A Decomposition-Based Approach to Uncertainty Quantification of Multicomponent Systems

by

Sergio Daniel Marques Amaral
B.S., Pennsylvania State University (2005)
S.M., Pennsylvania State University (2007)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
August 2015 [September 2015]

Signature redacted

Author .....
Department of Aeronautics and Astronautics
July 21, 2015

Signature redacted

Certified by .....
Karen E. Willcox
Professor of Aeronautics and Astronautics
Chair, Thesis Committee

Signature redacted

Certified by .....
Youssef Marzouk
Associate Professor of Aeronautics and Astronautics
Member, Thesis Committee

Signature redacted

Certified by ..
Douglas Allaire
Assistant Professor of Mechanical Engineering, Texas A&M University
Member, Thesis Committee

Signature redacted

Accepted by ...
Paulo Lozano
Associate Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

A Decomposition-Based Approach to Uncertainty Quantification of Multicomponent Systems

by

Sergio Daniel Marques Amaral

Submitted to the Department of Aeronautics and Astronautics
on August 18, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

To support effective decision making, engineers should comprehend and manage various uncertainties throughout the design process. In today's modern systems, quantifying uncertainty can become cumbersome and computationally intractable for one individual or group to manage. This is particularly true for systems comprised of a large number of components. In many cases, these components may be developed by different groups and even run on different computational platforms, making it challenging or even impossible to achieve tight integration of the various models.

This thesis presents an approach for overcoming this challenge by establishing a divide-and-conquer methodology, inspired by the decomposition-based approaches used in multidisciplinary analysis and optimization. Specifically, this research focuses on uncertainty analysis, also known as forward propagation of uncertainties, and sensitivity analysis. We present an approach for decomposing the uncertainty analysis task amongst the various components comprising a feed-forward system and synthesizing the local uncertainty analyses into a system uncertainty analysis. Our proposed decomposition-based multicomponent uncertainty analysis approach is shown to converge in distribution to the traditional all-at-once Monte Carlo uncertainty analysis under certain conditions. Our decomposition-based sensitivity analysis approach, which is founded on our decomposition-based uncertainty analysis algorithm, apportions the system output variance among the system inputs. The proposed decomposition-based uncertainty quantification approach is demonstrated on a multidisciplinary gas turbine system and is compared to the traditional all-at-once Monte Carlo uncertainty quantification approach.

To extend the decomposition-based uncertainty quantification approach to high dimensions, this thesis proposes a novel optimization formulation to estimate statistics from a target distribution using random samples generated from a (different) proposal distribution. The proposed approach employs the well-defined and determinable empirical distribution function associated with the available samples. The resulting optimization problem is shown to be a single linear equality and box-constrained quadratic program and can be solved efficiently using optimization algorithms that

scale well to high dimensions. Under some conditions restricting the class of distribution functions, the solution of the optimization problem yields importance weights that are shown to result in convergence in the L_1-norm of the weighted proposal empirical distribution function to the target distribution function, as the number of samples tends to infinity. Results on a variety of test cases show that the proposed approach performs well in comparison with other well-known approaches.

The proposed approaches presented herein are demonstrated on a realistic application; environmental impacts of aviation technologies and operations. The results demonstrate that the decomposition-based uncertainty quantification approach can effectively quantify the uncertainty of a multicomponent system for which the models are housed in different locations and owned by different groups.

Chair, Thesis Committee: Karen E. Willcox
Title: Professor of Aeronautics and Astronautics

Member, Thesis Committee: Youssef Marzouk
Title: Associate Professor of Aeronautics and Astronautics

Member, Thesis Committee: Douglas Allaire
Title: Assistant Professor of Mechanical Engineering, Texas A&M University

Acknowledgments

The work presented in this thesis would not be possible without the support of many people for whom I owe a great deal of thanks.

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Karen Willcox. I cannot begin to express in words my appreciation for accepting me as her student and providing me with this opportunity. Her guidance, encouragement, and compassion always raised my spirits and encouraged me to progress my research further. The knowledge I have acquired under her advisement, both academically and professionally, will always resonate with me throughout my future endeavors.

I would also like to thank the members of the thesis committee, Prof. Youssef Marzouk and Prof. Douglas Allaire. Their valuable insights and technical support continuously improved the quality of my thesis. I would like to especially thank Prof. Douglas Allaire who would always lend an ear to my broad ideas and offer support. My gratitude is also extended to Prof. Qiqi Wang and Dr. Elena de la Rosa Blanco for their comments and questions on the thesis, which have also served to improve the quality.

I also have to thank the many people who provided excellent support outside of my research. To my colleagues at ACDL and LAE, thank you for sharing this experience with me. I would like to especially recognize my lab members at ACDL including Chad Lieberman, Benjamin Peherstorfer, Tiangang Cui, Daniele Bigoni, Florian Augustin, Alessio Spantini, Alex Gorodetsky, Rémi Lam, Han Chen, Xun Huan, Eric Dow, Luico Di Ciaccio, and Giulia Pantalone. I would also like to thank my lab members at LAE including Steve Yim, Akshay Ashok, Fabio Caiazzo, Chris Gilmore, and Sebastian Eastham. I would like to thank AeroAstro staff members including Jean Sofronas, Joyce Light, Beth Marois, Melanie Burliss, and Meghan Pepin — I appreciate everything you have done to make my time here that much more rewarding and manageable.

Outside of these MIT walls, I was blessed to have the constant support of my close

friends. Madalina Persu, Domoinik Bongartz, Sebastian Eberle, Herve Courthion, Frank Dubois, and Marcel Tücking— thank you for pulling me away from work and helping me keep things in perspective. Ryan Carley, Katherine Li, Ryan Sullivan, Kel Elkins, Ryan Surmik, and Jay Hoffman — thank you for the support, dedication, and all the laughs throughout these years. To Sari Rothrock, I can not thank you enough for helping me fulfill my dream.

Finally, I would like to thank my family for their unyielding love and support. To my brothers, John Amaral and Eric Amaral, I may have left home at a young age but not a day goes by that I do not think about the times we spent together. Thank you for your unwavering support and know that it is reciprocated. To my parents, João Paulo Amaral and Beatriz Amaral, I am where I am today because of your steadfast love, guidance, and determination. Like many immigrants, they came in pursuit of better opportunities for themselves and their children. While it took me most of my youth to come to realize the scale of their sacrifice, I hope I achieved the dream you had for me, I love you.

The work presented in this thesis was funded in part by the United Stated Federal Aviation Administration Office of Environmental and Energy Award Number 09-C-NE-MIT, Amendment Numbers 028, 033, and 038, by the International Design Center at the Singapore University of Technology and Design, and by the DARPA META program through AFRL Contract Number FA8650-10-C-7083 and VU-DSR #21807-S7.

Contents

List of Figures

List of Tables

Nomenclature

d	Number of system variables
d_m	Number of inputs to Component m
f_m	Input-output function associated with Component m
g	Generic input-output function
h	Radon-Nikodym importance weight
k_m	Number of outputs of Component m
n_{eff}	Effective sample size
$p_{\mathbf{X}}$	Proposal density function
$p_{\mathbf{Y}}$	Target density function
t	Integration variable
w	Importance sampling weights
$\hat{w}$	L_2–norm optimal importance sampling weights
x_i	i^{th} component of the vector $\mathbf{x}$
$\mathbf{x}$	Vector of system variables
$\mathbf{x}_{\mathcal{A}_m}$	Vector of variables with indices in the set $\mathcal{A}_m$
$\mathbf{x}_{\mathcal{I}_m}$	Vector of variables with indices in the set $\mathcal{I}_m$
$\mathbf{x}_{\mathcal{J}_m}$	Vector of variables with indices in the set $\mathcal{J}_m$
$\mathbf{x}_{\mathcal{K}_m}$	Vector of variables with indices in the set $\mathcal{K}_m$
$\mathbf{x}_{\mathcal{O}_m}$	Vector of variables with indices in the set $\mathcal{O}_m$
$\mathbf{x}_{\mathcal{S}_m}$	Vector of variables with indices in the set $\mathcal{S}_m$
$\mathbf{x}_{\mathcal{T}_m}$	Vector of variables with indices in the set $\mathcal{T}_m$
$\mathbf{x}_{\mathcal{U}_m}$	Vector of variables with indices in the set $\mathcal{U}_m$
$\mathbf{x}_{\mathcal{V}_m}$	Vector of variables with indices in the set $\mathcal{V}_m$

$\mathbf{y}_m$	Vector of inputs to Component m
δ	Equality constraint Lagrange multiplier
$\boldsymbol{\lambda}$	Inequality constraint Lagrange multiplier
μ	Proposal measure
ν	Target measure
τ_i	Total sensitivity index of the i^{th} input
ω	L_2–norm distance metric between two distribution functions
$\mathbb{P}$	Probability measure
$\mathbf{X}$	Random variable associated with measure μ
$\mathbf{Y}$	Random variable associated with measure ν
$\mathbf{Y}_m$	Random vector of inputs to Component m
D	Variance
$D_{\mathcal{A}}$	Variance associated with the set $\mathcal{A}$
D_2	L_2–norm discrepancy metric
K	Kernel function
L	Bandwidth parameter in kernel density estimation
M	Number of Components which compose the system
$P_{\mathbf{X}}$	Proposal distribution function
$P_{\mathbf{X}}^n$	Proposal empirical distribution function
$P_{\mathbf{X};\mathbf{w}}^n$	Proposal importance weighted empirical distribution function
$P_{\mathbf{Y}}$	Target distribution function
$P_{\mathbf{Y}}^n$	Target empirical distribution function
$S_{\mathcal{A}}$	Sensitivity index of set $\mathcal{A}$
$\mathcal{A}_m$	Set of indices of the system variables that are a subset of inputs to Component m
$\mathcal{B}(a,b)$	Beta distribution with parameters a and b
$\mathcal{D}_m$	Domain of integration for Component m
$\mathcal{F}$	σ-algebra
$\mathcal{I}_m$	Set of indices of the system variables that are inputs to Component m
$\mathcal{J}_m$	Set of indices of all of the inputs and outputs associated with the first $m-1$ components, as well as the indices of the system inputs of Component m

$\mathcal{K}_m$	Set of indices in $\mathcal{J}_m$ with the exception of those indices in $\mathcal{I}_m$
$\mathcal{M}_i$	i^{th} set of components
$\mathcal{N}(\mu, \Sigma)$	Gaussian distribution with mean μ and covariance Σ
$\mathcal{L}$	Lagrangian
$\mathcal{O}_m$	Set of indices of the outputs of Component m
$\mathcal{P}_k$	Set parition $[t^k, t^{k+1})$
$\mathcal{S}_m$	Set of indices of new system inputs to Component m
$\mathcal{T}_m$	Set of indices of the shared inputs of Component m with any of the previous $m-1$ components' inputs or outputs
$\mathcal{U}_m$	Set of indices of the inputs and outputs of the first m components
$\mathcal{U}(a, b)$	Uniform distribution between a and b where $a < b$
$\mathcal{V}_m$	Set of indices of the inputs and outputs of Component m
Ω	Sample space
Π	Generic distribution function
π	Generic density function
$\hat{\pi}$	Estimate of a density function, π
$\mathbb{I}$	Indicator function
$\mathbf{1}_n^T$	Vector of size n containing entires equal to 1
$\mathbf{0}_n^T$	Vector of size n containing entires equal to 0

Chapter 1

Introduction

To support effective decision making, engineers should characterize and manage various uncertainties throughout the design process. Herein, the science of characterizing and managing uncertainties throughout the design process is referred to as uncertainty quantification [78]. Although uncertainty quantification is known to encompass a large scope [109], this research will focus on uncertainty analysis, also known as forward propagation of uncertainties, and sensitivity analysis. In today's modern systems, quantifying uncertainty can become cumbersome and computationally intractable for one individual or group to manage. This is particularly true for systems comprised of a large number of components. In many cases, these components may be developed by different groups and even run on different computational platforms. Recognizing the challenge of quantifying uncertainty in multicomponent systems, we establish a divide-and-conquer approach, inspired by the decomposition-based approaches used in multidisciplinary analysis and optimization [19, 64, 110, 61].

Motivation for uncertainty quantification of multicomponent systems is given in Section 1.1. In Section 1.2, the notation for subsequent developments and the problem statement are presented. The current practices in uncertainty analysis and sensitivity analysis for a single component are discussed in Section 1.3 and Section 1.4 respectively. The current practices in uncertainty quantification of multicomponent systems are discussed in Section 1.5. The objectives of this research are stated in Section 1.6, and an outline of the remainder of the thesis is given in Section 1.7.

1.1 Motivation for decomposition-based uncertainty quantification

Multidisciplinary analysis is an extensive area of research, intended to support today's modern engineered systems which are designed and developed by multiple teams. In addition to the difficulties associated with the design of such systems, the need to enhance performance and efficiency often drives the design to its physical limits. Therefore, the current methodology of modeling a baseline scenario and taking into account safety factors may no longer be sufficient. Instead, a rigorous characterization and management of uncertainty is needed, using quantitative estimates of uncertainty to calculate relevant statistics and failure probabilities. To estimate relevant statistics and failure probabilities requires an uncertainty quantification of the entire system. However, uncertainty quantification of the entire system may be cumbersome due to factors that result in inadequate integration of engineering disciplines, subsystems, and parts, which we refer to collectively here as *components*. Such factors include components managed by different groups, component design tools or groups housed in different locations, component analyses that run on different platforms, components with significant differences in analysis run times, lack of shared expertise amongst groups, and the sheer number of components comprising the system.

Here we present some real world examples illustrating the challenges and outcomes of designing today's modern engineered systems. Boeing's latest aircraft, the Boeing 787 "Dreamliner", is an example of a complex system that has been afflicted with unanticipated costs and delays due to system engineering errors [105, 37]. General Motors electric vehicle, the Chevy Volt, has also seen its assembly cost inflate to double the initial estimate [119]. NASA's International Space Station and the Constellation programs each experienced schedule delays and cost overruns due to factors such as organizational and technical issues [4]. These real world examples illustrate how our current design methodologies may no longer be adequate to analyze systems which have become increasingly complex and, therefore, progressively more difficult to design and manage. By characterizing and managing uncertainty

in complex systems, such as those presented, we can provide relevant statistics such as the probability of a schedule delay or cost overrun to further support decision- and policy-making processes. Government agencies, having acknowledged the importance of uncertainty quantification in design, have taken up the challenge to address these problems by launching new research initiatives in complex systems design under uncertainty [28, 107].

More specifically, in a National Science Foundation workshop on multidisciplinary design and optimization for complex engineered systems; dealing with uncertainty was listed as an overarching theme for future research [107]. The workshop also emphasized the importance of keeping humans in the loop. In the report, it was stressed that a team was necessary, due to the fact that people invariably specialize in a single discipline and must, therefore, work as a team to achieve multidisciplinary objectives. Given today's engineered systems are so complex that they are beyond the comprehension of a single engineer, even design convergence can become an issue depending on how the team is organized. These challenges are only heightened by the fact that globalization has spread the design of the complex engineered system across the world. These observations suggest there is a lacking but necessary aspect of system engineering and design that this research addresses, decomposition-based uncertainty quantification of multicomponent systems. The proposed approach decomposes the multicomponent uncertainty quantification task amongst the various components comprising the multicomponent system and synthesizes these local analyses to quantify the uncertainty of the multicomponent system.

1.2 Definitions

Multicomponent system

Illustrated in Figure 1-1 is a feed-forward multicomponent system whereby uncertainty in the system inputs are propagated throughout the system and ultimately to the system output. Here we formally define the feed-forward multicomponent system along with its respective components and introduce the notation required for

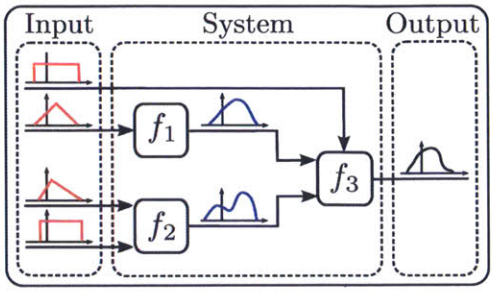

Figure 1-1: The feed-forward multicomponent system presented here illustrates how uncertainty in the *inputs* propagate throughout the *system* and ultimately to the system *outputs*. Different architectures of feed-forward multicomponent systems may also be represented with a similar diagram.

subsequent developments. We take this opportunity to discuss any assumptions and limitations imposed on our feed-forward multicomponent system.

Definition 1. *A system is a collection of M coupled components. Each component has an associated function that maps component input random variables to component output random variables. Let $\mathbf{x} = [x_1, x_2, \ldots, x_d]^\top$ be the vector of system variables, comprised of the inputs and outputs of each component of the system, where shared inputs are not repeated in the vector. For component m, where $m \in \{1, 2, ..., M\}$, let $\mathcal{I}_m \subset \{1, 2, \ldots, d\}$ denote the set of indices of the system variables corresponding to inputs to component m and let $\mathcal{O}_m \subset \{1, 2, \ldots, d\}$ denote the set of indices corresponding to the outputs from component m. Define $d_m = |\mathcal{I}_m|$ and $k_m = |\mathcal{O}_m|$. We denote the function corresponding to component m as $f_m : \mathbb{R}^{d_m} \to \mathbb{R}^{k_m}$, which maps that component's random input vector, $\mathbf{X}_{\mathcal{I}_m} : \Omega \to \mathbb{R}^{d_m}$, where Ω is the product sample space of the input random vector, into that component's random output vector, $\mathbf{X}_{\mathcal{O}_m} = f_m(\mathbf{X}_{\mathcal{I}_m})$. A system whose components can be labeled such that the inputs to the i^{th} component can be outputs from the j^{th} component only if $j < i$ is a feed-forward system.*

Additionally, for each component function, f_m, there exists sets $\{I_1, I_2, \ldots, I_j\}$ that partition the component input space $\mathbb{R}^{d_m}$, such that $f_m : I_i \to \mathbb{R}^{k_m}$ is strictly one-to-one and continuously differentiable for each $i \in \{1, 2, \ldots, j\}$. In later developments we may only be interested in a subset of the components input variables. Let $\mathcal{A}_m \subseteq \mathcal{I}_m$

with the complementary set $\mathcal{A}_m^c = \mathcal{I}_m \backslash \mathcal{A}_m$ and cardinality $0 \leq |\mathcal{A}_m| \leq d_m$, and let $\mathbf{X}_{\mathcal{A}_m} : \Omega \to \mathbb{R}^{|\mathcal{A}_m|}$ be a subset of $\mathbf{X}_{\mathcal{I}_m}$ with $\mathbf{X}_{\mathcal{A}_m^c}$ defining the complementary subset.

Probability framework

Let $\mathbf{t} \in \mathbb{R}^d$ be a generic point and designate entries of $\mathbf{t}$ by subscript notation as follows $\mathbf{t} = [t_1, t_2, \ldots, t_d]^\top$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω is a sample space, $\mathcal{F}$ is a σ-field, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$. Then the random variable $\mathbf{Y} : \Omega \to \mathbb{R}^d$ is associated with the continuous measure ν on $\mathbb{R}^d$, such that $\nu(A) = \mathbb{P}(\mathbf{Y}^{-1}(A))$ for $A \in \mathbb{R}^d$. Define $P_{\mathbf{Y}}(\mathbf{t})$ and $p_{\mathbf{Y}}(\mathbf{t})$ to be the distribution function and probability density function of $\mathbf{Y}$ evaluated at $\mathbf{t}$, respectively. The distribution function and probability density function are defined as

$$P_{\mathbf{Y}}(\mathbf{t}) = \nu((-\infty, \mathbf{t}]), \tag{1.1}$$

and

$$p_{\mathbf{Y}}(\mathbf{t}) = \frac{dP_{\mathbf{Y}}(\mathbf{t})}{d\mathbf{x}}, \tag{1.2}$$

respectively. Likewise, the random variable $\mathbf{X} : \Omega \to \mathbb{R}^d$ is associated with the measure μ on $\mathbb{R}^d$, such that $\mu(A) = \mathbb{P}(\mathbf{X}^{-1}(A))$ for $A \in \mathbb{R}^d$. Similarly, define $P_{\mathbf{X}}(\mathbf{t})$ and $p_{\mathbf{X}}(\mathbf{t})$ to be the distribution function and probability density function of $\mathbf{X}$ evaluated at $\mathbf{t}$, respectively. In addition, we confine the measure ν to be *absolutely continuous* with respect to measure μ.

Definition 2. *The measure ν is said to be absolutely continuous with respect to measure μ if $\mu(A) = 0$ implies $\nu(A) = 0$ for all finite sets $A \in \mathbb{R}^d$ [14].*

In Chapter 3 we constrain the measures to have *finite support.*

Definition 3. *A measure is said to have finite support if its support, $supp(\mu) = \{A \in \mathbb{R}^d \,|\, \mu(A) \neq 0\}$, is a compact set [14].*

This research does not account for discrete distributions (e.g., probability mass functions) since the absolute continuity condition would require that all proposal random

samples be positioned exactly according to the target random samples. To satisfy this condition would require that we know the target distribution function for all system variables prior to performing the uncertainty quantification. However, having knowledge of the target distribution function prior to performing the uncertainty quantification nullifies our decomposition-based approach. For this reason we only consider continuous distribution functions throughout this research.

For a given $\varnothing \neq \mathcal{A} \subseteq \{1, \ldots, d\}$, the marginal density function of $\mathbf{X}_{\mathcal{A}}$ is given by $p_{\mathbf{X}}(\mathbf{t}_{\mathcal{A}}) = \int_{\mathbb{R}^{d-|\mathcal{A}|}} p_{\mathbf{X}_{\mathcal{A}}}(\mathbf{t}) \mathrm{d}\mathbf{t}_{\mathcal{A}^c}$. Let $(\Omega', \mathcal{F}')$ be a measurable space such that the component mapping $f : \Omega \to \Omega'$ is measurable $\mathcal{F}/\mathcal{F}'$. Then the measure μ on $\mathcal{F}$, defines an output measure μf^{-1} on $\mathcal{F}'$ by

$$\mu f^{-1}(A') = \mu(f^{-1}A'), \; A' \in \mathcal{F}', \tag{1.3}$$

where f^{-1} is the push-back of the component mapping. This implies μf^{-1} assigns a value $\mu(f^{-1}A')$ to the set A'. Further, the real-valued component function, f, is a measurable square-integrable function with respect to the induced measure μ supported on $\mathbb{R}^d$.

Upon completing the uncertainty analysis, we may evaluate statistics of interest such as the moments of a quantity of interest or the probability of an event. The mean of a quantity of interest, g, is given by

$$\mathbb{E}_{\mathbf{X}}[g] = \int_{\mathbb{R}^d} g(\mathbf{t}) p_{\mathbf{X}}(\mathbf{t}) \mathrm{d}\mathbf{t}. \tag{1.4}$$

The variance of the quantity of interest, g, is given by

$$var_{\mathbf{X}}(g) = \int_{\mathbb{R}^d} \big(g(\mathbf{t}) - \mathbb{E}_{\mathbf{X}}[g]\big)^2 p_{\mathbf{X}}(\mathbf{t}) \mathrm{d}\mathbf{t} = \mathbb{E}_{\mathbf{X}}[g^2] - \mathbb{E}_{\mathbf{X}}[g]^2. \tag{1.5}$$

Lastly, the probability of an event A (e.g., $A = \{\mathbf{t} \in \mathbb{R}^d \,|\, g(\mathbf{t}) < \hat{g}\}$) is given by

$$\mathbb{E}_{\mathbf{X}}[A] = \int_{\mathbb{R}^d} \mathbb{I}(\mathbf{t} \in A) p_{\mathbf{X}}(\mathbf{t}) \mathrm{d}\mathbf{t}, \tag{1.6}$$

where $\mathbb{I}(\mathbf{t} \in A)$ is the indicator function,

$$\mathbb{I}(\mathbf{t} \in A) = \begin{cases} 1, & \text{if } \mathbf{t} \in A \\ 0, & \text{otherwise.} \end{cases} \tag{1.7}$$

1.3 Uncertainty analysis

Computational methods for uncertainty analysis of a single component can be classified into two groups: intrusive and nonintrusive approaches. Intrusive approaches, also known as embedded projection approaches, introduce a solution expansion into the formulation of the stochastic problem and projects the resulting stochastic equation onto the expansion basis to yield a set of equations that the expansion coefficients must satisfy. The assembly and solution of this stochastic problem requires access and modification to the existing computational model, which may not always be available [44, 115, 80, 68, 126, 48]. Nonintrusive approaches, also known as sampling-based methods, do not require modification of existing computational components and instead treat the components as a "black-box". This research focuses on nonintrusive approaches due to their broader applicability; that is, they can be applied to a wide range of models without requiring knowledge of or access to the underlying implementation details. Within the category of nonintrusive approaches exists a collection of sampling-based methods for the forward propagation of uncertainties. We list here the common sampling-based methods for a single component.

Monte Carlo simulation: Given the function $g : \mathbb{R}^d \to \mathbb{R}$, that takes random inputs $[\xi_1, \xi_2, \ldots, \xi_d]^\top$, we can estimate the mean of $g(\xi_1, \xi_2, \ldots, \xi_d)$ using Monte Carlo simulation as

$$\bar{g} = \frac{1}{n} \sum_{i=1}^{n} g(\xi_1^i, \xi_2^i, \ldots, \xi_d^i), \tag{1.8}$$

where $\{\xi_1^i, \xi_2^i, \ldots, \xi_d^i\}$ is the i^{th} sample realization of the random input to the function. By the strong law of large numbers, $\bar{g} \xrightarrow{a.s.} \mathbb{E}_{\boldsymbol{\xi}}[g(\xi_1, \xi_2, \ldots, \xi_d)]$ as $n \to \infty$ [14]. We may use Monte Carlo simulation to estimate other integral quantities, such as the variance, as well, with almost sure convergence guaran-

teed by the strong law of large numbers.

Full factorial numerical integration: With this approach, the statistical moments of the quantity of interest are calculated through a direct numerical integration using an appropriate quadrature rule [3, 39]. In numerical analysis, a quadrature rule is an approximation of the definite integral of a function, usually expressed as a weighted sum of function values at specified points in the domain of integration.

Projection Methods: With this approach, we express the function via orthogonal polynomials of the input random parameters [44, 68, 126, 48, 125]. The deterministic coefficients associated with each term of the expansion are evaluated through a (multivariate) integration scheme; via Monte Carlo simulation or quadrature rule.

Interpolatory Collocation Methods: With this approach, we express the function as a numerical surrogate model by interpolating between a set of solutions to the computational model [68, 126, 48, 43, 11, 127, 10, 87].

Other nonintrusive approaches not presented in detail here are expansion-based methods [118, 44], most probable point-based methods [52, 41], and nonprobabilistic-based methods [5, 82, 83]. Of the sampling-based methods available, we focus on Monte Carlo simulation. Monte Carlo simulation offers an approach which can more easily cope with dependent component inputs and high dimensional number of random inputs. Additionally, by an application of Skorokhod's representation theorem (see, e.g., Ref. [49]), the estimated mean and variance of any quantities of interest, if they exist, are guaranteed to converge to the true mean and variance.

To evaluate the performance of our decomposition-based multicomponent uncertainty analysis we will evaluate the full system uncertainty analysis. For the uncertainty analysis of a multicomponent feed-forward system, Monte Carlo simulation propagates uncertainty through the system's components by propagating realizations of the random inputs to the system in a serial manner. That is, realizations are

propagated through the system on a component-by-component basis, requiring components with outputs that are inputs to downstream components to be run prior to running downstream components. This can be problematic if components are housed in different locations or owned by different groups, due to communication challenges and the possible transfer of large datasets. Furthermore, any changes to upstream components (e.g., modeling changes, changes in input uncertainty distributions, etc.) will require recomputing all downstream uncertainty analyses.

1.4 Sensitivity analysis

Sensitivity analysis investigates the relationship between inputs and output. Moreover, it allows us to identify how the variability in an output quantity of interest is related to an input in the model and which input sources dominate the response of the system. Sensitivity analysis can be categorized as local or global. A local sensitivity analysis addresses sensitivity relative to point estimates of an input value and is quantified using derivatives of the computational model evaluated at the input value. A global sensitivity analysis quantifies sensitivity with regards to the input distribution rather than a point value. In this research we focus on global sensitivity analysis since we are interested in how the system behaves with respect to the system input distributions. We list here the common approaches to evaluate global sensitivity analysis of a single component.

Screening methods: This class of methods consist of evaluating the local sensitivity analysis whereby each input is varied "one-at-a-time". The sensitivity measures proposed in the original work of Morris are based on what is called an elementary effect [84]. The most appealing property of the screening methods is their low computational costs (i.e. a low required number of computational model evaluations). A drawback of this feature is that the sensitivity measure is only qualitative. It is qualitative in the sense that the input factors are ranked in order of importance, but they are not quantified on how much a given factor is

more important than others. Such methods include Morris's One-factor-At-a-Time.

Sampling-based methods: This class of methods utilize Monte Carlo simulation to investigate the input-output-relationship. These methods include graphical methods [13], regression analysis [55, 35], and correlation coefficients [55]. Unlike screening methods, which vary one-factor-at-a-time, these methods vary all inputs over their entire range.

Moment-independent importance measures: These methods evaluate the influence of the input uncertainty on the entire output distribution without reference to any specific moment of the model output. Moment-independent importance measures evaluate the influence the input has on the output using a distance metric on the output probability density function or output cumulative distribution function [25, 16, 17].

Variance-based methods: These methods quantify the amount of variance that each input factor contributes to the unconditional variance of the output. Variance-based methods offer an approach which captures the influence of the full range of variation of each input factor and the interaction effects among input factors. These methods include Sobol' indices [113], high dimensional model representation [114], Jansen winding stairs [20], and Fourier amplitude sensitivity test [102].

Of the global sensitivity analysis methods available, we will focus on variance-based methods. We perform the global sensitivity analysis using variance-based methods because these methods are well-studied, are easily interpreted, and are commonly used in practice. In this research we assume the multicomponent system inputs are independent. The motivation for the system-level global sensitivity analysis is research prioritization; which factor is the most deserving of further analysis or measurement?

1.5 Current practices for uncertainty quantification in systems

Uncertainty Analysis

The challenges of system uncertainty analysis, illustrated on the left in Figure 1-1, often lie in integrating the components and in the computational expense of simulating the full system. Past work has tackled these challenges through the use of surrogate modeling and/or a simplified representation of system uncertainty. Using surrogates in place of the higher fidelity components in the system provides computational gains and also simplifies the task of integrating components [77]. Using a simplified uncertainty representation (e.g., using mean and variance in place of full distributional information) avoids the need to propagate uncertainty from one component to another. Such simplifications are commonly used in uncertainty-based multidisciplinary design optimization methods as a way to avoid a system-level uncertainty analysis (see e.g., [129] for a review of these methods and their engineering applications). Such methods include implicit uncertainty propagation [50], reliability-based design optimization [24], moment matching [81], advanced mean value method [62], collaborative reliability analysis using most probable point estimation [38], and a multidisciplinary first-order reliability method [75].

Recent methods have exploited the structure of the multicomponent system to manage the complexity of the system uncertainty analysis. A likelihood-based approach has been proposed to decouple feedback loops, thus reducing the problem to a feed-forward system [104]. Dimension reduction and measure transformation to reduce the dimensionality and propagate the coupling variables between coupled components have been performed in a coupled feedback problem with polynomial chaos expansions [7, 8, 9]. Multiple models coupled together through a handful of scalars, which are represented using truncated Karhunen-Loève expansions, have been studied for multiphysics systems [27]. A hybrid method that combines Monte Carlo sampling and spectral methods for solving stochastic coupled problems has also been

proposed [6, 22]. The hybrid approach partitions the coupled problem into subsidiary subproblems which use Monte Carlo sampling methods if the subproblem depends on a very large number of uncertain parameters and spectral methods if the subproblem depends on only a small or moderate number of uncertain parameters. Another method solved an encapsulation problem, without any probability information; upon acquiring probabilistic information, solution statistics of the epistemic variables were evaluated at the post-processing steps [59, 23]. However, all the approaches presented in this review still require evaluating the system in its entirety. Since the approaches presented here do not allow for decomposing the system uncertainty quantification, these approaches do not take advantage of the decomposition-based benefits previously stated in Section 1.1.

Sensitivity Analysis

As was the case in system uncertainty analysis, previous works have simplified the system through the use of surrogate modeling and/or a simplified representation of system uncertainty to perform the system sensitivity analysis. In fact, methods for the forward propagation of uncertainty in systems can be implemented to evaluate the quantities of interest required by the system sensitivity analysis. However, performing a decomposition-based sensitivity analysis of a multicomponent system raises several challenging issues which we address in Chapter 3.

Past works have tackled decomposition-based sensitivity analysis in the application of feed-forward systems. A top-down (i.e., all system variables are independent) sensitivity analysis strategy was developed to determine critical components in the system and used a simplified formulation to evaluate the main sensitivity indices [130]. However, this approach can only be used for designing multicomponent systems with independent components (i.e., no shared variables as inputs to multiple components). To overcome the aforementioned limitations, an extended feed-forward sensitivity analysis method, was developed [74].

In the extended feed-forward sensitivity analysis method, two cases were investigated, dependent on whether there is a linear or nonlinear relation between upstream

component outputs and downstream component dependent coupling variables. In the case of dependent input variables, the Sobol' method has to be performed on all the independent variables and on an artificial subset variable which includes all the dependent coupling variables. Then, the covariance of the dependent coupling variables are required to compute the global sensitivity main effect indices. Lastly, in case of nonlinear dependency, a correction coefficient is used to compute the global sensitivity main effect indices. The extended feed-forward sensitivity analysis method is limited to the estimation of the main effect of the entire system and further efforts are necessary to extend this methodology for interaction effect terms. Our approach, presented in Chapter 3, avoids working with the correlation between system variables and instead evaluates the necessary statistics of interest required for the system sensitivity analysis in a decomposition-based manner.

1.6 Thesis objectives

Based on the motivation and past literature works, there is a need for an uncertainty quantification methodology to manage uncertainty in the complex settings of today's modern engineered systems. This research proposes a decomposition-based vision of the multicomponent uncertainty quantification task, performing uncertainty quantification on the respective components individually, and assembling the component-level uncertainty quantifications to quantify the system uncertainty. We propose a rigorous methodology with guarantees of convergence in distribution. Our decomposition-based approach is inspired by decomposition-based multidisciplinary optimization methods [19, 64, 110, 61]. This research specifically considers the problem of quantifying uncertainty through a feed-forward multicomponent system. To summarize, the high level objectives of this thesis are:

- to develop a decomposition-based uncertainty analysis methodology for feed-forward multicomponent systems with rigorous guarantees of convergence in distribution,

- to develop a decomposition-based global sensitivity analysis methodology for feed-forward multicomponent systems, and
- to demonstrate the decomposition-based uncertainty quantification methodologies on a real world application problem.

1.7 Thesis outline

The remainder of the thesis is organized as follows. In Chapter 2, we develop the decomposition-based uncertainty analysis algorithm and discuss its technical elements. We show that the decomposition-based uncertainty analysis algorithm is provably convergent in distribution and provide an illustrative example. In Chapter 3, we extend the ideas of Chapter 2 to develop a decomposition-based sensitivity analysis algorithm. We demonstrate the decomposition-based sensitivity analysis algorithm on the illustrate example presented in Chapter 2. In Chapter 4, we present an approach to change of measure which overcomes the high dimensional challenges we encounter in Chapter 2. We show that our new change of measure process, under moderate assumptions, is provably convergent in distribution. In Chapter 5, we apply the algorithms developed in this research on a real world application problem; environmental impact of aviation. Additionally, we examine how an individual component-level global sensitivity analysis can be integrated into the decomposition-based multicomponent uncertainty quantification process. Finally, we summarize the thesis contributions in Chapter 6.

Chapter 2

Decomposition-Based Uncertainty Analysis

Our approach tackles the complexity of uncertainty analysis in a feed-forward multicomponent system through decomposition. As illustrated on the right in Figure 2-1, we decompose the system uncertainty analysis into individual component-level uncertainty analyses that are then assembled in a provably convergent in distribution manner to the desired system uncertainty analysis. Many benefits that were not present in the previous works are gained through decomposing the system uncertainty analysis. Such benefits include managing the system through divide and conquer, exploiting team disciplinary expertise, avoiding the challenges of tight analysis integration among components, and being consistent with many organizational structures.

In Section 2.1, we provide, in conjunction with a simple example, an overview of the decomposition-based uncertainty analysis algorithm. The technical elements of the algorithm are discussed in greater detail in Sections 2.2–2.5. In Section 2.6, we discuss the convergence analysis and present an a posteriori indicator to assess the effects of the assumptions underlying the decomposition. The decomposition-based uncertainty analysis algorithm is demonstrated on an aerospace system application problem in Section 2.7. Lastly, the challenges facing the decomposition-based uncertainty analysis approach are discussed in Section 2.8.

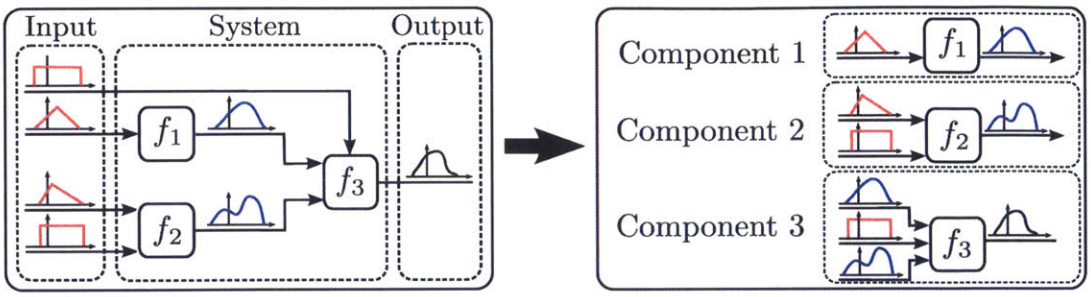

Figure 2-1: The proposed method of multicomponent uncertainty analysis decomposes the problem into manageable components, similar to decomposition-based approaches used in multidisciplinary analysis and optimization, and synthesizes the system uncertainty analysis without needing to evaluate the system in its entirety.

2.1 Introduction to decomposition-based uncertainty analysis

As shown in Figure 2-1, we wish to perform system uncertainty analysis by propagating uncertainty in system inputs to uncertainty in system outputs. We aim to do this by performing uncertainty analysis on system components in a local "offline phase" (ahead of time, decoupled from other component analyses) followed by a synthesis of the component-level analyses in an "online phase". This online synthesis step should ensure that system uncertainty analysis results are achieved in a provably convergent manner, while avoiding any evaluations of the system in its entirety. Specifically, our goal is to develop a decomposition-based uncertainty analysis methodology where the quantities of interest estimated from our decomposition-based approach converge in distribution to the true quantities of interest of the integrated feed-forward system.

Our proposed decomposition-based multicomponent uncertainty analysis approach comprises two main procedures: (1) Local uncertainty analysis: perform a local Monte Carlo uncertainty analysis on each component using their respective proposal distributions; and (2) Global compatibility satisfaction: resolve the coupling among the components *without any further evaluations of the components or of the system as a whole*. Figure 2-2 represents the local and global steps of our approach for a generic component.

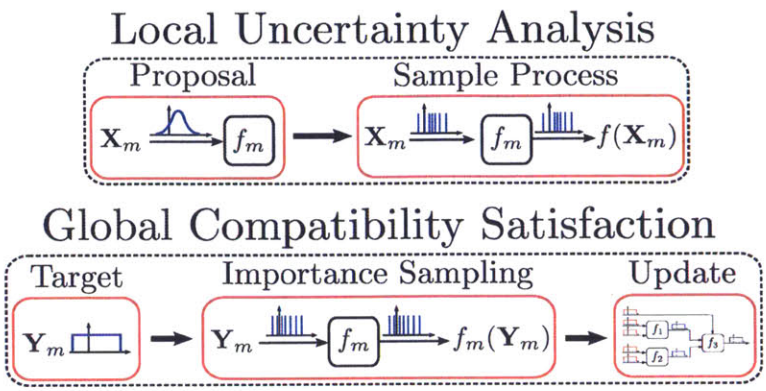

Figure 2-2: The process depicts the local uncertainty analysis and global compatibility satisfaction for component m. First, local uncertainty analysis is performed on the component. Second, global compatibility satisfaction uses importance sampling to update the proposal samples so as to approximate the target distribution. Finally, an update step accounts for dependency among variables.

Each local uncertainty analysis is done in a decoupled offline phase. The challenge created by decomposition is that the distribution functions of the inputs for each component are unknown when conducting the local uncertainty analysis. Therefore, we propose an initial distribution function for each component input, which we refer to as the *proposal distribution function*. Local uncertainty analysis uses the proposal distribution function to generate samples of the uncertain component inputs and propagate them through the component analysis to generate corresponding samples of component outputs.

In the online phase, we learn the true distribution function of the inputs of each component. We refer to these true distribution functions as the *target distribution functions*. For those component inputs that correspond to system inputs, the target distribution functions represent the particular specified scenario under which we wish to perform the system uncertainty analysis. For those component inputs that correspond to coupling variables (i.e., they are outputs from upstream components), the target distribution functions are specified by the uncertainty analysis results of the corresponding upstream component(s).

Global compatibility satisfaction is ensured by starting with the most upstream components of the system and approximating their respective target distribution functions using importance sampling on the corresponding proposal distribution functions.

The densities of updated output samples of these components are then represented using density estimation. We construct joint densities among the inputs and outputs of these upstream components to account for any dependencies that were not captured in the marginal densities of each component's outputs. Once this is complete, downstream components receive their respective target distribution functions and the process of importance sampling and accounting for dependency is repeated through the system.

The theoretical analysis presented in Section 2.6 requires that components can be represented by piecewise functions each of which are one-to-one and continuously differentiable on sets of finite measure. In addition, the density estimation steps of the approach will suffer loss of accuracy if the underlying densities are not sufficiently smooth. This restricts the class of problems for which we can expect good convergence of the decomposition-based approach. For smooth problems, methods such stochastic Galerkin and stochastic collocation can yield faster convergence than Monte Carlo simulation for problems of moderate dimension. In this research, we set up our mathematical framework using Monte Carlo simulation, due to its broader generality; while we do not develop the theory here, our general approach of decomposing into local uncertainty analysis and global compatibility satisfaction could also be combined with stochastic Galerkin and stochastic collocation methods. We also note that if importance weights could be computed without requiring density estimation (the subject of Section 2.4), then our decomposition approach will be applicable to a more general class of multidisciplinary engineering problems, including those that exhibit irregular dependencies, steep gradients, sharp transitions, etc.

To describe our decomposition-based approach more concretely, we consider the specific case of a three-component feed-forward system, shown in Figure 2-3, although our approach extends to multiple components and other forms of component feed-forward coupling. The system inputs are $(\xi_1, \xi_2, \xi_3)^\top$ and ξ_6 is the system output quantity of interest. The coupling variable ξ_4 is an output of component 1 and an input of component 3, and the coupling variable ξ_5 is an output of component 2 and an input of component 3. Thus, the local uncertainty analysis for component 1 involves

evaluating f_1 with sample realizations of the proposal distribution functions of ξ_1 and ξ_2 in order to generate samples of ξ_4. Similarly, the local uncertainty analysis for component 2 involves evaluating f_2 with sample realizations from the proposal distribution functions of ξ_2 (independently of the samples of ξ_2 drawn for component 1) and ξ_3, to generate samples of ξ_5, and the local uncertainty analysis of component 3 involves the evaluation of f_3 with sample realizations from proposal distribution functions of ξ_4 and ξ_5 to generate samples of ξ_6. The key challenges in decomposing the uncertainty analysis for this system are: (1) the local Monte Carlo simulation for f_3 is performed before the target distribution function of the coupling variables ξ_4 and ξ_5 is known, and (2) the dependency between ξ_4 and ξ_5 due to ξ_2 is not accounted for in the local analyses.

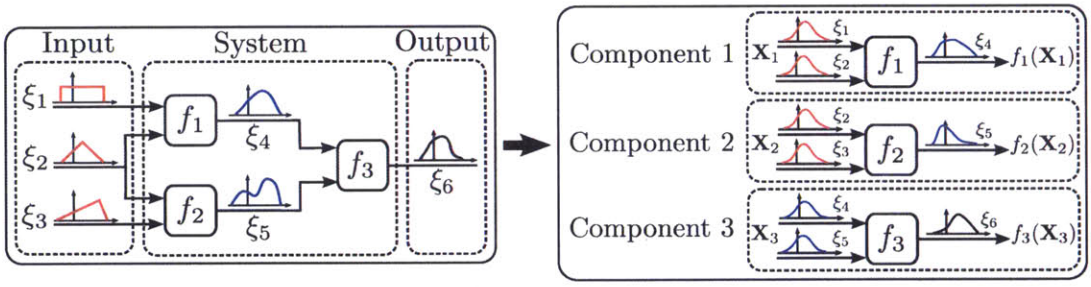

Figure 2-3: Three components of a feed-forward system shown from the system Monte Carlo perspective (left) along with the same components exercised concurrently from the perspective of the decomposition-based multicomponent uncertainty analysis (right).

Starting from the upstream components, here component 1 and component 2, importance sampling assigns weights to the computed samples so as to approximate the input target distribution functions of each component using the samples previously simulated during local analysis from the input proposal distribution functions. The result is an updated output target marginal distribution function of ξ_4 and ξ_5 that requires no further evaluations of the component 1 or component 2 functions. To account for the dependency between ξ_4 and ξ_5, we construct the joint density of $(\xi_4, \xi_5)^\mathsf{T}$ using a conditioning process that is described in detail in Section 2.5 and again requires no further evaluations of the component functions. We then use importance sampling for component 3 to yield an updated output target distribution function for

ξ_6. As a result, the system input target distribution functions are propagated downstream to quantify the distribution function of the output quantity of interest, here ξ_6, without having to perform a full system-level uncertainty analysis. As shown in Section 2.6.1, under some mild assumptions on the component functions, the results of the decomposition-based approach converge in distribution to the true variables of the feed-forward system.

2.2 Local uncertainty analysis

In this section we shall temporarily drop the subscript variable indicating the component of interest to simplify the notation. We define the proposal distribution function with respect to the proposal probability density function for the inputs as

$$P_{\mathbf{X}}(\mathbf{x}) = \int_{\mathcal{D}} p_{\mathbf{X}}(\mathbf{t})\, \mathrm{d}\mathbf{t}, \tag{2.1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_d)^\top$, $p_{\mathbf{X}}$ is the proposal density of the input random vector, $\mathbf{X} = (X_1, X_2, \ldots, X_d)^\top$, and x_j denotes the j^{th} component of the vector $\mathbf{x}$. The domain of integration is $\mathcal{D} = (-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_d)$, $\mathbf{t}$ is a dummy variable of integration, and we assume that the input random variables of each component are continuous. The local uncertainty analysis uses Monte Carlo simulation to generate n samples, $\{\mathbf{x}^i\}_{i=1}^n$, from the proposal distribution function $P_{\mathbf{X}}$, where $\mathbf{x}^i$ denotes the i^{th} sample. We propagate those samples through the component, computing the corresponding component outputs $\{f(\mathbf{x}^i)\}_{i=1}^n$.

The proposal distribution function $P_{\mathbf{X}}$ represents our "best guess" at describing the uncertainty in the inputs of component, made before we receive distributional information from upstream components or system inputs. Choosing an appropriate proposal distribution is important but can be difficult since the target is unknown at the time the proposal is specified. The difficulties are that the target density must be absolutely continuous with respect to the proposal density, and at the same time the proposal should adequately capture the dependency structure and high probability

regions of the target. If the condition of absolute continuity is not satisfied, then the computation of the importance sampling weights (described in more detail in the next section) will fail. Therefore, it is typical to choose a conservative proposal distribution function, ensuring that the support of the proposal is sufficiently wide to encompass the full range of input values expected from upstream components or from system inputs. If the proposal has appropriate support, but does not adequately capture the structure of the target, then our convergence results in Section 2.6 still hold, but the number of samples needed in the offline stage (to achieve a desired accuracy level) may be prohibitive. The quality of the proposal distribution and its impact on the uncertainty assessment results are discussed in Section 2.6.2.

2.3 Sample weighting via importance sampling

In keeping with the notation from Section 1.2, we define the target distribution function of the inputs as

$$P_{\mathbf{Y}}(\mathbf{x}) = \int_{\mathcal{D}} p_{\mathbf{Y}}(\mathbf{t})\, \mathrm{d}\mathbf{t}, \tag{2.2}$$

where $p_{\mathbf{Y}}$ is the target density of $\mathbf{Y}$. We use importance sampling to weight the precomputed proposal samples, $\{\mathbf{x}^i\}_{i=1}^n$, ensuring that the weighted proposal input empirical distribution function converges pointwise to the target input distribution function.

Lemma 1. *Let $P^n_{\mathbf{X};\mathbf{w}}(\mathbf{t})$ be the weighted proposal empirical distribution function for the inputs, computed using n samples. We write*

$$P^n_{\mathbf{X};\mathbf{w}}(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n} w(\mathbf{x}^i)\mathbb{I}(\mathbf{x}^i \leq \mathbf{t}), \tag{2.3}$$

where $\mathbb{I}(\mathbf{x}^i \leq \mathbf{t})$ is the indicator function of the event $\{x^i_j \leq t_j, \forall j \in \{1, 2, \ldots, d\}\}$, $w(\mathbf{x}^i) \propto \frac{p_{\mathbf{Y}}(\mathbf{x}^i)}{p_{\mathbf{X}}(\mathbf{x}^i)}$ and is subject to the condition that the weights sum to n, and ν is absolutely continuous with respect to μ. Then

$$\lim_{n\to\infty} P^n_{\mathbf{X};\mathbf{w}}(\mathbf{t}) = P_{\mathbf{Y}}(\mathbf{t}), \tag{2.4}$$

for all continuity points $\mathbf{t}$ *of* $P_{\mathbf{Y}}(\mathbf{t})$.

Proof. See, e.g., Ref. [108]. □

Lemma 1 implies that the weighted proposal input empirical distribution function converges to the input target distribution function. By applying Skorokhod's representation theorem and assuming that each f is bounded and discontinuous on sets of zero measure, it is straightforward to show (see, e.g., [49]) that the corresponding empirical distribution of the component outputs converges to the distribution function of the target outputs. Thus, all that is required to ensure that we can change from proposal to target distribution functions for each component, without further model evaluations, is the ability to estimate the proposal and target input densities pointwise to provide the sample weights. We discuss how we accomplish this in the next section.

Algorithm 5 describes the importance sampling procedure. This process is shown notionally in Figure 2-4. The left plot on Figure 2-4 shows the contours, in solid, of an example proposal density and samples from that density as dots. Target density contours are shown on the same plot as the dashed curves. Importance sampling provides a weight for each red sample, where the relative weights of the samples are reflected by the size of the blue dots on the right plot of Figure 2-4.

We note here that if a poor proposal distribution function is selected then the importance sampling algorithm may encounter sample impoverishment and therefore result in a poor convergence rate. Therefore, it is desirable to select a conservative proposal distribution function to account for the unknown target distribution function at the local uncertainty analysis step. Another possible source of error arises when the statistic of interest is strongly influenced by a region of low proposal probability. Since the proposal probability density function appears in the denominator of the importance weights, then the estimator of the statistic of interest, when appropriately weighted, may result in an unbounded variance. To prevent this from occurring, the recommended practice is to use, in place of the proposal distribution, a heavy tailed multivariate student's t distribution instead of a light tailed multivariate Gaussian

distribution [97].

Algorithm 1: Importance Sampling for component m.

Data: Target input density $p_{\mathbf{Y}}$, proposal distribution $P_{\mathbf{X}}$ and density $p_{\mathbf{X}}$.

Result: Sample importance weights, $\{w^i\}_{i=1}^n$.

Sampling:

Generate n samples, $\{\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^n\}$, i.i.d. from $P_{\mathbf{X}}$;

Importance Sampling:

for $i \in \{1, \ldots, n\}$ **do**

  Assign to sample $\mathbf{x}^i$ the weight $w^i \propto \frac{p_{\mathbf{Y}}(\mathbf{x}^i)}{p_{\mathbf{X}}(\mathbf{x}^i)}$;

end

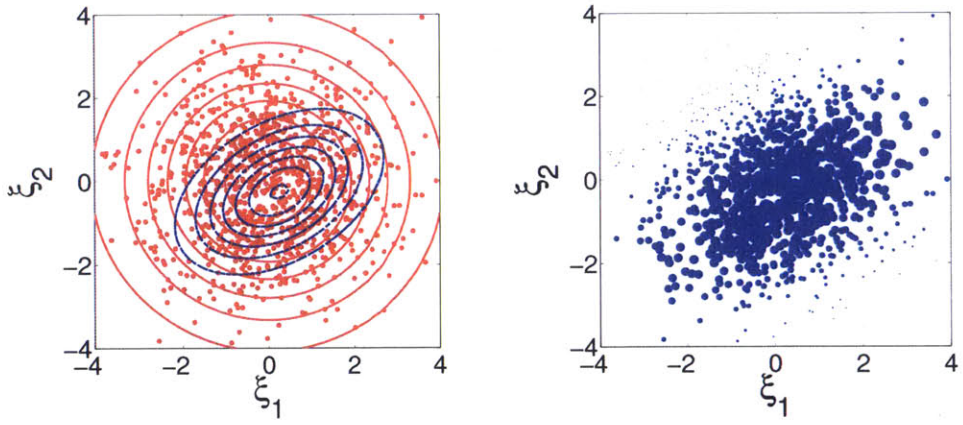

Figure 2-4: The importance sampling process uses the realizations (red dots on left figure) generated from a proposal distribution $P_{\mathbf{X}}(\xi_1, \xi_2)$ (corresponding density shown as red solid contour on left figure) to approximate a target distribution $P_{\mathbf{Y}}(\xi_1, \xi_2)$ (blue dash contour on left figure), by weighting the proposal realizations, (blue dots on right figure).

2.4 Density estimation for estimating importance weights

Algorithm 5 computes importance weights that are the ratio of the target density to the proposal density for a given sample. Since we employ here a sample-based approach to uncertainty propagation, we require a means of estimating proposal and target densities from a set of samples. In particular for some continuous random

variable Ξ with distribution function $\Pi(\xi)$ and a density $\pi(\xi)$, we require that for any density estimate, $\hat{\pi}^n(\xi)$,

$$\lim_{n\to\infty} \hat{\pi}^n(\xi) = \pi(\xi) \tag{2.5}$$

at all points of continuity of the density $\pi(\xi)$. For this we use kernel density estimation,

$$\hat{\pi}^n(\xi) := \frac{1}{nL^d} \sum_{i=1}^{n} K\left(\frac{\xi - \xi^j}{L}\right), \tag{2.6}$$

where $L > 0$ is a bandwidth parameter with the property that $\lim_{n\to\infty} L = 0$, and K is a kernel function satisfying

$$0 \leq K(\mathbf{t}) \leq \infty, \tag{2.7}$$

$$\int_{\mathbb{R}^d} K(\mathbf{t})\, \mathrm{d}\mathbf{t} = 1, \tag{2.8}$$

$$\int_{\mathbb{R}^d} K(\mathbf{t})\xi\, \mathrm{d}\mathbf{t} = 0, \tag{2.9}$$

$$\int_{\mathbb{R}^d} K(\mathbf{t})\, ||\mathbf{t}||^2\, \mathrm{d}\mathbf{t} < \infty, \tag{2.10}$$

where $\mathbf{t} \in \mathbb{R}^d$ and $||\cdot||$ is the Euclidean norm. Then, $\lim_{n\to\infty} \hat{\pi}^n(\xi) = \pi(\xi)$ at every point ξ of continuity of $\pi(\cdot)$ [91, 32]. The Gaussian kernel function and mean integrated squared error bandwidth parameter selection criteria are implemented throughout all examples in this research [51].

If the set of points of continuity of $\pi(\cdot)$ is of measure 1, then in the limit as $n \to \infty$, $\hat{\pi}^n(\xi)$ is a density of the distribution function $\Pi(\xi)$. To ensure this criterion, we require that the inputs to a given system be absolutely continuous random variables. Further, as discussed in Ref. [49], the component functions, f_m, must be such that there are sets $\{I_1, I_2, \ldots, I_j\}$ that partition the component input space $\mathbb{R}^{d_m}$, such that $f_m : I_i \to \mathbb{R}^{k_m}$ is strictly one-to-one and continuously differentiable for each set $i \in \{1, \ldots, j\}$. This constraint ensures that the output probability density function is also an absolutely continuous random variable.

2.5 Accounting for dependency among variables

The weighting of samples via importance sampling in the global compatibility step of our multicomponent uncertainty analysis approach ensures that we achieve the target marginal distribution functions of a given component's inputs. However, the dependencies among variables, such as ξ_4 and ξ_5 in the system shown in Figure 2-3, are not captured. These dependencies must be recovered in order to achieve the correct results.

For example, consider again the system presented in Figure 2-3. The dependency between ξ_4 and ξ_5 caused by the common dependence on ξ_2 can be accounted for by considering the target density of $(\xi_4, \xi_5)^\top$ and the dependency structure as follows [49],

$$\begin{aligned} p_{\mathbf{Y}}(\xi_4, \xi_5) &= \int_{\xi_2} p_{\mathbf{Y}}(\xi_2, \xi_4, \xi_5) d\xi_2 = \int_{\xi_2} p_{\mathbf{Y}}(\xi_2, \xi_4) \cdot p_{\mathbf{Y}}(\xi_5|\xi_2) d\xi_2 \\ &= \int_{\xi_2} \frac{p_{\mathbf{Y}}(\xi_2, \xi_4) \cdot p_{\mathbf{Y}}(\xi_2, \xi_5)}{p_{\mathbf{Y}}(\xi_2)} d\xi_2. \end{aligned} \tag{2.11}$$

Here, $p_{\mathbf{Y}}(\xi_2, \xi_4)$ is the target density associated with input ξ_2 and output ξ_4, which was constructed via importance sampling for component 1. Likewise, $p_{\mathbf{Y}}(\xi_2, \xi_5)$ is the target density associated with input ξ_2 and output ξ_5, which was constructed via importance sampling for component 2. We construct $p_{\mathbf{Y}}(\xi_4, \xi_5)$ using Equation 2.11, which uses the the system's inherent dependence structure to yield the correct target input distribution function to component 3.

To generalize the concept, consider the construction of the target density of the inputs to the first m components of an M-component feed-forward system, where the inputs to component i cannot be outputs of component j unless $j < i$. Recall from Definition 1 that $\mathbf{x} = (x_1, x_2, \ldots, x_d)^\top$ is the vector of system variables, $\mathcal{I}_m \subset \{1, 2, \ldots, d\}$ is the set of indices of the system variables corresponding to inputs to component m, and $\mathcal{O}_m \subset \{1, 2, \ldots, d\}$ is the set of indices corresponding to the outputs from component m. We further define $\mathcal{S}_m \subset \{1, 2, \ldots, d\}$ to be the set of indices of new system inputs to component m, where new system inputs refer

to those inputs that are not output from any component, are not inputs to any previous component, and are assumed independent. Then, let $\mathcal{V}_m = \mathcal{I}_m \cup \mathcal{O}_m$, which is the set of indices of the inputs and outputs of component m, let $\mathcal{U}_m = \cup_{i=1}^{m} \mathcal{V}_m$, which is the set of indices of the inputs and outputs of the first m components, and let $\mathcal{T}_m = (\cup_{i=1}^{m-1} \mathcal{V}_i) \cap \mathcal{I}_m$, which is the set of indices of the shared inputs of component m with any of the previous $m-1$ components' inputs or outputs. We also define $\mathcal{J}_m = \mathcal{U}_{m-1} \cup \mathcal{S}_m$, which is the set of indices of all of the inputs and outputs associated with the first $m-1$ components, as well as the indices of the system inputs of component m, and let $\mathcal{K}_m = \mathcal{J}_m \setminus \mathcal{I}_m$. By constructing the target density of the variables with indices in the set $\mathcal{J}_m$, we correctly capture the dependency among the inputs to component m. We denote by $\mathbf{x}_{\mathcal{J}_m}$ the vector of system variables corresponding to those indices in the set $\mathcal{J}_m$ (and similarly for the other sets defined above).

For the example system given in Figure 2-3, with $m = 3$, we have that $\mathbf{x} = (\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6)^\top$. The indices in each of the sets for $m = 1, 2$, and 3 are given in Table 2.1.

Table 2.1: Index sets for the system presented in Figure 2-3.

$m=1$	$m=2$	$m=3$
$\mathcal{I}_1 = \{1,2\}$	$\mathcal{I}_2 = \{2,3\}$	$\mathcal{I}_3 = \{4,5\}$
$\mathcal{O}_1 = \{4\}$	$\mathcal{O}_2 = \{5\}$	$\mathcal{O}_3 = \{6\}$
$\mathcal{S}_1 = \{1,2\}$	$\mathcal{S}_2 = \{3\}$	$\mathcal{S}_3 = \{\varnothing\}$
$\mathcal{V}_1 = \{1,2,4\}$	$\mathcal{V}_2 = \{2,3,5\}$	$\mathcal{V}_3 = \{4,5,6\}$
$\mathcal{U}_1 = \{1,2,4\}$	$\mathcal{U}_2 = \{1,2,3,4,5\}$	$\mathcal{U}_3 = \{1,2,3,4,5,6\}$
$\mathcal{T}_1 = \{\varnothing\}$	$\mathcal{T}_2 = \{2\}$	$\mathcal{T}_3 = \{4,5\}$
$\mathcal{J}_1 = \{1,2\}$	$\mathcal{J}_2 = \{1,2,3,4\}$	$\mathcal{J}_3 = \{1,2,3,4,5\}$
$\mathcal{K}_1 = \{\varnothing\}$	$\mathcal{K}_2 = \{1,4\}$	$\mathcal{K}_3 = \{1,2,3\}$

For this example, the target densities required for each component are $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_1}) = p_{\mathbf{Y}}(\xi_1, \xi_2)$, $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_2}) = p_{\mathbf{Y}}(\xi_1, \xi_2, \xi_3, \xi_4)$, and $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_3}) = p_{\mathbf{Y}}(\xi_1, \xi_2, \xi_3, \xi_4, \xi_5)$. The target densities, $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_m})$, contain the correct dependence structure for the inputs, $\mathbf{x}_{\mathcal{I}_m}$, to component m. We write the target density of these inputs as $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m})$ for each component m. The evaluation of the input target density, $p_{\mathbf{Y}}(\mathbf{x}^i_{\mathcal{I}_m})$, at proposal

input samples, $\mathbf{x}_{\mathcal{I}_m}^i$, for each component m, ensures that we can properly weight the proposal samples to each component of the system according to Algorithm 5.

Lemma 2. *The target density of the inputs and outputs of the first $m-1$ components and the inputs to component m, where $m \geq 3$, is given by*

$$p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_m}) = \frac{\prod_{k=1}^{m-1} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_k})}{\prod_{l=2}^{m-1} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{T}_l})} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{S}_m}), \tag{2.12}$$

where if $\mathcal{S}_m = \{\varnothing\}$, then $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{S}_m}) = 1$.

Proof.

$$\begin{aligned}
\frac{\prod_{k=1}^{m-1} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_k})}{\prod_{l=2}^{m-1} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{T}_l})} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{S}_m}) &= p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_1}) \frac{p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_2})}{p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{T}_2})} \cdots \frac{p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_{m-1}})}{p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{T}_{m-1}})} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{S}_m}) && (2.13)\\
&= p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_1}) \frac{p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_2})}{p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{T}_2})} \frac{\prod_{k=3}^{m-1} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_k})}{\prod_{l=3}^{m-1} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{T}_l})} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{S}_m}) && (2.14)\\
&= p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{U}_2}) \frac{\prod_{k=3}^{m-1} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_k})}{\prod_{l=3}^{m-1} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{T}_l})} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{S}_m}) && (2.15)\\
&= p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{U}_{m-1}}) p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{S}_m}) && (2.16)\\
&= p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_m}), && (2.17)
\end{aligned}$$

where Equation 2.15 follows from the definition of a conditional probability density (see, e.g., [49]) and Equation 2.17 follows because the system inputs of component m are considered independent from all other system variables. □

Once we have the density $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_m})$, we can evaluate the input target density, $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m})$, required by component m with the correct dependency among the inputs. We note here that for a system with $m = 1$ component there is no dependency structure to resolve. For $m \geq 2$, the input target density for component m is given by

$$p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m}) = \int_{\text{supp}(\mathbf{X}_{\mathcal{K}_m})} p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{K}_m}) p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_m} | \mathbf{x}_{\mathcal{K}_m}) \mathrm{d}\mathbf{x}_{\mathcal{K}_m}, \tag{2.18}$$

where $\text{supp}(\mathbf{X}_{\mathcal{K}_m})$ is the support of the random vector $\mathbf{X}_{\mathcal{K}_m}$. We note here that Equation 2.18 is the expected value of the conditional density, $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_m} | \mathbf{x}_{\mathcal{K}_m})$ with respect

to $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{K}_m})$. The density $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{K}_m})$ is obtained similarly to $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_m})$ using the densities in Equation 2.12 with the variables in $\mathbf{x}_{\mathcal{I}_m}$ marginalized out. The target input density, $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m})$, only needs to be evaluated at the proposal samples as specified in Algorithm 5. A procedure for evaluating this density at the proposal samples, using Monte Carlo simulation to evaluate Equation 2.18 and Lemma 2 to construct $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_m}|\mathbf{x}_{\mathcal{K}_m})$, is given in Algorithm 2. Algorithm 2 avoids the challenge of estimating high-dimensional densities with kernel density methods by assembling the large densities $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{J}_m})$ and $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{K}_m})$ using the smaller dimensional component densities $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_i})$ and $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{T}_i})$. As a result, the largest dimension estimated with kernel density methods is that of the component with the largest cardinality, $|\mathcal{V}_i|$.

Algorithm 2: Accounting for dependency for inputs to component m, where $m \geq 3$.

Data: Target densities $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_i})$ for $i = 1, \ldots, m-1$ and proposal samples $\{\mathbf{x}^i_{\mathcal{I}_m}\}_{i=1}^N$.

Result: Target density $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m})$ evaluated at the proposal samples $\{\mathbf{x}^i_{\mathcal{I}_m}\}_{i=1}^N$.

for $i = 1 : N$ **do**
  Initialize $p_{\mathbf{Y}}(\mathbf{x}^i_{\mathcal{I}_m}) = 0$;
  for $s = 1 : S$ **do**
    $\mathbf{x}^s_{\mathcal{J}_m} = \mathbf{x}^i_{\mathcal{I}_m}$.
    for $j = 1 : (m-1)$ **do**
      Generate sample $\mathbf{x}^s_{\mathcal{V}_j}$ from the density $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_j})$ conditioned on the known set $\mathbf{x}^s_{\mathcal{J}_m}$.
      $\mathbf{x}^s_{\mathcal{J}_m} = \mathbf{x}^s_{\mathcal{J}_m} \cup \mathbf{x}^s_{\mathcal{V}_j}$.
    end
    Generate sample $\mathbf{x}^s_{\mathcal{S}_m}$ from $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{S}_m})$ conditioned on the known set $\mathbf{x}^s_{\mathcal{J}_m}$.
    $\mathbf{x}^s_{\mathcal{J}_m} = \mathbf{x}^s_{\mathcal{J}_m} \cup \mathbf{x}^s_{\mathcal{S}_m}$.
    Evaluate $p_{\mathbf{Y}}(\mathbf{x}^i_{\mathcal{I}_m}) = p_{\mathbf{Y}}(\mathbf{x}^i_{\mathcal{I}_m}) + \frac{1}{S} p_{\mathbf{Y}}(\mathbf{x}^s_{\mathcal{J}_m}) / p_{\mathbf{Y}}(\mathbf{x}^s_{\mathcal{K}_m})$ with Equation 2.12;
  end
end

We note here that for $m = 2$, the input target density may be obtained in a similar fashion to what is done in Algorithm 2. In this case, the innermost for-loop is modified

so that samples are taken from the target density, $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{V}_1})$, and the new system input density, $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{S}_2})$.

2.6 Convergence analysis & a posteriori indicator

This section addresses the convergence properties of the decomposition-based multicomponent uncertainty analysis approach, describes an *a posteriori* indicator to assess proposal quality, and presents a simple example to demonstrate convergence.

2.6.1 Convergence theory

We prove here that the decomposition-based multicomponent uncertainty analysis approach leads to the convergence in distribution of all of the variables associated with a given feed-forward system.

Theorem 1. *Let f_m, for $m = 1, 2, \ldots, M$, be the functions comprising an M-component feed-forward system, where the input spaces of the functions can be partitioned such that on each partition, the functions are one-to-one and continuously differentiable. Let the system inputs be absolutely continuous random variables. Then the target random variables for all system variables estimated via the decomposition-based multicomponent uncertainty analysis procedure converge in distribution to their respective true target random variables as the number of samples tends to infinity.*

Proof. For each component $m = 1, 2, \ldots, M$, local uncertainty analysis using n samples drawn from the proposal distribution functions, $P_{\mathbf{X}}(\mathbf{x}_{\mathcal{I}_m})$, results in proposal empirical distribution functions, $P^n_{\mathbf{X}}(\mathbf{x}_{\mathcal{I}_m})$. Define the set $\mathcal{M}_1$ as the indices of the components with no dependence on upstream components, $\mathcal{M}_1 = \{m \in \{1, 2, \ldots, M\} : \mathcal{I}_m \cap (\cup_{i=1}^{M} \mathcal{O}_i) = \varnothing\}$. The target distribution functions, $P_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m})$, for each component $m \in \mathcal{M}_1$ are therefore known. We estimate the densities $\hat{p}^n_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m})$ for $m \in \mathcal{M}_1$ from samples of the target distribution functions using a kernel density estimation method

that is strongly uniform convergent [32]. Then, for each $m \in \mathcal{M}_1$,

$$\lim_{n\to\infty} \hat{p}^n_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m}) = p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m}), \tag{2.19}$$

for all points of continuity of the target density $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m})$. Since all inputs to the components in the set $\mathcal{M}_1$ are absolutely continuous random variables, the measure of the set of discontinuous points of $p_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m})$ is zero. Eq. (2.3) defines the weighted empirical distribution function $P^n_{\mathbf{X};\mathbf{w}}(\mathbf{x}_{\mathcal{I}_m})$ and by Lemma 1 we have

$$\lim_{n\to\infty} P^n_{\mathbf{X};\mathbf{w}}(\mathbf{x}_{\mathcal{I}_m}) = P_{\mathbf{Y}}(\mathbf{x}_{\mathcal{I}_m}). \tag{2.20}$$

Let $\hat{\mathbf{Y}}$ be a random variable with distribution function $P^n_{\mathbf{X};\mathbf{w}}$, then $\hat{\mathbf{Y}} \xrightarrow{d} \mathbf{x}_{\mathcal{I}_m}$. Then, for each set in the partition of the input space of f_m, we have by Skorokhod's representation theorem, $f_m(\hat{\mathbf{Y}}) \xrightarrow{d} f_m(\mathbf{x}_{\mathcal{I}_m})$. Since the boundaries of the sets of the partition comprise a set of measure zero, this convergence applies over the complete domain of the function.

Then, by Lemma 2, we can obtain samples from the joint distribution function of the inputs and outputs for all components in $\mathcal{M}_1$. We then define $\mathcal{M}_2$ as the indices of those components with no dependence on upstream components other than those components in $\mathcal{M}_1$. That is, $\mathcal{M}_2 = \{m \in \{1, 2, \ldots, M\} : \mathcal{I}_m \cap (\cup_{i\notin\mathcal{M}_1} \mathcal{O}_i) = \varnothing\}$. The target distribution functions of all inputs for components $m \in \mathcal{M}_2$ are now available; thus the analysis described above for $\mathcal{M}_1$ applies to $\mathcal{M}_2$. We proceed by defining in turn $\mathcal{M}_3, \mathcal{M}_4, \ldots, \mathcal{M}_k$, where $\cup_{i=1}^k \mathcal{M}_k = \{1, 2, \ldots, M\}$, and obtaining samples from the distribution functions of all of the inputs and outputs for all components in each $\mathcal{M}_i$. From the samples generated for the components with indices in $\mathcal{M}_k$ we can construct the empirical distribution function of all system variables. By the strong law of large numbers, this empirical distribution function converges pointwise to the true distribution function of all system variables. □

The rate of convergence of our decomposition-based uncertainty analysis depends on several elements: the rate of convergence of the underlying Monte Carlo sam-

pling, the rate of convergence of the kernel density estimation, and the quality of the proposal distribution functions relative to their corresponding targets. As already discussed in Section 2.2, choosing a good proposal distribution is particularly important for achieving satisfactory convergence rates; a poor choice of proposal can lead to needing a prohibitive number of samples in the offline phase. While one can provide only qualitative guidance for proposal selection as discussed in Section 2.2, Section 2.6.2 presents a quantitative metric that compares proposal and target distributions for all inputs and coupling variables, thus highlighting *a posteriori* when a poor proposal choice may have compromised accuracy.

We further note here that, though we have shown convergence in distribution of all variables associated with a feed-forward system, for many uncertainty analysis tasks we may care only about statistics such as the mean and variance of a quantity of interest. Generally, if we have a function $g : \mathbb{R}^s \to \mathbb{R}$, that takes random inputs $(\xi_1, \xi_2, \ldots, \xi_s)^\top$, we can estimate the mean of $g(\xi_1, \xi_2, \ldots, \xi_s)$ using Monte Carlo simulation as

$$\bar{g} = \frac{1}{n} \sum_{i=1}^{n} g(\xi_1^i, \xi_2^i, \ldots, \xi_s^i), \tag{2.21}$$

where $(\xi_1^i, \xi_2^i, \ldots, \xi_s^i)^\top$ is the i^{th} sample realization of the random input to the function. By the strong law of large numbers, $\bar{g} \xrightarrow{a.s.} \mathbb{E}[g(\xi_1, \xi_2, \ldots, \xi_s)]$ as $n \to \infty$. We may use Monte Carlo simulation to estimate other integral quantities, such as the variance, as well, with almost sure convergence guaranteed by the strong law of large numbers. In our decomposition-based approach to uncertainty analysis, if the functions, f_m, corresponding to each component in the feed-forward system are also bounded and discontinuous on sets of zero measure, then, by an application of Skorokhod's representation theorem (see, e.g., Ref. [49]), the estimated mean and variance of any quantities of interest will converge to the true mean and variance. Thus, our decomposition-based multicomponent uncertainty analysis methodology can perform typical uncertainty analysis tasks in a provably convergent manner.

2.6.2 A posteriori indicator

Selection of an adequate proposal distribution function should use expert opinion and/or previous knowledge from past analyses. However, a poor proposal distribution may detrimentally affect the convergence performance of the distributed uncertainty assessment approach. In the general case, we cannot analyze *a priori* the effects of a given proposal distribution (if we had such information, we would use it to select a better proposal distribution); instead, we use quantitative indicators to determine *a posteriori* if the results are satisfactory.

Drawing from sequential Monte Carlo methods, we evaluate the quality of our proposal distributions once the importance weights are known, using the effective sample size,

$$n_{\text{eff}} = \frac{1}{\sum_{i=1}^{N} \left(w(\mathbf{x}^i)\right)^2}, \tag{2.22}$$

where $w(\mathbf{x}^i)$ is the importance weight associated to proposal sample $\mathbf{x}^i$ [63, 73, 34]. The effective sample size can range in value from $n_{\text{eff}} = 1$ to $n_{\text{eff}} = N$. A value of $n_{\text{eff}} = N$ indicates that the proposal and target distributions are equivalent, while a value of $n_{\text{eff}} = 1$ indicates an extremely poor proposal distribution where only one sample bears any weight in the weighted empirical distribution. The effective sample size is thus a suitable measure of the degeneracy of a given proposal distribution relative to a given target distribution. However, the effective sample size in Equation 2.22 does not account for the statistic of interest. To account for the statistic of interest, the effective sample size should use in place of the importance weights in Equation 2.22, the adjusted importance weights [40],

$$\widetilde{w}_i = \frac{|f(\mathbf{x}^i)| p_{\mathbf{Y}}(\mathbf{x}^i)/p_{\mathbf{X}}(\mathbf{x}^i)}{\sum_{i=1}^{N} |f(\mathbf{x}^i)| p_{\mathbf{Y}}(\mathbf{x}^i)/p_{\mathbf{X}}(\mathbf{x}^i)}. \tag{2.23}$$

To assess the quality of a distributed uncertainty assessment result, we recommend computing the effective sample size for each component once its target distribution is known. If a component's effective sample size is below a user-specified threshold, this indicates that sample impoverishment has occurred to a potentially detrimen-

tal degree, and we recommend reevaluating the local uncertainty analysis for that component. Upon completing the local uncertainty analysis of the component in question, the global compatibility step is computed again and the component's new effective sample size is evaluated. In the reevaluation step, the component's input target distribution can be used in place of the poor proposal distribution—of course this reevaluation breaks the strictly offline/online decomposition of our approach, but this recourse is necessary to provide some robustness. A second approach would be to use the moments from the component's input target distribution to reassign the proposal distribution as a heavier tailed distribution such as the multivariate student's t distribution. This is a desirable approach if the system input target distributions are expected to change throughout the course of the design process. The threshold for n_{eff} is a user choice; we investigate its effect with a simple example in the next section. Guidance can also be found in the importance sampling literature [70].

2.6.3 Convergence example

The following example lays out a step-by-step application of the approach and demonstrates the convergence in distribution for the system shown in Figure 2-3. The component functions are:

$$\begin{aligned} f_1\colon \xi_4 &= \xi_1 + \xi_2 \\ f_2\colon \xi_5 &= \xi_2 + \xi_3 \\ f_3\colon \xi_6 &= \xi_4 + \xi_5. \end{aligned}$$

The first phase of the approach is to conduct the offline analysis for each local component, which requires selecting proposal distributions for each component's inputs. In this example, the proposal distributions selected are Gaussian with conservative variance estimates. For component 1 we use $\xi_1 \sim \mathcal{N}(-0.5, 1.5)$ and $\xi_2 \sim \mathcal{N}(1.5, 2.0)$. For component 2 we use $\xi_2 \sim \mathcal{N}(-1.0, 2.0)$ and $\xi_3 \sim \mathcal{N}(-0.5, 2.5)$. For component 3 we use $\xi_4 \sim \mathcal{N}(1.5, 5.0)$ and $\xi_5 \sim \mathcal{N}(-1.5, 4.5)$. (Note that the proposal distribution for ξ_2 in the local analysis for component 1 is not necessarily the same as the

proposal distribution for ξ_2 in the local analysis for component 2.) Based on these proposal distributions, we conduct a local uncertainty analysis for each of the three components. In each case, we use a Monte Carlo simulation with n ranging from 100 to 1000 samples. Each Monte Carlo simulation results in a set of samples of the component outputs (ξ_4 for component 1, ξ_5 for component 2, and ξ_6 for component 3). For each component, we store the input and output sample data sets.

The next phase of the approach is the online analysis, which uses those samples pre-computed in the offline phase and does not require additional evaluations of any of the components. The first step in the online analysis is to specify target distributions for all system inputs, in this case ξ_1, ξ_2, and ξ_3. These target distributions represent the particular scenario for which we wish to analyze the system. In this example, we specify all three system inputs to have standard Gaussian distributions: $\xi_1 \sim \mathcal{N}(0,1)$, $\xi_2 \sim \mathcal{N}(0,1)$, and $\xi_3 \sim \mathcal{N}(0,1)$. We then begin with the upstream components, here component 1 and component 2. Given the newly specified target distributions and the proposal distributions assumed in the offline phase, for each component we compute the importance weights using Algorithm 1. We apply these importance weights to the corresponding samples of component outputs, here ξ_4 and ξ_5. This gives updated estimates of the component output distributions. The next step is to resolve the dependency structure between ξ_4 and ξ_5, induced by the shared input variable ξ_2. We achieve this using Algorithm 2, which evaluates the joint target density $p_{\mathbf{Y}}(\xi_4, \xi_5)$ at the input proposal samples of component 3. The last step of the online phase is to compute the importance weights for component 3, using the newly acquired target density $p_{\mathbf{Y}}(\xi_4, \xi_5)$ evaluations from Algorithm 2. Applying these importance weights to the pre-computed samples of ξ_6 leads to the final updated estimate of the system output.

For this particular scenario, we can compute the true system output distribution analytically as a Gaussian distribution, $\xi_6 \sim \mathcal{N}(0,6)$. We compare the numerical results from our decomposition approach to this analytical solution. The convergence

in distribution is demonstrated with the Cramer von-Mises criterion,

$$\omega = \int_{-\infty}^{\infty} [P^n_{\mathbf{X};\mathbf{w}}(\xi_6) - P_{\mathbf{Y}}(\xi_6)]^2 \, \mathrm{d}P_{\mathbf{Y}}(\xi_6), \tag{2.24}$$

where $P^n_{\mathbf{X};\mathbf{w}}(\xi_6)$ and $P_{\mathbf{Y}}(\xi_6)$ are the weighted empirical and analytical distribution functions of ξ_6, respectively. The Cramer von-Mises criterion is estimated using Monte Carlo simulation, where samples of ξ_6 are drawn from the analytical distribution $P_{\mathbf{Y}}(\xi_6)$.

Figure 2-5 presents the averages over 100 independent simulations of an all-at-once system Monte Carlo uncertainty analysis and our decomposition-based approach. The decomposition-based results implemented the kernel density estimation method and used 25 samples, S, to resolve the dependency among variables in Algorithm 2. The result shows that for the same number of overall samples per component, n, we incur a larger error than the system-level Monte Carlo simulation. This error is due to Algorithm 5. Specifically the target measure needs to be absolutely continuous with respect to the proposal measure, which leads to a conservative choice of proposal density. This in turn means that there exist proposal samples that have negligible importance weight ("wasted" samples). The closer the proposal distribution to the target distribution, the smaller this offset. This is the price we pay for decomposition—it is important to emphasize that our goal is not an improvement in computational efficiency, but rather the ability to manage system complexity and to analyze uncertainty in systems for which an integrated Monte Carlo simulation approach may not be tractable or feasible. Furthermore, our approach can perform the local uncertainty analysis of each component concurrently. This could lead to significant further run time improvement compared to the system Monte Carlo uncertainty analysis, which can perform uncertainty analysis on downstream models only after their respective dependent upstream components' uncertainty analyses are complete.

Figure 2-6 demonstrates the effects of changing the proposal distribution on the convergence of the decomposition-based approach for this simple example. Here we consider modifying the proposal distribution of component 2 while keeping the pro-

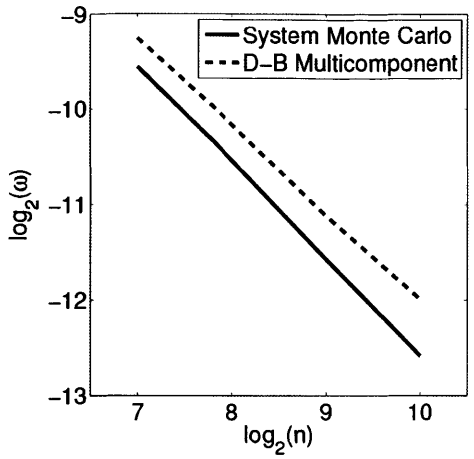

Figure 2-5: The results indicate the output of interest, ξ_6, Cramer von-Mises criterion converges with the number of samples. The system Monte Carlo weighted empirical distribution function uses $\mathbf{w} = \mathbf{1}_n^T$. The decomposition-based multicomponent weighted empirical distribution function uses weights computed via Algorithm 1.

posal distributions of component 1 and component 3 the same as the previous analysis, with $n = 256$. The proposal distribution for component 2 is a bivariate Gaussian distribution with zero mean and diagonal variance set to 1.0, 2.5, 5.0, and 10 representing a perfect, good, moderate, and poor proposal distribution, respectively. These four cases result respectively in values of $n_{\text{eff}} = 247, 158, 90$, and 48. As the proposal distribution improves, the ratio n_{eff}/n increases (for this example n is fixed), which in turn leads to improved estimation of the outputs of interest as shown by the decreasing Cramer von Mises criterion in Figure 2-6.

2.7 Application to a gas turbine system

In this section we present a demonstration of the decomposition-based uncertainty analysis approach for a gas turbine blade application. We compare the results of our method with all-at-once Monte Carlo system uncertainty analysis.

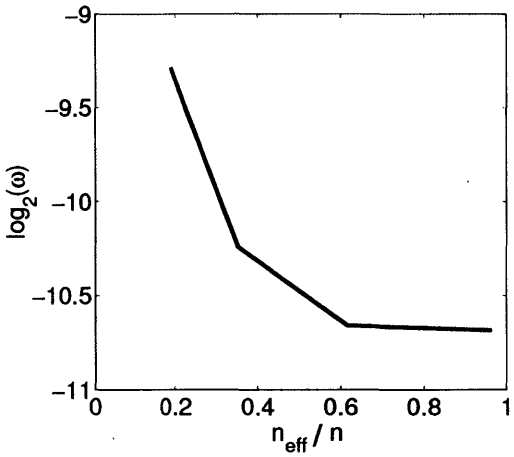

Figure 2-6: The results show the implications of selecting a poor proposal distribution for component f_2 with $n = 256$. As n_{eff} approaches n, indicating a better proposal distribution, the accuracy of our estimate improves.

2.7.1 Gas turbine system setup

Our application problem consists of four components, each representing a disciplinary analysis: blade heat transfer, blade lifetime, engine performance, and an economic model. The functional relationships and random variables are shown in Figure 2-7. This application is representative of an organizational multidisciplinary environment where different groups are responsible for different aspects of the gas turbine design and assessment. The specific objective of our analysis is to quantify the effects of uncertainties throughout the gas turbine design process on the output of interest, here the economics of the product. We consider the uncertain system inputs shown in Table 2.2. The distributions shown in the table are the target distributions used for our analysis (i.e., they represent the particular scenario of interest in the system uncertainty analysis). These target distributions are considered unknown when conducting the local uncertainty analysis for each of the four components.

Heat transfer model The blade heat transfer model simulates a cooled gas turbine blade in hot gas path flow using finite element analysis. The uncertain inputs to this subsystem are shown in Figure 2-7. We consider three blade passages, each with its own independent coolant temperature variable. Thus there are eight uncertain

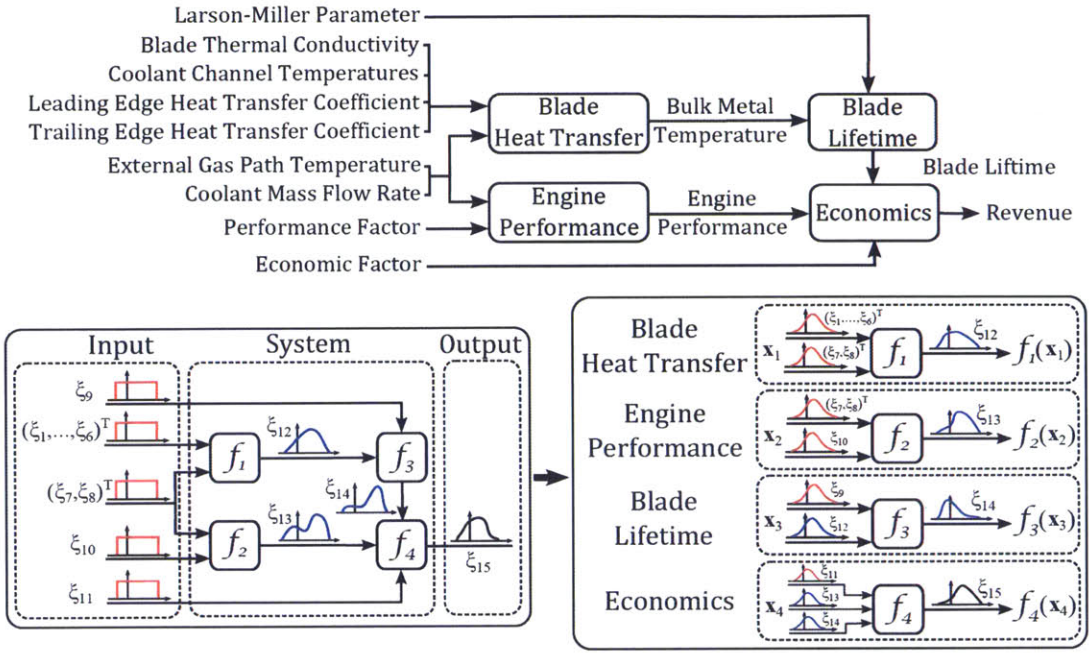

Figure 2-7: The gas turbine application problem contains four components, each representing a disciplinary analysis: heat transfer, structures, performance, and economics.

inputs to this component. External heat transfer along the pressure and suction side surfaces is computed as

$$htc(\zeta) = h_{TE} + (h_{LE} - h_{TE}) \cdot exp\big(-4 \cdot (\zeta/c)^2\big), \tag{2.25}$$

where ζ is the chordwise spatial coordinate and c is the blade chord length, here taken as $c = 0.04$ [m]. The output of the heat transfer model is bulk metal temperature, T_{bulk} [K]. The relationship between the input and output variables is computed using a finite element method to solve the heat equation. The blade profile and mesh along with the random variables are shown in Figure 2-8.

The local uncertainty analysis for this model is conducted using the proposal distributions shown in Table 2.3. Note that for our particular demonstration we have chosen the proposal variances conservatively to ensure adequate support in the proposal samples, as discussed in Section 2.3.

Table 2.2: Gas turbine system input uncertainty distributions where $\mathcal{U}(a, b)$ represents a uniform distribution between the lower limit a and upper limit b.

Variable	Name	Description	Units	Distribution
ξ_1	T_{c1}	First Passage Coolant Temperature	K	$\mathcal{U}(590, 610)$
ξ_2	T_{c2}	Second Passage Coolant Temperature	K	$\mathcal{U}(640, 660)$
ξ_3	T_{c3}	Third Passage Coolant Temperature	K	$\mathcal{U}(690, 710)$
ξ_4	k	Blade Thermal Conductivity	W/m/K	$\mathcal{U}(29, 31)$
ξ_5	h_{LE}	Leading Edge Heat Transfer Coefficient	$W/m^2/K$	$\mathcal{U}(1975, 2025)$
ξ_6	h_{TE}	Trailing Edge Heat Transfer Coefficient	$W/m^2/K$	$\mathcal{U}(975, 1025)$
ξ_7	$\dot{m}$	Coolant Mass Flow Rate	kg/sec	$\mathcal{U}(0.108, 0.132)$
ξ_8	T_g	External Gas Path Temperature	K	$\mathcal{U}(1225, 1275)$
ξ_9	LMP	Larson-Miller Parameter	–	$\mathcal{U}(2.45{\cdot}10^4, 2.55{\cdot}10^4)$
ξ_{10}	F_{perf}	Performance Factor	–	$\mathcal{U}(0.85, 0.95)$
ξ_{11}	F_{econ}	Economic Factor	–	$\mathcal{U}(0.9, 1.1)$

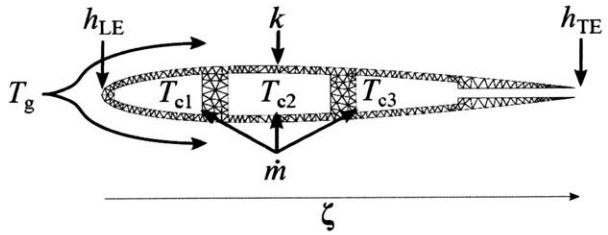

Figure 2-8: The gas turbine blade profile and mesh, along with the random input variables.

Lifetime model The lifetime model estimates the expected time until blade failure assuming a Larson-Miller [96] nickel super alloy stress-to-failure scenario. As shown in Figure 2-7, the inputs to this subsystem are bulk temperature, T_{bulk}, and the Larson-Miller failure parameter, LMP. The output is expected time until failure, t_{fail} [hr]. The relationship between the input and output variables is given by

$$t_{fail} = exp\big(LMP/T_{bulk} - 20\big). \tag{2.26}$$

The input proposal distributions assumed for the local uncertainty analysis of this component are given in Table 2.4.

Performance model A high-fidelity gas turbine performance model would account for compressor coolant flow extraction, leakage losses, and mixing losses which is

Table 2.3: Heat transfer model input proposal uncertainty distributions.

Variable	Name	Description	Units	Distribution
ξ_1	T_{c1}	First Passage Coolant Temperature	K	$\mathcal{N}(595, 75)$
ξ_2	T_{c2}	Second Passage Coolant Temperature	K	$\mathcal{N}(645, 75)$
ξ_3	T_{c3}	Third Passage Coolant Temperature	K	$\mathcal{N}(705, 75)$
ξ_4	k	Blade Thermal Conductivity	W/m/K	$\mathcal{N}(29, 1.5)$
ξ_5	h_{LE}	Leading Edge Heat Transfer Coefficient	$W/m^2/K$	$\mathcal{N}(2025, 1500)$
ξ_6	h_{TE}	Trailing Edge Heat Transfer Coefficient	$W/m^2/K$	$\mathcal{N}(1000, 500)$
ξ_7	$\dot{m}$	Coolant Mass Flow Rate	kg/sec	$\mathcal{N}(0.12, 10^{-4})$
ξ_8	T_g	External Gas Path Temperature	K	$\mathcal{N}(1260, 450)$

Table 2.4: Blade lifetime model input proposal uncertainty distributions.

Variable	Name	Description	Units	Distribution
ξ_{10}	LMP	Larson-Miller Parameter	–	$\mathcal{N}(2.5{\cdot}10^4, 2{\cdot}10^5)$
ξ_{12}	T_{bulk}	Bulk Metal Temperature	K	$\mathcal{N}(865, 400)$

beyond the scope of this work. Instead, a simplified low-fidelity model is implemented to evaluate the maximum power. The performance model rewards high external hot gas path temperatures and penalizes coolant flow usage. As shown in Figure 2-7, the inputs to this subsystem are external gas temperature, T_{gas}, performance factor, F_{perf}, and coolant mass flow, $\dot{m}$. The performance factor, F_{perf}, is introduced to account for the effects on engine performance of randomness associated with other gas turbine components. The output of the performance model is engine performance, P_{eng}, defined as

$$P_{eng} = F_{perf} \cdot (\dot{m}_o - N \cdot \dot{m}) \cdot C_p \cdot T_o \cdot \left(T_g/T_o - 2 \cdot \sqrt{T_g/T_o} + 1\right), \tag{2.27}$$

where T_o is the inlet compressor temperature, $\dot{m}_o$ is the inlet compressor flow rate, N is the number of gas turbine blades, and C_p is the specific heat at constant pressure. These parameters are treated deterministically and set to the values $T_o = 300$ [K], $\dot{m}_o = 430$ [kg/sec], $N = 90$, and $C_p = 1003.5$ [J/kg/K]. The input proposal distributions assumed for the local uncertainty analysis of this component are given in Table 2.5.

Table 2.5: Performance model input proposal uncertainty distributions.

Variable	Name	Description	Units	Distribution
ξ_7	$\dot{m}$	Coolant Mass Flow Rate	kg/sec	$\mathcal{N}(0.115,\ 10^{-4})$
ξ_8	T_g	External Gas Path Temperature	K	$\mathcal{N}(1240,\ 500)$
ξ_{10}	F_{perf}	Performance Factor	–	$\mathcal{N}(0.9,\ 7.5 \cdot 10^{-3})$

Economics model The economics model simulates the revenue from the operating gas turbine. The model rewards a high-performance gas turbine engine and penalizes a gas turbine engine that introduces risk of failure. As shown in Figure 2-7, the inputs to this subsystem are expected time until failure, t_{fail}, engine performance, P_{eng}, and economic factor, F_{econ}. The economic factor, F_{econ}, is introduced to account for randomness associated with other gas turbine components not represented in the models. The output is revenue, r_{econ}, defined as

$$r_{econ} = F_{econ} \cdot t_{fail} \cdot P_{eng} \cdot c_o, \tag{2.28}$$

where c_o is the cost of energy which is treated deterministically and set to the value $c_o = 0.07$ [\$/kWh]. The input proposal distributions assumed for the local uncertainty analysis of this component are given in Table 2.6.

Table 2.6: Economics model input proposal uncertainty distributions.

Variable	Name	Description	Units	Distribution
ξ_{11}	F_{econ}	Economic Factor	–	$\mathcal{N}(1.0,\ 0.01)$
ξ_{13}	t_{fail}	Blade Lifetime	year	$\mathcal{N}(425,\ 6{\cdot}10^4)$
ξ_{14}	P_{eng}	Engine Performance	MW	$\mathcal{N}(120,\ 150)$

2.7.2 Uncertainty analysis

In the "offline phase", the local uncertainty analyses are carried out for each component individually, using the input proposal distributions specified in Tables 2.3– 2.6. Each component uses n independent samples in its local Monte Carlo simulation. (Note that the number of samples does not need to be the same across components.) Output

samples for each component are stored in a database.

The "online phase" considers a system uncertainty analysis for the system input distributions shown in Table 2.2. Global compatibility satisfaction begins by considering the Heat Transfer and Performance components, which have only system inputs and thus require no information from upstream disciplines. Using Algorithm 5 and kernel density estimation, we obtain target densities $p_{\mathbf{Y}}(\xi_{12}, \xi_7, \xi_8)$ and $p_{\mathbf{Y}}(\xi_{13}, \xi_7, \xi_8)$. The same procedure is applied to the Lifetime component using the recently acquired density $p_{\mathbf{Y}}(\xi_{12})$ to obtain the target density $p_{\mathbf{Y}}(\xi_{14}, \xi_{12})$. Using target densities $p_{\mathbf{Y}}(\xi_{12}, \xi_7, \xi_8)$, $p_{\mathbf{Y}}(\xi_{13}, \xi_7, \xi_8)$, and $p_{\mathbf{Y}}(\xi_{14}, \xi_{12})$ along with Algorithm 2 with $S = 200$ samples, we obtain the desired target density $p_{\mathbf{Y}}(\xi_{13}, \xi_{14})$ evaluated at the Economic model's proposal samples. The global compatibility satisfaction procedure is then performed on the Economics model to obtain the system output of interest, revenue, with target density $p_{\mathbf{Y}}(\xi_{15})$. The Cramer von-Mises convergence plots for variables ξ_{12}, ξ_{13}, ξ_{14}, and ξ_{15} averaged over 100 independent simulations are shown in Figure 2-9. The true distribution is defined by the empirical distribution function generated using a system Monte Carlo simulation with 10^6 samples.

The system output of interest distribution function using the decomposition-based uncertainty analysis approach is given in Figure 2-10. For comparison, the proposal distribution function and the system Monte Carlo uncertainty analysis distribution function are also shown on Figure 2-10. The results show that the decomposition-based approach propagated, in the online phase, the target system input uncertainty distributions through the system to obtain an adequate representation of the system output of interest distribution. We emphasize that this online phase required no additional evaluations of any of the component models. Our decomposition-based approach therefore provides a quantitative means of calculating the system output of interest relevant statistics and failure probabilities.

The small discrepancy between the decomposition-based uncertainty analysis approach and the system Monte Carlo uncertainty analysis approach is due to the errors introduced by the finite number of samples used in the density estimation step, the Monte Carlo approximation used in Algorithm 2, and the sample impoverishment in-

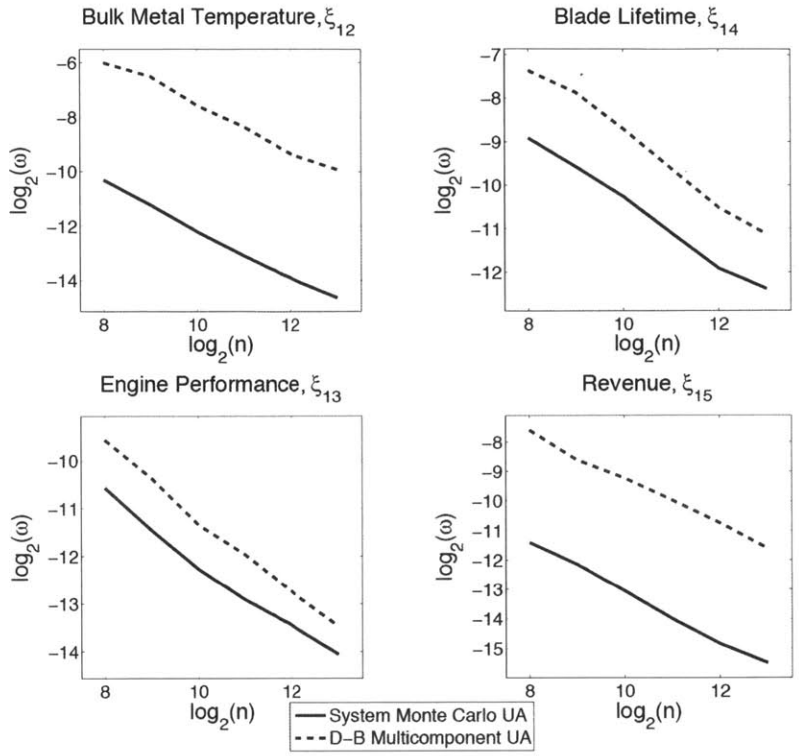

Figure 2-9: The Cramer von-Mises convergence plots are shown for the intermediate variables ξ_{12}, ξ_{13}, and ξ_{14} as well as for the system output of interest, revenue, ξ_{15}. The solid lines are the result obtained from a system Monte Carlo simulation. The dashed lines are the result obtained using our decomposition-based multicomponent uncertainty analysis.

troduced by the requirement that target distributions be absolutely continuous with respect to their proposal distributions. The sample impoverishment error can be minimized by using more appropriate proposal distributions. However, it is not always possible to correctly predict the range of the target distribution. This is one of the prices to pay for decomposition.

2.7.3 Flexibility of decomposition-based uncertainty analysis

A benefit of our decomposition-based approach is that if any system input distributions are modified, yet remain absolutely continuous with respect to their proposal distribution, then the system output of interest distribution function can be re-computed with no additional component analyses. For example, if the system input variables

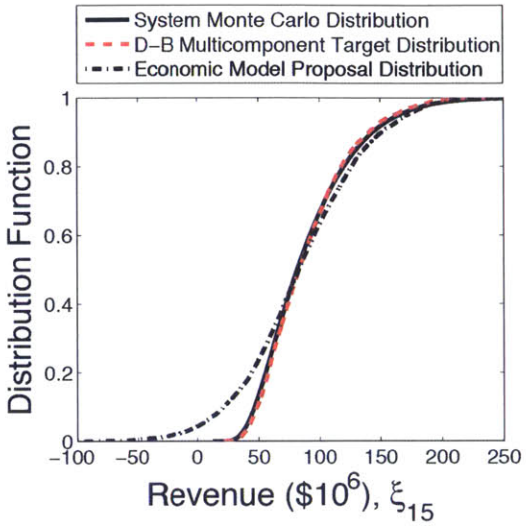

Figure 2-10: The system output of interest, revenue, distribution function using $n = 8192$ samples is shown in millions of dollars. The solid line is the result obtained from a system Monte Carlo simulation. The dashed line is the result obtained from the decomposition-based multicomponent uncertainty analysis. The dash-dot line is the result obtained from the local uncertainty analysis of the Economics model.

ξ_5, ξ_8, and ξ_9 are modified from those in Table 2.2 to those in Table 2.7, then the results given by the system Monte Carlo uncertainty analysis are invalid. However, our decomposition-based approach can, with no additional evaluations of any of the component models, evaluate the system output of interest distribution function as shown in Figure 2-11. For comparison, the previous system Monte Carlo uncertainty analysis distribution function and a new system Monte Carlo uncertainty analysis distribution function, which required evaluating the entire system again, are also shown on Figure 2-11. The results show the decomposition-based approach, without re-evaluating any component, approximated the distribution function accurately.

Table 2.7: Updated gas turbine system input uncertainty distributions.

Variable	Name	Description	Units	Distribution
ξ_5	h_{LE}	Leading Edge Heat Transfer Coefficient	$W/m^2/K$	$\mathcal{U}(2025, 2075)$
ξ_8	T_g	External Gas Path Temperature	K	$\mathcal{U}(1240, 1280)$
ξ_9	LMP	Larson-Miller Parameter	–	$\mathcal{U}(2.425{\cdot}10^4, 2.525{\cdot}10^4)$

Likewise, a modification to a component would require the system Monte Carlo

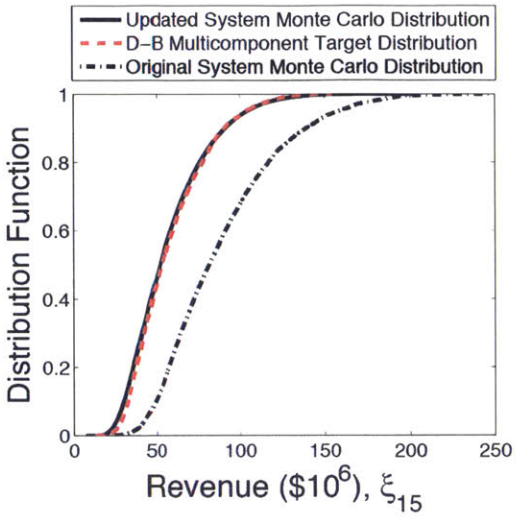

Figure 2-11: The system output of interest, revenue, distribution function using $n = 8192$ samples is shown in millions of dollars. The solid line is the result obtained from an updated system Monte Carlo simulation which required evaluating the entire system again. The dashed line is the result obtained from the decomposition-based multicomponent uncertainty analysis using the online phase only. The dash-dot line is the result from the previous Monte Carlo uncertainty analysis.

uncertainty analysis approach to recompute the samples associated with the modified component and any components that depend on the modified component. In contrast, our decomposition-based uncertainty analysis approach would only have to perform the local uncertainty analysis on the modified component and those components for which the target distribution is no longer absolutely continuous with respect to the proposal distribution. For example, if the Heat Transfer component modified the heat transfer enhancement in the cooling channels from a factor of 2.5 to 2.25, then the target density $p_{\mathbf{Y}}(\xi_{12})$ would still be absolutely continuous with respect to the Lifetime model proposal density $p_{\mathbf{X}}(\xi_{12})$ as shown in Figure 2-12. As a result, the decomposition-based approach would not require the Lifetime model or the Economics model to perform a local uncertainty analysis whereas the system Monte-Carlo uncertainty analysis approach would. Instead the decomposition-based approach evaluates the system output of interest distribution shown in Figure 2-12 using only the online phase.

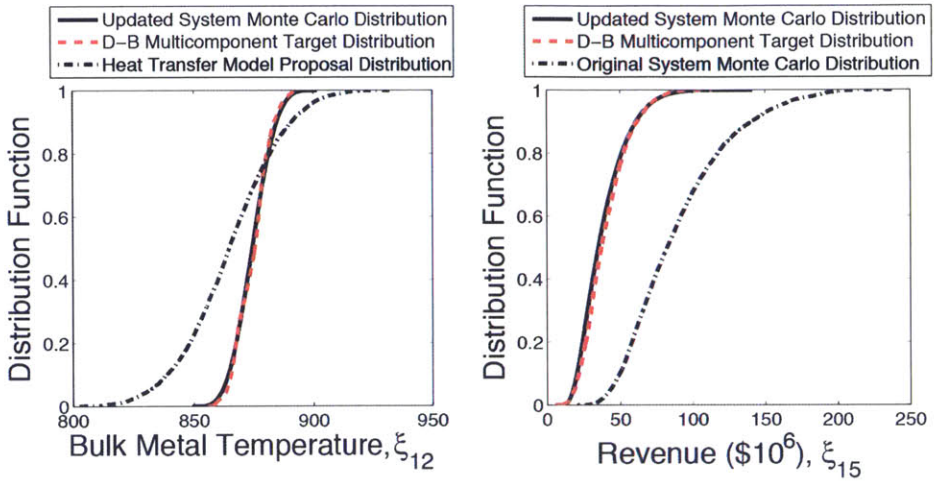

Figure 2-12: The bulk metal temperature, ξ_{12}, is shown on the left. The results shows that the proposal distribution (dash-dot line) of the bulk metal temperature of the Lifetime model supports the target distribution (dashed line) coming from the Heat Transfer model. The system Monte Carlo uncertainty analysis results, solid line, required evaluating the the Heat Transfer, Lifetime, and Economics model, whereas the decomposition-based multicomponent uncertainty analysis results were obtained using the online phase only. The revenue, ξ_{15}, in millions of dollars is shown on the right. The solid line is the result obtained from a system Monte Carlo uncertainty analysis. The dashed line is the result obtained from the decomposition-based multicomponent uncertainty analysis using the online phase only. The dash-dot line is the result obtained from the previous Monte Carlo uncertainty analysis.

2.8 Pitfalls of decomposition-based uncertainty analysis

The performance of our decomposition-based uncertainty analysis approach is largely influenced by how well we can perform the change of measure (e.g., approximating ν using samples drawn from μ). In Section 2.3, we accomplished the change of measure with the ratio of probability density functions where the probability density functions were constructed using density estimation. However, density estimation is particularly challenging in cases of high dimension [106, 53, 120]. As a result, in this research we have investigated methods to overcome this challenge. In Chapter 4, we present a novel approach to the change of measure problem which entirely avoids

the need for density estimation. Our approach allows us to perform the change of measure in dimensions which were previously not possible using density estimation. In Chapter 5, we use the component global sensitivity analysis to determine which inputs to the component are most important for the change of measure. This allows us to reduce the dimension associated with the change of measure to sizes which are feasible.

Chapter 3

Decomposition-Based Global Sensitivity Analysis

In this chapter we perform a decomposition-based global sensitivity analysis of the feed-forward multicomponent system. Our approach utilizes the decomposition-based uncertainty analysis algorithm presented in Chapter 2 to perform the variance-based global sensitivity analysis in a decomposition-based manner. The purpose of conducting a decomposition-based global sensitivity analysis is to identify key system inputs that significantly impact the output of interest variation while incorporating the benefits of a decomposition-based approach previously stated in Section 1.1.

This chapter presents a review of the variance-based global sensitivity analysis method by I.M. Sobol' and sets notation for subsequent developments in Section 3.1. Following that discussion we present the challenges associated with decomposition-based global sensitivity analysis and develop our decomposition-based global sensitivity analysis approach in Section 3.2. Lastly, in Section 3.3, our decomposition-based global sensitivity analysis algorithm is demonstrated on an aerospace system application problem.

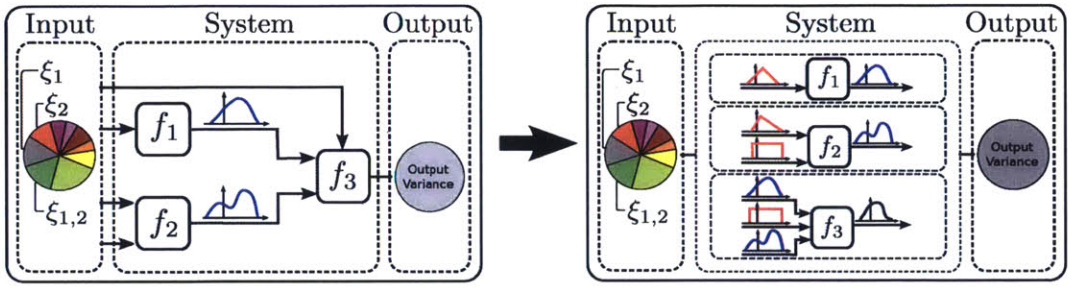

Figure 3-1: The proposed method for multicomponent global sensitivity analysis utilizes the decomposition-based uncertainty analysis algorithm presented in Chapter 2 to evaluate the statistics of interest necessary for the variance-based method of I.M. Sobol'. The objective of a variance-based global sensitivity analysis method is to apportion the output of interest variance across the system inputs and is depicted here using the pie chart.

3.1 Variance-based global sensitivity analysis

In this section we review the variance-based global sensitivity analysis method by I.M. Sobol' and set notation for subsequent developments [112, 111, 94, 66, 71]. The objective of performing a variance-based global sensitivity analysis is depicted on the left in Figure 3-1, where the pie chart representing the variance of the system output is decomposed using the variance-based global sensitivity analysis into the contributions from the system inputs. The outcome of a global sensitivity analysis permits a ranking of system inputs that can be used in different developments; here we are interested in the prioritization of the system inputs so that we may allocate future research efforts. As stated in Section 1.4, we assume the system inputs are independently distributed which allows us to implement the variance-based global sensitivity analysis method by I.M. Sobol'. This derivation follows from the work of Homma and Saltelli [103].

3.1.1 Formulation

We present the variance-based global sensitivity analysis formulation for a single component, denoted by f. However, f could represent the entire system, considered as an integrated single component, as illustrated on the left in Figure 3-1. As previously mentioned, let the system inputs be independently distributed, that is, we may express

the system's input distribution as a product-type distribution, $P_{\mathbf{Y}}(\mathbf{t}) = \prod_{i=1}^{d} P_{\mathbf{Y}}(t_i)$. Since f is a square-integrable function with respect to the induced measure μ, we can express it as a finite hierarchical expansion in terms of inputs with increasing dimensions [56]. This expansion is given by the compact form

$$\begin{aligned} f(\mathbf{t}) &= f_0 + \sum_{i=1}^{d} f_{\{i\}}(t_i) + \sum_{i=1}^{d}\sum_{j>i}^{d} f_{\{i,j\}}(t_i, t_j) + \cdots + f_{\{1,2,\ldots,d\}}(\mathbf{t}), \\ &= \sum_{\mathcal{A}\subseteq\{1,\ldots,d\}} f_{\mathcal{A}}(\mathbf{t}_{\mathcal{A}}), \end{aligned} \tag{3.1}$$

where f_0 is a constant, $f_{\{i\}}$ is a function of only $\mathbf{X}_{\{i\}}$, $f_{\{i,j\}}$ is a function of only $\mathbf{X}_{\{i\}}$ and $\mathbf{X}_{\{j\}}$, etc. The summation in Equation 3.1 comprises of 2^d subcomponent[1] functions, with each function depending on a group of variables indexed by a particular subset of $\{1,\ldots,d\}$, including the empty set. This expansion can be made unique by requiring all nonconstant subcomponent functions $f_{\mathcal{A}}$ to integrate to zero with respect to the marginal density of each random variable in $\mathcal{A}$ [111], that is

$$\int_{\mathbb{R}} f_{\mathcal{A}}(\mathbf{t}_{\mathcal{A}}) p_{\mathbf{Y}}(t_i) \mathrm{d}t_i = 0, \; i \in \mathcal{A} \neq \varnothing. \tag{3.2}$$

This constraint, which we refer to as the strong annihilating constraint, ensures that each nonconstant subcomponent has zero mean,

$$\mathbb{E}_{\mathbf{Y}}[f_{\mathcal{A}}] = 0, \tag{3.3}$$

and any two distinct subcomponents within the expansion are orthogonal,

$$\mathbb{E}_{\mathbf{Y}}[f_{\mathcal{A}} f_{\widetilde{\mathcal{A}}}] = 0, \tag{3.4}$$

where $\varnothing \neq \mathcal{A} \subseteq \{1,\ldots,d\}$, $\varnothing \neq \widetilde{\mathcal{A}} \subseteq \{1,\ldots,d\}$, and $\widetilde{\mathcal{A}} \neq \mathcal{A}$.

Integrating Equation 3.1 with respect to the marginal distribution $P_{\mathbf{Y}_{\mathcal{A}^c}}$, that is over all variables except $\mathcal{A}$, and using Equation 3.2 yields the subcomponent func-

[1]Not to be confused with the components which comprise the system.

tions [114, 66],

$$\begin{aligned} f_0 &= \int_{\mathbb{R}^d} f(\mathbf{t}) p_{\mathbf{Y}}(\mathbf{t}) \mathrm{d}\mathbf{t}, \\ f_{\mathcal{A}}(\mathbf{x}_{\mathcal{A}}) &= \int_{\mathbb{R}^{d-|\mathcal{A}|}} f(\mathbf{x}_{\mathcal{A}}, \mathbf{t}_{\mathcal{A}^c}) \prod_{i \in \mathcal{A}^c} p_{\mathbf{Y}}(t_i) \mathrm{d}t_i - \sum_{\tilde{\mathcal{A}} \subset \mathcal{A}} f_{\tilde{\mathcal{A}}}(\mathbf{x}_{\tilde{\mathcal{A}}}). \end{aligned} \tag{3.5}$$

Applying the expectation operator to f and recognizing the strong annihilating constraint we obtain the mean,

$$\mathbb{E}_{\mathbf{Y}}[f] = \int_{\mathbb{R}^d} f(\mathbf{t}) p_{\mathbf{Y}}(\mathbf{t}) \mathrm{d}\mathbf{t} = f_0. \tag{3.6}$$

Similarly, applying the expectation operator to $(f - \mathbb{E}_{\mathbf{Y}}[f])^2$ we obtain the variance which we denote by D,

$$D = \int_{\mathbb{R}^d} \left(f(\mathbf{t}) - \mathbb{E}_{\mathbf{Y}}[f]\right)^2 p_{\mathbf{Y}}(\mathbf{t}) \mathrm{d}\mathbf{t} = \int_{\mathbb{R}^d} f(\mathbf{t})^2 p_{\mathbf{Y}}(\mathbf{t}) \mathrm{d}\mathbf{t} - f_0^2. \tag{3.7}$$

Given the decomposition of f in Equation 3.1 and recognizing the strong annihilating constraint (i.e., orthogonality between any two distinct subcomponent functions) the expectation operator applied to f^2 results in,

$$\int_{\mathbb{R}^d} f(\mathbf{t})^2 p_{\mathbf{Y}}(\mathbf{t}) \mathrm{d}\mathbf{t} = f_0^2 + \sum_{\varnothing \neq \mathcal{A} \subseteq \{1,2,\ldots,d\}} D_{\mathcal{A}}, \tag{3.8}$$

where the partial variances are given by

$$D_{\mathcal{A}} = \int_{\mathbb{R}^{|\mathcal{A}|}} f_{\mathcal{A}}(\mathbf{t})^2 p_{\mathbf{Y}}(\mathbf{t}) \mathrm{d}\mathbf{t}. \tag{3.9}$$

Combining Equation 3.7 with Equation 3.8 implies

$$D = \sum_{i=1}^{d} D_{\{i\}} + \sum_{i,j=1}^{d} D_{\{i,j\}} + \cdots + D_{\{1,2,\ldots,d\}} = \sum_{\varnothing \neq \mathcal{A} \subseteq \{1,2,\ldots,d\}} D_{\mathcal{A}}, \tag{3.10}$$

which was notionally depicted on the left in Figure 3-1.

The global sensitivity indices are defined as

$$S_{\mathcal{A}} = \frac{D_{\mathcal{A}}}{D}, \tag{3.11}$$

where the sum of all global sensitivity indices is unity. Global sensitivity indices with only a single subscript, (e.g., $S_{\{i\}}$), are called main effect indices. Inputs with large main effect indices are known as the "most influential factors", or the inputs which, on average, once fixed, would result in the greatest reduction in variance. Global sensitivity indices with multiple subscripts, (e.g., $S_{\{i,j\}}$), are called interaction effect indices. Summing the main effect index of $\{i\}$ along with all the interaction effect indices that include input $\{i\}$ results in the total effect index, τ_i, which is defined as

$$\tau_{\{i\}} = S_{\{i\}} + \sum_{\varnothing \neq \mathcal{A} \subseteq \{1,2,\ldots,d\} \setminus \{i\}} S_{\{i,\mathcal{A}\}}. \tag{3.12}$$

Total effect indices quantify how influential an input is on the output of interest variance. If an input has a negligible total effect index (i.e., $\tau_{\{i\}} \ll 1$), then this is a necessary and sufficient condition for $\{i\}$ being a noninfluential input with respect to the output of interest variation [103].

3.1.2 Computing sensitivity indices

The main and total effect indices can be computed via Monte Carlo simulation as follows [103], where hat quantities denote an estimate of the corresponding true quantities. For d inputs, the calculation of sensitivity indices requires $(d+2)$ Monte Carlo simulations (each with n model evaluations) if both the main effect and total effect indices are desired [103]. The estimates for the mean and variance are computed by,

$$\begin{aligned} \hat{f}_0 &= \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{y}^i) \\ \hat{D} &= \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{y}^i)^2 - \hat{f}_0^2, \end{aligned} \tag{3.13}$$

where $\mathbf{y}^i = [y_1^i, \cdots, y_d^i]^\top$ is a vector of random samples drawn from their respective target distributions. To compute the main effect index for $\mathcal{A} = \{i\}$ would require evaluating the previous n model evaluations with the random samples $y_{\mathcal{A}^c}^j$ for all $j \in \{1, \ldots, n\}$ having been resampled from their respective target distributions. The main effect index is evaluated as follows:

$$\hat{S}_{\{i\}} = \frac{1}{n\hat{D}} \sum_{j=1}^{n} f(\mathbf{y}^j) f([\tilde{y}_1^j, \ldots, \tilde{y}_{\{i-1\}}^j, y_{\{i\}}^j, \tilde{y}_{\{i+1\}}^j, \ldots, \tilde{y}_d^j]) - \hat{f}_0^2, \tag{3.14}$$

where $\tilde{y}$ denotes the sample which has been resampled from its respective target distribution. Similarly, to compute the total effect index for $\mathcal{A} = \{i\}$ would require evaluating the initial n model evaluations with the random samples $y_{\mathcal{A}}^j$ for all $j \in \{1, \ldots, n\}$ having been resampled from its respective target distribution. The total effect index is evaluated as follows:

$$\hat{\tau}_i = \frac{1}{n\hat{D}} \sum_{j=1}^{n} f(\mathbf{y}^j) f([y_1^j, \ldots, y_{\{i-1\}}^j, \tilde{y}_{\{i\}}^j, y_{\{i+1\}}^j, \ldots, y_d^j]) - \hat{f}_0^2. \tag{3.15}$$

3.2 Decomposition-based global sensitivity analysis

From a system-level perspective, the objective is to identify which of the system inputs are the most influential with respect to the variance of the system output of interest. To rank the influential inputs in terms of importance we must evaluate the main effect index, Equation 3.14, for each system input. However, since we cannot assemble the system, we may not directly evaluate the main effect indices using Monte Carlo simulation [103]. Instead, we must evaluate Equation 3.14 using a decomposition-based approach. We rely on the decomposition-based uncertainty analysis algorithm presented in Chapter 2 to quantify the system's main effect indices. We assume the system inputs are independently distributed, therefore we can apply the variance-based method by I.M. Sobol'. This section provides the procedure to evaluate, using the decomposition-based uncertainty analysis algorithm, the main effect indices of the feed-forward multicomponent system.

3.2.1 Formulation

In Section 3.1, we derived the main effect index and presented an approach to estimate this quantity using Monte Carlo simulation. Here we shall express the main effect index with respect to the i^{th} input variable by combining Equation 3.9 with Equation 3.5 and the strong annihilating constraint as,

$$S_{\{i\}} = \frac{D_{\{i\}}}{D} = \frac{var_{Y_i}(\mathbb{E}_{Y_i^c}[f \mid Y_i])}{var_{\mathbf{Y}}(f)}. \tag{3.16}$$

Similarly, sensitivity indices of higher-order subcomponent functions can also be expressed in this fashion. Take for instance the sensitivity index associated with the interaction term $f_{\{i,j\}}$,

$$S_{\{i,j\}} = \frac{var_{\{Y_i,Y_j\}}(\mathbb{E}_{\{Y_i,Y_j\}^c}[f \mid Y_i, Y_j])}{var_{\mathbf{Y}}(f)}, \tag{3.17}$$

where $i \neq j$.

Since the equations above require evaluating the expectation of f conditioned on random variables we cannot directly implement a decomposition-based methodology. Instead, we overcome this aforementioned challenge by evaluating the expectation of f conditioned on finite measures and not on random variables. For example, in Equation 3.16, instead of conditioning on the random variable Y_i which has zero measure, we condition over a range which encompasses Y_i and by construction has a nonzero measure. By relaxing the conditional dependence in this way, we may evaluate the expectations contained within these equations shown above using our decomposition-based uncertainty analysis algorithm. To construct these conditional sets of finite measures we partition the input space into a finite number of bins [103].

3.2.2 Algorithm

Here we present the algorithm for evaluating the main effect indices of a multicomponent system in a decomposition-based approach. Afterwards, we discuss the extension of this algorithm to evaluate interaction effect indices. The algorithm presented here

requires that the "offline phase" of the decomposition-based uncertainty analysis algorithm presented in Chapter 2 has been completed (i.e., that local uncertainty analyses have been performed for each component of the system).

To evaluate the main effect index for the i^{th} system input we first discretize the range of the system input into a set of finite partitions $\{[t^k, t^{k+1})\}_{k=1}^{K}$ where K is the total number of partitions and the partitions are not overlapping (i.e., $[t^k, t^{k+1}) \cap [t^{k+1}, t^{k+2}) = \varnothing$). Beginning with the first partition (i.e., $k = 1$), we restrict the input's target measure to be non-zero only within the said partition and zero elsewhere. The i^{th} system input's new target measure is then given by,

$$\widetilde{\nu}(A) = \begin{cases} \frac{\nu(A)}{\int_{\mathcal{P}_{i,k}} \mathrm{d}\nu} & \text{if } A \in \mathcal{P}_{i,k} \\ 0 & \text{otherwise} \end{cases}, \tag{3.18}$$

where $\mathcal{P}_{i,k} = [t^k, t^{k+1})$. By defining the input's new target measure according to Equation 3.18 we are approximating the conditional constraint contained in Equation 3.16. After defining the input's new target measure we can now formulate this problem as a decomposition-based uncertainty analysis. This allows us to evaluate the statistics of interest, here an expectation, while satisfying our conditional constraint. For example, the conditional expectation presented in Equation 3.16 is now expressed as,

$$\mathbb{E}_{\{\widetilde{Y}_i, Y_i^c\}}[f], \tag{3.19}$$

where $\widetilde{Y}_i$ is associated with the continuous measure $\widetilde{\nu}$ in Equation 3.18.

Note, to evaluate these statistics of interest we only require the "online phase" of the decomposition-based uncertainty analysis algorithm where the i^{th} input target measure is provided by Equation 3.18. We emphasize that this procedure, which utilizes the global compatibility satisfaction step in Chapter 2, requires no further evaluation of the components. An equivalent procedure to the one performed here, that is adjusting the system's input target distribution function and reevaluating the system uncertainty analysis, was previously demonstrated in the flexibility of our decomposition-based uncertainty analysis in Section 2.7. Next, this procedure of

restricting the input's target measure and performing a decomposition-based uncertainty analysis is repeated for each partition. Upon evaluating, for each partition, the expectation of the output of interest we may at last evaluate the variation over these K expectations. Numerically this discretization approach approximates Equation 3.16 by

$$\hat{S}_i = \frac{\sum_{k=1}^{K} P_{\mathbf{Y}}(Y_i \in \mathcal{P}_{i,k})(\mathbb{E}_{\{\tilde{Y}_i, Y_i^c\}}[f] - \hat{f}_{\{0\},i})^2}{var_{\mathbf{Y}}(f)}, \tag{3.20}$$

where $\hat{f}_{\{0\},i}$ is given by

$$\hat{f}_{\{0\},i} = \sum_{k=1}^{K} P_{\mathbf{Y}}(Y_i \in \mathcal{P}_{i,k})\mathbb{E}_{\{\tilde{Y}_i, Y_i^c\}}[f]. \tag{3.21}$$

Here the hat quantities denote an estimate of the corresponding true quantities. By the strong law of large number, as the number of partitions goes to infinity (i.e, $K \to \infty$) we obtain Equation 3.16. However, increasing K will require being able to evaluate the decomposition-based uncertainty analysis algorithm on smaller and smaller partitions. Therefore, in practice, K should be selected with care since the number of proposal samples within each partition decreases (i.e., poorer approximation of $\mathbb{E}_{\{\tilde{Y}_i, Y_i^c\}}[f]$) with increasing K. We demonstrate this issue through an example in Section 3.3.

To evaluate all the main effect indices of the system requires Kd_S decomposition-based uncertainty analysis evaluations where d_S is the total number of system input variables. The algorithm to evaluate Equation 3.20 is given in Algorithm 3.

Algorithm 3: Decomposition-based global sensitivity analysis.

Data: Number of partitions K and the decomposition-based uncertainty analysis results (i.e., proposal samples, target measures, and $var_{\mathbf{Y}}(f)$).

Result: Main effect indices $S_{\{i\}}$ for all $i \in \{1, \ldots, d_S\}$.

for $i = 1 : d_S$ **do**
  Determine sets $\mathcal{P}_{i,k} = [t^k, t^{k+1})$ for $k \in \{1, 2, \ldots, K\}$;
  for $k = 1 : K$ **do**
    Adjust the i^{th} system input target measure via Equation 3.18;
    Perform a *decomposition-based uncertainty analysis*;
    Evaluate and store the statistic of interest, $\mathbb{E}_{\{\widetilde{Y}_i, Y_i^c\}}[f]$;
  end
  Evaluate the mean and main effect index in Equation 3.20;
end

We may also quantify higher-order indices through a similar approach. For example, evaluating Equation 3.17 for any two distinct system inputs requires partitioning the input space into K^2 partitions, where for each partition, a decomposition-based uncertainty analysis is performed. This procedure needs to be repeated for all possible combinations of two distinct system inputs (i.e., $\binom{d_S}{2}$). Evaluating these interaction effect indices and higher-order indices is possible but computationally expensive.

3.3 Application to a gas turbine system

This section demonstrates the decomposition-based global sensitivity analysis approach on a gas turbine blade application. We compare the results from our method to the all-at-once Monte Carlo system global sensitivity analysis.

3.3.1 Gas turbine system setup

This application problem is a continuation of the gas turbine system presented in Section 2.7. The objective of our global sensitivity analysis is to quantify the main

effect indices of the system inputs with respect to the system output of interest, here the economics of the product. By quantifying the main effect indices we may guide the allocation of future research efforts aimed at reducing the system output of interest variability.

As was the case for the decomposition-based uncertainty analysis presented in Section 2.7, here for the "offline phase", each component individually performs a local uncertainty analysis using the input proposal distributions specified in Tables 2.3–2.6. Each component uses $n = 10,000$ independent proposal samples in its local Monte Carlo simulation. The number of proposal samples for this study was deemed adequate in Section 2.7 for the decomposition-based uncertainty analysis. Output samples for each component are stored in a database. For the "online phase", here Algorithm 3, we consider the uncertain system inputs provided in Table 2.2.

To evaluate the main effect index for the i^{th} system input we first partition the i^{th} system input range into K bins. For this study we used $K = 10$ partitions where each partition encloses an equal target probability (i.e., $\nu(\mathcal{P}_{:,i}) = \nu(\mathcal{P}_{:,j})$ for all $i, j \in \{1, \ldots, K\}$).

3.3.2 Global sensitivity analysis

The main effect indices of the gas turbine blade application are provided in Figure 3-1. Our results are compared to the all-at-once Monte Carlo system global sensitivity analysis, Equation 3.14, using $n = 100,000$ samples. The results provided in Figure 3-1 demonstrate that our decomposition-based global sensitivity analysis algorithm can accurately identify the most influential system input variables.

Table 3.1 presents the performance of the decomposition-based global sensitivity analysis algorithm with respect to the number of partitions and number of proposal samples. The performance is evaluated by,

$$\epsilon = \sum_{i=1}^{d_S} |S_{\{i\}} - \hat{S}_{\{i\}}|, \tag{3.22}$$

where $S_{\{i\}}$ is given by the all-at-once Monte Carlo global sensitivity analysis approach

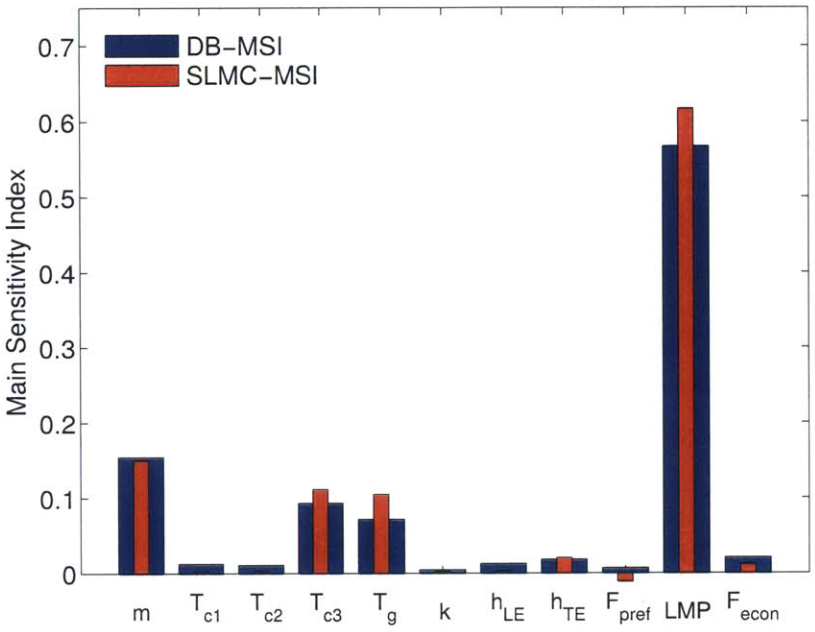

Figure 3-2: The decomposition-based main effect indices (DB-MSI) are plotted against the all-at-once Monte Carlo main effect indices (SLMC-MSI). The results demonstrate that the decomposition-based approach quantifies the main effect indices accurately with respect to the standard all-at-once Monte Carlo approach.

and $\hat{S}_{\{i\}}$ is given by our decomposition-based global sensitivity analysis approach. The results indicate that the accuracy of our estimate improves with increasing number of bins and increasing number of proposal samples as expected.

Figure 3-3 presents two forms of error which might arise when the decomposition-based global sensitivity analysis algorithm performs inadequately. On the left are the results using $K = 2$ partitions and $n = 10,000$ proposal samples. Here the summation of the main effect indices is greater than unity which indicates a failure in our analysis since the sum of all global sensitivity indices is unity and all variance-based sensitivity indices presented here are by definition nonnegative. On the right are the results using $K = 2$ partitions and $n = 5,000$ proposal samples. Here the decomposition-based sensitivity analysis incorrectly characterizes T_{c2} as an influential system input variable. Clearly, the latter of the two failures is more difficult to identify since it cannot be assessed using strict variance-based global sensitivity analysis conditions.

These two forms of failure arise due to two interacting effects. These effects in-

clude using a small number of partitions (i.e., small K) and having a small effective sample size per partition (i.e., small n). Using too few partitions results in inaccurately estimating Equation 3.16 using the approximate main effect index presented in Equation 3.20. However, having a small effective sample size results in inaccurately estimating the expectation in Equation 3.20 using the decomposition-based uncertainty analysis algorithm. These two issues are inversely related; a high number of partitions implies a low effective sample size per partition and vice versa. As demonstrated in Section 2.6, we can gradually increase the number of partitions while simultaneously using an a posterior indicator, such as the effective sample size, to provide confidence in our results.

Table 3.1: The performance of the decomposition-based global sensitivity analysis algorithm as quantified by Equation 3.22 is presented. The results suggest the decomposition-based global sensitivity analysis degrades with decreasing number of partitions and decreasing number of proposal samples.

	$K = 10$	$K = 5$	$K = 2$
$n = 10000$	0.1645	0.1921	0.4976
$n = 5000$	0.1807	0.2246	0.4165
$n = 2000$	0.3074	0.5001	0.5679

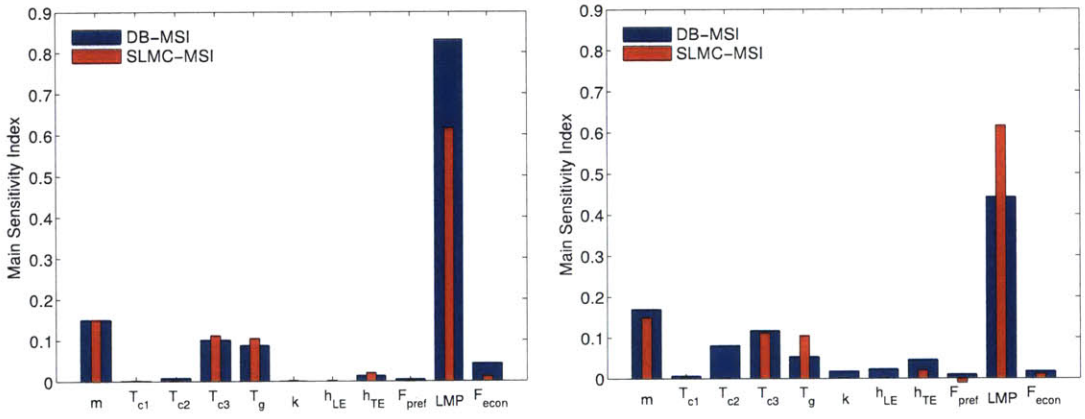

Figure 3-3: The decomposition-based main effect indices (DB-MSI) are plotted against the all-at-once Monte Carlo main sensitivity indices (SLMC-MSI). The plot on the (left) implements $K = 2$ partitions and $n = 10,000$ proposal samples. The plot on the (right) implements $K = 2$ partitions and $n = 5,000$ proposal samples. These results show two scenarios for which the decomposition-based global sensitivity analysis algorithm performs inadequately. On the (left) our approach suggests the sum of the main effect indices are greater than unity. On the (right) our approach erroneously suggests T_{c2} is an influential system input.

Chapter 4

Optimal L_2–norm empirical importance weights

To estimate statistics from a target distribution one may apply standard Monte Carlo simulation, using random samples that are generated from the target distribution [98]. However, if one only has available random samples generated from a proposal distribution, then evaluating statistics from a target distribution given random samples generated from a proposal distribution is acknowledged as the change of measure. This problem arises in a host of domains such as information divergence, particle filtering, and importance sampling as previously presented in Section 2.3 (see e.g., [116] for a fuller discussion of applications). If both proposal and target distributions are known and satisfy additional conditions, then the Radon-Nikodym Theorem provides a solution [14]. In this chapter we consider the scenario whereby the distributions from which the random samples are generated are unknown; that is, we have available the samples but no explicit description of their underlying distributions.

In Section 4.1, we present the current practices for the change of empirical measure and introduce our approach. Section 4.2 sets nomenclature, formalizes the objective of this chapter, and presents the proposed optimization formulation. In Section 4.3, we present the numerical formulation and examine properties of the optimization statement. In Section 4.4, we prove that the proposed approach achieves convergence in the L_1–norm for multidimensional distributions and weak convergence for

one-dimensional distributions. In Section 4.5, we examine the analytic solution to the optimization statement for the case of a one-dimensional distribution problem. Section 4.5 also presents a numerical solution to the optimization statement and discusses techniques that extend our approach to large-scale applications. In Section 4.6, we demonstrate the properties of the proposed approach on a one-dimensional distribution problem. Section 4.6 also compares our approach to previous approaches on an importance sampling problem over a range of parameters, evaluates the performance of the optimization algorithms, and examines the relationship between discrepancy theory and the proposed approach when the proposal and target are distributed according to the uniform distribution.

4.1 Introduction to empirical change of measure

In this section we set the notation for the subsequent developments, establish the problem statement, and review the current practices. The section concludes with a description of our proposed approach.

4.1.1 Definitions and Problem Setup

In our setting, we assume the proposal measure μ has finite support and is accessible to us only through sampling; that is, we are provided with random samples of the random variable $\mathbf{X}$ but we cannot evaluate $P_{\mathbf{X}}$ or $p_{\mathbf{X}}$ explicitly. Let $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$ be random samples of $\mathbf{X}$, where n is the number of random samples. The objective is to estimate statistics from the target measure ν given random samples $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$ generated from the proposal measure μ. The challenge with this objective, recognized as a change of measure, is that the proposal measure μ is accessible to us only through sampling.

Although the Radon-Nikodym Theorem is still valid (if the underlying distributions satisfy the appropriate conditions), it is indeterminable because we cannot compute the Radon-Nikodym derivative (i.e., the ratio of the target probability density function to the proposal probability density function), herein referred to as the

probability density ratio. To complete the change of measure, an importance weight proportional to the probability density ratio needs to evaluated and associated with each random sample. However, the importance weights cannot be computed directly in the usual way, since the probability density ratio is indeterminable. In this chapter we present an approach that overcomes this challenge by formulating and solving a scalable optimization problem to determine the importance weights. We first discuss several previously proposed solutions to this problem.

4.1.2 Current Practices

The previous approaches summarized here all assume that the random samples are generated from an unknown distribution (see e.g., [116] for a detailed discussion of these approaches). As a result, these approaches seek to estimate the probability density ratio from the random samples. The approach implemented in Section 2.4 is a commonly used approach which estimates the unknown probability density function from the random samples [106]. By estimating the proposal and target probability density functions, we can then estimate the probability density ratio. The solution to the change of measure problem then follows from the Radon-Nikodym Theorem along with the estimated probability density ratio. However, estimating the unknown probability density function from the random samples is difficult and is particularly challenging in cases of high dimension as pointed out in Section 2.8 [106, 53, 120]. In practice, this challenge can be overcome if the random samples are known to be generated from a parametric distribution family, in which case a parametric density estimation method can be employed.

As a result, other approaches have avoided estimating the probability density function and instead estimate directly the probability density ratio using the random samples. The kernel mean matching approach matches the moments using a universal reproducing kernel Hilbert function [58, 47]. The probabilistic classification approach computes the probability density ratio by applying Bayes' Theorem [93]. The importance estimation filtering approach minimizes the Kullback-Leibler divergence metric between the estimated and actual probability density ratios [116]. The unconstrained

least squares importance filtering approach minimizes the L_2–norm between the estimated and actual probability density ratios [60]. The direct density ratio estimation with dimension reduction solves the previous approach on a lower-dimensional space [117]. These approaches share in common multiple attributes. They each present a means of computing the probability density ratio using the random samples. They each represent the probability density ratio using a set of basis functions, thereby constraining the solution to exist within a specified basis representation. Finally, these approaches require tuning parameters, which one can choose using a variant of cross-validation.

4.1.3 Proposed Approach

Our approach avoids estimating or working with the unknown distribution function or its probability density function. Instead, we work with the well-defined and determinable *empirical distribution function* associated with the random samples. Specifically, our approach, illustrated in Figure 5-1, formulates and solves an optimization problem to determine a set of empirical importance weights that minimize the L_2–norm between the weighted proposal empirical distribution function and the target distribution function. In the example in Figure 5-1, the target is the beta distribution function, $\mathcal{B}(0.5, 0.5)$, and the proposal random samples are generated from the uniform distribution function, $\mathcal{U}(0, 1)$. The core idea of our approach is to compute importance weights associated with the proposal random samples that transform the weighted proposal empirical distribution function to the target distribution function. We also constrain the importance weights to define a probability measure. This requires that these importance weights are non-negative and that the empirical probability measure assigns a unit value to the entire probability space.

The approach proposed in this chapter shares resemblance to the recent constructive setting of the density ratio estimate [121]. That work minimizes the regularized L_2–norm between the weighted proposal empirical distribution function and the empirical target distribution function, where the importance weights are defined on a set of basis functions. Those importance weights are shown in [121] to converge in

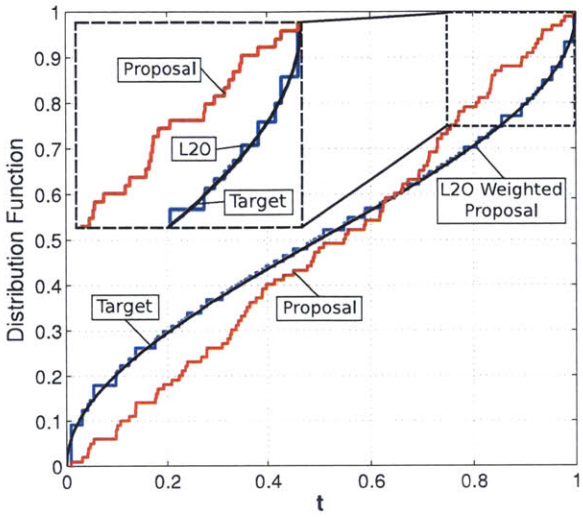

Figure 4-1: The proposed approach minimizes, with respect to empirical importance weights associated with the proposal random samples, the L_2–norm between the weighted proposal empirical distribution function and the target distribution function. In this example, we generated $n = 100$ random samples from the proposal uniform distribution function, $\mathcal{U}(0,1)$. The results show our weighted proposal empirical distribution function, labeled "L2O Weighted Proposal", accurately represents the target beta distribution function, $\mathcal{B}(0.5, 0.5)$.

probability to the Radon-Nikodym derivative, as the number of proposal and target random samples tend to infinity. Our approach does not use a basis function representation of the importance weights, since we are only interested in evaluating the importance weights at the random sample locations (i.e., we associate one weight with each random sample). We also do not include regularization, since this modifies the solution and introduces smoothness that may not be desirable. Instead, we rely on the optimization solvers to exploit the structure of the problem. Avoiding regularization allows us to avoid tuning parameters, yet our formulation maintains convergence in the L_1–norm of the weighted proposal empirical distribution function to the target distribution function, as the number of random samples tends to infinity. Moreover, our optimization approach can be implemented at large scale (both high dimensional distribution functions and a large number of random samples). Our approach has an analytic closed-form solution in the case of a one-dimensional distribution problem, which has a connection to the trapezoidal integration rule and achieves convergence

almost everywhere of the weighted proposal empirical distribution function to the target distribution function, as the number of random samples tends to infinity. Our approach is also shown to relate to discrepancy theory.

4.2 Optimization Statement

In this section we formulate our solution to the change of measure objective by casting the change of measure problem as an optimization statement. In introducing the optimization statement we assume that the target distribution function and target probability density function are known.

The typical approach to overcome the change of measure challenge, and the approach implemented in Section 2.4, is to apply density estimation to the random samples $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$, yielding an estimate of the proposal density $p_{\mathbf{X}}$. However, density estimation in high dimensions is notoriously difficult, and state-of-the-art approaches often perform poorly for high-dimensional problems. Therefore, we approach the change of measure challenge in a different way—using instead the well-defined proposal empirical distribution function,

$$P^n_{\mathbf{X}}(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\mathbf{x}^i \leq \mathbf{t}), \tag{4.1}$$

where $\mathbb{I}(\mathbf{x}^i \leq \mathbf{t})$ is the maximum convention Heavyside step function defined as

$$\mathbb{I}(\mathbf{x} \leq \mathbf{t}) = \begin{cases} 1, & \text{if } x_i \leq t_i,\ \forall\, i \in \{1, 2, \ldots, d\} \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

Here we have used the subscript and superscript notation for the empirical distribution function, $P^n_{\mathbf{X}}$, to identify the measure of the random samples from which it is built, μ, and the number of random samples, n. The strong law of large numbers (SLLN) states that the estimator $P^n_{\mathbf{X}}$ converges to the proposal distribution function $P_{\mathbf{X}}$ defined as

$$P_{\mathbf{X}}(\mathbf{t}) = \mu((-\infty, \mathbf{t}]), \tag{4.3}$$

as n tends to infinity almost everywhere (a.e.) for all continuity points $\mathbf{t}$ of $P_{\mathbf{X}}(\mathbf{t})$ [14].

To accomplish the change of measure objective, we propose to compute a set of importance weights, defined here as *empirical importance weights*, to transform the proposal empirical distribution function into the target distribution function. We introduce n empirical importance weights, denoted by the vector $\mathbf{w} = [w_1, w_2, \ldots, w_n]^\top$. Each empirical importance weight w_i is associated with a random sample $\mathbf{x}^i$. We use the notation

$$P^n_{\mathbf{X};\mathbf{w}}(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n} w_i \mathbb{I}(\mathbf{x}^i \leq \mathbf{t}) \tag{4.4}$$

to represent a weighted empirical distribution function that is composed of n random samples generated from the measure μ and weighted by $\mathbf{w}$. The empirical importance weights are dependent on the random samples, $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$; however, for simplicity we do not show the dependency in the notation.

We now cast the change of measure objective as an optimization statement. The objective is to minimize, with respect to the empirical importance weights, the distance between $P^n_{\mathbf{X};\mathbf{w}}$, defined in Equation (4.4), and the target distribution function, $P_{\mathbf{Y}}$. The criterion selected is the L_2–norm distance metric. Thus, the L_2–norm objective function is defined as

$$\omega^2(\mathbf{w}) = \frac{1}{2}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(P^n_{\mathbf{X};\mathbf{w}}(\mathbf{t}) - P_{\mathbf{Y}}(\mathbf{t})\right)^2 \mathrm{d}\mathbf{t}, \tag{4.5}$$

conditioned on the scaled empirical importance weights being a probability measure. That is, $\mathbf{w}$ satisfies the non-negativity box-constraint, $w_i \geq 0$, $\forall i \in \{1, 2, \ldots, n\}$, and the single equality constraint, $\mathbf{1}_n^\top \mathbf{w} = n$, where $\mathbf{1}_n \in \mathbb{R}^n$ is a vector with all entries equal to 1. The optimization statement that determines the empirical importance weights associated with the proposal random samples for the change of measure is

thus stated as follows:

$$\begin{aligned} \arg\min_{\mathbf{w}} \ & \omega^2(\mathbf{w}) \\ s.t. \ & w_i \geq 0, \ \forall i \in \{1, 2, \ldots, n\} \\ & \mathbf{1}_n^\top \mathbf{w} = n. \end{aligned} \tag{4.6}$$

In the above optimization statement, we have assumed that the target distribution $P_{\mathbf{Y}}$ is known explicitly. However, our approach can be applied to the case where the target measure is represented only through random samples of the random variable $\mathbf{Y}$. In that case, we replace $P_{\mathbf{Y}}$ in Equation (4.5) with the target empirical distribution function $P_{\mathbf{Y}}^m$, where m is the number of random samples of the random variable $\mathbf{Y}$. In the following development, we work mostly with the formulation defined in Equations (4.5) and (4.6); when applicable we introduce the target empirical distribution function into the optimization statement.

4.3 Numerical Formulation

This section describes how the optimization statement (4.6) can be formulated as a single linear equality and box-constrained quadratic program. Afterwards, we examine the properties of the optimization statement using the Karush Kuhn Tucker (KKT) conditions.

4.3.1 Single Linear Equality and Box-Constrained Quadratic Program

Upon substituting Equation (4.4) into Equation (4.5) and, without loss of generality, confining the support of μ to the unit hypercube, we obtain

$$\omega^2(\mathbf{w}) = \frac{1}{2} \int_0^1 \cdots \int_0^1 \left(\frac{1}{n} \sum_{i=1}^{n} w_i \mathbb{I}(\mathbf{x}^i \leq \mathbf{t}) - P_{\mathbf{Y}}(\mathbf{t}) \right)^2 \mathrm{d}\mathbf{t}. \tag{4.7}$$

This expression can be expanded as follows:

$$\omega^2(\mathbf{w}) = \frac{1}{2}\int_0^1 \cdots \int_0^1 \left[\left(\frac{1}{n}\sum_{i=1}^{n} w_i \mathbb{I}(\mathbf{x}^i \leq \mathbf{t})\right)^2 - \frac{2P_{\mathbf{Y}}(\mathbf{t})}{n}\sum_{i=1}^{n} w_i \mathbb{I}(\mathbf{x}^i \leq \mathbf{t}) + (P_{\mathbf{Y}}(\mathbf{t}))^2\right] \mathrm{d}\mathbf{t}. \tag{4.8}$$

The third term in the integrand of Equation (4.8) is independent of the optimization parameter and thus can be discarded from the optimization statement without affecting the optimal solution $\mathbf{w}$. We now examine the first term and second term individually and formulate their respective numerical representations.

The first term of the integrand in Equation (4.8) can be represented as

$$\begin{aligned}
&\int_0^1 \cdots \int_0^1 \left(\frac{1}{n}\sum_{i=1}^{n} w_i \mathbb{I}(\mathbf{x}^i \leq \mathbf{t})\right)^2 \mathrm{d}\mathbf{t} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \int_0^1 \cdots \int_0^1 \mathbb{I}(\mathbf{x}^i \leq \mathbf{t})\mathbb{I}(\mathbf{x}^j \leq \mathbf{t})\ \mathrm{d}\mathbf{t} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \prod_{k=1}^{d} \int_0^1 \mathbb{I}(x_k^i \leq t)\mathbb{I}(x_k^j \leq t)\ \mathrm{d}t_k \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \prod_{k=1}^{d} \int_{z_k^{i,j}}^1 \mathrm{d}t_k = \mathbf{w}^\top H \mathbf{w},
\end{aligned} \tag{4.9}$$

where $z_k^{i,j} = \max(x_k^i, x_k^j)$ and x_k^i is the k^{th} entry of random sample $\mathbf{x}^i$. Note that $H \in \mathbb{R}^{n\times n}$ is a reproducing kernel and by definition a positive definite matrix (see e.g., [89] for a review of this analysis). Additionally, the H matrix is the Hadamard product of d individual matrices. To obtain the Hadamard construction of H, we define the matrix corresponding to the single dimension k, H^k, where the $(i,j)^{th}$ entry of H^k is

$$H_{i,j}^k = \int_{z_k^{i,j}}^1 \mathrm{d}t_k, \tag{4.10}$$

and $k \in \{1, 2, \ldots, d\}$. Then the $(i,j)^{th}$ entry of H can be defined as

$$H_{i,j} = \frac{1}{n^2} \prod_{k=1}^{d} H_{i,j}^k, \tag{4.11}$$

which allows us to construct matrix H as

$$H = \frac{1}{n^2} \left(H^1 \circ H^2 \circ \cdots \circ H^d \right), \tag{4.12}$$

where "$\circ$" represents the Hadamard product.

The second term of the integrand in Equation (4.8) can be represented as

$$\begin{aligned} &\int_0^1 \cdots \int_0^1 \frac{2P_{\mathbf{Y}}(\mathbf{t})}{n} \sum_{i=1}^{n} w_i \mathbb{I}(\mathbf{x}^i \leq \mathbf{t}) \, \mathrm{d}\mathbf{t} \\ &= \frac{1}{n} \sum_{i=1}^{n} w_i \int_0^1 \cdots \int_0^1 P_{\mathbf{Y}}(\mathbf{t}) \mathbb{I}(\mathbf{x}^i \leq \mathbf{t}) \, \mathrm{d}\mathbf{t} \\ &= \frac{1}{n} \sum_{i=1}^{n} w_i \int_{x_1^i}^1 \cdots \int_{x_d^i}^1 P_{\mathbf{Y}}(\mathbf{t}) \, \mathrm{d}\mathbf{t}, \\ &= \mathbf{w}^\top \mathbf{b}, \end{aligned} \tag{4.13}$$

where the i^{th} entry of $\mathbf{b} \in \mathbb{R}^n$ is

$$b_i = \frac{1}{n} \int_{x_1^i}^1 \cdots \int_{x_d^i}^1 P_{\mathbf{Y}}(\mathbf{t}) \, \mathrm{d}\mathbf{t}. \tag{4.14}$$

If the target distribution function, $P_{\mathbf{Y}}$, is unknown and instead we have m random samples of the random variable $\mathbf{Y}$, $\{\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^m\}$, then the i^{th} entry of $\mathbf{b}$ is

$$b_i = \frac{1}{n} \int_{x_1^i}^1 \cdots \int_{x_d^i}^1 \frac{1}{m} \sum_{j=1}^{m} \mathbb{I}(\mathbf{y}^j \leq \mathbf{t}) \, \mathrm{d}\mathbf{t}. \tag{4.15}$$

Our modified optimization statement is now

$$\begin{aligned} \hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \ & \hat{\omega}^2(\mathbf{w}) \\ s.t. \ & w_i \geq 0, \ \forall \, i \in \{1, \ldots, n\} \\ & \mathbf{1}_n^\top \mathbf{w} = n, \end{aligned} \tag{4.16}$$

where

$$\hat{\omega}^2(\mathbf{w}) = \frac{1}{2}\big(\mathbf{w}^\top H \mathbf{w} - 2\mathbf{w}^\top \mathbf{b}\big). \tag{4.17}$$

Solving (4.16) yields the optimal empirical importance weights $\hat{\mathbf{w}}$ that minimize our original L_2-norm distance metric while satisfying the requirement of $\hat{\mathbf{w}}/n$ forming a probability measure.

4.3.2 Karush Kuhn Tucker Conditions

The Lagrangian of the optimization statement (4.16) is

$$\mathcal{L}(\mathbf{w}, \delta, \boldsymbol{\lambda}) = \frac{1}{2}\big(\mathbf{w}^\top H \mathbf{w} - 2\mathbf{w}^\top \mathbf{b}\big) + \delta(\mathbf{1}_n^\top \mathbf{w} - n) - \boldsymbol{\lambda}^\top \mathbf{w}, \tag{4.18}$$

where $\delta \in \mathbb{R}$ and $\boldsymbol{\lambda} \in \mathbb{R}^n$ are the equality and inequality constraint Lagrange multipliers, respectively. The optimal solution to (4.16) satisfies the following KKT conditions:

$$\begin{aligned} \frac{\partial \mathcal{L}(\hat{\mathbf{w}}, \delta, \boldsymbol{\lambda})}{\partial \mathbf{w}} = \ & \mathbf{0}_n = H\hat{\mathbf{w}} - \mathbf{b} + \delta \mathbf{1}_n - \boldsymbol{\lambda} \\ & \hat{w}_i \geq 0, \ \forall \, i \in \{1, 2, \ldots, n\} \\ & \lambda_i \geq 0, \ \forall \, i \in \{1, 2, \ldots, n\} \\ & \mathbf{1}_n^\top \mathbf{w} = n \\ & \delta \text{ is sign unrestricted} \\ & \lambda_i \hat{w}_i = 0 \ \ \forall i \in \{1, 2, \ldots, n\} \end{aligned} \tag{4.19}$$

where $\mathbf{1}_n \in \mathbb{R}^n$ and $\mathbf{0}_n \in \mathbb{R}^n$ are vectors with all entries equal to 1 and 0 respectively.

Since the optimization statement is a strictly convex quadratic program with linear constraints, one may show that the solution $\hat{\mathbf{w}}$ of (4.19) is the global solution to (4.16)

[18]. This implies that for all n the following inequality holds,

$$\int_A (P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{w}}}(\mathbf{t}))^2 \, \mathrm{d}\mathbf{t} \leq \int_A (P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\bar{\mathbf{w}}}(\mathbf{t}))^2 \, \mathrm{d}\mathbf{t}. \tag{4.20}$$

where $\bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_n]^\top$ is any set of importance weights that satisfies the constraints of the optimization statement (4.16).

The active set method is one numerical method that solves (4.16), and has been shown to converge and terminate in a finite number of steps [67]. This method employs an iterative approach that splits the solution space into an active set, $\mathcal{A} = \{i : w_i = 0\}$, and a passive set, $\mathcal{P} = \{i : w_i > 0\}$. The active and passive sets are updated iteratively until the KKT conditions are satisfied. At each iteration, the method solves an optimization problem for the passive set importance weights that has a closed-form solution. We use this closed-form solution to derive an analytic solution for the special case $d = 1$ (Section 4.5.1); however, our general numerical results employ optimization methods that are more amenable to large-scale problems, as described in Section 4.5.2. Before discussing the optimization solution strategies in detail, we first analyze the convergence properties of our approach.

4.4 Convergence

This section demonstrates that our approach, based on (4.16), exhibits convergence in the L_1–norm as the number of random samples tends to infinity. To demonstrate convergence in the L_1–norm we require the Radon-Nikodym derivative, which we recall in this section. The section concludes with the convergence theorem and proof.

The Radon-Nikodym Theorem states that

$$\nu(A) = \int_A h \, \mathrm{d}\mu \tag{4.21}$$

for any measurable subset $A \in \mathcal{F}$, where the measurable function $h : \mathbb{R}^d \to \mathbb{R}$ is called the Radon-Nikodym derivative and is defined by the probability density ratio, $h = p_{\mathbf{Y}}/p_{\mathbf{X}}$ [14]. In our problem setting, the Radon-Nikodym derivative exists but

is unknown. Let $\{h(\mathbf{x}^1), h(\mathbf{x}^2), \ldots, h(\mathbf{x}^n)\}$ be the Radon-Nikodym derivatives corresponding to proposal random samples $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$. To construct a probability measure, define the Radon-Nikodym importance weights as $\hat{h}(\mathbf{x}^i) = h(\mathbf{x}^i)/\bar{h}$ where $\bar{h} = \frac{1}{n}\sum_{i=1}^{n} h(\mathbf{x}^i)$. If weighted by $\hat{\mathbf{h}} = [\hat{h}(\mathbf{x}^1), \hat{h}(\mathbf{x}^2), \ldots, \hat{h}(\mathbf{x}^n)]^\top$, the Radon-Nikodym importance weighted empirical distribution function,

$$P^n_{\mathbf{X};\hat{\mathbf{h}}}(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n} \hat{h}(\mathbf{x}^i)\mathbb{I}(\mathbf{x}^i \leq \mathbf{t}), \tag{4.22}$$

converges to the distribution function $P_{\mathbf{Y}}$ by the SLLN as n tends to infinity almost everywhere for all continuity points $\mathbf{t}$ of $P_{\mathbf{Y}}(\mathbf{t})$.

We now present the convergence proof using our empirical importance weight vector $\hat{\mathbf{w}}$. We emphasize that Theorem 2 given below does not imply that the empirical importance weights converge pointwise to the Radon-Nikodym importance weights as the number of random samples tends to infinity. The proof establishes that the sequence of functions $\{P^1_{\mathbf{X};\hat{\mathbf{w}}}, P^2_{\mathbf{X};\hat{\mathbf{w}}}, \ldots\}$, defined by Equation (4.4), as the number of random samples tends to infinity converges to the target distribution function in the L_1–norm.

Theorem 2. *Let $P_{\mathbf{Y}}$ be the distribution function of ν and $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$ be random samples generated from the finite support probability measure μ where ν is absolutely continuous with respect to μ. Then there exists a set of empirical importance weights $\hat{\mathbf{w}} = [\hat{w}_1, \hat{w}_2, \cdots, \hat{w}_n]^\top$ satisfying (4.16) such that*

$$\lim_{n\to\infty} \int_A \left| P^n_{\mathbf{X};\hat{\mathbf{w}}}(\mathbf{t}) - P_{\mathbf{Y}}(\mathbf{t}) \right| \, d\mathbf{t} = 0. \tag{4.23}$$

where $A = \{\mathbf{t} \in \mathbb{R}^d \mid p_{\mathbf{X}}(\mathbf{t}) > 0\}$.

Proof. We begin by applying the Radon-Nikodym importance weights $\hat{\mathbf{h}}$, which satisfy the constraints in the optimization statement (4.16). As stated previously, by the SLLN we have

$$\lim_{n\to\infty} P^n_{\mathbf{X},\hat{\mathbf{h}}}(\mathbf{t}) \overset{a.e.}{=} P_{\mathbf{Y}}(\mathbf{t}), \tag{4.24}$$

for every continuity point $\mathbf{t}$ of $P_{\mathbf{Y}}(\mathbf{t})$. Since there exists an integrable function dom-

inating $P^n_{\mathbf{X},\hat{\mathbf{h}}}(\mathbf{t}) \leq 1$ for all $\mathbf{t} \in A$ and n, we can apply the dominated convergence theorem to obtain convergence in the L_1–norm:

$$\lim_{n\to\infty} \int_A \left| P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{h}}}(\mathbf{t}) \right| \, \mathrm{d}\mathbf{t} = 0. \tag{4.25}$$

Using the inequality $\left| P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{h}}}(\mathbf{t}) \right| \leq 1$, for all $\mathbf{t} \in A$ and all n, we obtain a bound on the L_2-norm,

$$\int_A (P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{h}}}(\mathbf{t}))^2 \, \mathrm{d}\mathbf{t} \leq \int_A 1 \cdot \left| P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{h}}}(\mathbf{t}) \right| \, \mathrm{d}\mathbf{t}. \tag{4.26}$$

Combining Equation (4.26) with Equation (4.25) we show convergence in the L_2–norm:

$$\lim_{n\to\infty} \int_A (P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{h}}}(\mathbf{t}))^2 \, \mathrm{d}\mathbf{t} = 0. \tag{4.27}$$

Since $\hat{\mathbf{h}}$ satisfies the constraints of the optimization statement (4.16), we use Equation (4.20) to show that

$$\int_A (P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{w}}}(\mathbf{t}))^2 \, \mathrm{d}\mathbf{t} \leq \int_A (P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{h}}}(\mathbf{t}))^2 \, \mathrm{d}\mathbf{t}. \tag{4.28}$$

This result coupled with Equation (4.27) states that convergence of $P^n_{\mathbf{X},\hat{\mathbf{h}}}$ to $P_{\mathbf{Y}}$ in the L_2–norm implies convergence of $P^n_{\mathbf{X},\hat{\mathbf{w}}}$ to $P_{\mathbf{Y}}$ in the L_2–norm,

$$\lim_{n\to\infty} \int_A (P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{w}}}(\mathbf{t}))^2 \, \mathrm{d}\mathbf{t} = 0. \tag{4.29}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} &\int_A \left| P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{w}}}(\mathbf{t}) \right| \, \mathrm{d}\mathbf{t} \\ &\leq \left(\int_A (P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{w}}}(\mathbf{t}))^2 \mathrm{d}\mathbf{t} \right)^{1/2} \cdot \left(\int_A (1)^2 \, \mathrm{d}\mathbf{t} \right)^{1/2} \\ &\leq M \cdot \left(\int_A (P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{w}}}(\mathbf{t}))^2 \, \mathrm{d}\mathbf{t} \right)^{1/2}, \end{aligned} \tag{4.30}$$

where $M < \infty$. Coupling this statement with (4.29), we show convergence in the

L_1–norm,

$$\lim_{n\to\infty} \int_A \left| P_{\mathbf{Y}}(\mathbf{t}) - P^n_{\mathbf{X},\hat{\mathbf{w}}}(\mathbf{t}) \right| \, \mathrm{d}\mathbf{t} = 0. \tag{4.31}$$

□ □

For the one-dimensional case (i.e., $d = 1$), Equation 4.31, is the Kantorovich or L_1-Wasserstein distance metric [45]. Convergence in the L_1-Wasserstein distance metric under our stated assumption, that the set A is finitely supported, establishes weak convergence. Weak convergence, or convergence in distribution, for this work establishes that

$$\lim_{n\to\infty} P^n_{X;\hat{\mathbf{w}}}(t) \overset{a.e.}{=} P_Y(t), \tag{4.32}$$

for every continuity point $t \in A$ of $P_Y(t)$ where $A = \{t \in \mathbb{R} \mid p_X(t) > 0\}$.

In Theorem 2, we required that the proposal measure have finite support. This constraint was necessary in order to invoke the dominated convergence theorem, Equation 4.25, and the Cauchy-Schwarz inequality, Equation 4.30. However, this constraint can be an artifact of the theorem and future work may one day show similar convergence without constraining the proposal measure to have finite support. In practice and as demonstrated in Section 4.6, this constraint does not need to be strictly enforced since we work with a finite number of samples. Further, one may approximate unsupported distributions using their respective truncated distributions.

Corollary 3. *Let* $\{\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^m\}$ *be* m *random samples generated from the probability measure* ν *and* $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$ *be* n *random samples generated from the finite support probability measure* μ *where* ν *is absolutely continuous with respect to* μ*. Then there exists a set of empirical importance weights* $\hat{\mathbf{w}} = [\hat{w}_1, \hat{w}_2, \cdots, \hat{w}_n]^\top$ *satisfying (4.16) with vector* $\mathbf{b}$ *defined by Equation (4.15) such that*

$$\lim_{\min n,m\to\infty} \int_A \left| P^n_{\mathbf{X},\hat{\mathbf{w}}}(\mathbf{t}) - P^m_{\mathbf{Y},\hat{\mathbf{w}}}(\mathbf{t}) \right| \, d\mathbf{t} = 0. \tag{4.33}$$

where $A = \{\mathbf{t} \in \mathbb{R}^d \mid p_{\mathbf{X}}(\mathbf{t}) > 0\}$.

Proof. Application of SLLN and the dominated convergence theorem we can establish that the estimator $P^m_{\mathbf{Y}}$ converges in the L_1-norm to the target distribution function

$P_{\mathbf{Y}}$, noted as

$$\lim_{\min(m)\to\infty} \int_A |P_{\mathbf{Y}}(\mathbf{t}) - P^m_{\mathbf{Y}}(\mathbf{t})| \ \mathrm{d}\mathbf{t} = 0. \tag{4.34}$$

By Theorem 2 in combination with Equation (4.34) and the triangle inequality we define a bound on the quantity of interest and conclude the proof [100],

$$\begin{aligned} &\int_A \left|P^n_{\mathbf{X},\hat{\mathbf{w}}}(\mathbf{t}) - P^m_{\mathbf{Y},\hat{\mathbf{w}}}(\mathbf{t})\right| \ \mathrm{d}\mathbf{t} \\ &\leq \int_A \left|P^n_{\mathbf{X},\hat{\mathbf{w}}}(\mathbf{t}) - P_{\mathbf{Y}}(\mathbf{t})\right| \ \mathrm{d}\mathbf{t} + \int_A |P_{\mathbf{Y}}(\mathbf{t}) - P^m_{\mathbf{Y}}(\mathbf{t})| \ \mathrm{d}\mathbf{t}. \end{aligned} \tag{4.35}$$

□ □

4.5 Solving the Optimization Statement

In this section we examine the solution to the optimization statement (4.16). We begin by presenting the analytical solution to the optimization statement for $d = 1$ as this solution provides a better understanding of the optimization statement. The section concludes with the general solution to the optimization statement by numerical methods. Here we introduce methods that extend our approach to large-scale applications and demonstrate how to incorporate a target empirical distribution function.

4.5.1 Analytic Solution for $\mathbb{R}$

For the case when $d = 1$, we present the analytic solution to (4.16) and demonstrate that this solution satisfies the KKT conditions (4.19). Note that for this case the random variable is one-dimensional, but the dimension of the optimization problem is still n, the number of proposal random samples. Without loss of generality, let the random samples of $X : \Omega \to \mathbb{R}$, $\{x^1, x^2, \ldots, x^n\}$, be ordered such that $x^i <$

x^{i+1}, $\forall\, i \in \{1, 2, \ldots, n-1\}$. Using Equation (4.12), the matrix H is

$$H = \frac{1}{n^2} \begin{bmatrix} (1-x^1) & (1-x^2) & (1-x^3) & \ldots & (1-x^n) \\ (1-x^2) & (1-x^2) & (1-x^3) & \ldots & (1-x^n) \\ (1-x^3) & (1-x^3) & (1-x^3) & \ldots & (1-x^n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (1-x^n) & (1-x^n) & (1-x^n) & \ldots & (1-x^n) \end{bmatrix}. \tag{4.36}$$

Similarly, using Equation (4.14), the vector $\mathbf{b}$ is

$$\mathbf{b} = \frac{1}{n} \cdot \begin{bmatrix} \int_{x^1}^{1} P_Y(t)\ \mathrm{d}t \\ \int_{x^2}^{1} P_Y(t)\ \mathrm{d}t \\ \ldots \\ \int_{x^n}^{1} P_Y(t)\ \mathrm{d}t \end{bmatrix}. \tag{4.37}$$

Then the solution to (4.16) is

$$\boldsymbol{\lambda} = \mathbf{0}_n, \tag{4.38}$$

$$\delta = \frac{1}{n} \int_{x^n}^{1} P_Y(t)\ \mathrm{d}t + \frac{x^n - 1}{n}, \tag{4.39}$$

and

$$\hat{\mathbf{w}} = \begin{bmatrix} \frac{n}{(x^2-x^1)} \int_{x^1}^{x^2} P_Y(t)\ \mathrm{d}t & \\ \frac{n}{(x^3-x^2)} \int_{x^2}^{x^3} P_Y(t)\ \mathrm{d}t & -\sum_{i=1}^{1} w_i \\ \vdots & \\ \frac{n}{(x^n-x^{n-1})} \int_{x^{n-1}}^{x^n} P_Y(t)\ \mathrm{d}t & -\sum_{i=1}^{n-2} w_i \\ n & -\sum_{i=1}^{n-1} w_i \end{bmatrix}. \tag{4.40}$$

This solution can be derived using the active set method [67]; we omit the details of the derivation here for brevity, but show that this solution satisfies the KKT conditions (4.19).

First, it can be seen that the empirical importance weights (4.40) are by construction non-negative, since P_Y is a monotonically non-decreasing function. Thus, in this $d = 1$ case, the non-negativity constraints on the importance weights do not play a role in constraining the optimal solution and all the corresponding Lagrange multipliers are zero, $\lambda_i = 0,\ \forall\, i \in \{1, 2, \ldots, n\}$. This result means that the complementarity conditions are satisfied. Second, summing the terms in (4.40), it is easy to show that the equality constraint $\mathbf{1}_n^\top \hat{\mathbf{w}} = n$ is satisfied. Lastly, we show that $H\hat{\mathbf{w}} = \mathbf{b} - \delta \mathbf{1}_n$ holds for each row entry $j \in \{1, 2, \ldots, n\}$. That is, we show

$$\begin{aligned} &\frac{1 - x^j}{n^2} \sum_{i=1}^{j} \hat{w}^i + \frac{1}{n^2} \sum_{i=j+1}^{n-1} \hat{w}^i (1 - x^i) + \frac{\hat{w}^n (1 - x^n)}{n^2} \\ &\quad = \frac{1}{n} \int_{x^j}^{1} P_Y(t)\ \mathrm{d}t - \left(\frac{1}{n} \int_{x^n}^{1} P_Y(t)\ \mathrm{d}t + \frac{x^n - 1}{n} \right). \end{aligned} \tag{4.41}$$

By substituting the empirical importance weights (4.40) into the left-hand side of Equation (4.41) and simplifying, we obtain,

$$\begin{aligned} &\frac{1 - x^j}{n^2} \sum_{i=1}^{j} \hat{w}^i + \frac{1}{n^2} \sum_{i=j+1}^{n-1} \hat{w}^i (1 - x^i) + \frac{\hat{w}^n (1 - x^n)}{n^2} \\ &\quad = \frac{1}{n} \int_{x^j}^{x^n} P_Y(t)\ \mathrm{d}t + \frac{1 - x^n}{n}. \end{aligned} \tag{4.42}$$

We obtain Equation (4.41) upon adding and subtracting b_n to Equation (4.42),

$$\begin{aligned} &\frac{1}{n} \int_{x^j}^{x^n} P_Y(t)\ \mathrm{d}t + \frac{1}{n}(1 - x^n) + b_n - b_n \\ &\quad = \frac{1}{n} \int_{x^j}^{1} P_Y(t)\ \mathrm{d}t - \left(\frac{1}{n} \int_{x^n}^{1} P_Y(t)\ \mathrm{d}t + \frac{x^n - 1}{n} \right). \end{aligned} \tag{4.43}$$

Since the KKT conditions are satisfied, (4.38)–(4.40) represent the solution to the optimization problem (4.16) for $d = 1$.

If instead we are given a target empirical distribution function represented by m random samples $\{y^1, y^2, \ldots, y^m\}$ generated from ν, then the optimal solution remains the same, with P_Y in (4.39)–(4.40) replaced by P_Y^m.

We conclude this subsection with a demonstration of Theorem 2 using the empirical importance weights defined in (4.40). That is, we show that

$$\lim_{n\to\infty} P^n_{X;\hat{\mathbf{w}}}(t) \overset{a.e.}{=} P_Y(t), \tag{4.44}$$

for every continuity point $t \in A$ of $P_Y(t)$ where $A = \{t \in \mathbb{R} \mid p_X(t) > 0\}$ has finite support. Given ν is absolutely continuous with respect to μ, let $i = \{j \in \{1, 2, \ldots, n-1\} \mid \hat{t} \in [x^j, x^{j+1})\}$ where $\hat{t}$ is a continuity point of $P_Y(\hat{t})$. We expand $P^n_{X;\hat{\mathbf{w}}}(\hat{t})$ using the empirical importance weights from (4.40):

$$P^n_{X;\hat{\mathbf{w}}}(\hat{t}) = \frac{1}{n}\sum_{i=1}^{n} \hat{w}^i \mathbb{I}(x^i \leq \hat{t}) = \frac{1}{x^{i+1} - x^i}\int_{x^i}^{x^{i+1}} P_Y(t)\ \mathrm{d}t. \tag{4.45}$$

Given that P^n_Y is monotonically non-decreasing and using Equation (4.45), we obtain the following inequality:

$$P_Y(x^i) \leq P_Y(\hat{t}), P^n_{X;\hat{\mathbf{w}}}(\hat{t}) < P_Y(x^{i+1}). \tag{4.46}$$

Since the target distribution is continuous at $\hat{t}$, this ensures for every $\epsilon > 0$ there exists a $\delta > 0$ such that $|P_Y(x) - P_Y(\hat{t})| \leq \epsilon$ for all points $x \in A$ for which $|x - \hat{t}| \leq \delta$. Now, since ν is absolutely continuous with respect to μ, there exists a finite n which is sufficiently large that we can find an $i = \{j \in \{1, 2, \ldots, n-1\} \mid \hat{t} \in [x^j, x^{j+1}]\}$ that yields $|x^i - \hat{t}| \leq \delta$ and $|x^{i+1} - \hat{t}| \leq \delta$. This implies $|P_Y(x^i) - P_Y(\hat{t})| \leq \epsilon$ and $|P_Y(x^{i+1}) - P_Y(\hat{t})| \leq \epsilon$. Lastly, by Equation (4.46) and application of the triangle inequality, we obtain

$$\begin{aligned} &|P^n_{X;\hat{\mathbf{w}}}(\hat{t}) - P_Y(\hat{t})| \\ &\quad < |P_Y(x^i) - P_Y(x^{i+1})| \\ &\quad \leq |P_Y(x^i) - P_Y(\hat{t})| + |P_Y(\hat{t}) - P_Y(x^{i+1})| \\ &\quad \leq 2\epsilon, \end{aligned} \tag{4.47}$$

which yields the desired result for every continuity point $\hat{t} \in A$ of $P_Y(\hat{t})$ as n tends to

infinity.

4.5.2 Optimization Algorithm

Here we focus on the optimization statement for the general case when $d > 1$ and examine algorithms that extend our approach to large-scale applications (i.e., a large number proposal random samples, n). The challenge with solving the optimization statement (4.16) when $d > 1$ is that the matrix H is not analytically invertible as was the case for $d = 1$. As a result, we rely on a numerical optimization routine to solve (4.16).

The optimization statement in (4.16) is classified as a single linear equality and box-constrained quadratic program. A popular application which falls into this class of problems is the dual form of the nonlinear support vector machine optimization statement [120]. That application resulted in algorithms to extend single linear equality and box-constrained quadratic programs to large-scale applications [92, 31, 72, 131]. For this work we have selected two large-scale optimization algorithms that exploit our problem structure: the Frank-Wolfe algorithm [42] and the Dai-Fletcher algorithm [31].

The Frank-Wolfe algorithm is well-suited for solving (4.16) since the objective is a differentiable convex function and the constraints are a bounded convex set. The core idea behind the Frank-Wolfe algorithm is to approximate the objective with a linear function and then take a step in the descent direction. The Frank-Wolfe algorithm is particularly attractive because it has well-established convergence rates, low computational complexity, and can generate sparse solutions. The pseudo algorithm describing the Frank-Wolfe algorithm tailored to the optimization statement (4.16) is given in Algorithm 4. Note that the step length α can be chosen to be the deterministic value $2/(2+k)$, where k is the iteration number, or alternatively α can be chosen such that it minimizes the objective function of (4.16) at that particular iteration. The computational complexity of the Frank-Wolfe algorithm per iteration is low since it requires only a rank-one update to the gradient vector at each iteration. With the structure of our problem, this update can be computed very efficiently.

Algorithm 4: Frank-Wolfe Algorithm for solving (4.16).

Data: Random samples $\mathbf{x}$, vector $\mathbf{b}$, initial feasible solution $\mathbf{w}_0$, and termination criteria.

Result: Empirical importance weight vector $\hat{\mathbf{w}}$.

Initialization: $\hat{\mathbf{w}} = \mathbf{w}_0$

$\mathbf{a} = H\hat{\mathbf{w}}$,

$\mathbf{g} = \mathbf{a} - \mathbf{b}$,

for $k = 1, 2, \ldots$ **do**

- Steepest descent direction: $\ell = \arg\min_{i \in \{1,2,\ldots,n\}}(g_i)$,
- $\bar{w}_i = \begin{cases} 1, & \text{if } i = \ell \\ 0, & \text{otherwise} \end{cases}$
- Set $\hat{\mathbf{w}} = \hat{\mathbf{w}} + \alpha(\bar{\mathbf{w}} - \hat{\mathbf{w}})$, where $\alpha \in [0, 1]$,
- $\mathbf{a} = (1 - \alpha)\mathbf{a} + \alpha H_{(\cdot,\ell)}$,
- $\mathbf{g} = \mathbf{a} - \mathbf{b}$,

if *(termination criteria satisfied)* **then**

Exit

end

end

As a second option, we examine the Dai-Fletcher optimization algorithm. The general idea of the Dai-Fletcher algorithm is to construct the Lagrangian penalty function

$$L(\mathbf{w}; \delta) = \frac{1}{2}\left(\mathbf{w}^\top H\mathbf{w} - 2\mathbf{w}^\top \mathbf{b}\right) - \delta(\mathbf{1}_n^\top \hat{\mathbf{w}} - n), \tag{4.48}$$

where δ is the equality constraint Lagrangian multiplier. Then for any fixed δ, the box-constrained quadratic program [30],

$$\begin{aligned} \hat{\mathbf{w}}(\delta) = \arg\min_{\mathbf{w}} \; & L(\mathbf{w}; \delta) \\ s.t. \; & w_i \geq 0, \; \forall\, i \in \{1, 2, \ldots, n\}, \end{aligned} \tag{4.49}$$

is solved. Next, δ is adjusted in an outer secant-like method to solve the single

nonlinear equation,

$$r(\delta) = \mathbf{1}_n^\top \hat{\mathbf{w}}(\delta) - n = 0. \tag{4.50}$$

That is, Equation (4.50) enforces that the solution of (4.49) satisfies the equality constraint (i.e., exists in the feasible region). In summary, for each iteration of the Dai-Fletcher algorithm, a box-constrained projected gradient-based algorithm is used to compute a new solution for (4.49). This solution is projected into a feasible region using a secant projection approximation method, thereby satisfying Equation (4.50). A summary of the Dai-Fletcher algorithm is given in Algorithm 5.

Algorithm 5: Dai-Fletcher Algorithm for solving (4.16).

Data: Random samples $\mathbf{x}$, vector $\mathbf{b}$, initial solution $\mathbf{w}_0$, and termination criteria.

Result: Empirical importance weight vector $\hat{\mathbf{w}}$.

Initialization: $\hat{\mathbf{w}} = \mathbf{w}_0$

for $k = 1, 2, \ldots$ **do**

- Compute gradient of (4.48),
- Take a steepest descent step,
- Project into feasible region by (4.50),
- Possibly carry out a linc scarch,
- Calculate a Barzilai-Borwein step length,
- Update the line search control parameters,

if *(termination criteria satisfied)* **then**

- Exit

end

end

The termination criteria in Algorithm 4 or Algorithm 5 may incorporate a maximum number of iterations and a minimum tolerance associated with the gradient of the objective function in (4.16) or the Lagrangian penalty function, Equation (4.48), respectively. Although Algorithm 4 and Algorithm 5 may in some cases terminate

prior to locating the global optimal solution, by construction they generate a sequence of feasible iterates. In Section 4.6, we evaluate the performance of these two algorithms over a range of parameters. The remainder of this section discusses numerical techniques to extend our approach to large-scale applications and to incorporate the target empirical distribution function.

The largest computational expense in Algorithm 5 is in the calculation of the matrix-vector product, $H\mathbf{w}$. The matrix-vector product, $H\mathbf{w}$, is also required in Algorithm 4, but since it only needs to be evaluated once, it has less impact on the computational performance of Algorithm 4. In the circumstance where the matrix H is small, the matrix can be assembled and stored for computations; however, large-scale applications (many samples) may prohibit assembly of the matrix H. In these cases, one option is to use the Frank-Wolfe algorithm and avoid repeated matrix-vector products altogether. Since in some cases the Dai-Fletcher algorithm may yield improved convergence rates, another option is to exploit the structure in the problem to reduce the numerical complexity of the matrix-vector product calculations. In particular, we recognize that since active set empirical importance weights are zero, they do not contribute to the matrix-vector product. As a result, only the columns of matrix H associated with passive set empirical importance weights are required for the matrix-vector product calculation. Thus, the numerical complexity of the gradient evaluation is $\mathcal{O}(n|\mathcal{P}|d^2+n|\mathcal{P}|)$, where the first term captures the construction of matrix H, the second term captures the matrix-vector product, and $|\mathcal{P}|$ denotes the cardinality of the passive set. In addition, efficient algorithms which rely on the divide-and-conquer technique have been developed and applied successfully to Equation (4.9) [12, 54]. Lastly, one may take advantage of parallel routines to divide and conquer the matrix-vector product [132, 85].

Solving the optimization problem also requires evaluating the vector $\mathbf{b}$. Here we will describe two special circumstances for which the vector $\mathbf{b}$ can be directly evaluated: an independently distributed target distribution function and a target empirical distribution function. For an independently distributed target distribution function we can define the target measure ν as the product of d individual measures,

$\nu = \nu_1 \otimes \cdots \otimes \nu_d$, where ν_i is the i^{th} target measure on $\mathbb{R}$. Then the resulting target distribution function can be expanded using a product series as $P_Y(\mathbf{t}) = \prod_{k=1}^{d} P_{Y_k}(t_k)$. The vector $\mathbf{b}$, Equation (4.14), can then be evaluated as a Hadamard product over each dimension:

$$\mathbf{b} = \begin{bmatrix} \int_{x_1^1}^1 P_{Y_1}(t)\ \mathrm{d}t \\ \int_{x_1^2}^1 P_{Y_1}(t)\ \mathrm{d}t \\ \cdots \\ \int_{x_1^n}^1 P_{Y_1}(t)\ \mathrm{d}t \end{bmatrix} \circ \cdots \circ \begin{bmatrix} \int_{x_d^1}^1 P_{Y_d}(t)\ \mathrm{d}t \\ \int_{x_d^2}^1 P_{Y_d}(t)\ \mathrm{d}t \\ \cdots \\ \int_{x_d^n}^1 P_{Y_d}(t)\ \mathrm{d}t \end{bmatrix}. \tag{4.51}$$

If the target distribution function is unknown and is instead estimated by the target empirical distribution function given m random samples $\{\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^m\}$ generated from ν, then there also exists an approach to directly construct the vector $\mathbf{b}$. The approach requires expanding Equation (4.15) as follows:

$$\begin{aligned} \mathbf{b} &= \frac{1}{nm} \sum_{j=1}^{m} \int_0^1 \cdots \int_0^1 \mathbb{I}(\mathbf{x}^i \leq \mathbf{t}) \mathbb{I}(\mathbf{y}^j \leq \mathbf{t})\ \mathrm{d}\mathbf{t}, \\ &= \hat{H}\mathbf{v}, \end{aligned} \tag{4.52}$$

and noting the similarities with the matrix-vector product $H\mathbf{w}$. Here we define $\mathbf{v} \in \mathbb{R}^m$ as the importance weights of the target random samples (i.e., $v_i = 1,\ \forall i \in \{1, 2, \ldots, m\}$). Additionally, we define an entry of matrix $\hat{H} \in \mathbb{R}^{n \times m}$ as

$$\hat{H}_{i,j} = \frac{1}{nm} \prod_{k=1}^{d} \int_{\hat{z}_k^{i,j}}^1 \mathrm{d}t_k, \tag{4.53}$$

where $\hat{z}_k^{i,j} = max(x_k^i, y_k^j)$. The vector $\mathbf{b}$ is then computed by the matrix-vector product (4.52).

4.6 Applications

In this section we apply the proposed approach to a number of numerical experiments. In the first application, we demonstrate the properties of the proposed approach on a one-dimensional distribution problem. For the second application, we compare the proposed approach to previous approaches on an importance sampling problem over a range of parameters. The last application examines the relationship between discrepancy theory and the proposed approach when the proposal and target are distributed according to the uniform distribution. We also use this opportunity to evaluate the performance of the Frank-Wolfe algorithm and Dai-Fletcher algorithm.

4.6.1 One-Dimensional Numerical Example

This analysis revisits the problem presented in Figure 5-1. However, instead of using the analytic empirical importance weights (4.40), as was done in Figure 5-1, this example uses the Frank-Wolfe algorithm with a step length $\alpha = 2/(2+k)$ and premature termination to obtain sparse solutions (recall that the Frank-Wolfe algorithm updates only one weight at each iteration). To initialize the Frank-Wolfe algorithm (i.e., $\mathbf{w}_0$), we choose an empirical importance weight vector with entries equal to

$$w_{0,i} = \begin{cases} n, & \text{if } i = \ell \\ 0, & \text{otherwise} \end{cases}, \tag{4.54}$$

where $\ell \in \{1, 2, \ldots, n\}$ is selected uniformly at random. The results of this numerical experiment using $n = 100$ proposal random samples are presented in Figure 4-2. The top and bottom plots show the results after 25 and 100 iterations, respectively, of the Frank-Wolfe algorithm.

These results illustrate that the proposed approach produces accurate representations of the target distribution function. Since the support of the proposal distribution function is finite, we can guarantee weak convergence by Theorem 2 (i.e., L_1-Wasserstein distance metric); permitting the Frank-Wolfe algorithm to run for more iterations would recover the analytic empirical importance weights (4.40). However,

the sparse empirical importance weights, shown on the top plot of Figure 4-2, are already a good approximation and may be advantageous if one wishes to evaluate computationally expensive statistics with respect to a complex or unknown target distribution function. That is, with the proposed approach, we have demonstrated one can approximate a target distribution function using a small set of optimally weighted proposal random samples. These results also illustrate that the proposed approach naturally accounts for clustering of the proposal random samples and other deviations from the original proposal distribution function. In the next section we compare our approach to previous approaches over a range of multiple-dimensional distributions with the application of the target empirical distribution function.

4.6.2 Importance Sampling

Importance sampling is a commonly used technique for estimating statistics of a target distribution given random samples generated from a proposal distribution. In this importance sampling example, the target and proposal probability density functions are $p_{\mathbf{Y}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $p_{\mathbf{X}} \sim \mathcal{N}(\mathbf{1}/\sqrt{d}, \mathbf{I})$, respectively, where $\mathbf{I} \in \mathbb{R}^{d \times d}$ is the identity matrix. Since these measures have infinite support, although our approach is still applicable as demonstrated in this application, we cannot guarantee weak convergence. For this example we assume we do not known $p_{\mathbf{Y}}$ or $p_{\mathbf{X}}$; but, are provided with proposal random samples $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$ and target random samples $\{\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^m\}$. To compare our approach to other existing approaches, we assume the target distribution function is unknown and instead estimate it using the target empirical distribution function.

To assess the performance of the various approaches, we consider the statistic

$$E = \int_{\Omega} \mathbb{I}(\mathbf{t} \in A) p_{\mathbf{Y}}(\mathbf{t}) \, \mathrm{d}\mathbf{t}, \tag{4.55}$$

where $A = \{\mathbf{t} \in \mathbb{R}^d \mid \langle \mathbf{t}, \mathbf{1} \rangle \geq \sqrt{d}\}$ is the event of interest. Regardless of dimension, the solution to Equation (4.55) is given by $E = \Phi(-1)$ where Φ is the standard normal cumulative distribution function. We will estimate Equation (4.55) using the

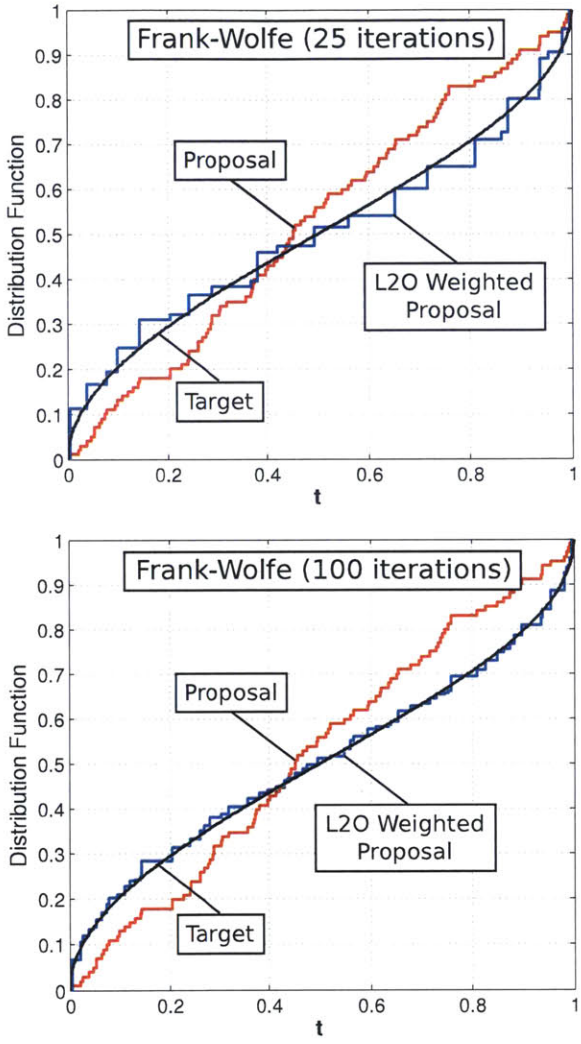

Figure 4-2: Performance of our optimal empirical importance weights determined using the Frank-Wolfe algorithm with step length $\alpha = 2/(2+k)$ and premature termination. This example uses $n = 100$ proposal random samples generated from a uniform distribution function, $\mathcal{U}(0,1)$. The target distribution function is the beta distribution function, $\mathcal{B}(0.5, 0.5)$. Terminating the Frank-Wolfe algorithm after 25 iterations (top) results in a sparse empirical importance weight vector. Terminating the Frank-Wolfe algorithm after 100 iterations (bottom) results in a dense solution and as a result a more accurate representation of the target distribution function.

following weighted Monte Carlo integration rule:

$$E_n = \sum_{i=1}^{n} \hat{w}^i \mathbb{I}(\mathbf{x}^i \in A), \tag{4.56}$$

where $\hat{\mathbf{w}}$ is obtained using one of the approaches described below. Note that although we may estimate Equation (4.55) using the available target random samples, our goal here is to assess the estimate E_n obtained using the weighted proposal samples. Thus, the available target random samples are used only in determining the empirical importance weights, but do not contribute to our estimate of E.

Our numerical experiment compares the following approaches:

1. **L_2–norm Optimal Weight (L2O)**: Our optimal empirical importance weights are obtained by solving the optimization statement developed in this chapter (4.16). For $d = 1$, we use analytic empirical importance weights (4.40) where $P_{\mathbf{Y}}$ is replaced by the target empirical distribution function $P_{\mathbf{Y}}^m$. For $d > 1$, we use the Dai-Fletcher algorithm and terminate after $2\max(n, m)$ iterations where n and m are the number of proposal and target random samples, respectively. For the implementation of the Dai-Fletcher algorithm, we compute the matrix H once for each case considered and store it for use at each optimization iteration.

2. **Kernel Density Estimation (KDE)**: The kernel density estimation [106] approach is applied to approximate $p_{\mathbf{Y}}$ and $p_{\mathbf{X}}$, denoted by $\tilde{p}_{\mathbf{Y}}$ and $\tilde{p}_{\mathbf{X}}$, from their respective random samples. We compute the Radon-Nikodym importance weights by approximating the Radon-Nikodym derivative with $\tilde{p}_{\mathbf{Y}}/\tilde{p}_{\mathbf{X}}$. The KDE uses Gaussian kernels where the kernel bandwidth is selected using the minimal mean squared error.

3. **Ratio Fitting (uLS)**: The unconstrained Least Squares Importance Fitting [60] approach is applied to approximate $h = p_{\mathbf{Y}}/p_{\mathbf{X}}$. Here h is represented by

the linear model,

$$\tilde{h}(\mathbf{t}) = \sum_{i=1}^{b} \hat{\beta}_i \phi_i(\mathbf{t}), \tag{4.57}$$

where b is the number of basis functions, $\{\phi_i\}_{i=1}^{b}$ are the basis functions, and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_b)^\top$ are the parameters to be learned. The parameters are obtained by solving the following optimization statement,

$$\begin{aligned} \hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \ & \frac{1}{2} \int \left(\tilde{h}(\mathbf{t}) - h(\mathbf{t})\right)^2 p_{\mathbf{Y}}(\mathbf{t}) \, \mathrm{d}\mathbf{t} + \gamma \boldsymbol{\beta}^\top \mathbf{1} \\ s.t. \ & \beta_i \geq 0, \ \forall \, i \in \{1, \ldots, b\}, \end{aligned}$$

where γ is the regularization parameter. The basis functions are Gaussian kernel models centered at the target random samples. The Gaussian kernel variance and regularization parameter are selected based on a 5-fold cross validation. Note that although the unknown Radon-Nikodym derivative appears in the objective, it is not explicitly evaluated.

4. **Divergence Fitting (KLD)**: The Kullback-Liebler (divergence) importance estimation [116] approach applies the linear model in Equation (4.57). The parameters are obtained by solving the following optimization statement,

$$\begin{aligned} \hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \ & \int p_{\mathbf{Y}}(\mathbf{t}) log\left(\frac{h(\mathbf{t})}{\tilde{h}(\mathbf{t})}\right) \mathrm{d}\mathbf{t} + \lambda \boldsymbol{\beta}^\top \mathbf{1} \\ s.t. \ & \sum_{i=1}^{n} \sum_{j=1}^{b} \beta_j \phi(\mathbf{x}^i) = n \\ s.t. \ & \beta_i \geq 0, \ \forall \, i \in \{1, \ldots, b\}, \end{aligned}$$

where the equality constraint ensures that $\tilde{h}$ defines a probability density function. The basis functions are Gaussian kernel models centered at the target random samples. The Gaussian kernel variance and regularization parameter are selected based on a 5-fold cross validation. Note that although the unknown Radon-Nikodym derivative appears in the objective, it is not explicitly evalu-

ated.

The four approaches presented above are tested on the following four scenarios:

1. $n = 2^{10}, m = 2^{10}$ and $d = \{1, 2, 5, 10\}$,
2. $n = 2^{10}, m = 2^{12}$ and $d = \{1, 2, 5, 10\}$,
3. $n = 2^{12}, m = 2^{10}$ and $d = \{1, 2, 5, 10\}$,
4. $n = 2^{12}, m = 2^{12}$ and $d = \{1, 2, 5, 10\}$.

For all scenarios, the results are the average over 100 independent trials and the quality of the results is quantified by

$$r_n = \frac{|E_n - E|}{E}. \tag{4.58}$$

Table 4.1 presents the results for each scenario. The quality of the result, Equation (4.58), contains two sources of error. The first source of error stems from how well the target empirical distribution function approximates the target distribution function. The second source of error stems from how well the proposal empirical distribution function, when weighted by the importance weights discussed above, approximates the target empirical distribution function.

The results presented in Table 4.1 allow us to differentiate the impact of these sources of error on our proposed approach. Comparing the results in scenario $\{n = 2^{10}, m = 2^{10}\}$ to scenario $\{n = 2^{10}, m = 2^{12}\}$ we observe that increasing the number of target random samples improves the accuracy of the estimate for $d \leq 5$. For $d = 10$, the results do not improve, which indicates that we do not have enough proposal random samples to accurately represent the target empirical distribution function. This conclusion is confirmed by comparing scenario $\{n = 2^{10}, m = 2^{12}\}$ to scenario $\{n = 2^{12}, m = 2^{12}\}$. By increasing the number of proposal random samples we improve the accuracy for $d = 10$. Comparing the results in scenario $\{n = 2^{10}, m = 2^{10}\}$ to scenario $\{n = 2^{12}, m = 2^{10}\}$ we observe an improvement for $d = 10$ but similar

Table 4.1: The error metric r_n, Equation (4.58), measured as a percentage, for the four methods and all four scenarios. Results are averaged over 100 independent trials and the term in parentheses is the corresponding standard deviation. Bold text indicates the best estimate for all the methods not including importance sampling (IS). The importance sampling results use the unknown proposal and target probability density functions to define the density ratio. The importance sampling results are provided here in order to compare to the standard solution using the unknown probability density functions. The results demonstrate that the proposed approach (L2O) outperforms the previous approaches. The proposed approach degrades with increasing dimensions and decreasing number of proposal and target random samples, however, less so than the other approaches.

	$\{$Low: $n = 2^{10}$, Low: $m = 2^{10}\}$			
	$d = 1$	$d = 2$	$d = 5$	$d = 10$
IS	*4.79(3.56)*	*5.45(4.32)*	*5.27(3.66)*	*4.34(3.22)*
L2O	**4.77(3.52)**	**5.51(3.96)**	**7.97(6.44)**	**37.7(11.0)**
KDE	5.72(3.82)	17.1(7.83)	48.9(11.9)	84.1(13.5)
uLS	7.92(7.14)	14.3(9.88)	31.5(9.94)	173(24.4)
KLD	19.2(8.33)	10.6(6.82)	41.8(21.3)	111(14.1)

	$\{$Low: $n = 2^{10}$, High: $m = 2^{12}\}$			
	$d = 1$	$d = 2$	$d = 5$	$d = 10$
IS	*4.79(3.56)*	*5.45(4.32)*	*5.27(3.66)*	*4.34(3.22)*
L2O	**3.27(2.31)**	**3.41(2.23)**	**6.84(4.79)**	**35.8(9.48)**
KDE	4.79(3.19)	14.7(5.13)	47.2(10.4)	74.3(12.0)
uLS	5.93(6.18)	16.3(10.9)	29.4(10.1)	175(11.2)
KLD	18.0(9.68)	12.0(11.9)	47.0(23.0)	112(13.2)

	$\{$High: $n = 2^{12}$, Low: $m = 2^{10}\}$			
	$d = 1$	$d = 2$	$d = 5$	$d = 10$
IS	*2.60(1.75)*	*2.84(2.02)*	*2.61(1.74)*	*2.46(1.84)*
L2O	**5.40(4.11)**	**6.47(4.82)**	**7.82(5.82)**	**7.78(5.05)**
KDE	5.83(4.46)	11.7(6.75)	48.0(9.35)	83.1(8.85)
uLS	10.4(7.85)	11.2(8.37)	33.6(9.66)	176(5.35)
KLD	18.8(9.55)	10.4(7.31)	43.7(24.5)	112(9.08)

	$\{$High: $n = 2^{12}$, High: $m = 2^{12}\}$			
	$d = 1$	$d = 2$	$d = 5$	$d = 10$
IS	*2.60(1.75)*	*2.84(2.02)*	*2.61(1.74)*	*2.46(1.84)*
L2O	**2.81(2.23)**	**3.12(2.05)**	**6.67(4.52)**	**6.72(4.88)**
KDE	2.98(2.51)	10.0(4.16)	45.0(5.93)	73.6(6.51)
uLS	6.25(5.77)	8.20(6.83)	32.7(9.25)	176(5.14)
KLD	19.0(9.51)	11.9(10.9)	45.9(21.7)	114(10.2)

results for $d \leq 5$. This indicates the proposal can accurately represent the target empirical distribution function; however, the target empirical distribution function may not accurately represent the actual target distribution function. Comparing the results in scenario $\{n = 2^{12}, m = 2^{10}\}$ to scenario $\{n = 2^{12}, m = 2^{12}\}$ we observe that increasing the number of target random samples produces a reduction in the standard deviation. This is because increasing the number of target random samples improved the target empirical distribution function estimate of the target distribution function and as a result improved the accuracy of the quantity of interest. We also note that the dependencies between the error metric defined in Equation (4.58) and the number of proposal and target random samples are not as obvious for the other three approaches.

The results presented here have also demonstrated that for the proposed method to perform the change of measure as expected, it is important that the target empirical distribution function accurately represents the target distribution function. However, selecting an adequate number of proposal and target random samples to obtain satisfactory results will require quantifying the rate of convergence for our proposed method. Although not performed in this thesis, guidance for quantifying the rate of convergence can come from inequalities such as the multivariate Dvoretzky-Kiefer-Wolfowitz inequality [124].

4.6.3 Uniform Distribution and the L_2–norm Discrepancy

In this example we present the relationship between our proposed approach and discrepancy theory [33]. To illustrate this relationship, the proposal and target distributions are the uniform distribution on the unit hypercube. We also take this opportunity to evaluate the performance of the Frank-Wolfe algorithm and Dai-Fletcher algorithm over a range of parameters.

Substituting the uniform distribution function, $P_{\mathbf{Y}}(\mathbf{t}) = \prod_{i=1}^{d} t_i$, for the target

distribution function in Equation (4.8), we obtain

$$\begin{aligned}\tilde{\omega}^2(\hat{\mathbf{w}}) =& \frac{1}{2}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{w}_i\hat{w}_j\prod_{k=1}^{d}\left(1-\max(x_k^i,x_k^j)\right)\right.\\ &\left.-\frac{2}{n}\sum_{i=1}^{n}\hat{w}_i\prod_{k=1}^{d}\frac{1-(x_k^i)^2}{2}+\frac{1}{3^d}\right),\end{aligned} \tag{4.59}$$

where we use $\tilde{\omega}$ to denote our L_2–norm distance metric in the special case of a uniform target distribution. If the proposal random samples are uniformly weighted (i.e., $\hat{w}_i = 1, \forall i \in \{1, 2, \ldots, n\}$), then Equation (4.59) relates directly to the L_2–norm discrepancy. The L_2–norm discrepancy is defined as

$$D_2 = \sqrt{2}\tilde{\omega}(\mathbf{1}_n), \tag{4.60}$$

and is sometimes referred to as Warnock's formula [79, 123].

In the following numerical study, we compare the ratio between the weighted L_2–norm discrepancy that results from using (4.59) with our optimal empirical importance weights and Warnock's formula (4.60),

$$r = \frac{\sqrt{2}\tilde{\omega}(\hat{\mathbf{w}})}{D_2} = \frac{\tilde{\omega}(\hat{\mathbf{w}})}{\tilde{\omega}(\mathbf{1}_n)}. \tag{4.61}$$

We investigate two scenarios: proposal random samples drawn from a pseudo-random (PR) sequence and from a randomized Sobol' low discrepancy (i.e., quasi-random, QR) sequence [86]. A pseudo-random number generator combines randomness from various low-entropy input streams to generate a sequence of outputs that are in practice statistically indistinguishable from a truly random sequence, whereas a quasi-random number generator constructs a sequence of outputs deterministically such that the output produces a small discrepancy.

For the case $d = 1$, the analytic empirical importance weights (4.40) are

$$\hat{\mathbf{w}} = \frac{1}{2} \begin{bmatrix} x^2 + x^1 \\ x^3 - x^1 \\ \cdots \\ x^n - x^{n-2} \\ 2 - x^n - x^{n-1} \end{bmatrix}. \tag{4.62}$$

Table 4.2 presents the results for the $d = 1$ case. Shown are the ratios r (in percentages), averaged over 100 independent trials. The results illustrate that the optimal empirical importance weights consistently reduce the L_2–norm discrepancy with respect to the uniformly weighted proposal random samples (i.e., $r < 1$). The reduction is more pronounced for the pseudo-random samples than the quasi-random samples. This is expected because quasi-random samples are constructed to reduce the discrepancy among the samples.

Table 4.2: The ratio of discrepancy computed using our optimal empirical importance weights and uniform importance weights, Equation (4.61) measured as a percentage. Shown are results for the $d = 1$ case, averaged over 100 independent trials. The term in parentheses is the corresponding standard deviation. n is the number of proposal random samples.

	$n = 2^8$	$n = 2^{10}$	$n = 2^{12}$
PR	12.2(4.80)	6.96(2.45)	3.38(1.17)
QR	86.4(6.48)	86.7(6.10)	85.9(6.86)

Since we have available the analytic representation of the empirical importance weights (4.62), we can also see that the resulting weighted Monte Carlo integration rule for an integrable function g is

$$\begin{aligned}\int_0^1 g(t)\,\mathrm{d}t &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\hat{w}_i g(x^i)\\ &= \lim_{n\to\infty}\frac{1}{2}\Bigg((x^2+x^1)g(x^1)+\sum_{i=1}^{n}(x^{i+1}-x^{i-1})g(x^i)\\ &\qquad +(2-x^n-x^{n-1})g(x^n)\Bigg),\end{aligned} \tag{4.63}$$

which was previously shown to be the trapezoidal integration rule [128].

For the general case $d > 1$, the empirical importance weights are computed using the Frank-Wolfe algorithm with an optimal step length α, and the Dai-Fletcher algorithm. For all simulations presented the Dai-Fletcher algorithm computes the matrix H once and stores it. The Frank-Wolfe algorithm using a deterministic step length α halves the computational time compared to using an optimal step length, but leads to poor results early in the optimization process. We selected a maximum number of iterations as the termination criterion for both algorithms. The maximum number of iterations were selected such that both algorithms have similar computational run times.[1] The purpose of this study is to evaluate our proposed approach and to compare the computational performance of the Frank-Wolfe algorithm to the Dai-Fletcher algorithm over a range of parameters. These parameters include the number of proposal random samples n, the initial solution $\mathbf{w}_0$, and dimension d. The initial solution for all simulations is uniform importance weights (i.e., $\mathbf{w}_0 = \mathbf{1}_n$). Figures 4-3, 4-4, and 4-5 show the results averaged over 100 independent trials for $d = 2$, $d = 5$, and $d = 10$, respectively.

As was the case for $d = 1$, these results illustrate that the optimal empirical importance weights consistently reduce the L_2–norm discrepancy with respect to uniformly weighted proposal random samples. Again, the reduction is more pronounced for the

[1] Intel® Xeon® E5410 (2.33GHz) processor & 6 GB RAM

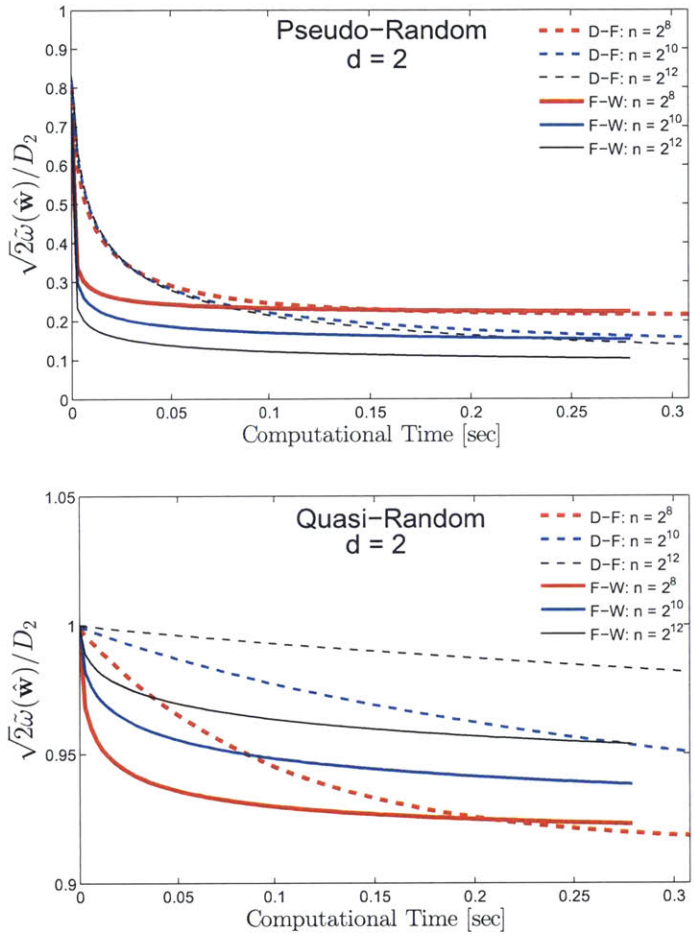

Figure 4-3: Discrepancy reduction for $d = 2$. Both algorithms reduce the L_2–norm discrepancy (i.e., $r < 1$) in both scenarios. The Frank-Wolfe algorithm converges more quickly than the Dai-Fletcher algorithm.

pseudo-random samples than the quasi-random samples. In general, if the proposal random samples are drawn from a pseudo-random sequence, then increasing n leads to further decrease in the discrepancy (r decreases further); however, if the proposal random samples are drawn from a quasi-random sequence, then increasing n leads to less discrepancy reduction (r shows less decrease). This can be explained since the pseudo-random proposal samples have poor (high) initial discrepancy and including more proposal samples gives our approach more degrees of freedom over which to optimize. Conversely, the quasi-random proposal samples already have low discrepancy; including more samples in this case makes it more difficult for the optimization to

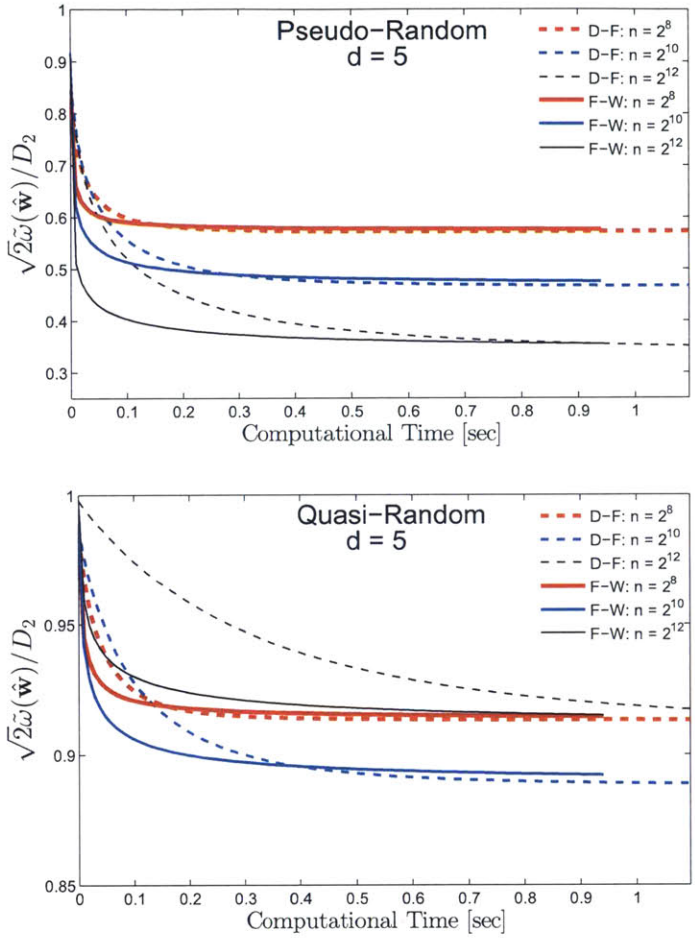

Figure 4-4: Discrepancy reduction for $d = 5$. Both algorithms reduce the L_2–norm discrepancy (i.e., $r < 1$) in both scenarios. The Frank-Wolfe algorithm converges more quickly than the Dai-Fletcher algorithm, although the final results are similar.

find a lower-discrepancy solution.

The results generally show that the Frank-Wolfe algorithm converges more quickly for cases using pseudo-random samples, while the Dai-Fletcher exhibits better performance for quasi-random samples. This suggests that the Frank-Wolfe algorithm may be preferred when the initial proposal empirical distribution function is far from the target distribution function, while the Dai-Fletcher algorithm is a better choice when the initial empirical importance weights are already close to optimal. Examining the results with increasing dimension d (i.e., increasing condition number of matrix H [122]), illustrates that both algorithms require more computational time to converge.

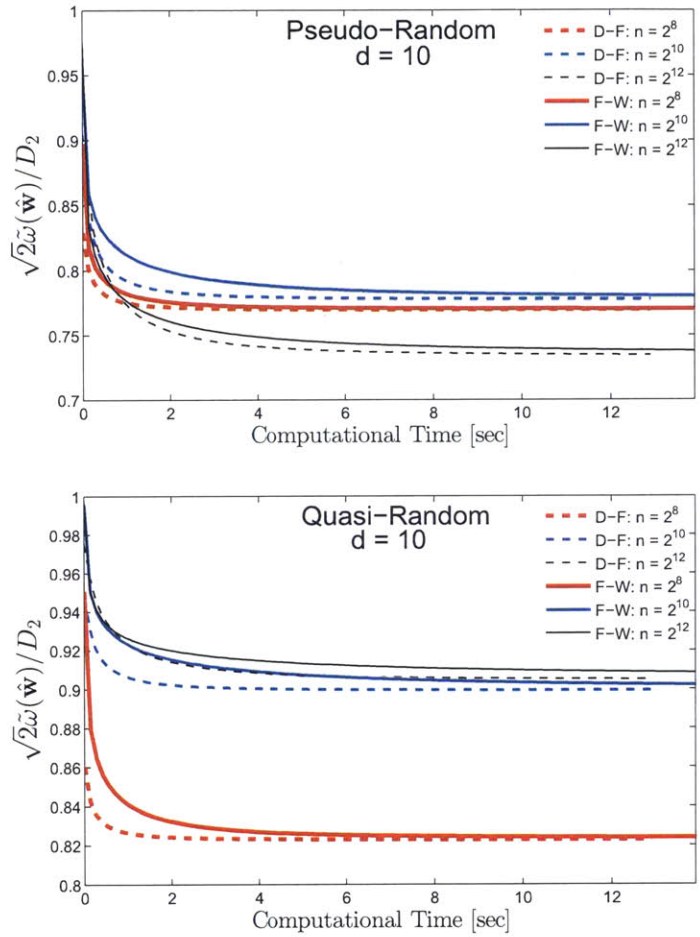

Figure 4-5: Discrepancy reduction for $d = 10$. Both algorithms reduce the L_2–norm discrepancy (i.e., $r < 1$) in both scenarios. The Dai-Fletcher algorithm converges more quickly the Frank-Wolfe algorithm, although the final results are similar.

This is expected since both algorithms implement gradient descent techniques whose convergence rates are expected to depend on the condition number of H.

The results presented in Figure 4-6 demonstrate our approach on a large-scale application problem. In this example we extended the results presented in Figure 4-4 using the Frank-Wolfe algorithm to proposal sample sizes $n = [8192, 32768, 131072]$. The computational times presented do not include the time required to evaluate the initial gradient (i.e., initial matrix-vector product; $\mathbf{a} = H\hat{\mathbf{w}}$). The results suggest our approach scales well with large number of samples. Numerical strategies such as divide-and-conquering methods and parallelization can be implemented to further

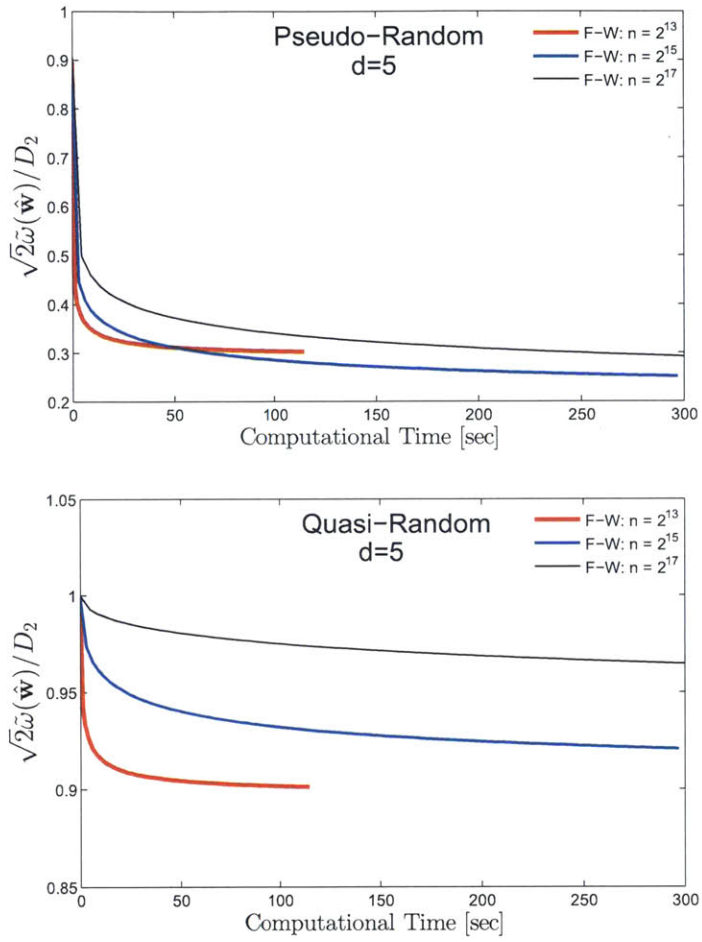

Figure 4-6: Discrepancy reduction for $d = 5$ and a large number of samples. The Frank-Wolfe algorithm reduces the L_2–norm discrepancy (i.e., $r < 1$) in both scenarios for a large-scale application problem (i.e., large n). The results presented are the average over 100 simulations.

improve the computational run times.

From these results, we recommend using the Frank-Wolfe algorithm when the dimension d is small or when the initial proposal empirical distribution function is far from the target distribution function. Otherwise, we recommend the Dai-Fletcher algorithm if the dimension d is large or if the initial proposal empirical distribution function is close to the target distribution function. If the number of proposal samples n is so large such that the matrix H cannot be stored, then we recommend using the Frank-Wolfe algorithm since the Dai-Fletcher algorithm will require constructing the

matrix H on the fly at each iteration, which will drastically increase computational time.

Chapter 5

Environmental Impacts of Aviation

The Federal Aviation Administration (FAA) is responsible for setting national aviation strategies and policies, including but not limited to aviation regulations and environmental policies. To support the FAA in effective decision- and policy-making we must evaluate an aircraft's fuel consumption, emissions, and performance. However, these quantities of interest are inherently uncertain since they are largely dependent on other uncertain variables such as aviation technologies and operations. Therefore, it is important that these quantities of interest come with an assessment of the associated uncertainties. In this chapter we quantify the uncertainty of an aircraft's fuel consumption performance using our decomposition-based uncertainty quantification approach.

In Section 5.1, we introduce the problem statement and motivate the need for decomposition-based uncertainty quantification. In Section 5.2 and Section 5.3 we present the aircraft technology component and aviation environmental impacts component, respectively. We discuss the interface between the aircraft technology component and the aviation environmental impacts component in Section 5.4. A demonstration on how a component-level global sensitivity analysis can be integrated within our decomposition-based uncertainty quantification framework to reduce the number of component-to-component interface variables is given in Section 5.5. Lastly, in Section 5.6 we present the uncertainty quantification results using our decomposition-based approach and compare it the all-at-once Monte Carlo simulation approach.

5.1 Introduction to environmental impacts of aviation

The aviation sector is projected to be one of the fastest growing contributors to anthropogenic greenhouse gas emissions [69]. If left unconstrained, aircraft emissions are projected to quadruple by 2050 [1]. As a measure to manage the climate impact of aviation, the Committee for Environmental Protection (CAEP) under the International Civil Aviation Organization (ICAO) and supported by the FAA, adopted a 2% annual efficiency improvement goal for aviation through 2050 [2]. To satisfy this fast paced fleet-wide improvement requires significant enhancements to aviation technology, sustainable fuels with low CO2 emissions, and efficient operational procedures [101].

To meet these demanding requirements, CAEP assembled a panel of independent experts with varying backgrounds to establish long-term technology goals for aviation fuel consumption [29]. Within their study they investigated future aviation technology scenarios, which represented varying regulatory pressure to reduce fuel consumption. The future aircraft technology scenarios were then applied in analysis tools as "technology packages" to assess the technology improvement on aircraft fuel consumption. Due to resource limitations the independent experts were unable to address the following issues;

- the impact of interdependencies between technologies due to the lack of model integration, and
- the impact of uncertainties in aviation technology on fuel consumption and implications on policy assessment.

In this chapter we address these topics by performing a system-level uncertainty quantification using a conceptual-level aircraft design tool and an aviation environmental impacts tool. The system of interest consists of the Transport Aircraft System OPTimization (TASOpt) [36] component and the Aviation Environment Design Tool (AEDT) version 2a [99] component as depicted in Figure 5-1. A decomposition-based

uncertainty quantification approach is warranted in this application since these components are housed in different locations and owned by different groups. The system output of interest is fuel consumption performance (i.e., fuel energy consumption per payload-range) and is defined as,

$$PFEI = \frac{\sum_i W_{fuel,i} h_{fuel}}{\sum_i W_{pay,i} R_{total,i}}, \tag{5.1}$$

where the summation is over number of missions, $W_{fuel,i}$ is the total fuel consumption of the i^{th} mission, $R_{total,i}$ is the total range of the i^{th} mission, W_{pay} is the payload weight of the i^{th} mission, and h_{fuel} is the specific heating value of Kerosene.

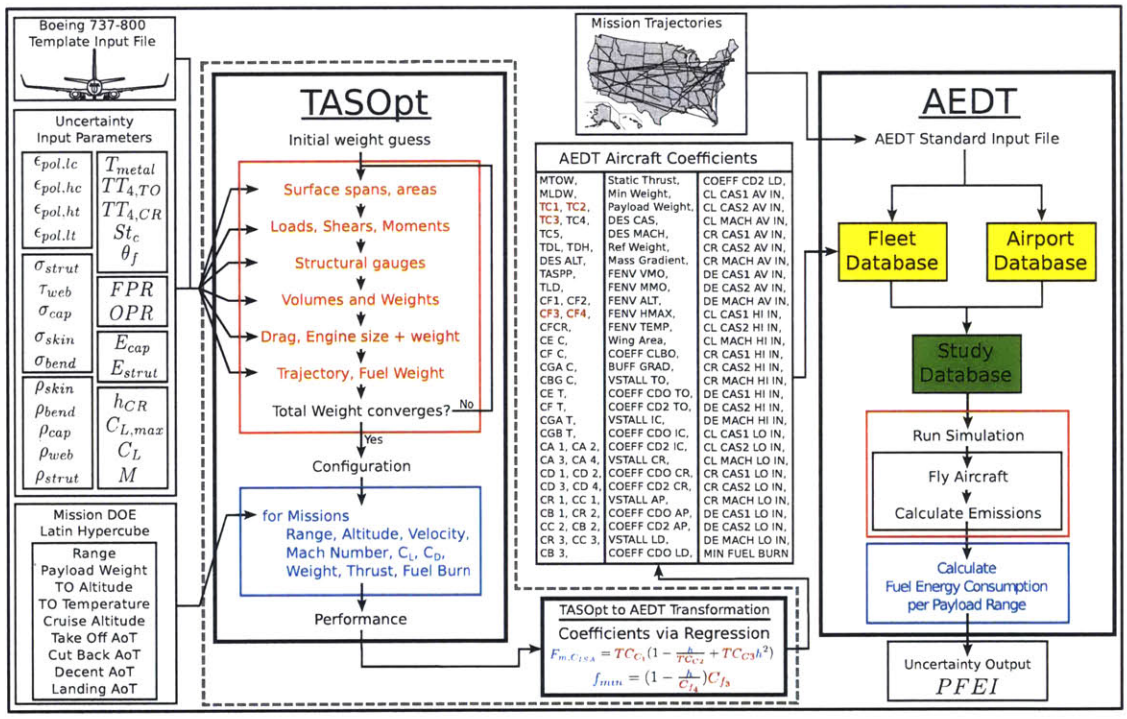

Figure 5-1: System-level uncertainty quantification of the toolset consists of quantifying how uncertainty in aircraft technologies and operations impact the uncertainty in the outputs of interest, here the aircrafts fuel consumption performance. The descriptions for the TASOpt input variables and AEDT input variables are provided in Appendix A and Appendix B, respectively.

5.2 Transport Aircraft System Optimization

The TASOpt component is an aircraft performance tool which allows users to evaluate and size future aircraft with potentially unconventional airframe, aerodynamic, engine, or operation variables using low-order physical models implementing fundamental structural, aerodynamic, and thermodynamic theory. TASOpt uses historical based correlations only when necessary, in particular only for some of the secondary structure and aircraft equipment. The TASOpt component takes as input aircraft technology and operational variables and can either optimize an aircraft over a specified set of constraints or resize an aircraft to meet a desired mission requirement. The aircraft configuration selected for this study is the Boeing 737-800W shown in Figure 5-2. This aircraft operates in the short to medium range while seating approximately 180 passengers.

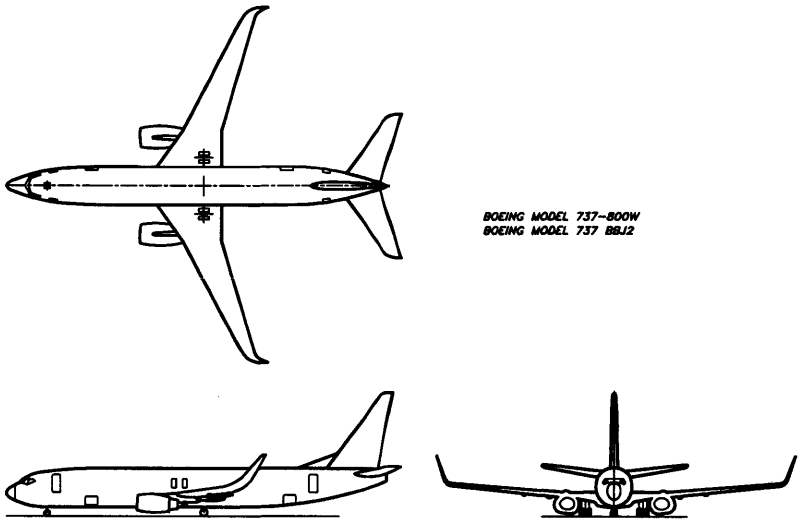

Figure 5-2: Boeing 737-800W airframe configuration. Schematic taken from [26].

5.2.1 TASOpt Inputs

Table 5.1 contains the 27 uncertain TASOpt random input variables selected for this study and their respective distributions. The description of the TASOpt random

input variables are provided in Appendix A. These input variables represent the technological and operational variables of an aircraft which were considered to be uncertain in the design phase. The uncertainty associated with the technology input variables captures our lack of knowledge due to material properties and measurements. The uncertainty associated with the operational input variables captures the designer lack of knowledge in the design phase of an aircraft. Given a single realization vector containing the 27 input variables in Table 5.1, the baseline aircraft's first mission profile and technology variables are populated. The baseline aircraft is created using TASOpt's optimization capabilities to best represent the Boeing 737-800W aircraft configuration (e.g., see Figure 5-1 "Boeing 737-800 Template Input File").

Table 5.1: TASOpt random input variables and their respective distributions.

Input	Units	Distribution	Input	Units	Distribution
$\epsilon_{pol,lc}$	[-]	$\mathcal{U}[0.936, 0.938]$	T_{metal}	[K]	$\mathcal{U}[1172, 1272]$
$\epsilon_{pol,hc}$	[-]	$\mathcal{U}[0.903, 0.904]$	$TT_{4,TO}$	[K]	$\mathcal{U}[1783, 1883]$
$\epsilon_{pol,lt}$	[-]	$\mathcal{U}[0.870, 0.872]$	$TT_{4,CR}$	[K]	$\mathcal{U}[1541, 1641]$
$\epsilon_{pol,ht}$	[-]	$\mathcal{U}[0.875, 0.877]$	St_c	[-]	$\mathcal{U}[0.094, 0.096]$
σ_{strut}	[psi]	$\mathcal{U}[28500, 31500]$	Θ_f	[-]	$\mathcal{U}[0.315, 0.325]$
τ_{web}	[psi]	$\mathcal{U}[19000, 21000]$	FPR	[-]	$\mathcal{U}[1.60, 1.62]$
σ_{cap}	[psi]	$\mathcal{U}[28500, 31500]$	OPR	[-]	$\mathcal{U}[24.2, 28.2]$
σ_{skin}	[psi]	$\mathcal{U}[14500, 15500]$	E_{cap}	[psi]	$\mathcal{U}[9.5e6, 10.5e6]$
σ_{bend}	[psi]	$\mathcal{U}[28500, 31500]$	E_{strut}	[psi]	$\mathcal{U}[9.5e6, 10.5e6]$
ρ_{strut}	[kg/m^3]	$\mathcal{U}[2672, 2726]$	h_{CR}	[ft]	$\mathcal{U}[34000, 36000]$
ρ_{web}	[kg/m^3]	$\mathcal{U}[2672, 2726]$	$C_{L,max}$	[-]	$\mathcal{U}[2.2, 2.3]$
ρ_{cao}	[kg/m^3]	$\mathcal{U}[2672, 2726]$	C_L	[-]	$\mathcal{U}[0.576, 0.578]$
ρ_{skin}	[kg/m^3]	$\mathcal{U}[2672, 2726]$	$Mach$	[-]	$\mathcal{U}[0.77, 0.79]$
ρ_{bend}	[kg/m^3]	$\mathcal{U}[2672, 2726]$	-	-	-

5.2.2 TASOpt Outputs

With the aircraft's first mission profile and technology variables populated, the TASOpt component then sizes the sampled aircraft configuration such that it satisfies the first mission profile. Next we quantify the sampled aircraft's performance by flying an additional 99 mission profiles which are generated using a Latin hypercube design of

experiments. The Latin hypercube design of experiments populates the 12 mission input variables contained in Table 5.2. For any mission variable (e.g., *Range*) we generate 99 realizations from the uniform distribution (i.e., $\mathcal{U}(a, b)$) using the Latin hypercube random sampling scheme and the parameters provided in Table 5.2. This process allows, for example, the *Range* to sweep between 750 [nmi] and 3250 [nmi].

Lastly, the aircraft's configuration, operational procedure, and performance over multiple flight segments and atmospheric conditions are quantified using 100 variables. These TASOpt output variables which characterize the aircraft's configuration, operational procedure, and performance are described in Section 5.4, where we discuss the coupling between TASOpt and AEDT.

Table 5.2: The performance of each sampled aircraft configuration is evaluated using a Latin hypercube design of experiments. Presented here are the TASOpt mission input variables and their respective uniform distribution parameters (i.e., $\mathcal{U}(a, b)$). Parameters containing an asterisk are also TASOpt random input variables. Therefore, the parameters represent differences from their respective realization (i.e., $\mathcal{U}(x - a, x + b)$ where x is the variables sample realization).

Input	Units	a	b
Range	[nmi]	750	3250
W_{max}	[lbs]	165	265
h_{TO}	[ft]	-4000	4000
ΔT	[K]	-12.5	12.5
h_{CR}	[ft]	-4000^*	4000^*
$C_{L,max}$	[-]	-0.05^*	0.05^*
Θ_{TO}	[deg]	39	41
Θ_{IC}	[deg]	2.8	3.2
$\Theta_{DE,1}$	[deg]	-3.2	-2.8
$\Theta_{DE,5}$	[deg]	-3.2	-2.8
C_L	[-]	-0.025^*	0.025^*
Mach	[-]	-0.02^*	0.02^*

5.3 Aviation Environmental Design Tool

The AEDT component is a suite of integrated aviation environmental impact tools. AEDT provides users with the ability to assess the interdependencies among aviation-produced fuel consumption, emissions, and noise. For cruise conditions the AEDT component implements the EUROCONTROL's Base of Aircraft Data (BADA) [90] which uses an energy-balance thrust model and thrust specific fuel consumption modeled as a function of airspeed. The BADA fuel consumption model has been shown to work well in cruise, with differences from airplane reported fuel consumption of about 5% [90]. For terminal conditions (e.g., departure/arrival flights until 10,000 [ft] above ground level), the AEDT component implements a set of energy-balance equations to support a higher level of fidelity in fuel consumption modeling.

5.3.1 AEDT Inputs

The AEDT component characterizes an aircraft using 100 input variables and depicted in Figure 5-1 as the "AEDT Aircraft Coefficients". The descriptions of the AEDT random input variables are provided in Appendix B. These input variables characterizes the aircraft's configuration and operational procedure, and defines the aircraft performance over multiple flight segments and atmospheric conditions. To initialize the AEDT component we first create a duplicate of the Boeing 737-800W aircraft within the AEDT fleet database. Any TASOpt generated aircraft may replace the 100 aircraft input variables in the temporary Boeing 737-800 AEDT fleet database through the component-to-component transformation process discussed in Section 5.4. The objective of the component-to-component transformation process is to ensure that the temporary AEDT aircraft characterized through these 100 input variables is a suitable representation of the TASOpt generated aircraft.

5.3.2 Flight Trajectories

For each randomly generated aircraft we must quantify its respective environmental impacts. To do so we fly each sampled aircraft over a set of deterministic flight

trajectories using the AEDT component. The flight trajectories for this study were selected from a 2006 representative day flight scenario database [88]. Using the representative day, the flights associated with the Boeing 737-700 aircraft were extracted as possible flight trajectories since the Boeing 737-800W was not comprehensively represented in the 2006 day flight scenario. Of the 900 possible flight trajectories (e.g., departure–arrival airport combinations) only 20 representative flight trajectories were selected due to limited computational resources. The 20 flight trajectories are presented in Table 5.3 and illustrated in Figure 5-3. For computational purposes, these flight trajectories are approximated by a great circle path from the departure airport to the arrival airport and were generated by the TASOpt component using the baseline aircraft configuration with the TASOpt random input variables set to their respective means.

Table 5.3: Presented here are the 20 representative flight trajectories (i.e., departure, arrival, and range) flown by the Boeing 737-800W in the TASOpt-AEDT uncertainty quantification study.

Depart Airport	Depart Runway	Arrival Airport	Arrival Runway	Range [nmi]
KDTW	04L	KPVD	21	535
KIAH	26L	KLAX	24L	1197
KLGA	22	KMEM	27	835
KDTW	04L	KSFO	28R	1801
KPDX	28R	KLAX	24L	725
KMIA	08L	KDEN	35R	1482
KPDX	28L	KABQ	08	964
KJFK	31R	KLGB	16L	2138
KIAD	01R	KORD	28	511
KPHX	26	KMSP	35	1106
KBWI	28	KFLL	09L	806
KPHX	26	KFLL	09L	1710
KMCO	35L	KDCA	01	662
KIAH	26L	KBOS	27	1387
KMCO	17R	KMKE	07R	928
KSJC	30R	KIAD	19L	2082
KSFO	28L	KPHX	25L	565
KDFW	35L	KSFO	28R	1270
KPHL	09R	KFLL	09L	864
KCLE	24L	KSFO	28R	1874

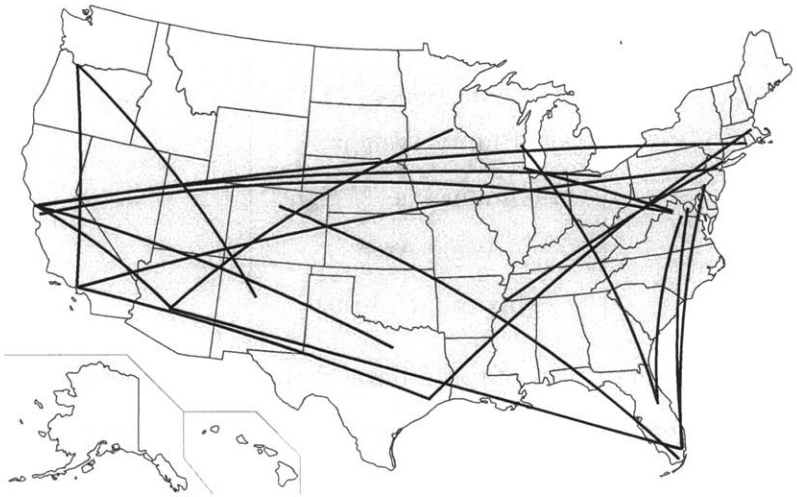

Figure 5-3: Illustrated here are the 20 representative flight trajectories flown by the Boeing 737-800W in the TASOpt-AEDT uncertainty quantification study.

5.4 TASOpt-AEDT Interface

In this section we summarize the transformation between the TASOpt outputs (i.e., aircraft configuration and performance) and the AEDT inputs. For a detailed discussion on this component-to-component transformation we refer to the ACDL Report [REF]. The objective of the component-to-component transformation is to determine the AEDT input variables such that the AEDT aircraft performs similar to the TASOpt generated aircraft. In this section we also validate the component-to-component transformation by comparing for similar scenarios (i.e., aircraft and flight trajectories) the fuel consumption, fuel burn rate, and net corrected thrust from the TASOpt component and from the AEDT component.

5.4.1 Transformation Procedure

The TASOpt component outputs the aircraft's configuration variables,

$$\begin{bmatrix} \text{Maximum Takeoff Weight} \\ \text{Empty Weight} \\ \text{Maximum Fuel Weight} \\ \text{Wing Area} \\ \text{Maximum Thrust} \end{bmatrix}.$$

The TASOpt component also outputs data characterizing the aircraft's performance. For each mission the aircraft performance is provided on 15 flight segments (i.e., 3 takeoff, 5 climb, 2 cruise, and 5 descent segments). In total TASOpt outputs 1,500 (i.e., 100 missions by 15 segments) individual performance data points. Each performance data point contains the variables,

$$\begin{bmatrix} \text{Range} \\ \text{Altitude} \\ \text{True Airspeed} \\ \text{Mach Number} \\ \text{Lift Coefficient} \\ \text{Drag Coefficient} \\ \text{Aircraft Weight} \\ \text{Thrust} \\ \text{Fuel Burn Rate} \\ \text{Angle of Attack} \\ \text{Total Temperature at Engine Inlet} \\ \text{Total Pressure at Engine Inlet} \end{bmatrix}.$$

Of the 100 mission profiles flown by TASOpt the first 50 are flown under international standard atmosphere (ISA) conditions while the remaining 50 are flown under non-ISA conditions. The component-to-component transformation procedure relies on linear and nonlinear regression routines over the appropriate mission segments to extract the AEDT aircraft performance coefficients from the TASOpt aircraft performance data. We present here an example of how to extract the AEDT thrust

coefficients TC_{C_1}, TC_{C_2}, and TC_{C_3} during the high altitude climb segment under ISA conditions. These AEDT aircraft performance coefficients are related to the TASOpt outputs thrust, F, and altitude, h, by,

$$F = TC_{C_1} \cdot (1 - \frac{h}{TC_{C_2}} + TC_{C3} \cdot h^2). \tag{5.2}$$

Since we are concerned with the high altitude climb segment under ISA conditions we only use the TASOpt performance data contained in missions 1 through 50 and flight segments 5 through 8. With the TASOpt performance data collected we perform a linear regression to obtain the AEDT coefficients from Equation 5.2. A similar procedure to that shown here is performed for the all 100 AEDT random input variables.

5.4.2 Validation

In the validation study we compare the fuel consumption, fuel burn rate, and net corrected thrust between an aircraft generated by and flown in the TASOpt component to that aircraft imported and flown in the AEDT component. The flight trajectories for this study are great circle flight paths from Boston to Atlanta, Boston to Denver, and Boston to Los Angeles. The objective of this validation study is to determine if the component-to-component transformation has suitably represented the TASOpt generated aircraft in the AEDT fleet database. Presented in Figure 5-4 are the results produced by the TASOpt component and the AEDT component for the Boston to Atlanta flight. The results demonstrate that the fuel burn rate and net corrected thrust computed by TASOpt and AEDT agree well throughout most the the flight. The largest discrepancy, which occurs in the arrival segment, can be accounted for by the difference between how AEDT 2a and TASOpt model an arrival procedure. In AEDT 2a the aircraft engines are assumed to be idle which is an unrealistic modeling assumption and has been corrected in AEDT version 2b. As a result AEDT 2a has a visibly lower fuel burn rate than TASOpt in the arrival segment. The Boston to Denver and Boston to Los Angeles flight trajectories produced similar results.

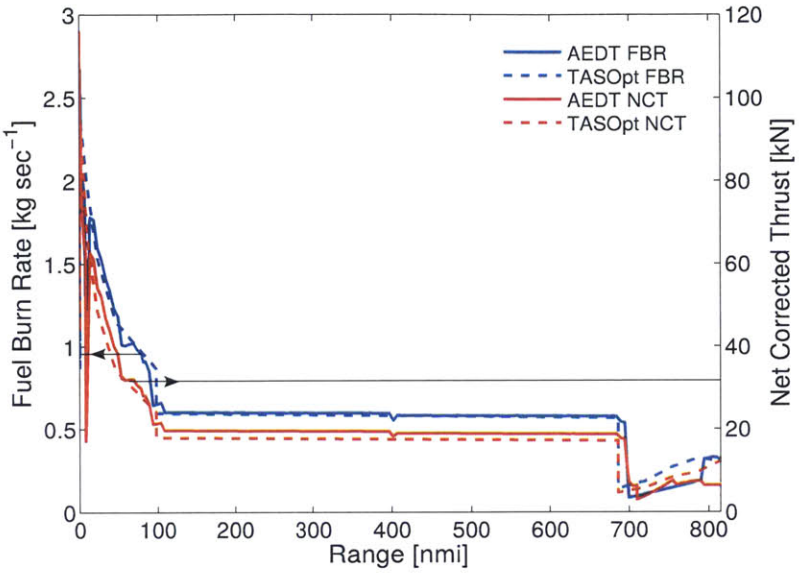

Figure 5-4: Validation study of the TASOpt component to AEDT component transformation for the Boston to Atlanta great circle flight. The results illustrate that the fuel burn rate (FBR) and net corrected thrust (NCT) match well throughout most of the flight trajectory.

The total fuel consumption and relative error for each flight trajectory are given in Table 5.4. The results indicate a good comparison between the components, indicating that the component-to-component transformation performed well. Of the three trajectories, the flight from Boston to Denver performed the worst however; this discrepancy can be attributed to the flight trajectory's range. The AEDT 2a component assumes a constant takeoff weight which is dependent only on the flight trajectory range whereas TASOpt evaluates the takeoff weight continuously. As a result, the Boston to Denver flight which has a range of 1520 [nmi] is on the lower end of the AEDT takeoff profile spectrum (i.e., 1500–2499 [nmi]). This causes the AEDT component to set a takeoff weight which is significantly larger than the corresponding TASOpt takeoff weight. As a result, the aircraft belonging to the AEDT component's requires more fuel than the corresponding aircraft belonging to the TASOpt component. The Boston to Atlanta and Boston to Los Angeles flights have flight trajectory ranges of 850 [nmi] and 2300 [nmi] respectively which resulted in a better agree-

Table 5.4: Presented here are the fuel consumption results over the three flight trajectories. The TASOpt row represents an aircraft generated by and flown in TASOpt. The AEDT row represents the TASOpt aircraft imported into the AEDT component through the component-to-component transformation and then flown on the same flight trajectory as the TASOpt flight trajectory.

	Boston-Atlanta		Boston-Denver		Boston-Los Angeles	
	Fuel [lb]	% Rel Error	Fuel [lb]	% Rel Error	Fuel [lb]	% Rel Error
TASOpt	5015.6	-	8715.3	-	13186.4	-
AEDT	5284.7	5.37	9737.4	9.62	13594.4	3.09

ment between the AEDT takeoff weight and the TASOpt takeoff weight. In turn this resulted in a better agreement between the TASOpt and AEDT fuel consumptions.

5.5 Dimension Reduction

The difficulty with performing a decomposition-based uncertainty quantification of the system illustrated Figure 5-1 lies in the 100-dimensional component-to-component transformation. The reason is because, as stated in Section 2.8, the high-dimensional component-to-component transformation requires a challenging high-dimensional change of measure. Although in Chapter 4 we improved the dimensional scalability of the change of measure problem, the dimension of this component-to-component transformation remains beyond our capabilities. As a result, we make here the assumption that many of the AEDT input variables are expected to have an insignificant impact on the system output of interest distribution. If our assumption is valid it implies that we can discard most of the AEDT input variables when performing the change of measure procedure. Discarding an AEDT input variable from the change of measure procedure implies that we do not require that its distribution function weakly converges to its respective target distribution function. In this section we quantify the influence of the AEDT input variables on the system output of interest and use this information to determine which of the AEDT input variables should take part in the change of measure procedure.

5.5.1 Problem Setup

To quantify the influence of the AEDT input variables on the system output of interest we perform a component-level global sensitivity analysis of the AEDT component. However, a component-level global sensitivity analysis of the AEDT component is complicated by the fact that the TASOpt output variables are dependent. This dependency structure renders the classical ANOVA dimensional decomposition (ADD) of the AEDT component inapplicable since the primary assumption in Section 3.1 is that the input variables are independent. Fortunately, recent efforts have investigated dependently distributed global sensitivity analysis and ADD [21, 57, 71, 76, 65, 95]. In this section we follow the ADD implementation of Rahmen [95], construct an ADD of the AEDT component, and use this information to identify the influential AEDT random input variables. The ADD construction of the AEDT component is termed generalized ADD to differentiate it from the classical ADD presented in Section 3.1.

Before constructing a generalized ADD of the AEDT component we first reduce the dimension of the AEDT random input variables using engineering judgment. This primary reduction in dimension is illustrated in Figure 5-5. The rationale behind this primary dimension reduction is that the quantity of interest (i.e., fuel consumption performance) is largely dependent on the takeoff and cruise segments of the flight trajectory. All AEDT coefficients not present in this set do not take part in the change of measure procedure of the decomposition-based uncertainty quantification approach. This implies that although we still treat these variables as uncertain, we do not require that their distribution functions weakly converge to their target distribution functions from the upstream TASOpt component.

With the primary dimension reduction completed the task now is to generate a generalized ADD of the AEDT component in Figure 5-5. In order to generate the generalized ADD we must first select a distribution for the 50 AEDT random input variables. The distribution selected is a multivariate Gaussian distribution since it simplifies the construction of the generalized ADD (e.g., Hermite polynomial basis) and can be easily parameterized in high-dimensional applications such as this. This

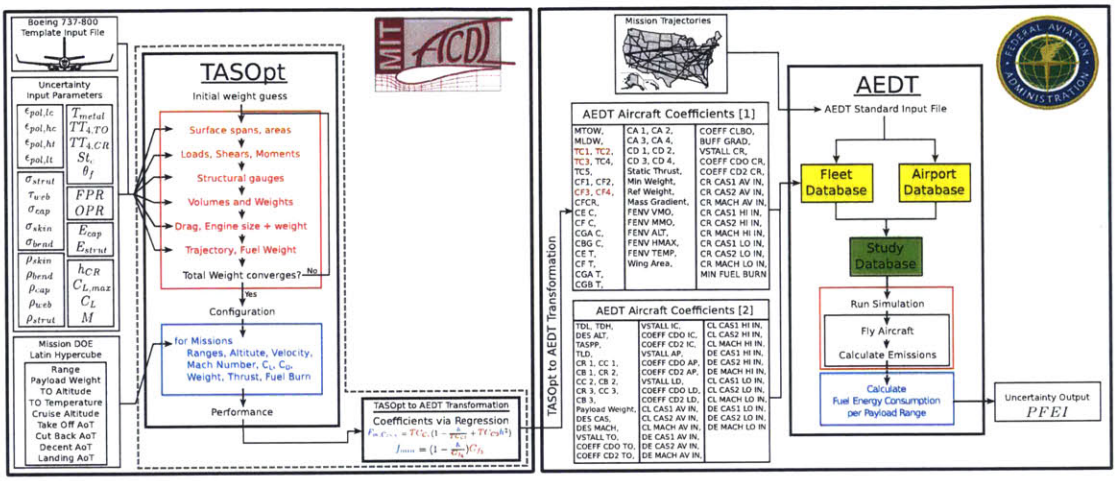

Figure 5-5: We partitioned the AEDT input variables into two sets; an influential set and a noninfluential set. The distribution function of the AEDT inputs in the second set are not required to converge weakly to the target distribution function.

multivariate Gaussian distribution is a guess of the unknown target distribution. Whether this distribution is adequate for ranking the inputs in order of influence and if an a posterior error check exists are topics that require further research. The probability density function of this distribution is expressed by

$$p_{\hat{\mathbf{Y}}}(\mathbf{t}) = (2\pi)^{-\frac{d}{2}}(|\Sigma_{\hat{\mathbf{Y}}}|)^{-\frac{1}{2}} exp\bigg(-\frac{1}{2}\mathbf{t}^{\top}\Sigma_{\hat{\mathbf{Y}}}^{-1}\mathbf{t}\bigg), \tag{5.3}$$

where $\Sigma_{\hat{\mathbf{Y}}} = \mathbb{E}_{\hat{\mathbf{Y}}}[\hat{\mathbf{Y}}\hat{\mathbf{Y}}^{\top}]$ is the covariance matrix. Here we use the notation $\hat{\mathbf{Y}}$ to emphasize that this is a guess to the unknown target distribution. To construct the multivariate Gaussian distribution, Equation 5.3, necessary for the generalized ADD we used $10,000$ target samples from the upstream TASOpt component. The reason for this step is because we did not have adequate experience or prior knowledge of this system to guess the target distribution.

5.5.2 Generalized ANOVA Dimensional Decomposition

Similar to the classical ADD the square-integrable multivariate function f admits a unique, finite, hierarchical expansion

$$f(\mathbf{t}) = \sum_{\mathcal{A}\subseteq\{1,\ldots,d\}} f_{\mathcal{A}}(\mathbf{t}_{\mathcal{A}}), \tag{5.4}$$

in terms of subcomponent functions $f_{\mathcal{A}}$ of input variables with increasing dimensions [95]. However, unlike the subcomponent functions in the classical ADD, the subcomponent functions in the generalized ADD cannot be derived from the strong annihilating conditions. This is because the orthogonal properties that stem from the classical ADD cannot be duplicated when the random variables are dependent. Yet, the subcomponent functions $f_{\mathcal{A}}$ can be obtained from a similar approach by weakening the annihilating conditions (see Rahman [95] for a thorough discussion of the generalized ADD).

Given the subcomponent functions of the generalized ADD, subsequent evaluations of their second-moment characteristics, including global sensitivity indices, can be readily computed. Applying the expectation operator on the generalized ADD results in the mean

$$\mu = \mathbb{E}_{\mathbf{Y}}[f(\mathbf{t})] = f_{\{0\}}. \tag{5.5}$$

Applying the expectation operator again, this time on $(f(\mathbf{t}) - \mu)^2$, along with the weakened annihilating conditions results in the variance [95]

$$\sigma^2 = \mathbb{E}_{\mathbf{Y}}[(f(\mathbf{t}) - \mu)^2] = \sum_{0\neq\mathcal{A}\subseteq\{1,\ldots,d\}} \mathbb{E}_{\mathbf{Y}}[f_{\mathcal{A}}^2] + \sum_{\substack{0\neq\mathcal{A},\bar{\mathcal{A}}\subseteq\{1,\ldots,d\} \\ \mathcal{A}\nsubseteq\bar{\mathcal{A}}\nsubseteq\mathcal{A}}} \mathbb{E}_{\mathbf{Y}}[f_{\mathcal{A}} f_{\bar{\mathcal{A}}}]. \tag{5.6}$$

The first sum represents variance contributions from all nonconstant subcomponent functions. In contrast, the second sum, which is not present in the classical ADD, contains covariance contributions from two distinct nonconstant subcomponent functions that are not orthogonal—a ramification of imposing the weak annihilating conditions appropriate for the generalized ADD.

For input variables involving dependent probability distributions, a triplet of global sensitivity indices is defined. The three $|\mathcal{A}|$-variate global sensitivity indices of $f_{\mathcal{A}}$ are denoted by $S_{\mathcal{A},v}$, $S_{\mathcal{A},c}$, and $S_{\mathcal{A}}$, and defined by the ratios

$$S_{\mathcal{A},v} = \frac{\mathbb{E}_{\mathbf{Y}}[f_{\mathcal{A}}^2]}{\sigma^2}, \tag{5.7}$$

$$S_{\mathcal{A},c} = \frac{\sum_{0 \neq \bar{\mathcal{A}} \subseteq \{1,\ldots,d\}} \mathbb{E}_{\mathbf{Y}}[f_{\mathcal{A}} f_{\bar{\mathcal{A}}}]}{\sigma^2}, \tag{5.8}$$
$$\mathcal{A} \nsubseteq \bar{\mathcal{A}} \nsubseteq \mathcal{A}$$

$$S_{\mathcal{A}} = S_{\mathcal{A},v} + S_{\mathcal{A},c}. \tag{5.9}$$

The first two indices, $S_{\mathcal{A},v}$ and $S_{\mathcal{A},c}$, represent the normalized versions of the variance contribution from $f_{\mathcal{A}}$ to σ^2 and of the covariance contributions from $f_{\mathcal{A}}$ and all $f_{\bar{\mathcal{A}}}$, such that $\mathcal{A} \nsubseteq \bar{\mathcal{A}} \nsubseteq \mathcal{A}$, to σ^2. They are termed the variance-driven global sensitivity index and the covariance-driven global sensitivity index, respectively, of $f_{\mathcal{A}}$. The third index, $S_{\mathcal{A}}$, referred to as the total global sensitivity index of $f_{\mathcal{A}}$, is the sum of variance and covariance contributions from or associated with $f_{\mathcal{A}}$ to σ^2. Since $0 \neq \mathcal{A} \subseteq \{1, \ldots, d\}$, there exist $2^d - 1$ such triplets of indices, which also sums to unity,

$$\sum_{0 \neq \mathcal{A} \subseteq \{1,\ldots,d\}} S_{\mathcal{A}} = \sum_{0 \neq \mathcal{A} \subseteq \{1,\ldots,d\}} S_{\mathcal{A},v} + \sum_{0 \neq \mathcal{A} \subseteq \{1,\ldots,d\}} S_{\mathcal{A},c} = 1. \tag{5.10}$$

From the definitions, the variance-driven sensitivity index $S_{\mathcal{A},v}$ is a nonnegative, real valued number. It reflects the variance contribution from $\mathbf{t}_{\mathcal{A}}$ to the total variance of f. In contrast, the covariance-driven sensitivity index $S_{\mathcal{A},c}$ can be negative, positive, or zero, depending on the correlation between $\mathbf{t}_{\mathcal{A}}$ and $\mathbf{t}_{\bar{\mathcal{A}}}$. It represents the variance contribution from $\mathbf{t}_{\mathcal{A}}$ by the interaction to $\mathbf{t}_{\bar{\mathcal{A}}}$, when $\mathcal{A} \nsubseteq \bar{\mathcal{A}} \nsubseteq \mathcal{A}$, due to the dependence structure. An individual index $S_{\mathcal{A}}$ may exceed unity or be negative, but the sum of all these indices is always equal to unity. When the random variables are independent, the covariance-driven contribution to the total sensitivity index vanishes, leaving behind only the variance-driven sensitivity contribution. The global sensitiv-

ity indices, whether derived from the generalized or classical ADD, can be used to rank variables in order of influence on the output of interest variance, fix unessential variables, and reduce dimensions of large-scale problems.

5.5.3 Results

Following the generalized ADD implementation of Rahmen [95], a generalized ADD of the AEDT component was constructed. The generalized ADD hierarchical expansion given by Equation 5.4 was terminated upon capturing 99% of the output of interest variance, Equation 5.6. The generalized ADD of the AEDT component is composed of subcomponent functions which are first-order basis functions (i.e., $\mathcal{A} = \{i\}$) with at most first-order polynomials (i.e., $f_{\{i\}}(t) = m_i t + b_i$). With the subcomponent functions in hand we compute the absolute maximum allowable variation (i.e., $|S_{\mathcal{A},v}| + |S_{\mathcal{A},c}|$) due to each AEDT random input variable. Using this criteria we rank the AEDT random input variables in decreasing order of influence as illustrated in Figure 5-6. These results reinforce our assumption that only a small subset of the 50 AEDT random input variables has an influence on the system output of interest variation.

The sensitivity matrix for the first 15 most influential AEDT input variables variables is illustrated in Figure 5-7. The quantities on the diagonal entries are the variance-driven sensitivity indices and are all nonnegative. The quantities on the nondiagonal entries are the covariance-driven sensitivity indices and can be any real valued number. The remaining inputs not presented here do not have a significant impact on the output of interest variation. Note, in this study we have ranked the influence of an input variable on the output variation, which is not guaranteed to be the same if we were to rank the influence of the input variable on the output distribution [17]. This is an important observation since we are interested in quantifying the output of interest distribution.

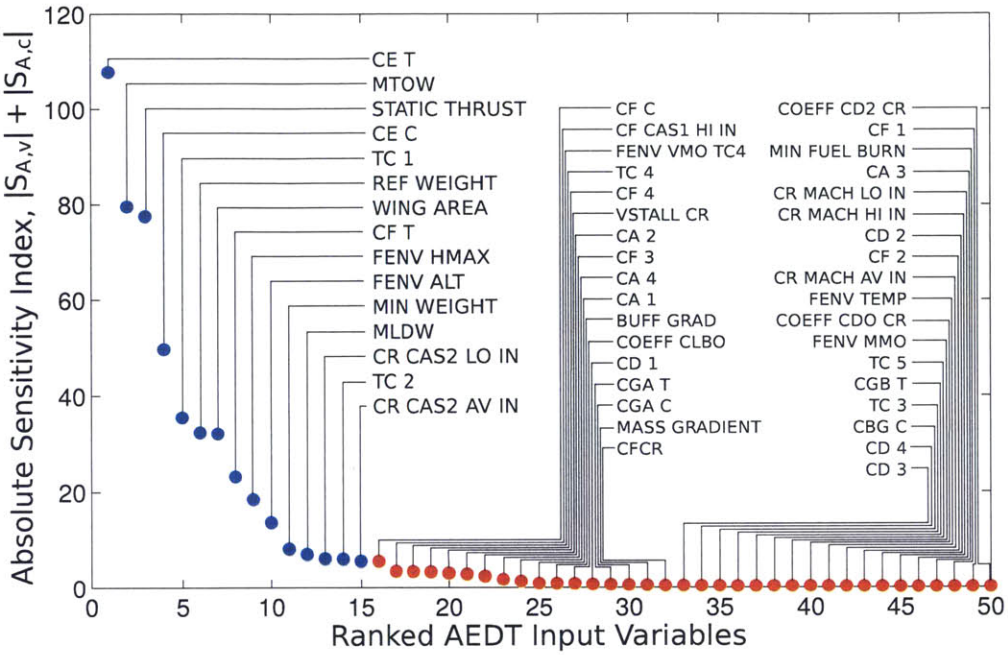

Figure 5-6: This plot presents the absolute variance- and covariance-driven sensitivity indices from each of the 50 AEDT random input variables on the AEDT fuel consumption performance variance. This plot illustrates that the absolute variance- and covariance-driven sensitivity indices decays very rapidly and that 15 AEDT random input variables capture almost all of the AEDT fuel consumption performance variance.

5.6 Uncertainty quantification

In the previous section we identified 15 AEDT random input variables which are expected to have the greatest influence on the AEDT output of interest variation. From these results the remaining noninfluential AEDT random input variables are not incorporated into the change of measure procedure of the decomposition-based uncertainty quantification approach. As before, this implies that although we still treat these variables as uncertain, we do not require that their distribution functions weakly converge to their target distribution functions from the upstream TASOpt component. The final reduced system layout for the decomposition-based uncertainty quantification study is illustrated in Figure 5-8.

The proposal distribution for the AEDT component is selected to be the distribu-

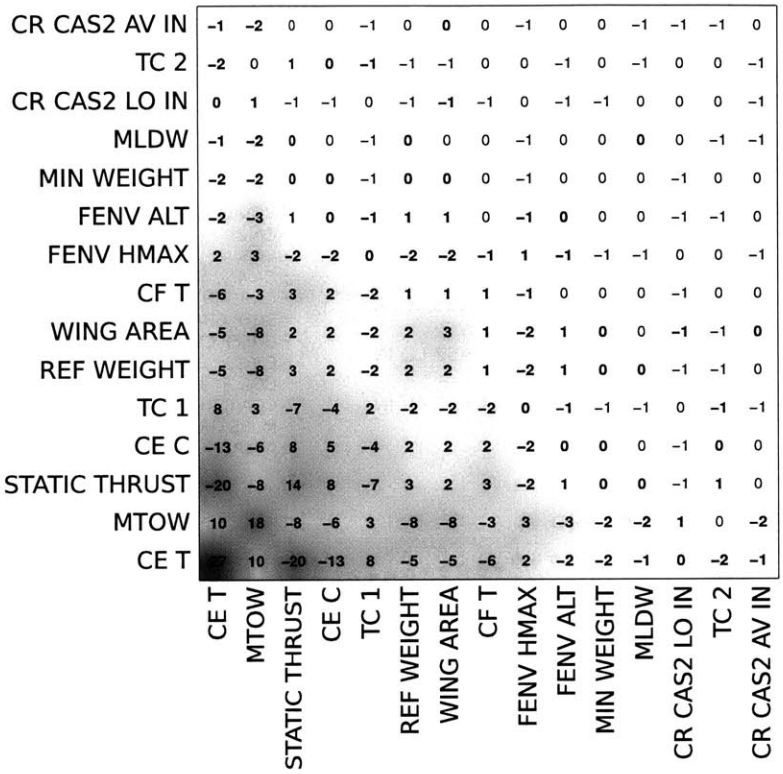

Figure 5-7: Sensitivity matrix of the 15 most influential AEDT random input variables impacting the AEDT fuel consumption performance variance.

tion used for the construction of the generalized ADD in Section 5.5 but with three times the covariance term. The proposal covariance term was multiplied by a factor of three to ensure the unknown target distribution from the upstream TASOpt component is supported by the AEDT input proposal distribution.

5.6.1 Uncertainty Analysis

The uncertainty analysis results are presented in Figure 5-9 and Figure 5-10. The results illustrate the system output of interest distribution function under three different scenarios. The first scenario is the outcome of running the AEDT component under the proposal distribution assumption. The second scenario is produced by performing an all-at-once uncertainty analysis of the system illustrated in Figure 5-1 using Monte Carlo simulation. The last scenario is the result of performing a decomposition-based uncertainty analysis of the system illustrated in Figure 5-8.

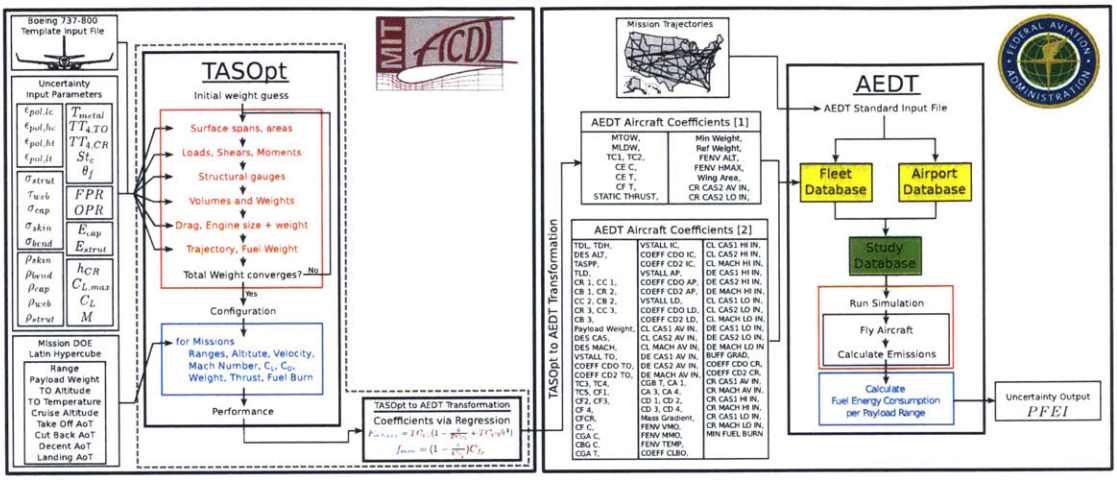

Figure 5-8: We partitioned the AEDT input variables into two sets; an influential set and a noninfluential set. The noninfluential set contains AEDT input variables which were deemed not to influence the cruise and takeoff segments as well as input variables which were labeled as noninfluential by the AEDT component global sensitivity analysis. The influential set contains AEDT random input variables which were labeled as influential by the AEDT component global sensitivity analysis.

These results demonstrate that our decomposition-based uncertainty analysis of the reduced system in Figure 5-8 accurately predict the results from the standard approach, the all-at-once Monte Carlo simulation of the full system in Figure 5-1. This implies that we not only performed the change of measure across the 15 dimensional interface between the TASOpt component and the AEDT component accurately but that we successfully identified the 15 influential AEDT random input variables. If either one of these procedures had performed poorly (i.e., a poor change of measure or a failure in identifying the influential AEDT random input variables) we may not have accurately captured the target distribution using the decomposition-based uncertainty analysis approach. Therefore, by reducing the dimensions of the component-to-component interface we illustrated that our decomposition-based uncertainty analysis approach can be extended to calculate the relevant statistics and failure probabilities of complex and high-dimensional systems.

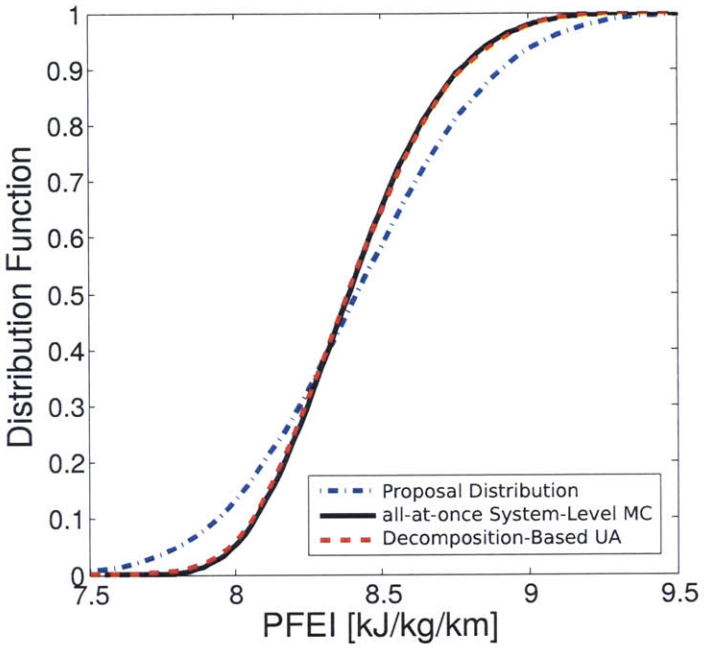

Figure 5-9: The AEDT fuel consumption performance distribution is shown here using the AEDT output proposal distribution, all-at-once Monte Carlo uncertainty analysis, and the decomposition-based uncertainty analysis. These results suggest that our decomposition-based uncertainty analysis performed adequately which implies the change of measure across the 15 AEDT random input variables was successful and that the correct 15 AEDT random input variables were selected by the AEDT component-level global sensitivity analysis.

5.6.2 Global Sensitivity Analysis

The next step in the uncertainty quantification is to identify which of the system inputs have the largest influence, on average, on the system output of interest variation. The results of the system-level main sensitivity indices are presented in Figure 5-11. The results demonstrate the system-level main sensitivity indices computed using the all-at-once Monte Carlo simulation approach of the system illustrated in Figure 5-1 and the decomposition-based approach of the system illustrated in Figure 5-8. These result confirm that our decomposition-based sensitivity analysis algorithm can accurately quantify the main sensitivity indices of the system in Figure 5-1. As previously mentioned, the decomposition-based sensitivity analysis algorithm hinged on the fact that we could evaluate the decomposition-based uncertainty analysis. The main sen-

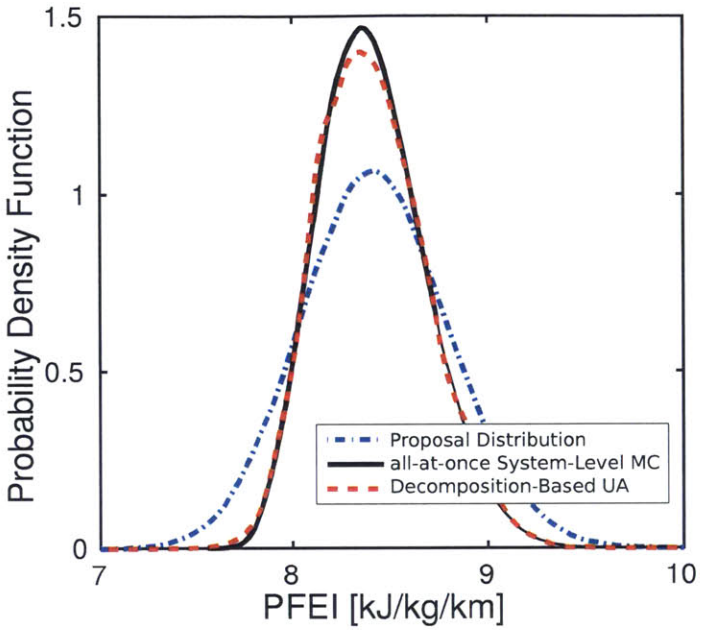

Figure 5-10: The AEDT fuel consumption performance probability density function is shown here using the AEDT output proposal probability density function, all-at-once Monte Carlo uncertainty analysis, and the decomposition-based uncertainty analysis. These results complement the distribution function results provided in Figure 5-9 and suggest that our decomposition-based uncertainty analysis performed adequately which implies the change of measure across the 15 AEDT random input variables was successful and that the correct 15 AEDT random input variables were selected by the AEDT component-level global sensitivity analysis.

sitivity indices also suggest that the output of interest variation is mostly dominated by a handful of aircraft technological and operational variables. That is, by improving our understanding of these TASOpt random input variables through research we can reduce the variation of the system output of interest which is beneficial for decision- and policy-making.

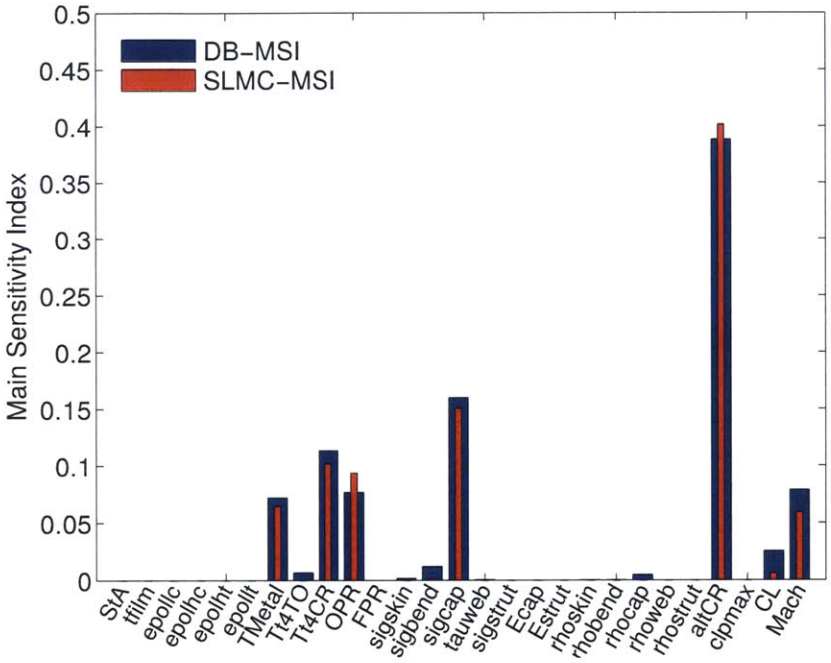

Figure 5-11: The system-level main sensitivity indices are shown here using the all-at-once Monte Carlo global sensitivity analysis and the decomposition-based global sensitivity analysis. These results suggest that our decomposition-based global sensitivity analysis performed adequately and that only a handful of technological and operational system input variables have a significant influence, on average, on the system output of interest. A description of the system inputs are provided in Appendix A

Chapter 6

Conclusions

In this chapter we provide a summary of the work done to meet the thesis objectives, enumerate the thesis contributions, and discuss future research that should be considered.

6.1 Summary

A decomposition-based uncertainty quantification approach intended to support the decision- and policy-making process in today's modern systems was established. Our decomposition-based uncertainty quantification approach investigated two topics; uncertainty analysis and sensitivity analysis. In traditional approaches, quantifying uncertainty in multicomponent systems either approximated the multicomponent system through the use of surrogate modeling or simplified the representation of the multicomponent system's uncertainty. Instead, our approach was motivated by the advantages brought about by decomposition: managing complexity through a divide-and-conquer strategy, exploiting specific disciplinary expertise through local analyses, promoting disciplinary/component autonomy while maintaining an awareness of system-level issues, and being consistent with many organizational structures. These are essential characteristics to achieve a sustainable strategy that manages uncertainty in the complex settings of today's modern engineered systems.

These characteristics are emphasized by drawing analogies between our decomposition-

based approach to uncertainty quantification and decomposition-based strategies for multidisciplinary optimization. In the multidisciplinary optimization literature, approaches are often categorized as monolithic versus distributed, and open consistency versus closed consistency. Our approach has a distributed architecture, that is, the multicomponent system uncertainty analysis is partitioned into multiple uncertainty analyses. This is in contrast to a monolithic architecture, which solves the problem in its entirety. Our approach has open consistency, that is, our samples are initially observations from the incorrect probability distributions but upon convergence become consistent with the desired distributions. In contrast, a closed consistency formulation requires that each sample satisfy global compatibility constraints at every stage of the algorithm.

In Chapter 2, we developed a decomposition-based uncertainty analysis approach to quantify the distribution function of all the variables belonging to the system. The proposed approach decomposed the multicomponent uncertainty analysis into manageable components, and synthesized the system uncertainty analysis without needing to evaluate the system in its entirety. Further, the proposed approach was shown to be provably convergent in distribution. Building upon the decomposition-based uncertainty analysis algorithm we developed, in Chapter 3, a decomposition-based sensitivity analysis algorithm. The proposed approach evaluated the system's main effect indices. The main effect indices may be used for factor prioritization since they provide the percentage of how much system output variability can be reduced, on average, by fixing a particular system input somewhere on its domain.

In Chapter 4, we investigated topics which might hinder implementing our decomposition-based uncertainty quantification approach to today's modern engineered systems. In particular, we focused on the high-dimensional component-to-component interfaces and the challenges brought about through the change of measure procedure. This thesis presented a novel approach that defines and computes empirical importance weights, and shows its connections to other discrepancy metrics and studies. A key attribute of the approach is its scalability: it lends itself well to handling a large number of samples through a scalable optimization algorithm. The proposed approach

also scales to problems with higher dimensions than was previously possible using the current practices. Our work also implemented, in Chapter 5, a component-level global sensitivity analysis to assist in reducing the dimensions of a component-to-component interface. This procedure, combined with the empirical change of measure, allowed us to implement our decomposition-based uncertainty quantification approach to a realistic application.

The realistic application presented in Chapter 5 investigated the environmental impacts of aviation technologies and operations. The multicomponent system was composed of a conceptual-level aircraft design tool and an environmental impacts tools. The challenges of this multicomponent system where the long computational run times and the high-dimensional component-to-component interface. By using the component-level global sensitivity analysis we were able to identify the most influential component-to-component interface variables and thereby reduce the dimensionality of the component-to-component interface. With this reduction in dimensions, our decomposition-based uncertainty quantification approach effectively determined the quantities deemed necessary for supporting the decision- and policy-making process.

The contributions of this thesis are:

1. A methodology that enables decomposition-based uncertainty analysis of feed-forward multicomponent systems which is provably convergent in distribution to the integrated system-level uncertainty analysis.

2. A methodology that enables decomposition-based global sensitivity analysis of feed-forward multicomponent systems.

3. An approach for the change of empirical measure which used the observable and well-defined empirical distribution function and demonstrated scalability to high dimensional distributions.

4. A demonstration of the decomposition-based uncertainty quantification of feed-forward multicomponent systems on a realistic application; environmental impacts of aviation.

6.2 Future work

In Chapter 2, we focused on a method for the forward propagation of uncertainties through a feed-forward system. Further research is required to extend the decomposition-based methodology to systems with two-way coupling and/or feedback loops. Such systems will require iteration between components in the online phase, thus destroying the clean partition between offline and online phases that we achieve in the feedforward case. Despite this drawback, the decomposition can still realize many of the advantages discussed in Section 1.1. A further limitation is that the current approach has been presented only for systems with random variables and random vectors. An important area of future work is to extend the method to systems that include random fields. This will require drawing on recent work to couple Karhunen-Loève expansion representations of random fields.

Regarding the decomposition-based sensitivity analysis in Chapter 3, our approach was limited computationally to the main effect indices. However, one may find it necessary to extend the decomposition-based sensitivity analysis to include indices which capture interaction effects. Additionally, from a divide-and-conquer perspective, one may also consider performing the component-level global sensitivity analyses concurrently and synthesizing this information to perform the decomposition-based multicomponent global sensitivity analysis. Lastly, an added extension would be to incorporate these component-level global sensitivity analyses to reduce the complexity (i.e., reducing the dimensions of the component-to-component interfaces) of the multicomponent system without sacrificing the accuracy of the multicomponent system quantities of interest. Although this thesis did touch on this topic in Section 5.5, our study used a variance-based method to perform the dimension reduction, which is not guaranteed to produce the same results had we ranked the influence according to the variation in the output distribution function.

In our empirical change of measure process, presented in Chapter 4, numerical challenges regarding high dimensions arose due to ill-conditioning of the matrix H. These challenges can be addressed, as they have in other fields such as optimization

of systems governed by partial differential equations [15], through a combination of preconditioning techniques and use of optimization solvers that are tolerant to ill-conditioned matrices. Other future directions of interest include exploitation of the optimization solution process to generate sparse solutions, which may yield a way to derive efficient Monte Carlo integration rules that rely on a condensed set of samples [46], and exploring different objective function metrics (in particular replacing the L_2–norm metric with an L_1–norm metric).

In Chapter 5, our decomposition-based uncertainty quantification approach, in combination with the empirical change of measure, was performed on the environmental impacts of aviation tool suite. However, this multicomponent system was composed of only two components. To extend this work to different and more complex architectures of feed-forward multicomponent systems will require accounting for dependency among variables using only the samples but no explicit description of their underlying probability density functions. This will require that Algorithm 2 be rewritten without any explicit description of the underlying probability density functions or distribution functions.

Appendix A

TASOpt Random Input Variables

$\epsilon_{pol,lc}$	[-], Low pressure compressor polytropic efficiency.
$\epsilon_{pol,hc}$	[-], High pressure compressor polytropic efficiency.
$\epsilon_{pol,lt}$	[-], Low pressure turbine polytropic efficiency.
$\epsilon_{pol,ht}$	[-], High pressure turbine polytropic efficiency.
σ_{strut}	[psi], Maximum allowable strut stress.
τ_{web}	[psi], Maximum allowable wing spar web shear stress.
σ_{cap}	[psi], Maximum allowable wing spar cap stress.
σ_{skin}	[psi], Maximum allowable fuselage skin pressurization stress.
σ_{bend}	[psi], Maximum allowable shell bending stress.
ρ_{strut}	[kg/m^3], Strut material density.
ρ_{web}	[kg/m^3], Wing box web material density.
ρ_{cap}	[kg/m^3], Wing box cap material density.
ρ_{skin}	[kg/m^3], Fuselage pressure-skin material density.
ρ_{bend}	[kg/m^3], Fuselage bending material density.
T_{metal}	[K], Turbine material temperature.
$TT_{4,TO}$	[K], Turbine inlet total temperature for takeoff.
$TT_{4,CR}$	[K], Turbine inlet total temperature for cruise.
St_c	[-], Turbine area-weighted Stanton number.
Θ_f	[-], Cooling effectiveness ratio.
FPR	[-], Design fan pressure ratio.

OPR	[-], Overall design pressure ratio.
E_{cap}	[psi], Wing spar cap modulus of elasticity.
E_{strut}	[psi], Strut modulus of elasticity.
h_{CR}	[ft], Cruise altitude.
$C_{L,max}$	[-], Maximum lift coefficient.
C_L	[-], Cruise lift coefficient.
$Mach$	[-], Cruise Mach number.

Appendix B

AEDT Random Input Variables

MTOW (MAX_TAKEOFF_WGT) [kg], Maximum takeoff weight.

MLDW (MAX_LANDING_WGT) [kg], Maximum landing weight.

TC1 (COEFF_TC1) [N], 1^{st} maximum climb thrust coefficient.

TC2 (COEFF_TC2) [ft], 2^{nd} maximum climb thrust coefficient.

TC3 (COEFF_TC3) [ft^2], 3^{rd} maximum climb thrust coefficient.

TC4 (COEFF_TC4) [K], 4^{th} thrust temperature coefficient.

TC5 (COEFF_TC5) [K^{-1}], 5^{th} thrust temperature coefficient.

DES_ALT [ft], Descent altitude.

TDH (COEFF_TDH) [-], High altitude descent thrust coefficient.

TLD (COEFF_TDL) [-], Low altitude descent thrust coefficient.

TAPP (COEFF_TAPP) [-], Approach thrust coefficient.

TLD (COEFF_TLD) [-], Landing thrust coefficient.

DES_CAS () [kts], Descent calibrated air speed.

DES_MACH () [-], Descent Mach number.

MASS_REF () [lb-ton], Aircraft reference weight.

MASS_MIN () [lb-ton], Aircraft minimum weight.

MASS_MAX () [lb-ton], Aircraft maximum weight.

MASS_PAYLD () [lb-ton], Aircraft payload weight

MASS_GRAD () [ft], Mass gradient on maximum altitude.

FENV_VMO () [kts], Maximum operational speed.

FENV_MMO () [-], Maximum operational Mach number.

FENV_ALT () [ft], Cruise altitude at MTOW in ISA conditions.

FENV_HMAX () [ft], Maximum altitude at MTOW in ISA conditions.

FENV_TEMP () [ft K^{-1}], Temperature gradient on maximum altitude.

WING_AREA () [ft^2], Aircraft lifting surface area.

COEFF_CLBO () [-], Initial buffet onset lift coefficient for $Mach = 0$.

BUFF_GRAD () [-], Lift coefficient gradient for $Mach = 0$.

VSTALL_TO () [kts], Aircraft stall calibrated air speed at take off.

COEFF_CD0_TO () [-], Parasitic drag coefficient at take off.

COEFF_CD2_TO () [-], Induced drag coefficient at take off.

VSTALL_IC () [kts], Aircraft stall calibrated air speed at initial climb.

COEFF_CD0_IC () [-], Parasitic drag coefficient at initial climb.

COEFF_CD2_IC () [-], Induced drag coefficient at initial climb.

VSTALL_CR () [kts], Aircraft stall calibrated air speed at cruise.

COEFF_CD0_CR () [-], Parasitic drag coefficient at cruise.

COEFF_CD2_CR () [-], Induced drag coefficient at cruise.

VSTALL_AP () [kts], Aircraft stall calibrated air speed at approach.

COEFF_CD0_AP () [-], Parasitic drag coefficient at approach.

COEFF_CD2 AP () [-], Induced drag coefficient at approach

VSTALL_LD () [kts], Aircraft stall calibrated air speed at landing.

COEFF_CD0_LD () [-], Parasitic drag coefficient at landing.

COEFF_CD2_LD () [-], Induced drag coefficient at landing.

CF1 (COEFF_CF1) [kg min^{-1} kN^{-1}], Aircraft specific 1^{st} TSFC coefficient.

CF2 (COEFF_CF2) [kts], Aircraft specific 2^{nd} TSFC coefficient.

CF3 (COEFF_CF3) [kg min^{-1}], Aircraft specific 1^{st} descent fuel flow coefficient.

CF4 (COEFF_CF4) [ft], Aircraft specific 2^{nd} descent fuel flow coefficient.

CFCR (COEFF_CF5) [-], Aircraft specific cruise flue flow correction coefficient.

CL_CAS1_AV () [kt], Standard climb average mission range calibrated airspeed 1.

CL_CAS2_AV () [kt], Standard climb average mission range calibrated airspeed 2.

CL_MACH_AV () [-], Standard climb average mission range Mach number.

CR_CAS1_AV () [kt], Standard cruise average mission range calibrated airspeed 1.

CR_CAS2_AV () [kt], Standard cruise average mission range calibrated airspeed 2.

CR_MACH_AV () [-], Standard cruise average mission range Mach number.

DE_CAS1_AV () [kt], Standard descent average mission range calibrated airspeed 1.

DE_CAS2_AV () [kt], Standard descent average mission range calibrated airspeed 2.

DE_MACH_AV () [-], Standard descent average mission range Mach number.

CL_CAS1_HI () [kt], Standard climb high mission range calibrated airspeed 1.

CL_CAS2_HI () [kt], Standard climb high mission range calibrated airspeed 2.

CL_MACH_HI () [-], Standard climb high mission range Mach number.

CR_CAS1_HI () [kt], Standard cruise high mission range calibrated airspeed 1.

CR_CAS2_HI () [kt], Standard cruise high mission range calibrated airspeed 2.

CR_MACH_HI () [-], Standard cruise high mission range Mach number.

DE_CAS1_HI () [kt], Standard descent high mission range calibrated airspeed 1.

DE_CAS2_HI () [kt], Standard descent high mission range calibrated airspeed 2.

DE_MACH_HI () [-], Standard descent high mission range Mach number.

CL_CAS1_LO () [kt], Standard climb low mission range calibrated airspeed 1.

CL_CAS2_LO () [kt], Standard climb low mission range calibrated airspeed 2.

CL_MACH_LO () [-], Standard climb low mission range Mach number.

CR_CAS1_LO () [kt], Standard cruise low mission range calibrated airspeed 1.

CR_CAS2_LO () [kt], Standard cruise low mission range calibrated airspeed 2.

CR_MACH_LO () [-], Standard cruise low mission range Mach number.

DE_CAS1_LO () [kt], Standard descent low mission range calibrated airspeed 1.

DE_CAS2_LO () [kt], Standard descent low mission range calibrated airspeed 2.

DE_MACH_LO () [-], Standard descent low mission range Mach number.

THR_STATIC () [lbf], Engine static thrust at sea-level and ISA conditions.

MIN_BURN () [kg sec^{-1}], Single engine minimum fuel burn rate.

CE_C (COEFF_E) [lbf], Jet thurst coefficients in climb.

CF_C (COEFF_F) [lbf kt^{-1}], Jet thurst coefficients in climb.

CGA_C (COEFF_GA) [lbf ft^{-1}], Jet thurst coefficients in climb.

CGB_C (COEFF_GB) [lbf ft^{-2}], Jet thurst coefficients in climb.

CE_T (COEFF_E) [lbf], Jet thurst coefficients in takeoff.

CF_T (COEFF_F) [lbf kt^{-1}], Jet thurst coefficients in takeoff.

CGA_T (COEFF_GA) [lbf ft^{-1}], Jet thurst coefficients in takeoff.

CGB_T (COEFF_GB) [lbf ft^{-2}], Jet thurst coefficients in takeoff.

CA_1 (COEFF_1) [kg min^{-1} kN^{-1}], 1^{st} approach TSFC coefficient.

CA_2 (COEFF_2) [kg min^{-1} kN^{-1}], 2^{nd} approach TSFC coefficient.

CA_3 (COEFF_3) [kg min^{-1} kN^{-1} h^{-1}], 3^{rd} approach TSFC coefficient.

CA_4 (COEFF_4) [kg min^{-1} kN^{-2}], 4^{th} approach TSFC coefficient.

CD_1 (COEFF_1) [kg min^{-1} kN^{-1}], 1^{st} descent TSFC coefficient.

CD_2 (COEFF_2) [kg min^{-1} kN^{-1}], 2^{nd} descent TSFC coefficient.

CD_3 (COEFF_3) [kg min^{-1} kN^{-1} h^{-1}], 3^{rd} descent TSFC coefficient.

CD_4 (COEFF_4) [kg min^{-1} kN^{-2}], 4^{th} descent TSFC coefficient.

CR_1 (COEFF_R) [-], Drag-over-lift coefficient, depends on the flaps setting (1).

CC_1 (COEFF_C) [-], Coefficient of Lift, depends on the flaps setting (1).

CB_1 (COEFF_B) [-], Ground-roll coefficient, depends on the flaps setting (1).

CR_2 (COEFF_R) [-], Drag-over-lift coefficient, depends on the flaps setting (2).

CC_2 (COEFF_C) [-], Drag-over-lift coefficient, depends on the flaps setting (2).

CB_2 (COEFF_B) [-], Ground-roll coefficient, depends on the flaps setting (2).

CR_3 (COEFF_R) [-], Drag-over-lift coefficient, depends on the flaps setting (3).

CC_3 (COEFF_C) [-], Drag-over-lift coefficient, depends on the flaps setting (3).

CB_3 (COEFF_B) [-], Ground-roll coefficient, depends on the flaps setting (3).

Bibliography

[1] ICAO 2010a. Report of the committee on aviation environmental protection, eighth meeting. Technical Report CAEP/8-IP/30, Montreal, Canada, February 2010.

[2] ICAO 2010b. 37th session of the ICAO assembly, resolution A37-19: Consolidated statement of continuing ICAO policies and practices related to environmental protection climate change. Technical report, Montreal, Canada, November 2010.

[3] Milton Abramowitz, Irene Stegun, and Donald A McQuarrie. Handbook of mathematical functions. *American Journal of Physics*, 34(2):177–177, 1966.

[4] Joel Achenbach. NASA's trajectory unrealistic, panel says. *Washington Post*, 14, 2009.

[5] Harish Agarwal, John E Renaud, Evan L Preston, and Dhanesh Padmanabhan. Uncertainty quantification using evidence theory in multidisciplinary design optimization. *Reliability Engineering & System Safety*, 85(1):281–294, 2004.

[6] Maarten Arnst, Soize Craig, and Roger Ghanem. Hybrid sampling/spectral method for solving stochastic coupled problems. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):218–243, 2013.

[7] Maarten Arnst, Roger Ghanem, Eric Phipps, and John Red-Horse. Dimension reduction in stochastic modeling of coupled problems. *International Journal for Numerical Methods in Engineering*, 92(11):940–968, 2012.

[8] Maarten Arnst, Roger Ghanem, Eric Phipps, and John Red-Horse. Measure transformation and efficient quadrature in reduced-dimensional stochastic modeling of coupled problems. *International Journal for Numerical Methods in Engineering*, 92(12):1044–1080, 2012.

[9] Maarten Arnst, Roger Ghanem, Eric Phipps, and John Red-Horse. Reduced chaos expansions with random coefficients in reduced-dimensional stochastic modeling of coupled problems. *International Journal for Numerical Methods in Engineering*, 97(5):352–376, 2014.

[10] Ivo Babuška, Fabio Nobile, and Raul Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM review*, 52(2):317–355, 2010.

[11] Volker Barthelmann, Erich Novak, and Klaus Ritter. High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics*, 12(4):273–288, 2000.

[12] Jon Louis Bentley. Multidimensional divide-and-conquer. *Communications of the ACM*, 23(4):214–229, 1980.

[13] JW Berglund, JW Garner, Jon C Helton, Jay D Johnson, and LN Smith. Direct releases to the surface and associated complementary cumulative distribution functions in the 1996 performance assessment for the waste isolation pilot plant: cuttings, cavings and spallings. *Reliability Engineering & System Safety*, 69(1):305–330, 2000.

[14] Patrick Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 2012.

[15] George Biros and Omar Ghattas. Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part I: The Krylov–Schur solver. *SIAM Journal on Scientific Computing*, 27(2):687–713, 2005.

[16] Emanuele Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, 2007.

[17] Emanuele Borgonovo, William Castaings, and Stefano Tarantola. Moment independent importance measures: New results and analytical test cases. *Risk Analysis*, 31(3):404–428, 2011.

[18] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, UK, 2004.

[19] Robert Braun and Ilan Kroo. Development and application of the collaborative optimization architecture in a multidisciplinary. *In: Alexandrov, N., Hussaini, M.Y. (eds.) Multidisciplinary design optimization: state of the art. SIAM Journal of Optimization*, 80:98–116, 1997.

[20] Karen Chan, Andrea Saltelli, and Stefano Tarantola. Winding stairs: a sampling tool to compute sensitivity indices. *Statistics and Computing*, 10(3):187–196, 2000.

[21] Gaëlle Chastaing, Fabrice Gamboa, and Clémentine Prieur. Generalized Hoeffding-Sobol decomposition for dependent variables-application to sensitivity analysis. *Electronic Journal of Statistics*, 6:2420–2448, 2012.

[22] Xiao Chen, Brenda Ng, Yunwei Sun, and Charles Tong. A flexible uncertainty quantification method for linearly coupled multi-physics systems. *Journal of Computational Physics*, 248(1):383–401, 2013.

[23] Xiaoxiao Chen, Eun-Jae Park, and Dongbin Xiu. A flexible numerical approach for quantification of epistemic uncertainty. *Journal of Computational Physics*, 240(1):211–224, 2013.

[24] Anukal Chiralaksanakul and Sankaran Mahadevan. Decoupled approach to multidisciplinary design optimization under uncertainty. *Optimization and Engineering*, 8(1):21–42, 2007.

[25] Moon-Hyun Chun, Seok-Jung Han, and Nam-IL Tak. An uncertainty importance measure using a distance metric for the change in a cumulative distribution function. *Reliability Engineering & System Safety*, 70(3):313–321, 2000.

[26] The Boeing Company. Boeing: Airport compatibility - CAD 3-view drawings for airport planning purposes. Boeing.com [Online; posted 15-May-2015].

[27] Paul Constantine, Eric Phipps, and Timothy Wildey. Efficient uncertainty propagation for network multiphysics systems. *International Journal for Numerical Methods in Engineering*, 2014.

[28] Luis G Crespo, Sean P Kenny, and Daniel P Giesy. The NASA Langley multidisciplinary uncertainty quantification challenge, 2013.

[29] N Cumpsty, JJ Alonso, S Eury, L Maurice, B Nas, M Ralph, and R Sawyer. Report of the independent experts on fuel burn reduction technology goals. In *International Civil Aviation Organization (ICAO), Committee on Aviation Environmental Protection (CAEP), Steering Group Meeting*, 2010.

[30] Yu-Hong Dai and Roger Fletcher. Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming. *Numerische Mathematik*, 100(1):21–47, 2005.

[31] Yu-Hong Dai and Roger Fletcher. New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. *Mathematical Programming*, 106(3):403–421, 2006.

[32] Luc Devroye and Clark S Penrod. The strong uniform convergence of multivariate variable kernel estimates. *Canadian Journal of Statistics*, 14(3):211–220, 1986.

[33] Josef Dick and Friedrich Pillichshammer. *Digital nets and sequences: discrepancy theory and quasi-Monte Carlo integration.* Cambridge University Press, New York, NY, USA, 2010.

[34] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential Monte Carlo sampling methods for bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.

[35] Norman Richard Draper, Harry Smith, and Elizabeth Pownell. *Applied regression analysis*, volume 3. Wiley New York, 1966.

[36] Mark Drela. Simultaneous optimization of the airframe, powerplant, and operation of transport aircraft. *Hamilton Place, London*, 2010.

[37] Christopher Drew. A dream interrupted at Boeing. *The New York Times*, 5, 2009.

[38] Xiaoping Du and Wei Chen. Collaborative reliability analysis under the framework of multidisciplinary systems design. *Optimization and Engineering*, 6(1):63–84, 2005.

[39] David H Evans. An application of numerical integration techniques to statistical tolerancing, III-general distributions. *Technometrics*, 14(1):23–35, 1972.

[40] Michael Evans and Tim Swartz. Methods for approximating integrals in statistics with special emphasis on bayesian integration problems. *Statistical Science*, pages 254–272, 1995.

[41] Bernd Fiessler, Rudiger Rackwitz, and Hans-Joachim Neumann. Quadratic limit states in structural reliability. *Journal of the Engineering Mechanics Division*, 105(4):661–676, 1979.

[42] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956.

[43] Roger Ghanem. Hybrid stochastic finite elements and generalized Monte Carlo simulation. *Transctions-American Society of Mechanical Engineers Journal of Applied Mechanics*, 65:1004–1009, 1998.

[44] Roger Ghanem and Pol Spanos. *Stochastic finite elements: a spectral approach.* Dover publications, 2003.

[45] Alison L Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002.

[46] Mark Girolami and Chao He. Probability density estimation from optimally condensed data samples. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(10):1253–1264, 2003.

[47] Arthur Gretton, Alex Smola, Jiayuan Huang, Marcel Schmittfull, Karsten Borgwardt, and Bernhard Schölkopf. Covariate shift by kernel mean matching. *Dataset Shift in Machine Learning*, 3(4):1–38, 2009.

[48] Mircea Grigoriu. *Stochastic systems: Uncertainty quantification and propagation.* Springer Science & Business Media, 2012.

[49] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes.* Texts from Oxford University Press. Oxford University Press, 2001.

[50] Xiaoyu Stacey Gu, John Renaud, and Charles Penninger. Implicit uncertainty propagation for robust collaborative optimization. *Journal of Mechanical Design*, 128(4):1001–1013, 2006.

[51] Peter Hall, Simon Sheather, MC Jones, and JS Marron. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, 78(2):263–269, 1991.

[52] Abraham M Hasofer and Niels C Lind. Exact and invariant second-moment code format. *Journal of the Engineering Mechanics division*, 100(1):111–121, 1974.

[53] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*, volume 2. Springer, 2009.

[54] Stefan Heinrich. Efficient algorithms for computing the l_2-discrepancy. *Mathematics of Computation of the American Mathematical Society*, 65(216):1621–1633, 1996.

[55] JC Helton and FJ Davis. Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Analysis*, 22(3):591–622, 2002.

[56] Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, pages 293–325, 1948.

[57] Giles Hooker. Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables. *Journal of Computational and Graphical Statistics*, 16(3), 2007.

[58] Jiayuan Huang, Alexander J Smola, Arthur Gretton, Karsten M Borgwardt, and Bernhard Schölkopf. Correcting sample selection bias by unlabeled data. *Advances in Neural Information Processing Systems*, 19:601–608, 2007.

[59] John Jakeman, Michael Eldred, and Dongbin Xiu. Numerical approach for quantification of epistemic uncertainty. *Journal of Computational Physics*, 229(12):4648–4663, 2010.

[60] Takafumi Kanamori, Shohei Hido, and Masashi Sugiyama. A least-squares approach to direct importance estimation. *The Journal of Machine Learning Research*, 10:1391–1445, 2009.

[61] Hyung Min Kim, Nestor Michelena, Panos Papalambros, and Tao Jiang. Target cascading in optimal system design. *Journal of Mechanical Design*, 125(3):474–480, 2003.

[62] Michael Kokkolaras, Zissimos P Mourelatos, and Panos Y Papalambros. Design optimization of hierarchically decomposed multilevel systems under uncertainty. *Journal of Mechanical Design*, 128(2):503–508, 2006.

[63] Augustine Kong, Jun Liu, and Wing Wong. Sequential imputations and bayesian missing data problems. *Journal of the American statistical association*, 89(425):278–288, 1994.

[64] Ilan Kroo. Distributed multidisciplinary design and collaborative optimization. In *VKI lecture series on Optimization Methods & Tools for Multicriteria/Multidisciplinary Design*, number November 15–19, pages 1–22, 2004.

[65] Sergei Kucherenko, Miguel Munoz Zuniga, Stefano Tarantola, and Paola Annoni. Metamodelling and global sensitivity analysis of models with dependent variables. In *Numerical Analysis and Applied Mathmatics ICNAAM 2011: International Conference on Numerical Analysis and Applied Mathematics*, volume 1389, pages 1913–1916. AIP Publishing, 2011.

[66] Frances Kuo, Ilan Sloan, Grzegorz Wasilkowski, and Henryk Woźniakowski. On decompositions of multivariate functions. *Mathematics of computation*, 79(270):953–966, 2010.

[67] Charles L Lawson and Richard J Hanson. *Solving least squares problems*, volume 161. SIAM, 1974.

[68] Olivier Le Maître and Omar Knio. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer, 2010.

[69] David S Lee, David W Fahey, Piers M Forster, Peter J Newton, Ron CN Wit, Ling L Lim, Bethan Owen, and Robert Sausen. Aviation and global climate change in the 21st century. *Atmospheric Environment*, 43(22):3520–3537, 2009.

[70] Russell Lenth. Some practical guidelines for effective sample size determination. *The American Statistician*, 55(3):187–193, 2001.

[71] Genyuan Li and Herschel Rabitz. General formulation of HDMR component functions with independent and correlated variables. *Journal of Mathematical Chemistry*, 50(1):99–130, 2012.

[72] Chih-Jen Lin, Stefano Lucidi, Laura Palagi, Arnaldo Risi, and Marco Sciandrone. Decomposition algorithm model for singly linearly-constrained problems subject to lower and upper bounds. *Journal of Optimization Theory and Applications*, 141(1):107–126, 2009.

[73] Jun Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119, 1996.

[74] Yu Liu, Xiaolei Yin, Paul Arendt, Wei Chen, and Hong-Zhong Huang. A hierarchical statistical sensitivity analysis method for multilevel systems with shared variables. *Journal of Mechanical Design*, 132(3):031006, 2010.

[75] Sankaran Mahadevan and Natasha Smith. Efficient first-order reliability analysis of multidisciplinary systems. *International Journal of Reliability and Safety*, 1(1):137–154, 2006.

[76] Thierry A Mara and Stefano Tarantola. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107:115–121, 2012.

[77] Jay Martin and Timothy Simpson. A methodology to manage system-level uncertainty during conceptual design. *Journal of Mechanical Design*, 128(4):959–968, 2006.

[78] Youssef Marzouk and Karen Willcox. Uncertainty quantification. In Nicholas Higham, editor, *The Princeton Companion to Applied Mathematics*. Princeton University Press, 2015.

[79] Jiří Matoušek. On the L2–discrepancy for anchored boxes. *Journal of Complexity*, 14(4):527–556, 1998.

[80] Hermann G Matthies and Andreas Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 194(12):1295–1331, 2005.

[81] Mark McDonald and Sankaran Mahadevan. Uncertainty quantification and propagation for multidisciplinary system analysis. In *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Paper 2008-6038*, 2008.

[82] Sherif Mohamed and Alison K McCowan. Modelling project investment decisions under uncertainty using possibility theory. *International Journal of Project Management*, 19(4):231–241, 2001.

[83] Ramon Moore. *Methods and applications of interval analysis*, volume 2. SIAM, 1979.

[84] Max D Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[85] John Nickolls, Ian Buck, Michael Garland, and Kevin Skadron. Scalable parallel programming with CUDA. *Queue*, 6(2):40–53, 2008.

[86] Harald Niederreiter. Quasi-Monte Carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society*, 84(6):957–1041, 1978.

[87] Fabio Nobile, Raúl Tempone, and Clayton G Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.

[88] George Noel, Doug Allaire, Stuart Jacobson, Karen Willcox, and Rebecca Cointin. Assessment of the aviation environmental design tool. In *Proceedings of the Eighth USA/Europe Air Traffic Management Research and Development Seminar (ATM2009), Napa, CA, June*, volume 29, 2009.

[89] Erich Novak and Henryk Wozniakowski. L_2-Discrepancy and multivariate integration. *Analytic Number Theory: Essays in Honour of Klaus Roth*, pages 359–388, 2009.

[90] A Nuic. User manual for the base of aircraft data (bada) revision 3.10. *Atmosphere*, 2010:001, 2010.

[91] Emanuel Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[92] John Platt et al. Fast training of support vector machines using sequential minimal optimization. *Advances in kernel methodsâĂŤsupport vector learning*, 3, 1999.

[93] Jing Qin. Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*, 85(3):619–630, 1998.

[94] Herschel Rabitz and Ömer F Aliş. General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, 25(2-3):197–233, 1999.

[95] Sharif Rahman. A generalized ANOVA dimensional decomposition for dependent probability measures. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):670–697, 2014.

[96] R.C. Reed. *The Superalloys: Fundamentals and Applications*. Cambridge University Press, 2006.

[97] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, 1 edition, 1999.

[98] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[99] Christopher Roof, Andrew Hansen, Gregg Fleming, Ted Thrasher, Alex Nguyen, Cliff Hall, F Grandi, B Kim, S Usdrowski, and P Hollingsworth. Aviation environmental design tool (AEDT) system architecture. *Doc# AEDT-AD-01*, 2007.

[100] Walter Rudin. *Real and complex analysis, 3rd Ed.* McGraw-Hill, Inc., New York, NY, USA, 1987.

[101] Daniel Rutherford and Mazyar Zeinali. Efficiency trends for new commercial jet aircraft 1960 to 2008. *International Council on Clean Transportation.*

[102] Andrea Saltelli and Ricardo Bolado. An alternative way to compute fourier amplitude sensitivity test (fast). *Computational Statistics & Data Analysis*, 26(4):445–460, 1998.

[103] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global sensitivity analysis: the primer.* John Wiley & Sons, 2008.

[104] Shankar Sankararaman and Sankaran Mahadevan. Likelihood-based approach to multidisciplinary analysis under uncertainty. *Journal of Mechanical Design*, 134(3):031008, 1–12, 2012.

[105] Adrian Schofield. End of year is new goal for 787 first flight. *Aviation Week*, 27, 2009.

[106] David Scott. *Multivariate density estimation: theory, practice, and visualization.* A Wiley-Interscience publication. Wiley, 1992.

[107] Timothy W Simpson and Joaquim RRA Martins. Multidisciplinary design optimization for complex engineered systems: report from a national science foundation workshop. *Journal of Mechanical Design*, 133(10):101002, 2011.

[108] Adrian Smith and Alan Gelfand. Bayesian statistics without tears: a sampling–resampling perspective. *The American Statistician*, 46(2):84–88, 1992.

[109] Ralph C Smith. *Uncertainty quantification: theory, implementation, and applications*, volume 12. SIAM, 2013.

[110] Jaroslaw Sobieszczanski-Sobieski, Jeremy Agte, and Robert Sandusky. Bilevel integrated system synthesis. *AIAA Journal*, 38(1):164–172, 2000.

[111] Ilya M Sobol'. Multidimensional quadrature formulas and Haar functions. *Izdat Nauka, Moscow*, 1969.

[112] Ilya M Sobol'. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2(1):112–118, 1990.

[113] Ilya M Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280, 2001.

[114] Ilya M Sobol'. Theorems and examples on high dimensional model representation. *Reliability Engineering & System Safety*, 79(2):187–193, 2003.

[115] Christian Soize and Roger Ghanem. Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM Journal on Scientific Computing*, 26(2):395–410, 2004.

[116] Masashi Sugiyama, Taiji Suzuki, and Takafumi Kanamori. *Density ratio estimation in machine learning.* Cambridge University Press, Cambridge, UK, 2012.

[117] Masashi Sugiyama, Makoto Yamada, Paul Von Buenau, Taiji Suzuki, Takafumi Kanamori, and Motoaki Kawanabe. Direct density-ratio estimation with dimensionality reduction via least-squares hetero-distributional subspace search. *Neural Networks*, 24(2):183–198, 2011.

[118] Palle Thoft-Christensen and Michael J Baker. Structural reliability theory and its applications. 1982.

[119] Peter Valdes-Dapena. Chevy volt to get 230 mpg rating. CNNMoney.com [Online; posted 09-Aug-2009].

[120] Vladimir Vapnik. *Statistical learning theory*, volume 2. Wiley New York, New York, NY, USA, 1998.

[121] Vladimir Vapnik, Igor Braga, and Rauf Izmailov. A constructive setting for the problem of density ratio estimation. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 434–442, 2014.

[122] George Visick. A quantitative version of the observation that the Hadamard product is a principal submatrix of the Kronecker product. *Linear Algebra and Its Applications*, 304(1):45–68, 2000.

[123] Tony Warnock. *Computational investigations of low-discrepancy point sets, applications of number theory to numerical analysis (S. K. Zaremba, ed.).* Academic Press, 1972.

[124] Jon A Wellner. Empirical processes in action: a review. *International Statistical Review/Revue Internationale de Statistique*, pages 247–269, 1992.

[125] Dongbin Xiu. Efficient collocational approach for parametric uncertainty analysis. *Communications in computational physics*, 2(2):293–309, 2007.

[126] Dongbin Xiu. *Numerical Methods for Stochastic Computations: a spectral method approach.* Princeton University Press, 2010.

[127] Dongbin Xiu and Jan S Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, 2005.

[128] S Yakowitz, JE Krimmel, and F Szidarovszky. Weighted Monte Carlo integration. *SIAM Journal on Numerical Analysis*, 15(6):1289–1300, 1978.

[129] Wen Yao, Xiaoqian Chen, Wencai Luo, Michel van Tooren, and Jian Guo. Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles. *Progress in Aerospace Sciences*, 47(6):450–479, 2011.

[130] Xiaolei Yin and Wei Chen. A hierarchical statistical sensitivity analysis method for complex engineering systems design. *Journal of Mechanical Design*, 130(7):071402, 2008.

[131] Luca Zanni. An improved gradient projection-based decomposition technique for support vector machines. *Computational Management Science*, 3(2):131–145, 2006.

[132] Luca Zanni, Thomas Serafini, and Gaetano Zanghirati. Parallel software for training large scale support vector machines on multiprocessor systems. *The Journal of Machine Learning Research*, 7:1467–1492, 2006.