Identification and robust control methods using ellipsoidal parametric uncertainty descriptions
by Steven Michael Hietpas

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering
Montana State University

Abstract:
This dissertation addresses the areas of identification and robust control of uncertain systems. The first portion of this dissertation concerns system identifica-tion. The focus of system identification is grounded in the signal analysis technique developed by G. R. B. Prony in 1795. This approach uses two separate least-squares solutions, with the first least-squares solution resulting in the eigenvalues of the out-put signal and the second yielding the output residues. Previous system identification methods based on Prony signal analysis apply to systems utilizing a continuous-time input. This research develops a method that allows for a sampled-and-held version of the previously used continuous-time input. The new method results in a more simplified algorithm due to the nature of sampled-and-held input signals. Various examples are provided to show the characteristics of the method. The second portion of this dissertation concerns control of systems which are modeled through system identification techniques that provide models with a parametric uncertainty set. The robust control strategy is viewed from dynamic game theory such that a saddle point solution is desired. The game theory method employed has two players: the first player, the feedback control signal, minimizes an energy cost function; and the second player: the set of uncertain parameters, maximizes an energy cost function. It is shown that the optimal control signal necessarily is derived from Linear Quadratic Regulation (LQR) theory. The maximization is performed on the final cost function from the LQR solution and is constrained by an ellipsoidal parameter bound and an appropriate Lyapunov equation. Iterative algorithms for both continuous-time and discrete-time systems are developed. A sufficient condition for the existence of the maximization portion of the discrete-time algorithm is given. Examples show the fixed-point solution of the algorithm is a saddle point. A closed-loop discrete controlled servo-mechanism problem uses the Optimal Volume Ellipsoid (OVE) algorithm of Man-Fung Cheung (1990) for identification of a model and an ellipsoid uncertainty region. The discrete minimax algorithm is applied to determine the optimal state-feedback gain. The servo example is stable and shows the worst-case tracking for the noisy, uncertain system and is compared to the known system.

IDENTIFICATION AND ROBUST CONTROL METHODS USING ELLIPSOIDAL PARAMETRIC UNCERTAINTY DESCRIPTIONS

by

Steven Michael Hietpas

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Electrical Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 1994

APPROVAL

of a thesis submitted by

Steven Michael Hietpas

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

7/14/94
Date

Donald A. Pierre
Donald A. Pierre
Chairperson, Graduate Committee

Approved for the Major Department

7/14/94
Date

Victor Gerez
Head, Electrical Engineering

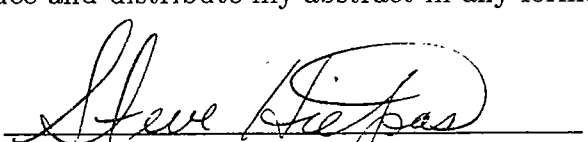
Approved for the College of Graduate Studies

7/23/94
Date

Robert Brown
Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U. S. Copyright Law. Requests for extensive copying or reproduction of this thesis should be referred to University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation for sale in and from microform or electronic format, along with the right to reproduce and distribute my abstract in any format in whole or in part."

Signature 

Date 7-14-94

ACKNOWLEDGEMENTS

This research was made possible through financial support from the National Science Foundation, the Montana Electric Power Research Association and the Engineering Experiment Station of Montana State University. I express my gratitude to Professor Donald Pierre, my principal advisor, for his continuous support, and certainly for his enduring patience. Dr. Pierre has given me a great example of a professional educator and researcher. Through the efforts of Professor Roy Johnson in his continual quest for upgrading existing computer equipment to the best possible – I had it very nice. Dr. Ming Lau of Sandia Laboratories, New Mexico, played a strong part in the robust control portion of this research, and offered much of his time to the review of earlier work and carefully steered me in the correct direction at crucial times. To the rest of my committee, Dr. James Smith, Dr. Hashem Nehrir, Dr. Curtis Vogel and Dr. Richard Swanson, thank you for the time you spent listening to my concerns and offering advice where needed. I extend my appreciation to all of my fellow graduate students, Cesar Amicarella, Fereshteh Fatehi, Don Hammerstrom, Ed Ray, Tia Sharpe and Stott Woods, who have always shown confidence in me when things were not going as well as I desired. A special thanks to Don Hammerstrom, a great engineer and a good friend, for all the helpful discussions, assistance in mastering LaTeX, and for all the exciting racquet ball games. My pastors, Blake Updike and Dave Green, were always available to listen and offer their wisdom and advice when I needed it most. Of course, without a doubt, I thank my parents, Jerry and JoAnne Hietpas, and Jack and Mary Gay Wells for their encouragement. I am fortunate to have wonderful in-laws who also encouraged me throughout this dissertation; thanks

Lance and Marlys Johnson and Collin. I also have the best brother, Tom and sisters, Elise and Tara, who have always been there for me, thanks. My special buddies, Taylor and Aaron, you have always been and always will be an inspiration and joy to me. In addition to these wonderful boys, on June 16, 1994, I was blessed with a beautiful girl, Megan Lynn. Lastly, and most importantly, to my wife Tara, your unfailing love, commitment and confidence in me has kept me going. I owe much of this work to you; thank you so much for being at my side through everything while taking care of the home front. At the beginning of this dissertation, Tara gave me the following quote, Philippians 4:13, New International Version of the Holy Bible; and I close, acknowledging that Christ Jesus is truly the sustainer of all things on this earth.

I can do everything through Him who gives me strength.

TABLE OF CONTENTS

TABLE OF CONTENTS — Continued

TABLE OF CONTENTS — Continued

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This dissertation addresses the areas of identification and robust control of uncertain systems. The first portion of this dissertation concerns system identification. The focus of system identification is grounded in the signal analysis technique developed by G. R. B. Prony in 1795. This approach uses two separate least-squares solutions, with the first least-squares solution resulting in the eigenvalues of the output signal and the second yielding the output residues. Previous system identification methods based on Prony signal analysis apply to systems utilizing a continuous-time input. This research develops a method that allows for a sampled-and-held version of the previously used continuous-time input. The new method results in a more simplified algorithm due to the nature of sampled-and-held input signals. Various examples are provided to show the characteristics of the method. The second portion of this dissertation concerns control of systems which are modeled through system identification techniques that provide models with a parametric uncertainty set. The robust control strategy is viewed from dynamic game theory such that a saddle point solution is desired. The game theory method employed has two players: the first player, the feedback control signal, minimizes an energy cost function; and the second player: the set of uncertain parameters, maximizes an energy cost function. It is shown that the optimal control signal necessarily is derived from Linear Quadratic Regulation (LQR) theory. The maximization is performed on the *final cost function* from the LQR solution and is constrained by an ellipsoidal parameter bound and an appropriate Lyapunov equation. Iterative algorithms for both continuous-time and discrete-time systems are developed. A sufficient condition for the existence of the maximization portion of the discrete-time algorithm is given. Examples show the fixed-point solution of the algorithm is a saddle point. A closed-loop discrete controlled servo-mechanism problem uses the Optimal Volume Ellipsoid (OVE) algorithm of Man-Fung Cheung (1990) for identification of a model and an ellipsoid uncertainty region. The discrete minimax algorithm is applied to determine the optimal state-feedback gain. The servo example is stable and shows the worst-case tracking for the noisy, uncertain system and is compared to the known system.

CHAPTER 1

INTRODUCTION

Very often, a dynamical system is sufficiently complex such that linear or nonlinear sets of equations that adequately describe the system are not available. This is true for many systems, including sophisticated aircraft, chemical processes, large power systems, etc. Under these circumstances it may be necessary to assume a particular parameterized model and apply system identification techniques to arrive at a useful mathematical model of the system to be controlled. The subject of this dissertation addresses system identification and robust control of uncertain systems.

System identification deals with the problem of building mathematical models of dynamical systems based on observed data from the system. A system is an object in which variables of different kinds interact and produce observable signals. Observable signals that emanate from the system are often referred to as outputs. External signals that can be manipulated by the observer and applied to the system are called controlled inputs. Other system input signals are called disturbances, some of which can be directly measured, while others can only be observed through their influence on the output. Systems such as aircraft, space structures, bridges, ships and power grids, to name a few, are considered dynamical, which means that the current outputs depend not only on the current inputs but also on their earlier values. Outputs of dynamical systems whose external stimuli are not observed, e.g., the sound of a human voice generated by the vibration of the vocal chords, are often called *time series*. This term is common in economic applications. Clearly, the list of examples of dynamical systems can be very long and it stretches over many fields of human endeavor.

A major part of the engineering field deals with how to make good designs based on mathematical models. A model describes the relationship among system variables in terms of mathematical expressions like difference or differential equations. Such models are called *mathematical* (or *analytical*) *models.* Mathematical models may be further characterized by a number of adjectives: time continuous or time discrete, lumped or distributed, deterministic or stochastic, linear or nonlinear, single or multi-input/output, etc. When input and output signals from a system that are subjected to data analysis in order to infer a model, the activity is known as *system identification.*

The construction of a model from data involves three basic entities [1]:

I. The data.

II. A set of candidate models.

III. A rule by which candidate models can be assessed using the data.

The data record should be selected to be maximally informative, subject to constraints that may be at hand. Obtaining maximally informative input-output data is affected by which output signals are measured, when and how the signals are measured, and the choice of input signals for external stimuli. The set of candidate models to choose from is certainly the most important step of the identification procedure. It is here where prior knowledge, engineering intuition and insight have to be combined with the formal properties of models. After selection of an appropriate model one is faced with various methods of fitting the model to the data. One approach to the assessment of the model quality is based on how the model performs when it attempts to reproduce the measured data.

The first part of this dissertation is concerned with system identification. The focus of system identification in this research is grounded in the signal analysis tech-

nique developed by G. R. B. Prony [2] in 1795. In this work, Prony proposed a basic signal analysis method to approximate a signal $y(t)$ by a summation of n weighted exponentials, i.e.,

$$y(t) \approx \hat{y}(t) = \sum_{i=1}^{n} B_i e^{\lambda_i t}$$

for continuous time $t \geq 0$, where $B_i \in \mathbb{C}$. The approach uses two separate least-squares solutions, each of order n, with the first least-squares solution resulting in the eigenvalues of the signal $\{\lambda_i\}$, and the second yielding the weighting terms (or output residues) $\{B_i\}$. The basic Prony method is a signal analysis method rather than a system identification method. Extending this signal analysis method of Prony to system identification was possibly first presented by Poggio, et al., [3] in 1978. The system input in [3] is a summation of two real exponential inputs.

Typically the number of eigenvalues (or poles) n is chosen sufficiently large (greater than the 'actual' number of eigenvalues) prior to the formulation of the first least-square problem. The problem of solving the first set of equations is equivalent to a linear prediction (LP) problem. Kumaresan and Tufts [4]-[7] address the effect of noise on the solution to the LP problem in the first least-squares formulation. Their work shows, that from the over-determinization (i.e., large n), the effect of using a truncated singular value decomposition (SVD) essentially increases the signal-to-noise-ratio (SNR) in the data prior to obtaining the solution. The sample variance of the parameter estimates are compared to the Cramér-Rao (CR) bound for the pole damping factors and pole frequencies. A description of the Cramér-Rao bound can be found in [1, pages 183-187] and is briefly described here. Suppose we are interested in estimating a set of parameters θ from a set of data denoted by y^N. Let the estimated set be defined as $\hat{\theta}(y^N)$ and the "true value" of θ be denoted by θ_0. The quality of the estimator can be assessed by its mean-square error matrix:

$$P = \mathbb{E}[\hat{\theta}(y^N) - \theta_0][\hat{\theta}(y^N) - \theta_0]^T$$

where $\mathbb{E}$ evaluates the mean. We may be interested in selecting estimators that make P small. It is then interesting to note that there is a lower limit to the values of P that can be obtained with the various unbiased estimators. Let the probability density function of y^N be given by $f_y(\theta_0; y^N)$ and assumed to be known. The definition of the *Cramér-Rao inequality* is stated as follows:

Definition 1.1 (Cramér-Rao inequality) *Let $\theta(y^N)$ be an estimator of θ such that $\mathbb{E}\theta(y^N) = \theta_0$ and suppose that y^N may take values in a subset of $\mathbb{R}^N$, whose boundary does not depend on θ_0. Then*

$$\mathbb{E}[\hat{\theta}(y^N) - \theta_0][\hat{\theta}(y^N) - \theta_0]^T \geq M^{-1}$$

where

$$M = \mathbb{E}\left[\frac{d}{d\theta}\log f_y(\theta; y^N)\right]\left[\frac{d}{d\theta}\log f_y(\theta; y^N)\right]^T\Bigg|_{\theta=\theta_0} = -\mathbb{E}\frac{d^2}{d\theta^2}\log f_y(\theta; y^N)\Bigg|_{\theta=\theta_0}.$$

The matrix M is known as the *Fisher information matrix* and evaluation requires the knowledge of θ_0, so the exact value of M may not be available to the user. More information on the Cramér-Rao bound when dealing with dynamical systems can be found in [1, pages 186-187].

Further analysis of the effect of the truncated SVD and the variance of the parameter estimates is given in [8] and references therein. Although much of the work in [4]-[7] deals with signal analysis aspects, a transfer function identification example is given in [7]. In this example, however, the input is an impulse signal and after the poles of the system are determined using Prony's method, the zeros of the transfer function (roots of the numerator polynomial) are found using Shank's method [9]. Further investigation of Prony's method for signal modeling of noisy data and model order selection can be found in [10]-[12]. Parameter estimation of exponentially damped sinusoids using higher order statistics (HOS) utilizing Prony's method is explored in [13].

The work in [14]- [20] progressively extends the flexibility of using Prony analysis to perform system identification. In [14] and [15] the input is restricted to be a single square-wave pulse. This approach was improved in [16] where the input is allowed to exhibit a series of step changes, with the input characterized by a single real eigenvalue between step changes. In [17], using an input similar to [16], it is shown that nonzero initial conditions can be identified in addition to the system transfer function. For the results in [14] through [17], the assumed system is not allowed to have a feedthrough term, and the system input is not allowed to have an eigenvalue that is identical to an eigenvalue of the system. Feedthrough terms are included in [18] and [19], however, and in [19] the assumed conditions are relaxed to allow the system to have an eigenvalue that is identical to the one associated with the input. In [20] the class of allowed input signals is expanded: the signals can exhibit jump discontinuities and can be characterized by a finite number of eigenvalues between discontinuities. The advantage of this more general input form is that it can be tailored to excite system modes in frequency ranges of interest. A point in common with several of the later methods mentioned above is that all input signals are required to be time-continuous during each interval of discontinuity, which does not lend well to systems that are microprocessor controlled. This research addresses this point and results in a method that allows for a sampled-and-held version of the general input of [20]. The method developed here provides a more simplified algorithm than that of [20] as a result of the sampled-and-held input.

While there are a large number of identification methods available in the literature, only a few methods of system identification techniques are reviewed in Chapter 2. Following this review of system identification techniques the method based on Prony signal analysis is developed for a discrete controlled system. Once the method is developed, several examples are provided to show various attributes

of the method. Further analysis is provided by comparing the Prony based method to another technique based on a signal energy analysis method that utilizes singular value decomposition. While the method that utilizes SVD for signal decomposition is easier to implement than Prony, there are fewer 'knobs' available for the user to work with in arriving at an optimal model. From an engineering perspective, it is often possible, that the more 'knobs' there are available to use, the better the chance to obtain an acceptable solution. Under the noisiest conditions the Prony method performs slightly better than the SVD method. As a last example, *discrete Prony* identification is applied to a one-machine infinite-bus power system to obtain a low-order model. Linear quadratic control synthesis based on this model is used for transient damping while the system experiences a transmission line fault. The control based on the *discrete Prony* identified model provides significant damping.

The second portion of this dissertation is concerned with control of systems which are modeled through system identification techniques that result in a model with an associated uncertainty set. Control systems are designed so that certain designated signals, such as tracking errors and actuator inputs, do not exceed prespecified levels. Hindering the achievement of this goal are uncertainty about the plant to be controlled and errors in measuring output signals. Uncertainty about the plant is a direct result of the fact that we use mathematical models in representing real physical systems. Furthermore, because sensors can measure signals only to a certain accuracy and because of the natural presence of noise in a system, additional uncertainty through identification occurs in the mathematical modeling process. The issue of uncertainty in the mathematical model leads to the *robust performance problem*, with the goal of achieving specified signal levels in the face of plant uncertainty.

Prior to the 1950s, classical control theory was well established based, in part, on the works of Nyquist [21] and Bode [22]. Following classical methods, control

theory entered the modern control era with emphasis placed on "state-space methods" (a name for all kinds of methodical procedures that has already been used for a long time, e.g., in analytical dynamics, quantum mechanics, theory of stability, in the solution of ordinary differential equations and in other fields). The application of these methods to automatic control was stimulated in the second half of the 1950s mainly by the work of L. S. Pontryagin [23, 24], by the method of dynamic programming as suggested by R. E. Bellman [25] and by the general theory of filtering and control theory elaborated by R. E. Kalman [26]. The recent literature on control systems that is based on state -space methods is very extensive.

Despite the seemingly obvious requirement of bringing plant uncertainty explicitly into control problems, it was only in the early 1980s that 'modern' control researchers re-established the link to the classical work of Nyquist, Bode and others by formulating a tractable mathematical notion of uncertainty in an input-output framework and developing rigorous mathematical techniques to cope with it. The basic technique is to model the plant as belonging to a set $\mathcal{P}$, where such a set can be either *structured* or *unstructured.* The important work of Doyle and Stein [27] established the basic framework for robust control synthesis for systems with unstructured uncertainty.

For an example of a structured set consider the plant model [28]

$$\frac{1}{s^2 + as + 1}.$$

This is a standard second-order transfer function with natural frequency 1 rad/sec. and damping ratio $a/2$ – it could represent, for example, a mass-spring damper or an R-L-C circuit. Suppose that the constant a is known only to the extent that it lies in some interval $[a_{min}, a_{max}]$. Then the plant belongs to the structured set

$$\mathcal{P} = \left\{ \frac{1}{s^2 + as + 1} : \; a_{min} \leq a \leq a_{max} \right\}.$$

For an example of unstructured uncertainty consider the nominal transfer function plant P with the perturbed plant transfer function of the form $\mathcal{P} = \{(1 + \Delta W)P\}$. Here W is a fixed stable transfer function, the weight, and Δ is a variable stable transfer function satisfying $\|\Delta\|_\infty \leq 1$. The uncertainty may be the result of unmodeled dynamics, particularly at high frequency. The infinity-norm on a transfer function is defined as

$$\|G\|_\infty \triangleq \sup_\omega |G(j\omega)|$$

and in simple terms says compute the maximum magnitude of the transfer function G over the entire frequency domain (i.e., the maximum gain of the transfer function).

As elaborated in [28], $|W(j\omega)|$ provides the uncertainty profile and at each frequency point the evaluation of $\mathcal{P}/P$ lies in the disk with center 1, radius $|W|$. Typically, $|W(j\omega)|$ is an increasing function of ω such that uncertainty increases with increasing frequency. The main purpose of Δ is to account for phase uncertainty and to act as a scaling factor on the magnitude of the perturbation (i.e., $|\Delta|$ varies between 0 and 1). This type of uncertainty is referred to as multiplicative uncertainty. More detail on other types of frequency domain (*unstructured*) uncertainty can be found in [29, 30] and references therein.

The focus of the robust control research in this dissertation is directed towards those dynamical systems that involve *structured* uncertainty in the form of a bounded ellipsoidal domain on the parameters of the model. In particular, the uncertainty is a direct result of system identification in the presence of noise.

Most identification methods when applied to a set of input/output data provide a model that is usually assumed to be exact for the particular operating point of interest. Quite often, the model is adequate for control design and implementation. In some instances, however, this approach may not be adequate. Thus, we consider the problem where the identified model is given in the form of a nominal set of

parameters including a range of uncertainty associated with the parameters. This type of uncertainty is sometimes referred to as "parametric uncertainty" and indeed falls under the class of *structured uncertainty.*

Recently, the works of [31]-[33] provide system identification methods where the uncertainty on the identified parameters is given in the form of a bounded ellipsoidal region around a nominal set of parameters. The ellipsoid uncertainty region resulting from the method given in [31] is strictly due to additive noise at the output of the system. The method given in [32] and [33] is due to a combination of including a model of unstructured uncertainty in the parameterized model and of additive noise at the output of the system. These methods of identification that result in a nominal set of parameters bounded by an ellipsoidal uncertainty domain are reviewed in Chapter 3.

A brief discussion on set-membership robust control is provided in Chapter 3. Set-membership robust control is simply a control strategy that assumes the system model is given in the form of a nominal set of parameters and an associated bounded uncertainty region. The idea of set-membership control may first have been introduced by Bertsekas and Rhodes in 1971 [34]. While set-membership is a general term, we consider a method of control where the parameters of uncertainty are known to lie within a bounded ellipsoidal domain. Recent results in set-membership control can be found in [35]-[37].

The robust control strategy is approached from a dynamic game theory point of view such that a saddle point solution is desired. For excellent discussion on differential and dynamic game theory see, for example, [38]-[40]. The saddle point is the result of a minimaximization of an energy (cost) function. The game theory method here has two players. Player one is the feedback control signal that attempts to minimize the cost. Player two is the set of uncertain parameters which attempts

to maximize the cost.

It is shown that the optimal control signal necessarily relates to Linear Quadratic Regulation (LQR). The theory of LQR formed a corner stone of the modern control era of the 1960s. The *final cost function* associated with a given plant and the LQR solution provides the means for an iterative approach to the minimaximization solution. The maximization is performed on the *final cost function* and is constrained by the ellipsoidal parameter bound and an appropriate Lyapunov equation from the LQR solution.

An iterative algorithm for both continuous-time and discrete-time systems is developed in Chapter 4. That the fixed-point solution of the algorithm is a saddle point is illustrated by several examples. A sufficient condition for the existence of the maximization portion of the discrete-time algorithm is given.

A closed-loop discrete controlled servo-mechanism example is also provided. The second order system is assumed to be unknown (i.e., a *black box*). Through an identification technique known as the *Optimal Volume Ellipsoid* (OVE) algorithm [31] (discussed in Chapter 3) a model and ellipsoid uncertainty region is found using noisy input/output data. A Kalman filter is used for state estimation. The discrete minimax algorithm is applied to determine the optimal state-feedback gain and the loop is closed. The servo example is stable and shows the worst-case tracking for the noisy, uncertain system. Results are compared to the case where the system is known (i.e., actual plant parameters are used).

A discussion of the results is provided in Chapter 5 and is followed by a section on possible future research in Chapter 6.

CHAPTER 2

DISCRETE PRONY IDENTIFICATION

In this chapter we extend previous results on transfer function identification using Prony signal analysis methods [20]. Recent work resulted in a generalized Prony identification method that incorporates piecewise continuous system input signals; between points of discontinuity, the input is characterized by input eigenvalues and input residues. We refer to the development of [20] as the *continuous Prony* identification method. Here we extend this work to account for sampled-and-held inputs that arise in digital filtering and control. This new method is referred to as *discrete Prony* identification. Using sampled input and output data, initial condition residues are estimated along with transfer function residues and eigenvalues. The two major phases of the *continuous Prony* method are modified to account for the sampled-and-held input; a major contribution is made to the way in which the residues are fit to corresponding eigenvalues in the second phase.

The organization of this chapter is as follows. First, a review of various *parametric* identification techniques is given as a means to establish the category in which Prony identification falls. Following this review of identification techniques the method of *discrete Prony* identification based on Prony signal analysis is presented. The *discrete Prony* algorithm was implemented using MATLAB, and examples are provided. The chapter ends with some concluding remarks.

Review of System Identification Methods

As a means of motivation for this chapter, a review of several methods of identification

is provided. This review is not meant to be exhaustive as this subject area is covered thoroughly in the literature, (e.g., see [1], [41] and [42]).

We can categorize identification methods in a number of ways. In this chapter we consider, among others, the attributes associated with model order selection and whether the method may be used on-line (real-time) or off-line (batch mode). Occasionally, exact knowledge of the system order is known, and identification methods that assume a priori system order should be used. For those situations where exact knowledge of the system order is not known, methods which make no a priori assumption on system order must be used. These methods start over-parameterized and must employ a stage within the identification process where model order selection (reduction) can be accomplished. Methods are available that optimally select the model order of the over-parameterized system. These order selection methods are well suited for on-line identification since they can often be coded. For off-line methods, optimal procedures for model order selection can be used but are not always necessary. Some applications, e.g., those that apply adaptive control require identification methods that can be implemented on-line. The nature of the system and the type of control strategy implemented often will dictate the type of identification to be used; thus, there are always trade-offs between identification schemes.

Identification methods can also be categorized further by considering the effects of noise in the system or the effects of modeled uncertainty. Those models that do not take into account noise are considered deterministic while those that do account for noise are considered stochastic models. Methods of identification that utilize models which include modeled uncertainty are referred to as robust methods of identification.

In the following subsections, we consider two popular models used in the control community and review some of the methods of analysis used to arrive at realizations of the model from the measured input and output data. Following this

review the *continuous Prony* method is introduced and the *discrete Prony* method is developed.

Autoregressive Models (ARX)

Consider the autoregressive model with exogenous input (ARX). The exogenous input is $u : \mathbb{N} \mapsto \mathbb{R}$. Assume that n_a parameters are associated with the AR part of the model, and n_b parameters are associated with the exogenous input. The ARX model is single-input/single-output (SISO) and can be given as

$$\begin{aligned} y(k) &= -\sum_{i=1}^{n_a} a_i y(k-i) + \sum_{j=1}^{n_b} b_j u(k-j) \\ &= \theta^T \phi(k) \end{aligned} \tag{2.1}$$

where the parameter vector to be estimated is

$$\theta^T = [a_1, \cdots, a_{n_a}, b_1, \cdots, b_{n_b}]$$

and the regression vector containing the past input and output values is given by

$$\phi(k) = [-y(k-1), \cdots, -y(k-n_a), u(k-1), \cdots, u(k-n_b)]^T.$$

The model order is assumed to be known a priori. Equation (2.1) is considered deterministic since there is no inclusion of noise in the model. The stochastic version can be obtained by introducing an additional noise term, say $\nu(k)$ to the right hand side of (2.1) as seen below.

$$\begin{aligned} y(k) &= -\sum_{i=1}^{n_a} a_i y(k-i) + \sum_{j=1}^{n_b} b_j u(k-j) + \nu(k) \\ &= \theta^T \phi(k) + \nu(k) \end{aligned} \tag{2.2}$$

where θ^T and $\phi(k)$ are defined above.

An expanded version of the ARX model is the autoregressive, moving average model with exogenous input (ARMAX). The moving average part comes from a third

set of parameters operating on the noise terms $\nu(k)$, and is given in (2.3).

$$\begin{aligned} y(k) &= -\sum_{i=1}^{n_a} a_i y(k-i) + \sum_{j=1}^{n_b} b_j u(k-j) + \sum_{l=1}^{n_c} c_l \nu(k-l) + \nu(k) \\ &= \theta^T \phi(k) + \nu(k) \end{aligned} \tag{2.3}$$

where the parameter vector to be estimated is

$$\theta^T = [a_1, \cdots, a_{n_a}, b_1, \cdots, b_{n_b}, c_1, \cdots, c_{n_c}]$$

and the regression vector containing the past input, output and noise values is given by

$$\phi(k) = [-y(k-1), \cdots, -y(k-n_a), u(k-1), \cdots, u(k-n_b), \nu(k-1), \cdots, \nu(k-n_c)]^T.$$

Here the size of the moving average vector, n_c, must be given a priori. Also, past noise terms are seldom known exactly, in which case estimates of them are used in place of actual noise values in the regression vector. It should be noted that the ARMAX model has become a standard tool for both system description and control design.

Least Squares Error Method After the model is established the method of determining the estimate of θ can be formulated. Consider the *least-squares* method. With (2.1) the prediction error becomes

$$\epsilon(k, \theta) = y(k) - \phi^T(k)\theta. \tag{2.4}$$

The *least-squares criterion* for (2.4) is

$$V_N(\theta) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{2} [y(k) - \phi^T(k)\theta]^2. \tag{2.5}$$

The *least-squares criterion* is a quadratic function in θ and can be minimized analytically, which gives, provided the indicated inverse exists,

$$\hat{\theta}_N^{LS} = \arg\min_{\theta \in \mathbb{R}} \left[\frac{1}{N} \sum_{k=1}^{N} \phi(k)\phi^T(k) \right]^{-1} \frac{1}{N} \sum_{k=1}^{N} \phi(k) y(k) \tag{2.6}$$

the *least-squares estimate* (LSE). From this development, one could introduce different weights on different measurements in the *least-squares criterion* resulting in the *weighted least-squares criterion.*

This method has not appealed to any statistical arguments for the estimation of θ. In fact, the framework of fitting these models to the input/output data makes sense regardless of the stochastic setting. Methods such as *maximum likelihood* are considered statistical parameter estimation techniques. Recursive forms of the ARX/LS and ARMAX/LS methods are available for on-line identification and adaptive control [41]. For a review of these and other methods see [1] and references therein.

State-Space Models

In the state-space form the relationship between the input and output signals is written as a system of first-order differential or difference equations using an auxiliary state vector $\mathbf{x}(t)$. This description of linear dynamical systems became an increasingly dominating approach after Kalman's work in 1960 [43] on prediction and linear quadratic control. This model is especially useful in that insights into physical mechanisms of the system can usually more easily be incorporated into state-space models than into models of the ARX form.

Consider the multi-input multi-output (MIMO) state-space model given by

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t), \tag{2.7}$$

where $A \in \mathbb{R}^{n_a \times n_a}$, $B \in \mathbb{R}^{n_a \times m_i}$, the number of inputs is given by m_i and control $\mathbf{u} : \mathbb{R}^+ \mapsto \mathbb{R}^{m_i \times 1}$, and state $\mathbf{x} : \mathbb{R}^+ \mapsto \mathbb{R}^{n_a \times 1}$. The output of the system is given by

$$\mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{u}(t), \tag{2.8}$$

where $C \in \mathbb{R}^{m_o \times n_a}$ and $D \in \mathbb{R}^{m_o \times m_i}$; the number of outputs is given by m_o.

Consider the single-input single-output (SISO) version of (2.7) and (2.8) with $m_i = m_o = 1$. Let the Laplace transform of $u(t)$ and $y(t)$ be given as $U(s)$ and $Y(s)$, where $s \in \mathbb{C}$. The transfer function of the input/output model from $U(s)$ to $Y(s)$ is given in the Laplace variable s as

$$G(s) = \frac{b_{n_b+1}s^{n_b} + b_{n_b}s^{n_b-1} + b_{n_b-1}s^{n_b-2} + \cdots + b_2 s + b_1}{s^{n_a} + a_{n_a}s^{n_a-1} + \cdots + a_2 s + a_1} \tag{2.9}$$

where $n_b \leq n_a$. The poles of (2.9) are defined as the roots of the denominator polynomial which are equivalent to the eigenvalues of A from (2.7). For multi-input multi-output (MIMO) transfer functions and their characteristics, see [44, 45] and references therein.

A discretized version of the continuous time system of equations of (2.7) and (2.8) is given by (2.10) and (2.11) [46]. This discretized version is obtained when the input $\mathbf{u}(t)$ is sampled-and-held periodically for a time period T.

$$\mathbf{x}((k+1)T) = F(T)\mathbf{x}(kT) + H(T)\mathbf{u}(kT) \tag{2.10}$$

and

$$\mathbf{y}(kT) = C\mathbf{x}(kT) + D\mathbf{u}(kT). \tag{2.11}$$

Here we have

$$F(T) = e^{AT} \tag{2.12}$$

and

$$H(T) = \left(\int_0^T e^{A\lambda} d\lambda \right) B. \tag{2.13}$$

If matrix A is nonsingular, $H(T)$ given by (2.13) can be simplified to

$$H(T) = A^{-1}(e^{AT} - I)B.$$

Note that the values of C and D of the continuous time state-space model are not affected by the sampled-and-held input signal, $\mathbf{u}(kT)$.

There have been various approaches to system identification using the state-space model. Most of the methods involve extensive use of singular value decomposition (SVD). The derivations of these procedures are relatively lengthy and only a few are described below. It is noted that each of the identification approaches is based on some distinctive method of analysis, as can be seen in the following sections.

Principal Component Analysis Method Consider the approach based on the pulse response matrix of the strictly proper ($D = 0$) discrete system of (2.10) and (2.11). The response $\mathbf{y}(k)$ of the system is given by

$$\mathbf{y}(k) = CF^k\mathbf{x}(0) + \sum_{i=1}^{k} CF^{i-1}H\mathbf{u}(k-i), k \geq 1. \tag{2.14}$$

The pulse response matrix (also termed Markov parameters) is given by

$$\mathbf{y}_M(k) = CF^{k-1}H, \ k = 1, 2, \ldots \tag{2.15}$$

The system Hankel matrix can be constructed from these Markov parameters as follows

$$\Gamma_{rs}(k-1) = \begin{bmatrix} \mathbf{y}_M(k) & \mathbf{y}_M(k+t_1) & \cdots & \mathbf{y}_M(k+t_{s-1}) \\ \mathbf{y}_M(j_1+k) & \mathbf{y}_M(j_1+k+t_1) & \cdots & \mathbf{y}_M(j_1+k+t_{s-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_M(j_{r-1}+k) & \mathbf{y}_M(j_{r-1}+k+t_1) & \cdots & \mathbf{y}_M(j_{r-1}+k+t_{s-1}) \end{bmatrix}$$

where $j_i (i = 1, \ldots, r-1)$ and $t_i (i = 1, \ldots, s-1)$ are arbitrary integers which can be adjusted to include (or eliminate) specific observations in the Hankel matrix. In [47] a realization (identification) method is developed which provides a full signal model in a single stage, from a single SVD of the system Hankel matrix. Let us suppose for the sake of simplicity that $\Gamma_{rs}(0)$ is a square matrix of order $N = m_o r = m_i s$, where m_o and m_i are defined in (2.7) and (2.8). The SVD then consists of finding an $N \times N$ orthonormal matrix P_N and an $N \times N$ orthonormal matrix Q_N so that

$$\Gamma_{rs}(0) = P_N \Lambda_N Q_N^T$$

with $\Lambda_N = diag(\lambda_1, \lambda_2, \ldots, \lambda_N)$ and $\{\lambda_i\} : \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_\eta \geq \epsilon \geq \lambda_{\eta+1} \geq \cdots \geq \lambda_N \geq 0$. Based on this decomposition, the rank η of $\Gamma_{rs}(0)$ is found as the index of the first nil singular value (for noisy matrices and the relative sizes of the singular values, the nil value is subject to interpretation and is here shown as ϵ).

If the order is chosen as η, the realization becomes

$$\begin{aligned} F &= \Lambda_\eta^{-1/2} P_\eta^T \Gamma_{rs}(1) Q_\eta \Lambda_\eta^{-1/2} \\ H &= \Lambda_\eta^{1/2} Q_\eta^T E_{m_i} \\ C &= E_{m_o}^T P_\eta \Lambda_\eta^{1/2} \end{aligned}$$

where $E_{m_i}^T = [I_{m_i}, \odot]$ and $E_{m_o}^T = [I_{m_o}, \odot]$. The terms I_{m_i} and I_{m_o} are identity matrices and the term $\odot$ is the null matrix of appropriate dimension. The construction of P_η and Q_η are from P_N and Q_N, respectively; further details can be found in [47].

The method may be considered to be a two-stage algorithm since the model order must be determined prior to calculating the final model. However, once the model order is chosen, all parameters are calculated in one step; thus, the algorithm is, in truth, a single stage algorithm. This approach to system identification, although simple, is not realistic for systems with significant amounts of noise. The model order selection, which is based on the singular values of the Hankel matrix, becomes unreliable under noisy conditions. Also the type of input used for probing is limited to that of a pulse. However, the models to be realized can be SISO, MIMO, or SIMO and are all minimal, which is an advantage over ARX methods.

Canonical Variate Analysis Methods The fundamental concept in the canonical variate analysis (CVA) approach is the *past* and *future* of the process [48], and is based on statistical principles. Associated with each time k is a past vector $\mathbf{p}_k$ consisting of the past outputs and inputs occurring prior to time k as well as a future

vector $\mathbf{f}_k$ consisting of outputs at time k or later,

$$\begin{aligned} \mathbf{p}_k &= \left[\mathbf{y}^T(k-1), \mathbf{y}^T(k-2), \ldots, \mathbf{u}^T(k-1), \mathbf{u}^T(k-2), \ldots\right]^T, \\ \mathbf{f}_k &= \left[\mathbf{y}^T(k), \mathbf{y}^T(k+1), \ldots\right]^T. \end{aligned} \tag{2.16}$$

The vector processes $\mathbf{y}(k)$ and $\mathbf{u}(k)$ are assumed to be jointly stationary and the covariance matrices among the past $\mathbf{p}_k$ and the future $\mathbf{f}_k$ are denoted as Σ_{ff}, Σ_{pp}, and Σ_{fp}. Now suppose that for a specified state order η, we wish to determine η linear combinations of the past $\mathbf{p}_k$ which allow the optimal prediction of the future $\mathbf{f}_k$. The set of η linear combinations of the past $\mathbf{p}_k$ is denoted as an $\eta \times 1$ vector $\mathbf{m}_k$ and considered as η-order memory of the past. For any selection of the memory $\mathbf{m}_k$, the optimal linear prediction $\hat{\mathbf{f}}_k$ is the function [48]

$$\hat{\mathbf{f}}_k(\mathbf{m}_k) = \Sigma_{fm}\Sigma_{mm}^{-1}\mathbf{m}_k \tag{2.17}$$

of the reduced order memory $\mathbf{m}_k$. The prediction error is measured using the quadratic weighting

$$\mathbb{E}\{\|\mathbf{f}_k - \hat{\mathbf{f}}_k\|_{\Lambda^\dagger}^2\} = \mathbb{E}\{(\mathbf{f}_k - \hat{\mathbf{f}}_k)^T\Lambda^\dagger(\mathbf{f}_k - \hat{\mathbf{f}}_k)\} \tag{2.18}$$

where $\mathbb{E}$ is the expectation operation and Λ is an arbitrary positive semidefinite symmetric matrix so that the pseudo-inverse $\Lambda^\dagger$ is an arbitrary quadratic weighting that is possibly singular. For system identification, the use of the weighting matrix $\Lambda = \Sigma_{ff}$ results in a near maximum likelihood system identification procedure.

In terms of these quantities, the following problem is stated. For a given order η, determine an optimal η-order memory

$$\mathbf{m}_k = J_\eta \mathbf{p}_k \tag{2.19}$$

by choosing the η rows of J_η such that the optimal linear predictor $\hat{\mathbf{f}}_k(\mathbf{m}_k)$ based on $\mathbf{m}_k$ minimizes the prediction error (2.18). The method to arrive at J_η is the subject of

the following discussion. The vector $\mathbf{m}_k$ is intentionally called 'memory' rather than 'state'. A given selection of memory $\mathbf{m}_k$ normally will not correspond to the state of any well defined η-order Markov process. For the system identification problem, this is not a problem since many orders of η will be considered and the one giving the best prediction will be chosen as the optimal order.

Theorem 2.1 (Larimore [48]) Consider the problem of choosing η linear combinations $\mathbf{m}_k = J_\eta \mathbf{p}_k$ of $\mathbf{p}_k$ for predicting $\mathbf{f}_k$, such that (2.18) is minimized where Σ_{pp}, and Λ are possibly singular positive semidefinite symmetric matrices with ranks m and n respectfully. Then the existence and uniqueness of solutions are completely characterized by the (Σ_{pp}, Λ)-generalized singular value decomposition which guarantees the existence of matrices J, L, and generalized singular values $\gamma_1, \cdots, \gamma_q$ such that

$$\begin{aligned} J\Sigma_{pp}J^T &= I_m, \quad L\Lambda L^T = I_n, \\ J\Sigma_{pf}L^T &= diag(\gamma_1 \geq \cdots \geq \gamma_q > 0, \cdots, 0). \end{aligned} \tag{2.20}$$

The solution is given by choosing the rows of J_η if the η-th singular value satisfies $\gamma_\eta > \gamma_{\eta+1}$. If there are r repeated singular values equal to γ_η, then there is an arbitrary selection from among the corresponding singular vectors, i.e. rows of J. Thus, the condition that $\gamma_\eta > \gamma_{\eta+1}$ merely defines a decision point for model order selection. The model order η is to some degree arbitrary, however, the optimal state order can be determined by use of the Akaike information criterion. The minimum value is

$$\min_{rank(J_\eta \Sigma_{pp} J_\eta^T)=\eta} \mathbb{E}\{\|\mathbf{f}_k - \hat{\mathbf{f}}_k\|^2_{\Lambda^\dagger}\} = tr\Lambda^\dagger \Sigma_{ff} - \gamma_1^2 - \cdots - \gamma_\eta^2. \tag{2.21}$$

The vectors of variables $\mathbf{c} = J\mathbf{p}_k$ and $\mathbf{d} = L\mathbf{f}_k$ are canonical variables and (2.20) characterizes a *canonical variate analysis*. Now consider the system described by (2.10) and (2.11) where $\mathbf{u}(k)$ and $\mathbf{y}(k)$ are random processes. In [48] models of

noise are included; here we have omitted them for ease of presentation. We wish to model and predict the future of $\mathbf{y}(k)$ by an η-state $\mathbf{x}(k)$. The particular multivariate regression equations are expressed in terms of covariances, denoted by Σ, among various vectors as

$$\begin{bmatrix} F & H \\ C & D \end{bmatrix} = \Sigma\left[\begin{pmatrix} \mathbf{m}(k+1) \\ \mathbf{y}(k) \end{pmatrix}, \begin{pmatrix} \mathbf{m}(k) \\ \mathbf{u}(k) \end{pmatrix}\right] \Sigma^{-1} \left[\begin{pmatrix} \mathbf{m}(k) \\ \mathbf{u}(k) \end{pmatrix}, \begin{pmatrix} \mathbf{m}(k) \\ \mathbf{u}(k) \end{pmatrix}\right]. \quad (2.22)$$

Explicit computation is obtained by the substitution of $\mathbf{m}_k = J_\eta \mathbf{p}_k$.

To decide on the model state order or model structure, recent developments based upon entropy or information measures are used. Such methods were originally developed by Akaike [49] and involve the use of the Akaike Information Criterion (AIC) for deciding the appropriate order of a statistical model. In simplest terms the AIC is a procedure that minimizes negative entropy. The AIC for each order η is defined by

$$AIC(\eta) = -2 \log p(Y^N, U^N; \hat{\theta}_\eta) + 2M_\eta \quad (2.23)$$

where p is the likelihood function based on the observations (Y^N, U^N) at N time points, and where $\hat{\theta}_\eta$ is the maximum likelihood parameter estimate using an η-order model with M_η parameters. For more information on likelihood functions see [1, pages 181-190]. The model order η is chosen corresponding to the minimum value of the $AIC(\eta)$. The number of parameters in (2.10) and (2.11) is

$$M_\eta = (2\eta + m_i)m_o + m_i m_o + m_o(m_o + 1)/2$$

where m_i and m_o are the number of input and output variables, respectively. This result is developed by considering the size of the equivalent class of state space models having the same input/output and noise characteristics [50]. Thus the number of functionally independent parameters in a state-space model is far less than the number of elements in the various state-space matrices. Effectively, the AIC imposes a statistical penalty on the order (parameter size) of the model.

State Variable Analysis Methods An approach that is in some ways similar to CVA, referred to as State Variable Analysis (SVA), attempts to estimate the state of the system, and from the state, produce a realization of the state-space model. A particular formulation of an SVA algorithm is given in [51] and is briefly discussed as follows. First, a time-series state-vector basis for the dynamical system is realized as the intersection of the row spaces of two block Hankel matrices constructed with measured input/output (I/O) data. This step is done by repeated use of SVD. The second step solves a set of linear equations for F, H and C. The set of linear equations for step two is constructed from the basis vector estimated in step one.

The algorithm assumes that the data are consistent with the discrete-time state-space model of (2.10) and (2.11). The algorithm utilizes a Hankel matrix Υ,

$$\Upsilon = \begin{bmatrix} \Upsilon_1 \\ \Upsilon_2 \end{bmatrix} \tag{2.24}$$

where

$$\Upsilon_1 = \begin{bmatrix} \mathbf{u}(k) & \mathbf{u}(k+1) & \cdots & \mathbf{u}(k+j-1) \\ \mathbf{y}(k) & \mathbf{y}(k+1) & \cdots & \mathbf{y}(k+j-1) \\ \mathbf{u}(k+1) & \mathbf{u}(k+2) & \cdots & \mathbf{u}(k+j) \\ \mathbf{y}(k+1) & \mathbf{y}(k+2) & \cdots & \mathbf{y}(k+j) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}(k+i-1) & \mathbf{u}(k+i) & \cdots & \mathbf{u}(k+j+i-2) \\ \mathbf{y}(k+i-1) & \mathbf{y}(k+i) & \cdots & \mathbf{y}(k+j+i-2) \end{bmatrix}$$

and

$$\Upsilon_2 = \begin{bmatrix} \mathbf{u}(k+i) & \mathbf{u}(k+i+1) & \cdots & \mathbf{u}(k+j+i-1) \\ \mathbf{y}(k+i) & \mathbf{y}(k+i+1) & \cdots & \mathbf{y}(k+j+i-1) \\ \mathbf{u}(k+i+1) & \mathbf{u}(k+i+2) & \cdots & \mathbf{u}(k+j+i) \\ \mathbf{y}(k+i+1) & \mathbf{y}(k+i+2) & \cdots & \mathbf{y}(k+j+i) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}(k+2i-1) & \mathbf{u}(k+2i) & \cdots & \mathbf{u}(k+j+2i-2) \\ \mathbf{y}(k+2i-1) & \mathbf{y}(k+2i) & \cdots & \mathbf{y}(k+j+2i-2) \end{bmatrix}.$$

Application of singular value decomposition is used to estimate a basis, X, from which the time-series state vector can be obtained. It is shown in [51] that

$$span_{row}(X) = span_{row}(\Upsilon_1) \cap (\Upsilon_2),$$

where the basis, X, is given as

$$X = [\mathbf{x}_{k+i}\ \mathbf{x}_{k+i+1}\ \cdots\ \mathbf{x}_{k+i+j-2}].$$

Note, we have replaced $\mathbf{x}(k)$ with subscript notation $\mathbf{x}_k$. As stated in [51], i and j should be chosen sufficiently large (insuring sufficient information about the system), and $j \gg i$ so that computational load and noise sensitivity are reduced. With the state-vector basis determined the order of the system must be decided before the state-space matrices of (2.10) and (2.11) can be estimated. If the system is completely noise free the singular values of the SVD will reveal the exact order of the system. If noise is present one must observe the relative magnitudes of the singular values and determine an appropriate order for the model. This problem of the effect of noise on the singular values of the system Hankel matrix has been addressed in [52]. In [52] it is shown that small perturbations in the input and output matrices H and C can reveal those modes that are associated with the actual system and those modes associated with noise. Also, a selection procedure similar to the AIC may be employed for model order reduction.

With the order and resulting basis X established, the following set of linear equations may be solved in a least-squares sense for F, H, C and D.

$$\begin{bmatrix} X_{k+i+1}, \cdots, X_{k+j+i-1} \\ \mathbf{y}_{k+i}, \cdots, \mathbf{y}_{k+j+i-2} \end{bmatrix} = \begin{bmatrix} F & H \\ C & D \end{bmatrix} \begin{bmatrix} X_{k+i}, \cdots, X_{k+j+i-2} \\ \mathbf{u}_{k+i}, \cdots, \mathbf{u}_{k+j+i-2} \end{bmatrix}. \tag{2.25}$$

This method can be adapted for on-line identification.

This review has described some of the modern techniques of identification being used and it revealed issues that one must be concerned with when choosing a method. Each of the methods was presented in its simplest form. Each method has been extended to account for various types of noise, see, e.g., [53, 54]. Following is the development of the *discrete Prony* method. At the end of this chapter *discrete Prony* will be compared to the SVA method just described.

Prony Signal Identification for Discrete Control Systems

Introduction: Discrete Prony

The Prony method for system identification is based on a signal analysis method for approximating a signal with a weighted sum of exponentials. A complete survey of the development and uses of Prony methods in system identification is given in [20]. Since the original work of Prony [2], much has been accomplished in the way of system identification based on this analysis technique.

Most recently [20] has provided a general Prony approach for system identification where the probing input is piecewise continuous; between points of discontinuity, the input is characterized by input eigenvalues and input residues. The procedure is given in essentially two steps: a) modal (eigenvalue) identification and b) residue fitting. Within each of these steps there are additional steps that can be taken to aid in the final transfer function model optimal selection. The method results in estimates of initial condition residues, transfer-function residues, and system eigenvalues. The development in [20] is set in the continuous time domain and leads to a relatively complicated procedure in step b.

Unlike the method in [20] the development here is for those systems that are implemented using sampled-and-held inputs. This development leads to a straightforward procedure for step b.

System Characterization

System Model The system model shown in Figure 1 is a SISO system. Through partial fraction expansion the system model from (2.9) has the Laplace transform function represented in standard parallel form:

$$\hat{G}(s) = R_0 + \sum_{j=1}^{\eta} \frac{R_j}{s - \xi_j}. \tag{2.26}$$

In Figure 1 the initial condition terms are included explicitly in the summation preceding the output $\hat{y}(t)$, so that the input $u(t)$ can be taken as 0 for $t < 0$. The input $u(t)$ is sampled-and-held prior to its application to both the model and the system. The constant sample period is T. Although a zero-order-hold (ZOH) circuit is indicated in Figure 1, the approach that follows can be modified to account for other hold circuits. The ξ_j's are the eigenvalues of the system model, R_0 is a feed-through gain, R_1 through R_η are the model residues, and E_1 through E_η are initial condition residues. The ξ_j's are assumed to be distinct and can occur in complex conjugate pairs. Residues corresponding to complex conjugate eigenvalues also occur in complex conjugate pairs. The objective of the identification procedure is to find values of ξ_j's, R_j's, E_j's and η so that the model's output $\hat{y}(t)$ is close as possible, in some appropriate sense, to the actual system output $y(t)$.

If there is a dc offset in the output, then some ξ_j will equal zero with the associated $E_j \neq 0$; and if there is no eigenvalue with value zero in $G(s)$, then the associated $R_j \cong 0$.

General System Input For the identification method that is developed in this thesis, the input $u(t)$ for $t \geq 0$ is assumed to be of the general form:

$$u(t) = \sum_{k=1}^{q} \sum_{l=1}^{m_k} c_{k,l}(e^{\mu_{k,l}(t-\tau_{k-1})}) \left[u_s(t - \tau_{k-1}) - u_s(t - \tau_k)\right], \tag{2.27}$$

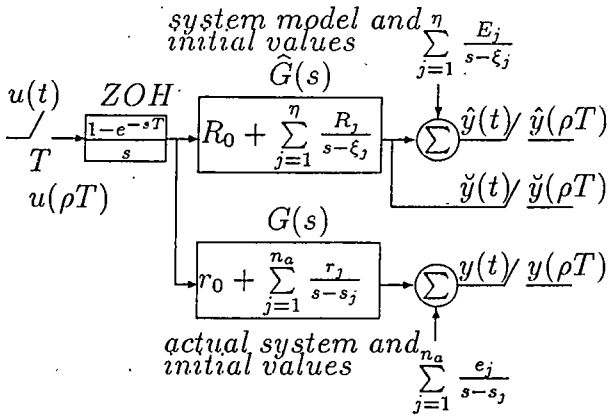


Figure 1. System and actual model in parallel form.

which is discontinuous at a finite number $q+1$ points in time. Here the step function is represented by $u_s(\cdot)$. The k^{th} input time interval is characterized by $t \in [\tau_{k-1}, \tau_k)$ where $\tau_0 = 0$ without loss of generality. A total of $q+1$ time intervals exist for $t \geq 0$, where the $(q+1)^{th}$ time interval corresponds to $t \geq \tau_q$ in which $u(t) = 0$.

For each input interval, $(\tau_k - \tau_{k-1})/T$ is assumed to be an integer value. Let the integer M_k be defined by

$$M_k \triangleq \frac{\tau_k}{T}, \qquad k = 0, 1, \ldots, q.$$

The input time intervals can be characterized in terms of the M_k's: letting $t = \rho T, \rho = 0, 1, \ldots$, the input time interval is characterized by integer time $\rho \in [M_{k-1}, M_k)$ for $k = 1, 2, \ldots, q$. During the k^{th} time interval, the input signal is characterized by the set of input eigenvalues $\{\mu_{k,l}\}_1^{m_k}$ with corresponding amplitude set $\{c_{k,l}\}_1^{m_k}$. The input eigenvalues can be selected to excite specific frequency ranges of interest. All values of the input signal parameters are assumed known.

Output Response Using $u(t)$ of (2.27) as the input in Figure 1, the model output $\hat{y}(t)$ is given by

$$\hat{y}(t) = \sum_{j=1}^{\eta} E_j e^{\xi_j t} + \breve{y}(t), \tag{2.28}$$

where $\breve{y}(t)$ is shown in Figure 1.

In order to employ the method of Prony in modal identification a thorough understanding of the sampled output signal in terms of the input and system modes is necessary. The analysis can be quite tedious because the system is being driven by a zero-order hold circuit and the input signal is piece-wise continuous. Therefore, for purposes of explanation, consider a first-order model

$$\widehat{G}(s) = \frac{R}{s-\xi},$$

with initial conditions IC,

$$IC(s) = \frac{E}{s-\xi}.$$

Allow the input to be given by

$$u(t) = e^{\mu(t-\tau_{k-1})}[u_s(t-\tau_{k-1}) - u_s(t-\tau_k)]. \tag{2.29}$$

The following derivations make use of various properties and theorems of the z-transform and may be found in [46, pages 51–57, 86 and 152]. Let the z-transform operator be represented by $\mathcal{Z}\{\cdot\}$; then the z-transform of the input of (2.29) is

$$U(z) = \mathcal{Z}\{u(t)\} = z^{-M_{k-1}}[1 - z^{-\triangle_M}\gamma^{\triangle_M}]\frac{z}{z-\gamma}, \tag{2.30}$$

where

$$\triangle_M \triangleq M_k - M_{k-1} \text{ and } \gamma \triangleq e^{\mu T}.$$

The z-transform of IC(s) is

$$IC(z) = \mathcal{Z}\left\{\frac{E}{s-\xi}\right\} = E\frac{z}{z-\lambda}, \tag{2.31}$$

where

$$\lambda \triangleq e^{\xi T}.$$

Next, consider Figure 1 and (2.30) and (2.31); then the z-transform of $\hat{y}(t)$ is given by

$$\begin{aligned} \hat{Y}(z) &= IC(z) + U(z)(1-z^{-1})\mathcal{Z}\left\{\frac{R}{s(s-\xi)}\right\} \\ &= E\frac{z}{z-\lambda} + z^{-M_{k-1}}[1 - z^{-\triangle_M}\gamma^{\triangle_M}]\frac{\frac{R}{\xi}(\lambda-1)z}{(z-\lambda)(z-\gamma)}. \end{aligned} \tag{2.32}$$

There are three cases to consider when analyzing the output signal.

Case 1) $\rho:\ \rho T < M_{k-1}T$

It can be seen from (2.32) that the only part of $\hat{Y}(z)$ that has nonzero inverse when $\rho < M_{k-1}$ is

$$E\frac{z}{z-\lambda},$$

and thus

$$\hat{y}(\rho T) = E\lambda^{\rho}, \quad \rho < M_{k-1}.$$

Case 2) $\rho:\ M_{k-1}T \leq \rho T < M_k T$

In this case, those parts of $\hat{Y}(z)$ having nonzero inverse terms over the time interval are

$$E\frac{z}{z-\lambda} + z^{-M_{k-1}}\frac{\frac{R}{\xi}(\lambda-1)z}{(z-\lambda)(z-\gamma)}$$

and thus

$$\hat{y}(\rho T) = E\lambda^{\rho} + \frac{R(\lambda-1)}{\xi(\lambda-\gamma)}[\lambda^{\rho-M_{k-1}} - \gamma^{\rho-M_{k-1}}], \quad M_{k-1} \leq \rho < M_k. \tag{2.33}$$

Case 3) $\rho:\ \rho T \geq M_k T$

It can be seen from (2.32) and (2.33) that $\hat{y}(\rho T)$ takes on the following form

$$\hat{y}(\rho T) = E\lambda^{\rho} + \frac{R(\lambda-1)}{\xi(\lambda-\gamma)}\left[\lambda^{\rho-M_{k-1}} - \gamma^{\triangle_M}\lambda^{\rho-M_k}\right], \quad \rho \geq M_k.$$

Hence, it can be seen that while the input is off the output signal contains only those modes associated with the system, whereas, while the input is on, as in case 2, the output signal is comprised of input and system modes. Further note that in case 3 the modal content of the output signal is comprised of linear combinations of the system mode, λ, where the constant multipliers of this system mode are a function of the past input mode, γ.

In general it can be said that during a given interval k the modal content of the output signal consists of linear combinations of both the input and system modes, and while the input is off, the modal content of the output signal consists of linear combinations of the system modes.

Thus, for any interval defined by $[\tau_{k-1}, \tau_k)$ such that the probing input is on, (i.e. for $k = 1, 2, \ldots, q$), the part of $\hat{Y}(z)$ that has a nonzero inverse can be characterized by

$$\left[\hat{Y}(z)\right]_k \triangleq \left[\frac{n_{\hat{y}}(z^{-1})}{d_{\hat{y}}(z^{-1})}\right]_k = \frac{[n_{\hat{y}}(z^{-1})]_k}{d_{\widehat{G}}(z^{-1})\,[d_u(z^{-1})]_k}, \tag{2.34}$$

where $d_{\widehat{G}}(z^{-1})$ is the characteristic polynomial of the model, $[d_u(z^{-1})]_k$ is the characteristic polynomial for the input over the k^{th} interval and $[n_{\hat{y}}(z^{-1})]_k$ is an appropriate numerator polynomial. Otherwise, for $k = q + 1$ or while the input is off

$$\left[\hat{Y}(z)\right]_k = \frac{[n_{\hat{y}}(z^{-1})]_k}{d_{\widehat{G}}(z^{-1})}. \tag{2.35}$$

From (2.34) it can be seen that the modal content of the sampled output signal consists of a combination of input and system modes, whereas, from (2.35) the modal content of the sampled output signal consists only of those contributed by the system.

Prony Based Analysis

Basic Properties The objective of the first step of Prony analysis is to find the eigenvalues of the sampled output signal $y(t)$. We assume that the signal $y(t)$ is

sampled at a sample period T smaller than the Nyquist period. During the k^{th} time interval the sampled signal $\hat{y}(\rho T)$ can be rewritten in discrete-time form as

$$\hat{y}_\rho = \sum_{i=1}^{\nu_k} \beta_{k,i} z_i^{\rho - M_{k-1}}, \ \rho = M_{k-1}, \ldots, M_k - 1,$$

where ρ is integer time, $\hat{y}_\rho \equiv \hat{y}(\rho T)$, $\beta_{k,i} \in \mathbb{C}$ for $i = 1, 2, \ldots, \nu_k$, $\nu_k \triangleq \eta + m_k$, and each z_i is a discrete-time eigenvalue of the system or of the input. It is well known that $\hat{y}_\rho$ satisfies its own characteristic equation, [2, 9]; thus from (2.34) and (2.35), it follows that

$$\left[d_{\hat{y}}(z^{-1})\right]_k \hat{y}_\rho \cong 0, \quad \rho = M_{k-1} + \nu_k, \ldots, M_k - 1, \tag{2.36}$$

where the characteristic polynomial of (2.34) and (2.36) is defined as

$$\left[d_{\hat{y}}(z^{-1})\right]_k \triangleq 1 - \left(\phi_{k,1} z^{-1} + \phi_{k,2} z^{-2} + \cdots + \phi_{k,\nu_k} z^{-\nu_k}\right). \tag{2.37}$$

Thus, for $\rho \in [M_{k-1} + \eta + m_k, M_k - 1]$, where the definition for ν_k has been substituted, we have from (2.36) and (2.37)

$$\hat{y}_\rho \cong \phi_{k,1} \hat{y}_{\rho-1} + \phi_{k,2} \hat{y}_{\rho-2} + \cdots + \phi_{k,\eta+m_k} \hat{y}_{\rho-\eta-m_k}. \tag{2.38}$$

Allowing ρ to range over the defined interval and substituting y_ρ for $\hat{y}_\rho$ provides a system of over-determined linear equations (i.e., more parameters than necessary to describe the 'true' system) that can be solved in a least-squares sense for the coefficients $\phi_{k,i}$. Factoring (2.37) then results in estimates for the eigenvalues of the signal. By removing the eigenvalues $\{\mu_{k,l}\}_1^{m_k}$ associated with the input signal, the desired estimates for the eigenvalues $\{\xi_j\}_1^\eta$ of the system are obtained along with residual eigenvalues due to the over-determination.

<u>Modal Identification</u> Three different approaches to obtaining the set $\{\xi_j\}_1^\eta$ are described in detail in [20] and are referred to as i) Separate least-squares solutions; ii) Combined least-squares solution for system eigenvalues; and iii) Combined

least-squares solution for all eigenvalues. Equations (2.36) - (2.38) are the basis for method i.

The second approach provides a good degree of insight to the first stage of Prony analysis and will be discussed below. Method ii requires the construction of auxiliary equations with knowledge of the input signal and solves directly for the system eigenvalues, ξ_j's. Let the characteristic polynomial of the system be given by

$$d_{\widehat{G}}(z^{-1}) \triangleq 1 - \left(\psi_1 z^{-1} + \psi_2 z^{-2} + \cdots + \psi_\eta z^{-\eta}\right) \tag{2.39}$$

and in factored form

$$d_{\widehat{G}}(z^{-1}) \equiv \prod_{j=1}^{\eta} \left(1 - e^{\xi_j T} z^{-1}\right). \tag{2.40}$$

The characteristic polynomial of the input for the k^{th} interval is given by

$$\left[d_u(z^{-1})\right]_k \triangleq 1 + \alpha_{k,1} z^{-1} + \alpha_{k,2} z^{-2} + \cdots + \alpha_{k,m_k} z^{-m_k}$$

and in factored form

$$\left[d_u(z^{-1})\right]_k \equiv \prod_{l=1}^{m_k} \left(1 - e^{\mu_{k,l} T} z^{-1}\right). \tag{2.41}$$

Over the k^{th} time interval, from (2.34) we have

$$d_{\widehat{G}}(z^{-1}) \left(\left[d_u(z^{-1})\right]_k y_\rho\right) \cong 0$$

or

$$d_{\widehat{G}}(z^{-1}) p_k(\rho) \cong 0, \tag{2.42}$$

where $p_k(\rho)$ is generated based on the known $[d_u(z^{-1})]_k$ of (2.41). The form of $p_k(\rho)$ is given by

$$p_k(\rho) = y_\rho + \alpha_{k,1} y_{\rho-1} + \alpha_{k,2} y_{\rho-2} + \cdots + \alpha_{k,m_k} y_{\rho-m_k}, \tag{2.43}$$

where $\rho \in [M_{k-1} + m_k, M_k - 1]$. Using (2.39), (2.42) and (2.43), and appropriately narrowing the range on ρ, we have

$$p_k(\rho) \cong \psi_1 p_k(\rho - 1) + \psi_2 p_k(\rho - 2) + \cdots + \psi_\eta p_k(\rho - \eta), \tag{2.44}$$

where $\rho \in [M_{k-1} + m_k + \eta, M_k - 1]$.

For some input interval, say $k = r$, if $M_r - 1$ is less then $M_{r-1} + m_r + \eta$, then the r^{th} interval of data cannot be used in the construction of the above set of equations. Also, for interval $k = q + 1$ or any interval where the input $u(t) = 0$, it follows from (2.35) and (2.39) that

$$y_\rho \cong \psi_1 y_{\rho-1} + \psi_2 y_{\rho-2} + \cdots + \psi_\eta y_{\rho-\eta}, \tag{2.45}$$

where

$$\rho : \begin{cases} \rho \in [M_{k-1} + \eta, M_k - 1] & if\ u(t) = 0, \\ \rho \geq M_{k-1} + \eta & if\ k = q + 1. \end{cases}$$

The unknowns in (2.44) and (2.45) are the ψ_i's. For each time interval of sufficient width, and where $u(t) \neq 0$, a set of linear equations in the following form result:

$$\mathbf{v}_k \cong \Phi_k \psi, \tag{2.46}$$

where $\psi \in \mathbb{R}^\eta$ consists of ψ_i's, the Toeplitz matrix $\Phi_k \in \mathbb{R}^{(M_k - M_{k-1} - m_k - \eta, \eta)}$ consists of appropriate ordering of $p_k(\rho-1), \ldots, p_k(\rho-\eta)$, and $\mathbf{v}_k \in \mathbb{R}^{(M_k - M_{k-1} - m_k - \eta)}$ consists of appropriate ordering of $p_k(\rho)$. For intervals where $u(t) = 0$ or $k = q+1$, we have the same ψ, however, the Toeplitz matrix $\Phi_k \in \mathbb{R}^{(M_k - M_{k-1} - \eta, \eta)}$ and $\mathbf{v}_k \in \mathbb{R}^{(M_k - M_{k-1} - \eta)}$ each consist of appropriate ordering of y_ρ. As suggested in [20], it may be necessary for numerical reasons to row scale each interval set of equations such that the matrices have same order of magnitude. In any case, with all input intervals of sufficient length used in (2.46) the concatenation results in an over-determined set of equations:

$$\mathbf{v} \cong \Phi \psi.$$

To arrive at a set of parameters ψ_j's, the following least-squares problem can be solved

$$\psi = arg \min_{\psi} \|\mathbf{v} - \Phi\psi\|_2.$$