UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MULTI-MODAL BEHAVIOR AND CLUSTERING IN DYNAMICAL SYSTEM WITH APPLICATIONS TO WIND FARMS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

BY

YONG MA
NORMAN, OKLAHOMA
2009

MULTI-MODAL BEHAVIOR AND CLUSTERING IN DYNAMICAL SYSTEM WITH APPLICATIONS TO WIND FARMS

A DISSERTATION APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

_____
Dr. Thordur Runolfsson, Chair

_____
Dr. Ning Jiang

_____
Dr. Joseph P. Havlicek

_____
Dr. Mark B. Yeary

_____
Dr. Kyran D. Mish

ACKNOWLEDGEMENTS

I would first like to express my deepest appreciation to my academic advisor Professor Thordur Runolfsson for his support, assistance, and guidance over the past five years. Without his persistent help and encouragement, I would have not had the chance to explore such interesting and fruitful research topics, nor would the dissertation have been possible.

I would also like to thank my committee members, Professor Joseph Havlicek, Professor Mark Yeary, Professor John Jiang and Professor Kyran Mish for making my dream with "a doctoral cap" come true. Specifically, I would like acknowledge Professor John Jiang for his extraordinary insight in power industry, which has resulted in the publication of valuable research.

Finally, I would like to thank my family for their continuous support, sacrifice and encouragement in all the years of my graduate study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The objective of this research is to develop a comprehensive model identification approach for complex multi-modal systems based on spectral theory for nonreversible Markov process that entails (i) model reduction techniques for a nonreversible Markov chain, (ii) the identification of the modal dynamics, and (iii) modeling and identification of local dynamics. This dissertation addresses the theoretical approach, algorithmic development, computational efficiency and numerical examples of the developed techniques.

The dissertation then presents a novel methodology for clustering wind turbines of a wind farm into different groups. The method includes creation of a Markov transition matrix given the power output of each turbine, spectral analysis of the transition matrix and identification approach of each group. An application of the method is provided based on real data of a wind farms consisting of 25 turbines and 79 turbines, respectively. The application shows that those distinct wind farm groups with different dynamic output characteristics can be identified and the turbines in each group can also be determined.

1 INTRODUCTION

1.1 Motivation

Although computer and network technologies are increasingly becoming powerful, there are many problems in simulation and control of complex dynamical system, climate modeling, machine learning, bio-informatics, chemistry, etc., where the dimensions and time scales of interest remain entirely out of reach of current technologies capability, and will remain beyond the available capability in the foreseeable future. Consequently, in order to obtain models for analysis and decision there is a need for model identification and reduction/simplification. Many dynamic systems of interest exhibit multi-modal behavior and clustering in state space. The systems we study in this research are systems that are subject to uncertainty, either internal (e.g. parametric) or external (e.g. disturbance), that contributes to the modal behavior. For instance, the system may be subject to an external random driving force that drives it from a desired to an undesired mode of operation or be subject to an uncertain internal parameter that may cause structural change, such as bifurcation, from a desired

mode of operation to an undesired mode. Multi-modal behavior and clustering in a complex system is characterized by a time scale separation between the infrequent changes in mode and "normal" system dynamics. In addition to the time-scale separation, the modal behavior induces a spatial decomposition of the system dynamics as well. The identification process seeks to replace a complex system with modal dynamics and subsystems of substantially lower complexity that capture the essential or dominate characteristics of the input-output behavior of the system response.

Complex engineered systems that comprise of many different components that are subject to performance and operational constraints often exhibit complex behavior that was not anticipated at design time. The root of this behavior is often due to complex dynamic interactions that are difficult predict and even harder to control. Frequently, one mode of behavior is the desired or normal or optimal one while the others represent either undesired or abnormal, or suboptimal or even failed operation. In order to develop a control strategy for maintaining the system in the desired or optimal mode, a model of the system behavior that is sufficiently complex to capture the multi-modal behavior but simple enough for control system design is required.

Examples of systems that exhibit behavior of this sort arise in numerous applications [1]. For instance in power networks [2] [3], the system is subject to

constraints set by load limits on lines and exceeding these limits can cause a change in the network topology. Furthermore, nonlinear effect due to reactive loads can lead to bifurcations causing voltage collapse [79] [80]. Finally, loss of critical elements can cause loss of synchronization leading severe power flow instabilities [81]. Similar effects can be found in the dynamics of other networks, including communication networks [82] [83]. In many chemical processes individual system components (e.g. reactors) are subject to bifurcating from an efficient operation to an inefficient (i.e. non productive) state as a function of other system variables that often are subject to dynamic laws themselves. When the system is further integrated through physical feedback to increase efficiency (e.g. using waste heat to preheat supply streams) the overall dynamics become very complex and can exhibit unexpected transitions to undesired operating state [4]. Similar behavior has been observed in the thermodynamics of highly efficient, environmentally friendly heat pumps [5].

Most multi-modal systems are modeled as hybrid systems, i.e. their behavior is modeled by interacting continuous and discrete dynamics. Here continuous and discrete refers to the spatial components, i.e. the continuous dynamics take values in a continuous set while the discrete dynamics take values in a discrete (usually finite) set. Numerous papers on analysis and control of hybrid systems have been published and the understanding of the dynamics

behavior of such systems, including stability, is fairly well developed. Most, if not all, of the research on the control of hybrid systems relies on the assumption that there is available a model of the system. However, only in rare cases is it possible to obtain such models from first principles and, in fact, most systems that admit a hybrid systems formulation are an approximation of complex dynamical systems whose behavior has a separation into continuous and discrete components through spatial scale separation as well as time scale separation. For such systems modeling from first principles is next to impossible and control oriented system models must be based on experimental or large scale simulation data [6].

Based on the data either from real system operation or large scale dynamic simulation models, we will develop a comprehensive identification strategy for modeling complex multi-modal systems as hybrid systems that entails (i) the identification of the modal dynamics (the number of modes, discrete sets); (ii) the partition of the system state space into components and identification of dynamics between components (dynamics in slow time scale); (iii) modeling and identification of local dynamics (dynamics in fast time scale). For part (i) and (ii) the identification has received considerable attention in the literature [7] [20] [21]. Part (iii) includes conventional identification techniques and has been studied in

great detail for linear time invariant models as well as several nonlinear models [22] [60] [61] [72] [73] [74] [75].

The amount of wind energy being harvest into electric power in the world is growing rapidly. With construction of more and more large-scale wind farms, the integration of such wind farms into the power grid is a challenge to the current power system of paramount importance. In that context, in order to study the impacts of wind farms on quality measures of electric power such as reliability assessment, system modeling, power output forecasting, etc., understanding the dynamics of the power output of a wind farm is important to the integration of large scale wind energy into the power system. In a large complex dynamic engineering system, such as a wind farm, clustering is an effective way to reduce the model complexity and improve the understanding of its local dynamics. Each cluster is a collection of wind turbines that behave similarly in terms of the dynamics of their power output.

1.2 Current State of the Art in the Field

In recent years, there has been considerable research for approximation of the essential behavior of complex deterministic and stochastic dynamical systems [8] [9] [10] [11] [12] [13]. The approach in all of these papers is based on an operator approach for modeling the overall dynamics and spectral theory of the

operator for characterizing certain features of the dynamics. In particular, eigenvalues of the operator which are on or close to the unit circle and their corresponding eigenfunctions play an important role. An efficient technique in the numerical approximation of complicated dynamical behavior has been proposed by Dellnitz and Junge [8] [9]. In their papers, the authors model a statistical description of the essential dynamical behavior by an underlying invariant measures (i.e. SRB measure). The main idea of the approximation is to define an operator (the Frobenius-Perron operator) on the space of probability measures whose fixed points are invariant measures, then to discretize this operator via a finite dimensional Galerkin projection and finally to solve an eigenvalue problem for the resulting matrix to obtain an approximation to the invariant measures. The eigenvectors of the resulting matrix corresponding to eigenvalues on (or close to) the unit circle can be used to identify (almost) invariant sets which are regions in state space where the system spends a long time before the dynamic system leads to different region and (almost) cyclic behavior, which are regions where the system dynamics permutes cyclically.

Huisinga, utilizes the concept of metastability and develops theoretical justification for a algorithmic strategy for the identification of metastable subsets [10] [11]. In order to make the identification strategy numerically applicable, Huisinga extends deterministic transfer operators to stochastic transition

functions through a randomizing process and then decomposes the state space into metastable subsets and captures essential statistical behavior based on the eigenfunctions of the transition functions corresponding to eigenvalues close to one [10] [11]. They successfully demonstrate the algorithmic approach in the study of molecular dynamics and outline strategies for studying larger molecular systems.

Coifman, Lafon and Nadler provide a framework based on diffusion processes for capturing the long time evolution characteristics of some complex dynamic system using data which is sampled from the system [12] [13]. Through a data driven construction of diffusion kernels they first construct a stochastic transition matrix defining a random walk. Then they construct a so called diffusion map that maps the original data points into a Euclidean space and obtain a new description of data sets based on spectrum of the Markov transition matrix. The diffusion map induces the so called diffusion distance to classify points in terms of their connectivity in the new Euclidean space and as a result classifies the system dynamics into clustering components [12] [13].

1.3 Contribution of the Dissertation

Most of the above approaches rely on spectral properties of a reversible Markov process. Reversibility describes the property that the Markov process

and its time-reversed counterpart are statistically the same. For a reversible process the transfer operator or transition matrix is self-adjoint. However, most systems of interest in engineering applications are not reversible process and thus an extension of the above methods is needed. The framework examined in this research is for identifying a metastable partition in terms of the spectral analysis of a general non-reversible Markov processes using diffusion distance and sign structure concepts. Specifically, we assume that the transitive behavior between different metastable subsets is disturbance or noise induced and that the overall system can be modeled as a Markov process. We also define a multiplicative reversible Markov process for the nonreversible case which has same stationary distribution as the original process. Thus, the spectrum of the multiplicative reversibilization has a well defined relationship with that of the original process. We utilize this relationship to obtain a reduced order approximate operator for the original Markov process which is shown to be a good approximation theoretically. We also identify metastable components of the state space using diffusion distance method and/or sign structure method. Once metastable components have been identified, the original data in each metastable component can be used to estimate local dynamics in terms of subspace model and noise sequence estimation.

Finally, we propose a novel methodology to cluster wind turbines of a wind farm into different groups based on our preliminary research. We first build a weighted graph to represent the complex relationships between power outputs of wind turbines. The graph is used to construct a Markov Chain and estimate the likelihood of any two wind turbines belong to the same cluster. We analyze the spectral properties of the Markov chain to identify the number of clusters. With the proposed method, the elements of each cluster can be identified in both diffusion distance method and sign structure method. Theoretical study and case studies show that the proposed methodology simplifies the model of the dynamics of power output of a wind farm without compromising the overall dynamic characteristics of the original system asymptotically.

1.4 Organization of the Dissertation

This dissertation is focused on developing a novel approach based on spectral theory for nonreversible Markov process (i.e. transfer operator formulation for system dynamics) for identification of modal transition dynamics and identification of local dynamics. The remainder of the dissertation is organized as follows: Chapter 2 introduces the fundamentals of multi-modal behavior and clustering which include problem formulation and overview of mathematical approach. Chapter 3 discusses some asymptotic properties of

probability distributions and convergence rates for both reversible and nonreversible process, develops spectral properties for nonreversible processes and presents the low dimensional approximations for general nonreversible Markov chains. Chapter 4 provides a comprehensive approach for identification of metastable components and local dynamics and presents some illustrative numerical examples. Chapter 5 proposes a novel method to clustering wind turbines of a wind farm into different groups and presents results of cluster analysis of large-scale wind farm located in northwest of Oklahoma.

2 FUNDAMENTALS OF MULTI-MODAL BEHAVIOR AND CLUSTERING

2.1 Introduction

As we stated earlier we assume that the system of interest has characteristic multi-modal behavior, i.e. the system spends considerable time in one region of state space before transitioning to another region where it again spends a long time before transitioning again. To model this behavior requires the identification of the following three system characteristics: (i) the domains or regions in state space where the system spends a long time between transitions, (ii) the dynamic laws that characterize the transitive behavior between the domains and (iii) the dynamics of the system inside each domain. The basis of the approach of the identification of the domain and the transitive or modal dynamics lie in the spectral theory of Markov operators. In particular, the transitive nature of the dynamics is characterized as metastable behavior and the metastability identified through the spectral properties of the operator. In

addition, we assume the local dynamic behavior is characterized by a single point attractor and can be approximated locally by a linear system. Subspace identification is a convenient identification technique which allows the estimation of state space equation directly from the given data, i.e. from the data of each domain or region. In this chapter, we present the model formulation as well as the basic theoretical foundations of the proposed approach.

2.2 Problem Formulation

Consider a discrete time dynamic system with state vector $x(t),\ t \in \mathbf{Z}_+$. The system is subject to uncertainties, either external or internal. We assume that as function of the uncertainties, the system undergoes abrupt changes that result in dramatic changes in dynamic behavior. However, the frequency of the abrupt changes is relatively low compared to the normal system dynamics. We consider a discrete time dynamic system in a state space model for the system of the form

$$x_{k+1} = A(r_k)x_k + f(r_k) + w_k \tag{2.1}$$

where $x_k \in \mathbf{R}^n$ is the continuous state, $w_k \in \mathbf{R}^n$ is a noise input modeled as Gaussian white noise $N(0, R(r_k))$ and r_k, the modal variable, is a discrete valued variable taking values in a finite set $\{1,\ldots,q\}$. The component $f(r_k)$ is mode dependent bias term. We also assume that the state space can be partitioned into a partition $A_1,\ldots,A_q$ (the A_i's do not overlap and their union is all of $\mathbf{R}^n$)

where the set A_i corresponds the discrete value $r_k = i$. Given a collection of N measurements x_k, we want to identify the number of subsystems q, the state space partitions $A_1, \ldots, A_q$ and the model parameters $A(i)$, $f(i)$ and $R(i)$. In this dissertation, we propose a two step process method for the identification of the system. The first step is based on discretization of the original process and modeling the process as a Markov chain on a finite dimensional state space $X = \{x^1, \ldots, x^N\}$. Based on the resulting Markov chain, we identify the number q of partition components, the partition components themselves and the modal dynamics between partition components. In the second step we map the original data onto the partition components and identify the local dynamic models using conventional subspace identification methods with noise estimation.

2.3 Overview of Mathematical Approach

In this section we give an overview of the mathematical methods that will be employed in the solution of the identification problem of complex dynamics. We start with a general formulation of the mathematical problem and then present the idea behind the proposed approach to the solution of the problem as well as some preliminary findings. For more details see, e.g., [9][11][14][15].

2.3.1 Stochastic Transition Functions and the Perron-Frobenious Operator

Consider a Markov chain $(\xi_0, \xi_1, \ldots)$ defined on $\mathbf{R}^n$ with transition function $p(\xi, A)$ defined for all $\xi \in \mathbf{R}^n$ and Borel sets $A \in \mathcal{B}$, $\mathcal{B} = \mathcal{B}(\mathbf{R}^n)$ i.e. $p(\xi, A)$ satisfies

1) $p(\cdot, A)$ is measurable for each fixed set $A \in \mathcal{B}$,

2) $p(\xi, \cdot)$ is probability measure on $(\mathbf{R}^n, \mathcal{B})$ for each fixed $\xi \in \mathbf{R}^n$.

Consider a time invariant discrete time stochastic dynamical system on $\mathbf{R}^n$ that can be represented by the Markov chain with transition function $p(x, A) = \Pr(x_{k+1} \in A \mid x_k = x)$. Let $\mathcal{M}_1 = \mathcal{M}_1(\mathbf{R}^n)$ be the space of all probability measures on $\mathbf{R}^n$ and assume that the initial state x_0 has distribution $\nu \in \mathcal{M}_1$. Then the distribution of x_1 is

$$\nu_1(A) = \int_{\mathbf{R}^n} p(\xi, A)\nu(d\xi) = (P\nu)(A) \tag{2.2}$$

where $P : \mathcal{M}_1 \to \mathcal{M}_1$ is the Perron-Frobenius operator [32]. It follows that the distribution of x_n is $\nu_n(A) = (P^n\nu)(A)$. A measure μ is called an invariant measure for the Markov chain if

$$\mu(A) = \int_{\mathbf{R}^n} p(\xi, A)\mu(d\xi) = (P\mu)(A) \tag{2.3}$$

for all Borel sets A. We assume that μ is a unique invariant probability measure for $p(x, A)$, i.e. the Markov chain is assumed to be ergodic.

A transition function p is called reversible w.r.t the invariant probability measure μ if

$$\int_A p(\xi, B)\mu(d\xi) = \int_B p(\xi, A)\mu(d\xi) \tag{2.4}$$

The value

$$p(B, A) = \frac{1}{\mu(B)} \int_B p(\xi, A)\mu(d\xi) \tag{2.5}$$

is the probability of transitioning from set B into set A in one step [14] [32]. Note that if $\mu(B) = 0$, we let $p(B, A) = 0$. A time reversed system can be defined by

$$\tilde{p}(A, B) = \frac{\mu(B)p(B, A)}{\mu(A)} \tag{2.6}$$

A stochastic transition function with invariant measure μ is called uniformly ergodic [14], if there are constants $r < 1$ and $K > 0$, such that $\left\| p^n(\xi, \cdot) - \mu \right\|_{\text{var}} \leq Kr^n$ for $n = 0, 1, 2, \ldots$ and all $\xi \in X$ where $\|\cdot\|_{\text{var}}$ is the variation norm as defined in [14].

2.3.2 Metastability

Let A and B be measurable sets on $\mathbf{R}^n$, assume that the distribution of the initial state is the invariant measure μ and let $p(B, A)$ be the transition probability from B to A in one step. We note that $p(B, A)$ characterizes the

dynamical fluctuations of the distribution the Markov chain within the invariant distribution μ. The flowing definition is from [10].

DEFINITION 2.1 *A Borel set* A *is said to be invariant if* $p(A,A)=1$ *and metastable if* $p(A,A)\approx 1$. *Therefore, a metastable set is almost invariant.*

We note that if set A is invariant and if the system starts in A, it will stay in there forever. On the other hand if the initial state belongs to a metastable set then the system state will stay there for a long time but will eventually exit the set with positive probability. Consider a partition $A_1,\ldots,A_q$ of the state space $\mathbf{R}^n$, i.e. $\bigcup_{i=1}^{q} A_i = \mathbf{R}^n$ and $A_i \cap A_j = 0$ unless $i = j$. We are interested in finding a partition such that $\sum_{i=1}^{q} p(A_i \mid A_i) \approx q$, i.e. a metastable partition. For such a partition, if it exits, we can approximate the slow time scale transitions of the Markov process between the metastable components with a finite dimension Markov chain defined on a state space S of dimension q. The characterization of the metastable partition can be related to spectral properties of P and the indicator (characteristic) functions of the metastable sets which may be identified with the eigenfunctions corresponding to the dominant (close to one) eigenvalues of P [10] [16].

2.3.3 Discretization of Transfer Operator

We start with a Markov process with transition function $p(x, A)$ on $\mathbf{R}^n$ and assume that the process is defined on a compact subset X of $\mathbf{R}^n$. Let v be a probability measure on X and let $X_1, \ldots, X_N$ be a partition of X (not necessary metastable). Let m be the Lebegue measure on X and define a discrete measure on X by

$$v_N(dx) = \sum_{i=1}^{N} v(X_i)\chi_{X_i}(x)\frac{m(dx)}{m(X_i)} \tag{2.7}$$

where $\chi_{X_i}(x)$ is the indicator function for X_i. We note that $v_N(X_j) = v(X_j)$, $j = 1, \ldots, N$. Furthermore, for any Borel set A,

$$\begin{aligned}(Pv_N)(A) &= \int_X p(x, A)v_N(dx) \\ &= \sum_{i=1}^{N} v(X_i)\frac{\int_{X_i} p(x, A)m(dx)}{m(X_i)} \\ &= \sum_{i=1}^{N} v(X_i)p_m(X_i, A)\end{aligned} \tag{2.8}$$

If we define $p_{i,j} = p_m(X_i, X_j)$ then the matrix P_N with entries $p_{i,j}$ is a stochastic matrix and $(Pv_N)(X_j) = (\overline{v}_N P_N)_j$ where $\overline{v}_N = (v(X_1)\cdots v(X_N))$. Thus, the stochastic matrix (transition matrix) P_N is a discretization of the operator P that agrees with P on the discretization components. We remark that if X is not compact then the above construction still works provided we replace the Lebegue measure with a finite measure (e.g. probability measure). We also note that as $N \to \infty$ it can be shown that the "error" between the original and discretized operators converges to zero in an appropriate sense. Let the

sequence of data $\{x_k, k = 0,1,\ldots\}$ be obtained from a realization of the discrete time Markov process. For an event C, let $\#[C]$ denote the number of times the event happened, then

$$p(A,B) \approx \frac{\#[x_k \in A \text{ and } x_{k+1} \in B]}{\#[x_k \in A]} \tag{2.9}$$

where the right hand side converges to the left hand side under the appropriate conditions as the number of data points goes to infinity [14]. An $N \times N$ matrix C with entries c_{ij} that count the number of times there is a transition from element X_i to element X_j, i.e. starting with $c_{ij} = 0$ we increase c_{ij} by one if $x_k \in X_i$ and $x_{k+1} \in X_j$ for some k is defined. After we have run through all of the data we obtain a transition matrix $\tilde{P}_N$ by row normalization of C as

$$c_{ij} \to \tilde{p}_{ij} = \frac{c_{ij}}{\sum_i c_{ij}} \tag{2.10}$$

We note that in large N $\tilde{P}_N = P_N$ and, with some abuse of notation, we use P_N in place of $\tilde{P}_N$.

2.3.4 The Nyström Extension Method

The Nyström extension method [23] [24] [25] is a technique for approximating an integral equation using a quadrature rule, i.e.

$$\int_a^b h(x)dx \approx \sum_{k=1}^{n} w_k h(x_k) \tag{2.11}$$

The Nyström method can be used for finding numerical approximations to eigenfunction problems of the form

$$\int_a^b P(x,y)\phi(y)dy = \lambda\phi(x) \tag{2.12}$$

Indeed, letting $h(y) = P(x,y)\phi(y)$ in (2.11), we employ the simple quadrature rule by evaluating the equation at a set of evenly sampled points $\xi_1, \xi_2, \ldots, \xi_n$ on the interval $[a, b]$. The resulting equation of the eigenfunction problem becomes

$$\frac{b-a}{n}\sum_{k=1}^{n} P(x,\xi_k)\hat{\phi}(\xi_k) = \lambda\hat{\phi}(x) \tag{2.13}$$

Here $\hat{\phi}(x)$ is an approximation to the true $\phi(x)$. Assume $P(x,y)$ has n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and denote the corresponding eigenfuctions as $\phi_1, \phi_2, \ldots, \phi_n$. The approximation equation (2.13) yields

$$\frac{b-a}{n}\sum_{k=1}^{n} P(x,\xi_k)\hat{\phi}_i(\xi_k) = \lambda_i\hat{\phi}_i(x) \tag{2.14}$$

Without loss of generality, let $[a, b]$ be $[0, 1]$. Let P be the $n \times n$ matrix with elements $P_{ij} = P(\xi_i, \xi_j)$ for $i, j = 1, \ldots n$, $U \in \mathbf{R}^{n\times n}$ and Λ be diagonal matrix with entries $\lambda_1, \lambda_2, \ldots, \lambda_n$. Consider the matrix eigenproblem

$$PU = U\Lambda \tag{2.15}$$

Comparing (2.14) and (2.15), we arrive at the following,

$$\hat{\phi}_i(\xi_j) = \sqrt{n} U_{j,i}, \qquad \lambda_i = \frac{\lambda_i}{n} \tag{2.16}$$

Substituting these back into (2.13) gives the Nyström extension for the ith eigenfunction [23],

$$\hat{\phi}_i(x) = \frac{1}{n\lambda_i} \sum_{k=1}^{n} P(x, \xi_k) \hat{\phi}_i(\xi_k) = \frac{\sqrt{n}}{\lambda_i} \sum_{k=1}^{n} P(x, \xi_k) U_{k,i} \tag{2.17}$$

2.3.5 Subspace Identification

Subspace identification is used to estimate linear stationary state space models from given input and output data. A general stochastic linear state space model can be written as

$$\begin{aligned} \tilde{x}_{k+1} &= A\tilde{x}_k + Bu_k + e_k \\ y_k &= C\tilde{x}_k + Du_k + f_k \end{aligned}$$

where $\tilde{x}_k \in \mathbf{R}^n$, $u_k \in \mathbf{R}^m$, $y_k \in \mathbf{R}^l$ are the system state, input and output, respectively, and $A \in \mathbf{R}^{n \times n}$, $B \in \mathbf{R}^{n \times m}$, $C \in \mathbf{R}^{l \times n}$ and $D \in \mathbf{R}^{l \times m}$ are system matrices. The inputs $e_k \in \mathbf{R}^n$ and $f_k \in \mathbf{R}^l$ are assumed to be zero mean Gaussian white noise processes with covariances R_e and R_f, respectively. Assuming that the y_k is an observed output define the one step ahead predictor for the state $\tilde{x}_k$ as $x_k = E\left[\tilde{x}_k \mid y_p, p \le k-1\right]$. Then it can be shown that x_k satisfies the innovation equation [22],

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Kw_k \\ y_k &= Cx_k + Du_{kl} + w_k \end{aligned} \tag{2.18}$$

The process $w_k = y_k - Cx_k$ is the so-called innovation sequence that can be shown to be a zero mean Gaussian white noise with covariance $E(w_k w_k^T) = R = CPC^T + R_f$ where $P = E\left[\left(\tilde{x}_k - x_k\right)\left(\tilde{x}_k - x_k\right)^T\right]$ and $K \in \mathbf{R}^{n \times l}$ is a Kalman filter gain [22]. The order n is assumed known or estimated by methods proposed in [62] [63] [64] [65]. In order to get consistent estimation, we introduce the following basic assumptions for (2.18): (i) (A, C) is observable and $(A, [B \ K])$ is controllable, (ii) the input u and the innovation sequence w are uncorrelated.

If we have the measurements $y(k)$, $u(k)$, $k = 1, \ldots, N + r - 1$ available where r denotes future horizon, we introduce the input matrix Y , output matrix U and past data matrix Φ as

$$Y = \begin{bmatrix} Y_r(1) & Y_r(2) & \ldots & Y_r(N) \end{bmatrix} \tag{2.19}$$

$$U = \begin{bmatrix} U_r(1) & U_r(2) & \ldots & U_r(N) \end{bmatrix} \tag{2.20}$$

$$\Phi = \begin{bmatrix} \varphi_s(1) & \varphi_s(2) & \ldots & \varphi_s(N) \end{bmatrix} \tag{2.21}$$

$$Y_r(k) = \begin{bmatrix} y(k) \\ y(k+1) \\ \vdots \\ y(k+r-1) \end{bmatrix} \tag{2.22}$$

$$U_r(k) = \begin{bmatrix} u(k) \\ u(k+1) \\ \vdots \\ u(k+r-1) \end{bmatrix} \tag{2.23}$$

$$\varphi_s(k) = \begin{bmatrix} y(k-1) \\ \vdots \\ y(k-s_1) \\ u(k-1) \\ \vdots \\ u(k-s_2) \end{bmatrix} \tag{2.24}$$

s_1 and s_2 represent the number of data prior to time k. For general case considered in [22], the overall estimation method can be outlined in four steps as follows:

Step 1 From the input-output data, formulate G matrix as

$$G = \frac{1}{N} Y \widetilde{U} \Phi^T \tag{2.25}$$

where $\widetilde{U} = I - U^T (UU^T)^{-1} U$.

Step 2 Select weighting matrices $W_1 (rl \times rl)$ which is invertible and W_2 and perform singular value decomposition (SVD)

$$\hat{G} = W_1 G W_2 \approx U_1 S_1 V_1^T \tag{2.26}$$

where S_1 are first n (system order) significant singular values. U_1 and V_1 are matrices of corresponding right and left eigenvectors.

Step 3 Select a full rank matrix Z and estimate the extended observability matrix

$$\hat{O}_r = W_1^{-1} U_1 Z \tag{2.27}$$

In terms definition of observability matrix, A and C can be solved from $\hat{O}_r$.

Step 4 Estimate B and D by solving least squares problem

$$y(k) = C(qI - A)^{-1}Bu(k) + Du(k) + C(qI - A)^{-1}w(k) + w(k) \qquad (2.28)$$

The biggest challenge is to estimate the noise input w_k. Given a set of data, the noise covariance matrix R needs to be estimated as well. Lin, Qin and Ljung [21] show that consistency of (closed-loop) subspace identification methods can be achieved through innovation estimation to estimate the noise sequence. We summarize the noise estimation as follows.

For simplicity, assume we have the N measurements. By choosing the future horizon $r = 1$ in (2.22), define output matrix Y as

$$Y = \begin{bmatrix} Y_1(s+1) & Y_1(s+2) & \cdots & Y_1(N) \end{bmatrix} \qquad (2.29)$$

Assuming the past horizon is s which means there are s data prior to time k, define matrix Z as

$$Z = \begin{bmatrix} Z_s(1) & Z_s(2) & \cdots & Z_s(s) \end{bmatrix} \qquad (2.30)$$

where $Z_s(k) = \begin{bmatrix} y(k) \\ \vdots \\ y(N-s-1+k) \\ u(k) \\ \vdots \\ u(N-s-1+k) \end{bmatrix}$

In our work, we consider systems without an input sequence $u(k)$. Therefore, a least squares estimate of the noise sequence is

$$\hat{E}_1 = Y - YZ(Z^T Z)^{-1} Z^T \qquad (2.31)$$

3 MODEL REDUCTION OF NONREVERSIBLE MARKOV CHAINS

3.1 Introduction

As we stated earlier, we assume that the system of interest has characteristic multi-modal behavior, i.e. the system trajectories cluster in several subsets of the state space. In the dissertation, we model the system behavior as a Markov process and consider the problem of finding a lower dimensional approximation of the Markov process first before we try to identify the clusters and the system local dynamics in each cluster. This type of a problem arises in many applications that all share the common characteristic that the sample trajectories of the Markov process cluster and the approximation problem is to find a representation of the clustering and an approximate operator that represents the system behavior on the corresponding lower dimensional space. The problem of finding an approximate operator is much simpler when the Markov chain is reversible and several solution approaches have been developed for this case [10]

[11] [12] [13]. Most of these approaches rely on spectral properties of Markov chains. In this chapter, we consider the general nonreversible case and develop a reduction techniques that parallels in some respects the methods developed in [12] for reversible case.

3.2 Multiplicative Reversibilization

Let X be a finite set and consider a Markov chain $\{x_k, k = 0,1,\ldots\}$ on X with transition probabilities $p(x,y) = \Pr(x_{k+1} = y \mid x_k = x)$. Let P be the $N \times N$ transition matrix with entries $p(x,y)$ ($N = \dim(X)$). We assume that the Markov chain is aperiodic and irreducible, i.e. the chain is ergodic. Then there exists an unique invariant or stationary distribution π such that

$$\lim_{n\to\infty} p^n(x,y) = \pi(y) \tag{3.1}$$

where $p^n(x,y)$ is the (x,y) entry of P^n. It is well known that the stationary distribution satisfies the identity $\pi P = \pi$, i.e. π is the left eigenvector of P corresponding to the eigenvalue 1 [14]. The rate of convergence in (3.1) is of considerable interest. In particular, if the rate of convergence is very fast then the Markov chain may be approximated by its stationary distribution. For reversible Markov chains, i.e. chains that satisfy the detailed balance condition

$$\pi(x)\,p(x,y) = \pi(y)\,p(y,x) \tag{3.2}$$

It can be shown that the rate of convergence in (3.1) is determined by $\beta_1 = \max(\lambda_1, |\lambda_{N-1}|)$, where λ_1 is the second largest eigenvalue of P and λ_{N-1} the smallest one [17] (recall that an ergodic reversible chain P has all real eigenvalues belonging to the interval (-1, 1]). In fact the rate of convergence in (3.1) behaves like β_1^n. For nonreversible chains a similar result was proven by Fill [17].

Let $\tilde{P} = (\tilde{p}(x, y))$ be the time reversal of P, i.e.

$$\tilde{p}(x, y) = p(y, x)\frac{\pi(y)}{\pi(x)} \tag{3.3}$$

Then $\tilde{P}$ is an ergodic Markov transition matrix that has the same unique stationary distribution π as P. Define the multiplicative reversibilization $M(P)$ of P by

$$M(P) = P\tilde{P} \tag{3.4}$$

It is easy to see that $M(P)$ is a reversible transition matrix that also has the same stationary distribution π. Furthermore, the eigenvalues of $M(P)$ are real and nonnegative, i.e. they belong to the interval [0, 1]. Indeed, let $D = diag(\pi_1, \ldots, \pi_N)$ and define $S(M) = D^{1/2}MD^{-1/2}$. We note that $S(M)$ and M have the same eigenvalues, i.e. are algebraically isomorphic. Using the fact that $\tilde{P} = D^{-1}P^T D$ we have $S(M(P)) = S(P)(S(P))^T$ is nonnegative and thus has nonnegative eigenvalues.

3.3 Asymptotic Properties

Since $M(P)$ is a transition matrix it follows that its maximum eigenvalue is 1. Let $\beta_1(\mathrm{M})$ be the second largest eigenvalue of $M(P)$. For probability vectors μ, π define variation distance

$$\|\mu-\pi\|_{\text{var}} = \max_{A\subset X}|\mu(A)-\pi(A)| = \frac{1}{2}\sum_{x\in X}|\mu(x)-\pi(x)| \tag{3.5}$$

and following [17] define the chi-square distance from stationarity at time n as

$$\chi_n^2(\mu) = \sum_{x\in X}\frac{(\pi_n^\mu(x)-\pi(x))^2}{\pi(x)} \tag{3.6}$$

where $\pi_n^\mu(x)$ is the distribution of the chain at time n with initial distribution μ. The proof of the following result can be found in [17].

THEOREM 3.1 *Let P be an ergodic transition matrix on a finite state space X and let π be its stationary distribution. Then for any initial distribution μ*

$$4\left\|\pi_n^\mu-\pi\right\|_{\text{var}}^2 \le \chi_n^2(\mu) \le (\beta_1(M)^n)\chi_0^2(\mu) \tag{3.7}$$

Now let μ and v be two initial distributions and π_n^μ and π_n^v the corresponding distributions at time n. Then it is easy to see that

$$4\left\|\pi_n^\mu-\pi_n^v\right\|_{\text{var}}^2 \le (\beta_1(M)^n)(\sqrt{\chi_0^2(\mu)}+\sqrt{\chi_0^2(v)})^2 \tag{3.8}$$

In particular, asymptotically the variation distance between the distributions corresponding to different initial distributions behaves like $\beta_1(M)^{n/2}$.

3.4 Spectral Properties

We now discuss the relationship between the spectrum of P and the multiplicative reversibilization $M(P)$. In particular, we note that in the previous results the asymptotic convergence of the distribution of the chain with transition probabilities P is determined by an eigenvalue of the multiplicative reversibilization $M(P)$ and thus the relationship between the spectra of the two processes is of interest. Let $\lambda_0,\ldots,\lambda_{N-1}$ be the eigenvalues of P and $\beta_0,\ldots,\beta_{N-1}$ be the eigenvalues of $M(P)$ and assume that $\beta_i \geq \beta_{i+1},\ i=0,\ldots,N-1.$

PROPOSITION 3.1 *Assume that P has a unique stationary distribution π with $\pi_i > 0,\ i=0,\ldots,N-1$. Then there exist $\theta_0,\ldots,\theta_{N-1} \in [0,2\pi)$ such that $\lambda_i = \sqrt{\beta_i}e^{j\theta_i},\ i=0,\ldots,N-1$. In particular, $\beta_i = |\lambda_i|^2,\ i=0,\ldots,N-1.$*

Proof: We note the obvious fact that $\lambda_0 = \beta_0 = 1$. Recall that $M(P) = D^{-1/2}S(M)D^{1/2} = D^{-1/2}S(P)(S(P))^T D^{1/2}$ where $S(P) = D^{1/2}PD^{-1/2}$ is a square root of $S(M)$. Note that $M(P)$ and $S(M)$ are algebraically isomorphic and P and $S(P)$ are algebraically isomorphic. Since $S(M)$ is symmetric we can write $S(M) = V\Sigma V^T$ where $\Sigma = diag(\beta_0,\ldots,\beta_{N-1})$ and V is orthogonal. Define $\tilde{R} = V\sqrt{\Sigma}V^T$. Then $\tilde{R}$ is a square root of $S(M)$ and the eigenvalues of $\tilde{R}$ are $\sqrt{\beta_i},\ i=0,\ldots,N-1.$ Note that any other square root R of $S(M)$ is related to $\tilde{R}$ through $R = \tilde{R}U$ where U is a unitary matrix. Thus,

since $S(P) = D^{1/2} P D^{-1/2}$ is a square root of $S(M)$ we know that there exists a unitary $\tilde{U}$ such that

$$S(P) = D^{1/2} P D^{-1/2} = \tilde{R}\tilde{U} = V\sqrt{\Sigma} V^T \tilde{U} \tag{3.9}$$

Let $\lambda = \sigma + jw$ be an eigenvalue of $S(P) = \tilde{R}\tilde{U}$ (and P). Then there exist an eigenvector $v = y + jz$ corresponding to λ such that $\tilde{R}\tilde{U}v = S(P)v = \lambda v$ and therefore

$$(\tilde{R}\tilde{U}v)^H \tilde{R}\tilde{U}v = v^H U^H \tilde{R}^T \tilde{R}\tilde{U}v = |\lambda|^2 \|v\|^2 \tag{3.10}$$

Therefore, $|\lambda|^2$ is an eigenvalue of $\tilde{R}^T\tilde{R} = S(M)$ with eigenvector $\tilde{U}v$. Finally, since $\beta_i > 0$ we have $|\lambda_i| = \sqrt{\beta_i}$ and since $\lambda \in \boldsymbol{C}$ there exists an angle θ_i such that $\lambda_i = \sqrt{\beta_i} e^{j\theta_i}$, $i = 0, \ldots, N-1$. ■

This relationship between the spectrum of P and the multiplicative reversibilization $M(P)$ makes it possible to characterize the metastable partition by relating to spectral properties of $M(P)$. The indicator (characteristic) functions of the metastable sets can be identified with eigenfunctions corresponding to the dominant (closet to one) eigenvalues of $M(P)$ which correspond to eigenvalues of P close to the unit circle.

3.5 Diffusion Distance

We are interested in obtaining a reduced order approximate model and the identification of metastable components for the Markov chain with transition

matrix P. We noted earlier that if the underlying process is ergodic and the second largest eigenvalue of $M(P)$ is small then the convergence to π is fast and the stationary distribution is a good low dimensional approximate model. However, if $M(P)$ has several eigenvalues close to one then this convergence is not as fast. Let ν be the distribution of the initial state x_0 and $\pi_n^\nu = \nu P^n$ be the distribution of the Markov chain at time n. We now consider the problem of finding an approximate operator π_{an}^ν for π_n^ν that is good for all initial distributions ν and for which the error $\pi_n^\nu - \pi_{an}^\nu$ converges to zero considerably faster than $\beta_1(M)^{n/2} = \left|\lambda_1(P)^n\right|$.

We begin by introducing some notation from [12]. Let μ and σ be two probability distribution vectors and define the distance between them by the weighted L_2 norm

$$\left\|\mu - \sigma\right\|_w^2 = \sum_{x \in X} (\mu(x) - \sigma(x))^2 w(x) \tag{3.11}$$

where $w(x),\ x \in X$ are weights. Following [12] we select $w(x) = \dfrac{1}{\pi(x)}$ to account for variations in the stationary distribution of states. If a state x has a low stationary probability the corresponding weight is going to be large and therefore $\mu(x),\ \sigma(x)$ have to be relatively close for $\left\|\mu - \sigma\right\|_w^2$ to be small. We note that for this selection of $w(x)$, the weighted L_2 norm equals the chi-squared distance at time $n = 0$. As before let μ and ν be two initial distributions and π_n^μ and π_n^ν the corresponding distributions at time n. Then

the L_2 distance between π_n^μ and π_n^ν, the diffusion distance in the terminology of [12] , is

$$D_n^2(\mu,\nu) = \left\|\pi_n^\mu - \pi_n^\nu\right\|_w^2 = \sum_{x\in X} \frac{(\pi_n^\mu - \pi_n^\nu)^2}{\pi(x)} \tag{3.12}$$

We note that by (3.8) the distance between π_n^μ and π_n^ν in (variation norm) decays at the rate of $\beta_1(M)^n$ and consequently (due to equivalence of norms on finite dimensional space) so does $D_n^2(\mu,\nu)$. The diffusion distance can be rewritten as

$$D_n^2(\mu,\nu) = \left\|\pi_n^\mu - \pi_n^\nu\right\|_w^2 = (\pi_n^\mu - \pi_n^\nu)D^{-1}(\pi_n^\mu - \pi_n^\nu)^T \tag{3.13}$$

where, as before, D is a diagonal matrix with the entries of π on the diagonal. Using the relationship $\pi_n^\nu = \nu P^n$ we get

$$D_n^2(\mu,\nu) = (\mu-\nu)P^n D^{-1}(P^n)^T(\mu-\nu)^T = (\mu-\nu)M(P^n)D^{-1}(\mu-\nu)^T \tag{3.14}$$

Define $Q_n = D^{1/2}M(P^n)D^{-1/2}$, and note that

$$Q_n = S(M(P^n)) = D^{1/2}P^n D^{-1}(P^n)^T D^{1/2} = S(P^n)S(P^n)^T \geq 0 \tag{3.15}$$

Since $Q_n \geq 0$, it has positive real eigenvalues and a complete set of orthonormal eigenvectors. Furthermore since Q_n and $M(P^n)$ have the same eigenvalues and $M(P^n)$ is a transition matrix it follows that the eigenvalues of Q_n (and $M(P^n)$) satisfy $0 \leq \beta_{N-1}^n \leq \cdots \leq \beta_0^n = 1$. Let $v_0^n,\ldots,v_{N-1}^n$ be the eigenvectors of Q_n. Then $M(P^n)$ has right and left eigenvectors $\psi_i^n = D^{-1/2}v_i^n$ and $\varphi_i^n = D^{1/2}v_i^n, i = 0,\ldots,N-1$, respectively. We note that $\psi_0^n = [1\cdots1]^T$ and $\varphi_0^n = \pi^T$. From the spectral decomposition of Q_n we get the decomposition

$$M(P^n) = D^{-1/2} Q_n D^{1/2} = D^{-1/2} (\sum_{k=0}^{N-1} \beta_k^n v_k^n (v_k^n)^T) D^{1/2} = \sum_{k=0}^{N-1} \beta_k^n \psi_k^n (\varphi_k^n)^T \quad (3.16)$$

Therefore,

$$D_n^2(\mu, \nu) = (\mu - \nu) M(P^n) D^{-1} (\mu - \nu)^T = (\mu - \nu) \sum_{k=0}^{N-1} \beta_k^n \psi_k^n (\varphi_k^n)^T (\mu - \nu)^T \quad (3.17)$$

Define a map

$$\Psi_n : v \to \begin{bmatrix} \sqrt{\beta_0^n} v \psi_0^n \\ \vdots \\ \sqrt{\beta_{N-1}^n} v \psi_{N-1}^n \end{bmatrix} \quad (3.18)$$

i.e. $\Psi_n : \mathbf{R}^n \to \mathbf{R}^n$. Then it is easy to see that $D_n^2(\mu, \nu) = \left\| \Psi_n(\mu) - \Psi_n(\nu) \right\|^2$ where $\|\cdot\|$ is the Euclidean norm on $\mathbf{R}^n$. The above construction of Ψ_n and the representation of the diffusion distance as a Euclidean distance in the coordinates Ψ_n parallels the corresponding construction for reversible processes in [12].

Let $x^1, \ldots, x^N$ be the elements of X and let the unit vectors $e_1, \ldots, e_N$ of $\mathbf{R}^n$ denote the distributions concentrated at $x^1, \ldots, x^N$. Then, the diffusion distance between x^i and x^j can be written as

$$\begin{aligned} D_n^2(x^i, x^j) &= D_n^2(e_i^T, e_j^T) \\ &= \left\| \Psi_n(e_i^T) - \Psi_n(e_j^T) \right\|^2 \\ &= \left\| \begin{bmatrix} \sqrt{\beta_0^n} (\psi_0^n(i) - \psi_0^n(j)) \\ \vdots \\ \sqrt{\beta_{N-1}^n} (\psi_{N-1}^n(i) - \psi_{N-1}^n(j)) \end{bmatrix} \right\|^2 \end{aligned} \quad (3.19)$$

where i denotes the ith element of the right eigenvector ψ^n of $M(P^n)$.

3.6 Lower Dimensional Approximate Model

Let $q < N$ be the number of dominant eigenvalues of $M(P^n)$, i.e. we assume that $\beta_0, \ldots, \beta_{q-1}$ are of comparable size (close to one) and $\beta_q << \beta_{q-1}$. Define the approximate symmetrized model as

$$M_a(P^n) = \sum_{k=0}^{q-1} \beta_k^n \psi_k^n (\varphi_k^n)^T \tag{3.20}$$

Let $Q_{an} = \sum_{k=0}^{q-1} \beta_k^n v_k^n (v_k^n)^T$ and note that $Q_n - Q_{an} = \sum_{k=q}^{N-1} \beta_k^n v_k^n (v_k^n)^T$ and $\|Q_n - Q_{an}\| = \beta_q^n$. Furthermore, it is easy to see that $\|M(P^n) - M_a(P^n)\|$ is bounded above by β_q^n as well. Since $Q_{an} \geq 0$ there exists a matrix S_{an} such that $Q_{an} = S_{an} S_{an}{}^T$. Define $P_a^n = D^{-1/2} S_{an} D^{1/2}$. We now discuss how to select the square root S_{an} so that the operator P_a^n is a good approximation of P^n. In particular, we want to select the square root so that if $q = N$ then $P_a^n = P^n$, i.e. if there is no reduction in dimension we recover the original system.

PROPOSITION 3.2 *Let V be the matrix of orthonormal eigenvectors of Q_n, let $J_a = diag(I_q, 0)$, where I_q is the $q \times q$ identity matrix, and define $\overline{S}_{an} = V J_a V^T D^{1/2} P^n D^{-1/2}$, then the approximate operator*

$$\overline{P}_{an} = D^{-1/2} \overline{S}_{an} D^{1/2} = D^{-1/2} V J_a V^T D^{1/2} P^n = L_a P^n \tag{3.21}$$

has the property that $\overline{P}_{an} = P^n$ for $q = N$. Furthermore, L_a is a projection.

Proof: Recall that $M(P^n) = D^{-1/2} Q_n D^{1/2}$ with $Q_n \geq 0$. Furthermore, since $Q_n = S(M(P^n)) = S(P^n) S(P^n)^T \geq 0$ we see that $S(P^n) = D^{1/2} P^n D^{-1/2}$ is

one of the square roots of Q_n. Next note that we can write $Q_n = V\Sigma V^T$ where $\Sigma = diag(\beta_0, \ldots, \beta_{N-1})$ and V is orthogonal. Define $S_n = V\sqrt{\Sigma}V^T$. Then any other square root of Q_n satisfies $\bar{S}_n = S_n U$ where U is a unitary matrix. Define $\bar{P}^n = D^{-1/2}\bar{S}_n D^{1/2}$. We want to select the square root of Q_n so that $P^n = \bar{P}^n$. Therefore, consider the equation

$$P^n = D^{-1/2}\bar{S}_n D^{1/2} = D^{-1/2}S_n U D^{1/2} \tag{3.22}$$

Assuming that S_n (i.e. $M(P^n)$) is nonsingular we get

$$U = V(\sqrt{\Sigma})^{-1}V^T D^{1/2}P^n D^{-1/2} \tag{3.23}$$

Now let q be selected as above. Then we can write $Q_{an} = V\Sigma_a V^T$ and the corresponding square root is $S_{an} = V\sqrt{\Sigma_a}V^T$. Consequently,

$$\begin{aligned} \bar{S}_{an} &= S_{an}U \\ &= V\sqrt{\Sigma_a}V^T V(\sqrt{\Sigma})^{-1}V^T D^{1/2}P^n D^{-1/2} \\ &= V\sqrt{\Sigma_a}(\sqrt{\Sigma})^{-1}V^T D^{1/2}P^n D^{-1/2} \\ &= VJ_a V^T D^{1/2}P^n D^{-1/2} \end{aligned} \tag{3.24}$$

and the approximate operator

$$\bar{P}_{an} = D^{-1/2}\bar{S}_{an}D^{1/2} = D^{-1/2}VJ_a V^T D^{1/2}P^n = L_a P^n \tag{3.25}$$

We note that for $q = N$ we have for $\bar{P}_{an} = P^n$. Furthermore, $L_a{}^2 = D^{-1/2}VJ_a V^T D^{1/2}D^{-1/2}VJ_a V^T D^{1/2} = L_a$, and thus L_a is a projection. ■

THEOREM 3.2 *Consider the reduced system* $\bar{P}_{an}$ *and let* ν *be any initial distribution. Assume that* P *has a unique stationary distribution* π *with*

$\pi_i > 0,\ i = 0,\ldots,N-1$. *Then the weighted distance between the distributions* vP^n *and* $v\bar{P}_{an}$ *with weights* $w_i = \dfrac{1}{\pi_i}, i = 0,\ldots,N-1$ *satisfies*

$$\left\| vP^n - v\bar{P}_{an} \right\|_w \le \left|\lambda_q^n\right| \left\| v \right\|_w \tag{3.26}$$

where q *is the number of dominant eigenvalues of* $M(P^n)$ *(and* P^n*)* *and* λ_q *is the eigenvalue of* P *such that* $\left|\lambda_q^n\right|^2 = \beta_q^n$ *where* β_q^n *is the* $q+1$ *largest eigenvalue of* $M(P^n)$

Proof:

$$\begin{aligned}
\left\| vP^n - v\bar{P}_{an} \right\|_w^2 &= (vP^n - v\bar{P}_{an}) D^{-1} (vP^n - v\bar{P}_{an})^T \\
&= vD^{-1/2}(S_n - S_{an}) UU^T (S_n - S_{an}) D^{-1/2} v^T \\
&= vD^{-1/2} V (\Sigma - \Sigma_a) V^T D^{-1/2} v^T = \sum_{i=q}^{N-1} \beta_i^n (v\psi_i)^2 \\
&\le \sum_{i=q}^{N-1} \beta_q^n (v\psi_i)^2 \le \beta_q^n v\Psi\Psi^T v^T \\
&= \beta_q^n vD^{-1/2} VV^T D^{-1/2} v^T = \beta_q^n vD^{-1} v^T \\
&= \left|\lambda_q^n\right|^2 \left\| v \right\|_w^2
\end{aligned} \tag{3.27}$$

where Ψ is the matrix of right eigenvectors of $M(P^n)$. Here we have used the fact that if λ is an eigenvalue of P then λ^n is an eigenvalue of P^n. ■

The above result shows in particular that approximate operator $\bar{P}_{an}$ constructed above is a very good approximation of the original P at time n provided q is chosen so that $\left|\lambda_q\right|$ is small.

PROPOSITION 3.3 *Let* $e \in \mathbf{R}^n$ *be the vector of all ones and* π *be stationary distribution for* P. *The approximate matrix* $\bar{P}_{an}$ *satisfies* $\bar{P}_{an} e = e$ *and* $\pi \bar{P}_{an} = \pi$

Proof: First we note that $M(P^n)$ is a stochastic matrix with stationary distribution π. Therefore,

$$\pi M(P^n) = \pi D^{-1/2} Q_n D^{1/2} = \left[\sqrt{\pi_0}, \ldots, \sqrt{\pi_{N-1}}\right] Q_n D^{1/2} = \pi \tag{3.28}$$

and

$$\left[\sqrt{\pi_0}, \ldots, \sqrt{\pi_{N-1}}\right] Q_n = \left[\sqrt{\pi_0}, \ldots, \sqrt{\pi_{N-1}}\right] \tag{3.29}$$

i.e. $v_0^T = \left[\sqrt{\pi_0}, \ldots, \sqrt{\pi_{N-1}}\right]$ is left eigenvector of Q_n with eigenvalue 1. Furthermore, due to the symmetry of Q_n we note that v_0 is right eigenvector of Q_n as well. We also recall that Q_{an} may be written as $Q_{an} = V \Sigma_a V^T$ and note that v_0 is the first column of V. Recall that $\bar{P}_{an} = D^{-1/2} V J_a V^T D^{1/2} P^n = L_a P^n$

$$\begin{aligned} \pi \bar{P}_{an} &= \pi D^{-1/2} V J_a V^T D^{1/2} P^n \\ &= v_0^T V J_a V^T D^{1/2} P^n = \left[1, \cdots, 0\right] V^T D^{1/2} P^n \\ &= v_0^T D^{1/2} P^n = \pi P^n \\ &= \pi \end{aligned} \tag{3.30}$$

Also,

$$\begin{aligned} \bar{P}_{an} e &= D^{-1/2} V J_a V^T D^{1/2} P^n e \\ &= D^{-1/2} V J_a V^T D^{1/2} e = D^{-1/2} V J_a V^T v_0 \\ &= D^{-1/2} V J_a \left[1, \cdots, 0\right]^T = D^{-1/2} v_0 \\ &= e \end{aligned} \tag{3.31}$$

■

We note that by above proposition the approximate operator has the same stationary distribution as the original Markov chain, i.e. the limiting (stationary) behavior of the original process is retained in the model reduction process. We note also that in general the approximate operator is not Markov, i.e. the entries of $\overline{P}_{an}$ are not guaranteed to belong to $[0, 1]$. Note however that rows of $\overline{P}_{an}$ do sum to 1.

4 IDENTIFICATION OF METASTABLE COMPONENTS AND LOCAL DYNAMICS

4.1 Introduction

In Chapters 1 and 2, we stated that we need to identify three system characteristics when we model a complex dynamic system which has multi-modal behavior. These three system characteristics are: (i) the domains or regions in state space where the system spends a long time between transitions, (ii) the dynamic laws that characterize the transitive behavior between the domains and (iii) the dynamics of the system inside each domain. Most existing papers that address problems of this type model both local dynamics and the transitive behavior in a single identification step [66] [67] [68]. We note that the identification problem is further complicated if the local dynamics as well as the characteristic domains and the transitive law depend on a control input. In this dissertation we only consider autonomous case, i.e. without a control input. The real challenge in characterizing the system behavior is the identification of the

splitting of the underlying space into clustering domains and dynamic law for the transitive behavior between those domains. This problem has been studied in some considerable details in recent years for the case of linear or affine local dynamics and each domain being modeled as a polyhedral region or linear subspace [27] [28]. In this chapter, we will illustrate two methods for identifying metastable components based on the approximate operator which we discussed in Chapter 3. Then in a separate step, we identify the local dynamics for each metastable component. A couple of numerical examples will be presented for illustration and verification purpose.

4.2 Identification of Metastable Components

We next outline how to construct a metastable decomposition from $\overline{P}_{an}$ as well as the transition or modal dynamics between the metastable components. This provides the basis for the identification of the partition of the underlying space for hybrid system identification. We note that $\overline{P}_{an}$ given by (3.25) can be written as

$$\overline{P}_{an} = L_a P^n = D^{-1/2} V J_a V^T D^{1/2} P^n = \Psi J_a \Phi^T P^n = \Psi_q \Phi_q{}^T P^n \qquad (4.1)$$

where the matrices Ψ and Φ consist of the right and left eigenvectors of $M(P^n)$ and Ψ_q, Φ_q consist of the first q corresponding eigenvectors. For a given initial state v the distribution at time n is approximately $v\overline{P}_{an}$.

Obviously, $v\bar{P}_{an} \in Range(\bar{P}_{an}^T)$ for all $n > 0$, so the distribution of the original system approximately belongs to the row space of $\bar{P}_{an}$. We note that row space is spanned by the column of Φ_q. Therefore the evolution of the distribution of the approximate operator is determined by first q left eigenvctors of $M(P^n)$.

The matrix Φ_q can be used to identify the division of the original state space into the q disjoint components $A_1, \ldots, A_q$ that the system dynamics cluster on, i.e. the metastable decomposition. Obviously, the transition matrix P_q that describes the transition dynamics of the system between these components is $P_q(i, j) = \Pr(A_j \mid A_i)$. The sets A_i are identified by examining the rows of the matrix Φ_q and grouping rows that have similar structure into q equivalence classes.

In the dissertation we will introduce two ways to identify sets A_i. One way is to utilize the sign structure of rows of Φ_q to classify elements into equivalence classes. The other method to identify the set A_i is by utilizing the diffusion distance which is formulated in Chapter 3.

4.2.1 Sign Structure Identification Method

Recall that $\bar{P}_{an} = \Psi_q \Phi_q{}^T P^n$ where Ψ_q, Φ_q consist of the first q right and left eigenvectors of $M(P^n)$. We also note $\Phi_q = D^{1/2} V_q$, where $V_q = [v_0 \cdots v_{q-1}]$ consists of the first q orthogonal eigenvectors of Q_n. From

[19], we know the collection of right eigenvectors Φ can be applied to identify almost invariant aggregates in reversible nearly uncoupled Markov chains. We begin by assuming that M is a transition matrix of an uncoupled Markov chain with q disjoint clusters, i.e. M has q dominant eigenvalues which are all equal to one. By permutation, M can be put into a block-diagonal form

$$M = \begin{bmatrix} M_{11} & 0 & \cdots & 0 \\ 0 & M_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & M_{qq} \end{bmatrix} \tag{4.2}$$

where each block M_{ii} is assumed to be a reversible, aperiodic and irreducible stochastic matrix. Thus, for the eigenvalue $\beta_i = 1$ of block M_{ii}, there is a corresponding right eigenvector $(1,...,1)^{\mathrm{T}}$ and stationary distribution (left eigenvector of M_{ii}) π_i of the length $\dim(M_{ii})$. As is shown in [19], the eigenspace basis $\{\varphi_i\}_{i=1,\ldots,q}$ corresponding to $\beta_i = 1$ can be written as a linear combination of characteristic functions χ_{A_i} of the sets A_i with coefficients $\alpha_{ij} \in \mathbf{R}$ such that

$$\begin{aligned} \varphi_i &= \sum_{j=1}^{q} \alpha_{ij} \chi_{A_j}, \quad i = 1,\ldots,q \\ \chi_{A_i} &= (0,\ldots,0,e_i^T,0,\ldots,0)^T \end{aligned} \tag{4.3}$$

Therefore, in terms of the stochastic matrix M, the left eigenvector corresponding each $\beta_i = 1$ is constant on the corresponding block M_{ii}. In order to identify the clusters A_i, define a map

$$X_i \to (sign((\varphi_1)_i),\ldots,sign((\varphi_q)_i)) \tag{4.4}$$

where X_i are Markov states. It turns out that each cluster is a collection of Markov states with common sign structure and different clusters exhibit different sign structure. Indeed the following is proved in [19].

PROPOSITION 4.1 *Assume that eigenvalue* $\beta = 1$ *has multiplicity* q*, i.e., the Markov chain has* q *invariant sets. Then (i) all states that belong to the same invariant sets has same sign structure and (ii) elements from different invariant sets exhibit different sign structure.*

Consider now the case when M is not block diagonal but has off-diagonal elements that are "small". In that case, the stochastic matrix becomes block-diagonally dominant rather than block-diagonal matrix. Matrix perturbation theory [26] states stability of the eigenvectors of a matrix is determined by the eigengap ($\beta_{q-1} - \beta_q$). That is, $\varphi_1, \ldots, \varphi_q$ will be stable with respect to small changes in off diagonal elements of M if the eigengap is large. Consequently clusters can be identified by exploiting the right eigenvectors in terms of their constant sign structure. However, in some cases, the constant pattern of eigenvectors' sign structure becomes fluctuant due to the perturbations of the eigenvectors. The fluctuation may happen on some elements of the eigenvectors which are close to zero. In that case, the sign structure of some eigenvectors switches between positive and negative. Generally speaking, the pure sign identification of eigenvectors can't reliably classify the states into

different clusters correctly and efficiently unless the stochastic matrix is block diagonal or block-diagonally dominant (small perturbation) and in the general case additional measures must be taken. Referring to the algorithm proposed in [19], a simplified and effective algorithm presented as follows.

Step 1 Rescale the eigenvectors. In order to make the positive and negative elements of all eigenvectors comparable at same size, we first split each eigenvector to positive part and negative part. For example, the i th eigenvector $\varphi_i = \varphi_i^+ + \varphi_i^-$, where $\varphi_i^+ = \max(0, \varphi_i)$ and $\varphi_i^- = \min(0, \varphi_i)$. Then we rescale the positive part divided by its maximum value and rescale the negative part by divided its maximum absolute value. That is $\overline{\varphi}_i^+ = \dfrac{\varphi_i^+}{\left\|\varphi_i^+\right\|_\infty}$ and $\overline{\varphi}_i^- = \dfrac{\varphi_i^-}{\left\|\varphi_i^-\right\|_\infty}$. The rescaled matrix becomes $\overline{\Phi} = \overline{\Phi}^+ + \overline{\Phi}^-$.

Step 2 Find the q "most likely" different stable sign structure. Let $SR_i(j)$ be the entry in i th row and j th column of $\overline{\Phi}$ $i = 1, \ldots, n,\ j = 1, \ldots, q$. Then sort $SR_i(\cdot)$ descending for all rows based on $\max(\left|SR_i(j)\right|)$ for $j = 1, \ldots, q$. The sub algorithm for this step is in Table 4.1. Finally check the sign structure of SR found from the above algorithm. If the result has repeated sign structure, delete the repeated one(s) and run the algorithm until q different stable sign structure are found denoted by $SS_1, \ldots, SS_q$ where SS_i are a given row of $\overline{\Phi}$.

Step 3 Construct disjoint clusters $A_1,\ldots,A_q$. Each sign structure of $SS_1,\ldots,SS_q$ represents a cluster. This step is to allocate all other $SR_i(\cdot)$ to the closest sign structure based on the below sub algorithm in Table 4.2.

Table 4.1: Sub algorithm 1 of sign stucture method

Set $T^- = 0$ *and* $T^+ = 1$ *to perform a bisection and* $m = 0$

While $m \neq q$

Set $T = (T^- + T^+)/2$ *as threshold*

m = *count (* $\min(|SR_i(\cdot)|) \geq T$ *)*

If $m > q$ *then*

$T^- = T$

Else

$T^+ = T$

End

End

Table 4.2: Sub algorithm 2 of sign stucture method

For $i = 1$ *to* n

For $j = 1$ *to* q

Distance(i, j)= $\dfrac{\left\|\overline{\varphi}_i - SS_j\right\|}{\left\|SS_j\right\|}$

End

Find minimum of Distance (i, j) *for* $j = 1$ *to* q, *and set the minimum index* $k \in \{1,...,q\}$ *assign* $\overline{\varphi}_i$ *to cluster* A_k

End

Summarizing, the sign structure identification method may assign elements even if which may have different pure signs into a cluster. On the other hand, it's possible that some elements may not be allocated correctly if off-diagonal elements of the M matrix are not "small" and spoil the sign structure.

4.2.2 Diffusion Distance Identification Method

Recall that in the approximate and reduced model, the diffusion distance of any two points x^i and x^j can be written as

$$D_q^2(x^i, x^j) = \left\| \begin{array}{c} \sqrt{\beta_0^n}\,(\psi_0^n(i) - \psi_0^n(j)) \\ \vdots \\ \sqrt{\beta_{q-1}^n}\,(\psi_{q-1}^n(i) - \psi_{q-1}^n(j)) \end{array} \right\|^2 \tag{4.5}$$

If $D_q^2(x^i, x^j)$ is smaller than some threshold δ, then x^i and x^j are considered to belong to the same set. Note that the diffusion distance and sign structure are similar in that both defined proximity of points in terms of the closeness of the eigenvectors of $M(P^n)$. The main challenge for this identification method is how to select threshold δ efficiently if the constructed matrix $M(P^n)$ is not block-diagonal form but block-diagonally dominant and the $N \times N$ matrix $D_q^2 = (D_q^2(x^i, x^j))_{i,j=1,\ldots,N}$ has large dimension. This challenge actually becomes the question of how to cluster the elements of D_q^2 if the diffusion distance is calculated for all data point pairs. Past researches [69]

[70] [71] show that the K-nearest neighbor (KNN) is one of the most fundamental and simple clustering method when there is little or no prior knowledge about the distribution of the data. The principle difference between diffusion distance method and traditional KNN clustering is that diffusion distance method maps (diffusion map) the original data points into a Euclidean space and obtain a new description of data sets based on spectrum of the Markov transition matrix. Then diffusion distance method classifies points in terms of their connectivity in the new Euclidean space. The traditional KNN clustering method considers those points close or not in terms of the Euclidean distance or Hamming distance in original space. In addition, since we don't have any prior training points, there is no any prior label for each metastable component. Thus we propose a modified KNN algorithm applied to diffusion distance method which is presented as follows.

Step 1 Formulate diffusion distance matrix. In order to compare all the points, we first calculate diffusion distance for all two point pairs and formulate a diffusion distance matrix $D = D_q^2$, where D_{ij} represents $D_q^2(x^i, x^j)$. Then we set the diagonal entries $D_{ii} = \infty$.

Step 2 Order and group diffusion distances. In this step we first create a matrix G, where $G \in R^{N/2 \times N}$ (for simplicity assuming N is even) and initially

set all entries to zero. The main idea of this method is to find the minimal diffusion distance, record the indices of the points and put the indices as a group

Table 4.3: Sub algorithm of diffusion distance method

For $k = 1$ *to* $N - 1$

Find the minimal diffusion distance in D_{ij} *and record indices as* r, c.

If $k == 1$

Create G *matrix with all zero entries, and assign* r, c *to first row.*

Else

If r, c *does not belong a group in* G

Add r, c *to a new group in* G

Else if r *doesn't belong to a group but* c *does*

Add r *to the group* c *exists in*

Else if c *doesn't belong to a group but* r *does*

Add c *to the group* r *exists in*

Else both r, c *belong to a group in* G

Add the group c *exists in to the group* r *exists in*

End

End

Set D_{rc} *and* D_{cr} *as infinity.*

End

in matrix G. Then next find the minimal diffusion distance, and assign the indices to one of the existing groups or a new group created in G. The sub algorithm for this step is in Table 4.3.

Step 3 Clustering Ordered diffusion distances into q groups. Since all the points are ordered based on the algorithm of the last step, we just need to find the q maximal ones in terms of diffusion distance corresponding to the ordered points, of which the indices are the segment lines to clustering q groups.

4.2.3 Identification of Transition Dynamics

Given the Markov state data in each metastable set which are identified in last subsection, we can identify the transition dynamics between the metastable sets.

PROPOSITION 4.2 *Given $x^1,\ldots,x^N$ on X and a cluster A_i, define the characteristic vector* $e_{A_i}=\begin{bmatrix}\chi_{A_i}(x^1)\\ \vdots \\ \chi_{A_i}(x^N)\end{bmatrix}$ *where* $\chi_{A_i}(x^j)=1$ *for* $x^j\in A_i$ *and* $\chi_{A_i}(x^j)=0$ *otherwise. If we have q disjoint clusters $A_1,\ldots,A_q$, formulate a matrix* $E=[e_{A_1}\ \cdots\ e_{A_q}]$ *then the transition matrix characterizing the transition dynamics between $A_1,\ldots,A_q$ is P_q with entries*

$$P_q(i,j)=\frac{(E^TP^nD^{-1}E)(i,j)}{\sum_{j=1}^{q}(E^TP^nD^{-1}E)(i,j)} \tag{4.6}$$

Proof: Referring to the definition of the (conditional) transition probability from cluster A_i to cluster A_j, i.e. (2.5) and (2.8) with respect to the stationary distribution π in [19], (4.6) is easily proved. ■

4.3 Nyström Extension Method

The discretization of the transfer operator P is subject to the curse of dimensionality, i.e. as the number of discrete components increases the dimension of discretized operator grows exponentially. In particular, if the process belongs to a n dimensional space and is uniformly discretized to size m then the number of discretization components, i.e. Markov states, is $N = m^n$ and P_N is an $N \times N$ matrix. In addition, due to the characteristics of transfer operator formulation of system dynamics, the P_N matrix is highly sparse as well. In Chapter 3, we obtained the reduced order approximate model for the Markov chain with transition matrix P using spectral decomposition methods. In order to alleviate the computation burden of eigenvalues and eigenvectors of such large-size and highly-sparse matrix, we present an approximation technique which is based on the Nyström extension method. In short the technique finds eigenvalues and eigenvectors of some small sub matrix of the sparse matrix and then approximates the eigenvectors of the original matrix employing Nyström extension.

Consider the matrix $Q \in \mathbf{R}^{N \times N}$, and let M columns and M rows be chosen (without replacement) in some manner. Then Q can be partitioned as

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \tag{4.7}$$

Where $Q_{11} \in \mathbf{R}^{M \times M}$ represents the subblock of matrix entries are corresponding to the selected columns and selected rows (maybe after reordering). Q_{12} is matrix consisting of entries with a non-selected column label and selected row label and Q_{21} consists of entries with a selected column label and non-selected row label, respectively. $Q_{22} \in \mathbf{R}^{(N-M)\times(N-M)}$ is matrix consisting of the remaining entries.

The Q matrix in our case is symmetric and positive semidefinite, and Q_{11} is as well. Therefore, its singular value decomposition (SVD) may be written as $Q_{11} = U\Sigma U^T$ and we know that $Q_{12} = Q_{21}^T$. From [29] [30], the Nyström extension of U gives the approximation of the eigenvectors of the full matrix Q as

$$\overline{U} = \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} U\Sigma^{-1} = \begin{bmatrix} U \\ Q_{21}U\Sigma^{-1} \end{bmatrix} \tag{4.8}$$

Then the approximation of Q may be written as

$$
\begin{aligned}
\overline{Q} &= \overline{U}\Sigma\overline{U}^T \\
&= \begin{bmatrix} U \\ Q_{21}U\Sigma^{-1} \end{bmatrix} \Sigma \begin{bmatrix} U \\ Q_{21}U\Sigma^{-1} \end{bmatrix}^T \\
&= \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} Q_{11}^{-1} \begin{bmatrix} Q_{11} & Q_{12} \end{bmatrix} \\
&= \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{21}Q_{11}^{-1}Q_{12} \end{bmatrix}
\end{aligned}
\tag{4.9}
$$

More generally, if Q is an arbitrary $m \times n$ matrix, then $Q_{11} = U\Sigma V^T$ and the Nyström extension U and V gives the following approximation for the left and right eigenvectors of SVD of the full matrix Q as

$$
\overline{U} = \begin{bmatrix} U \\ Q_{21}V\Sigma^{-1} \end{bmatrix}, \text{ and}
$$

$$
\overline{V} = \begin{bmatrix} V \\ Q_{12}^T U\Sigma^{-1} \end{bmatrix} \tag{4.10}
$$

Then

$$
\overline{Q} = \overline{U}\Sigma\overline{V}^T = \begin{bmatrix} U \\ Q_{21}V\Sigma^{-1} \end{bmatrix} \Sigma \begin{bmatrix} V \\ Q_{12}^T U\Sigma^{-1} \end{bmatrix} = \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} Q_{11}^{+} \begin{bmatrix} Q_{11} & Q_{12} \end{bmatrix} \tag{4.11}
$$

If Q_{11} is nonsingular, then the equation becomes

$$
\overline{Q} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{21}Q_{11}^{-1}Q_{12} \end{bmatrix} \tag{4.12}
$$

Thus, in the Nyström method the block Q_{22} is approximated by $Q_{21}Q_{11}^{-1}Q_{12}$ and Q is approximated by $\overline{Q}$. The approximation quality can be quantified by the norm of Schur complement $\left\| Q_{22} - Q_{21}Q_{11}^{-1}Q_{12} \right\|$.

The principal question is how to choose the M columns of interest. In [30], M columns of Q matrix are uniformly picked as sampled columns. In [31]

[34] [35], the authors utilized the important sampling of columns and/or rows of the matrix with carefully chosen nonuniform probability distribution in order to obtain provable error bounds. Furthermore in [31] [34] [35], the discrete operator P is obtained from data of a discrete time Markov process as in (2.9). This leads a stochastic matrix whose elements are not uniformly distributed nor are the columns uniformly distributed. Referring to the main approximation algorithm in [36], we propose a Nyström extension algorithm with nonuinformly picked columns and rows to approximate eigenvectors.

Step 1 Formulate permutation matrix. As we stated in Chapter 2, in order to approximate transfer operator P we first create an $N \times N$ matrix C with entries c_{ij} that count the number of times there is a transition from element X_i to element X_j, i.e. starting with $c_{ij} = 0$ we increase c_{ij} by one if $x_k \in X_i$ and $x_{k+1} \in X_j$ for some k is defined. In matrix C, the number of times a state in X_j is entered is defined by $\sum_{i=1}^{N} c_{ij}, j = 1, \ldots, N$. We select the M sampled columns as the rows that have maximal sum values in matrix C. This selection makes sure that elements of the dynamic systems that have high probability are selected. The sub algorithm for this step is in Table 4.4.

Table 4.4: Sub algorithm of Nyström extension method

Create C matrix and permutation matrix S with zero entries
Sort $sum(C^T)$ descend and record indices in matrix I
For $i = 1$ to N
$S(i, I(i)) = 1$
End

Step 2 Decompose Q matrix. Firstly Q is permutated as $Q = SQS^T$ and then Q is decomposed as form in (4.7).

Step 3 Calculate eigenvalues and eigenvectors of Q. LAPACK routine [37] is used to calculate the eigenvalues and eigenvectors of matrix Q_{11}. Then calculate $\overline{U}$, $\overline{V}$ according to (4.8).

The amount of computation required for finding the eigenvectors of matrix Q is $O(N^3)$. By applying Nyström extension method to find the approximate eigenvectors, the computational burden decreases to $O(M^2 N)$ operations.

4.4 Identification of Local Dynamics

Once the metastable components A_i have been identified, the original data can be segmented and classified into each component. Using the data in each component, a model of the local dynamics can be constructed. We consider

models in a state space form and assume that the system can be represented in each clustering component by an affine model of the form

$$x_{k+1} = A(r_k)x_k + f(r_k) + w_k \tag{4.13}$$

This implies in particular that locally in each component the dynamics have a stable point attractor and in vicinity of that point the dynamics are approximately linear. We choose subspace identification methods (SIMs) to identify the system parameter matrix $A(i)$, the bias term $f(i)$ and the statistics of w_k. The noise input sequence w_k is assumed to be a stationary, zero mean, white noise process with covariance $E(w_i w_i^T) = R(r_k)$ as we stated in Section 2.3.5. Under certain assumptions subspace identification introduced in Section 2.3.5 can successfully estimate the system order and parameter matrices. Furthermore, we use the identification method to get the noise sequence and then calculate covariance matrix R. We obtain a covariance matrix $R(r_k)$ for each clustering component and an overall resulting model of the form

$$x_{k+1} = A(r_k)x_k + f(r_k) + \sqrt{R(r_k)}e_k \tag{4.17}$$

where e_k is standard Gaussian white noise process $N(0, I)$.

We remark that if the local dynamics in some component A_i do not have a point attractor, then the above approach doesn't work and nonlinear identification techniques are required.

4.5 Numerical Examples

4.5.1 Example 1

Consider a Markov chain with $N = 12$ that is constructed as follows. Let $P_1 \in \mathbf{R}^{12\times 12}$ be a block diagonal matrix with three blocks with positive entries, $P_2 \in \mathbf{R}^{12\times 12}$ be a matrix with positive entries, ε is a small parameter and define $\hat{P} = (1-\varepsilon)P_1 + \varepsilon P_2$. We select the entries P_1 and P_2 randomly from a uniform distribution on $[0, 1]$. Define the normalized matrix with entries $p_{ij} = \dfrac{\hat{p}_{ij}}{\sum_{j=1}^{N} \hat{p}_{ij}}$. Then P is a transition matrix of a Markov chain. We assume that P so chosen is a irreducible and aperiodic. For a particular choice of P with $\varepsilon = 0.01$ we obtain multiplicative reversibilization $M(P)$ with eigenvlaues

[1, 0.9391, 0.9141,0.2971,0.0895,0.0846,0.0475,0.0441,0.0376,0.0313,0.0008,0]

Clearly there is a big gap in the spectrum after the first three eigenvalues and thus we obtain $q = 3$ (as expected). Furthermore,

$$\Phi_3 = \begin{bmatrix} 0.0955 & 0.0909 & -0.0203 \\ 0.0912 & 0.0876 & -0.0192 \\ 0.0703 & 0.0679 & -0.0148 \\ 0.1249 & 0.1194 & -0.0266 \\ 0.1289 & 0.1222 & -0.0275 \\ 0.0993 & -0.0625 & 0.1978 \\ 0.0502 & -0.0305 & 0.0967 \\ 0.0431 & -0.0251 & 0.0819 \\ 0.0458 & -0.0581 & -0.0421 \\ 0.0587 & -0.0689 & -0.0498 \\ 0.0971 & -0.1224 & -0.0886 \\ 0.0950 & -0.1204 & -0.0873 \end{bmatrix}$$

and

$$\Psi_3 = \begin{bmatrix} 1 & 0.9517 & -0.2128 \\ 1 & 0.9609 & -0.2111 \\ 1 & 0.9647 & -0.2111 \\ 1 & 0.9559 & -0.2133 \\ 1 & 0.9480 & -0.2133 \\ 1 & -0.6288 & 1.9907 \\ 1 & -0.6069 & 1.9236 \\ 1 & -0.5807 & 1.8974 \\ 1 & -1.2688 & -0.9193 \\ 1 & -1.1747 & -0.8490 \\ 1 & -1.2613 & -0.9125 \\ 1 & -1.2670 & -0.9185 \end{bmatrix}$$

First we apply sign structure method to classify different groups. From Φ_3, we note there are three different sign structure ([+ + -], [+ - +] and [+ - -]) thus we can group the states into three classes, i.e. 1-5, 6-8 and 9-12.

For comparison we apply the diffusion distance method. We calculate diffusion distance matrix D with entries $D_3^2(x^i, x^j)$ from Ψ_3 as

$$D=\begin{bmatrix}
0 & .0001 & .0002 & 0 & 0 & 6.7846 & 6.4537 & 6.2759 & 5.0868 & 4.6165 & 5.0468 & 5.0785 \\
.0001 & 0 & 0 & 0 & 0.0002 & 6.8050 & 6.4740 & 6.2958 & 5.1272 & 4.6551 & 5.0870 & 5.1189 \\
.0002 & 0 & 0 & 0.0001 & 0.0003 & 6.8167 & 6.4854 & 6.3070 & 5.1435 & 4.6707 & 5.1033 & 5.1351 \\
0 & 0 & .0001 & 0 & 0.0001 & 6.7989 & 6.4678 & 6.2897 & 5.1035 & 4.6326 & 5.0635 & 5.0952 \\
0 & .0002 & .0003 & 0.0001 & 0 & 6.7755 & 6.4447 & 6.2670 & 5.0708 & 4.6013 & 5.0309 & 5.0626 \\
6.7846 & 6.8050 & 6.8167 & 6.7989 & 6.7755 & 0 & 0.0046 & 0.0101 & 8.1257 & 7.6513 & 8.0804 & 8.1194 \\
6.4537 & 6.4740 & 6.4854 & 6.4678 & 6.4447 & 0.0046 & 0 & 0.0013 & 7.7995 & 7.3298 & 7.7547 & 7.7933 \\
6.2759 & 6.2958 & 6.3070 & 6.2897 & 6.2670 & 0.0101 & 0.0013 & 0 & 7.6974 & 7.2264 & 7.6525 & 7.6911 \\
5.0868 & 5.1272 & 5.1435 & 5.1035 & 5.0708 & 8.1257 & 7.7995 & 7.6974 & 0 & 0.0128 & 0.0001 & 0 \\
4.6165 & 4.6551 & 4.6707 & 4.6326 & 4.6013 & 7.6513 & 7.3298 & 7.2264 & 0.0128 & 0 & 0.0107 & 0.0124 \\
5.0468 & 5.0870 & 5.1033 & 5.0635 & 5.0309 & 8.0804 & 7.7547 & 7.6525 & 0.0001 & 0.0107 & 0 & 0.0001 \\
5.0785 & 5.1189 & 5.1351 & 5.0952 & 5.0626 & 8.1194 & 7.7933 & 7.6911 & 0 & 0.0124 & 0.0001 & 0
\end{bmatrix}$$

Then we apply the K-nearest neighbor (KNN) algorithm proposed in Section 4.2.2 to group the diffusion distance matrix. The formulated G matrix evolves as

$$G=\begin{bmatrix}
9 & 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

$$\vdots$$

$$G=\begin{bmatrix}
9 & 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 2 & 4 & 1 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

$$G=\begin{bmatrix}
9 & 12 & 11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 2 & 4 & 1 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

$$\vdots$$

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6 & 7 & 8 & 9 & 12 & 11 & 10 & 3 & 2 & 4 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So based on the forth row of matrix G , the ordered diffusion distance is computed as

$$\begin{bmatrix} D_{67} & D_{78} & D_{89} & D_{912} & D_{1211} & D_{1110} & D_{103} & D_{32} & D_{24} & D_{41} & D_{15} & D_{56} \end{bmatrix}$$

$$= \begin{bmatrix} 0.0046 & 0.0013 & 7.6974 & 0 & 0.0001 & 0.0107 & 4.6707 & 0 & 0 & 0 & 0 & 6.7755 \end{bmatrix}$$

It's obviously the states group into three classes which are same the one obtained using the sign structure method. Also color map of the grouped diffusion distance matrix is shown in Fig.4-1.

As expected, the resulting division of the state space into three subsets agrees with the dimensions of the diagonal blocks in P_1. Furthermore, the reduced system has transition matrix

$$P_3 = \begin{bmatrix} 0.9623 & 0.0227 & 0.0150 \\ 0.0065 & 0.9847 & 0.0089 \\ 0.0124 & 0.0238 & 0.9639 \end{bmatrix}$$

We note that reduced transition matrix characterizes the transition between the three subsets of states defined by the equivalence structure.

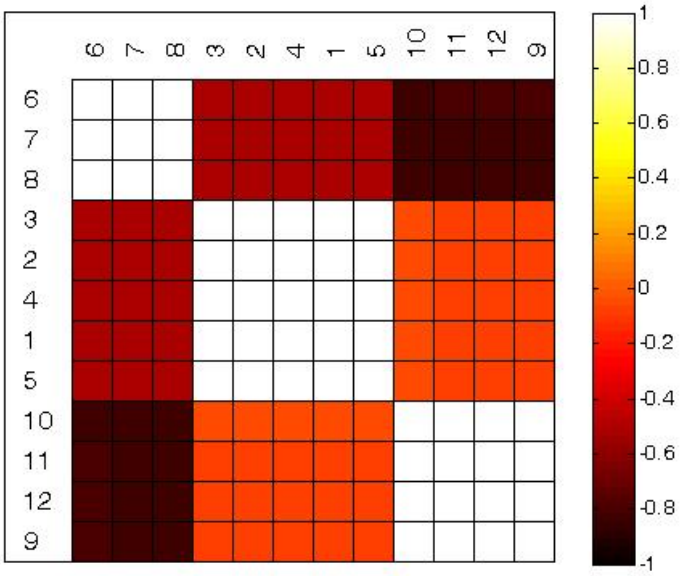

Fig. 4-1 Color map of grouped diffusion distance matrix ($\varepsilon = 0.01$)

If $\varepsilon = 0.1$, we obtain multiplicative reversibilization $M(P)$ with eigenvlaues $[1, 0.6052, 0.5090, 0.1394, 0.0923, \ldots, 0.005]$. Clearly there is also a big gap in the spectrum after the first three eigenvalues and thus we still obtain $q = 3$. Transition matrix of the reduced system becomes

$$P_3 = \begin{bmatrix} 0.8292 & 0.0745 & 0.0962 \\ 0.1171 & 0.7850 & 0.0978 \\ 0.0800 & 0.1158 & 0.8041 \end{bmatrix}$$

Using sign structure method for Φ_3, we can still get three classes (1-5, 6-8 and 9-12). For diffusion distance method, we get similar results as the sign structure method. Color map of the grouped diffusion distance matrix is shown in Fig. 4-2.

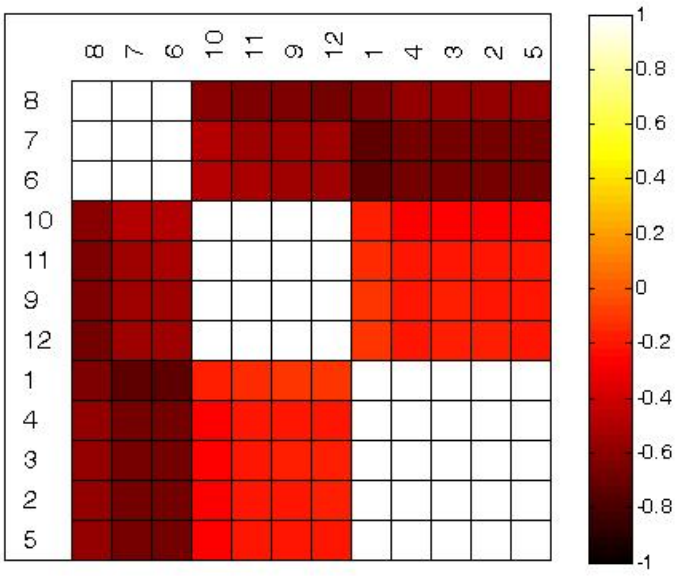

Fig. 4-2 Color map of grouped diffusion distance matrix ($\varepsilon = 0.1$)

4.5.2 Example 2

Consider a nonlinear discrete time stochastic dynamical system on $\mathbf{R}^2$ described by the equations

$$x(t+1) = f(x(t)) + \varepsilon w(t)$$

where $x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, w(t) = \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}$ is a sequence of i.i.d standard Gaussian random variables and

$$f(x) = \begin{bmatrix} x_1 + \delta t x_2 \\ x_2 + \delta t(-x_2 + \alpha(\beta x_1 - x_1^3)) \end{bmatrix}$$

The coefficients are chosen to be $\alpha = \beta = 1, \delta t = 0.2$ and $\varepsilon = 0.1$. The system has three equilibrium points, an unstable one at the origin and stable equilibrium

at $\begin{bmatrix} 0 & \pm\sqrt{\beta} \end{bmatrix}^T$. The dynamics of the unperturbed nonlinear system is shown in Fig. 4-3 for two different initial conditions. A simulation of the stochastic system for a typical initial condition is shown in Fig. 4-4. A uniform grid of size is 30×30 placed over the image of the process and a 900 state Markov chain generated from the discretized process defined by the system dynamics. The reversibilization procedure is performed and the first few eigenvalues of $M(P^n)$ found to be

$$\begin{bmatrix} 1.0 & 0.9934 & 0.9288 & 0.9240 & 0.9136 & \cdots \end{bmatrix} \text{ for } n = 1.$$

The first gap in the spectrum is found to be between the 2nd and 3rd eigenvalues (we note that as n increases the first gap becomes more predominant). The model reduction procedure is performed for $q = 2$ and the resulting reduced 2×2 transition matrix is

$$P_2 = \begin{bmatrix} 0.9850 & 0.0150 \\ 0.0118 & 0.9882 \end{bmatrix}$$

The matrix Ψ_2 is examined and the equivalence structure shown in Fig. 4-5 constructed by utilizing the diffusion distance method. In Fig. 4-5 we show the division of the image into the two metastable sets (shown in blue and red). Finally, in Fig. 4-6 we have overlaid the system trajectories from Fig. 4-4 onto the image Fig. 4-5. If we utilize the sign structure to identify the metastable components, we get similar decomposition as shown on Fig. 4-7 and Fig. 4-8. We remark that, as expected, the two metastable sets closely resemble the domains of attraction of

the original (unperturbed) dynamical system. In order to identify the local dynamics of each set, we classify the original data into segments belonging to the metastable components and utilize this data to identify local dynamic models of the form (4-12) on each component. A simulation of the resulting hybrid affine system is shown in Fig. 4-9 which is overaid on the original data.

Finally we apply the Nyström extension algorithm proposed in Section 4.3 to analyze and identify the system dynamics. Fig. 4-10 shows the ordered numbers which are sum values of each column in matrix C. From the figure, we may choose first 10 or 100 sampled columns to approximate matrix Q. For $M = 10$ the reversibilization procedure is similarly performed and the first few eigenvalues of $M(P^n)$ which are same as those of Q_{11} found to be

$$\begin{bmatrix} 1.0 & 0.9757 & 0.3312 & 0.3255 & 0.1741 & \cdots \end{bmatrix}$$

The first gap in the spectrum is between the 2nd and 3rd eigenvalues as expected. Φ_2 and Ψ_2 are approximated by Nyström extension algorithm. The matrix Ψ_2 is examined and the equivalence structure shown in Fig. 4-11 constructed by utilizing the diffusion distance method. The model reduction procedure is performed for the two disjoint metastable sets and the resulting reduced 2×2 transition matrix is

$$P_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

P_2 shows there is no transition dynamics between the two attraction sets. Thus, a simulation of the resulting hybrid affine system for one initial condition will stay in one set forever which is show in Fig. 4-12. In particular, applying Nyström extension overestimates the diagonal entries of P_2 and using $M = 10$ is too coarse in capturing the transition dynamics.

For $M = 100$ the reversibilization procedure is similarly performed and the first few eigenvalues of $M(P^n)$ which are same as those of Q_{11} found to be

$$\begin{bmatrix} 1.0 & 0.9901 & 0.8342 & 0.8115 & 0.7308 & \cdots \end{bmatrix}$$

The first gap in the spectrum is between the 2nd and 3rd eigenvalues as expected. Fig.4-13 and Fig.4-14 show the similar shape of eignenvalue between original and approximate eigenvalues. Φ_2 and Ψ_2 are approximated by Nyström extension algorithm. The matrix Ψ_2 is examined and the equivalence structure shown in Fig. 4-15 constructed by utilizing the diffusion distance method. The model reduction procedure is performed for the two metastable sets and the resulting reduced 2×2 transition matrix is

$$P_2 = \begin{bmatrix} 0.9904 & 0.0096 \\ 0.0082 & 0.9918 \end{bmatrix}.$$

A simulation of the resulting hybrid affine system is shown in Fig. 4-16 which is overaid on the original data.

From above simulation results, choosing $M = 10$ or less in Nyström extension algorithm is enough to capture the two mestastable sets. We note

however that the Nyström extension results in transition dynamics have considerable large diagonal entries and thus there is some error in applying the Nyström extension to the identification of transitive probabilistics. Thus, in order to capture the transition dynamics, we have to choose M around 100. The norm values of Schur complement $\left\| Q_{22} - Q_{21} Q_{11}^{-1} Q_{12} \right\|$ are calculated on condition of $M = 1$ to $M = 450$, which are shown in Fig. 4-17. From Fig. 4-17, we note that choosing M around 100 is good approximation.

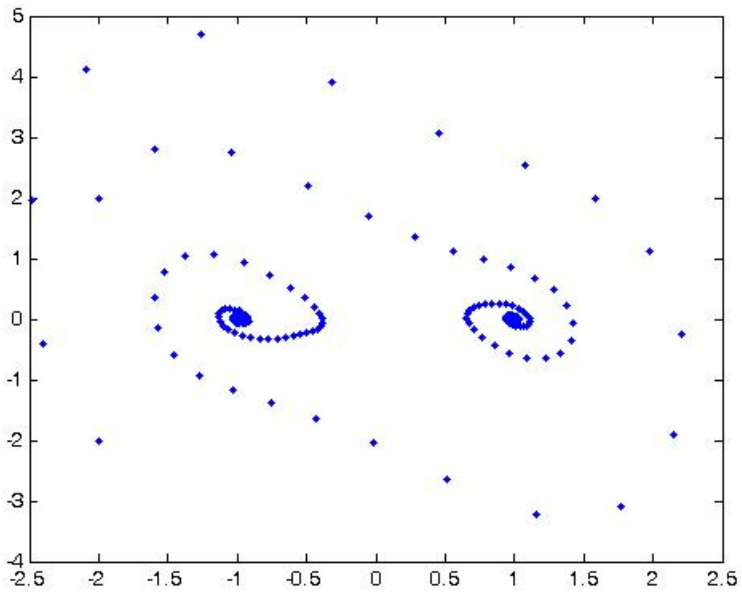

Fig. 4-3 Unperturbed system

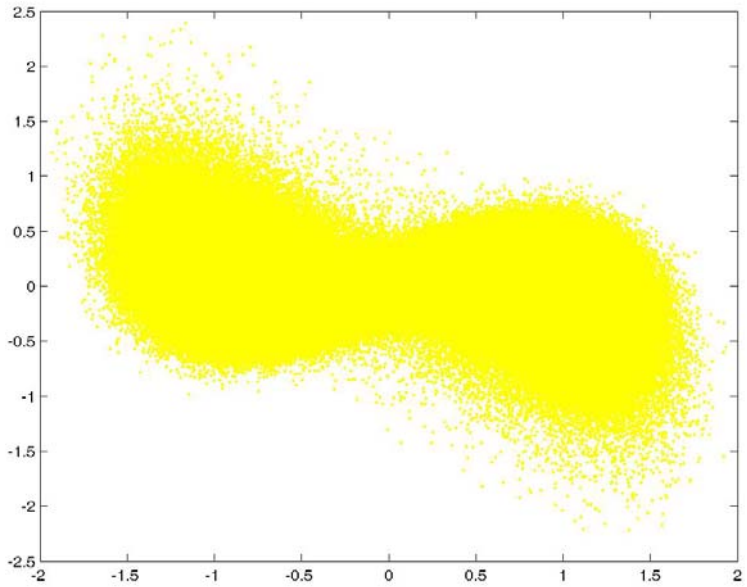

Fig. 4-4 Original trajectories

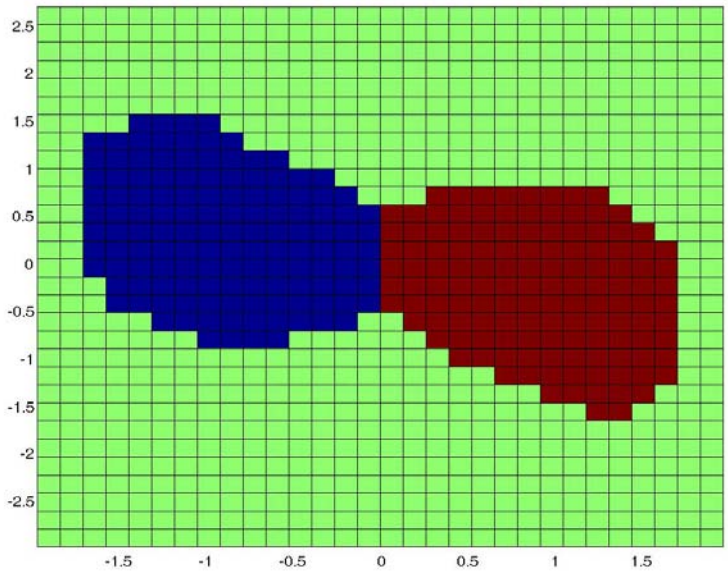

Fig. 4-5 Identification of two invariant sets

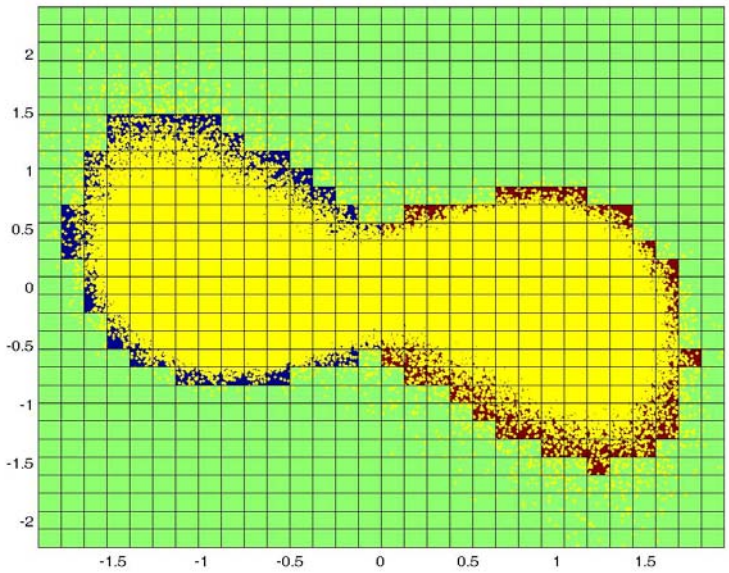

Fig. 4-6 Invariant set grid overlaid with original trajectories

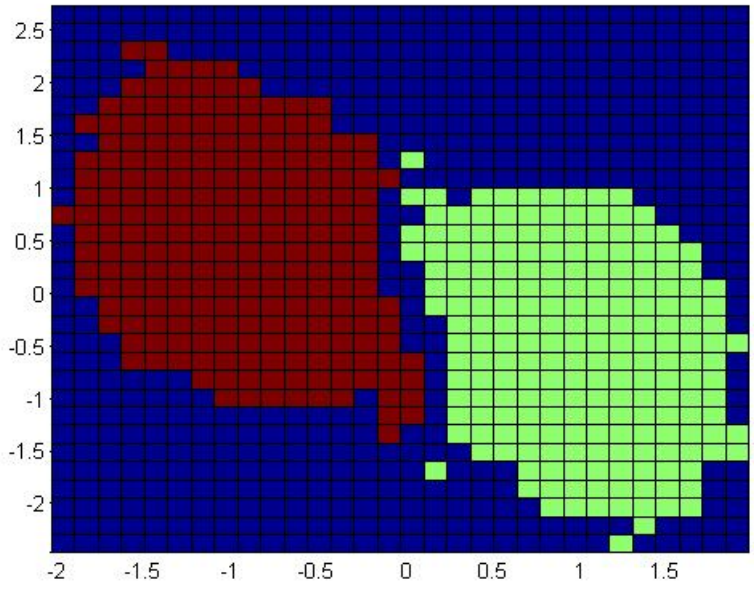

Fig. 4-7 Identification of two invariant sets by sign structure

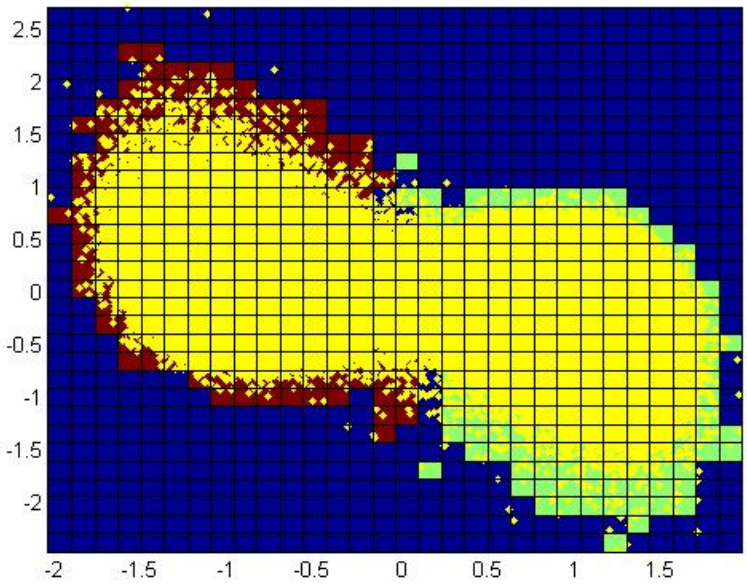

Fig. 4-8 Invariant set grid overlaid with original trajectories by sign structure

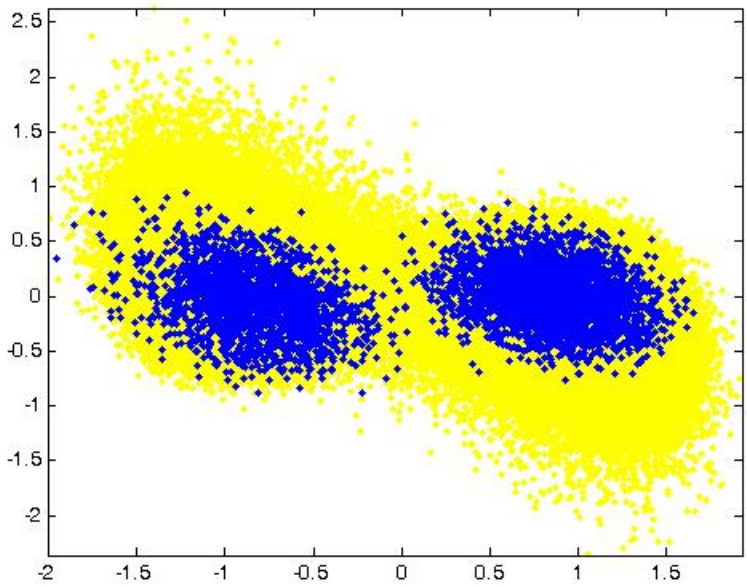

Fig. 4-9 Estimated model trajectories overlaid with original trajectories

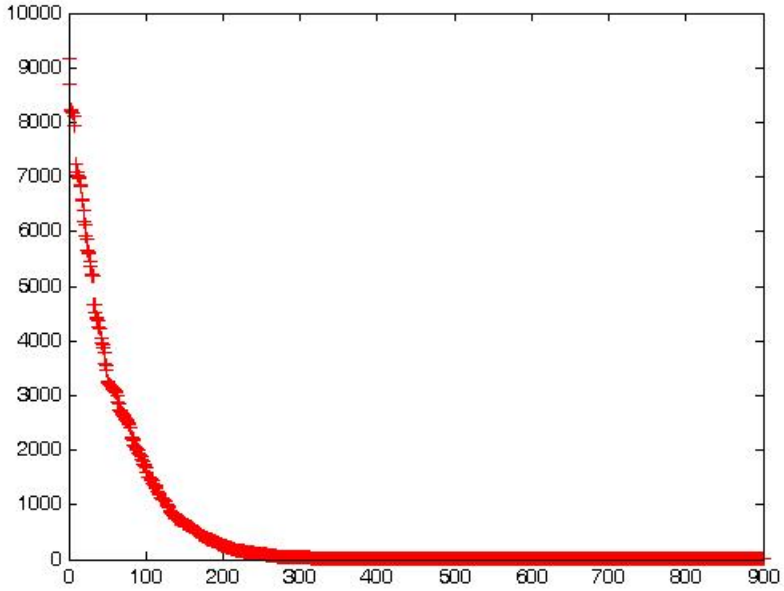

Fig. 4-10 Ordered sum value of each column

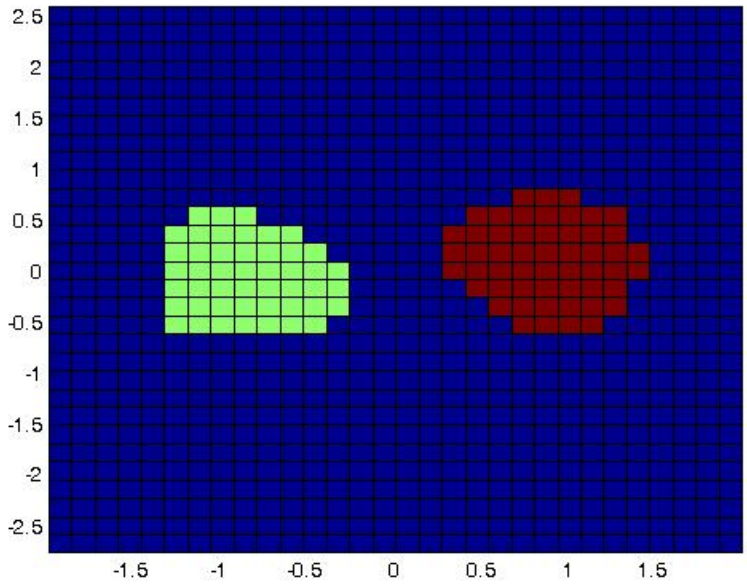

Fig. 4-11 Identification of two invariant sets (M=10)

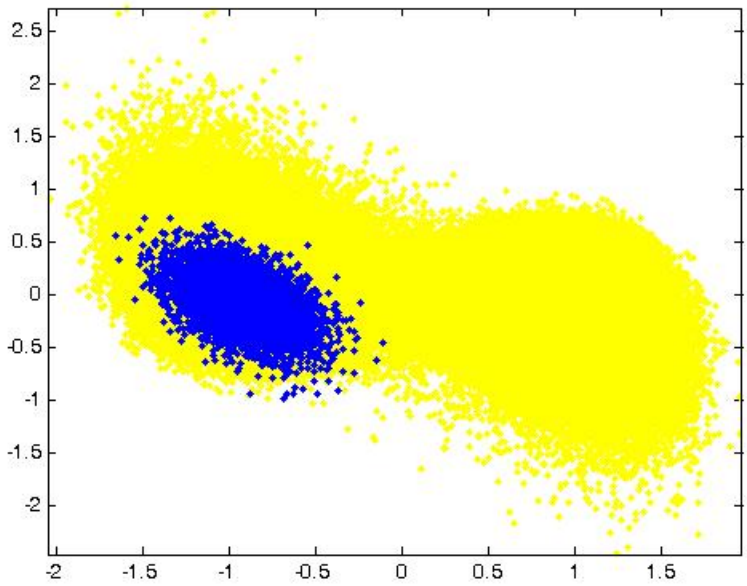

Fig. 4-12 Estimated model trajectories overlaid with original trajectories (M=10)

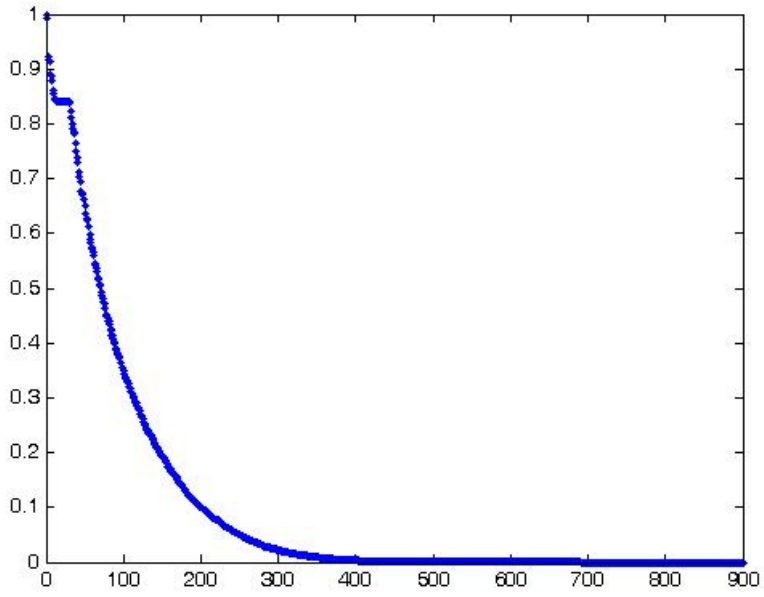

Fig. 4-13 Eigenvalues of the original M(P)

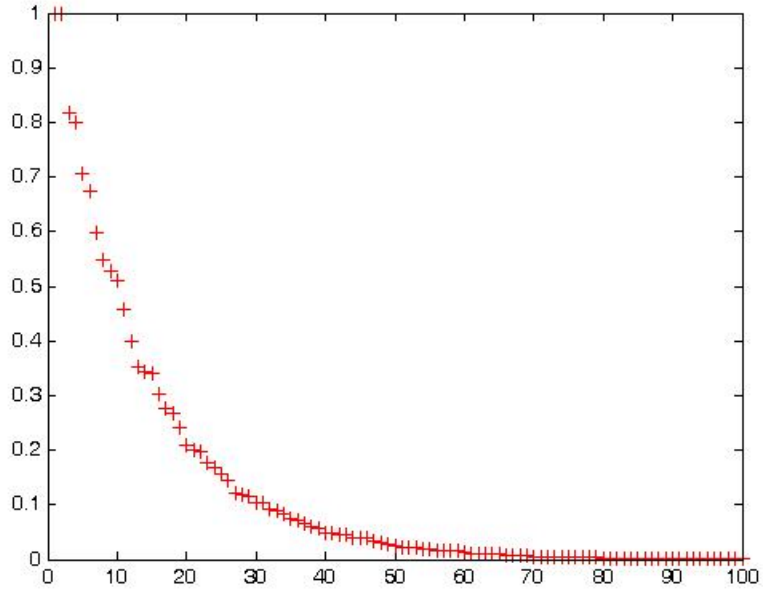

Fig. 4-14 Eigenvalues of the approximate M(P) (M=100)

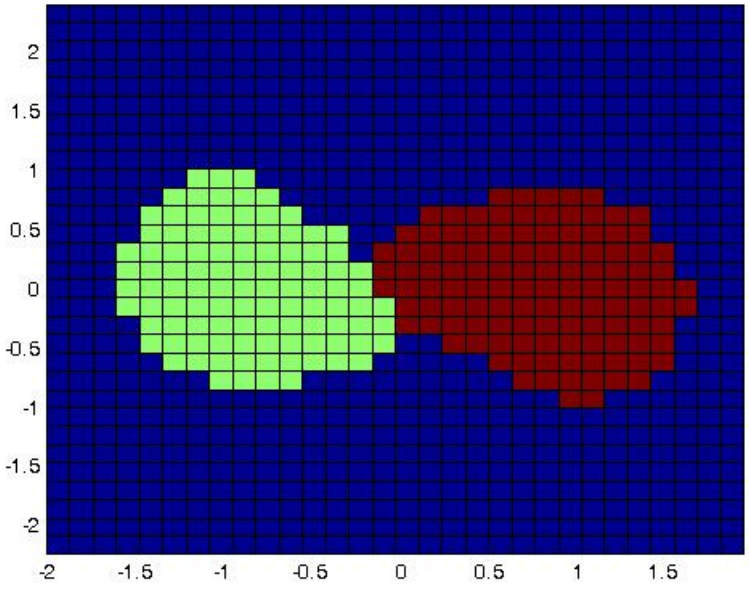

Fig. 4-15 Identification of two invariant sets (M=100)

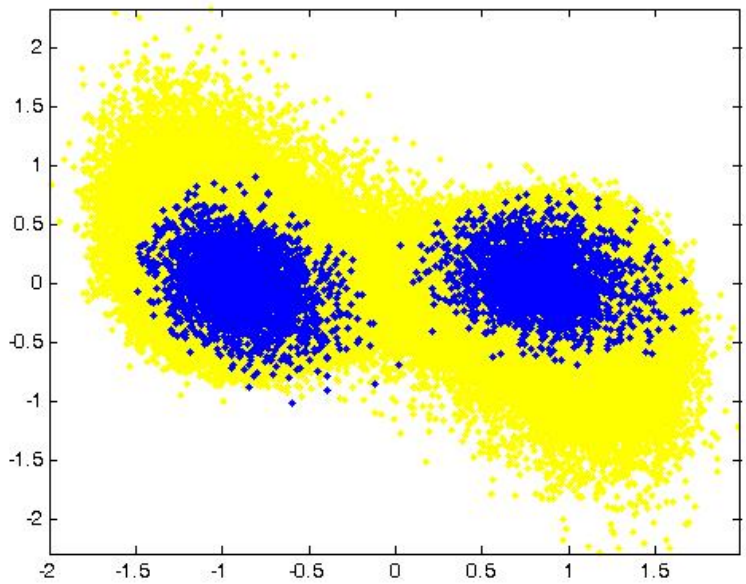

Fig. 4-16 Estimated model trajectories overlaid with original trajectories (M=100)

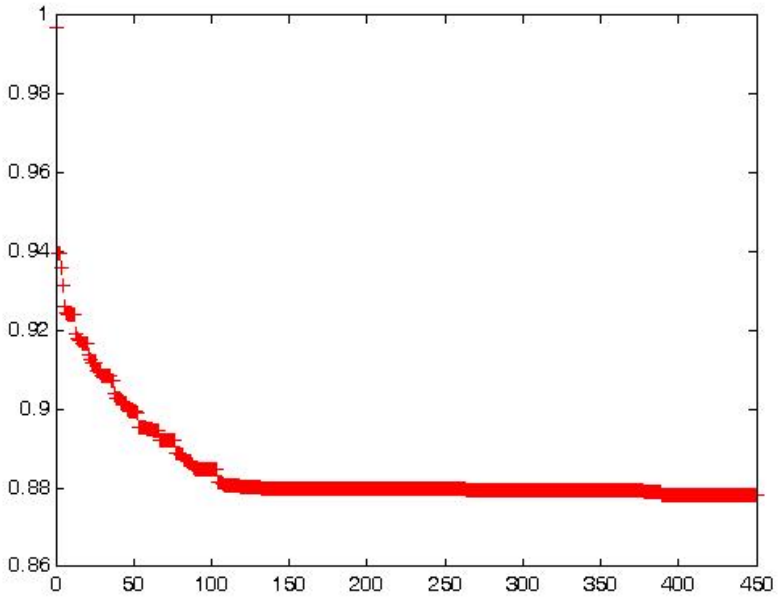

Fig. 4-17 Norm of of Schur complement (M=1 to 450)

5 CLUSTER ANALYSIS OF WIND TURBINES OF LARGE WIND FARM

5.1 Introduction

The modern wind power industry has its origins in the 1980s, but in the last decade it has become the fastest growing energy industry in the world. In 2007, the global wind power industry installed over 20,000 MW of wind turbine capacity which was an increase of 31% compared to in 2006 [38]. The rapid growth was lead by the US, Europe and China. In the U.S., 5,244 MW of installed power was reported which made the overall wind power capacity grow by 45% in 2007 [38]. The American Wind Energy Association (AWEA) expects that 20% of the overall power production of the U.S. will be wind by 2030, with persistent aid of federal Production Tax Credit (PTC) and state renewable portfolio standards (RPS) [38]. This means that a large number of large-scale wind power farms will be built.

The problem of how to model and control the power output of a wind farm

is becoming a pressing issue that challenges power system reliability. Past experience and lessons learned indicate that the variability of the power output of a wind farm could pose a substantial negative impact on the power system reliability, especially for a power system with high penetration of wind generation [39] [40] [41]. In order to address this issue, accurate forecast and control models of the power output of wind farm have to be developed.

Dimensionality challenges the modeling of the total power output of all turbines of a large wind farm. In previous research, a number of wind turbine level models have been developed for controlling voltages, seeking maximum utilization of wind power based on single wind turbine model or model of a small number of turbines. For example, [42] presented some concepts of evaluation of the system's reliability based on a simplified wind power generation model; [43] as well as many others presented controller design for single wind turbine. However, the control of the total output of all turbines of a wind farm hasn't been explored. A modern wind farm usually consists of hundreds of wind turbines and each wind turbine is a nonlinear dynamic system. The behavior of the total power output of all wind turbines of an entire wind farm is not a simple aggregation of behaviors of individual turbines. The modeling is a real challenge because of dimension. For instance, it is practically impossible to control the output of wind farm by sending control signals to each turbine for the desired aggregate output of

wind farm [33] [84].

One possible solution to this dimensionality issue is model reduction. Model reduction seeks to replace a large-scale system by one or several lower order systems that maintain the dominating characteristics of the input-out behavior of the overall system response [44] [45] [46]. Various model reduction methods have been developed and applied to complex engineering systems which exhibit complex behavior. For example, [47] used Krylov subspace methods to simplify the model of a power system; [5] represented the complicated thermodynamic behavior of heat pumps with a reduced model of two steady operation states. In biomolecular dynamics, clustering of different configurations has been suggested to be an effective method to shed light on the nature of bimolecular dynamic behavior and their influence in biochemical reactions [10].

It is of great interest to explore if cluster analysis can be used to reduce the dimension of a dynamic wind farm model. If the turbines can be clustered, then the entire wind farm can be represented by several representative turbine models that capture the dominant system characteristics, which significantly reduces the dimension of the system.

This chapter presents a methodology of cluster analysis based on the theoretical research results presented in earlier chapters. We first assume the

power output of each wind turbine is a random process with Markovian characteristics, and the overall process of all turbines is then represented by a Markov transition matrix that is constructed from real data by building a graph with Gaussian weights. Then, spectral theory is applied to identify the number of clusters and we map the original wind turbines to the appropriate cluster using identification methods introduced in previous chapter. Then we present the results of clustering of 25 wind turbines located in three distinct locations of a wind farm based on the real power outputs for illustration and verification purpose of the proposed methodology. Finally, results of a comprehensive study of all turbines of the wind farm are also included.

We remark that assuming that the output of a wind turbine is a random process is based on the random character of the power source of the turbine, i.e. the wind.

5.2 Overview of Cluster Analysis

Clustering a complex dynamic system is characterized by a time scale separation and a spatial decomposition of the system dynamics. Cluster analysis is one of the model reduction techniques used for identification of sub-group feature of system dynamics. That is, the system output is partitioned into different groups on the basis of the proximity of individual dynamics of each group.

Clustering methods [48] can be divided into two basic types: hierarchical clustering and partitional clustering. Hierarchical methods find successive clusters using previously established clusters. That is, the method either merges smaller clusters into larger ones, or splits larger clusters into smaller ones. Hierarchical clustering is based on a certain measure of distance, and selects linkage method to form clusters. The distance measures include Euclidean distance, Manhattan distance, Mahalanobis distance, Hamming distance, etc [86].

Partitional method, on the other hand, intends to determine all the clusters at once. K-means clustering is one of the most commonly used methods in partitional clustering [85]. K-means clustering specifies the number of clusters in advance, then iterates groups observations based on nearest Euclidean distance to the mean of the cluster and calculates the K clusters until cluster means do not shift more than a given threshold value or the iteration limit is reached. However, two major drawbacks of K-means method are recognized by many recent studies such as [49] and [50]. First, the results could be very sensitive to choice of the number of clusters, which makes the method less stable: quite different kinds of clusters may emerge when K is changed; second, K-means method can't find solutions when the clusters are non-linearly separated in output space. Spectral methods, a powerful way to separate

non-convex groups of data has recently emerged in different fields [74] [75] [76] [77] [78]. Here, we build weighted graphs to represent a notion of geometry based on the local similarity or proximity between the data points. This method has shown great promise on data clustering in several fields [76] [77] [78]. The proposed clustering method is carried out in several steps. First, we define a weighted graph to represent the similarity of data sets. The graph is then used to construct a Markov Chain. Then we analyze the spectral properties of the Markov chain to identify the number of clusters. Finally, the elements of each cluster can be identified by some measures in eigenspace basis. Several important concepts for the proposed method are introduced below.

5.2.1 Construction of Markov Chain

In the field of spectral clustering, Markov random walks on graphs have proven to be very useful for identification of relevant structure when the underlying clusters have nonlinear shapes, see [52] and [13].

Let a real-valued random process be defined by $x(t),\ t \in T = \{1,\ldots,m\}$ and assume it has a stationary distribution. Given a set of samples from such random process $x_i(t),\ t \in T$, define a pairwise similarity matrix $A \in \mathbf{R}^{n\times n}$ by building a graph with Gaussian weights with entries

$$A_{ij} = e^{-\|x_i - x_j\|^2 / 2\sigma^2} \tag{5.1}$$

where σ is width parameter which represents the closeness of the data configuration [53]. Then A is row normalized to produce a Markov transition matrix $P \in \mathbf{R}^{n\times n}$ with entries

$$P_{ij} = \frac{A_{ij}}{\sum_j A_{ij}} \tag{5.2}$$

5.2.2 Diffusion Process

Let $\Omega \subset \mathbf{R}^m$ be a compact connected set with smooth boundary $\partial\Omega$. Assume that the samples x_i considered previously are i.i.d. random variables with common distribution with probability density $p(x)$ defined on Ω. For $x, y \in \Omega$ define the transition probability density

$$P_\varepsilon(x,y) = \frac{e^{-\|x-y\|^2/2\varepsilon}}{\int e^{-\|x-y\|^2/2\varepsilon} p(x)dx} \tag{5.3}$$

We note that $P_\varepsilon(x,y)$ is the continuous space analog of (5.2). In particular, it is indicated in [12] that in the limit $n \to \infty$ the Markov chain with transition matrix (5.2) converges to the Markov process with transition density $P_\varepsilon(x,y)$. Following [12] define the forward and backward operators T_f and T_b as

$$\begin{aligned} T_f[\varphi](x) &= \int_\Omega P_\varepsilon(x,y)\varphi(y)\,p(y)dy \\ T_b[\psi](x) &= \int_\Omega P_\varepsilon(y,x)\psi(y)\,p(y)dy \end{aligned} \tag{5.4}$$

We note that the forward operator T_f characterizes the propagation of the distribution of a discrete time Markov process on Ω with transition function

$$P(\varepsilon, x, A) = \int_A P_\varepsilon(x, y)\, p(y) dy \tag{5.5}$$

i.e., if φ is the probability density of the distribution of the process at time 0 then $T_f[\varphi](x)$ is the density at time ε. Similarly, $T_b[\psi](x)$ is the expected value of $\psi(x_\varepsilon)$ given that $x_0 = x$. We note that the eigenvalues and left and right eigenvectors of the transition matrix P given by (5.2) converge to the corresponding eigenvalues and eigenfunctions of the operators T_f and T_b [54].

We next consider the limit as the time step ε converges to zero, i.e. the continuous time limit. Thus, if we let $p(x,t)$ be the density of the discrete time process at time t and $p(x,t+\varepsilon) = T_f[p(\cdot,t)](x)$ be the distribution at time $t+\varepsilon$ then

$$\frac{\partial p(x,t)}{\partial t} = \lim_{\varepsilon \to 0} \frac{p(x,t+\varepsilon) - p(x,t)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{T_f - I}{\varepsilon} p(x,t) \tag{5.6}$$

As is defined in [12], the generators of forward operator and backward operator for the continuous time process are,

$$H_f = \lim_{\varepsilon \to 0} \frac{T_f - I}{\varepsilon} \text{ and } H_b = \lim_{\varepsilon \to 0} \frac{T_b - I}{\varepsilon} \tag{5.7}$$

We note that the eigenfunctions of T_f and T_b converge to those of H_f and H_b, respectively. Therefore, provided n is large enough, the structure and characteristics of eigenvalues and eigenvectors of the discrete Markov chain P are similar to those of eigenvalues and eigenvectors of H_f and H_b. We note that the continuous time process with generator H_b is a diffusion process on Ω.

Furthermore, as is shown in [12], if we let $p(x)$ be on the Boltzman form $p(x) = e^{-U(x)/2}$ then the backward operator becomes the Fokker-Planck operator for the diffusion process

$$\dot{x} = -\nabla U + \sqrt{2}\dot{w} \tag{5.8}$$

where w is a standard Brownian motion.

To conclude, the probability density $p(x)$ of the samples X represents the inherent data structure and is the density for the stationary distribution of the diffusion process that is assumed to govern the dynamics of the underlying system. The data is generated by the diffusion process with forward and backward operators whose eigenfunctions are similar to the eigenvectors of the finite dimensional Markov chain generated by the data. Consequently, the structure and characteristics of eigenvalues and eigenvectors of discrete Markov chain P capture the intrinsic properties of the sampled data set.

5.2.3 Spectral Analysis

The P matrix is the object of interest for finding the clusters. Spectral analysis of the Markov transition matrix, namely analysis of eigenvalues and eigenvectors, is employed to find the geometric structure of the data. In order to apply spectral theory, we assume the Markov chain is aperiodic and irreducible, i.e. the chain is ergodic. Then there exists an unique stationary distribution π

whose definition and properties is introduced in Chapter 2. Since the similarity matrix A defined by a Gaussian weighted graph is symmetric, the Markov matrix P is reversible.

Let λ_i, φ_i and ψ_i, $i = 0,\ldots,n-1$, denote the ith eigenvalue, left eigenvector and right eigenvector of P, respectively. Since P is the transition matrix of a reversible Markov chain all of its eigenvalues are real. If the eigenvalues are arranged in decreasing order:

$$1 \geq \lambda_0 \geq \lambda_1 \geq \lambda_2 \cdots \geq \lambda_i \geq \cdots \geq \lambda_{n-1} \geq 0$$

then the spectral decomposition of the Markov matrix P may be written as

$$P = \sum_{k=0}^{n-1} \lambda_k \psi_k \varphi_k^T \tag{5.9}$$

and a lower dimensional approximate model can be defined as

$$P_a = \sum_{k=0}^{q-1} \lambda_k \psi_k \varphi_k^T \tag{5.10}$$

where we have retained the first q components of the spectral decomposition.

There are two conditions that need to be satisfied to find good approximation P_a. If P has $q < n$ dominant eigenvalues, i.e. we assume $\lambda_0,\ldots,\lambda_{q-1}$ are of comparable size (close to one) and $\lambda_q << \lambda_{q-1}$, then a good approximate model can be defined by the first q eigenvalues and corresponding left and right eigenvectors. In this case there exist q well-identifiable clusters characterized by the dominant eigenvectors. However, frequently P has not

obvious dominant eigenvalues all close to one. In this case we look for a spectral gap in the eigenvalues, i.e. we look for the first value of q such that eigenvalues starting from λ_q are small and of comparable same size and $\lambda_{q-1} - \lambda_q >> \lambda_{i-1} - \lambda_i$ for all $q+1 \le i \le n-1$. Note that $\|P - P_a\|$ is bounded above by λ_q, i.e. if λ_q is small the approximate model is a good representation of the original P [26].

5.3 Cluster Analysis of Wind Farm Power Output

In this section we present a cluster analysis for wind farm power generation based on real data using the method introduced in the previous sections and chapters.

5.3.1 The Data

The data used in this study are from the Centennial Wind Farm which is one of largest wind farms in Oklahoma. The wind farm has an array of 80 GE 1.5MW SLE turbines, each standing 262 feet tall at its hub and the overall height of the structure is 389 feet. The turbines' location in the wind farm is shown in Fig. 5-1.

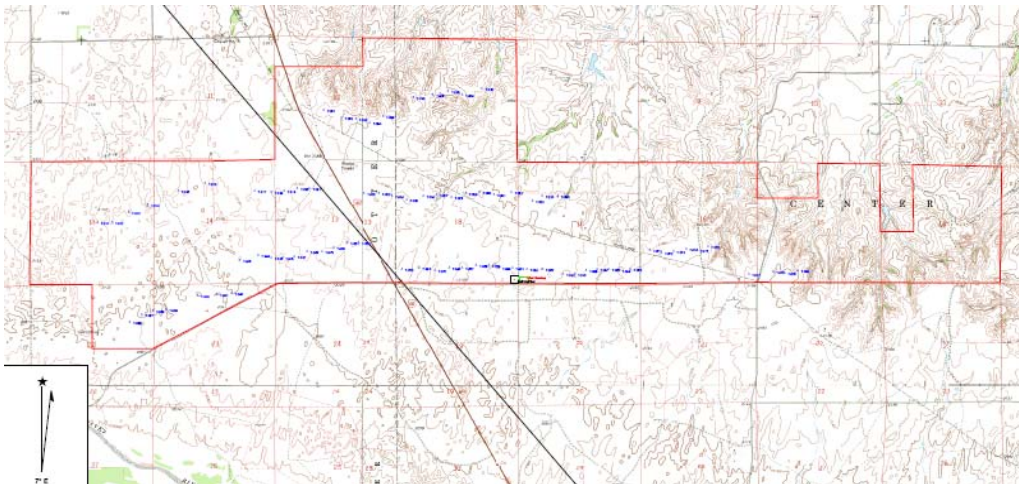

Fig. 5-1 80 turbines' site in the wind farm

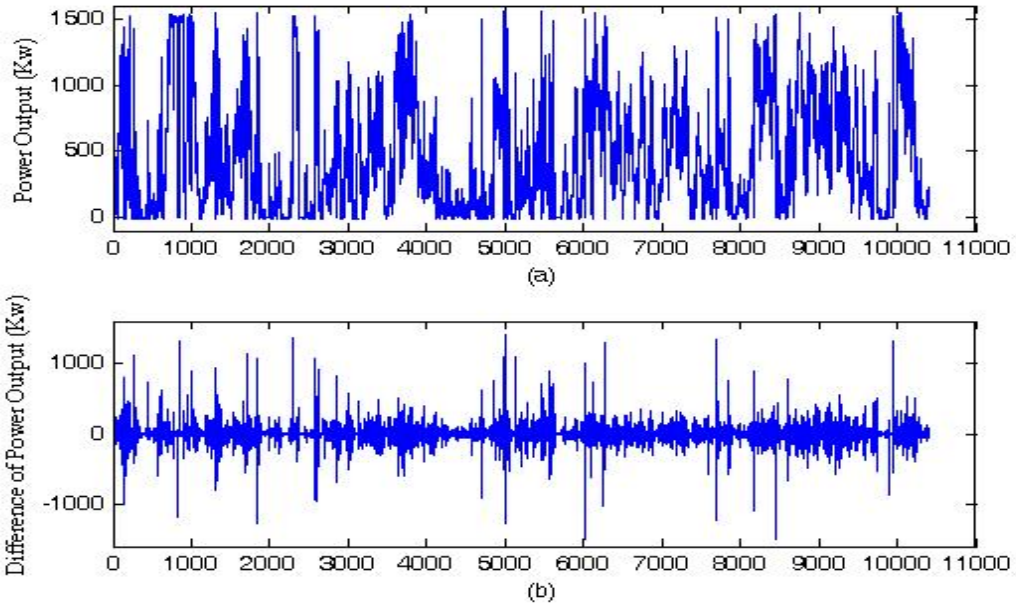

Fig. 5-2 A turbine's power output and its difference of power output

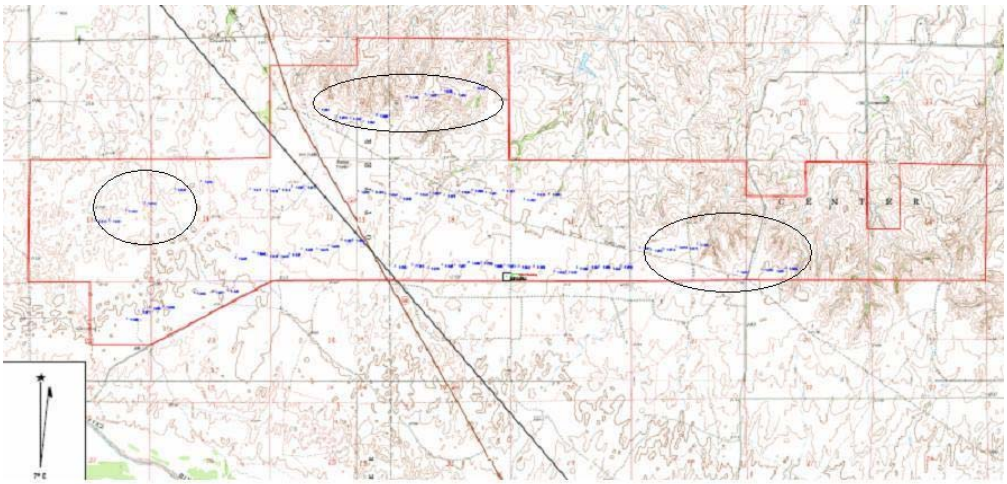

Fig. 5-3 The locations of 25 wind turbines

The turbines in the wind farm are equipped with Supervisory Control and Data Acquisition (SCADA) system. The raw data in each channel is generated as the mean value of 600 sampled data points in each 10 minutes interval. Thus the sampled data sequence can be treated as a random process. The time-series power output data used for the study is based on the observations between June 1st and August 31st in 2007. One of the turbines had system problems during the time of study so there are actually 79 turbines studied. Fig. 5-2 (a) shows the power output of a single turbine.

For the cluster analysis of the wind turbines of the wind farm, we found that the difference of SCADA time-series data between two time intervals is a better representation of the dynamics of the turbine than the absolute output level. Such difference removes drift and reflects the inherent dynamics. The difference is defined as

$$x(t) = y(t+1) - y(t) \tag{5.11}$$

where $y(t)$ is time series power output data of a wind turbine and $x(t)$ is the corresponding difference. We note that $x(t)$ is Markov process. Table 5.1 shows the original and difference time series data. Fig. 5-2 (b) shows the time-series of the difference for the turbine shown in Fig. 5-2 (a).

The analysis of 25 turbines at three distinct locations of the wind farm is presented for illustration and verification purposes. The complete analysis of all wind turbines of the wind farms is presented afterwards. The locations of the wind turbines are shown in Fig. 5-3.

Table 5.1: Original and difference series data

Time series	Turbine 1		Turbine 2		. . .	Turbine 25	
	y(t)	x(t)	y(t)	x(t)		y(t)	x(t)
1	-3.7	0.46	33.81	-28.9		91.38	-52.2
2	-3.24	0.39	4.847	1.473		39.17	17.69
3	-2.85	.	6.382	.	.	56.86	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.		.			.	
10409	150.6	-53.3	1.552	-5.75		133.6	2.8
10410	97.34		-4.20			136.4	

5.3.2 Construction of Markov Chain

In order to measure the likelihood of any two turbines having similar output dynamics, a similarity matrix and the corresponding Markov transition matrix needs to be constructed. The data of the difference is defined by (5.11), which can be seen as a set of random processes $X = \{x_1, \ldots, x_{25}\}$, where x_i is a discrete random process for the ith turbine. The data corresponding to each time interval are listed in Table 5.1. The time-series data are used to construct the Markov matrix. In order to weigh each difference at same size and avoid choosing very big value of σ in (5.1), the processes are rescaled by $X = X / \max(abs(X))$. Based on (5.1), the similarity matrix $A \in \mathbf{R}^{25\times 25}$ is calculated as

$$\begin{bmatrix} 1 & 0.3575 & \cdots & 0.0591 & 0.058 \\ 0.3575 & 1 & \cdots & 0.0587 & 0.0565 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0.0591 & 0.0587 & \cdots & 1 & 0.3775 \\ 0.058 & 0.0565 & \cdots & 0.3775 & 1 \end{bmatrix}$$

where we have chosen $\sigma = 18$.

After normalization of A as in (5.2), we can get Markov transition matrix $P \in \mathbf{R}^{25\times 25}$ and

$$P = \begin{bmatrix} 0.2657 & 0.0950 & \cdots & 0.0157 & 0.0154 \\ 0.0864 & 0.2416 & \cdots & 0.0142 & 0.0137 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0.0159 & 0.0158 & \cdots & 0.2684 & 0.1013 \\ 0.0162 & 0.0158 & \cdots & 0.1054 & 0.2791 \end{bmatrix}$$

5.3.3 Spectral Analysis

This subsection describes how to determine the number of disjoint wind turbines' clusters. Based on the spectral analysis in the Section 5.2, the eigenvalues and eigenvectors of P contain the information about the characteristics of cluster partitions. LAPACK routine [37] is used to calculate the eigenvalues, eigenvectors of matrix P and stationary distribution π. Sorting the eigenvalues of the Markov transition matrix P in descending order, i.e. $\lambda_{i-1} \geq \lambda_i$, $i = 0,\ldots,24$ gives

$$\begin{aligned}\lambda &= [\lambda_0 \quad \lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5 \quad \cdots \quad \lambda_{24}] \\ &= [1 \quad 0.5707 \quad 0.4767 \quad 0.3767 \quad 0.3439 \quad 0.3002 \quad \cdots \quad 0.1344]\end{aligned}$$

We note that P has not q $(q > 1)$ dominant eigenvalues which are close to one and $\lambda_q << \lambda_{q-1}$. However we note that the gap between the third and fourth eigenvalues is much larger than the gap between all higher indexed eigenvalues and thus we pick $q = 3$. Furthermore, $\|P - P_a\|$ is small and changes only slightly for q larger than 3. Therefore a good approximate model is obtained with $q = 3$, i.e. the twenty-five wind turbines can be grouped into 3 disjoint clusters based on similarity of their dynamics.

5.3.4 Sign Structure Method

Once the number of clusters is determined, we need to identify the wind turbines belonging to each cluster based on the sign structure identification

algorithm described in Section 4.2.1. The sign structure is characterized by the right eigenvectors of P. In this Example, the 3 most stable's sign structure collection SS after rescaling are {[1 -1 -0.856], [1 -0.742 1], [1 1 -0.1856]} which represent sign structure {[+ - -], [+ - +], [+ + -]} and the threshold T is 0.1125. Then three clusters can be identified effectively by assigning each wind turbines to "nearest" cluster.

5.3.5 Diffusion Distance Method

Once the number of clusters is determined, we need to identify the wind turbines belonging to each cluster based on the diffusion distance method described in Section 4.2.2 The defined diffusion distance is characterized by the eigenvalues and the corresponding right eigenvectors of P.

The diffusion distance for any two wind turbines is calculated using (4.5). For example, the distance between turbine 1 and turbine 2 can be calculated as follows,

$$\begin{aligned} D^2(x_1, x_2) &= \left\| \Psi(e_1^T) - \Psi(e_2^T) \right\|^2 \\ &\approx \left\| \begin{matrix} \lambda_0(\psi_0(1) - \psi_0(2)) \\ \lambda_1(\psi_1(1) - \psi_1(2)) \\ \lambda_2(\psi_2(1) - \psi_2(2)) \end{matrix} \right\|^2 \\ &= \left\| \begin{matrix} 0 \\ -0.009 \\ -0.0079 \end{matrix} \right\|^2 = 0.0001 \end{aligned}$$

Similarly, the diffusion distance for other turbines can be calculated one by one. In this case, three clusters can be identified effectively if using the algorithm proposed in Section 4.2.2.

5.3.6 Results and Discussion

The proposed method has been applied to cluster analysis of wind turbines of the wind farm. The results are shown in Table 5.2, Table 5.3 and Table 5.4. In order to verify the effectiveness of the method, twenty-five turbines are randomly ordered when we construct the Markov matrix, and the result of clustering is shown in Table 5.3. Fig. 5-4 and Fig. 5-5 show the average power output of each turbine in each cluster as well as the standard deviations. The different level of the average and the deviation in each cluster shows the clustering analysis for wind turbines exactly captures the different characteristics of power output. The result of cluster analysis of 79 turbines of entire wind farm is shown in Table 5.4.

There are some unique features identified when the proposed method was implemented for cluster analysis of wind turbines in terms of real power output. First of all, since wind speed fluctuates sharply from minute to minute, the power output of wind turbines varies fast even in the average value of 10-mintues interval. Whether power output or difference of power output is applied, the data range is always [0, 1500] or [-1500, 1500], respectively. When we construct the

Markov matrix by building a graph with Gaussian weight, such large data values will make it difficult to select proper σ to avoid a sparse P matrix or reducible P matrix. So in the implementation, we rescale the data into the range [-1, 1]. We note that from (5.1) this is equivalent to choosing a scaling value for σ. The other issue is how to efficiently and correctly assign wind turbines to each cluster. In order to avoid some wind turbines around the "separation line" between clusters which may not be correctly clustered, the algorithm proposed in Section 4.2.1 is recommended.

In summary, a large scale wind farm is divided into different clusters by the method developed in Section 5.2. Each cluster is a collection of wind turbines which have similar power output dynamics. So the wind farm can be modeled by several typical turbines each of which represents a cluster. The turbines and the coordination between the turbines will play an important role in solving the problems addressed by current research [55] [56] [57] [58] [59]. Thus, the clustering of large wind farms will promote and improve integration of wind farms.

Table 5.2: Clustering results for 25 wind turbines

Index of 25 wind turbines	1 2 3 … 24 25		
q	3		
Sorted eigenvalues	1 0.5707 0.4767		
	Cluster 1	Cluster 2	Cluster 3
Index for each cluster	1-10	11-15	16-25

Table 5.3: Clustering results for 25 wind turbines (2)

Reindex of 25 wind turbines corresponding each one in Table I	8 18 14 4 10 22 15 13 23 2 21 17 9 19 12 16 6 20 3 24 5 25 7 1 11		
q	3		
Sorted eigenvalues	1, 0.5707, 0.4767		
	Cluster 1	Cluster 2	Cluster 3
Index for each cluster	8 18 14 4 10 22 15 13 23 2	21 17 9 19 12	16 6 20 3 24 5 25 7 1 11

Table 5.4: Clustering results for 79 wind turbines

Index of 79 wind turbines	1 2 3 … 78 79		
q	2		
Sorted eigenvalues	1 0.1912 0.1244		
	Cluster 1	Cluster 2	Cluster 3
Index for each cluster	1 2 3 4 5 6 7 8 9 10 20 21 22 23 24 25 26 27 28 29 30 31 32 48 49 50 51 52 53 54 55	11 12 13 14 15 16 17 18 19 35 36 37 38 39 40 41 42 43 44 45 46 47	33 34 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79

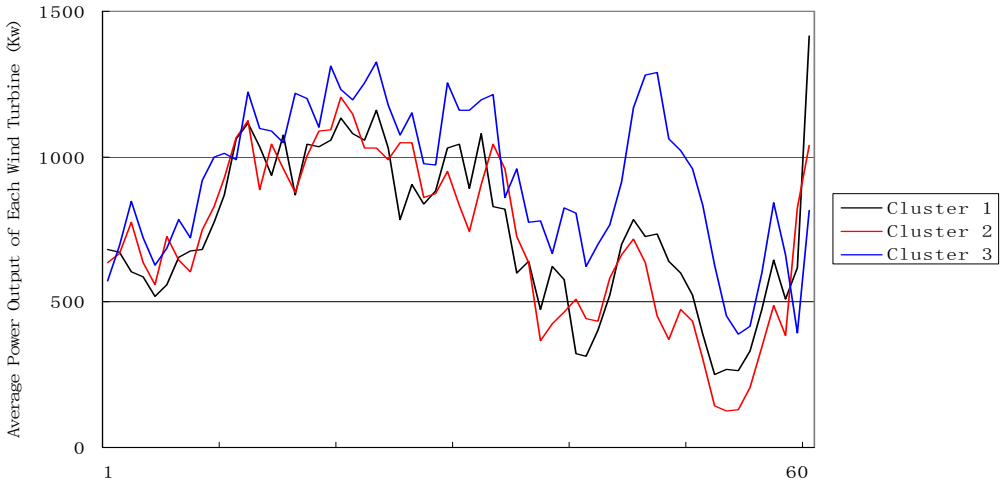

Fig. 5-4 Average power output of each turbine

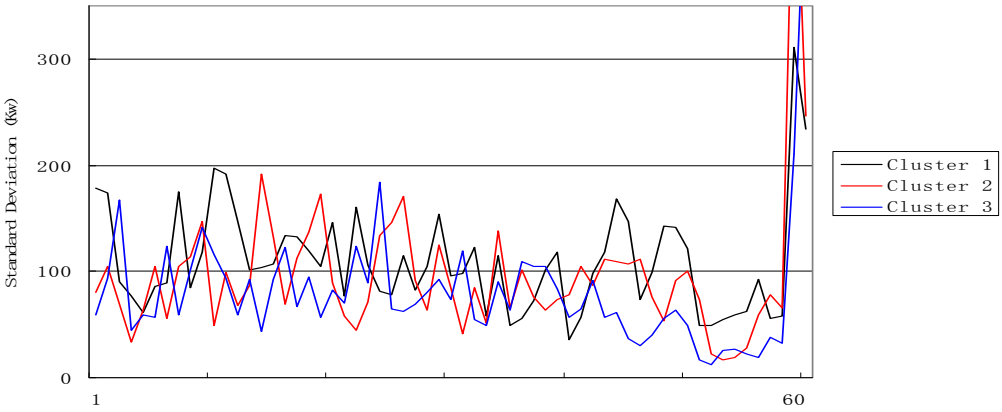

Fig. 5-5 STD of power output of each turbine

6 CONCLUDING REMARKS AND FUTURE RESEARCH

In this dissertation, we develop a comprehensive model identification approach for complex multi-modal system given data either from real system operation or large scale dynamic simulation models. The approach is based on a transfer operator approach for the description of the system dynamics and spectral theory for this operator for characterizing the system modal behavior. We don't only address the theoretical aspects of the developed techniques but also algorithmic development, numerical implementations and case studies.

The developed identification strategy entails (i) model reduction of nonreversible Markov chains and the identification of number of modes; (ii) the identification of the domains or regions in state space where the system spends a long time between transitions and the identification of dynamics that characterize the transitive behavior between components (dynamics in slow time scale); (iii) modeling and identification of local dynamics inside each domain (dynamics in fast time scale). In part (i), we formulate nonreversible Markov process for modeling the complex system dynamics, develop spectral properties for nonreversible processes and present the low dimensional

approximations for general nonreversible Markov chains. In part (ii), we propose diffusion distance method and sign structure method both of which are related to eigenfunctions of a multiplicative reversible Markov process to identify metastable components of the state space and transition dynamics between these metastable components. In order to alleviate the computation burden of eigenvalues and eigenvectors of the multiplicative reversible Markov chain which is subject to the exponential increase of dimensionality, we also present an approximation technique which is based on Nyström extension method. In part (iii), in each metastable component, we choose subspace identification methods (SIMs) with estimation of noise sequence to identify local system dynamics which have a stable point attractor and are approximately linear around the point. Furthermore, we illustrate the construction of the approximate and reduced systems, identification of multi-modal dynamics and computation of approximate eigenfunctions in a couple of numerical examples.

We then develop a novel approach to address a pressing issues associated with penetrations of large scale wind farms. The method combines Markov chain techniques to reduce the complexity of power output dynamics of a large scale wind farm. The proposed method uses time-series power output of all turbines of the wind farm to construct a Markov matrix by building a graph with

Gaussian weights. The number of clusters is identified by spectral properties of the Markov chain and wind turbines classified into different clusters by both diffusion distance method and sign structure method. The implementation of the approach in a large scale wind farm in Oklahoma is presented step by step. The results of clustering for both twenty-five and seventy-nine wind turbines demonstrate that the proposed method is very effective for clustering the different characteristics of power output in the wind farm.

In this dissertation, we complete the theory for identification of metastable components and modal dynamics for nonreversible Markov chains and develop subspace type identification procedures for identification of local dynamics for systems that exhibit strong point attractor behavior in each component. In the future, the research will be extended to systems with noisy output measurements and control inputs. The future research will extend the theory developed for uncontrolled Markov processed to the case of controlled nonreversible Markov process. This will involve the identification of metastable components in the state-control space and the identification control dependent modal dynamics. Furthermore, the research will extend identification of local nonlinear dynamics (e.g. periodic or chaotic behavior) to both of the non controlled case and the controlled case in each component.

In addition, we present a novel methodology for cluster analysis of wind

turbines in a wind farm based on its power output dynamics. In the next step, we will develop a wind turbine model with control input to represent each cluster and develop a control strategy for the whole wind farm. For example, if the control input is a function of the instantaneous wind speed, we may use an equivalent wind velocity in the cluster to drive the representative wind turbine model in some cluster. Then control strategy can be developed based on estimation of wind speed and coordination of each representative dynamic wind turbine model. In addition, the power prediction of a wind farm can be made more efficient and accurate by using artificial neural networks (ANN), mixture of experts (ME), or support vector machine (SVM), if those methods are applied to each cluster independently.

REFERENCES

[1] S. H. Strogatz, "Exploring complex networks," *Nature*, 410:268-276, 2001.

[2] P. M. Anderson and A. A. Fouad, *Power System Control and Stability*, John Wiley and Sons, New York, 2002.

[3] A. S. Willsky and B. C. Levy, "Stochastic stability research for complex power systems," *Lab. Inf. Decision Systems, MIT*, *Report ET-76-C-01-2295*, 1979.

[4] J. T. Pukrushpan, A. G. Stefanopoulou, S. Varigonda, L. M. Pedersen, S. Ghosh and H. Peng, "Control of natural gas catalytic partial oxidation for hydrogen generation in fuel cell applications," *IEEE Trans. on Control Systems Technology*, 13:3-14, 2005.

[5] B. Eisenhower and T. Runolfsson, "Modeling and analysis of bistable behavior in a transcritical heart pump," *Proceedings of the 2004 IEEE CDC*, 2004.

[6] E. D. Sontag, "Nonlinear regulation: The piecewise linear approach," *IEEE Trans. Automat. Contr.*, AC-26:326-356, 1981.

[7] E. Amaldi and M. Mettavelli, "The min pfs problem and piecewise linear model estimation," *Discrete Appl. Math*, 118:115-143, 2002.

[8] M. Dellnitz and O. Junge, "On the approximation of complicated dynamical behavior," *SIAM J. Numerical Analysis*, 36, 491-515, 1999.

[9] M. Dellnitz and O. Junge, *Set Oriented Numerical Methods for Dynamical Systems, Handbook of Dynamical Systems II: Towards Applications*, World Scientific, 2000.

[10] W. Huisinga and B. Schmidt, "Metastability and dominant eigenvalues of transfer operators, advances in algorithms for macromolecular

simulation," *Lecture Notes in Computational Science and Engineering,* 2005.

[11] W. Huisinga, "Metastability of Markovian systems - a transfer operator approach in application to molecular dynamics," *Workshop on Atomistic to Continuum Models for Long Molecules and Thin Films*, Monte Verita, Ascona, Switzerland, July 20, 2001.

[12] B. Nadler, S. Lafon, R. Coifman, and I. Kevrekidis, "Diffusion maps, spectral clustering and eigenfunctions of fokker-planck operators," 2005, preprint.

[13] R. R. Coifman, and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis: Special Issue on Diffusion Maps and Wavelets,* vol. 21, pp. 5-30, July 2006.

[14] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*, Springer, Berlin, 1993.

[15] A. Lasota and M.C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, Springer-Verlag, 1994.

[16] C. Schütte, W. Huisinga, and P. Deuhard, "Transfer operator approach to conformational dynamics in biomolecular system," *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, pp.191-223 Springer, 2001.

[17] J. Fill, "Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process," *The Annals of Applied Probability,* vol.1, pp. 62-87, 1991.

[18] T. Runolfsson and Y. Ma. "Model reduction of non-reversible Markov chains," *In Proceedings of the IEEE CDC, New Orlean*s, 2007.

[19] A. F. P. Deuflhard, W. Huisinga and C. Schütte, "Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains," *Lin. Alg. Appl.*, 2000.

[20] K. P. Bennett and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets," *Optim. Meth. Software*, 1:23-34, 1992.

[21] W. Lin, S. Qin, and L. Ljung, "On consistency of closed-loop subspace identification with innovation estimation," *In Proceedings of the IEEE CDC, Bahamas*, 2004.

[22] L. Ljung, *System Identification - Theory for the User, 2nd ed.*, PTR Prentice Hall, Upper Saddle River, N.J., 1999.

[23] C. T. H. Baker, *The Numerical Treatment of Integral Equations*, Oxford: Clarendon Press, 1977.

[24] E. J. Nyström, "Über die praktische auflösung von linearen Integralgleichungen mit anwendungen auf randwertaufgaben der potentialtheorie," *Commentationes Physico-Mathematicae*, vol. 4, no. 15, pp. 1-52, 1928.

[25] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, second ed.*, Cambridge Univ. Press, 1992.

[26] G. W. Stewart and J. G. Sun, *Matrix Perturbation Theory,* Academic Press, 1990.

[27] G. Ferrari- Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine systems," *Automatica*, 29:205-217, 2003.

[28] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (gpca)," *IEEE Trans Pattern Anal. and Machine Learn.*, 27:1945-1959, 2005.

[29] P. Drineas, R. Kannan, and M. W. Mahoney, "Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix," *Technical Report YALEU/DCS/TR-1270,Yale University Department of Computer Science,* New Haven, CT, February 2004.

[30] P. Drineas, R. Kannan, and M. W. Mahoney, "Fast Monte Carlo algorithms for matrices III: Computing a compressed approximate matrix decomposition," *Technical Report YALEU/DCS/TR-1271, Yale University Department of Computer Science*, New Haven, CT, February 2004.

[31] P. Drineas, R. Kannan, and M. W. Mahoney, "Fast Monte Carlo algorithms for matrices I: Approximating matrix multiplication," *Technical Report YALEU/DCS/TR-1269, Yale University Department of Computer Science*, New Haven, CT, February 2004.

[32] O. Junge, J. E. Marsden and I. Mezic, "Uncertainty in the dynamics of conservative maps," in *Proc. 43rd IEEE Conf. on Decision and Cont.*, vol. 2 pp. 2225-2230, 2004.

[33] P. B. Eriksen, T. Ackermann, H. Abildgaard, P. Smith, W. Winter, and J. R. Garcia, "System operation with high wind penetration," IEEE Power Energy Mag., vol. 3, no. 6, pp. 65 – 74, Nov./Dec. 2005

[34] P. Drineas, R. Kannan, and M. W. Mahoney, "Sampling sub-problems of heterogeneous Max-Cut problems and approximation algorithms," *Technical Report YALEU/DCS/TR-1283, Yale University Department of Computer Science,* New Haven, CT, April 2004.

[35] P. Drineas, R. Kannan, and M. W. Mahoney, "Sampling sub-problems of heterogeneous Max-Cut problems and approximation algorithms," *In Proceedings of the 22nd Annual International Symposium on Theoretical Aspects of Computer Science*, pages 57–68, 2005.

[36] P. Drineas and M. W. Mahoney, "Approximating a Gram matrix for improved kernel-based learning," *In Proceeding of the 18th Annual Conference on Learning Theory (COLT)*, pp. 323–337, 2005.

[37] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK User's Guide*, http://www.netlib.org/lapack/lug/lapack_lug.html, [Online] . Third Edition, SIAM, Philadelphia, 1999.

[38] *GWEC Global Wind 2007 Report*, [Online]. Available: http://www.gwec.net.

[39] *EWEA 2006 Annual Report*, [Online]. Available: http://www.ewea.org.

[40] P. Gardner, H. Snodin, A.Higgins, and S. McGoldrick, "The impacts of increased levels of wind penetration on the electricity systems of the republic of Ireland and Northern Ireland: final report," [Online]. Available: http://www.cer.ie.

[41] *Final Report System Disturbance on 4 November 2006*, Published by UCTE (Union for the Co-ordination of Transmission of Electricity), December 2007.

[42] R. Karki, P. Hu and R. Billinton, "A simplified wind power generation model for reliability evaluation," *IEEE Trans. Energy Convers.*, vol. 21, no. 2, pp. 533-540, Jun. 2006.

[43] M. Jelavic, N. Peric, I. Petrovic, "Identification of wind turbine model for controller design," *12th international Power Electronics and Motion Control Conference(EPE-PEMC),* pp. 1608-1613, Aug. 2006.

[44] A. C. Antoulas and D. C. Sorensen, "Approximation of large-scale dynamical systems: an overview," *Technical Report,* February*,* 2001.

[45] S. Lall, J. E. Marshden, and S. Galvaski, "A subspace approach to balanced truncation for model reduction of nonlinear control systems*," Int. J on Robust and Nonlin. Contr.*, 2002.

[46] J. M. A. Scherpen and W. S. Gray, "Minimality and local state decomposition of a nonlinear state space realization using energy functions*," IEEE Tran. Automat. Contr.*, AC-45, 2000.

[47] D. Chaniotis and M. A. Pai, "Model reduction in power systems using Krylov subspace methods," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 888-894, May 2005.

[48] J. Abonyi and B. Feil, *Cluster Analysis for Data Mining and System Identification*, Boston and Basel, Switzerland: Birkhäuser Basel, 2007.

[49] I. S. Dhillon, Y. Guan, and J. Kogan, "Iterative clustering of high dimensional text data augmented by local search," *In Proceedings of The 2002 IEEE International Conference on Data Mining,* pages 131–138, 2002.

[50] R. Kannan, S. Vempala, and A. Vetta, "On clusterings good, bad, and spectral," *In Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, 2000.

[51] F. Bach and M. I. Jordan, “Blind one-microphone speech separation: a spectral learning approach,” Advances in Neural Information Processing Systems, 17, 2002.

[52] M. Szummer, and T. Jaakkola, “Partially labeled classification with Markov random walks,” *Neural Inf. Process*, vol. 14, pp. 945–952, 2001.

[53] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On spectral clustering: analysis and an algorithm,” *Proc. Neural Info. Processing Systems (NIPS 2001)*, 2001.

[54] B. Nadler, S. Lafon, R. R. Coifman and I. G. Kevrekidis, “Diffusion maps, spectral clustering, and the reaction coordinates of dynamic systems,” *Applied and Computational Harmonic Analysis: Special Issue on Diffusion Maps and Wavelets*, vol. 21, pp 113-127, July 2006.

[55] T.-H. Yeh, and L. Wang, “Study on generator capacity for wind turbines under various tower heights and rated wind speeds using Weibull distribution,” *IEEE Trans. Energy Convers.*, vol. 23, no. 2, pp. 592-602, Jun. 2008.

[56] H. Banakar, C. Luo and B. T. Ooi, **“**Impacts of wind power minute-to-minute variations on power system operation,” *IEEE Trans. Power Syst.*, vol. 23, no. 1, pp. 150-160, Feb. 2008.

[57] A. Abo-Khalil and D. Lee, “Dynamic modeling and control of wind turbines for grid-connected wind generation system,” *In Proc. IEEE Power Electronics Specialists Conference,* pp. 1-6, Korea, Jun. 2006.

[58] R. Billinton and W. Wangdee, “Reliability-based transmission reinforcement planning associated with large-scale wind farms,” *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 24-41, Feb. 2007.

[59] R. Spee, S. Bhowmik, and J. H. R. Enslin, “Novel control strategies for variable-speed doubly fed wind power generation systems,” *Renewable Energy*, vol. 6, no. 8, pp. 907–915, 1995.

[60] J. Sjoberg, Q. Zhang, and L. Ljung, “Nonlinear black-box modeling is system identification: a unified overview,” *Automatica*, 31:1691-1924, 1995.

[61] A. C. Antoulas and B. D. O. Anderson, “State space and polynomial approaches to rational interpolation”, *Progress in Systems and Control Theory III: Realization and Modeling in System Theory*, M.A. Kaashoek, J.H. van Schuppen and A.C.M. Ran Editors, Birkhäuser, pp.73-82,1990.

[62] G. Picci, “Statistical properties of certain subspace identification methods,” *Proceedings of the SYSID 97*, vol. 3, pp. 1093-1099, Fukuoka, Japan, 1997.

[63] J. Sorelius, T. Soderstrom, P. Stoica, and M. Cedervall. “Order estimation method for subspace based system identification,” *In Proc. System Identification (SYSID)*, 1997.

[64] K. Peternell, *Identification of Linear Dynamic Systems by Subspace and Realization-Based Algorithms*, PhD thesis, TU Wien, Austria, 1995.

[65] D. Bauer, *Some Asymptotic Theory for the Estimation of Linear Systems Using Maximum Likelihood Methods or Subspace Algorithms*, PhD thesis, TU Wien, Austria, 1998.

[66] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, “A bounded-error approach to piecewise affine system identification,” *IEEE Trans Automat. Contr.*, AC-50:1567-1580, 2005.

[67] A. Juloski, W. P. M. H. Hemmels and S. Wieland, “A Bayesian approach to the identification of hybrid systems,” *IEEE Trans. Automat. Contr.*, AC-50:1520-1533, 2005.

[68] R. Vidal, “Identification of PWARX hybrid models with unknown and possibly different orders,” *In Proceedings of the 2004 American Control Conference,* Boston, MA, 2004.

[69] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, Inc., 1990.

[70] T. M. Mitchell, *Machine Learning*, McGraw Hill, 1997.

[71] V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley & Sons, Inc., 1976.

[72] S. Van Vaerenbergh, J. Via, and I. Santamaria, "A sliding window kernels algorithm and its application to nonlinear channel identification," *In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* Toulouse, France, May 2006.

[73] B. Scholkopf, A. Smola, and K. R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[74] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *Journal of Machine Learning Research*, vol. 3, pp. 1–48, 2003.

[75] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least squares-algorithm," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, Aug. 2004.

[76] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. PAMI*, vol. 22, no. 8, pp. 888-905, 2000.

[77] F. Chung, *Spectral Graph Theory*, In: CBNS-AMS, vol. 92, Amer. Math. Soc., Providence, RI.

[78] T. Haveliwala, "Topic-sensitive PageRank: A context-sensitive ranking algorithm for web search," *IEEE Trans. Knowl. Data Eng.,* vol. 25, no. 4, pp. 784-796, 2003.

[79] H.-D. Chiang, I. Dobson, R. J. Thomas, J. S. Thorp, and L. Fekih-Ahmed, "On voltage collapse in electric power systems," *IEEE Trans. Power Syst.,* vol. 5, pp. 601-611, 1990.

[80] E. H. Abed, A. M. A. Hamdan, H.-C. Lee, and A. G. Parlos, "On bifurcations in power system models and voltage collapse," in *Proc. 29th IEEE Conf. on Decision and Cont.*, pp. 3014-3015, 1990.

[81] J. F. Hauer, "Advanced topics in power system dynamics," in *Proc.: EPRINSF Workshop on Appl. of Advanced Math.* to *Power Syst.: Report TR-101795,* A.B. Ranjit Kumar, A. Ipakchi, F. Alvarado, I. Dobson, *S.* Mulcherjee and W.H. Esselman, Eds., Electric Power Research Institute, 1993.

[82] C. Li, G. Chen, X. Liao, and J. Yu, "Hopf bifurcation in an internet congestion control model," *Chaos, Solitons, and Fractals,* 19:853–862, 2004.

[83] R. Srikant, *The Mathematics of Internet Congestion Control Systems & Control: Foundations & Applications,* Birkhauser, Boston, MA, 2004.

[84] P. Giorsetto and K. F. Utsurogi, "Development of a new procedure for reliability modeling of wind turbine generators," *IEEE Trans. PAS*, vol. 102, no. 1, pp. 134–143, Jan. 1983.

[85] P. S. Bradley and U. Fayyad, "Refining initial points for K-means clustering," *Proc. 15th Intl. Conf. Machine Learning*, pp. 91-99, 1998.

[86] V. Capoyleas, G. Rote and G. Woeginger, "Geometric clusterings," *J. Algorithms*, vol. 12, pp. 341-356, 1991.