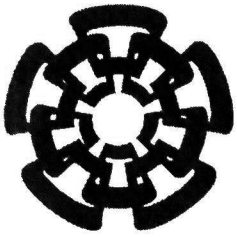
BC-635
Don.- 2011

xx(178999.1)



Centro de Investigación y de Estudios Avanzados
del Instituto Politécnico Nacional
Unidad Guadalajara

Monitoreo en tiempo real de dinámicas oscilatorias de sistemas de potencia usando algoritmos canónicos y recursivos

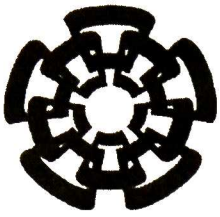
Tesis que presenta:
Ismael Moreno Nuñez

para obtener el grado de:
Maestro en Ciencias

en la especialidad de:
Ingeniería Eléctrica

Director de Tesis
Dr. Arturo Román Messina

CINVESTAV del IPN Unidad Guadalajara, Guadalajara, Jalisco, Agosto de 2010.



Centro de Investigación y de Estudios Avanzados

del Instituto Politécnico Nacional

Unidad Guadalajara

Real-time monitoring of power system oscillatory dynamics using canonical and recursive algorithms

A thesis presented by:
Ismael Moreno Núñez

to obtain the degree of:
Master in Science

in the subject of:
Electrical Engineering

Thesis Advisors:
Dr. Arturo Román Messina

CINVESTAV del IPN Unidad Guadalajara,Guadalajara, Jalisco, August 2010

Monitoreo en tiempo real de dinámicas oscilatorias de sistemas de potencia usando algoritmos canónicos y recursivos

Tesis de Maestría en Ciencias
Ingeniería Eléctrica

Por:
Ismael Moreno Nuñez
Ingeniero en Electrónica
Instituto Tecnológico de Minatitlán 2003-2007

Becario de CONACYT, expediente no. 219117

Director de Tesis
Dr. Arturo Román Messina

CINVESTAV del IPN Unidad Guadalajara, Agosto de 2010.

Real-time monitoring of power system oscillatory dynamics using canonical and recursive algorithms

Master of Science Thesis
In Electrical Engineering

By:
Ismael Moreno Núñez
Electronic Engineer
Instituto Tecnológico de Minatitlán 2003-2007

Scholarship granted by CONACYT, No. 219117

Thesis Advisors:
Dr. Arturo Román Messina

CINVESTAV del IPN Unidad Guadalajara, August, 2010.

Acknowledgments

Thanks to almighty God for the completion of this thesis. Only due to His blessings I could finish my thesis.

Thanks to my beloved mother Guadalupe whose love and dedication towards me and my brothers were fundamental in forming a great family.

Thanks to my family for their encouragement who are so supportive to me throughout my life.

I would like to express my deepest gratitude to my advisor Ph. D. Arturo Róman Messina by her patience, invaluable advices and suggestions, for making this thesis a reality.

I also would like to express my thanks to my professors Ph. D. Juan Manuel Ramirez, Ph. D. José Manuel Cañedo, Ph. D. Pablo Moreno Villalobos, and Ph. D. Abner Ramirez.

I would like to thank all people who have helped me during my study in Cinvestav Guadalajara but I cannot mention their names here one by one.

Finally, I would like express my gratitude to CONACYT, since this research would not been possible without its financial assistance.

Resumen

El análisis y caracterización de mediciones de área amplia sincronizadas en tiempo requiere de herramientas matemáticas que sean adaptables a las condiciones variantes del sistema, exactas y rápidas, y que al mismo tiempo reduzcan la complejidad de los datos para hacerlos útiles en decisiones de control y de tiempo real. En esta tesis, nuevos algoritmos basados en análisis de variación canónica y mínimos-cuadrados recursivos son propuestos para extraer propiedades modales a partir de mediciones sincronizadas en el tiempo. Los métodos toman en cuenta la naturaleza estocástica del comportamiento dinámico del sistema y pueden ser aplicados en tiempo real.

Se proponen y se evaluan técnicas rápidas de monitoreo de sistemas de área amplia para la evaluación y seguridad del sistema utilizando señales sintéticas y medidas. Las comparaciones con enfoques bien establecidos tales como algoritmos medidores de modos por procesamiento a bloques demuestran la exactitud y eficiencia de los algoritmos propuestos.

Abstract

Analysis and characterization of wide-area time-synchronized measurements requires mathematical tools that are adaptable to the varying system conditions, accurate and fast while reducing the complexity of the data to make them useful for control and real-time decisons. In this thesis, novel algorithms based on canonical variation analysis and recursive least-square algorithms are proposed to extract modal properties from time-synchronized measurements. The methods take into account the stochastic nature of system dynamic behavior and can be applied in real time.

Fast wide-area system monitoring techniques for both system monitoring and security are proposed and tested using synthetic and measured signals. Comparisons with well-established approaches such as mode-meter block-processing algorithms demonstrate the accuracy and efficiency of the proposed algorithms..

Index

List of Figures

Chapter 1
Introduction

1.1 Research Motivation and Problem Statement

Modal identification is a neccesary first step towards improved modeling and control of power system dynamic processes. It complements theoretical modeling by allowing the analyst to investigate several aspects of system behavior that cannot otherwise be explored.

Recent advances in comprehensive monitoring of power system behavior by means of properly placed time-synchronized phasor measurement units (PMUs) along with technical developments in communication technologies provide the opportunity to analyze and characterizing inter-area swing dynamics. Wide-area real-time monitoring, in particular, may provide invaluable in power system dynamic analysis by giving a quick assessment of the damping and frequency content of dominant system modes. Prediction of temporal dynamics, with the ultimate application to real-time system monitoring , protection and control, remains a major research challenge due to the complexity of the driving dynamic and control processes operating on various temporal scales that can become dynamically involved. In addition, measured data, may exhibit quite different dynamics at each system location or exhibit abrupt changes, dynamic irregularities, or be complicated by nonlinear trends or noise.

Measurement-based signal processing techniques are becoming a standard tool for analyzing the dynamic properties of the system. Determining the conditions for the onset of system oscillations can be a challenging task because of the complexity of observed system behavior.

Measured ambient data , in particular, are known to exhibit noisy, nonstationary fluctuations resulting primarily from small magnitude, random changes in load, driven by low-scale motions or nonlinear trends originating from slow control actions or changes in operating conditions.

Traditional analysis approaches for system identification which rely on stationarity are unable to resolve the localized nature of these processes and hence provide little useful information concerning the nature of noisy, time-varying oscillatory proceses.

In this thesis, a new method for analyzing the temporal dynamics of

inter-area oscillations using canonical and recursive least-squares methods is proposed. This research is motivated by the limitations of a wide variety of techniques proposed in the literature to deal with time-varying behavior. Two novel algorithms are developed to address identification issues that present several conceptual and numerical advantages. The first method is a canonical variate algorithm. The second is an algorithm to compute modal information using recursive least-squares algorithms.

As specific applications, data obtained from PMU measurements from real events in various power systems are used to examine the potential usefulness and accuracy of time series analysis techniques to characterize the modal structure of the observed oscillations.

The methods can be implemented for on-line estimation of modal damping and frequency using synchronized wide-area measurement systems.

1.2 Review of previous work

Recently, the problem of modal identification of electromechanical data from time-synchronized measurements has received considerable attention. Several representations have been explored over the last years to analyze natural and synthetic signals which are characterized by non-stationary behavior.

The detection of instantaneous or locally-occurring transient oscillations is crucial to protection and control strategies [3]. The past few years witnessed a remarkable breakthrough in numerical algorithms to estimate modal properties from ambient data. Algorithms with the ability to extract modal information in the presence of noise, and changing operating conditions are being developed and tested using measured data [24],[29] with varied success.

A variety of techniques have been proposed for automatic extraction of dynamic features from measured data. Block- processing techniques have been used for extracting and characterizing dynamic features. Extending these approaches to extract modal parameters in near real-time, however, is very challenging [32], [33].

In parallel to this effort, ambient-noise driven based electromechanical mode estimation techniques with the ability to track modes under ambient conditions have been developed and tested using real measured data. As pointed out in [13], the accuracy of any mode estimation technique is limited and may be affected by the very nature of system dynamic behavior as well as the approximation made in the computation.

Currently the concept of treating power system motion stochastically has been widely accepted. In this connection, many studies have been undertaken to analyze the problems of characterization, simulation and response of large,

complex power systems. Preliminary , analytical explorations on the use of these techniques indicate that these methods have the potential to be used in real-time wide-area monitoring and control systems. Several technical challenges, however, need to be solved to enable efficient implementation of these techniques in existing wide-area measurement, control and protection systems.

This has provided the thrust of this work.

1.3 Thesis Objectives and Overview of Contributions

The main objectives of this research can be summarized as follows:

The development of an analytical framework for estimating electromechanical modes from ambient data.

The development of efficient numerical techniques to assess modal structure in near real-time. In particular, novel algorithms based on stochastic subspace identification techniques are developed that improve numerical characteristics of existing implementations. The key innovation in the new method is the use of an statistical framework which incorporates linear filtering techniques. In addition, the new model leads to a more realistic power system representation in the presence of noise and random load variations. Because of its adaptive nature, the proposed techniques can capture temporal variations better than other approaches

The original contributions of this thesis are:

1. The development of a methodology for automatically implementing a modemeter algorithm in near real-time that incorporates predictive techniques

2. To this end, a mathematical framework is developed, which extends current formulations. Unlike existing approaches, the technique has a solid analytical foundation and incorporates dynamic features in an stochastic framework. Other methods such as canonical variate analysis (CVA) can be shown to be a particular case of the developed techniques.

3. The comparison of various adaptive filtering methods for modal identification.

1.4 Thesis structure

This thesis is comprised of eight chapters.

This introductory chapter gives the motivation and background neccesary for this research, and gives a general overview of existing analysis methods.

The remainder of the thesis is organized as follows:

Chapter 2 provides a general overview of modal identification methods in power systems. The theoretical basis for these methods is described as well as application, properties and performance.

In Chapter 3, a review of previous work on modal identification is presented with an emphasis on its application to power system measured data. Traditional identification methods used in the power system literature are examined and models which adequately characterize the underlying oscillatory behavior are reviewed.

Chapter 4 presents a conceptual modeling approach based on the use of stochastic subspace identification methods to estimate modal structure that uses wide-area measurements. Methods related to stochastic subspace identification are also discussed and their relative advantages and limitations are pointed out.

Chapter 5 discusses the application of recursive least-squares techniques to monitor power system oscillatory dynamics. Numerical algorithms for the system identification problem are given.

Chapter 6 discusses the application of the proposed techniques to extract modal damping and frequency from time-synchronized measured data. Both simulated and measured data are considered.

Finally, Chapter 7 presents the general conclusions and suggestions for future work. Possible modifications and improvement are considered.

Chapter 2
Modal Identification in Power System

Ambient-noise driven based electromechanical mode estimation techniques with the ability to track modes under ambient conditions have been developed and tested using real measured data.
This Chapter provides a perspective on the use of wide-area measurement systems analysis tools for tracking power system oscillatory characteristics. Sources of dynamic information are reviewed and system identification methods are classified. The performance properties of parametric and non-parametric methods are reviewed in the light of their application to the analysis of power system oscillatory dynamics. Factors affecting the performance of numerical algorithms for automated dynamic stability assessment are also reviewed.

2.1 Modal Identification

Advances in signal processing algorithms, along with continuously growing computational resources and monitoring systems are beginning to make feasible the analysis of transient processes using real-time information. Much of the recent work has been driven by interest in near real-time estimation of electromechanical modal properties from measured ambient data collected using Phasor Measurement Systems (PMUs) [4],[3]. Wide-area measurements improve the visibility and observability of critical system modes and may lead to improve controllability [23].

By tracking the modal attributes of critical modes, the onset of system instability can be determined and this information may be used to trigger control actions in the system [23] and provide valuable information to system operator [24], [35].

In recent years, the notion a mode-meter has been introduced [4]. The mode meter algorithms automatically estimate the instantaneous damping and frequency of critical modes. To be of practical use, mode meters must be able to process, in near real time, all types of data.

Sources of dynamic information include [4] [35]:

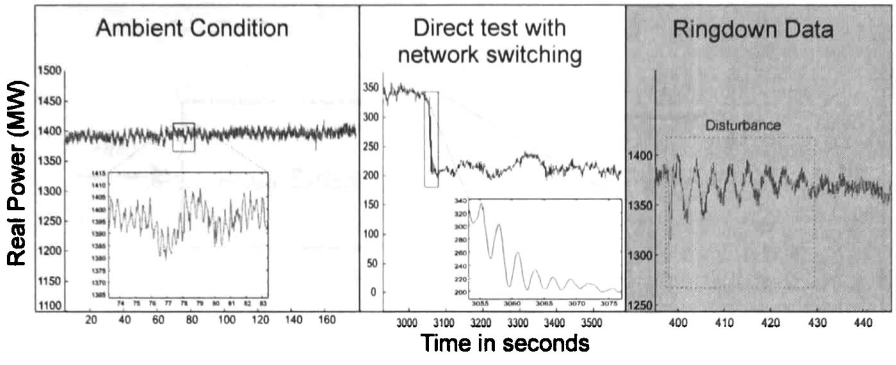


Figure 2.1: Data types according to type of exciting in inputs of system

- Ambient noise measurements
- Ringdown measurements
- Probing test

Ambient analysis estimate the modes when the primary excitation to the system is random load variations. Short term (noise) fluctuations are a consequence of load variations whose characteristics vary in both the short and longer term. Ringdown analysis, on the other hand, is aimed at estimating modal properties from a transient data arising from large perturbations in the system. By tracking the evolving dynamics of the underlying oscillations, the onset of system stability can be determined and the critical stages for analysis and control can be identified.

Finally, probing data is obtained from the system response to test signals of known characteristics [35]. Figure (2.1) gives a conceptual representation of measured data showing the system response to various operating conditions.

Ambient-based mode estimation, in particular, can be conducted in both the time domain and the frequency domain. As discussed above, extraction of modal information may involve different types of analysis:

- Disturbance analysis
- Ambient data analysis (spectral and correlation analysis, parametric modal analysis)
- Analysis of direct tests involving topology switching tests and low-level excitation test

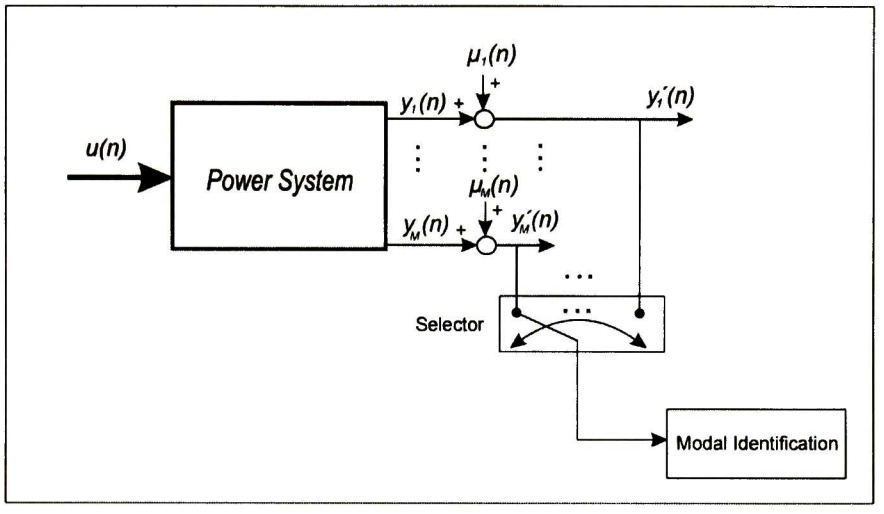


Figure 2.2: Block diagram illustrating information sources for system identification at system locations where the mode of interest is most observable

The modal properties of specific electromechanical modes are obtained using time-synchronized measurements at system locations where the mode of interest is most observable as explained in [33] Figure (2.2) gives a conceptual representation of a wide-area measurement system showing inputs and outputs of interest and sources of noise.

In the case when the input $u(n)$ see (2.2) is any deterministic input, all topology characteristics are known. Tipically, the probing signals $u(n)$ is designed to excite a given set of modes or frequencies of concern [4].

Referring to Figure (2.2), when the input $u(n)$ is natural input noise, the power system produces a colored or process noise component in the output $y(n)$. An example of input noise is the random variation in the loads. The output signal also contains measurement noise $\mu(n)$ produced by instruments. In addition, topology changes result in noise in system measurements that has to be accounted for.

A motivation for this research is the fact that process (colored) noise is particulary attractive for nonintrusive identification of system dynamics, it is identification without changes to the system, and it is an advantage of ambient analysis. However, the natural input noise $u(n)$ is not observable and its true nature is not known. A common assumption is that input noise is white and wide-sense stationary.

2.2 Classification of System Identification Methods used for Modal Identification

System identification deals with the problem of building mathematical models of dynamical systems based on observed data from the system. The models more used are rational functions and state space models. From these models it is possible obtain the low frequency electromechanical modes in power system [16]. *In the sections that follow, the main characteristics of methods used for modal identification are examined. These methods are classified into two main groups; nonparametric and parametric.*

2.2.1 Nonparametric Estimation

A linear time-invariant (LTI) model can be described by its transfer functions or by the corresponding impulse responses. Methods that aim at determining these functions by direct techniques without first selecting a confined set of possible models are often called nonparametric, that is to say, they do not employ a finite-dimensional parameter vector in the search for a best description [17] [16]. Before presenting the characteristics of these methods, some definitions are needed.

Nonrecursive Representation. Consider a stable LTI system with impulse response $g(n)$ and input $u(n)$. The output $y(n)$ can be written formally as the infinite sum

$$y(n) = \sum_{k=-\infty}^{\infty} g(k)u(n-k)$$

where the output is a weighted combination of system inputs.

Recursive Representation. Let the inverse system $\overline{G}(z) = 1/G(z)$ be causal and stable. If it is further assumed, without any loss of generality, that $g(0) = 1$, then $\overline{g}(n) = Z^{-1}\left\{\overline{G}(z)\right\}$ has $\overline{g}(0) = 1$ and the input $u(n)$ can be written as

$$u(n) = g(n) + \sum_{k=1}^{\infty} \overline{g}(k)y(n-k) \tag{2.1}$$

Solving for $y(n)$, yields

$$y(n) = -\sum_{k=1}^{\infty} \overline{g}(k)y(n-k) + u(n)$$

in which the output y(n) is a linear combination of all past output values plus the present input value.

Innovation Representation. If the system $G(z)$ is minimun-phase, then $g(n)$ and $\overline{g}(n)$ are causal and stable. Therefore, the output signal can be expressed nonrecursively by

$$y(n) = \sum_{k=0}^{\infty} g(k)u(n-k) = \sum_{k=-\infty}^{n} g(n-k)u(k)$$

From this expression, and using Equation (2.1) it is possible to show that

$$\begin{aligned} y(n+1) &= \sum_{k=-\infty}^{n} g(n+1-k)u(k) + u(n+1) \\ &= \sum_{k=-\infty}^{n} g(n+1-k)y(k) + u(n+1) \end{aligned} \tag{2.2}$$

Equation (2.2) indicates that if the system generating $y(n)$ is minimum-phase, the sample $u(n+1)$ captures all the new information (innovation) in the sample $y(n+1)$. All other information can be predicted from the past samples $y(n), y(n-1), ...$ of the signal.

In practical applications, where only a finite segment of a signal is available, it forces previous representations to become parametric.

The nonparametric techniques that do not assume a particular functional form, but allow the form of the estimator to be determined entirely by the data, are those nonparametric methods based on the discrete Fourier transform of either the signal segment or its autocorrelation sequence [17].

Random Signal Model. If the system has a zero-mean white noise input $e(n)$ with variance σ_e^2, autocorrelation $r_e(l) = \sigma_e^2\delta(l)$, and PSD $R_e(e^{j\omega}) = \sigma_e^2$, $-\pi < \omega \leq \pi$, then the autocorrelation and PSD of the output $y(n)$ are given by [17] [11]

$$r_y(l) = \sigma_e^2 \sum_{k=-\infty}^{\infty} h(k)h^*(k-l) = \sigma_e^2 r_h(l)$$

$$R_y(e^{j\omega}) = \sigma_e^2 \left|H(e^{j\omega})\right|^2 = \sigma_e^2 R_h(e^{j\omega})$$

Then, the shape of the autocorrelation and the power spectrum of the output signal are completely characterized by the system.

To deal with the problem of estimating the power spectrum $R_y(e^{j\omega})$ of a stationary process $y(n)$ from a finite record of observations of a single realization,

the power spectral density of a zero-mean stationary stochastic process is defined as

$$R_y(e^{j\omega}) = \sum_{l=-\infty}^{\infty} r_y(l)e^{-j\omega l} \tag{2.3}$$

where it is assumed that the autocorrelation sequence $r_y(l)$ is absolutely summable.

The objective is to find an estimate that can characterize the power-versus-frequency distribution of the stochastic process using only a segment of a single realization. For this to be possible, the estimate should typically involve some kind of averaging among several realizations or along a single realization.

In some practical applications, it is possible to directly measure the autocorrelation $r_y(l)$ with great accuracy but focus on the stochastic version of the problem where $R_y(e^{j\omega})$ is estimated from the data segment $y(n)$.One approach to estimate $R_y(e^{j\omega})$ is using Equation (2.3) where $r_y(l)$ is estimated from the available data. Another form is using the periodogram. The periodogram of the data segment $y(n)$ is defined by [17]

$$\widehat{R}_y(e^{j\omega}) = \frac{1}{N}\left|\sum_{n=0}^{N-1} p(n)e^{-j\omega n}\right|^2 = \frac{1}{N}\left|P\left(e^{j\omega}\right)\right|^2$$

where $P\left(e^{j\omega}\right)$ is the DTFT of the windowed sequence

$$p(n) = y(n)w(n) \qquad 0 \leq n \leq N-1$$

In practice, the second-order moments of the signal to be modeled are not known a priori and have to be estimated from a set of signal observations. This element introduces a new dimension and additional complications to the system identification problem.

2.2.2 Parametric Estimation

The parametric estimation is based on estimate a finite number of parameters of a functional form such as a rational transfer function. The modal identification in power system is obtained of models as rational transfer functions or finite dimensional state-space descriptions.

A linear time-invariant model is specified by the impulse response $g(n)$, the spectrum $R_{y_e}(e^{j\omega}) = \sigma_e^2 \left|H(e^{j\omega})\right|^2$ of the additive disturbance, and, possibly, the probability density function (PDF) of the disturbance $e(t)$ [16] [11]. A complete model is thus given by

$$y(n) = G(z)u(n) + H(z)e(n) \qquad (2.4)$$
$$f_e(\cdot), \text{ the PDF of } e$$

with

$$G(z) = 1 + \sum_{k=1}^{\infty} g(k)z^{-k} \qquad H(z) = 1 + \sum_{k=1}^{\infty} h(k)z^{-k}$$

A particular model thus requires the specification of the three functions G, H, and f_e It is often impractical to make this specification by enumerating the infinite sequences $g(k), h(k)$ together with the function $f_e(\cdot)$. Instead, one chooses to work with structures that permit the specification of G and H in terms of a finite number of numerical values. Rational transfer functions and finite dimensional state-space descriptions are typical examples of this. Also, frequently, the PDF f_e is not specified as a function, but described in terms of the first and second moments.

The specification of Equation (2.4) in terms of a finite number of numerical values, or coefficients, has another and most important consequence for the purposes of system identification. This means that the coefficient in question enter in the model (Equation (2.4)) as parameter to be determined. Such parameter is denoted by the vector Θ.

Assume that a set of candidate models has been selected, and it is parametrized as a model structure, using a parameter vector Θ. The search for the best model within the set then becomes a problem of determining or estimating Θ. There are many different ways of organizing such a search and also different views on what one should search for, [16].

Principles of parametric estimation methods

As was discussed earlier in this section, the nonparametric representations such as recursive and innovation in practice adjusts a model using a set of data, then these representations taking a parametric approach. The models have a structure S (model estructure). Particular models are denoted by $S(\Theta)$ and are parametrized using the parameter vector $\Theta \in D_S \subset \mathbb{R}^d$

The set of models is defined by

$$S_{set} = \{S(\Theta) \mid \Theta \in D_S\}$$

The model $S(\Theta)$ can contain assumptions about the character of the associated prediction errors, such as their variances or their probability distribution.

Assume we are given a set of data

$$\mathbf{z}^{\eta}=[y(1),u(1),...,y(\eta),u(\eta)]$$

the problem that arises in parametric estimation is to decide how to use the information contained in $\mathbf{z}^{\eta}$ to obtain an adecuate value $\widehat{\Theta}_{\eta}$ of the parameter vector, and hence an adecuate member $S(\widehat{\Theta}_{\eta})$ in the set S_{set}. This represents a mapping $\mathbf{z}^{\eta} \rightarrow \widehat{\Theta}_{\eta} \in D_S$, and forms the basis of parametric estimation methods.

Each model of S_{set} is a linear predictor for the system in Eq(2.4) of the form,

$$S(\Theta) : \widehat{y}(n \mid \Theta) = [1 - H^{-1}(z,\Theta)]y(n) + [H^{-1}(z,\Theta)G(z,\Theta)]u(n)$$

Then the prediction error given by a certain model $S(\Theta_p)$is given by

$$\varepsilon(n,\Theta_p) = y(n) - \widehat{y}(n \mid \Theta_p) \tag{2.5}$$

Based on $\mathbf{z}^{\eta}$, the prediction error can be computed for $n = 1,2,...,\eta$. The principal objective of model identification based on prediction error is to adjust the error to be zero or as small as possible. The general procedure is to compute the error $\varepsilon(n,\Theta)$ for a set of data $\mathbf{z}^{\eta}$, at time $n = \eta$, select $\widehat{\Theta}_{\eta}$ so that the prediction errors $\varepsilon(n,\widehat{\Theta}_{\eta})$, $n = 1,2,...,\eta$, become as small as possible.

There are two general approaches to determine the minimization error. One approach is to form a scalar-valued norm or criterion function that measures the size of ε other approach is achieve that $\varepsilon(t,\widehat{\Theta}_n)$ be uncorrelated with a given data sequence, that is to say, that certain projections of $\varepsilon(t,\widehat{\Theta}_n)$ are zero [17] [16].

Prediction Error Identification Approach

Prediction error identification methods are based on criterion functions that measure the size of ε using any norm applied to the prediction error sequence vector in $\mathbb{R}^n$ The way of evaluating how large is as follows:

The prediction error sequence is filtered through a stable linear filter

$$\varepsilon_f(n,\Theta) = l(z)\varepsilon(n,\Theta) \qquad 1 \leq n \leq \eta \tag{2.6}$$

If the predictor is linear and time invariant with $y(n)$ and $u(n)$ being scalars, then filtering ε is the same result as first filtering the input and output. With the filter applied to prediction error sequence, the norm is applied to ε_f as follows:

$$N_{\eta}(\Theta,\mathbf{z}^{\eta}) = \frac{1}{\eta}\sum_{n=1}^{\eta} p(\varepsilon_f(n,\Theta)) \tag{2.7}$$

where $p(\cdot)$ is a scalar-valued function. Because $N_\eta(\Theta, \mathbf{z}^\eta)$ is a well-defined scalar-valued function for any $\mathbf{z}^\eta$of the model parameter Θ, it is possible to estimate $\widehat{\Theta}_n$ by minimization of Equation (2.7):

$$\widehat{\Theta}_\eta = \widehat{\Theta}_\eta(\mathbf{z}^\eta) = \underset{\Theta \in D_S}{\arg\min}\ N_\eta(\Theta, \mathbf{z}^\eta) \tag{2.8}$$

The methods based on Equation (2.8) are classified as prediction-error identification methods, with the different methods varying depending on the choices of $p(\cdot)$ and $l(\cdot)$.

The Least Squares Method

A predictor using a linear regression model structure is defined by

$$\widehat{y}(n \mid \Theta) = \psi^T(n)\Theta \tag{2.9}$$

where the prediction is linear in the parameters and ψ is the vector of regressors:

$$\psi(n) = [-y(n-1) \quad -y(n-2)... \quad -y(n-a) \quad u(n-1)... \quad u(n-b)]^T$$

The associated prediction error is

$$\varepsilon(n, \Theta) = y(n) - \psi^T(n)\Theta$$

The Least-Squares criterion is a particular case of Equations (2.6), (2.7) with $l(z) = 1$ and $p(\varepsilon) = \frac{1}{2}\varepsilon^2$, namely:

$$N_\eta(\Theta, \mathbf{z}^\eta) = \frac{1}{\eta}\sum_{n=1}^{\eta}\frac{1}{2}\left[y(n) - \psi^T(n)\Theta\right]^2 \tag{2.10}$$

Equation (2.10) is a quadratic function in Θ and can be minimized analytically as follows

$$\widehat{\Theta}_\eta = \arg\min N_\eta(\Theta, \mathbf{z}^\eta) = \left[\frac{1}{\eta}\sum_{n=1}^{\eta}\psi(n)\psi^T(n)\right]^{-1}\frac{1}{\eta}\sum_{n=1}^{\eta}\psi(n)y^T(n)$$

It can be seen that this minimization consist of estimates of the covariances functions of $y(n)$ and $u(n)$ and is therefore related to correlation analysis.

Estimating State Space Models using Least-Squares. A linear system in state space form is represented by:

$$\begin{aligned} x(n+1) &= Ax(n) + Bu(n) + w(n) \\ y(n) &= Cx(n) + Du(n) + v(n) \end{aligned} \tag{2.11}$$

with plant noise $w(n)$ and observation noise $v(n)$ both are white noise. Being possible estimates the states $x(n)$ using subspace methods and having $u(n)$ and $y(n)$ measured, the model Equation (2.11) becomes a linear regression:

$$Y(t) = \Psi\Phi(n) + \mu(n) \tag{2.12}$$

with

$$Y(n) = \begin{bmatrix} x(n+1) \\ y(n) \end{bmatrix}, \ \Psi = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \ \Phi(n) = \begin{bmatrix} x(n) \\ y(n) \end{bmatrix} \quad \mu(n) = \begin{bmatrix} w(n) \\ v(n) \end{bmatrix} \tag{2.13}$$

From Equation (2.12), all the matrix elements in Ψ can be estimated using a least-squares approximation. Straightforward computation yields

$$N_\eta(\Psi, \mathbf{z}^\eta) = \frac{1}{\eta} \sum_{n=1}^{\eta} \left\| y(n) - \Psi^T \Phi(n) \right\|^2$$

with the estimate

$$\widehat{\Psi}_\eta = \left[\frac{1}{\eta} \sum_{t=1}^{\eta} \Phi(n)\Phi^T(n) \right]^{-1} \frac{1}{\eta} \sum_{n=1}^{\eta} \Phi(n) y^T(n)$$

The covariance matrix for $\mu(n)$ can also be estimated easily as the sample sum of the squared model residuals. This will give the covariance matrices as well as the cross covariance matrix for w and v [16] [21].

The Maximum Likelihood Method

Maximum likelihood is an approach that deals with the problem of extracting information from observation that themselves could be unreliable; these observations are described as realizations of stochastic variables [16]. The maximum likelihood is based on the probability density function (PDF) of $y^\eta = [y(1), ..., y(\eta)]$:

$$f(\Theta; x_1, x_2, ..., x_\eta) = f_y(\Theta; x_\eta)$$

where Θ is supposed to be unknown and the purpose is in fact to estimate the vector Θ using the observation y^η , that is to say, $\widehat{\Theta}(y^\eta)$. Then an estimate of Θ that maximize the probability of event observed is the maximum likelihood estimator. The probability that the realization indeed should take the value y^η is thus proportional to $f_y(\Theta; y^\eta)$ and an estimator of Θ is given by

$$\widehat{\Theta}(y^\eta) = \arg\max f_y(\Theta; y^\eta)$$

This expression is the maximum likelihood estimator for fixed y^η.

Relation of Maximum Likelihood with Error-Prediction Method. Having a model described by

$$S(\Theta) \;:\; \widehat{y}(n \mid \Theta)\;,\; \varepsilon(n,\Theta) = y(n) - \widehat{y}(n \mid \Theta) \text{ are independent and have the PDF } f_e(x, n; \Theta)$$

and replacing the dummy variables x_i by the corresponding observations y_i and using Baye's Rule [11] the likelihood function, one has

$$\begin{aligned} f_y(\Theta; y^\eta) &= \prod_{n=1}^{\eta} f_e(y(n) - \widehat{y}(n \mid \Theta), n; \Theta) \\ &= \prod_{n=1}^{\eta} f_e(\varepsilon(n,\Theta), n; \Theta) \end{aligned} \tag{2.14}$$

Noting that this function can be maximized by multiplying it by $\frac{1}{\eta}\log$:

$$\frac{1}{\eta}\log f_y(\Theta; y^\eta) = \frac{1}{\eta}\sum_{n=1}^{\eta} \log f_e\left(\varepsilon(n,\Theta), n; \Theta\right)$$

and defining $p(\cdot) = -\log f_e(\cdot)$ it is possible show that the maximum likelihood method is a special case of the prediction-error criterion:

$$\widehat{\Theta}(y^\eta) = \underset{\Theta}{\arg\min} \; -\frac{1}{\eta}\sum_{n=1}^{\eta} \log f_e\left(\varepsilon(n,\Theta), n; \Theta\right)$$

In practice, the exact likelihood function is quite complicated for time-series problems and several approximation have to be introduced.

Correlation Approach

The main idea behind this method is that it produces prediction errors that are independent of past data. That is to say, select a certain finite-dimensional vector sequence $\phi(n)$ derived from $\mathbf{z}^{n-1}$ and demand a certain transformation of $\varepsilon(n,\Theta)$ to be uncorrelated with $\phi(n)$ [17], [11], [21] [6]. More formally, we can write

$$\frac{1}{\eta}\sum_{n=1}^{\eta} \phi(n)\varepsilon(n,\Theta) = 0 \tag{2.15}$$

and the Θ-value that satisfies this equation would be the best estimation $\widehat{\Theta}_\eta$ based on the observed data, it is:

$$\widehat{\Theta}_\eta = \underset{\Theta \in S_M}{\text{sol}} \left(\frac{1}{\eta} \sum_{n=1}^{\eta} \phi(n)\varepsilon(n,\Theta) = 0 \right) \tag{2.16}$$

Would be the case when is useful to consider an augmented correlation sequence of higher dimension so that Equation (2.16) is an overdetermined set of equations, typically without any solution. Then the estimate is taken to be the value that minimizes some quadratic norm of Equation (2.15), these correlation approaches links with the minimization errors of last section.

Instrumental-Variable Methods [16] [11]. These methods use the linear regression model in Equation (2.9) where the solution using $\phi(n) = \psi(n)$ is:

$$\widehat{\Theta}_\eta = \text{sol} \left(\frac{1}{\eta} \sum_{n=1}^{\eta} \psi(n) \left[y(n) - \psi^T(n)\Theta \right] = 0 \right)$$

When the data is described by

$$y(n) = \psi^T(n)\Theta + v(n)$$

being $v(n)$ some sequence and having the restrictions of not be a sequence of independent random variables with zeros mean values or not be independent of the input sequence $u(n)$ and not has $a = 0$. Then $\widehat{\Theta}_\eta$ will not tend to real value Θ. The form of tend real value Θ is choice elements of $\phi(n)$ having the following properties:

$$\mathbf{E}\left[\phi(n)\psi^T(n)\right] \text{ be nonsingular}$$

$$\mathbf{E}\left[\phi(n)v^T(n)\right] = 0$$

Solving this expression for Θ yields

$$\widehat{\Theta}_\eta = \text{sol} \left(\frac{1}{\eta} \sum_{n=1}^{\eta} \phi(n) \left[y(n) - \psi^T(n)\Theta \right] = 0 \right)$$

or

$$\widehat{\Theta}_\eta = \left[\frac{1}{\eta} \sum_{n=1}^{\eta} \phi(n)\psi^T(n) \right]^{-1} \frac{1}{\eta} \sum_{n=1}^{\eta} \phi(n) y^T(n)$$

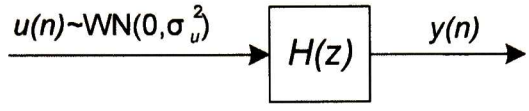


Figure 2.3: Synthesis or coloring filter

2.3 Ambient Data Pre-Processing

2.3.1 Assumptions of Ambient Data

In this research, ambient data of power system are assumed to be produced by white noise inputs $u(n)$ springing from random perturbations such as random load changes, these perturbations are enough for exciting the modes of the system to be observed for adecuate identification technique. The noise input $u(n)$ have the character of input, but it is not controllable and not is observable. It is not known beforehand. It is thus natural to employ a probabilistic framework to describe the process noise. A complete characterization would be to describe the conditional joint probability density function. This would, however, in most cases be too laborious and impractical. The approach used in this research is versatile enough for most practical purpose. The principal idea to consider the white noise input $u(n)$ injected to the system $H(z)$ is generates the signal $y(n)$ by introducing dependence in the white noise input $u(n)$ and it is known as the synthesis or coloring filter, the Figure (2.3) shows a block diagram of coloring filter.

Other aspect to be consider is that of the colored noise output $y(n)$ produced by the system $H(z)$ is contamined by measurement noise $\mu(n)$ as shows the Figure (2.2).

2.3.2 Pre-Processing

The data collected of power system $y(n)$ are not likely to be in shape for immediate use in identification algorithms. There are several deficiencies in the data that should be taking account, such as, high-frequency disturbances occasional that are not interest to the system dynamics, outliers, missing-data and offsets [16] [35]. One approach to this problem is the use of pre-processing techniques.

Signal Offset. An approach to deals with offset is removing the disturbances by explicit pretreatment of the data. The most natural way is to

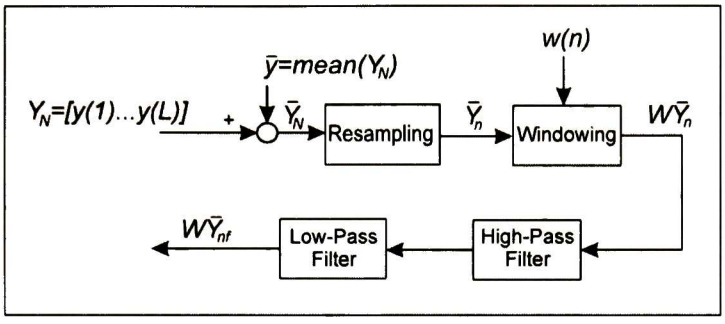


Figure 2.4: Block diagram of ambient data pre-processing

determine the level offset $\overline{y}$ that corresponds to a constant close to the desired operating point. A sound approach is to define:

$$\overline{y} = \frac{1}{\eta} \sum_{n=1}^{\eta} y(n)$$

Then an output that varies around $\overline{y}$ is likely to be close to a steady point, it is substract to data $y(n)$.

Outliers and Missing Data. In practice, the data collected may exhibit missing of singles values or portions [16] [4] [35], due to malfunctions in the measuring devices or communication channels.

Such bad values are often called outliers, and may have a substantial negative effect on the estimate. To deal with outliers and missing data, there are a few possibilities. One is to cut out segments of the data sequence so that portions with bad data are avoided. The segments can then be merged. The outliers is better to treat as missing data and view then as unknown parameters.

Prefiltering. The prefilter applied to data have the propose to remove disturbance of high and low frequencies that is not want to include in the modeling [16]. In this research is avoid the frequencies out of range 0.1-1 Hz. Prefiltering the data $y(n)$ is equal as filtering the prediction errors.

High-frequency disturbances in the data, above the frequencies of interest for the system dynamics, indicate that the choices of sampling interval and presampling filters were not thoughtful enough. This can however be remedied by low pass filtering of the data.

Figure (2.4) shows a block diagram of pre-processing ambient data before apply the modal identification algorithm.

Chapter 3
Previous Work on Modal Identification Techniques Using Ambient Data

Modal identification techniques on power system have increased and perfected using ringdown data. However, in the case to use ambient data, the techniques have some shortcomings. The ability to track the low frequency modes in equilibrium conditions of power system and before perturbations is very important because it provides evidence and possible prediction of oscillations and thus it is possible take corrective actions. This chapter discusses the experience in the use of ambient-based modal identification techniques. The analysis focuses on three major analysis techniques: Prony and matrix pencil methods, and parametric methods.
The analytical basis of these techniques are reviewed in the context of their application to power system data.

3.1 Improved Ringdown Methods To Deal With Ambient Data

3.1.1 Prony Method

Prony analysis is a method of fitting a linear combination of exponential terms to a signal. Consider the problem of fitting a sum of exponential functions

$$y(n) = \sum_{k=1}^{M} a(k)e^{s(k)n} \tag{3.1}$$

to real data, for $N > 2M$. where the parameters $a(k)$ and $s(k)$ for $k = 1, 2, ..., M$ are unknown complex numbers. Assume that the values of y are specified on a set of equally spaced points $1, 2, ..., N$. Evaluating this expression at $y = 0, 1, ..., N - 1$ results in [8].

$$
\begin{aligned}
a(1) + a(2) + ... + a(M) &= y(0) \\
a(1)e^{s(1)} + a(2)e^{s(2)} + ... + a(M)e^{s(M)} &= y(1) \\
a(1)e^{2s(1)} + a(2)e^{2s(2)} + ... + a(M)e^{2s(M)} &= y(2) \\
&\vdots \\
a(1)e^{(N-1)s(1)} + a(2)e^{(N-1)s(2)} + ... + a(M)e^{(N-1)s(M)} &= y(N-1)
\end{aligned}
\tag{3.2}
$$

For purposes of illustration, consider a four-sample case. It follows that

$$
\begin{aligned}
C_1 + C_2 &= M_0 \\
C_1\mu_1 + C_2\mu_2 &= M_1 \\
C_1\mu_1^2 + C_2\mu_2^2 &= M_2 \\
C_1\mu_1^3 + C_2\mu_2^3 &= M_3
\end{aligned}
\tag{3.3}
$$

representing four equations in the four unknowns μ_1, μ_2, C_1 and C_2. Assuming now μ_1 and μ_2 to be the zeros of $A(\mu)$

$$
A(\mu) = (\mu - \mu_1)(\mu - \mu_2) = \mu^2 + \alpha_1\mu + \alpha_2 = 0 \tag{3.4}
$$

the problem now is to determine the coefficients α_1 and α_2.To find these coefficients the third equation of (3.3) is multiplied by 1, the second by α_1 and the first by α_2, and using Eq. (3.4) one finds

$$
M_2 + M_1\alpha_1 + M_0\alpha_2 = 0
$$

Similary, from the fourth, third, and second equations we obtain

$$
M_3 + M_2\alpha_1 + M_1\alpha_2 = 0
$$

Returning now back to the set of equations (3.2) and generalizing the above concept, the roots of the algebraic equation may be written

$$
\begin{aligned}
A(e^{s(k)}) &= \prod_{k=1}^{M}(e^s - e^{s(k)}) = 0 \\
&= e^{Ms} + \alpha_1 e^{(M-1)s} + \alpha_2 e^{(M-2)s} + ... + \alpha_{M-1}e^s + \alpha_M
\end{aligned}
\tag{3.5}
$$

are solved for the coefficients $\alpha_1, ..., \alpha_M$. Following a similar procedure to that of the example, the first equation in (3.2) is multiplied by α_M the second equation by α_{M-1}; the process is continued until the Mth equation is multiplied by α_1, and the $(M+1)$th equation by 1, and the results are summed together. It can be proved that

$$
y(M) + y(M-1)\alpha_1 + ... + y(0)\alpha_M = 0
$$

This yields a set of $N - M - 1$ equations of the form

$$\begin{aligned} y(M) + y(M-1)\alpha_1 + ... + y(0)\alpha_M &= 0 \\ y(M+1) + y(M)\alpha_1 + ... + y(1)\alpha_M &= 0 \\ &\vdots \\ y(N-1) + y(N-2)\alpha_1 + ... + y(N-M-1)\alpha_M &= 0 \end{aligned} \tag{3.6}$$

This set can be solved directly if $N = 2M$ or solved approximately by the method of least squares if $N > 2M$. After the α's are determined, the M $e^{s(k)}$'s are found as the roots of Eq. (3.5). The Equations (3.2) then become linear equations in the M a's, thus they are determined from the first M of these equations or by applying the least square technique to the entire set [5] [8].

Several variations to this method to deal with noise have been applied to Prony Method. We next discuss two of these approaches.

Kumaresan and Tufts proposed in [13], [25] that the observed data sequence $y(n)$ consists of Eq. (3.1) plus white Gaussian noise $w(n)$:

$$y(n) = \sum_{k=1}^{M} a(k)e^{s(k)n} + w(n) \tag{3.7}$$

Then, a singular value decomposition (SVD) is applied to set of linear prediction equations using data in the backward direction:

$$\begin{bmatrix} y(1) & y(2) & \cdots & y(L) \\ y(2) & y(3) & \cdots & y(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ y(N-L) & y(N-L+1) & \cdots & y(N-1) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_L \end{bmatrix} = - \begin{bmatrix} y(0) \\ y(2) \\ \vdots \\ y(N-L-1) \end{bmatrix} \tag{3.8}$$

$$A\boldsymbol{\alpha} = -\mathbf{h}$$

where L is chosen to satisfy the inequality $M \leqslant L \leqslant N-M$. The SVD is applied because of the presence of noise in the data and the equations are solved in the least square sense injecting perturbations in $\boldsymbol{\alpha}$ [14], [13] However, moderately large values of L are essential in improving the accuracy of the pole location estimates [12], [13]. The main idea behind this approach is to apply SVD to A and then find a truncated SVD solution by setting the smaller singular values of A to zero. The SVD of $\boldsymbol{\alpha}$ is given by

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_L]^T = -\sum_{k=1}^{M} \sigma_k \mathbf{u}_k^{\dagger} \mathbf{h} \mathbf{v}_k \tag{3.9}$$

where $\sigma_k, k = 1, 2, .., L$ are the singular values of $A.\mathbf{v}_k, k = 1, 2, ..., L$ and $\mathbf{u}_k, k = 1, 2, ..., N - L$ are the eigenvectors of $A^\dagger A$ and $AA^\dagger$ respectively and $\dagger$ stands for matrix complex conjugate transpose. The polynomial

$$\Lambda(z) = 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + ... + \alpha_L z^{-L} \tag{3.10}$$

has $v_k, 1 \leq k \leq L$ roots.

Kumaresan and Tufts have shown in [12], [13] that when $\sigma_k = 0$ (when there is no measurement noise) this polynomial has exactly M roots outside the unit circle, which are the reciprocals of $e^{s(k)}, 1 \leq k \leq M$. The extra roots of $\Lambda(z)$ are guaranted to be inside the unit circle. Based on this observation, the roots of $\Lambda(z)$ of signals with noise are estimated from

$$e^{s(k)} = v_k^{-1} \; ; \; 1 \leq k \leq M$$

where $v_k \; ; \; 1 \leq k \leq M$ are the roots of $\Lambda(z)$ outside the unit circle.

The effect of using a truncated SVD is to increase the signal-to-noise-ratio (SNR) in the data prior to obtaining the solution vector $\boldsymbol{\alpha}$.

Another approach to deal with noisy data proposed in [14] is to use the conventional Prony method (Eq. 3.6) with L as in Eq. 3.8. The result is a set of L exponentials which are candidates for signal components. Then, with high-order Prony calculation, determine the best subset of size M. A best subset of M exponentials is one for which a linear combination of the M exponentials best approximates the observed data using a least squares criterion. If M is unknown a priori, a search with increasing $\widehat{M}$ in one to one and computed the error respectively, the stop criterion is when the rate of decrease of the error with increasing values of $\widehat{M}$ is small. The value $\widehat{M}$ that make this small error is taken as M.

3.1.2 Matrix Pencil Method

The singular values decomposition (SVD) step in Matrix Pencil method acts as a filtering process, thus this method can extract dynamic information from noisy data [26]. When noise is present, the zero singular values are perturbed by the noise and take small positive values [15]. Now, the singular value decomposition provides an effective way of noise filtering. The singular values below some specified threshold are considered to be caused by noise and need to be set as zero. In the matrix pencil method, L is a free parameter subject to $M \leq L \leq N - M$.

Another approach to deal with the problem of estimating the unknowns in Eq. (3.7) is the Matrix Pencil method. Unlike Prony Method the Matrix Pencil

Method is a one step process for determining the roots z_k in the polynomial Eq. (3.10) The matrix Pencil method is not only computationally more efficient, but it also has better statistical properties for the estimates of z_k than the Prony Method.

The matrix pencil arises of combining two functions defined on a common interval, with a scalar parameter, λ. A pencil function of $g(n)$ and $h(n)$ is given by [26]:

$$f(n, \lambda) = g(n) + \lambda h(n)$$

and parameterized by λ and where $g(n)$ is not permitted to be a scalar multiple of $h(n)$. The pencil-of-functions enables extracting information about z_k, given $y(n)$, when $g(n)$, $h(n)$, and λ are approximately selected.

Similar to Prony methods, the Matrix Pencil method is designed to fit the signal with a model of the form

$$y(n) = x(n) + w(n) = \sum_{k=1}^{M} a(k) e^{s(k)n} + w(n) = \sum_{k=1}^{M} a_k z_k^n + w(n)$$

where $a_k = a(k)$ and $z_k^n = e^{s(k)n}$

If the data is assumed noise free, then the following equalities hold

$$
\begin{aligned}
Y_2 &= \begin{bmatrix} x(1) & x(2) & \cdots & x(L) \\ x(2) & x(3) & \cdots & x(L+1) \\ \vdots & \vdots & & \vdots \\ x(N-L) & x(N-L+1) & \cdots & x(N-1) \end{bmatrix} \\
&= \begin{bmatrix} \sum_{k=1}^{M} a_k z_k^1 & \sum_{k=1}^{M} a_k z_k^2 & \cdots & \sum_{k=1}^{M} a_k z_k^L \\ \sum_{k=1}^{M} a_k z_k^2 & \sum_{k=1}^{M} a_k z_k^3 & \cdots & \sum_{k=1}^{M} a_k z_k^{L+1} \\ \vdots & \vdots & & \vdots \\ \sum_{k=1}^{M} a_k z_k^{N-L} & \sum_{k=1}^{M} a_k z_k^{N-L+1} & \cdots & \sum_{k=1}^{M} a_k z_k^{N-1} \end{bmatrix} \\
&= \begin{bmatrix} a_1 & a_2 & \cdots & a_M \\ a_1 z_1 & a_2 z_2 & \cdots & a_M z_M \\ \vdots & \vdots & & \vdots \\ a_1 z_1^{N-L} & a_2 z_2^{N-L} & \cdots & a_M z_M^{N-L} \end{bmatrix} \begin{bmatrix} 1 & z_1 & \cdots & z_1^{L-1} \\ 1 & z_2 & \cdots & z_2^{L-1} \\ \vdots & \vdots & & \vdots \\ 1 & z_M & \cdots & z_M^{L-1} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_M \\ \vdots & \vdots & & \vdots \\ z_1^{N-L-1} & z_2^{N-L-1} & \cdots & z_M^{N-L-1} \end{bmatrix}_{(N-L)\times M} \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ & & & a_M \end{bmatrix}_{M\times M} \times \\
&\quad \begin{bmatrix} z_1 & & & \\ & z_2 & & \\ & & \ddots & \\ & & & z_M \end{bmatrix}_{M\times M} \begin{bmatrix} 1 & z_1 & \cdots & z_1^{L-1} \\ 1 & z_2 & \cdots & z_2^{L-1} \\ \vdots & \vdots & & \vdots \\ 1 & z_M & \cdots & z_M^{L-1} \end{bmatrix}_{M\times L} \\
&= \Xi_1 A Z \Xi_2
\end{aligned}
\tag{3.11}
$$

and the same form that Equation (3.11) Y_1 is defined by:

$$
\begin{aligned}
Y_1 &= \begin{bmatrix} x(0) & x(1) & \cdots & x(L-1) \\ x(1) & x(2) & \cdots & x(L) \\ \vdots & \vdots & & \vdots \\ x(N-L-1) & x(N-L) & \cdots & x(N-2) \end{bmatrix} \\
&= \Xi_1 A \Xi_2
\end{aligned}
\tag{3.12}
$$

where L is referred to as the pencil parameter. Then, considering the matrix pencil

$$Y_2 - \lambda Y_1 = \Xi_1 A \left(Z - \lambda \mathbf{I}\right) \Xi_2$$

The rank of $Y_2 - \lambda Y_1$ will be M. Note that, if $\lambda = z_k, k = 1...M$, the kth row of $Z - \lambda \mathbf{I}$ is zero, and the rank of this matrix is $M - 1$. Hence, the

parameters z_k are determined by solving an ordinary eigenvalue problem of the form [15], [9], [26]:

$$Y_1^{\dagger} Y_2 - \lambda \mathbf{I}$$

To deal with noise, the total-squares Matrix Pencil has been found to be superior that Prony Method [15]. In the implementation of the method [26],[15], it is used the matrix Y from the noisy signal $y(n)$, i.e.

$$Y = \begin{bmatrix} y(0) & y(1) & \cdots & y(L) \\ x(1) & x(2) & \cdots & y(L+1) \\ \vdots & \vdots & & \vdots \\ y(N-L-1) & y(N-L) & \cdots & y(N-1) \end{bmatrix}$$

where L is chosen between $N/3$ to $N/2$ for an efficient noise filtering. Applying a SVD to Y

$$Y = USV^T$$

the value M is selected considering the eigenvalues σ_k of S such that

$$\frac{\sigma_k}{\sigma_{\max}} \approx 10^{-p}$$

where the ratio of singular values σ_k that are below 10^{-p} is considered noise singular values and they are not considered in modeling. In this form $M+1$ to L small eigenvalues are discarded. Therefore, S_M is from the M columns of S and V_M contains only M dominant right-singular vectors of V Then Y_1 and Y_2 are obtained by

$$Y_1 = US_M V_{M1}^T$$

$$Y_2 = US_M V_{M2}^T$$

where V_{M1} is obtained from V_M with the last row of V_M deleted and V_{M2} is obtained by removing the first row of V_M. Then, the eigenvalues are obtained from

$$Y_1^{\dagger} Y_2 - \lambda \mathbf{I}$$

Once M and the z_k are known, the residues a_k are solved from a least-square problem.

The results of Prony Method and Matrix Pencil in [15], [26] are different under noise. It can be shown in [15] that under noise, the statistical variance of the poles z_k for the MP method is always less than that of Prony Method.

3.2 Models of Random Process and Yule-Walker Equations

Models for process differ from those for deterministic signals in two ways. First, a random process may only be characterized statistically and the values of $y(n)$ are only known in a probabilistic sense. The second difference is in the characteristic of the signal that is used as the input to the system that is used to model $y(n)$. For a random process the input signal must be a random process.

3.2.1 Autoregressive Moving Average Processes

An Autoregressive Moving Average (ARMA) process is a causal linear shift-invariant filter with white noise $v(n)$ input. The rational system function with p poles and q zeros is given by

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{q} b(k) z^{-1}}{1 + \sum_{k=1}^{p} a(k) z^{-1}} \tag{3.13}$$

It can be shown that an ARMA process of order (p, q) referred to as an ARMA(p.q) process has a power spectrum for the output $y(n)$ of the form

$$R_y(e^{j\omega}) = \sigma_v^2 \frac{|B(e^{j\omega})|^2}{|A(e^{j\omega})|^2} \tag{3.14}$$

where it is assumed that the filter is stable, the output process $y(n)$ will be wide-sense stationary and with $R_v(e^{j\omega}) = \sigma_v^2$ [6].

Since $y(n)$ and $v(n)$ are related by the linear constant coefficient difference equation

$$y(n) + \sum_{l=1}^{p} a(l) y(n-l) = \sum_{l=0}^{q} b(l) v(n-l) \tag{3.15}$$

the autocorrelation of $y(n)$ and the cross-correlation between $y(n)$ and $v(n)$ satisfy the same difference equation. Then, multiplying both sides of Eq.(3.15) by $y^*(n-k)$ and taking the expected value, then it is obtained [32], we get

$$r_y(k) + \sum_{l=1}^{p} a(l) r_y(k-l) = \sum_{l=0}^{q} b(l) r_{vy}(k-l) \tag{3.16}$$

where it is assumed that $v(n)$ is wide-sense stationary and

$$r_{vy}(k-l) = \mathbf{E}\left[v(n-l)x^*(n-k)\right]$$

However, by writing the cross-correlation $r_{vy}(k-l)$ in terms of the unit sample response of the filter, it is obtained:

$$\mathbf{E}\left[v(n-l)x^*(n-k)\right] = \sigma_v^2 h^*(l-k)$$

Then, equation (3.16) can be expressed by

$$r_y(k) + \sum_{l=1}^{p} a(l) r_y(k-l) = \sigma_v^2 \sum_{l=0}^{q} b(l) h^*(l-k) \tag{3.17}$$

Because $h(n)$ is causal, equation (3.17) can be written as

$$r_y(k) + \sum_{l=1}^{p} a(l) r_y(k-l) = \begin{cases} \sigma_v^2 \sum_{l=0}^{q-k} b(l+k) h^*(l) & 0 \leq k \leq q \\ 0 & k > q \end{cases} \tag{3.18}$$

These are the Yule-Walker equations.

Modified Yule-Walker Equation method.

In this work, only the coefficients of polynomial $A(z)$ are of interest, since it is aimed at extracting the modal content of signal, which are the roots of $A(z)$. A [28] Modified Yule-Walker Equation (MYWE) method is obtained by expressing Equation (3.18) in matrix form for $k = q+1, ..., q+p$:

$$\begin{bmatrix} r_y(q) & r_y(q-1) & \cdots & r_y(q-p+1) \\ r_y(q+1) & r_y(q) & \cdots & r_y(q-p+2) \\ \vdots & \vdots & & \vdots \\ r_y(q+p-1) & r_y(q+p-2) & \cdots & r_y(q) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} r_y(q+1) \\ r_y(q+2) \\ \vdots \\ r_y(q+p) \end{bmatrix}$$

This procedure results in p lineal equations with p unknowns , $a(k)$ This set of lineal equations can be solved for $a(k)$ using different methods as discussed below.

Least Squares MYWE method.

A problem that arises in practical applications using the MYWE is that often the autocorrelations are unknown and they have to be estimated using a sample realization of the process. In [29], [28], [32] a better estimate of $r_y(k)$ is obtained by using a long data sequence; this gives a set of linear equations in the unknown $a(k)$:

$$\begin{bmatrix} r_y(q) & r_y(q-1) & \cdots & r_y(q-p+1) \\ r_y(q+1) & r_y(q) & \cdots & r_y(q-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ r_y(L-1) & r_y(L-2) & \cdots & r_y(L-p) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} r_y(q+1) \\ r_y(q+2) \\ \vdots \\ r_y(L) \end{bmatrix}$$

where $L > p$ and a least squares solution can be found.

3.2.2 Autoregressive Processes

An ARMA(p,0) process is called an autoregressive process of order p. In this case $y(n)$is generated by filtering white noise with an all-pole filter of the form:

$$H(z) = \frac{b(0)}{1+} \sum_{k=1}^{p} a(k) z^{-k}$$

The power spectrum in an AR(p) give for the output $y(n)$ is

$$R_y(e^{j\omega}) = \sigma_v^2 \frac{|b(0)|^2}{|A(e^{j\omega})|^2}$$

where $R_v(e^{j\omega}) = \sigma_v^2$.

The Yule-Walker equations for an AR(p) process are[6]:

$$r_y(l) + \sum_{l=1}^{p} a(l) r_y(k-l) = \sigma_v^2 |b(0)|^2 \delta(k) \quad ; \quad k \geq 0 \tag{3.19}$$

Yule-Walker Equation method

The Yule Walker Equation is obtained for expressing Equation (3.19) in matrix form for $k = 1, 2, ..p$,using the conjugate symmetry of $r_y(k)$:

$$\begin{bmatrix} r_y(0) & r_y^*(1) & \cdots & r_y^*(p-1) \\ r_y(1) & r_y(0) & \cdots & r_y^*(p-2) \\ \vdots & \vdots & & \vdots \\ r_y(p-1) & r_y(p-2) & \cdots & r_y(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} r_y(1) \\ r_y(2) \\ \vdots \\ r_y(p) \end{bmatrix}$$

3.3 Adaptive Algorithms

An adaptive algorithms approach has been applied in [31], [30], [34], [36] to track low frequency modes in near real-time using measured ambient power system data. These adaptive techniques are based on the assumption that there is random noise at the input. In what follows we give a brief review of various adaptive filtering techniques which are relevant to this research.

3.3.1 Least-Mean Squares (LMS) Adaptive Filtering Technique

With the assumption that input of the power system is approximately white noise, adaptive filtering techniques can be applied to track the frequency and damping associated with the electromechanical modes in the system. This is acomplished by finding the weights associated with the specific adaptive filter block applied to the ambient data.

The LMS filter is used in a linear predictor to make a whitening filtering. The dominant modes of the whitening filter correspond to the dominant modes of the time-varying power system. The LMS algorithm is used in [31] to adapt the weights in order to track time variations in the power system. The LMS process is described by:

$$\begin{aligned} \widehat{y}(n) &= \widehat{\mathbf{w}}^H(n)\mathbf{y}(n-1) \\ e(n) &= y(n) - \widehat{y}(n) \\ \widehat{\mathbf{w}}(n+1) &= \widehat{\mathbf{w}}(n) + \mu\mathbf{y}(n-1)e^*(n) \end{aligned} \tag{3.20}$$

where $\mathbf{y}(n-1) = [y(n-1)\ y(n-2)\ ...\ y(n-M)]^T$ is a vector of M past values of power system data $y(n)$, $\widehat{\mathbf{w}}(n) = [w_1(n)\ \ w_2(n)\ ...w_M(n)]^T$ is a vector of time-varying filter weights, $\widehat{y}(n)$ is filter output, $e(n)$ is the approximate white noise filter output, M is the order of the filter and μ is the step-size parameter. The figure (3.1) shows a diagram of LMS adaptive filtering.

Then, the LMS filtering process consist of generated a linear prediction $\widehat{y}(n)$ using a set of past data $\mathbf{y}(n-1)$ and a weight vector $\widehat{\mathbf{w}}(n)$, and error $e(n)$ is used to regenerate the adaptive process.

For the convergence of LMS tracking algorithm this is guaranted for wide-sense stationary data and small μ, the step-size can be adjusted to the interval $0 < \mu < (2/total\ MSV)$, where

$$total\ \text{MSV}=\frac{M}{N}\sum_{n=0}^{N-1} y^2(n)$$

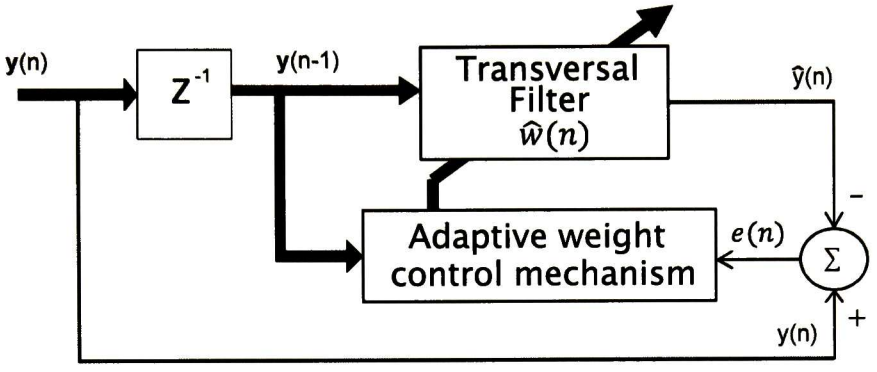


Figure 3.1: Diagram of LMS adaptive filtering

The vector $\widehat{\mathbf{w}}(n)$ have the coefficients of the polynomial on z-domain, its roots and converting they to s-plane give knowledge of the modes of the system, this polynomial is represented by:

$$A(z) = \begin{bmatrix} 1 & z^{-1} & \cdots & z^{-M} \end{bmatrix} \begin{bmatrix} 1 \\ -\widehat{\mathbf{w}}(n) \end{bmatrix}$$

3.3.2 Normalized Least Mean Squares (NLMS) Filtering

The main characteristic of this to improve the convergence time of the estimates compared to the LMS. In LMS algorithm [1], [7], the correction $\mu\mathbf{y}(n-1)e^*(n)$ applied to the weight vector $\widehat{\mathbf{w}}(n)$ at iteration $n+1$ is directly proportional to the input vector $\mathbf{y}(n-1)$. Therefore, when $\mathbf{y}(n-1)$ is large, the LMS algorithm experiences a gradient noise amplification problem. To deal with this problem is used the NMLS. In this method the correction applied to the weight vector $\widehat{\mathbf{w}}(n)$ at iteration $n+1$ is normalized with respect to the squared Euclidian norm of the input vector $\mathbf{y}(n-1)$.

The problem to solve in NLMS filtering is that given the input vector $\mathbf{y}(n-1)$ and $y(n)$, determine the weight vector $\widehat{\mathbf{w}}(n+1)$ so as to minimize the squared Euclidian norm of the change

$$\Delta\widehat{\mathbf{w}}(n+1) = \widehat{\mathbf{w}}(n+1) - \widehat{\mathbf{w}}(n)$$

in the weight vector $\widehat{\mathbf{w}}(n+1)$ with respect to its previous value $\widehat{\mathbf{w}}(n)$.

In [1], [7], the minimization is given by

$$\Delta\widehat{\mathbf{w}}(n+1) = \frac{1}{\|\mathbf{y}(n-1)\|^2}\mathbf{y}(n-1)e^*(n)$$

A positive real scaling factor ϵ is introduced to exercise control over the change in the weight vector from one iteration to the next without changing its direction. Then, $\Delta\widehat{\mathbf{w}}(n+1)$ is given by

$$\Delta\widehat{\mathbf{w}}(n+1) = \frac{\epsilon}{a + \|\mathbf{y}(n-1)\|^2}\mathbf{y}(n-1)e^*(n)$$

where $0 < \epsilon < 2$ to ensure the convergence in the mean square [7], and a is a positive constant to avoid numerical difficulties when $\|\mathbf{y}(n-1)\|^2$ be a small value. Now, it is possible write $\widehat{\mathbf{w}}(n+1)$ by

$$\widehat{\mathbf{w}}(n+1) = \widehat{\mathbf{w}}(n) + \frac{\epsilon}{a + \|\mathbf{y}(n-1)\|^2}\mathbf{y}(n-1)e^*(n) \tag{3.21}$$

Equation (3.21) replace equation (3.20) for $\widehat{\mathbf{w}}(n+1)$ for Normalized LMS. The normalized LMS algorithm exhibits a rate of convergence that is potentially faster than of LMS algorithm for both uncorrelated and correlated data input.

3.3.3 Adaptive Step-Size Least Mean Squares (ASLMS) Filtering

In [30], an adaptive step-size Least Mean Square algorithm (ASLMS) for estimating the electromechanical modes is proposed in which the step-size μ is adaptively controlled to same proposed that NLMS, namely, the time of the estimates is improved compared to the LMS. The main equations associated with the ASLMS, added to the set of three LMS equations in (3.20) is given by

$$\mu(n) = \mu(n-1) + \rho e(n)\gamma^H(n)\mathbf{y}(n-1)$$

where γ^H is the gradient vector and ρ is a small positive constant which controls the update of the step size parameter.

3.3.4 Recursive Least Squares (RLS) algorithms

In [36] and [34] R3LS and RRLS algorithms are proposed respectively. In [34] an Robust Recursive Least Square (RRLS) algorithm is proposed for find the parameters of AR in the case of ambient data. Zhou and Pierre said that the prediction model for the AR takes the same form as in the prony analysis, it is based on [27], [16]. Thus,

$$\begin{aligned}\widehat{y}(n \mid \Theta) &= -\alpha_1 y(n-1) - \alpha_2 y(n-2) - \ldots - \alpha_a y(n-a) \\ &= \psi^T(n)\Theta\end{aligned}$$

can be used as a prediction model for both signal types (Ringdown data and Ambient data).

The RRLS algorithm in[34] is described by

$$\begin{aligned}
\widehat{\Theta}(n) &= \widehat{\Theta}(n-1) + R(n)p'(\varepsilon(n)) \\
\varepsilon(n) &= y(n) - \psi^T(n)\widehat{\Theta}(n-1) \\
P(n) &= \lambda^{-1}P(n-1) - \lambda^{-1}R(n)\psi^T(n)P(n-1)p''(\varepsilon(n)) \\
R(n) &= \frac{P(n-1)\psi(n)}{\lambda + p''(\varepsilon(n))\psi^T(n)P(n-1)\psi(n)}
\end{aligned}$$

where $p(\varepsilon) = \varepsilon^2/2$ is a scalar-valued function as in Eq. (2.7) and $p'(\varepsilon) = \varepsilon$ if $|\varepsilon| \leq \Delta\sigma$ or $p'(\varepsilon) = 0$ if $|\varepsilon| > \Delta\sigma$. $p''(\varepsilon) = 1$ if $|\varepsilon| \leq \Delta\sigma$ or $p''(\varepsilon) = 0$ if $|\varepsilon| > \Delta\sigma$. σ is the standard deviation of ε, and Δ is a user chosen positive constant. For more details see [34].

3.4 Method of Stochastic Realization by using a linear matrix inequality

There are different approaches to solve for the unknown parameters of models, such as correlation methods and prediction error methods. Others methods that have been recently used, focus in determining the unknown parameters of stochastic state-space models. They are numerically robusts because they utilize techniques such as QR-factorization and singular value decomposition (SVD). As a result, the system modes are computed directly from system matrix A of the stochastic state space model obtained. In [11], it is presented a method of stochastic realization by using the deterministic realization theory and a linear matrix inequality (LMI) satisfied by the state covariance matrix. This results in a stochastic realization algorithm based on a finite covariance data.

Chapter 4
Method of Stochastic Subspace Identification for Ambient Modal Estimation

System identification provides a meaningful engineering alternative to physical modeling. Compared to models obtained from physics, system identification models have a limited validity and working range and in some cases have no direct physical meaning[21]. *But, they are relatively easy to obtain and use and even more importantly, these models are simple enough to make model-based control system design mathematically (and also practically) tractable.*
This chapter discusses the development of subspace identification methods for modal estimation from measured data. The underlying assumption is that small motions of the power system can be described by a set of ordinary difference equations in discrete time or differential equations in continue time.

4.1 Background

It is assumed that matrices $A \in \mathbb{R}^{p \times j}$, $B \in \mathbb{R}^{q \times j}$ are given. The elements of a row of one of given matrices can be considered as the coordinates of a vector in the $j-$dimensional ambient space.

4.1.1 Orthogonal Projections

Orthogonal projection projects the row space of the matrix A onto the row space of the matrix B[11]:

$$A/B = AB^T(BB^T)^\dagger B$$

where $\bullet^\dagger$ denotes the Moore-Penrose pseudo-inverse of the matrix, and $\bullet$. A/B is a shorthand for the projection of matrix $A \in \mathbb{R}^{p \times j}$ on the row space of the matrix B.

4.1.2 Statistical Assumptions

Consider two sequences $a_k \in \mathbb{R}^{n_a}$ and $e_k \in \mathbb{R}^{n_e}, k = 0, 1, ..., j$. If the sequence e_k is a zero mean sequence, independent of a_k it is straightforward to show that

$$\mathbf{E}\left[e_k\right] = 0 \tag{4.1}$$

$$\mathbf{E}\left[a_k e_k^T\right] = 0 \tag{4.2}$$

In subspace identification [20],[18] it is assumed that long time series of data available ($j \longrightarrow \infty$) are available, and that the data is ergodic. Due to ergodicity and the infinite number of data, the expectation operator $\mathbf{E}$ (average over an infinite number of experiments) can be replaced with the difference operator $\mathbf{E}_j$ applied to the sum of variables (average over one, infinitely long, experiment)[11]. Then, for the correlation between a_k and e_k, we have

$$\mathbf{E}\left[a_k e_k^T\right] = \lim_{j \longrightarrow \infty} \left[\frac{1}{j} \sum_{i=0}^{j} a_i e_i^T\right] = \mathbf{E}_j \left[\sum_{i=0}^{j} a_i e_i^T\right] \tag{4.3}$$

and

$$\mathbf{E}_j\left[\bullet\right] = \lim_{j \longrightarrow \infty} \frac{1}{j}\left[\bullet\right] \tag{4.4}$$

These equations lie at the heart of the subspace approach. Let u_k be a sequence of inputs and e_k be a disturbance. If it is assumed that an infinite number of data are available and that the data are ergodic and that u_k and e_k are independent, it can be proved that [21]:

$$\mathbf{E}_j \left[\sum_{i=0}^{j} u_i e_i^T\right] = 0 \tag{4.5}$$

or

$$\mathbf{E}_j\left[\mathbf{u}\mathbf{e}^T\right] = 0 \tag{4.6}$$

where $\mathbf{u} = (u_0 \;\; u_1 \;\; ... \;\; u_j)$; $\mathbf{e} = (e_0 \;\; e_1 \;\; ... \;\; e_j)$,which implies that the input vector $\mathbf{u}$ is perpendicular to the noise vector $\mathbf{e}$ [6]. So, geometrically (and for $j \longrightarrow \infty$) it is possible to state that the row vectors of disturbances are perpendicular to the row vector of inputs (and to other variables not correlated with the noise).

This property is used in subspace identification algorithms to filter out noise effects [19].

4.1.3 Covariance

The covariance $\Phi_{[A,B]}$ between two matrices $A \in \mathbb{R}^{p \times j}$ and $B \in \mathbb{R}^{q \times j}$ is defined in the statistical (stochastic) framework as [11]:

$$\Phi_{[A,B]} = \mathbf{E}_j \left[AB^T \right] \tag{4.7}$$

from which it follows that

$$\Phi_{[A,B]} \simeq \frac{1}{j} AB^T \tag{4.8}$$

Then, for the orthogonal projection:

$$\begin{aligned} A/B &= \Phi_{[A,B]} \Phi^{\dagger}_{[B,B]} B \\ &= \left[\frac{1}{j} AB^T \right] \left[\frac{1}{j} BB^T \right]^{\dagger} B \\ &= AB^T \left[BB^T \right]^{\dagger} B \end{aligned} \tag{4.9}$$

and

$$AB^T \longleftarrow \Phi_{[A,B]} \tag{4.10}$$

4.2 Stochastic System

Stochastic subspace identification algorithms compute state space models from given

output data. A stochastic linear system can be described by the state space model [11], [21]:

$$\begin{aligned} x_{k+1} &= Ax_k + \omega_k \\ y_k &= Cx_k + v_k \qquad k = 0, 1, ...s \end{aligned} \tag{4.11}$$

where $x \in \mathbb{R}^n$ is the state vector, $y \in \mathbb{R}^l$ is the observation vector, $\omega \in \mathbb{R}^n$ is the plant noise vector, and $v \in \mathbb{R}^l$ os the observation noise vector. The system matrices are $A \in \mathbb{R}^{n \times n}$ $C(t) \in \mathbb{R}^{l \times n}$ ω_k and v_k are zero mean, white noise vector with covariance matrices:

$$\mathbf{E} \left[\begin{pmatrix} \omega_p \\ v_p \end{pmatrix} \begin{pmatrix} \omega_q & v_q \end{pmatrix} \right] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq} \tag{4.12}$$

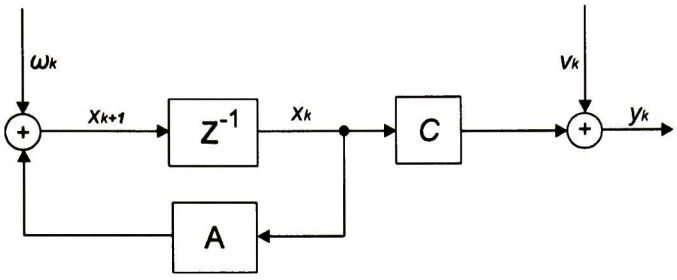


Figure 4.1: A linear time-invariant stochastic system with outputs y_k and states x_k, described by the matrices A, C and the covariances matrices Q, S, R.

where $Q \in \mathbb{R}^{n\times n}$,$S \in \mathbb{R}^{n\times l}$, $R \in \mathbb{R}^{l\times l}$ so that the second order statistics of the output of the model and of the given output are equal [18].

The system described by Equation(4.11) is schematically shown in Figure (4.1).

We note that, if the matrix A is stable the process becomes a stationary process [11] with:

$$\begin{aligned} \mathbf{E}\left[x_k\right] &= 0 \\ \mathbf{E}\left[x_k(x_k)^T\right] &= \Pi \end{aligned} \tag{4.13}$$

where the state covariance matrix Π is independent of time k.

4.2.1 Forward Model

Since ω_k and v_k are zero mean white noise vector sequences independent of x_k, we have

$$\begin{aligned} \mathbf{E}\left[x_k v_k^T\right] &= 0 \\ \mathbf{E}\left[x_k \omega_k^T\right] &= 0 \end{aligned} \tag{4.14}$$

and the Lyapunov equation for the state covariance matrix Π is:

$$\begin{aligned} \Pi &= \mathbf{E}\left[(Ax_k + \omega_k)(Ax_k + \omega_k)^T\right] \\ &= A\mathbf{E}\left[x_k(x_k)^T\right]A^T + \mathbf{E}\left[\omega_k \omega_k^T\right] \\ &= A\Pi A^T + Q \end{aligned} \tag{4.15}$$

To obtain the covariance matrices, let

$$\Lambda_i = \mathbf{E}\left[y_{k+i}y_k^T\right]$$

Then,

$$\begin{aligned}\Lambda_0 &= \mathbf{E}\left[y_k(y_k)^T\right] \\ &= \mathbf{E}\left[\left(Cx_k + v_k\right)\left(Cx_k + v_k\right)^T\right] \\ &= C\mathbf{E}\left[x_k(x_k)^T\right]C^T + \mathbf{E}\left[v_k v_k^T\right] \\ &= C\Pi C^T + R\end{aligned} \tag{4.16}$$

Further, defining

$$\begin{aligned}G &= \mathbf{E}\left[x_{k+1}y_k^T\right] \\ &= \mathbf{E}\left[\left(Ax_k + \omega_k\right)\left(Cx_k + v_k\right)^T\right] \\ &= A\mathbf{E}\left[x_k(x_k)^T\right]C^T + \mathbf{E}\left[\omega_k(v_k)^T\right] \\ &= A\Pi C^T + S\end{aligned} \tag{4.17}$$

From [21], for $i = 1, 2, ...$:

$$\begin{aligned}\Lambda_i &= CA^{i-1}G \\ \Lambda_{-i} &= G^T(A^{i-1})^T C^T\end{aligned} \tag{4.18}$$

This last observation, indicates that the output covariances can be considered as Markov parameters of the deterministic linear time invariant system A, G, C, Λ_0[11], [21].

4.2.2 Block Hankel Matrices

Output Hankel Matrices are useful in stochastic subspace identification algorithms. Based on the foregoing analysis, the output measurements are put into a block Hankel matrix in the form

$$Y_{0|i-1} = Y_{past} = \begin{pmatrix} y_0 & y_1 & y_2 & \cdots & y_{j-2} & y_{j-1} \\ y_1 & y_2 & y_3 & \cdots & y_{j-1} & y_j \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y_{i-1} & y_i & y_{i+1} & \cdots & y_{i+j-3} & y_{i+j-2} \end{pmatrix}$$

$$Y_{i|2i-1} = Y_{future} = \begin{pmatrix} y_i & y_{i+1} & y_{i+2} & \cdots & y_{i+j-2} & y_{i+j-1} \\ y_{i+1} & y_{i+2} & y_{i+3} & \cdots & y_{i+j-1} & y_{i+j} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y_{2i-1} & y_{2i} & y_{2i+1} & \cdots & y_{2i+j-3} & y_{2i+j-2} \end{pmatrix}$$

$$Y_{0|2i-1} = \begin{pmatrix} Y_{0|i-1} \\ Y_{i|2i-1} \end{pmatrix} = \begin{pmatrix} Y_{past} \\ Y_{future} \end{pmatrix} = \begin{pmatrix} Yp \\ Y_f \end{pmatrix} \tag{4.19}$$

The matrices Y_p (the past outputs) and Y_f (the future outputs) are structured as i block rows, Y_p and Y_f splitting $Y_{0|2i-1}$ into two equal parts of i block rows. The subscripts $Y_{0|2i-1}$ $Y_{0|i-1}$ and $Y_{i|2i-1}$ are the first and last element in the first column of the block Hankel matrix. The matrices Y_p^+ and Y_f^- are defined by shifting the border between past and future data one row down in Equation (4.19), as:

$$\begin{aligned} Y_p^+ &= Y_{0|i} \\ Y_f^- &= Y_{i+1|2i-1} \end{aligned}$$

The number of block rows i is typically equal to $2M/l$ [2], where l is the number of outputs of system, M represents the expected maximal order of the system, and it is assumed that $i > n$.The number of columns j is typically equal to $s - 2i + 1$, where s is the number of measurements of outputs, which implies that all given data samples are used. However, in subspace identification it is assumed that there are long time series of data available ($j \longrightarrow \infty$), for statistical reasons [19].

4.2.3 System Related Matrices

Another important matrix for stochastic subspace identification algorithms is the observability matrix. The extended ($i > n$) observability matrix $\Im_i$ (where the subscript i denotes the number of block rows) can be written as

$$\partial_i = \begin{pmatrix} C \\ CA \\ CA^2 \\ \cdots \\ CA^{i-1} \end{pmatrix} \tag{4.20}$$

where $\partial_i \in \mathbb{R}^{li \times n}$.Moreover, it is assumed that the pair $\{A, C\}$ is observable, that is $\partial(A, C)$ has rank n [18]. Also, the reversed extended stochastic controllability matrix $\hat{\mathbf{C}}_i$ is defined [21] as:

$$\hat{\mathbf{C}}_i = \begin{pmatrix} A^{i-1}G & A^{i-2}G & \cdots & AG & G \end{pmatrix} \tag{4.21}$$

where $\hat{\mathbf{C}}_i \in \mathbb{R}^{n \times li}$.It is assumed that the pair $\left\{A, Q^{1/2}\right\}$ is controllable. This implies that all the dynamical modes of the system are excited by the plant (or process) noise [18]. The block Toeplitz matrices are constructed from the output covariance matrices.

Denote the block Toeplitz covariance matrix L_i as

$$L_i = \begin{pmatrix} \Lambda_0 & \Lambda_{-1} & \Lambda_{-2} & \cdots & \Lambda_{1-i} \\ \Lambda_1 & \Lambda_0 & \Lambda_{-1} & \cdots & \Lambda_{2-i} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Lambda_{i-1} & \Lambda_{i-2} & \Lambda_{i-3} & \cdots & \Lambda_0 \end{pmatrix}$$

where $L_i \in \mathbb{R}^{li \times li}$.

Also, the corresponding block Toeplitz cross covariance matrix H_i is denoted by

$$H_i = \begin{pmatrix} \Lambda_i & \Lambda_{i-1} & \Lambda_{i-2} & \cdots & \Lambda_1 \\ \Lambda_{i+1} & \Lambda_i & \Lambda_{i-1} & \cdots & \Lambda_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Lambda_{2i-1} & \Lambda_{2i-2} & \Lambda_{2i-3} & \cdots & \Lambda_i \end{pmatrix}$$

where $H_i \in \mathbb{R}^{li \times li}$

Based on the above developments and using the theory in section 4.1.3 the ith output covariance matrix become:

$$\Lambda_i = \mathbf{E}_j \left[\sum_{k=0}^{j-1} y_{k+i} y_k^T \right] = \lim_{j \longrightarrow \infty} \frac{1}{j} \left[(Y_{i|i})(Y_{0|0})^T \right] = \Phi_{[Y_{i|i}, Y_{0|0}]}$$

The following equalities hold true:

$$\lim_{j \longrightarrow \infty} \frac{1}{j} \begin{pmatrix} Y_{0|i-1} \\ Y_{i|2i-1} \end{pmatrix} \begin{pmatrix} (Y_{0|i-1})^T & (Y_{i|2i-1})^T \end{pmatrix} = \begin{pmatrix} L_i & H_i^T \\ H_i & L_i \end{pmatrix}$$

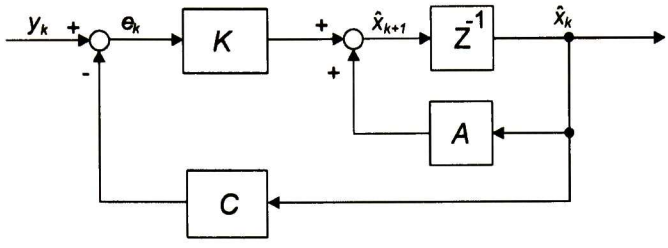


Figure 4.2: Block Diagram of Kalman filter

$$\begin{aligned} L_i &= \Phi_{[Y_p,Y_p]} = \Phi_{[Y_f,Y_f]} \\ H_i &= \Phi_{[Y_f,Y_p]} \end{aligned} \tag{4.22}$$

4.2.4 Kalman Filter States

The Kalman filter algorithm for the discrete-time stochastic system described by Equation (4.11) and (4.12) is given by an one-step-ahead estimator [11]:

$$\widehat{x}_{k+1} = A\widehat{x}_k + K_k(y_k - C\widehat{x}_k) \tag{4.23}$$

Figure(4.2) displays a block diagram of Kalman filter that produces the one-step ahead estimates $\widehat{x}_{k+1}$.

Therefore, the reconstruction error $\widetilde{x} = x - \widehat{x}$ is governed by:

$$\begin{aligned} \widetilde{x}_{k+1} &= x_{k+1} - \widehat{x}_{k+1} = A(x_k - \widehat{x}_k) + \omega_k - K_k(y_k - C\widehat{x}_k) \\ &= A\widetilde{x}_k + \omega_k - K_k(y_k - C\widehat{x}_k) \\ &= A\widetilde{x}_k + \omega_k - K_k(Cx_k + v_k - C\widehat{x}_k) \\ &= (A - K_kC)\widetilde{x}_k + \omega_k - K_kv_k \\ &= \begin{pmatrix} I & -K_k \end{pmatrix} \left(\begin{pmatrix} A \\ C \end{pmatrix} \widetilde{x}_k + \begin{pmatrix} \omega_k \\ v_k \end{pmatrix} \right) \end{aligned} \tag{4.24}$$

It should be observed that noise properties are taken into account; the criterion is to minimize the variance of the estimation error, which is denoted by $\widetilde{P}_{k+1}$is given by

$$\widetilde{P}_{k+1} = \mathbf{E}\left[\left(\widetilde{x}_{k+1} - \mathbf{E}\left[\widetilde{x}_{k+1} \right] \right) \left(\widetilde{x}_{k+1} - \mathbf{E}\left[\widetilde{x}_{k+1} \right] \right)^T \right] \tag{4.25}$$

where the mean value of $\widetilde{x}$ is obtained from Equation (4.24):

$$\mathbf{E}\left[\widetilde{x}_{k+1}\right] = (A - K_k C)\mathbf{E}\left[\widetilde{x}_k\right] \tag{4.26}$$

Note that since $\mathbf{E}\left[x_0\right] = 0$, the mean value of the reconstruction error is zero for all times $k \geq 0$, independent of K if $\widehat{x}_0 = 0$. Because $\widetilde{x}_k$ is independent of ω_k and v_k , it follows from Equation (4.24) that:

$$\begin{aligned}
\mathbf{a} &= \begin{pmatrix} I & -K_k \end{pmatrix}, \\
\mathbf{b} &= \begin{pmatrix} A \\ C \end{pmatrix}, \\
\mathbf{c} &= \begin{pmatrix} \omega_k \\ v_k \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
\widetilde{P}_{k+1} &= \mathbf{E}\left[\widetilde{x}_{k+1}(\widetilde{x}_{k+1})^T\right] \\
&= \mathbf{E}\left[(\mathbf{ab}\widetilde{x}_k + \mathbf{ac})(\mathbf{ab}\widetilde{x}_k + \mathbf{ac})^T\right] \\
&= \mathbf{abE}\left[\widetilde{x}_k(\widetilde{x}_k)^T\right]\mathbf{b}^T\mathbf{a}^T + \mathbf{acc}^T\mathbf{a}^T \\
\widetilde{P}_{k+1} &= \begin{pmatrix} I & -K_k \end{pmatrix}\left(\begin{pmatrix} A \\ C \end{pmatrix}\widetilde{P}_k\begin{pmatrix} A \\ C \end{pmatrix}^T + \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}\right)\begin{pmatrix} I \\ -K_k^T \end{pmatrix} \\
&= \begin{pmatrix} I & -K_k \end{pmatrix}\begin{pmatrix} AP_kA^T + Q & AP_kC^T + S \\ CP_kA^T + S^T & CP_kC^T + R \end{pmatrix}\begin{pmatrix} I \\ -K_k^T \end{pmatrix}
\end{aligned} \tag{4.27}$$

By using the idea of the completion of squares, it also follows that $\alpha^T\widetilde{P}_k\alpha$ is minimized by K_k satisfying:

$$K_k(C\widetilde{P}_kC^T + R) = A\widetilde{P}_kC^T + S \tag{4.28}$$

for any α.Further ,if $CP_kC^T + R$ is positive definite then:

$$K_k = \left(A\widetilde{P}_kC^T + S\right)(C\widetilde{P}_kC^T + R)^{-1} \tag{4.29}$$

Inserting Equation (4.29) into Equation (4.27), one has

$$\widetilde{P}_{k+1} = A\widetilde{P}_kA^T + Q - (A\widetilde{P}_kC^T + S)(R + C\widetilde{P}_kC^T)^{-1}(C\widetilde{P}_kA^T + S^T) \tag{4.30}$$

The reconstruction defined by recursive Equations (4.23),(4.29) and (4.30) is called the Kalman filter:

$$\widehat{x}_{k+1} = A\widehat{x}_k + K_k(y_k - C\widehat{x}_k) \tag{4.31}$$

$$K_k = \left(A\widetilde{P}_kC^T + S\right)(C\widetilde{P}_kC^T + R)^{-1} \tag{4.32}$$

$$\widetilde{P}_{k+1} = A\widetilde{P}_k A^T + Q - (A\widetilde{P}_k C^T + S)(R + C\widetilde{P}_k C^T)^{-1}(C\widetilde{P}_k A^T + S^T) \quad (4.33)$$

For the purpose of stochastic subspace identification algorithms [11], [21], a different form of these recursive Kalman filter equations is more useful. With A, C, Q, S, R given, the matrices Π, G and Λ_0 can be computed through the Equations (4.15), (4.16), and (4.17). Now define the transformation:

$$\widetilde{P}_k \longleftarrow \Pi - P_k$$

From Equation (4.32) we have

$$\begin{aligned} \Pi - P_{k+1} &= A\Pi A^T - AP_kA^T + (\Pi - A\Pi A^T) \quad (4.34) \\ &\quad - \left[(A\Pi C^T + S) - AP_kC^T\right]\left[(C\Pi C^T + R) - CP_kC^T\right]^{-1} \\ &\quad \times \left[(A\Pi C^T + S) - AP_kC^T\right]^T \\ \Pi - P_{k+1} &= \Pi - A\Pi A^T - AP_kA^T + (\Pi - A\Pi A^T) \\ &\quad -(G - AP_kC^T)(\Lambda_0 - CP_kC^T)^{-1}(G - AP_kC^T)^T \\ P_{k+1} &= AP_kA^T + (G - AP_kC^T)(\Lambda_0 - CP_kC^T)^{-1}(G - AP_kC^T)^T \end{aligned}$$

Then, the Kalman filter can be expresed as

$$\begin{aligned} x_k &= Ax_k + Ke_k \quad (4.35) \\ y_k &= Cx_k + e_k \end{aligned}$$

where e_k is the innovation process with covariance matrix

$$\mathbf{E}\left[e_k \ (e_k \ )^T\right] = (\Lambda_0 - CP_kC^T)$$

Equation (4.32) is the (forward) Kalman gain, and P_{k+1} in Equation (4.34) is the forward state covariance matrix, which can be determined as the stabilizing solution of the forward Riccati equation (4.34).The state space Equation (4.35) is called a forward innovation model for the process y.

Forward non-steady state Kalman filter.

The non-steady state Kalman filter state estimate $\widehat{x}_k$ is defined by the following recursive formulas:

$$\widehat{x}_{k+1} = A\widehat{x}_k + K_k(y_k - C\widehat{x}_k) \quad (4.36)$$

$$K_k = (G - AP_kC^T)(\Lambda_0 - CP_kC^T)^{-1} \quad (4.37)$$

$$P_{k+1} = AP_kA^T + (G - AP_kC^T)(\Lambda_0 - CP_kC^T)^{-1}(G - AP_kC^T)^T \quad (4.38)$$

in which

$$\widehat{x}_0 = 0$$

$$P_0 = \mathbf{E}\left[\widehat{x}_0(\widehat{x}_0)^T\right] = 0$$

The output measurement $y_0, ..., y_{k-1}$
can be adopted as initial conditions, then with this initial conditions is possible obtain the next.

From Equation (4.36) with $k = 0$,

$$\begin{aligned} \widehat{x}_1 &= A\widehat{x}_0 + K_0(y_0 - C\widehat{x}_0) \\ &= K_0y_0 \end{aligned}$$

with

$$\begin{aligned} K_0 &= (G - AP_0C^T)(\Lambda_0 - CP_0C^T)^{-1} \\ &= G\Lambda_0^{-1} \\ &= \hat{\mathbf{C}}_1L_1^{-1} \end{aligned}$$

In addition, $\widehat{x}_1$ and P_1 are calculated from

$$\begin{aligned} \widehat{x}_1 &= G\Lambda_0^{-1}y_0 \\ &= \hat{\mathbf{C}}_1L_1^{-1}y_0 \end{aligned} \quad (4.39)$$

$$\begin{aligned} P_1 &= AP_0A^T + (G - AP_0C^T)(\Lambda_0 - CP_0C^T)^{-1}(G - AP_0C^T)^T \\ &= G\Lambda_0^{-1}G^T \\ &= \hat{\mathbf{C}}_1L_1^{-1}\left(\hat{\mathbf{C}}_1\right)^T \end{aligned} \quad (4.40)$$

For $k = 1$,it is obtained

$$\begin{aligned} K_1 &= (G - AP_1C^T)(\Lambda_0 - CP_1C^T)^{-1} \\ &= (G - AG\Lambda_0^{-1}G^TC^T)(\Lambda_0 - CG\Lambda_0^{-1}G^TC^T)^{-1} \end{aligned} \quad (4.41)$$

Inserting Equation (4.18) into Equation (4.41), we obtain:

$$K_1 = (G - AG\Lambda_0^{-1}\Lambda_{-1})(\Lambda_0 - \Lambda_1\Lambda_0^{-1}\Lambda_{-1})^{-1} \tag{4.42}$$

Then, for $\widehat{x}_2$:

$$\begin{aligned}
\widehat{x}_2 &= A\widehat{x}_1 + K_1(y_1 - C\widehat{x}_1) \\
&= AG\Lambda_0^{-1}y_0 + (G - AG\Lambda_0^{-1}\Lambda_{-1})(\Lambda_0 - \Lambda_1\Lambda_0^{-1}\Lambda_{-1})^{-1}(y_1 - \Lambda_1\Lambda_0^{-1}y_0) \\
&= \begin{pmatrix} AG\Lambda_0^{-1} + (G - AG\Lambda_0^{-1}\Lambda_{-1})(\Lambda_0 - \Lambda_1\Lambda_0^{-1}\Lambda_{-1})^{-1}(-\Lambda_1\Lambda_0^{-1}) \\ (G - AG\Lambda_0^{-1}\Lambda_{-1})(\Lambda_0 - \Lambda_1\Lambda_0^{-1}\Lambda_{-1})^{-1} \end{pmatrix}^T \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} \\
&= \begin{pmatrix} AG & G \end{pmatrix} \begin{pmatrix} \Lambda_0^{-1} - (\Lambda_0^{-1}\Lambda_{-1})\Omega(-\Lambda_1\Lambda_0^{-1}) & (\Lambda_0^{-1}\Lambda_{-1})\Omega \\ \Omega(-\Lambda_1\Lambda_0^{-1}) & \Omega \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} \\
\text{with } \Omega &= (\Lambda_0 - \Lambda_1\Lambda_0^{-1}\Lambda_{-1})^{-1}
\end{aligned}$$

Using now the lemma of inverse of block matrices [10] , $\widehat{x}_2$ is calculated as

$$\begin{aligned}
\widehat{x}_2 &= \begin{pmatrix} AG & G \end{pmatrix} \begin{pmatrix} \Lambda_0 & \Lambda_{-1} \\ \Lambda_1 & \Lambda_0 \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} \\
&= \hat{\mathbf{C}}_2 L_2^{-1} \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}
\end{aligned} \tag{4.43}$$

Finally, using Equation (4.39) and (4.40), this expression is generalized to

$$\widehat{x}_k = \hat{\mathbf{C}}_k L_k^{-1} \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{k-1} \end{pmatrix} \tag{4.44}$$

This is an explicit form to obtain the state estimates.
For $k = 1$:

$$P_1 = \hat{\mathbf{C}}_1 L_1^{-1} \left(\hat{\mathbf{C}}_1\right)^T$$

Now, following a similar procedure used for $\widehat{x}_k$,it is possible to write

$$P_k = \hat{\mathbf{C}}_k L_k^{-1} \left(\hat{\mathbf{C}}_k\right)^T \tag{4.45}$$

Note that of this expression is true for $k = p$, it is also true for $k = p + 1$.
Using now the inverse lemma of block matrices, we obtain.

$$
\begin{aligned}
P_{p+1} &= \hat{\mathbf{C}}_{p+1} L_{p+1}^{-1} \left(\hat{\mathbf{C}}_{p+1}\right)^T \\
&= \begin{pmatrix} A\hat{\mathbf{C}}_p & G \end{pmatrix} \begin{pmatrix} L_p & \hat{\mathbf{C}}_p^T C^T \\ C\hat{\mathbf{C}}_p & \Lambda_0 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\mathbf{C}}_p^T A^T \\ G^T \end{pmatrix} \\
&= \begin{pmatrix} A\hat{\mathbf{C}}_p & G \end{pmatrix} \begin{pmatrix} L_p^{-1} + L_p^{-1}\hat{\mathbf{C}}_p^T C^T \Delta^{-1} C \hat{\mathbf{C}}_p L_p^{-1} & -L_p^{-1}\hat{\mathbf{C}}_p^T C^T \Delta^{-1} \\ -\Delta^{-1} C \hat{\mathbf{C}}_p L_p^{-1} & \Delta^{-1} \end{pmatrix} \\
&\quad \times \begin{pmatrix} \hat{\mathbf{C}}_p^T A^T \\ G^T \end{pmatrix} \\
\text{where } \Delta &= \Lambda_0 - C\hat{\mathbf{C}}_p L_p^{-1} \hat{\mathbf{C}}_p^T C^T = \Lambda_0 - C\hat{\mathbf{C}}_p C^T \\
P_{p+1} &= A\hat{\mathbf{C}}_p L_p^{-1} \hat{\mathbf{C}}_p^T A^T + (G - A\hat{\mathbf{C}}_p L_p^{-1} \hat{\mathbf{C}}_p^T C^T)\Delta^{-1}(G - A\hat{\mathbf{C}}_p L_p^{-1} \hat{\mathbf{C}}_p^T C^T)^T \\
&= AP_pA^T + (G - AP_pC^T)(\Lambda_0 - CP_pC^T)^{-1}(G - AP_pC^T)^T
\end{aligned}
$$

This last equation clearly indicates that the matrix P_{p+1} calculated from Equation (4.38) and from Equation (4.45) are the same.

Then, it is possible estimates the states in explicitly form from

$$
\widehat{x}_k = \hat{\mathbf{C}}_k L_k^{-1} \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{k-1} \end{pmatrix} \tag{4.46}
$$

and the explicit solution of the covariance matrix P_k can be written as

$$
P_k = \hat{\mathbf{C}}_k L_k^{-1} \left(\hat{\mathbf{C}}_k\right)^T \tag{4.47}
$$

The assumption $P_0 = 0$ is the same as the assumption $\widetilde{P}_0 = \Pi$. This means that the state will be estimated exactly when an infinite amount of output data is available [11].

The significance of Equation (4.46) and (4.47) is that it indicates how the Kalman filter state estimates $\widehat{x}_k$ can be written as a linear combination of the past output measurement [18]. In stochastic subspace identification [21], [11], the forward Kalman filter state sequence can be recovered from the expression

$$
\widehat{X}_i = \begin{pmatrix} \widehat{x}_i & \widehat{x}_{i+1} & \cdots & \widehat{x}_{i+j+1} \end{pmatrix} = \hat{\mathbf{C}}_i L_i^{-1} Y_p \tag{4.48}
$$

4.2.5 Stochastic Identification

In this subsection it is shown how calculate the row space of the state sequence $\widehat{X}_i$ and the column space of the extended observability matrix ∂_i directly from the output data, without any previous knowledge of the system matrices. The system matrices can then be extracted from $\widehat{X}_i$ and ∂_i.

Assume now that:

1. The plant noise ω_k and the observed noise v_k are not zero.
2. The number of measurement from outputs y_k goes to infinity $j \rightarrow \infty$.
3. The weighting matrices $W_1 \in \mathbb{R}^{li \times li}$ and $W_2 \in \mathbb{R}^{j \times j}$ are such that W_1 is of full rank and W_2 obeys: rank(Y_p)=rank($Y_p W_2$).

The orthogonal projection from the row space of the Hankel matrix Y_f onto the row space of the Hankel matrix Y_p is denoted by

$$Z_i = Y_f / Y_p \tag{4.49}$$

and the singular value decomposition is:

$$W_1 Z_i W_2 = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 S_1 V_1^T \tag{4.50}$$

From Equation (4.49), and using Equations (4.22), it is possible to show that:

$$\begin{aligned} Z_i &= Y_f / Y_p \\ &= \Phi_{[Y_f, Y_p]} \Phi^{\dagger}_{[Y_p, Y_p]} Y_p \\ &= H_i L_i^{-1} Y_p \end{aligned} \tag{4.51}$$

and

$$
\begin{aligned}
H_i &= \begin{pmatrix} \Lambda_i & \Lambda_{i-1} & \Lambda_{i-2} & \cdots & \Lambda_1 \\ \Lambda_{i+1} & \Lambda_i & \Lambda_{i-1} & \cdots & \Lambda_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Lambda_{2i-1} & \Lambda_{2i-2} & \Lambda_{2i-3} & \cdots & \Lambda_i \end{pmatrix} \\
&= \begin{pmatrix} CA^{i-1}G & CA^{i-2}G & \cdots & CAG & CG \\ CA^{i}G & CA^{i-1}G & \cdots & CA^{2}G & CAG \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ CA^{2i-2}G & CA^{2i-3}G & \cdots & CA^{i}G & CA^{i-1}G \end{pmatrix} \\
&= \begin{pmatrix} C \\ CA \\ CA^2 \\ \cdots \\ CA^{i-1} \end{pmatrix} \begin{pmatrix} A^{i-1}G & A^{i-2}G & \cdots & AG & G \end{pmatrix} \\
&= \partial_i \hat{\mathbf{C}}_i
\end{aligned} \tag{4.52}
$$

Replacing H_i into Z_i:

$$Z_i = H_i L_i^{-1} Y_p = \partial_i \hat{\mathbf{C}}_i L_i^{-1} Y_p$$

Using now Equation (4.48), yields

$$Z_i = Y_f / Y_p = \partial_i \widehat{X}_i \tag{4.53}$$

Equation (4.53), shows that Z_i is equal to the product of the extended observability matrix and the forward Kalman filter state sequence. In other words, the row space of the states $\widehat{X}_i$ can be found by orthogonally projecting the row space of the future outputs Y_f on the row space of the past outputs Y_p [20], [19].

Since $W_1 Z_i W_2$ is of rank n, the number of singular values different from zero will be equal to the order of the system in Equation (4.11) [11]. Also, since W_1 is of full rank, and since the rank of $Y_p W_2$ is equal to the rank of Y_p, then the rank of $W_1 Z_i W_2$ is equal to the rank of Z_i.which in turn is equal to n. This is due the fact that Z_i is equal to the product of a matrix with n columns and a matrix with n rows.

Combining equations (4.50) and (4.53), it is obtained

$$
\begin{aligned}
W_1 Z_i W_2 &= W_1 \partial_i \widehat{X}_i W_2 \\
&= U_1 S_1^{1/2} S_1^{1/2} V_1^T
\end{aligned} \tag{4.54}
$$

where

$$\widehat{X}_i = S_1^{1/2} V_1^T W_2^{-1} \tag{4.55}$$

is the state sequence $\widehat{X}_i$,and

$$\Theta_i = W_1^{-1} U_1 S_1^{1/2} \tag{4.56}$$

is the extended observability matrix.

Alternatively, the forward sequence $\widehat{X}_i$ can be written as

$$\widehat{X}_i = \Theta_i^{\dagger} Z_i \tag{4.57}$$

4.2.6 System Matrices

Using Equation (4.50) it is possible to determine the order of system by taking the largest singular values. Using a similar approach to that in Equation (4.49), it is possible to show that the following holds:

$$Z_{i-1} = Y_f^- / Y_p^+ = \Theta_{i-1} \widehat{X}_{i+1} \tag{4.58}$$

Deleting the last l (number of outputs) rows of Θ_i, result in

$$\Theta_{i-1} = \underline{\Theta}_i \tag{4.59}$$

From the above discussion and using Eq. (4.58) and Eq. (4.59) we can calculate $\widehat{X}_{i+1}$ as follows

$$\widehat{X}_{i+1} = \Theta_{i-1}^{\dagger} Z_{i-1} \tag{4.60}$$

From these results, we can partition the state vector in the extended form

$$\begin{pmatrix} \widehat{X}_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} \left(\widehat{X}_i \right) + \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} \tag{4.61}$$

This set of equations can be solved for A and C.Since the Kalman filter residuals ρ_w and ρ_v are uncorrelated with $\widehat{X}_i$, it seems natural to solve this set of equations in a least squares sense (since the least squares residuals are orthogonal and thus uncorrelated with the regressors $\widehat{X}_i$). The solution of A and C is

$$\begin{pmatrix} A \\ C \end{pmatrix} = \begin{pmatrix} \widehat{X}_{i+1} \\ Y_{i|i} \end{pmatrix} \left(\widehat{X}_i \right)^{\dagger}$$

An important observation is that the identified sequence determined by A, G, C, Λ_0 should be a positive sequence. If this is not true, a spectral factor can not be computed, and the set of forward realizations of the covariance sequence is empty [21]. It turns out that if one starts from raw data, even when it was generated by simulation of a linear stochastic system, the positive real

condition of the identified covariance sequence is hardly ever satisfied. This is due to either the finite amount of data available or to the real data is often not generated by a linear stochastic time invariant system.

An alternative way to ensure positive realness of the estimated covariance sequence is possible when the covariance of the process and measurement noise are recover from the residuals ρ_w and ρ_v of equation (4.61) as

$$E_j\left[\begin{pmatrix}\rho_w\\ \rho_v\end{pmatrix}\begin{pmatrix}\rho_w^T & \rho_v^T\end{pmatrix}\right] = \begin{pmatrix}Q_i & S_i\\ S_i^T & R_i\end{pmatrix}$$

The subscript i indicates that the estimated covariances are not the steady state covariances, but are the non-steady state covariance matrices of the non-steady state Kalman filter equation:

$$P_{i+1} = AP_iA^T + Q_i$$

$$G = AP_iC^T + S_i$$

$$\Lambda_0 = CP_iC^T + R_i$$

When $i \to \infty$, which is upon convergence of the Kalman filters, we have [21] that $Q_i \to Q, S_i \to S, R_i \to R$.This result guaranteed positive real covariance sequence.

The matrices G and Λ_0 can now be extracted from the solution of the Lyapunov equation for Π as follows:

$$\Pi = A\Pi A^T + Q$$

Further,

$$G = A\Pi C^T + S$$

$$\Lambda_0 = C\Pi C^T + R$$

The form can be converted into a forward innovation form by solving the Riccati equation:

$$P = APA^T + (G - APC^T)(\Lambda_0 - CPC^T)^{-1}(G - APC^T)^T$$

for matrix P.Also of interest, the Kalman gain can be computed from:

$$K_k = (G - AP_kC^T)(\Lambda_0 - CP_kC^T)^{-1}$$

4.2.7 Canonical Variation Algorithm

The weights in Eq(4.50) can be used to derive the Canonical Variation Algorithm as described below.

When $j \neq \infty$ or when the data-generating system is not linear, the singular values from $W_1 Z_i W_2$, are all different from zero. In that case the row space of Z_i is of dimension i and the order has to be chosen equal to the number of dominant singular values. The complexity reduction step is then truly a reduction of the dimension of the row space of Z_i, and the weights W_1 and W_2 play an important role in determining which part of the original row space of Z_i is retained.

The canonical correlation analysis (CCA) [11] is a technique of multivariable statistical analysis that clarifies the mutual dependence between two sets of variables by finding a new coordinate system in the space of each set of variables. The principal idea of canonical correlation analysis is find vectors w_1 and z_1 with the maximum mutual correlation in the linear spaces spanned by x and y (which are two vectors of zero mean random variables), and define (w_1, z_1) as the first coordinates in the new system. Then find w_2 and z_2 such that their correlation is maximum under the assumption that they are uncorrelated with the first coordinates (w_1, z_1). This procedure is continued until two new coordinate systems are determined.

Having obtained the covariances matrices of Y_p and Y_f,the canonical correlations are computed by the SVD as

$$\begin{aligned} \mu^C &= \Phi_{[Y_f,Y_f]}^{-1/2} \Phi_{[Y_f,Y_p]} \Phi_{[Y_p,Y_p]}^{-1/2} \\ &= \begin{pmatrix} \upsilon_1 & \upsilon_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \nu_1^T \\ \nu_2^T \end{pmatrix} \end{aligned} \tag{4.62}$$

The cosines of the principal angles between Y_f and Y_p are given by the elements of Σ_1 and the principal directions α in Y_p and β in Y_f are given by

$$\alpha = \nu_1^T \Phi_{[Y_p,Y_p]}^{-1/2}(Y_p)$$

$$\beta = \upsilon_1^T \Phi_{[Y_f,Y_f]}^{-1/2}(Y_f)$$

The order of the system is determined from the number of principal angles different from $\frac{1}{2}\pi$.

It should be remarked that equation (4.62) from Canonical Variation Algorithm corresponds to the Equation (4.51) of the subspace algorithm with the following weights [18]:

$$\begin{aligned} W_1 &= \Phi^{-1/2}_{[Y_f,Y_f]} \\ W_2 &= I_j \end{aligned} \tag{4.63}$$

From previous results, we also have

$$W_1 Z_i W_2 = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 S_1 V_1^T$$

which is denoted as M^C :

$$\begin{aligned} M^C &= \Phi^{-1/2}_{[Y_f,Y_f]} Z_i I_j \\ &= \Phi^{-1/2}_{[Y_f,Y_f]} \Phi_{[Y_f,Y_p]} \Phi^{\dagger}_{[Y_p,Y_p]} Y_p \end{aligned}$$

Finding the covariance matrix of the weighted projection yields

$$\begin{aligned} M^C (M^C)^T &= \Phi^{-1/2}_{[Y_f,Y_f]} Z_i Z_i^T \Phi^{-1/2}_{[Y_f,Y_f]} \\ &= \Phi^{-1/2}_{[Y_f,Y_f]} \Phi_{[Y_f,Y_p]} \Phi^{\dagger}_{[Y_p,Y_p]} Y_p (Y_p^T) \Phi^{\dagger}_{[Y_p,Y_p]} \Phi^T_{[Y_f,Y_p]} \Phi^{-1/2}_{[Y_f,Y_f]} \\ &= \Phi^{-1/2}_{[Y_f,Y_f]} \Phi_{[Y_f,Y_p]} \Phi^{-1}_{[Y_p,Y_p]} \Phi_{[Y_p,Y_f]} \Phi^{-1/2}_{[Y_f,Y_f]} \\ &= (U_1 S_1 V^T)(U_1 S_1 V^T)^T \\ &= U_1 S_1^2 U_1^T = \mu^C (\mu^C)^T = \upsilon_1 \Sigma_1^2 \upsilon_1 \end{aligned}$$

This implies that the left singular vectors and singular values of both matrices are equal.

CVA for Modal Identification

The numerical algorithm developed take advantage of numerical tools as the RQ-decomposition, and SVD decomposition. The algorithm can be summarized as follows:

1.- Calculate the orthogonal projections of Y_f onto Y_p :

$$Z_i = Y_f \, / \, Y_p$$

$$Z_{i-1} = Y_f^- / Y_p^+$$

Calculate the RQ^T decomposition with the matrices Q^T orthonormal and R lower triangular as follows. Defining the matrix H

$$H = \frac{1}{\sqrt{j}}\left(Y_{0|2i}\right) = RQ^T$$

with $R \in 2il \times 2il$ and $Q^T \in 2il \times j$, we can show that

$$\begin{aligned} H &= \frac{1}{\sqrt{j}}\begin{pmatrix} Y_p \\ Y_f \end{pmatrix} = \frac{1}{\sqrt{j}}\begin{pmatrix} Y_p^+ \\ Y_f^- \end{pmatrix} = \frac{1}{\sqrt{j}}\begin{pmatrix} Y_{0|i-1} \\ Y_{i|i} \\ Y_{i+1|2i-1} \end{pmatrix} \begin{matrix} li \\ l \\ l(i-1) \end{matrix} \\ &= RQ^T = \begin{matrix} li \\ l \\ l(i-1) \end{matrix} \overset{\begin{matrix} li & l & l(i-1) \end{matrix}}{\begin{pmatrix} R_{1,1} & 0 & 0 \\ R_{2,1} & R_{2,2} & 0 \\ R_{3,1} & R_{3,2} & R_{3,3} \end{pmatrix}} \overset{j}{\begin{pmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \end{pmatrix}} \end{aligned} \tag{4.64}$$

Using the orthogonal project concept in Eq (4.9) and using Eq (4.64) for Y_p and Y_f, the orthogonal projections are:

$$\begin{aligned} Z_i &= Y_f \,/\, Y_p \\ &= \Phi_{[Y_f,Y_p]}\Phi^{\dagger}_{[Y_p,Y_p]}Y_p \\ &= \left(R_{2:3,1:3}Q^TQR^T_{1:1,1:3}\right)\left(R_{1:1,1:3}Q^TQR^T_{1:1,1:3}\right)^{\dagger}R_{1:1,1:3}Q^T \\ &= R_{2:3,1:1}Q_1^T \end{aligned}$$

$$\begin{aligned} Z_{i-1} &= Y_f^- \,/\, Y_p^+ \\ &= \left(R_{3:3,1:3}Q^TQR^T_{1:2,1:3}\right)\left(R_{1:2,1:3}Q^TQR^T_{1:2,1:3}\right)^{\dagger}R_{1:2,1:3}Q^T \\ &= R_{3:3,1:2}Q^T_{1:2} \end{aligned}$$

2.- Calculate the SVD from projections with weights

$$W_1 Z_i W_2 = \Phi^{-1/2}_{[Y_f,Y_f]}Z_i = USV^T$$

3.- Determine the system order by taking the greatest singular values and partition the SVD for obtain U_1 and S_1.

4.-Determine $\Im_i$ and $\Im_{i-1}$using

$$\Im_i = W_1^{-1}U_1S_1^{1/2} = \Phi^{-1/2}_{[Y_f,Y_f]}U_1S_1^{1/2}$$

$$\Im_{i-1} = \underline{\Im}_i$$

5.-Determine $\widehat{X}_i$ and $\widehat{X}_{i+1}$

$$\widehat{X}_i = \Im_i^{\dagger}Z_i$$

$$\widehat{X}_{i+1} = \Theta_{i-1}^{\dagger} Z_{i-1}$$

6.-Solve the set of linear equations for A and C

$$\begin{pmatrix} \widehat{X}_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} \left(\widehat{X}_i \right) + \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix}$$

7.-The eigenvalues from matrix A are the characteristic values of the system, then the corresponding frequency and ratio are calculated, in this form:

$$\begin{aligned} s_i &= \alpha_i \pm j\beta_i, \text{ with } \alpha_i = \frac{1}{T} \ln |z_i| , \\ \beta_i &= \frac{1}{T} \tan^{-1} \left\{ \frac{z_{Ii}}{z_{Ri}} \right\} \text{ and } z_i = z_{Ri} + j z_{Ii} \end{aligned}$$

Futhermore, the natural frequency and damping are calculated from:

$$\omega_n = \left(\alpha_i^2 + \beta_i^2 \right)^{1/2}$$

and

$$\delta_i = -\alpha_i / \omega_n$$

Chapter 5
Monitoring of Power System Oscillatory Dynamics using RLS

Adaptive filtering algorithms are well suited for mode estimation in power system because they rely heavily on the ambient noise that characterizes the data[29] ,[24]. *The algorithms discussed in this chapter offer the potential of being a useful tool for feature extraction and modal identification of measured data and can be applied in real time. Advantages of these algorithms are the ability to track time varying parameters, small computational burden at each step thus enabling use of small computers for on-line data analysis, and as an aid in model building where form of parameter time variation may suggest cause of a model inadequacy.*

5.1 RLS for mode identification using Ambient Data

Power systems are random in nature and inherently nonstationary. Random load variations act as constant low-level excitation to the electromechanical dynamics of the system which shows up as ambient noise in field measurements [30]. We recall from Chapter 2 that the ambient noise in a power system is caused by a white noise input. Because white noise input is relatively small, the small-signal dynamics are assumed to dominate the response. This in turn, allows using a whitening filter approach.

Based on this analysis, the ambient noise output of the power system is the result of an approximately stationary white noise input in the frequency band of interest over an analysis window. Then, the output of power system can be filtered with a whitening filter [24].

Figure (5.1) shows a block diagram of whitening filtering of power system data $y(n)$.

Let the power system transfer function $H(z)$ be given by the AR model

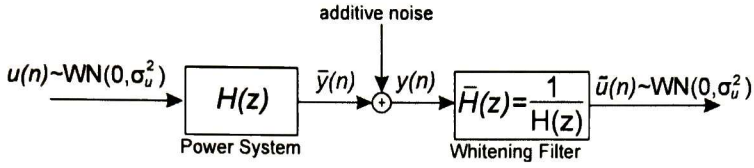


Figure 5.1: Block diagram of whitening filtering of power system data $y(n)$.

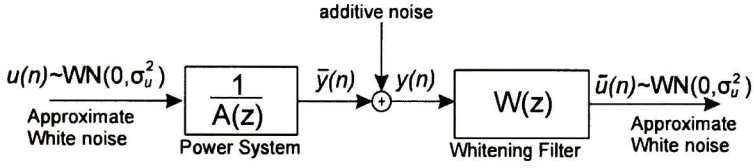


Figure 5.2: Block diagram of whitening filter where the power system transfer function is $1/A(z)$ and the whitening filter transfer function is given by $w(z)$.

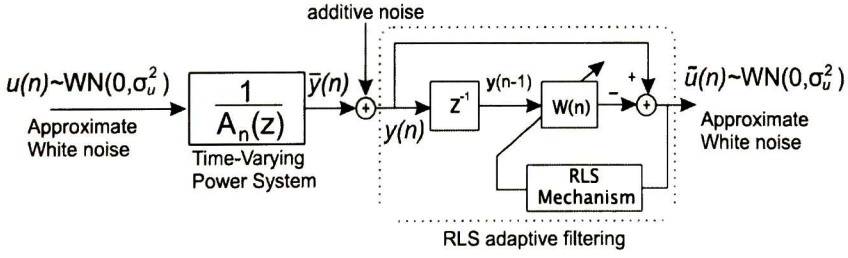


Figure 5.3: Diagram of Adaptive whitening filter of power system data using the RLS adaptive filtering.

$$H(z) = \frac{1}{A(z)}$$

It then follows that, it is possible to use a FIR filter described by the transfer function [29], [31], [30]

$$W(z) = 1 + w(1)z^{-1} + ... + w(M)z^{-M}$$

to whiten the power system data. As a consequence, the poles of system, roots of $A(z)$, can be estimated as the dominant roots of $W(z)$. Figure (5.2) clarifies this fact.

Alternatively, the whitening filter function transfer $W(z)$ can be computed by RLS adaptive filters where the coefficients of polynomial $W(z)$ is given by the RLS adaptive filter weights. Figure (5.3) shows a diagram of a RLS adaptive filter used to whitening filter; the details for RLS adaptive filters used to whitening filter will be explain in the next section.

5.2 Least-Saquares Method

The advantage of using the method of least squares to solve the linear filtering problem is that it does not make assumptions on the statistics of the inputs applied to the filter [1], [7].

The method of least-squares use a linear regression model structure (Eq(2.9)) to a lineal predictor where the response $y(n)$ is modeled as

$$\widehat{y}(n) = \sum_{k=1}^{M} w^*(k)y(n-k) + \mu(n)$$

where the $w(1), ..., w(M)$ are the unknown parameters of the model, and $\mu(n)$ represents the measurement error to which the statistical nature of the phenomenon is ascribed. The measurement error $\mu(n)$ is an unobservable random variable that is introduced into the model to account for its inaccuracy. It is assumed that the measurement error process $\mu(n)$ is white with zero mean [7]. That is, $\mathbf{E}[\mu(n)] = 0$ for all n.

The mean of the response $y(n)$, in theory, is uniquely determined by the model $\sum_{k=1}^{M} w^*(k)y(n-k)$.

The solution for the weights $w(1), ..., w(M)$ can be obtained by using a linear transversal filter:

$$e(n) = y(n) - \widehat{y}(n) \tag{5.1}$$

$$e(n) = y(n) - \sum_{k=1}^{M} w^*(k)y(n-k) \tag{5.2}$$

In the method of least-squares, the weights $w(1), ..., w(M)$ of the transversal filter are choose so as to minimize a criterion function that consists of the sum of error squares [1]:

$$N_{n_2-n_1}(\Theta) = \sum_{n=n_1}^{n_2} |e(n)|^2$$

where $\Theta = w(1), ..., w(M)$. This sum may also be viewed as an error energy.

The problem to solve in least-squares is to minimize the criterion function $N_{n_2-n_1}(\Theta)$ with respect Θ of the transversal filter and using Eq. (5.1). In the minimization, Θ is constant during the interval $n_1 \leq n \leq n_2$. The filter obtained is termed a linear least squares filter.

The criterion function for a set of data finite is

$$\begin{aligned} N_{N-M}(\Theta) &= \sum_{n=M}^{N} |e(n)|^2 \\ &= \sum_{n=M}^{N} e(n)e^*(n) \end{aligned} \tag{5.3}$$

From optimization theory [6], it can be shown that $N_{N-M}(\Theta)$ depends only on the coefficients, $w(k)$. Therefore, the coefficients that minimize this squared error may be found by setting the partial derivatives of $N_{N-M}(\Theta)$ with respect to $w^*(k)$ equal to zero. This yields

$$\begin{aligned} \frac{\partial N_{N-M}(\Theta)}{\partial w^*(k)} &= \sum_{n=M}^{N} \frac{\partial[e(n)e^*(n)]}{\partial w^*(k)} = 0 \\ &= \sum_{n=M}^{N} \frac{\partial[y(n) - \sum_{k=1}^{M} w^*(k)y(n-k)]}{\partial w^*(k)} e^*(n) = 0 \\ &= \sum_{n=M}^{N} y(n-k)e^*(n) = 0, \qquad k = 1, ..., M \end{aligned} \tag{5.4}$$

Based on principle of orthogonality (Eq. (5.4)), it is possible to demonstrate that:

$$\sum_{n=M}^{N} \widehat{y}(n)e^*(n) = 0 \tag{5.5}$$

which states that when a transversal filter operates in its least-squares condition, the least-squares estimate of the desired response, produced at the filter output and represented by the time series $\widehat{y}(n)$, and the minimum estimation error time series $e(n)$ are perpendiculars to each other over time n.

The principle of orthogonality in Eq. (5.4) is used to find the normal equations for a linear least-squares filter [6]. Substituting Eq. (5.2) into Eq. (5.4), and rearranging terms, a system of M equations is obtained

$$\sum_{t=1}^{M} w(t) \sum_{n=M}^{N} y(n-k)y^*(n-t) = \sum_{n=M}^{N} y(n-k)y^*(n) \qquad k = 1, ..., M \tag{5.6}$$

Defining

$$r_y(t,k) = \sum_{n=M}^{N} y(n-k)y^*(n-t), \qquad 1 \leq t,k \leq M$$

where $r_y(t,k)$ represents the time averaged autocorrelation function of inputs in the transversal filter, Eq. (5.6) becomes:

$$\sum_{t=1}^{M} w(t)r_y(t,k) = r_y(0,-k) \quad , \quad k = 1, ..., M \tag{5.7}$$

Equation (5.7) represents the expanded system of the normal equations for a linear least-squares filter. In matrix form, we can write

$$\begin{bmatrix} r_y(1,1) & r_y(2,1) & \cdots & r_y(M,1) \\ r_x(1,2) & r_x(2,2) & \cdots & r_x(2,p) \\ \vdots & \vdots & & \vdots \\ r_x(1,M) & r_x(2,M) & \cdots & r_x(M,M) \end{bmatrix} \begin{bmatrix} w(1) \\ w(2) \\ \vdots \\ w(M) \end{bmatrix} = \begin{bmatrix} r_y(0,-1) \\ r_x(0,-2) \\ \vdots \\ r_x(0,-M) \end{bmatrix}$$

or, in compact form

$$\mathbf{R}_y\mathbf{w} = \mathbf{r}_y \tag{5.8}$$

Assume now that $\mathbf{R}_y$ is nonsingular, and the inverse matrix $\mathbf{R}_y^{-1}$ exists. Solving Eq.(5.8) for the weight vector $\mathbf{w}$ of the linear least-squares filter yields

$$\mathbf{w} = \mathbf{R}_y^{-1}\mathbf{r}_y \tag{5.9}$$

Equation (5.9) is fundamental to the development of all recursive formulations of the linear-squares filter.

$$\mathbf{R}_y(n) = \sum_{k=1}^{n} \lambda^{n-k} \mathbf{y}(k)\mathbf{y}^H(k) \tag{5.11}$$

and the $M \times 1$ correlation vector $\mathbf{r}_y(n)$ is

$$\mathbf{r}_y(n) = \sum_{k=1}^{n} \lambda^{n-k} \mathbf{y}(k)y^*(k) \tag{5.12}$$

Isolating the term corresponding to $k = n$ from the rest of the summation in Eq. (5.11), we obtain

$$\begin{aligned} \mathbf{R}_y(n) &= \sum_{k=1}^{n} \lambda^{n-k} \mathbf{y}(k)\mathbf{y}^H(k) \\ &= \sum_{k=1}^{n-1} \lambda^{n-k} \mathbf{y}(k)\mathbf{y}^H(k) + \mathbf{y}(n)\mathbf{y}^H(n) \\ &= \lambda \left(\sum_{k=1}^{n-1} \lambda^{n-k-1} \mathbf{y}(k)\mathbf{y}^H(k) \right) + \mathbf{y}(n)\mathbf{y}^H(n) \\ &= \lambda \mathbf{R}_y(n-1) + \mathbf{y}(n)\mathbf{y}^H(n) \end{aligned} \tag{5.13}$$

where $\mathbf{R}_y(n) = \lambda \mathbf{R}_y(n-1) + \mathbf{y}(n)\mathbf{y}^H(n)$ is a recursive equation for updating the value of the correlation matrix of the inputs. Similary, Eq. (5.12) can be represented by recursive form:

$$\begin{aligned} \mathbf{r}_y(n) &= \sum_{k=1}^{n} \lambda^{n-k} \mathbf{y}(k)y^*(k) \\ &= \lambda \mathbf{r}_y(n-1) + \mathbf{y}(n)y^*(n) \end{aligned} \tag{5.14}$$

To avoid performing the inverse $\mathbf{R}_y^{-1}(n)$ to find $\mathbf{w}(n)$ it is possible to use the matrix inversion lemma [7]. Using this notion, the inverse of the correlation matrix can be written as

$$\mathbf{R}_y^{-1}(n) = \lambda^{-1}\mathbf{R}_y^{-1}(n-1) - \frac{\lambda^{-2}\mathbf{R}_y^{-1}(n-1)\mathbf{y}(n)\mathbf{y}^H(n)\mathbf{R}_y^{-1}(n-1)}{1 + \lambda^{-1}\mathbf{y}^H(n)\mathbf{R}_y^{-1}(n-1)\mathbf{y}(n)} \tag{5.15}$$

where, for convenience of notation, we define

$$F(n) = \mathbf{R}_y^{-1}(n) \tag{5.16}$$

$$\mathbf{K}(n) = \frac{\lambda^{-1}F(n-1)\mathbf{y}(n)}{1 + \lambda^{-1}\mathbf{y}^H(n)F(n-1)\mathbf{y}(n)} \tag{5.17}$$

Then

$$F(n) = \lambda^{-1}F(n-1) - \lambda^{-1}\mathbf{K}(n)\mathbf{y}^H(n)F(n-1) \tag{5.18}$$

The $M \times M$ matrix $F(n)$ is referred to as the inverse correlation matrix. The $M \times 1$ vector $\mathbf{K}(n)$ is referred to as the gain vector. Next, we develop a recursive equation for updating the least-squares estimate $\widehat{\mathbf{w}}(n)$ for the weight vector, based on Equations (5.14) and (5.16). More precisely,

$$\mathbf{R}_y(n)\widehat{\mathbf{w}}(n) = \mathbf{r}_y(n)$$

$$\widehat{\mathbf{w}}(n) = \lambda F(n)\mathbf{r}_y(n-1) + F(n)\mathbf{y}(n)y^*(n)$$

Now, based on Eq. (5.18) we get

$$\begin{aligned}
\widehat{\mathbf{w}}(n) &= F(n-1)\mathbf{r}_y(n-1) - \mathbf{K}(n)\mathbf{y}^H(n)F(n-1)\mathbf{r}_y(n-1) + \\
&\quad F(n)\mathbf{y}(n)y^*(n) \\
&= \mathbf{R}_y^{-1}(n-1)\mathbf{r}_y(n-1) - \mathbf{K}(n)\mathbf{y}^H(n)\mathbf{R}_y^{-1}(n-1)\mathbf{r}_y(n-1) + \\
&\quad F(n)\mathbf{y}(n)y^*(n) \\
&= \widehat{\mathbf{w}}(n-1) + \mathbf{K}(n)\left(y^*(n) - \mathbf{y}^H(n)\widehat{\mathbf{w}}(n-1)\right) \text{ due to } \mathbf{K}(n) = F(n)\mathbf{y}(n) \\
&= \widehat{\mathbf{w}}(n-1) + \mathbf{K}(n)\epsilon^*(n)
\end{aligned} \tag{5.19}$$

where

$$\begin{aligned}
\epsilon(n) &= y(n) - \mathbf{y}^T(n)\widehat{\mathbf{w}}^*(n-1) \\
&= y(n) - \widehat{\mathbf{w}}^H(n-1)\mathbf{y}(n)
\end{aligned} \tag{5.20}$$

is the a priori estimation error. Equations (5.17),(5.20),(5.19) and (5.18) in that order, form the Recursive Least-Squares (RLS) algorithm. Figure (5.4) give a block diagram of RLS algorithm. Equation (5.19) describes the adaptive operation of the algorithm, where the weight vector is updated by incrementing its old value by an amount equal to the complex conjugate of the priori estimation error $\epsilon(n)$ times the time- varying gain vector $\mathbf{K}(n)$ which is updated by Eq.(5.17) and (5.18).

To initialize of RLS algorithm, the value $F(0)$ has to ensure the nonsingularity of the correlation matrix $\mathbf{R}_y(n)$. An alternative [7], [1] to deal with this problem is to evaluate the inverse

$$\left(\sum_{k=-n}^{0} \lambda^{-k}\mathbf{y}(k)\mathbf{y}^H(k)\right)^{-1}$$

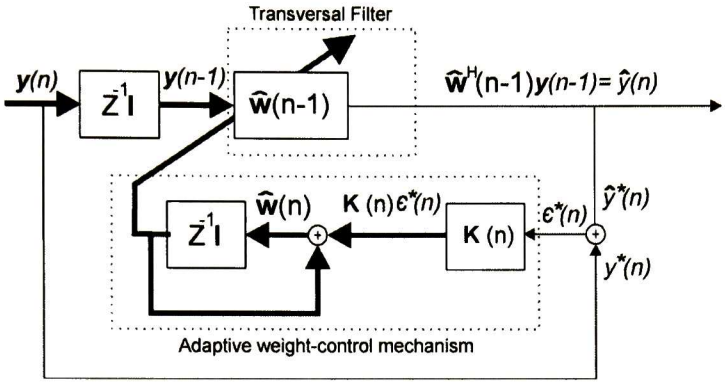


Figure 5.4: Representations of the RLS algorithm

from an initial block of data for $-n \leq k \leq 0$. Another approach is to use $F(0) = \delta^{-1}\mathbf{I}$, where δ is a small positive constant where it should be small compared to $0.01\sigma_y^2$. The vector $\widehat{\mathbf{w}}(0)$ can be initialized by [31], [29] $\widehat{\mathbf{w}}(0) = \mathbf{0}$.

In a stationary environment, the rate of convergence of the RLS algorithm is typically an order of magnitude faster than that of the LMS algorithm, the rate of convergence of the RLS algorithm is invariant to the eigenvalue spread of the ensemble-averaged correlation matrix of the input vector and the excess mean-squared error of the RLS algorithm converges to zero as the number of iterations n go to infinity.

5.4 Recursive Least-Squares algorithm with adaptive forgetting factor

Because the power system is a time varying system the LMS and RLS algorithms should calculate optimum values for the for the step-size parameter μ in the LMS algorithm and calculate optimum values for the exponential weighting factor λ in the RLS algorithm. The conventional form of the adaptive filters to track the time varying system requires knowledge of the correlation matrix of the process noise and the correlation matrix of the measurement noise. However, sometimes this may be a drawback since the data are not known. Certain modifications to the RLS algorithm can be made to improve its performance in non-stationary ambient. This consist in turning λ in the RLS.

RLS algorithm can be adapted for tuning the forgetting factor λ [7].The objective is to find the particular value λ that optimizes the criterion function

$$N(n) = \frac{1}{2}\mathbf{E}[|\epsilon(n)|^2] \tag{5.21}$$

where $\epsilon(n)$ is defined by Eq. (5.20). Differentiating the criterion function $N(n)$ with respect to λ yields

$$\begin{aligned} \nabla(n) &= \frac{\partial N(n)}{\partial \lambda} = \frac{1}{2}\mathbf{E}\left[\frac{\partial \epsilon(n)}{\partial \lambda}\epsilon^*(n) + \frac{\partial \epsilon^*(n)}{\partial \lambda}\epsilon(n)\right] \\ &= -\frac{1}{2}\mathbf{E}\left[\boldsymbol{\phi}^H(n-1)\mathbf{y}(n)\epsilon^*(n) + \mathbf{y}^H(n)\boldsymbol{\phi}^H(n-1)\epsilon(n)\right] \end{aligned} \tag{5.22}$$

$\nabla(n)$ is a scalar gradient ,where

$$\boldsymbol{\phi}(n) = \frac{\partial \mathbf{w}(n)}{\partial \lambda} \tag{5.23}$$

Having that

$$\mathbf{w}(n) = \mathbf{w}(n-1) + F(n)\mathbf{y}(n)\epsilon^*(n) \tag{5.24}$$

and defining the derivative of the inverse correlation matrix $F(n)$ with respect to λ by

$$\mathbf{D}(n) = \frac{\partial \mathbf{D}(n)}{\partial \lambda} \tag{5.25}$$

and using Eqs. (5.20), (5.24) and (5.25) in Eq. (5.23) it is possible to show that

$$\boldsymbol{\phi}(n) = \left(\mathbf{I} - \mathbf{K}(n)\mathbf{y}^H(n)\right)\boldsymbol{\phi}(n-1) + \mathbf{D}(n)\mathbf{y}(n)\epsilon^*(n) \tag{5.26}$$

Equations (5.15) and (5.16) allows us to write

$$F(n) = \lambda^{-1}F(n-1) - \frac{\lambda^{-2}F(n-1)\mathbf{y}(n)\mathbf{y}^H(n)F(n-1)}{1 + \lambda^{-1}\mathbf{y}^H(n)F(n-1)\mathbf{y}(n)} \tag{5.27}$$

where is possible obtain an expression for the recursive compute of $\mathbf{D}(n)$ differentiating Eq.(5.27) with respect to λ :

$$\begin{aligned} \mathbf{D}(n) &= \lambda^{-1}\left(\mathbf{I} - \mathbf{K}(n)\mathbf{y}^H(n)\right)\mathbf{D}(n-1)\left(\mathbf{I} - \mathbf{y}(n)\mathbf{K}^H(n)\right) \\ &\quad + \lambda^{-1}\mathbf{K}(n)\mathbf{K}^H(n) - \lambda^{-1}F(n) \end{aligned} \tag{5.28}$$

Then, using the instantaneous estimate $-\operatorname{Re}\left(\boldsymbol{\phi}^H(n-1)\mathbf{y}(n)\epsilon^*(n)\right)$ for the scalar gradient $\nabla(n)$ is possible adaptively compute the forgetting factor using the recursion [7]:

$$\begin{aligned}\lambda(n) &= \lambda(n-1) - \bigtriangledown(n) \\ &= \left[\lambda(n-1) + \mathrm{Re}\left(\boldsymbol{\phi}^H(n-1)\mathbf{y}(n)\epsilon^*(n)\ \right)\right]_{\lambda_-}^{\lambda_+}\end{aligned}$$

The bracket with λ_+ and λ_- indicates truncation. The RLS algorithm with adaptive forgetting factor is summarized as follows:

$$\begin{aligned}\mathbf{K}(n) &= \frac{\lambda^{-1}(n-1)F(n-1)\mathbf{y}(n)}{1+\lambda^{-1}(n-1)\mathbf{y}^H(n)F(n-1)\mathbf{y}(n)} \\ \epsilon(n) &= y(n) - \widehat{\mathbf{w}}^H(n-1)\mathbf{y}(n) \\ \widehat{\mathbf{w}}(n) &= \widehat{\mathbf{w}}(n-1) + \mathbf{K}(n)\epsilon^*(n) \\ F(n) &= \lambda^{-1}(n-1)F(n-1) - \lambda^{-1}(n-1)\mathbf{K}(n)\mathbf{y}^H(n)F(n-1) \\ \lambda(n) &= \left[\lambda(n-1) + \mathrm{Re}\left(\boldsymbol{\phi}^H(n-1)\mathbf{y}(n)\epsilon^*(n)\ \right)\right]_{\lambda_-}^{\lambda_+} \\ \mathbf{D}(n) &= \lambda^{-1}(n)\left(\mathbf{I} - \mathbf{K}(n)\mathbf{y}^H(n)\right)\mathbf{D}(n-1)\left(\mathbf{I} - \mathbf{y}(n)\mathbf{K}^H(n)\right) \\ &\quad +\lambda^{-1}(n)\mathbf{K}(n)\mathbf{K}^H(n) - \lambda^{-1}(n)F(n) \\ \boldsymbol{\phi}(n) &= \left(\mathbf{I} - \mathbf{K}(n)\mathbf{y}^H(n)\right)\boldsymbol{\phi}(n-1) + \mathbf{D}(n)\mathbf{y}(n)\epsilon^*(n)\end{aligned}$$

The RLS algorithm with adaptive forgetting factor is useful for time-varying systems due to the characteristic of tuning the forgetting factor.

The practical validity of the this idea has been fully supported by signal processing applications on adaptive equalization and Phase-Locked Loop, and proof of convergence based on a fairly strong result rooted in stochastic approximation theory [7].

Chapter 6
Applications

This chapter describes the application of canonical and least-squares algorithms to estimate low-frequency electromechanical modes through ambient analysis.
First, simulation data from a 16-machine test power system is used to test the performance of the method in the presence of random load fluctuations. Then, actual measured ambient data from actual system events is used to evaluate the ability of the method to characterize power system oscillatory dynamics. The non-stationary character of the recorded oscillations provides the thrust to examine the application of the method for on-line estimation of instantaneous attributes. Comparisons are provided with overdetermined ARMA methods and LMS algorithms, and power spectral density.
The analysis demonstrates that the nonstationary characteristics of ambient data can be reasonably assessed by simple subspace identification techniques and the information thus obtained can be helpful in developing improved real-time monitoring and control systems.

6.1 Dynamic model verification using simulated data

In this section the continuous modal parameter estimation techniques are validated on a practical 16-machine test system. To test the methods, simulated data containing small oscillations arising from changes is load behavior are used.

In this analysis, the system model is perturbed by random load variations at a number of points in the system, and the system response of the model is obtained at locations where the modes are observable in an attempt to replicate the actual power system dynamics.

6.1.1 Power system model

The test power system is a 16-machine system with 86 transmission lines and 68 buses model of the New England test system (NETS) and New York power

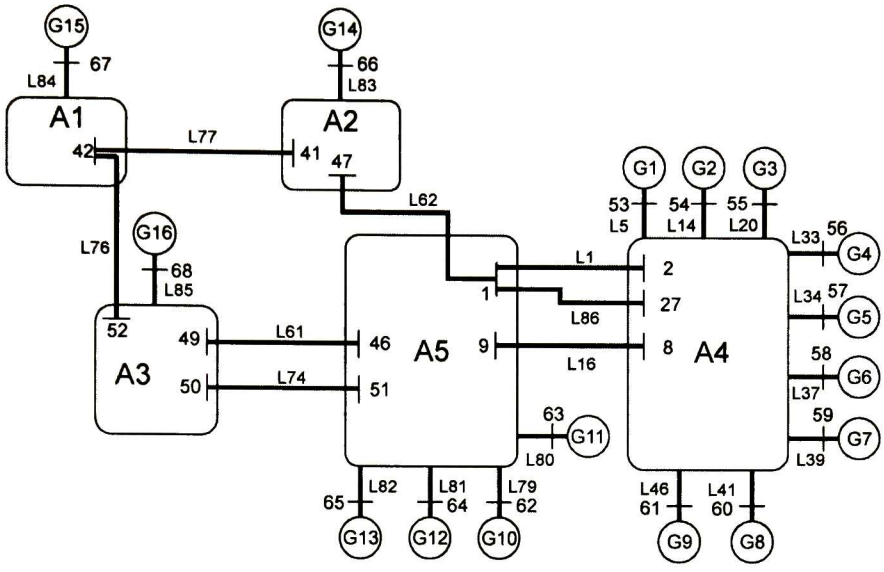


Figure 6.1: Diagram of 16 machine system

system (NYPS). Figure (6.1) gives a single-line diagram of this system showing major transmission resources. Each generator is fully represented using a detailed two-axis transient model and excitation and speed controllers. The detailed description of the test system including network data and dynamic data for the machines, buses, branches can be found in [22].

The system dynamic behavior of this system is characterized by several low-frequency oscillation modes in the range of 0.2-0.7 Hz. Table (6.1) summarizes the main characteristics of the slow-frequency modes showing their frequencies and damping ratios.

The simulations were conducted using the Power System Transient (PST) stability program.

Mode	Eigenvalue	Frequency (Hz)	Damping Ratio
1	-0.366 ± 2.677i	0.426	0.135
2	-0.290 ± 3.597i	0.572	0.080
3	-0.504 ± 4.206i	0.669	0.119
4	-0.3816 ± 5.178i	0.824	0.073

(6.1)

6.1.2 Simulated data

For the analysis of system response, the system model is perturbed by random load variations. Specifically, independent Gaussian random load modulations are added to the real and reactive loads at buses. The following considerations are introduced in this analysis:

- The smaller loads change frequently and this leads to an actual system load which varies almost randomly about a constant value.
- The loads change varies about of $\pm 1\%$ from its nominal value.
- Each load is taken to be independent of other loads.
- The load was assumed to change randomly every 0.01 S.

The time-domain response was calculated using PSTV2 for a duration of 10 minutes. A sampling rate of 0.01 seconds was used in all simulations.

Records were selected based on the following criteria:

- Selected outputs from these simulations are the active power flows across selected lines.
- Selected transmission lines interconnected major coherent groups
- Critical inter-area modes are observable at these interties. In this analysis, participation factors are used to select signals of interest.

  Table 6.1lists major inter-area modes of interest obtained using the software PSTV2.

For illustration, Fig. (6.2) shows the real power flow across transmission line 82. Note the stochastic nature of system behavior.

6.1.3 Selection of measurement locations

Transmission lines were selected based on participation factors but other techniques could also be used. Mode 1 was selected for analysis.

Figure (6.3) shows the participation factors for the 0.426 Hz mode showing the participation of dominant generators. As seen in this plot, generators 13-15 show the largest participation.

Based on these results, transmission lines having a dominant participation in this mode were selected. Table (6.2) identifies major transmission lines having an important contribution on the 0.426 Hz mode.

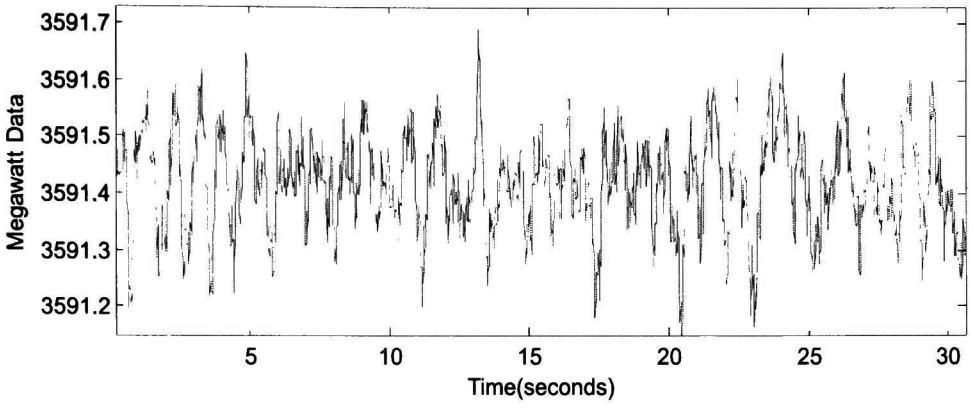


Figure 6.2: Real-power flow on line 82

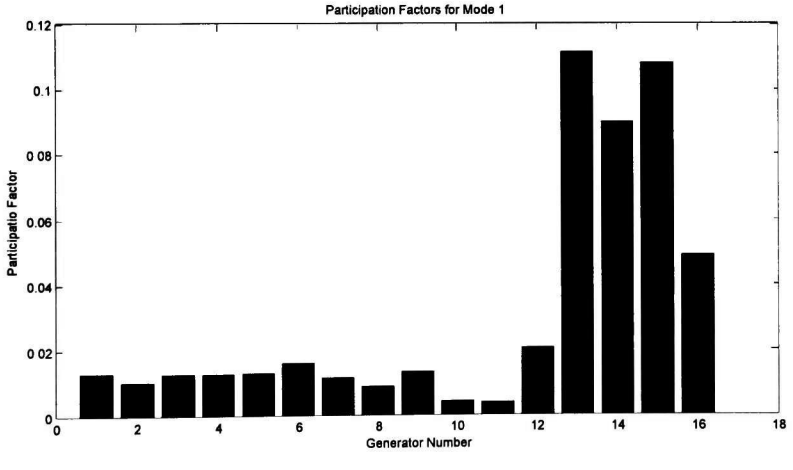


Figure 6.3: The mode 1 can be observed in the lines 82, 83, 84, 85 corresponding to generators 13, 14, 15 y 16 respectively.

Number of Line	Number of Mode		
82	1	3	
83	1	2	4
84	1	4	
85	1	2	4

(6.2)

6.1.4 Modal Identification using canonical variate analysis

The mode frequency and damping of the data generated was estimated using 1 and 5 minutes windows. Figures (6.4) and (6.5) show modal estimates as a function of the observation window for 1 minute block size. The solid bold lines represents the actual values obtained from PSTV2. Table (6.3) compares the estimated frequency and modal damping for 5 minutes block size with the results from PSTV2. As observed from these figures, modal estimates are in excellent agreement with actual system values for the entire observation period.

Applying the CVA, the system matrix A is obtained, the eigenvalues from matrix A are the characteristic values of the system, then the corresponding frequency and ratio are calculated. For the system, the stochastic subspace method with weights for CVA was applied to estimate both the frequency and damping ratio of the 0.426 Hz mode and the simulated signal used for this estimation was the Line 82. The method was applied a single trial of data. This is consistent with theoretical expectations.

		PSTV2	CVA
Mode 1	frequency	0.426	0.432
	Damping	0.135	0.1613

(6.3)

6.1.5 Modal Identification using Recursive Algorithms

To use the recursive algorithms, the data was pre-processed before using the adaptive algorithms. First, the mean value is removed from the data; then a low pass filter with a cutoff frequency of 2 Hz is used to remove the high frequency measurement noise from the data. Finally, the data is decimated from 100 samples/second to 5 samples/second. The main purpose for applying the decimated is achieve the Nyquist criterion in looking at the modes with frequencies below 2.5 Hz. Finally, the decimated data is high pass filtered to remove very low frequencies below 0.1 Hz and to remove any offset. Figure (6.6) shows a portion of preprocessed megawatt data of line 82.

The exponentially weighting RLS algorithm and RLS algorithm with adap-

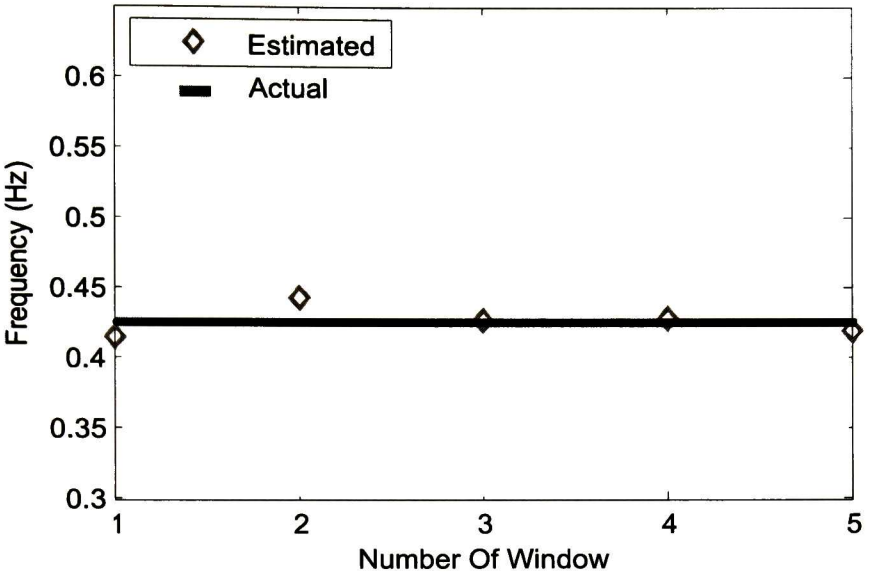


Figure 6.4: Mode frequency estimate using 1-minute block size

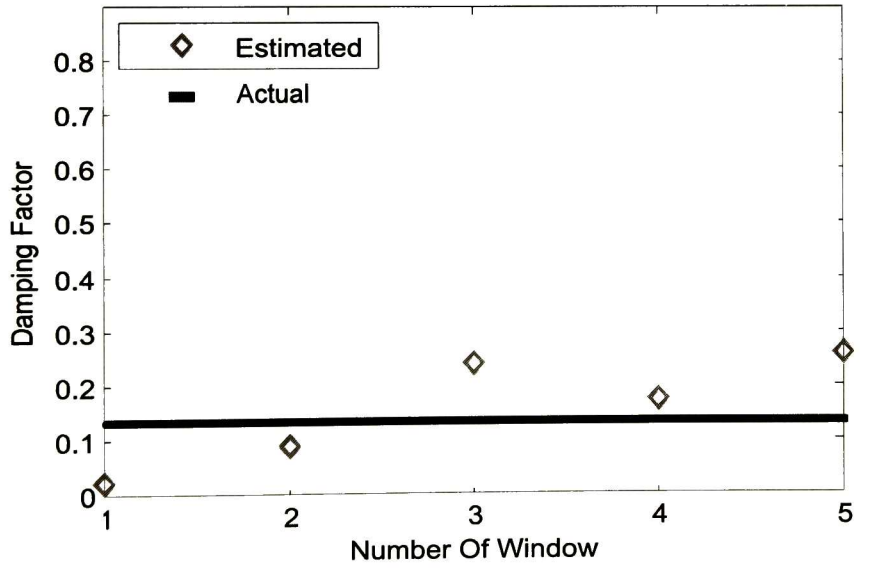


Figure 6.5: Mode Damping using 1-minute block size

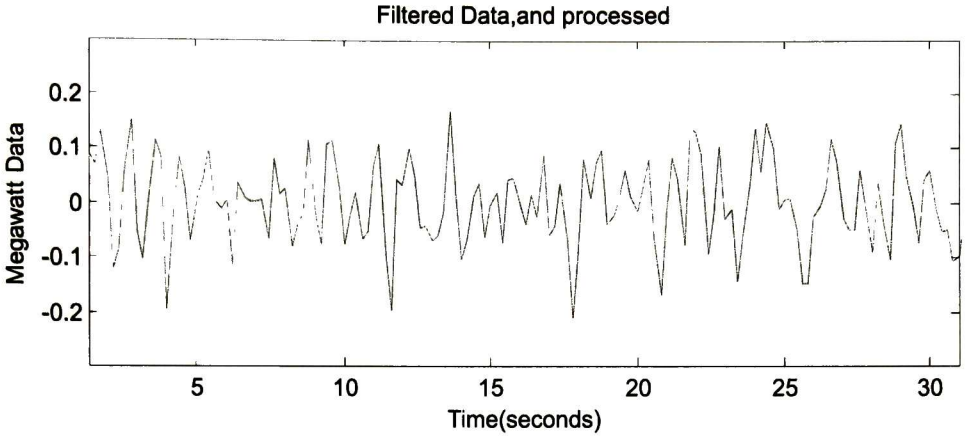


Figure 6.6: 30 seconds of preprocessed megawatt data of Line 82 of 16-machine model.

tive forgetting factor were applied to the data to track the 0.426 Hz. Figures (6.7), (6.8) show the frequency and damping ratio estimates, respectively obtained using the exponentially weighting RLS algorithm and RLS algorithm with adaptive forgetting factor. The comparison was made with LMS algorithm.

As can be seen from the plots, the accuracy and convergence time compare with LMS algorithm. It is worth noting that the RLS algorithms provide better estimates of damping ratios. RLS algorithms lead to similar estimates because the adaptation of the forgetting factor λ was limited to a narrow range (0.989 to 0.999) very near to the value λ (0.999) of the exponentially weighting RLS algorithm. We emphasize that the initial values of weight vectors was adjusted to zero (cold start).

6.2 Modal Estimation using Ambient Data

Several time series collected using phasor measurement systems are used as example to investigate the capacity of the proposed techniques to extract modal information. The data set includes recordings of various events, namely:

(a) A real signal from a transient event in the North American system

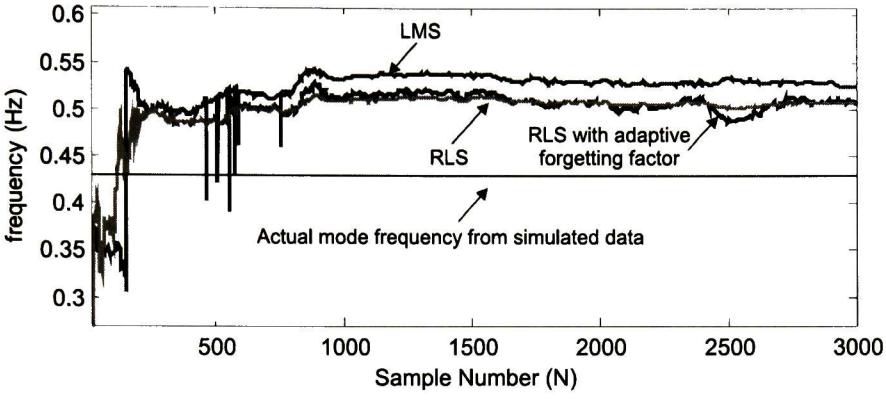


Figure 6.7: RLS algorithms frequency estimate using simulated data of 16-machine model (10 minutes of data) and compared with LMS frequency estimate,(all are initialized by zero weigth vectors).

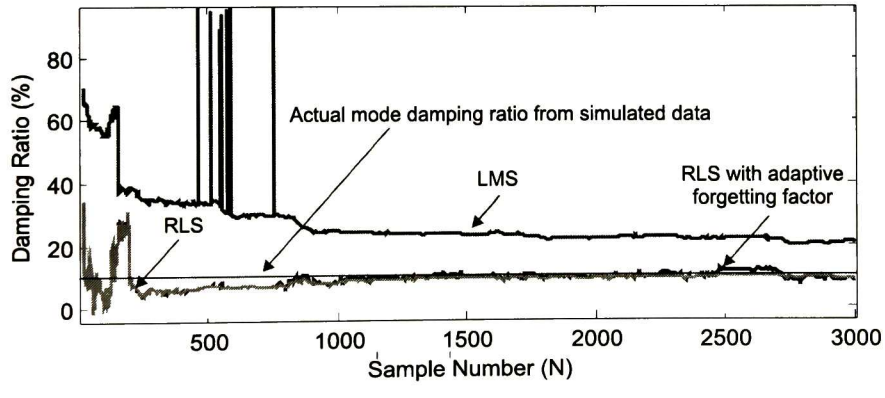


Figure 6.8: RLS algorithms damping ratio (%) estimate using simulated data of 16-machine model (10 minutes of data) and compared with LMS damping ratio estimate,(all are initialized by zero weigth vectors).

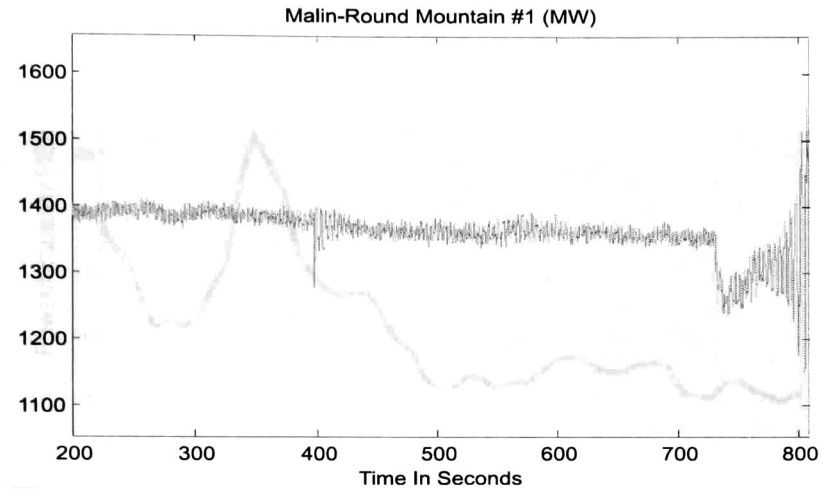


Figure 6.9: Oscillation buildup for the WSCC breakup of August 10, 1996

(b) Ambient data from the Nordic power system

This data is representative of many complex oscillations involving transient variations.

For completeness, the CVA method is compared with an overdetermined ARMA and the recursive least-squares methods are compared with LMS methods.

6.2.1 Malin-Round Mountain signal

The first signal analyzed is tie-line power from the August 10, 1996 event. The event shows a condition of high dynamic stress; the mechanism of failure was a transient oscillations, under conditions of high power transfer on long paths that had been progressively weakened through a series of seemingly routine transmission line outages. Figure (6.9) gives the time evolution of selected tie-line power at a critical interface.

Figure (6.10) shows the estimate of the power spectral density for the real-power flow Malin-Round Mountain. The analysis shows a dominant peak at about 0.253 Hz.

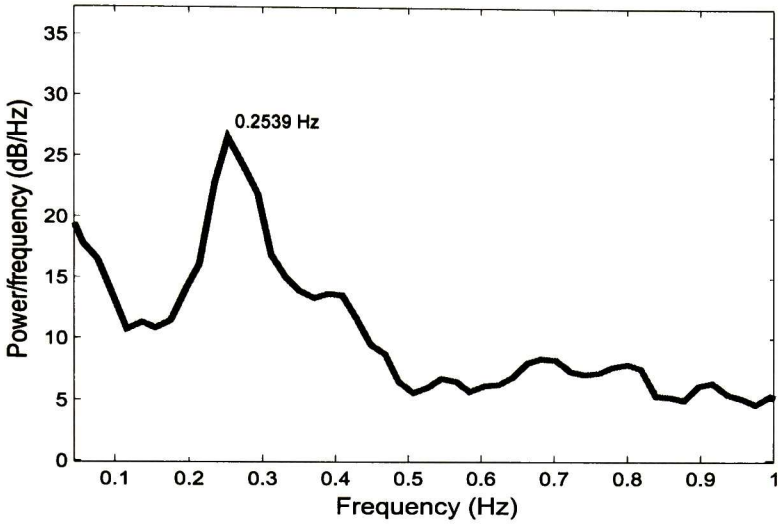


Figure 6.10: Power spectral density of test signal 1.

Modal Identification using canonical variate analysis

Canonical variate analysis technique and ARMA models were applied to the test signal. Figure (6.11) compares the instantaneous frequency estimates whilst Fig. (6.12) shows the corresponding damping ratio estimates. In both cases a 1-minute sliding window, updated every minute was used.

Comparison of the instantaneous attributes shows that the CVA and the ARMA models are able to extract the underlying mode of interest for the entire interval of interest with each of the methods yielding better results for some intervals of the signal.

Figures (6.13) and (6.14) show simulation results for a 2-minute sliding window updated every 2 minutes.

Modal Identification using Recursive Algorithms

The Malin-Roun Mountain #1 (MW) ambient power system data was preprocessed before applying the recursive algorithms. The results with exponentially weighting RLS algorithm and RLS with adaptive forgetting factor algorithm were compared with LMS, LMS normalized and overdetermined AR block processing method results. The analysis was made with initial values adjusted to zero, namely, cold start.

The analysis was applied to the first 18 minutes of the signal; this includes the portion of larger oscillation for also analysis of the methods to large pertur-

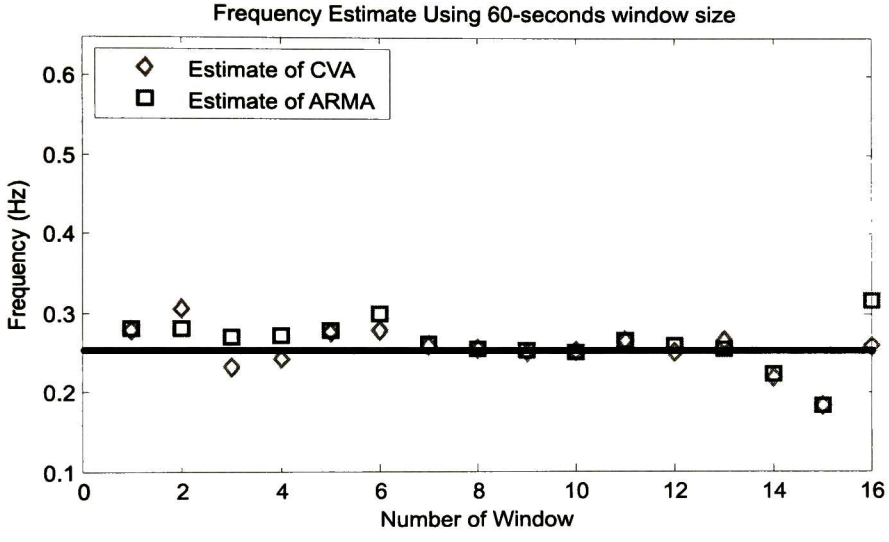


Figure 6.11: CVA and ARMA frequency estimates using Malin-Round Mountain megawatt data (1-minute window size)

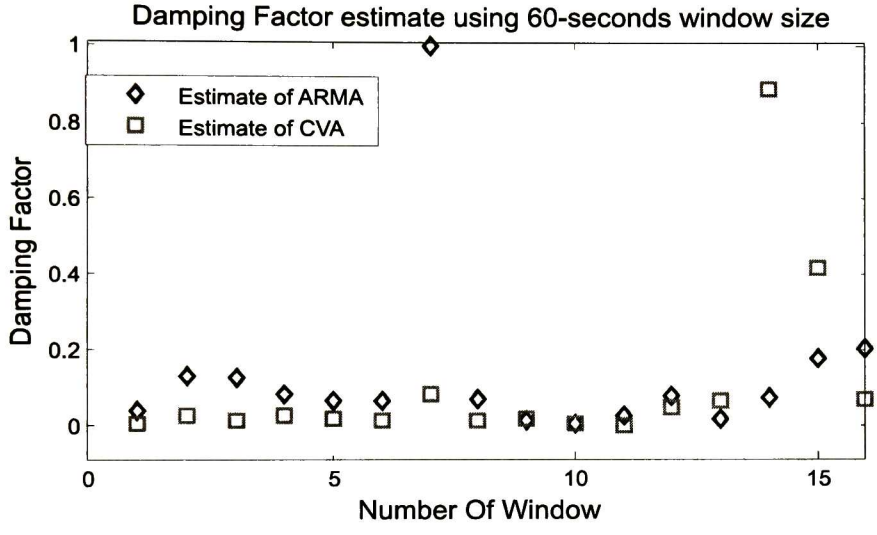


Figure 6.12: CVA and ARMA damping estimates using Malin-Round Mountain (1-minute window size)

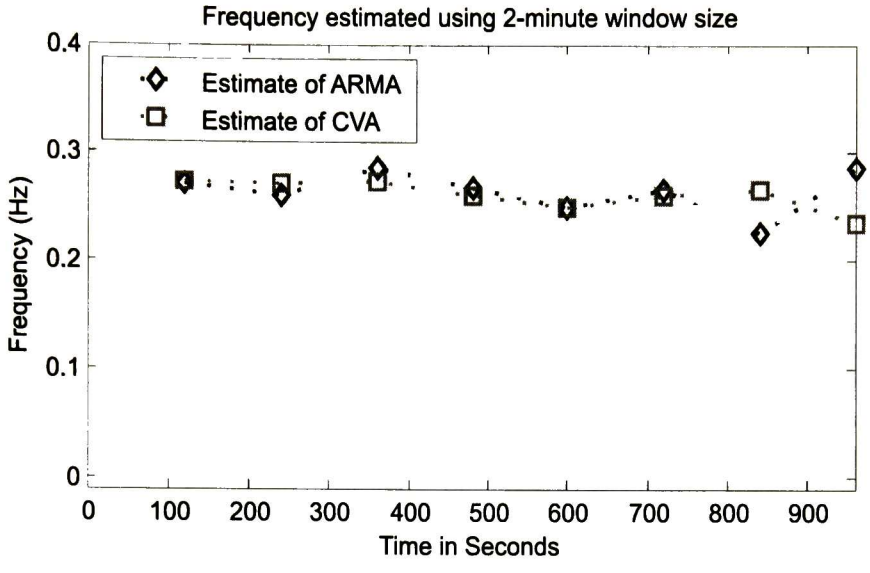


Figure 6.13: CVA and ARMA frequency estimate using Malin-Round Mountain (2-minute window size)

bations. The exponentially weighting RLS and RLS with adaptive forgetting algorithms were applied to the ambient data to track the 0. 2539 HZ mode during its equilibrium condition and when it is have large perturbations. The reference used to mode frequency was the peak in the PSD and the reference used to mode damping ratio was an overdetermined AR of order 24 using full block size (18 minutes). In all simulations of adaptive algorithms was used an order 28.

Figures (6.15) and (6.16) shows the mode frequency estimate using the RLS algorithms, these estimates are compared with LMS and normalized LMS algorithm estimates. The reference PSD for mode frequency shows that the normalized LMS algorithms and the RLS algorithms yield very good estimates. At about N=2000 and N=4000 it is possible observe the perturbations in the mode frequency estimates arising from large perturbations in the ambient data, following a large perturbation. This is more clearly observed in the damping ratio estimates in figures (6.15) and (6.16).

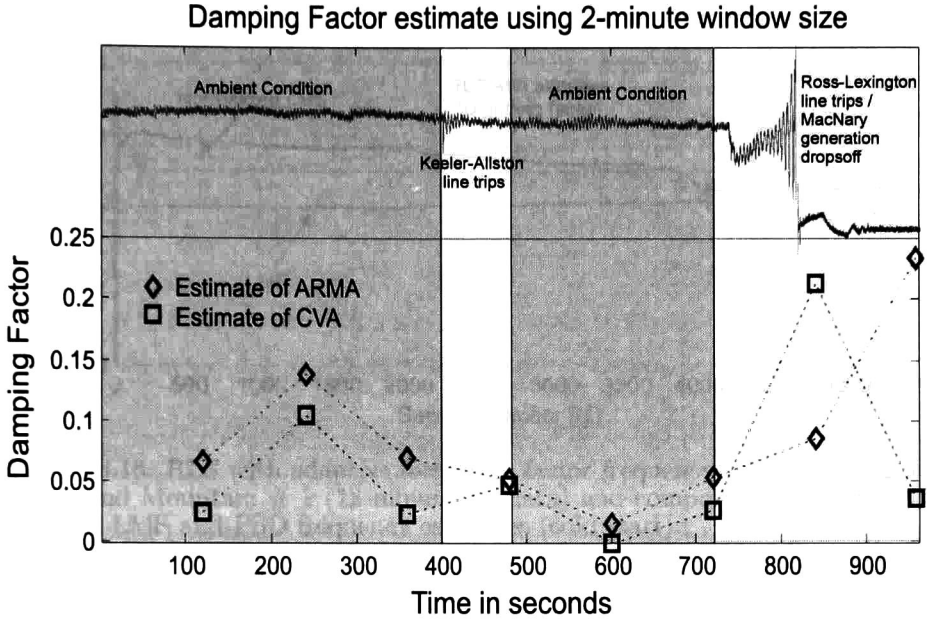


Figure 6.14: CVA and ARMA damping ratio estimates using Malin-Round Mountain (2-minute window size)

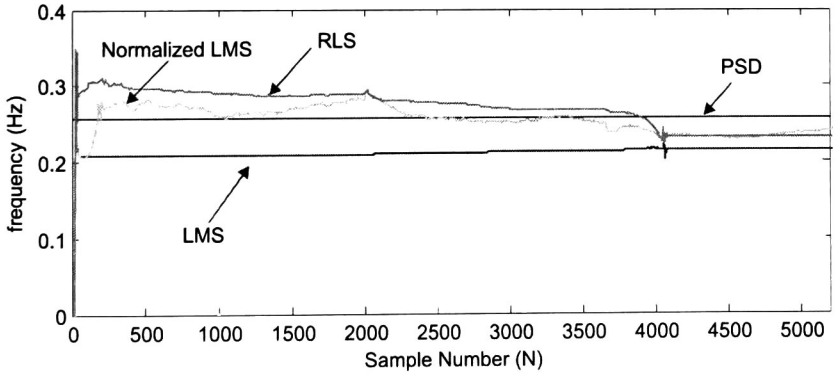


Figure 6.15: RLS frequency estimate using Malin-Round Mountain # 1 (18 minutes of data) and compared with LMS, Normalized LMS and PSD frequency estimates (cold start).

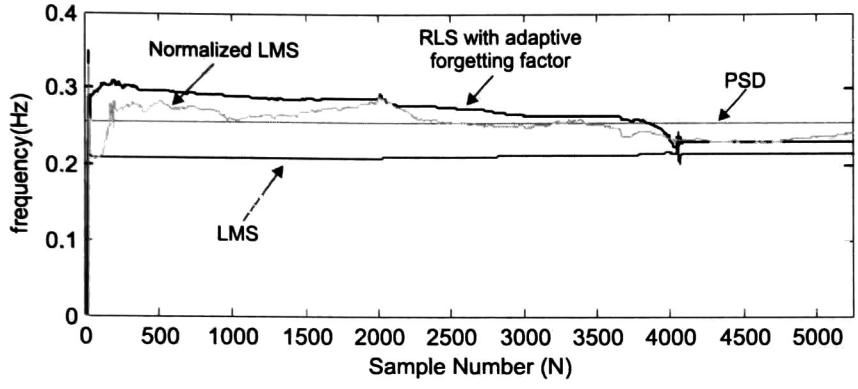


Figure 6.16: RLS with adaptive forgetting factor frequency estimate using Malin-Round Mountain # 1 (18 minutes of data) and compared with LMS, Normalized LMS and PSD frequency estimates (cold start).

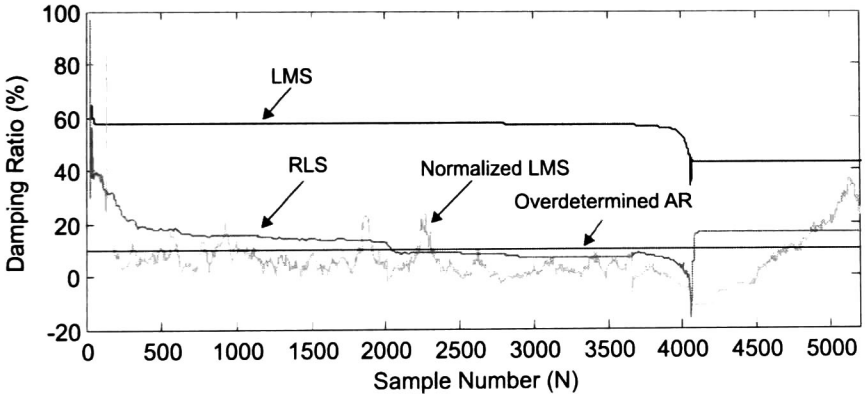


Figure 6.17: RLS damping ratio (%) estimate using Malin-Round Mountain # 1 (18 minutes of data) and compared with LMS, Normalized LMS and overdetermined AR damping ratio (%) estimates (cold start).

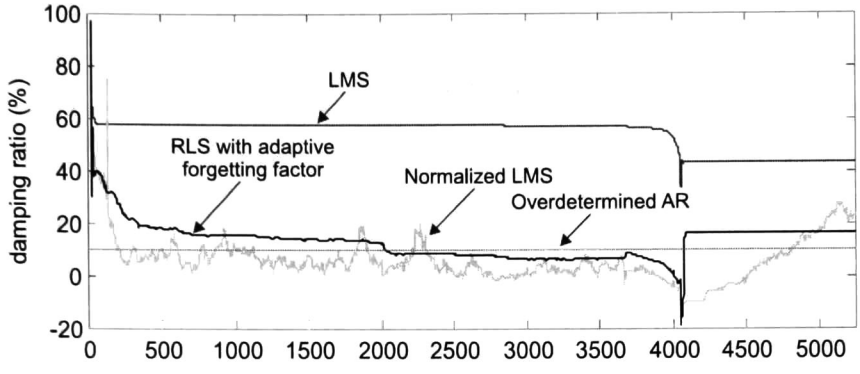


Figure 6.18: RLS with adaptive forgetting factor damping (%) estimate using Malin-Round Mountain # 1 (18 minutes of data) and compared with LMS, Normalized LMS and overdetermined AR damping (%) estimates (cold start).

6.2.2 Test Signal 2

The second signal represents ambient data from Nordic Power System and shows a strongly nonstationary ample content caused by switching actions in the system. Figure (6.19) shows the time evolution of this signal.

Also of interest, the spectra of this signal in Fig (6.20) shows a dominant mode at about 0.351 Hz. Other modes are practically negligible

Modal Identification using CVA

Following a procedure similar to that in the previous section, the CVA method was applied using various window sizes. Figures (6.21) and (6.22) shows modal estimates. In both cases, a 4-minute sliding window updated every 4 minutes was used.

Simulation results are again found to be in good agreement with theoretical expectations showing the correctness of results.

Modal Identification using Recursive Algorithms

The results with exponentially weighting RLS algorithm and RLS with adaptive forgetting factor algorithm were compared with LMS, LMS normalized and PSD results. The analysis was made with initial values adjusted to zero.

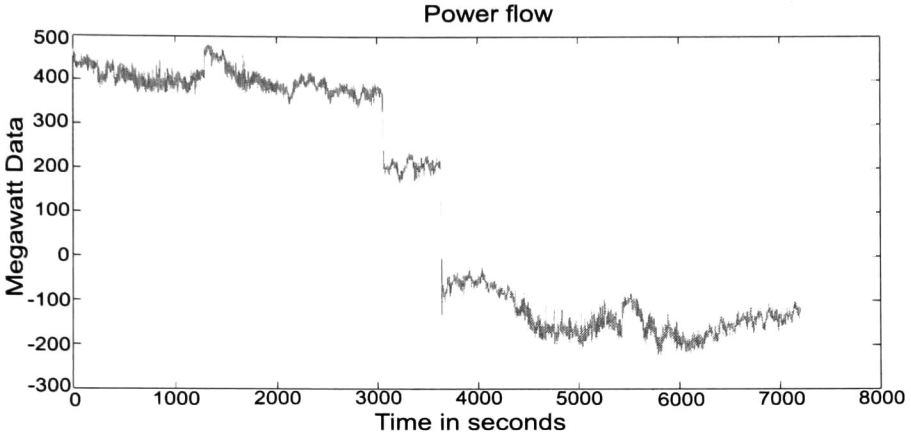


Figure 6.19: Time evolution of test signal 2

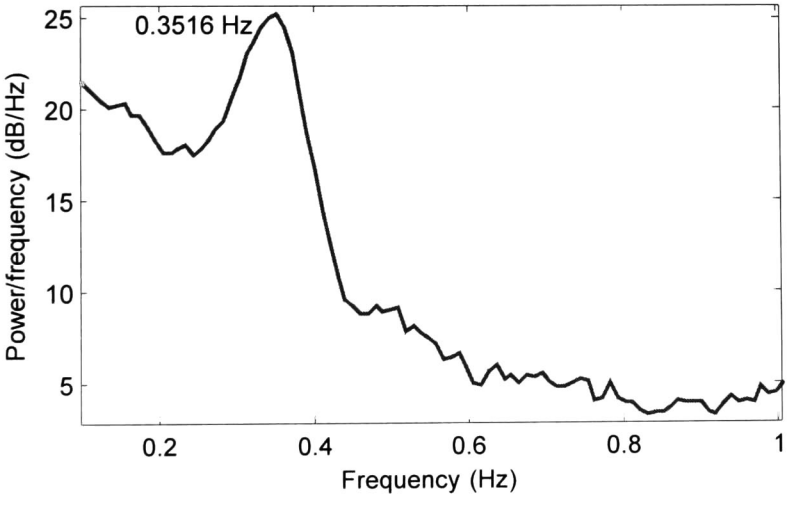


Figure 6.20: Power spectrum density of real-power flow on test signal 2.

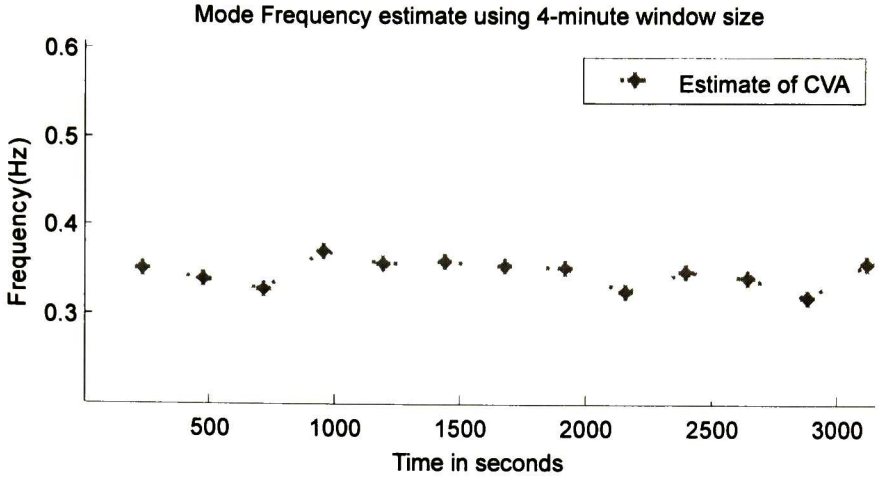


Figure 6.21: CVA frequency estimate using test signal 2 (4-minute window size)

The analysis was applied to the first 50 minutes of the signal. The exponentially weighting RLS and RLS with adaptive forgetting algorithms were applied to the ambient data to track the 0. 351 HZ mode. In all simulations of adaptive algorithms was used an order 28.

Figures (6.23) show the mode frequency estimate using the RLS algorithms, this estimate is compared with LMS and normalized LMS algorithm estimates. The reference PSD for mode frequency shows that the RLS algorithm yield very good estimate. The Figure (6.24) show that RLS algorithm yield very good estimate of damping ratio

In the literature, this mode is described as a mode of 0.33 Hz with a 4.34% damping factor being observable as power oscillations between Finland and South Norway (Hammond, 1968).

Several interesting observations from these results can now be formulated.

- Both, CVA and ARMA models are found to provide a good approximation to the temporal evolution of modal characteristics. Of particular importance, estimation of modal damping is the most challenging problem as pointed out in the above discussion

- The analysis of transient conditions on the other hand illustrates other aspects of interest in the application of the methods. As may be seen from these plots, canonical analysis appears to detect the onset of system

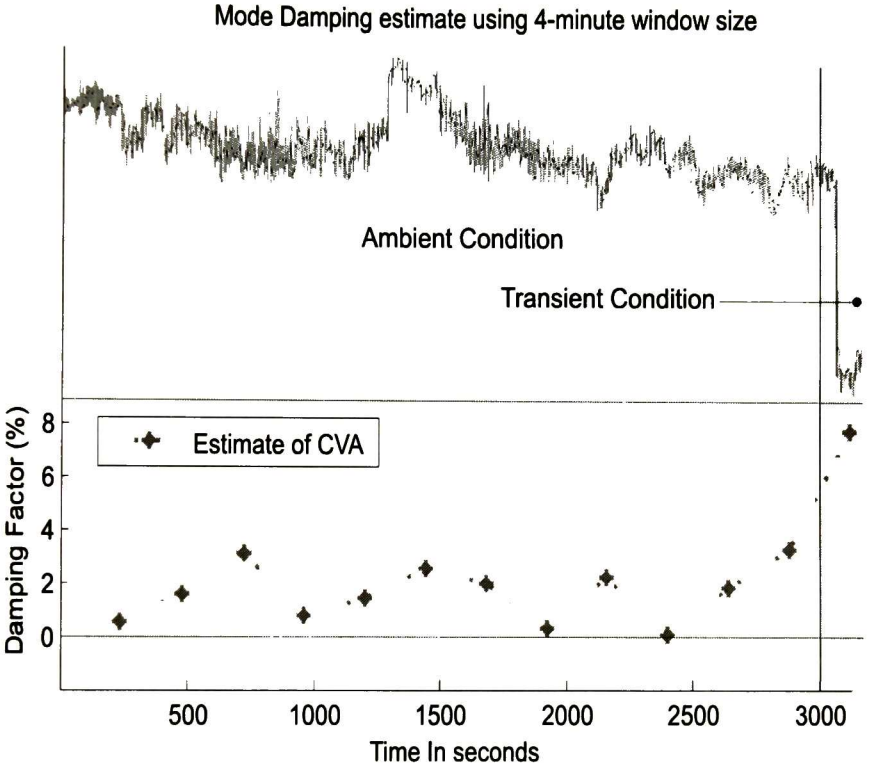


Figure 6.22: CVA damping factor estimate using test signal 2 (4-minute window size)

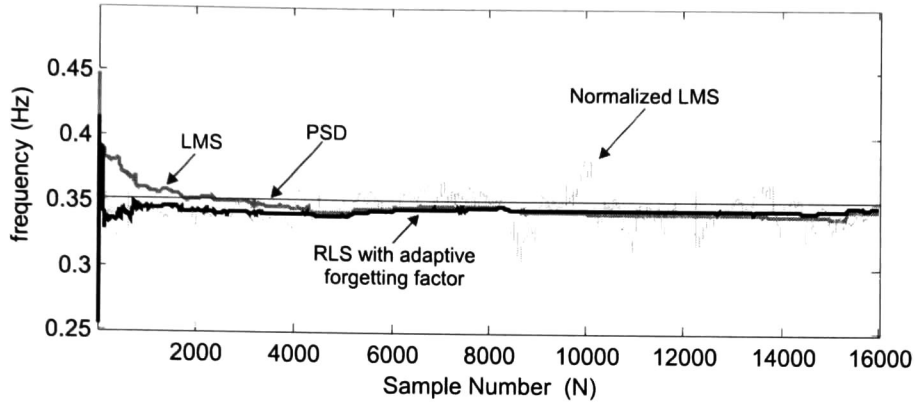


Figure 6.23: RLS with adaptive forgetting factor frequency estimate using test signal 2 (50 minutes of data) and compared with LMS and Normalized LMS frequency estimates (cold start).

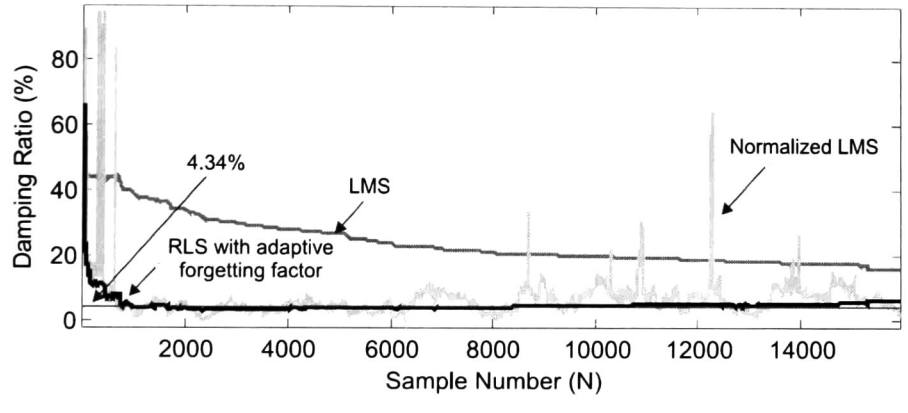


Figure 6.24: RLS with adaptive forgetting factor damping (%) estimate using test signal 2 (50 minutes of data) and compared with LMS and Normalized LMS damping (%) estimates (cold start).

oscillations. This is highly desirable feature since the system undergoes an abrupt transition

Chapter 7
General Conclusions

Transient processes in power systems are inherently non-stationary. In this thesis, continuous, block-meter processing techniques based on canonical and least-squares recursive techniques have been employed to characterize the dynamic behavior of ambient. Compared to other approaches, these techniques are thought to be more suitable to analyze ambient measurements. Simulation studies conducted on time-synchronized data demonstrate the potential of these methods to be used in real-time applications.

The two approaches lead to different methods for modal extraction and provide comparable information. Other aspects such as computation effort, sensitivity to outliers and missing data, and performance in the presence of noise, among other aspects need to be further investigated.

7.1 Future work

Several lines of investigation open from these studies. First, further examination is needed of numerical properties, especially in the presence of noise. Further study is also required to understand and characterize nonlinearities in system performance. The results are an initial step in this direction.

Future work including improved analytical models and validation under more general operating and testing conditions are needed. In addition, parametric investigations have to be performed to investigate the effects of sampling rates and other dynamic parameters of interest.

It is hoped that this information will be useful in the further development of nonstationary stochastic models when mathematical sophistication permits, and that eventually they can be used in solving on-line prediction problems.

References

[1] P S. R. DINIZ. "Adaptive Filtering: Algorithms and Practical Implementation" New York: Springer (2008).

[2] H. GHASEMI. "On-Line Monitoring and Oscillatory Stability Margin Prediction in Power Systems Based on System Identification" PhD thesis, University of Waterloo (2006).

[3] A. GHOSH AND G. LEDWICH. Opportunities in wide are control and measurements (WACAM). *Australasian Universities Power Engineering Conference AUPEC 2006* pp. 1–6 (Decemeber 2006).

[4] J. HAUER, D. TRUDNOWSKI, AND J. DESTEESE. A perspective on WAMS analysis tools for tracking of oscillatory dynamics. *IEEE Power Engineering Society General Meeting* pp. 1–10 (June 2007).

[5] J. F HAUER. Application of prony analysis to the determination of modal content and equivalent models for measured power system response. *Transactions on Power System* **6**(3), 1062–1068 (August 1991).

[6] M. H. HAYES. "Statistical Digital Signal Processing and Modeling". NewYork: Wiley (1996).

[7] S. HAYKIN. "Adaptive Filter Theory" NJ:Prentice-Hall (1991).

[8] F B. HILDEBRAND. "Introduction to Numerical Analysis" New York: Dover Publications,Inc (1987).

[9] Y. HUA AND T. K. SARKAR. Matrix pencil and system poles. *Elsevier Science Publisher B. V. signal processing* **21**, 195–198 (1990).

[10] T. KAILATH. "Linear Systems" Prentice-Hall (1980).

[11] T. KATAYAMA. "Subspace Method for System Identification: A Realization Approach" London: Springer-Verlag (2005).

[12] R. KUMARESAN. On the zeros of the linear prediction error filter for deterministic signals. *IEEE Transactions on acoustics, speech, and signal processing* **31**(1), 217–220 (February 1983).

[13] R. KUMARESAN AND D. W TUFTS. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Transactions on acoustics, speech, and signal processing* **30**(6) (December 1982).

[14] R. KUMARESAN AND D. W TUFTS. A prony method for noisy data: Chossing the signal components and selecting the order in exponential signal models. *Proceedings of IEEE* **72**(2), 230–233 (February 1984).

[15] G. LIU, J. QUINTERO, AND V. M. VENKATASUBRAMANIAN. Oscillation monitoring system based on wide are synchrophasors in power system. *2007 IREP symposium Bulk Power System Dynamics and Control* pp. 1–13 (August 2007).

[16] L. LJUNG. "System Identification: Theory for the User" New Jersey: Prentice-Hall (1999).

[17] D. G. MANOLAKIS, V. K. INGLE, AND S. M. KOGON. "Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing" Norwood, Massachusset: Artech House, Inc. (2005).

[18] P. V. OVERSCHEE AND B. DE MOOR. Subspace algorithms for the stochastic identification problem. *Proceedings of the 30th conference on decision and control* pp. 1321–1326 (december 1991).

[19] P V OVERSCHEE AND B. DE MOOR. A unifying theorem for three subspace system identification algorithms. *Automatica, Pergamon* **31**(12), 1853–1864 (1995).

[20] P V. OVERSCHEE AND B. DE MOOR. A subspace approach to stochastic realization and positive real sequence. pp. 1–19 (september 1996).

[21] P V OVERSCHEE AND B. D. MOOR. "Subspace Identification for Linear Systems: Theory, Implementation, Applications" London: Kluwer academic publishers (1996).

[22] B. PAL AND B. CHAUDHURI. "Robust Control in Power System" springer (2005).

[23] M. PARASHAR AND J. MO. Real time dynamics monitoring system: Phasor applications for the control room. *IEEE 42nd Hawaii International Conference on System Sciences* pp. 1–11 (January 2009).

[24] J. W. PIERRE, D. J. TRUDNOWSKI, AND M. K. DONNELLY. Initial results in electromechanical mode identification from ambient data. *IEEE Transactions on Power Systems* **12**(3), 1245–1250 (August 1997).

[25] B. PORAT AND B. FRIEDLANDER. On the accuracy of the kumaresan tufts method for estimating complex damped exponentials. *IEEE Transactions on acoustics, speech, and signal processing* **35**(2), 231–235 (February 1987).

[26] T. K. SARKAR AND O. PEREIRA. Using the matrix pencil method to estimate the parameters of a sum of complex exponentials. *IEEE Antennas and Propagation Magazine* **37**(1), 48–55 (February 1995).

[27] P STOICA. "Introduction to Spectral Analysis" NJ: Prentice-Hall (1997).

[28] D. J. TRUDNOWSKI, J. W. PIERRE, N. ZHOU, J. F. HAUER, AND M. PARASHAR. Performance of three mode-meter block-processing algorithms for automated dynamic stability assessment. *IEEE Transactions on Power Systems* **23**(2) (May 2008).

[29] R. W. WIES. "Estimating Low-Frequency Electromechanical Modes of Power Systems Using Ambient Data" PhD thesis, University of Wyoming (May 1999).

[30] R. W. WIES, A. BALASUBRAMANIAN, AND J. W PIERRE. Using adaptive step-size least mean squares (ASLMS) for estimating low-frequency electromechanical modes in power systems. *9th International Conference on Probabilistic Methods Applied to Power Systems* (June 2006).

[31] R. W. Wies and J. W. Pierre. Use of least-mean squares (LMS) adaptive filtering technique for estimating low-frequency electromechanical modes in power systems. *Proceedings of the America Control Conference* (May 2002).

[32] R. W. Wies, J. W. Pierre, and D. J. Trudnowski. Use of ARMA block processing for estimating stationary low-frequency electromechanical modes of power systems. *IEEE Transactions on Power Systems* **18**(1) (February 2003).

[33] D. H. Wilson, K. Hay, and G. J. Rogers. Dynamic model verification using a continuous modal parameter estimation. *IEEE Bologna PowerTech Conference, Bologna, Italy* pp. 1–6 (June 2003).

[34] N. Zhou, J. W. Pierre, D. J. Trudnowski, and R. T. Guttromson. Robust RLS methods for online estimation of power system electromechanical modes. *IEEE Transactions on Power Systems* **22**(3) (august 2007).

[35] N. Zhou, D. J. Trudnowski, J. W. Pierre, and W A. Mittelstadt. Electromechanical mode online estimation using regularized robust RLS methods. *IEEE Transactions on Power Systems* **23**(4) (Novemeber 2008).

[36] N. Zhou, D. J. Trudnowski, J. W. Pierre, and W. A. Mittelstadt. Electromechanical mode online estimation using regularized robust (RLS) methods. *IEEE Transactions on Power Systems* **23**(4) (November 2008).

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL I.P.N.
UNIDAD GUADALAJARA

"2010, Año de la Patria, Bicentenario del Inicio de la Independencia y Centenario del Inicio de la Revolución"

El Jurado designado por la Unidad Guadalajara del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional aprobó la tesis

Monitoreo en tiempo real de dinámicas oscilatorias de sistemas de potencia usando algoritmos canónicos y recursivos

Real-time monitoring of power system oscillatory dynamics using canonical and recursive algorithms

del (la) C.

Ismael MORENO NUÑEZ

el día 17 de Agosto de 2010.

Dr. Juan Manuel Ramírez Arredondo
Investigador CINVESTAV 3C
CINVESTAV Unidad Guadalajara

Dr. Arturo Román Messina
Investigador CINVESTAV 3C
CINVESTAV Unidad Guadalajara

Dr. Antonio Ramírez Treviño
Investigador CINVESTAV 3A
CINVESTAV Unidad Guadalajara