UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

PROBLEMS IN THE DESIGN AND OPERATION OF UNCERTAIN COMPLEX ENGINEERING SYSTEMS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

CHENXI LIN
Norman, Oklahoma
2011

PROBLEMS IN THE DESIGN AND OPERATION OF UNCERTAIN COMPLEX ENGINEERING SYSTEMS

A DISSERTATION APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

Dr. Thordur Runolfsson, Chair

Dr. Nikola Petrov

Dr. Joe Havlicek

Dr. John Jiang

Dr. Choon Yik Tang

To My Parents

Acknowledgements

First of all, I would like to thank my academic advisor Dr. Thordur Runolfsson for his guidance and supports during my Ph.D. studies. He is a smart, knowledgeable person with endless passions for his career. Thanks for his constant encouragement which drives me along the research study, and for his patience and tolerance which give me more freedom in my study and research interests. His profound impact is not only on my studies, but also to my life altitudes—to become a tolerant and knowledgeable person.

I would also like to thank Dr. John Jiang. Without his guidance, I won't have a chance to apply my knowledge in power systems and wind energy. He is well noted for his research and personality. I am under the shelter of his research experience and reputation in the industry. Thanks for his recommendation for my several internship opportunities.

I would like to extend my thanks to all the faculty and staff members in ECE for their hard working to make the department run smoothly. And also I would like to thank all my dear friends in OU who help me out through struggling time, and paint the color for my life.

Last but most importantly, I would like to give my greatest thanks to my family in China for their unconditional love, strong supports and trusts.

Contents

List of Tables

List of Figures

Abstract

In this dissertation, we consider two problems. The first one is a general approach to the optimal design of uncertain dynamical systems, where the uncertainty is represented by a random parameter. The problem is formulated using two types of performance criteria, that result in two different optimal design methods. However, both of them are difficult to solve analytically for most uncertain complex dynamical systems. A numerical scheme is developed for the optimal design that involves two steps. First, in order to obtain a numerical algorithm for the optimal solution, we apply randomized algorithms for average performance synthesis to approximate the optimal solution. Second, using the properties of the Perron-Frobenius operator we develop an efficient computation approach for calculating the stationary distribution for the uncertain dynamical systems and the average performance criteria. The proposed approach is demonstrated through numerical examples. The second problem is a novel approach for evaluating the short-term Loss of Load Probability (LOLP) in power systems that include wind generation resources that vary stochastically in time. We firstly introduce a mathematical model for calculating the short-term LOLP, and then a novel quantitative measure of its behavior when converging to its steady-state level is derived. In addition, the corresponding empirical formulas are offered which can be used in practice to estimate the convergence time of LOLP under different conditions. Finally, an application of the outcomes of the analytical work in estimation of the dynamic behavior of short-term LOLP with an actual wind generation profile is presented to show the significance of the developed measures.

Part I

Part I: Optimal Design of Uncertain Dynamical Systems

CHAPTER 1

Introduction

Demands for increased functionality, improved performance, increased efficiency and better utilization of resources frequently results in engineered systems that are large, distributed and highly interconnected. Typical systems are built as complex physical and functional interconnection of components that are sophisticated systems themselves that have been designed and optimized for performance and cost in isolation. The system components frequently have complex nonlinear behavior that when combined with the spatial distribution of components, large separation of time scales and high connectivity results in a system that can have very complex dynamic behavior.

The use of physics based models in the design of complex systems is becoming a part of standard engineering design process [3]. The value of model-based design includes the ability to quickly evaluate a large number of design alternatives, optimize system design and evaluate transient behavior. Additional benefits include early detection of design flaws, prevention of component damage and decreased hardware commissioning time.

However, all models in applications have some degree of uncertainty. This uncertainty can be attributed to errors in the physical description, due to either unknown or ignored phenomena, parametric uncertainty or unknown external conditions. In general, we can classify uncertainty into two classes, parametric uncertainty and model uncertainty. Parametric uncertainty is uncertainty in the parameters and initial/boundary conditions of a model of a fixed structure whereas model uncertainty is uncertainty in the structure of the model itself. The former type of uncertainty typically arises due to imperfect measurements or information about the physically parameters of a system. Model uncertainty is

usually either due to unknown or unmodeled behavior or approximation error. These uncertainties call for improved design methodologies that can account for the complexities and uncertainties in the model descriptions in the design process.

Many pessimistic results on the complexity-theoretic barriers of classical robust control problems have stimulated research in the direction of finding alternative solution [9]. One effective solution is to first shift the meaning of robustness from its usual deterministic sense to a probabilistic one. This shift in meaning implies a statistical description of the uncertainty, opposed to a purely unknown-but-bounded one. It has been shown that the probabilistic approach presents itself as a natural tool to deal with the random character of uncertainties affecting control systems [10]. In this work we consider the problem of optimal design of uncertain systems by a probabilistic robust design approach. In particular we assume that the system is described by a discrete time dynamical system

$$x_{i+1} = T(x_i, \theta, \xi)$$
$$y_i = f(x_i)$$

where ξ represents the uncertainty in the system which is assumed to be a random parameter with distribution ρ and θ is a vector of design parameters that can either be physical parameters and/or controller parameters.

Frequently the design objective is to minimize a cost function $g(\bar{y})$ over all admissible $\theta \in \Theta$, where $\bar{y}$ is the stationary output for the system (i.e. ignoring the effect of dynamics, see e.g. [60]). In this case the resulting optimization problem is

$$\min_{\theta \in \Theta} g(\bar{y}) \quad \text{s.t. } \bar{x} = T(\bar{x}, \theta, \xi),\ \bar{y} = f(\bar{x}) \ .$$

If the parameter ξ is random the steady state value $\bar{x}$ is random as well and consequently the objective function $g(\bar{y})$ is random. Following [48], [71] an average performance is used as the objective function to be minimized. Thus, we replace $g(\bar{y})$ with the average

cost resulting in the optimal design problem

$$\min_{\theta\in\Theta} \mathbb{E}^{\rho}\left[g(\bar{y})\right] \quad \text{s.t. } \bar{x} = T(\bar{x},\theta,\xi),\ \bar{y} = f(\bar{x})$$

where $\mathbb{E}^{\rho}[\cdot]$ is the expectation with respect to the distribution ρ.

The objective of this method is to select the design parameters so as to minimize the cost function based on the a-priori uncertainty information due to parameter uncertainty. Obviously, the above design methodology only makes sense when the system converges to the equilibrium $\bar{x} = \bar{x}(\theta,\xi)$ for all initial conditions. Indeed, if for some values of ξ and θ the equilibrium is unstable then the system will never reach the steady state and consequently the steady state optimization problem does not make sense. Furthermore, for some complex dynamical systems that have uncertain behavior generated by the dynamics of the system itself, e.g. chaotic motion, the above method also fails. Therefore, there is a strong need for efficient alternatives. In this dissertation, we propose an alternative optimal design approach by analysis of asymptotic dynamics of uncertain systems.

The dissertation is organized as follow: In Chapter 2 we introduce the mathematical setup that forms the basis for our proposed research. In Chapter 3 we formulate the optimal design of the uncertain dynamical systems. In Chapter 4 we introduce randomized algorithms for average performance synthesis in the design of uncertain systems. In Chapter 5 we solve the optimal design problem numerically. In Chapter 6 we review the Monte Carlo method as a traditional method for calculating the invariant measure. In Chapter 7 we discuss an efficient computational approach for the calculation of the invariant measure. In Chapter 8 we present illustrative examples for the proposed methods to calculate the invariant measure and optimal design of uncertain dynamical systems. In Chapter 9 we conclude the Part I of my dissertation. In Chapter 10 a study of short-term reliability analysis of power systems with wind generation is presented as Part II of my dissertation.

CHAPTER 2

Mathematical Setup

In this chapter, we introduce the basic mathematical concepts necessary for formulation and solution of the uncertain system design problem. These concepts may be studied in detail before continuing on to the core of our subject matter, which starts in Chapter 3, or, they may be skimmed on first reading to fix the location of important concepts for later reference.

2.1 Measure Theory

In this section, we briefly outline some essential concepts from measure theory applied to dynamical systems [39]. This material is in no sense exhaustive; more detailed treatments can be found in [30] and [56].

Definition 1 *A collection $\mathscr{B}$ of subsets of a set X is a $\sigma-$algebra if:*

(a) When $A \in \mathscr{B}$ then $X \setminus A \in \mathscr{B}$;

(b) For any finite or infinite sequence $\{A_k\}$ of subsets, $A_k \in \mathscr{B}$, the countable union $\cup_k A_k \in \mathscr{B}$;

(c) $X \in \mathscr{B}$.

From this definition it follows immediately, by properties (a) and (c), that the empty set $\emptyset$ belongs to $\mathscr{B}$ since $\emptyset = X \setminus X$. Further, given a sequence $\{A_k\}$, $A_k \in \mathscr{B}$, then the intersection $\cap_k A_k \in \mathscr{B}$, because

$$\cap_k A_k = X \setminus \cup_k (X \setminus A_k)$$

and then apply properties (a) and (b). Finally, the difference $A \setminus B$ of two sets A and B that belong to $\mathscr{B}$ also belongs to $\mathscr{B}$ because

$$A \setminus B = A \cap (X \setminus B).$$

If X is any metric space, or more generally any topological space, the $\sigma-$algebra generated by the family of open sets in X is called the *Borel* $\sigma-$*algebra* on X. Its members are called Borel sets. Borel $\sigma-$algebra includes open sets, closed sets, countable intersections of open sets, countable unions of closed sets, and so forth. It is the smallest $\sigma-$algebra containing intervals in X.

Definition 2 *A real valued function* μ *defined on a* $\sigma-$*algebra* $\mathscr{B}$ *is a measure if:*

(a) $\mu(\emptyset) = 0$;

(b) $\mu(A) \geq 0$ *for all* $A \in \mathscr{B}$;

(c) $\mu(\cup_k A_k) = \sum_k \mu(A_k)$ *if* $\{A_k\}$ *is a sequence of disjoint sets from* $\mathscr{B}$*, that is* $A_i \cap A_j = \emptyset$ *for* $i \neq j$.

This definition of a measure and the properties of a $\sigma-$algebra ensure that if we know the measure of a set X and a subset A of X we can determine the measure of $X \setminus A$; and if we know the measure of a collection disjoint subset A_k of $\mathscr{B}$ we can calculate the measure of their unions.

If $X = \mathbb{R}$, there is a large family of measures whose domain is the Borel $\sigma-$algebra and such measures are called *Borel measures.* The *Lebesgue measure* on $\mathbb{R}$ is the measure for which the measure of an interval is its length. In particular, the Lebesgue measure is usually denoted as m, and $m([a,b]) = b - a$. The Lebesgue measure is a completed Borel measure. The definition of a completion of a measure can be found in [24]. Whenever considering spaces $X = \mathbb{R}$ or $X = \mathbb{R}^n$ or subsets of these we always assume measures to be Borel measures.

Another important class of measures are supported on finite or countable sets. The simplest of these measures is the *Dirac measure* defined by

$$\delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Definition 3 *If $\mathscr{B}$ is a $\sigma-$algebra of subsets of X and if μ is a measure on X, then the triple $(X, \mathscr{B}, \mu)$ is called a measure space. The sets belonging to $\mathscr{B}$ are called measurable sets.*

Definition 4 *Let $(X, \mathscr{B}, \mu)$ be a measure space. A real-valued function $f : X \to \mathbb{R}$ is called measurable on X if $f^{-1}(\Delta) \in \mathscr{B}$ for every interval $\Delta \subset \mathbb{R}$. In particular, if $X = \mathbb{R}$, then the function f is called Lebesgue measurable.*

Definition 5 *A measure space $(X, \mathscr{B}, \mu)$ is called finite if $\mu(X) < \infty$. In particular, if $\mu(X) = 1$, then the measure space is said to be probabilistic and such a measure μ is called probability measure.*

In a probabilistic measure space, X is to be considered as the set of all possible outcomes of some process, such as an experiment or a gambling game, and a measure of a set $E \in \mathscr{B}$ is interpreted as the probability that the outcome lies in E. In the study of dynamical systems, it is a convention to use analysts' terms in measure theory instead of probabilists' terms in probability theory.

If $(X, \mathscr{B}, \mu)$ is a measure space, a set $E \in \mathscr{B}$ such that $\mu(E) = 0$ is called a null set. If a statement about points $x \in X$ is true except for x in some null set, we say that property is true *almost everywhere (a.e.)*.

One of the most important notations in analysis, measure theory, and topology, as well as other areas of mathematics, is that of the Cartesian product. To introduce this concept we start with a definition.

Definition 6 *Given two arbitrary sets A_1 and A_2, the Cartesian product of A_1 and A_2 is the set of all pairs (x_1,x_2) such that $x_1 \in A_1$ and $x_2 \in A_2$. This is written as*

$$A_1 \times A_2 = \{(x_1,x_2) : x_1 \in A_1, x_2 \in A_2\}.$$

In a natural way this concept may be extended to more than two sets. Thus the Cartesian product of the sets $A_1, ..., A_d$ is the set of all sequences $(x_1, ..., x_d)$ such that $x_i \in A_i, i = 1, ..., d$, or

$$A_1 \times \cdots \times A_d = \{(x_1, \cdots, x_2) : x_i \in A_i, \qquad \text{for } i = 1, ..., d\}.$$

An important consequence following the concept of the Cartesian product is to extend the property of measure space to the Cartesian product. Thus, given d measure spaces $(X_i, \mathscr{B}_i, \mu_i)$, $i = 1, ..., d$, we define

$$X = X_1 \times \cdots \times X_d, \tag{2.1}$$

$\mathscr{B}$ to be the smallest $\sigma-$algebra of subsets of X containing all sets of the form

$$A_1 \times \cdots \times A_d \quad \text{ with } A_i \in \mathscr{B}_i,\ i = 1, ..., d \tag{2.2}$$

and

$$\mu(A_1 \times \cdots \times A_d) = \mu_1(A_1) \cdots \mu_d(A_d) \quad \text{ for } A_i \in \mathscr{B}_i. \tag{2.3}$$

Theorem 1 *If measure spaces $(X_i, \mathscr{B}_i, \mu_i)$, $i = 1, ..., d$ are given and X, $\mathscr{B}$, and μ are defined by equations (2.1), (2.2) and (2.3), respectively, then there exists a unique extension of μ to a measure defined on $\mathscr{B}$.*

The measure space $(X, \mathscr{B}, \mu)$ whose existence is guaranteed by Theorem 1 is called the *product space* and the measure is called the *product measure.* Moreover, if all the measure space $(X_i, \mathscr{B}_i, \mu_i)$ are finite or probabilistic, then $(X, \mathscr{B}, \mu)$ is also finite or probabilistic.

Finally, we conclude this section with the definition of absolutely continuity of a measure.

Definition 7 *Assume two measures ν and μ are defined on $(X, \mathscr{B})$. Then we say that ν is absolutely continuous with respect to μ denoted as*

$$\nu \ll \mu$$

if $\nu(E) = 0$ for every $E \in \mathscr{B}$ for which $\mu(E) = 0$.

Theorem 2 ***(**Radon-Nikodym theorem**)** *Let $(X, \mathscr{B}, \mu)$ be a measure space and let ν be a finite measure with $\nu \ll \mu$. Then there exists a non-negative integrable function $f \in L^1(X, \mu)$ such that*

$$\nu(E) = \int_E f(x)\, \mu(dx)$$

for all $E \in \mathscr{B}$.

In the following sections, we will introduce some notations used in the study of dynamical systems, including both deterministic and uncertain cases.

2.2 Deterministic Dynamical Systems

In this section, some notations used in the study of deterministic dynamical systems will be defined. Consider a deterministic discrete time dynamical system

$$\begin{aligned} x_{i+1} &= T(x_i) \\ y_i &= f(x_i) \end{aligned} \tag{2.4}$$

where $T : X \to X$ is a diffeomorphism on a compact subset $X \subset \mathbb{R}^n$. We denote by $\mathscr{B}$ the Borel $\sigma-$algebra on X and by m the Lebesgue measure on $\mathscr{B}$. Moreover, let $\mathscr{M}$ be the space of all real value measures on $\mathscr{B}$.

Definition 8 *A measure $\mu \in \mathscr{M}$ is said to be an invariant measure of T if*

$$\mu(B) = \mu\left(T^{-1}(B)\right)$$

for all sets $B \in \mathscr{B}$.

Definition 9 *A function $p : X \times \mathscr{B} \to \mathbb{R}$ is a stochastic transition function, if (i) $p(x,\cdot)$ is a probability measure for every $x \in X$. (ii) $p(\cdot,A)$ is Lebesgue-measurable for every $A \in \mathscr{B}$.*

Let δ_y denote the Dirac measure supported at the point $y \in X$. Then $p(x,A) = \delta_{h(x)}(A)$ is a stochastic transition function for every m-measurable function h. We will see below that the specific choice $h = T$ represents the deterministic situation as system (2.4) in this more general set up.

We set $p^{(1)}(x,A) = p(x,A)$ and define recursively the i-step stochastic transition function $p^{(i)} : X \times \mathscr{B} \to \mathbb{R}$ by

$$p^{(i+1)}(x,A) = \int p^{(i)}(y,A)\, p(x,dy), \quad i = 1,2,...$$

It is easy to see that $p^{(i)}$ is indeed a stochastic transition function. In particular, for the case where $p(x,A) = \delta_{T(x)}(A)$ we obtain for $i \geq 1$

$$p^{(i)}(x,A) = \delta_{T^i(x)}(A)$$

Definition 10 *Let p be a stochastic transition function. If $\mu \in \mathscr{M}$ satisfies*

$$\mu(A) = \int p(x,A)\, d\mu$$

for all $A \in \mathscr{B}$, then μ is an invariant measure of p.

Remark 1 *(a) If μ is an invariant measure of p then if follows that*

$$\mu(A) = \int p^{(i)}(x,A)\, d\mu$$

for all $i = 1,2,...$

(b) For system (2.4) the stochastic transition function is chosen as $p(x,\cdot) = \delta_{T(x)}$. If μ be an invariant measure of p, then μ is also an invariant measure of T.

If we assume that for every for $x \in X$ the probability measure $p(x,\cdot)$ is a absolutely continuous with respect to the Lebesgue measure m, then we can write $p(x,\cdot)$ with an appropriate transition density function $k : X \times X \to \mathbb{R}$ by

$$p(x,A) = \int_A k(x,y) dm(y)$$

for all $A \in \mathscr{B}$. In this case, we also call the stochastic transition function p absolutely continuous.

Definition 11 *Let p be a stochastic transition function. Then the Perron-Frobenius (P-F) operator P: $\mathscr{M} \to \mathscr{M}$ is defined by,*

$$P\mu(A) = \int p(x,A)\, d\mu(x)$$

where $\mathscr{M}$ is the space of bounded complex valued measures on $\mathscr{B}$. If p is absolutely continuous with density function k then we can define the Perron-Frobenius operator P on L^1 by,

$$Pg(y) = \int k(x,y)\, g(x) dm(x)$$

for all $g \in L^1$. In particular, for the deterministic dynamical system $T : X \to X$ the P-F

operator is defined by,

$$P\mu(A) = \int \delta_{T(x)}(A) d\mu(x) = \mu(T^{-1}(A))$$

for all sets $A \in \mathscr{B}$.

Remark 2 *(a) By definition a measure* $\mu \in \mathscr{M}$ *is invariant if and only if it is a fixed point of the P-F operator P, that is, for all sets* $A \in \mathscr{B}$,

$$P\mu(A) = \mu(A)$$

(b) The P-F operator characterizes the evolution of the distribution of the state x_i*, i.e. if the initial state has distribution* $\mu_0 \in \mathscr{M}$*, the distribution of* x_i *is* $P^i\mu_0$

(c) Note that in the case when p is absolutely continuous with density function k we have P: $L^1 \to L^1$ *and for each* $g \in L^1$

$$\begin{aligned} \int Pg(y)dm(y) &= \int\int k(x,y)g(x)dm(x)dm(y) \\ &= \int g(x)\int k(x,y)dm(y)dm(x) \\ &= \int g(x)dm(x) < \infty \end{aligned}$$

Since our goal is to characterize the stationary distribution for uncertain systems, the same notations used for deterministic dynamical systems will be extended to random dynamical systems with parametric uncertainty in the next section.

2.3 Random Dynamical Systems

Consider an uncertain dynamical system

$$\begin{aligned} x_{i+1} &= T(x_i, \xi) \\ y_i &= f(x_i) \end{aligned} \tag{2.5}$$

In order to simplify the discussion we assume that X is a compact subset of $\mathbb{R}^n$ and N is a compact subset of $\mathbb{R}^r$. The random parameter ξ can be very general and is specified in more detail later. We assume that $T(x,\xi)$ is $C^r, r \geq 1$ in x for every $\xi \in N$ and assume that $f : X \to \mathbb{R}$ satisfies $f \in L^1(X)$. Denote $T_\xi^i(x) = T_\xi \circ ... \circ T_\xi$ where $T_\xi(x) = T(x,\xi)$. Note that for a given parameter ξ the uncertain dynamical system (2.5) becomes the deterministic case (2.4). Therefore, the notations used in Section 2.2 can be defined in a similar way for uncertain dynamical systems. The discussion that follows is based on some recent results [2], [42], [57], [43] and is not exhaustive.

Definition 12 *Let $\mathscr{M}$ be the vector space of real valued measures on X. For a fixed value of ξ the Perron-Frobenius (P-F) operator $P_\xi : \mathscr{M} \to \mathscr{M}$ corresponding to the dynamical system $T_\xi : X \to X$ is defined as*

$$P_\xi \mu(B) = \int_{T_\xi^{-1}(B)} d\mu = \mu\left(T_\xi^{-1}(B)\right)$$

for all sets $B \in \mathscr{B}$.

Definition 13 *A measure $\mu_\xi \in \mathscr{M}$ is said to be a T_ξ invariant measure if*

$$\mu_\xi(B) = P_\xi \mu_\xi(B) = \mu_\xi\left(T_\xi^{-1}(B)\right)$$

for all sets $B \in \mathscr{B}$.

Thus, $\mu_\xi \in \mathscr{M}$ is invariant if and only if it is a fixed point of the Perron-Frobenius operator P_ξ.

Let $\mathbf{P} = X \times N$ be the state-uncertainty product space and endow it with the product σ-algebra $\mathscr{P}$ in the usual way, i.e. if $\mathscr{B}$ is the Borel $\sigma-$algebra on X and $\mathscr{F}$ is the $\sigma-$algebra on N, then $\mathscr{P} = \mathscr{B} \times \mathscr{F}$.

Definition 14 *A probability measure η on $\mathscr{P}$ is called an input measure.*

Example 1 *Consider the evolution given by*

$$x_{i+1} = x_i + \xi x_i - x_i^3. \tag{2.6}$$

System (2.6) has a bifurcation point at $\xi = 0$. In particular, for $\xi < 0$ there is an unique stable equilibrium at $x = 0$ and for $\xi > 0$ there are three equilibria, two stable at $x = \pm\sqrt{\xi}$ and an unstable one at $x = 0$. The input measure η is given by the product measure $\nu \times \rho$, where $\nu(E) = l(E \cap [-a,a])/l([-a,a])$, l the Lebesgue measure on the real line, $a > 0$. This measure is given by the uniform density function $f_\nu(x) = 1/2a$ for $x \in [-a,a]$ and represents the uncertainty in initial conditions. The measure ρ is a measure on $\mathbb{R}$ defined by $\rho(E) = l(E \cap [-b,b])/l([-b,b])$, $0 < b < 1$. This measure is given by the uniform density function $f_\rho(\xi) = 1/2b$ for $\xi \in [-b,b]$ and is the density for the "parameter input measure".

We are interested in the question of how does the uncertainty in the "output" of the process depend on the input measure. For the observable defined by $f : X \to \mathbb{R}$, the "initial" uncertainty is described by a probabilistic measure ϖ_i on $\mathbb{R}$ (endowed with the Borel σ-algebra $\mathscr{B}$) defined by

$$\varpi_i(E) = \eta((f)^{-1}(E)),$$

where $E \in \mathscr{B}$. This measure evolves in time, becoming

$$\begin{aligned} \varpi^n(E) &= \eta((f \circ T_\xi^n)^{-1}(E)) \\ &= \eta((T_\xi^n)^{-1} f^{-1}(E)) = P_\xi^n \eta(f^{-1}(E)), \end{aligned}$$

We call ϖ^n an *output measure*. It describes the uncertainty of the system output at the n-th step of the process given the input measure η.

Frequently we are mostly interested in the long term behavior of the solution of the system. In this case the uncertainty in the system output is best studied in terms of the uncertainty in the asymptotic properties of the system. In particular, define the time-average

$$f^*(x,\xi) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T(x_i,\xi)) \tag{2.7}$$

and the *asymptotic output measure* ϖ_a

$$\varpi_a(E) = \lim_{n\to\infty} \varpi^n(E) = \eta((f^*)^{-1}(E)). \tag{2.8}$$

Return now back to the uncertain system (2.5) with assumption that $\xi = c$ where c represents a (certain) parameter. In other words, there is no parametric uncertainty and the input measure η is replaced by uncertainty distribution ν for the initial conditions of the states. Then we can rewrite the asymptotic output measure as

$$\varpi_a(E|\xi = c) = \lim_{n\to\infty} P_c^n \nu(f^{-1}(E)) = \mu_c(f^{-1}(E)) \tag{2.9}$$

Therefore, if the random parameters in the dynamical systems are fixed, the asymptotic output measure can be characterized by the invariant measure of the system.

Next we discuss methods for characterizing the asymptotic output (2.7) in terms of the invariant measures of the random dynamical system. In particular, we are interested

in situations when the system has a physical measure μ in the sense that for almost all $(x,\xi) \in X \times N$

$$f^*(x,\xi) = \int_{X\times N} f d\mu \tag{2.10}$$

We begin by rewriting system (2.5) so that it can be studied within the framework of Discrete Random Dynamical System (DRDS) [2]. We are interested in characterizing invariant measures for (2.5). For the purpose of presenting the formulation in [2] full generality we allow the uncertainty to vary as a function of time. Let $\Omega = N^{\mathbb{Z}}$ be the space of all sequences taking values in N, denote an element of Ω as ω and let $\mathscr{F}$ be the Borel $\sigma-$algebra on Ω. Let $S(k)$ be the shift transformation on Ω, i.e. $S(k)\,\omega_i = \omega_{i+k}$, define the map $\pi : \Omega \to N$ by

$$\pi(\omega) = \omega(0)$$

and the coordinate process

$$\xi_i = \omega(i) = \pi(S(i)\,\omega) = (\pi \circ S(i))(\omega).$$

Let $\mathbb{P}$ be any S invariant (probability) measure. Then $\xi_i = \omega(i)$ is a stochastic process on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Now we let $\hat{T}(x,\omega) = T(x,\omega(0))$ and get

$$\begin{aligned} x_{i+1} &= T(x_i,\xi_i) = T(x_i,\omega(i)) = \hat{T}(x_i, S(i)\,\omega) \\ y_i &= f(x_i) \end{aligned} \tag{2.11}$$

Let $\psi(\omega) : X \to X$ be the operator defined by $\psi(\omega)x = \hat{T}(x,\omega)$. Then the solution of (2.11) can be represented as $x_i = \varphi(i,\omega)x$ where

$$\varphi(i,\omega) = \begin{cases} \psi(S_{i-1}\omega) \circ \cdots \circ \psi(\omega) & i \geq 1 \\ \mathrm{id}_X & i = 0 \end{cases}$$

where id_X is the identity operator on X. The mapping

$$(\omega, x) \longmapsto (S_i\omega, \varphi(i,\omega)x) = \Theta(i)(\omega, x)$$

is called the skew product of S_i and $\varphi(i,\cdot)$ and is measurable dynamical system on $(\Omega \times X, \mathscr{F} \times \mathscr{B}_X)$. Let π_Ω denote the projection by $\pi_\Omega(\omega, x) = \omega$.

Definition 15 *A probability measure μ on $(\Omega \times X, \mathscr{F} \times \mathscr{B})$ is said to be an invariant measure of the DRDS defined by (2.11) if it satisfies (i) $\Theta(1)\mu = \mu$ and (ii) $\pi_\Omega \mu = \mathbb{P}$.*

Note that any measure that is invariant with respect to Θ has a marginal $\pi_\Omega\mu$ that is invariant with respect to S. Furthermore, the second condition in the above definition is imposed since the measure is an a-priori specified invariant measure for S. Now let $\mathscr{P}(\Omega \times X)$ be the set of all probability measures on $\Omega \times X$ and define

$$\mathscr{P}_{\mathbb{P}}(\Omega \times M) = \{\nu \in \mathscr{P}(\Omega \times X) : \pi_\Omega \nu = \mathbb{P}\}$$

$$\mathscr{I}_{\mathbb{P}} = \{\mu \in \mathscr{P}_{\mathbb{P}}(\Omega \times X) : \mu \text{ is invariant for (2.11)}\}$$

Assume $\mu \in \mathscr{P}_{\mathbb{P}}(\Omega \times X)$. A function $\mu_\cdot(\cdot) : \Omega \times \mathscr{B} \longrightarrow [0,1]$ is called a *factorization* of μ with respect to $\mathbb{P}$ if

1. for all $A \in \mathscr{B}$, $\mu_\cdot(A)$ is $\mathscr{F}$ measurable,
2. $\mu_\omega(\cdot)$ is a probability measure on $(X, \mathscr{B})$ for $\mathbb{P}$ almost all ω,
3. for all $C \in \mathscr{F} \times \mathscr{B}$ we have

$$\mu(C) = \int_\Omega \int_X \chi_C(\omega, x)\, \mu_\omega(dx)\, \mathbb{P}(d\omega)$$

where χ_C is the indicator function for C.

Note that it follows that for $h \in L^1(\mu)$ we have

$$\int_{\Omega \times X} h d\mu = \int_{\Omega} \int_{X} h(\omega, x) \mu_{\omega}(dx) \mathbb{P}(d\omega)$$

It can be shown that under the assumptions we have made about X the factorization exists and is $\mathbb{P}$ almost surely unique [2].

Return now back to the uncertain system (2.5), i.e. assume that for all i we have $\xi_i = \xi$ where ξ is a random parameter that has distribution ρ on N. Then the invariant measure of $S(1)$ is a (random) measure concentrated at ξ and if μ is a physical measure in the sense (2.10)

$$\begin{aligned} f^*(x, \xi) &= \int_{\Omega \times X} f d\mu = \int_{\Omega} \int_{X} f(z) \mu_{\omega}(dz) \delta(\omega - \xi) d\omega \\ &= \int_{X} f(z) \mu_{\xi}(dz) \end{aligned}$$

We note that the through the factorization of the physical measure the time average $f^*(x, \xi)$ is an explicit function of the random parameter ξ. The dependence on the (random) initial state can be characterized in terms of the ergodic properties of the system.

Definition 16 *Let ρ be an a-priori measure on the Borel σ-algebra on X. System (2.5) is called* B_ρ-regular *if for each fixed ξ there exists a finite set of ergodic invariant measures μ_{ξ}^{i}, $i = 1, \ldots, K_{\xi}$ such that for almost every $x \in X$ and every $f \in C(X)$ there exists a $j \in \{1, \ldots, K_{\xi}\}$ such that the time average satisfies*

$$f^*(x, \xi) = \int_{X} f(z) \mu_{\xi}^{j}(dz)$$

Furthermore, there exists an ergodic partition, i.e. disjoint (random) sets $D_1^{\xi}, \ldots, D_{K_{\xi}}^{\xi}$ such

that $\rho\left(X-\cup_{i=1}^{K_\xi} D_i^\xi\right)=0$ *and*

$$D_i^\xi=\left\{x\in X\,\middle|\, f^*(x,\xi)=\int_X f(z)\,\mu_\xi^i(dz)\ \forall\ f\in C(X)\right\}$$

The following proposition from [43] characterizes the calculation of the asymptotic output measure in terms of the ergodic invariant measure.

Proposition 3 *Assume that (2.5) is* B_ρ-regular *and the family of measures* $\mu_i^\xi, i=1,...,K_\xi$ *has the property that each* $\mu_i^\xi(B)$ *is continuous as a function of* ξ *for any* $B\in\mathscr{B}_X$. *Assume that the input measure* η *is absolutely continuous with respect to the a-priori measure* ρ. *Then the cumulative distribution function* F_{ϖ_a} *for* ϖ_a *is piecewise continuous with a finite number of steps.*

If we define $\mu_\xi(A)=\sum_{i=1}^{K_\xi}\mu_\xi^i\left(A\cap D_i^\xi\right)$ then $\mu_\xi(A)$ is an invariant measure for T_ξ and factorization of the invariant measure μ for (2.5).

The following result proven in [2] provides further insight into the conditions under which the invariant measure of (2.11) is a physical measure in the sense of (2.10).

Definition 17 *The DRDS is said to be continuous if for each fixed* ω *the mapping* $\varphi(\cdot,\omega)\cdot:\mathbb{Z}\times X\to X$ *is continuous.*

We note that by Theorem 1.5.10 in [2] if X is a compact metric space and the DRDS is continuous then $\mathscr{I}_\mathbb{P}$ is a non-empty convex compact subset of $\mathscr{P}_\mathbb{P}(\Omega\times X)$. For a measure $\nu\in\mathscr{P}(\Omega\times X)$ and $f\in L_1(\nu)$ define $\nu(f)=\int f d\nu$.

Theorem 4 *Assume that DRDS is continuous. For* $\nu\in\mathscr{P}_\mathbb{P}(\Omega\times X)$ *define*

$$\mu_N(f)=\frac{1}{N}\sum_{n=0}^{N-1}\Theta(n)\,\nu(f)=\frac{1}{N}\sum_{n=0}^{N-1}\nu(f\circ\Theta(n))$$

Then as $N\to\infty$ *every limit point of* μ_N *converges weakly to some* $\mu\in\mathscr{I}_\mathbb{P}$ *and every* $\mu\in\mathscr{I}_\mathbb{P}$ *arises in this way.*

Finally we have the following general result taken from [2] that further characterizes the factorization of the invariant measure.

Theorem 5 *Let $\mu \in \mathscr{P}_{\mathbb{P}}(\Omega \times X)$. Then (i) $\mu \in \mathscr{I}_{\mathbb{P}}$ if and only if for all $i \in \mathbb{N}$*

$$E\left((\varphi(i,\cdot)\,\mu_{\cdot} \mid S(i)^{-1}\mathscr{F}\right)_{\omega} = \mu_{S(i)\omega} \qquad \mathbb{P}-a.s.$$

(ii) If S is measurably invertible then $S(i)^{-1}\mathscr{F} = \mathscr{F}$ for all $i \in \mathbb{Z}$, and $\mu \in \mathscr{I}_{\mathbb{P}}$ if and only if for all $i \in \mathbb{Z}$

$$\varphi(i,\omega)\mu_{\omega} = \mu_{S(i)\omega} \qquad \mathbb{P}-a.s.$$

Consider now the special case of an uncertain system (2.5), i.e. $\xi_i = \xi$ for all i where ξ has the distribution ρ on N. In this case, if $\mu \in \mathscr{I}_{\mathbb{P}}$ then $\varphi(i,\xi)\mu_{\xi} = \mu_{\xi}$. Thus, since $\varphi(1,\xi)x = T_{\xi}(x)$, for each fixed value of $\xi \in N$ the marginal $\mu_{\xi} = \pi_X \mu$ is the invariant measure of (2.5) on $(X, \mathscr{B}_X)$. Furthermore, if the dynamical system T_{ξ} is ergodic for each $\xi \in N$, then the φ-invariant measure μ_{ξ} is called a random Dirac measure, i.e. there exists a map $h : N \to X$ with $\mu_{\xi} = \delta_{h(\xi)}$.

In the following chapters, we will present the main contribution of the paper, i.e. an efficient optimal design method of uncertain dynamical systems.

CHAPTER 3

Uncertain System Design

The use of mathematical models in the design of engineering products and systems has become widespread in the last few decades. Depending on the complexity of the product or process the models range from simple physics based and empirical models to large computational models. Most of the models used in design are algebraic models representing the steady state or static relationship between variables. In [1], the quantities of interest in a typical design problem are the following:

- Design Parameter [DP]: Any free or independent parameter whose value is determined during the design process.
- Performance Parameter [PP]: Any parameter used in the design process that has a specified value [FR] determined independent of (and usually in advance of) the design process. The performance parameters [PPs] are usually dependent on the design parameters [DPs], and possibly some other PPs.
- Output Parameter [OP]: Any parameter used in the design process that is dependent on the design parameters [DPs], and possibly some performance parameters [PPs], but has no specified functional requirement [FR] value.
- Functional Requirement [FR]: A value, or range of values, or fuzzy number that is the specified value for a Performance Parameter [PP]. This value is determined independent of (and usually in advance of) the design process. Each Performance Parameter has a FR.

- Performance Parameter Expression [PPE]: An expression, relationship, or equation relating some or all of the Design Parameters to a Performance Parameter. Each PP has a PPE.

In terms of a mathematical design process the problem is formulated as an optimization problem where the objective is to optimize over the Design Parameters the deviation between Performance Parameters and their corresponding Functional Requirements subject to constraints set by the Performance Parameter Expressions. Let y represent the Performance Parameters, d denote the Design Parameter, $P(y,d) = 0$ represent the Performance Parameter Expression and $g(y)$ be a cost function or performance function that represents the level to which the Performance Parameters meet the Functional Requirements. Then the optimal design problem becomes the optimization of $g(y)$ subject to the constraint $P(y,d) = 0$.

In many practical problems the Performance Parameter Expression is subject to uncertainties that can be attributed to inaccuracies in physical parameters, lack of knowledge about physical characteristics as well as modeling inaccuracies and approximation errors. Frequently these uncertainties are modeled as random parameters with a known distribution and incorporated into the Performance Parameter Expression [48].

In some design problems the Performance Parameter Expression is a true static relationship. This is in particular true for many component design problems. In other problems the system is truly dynamic and the effect of the dynamics cannot be ignored. This is the case for many system design problems that can exhibit complex nonlinear behavior. In this section we discuss the incorporation of uncertainty into the design specifications in such problems. In particular, we consider the uncertain discrete dynamical system (2.5) but assume that the system depends on a design parameter θ. Thus the system is now

$$x_{i+1} = T(x_i, \theta, \xi) \tag{3.1}$$
$$y_i = f(x_i)$$

where $x_i \in G \subset \mathbb{R}^n$ is the system state, $\theta \in \Theta$ is the design parameter, and the uncertainty $\xi \in H$ is as before a random parameter with distribution ρ. In order to simplify the discussion we assume that G is a compact subset of $\mathbb{R}^n$ and H is a compact subset of $\mathbb{R}^r$. Also, we assume that the system is stable for all initial states, all $\theta \in \Theta$ and all parameters $\xi \in H$. Of particular interest is the case when the parameters θ arise from a parameterization of controllers. In this case we assume we start with the controlled system with control input u_i. Thus the system is now

$$\begin{aligned} x_{i+1} &= \hat{T}(x_i, u_i, \xi) \\ y_i &= f(x_i) \end{aligned} \tag{3.2}$$

Furthermore, we assume the control input u_i has the form $u_i = u(x_i, \theta)$ where the control parameter $\theta \in \tilde{\Theta}$. We denote the class of all such control laws by $\widetilde{\mathscr{U}}(\theta)$. We assume that there exists a nontrivial compact subset $\Theta \subset \tilde{\Theta}$ and a subset $\mathscr{U}(\theta) \subseteq \widetilde{\mathscr{U}}(\theta)$ that has the property that any control law $u \in \mathscr{U}(\theta)$ with $\theta \in \Theta$ stabilizes the closed loop system for all initial states and all parameters $\xi \in H$. In other words, we focus our attention on stabilizing state feedback control laws. For $u \in \mathscr{U}(\theta)$, $\theta \in \Theta$ the dynamical system equation can be rewritten as

$$\begin{aligned} x_{i+1} &= \hat{T}(x_i, u(x_i, \theta), \xi) = T(x_i, \theta, \xi) \\ y_i &= f(x_i) \end{aligned} \tag{3.3}$$

Assume that the objective of the design for system (3.3) is to minimize the performance function $g(y)$ over all admissible $\theta \in \Theta$, where y is the Performance Parameter for the system. In the design literature y is frequently evaluated at an equilibrium point of the system. In that case the resulting optimization problem becomes

$$\min_{\theta \in \Theta} g(\bar{y}) \quad \text{s.t. } \bar{x} = T(\bar{x}, \theta, \xi), \ \bar{y} = f(\bar{x})$$

where $\bar{y}$ is the stationary output of the system. Since the parameter ξ is random the steady state value $\bar{x}$ is random as well and consequently the objective function $g(\bar{y})$ is random. Following [48], [71] we use an average performance as the objective function to be minimized. Thus, we replace $g(\bar{y})$ with the average cost resulting in the optimal design problem

$$\min_{\theta \in \Theta} \mathbb{E}^{\rho}\left[g(\bar{y})\right] \quad \text{s.t. } \bar{x} = T(\bar{x}, \theta, \xi), \ \bar{y} = f(\bar{x}) \tag{3.4}$$

where $\mathbb{E}^{\rho}[\cdot]$ is the expectation with respect to the distribution ρ. If we define a a-priori equilibrium measure ϖ_s^{θ} by

$$\varpi_s^{\theta}(E) = \rho\left(\{\xi \in H | f(x) \in E, x = T(x, \theta, \xi)\}\right) \tag{3.5}$$

then (3.4) can be rewrite as

$$\min_{\theta \in \Theta} \int_{\mathbb{R}} g(z)\, \varpi_s^{\theta}(dz) \tag{3.6}$$

In the remainder of the paper we refer to (3.6) as the *a-priori design criteria*.

For a dynamical system evolving from an (unknown) initial condition the above formulation only makes sense if the solution $\bar{x}$ of the constraint equations corresponds to an attractor of the dynamical system. Indeed, if for some values of ξ and θ the equilibrium is unstable then the system will never reach the steady state and consequently the steady state optimization problem does not make sense. Furthermore, the system may exhibit uncertain behavior that is "generated" by the dynamics of the system itself. All uncertainty effects can be correctly accounted for by reformulating the design problem utilizing the asymptotic output measure corresponding to the time-averages of the output y. Indeed, formulating the optimal design problem as

$$\min_{\theta \in \Theta} \int_{\mathbb{R}} g(z)\, \varpi_a^{\theta}(dz) \tag{3.7}$$

captures all uncertainty effects through the definition of *asymptotic output measure* $\varpi_a^\theta(E) = \eta\left(\left(f_\theta^*\right)^{-1}(E)\right)$ where

$$f_\theta^*(x,\xi) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(T(x_i,\theta,\xi)) \tag{3.8}$$

and η is the input measure. We call (3.7) the *a-posteriori design criteria*.

It is of interest to compare the designs obtained from the above two formulations for the design optimization. In the former one (3.6) the objective is to design a controller that minimizes the cost function based on the a-priori uncertainty information due to initial conditions and parameter uncertainty, while the latter one (3.7) incorporates a-posteriori uncertainty induced by the system dynamics [57]. We call the former design method a-priori design and the latter one a-posteriori design. The difference between the two design measures is illustrated in the following simple example.

Example 2 *Consider the system*

$$x_{i+1} = x_i + \xi(x_i-\theta) - (x_i-\theta)^3 \tag{3.9}$$

where $\theta \in [-a,a]$ is a design parameter and $\xi \in [-b,b]$ is an uncertain parameter, assumed to uniformly distributed on $[-b,b]$. The initial state is assumed to be uniformly distributed on $[-a,a]$. System (3.9) has equilibrium at $\bar{x} = \theta,\ \theta \pm \sqrt{\xi}$. The cumulative distribution function of the output measure is given by

$$F_a(z,\theta) = \begin{cases} 0, & z \le -\sqrt{b}+\theta \\ \frac{(a+\theta)\left(b-(z-\theta)^2\right)}{4ab}, & -\sqrt{b}+\theta < z \le \theta \\ \frac{3a+\theta}{4a} + \frac{(a-\theta)(z-\theta)^2}{4ab}, & \theta < z \le \sqrt{b}+\theta \\ 1, & z > \sqrt{b}+\theta \end{cases}$$

and it is easy to see that $F_a(z,\theta)$ has density

$$f_a(z,\theta) = \frac{1}{2}\delta(z-\theta) - \frac{(a+\theta)(z-\theta)}{2ab}\chi_{\{z|-\sqrt{b}+\theta<z<\theta\}} + \frac{(a-\theta)(z-\theta)}{2ab}\chi_{\{z|\theta<z\leq\sqrt{b}+\theta\}}$$

On the other hand the equilibrium measure (3.5) has density

$$\begin{aligned} f_s(z,\theta) &= -\frac{(z-\theta)}{b}\chi_{\{z|-\sqrt{b}+\theta<z<\theta\}} \\ &\quad + \frac{(z-\theta)}{b}\chi_{\{z|\theta<z\leq\sqrt{b}+\theta\}} \end{aligned}$$

Comparing the two measures shows that they will result in quite different cost functions. Furthermore, the dependence of the density f_a for the average output measure on the initial conditions (reflected in the appearance of a) and the presence of the term $\frac{1}{2}\delta(z-\theta)$ shows shows the importance of the initial condition and the dynamic evolution of the system.

We remark that both (3.6) and (3.7) are in general difficult to compute analytically, since the a-priori equilibrium measure ϖ_s^θ and asymptotic output measure ϖ_a^θ cannot generally be characterized analytically for complex systems. Consequently, since it is difficult to derive the performance function analytically it is even harder to obtain the optimal parameters that minimize the performance functions. Therefore, we need an effective computational method for the optimal design of the uncertain system. In this dissertation research, we have developed an efficient approach called *randomized algorithms* that is based on statistical learning theory to solve this problem.

3.1 Randomized Algorithms for Average Performance Synthesis

In recent years, probabilistic and randomized methods have been developed as effective tools to deal with uncertain complex systems, see e.g. the book [63] and the papers [12], [64]. The starting point of these methods is the assumption that the uncertainty affecting the system is random. The objective is then to provide probabilistic assessments on the system characteristics. More precisely, we say that a certain performance level of the system is robustly satisfied in a probabilistic sense if it is guaranteed against most, albeit not all, possible uncertainty outcomes. In other words, we accept the risk that a given system property is violated for a set of uncertainties having small probability. Such systems may be viewed as being "practically robust" from an engineering point of view.

One of the advantages of this approach for control purposes is to provide a reapproachement between the stochastic and the robust paradigms, utilizing classical worst-case bounds of robust control together with probabilistic information, which is often neglected in a deterministic context. The interplay of probability and robustness also leads to innovative concepts such as the probabilistic robustness margin and the probability degradation function. However, it should be noted that moving from deterministic to probabilistic robustness does not imply a simplification of the problem. Indeed, assessing probabilistic robustness may be even computationally harder than establishing robustness in the usual deterministic sense, since it requires the evaluation of multidimensional probability integrals, which can be computed exactly only in very special cases of limited practical interest.

This computational problem is often resolved by means of randomization techniques, which have been used extensively in various branches of science and engineering to tackle difficult problems that are too hard to be treated via exact deterministic methods. Specific examples include the Monte Carlo methods in computational physics, simulations and financial risk analysis, and the Las Vegas techniques in computer science [64]. In the context of systems and control, the key idea is to introduce "uncertainty randomization," and

this requires the development of specific techniques for generating random samples of the structured uncertainty acting on the system [63]. The probability of performance is estimated using a finite number of random samples, and tail inequalities are used to bound the estimation error. Since the estimated probability is itself a random quantity, this method always entails a certain risk of failure, i.e. there exists a nonzero probability of making an erroneous estimation. The resulting algorithms, called *randomized algorithms (RA),* often have low complexity and the resulting robustness bounds are generally less conservative than classical ones [11].

The randomized algorithms used for the probabilistic approach are particularly effective in the context of control design where the system is not fixed a a-priori, but depends on some controller parameters which should be determined. Specifically, a RA for average performance synthesis is an algorithm that returns a parameter vector which guarantees the average performance is minimized with an a-priori specified accuracy characterized by probabilities. In other words, this probabilistic approach aims at designing a controller that performs well on average. Let $\xi \in H$ represent the random uncertainty acting on the system and let ρ be the distribution of ξ. The design parameter is a vector $\theta \in \Theta$, where $\Theta \subseteq \mathbb{R}^{n_\theta}$ is the set of allowable design parameters. Furthermore, let $J : H \times \Theta \to [0,1]$ be a performance function for the uncertain system. The following definition of RA for average performance synthesis is from [63].

Definition 18 *Let $\varepsilon \in (0,1)$, $\delta \in (0,1)$ be given probability levels. Let*

$$\phi(\theta) = \int_H J(\xi,\theta)\rho\,(d\xi) = \mathbb{E}^\rho[J(\xi,\theta)] \tag{3.10}$$

denote the average performance cost (with respect to ρ) of the uncertain system, and

$$\phi^* = \min_{\theta\in\Theta} \phi(\theta)$$

denote the optimal achievable average performance. A randomized average synthesis al-

gorithm should return with probability $1-\delta$ *a design vector* $\hat{\theta}_N \in \Theta$, *such that*

$$\phi(\hat{\theta}_N) - \phi^* \leq \varepsilon$$

The parameter $\hat{\theta}_N$ *is constructed based on a finite number N of random samples of* ξ .

The RA for average performance synthesis design procedure follows two steps: In the first step, the average performance cost $\phi(\theta) = \mathbb{E}^{\rho}[J(\xi,\theta)]$ is approximated, and in the second step a minimization is performed on the approximate cost to obtain the "optimal" solution. In the first step of the procedure, since the exact computation of the expected value is, in general, computationally difficult, we approximate the expectation $\phi(\theta)$ by its empirical version. That is, N samples $\xi_1, ..., \xi_N$ are collected and the empirical mean is obtained as

$$\hat{\phi}_N(\theta) = \hat{\mathbb{E}}_N[J(\xi,\theta)] = \frac{1}{N}\sum_{i=1}^{N} J(\xi_i,\theta).$$

Obviously, the approximation error will be affected by sample size N. From Statistical Learning Theory in [54], [69], [72], we have the following result for sample size bound.

Theorem 6 *Given* $\varepsilon_1, \delta_1 \in (0,1)$, *let*

$$N \geq \frac{32}{\varepsilon_1^2}\left[\ln\frac{8}{\delta_1} + d\left(\ln\frac{16e}{\varepsilon_1}\ln\ln\frac{16e}{\varepsilon_1}\right)\right] \tag{3.11}$$

where d is an upper bound on the so-called Pollard (or P) dimension of the function family $\{J(\cdot,\theta), \theta \in \Theta\}$, *then, it can be asserted with confidence* $1-\delta_1$ *that*

$$\left|\phi(\theta) - \hat{\phi}_N(\theta)\right| = \left|\mathbb{E}^{\rho}[J(\xi,\theta)] - \hat{\mathbb{E}}_N[J(\xi,\theta)]\right| \leq \varepsilon_1 \tag{3.12}$$

for all $\theta \in \Theta$.

In the second step, we have to find the solution $\hat{\theta}_N \in \Theta$ such that minimize the approx-

imate expectation $\hat{\phi}_N(\theta)$, that is

$$\hat{\theta}_N = \arg \inf_{\theta \in \Theta} \hat{\phi}_N(\theta).$$

In principle the minimization of $\hat{\phi}_N(\theta)$ over $\theta \in \Theta$ can be performed by any numerical optimization method. However, if the performance function $J(\xi, \theta)$ is non-convex in θ for some fixed ξ, then the empirical mean $\hat{\phi}_N(\theta)$ is also non-convex, and there are obvious difficulties in finding a global minimum. Thus, a viable approach would be to use a randomized algorithm also for the determination of a probable minimum of $\hat{\phi}_N(\theta)$. In order to apply a randomized algorithm to find the optimal solution, we introduce an artificial probability distribution π over the set Θ of design parameters (sometimes based on some a-priori knowledge about the parameters). Note that the sampling distribution π is often chosen to be the uniform distribution on Θ if we have no a-priori knowledge about the parameters. Correspondingly, the performance function $\hat{\phi}_N(\theta_k)$ is evaluated for each sample θ_k, $k = 1, ..., M$ and the empirical optimal solution is obtained as

$$\hat{\theta}_{NM} = \arg\min_{k=1,...,M} \hat{\phi}_N(\theta_k). \tag{3.13}$$

This randomized algorithm to find the approximate optimal solution is also called a random search algorithm.

A simple randomized algorithm to search the optimal solution is a so-called pure random search (PRS) that generates a sequence of independent, identically distributed sample points in the feasible region Θ, and the objective function has no impact on the technique of generating the next sample points. However, it has been shown that PRS is inefficient when the search region is high dimensional since the expected number of sample points in PRS is exponential in dimensionality. In [61], a more efficient search procedure is developed called adaptive partitioned random search (APRS) which is shown to be at least hundreds

of times more efficient compared with PRS. In this dissertation, we still use PRS to present our randomized algorithms for simplicity. Furthermore, the following theorem in [35], [62] gives us an estimation of sample size M required for a specific approximation error .

Theorem 7 *For any* $\varepsilon_2 \in (0,1)$ *and* $\delta_2 \in (0,1)$, *if*

$$M \geq \frac{\ln \frac{1}{\delta_2}}{\ln \frac{1}{1-\varepsilon_2}}$$

then, with probability greater than $1-\delta_2$, *we have*

$$\Pr\left\{\hat{\phi}_N(\theta) \leq \hat{\phi}_N(\hat{\theta}_{NM})\right\} \leq \varepsilon_2 \tag{3.14}$$

Finally, combining two steps randomized algorithm presented in Theorem 6 and Theorem 7, we obtain the following corollary.

Corollary 8 *Given* $\varepsilon_{1,}\ \varepsilon_2, \delta \in (0,1)$, *let*

$$M \geq \frac{\ln \frac{2}{\delta}}{\ln \frac{1}{1-\varepsilon_2}}$$

and

$$N \geq \frac{32}{\varepsilon_1^2}\left[\ln\frac{16}{\delta} + d\left(\ln\frac{16e}{\varepsilon_1}\ln\ln\frac{16e}{\varepsilon_1}\right)\right] \tag{3.15}$$

Then, with confidence $1-\delta$, *it holds that*

$$\Pr\left\{\phi(\theta) < \hat{\phi}_N(\hat{\theta}_{NM}) - \varepsilon_1\right\} \leq \varepsilon_2.$$

The quantity $\hat{\theta}_{NM}$ in (3.13) is called in [70] a probably approximate near minimizer of $\phi(\theta)$. In the same paper, it is observed that the sample size in (3.11) can be reduced if we use a randomized approach to find the optimal parameters since it is not strictly necessary to guarantee uniformity of the bound (9.6) for all values of $\theta \in \Theta$. Actually, it

is sufficient to require that the inequality (9.6) holds for $\theta \in \{\theta_1, ..., \theta_M\}$. This amounts to guaranteeing the convergence of the empirical estimate uniformly with respect to a family of performance functions having finite cardinality M. It follows that the bound (9.6) is satisfied for $\theta \in \{\theta_1, ..., \theta_M\}$ if we draw at least

$$N \geq \frac{\ln \frac{2M}{\delta_1}}{2\varepsilon_1^2}$$

samples of ξ. This leads to the following theorem, also stated in [70].

Theorem 9 *Given* $\varepsilon_1, \varepsilon_2, \delta \in (0,1)$, *let*

$$M \geq \frac{\ln \frac{2}{\delta}}{\ln \frac{1}{1-\varepsilon_2}} \tag{3.16}$$

and

$$N \geq \frac{\ln \frac{4M}{\delta}}{2\varepsilon_1^2} \tag{3.17}$$

Then, with confidence $1-\delta$, *it holds that*

$$\Pr\left\{\phi(\theta) < \hat{\phi}_N(\hat{\theta}_{NM}) - \varepsilon_1\right\} \leq \varepsilon_2. \tag{3.18}$$

It is easy to verify that the bounds required by (3.17) are substantially better than is required by (3.15). For example, taking $\delta = \varepsilon_1 = \varepsilon_2 = 0.01$, Theorem 9 leads to $M \geq 528$ and $N \geq 61,295$, whereas Corollary 8, with $d = 1$, would require $N \geq 4,248,297$. Now, we present the randomized algorithms for average performance synthesis based on Theorem 9 as follows [63]:

Algorithm 1: (RA for average performance synthesis) Let $J : H \times \Theta \to [0,1]$. Assume θ and ξ are random with distribution π and ρ. Given $\varepsilon_1, \varepsilon_2, \delta \in (0,1)$, this RA returns with probability at least $1-\delta$ a design vector $\hat{\theta}_{NM}$ such that (3.18) holds. The steps of the algorithm are:

1) Determine $M = \overline{M}(\varepsilon_2, \delta)$ and $N = \overline{N}(\varepsilon_1, \delta, M)$ according to (3.16) and (3.17);

2) Draw M iid samples $\theta_{1,\ldots,}, \theta_M$ from π;

3) Draw N iid samples $\xi_{1,\ldots,}\xi_N$ from ρ;

4) Return the empirical optimal parameter:

$$\overset{\wedge}{\theta}_{NM} = \underset{k=1,\ldots,M}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} J(\xi_i, \theta_k). \tag{3.19}$$

Obviously, the computational burden of a randomized algorithm is strictly related to its sample complexity. One of the main critical issues of Algorithm 1 is that the bounds obtained are conservative for extremely small values of ε_2 resulting in difficult practical application. Therefore, a challenging topic of great interest is to reduce this conservatism. Along this direction, an attempt is given in [15], where a bootstrap technique is used to determine a stopping rule for the number of required samples.

In the next section, we present an efficient optimal design approach for uncertain systems based on RA for average performance synthesis.

3.2 Optimal Design of Uncertain Systems

In this section we apply the algorithm presented in Section 3.1 to both the a-priori and a-posteriori design criteria for optimal design for the uncertain systems. We assume that all performance functions are normalized to satisfy the requirement in Algorithm 1.

3.2.1 The a-priori Design Criteria

As we discuss earlier, for the a-priori design criteria, by comparing (3.4) and (3.10) we get the performance function

$$J_s(\xi, \theta) = g(\bar{y}_{\theta,\xi})$$

where $\bar{y}_{\theta,\xi}$ is the stationary output of the system equations (3.3) determined by θ and ξ. In applying Algorithm 1 we sample the parameter and controller spaces to obtain ξ_i and θ_k for $i = 1, ..., N$, $k = 1, ..., M$. Thus, we obtain

$$J_s(\xi_i, \theta_k) = g(\bar{y}_{k,i})$$

where $\bar{y}_{k,i}$ is the output solution of steady state equations for θ_k, ξ_i, that is,

$$\begin{aligned} \bar{x} &= T(\bar{x}, \theta_k, \xi_i) \\ \bar{y}_{k,i} &= f(\bar{x}) \end{aligned} \tag{3.20}$$

If $\bar{y}_{k,i}$ is not unique, since the a-priori equilibrium measure is only determined by the system steady state equations, and has nothing to do with system dynamics, $\varpi_s^{\theta_k}$ is a Dirac delta measure concentrated at every stationary output with equal weights, i.e.

$$J_s(\xi_i, \theta_k) = \frac{1}{n(\bar{y}_{k,i})} \sum_{j=1}^{n(\bar{y}_{k,i})} g(\bar{y}_{k,i,j})$$

where $n(\bar{y}_{k,i})$ is the number of distinct solution $\bar{y}_{k,i}$ for equation (3.20). Furthermore, by Algorithm 1 we obtain the a-priori design controller,

$$\overset{\wedge^s}{\theta}_{NM} = \arg\min_{k=1,\ldots,M} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n(\bar{y}_{k,i})} \frac{g\left(\bar{y}_{k,i,j}\right)}{n(\bar{y}_{k,i})}. \tag{3.21}$$

3.2.2 The a-posteriori Design Criteria

For the a-posteriori design criteria, it is not as straightforward to derive the suboptimal solution. In particular, in order to apply the randomized design algorithm we rewrite the a-posteriori design criteria (3.7) as follows.

Proposition 10 *The a-posteriori design criteria (3.7) satisfies*

$$\int_{\mathbb{R}} g(z)\, \varpi_a^{\theta}(dz) = \int_H J_a(\xi,\theta)\rho(d\xi)$$

where the performance function $J_a(\xi,\theta)$ is given by

$$J_a(\xi,\theta) = \int_{\mathbb{R}} g(z)\nu\left(x \in G | f_{\theta}^{*}(x,\xi) \in dz\right).$$

Proof. We note that

$$\begin{aligned} & \int_{\mathbb{R}} g(z)\, \varpi_a^{\theta}(dz) \\ = & \int_{\mathbb{R}} g(z)\, \eta\left((f_{\theta}^{*})^{-1}(dz)\right) \\ = & \int_H \int_G \int_{\mathbb{R}} g(z) \chi_{(f_{\theta}^{*})^{-1}(dz)}(x,\xi)\nu(dx)\rho(d\xi) \end{aligned}$$

where $\chi_{(\cdot)}$ is the indicator function and ν is the initial distribution defined on state space G. Thus, we have

$$\begin{aligned} J_a(\xi,\theta) &= \int_{\mathbb{R}}\int_G g(z)\chi_{\left(f_\theta^*\right)^{-1}(dz)}(x,\xi)\nu(dx) \\ &= \int_{\mathbb{R}} g(z)\nu\left(x\in G|f_\theta^*(x,\xi)\in dz\right). \end{aligned}$$

■

Now that we have rewritten the design criteria in the standard form (3.10) we proceed to apply the randomized algorithm for optimal design. We sample parameter and controller space to obtain $J_a(\xi_i,\theta_k)$,

$$J_a(\xi_i,\theta_k)=\int_{\mathbb{R}} g(z)\nu\left(x\in G|f_{\theta_k}^*(x,\xi_i)\in dz\right). \tag{3.22}$$

From the definition of the asymptotic output measure ϖ_a^θ we have

$$\varpi_a^{\theta_k}(E|\xi=\xi_i)=\nu\left(x\in G|f_{\theta_k}^*(x,\xi_i)\in E\right).$$

Also, by (2.9), we have

$$\varpi_a^{\theta_k}(E|\xi=\xi_i)=\mu_{\xi_i}^{\theta_k}(f^{-1}(E))$$

where $\mu_{\xi_i}^{\theta_k}$ is the invariant measure of the dynamical system for given ξ_i and θ_k. Thus, we can rewrite (3.22) with respect to invariant measure as

$$\begin{aligned} J_a(\xi_i,\theta_k) &= \int_{\mathbb{R}} g(z)\mu_{\xi_i}^{\theta_k}(f^{-1}(dz)) \\ &= \int_G g\circ f(x)\mu_{\xi_i}^{\theta_k}(dx). \end{aligned}$$

By (3.19), the a-posteriori suboptimal design vector is

$$\overset{\wedge}{\theta}{}^{a}_{NM} = \underset{k=1,\ldots,M}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} \int_{G} g \circ f(x) \, \mu_{\xi_i}^{\theta_k}(dx) \,. \tag{3.23}$$

In order to evaluate the cost function (3.23) we need to evaluate the integrals $\int_G g \circ f(x) \, \mu_{\xi_i}^{\theta_k}(dx)$ which requires the invariant measures $\mu_{\xi_i}^{\theta_k}$. A common approach for the computation of $\mu_{\xi_i}^{\theta_k}$ is to use a Monte Carlo type method that samples the distribution of the initial state and then simulates the system until it reaches stationarity (steady state), which is discussed in the next chapter. Then we present an alternative operator based approach for the computation of the invariant measure that is more accurate and efficient than the Monte Carlo approach.

CHAPTER 4

Computation of Invariant Measure

4.1 Monte Carlo Method

In this section we discuss Monte Carlo (MC) and quasi Monte Carlo (QMC) methods. The Monte Carlo method has a very long history, which began in 1949 with the seminal paper of Metropolis and Ulam [41]. The quasi Monte Carlo (QMC) is an improved Monte Carlo method, which is more recent and regarded as a deterministic version of MC [46].

Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling to compute their results. Monte Carlo methods are often used in simulating physical systems and evaluation of mathematical objects. Because of their reliance on repeated computation, these methods are most suited for calculation by a computer and tend to be used when it is unfeasible or impossible to compute an exact result with a deterministic algorithm.

Monte Carlo simulation methods are especially useful in studying systems with a large number of coupled degrees of freedom, such as fluids, disordered materials, strongly coupled solids, and cellular structures. More broadly, Monte Carlo methods are useful for modeling phenomena with significant uncertainty in inputs, such as the calculation of risk in business. These methods are also widely used in mathematics: a classic use is for the evaluation of definite integrals, particularly multidimensional integrals with complicated boundary conditions. It is a widely successful method in risk analysis when compared with alternative methods or human intuition.

4.1.1 Monte Carlo Methods for Calculating Invariant Measure

The most straightforward method for obtaining the stationary distribution (invariant measure) for the uncertain dynamical system (3.3) is Monte Carlo Simulation. We assume that the probability distribution of the initial state $x_0 = x$ is known and want to find the invariant measure $\mu_{\xi_i}^{\theta_k}$ for (3.3).

We propose the following nested Monte Carlo algorithm for calculating the invariant distribution. In particular we sample L points x_0^l from the initial distribution ν and for each sample we simulate the system equation until it reaches steady state $\bar{x}\left(x_0^l, \xi_i, \theta_k\right)$. Let χ_A be the indicator function for the set A and define random variables $z_l\left(\xi_i, \theta_k\right) = \chi_A\left(\bar{x}\left(x_0^l, \xi_i, \theta_k\right)\right)$. Then compute $\frac{1}{L}\sum_{l=1}^{L} z_l\left(\xi_i, \theta_k\right)$ as an approximation to $\mu_{\xi_i}^{\theta_k}(A)$.

The accuracy of this method is determined by the number of the samples we pick from the distribution ν. In order to evaluate the error in the Monte Carlo approximation scheme we compare simulation result with the true value $\mu_{\xi_i}^{\theta_k}(A)$. Consider the independent samples $\bar{x}\left(x_0^l, \xi_i, \theta_k\right), l = 1, ..., L$ and assume that the distribution of the $\bar{x}\left(x_0^l, \xi_i, \theta_k\right)$ is identical to that of the steady state $\bar{x}\left(x, \xi_i, \theta_k\right)$, i.e. each simulation has reached the steady state. Then it follows that $E[z_l\left(\xi_i, \theta_k\right)] = E[\chi_A\left(\bar{x}\left(x_0^l, \xi_i, \theta_k\right)\right)] = \mu_{\xi_i}^{\theta_k}(A)$ and by the central limit theorem,

$$\Pr\left(\left|\frac{1}{L}\sum_{l=1}^{L} z_l\left(\xi_i, \theta_k\right) - \mu_{\xi_i}^{\theta_k}(A)\right| < \frac{3v\left(\xi_i, \theta_k\right)}{\sqrt{L}}\right) \approx 0.997 \tag{4.1}$$

and the mean square error between the estimate $\frac{1}{L}\sum_{l=1}^{L} z_l\left(\xi_i, \theta_k\right)$ and true value $\mu_{\xi_i}^{\theta_k}(A)$ is given by,

$$e_L^{MC} = \left(E\left[\left|\frac{1}{L}\sum_{l=1}^{L} z_l\left(\xi_i, \theta_k\right) - \mu_{\xi_i}^{\theta_k}(A)\right|^2\right]\right)^{\frac{1}{2}} = \frac{v\left(\xi_i, \theta_k\right)}{\sqrt{L}} \tag{4.2}$$

where $v(\xi_i, \theta_k)$ is the standard deviation of the $z_l(\xi_i, \theta_k)$, that is,

$$\begin{aligned} v(\xi_i, \theta_k)^2 &= E\left[\left(z_l(\xi_i, \theta_k) - \mu_{\xi_i}^{\theta_k}(A)\right)^2\right] \\ &= E\left[z_l(\xi_i, \theta_k)^2\right] - 2E\left[z_l(\xi_i, \theta_k)\right]\mu_{\xi_i}^{\theta_k}(A) + \left(\mu_{\xi_i}^{\theta_k}(A)\right)^2 \\ &= \mu_{\xi_i}^{\theta_k}(A) - 2\left(\mu_{\xi_i}^{\theta_k}(A)\right)^2 + \left(\mu_{\xi_i}^{\theta_k}(A)\right)^2 \\ &= \mu_{\xi_i}^{\theta_k}(A) - \left(\mu_{\xi_i}^{\theta_k}(A)\right)^2 \end{aligned}$$

Note that since $\mu_{\xi_i}^{\theta_k}(A) \in [0,1]$ we have $v(\xi_i, \theta_k)^2 \leq \frac{1}{4}$. Also, we note that the variance depends on the uncertain parameter.

At the heart of the Monte Carlo method is the generation of the independent statistical samples x_0 from distribution ν. It can be shown that an arbitrary random sequence $(\phi_i)_{i=1,2,...}$ can always be represented as a sequence of functions $\psi_i(\alpha_1, ..., \alpha_i), i = 1, 2...$ where $(\alpha_i)_{i=1,2,...}$ is a sequence of independent uniformly distributed random variables on unit interval $U = [0,1]$ [31]. In particular, assume that we are interested in generating a sequence of independent and identically distributed random variables with cumulative distribution function F_ν and random variable α is uniformly distributed on U. Define

$$g(\alpha) = \inf\{z : F_\nu(z) \geq \alpha\} \tag{4.3}$$

Then we have,

$$\begin{aligned} \Pr\{g(\alpha) < x\} &= \Pr\{\inf\{z : F_\nu(z) \geq \alpha\} < x\} \\ &= \Pr\{F_\nu(x) \geq \alpha\} \\ &= F_\nu(x) \end{aligned}$$

Therefore, the random variable $\phi = g(\alpha)$ has cumulative distribution function F_ν. Using this method, a sequence of independent random variables $(\phi_i)_{i=1,2,...}$ with an arbitrary

distribution F_v can be generated from a sequence of uniformly distributed random variables $(\alpha_i)_{i=1,2,...}$. Other methods, that may be computationally superior, for generating the sequences $(\phi_i)_{i=1,2,...}$ are discussed in [32].

A device that generates the sequence $(\alpha_i)_{i=1,2,...}$ of independent uniformly distributed random variables is called a random generator. The most common procedure for generating $(\alpha_i)_{i=1,2,...}$ is to use efficient computer codes that generate pseudo-random sequences, i.e. long sequences that have many of the properties of the independent uniformly distributed random sequences, as measured by statistical tests.

One of the drawback of Monte Carlo method is that the mean square error between the approximate and true value is proportional to $\frac{1}{\sqrt{L}}$. Therefore, in order to achieve a mean square error of ε we need of the order of $\frac{1}{\varepsilon^2}$ simulations, i.e. $L = o\left(\frac{1}{\varepsilon^2}\right)$. In many real applications, particularly for complex dynamical systems, each simulation run can be computationally expensive as well as very time-consuming.

For problems where each simulation run is computationally very expensive more efficient methods are needed. Various improved statistical sampling techniques such as importance sampling and stratified sampling) exists that result in reduction in the variance in the error estimate (4.2) but still have an error of the order $\frac{1}{\sqrt{L}}$ [32]. An attempt for increasing the computational efficiency that has shown great promise in several applications are so-called quasi Monte Carlo methods which arise in information based complexity theory [66].

4.1.2 Quasi Monte Carlo Methods for Calculating Invariant Measure

The quasi Monte Carlo (QMC) method is a deterministic version of Monte Carlo with guaranteed approximation errors instead of probabilistic errors (4.2). In the quasi Monte Carlo method, deterministic points chosen according to some optimal criterion are used instead of random samples generated according to a given probability density function. In particular, in quasi Monte Carlo methods, the sample sequence with some distribution

F_V is generated by a low discrepancy sequence of random variables instead of uniformly distributed random sequences in regular Monte Carlo methods [46].

Let d denote the dimension of random variable we want to generate, let $x = (x_1 x_2 ... x_d)^T \in U^d$ and let $\chi_{[0,x)}$ be the indicator function for the cube $[0,x)$. For $\varsigma_1, ..., \varsigma_L \in U^d$ define

$$\begin{aligned} R_L(x; \varsigma_1, ..., \varsigma_L) &= \frac{1}{L} \sum_{l=1}^{L} \chi_{[0,x)}(\varsigma_l) - m([0,x)) \\ &= \frac{1}{L} \sum_{l=1}^{L} \chi_{[0,x)}(\varsigma_l) - x_1 x_2 ... x_d \end{aligned}$$

where m is the Lebesgue measure. The L_2 and L_∞ discrepancy of $\varsigma_1, ..., \varsigma_L$ are defined as follows

$$\begin{aligned} \|R_L(\cdot; \varsigma_1, ..., \varsigma_L)\|_2 &= \left(\int_{U^d} R_L^2(x; \varsigma_1, ..., \varsigma_L)\, dx \right)^{\frac{1}{2}} \\ \|R_L(\cdot; \varsigma_1, ..., \varsigma_L)\|_\infty &= \sup_{x \in U^d} |R_L(x; \varsigma_1, ..., \varsigma_L)| \end{aligned}$$

It can be shown that

$$\inf_{\varsigma_1, ..., \varsigma_L} \|R_L(\cdot; \varsigma_1, ..., \varsigma_L)\|_2 = o\left(\frac{(\log L)^{\frac{d-1}{2}}}{L} \right)$$

A low discrepancy sequence $(\varsigma_l)_{l=1,2,...}$ is an infinite sequence for which

$$\|R_L(\cdot; \varsigma_1, ..., \varsigma_L)\|_2 \leq \|R_L(\cdot; \varsigma_1, ..., \varsigma_L)\|_\infty = o\left(\frac{(\log L)^d}{L} \right)$$

The basic idea behind a low discrepancy sequence is that they cover the unit cube U^d as uniformly as possible. Examples of low discrepancy sequences are the Halton and Sobol sequences [53].

Assume that we use a low discrepancy sequence $(\varsigma_l)_{l=1,2,...}$ to generate the sequence of initial states $\left(x_0^l\right)_{l=1,2,...}$, e.g. $x_0^l = g(\varsigma_l)$ where $g(\cdot)$ is defined by (4.3). Then we simulate system (3.3) L times giving $z_l(\xi_i, \theta_k) = \chi_A(\bar{x}\left(x_0^l, \xi_i, \theta_k\right))$ where $\bar{x}\left(x_0^l, \xi_i, \theta_k\right)$ is the steady

state of (3.3) for each sample ξ_i, θ_k and initial condition x_0^l. Using the fact that $(\varsigma_l)_{l=1,2,...}$ is a low discrepancy sequence it can then be shown that the mean square error for the quasi Monte Carlo method is given by

$$\begin{aligned} e_L^{QMC} &= \left| \frac{1}{L} \sum_{l=1}^{L} z_l(\xi_i, \theta_k) - \mu_{\xi_i}^{\theta_k}(A) \right|^2 \\ &\leq C \| R_L(\cdot; \varsigma_1, ..., \varsigma_L) \|_\infty \\ &\leq o\left(\frac{(\log L)^d}{L} \right) \end{aligned} \tag{4.4}$$

where the constant C depends only on the parameters of the system and is independent of the number of samples L.

Compared to (4.2), the quasi Monte Carlo method outperforms the Monte Carlo method in terms of convergence speed if the dimension d is small. For large d the upper bound for the quasi Monte Carlo method exceeds the Monte Carlo error bound for large value of L. Several simulation studies for large values of d indicate that quasi Monte Carlo methods still outperform standard Monte Carlo methods [65], [53], [49], [51]. However, these results are empirical and rigorous explanation for this good performance of quasi Monte Carlo methods for large values of d do not exist. Although there exist other methods (such as variance reduction schemes) for improving the computational efficiency of Monte Carlo methods, these methods only result in marginal improvements or are only applicable to a limited class of dynamical systems. Thus, there is a considerable need for efficient alternatives. In the next section, we introduce a new approach to solve this problem using some recently developed results from Random Dynamical Systems.

4.2 Operator Approach

In this section we discuss an approach for calculating the invariant measure in (3.23) using the theory of random dynamical systems in Section 2.3. In particular we note that under the appropriate conditions the system has an ergodic invariant measure that is characterized as a fixed point of the Perron Frobenius operator P_ξ for T_ξ.

We note that the invariant measure may not be unique. However, it can be shown that, under mild conditions on the dynamical system, by adding small (localized) noise, the resulting system possesses an unique invariant measure [37] that converges to the true ergodic measure as the noise intensity converges to zero. The computational approach relies on the discretization of the P-F operator that we discuss next.

4.2.1 Discretization

In order to obtain a finite-dimensional (discrete) approximation of the P-F operator, we consider a finite partition of the state space G, denoted as $B_{1,}B_2,...B_L$, where $B_i \cap B_j = \phi$ and $\cup_j B_j = G$. Corresponding to each partition element we associate a positive number $\mu_j \in [0,1]$ with $\sum_{j=1}^{L} \mu_j = 1$, i.e. $\mu = (\mu_{1,}...\mu_L) \in \mathbb{R}^L$ is a probability vector. Define a probability measure on G as

$$\overline{\mu}(dx) = \sum_{i=1}^{L} \overline{\mu}_i \chi_{B_i}(x) \frac{m(dx)}{m(B_i)} \tag{4.5}$$

where m is the Lebesgue measure and χ_{B_i} is the indicator function for B_i. Then, the action of the Perron Frobenius operator P_ξ on $\overline{\mu}$ on the element B_j is

$$\begin{aligned} P_\xi \overline{\mu}(B_j) &= \overline{\mu}\left(T_\xi^{-1}(B_j)\right) \\ &= \sum_{i=1}^{L} \overline{\mu}_i \frac{m(T_\xi^{-1}(B_j) \cap B_i)}{m(B_i)} = \sum_{i=1}^{L} \overline{\mu}_i \overline{P}_{ij}(\xi) \end{aligned}$$

where the $L \times L$ matrix with entries

$$\overline{P}_{ij}(\xi) = \frac{m(T_\xi^{-1}(B_j) \cap B_i)}{m(B_i)} \tag{4.6}$$

is a stochastic transition matrix. We will see below that the operator $\overline{P}(\xi)$ is a "good" approximation of P_ξ and the invariant measure for P_ξ can be approximated by a measure $\overline{\mu}$ defined by (4.5) where the coefficients of $\overline{\mu}$ are invariant for $\overline{P}(\xi)$, i.e. satisfy $\overline{\mu}_i = (\pi_L)_i$ where $\pi_L = \pi_L \overline{P}(\xi)$. We note that the computation of the entries of $\overline{P}(\xi)$ is much more efficient than the Monte Carlo method.

The basic justification for using a finite dimensional approximation for the calculation of the invariant measure lies in the theory of finite dimensional approximations for compact operators [19], [47]. For the Perron Frobenius operator P_ξ we will define an approximate compact operator $P_\varepsilon : L_2(G) \to L_2(G)$ and then use finite dimensional approximations for compact operators to obtain the finite dimensional approximation for P_ξ. Here $L_2(G)$ denotes the space of functions that are square integrable on G. Define a kernel

$$k_\varepsilon(y,z) = \frac{1}{\varepsilon^n m(B)} \chi_B\left(\frac{z-y}{\varepsilon}\right), \qquad x,z \in G$$

where $B \subset G$ is the ball of radius one and center at zero. We note that $k_\varepsilon\left(T_\xi(x), z\right)$ is a transition density for the transition function

$$p_\varepsilon(x,A) = \int_A k_\varepsilon\left(T_\xi(x), z\right) m(dz)$$

It can be shown that $p_\varepsilon(x,\cdot) \to \delta_{T_\xi(x)}(\cdot)$ as $\varepsilon \to 0$ in a weak sense, i.e. p_ε is the transition function for a Markov process that is a small random perturbation of a discrete dynamical system defined by T_ξ [36]. We note that the evolution of the distribution for the Markov process is given by the operator $P_\varepsilon \nu(A) = \int_G p_\varepsilon(x,A)\, \nu(dx)$. If the initial measure ν has density g with respect to m then P_ε can be viewed as an operator mapping $L_2(G) \to L_2(G)$,

i.e. the density evolves according to

$$P_\varepsilon g(y) = \int k_\varepsilon \left(T_\xi(x), y\right) g(x)\, m(dx)$$

Next note that

$$\int \int \left|k_\varepsilon \left(T_\xi(x), z\right)\right| m(dx)\, m(dz) \leq \frac{m(G)}{\varepsilon^n m(B)}$$

Therefore, the transition operator P_ε is a compact operator on $L_2(G)$ [38].

Next we describe how to construct the finite dimensional approximation (4.6) for the compact operator P_ε that (for small ε) gives a finite dimensional approximation for P_ξ as well [19]. Let V_L be a $L-$dimensional approximation of $L_2(G)$, e.g. $V_L = span\{\varphi_1, \ldots, \varphi_L\}$ for some "independent" functions $\varphi_1, \ldots, \varphi_L \in L_2(G)$. Let $Q_L : L_2(G) \to V_L$ be a projection such that Q_L converges pointwise to the identity in $L_2(G)$ as $L \to \infty$. Define an approximate operator $P_{\varepsilon,L} = Q_L P_\varepsilon$ where P_ε is the compact operator defined previously. Then $\|P_{\varepsilon,L} - P_\varepsilon\|_2 \to 0$ as $L \to \infty$. We use the *finite dimensional* operator $P_{\varepsilon,L}$ as an approximation of the Perron-Frobenius operator P_ξ.

Let V_L be defined by $\varphi_i(y) = \chi_{B_i}(y)$ where the sets B_i form the partition of G discussed earlier. Define the Galerkin projection Q_L of $g \in L_2(G)$ by

$$\langle Q_L g, \varphi_i \rangle = \langle g, \varphi_i \rangle, \quad i = 1, \ldots, L \tag{4.7}$$

where $\langle \cdot, \cdot \rangle$ is the inner product on $L_2(G)$. Since $\varphi_i(y) = \chi_{B_i}(y)$ we have

$$\int_{B_i} Q_L g = \int_{B_i} g, \quad i = 1, \ldots, L$$

For $g \in V_L = span\{\varphi_1, \ldots, \varphi_L\}$ we write $g(y) = \sum_{i=1}^{L} \varphi_i(y) g_i$. It is easy to see that for any such g we have $\left\langle g, \varphi_j \right\rangle = g_j$. Now for any $g \in L_2(G)$ we have by definition $P_{\varepsilon,L} g = Q_L P_\varepsilon g$.

Therefore, we get from (4.7) with $P_\varepsilon g$ replacing g

$$\langle P_\varepsilon g, \varphi_i \rangle = \langle Q_L P_\varepsilon g, \varphi_i \rangle = \langle P_{\varepsilon,L} g, \varphi_i \rangle, \quad i = 1, \ldots, L \tag{4.8}$$

Since $P_{\varepsilon,L} g \in V_L$ we know that there exist constants $m_i, i = 1, \ldots, L$ such that $P_{\varepsilon,L} g = \sum_{i=1}^{L} \varphi_i(y) m_i$. Furthermore, for $j = 1, \ldots, L$ we get from (4.8)

$$m_j = \left\langle P_{\varepsilon,L} g, \varphi_j \right\rangle = \left\langle P_\varepsilon g, \varphi_j \right\rangle$$

Now if in addition $g \in V_L$ then

$$P_{\varepsilon,L} g = P_{\varepsilon,L} \sum_{j=1}^{L} \varphi_j(y) g_j = \sum_{j=1}^{L} P_{\varepsilon,L} \varphi_j(y) g_j$$

Note that if $g = \varphi_j$ then $P_{\varepsilon,L} \varphi_j(y) = \sum_{i=1}^{L} \varphi_i(y) m_{ij}^\varepsilon$ where $m_{ij}^\varepsilon = \left\langle P_{\varepsilon,L} \varphi_j, \varphi_i \right\rangle = \left\langle Q_L P_\varepsilon \varphi_j, \varphi_i \right\rangle = \left\langle P_\varepsilon \varphi_j, \varphi_i \right\rangle$. Thus

$$P_{\varepsilon,L} g = \sum_{j=1}^{L} \sum_{i=1}^{L} \varphi_i(y) m_{ij}^\varepsilon g_j$$

We note that when restricted to the finite dimensional subspace V_L the action of the the operator P_ξ is fully represented by the $L \times L$ matrix M^ε with coefficients m_{ij}^ε. We finally note that in the limit $\varepsilon \to 0$ we have $m_{ij}^\varepsilon \to m(T^{-1}(B_i) \cap B_j)$ which after renormalization to a stochastic matrix agrees with (4.6). Furthermore, it follows from the results in [37], [19] that when the invariant measure of (2.5) is an ergodic invariant measure in the sense of Definition 16 then the approximate invariant measure π_L converges to μ_ξ as $L \to \infty$.

4.2.2 Subdivision

In this section, we will discuss the application of subdivision methods for the partition of the state space in the computation of the approximate invariant measure. Recently subdivision methods have been successfully applied to the numerical analysis of complex dynam-

ical behavior [21], [17], [16], [19], [25]. These methods can be used for two essentially different purposes: the first is to understand the geometric structure of an underlying attractor. Secondly the goal may be to approximate the observable dynamical behavior of the underlying system in a specific region of state space by the computation of invariant measures. Here, we concern the application of subdivision algorithms for the second purpose.

The subdivision algorithm has been used to approximate the unstable manifolds of invariant sets in discrete dynamical systems. Existing methods for the computation of an unstable manifolds are local in the sense that one starts the computation at a hyperbolic periodic point and then computes part of the unstable manifold by some sort of continuation procedure [52]. In contrast to these methods, the subdivision algorithm is a global approach in the estimation of unstable manifolds.

In particular, we consider discrete dynamical systems of the form

$$x_{j+1} = f\left(x_j\right), \qquad j = 0, 1, 2, ...,$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism. The central object which is approximated by the subdivision algorithm developed in [17] is the so-called *relative global attractor*,

$$A_Q = \bigcap_{j \geq 0} f^j\left(Q\right)$$

where $Q \subset \mathbb{R}^n$ is a compact subset. Roughly speaking, the set A_Q should be viewed as the union of unstable manifolds of invariant sets inside Q. Moreover, the relative global attractor is in general "invisible" in the sense that it cannot be computed by direct simulation of the underlying dynamical system.

The subdivision algorithm for the approximation of A_Q generates a sequence $\mathscr{B}_0, \mathscr{B}_1, \mathscr{B}_2, ...$ of finite collections of boxes with the property that for all integers k the set $Q_k = \bigcup_{B \in \mathscr{B}_k} B$ is a covering of the relative global attractor under consideration. The sequence of coverings

is constructed in such a way that the diameter of the boxes,

$$\operatorname{diam}(\mathscr{B}_k) = \max_{B \in \mathscr{B}_k} \operatorname{diam}(B)$$

converges to zero for $k \to \infty$. Moreover, it is shown in [17] that the finite collection of closed subsets Q_k converges to the relative global attractor A_Q as $k \to \infty$.

Given an initial collection $\mathscr{B}_0$, one can inductively obtain $\mathscr{B}_k$ from $\mathscr{B}_{k-1}$ for $k = 1, 2, ...$ by any subdivision method. The simplest approach is to use bisection for all boxes and such algorithm called a *standard subdivision algorithm*. In particular, we assume that the dynamical system is defined on a compact subset of $\mathbb{R}^n$. We start by specifying one box in $\mathbb{R}^n$ on which we want to analyze the dynamical behavior. For a given fixed integer L we interactively obtain 2^L boxes in $\mathbb{R}^n$, $B_1, ..., B_{2^L}$, of equal size by a bisection algorithm.

In the calculation of the approximate invariant measure, the principal factor affecting the computational complexity is the discretization level on the state space. Using the standard subdivision algorithm with the discretization level L, the size of the approximate stochastic transition matrix $\overline{P}$ is then $2^L \times 2^L$. We note that if the requirement for computational accuracy is stringent then L will be large and the resulting $\overline{P}$ will be huge requiring excessive storage and computational effort.

The standard subdivision algorithm described above leads to a partition with boxes of equal size. We note that the subdivision is done without utilizing any information about the system dynamics or the invariant measure. However, frequently there exist subsets in the state space that have a very small invariant measure. Consequently, the subdivision of these subsets is not necessary for the computational determination of the invariant measure and their subdivision will lead to unnecessary computational effort. By incorporating information about the invariant measure it is possible to produce more efficient partitioning schemes that result in an adaptive subdivision into boxes of unequal sizes. Here, we introduce an *adaptive subdivision algorithm* that is a variant of the algorithm developed in

[18].

Roughly speaking, the idea of the adaptive subdivision algorithm is as follows. Suppose that the dynamical system is defined on G and we start by one box covering the whole space. In a first step we subdivide this box and throw away boxes which have empty intersection with the support of the invariant measure. Then we subdivide the remaining boxes again and proceed in the same manner. Since we use the information on the actual approximation of the invariant measure to decide whether or not a box should be subdivided, the adaptive subdivision algorithm is obviously more efficient in the calculation of the approximate invariant measure.

In particular, let $\{\delta_k\}$ be a sequence of positive real numbers such that $\delta_k \to 0$ for $k \to \infty$ and let $\mathscr{B}_k$ be a finite collection of compact subsets of $\mathbb{R}^n$ at step k (the partition at step k). Let π_k be invariant measure of stochastic matrix $\overline{P}_k$ obtained at step k. Given an initial pair $(\mathscr{B}_0, \pi_0)$, one inductively obtains $(\mathscr{B}_k, \pi_k)$ from $(\mathscr{B}_{k-1}, \pi_{k-1})$ for $k = 1, 2, \ldots$ in three steps:

(i) *Subdivision:* Define

$$\begin{aligned} \mathscr{B}_{k-1}^{-} &= \{B \in \mathscr{B}_{k-1} : \pi_{k-1}(B) < \delta_{k-1}\} \text{ and} \\ \mathscr{B}_{k-1}^{+} &= \mathscr{B}_{k-1} \backslash \mathscr{B}_{k-1}^{-} \end{aligned}$$

Construct a new collection $\hat{\mathscr{B}}_k^{+}$ such that

$$\bigcup_{B \in \hat{\mathscr{B}}_k^{+}} B = \bigcup_{B \in \mathscr{B}_{k-1}^{+}} B$$

where

$$\operatorname{diam}(\hat{\mathscr{B}}_k^{+}) \leq \theta \operatorname{diam}(\mathscr{B}_{k-1}^{+})$$

for some $0 < \theta < 1$.

(ii) *Calculation of the invariant measure:* Set

$$\hat{\mathscr{B}}_k = \mathscr{B}_{k-1}^- \cup \hat{\mathscr{B}}_k^+$$

For the collection $\hat{\mathscr{B}}_k$ calculate the approximate invariant measure as the fixed point $\hat{\pi}_k$ of the discretized P-F operator defied by (4.6).

(iii) *Selection:* Set

$$\mathscr{B}_k = \left\{ B \in \hat{\mathscr{B}}_k : \hat{\pi}_k(B) > 0 \right\}$$

and

$$\pi_k = \hat{\pi}_k|_{\mathscr{B}_k}.$$

In the first step, we subdivide the boxes in the collections $\hat{\mathscr{B}}_k^+$, which is consist of successively finer boxes as k increasing. Meanwhile, we remove each subset with small invariant measure based on the information in the second step. In the third step, a new collection $\mathscr{B}_k$ is constructed for the next subdivision. As we shall see, this algorithm is an efficient way for calculating the invariant measure.

Example 3 *Let us illustrate the algorithm for the simplest possible case, namely for the mapping* $f : \mathbb{R} \to \mathbb{R}$*,*

$$f(x) = \alpha x$$

with $\alpha \in \left(0, \frac{1}{2}\right)$*. With this choice of* α *the invariant measure of the dynamical system is a Dirac measure on origin, i.e.* $\delta_{\{x=0\}}$*. We start with a single interval* $\mathscr{B}_0 = \{[-1,1]\}$ *and construct* $\hat{\mathscr{B}}_k^+$ *by bisection of the intervals in* $\mathscr{B}_{k-1}^+$*. Hence,*

$$\mathscr{B}_0^- = \emptyset$$

$$\mathscr{B}_1 = \hat{\mathscr{B}}_1^+ = \{[-1,0],[0,1]\}.$$

No interval is removed, since each of them has equal invariant measure $\frac{1}{2}$. *Subdividing again, we get four intervals*

$$\mathscr{B}_2 = \hat{\mathscr{B}}_2 = \left\{ \left[-1,-\frac{1}{2}\right], \left[-\frac{1}{2},0\right], \left[0,\frac{1}{2}\right], \left[\frac{1}{2},1\right] \right\}.$$

Applying the selection rule, the two boundary intervals are removed from subdivision, i.e.

$$\mathscr{B}_2^- = \left\{ \left[-1,-\frac{1}{2}\right], \left[\frac{1}{2},1\right] \right\}$$

$$\mathscr{B}_2^+ = \left\{ \left[-\frac{1}{2},0\right], \left[0,\frac{1}{2}\right] \right\}$$

Keep the interval in $\mathscr{B}_2^-$ *and subdivide the interval in* $\mathscr{B}_2^+$, *we get*

$$\hat{\mathscr{B}}_3 = \left\{ \left[-1,-\frac{1}{2}\right], \left[-\frac{1}{2},-\frac{1}{4}\right], \left[-\frac{1}{4},0\right], \left[0,\frac{1}{4}\right], \left[\frac{1}{4},\frac{1}{2}\right], \left[\frac{1}{2},1\right] \right\}$$

$$\mathscr{B}_3 = \left\{ \left[-\frac{1}{2},-\frac{1}{4}\right], \left[-\frac{1}{4},0\right], \left[0,\frac{1}{4}\right], \left[\frac{1}{4},\frac{1}{2}\right] \right\}$$

Proceeding this way, we obtain after k subdivision steps

$$\mathscr{B}_k = \left\{ \left[-\frac{1}{2^{k-2}},-\frac{1}{2^{k-1}}\right], \left[-\frac{1}{2^{k-1}},0\right], \left[0,\frac{1}{2^{k-1}}\right], \left[\frac{1}{2^{k-1}},\frac{1}{2^{k-2}}\right] \right\}.$$

We see that the union $\cup_{B\in\mathscr{B}_k} B$ *is indeed approaching the global attractor* $x = 0$ *for* $k \to \infty$. *Moreover, the number of intervals in the collection remains constant during the adaptive subdivision algorithm whereas the standard subdivision algorithm has exponential growth of intervals.*

Compared with standard subdivision algorithm with an exponential growth of boxes, it is easy to see that adaptive subdivision algorithm even in the worst case, that is the invariant

measure is positive everywhere, has a polynomial growth of boxes:

$$|\mathscr{B}_k| = 1 + |\mathscr{B}_0^+| + |\mathscr{B}_1^+| + ... + |\mathscr{B}_{k-1}^+| \qquad \text{at step } k$$

where $|\cdot|$ denote the number of boxes in the collection and

$$|\mathscr{B}_i^+| \leq 2^i \qquad \text{for } i = 0, 1, ..., k-1$$

In the realization of the algorithm, we typically subdivide the boxes in the collection $\mathscr{B}_k^+$ by bisection. This guarantees that the number of boxes is not growing too fast. The details concerning the implementation are discussed in [17], [19]. Moreover, there is some freedom in choosing the sequence $\{\delta_k\}$ of positive numbers used in the subdivision step. Note however that this sequence determines the number of boxes which will be subdivided and hence it has a significant influence on the storage requirement. In the computations it turned out to be quite efficient to choose the average

$$\delta_k = \frac{1}{N_k} \sum_{B \in \hat{\mathscr{B}}_k} \pi_k(B) = \frac{1}{N_k}$$

where N_k is the number of boxes in $\mathscr{B}_k$.

Using the proposed P-F method in this section, the invariant measure $\mu_{\xi_i}^{\theta_k}$ is approximated by a measure $\pi_{\xi_i}^{\theta_k}$ calculated by the corresponding transition matrix $\overline{P}(\theta_k, \xi_i)$ defined in a similar way as (4.6):

$$\overline{P}_{ij}(\theta_k, \xi_i) = \frac{m(T_{\xi_i, \theta_k}^{-1}(B_j) \cap B_i)}{m(B_i)} \tag{4.9}$$

Utilizing the approximate invariant measure $\pi_{\xi_i}^{\theta_k}$ the approximate optimization problem for (3.23) becomes

$$\overset{\wedge}{\theta}{}_{NM}^{a} = \underset{k=1,...,M}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{L} g \circ f\left(m_j\right) \cdot \pi_{\xi_i}^{\theta_k}(j) \tag{4.10}$$

where L is the number of boxes in the partition of the state space G and m_j is the central point of the j-th box.

We illustrate both the Monte Carlo and P-F based approach for calculating the invariant measure and the uncertain system design in numerical examples in the next chapter.

CHAPTER 5

Example

5.1 Computation of Invariant Measure

5.1.1 Pitchfork

Consider the system in Example 1, i.e.

$$x_{i+1} = x_i + \xi x_i - x_i^3$$

where $x_i \in G \subset \mathbb{R}$ is the system state, $\xi \in H \subset \mathbb{R}$ is a random parameter. As before for initial conditions the initial measure is chosen as $\nu(E) = m(E \cap [-a,a])/2a$ and the uncertain parameter is assumed to have the distribution $\rho(E) = m(E \cap [-b,b])/2b$. We want to compute the invariant measure μ_ξ for any ξ. Below we illustrate the numerical determination of the invariant measure using both Monte Carlo Simulation and the suggested P-F method for random dynamical system. We pick $a = 1$ and $b = 0.5$ in the example.

Monte Carlo Method

First we sample $L = 100$ points from the distribution of the parameter to obtain ξ_i, $i = 1,2...100$. Second, for each ξ_i, we perform $K = 100$ simulations with initial conditions drawn from the initial condition distribution and obtain the distribution of the limiting steady states as the approximation of the invariant measure μ_{ξ_i}. The results are displayed in Figure 5.1. To display the invariant measure we use a square to denote the steady states and color the squares with respect to their probability (i.e. frequency).

From the Monte Carlo simulation we see that for each fixed value of ξ the system has

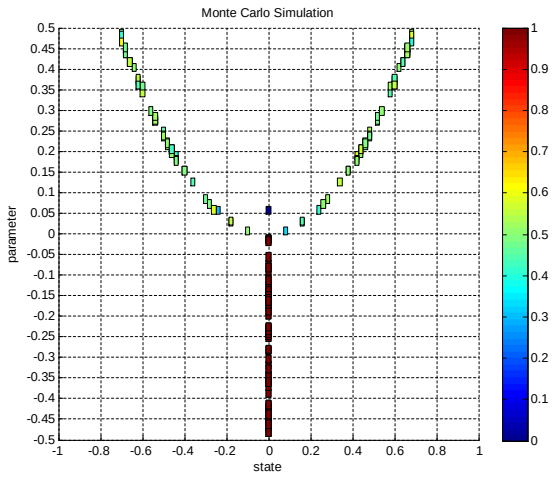

Figure 5.1: Invariant Measure by Monte Carlo

either one or two equilibrium points at $\bar{x} = 0$ and $\bar{x} = \pm\sqrt{\xi}$, respectively, and and the corresponding invariant measure μ_ξ is approximated as

$$\mu_\xi = \begin{cases} \frac{1}{2}\delta_{\{x=-\sqrt{\xi}\}} + \frac{1}{2}\delta_{\{x=\sqrt{\xi}\}} & \text{for } \xi \geq 0 \\ \delta_{\{x=0\}} & \text{for } \xi < 0 \end{cases}$$

Since the invariant measure μ_ξ is the combination of finite Dirac measure we note that the system is a parametric B_ρ-regular system.

P-F Method

The adaptive subdivision algorithm and the approximate P-F method was used to calculate the invariant measure μ_ξ. In order to the make the algorithm even more efficient we considered the parameter to be a dynamic state (sitting at a fixed point) and calculated the invariant measure for the skew product system. This allows us to adaptively select the size of the partition in both the state and uncertainty dimension.

The results using a standard subdivision algorithm and 1024 boxes are displayed in Figure 5.2.

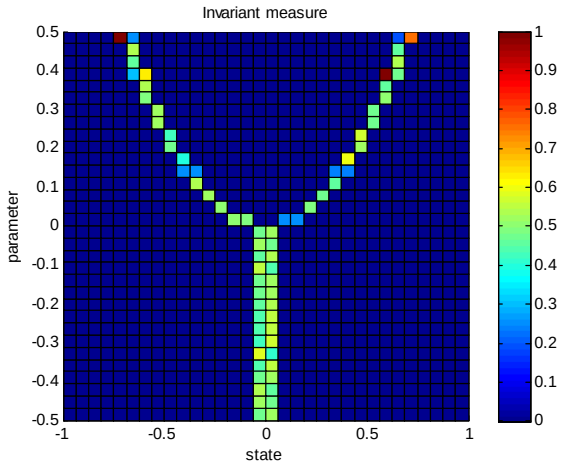

Figure 5.2: Invariant measure by standard subdivision

The savings in computational effort when using the adaptive subdivision algorithm is considerable. Indeed, after 10 subdivision steps (note $2^{10} = 1024$ so this corresponds to the previous non adaptive case) we obtain only 388 boxes so there is considerable reduction in computational load but at the same time the result are much more precise as is seen in Figure 5.3.

We note that the Monte Carlo method and P-F method give us the same results. Moreover, since the P-F method requires one step simulation for the determination of the elements of the discretized operator it was found to be much more efficient than the Monte Carlo method. This is illustrated in the following catalytic reactor example with S-shape steady state characteristic, which is referred to as a system with Arrhenius dynamics. The negative slope part of the S-shape curve represents a set of unstable steady states.

5.1.2 Catalytic Reactor

In [14], [20], a set of nonlinear partial differential equations, which describe heat and mass transfer in a spherically shaped catalytic pellet, was reduced to a dimensionless first-order ordinary differential equation which we represent in the form

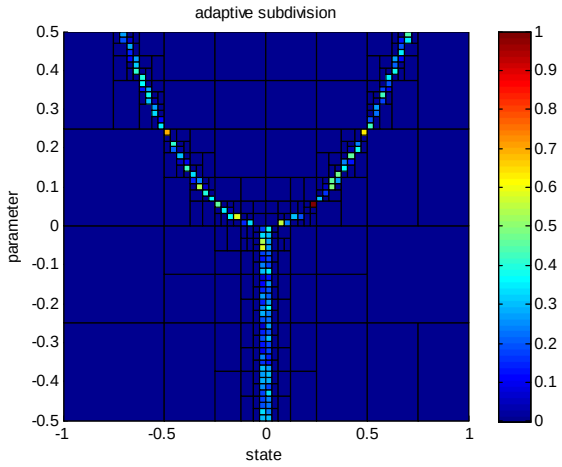

Figure 5.3: Invariant measure by adaptive subdivision

$$dx/dt = A - g(x,B)$$

where x is the dimensionless temperature in the reactor and t is time expressed in units of the rise time of the system. The rise time is defined as the time necessary to reach a sufficiently small neighborhood of the equilibrium position. A is the dimensionless external temperature, which is the random parameter in our example. B is a parameter defined by the reacting substances viewed as a constant and $g(x,B)$ is of the form

$$e^{-Bx}(\varphi_1(x)/\varphi_2(x))$$

where $\varphi_s(x)$, $s = 1,2$ are polynomials in x. In our simulation, we choose $B = 5.5$ and $g(x,B) = (x/(1-x))e^{-5.5x}$. For initial uncertainty of the system, we assume x has an uniform distribution in $[0,1]$ and A is also uniform in $[0,0.16]$. The discretization time step is 0.01.

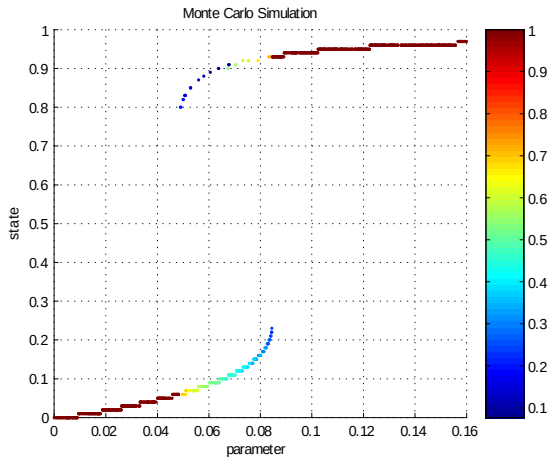

Figure 5.4: Invariant measure by Monte Carlo

Monte Carlo Method

We sample 1000 points both in state space and parameter space. As we discuss before, the mean square error is proportional to $\frac{1}{\sqrt{N}}$, that is about 0.03 in this example. The results are shown in Figure 5.4 and the simulation costs 50.94 sec.

P-F Method

The standard subdivision is shown in Figure 5.5 and the adaptive subdivision is shown in Figure 5.6. For standard subdivision, we have 4096 boxes to cover the whole space with 21.57 sec. simulation time while we only have 286 boxes with 8.49 sec. in adaptive subdivision. Moreover, the adaptive subdivision method has much higher accuracy than standard method in the computation of the invariant measure. The average size of boxes that cover the invariant measure is $2.5 * 10^{-6}$ in the adaptive subdivision while the size of boxes is $3.9 * 10^{-5}$ in the standard subdivision. The adaptive subdivision algorithm shows great efficiency in this example.

The negative slope part of the S-shape curve represents a set of unstable steady state. It has been shown that the system with smaller negative slope part gives an increase in productivity of the plants. Studying the invariant measure for the system can help us in

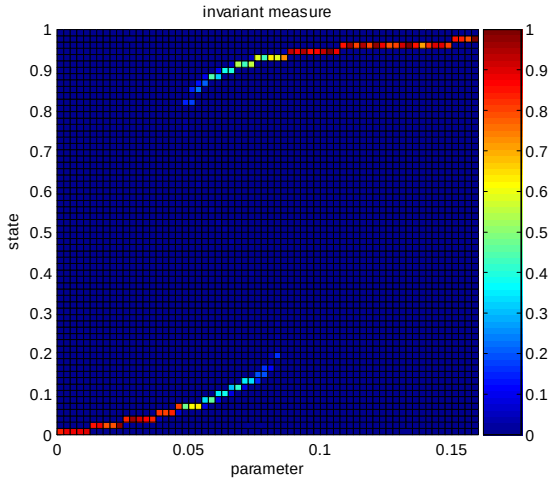

Figure 5.5: Invariant measure by standard subdivision

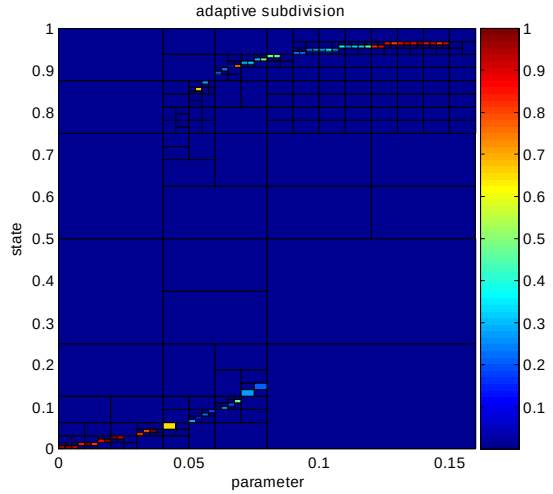

Figure 5.6: Invariant measure by adaptive subdivision

analysis of the domains of attraction and in the design of control methods to shape the curve so to increase the reaction productivity.

5.2 Uncertain System Design

Consider an optimal design problem for a continuous stirred tank reactor (CSTR) system. In [68], the kinetic equations of the reaction can be reduced to a dimensionless second-order ordinary differential equation which we represent in the form as (3.2)

$$\begin{aligned} \dot{x}_1 &= -x_1 + u \cdot \exp\{x_2\} \\ \dot{x}_2 &= -(1+\beta)x_2 + Bu \cdot \exp\{x_2\} \end{aligned} \tag{5.1}$$

where $x_1 \in [0,1]$, $x_2 \in [0,10]$ are the system states denoting the reactant concentration and temperature respectively, u is the control input for the reactant concentration. The constant $B \in [5,9]$ is defined by the dimensionless adiabatic temperature rise and related to the inverse of the reactant feed temperature which, in turn, is viewed as a random parameter. Finally, $\beta = 0.5$ is a dimensionless heat transfer coefficient viewed as a constant parameter in our example. It is well known that the norm of the steady state $\|x_s(B,u)\|_2$ has an interpretation as the productivity of the reaction process in the system, and maximization of this productivity in a stable steady state by an appropriate control method is an important practical goal. Here, we consider the simple state feedback control strategy,

$$u = D_a \cdot (1 - x_1)$$

where D_a is the Damkohler number is identified with the inverse of the input flow rate and is considered here as a control parameter. Then, we modify the dynamical system (5.1) by

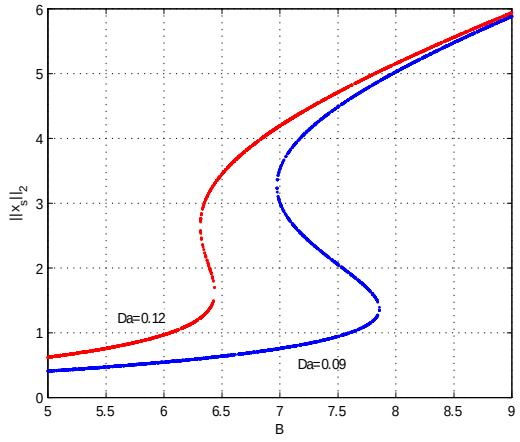

Figure 5.7: Arrhenius Dynamics

adding the control parameter D_a and an output equation related to the objective function,

$$\begin{aligned} \dot{x}_1 &= -x_1 + D_a(1-x_1)\exp\{x_2\} \\ \dot{x}_2 &= -(1+\beta)x_2 + BD_a(1-x_1)\exp\{x_2\} \\ y &= \sqrt{x_1^2 + x_2^2} \end{aligned} \tag{5.2}$$

It can be shown that steady state $x_s = (x_{1s}\, x_{2s})^T$ of (5.2) satisfies

$$\begin{aligned} D_a &= (x_{1s}/(1-x_{1s}))\exp\{-(B/(1+\beta))x_{1s}\} \\ x_{2s} &= (B/(1+\beta))x_{1s} \end{aligned} \tag{5.3}$$

Therefore,

$$\bar{y} = ||x_s||_2 = x_{1s}(1 + B^2/(1+\beta)^2)^{1/2}$$

which results, for all values of the control parameter D_a, in the S-shape steady state characteristic shown in Figure 5.7, referred to as a system with Arrhenius dynamics. The negative slope part of the S-shape curve represents a set of unstable steady states.In reality, the upper stable branch of the curve is not acceptable because of technological reasons. Hence, our

performance function is chosen to maximize the norm of steady states bounded to some acceptable range. Specifically, the resulting optimization problem is

$$\min_{\theta\in\Theta} g(\bar{y}) = \min_{D_a\in\Theta} \frac{1}{\chi_E(||x_s||_2)}$$

where $\chi_E(\cdot)$ is the indicator function for some acceptable system operation range E determined by reaction device parameters. Due to the random parameter B, the steady states are random also. The problem can be considered as an uncertain system design problem of the type discussed in the previous Sections. Below we will apply both the a-priori design method and the a-posteriori design method. In our design, we assume initial states have uniform distribution in $[0,1]\times[0,10]$ and the control parameter D_a is uniform on $[0.06,0.12]$ and choose $E=[0,2]$. The discretization step is 10^{-4}.

1) a-priori design method:

Choosing $\varepsilon_1=\varepsilon_2=0.1$, $\delta=0.01$, the design of the optimal controller according to Algorithm 1 requires

$$M = 51$$

samples of the controller parameter D_a, and

$$N = 496$$

samples of the uncertainty B. For every sample of D_a and B, we solve the steady state equation (5.3) and find the optimal controller parameter by (3.21),

$$D_a^* = 0.0784$$

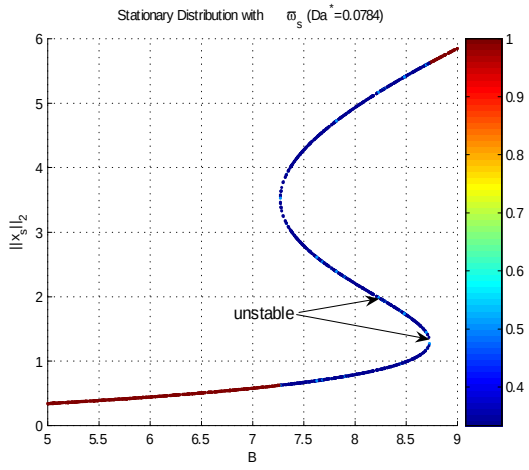

Figure 5.8: Stationary distribution by apriori design method

The corresponding average optimal performance value is

$$\begin{aligned}\widehat{\phi}_N^s(D_a^*) &= \frac{1}{N}\sum_{i=1}^{N} J_s(\xi_i, D_a^*) \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{n(\bar{y}_i)} \frac{1}{\chi_E\left(\bar{y}_{i,j}\right)\cdot n(\bar{y}_i)} = 2.1872\end{aligned}$$

The stationary distribution of system output $||x_s||_2$ colored by a-priori equilibrium measure is shown in Figure 5.8.

2) a-posteriori design method:

Setting the same probability levels as in the a-priori design method results in the same number for samples $M = 51$ and $N = 496$. For every sample of D_a, using the P-F method and adaptive subdivision algorithm, we obtain the approximate invariant measure $\pi_{\xi_i}^{\theta_k}$ for $k = 1,2,...,M, i = 1,2,...,N$ with the number of boxes $L = 472$. Finally, (4.10) gives us the optimal controller parameter

$$D_a^* = 0.0642$$

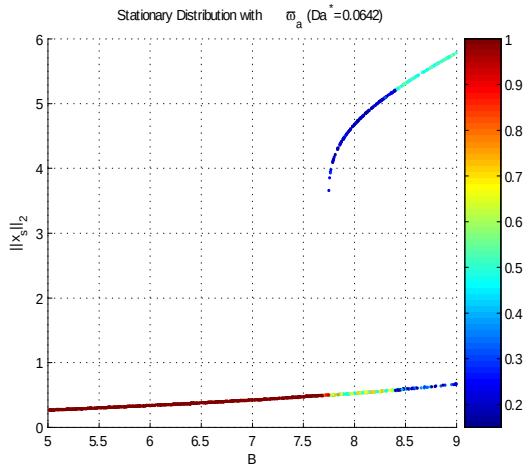

Figure 5.9: Stationary distribution by aposteriori design method

and the corresponding average optimal performance value

$$\begin{aligned}\widehat{\phi}_N^a(D_a^*) &= \frac{1}{N}\sum_{i=1}^{N} J_a(\xi_i, D_a^*) \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{L} \frac{1}{\chi_E\left(||m_j||_2\right)} \cdot \pi_{\xi_i}^{\theta_k}(j) \\ &= 2.8409\end{aligned}$$

The stationary distribution of output with respect to asymptotic output measure is shown in Figure 5.9.

Although the numerical calculations show that $\widehat{\phi}_N^s(D_a^*) < \widehat{\phi}_N^a(D_a^*)$, the a-priori equilibrium measure ϖ_s^θ does not discriminate between stable and unstable points and as result the optimal calculation includes values evaluated at some unstable steady states as shown in Figure 5.8. Obviously, unstable points are not allowed as part of the optimal solution. Consequently, since the support of asymptotic output measure ϖ_a^θ is only on the stable branches of the steady states characteristic and a-posteriori design method does not include any such unstable points we conclude that it still yield a better result. Moreover, the a-posteriori design method is more robust than a-priori design method since the norm of steady states stays in the accessible range for a larger range of values of the uncertain parameter.

CHAPTER 6

Conclusions

In this part of the dissertation we discussed two types of optimal design based on a probabilistic approach for uncertain system design. We sought a controller that minimizes the average performance function with respect to the uncertainty. The first approach, called a-priori design, is based on a-priori equilibrium measure for the design of the system on the basis of a-priori uncertainty information for the parameter uncertainty. However, this method may fails for the dynamical systems that have initial condition dependent uncertain behavior. The second design method, a-posteriori design, utilizes an asymptotic output measure for the design of an optimal controller based on a-posteriori uncertainty generated by the dynamics of the system itself. It is shown that the a-posteriori design approach can capture all uncertainty effects. In order to obtain the optimal controller numerically, we applied randomized algorithms for average performance synthesis for the two design methods by sampling both the uncertain parameters space and the design parameters space. Furthermore, an approach using properties of the Perron-Frobenius operator is presented to efficiently compute the invariant measure of dynamical systems with random uncertain parameters. Finally, to illustrate the approach, we applied the a-priori and a-posteriori design methods to a CSTR example system and compared the results.

Part II

Part II: Characteristics of Short-term LOLP with Wind Generation

CHAPTER 7

Introduction

Large-scale intermittent and variable wind generation in the future electric energy supply portfolio challenges the assessment of generation adequacy, especially when the penetration is high. Although the capacity provided by the wind power improves the system adequacy level to a certain extent in a long-term measure, the variability of its output can affect the system adequacy in a short-term, which imposes an immediate risk to system reliability operation [29], [67].

Loss of Load Probability (LOLP) is an important measure of generation adequacy. By definition, for a power system with both wind generation and conventional generation, for any time $t \in [0, T]$ and a given peak load l during this period of T, LOLP can be described as follows

$$LOLP(t) = \Pr\{C_T(t) + C_W(t) < l\}$$

where $C_T(t)$ is a variable representing the total available conventional generation at time t and $C_W(t)$ represents the wind generation at time t. Conventionally, controllable generation is usually modeled with a two-state process including an available state with planned capacity and an outage state with zero or reduced amount of output, while wind generation can be viewed as a stochastic process driven by the wind.

Incorporation of wind generation in LOLP calculation has been explored in several studies. For example, [34] and [6] used the Monte Carlo method to simulate hourly stochastic change of generation availability. An auto-regressive and moving average (ARMA) time series model is used in [7] to simulate hourly wind speed and available wind power in consideration of chronological characteristics. [26] and [8] presented some analytical

methods for estimation of wind output with multi-state models of wind speed.

These existing techniques mainly focus on the long-term reliability evaluation ranging from several months to several years using the stationary LOLP. In long-term LOLP calculation, wind generation has been viewed as a multi-state power plant and probabilities for each output level are time-invariant, which cannot be directly used to describe the short-term change of LOLP. Lack of an appropriate techniques to assess the dynamics of short-term LOLP makes it difficult to quantify the impact of the wind on generation adequacy.

The concerns about the impact of variable wind generation on short-term generation adequacy are expressed in several recent articles, where some new methods are also proposed. For example, [45] developed an empirical sliding window to update LOLP on an hourly basis. This approach was extended and applied to assess the fraction of system reserve that can be allocated to wind farms [44]. In [27] and [28], the authors applied a Markov chain model to represent variable wind generation in operational risk evaluation. However, it is found that the short-term LOLP converges to its steady-state value, i.e. its long-term level. This finding indicates that the information about variable and intermittent wind generation can be lost if the updating interval of LOLP is too long [33].

In this dissertation, we will introduce a method for calculation of the short-term LOLP. An instantaneous multi-state model is constructed to characterize the output of wind generation. The instantaneous state probabilities are estimated by Markov chains. Moreover, a novel analytical description of convergence time is developed and empirical formulas for calculation of convergence time are derived, which can be used to determine the appropriate period for updating LOLP and understanding its dynamic behavior. Finally, an application of the methods and measures are shown using the output profile of an actual wind farm. The discussion of the impact of short-term LOLP on generation adequacy under different wind penetration scenarios are also presented.

CHAPTER 8

Computation of Short-term LOLP with Wind Generation

8.1 Instantaneous Multi-State Wind Generation Model

A multi-state model has been used to study the impact of wind generation on LOLP calculation [26]. More specifically, wind generation is assumed to have multiple output states, while conventional generation has only two states: "on" for available state and "off" for outage state.

Due to the variability of the wind input, a time-invariant multi-state model is not suitable for short-term LOLP calculation. In order to describe the change of wind generation, an instantaneous k-state wind generation model is proposed as follows

$$C_W(t) = \begin{cases} c_w^0 & \textit{with probability } \mu_t(0) \\ c_w^1 & \textit{with probability } \mu_t(1) \\ \quad \vdots & \qquad\qquad \vdots \\ c_w^{k-1} & \textit{with probability } \mu_t(k-1) \end{cases} \tag{8.1}$$

where $c_w^j,\ j = 0,...,k-1$ are discretized states for wind generation levels while $\mu_t(j)$ are the corresponding instantaneous state probabilities for the output state at time t, which are defined by

$$\mu_t(j) = \Pr\left\{C_W(t) \in \left(c_w^j, c_w^{j+1}\right)\right\},\ j = 0,...k-1 \tag{8.2}$$

with $c_w^k = c_w^{\max}$ the maximum output of wind generation.

Usually, we rewrite $\mu_t(j),\ j = 0,...,k-1$ in a probability row-vector form, i.e. $\mu_t = [\mu_t(0),...,\mu_t(k-1)]$, to describe the distribution for different levels of wind generation at

time t.

8.2 Formulation of Short-term LOLP

The LOLP model considering only conventional generation can be computed through probability convolution. In long-term LOLP computation, the force outage rate (FOR) is used to describe the unavailability of conventional units. Since we are interested in the probability of generation outage during a short-term period T, the force outage rate is replaced by the outage replacement rate (ORR) used in PJM method [5]. Specifically, the ORR of a unit during period T is given by

$$ORR_T = \Pr\{\text{unit is out during } T\} = 1 - \exp(-\lambda T) \tag{8.3}$$

where λ is the failure rate. For a two-state conventional generation with capacity c, let ORR_T be the generation outage replacement rate, then the probabilities of available state and outage state are $1 - ORR_T$ and ORR_T, respectively. Thus the cumulative probability of having a generation capacity less than or equal to x during period T *after* adding this conventional generation, $P_T(x)$ can be calculated as follows

$$P_T(x) = ORR_T \cdot \hat{P}_T(x) + (1 - ORR_T) \cdot \hat{P}_T(x - c) \tag{8.4}$$

where $\hat{P}_T(x)$ is the cumulative probability of having a system capacity less than or equal to x *before* adding this generation unit with capacity x. Using the above equation iteratively, we can obtain a capacity outage table where the cumulative probability $P_T(x)$ is deduced directly for any given x.

Similar to the convolution of a two-state conventional generation, a time-variant cumulative probability, $P_t(x)$, of having a system capacity less than or equal to x after adding a

k-state wind generation can be found as follows

$$\begin{aligned} P_t(x) &= \Pr\{C_T(t) + C_W(t) \leq x\} \\ &= \sum_{j=0}^{k-1} \mu_t(j) \hat{P}_t(x - c_w^j) \end{aligned}$$

where $\mu_t(j)$ is the instantaneous state probability of wind generation for state j, while $\hat{P}_t(x)$ is the instantaneous cumulative probability of having conventional generation less than or equal to x before adding any wind generation. Then, the instantaneous LOLP at time $t \in [0, T]$ for peak load l can be written as

$$\begin{aligned} LOLP(t) &= P_t(l) = \sum_{j=0}^{k-1} \mu_t(j) \hat{P}_t\left(l - c_w^j\right) \\ &= \sum_{j=0}^{k-1} \mu_t(j) \hat{P}_T\left(l - c_w^j\right). \end{aligned} \tag{8.5}$$

Equation (8.5) assumes that the system has two types of supplies, conventional generation (two-states) and wind generation (k-state). The cumulative probability $\hat{P}_t(x)$ describes the probability of generation capacities less than or equal to x with only conventional generation. Since the generation outage replacement rate ORR_T is time-invariant once a certain period T is fixed, we can replace $\hat{P}_t(\cdot)$ with $\hat{P}_T(\cdot)$, which is known a-priori from the capacity outage table. Therefore, the key to the computation of instantaneous LOLP described by Equation (8.5) is to calculate the instantaneous state probabilities of wind generation, i.e., $\mu_t(j)$, for all state j at time t.

8.3 Probability Estimation by Markov Chain

In steady state LOLP calculation, the probability of the wind generation output staying at each output level can be computed using the power curve of wind turbines and the Weibull distribution of hourly wind speeds as discussed in [26] and [8]. However, since we want to compute the instantaneous transition probabilities between states, those computation

methods for long-term LOLP cannot be used for short-term LOLP estimation because the state probabilities estimated by these methods are time-invariant, thus cannot characterize the impact of variable wind on LOLP.

A number of articles have adopted Markov chains to describe the variation of wind speed, where each state represents a discrete wind speed level. In some recent studies, such as [58], a first-order Markov chain is used to generate synthetic series of wind speed and the results show that short-term dynamics by the Markov chain are very close to the actual wind speed. In [50], a method for direct generation of synthetic time series of wind power output by Markov chain Monte Carlo is proposed. Another application of Markov chains in evaluation of the reliability of distribution networks containing embedded wind generation can be found in [59]. In addition, a study reported in [28] used a Markov process to model wind generation to evaluate operational risk in a power system with high wind penetration.

Note that for a process to be represented by a Markov chain, it need to be stationary. In other words, the transition rates between different states remain constant throughout the period of interest. Since the wind speed usually has seasonal patterns, the mean and standard deviation of wind speed cannot remain constant all the time. Therefore, the wind speed cannot be described as a stationary process. In some recent studies, such as [22], authors have proposed an approach to overcome this problem of non-stationary by partitioning the annual cycle into months and model monthly wind data.

Based on the idea proposed in previous studies, we use a Markov process to model wind generation and estimate the instantaneous state probabilities in Equation (8.5). Assume the wind generation, $C_W(t)$, is a homogenous Markov process with states described by Equation (8.2). Then, the instantaneous state probabilities, $\mu_t(j)$, can be evaluated by

solving the differential equations [40]

$$\frac{d\mu_t(j)}{dt} = \sum_{\substack{i=0 \\ i\neq j}}^{k-1} \mu_t(i)\alpha_{ij} - \mu_t(j) \sum_{\substack{i=0 \\ i\neq j}}^{k-1} \alpha_{ji} \quad j = 0,...,k-1 \tag{8.6}$$

where α_{ij} is the transition rate from state i to state j. If we rewrite Equation (8.6) with row-vector μ_t, we have

$$\dot{\mu}_t^T = A \cdot \mu_t^T \tag{8.7}$$

where matrix A is given by

$$A = \begin{bmatrix} -\sum_{i=1}^{k-1} \alpha_{0i} & \alpha_{10} & \cdots & \alpha_{k-1,0} \\ \alpha_{01} & -\sum_{\substack{i=0 \\ i\neq 1}}^{k-1} \alpha_{1i} & \cdots & \alpha_{k-1,1} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{0,k-1} & \alpha_{1,k-1} & \cdots & -\sum_{i=0}^{k-2} \alpha_{k-1,i} \end{bmatrix}$$

It is shown in [40] that solving Equation (8.6) under initial condition $\mu_0(j), j = 1,2,\ldots,k-1$, the instantaneous state probability, $\mu_t(j)$, can be calculated for any time instant $t \in [0,T]$. The solution of Equation (8.7) is given by [13]

$$\mu_t^T = e^{A(t-t_0)} \cdot \mu_{t_0}^T \tag{8.8}$$

for any $t_0 \in [0,T]$. However, the practical challenge is that, in order to have high computational accuracy, a large number of states k may be needed, which makes computation complexity high in Equation (8.7).

Here we use a discrete approximation so that Equation (8.7) can be solved easily

without compromising much accuracy. Let Δt be a very small time step compared to the minimum residence time in all states of wind generation. Then the time period of interest, $[0,T]$, can be divided to N intervals of equal length $\Delta t = T/N$. During each small time interval of Δt, the state probabilities, $\mu_t(j)$, are assumed constant. Let $\hat{\mu}_n(j)$ be the probability that wind generation is in j state during the $(n+1)$th time interval, and $\hat{\mu}_N(j) = \mu_T(j)$. Denote $\hat{\mu}_n$as the row vector representation of state probabilities $\hat{\mu}_n(j)$ for all states. Thus, for a fixed Δt, a discrete form of Equation (8.8) can be obtained for interval $n = 1,\ldots,N-1$

$$\hat{\mu}_n^T = e^{A\Delta t} \cdot \hat{\mu}_{n-1}^T = \mathbb{P} \cdot \hat{\mu}_{n-1}^T \tag{8.9}$$

if initial condition $\hat{\mu}_0$ is known. Equation (8.9) represents a k-state first-order Markov chain and $e^{A\Delta t}$ is called transition probability matrix denoted as $\mathbb{P}$ characterizing the probability of going from one state to other states during Δt. Once the initial distribution $\hat{\mu}_0$ and the transition probability matrix $\mathbb{P}$ are given, the state distribution of the Markov chain for any step can be found.

For empirical study on the short-term LOLP, we propose the following formula for the estimation of the state distribution and transition probability matrix $\mathbb{P}$ based on real wind data. The method is based on matrix multiplication approach discussed in [4].

Using Taylor series, we have

$$\mathbb{P} = e^{A\Delta t} = I + A\Delta t + \frac{(A\Delta t)^2}{2!} + \frac{(A\Delta t)^3}{3!} + \ldots$$

Since Δt is small, first two terms in Taylor series are used as an approximation for Equation (8.9), that is

$$\begin{aligned} \hat{\mu}_n &= \hat{\mu}_{n-1} \cdot (I + A^T \Delta t) = \hat{\mu}_{n-1} \cdot \hat{\mathbb{P}} \\ &= \hat{\mu}_0 \cdot \hat{\mathbb{P}}^n \end{aligned} \tag{8.10}$$

where $\hat{\mathbb{P}}$ is a $k \times k$ stochastic matrix as an approximation of $\mathbb{P}$ with entries $\hat{p}_{ij}$ defined by

$$\hat{p}_{ij} = \begin{cases} 1 - \sum_{\substack{j=0 \\ j \neq i}}^{k-1} \alpha_{ij} \cdot \Delta t & for\ i = j \\ \alpha_{ij} \cdot \Delta t & for\ i \neq j \end{cases} \tag{8.11}$$

Note that the transition rate between two states describes the frequency of occurrence of a transition per unit time independent from the time step Δt. For a metered wind generation data over time period T_t, the maximum likelihood estimator for the transition rate α_{ij} from state i to state j is given by

$$\alpha_{ij} = \frac{n_{ij}}{T_t}$$

where n_{ij} is the number of transitions of samples from state i to state j observed in the time period, T_t.

Since the time period of interest $[0, T]$ is divided into N intervals with equal length of Δt, there exists a unique $n \in \{0, 1, \ldots, N-1\}$ with which the corresponding row vector of state probabilities, $\hat{\mu}_n$, that can characterize the instantaneous state distribution, μ_t for any $t \in [0, T]$. In other words, we have

$$\mu_t = \hat{\mu}_{\lfloor \frac{t}{\Delta t} \rfloor} = \hat{\mu}_0 \cdot \hat{\mathbb{P}}^{\lfloor \frac{t}{\Delta t} \rfloor} \tag{8.12}$$

where $\lfloor \cdot \rfloor$ is the floor function mapping a real number to the next smallest integer.

The instantaneous LOLP described by Equation (8.5) can be written by using the approximate discrete time state probabilities $\hat{\mu}_n$as,

$$LOLP(t) = \sum_{j=0}^{k-1} \hat{\mu}_{\lfloor \frac{t}{\Delta t} \rfloor}(j) \hat{P}_T \left(l - c_w^j \right). \tag{8.13}$$

Note that $\hat{\mu}_{\lfloor \cdot \rfloor}$is the estimated instantaneous state probability by Equation (8.12), and $\hat{P}_T(\cdot)$

is the cumulative probability of the availability of conventional generation that can be computed by (8.4). Hence, once the time step Δt is determined and the transition matrix $\hat{\mathbb{P}}$ is constructed by (8.11), the instantaneous LOLP during the time period $[0, T]$ for a given initial wind generation condition can be obtained by Equation (8.13).

CHAPTER 9

Time Period for Updating Short-term LOLP

9.1 Estimation of Convergence Time

Figure 9.1 shows a short-term LOLP profile for a 6 hour time period with different initial conditions. It is observed that different initial conditions can cause the different trajectories of short-term LOLP, since the initial conditions of wind power affect the instantaneous state probabilities according to Equation (8.12). Moreover, it is also observed that the convergences of instantaneous LOLP under both high and low initial wind speeds: $LOLP(t)$ will converge to a steady state if the time period T for estimation is long enough. For example, as shown by the diamond marked line, if the estimation period is $T = 200$ min, the $LOLP(t)$ with a high initial wind speed will increase to its stationary level 0.054. On the other hand, as shown by the square marked line, $LOLP(t)$ with a low initial wind speed will decrease to its steady state level in about $T = 180$ min.

The reason for the convergence of short-term LOLP is due to the fact that the instantaneous state distribution converges. In particular, as $t \to \infty$, it can be shown that the instantaneous state distribution μ_t will converge to the stationary distribution π of the Markov chain. The corresponding steady state LOLP is given by

$$LOLP(\infty) = \sum_{j=0}^{k-1} \pi(j) \hat{P}_T \left(l - c_w^j\right). \tag{9.1}$$

Since the stationary distribution π characterizes the long-term behavior of a Markov chain, Equation (3.20) also provides another way to look at long-term LOLP.

Once the LOLP converges, it will not change significantly as the time period further

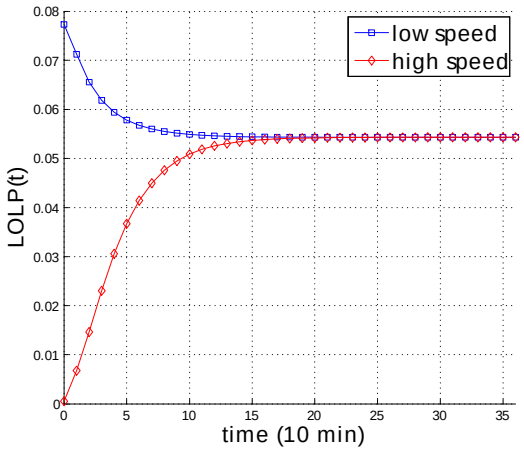

Figure 9.1: Short-term LOLP

increases. This implies that the short-term effect of wind generation will be lost if the period of estimation is too long. For example, as shown in Figure 9.1, for a given initial wind speed, instantaneous $LOLP(t)$ with $t \in [0, T]$ for a period of $T = 200$ min is almost the same as the one estimated for $T = 350$ min. In order to have a better assessment, it is necessary to update the LOLP with a new initial wind generation condition before it converges to the steady-state.

We found that the appropriate updating frequency can be found by applying some properties of Markov chain. It is known that instantaneous state probabilities of an ergodic Markov chain converge to its steady state distribution as time increases, and the corresponding convergence rate is determined by the absolute value of the second largest eigenvalue of the transition matrix [55]. More specifically, if the absolute value of the second largest eigenvalue is small, the short-term LOLP for different initial wind power will quickly converge to the steady state LOLP. Therefore, choosing a suitable time period $[0, T]$ is important to the calculation, since it will impact the accuracy of assessment of generation adequacy.

In general, for short-term LOLP calculation, we have a maximum time period $T_{\max}$, say 12 hours. Moreover, we define the *convergence time* of instantaneous LOLP as the shortest

time period T when the difference between $LOLP(T)$ and $LOLP(\infty)$ stays within a certain tolerance level, which is the solution to the following optimization problem

$$\min\ T \quad \text{s.t. } T \leq T_{\max},\ |LOLP(T) - LOLP(\infty)| \leq \delta \tag{9.2}$$

where $T_{\max}$ is the maximum time period for the short-term LOLP calculation and δ is the given tolerance level.

The optimal solution of (9.2), i.e. convergence time of instantaneous LOLP, is denoted as T^*. The tolerance level used in this study is given by

$$\delta = \alpha \cdot |LOLP(0) - LOLP(\infty)| \tag{9.3}$$

where $\alpha \in (0,1)$ is a coefficient.

According to equations (8.5) and (3.20), we define the discrete form of convergence time N^* with time step Δt

$$\min\ N \quad \text{s.t. } \left| \sum_{j=0}^{k-1} \hat{P}_{N\cdot\Delta t}\left(l - c_w^j\right) \cdot \left(\hat{\mu}_N(j) - \pi(j)\right) \right| \leq \delta \tag{9.4}$$

where $\hat{P}_{N\cdot\Delta t}(\cdot)$ is the cumulative probability of conventional generation, $\hat{\mu}_N(j)$ is the approximated state probability of wind generation at level j in step N, and π is the stationary distribution of the Markov chain. The discrete convergence time is the optimal solution N^* to the problem described by Equation (9.4). The corresponding convergence time is $T^* = N^* \cdot \Delta t$.

Definition 19 *We refer the optimal solution N^* as the* time window *or* window.

In order to obtain a more accurate results in practical estimation, numerous states of wind generation might be needed. Therefore, it could be very challenging and time-consuming to solve for the window N^* through iterative simulation. In order to address

this issue, we will propose a simplified method for quick assessment of the convergence time or window using the properties of ergodic Markov chains.

9.2 Empirical Formula

We found that the convergence properties of an ergodic Markov chain can be used to simplify the numerical problem described by Equation (9.4). A Markov chain is *ergodic* if the chain is aperiodic and irreducible. Let $\mathbb{P}$ be the transition matrix of a k-state Markov chain and $\lambda_0, \lambda_1, \ldots, \lambda_{k-1}$ be the eigenvalues of $\mathbb{P}$. Then, the stationary distribution π of the chain satisfies the equation [55]

$$\pi\mathbb{P} = \pi$$

The chain is ergodic if and only if π is unique which is equivalent to the condition that there is a unique eigenvalue equal to 1.

We notice that in this case the magnitude of all the remaining eigenvalues of $\mathbb{P}$ are strictly less than one, i.e. if $\lambda_0 = 1$ then $|\lambda_i| < 1$ for $i = 1, .., k-1$. This property allows us to use the unique stationary distribution π to define a metric, so that the distance between probability distributions of the Markov chain at any time instant and the its stationary distribution can be characterized and then, used to simplify Equation (9.4).

Definition 20 *For two probability vectors μ, π on a finite state space X, the variation distance between these two probability vectors is defined as follows,*

$$||\mu - \pi||_{var} = \max_{A \subset X} |\mu(A) - \pi(A)| = \frac{1}{2} \sum_{j \in X} |\mu(j) - \pi(j)|.$$

The proof of the following result can be found in [23].

Theorem 11 *Let $\mathbb{P}$ be an ergodic transition matrix on a finite state space X and let π be its stationary distribution. Then for any initial distribution μ_0, the distribution of the chain*

at time n, denoted as μ_n, satisfies the following inequality:

$$||\mu_n - \pi||_{var}^2 \leq \frac{1}{4}\lambda_*^{2n} \sum_{j \in X} \frac{(\mu_0(j) - \pi(j))^2}{\pi(j)}$$

where λ_* is the absolute value of the second largest eigenvalue of $\mathbb{P}$.

For an ergodic Markov chain, the stationary distribution is unique. Therefore, the variation distance in Definition 20 can be used to measure the convergence time between the distribution $\hat{\mu}_N$ and the stationary distributions π, i.e., the two distributions used in computing the LOLP convergence time.

In many practical problems such as those associated with variable wind generation, the transition matrix $\mathbb{P}$ has distinct eigenvalues and is therefore diagonalizable. For a diagonalizable transition matrix, we found that estimation of the variation distance by Theorem 11 can be further simplified to characterize the absolute difference between the current state distribution of Markov chain μ_n and the stationary distribution π for any point in the state spaces. This simplification is discussed in the following proposition.

Proposition 12 *For an ergodic Markov chain defined on a finite state space X, assume the initial state distribution μ_0 is given and transition matrix $\mathbb{P}$ is diagonalizable with distinct eigenvalues $\lambda_{0,\ldots,}\lambda_{k-1}$, and define $\lambda_0 = 1$, $\lambda_* = \max_{1 \leq j \leq k-1} |\lambda_j|$. Then, the absolute difference between the state distribution μ_n at time n and the stationary distribution π for point $j \in X$ satisfies:*

$$\begin{aligned} |\mu_n(j) - \pi(j)| &\leq \sum_{i=1}^{k-1} |a_i v_i(j)| \cdot |\lambda_i|^n \\ &\leq \left(\sum_{i=1}^{k-1} |a_i v_i(j)| \right) \cdot \lambda_*^n \end{aligned}$$

where $v_0, .., v_{k-1}$ is a basis of left eigenvectors of the transition matrix $\mathbb{P}$ corresponding to

$\lambda_{0,...,}\lambda_{k-1}$, *and* $a_0,...,a_{k-1}$ *are the unique coefficients such that*

$$\mu_0 = a_0 v_0 + ... + a_{k-1} v_{k-1}.$$

Proof. For any eigenvalue λ_i of the transition matrix $\mathbb{P}$ with the corresponding left eigenvector v_i, we have

$$v_i \mathbb{P} = \lambda_i v_i$$

Furthermore, by iteration, for any $n \geq 0$, we also have

$$v_i \mathbb{P}^n = \lambda_i^n v_i$$

Assume $\lambda_0 = 1$ and the initial distribution μ_0 is given by

$$\mu_0 = a_0 v_0 + a_1 v_1 + ... + a_{k-1} v_{k-1}$$

then, for any $n \geq 0$, we have

$$\mu_n = \mu_0 \mathbb{P}^n = a_0 v_0 + a_1 v_1 (\lambda_1)^n + ... + a_{k-1} v_{k-1} (\lambda_{k-1})^n.$$

Since $|\lambda_1|, ..., |\lambda_{k-1}| < 1$,

$$\pi = \lim_{n \to \infty} \mu_n = a_0 v_0.$$

For a point $x \in X$,

$$\begin{aligned} |\mu_n(x) - \pi(x)| &= \left| a_1 v_1(x) \lambda_1^n + ... + a_{k-1} v_{k-1}(x) \lambda_{k-1}^n \right| \\ &\leq \sum_{i=1}^{k-1} |a_i v_i(x)| \cdot |\lambda_i|^n \\ &\leq \left(\sum_{i=1}^{k-1} |a_i v_i(x)| \right) \cdot \lambda_*^n \end{aligned}$$

where $\lambda_* = \max_{1\le j\le k-1} |\lambda_j|$. ■

Now, the time window N^* for the empirical estimation of the convergence time of the short-term LOLP will be derived using these analytical results. Note that the time period T affects the outage replacement rate ORR_T and cumulative probability $\hat{P}_T$ in Equations (8.3) and (8.5). It is obvious that the following condition holds

$$\hat{P}_T(x) \le \hat{P}_{T_{\max}}(x) \tag{9.5}$$

for any given x.

Using Equation (9.5) and Theorem 11, we can have

$$\begin{aligned} & \left| \sum_{j=0}^{k-1} \hat{P}_{N\cdot\Delta t}\left(l - c_w^j\right) \cdot \left(\hat{\mu}_N(j) - \pi(j)\right) \right| \\ \le & \sum_{j=0}^{k-1} \hat{P}_{T_{\max}}\left(l - c_w^j\right) \cdot ||\hat{\mu}_N - \pi||_{var} \\ \le & \frac{1}{2} \sum_{j=0}^{k-1} \hat{P}_{T_{\max}}\left(l - c_w^j\right) \left(\sum_{j=0}^{k-1} \frac{\left(\hat{\mu}_0(j) - \pi(j)\right)^2}{\pi(j)} \right)^{\frac{1}{2}} \lambda_*^N \end{aligned} \tag{9.6}$$

The smallest N, i.e. window N^*, is approximated by letting the upper bound of (9.6) equal to the boundary δ and the approximate window $\tilde{N}^*$ is given by

$$\tilde{N}^* = \left\lfloor \log_{\lambda_*} \left(\frac{2\delta}{\sum_{j=0}^{k-1} \hat{P}_{T_{\max}}\left(l - c_w^j\right) \left(\sum_{j=0}^{k-1} \frac{\left(\hat{\mu}_0(j) - \pi(j)\right)^2}{\pi(j)} \right)^{\frac{1}{2}}} \right) \right\rfloor \tag{9.7}$$

where $\lfloor \cdot \rfloor$ is the floor function.

By Proposition 12, if the transition matrix $\mathbb{P}$ is diagonalizable, we have

$$\tilde{N}^* = \left\lfloor \log_{\lambda_*} \left(\frac{\delta}{\sum_{j=0}^{k-1 \wedge} P_{T_{\max}} \left(l - c_w^j \right) \left(\sum_{i=1}^{k-1} |a_i v_i(j)| \right)} \right) \right\rfloor. \tag{9.8}$$

Since the transition matrix $\mathbb{P}$ is diagonalizable, the left eigenvectors v_i, $i = 0, .., k-1$ are independent and the coefficients $a_0, ..., a_{k-1}$ can be obtained by

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{bmatrix} = V^{-1} \cdot \begin{bmatrix} \mu_0(0) \\ \mu_0(1) \\ \vdots \\ \mu_0(k-1) \end{bmatrix}$$

where $V = [v_0, ..., v_{k-1}] \in \mathbb{R}^{k \times k}$ is nonsingular.

The $\tilde{N}^*$ found in Equations (9.7) and (9.8) is a suboptimal solution of (9.4) and the corresponding convergence time $\tilde{T}^* = \tilde{N}^* \cdot \Delta t$ is slightly smaller than the actual optimal solution of (9.2). By Equations (9.7) and (9.8), it is convenient to estimate the time window for updating LOLP in empirical analysis.

In the next chapter, the analytical results obtained will be used to study the dynamic behavior of the short-term LOLP and its convergence time based on an actual wind profiles in different situations.

CHAPTER 10

Application

This section presents three case studies on the characteristics of the short-term LOLP based on wind data measured at 10 min time interval at a wind farm located in northwest Oklahoma, with the concepts and the methodology developed in previous sections.

Case 1 only considers conventional generation in estimation of short-term LOLP. Case 2 studies the short-term LOLP with wind generation added. The wind generation profile in winter and in summer are considered separately to overcome the problem of non-stationary due to seasonal patterns, as suggested in [22]. Case 3 studies the impact of wind penetration level on the convergence time.

10.1 Case 1: LOLP with Conventional Generation Only

Assume that the power system has 6 conventional generating units, each with capacity of 250 MW. The time horizon of interest, T, is set to be 6 hours. The generation failure rate is $\lambda = 0.0139\text{h}^{-1}$. Equation (8.3) gives an outage replacement rate $ORR_T = 0.08$. The convolution algorithm can be used to generate the capacity outage table by adding one generator at a time. The results are shown in Table 10.1.

If the peak load l is assumed to be 1000 MW, the LOLP with conventional generation is 0.077, which remains constant over time period T.

Available Gen.	Forced Outage Gen.	Cumulative Probability
1500 MW	0 MW	1.000000000
1250 MW	250 MW	0.39364500
1000 MW	500 MW	0.07728587
750 MW	750 MW	0.00851214
...	...	...

Table 10.1: LOLP with conventional generation

10.2 Case 2: Short-term LOLP with Wind Generation

In this case, 500 MW wind generation is added to the conventional generation portfolio described in Case 1. The wind generation consists of 100 wind turbines, each with capacity of 5 MW. The outage rate of wind turbine is ignored in this short-term performance study.

The time series of wind generation is based on a real wind profiles at a wind farm located in northwest Oklahoma measured at 10 min time interval. A simple cubic function below is assumed to describe the power curve, given by

$$P_t = \begin{cases} 0 & w_t \leq w_{ci}, w_t > w_{co} \\ P_\tau \cdot \left(\frac{w_t - w_{ci}}{w_\tau - w_{ci}}\right)^3 & w_{ci} < w_t \leq w_\tau \\ P_\tau & w_\tau < w_t \leq w_{co} \end{cases}$$

where P_t is the power output of wind generation at time t, w_t is wind speed at time t, w_τ and P_τ are rated wind speed and power output, w_{ci} and w_{co} are the cut-in and cut-out wind speeds. In the study, we choose $w_{ci} = 3\ m/s$, $w_\tau = 14\ m/s$, $w_{co} = 25\ m/s$. The generated wind power output time series is shown in Figure 10.1. Figure 10.1 (a) shows that the wind generation output in winter (Dec, 2009), while Figure 10.1 (b) shows that in summer (Aug, 2009).

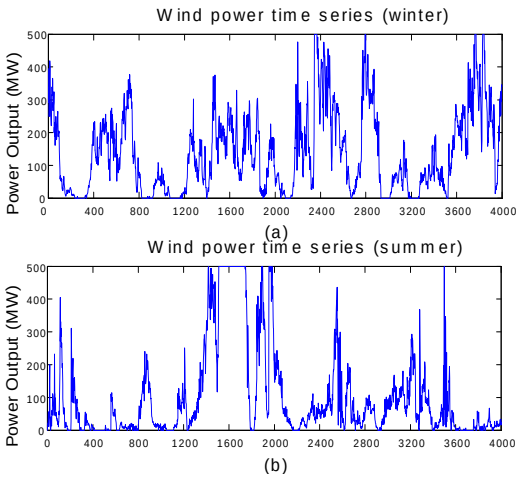

Figure 10.1: Time series of wind generation

In order to calculate the short-term LOLP numerically, we discretized the time series into 10 states with equal range between the adjacent states, i.e., $c_w^0 = 0$ MW, $c_w^1 = 50$ MW, ..., $c_w^9 = 450$ MW. The time step chosen to be $\Delta t = 1$ min and we estimated the transition rate between different wind levels from the time series to construct a 10×10 transition matrix $\mathbb{P}$ for the Markov chain.

In this study, we found that all eigenvalues of $\mathbb{P}$ matrix are distinct, which means it is diagonalizable. Therefore, the assumption for Proposition 12 is valid for the study.

The variation of short-term LOLP depends on the initial state probability of wind generation, as described in Equation (8.12). In this study, three different initial state distributions are presented for low, medium and high initial wind generation levels, given by $\hat{\mu}_0^L = [1,0,...,0]$, $\hat{\mu}_0^M = [0...,1,...0]$, $\hat{\mu}_0^H = [0,...0,1]$. The result of the short-term LOLP with wind generation in winter is shown in Figure 10.2.

As shown in Figure 10.2, the short-term LOLP converges to its stationary level regardless the initial levels, i.e., $LOLP(\infty) = 0.0378$. Note that for 6 hours time period, LOLP has already been in its steady state for about 1 hours. This means that, if the time period for LOLP updating is 6 hours, it will have lost all short-term information. Therefore, LOLP has to be updated much more frequently than 6 hours.

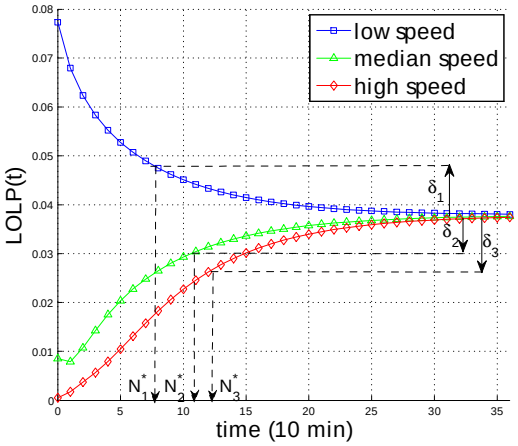

Figure 10.2: Short-term LOLP in winter

The variation ranges for three different initial distributions are: $\delta_1 = 0.0099$, $\delta_2 = 0.0073$, $\delta_3 = 0.0093$, with the tolerance factor equal to 0.25 in Equation (9.3). To simplify the problem, we assume $T_{\max} = T = 6$h. The corresponding window N_1^*, N_2^*, N_3^* are also shown in Figure 10.2. The approximated windows obtained by (9.8) are: $\tilde{N}_1^* = 52$, $\tilde{N}_2^* = 87$, $\tilde{N}_3^* = 110$ for low, medium and high initial wind conditions, respectively. The corresponding approximated convergence times are: $\tilde{T}_1^* = 52$ min, $\tilde{T}_2^* = 87$ min, $\tilde{T}_3^* = 110$ min.

The result of the short-term LOLP with wind generation in summer is shown in Figure 10.3.In this case, the states of Markov chain are the same as that for wind generation in winter while the transition matrix $\mathbb{P}$ is constructed using time series in Figure 10.1 (b). Compared to that in Figure 10.2, the short-term LOLP in summer converges to its steady state level at a slower rate. The time windows for LOLP updates are shown in Figure 10.3. Using the same tolerance level α, the approximated convergence times are: $\tilde{T}_1^* = 64$ min, $\tilde{T}_2^* = 92$ min, $\tilde{T}_3^* = 131$ min.

The time window or convergence time for updating LOLP in summer is larger than that in winter due to the "smoother" time series of wind generation shown in Figure 10.1 (b) compared with Figure 10.1 (a). It means that for a fast-changing intermittent wind

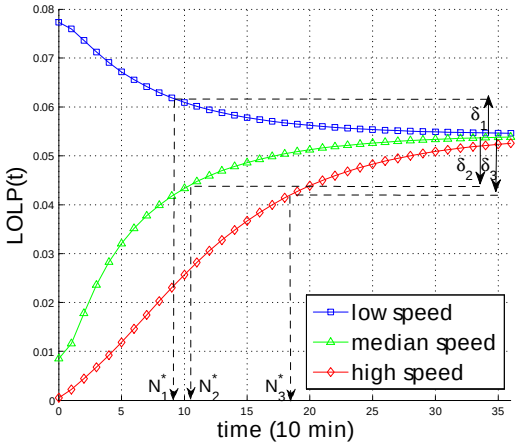

Figure 10.3: Short-term LOLP in summer

generation, LOLP needs to be updated much more frequently to better reflect the actual system generation adequacy condition in short-term.

It is also shown that the steady state LOLP in Figure 10.2 is better than that in Figure 10.3 due to the fact that the wind generation in winter remains at a high level for longer time than in summer. In other words, the wind generation in winter has more contribution to generation adequacy than in summer for the time period of this case study.

Based on the observations in Case 2 study, it is also suggested that although the capacity benefit of wind generation is better in winter in a long-term, the LOLP needs to be updated more frequently to reflect the actual generation adequacy condition.

10.3 Case 3: Impact of Wind Penetration Level

Case 3 studies the impact of wind penetration level on the convergence time of short-term LOLP by changing the number of wind turbines in wind farms. The wind penetration is defined as the percentage of total wind generation over total system generation. The results are shown in Figure 10.4.

Figure 10.4 indicates that the impact of wind penetration level on the short-term LOLP is non-linear. By increasing the penetration, the time window decreases, which implies that

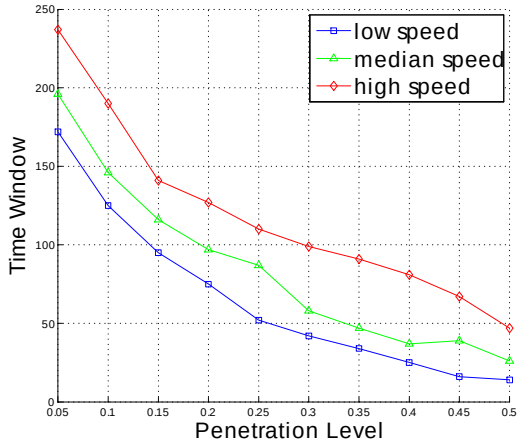

Figure 10.4: Impact of wind penetration level

we should update the short-term LOLP more frequently. Note that when the penetration level of wind generation is increased, the impact of variable wind generation on short-term LOLP becomes larger, thus more frequent update of LOLP is desired.

The finding and suggestion presented in above three case studies show that the proposed method and measures improve the understanding about the impact of variable wind generation on short-term generation adequacy, and provide quantitative support for power system reliability operation.

CHAPTER 11

Conclusions

In this part of the dissertation, a method for understanding the short-term impact of wind generation on LOLP is proposed. The short-term LOLP is calculated using an instantaneous multi-state wind generation model. A discrete method based on a corresponding Markov chain is proposed for estimation of short-term LOLP. Furthermore, by using properties of an ergodic Markov chains, several methods for determining the appropriate time interval for updating the short-term LOLP are provided. Finally, the methods are applied to a study of the short-term LOLP under different initial wind generation levels, different wind generation output profiles and different penetration rates.

The novel measures of LOLP convergence time presented in this paper can be used to better understand the impact of wind generation on system reliability and are very useful for short-term generation adequacy assessment. More specifically, the methods and measures enable a quantitative estimation of convergence characteristics of short-term LOLP, which provides scientific support for the system operator to understand the accountability of wind generation.

Additional enhancements could be made in our future research. For example, the failure rate of the wind turbines could be taken into consideration. Also, due to the daily and seasonality patterns of wind energy, different transition matrices for different time periods may be needed for more accurate results.

Bibliography

[1] E. Antonsson and K. N. Otto, "Imprecision in engineering design," *Special Combined Issue of the Transactions of the ASME commemorating the 50th anniversary of the Design Engineering Division of the ASME*, vol. 117, pp. 25–32, 1995.

[2] L. Arnold, *Random Dynamical Systems*. Berlin: Springer, 1998.

[3] J. Axelsson, "Model based systems engineering using a continuous-time extension of the unified modeling language (uml)," *Systems Engineering*, vol. 5, no. 3, pp. 165–179, 1995.

[4] R. Billinton and R. Allan, *Reliability Evaluation of Engineering Systems*. Springer, 1992.

[5] ——, *Reliability evaluation of power systems*. Plenum Press, 1996.

[6] R. Billinton and H. Chen, "Assessment of risk-based capacity benefit factors associated with wind energy conversion systems," *IEEE Transactions on Power Systems*, vol. 13, pp. 1191–1196, 1998.

[7] R. Billinton, H. Chen, and R. Ghajar, "Time-series models for reliability evaluation of power systems including wind energy," *Microelectron. Reliab.*, vol. 36, pp. 1253–1261, 1996.

[8] R. Billinton and A. A. Chowdhury, "Incorporation of wind energy conversion system in conventional generating capacity adequacy assessment," *IEE Proceedings-C*, vol. 139, pp. 47–56, 1992.

[9] V. Blondel and J. Tsitsiklis, "A survey of computational complexity results in systems and control," *Automatica*, vol. 36, pp. 1249–1274, 2000.

[10] G. Calafiore, F. Dabbene, and R. Tempo, "Randomized algorithms for probabilistic robustness with real and complex structured uncertainty," *IEEE Transactions on Automatic Control*, vol. 45, pp. 2218–2235, 2000.

[11] ——, "Randomized algorithms in robust control," *Proceedings of the 2003 IEEE CDC*, 2003.

[12] ——, "A survey of randomized algorithms for control synthesis and performance verification," *Journal of Complexity*, vol. 23, pp. 301–316, 2007.

[13] C.-T. Chen, *Linear System Theory and Design*. USA: Oxford University Press, 1998.

[14] C.McGreavy and J.M.Thornton, "Stability studies of single catalyst particles," *Chem.Eng.J.*, vol. 1, pp. 296–301, 1970.

[15] A. R. Daniel and A. A. Chen, "Stochastic simulation and forecasting of hourly average wind speed sequences in jamaica," *Solar Energy*, vol. 46, pp. 1–11, 1991.

[16] M. Dellnitz and A. Hohmann, "The computation of unstable manifolds using subdivision and continuation," *Nonlinear Dynamical Systems and Chaos*, vol. 19, pp. 449–459, 1996.

[17] ——, "A subdivision algorithm for the computation of unstable manifolds and global attractors," *Numerische Mathematik*, vol. 75, pp. 293–317, 1997.

[18] M. Dellnitz and O. Junge, "An adaptive subdivision technique for the approximation of attractors and invariant measure," *Computing and Visualization in Science*, vol. 1, pp. 63–68, 1998.

[19] ——, "On the approximation of complicated dynamical behavior," *SIAM Journal on Numerical Analysis*, vol. 36, pp. 491–515, 1999.

[20] D.S.Cohen and B.J.Matkowsky, "On inhibiting runaway in catalytic recator," *SIAM J.Appl.Math.*, vol. 35, pp. 307–314, 1978.

[21] M. Eidenschink, *Exploring Global Dynamics: A Numerical Algorithm Based on the Conley Index Theory, Ph.D. Thesis*. Atlanta, USA: Georgia Institute of Technology, 1995.

[22] L. Fawcett and D. Walshaw, "Markov chain models for extreme wind speeds," *Environmetrics*, vol. 17, pp. 795–809, 2006.

[23] 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.

[24] G. B. Folland, *Real analysis: modern techniques and their applications*. New York: Wiley, 1999.

[25] G. Froyland and M. Dellnitz, "Detecting and locating near-optimal, almost-invariant sets and cycles," *SIAM J. Sci. Comput.*, vol. 24, pp. 1839–1863, 2003.

[26] P. Giorsetto and K. F. Utsurogi, "Development of a new procedure for reliability modeling of wind turbine generators," *IEEE Transactions on Power Apparatus and Systems*, vol. 102, pp. 134–143, 1983.

[27] E. M. Gouveia and M. A. Matos, "Operational reserve of a power system with a large amount of wind power," *International Conference on Probabilistic Methods Applied to Power Systems*, 2004.

[28] ——, "Evaluating operational risk in a power system with a large amount of wind power," *Electric Power Systems Research*, vol. 79, no. 5, pp. 734–739, 2009.

[29] R. Gross, *The costs and impacts of intermittency: An Assessment of the Evidence on the Costs and Impacts of Intermittent Generation on the British Electricity Network*. London: Imperial College, 2006.

[30] P. R. Halmos, *Measure Theory*. New York: Springer Verlag, 1974.

[31] J. Halton, *On the Generation of Random Sequences in Frechet Spaces*. Upton, NY: Brookhaven National Laboratory, 1965, vol. Rep. AMD 374/BBNL 8996.

[32] ——, "A retrospective and prospective survey of the monte carlo method," *SIAM Review*, vol. 12, no. 1, pp. 1–63, 1970.

[33] J. Jiang, C. Lin, and T. Runolfsson, "A study of short-term impact of wind generation on lolp," *Proceedings of the IEEE Transmission and Distribution Conference and Exposition*, 2010.

[34] R. Karki and R. Billinton, "Cost-effective wind energy utilization for reliable power supply," *IEEE Transactions on Energy Conversion*, vol. 19, pp. 435–440, 2004.

[35] P. Khargonekar and A. Tikku, "Randomized algorithms for robust control analysis and synthesis have polynomial complexity," *Proceedings of the 1996 IEEE CDC*, 1996.

[36] Y. Kifer, "General random perturbations of hyperbolic and expanding trans- formations," *J. Analyse Math*, vol. 29, pp. 47–111, 1986.

[37] ——, *Random Perturbations of Dynamical Systems*. New York: Birkhauser, 1988.

[38] K.Yosida, *Functional Analysis*. Springer, 1980.

[39] A. Lasota and M. C. Mackey, *Chaos, Fractals and Noise*. New York: Springer-Verlag, 1994.

[40] G. Levitin, *Universal Generating Function and its Applications*. London: Springer, 1985.

[41] N. Metroplois and S. Ulam, "The monte carlo method," *J. Amer. Statist. Assoc.*, vol. 44, pp. 333–341, 1949.

[42] I. Mezic and T. Runolfsson, "Uncertainty analysis of complex dynamical systems," in *Proceedings of the 2004 American Control Conference, Boston, MA*, 2004.

[43] ——, "Uncertainty propagation in dynamical systems," *Automatica*, vol. 44, pp. 3003–3013, 2008.

[44] M. R. Milligan, "A chronological reliability model to assess operating reserve allocation to wind power plants," *Proceedings of the European Wind Energy Conference*, 2001.

[45] ——, "A sliding window technique for calculating the system lolp contributions of wind powers plants," *Proceedings of the American Wind Energy Association Windpower Conference*, 2001.

[46] H. Niederreiter, "Random number generation and quasi-monte carlo methods," in *CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia*, 1992.

[47] J. Osborn, "Spectral approximation for compact operators. mathematics of compuation," *Mathematics of Compuation*, vol. 29, pp. 712–725, 1975.

[48] K. Otto and E. Antonsson, "Design parameter selection in presence of noise," *Research in engineering Design*, vol. 6, no. 4, pp. 234–246, 1994.

[49] P. G. P. Acworth, M. Broadie, "A comparison of some monte carlo and quasi monte carlo techniques for option pricing," in *Monte Carlo and Quasi-Monte Carlo Methods: Proceedings of a Conference at the University of Salzburg*, 1996, pp. 1–18.

[50] G. Papaefthymiou and B. Klockl, "Mcmc for wind power simulation," *IEEE Transactions on Energy Conversion*, vol. 23, pp. 234–240, 2008.

[51] A. Papageorgiou and J. Traub, "Faster evaluation of multidimensional integrals," *Computers in Physics*, vol. 6, no. 4, pp. 574–578, 1997.

[52] T. S. Parker and L. O. Chua, *Practical Numerical Algorithms for Chaotic Systems*. New York: Springer, 1989.

[53] S. Pliska and e. M. Dempster, *Mathematics of Derivative Securities*. Cambridge, UK: Cambridge University Press, 1997, vol. S. Paskov, New Methodologies for Valuing Derivatives.

[54] D. Pollard, *Convergence of Stochastic Processes*. New York: Springer-Verlag, 1984.

[55] J. S. Rosenthal, "Convergence rates for markov chains," *SIAM Review*, vol. 37, pp. 387–405, 1995.

[56] H. L. Royden, *Real Analysis*. London: Macmillan, 1968.

[57] T. Runolfsson, "Optimal design of uncertainty complex dynamical systems," *Proceedings of the 2004 IEEE CDC*, 2004.

[58] A. Sahin and Z. Sen, "First-order markov chain approach to wind speed modeling," *Wind Engineering and Industrial Aerodynamics*, vol. 89, pp. 263–269, 2001.

[59] F. Sayas, *Reliability assessment of distribution networks containing embedded wind generation, Ph.D. Dissertation*. United Kindom: University of Manchester, 1996.

[60] R. Suri and K. Otto, "Manufacturing system robustness through integrated modeling," *Journal of Mechanical Design*, vol. 123, no. 4, pp. 630–636, 2001.

[61] Z. B. Tang, "Adaptive partitioned random search to global optimization," *IEEE Transactions on Automatic Control*, vol. 39, pp. 2235–2244, 1994.

[62] R. Tempo, E. W. Bai, and F. Dabbene, "Probabilistic robustness analysis: explicit bounds for the minimum number of samples," *Systems and Controls Letters*, vol. 30, pp. 237–242, 1997.

[63] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems*. New York: Springer, 2004.

[64] R. Tempo and H. Ishii, "Monte carlo and las vegas randomized algorithms for systems and control: an introduction," *European Journal of Control*, vol. 13, pp. 189–203, 2007.

[65] J. Traub and S. Paskov, "Faster valuation of financial derivatives," *Journal of Portfolio Management*, vol. 22, no. 1, pp. 113–120, 1995.

[66] J. Traub and A. Werschultz, *Complexity and Information*. Cambridge, UK: Cambridge University Press, 1998.

[67] B. C. Ummels, *Integration of wind power in the Liberalized Dutch Electricity Market*. Wiley Interscience, 2006.

[68] A. Uppal, W. H. Ray, and A. B. Poore, "On the dynamic behavior of continuous stirred tank reactor," *Chem. Eng. Sci.*, vol. 29, pp. 967–985, 1974.

[69] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1996.

[70] M. Vidyasagar, "Statistical learning theory and randomized algorithms for control," *IEEE Control Systems Magazine*, vol. 18, pp. 69–85, 1998.

[71] ——, "Randomized algorithms for robust controller synthesis using statistical learning theory," *Automatica*, vol. 37, pp. 1515–1528, 2001.

[72] ——, *Learning and Generalization: With Applications to Neural Networks*. New York: Springer-Verlag, 2002.