) = \frac{1}{(i\omega)^{\ell+1}} + \sum_{k \neq 0} \frac{b_0}{(i\omega + \frac{i2\pi k}{T_s})^{\ell+1}}.$$

By putting the two previous expressions on a common denominator, we get the following relation

$$\frac{G_c(i\omega) H(i\omega)}{G_d(e^{i\omega T_s})} - F_{\ell+1}^{c,T_s}(i\omega) \rightarrow F_{\ell+1}^{c,T_s}(i\omega) R(i\omega) S(i\omega)$$

where

$$R(i\omega) = \frac{1 - G_c(i\omega) \frac{(i\omega)^\ell}{b_0}}{G_c(i\omega) + \sum_{k \neq 0} \frac{b_0}{(i\omega + \frac{i2\pi k}{T_s})^{\ell+1}}}$$

and

$$S(i\omega) = \sum_{k \neq 0} \frac{b_0}{(i\omega + \frac{i2\pi k}{T_s})^{\ell+1}}.$$

Since $F$ and $R$ are bounded in $\omega$ and the terms of $S$ are bounded as

$$\left|\frac{1}{(i\omega + \frac{i2\pi k}{T_s})^{\ell+1}}\right| \leq \frac{T_s^{\ell+1}}{(i\omega T_s + i2\pi k)^{\ell+1}} \leq C(\omega) \left(\frac{T_s}{k}\right)^{\ell+1}$$
if $k \neq 0$, the result
\[
\left| \frac{G_c(i\omega)H(i\omega)}{G_d(e^{i\omega T_s})} - F_{\ell+1,T_s}^c(i\omega) \right| \leq \left| F_{\ell+1,T_s}^c(i\omega) \right| |R(i\omega)||S(i\omega)| \leq C(\omega)T_s^{\ell+1}
\]
follows.

This result opens for the estimation of the continuous-time Fourier transform of the output as
\[
\hat{Y}_c(i\omega) = F_{\ell+1,T_s}^c(i\omega)Y_d(e^{i\omega T_s})
\]
which can be interpreted as assuming that $y(t)$ behaves as an $\ell$ times continuously differentiable function between the sampling instants. The parameters can then be estimated from $\hat{Y}_c$ by the continuous-time method
\[
\hat{\theta} = \arg \min_{\theta} \sum_{k=1}^{N_\omega} \left| \hat{Y}_c(i\omega_k) - G_c(i\omega_k, \theta)U_c(i\omega_k) \right|^2.
\]

This two-stage process of first estimating the continuous-time spectrum and then estimating the system parameters, has the advantage of not directly involving the exact discrete-time frequency response $G_d(e^{i\omega T_s})$.

### 4.7.1 Numerical Illustration

In Figure 4.6 the parameters of the continuous-time system in (4.7)
\[
G(s) = \frac{b_1 s + b_2}{s^2 + a_1 s + a_2}
\]
where $b_1 = 1$, $b_2 = 0.5$, $a_1 = 2$ and $a_2 = 3$ have been estimated using the method described in (4.34) and (4.35) using different sampling intervals $T_s$. The process has always been observed during $T = N_t T_s = 1000s$ and the excitation signal is random and binary. Frequency domain data up to the Nyquist frequency have been used, and therefore we have
\[
\omega_k = \frac{2\pi}{T_s} k, \quad k = 1 \ldots N_\omega
\]
where $N_\omega = \frac{N_t - 1}{2}$. The estimated parameter values are also found in Table 4.5. As can be seen from both Figure 4.4 and Table 4.4 the approximation will be quite good at moderate sampling rates.

### 4.8 Comparisons

In Sections 4.4, 4.6 and 4.7 new, approximation based, methods for identification of continuous-time input-output models have been introduced. In this section we will illustrate how the different approaches perform for some systems with different sampling
Figure 4.6: Identified parameters for the system $G_c(s) = \frac{s + 0.5}{s^2 + 2s + 3}$ versus $T_s$. The method used is that of (4.35).

intervals. In all cases we will simulate the models

\[ G_1(s) = \frac{s + 0.5}{s^2 + 2s + 3} \]
\[ G_2(s) = \frac{1}{s^3 + 2s^2 + 3s + 4} \]
\[ G_3(s) = \frac{4}{s^3 + 4s^2 + 4s + 4} \],

with a piecewise constant input where the amplitudes of the constant segments have a Gaussian distribution. No noise was added to output in the simulations. The results of this comparison is provided in Table 4.6, Table 4.7 and Table 4.8.
Table 4.5: Identified parameters for the system $G_c(s) = \frac{s+0.5}{s^2+2.5s+3}$ versus $T_s$. The method used is that of (4.35).

<table>
<thead>
<tr>
<th>$T_s$</th>
<th>$a_1 = 2$</th>
<th>$a_2 = 3$</th>
<th>$b_0 = 1$</th>
<th>$b_1 = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.9910</td>
<td>3.0021</td>
<td>0.9949</td>
<td>0.5033</td>
</tr>
<tr>
<td>0.2</td>
<td>1.9684</td>
<td>3.0170</td>
<td>0.9800</td>
<td>0.5152</td>
</tr>
<tr>
<td>0.3</td>
<td>1.9270</td>
<td>3.0284</td>
<td>0.9539</td>
<td>0.5274</td>
</tr>
<tr>
<td>0.4</td>
<td>1.8824</td>
<td>3.0542</td>
<td>0.9202</td>
<td>0.5531</td>
</tr>
<tr>
<td>0.5</td>
<td>1.8180</td>
<td>3.0436</td>
<td>0.8767</td>
<td>0.5579</td>
</tr>
<tr>
<td>0.6</td>
<td>1.7498</td>
<td>3.0501</td>
<td>0.8232</td>
<td>0.5737</td>
</tr>
<tr>
<td>0.7</td>
<td>1.6830</td>
<td>3.0506</td>
<td>0.7657</td>
<td>0.5904</td>
</tr>
<tr>
<td>0.8</td>
<td>1.6036</td>
<td>3.0727</td>
<td>0.6994</td>
<td>0.6126</td>
</tr>
<tr>
<td>0.9</td>
<td>1.5101</td>
<td>3.0596</td>
<td>0.6272</td>
<td>0.6164</td>
</tr>
<tr>
<td>1.0</td>
<td>1.4255</td>
<td>2.9974</td>
<td>0.5526</td>
<td>0.6040</td>
</tr>
</tbody>
</table>

Approach 1

The first, and simplest, approach to is to push one’s luck and assume that the data is sampled so fast that it can be considered band-limited. Thereby one would use the method

$$\hat{\theta} = \arg \min_\theta \sum_{k=1}^{N_\omega} \left| Y_d(e^{i\omega_k T_s}) - G_c(i\omega_k, \theta)U_d(e^{i\omega_k T_s}) \right|^2.$$  (4.42)

If the true system is of low pass character, this assumption may be more plausible for the output than for the input. We will label this Approach 1.

Approach 2

Another reasonable approximation which was presented earlier, is to use just the central term (i.e. $N_f = 0$ in (4.8)) in the approximation of the discrete-time frequency response. This would then mean that we can apply the method

$$\hat{\theta} = \arg \min_\theta \sum_{k=1}^{N_\omega} \left| Y_d(e^{i\omega_k T_s}) - G_c(i\omega_k, \theta) \frac{H(i\omega_k)}{T_s} U_d(e^{i\omega_k T_s}) \right|^2.$$  (4.43)

Compared to Approach 1, we have a continuous-time method where the piecewise constant input is correctly translated to continuous time and a band limited assumption on the output is used. This is in line with the assumption that the system is low pass in relation to the sampling interval. We label this Approach 2.

Approach 3

An obvious variant of the above approach is to involve more terms in (4.8). We call this Approach 3. Clearly, as $N_f \to \infty$ we approach the “correct” method.

$$\hat{\theta} = \arg \min_\theta \sum_{k=1}^{N_\omega} \left| Y_d(e^{i\omega_k T_s}) - G_d(i\omega_k, \theta)U_d(e^{i\omega_k T_s}) \right|^2.$$  (4.44)
where

\[ G_d(e^{i\omega T_s}, \theta) = \left( \frac{1 - e^{-i\omega T_s}}{T_s} \right) \sum_{k=-N_f}^{N_f} \frac{G_c(i\omega + i\frac{2\pi k}{T_s}, \theta)}{i\omega + i\frac{2\pi k}{T_s}}. \]  

(4.45)

**Approach 4**

In (4.34), a way to estimate the continuous-time Fourier transform of the output \( Y_c(i\omega) \) using the pre-filter in (4.30) was devised.

\[ \hat{Y}_c(i\omega) = F_{\ell+1, T_s}(i\omega) Y_d(e^{i\omega T_s}) \]  

(4.46)

After the transform was found, the continuous-time parameters could be acquired using the relationship in (4.35),

\[ \hat{\theta} = \arg \min_{\theta} \sum_{k=1}^{N_\omega} \left| \hat{Y}_c(i\omega_k) - G_c(i\omega_k, \theta) U_c(i\omega_k) \right|^2 \]  

(4.47)

where

\[ U_c(i\omega_k) = H(i\omega_k) U_d(e^{i\omega_k T_s}) \]  

(4.48)

We call this **Approach 4**.

**Approach 5**

Finally, the method in (4.27) from Section 4.6

\[ \hat{\theta} = \arg \min_{b_0, \theta} \frac{1}{N_\omega} \sum_{k=1}^{N_\omega} \left| Y_d(e^{i\omega_k T_s}) - \left( G_c(i\omega_k, \theta) \frac{H(i\omega)}{T_s} + b_0 F_{dc}^{(\ell)}(i\omega_k) \right) U_d(e^{i\omega_k T_s}) \right|^2 \]  

where the higher order terms in (4.3) \((k \neq 0)\) are approximated by \(\ell\) order integrators

\[ F_{dc}^{(\ell)}(i\omega) = \frac{\Pi_{\ell}(e^{i\omega T_s})}{\ell! \left( e^{i\omega T_s} - 1 \right)^{\ell}} - \frac{1}{(i\omega)^\ell} \frac{H(i\omega)}{T_s} \]  

(4.49)

we call **Approach 5**.

**Final Remark**

It should be noted that in practice it may be essential to limit the fit of all the estimation methods to frequencies that do not extend all the way to the Nyquist frequency, since the observations may be less reliable at higher frequencies. Another reason is that \(F_{\ell+1, T_s}^{C}\) in (4.30) will tend to infinity at the Nyquist frequency for \(\ell\) being even (the sampled multi-integrator will then have a zero at the Nyquist frequency).
4.9 Summary

In this chapter, different approaches to direct frequency domain estimation of continuous-time transfer functions based on sampled data have been presented. If the input intersample behavior is known this can be done without approximation. In particular for piecewise constant excitations, there are well known, but somewhat complicated formulas for this. We have investigated various frequency domain approximation of the exact transformation that are simpler to use and give good approximations, at least for sufficiently fast sampling rates. Essentially, these approximations are based on replacing the true parameter dependent system with a number of integrators that equal the pole excess of the system. This line of thought is continued in the following chapter.

Table 4.6: Results for the system \( \frac{s + 0.5}{s^2 + 2s + 3} \). This system has a pole excess of 1 and a bandwidth of 8.60 rad/s.

<table>
<thead>
<tr>
<th></th>
<th>( T_s = 0.02 )</th>
<th>( T_s = 0.1 )</th>
<th>( T_s = 0.5 )</th>
<th>( T_s = 1 )</th>
<th>( T_s = 1.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appr 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.9753</td>
<td>0.8794</td>
<td>0.2614</td>
<td>-0.5413</td>
<td>-0.3764</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5137</td>
<td>0.5529</td>
<td>1.4552</td>
<td>1.4623</td>
<td>0.5553</td>
</tr>
<tr>
<td>2</td>
<td>1.9443</td>
<td>1.7432</td>
<td>1.3928</td>
<td>1.5599</td>
<td>1.0350</td>
</tr>
<tr>
<td>3</td>
<td>2.9972</td>
<td>2.9670</td>
<td>4.8357</td>
<td>5.6280</td>
<td>3.1000</td>
</tr>
<tr>
<td>Appr 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.0002</td>
<td>1.0020</td>
<td>0.9879</td>
<td>0.5292</td>
<td>-0.1094</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5002</td>
<td>0.5046</td>
<td>0.6266</td>
<td>1.4389</td>
<td>1.0510</td>
</tr>
<tr>
<td>2</td>
<td>2.0006</td>
<td>2.0003</td>
<td>1.9033</td>
<td>1.3682</td>
<td>0.7592</td>
</tr>
<tr>
<td>3</td>
<td>3.0005</td>
<td>3.0046</td>
<td>3.1556</td>
<td>4.8057</td>
<td>4.2858</td>
</tr>
<tr>
<td>Appr 3 (( N_f = 10 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.0007</td>
<td>1.0029</td>
<td>1.0088</td>
<td>1.0070</td>
<td>0.8288</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4998</td>
<td>0.5058</td>
<td>0.5112</td>
<td>0.5180</td>
<td>0.4944</td>
</tr>
<tr>
<td>2</td>
<td>2.0018</td>
<td>2.0084</td>
<td>2.0133</td>
<td>2.0073</td>
<td>1.7852</td>
</tr>
<tr>
<td>3</td>
<td>2.9988</td>
<td>3.0085</td>
<td>3.0154</td>
<td>3.0123</td>
<td>2.8269</td>
</tr>
<tr>
<td>Appr 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.9999</td>
<td>0.9949</td>
<td>0.8791</td>
<td>0.5541</td>
<td>0.1864</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5000</td>
<td>0.5013</td>
<td>0.5272</td>
<td>0.6046</td>
<td>0.5351</td>
</tr>
<tr>
<td>2</td>
<td>2.0003</td>
<td>1.9929</td>
<td>1.8353</td>
<td>1.4233</td>
<td>1.0898</td>
</tr>
<tr>
<td>3</td>
<td>3.0000</td>
<td>2.9986</td>
<td>2.9659</td>
<td>3.0044</td>
<td>2.8335</td>
</tr>
<tr>
<td>Appr 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.0047</td>
<td>0.9994</td>
<td>0.9866</td>
<td>0.8223</td>
<td>0.4488</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4956</td>
<td>0.5040</td>
<td>0.5020</td>
<td>0.5244</td>
<td>0.4341</td>
</tr>
<tr>
<td>2</td>
<td>2.0288</td>
<td>2.0015</td>
<td>1.9760</td>
<td>1.6736</td>
<td>1.1949</td>
</tr>
<tr>
<td>3</td>
<td>2.9966</td>
<td>3.0093</td>
<td>3.0037</td>
<td>2.9537</td>
<td>2.3787</td>
</tr>
</tbody>
</table>
Table 4.7: Results for the system \( \frac{1}{s^3 + 2s^2 + 3s + 4} \). This system has a pole excess of 3 and a bandwidth of 2.1 rad/s.

<table>
<thead>
<tr>
<th></th>
<th>( T_s = 0.02 )</th>
<th>( T_s = 0.1 )</th>
<th>( T_s = 0.5 )</th>
<th>( T_s = 1 )</th>
<th>( T_s = 1.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appr 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.9805</td>
<td>0.9096</td>
<td>0.6692</td>
<td>0.4873</td>
<td>0.3584</td>
</tr>
<tr>
<td>2</td>
<td>1.9550</td>
<td>1.7887</td>
<td>1.2413</td>
<td>0.8993</td>
<td>0.7554</td>
</tr>
<tr>
<td>3</td>
<td>2.9805</td>
<td>2.9084</td>
<td>2.6481</td>
<td>2.4271</td>
<td>2.2400</td>
</tr>
<tr>
<td>4</td>
<td>3.8937</td>
<td>3.5078</td>
<td>2.2317</td>
<td>1.4376</td>
<td>1.0490</td>
</tr>
<tr>
<td><strong>Appr 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.0017</td>
<td>1.0000</td>
<td>0.9996</td>
<td>0.9857</td>
<td>0.9050</td>
</tr>
<tr>
<td>2</td>
<td>2.0031</td>
<td>2.0000</td>
<td>1.9993</td>
<td>1.9729</td>
<td>1.8921</td>
</tr>
<tr>
<td>3</td>
<td>3.0018</td>
<td>3.0000</td>
<td>2.9987</td>
<td>2.9900</td>
<td>2.9346</td>
</tr>
<tr>
<td>4</td>
<td>4.0061</td>
<td>4.0000</td>
<td>3.9987</td>
<td>3.9458</td>
<td>3.7830</td>
</tr>
<tr>
<td><strong>Appr 3 (( N_f = 2 ))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.9983</td>
<td>1.0039</td>
<td>1.0092</td>
<td>0.9957</td>
<td>0.9979</td>
</tr>
<tr>
<td>2</td>
<td>2.0001</td>
<td>2.0100</td>
<td>2.0161</td>
<td>1.9907</td>
<td>2.0124</td>
</tr>
<tr>
<td>3</td>
<td>3.0003</td>
<td>3.0032</td>
<td>3.0080</td>
<td>3.0033</td>
<td>3.0019</td>
</tr>
<tr>
<td><strong>Appr 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.0018</td>
<td>1.0000</td>
<td>1.0015</td>
<td>1.0268</td>
<td>0.9641</td>
</tr>
<tr>
<td>2</td>
<td>2.0032</td>
<td>1.9998</td>
<td>2.0002</td>
<td>1.9821</td>
<td>1.5172</td>
</tr>
<tr>
<td>3</td>
<td>3.0019</td>
<td>3.0000</td>
<td>3.0002</td>
<td>3.0044</td>
<td>2.8956</td>
</tr>
<tr>
<td>4</td>
<td>4.0063</td>
<td>3.9996</td>
<td>4.0013</td>
<td>3.9758</td>
<td>3.1243</td>
</tr>
<tr>
<td><strong>Appr 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.0346</td>
<td>0.9993</td>
<td>1.0028</td>
<td>0.9949</td>
<td>0.9900</td>
</tr>
<tr>
<td>2</td>
<td>2.1001</td>
<td>2.0012</td>
<td>2.0092</td>
<td>1.9930</td>
<td>2.0436</td>
</tr>
<tr>
<td>3</td>
<td>3.0554</td>
<td>2.9998</td>
<td>3.0040</td>
<td>2.9966</td>
<td>2.9994</td>
</tr>
<tr>
<td>4</td>
<td>4.1943</td>
<td>4.0038</td>
<td>4.0172</td>
<td>3.9863</td>
<td>4.0891</td>
</tr>
</tbody>
</table>

Table 4.8: Estimated values of \( a \) for the model \( \frac{1}{s^3 + a s^2 + a s + a} \) with true value \( a = 4 \). This system has a pole excess of 3 and a bandwidth of 0.75 rad/s.

<table>
<thead>
<tr>
<th></th>
<th>( T_s = 0.02 )</th>
<th>( T_s = 0.1 )</th>
<th>( T_s = 0.5 )</th>
<th>( T_s = 1 )</th>
<th>( T_s = 1.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appr 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4.0081</td>
<td>4.0388</td>
<td>4.2330</td>
<td>4.4837</td>
<td>4.7438</td>
</tr>
<tr>
<td>2</td>
<td>3.9988</td>
<td>3.9935</td>
<td>3.9988</td>
<td>3.9995</td>
<td>4.0020</td>
</tr>
<tr>
<td><strong>Appr 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4.0123</td>
<td>4.0041</td>
<td>4.0092</td>
<td>4.0017</td>
<td>4.0501</td>
</tr>
<tr>
<td>2</td>
<td>3.9987</td>
<td>3.9935</td>
<td>3.9988</td>
<td>3.9999</td>
<td>4.0038</td>
</tr>
<tr>
<td><strong>Appr 3 (( N_f = 5 ))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.9444</td>
<td>3.9855</td>
<td>3.9989</td>
<td>3.9951</td>
<td>3.9689</td>
</tr>
</tbody>
</table>
Identification of Input-Output Models
From Non-Uniformly Sampled Data

“It’s not so much that we’re afraid of change or so in love with the old ways, but it’s that place in between that we fear.... It’s like being between trapezes. It’s Linus when his blanket is in the dryer. There’s nothing to hold on to.”

Marilyn Ferguson (1938 - )

5.1 Introduction

In the beginning of the previous chapter a, more or less, explicit promise was made, that some of the methods that were presented, would open the way for frequency domain identification from non-uniformly sampled data. The method and line of thought from the previous chapter that seem to have the most potential in this matter, is the one which tries to estimate the continuous-time Fourier transform as a first step of the procedure. In (4.34) this was accomplished by multiplying the discrete-time Fourier transform of the output $y$ by a particular factor $F_{c,\ell+1,T_s}(i\omega)$ defined in (4.30). That function would then only depend on the sampling interval $T_s$ and the relative degree $\ell$ of the transfer function $G_c(s)$ to be identified. In order to summarize, one would say that the transform could be estimated as

$$Y_c(i\omega) = F_{c,\ell+1,T_s}(i\omega) Y_d(e^{i\omega T_s}).$$

where $Y_d(e^{i\omega T_s})$ is the discrete-time Fourier transform computed from equidistantly sampled data. Now, if we for a moment would assume that, $F_{c,\ell+1,T_s}(i\omega)$ is the continuous-
time Fourier transform of some function \( F_{\ell+1,T_s}^c(t) \) such that

\[
F_{\ell+1,T_s}^c(i\omega) = \int_{-\infty}^{\infty} F_{\ell+1,T_s}^c(t) e^{-i\omega t} dt \tag{5.2}
\]

then we would automatically get

\[
Y_c(i\omega) = \int_{-\infty}^{\infty} \hat{y}_c(t) e^{-i\omega t} dt \tag{5.3}
\]

where

\[
\hat{y}_c(t) = \sum_{k=0}^{N_\ell-1} y(kT_s) F_{\ell+1,T_s}^c(t - kT_s). \tag{5.4}
\]

The interpretation of this relationship is straightforward, and points towards interpolation with the kernel function \( F_{\ell+1,T_s}^c(t) \) in order to reconstruct the intersample behavior of the function \( y(t) \). In particular, interpolation in terms of polynomial splines (deBoor, 1978; Schoenberg, 1973). Elucidating the mechanisms behind these statements and their consequences for the frequency domain estimation of deterministic continuous-time systems from non-uniformly sampled data, will be the objective of the remaining part of this chapter.

### 5.2 Outline

The outline of the chapter is the following. First, in Section 5.3 there will be a brief introduction to polynomial interpolation. Mainly in order to prepare the reader for the basics of polynomial splines which are introduced in Section 5.4. The exposition found in these sections is mostly based on classical material, which can be found in basic textbooks on numerical analysis and on spline interpolation, i.g. Eldén and Wittmeyer-Koch (1996), deBoor (1978), Schoenberg (1973) and Schumaker (1981). For a more extensive empirical investigation on interpolation using splines, we refer to the paper by Rolain (Rolain et al., 1998). The discussion in Section 5.3 and Section 5.4 is quite general by nature and applies equally well to uniform and non-uniform sampling. However, the degree of generality found here will obscure some of the connections that can be made with the methods in Chapter 4, where data is presented uniformly. Therefore, in Section 5.5 we will lend ourselves to the domain of equidistant sampling where the linear system techniques found in the papers by Unser et al. (1991, 1993a,b) can be used to reveal this connection. In Section 5.6 the face of the function \( F_{\ell+1,T_s}^c(t) \) in (5.2) is unveiled and the special interpolation form in (5.4) will have its explanation. Finally, in Section 5.8 a frequency domain method based on splines will be tested numerically.
5.3 Polynomial Interpolation

Assume that function values \( y(t_k) = y_k \) are known at \((\ell + 1)\) different points \( \{t_k\}_{k=1}^{\ell+1} \) and we seek a function \( P \) such that

\[
P(t_k) = y_k, \quad k = 1, 2, \ldots, \ell + 1.
\]

That is, \( P \) should interpolate \( y \) at the points \( \{t_k\}_{k=1}^{\ell+1} \). The first question that comes to one's mind after this definition, is which class of functions one should use. An obvious choice would be to use polynomials of degree \( \ell \), maybe represented in Newton's form

\[
P_\ell(t) = c_0 + \sum_{k=1}^{\ell} c_k \prod_{l=1}^{k} (t - t_l)
\]

because of its simple structure and easy evaluation.

For instance, let \( \ell = 2 \). Then the polynomial \( P_2 \) will have the representation

\[
P_2(t) = c_0 + c_1(t - t_1) + c_2(t - t_1)(t - t_2).
\]

and the interpolation conditions in (5.5) together with the polynomial representation in (5.7) will give the following system of equations for the extraction of the coefficients \( c_0, c_1 \) and \( c_2 \),

\[
\begin{align*}
    c_0 &= y_1 \\
    c_0 + c_1(t_2 - t_1) &= y_2 \\
    c_0 + c_1(t_3 - t_1) + c_2(t_3 - t_1)(t_3 - t_2) &= y_3.
\end{align*}
\]

Solving the particular triangular set of equations in (5.8) will then yield the coefficients

\[
\begin{align*}
    c_0 &= y_1 \\
    c_1 &= \frac{y_2 - y_1}{t_2 - t_1} \\
    c_2 &= \frac{y_3 - y_1 - \frac{t_3 - t_1}{t_2 - t_1}(y_2 - y_1)}{(t_3 - t_1)(t_3 - t_2)}.
\end{align*}
\]

Generally, the system in (5.8) will have a triangular shape due to the interpolation conditions and the special structure of the polynomial in (5.6). This in turn will produce a recursive expression for an \( \ell \) order polynomial \( P_\ell \) that would interpolate at the points \( \{t_k\}_{k=1}^{\ell+1} \). This formula that can be stated as

\[
P_0(t) = c_0 \quad (5.12)
\]

\[
P_k(t) = P_{k-1}(t) + c_k \prod_{l=1}^{k} (t - t_l). \quad (5.13)
\]

where \( P_k \) is the polynomial of order \( k \) which interpolates at the points \( \{t_l\}_{l=1}^{k+1} \) and \( k \leq \ell \). This means that \( c_k \) will only depend on the values \( \{y(t_k)\}_{k=1}^{k+1} \) and we can therefore write

\[
c_k = [t_1, t_2, \ldots, t_{k+1}] y
\]

(5.14)
where \([t_1, t_2, \ldots, t_{k+1}]\) is known as the \(k\)th divided difference of \(y\). By means of further trivial calculations, it is possible to show that (Eldén and Wittmeyer-Koch, 1996)

\[
[t_k]y = y(t_k)
\]

\[
[t_1, t_2, \ldots, t_{k+1}]y = \frac{[t_2, t_3, \ldots, t_{k+1}]y - [t_1, t_3, \ldots, t_k]y}{t_{k+1} - t_1}
\]

and hence, all divided differences can be computed in the recursive manner above. In case the distance between the data points is uniform such that \(\Delta t_k = T_s\) the particularly simple form

\[
{[t_k, t_{k+1}, \ldots, t_{k+\ell+1}]y = \frac{\Delta^{\ell+1} y(t_k)}{(\ell + 1)!}}
\]

for the \((\ell + 1)\)th divided difference at the point \(t_k\) can be derived. Here, we define the forward difference operator \(\Delta\) (or delta operator) (Middleton and Goodwin, 1990) acting on a function as

\[
\Delta y(t) = \frac{y(t + T_s) - y(t)}{T_s}
\]

This means, that in general we will have the binomial expansion (Abramowitz and Stegun, 1972)

\[
\Delta^{\ell+1} y(t_k) = \sum_{l=0}^{\ell+1} \frac{(-1)^l}{T_s^{\ell+1}} \binom{\ell + 1}{l} y(t_k + (\ell + 1 - l)T_s)
\]

and the \(k\)th divided difference can be computed as

\[
{[t_k, t_{k+1}, \ldots, t_{k+\ell+1}]y = \sum_{l=0}^{\ell+1} \frac{(-1)^l}{(\ell + 1)!T_s^{\ell+1}} \binom{\ell + 1}{l} y(t_k + (\ell + 1 - l)T_s)}
\]

\[\text{5.4 Spline Interpolation}\]

Unfortunately, interpolation with very high order polynomials suffer from a serious problem. Imagine that we would like to interpolate the function

\[
f(t) = \frac{1}{1 + 25t^2}
\]

at the \(\ell\) equidistantly distributed points

\[
t_k = -1 + (k - 1)\frac{2}{\ell} \quad i = 1 \ldots \ell + 1
\]

with a polynomial \(P_\ell\) of degree \(\ell\). Then, near the endpoints, the polynomial solution will oscillate wildly and the error

\[
\lim_{\ell \to \infty} \max_{-1 \leq t \leq 1} |f(t) - P_\ell(t)| = \infty
\]
will be out of control. This effect is known as Runge’s phenomenon and indicates that interpolation by a high order polynomial is not advisable. A way to resolve this issue is to allow piecewise polynomial functions such as splines.

Let therefore \( \{ \tau_k \}_{k=1}^{N_t+1} \) be a strictly increasing sequence of points, which are called the breakpoints of the function \( f \). Also define a sequence of polynomials \( \{ P_\ell^k \}_{k=1}^{N_t+1} \) each of order \( \ell \). Then, we can define the corresponding piecewise polynomial \( \hat{y}_c \) of order \( k \) by the expression

\[
\hat{y}_c(t) = \begin{cases} 
  P_\ell^1(t) & \text{if } t < \tau_1 \\
  P_\ell^k(t) & \text{if } \tau_k < t < \tau_{k+1} \quad \forall k = 1, \ldots, N_t \\
  P_\ell^{N_t}(t) & \text{if } \tau_{N_t+1} < t.
\end{cases}
\]

The class of all such piecewise polynomial functions of order \( \ell \) with the breakpoint sequence \( \{ \tau_k \}_{k=1}^{N_t+1} \) is then denoted with \( \mathbb{P}_\ell,\tau \).

The values \( \{ \hat{y}_c(\tau_k) \}_{k=1}^{N_t+1} \) of the piecewise polynomial \( \hat{y}_c \) at the breakpoints are actually not defined by expression (5.24). Therefore we treat the function \( \hat{y}_c \) as right continuous and let

\[
\hat{y}_c(\tau_k) = \hat{y}_c(\tau_k^+) \quad k = 1, \ldots, N_t + 1
\]

where

\[
\hat{y}_c(\tau^+) = \lim_{h \to 0^+, h > 0} \hat{y}_c(\tau + h) \quad (5.26)
\]

\[
\hat{y}_c(\tau^-) = \lim_{h \to 0^-, h < 0} \hat{y}_c(\tau + h). \quad (5.27)
\]

Now, assume again that we are given function values at the points \( t_1 < t_2 < \cdots < t_{N_t+1} \). This time, we wish to interpolate these points using a piecewise polynomial function \( \hat{y}_c \) of order 2 with a continuous first derivative. Also, we wish to choose the breakpoint sequence such that

\[
\tau_k = t_k, \quad k = 1, \ldots, N_t + 1. \quad (5.28)
\]

Since each individual polynomial in \( \hat{y}_c \) has 3 individual degrees of freedom we will get \( 3(N_t + 1) \) degrees of freedom in total that has to be reduced in order to produce uniqueness. The interpolation conditions together with the continuity requirements would then yield the following \( 3(N_t + 1) \) equations

\[
\begin{align*}
\hat{y}_c(\tau_k^+) &= y_k \\
\hat{y}_c(\tau_k^-) - \hat{y}_c(\tau_k^+) &= 0 \\
\cdots
\end{align*}
\]

\[ (5.29) \]

From (5.29) it is evident that the \( 2(N_t + 1) \) homogenous equations do not directly depend on the data \( \{ y_k \}_{k=1}^{N_t+1} \). Therefore it would be possible to use these conditions in order to reduce the number of degrees of freedom in (5.29) to an absolute minimum \( N_t + 1 \).
Such a construction is equivalent to constructing a linear basis \( \{ \phi_k(t) \}_{k=1}^{N_t+1} \) for the linear mapping from the coefficients \( c_k, k = 1, \ldots, N_t + 1 \) to the interpolations points \( \hat{y}_c(t_k), k = 1, \ldots, N_t + 1 \). That interpretation means that the interpolation can be represented as

\[
\hat{y}_c(t) = \sum_{k=1}^{N_t+1} c(k) \phi_k(t).
\] (5.30)

One can therefore say that the homogenous equations will define a subspace of the space of functions \( P_{k,\tau} \) satisfying these homogenous conditions.

One particularly useful basis is the one made up of the so called \( B\)-splines ("basis splines")

\[
B_{k,\ell+1,\tau}(t) = (\tau_{k+\ell+1} - \tau_{k} \, \ldots \, \tau_{k+\ell+1} - \cdot)_{+}^\ell
\] (5.31)
of order \( \ell \) for the knot sequence \( \tau_1, \tau_2, \ldots, \tau_{n+1} \). Here, we have

\[
(\tau - t)_{+}^\ell = \begin{cases} 
(\tau - t)^\ell & \text{if } t \leq \tau \\
0 & \text{if } \tau < t.
\end{cases}
\] (5.32)

It should be noted that the knots are not equivalent to the breakpoints. There could be several knots at the same point which thereby dictate the number of continuous derivatives found there. We will however not go deeper into this issue and from now on we will assume that there is one knot at each break point. A cubic \( B\)-spline is illustrated in Figure 5.4.

Since these \( B\)-splines constitute a finite basis such that

\[
\hat{y}_c(t) = \sum_{k=1}^{N_t+1} c(k) B_{k,\ell,\tau}^c(t),
\] (5.33)

finding a function which satisfies the interpolation conditions \( \hat{y}_c(t_k) = y_k, k = 1, \ldots, N_t + 1 \), is just a matter of solving the linear system of equations

\[
\begin{pmatrix}
B_{1,\ell+1,\tau}^c(t_1) & \cdots & B_{N_t+1,\ell+1,\tau}^c(t_1) \\
\vdots & \ddots & \vdots \\
B_{1,\ell+1,\tau}^c(t_{N_t+1}) & \cdots & B_{N_t+1,\ell+1,\tau}^c(t_{N_t+1})
\end{pmatrix}
\begin{pmatrix}
c_1 \\
\vdots \\
c_{N_t+1}
\end{pmatrix} =
\begin{pmatrix}
\hat{y}_1 \\
\vdots \\
\hat{y}_{N_t+1}
\end{pmatrix}.
\] (5.34)

Since the functions \( B_k \) are known to have a small support, i.e.

\[
B_{k,\ell+1,\tau}^c(t) = 0 \quad x \notin [t_k, t_k + \ell]
\] (5.35)
the matrix \( B \) will actually have a banded structure, which makes the computation of \( c \) less demanding. In the following section, we will actually see that, if the interpolation and knot points coincide and are equidistantly distributed, the system in (5.34) can be solved at lightning speed by means of linear filtering methods.
5.5 Linear Filtering and Spline Functions

As mentioned above, the knot sequence and data points in this section are always equidistantly distributed, i.e.

\[ \tau_k = t_k = kT_s, \quad k = 1, \ldots, n + 1. \]  

(5.36)

Consequently, an \( \ell + 1 \) order spline basis with that knot sequence will be denoted by \( B_{\ell+1,T_s}^{c} \) and each individual basis functions indexed by \( k \) will only be a translation of the others such that

\[ B_{k,\ell+1,T_s}^{c}(t) = B_{0,\ell+1,T_s}^{c}(t - kT_s). \]  

(5.37)

From now on we will therefore define

\[ B_{\ell+1,T_s}^{c}(t) = B_{0,\ell+1,T_s}^{c}(t). \]  

(5.38)

Further, let us also define the sampled version of a spline function as

\[ B_{\ell+1,\tau}^{d}(k) = B_{\ell+1,\tau}^{c}(kT_s) \]  

(5.39)

Then, the expression

\[ y(lT_s) = \hat{y}_c(lT_s) \]  

(5.40)

\[ = \sum_{k=-\infty}^{\infty} c(k)B_{0,\ell+1,\tau}^{c}(T_sl - T_sk), \quad l = -\infty \ldots \infty \]  

(5.41)

will contain the interpolation conditions for a bi-infinite sequence \( \{y(kT_s)\}_{k=-\infty}^{\infty} \) of equidistantly distributed data points of some \( \ell \) times continuously differentiable function \( y(t) \). What this means is, that in order to find a spline function of order \( \ell + 1 \) with the interpolation property, one does not have to solve a system of equations as the one in (5.34) in order to find the coefficients \( c(k) \). Instead, one can concentrate on the discrete convolution equation

\[ y(lT_s) = \sum_{k=-\infty}^{\infty} c(k)B_{\ell+1,\tau}^{d}(l - k). \]  

(5.42)

which captures the Toepliz structure of the matrix in \( \mathbf{B} \) in (5.34) and utilize methods from linear filtering and linear systems. Taking the z-transform of the expression above we will then get

\[ Y(z) = \sum_{k=-\infty}^{\infty} y(kT_s)z^{-k} = C(z)B_{\ell+1,\tau}^{d}(z), \]  

(5.43)

where

\[ C(z) = \sum_{k=-\infty}^{\infty} c(k)z^{-k}, \]  

(5.44)

\[ B_{\ell+1,\tau}^{d}(z) = \sum_{k=-\infty}^{\infty} B_{\ell+1,\tau}^{d}(k)z^{-k}. \]  

(5.45)
Extracting the $z$-transform of the coefficients is then just a matter of inversion, and we will get

$$C(z) = \left[ B_{\ell+1,\tau}(z) \right]^{-1} Y(z).$$  \hspace{1cm} (5.46)

What remains is to compute the $Z$-transform of $B_{\ell+1,\tau}^d(z)$.

### 5.5.1 Z-transform of a B-Spline

Because of what we know from (5.20) about the properties of the divided difference in (5.14) for uniformly distributed data points with inter-point distance $T_s$, it is possible to

---

**Figure 5.1:** Cubic spline basis function ($\ell = 3$).
show that (Unser et al., 1993a)

\[ B_{\ell+1, T_s}(t) = (\tau_{\ell+1} - \tau_0)[\tau_0, \ldots, \tau_{\ell+1}](\cdot - t)^\ell \]

\[ = (\ell + 1)T_s[0, \ldots, (\ell + 1)T_s](\cdot - t)^\ell \]

\[ = (\ell + 1)T_s \frac{\Delta^{\ell+1}(\cdot - t)}{(\ell + 1)!} \quad \text{(5.47)} \]

\[ = \sum_{l=0}^{\ell+1} \frac{(-1)^l}{l!T_s^l} \binom{\ell + 1}{l} ((\ell + 1 - l)T_s - t)^\ell. \quad \text{(5.48)} \]

In turn, this means that the transformed version of the spline basis function can be computed as (Unser et al., 1993a)

\[ B_d^{\ell+1, T_s}(z) = \sum_{k=-\infty}^{\infty} B_d^{\ell+1, T_s}(k) z^{-k} \]

\[ = z^{-(\ell+1)} \sum_{l=0}^{\ell+1} \frac{(-1)^l}{l!} \binom{\ell + 1}{l} z^\ell \sum_{k=0}^{\infty} k^\ell z^k \]

\[ = z^{-(\ell+1)} \frac{1 - z^{\ell+1}}{\ell!} \sum_{k=0}^{\infty} k^\ell z^k. \quad \text{(5.51)} \]

Now, because of the relationship (Schoenberg, 1973)

\[ \sum_{k=0}^{\infty} k^\ell z^k = z^{\ell+1} \frac{\Pi_\ell(z)}{(1-z)^{\ell+1}}, \quad |z| \leq 1 \]

between the Euler-Frobenius polynomials \( \Pi_\ell(z) \) and the summation in (5.51), the following explicit expression for \( B_d^{\ell+1, T_s}(z) \) is possible,

\[ B_d^{\ell+1, T_s}(z) = \frac{z^{-\ell} \Pi_\ell(z)}{\ell!}. \quad \text{(5.53)} \]

Using the expressions for \( \Pi_\ell(z) \) found in Section 4.5.1 we can then show that

\[ B_d^{2, T_s}(z) = \frac{1}{z} \quad \text{(5.54)} \]

\[ B_d^{3, T_s}(z) = \frac{z + 1}{2z^2} \quad \text{(5.55)} \]

\[ B_d^{4, T_s}(z) = \frac{z^2 + 4z + 1}{6z^3} \quad \text{(5.56)} \]

\[ B_d^{5, T_s}(z) = \frac{z^3 + 11z^2 + 11z + 1}{24z^4} \quad \text{(5.57)} \]
5.5.2 Fourier Transform of a Spline

Altogether, the discussion above means that when a function is represented as a spline in (5.33), its continuous time Fourier transform will be

\[ \hat{Y}_c(i\omega) = \int_{-\infty}^{\infty} \hat{y}_c(t) e^{-i\omega t} dt \]

\[ = \int_{-\infty}^{\infty} \left( \sum_{k=-\infty}^{\infty} c(k) B_{k,\ell+1,\tau}^c(t) \right) e^{-i\omega t} dt \]

\[ = \sum_{k=-\infty}^{\infty} c(k) \int_{-\infty}^{\infty} B_{0,\ell+1,\tau}^c(t - T_s k) e^{-i\omega t} dt \]

\[ = \int_{-\infty}^{\infty} B_{0,\ell+1,\tau}^c(t) e^{-i\omega t} dt \sum_{k=-\infty}^{\infty} c(k) e^{-i\omega T_s k}. \quad (5.58) \]

If we use the relationship (Schoenberg, 1973)

\[ B_{\ell+1,\tau}^c(i\omega) = \int_{-\infty}^{\infty} B_{\ell+1,\tau}^c(t) e^{-i\omega t} dt = \left( \frac{1 - e^{-i\omega T_s}}{i\omega} \right)^{\ell+1}, \quad (5.59) \]

and insert this expression into (5.58) we get

\[ \hat{Y}_c(i\omega) = F_{\ell+1,T_s}^c(i\omega) Y_d(e^{i\omega T_s}) \]

\[ = \frac{B_{\ell+1,\tau}^c(i\omega)}{B_{\ell+1,\tau}^d(e^{i\omega T_s})} Y_d(e^{i\omega T_s}), \quad (5.61) \]

where

\[ F_{\ell+1,T_s}^c(i\omega) = \left( \frac{e^{i\omega T_s} - 1}{e^{i\omega T_s} \Pi_{\ell}(z)} \right)^{\ell+1}, \quad (5.62) \]

is identical to the function found in (4.30) in Chapter 4. This means, that estimating the continuous-time Fourier transform as in Chapter 4 is equivalent to interpolating your uniformly sampled data with an \( \ell + 1 \) order spline function with inter-knot distance \( T_s \).

5.6 Fundamental Polynomial B-Spline Function

An interesting consequence of the above line of reasoning is that if we interpret \( F_{\ell+1,T_s}^c(i\omega) \) as the continuous-time Fourier transform of some function \( F_{\ell+1,T_s}^c(t) \) such that

\[ F_{\ell+1,T_s}^c(i\omega) = \frac{B_{0,\ell+1,\tau}^c(i\omega)}{B_{0,\ell+1,\tau}^d(e^{i\omega T_s})} = \int_{-\infty}^{\infty} F_{\ell+1,T_s}^c(t) e^{-i\omega t} dt. \quad (5.63) \]
Then, $F_{\ell+1,T_s}^c(t)$ is the so-called fundamental spline function of order $\ell + 1$ (see for instance Lecture 4 in (Schoenberg, 1973)) which corresponds to the solution of the interpolation problem

$$\delta(l) = \sum_{k=-\infty}^{\infty} c(k)\varphi(l)(T_s l - T_s k) \quad l = -\infty, \ldots, \infty$$

(5.64)

where

$$\delta(l) = \begin{cases} 1 & l = 0 \\ 0 & l \neq 0 \end{cases}$$

(5.65)

is the Kronecker delta function. This means that we can also write our interpolation function in (5.33) as

$$\hat{y}_c(t) = \sum_{k=-\infty}^{\infty} y(T_s k)F_{\ell+1,T_s}^c(t - T_s k)$$

(5.66)

which is called the Lagrange form (deBoor, 1978) of the spline representation. This explains expression (5.4).

The functions $F_{\ell+1,T_s}^c(t - kT_s)$ can be thought of as an orthogonal basis for the linear mapping from the interpolation points to the spline of order $\ell + 1$ with equidistantly distributed knots. In Figure 5.2 we have portrayed the fundamental spline function $F_{\ell+1,T_s}^c(t - kT_s)$ of cubic order $\ell + 1 = 4$.

5.7 Non-Uniform Sampling and Splines

Assume again that we have the setting described in Figure 4.1. That is, a continuous-time system feeded with a piecewise constant input signal $u(t)$. Then, the conclusion of the discussion held in the sections above is that, estimating the continuous-time Fourier transform from sampled data $\{y(kT_s)\}_{k=1}^{N+1}$ as

$$\hat{Y}_c(i\omega) = F_{\ell+1,T_s}^c(i\omega)Y_d(e^{i\omega T_s})$$

(5.67)

where

$$F_{\ell+1,T_s}^c(i\omega) = \left(\frac{e^{i\omega T_s} - 1}{i\omega} \right)^{\ell + 1}$$

(5.68)

is equivalent to interpolating the data using polynomial splines of order $\ell + 1$ with equidistant knots, such that

$$\hat{y}_c(t) = \sum_{k=-\infty}^{\infty} c(k)B_{k,\ell+1,T_s}^c(t)$$

(5.69)
in a B-spline basis or

\[ \hat{y}_c(t) = \sum_{k=-\infty}^{\infty} y(T_s k) F_{c}^{e} \phi_{k, \ell+1, T_s}(t) \]  

(5.70)

in the fundamental spline basis. The reason for this is mainly that the input is a known piecewise constant function produced by running a discrete signal \( u(kT_s) \) through a zero-order hold circuit. Thereby producing an input signal \( u(t) \) such that

\[ u(t) = \sum_{k=0}^{\infty} u(kT_s) \left( H(t - kT_s) - H(t - (k + 1)T_s) \right) \]  

(5.71)

as in (2.19). The sampling is then fast enough to allow the intersample behavior to be approximated by a gain followed by \( \ell \) integrators.

In the case of irregular sampling, it seems sensible to assume that the user is still in command of the input, which is choosen as in (5.71). On the other hand, there is no control on the sampling of the output which will occur at the time instances \( \{t_k\}_{k=1}^{N_t+1} \). This means that the right way to do the interpolation would again be to use an \( \ell + 1 \) order

\[ \begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image}
\caption{Fundamental cardinal spline basis function \( f^{(\ell)} \) of cubic order \( \ell = 3 \).
\end{figure} \]
polynomial spline function as in (5.33) with equidistantly distributed knots with sample
distance $T_s$. The coefficients would be computed as

$$
\begin{pmatrix}
B_{1,\ell+1,T_s}(t_1) & \ldots & B_{N_t+1,\ell+1,T_s}(t_1) \\
B_{1,\ell+1,T_s}(t_2) & \ldots & B_{N_t+1,\ell+1,T_s}(t_2) \\
\vdots & \ddots & \vdots \\
B_{1,\ell+1,T_s}(t_{N_t+1}) & \ldots & B_{N_t+1,\ell+1,T_s}(t_{N_t+1})
\end{pmatrix}
\begin{pmatrix}
c_1 \\
c_2 \\
\vdots \\
c_{N_t+1}
\end{pmatrix}
= 
\begin{pmatrix}
y(t_1) \\
y(t_2) \\
\vdots \\
y(t_{N_t+1})
\end{pmatrix}
$$

(5.72)

and reconstructed values at the sampling point could be found as

$$
\begin{pmatrix}
\hat{y}_c(0) \\
\hat{y}_c(T_s) \\
\vdots \\
\hat{y}_c(N_t T_s)
\end{pmatrix}
= 
\begin{pmatrix}
B_{1,\ell+1,T_s}(0) & \ldots & B_{N_t+1,\ell+1,T_s}(0) \\
B_{1,\ell+1,T_s}(T_s) & \ldots & B_{N_t+1,\ell+1,T_s}(T_s) \\
\vdots & \ddots & \vdots \\
B_{1,\ell+1,T_s}(nT_s) & \ldots & B_{N_t+1,\ell+1,T_s}(N_t T_s)
\end{pmatrix}
\begin{pmatrix}
c_1 \\
c_2 \\
\vdots \\
c_{N_t+1}
\end{pmatrix}
$$

(5.73)

From these values an estimate of the discrete-time Fourier transform can then be produced as

$$
\hat{Y}_d(e^{i\omega T_s}) = \sum_{k=0}^{N_t} \hat{y}_c(k T_s) e^{i\omega T_s}
$$

(5.74)

and the continuous-time Fourier transform is estimated using

$$
\hat{Y}_c(i\omega) = F_{\ell+1,T_s}^c(i\omega) \hat{Y}_d(e^{i\omega T_s}).
$$

(5.75)

The parameter estimates are then found using the method in (4.35) of Chapter 4. Reconstruction to an equidistant grid can also be accomplished using the fundamental spline function such that

$$
\begin{pmatrix}
\hat{y}_c(0) \\
\hat{y}_c(T_s) \\
\vdots \\
\hat{y}_c(N_t T_s)
\end{pmatrix}
= 
\begin{pmatrix}
F_{1,\ell+1,T_s}^c(t_1) & \ldots & F_{N_t+1,\ell+1,T_s}^c(t_1) \\
F_{1,\ell+1,T_s}^c(t_2) & \ldots & F_{N_t+1,\ell+1,T_s}^c(t_2) \\
\vdots & \ddots & \vdots \\
F_{1,\ell+1,T_s}^c(t_{N_t+1}) & \ldots & F_{N_t+1,\ell+1,T_s}^c(t_{N_t+1})
\end{pmatrix}^{-1}
\begin{pmatrix}
y(t_1) \\
y(t_2) \\
\vdots \\
y(t_{N_t+1})
\end{pmatrix}
$$

5.8 Numerical Examples

In the following section we will illustrate how one can identify continuous-time
deterministic models from non-uniformly sampled output data using spline interpolation. The example systems

$$
G_1(s) = \frac{s + 0.5}{s^2 + 2s + 3}
$$

(5.76)

$$
G_2(s) = \frac{1}{s^3 + 2s^2 + 3s + 4}
$$

(5.77)

$$
G_3(s) = \frac{a}{s^3 + as^2 + as + a}, \quad a = 4
$$

(5.78)
are the same as those considered in the numerical illustrations and the examples in Chapter 4.

The input signal is uniform piecewise constant with randomly distributed amplitudes. The output \( y \) will be sampled at uniform time instances subject to jitter such that

\[
t_k = kT_s + \delta_k, \quad k = 1, \ldots, N_t - 1
\]

where

\[
\delta_k \in U(-\delta_0, \delta_0) \quad k = 1, \ldots, N_t - 1.
\]

The parameter value \( \delta_0 = T_s/3 \) will be used in order to generate enough non-uniformity in the sampling. For the sake of simplicity, the initial and endpoint time instances have been chosen such that \( t_0 = 0 \) and \( t_{N_t} = N_tT_s \). The output \( y(t_k) \), \( k = 1, \ldots, N_t \) is then interpolated by a polynomial a spline of order \( \ell + 1 \) using uniformly distributed knots \( \tau \) on the grid

\[
\tau_k = kT_s, \quad k = -1, \ldots, N_t + \ell.
\]

The continuous-time Fourier transform is then estimated using the method found in (4.34) of Section 4.7, and parameter estimates are found using the method in (4.35).

Illustrations of parameter estimates versus the nominal sampling time \( T_s \) are illustrated in Figure 5.3 and Table 5.1 for the system in (5.76), in Figure 5.4 and Table 5.2 for the system in (5.77) and in Figure 5.5 and Table 5.3 for the system in (5.78)

**Table 5.1:** Parameter estimates for the model \( G_1(s) = \frac{s^{1.5}}{s^2 + 2s + 3} \) versus the sampling time \( T_s \). Sampling instances have been generated such that \( t_k = kT_s + \delta_k \) where \( \delta_k \sim U(-T_s/3, T_s/3) \). The number of samples used are \( N_t = 10000 \). Relative degree of this system is \( \ell = 3 \) and reconstruction to an uniform grid has been accomplished using an \( \ell + 1 = 2 \) order polynomial spline.

<table>
<thead>
<tr>
<th>( T_s )</th>
<th>( a_1 = 2 )</th>
<th>( a_2 = 3 )</th>
<th>( b_1 = 1 )</th>
<th>( b_2 = 0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>2.0023</td>
<td>3.0071</td>
<td>0.9996</td>
<td>0.5029</td>
</tr>
<tr>
<td>0.2</td>
<td>1.9921</td>
<td>3.0098</td>
<td>0.9959</td>
<td>0.5056</td>
</tr>
<tr>
<td>0.3</td>
<td>1.9778</td>
<td>3.0232</td>
<td>0.9896</td>
<td>0.5129</td>
</tr>
<tr>
<td>0.4</td>
<td>1.9514</td>
<td>3.0425</td>
<td>0.9766</td>
<td>0.5239</td>
</tr>
<tr>
<td>0.5</td>
<td>1.9006</td>
<td>3.0459</td>
<td>0.9525</td>
<td>0.5325</td>
</tr>
<tr>
<td>0.6</td>
<td>1.8389</td>
<td>3.0653</td>
<td>0.9186</td>
<td>0.5429</td>
</tr>
<tr>
<td>0.7</td>
<td>1.7690</td>
<td>3.0734</td>
<td>0.8842</td>
<td>0.5520</td>
</tr>
<tr>
<td>0.8</td>
<td>1.6818</td>
<td>3.0696</td>
<td>0.8314</td>
<td>0.5586</td>
</tr>
<tr>
<td>0.9</td>
<td>1.5941</td>
<td>3.0486</td>
<td>0.7733</td>
<td>0.5690</td>
</tr>
<tr>
<td>1.0</td>
<td>1.4963</td>
<td>2.9653</td>
<td>0.7196</td>
<td>0.5525</td>
</tr>
</tbody>
</table>

5.9 Summary

The conclusion of this chapter, is that interpolation in terms of polynomial spline functions can be a feasible mean of reconstructing the output of continuous-time input-output
Figure 5.3: Parameter estimates for the model $G_1(s) = \frac{s^{0.5}}{s^2 + 2s + 3}$ versus the sampling time $T_s$. Sampling instances have been generated such that $t_k = kT_s + \delta_k$ where $\delta_k \sim \mathcal{U}(-T_s/3, T_s/3)$. The number of samples used are $N_t = 10000$. Relative degree of this system is $\ell = 3$ and reconstruction to an uniform grid has been accomplished using an $\ell + 1 = 2$ order polynomial spline.

models from sampled data. Splines of the order $\ell$, as the relative degree of the system could be used. The continuous-time Fourier transform is then estimated using the method found in (4.34) of Section 4.7, and parameter estimates are then found using the method in (4.35). ??
Figure 5.4: Parameter estimates for the model \( G_2(s) = \frac{1}{s^3 + 2s^2 + 3s + 4} \) versus the sampling time \( T_s \). Sampling instances have been generated such that \( t_k = kT_s + \delta_k \) where \( \delta_k \sim \mathcal{U}(-T_s/3, T_s/3) \). The number of samples used are \( N_t = 1000 \). Relative degree of this system is \( \ell = 3 \) and reconstruction to an uniform grid has been accomplished using an \( \ell + 1 = 4 \) order polynomial spline.
Table 5.2: Parameter estimates for the model $G_2(s) = \frac{1}{s^3 + 2s^2 + 3s + 4}$ versus the sampling time $T_s$. Sampling instances have been generated such that $t_k = kT_s + \delta_k$ where $\delta_k \sim \mathcal{U}(-T_s/3, T_s/3)$. The number of samples used are $N_t = 1000$. Relative degree of this system is $\ell = 3$ and reconstruction to an uniform grid has been accomplished using an $\ell + 1 = 4$ order polynomial spline.

<table>
<thead>
<tr>
<th>$T_s$</th>
<th>$a_1 = 2$</th>
<th>$a_2 = 3$</th>
<th>$a_3 = 4$</th>
<th>$b_1 = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>2.0069</td>
<td>3.0015</td>
<td>4.0117</td>
<td>1.0012</td>
</tr>
<tr>
<td>0.2</td>
<td>2.0001</td>
<td>3.0013</td>
<td>3.9995</td>
<td>0.9995</td>
</tr>
<tr>
<td>0.3</td>
<td>2.0057</td>
<td>3.0047</td>
<td>4.0104</td>
<td>1.0027</td>
</tr>
<tr>
<td>0.4</td>
<td>1.9992</td>
<td>3.0009</td>
<td>3.9991</td>
<td>0.9979</td>
</tr>
<tr>
<td>0.5</td>
<td>2.0000</td>
<td>3.0007</td>
<td>4.0003</td>
<td>0.9999</td>
</tr>
<tr>
<td>0.6</td>
<td>1.9983</td>
<td>3.0006</td>
<td>3.9962</td>
<td>0.9988</td>
</tr>
<tr>
<td>0.7</td>
<td>1.9955</td>
<td>2.9986</td>
<td>3.9910</td>
<td>0.9972</td>
</tr>
<tr>
<td>0.8</td>
<td>1.9920</td>
<td>2.9980</td>
<td>3.9827</td>
<td>0.9942</td>
</tr>
<tr>
<td>0.9</td>
<td>1.9810</td>
<td>2.9926</td>
<td>3.9599</td>
<td>0.9880</td>
</tr>
<tr>
<td>1.0</td>
<td>1.9646</td>
<td>2.9867</td>
<td>3.9250</td>
<td>0.9770</td>
</tr>
</tbody>
</table>

Table 5.3: Parameter estimates for the model $G_3(s) = \frac{a}{s^3 + as^2 + as + a}$ versus the sampling time $T_s$. The sampling instances have been generated such that $t_k = kT_s + \delta_k$ where $\delta_k \sim \mathcal{U}(-T_s/3, T_s/3)$. The number of samples used are $N_t = 1000$. Relative degree of this system is $\ell = 3$ and reconstruction to an uniform grid has been accomplished using an $\ell + 1 = 4$ order polynomial spline.

<table>
<thead>
<tr>
<th>$T_s$</th>
<th>$a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>4.0125</td>
</tr>
<tr>
<td>0.2</td>
<td>4.0171</td>
</tr>
<tr>
<td>0.3</td>
<td>4.0015</td>
</tr>
<tr>
<td>0.4</td>
<td>4.0114</td>
</tr>
<tr>
<td>0.5</td>
<td>4.0172</td>
</tr>
<tr>
<td>0.6</td>
<td>3.9900</td>
</tr>
<tr>
<td>0.7</td>
<td>3.9797</td>
</tr>
<tr>
<td>0.8</td>
<td>3.9635</td>
</tr>
<tr>
<td>0.9</td>
<td>3.9486</td>
</tr>
<tr>
<td>1.0</td>
<td>3.9303</td>
</tr>
</tbody>
</table>
Figure 5.5: Parameter estimates for the model $G_3(s) = \frac{a}{s^3 + as^2 + bs + c}$ versus the sampling time $T_s$. Sampling instances have been generated such that $t_k = kT_s + \delta_k$ where $\delta_k \sim U(-T_s/3, T_s/3)$. The number of samples used are $N_t = 1000$. Relative degree of this system is $\ell = 3$ and reconstruction to an uniform grid has been accomplished using an $\ell + 1 = 4$ order polynomial spline.
Interpolation and the Estimation of the Continuous-Time Power Spectrum

“For humans, honesty is a matter of degree. Engineers are always honest in matters of technology and human relationships. That’s why it’s a good idea to keep engineers away from customers, romantic interests, and other people who can’t handle the truth.”

Scott Adams (1957 - )

6.1 Introduction

In the two previous chapters, a framework for the identification of continuous-time input-output models from uniformly and non-uniformly sampled data was developed. The following chapter represents the start of a journey towards the identification of time series models from non-equidistantly sampled measurements. The class of models we would like to identify are the continuous time series or stochastic differential equations. In particular we will focus on the continuous-time autoregressive moving average (CARMA) models, which can be defined as

\[ y(t) = H_c(s)e(t) \]
\[ H_c(s) = \frac{B(s)}{A(s)} \]

where \( e(t) \) is used informally to denote continuous-time white noise such that

\[ E(e(t)) = 0 \]
\[ E(e(t)e(s)) = \sigma^2 \delta(t - s). \]

For a more detailed discussion on continuous-time stochastic processes, the reader is referred to Section 2.5 of Chapter 2.
The numerator and the denominator in the transfer function are
\[
A(s) = s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_n \tag{6.3}
\]
\[
B(s) = s^m + b_1 s^{m-1} + \cdots + b_m \tag{6.4}
\]
and the vector of parameters is defined as \( \theta = [a_1 \ a_2 \ \ldots \ a_n \ b_1 \ b_2 \ \ldots \ b_m \ \lambda]^T \) where \( \lambda = \sigma^2 \) is the variance of the driving noise. Here \( m < n \), so the system is strictly proper and \( E y^2(t) < \infty \). A schematic view of the setup is illustrated in Figure 6.1.

![Figure 6.1: Input and sampling setup for a continuous-time system.](image)

The perspective is again from the frequency domain, and the estimation method which the reader should have in mind is the continuous-time Whittle likelihood method
\[
\hat{\theta} = \arg \min_\theta \sum_{k=1}^{N_c} \frac{\hat{\Phi}_{c,T}(i\omega_k)}{\Phi_{c}(i\omega_k, \theta)} + \log \Phi_{c}(i\omega_k, \theta) \tag{6.5}
\]
introduced in (2.168) and (2.169) of Chapter 2. The text will mainly serve as an introduction to the problems associated with continuous-time identification of continuous-time series. The purpose is more directed towards asking questions than answering them. Answers are left to Chapter 10.

### 6.2 Outline

The chapter will be organized as follows. First, in Section 6.3 a number of different spline based methods for estimating the continuous-time Fourier transform and spectrum are introduced. These methods can be interpreted as interpolation of the covariance function, and in Section 6.4 of this interpolation in terms of spectral bias is analyzed. Here expressions both for the bias due to the interpolation and bias due to the leakage are derived. This takes up a substantial part of the chapter.

At the end, in Section 6.5, the spectral weighting at different frequencies caused by interpolation is studied. When the sampling is uniform, an optimal weighting for the estimation of the continuous-time power spectrum can be derived. It is shown that spline interpolation will cause some bias in the estimated spectrum, especially at high frequencies. Although, for a given individual frequency, the bias will decrease as \( T_s \to 0 \).
6.3 Estimation of Power Spectrum

A prerequisite for using (6.5) for identification is a good estimate \( \hat{\Phi}_{c,T} \) of the continuous-time power spectral density \( \Phi_c \). To begin with, we will therefore try to draw on the ideas developed in Chapter 5 and use interpolation (smoothing) in order to reconstruct the intersample behavior of the time series. We will restrict ourselves to using piecewise constant and piecewise linear functions to exemplify the process and it will be shown that for these cases, the reconstruction \( \hat{y}(t) \) of the output \( y(t) \) can be interpreted as a way to interpolate the continuous-time covariance kernel \( r(s, t) = E[y(t)y(s)] \) and the associated covariance function \( r(\tau) \). This will put a bound on the bias of the continuous-time periodogram

\[
\hat{\Phi}_{c,T}(i\omega) = \left| \hat{Y}_{c,T}(i\omega) \right|^2
\]

(6.6)

as an estimate of the continuous-time spectrum

\[
\Phi_c(i\omega) = \int_{-\infty}^{\infty} r(\tau) e^{-i\omega \tau} d\tau
\]

(6.7)

where

\[
\hat{Y}_{c,T}(i\omega) = \frac{1}{\sqrt{T}} \int_{0}^{T} \hat{y}(t)e^{-i\omega t} dt.
\]

(6.8)

A consequence of the method in (6.5) is that the bias in the spectral estimate \( \hat{\Phi}_{c,T}(i\omega) \) in (6.6) will end up as bias in the parameter estimates \( \hat{\theta} \). A discussion about this translation from spectral bias to parameter bias will be the topic of Chapter 7. Here, we will concentrate on the spectral bias that originates from the interpolation process.

At the end of the chapter we will return to the ideas in Chapter 4, where we used the properties of the sampled continuous-time system in order to create an estimate of the continuous-time Fourier transform. This discussion will lead to the conclusion that piecewise constant and piecewise linear interpolation is probably not the best choice for estimating the continuous-time spectrum. A method similar to the one in Chapter 4, which is based on the properties of the sampled time series models will then be devised in Chapter 10 and more fully developed for the case of non-equidistant sampling in Chapter 10.

The exact method for continuous-time identification in (6.5) requires knowledge of the continuous-time realizations of the output \( y(t) \) via (6.8) and (6.6). This form of continuous-time sampling is of course not practically possible since it would require an infinite amount of storage. Instead, during system identification, one has to be content with a number of discrete samples distributed uniformly or non-uniformly during the time of observation \([0, T]\). This sampling will cause a loss of information if there is little or no knowledge of what happens in between samples.

In order to fill this gap, we will therefore resort to interpolation as a means of reconstructing the output of a time series model in continuous-time. In this chapter, as is
Figures 6.2: Piecewise constant and piecewise linear interpolation.

Illustrated in Figure 6.2, we will reconstruct the spline functions defined in Chapter 5

\[ \hat{y}_c(t) = \sum_{k=1}^{N_t} y(t_k) F_{k,\ell+1,\tau}(t) \]  

(6.9)

where we use a non-uniform knot sequence \( \{\tau_k\}_{k=1}^{N_t} \) such that

\[ \tau_k = t_k \quad k = 1, \ldots, N_t. \]  

(6.10)

Here, we will restrict the order of the spline such that \( \ell \in \{-1, 0, 1\} \), and by \( \ell = -1 \) we mean that we interpolate such that

\[ F_{k,0,\tau}(t) = \delta(t - t_k) \]  

(6.11)

which we term Riemann interpolation. Here \( \delta(t) \) is the Dirac delta function. In case of piecewise constant interpolation (\( \ell = 0 \)) we will have the B-splines

\[ F_{k,1,\tau}(t) = \begin{cases} 
0 & t < t_k \\
1 & t_k \leq t \leq t_{k+1} \\
0 & t_{k+1} < t.
\end{cases} \]  

(6.12)

Finally, the basis function for piecewise linear interpolation (\( \ell = 1 \)) will be the classical tent shaped functions

\[ F_{k,2,\tau}(t) = \begin{cases} 
0 & t < t_k \\
\frac{t-t_k}{t_{k+1}-t_k} & t_k \leq t < t_{k+1} \\
\frac{t_{k+2}-t}{t_{k+2}-t_{k+1}} & t_{k+1} \leq t < t_{k+2} \\
0 & t_{k+2} \leq t.
\end{cases} \]  

(6.13)
The Fourier transform of the interpolated realization can then be computed from the reconstructed continuous-time output $\hat{y}_c(t)$ as

$$\hat{Y}_{c,T}(i\omega) = \frac{1}{\sqrt{T}} \int_0^T \hat{y}_c(t)e^{-i\omega t} dt.$$  

(6.14)

If we sample the interpolated output $\hat{y}_c(t)$ continuously and perform the Fourier transform as in the resulting formulas for interpolation will be

$$\hat{Y}_{c,T}(i\omega_n) = \frac{1}{\sqrt{T}} \sum_{k=1}^{N_t} y(t_k)e^{-i\omega_n t_k}$$  

(6.15)

$$\hat{Y}_{c,T}(i\omega_n) = \frac{1}{\sqrt{T}} \sum_{k=1}^{N_t-1} y(t_k)\frac{1-e^{-i\omega_n \Delta t_k}}{i\omega_n} e^{-i\omega_n t_k}$$  

(6.16)

$$\hat{Y}_{c,T}(i\omega_n) = y(t_{N_t})\frac{e^{i\omega_n \Delta t_{N_t-1}} - 1}{(i\omega_n)^2 \Delta t_{N_t-1}} + y(t_1)\frac{e^{-i\omega_n \Delta t_1} - 1}{(i\omega_n)^2 \Delta t_1}$$

$$+ \frac{1}{\sqrt{T}} \sum_{k=2}^{N_t-1} y(t_k)\left(\frac{e^{i\omega_n \Delta t_{k-1}} - 1}{(i\omega_n)^2 \Delta t_{k-1}} + \frac{e^{-i\omega_n \Delta t_k} - 1}{(i\omega_n)^2 \Delta t_k}\right) e^{-i\omega_n t_k}$$  

(6.17)

where the sequence $\{\omega_n\}_{n=1}^{N_{\omega}}$ contains the frequencies where the transform is evaluated. For non-uniformly sampled data, this procedure will be costly, because the effective algorithms of the fast fourier transform cannot be used. Instead the transform will be computed at the cost of $O(N_{\omega}N_t)$ operations. It should also be noted that the interpolation will cause a spectral bias, which will be the topic of the next section.

### 6.4 Interpolation and Spectral Bias

The approximation $\hat{y}_c(t)$ of a realization of the stationary process $y(t)$ will have the following second order properties

$$\hat{m}(t) = E[\hat{y}_c(t)] = 0$$  

(6.18)

$$\hat{r}(t, s) = E[\hat{y}_c(t) - \hat{m}(t)][\hat{y}_c(s) - \hat{m}(s)]$$

$$= \sum_{k=1}^{N} \sum_{l=1}^{N} E y(t_k) y(t_l) F_{k,\ell+1,\tau}(t) F_{\ell,\ell+1,\tau}(s)$$

such that

$$\hat{r}(t, s) = \sum_{k=1}^{N} \sum_{l=1}^{N} r(t_k, t_l) F_{k,\ell+1,\tau}(t) F_{\ell,\ell+1,\tau}(s)$$  

(6.20)

where $r$ is the covariance kernel defined in (2.7). An interesting question is how well $\hat{r}(t, s)$ approximates $r(t, s)$. 
From the representation in (6.20) we see that piecewise constant interpolation of \( y(t) \) will result in a piecewise constant interpolation of \( r(t,s) \). Piecewise linear interpolation of \( y(t) \) on the other hand, will result in piecewise bilinear interpolation of \( r(t,s) \). In general we will get a so called tensor-product spline (Schumaker, 1981).

If we define

\[
    h_{\text{max}} = \max_{1 \leq i < N_t-1} (t_{i+1} - t_i).
\]

and

\[
    h_{\text{min}} = \min_{1 \leq i < N_t-1} (t_{i+1} - t_i).
\]

a bound on the interpolation error caused by such a two-dimensional interpolation is defined by the following theorem

**Theorem 6.1**

Assume that \( r \) is the covariance function defined in (2.7) and \( \hat{r} \) is the tensor-product spline interpolant defined in (6.20). Then there exists a constant \( C \) depending only on

\[
    \gamma = \frac{h_{\text{max}}}{h_{\text{min}}}
\]

such that

\[
    \max_{s,t \in [0,T]} |\hat{r}(t,s) - r(t-s)| \leq C h_{\text{max}}^{\ell+1}.
\]

**Proof:** See p. 491 Schumaker (1981).

The error in this estimate of the continuous-time covariance will translate into bias in the estimate of the continuous-time spectrum. Let us therefore define the continuous-time periodogram as

\[
    \hat{\Phi}_c,T(i\omega) = \left| \hat{Y}_{c,T}(i\omega) \right|^2
\]

when the signal \( y \) is interpolated and

\[
    \hat{\Phi}_c,T(i\omega) = \left| Y_{c,T}(i\omega) \right|^2
\]

when the signal \( y(t) \) is sampled continuously. The approximate periodogram, which is an estimate of the power spectrum, will be biased due to the interpolation and the limited time of observation \( T \). This bias, as we will see in Chapter 7, translates into the bias in the parameter estimates. We are therefore interested in controlling the size of the bias quantity

\[
    \Delta \Phi(i\omega) \triangleq \hat{\Phi}_c,T(i\omega) - \Phi_c(i\omega).
\]

In this process, a bound on the expected value of the difference between the spectral estimate from reconstructed continuous-time data \( \hat{\Phi}_c,T \) and from continuously sampled data \( \hat{\Phi}_c,T \)

\[
    E \left( \left| \hat{\Phi}_c,T(i\omega) - \hat{\Phi}_c,T(i\omega) \right| \right)
\]
Then a bound on the expected value of the difference between the spectral estimate from continuously sampled data $\hat{\Phi}_{c,T}$ and the true spectrum $\Phi_c$

$$|E\hat{\Phi}_{c,T}(i\omega) - \Phi_c(i\omega)|$$  \hspace{1cm} (6.28)

will be derived. The result will then follow by use of the triangle inequality.

### 6.4.1 Bias Due to Interpolation

The following expression relates the bias in the approximate periodogram to the error from the interpolation of the covariance function.

**Lemma 6.1**

Let $y(t)$ be a stationary stochastic process. Let $\hat{\Phi}_{c,T}(i\omega)$ be defined by (6.26) and let $\hat{\hat{\Phi}}_{c,T}(i\omega)$ be defined by (6.25). Then, for

$$\Delta \hat{\Phi}_1 = \left| E\left(\hat{\Phi}_{c,T}(i\omega) - \hat{\Phi}_{c,T}(i\omega)\right) \right|$$

we have

$$\Delta \hat{\Phi}_1(i\omega) \leq C_1 \max_{t,s \in [0,T]} |r(t-s) - \hat{r}(t,s)|, \quad \forall \omega$$ \hspace{1cm} (6.30)

when $\hat{r}(t,s)$ is defined as in (6.20).

**Proof:** From the definitions we get

$$\Delta \hat{\Phi}_1 = \left| E|\hat{Y}_{c,T}(i\omega)|^2 - E|Y_c(i\omega)|^2 \right|$$

$$= \left| \frac{1}{T} \int_0^T \int_0^T (\hat{r}(t,s) - r(t-s)) e^{i\omega(t-s)} dt ds \right|$$

$$\leq \frac{1}{T} \int_0^T \int_0^T |\hat{r}(t,s) - r(t-s)| dt ds.$$

Let $\Delta r(t,s) = |\hat{r}(t,s) - r(t-s)|$ and make a change of variables. Then we get

$$\frac{1}{T} \int_0^T \int_0^T \Delta r(t,s) dt ds = \frac{1}{T} \int_0^T \int_{-t}^{T-t} \Delta r(t,t+\tau) d\tau dt$$

$$= \frac{1}{T} \int_0^{\min\{T-t,\alpha\}} \int_{-\alpha,-t}^{T-\alpha} \Delta r(t,t+\tau) d\tau dt$$

$$+ \frac{1}{T} \int_0^{\alpha} \int_{T-\alpha}^{T-t} \Delta r(t,t+\tau) d\tau dt$$

$$+ \frac{1}{T} \int_{-\alpha}^{\alpha} \int_{T-\alpha}^{T-t} \Delta r(t,t+\tau) d\tau dt$$

$$+ \frac{1}{T} \int_{-\alpha}^{\alpha} \int_0^{\min\{\alpha,T-t\}} \Delta r(t,t+\tau) d\tau dt$$
These terms can be separately bounded. For the first term we have

\[
\frac{1}{T} \int_0^T \int_0^{\min\{\alpha, T-t\}} \Delta r(t, t+\tau) d\tau dt \leq \frac{1}{T} \int_0^T \int_0^{\max\{-\alpha, -t\}} Ch^m_{\max} d\tau dt
\]

\[
\leq \alpha Ch^m_{\max}
\]

For some \( \alpha > 0 \) and \( \lambda > 0 \) we have

\[
\Delta r(\tau) \leq |r(\tau)| + |\hat{r}(t, t+\tau)| < e^{-\lambda|\tau|}, \quad |\tau| \geq \alpha
\]

The second term can then be bounded the following way

\[
\frac{1}{T} \int_{\alpha}^{-\alpha} \int_{-t}^{T-\alpha} \Delta r(t, t+\tau) d\tau dt \leq \frac{1}{T} \int_{\alpha}^{-\alpha} \int_{-t}^{T-\alpha} e^{-\lambda|\tau|} d\tau dt
\]

\[
\leq \int_{-\infty}^{\alpha} e^{-\lambda|\tau|} d\tau.
\]

We then choose \( \alpha > 0 \) such that

\[
\alpha Ch^m_{\max} \geq \int_{-\infty}^{\alpha} e^{-\lambda|\tau|} d\tau
\]

and we have our result. \( \square \)

### 6.4.2 Bias Due to Leakage

Since we only observe our process during a finite time interval \([0, T]\) the expected periodogram and the spectrum will be slightly different. The following lemma by Ljung (Ljung, 1999) quantifies this difference.

**Lemma 6.2**

*Let \( y(t) \) be a stationary stochastic process with spectrum \( \Phi_c \) and let \( \hat{\Phi}_{c,T} \) be defined by (6.26), then*

\[
\Delta \hat{\Phi}_2(i\omega) = \left| E \hat{\Phi}_{c,T}(i\omega) - \Phi_c(i\omega) \right| \leq \frac{C_2}{T}, \quad \forall \omega \tag{6.31}
\]

**Proof:** Lemma 6.1 in (Ljung, 1999) adopted to continuous-time.

\[
EY_T(i\omega)Y_T(-i\xi) = \frac{1}{T} \int_0^T \int_0^T \int_0^T e^{i\omega(\omega r - \xi s)} d\sigma dr ds
\]

\[
= \frac{1}{T} \int_0^T \int_0^T r(t-s) e^{-i(\omega t - \xi s)} ds dt
\]

\[
= \frac{1}{T} \int_0^T e^{-i(\omega-\xi)t} \int_{t-T}^t r(\tau) e^{-i\xi \tau} d\tau dt
\]
Now
\[ \int_{t-T}^{t} r(\tau)e^{-i\xi \tau} d\tau = \Phi_c(i\xi) - \int_{-\infty}^{-\infty} r(\tau)e^{-i\xi \tau} d\tau \]
and
\[ \int_{t}^{t} \int_{-\infty}^{\infty} e^{-i(\omega-\xi)\tau} d\tau dt = \begin{cases} 1, & \text{if } \omega = \xi \\ 0, & \text{if } (\omega - \xi) = \frac{2\pi}{T} k, \quad k = \pm 1, \pm 2, \ldots, \pm \infty. \end{cases} \]

Consider
\[ \left| \frac{1}{T} \int_{0}^{T} e^{-i(\omega-\xi)\tau} \int_{-\infty}^{t-T} r(\tau)e^{-i\xi \tau} d\tau dt \right| \leq \frac{1}{T} \int_{0}^{T} \int_{-\infty}^{t-T} |r(\tau)| d\tau dt \]
\[ \leq \frac{1}{T} \int_{0}^{\infty} |\tau| |r(\tau)| d\tau \leq \frac{C}{T} \]
provided
\[ \int_{-\infty}^{\infty} |\tau r(\tau)| d\tau. \]

Similarly
\[ \left| \frac{1}{T} \int_{0}^{T} e^{-i(\omega-\xi)\tau} \int_{-\infty}^{t-T} r(\tau)e^{-i\xi \tau} d\tau dt \right| \leq \frac{1}{T} \int_{0}^{\infty} |\tau| |r(\tau)| d\tau \leq \frac{C}{T} \]

\[ \square \]

6.4.3 Periodogram Bias

The previous two lemmas can now be used to estimate the difference between the expected approximate periodogram and the spectrum.

Theorem 6.2

Let
\[ \Delta \hat{\Phi}(i\omega) = \left| E\hat{\Phi}_{c,T}(i\omega) - \Phi_c(i\omega) \right| \]
where \( \Phi_c \) is the continuous-time spectrum defined in (2.73), \( \hat{\Phi}_{c,T} \) is the continuous-time periodogram by interpolation, and

\[
h_{\text{max}} = \max_{1 \leq i \leq N-1} t_{i+1} - t_i
\]

Then

\[
\Delta \hat{\Phi}(i\omega) \leq C_1 h_{\text{max}}^{\ell+1} + \frac{C_2}{T}, \quad \forall \omega
\]

(6.32)

where \( C_1 \) and \( C_2 \) are positive numbers.

**Proof:** By Lemma 6.1, Lemma 6.2, and the triangle inequality, the result

\[
\Delta \hat{\Phi}(i\omega) = \left| E \hat{\Phi}_{c,T}(i\omega) - \Phi_c(i\omega) \right|
\]

\[
\leq \left| E \hat{\Phi}_{c,T}(i\omega) - E \hat{\Phi}_{c,T}(i\omega) + E \hat{\Phi}_{c,T}(i\omega) - \Phi_c(i\omega) \right|
\]

\[
\leq \left| E \left[ \hat{\Phi}_{c,T}(i\omega) - \Phi_c(i\omega) \right] \right| + \left| E \left[ \hat{\Phi}_{c,T}(i\omega) - \Phi_c(i\omega) \right] \right|
\]

\[
\leq C_1 \max_{s,t \in [0,T]} |\hat{r}(s,t) - r(s-t)| + \frac{C_2}{T}
\]

follows. \( \square \)

### 6.5 Uniform Sampling

Until now, we have studied the general case of non-uniformly sampled data. The interpolation scheme has been based on piecewise constant and piecewise linear functions. We have shown that for each individual frequency \( \omega \), the bias of the estimated spectrum will decrease with \( T_s \). The question that remains to be asked is if we could do better than this.

In this section we will go back to the basics as in Chapter 4 in order to sort things out. This means that we will concentrate on uniformly sampled data in order to interpret the effect of the interpolation in the frequency domain. We will then show that interpolation by splines might not be the most logical choice.

In the case of uniformly sampled data where the sequence of sampling points is chosen as

\[
t_k = kT_s \quad k = 1, \ldots, N_t
\]

(6.33)

the truncated Fourier transform of the interpolated data will become

\[
\hat{Y}_{c,T}(i\omega) = \frac{1}{\sqrt{T}} \int_0^T \sum_{k=1}^N y(kT_s) F_{\ell+1,T_s}(t - kT_s)e^{-i\omega t} dt
\]

\[
= F_{\ell+1,T_s}(i\omega) \frac{1}{\sqrt{N}} Y_d(e^{i\omega T_s})
\]

(6.34)
where

\[ F_{\ell+1, T_s}(i\omega) = \frac{1}{\sqrt{T_s}} \int_{-\infty}^{\infty} F_{\ell+1, T_s}(t)e^{-i\omega t} dt \]  

(6.35)

and the discrete-time Fourier transform is

\[ Y_d(e^{i\omega T_s}) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} y(kT_s)e^{-i\omega kT_s}. \]  

(6.36)

In connection with the expressions, it is worth mentioning that when we move from discrete-time to continuous-time the Nyquist frequency will become infinite. Since we have made an assumption about the intersample behavior it is possible for frequencies above the Nyquist frequency to carry information. However, as will be seen later on, these frequencies should be left out for several reasons.

First however, we will focus on the frequency-domain effects of the interpolation. If we define the discrete-time periodogram as

\[ \hat{\Phi}_d(e^{i\omega T_s}) = |Y_d(e^{i\omega T_s})|^2 \]  

(6.37)

the continuous-time periodogram estimate based on (6.34), (6.35) and (6.36) can be written as

\[ \hat{\Phi}_{c,T}(i\omega) = |F_{\ell+1, T_s}(i\omega)|^2 \hat{\Phi}_d(e^{i\omega T_s}) \]  

(6.38)

for uniform sampling and interpolated data. For the interpolation orders determined by \(\ell = \{-1, 0, 1\}\), this corresponds to

\[ |F_{0, T_s}(i\omega)|^2 = T_s \]

\[ |F_{1, T_s}(i\omega)|^2 = T_s \left| \frac{1 - e^{-i\omega T_s}}{i\omega T_s} \right|^2 \]

\[ = \frac{T_s}{2} \left| e^{i\frac{\omega T_s}{2}} - e^{-i\frac{\omega T_s}{2}} \right|^2 \]

\[ = \frac{T_s}{2} \sin^2 \frac{\omega T_s}{2} \]

\[ |F_{2, T_s}(i\omega)|^2 = \frac{1}{T_s} \left| \frac{e^{i\omega T_s} - 1}{(i\omega)^2 T_s} + \frac{e^{-i\omega T_s} - 1}{(i\omega)^2 T_s} \right|^2 \]

\[ = \frac{1}{T_s} \left| \frac{e^{i\omega T_s} - 2 + e^{-i\omega T_s}}{(i\omega)^2 T_s} \right|^2 \]

\[ = \frac{T_s}{16} \left| e^{i\frac{\omega T_s}{2}} - e^{-i\frac{\omega T_s}{2}} \right|^4 \]

\[ = \frac{T_s}{16} \sin^4 \frac{\omega T_s}{2} \]
For higher order spline interpolation we will get
\[
\hat{y}_c(t) = \sum_{k=1}^{N_t} y(kT_s) F_{\ell+1,T_s}(t - kT_s)
\] (6.39)
will as in Chapter 5. Here, \(F_{\ell+1,T_s}\) is the fundamental spline function in (5.63).

The objective of the interpolation is to produce a consistent estimate of the continuous-time power spectrum. Is this kind of interpolation the most logical choice?

If we think in the same way as in Chapter 4 and assume that we know the true parameters \(\theta_0\) of the continuous time series model, then we would know the power spectral densities \(\Phi_c(i\omega, \theta_0)\) and \(\Phi_d(e^{i\omega T_s}, \theta_0)\) of the continuous and sampled output. Therefore, the analogous reasoning to that found in Section 4.7 is to estimate the continuous-time spectrum as
\[
\hat{\Phi}_{c,T}(i\omega) = |F_{opt}(i\omega)|^2 \hat{\Phi}_d(e^{i\omega T_s})
\] (6.40)
where
\[
|F_{opt}(i\omega)|^2 = \frac{\Phi_c(i\omega, \theta_0)}{\Phi_d(i\omega, \theta_0)}
\] (6.41)

In Figure 6.3 we have compared the spectrum \(|F_{opt}(i\omega)|^2\) for the time series model
\[
y(t) = H(s)e(t)
\] (6.42)
\[
H(s) = \frac{1}{s^2 + 2s + 1}
\] (6.43)
to that of piecewise constant \(|F_{1,T_s}(i\omega)|^2\) and piecewise linear \(|F_{2,T_s}(i\omega)|^2\). From this perspective we see that the piecewise linear interpolation is actually worse for the particular system than piecewise constant interpolation. Anyway neither of them can compete with the optimal choice, \(|F_{opt}(i\omega)|^2\) in (6.41), if the true parameter \(\theta_0\) is known.

### 6.6 Summary

In this chapter, estimation of the continuous-time spectrum from sampled data has been treated. The main approach gas been the use of spline interpolation in order to bridge the gap between the continuous and discrete domains. Here it was discovered that interpolating the output is actually equivalent to interpolating the covariance function. It could be shown that for each individual frequency, the spectral bias caused by the interpolation would decrease as \(T_s \to 0\). If, on the other hand kept the sampling time \(T_s\) constant, there would always be a significant high frequency bias present. In the next chapter, ways to limit the effect of bias like this on parameter estimates will be treated.
Figure 6.3: Absolute square of the Fourier transforms of the interpolation kernels for $T_s = 0.1$. The optimal kernel (solid) is for a second order model where $H(s) = s^2 + 2s + 1$ is compared with piecewise constant (dotted) and piecewise linear (dash dot).
“Compromise, if not the spice of life, is its solidity. It is what makes nations great and marriages happy.”

Phyllis McGinley (1905-1978)

7.1 Introduction

In practical identification applications, the amount and quality of available data can vary from situation to situation. This makes tools and rules of thumb valuable if they can make the procedure of designing identification algorithms easier. The following chapter includes a set of such qualitative tools aiding the user in his efforts.

First of all, it will be shown what the asymptotic bias of the parameter estimates obtained by the continuous-time Whittle likelihood method in (2.168) and (2.169) will be if the spectrum is disturbed in some sense. In this context we will assume that the measured signal is corrupted by a random Gaussian disturbance, such that

\[
y_m(t) = y(t) + e(t). \tag{7.1}
\]

and \( y(t) \) is independent of \( e(t) \). This means that the continuous-time periodogram in (6.26) of the measured output will take the form of

\[
\hat{\Phi}_m(i\omega) = \hat{\Phi}(i\omega, \theta_0) + \hat{\Phi}_e(i\omega) \tag{7.2}
\]

where the signal \( y \) and the disturbance \( e \) are distributed such that

\[
\hat{\Phi}(i\omega_k) \sim \text{AsExp} \Phi(i\omega_k, \theta_0) \quad k = 1, \ldots, N_\omega \tag{7.3}
\]

\[
\hat{\Phi}_e(i\omega_k) \sim \text{AsExp} \Phi_e(i\omega_k). \tag{7.4}
\]
The continuous-time Whittle likelihood method in (2.168) and (2.169), then simply consists of minimizing the negative log likelihood $L$, where

$$
\hat{\Phi}_m = \left( \hat{\Phi}_m(i\omega_1) \ldots \hat{\Phi}_m(i\omega_{N_\omega}) \right),
$$

such that

$$
\hat{\theta} = \arg \min_{\theta} L(\theta, \hat{\Phi}_m)
$$

$$
L(\theta, \hat{\Phi}_m) = \sum_{k=1}^{N_\omega} \frac{\hat{\Phi}_m(i\omega_k)}{\Phi(i\omega_k, \theta)} + \log \Phi(i\omega_k, \theta)
$$

In that case, it will be shown that the parameter bias produced by the disturbance $\hat{\Phi}_e$ will be

$$
b(\theta_0) = E(\hat{\theta} - \theta_0) \approx \sum_{k=1}^{N_\omega} S(i\omega_k) \Delta \Phi(i\omega_k),
$$

where

$$
\Delta \Phi(i\omega_k) = \frac{E\hat{\Phi}_m(i\omega_k) - \Phi(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)} = \frac{\Phi_e(i\omega_k)}{\Phi(i\omega_k, \theta_0)}
$$

denotes the relative bias of the periodogram (observe that the expectation is over both $y$ and $e$). In this expression we also have

$$
S(i\omega_k) = J(\theta_0)^{-1} J_k(\theta_0)
$$

where the asymptotic “information” is

$$
J(\theta_0) = \sum_{k=1}^{N_\omega} J_k(\theta_0) J_k(\theta_0)^T
$$

and where the “information content” from each individual frequency is

$$
J_k(\theta_0) = \frac{\Phi_e(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)}.
$$

We call this last quantity the relative sensitivity. The parameter variance will be approximately inversely proportional to $J(\theta_0)$. A related and interesting discussion concerning approximate bias and variance properties of non-linear estimators in image processing can be found in an article by Fessler (1996).

### 7.2 Outline

The chapter is organized as follows. First, theoretical asymptotic expressions for the bias and variance of parameter estimates are derived in Section 7.3 and 7.4. These first results
basically states that it is the relative error (bias) in the spectrum that affects the bias found in the parameters. Similarly, it is the “relative sensitivity” of the spectrum to changes in the parameters that dictates the expression for the variance. These insights are then used in Section 7.5 in order to provide practical rules of thumb for frequency selection. These ideas are also explained and illustrated by two examples.

7.3 Bias Expression

In the previous chapter, various ways of estimating the continuous-time periodogram using spline interpolation were investigated. A common property of this approach was that it would yield a biased estimate of the continuous-time spectrum at high frequencies. Such wide band spectral bias would have an impact on the bias of parameter estimates, and the purpose of this section is therefore to derive an expression which relates the spectral bias to this parameter bias.

**Theorem 7.1**

Let $L$ be defined as in (7.7) and let the estimated parameters be defined by

$$
\hat{\theta} = \arg \min_{\theta} L(\theta, \hat{\Phi}_m),
$$

(7.13)

with the expected value

$$
\theta^* = \arg \min_{\theta} E L(\theta, \hat{\Phi}_m) = \arg \min_{\theta} L(\theta, E\hat{\Phi}_m).
$$

(7.14)

If we assume that the true signal model is in the model class, where

$$
\Phi(\theta_0) = (\Phi(i\omega_1, \theta_0) \ldots \Phi(i\omega_N, \theta_0))
$$

(7.15)

then

$$
\theta_0 \triangleq \arg \min_{\theta} L(\theta, \Phi(\theta_0)).
$$

(7.16)

Hence, the spectral bias translates into parameter bias such that

$$
b(\theta_0) = \theta^* - \theta_0 = \sum_{k=1}^{N_\omega} S(i\omega_k)\Delta\Phi(i\omega_k) + O \left( \|\theta^* - \theta_0\|^2 \right).
$$

(7.17)

where $S$ and $\Delta\Phi$ are defined as in (7.9) and (7.10) respectively.

**Proof:** From the definition in (7.14) and (7.16) we have

$$
L'_\theta(\theta^*, E\hat{\Phi}_m) = 0
$$

$$
L'_\theta(\theta_0, \Phi(\theta_0)) = 0.
$$

Let

$$
\hat{\Phi} = \left( \hat{\Phi}(i\omega_1), \ldots, \hat{\Phi}(i\omega_N) \right)
$$

(7.18)
A Taylor expansion of $L'(\theta, \hat{\Phi})$ around $\theta_0$ and $\Phi(\theta_0)$ then yields

$$L'(\theta^*, E\hat{\Phi}_m) = L'(\theta_0, \Phi(\theta_0)) + L''_{\theta\theta}(\theta_0, \Phi(\theta_0))(\theta^* - \theta_0)$$

$$+ \sum_{k=1}^{N_\omega} L''_{\Phi_k\theta}(\theta_0, \Phi(\theta_0)) \left( E\Phi_m(i\omega_k) - \Phi(i\omega_k) \right)$$

$$+ \mathcal{O}(\|\theta^* - \theta_0\|^2)$$

if $\hat{\Phi}_k$ is the $k$th component of $\hat{\Phi}$, since $L$ is linear in each $\hat{\Phi}_k$. From Corollary 7.1 and Lemma 7.2 in Appendix A, it then follows that

$$L''_{\theta\theta}(\theta_0, \Phi(\theta_0)) = \sum_{k=1}^{N_\omega} \Phi'(i\omega_k, \theta_0) \left( \frac{\Phi'_\theta(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)} \right)^T$$

and

$$L''_{\Phi_k\theta}(\theta_0, \Phi(\theta_0)) = -\frac{\Phi'_\theta(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)^2}.$$ 

Using the definitions of $\Delta\Phi$ in (7.9), $S$ in (7.10), $J$ in (7.11) and $J_k$ in (7.12) we can write

$$\theta^* - \theta_0 = \sum_{k=1}^{N_\omega} S(i\omega_k)\Delta\Phi(i\omega_k) + \mathcal{O} \left( \|\hat{\theta} - \theta_0\|^2 \right).$$

\[\Box\]

### 7.4 Variance Expression

For a user of estimation methods, the mean square error (MSE)

$$MSE(\theta_0) = E(\hat{\theta} - \theta)(\hat{\theta} - \theta)$$

is common measure of quality. In the case of an unbiased estimator we know from the Cramér-Rao lower bound\(^{(Cramér, 1946; Rao, 1946)}\) that

$$MSE(\theta_0) \geq M(\theta_0)^{-1}$$

where

$$M = EL'_\theta(\theta_0, \hat{\Phi}) L'_\theta(\theta_0, \hat{\Phi})^T$$

if $L$ is the negative log likelihood function. For the Whittle likelihood estimator in Lemma 7.1 in Appendix A tells us that

$$L'_\theta(\theta_0, \hat{\Phi}) = \sum_{k=1}^{N_\omega} \Phi'_\theta(i\omega_k, \theta) \left( 1 - \frac{\hat{\Phi}(i\omega_k)}{\Phi(i\omega_k, \theta)} \right).$$

(7.22)
and therefore, we will have

\[ M(\theta_0) = E \sum_{k=1}^{N_\omega} \sum_{l=1}^{N_\omega} \frac{\Phi'_{\omega}(i\omega_k, \theta) \Phi'_{\omega}(i\omega_l, \theta)}{\Phi(i\omega_k, \theta) \Phi(i\omega_l, \theta)} \left( 1 - \frac{\hat{\Phi}(i\omega_k)}{\Phi(i\omega_k, \theta)} \right) \left( 1 - \frac{\hat{\Phi}(i\omega_l)}{\Phi(i\omega_l, \theta)} \right). \]  

(7.23)

Assume further that there is no spectral bias, and that \( \hat{\Phi}(i\omega_k) \) and \( \hat{\Phi}(i\omega_l) \) are independent if \( k \neq l \). Then,

\[ M(\theta_0) = \sum_{k=1}^{N_\omega} \frac{\Phi'_{\omega}(i\omega_k, \theta) \Phi'_{\omega}(i\omega_l, \theta)}{\Phi(i\omega_k, \theta) \Phi(i\omega_l, \theta)} = J(\theta_0) \]  

(7.24)

since

\[ E \left( 1 - \frac{\hat{\Phi}(i\omega_k)}{\Phi(i\omega_k, \theta)} \right) \left( 1 - \frac{\hat{\Phi}(i\omega_l)}{\Phi(i\omega_l, \theta)} \right) = 1 \]  

(7.25)

when \( E(\hat{\Phi}(i\omega_k) - \Phi(i\omega_k))^2 = \Phi(i\omega_k)^2 \). This means that for the Whittle likelihood estimator with unbiased spectral information, we will have

\[ MSE(\theta_0) \geq J(\theta_0)^{-1}. \]  

(7.26)

where \( J(\theta_0) \) is defined as in (7.11) Therefore, the relative sensitivity will play a role in the expression for the variance as well as for the bias.

In the presence of bias the expression in (7.26) changes and the MSE will be the sum of a bias and variance contribution

\[ MSE(\theta_0) = b(\theta_0)b(\theta_0)^T + (I + \nabla_{\theta}b(\theta_0)) J^{-1}(\theta_0) (I + \nabla_{\theta}b(\theta_0))^T. \]  

(7.27)

Hence, the user should be aware that the MSE will increase due to the bias. First, because of the bias itself, but also since the variance will increase if the bias is sensitive to changes in the true parameters \( \theta_0 \). Here however, we are content with using the unbiased version of the variance such that

\[ MSE(\theta_0) \approx b(\theta_0)b(\theta_0)^T + J^{-1}(\theta_0). \]  

(7.28)

For a more complicated but more precise bound, one should resort to (7.27) and the interesting paper by Eldar (2004) on biased estimators. We will now move on and provide more practical advices based on the findings above.

### 7.5 Practical Considerations for Frequency Selection

So far we have said very little about which frequencies to use in the Whittle likelihood method in order to make a tradeoff between bias and variance. We have only pointed out earlier that we are restricted to the frequencies

\[ \omega_k = \frac{2\pi}{T} k, \quad k \in \mathbb{Z} \]
where \( k = 1, \ldots, N_\omega \) in order to keep the spectral estimates at different frequencies asymptotically uncorrelated. It is intuitive that as many frequencies as possible should be used if the periodogram estimate is unbiased. More information reduces the uncertainty. At high frequencies the periodogram is however often biased due to interpolation or un-modelled dynamics. Including high frequencies will then reduce the quality of the estimates and a bias-variance tradeoff has to be made.

### 7.5.1 Minimizing the Variance

According to the expression in (7.4) the variance of the parameter estimates of the Whittle likelihood method is roughly inversely proportional to the quantity

\[
J(\theta_0) = \sum_{k=1}^{N_\omega} J_k(\theta_0, \Phi) J_k(\theta_0, \Phi)^T
\]

where

\[
J_k(\theta_0) = \frac{\Phi'(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)}.
\]

For a second-order continuous-time autoregressive (CAR) model such as

\[
y(t) = H(s)e(t) \quad (7.29)
\]

\[
H(s) = \frac{\sigma}{s^2 + a_1 s + a_2} \quad (7.30)
\]

with the power spectral density

\[
\Phi(i\omega_k, \theta) = \frac{\sigma^2}{(a_2 - \omega_k^2)^2 + a_1^2 \omega_k^2} \quad (7.31)
\]

and the parameter vector \( \theta = (a_1, a_2, \sigma)^T \), the relative sensitivity will take on the form of

\[
J_k(\theta) = \begin{pmatrix}
\frac{\partial}{\partial a_1} \Phi(i\omega_k, \theta) \\
\frac{\partial}{\partial a_2} \Phi(i\omega_k, \theta) \\
\frac{\partial}{\partial \sigma} \Phi(i\omega_k, \theta)
\end{pmatrix} = \begin{pmatrix}
-2a_1 \omega_k^2 \\
-2(a_2 - \omega_k^2)^2 + a_1^2 \omega_k^2 \\
-2(2a_2 - \omega_k^2) \frac{\sigma}{2}
\end{pmatrix}.
\]

The three elements of this vector are shown in Figure 7.1, where we have set \( a_1 = 2 \), \( a_2 = 1 \) and \( \sigma = 1 \). From the figure we see that for the first two elements, corresponding to \( a_1 \) and \( a_2 \), the relative sensitivity is only significantly different from zero on a limited frequency interval. For the standard deviation \( \sigma \) however the same quantity is a constant value for all frequencies.

The expression in (7.26) basically states that the parameter variances are inversely proportional to the sum of the squares of the relative sensitivities for each parameter. Hence, in order to decrease the variance of \( a_1 \) and \( a_2 \) one should use as many frequencies
as possible in the interval where the parameter sensitivities differ from zero. For \( \sigma \) we could use as many and as high frequencies as possible.

In Figure 7.2 we have illustrated the effect of frequency choice on the parameter variance. The parameters of the model (7.31), where \( a_1 = 2, a_2 = 1 \) and \( \sigma = 1 \), have been estimated using different frequency intervals. The time of observation was set to \( T = 200 \text{ s} \) and the sampling interval was set to \( T_s = 0.1 \text{ s} \). First, we used the frequency span \( \{2\pi \frac{1}{200}, \ldots, 2\pi \frac{200}{200}\} \) and indicated those with dots. In the second case we used the frequencies \( \{2\pi \frac{8}{200}, \ldots, 2\pi \frac{208}{200}\} \), indicated by plus signs. This means that we used the same number of frequencies, but a slightly different set.

As can be seen from the picture this doubled the variance of the \( a_2 \) parameter estimate. The standard deviation for \( a_2 \) was 0.0201 for the first case which was indicated by dots. In the second case which was indicated by plus signs the variance was 0.0491. The reason for this can be found in Figure 7.1 where it is shown that the parameter \( a_2 \) is quite sensitive at the lowest frequencies.
Figure 7.2: Illustration of the influence of frequency choice on the variance of parameter estimates. The model is \( H(s) = \frac{\sigma}{s^2 + a_1 s + a_2} \), where \( a_1 = 2 \) and \( a_2 = 1 \). Parameters denoted by plus signs are identified using slightly different frequencies than those indicated by dots. This causes a standard deviation in \( a_2 \) that is twice as large for the plus sign estimates. Here the number of estimates is \( N_{MC} = 400 \), the time of observation is \( T = 200 \) s and the sample interval is \( T_s = 0.1 \) s.

### 7.5.2 Minimizing the Bias

By now, it is known from Theorem 7.1 that the contribution to the bias from each individual frequency component of the power spectrum is roughly explained by the relationship

\[
b(\theta_0) = E(\hat{\theta} - \theta_0) \approx \sum_{k=1}^{N_\omega} S(i\omega_k) \Delta \Phi(i\omega_k)
\]  

(7.32)

where (7.9)

\[
\Delta \Phi(i\omega_k) = \frac{\hat{\Phi}_m(i\omega_k) - \Phi(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)} = \frac{\Phi_e(i\omega_k)}{\Phi(i\omega_k, \theta_0)},
\]

(7.33)

and (7.10)

\[
S(i\omega_k) = J(\theta_0)^{-1} J_k(\theta_0), \quad k = 1 \ldots N_\omega
\]

(7.34)
is determined by the relative sensitivity. In Figure 7.3 the relative bias $\Delta \Phi(\omega_k)$ is plot-

ted as a function of the frequency $\omega$. The system is the one in (7.31) and the output is
uniformly sampled with sampling interval $T_s = 0.5 \, s$. The solid line indicates the relative
spectral bias when the discrete-time spectrum is used instead of the continuous-time spectrum.
The dotted line represents the same bias for piecewise constant interpolation while the dash-dotted illustrates it for piecewise linear interpolation.

From this picture one can see that for large frequencies this relative bias can be quite
large, and because the relative sensitivity of the standard deviation $\sigma$ is constant even for
large frequencies, the impact of the spectral bias there can be significant.

While there are negative effects in terms of increased parameter variance due to the
exclusion of low frequencies, the use of high frequencies can be equally detrimental in
terms of bias. This is something that is illustrated in Figure 7.4, where we have used
frequencies up to $\omega_{N_s}$ in the estimation procedure. On the vertical axis one will find the

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.3}
\caption{Relative spectral bias for spectral estimates of the model $H(s) = \frac{\sigma}{s^2 + a_1 s + a_2}$, where $a_1 = 2$, $a_2 = 1$ and $\sigma = 1$. The sampling interval is set to $T_s = 0.5 \, s$. Here the solid line represents the case when the discrete-time spectrum is used to estimate the continuous-time spectrum. The case when the output is interpolated by a piecewise constant function is represented by the dotted line. Piecewise linear interpolation is shown as dashed-dotted.}
\end{figure}
Figure 7.4: RMSE for parameter estimates for a second order CAR model with respect to maximum frequency. Here $N_{MC} = 200$, $T = 200$, $T_s = 0.1$.

root mean square error of the parameter estimates

$$RMSE = \sqrt{\frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} (\hat{\theta} - \theta_0)^2}. \quad (7.35)$$

The reason for the large RMSE at high frequencies is that there is a very large relative bias at these frequencies. This fact suggests that there is a tradeoff to be made, at least in theory. In order to reduce the variance we should include higher and higher frequencies. On the other hand there will be bias at higher frequencies due to interpolation and model errors. Therefore, as illustrated in Figure 7.4, there will be a point where including higher frequencies will actually increase the RMSE of the parameter estimates. This observation has also been done in the work by Yuz et al. (2005).

### 7.6 Summary

In this chapter the focus has been on bias and variance issues related to the parameter estimates. First, variance and bias were proved to rely on the relative sensitivity of the
parameters and relative bias of the spectral estimate. Then, practical considerations concerning which frequencies to use in the estimation procedure were discussed. Finally, the discussions in the previous sections were illustrated by two examples. The first one showing that some frequencies are more valuable when it comes to reducing the variance of certain parameter estimates. The other example illustrated the bias-variance tradeoff which is connected to the selection of frequencies.

The conclusion of this chapter is that the relative bias in the spectrum that cause the absolute bias in the parameter estimates. At high frequencies, the value of the spectrum is often very small and this could explain why the high frequency bias could cause a large parameter bias. In this chapter it was also found, that it is the relative sensitivity of the spectrum to parameter changes that determines the degree of information content of a particular frequency. Therefore, frequencies with a high relative sensitivity are especially important to use when one seeks accurate parameter estimates.

### 7.7 Appendix A

In both the expressions for the bias and the asymptotic covariance, the Hessian with respect to the parameters is needed. We will show what these Hessians are in the following lemmas.

**Lemma 7.1**

Let $L$ be defined as in (7.7). Then

$$L''_{\theta\theta}(\hat{\theta}, \hat{\Phi}) = \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(i\omega_k, \hat{\Phi}) \hat{\Phi}(i\omega_k) \Phi'_\theta(i\omega_k, \hat{\Phi})^T}{\Phi(i\omega_k, \hat{\Phi})^3} + \sum_{k=1}^{N_\omega} \Phi''_{\theta\theta}(i\omega_k, \hat{\Phi}) \left( \frac{\hat{\Phi}(i\omega_k)}{\Phi(i\omega_k, \hat{\Phi})} \right)^2 - \left( \frac{\hat{\Phi}(i\omega_k)}{\Phi(i\omega_k, \hat{\Phi})} \right)^2$$

**Proof:** From the definition in (7.6) we have

$$L'_\theta(\theta, \hat{\Phi}) = \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(i\omega_k, \theta) \hat{\Phi}(i\omega_k) \Phi'_\theta(i\omega_k, \theta)^T}{\Phi(i\omega_k, \theta)^2} \left( 1 - \frac{\hat{\Phi}(i\omega_k)}{\Phi(i\omega_k, \theta)} \right).$$

The result follows from another differentiation with respect to $\theta$. \hfill \Box

**Corollary 7.1**

Let $L$ be defined as in (7.7). Then

$$L''_{\theta\theta}(\theta_0, \Phi(\theta_0)) = \sum_{k=1}^{N_\omega} \Phi'_\theta(i\omega_k, \theta_0) \left( \frac{\Phi'_\theta(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)} \right)^T.$$
Proof: From Lemma 7.1 we have
\[
L''_{\theta}(\theta_0, \Phi) = \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(i\omega_k, \theta_0)\Phi(i\omega_k, \theta_0)\Phi'_\theta(i\omega_k, \theta_0)^T}{\Phi(i\omega_k, \theta_0)^3} + \sum_{k=1}^{N_\omega} \frac{\Phi''_{\theta\theta}(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)^2} \left( 1 - \frac{\Phi(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)} \right) - \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(i\omega_k, \theta_0)\Phi'_\theta(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)^2} \left( 1 - \frac{\Phi(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)} \right). \tag{7.36}
\]

Hence the result follows. \qed

The derivatives of the negative log likelihood \( L \) with respect to the parameters and the information from the estimated spectrum are also needed. This is summarized in the following lemma

Lemma 7.2
Let \( \hat{\Phi}_k = \hat{\Phi}(i\omega_k) \) and \( L \) be defined as in (7.7), then
\[
L''_{\theta}(\theta_0, \hat{\Phi}) = -\frac{\Phi'_\theta(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)^2}. \tag{7.37}
\]

Proof: As in the previous lemma we have
\[
L'_\theta(\theta, \hat{\Phi}) = \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(i\omega_k, \theta)}{\Phi(i\omega_k, \theta)} \left( 1 - \frac{\hat{\Phi}(i\omega_k)}{\Phi(i\omega_k, \theta)} \right)
\]
and hence
\[
L''_{\theta}(\theta, \hat{\Phi}) = -\frac{\Phi'_\theta(i\omega_k, \theta)}{\Phi(i\omega_k, \theta)^2}. \tag{7.38}
\]

\qed
Application to Estimation of Tire Pressure

“The one thing that unites all human beings, regardless of age, gender, religion, economic status or ethnic background, is that, deep down inside, we ALL believe that we are above average drivers.”

Dave Barry, “Things That It Took Me 50 Years to Learn”

8.1 Introduction

That an accurate tire pressure is important is probably known to almost everyone with a driver’s licence. The wrong pressure can make the vehicle one drives very difficult to manoeuvre and could induce an unnecessary amount of tire wear. It can even cause deadly accidents if the tire breaks down due to low pressure, when the vehicle travels at high speed.

A way to monitor the tire would be to attach an electro-mechanical device to it, which would simultaneously measure the pressure and display the value to the driver. This is however a costly solution since it includes additional hardware. Increased cost due to extra hardware is not an attractive option in these days of overcapacity in the automobile industry. Therefore, innovators are left with extracting as much data as possible from existing measurements. One important source of information is the wheel speed measurements used in connection with the Anti-lock Braking System (ABS). As pointed out in Chapter 1 a possible way to extract more information out of existing data is by using a model. It also requires an understanding of the physics of the tire, i.e. a continuous-time physically parameterized model.

In Chapter 6 a theoretical framework for system identification of continuous time series models was constructed. The methods developed there were then analyzed in Chapter 7 with respect to the quality of the estimates. The focus in this chapter, will therefore be on applications of these ideas to the estimation of tire pressure or axel vibrations.
8.2 Outline

First, in Section 8.3, a simple continuous-time vibrational model of a pneumatic tire is introduced. It is also explained how information about the tire pressure can be derived from that model. Then, in Section 8.4, a number of objectives for a high-sensitivity vibration analysis procedure are introduced. In Section 8.5 an existing time domain method is compared to a frequency domain approach. Then in Section 8.6 a method for combining several spectral estimates into one is introduced together with a discussion in Section 8.7 related to the material on bias and variance in Chapter 7. Finally, the methods are used to estimate the vibrations in real pneumatic tire from ABS system signals in Section 8.8.

8.3 Tire Pressure Modelling

In this chapter the tire is modelled as a spring-damper system in the torsional direction (Wong, 1993). Road irregularities which can be modelled as a force with a white normal noise characteristics, excite vibrations in the tire which affect the wheel speed directly. Sensors measuring the velocity of the wheel can then pick up these vibrations which are usually in the area of 0-100 Hz for a normal car.

Certain vibrations in the range of 30 - 60 Hz can be modelled as a mass-spring-damper system which motivates the following relationship between the torsional oscillations \( y(t) \) and the effect from road irregularities \( e(t) \)

\[
\ddot{y}(t) + \frac{b}{m} \dot{y}(t) + \frac{k}{m} y(t) = e(t).
\]

Here \( m \) is the mass of a piece of the tire rubber rim and \( k \) is a spring constant connected to the elasticity of the material. The model can also be rewritten as

\[
\ddot{y}(t) + 2\gamma \dot{y}(t) + \omega_0^2(t) = e(t)
\]

where

\[
\gamma = \frac{b}{2m}, \quad \omega_0 = \sqrt{\frac{k}{m}}.
\]

and therefore the oscillations can be modelled as

\[
y(t) = H(s)e(t) = \frac{1}{s^2 + 2\gamma s + \omega_0^2} e(t).
\] (8.1)

The tire pressure will affect the spring constant and the damping coefficient and hence also the resonance peak of the system. Tracking changes in this peak will then make it possible to track the tire pressure.

The continuous-time spectrum of the oscillatory process will be

\[
\Phi(i\omega, \theta) = \frac{\sigma^2}{|H(i\omega)|^2} = \frac{\sigma^2}{(\omega^2 - \omega_0^2)^2 + 4\gamma^2\omega^2}
\] (8.2)
where \( \sigma \) is the standard deviation of the driving white noise. This means that the location of the resonance peak is

\[
\omega_{res} = \sqrt{\omega_0^2 - 2\gamma^2}. \tag{8.3}
\]

The original model in (8.1) is given in continuous-time, and the objective will be to estimate the continuous-time parameters \( \gamma \) and \( \omega_0 \). Then, an estimate of the resonance frequency \( \omega_{res} \) can be acquired through (8.3) (Persson, 2002).

8.4 Problem Specifics and Objectives

A characteristic problem with signal processing using tachometer measurements on rotating axles is that the measurements are uniform in the angle domain but non-uniform (speed dependent) in the time domain. This comes from the fact that most common sensors for such applications measure the time between certain angle displacements, which are speed dependent. One can for instance illustrate this with the ABS sensors in a car, which gives between 50 and 100 pulses per revolution of each wheel. If tire vibration analysis and other similar problems are to be approached in the time sampled domain, either one has to rely on data interpolation to uniform time sampling, or derive dedicated algorithms. Motivated by the developments in previous chapters a frequency domain approach to the problem is presented and compared theoretically to a time domain algorithm proposed in (Persson, 2002; Persson and Gustafsson, 2001).

We believe that the main specifications on a procedure aimed at high-sensitivity tire vibration analysis are the following:

1. Being based on parametric physical models of the vibrations such as the one described in (8.1).
2. Operate on short data batches in a pre-specified speed interval where the data pass several quality checks.
3. Have the potential to efficiently reject wide band disturbances that are non-interfering with the important vibrations.
4. Have the potential to reject narrow band disturbances that are interfering with the vibrations.

The time domain method given in (Persson, 2002; Persson and Gustafsson, 2001) successfully solves the first three specifications, but not the last one. The method proposed here, on the other hand, has the potential to be modified to eliminate outliers in the frequency domain. This general disturbance problem occurs in several other applications as in an automotive drive-line where the vibrations are due to engine knock or resonant modes. In robotics for instance, the vibrations may originate from the load. This chapter will however focus on Point 1, 2 and 3 in the list above. In the subsection below, there will only be a short discussion of time and frequency-domain methods with respect to Point 4. Issues connected to narrow band disturbances are left to Chapter 9.
8.5 Time and Frequency Domain Approaches

Table 8.1 summarizes the notation and signal models that are used in the time and frequency domain. Basically, the tire vibration analysis is approached by a continuous-time autoregressive model motivated by the spring-damper model of the axel in (8.1). Superimposed on this signal are other vibrations and external disturbances \( d(t) \), and the speed signal itself. Measurements are taken each time \( t_k \) a pulse is received from the ABS signal. These pulses represent a certain fixed angle displacement, which explains the special appearance of \( y[k] = y(t_k) \) in Table 8.1.

<table>
<thead>
<tr>
<th>Time domain</th>
<th>Frequency domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y(t) = \frac{1}{A(p; \theta)} e(t) + d(t) )</td>
<td>( \Phi_H(i\omega) = \frac{1}{</td>
</tr>
<tr>
<td>( y(t_k) = \frac{2\pi}{L} \int_{t_{k-1}}^{t_k} d(t) dt )</td>
<td>( \Phi_y(i\omega) = \Phi_H(i\omega) \Phi_e(i\omega) + \Phi_d(i\omega) )</td>
</tr>
</tbody>
</table>

\( e(t) \) white noise,
\( A(p; \theta) \) AR model,
\( \theta \) parameters in the AR model,
\( d(t) \) disturbance,
\( y[k] \) measured non-uniform samples of angle,
\( L \) number of cogs per revolution,
Angle uniform sampling, not time uniform sampling.
\( \Phi_e(i\omega) = \sigma \) white noise spectrum,
\( A(i\omega; \theta) \) AR model,
\( \theta \) parameters in the AR model,
\( \Phi_d(i\omega) \) disturbance spectrum,
\( \Phi_y(i\omega) \) 'measured' spectrum.

**Table 8.1**: Signal models and assumption in time and frequency domains, respectively.

The time domain algorithm proposed in (Persson, 2002) estimates the parameters of the autoregressive model in the following steps:

1. Interpolate data to a high sampling rate to avoid aliasing,
2. Band-pass filter the signal to get rid of broad band disturbances and to focus on the region 30 - 60 Hz,
3. Down-convert the signal utilizing deliberate aliasing,
4. Estimate a discrete-time AR model and
5. Extract vibration data from this model.

It is not easy to modify this algorithm to remove the narrow-band interference \( d(t) \), so the most practical solution is to turn off the algorithm when such a disturbance is detected.
Algorithm 1 on the other hand, outlines the frequency domain approach which will be described in the section below. Since the method operates in the frequency domain, narrow band disturbances $\Phi_d(i\omega)$ as those illustrated in Table 8.1, can be seen as statistical outliers in the frequency domain. Outliers in the time-domain have traditionally been dealt with by introducing more robust norms in the estimation criteria e.g. (Huber, 1981). This approach will be transferrable to the case of frequency domain data. Here, on the other hand, the focus will be on rejection of wide band disturbances, short data batches and physical modeling. Managing these issues is a necessary condition for the overall usefulness of the method in the context of tire vibration analysis.

### 8.6 Frequency Domain Estimation

The approach to frequency domain parameter estimation in this section is similar to the framework described in Chapter 6. Since the entire continuous-time realization of the output is unavailable and only a finite number of samples at time instances $\{t_k\}_{k=1}^{N_t}$ are available, piecewise constant interpolation as in (6.9)

$$
\hat{y}(t) = \sum_{i=1}^{N} y(t_i) \phi_i(t - t_i)
$$

will be used to reconstruct the realization. In order to reduce the variance of the periodogram data, the output data set is split into $N_b$ batches of duration $\{T_k\}_{k=1}^{N_b}$. Then a periodogram

$$
\hat{\Phi}_n(y)(i\omega) = |\hat{Y}_{T_n}(i\omega)|^2 \quad n = 1 \ldots N_b
$$

is calculated for each batch. The Fourier transform for piecewise constant interpolation is computed as in (6.16) such that

$$
\hat{Y}_{T_n}(i\omega) = \frac{1}{\sqrt{T_n}} \sum_{k=1}^{N_n} y(t_k) e^{-i\omega t_k} - e^{-i\omega t_k} \quad n = 1 \ldots N_b
$$

where $\{N_n\}_{n=1}^{N_b}$ is the number of samples in each batch. A periodogram with reduced variance is then formed as

$$
\hat{\Phi}_y(i\omega) = \frac{1}{N_b} \sum_{n=1}^{N_b} \hat{\Phi}_n(y)(i\omega).
$$

This method is analogous to the method by Welch (Welch, 1967) for the smoothing of spectral estimates.

When the periodogram estimate $\hat{\Phi}$ of the wheel speed $y(t)$ is available a continuous-time autoregressive model can be identified by using the continuous-time Whittle likelihood method described in Chapter 2

$$
\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N_\omega} \frac{\hat{\Phi}_y(i\omega_k)}{\Phi(i\omega_k, \theta)} + \log \Phi(i\omega_k, \theta).
$$
Algorithm 1 Tire Pressure Estimation Method

1. Divide wheel speed information \( \{y(t_k)\}_{k=1}^{N_t} \) into \( N_b \) batches of data of length \( T_b \).

2. Approximate the continuous time Fourier transform of each individual batch using interpolation

\[
\hat{Y}_n(i\omega_k) = \frac{1}{\sqrt{N_b}} \int_{(n-1)T_b}^{nT_b} \hat{y}(t)e^{i\omega_k t} dt \quad n = 1 \ldots N_b, \quad k = 1 \ldots N_\omega \\
\hat{y}(t) = \sum_{k=1}^{N_t} y(t_k)\phi_k(t-t_k) \quad n = 1 \ldots N_b, \quad k = 1 \ldots N_\omega.
\]

3. Form the continuous-time periodogram \( \hat{\Phi}_y^n(i\omega_k) = \left|\hat{Y}_n(i\omega_k)\right|^2 \) for each individual batch and frequency.

4. Compute the average periodogram

\[
\hat{\Phi}_y(i\omega_k) = \frac{1}{N_b} \sum_{n=1}^{N_b} \hat{\Phi}_y^n(i\omega_k) \quad k = 1 \ldots N_\omega. \quad (8.4)
\]

5. Use the Whittle likelihood method to estimate the parameters

\[
\hat{\theta} = \arg \min_{\theta} \sum_{k=1}^{N_\omega} \frac{\hat{\Phi}_y(i\omega_k)}{\Phi(i\omega_k, \theta)} + \log \Phi(i\omega_k, \theta)
\]

of the vibrational model

\[
\Phi(i\omega_k, \theta) = \frac{\sigma^2}{\left[(i\omega_k)^2 + 2i\omega_k + \omega_0^2\right]^2}.
\]

6. Compute the location of the resonance peak

\[
\omega_{res} = \sqrt{\omega_0^2 - 2\gamma^2},
\]

which is related to the tire pressure.
8.7 Properties of Bias and Variance

In Chapter 7 statistical properties of estimators such as the one in (8.5) were discussed. In this section we will use the results from that chapter, but in a slightly different manner.

In the case of tire vibration analysis there can be wide band disturbances corrupting the measurements of the output $y$. From (7.8) we know that

$$E(\hat{\theta} - \theta_0) \approx \sum_{k=1}^{N_\omega} S(i\omega_k) \Delta \Phi_y(i\omega_k).$$

where $\hat{\theta}$ are the estimated and $\theta_0$ are the true parameter values. The relative bias in the periodogram estimate of the power spectrum is defined as

$$\Delta \Phi(i\omega_k) = \frac{E\hat{\Phi}_y(i\omega_k) - \Phi(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)}.$$

The sensitivity of the parameter estimates to the relative bias in the periodogram is

$$S(i\omega_k) = J(\theta_0, \Phi)^{-1} J_k(\theta_0, \Phi)$$

where the relative sensitivity is

$$J_k(\theta_0, \Phi) = \frac{\Phi'(i\omega_k, \theta_0)}{\Phi(i\omega_k, \theta_0)}$$

and

$$J(\theta^*, \Phi) = \sum_{k=1}^{N_\omega} J_k(\theta_0)J_k(\theta_0)^T$$

In the case of tire vibration analysis there can be wide band disturbances corrupting the spectrum of the output $y$. Those disturbances are in some cases not interfering with the important vibrations. Therefore, the spectrum $\Phi$ of the model is very small in the area of the disturbances and the relative bias in the spectrum can be quite significant. Hence, in order to avoid bias it is necessary to ignore information from frequencies where the relative bias is large.

In an online automotive application computational power and available memory will always be an important design constraint. Calculating the periodogram for a particular frequency can be cumbersome and therefore it is important to know which frequencies carry the most information. In the case of variance it was established in (7.26) in Chapter 7 that

$$E(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^T \approx J(\theta_0)^{-1}$$

and again the relative sensitivity plays an important role. Therefore, in order to reduce the variance, information from frequencies where the relative sensitivity is large should be prioritized.
8.8 Experimental Results

In this section the theory presented above is applied to the estimation of the resonance peak of the torsional vibrations of a pneumatic tire. The samples $y(t_k)$ are pre-processed measurements from an axle-angle measurement device in the ABS system of a car. The frequency spectrum of $y(t)$ is approximately divided as summarized in Table 8.2. The vibrations in the range 30-60 Hertz are modelled as a spring-damper system excited by white noise $e(t)$ as in (8.1).

<table>
<thead>
<tr>
<th>Speed</th>
<th>Mode 1</th>
<th>Noise</th>
<th>Mode 2</th>
<th>Noise</th>
<th>Mode 3</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10</td>
<td>Mode 1</td>
<td>Noise</td>
<td>Mode 2</td>
<td>Noise</td>
<td>Mode 3</td>
<td>Noise</td>
</tr>
</tbody>
</table>

Narrow-band noise components

For the special parametrization of the spectrum in (8.2) the relative sensitivity functions with respect to the parameters are shown in Figure 8.2. Here we have chosen $\gamma = 33.88$ and $\omega_0 = 289.687$. This means that $w_{res} = 285 \text{ rad/s}$ or $f_{res} = 45.47\text{ Hz}$. From this figure, we conclude that the parameter $\gamma$ in (8.1) is sensitive near the natural resonance frequency of the system. The frequency $\omega_0$ in (8.1) on the other hand, is par-
particularly sensitive at low frequencies. According to Table 8.2 there is noise between 15 and 30 Hertz and 60 and 80 Hertz. Therefore we restrict the frequencies used to those between 30 and 60 Hertz in order to avoid bias in the parameters.

In Figure 8.3 we have estimated the resonance frequencies from the refined set of real data from an ABS sensor. The data has been divided into four parts. These parts have then been subdivided into a set of batches with a duration of a certain number of revolutions or laps of the tire. The number of laps per sub-batch is indicated on the horizontal axis of the figure. Periodograms have been estimated using piecewise constant interpolation for each batch and subsequently averaged in order to yield four estimates of the spectrum. Parameters have then been estimated using the Whittle likelihood approach and the mean value of the estimated resonance frequency \( f_{\text{res}} \) is plotted. The figure indicates that the method is feasible and that a sub-batch size of about 10 laps is sufficient to yield a stable estimate of the resonance frequency. Below 10 laps per sub-batch the mean value of parameter estimates decrease rapidly due to bias from the small number of frequency domain samples.

\[ \text{Figure 8.2: Relative sensitivities } J_k \text{ for } \gamma \text{ (upper) and } \omega_0 \text{ (lower) in (8.2).} \]
Figure 8.3: Resonance frequency $f_{\text{res}}$ in Hertz versus batch size in number of tire laps. Bars indicate one standard deviation.

8.9 Summary

In this chapter, the ideas in Chapter 6 and 7 were applied to a set of real data emanating from the ABS system of a car. The purpose was to estimate the tire pressure by monitoring the position of a particular resonance peak found in the wheel speed information. Since there were disturbances outside the frequency interval of interest, these frequencies were not used in estimation in order to avoid severe parameter bias. The model was of the mass-spring-damper type, and in order to reduce the variance in the parameter estimates, the data set was divided into a number of sub-batches. The spectrum was estimated for each individual batch and all the estimates were then averaged. The average spectral estimate was then used to find the resonance peak of the model.
9 Robust Frequency Domain ARMA Modelling

“Orr would be crazy to fly more missions and sane if he didn’t, but if he was sane he had to fly them. If he flew them he was crazy and didn’t have to; but if he didn’t want to he was sane and had to.”

Joseph Heller (1923 - 1999), Catch 22

9.1 Introduction

In Chapter 7 the problem of dealing with wide-band disturbances was discussed in some depth. There, it was showed that their influence in terms of bias could be reduced by considering a limited frequency interval. In this chapter we will focus on narrow-band disturbances. This is a problem of frequency domain outliers, which can occur in several applications. For instance in automotive applications, see Figure 9.1, where the vibrations can be caused by resonances in the drive-line. The oscillations are characteristically localized in the frequency domain and are distributed over relatively large time intervals. Therefore, a frequency domain approach is suitable.

The classical way of reducing the effect of outliers in statistics, signal processing and identification in the time-domain is to introduce a robust norm into the criterion function of the estimation method (Huber, 1981; Poljak and Tsypkin, 1980; Akcay et al., 1996). These ideas carry over to the frequency-domain with a few modification.

Unfortunately the old saying that “nothing is for free” applies to the robustification of estimators. Making a method more robust to outliers means that it will, so to speak, be less efficient at extracting information from the data. It will trust the data to a lesser degree. Therefore the variance of the parameter estimates may increase and the user of the methods presented in this chapter will have to make a tradeoff between bias and variance.
Figure 9.1: Smoothed spectrum for vibrations extracted from ABS system signals. Notice the powerful outliers interfering with the spectrum.

9.2 Outline

The chapter will be arranged as follows. First, in Section 9.3, there will be a short introduction to the infinitesimal approach to robust statistics. Concepts such as the influence function, M-estimators, gross error sensitivity and optimally robust estimators are introduced. Then, in Section 9.4 this material will be extended to the frequency domain identification of ARMA models. It is shown that the Whittle likelihood method, that has been used throughout this thesis, can be robustified in order to manage outliers in the frequency domain. The rest of the chapter is dedicated to bounding the asymptotic bias and deriving the variance of the estimator. Also, guaranteeing consistency under normal conditions is an important issue that is dealt with.

9.3 Infinitesimal Approach to Robust Estimation

The statistics framework which the material of this chapter will rely on is mainly the so called infinitesimal approach (Hampel et al., 1986). Therefore, the current section will provide an introduction to the elements of this theory. We will therefore begin with ex-
plaining the concept of an estimator in this particular context together with the important influence function.

### 9.3.1 Estimators

As a start we will introduce the estimator, which is defined as a functional $T$

$$
\hat{\theta} = T(F_\theta)
$$

(9.1)

that takes a probability distribution $F_\theta$ of the data, parameterized in terms of $\theta$, and provides an estimate of $\theta$. If the distribution is the so called empirical distribution derived from observations $\{x_k\}_{k=1}^N$, with the density function

$$
f_N(x) = \frac{1}{N} \sum_{k=1}^N \delta(x - x_k).
$$

(9.2)

the functional can be written in the more recognizable data dependent form

$$
\hat{\theta}_N = T_N(x_1, \ldots, x_N) = T(F_N).
$$

(9.3)

A well known and common estimator is the maximum likelihood (ML) method in (2.83). Here, the estimated parameters are defined explicitly as

$$
\hat{\theta}^{ML}_N = T_{ML}(x_1, \ldots, x_N) = \arg \min_{\theta} \sum_{k=1}^N - \log f(x_k|\theta).
$$

(9.4)

where $f(x_k|\theta)$ is the likelihood function in (2.82) for the measurements $\{x_k\}_{k=1}^N$, if these are assumed independent and identically distributed. The estimator can also be implicitly defined as

$$
\sum_{k=1}^N s(x_k, \hat{\theta}^{ML}_N) = 0
$$

(9.5)

where

$$
s(x, \theta) = \frac{\partial}{\partial \theta} \log f(x, \theta)
$$

(9.6)

is known as the maximum likelihood score function.

Assume, for instance, that the data $\{x_k\}_{k=1}^N$ originates from an exponential distribution, such that

$$
f(x_k, \theta) = \frac{1}{\theta} e^{-\frac{x_k}{\theta}}.
$$

(9.7)

Then, the ML estimate of the parameter $\theta$ would be defined by the expression

$$
\hat{\theta}_N = T_{ML}(x_1, \ldots, x_N) = \arg \min_{\theta} \sum_{k=1}^N \frac{x_k}{\theta} + \log \theta.
$$

(9.8)
or as the solution $\hat{\theta}^M_N$ to the equation

$$\sum_{k=1}^{N} x_k - \hat{\theta}^M_N = 0. \quad (9.9)$$

From now on, we will focus on estimates $\hat{\theta}$ that are defined as solutions to equations such as

$$\sum_{k=1}^{N} \psi(x_k, \hat{\theta}^M_N) = 0 \quad (9.10)$$

and generalizations of these. The methods in (9.10) are known as $M$-estimators because of their close relationship with the Maximum Likelihood (ML) estimators.

### 9.3.2 The Influence Function

In the classical theory of statistical inference, one assumes that the observations which the estimates are based on, originate from a particular model distribution $F_\theta$ where $\theta$ belongs to some parameter set. In robustness theory however, it is acknowledged that the model $F_\theta$ is only a mathematical abstraction of reality. This means that when we study the distributions of the estimators, we should also consider deviations from $F_\theta$. For this purpose we need a tool to study how deviations in the distribution of the data can affect the distribution of the estimates. A central tool in this effort, is the so-called influence function (Hampel et al., 1986)

$$IF(x; T, F) = \lim_{t \to 0} \frac{T((1-t)F + t\Delta_x) - T(F)}{t} \quad (9.11)$$

where the density function of $\Delta_x$ is

$$\delta_x(z) = \delta(x - z). \quad (9.12)$$

This means that if some distribution $G$ is near $F$, the influence function can be thought of as the first order term of a Taylor (actually von Mises) expansion of the estimator $T$ around the distribution $F$ such that

$$T(G) = T(F) + \int IF(x; T, F)d(G - F)(x) + \ldots. \quad (9.13)$$

If $G$ is replaced by the empirical distribution $F_N$ and the integral in (9.13) is evaluated, then under certain conditions

$$\sqrt{N}(T(F_N) - T(F)) \approx \frac{1}{\sqrt{N}} \sum_{k=1}^{N} IF(x_k; T, F) + \cdots \to N(0, V(T, F)) \quad (9.14)$$

as $N \to \infty$ where the asymptotic variance $V(T, F)$ will be

$$V(T, F) = \int IF(x; T, F)^2 dF(x). \quad (9.15)$$
Another way to define the estimator $T$ is as the functional given by

$$\int \psi(x, T(G))dG(x) = 0 \tag{9.16}$$

for all distributions $G$. Using this approach it is possible to compute the influence function by replacing the distribution $G$ by the convex combination $F_{t,x} = (1-t)F + t\Delta x$ and differentiate with respect to $t$. This yields

$$0 = \int \psi(y, T(F))d(\Delta x - F)(y) + \int \frac{\partial}{\partial \theta} \left[ \psi(y, \theta) \right]_{\theta = T(F)} dF(y) \frac{\partial}{\partial t} [T(F_{t,x})]_{t=0} \tag{9.17}$$

which means that

$$IF(x; \psi; F) = \frac{\partial}{\partial t} [T(F_{t,x})]_{t=0} = -\frac{\int \psi(y, T(F))d(\Delta x - F)}{\int \frac{\partial}{\partial \theta} \left[ \psi(y, \theta) \right]_{\theta = T(F)} dF(y)} \tag{9.18}$$

For maximum likelihood methods, expressions for the influence function and the asymptotic variance can be computed explicitly as

$$IF(x; \psi, F) = \frac{\psi(x, \theta)}{\int \psi(y, \theta)s(y, \theta)dF(y)} \tag{9.20}$$

$$VF(T, \theta) = \frac{\int \psi(y, \theta)^2dF(y)}{\left(\int \psi(y, \theta)s(y, \theta)dF(y)\right)^2} \tag{9.21}$$

where $s$ is the maximum likelihood score function defined in (9.6) and

$$\psi(x, \theta) = \log f(x|\theta) \tag{9.22}$$

In the case of the maximum likelihood estimation of the parameter $\theta$ in the exponential distribution where

$$\psi(x, \theta) = s(x, \theta) = \frac{\partial}{\partial \theta} \log f(x_i, \theta) = \frac{x - \theta}{\theta^2}, \tag{9.23}$$

we will have

$$IF(x; \psi, F) = \frac{\psi(x, \theta)}{\int \psi(y, \theta)s(y, \theta)dF(y)} = \frac{x - \theta}{\int \frac{y - \theta}{\theta^2}dF(y)} = x - \theta \tag{9.24}$$

and

$$VF(T, \theta) = \frac{\int \psi(y, \theta)^2dF(y)}{\left(\int \psi(y, \theta)s(y, \theta)dF(y)\right)^2} = \frac{\int \left(\frac{y - \theta}{\theta^2}\right)^2dF(y)}{\left(\int \frac{y - \theta}{\theta^2}dF(y)\right)^2} = \theta^2. \tag{9.25}$$
9.3.3 M-estimators of Scale

If a family of probability densities with parameter $\theta$ is of the form

$$f_\theta(x) = \frac{1}{\theta}f\left(\frac{x}{\theta}\right)$$  \hspace{1cm} (9.26)

where $f$ is a probability density function, then $\theta$ is called a scale parameter. This also means that the distribution function can be written as

$$F_\theta(x) = F\left(\frac{x}{\theta}\right).$$  \hspace{1cm} (9.27)

In the exponential case, the distribution function has the particular structure

$$F_\theta(x) = 1 - e^{-\frac{x}{\theta}} = F\left(\frac{x}{\theta}\right)$$  \hspace{1cm} (9.28)

and estimating $\theta$ is therefore a matter of estimating a scale parameter. An M-estimate of scale $T$ is then defined by the relation

$$\int \psi\left(\frac{y}{T(G)}\right) dG(y) = 0$$  \hspace{1cm} (9.29)

for any distribution $G$. Using the expression in (9.18) it is possible to conclude that

$$IF(x; F, T) = \frac{\psi(x)T(F)}{\int T(F) \psi\left(\frac{y}{T(F)}\right) dF(y)}.$$  \hspace{1cm} (9.30)

In the case of maximum likelihood estimation for scale families such as those in (9.26) is an M-estimate with

$$\psi\left(\frac{x}{\theta}\right) = \frac{1 + \frac{x}{\theta} f'\left(\frac{x}{\theta}\right)}{\theta}$$  \hspace{1cm} (9.31)

which in the exponential case will become

$$\psi\left(\frac{x}{\theta}\right) = \frac{1 - \frac{x}{\theta}}{\theta}$$  \hspace{1cm} (9.32)

since $f\left(\frac{x}{\theta}\right) = e^{-\frac{x}{\theta}}$ and $f'\left(\frac{x}{\theta}\right) = -e^{-\frac{x}{\theta}}$. The estimator $\hat{\theta}_N$ of the scale parameter $\theta$ of the exponential distribution will therefore be defined as

$$\sum_{k=1}^{N} \psi\left(\frac{x}{\hat{\theta}_N}\right) = \frac{1}{\hat{\theta}_N} \sum_{k=1}^{N} \left(\frac{x}{\hat{\theta}_N} - 1\right) = 0$$  \hspace{1cm} (9.33)

which is identical to what was derived in (9.9).
9.3.4 The Gross-Error Sensitivity

In robust statistical theory, as pointed out earlier, the objective is to quantify and limit the effect of unexpected properties of the data. One such measure of quality is the gross-error sensitivity

\[ \gamma^*(T, F) = \sup_x |IF(x; T, F)|. \]  

(9.34)

which signifies the worst influence which a small amount of contamination at \( x \) can have on the value of the estimator. It is regarded as an upper bound on the asymptotic bias and having a bound on \( \gamma^* \) is therefore a desirable robustness property of an estimator, which is then called \( B \)-robust (bias robust).

In the case of our previous example, the estimation of the parameter in the exponential distribution, we have from (9.24) that

\[ \gamma^*(T, F) = \sup_x |x - \theta| = \infty \]  

(9.35)

and the effect of outliers can therefore be arbitrarily severe.

9.3.5 Optimally B-Robust Estimators

A class of estimators, which are as efficient as possible subject to an upper bound on the gross-error sensitivity are called optimally \( B \)-robust. Fortunately, these can be found efficiently from an underlying maximum likelihood method by choosing

\[ \psi(x, \theta) = [s(x, \theta) - a(\theta)]_{-b(\theta)}, \]

such that

\[ \int \psi(x) dF(x) = 0 \]  

(9.37)

for each \( b(\theta) \), where \( s \) is the score function in (9.6). This is result of the following theorem.

**Theorem 9.1 (Hampel et al. (1986))**

Let \( b > 0 \) be some constant and define

\[ s(x, \theta^*_+) = \frac{\partial}{\partial \theta} \log f_\theta(x)|_{\theta^*_+}. \]  

(9.38)

Then there exists a real number \( a \) such that

\[ \tilde{\psi}(x) = [s(x, \theta^*_+) - a]_{-b} \]  

(9.39)
satisfies

\[ \int \tilde{\psi}(x) dF_*(x) = 0 \quad (9.40) \]

\[ d = \int \tilde{\psi}(x)s(x, \theta_*) dF_*(x) > 0. \quad (9.41) \]

Now \( \tilde{\psi} \) minimizes

\[ \frac{\int \psi(x)^2 dF_*(x)}{(\int \psi(x)s(x, \theta_*) dF_*(x))^2} \quad (9.42) \]

among all mappings \( \psi \) that satisfy

\[ \int \psi(x) dF_*(x) = 0 \quad (9.43) \]

\[ \int \psi(x)s(x, \theta_*) dF_*(x) \neq 0 \quad (9.44) \]

\[ \sup_x \left| \frac{\psi(x)}{\int \psi(x)s(x, \theta_*) dF_*(x)} \right| \leq c = \frac{b}{d}. \quad (9.45) \]

Any other solution of this extremal problem coincides with a nonzero multiple of \( \tilde{\psi} \).

**Proof:** Hampel et al. (1986)

### 9.3.6 Optimally B-Robust Estimators of Scale

Because \( \psi(x, \theta) = \psi \left( \frac{x}{\theta} \right) \) in the case of estimation of scale, the \( \psi \)-function at \( \theta_0 = 1 \) determines everything. The optimal B-robust scale M-estimator is then given by

\[ \tilde{\psi}(y) = \left[ -y \frac{f'(y)}{f(y)} - 1 - a \right]_b^{-b} . \quad (9.46) \]

In the case of the exponential distribution this will become

\[ \tilde{\psi}(y) = [y - 1 - a]_b^{-b} \quad (9.47) \]

since \( f(x) = e^{-y} \) and \( f'(x) = -e^{-y} \). Therefore, the robust estimate can be computed as the solution to the equation

\[ \sum_{k=1}^{N} \tilde{\psi} \left( \frac{x_k}{\hat{\theta}_N} \right) = \sum_{k=1}^{N} \left[ \frac{x_k}{\hat{\theta}_N} - 1 - a \right]_b^{-b} = 0 \quad (9.48) \]

where \( a \) is computed as indicated in Theorem 9.1. We will now use inspiration from this estimator in order do robust ARMA modelling in the frequency domain.
9.4 Robust Frequency Domain ARMA Estimation

Until now we have merely repeated the basics of robust estimation. From now on we will attempt to generalize and extend these ideas in the context of frequency-domain identification of autoregressive moving average (ARMA) models. That is, when observed data \( t \) is modelled as Gaussian noise passed through a continuous- or discrete-time dynamical system. This can also be interpreted such that the true output would have the spectrum \( \Phi(i\omega, \theta_0) \) or \( \Phi(e^{i\omega T_s}, \theta_0) \), where \( T_s \) is the sampling interval in case of discrete-time modelling.

If we for instance would compute the continuous-time Fourier transform of this true output, its periodogram would be distributed as

\[
\hat{\Phi}(i\omega_k) = |Y_T(i\omega_k)|^2 \sim \text{AsExp} \Phi(i\omega_k, \theta_0)
\]

and the transform would be asymptotically independent at the frequencies \( \{\omega_k\}_{k=1}^{N_\omega} \) where

\[
\omega_k = \frac{2\pi}{T} k
\]

if \( T \) is the time of observation of the output \( y(\cdot) \). We assume that \( N_\omega \) is large so that the asymptotic expression can be considered valid. We will also treat the continuous-time periodogram estimates as our “measurements” and will consider the element of the sequence

\[
\hat{\Phi}(i\omega_k) \ldots \hat{\Phi}(i\omega_{N_\omega})
\]

as mutually uncorrelated. When an estimate of the power spectrum is available, a model can be identified by the following Maximum-Likelihood (ML) procedure described in Chapter 2 or in the papers by Whittle (1961) or Dzhaparidze (1970)

\[
L(\theta, \hat{\Phi}) = \sum_{k=1}^{N_\omega} \frac{\hat{\Phi}(i\omega_k)}{\Phi(i\omega_k, \theta)} + \log \Phi(i\omega_k, \theta)
\]

\[
\hat{\theta} = \arg \min_{\theta} L(\theta, \hat{\Phi}).
\]

where the asymptotic Fisher information at the true parameters \( \theta_0 \) is

\[
J(\theta_0) = \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(\omega_k, \theta_0) \Phi'_\theta(\omega_k, \theta_0)^T}{\Phi(\omega_k, \theta_0)^2}.
\]

This approach forms the basis for the work presented in the remaining part of the chapter. The estimate in (9.52) and (9.53) can also be described as the \( \hat{\theta} \) which is the solution to the vector equation

\[
\sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(\omega_k, \hat{\theta})}{\Phi(\omega_k, \hat{\theta})} \left( \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta) - 1} \right) = 0
\]
where $\Phi'(\omega_k, \theta)$ is the gradient of the spectrum with respect to the parameters.

If the model is the correct one, the maximum-likelihood method is known to be the best possible (optimal) estimator with consistency and asymptotic efficiency (Wald, 1949; Cramér, 1946). If there, on the other hand, are unmodelled outliers present, the method might not be optimal. This leads us to consider robust estimators which give up efficiency at the true model in exchange for reasonable performance if the model is not the true one (Casella and Berger, 2002). Therefore we introduce the following method

$$\Psi(\hat{\Phi}, \theta) = \sum_{k=1}^{N_\omega} \frac{\Phi'(\omega_k, \theta)}{\Phi(\omega_k, \theta)} \psi \left( \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta)} \right) = 0$$

(9.55)

inspired by the M-estimators for scale introduced by Huber (1981). For another interesting application of robust estimation of see the work by Brcich et al. (2004).

The unknown in (9.55) is a function $\psi$ which will be found using the influence function approach introduced by Hampel (Hampel et al., 1986). Certain measures must however be taken in order to assure that the estimates are consistent when the periodogram originates from the assumed model in (9.49). This implies that

$$\int \psi \left( \frac{\hat{\Phi}(i\omega_k)}{\Phi(\omega_k, \theta_0)} \right) dF_k \left( \hat{\Phi}(\omega_k) \right) = \left[ x \right]_{x = \hat{\Phi}(i\omega_k)} = \int \psi (x) dG (x) = 0$$

(9.56)

where we from now on define the distributions

$$F_k \sim \text{Exp} \Phi(\omega_k, \theta_0),$$

$$G \sim \text{Exp} 1.$$

### 9.4.1 Multivariable Influence Function

As in Section 9.3, the influence function will play an important role. If we define the solution to equation (9.55) as a vector valued stochastic variable $\hat{\theta} = T(F)$ with a distribution dependent on the multivariable distribution $F = (F_1, \ldots, F_{N_\omega})$ of the periodogram $\hat{\Phi}(\omega_k)$, we can define the influence function for (9.55) as

$$IF(\hat{\Phi}; T, F) = \lim_{h \to 0} \frac{T((1 - h)F + h\delta_\hat{\Phi}) - T(F)}{h}$$

where $\delta_\hat{\Phi}$ is the multi-variable measure which puts the probability mass 1 at

$$\hat{\Phi} = \left( \hat{\Phi}(i\omega_1), \ldots, \hat{\Phi}(i\omega_{N_\omega}) \right)^T.$$

Hence, in order to find the parameters, we have to solve the equation

$$\Psi \left( \hat{\Phi}, \hat{\theta} \right) = 0$$

(9.57)

which means for any distribution $G$ of the data $\hat{\Phi}$ we have

$$\int \Psi \left( \hat{\Phi}, T(F) \right) dG \left( \hat{\Phi} \right) = 0.$$  

(9.58)
We can therefore derive the influence function the same way as in Section 9.3.2 and this means that we will have

\[ IF(\hat{\Phi}; T, F) = M(\Psi, T(F))^{-1}\Psi(\hat{\Phi}, T(F)) \]

where

\[ M(\Psi, F) = -\int \left( \Psi(\hat{\Phi}, \theta) \right)'_{\theta=\theta_0} dF(\hat{\Phi}) = J(\theta_0) \int x\psi'(x)dF(x). \]

for \( J \) defined as in (9.54) by Lemma 9.1 of Appendix A. The asymptotic parameter variance will then become

\[ V(T, F) = \int IF(\hat{\Phi}; T, F)IF(\hat{\Phi}; T, F)dF(\hat{\Phi}) \]

\[ = M(\psi, F)^{-1}Q(\psi, F)M(\psi, F)^{-T} \]

\[ = J^{-1}(\theta) \int \psi^2(x)dG(x) \]

\[ J^{-1}(\theta) \int \frac{\psi^2(x)dG(x)}{(\int x\psi'(x)dG(x))^2}, \] (9.59)

since

\[ Q(\Psi, F) = \int \Psi(\hat{\Phi}, T(F))\Psi^T(\hat{\Phi}, T(F))dF(\hat{\Phi}) = J(\theta) \int \psi^2(x)dG(x) \]

by Lemma 9.2 in Appendix A.

### 9.4.2 Multivariable Optimally B-Robust Estimator

Assume now that we want to find a function \( \psi \) that limits the gross sensitivity measured as

\[ \gamma_u^*(T, F) = \sup_{\hat{\Phi}} \{ |IF(\hat{\Phi}; T, F)| \} \]

while at the same time minimizes the trace of the variance \( \text{Tr} V(T, F) \). Here \(| \cdot |\) means taking the absolute value of each vector component. According to Theorem 9.2 below, this can be accomplished by using the function

\[ \psi(x) = \begin{cases} \frac{b}{-b} & \text{if } b < x - 1 - a \\ x - 1 - a & \text{if } -b < x - 1 - a \leq b \\ -b & \text{if } x - 1 - a \leq -b \end{cases} \] (9.60)

in the estimation method in (9.55).

**Theorem 9.2**

*Using the function in (9.60) in*

\[ \sum_{k=1}^{N_\omega} \Phi(\omega_k, \theta) \psi \left( \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta)} \right) = 0 \]
will minimize

\[ Tr \ V(T, F) = \frac{\int \psi^2(x)dG(x)}{(\int x\psi'(x)dG(x))^2} Tr \ J^{-1}(\theta) \]

while subject to

\[ \sup_{\hat{\Phi}} \left| IF(\hat{\Phi}; T, F) \right| \leq b \sum_{k=1}^{N_\omega} \left| J(\theta)^{-1} \frac{\Phi'(\omega_k, \theta)}{\Phi(\omega_k, \theta)} \right| \frac{\int x\psi'(x)dG(x)}{\int x\psi'(x)dG(x)} \]

Proof: See Appendix B. \qed

\[ \text{Figure 9.2: The parameter } a \text{ as a function of the parameter } b. \text{ The breakpoint between the two solution occurs at } b \approx 0.7968 \]

The parameter \( b \) in (9.60) can be considered a tuning factor which is selected by the user in order to strike a balance between the asymptotic bias and variance of the parameter estimates. The choice of \( b \) will automatically have an effect on \( a \) in order to maintain consistency. The theorem below explains how one should choose \( a \) from \( b \).
Theorem 9.3
The function \( a(b) \) is defined by the equations
\[
\begin{cases}
  e^{-b}e^{-(1+a)} + a = 0 & \text{if } 1 + a \leq b \\
  e^{1+a} = \ln \left( \frac{e^{-b} - e^{-b}}{b} \right) & \text{if } 1 + a > b.
\end{cases}
\] (9.61)

Proof: See Appendix C.

In Figure 9.2 the parameter \( a \) is presented as a function of the parameter \( b \). The breakpoint between the two sets of solutions in (9.61) occurs when \( b \approx 0.7968 \).

9.5 Numerical Example

In this section we will compare the efficiency and robustness of the proposed estimation method in (9.55) and (9.60) with that of the ML approach in (9.53). The objective is to estimate the parameters of the continuous-time spectrum
\[
\Phi(i\omega) = \frac{\lambda}{(i\omega)^2 + a_1(i\omega) + a_2}^2
\] (9.62)
where \( a_1 = 3 \), \( a_2 = 2 \) and \( \lambda = 1 \). The underlying continuous-time system is the autoregressive model
\[
y(t) = \frac{\sigma}{s^2 + a_1 s + a_2} e(t)
\] (9.63)
where \( e(t) \) is continuous-time Gaussian white noise such that \( \lambda = \sigma^2 \).

In Figure 9.3 the standard deviation for the parameter estimates of the continuous-time model found in (9.63) have been computed. Both the new (robust) and the original ML method. Simulations have been performed using \( N_{MC} = 100 \) Monte-Carlo runs for each value of \( b \). In the estimation process, the frequencies \( w = \{0, 0.01, \ldots, 2\pi\} \) have been used. In order to simulate outliers, an additive, random and exponentially distributed disturbance was introduced at 5% of all frequencies. The magnitude of this disturbance was set to twice the maximum value of the resonance peak of the spectrum (9.62). From the figure one can see that the outliers cause bias in the parameter estimates when the ML method is utilized. This bias is then reduced by approximately 50% with the use of the more robust criterion.

9.6 Summary

In this chapter a method for the rejection of frequency domain outliers is proposed. The algorithm is based on the work by Huber (1981) on M-estimators and the concept of influence function introduced by Hampel et al. (1986). The estimation takes place in the context of frequency domain continuous-time ARMA modelling, but the method can be also be applied to the discrete-time case. It is also proved that a certain choice of criterion will produce an optimal tradeoff between bias and variance under certain assumptions. Finally the method was illustrated by a numerical example which shows that the bias can be reduced by approximately 50%.
Figure 9.3: The standard deviations (upper) and bias (lower) of the parameter estimates. Robust (dotted) and ML method (dashed).
9.7 Appendix A

In this appendix we compute two quantities which are needed to compute the influence function and variance of the novel estimator in Section 9.4.1.

Lemma 9.1

\[ M(\Psi, F) = -\int (\Psi(\hat{\Phi}, \theta))' dF(\hat{\Phi}) = J(\theta) \int x\psi'(x) dG(x). \]

**Proof:** From the definition, we have

\[ M(\Psi, F) = -\int (\Psi(\hat{\Phi}, \theta))' dF(\hat{\Phi}) \]

\[ = -\sum_{k=1}^{N\omega} \int \left( \frac{\Phi'(\omega_k, \theta)}{\Phi(\omega_k, \theta)} \psi \left( \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta)} \right) \right)' \theta = \theta_0 \]

and

\[ -\left( \frac{\Phi'(\omega_k, \theta)}{\Phi(\omega_k, \theta)} \psi \left( \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta)} \right) \right)' = \]

\[ \left( \frac{\Phi'(\omega_k, \theta_0)}{\Phi(\omega_k, \theta_0)} \Phi(\omega_k, \theta_0)^T - \frac{\Phi''(\omega_k, \theta_0)}{\Phi(\omega_k, \theta_0)} \right) \psi \left( \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta_0)} \right) \]

\[ + \frac{\Phi'(\omega_k, \theta_0)}{\Phi(\omega_k, \theta_0)} \Phi'(\omega_k, \theta_0)^T \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta_0)} \psi' \left( \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta_0)} \right). \]

Because of (9.56) the expected value of first expression on the right hand side of the equation above will be zero. Therefore we will have

\[ M(\Psi, F) = J(\theta) \int x\psi'(x) dG(x). \]

\[ \square \]

Lemma 9.2

\[ Q(\Psi, F) = \int \Psi(\hat{\Phi}, T(F))\Psi^T(\hat{\Phi}, T(F)) dF(\hat{\Phi}) = J(\theta) \int \psi^2(x) dG(x) \]  

(9.64)

**Proof:**

\[ Q(\Psi, F) = \int \Psi(\hat{\Phi}, T(F))\Psi^T(\hat{\Phi}, T(F)) dF(\hat{\Phi}) \]

\[ = J(\theta) \int \psi^2(x) dG(x) \]

where \( J \) is defined as in (9.54).

\[ \square \]
Appendix B Proof of Theorem 9.2

Proof: In our case we will have

\[ \text{Tr } V(T, F) = \text{Tr } \frac{\int \psi^2(x) dG(x)}{\left( \int x \psi'(x) dG(x) \right)^2} J^{-1}(\theta) \]
\[ = \frac{\int \psi^2(x) dG(x)}{\left( \int x \psi'(x) dG(x) \right)^2} \text{Tr } J^{-1}(\theta) \]

and

\[ |IF(\hat{\Phi}; T, F)| = |M(\Psi, T(G))^{-1} \Psi(\hat{\Phi}, T(F))| \]
\[ = \left| \frac{J(\theta)^{-1}}{\int x \psi'(x) dG(x)} \sum_{k=1}^{N_\omega} \frac{\Phi'_b(\omega_k, \theta)}{\Phi(\omega_k, \theta)} \psi \left( \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta)} \right) \right| \]
\[ \leq \left| \frac{b}{\int x \psi'(x) dG(x)} \sum_{k=1}^{N_\omega} J(\theta)^{-1} \frac{\Phi'_b(\omega_k, \theta)}{\Phi(\omega_k, \theta)} \right|. \]

According to Theorem 1 and Equation (2.4.10) on pp 122 in (Hampel et al., 1986) a version of the so called Huber function

\[ \psi(x) = \begin{cases} 
  b & \text{if } b < x - 1 - a \\
  x - 1 - a & \text{if } -b < x - 1 - a \leq b \\
  -b & \text{if } x - 1 - a \leq -b
\end{cases} \]

minimizes

\[ \frac{\int \psi^2(x) dG(x)}{\left( \int x \psi'(x) dG(x) \right)^2} \]

among all mappings \( \psi \) that satisfy

\[ \int \psi(x) dG(x) = 0 \]
\[ \sup_x \left| \frac{\psi(x)}{\int x \psi'(x) dG(x)} \right| \leq \frac{b}{\left| \int x \psi'(x) dG(x) \right|}. \]

Therefore, using the function in (9.60) in

\[ \sum_{k=1}^{N_\omega} \frac{\Phi'_b(\omega_k, \theta)}{\Phi(\omega_k, \theta)} \psi \left( \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta)} \right) = 0 \]

will minimize

\[ \text{Tr } V(T, F) = \frac{\int \psi^2(x) dG(x)}{\left( \int x \psi'(x) dG(x) \right)^2} \text{Tr } J^{-1}(\theta) \]
while

$$\sup_{\Phi} |IF(\hat{\Phi}; T, F)| \leq b \sum_{k=1}^{N} \left| \frac{J(\theta) - 1}{\Phi(\omega_k, \theta)} \right| \int x\psi'(x)dG(x).$$

\[\square\]

\section*{9.9 Appendix C Proof of Theorem 9.3}

\textbf{Proof:} The parameter $a$ is then selected such that the estimator is asymptotically unbiased when the data is produced by the model. That is

$$\int \psi(x)dG(x) = \int_{0}^{\infty} [x - 1 - a]_{-b}^b e^{-x}dx = 0.$$  

This means that three different cases have to be considered.

\textit{Case 1}

The first, when $b \leq -1 - a$ which is trivial since

$$\int_{0}^{\infty} [x - 1 - a]_{-b}^b e^{-x}dx = b = 0$$

yields $b = 0$.

\textit{Case 2}

The second case is when $-b < -1 - a \leq b$. Here

$$\int_{0}^{\infty} [x - 1 - a]_{-b}^b e^{-x}dx = \int_{0}^{1+a+b} (x - 1 - a)e^{-x}dx + \int_{1+a+b}^{\infty} (x - 1 - a)e^{-x}dx.$$  

The first integral in this expression will be

$$\int_{0}^{1+a+b} (x - 1 - a)e^{-x}dx =$$

$$[e^{-x} - xe^{-x} + (1 + a)]_{0}^{1+a+b}$$

$$= [(a - x)e^{-x}']_{0}^{1+a+b}$$

$$= (-1 - b)e^{-b}e^{-(1+a)} - a.$$
The second integral is
\[
\int_{1+a+b}^{\infty} be^{-x} dx
= \left[ -be^{-x} \right]_{1+a+b}^{\infty} = be^{-b} e^{-(1+a)}.
\]
This means that given \( b \) we have to chose \( a(b) \) such that
\[
(-1 - b)e^{-b} e^{-(1+a)} - a + be^{-b} e^{-(1+a)} = -e^{-b} e^{-(1+a)} - a = 0
\]
which in turn means that
\[
\begin{align*}
-1 + a & \leq b \\
e^{-b} e^{-(1+a)} & + a = 0
\end{align*}
\]
in this case.

**Case 3**

In the third case we have \(-1 - a \leq -b \) and
\[
\int_{0}^{\infty} \left[ x - 1 - a \right]_{0}^{b} e^{-x} dx = - \int_{0}^{1+a+b} be^{-x} dx
\]
\[
\int_{1+a-b}^{\infty} (x - 1 - a)e^{-x} dx + \int_{1+a-b}^{1+a+b} (x - 1 - a)e^{-x} dx.
\]
The first integral will in this case be
\[
- \int_{0}^{1+a-b} be^{-x} dx = \left[ be^{-x} \right]_{0}^{1+a-b} = be^{-b} e^{-(1+a)} - b
\]
while the second is
\[
\int_{1+a-b}^{1+a+b} (x - 1 - a)e^{-x} dx = \]
\[
= \left[ (a - x)e^{-x} \right]_{1+a-b}^{1+a+b} = (a - 1 - a - b)e^{-b} e^{-(1+a)} - (a - 1 - a + b)e^{b} e^{-(1+a)}
\]
\[
= ((-b - 1)e^{-b} - (b - 1)e^{b}) e^{-(1+a)}.
\]
The third integral is the same as previously
\[
\int_{1+a+b}^{\infty} (x - 1 - a)e^{-x} = be^{-b}e^{-(1+a)}.
\]
This means that we have
\[
\int_{0}^{\infty} [x - 1 - a]_0^b e^{-x} dx =
be^b e^{-(1+a)} - b + ((-b - 1)e^{-b} - (b - 1)e^b)e^{-(1+a)}
+ be^{-b}e^{-(1+a)}
= (e^b - e^{-b})e^{-(1+a)} - b = 0
\]
and the solution is
\[
\begin{cases}
a = \ln \frac{e^b - e^{-b}}{b} - 1 \\
1 + a \geq b.
\end{cases}
\]
Together, the three cases means that \( a(b) \) is defined by the equations
\[
\begin{cases}
 e^{-b}e^{-(1+a)} + a = 0 & \text{if } 1 + a \leq b \\
 e^{1+a} = \ln \frac{e^b - e^{-b}}{b} & \text{if } 1 + a \geq b.
\end{cases}
\]
Identification of Continuous-Time ARMA Models

“In off the moors, down through the mist bands, God-cursed Grendel came greedily loping. The bane of the race of men roamed forth, hunting for prey in the high hall.”

Beowulf (800th century)

10.1 Introduction

In Chapter 6 a spline interpolation method for continuous-time spectral estimation from non-uniformly sampled data was introduced. Later in the same chapter it was shown, that in the case of uniformly sampled data, the interpolation produced an estimate of the continuous-time Fourier transform by frequency weighting of the discrete-time Fourier transform such that

\[ \hat{Y}_c(i\omega) = F_{\ell+1,T_s}(i\omega)Y_{d,N_t}(e^{i\omega T_s}). \]  

(10.1)

From this transform, an estimate

\[ \hat{\Phi}_{c,T}(i\omega) = |F_{\ell+1,T_s}(i\omega)|^2 \hat{\Phi}_{d,N_t}(e^{i\omega T_s}) \]  

(10.2)

of the continuous-time power spectrum could be fashioned, where the weighting would be on the discrete-time periodogram. An optimal spectral weighting function

\[ F_{c,2T_s}(i\omega) = \frac{\Phi_c(i\omega, \theta_0)}{\Phi_d(e^{i\omega T_s}, \theta_0)} \]  

(10.3)

was also derived using the relationship

\[ \Phi_d(e^{i\omega T_s}, \theta_0) = \sum_{k=-\infty}^{\infty} \Phi_c(i\omega + \frac{2\pi}{T_s}k, \theta_0) \]  

(10.4)
between the continuous- and discrete-time spectrums $\Phi_c(i\omega, \theta_0)$ and $\Phi_d(e^{i\omega T_s}, \theta_0)$, and it indicated that spline interpolation of the sampled output $\{y(kT_s)\}_{k=0}^{N_t-1}$ might not be the optimal way to estimate the continuous-time spectrum.

Unfortunately, the optimal weighting depended on the true parameters of the time series model, which of course are unknown during identification. The solution presented in this chapter to the above dilemma is similar to the one in Chapter 4 where the system is replaced by a chain integrations such that

$$H_c(s) \approx \frac{1}{s^\ell},$$

(10.5)

and where $\ell = n - m$ is the relative degree or pole excess. This approximation can then be used in the summation formula in (10.4) and will yield another weighting factor

$$F_{2\ell, T_s}^c(i\omega) = \frac{e^{i\omega T_s} \Pi_{2\ell-1}^0 e^{i\omega T_s}}{(2\ell-1)!}$$

(10.6)

such that the spectrum can be estimated as

$$\hat{\Phi}_{c,T}(i\omega) = F_{2\ell, T_s}^c(i\omega) \hat{\Phi}_{d,N_t}(i\omega)$$

(10.7)

where the weighting only depends on $\ell$. The parameters can then be estimated using the continuous-time Whittle likelihood approach

$$\hat{\theta} = \arg \min_\theta \sum_{k=1}^{N_\omega} \hat{\Phi}_{c,T}(i\omega_k) + \log \Phi_c(i\omega_k, \theta).$$

(10.8)

It will be shown that the direct interpretation of this type of spectral weighting is equivalent to interpolating an estimate of the covariance function $\{\hat{r}(kT_s)\}_{k=0}^{N_t-1}$ instead of the output $\{y(t)\}_{k=0}^{N_t-1}$. It will also be proved, that this setup can be interpreted as time-domain interpolation with a new spline based kernel $G_{\ell, T_s}^c(t)$ such that

$$\hat{y}_c(t) = \sum_{k=0}^{N_t-1} y(kT_s) G_{\ell, T_s}^c(t - kT_s).$$

(10.9)

### 10.2 Outline

First, the model structure for continuous-time series models used in the chapter is introduced in Section 10.3. Then, in Section 10.4, an idea similar to the truncation of the expression for the pulse transfer function in Chapter 4 is explained. A procedure for the estimation of the continuous-time power spectrum is then introduced in Section 10.6. In Section 10.7 it is shown that for uniformly sampled data this method is equivalent to interpolation of the covariance function by polynomial splines. A function for time domain estimation of the output of the process is then computed by spectral factorization.
The question of whether interpolation, which in the time series case is named smoothing, is such a brilliant idea for spectral estimation is finally answered in Section 10.8. Here it is shown that smoothing estimates will always be troubled by a high frequency bias.

In Section 10.9, the factorization approach in Section 10.7 is then extended to the case of effective spectral and parameter estimation from non-uniformly sampled data. The key will be an efficient factorization the covariance matrix instead of the spectrum in order to reconstruct data on an uniform grid. This material can be seen as a continuation of the work started in the thesis by Erik K. Larsson and the material surrounding it (Larsson and Söderström, 2002; Larsson, 2003; Larsson and Mossberg, 2003).

### 10.3 Model and Representations

In this chapter we shall consider continuous-time ARMA models represented as

\[ y(t) = H_c(s)e(t) \]  

(10.10)

where \( e(t) \) is continuous time white noise such that

\[ Ee(t) = 0 \]

\[ Ee(t)e(s) = \sigma^2 \delta(t - s). \]

We assume that \( G(s) \) is strictly proper, so \( y(t) \) itself does not have a white-noise component, but is a well defined second order, stationary process. Its spectrum (spectral density) can be written as

\[ \Phi_c(i\omega) = \sigma^2 |H_c(i\omega)|^2 \]  

(10.11)

We shall again consider a general model parametrization

\[ H_c(s, \theta) \]  

(10.12)

where the model parameter vector \( \theta \) includes the noise variance \( \lambda \) (whose true value is \( \sigma^2 \)). The transfer function \( H_c \) can be parameterized by \( \theta \) in an arbitrary way, for example by the conventional numerator and denominator parameters:

\[ H_c(s, \theta) = \frac{B(s)}{A(s)} \]

\[ A(s) = s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_n \]

\[ B(s) = s^m + b_1 s^{m-1} + \cdots + b_m \]

\[ \theta = (a_1 \ a_2 \ \cdots \ a_n \ b_1 \ b_2 \ \cdots \ b_m \ \lambda)^T. \]  

(10.13)

### 10.4 Truncating the Poisson Summation Formula

A straightforward way to do frequency domain identification of a continuous time series model from uniformly sampled data would be to compute the discrete-time periodogram from sampled data such that

\[ \hat{\Phi}_{d,N_1}(e^{i\omega T}) = |Y_{d,N_1}(e^{i\omega T})|^2. \]  

(10.14)
where

\[ Y_{d,N_t}(e^{i\omega T_s}) = \frac{1}{\sqrt{N_t}} \sum_{k=1}^{N_t} y(kT_s) e^{-i\omega kT_s} \]  \hspace{1cm} (10.15)

Parameters, would then be estimated using the discrete-time Whittle likelihood procedure

\[ \hat{\theta} = \arg \min_{\theta} \sum_{k=1}^{N_f} \frac{\hat{\Phi}_{d,N_t}(e^{i\omega kT_s})}{\hat{\Phi}_{d}(e^{i\omega kT_s}, \theta)} + \log \Phi_{d}(e^{i\omega kT_s}, \theta) \]  \hspace{1cm} (10.16)

where the discrete-time spectrum is expressed in terms of continuous-time parameters by the relationship

\[ \Phi_{d}(e^{i\omega T_s}, \theta) = \sum_{k=-\infty}^{\infty} \Phi_c(i\omega + i\frac{2\pi}{T_s} k, \theta). \]  \hspace{1cm} (10.17)

A drawback connected with using the formula in (10.17) is of course the infinite sum. Good approximations can however be achieved with a limited number of terms

\[ \Phi_{d}(e^{i\omega T_s}) = \sum_{k=-N_f}^{N_f} \Phi_c(i\omega + i\frac{2\pi}{T_s} k) \]  \hspace{1cm} (10.18)

when the continuous-time system is strictly proper.

### 10.4.1 Numerical Example

In Figure 10.1 and Table 10.4.1, we have estimated the second-order continuous-time autoregressive (AR) model

\[ y(t) = H_c(s)e(t) \]

\[ H_c(s) = \frac{1}{s^2 + a_1 s + a_2} e(t) \]  \hspace{1cm} (10.19)

with the true parameters \( \sigma = 1, \ a_1 = 3 \) and \( a_2 = 2 \). The duration of the data set was \( T = 1000s \) with the sampling interval \( T_s = 1s \). Estimates were produced using \( N_{MC} = 250 \) Monte-Carlo simulations and the information in the table and figure is based on the mean of those values. The method that has been employed is that of (10.16) where the discrete time spectrum has been approximated using (10.18).

Figure 10.1 illustrates the frequency-domain bias which could occur in the transfer function if only the central term \( (N_f = 0) \) in (10.18) have been used, which means that we have assumed that

\[ \Phi_{d}(e^{i\omega T_s}, \theta) = \Phi_c(i\omega, \theta). \]  \hspace{1cm} (10.20)

This will produce a biased estimate which is illustrated by the dashed-dotted line in the Bode diagram for the system. A set of systems with virtually no bias has also been
Figure 10.1: Bode diagram comparing the Whittle likelihood estimator with $N_f = 0$ (dashdot) and $N_f = 5$ (dashed) in (10.18) to the true system (solid). The system is $H_c(s) = \frac{\sigma}{s^2 + a_1s + a_2}$ where $\sigma = 1$, $a_1 = 2$ and $a_2 = 2$. The sampling interval is $T_s = 0.8$ and frequencies up to the Nyquist frequencies have been used. When $N_f = 0$ the estimated system will be biased. For the case $N_f = 5$ the dashed curve is almost identical to the solid curve of the true system.

estimated using $N_f = 5$ in (10.18). The mean value of these are also illustrated in the figure as a dotted line which is almost identical to the true system which is represented by the solid line.

In Table 10.4.1 the mean parameter values are also illustrated for $N_f = 0$ and $N_f = 5$ in (10.18). From the table we see that assuming that the discrete-time spectrum is equal to the continuous-time spectrum will also produce bias in the parameters. However, including $N_f = 5$ terms in (10.18) will almost entirely remove this.

10.5 Direct Estimation

In the previous section we have estimated continuous-time parameters by parameterizing the discrete-time spectrum in terms of the continuous-time parameters. Then, we used the
Identification of Continuous-Time ARMA Models

Table 10.1: Mean values of $N_{MC} = 250$ parameter estimates of the system is $H_c(s) = \frac{\sigma}{\sigma^2 + a_1 s + a_2}$ where $\sigma = 1$, $a_1 = 2$ and $a_2 = 2$ using the Whittle likelihood estimator with (10.18). The sampling interval is $T_s = 0.8$ and frequencies up to the Nyquist frequencies have been used. When $N_f = 0$, the parameters will carry bias. For the case $N_f = 5$ the parameters are close those of the true system.

<table>
<thead>
<tr>
<th>Method</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_f = 0$</td>
<td>5.090122</td>
<td>3.217543</td>
<td>1.675111</td>
</tr>
<tr>
<td>$N_f = 5$</td>
<td>3.077730</td>
<td>2.044676</td>
<td>1.020823</td>
</tr>
</tbody>
</table>

discrete-time Whittle estimator. In this section, the approach is slightly different. Instead of using a discrete-time identification method we would like to employ the continuous-time Whittle likelihood estimator developed in Chapter 2. Mainly, because this would allow us to use the continuous-time spectrum $\Phi_c(i\omega)$ directly, without going via the summation formula in (10.17) or (10.18), and the discrete-time spectrum $\Phi_d(e^{i\omega T_s})$. This means that we would like to use the method

$$\hat{\theta} = \arg \min_\theta \sum_{k=1}^{N_\omega} \hat{\Phi}_{c,T}(i\omega_k) + \log \Phi_c(i\omega_k, \theta).$$

(10.21)

where the continuous-time spectrum is estimated as

$$\hat{\Phi}_{c,T}(i\omega) = |Y_{c,T}(i\omega)|^2.$$

and

$$Y_{c,T}(i\omega) = \frac{1}{\sqrt{T}} \int_0^T y(t)e^{-i\omega t} dt.$$

This approach is similar to that found in (6.14) Chapter 6. The problem of computing the estimate $\hat{\Phi}_{c,T}$ will be treated below.

10.6 Estimating the Continuous-Time Spectrum

The main problem here, is that the discrete-time periodogram $\hat{\Phi}_{d,N_i}(e^{i\omega T_s})$ can be readily found from the uniformly sampled data in (10.14) and (10.15), while we need the continuous-time periodogram $\hat{\Phi}_{c,T}(i\omega)$ in the criterion (10.21). How can the latter then be estimated or computed from $\hat{\Phi}_{d,N_i}(e^{i\omega T_s})$? We will be looking for relationships like

$$\hat{\Phi}_{c,T}(i\omega) = \Phi_f(i\omega)\hat{\Phi}_{d,N_i}(e^{i\omega T_s})$$

(10.22)

for a suitable function $\Phi_f(i\omega)$. We have already in Chapter 6 studied such transformations that were based on the assumptions that the signal is piecewise constant (ZOH) or
10.6 Estimating the Continuous-Time Spectrum

piecewise linear (FOH) between samples, see (6.38). To do better than that we could argue as follows: If the true signal parameters \( \theta_0 \) were known, we would have

\[
E\hat{\Phi}_{d,N_t}(e^{i\omega T_s}) = \Phi_d(e^{i\omega T_s}, \theta_0)
\]

and

\[
E\hat{\Phi}_{c,T}(i \omega) = \Phi_c(e^{i\omega T_s}, \theta_0).
\]

This means that the ideal spectral weighting in (10.22) would be characterized by

\[
\Phi_f(i \omega) = \frac{\Phi_c(i \omega, \theta_0)}{\Phi_d(e^{i\omega T_s}, \theta_0)}.
\]

Since \( \theta_0 \) is unknown, we cannot construct \( \Phi_f(i \omega) \) in this way, but the point is that when \( T_s \to 0 \), \( \Phi_f(i \omega) \) in (10.23) will approach \( F_{2\ell,T_s}^c(i \omega) \), which is defined in (5.62). It turns out that \( F_{2\ell,T_s}^c(i \omega) \) is real, positive, and does not depend on the signal parameters \( \theta_0 \), but only on the relative degree (pole excess) \( \ell = n - m \) of the time series model. This is what we will show now.

Let the model be strictly proper, stable and \( \ell = n - m \) be its relative degree (or pole excess), i.e. the difference between the number of poles and zeros of the system. Then, at high frequencies, we can approximate the system as a chain of \( \ell \) integrators with the continuous-time spectrum

\[
\Phi_c(i \omega) = \frac{1}{|i \omega|^{2\ell}}.
\]

The discrete-time spectrum for this model is then well known as (Wahlberg, 1988)

\[
\Phi_d(e^{i\omega T_s}) = \frac{1}{|i \omega|^{2\ell}} + \sum_{k \neq 0} \frac{1}{|i \omega + i \frac{2\pi}{T_s} k|^{2\ell}} = \frac{T_s^{2\ell}}{2\ell} (-1)^\ell \frac{e^{i\omega T_s} \Pi_{2\ell-1}(e^{i\omega T_s})}{(2\ell - 1)!} (e^{i\omega T_s} - 1)^{2\ell}
\]

which means that we get the spectral weighting

\[
\Phi_f(i \omega) = \frac{\Phi_c(i \omega)}{\Phi_d(e^{i\omega T_s})} = \frac{1}{|i \omega|^{2\ell}} + \sum_{k \neq 0} \frac{1}{|i \omega + i \frac{2\pi}{T_s} k|^{2\ell}}
\]

\[
= \left( \frac{e^{i\omega T_s} - 1}{i \omega T_s} \right)^{2\ell} \frac{\Pi_{2\ell-1}(e^{i\omega T_s})}{(2\ell - 1)!} = F_{2\ell,T_s}^c(i \omega),
\]

where \( \Pi_{2\ell-1}(z) \) are the Euler-Frobenius polynomials found in Section 4.5.1 of Chapter 4. Observe that since \( \Phi_f(i \omega) \geq 0 \) this actually means that

\[
F_{2\ell,T_s}^c(i \omega) = \left( \frac{e^{i\omega T_s} - 1}{i \omega T_s} \right)^{2\ell} \frac{\Pi_{2\ell-1}(e^{i\omega T_s})}{(2\ell - 1)!} \frac{1}{|i \omega T_s|^{2\ell}}
\]

In Figure 10.2 we see that there is a very good correspondence between \( \Phi_c(i \omega, \theta_0)/\Phi_d(i \omega, \theta_0) \) and \( F_{2\ell,T_s}^c(i \omega) \) for the system in (10.19). This observation is verified by the following theoretical result.
Theorem 10.1

Assume \( F_{2\ell,T_s}^c(i\omega) \) is defined as in (10.26) and that \( \ell \geq 1 \). Then

\[
\frac{\Phi_c(i\omega, \theta_0)}{\Phi_d(e^{i\omega T_s}, \theta_0)} \to F_{2\ell,T_s}^c(i\omega)
\]
as \( T_s \to 0 \).

Proof: First, as \( T_s \to 0 \)

\[
\frac{\Phi_c(i\omega + i\frac{2\pi}{T_s}k, \theta_0)}{\sigma^2 |i\omega + i\frac{2\pi}{T_s}k|^{2\ell}} \to 1
\]
if \( k \neq 0 \). This has the consequence that

\[
\frac{\Phi_c(i\omega, \theta_0)}{\Phi_d(e^{i\omega T_s}, \theta_0)} \to \frac{\Phi_c(i\omega, \theta_0)}{\Phi_c(i\omega, \theta_0) + \sum_{k \neq 0} \frac{\sigma^2}{|i\omega + i\frac{2\pi}{T_s}k|^{2\ell}}}
\]
as \( T_s \to 0 \). From Lemma 3.2 in (Wahlberg, 1988)

\[
F_{2\ell,T_s}^c(i\omega) = \frac{1}{|i\omega|^{2\ell} + \sum_{k \neq 0} \frac{\sigma^2}{|i\omega + i\frac{2\pi}{T_s}k|^{2\ell}}}
\]

By putting the two previous expressions on a common denominator, we get the following relation

\[
\frac{\Phi_c(i\omega, \theta_0)}{\Phi_d(e^{i\omega T_s}, \theta_0)} - F_{2\ell,T_s}^c(i\omega) \to \Phi_c(i\omega, \theta_0)\Phi_r(i\omega, \theta_0)\Phi_s(i\omega)
\]
where

\[
\Phi_r(i\omega, \theta_0) = \frac{1 - \Phi_c(i\omega, \theta_0)}{\Phi_c(i\omega, \theta_0) + \sum_{k \neq 0} \frac{\sigma^2}{|i\omega + i\frac{2\pi}{T_s}k|^{2\ell}}}
\]

and

\[
\Phi_s(i\omega) = \sum_{k \neq 0} \frac{\sigma^2}{|i\omega + i\frac{2\pi}{T_s}k|^{2\ell}}.
\]

Since \( F_{2\ell,T_s}^c \) and \( \Phi_r \) are bounded and

\[
\frac{1}{|i\omega + i\frac{2\pi}{T_s}k|^{2\ell}} \leq C \left( \frac{T_s}{k} \right)^{2\ell}
\]
if \( k \neq 0 \), the result

\[
\left| \frac{\Phi_c(i\omega, \theta_0)}{\Phi_d(e^{i\omega T_s}, \theta_0)} - F_{2\ell,T_s}^c(i\omega) \right| \leq CT_s^{2\ell}
\]
will follow.
Figure 10.2: Comparison of $\Phi_c(i\omega, \theta_0)/\Phi_d(e^{i\omega T_s})$ (solid) to $F_{2\ell,T_s}^c(i\omega)$ (dashdot) for the system $H_c(s) = \frac{\sigma}{s^2 + a_1 s + a_2}$ where $\sigma = 1$, $a_1 = 2$ and $a_2 = 2$. In the inner plot $T_s = 2$ and in the outer plot $T_s = 0.5$. As can be seen from the figure, the discrepancy is small at these sampling intervals and decreases as $T_s$ decrease.
Therefore a reasonable estimate of the continuous-time spectrum would be

\[ \hat{\Phi}_{c,T}(i\omega) = \frac{e^{i\omega T_s} - 1}{i\omega T_s} \frac{2^\ell}{\Pi_{2\ell-1}((e^{i\omega T_s})^{2\ell-1})} \hat{\Phi}_{d,N_t}(e^{i\omega T_s}). \]

### 10.6.1 Numerical Illustration

In Table 10.6.1 we have compared the performance of the approach in the section above, which we call Method 2, to using the discrete-time Whittle likelihood approach with the exact discrete time spectrum in (10.17), which we call Method 1. We have used different sampling intervals, and the mean parameter values have been estimated by \( N_{MC} = 250 \) Monte-Carlo simulations. The system is the one in (10.19) and the correspondence between the mean parameter estimates is good.

**Table 10.2: Comparison of mean values for \( N_{MC} = 250 \) Monte-Carlo simulations of parameter estimates versus the sample time \( T_s \). Method 1 is using the discrete-time Whittle likelihood approach with the exact discrete time spectrum. Method 2 employs the continuous-time Whittle likelihood approach with the spectral estimator (10.26). The system is \( H_c(s) = \frac{\sigma}{s^2 + a_1 s + a_2} \) where \( \sigma = 1, a_1 = 3 \) and \( a_2 = 2 \) and the performance of the methods are similar.**

<table>
<thead>
<tr>
<th>( T_s )</th>
<th>Method</th>
<th>( a_1 = 3 )</th>
<th>( a_2 = 2 )</th>
<th>( \sigma = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>1</td>
<td>3.0355</td>
<td>2.0189</td>
<td>1.0071</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>2.9935</td>
<td>2.0166</td>
<td>1.0007</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>3.0186</td>
<td>2.0206</td>
<td>1.0054</td>
</tr>
<tr>
<td>0.5</td>
<td>2</td>
<td>3.0476</td>
<td>2.0376</td>
<td>1.0150</td>
</tr>
<tr>
<td>0.4</td>
<td>1</td>
<td>3.0266</td>
<td>2.0150</td>
<td>1.0055</td>
</tr>
<tr>
<td>0.4</td>
<td>2</td>
<td>3.0449</td>
<td>2.0240</td>
<td>1.0110</td>
</tr>
<tr>
<td>0.3</td>
<td>1</td>
<td>3.0276</td>
<td>2.0143</td>
<td>1.0071</td>
</tr>
<tr>
<td>0.3</td>
<td>2</td>
<td>3.0327</td>
<td>2.0168</td>
<td>1.0086</td>
</tr>
<tr>
<td>0.2</td>
<td>1</td>
<td>3.0009</td>
<td>2.0118</td>
<td>1.0009</td>
</tr>
<tr>
<td>0.2</td>
<td>2</td>
<td>3.0017</td>
<td>2.0122</td>
<td>1.0011</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>3.0245</td>
<td>2.0210</td>
<td>1.0071</td>
</tr>
<tr>
<td>0.1</td>
<td>2</td>
<td>3.0246</td>
<td>2.0211</td>
<td>1.0072</td>
</tr>
</tbody>
</table>

In Figure 10.3 we have compared the mean values of \( N_{MC} = 250 \) Monte-Carlo simulations of parameter estimates versus the sample time \( T_s \). The dotted line represents the method where we have assumed that \( \hat{\Phi}_{c,T}(i\omega) = \hat{\Phi}_{d,N_t}(e^{i\omega T_s}) \). The solid line shows the result when \( \hat{\Phi}_{c,T}(i\omega) = F_{2\ell,T_s}^c(e^{i\omega T_s}) \hat{\Phi}_{d,N_t}(e^{i\omega T_s}) \) as in (10.26). Frequencies up to \( 2\pi \) rad/s have been used and the time of observation is \( T = 1000 \) s. For low sampling rates the difference between the two estimates is significant. This is due to the bias that occurs if no spectral weighting by \( F_{2\ell,T_s}^c(i\omega) \) is used.
Figure 10.3: Comparison of mean values for $N_{MC} = 250$ Monte-Carlo simulations of parameter estimates versus the sample time $T_s$. The dotted line represents the method where we have assumed that $\hat{\Phi}_{c,T}(i\omega) = \hat{\Phi}_{d,N_t}(e^{i\omega T_s})$. The solid line shows the result when $\hat{\Phi}_{c,T}(i\omega) = F_{c,T_s}^\ell(e^{i\omega T_s})\hat{\Phi}_{d,N_t}(e^{i\omega T_s})$ as in (10.26). The system is $H_c(s) = \frac{\sigma}{s^2 + a_1s + a_2}$ where $\sigma = 1$, $a_1 = 2$ and $a_2 = 2$. Frequencies up to $2\pi$ rad/s have been used and the time of observation is $T = 1000$ s. For low sampling rates the difference between the two estimates is significant.
10.7 Interpretations in Terms of Splines

Imagine that it was possible to compute the spectral factor \( G_{c,\ell,T_s}^c(i\omega) \) of \( F_{2\ell,T_s}^c(i\omega) \) such that
\[
F_{2\ell,T_s}^c(i\omega) = |G_{c,\ell,T_s}^c(i\omega)|^2 .
\] (10.29)

Then it would be possible to rewrite the estimate of the continuous-time spectrum as
\[
\hat{\Phi}_{c,T}(i\omega) = F_{2\ell,T_s}^c(i\omega) \hat{\Phi}_{d,N_t}(e^{i\omega T_s}) = |G_{c,\ell,T_s}^c(i\omega)Y_{d,N_t}(e^{i\omega T_s})|^2
\] (10.30)
\[
= |G_{c,\ell,T_s}^c(i\omega)Y_{d,N_t}(e^{i\omega T_s})|^2 .
\] (10.31)

This in turn would mean that we could define a function
\[
\hat{y}_c(t) = \sum_{k=0}^{N_t-1} y(kT_s)G_{c,\ell,T_s}^c(t - kT_s).
\] (10.32)

Now, as the reader might have noticed, the entity \( F_{2\ell,T_s}^c(i\omega) \) is just the Fourier transform of the fundamental spline function \( F_{c,2\ell,T_s}^c(i\omega) \) introduced in Section 5.6 and its factorization would be
\[
F_{2\ell,T_s}^c(i\omega) = \left| \frac{e^{i\omega T_s} - 1}{i\omega T_s} \right|^{2\ell} = \left| C\left(e^{i\omega T_s}\right)\left(\frac{e^{i\omega T_s} - 1}{i\omega T_s}\right)^{\ell}\right|^2
\] (10.33)

where
\[
|C(z)|^2 = \left| \frac{(2\ell - 1)!}{\Pi_{2\ell-1}(z)} \right| .
\] (10.34)

This is possible, since if \( z \) is a root of \( \Pi_{2\ell-1}(z) \) the so is \( 1/z \) (Sobolev, 1977). Hence we have
\[
G_{c,\ell,T_s}^c(i\omega) = C(e^{i\omega T_s})\left(\frac{e^{i\omega T_s} - 1}{i\omega T_s}\right)^{\ell} ,
\] (10.35)
and since the Fourier transform of a traditional B-spline of order \( \ell \) is
\[
B_{\ell,T_s}^c(i\omega) = \left(\frac{e^{i\omega T_s - 1}}{i\omega T_s}\right)^\ell
\]
(10.40)
the function \( G_{\ell,T_s}^c \) can be expressed as
\[
G_{\ell,T_s}^c(t) = \sum_{k=0}^{\infty} c(k) B_{\ell,T_s}^c(t - kT_s)
\]
(10.41)
with the convolution property
\[
F_{2\ell,T_s}^c(i\omega) = \int_{-\infty}^{\infty} G_{\ell,T_s}^c(t - \tau) G_{\ell,T_s}^c(\tau).
\]
(10.42)
The following example will illustrate the above line of reasoning for the case when \( \ell = 2 \).

**Example 10.1**

Let \( \ell = 2 \). Then we have
\[
F_{2\ell,T_s}^c(i\omega) = F_{4,T_s}^c(i\omega) = \left(\frac{e^{i\omega T_s - 1}}{i\omega T_s}\right)^4 \frac{\Pi_4(e^{i\omega T_s})}{6}
\]
(10.43)
where
\[
\Pi_3(z) = z^2 + 4z + 1.
\]
(10.44)
Now, we can write
\[
\Pi_3(z) = z^2 + 4z + 1 = (z + 2 - \sqrt{3})(z + 2 + \sqrt{3})
\]
(10.45)
\[
= \frac{z}{2 - \sqrt{3}}(1 + z^{-1}(2 - \sqrt{3}))(1 + z(2 - \sqrt{3}))
\]
(10.46)
which means that we can choose
\[
C(z) = \sqrt{6}\sqrt{2 - \sqrt{3}} \frac{1}{1 + z^{-1}(2 - \sqrt{3})}.
\]
(10.47)
Hence we will have
\[
G_{\ell,T_s}^c(i\omega) = \sqrt{6}\sqrt{2 - \sqrt{3}} \left(\frac{e^{i\omega T_s - 1}}{e^{i\omega T_s}}\right)^{2} \frac{e^{i\omega T_s}}{e^{i\omega T_s} + 2 - \sqrt{3}}
\]
(10.48)
and
\[
G_{2,T_s}^c(i\omega) = \sqrt{6}\sqrt{2 - \sqrt{3}} \sum_{k=0}^{\infty} (-1)^k (2 - \sqrt{3})^k B_{2,T_s}^c(t - kT_s).
\]
(10.49)
This function is illustrated in Figure 10.4.

It is also worthwhile to notice that by (2.2.2), (2.2.4) and (2.2.6) in Stoica and Moses
Figure 10.4: The function $G_{2,1}^c(t)$ (upper figure) defined in (10.41) and (10.49) its convolution $F_{4,1}^c(\tau) = \int_{-\infty}^{\infty} G_{2,1}^c(\tau - t)G_{2,1}^c(t)dt$ (lower figure).
where

$$\hat{r}_d(kT_s) = \frac{1}{N_t} \sum_{l} N_t y((k+l)T_s), \quad 0 \leq k \leq N_t$$ (10.51)

one will have

$$\hat{\Phi}_{c,T}(i\omega) = F_{2\ell,T_s}(i\omega) \hat{\Phi}_{d,N_t}(e^{i\omega T_s})$$ (10.52)

$$= F_{2\ell,T_s}(i\omega) \sum_{k=-N_t}^{N_t-1} \hat{r}_d(kT_s) e^{-i\omega kT_s}$$ (10.53)

$$= \int_{-T}^{T} \hat{r}_c(t) e^{-i\omega t} dt$$ (10.54)

where

$$\hat{r}_c(t) = \sum_{k=-N_t}^{N_t-1} \hat{r}_d(kT_s) F_{2\ell,T_s}^c(t - kT_s)$$ (10.55)

and $F_{2\ell,T_s}^c(t)$ is the fundamental spline basis function of order $2\ell$. Hence, in the case of continuous time series models and uniformly sampled data, one is actually interpolating the covariance function instead of the output as in the case of input-output models.

## 10.8 Reconstruction by Smoothing

In Section 10.5 a method for approximating the continuous-time periodogram using the discrete-time periodogram

$$\hat{\Phi}_{c,T}(i\omega) = \Phi_f(e^{i\omega T_s}) \hat{\Phi}_{d,N_t}(e^{i\omega T_s})$$ (10.56)

was presented for the case of uniformly sampled data $t_k = kT_s, \ k = 1, \ldots, N$. Here the optimal choice of $\Phi_f$ would be

$$\Phi_f(i\omega) = \frac{\Phi_c(i\omega, \theta_0)}{\Phi_d(e^{i\omega T_s}, \theta_0)}$$ (10.57)

if the parameters of the true system were known. What the following theorem will show, is that when the inter sample behavior is estimated by smoothing we will in fact have

$$\Phi_f(i\omega) = \frac{\Phi_c(i\omega, \theta_0)^2}{\Phi_d(e^{i\omega T_s}, \theta_0)^2},$$ (10.58)

which will produce a spectral estimate $\hat{\Phi}_{c,T}(i\omega)$ that will be biased at high frequencies.
Theorem 10.2
Assume that we want to create a linear estimator of the continuous-time process \( y(t) \) when we only know equidistantly distributed samples. This estimator would then be expressed as

\[
\hat{y}_c(t) = \sum_{k=-\infty}^{\infty} f(t, t_k, \theta_0) y(t_k).
\]

The best such estimator, in a mean square sense, would be characterized by the coefficients \( \{ f(t, t_k, \theta_0) \}_{k=-\infty}^{\infty} \) minimizing

\[
E \left| y(t) - \sum_{k=-\infty}^{\infty} f(t, t_k, \theta_0) y(t_k) \right|^2.
\]

Then, if we define

\[
\hat{Y}_c(i\omega) = \int_{-\infty}^{\infty} \hat{y}_c(t) e^{-i\omega t} dt \tag{10.59}
\]

\[
\hat{\Phi}_c(i\omega) = \left| \hat{Y}_c(i\omega) \right|^2 \tag{10.60}
\]

and

\[
Y_d(e^{i\omega T_s}) = \sum_{-\infty}^{\infty} y_c(kT_s) e^{-i\omega kT_s} dt \tag{10.61}
\]

\[
\hat{\Phi}_d(e^{i\omega T_s}) = \left| Y_d(e^{i\omega T_s}) \right|^2 \tag{10.62}
\]

we have

\[
\hat{\Phi}_c(i\omega) = \frac{\Phi_c(i\omega, \theta_0)^2}{\Phi_d(e^{i\omega T_s}, \theta_0)^2} \hat{\Phi}_d(e^{i\omega T_s}). \tag{10.63}
\]

Proof: The solution to this problem are the classical Yule-Walker equations

\[
E \left( y(t) - \sum_{k=-\infty}^{\infty} f(t, t_k, \theta_0) y(t_k) \right) y(t_l) \quad \forall l \in \mathbb{Z}
\]

which are equivalent to

\[
r(t - t_l, \theta_0) = \sum_{k=-\infty}^{\infty} f(t, t_k, \theta_0) r(t_k - t_l, \theta_0) \quad \forall l \in \mathbb{Z} \tag{10.64}
\]

If we define \( t' = t - t_l \) and \( t'_\tau = t_k - t_l \) the expression will become

\[
r(t', \theta_0) = \sum_{\tau=-\infty}^{\infty} f(t' + t_l, t'_\tau + t_l, \theta_0) r(t'_\tau, \theta_0) \quad \forall l \in \mathbb{Z}
\]
Since the left side is independent of $t_l$ the same must be true for the right. Hence
\[ f(t' + t_l, t + t_l, \theta_0) = f(t', t', \theta_0) \quad \forall t \in \mathbb{Z} \]
which in turn means that
\[ f(t', t, \theta_0) = f(t' - t, \theta_0). \]
The expression in (10.64) will therefore become
\[ r(t', \theta_0) = \sum_{\tau = -\infty}^{\infty} f(t' - t', \theta_0) r(t', \theta_0). \]
Taking the continuous-time Fourier transform of this expression will lead to
\[ \Phi_c(i \omega, \theta_0) = \int_{-\infty}^{\infty} r(t, \theta_0) e^{-i \omega t'} dt' = \sum_{\tau = -\infty}^{\infty} \int_{-\infty}^{\infty} f(t' - t', \theta_0) r(t', \theta_0) e^{-i \omega t'} dt' r(t', \theta_0) \]
\[ = \sum_{\tau = -\infty}^{\infty} F(i \omega, \theta_0) e^{-i \omega t'} r(t', \theta_0) \]
\[ = F(i \omega, \theta_0) \sum_{\tau = -\infty}^{\infty} r(t', \theta_0) e^{-i \omega t'} = F(i \omega, \theta_0) \Phi_d(e^{i \omega T_s}, \theta_0). \]
Therefore we will have the expression
\[ F(i \omega, \theta_0) = \frac{\Phi_c(i \omega, \theta_0)}{\Phi_d(e^{i \omega T_s}, \theta_0)}. \]
and this means in fact that we get
\[ \hat{Y}_c(i \omega) = F(i \omega, \theta_0) Y_d(e^{i \omega T_s}) = \frac{\Phi_c(i \omega, \theta_0)}{\Phi_d(e^{i \omega T_s}, \theta_0)} Y_d(e^{i \omega T_s}) \quad (10.65) \]
and
\[ \hat{\Phi}_c(i \omega) = \left| \hat{Y}_c(i \omega) \right|^2 = \frac{\Phi_c(i \omega, \theta_0)^2}{\Phi_d(e^{i \omega T_s}, \theta_0)^2} \hat{\Phi}_d(e^{i \omega T_s}). \quad (10.66) \]

Unfortunately the smoothing-based approach above leads to a biased estimate of the spectrum. In particular at high frequencies. The reason is that for equidistant sampling, we get an estimate of the continuous-time spectrum which is
\[ \hat{\Phi}_{c,T}(i \omega) = |Y_c(i \omega)|^2 = \frac{\Phi_c(i \omega, \theta_0)^2}{\Phi_d(e^{i \omega T_s}, \theta_0)^2} \hat{\Phi}_d,e_{N_t}(e^{i \omega T_s}) \quad (10.67) \]
where the quotient
\[ \frac{\Phi_c(i\omega, \theta_0)^2}{\Phi_d(e^{i\omega T_s}, \theta_0)^2} \approx 1 \]  
(10.68)

at low frequencies, but where the attenuation at high frequencies is too strong. Instead we would prefer the expression
\[ \hat{\Phi}_{c,T}(i\omega) = \frac{\Phi_c(i\omega, \theta_0)}{\Phi_d(e^{i\omega T_s}, \theta_0)} \]  
(10.69)

which could be realized through spectral factorization
\[ \frac{\Phi_c(i\omega, \theta_0)}{\Phi_d(e^{i\omega T_s}, \theta_0)} = \left| \frac{H_c(i\omega, \theta_0)}{H_d(e^{i\omega T_s}, \theta_0)} \right|^2 = \left| G_c(i\omega, \theta_0) \right|^2, \]  
(10.70)

such that
\[ G_c(i\omega, \theta_0) = \frac{H_c(i\omega)}{H_d(e^{i\omega T_s})}, \]  
(10.71)

as was illustrated in Section 10.7. The question that remains however, is whether this thinking can be transferred into the case of non-uniformly sampled data. The answer is that we can, but we have to leave the domain of transfer and covariance functions, and return to the underlying state-space descriptions and covariance matrices.

### 10.9 Reconstruction by Factorization

Assume that we have sampled the output of the model (10.10) non-uniformly, such that
\[ Y_n = (y(t_1), y(t_2), \ldots, y(t_{N_t}))^T. \]  
(10.72)

where \( n \) stands for non-uniform. Then, \( Y_n \) will have a multi-variable normal distribution, such that
\[ Y_n \sim \mathcal{N}(0, R_n). \]  
(10.73)

with
\[ R_n = EY_nY_n^T = \begin{pmatrix} r(t_1 - t_1) & r(t_1 - t_2) & \cdots & r(t_1 - t_{N_t}) \\ r(t_2 - t_1) & r(t_2 - t_2) & \cdots & r(t_2 - t_{N_t}) \\ \vdots & \vdots & \ddots & \vdots \\ r(t_{N_t} - t_1) & r(t_{N_t} - t_2) & \cdots & r(t_{N_t} - t_{N_t}) \end{pmatrix} \]  
(10.74)

Here \( r(\tau) \) is the continuous-time covariance function defined in (2.44). Assume now, that we want to reconstruct a new vector \( Y_r \) from \( Y_n \) defined as
\[ Y_r = (y_r(T_s), y_r(2T_s), \ldots, y_r(N_tT_s))^T \]  
(10.75)
which is a linear transformation of the values of \( Y_n \) such that
\[
Y_r = SY_n
\]  
(10.76)
for some \( N_t \times N_t \) matrix \( S \). We also want this \( Y_r \) to have the same statistical properties as the vector
\[
Y_u = (y(T_s), y(2T_s), \ldots, y(N_tT_s))^T.
\]  
(10.77)
of uniformly distributed samples of \( y(t) \). This means, that we want the following equation to be true
\[
R_r = EY_rY_r^T = EY_uY_u^T = R_u,
\]  
(10.78)
where
\[
R_u = \begin{pmatrix}
  r(0) & r(-T_s) & \ldots & r(-(N_t-1)T_s) \\
  r(T_s) & r(0) & \ldots & r(-(N_t-2)T_s) \\
  \vdots & \vdots & \ddots & \vdots \\
  r((N_t-1)T_s) & r((N_t-2)T_s) & \ldots & r(0)
\end{pmatrix}.
\]  
(10.79)
If we substitute expression (10.76) for \( Y_r \) into the relationship (10.78) we get the following expression for the transformation matrix \( S \)
\[
R_u = EY_rY_r^T = ESY_nY_n^T S^T = SR_nS^T = R_u.
\]  
(10.80)
Therefore, we have to solve for \( S \) in the following equation
\[
SR_nS^T = R_u
\]  
(10.82)
in order to give \( Y_r \) the same statistical properties as \( R_u \). Now, in Section 2.7.2, it was shown that by means of the Kalman filter, it is possible to decompose the covariance matrices \( R_n \) and \( R_u \) such that
\[
R_n = L_n R_{e,n} L_n^T
\]  
(10.83)
\[
R_u = L_u R_{e,u} L_u^T
\]  
(10.84)
if the true underlying state space model is known. Here, \( L_n \) and \( L_u \) are lower triangular and \( R_{e,n} \) and \( R_{e,u} \) are diagonal matrices. This, in turn, means that the equation for \( S \) in (10.82) will become
\[
SL_n R_{e,n} L_n^T S^T = L_u R_{e,u} L_u^T,
\]  
(10.85)
and the straightforward solution to this problem is then to choose
\[
S = L_u R_{e,u}^{1/2} R_{e,n}^{-1/2} L_n^{-1}.
\]  
(10.86)
10.9.1 Reconstruction via Innovations

The solution above in (10.86) can be interpreted as first running the non-uniformly sampled measurements \( \{y(t_k)\}_{k=1}^{N_t} \) through the Kalman filter for the non-uniformly sampled continuous-time system such that

\[
\begin{cases}
\dot{x}(t_{k+1}) = F(\Delta t_k)\dot{x}(t_k) + K_p(t_k)(y(t_k) - C\dot{x}(t_k)) \\
e_n(t_k, \theta) = y(t_k) - C\dot{x}(t_k).
\end{cases}
\] (10.87)

Thereby one would generate a sequence of innovations \( \{e_n(t_k)\}_{k=1}^{N_t} \). These innovations, would then be scaled in order to produce a new “uniformly sampled set of innovations” \( \{e(kT_s)\}_{k=1}^{N_t} \) such that

\[
e_u(kT_s) = \sqrt{R_{e,u}(T_s)}e_n(t_k)
\] (10.88)

The new innovations could then act as input to the true model when it is put in a discrete-time innovations form

\[
\begin{cases}
\dot{x}((k+1)T_s) = F(T_s)\dot{x}(kT_s) + K_p(T_s)e_u(kT_s) \\
y_r(kT_s) = C\dot{x}(kT_s) + e_u(kT_s).
\end{cases}
\] (10.89)

10.9.2 Spectral Estimation

When the uniformly sampled data is reconstructed, it is possible to compute the discrete-time Fourier transform such that

\[
Y_{d,N_t}(e^{i\omega T_s}) = \frac{1}{\sqrt{N_t}} \sum_{k=0}^{N_t-1} y_r(kT_s)e^{-i\omega kT_s}
\] (10.90)

using the Fast Fourier Transform (FFT). From this we can estimate the discrete-time spectrum

\[
\hat{\Phi}_{d,N_t}(e^{i\omega T_s}) = \left|Y_{d,N_t}(e^{i\omega T_s})\right|^2
\] (10.91)

and can then use the methods of Section 10.5 in order to estimate the continuous-time spectrum such that

\[
\hat{\Phi}_{c,T}(i\omega) = F_{2c,T_s}^T(i\omega)\hat{\Phi}_{d,N_t}(e^{i\omega T_s}).
\] (10.92)

10.9.3 Approximation

The reconstruction procedure described above basically consists of two steps. First the “non-uniform innovations” are computed, then they are scaled and run through the discrete-time version of the model in innovations form. This is illustrated in Figure 10.5

Since we are going to use our spectral estimate for identification we are again approached by the paradox of knowing the true parameters. It is therefore tempting to
resolve this issue by resorting to the same approximations as before, namely replacing the true system by

\[ H_c(s) = \frac{1}{s^{\ell}} \]  

(10.93)

in the reconstruction process. However, due to the separation into stages in Figure 10.5, this approach will prove to be ill-conditioned. The reason for this is that the first step in the process will then basically filter the incoming sequence \( \{y(t_k)\}_{k=1}^{N_t} \) through the non-uniformly sampled discrete-time version of the filter

\[ H_c^{-1}(s) = s^{\ell}. \]  

(10.94)

This would mean that the low frequency components of \( \{y(t_k)\}_{k=1}^{N_t} \) would almost disappear. The innovations would then be scaled and run through the uniformly sampled and discrete-time version of the system

\[ H_c(s) = \frac{1}{s^{\ell}}, \]  

(10.95)

where low frequency components would be almost infinitely amplified. This will lead to the numerical ill-conditioning. In order to introduce some form of regularization, we will therefore approximate the true system by

\[ H_c(s) = \frac{1}{(s - \lambda_b)^{\ell}}, \]  

(10.96)

where \( \lambda_b \) should have some resemblance with the bandwidth of the true system.

### 10.9.4 Numerical Experiments

In the following examples we have used the continuous-time model

\[ y(t) = \frac{1}{s^2 + a_1 s + a_2} e(t). \]  

(10.97)

where \( a_1 = 3 \) and \( a_2 = 2 \). The output \( y(t) \) has been sampled at time instances subject to jitter such that

\[ t_k = kT_s + \delta_k, \quad k = 1, \ldots, N_t - 1 \]
Figure 10.6: Parameter estimates for the model $H_c(s) = \frac{1}{s^2 + 3s + 2}$ versus the sampling time $T_s$. Sampling instances have been generated such that $t_k = kT_s + \delta_k$ where $\delta_k \sim \mathcal{U}(-T_s/3, T_s/3)$. Frequencies up to $2\pi \text{ rad/s}$ have been used.

where

$$\delta_k \sim \mathcal{U}(-\delta_0, \delta_0) \quad k = 1, \ldots, N_t - 1.$$  

For the sake of simplicity, the initial and endpoint time instances have been chosen such that $t_0 = 0$ and $t_{N_t} = N_t T_s$.

In Figure 10.6 we estimated the parameters of the process in (10.97). The sampling rate have been varied from $T_s = 0.1s$ to $T_s = 0.6s$ while the jitter factor is $\delta_0 = T_s/3$. The output have then been observed during $T = 1000s$ and the number of samples is $N = T/T_s$. The parameter estimates presented in the figure are the averages of $N_{MC} = 1000$ Monte-Carlo simulations. Frequencies up to $2\pi \text{ rad/s}$ have been used. In Table 10.3 the corresponding parameters are shown.

## 10.10 Summary

In this chapter, the problem of frequency domain estimation of continuous time series models was approached. Here, a method for the estimation of the continuous-time spectrum from uniformly sampled data was first devised. Later on it could be proved that
Table 10.3: Parameter estimates for the model \( H_c(s) = \frac{1}{s^2 + 3s + 2} \) versus the sampling time \( T_s \). Sampling instances have been generated such that \( t_k = kT_s + \delta_k \) where \( \delta_k \sim U(-T_s/3, T_s/3) \). Frequencies up to \( 2\pi \) rad/s have been used.

<table>
<thead>
<tr>
<th>( T_s )</th>
<th>( a_1 = 3 )</th>
<th>( a_2 = 2 )</th>
<th>( \sigma = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>3.0084</td>
<td>1.9926</td>
<td>1.0046</td>
</tr>
<tr>
<td>0.2</td>
<td>2.9996</td>
<td>1.9852</td>
<td>1.0032</td>
</tr>
<tr>
<td>0.3</td>
<td>3.0105</td>
<td>1.9921</td>
<td>1.0076</td>
</tr>
<tr>
<td>0.4</td>
<td>3.0044</td>
<td>1.9889</td>
<td>1.0069</td>
</tr>
<tr>
<td>0.5</td>
<td>2.9997</td>
<td>1.9927</td>
<td>1.0072</td>
</tr>
<tr>
<td>0.6</td>
<td>2.9353</td>
<td>1.9752</td>
<td>0.9929</td>
</tr>
</tbody>
</table>

This method was actually a way of interpolating the continuous-time covariance function in terms of polynomial spline functions. This was in contrast to the case of input-output models, where instead, the output was interpolated. This lead to a proof which stated that interpolation, which is termed smoothing for time series, will actually always cause a high frequency bias. Finally, it was showed, that the “correct” way to tackle the problem of spectral and parameter estimation was by factorization. For uniformly sampled data this could be accomplished by spectral factorization. For non-uniformly sampled data however, the path to success was via an efficient factorization of the covariance matrix by means of a Kalman filter. The innovations from the output sequence could then be computed, scaled and injected into an equidistantly sampled innovations form. This would require little knowledge of the true underlying system.
Conclusions and Further Research

“He has done his worst but the wound will end him. He is hasped and hooped and hirpling with pain, limping and looped in it. Like a man outlawed, for wickedness, he must await, the mighty judgement of God in majesty”

Beowulf, after killing Grendel (800th century)

This thesis has been propelled by the will to find alternative ways of identifying continuous-time input-output and continuous time series models in the frequency-domain. In Chapter 2 maximum-likelihood methods for doing this were identified. Unfortunately these methods required an accurate estimate of the continuous-time Fourier transform or spectrum of the output while there were only discrete-time measurements of the output available. Producing an estimate of the continuous-time spectrum for uniformly and non-uniformly sampled data then became the objective of Chapter 4 and Chapter 5 for input-output models, and Chapter 10 for time series.

The conclusion of Chapter 5 was that for continuous-time input-output models, interpolation by means of polynomial spline functions was a feasible means of reconstructing the output from discrete measurement. Therefore, in Chapter 6, this idea was applied to continuous time series models, and here the interpolation actually meant interpolating the covariance function. It could then be shown that for each individual frequency, the spectral bias caused by the interpolation would decrease as $T_s \to 0$. If, on the other hand the sampling time $T_s$ was constant, there would always be a high frequency bias present. This apparent paradox was not resolved until Chapter 10.

In Chapter 7 the focus turned to understanding how spectral bias would transform into parameter bias. The conclusion was that the relative bias in the spectrum caused the bias in the parameter estimates. At high frequencies, the value of the spectrum is often very small and this could explain why the high frequency bias in the spectral estimate could cause a large parameter bias. In this chapter, it was also found that it is the relative sensitivity of the spectrum to parameter changes that determines the degree of information
content of a particular frequency. Therefore frequencies with a high relative sensitivity are especially important to use when one seeks accurate parameter estimates.

In Chapter 8 the ideas in Chapter 6 and Chapter 7 were applied to a set of real life data emanating from the ABS system of a car. The purpose was to estimate the tire pressure by monitoring the position of a particular resonance peak found in the wheel speed information. Since there were disturbances outside the frequency interval of interest, these frequencies were not used for estimation in order to avoid severe parameter bias. The model was of the mass-spring-damper kind, and in order to reduce the variance in the parameter estimates and the computational burden, the data set was divided into a number of sub batches. The spectrum was estimated for each individual batch and all the estimates were averaged. The average spectral estimate was then used to find the resonance peak of the model.

While the problem of broad-band disturbances were tackled in Chapter 7 and Chapter 8 the topic of Chapter 9 was the rejection of narrow band disturbances. Here, methods from robust statistics in general, and the infinitesimal approach in particular, were used to make the Whittle likelihood method more robust to outliers in the frequency domain.

Finally, in Chapter 10, the problem of frequency domain estimation of continuous time series models was approached. Here, a method for the estimation of the continuous-time spectrum from uniformly sampled data was first devised. Later on, it could be proved that this method was actually a way of interpolating the continuous-time covariance function in terms of polynomial spline functions. This was in contrast to the case of input-output models, where instead, the output was interpolated. This lead to a proof which stated that interpolation, which is termed smoothing for time series, will actually always cause a high frequency bias. Finally, it was showed, that the “correct” way to tackle the problem of spectral and parameter estimation was by factorization. For uniformly sampled data this could be accomplished by spectral factorization. For non-uniformly sampled data however, the path to success was via an efficient factorization of the covariance matrix by means of a Kalman filter. The innovations from the output sequence could then be computed, scaled and injected into an equidistantly sampled innovations form. This could be accomplished with little knowledge of the true underlying system.

“Vintery, mintery, cutery, corn,
Apple seed and apple thorn;
Wire,briar, limber lock,
Three geese in a flock.
One flew east,
And one flew west,
And one flew over the cuckoo’s nest.”
Bibliography


LDLT-factorization, 24
analytical approach, 3
B-robust, 137
optimal, 137
B-splines, 82
block pulse functions, 51
CARMA models, 95
Cholesky factorization, 24
banded, 25
continuous-time, 4
covariance function, 11
covariance kernel, 10
delta operator, 80
difference operator
forward, 80
Dirac comb function, 57
Dirac delta function, 57
discrete-time, 4
delta operator, 80
difference operator
forward, 80
Dirac comb function, 57
Dirac delta function, 57
discrete-time, 4
empirical distribution, 133
estimator, 22
Euler-Frobenius polynomials, 64
Fast Fourier Transform, 34
FFT, 34
Fourier modulating functions, 46
Fourier series, 51
Fourier transform
continuous-time, 18
TDFT, 19
time-discrete, 19
fundamental B-spline function, 87
Gaussian process, 10
gross-error sensitivity, 137
Hartley modulating functions, 47
impulse response, 19
continuous-time, 12
discrete-time, 13
influence function, 134
innovations, 26
innovations form, 26
input signal, 11
integration methods, 49
Kalman filter, 25
Kronecker delta, 87
Lagrange form, 87
Laguerre polynomials, 51
Laplace transform, 18
inversion formula, 18
Legendre polynomials, 51
likelihood function, 22
linear filter methods, 47
linear integral filter methods, 51
log likelihood function, 22

M-estimator, 134
  of scale, 136
maximum likelihood
  estimator, 22
  score function, 133
mean square error
  MSE, 112
measurement noise, 25
model, 1
modulating functions, 45

Newton’s form, 79
numerical integration methods, 51
orthogonal functions, 50
output signal, 11
periodogram
  continuous-time, 97, 100
  discrete-time, 105
Poisson moment functionals, 48
Poisson summation formula, 19
pole excess, 63, 152
process noise, 25
pulse transfer function, 56, 57
relative bias, 110
relative degree, 63, 152
relative sensitivity, 110
Riccati
  equation, 31
  recursion, 26
Riemann interpolation, 98
roll-off, 64
Runge’s phenomenon, 81
sampling instants, 20
sampling zeros, 64
scale parameter, 136
simplified instrumental variable method, 52
spectral density function, 20
spectrum, 20
state variable filters, 47
state variables, 11
state vector, 11
states, 11
stationary, 11
stationary process, 11
stochastic differential equation, 14
stochastic process, 10
systems approach, 3
Töplitz, 33
tensor-product spline, 100
time series, 14
time-invariant, 11
time-varying, 11
transfer function
  continuous-time, 20
  discrete-time, 20
transition matrix, 12
trapezoidal pulse functions, 51
weak stationarity, 11
z-transform, 18