MODELING AND MANAGEMENT OF EPISTEMIC UNCERTAINTY FOR MULTIDISCIPLINARY SYSTEM ANALYSIS AND DESIGN

By

KAIS ZAMAN

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Civil Engineering

August, 2010

Nashville, Tennessee

Approved:

Dr. Sankaran Mahadevan

Dr. Mark McDonald

Dr. Mark Ellingham

Dr. Prodyot K. Basu

ABSTRACT

The role of uncertainty management is increasingly being recognized in the design of complex systems that require multi-level multidisciplinary analyses. Most previous studies in this direction have only dealt with aleatory uncertainty (i.e., natural or physical variability). However, various modeling errors, assumptions and approximations, measurement errors, and sparse and imprecise information introduce additional epistemic uncertainty in model prediction. Therefore, an approach to multidisciplinary uncertainty analysis and system design that addresses both aleatory and epistemic uncertainty is needed. The objective of this dissertation is to develop a methodology that provides decision support to engineers for multidisciplinary design and analysis of systems under aleatory uncertainty (i.e., natural or physical variability) and epistemic uncertainty (due to sparse and imprecise data).

Specifically, the dissertation accomplishes this objective through: (1) Development of a probabilistic approach for the representation of epistemic uncertainty; (2) Development of a probabilistic framework for the propagation of both aleatory and epistemic uncertainty; (3) Development of formulations and algorithms for design optimization under aleatory and epistemic uncertainty, from the perspective of system robustness and reliability; (4) Development of a framework for uncertainty propagation in multidisciplinary system analysis; and (5) Development of formulations and algorithms for design optimization under aleatory and epistemic uncertainty for multidisciplinary systems, from the perspective of system robustness and reliability.

The methodology developed in this dissertation is illustrated through problems related to spacecraft design and analysis, such as the conceptual upper-stage design of a two-stage-to-orbit vehicle, and design and analysis of a fire detection satellite.

To my parents, Dil Roushan Ara and Late AKM Asaduzzaman

ACKNOWLEDGEMENTS

First, I would like to express my sincere respect and gratitude to my thesis advisors, Dr. Sankaran Mahadevan and Dr. Mark McDonald, for their thoughtful suggestions, constant guidance and encouragement throughout the progress of this research. Dr. McDonald spent a lot of his time during my early years to help me stay focused on my research. Dr. Mahadevan was always available for discussion. His critical comments and useful advice helped a lot to guide this research in the right direction. I would also like to record my profound thanks and gratitude to the rest of my dissertation committee, Dr. Mark Ellingham and Dr. Prodyot K. Basu, for their insightful comments and advice to enrich this research.

The research reported in this dissertation was supported by funds from NASA Langley Research Center (LaRC), Hampton, VA. I would like to thank the researchers at NASA Langley, especially Mr. Lawrence Green, who provided valuable information and advice. I would like to thank Vanderbilt Civil Engineering department for supporting me as a Teaching Assistant in my first year of study. I would also like to thank the Integrative Graduate Education, Research, and Training (IGERT) Program in Reliability and Risk Engineering and Management at Vanderbilt University for providing a forum of researchers that helped me a lot in communicating and sharing ideas.

I would also like to thank the members of our research group for their support, especially Shankar Sankararaman, Venkata Sura, and Dr. Sirisha Rangavajhala. I am grateful to all of my friends for their cooperation and encouragement during this study. Finally, I would like to express my deepest appreciation to my mother for continuous support and love. I am also thankful to God for giving me the opportunity to extend my education. I would also like to thank all the people who are not listed here but have contributed in any way, being by my side or in any other sense.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER VIII

CHAPTER IX

CHAPTER X

CHAPTER XI

LIST OF FIGURES

CHAPTER V

CHAPTER VI

CHAPTER VII

CHAPTER VIII

CHAPTER IX

CHAPTER X

CHAPTER I

INTRODUCTION

The role of systematic uncertainty quantification is increasingly being recognized in assessing the performance, safety, and reliability of complex physical systems, often in the absence of an adequate amount of experimental data for many applications. Further, simulation of a complex physical system often involves multiple levels of modeling, such as material to component to subsystem to system. Interacting models and simulation codes from multiple disciplines (multiple physics) may be required at some of the levels. As the models are integrated across multiple disciplines and levels, the complexity and sophistication of the models increase, and assessing the predictive capability of the overall system model becomes a difficult challenge. The variability in the input parameters is propagated through the simulation codes, between individual disciplines, and from one level to next level. Various modeling errors, assumptions and approximations, measurement errors, and sparse and imprecise information, further compound the uncertainty in the predictive capability of the system model. An efficient methodology that accounts for all sources of uncertainty in multidisciplinary systems awaits development. Therefore, the overall research objective of this dissertation is to pursue computational methods to quantify, propagate and manage the uncertainty in multi-disciplinary system analysis models. In order to assess the uncertainty in estimates of system performance and system-level figures of merit, it is important that various types of known uncertainties be accounted for appropriately. Uncertainty in engineering analysis and design arises from several different sources (Oberkampf et al, 1999) and must be propagated through the system model. Some of the "known" sources are:

(1) Physical uncertainty or inherent variability: The demands on an engineering system as well as its properties always have some variability associated with them, due to environmental factors and variations in operating conditions, manufacturing processes, quality control etc. Such quantities are represented in engineering analysis as random variables, with statistical parameters such as mean values, standard deviations, distribution types etc. estimated from observed data and there exist well established methods for handling such uncertainty.

(2) Epistemic Uncertainty: Epistemic uncertainty represents a lack of knowledge about the system due to limited data, measurement limitations, or simplified approximations in modeling system behavior. This type of uncertainty can be typically reduced by gathering more information. Epistemic uncertainty can be viewed in two ways. It can be defined with reference to a stochastic but poorly known quantity (Baudrit and Dubois, 2006) or with reference to a fixed (deterministic) but poorly known physical quantity (Helton et al, 2004). An example of the first type of epistemic uncertainty is an expert giving a range of values for a physical quantity (such as elastic modulus of a foam material). An example of the second type of epistemic uncertainty is a measurement of the size of a crack within a mechanical component; the crack has a fixed length, but due to measurement difficulties, only an interval might be reported. Non-probabilistic representations such as fuzzy sets, evidence theory, etc. are available for describing such uncertainties (Ferson et al, 2007; Mourelatos and Zhou, 2006; Rao and Annamdas, 2009). This dissertation focuses on the first type of epistemic uncertainty, where the probability

distributions of input random variables may need to be inferred from data such as intervals given by experts or sparse point data. In this case, the distribution parameters of the random variables are also uncertain. In this dissertation, a probabilistic framework has been developed for the representation of epistemic uncertainty arising from sparse point and interval data on random variables.

(3) Model Error: This results from approximate mathematical models of the system behavior and from numerical approximations during the computational process, resulting in two types of error in general – solution approximation error, and model form error. For new and complex engineering systems, this type of uncertainty is not quantifiable a priori. This dissertation, however, does not include model errors in the uncertainty analysis.

The uncertainty described by sparse point or interval data (either regarding the variables or their distribution parameters) should be represented in a manner that facilitates ease of use in efficient algorithms for reliability analysis or design optimization. In this dissertation, this uncertainty has been represented through a flexible family of probability distributions. Such conversion of epistemic uncertainty to a probabilistic format enables the use of computationally efficient methods for probabilistic uncertainty propagation.

After finding an appropriate representation strategy for aleatory and epistemic uncertainty in the system input, it must be propagated through the system model if a statement about the uncertainty in model output is to be made. Many probabilistic uncertainty propagation methods have been developed for single discipline problems involving expensive computational codes in order to propagate physical variability in the input, typically expressed through random variables and random processes and/or fields. Most of these techniques have only been studied with respect to physical variability represented by probability distributions, but are not able to include both aleatory and epistemic uncertainty.

Uncertainty analysis often assumes independence among input random variables. However, intervariable dependencies or statistical correlations might have significant impact on the results of uncertainty analysis. Multivariate input modeling methods have been developed for some known marginal distributions (Der Kiureghian and Liu, 1986; Liu and Der Kiureghian, 1986; Minhajuddin et al; 2004 and Haas, 1999). However, in practice, it is likely that the marginal distribution types for the input variables are not known or cannot be specified accurately due to the presence of limited or interval data and in this case, the correlation itself is uncertain. Moreover, correlations among the distribution parameters have also significant impact on uncertainty analysis. Little to no work exists in the literature that considers uncertainty in correlation coefficients and correlations among distribution parameters in the presence of sparse point data or interval data. Again, for interval data, the correlations among the input variables themselves are unknown and computationally efficient methods are needed for the propagation of both aleatory and statistical uncertainty that account for correlations among random variables for which the information is only available in the form of intervals.

There has been an increased emphasis focused on accounting for uncertainty in design inputs used for design optimization. In deterministic design optimization, it is generally assumed that all the design variables and model inputs are precisely known; the influence of data or distribution parameter uncertainty on the optimality and feasibility of the models is not explicitly considered. However, real-life engineering problems are not deterministic and this deterministic assumption about inputs may lead to infeasibility or poor performance (Sim, 2006). In recent years, several methods have been developed for design under uncertainty. Reliability-based design (Chiralaksanakul and Mahadevan, 2005) and robust design (Du and Chen, 2000, Huang and Du, 2007) are two major developments among these. While reliability-based design aims to maintain design feasibility at desired reliability levels, robust design optimization attempts to minimize variability in the system performance due to variations in the inputs (Lee et al, 2008). In recent years, several methods have also been proposed to integrate these two paradigms of design under uncertainty (Lee et al, 2008, Du et al, 2004). All these methods developed so far work under aleatory uncertainty (i.e., precise probabilistic information). However, such precise knowledge about probability distribution is rarely available in practice.

In recent years, multidisciplinary reliability analysis and design optimization under uncertainty have received increased attention in order to account for uncertainties in the system and design variables. Several solution techniques are reported in the literature for multidisciplinary design optimization (MDO) under uncertainty (e.g., Du and Chen, 2002; Du and Chen, 2005; Mahadevan and Smith, 2006; Chiralaksanakul and Mahadevan, 2007; Du et al, 2008). These studies have dealt with aleatory uncertainty only. However, in practice, sufficient data are not available to construct the probability distributions of some of the input variables. Sometimes the only information available for an input variable is given by one or more intervals. If the system design can accommodate both aleatory and epistemic uncertainty, the resulting systems will be safer and more robust. Therefore, it is necessary to develop algorithms for multidisciplinary design optimization that deal with both aleatory and data uncertainty. Computational methods for multidisciplinary analysis and design under both aleatory and epistemic uncertainty are in their infancy. A few methods exist for multidisciplinary design optimization under both aleatory and epistemic uncertainty. Many of these methods use non-probabilistic methods to handle epistemic uncertainty and are computationally expensive. This dissertation research advances the state of the art in multidisciplinary system design under uncertainty.

Objectives

The overall goal of this dissertation is to develop and demonstrate effective methodologies for quantifying, propagating, and designing for uncertainty in multidisciplinary systems. Both probabilistic and non-probabilistic formats of uncertainty data have been included and integrated. In developing the methodology, this dissertation research addresses fundamental questions focused on the following five research objectives:

1. Input uncertainty representation
2. Uncertainty propagation
3. Design Optimization under uncertainty
4. Multidisciplinary uncertainty propagation analysis
5. Multidisciplinary design optimization under uncertainty

These five objectives and solution approaches are discussed below, along with the organization of the dissertation.

Objective 1: Input uncertainty representation

Input uncertainty representation is the first step for reliability analysis and probabilistic design optimization for any system. A mathematical model of the physical system must account for uncertainty. This dissertation represents uncertain quantities as random variables, described through probability distributions. However, sometimes the data on the random variable is sparse, imprecise, or incomplete and this results in uncertainty about the distribution type and distribution parameters. Again, intervariable dependencies or statistical correlations might have a significant impact on the results of uncertainty analysis. In practice, it is likely that the marginal distribution types for the input variables are not known or cannot be specified accurately due to the presence of limited or interval data which results in uncertainty in correlations among model inputs as well as their distribution parameters. This objective focuses on the following questions: (1) How can uncertainty in model inputs be quantified? (2) How can uncertainty in distribution type be addressed? (3) How can uncertainty in distribution parameters be quantified? (4) How can an efficient multivariate input modeling technique be developed in the presence of sparse and imprecise probabilistic information? (5) How can uncertainty in correlations among model inputs, and among distribution parameters of model inputs be addressed?

In order to address this uncertainty in distribution type, this dissertation proposes the use of a flexible family of distributions. Next the uncertainty in the distribution parameters themselves is considered, and the use of computational resampling methods to determine Johnson distributions for the distribution parameters is proposed. A methodology is proposed to convert uncertainty arising from interval data to a probabilistic format. This dissertation also proposes a methodology for multivariate input modeling of random variables by using a four parameter flexible Johnson family of distributions for the marginals that also accounts for data uncertainty. This multivariate input model is particularly suitable for uncertainty quantification problems that contain both aleatory and data uncertainty. In this dissertation, a computational framework is developed to consider correlations among basic random variables as well as among their distribution parameters. Chapters III and IV of this dissertation address questions 1 to 4 and Chapter VI addresses questions 4 and 5 in detail.

Objective 2: Uncertainty propagation

Once the uncertainty in model inputs, their distributions, and correlations among model inputs is quantified, it must be propagated through the system model if a statement about the uncertainty in model output is to be made. This objective focuses on the following questions: (1) How can a computationally efficient method be developed for the propagation of uncertainty through system models? (2) How can uncertain correlations among model inputs and their distribution parameters be included in uncertainty analysis?

An optimization-based approach is proposed for computing the bounds on the reliability of a design that allows for the decoupling of epistemic and aleatory uncertainty analysis, enabling computationally affordable approaches to reliability analysis under aleatory and epistemic uncertainty arising from sparse point data. This dissertation develops and illustrates a probabilistic approach for propagation in system analysis, when the information on the uncertain input variables and/or their distribution parameters may be available as either probability distributions or simply intervals (single or multiple). A methodology for propagating both aleatory and data uncertainty arising from sparse point data through computational models of system response that assigns probability distributions to the distribution parameters and quantifies the uncertainty in correlation coefficients by use of computational resampling methods is also proposed. For interval data, the correlations among the input variables are unknown. This dissertation formulates the optimization problems of deriving bounds on the cumulative probability distribution of system response, using correlations among the input variables that are described by interval data. Chapters III and V of this dissertation address question 1 and Chapter VI addresses question 2 in detail.

Objective 3: Design Optimization under uncertainty

Now that the uncertainty in the input is quantified, and an uncertainty propagation method to quantify the uncertainty in the output is developed, the next step is to develop formulations and algorithms for design optimization under data uncertainty, both from the perspective of system robustness so that the resulting solutions are least sensitive to variations in the model inputs and from the perspective of satisfying the system reliability. This objective addresses the following questions: (1) How can a methodology be developed for design optimization that can handle data uncertainty (i.e., imprecise probabilistic information)? (2) How can the proposed methodology improve the robustness and reliability of the design?

This dissertation proposes formulations and algorithms for design optimization under both aleatory (i.e., natural or physical variability) and epistemic uncertainty (i.e., imprecise probabilistic information), from the perspective of system robustness and reliability. An approach is proposed in this dissertation to decouple the robustness-based and reliability-based design from the analysis of non-design epistemic variables to achieve computational efficiency. Chapters VII and VIII of this dissertation address questions 1 and 2 in detail.

Objective 4: Multidisciplinary uncertainty propagation analysis

Multidisciplinary system analysis, even deterministic, is computationally expensive. Uncertainty analysis multiplies the computational effort even further. Inclusion of data uncertainty within the analysis again multiplies the computational effort. This objective focuses on the following fundamental questions: (1) How can the uncertainty quantification methods developed in Objectives 1-2 be extended for multidisciplinary systems? (2) How can an efficient method for uncertainty quantification be developed for a multidisciplinary system that includes imprecise probabilistic information and remains computationally tractable?

This dissertation develops an efficient probabilistic approach for uncertainty propagation in multidisciplinary system analysis, when the information on the uncertain input variables may be available as either sparse point data or as intervals (single or multiple). A decoupled approach is used in this dissertation to un-nest the multidisciplinary system analysis from the probabilistic analysis to achieve computational efficiency. This approach uses deterministic optimization to first quantify the uncertainty in the coupling variables, without any coupled system level analysis. Once the uncertainty in the coupling variables is quantified, the system level uncertainty propagation analysis is similar to single discipline problems. The proposed methods are equally applicable to both sampling and analytical approximation-based reliability analysis methods. Chapter IX of this dissertation addresses questions 1 and 2 in detail.

Objective 5: Multidisciplinary design optimization under uncertainty

Multidisciplinary design optimization under aleatory uncertainty itself is a challenging problem. Inclusion of epistemic uncertainty makes this problem more difficult. This objective focuses on the following fundamental questions: (1) How can the design optimization methods developed in Objective 3 be extended for multidisciplinary systems? (2) How can an efficient method be developed for multidisciplinary system design that includes sparse and imprecise probabilistic information and remains computationally tractable?

This dissertation proposes formulations and algorithms for design optimization of multidisciplinary systems under both aleatory uncertainty (i.e., natural or physical variability) and epistemic uncertainty (i.e., imprecise probabilistic information), from the perspective of system robustness and reliability. A single loop approach is used for the robustness-based design optimization, which does not require any explicit coupled multidisciplinary uncertainty propagation analysis. Thus the computational complexity and cost involved in estimating the mean and variation of the performance function is greatly reduced. The proposed methodology for reliability-based design of multidisciplinary systems also does not require any coupled system level analysis. An approach is proposed in this dissertation to decouple the robustness-based and reliability-based design from the analysis of non-design epistemic variables to achieve further computational efficiency. Chapters X and XI address questions 1 and 2 in detail.

The uncertainty quantification and design optimization methodologies developed in this dissertation are illustrated through problems related to spacecraft design and analysis, such as the conceptual level upper stage design of a two-stage-to-orbit (TSTO) vehicle and a simplified three-disciplinary version of a fire satellite (FireSat).

Two-stage-to-orbit (TSTO) Vehicle

The two-stage-to-orbit (TSTO) vehicle involves a multidisciplinary system analysis consisting of geometric modeling, aerodynamics, aerothermodynamics, engine performance analysis, trajectory analysis, mass property analysis and cost modeling (Stevenson et al, 2002). The Two-Stage-To-Orbit (TSTO) is a Highly Reliable Reusable Launch Systems (HRRLS) concept vehicle, as shown in Figure 1. This concept vehicle is used in the NASA Aeronautics Research Mission Directorate (ARMD) Fundamental Aeronautics Program (Hypersonics Project). The first (launch) stage (shown in blue in the figure), employs a turbine-based, combined cycle propulsion system. The second (upper) stage is (shown in red in the figure) employs a rocket powered propulsion system.

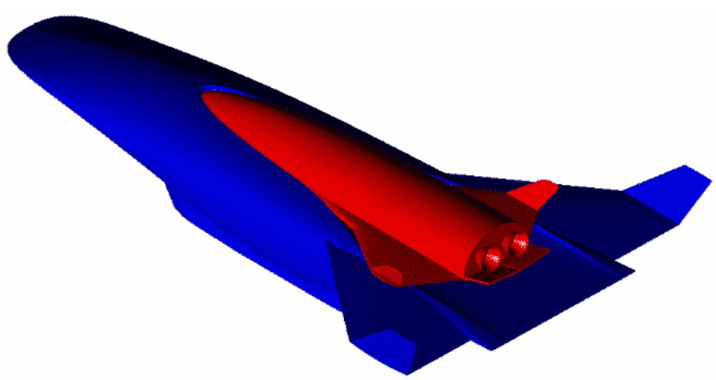


Figure 1: The Two-Stage-To-Orbit (TSTO) Highly Reliable Reusable Launch Systems (HRRLS) concept vehicle.

In this dissertation, a simplified version of the upper stage design process of a TSTO vehicle is used to illustrate the proposed methods. High fidelity codes of individual disciplinary analysis are replaced by inexpensive surrogate models. Figure 2 illustrates the analysis process of a TSTO vehicle.

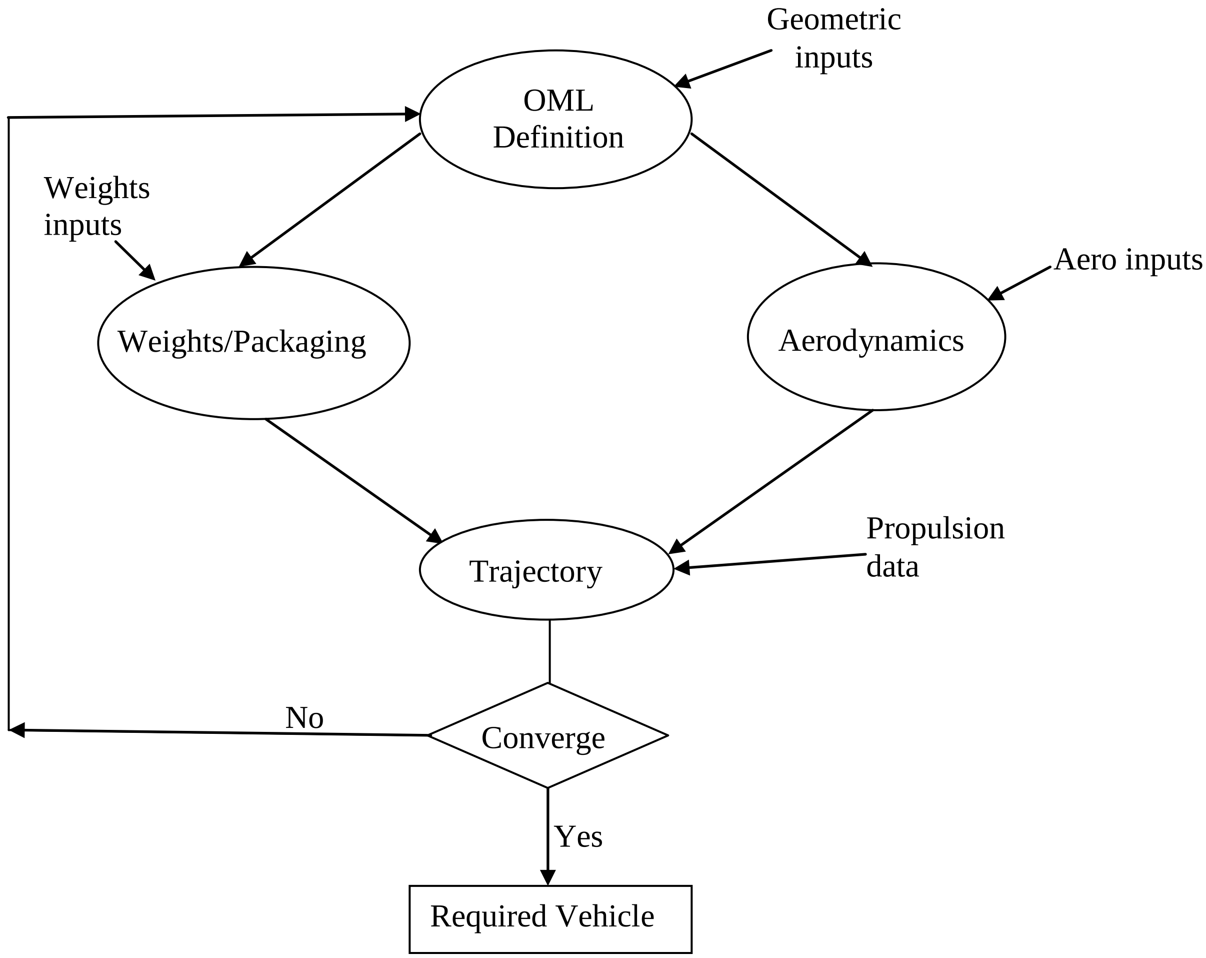


Figure 2: TSTO vehicle concept

Fire Satellite (FireSat) Design and Analysis

This problem has been originally described in Wertz and Larson (1999). This is a hypothetical but realistic spacecraft consisting of a large number of subsystems with both feedback and feed-forward couplings. The primary objective of the fire satellite (FireSat) is to detect, identify, and monitor forest fires in near real time. This satellite is intended to carry a large and accurate optical sensor of length 3.2 m and weight 720 kg, and has an angular resolution of 8.8e-7 radians. In this dissertation, a simplified subset of FireSat subsystems consisting of *i)* Orbit Analysis, *ii)* Attitude Control and *iii)* Power, based on Ferson et al (2009) has been used . This three-discipline problem is sketched in Figure 3.

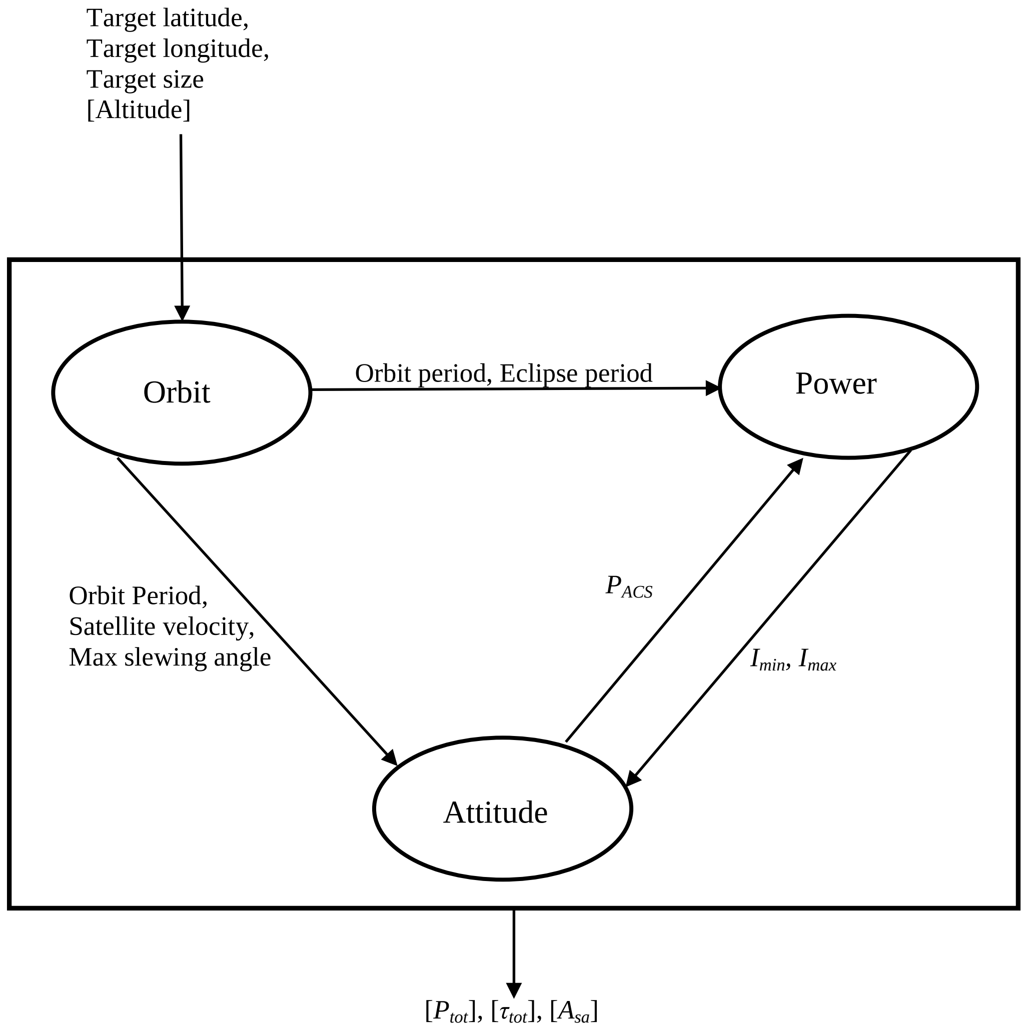


Figure 3: FireSat: A three-subsystem representation

The following chapter discusses the existing methods for handling the above objectives and explains the scope of the proposed methods.

CHAPTER II

LITERATURE REVIEW

1. Introduction

As discussed in Parry (1996), there are three elements in a model-based uncertainty analysis: *i)* characterizing uncertainty in individual elements of the model, i.e., representing input uncertainty regarding individual elements of the model, *ii)* propagating the uncertainty thus represented through a model of system response to obtain a representation of the output uncertainty and *iii)* communicating the results thus obtained to the decision makers and other stakeholders. Therefore, it is important that the different types of uncertainty in the system and the model be represented in a way that it can be efficiently used in further analysis i.e., in algorithms for reliability analysis and design optimization and the results can be easily communicated to the stakeholders. It is now well recognized that both aleatory and epistemic uncertainty must be represented in an appropriate manner so that it can be used in any decision support analysis (Helton and Burmaster, 1996; Parry, 1996; Pate´-Cornell, 1996).

Many important physics-based engineering analyses require the use of computationally expensive codes, and often involve uncertain inputs that are described using various probabilistic and non-probabilistic methods. Many probabilistic uncertainty propagation methods have been developed for many single discipline problems involving computationally expensive simulation models in order to propagate physical variability in the input, expressed through random variables and random processes and/or fields. In recent years, several methods have been developed for design under uncertainty for both single and multidisciplinary problems.

All these methods developed so far work under precise probabilistic information on the random variables. This dissertation specifically focuses on epistemic uncertainty arising from imprecise probabilistic information (especially *sparse point data* and *interval data*) on the random variables. In particular, this dissertation develops methods for different uncertainty management tools, namely uncertainty quantification and design optimization for both single and multidisciplinary systems under both aleatory and epistemic uncertainty arising from sparse point and interval data.

The following sections present a review of existing methods in the literature for uncertainty quantification, propagation, and design optimization under uncertainty. This review is followed by an outline of the methods proposed in the subsequent chapters to address some of the unfulfilled research needs in the current literature, especially with respect to imprecise probabilistic information.

2. Uncertainty quantification with sparse point data

One approach for uncertainty representation under data uncertainty is evidence theory (Shafer, 1976). Evidence theory has been used with interval data for reliability-based design optimization (Mourelatos and Zhou, 2006) and multidisciplinary systems design (Agarwal et al, 2004), where a belief measure is used to formulate the non-deterministic design constraints. Other approaches for epistemic uncertainty quantification based on evidence theory include Guo and Du (2007) and Guo and Du (2009). However, evidence theory requires basic probability assignments (bpa) and it is not clear how to construct bpa from sparse data.

In some cases, random variables with sparse data can be modeled using convex models of uncertainty (Ben-Haim and Elishakoff, 1990). Examples of convex models include intervals, ellipses or any convex sets. Convex models usually require less detailed information to characterize uncertainties than probabilistic models. They require a worst-case analysis in design applications that can be formulated as a constrained optimization problem. When the convex models are intervals, techniques in interval analysis can be used, though they are computationally expensive. As an extension of interval analysis, some research in uncertainty representation and propagation under data uncertainty has focused on the use of possibility/fuzzy set theory (Dubois and Prade, 1988). The drawback of these approaches is that they require combinatorial interval analysis, and the computational expense increases exponentially with the number of uncertain variables and with the nonlinearity of the function. Further, the use of interval analysis methods for problems with sparse point data requires that interval information be inferred from point data, and this introduces additional uncertainty to the problem.

It is clear from the above discussion that probability theory might be easier and more intuitive in handling the information available from sparse point data. While probability theory is a widely understood and perfect description of aleatory uncertainty, knowledge of the exact probability distribution type and/or parameters for random variables is usually imperfect due to limited data. Extensive probabilistic techniques for uncertainty quantification and propagation also exist, which usually rely on existence of sufficient data. For cases where data is limited, it is impossible to define a unique probability distribution function to adequately describe the random variable. Hence, the probability distributions are imprecise. Two types of approaches are available to handle this situation: (1) Bayesian methods, and (2) P-boxes, as discussed below.

Bayesian methods (e.g. Der Kiureghian 1984, McFarland 2007) have been used to leverage expert opinion in cases where data is sparse, while including the information gained from the data. However, under sparse data, the distributions selected are sensitive to the choice of prior distributions. Alternately, other studies within the context of imprecise probability theory have focused on representing uncertainty in the probability distribution by using a probability box, or p-box (e.g., Ferson et al, 2007), which is the collection of all possible empirical distributions for the random variable. Other research has focused on developing bounds, e.g., on CDFs. Halperin (1986) extensively developed the idea of interval bounds on CDFs as well as methods for propagation of these probability intervals through simple expressions. Hyman (1982) developed similar ideas for probabilistic arithmetic expressions in the density domain. Williamson and Downs (1990) described algorithms to compute arithmetic operations (addition, subtraction, multiplication and division) on pairs of p-boxes. These operations generalize the notion of convolution between probability distributions (Berleant, 1993; 1996; Berleant and Goodman-Strauss, 1998).

Several of these current methods of uncertainty propagation under data and distribution uncertainty can be computationally expensive. One reason is that for every combination of distribution parameters, the probabilistic analysis for aleatory variables has to be repeated, which results in a computationally expensive *nested* analysis. Various approaches can be used to reduce the computational expense of the nested approach; for example, Monte Carlo methods leveraging importance sampling, the first-order reliability method (FORM), and second-order reliability method (SORM) (Haldar and Mahadevan,

2000) can be used. The system analysis may also be replaced with an inexpensive surrogate model (e.g., polynomial chaos (Ghanem and Spanos 1991; Cheng and Sandu, 2009) or Gaussian process model (Bichon et al, 2008)) to achieve computational efficiency. While these propagation methods are useful for problems dealing with uncertainties having probabilistic representation arising primarily from inherent variability in physical parameters, decoupled methods to efficiently represent and propagate aleatory and data uncertainty (or a mixture of aleatory and epistemic uncertainty) are yet to be developed.

Therefore, Chapter III of this dissertation develops and illustrates an approach for the propagation of both aleatory and epistemic uncertainty in such a way that the epistemic and aleatory uncertainty analyses are not nested, thus enabling computationally efficient calculation of bounds on reliability estimates under epistemic and aleatory uncertainty.

3. Uncertainty quantification with interval data

3.1 Sources of Interval data

Interval data are encountered frequently in practical engineering problems. Several such situations where interval data arise are discussed in (Du et al 2005; Ferson et al 2007), for example: (a) physical limits and theoretical constraints may be the only sources of information, which can only circumscribe possible ranges of quantities resulting in interval data. (b) Interval data arises when the only information available for a variable is a collection of expert opinions, which specify a range of possible values for a variable. (c) Reporting data with plus-or-minus uncertainties associated with the calibration of measuring devices also leads to interval data. (d) Some tests in chemical or purity quantification can only state that an observation is below a certain detection limit, resulting in an interval observation for the amount of impurity between zero and the threshold. (e) Intervals may arise in the detection of a fault when observations are spaced temporally; as the fault occurs between two consecutive observations, the time of failure is given by a window of time. Interval data requires careful treatment, especially if the width of the interval cannot be ignored when compared to the magnitude of the variable.

Two types of interval data are considered in this dissertation, based on computational methods: single or multiple intervals. When compared to single interval cases, multiple intervals require consideration of two additional issues: (1) From the context of computational expense, estimating statistics from multiple intervals can be more challenging, (2) From the context of aggregation of information represented in the multiple intervals, there may be no basis to believe that the "true" value of the variable lies at any particular location of any intervals, such as endpoints or midpoints of the intervals. Although not necessarily true, a common assumption in the literature is that all the intervals are equally likely to enclose the "true" value of the variable, i.e., all intervals have an equal weight (Ferson et al 2007).

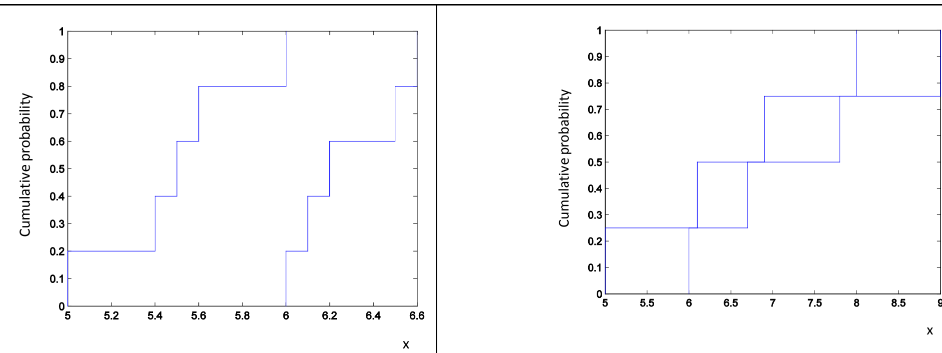
When data is available in multiple intervals (e.g., given by multiple experts), the information contained in one interval could contradict that in the other interval(s), or could be contained by other interval(s). In this context, intervals can be broadly categorized as *non-overlapping* and *overlapping* intervals.

3.2 What does an interval represent?

In order to propagate uncertainty through models of system response, it is necessary to first have a valid representation of the input uncertainty that can lead to meaningful quantification of the uncertainty in the system response. In this context, there are two broadly categorized interpretations of what interval data represents in the literature.

The first is the so-called equi-probability model, which corresponds to the Laplacian principle of indifference (Howson and Urbach, 1993) and considers each interval as a uniform distribution (Bertrand and Groupil, 2000). Each possible value in every interval is assumed equally likely, resulting in a single probability mass and/or density for each possible realization of a random variable. We note that there might not be a justification to assume uniform distribution or any other distribution within a particular interval, which can be viewed as a limitation of the equi-probability model. Also, the equi-probability model results in a precise probabilistic representation of interval data, thereby failing to capture the inherent imprecision in the data.

The second popular interpretation of interval data, which is adopted in this dissertation, is that it represents *incertitude* in the data (Ferson et al 2007). As a result, the possible values for quantities of interest such as probability of an event will in general be an *interval*, unlike a single value for point data. Unlike the equi-probability model that results in a single probabilistic representation of the interval, the notion of incertitude leads to a collection of distribution functions that could arise from different possible combinations of values from within the intervals.



Overlapping data with 5 intervals [5, 6; 5.5, 6.1; 6, 6.5; 5.4, 6.2; 5.6, 6.6]	Non-overlapping data with 4 intervals [5, 6; 6.1, 6.7; 6.9, 7.8; 8, 9]

Figure 1: Examples of an empirical p-box for multiple intervals

The set of all possible probability distributions of a particular distribution type (e.g., empirical, normal) for a variable described by interval data is known as a probability box, or a p-box for short (Williamson and Downs 1990). To illustrate, we explain the notion of an empirical p-box that exists in the literature (Ferson et al 2007), which is the collection of all possible empirical distributions for the given set of intervals. An empirical p-box summarizes the interval data set graphically. It is constructed as an increasing step function with a constant vertical step height of 1/N, where N is the number of intervals. The construction of the empirical p-box requires sorting the lower and upper bounds for the set of intervals, followed by plotting the empirical cumulative distribution function (CDF) of each of the sorted bounds as shown in Figure 1 for overlapping interval data with five intervals [5, 6; 5.5, 6.1; 6, 6.5; 5.4, 6.2; 5.6, 6.6] and non-overlapping interval data with four intervals [5, 6; 6.1, 6.7; 6.9, 7.8; 8, 9]. The step height at each data point for the empirical CDF in Figure 1 is equal, which reflects the assumption that the intervals are all equally weighed. Note that the p-box and the

Dempster-Shafer structure (discussed in the following subsection) are equivalent and each representation can be converted to the other (Regan et al, 2004).

Several other approaches such as evidence theory and fuzzy logic are also used within the interpretation of incertitude. A brief discussion of the various techniques to represent interval data within the scope of incertitude is presented next.

3.3 Existing Methods for Treatment of Interval Uncertainty

The Sandia epistemic uncertainty project (Oberkampf et al 2004) conducted a workshop that invited various views on quantification and propagation of epistemic data uncertainty (includes interval data), which are summarized in (Ferson et al 2004). Many uncertainty theories for representation and propagation of interval uncertainty have been discussed at the workshop, which include Dempster–Shafer structures (Helton et al, 2004; Klir, 2004), probability distributions (Helton et al, 2004), possibility distributions (Helton et al, 2004), random intervals (Fetz and Oberguggenberger, 2004), sets of probability measures (Fetz and Oberguggenberger, 2004), fuzzy sets (Fetz and Oberguggenberger, 2004), random sets (Berleant and Zhang, 2004; Hall and Lawry, 2004), imprecise coherent probabilities (Kozine and Utkin, 2004), coherent lower previsions (De Cooman and Troffaes, 2004), p-boxes (Ferson and Hajagos, 2004), families of polynomial chaos expansions (Red- Horse and Benjamin, 2004), info-gap models (Ben-Haim, 2004), etc. A brief discussion of some of the popular uncertainty theories discussed in the above workshop, and interval data in general, follows.

In addition to the p-box representation discussed previously, other research within the realm of probability theory for interval data has focused on developing bounds, e.g., on CDFs. Hailperin (1986) extensively developed the idea of interval bounds on CDFs as well as methods for propagation of these probability intervals through simple expressions. Hyman (1982) developed similar ideas for probabilistic arithmetic expressions in the density domain. Williamson and Downs (1990) described algorithms to compute arithmetic operations (addition, subtraction, multiplication and division) on pairs of p-boxes. These operations generalize the notion of convolution between probability distributions (Berleant 1993; 1996; Berleant and Goodman-Strauss, 1998). Additional results involving bounds on CDFs are available in Helton et al (2004) and Helton et al (2008). Epistemic uncertainty has also been expressed using subjective probability (e.g., (Apeland et al, 2002; Hofer et al, 2002)). On the other hand, some researchers believe that a probabilistic representation is not appropriate for interval data because information may be added to the problem (Du et al, 2005; Agarwal et al, 2004).

A commonly used approach for representation of interval data is Dempster-Shafer evidence theory (Shafer, 1976). In the context of evidence theory, there exist many rules to aggregate different sources of information. Among different rules of combination, the Dempster's rule is one of the most popular, however, this approach may not be suitable particularly for cases where there is inconsistency in the available evidence (Oberkampf et al, 2001; Agarwal et al, 2004), e.g., in the case of non-overlapping intervals. In such cases, a mixture or averaging rule may be appropriate (Oberkampf et al, 2001). Evidence theory has been applied to quantify epistemic uncertainty in the presence of interval data for multidisciplinary systems design (Agarwal et al, 2004), where a belief measure is used to formulate the non-deterministic design constraints. Others have developed approaches for epistemic uncertainty quantification based on evidence theory, including Guo and Du (2007) and Guo and Du (2009).

In some cases, uncertain events form patterns that can be modeled using convex models of uncertainty (Ben-Haim and Elishakoff, 1990). Examples of convex models include intervals, ellipses or any convex sets. Convex models usually require less detailed information to characterize uncertainties than probabilistic models. They require a worst-case analysis in design applications which can be formulated as a constrained optimization problem. When the convex models are intervals, techniques in interval analysis can be used, though they are computationally expensive.

Some research has focused on the use of possibility/fuzzy set theory for interval data. The possibility distribution (membership function) of a function of an interval variable with a given possibility distribution can be found using Zadeh's Extension Principle (Dubois and Prade, 1988). The drawback of this approach is that it requires combinatorial interval analysis, and the computational expense increases exponentially with the number of uncertain variables and with the nonlinearity of the function. Within the realm of fuzzy representation, Rao and Annamdas (2009) present the idea of weighted fuzzy theory for intervals, where fuzzy set based representations of interval variables from evidences of different credibilities are combined to estimate the system margin of failure.

The aggregation of multiple sources of information as seen with multiple interval data is an important issue in characterizing input uncertainty. There is now an extensive list of literature that discuss different aggregation methods, which include stochastic mixture modeling (Ferson and Hajagos, 2004; Helton et al, 2004), Dempster's rule (Agarwal et al, 2004; Rutherford, 2004), a posteriori mixture (Red- Horse and Benjamin, 2004), natural extension of pointwise maximum (De Cooman and Troffaes, 2004), etc.

However, the aggregation method used in uncertainty representation must be consistent with the nature of the uncertainty as well as the specific uncertainty theory used (Helton et al, 2004).

Helton et al (2004) discussed and illustrated the use of different uncertainty theories, namely, probability theory, evidence theory, possibility theory, and interval analysis for the representation and propagation of epistemic uncertainty. This paper used a sampling-based approach with each of the uncertainty theories. For probability theory, they defined the probability spaces by assuming uniform distributions over the sets of the possible values of the input variables. Multiple sources of information are aggregated by simply averaging the distributions for the number of sources assigning equal weight to each source. Baudrit and Dubois (2006) proposed a methodology to represent imprecise probabilistic information described by intervals using different uncertainty approaches, such as probability theory, possibility theory and belief functions, etc.

Within the context of uncertainty propagation with interval variables, there exists literature that considers both interval and aleatory uncertainties. Approaches such as evidence theory or possibility theory are commonly used to represent interval variables, while probabilistic representation is typically used to represent aleatory uncertainties. The propagation of an evidence theory representation of uncertainty through a model of system response is computationally more expensive than that of probability theory (Helton et al, 2007). Helton et al (2008) discussed the efficiency of different alternatives for the representation and propagation of epistemic uncertainty and argued that propagation of epistemic uncertainty using evidence theory and possibility theory required more computational effort than that of probability theory. In uncertainty propagation analysis, for every combination of interval values, the probabilistic analysis for aleatory variables is repeated, which results in a computationally expensive *nested* analysis. Some research in the literature focuses on managing this computational expense (Penmetsa and Grandhi, 2002, Rao and Cao, 2002). Representation and propagation of interval uncertainty has been studied from the context of structural problems (Langley, 2000) and multidisciplinary problems (Du and Chen, 2000). Besides their computational complexity, another disadvantage of using non-probabilistic methods is that the end users of the uncertainty analysis are little aware of these methods and therefore, it may involve huge educational effort to make them familiar with these non-traditional uncertainty analysis methods (Helton et al, 2008).

As discussed above, there are various approaches for treating interval data, each with their own advantages and limitations. One of the drawbacks of the current approaches is the need for nested analysis in the presence of interval variables. To alleviate this issue, Chapter IV develops a probabilistic representation for interval data using a collection of flexible probability distributions.

If non-probabilistic methods are to be used for epistemic uncertainty propagation, new efficient approaches have to be developed. However, if the uncertainty described by intervals can be represented through probability distributions, the computational expense of interval analysis is avoidable as it allows for treatment of aleatory and epistemic uncertainty together without nesting, and already well established probabilistic methods of uncertainty propagation, for example, Monte Carlo methods (Robert and Cesalla, 2004) and optimization-based methods such as first-order reliability method (FORM), second-order reliability method (SORM) etc. (Haldar and Mahadevan, 2000) can be used.

The system may also be replaced with an inexpensive surrogate model (e.g., polynomial chaos (Ghanem and Spanos 1991; Cheng and Sandu, 2009) or Gaussian process model (Bichon et al, 2008)) to achieve computational efficiency. While these uncertainty propagation methods are useful for problems dealing with uncertainties having probabilistic representation arising primarily from inherent variability in physical parameters, methods to efficiently represent and propagate epistemic uncertainty (or a mixture of aleatory and epistemic uncertainty) are yet to be developed.

It should be noted that some researchers argue that a probabilistic representation is not appropriate for epistemic uncertainty because information may be added to the problem (Du et al, 2005; Agarwal et al, 2004). This may be true when a single fixed probability distribution is assumed for the epistemic variable. In this dissertation, we alleviate this concern by using a flexible family of Johnson distributions. The use of a family of distributions for the underlying basic random variable avoids the problem of incorrect classification of the distribution type and thus minimizes the risk of adding information to the problem. However, we observe that some non-probabilistic methods may also add subjective information to the problem. For example, when evidence theory is used for the representation of interval uncertainty, the use of a combination rule adds an assumption about combining evidence; different combination rules exist in the literature (Agarwal et al, 2004). The commonly used Dempster's rule also requires some consistency or agreement among the intervals (Oberkampf et al, 2001; Agarwal et al, 2004). The evidence theory also requires that an interval for a random variable be associated with the basic probability assignments (BPA) associated with the intervals.

However, in practice, such consistency among the different intervals may not be possible and any assumption about the BPA can add information to the problem.

Chapter V of this dissertation develops a new probabilistic approach for the propagation of both probabilistic and interval variables.

4. Uncertainty quantification considering correlations

As mentioned in Chapter I, uncertainty analysis studies often assume independence among input random variables for the sake of convenience and due to lack of multivariate data. However, intervariable dependencies or statistical correlations might have significant impact on the results of uncertainty analysis. Uncertainty analyses with correlated variables require the joint PDF of input variables. However, it is almost impossible to obtain the joint PDF of the input variables, as it requires joint multivariate observations. Therefore, uncertainty analyses tend to use only information on marginal distributions and covariances. Correlation information can be used to transform the correlated variables to an uncorrelated reduced normal space in the case of analytical reliability methods (e.g., FORM) or to simulate correlated random variables for use in Monte Carlo simulation.

There exist various methods to transform correlated variables to uncorrelated standard normal space and to simulate correlated random variables, e.g., Rosenblatt transformation (Rosenblatt, 1952), Nataf transformation (Nataf, 1962), Power and Modulus transformations (Box and Cox, 1964; John and Draper, 1980), etc. The Rosenblatt transformation is quite accurate, but it requires closed form conditional distributions which are almost impossible to obtain in practice. The Nataf transformation requires only information on marginal distributions and the correlation matrix (Rebba, 2005). Methods of generating correlated variables or transforming correlated variables to uncorrelated standard normal space have been discussed and illustrated in many studies (Der Kiureghian and Liu 1986; Liu and Der Kiureghian 1986 ; Haas 1999 and Minhajuddin et al 2004) for known marginals such as normal, lognormal, shifted exponential, shifted Rayleigh, Gamma, beta, etc. Der Kiureghian and Liu (1986) presented semi-empirical formulas that relate the correlation coefficients in the reduced normal space $\rho_{0,ij}$ to the original correlation coefficients ρ_{ij} for several known two-parameter marginal distributions.

In practice, it is likely that the marginal distribution types are not known or cannot be specified accurately due to the presence of limited or interval data. For such cases, the Johnson family of distributions (Johnson, 1949a) is a convenient choice as it has the flexibility to fit data with a large range of different distribution function shapes and thus eliminates the need to forcibly assume a fixed distribution type. While there are several other viable four-parameter distributions that may also be used with this approach, such as the Pearson (Pearson, 1895), Beta (McDonald, 1984), and Lambda distributions (Ramberg and Schmeiser, 1974), the Johnson family is a convenient choice. This is because the Johnson distribution lends itself to easy transformation to a standard normal space, which then can be conveniently applied in well known reliability analysis and reliability-based design optimization methods. However, for random variables having Johnson marginal distributions, an efficient methodology to transform correlated variables to uncorrelated standard normal space or to simulate correlated variables is yet to be developed. Johnson (1949b) proposed a bivariate distribution based on the univariate Johnson distributions. This method can be extended for simulating multivariate Johnson distributions as discussed in Stanfield et al (1996). However, this multivariate Johnson distribution cannot match the sample correlation matrix of the original data set if some of the marginal distributions possess large skewness. Stanfield et al (1996) proposed an improved method to model multivariate Johnson distributions that can match the first three marginal moments and correlation structure of the data but fail to match the kurtosis of the data.

Most of the existing methods that use statistical correlation in uncertainty analysis have been developed only in the context of aleatory uncertainty in the input random variables (e.g., Noh et al, 2009). These methods consider correlations among basic random variables that are described by well known two-parameter probability distributions (e.g., normal, lognormal, exponential, Rayleigh, Gamma, etc.). Some uncertainty quantification methods exist that deal with unknown dependencies among the input variables (Berleant and Zhang, 2004). Ferson and Kreinovich (2006) described dependence among input variables described by interval data in the context of interval analysis. Recently, uncertainty quantification methods under both aleatory and data uncertainty have been developed where input uncertainty is represented by a flexible family of distributions, e.g., Johnson distributions (McDonald et al, 2009 and Zaman et al, 2009a, 2009b). These uncertainty quantification methods were developed assuming independence among the input variables. Chapter V of this dissertation develops a multivariate input model of random variables and extends these methods to include correlations.

5. Design optimization with epistemic uncertainty

In deterministic design optimization, it is generally assumed that all design variables and system variables are precisely known; the influence of natural variability and data uncertainty on the optimality and feasibility of the design is not explicitly considered. However, real-life engineering problems are not deterministic and this deterministic assumption about inputs may lead to infeasibility or poor performance (Sim, 2006). In recent years, many methods have been developed for design under uncertainty. Reliability-based design (e.g., Chiralaksanakul and Mahadevan, 2005; Ramu et al, 2006; Agarwal et al, 2007and Du and Huang, 2007) and robust design (e.g., Parkinson et al, 1993; Du and Chen, 2000; Doltsinis and Kang, 2004 and Huang and Du, 2007) are two directions pursued by these methods. While reliability-based design aims to maintain design feasibility at desired reliability levels, robust design optimization attempts to minimize variability in the system performance due to variations in the inputs (Lee et al, 2008). In recent years, several methods have also been proposed to integrate these two paradigms of design under uncertainty (e.g., Du et al, 2004, Lee et al, 2008).

Taguchi proposed robust design methods for selecting design variables in a manner that makes the product performance insensitive to variations in the manufacturing process (Taguchi, 1993). Taguchi's methods have widespread applications in engineering; however, these methods are implemented through statistical design of experiments and cannot solve problems with multiple measures of performances and design constraints (Wei et al, 2009). With the introduction of nonlinear programming to robust design, it has become possible to achieve robustness in both performance and design constraints (Du and Chen, 2000).

Although there is now an extensive volume of literature for robust optimization methods and applications, all these methods have only been studied with respect to physical or natural variability represented by probability distributions. However, uncertainty in system design also arises from other contributing factors as discussed in Chapter I. A few studies on robust design optimization are reported in the literature to deal with epistemic uncertainty arising from lack of information. Youn et al (2007) used a possibility-based method, and redefined the performance measure of robust design using the most likely values of fuzzy random variables. Dai and Mourelatos (2003) proposed two two-step methods for robust design optimization that can treat aleatory and epistemic uncertainty separately using a range method and a fuzzy sets approach.

There is now also an extensive volume of literature available for RBDO methods and applications. However, all these methods have only been studied with respect to physical or natural variability represented by probability distributions. RBDO is a challenging problem in presence of epistemic uncertainty, because the design methodology requires employing a search among the possible values of epistemic variables in order to find a conservative design. A few studies on RBDO are reported in the literature to deal with epistemic uncertainty arising from lack of information. Agarwal et al (2004) developed an evidence theory based approach to multidisciplinary RBDO using response surfaces for uncertain measures represented by the belief and plausibility functions and a sequential approximate optimization approach. However, this method cannot handle both alatory and epistemic uncertainty together. Mourelatos and Zhou (2006) developed an evidence theory based design optimization (EBDO) methodology for single discipline system that can handle both aleatory and epistemic uncertainty. Mourelatos and Zhou

(2005) proposed a possibility based design optimization (PBDO) methodology for single discipline system, which is a formulation of triple loop optimization sequence and therefore, is computationally expensive. Zhou and Mourelatos (2008) proposed double loop and sequential strategies to manage the computational expense in the PBDO methodology.

Most of the current methods of robust optimization and RBDO with epistemic uncertainty need additional non-probabilistic formulations to incorporate epistemic uncertainty into the robust optimization framework, which may be computationally expensive. Therefore, there is a need for efficient robust design optimization and RBDO methodologies that deal with both aleatory and epistemic uncertainty.

Chapter VII of this dissertation develops an efficient robust optimization methodology that includes both aleatory and epistemic uncertainty arising from both sparse point data and interval data. Chapter VIII of this dissertation develops an efficient RBDO methodology that includes both aleatory and epistemic uncertainty arising from both sparse point data and interval data.

6. Uncertainty quantification in multidisciplinary systems

Efficient uncertainty propagation methods are available to include both aleatory and epistemic uncertainty in uncertainty propagation analysis, but for single discipline problems only, for example see Zaman et al (2009b) and the references cited therein. Uncertainty propagation for multidisciplinary systems, even with aleatory uncertainty alone, is expensive as it involves coupled system analysis that is achieved through iterative executions of individual disciplinary analysis codes. Some efficient methods such as Du and Chen (2005) and Du et al (2008) are available for handling aleatory uncertainty in multidisciplinary analysis. These methods take advantage of optimization to construct analytical approximations to evaluate the system compatibility requirement. When both aleatory and epistemic uncertainty are present, propagation of uncertainty through multidisciplinary system models becomes even more difficult. This dissertation focuses on the handling of sparse point and interval data in a manner that facilitates efficient algorithms for reliability analysis or design optimization of multidisciplinary systems.

There is now an extensive volume of literature available for deterministic multidisciplinary design optimization (MDO) methods and applications (e.g., Cramer et al, 1994; Sobieszczanski-Sobieski, 1995; Sobieszczanski-Sobieski and Haftka, 1997). In recent years, multidisciplinary reliability analysis and design optimization under uncertainty have received increased attention in order to account for uncertainties in the system and design variables. Several solution techniques are reported in the literature for multidisciplinary design optimization (MDO) under uncertainty (e.g., Du and Chen, 2002; Du and Chen, 2005; Mahadevan and Smith, 2006; Chiralaksanakul and Mahadevan, 2007; Du et al, 2008). These studies have dealt with aleatory uncertainty only. However, in practice, sufficient data are not available to construct the probability distributions of some of the input variables. Sometimes the only information available for an input variable is given by one or more intervals. Therefore, it is necessary to develop algorithms for multidisciplinary reliability analysis and design optimization that deal with both physical variability and data uncertainty.

A few methods exist for MDO under both aleatory and epistemic uncertainty. Zhang and Huang (2009) proposed algorithms that considered both random and fuzzy variables. Agarwal et al (2004) proposed a methodology for uncertainty quantification using evidence theory. Li and Azarm (2008) proposed methods for interdisciplinary uncertainty propagation embedded within a multidisciplinary robust optimization framework for interval variables. Gu et al (2006) proposed an implicit uncertainty propagation method considering aleatory uncertainty in the design variables and prediction error in disciplinary simulation-based design tools. An efficient methodology for multidisciplinary uncertainty propagation with both aleatory and epistemic uncertainty that works within a probabilistic framework of uncertainty representation awaits development. This is the focus and contribution of Chapter IX of this dissertation.

The efficiency of multidisciplinary uncertainty propagation analysis depends on how the system analysis is handled. Several methods are available for system analysis within the MDO literature, namely the multidisciplinary feasibility (MDF) method, the all-at-once (AAO) method, and the individual disciplinary feasibility (IDF) method (Cramer et al, 1994). All these methods have their own advantages and limitations.

A decoupled approach for multidisciplinary reliability analysis was previously developed in Mahadevan and Smith (2006). This approach quantifies the uncertainty associated in coupling variables and therefore un-nests the system analysis from the algorithms of probabilistic analysis. However, this method has been developed for handling aleatory uncertainty only. In this disstertation, we extend this idea of a decoupled formulation and propose probabilistic methods for multidisciplinary reliability analysis under both aleatory and epistemic uncertainty.

Chapter VII of this dissertation develops a probabilistic framework for the propagation of both aleatory and epistemic uncertainty in multidisciplinary systems that can deal with both sparse point data and any type of interval data (nested, un-nested and mixed).

7. Multidisciplinary design optimization with epistemic uncertainty

MDO is the optimization of systems of coupled simulations (Cramer et al, 1994). There is now an extensive volume of literature available for MDO methods and applications (e.g., Cramer et al, 1994; Sobieszczanski-Sobieski, 1995; Sobieszczanski-Sobieski and Haftka, 1997). However, these deterministic methods can be inadequate in real-world applications since they do not explicitly take uncertainty into account. Robustness-based design optimization and RBDO account for this uncertainty in design parameters.

Robustness-based design optimization of a multidisciplinary system aims to simultaneously optimize the mean value of the objective function and minimize its variation while satisfying the system compatibility requirements of the multidisciplinary system. Although there is now an extensive volume of literature for robust optimization methods and applications, all these methods have only been studied for single discipline problems. As mentioned earlier, some of these methods can only handle aleatory uncertainty, while others can handle both aleatory and epistemic uncertainty.

Multidisciplinary robustness-based design integrates the concept of robust design with multidisciplinary design optimization (MDO). The difficulties lie in estimating the mean and variation of the performance functions considering the multidisciplinary nature of the system. The term *performance function* refers to the objective function as well as the constraint functions of the robustness-based design optimization. Generally, multidisciplinary robustness-based design optimization requires uncertainty analysis of the coupled system for estimating the mean and variation of the performance function. Therefore, the efficiency of the robust design methodology depends on the efficiency of the uncertainty analysis method. Du and Chen (2002) proposed efficient uncertainty analysis methods for multidisciplinary problems, namely, the system uncertainty analysis (SUA) method and the concurrent subsystem uncertainty analysis (CSSUA) method. They used these uncertainty analysis methods in the framework of robust design for multidisciplinary systems to achieve computational efficiency. However, these methods have two limitations. Firstly, SUA requires at least one coupled multidisciplinary system level analysis at each iteration of the robust optimization problem, and CSSUA requires a nested double loop formulation when used in a robust optimization framework. Secondly, these methods are developed to account for aleatory and model uncertainty only; no data uncertainty is considered. Du and Chen (2002) proposed another hierarchical collaborative approach to multidisciplinary robust optimization. However, this approach may suffer from convergence issues and like SUA and CSSUA, this method also does not consider data uncertainty. Gu et al (2000) proposed a worst case uncertainty propagation method for multidisciplinary systems and then applied this method to robust design optimization; however, they did not consider data uncertainty. Robust design optimization of multidisciplinary systems has also been studied using game theory methods (Chen and Lewis, 1999; Kalsi et al, 2001). A detailed review of methods for multidisciplinary robust optimization is found in Allen et al (2006).

Most of these multidisciplinary methods deal with aleatory uncertainty only and a few of them deal with both aleatory and model uncertainty. However, uncertainty in system performance may arise from many contributing factors as discussed in Chapter I. The sources of errors in the models of physical systems can be divided into two types: model form error and solution approximation or numerical error (Mahadevan and Rebba, 2006). Model form errors result from approximation about system behavior model, boundary conditions, etc. Solution approximation error may include discretization error as seen in finite element or finite difference methods, truncation error as seen in lower-order approximations in response surface methods, numerical round-off error, etc. There are two ways to include model form errors in design optimization. The first approach assumes model error as a stochastic variable with a mean value 0 and standard deviation being proportional to the corresponding function value (Du and Chen, 2002; Smith and Mahadevan, 2003). The second approach quantifies the model error based on the comparison of model prediction with physical observations (Mahadevan and Rebba, 2006). Mahadevan and Rebba (2006) also developed method to quantify the solution approximation error based on the model itself, using the Richardson extrapolation method. Although all these model uncertainty quantification methods can be conveniently incorporated to our proposed robustness-based design optimization framework, our focus in this dissertation is on the epistemic uncertainty arising from sparse point data and interval data.

A few methods exist for multidisciplinary problems under both aleatory and epistemic uncertainty. Li and Azarm (2008) proposed methods for multidisciplinary robust optimization with interval uncertainty using collaborative optimization. In this method, the uncertain parameters are given as single intervals. This method requires a tolerance region for the coupling variables. The system compatibility requirement is assumed to be satisfied when this tolerance region for the coupling variable is smaller than a predefined tolerance region of the target variable. However, for a multidisciplinary system, the system compatibility requirement should be satisfied at every single point value. Also, this method needs additional non-probabilistic formulations to incorporate epistemic uncertainty into the design optimization framework, which may be computationally expensive. Therefore, an efficient methodology for multidisciplinary robust design optimization under both aleatory and epistemic uncertainty awaits development.

As discussed earlier, most of the existing methods for RBDO can handle only single discipline problems. Multidisciplinary RBDO under aleatory uncertainty alone is a computationally challenging problem. The inclusion of epistemic uncertainty in multidisciplinary RBDO further multiplies this computational effort. Little or no method for multidisciplinary RBDO exists in the literature that can handle both aleatory and epistemic uncertainty. Therefore, there is a need for an efficient RBDO methodology that deals with both aleatory and epistemic uncertainty for multidisciplinary problems.

Chapter IX of this dissertation develops a methodology for robustness-based design optimization for multidisciplinary systems that includes both aleatory and epistemic uncertainty. Chapter X of this dissertation develops a methodology for RBDO for multidisciplinary systems that includes both aleatory and epistemic uncertainty.

CHAPTER III

PROBABILISTIC SYSTEM ANALYSIS WITH SPARSE DATA

1. Introduction

In this Chapter, the problem of reliability analysis under both aleatory uncertainty (natural variability), and epistemic uncertainty (arising when our only knowledge about the random variables is sparse point data) is addressed. First, the epistemic uncertainty arising from a lack of knowledge of the distribution type of the random variables is considered. To address this uncertainty in distribution type, the use of a flexible family of distributions is proposed. The Johnson family of distributions has the ability to reproduce the shape of many named continuous probability distributions, and therefore alleviate the difficulty of determining an appropriate named distribution type for the random variable. We next consider uncertainty in the distribution parameters themselves, and propose the use of computational resampling methods to determine Johnson distributions for the distribution parameters. As a result, we compute the uncertainty in reliability estimates for limit state functions having random variables with imprecise probability distributions as their arguments. We propose an optimization-based approach for computing the bounds on the reliability of a design that allows for the decoupling of epistemic and aleatory uncertainty analysis, enabling computationally affordable approaches to reliability analysis under aleatory and epistemic uncertainty. The proposed methods are illustrated for a problem of uncertainty quantification for drag prediction, where the drag coefficient of a hypersonic aerospace vehicle is to be estimated as a function of its velocity and angle of attack.

The contribution of this chapter is to develop and illustrate an approach for the propagation of both aleatory and epistemic uncertainty in such a way that the epistemic and aleatory uncertainty analyses are not nested, thus enabling computationally efficient calculation of bounds on reliability estimates under epistemic and aleatory uncertainty. In this chapter, we specifically address two types of epistemic uncertainty that arise from sparse data. We first consider epistemic uncertainty arising from a lack of knowledge of the distribution type of the random variables. To address this uncertainty in distribution type, we propose the use of the Johnson family of distributions. The Johnson family of distributions has the ability to reproduce the shape of many named continuous probability distributions, and therefore alleviate the difficulty of determining an appropriate named distribution type for the random variable. We also consider uncertainty in the distribution parameters themselves. We propose the use of computational resampling methods to determine Johnson distributions for the parameters of the Johnson distribution. Finally, we address the uncertainty in the reliability estimates for limit state functions with random variables that have imprecise probability distributions. We propose an optimization-based approach to compute the bounds on the reliability without nesting the epistemic and aleatory uncertainty analyses, thus enabling a computationally affordable approach to reliability analysis under aleatory and epistemic uncertainty. The resulting approach allows for computationally affordable, though approximate, methods of defining and ultimately propagating imprecise probability distributions through computationally expensive simulation models.

The rest of the chapter is organized as follows. Section 2 describes the proposed methodologies for uncertainty representation. Section 3 describes the proposed methods for propagating both epistemic and aleatory uncertainty, using a first-order-reliability method (FORM). Section 4 gives a numerical illustration of the proposed techniques for a problem of uncertainty quantification for drag prediction, where the drag coefficient of a hypersonic aerospace vehicle is to be estimated as a function of its velocity and angle of attack. Section 5 provides conclusions and suggestions for future work.

2. Proposed Methods for Uncertainty Representation

2.1 Fitting the Johnson Distribution to Point Data

The Johnson distribution is a four parameter distribution, and as such it can match the first four moments of a wide variety of probability distribution shapes, thereby allowing it to replicate the shape of many named probability distributions. The shape of the fitted distribution is controlled by the parameters in the function as well as by the transformation function used. A brief description of the Johnson distribution function is provided here. Since Johnson family distribution has the flexibility to fit data with a large range of different distribution function shapes, this eliminates the need to test different distributions that will give the best fit to a set of sample data. Fitting data of a random variable with Johnson distribution involves transforming a continuous random variable x whose distribution is unknown into a standard Normal (z) with one of the four normalizing translations proposed by Johnson (Johnson, 1949). The general form of the translation is:

$$z = \gamma + \delta f\left(\frac{x-\xi}{\lambda}\right) \qquad (1)$$

where $z \sim N(0,1)$. f is the translation functions that map different distributions to the standard Normal distribution. The Johnson's distribution functions are as follows:

$$\begin{aligned} f(y) &= \ln(y), \text{ for lognormal } (\mathrm{S_L}) \text{ distribution} \\ &= \ln\left[y + \sqrt{y^2+1}\right], \text{ for unbounded } (\mathrm{S_U}) \text{ Johnson distribution} \\ &= \ln\left[y/(1-y)\right], \text{ for bounded } (\mathrm{S_B}) \text{ Johnson distribution, and} \\ &= y, \text{ for Normal } (\mathrm{S_N}) \text{ distribution} \end{aligned} \qquad (2)$$

where y = $(x\text{-}\xi)/\lambda$.

DeBrota et al (1988) present four methods to estimate the Johnson distribution parameters. These methods include the method of moments (requiring the first four moments of the data), percentile matching (by using four points and solving a system of nonlinear equations for the distribution parameters), least squares estimation (by minimizing the sum of squared errors in the percentile values of the probability distribution), and minimizing the error norm of the Johnson distribution when compared with the empirical CDF.

Venkataraman and Wilson (1987) implement the above methods, and determine the distribution using the following procedure:

1. Calculate the moments of x: m_2, m_3, and m_4.
2. Calculate the skewness and kurtosis of x: $\beta_1 \equiv m_3^2 / m_2^3$ and $\beta_2 \equiv m_4 / m_2^2$.
3. Use the chart in Figure 1 to determine the appropriate distribution family.

After the distribution parameters of the experimental data have been estimated, regenerating random variable *x* that follows this distribution is easy. The first step is to generate standard Normal random variable *z*. Then *x* can be generated by performing the inverse translation to *z*:

$$x = \xi + \lambda f^{-1}\left(\frac{z-\gamma}{\delta}\right) \quad (3)$$

Note that cumulative probability calculations are much simpler in the standard normal space, allowing for relatively simple calculations of the PDF and CDF of *x*. Examples of PDFs for different Johnson distributions are shown in Figure 2.

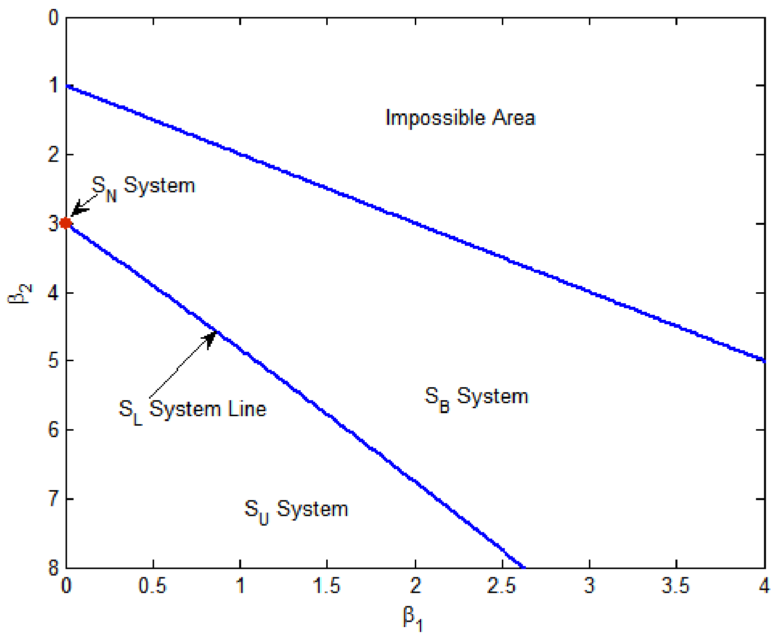
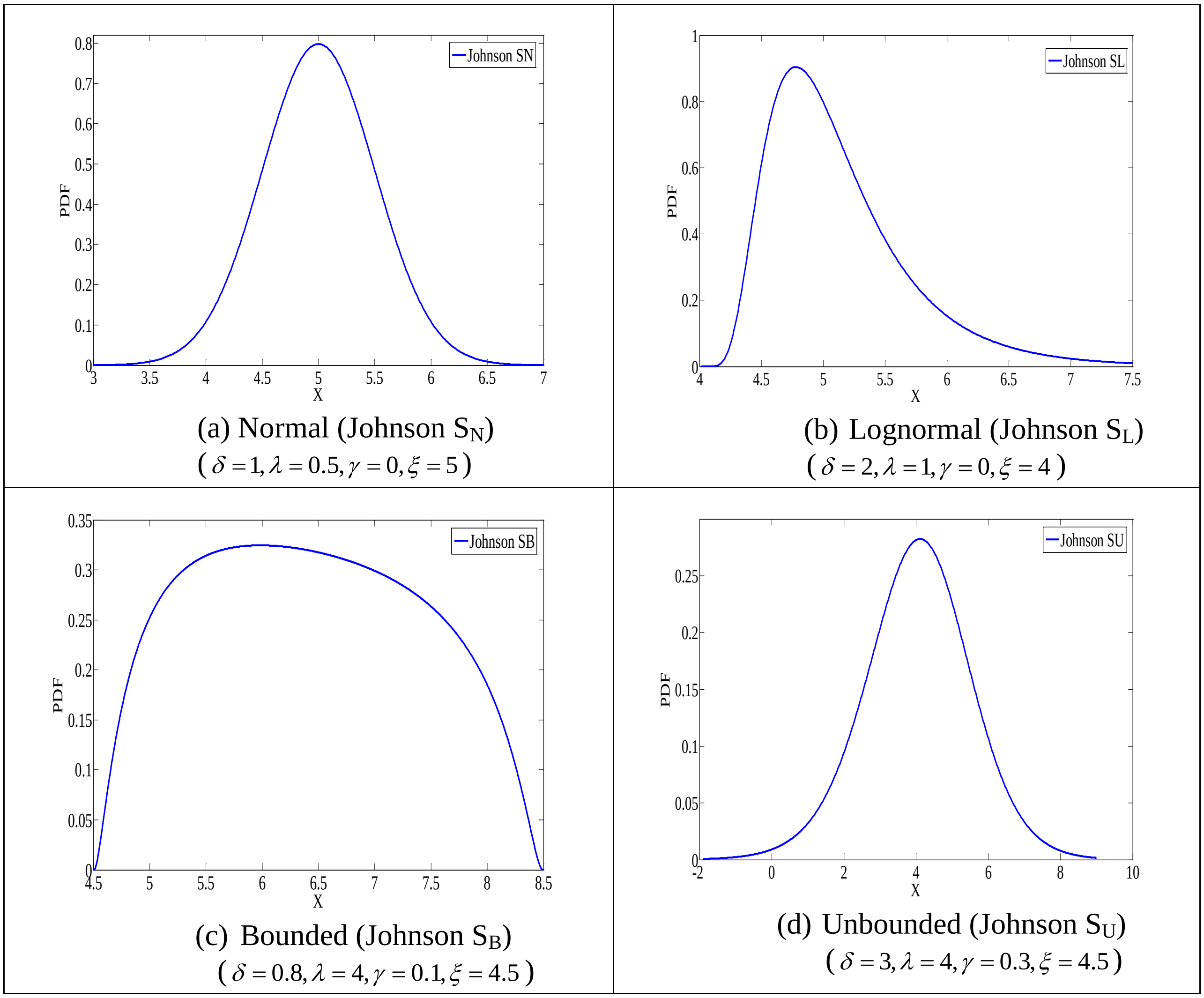


Figure 1: Johnson distribution family identification. $\beta_1 \equiv m_3^2 / m_2^3$ and $\beta_2 \equiv m_4 / m_2^2$. (Marhadi, 2007)



(a) Normal (Johnson S_N) ($\delta = 1, \lambda = 0.5, \gamma = 0, \xi = 5$)

(b) Lognormal (Johnson S_L) ($\delta = 2, \lambda = 1, \gamma = 0, \xi = 4$)

(c) Bounded (Johnson S_B) ($\delta = 0.8, \lambda = 4, \gamma = 0.1, \xi = 4.5$)

(d) Unbounded (Johnson S_U) ($\delta = 3, \lambda = 4, \gamma = 0.3, \xi = 4.5$)

Figure 2: Examples of PDFs for different Johnson distributions

2.2 Statistical Uncertainty Quantification via Jackknife

With the assumption of the Johnson distribution, we are able to alleviate the issue of uncertainty in distribution type. However, it is not possible under small sample sizes to know the precise values of the distribution parameters. Therefore, we introduce a novel and versatile approach for the uncertainty quantification of distribution parameters. This approach assumes that both the basic random variables and their distributions are Johnson distributed, and uses a jackknife technique to estimate the distribution of the distribution parameters. The assumption of the distribution parameters also having the Johnson distribution allows for the possibility of a non-normal distribution for the distribution parameters. This is important, particularly if moment matching is used to estimate the distribution parameters given small sample sizes, for two reasons. First, the estimates resulting from a moment matching approach do not necessarily have asymptotic normality properties as would be the case, for instance, when using a maximum likelihood estimator. Second, even if the estimator had an asymptotic normality property, the sample size may be too small to assume that it holds. In that case, common assumptions that the unknown population mean takes on a normal distribution and that the unknown population variance assumes a chi-square distribution are unwarranted. Our proposed method does not assume any particular distribution type for the distribution parameters, and therefore can be used with any method for distribution parameter estimation, including the method of moments, maximum likelihood, or Bayesian estimation techniques.

Jackknifing (Miller, 1964, 1968, 1974; Arvesen, 1969; Efron, 1979) is used in statistical inferencing to estimate the bias and standard error in a statistic, when a random sample of observations is used to calculate it. The basic idea behind the jackknife estimator lies in systematically recomputing the distribution parameter estimate, leaving out one observation at a time from the sample set. From this new set of "observations" for the statistic an estimate for the bias can be calculated and an estimate for the variance of the parameter. We propose the following algorithm for uncertainty quantification of the distribution parameters:

Algorithm for Uncertainty Quantification in Distribution Parameters

Set i = 1

while (i <= N)

  Delete observation *i* from the original set of observations

  Estimate the Johnson distribution parameters on the basis of the *N*-1 remaining points.

  Record as estimate *i*.

  Restore observation *i* to the set of original observations.

  i = *i* + 1

end while

Fit a Johnson Distribution to the set of parameter estimates obtained in the while loop.

As an illustration of this approach, consider the following set of observations of a random variable X: [5.0, 5.2, 5.5, 6.0, 6.3, 6.5, 7.0, 7.2, 7.5, 8.0]. The PDFs and CDFs of the Johnson distributions estimated on the basis of leaving one observation out are shown below in Figures 3 and 4, respectively.

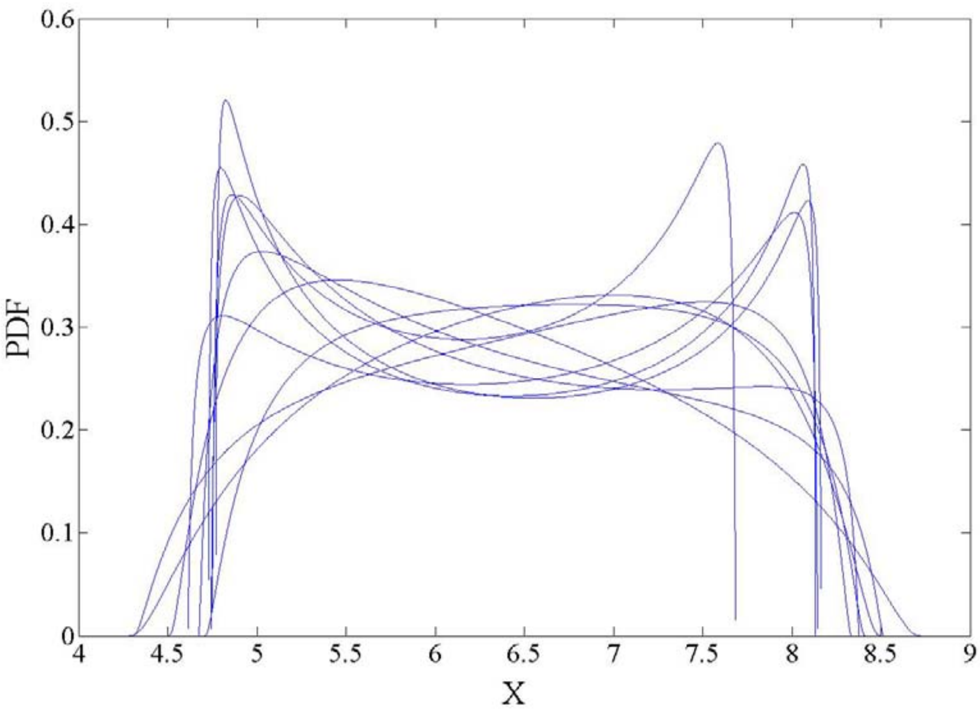


Figure 3: Jackknifed PDF estimates given sparse data

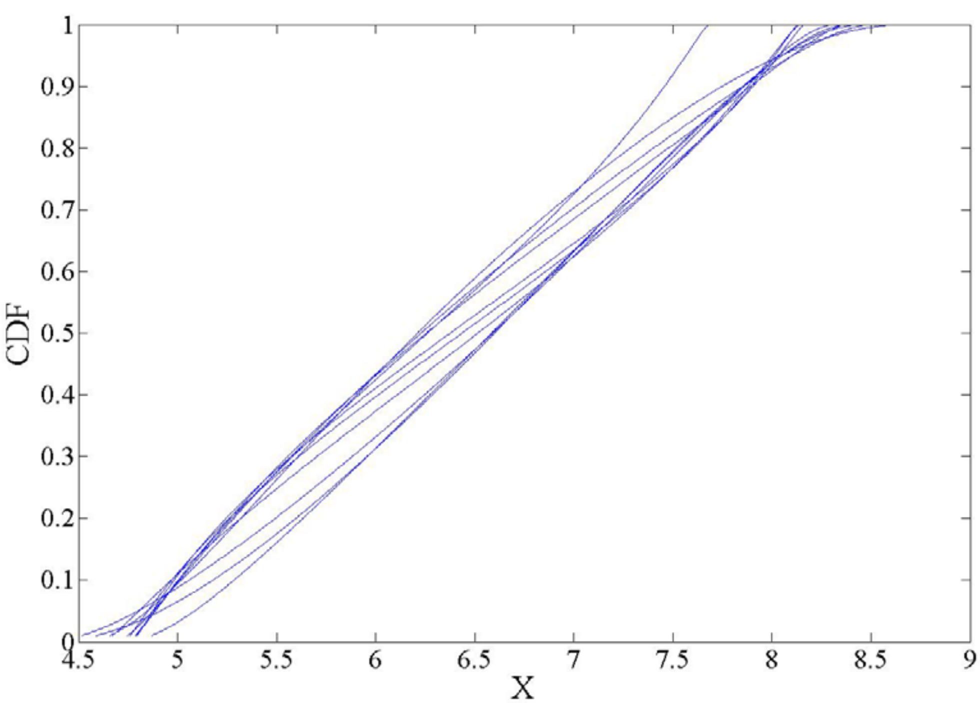


Figure 4: Jackknifed CDF estimates given sparse data

3. Proposed Methods for Uncertainty Propagation

The proposed methods in this section propagate both aleatory and epistemic input uncertainty to calculate the resulting uncertainty in the output. The proposed methods are based on the concepts of FORM, inverse FORM, and sensitivity analysis. A brief overview of these concepts is provided first, and the proposed methods are developed subsequently.

3.1 FORM, Inverse FORM, and Sensitivity Analysis

In model-based reliability analysis, the failure probability estimation problem is given as

$$P_F = P(g(x) \leq k) \qquad (4)$$

It is customary to formulate this problem such that the condition $g < 0$ corresponds to failure, while $g > 0$ corresponds to a condition of safety. The limit state "surface" corresponds to points where $g = 0$.

The rigorous mathematical definition of failure probability requires the evaluation of the integral of the joint probability density function (pdf) of all the random variables over the failure domain as:

$$P_F = P(g < 0) = \int \ldots \int_{g \leq 0} f(x) dx \qquad (5)$$

This integral poses computational hurdles as it can be difficult to formulate the joint probability density explicitly and integration of a multidimensional integral may be difficult. Alternatively, P_F can be evaluated using several methods (first-order second moment (FOSM), first-order reliability method (FORM), second-order reliability method (SORM), inverse FORM, and Monte Carlo simulation), all of them iterative (Haldar and

Mahadevan, 2002). Further details on these methods and computational issues are provided in Haldar and Mahadevan (2002).

In the first order reliability method (FORM), the variables, $\boldsymbol{x}$, which may each be of a different probability distribution, and may be correlated, are first transformed to a space of uncorrelated reduced normal variables $\boldsymbol{u}$. Well-known methods (Haldar and Mahadevan, 2002) are available to transform $\boldsymbol{x}$ to $\boldsymbol{u}$. The closest point to the origin on the function g = 0 in the reduced normal space is then found. This minimum distance point is referred to as the most probable point (MPP) of this limit state, and the distance β is referred to as the reliability index. Then the first-order estimate of P_F is the same as in Eq. 4, i.e. $P_F = \Phi(-\beta)$. The MPP can be calculated as the optimal solution of:

$$\min \| u \| \qquad (6)$$

$$\text{s.t. } g(u) = 0$$

It is also possible to find the extreme value of the response function g for which the probability of exceedence will be equal to $\Phi(\pm\beta_T)$. This is done by solving the following inverse FORM problem:

$$\text{min/max } g(u) \qquad (7)$$

$$\text{s.t. } \| u \| = \beta_T$$

The inverse FORM problem has a very important role in this chapter because it returns a worst-case point at a certain probability level. The inverse FORM formulation is particularly useful in dealing with reliability analysis under uncertainty in the distribution parameters, when the uncertainty in the distribution parameters is described probabilistically using the Jackknife technique. The inverse FORM formulation could be used with the failure probability, conditioned on the values of the distribution parameters, as the *g* function. In this case, inverse FORM would yield the solution to the problem of estimating confidence bounds on the failure probability. Optimization-based approaches to obtain confidence bounds on the reliability estimate are described in detail in the following section.

An additional by-product of FORM is the sensitivity vector $\boldsymbol{\alpha}$. The sensitivity vector is defined as:

$$\alpha = -\frac{\nabla_u G(u)}{\|\nabla_u G(u)\|} \tag{8}$$

The sensitivity vector is collinear with the MPP vector, and its components quantify the influence of each random variable on the reliability index. These components are referred to as probabilistic sensitivity factors. This sensitivity vector shows the relative contribution of each of the random variables to the variance in the limit state function. As such, the alpha vector gives quantitative guidance about which random variables to collect further information. When applied to the parameter space in terms of the inverse FORM problem of finding the distribution parameters that maximize or minimize the failure probability, this vector provides information about the sensitivity of the failure probability estimate to each distribution parameter.

3.2 Optimization-Based Confidence Intervals for CDF and Reliability Estimates

In this chapter, uncertainty analysis is carried out using the probabilistic techniques for reliability analysis described in the last section. We treat epistemic and aleatory uncertainties separately, performing reliability analysis conditioned on a realization of the distribution parameters. Thus there are two sets of uncertain variables in the problem. The first set of uncertain variables, **x**, has aleatory or irreducible uncertainty and is basic to the limit state function, i.e., these variables correspond to quantities such as capacity and load for a structure. The second set of variables has epistemic uncertainty, and is the distribution parameters $\boldsymbol{\theta}$, selected from a set of admissible values $\boldsymbol{\Theta}$. It should be noted that given the presence of epistemic uncertainty, the failure probability is itself uncertain because of the uncertainty in the distributions of the basic random variables. It is desired to determine bounds on this failure probability, given uncertainty in the distribution parameters. Explicit and separate treatment of the epistemic and aleatory variables allows for the calculation of probability distributions of and confidence intervals for the failure probability.

In general, the aleatory uncertainty is propagated using any appropriate probabilistic technique. However, the failure probability is conditioned on a set of distribution parameter values. This conditioning has necessitated nested methods for uncertainty propagation, where a set of distribution parameters would be selected first, and then given these distribution parameters, a reliability analysis would be performed. Mehta et al (1993) proposed formulations that allow for the use of FORM in such a nested manner. The most general problem of calculating bounds on the failure probability would thus be stated as

$$\begin{aligned} &\min/\max \quad P_F(\boldsymbol{\theta}) \\ &w.r.t.\ \boldsymbol{\theta} \\ &s.t.\ \boldsymbol{\theta} \in \boldsymbol{\Theta} \end{aligned} \tag{9}$$

In the reliability analysis, the distributions of the basic random variables $\mathbf{x}$ are conditioned on $\boldsymbol{\theta}$. The cumulative distribution functions for the basic random variables with uncertain probability distributions are calculated by conditioning on a particular realization of the uncertain distribution parameters. Their optimum values are chosen to minimize or to maximize the failure probability.

If FORM is to be used in confidence bounds calculation, then the MPP is given below as a mathematical programming problem with the following generalized statement:

$$
\begin{aligned}
&\min_{\boldsymbol{\theta}}/\max\{\min_{\mathbf{x}} \beta(\mathbf{x},\boldsymbol{\theta})\} \\
&s.t. \\
&g(\mathbf{x}) = 0 \\
&\boldsymbol{\theta} \in \Theta
\end{aligned}
\tag{10}
$$

This nested optimization problem can be decoupled and expressed as:

$$
\begin{aligned}
&\mathbf{x}^* = \arg\min_{\mathbf{x}}(\beta(\mathbf{x},\boldsymbol{\theta}^*) \big| g(\mathbf{x}) = 0) \\
&where \\
&\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}/\max\, \beta(\mathbf{x}^*,\boldsymbol{\theta})
\end{aligned}
\tag{11}
$$

Each optimization problem in Eq. (11) is solved iteratively until convergence.

If the uncertainty in the distribution parameters is represented probabilistically, then it is possible to use the approach of Eq. (11) to calculate confidence bounds on the failure probability. In calculating these confidence bounds, it is useful to define a transformation of the distribution parameters to the standard normal parameter space $\mathbf{u}^{\theta}$. Once this transformation is defined, the second optimization problem can be defined such that the set $\boldsymbol{\Theta}$ becomes a hypersphere in the transformed space of radius β_T. With this definition, Eq. (11) then becomes

$$
\begin{aligned}
&\mathbf{x}^* = \arg\min_{\mathbf{x}}\{\beta(\mathbf{x},\boldsymbol{\theta}^*) \big| g(\mathbf{x}) = 0)\} \\
&where \\
&\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}}/\max\{\beta(\mathbf{x}^*,\boldsymbol{\theta}) \mid \left\|\mathbf{u}^{\theta}\right\| = \beta_T\}
\end{aligned}
\tag{12}
$$

We note that the solution of Eq. (12) guarantees that to first order accuracy the probability of the reliability index associated with the system's limit state exceeding $\beta(\mathbf{x^*}, \boldsymbol{\theta^*})$ is $\Phi(\pm\beta_T)$. Hence the solution of the problem gives confidence bounds on the failure probability with the 1 - $\alpha/2$ confidence level equal to $\Phi(-|\beta_T|)$.

If the failure probability of an entire series or parallel system is of concern, MCS could be used directly with Eq. (9) where the failure or safety of all components in the system is evaluated for each randomly generated sample point. Alternatively, the MPP for each component could be determined using FORM for each limit state function, and the system reliability would become the objective function for the second optimization problem in Eq. (12).

It should be noted that there are no system response function evaluations required for the inverse FORM analysis with the epistemic variables. In other words, if expensive structural or CFD codes are required to evaluate the limit state function for the purposes of reliability analysis, in this decoupled formulation, no evaluations are required to find the values of the distribution parameters which minimize or maximize the likelihood of the MPP. This is because the second problem (the inverse FORM problem) in Eq. (11) manipulates the transformation to normality only, and does not involve solution of the first problem (the direct FORM problem) in which limit state functions are required to evaluate the gradients of the objectives and constraints. The inverse FORM reliability analysis finds the worst-case values of the distribution parameters so that the failure probability is maximized, or best case parameters such that the failure probability is minimized. When the two optimization problems converge, we have first order estimates of the failure probability by solving the reliability analysis, where the expensive function evaluations are encountered, only a few times in this decoupled formulation.

As in direct FORM for the case of certain probability distributions, sensitivity analysis can be performed on both the epistemic and aleatory uncertainties using the sensitivity vector α. The interpretation of the sensitivity vector α (see Eq. 8) for the aleatory random variables is much the same as in the case with probability distributions with no randomness. However, the alpha vector for the distribution parameters also lends important information to the decision maker. This vector gives an indication of the sensitivity of the failure probability to the uncertainty in each distribution parameter. Sensitivities of distribution parameters near zero indicate that the outcome of the design problem is unlikely to change, regardless of the value of the distribution parameter. High sensitivities, however, indicate the distribution parameter has a large influence on the reliability estimate. This information can be used in determining the variables for which to pursue more intensive data collection.

4. Numerical Illustration

In this section, the proposed methods are applied to a single aerodynamic data set for the upper stage of the Two-Stage-To-Orbit (TSTO) concept vehicle, as shown in Figure 1 of Chapter I. The objective is to quantify the uncertainty in the predicted drag, given uncertainty in the flight conditions.

Because the fine grid (3,800,000 grid points) computational fluid dynamics code is too expensive for uncertainty propagation analysis, a design of computational experiments has been conducted to construct a surrogate model. The following cases have been analyzed to develop a surrogate model for drag prediction. The following values for the Mach number (Mach) were selected: 0.5, 0.9, 1.1, 1.2, 1.6, 2.0, 4.0, 6.0, 8.0, 10.0, and 12.0. The following values of Angle of Attack (AoA) were selected: 0, 5, 10, 15, 20, 30, and 40 degrees. Instead of the expected $11 \times 7 = 77$ data points for every combination of Mach and AoA, only 68 points were used as the analysis code failed to converge for remaining 9 points.

Centerline and surface pressure contours for a representative case (Mach = 2.0, and angle of attack = 10 degrees) are shown in Figure 5. From the uncertainty analysis point of view of this chapter, only two variables are of interest: Mach number and angle of attack. Each calculation by the Cart3D code is presumed here to be deterministic. Any issues related to the repeatability of individual results from this code, or any other data source, are beyond the scope of this chapter and are within the domain of code verification, rather than uncertainty quantification. Given a Mach number and an angle of attack, a selection of Cart3D options to be used within the calculation, and a prescription for the process used to capture the results, there is no uncertainty within any individual result.

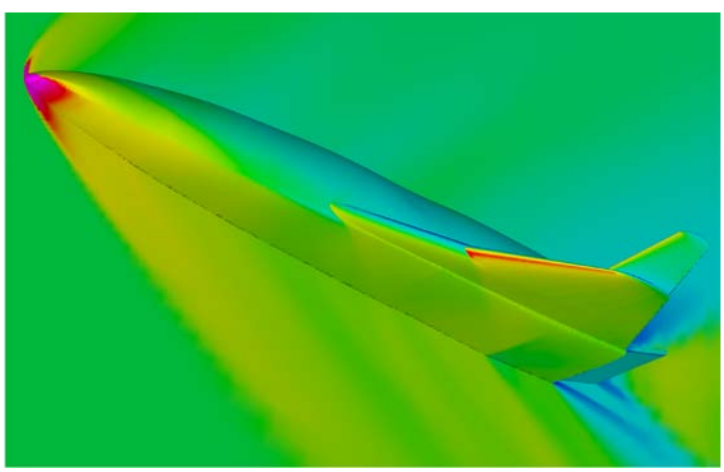


Figure 5: Centerline pressure contours for TSTO upper stage at Mach = 2.0, AoA = 10 degrees

A response surface (Equation 13) for a model-predicted drag coefficient (C_D) has been created as a function of Mach number (Mach) and angle of attack (AoA).

$$C_D = 0.050269 - 0.015291 * \text{Mach} + 4.02211\text{E} - 004 * \text{AoA} - \\ 7.04277\text{E} - 004 * \text{Mach} * \text{AoA} + 1.44863\text{E}003 * \text{Mach}^2 + 4.61108\text{E}\ 004 * \text{AoA}^2 \tag{13}$$

A surface plot of the drag coefficient is given in Figure 6. We wish to determine the 95 percent confidence interval for the probability that the drag coefficient exceeds 0.15. Thus, we will use a limit state function of

$$g(\text{Mach}, \text{AoA}) = 0.15 - (0.050269 - 0.015291 * \text{Mach} + 4.02211\text{E} - 004 * \text{AoA} - \\ 7.04277\text{E} - 004 * \text{Mach} * \text{AoA} + 1.44863\text{E}003 * \text{Mach}^2 + 4.61108\text{E}\ 004 * \text{AoA}^2) \tag{14}$$

and use analytical reliability methods to evaluate the exceedence probability. Mach and AoA are described by sparse point data as given in Table 1.

Table 1: Data for Mach and AoA

Data	
Mach	AoA
6.52	21.73
6.06	20.19
5.49	18.30
6.52	21.75
5.74	19.13
5.74	19.14
5.34	17.79
6.24	20.79

5.42	18.07
6.10	20.34

As the variability of the mission parameters are described by sparse point data, this creates uncertainty about the distribution parameters of Mach and AoA. In this example, it is assumed that Mach and AoA as well as their distribution parameters are characterized by bounded Johnson distributions. We follow the procedure described in Section 2 to obtain the distributions of each distribution parameter of Mach AoA as given in Tables 2 and 3.

Table 2: Distribution parameters for distribution of Mach

	δ	λ	γ	ξ
δ^{Mach}	0.2194	0.2330	-0.0250	0.3991
λ^{Mach}	0.3493	0.1905	0.2098	1.3401
γ^{Mach}	0.2555	0.3528	-0.0014	-0.0811
ξ^{Mach}	0.5263	0.1652	-0.1672	5.1554

Table 3: Distribution parameters for distribution of AOA

	δ	λ	γ	ξ
δ^{AoA}	0.2194	0.2330	-0.0250	0.3991
λ^{AoA}	0.3493	0.6350	0.2098	4.4670
γ^{AoA}	0.2555	0.3528	-0.0014	-0.0811
ξ^{AoA}	0.5263	0.5506	-0.1672	17.1846

The set of admissible distribution parameter values is found by transforming the distribution parameters to the standard normal space and considering only those distribution parameters for which their image in the $\mathbf{u}^{\theta}$ space fall on a sphere centered at the origin having radius 1.96. Thus, we will use the problem statement of Eq. (10) to calculate a 95 percent confidence interval for the probability of the drag coefficient taking on a value of 0.15 or greater.

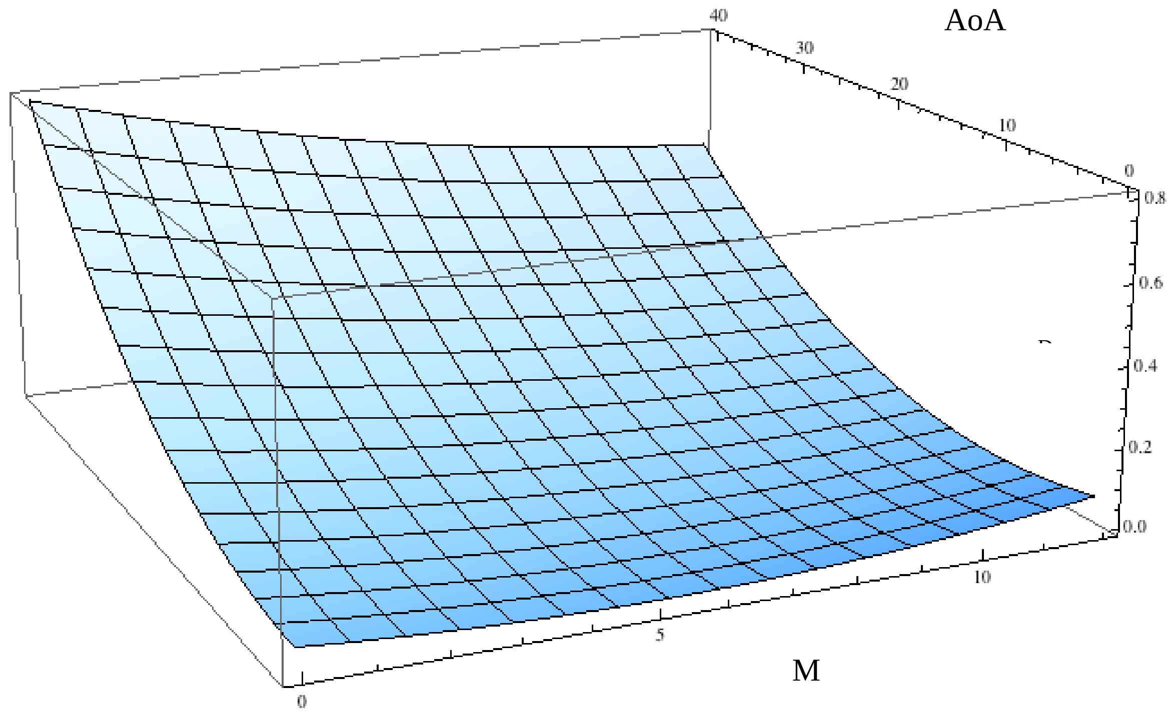


Figure 6: Drag Coefficient Response Surface

By solving Eq. (10), the 95 percent confidence interval for the exceedance probability is found to be (0.0126, 0.2732). The worst-case distribution parameters and sensitivities are given in Tables 4 and 5. From Table 5 we see that uncertainty in the distribution of AoA is more important than the uncertainty in the distribution of Mach. This is intuitive considering the larger gradients of the response surface in the AoA direction and the wider distribution of AoA.

Table 4. Worst Case Distribution Parameters
(P_f = 0.2732)

	δ	λ	γ	ξ
Mach	0.4500	1.3957	0.2091	5.2328
AoA	0.4144	4.9903	-0.0529	17.6268

Table 5. Worst-Case Sensitivities
(P_f = 0.2732)

Aleatory Sensitivities	
Variable	Sensitivity
Mach	-0.2999
AoA	0.9541
Epistemic Sensitivities	
Variable	Sensitivity
δ^{Mach}	-0.1554
λ^{Mach}	-0.0509
γ^{Mach}	0.1992
ξ^{Mach}	-0.1192
δ^{AoA}	-0.3100
λ^{AoA}	0.3823
γ^{AoA}	-0.3192
ξ^{AoA}	0.2922

The best-case parameter values (where P_f = 0.0126) and sensitivities are given in Tables 6 and 7.

Table 6. Best Case Distribution Parameters
(P_f = 0.0126)

	δ	λ	γ	ξ
Mach	0.5954	1.4179	0.0000	5.2729
AoA	0.6122	4.5163	0.2130	17.2949

Table 7. Best-Case Sensitivities
(P_f = 0.0126)

Aleatory Sensitivities	
Variable	Sensitivity
Mach	-0.5101

AoA	0.8601
Epistemic Sensitivities	
Variable	Sensitivity
δ^{Mach}	0.1750
λ^{Mach}	0.0410
γ^{Mach}	-0.1583
ξ^{Mach}	0.1568
δ^{AoA}	0.2527
λ^{AoA}	-0.3340
γ^{AoA}	0.2094
ξ^{AoA}	-0.4570

From Table 7 we see that uncertainty in the distribution of AoA is more important than the uncertainty in the distribution of Mach. This is intuitive considering the larger gradients of the response surface in the AoA direction and the larger variance in AoA. Because the limit state function is very sensitive to AoA, and the scatter of the distribution is wide, it is obvious that the failure probability is very sensitive to the uncertainty in the distribution of AoA. The CDFs of the worst and best case distributions of Mach and AoA are shown in Figures 7 and 8, respectively.

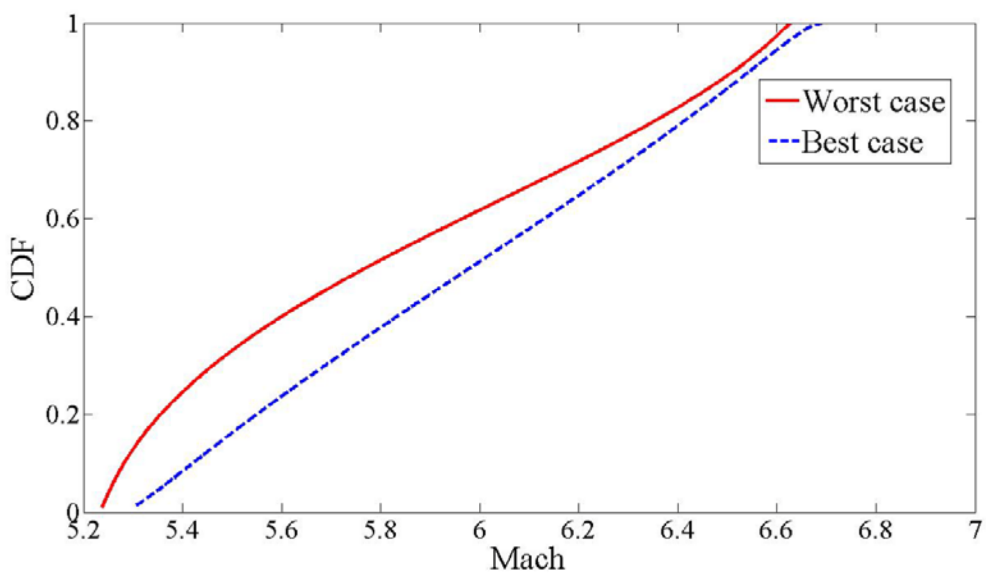


Figure 7: The worst and best case distribution parameters of Mach

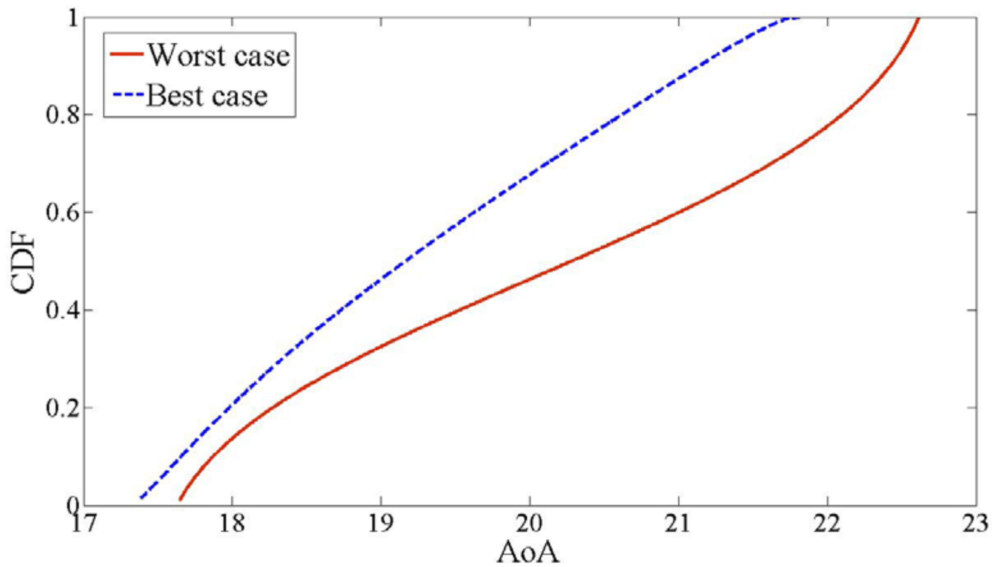


Figure 8: The worst and best case distribution parameters of AoA

The use of optimization methods in calculating confidence bounds on the failure probability makes the proposed method computationally efficient (483 function evaluations) as compared to a sampling-based method (e.g., MCS). If we used the sampling method to calculate the confidence bounds, we would require $N \times n$ function evaluations (e.g., 10×10000), where N is the sample size of sparse point data and n is the MCS sample size.

5. Conclusion

This chapter developed a methodology for propagating both aleatory and epistemic uncertainty arising from sparse data through computational models of system response. A flexible Johnson family of distributions is used to represent variables with sparse data. The methodology differs from existing approaches in that it infers Johnson probability distributions to the distribution parameters also by use of computational resampling methods. Once the uncertainty in the distribution parameters is quantified, the reliability analysis of the system uses probability distributions conditioned on the distribution parameter values. An efficient optimization-based method for calculating the confidence intervals of the failure probability is developed based on FORM. This method eliminates the computationally expensive process of nesting an aleatory uncertainty analysis inside an epistemic uncertainty analysis. This methodology also affords sensitivity analysis information with regard to each of the distribution parameters as well as the basic random variables. The results of the sensitivity analysis give quantitative guidance regarding data collection for the random variables.

CHAPTER IV

A PROBABILISTIC APPROACH FOR REPRESENTATION OF INTERVAL UNCERTAINTY

1. Introduction

In this chapter, we propose a probabilistic approach to represent interval data for input variables in reliability and uncertainty analysis problems, using flexible families of continuous Johnson distributions. Such a probabilistic representation of interval data facilitates a unified framework for handling aleatory and epistemic uncertainty. For fitting probability distributions, methods such as moment matching are commonly used in the literature. However, unlike point data where single estimates for the moments of data can be calculated, moments of interval data can only be computed in terms of upper and lower bounds. Finding bounds on the moments of interval data even within some given finite accuracy has been conjectured to be an NP hard problem because it includes a search among the combinations of multiple values of the variables, including interval endpoints. In this chapter, we present efficient algorithms based on continuous optimization to find the bounds on second and higher moments of interval data. With numerical examples, we show that the proposed bounding algorithms are scalable in polynomial time with respect to increasing number of intervals. Using the bounds on moments computed using the proposed approach, we fit a family of Johnson distributions to interval data. Furthermore, using an optimization approach based on percentiles, we find the bounding envelopes of the family of distributions, termed as a Johnson p-box.

The idea of bounding envelopes for the family of Johnson distributions is analogous to the notion of empirical p-box in the literature. Several sets of interval data with different numbers of intervals and type of overlap are presented to demonstrate the proposed methods. As against the computationally expensive nested analysis that is typically required in the presence of interval variables, the proposed probabilistic representation enables inexpensive optimization-based strategies to estimate bounds on an output quantity of interest.

Within the context of reliability analysis, it is often required that a certain function $g(\mathbf{x})$ of input variables $\mathbf{x}$, representing a response of the designed system, lie within given bounds. In many cases, the values of some elements of $\mathbf{x}$ are uncertain, and this uncertainty may be of aleatory or epistemic type. Aleatory uncertainty can be represented by using probability distributions. In some cases of epistemic uncertainty, the distribution for $\mathbf{x}$ must be determined from imprecisely available data, such as intervals given by experts. This implies that the cumulative distribution function of $\mathbf{x}$, and subsequently that of $g(\mathbf{x})$, denoted as $F(g(\mathbf{x}))$, cannot be known precisely. Instead of formulating design requirements in terms of failure probabilities, the requirements may then have to be formulated as bounds on the cumulative distribution function $F(\mathrm{g}(\mathbf{x}))$ of the function $g(\mathbf{x})$. In this chapter, we focus on the representation of epistemic uncertainty arising from interval data in the input variables $\mathbf{x}$, where a variable's possible values are described by intervals.

As discussed in Chapter II above, there are various approaches for treating interval data, each with their own advantages and limitations. One of the drawbacks of the current approaches is the need for nested analysis in the presence of interval variables. To alleviate this issue, we propose a probabilistic representation for interval data using a family of Johnson distributions. A new aggregation technique is proposed to combine multiple intervals. This aggregation technique enables the use of the method of matching moments to represent the uncertainty described by the multiple intervals through a family of probability distributions. An important advantage of the proposed approach is that it allows for a unified probabilistic framework to be applied that can jointly handle aleatory and epistemic uncertainties, thereby allowing for well developed and efficient analytical probabilistic methods such as FORM and SORM to be used in uncertainty propagation. The proposed representation avoids the expensive nested analysis by enabling the use of an optimization-based strategy that can estimate the distribution parameters of the input variables that maximize or minimize an output quantity of interest.

It is a common practice in the literature to use methods such as moment matching and percentile matching to fit probability distributions to data sets. However, describing interval data in a probabilistic format is not straightforward. Unlike point data, where statistics such as moments have precise values, statistics for interval data are usually described by their upper and lower bounds. Finding bounds on the statistics of interval data is a computationally challenging problem because it typically involves interval analysis that is conducted using a combinatorial search. It has been reported that computing the upper bound on second moment of overlapping intervals is conjectured to be an NP-hard problem even if some given finite accuracy in the moment bounds is of interest (Kreinovich 2004, Ferson et al 2007), although polynomial time algorithms have been reported for some special cases (Kreinovich et al 2006). Little to no work exists in the literature about bounds on higher moments. Most previous approaches that calculate bounds on moments combinatorially search for points within the intervals that minimize or maximize the moments of the data. A major contribution of this chapter is the development of algorithms based on continuous optimization methods which scale polynomially in computational effort with respect to the number of intervals. Knowledge of the bounds on moments on the interval data is useful because it provides restrictions on the possible distributions the underlying random variable may assume. Using the moment bounds computed using the proposed algorithms, we develop a probabilistic representation of the interval as a Johnson p-box, which is an ensemble of bounded Johnson distributions.

The main contributions of this chapter are summarized as follows. First, we present approaches based on continuous optimization to find the bounds on second and higher moments of interval data with single and multiple intervals. Second, we demonstrate using numerical examples that these algorithms are scalable in polynomial time with respect to increasing number of intervals. Third, using the bounds on moments, we fit a family of Johnson distributions to interval data. Analogous to the notion of empirical p-box as the bounding envelope for empirical distributions, we construct a Johnson p-box, which represents the bounding envelope for Johnson distributions for interval data.

The remainder of the chapter is organized as follows. Section 2 develops the methods for estimating moment bounds for interval data and Section 3 develops a probabilistic approach for the representation of interval uncertainty. Section 4 illustrates the proposed developments using different examples of interval data, where comparisons with alternate representations such as the empirical p-box are made. Section 5 concludes the chapter with summary and future work.

2. Estimating Bounds on Moments for Interval Data

This section discusses the proposed algorithms that estimate bounds on moments for interval data for single and multiple interval cases. A brief background is provided first.

In this dissertation, we fit a family of Johnson distributions to interval data using the moment matching approach. Moment matching involves equating the moments derived from data to those of the probability distribution being fit. The Johnson family is a generalized family of distributions that can represent normal, lognormal, bounded, or unbounded distributions. While there are several other viable four-parameter distributions that may also be used with this approach, such as the Pearson, Beta, and Lambda distributions, the Johnson family is a convenient choice. This is because the Johnson distribution lends itself to easy transformation to a standard normal space, which then can be conveniently applied in well known reliability analysis and reliability-based design optimization methods.

Among other methods (see Chapter III), we use the moment matching approach in this dissertation to take advantage of the moment bounding algorithms developed in this section. Moreover, to determine the appropriate type of Johnson distribution (bounded, unbounded, normal, lognormal), we need to compute the moments of the data set. While it is possible to have point estimates for the moments of point data, moments on interval data must be described using upper and lower bounds. As discussed in Section 1, it is challenging to compute bounds on moments of a variable described by multiple intervals. Note that in this dissertation, we assume that the multiple interval data are obtained from equally credible sources. As discussed in Section 1, this is a common assumption in the literature. The reason is that in absence of any additional information regarding the relative credibility of each source; it is reasonable to assume that all sources of information are equally credible.

In the following subsections, we propose methods that can compute lower and upper bounds on the first four moments for single and multiple interval cases.

2.1 Bounds on Moments for Single Interval

In this subsection, we outline the proposed method to estimate bounds on moments for a single interval case.

In order to estimate the bounds on moments, we first find the probability mass function (PMF) of the end points of the interval that minimize or maximize the moments of the single interval data. The following procedure is used:

1. Sample *ns* data points from the given interval (both endpoints included)

2. Solve the following optimization problems with the PMFs, $p(x_i)$, $i = \{1, \ldots, ns\}$, as the decision variables:

$$\min_{p(x_i)}/\max \quad M_k \quad k = 1,2,3, or\ 4 \tag{1}$$

$$s.t. \quad \sum_{i=1}^{ns} p(x_i) = 1 \tag{2}$$

Here, $M_1 = E(x)$

$$M_2 = E(x^2) - (E(x))^2 \tag{3}$$

$$M_3 = E(x^3) - 3E(x^2)E(x) + 2(E(x))^3$$

$$M_4 = E(x^4) - 4E(x^3)E(x) + 6E(x^2)(E(x))^2 - 3(E(x))^4$$

where, $$E(x) = \sum_{i=1}^{ns} x_i p(x_i)$$

$$E(x^2) = \sum_{i=1}^{ns} x_i^2 p(x_i) \qquad (4)$$

$$E(x^3) = \sum_{i=1}^{ns} x_i^3 p(x_i)$$

$$E(x^4) = \sum_{i=1}^{ns} x_i^4 p(x_i)$$

Note that the above formulas for the third and fourth moments have been derived from the definition of moments as given below (DeGroot, 1984):

Consider a random variable X for which the first moment i.e., the expectation of X is $E(X) = \mu$. Then for any positive integer *k*, the expectation $E\left[(X-\mu)^k\right]$ is called the *k*th central moment of the variable X or the *k*th moment of X about the mean value.

2.1.1 Bounds on first moment for single interval

For the lower bound on the first moment, the above minimization yields that the probability mass function (PMF) at the lower endpoint of the interval is the Dirac delta function, i.e., PMF is equal to one at this point and zero elsewhere. Thus the lower bound on the mean for a single interval is the lower bound of the interval itself. Similarly, the upper bound on the mean for a single interval occurs when the probability mass function (PMF) at the upper endpoint of the interval is the Dirac delta function. The upper bound on the mean for a single interval therefore is the upper bound of the interval. If we estimated the bounds on the first moment of single interval data based on observation, we would get the exact same results.

2.1.2 Bounds on second moment for single interval

For the lower bound on the second moment, the above minimization yields that the PMF at any point within the interval is the Dirac delta function, which implies that the lower bound on variance for a single interval is zero. Similarly, for the upper bound on the second moment, the above maximization yields a PMF of 0.5 at the both endpoints of the single interval.

2.1.3 Bounds on third moment for single interval

For the lower bound on the third moment, the above minimization yields a PMF of 0.2113 for the lower endpoint and 0.7887 for the upper endpoint. Similarly, a PMF of 0.7887 for the lower endpoint and 0.2113 for the upper endpoint is obtained for the upper bound on the third moment (maximization).

2.1.4 Bounds on fourth moment for single interval

For the lower bound on the fourth moment, the above minimization yields that the PMF at any point within the interval is the Dirac delta function, which implies that the lower bound on the fourth moment for a single interval is zero. For the upper bound on the fourth moment, the above optimization yields a PMF of 0.7887 for one of the endpoints and 0.2113 for the other.

We summarize these methods in Table 1 below. Note that these values of PMFs for the end points hold irrespective of the actual data represented by the single interval. For a given single interval, one could therefore directly use the above PMFs to estimate the lower and upper bounds on the moments, without having to repeat the optimization for each problem. We also note that we have solved the above mentioned optimization problems for four different sample sizes i.e., by discretizing the single interval into four different sizes (10, 100, 500, and 1000) and obtained the exact same results with linear computational efforts. The nature of the sampling or discretization does not have any effect on the end results as long as the samples include the two endpoints of the single interval data.

Table 1: Methods for calculating moment bounds for single interval data

Moment	Condition		Formula
	Lower bound	Upper bound	
1	PMF = 1 at lower endpoint = 0 elsewhere	PMF = 1 at upper endpoint = 0 elsewhere	$M_1 = E(x)$
2	PMF = 1 at any point = 0 elsewhere	PMF = 0.5 at each endpoint	$M_2 = E(x^2) - (E(x))^2$
3	PMF = 0.2113 at lower endpoint = 0.7887 at upper endpoint	PMF = 0.7887 at lower endpoint = 0.2113 at upper endpoint	$M_3 = E(x^3) - 3E(x^2)E(x) + 2(E(x))^3$
4	PMF = 1 at any point = 0 elsewhere	PMF = 0.7887 at one of the endpoints = 0.2113 at the other endpoint	$M_4 = E(x^4) - 4E(x^3)E(x) + 6E(x^2)(E(x))^2 - 3(E(x))^4$

Note: $E(x) = \sum_{i=1}^{2} x_i p(x_i)$ $E(x^2) = \sum_{i=1}^{2} x_i^2 p(x_i)$ $E(x^3) = \sum_{i=1}^{2} x_i^3 p(x_i)$ $E(x^4) = \sum_{i=1}^{2} x_i^4 p(x_i)$
where $p(x_i)$ = Probability Mass Function (PMF)

It is seen from the optimization results that the minimum and maximum of the moments occur, when all the probability masses are concentrated at the two endpoints only, with two exceptions as seen for the lower bounds on the second and fourth moments. This is intuitive for the lower and upper bounds on the first moment. However, for the other cases, we investigate this issue as follows:

With regard to the proposed algorithm, the following can be stated from the definition of moments as mentioned earlier in this section:

1. As the second and fourth moments are by definition positive, the lower bounds on these moments are zero with the Probability Mass Function PMF being the Dirac delta function at any point within the interval.

2. As the moments are by definition, the expectation of powers of deviation from the mean value, these expectations are essentially minimum (for the third moment) or maximum (for the second, third and fourth moments), when the data points are located at the endpoints of the interval i.e., when the PMFs are concentrated only at the endpoints of the single interval.

Once we know that for minimum and maximum of some moments, the PMFs concentrate only on the two endpoints of the single interval, it might be interesting to investigate the nature of the solutions. We plot the values of the moments as a function of the pair (w_1, w_2), where w_1 is the PMF at the lower endpoint and $w_2 = 1-w_1$ is the PMF at the upper endpoint of the interval. It is seen in Figure 1 that the second moment reaches its maximum when PMFs at both the endpoints are 0.5 each, which are consistent with our optimization results. For the third moment, we get a symmetric shape, which is consistent with our optimization results, where we have found that the PMFs at both the end points get flipped for the minimization (0.2113, 0.7887) and maximization problems (0.7887, 0.2113). For the fourth moment, we get a bi-modal shape. The curve reaches its maximum for two sets of PMF pairs (0.2113, 0.7887) and (0.7887, 0.2113), which are consistent with our optimization results. These two sets of PMFs also correspond to the minimum and maximum of the third moments, respectively as seen in Figure 1.

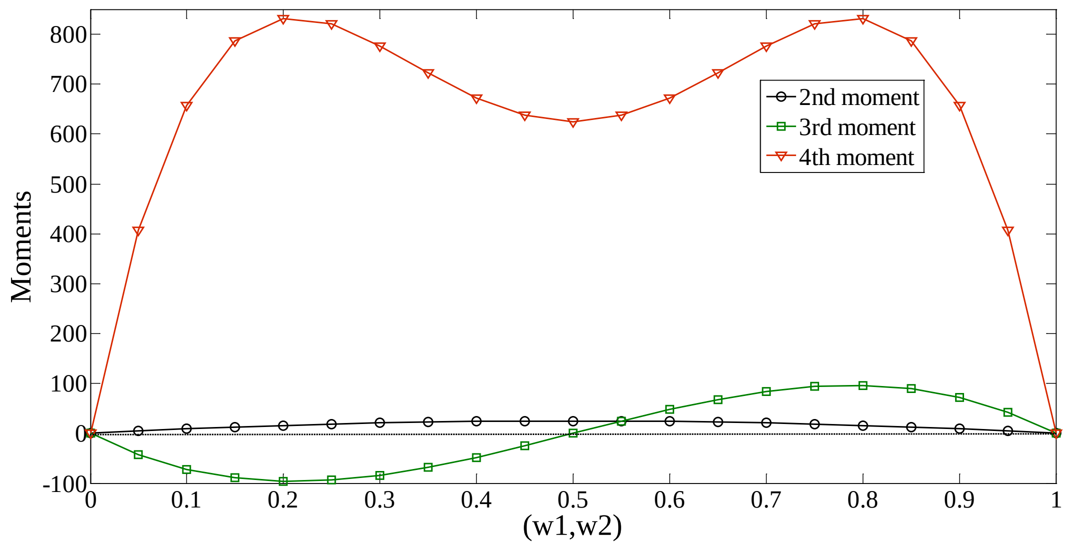


Figure 1: Moments vs. PMFs at the interval endpoints

2.1.5 Numerical Example

We apply the proposed method of estimating bounds on a single interval to the following example: [5, 15]. The bounds on the first moment are calculated to be [5, 15], those on the second moment are [0, 25], those on the third moment are [-96.225, 96.225], and those on the fourth moment are [0, 833.333]. We use this example later in the chapter to illustrate subsequent steps in the proposed methodology.

2.2 Bounds on moments for multiple intervals

As discussed in Section 1, the computation of bounds on moments for multiple intervals is computationally expensive as it is usually treated as a combinatorial problem, where the moments are calculated at the combinations of possible values of the interval variable. Rather than deal with this problem combinatorially, we have formulated this computation as a nonlinear programming problem with the objective being minimization or maximization of the moments of data points that are constrained to fall within each of the respective intervals. The computational effort of this approach with increasing number of variables is demonstrated to be of polynomial order in the number of intervals. The proposed formulations are valid for any type of interval data, i.e. overlapping or non-overlapping intervals. The bounds on moments thus found are rigorous, i.e., they completely enclose all possible moments generated from various combinations of the interval data.

2.2.1 Bounds on first moment for multiple intervals

Consider a set of intervals given as $a_i \le x_i \le b_i$, $i = \{1, ..., n\}$ where n is the number of intervals. Estimating the bounds on the first moment (arithmetic mean) involves identifying a configuration of scalar points $\{x_i, i = \{1, ..., n\}\}$, (where x_i indicates the true value of the observation within the interval) within the respective intervals that yield the smallest possible mean, and a configuration that yield the largest possible mean. Because the mean is proportional to the sum of the interval data, the configuration for the lower bound on the mean is the set of left endpoints of the interval, and that for the upper bound on the mean is the set of right interval endpoints. The formula for the arithmetic mean of interval data x_i is therefore

$$\left[\underline{M} \quad \overline{M}\right] = \left[\frac{1}{n}\sum_{i}^{n} a_i, \frac{1}{n}\sum_{i}^{n} b_i\right] \tag{5}$$

where $\left[\underline{M}, \overline{M}\right]$ are the lower and upper bounds on the mean, respectively.

2.2.2 Bounds on second moment for multiple intervals

The second central moment (variance) is a quadratic function of each of the values of its data. We search for the configuration of scalar points, x_i, constrained to lie within their respective intervals that minimizes (or maximizes) the function shown below to yield the lower (or upper) bound on the variance. Therefore, we construct a linearly constrained optimization problem as follows:

$$\min_{x_1,\ldots,x_n}/\max \quad M_2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right) - \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i\right)^2 \tag{6}$$

$$\text{s.t.} \qquad lb_i \le x_i \le ub_i \quad i = \{1,\ldots,n\} \tag{7}$$

2.2.3 Bounds on third and fourth moments for multiple intervals

The third and fourth central moments are third and fourth order polynomial functions of each of the values of the data, respectively. We search for the configuration of points $\{x_i,\ i = \{1, 2,.., n\}\}$ constrained to lie within their respective intervals that minimizes (or maximizes) the function shown below to yield the lower (or upper) bound of the third/fourth moment.

$$\min_{x_1,\ldots,x_n}/\max \quad M_k = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)^k \tag{8}$$

$$\text{s.t.} \qquad lb_i \le x_i \le ub_i \quad i = \{1,\ldots,n\} \tag{9}$$

where minimizing (or maximizing) the above problem with $k = 3$ and $k = 4$ yields the lower (or upper) bound on the third and fourth moments, respectively.

We have implemented the formulations to calculate the lower and upper bounds on the second, third and fourth moments for various test cases with increasing number of intervals. We considered both overlapping and non-overlapping interval examples to demonstrate the performance of the proposed formulations. The following procedure was used to generate the intervals for overlapping interval test cases. The interval extremes (lowest of the lower bound and the highest of the upper bound) were arbitrarily assumed. In order to generate a desired number of intervals for each test case, a uniform random number generator was used to generate overlapping intervals between interval extremes. To generate non-overlapping interval data with *n* intervals for the test problems, we used the following procedure. First, a sequence of monotonically increasing random numbers is generated, {1 ,…, 2×*n*}. The *i*-th interval is generated by collecting the (2*i*-1)-th and (2*i*)-th random number. Thus the interval widths and the end points are generated randomly.

We solved the above optimization formulations in Eqs. (6)-(9) using the MATLAB function *fmincon*, which implements a sequential quadratic programming algorithm. The plots in the Figures 2 and 3 illustrate the scalability of the proposed formulations with increasing number of intervals for overlapping and non-overlapping cases, respectively. For each plot shown in Figures 2 and 3, we fit a linear or quadratic function as well as an exponential function (solely for comparison purposes). The regression coefficients (i.e., the values of R^2) indicate a strong linear/quadratic trend for the scalability of the algorithms.

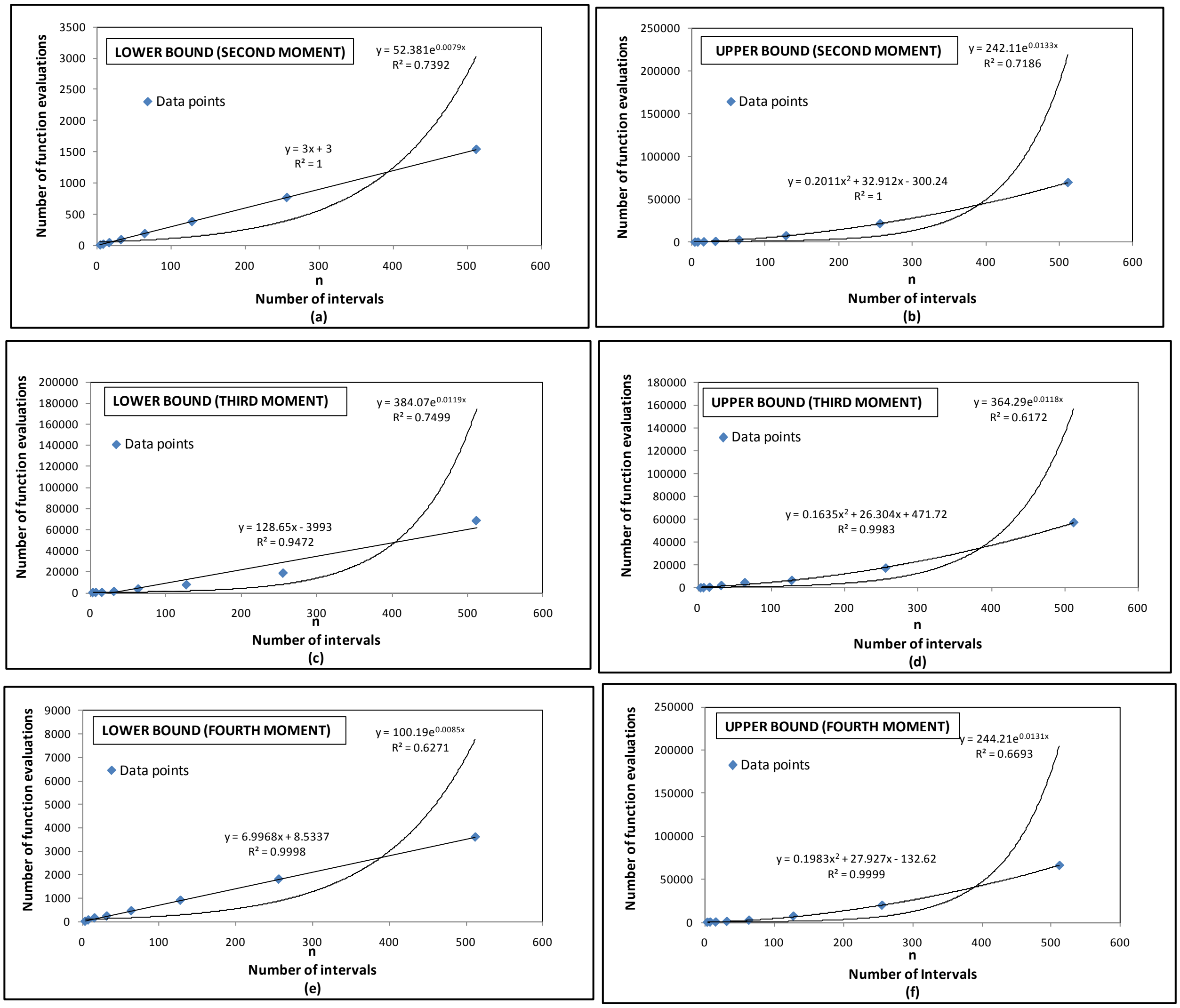


Figure 2: Computational effort for the estimation of bounds on second, third, and fourth moments for overlapping intervals

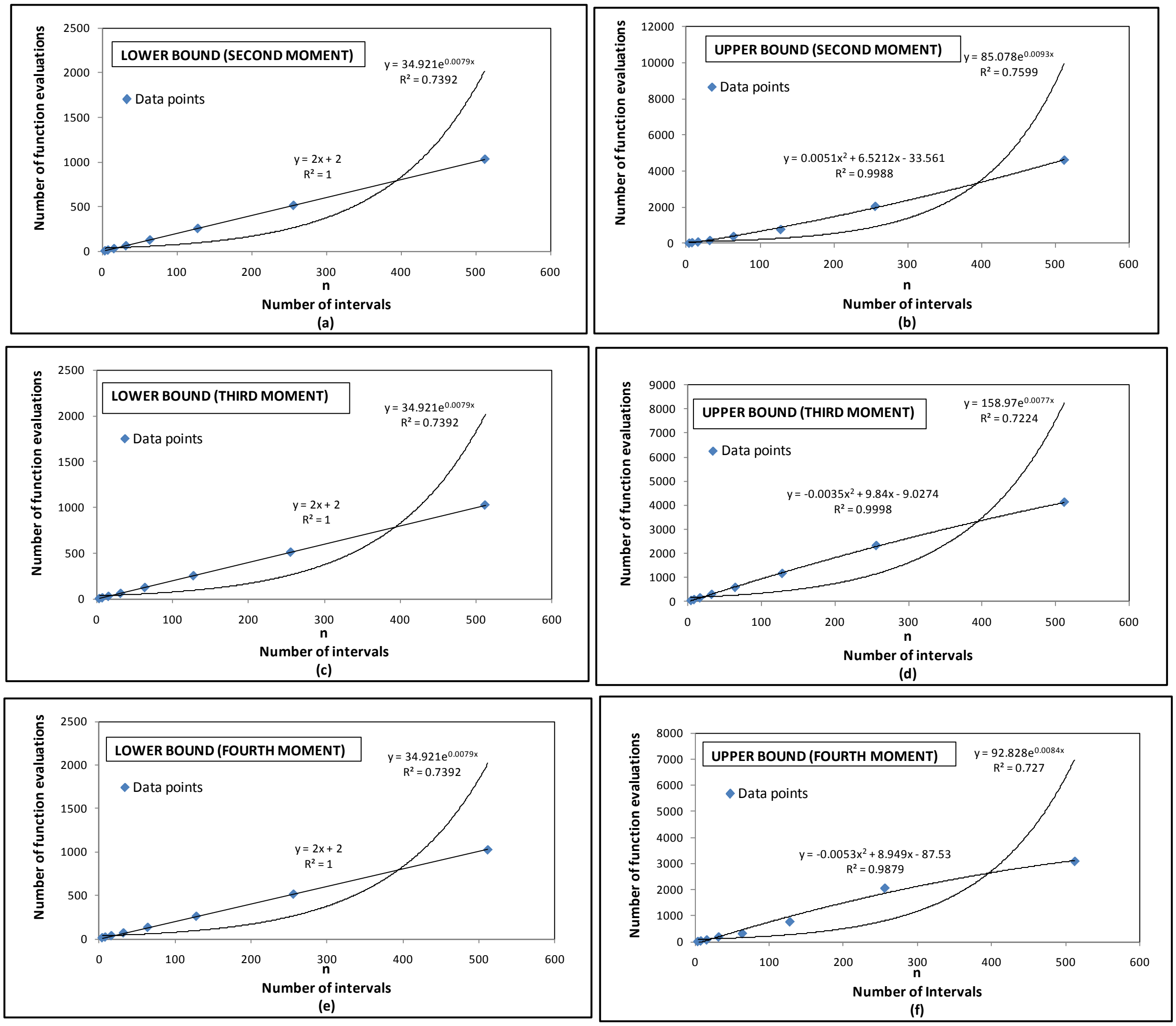


Figure 3: Computational effort for the estimation of bounds on second, third, and fourth moments for non-overlapping intervals

Observe that the computational effort for estimating the lower bound on second moment increases linearly with increasing number of intervals for both overlapping and non-overlapping data (subplots (a) in both Figures 2 and 3). The computational effort to estimate the upper bound on second moment with increasing number of intervals is observed to be $O(n^2)$, making this a computationally affordable procedure, even for relatively large data sets(subplots (b) in both Figures 2 and 3).

The computational effort is also found to scale polynomially with the number of intervals for both minimization and maximization of third and fourth moments, as seen from subplots (c)-(f) in both Figures 2 and 3. These plots show the best fitting polynomial and exponential trend lines to show that the trend is indeed polynomial in the number of intervals.

So far, we discussed the proposed optimization formulations to estimate bounds on the second, third, and fourth moments of interval data, which is the first important contribution of this chapter. The moment bounds estimated in this section can be used to fit a family of Johnson distributions to interval data, as discussed in the next section.

3. Fitting Johnson Distributions to Interval Data

As discussed in Chapter III, there are several approaches to fit Johnson distributions to point data using statistics such as moments or percentiles. Unlike for point data where there can be a single probability distribution as the uncertainty description (when a large amount of samples is available), multiple probability distributions could describe interval data. Once the bounds on the moments of the interval data are calculated using the approach outlined in the previous section, we can now fit the Johnson distributions whose moments fall within the bounds of the moments of the interval data.

Within the proposed framework, two procedures could be adopted for the uncertainty quantification of interval data: (1) *sampling-based*, which involves taking random samples of moments from within the bounds computed earlier, and fitting a Johnson distribution to each set of sampled moments, and (2) *optimization-based*, where a bounding envelope of the family of distributions can be constructed using an optimization approach using percentiles. The sampling based approach is discussed next.

3.1 Sampling-based procedure

The proposed sampling-based procedure for constructing the family of Johnson distributions is as follows:

1. Calculate the bounds on the first four moments of single or multiple interval data (Section 2).

2. Randomly select a set of moments from within the bounds of the first four moments. This sampling can be done using uniform distribution or by any discretization method. In this chapter, we use uniform distribution. We note here that the type of sampling or discretization method used might have impact on the end results. However, this issue is not investigated in this dissertation.

3. From Figure 1 (see Chapter III), infer the type of distribution to be fitted (e.g. bounded, unbounded, etc.) We only select those samples that suggest a bounded Johnson distribution fit, so that the resulting distribution lies within the bounds of the interval data specified, because, for interval uncertainty, it may be reasonable to argue that the true measurement has zero probability of lying outside the given interval for the single interval case or outside the overall bounds ([*Min(Lower bounds) Max(Upper bounds)*]) for the multiple interval case.

4. Using the bounds of the interval data, two parameters of the bounded Johnson distribution, ξ and λ, are estimated as $\xi = \min \{a_i, i = \{1,\ldots, n\}\}$, and $\lambda = \max \{b_i, i = \{1, \ldots, n\}\} - \min \{a_i, i = \{1 ,\ldots, n\}\}$. The parameters ξ and λ, which are the

location parameters (DeBrota et al, 1989), determine the lower end point and the range, respectively, of the bounded Johnson distribution.

5. The remaining two unknown parameters γ and δ, which govern the shape of the bounded Johnson distribution, are computed by solving the following optimization problem.

$$\min_{\gamma,\delta} \quad f(x) = \sum_{i=1}^{4} \left(M_{i(sampled)} - M_{i(johnson)} \right)^2 \tag{10}$$

$$s.t. \quad -50 \le \gamma \le 50 \tag{11}$$

$$0.2 \le \delta \tag{12}$$

where $M_{i(sampled)}$ is the set of moments sampled from step 2, and are the set of moments for a $M_{i(johnson)}$ Johnson distribution. Constraints on the Johnson parameters are imposed for numerical reasons (discussed later). Note that the objective function of the above optimization problem may require scaling since the moments can be of largely different magnitudes.

6. Repeat steps 2, 3, and 4 for a desired number of times. Each repetition of steps 3, 4, and 5 yields a single Johnson distribution.

The above sampling-based procedure can be repeated as many times as desired to obtain a family of Johnson distributions. The issue of sampling size can be problem dependent. The sampling-based procedure of uncertainty representation cannot guarantee rigorous bounds on input distributions, as it might underestimate the uncertainty due to practical limitations or computational expense. The sample size is a more critical issue when this uncertainty has to be propagated through some models of system response. In order to alleviate the issue of sampling size in uncertainty representation, we have proposed an optimization-based strategy to represent interval uncertainty.

Note that the above procedure is the same for both overlapping and non-overlapping intervals. The optimization-based procedure to generate a probabilistic representation for interval data is discussed next.

3.2 Optimization-based procedure: Johnson p-box

Theoretically, infinitely many distributions can be fit to the given interval data. It is of interest for practical reasons to compute bounding envelopes for the family of Johnson distributions, which we call the Johnson p-box. Note that the Johnson p-box is analogous to the empirical p-box (Figure 1 of Chapter II), which is the bounding envelope of empirical distributions to fit the interval data. In this subsection, we present an optimization formulation based on percentiles to construct the Johnson p-box.

In order to compute the bounding envelope, we solve a set of optimization problems, each for a different percentile value, α, where $0.01 \le \alpha \le 0.99$ Each optimization problem for a chosen α finds the parameters of the Johnson distribution that maximize or minimize the Johnson variable, x^{α}, such that the moments of the Johnson distribution fall within the bounds computed in Section 2. The following optimization formulation is used to compute the Johnson p-box. Note that the minimization yields the left most bound of the family of distributions for each α. Similarly, maximization of the optimization problem below yields the right most bound for each α.

$$\min_{\gamma,\delta}/\max \quad x^{\alpha} \tag{13}$$

$$s,t. \quad m1_{lb} \le m1_{johnson} \le m1_{ub} \tag{14}$$

$$m2_{lb} \le m2_{johnson} \le m2_{ub} \tag{15}$$

$$m3_{lb} \le m3_{johnson} \le m3_{ub} \tag{16}$$

$$m4_{lb} \le m1_{johnson} \le m4_{ub} \tag{17}$$

$$-50 \le \gamma \le 50 \tag{18}$$

$$0.2 \le \delta \tag{19}$$

where x^{α} is the α - th percentile point, $0.01 \le \alpha \le 0.99$, $m1_{\text{johnson}}, \ldots m4_{\text{johnson}}$ are the first four moments of the Johnson distribution with parameters ξ, λ, γ, and δ, respectively, which can be computed using simulation; $m1_{\text{lb}}$, …, $m4_{\text{lb}}$ respectively are the lower bounds on the first four moments of the interval computed using the proposed approach; and $m1_{\text{ub}}$, …, $m4_{\text{ub}}$ respectively are the upper bounds on the first four moments of the interval computed using the proposed approach.

The value of the objective function, x^{α}, can be found by applying the Johnson transformation (see Eq. 1 in Chapter III) to a standard normal variable corresponding to the given α. Constraints in Eq. (18) and (19) are imposed on the Johnson parameters for numerical reasons. The bounded Johnson transformation (DeBrota et al, 1989) is given as

$x = \xi + \lambda\left[1 + \exp\left(-\frac{z-\gamma}{\delta}\right)\right]^{-1}$, where x is the Johnson variable, and z is the standard normal variable. The δ parameter is restricted to be greater than 0.2: as $\delta \to 0$, the moments approach the impossible region for the Johnson family of distributions (Figure 1 in Chapter III) and can cause division by zero problems with the bounded Johnson transformation (Eqs. 1 and 2 in Chapter III). The bounds on γ have been chosen so that the bounded Johnson transformation function, $\left[1+\exp\left(-\frac{z-\gamma}{\delta}\right)\right]^{-1}$, has a finite non-zero value.

4. Numerical Examples

In this section, we apply the proposed approaches to five example problems. We consider four multiple interval examples, each with different numbers of intervals and overlaps, and one single interval example. Note that the examples used in this chapter may not cover all types of overlaps; however, the proposed methods work in more general situations. Comparisons with alternate representations, such as the empirical p-box, are also discussed.

4.1 Illustration of the proposed methodology

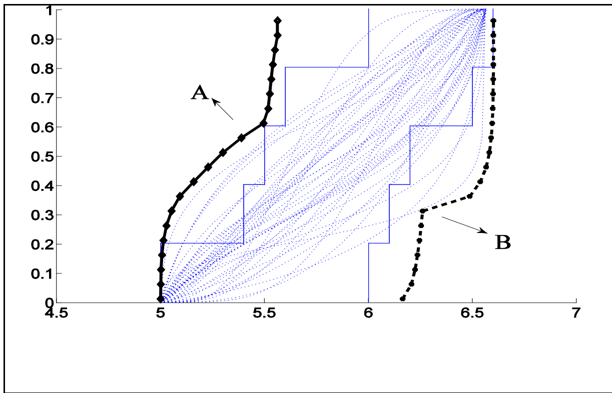
We consider two examples each for overlapping and non-overlapping multiple interval data, each with different numbers of intervals (Table 2). We follow the procedure outlined in Section 4.1 to fit a family of bounded Johnson distributions to each multiple interval data set in Table 2. The cumulative distribution functions of the family of Johnson distributions for each multiple interval data set are shown by thin dotted lines in Figure 4. The corresponding single interval results, where the moment bounds are computed using the methods outlined in Section 2.1, are shown in the left hand side plot in Figure 5.

Table 2: Interval data for the five numerical examples

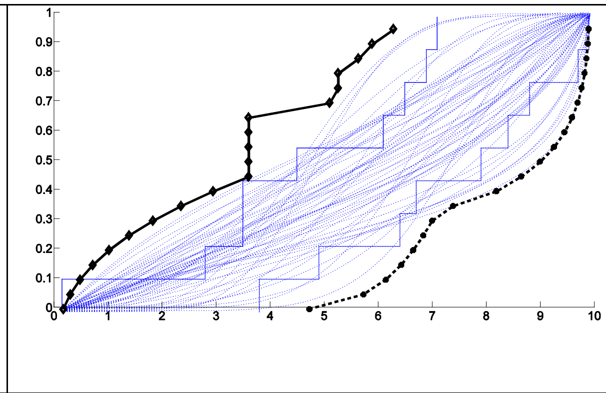
Example	Data
Example 1 with 5 overlapping intervals	[5, 6; 5.5, 6.1; 6, 6.5; 5.4, 6.2; 5.6, 6.6]
Example 2 with 9 overlapping intervals [Ferson et al 2007]	[3.5, 6.4; 6.9, 8.8; 6.1, 8.4; 2.8, 6.7; 3.5, 9.7; 6.5, 9.9; 0.15, 3.8; 4.5, 4.9; 7.1, 7.9]
Example 3 with 4 non-overlapping intervals	[5, 6; 6.1, 6.7;6.9, 7.8; 8, 9]
Example 4 with 6 non-overlapping intervals [Ferson et al 2007]	[1, 1.52; 2.68, 2.98; 7.52, 7.67; 7.73, 8.35; 9.44, 9.99;3.66, 4.58]
Example 5 with a single interval	[5, 15]

The Johnson p-box optimization problem is solved for each set of interval data in Table 2 using Matlab's *fmincon* solver. We use 20 equally spaced points for the percentile values, α, ranging between 0.01 and 0.99. Note that the selectetion of the number of percentile points is arbitray. However, solving the optimization problem at increased number of percentile points results in more accrutare bounds on uncertainty but with increased computational efforts. For each α, the minimization and maximization problems yield the left and right bounds on the p-box in Figure 4, respectively. At each α value, we repeated the maximization/minimization using 15 different starting points to avoid local optima; the best results among the 15 runs are reported.

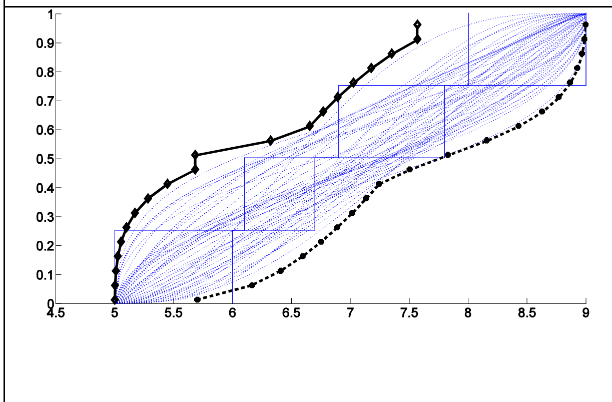
It is interesting to note that the Johnson p-boxes in Figure 4 for all the multiple interval examples shown have discontinuities. It is noted that the set of active constraints in the optimization (particularly, those with the moment bounds (Eqs. 14- 17)) changes at the point of discontinuity. For example, at the point A for Example 1 in Figure 4, the set of active constraints changes.



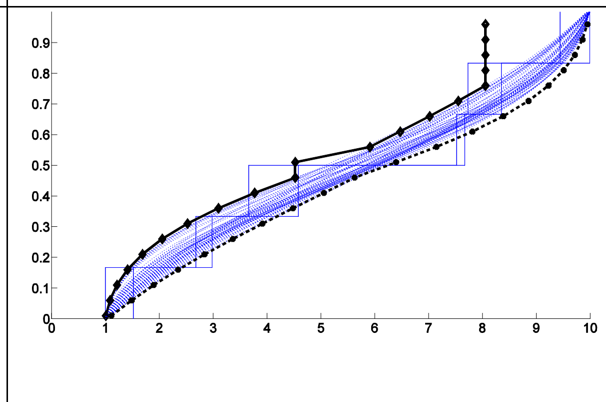
Example 1



Example 2

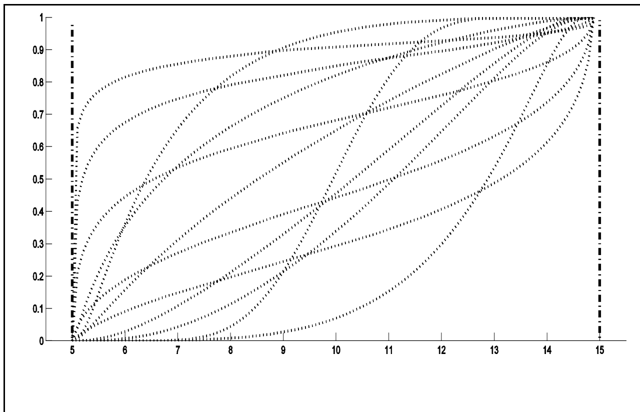


Example 3

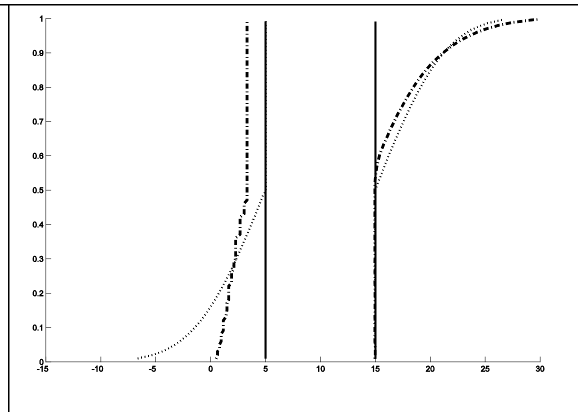


Example 4

Figure 4: Samples from the family of Johnson cumulative distributions for overlapping and non-overlapping examples for multiple interval examples (thick solid lines – Johnson p-box, thin solid lines – empirical p-box, dashed thin lines – family of Johnson CDFs)



Samples from the family of Johnson distributions (thin dotted lines) with Johnson p-box (thick dotted line)



Comparison of Johnson p-box (thick solid line) with normal (thin dotted line) and lognormal (thick dotted line) p-boxes

Figure 5: Single interval example

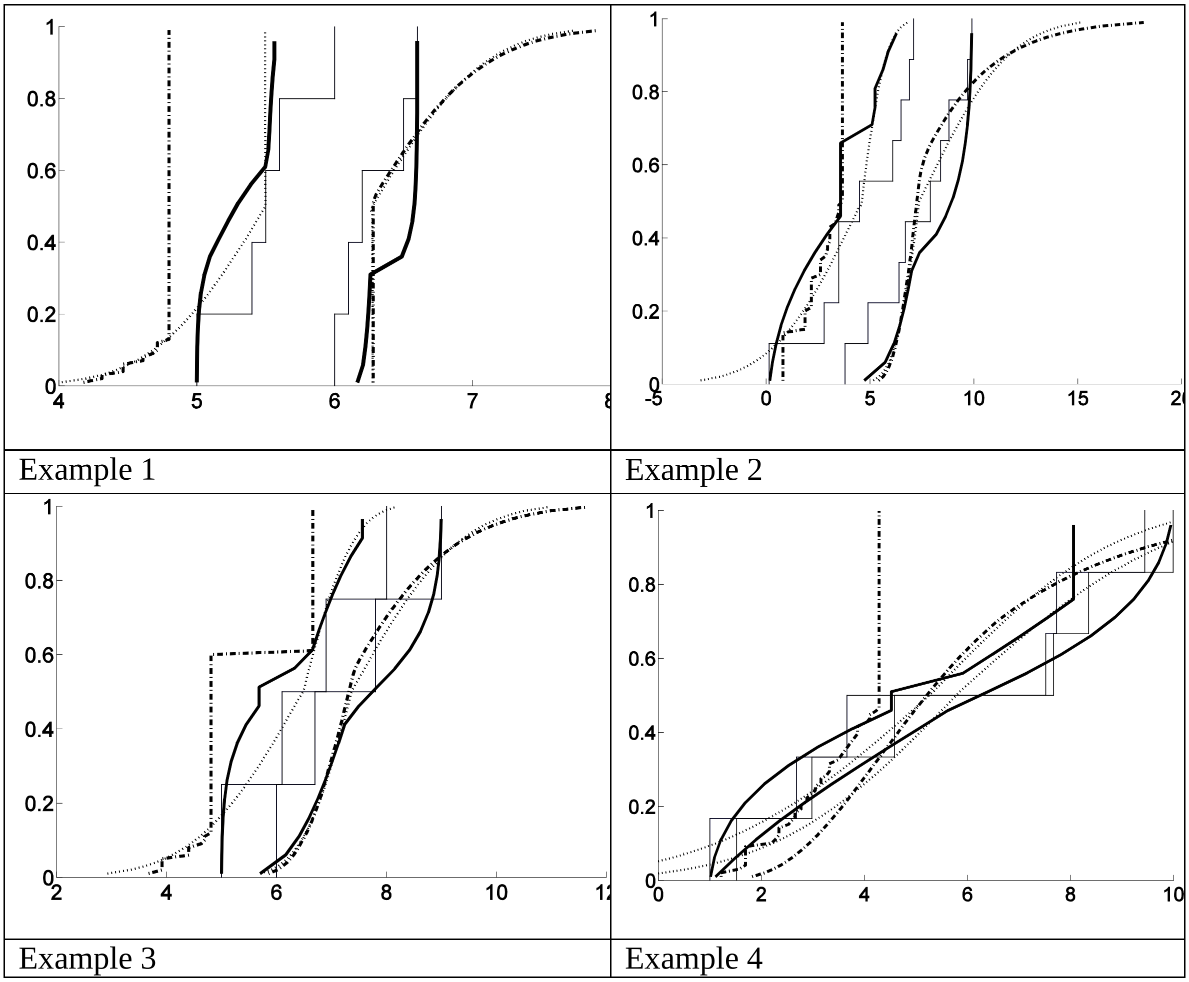


Figure 6: Comparison of empirical, bounded Johnson, normal, and lognormal p-boxes for multiple interval examples (thick dashed line – lognormal, thin dotted line – normal, thick solid line – bounded Johnson, thin solid line – empirical distribution)

Below the point A, the constraints on the upper bound of the third moment (upper bound in Eq. 16) and on the lower bound of the first moment (lower bound in Eq. 14) are active. Above the point A, the constraints on the lower bound of the fourth moment (lower bound in Eq. 17) and on the lower bound of the first moment (lower bound on Eq. 14) are active. Similar trend was observed at point B in Example 1, where a discontinuity occurs in the bounding envelope. Below the point B, the lower bound of the fourth moment (lower bound in Eq. 17) and the upper bound of the first moment (upper bound in Eq. 14) are the active constraints. Above the point B, the lower bound of the third moment (lower bound in Eq. 16) and the upper bound of the first moment (upper bound in Eq. 14) are the active constraints. For the single interval example, the Johnson p-box coincides with the left and right end points of the interval data.

4.2 Comparison with other representations

In this subsection, we present a comparison of the Johnson p-box with the empirical p-box idea available in the literature. We also compare how the choice of Johnson family of distributions impacts the probabilistic representation of interval data. Using an optimization formulation similar to that of the Johnson p-box, we compute the corresponding bounding envelopes for normal and lognormal distributions for single and multiple interval examples.

The empirical p-boxes for the multiple interval data cases, obtained by sorting the endpoints of the intervals, are also plotted in Figures 4 and 6 for comparison purposes (thin solid lines). Note that for all examples presented in this section, not all members of the Johnson family of distributions fall inside the empirical p-box. The moments of the family of Johnson distributions fall within the moment bounds computed earlier; however, the distributions do not necessarily fall within the empirical p-box.

In order to study the effect of the choice of Johnson family, we compare the Johnson p-box to the bounding envelopes obtained for normal and lognormal distributions. The following optimization formulation is used to find the bounding envelopes for normal and lognormal distributions, where constraints are imposed on the first two moments.

$$\min_{d}/\max \quad x^{\alpha} \tag{20}$$

$$s,t. \quad m1_{lb} \le m1_{dist} \le m1_{ub} \tag{21}$$

$$m2_{lb} \le m2_{dist} \le m2_{ub} \tag{22}$$

where x^{α} is the α - th percentile point, $0.01 \le \alpha \le 0.99$, d = (μ_Y, σ_Y) is the design variable vector, where Y is the normal random variable ; $m1_{dist}$ and $m2_{dist}$ are the first and the second moments for normal/lognormal distributions, respectively; $m1_{lb}$ and $m2_{lb}$ respectively are the lower bounds on the first two moments of the interval computed using the proposed approach; and $m1_{ub}$ and $m2_{ub}$ respectively are the upper bounds on the first two moments of the intervals computed using the proposed approach.

The quantities $m1_{dist}$ and $m2_{dist}$ for the normal p-box are related to the design variable vector, $m1_Y = \mu_Y$ and $m2_Y = \sigma_Y^2$. The moments of the lognormal variable (X), $m1_{dist}$ and $m2_{dist}$ are computed in terms of the corresponding normal variable moments, (μ_Y, σ_Y), as follows.

$$m1_X = e^{\mu Y + 0.5\sigma_Y^2} \tag{23}$$

$$m2_X = \mu_X^2\left(e^{\sigma_Y^2} - 1\right) \tag{24}$$

The maximization and minimization at each percentile point for the normal and lognormal have been repeated with 15 different starting points to avoid local optima. The best results from within the 15 starting points have been plotted in Figure 6. Note that the bounded Johnson p-box remains close to the empirical p-box for all the four multiple interval examples, which is not necessarily the case for normal and lognormal p-boxes. One possible reason for this behavior could be the theoretical bounds that exist on normal, lognormal, and bounded Johnson distributions. The normal distribution is unbounded, and can lie between [-∞, +∞], whereas the lognormal distribution is bounded between [0, +∞]. The bounded Johnson distribution is restricted to lie within the interval bounds (discussed in Step 4 of Section 3.2).

As shown from the examples above, the proposed probabilistic representation of interval data using a family of bounded Johnson distributions is a viable approach for uncertainty quantification for interval uncertainty. Once such a family of distributions is constructed, it could be used in the context of uncertainty/reliability analysis using Monte Carlo simulations or FORM/SORM, resulting in set of values for an output quantity. This notion is unlike the case with aleatory uncertainties, where usually a single probabilistic representation describes the uncertainty, which yields a single quantity of interest from the uncertainty propagation stage. The proposed uncertainty representation is particularly suitable for use in FORM/SORM, since these methods require that the random variables are represented by probability distributions. These methods also require transforming the random variables into standard normal space, which is easy with Johnson distributions.

The state-of-the-art in uncertainty propagation in the presence of interval data requires a nested analysis – instances of interval variable are considered in an outer loop, each iteration of which requires a probabilistic analysis for the aleatory uncertainties – inner loop. Instead, one could use an *optimization-based* uncertainty propagation approach, where the parameters of the input interval variables (probabilistically described) that either maximize or minimize an output quantity of interest, e.g., probability of failure, can be found. We have proposed such optimization-based approaches for cases where the input variables are described by sparse point data (McDonald et al, 2009). Similar ideas can be extended to variables described by intervals, which will be studied in the future.

5. Conclusion

In this chapter, we propose a probabilistic framework for representing uncertainty information available through interval data. The main contributions of this chapter are: (1) development of algorithms to estimate bounds on the second, third, and fourth moments of single and multiple interval data, (2) demonstration that the proposed moment bounding algorithms are scalable in polynomial time, (3) use of the moment bounds thus estimated to fit a family of flexible Johnson distributions, (4) definition of a Johnson p-box, which is the bounding envelope of the family of Johnson distributions, and (5) development of an optimization-based method to construct the Johnson p-box.

Through scalability testing, we have shown that the algorithms to compute bounds on the second, third and fourth moment of interval data scale polynomially in the number of intervals. This is important because these problems have been generally considered earlier to be NP-hard. We have also shown how a probabilistic description for interval data can be provided by a *family of distributions.* Due to the nature of the interval data, however, we make no assumptions about the relative likelihood of any of these CDFs to be the true CDF. For point data, statistics such as moments or percentiles which are used to fit probability distributions assume single values. However, for interval data, we can only estimate bounds on the statistics such as moments or percentiles. Therefore, unlike for point data where there can be a single probability distribution as the uncertainty description, multiple probability distributions should describe interval data.

This chapter presented an approach that can be used to fit a family of Johnson distributions using moment bounds obtained as discussed above. The family of Johnson distributions thus fit can be used as the probabilistic representation of the interval data. This process could also be performed using several other distributions. Johnson distributions offer an advantage because they have convenient transformations to be mapped into the normal space, which facilitates the use of popular analytical reliability methods such as FORM and SORM.

The proposed probabilistic framework of handling interval data can be applied for a combined treatment of aleatory and epistemic input uncertainties from the perspective of uncertainty propagation or reliability based design. This approach to uncertainty representation given interval data can allow for computationally efficient propagation by avoiding the nested analysis that is typically performed in the presence of interval variables.

CHAPTER V

PROBABILISTIC FRAMEWORK FOR UNCERTAINTY PROPAGATION WITH BOTH PROBABILISTIC AND INTERVAL VARIABLES

1. Introduction

This chapter develops and illustrates a probabilistic approach for uncertainty representation and propagation in system analysis, when the information on the uncertain input variables and/or their distribution parameters may be available as either probability distributions or simply intervals (single or multiple). A unique aggregation technique is used to combine multiple interval data and to compute rigorous bounds on the system response CDF. The uncertainty described by interval data is represented through a flexible family of probability distributions. Conversion of interval data to a probabilistic format enables the use of computationally efficient methods for probabilistic uncertainty propagation. Two methods are explored for the implementation of the proposed approach, based on: (1) sampling and (2) optimization. The sampling based strategy is more expensive and tends to underestimate the output bounds. The optimization based methodology improves both aspects. The proposed methods are used to develop new solutions to challenge problems posed by the Sandia Epistemic Uncertainty Workshop (Oberkampf et al, 2004). Results for the challenge problems are compared with earlier solutions.

As mentioned earlier in Chapter I, if the uncertainty described by intervals can be converted to a probabilistic format, the computational expense of interval analysis is avoidable as it allows for treatment of aleatory and epistemic uncertainty together without nesting, and already well established probabilistic methods of uncertainty propagation can be used. This chapter develops and illustrates a new approach for the representation and propagation of uncertainty available in both probabilistic and non-probabilistic formats. The proposed representation avoids the expensive nested analysis by enabling the use of an optimization-based strategy that can estimate the distribution parameters of the input variables that minimize or maximize an output quantity of interest, e.g., probability of failure or expectation of system response. Note that this optimization is done not to change the design but rather to determine the endpoints of intervals that bound the *output* estimates. The system design is considered static. A new aggregation technique is used to combine multiple intervals and to compute rigorous bounds on the system response CDF. This aggregation technique enables the use of the method of matching moments to represent the uncertainty described by the multiple intervals through a family of probability distributions (see Chapter IV).

The rest of the chapter is organized as follows. Section 3 describes the proposed methodologies for representation and propagation of epistemic and aleatory uncertainty. Section 3 describes numerical examples, specifically Sandia Challenge Problems (Oberkampf et al, 2004) and solves them using the two proposed approaches: (1) sampling-based, and (2) optimization-based. Solutions from the proposed methods are compared to those obtained with earlier methods. Section 4 provides concluding remarks and suggestions for future work.

2. Uncertainty Propagation using Probabilistic Analysis

Chapter IV proposed a methodology for representation of interval uncertainty using a flexible family of Johnson distributions. In this chapter, we have used Monte Carlo simulation (MCS) to achieve the propagation of both probabilistic and interval uncertainty through system models. Our purpose here is to develop a unified probabilistic framework that can represent and propagate both aleatory and epistemic uncertainty, no matter what uncertainty propagation method is used. However, we note here that analytical approximation methods (e.g., FORM, SORM) and efficient sampling methods (e.g., importance sampling) can also be used within the proposed uncertainty propagation framework.

Problems involving interval uncertainty can be divided into two cases: *1)* Input variable is described by a single interval or multiple intervals; or *2)* the distribution parameter of the input variable is described by a single interval or multiple intervals. In the following subsections, we propose sampling and optimization-based approaches for propagation of aleatory and epistemic uncertainty for each of the cases.

2.1 Sampling-based methodology for uncertainty propagation

2.1.1 Case 1: Input variable described by interval data

In this case, the uncertainty is modeled probabilistically by fitting a Johnson distribution, using values of the moments sampled from within the moment bounds of the interval data. The following computational procedure can be used to implement uncertainty quantification by sampling:

1. Generate a family of CDFs for each of the input variables described by single or multiple interval data by the procedure described in section 2.
2. Propagate each of the CDFs from the input family of CDFs through the system response equation by any probabilistic uncertainty propagation method (e.g, FORM, SORM or MCS).
3. Construct the CDF of the system response given a realization of the distribution parameters from the family of CDFs by repeating step 2 for a range of threshold values and thus obtain a family of CDFs for system response.

2.1.2 Case 2: Input variable <u>distribution parameters</u> described by interval data

As in Case 1, the uncertainty is again modeled probabilistically by fitting a family of Johnson distributions, using values of the moments sampled from within the moment bounds of the interval data. The following computational procedure can be used to implement uncertainty quantification by sampling:

1. Generate a family of *n* CDFs for each of the distribution parameters of the input variables described by a single or multiple interval data by the procedure described in section 2.
2. Take *m* samples from each of the *n* CDFs of distribution parameters of an input variable. Now the input variable of interest has a family of $m \cdot n$ CDFs. Note that *n* is the number of CDFs for each distribution parameter. We can sample as many sets of distribution parameter values as we want from each of the *n* CDFs. Each set of the sampled distribution parameter values now gives a single CDF for the input random variable of interest.

Therefore, by sampling *m* sets of distribution parameter values from each of the *n* CDFs of the distribution parameters, we have a total of $m \cdot n$ number of CDFs for the input random variable of interest. . Now if we generate *p* samples from each of the $m \times n$ CDFs, the overall sample size will be $m \times n \times p$.

3. Propagate each of the CDFs from the family of CDFs of each of the input variables through the system response equation by any probabilistic uncertainty propagation method (e.g, FORM or MCS).
4. Construct the CDF of the system response given a realization of the distribution parameters from the family of CDFs by repeating step 3 for a range of threshold values and thus obtain a family of CDFs for system response.

2.2 Optimization-based methods for uncertainty propagation

The above methodology to convert interval uncertainty into a probabilistic format is based on a sampling strategy. A sampling strategy might underestimate the output bounds since the sampling is not exhaustive due to practical limitations or computational expense. Therefore, in this subsection, we develop an optimization-based strategy to convert uncertainty described by interval data into a probabilistic framework. The optimization approach is also much less expensive compared to the sampling-based approach. We propose two types of optimization – percentile-based and expectation-based.

2.2.1 Case 1: Input variable described by interval data

Percentile-Based Optimization (PBO)

This method minimizes and maximizes the system output $g_\alpha(x|m)$ conditioned on a set of moments *(m_i)* for the input variables at different percentile values *(α)* of the output CDF and thus obtains bounds on the system output CDF. Its implementation is as follows:

1. Calculate the bounds on the first four moments of single or multiple interval data by the methods described in Chapter IV.
2. Solve the following optimization problems at different percentile values *(α)* to obtain bounds on output CDF. Minimizing the objective function gives the lower bound on the output and maximizing the objective function gives the upper bound on the output.

$$\begin{aligned}
&\min_m/\max \, g_\alpha(x|m) \\
&s.t. \, m_i \geq a_i \\
&\quad m_i \leq b_i \qquad i = 1,2.....,4 \\
&\quad \beta_2 - \beta_1 - 1 \geq 0 \\
&\quad \beta_2 - 2\beta_1 - 3 \leq 0 \\
&\quad \text{where } \beta_1 = m_3^2 / m_2^3 \\
&\qquad\qquad \beta_2 = m_4 / m_2^2
\end{aligned} \tag{1}$$

Here, the decision variable set *m* is the set of moments (m= [m_1 m_2 m_3 m_4]). The last two nonlinear constraints ensure that the optimizer only selects those values of moments that suggest a bounded Johnson distribution fit (See Figure 1 of Chapter III), so that the resulting distribution lies within the bounds of the interval data specified. It is noted here that the objective function in this optimization problem is conditioned on a set of moments for each of the input variables and estimates the parameters of Johnson distribution from the set of moments in each iteration by the method described in Chapter IV.

Expectation-Based Optimization (EBO)

The optimization formulation as described above is rigorous, but expensive as it requires solving the problem repeatedly at different α-levels. Therefore, in the following discussion, we propose an expectation-based optimization strategy to obtain approximate bounds on system response CDFs which is computationally less expensive. This formulation is based on the assumption that the sets of distribution parameters of input variables which result in minimum or maximum expectation of the system response *(E(g(x)))*, can also give an upper bound on the entire CDF of the system response *(g(x))* for the minimization problem and a lower bound for the maximization problem, respectively. A proof in support of this statement is given below:

Theorem: CDF lower bound obtained by EBO will be no less than that obtained by PBO. The variables and the constraints for the PBO and EBO optimization problems are identical.

Proof: The most general problem of calculating lower bounds on the system response can be stated as follows:

EBO	PBO
$\underset{\theta}{Min} \quad E(g(x))$	$\underset{\theta}{Min} \quad g^{\alpha}(x)$
$s.t. \quad \theta \in \Theta$	$s.t. \quad \theta \in \Theta$

where θ is the set of distribution parameters, selected from a set of admissible values $\boldsymbol{\Theta}$.

There are two possibilities:

i) $\theta_E^* = \theta_P^*$: EBO curve is coincident with PBO curve, which implies $g^\alpha\left(x\middle|\theta_P^*\right) = g^\alpha\left(x\middle|\theta_E^*\right)$.

ii) $\theta_E^* \neq \theta_P^* : \theta_E^*$ is not the optimal solution. Note that in both EBO and PBO, the feasible sets are identical, however, θ_E^* minimizes the expectation of the system response ($E(g(x))$), whereas θ_P^* minimizes the quantile value of the system response ($g^\alpha(x)$). This implies that there is a set of parameters $\theta_P^* \in \Theta$ for which $g^\alpha\left(x\middle|\theta_P^*\right) \leq g^\alpha\left(x\middle|\theta_E^*\right)$. This is illustrated in Figure 1. The left two curves are obtained by solving the minimization problems and the right two curves are obtained by solving the maximization problems. For the minimization problems, it is seen from the figure that at fixed α-level the EBO solution gives a higher value of system response ($g^\alpha(x)\middle|\theta_E^*$) than that of the PBO solution ($g^\alpha(x)\middle|\theta_P^*$).

This proves the theorem that CDF lower bound obtained by EBO will be no less than that obtained by PBO.

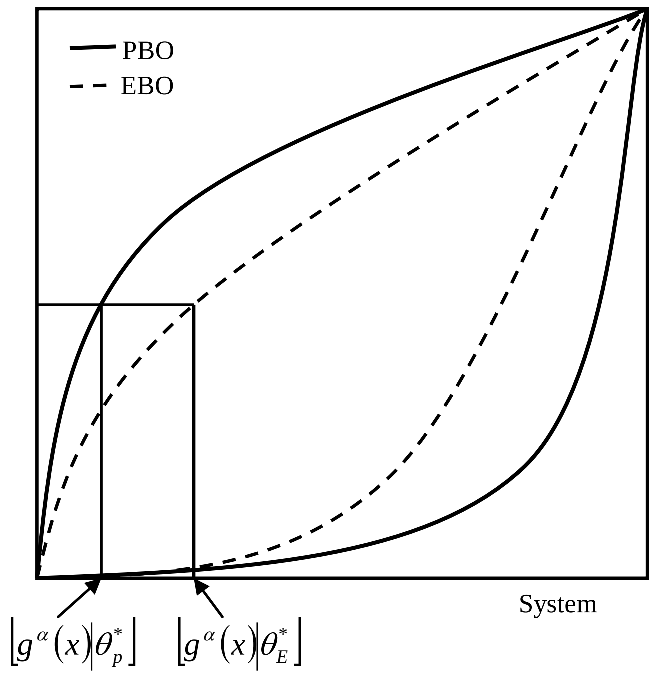


Figure 1: PBO and EBO bounds

Therefore, EBO gives an upper bound (to the right of PBO lower bound) of output uncertainty for the minimization problem. Similarly, it can be proved that EBO gives a lower bound (to the left of PBO upper bound) of output uncertainty for the maximization problem.

EBO has the same formulation as in Eq. (1) but with a different objective function $\min/\max_{m} E(g(x|m))$. All the constraints remain the same. In this case, the optimization formulation yields sets of moments each corresponding to a set of Johnson distribution parameters. Once the distribution parameters are obtained, any probabilistic uncertainty propagation method (e.g., FORM or MCS) can be used to construct approximate bounds on the CDF of the system response. Figure 2 illustrates the two optimization methods for Case 1.

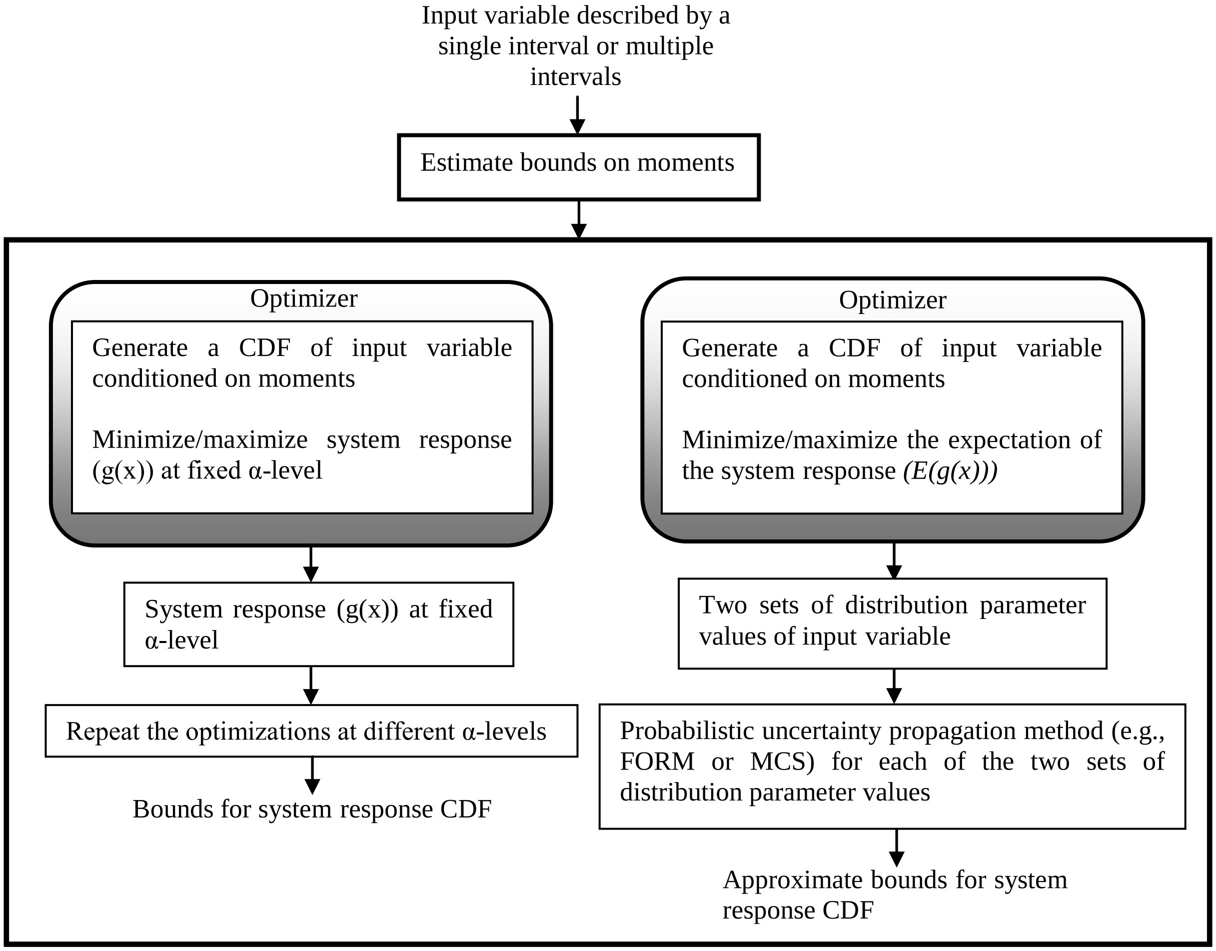


Figure 2: Optimization methods for output uncertainty quantification (Case 1)

2.2.2 Case 2: Input variable distribution parameters described by interval data

In this case, the implementation is more involved than in Case 1 where the input variable itself was described by interval data. In Case 1, input uncertainty was represented by a *family of distributions for the input variable*. In Case 2, we have *a family of distributions for each distribution parameter of the input variable*.

<u>Percentile-Based Optimization (PBO)</u>

The proposed PBO formulation involves two nested uncertainty analysis procedures. The inner loop uncertainty analysis calculates the conditional CDF of the system response given a value of the distribution parameter. The outer loop uncertainty analysis calculates the distribution parameters which minimize or maximize the conditional CDF of the system response.

In the inner loop, the cumulative distribution functions for the basic random variables with uncertain probability distributions are calculated by conditioning on a particular realization of the uncertain distribution parameters. Their optimum values are chosen to minimize or to maximize the system response $g_\alpha\left(x\middle|\theta^D\right)$ conditioned on a realization of distribution parameters for the input variables at different percentile values *(α)*.

In the outer loop, the cumulative distribution functions for the basic random variables with uncertain probability distributions are calculated by conditioning on a particular distribution of distributions of the uncertain distribution parameters. Their optimum values are chosen to minimize or to maximize the system response $g_\alpha\left(x\middle|m\right)$ conditioned on set of moments *(m_i)* for each of the input variables and/or distribution parameters at different percentile values *(α)*. The lower bound of the α-percentile value of the output *g(x)* is obtained by solving:

$$
\begin{aligned}
&\min_m\left(\min_{\theta^D}(g_\alpha(x|\theta^D)|\theta^D \in \theta^B(m))\right)\\
&s.t. m_i \geq a_i\\
&\quad m_i \leq b_i \qquad i = 1,2.....,4 \qquad\qquad (2)\\
&\quad \beta_2 - \beta_1 - 1 \geq 0\\
&\quad \beta_2 - 2\beta_1 - 3 \leq 0\\
&\quad \text{where } \beta_1 = m_3^2 / m_2^3\\
&\qquad\qquad \beta_2 = m_4 / m_2^2
\end{aligned}
$$

The upper bound of the α-percentile value of the output *g(x)* is obtained by solving:

$$
\begin{aligned}
&\max_{m}\left(\max_{\theta^D}(g_\alpha(x|\theta^D)|\theta^D \in \theta^B(m))\right)\\
&s.t.\, m_i \geq a_i\\
&\quad m_i \leq b_i \qquad i = 1,2.....,4 \qquad (3)\\
&\quad \beta_2 - \beta_1 - 1 \geq 0\\
&\quad \beta_2 - 2\beta_1 - 3 \leq 0\\
&\quad \text{where } \beta_1 = m_3^2 / m_2^3\\
&\qquad\quad \beta_2 = m_4 / m_2^2
\end{aligned}
$$

In the above optimization formulations, θ^D corresponds to the realization of distribution parameters and θ^B corresponds to the hyper parameters i.e., distribution parameters of distribution parameters of the input variable. The constraints are the same as in Eq. (1).

The most general way of solving this optimization formulation can be outlined as follows:

1. Calculate the bounds on the first four moments of single or multiple interval data by the methods described in Chapter IV.
2. The outer loop optimizer passes a single CDF for the distribution parameter to the inner loop optimization problem. This single CDF is sampled for realizations of the distribution parameters. The inner loop optimizer solves for the particular realization of the distribution parameter which leads to the minimum or maximum system response $g_\alpha(x|\theta^D)$ at different percentile values *(α)*. Therefore, solving these nested formulations of Eqs. (2) and (3) yields the realization of distribution parameters which minimizes or maximizes the system response $g_\alpha(x|\theta^D)$, respectively at different percentile values *(α)* to obtain bounds on output CDF.

Expectation-Based Optimization (EBO)

As in Case 1, we also propose an expectation-based strategy for Case 2 to obtain approximate bounds on system response CDF based on the same assumption that was made for Case 1.

The formulations are the same as in Eqs. (2) and (3) but with a different objective functions $\min_{m}\left(\min_{\theta^D}\left(E\left(g(x|\theta^D)|\theta^D \in \theta^B(m)\right)\right)\right)$ and $\max_{m}\left(\max_{\theta^D}\left(E\left(g(x|\theta^D)|\theta^D \in \theta^B(m)\right)\right)\right)$, respectively. All the constraints remain the same. Solving these nested formulations yields the realization of distribution parameters that minimize or maximize the expectation of the system response. Once the realizations of the distribution parameters are obtained, any probabilistic uncertainty propagation method (e.g., FORM, SORM or MCS) can be used to construct approximate bounds on the CDF of the system response. Figure 3 illustrates both the optimization methods for Case 2.

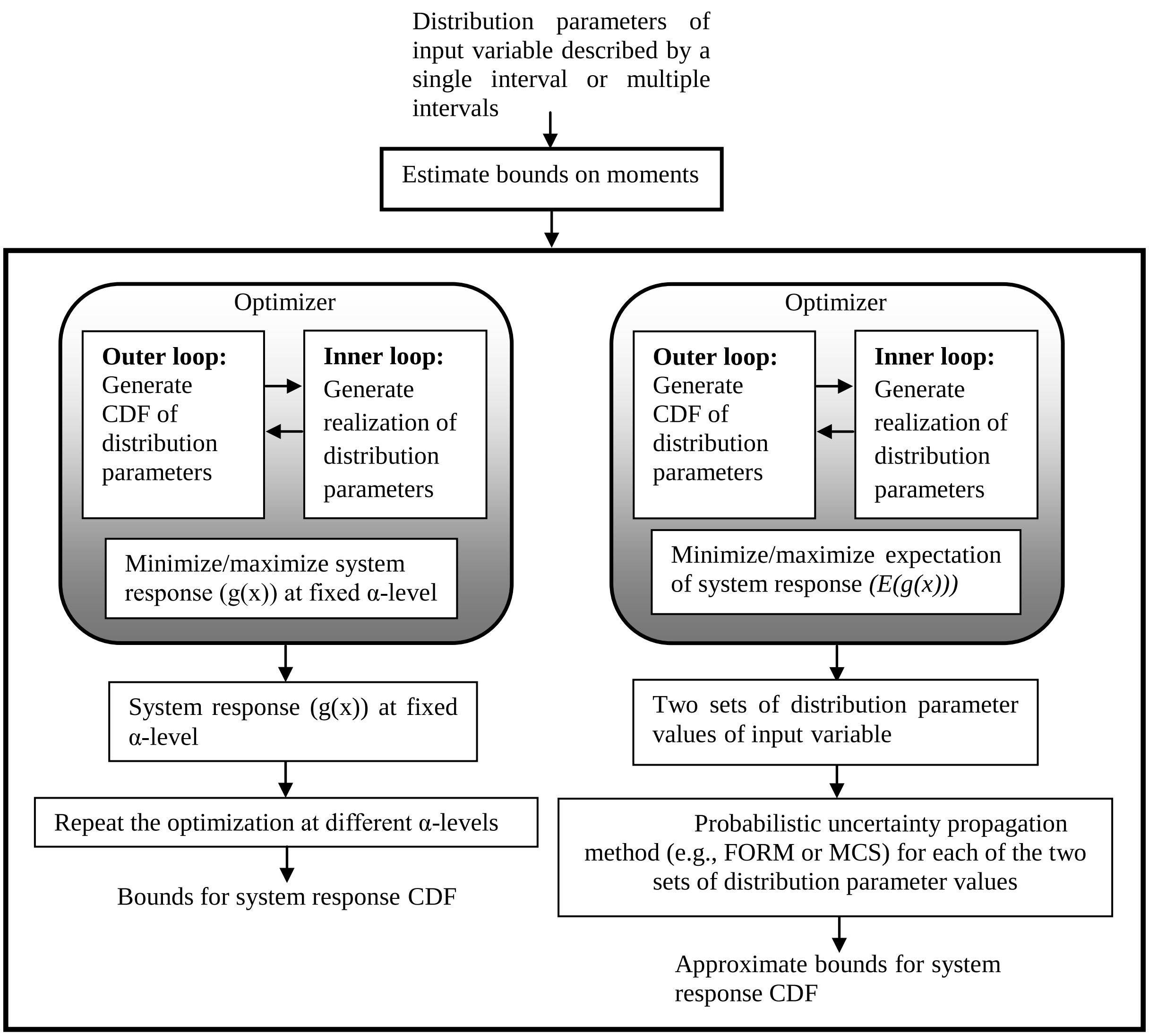


Figure 3: Optimization methods for output uncertainty quantification (Case 2)

Note that the PBO and EBO methods are developed to propagate uncertainty described by only interval data. However, in practice, a mixture of both sparse point and interval data could be available for the same variable, or some variables might be described by sparse point data and the others might be described by interval data. If both sparse point and interval data are available for the same variable, the optimization-based moment bounding algorithms developed in Chapter IV can still be used to calculate moment bounds for the mixed data. In the moment bounding algorithms, each decision variable corresponds to an interval. When both sparse point and interval data are available for the same variable, the number of decision variables in the optimization problems is still equal to the number of intervals only, and the sparse point data are used as fixed quantities in calculating moments in the objective functions of the optimization problems. Once the bounds on the moments for the mixed data are obtained, the uncertainty propagation can be achieved by using both PBO and EBO.

When some variables are described by sparse point data and others are described by interval data, it is necessary to estimate the confidence bounds on the first four moments for the variable described by sparse point data. Efficient methods are available to estimate the confidence bounds on mean values and variances in the presence of limited data (see Chapter VII). It is also possible to estimate bounds on the third and fourth moments for sparse point data using bootstrap methods. The two types of bounds are treated in the same manner. Once the confidence bounds on the moments for sparse point data and the bounds on moments for interval data are obtained, the uncertainty propagation can be achieved by using both PBO and EBO.

3. Numerical Examples

In this section, the proposed methods for propagation of aleatory and epistemic uncertainty are illustrated with Challenge Problems from the Sandia Epistemic Uncertainty Workshop (Oberkampf et al, 2004). Section 4.1 briefly describes the two problem sets. Section 4.2 presents the solutions for the challenge problems using both sampling and optimization-based approaches.

3.1 Challenge Problems

The two problem sets involve (A) a simple algebraic function, and (B) solution to a linear ordinary differential equation (ODE).

3.1.1 Challenge Problem A: Algebraic problem set

Consider the algebraic function

$$y = (a + b)^a \tag{4}$$

where, *y* is the output. The input variables *a* and *b* are assumed to be independent of each other and both *a* and *b* are positive real numbers. The task for each problem in the set is to quantify the uncertainty in *y* given the information concerning *a* and *b*. It is assumed that there is no uncertainty about the model form. Only uncertainty in the model input variables is considered.

Six problems are specified in sequence. The sequence is structured by the type and quantity of information specified for *a* and *b*. The structure of the sequence is given here:

Problem 1: *a* and *b* are both uncertain and must lie within given single intervals.

Problem 2: *a* is uncertain and must lie within a single interval, and *b* is characterized by multiple intervals.

Problem 3: Both *a* and *b* are characterized by multiple intervals.

Problem 4: *a* is uncertain and must lie within a single interval, and the uncertainty in *b* is specified by a probability distribution with imprecise parameters.

Problem 5: *a* is characterized by multiple intervals, and the uncertainty in *b* is specified by a probability distribution with imprecise parameters.

Problem 6: *a* is uncertain and must lie within a single interval, and the uncertainty in *b* is described by a precise probability distribution.

Problems 2, 3 and 5 each are further divided into three sub-problems based on the nature of the multiple intervals. The types of the multiple intervals are classified as *i)* consonant collection of intervals (intervals are nested), *ii)* consistent collection of intervals (no overlaps among the intervals), and *iii)* arbitrary collection of intervals (no assumption about the overlap or relationship among the intervals). The complete description of each problem set and the numerical data can be found in Oberkampf et al (2004). In our approach, we use an optimization technique to obtain the bounds on moments of interval data, and our method does not depend on the type of intervals. Since our proposed methods can handle all three classes of interval data in the same manner, we only choose to solve the first sub-problem for each of Problems 2, 3 and 5.

3.1.2 Challenge Problem B: ODE Problem

The ODE problem is described by a spring mass - damper system acted on by a forcing function $Y \cos \omega t$ as shown in Figure 4. The displacement and the velocity of the mass relative to a fixed reference frame are given by x and $\dot{x}$, respectively.

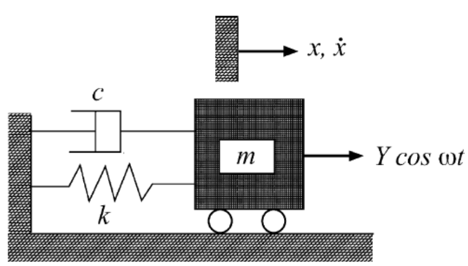


Figure 4: Mass-spring-damper system acted on by an excitation function (Oberkampf et al, 2004)

The equation of motion for the mass is given in Eq. (5)

$$m\ddot{x} + c\dot{x} + kx = Y\cos\omega t \quad (5)$$

The analytical expression for the steady-state magnification factor can be obtained as

$$D_s = \frac{k}{\sqrt{\left(k - m\omega^2\right)^2 + \left(c\omega\right)^2}} \quad (6)$$

The task for this problem is to quantify the uncertainty in D_s given the information concerning *k*, *m* and *ω*.

In the prescribed problem set, parameter *m* is given by a triangular probability distribution defined on the interval $[m_{\min}, m_{\max}]$ with a mode of m_{mod}. The values of m_{min}, m_{max}, and m_{mod} are precisely known.

Parameter *k* is described by a triangular distribution with imprecise k_{min}, k_{max}, and k_{mod}. The values of k_{min}, k_{max}, and k_{mod} are described by multiple intervals.

Parameter *c* is described by multiple intervals, and *ω* is given by a triangular probability distribution defined on the interval $[\omega_{\min}, \omega_{\max}]$ with a mode of ω_{mod}. The values of ω_{min}, ω_{max}, and ω_{mod} are described by single intervals.

The complete description of the problem and the numerical data can be found in Oberkampf et al (2004).

3.2 Numerical Results

Bounds on the CDF of system response for each of the problems are constructed using the optimization methods described in Section 3.2. A family of CDFs of system response is also constructed using the sampling strategy described in section 3.1. In the sampling approach, we have used 100,000 samples of system response for each of 10 sets of distribution parameters for each problem under Case 1 and 100,000 samples for each of 100 sets of distribution parameters for each problem under Case 2. The optimization-based strategy used 1000 samples of system response for each problem under Case 1 and 1000 samples for each of 100 sets of distribution parameters for each problem under Case 2.

3.2.1 Challenge Problem A

Problem A-1

For this problem, both input variables *a* and *b* are described by single intervals [0.1, 1.0] and [0.0, 1.0], respectively. We follow the procedure outlined in Chapter IV to fit a family of bounded Johnson distributions to each single interval data set. As an example, samples of cumulative density functions for the family of Johnson distributions for input variable *a* are shown in Figure 5.

This problem belongs to Case 1 as described in Section 2 and is solved by both optimization and sampling-based strategies and the results are shown in Figure 6.

This particular problem can also be solved by a simple deterministic optimization approach as shown below:

$$\begin{aligned} &\min/\max \;\; y = (a+b)^{a} \\ &s.t. \quad lb \le a \le ub \\ &\qquad\;\; lb \le b \le ub \end{aligned} \tag{6}$$

This optimization formulation yields the bounds on the system response as [0.6922, 2] which is exactly the same as the lowermost and uppermost bounds obtained by the proposed probabilistic approach, corresponding to CDF values of 0 and 1.

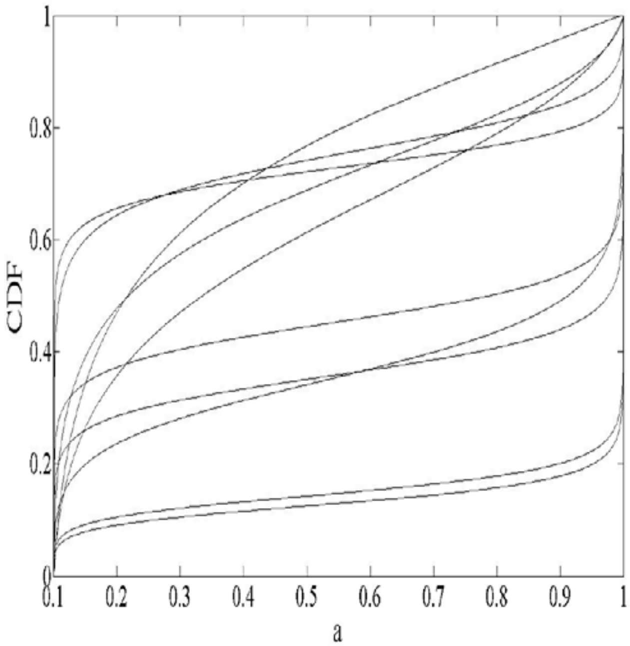


Figure 5: Family of Johnson distributions for input variable *a* for Problem A-1

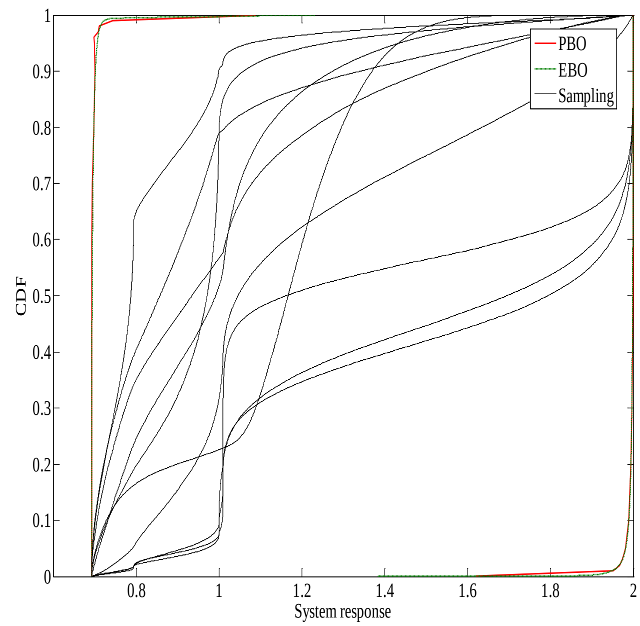


Figure 6: Bounds on CDF of system response for Problem A-1

Problem A-2

For this problem, the input variable *a* is described by a single interval [0.1, 1] and has the same uncertainty representation as shown in Figure 5. Input variable *b* is described by multiple interval data ([0.6, 0.8], [0.4, 0.85], [0.2, 0.9], [0.0, 1.0]) and we follow the procedure outlined in Chapter IV to fit a family of bounded Johnson distributions to the multiple interval data set of input variable *b*. Several sample cumulative density functions from the family of Johnson distributions for input variable *b* are shown in Figure 7.

This problem belongs to Case 1 as described in Section 2 and is solved by both optimization and sampling-based strategies and the results are shown in Figure 8.

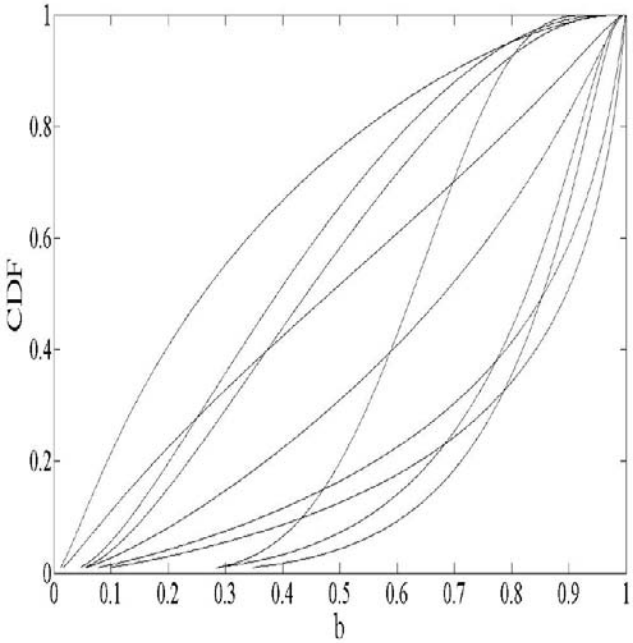


Figure 7: Family of Johnson distributions for input variable *b* for Problem A-2

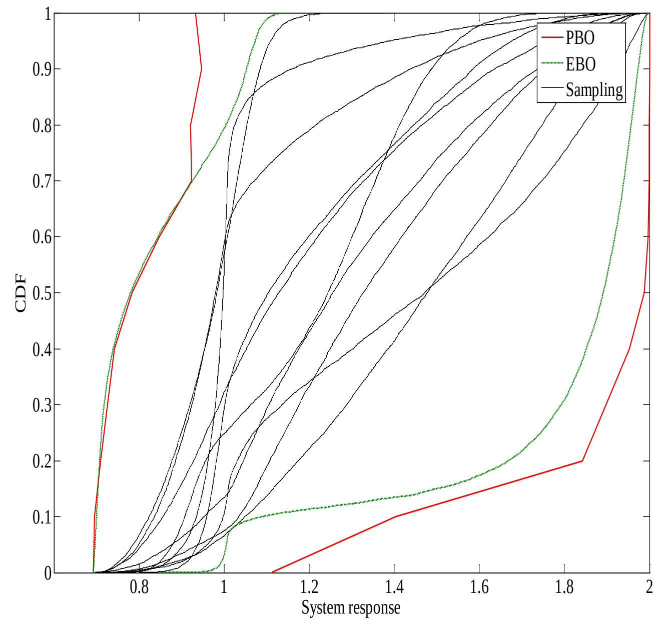


Figure 8: Bounds on CDF of system response for Problem A-2

Problem A-3

For this problem, both input variables *a* and *b* are described by multiple interval data ([0.5, 0.7], [0.3, 0.8], [0.1, 1.0]) and ([0.6, 0.6], [0.4, 0.85], [0.2, 0.9], [0.0, 1.0]), respectively, and have similar representations of uncertainty as shown in Figure 7. This problem belongs to Case 1 as described in Section 2 and is solved by both optimization and sampling-based strategies and the results are shown in Figure 9.

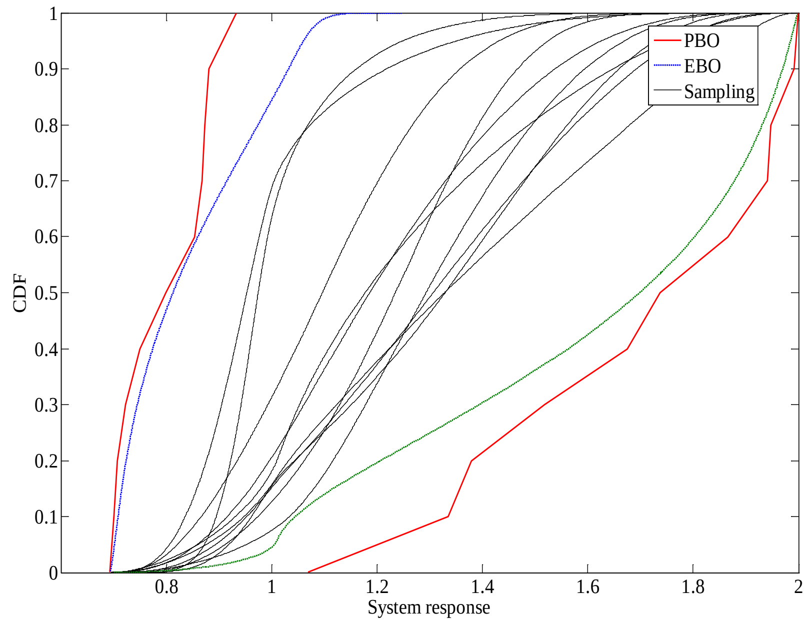


Figure 9: Bounds on CDF of system response for Problem A-3

Problem A-4

For this problem, the input variable *a* is described by a single interval [0.1, 1.0] and has the same uncertainty representation as shown in Figure 5. Input variable *b* is given by a log-normal probability distribution with imprecise parameters. These parameters are given by single intervals [0.0, 1.0] and [0.1, 0.5], respectively, and have similar uncertainty representations as shown in Figure 5. We follow the procedure outlined in Section 2.1 to obtain a family of log-normal distributions for input variable *b* given that the distribution parameters are represented as families of Johnson distributions. As an example, samples of cumulative density functions of the family of log-normal distributions for input variable *b* are shown in Figure 10.

This problem belongs to Case 2 as described in Section 2 and is solved by both optimization and sampling-based strategies and the results are shown in Figure 11.

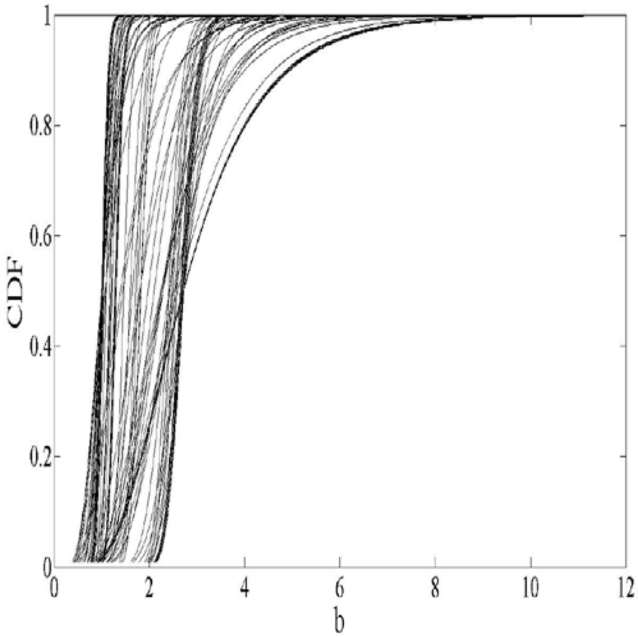


Figure 10: Family of log-normal distributions for input variable *b* for Problem A-4

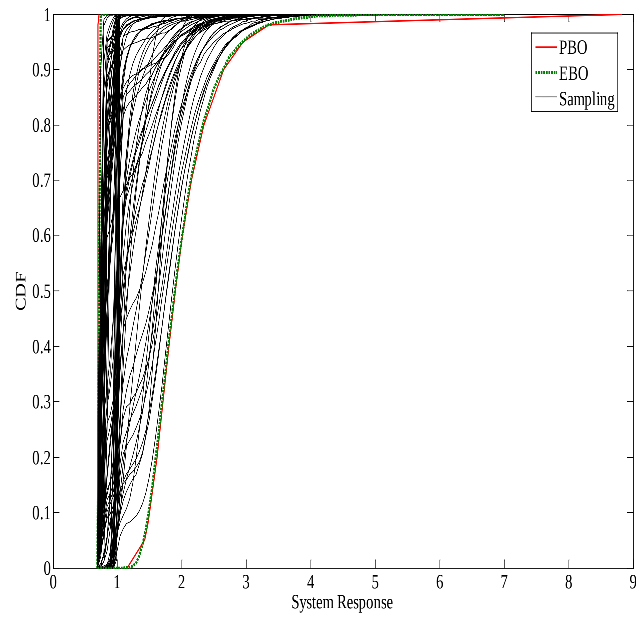


Figure 11: Bounds on CDF of system response for Problem A-4

Problem A-5

For this problem, the input variable *a* is described by multiple interval data ([0.5, 0.7], [0.3, 0.8], [0.1, 1.0]) and has a similar uncertainty representation as shown in Figure 7. Input variable *b* is given by a log-normal probability distribution with imprecise parameters. These parameters are described by multiple intervals ([0.6, 0.8], [0.2, 0.9], [0.0, 1.0]) and ([0.3, 0.4], [0.2, 0.45], [0.1, 0.5]), respectively, and have similar uncertainty representations as shown in Figure 7. We follow the procedure outlined in Section 2.1 to obtain a family of log-normal distributions for input variable *b* given that the distribution parameters are represented as families of Johnson distributions. Several sample cumulative density functions from the family of log-normal distributions for input variable *b* are shown Figure 12.

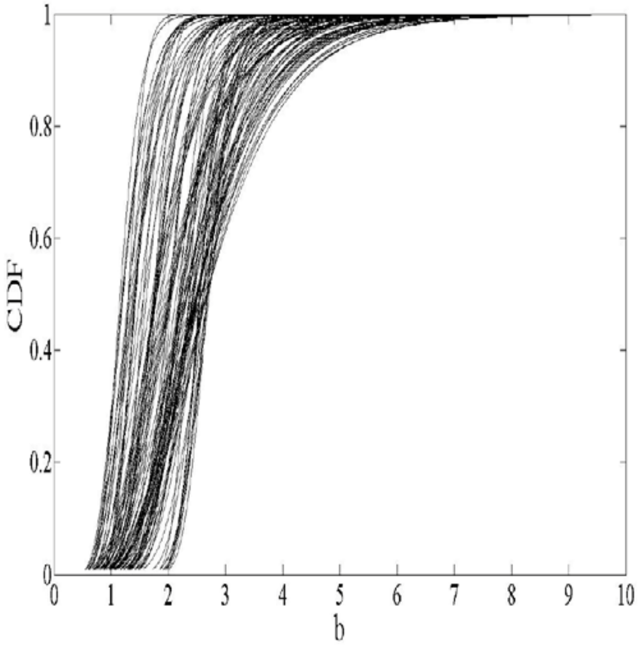


Figure 12: Family of log-normal distributions for input variable *b* for Problem A-5

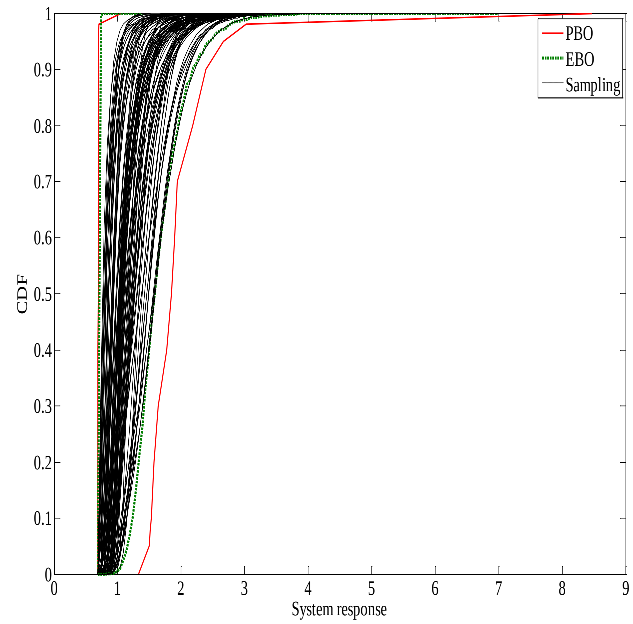


Figure 13: Bounds on CDF of system response for Problem A-5

This problem belongs to Case 2 as described in Section 2 and is solved by both optimization and sampling-based strategies and the results are shown in Figure 13.

Problem A-6

For this problem, the input variable *a* is described by a single interval [0.1, 1.0] and has the same uncertainty representation as shown in Figure 5. Input variable *b* is given by a log-normal probability distribution with precise parameters, 0.5 for each. This problem belongs to Case 1 as described in Section 2 and is solved by both optimization and sampling-based strategies and the results are shown in Figure 14.

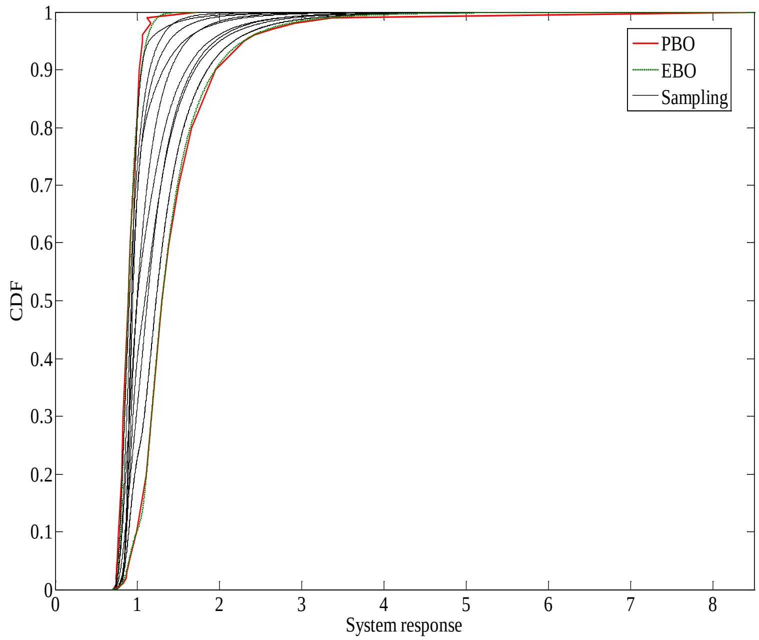


Figure 14: Bounds on CDF of system response for Problem A-6

It is seen in Figures 6, 11 and 14 that the bounds obtained by the expectation-based optimization (EBO) formulation and the percentile-based optimization (PBO) formulation almost coincide with each other. It is seen in Figures 8, 9 and 13 that the percentile-based optimization (PBO) formulation generates rigorous bounds compared to those obtained by the expectation-based optimization (EBO) formulation. The bounds obtained by EBO are still wider than those obtained by the sampling method.

The computational efforts for both PBO and EBO methods are listed in Table 1. It is seen from Table 1 that EBO is less expensive compared to PBO for each problem.

Table 1: Computation effort for Challenge Problem A

Challenge Problem A	PBO		EBO
	Percentile Points	Function Evaluations	Function Evaluations
A-1	21	7286	526
A-2	12	2962	148
A-3	11	3750	446
A-4	15	6123	349
A-5	15	10271	556
A-6	21	2494	127

3.2.2 Challenge Problem B

For this problem, the input variable k is given by a triangular distribution with imprecise k_{min}, k_{max}, and k_{mod}. The distribution parameters k_{min}, k_{max}, and k_{mod} are described by multiple interval data ([90, 100], [80, 110], ([200, 210], [200, 220], [190, 230]) and ([150, 160], [140, 170], [120, 180]) respectively, and have similar uncertainty representations as shown in Figure 7. We follow the procedure outlined in Section 2.1 to obtain a family of triangular distributions for input variable k given that the distribution parameters k_{min}, k_{max}, and k_{mod} are represented as families of Johnson distributions. Several sample cumulative density functions from the family of triangular distributions for input variable k are shown in Figure 15.

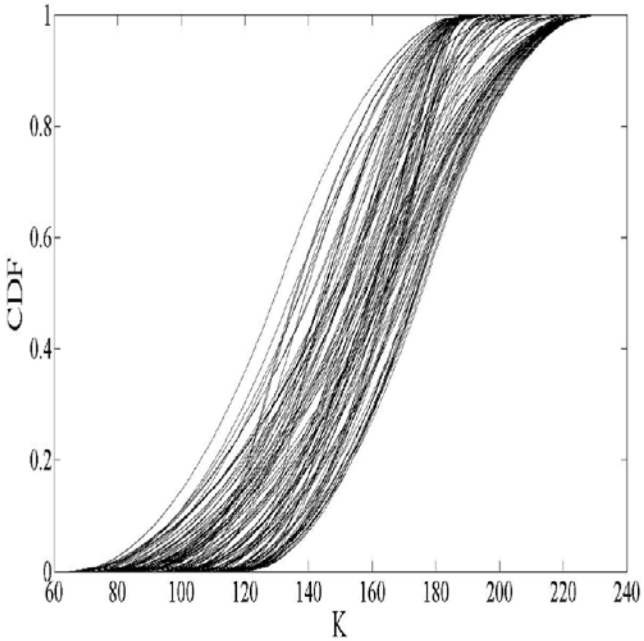


Figure 15: Family of triangular distributions for input variable *k* for Problem B

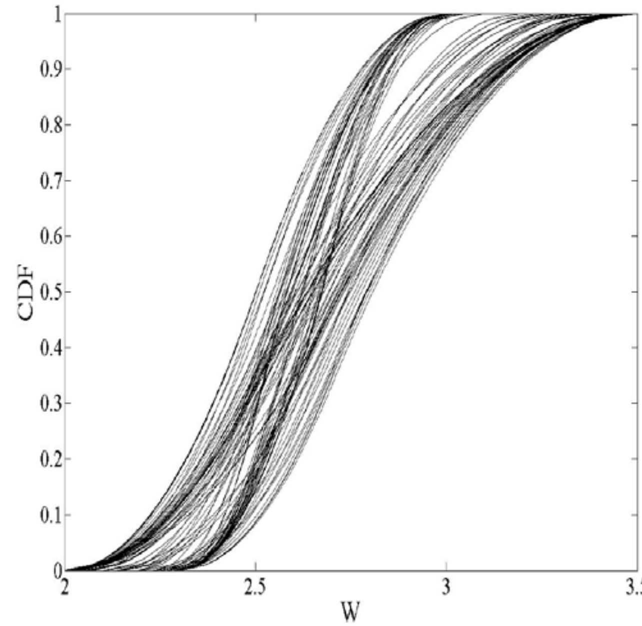


Figure 16: Family of triangular distributions for input variable ω for Problem B

Input variable *c* is described by multiple intervals ([5, 10], [15, 20], [25, 25]) and has a representation similar to that shown in Figure 7. Input variable ω is given by a triangular probability distribution defined on the interval $[\omega_{\min}, \omega_{\max}]$ with a mode of ω_{mod}. The distribution parameters ω_{min}, ω_{max}, and ω_{mod} are described by single intervals [2, 2.3], [2.5, 2.7] and [3.0, 3.5], respectively, and have similar representations as shown in Figure 5. We follow the procedure outlined in Section 2.1 to obtain a family of triangular distributions for input variable ω given that the distribution parameters ω_{min}, ω_{max}, and ω_{mod} are represented as families of Johnson distributions. Several sample cumulative density functions from the family of triangular distributions for input variable ω are shown in Figure 16.

This problem belongs to Case 2 as described in Section 2 and is solved by both expectation-based optimization (EBO) and sampling-based strategies and the results are shown in Figure 17. It is noted here that we did not find any converged solutions for this problem by the percentile-based optimization (PBO) method.

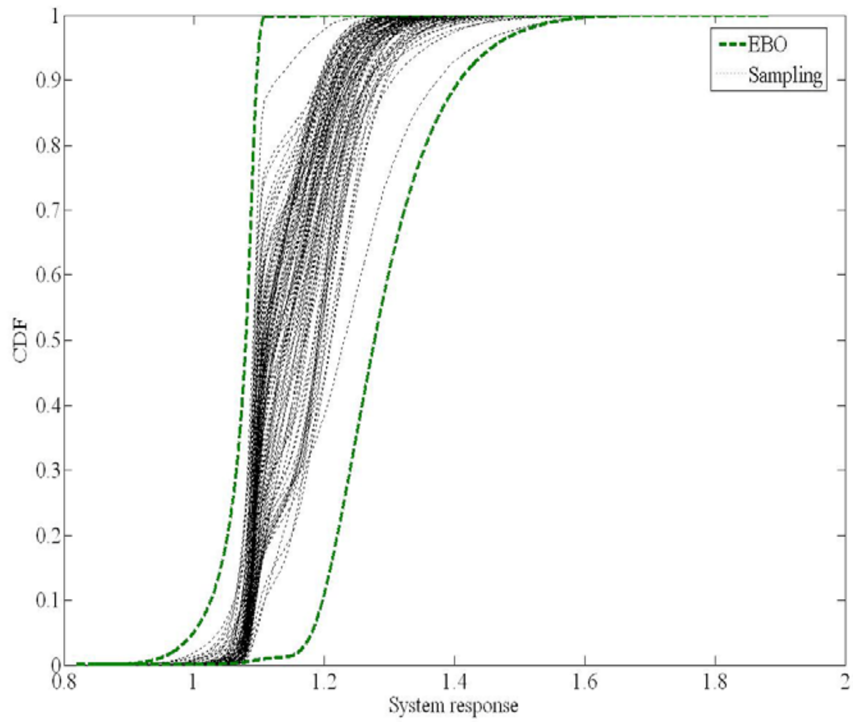


Figure 17: Bounds on CDF of system response for Problem B

It is seen in Figure 17 that the results obtained by EBO provide an envelope for the CDFs obtained by sampling.

Comparison with results from earlier studies

The results obtained by the proposed optimization-based methodology are compared with earlier solutions (Helton et al, 2004; Kozine and Utkin, 2004; De Cooman and Troffaes, 2004; Ferson and Hajagos, 2004 and Red- Horse and Benjamin, 2004). Ferson et al (2004) compared these earlier solutions in a tabular form. We have added an extra column to their table with the solutions from our approaches as shown in Table 2. The earlier solutions given in Table 2 are in terms of bounds on the expected values of the output, whereas our approaches give bounds on the entire output CDF. It can be mentioned here that bounds on the expected value of the system response are same as the bounds on the system response itself (Ferson et al, 2004 and Zaman et al, 2009a).

Table 2: Comparison of bounds on expected values

	Helton et al. (2004)	**Kozine and Utkin (2004)**	**De Cooman and Troffaes (2004)**	**Ferson and Hajagos (2004)**	**Red- Horse and Benjamin (2004)**	**Approach in this chapter**
1	-	[0.69, 2.0]	[0.692201, 2 .0]	[0.692, 2]	-	[0.6922,2]
2a	-	[0.93,1.84]	[0.956196, 1.8]	[0.84, 1.89]	-	[0.6922, 2]
3a	-	[0.944, 1.473]	[1.04881, 1.2016]	[0.83, 1.56]	-	[0.6922,2]
4	[1, 3.7]	[0.859, 1.108]	[1.00966, 4.08022]	[0.9944, 4.416]	-	[0.6922, 8.8329]
5a	(Graphical)	[1.45, 2.824]	[1.54027, 2.19107]	[1.05, 3.79]	-	[0.6922, 8.4681]
6	[1.05, 3]	[1.019, 2.776]	[1.05939, 2.86825]	[1.052, 2.89]	-	[0.7050, 8.4066]
B	(Graphical)	-	-	[1.17, 3.72]	(Graphical)	[0.8192, 1.8869]

Some quantitative agreement on the expected values is found among the earlier five studies, particularly for Problem A-1 as shown in Table 2. The results with the proposed approach (last column in Table 2) show some overlaps with the results of the earlier studies. Ferson et al (2004) argued that the disagreements in results among different studies are mostly due to the approaches by which the uncertainty described by multiple intervals is aggregated. Moreover, disagreement has also been observed regarding the answers to the Problems A-4 and A-6, though these problems did not have any multiple estimates for any of the input variables and thus should not have any problem due to the aggregation methods. Ferson et al (2004) mentioned four possible reasons for the observed discrepancies among the answers: *i)* nesting (due to difference in approaches, one result may be nested in others), *ii)* differences in truncation about whether or where the distributions were truncated to finite ranges, *iii)* numerical approximation error, and *iv)* different representations of independence.

Some authors mention repeated parameters in the system response expression as an issue when non-probabilistic methods are used for uncertainty quantification, as it can introduce the uncertainty of the repeated parameters more than once in the analysis (Ferson et al, 2004). Different authors employed different strategies for handling the issue of repeated parameters. These include sampling, exact evaluation, Mathematical programming, Independent natural extension, Subinterval reconstitution, Systematic sampling, Dependency tracking, vertex method, etc. The approach proposed in this chapter uses probabilistic uncertainty propagation methods, where the effect of repeated parameters is not an issue.

For Problem B, Helton et al (2004) used a sampling strategy and a Dempster-Shafer structure to compute the output bounds as [1.44, 2.86]. Red-Horse and Benjamin (2004) gave bounding distributions for D_s which had the support [1.4, 3.6]. Note that Helton et al (2004) and Red-Horse and Benjamin (2004) did not provide any numerical value for the output bounds in their papers. The numerical values for the output bounds mentioned in this chapter are estimated by Ferson et al (2004) from the graphical solutions. Ferson and Hajagos (2004) computed an interval ([1.17, 3.72]) based on moment propagation and argued that it might overestimate the uncertainty of the answer, whereas a sampling-based strategy might underestimate the uncertainty. We employed both sampling and optimization-based strategies but with a different aggregation method and computed the output interval as [0.8192, 1.8869].

It is seen that the optimization-based methodology proposed in this chapter gives wider bounds than other methods for the Problems A and narrower bounds for Problem B. However, instead of considering whether the bounds are narrower or wider, it is more helpful to evaluate bounds in terms of “rigor” and “optimality” as conceptually sketched in Figure 18. By rigorous, it is meant that the true interval of the possible quantile values lies within the computed bounds. By optimal, it is meant that the bounds are the narrowest possible, while still being rigorous.

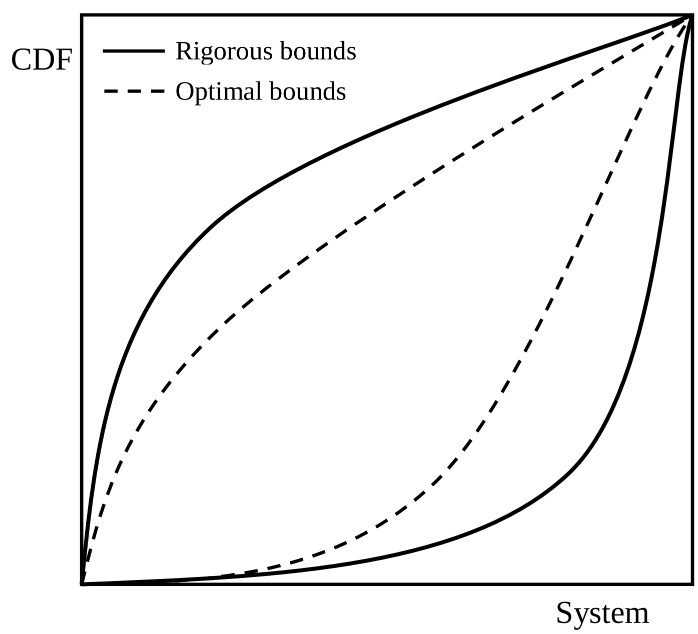


Figure 18: Rigorous vs. optimal bounds

The proposed PBO bounds are rigorous, provided that the set Θ encompasses all admissible distribution parameter values. This is because the quantile values are minimized or maximized over the entire set Θ. If the set of all admissible distribution parameter values is equal to Θ, then the bounds obtained by PBO are optimal. Suppose θ^* is the solution to the PBO problem. Because θ^* is an element of Θ, rigor requires that θ^* minimizes or maximizes $g^{\alpha}(x|\theta)$ over the set Θ. And it is impossible to construct a wider interval without violating the constraint $\theta \in \Theta$. In this instance the bounds are both rigorous and optimal. Again, if the set Θ is a superset of all actually admissible distribution parameter values, the bounds will still be rigorous, as the search over Θ includes a search over the set of all actually admissible distribution parameter values; however, the bounds will not be optimal because Θ is larger than the set of all admissible parameter values.

The differences in the results obtained by the different solution methods appear to create another type of epistemic uncertainty, which may be referred to as *method uncertainty*. The output intervals given by multiple methods may also be aggregated using the method described in section 2.

4. Conclusion

This chapter developed a probabilistic framework for the representation and propagation of uncertainty available as interval data. Both sampling and optimization-based methods are developed for two cases: (1) when the input variable is described by interval data, and (2) when the distribution parameters of the input variable are described by interval data. The methodology proposed in this chapter can handle all three classes of interval data mentioned in Section 4 in the same manner.

It is obvious that there is no unique or right answer to problems involving interval uncertainty (Oberkampf et al, 2004). However, the probabilistic methodology proposed in this chapter is flexible, and conversion of interval data to a probabilistic format enables the use of computationally efficient methods for probabilistic uncertainty propagation. The optimization-based approach adds further efficiency and ensures more rigorous bounds compared to the sampling-based approach. Further, the aggregation method for multiple intervals used here is also computationally efficient, and only scales polynomially in computational effort with respect to the number of interval data. The proposed approach facilitates the implementation of design optimization under uncertainty using efficient reliability-based design optimization (RBDO) methods, e.g., single loop, decoupled, etc., due to the use of a probabilistic format to represent all the uncertain variables. Note that the example problems assume statistical independence among the input random variables. However, the proposed approach will also work for any correlated interval variable with any appropriate multivariate input modeling method. Also, the accuracy of the optimization methods depends on the solver used. As in the case in nonlinear optimization, the proposed optimization-based strategies do not always guarantee convergence. Sometimes the problems might not have a unique solution and a non-gradient based solver (e.g., genetic algorithm) might help when convergence problems are encountered.

CHAPTER VI

INCLUSION OF CORRELATION EFFECTS IN MODEL PREDICTION UNDER DATA UNCERTAINTY

1. Introduction

In many uncertainty propagation analyses, it is likely that the marginal distribution types for the input variables are not known or cannot be specified accurately due to the presence of sparse point or interval data. This chapter proposes a methodology for multivariate input modeling of random variables by using a four parameter flexible Johnson family of distributions for the marginals that also accounts for data uncertainty. Semi-empirical formulas in terms of the Johnson marginals and covariances are presented to estimate the model parameters (reduced correlation coefficients). This multivariate input model is particularly suitable for uncertainty quantification problems that contain both aleatory and data uncertainty. In this chapter, a computational framework is developed to consider correlations among basic random variables as well as among their distribution parameters. We present a methodology for propagating both aleatory and data uncertainty arising from sparse point data through computational models of system response that assigns probability distributions to the distribution parameters and quantifies the uncertainty in correlation coefficients by use of computational resampling methods. For interval data, the correlations among the input variables are unknown. We formulate the optimization problems of deriving bounds on the cumulative probability distribution of system response, using correlations among the input variables that are described by interval data.

This chapter develops a multivariate input model for random variables having Johnson marginal distributions based on the Nataf transformation. We present semi-empirical formulas that relate $\rho_{0,ij}$ to ρ_{ij} in terms of the prescribed marginal distributions and covariances. This $\rho_{0,ij}$ can be used to generate correlated standard normal variates which are later transformed to an uncorrelated standard normal space for use in analytical reliability methods (e.g., FORM), or used to simulate correlated random variables for use in MCS.

It should be noted that given the presence of limited or interval data, the marginal distributions of the input variables and their correlation coefficients are also uncertain. Little to no work exists in the literature that considers uncertainty in correlation coefficients, and statistical correlations among distribution parameters. Moreover, for interval data, the correlations among the input variables are unknown and very few computationally efficient methods exist for propagation of both aleatory and statistical uncertainty that account for correlations among interval variables. Therefore, the contributions of this chapter are to (i) derive semi-empirical formulas that relate $\rho_{0,ij}$ to ρ_{ij} in terms of the Johnson marginal distributions and covariances and hence, develop a framework for multivariate input modeling of random variables modeled with Johnson marginal distributions; (ii) develop a method for the propagation of both aleatory and data uncertainty arising from sparse point data, by taking into account the uncertainty in correlations among basic random variables as well as correlations among distribution parameters; and (iii) develop a method for the propagation of both aleatory and data uncertainty arising from interval data by taking into account the correlations among basic random variables.

The rest of the chapter is organized as follows. Section 2 describes the computational framework for input modeling with Johnson distributions using correlations. Section 3 describes the proposed methods for the representation and propagation of both statistical and aleatory uncertainty using correlations. Section 4 gives the numerical results using two proposed methods: (1) for sparse point data, and (2) for interval data. Section 5 provides concluding remarks and suggestions for future work.

2. Input modeling with Johnson distributions using correlations

In this section, we propose a methodology to simulate correlated random variables (or uncorrelated standard normal variates) when the marginals and the correlation matrix [C] are the only information available. We use the Nataf transformation to calculate the reduced correlation coefficient ($\rho_{0,ij}$), similar to Der Kiureghian and Liu (1986). The Nataf transformation assumes that if Z_1 and Z_2 are standard normal variates obtained by marginal transformations of X_1 and X_2, and if we assume Z_1 and Z_2 are jointly normal, then X_1 and X_2 are jointly Nataf distributed. This process involves solving the following integral equation:

$$\rho_{12} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x_1 - m_1}{\sigma_1}\right)\left(\frac{x_2 - m_2}{\sigma_2}\right)\phi_2\left(z_1, z_2, \rho_{0,12}\right)dz_1 dz_2 \qquad (1)$$

where ρ_{12} is the correlation coefficient of the basic random variables, $\rho_{0,12}$ is the reduced correlation coefficient of the standard normal variates obtained by the following transformation:

$$Z_i = \Phi^{-1}\left[F_{X_i}\left(X_i\right)\right] \qquad \mathrm{i} = 1,2 \qquad (2)$$

$\phi_2(z_1, z_2, \rho_{0,12})$ is the bivariate normal PDF of zero means, unit standard deviations and correlation coefficient $\rho_{0,12}$.

$$\phi_2 = \frac{1}{2\pi\sqrt{1-\rho_{0,12}^2}} \exp\left[\frac{z_1^2 - 2\rho_{0,12} z_1 z_2 + z_2^2}{2\left(1-\rho_{0,12}^2\right)}\right] \qquad (3)$$

The integral equation in Eq. (1) has to be solved iteratively for given marginal distributions and correlation coefficient ρ_{ij}. To avoid solving an integral equation iteratively, Der Kiureghian and Liu (1986) proposed semi-empirical formulas for the ratio

$$F = \frac{\rho_{0,ij}}{\rho_{ij}} \qquad (4)$$

for some two-parameter marginal distributions, e.g., uniform, shifted exponential, shifted Rayleigh, Type-I largest value, log-normal, gamma, and Type-II largest value.

In this section, we develop similar semi-empirical formulas to calculate reduced correlation coefficient ($\rho_{0,ij}$) for random variables having Johnson marginal distributions. As in the case of Nataf transformation, we also assume that the transformed standard normal variates are jointly normal. Then instead of solving the integral equation in Eq. (1), we calculate the reduced correlation coefficient $\rho_{0,ij}$ by a numerical technique based on optimization. The procedure can be outlined as follows:

1. Define standard normal variates $Z = (Z_1, Z_2)$ obtained by marginal transformations of $X = (X_1, X_2)$ given by Eq. (2).

2. Z_1 and Z_2 are now assumed jointly normal with joint PDF given by Eq. (3).

3. Choose an initial value for $\rho_{0,ij}$.

4. Calculate the correlation coefficient ρ_{ij} from the simulated correlated variables X_1 and X_2.

5. Iterate until the correlation coefficients calculated from the original data ($\rho_{ij, data}$) and the simulated correlated variables ($\rho_{ij, simulation}$) become equal. Obtain the reduced correlation coefficient $\rho_{0,ij}$. This is achieved by the following optimization problem:

$$\min_{\rho_{0,ij}}\left(\rho_{ij,data} - \rho_{ij,simulation}\left(\rho_{0,ij}\right)\right)^2 \quad (5)$$
$$s.t. -0.99 \le \rho_{0,ij} \le 0.99$$

We solved this optimization problem using the MATLAB function *fmincon*, which implements a sequential quadratic programming algorithm.

This calculation can be tedious; therefore, we also present semi-empirical formulas for the correction factor F when marginal distributions come from the Johnson family of distributions. These formulas are based on the following properties (Der Kiureghian and Liu, 1986):

1. F is a function of ρ_{ij} and the parameters of the two marginal distributions.

2. F is always greater than one for any arbitrary ρ_{ij} and marginal distributions.

Based on the above properties, we propose three semi-empirical formulas for the following three cases:

1. Both X_i and X_j are bounded Johnsons (SB) (Eq. (6))

2. Both X_i and X_j are unbounded Johnsons (SU) (Eq. (7))

3. X_i is unbounded Johnson (SU) and X_j is bounded Johnson (SB) (Eq. (8))

$$\begin{aligned}F = {} & 1.2924 + 0.0443\rho_{ij} - 0.1316(V_i + V_j) + 0.1508(\beta_{1i} + \beta_{1j}) - 0.0864(\beta_{2i} + \beta_{2j}) \\ & + 0.0138\rho_{ij}(V_i + V_j) - 0.0205\rho_{ij}(\beta_{1i} + \beta_{1j}) - 0.0118\rho_{ij}(\beta_{2i} + \beta_{2j}) - 0.0203V_iV_j \\ & + 0.0031(V_i(\beta_{1i} + \beta_{1j}) + V_j(\beta_{1i} + \beta_{1j})) + 0.0128(V_i(\beta_{2i} + \beta_{2j}) + V_j(\beta_{2i} + \beta_{2j})) \\ & + 0.0519\beta_{1i}\beta_{1j} - 0.0083\beta_{2i}\beta_{2j} - 0.0388(\beta_{1i}\beta_{2i} + \beta_{1j}\beta_{2j}) - 0.0111(\beta_{1i}\beta_{2j} + \beta_{1j}\beta_{2i}) \\ & - 0.0488\rho_{ij}^2 + 0.1814(V_i^2 + V_j^2) + 0.0226(\beta_{1i}^2 + \beta_{1j}^2) + 0.0199(\beta_{2i}^2 + \beta_{2j}^2)\end{aligned} \tag{6}$$

$$\begin{aligned}F = {} & 1.0375 + 0.0133\rho_{ij} + 0.0401(V_i + V_j) + 0.0630(\beta_{1i} + \beta_{1j}) - 0.0055(\beta_{2i} + \beta_{2j}) \\ & + 0.0135\rho_{ij}(V_i + V_j) + 0.0273\rho_{ij}(\beta_{1i} + \beta_{1j}) - 0.0042\rho_{ij}(\beta_{2i} + \beta_{2j}) - 0.1649V_iV_j \\ & - 0.0212(V_i(\beta_{1i} + \beta_{1j}) + V_j(\beta_{1i} + \beta_{1j})) + 0.0028(V_i(\beta_{2i} + \beta_{2j}) + V_j(\beta_{2i} + \beta_{2j})) \\ & - 0.0200\beta_{1i}\beta_{1j} + 0.0079\beta_{2i}\beta_{2j} - 0.0016(\beta_{1i}\beta_{2i} + \beta_{1j}\beta_{2j}) + 0.0003(\beta_{1i}\beta_{2j} + \beta_{1j}\beta_{2i}) \\ & + 0.0001\rho_{ij}^2 - 0.0223(V_i^2 + V_j^2) + 0.0146(\beta_{1i}^2 + \beta_{1j}^2) - 0.0049(\beta_{2i}^2 + \beta_{2j}^2)\end{aligned} \tag{7}$$

$$\begin{aligned}F = {} & 1.1432 - 0.6575\rho_{ij} + 0.8593V_i + 0.1864V_j + 0.2871\beta_{1i} + 0.7429\beta_{1j} + 0.2446\beta_{2i} - 0.8899\beta_{2j} \\ & - 0.0411\rho_{ij}V_i + 0.0013\rho_{ij}V_j - 0.3753\sigma_{ij}\beta_{1i} - 0.0137\rho_{ij}\beta_{1j} + 0.2135\rho_{ij}\beta_{2i} + 0.0076\rho_{ij}\beta_{2j} \\ & - 0.2269V_iV_j + 0.2459V_i\beta_{1i} + 0.0127V_i\beta_{1j} - 0.2800V_i\beta_{2i} + 0.0292V_i\beta_{2j} + 0.3745V_j\beta_{1i} \\ & - 0.1167V_j\beta_{1j} - 0.1528V_j\beta_{2i} + 0.1532V_j\beta_{2j} + 0.1126\beta_{1i}\beta_{1j} - 0.0661\beta_{1i}\beta_{2i} - 0.2850\beta_{1i}\beta_{2j} \\ & - 0.0798\beta_{1j}\beta_{2i} - 0.1778\beta_{1j}\beta_{2j} + 0.1670\beta_{2i}\beta_{2j} - 0.0195\rho_{ij}^2 + 0.0143V_i^2 + 0.0466V_j^2 \\ & + 0.2770\beta_{1i}^2 + 0.1045\beta_{1j}^2 - 0.0450\beta_{2i}^2 + 0.0723\beta_{2j}^2\end{aligned} \tag{8}$$

In Eqs. (6), (7) and (8), F is the correction factor, ρ_{ij} is the original correlation coefficient, V_i and V_j are the coefficients of variation, β_{1i}, β_{1j} are the skewness and β_{2i}, β_{2j} are the kurtosis for X_i and X_j, respectively. These equations are obtained by least-square fitting to a general second degree polynomial. Note that in Eqs. (6) and (7), X_i and X_j are the same Johnson type (both bounded, or both unbounded), therefore the formulas for F are symmetric in i and j. In Eq. (8) X_i and X_j are of different type (X_i is unbounded, X_j is bounded), therefore this formula is not expected to be symmetric.

In order to validate the fitted models, the coefficients of determination R^2 for each of Eqs. (6)-(8) are calculated. However, in the presence of multiple regressor variables, the R^2 value might overestimate the strength of the model, as it always increases with increasing number of regressor variables in the models (Haldar and Mahadevan, 2000). Therefore, another statistic, the adjusted R^2 is also computed for each of the models, which also accounts for the sample size of the regressor variables (Cramer, 1987). The overall significance of the proposed regression models is also tested using the F-test. The F-tests are performed at 0.05 significance level. The values of *p* in Table 1 reflect the significance of F-statistics. For example, a value of *p* less than the significance level of 0.05 indicates a good fit. The regression statistics are given in Table 1.

Table 1: Regression statistics for the semi-empirical formulas

Distributions	R^2	R^2-adjusted	F	p	$F_{critical}$
Johnson SB-SB	0.9597	0.9350	14.3740	1.8162e-010	1.9332
Johnson SU-SU	0.8959	0.8452	5.4608	3.7955e-006	1.8599
Johnson SU-SB	0.9818	0.9645	27.1762	0	1.7390

It is seen in Table 1 that the values of both R^2 and R^2-adjusted are reasonably high for each of the fitted models, the values of F are well above the corresponding critical values, and the respective *p* values are much less than the significance level 0.05, indicating that the models in Eqs. (6)-(8) provide a good fit to the data.

Table 2 lists the allowable domains for the correlation coefficients for the semi-empirical formulas given above. The proposed formulas are valid for only these ranges of the correlation coefficients. The other coefficients in the formulas are functions of the first four moments of the marginal Johnson distributions and are constrained by the Johnson translation system.

Table 2: Allowable domains for the correlation coefficients

Distributions	ρ_{ij}
Johnson SB-SB	-0.95 to 0.92
Johnson SU-SU	-0.93 to 0.93
Johnson SU-SB	-0.94 to 0.93

Once the reduced correlation coefficients are obtained by one of the semi-empirical formulas presented above, the next step is to transform the correlated variables into uncorrelated standard normal variates for use in analytical reliability methods or to generate correlated random variates for use in MCS. The procedure of transformation to uncorrelated standard normal variates is as follows:

1. Calculate the correction factor *F* for the given marginals and correlation coefficient ρ_{ij} and thus obtain the reduced correlation coefficient $\rho_{0,ij}$ and reduced correlation matrix $[C']$.

$$[C'] = \begin{bmatrix} 1 & \rho_{0,12} \\ \rho_{0,12} & 1 \end{bmatrix} \quad (9)$$

2. Generate correlated standard normal variates *(Z)* from the joint PDF given in Eq. (3) with the reduced correlation matrix $[C']$.

3. Transform correlated standard normal variates *(Z)* to uncorrelated standard normal *(u)* space by the transformation

$$u = L^{-1}(Z)^t \quad (10)$$

where L is the lower triangular matrix obtained by Cholesky factorization of the reduced correlation matrix $[C^{'}]$.

Since Monte Carlo methods have widespread applications in uncertainty analysis, it is important to include correlations into the computational process, when the input variables are correlated. As mentioned earlier that there are several ways to generate correlated random variables with given marginals and correlation matrix. In this chapter, we have used the following procedure to generate correlated random variables:

1. Calculate the correction factor F for the given marginals and correlation coefficient ρ_{ij} and thus obtain the reduced correlation coefficients $\rho_{0,ij}$ and reduced correlation matrix $[C^{'}]$.

2. Generate correlated standard normal variates (Z) from the joint PDF given in Eq. (3) with the reduced correlation matrix $[C^{'}]$.

3. Generate correlated random variables (X) with given marginals by the following transformation:

$$F_{X_i}(X_i) = \Phi\ [Z_i] \qquad \mathrm{i} = 1,2 \tag{11}$$

We note here that the procedures described above require that the reduced correlation matrix $[C']$ be at least positive semi-definite, if not positive definite. This condition is satisfied in almost all practical cases, because the original correlation matrix $[C]$ is by definition positive define and the differences between the original correlation coefficients ρ_{ij} and the reduced correlation coefficients $\rho_{0,ij}$ are usually small (Liu and Der Kiureghian, 1986). However, in some practical cases, when we construct the reduced correlation matrix $[C']$ by estimating the pairwise correlation coefficients independently, it is likely that the correlation matrix $[C']$ will be non-positive semi-definite. Methods (e.g., Higham, 2002; Mishra, 2007) exist for adjusting a non-positive semi-definite matrix so that it can be positive semi-definite and remains as close as possible to the original matrix.

Once we have transformed the original random variables to uncorrelated standard normal space or we have generated correlated input variables, the next step is to propagate this uncertainty through models of system response by any uncertainty propagation method (e.g., FORM or MCS).

3. Proposed Methodology for uncertainty propagation under uncertain correlations

In this section we describe our proposed methodology for the propagation of epistemic and aleatory uncertainty using correlations. First, we fit a family of Johnson distributions to sparse point and interval data on the input variables using the moment matching approach. Moment matching involves equating the moments derived from data to those of the probability distribution being fit. A detailed discussion on fitting Johnson distributions to sparse point and interval data can be found in McDonald et al (2009) and Zaman et al (2009a), respectively.

The Johnson family is a generalized family of distributions that can represent normal, lognormal, bounded, or unbounded distributions. Because of their flexibility, Johnson distributions can be used for probabilistic representation of sparse point data or interval data when the underlying probability distribution is not known. As discussed earlier in

Section 1, the Johnson family is a convenient choice for this purpose among other four parameter distributions, as it has easy transformation to standard normal space, which then can be conveniently used for further analysis.

In Section 3.1 we describe novel approaches for uncertainty quantification with sparse point data. In Section 3.2 we describe the methods for quantification of both aleatory and interval uncertainty using correlations among interval variables.

3.1 Statistical Uncertainty Quantification via Jackknife for sparse point data

It should be noted that given the presence of limited data, the marginal distributions of the input variables and their correlation coefficients are also uncertain. We introduce a versatile approach for uncertainty quantification of distribution parameters and correlation coefficients among basic random variables as well as their distribution parameters for sparse point data. This approach assumes that both the basic random variables and their distribution parameters are Johnson distributed, and uses a jackknife technique to estimate the distribution of the distribution parameters and correlation coefficients among basic random variables. The assumption of the distribution parameters having the Johnson distribution allows for both the possibility of a non-normal distribution for the small sample size as well as the distribution asymptotically approaching normality

Jackknifing (Arvesen, 1969 and Miller, 1974) is used to estimate the bias and standard error in a statistic, when a random sample of observations is used to calculate it. The basic idea behind the jackknife estimator lies in systematically recomputing the statistic estimate, leaving out one observation at a time from the sample set. From this new set of "observations" for the statistic an estimate for the bias can be calculated and an estimate for the variance of the parameter. We propose the following algorithm for uncertainty quantification of the distribution parameters and correlation coefficients.

Note that any appropriate point estimation technique may be used for this procedure. The use of the Johnson distribution for the underlying basic random variable avoids the problem of incorrect classification of the distribution type. The use of the Johnson distribution for characterizing parameter uncertainty allows for relaxation of the assumption of asymptotic normality. The Johnson distribution can much more closely match the shape of the parameter's distribution even if it is non-normal, as it may be under small sample sizes, and will still be appropriate for large samples.

Algorithm for Uncertainty Quantification in Distribution Parameters and Correlation Coefficients

Set i = 1

while (i <= N)

> **Delete** observation *i* from the original set of observations
>
> **Estimate** the Johnson distribution parameters and correlation coefficient of the basic random variables on the basis of the *N*-1 remaining points.
>
> **Record** as estimate *i*.
>
> **Restore** observation *i* to the set of original observations.
>
> *i* = *i* + 1

end while

Obtain a set of distribution parameters and correlation coefficients

Fit a Johnson Distribution to the set of parameter estimates obtained in the while loop.

Once the uncertainty in the distribution parameters and correlation coefficients of the basic random variables are quantified, the next step is to consider the correlations among the distribution parameters. As we have a set of distribution parameters for which we fit Johnson distribution for the distributions of distribution parameters, we are now able to generate correlated distribution parameters from their marginals using the approach described in section 2. Therefore, the output uncertainty quantification procedure with sparse point data on the input, considering correlations among basic random variables as well as among their distribution parameters can be outlined as follows:

1. Obtain N sets of distribution parameters and correlation coefficients for the basic random variables (X) of sample size N via jackknife.

2. Fit Johnson distributions to the set of distribution parameters obtained in step 1. Now, we have four marginal distributions for the distribution parameters of the basic random variables.

3. Calculate the correlation coefficients ρ_{ij} for the distribution parameters from the set of distribution parameters obtained in step 1.

4. Obtain reduced correlation coefficients $\rho_{0,ij}$ for the distribution parameters by the procedure described in section 2.

5. Generate N sets of correlated distribution parameters using the marginal distributions obtained in step 2 and reduced correlation coefficients obtained in step 4 by the procedure described in section 2.

6. Generate correlated input variables using each set of distribution parameters obtained in step 5 and correlation coefficients obtained in step 1 by the procedure

described in section 2 and propagate through model of system response by MCS to obtain a CDF of system response.

7. Repeat step 6 *N* times and thus obtain a family of CDFs for the system response.

This procedure is illustrated in Section 4 through an example problem.

3.2 Uncertainty quantification with interval data

In order to express and propagate interval data using probabilistic methods, it is necessary to fit probability distributions to interval data. An approach for fitting a family of Johnson distributions to interval data has been discussed in Chapter IV. As we are able to calculate the bounds on the moments of an uncertain quantity characterized by interval data, we can require that the moments of the distribution fall between the upper and lower bounds given from the estimation procedures. With interval data, it is impossible to know the true moments of the data, thus there are infinitely many possible probability distributions that can represent the interval data. This uncertainty in the moments of the data also creates uncertainty in the parameters of the Johnson distribution. Chapter IV proposed algorithms to compute bounds on moments for single interval and multiple interval data. Chapter V proposed an optimization-based methodology for uncertainty propagation with interval data.

Interval data are encountered frequently in practical engineering problems as discussed in Chapter II. In many problems, it is likely that interval data for individual input variables are not observed simultaneously. Therefore, it is impractical to calculate the correlation coefficients among the input variables which are described by interval data. Rather it is assumed that with interval data the correlations among the input variables are unknown and therefore can range from -1 to +1. In the following discussion, we reformulate the optimization-based approaches proposed in Chapter V to include correlations among input random variables.

The first approach is a percentile-based optimization (PBO) method which minimizes and maximizes the system response $g_\alpha(x|m)$ conditioned on a set of moments *(m_i)* for the input variables at different percentile values *(α)* and thus bounds on system response CDF is obtained. We include correlation in this analysis by adding the reduced correlation coefficient $\rho_{0,ij}$ as a decision variable, which ranges from r_{min} to r_{max}. Note that for a pair of input random variables, there exists either positive or negative correlation. The proposed approach requires that the designer has the knowledge about the correlation type among the input random variables. Therefore, the quantities r_{min} and r_{max} are specified by the designer in the optimization formulation. For example, r_{min} and r_{max} may assume values between -1 to -0.1 for negatively correlated variables and 0.1 to 1 for positively correlated variables.

The implementation of this uncertainty quantification approach considering correlations is as follows:

1. Calculate the bounds on the first four moments of single or multiple interval data by the methods described in Chapter IV.

2. Solve the following optimization problems at different percentile values *(α)* to obtain bounds on output CDF.

$$\begin{aligned}
&\min_{m,\rho_{0,ij}}/\max\, g_{\alpha}\left(x \middle| m, \rho_{0,ij}\right) \\
&s.t. m_i \geq a_i \\
&\quad m_i \leq b_i \qquad i = 1,2.....,4 \\
&\quad r_{\min} \leq \rho_{0,ij} \leq r_{\max} \\
&\quad \beta_2 - \beta_1 - 1 \geq 0 \\
&\quad \beta_2 - 2\beta_1 - 3 \leq 0 \\
&\quad \text{where } \beta_1 = m_3^2 / m_2^3 \\
&\qquad\qquad \beta_2 = m_4 / m_2^2
\end{aligned} \tag{12}$$

The last two nonlinear constraints ensure that the optimizer only selects those values of moments that suggest a bounded Johnson distribution fit, so that the resulting distribution lies within the bounds of the interval data specified. It is noted here that the objective function in this optimization problem is conditioned on a set of moments and reduced correlation coefficient $\rho_{0,ij}$ for the input variables and estimates the parameters of Johnson distribution from the set of moments in each iteration by the method described in Chapter IV.

Percentile-based optimization is expensive as it requires solving the problems repeatedly at different α-levels. Therefore, Chapter V proposed another expectation-based optimization (EBO) strategy to obtain approximate bounds on system response CDFs which is computationally less expensive. This formulation is based on the assumption that the sets of distribution parameters of input variables which result in minimum or maximum expectation of the system response (*E(g(x))*) can also give an upper bound on the entire CDF of the system response (*g(x)*) for the minimization problem and a lower bound for the maximization problem, respectively. A proof in support of this statement is given in Chapter V. As in the case of PBO, we also reformulate this expectation-based optimization (EBO) problem by including correlations in the analysis. Its implementation is as follows:

1. Calculate the bounds on the first four moments of single or multiple interval data by the methods described in Chapter IV.

2. Obtain two set of moments and reduced correlation coefficients $\rho_{0,ij}$ that minimize and maximize the expected value of the system response *(E(g(x)))* conditioned on a set of moments *(m_i)* and reduced correlation coefficient $\rho_{0,ij}$ for the input variables. These are obtained by the following optimization problems:

$$\begin{aligned}
&\min_{m,\rho_{0,ij}}/\max E(g(x|m,\rho_{0,ij})) \\
&s.t. m_i \geq a_i \\
&\quad m_i \leq b_i \qquad i = 1,2.....,4 \\
&\quad r_{\min} \leq \rho_{0,ij} \leq r_{\max} \\
&\quad \beta_2 - \beta_1 - 1 \geq 0 \\
&\quad \beta_2 - 2\beta_1 - 3 \leq 0 \\
&\quad \text{where } \beta_1 = m_3^2 / m_2^3 \\
&\qquad\qquad \beta_2 = m_4 / m_2^2
\end{aligned} \tag{13}$$

The last two nonlinear constraints ensure that the optimizer only selects those values of moments that suggest a bounded Johnson distribution fit as mentioned earlier.

3. Obtain two sets of parameters of Johnson distribution for the input variables from the sets of moments obtained in step 2.

4. Construct the CDF of the system response given a set of reduced correlation coefficients and distribution parameters for the input variables by any probabilistic uncertainty propagation method (e.g., FORM or MCS) and thus obtain approximate bounds on the CDF of the system response.

In summary, this chapter developed a multivariate input model of random variables by using a four parameter flexible family of distributions for the marginals to account for data uncertainty. The proposed multivariate input model is then used to develop a computational framework for the uncertainty propagation that considers statistical correlations among basic random variables as well as among their distribution parameters.

4. Example Problems

In this section, the proposed methods are applied to a single aerodynamic data set for the upper stage of the Two-Stage-To-Orbit (TSTO) Highly Reliable Reusable Launch Systems (HRRLS) concept vehicle, as described in Chapter III. A response surface (Eq. 13 in Chapter II) for a model-predicted drag coefficient (C_D) is used for uncertainty propagation, which is a function of Mach number (Mach) and angle of attack (AoA). We wish to quantify the uncertainty in C_D given available information concerning Mach and AoA. The information on Mach and AOA is available as either sparse point data or interval data.

4.1 Uncertainty propagation with sparse point data

In this case, Mach and AoA are assumed to be given by sparse point data as shown in Table 3. The distributions of Mach and AoA are inferred from point data and fitted to bounded Johnson distributions by the method of matching moments. Since the data sets of Mach and AoA are small, they are jackknifed and a Johnson distribution is fitted using each of the jackknifed parameter estimates with one observation deleted in order to quantify the statistical uncertainty in the distribution parameters. Uncertainty in correlation coefficients is also quantified via jackknife.

Table 3: Sparse point data for Mach and AoA

Data	
Mach	AoA
6.52	18.52
6.06	16.04
5.49	17.52
6.52	16.53
5.74	18.63
5.74	15.94
5.34	16.32
6.24	17.40
5.42	16.00
6.10	18.54

The output is a family of CDFs of system response conditioned on each jackknifed observation of the distribution parameters of Mach and AoA and the correlation coefficient ρ_{ij}. The results are shown in Figure 1.

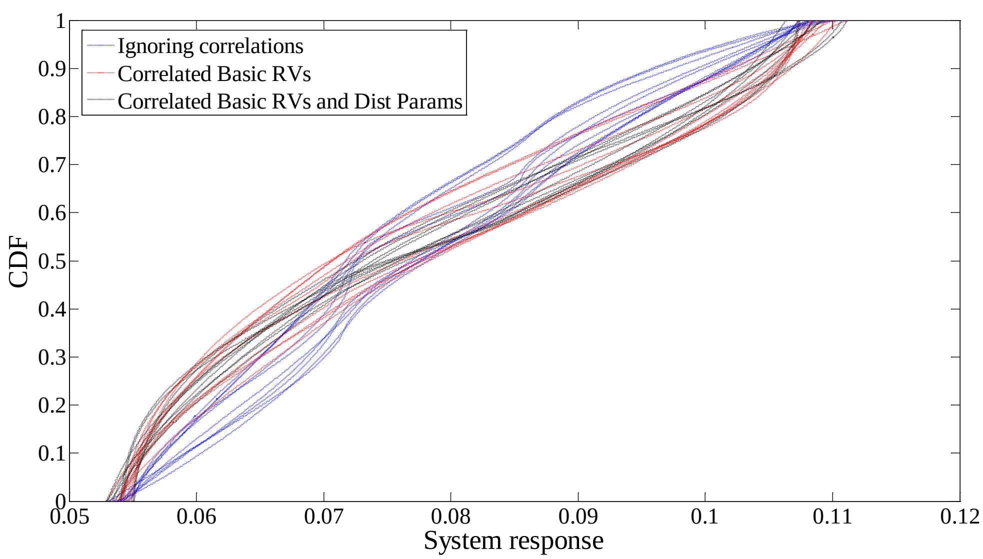


Figure 1: Family of CDFs of system response for sparse point data

Figure 1 shows the results for three cases: *i*) ignoring correlations, *ii*) including correlations in basic random variables and *iii*) including correlations in basic random variables and their distribution parameters. It is seen in Figure 1 that correlations among the basic random variables as well as their distribution parameters have significant impact on the distributions of system response, especially in the tails of the distributions, which are the regions of interest to the decision maker. It is also seen that we obtain a tighter scatter in output distributions, i.e., narrower bounds on the output CDFs when we consider correlations among basic random variables as well as among their distribution parameters. We further obtain narrower bounds on the output CDF in Case (*iii*) as compared to that obtained in Case (*ii*), which suggests that statistical correlations among the distribution parameters should be included in predicting the system response.

4.2 Uncertainty propagation with multiple interval data

In this case, the uncertainty in both Mach and AoA are described by multiple interval data as given in Table 4. Here, it is assumed that there exists positive correlation between Mach and AoA and their correlation coefficients range from 0.1 to 1.This problem is solved by both PBO and EBO approaches presented in section 3.2. The output is in the form of bounds on the system response CDF. The results are shown in Figure 2.

Table 4: Multiple interval data for Mach and AoA

Data	
Mach	AoA
[5, 6; 5.5, 6.1; 6, 6.5; 5.4, 6.2; 5.6, 6.6]	[18, 19; 18.5, 20; 19, 20; 19.5, 21; 18, 20.5]

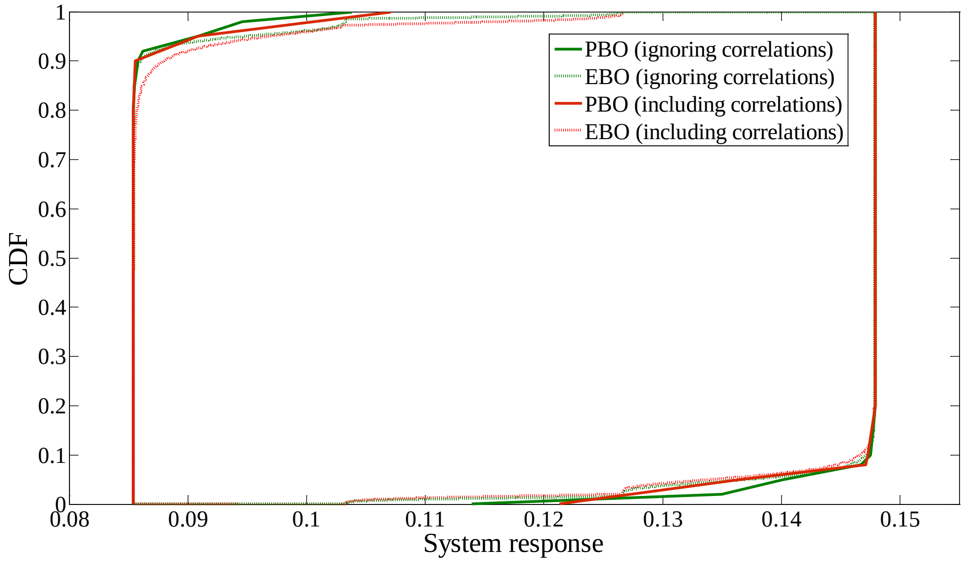


Figure 2: Bounds on CDF of system response for interval data

Figure 2 shows the results for two cases: *i*) ignoring correlations and *ii*) including correlations in basic random variables. It is seen in Figure 2 that correlations among the basic random variables have significant impact on the distributions of system response, especially in the tails of the distributions. It is also seen that we obtain narrower bounds on the output CDFs when we consider correlations among basic random variables. As expected, EBO is less expensive (65 function evaluations) for correlated input modeling as compared to PBO (850 function evaluations) which was solved at 15 different percentile values. Similarly, for uncorrelated input modeling, EBO required only 60 function evaluations as compared to 810 function evaluations by PBO (solved at 15 different percentile values).

5. Conclusion

This chapter developed a methodology for multivariate input modeling of random variables by using a flexible Johnson family of distributions for the marginals that also accounts for data uncertainty. Semi-empirical formulas in terms of the Johnson marginals and covariances are presented to estimate the reduced correlation coefficients. This reduced correlation coefficient is then used to transform correlated random variables to uncorrelated standard normal space for use in analytical reliability analysis methods and to generate correlated random variables for use in MCS for Johnson distributed marginal distributions. This chapter also developed a methodology for propagating both aleatory and data uncertainty arising from sparse point data through computational models of system response. The methodology differs from existing approaches in that it assigns probability distributions to the distribution parameters and quantifies the uncertainty in correlation coefficients through the use of computational resampling methods. A methodology has also been developed for the propagation of both aleatory and interval uncertainty in the presence of correlations among interval variables. These methods are illustrated with example problems. The results show that statistical correlations have significant impact on uncertainty quantification, especially in the tails of the output distributions, which are the regions of interest to the decision maker. The proposed approach facilitates the implementation of design optimization under uncertainty considering correlations.

CHAPTER VII

ROBUSTNESS-BASED DESIGN OPTIMIZATION UNDER DATA UNCERTAINTY

1. Introduction

This chapter proposes formulations and algorithms for design optimization under both aleatory (i.e., natural or physical variability) and epistemic uncertainty (i.e., imprecise probabilistic information), from the perspective of system robustness. The proposed formulations deal with epistemic uncertainty arising from both sparse and interval data without any assumption about the probability distributions of the random variables. A decoupled approach is proposed in this chapter to un-nest the robustness-based design from the analysis of non-design epistemic variables to achieve computational efficiency. The proposed methods are illustrated for the upper stage design problem of a two-stage-to-orbit (TSTO) vehicle, where the information on the random design inputs are only available as sparse point and/or interval data. As collecting more data reduces uncertainty but increases cost, the effect of sample size on the optimality and robustness of the solution is also studied. A method is developed to determine the optimal sample size for sparse point data that leads to the solutions of the design problem that are least sensitive to variations in the input random variables.

The essential elements of robust design optimization are: (1) maintaining robustness in the objective function (objective robustness); (2) maintaining robustness in the constraints (feasibility robustness); (3) estimating mean and measure of variation (variance) of the performance function; and (4) multi-objective optimization. The rest of this section briefly reviews the literature with respect to these four elements and establishes the motivation for the current study.

Objective robustness

In robust optimization, the robustness of the objective function is usually achieved by *simultaneously optimizing its mean and minimizing its variance*. Two major robustness measures are available in the literature: one is the variance, which is extensively discussed in the literature (Du and Chen, 2000; Lee and Park, 2001 and Doltsinis and Kang, 2004) and the other is based on the percentile difference (Du et al, 2004). Although the percentile difference method has the advantage that it contains the information of probability in the tail regions of the performance distribution, this method is only applicable to unimodal distributions. Variance as a measure of variation of the performance function can be applied to any distribution (unimodal or multimodal), but it only characterizes the dispersion around the mean (Huang and Du, 2007).

Feasibility robustness

Feasibility robustness i.e., robustness in the constraints can be defined as *satisfying the constraints of the design in the presence of uncertainty*. Du and Chen (2000) classified the methods of maintaining feasibility robustness into two categories, methods that use probabilistic and statistical analysis, and methods that do not require them. Among the methods that require probabilistic and statistical analysis, a probabilistic feasibility formulation (Du and Chen, 2000 and Lee et al, 2008), and a moment matching formulation (Parkison et al, 1993) have been proposed. Du and Chen (2000) used a most probable point (MPP)-based importance sampling method to reduce the computational burden associated with the probabilistic feasibility formulation. The moment matching formulation is a simplified approach which requires only the constraints on the first and second moments of the performance function to be satisfied, and assumes that the performance function is normally distributed. A variation of this approach, *the feasible region reduction method* has been described in Park et al (2006), which is more general and does not require the normality assumption. This is a tolerance design method, where width of the feasible space in each direction is reduced by the amount $k\sigma$, where k is a user-defined constant and σ is the standard deviation of the performance function. This method only requires the mean and variance of the performance function.

Methods that do not require probabilistic and statistical analysis are also available, for example, worst case analysis (Parkinson et al, 1993), corner space evaluation (Sundaresan et al, 1995), and manufacturing variation patterns (MVP) (Yu and Ishii, 1998). A comparison study of the different constraint feasibility methods can be found in Du and Chen (2000).

Estimating mean and variance of the performance function

Various methods have been reported in the literature to estimate the mean and standard deviation of the performance function. These methods can be divided into three major classes: (i) Taylor series expansion methods, (ii) sampling-based methods and (iii) point estimate methods (Huang and Du, 2007).

The Taylor series expansion method (Haldar and Mahadevan, 2000; Du and Chen, 2000; and Lee et al, 2001) is a simple approach. However, for a nonlinear performance function, if the variances of the random variables are large, this approximation may result in large errors (Du et al., 2004). Although a second-order Taylor series expansion is generally more accurate than the first-order approximation, it is also computationally more expensive.

Sampling-based methods require information on distributions of the random variables, and are expensive. Efficient sampling techniques such as importance sampling, Latin hypercube sampling, etc. (Robert and Cesalla, 2004) can be used to reduce the computational effort, but are still prohibitive in the context of optimization. Surrogate models (Ghanem and Spanos 1991; Bichon et al, 2008; Cheng and Sandu, 2009) may be used to further reduce computational effort.

In an attempt to overcome the difficulties associated with the computation of derivatives required in Taylor series expansion, Rosenlblueth (1975) proposed a point estimate method to compute the first few moments of the performance function. Different variations of this point estimate method (Hong, 1998; Zhao and Ono, 2000 and Zhao and Ang, 2003) have been studied. Although point estimate methods are easier to implement, the accuracy may be low and may generate points that lie outside the domain of the random variable.

Multi-objective optimization

Robustness-based optimization considers two objectives: optimize the mean of the objective function and minimize its variation. An extensive survey of the multi-objective optimization methods can be found in Marler and Arora (2004). Among the available methods, the weighted sum approach is the most common approach to multi-objective optimization and has been extensively used in robust design optimization (Lee and Park, 2001; Doltsinis and Kang, 2004; Zou and Mahadevan, 2006). The designer can obtain alternative design points by varying the weights and can select the one that offers the best trade-off among multiple objectives. Despite its simplicity, the weighted sum method may not obtain potentially desirable solutions (Park et al, 2006). Another common approach is the ε-constraint method in which one of the objective functions is optimized while the other objective functions are used as constraints. Despite its advantages over weighted sum method in some cases, the ε-constraint method can be computationally expensive for more than two objective functions (Mavrotas, 2009).

Other methods include goal programming (Zou and Mahadevan, 2006), compromise decision support problem (Bras and Mistree, 1993, 1995; Chen et al, 1996), compromise programming (CP) (Zalney, 1973; Zhang, 2003; Chen et al, 1999) and physical programming (Messac, 1996; Messac et al, 2001; Messac and Ismail-Yahaya, 2002; Chen et al, 2000). Each of these methods has its own advantages and limitations.

As discussed in Chapter II, most of the current methods of robust optimization for epistemic uncertainty use non-probabilistic methods to represent epistemic uncertainty. These methods need additional non-probabilistic formulations to incorporate epistemic uncertainty into the robust optimization framework and thus, are computationally expensive. However, if the epistemic uncertainty can be converted to a probabilistic format, the need for these additional formulations is avoidable, and well established probabilistic methods of robust design optimization can be used. Therefore, there is a need for an efficient robust design optimization methodology that deals with both aleatory and epistemic uncertainty. In this chapter, we propose robustness-based design optimization formulations that work under both aleatory and epistemic uncertainty using probabilistic representations of different types of uncertainty. Our proposed formulations deal with both sparse point and interval data without any strict assumption about probability distributions of the random variables.

The performance of robustness-based design can be defined by the mean and variation of the performance function. In our proposed formulations, we obtain the optimum mean value of the objective function (e.g., gross weight) while also minimizing its variation (e.g., standard deviation). Thus, the design will meet target values in terms of both design bounds and standard deviations of design objectives and design variables thereby ensure feasibility robustness.

A Taylor series expansion method is used in this dissertation to estimate the mean and standard deviation of the performance function, which requires means and standard deviations of the random variables. However, with sparse point data and interval data, it is impossible to know the true moments of the data, and there are many possible probability distributions that can represent these data (see Chapter IV). In this chapter, we propose methods for robustness-based design optimization that account for this uncertainty in the moments due to sparse point data and interval data and thereby include epistemic uncertainty into the robust design optimization framework. As collecting more data reduces uncertainty but increases cost, the effect of sample size on the optimality and the robustness of the solution is also studied. A method to determine the optimal sample size for sparse point data that will lead to the minimum scatter on solutions to the design problem is also presented in this chapter.

In some existing methods for robust design under epistemic uncertainty, all the epistemic variables are considered as design variables (Youn et al, 2007). However, if the designer does not have any control on an epistemic variable (e.g., Young's modulus in beam design), considering that variable as a design variable might lead to a solution that could underestimate the design objectives. Therefore, in this chapter, we propose a general formulation for robust design that considers some of the epistemic variables as non-design variables, which leads to a conservative design under epistemic uncertainty. An example of epistemic uncertainty in a design variable is the geometric dimension of a component, whose manufactured value is different from the design value. This difference might be specified as an interval by an expert, or only a few instances of historic values of this difference might be available. Note that the sparse point and/or interval data for the epistemic design variables are used only to estimate the variances; the mean values of such variables are controlled by the design.

Note that the proposed robustness-based design optimization method is general and capable of handling a wide range of application problems under data uncertainty. The proposed methods are illustrated for the conceptual level design process of a two-stage-to-orbit (TSTO) vehicle, where the distributions of the random inputs are described by sparse point and/or interval data.

The rest of the chapter is organized as follows. Section 2 proposes robustness-based design optimization framework for sparse point data and interval data. In Section 3, we illustrate the proposed methods for the conceptual level design process of a TSTO vehicle. Section 4 provides conclusions and suggestions for future work.

2. Proposed methodology

Deterministic design optimization

In a deterministic optimization formulation, all design variables and system variables are considered deterministic. No random variability or data uncertainty is taken into account. The deterministic optimization problem is formulated as follows:

$$\begin{aligned} &\min_{x} \quad f(x) \\ &s.t. \quad LB \le g_i(x) \le UB \quad \text{for all } i \\ &\qquad lb \le x \le ub \end{aligned} \tag{1}$$

where $f(x)$ is the objective function, $\boldsymbol{x}$ is the vector of design variables, $g_i(x)$ is the *i*th constraint, *LB* and *UB* are the vectors of lower and upper bounds of constraints $g_i's$ and *lb* and *ub* are the vectors of lower and upper bounds of design variables.

In practice, the input variables might be uncertain and solutions of this deterministic formulation could be sensitive to the variations in the input variables. Robustness-based design optimization takes this uncertainty into account. The optimal design points obtained using the deterministic method could be used as initial guesses in robustness-based optimization.

Robustness-based design optimization

In the proposed methodology, we use *variance* as a measure of variation of the performance function in order to achieve *objective robustness, the feasible region reduction method* to achieve *feasibility robustness, a first-order Taylor series expansion* to estimate the mean and variance of the performance function, and *a weighted sum method* for the aggregation of multiple objectives. This combination of methods is only used for the sake of illustration. Other approaches can be easily substituted in the proposed methodology. The robustness-based design optimization problem can now be formulated as follows:

$$
\begin{aligned}
&\min_{d} \; f(\mu,\sigma) = (w * \mu_f + v * \sigma_f) \\
&s.t. \quad LB + k\sigma(g_i(d,z)) \le E(g_i(d,z)) \le UB - k\sigma(g_i(d,z)) \;\; \text{for all } i \\
&\qquad lb + k\sigma(x_i) \le d_i \le ub - k\sigma(x_i) \quad \text{for } i = 1,2,...,nrdv \\
&\qquad lb \le d_i \le ub \quad \text{for } i = 1,2,...,nddv
\end{aligned} \tag{2}
$$

where μ_f and σ_f are the mean value and standard deviation of the objective function, respectively; ***d*** is the vector of deterministic design variables as well as the mean values of the uncertain design variables ***x***; *nrdv* and *nddv* are the numbers of the random design variables and deterministic design variables, respectively; and ***z*** is the vector of non-design input random variables, whose values are kept fixed at their mean values as a part of the design. $w \ge 0$ and $v \ge 0$ are the weighting coefficients that represent the relative importance of the objectives μ_f and σ_f in Eq.(2); $g_i(d,z)$ is the *i*th constraint; $E(g_i(d,z))$ is the mean and $\sigma(g_i(d,z))$ is the standard deviation of the *i*th constraint. *LB* and *UB* are the vectors of lower and upper bounds of constraints g_i's; *lb* and *ub* are the vectors of lower and upper bounds of the design variables; $\sigma(x)$ is the vector of standard deviations of the random variables and *k* is some constant. The role of the constant *k* is to adjust the robustness of the method against the level of conservatism of the solution. It reduces the feasible region by accounting for the variations in the design variables and is related to the probability of constraint satisfaction. For example, if a design variable or a constraint function is normally distributed, *k* =1 corresponds to the probability 0.8413, *k* =2 to the probability 0.9772, etc.

Note that the robust design formulation in Eq. (2) is a standard nonlinear multi-objective optimization formulation. The optimality conditions of such a formulation have been extensively described in the literature including Cagan and Williams (1993) and Marler and Arora (2004).

In the proposed formulation, the performance functions considered are in terms of the model outputs. The means and standard deviations of the objective and constraints are estimated by using a first-order Taylor series approximation as follows:

Performance function: $Y = g(X_1, X_2, ..., X_n)$ (3)

First-order approximate mean of y: $E(Y') \approx g(\mu_{X_1}, \mu_{X_2}, ..., \mu_{X_n})$ (4)

First-order variance of y: $Var(Y') \approx \sum_{i=1}^{n} \left(\frac{\partial g}{\partial X_i} \right)^2 Var(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} \frac{\partial g}{\partial X_i} \frac{\partial g}{\partial X_j} Cov(X_i, X_j)$ (5)

The implementation of Eq. (2) requires that variances of the random design variables X_i and the means and variances of the random non-design variables Z_i be precisely known, which is possible only when a large number of data points are available. In practical situations, only a small number of data points may be available for the input variables. In other cases, information about random input variables may only be specified as intervals, as by expert opinion. This is input data uncertainty, causing uncertainty regarding the distribution parameters (e.g., mean and variance) of the inputs X_i and Z_i. Robustness-based optimization has to take this into account. In the following subsections, we propose a new methodology for robustness-based design optimization that accounts for data uncertainty.

2.1 Robustness-based design optimization under data uncertainty

The inclusion of epistemic uncertainty in robust design adds another level of complexity in the design methodology. The design variables ***d*** and/or the input random variables ***z*** in Eq. (2) might have epistemic uncertainty. Since the designer does not have any control on the non-design epistemic variables ***z***, the design methodology has to employ a search among the possible values of such epistemic variables in order to find an optimal solution. In such case, we get a conservative robust design. The robustness-based design optimization problem can now be formulated with the following generalized statement:

$$\begin{aligned}
&\min_{d}\left(\max_{\mu_z} f(\mu,\sigma)=(w*\mu_f+v*\sigma_f)\right) \\
&s.t. \quad LB+k\sigma(g_i(d,\mu_z))\le E(g_i(d,z))\le UB-k\sigma(g_i(d,\mu_z)) \ \text{ for all } i \\
&\qquad lb+k\sigma(x)\le d\le ub-k\sigma(x) \\
&\qquad Z_l\le \mu_z\le Z_u
\end{aligned} \tag{6}$$

where Z_l and Z_u are the vectors of lower and upper bounds of the decision variables μ_z of the inner loop optimization problem.

Note that in this formulation, the outer loop decision variables ***d*** may consist of stochastic design variables as well as epistemic design variables. The outer loop optimization is a design optimization problem, where a robust design optimization is carried out for a fixed set of non-design epistemic variables. The inner loop optimization is the analysis for the non-design epistemic variables, where the optimizer searches among the possible values of the non-design epistemic variables for a conservative solution of the robust design problem.

This nested optimization problem can be decoupled and expressed as:

$$\begin{aligned}
&d^*=\arg\min_{d}\left(w*\mu_f\left(d,\mu_z^*\right)+v*\sigma_f\left(d,\mu_z^*\right)\right) \\
&s.t. \quad LB+k\sigma(g_i(d,\mu_z^*))\le E(g_i(d,z))\le UB-k\sigma(g_i(d,\mu_z^*)) \ \text{ for all } i \\
&\qquad lb+k\sigma(x)\le d\le ub-k\sigma(x)
\end{aligned} \tag{7}$$

$$\mu_z^* = \arg\max_{\mu_z} \left(w * \mu_f\left(d^*, \mu_z\right) + v * \sigma_f\left(d^*, \mu_z\right) \right) \qquad (8)$$

$$s.t. \quad LB + k\sigma(g_i(d^*, \mu_z)) \le E(g_i(d, z)) \le UB - k\sigma(g_i(d^*, \mu_z)) \text{ for all } i$$

$$Z_l \le \mu_z \le Z_u$$

The optimization problems in Eqs. (7) and (8) are solved iteratively until convergence. Note that the first constraint (i.e., the robustness constraint) in Eq. (8) is required to ensure that the optimization is driven by all non-design epistemic variables, because sometimes the objective function may not be a function of all non-design epistemic variables. In cases when the objective function is the function of all non-design epistemic variables, this constraint is not required. Figure 1 illustrates the decoupled approach for robustness-based design optimization under both aleatory and epistemic uncertainty.

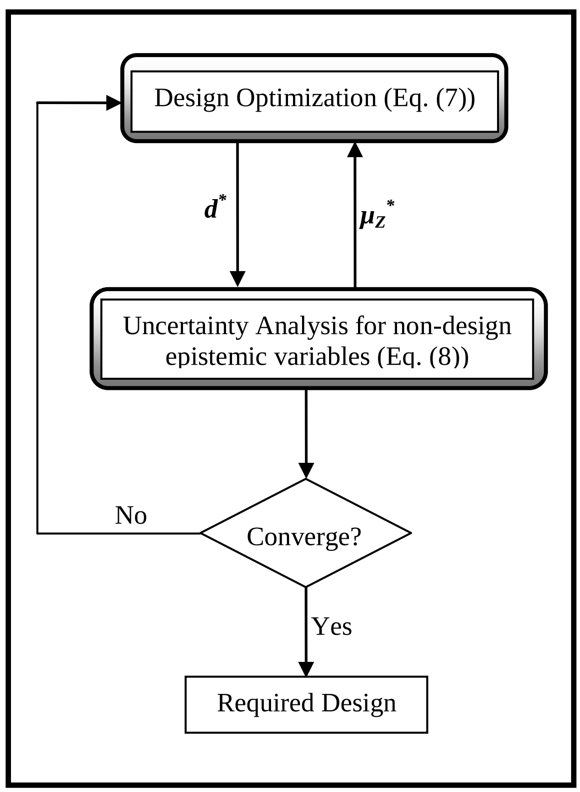


Figure 1: Decoupled approach for robustness-based design optimization

Note that $\boldsymbol{d}^*$ are fixed quantities in the optimization in Eq. (8) and $\boldsymbol{\mu_z}^*$ are the fixed quantities in the optimization in Eq. (7).

2.1.1 Robustness-based design with sparse point data

This section develops a methodology for robustness-based design optimization with sparse point data, using the formulations in Eqs. (7) and (8). It is assumed that only sparse point data are available for the uncertain design variables as well as non-design epistemic variables.

When a variable, either design or non-design, is described by sparse point data, there is uncertainty about the mean and variance calculated from the samples. In the design optimization (Eq. (7)), the mean values of the design variables (either aleatory or epistemic) are controlled by the given design bounds. As in design optimization under aleatory uncertainty only, here also it is assumed that the variances of the epistemic design variables do not change as their mean values change. However, since the mean values of the non-design variables cannot be controlled in the design optimization, the proposed robustness-based design optimization methodology accounts for the uncertainty about mean values of such epistemic variables through the optimization in Eq. (8).

The constraints on the non-design epistemic variables in Eq. (8) are implemented through the construction of confidence intervals about mean values. As these variables are described by the sparse point data, it is possible that the underlying distributions of the variables might have major deviations from normality. Therefore, we have used the Johnson's modified *t* statistic (Johnson, 1978) to construct the confidence bounds on mean values of the non-design epistemic variables as follows:

$$Z_l = \bar{z} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} + \frac{\mu_3}{6s^2 n}$$
$$Z_u = \bar{z} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} + \frac{\mu_3}{6s^2 n} \qquad (9)$$

where $\bar{z}$ is the vector of means of the epistemic variables, s is the vector of standard deviations, n is the sample size of the sparse point data, μ_3 is the third central moment and $t_{\alpha/2,n-1}$ is obtained from the Student t distribution at (n-1) degrees of freedom and α significance level. This modified statistic takes into account the skewness of the distribution and thus provides a better estimate of the confidence bound in the presence of limited data.

The proposed robustness-based design optimization methodology accounts for the uncertainty about the variances for all epistemic variables by first estimating confidence bounds on variances and then solving the optimization formulations in Eqs. (7) and (8) using the upper bound variances for the input random variables x_i and z_i. Solving the optimization formulations in Eqs. (7)-(8) using the upper bound variances for all the epistemic variables ensures that the resulting solution is least sensitive to the variations in the input random variables.

The chi-square distribution is a good assumption for the distribution of the variance, especially if the underlying population is normal. The two-sided (1-α) confidence interval for the population variance σ^2 can be expressed as (Haldar and Mahadevan, 2000):

$$\left[\frac{(n-1)s^2}{c_{1-\alpha/2,n-1}};\frac{(n-1)s^2}{c_{\alpha/2,n-1}}\right] \qquad (10)$$

where n is the sample size, s is the sample standard deviation of sparse point data, and $c_{\alpha/2,n-1}$ is obtained from the chi-square distribution at (n-1) degrees of freedom and α significance level. Note that Eq. (10) can still be used to obtain approximate confidence bounds for variance if the underlying population is not normal. However, in such cases, other approximation methods (Bonett, 2006; Cojbasic and Tomovic, 2007) can be used to obtain more reliable estimates of confidence bounds.

The optimization formulation shown in Eqs. (7)-(8) involves aggregation of multiple objectives. In the proposed formulations, the aggregate objective function consists of two types of objectives, expectation and standard deviation of model outputs. Since different objectives have different magnitudes, a scaling factor has to be used in the formulation.

2.1.2 Determination of optimal sample size for sparse point data

The optimal solutions depend on the sample size of the sparse data as will be discussed in Section 3.1. Therefore, it is of interest to determine the optimal sample size of the sparse data that leads to the solution of the design problem that is least sensitive to the variations of design variables. This will facilitate resource allocation decision for data collection. The following two optimization formulations are solved iteratively until convergence for the optimal sample sizes of the epistemic design variables (n_d^*) and epistemic non-design variables(n_e^*).The formulations in Eq. (11)-(12) are the weighted sum formulations of a three-objective optimization problem, where the first and second objectives are the mean and standard deviation of GW respectively and the third objective is the total cost of obtaining samples for all the random variables.

$$\left[d^*, n_d^*\right] = \arg\min_{d, n_d} \left(w * E(g_i(d, \mu_z^*)) + v * \sigma(g_i(d, \mu_z^*), n_d, n_e^*) + (1 - w - v) * \left(\sum_{j=1}^{m} n_{d_j} c_{d\,j} + \sum_{j=1}^{q} n_{e_j}^* c_{e\,j} \right) \right)$$

$$\begin{aligned} s.t. \quad & LB + k\sigma(g_i(d, \mu_z^*), n_d, n_e^*) \le E(g_i(d, z)) \le UB - k\sigma(g_i(d, \mu_z^*), n_d, n_e^*) \ \text{ for all } i \\ & lb + k\sigma(x, n_d) \le d \le ub - k\sigma(x, n_d) \\ & \sum_{j=1}^{m} n_{d_j} c_{d\,j} + \sum_{j=1}^{q} n_{e_j}^* c_{e\,j} \le C \\ & n_{d\,j} \le b_{d\,j} \quad \text{for all j} \end{aligned} \tag{11}$$

$$\left[\mu_z^*, n_e^*\right] = \arg\max_{\mu_z, n_e} \left(w * E(g_i(d^*, \mu_z)) + v * \sigma(g_i(d^*, \mu_z), n_d^*, n_e) + (1 - w - v) * \left(\sum_{j=1}^{m} n_{d_j}^* c_{d\,j} + \sum_{j=1}^{q} n_{e_j} c_{e\,j} \right) \right)$$

$$\begin{aligned} s.t. \quad & LB + k\sigma(g_i(d^*, \mu_z), n_d^*, n_e) \le E(g_i(d, z)) \le UB - k\sigma(g_i(d^*, \mu_z), n_d^*, n_e) \ \text{ for all } i \\ & Z_l(n_e) \le \mu_z \le Z_u(n_e) \\ & \sum_{j=1}^{m} n_{d_j}^* c_{d\,j} + \sum_{j=1}^{q} n_{e_j} c_{e\,j} \le C \\ & n_{e_j} \le b_{e\,\mathrm{j}} \quad \text{for all j} \end{aligned} \tag{12}$$

where $w \ge 0$ and $v \ge 0$ are the weighting coefficients that represent the relative importance of the objectives; n_{d_j} and n_{e_j} are the sample sizes and b_{d_j} and b_{e_j} are the maximum sample size possible for the *j*th design and non-design random variables, respectively. *m* and *q* are the number of design and non-deisgn random variables, respectively. c_{d_j} and c_{e_j} are the cost of obtaining one sample for the *j*th random design and non-design variables, respectively and *C* is the total cost allocated for obtaining samples for all the random variables. Note that as in Eq. (8), the robustness constraint in Eq. (12) is only required if the objective function is not a function of all non-design epistemic variables. The optimization formulation presented above is a mixed-integer nonlinear problem. A relaxed problem is solved in Section 3.

2.1.3 Robustness-based design with interval data

This section develops a methodology for robustness-based design optimization with interval data, using the formulations in Eqs. (7) and (8). In this case, the only information available for one or more input random variables is in the form of single interval or multiple interval data.

The methodology for robustness-based design optimization with interval data is similar to sparse point data as described in Section 2.1.1. However, the estimation of mean values and variances for interval data is not straightforward. For interval data, the moments (e.g., mean and variance) are not single-valued, rather only bounds can be given (see Chapter IV). We have proposed methods to compute the bounds of moments for both single and multiple interval data in Chapter IV. Once the bounds on the mean and variance of interval data are estimated, we use the upper bounds of sample variance to solve the formulations of robust design under uncertainty represented through single interval or multiple interval data. Therefore, the resulting solution becomes least sensitive to the variations in the uncertain variables.

For non-design epistemic variables described by interval data, the constraints on the decision variables in Eq. (8) are implemented through estimating the bounds of the means by the methods as described in Chapter IV.

Once the bounds on the mean and variance of interval data are estimated by the methods described in Chapter IV, we can now use these bounds to solve the formulations of robustness-based design optimization under uncertainty represented through single interval or multiple interval data. In the following section, we illustrate our proposed formulations for robustness-based design optimization with both sparse point and interval data.

3. Example Problem

In this section, the proposed methods are illustrated for the conceptual level design process of a TSTO vehicle as discussed in Chapter I. The TSTO concept vehicle is shown in Figure 1 of Chapter I. The analysis process of a TSTO vehicle is illustrated in Figure 2 of Chapter I.

The analysis outputs (performance functions) are Gross Weight (GW), Engine Weight (EW), Propellant Fraction Required (PFR), Vehicle Length (VL), Vehicle Volume (VV) and Body Wetted Area (BWA). Each of the analysis outputs is approximated by a second-order response surface and is a function of the random design variables Nozzle Expansion Ratio (ExpRatio), Payload Weight (Payload), Separation Mach (SepMach), Separation Dynamic Pressure (SepQ), Separation Flight Path Angle (SepAngle), and Body Fineness Ratio (Fineness). Each of the random variables is described by either sparse point data or interval data.

The objective is to optimize an individual analysis output (e.g., Gross Weight) while satisfying the constraints imposed by each of the design variables as well as all the analysis outputs. We note here that we have assumed independence among the uncertain input variables and thereby ignored the covariance terms in Eq. (5) to estimate the variance of the performance function in each of the following examples. The numerical values of the design bounds for the design variables and analysis outputs are given in Tables 1 and 2, respectively.

Table 1: Design bounds for the design variables

Design Variable	lb	ub
ExpRatio	40	150
Payload	8000	40000
SepMach	7	12
SepQ	40	200
SepAngle	7	12
Fineness	4	6

Table 2: Design bounds for the analysis outputs

Analysis output	LB	UB
GW	0	100e+005
EW	0	100e+005
PFR	0.4	0.95
VL	0	100e+002
VV	0	100e+003
BWA	0	100e+005

3.1. Robustness-based design optimization with sparse point data

The methodology proposed in Section 2.1.1 is illustrated here for the TSTO problem. It is assumed that all the input variables ***x*** are described by sparse point data as given in Table 3. For this example, the input variable SepQ is assumed to be a non-design epistemic variable and all the remaining variables are assumed to be design variables. The design bounds for the respective design variables and the analysis outputs are given in Tables 1 and 2.

Table 3: Sparse Point Data for the random input variables

Sample	ExpRatio	Payload	SepMach	SepQ	SepAngle	Fineness
01	85.23	2.8952e+004	10.85	115.38	9.12	4.07
02	82.25	2.9747e+004	10.56	111.63	9.49	4.02
03	88.79	2.6638e+004	10.93	118.57	9.85	4.47
04	83.93	2.8356e+004	10.70	111.60	9.87	4.15
05	80.67	2.7193e+004	10.58	100.34	9.27	4.15

06	91.32	2.9168e+004	10.82	102.42	9.21	4.17
07	83.64	2.8844e+004	10.88	117.25	9.57	4.23
08	86.64	2.5836e+004	10.99	109.69	9.64	4.32
09	90.32	2.9310e+004	10.00	116.90	9.42	4.01
10	85.39	2.9949e+004	10.87	104.19	9.21	4.42

The design problem becomes:

$$
\begin{aligned}
& d^* = \arg\min_d \left(w * E(GW) + (1-w) * \sigma(GW) \right) \\
& s.t. \quad LB_1 + k\sigma(GW) \le E(GW) \le UB_1 - k\sigma(GW) \\
& \quad LB_2 + k\sigma(EW) \le E(EW) \le UB_2 - k\sigma(EW) \\
& \quad LB_3 + k\sigma(PFR) \le E(PFR) \le UB_3 - k\sigma(PFR) \\
& \quad LB_4 + k\sigma(VL) \le E(VL) \le UB_4 - k\sigma(VL) \\
& \quad LB_5 + k\sigma(VV) \le E(VV) \le UB_5 - k\sigma(VV) \\
& \quad LB_6 + k\sigma(BWA) \le E(BWA) \le UB_6 - k\sigma(BWA) \\
& \quad lb + k\sigma(x) \le d_i \le ub - k\sigma(x) \qquad \text{for } i = 1,2,...,5
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
& \mu_z^* = \arg\max_{\mu_z} \left(w * E(GW) + (1-w) * \sigma(GW) \right) \\
& s.t. \qquad Z_l \le \mu_{z_i} \le Z_u \qquad \text{for } i = 1
\end{aligned} \tag{14}
$$

where the bounds Z_l and Z_u for the mean of the non-design epistemic variable SepQ are calculated by Eq. (9) as given in Section 2.1.1. Note that in Eq. (14), we do not use the robust design constraints, since the objective function in this case is a function of all non-design epistemic variables.

As mentioned earlier in Section 2, $w \ge 0$ is the weight parameter that represents the relative importance of the objectives and k is a constant that adjusts the robustness of the method against the level of conservatism of the solution. In this dissertation, k is assumed to be unity.

Variances of the random variables ***x*** and ***z*** are estimated as single point values. Confidence intervals for the variances are estimated for each random variable described by the sparse point data. The weight parameter w is varied (from 0 to 1) and the optimization problem in Eqs. (13)-(14) are solved iteratively until convergence by the Matlab solver 'fmincon' for different sample sizes (n) of the sparse point data. The formulations are relaxed by assuming that standard deviations estimates of the variables do not change significantly as the sample size changes. Therefore, the same standard deviations as estimated from the data given in Table 3 are used in each case. As the sample size (n) changes, the confidence bounds on the variance also change (see Eq. (10)). In each case, the optimization problems converged in less than 5 iterations. Here, 'fmincon' uses a sequential quadratic programming (SQP) algorithm. The estimate of the Hessian of the Lagrangian is updated using the BFGS formula at each iteration. The convergence properties of SQP have been discussed by many authors including Fletcher (1987) and Panier and Tits (1993).

The solutions are obtained by solving the problem using the upper confidence bound for the variances of the random variables ***x*** and ***z***. The solutions are presented in Figure 2.

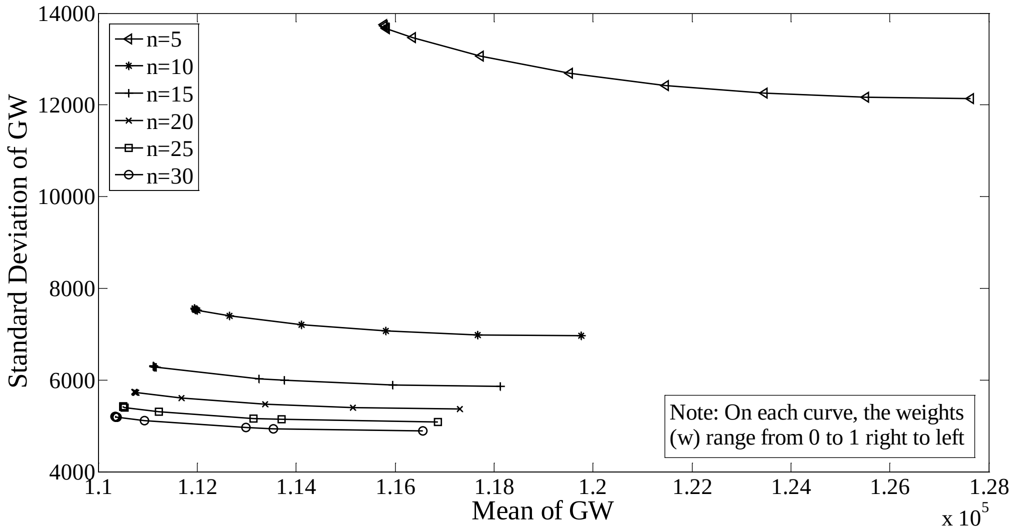


Figure 2: Robustness-based design optimization with sparse data for different sample sizes (*n*)

It is seen in Figure 2 that the solutions become more conservative (i.e., the mean and standard deviation of GW assume higher values) as we add uncertainty to the design problem. As gathering more data reduces data uncertainty, the solutions become less sensitive (i.e., the standard deviation of GW assumes lower value) to the variations of the input random variables as the sample size (*n*) increases. Also, looking at the mean of GW, it is seen that as the uncertainty decreases with sample size, the optimum mean weight required is less.

3. 2 Determination of optimal sample size for sparse point data

The optimal sample size formulations are illustrated here for the TSTO design problem. The formulations are relaxed by assuming that standard deviations of the data do not change significantly as sample size changes. To make the problem simpler, we first relax the integer requirement on the optimal sample size *n* and then round off the solution for *n* to the nearest integer value. The input variable SepQ is assumed to be a non-design epistemic variable and all the remaining variables are assumed to be design variables. The design bounds for the respective design variables and the analysis outputs remain the same as in Tables 3 and 4.

Therefore, the design problem becomes as follows:

$$\begin{aligned}
&\left[d^*, n_d^*\right] = \arg\min_{d, n_d} \; \left(w * E(GW) + v * \sigma(GW) + (1 - w - v) * \left(5n_{d1} + 10n_{d2} + 5n_{d3} + 5n_{d4} + 4n_{d5} + 6n_e^*\right)\right) \\
&s.t. \quad LB_1 + \sigma(GW) \le E(GW) \le UB_1 - \sigma(GW) \\
&\qquad LB_2 + \sigma(EW) \le E(EW) \le UB_2 - \sigma(EW) \\
&\qquad LB_3 + \sigma(PFR) \le E(PFR) \le UB_3 - \sigma(PFR) \\
&\qquad LB_4 + \sigma(VL) \le E(VL) \le UB_4 - \sigma(VL) \\
&\qquad LB_5 + \sigma(VV) \le E(VV) \le UB_5 - \sigma(VV) \\
&\qquad LB_6 + \sigma(BWA) \le E(BWA) \le UB_6 - \sigma(BWA) \\
&\qquad lb + k\sigma(\underline{x}) \le d_i \le ub - k\sigma(\underline{x}) \qquad \text{for all } i = 1,2,\ldots,5 \\
&\qquad 5n_{d1} + 10n_{d2} + 5n_3 + 5n_{d3} + 4n_{d5} + 6n_e^* \le 1050 \\
&\qquad n_{d_j} \le 30 \quad \text{for } \; \mathrm{j} = 1,2,\ldots,5
\end{aligned} \tag{15}$$

$$\begin{aligned}
&\left[\mu_z^*, n_e^*\right] = \arg\max_{\mu_z, n_e} \; \left(w * E(GW) + v * \sigma(GW) + \left(1 - w - v\right) * \left(5n^*_{d1} + 10n^*_{d2} + 5n^*_{d_3} + 5n^*_{d3} + 4n^*_{d5} + 6n^e\right)\right) \\
&s.t. \qquad Z_l\left(n_e\right) \le \mu_z \le Z_u\left(n_e\right) \\
&\qquad\quad 5n^*_{d1} + 10n^*_{d2} + 5n^*_{d_3} + 5n^*_{d3} + 4n^*_{d5} + 6n^e \le 1050 \\
&\qquad\quad n_{e_j} \le 30 \quad \text{for } \; \mathrm{j} = 1
\end{aligned} \tag{16}$$

We have solved this problem for different combinations of weights *w* and *v* and the optimal solutions are presented in Table 4. In each case, the optimization problems converged in less than 4 iterations.

Table 4: Objective function values at optimal solutions and optimal sample sizes

Weights			Objective function Value			Optimal Sample Sizes					
w	v	$1-w-v$	Mean GW	Std GW	Total Cost	n_{d1}	n_{d2}	n_{d3}	n_{d4}	n_{d5}	n_e
0	0	1	1.6118e+005	6.3732e+004	455.3008	5	10	15	8	9	30
0.6	0.2	0.2	1.4684e+005	5.3219e+004	539.8948	6	10	30	8	10	30
0.5	0.4	0.1	1.4878e+005	5.0526e+004	593.6961	7	10	30	14	15	30
0.5	0.5	0	1.5143e+005	4.7604e+004	886.9363	25	25	30	30	30	15

It is seen in Table 6 that the total cost incurred in obtaining samples is the minimum when we solve the problem giving the maximum importance on the total cost. In this case, we get the most conservative robust design i.e., the mean and the standard deviation of GW assume the maximum of all possible values. Note that the optimal sample size required is also the minimum in this case. As we give more importance on the mean and standard deviation of GW, the total cost and also the optimal sample size increase with a decrease in both the mean and standard deviation of GW.

3. 3 Robustness-based design optimization with sparse point and interval data

The methodology proposed in Section 2.1 is illustrated here for the same TSTO problem. Here, it is assumed that the design variable ExpRatio is described by sparse point data as given in Table 3, the design variable Payload is described by multiple interval data as given in Table 5 and the design variables SepMach and SepQ are described by single interval data as given in Table 6. The non-design epistemic variables SepAngle and Fineness are described by the sparse point data (as given in Table 3) and the single interval data (as given in Table 6), respectively. The design bounds for the respective design variables and the analysis outputs remain the same as in Tables 3 and 4.

Table 5: Multiple Interval Data for the random input variables

Payload	[25000, 28000], [26000, 29000], [25000, 29000], [26000, 30000], [25000, 30000]

Table 6: Single Interval Data for the random input variables

SepMach	[9, 10]
SepQ	[100, 120]
Fineness	[4, 4.5]

The design problem is now formulated as follows:

$$
\begin{aligned}
d^* &= \arg\min_{d} \left(w * E(GW) + (1-w) * \sigma(GW) \right) \\
s.t. \quad & LB_1 + k\sigma(GW) \le E(GW) \le UB_1 - k\sigma(GW) \\
& LB_2 + k\sigma(EW) \le E(EW) \le UB_2 - k\sigma(EW) \\
& LB_3 + k\sigma(PFR) \le E(PFR) \le UB_3 - k\sigma(PFR) \\
& LB_4 + k\sigma(VL) \le E(VL) \le UB_4 - k\sigma(VL) \\
& LB_5 + k\sigma(VV) \le E(VV) \le UB_5 - k\sigma(VV) \\
& LB_6 + k\sigma(BWA) \le E(BWA) \le UB_6 - k\sigma(BWA) \\
& lb + k\sigma(x) \le d_i \le ub - k\sigma(x) \qquad \text{for } i = 1,2,3,4
\end{aligned} \tag{15}
$$

$$
\begin{aligned}
\mu_z^* &= \arg\max_{\mu_z} \left(w * E(GW) + (1-w) * \sigma(GW) \right) \\
s.t. \quad & Z_l \le \mu_{z_i} \le Z_u \qquad \text{for } i = 1,2
\end{aligned} \tag{16}
$$

where the bounds Z_l and Z_u for the mean value of the non-design epistemic variable SepAngle are calculated by Eq. (9) as given in Section 2.1.1 and those for the epistemic variable Fineness are calculated by the method described in Section 2.1.3. Note that in Eq. (16), we do not use the robust design constraints, since the objective function in this case is a function of all non-design epistemic variables.

Variances of the random variables ExpRatio and SepAngle are estimated as single point values. Confidence intervals for the variances are estimated for each random variable described by sparse point data. Bounds on the variances of the random variables SepMach, SepQ, Fineness, and Payload are estimated by the methods described in Sections 2.1.3. The free parameter *w* is varied (from 0 to 1) and the optimization problems in Eqs. (15) and (16) are solved iteratively until convergence. In each case, the optimization problems converged in less than 5 iterations. The solutions are obtained by solving the problems using the upper confidence bound on sample variance for the random variables ExpRatio and SepAngle, and the upper bound on sample variances for the random variables Payload, SepMach, SepQ and Fineness. The solutions are presented in Figure 3.

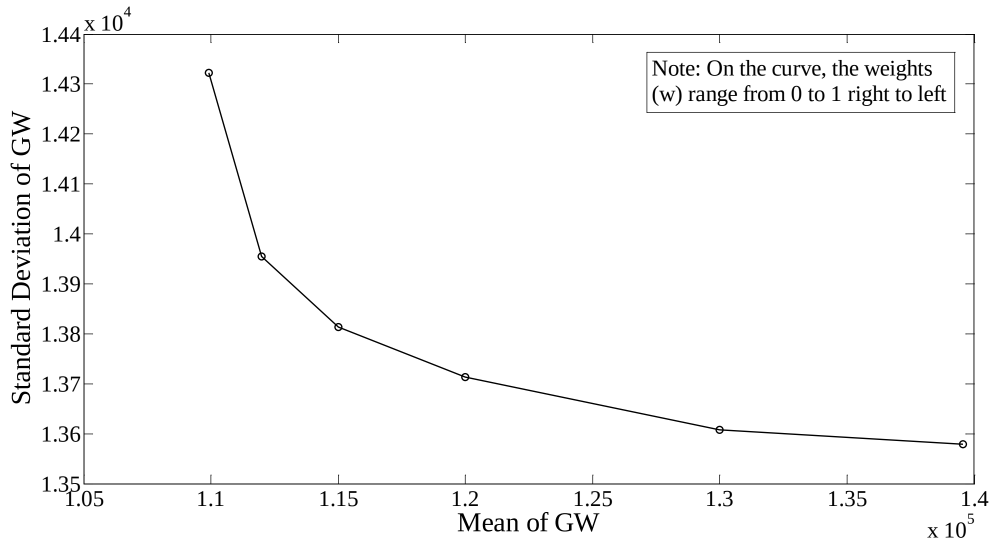


Figure 3: Robustness-based design optimization with sparse point and interval data

Figure 3 shows the solutions of the conservative robust design in presence of uncontrollable epistemic uncertainty described through mixed data i.e., both sparse point data and interval data, which is seen frequently in many engineering applications.

4. Conclusion

This chapter proposed several formulations for robustness-based design optimization under data uncertainty. Two types of data uncertainty – sparse point data and interval data – are considered. The proposed formulations are illustrated for the upper stage design problem of a TSTO space vehicle. A decoupled approach is proposed in this chapter to un-nest the robustness-based design from the analysis of non-design epistemic variables to achieve computational efficiency. As gathering more data reduces uncertainty but increases cost, the effect of sample size on the optimality and the robustness of the solution is also studied. This is demonstrated by numerical examples, which suggest that as the uncertainty decreases with sample size, the resulting solutions become more robust. We have also proposed a formulation to determine the optimal sample size for sparse point data that leads to the solution of the design problem that is least sensitive (i.e., robust) to the variations of design variables. In this chapter, we have used the weighted sum approach for the aggregation of multiple objectives and to examine the trade-offs among multiple objectives. Other multi-objective optimization techniques can also be explored within the proposed formulations.

The major advantage of the proposed methodology is that unlike existing methods, it does not use separate representations for aleatory and epistemic uncertainties and does not require nested analysis. Both types of uncertainty are treated in a unified manner using a probabilistic format, thus reducing the computational effort and simplifying the optimization problem. The results regarding robustness of the design versus data size are valuable to the decision maker. The design optimization procedure also optimizes the sample size, thus facilitating resource allocation for data collection efforts. Due to the use of a probabilistic format to represent all the uncertain variables, the proposed robustness-based design optimization methodology facilitates the implementation of multidisciplinary robustness-based design optimization, which is a challenging problem in presence of epistemic uncertainty.

CHAPTER VIII

RELIABILITY-BASED DESIGN OPTIMIZATION (RBDO) UNDER EPISTEMIC UNCERTAINTY

1. Introduction

This chapter proposes formulations and algorithms for reliability-based design optimization (RBDO) under both aleatory uncertainty (i.e., natural or physical variability) and epistemic uncertainty (i.e., imprecise probabilistic information). The proposed formulations specifically deal with epistemic uncertainty arising from sparse point data and interval data. An efficient decoupled approach is proposed that un-nests the design analysis from the epistemic analysis. The proposed methods are illustrated through an example problem.

As mentioned in Chapter II, most of the existing methods are based on non-probabilistic theory. Many of these methods need additional non-probabilistic formulations to incorporate epistemic uncertainty into the design optimization framework, which may be computationally expensive. However, if the epistemic uncertainty can be converted to a probabilistic format, the need for these additional formulations is avoidable, and well established probabilistic methods of RBDO can be used. Therefore, there is a need for an efficient RBDO methodology that deals with both aleatory and epistemic uncertainty.

The contribution of this chapter is to develop a methodology for RBDO that includes both aleatory and epistemic uncertainty. This chapter specifically focuses on epistemic uncertainty arising from *sparse point data* and *interval data*. This chapter proposes an efficient decoupled approach that un-nests the design analysis from the epistemic analysis

The rest of the chapter is organized as follows. Section 2 proposes an RBDO framework that considers sparse point data and interval data for the random variables. Section 3 illustrates the proposed methods with an example problem. Section 4 provides conclusions and suggestions for future work.

2. RBDO for single discipline systems

2.1 Deterministic design optimization

In a deterministic optimization formulation, all design variables are considered deterministic. No random variability or data uncertainty is taken into account. The deterministic optimization problem is formulated as follows:

$$\begin{aligned} &\min_{x} \; f(x) \\ &s.t. \quad g_i(x) \leq 0 \quad \text{for all } i \\ &\qquad lb \leq x \leq ub \end{aligned} \tag{1}$$

where $f(x)$ is the objective function, $\boldsymbol{x}$ is the vector of design variables, $g_i(x)$ is the ith constraint and *lb* and *ub* are the vectors of lower and upper bounds of design variables.

In practice, the input design variables might be uncertain and solutions of this deterministic formulation could be sensitive to the uncertainty of input design variables. Reliability-based design optimization (RBDO) takes this uncertainty into account. The optimal design points obtained using the deterministic method could be used as initial guess in RBDO.

2.2 Reliability-based design optimization

A simple, typical RBDO formulation with only component level reliability constraints is as follows:

$$\begin{aligned} &\min \quad f(d,Z) \\ &s.t. \quad p_{f_i} = P\left(g_i(X,Z) \le 0\right) < p_i \quad \text{for } i = 1,2,\ldots,k \end{aligned} \tag{2}$$

where $f(d,Z)$ is the objective function, $\boldsymbol{d}$ is a set of design variables, $\boldsymbol{Z}$ is a set of input random variables, and p_i could be ith threshold failure probability. The vector $\boldsymbol{d}$ may include both deterministic design variables as well as distribution parameters of random design variables $\boldsymbol{x}$. Note that in RBDO, the objective function value is the nominal value, which is estimated at the mean values of the random variables $\boldsymbol{x}$ and $\boldsymbol{z}$.

RBDO methods fall into three groups depending on how reliability analysis is incorporated into the optimization process. Tu et al (2001) referred to the RBDO methods that use the reliability index directly as a reliability index approach (RIA) and to those based on quantile functions of the probability distributions as the performance measure approach (PMA). The RIA uses a direct FORM, whereas the PMA uses an inverse FORM for reliability analysis. Nested algorithms, which were used before the 1990s, include a full reliability analysis at every step of the design optimization algorithm. It is well known that nesting these two procedures results in a large number of function evaluations, and studies performed by Agarwal and Renaud (2004), Liang et al (2004), Du and Chen (2004), and Yang and Gu (2004) have confirmed that nested methods require many more function evaluations than RBDO methods in which the reliability analysis loop is either decoupled or eliminated via single loop methods. To reduce the computational expense associated with nested methods, many researchers have developed single-loop approaches to RBDO (Madsen and Hansen, 1992; Chen et al, 1997; Wang and Kodiyalam , 2002; Agarwal and Renaud, 2004). The methodologies are focused on removing the inner reliability analysis loop by making the optimality conditions of either FORM or inverse FORM constraints in the optimization loop. Although a number of RBDO studies have focused on developing computationally efficient methods to solve Eq. (2), a very few methods exist for reliability-based design under epistemic uncertainty as mentioned in Section 1. The focus of this chapter is not on efficiency, but on the inclusion of epistemic uncertainty in the design optimization. Therefore, in developing the methodology for RBDO under epistemic uncertainty, we use the classical nested loop RBDO formulation. In this formulation, the reliability analysis required for evaluating the reliability constraints is done inside the RBDO framework using direct FORM.

The FORM estimates the failure probability as $P_f = \Phi(-\beta)$ where $\mathbf{\Phi}$ is the cumulative distribution function (CDF) for the standard normal probability distribution and β is the minimum distance from the origin to the limit state in the uncorrelated reduced normal space (Hasofer and Lind, 1974). The limit state function g is derived from a system performance criterion and formulated such that $g < 0$ indicates failure. The minimum distance point on the limit state is referred to as the most probable point (MPP), and β is referred to as the reliability index. The FORM method is able to handle correlated, non-normal random variables and nonlinear limit states; however, the probability estimate is based on a first-order approximation of the limit state at the MPP. The following formulation is used to estimate the failure probability:

$$\min \beta = \sqrt{(Y)^T (Y)} \qquad (3)$$
$$s.t. \;\; g_Y(Y) = 0$$

In Eq. (3), ***Y*** denotes all the random variables in uncorrelated standard normal space. Function g_Y is transformed functions such that $g_Y(Y) = g\left(T^{-1}(x)\right)$ where T is the transformation function from original space, ***x***, to standard normal space ***Y***. For more details about the implementation of FORM, the reader is referred to Ditlevsen and Madsen (1979), Haldar and Mahadevan (2000), and Nowak and Collins (2000).

In the following section, we develop the methodology for RBDO under epistemic uncertainty for single discipline problems.

2.3 RBDO under epistemic uncertainty

The inclusion of epistemic uncertainty in RBDO adds another level of complexity in the design methodology. The design variables ***d*** and/or the input random variables ***Z*** in Eq. (2) might have epistemic uncertainty. Since the designer does not have any control on the non-design epistemic variables, the RBDO methodology has to employ a search among the possible values of such epistemic variables in order to find an optimal solution. In such case, we get a conservative reliability-based design. The RBDO problem can now be formulated with the following generalized statement:

$$\min_d \left(\max_{\mu_z} f(d, Z) \right)$$
$$s.t. \qquad p_{f_i} = P\left(g_i(X, Z) \le 0\right) < p_i \quad \text{for } i = 1,2,\ldots,k \qquad (4)$$
$$Z_l \le \mu_z \le Z_u$$

where Z_l and Z_u are the vectors of lower and upper bounds of the decision variables μ_z of the inner loop optimization problem.

Note that in this formulation, the outer loop decision variables d may consist of stochastic design variables as well as epistemic design variables. The outer loop optimization is a design optimization problem, where an RBDO is carried out for a fixed set of non-design epistemic variables. The inner loop optimization is the analysis for the non-design epistemic variables, where the optimizer searches among the possible values of the non-design epistemic variables for a conservative solution of the RBDO.

This nested optimization problem can be decoupled and expressed as:

$$\begin{aligned} & d^* = \arg\min_{d} \left(f\left(d, \mu_z^*\right)\right) \\ & s.t. \qquad p_{f_i} = P\left(g_i\left(X, \mu_z^*\right) \le 0\right) < p_i \quad \text{for } i = 1,2,\ldots,k \end{aligned} \tag{5}$$

$$\begin{aligned} & \mu_z^* = \arg\max_{\mu_z} \left(f\left(d^*, \mu_z\right)\right) \\ & s.t. \quad p_{f_i} = P\left(g_i\left(X^*, \mu_z\right) \le 0\right) < p_i \quad \text{for } i = 1,2,\ldots,k \\ & \qquad Z_l \le \mu_z \le Z_u \end{aligned} \tag{6}$$

The optimization problems in Eqs. (5) and (6) are solved iteratively until convergence. Note that the first constraint (i.e., the reliability constraint) in Eq. (6) is required to ensure that the optimization is driven by all non-design epistemic variables, because sometimes the objective function may not be a function of all non-design epistemic variables. In cases when the objective function is the function of all non-design epistemic variables, this constraint is not required.

Since Eq. (5) is solved with a fixed set of non-design epistemic variables, Eq. (5) is equivalent to an RBDO problem under aleatory uncertainty alone. Eq. (6) is referred to as uncertainty analysis for the non-design epistemic variables throughout this dissertation. The RBDO formulations presented above are general and can handle all varieties of design and non-design variables, such as one or more design or non-design variables being deterministic, aleatory or epistemic. Since Eq. (5) is equivalent to traditional RBDO under aleatory uncertainty, it can accommodate both deterministic and aleatory design variables as well as both deterministic and aleatory non-design variables. Eq. (5) also accommodates epistemic design variables. The propose methodology accommodates non-design epistemic variables by employing a search among the possible values of non-design epistemic variables through the formulation in Eq. (6).

RBDO with sparse data

This section develops a methodology for RBDO with sparse point data, using the formulations in Eqs. (5) and (6). It is assumed that only sparse point data are available for some of the design variables as well as non-design epistemic variables.

When a variable, either design or non-design, is described by sparse point data, there is uncertainty about the mean and variance calculated from the samples. In the design optimization (Eq. (5)), the mean values of the design variables (either aleatory or epistemic) are controlled by the given design bounds. However, since the mean values of the non-design variables cannot be controlled in the design optimization, the proposed RBDO methodology accounts for the uncertainty about mean values of such epistemic variables through the optimization in Eq. (6).

The constraints on the non-design epistemic variables in Eq. (6) are implemented through the construction of confidence intervals about mean values using Eq. (9) of Chapter VII.

The proposed RBDO methodology accounts for the uncertainty about the variances for all epistemic variables by first estimating confidence bounds on variances and then solving the optimization formulations in Eqs. (5) and (6) using the upper bound variances for the input random variables x_i and z_i. Solving the optimization formulations in Eqs. (5)-(6) using the upper bound variances for all the epistemic variables ensures that the resulting solution is least sensitive to the variations in the input random variables. The confidence bounds on variances are estimated using Eq. (10) of Chapter VII.

RBDO with interval data

This section develops a methodology for RBDO with interval data, using Eqs. (5) and (6). In this case, the only information available for one or more input random variables is in the form of single interval or multiple interval data.

The methodology for RBDO with interval data is similar to sparse point data as described earlier. However, the estimation of mean values and variances for interval data is not straightforward. For interval data, the moments (e.g., mean and variance) are not single-valued, rather only bounds can be given (see Chapter IV). We have proposed methods to compute the bounds of moments for both single and multiple interval data in Chapter IV. Once the bounds on the mean and variance of interval data are estimated, we use the upper bounds of the variances to solve the formulations of RBDO under epistemic uncertainty in Eqs. (5) and (6). Therefore, the resulting solution becomes least sensitive to the variations in the uncertain variables.

For non-design epistemic variables described by interval data, the constraints on the decision variables in Eq. (6) are implemented through estimating the bounds on the mean values by the methods as described in Chapter IV.

In the following section, the proposed RBDO formulations are illustrated for a Shaft-Gear Assembly.

3. Numerical Example

Shaft-Gear Assembly

This problem is adapted from Mahadevan and Rebba (2006), and modified in this example to include epistemic uncertainty. Consider a mechanical drive shaft assembled into a press-fit gear wheel as shown in Figure 1. The objective is to determine the radii of the solid shaft R and the gear wheel R_0 such that the assembly meets the design torque requirements reliably without slipping at the fit interface (Cruse, 1997). The interface length L is known and the interference fit tolerated in this assembly Δ is also deterministic. The maximum torque T that can be transmitted by the assembly (fit) without any slippage can be given in terms of the coefficient of friction η at the fit, interface length L (or gear wheel width in this case), shaft radius R, and the interference pressure p as (Shigley et al, 2004)

$$T = 2\pi\eta pLR^2 \qquad (7)$$

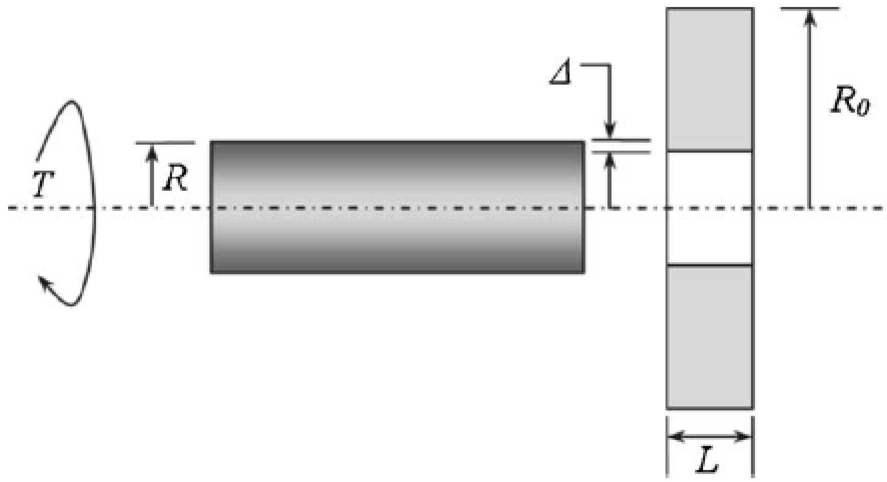


Figure 1: Schematic diagram for the torque shaft assembly (Mahadevan and Rebba, 2006)

The interface pressure can be derived using the assumption of a thick cylinder for the gear wheel and the shaft as

$$p = \frac{\Delta}{R\left[\frac{1}{E_0}\left(\frac{R_0^2 + R^2}{R_0^2 - R^2} + \upsilon_0\right) + \frac{1}{E_i}(1 - \upsilon_i)\right]} \tag{8}$$

where E_0 and E_i are Young's moduli, v_0 and v_i are the Poisson ratios of the gear wheel and the drive shaft, respectively.

The two design variables are bounded as $5 \leq R \leq 9$ and $10 \leq R_0 \leq 20$, respectively. Suppose we wish to ensure that the maximum torque transmitted by the assembly fit exceeds a threshold value T_0. The probability of achieving the design requirement needs to be evaluated first. A limit state is defined as $g = T - T_0$ and failure is defined when the torque delivered (T) is less than T_0, i.e., when $g < 0$.

Knowing the specific densities of the shaft ρ_i and the gear wheel ρ_0, the total weight of the assembly can be estimated as

$$W = \pi L g_a \left[\rho_0 R_0^2 + (\rho_i - \rho_0) R^2\right] \tag{9}$$

where g_a is the acceleration due to gravity. In this illustrative example, ρ_0 and ρ_i are assumed to be 7.85 and 7.95, respectively and T_0 is assumed to be 5000 units.

The general formulation for this RBDO problem is as follows:

$$\begin{aligned} \min \quad & W = \pi L g_a \left[\rho_0 R_0^2 + (\rho_i - \rho_0) R^2\right] \\ s.t. \quad & P(T \le T_0) \le p_0 \end{aligned} \tag{10}$$

where p_0 is assumed to be 0.0062 (β=2.5) in this example. Since the torque T transmitted by the mechanical assembly depends on both R and R_0, the probability P ($T \le T_0$) also depends on those respective radii.

The values for the deterministic variables and statistics of the various uncertain parameters are given in Table 1.

Table 1: Data of input variables in torque shaft assembly design

Variable	Distribution type	Data
E_0	Lognormal	[10,178; 9771; 9786; 9838; 9411; 10,288; 10,065; 9849; 10, 274; 9658]
E_i	Lognormal	[7980; 7952; 8064; 8063; 7827; 7994; 7967; 8126; 8219; 8222]
v_0	Lognormal	[0.1, 0.2]
v_i	Lognormal	Mean: 0.25, Std: 0.05

η	Lognormal	[0.7, 0.75; 0.73, 0.76; 0.72, 0.78]
Δ	Deterministic	0.01
L	Deterministic	4

In this case, the input random variable v_i is assumed to have aleatory uncertainty. All the remaining input random variables are considered as non-design epistemic variables, where E_0 and E_i are assumed to be described by sparse point data, v_0 is assumed to be described by single interval data, and η is assumed to be described by multiple interval data as given in Table 1. The confidence bounds on the mean for the variables E_0 and E_i are estimated by the methods described in Section 2. Bounds on the mean values and variances of the epistemic variables v_0 and η are estimated by the methods described in Chapter IV. Since this problem contains non-design epistemic variables, this problem is solved by the RBDO methodology developed in Section 2 by solving the following two optimization problems iteratively until convergence and the solutions are given in Table 2.

$$\begin{aligned} &[R^*, R_0^*] = \arg\min_{R,R_0} W(R, R_0) \\ &s.t. \quad P\left(T\left(R, R_0, \mu_z^*\right) \le T_0\right) \le p_0 \\ &\qquad 5 \le R \le 9 \\ &\qquad 10 \le R_0 \le 20 \end{aligned} \tag{11}$$

$$\begin{aligned} &\mu_z^* = \arg\max_{\mu_z} W\left(R^*, R_0^*\right) \\ &s.t. \quad P\left(T\left(R^*, R_0^*, \mu_z\right) \le T_0\right) \le p_0 \\ &\qquad Z_l \le \mu_{z_i} \le Z_u \qquad \text{for all } i \end{aligned} \tag{12}$$

where $\boldsymbol{Z_l}$ and $\boldsymbol{Z_u}$ are are the vectors of the bounds on the mean values of the epistemic variables.

Table 2: Optimal design solution for the torque shaft problem

Optimum (R, R_0)	W	No. of g function evaluations		
		Design Analysis (Eq. (11))	Epistemic Analysis (Eq. (12))	Total
(6.5297, 12.2412)	1.4539e+005	7,302	1,680	8,982

The optimizations in Eqs. (11) and (12) required only 2 iterations between the design problem (Eq. (11)) and the uncertainty analysis for the non-design epistemic variables (Eq. (12)) for convergence. Number of g function evaluations for both the design and epistemic analyses are listed in Table 2 for future reference. It is seen in Table 2 that the proposed RBDO methodology can solve this design problem with only 8,982 function evaluations, of which only 1,680 evaluations are required for the epistemic analyses and only 7, 302 evaluations are required for the design analyses. Note that the design analysis (Eq. (11)) is equivalent to an RBDO problem under aleatory uncertainty alone, since it is solved with a fixed set of non-design epistemic variables. If this example problem involved only aleatory uncertainty, the number of g function evaluation would be approximately half of 7,302, because it would require solving Eq. (11) only once instead of twice in the current example. Therefore, the proposed RBDO methodology under epistemic uncertainty can solve this problem with a reasonably increased number of function evaluations.

4. Conclusion

This chapter has developed formulations for reliability-based design optimization (RBDO) for single systems under both aleatory and epistemic uncertainty on the data of the random variables. Two types of data uncertainty – sparse point data and interval data – are considered. The computational efficiency of the proposed formulations is demonstrated with a number of example problems considering the number of individual disciplinary analyses.

The huge computational expense required for the epistemic analysis is reduced by decoupling the design analysis from the epistemic analysis. Unlike existing methods, it does not use separate representations for aleatory and epistemic uncertainties and does not require nested analysis. Both types of uncertainty are treated in a unified manner using a probabilistic format, thus reducing the computational effort and simplifying the optimization problem. The numerical example in this chapter was carried out using the classical nested loop RBDO formulation and the number of function evaluations needed was reported in Section 3. The focus of this chapter is not on efficiency, but on the inclusion of epistemic uncertainty in the design optimization. Several more efficient RBDO methods (single loop and sequential) have been developed in recent years, and all these methods can be enhanced to incorporate epistemic uncertainty. Future work in this direction also needs to include system reliability constraints.

CHAPTER IX

MULTIDISCIPLINARY SYSTEM ANALYSIS UNDER ALEATORY AND EPISTEMIC UNCERTAINTY

1. Introduction

This chapter develops an efficient probabilistic approach for uncertainty propagation in multidisciplinary system analysis, when the information on the uncertain input variables may be available as either sparse point data or as intervals (single or multiple). A decoupled approach is used in this chapter to un-nest the multidisciplinary system analysis from the probabilistic analysis to achieve computational efficiency. This approach uses deterministic optimization to first quantify the uncertainty in the coupling variables, without any coupled system level analysis. Once the uncertainty in the coupling variables is quantified, the system level uncertainty propagation analysis is similar to single discipline problems. The proposed methods are equally applicable to both sampling and analytical approximation-based reliability analysis methods. A mathematical problem and a practical engineering problem are used to illustrate the proposed methods. The accuracy of the proposed decoupled approach is verified by Monte Carlo simulation using a multi-discipline feasible (MDF) analysis approach.

As mentioned in Chapter II, the proposed method extends the idea of a decoupled formulation as developed in Mahadevan and Smith (2006) and proposes probabilistic methods for multidisciplinary reliability analysis under both aleatory and epistemic uncertainty. This approach uses deterministic optimization to first quantify the uncertainty in the coupling variables. No coupled system level analysis is required. Once the uncertainty in the coupling variables is quantified, the system level uncertainty propagation is achieved based on the single discipline uncertainty propagation methods that include both physical variability and data uncertainty, using a probabilistic approach.

The rest of the Chapter is organized as follows. Sections 2 and 3 describe the proposed methodology for the propagation of both epistemic and aleatory uncertainty through multidisciplinary analysis. Section 4 gives the numerical results using the proposed methods for a simple mathematical problem and a practical engineering problem. Section 5 provides concluding remarks and suggestions for future work.

2. Probabilistic Uncertainty Propagation for Single Discipline Problems

In the case of random variables for which only sparse point data or interval data are available, a flexible family of Johnson distributions is used to develop a probabilistic representation using the moment matching approach. Moment matching involves equating the moments derived from the data to those of the probability distribution being fit. A detailed discussion on fitting Johnson distributions to sparse point data and interval data can be found in Chapter III and Chapter IV, respectively.

An approach for uncertainty propagation with sparse point data has been developed in Chapter III. Chapter IV developed the method for propagation with interval data.

Note that the PBO and EBO methods presented in Chapter IV do not consider dependence among moments and are able to give rigorous bounds on the system response. However, it is more helpful to evaluate bounds in terms of both “rigor” and “optimality” as discussed and conceptually sketched in Figure 19 in Chapter IV. As mentioned in Chapter IV, by rigorous, it is meant that the true interval of the possible quantile values lies within the computed bounds. By optimal, it is meant that the bounds are the narrowest possible, while still being rigorous. The optimal bounds preserve the dependence among moments of interval data.

For a random variable, the moments are not independent to each other. For example, when the first moment is selected from a configuration of multiple interval data, it is obvious that the other moments will be estimated using the same configuration of multiple interval data. Therefore, if the moments are selected independently like the PBO and EBO methods presented in this section, the set Θ (see Chapter IV) becomes a superset of all actually admissible distribution parameter values resulting rigorous bounds on the system response, the lower and upper bounds of which may underestimate the output uncertainty. In the following discussion, we propose formulations for PBO and EBO that result in optimal bounds on the system response for multiple interval data. Note that this is not an issue for the single interval data. For multiple interval data, a particular value of moments within the moment bounds corresponds to a fixed set of configuration of multiple interval data. However, for single interval data, the moment bounds are calculated using closed-form formulas. See Chapter IV for details.

Optimal PBO formulation:

The approach is the same as in the percentile-based optimization (PBO) method presented earlier in this section, which minimizes and maximizes the system output $g_\alpha(x|m)$ conditioned on a set of moments *(m_i)* for the input variables at different percentile values *(α)* of the output CDF and thus obtains bounds on the system output

CDF. The optimal PBO solves the following optimization problems at different percentile values (α) to obtain bounds on the output CDF.

$$
\begin{aligned}
&\min_x / \max\, g_\alpha(x|m) \\
&s.t.\ a_i \le x_i \le b_i \qquad i = 1,2.....,N \\
&\quad \beta_2 - \beta_1 - 1 \ge 0 \\
&\quad \beta_2 - 2\beta_1 - 3 \le 0 \\
&\quad \text{where } \beta_1 = m_3^2 / m_2^3 \\
&\qquad\qquad \beta_2 = m_4 / m_2^2
\end{aligned}
\tag{1}
$$

where N is the number of intervals. Note that in PBO, the decision variables were the set of moments (m= [m_1 m_2 m_3 m_4]); however, in Eq. (1), the decision variables are configurations of multiple interval data (x = [x_1 x_2 x_3 … x_N]). The set of moments m are estimated using this configuration x of interval data inside the optimizer and thus, the dependency relationships among the moments are preserved resulting in optimal bounds on the system response.

Optimal EBO formulation:

The Optimal EBO method has the same formulation as in Eq. (1) but with a different objective function $\min_x / \max E(g(x|m))$. All the constraints remain the same. Note that in the EBO formulation, the decision variables were the set of moments (m= [m_1 m_2 m_3 m_4]), however, in optimal EBO, the decision variables are configurations of multiple interval data (x = [x_1 x_2 x_3 … x_N]). In this case, the optimization formulation yields configurations of multiple interval data, which are then used to estimate sets of moments for the uncertain design variables corresponding to a set of Johnson distribution parameters. Once the distribution parameters are obtained, any probabilistic uncertainty propagation method (e.g., FORM or MCS) can be used to construct approximate bounds on the CDF of the system response.

The uncertainty propagation methods described above are extended for multidisciplinary systems in Section 3.

3. Probabilistic Uncertainty Propagation for Multidisciplinary Problems

The computational effort required for multidisciplinary reliability analysis and design optimization depends on the type of formulation required for probabilistic system analysis. In this section, a decoupled approach adapted from Mahadevan and Smith (2006) is used to develop a method for multidisciplinary reliability analysis with sparse point data as well as interval data.

3.1 Multidisciplinary System Analysis

In many practical applications, multidisciplinary system analysis (MDA) makes use of individual disciplinary analysis codes that interact with each other through shared input and output data. A feasible multidisciplinary analysis yields a solution that simultaneously satisfies all individual disciplinary analyses (Du and Chen, 2005). Figure 1 shows a two-discipline system for the sake of illustration.

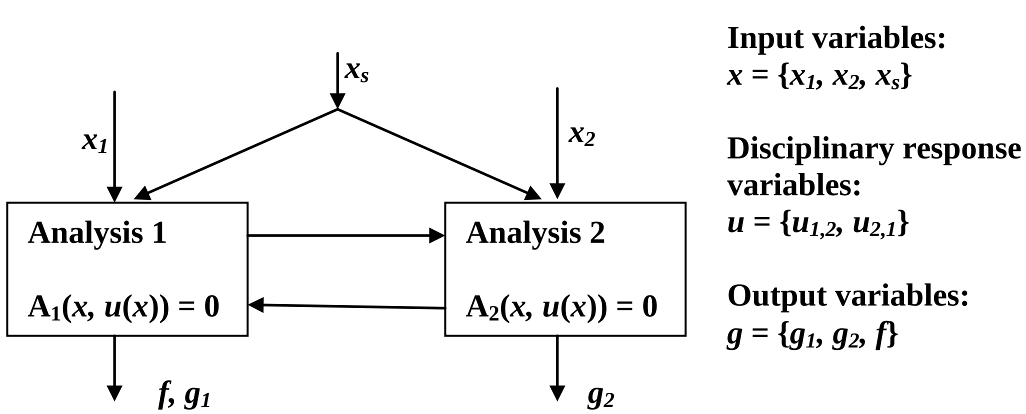


Figure 1: A two-disciplinary system with feedback coupling (Mahadevan and Smith, 2006)

In Figure 1, $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ are input variables to disciplines 1 and 2, respectively; $\boldsymbol{x_s}$ are the input variables common to each discipline. $\boldsymbol{u_{1,2}}$ and $\boldsymbol{u_{2,1}}$ are disciplinary response variables that couple the two disciplines. f, g_1 and g_2 are the system output variables. f may represent an objective function in the context of optimization and g_1 and g_2 may represent limit state functions for reliability analysis. In order to achieve feasibility in multidisciplinary system analysis, the non-linear equations shown in Eq. (2) below have to be solved simultaneously.

$$A_i(x, u(x)) = 0, \quad \text{for i} = 1,2 \tag{2}$$

In the following subsection, an approach is developed to decouple the system analysis from the probabilistic analysis.

3.2 Decoupled approach for probabilistic analysis

The coupling variables in a multidisciplinary analysis depend on the random input variables and therefore are random themselves. Mahadevan and Smith (2006) quantified the uncertainty in the coupling variables by using a first-order second moment (FOSM) approximation. Once the uncertainty in the coupling variables is quantified, probabilistic system analysis only needs uncertainty propagation through the individual disciplinary analyses, as shown in Figure 2. The uncertainty propagation can be achieved using already well established probabilistic methods of uncertainty propagation, for example, Monte Carlo methods (Robert and Cesalla, 2004) and optimization-based methods such as first-order reliability method (FORM), second-order reliability method (SORM) etc. (Haldar and Mahadevan, 2000).

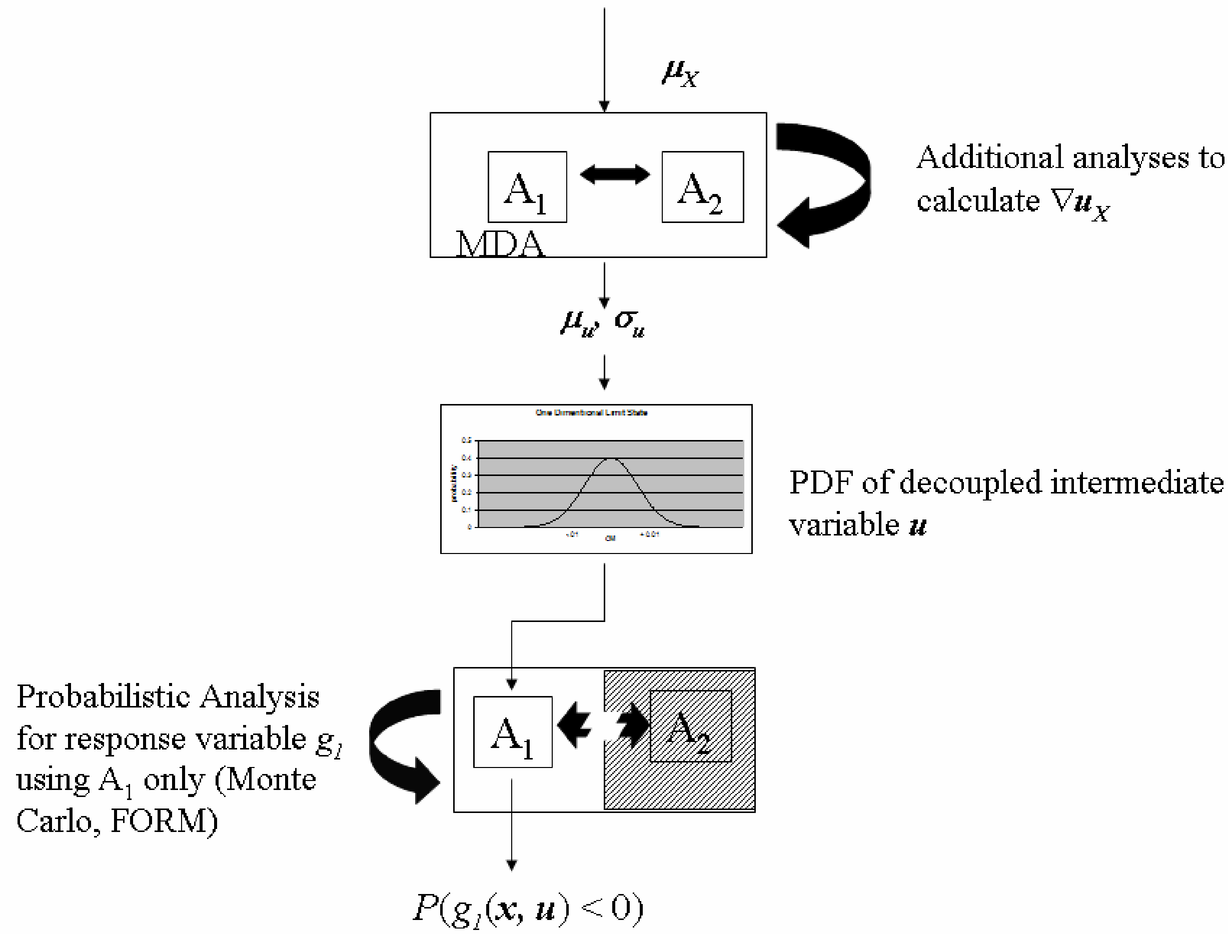


Figure 2: Decoupled formulation (Mahadevan and Smith, 2006)

The decoupled formulation described above has been developed for handling aleatory uncertainty only. In the following subsections, we extend this decoupled approach to develop methods for multidisciplinary uncertainty propagation analysis under epistemic uncertainty arising from sparse point data and interval data.

3.3 Multidisciplinary uncertainty propagation analysis with sparse point data

In this case, the only information available for the input random variables is sparse point data. In the first step, analysis is performed to generate data for the coupling variables. The second step performs uncertainty propagation through the individual disciplinary analyses using sparse point data for each of the input random variables. Du and Chen (2002) proposed uncertainty analysis methods for multidisciplinary problems, namely, the system uncertainty analysis (SUA) method and the concurrent subsystem uncertainty analysis (CSSUA) method for handling aleatory uncertainty. They used these methods to estimate the mean values of the coupling variables of the multidisciplinary system. The SUA requires a coupled system analysis and CSSUA requires solving a deterministic optimization problem using individual disciplinary analyses to estimate the mean of the coupling variable. Since coupled system analysis may be expensive for a large and complex system, we only use the CSSUA method in this chapter in order to generate sparse point data for the coupling variables as follows:

$$\min_{u^*} d = \sum_{i=1}^{n} \left(u_i - u_i^*\right)^2 \qquad (3)$$

where $\boldsymbol{u}^*$ are the unknown target values of the coupling variables and $\boldsymbol{u}$ are the values of the coupling variables obtained by subsystem analysis only. The optimizer minimizes the deviations between $\boldsymbol{u}^*$ and $\boldsymbol{u}$ and thus generates a set of data for the coupling variables. By solving the optimization problem in Eq. (3) N times, we can generate N number of sparse point data for each of the coupling variables.

Note that Du and Chen (2002) developed CSSUA for estimating the mean values of the coupling variables. However, in this chapter, we use this method only to generate sparse point data for the coupling variables. No coupled system level analysis is required. The system compatibility requirement is already satisfied through the system level optimizer as shown in Eq. (3). Once the data for the coupling variables is obtained, the uncertainty propagation method is straightforward. Its implementation is as follows:

1. Generate N sparse point data for the coupling variables by solving Eq. (3) N times, where N is the sample size for each of the input random variables $\boldsymbol{x}$.
2. Generate families of flexible probability distributions for each of the input random and coupling variables by the method described in Chapter III.
3. Propagate each of the input distributions through the corresponding individual disciplinary analysis to obtain a family of output distributions.

The three steps are illustrated in Figure 3.

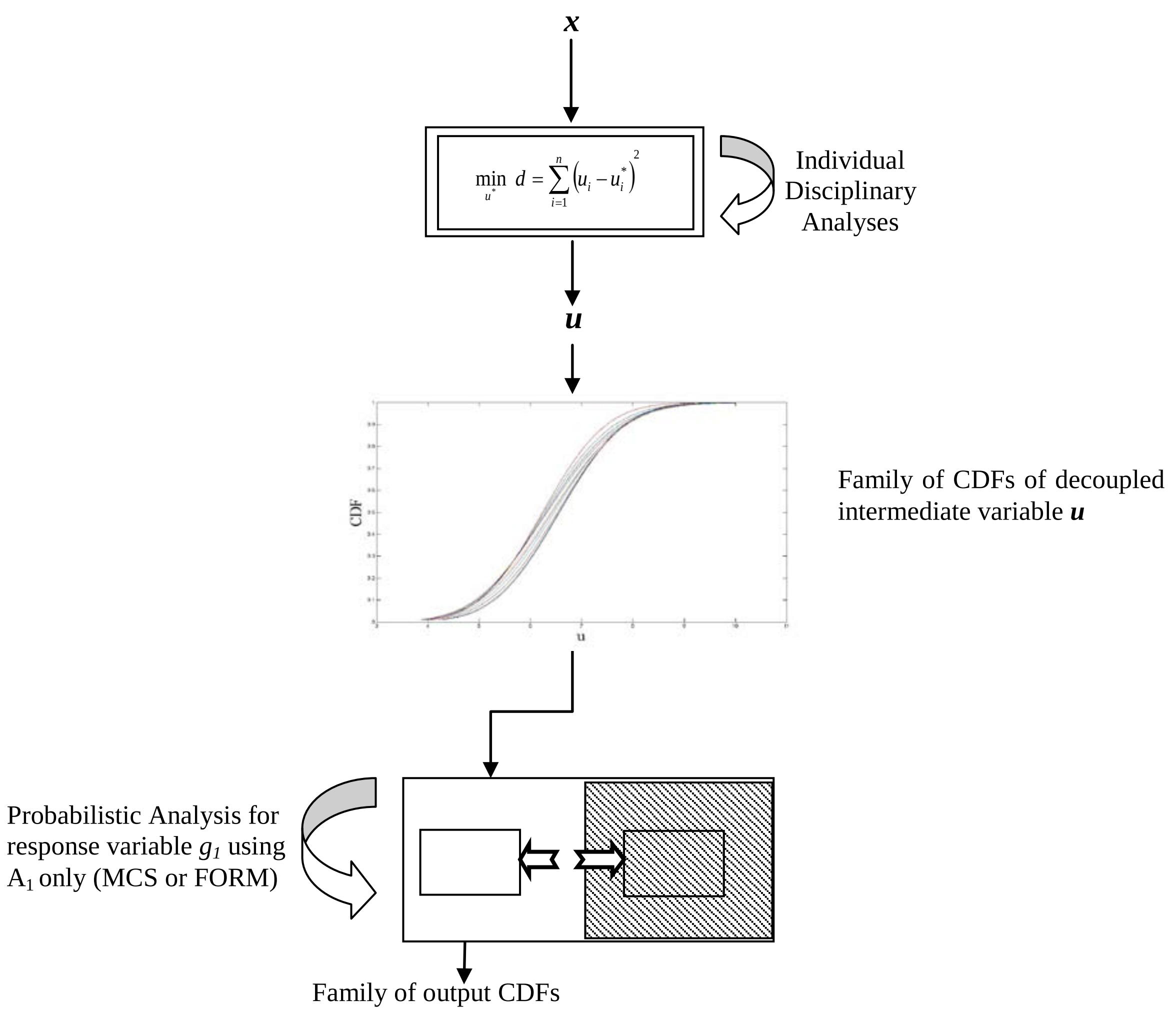


Figure 3: Decoupled approach for multidisciplinary uncertainty propagation with sparse point data

3.4 Multidisciplinary uncertainty propagation analysis with interval data

In this case, the only information available for the input random variables is in the form of single interval or multiple interval data. In the first step, a deterministic optimization is performed to generate interval data for the coupling variables. The second step performs uncertainty propagation through the individual disciplinary analyses using interval data for each of the input and coupling random variables. No coupled system level analysis is required. The system compatibility requirement is already satisfied through the constraints of the deterministic optimization. The steps of implementation are as follows:

1. Generate interval data for the coupling variables by solving:

$$\begin{aligned} &\min/\max \quad u \\ &s.t.\ a \le x \le b \\ &\quad A_i(x, u(x)) = 0, \text{for i} = 1, 2, .. \end{aligned} \tag{4}$$

   where a and b are the vectors of lower and upper endpoints of the given intervals for the random input variables $\boldsymbol{x}$. For multiple interval data, repeat the optimization in Eq. (4) N times to obtain N intervals for the coupling variables, where N is the number of intervals for the input random variables.

2. Calculate bounds on moments of the interval data for the input random and coupling variables by the method described in Chapter IV.

3. Obtain bounds on the system output by the optimization methods (PBO, EBO) mentioned in Section 2.

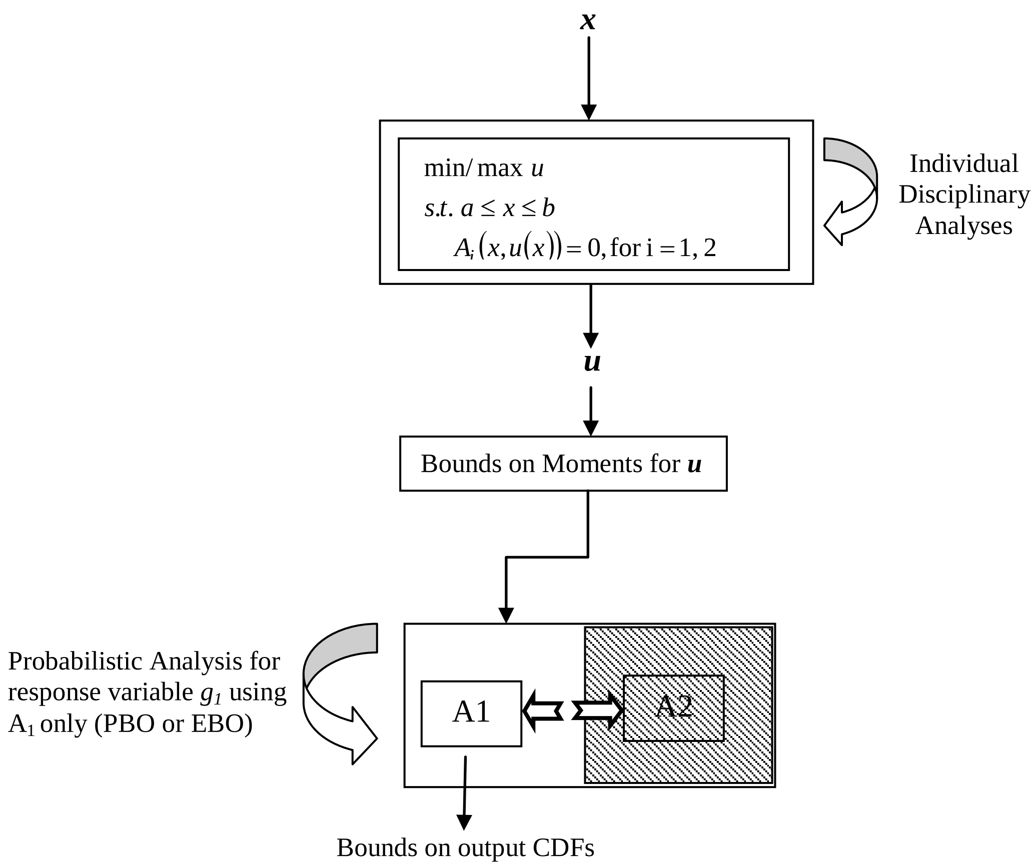


Figure 4: Decoupled approach for multidisciplinary uncertainty propagation with interval data

4. Numerical Examples

The proposed uncertainty propagation methods for multidisciplinary analysis under both aleatory and epistemic uncertainty on the input random variables are illustrated with two example problems: (1) a simple mathematical problem, and (2) an engineering problem.

4.1 Mathematical Problem

This mathematical problem is taken from Du and Chen (2005). This is a two-disciplinary problem with feedback coupling. The functional relationships for the individual disciplinary analyses are given as follows:

Analysis 1

$$X_s = \{x_1\}, X_1 = \{x_2, x_3\}$$
$$u_{1,2} = x_1^2 + 2x_2 - x_3 + 2\sqrt{u_{2,1}}$$
$$g_1 = x_1^2 + 2x_2 + x_3 + x_2 e^{-u_{2,1}}$$

Analysis 2

$$X_s = \{x_1\}, X_2 = \{x_4, x_5\}$$
$$u_{2,1} = x_1 x_4 + x_4^2 + x_5 + u_{1,2}$$
$$g_2 = \sqrt{x_1} + x_4 + x_5(0.4x_1)$$

In this example problem, the disciplinary response variables $u_{1,2}$ and $u_{2,1}$ couple the two analyses. These response variables are defined such that $u_{i,j}$ is an output of analysis *i* and an input to analysis *j*. The output g_1 of disciplinary analysis 1 will be used to illustrate the proposed multidisciplinary uncertainty propagation methods.

In this example problem, the disciplinary response variables $u_{1,2}$ and $u_{2,1}$ couple the two analyses. These response variables are defined such that $u_{i,j}$ is an output of analysis *i* and an input to analysis *j*. The output g_1 of disciplinary analysis 1 will be used to illustrate the proposed multidisciplinary uncertainty propagation methods.

Example 1(a): Sparse point data

In this case, the input random variables $\{x_1, ..., x_5\}$ are given by sparse point data. Each input random variable is described by the data set as given in Table 1. This problem is solved by the decoupled approach developed in Section 3.3 in order to obtain sparse point data for the coupling variables $u_{1,2}$ and $u_{2,1}$. Once the data of the coupling variables are obtained, the uncertainty propagation through the system output is achieved through the individual disciplinary analyses only by the method mentioned in Section 2 and the results are shown in Figure 5.

Table 1: Sparse Point Data for the random input variables

Sample	x_1	x_2	x_3	x_4	x_5
01	0.9567	0.9813	1.0294	0.9600	0.8396
02	0.8334	1.0726	0.8664	1.0690	1.0257
03	1.0125	0.9412	1.0714	1.0816	0.8944

04	1.0288	1.2183	1.1624	1.0712	1.1415
05	0.8854	0.9864	0.9308	1.1290	0.9195
06	1.1191	1.0114	1.0858	1.0669	1.0529
07	1.1189	1.1067	1.1254	1.1191	1.0219
08	0.9962	1.0059	0.8406	0.8798	0.9078
09	1.0327	0.9904	0.8559	0.9980	0.7829
10	1.0175	0.9168	1.0571	0.9843	0.9941

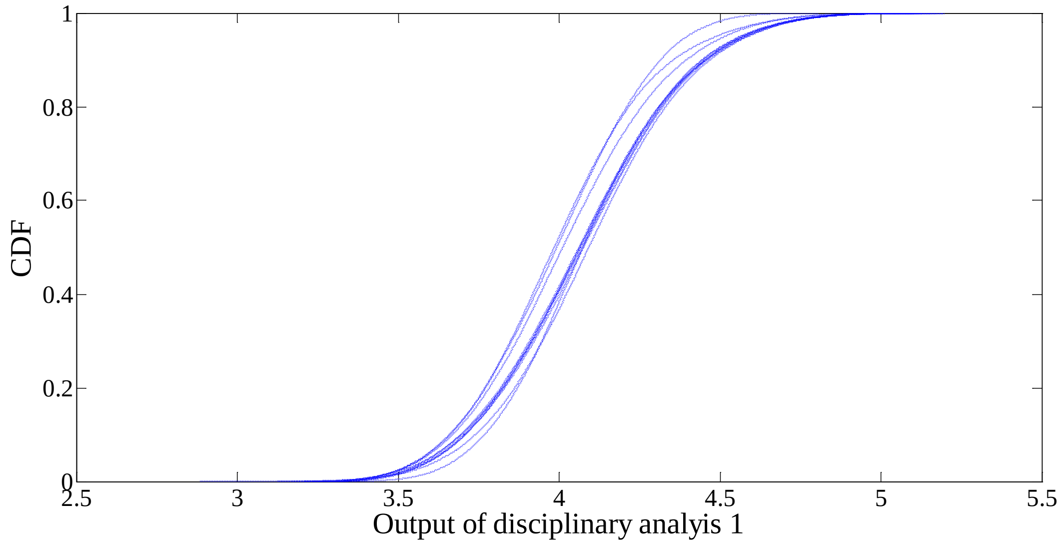


Figure 5: Propagation through multidisciplinary analysis of sparse point data

As seen in Figure 5, the proposed method quantifies the data uncertainty for the input random variables by generating a family of CDFs for the system response at the disciplinary level 1 of the two-discipline system. In this example problem, the generation of data for the coupling variables requires only 414 individual disciplinary analyses. The uncertainty propagation analysis has generated 100,000 (10×10,000) Monte Carlo samples for each of the input variables to construct 10 output CDFs, which requires 100,000 individual disciplinary analyses.

Example 1(b): Single interval data

In this case, the input random variables $\{x_1, \ldots, x_5\}$ are given by single interval data. Each input random variable ranges from 0.5 to 1.5. This problem is solved by the decoupled approach developed in Section 3.4 in order to obtain single interval data for the coupling variables $u_{1,2}$ and $u_{2,1}$. Once the data of the coupling variables are obtained, the uncertainty propagation through the system output is achieved through the individual disciplinary analyses only by the PBO and EBO methods mentioned in Section 2 and the results are shown in Figure 6. Note that this problem involves only single interval data and therefore, is solved by the basic PBO and PBO as described in Section 2. It is seen in Figure 6 that the bounds calculated by EBO and PBO almost coincide with each other for this problem. This problem is also solved by MCS using an MDF approach and the results are shown in Figure 6.

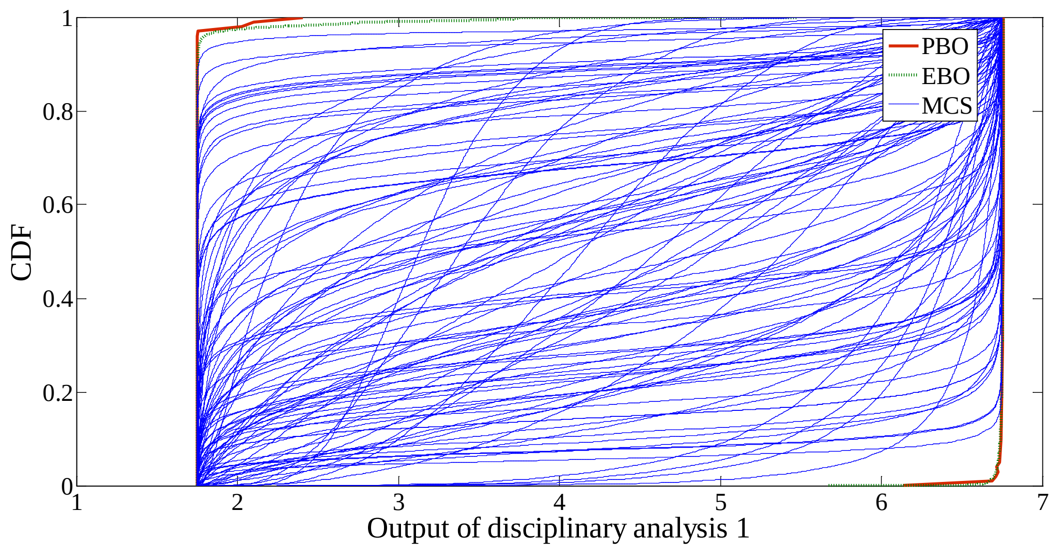


Figure 6: Propagation through multidisciplinary analysis of single interval data

Since this particular problem involves only single interval data, the uncertainty bounds for this problem can be obtained by a simple deterministic optimization as shown below:

$$\begin{aligned} &\min/\max g_1 \\ &s.t. \quad lb \le x_1 \le ub \\ &\qquad\; lb \le x_2 \le ub \\ &\qquad\; lb \le x_3 \le ub \\ &\qquad\; lb \le u_{2,1} \le ub \end{aligned} \qquad (5)$$

Note that the bounds on $u_{2,1}$ are obtained by the decoupled approach described in Section 3.4.

This optimization formulation yields the bounds on the system response as [1.75, 6.76] which is exactly the same as the lowermost and uppermost bounds obtained by the proposed probabilistic approach, corresponding to CDF values of 0 and 1. This approach requires only 15 individual disciplinary analyses. Note that if we solved this problem by a deterministic optimization using an MDF approach, it would give the same bounds as obtained by the decoupled deterministic optimization.

<u>Example 1(c): Multiple interval data</u>

In this case, the input random variables $\{x_1, ..., x_5\}$ are given by multiple interval data. Each input random variable is described by the following data set: ([0.5, 1.2], [0.8, 1.5], [0.75, 1.75], [0.5, 1.75], [0.7, 1.4]). This problem is solved by the decoupled approach developed in Section 3.4 in order to obtain multiple interval data for the coupling variables $u_{1,2}$ and $u_{2,1}$. Once the data of the coupling variables are obtained, the uncertainty propagation through the system output is achieved through the individual disciplinary analyses only by the PBO and EBO methods mentioned in Section 2 and the results are shown in Figure 7. This problem is also solved by MCS using an MDF approach and the results are shown in Figure 7.

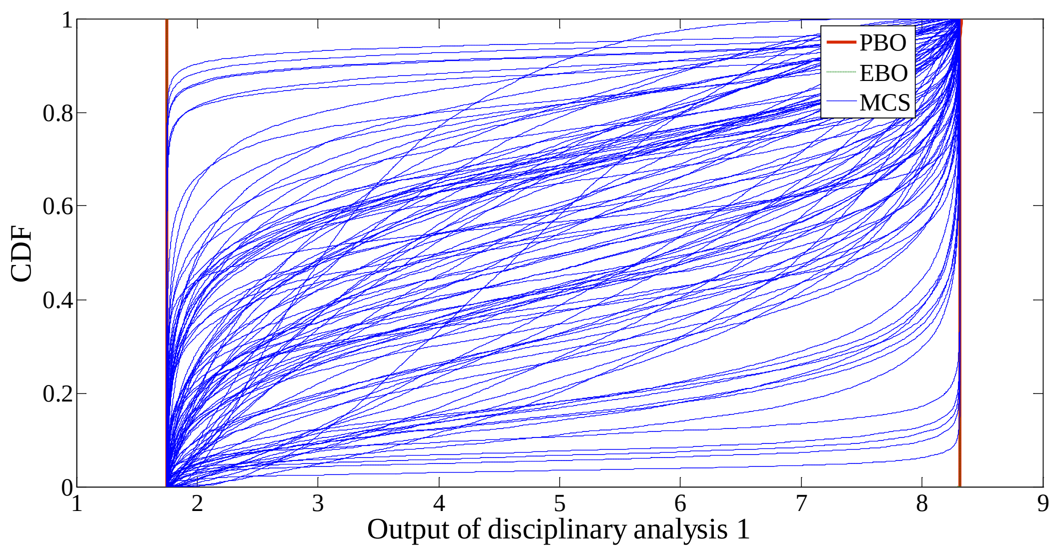


Figure 7: Propagation through multidisciplinary analysis of multiple interval data

As mentioned earlier in Section 2, the PBO and EBO methods give rigorous bounds on the system response for multiple interval data. Therefore, this problem is also solved by the proposed Optimal PBO and EBO formulations as discussed in Section 2 and the results are shown in Figure 8.

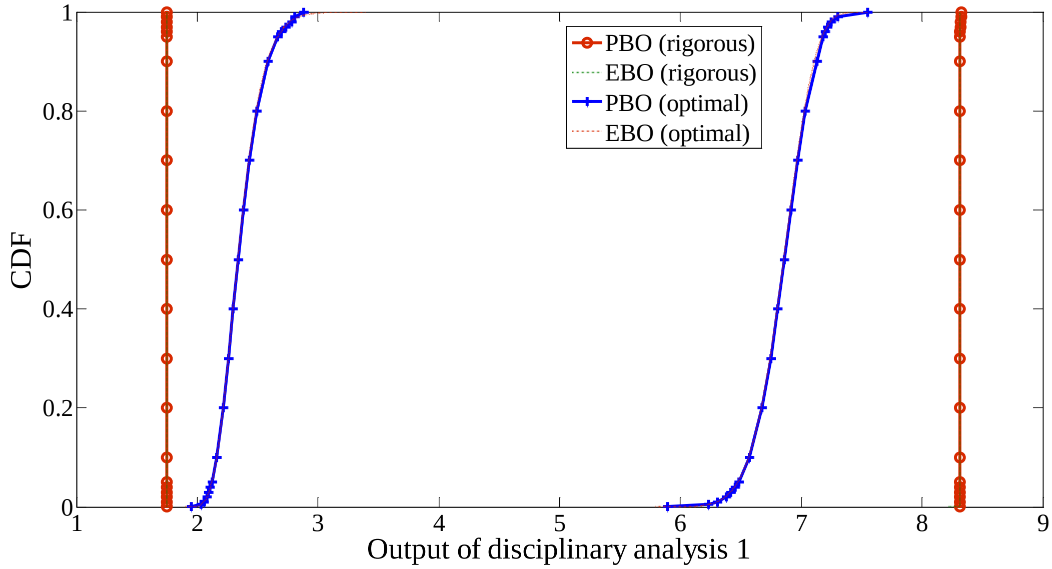


Figure 8: Rigorous vs. Optimal bounds for multiple interval data

It is seen in Figure 8 that the optimal bounds are in good agreement with the rigorous bounds. The bounds calculated by EBO and PBO coincide with each other for this problem.

The generation of data for the coupling variables requires only 56 and 280 individual disciplinary analyses for the single and multiple interval cases, respectively. The computational efforts for the PBO and EBO methods with both single and multiple interval data and the computational efforts for the Optimal PBO and Optimal EBO methods with multiple interval data are listed in Table 2. Obviously, EBO is less expensive compared to PBO for each problem. It is also seen that for the same problem, the Optimal PBO and Optimal EBO require more function evaluations than the basic PBO and EBO. This is expected due to the larger number of decision variables required in the former case.

Table 2: Computational effort for PBO and EBO for Example 1 with interval data

	Basic PBO		Basic EBO	Optimal PBO		Optimal EBO
	Percentile Points	**Function Evaluations**	**Function Evaluations**	**Percentile Points**	**Function Evaluations**	**Function Evaluations**
Single interval data	21	1482	295	-	-	-
Multiple interval data	21	1692	47	21	2394	243

4.2 Example 2: Engineering Problem (FireSat)

This problem has been is sketched in Figure 3 of Chapter I. As seen in Figure 3 (see Chapter I), the Orbit subsystem has feed-forward coupling with both Attitude Control and Power subsystems. Further, the Attitude Control and Power subsystems are coupled through three coupling variables P_{ACS}, I_{min}, and I_{max}. The functional relationships for the disciplinary analyses are given in Table 3. A satellite configuration is assumed in which two solar panels extend out from the spacecraft body. Each solar panel has dimensions *L* by *W* and the edge of the solar panel is at a distance *D* from the centerline of the satellite's body as sketched in Figure 9.

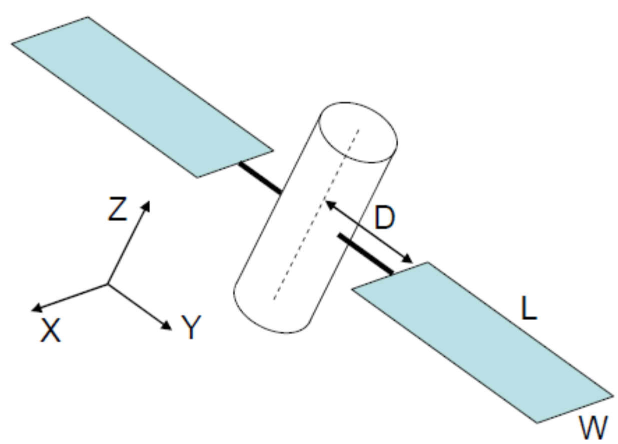


Figure 9: Schematic diagram for the spacecraft solar array (Ferson et al, 2009)

Table 3: Functional relationships among the disciplinary analyses

Subsystem-1 (Orbit)	**Subsystem-2 (Attitude Control)**	**Subsystem-3 (Power)**
The satellite velocity: $\upsilon = \sqrt{\frac{\mu}{R_E + H}}$ where, μ = Earth's gravity constant, R_E = Earth's radius, and H = orbit altitude. The orbit period: $\Delta t_{orbit} = 2\pi\sqrt{\frac{(R_E+H)^3}{\mu}} = \frac{2\pi(R_E+H)}{\upsilon}$ The maximum eclipse time: $\Delta t_{eclipse} = \frac{\Delta t_{orbit}}{\pi}\arcsin\left(\frac{R_E}{R_E+H}\right)$ The maximum slewing angle: $\theta_{slew} = \arctan\left(\frac{\sin(\phi_{target}/R_E)}{1-\cos(\phi_{target}/R_E)+H/R_E}\right)$ where, ϕ_{target} = target diameter.	The slewing torque: $\tau_{slew} = \frac{4\theta_{slew}}{\Delta t_{slew}^2} I_{\max}$ where, I_{max} = maximum moment of inertia of the spacecraft calculated in Power subsystem. Total disturbance torque: $\tau_{dist} = \sqrt{\tau_g^2 + \tau_{sp}^2 + \tau_m^2 + \tau_a^2}$ where, τ_g, τ_{sp}, τ_m, and τ_a = Torques due to gravity gradients, solar radiation, magnetic field interactions, and aerodynamic drag, respectively. $\tau_g = \frac{3\mu}{2(R_E+H)^3}\lvert I_{\max} - I_{\min}\rvert \sin(2\theta)$ Here, I_{max} and I_{min} = maximum and minimum moment of inertia for the spacecraft calculated in Power subsystem, θ = the deviation of the major moment axis from the local vertical (nadir). $\tau_{sp} = L_{sp}\frac{F_s}{c}A_s(1+q)\cos i$ Here, L_{sp} = moment arm for the solar radiation torque – the distance between the center of the solar pressure and the center of gravity of the spacecraft, F_s = average solar flux, c = speed of light (2.9979e8 m/s), q = reflectance factor or surface reflectivity, and i = sun incidence angle (angle at which the sun radiation hits the spacecraft surface), A_s = surface area off which the solar radiation is reflected. For cylindrical solar arrays, $A_s = A_{sa}/\pi$ $\tau_m = \frac{2MR_D}{(R_E+H)^3}$ Here, M is the magnetic moment of	The total power: $P_{tot} = P_{ACS} + P_{other}$ Here, the Attitude Control subsystem is only considered explicitly as a power consumer. All other power consumers are lumped into one bin as P_{other}. P_{ACS} is calculated in the Attitude subsystem. The total solar array size: $A_{sa} = \frac{P_{sa}}{P_{EOL}}$ where, Required Power Output, $P_{sa} = \frac{\left(\frac{P_e T_e}{X_e} + \frac{P_d T_d}{X_d}\right)}{T_d}$ Here, P_e and P_d = spacecraft's power requirements during eclipse and daylight, respectively. For this example problem, $P_e = P_d = P_{tot}$ T_e and T_d = time per orbit spent in eclipse and in sunlight, respectively. For this example problem, $T_e = \Delta t_{eclipse}$ and $T_d = \Delta t_{orbit} - T_e$ The power production capability at the end of life, $P_{EOL} = P_{BOL}(1-\varepsilon_{\deg})^{LT}$ where, LT = lifetime of the spacecraft in years, ε_{deg} = degradation per year in %/year. The power production capability at the beginning of life, $P_{BOL} = \eta F_s I_d \cos\theta$. Here, I_d = inherent degradation of the array — It lumps together temperature effects, shadowing, and uncovered areas in the physical layout, θ = sun incidence angle — typically a worst-case angle is used. The equations for the moment of inertia for the configuration as shown in Figure 6:

	the Earth expressed in Am^2 and R_D is the residual dipole of the spacecraft. $\tau_a = \frac{1}{2} L_a \rho C_d A V^2$ Here, L_a is the moment arm for the aerodynamic drag torque – the distance between the center of the aerodynamic pressure and the center of gravity of the spacecraft, ρ is the atmospheric density, C_d is the drag coefficient, A is the cross-sectional surface area in the direction of flight, and V is the velocity of the spacecraft in orbit. The Attitude control power: $P_{ACS} = \tau_{tot}\omega_{\max} + nP_{hold}$ where, $\tau_{tot} = \max(\tau_{slew}, \tau_{dist})$ ω = maximum rotational velocity of a reaction wheel (typically 5000-6000 rpm), n = number of reaction wheels that could be simultaneously active (in this case n=3), and P_{hold} = holding power — the power necessary to maintain a constant velocity of $\omega_{\max}$.	$L = \sqrt{\frac{A_{sa} r_{lw}}{n_{sa}}}$ $W = \sqrt{\frac{A_{sa}}{r_{lw} n_{sa}}}$ $m_{sa} = 2\rho LWt$ $I_{saX} = m_{sa}\left[\frac{1}{12}\left(L^2 + t^2\right) + \left(D + \frac{L}{2}\right)^2\right]$ $I_{say} = \frac{m_{sa}}{12}\left(H^2 + W^2\right)$ $I_{saZ} = m_{sa}\left[\frac{1}{12}\left(L^2 + W^2\right) + \left(D + \frac{L}{2}\right)^2\right]$ where, r_{lw} = the length to width ratio of the solar array, n_{sa} = number of solar arrays, ρ = average mass density of the arrays, t = thickness of the solar panels. The distance D can be independently chosen, but is should be larger than the radius of the spacecraft body. The total moment of inertia: $I_{tot} = I_{sa} + I_{body}$ I_{body} = MoI of the main body of the spacecraft (one for each axis x, y, and z). $I_{\min} = \min\left(I_{tot,X}, I_{tot,Y}, I_{tot,Z}\right)$ $I_{\max} = \max\left(I_{tot,X}, I_{tot,Y}, I_{tot,Z}\right)$ Note that in the full system analysis, the overall moments of inertia of the system are computed in the "Structures" subsystem. To limit the scope of this example problem, the overall moments of inertia are now calculated in the "Power" subsystem and the moments of inertia of the main body I_{body} are kept constant as follows: $I_{bodyX} = 6200 \text{ kg} - \text{m}^2$ $I_{bodyY} = 6200 \text{ kg} - \text{m}^2$ $I_{bodyZ} = 4700 \text{ kg} - \text{m}^2$

The objective is to quantify the uncertainty in 3 output variables that are the result of the 3-disciplinary analysis – total power P_{tot}, required solar array area A_{sa}, and total torque τ_{tot} as shown in Figure 3 of Chapter I. The uncertain variables involved in each subsystem and their corresponding single interval data are listed in Table 4. Note that this problem involves only single interval data and therefore, is solved by the basic PBO and PBO as mentioned in Section 2.

Table 4: Uncertain variables and data

No	Variable	Symbol	Unit	Data
1	Earth's radius	R_E	m	[6378135, 6378145]
2	Altitude	H	m	[2×10^5, 3.5787×10^7]
3	Power other than ACS	P_{other}	W	[825, 1375]
4	Avg solar flux	F_s	W/m^2	[1326, 1481]
5	Deviation of major moment axis from local vertical	θ	deg	[10, 19]
6	Moment arm for solar radiation torque	L_{sp}	m	[0, 3.75]
7	Reflectance factor	q		[0.1, 0.99]
8	Residual dipole of the space craft	R_D	Am^2	[0, 10]
9	Moment arm for aerodynamic torque	L_a	m	[0, 3.75]
10	Drag coefficient	C_d		[2, 4]

This problem is solved by the decoupled approach developed in Section 3.4 in order to obtain the interval data for the coupling variables P_{ACS}, I_{min}, and I_{max}. Once the data of the coupling variables are obtained, the uncertainty propagation through the system outputs P_{tot}, A_{sa}, and τ_{tot} are achieved through the individual disciplinary analyses only. Each of the problems is also solved by MCS using an MDF approach.

i) <u>System Output 1: Total power, P_{tot}</u>: This problem is solved by the PBO and EBO methods mentioned in Section 2.2 and the results are shown in Figure 10.

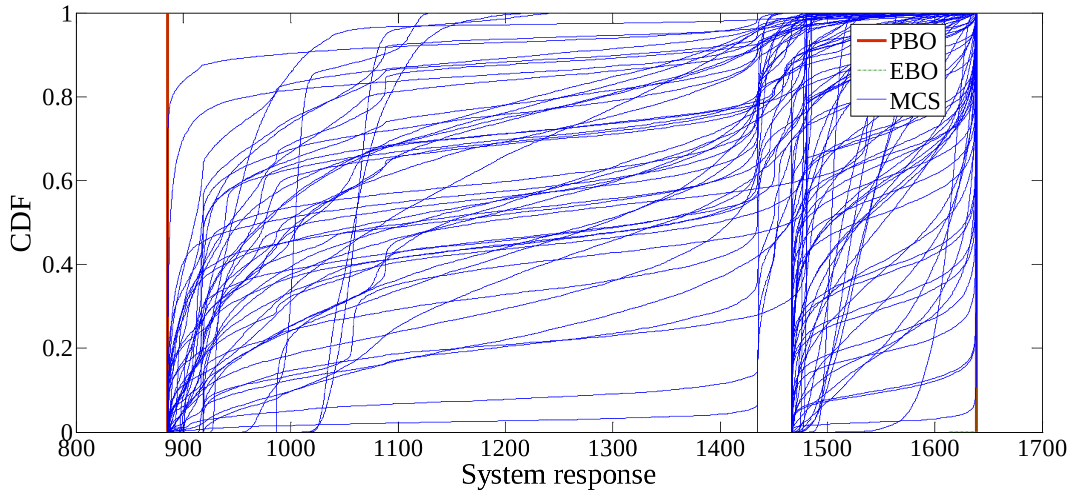


Figure 10: Bounds on P_{tot}

Since this problem involves only single interval data, the uncertainty bounds for this problem can be obtained by a simple deterministic optimization as shown below:

$$\begin{aligned} &\min_x / \max P_{tot} \\ &s.t.\ lb \le P_{ACS} \le ub \\ &\quad lb \le P_{other} \le ub \end{aligned} \tag{6}$$

where the design variables x are P_{ACS} and P_{other}. Note that the bounds on P_{ACS} are obtained by the decoupled approach described in Section 3.4.

This optimization formulation yields the bounds on the system response as [885.206, 1638.7] which is exactly the same as the lowermost and uppermost bounds obtained by the proposed probabilistic approach, corresponding to CDF values of 0 and 1. This approach requires only 6 individual disciplinary analyses.

ii) System Output 2: Total array size, A_{sa}: This problem is solved by the PBO and EBO methods mentioned in Section 2.2 and the results are shown in Figure 11.

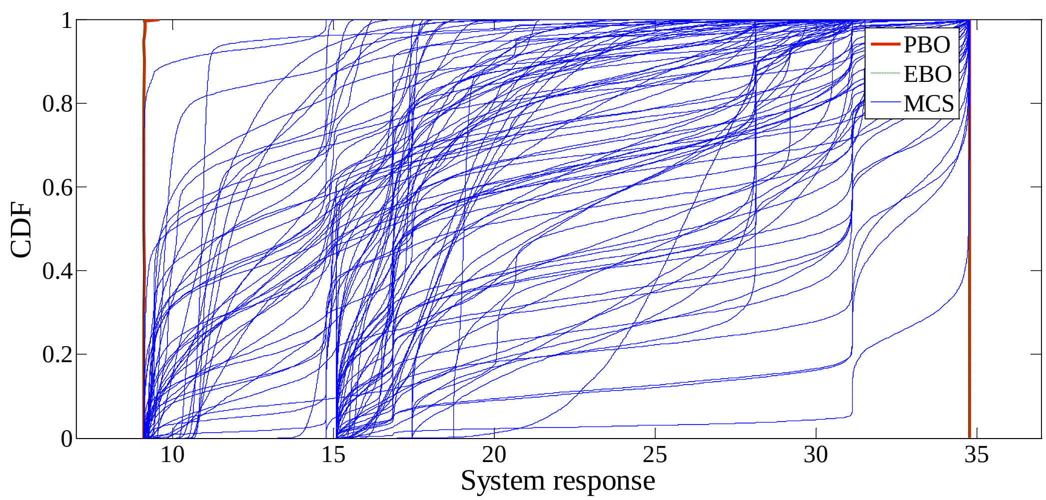


Figure 11: Bounds on A_{sa}

Since this problem involves only single interval data, the uncertainty bounds for this problem can be obtained by a simple deterministic optimization as shown below:

$$
\begin{aligned}
& \min_{x}/\max A_{sa} \\
& s.t.\ lb \le R_E \le ub \\
& \quad lb \le H \le ub \\
& \quad lb \le P_{other} \le ub \\
& \quad lb \le F_s \le ub \\
& \quad lb \le P_{ACS} \le ub
\end{aligned} \tag{7}
$$

where the design variables x are R_E, H, P_{other}, F_s, and P_{ACS}. Note that the bounds on P_{ACS} are obtained by the decoupled approach described in Section 3.4.

This optimization formulation yields the bounds on the system response as [9.11, 34.77] which is exactly the same as the lowermost and uppermost bounds obtained by the proposed probabilistic approach, corresponding to CDF values of 0 and 1. This approach requires only 24 individual disciplinary analyses.

iii) System Output 3: Total torque, τ_{tot}: This problem is solved by the PBO and EBO methods mentioned in Section 2.2and the results are shown in Figure 12.

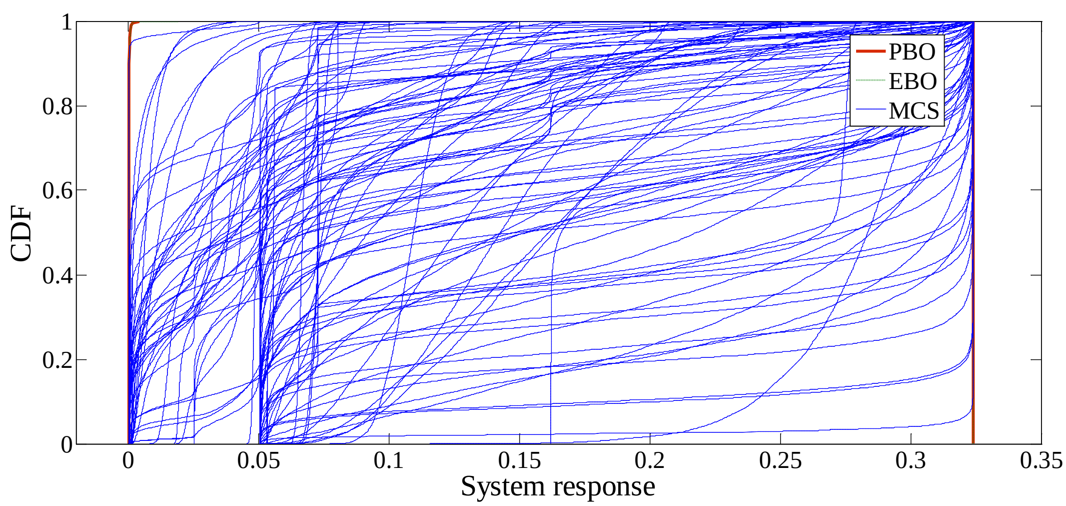


Figure 12: Bounds on τ_{tot}

Since this problem involves only single interval data, the uncertainty bounds for this problem can be obtained by a simple deterministic optimization as shown below:

$$
\begin{aligned}
\min_x/\max \quad & \tau_{tot} \\
s.t. \quad & lb \le R_E \le ub \\
& lb \le H \le ub \\
& lb \le \theta \le ub \\
& lb \le L_{sp} \le ub \\
& lb \le F_s \le ub \\
& lb \le q \le ub \\
& lb \le R_D \le ub \\
& lb \le L_a \le ub \\
& lb \le C_d \le ub \\
& lb \le I_{\min} \le ub \\
& lb \le I_{\max} \le ub
\end{aligned} \tag{8}
$$

where the design variables x are R_E, H, θ, L_{sp}, F_s, q, R_D, L_a, C_d, I_{min}, and I_{max}. Note that the bounds on I_{min}, and I_{max} are obtained by the decoupled approach described in Section 3.4.

This optimization formulation yields the bounds on the system response as [0.00033, 0.3240] which is exactly the same as the lowermost and uppermost bounds obtained by the proposed probabilistic approach, corresponding to CDF values of 0 and 1. This approach requires only 24 individual disciplinary analyses.

Note that we have used single interval data of the coupling variables in Eqs. (6)-(8) as obtained by the decoupled approach. If we solved these problems by a deterministic optimization using an MDF approach, it would give the same bounds as obtained by the decoupled deterministic optimization.

It is seen in Figures 10-12 that the bounds calculated by EBO and PBO coincide with each other. In this example problem, the generation of data for the coupling variables requires only 84 individual disciplinary analyses. The computational efforts for both the PBO and EBO methods are listed in Table 5. As expected, EBO is less expensive compared to PBO for each problem.

Table 5: Computational effort for the FireSat problem

	PBO		EBO Function Evaluations
	Percentile Points	**Function Evaluations**	
P_{tot}	26	559	18
A_{sa}	23	966	121
τ_{tot}	23	2484	108

Discussion

Each problem is also solved by MCS using an MDF approach in order to verify the results of the proposed PBO and EBO methods. It is seen in Figures 6-7 and 10-12 that the proposed PBO and EBO bounds are in good agreement with the MCS results. Note that this MCS approach is computationally very expensive. In order to generate 100 output CDFs, we have used 1 million (100×10,000) samples for each input variable, which requires 1 million system analyses. The proposed decoupled approach for multidisciplinary uncertainty propagation does not require any coupled system level analysis, which makes the proposed methods computationally feasible.

The deterministic optimizations with single interval data are very efficient in estimating the output uncertainty bounds with a very few individual disciplinary analyses. However, this approach is not able to give any probabilistic information about the output uncertainty and therefore this approach is recommended when the only quantity of interest is the bounds on output uncertainty. When the uncertainty propagation analysis is required to produce any probabilistic information, the PBO and EBO methods are recommended. Note that the deterministic optimization approach is applicable with single interval data only, whereas the PBO and EBO methods are equally applicable with both single and multiple interval data.

5. Conclusion

This chapter developed a probabilistic framework for the propagation of uncertainty through multidisciplinary systems, when the information is available as sparse point data or interval data. The uncertainty described by sparse point and interval data is represented through a flexible Johnson family of distributions. An optimization-based approach is used to decouple the probabilistic analysis from the system analysis. This approach uses deterministic optimization to first quantify the uncertainty in the coupling variables. No coupled system level analysis is required. This chapter also discussed the concepts of rigor and optimality with regard to the bounds on the system response and proposed optimization formulations that give optimal bounds on the output uncertainty. The proposed decoupled approach is illustrated for a mathematical problem and for a practical engineering problem.

The major advantage of the proposed methodology is that it does not require any coupled system level analysis, which makes it computationally efficient for large and complex multidisciplinary systems where only individual analysis codes are available. Unlike existing methods, it does not use separate representations for aleatory and epistemic uncertainties and does not require nested analysis. Both types of uncertainty are treated in a unified manner using a probabilistic format, thus reducing the computational effort and simplifying the optimization problem. The results regarding the uncertainty in the coupling variables are valuable to the designer as it can help select the initial guesses in an all-at-once approach to multidisciplinary design optimization. Due to the use of a probabilistic format to represent all the uncertain variables, the proposed uncertainty propagation framework facilitates the implementation of multidisciplinary design optimization in the presence of both aleatory and epistemic uncertainty.

CHAPTER X

ROBUSTNESS-BASED DESIGN OPTIMIZATION OF MULTIDISCIPLINARY SYSTEM UNDER EPISTEMIC UNCERTAINTY

1. Introduction

This chapter proposes formulations and algorithms for design optimization of multidisciplinary systems under both aleatory uncertainty (i.e., natural or physical variability) and epistemic uncertainty (i.e., imprecise probabilistic information), from the perspective of system robustness. The proposed formulations specifically deal with epistemic uncertainty arising from sparse and interval data without any assumption about the probability distributions of the random variables. A single loop approach is used for the design optimization, which does not require any explicit coupled multidisciplinary uncertainty propagation analysis. Thus the computational complexity and cost involved in estimating the mean and variation of the performance function is greatly reduced. A decoupled approach is proposed in this chapter to un-nest the robustness-based design from the analysis of non-design epistemic variables to achieve further computational efficiency. The proposed methods are illustrated for a mathematical problem and a practical engineering problem, where the information on the random inputs is only available as sparse point and/or interval data.

The contribution of this chapter is to develop a methodology for robustness-based design optimization for multidisciplinary systems that includes both aleatory and epistemic uncertainty. This chapter specifically focuses on epistemic uncertainty arising from *sparse point data* and *interval data*. In this chapter, we propose an efficient single loop formulation for the robust design problem. The proposed single loop formulation eliminates the need for explicit interdisciplinary uncertainty propagation for estimating the mean and variation of the output. A decoupled approach is proposed in this chapter to un-nest the robustness-based design from the analysis of non-design epistemic variables to achieve further computational efficiency. The proposed robustness-based MDO approach is based on the framework for single discipline systems developed in Chapter VII. In order to demonstrate the efficiency of the proposed method, the robust optimization methods based on SUA and CSSUA developed in Du and Chen (2002) are also used and modified in this chapter to include data uncertainty. The proposed method is illustrated by using a mathematical example and an engineering example.

The rest of the chapter is organized as follows. Section 2 proposes a multidisciplinary robustness-based design optimization framework that considers sparse point data and interval data for the random variables. In Section 3, we illustrate the proposed methods for a mathematical example and an engineering example. Section 4 provides conclusions and suggestions for future work.

2. Robustness-based design optimization for multidisciplinary systems

In Chapter VII, a methodology for robustness-based design optimization is proposed for single discipline systems. In this chapter, a methodology for robustness-based design optimization for multidisciplinary systems is developed, based on the methodology developed in Chapter VII.

As mentioned in Chapter IX, in order to achieve feasibility in multidisciplinary system analysis, the non-linear equations shown in Eq. (7) in Chapter IX have to be solved simultaneously.

Existing methods for multidisciplinary robust design optimization solve either Eq. (7) (see Chapter IX) or a sub-optimization problem nested within the framework of the design optimization to estimate the means of the disciplinary response variables and thereby estimate the mean of the performance function. The means of the input design variables as well as the disciplinary response variables are then used to estimate the variance of the performance function. This makes the current methods computationally expensive for coupled multidisciplinary systems. In this chapter, we propose a single loop formulation for the multidisciplinary robustness-based design optimization that eliminates the need for coupled interdisciplinary uncertainty analysis for estimating the mean and variance of the performance function.

Section 2.1 proposes a single loop formulation for multidisciplinary robustness-based design optimization and Section 2.2 proposes formulations for multidisciplinary robustness-based design optimization that account for input data uncertainty.

2.1 Multidisciplinary robustness-based design optimization

Existing robustness-based design optimization frameworks use different multidisciplinary optimization methods including the all-in-one approach (Du and Chen, 2002), collaborative optimization (Li and Azarm, 2008), etc. In this chapter, we compare the efficiency of the proposed method and the all-in-one approach of multidisciplinary robust optimization. The all-in-one approach is more commonly known as the multidisciplinary feasible method (MDF) in the literature (Cramer et al, 1994). The formulation of robustness-based design optimization using the all-in-one approach is as follows:

$$\begin{aligned} \min_{d} \quad & f(\mu,\sigma) = w*\mu_f(d,z) + (1-w)*\sigma_f(d,z) \\ s.t. \quad & LB + k\sigma_{g_i}(d,z) \le E\left(g_i(d,z)\right) \le UB - k\sigma_{g_i}(d,z) \text{ for all } i \\ & lb + k\sigma_{x/x_s} \le d \le ub - k\sigma_{x/x_s} \end{aligned} \tag{1}$$

The formulation in Eq. (1) is similar to that in Eq. (1) in Chapter VII. Here, the design variables $\boldsymbol{d}$ are the deterministic design variables and the mean values of the uncertain local input variables $\boldsymbol{x}$ as well as the shared input variables $\boldsymbol{x_s}$ (see Chapter IX). Note that one or more of the input random variables $\boldsymbol{x}$ and $\boldsymbol{x_s}$ may be non-design variables and referred to as $\boldsymbol{z}$ throughout this chapter. g_i is the constraint of the ith discipline. This robust design formulation requires estimating the mean values μ_f and μ_g as well as the standard deviations σ_f and σ_g considering the multidisciplinary nature of the system. In order to estimate the mean values μ_f and μ_g and the standard deviations σ_f and σ_g, it is necessary to estimate the mean values and the standard deviations of the coupling variables.

This all-in-one approach of robust design estimates the mean values of the coupling variables using either SUA or the CSSUA methods of interdisciplinary uncertainty propagation (Du and Chen, 2002). It has been mentioned earlier that SUA requires a coupled system level analysis at each iteration of the robust optimization problem and CSSUA requires a nested double loop formulation when used for robust optimization. The standard deviations of the coupling variables are estimated by approximating the system equations in Eq. (1) through a first-order Taylor series approximation at the mean values (Haldar and Mahadevan, 2000) of all the input and the coupling variables and then solving the following system of linear equations:

$$\sigma_{u_i}^2 = \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\frac{\partial F_{ui}}{\partial u_j} \right)^2 \sigma_{uj}^2 + \left(\frac{\partial F_{ui}}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial F_{ui}}{\partial x_s} \right)^2 \sigma_{x_s}^2 \quad \text{for all } i \tag{2}$$

where, $\boldsymbol{\sigma}_{u_i}^2, \boldsymbol{\sigma}_x^2$ and $\boldsymbol{\sigma}_{x_s}^2$ are the variances of *i*th coupling variable, local input variables and shared input variables, respectively.

In the following discussion, we propose a single loop formulation for the robustness-based design optimization that does not require any explicit interdisciplinary uncertainty propagation.

<u>Single loop formulation</u>

The all-in-one approach of multidisciplinary robust optimization satisfies the system compatibility requirements through a coupled interdisciplinary uncertainty analysis for SUA-based uncertainty propagation or by solving a nested double loop formulation for CSSUA-based uncertainty propagation as mentioned earlier. This approach also requires solving a system of linear equations at least once at each iteration of the design optimization. The idea behind the single loop formulation is that if the system compatibility requirement can be satisfied within the optimization algorithm by including the coupling variables as the optimization design variables as in the all-at-once approach, the only difficulty left is in estimating the standard deviations $\boldsymbol{\sigma_f}$ and $\boldsymbol{\sigma_g}$ of the objective function and the constraints, respectively. This difficulty can be overcome by including also the standard deviations of the disciplinary response variables $\boldsymbol{\sigma_u}$ as the optimization design variables. These additional design variables $\boldsymbol{\sigma_u}$ will be constrained by Eq. (9).

The single loop formulation of the multidisciplinary robustness-based design optimization is now as follows:

$$
\begin{aligned}
&\min_{d,\mu_u,\sigma_u} \quad f(\mu,\sigma) = w * \mu_f(d,z,u) + (1-w) * \sigma_f(d,z,u,\sigma_u) \\
&s.t. \qquad LB + k\sigma_{g_i}(d,z,u,\sigma_u) \le E(g_i(d,z,u)) \le UB - k\sigma_{g_i}(d,z,u,\sigma_u) \text{ for all } i \\
&\qquad\qquad lb + k\sigma_{x/x_s} \le d \le ub - k\sigma_{x/x_s} \\
&\qquad\qquad A_i(d,z,u(d,z)) = 0 \quad \text{for all } i \\
&\qquad\qquad \sigma_{u_i}^2 = \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(\frac{\partial F_{ui}}{\partial u_j}\right)^2 \sigma_{uj}^2 + \left(\frac{\partial F_{ui}}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial F_{ui}}{\partial x_s}\right)^2 \sigma_{x_s}^2
\end{aligned}
\tag{3}
$$

In the above optimization formulation, the design variables are the mean values of all input variables $\boldsymbol{x}$ and $\boldsymbol{x_s}$ and coupling variables $\boldsymbol{\mu_u}$ as well as the standard deviations of the coupling variables $\boldsymbol{\sigma_u}$.

In the following subsections, the methodology for robustness-based design optimization under epistemic uncertainty described in Chapter VII is extended for the multidisciplinary systems.

2.2 Multidisciplinary robustness-based design optimization under epistemic uncertainty

As in single discipline problem, in this case, the design variables ***d*** and/or the input random variables ***z*** in Eqs. (1) and (3) might have epistemic uncertainty. Since the designer does not have any control on the non-design epistemic variables ***z***, the design methodology has to employ a search among the possible values of such epistemic variables in order to find an optimal solution. In such case, we get a conservative robust design. When one or more epistemic variables cannot be treated as design variables, the design methodology has to solve two optimization problems iteratively until convergence in order to find a conservative robust design. For SUA and CSSUA-based multidisciplinary robust design, this approach requires solving the two decoupled optimization problems as given below.

$$
\begin{aligned}
& d^* = \arg\min_d \left(w * \mu_f\left(d, \mu_z^*\right) + (1-w) * \sigma_f\left(d, \mu_z^*\right)\right) && (4)\\
& s.t. \qquad LB + k\sigma(g_i(d, \mu_z^*)) \le E(g_i(d, z)) \le UB - k\sigma(g_i(d, \mu_z^*)) \ \text{ for all } i \\
& \qquad\qquad lb + k\sigma_{x/x_s} \le d \le ub - k\sigma_{x/x_s}
\end{aligned}
$$

$$
\begin{aligned}
& \mu_z^* = \arg\max_{\mu_z} \left(w * \mu_f\left(d^*, \mu_z\right) + (1-w) * \sigma_f\left(d^*, \mu_z\right)\right) && (5)\\
& s.t. \qquad Z_l \le \mu_z \le Z_u
\end{aligned}
$$

The optimization problems in Eqs. (4) and (5) are solved iteratively until convergence. Note that the robust constraint is satisfied only in Eq. (4). As mentioned earlier in Section 2.1, Eqs. (4)-(5) satisfy the system compatibility requirements through a coupled interdisciplinary uncertainty analysis for SUA-based uncertainty propagation or by solving a nested double loop formulation for CSSUA-based uncertainty propagation. This approach also requires solving a system of linear equations as shown in Eq. (2) at least once at each iteration of Eqs. (4)-(5).

<u>Single loop formulation</u>

When one or more epistemic variables cannot be treated as design variables, the single loop formulation for the robustness-based design optimization now requires solving the two decoupled optimization problems as given below.

$$d^* = \underset{d,\mu_u,\sigma_u}{\arg\min}\left(w*\mu_f\left(d,u,\mu_z^*\right)+(1-w)*\sigma_f\left(d,u,\sigma_u,\mu_z^*\right)\right) \quad (6)$$

$$\begin{aligned}
s.t. \quad & LB + k\sigma(g_i(d,u,\sigma_u,\mu_z^*)) \le E(g_i(d,u,z)) \le UB - k\sigma(g_i(d,u,\sigma_u,\mu_z^*)) \text{ for all } i \\
& lb + k\sigma_{x/x_s} \le d \le ub - k\sigma_{x/x_s} \\
& A_i\left(d,u\left(d,z^*\right),\mu_z^*\right) = 0 \quad \text{for all } i \\
& \sigma_{u_i}^2 = \sum_{\substack{j=1 \\ j\ne i}}^{n}\left(\frac{\partial F_{ui}}{\partial u_j}\right)^2\sigma_{uj}^2 + \left(\frac{\partial F_{ui}}{\partial x}\right)^2\sigma_x^2 + \left(\frac{\partial F_{ui}}{\partial x_s}\right)^2\sigma_{x_s}^2
\end{aligned}$$

$$\mu_z^* = \underset{\mu_z,\mu_u,\sigma_u}{\arg\max}\left(w*\mu_f\left(d^*,u,\mu_z\right)+(1-w)*\sigma_f\left(d^*,u,\sigma_u,\mu_z\right)\right) \quad (7)$$

$$\begin{aligned}
s.t. \quad & A_i\left(d^*,u\left(d^*,z\right),\mu_z\right) = 0 \quad \text{for all } i \\
& \sigma_{u_i}^2 = \sum_{\substack{j=1 \\ j\ne i}}^{n}\left(\frac{\partial F_{ui}}{\partial u_j}\right)^2\sigma_{uj}^2 + \left(\frac{\partial F_{ui}}{\partial x}\right)^2\sigma_x^2 + \left(\frac{\partial F_{ui}}{\partial x_s}\right)^2\sigma_{x_s}^2 \\
& Z_l \le \mu_z \le Z_u
\end{aligned}$$

Note that the optimization in Eq. (6) is a single loop formulation, which solves the design problem and the second optimization adjusts the design in presence of uncontrollable epistemic uncertainty. The optimization problems in Eqs. (6) and (7) are solved iteratively until convergence. Note that the robust constraint is satisfied only in Eq. (6). The first constraint (i.e., the system compatibility requirements) in Eq. (7) is only required if the objective function is not a function of all non-design epistemic variables.

In order to obtain solutions that are least sensitive to data uncertainty, the robustness-based design optimization formulations in Eqs. (4)-(7) have to be solved using the approach described in Chapter VII. In the following section, the proposed robustness-based design formulations are illustrated for a mathematical example and an engineering example.

3. Numerical Examples

The proposed robustness-based design optimization formulations for multidisciplinary systems are illustrated with two numerical examples: (1) a simple mathematical example and (2) an engineering problem.

3.1 Example 1: Mathematical Example

The two-disciplinary problem with feedback coupling as discussed in Chapter IX is used here. The output of the function g_2 in disciplinary analysis 2 will be used as objective function to illustrate the proposed multidisciplinary robust optimization methods.

In this case, the input random variable x_1 and x_4 are considered as non-design epistemic variable and the remaining input random variables $\{x_2, x_3, x_5\}$ are considered as design variables. The input random variables x_1 and x_2 are assumed to be described by sparse point data as given in Table 1. The input random variables x_3 and x_4 are assumed to be described by single interval data ([0.5; 1.5]) and ([2; 5]), respectively. The input random variable x_5 is assumed to be described by multiple interval data ([0.5, 1.2], [0.8, 1.5], [0.75, 1.75], [0.5, 1.75], [0.7, 1.4]). The design bounds for the input design variables are given in Table 2. Each disciplinary constraint has a lower bound of 2 and an upper bound of 20.

Table 1: Sparse Point Data for the random input variables

Sample	x_1	x_2
01	0.9567	0.9813
02	0.8334	1.0726
03	1.0125	0.9412
04	1.0288	1.2183
05	0.8854	0.9864
06	1.1191	1.0114
07	1.1189	1.1067
08	0.9962	1.0059
09	1.0327	0.9904
10	1.0175	0.9168

Table 2: Design variables and design bounds for mathematical example

No	**Variable**	**Design bounds**
1	x_2	[0, 10]
2	x_3	[0, 10]
3	x_5	[2, 10]

This problem is solved by both all-in-one and single loop formulations. The all-in-one robustness-based design formulation for this problem is as follows:

$$d^* = \arg\min_d \left(w * \frac{\mu_{g_2}}{\mu^*_{g_2}} + (1-w) * \frac{\sigma_{g_2}}{\sigma^*_{g_2}} \right) \tag{8}$$

$$s.t. \qquad LB + k\sigma(g_i(d, \mu^*_z)) \le E(g_i(d, z)) \le UB - k\sigma(g_i(d, \mu^*_z)) \ \text{ for all } i$$

$$lb + k\sigma_{x/x_s} \le d \le ub - k\sigma_{x/x_s}$$

where $\boldsymbol{d}$ is the vector of means of the input design variables.

$$\mu_z^* = \arg\max_{\mu_z} \left(w * \frac{\mu_{g_2}}{\mu_{g_2}^*} + (1-w) * \frac{\sigma_{g_2}}{\sigma_{g_2}^*} \right) \tag{9}$$

$$s.t. \qquad Z_l \le \mu_z \le Z_u$$

where the bounds Z_l and Z_u for the epistemic variables x_1 and x_4 are calculated using Eq. (9) as given in Chapter VII and by the moment bounding method described in Chapter IV, respectively.

The single loop formulation of the robustness-based design problem is as follows:

$$\begin{aligned}
\left[d^*, \mu_u^*\right] &= \arg\min_{d,\mu_u,\sigma_u} \left(w * \frac{\mu_{g_2}}{\mu_{g_2}^*} + (1-w) * \frac{\sigma_{g_2}}{\sigma_{g_2}^*} \right) \\
s.t. \quad & LB_1 + k\sigma_{g_1} \le \mu_{g_1} \le UB_1 - k\sigma_{g_1} \\
& LB_2 + k\sigma_{g2} \le \mu_{g2} \le UB_2 - k\sigma_{g_2} \\
& lb + k\sigma_x \le d \le ub - k\sigma_x \\
& u_{1,2} - \left(\left(z_1^*\right)^2 + 2x_2 - x_3 + 2\sqrt{u_{2,1}} \right) = 0 \\
& u_{2,1} - \left(z_1^* z_2^* + \left(z_2^*\right)^2 + x_5 + u_{1,2} \right) = 0 \\
& \sigma_{u_{1,2}}^2 - \left(\left(\frac{\partial F_{u_{1,2}}}{\partial u_{2,1}} \right) \sigma_{u_{2,1}}^2 + \left(\frac{\partial F_{u_{1,2}}}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial F_{u_{1,2}}}{\partial x_s} \right)^2 \sigma_{x_s}^2 \right) = 0 \\
& \sigma_{u_{2,1}}^2 - \left(\left(\frac{\partial F_{u_{2,1}}}{\partial u_{1,2}} \right) \sigma_{u_{2,1}}^2 + \left(\frac{\partial F_{u_{2,1}}}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial F_{u_{2,1}}}{\partial x_s} \right)^2 \sigma_{x_s}^2 \right) = 0
\end{aligned} \tag{10}$$

$$\begin{aligned}
\mu_z^* &= \arg\max_{\mu_z} \left(w * \frac{\mu_{g_2}}{\mu_{g_2}^*} + (1-w) * \frac{\sigma_{g_2}}{\sigma_{g_2}^*} \right) \\
s.t. \quad & \sigma_{u_{1,2}}^2 - \left(\left(\frac{\partial F_{u_{1,2}}}{\partial u_{2,1}} \right) \sigma_{u_{2,1}}^2 + \left(\frac{\partial F_{u_{1,2}}}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial F_{u_{1,2}}}{\partial x_s} \right)^2 \sigma_{x_s}^2 \right) = 0 \\
& \sigma_{u_{2,1}}^2 - \left(\left(\frac{\partial F_{u_{2,1}}}{\partial u_{1,2}} \right) \sigma_{u_{2,1}}^2 + \left(\frac{\partial F_{u_{2,1}}}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial F_{u_{2,1}}}{\partial x_s} \right)^2 \sigma_{x_s}^2 \right) = 0 \\
& Z_l \le \mu_z \le Z_u
\end{aligned} \tag{11}$$

Note that in Eq. (11), the objective function is a function of non-design epistemic variables x_1 and x_4 and therefore, the system compatibility equations are not used here. As mentioned earlier in Chapter VII, $w \geq 0$ is the weight parameter that represents the relative importance of the objectives and k is a constant that adjusts the robustness of the method against the level of conservatism of the solution. In this dissertation, k is assumed to be unity. $\mu_{g_2}^*$ and $\sigma_{g_2}^*$ are scaling factors used to normalize the two objectives in terms of mean value and standard deviation of the objective functions. The weight parameter w is varied (from 0 to 1) and the optimization formulations in Eqs. (8)-(11) are solved by the Matlab solver 'fmincon' by the methodology described in Chapter VII. The solutions are presented in Figure 1.

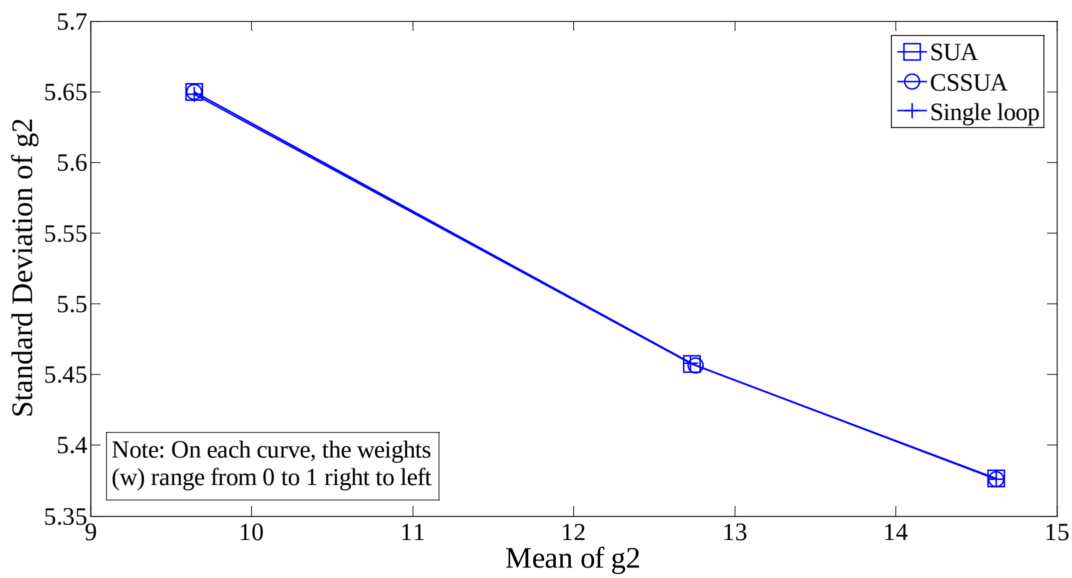


Figure 1: Robustness-based optimization for Example 1

Figure 1 shows the solutions of the conservative robust design in presence of uncontrollable epistemic uncertainty. It is seen from Figure 1 that the single loop formulation generates optimal solutions that are almost the same as the solutions obtained by SUA and CSSUA-based robust optimization methods.

The computational efforts of the different methods are compared in Table 3. It is seen that compared to both SUA and CSSUA-based optimization methods, the single loop formulation is much less expensive in terms of both function evaluations and computational time.

Table 3: Computational effort for different methods for Example 1

All-in-one						Single loop		
SUA			CSSUA					
DA	SA	CT	DA	SA	CT	DA	SA	CT
11588	454	17.53118	15504	0	14.804068	1600	0	3.037412

Note: DA = Disciplinary analysis   SA = System analysis   CT = Computational time in seconds

3.2 Example 2: Engineering Problem (FireSat)

The same FireSat problem as described in Chapters I and IX is used here. The output P_{tot} of disciplinary analysis 3 (Power subsystem) will be used as an objective function to illustrate the proposed multidisciplinary robust optimization methods. The objective is to simultaneously minimize the mean value of the total power consumption, P_{tot} and its standard deviation. The uncertain variables involved in each subsystem and their corresponding single interval data are given in Table 4.

Table 4: Uncertain variables and data for FireSat problem

No	Variable	Symbol	Unit	Data
1	Earth's radius	R_E	m	[6378135, 6378145]
2	Power other than ACS	P_{other}	W	[825, 1375]
3	Avg solar flux	F_s	W/m^2	[1326, 1481]
4	Deviation of major moment axis from local vertical	θ	deg	[10, 19]
5	Moment arm for solar radiation torque	L_{sp}	m	[0, 3.75]
6	Reflectance factor	q		[0.1, 0.99]
7	Residual dipole of the space craft	R_D	Am^2	[0, 10]
8	Moment arm for aerodynamic torque	L_a	m	[0, 3.75]
9	Drag coefficient	C_d		[2, 4]

For the sake of illustration, in this example problem, the following epistemic variables are considered as design variables with the design bounds given in Table 5 below. Note that the design variables q and C_d are function of other design variables of the original problem, i.e., the FireSat problem consisting of all the subsystems. In this paper, a simplified three disciplinary problem has been used. Therefore, these variables are considered here as design variables.

Table 5: Design variables and design bounds for FireSat problem

No	Variable	Symbol	Design bounds
1	Power other than ACS	P_{other}	[500, 1500]
2	Deviation of major moment axis from local vertical	θ	[0, 90]
3	Moment arm for solar radiation torque	L_{sp}	[0, 20]
4	Reflectance factor	q	[0, 1]
5	Moment arm for aerodynamic torque	L_a	[0, 10]
6	Drag coefficient	C_d	[1, 8]

This problem has six epistemic design variables and three epistemic non-design variables. This problem is solved by both all-in-one and single loop formulations.

The all-in-one robustness-based design formulation for this problem is as follows:

$$
\begin{aligned}
& d^{*} = \arg\min_{d} \left(w * \frac{\mu_{P_{tot}}}{\mu^{*}_{P_{tot}}} + (1-w) * \frac{\sigma_{P_{tot}}}{\sigma^{*}_{P_{tot}}} \right) \\
& s.t. \qquad LB_1 + k\sigma_{A_{sa}} \le \mu_{A_{sa}} \le UB_1 - k\sigma_{A_{sa}} \\
& \qquad\quad LB_2 + k\sigma_{\tau_{tot}} \le \mu_{\tau_{tot}} \le UB_2 - k\sigma_{\tau_{tot}} \\
& \qquad\quad lb + k\sigma_x \le d \le ub - k\sigma_x \qquad \text{for } i = 1,2,\ldots,6
\end{aligned}
\tag{12}
$$

where ***d*** is the vector of means of the input design variables.

$$\mu_z^* = \arg\max_{\mu_z}\left(w * \frac{\mu_{P_{tot}}}{\mu^*_{P_{tot}}} + (1-w) * \frac{\sigma_{P_{tot}}}{\sigma^*_{P_{tot}}} \right) \tag{13}$$
$$s.t. \qquad Z_l \le \mu_{z_i} \le Z_u \qquad \text{for } i = 1,2,3$$

where the bounds Z_l and Z_u for the epistemic variables R_E, F_s and R_D are calculated by the method described in Chapter IV.

The single loop formulation of the robustness-based design problem is as follows:

$$d^* = \arg\min_{d,\mu_u,\sigma_u}\left(w * \frac{\mu_{P_{tot}}}{\mu^*_{P_{tot}}} + (1-w) * \frac{\sigma_{P_{tot}}}{\sigma^*_{P_{tot}}} \right) \tag{14}$$
$$\begin{aligned}
s.t. \qquad & LB_1 + k\sigma_{A_{sa}} \le \mu_{A_{sa}} \le UB_1 - k\sigma_{A_{sa}} \\
& LB_2 + k\sigma_{\tau_{tot}} \le \mu_{\tau_{tot}} \le UB_2 - k\sigma_{\tau_{tot}} \\
& lb + k\sigma_x \le d \le ub - k\sigma_x \qquad \text{for } i = 1,2,\ldots,6 \\
& P_{ACS} - (\tau_{tot}\omega_{\max} + nP_{hold}) = 0 \\
& I_{\min} - \min(I_{tot,X}, I_{tot,Y}, I_{tot,Z}) = 0 \\
& I_{\max} - \max(I_{tot,X}, I_{tot,Y}, I_{tot,Z}) = 0 \\
& \sigma_{u_1}^2 - \left(\left(\frac{\partial P_{ACS}}{\partial u_2}\right)\sigma_{u_2}^2 + \left(\frac{\partial P_{ACS}}{\partial u_3}\right)\sigma_{u_3}^2 + \left(\frac{\partial P_{ACS}}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial P_{ACS}}{\partial x_s}\right)^2 \sigma_{x_s}^2 \right) = 0 \\
& \sigma_{u_2}^2 - \left(\left(\frac{\partial I_{\min}}{\partial u_1}\right)\sigma_{u1}^2 + \left(\frac{\partial I_{\min}}{\partial u_3}\right)\sigma_{u_3}^2 + \left(\frac{\partial I_{\min}}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial I_{\min}}{\partial x_s}\right)^2 \sigma_{x_s}^2 \right) = 0 \\
& \sigma_{u_3}^2 - \left(\left(\frac{\partial I_{\max}}{\partial u_1}\right)\sigma_{u_1}^2 + \left(\frac{\partial I_{\max}}{\partial u_2}\right)\sigma_{u_2}^2 + \left(\frac{\partial I_{\max}}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial I_{\max}}{\partial x_s}\right)^2 \sigma_{x_s}^2 \right) = 0
\end{aligned}$$

$$\mu_z^* = \underset{\mu_z, \mu_u, \sigma_u}{\arg\max} \left(w * \frac{\mu_{P_{tot}}}{\mu_{P_{tot}}^*} + (1-w) * \frac{\sigma_{P_{tot}}}{\sigma_{P_{tot}}^*} \right) \tag{15}$$

$$\begin{aligned}
s.t. \quad & P_{ACS} - \left(\tau_{tot}\omega_{\max} + nP_{hold}\right) = 0 \\
& I_{\min} - \min\left(I_{tot,X}, I_{tot,Y}, I_{tot,Z}\right) = 0 \\
& I_{\max} - \max\left(I_{tot,X}, I_{tot,Y}, I_{tot,Z}\right) = 0 \\
& \sigma_{u_1}^2 - \left(\left(\frac{\partial P_{ACS}}{\partial u_2} \right) \sigma_{u_2}^2 + \left(\frac{\partial P_{ACS}}{\partial u_3} \right) \sigma_{u_3}^2 + \left(\frac{\partial P_{ACS}}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial P_{ACS}}{\partial x_s} \right)^2 \sigma_{x_s}^2 \right) = 0 \\
& \sigma_{u_2}^2 - \left(\left(\frac{\partial I_{\min}}{\partial u_1} \right) \sigma_{u1}^2 + \left(\frac{\partial I_{\min}}{\partial u_3} \right) \sigma_{u_3}^2 + \left(\frac{\partial I_{\min}}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial I_{\min}}{\partial x_s} \right)^2 \sigma_{x_s}^2 \right) = 0 \\
& \sigma_{u_3}^2 - \left(\left(\frac{\partial I_{\max}}{\partial u_1} \right) \sigma_{u_1}^2 + \left(\frac{\partial I_{\max}}{\partial u_2} \right) \sigma_{u_2}^2 + \left(\frac{\partial I_{\max}}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial I_{\max}}{\partial x_s} \right)^2 \sigma_{x_s}^2 \right) \\
& Z_l \le \mu_{z_i} \le Z_u \qquad \text{for } i = 1,2,3
\end{aligned}$$

For this example problem, only single interval data is available for the input design variables as given in Table 2. The two disciplinary constraints are assumed to have lower bounds of 20 and 0.09 and upper bounds of 50 and 0.4, respectively. The weight parameter *w* is varied (from 0 to 1) and the optimization formulations in Eqs. (12)-(15) are solved by the Matlab solver 'fmincon' for sparse point and interval data. The solutions are presented in Figure 2.

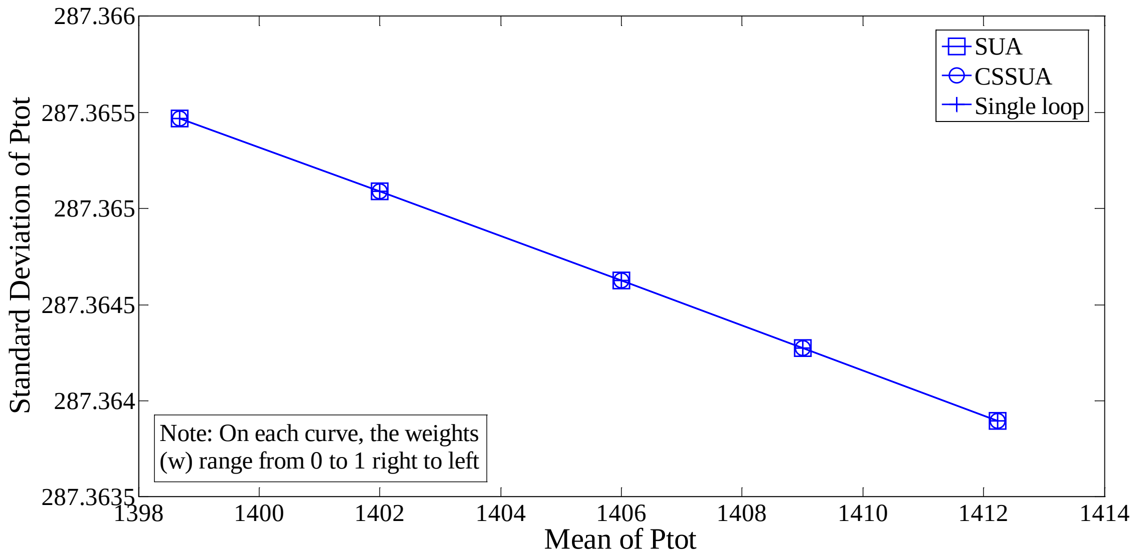


Figure 2: Robustness-based optimization for the FireSat problem

Figure 2 shows the solutions of the conservative robust design in presence of uncontrollable epistemic uncertainty. It is seen from Figure 2 that the single loop formulation generates optimal solutions that are almost the same as the solutions obtained by SUA and CSSUA-based robust optimization methods.

The computational efforts of the different methods are compared in Table 6. It is seen that compared to both SUA and CSSUA-based optimization methods, the single loop formulation is much less expensive in terms of both function evaluations and computational time.

Table 6: Computational effort for different methods for FireSat problem

All-in-one						Single loop		
SUA			CSSUA					
DA	SA	CT	DA	SA	CT	DA	SA	CT
2170	640	28.9766	28090	0	26.6426	1080	0	12.1503

Note: DA = Disciplinary analysis SA = System analysis CT = Computational time in seconds

4. Conclusion

This chapter has developed formulations for multidisciplinary robustness-based design optimization under data uncertainty. Two types of data uncertainty – sparse point data and interval data – are considered. A single loop approach is used for the design optimization, which does not require any explicit coupled multidisciplinary uncertainty propagation analysis. Thus the computational complexity and cost involved in estimating the mean and variation of the performance function is greatly reduced. A decoupled approach is proposed in this chapter to un-nest the robustness-based design from the analysis of non-design epistemic variables to achieve further computational efficiency. The computational efficiency of the proposed formulations is demonstrated by a mathematical and an engineering example problems considering the number of individual disciplinary analyses, number of system level analyses, and the overall computational time. The selection of the method may depend on the number of system level analysis as well as the disciplinary analysis and the computational time required. CSSUA-based method may be preferable over the SUA-based method if the system level analysis is computationally expensive, and the individual disciplinary analyses are more affordable. However, in both examples, the single loop formulation appears to be more efficient as it requires no integrated system level analysis and the number of individual disciplinary analyses as well as the computational time required are much less. Due to the use of a probabilistic format to represent all the uncertain variables, the proposed multidisciplinary robustness-based design optimization methodology facilitates the implementation of multidisciplinary reliability-based design optimization, which is a challenging problem in presence of epistemic uncertainty.

CHAPTER XI

RELIABILITY-BASED DESIGN OPTIMIZATION (RBDO) OF MULTIDISCIPLINARY SYSTEM UNDER EPISTEMIC UNCERTAINTY

1. Introduction

This chapter proposes formulations and algorithms for reliability-based design optimization (RBDO) of multidisciplinary systems under both aleatory uncertainty (i.e., natural or physical variability) and epistemic uncertainty (i.e., imprecise probabilistic information). The proposed formulations specifically deal with epistemic uncertainty arising from sparse point data and interval data. An efficient decoupled approach is proposed that un-nests the design analysis from the epistemic analysis. The proposed methodology for multidisciplinary systems does not require any coupled system level analysis. The proposed methods are illustrated for a mathematical problem and a practical engineering problem.

As mentioned in Chapter II, most of the existing methods are based on non-probabilistic theory and can handle only single discipline problems. Many of these methods need additional non-probabilistic formulations to incorporate epistemic uncertainty into the design optimization framework, which may be computationally expensive. However, if the epistemic uncertainty can be converted to a probabilistic format, the need for these additional formulations is avoidable, and well established probabilistic methods of RBDO can be used. Therefore, there is a need for an efficient RBDO methodology that deals with both aleatory and epistemic uncertainty.

The contribution of this chapter is to develop a methodology for RBDO for multidisciplinary systems that includes both aleatory and epistemic uncertainty. This chapter specifically focuses on epistemic uncertainty arising from *sparse point data* and *interval data*. In this chapter, we propose an efficient decoupled approach that un-nests the design analysis from the epistemic analysis. The proposed methodology for multidisciplinary RBDO does not require any coupled system level analysis.

The rest of the chapter is organized as follows. Section 2 extends the methodology for single discipline system as developed in Chapter VIII to multidisciplinary system. In Section 3, we illustrate the proposed methods for a number of example problems. Section 4 provides conclusions and suggestions for future work.

2. RBDO for multidisciplinary systems

As mentioned in Chapter IX, in order to achieve feasibility in multidisciplinary system analysis, the non-linear equations shown in Eq. (7) in Chapter IX have to be solved simultaneously.

Consider the following MDO formulation:

$$\begin{aligned} &\min_{x} \quad f\left(x,u(x)\right) \\ &s.t. \quad g\left(x,u(x)\right)\leq 0 \end{aligned} \tag{1}$$

In addition to satisfying the design constraints, MDO in Eq. (1) requires that the system compatibility among the disciplines in Eq. (7) in Chapter IX is also satisfied. Several methods are available for multidisciplinary optimization based on how the system analysis is handled, namely the multidisciplinary feasibility (MDF) method, the all-at-once (AAO) method, and the individual disciplinary feasibility (IDF) method (Cramer et al, 1994). All these methods have their own advantages and limitations.

2.1 Multidisciplinary RBDO

Now, consider the following probabilistic variation of Eq. (1)

$$\begin{aligned} &\min_{x} \quad f\left(x,u(x)\right) \\ &s.t. \;\; P\left(g\left(x,u(x)\right)\right) \le \alpha \end{aligned} \tag{2}$$

In Eq. (2), all or some of the design variables are random design variables. Like Eq. (1), Eq. (2) also requires satisfying the system compatibility requirements as shown in Eq. (7) in Chapter IX, in addition to satisfying the reliability constraints.

As mentioned in Chapter VIII, there exist different combinations of methods to solve singe disciplinary RBDO. Each of these combinations can be used with different MDO strategies, namely the MDF, AAO, or IDF method to handle the multidisciplinary system analysis. Therefore, a multidisciplinary RBDO problem of Eq. (2) can be solved by several combinations of methods (Chiralaksanakul and Mahadevan, 2007). All these methods have their own advantages and limitations. A detailed discussion of different RBDO methods for multidisciplinary systems can be found in Chiralaksanakul and Mahadevan (2007) and Smith (2007).

In this chapter, we use the RBDO/AAO method to develop the methodology for multidisciplinary RBDO under epistemic uncertainty. In RBDO/AAO method, the design formulation in Eq. (2) becomes:

$$\min_{x} \quad f(x,u(x))$$
$$s.t.\ \ P(g(x,u(x))) \le \alpha$$
$$A_i(x,u(x)) = 0 \qquad (3)$$

Note that in Eq. (3), the system compatibility requirement is used as constraints in the design optimization formulation. The reliability analysis required for estimating the reliability constraints in Eq. (3) is done as follows:

$$\min \beta = \sqrt{(Y)^T (Y)}$$
$$s.t.\ \ g_Y(Y, u_Y(Y)) = 0$$
$$A_i(Y, u_Y(Y)) = 0 \qquad (4)$$

where $\boldsymbol{Y}$ denotes all the random input variables of the system in uncorrelated standard normal space. Functions g_Y and $\boldsymbol{u_Y}$ are transformed functions such that $g_Y(Y) = g(T^{-1}(x))$ where T is the transformation function from original space, $\boldsymbol{x}$, to standard normal space $\boldsymbol{Y}$. The system compatibility requirements $A_i(Y, u_Y(Y)) = 0$, are included in Eq. (4) to ensure system compatibility in multidisciplinary reliability analysis.

The above mentioned formulation of multidisciplinary reliability is known as collaborative reliability analysis in Du and Chen (2005). Mahadevan and Smith (2005) proposed an efficient approach to solving Eq. (4), namely multi-constraint FORM for multidisciplinary reliability analysis. In this chapter, Eq. (4) is used within the multidisciplinary RBDO framework under epistemic uncertainty to evaluate the reliability constraints.

2.2 Multidisciplinary RBDO under epistemic uncertainty

As discussed in Chapter VIII, the inclusion of epistemic uncertainty in RBDO adds another level of complexity in the design methodology. Multidisciplinary RBDO under aleatory uncertainty alone is a computationally challenging problem. The inclusion of epistemic uncertainty in multidisciplinary RBDO further multiplies this computational effort. In Chapter VIII, we have proposed an approach that decouples the uncertainty analysis of the epistemic non-design variables from the design optimization problem. The same approach is used here for the multidisciplinary problem as follows:

As in Eq. (3), the general problem of multidisciplinary RBDO can be expressed as follows:

$$\begin{aligned} &\min_{d,u}\left(\max_{\mu_z} f(d,u,\mu_z)\right) \\ &s.t. \qquad p_{f_i} = P\left(g_i(X,Z) \le 0\right) < p_i \quad \text{for } i = 1,2,\dots,k \\ &\qquad\qquad Z_l \le \mu_z \le Z_u \\ &\qquad\qquad A_i\left(d, u(d,z), \mu_z\right) = 0 \quad \text{for all } i \end{aligned} \tag{5}$$

This nested optimization problem can be decoupled and expressed as:

$$\begin{aligned} &d^* = \arg\min_{d,u}\left(f\left(d,u,\mu_z^*\right)\right) \\ &s.t. \quad p_{f_i} = P\left(g_i\left(X,\mu_z^*\right) \le 0\right) < p_i \quad \text{for } i = 1,2,\dots,k \\ &\qquad A_i\left(d, u\left(d,z^*\right), \mu_z^*\right) = 0 \quad \text{for all } i \end{aligned} \tag{6}$$

$$\begin{aligned} &\mu_z^* = \arg\max_{\mu_z,u}\left(f\left(d^*,u,\mu_z\right)\right) \\ &s.t. \quad A_i\left(d^*, u\left(d^*,z\right), \mu_z\right) = 0 \quad \text{for all } i \\ &\qquad Z_l \le \mu_z \le Z_u \end{aligned} \tag{7}$$

The optimization problems in Eqs. (6) and (7) are solved iteratively until convergence. Note that the reliability constraint is satisfied only in Eq. (6). The first constraint (i.e., the system compatibility equations) in Eq. (7) is required to ensure that the optimization is driven by all non-design epistemic variables, because sometimes the objective function may not be a function of all non-design epistemic variables. In cases when the objective function is the function of all non-design epistemic variables, this constraint is not required.

We have developed the methodology of solving single discipline RBDO problem under both sparse point and interval data in Chapter VIII. The same methodology is used to solve the multidisciplinary RBDO problem under epistemic uncertainty. In the following section, the proposed RBDO formulations are illustrated for multidisciplinary example problems.

3. Numerical Examples

3.1 Example 1: Mathematical Example

The two-disciplinary problem with feedback coupling as discussed in Chapter IX is used here. The output of the function g_2 in disciplinary analysis 2 will be used as objective function to illustrate the proposed multidisciplinary RBDO method. Each input design variable has a lower bound of 0.001 and an upper bound of 10. A limit state is defined as $g = g_1 - g_{1,0}$ and failure is defined when $g < 0$. Here, $g_{1,0}$ is assumed to be 5.

The general formulation for this RBDO problem is as follows:

$$\begin{aligned} &\min \quad g_2 \\ &s.t. \;\; P\left(g_1 \le g_{1,0}\right) < p_0 \end{aligned} \tag{8}$$

where p_0 is assumed to be 0.0062 (β=2.5) in this example. In this example problem, the probability $P\ (g_1 \le g_{1,0})$ depends on all the random design variables $\boldsymbol{x}$.

In this case, the input random variable x_1 is considered as non-design epistemic variable and the remaining input random variables $\{x_2, ..., x_5\}$ are considered as design variables. The input random variables are assumed to be described by single interval data. Each input random variable ranges from 0.5 to 1.5. Bounds on the mean for the epistemic variable x_1 and bounds on the variances of all the random variables x are estimated by the methods described in Chapter IV. Since this problem contains non-design epistemic variables, this problem is solved by the RBDO methodology developed in Section 2 by solving the following two optimization problems iteratively until convergence and the solutions are given in Table 1.

$$\begin{aligned} &\left[d^*, u^*\right] = \arg\min_{d,u} \quad g_2\left(d, u, \mu_z^*\right) \\ &s.t. \qquad P\left(g_1\left(d, u, \mu_z^*\right) \le g_{1,0}\right) < p_0 \\ &\qquad\qquad u_{1,2} - \left(\left(z^*\right)^2 + 2x_2 - x_3 + 2\sqrt{u_{2,1}}\right) = 0 \\ &\qquad\qquad u_{2,1} - \left(z^* x_4 + x_4^2 + x_5 + u_{1,2}\right) = 0 \\ &\qquad\qquad lb \le d \le ub \end{aligned} \tag{9}$$

$$\begin{aligned} &\mu_z^* = \arg\max_{\mu_z} \quad g_2\left(d^*, u^*, \mu_z\right) \\ &s.t. \quad Z_l \le \mu_{z_i} \le Z_u \end{aligned} \tag{10}$$

where Z_l and Z_u are the bounds on the mean value of the non-design epistemic variable z. Note that in Eq. (10), the objective function is a function of non-design epistemic variable x_1 and therefore, the system compatibility equations are not used here.

Table 1: Optimal design solution for the mathematical problem

Optimum ***x***	g_2	No of analyses					
		Design Analysis (Eq. (9))		Epistemic Analysis (Eq. (10))		Total	
		DA	SA	DA	SA	DA	SA
(2.2436, 2.6628, 0.0010, 0.0010)	1.2300	13,846	0	8	0	13,854	0

Note: DA = Disciplinary analysis   SA = System analysis

The optimizations in Eqs. (9) and (10) required only 2 iterations between the design problem (Eq. (9)) and the uncertainty analysis for the non-design epistemic variables (Eq. (10)) for convergence. Number of function evaluations in terms of disciplinary analysis (DA) and system analysis (SA) for both the design and epistemic analyses are listed in Table 1 for future reference. It is seen in Table 1 that the proposed RBDO methodology can solve this design problem with only 13,854 disciplinary analyses, of which only 8 evaluations are required for the epistemic analyses and only 13,846 evaluations are required for the design analyses. If this example problem involved only aleatory uncertainty, the number of function evaluation would be approximately half of 13,846. Therefore, the proposed RBDO methodology under epistemic uncertainty can solve this problem with a reasonably increased number of function evaluations.

3.2 Example 2: Engineering Problem (FireSat)

The same FireSat problem as described in Chapters I, IX and X is used here. The output P_{tot} of disciplinary analysis 3 (Power subsystem) will be used as the objective function to illustrate the proposed multidisciplinary RBDO method. The uncertain variables involved in each subsystem and their corresponding single interval data are given in Table 4 in Chapter X. The design bounds for the design variables are given in Table 5 in Chapter X. In this example, it is assumed that all the input random variables have log-normal distributions, moments of which are estimated from the single interval data given in Table 4 in Chapter X.

The limit states are defined as $g_1 = A_{sa} - A_{sa,0}$ and $g_2 = \tau_{tot} - \tau_{tot,0}$ and failures are defined when $g_1 > 0$ and $g_2 > 0$. Here, $A_{sa,0}$ and $\tau_{tot,0}$ are assumed to be 50 and 0.35, receptively.

The general formulation for this RBDO problem is as follows:

$$\begin{aligned} \min \quad & P_{tot} \\ s.t. \quad & P\left(A_{sa} \geq A_{sa,0}\right) < p_{0,1} \\ & P\left(\tau_{tot} \geq \tau_{tot,0}\right) < p_{0,2} \end{aligned} \tag{11}$$

where $p_{0,1}$ and $p_{0,2}$ are assumed to be 0.0062 (β=2.5) each. In this example problem, the probabilities $P\left(A_{sa} \geq A_{sa,0}\right)$ and $P\left(\tau_{tot} \geq \tau_{tot,0}\right)$ depend on all the random input variables $\boldsymbol{x}$.

This problem has six epistemic design variables and three epistemic non-design variables. Bounds on the mean for the non-design epistemic variables R_E, F_s, R_D and bounds on the variances of all the random variables $\boldsymbol{x}$ are estimated by the methods described in Chapter IV. Since this problem contains non-design epistemic variables, this problem is solved by the RBDO methodology developed in Section 2 by solving the following two optimization problems iteratively until convergence and the solutions are given in Table 2.

$$
\begin{aligned}
d^* = \arg\min_{d,u} \quad & P_{tot}\left(d,u,\mu_z^*\right) \\
s.t. \quad & P\left(A_{sa}\left(d,u,\mu_z^*\right) \geq A_{sa,0}\right) < p_{0,1} \\
& P\left(\tau_{tot}\left(d,u,\mu_z^*\right) \geq \tau_{tot,0}\right) < p_{0,2} \\
& P_{ACS} - \left(\tau_{tot}\omega_{\max} + nP_{hold}\right) = 0 \\
& I_{\min} - \min\left(I_{tot,X}, I_{tot,Y}, I_{tot,Z}\right) = 0 \\
& I_{\max} - \max\left(I_{tot,X}, I_{tot,Y}, I_{tot,Z}\right) = 0 \\
& lb \leq d \leq ub
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
\mu_z^* = \arg\max_{\mu_z,u} \quad & P_{tot}\left(d^*,u,\mu_z\right) \\
s.t. \quad & P_{ACS} - \left(\tau_{tot}\omega_{\max} + nP_{hold}\right) = 0 \\
& I_{\min} - \min\left(I_{tot,X}, I_{tot,Y}, I_{tot,Z}\right) = 0 \\
& I_{\max} - \max\left(I_{tot,X}, I_{tot,Y}, I_{tot,Z}\right) = 0 \\
& Z_l \leq \mu_{z_i} \leq Z_u
\end{aligned}
\tag{13}
$$

where Z_l and Z_u are are the bounds on the mean values of the non-design epistemic variables $\boldsymbol{z}$.

Table 2: Optimal design solution for the FireSat problem

Optimum $\boldsymbol{x}$	P_{tot}	No of analyses					
		Design Analysis (Eq. (12))		Epistemic Analysis (Eq. (13))		Total	
		DA	SA	DA	SA	DA	SA
(700, 5, 3.6192, 0.3792, 1.0571, 1)	782.424	75,702	0	378	0	76,080	0

Note: DA = Disciplinary analysis   SA = System analysis

The optimizations in Eqs. (12) and (13) required only 2 iterations between the design problem (Eq. (12)) and the uncertainty analysis for the non-design epistemic variables (Eq. (13)) for convergence. Number of function evaluations in terms of disciplinary analysis (DA) and system analysis (SA) for both the design and epistemic analyses are listed in Table 6 for future reference. It is seen in Table 2 that the proposed RBDO methodology can solve this design problem with only 76,080 disciplinary analyses, of which only 378 evaluations are required for the epistemic analyses and only 75,702 evaluations are required for the design analyses. If this example problem involved only aleatory uncertainty, the number of function evaluation would be approximately half of 75,702. Therefore, the proposed RBDO methodology under epistemic uncertainty can solve this problem with a reasonably increased number of function evaluations.

4. Conclusion

This chapter has developed formulations for reliability-based design optimization (RBDO) for both multidisciplinary systems under both aleatory and epistemic uncertainty on the data of the random variables. Two types of data uncertainty – sparse point data and interval data – are considered. The computational efficiency of the proposed formulations is demonstrated with a number of example problems considering the number of individual disciplinary analyses.

The proposed RBDO methodology does not require any coupled system level analysis. The huge computational expense required for the epistemic analysis is reduced by decoupling the design analysis from the epistemic analysis. Unlike existing methods, it does not use separate representations for aleatory and epistemic uncertainties and does not require nested analysis. Both types of uncertainty are treated in a unified manner using a probabilistic format, thus reducing the computational effort and simplifying the optimization problem. The numerical examples in this chapter were carried out using the classical nested loop RBDO formulation and the number of function evaluations needed in each case was reported in Section 3. The focus of this chapter is not on efficiency, but on the inclusion of epistemic uncertainty in the design optimization. Several more efficient RBDO methods (single loop and sequential) have been developed in recent years, and all these methods can be enhanced to incorporate epistemic uncertainty. Future work in this direction also needs to include system reliability constraints, and the multi-level nature of the multidisciplinary systems.

CHAPTER XII

SUMMARY AND FUTURE NEEDS

Summary of Contributions

In order to design reliable complex systems, it is necessary that the design process accounts for all forms of uncertainty and ensures that the reliability targets are satisfied throughout all stages of design. In this dissertation, efficient methods are developed to incorporate uncertainty in the design of complex and multidisciplinary systems. This has been done in two ways. First, this dissertation developed efficient uncertainty representation and propagation methods for both single and multidisciplinary systems under epistemic uncertainty. Second, efficient design optimization methods, addressing both robustness and reliability, are developed for both single and multidisciplinary systems under epistemic uncertainty.

Objective 1 of this dissertation was to develop efficient methods to represent epistemic uncertainty arising from sparse point data and interval data. Chapters III, IV and VI of this dissertation developed efficient uncertainty representation methods using a flexible family of Johnson distributions to achieve this objective.

Objective 2 was to develop efficient uncertainty propagation methods under epistemic uncertainty. Chapters III, V and VI of this dissertation achieved this objective by developing both sampling and optimization-based uncertainty propagation methods.

Chapter III developed a methodology for propagating both aleatory and epistemic uncertainty arising from sparse point data through computational models of system response. This method eliminates the computationally expensive process of nesting an aleatory uncertainty analysis inside an epistemic uncertainty analysis. This methodology also affords sensitivity analysis information with regard to each of the distribution parameters as well as the basic random variables. The results of the sensitivity analysis give quantitative guidance regarding data collection for the random variables.

Chapter IV developed a probabilistic approach to represent interval data for input variables in reliability and uncertainty analysis problems. The proposed probabilistic framework of handling interval data can be applied for a combined treatment of aleatory and epistemic input uncertainties from the perspective of uncertainty propagation or reliability based design. This approach to uncertainty representation given interval data can allow for computationally efficient propagation by avoiding the nested analysis that is typically performed in the presence of interval variables.

Chapter V developed a probabilistic approach for uncertainty representation and propagation in system analysis, when the information on the uncertain input variables and/or their distribution parameters may be available as either probability distributions or simply intervals (single or multiple). Two methods are explored for the implementation of the proposed approach, based on: (1) sampling and (2) optimization. The sampling based strategy is more expensive and tends to underestimate the output bounds. The optimization based methodology improves both aspects. The proposed approach facilitates the implementation of design optimization under uncertainty using efficient reliability-based design optimization (RBDO) methods, e.g., single loop, decoupled, etc., due to the use of a probabilistic format to represent all the uncertain variables.

Chapter VI developed a methodology for multivariate input modeling of random variables by using a four parameter flexible Johnson family of distributions for the marginals that also accounts for data uncertainty. Semi-empirical formulas in terms of the Johnson marginals and covariances are presented to estimate the model parameters. This multivariate input model is particularly suitable for uncertainty quantification problems that contain both aleatory and data uncertainty. A computational framework is developed to consider correlations among basic random variables as well as among their distribution parameters. A methodology is developed for propagating both aleatory and data uncertainty arising from sparse point data and interval data through computational models of system response. The proposed approach facilitates the implementation of design optimization under uncertainty considering correlations.

Objective 3 of this dissertation was to develop efficient design optimization methods under epistemic uncertainty arising from sparse point data and interval data. Chapters VII and VIII of this dissertation achieved this objective by developing robustness and reliability-based design optimization methods under epistemic uncertainty, respectively.

Chapter VII developed formulations and algorithms for design optimization under both aleatory and epistemic uncertainty, from the perspective of system robustness. A decoupled approach is proposed in this dissertation to un-nest the robustness-based design from the analysis of non-design epistemic variables to achieve computational efficiency. As collecting more data reduces data uncertainty but increases expenses, the effect of sample size on the optimality and the robustness of the solution is also studied. A method is also presented to determine the optimal sample size for sparse point data that leads to the solutions of the design problem that are least sensitive to variations in the design variables. The major advantage of the proposed methodology is that unlike existing methods, it does not use separate representations for aleatory and epistemic uncertainties and does not require nested analysis. Both types of uncertainty are treated in a unified manner using a probabilistic format, thus reducing the computational effort and simplifying the optimization problem. The results regarding robustness of the design versus data size are valuable to the decision maker. The design optimization procedure also optimizes the sample size, thus facilitating resource allocation for data collection efforts. Due to the use of a probabilistic format to represent all the uncertain variables, the proposed robustness-based design optimization methodology facilitates the implementation of multidisciplinary robustness-based design optimization, which is a challenging problem in presence of epistemic uncertainty.

Chapter VIII developed formulations and algorithms for reliability-based design optimization (RBDO) for single discipline systems under both aleatory uncertainty and epistemic uncertainty. An efficient decoupled approach is proposed that un-nests the design analysis from the epistemic analysis. The huge computational expense required for the epistemic analysis is reduced by decoupling the design analysis from the epistemic analysis. Unlike existing methods, it does not use separate representations for aleatory and epistemic uncertainties and does not require nested analysis. Both types of uncertainty are treated in a unified manner using a probabilistic format, thus reducing the computational effort and simplifying the optimization problem.

Objective 4 of this dissertation was to develop efficient uncertainty propagation methods for multidisciplinary systems under epistemic uncertainty. Chapter IX of this dissertation developed efficient optimization-based uncertainty propagation methods for multidisciplinary systems to achieve this objective.

Chapter IX developed an efficient probabilistic approach for uncertainty propagation in multidisciplinary system analysis, when the information on the uncertain input variables may be available as either sparse point data or as intervals (single or multiple). A decoupled approach is used in this dissertation to un-nest the system analysis from the probabilistic analysis to achieve computational efficiency. This approach uses deterministic optimization to first quantify the uncertainty in the coupling variables. No coupled system level analysis is required. The proposed methods are equally applicable with both sampling and analytical approximation-based reliability analysis methods. Due to the use of a probabilistic format to represent all the uncertain variables, the proposed uncertainty propagation framework facilitates the implementation of multidisciplinary design optimization in the presence of both aleatory and epistemic uncertainty.

Objective 5 of this dissertation was to develop efficient design optimization methods for multidisciplinary systems under epistemic uncertainty. Chapters X and XI of this dissertation achieved this objective by developing robustness and reliability-based design optimization methods for multidisciplinary systems under epistemic uncertainty, respectively.

Chapter X developed formulations and algorithms for design optimization for multidisciplinary systems under both aleatory and epistemic uncertainty, from the perspective of system robustness. A single loop approach is used for the design optimization, which does not require any explicit interdisciplinary uncertainty propagation and thus the computational complexity and cost involved in estimating the mean and variation of the performance function is greatly reduced. A decoupled approach is proposed to un-nest the robustness-based design from the analysis of non-design epistemic variables to achieve further computational efficiency.

Chapter XI extended the RBDO methodology for single discipline system developed in Chapter VIII to multidisciplinary systems. The proposed RBDO methodology does not require any coupled system level analysis. The huge computational expense required for the epistemic analysis is reduced by decoupling the design analysis from the epistemic analysis.

In summary, the methodologies developed in this dissertation will allow engineers to comprehensively account for different types of uncertainty relevant to the design of multidisciplinary systems, and perform multidisciplinary design analysis under physical, and data uncertainty. The broader impact of this research includes (1) Stimulating new directions for modeling epistemic uncertainty, (2) Development of new methods and algorithms for design optimization under epistemic uncertainty, and (3) Application to multidisciplinary systems encountered in aerospace engineering, automobile design, and other domains that can use model-based reliability analysis and design optimization.

Future Research Needs

The short-term research needs are as follows. As mentioned in Chapter II, epistemic uncertainty can be viewed in two ways. It can be defined with reference to a stochastic but poorly known quantity or with reference to a fixed but poorly known physical quantity. This dissertation focuses on handling the first definition of epistemic uncertainty i.e., epistemic uncertainty with reference to a stochastic but poorly known quantity in a straightforward manner, as the uncertainty representation methods proposed in this dissertation are purely probabilistic, resulting in a family of probability distributions. However, the second definition of epistemic uncertainty i.e., epistemic uncertainty with reference to a fixed but poorly known quantity can also be managed using the probabilistic methods as can be found in Helton et al (2004) and Helton et al (2008), though the implications of probability distributions for the representation of this type of epistemic uncertainty merit further investigation. Following Helton et al (2004) and Helton et al (2008), the proposed methods can also handle this second definition of epistemic uncertainty. It would be worthwhile to investigate this issue of using probability theory for the second definition of epistemic uncertainty.

This dissertation specifically focuses on epistemic uncertainty arising from sparse point and interval data. However, as mentioned in Chapter I, epistemic uncertainty can also arise from other sources, for example, model error. The methodologies developed in this dissertation need to be extended to include other sources of epistemic uncertainty.

In this dissertation, uncertainty propagation methods are developed, which can handle either sparse point data or interval data. However, in practice, a mixture of both sparse point and interval data could be available for the same variable, or one or more variable might be described by sparse point data and the others might be described by interval data. The methods developed in this dissertation are capable of handling such cases; however, the uncertainty propagation methods developed in this dissertation have not been illustrated to solve such problems. In future, it would be worthwhile to solve such problems using the developed methods.

In this dissertation, design optimization methods are developed assuming independence among input random variables. However, intervariable dependencies or statistical correlations might have significant impact on the results of the design optimization. Correlations may also exist among multiple constraints and objectives, which may also affect the design optimization results. The design optimization methods developed in this dissertation need to be extended to include correlations among input randon variables as well as among multiple constraints and objectives.

Finally, this dissertation develops uncertainty analysis and design optimization methods for multidisciplinary systems. However, in practice, the multidisciplinary models might have multiple levels. As the models are integrated across multiple levels, the complexity and sophistication of the models increases, and assessing the predictive capability of the overall system model becomes a more difficult challenge. The methods developed in this dissertation need to be extended for the multi-level multidisciplinary systems.

In the long term, the methodologies developed in this dissertation can also be extended to solve problems in economics and finance, for example, portfolio optimization, product family optimization, probabilistic budget estimation, etc and problems in systems of systems (SoS), for example, transportation systems, emergency response, network optimization, etc. Most of the existing solution approaches to these problems deal with aleatory uncertainty only (McDonald, 2008; McInvale, 2009; Touran, 2010). There exist a few methods that deal with both aleatory and epistemic uncertainty in portfolio management (Garlappi et al, 2007, Berleant et al, 2008). These methods primarily focus on epistemic uncertainty arising from model error. However, if these problems can be solved taking into account both aleatory and epistemic uncertainty arising from all sources, the resulting solutions will be more robust, which may assist in more realistic decision making under uncertainty.

BIBLIOGRAPHY

[1] Agarwal, H, Renaud, J E, Preston, E L, and Padmanabhan, D, *Uncertainty quantification using evidence theory in multidisciplinary design optimization*, Reliability Engineering and System Safety 85 (2004) 281–294.

[2] Agarwal, H., and Renaud, J., 2004, *A Unilevel Method for Reliability Based Design Optimization*, Proceedings of 45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Material Conference, Paper No. AIAA- 2004-2029.

[3] Agarwal, H, Mozumder, C K, Renaud, J E, and Watson, L T, *An inverse-measure-based unilevel architecture for reliability-based design optimization*, Struct Multidisc Optim (2007) 33: 217–227.

[4] Allen, J K, Seepersad, C, Choi, H, and Mistree, F, *Robust Design for Multiscale and Multidisciplinary Applications*, J. Mech. Des. 128, 832 (2006).

[5] Apeland S, Aven T, Nilsen T. *Quantifying uncertainty under a predictive, epistemic approach to risk analysis*. Reliability Engineering and System Safety 2002; 75: 93-102.

[6] Arvesen, J.N. 1969. *Jackknifing U-Statistics*. Annals of Mathematical Statistics 40:2076-2100.

[7] Ayyub BM., *From dissecting ignorance to solving algebraic problems*, Reliability Engineering and System Safety, 85 (2004) 223-238.

[8] Baudrit, C., Dubois, D., *Practical representations of incomplete probabilistic knowledge*, Computational Statistics & Data Analysis 51 (2006) 86 **–** 108.

[9] Ben-Haim, Y and Elishakoff, I. 1990, *Convex Models of Uncertainty in Applied Mechanics*. Studies in Applied Mechanics, 25.

[10] Ben-Haim, Y, 2004, *Uncertainty, probability and information-gaps*, Reliability Engineering & System Safety**,** 85, 249–266.

[11] Berleant, D, Andrieu, L, Argaud, J.-P., Barjon, F, Cheong, M.-P., Dancre, M, Sheble, G and Teoh, C.-C., *Portfolio management under epistemic uncertainty using stochastic dominance and information-gap theory*, International Journal of Approximate Reasoning **49** (1) (2008), pp. 101–116.

[12] Berleant, D, 1993, *Automatically verified reasoning with both intervals and probability density functions*. Interval Computations, pp 48-70.

[13] Berleant, D, 1996, *Automatically verified arithmetic on probability distributions and intervals*. Applications of Interval Computations, B. Kearfott and V. Kreinovich (eds.), Kluwer Academic Publishers, pp 227-244.

[14] Berleant, D. and C. Goodman-Strauss. 1998, *Bounding the results of arithmetic operations on random variables of unknown dependency using intervals*. Reliable Computing 4, pp 147-165.

[15] Berleant D, Zhang J., Representation and problem solving with distribution envelope determination (DEnv), Reliability Engineering and System Safety, 85 (2004) 153-168.

[16] Berleant, D, and Zhang, J, *Using Pearson Correlation to Improve Envelopes around the Distributions of Functions*, Reliable Computing, 10:139-161, 2004.

[17] Bertrand, P., and Groupil, F., 2000, *Descriptive statistics for symbolic data*, in Analysis of Symbolic Data, H. H. Bock and E. Diday (editors), Springer, Berlin.

[18] Bichon, B J, McFarland, J M and Mahadevan, S, *Using Bayesian Inference and Efficient Global Reliability Analysis to Explore Distribution Uncertainty*, 49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 7 - 10 April 2008, Schaumburg, IL.

[19] Bonett, D G., *Approximate confidence interval for standard deviation of nonnormal distributions*, Computational Statistics & Data Analysis 50 (2006) 775 – 782.

[20] Box, G. E. P. and Cox, D. R., 1964, *An analysis of transformations*, Journal of the Royal Statistical Society, Series B, 26, 211-52.

[21] Bras, B. A., and Mistree, F., 1993, *Robust Design using Compromise Decision Support Problems*, Engineering Optimization, Vol. 21, pp. 213-239.

[22] Bras, B. A. and Mistree, F., 1995, *A Compromise Decision Support Problem for Robust and Axiomatic Design*, ASME Journal of Mechanical Design, Vol. 117, No. 1, pp. 10-19.

[23] Cagan, J., Williams, B. C., 1993, *First-Order Necessary Conditions for Robust Optimality*, in *ASME Advances in Design Automation*, Albuquerque, NM, ASME DE-Vol. 65-1.

[24] Chen, W., Allen, J. K., Mistree, F. and Tsui, K.-L., 1996, *A Procedure for Robust Design: Minimizing Variations Caused by Noise Factors and Control Factors*, ASME Journal of Mechanical Design, Vol. 118, pp.478-485.

[25] Chen, W, and Lewis, K, *A Robust Design Approach for Achieving Flexibility in Multidisciplinary Design*, AIAA Journal, 37(8), pp. 982–989, 1999.

[26] Chen W, Wiecek MM, Zhang J., *Quality utility—a compromise programming approach to robust design*, Journal of Mechanical Design (ASME) 1999; 121:179–187.

[27] Chen, W., Sahai, A., Messac, A., Sundararaj, G.J., *Exploration of the effectiveness of physical programming in robust design*, J. Mech. Des. 122, 155–163, 2000.

[28] Cheng, H and Sandu, A, *Efficient uncertainty quantification with the polynomial chaos method for stiff systems*, Mathematics and Computers in Simulation 79 (2009) 3278–3295.

[29] Chen, X., Hasselman, T. K., and Neill, D. J., 1997, "Reliability Based Structural Design Optimization for Practical Applications," *Proceedings of the 38th IAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Material Conference*, Kissimmee, FL, Paper No. AIAA-97-1403.

[30] Chiralaksanakul, A., and Mahadevan, S., *First-Order Approximation Methods in Reliability-Based Design Optimization*, J. Mech. Des.,Volume 127, Issue 5, 2005.

[31] Chiralaksanakul, A, and Mahadevan, S, *Decoupled approach to multidisciplinary design optimization under uncertainty*, Optim Eng (2007) 8: 21–42.

[32] Cojbasic, V., Tomovic, A., *Nonparametric confidence intervals for population variance of one sample and the difference of variances of two samples*, Computational Statistics & Data Analysis 51 (2007) 5562 – 5578.

[33] Cramer, J S., *Mean and variance of R^2 in small and moderate samples*, Journal of Econometrics, 35, 253-266, 1987.

[34] Cramer, E. J., Dennis, Jr., J. E., Frank, P.D., Lewis, R.M. and Shubin, G.R., *Problem Formulation for Multi-disciplinary Design Optimization*, SIAM Journal on Optimization, 4(4), pp. 754-776, November 1994.

[35] Cruse, T. A., 1997, *Mechanical Reliability Design and Models*, In Reliability-Based Mechanical Design (Ed: T.A. Cruse), Dekker, Inc, New York, pp. 49-90.

[36] Dai, Z, and Mourelatos, Z P, *Incorporating Epistemic Uncertainty in Robust Design*, Proceedings of DETC, 2003 ASME Design Engineering Technical Conferences September 2-6, 2003, Chicago, Illinois, USA.

[37] DeBrota, D. J., Swain, J. J., Roberts, S. D., Venkataraman, S. 1988, Input modeling with the Johnson System of distributions, *Proceedings of the 1988 Winter Simulation Conference.*

[38] DeBrota, D. J., Roberts, S. D., Dittus, R. S., and Wilson, J. R., 1989, "Visual interactive fitting of bounded Johnson distributions", Simulation, 52 (5), pp. 199-205.

[39] De Cooman G, and Troffaes MCM, Coherent lower previsions in systems modeling: products and aggregation rules, Reliability Engineering and System Safety, 85 (2004) 113-134.

[40] DeGroot, M H, 1984, Probability and Statistics, 2nd edition, Addison-Wesley Publishing Company, ISBN 0-201-11366-X.

[41] Der Kiureghian, A, and Liu, P, *Structural Reliability under Incomplete Probability Information*, Journal of Engineering Mechanics, Vol. 112, No. 1, January 1986, pp. 85-104.

[42] Ditlevsen, O., 1979, "Narrow Reliability Bounds for Structural Systems," J. Struct. Mech., **7**(4), pp. 453–472.

[43] Doltsinis, I, and Kang, Z, *Robust design of structures using optimization methods, Comput*, Methods Appl. Mech. Engrg. 193 (2004) 2221–2237.

[44] Du, X., and Chen, W., *Towards a better understanding of Modeling Feasibility Robustness in Engineering*, ASME J. Meach. Des., 2000, 122(4).

[45] Du, X., and Chen, W., 2000, An integrated methodology for uncertainty propagation and management in simulation-based systems design, AIAA Journal, 38(8), pp 1471-1478.

[46] Du, X, and Chen, W, *Efficient Uncertainty Analysis Methods for Multidisciplinary Robust Design*, AIAA JOURNAL, Vol. 40, No. 3, March 2002.

[47] Du, X., and Chen, W., 2004, *Sequential Optimization and Reliability Assessment Method for Efficient Probabilistic Design*, ASME J. Mech. Des., **126**, pp. 225–233.

[48] Du. X, Sudjianto, A., Chen. W., *An Integrated Framework for Optimization Under Uncertainty Using Inverse Reliability Strategy*, ASME, 2004.

[49] Du, X, and Chen, W, *Collaborative Reliability Analysis under the Framework of Multidisciplinary Systems Design*, Optimization and Engineering, 6, 63–84, 2005.

[50] Du, X., Sudjianto, A., and Huang, B., 2005, *Reliability based design with mixture of random and interval variables*, Journal of Mechanical Design, 127, pp 1068 – 1076.

[51] Du, X, and Beiqing Huang, B, *Reliability-based design optimization with equality constraints*, *Int. J. Numer. Meth. Engng* 2007; 72:1314–1331.

[52] Du, X., 2008, *Unified Uncertainty Analysis by the first Order Reliability Method*, Journal of Mechanical Design, Volume 130, Issue 9, 091401 (10).

[53] Du, X, Guo, J and Beeram, H, *Sequential optimization and reliability assessment for multidisciplinary systems design*, Struct Multidisc Optim (2008) 35:117–130.

[54] Dubois, D., and Prade, H., 1988, Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, First edition.

[55] Efron, B. 1979. *Bootstrap Methods: another look at the jackknife*. Annals of Statistics7: 1-26.

[56] Ferson S, and Hajagos J, Arithmetic with uncertain numbers: rigorous and (often) best-possible answers, Reliability Engineering and System Safety, 85(2004) 135-152.

[57] Ferson, S. ., Joslyn, C. A. ., Helton, J. C.., Oberkampf, W. L. ., and Sentz, K. 2004 ., Summary from the epistemic uncertainty workshop: consensus amid diversity, Reliability Engineering and System Safety 85, pp 355–369.

[58] Ferson, S, and Kreinovich, V, *Modeling Correlation and Dependence Among Intervals*, 2006, Available online, http://www.cs.utep.edu/vladik/2006/tr06-04.pdf.

[59] Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf W., and Ginzburg, L. *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*. Sandia National Laboratories Technical report SAND2007-0939, Albuquerque, New Mexico, 2007.

[60] Ferson, S, Tucker, W, T, Paredis, C J. J., Bulgak, Y, and Kreinovich, V, *Accounting for epistemic and aleatory uncertainty in early system design,* NASA SBIR Phase 2 Final Report, July 2009.

[61] Fetz T, Oberguggenberger M., Propagation of uncertainty through multivariate functions in the framework of sets of probability measures, Reliability Engineering and System Safety, 85(2004) 73-87.

[62] Fletcher, R, *Practical Methods of Optimization*, 2nd ed. John Wiley, New York, 1987.

[63] Garlappi L., Uppal R., Wang T, *Portfolio Selection with Parameter and Model Uncertainty:A Multi-Prior Approach*, The Review of Financial Studies (2007) 20:41–81.

[64] Ghanem, R., and Spanos, P., *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 1991.

[65] Guo. J and Du. X, 2007, *Sensitivity Analysis with Mixture of Epistemic and Aleatory Variables*, AIAA Journal, 45(9).

[66] Guo. J and Du. X, 2009, *Reliability sensitivity analysis with random and interval variables*, International Journal for Numerical Methods in Engineering, Available online.

[67] Gu, X, Renaud, J E, Batill, S M, Brach, R M, and Budhiraja, A S, *Worst case propagated uncertainty of multidisciplinary systems in robust design optimization*, Struct Multidisc Optim 20, 190–213, 2000.

[68] Gu, X, Renaud, J E, and Penninger, C L, *Implicit Uncertainty Propagation for Robust Collaborative Optimization*, Journal of Mechanical Design, Vol. 128 / 1001, JULY 2006.

[69] Haas, C.N, *On Modeling Correlated Random Variables in Risk Assessment*, Risk Analysis, Vol. 19, No. 6, 1999.

[70] Hailperin, T., 1986, *Boole's Logic and Probability*, North-Holland, Amsterdam.

[71] Haldar, A., Mahadevan, S., 2000, *Probability, Reliability and Statistical Methods in Engineering Design,* J. Wiley & Sons, New York, 304 Pages, ISBN-10: 0-471-33119-8.

[72] Hall JW, Lawry J., *Generation, combination and extension of random set approximations to coherent lower and upper probabilities*, Reliability Engineering and System Safety, 85 (2004) 89-101.

[73] Hasofer, A. M., and Lind, N. C., *Exact and Invariant Second Moment Code Format,* Journal of the Engineering Mechanics Division, ASCE, Vol. 100, No. EM1, pp. 111-121, 1974.

[74] Helton JC, Burmaster DE. Guest editorial: treatment of aleatory and epistemic uncertainty in performance assessments for complex systems, Reliability Engineering and System Safety, 54 (1996):91–94.

[75] Helton JC, Johnson JD, and Oberkampf WL, *An exploration of alternative approaches to the representation of uncertainty in model predictions*, Reliability Engineering and System Safety, 85 (2004) 39-71.

[76] Helton J C, Johnson, J D, Oberkampf, W L and Storlie, C B,A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory, Comput. Methods Appl. Mech. Eng. **196** (37–40) (2007), pp. 3980–3998.

[77] Helton,JC, Johnson, JD, Oberkampf, WL, and Sallaberry, C.J, 2008, *Representation of Analysis Results Involving Aleatory and Epistemic Uncertainty,* SANDIA REPORT, SAND2008-4379.

[78] Higham, N J, *Computing the nearest correlation matrix—a problem from finance*, IMA Journal of Numerical Analysis, 22 (2002), pp. 329–343.

[79] Hofer E, Kloos M, Hausmann BK, Peschke J, Woltereck M, An approximate epistemic uncertainty analysis approach in the presence of epistemic and aleatory uncertainties, Reliability Engineering and System Safety, 2002; 77:229–238.

[80] Hong, HP, *An efficient point estimate method for probabilistic analysis*, Reliability Engineering and System Safety 59 (1998) 261-267.

[81] Huang, B., and Du, X, *Analytical robustness assessment for robust design*, Struct Multidisc Optim (2007) 34:123–137.

[82] Howson C, Urbach P. *Scientific Reasoning: The Bayesian Approach*, 2nd edn. Chicago, IL: Open Court, 1993.

[83] Hyman, J.M,. 1982, *FORSIG: An Extension of FORTRAN with Significance Arithmetic*, *Report LA-9448-MS*, Los Alamos National Laboratory, Los Alamos, New Mexico. See also the website http://math.lanl.gov/ams/report2000/significancearithmetic.html.

[84] John, J. and Draper, N., 1980, *An alternative family of transformations*, Applied Statistics, 29, 190-7.

[85] Johnson N.L., 1949a, *Systems of frequency curves generated by methods of translation*. *Biometrika*; 36: 149–76.

[86] Johnson, N. L., *Bivariate distributions based on simple translation systems*, Biometrika 36:297–304, 1949b.

[87] Johnson, N J., *Modified t Tests and Confidence Intervals for Assymmetrical Populations*, Journal of the American Statistical Association, Vol. 73, No. 363, Sep., 1978, pp. 536-544.

[88] Kalsi, M, Hacker, K, and Lewis, K, *Comprehensive Robust Design Approach for Decision Trade-Offs in Complex System Design*, ASME J. Mech. Des., 123, pp. 1–10, 2001.

[89] Klir GJ., *Generalized information theory: aims, results, and open problems*, Reliability Engineering and System Safety, 85 (2004) 21-38.

[90] Kozine IO, and Utkin LV, *An approach to combining unreliable pieces of evidence and their propagation in a system response analysis,* Reliability Engineering and System Safety, 85 (2004) 103-112.

[91] Kreinovich, V. 2004. *Probabilities, intervals, what next? Optimization problems related to extension of interval computations to situations with partial information about probabilities.* Journal of Global Optimization, 29 (3), pp 265-280.

[92] Kreinovich, V., Xiang, G., Starks, S. A., Longpre, L., Ceberio, M., Araiza, R., BECK, J., Kandathi, R., Nayak, A., Torres, R., 2006, *Towards combining probabilistic and interval uncertainty in engineering calculations: Algorithms for computing statistics under interval uncertainty, and their computational complexity*, Reliable Computing, 12 (4), pp. 273–280.

[93] Langley, R. S., 2000, *Unified approach to probabilistic and possibilistic analysis of uncertain systems*, Journal of Engineering Mechanics, 126(11), pp 1163-1172.

[94] Lee, K-H, and Park, G-J, *Robust optimization considering tolerances of design variables*, Computers and Structures 79 (2001) 77-86.

[95] Lee. I., Choi. K. K., Du. L., Gorsich. D., *Dimension reduction method for reliability-based robust design optimization*, Computers and Structures, 2008.

[96] Li W, Hyman JM., *Computer arithmetic for probability distribution variables*, Reliability Engineering and System Safety, 85 (2004) 191-209.

[97] Li, M and Azarm, S, *Multiobjective Collaborative Robust Optimization with Interval Uncertainty and Interdisciplinary Uncertainty Propagation*, ASME Journal of Mechanical Design, AUGUST 2008, Vol. 130.

[98] Liang, J., Mourelatos, Z., and Tu, J., 2004, *A Single-Loop Method for Reliability-Based Design Optimization*, Proceedings of the ASME Design Engineering Technical Conferences.

[99] Liu, P, and Der Kiureghian, A, *Multivariate distribution models with prescribed marginals and covariances,* Probabilistic Engineering Mechanics, 1986, Vol. 1, No. 2, pp. 105-112.

[100] Madsen, H., and Hansen, P., 1992, *A Comparison of Some Algorithms for Reliability Based Structural Optimization and Sensitivity Analysis, Proceedings of the Fourth IFIP WG 7.5 Conference*, Munich, Germany, R. Rackwitz and P. Thoft-Christensen, eds., Springer-Verlag, Berlin, pp. 443–451.

[101] Mahadevan, S, and Rebba, R, *Inclusion of Model Errors in Reliability-Based Optimization*, J. Mech. Des. 128, 936 (2006).

[102] Mahadevan, S, and Smith, N, *Efficient first-order reliability analysis of multidisciplinary systems,* Int. J. Reliability and Safety, Vol. 1, Nos. 1/2, 2006.

[103] Marhadi, K.S., *Quantifying reliability estimate due to statistical uncertainty*. Independent Study Report, School of Mathematics, Claremont Graduate University, Claremont, California, May 2007.

[104] Marler, R T, and Arora, J S, *Survey of multi-objective optimization methods for engineering*, Struct Multidisc Optim 26, 369–395 (2004).

[105] Mavrotas, G, *Effective implementation of the e-constraint method in Multi Objective Mathematical Programming problems*, Applied Mathematics and Computation 213 (2009) 455–465.

[106] McDonald, J.B., *Some generalized functions for the size distribution of income*, Econometrica, 52 (1984), pp. 647–663.

[107] McDonald, M, *Optimization Under Uncertainty for Family of Systems and System of Systems Engineering*, PhD Dissertation, Submitted to the Faculty of the Graduate School of Vanderbilt University, 2008.

[108] McDonald, M., Zaman, K., Mahadevan, S., 2009, *Representation and First-Order Approximations for Propagation of Aleatory and Distribution Parameter Uncertainty*, In *Proceedings of the 50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Palm Springs, California, Paper number AIAA-2009-2250.

[109] McInvale, H D, *Optimal Control Policies for Stochastic Networks with Multiple Decision Makers*, PhD Dissertation, Submitted to the Faculty of the Graduate School of Vanderbilt University, 2009.

[110] Messac, A., *Physical Programming Effective Optimization for Computational Design*, AIAA Journal, Vol. 34, No. 1, 1996, pp. 149–158.

[111] Messac, A., Melachrinoudis, E., and Sukam, C. P., *Mathematical and Pragmatic Perspectives of Physical Programming*, AIAA Journal, Vol. 39, No. 5, 2001, pp. 885-893.

[112] Messac, A, and Ismail-Yahaya, A, *Multiobjective robust design using physical programming*, Struct Multidisc Optim 23, 357–371, 2002.

[113] Miller, R.G. 1974. *The jackknife - a review*. Biometrika 61: 1-15.

[114] Miller, R.G. 1964. *A trustworthy jackknife*. Annals of Mathematical Statistics 35: 1594-1605.

[115] Miller, R.G. 1968. *Jackknifing variances*. Annals of Mathematical Statistics 39: 567-582.

[116] Minhajuddin, A.T.M, Harris, I.R, and Schucany, W.R, *Simulating multivariate distributions with specific correlations, Journal of Statistical Computation and Simulation*, 1563-5163, Volume 74, Issue 8, 2004, Pages 599 – 607.

[117] Mishra S, *The nearest correlation matrix problem: solution by differential evolution method of global optimization*, Research Papers in Economics, April 14, 2007. <http://ideas.repec.org/p/pra/mprapa/2760.html>.

[118] Mourelatos, Z P., Zhou, 2005, *Reliability Estimation With Insufficient Data Based on Possibility Theory*, AIAA J., **43**(8), pp. 1696–1705.

[119] Mourelatos, Z.P., and Zhou J, 2006, *A Design Optimization Method Using Evidence Theory*, *Journal of Mechanical Design*, ASME, 128(4), pp 901 – 908.

[120] Nataf, A., *D´etermination des distributions de probalit´es dont les marges sont donn´ees*, Comptes Rendus de l'Academie des Sciences, Vol. 225, pp: 42–43, 1962.

[121] Noh, Y, Choi, K.K, and Du, L, *Reliability-based design optimization of problems with correlated input variables using a Gaussian Copula*, Struct Multidisc Optim (2009) 38:1–16.

[121] Nowak, A., and Collins, K., 2000, *Reliability of Structures*, McGraw-Hill, New York, p. 360.

[122] Oberkampf WL, Helton JC and Sentz K, 2001, *Mathematical representation of uncertainty*, In *Proceedings of the 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference & Exhibit*, Seattle, WA, Paper number AIAA 2001-1645.

[123] Oberkampf WL, Helton JC, Joslyn CA, Wojtkiewicz SF, Ferson S, *Challenge Problems: uncertainty in system response given uncertain parameters*, Reliability Engineering and System Safety, 85 (2004) 11-19.

[124] Panier, E R, and Tits, A L, *On combining feasibility, descent and superlinear convergence in inequality constrained optimization*, Mathematical Programming, 59:261–276, 1993.

[125] Park, G-J, Lee, T-H, Lee, KH, and Hwang, K-H, *Robust Design: An Overview*, AIAA JOURNAL, Vol. 44, No. 1, January 2006.

[126] Parkinson, A., Sorensen, C., and Pourhassan, N., 1993, *A General Approach for Robust Optimal Design*, Transactions of the ASME, Vol. 115, pp.74-80.

[127] Parry, G.W., *The characterization of uncertainty in probabilistic risk assessments of complex systems*, Reliability Engineering and System Safety, 54 (1996), pp. 119–126.

[128] Pate´-Cornell ME. *Uncertainties in risk analysis: six levels of treatment*, Reliability Engineering and System Safety, 1996; 54(2–3):95–111.

[129] Pearson, K, 1895, *Contributions to the Mathematical Theory of Evolution. II. Skew Variations in Homogeneous Material*, Philosophical Transactions of the Royal Society of London, Series A, 186, 343-414.

[130] Penmetsa, R..C., and Grandhi, R.V., 2002, *Efficient estimation of reliability for problems with uncertain intervals*, Computers and Structures, 80 (12), pp 1103 – 1112, 2002.

[131] Rao, S.S, and Cao, L., 2002, Optimum design of mechanical systems involving interval parameters, Journal of Mechanical Design, 124(3), pp 465 – 472.

[132] Ramberg, J. and B. Schmeiser, 1974, *An approximate method for generating asymmetric random variables*, Communications of the ACM, 17(2): 78–82.

[133] Ramu, P., Qu, X., Youn, B.D., Haftka, R.T. and Choi, K.K. (2006) *Inverse reliability measures and reliability-based design optimisation*, Int. J. Reliability and Safety, Vol. 1, Nos. 1/2, pp.187–205.

[134] Rao, S. S., and Annamdas, K. K., 2009, An Evidence-Based Fuzzy Approach for the Safety Analysis of Uncertain Systems, *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Palm Springs, California, Paper number AIAA-2009-2263.

[135] Rebba, R, *Model Validation and Design under Uncertainty*, PhD Dissertation, Submitted to the Faculty of the Graduate School of Vanderbilt University, 2005.

[136] Red-Horse JR, and Benjamin AS, *A probabilistic approach to uncertainty quantification with limited information*, Reliability Engineering and System Safety, 85 (2004) 183-190.

[137] Regan, H M, Ferson, S, Berleant, D, *Equivalence of methods for uncertainty propagation of real-valued random variables*, International Journal of Approximate Reasoning 36 (2004) 1–30.

[138] Robert, C. P. and Casella, G., *Monte Carlo Statistical Methods*. 2nd ed. Springer-Verlag, New York, 2004.

[139] Rosenblatt, M., *Remarks on Multivariate Transformation*, Ann. Math. Stat, Vol. 23(3), pp: 470-472, 1952.

[140] Rosenblueth E (1975), *Point estimates for probability moment*, Proc Natl Acad Sci U S A 72(10):3812–3814.

[141] Rutherford B., *A response-modeling approach to characterization and propagation of uncertainty specified over intervals*, Reliability Engineering and System Safety, 85 (2004) 211-222.

[142] Shafer, Glenn, 1976. *A Mathematical Theory of Evidence*, Princeton University Press, 1976, ISBN 0-608-02508-9.

[143] Shigley, J. E., Mischke, C. R., and Budynas, R. G., 2004, 7th ed. *Mechanical Engineering Design*, McGraw-Hill, New York.

[144] Sim M., *Robust Optimization*, *PhD dissertation submitted to the Sloan School of Management*, Massachusetts Institute of Technology, June 2004.

[145] Smith, N., and Mahadevan, S., *Probabilistic Methods for Aerospace System Conceptual Design*, Journal of Spacecraft and Rockets, 40 (2003), 411-418.

[146] Smith, N, *Probabilistic Design of Multidisciplinary Systems*, PhD dissertation submitted to the Faculty of the Graduate School of Vanderbilt University, May, 2007.

[147] Sobieszczanski-Sobieski, J, 1995, *Multidisciplinary design optimization: an emerging new engineering discipline*, In: Herskovits, J. (ed.) Adv. Struct. Opt.. Amsterdam: Kluwer. pp. 483–496.

[148] Sobieszczanski-Sobieski, J. and Haftka, R. T., 1997, *Multidisciplinary Aerospace Design Optimization: Survey of Recent Developments*, Structural Optimization, 14, pp. 1-23.

[149] Stanfield, P.M, Wilson, J.R, and Mirka, G.A, *Multivariate Input Modeling with Johnson Distributions*, Proceedings of the 1996 Winter Simulation Conference.

[150] Stevenson M D, Hartong A R, Zweber J V, Bhungalia A A, Grandhi R V, *Collaborative Design Environment for Space Launch Vehicle Design and Optimization*, Paper presented at the RTO AVT Symposium on "Reduction of Military Vehicle Acquisition Time and Cost through Advanced Modeling and Virtual Simulation", held in Paris, France, April 2002.

[151] Sundaresan, S., Ishii, K., and Houser, D. R., *A Robust Optimization Procedure with Variations on Design Variables and Constraints*, Engineering Optimization, Vol. 24, No. 2, 1995, pp. 101–117.

[152] Taguchi, G., 1993, *Taguchi on Robust Technology Development: Bringing Quality Engineering Upstream*, ASME Press, New York.

[153] Touran, A, *Probabilistic Approach for Budgeting in Portfolio of Projects*, Journal of Construction Engineering and Management, March 2010.

[154] Tu, J., Choi, K., and Park, Y., 2001, "Design Potential Method for Robust System Parameter Design," AIAA J., **39**, pp. 667–677.

[155] Venkataraman, S. and Wilson, J.R., *Modeling univariate populations with Johnson's translation system-Description of the FITR1 software*. Research Memorandum, School of Industrial Engineering, Purdue University, West Lafayette, Indiana, 1987.

[156] Wang, L., and Kodiyalam, S., 2002, *An Efficient Method for Probabilistic and Robust Design With Non-Normal Distributions*, Proceedings of the 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Denver, CO.

[157] Wei D L, Cui, Z S, and Chen, J, *Robust optimization based on a polynomial expansion of chaos constructed with integration point rules*, J. Mechanical Engineering Science, Vol. 223 Part C, 2009.

[158] Wertz, J.R. and W.J. Larson (eds). 1999. *Space Mission Analysis and Design* (*SMAD*), 3rd Ed. Microcosm, Inc.

[159] Williamson R.C. and T. Downs. (1990). *Probabilistic arithmetic I: numerical methods for calculating convolutions and dependency bounds*. International Journal of Approximate Reasoning 4, pp 89–158.

[160] Yang, R., and Gu, L., 2004, *Experience With Approximate Reliability-Based Optimization Methods*, Struct. Multidiscip. Optim., **26**, pp. 152–159.

[161] Youn, B D, Choi, K K, and Du, L, *Integration of Possibility-Based Optimization and Robust Design for Epistemic Uncertainty*, ASME Journal of Mechanical Design, Vol. 129, AUGUST 2007.

[162] Youn, BD and Wang, P, 2008, *Bayesian reliability-based design optimization using eigenvector dimension reduction (EDR) method*, Struct Multidisc Optim 36:107–123.

[163] Yu, J-C and Ishii, K., 1998, *Design for Robustness Based on Manufacturing Variation Patterns*, *Transactions of the ASME*, Vol.120, pp. 196-202.

[164] Zaman, K, Rangavajhala, S, McDonald, PM., and Mahadevan, S, 2009a, *A probabilistic approach for representation of interval uncertainty*, Reliability Engineering and System Safety. (under review)

[165] Zaman, K, McDonald, M and Mahadevan, S, *Probabilistic Framework for Uncertainty Propagation with both Probabilistic and Interval Variables*, ASME Journal of Mechanical Design 2009b. (Accepted)

[166] Zeleny, M. 1973, *Compromise programming, in: Multiple Criteria Decision Making*, eds: J. L. Cochrane and M. Zeleny, University of South Carolina Press, Columbia, SC, pp. 262-301.

[167] Zhao, Y-G, Ono, T (2000), *New point estimates for probability moments,*, J Eng Mech 126(4):433–436.

[168] Zhao Y-G, Ang AH-S (2003), *System reliability assessment by method of Moments*, J Struct Eng 129(10):1341–1349.

[169] Zhang R., Mahadevan S., 2000, *Model uncertainty and Bayesian updating in reliability-based inspection*, Struct Saf 22:145–160

[170] Zhang, W H, *A compromise programming method using multibounds formulation and dual approach for multicriteria structural optimization*, Int. J. Numer. Meth. Engng 2003; 58:661–678.

[171] Zhang, X, and Huang, H Z, *Sequential optimization and reliability assessment for multidisciplinary design optimization under aleatory and epistemic uncertainties*, Struct Multidisc Optim, 2009.

[172] Zhou, J, Mourelatos, Z P., *A Sequential Algorithm for Possibility-Based Design Optimization*, Journal of Mechanical Design, JANUARY 2008, Vol. 130.

[173] Zou, T, and Mahadevan, S, *Versatile Formulation for Multiobjective Reliability-Based Design Optimization*, J. Mech. Des. 128, 1217 (2006).