

Optimal Approximations of Coupling in Multidisciplinary Models

by

Ricardo Miguel Santos Baptista

B.A.Sc., University of Toronto (2015)

Submitted to the School of Engineering
in partial fulfillment of the requirements for the degree of
Master of Science in Computation for Design and Optimization

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

© Massachusetts Institute of Technology 2017. All rights reserved.

Signature redacted

Author
School of Engineering

Signature redacted May 24, 2017

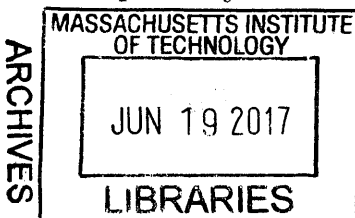
Certified by.....
Youssef Marzouk
Associate Professor of Aeronautics and Astronautics
Thesis Supervisor

Signature redacted

Certified by..
Karen Willcox
Professor of Aeronautics and Astronautics
Thesis Supervisor

Signature redacted

Accepted by.....
Nicolas Hadjicostantinou
Director, Computation for Design and Optimization



Optimal Approximations of Coupling in Multidisciplinary Models

by

Ricardo Miguel Santos Baptista

Submitted to the School of Engineering
on May 25, 2017, in partial fulfillment of the
requirements for the degree of
Master of Science in Computation for Design and Optimization

Abstract

Design of complex engineering systems requires coupled analyses of the multiple disciplines affecting system performance. The coupling among disciplines typically contributes significantly to the computational cost of analyzing a system, and can become particularly burdensome when coupled analyses are embedded within a design or optimization loop. In many cases, disciplines may be weakly coupled, so that some of the coupling or interaction terms can be neglected without significantly impacting the accuracy of the system output. However, typical practice derives such approximations in an ad hoc manner using expert opinion and domain experience. In this thesis, we propose a new approach that formulates an optimization problem to find a model that optimally balances accuracy of the model outputs with the sparsity of the discipline couplings. An adaptive sequential Monte Carlo sampling-based technique is used to efficiently search the combinatorial model space of different discipline couplings. Finally, an algorithm for optimal model selection is presented and combined with three tractable approaches to quantify the accuracy of the system outputs with approximate couplings. These algorithms are applied to identify the important discipline couplings in three engineering problems: a fire detection satellite model, a turbine engine cycle analysis model, and a lifting surface aero-structural model.

Thesis Supervisor: Youssef Marzouk

Title: Associate Professor of Aeronautics and Astronautics

Thesis Supervisor: Karen Willcox

Title: Professor of Aeronautics and Astronautics

Acknowledgments

I would first and foremost like to thank Professors Karen Willcox and Youssef Marzouk for accepting me as their student and providing me with this opportunity. I am extremely grateful for their support, words of encouragement, and guidance on directions to pursue with this study. Their advice has provided me with the foundations for this thesis project and motivated my interest for future research in this field.

I would also like to thank the many people in the ACDL who provided excellent support over the past two years. I want to especially recognize Sergio Amaral, Anirban Chaudhuri, Alex Feldstein, Rémi Lam, Rebecca Morrison, Benjamin Peherstorfer, Elizabeth Qian, Alessio Spantini, Ben Zhang, and Olivier Zahm for entertaining my endless questions and for the many fruitful and engaging discussions. I would also like to thank those at MIT outside of the ACDL including Matthew Li and Adam Rosenfield, and my dear friends and family back home for their continual support.

I am grateful for the financial support from the Air Force Office of Scientific Research (AFOSR) MURI on "Managing multiple information sources of multi-physics systems," Program Officer Jean-Luc Cambier, Award Number FA9550-15-1-0038.

Finally, I could not be where I am today without the help and guidance of my family. I dedicate this thesis to my parents who have always supported all of my academic endeavors. I love you both.

Contents

1	Introduction	19
1.1	Coupling in Multidisciplinary Models	19
1.2	Thesis Contributions	21
2	Background and Problem Setup	23
2.1	Multidisciplinary Models	23
2.2	Uncertainty Quantification for MDAO	25
2.3	Model Coupling Approximations	26
2.3.1	Decoupling Discipline Inputs	27
2.3.2	Measures of Statistical Distance	28
2.3.3	Search for Optimal Coupling	29
2.4	Combinatorial Optimization	30
2.4.1	Sequential Monte Carlo	30
2.4.2	Model Selection Procedure	34
3	Identifying Approximate Coupling	37
3.1	Model-Linearization Approach	37
3.1.1	Output Uncertainty in Linearized Model	38
3.1.2	Computational Complexity	41
3.2	Improvements to Model Output Uncertainty	43
3.2.1	Higher-Order Approximations	43
3.2.2	Control Variate Corrections	44
3.3	Extension to Bayesian Approach	46

3.3.1	Posterior over Model Space	47
3.3.2	Approximate Bayesian Computation	48
4	Numerical Results	51
4.1	Fire Detection Satellite Model	51
4.1.1	Nominal Input Uncertainty	51
4.1.2	Different Input Uncertainty	59
4.1.3	Control Variate Corrections	63
4.2	Turbine Engine Model	68
4.3	Aero-Structural Lifting Surface Model	75
5	Analysis of Approximate Coupling	85
5.1	Linear Systems	85
5.2	Solution Uncertainty	86
5.3	Feed-forward Systems	87
5.4	Feedback Systems	88
5.5	Connections to Probabilistic Graphical Models	91
5.5.1	Decoupling and Conditional Independence	91
5.5.2	Gaussian Distributions	94
5.5.3	Exponential Family Distributions	97
5.5.4	More General Distributions	100
6	Conclusions and Future Work	103
6.1	Conclusions	103
6.2	Future Work	105
6.2.1	Improved Optimization Objective	105
6.2.2	Preconditioning with Decoupled Models	107
6.2.3	Further Connections to Probabilistic Graphical Models	109

List of Figures

2-1	Graphical representation of a multidisciplinary model	24
2-2	Feedback coupled two discipline model, M_0	28
2-3	One-way coupled two discipline model, M	28
4-1	Fire detection satellite model from Ref. [51]	52
4-2	Joint distribution of the reference model	52
4-3	Comparison of Gaussian approximation and Monte Carlo samples	53
4-4	Optimal model coupling for $\lambda = 1.0 \times 10^{-4}$	55
4-5	Joint distribution for $\lambda = 1.0 \times 10^{-4}$	55
4-6	Optimal model coupling for $\lambda = 1.0 \times 10^{-3}$	55
4-7	Joint distribution for $\lambda = 1.0 \times 10^{-3}$	55
4-8	Optimal model coupling for $\lambda = 1.0 \times 10^{-1}$	55
4-9	Joint distribution for $\lambda = 1.0 \times 10^{-1}$	55
4-10	Optimal model coupling for $\lambda = 1.0 \times 10^0$	56
4-11	Joint distribution for $\lambda = 1.0 \times 10^0$	56
4-12	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-4}$	57
4-13	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-3}$	58
4-14	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-1}$	58
4-15	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^0$	59
4-16	Gaussian approximation and Monte Carlo samples with different inputs	60
4-17	Optimal model coupling for $\lambda = 1.0 \times 10^{-4}$	61
4-18	Joint distribution for $\lambda = 1.0 \times 10^{-4}$	61
4-19	Optimal model coupling for $\lambda = 1.0 \times 10^{-3}$	61

4-20	Joint distribution for $\lambda = 1.0 \times 10^{-3}$	61
4-21	Optimal model coupling for $\lambda = 10^{-4}$ to 10^{-1}	62
4-22	Optimal model coupling for $\lambda = 10^0$ to 10^1	62
4-23	Correlation of nonlinear and linearized models	64
4-24	Control variate estimates with increasing sample size	64
4-25	Linearized Gaussian approximation and control variate estimates in comparison to Monte Carlo samples	64
4-26	Optimal model coupling for $\lambda = 1.0 \times 10^{-6}$	65
4-27	Joint distribution for $\lambda = 1.0 \times 10^{-6}$	65
4-28	Optimal model coupling for $\lambda = 1.0 \times 10^{-5}$	65
4-29	Joint distribution for $\lambda = 1.0 \times 10^{-5}$	65
4-30	Control variate estimates for variance of τ_{tot}	66
4-31	Control variate estimates for variance of P_{tot}	66
4-32	Optimal model coupling for $\lambda = 1.0 \times 10^{-6}$	67
4-33	Optimal model coupling for $\lambda = 1.0 \times 10^{-5}$	67
4-34	Turbine engine cycle model from Ref. [27]	68
4-35	Joint distribution of the reference model	69
4-36	Optimal model coupling for $\lambda = 1.0 \times 10^{-1}$	70
4-37	Optimal model coupling for $\lambda = 1.0 \times 10^0$	70
4-38	Optimal model coupling for $\lambda = 1.0 \times 10^1$	71
4-39	Optimal model coupling for $\lambda = 1.0 \times 10^3$	71
4-40	Joint distribution for $\lambda = 1.0 \times 10^{-1}$	72
4-41	Joint distribution for $\lambda = 1.0 \times 10^0$	72
4-42	Joint distribution for $\lambda = 1.0 \times 10^1$	72
4-43	Joint distribution for $\lambda = 1.0 \times 10^3$	72
4-44	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-1}$	74
4-45	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^0$	74
4-46	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^1$	75
4-47	Wing planform from Ref. [29]	76
4-48	Aero-structural lifting surface model from Ref. [29]	76

4-49	Spatial locations for coupling variables in the aero-structural model	77
4-50	Joint distribution of the reference model	78
4-51	Optimal model coupling for $\lambda = 1.0 \times 10^{-6}$	78
4-52	Joint distribution for $\lambda = 1.0 \times 10^{-6}$	78
4-53	Optimal model coupling for $\lambda = 1.0 \times 10^{-3}$	79
4-54	Joint distribution for $\lambda = 1.0 \times 10^{-3}$	79
4-55	Optimal model coupling for $\lambda = 1.0 \times 10^{-1}$	79
4-56	Joint distribution for $\lambda = 1.0 \times 10^{-1}$	79
4-57	Optimal model coupling for $\lambda = 1.0 \times 10^1$	79
4-58	Joint distribution for $\lambda = 1.0 \times 10^1$	79
4-59	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-6}$	81
4-60	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-3}$	81
4-61	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-1}$	82
4-62	Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^1$	82
4-63	Probability distribution of chordwise loads	83
4-64	Probability distribution of spanwise loads	83
6-1	Model with only feed-forward connections and local feedback loop after decoupling disciplines 1 and 4	106
6-2	Decoupled model with parallel structure—disciplines 2, 3 and 4 com- pute feedback to discipline 1 independently and run in parallel	106

List of Tables

4.1	Random variable parameters in the satellite model from Ref. [51]	53
4.2	Linearized KL divergence for decoupled satellite models with nominal inputs	56
4.3	Linearized KL divergence for decoupled satellite models with different inputs	62
4.4	Linearized KL divergence with sub-optimal decoupled models	62
4.5	KL divergence estimates for decoupled satellite models based on control variate estimators for the mean and variance	67
4.6	Random variable parameters in the turbine engine cycle model	69
4.7	Linearized KL divergence for decoupled turbine engine models	73
4.8	Random variable parameters in the aero-structural lifting surface model	77
4.9	Linearized KL divergence for decoupled aero-structural models	80

Nomenclature

\mathbf{f}	Model output variables (QoI)
\mathbf{x}	Model input variables
\mathbf{y}	Model state variables
\mathcal{F}	Function for output variables
\mathcal{R}_i	Residual function for state variable i
$\pi_{\mathbf{f}}$	Probability density function of outputs
$\pi_{\mathbf{x}}$	Probability density function of inputs
m	Number of input variables
n	Number of state variables
p	Number of output variables
Λ	Vector of tuning parameters
λ	Tuning parameter in combinatorial optimization
\mathcal{M}	Model space
$\mathcal{P}(M)$	Function to count removed couplings
\mathcal{S}_k	Input state variables to discipline k
d	Number of discipline couplings
D_{KL}	Kullback-Leibler divergence
K	Number of disciplines
M	Discipline couplings in the decoupled model
M_0	Discipline couplings in the reference model
δ	Threshold for minimum sample diversity
$\rho(t)$	Tempering parameter at step t
ζ_t	Particle diversity at step t

$h(M)$	Model objective function
L	Number of particles in SMC
$P_t(M)$	Probability of model M at step t
$w_t^{(l)}$	Weight of model l at step t
$\hat{\mathbf{f}}_N$	Control variate estimator
$\mu_{\mathbf{f}}$	Mean of model outputs
$\mu_{\mathbf{x}}$	Mean of model inputs
$\Sigma_{\mathbf{f}}$	Covariance matrix of model outputs
$\tilde{\mathbf{f}}$	Linearization of model outputs
c	Cost of solving nonlinear model
N	Control variate sample size
ϵ	Tolerance for data statistics
η	Summary statistics for data
$Q(M)$	Model prior in Bayesian approach
ν	Distance function for summary statistics
D	Collection of data from reference model
$\Delta t_{eclipse}$	Satellite: Eclipse period
Δt_{orbit}	Satellite: Orbit period
τ_{tot}	Satellite: Total torque
θ_{slew}	Satellite: Maximum slewing angle
A_{sa}	Satellite: Area of the solar array
C_d	Satellite: Drag coefficient
F_s	Satellite: Average solar flux
H	Satellite: Altitude
H	Satellite: Deviation of moment axis
I	Satellite: Moment of inertia
L_a	Satellite: Aerodynamic torque moment arm
L_{sp}	Satellite: Radiation torque moment arm
P_{ACS}	Satellite: Power of attitude control system
P_{other}	Satellite: Power other than P_{ACS}

P_{tot}	Satellite: Total power
q	Satellite: Reflectance factor
R_D	Satellite: Residual dipole of spacecraft
v	Satellite: Satellite velocity
BPR	Turbine: Bypass ratio
CPR	Turbine: Compressor pressure ratio
F_n	Turbine: Net thrust
FPR	Turbine: Fan pressure ratio
OPR	Turbine: Overall pressure ratio
T_4	Turbine: Engine burner temperature
$TSFC$	Turbine: Thrust-specific fuel consumption
W	Turbine: Mass-flow rate
α	Aero-structural: angle of attack
ρ	Aero-structural: air density
a	Aero-structural: Speed of sounds
M_a	Aero-structural: Mach number
Δ	Matrix perturbation to inverse covariance
Γ	Original inverse covariance matrix
γ	Lagrange multipliers
\mathbf{A}	Linear model matrix
\mathbf{b}	Linear model random inputs
\mathbf{e}_i	Euclidean unit vector
\mathcal{E}	Edge set in original graph, \mathcal{G}
\mathcal{J}	Set of decoupled coupling variables
Φ	Log-partition function
$\phi(\mathbf{y})$	Sufficient statistics in exponential family density
Π	Constrained set of Gaussian distributions
θ	Parameters in exponential family density
$\tilde{\Gamma}$	Perturbed inverse covariance matrix
$\tilde{\mathcal{E}}$	Edge set in sparser graph, $\tilde{\mathcal{G}}$

Chapter 1

Introduction

We begin this thesis by introducing the field of multidisciplinary modeling and the computational challenges of working with coupled systems. After presenting our motivation for this work, we describe our contributions in Section 1.2.

1.1 Coupling in Multidisciplinary Models

Multidisciplinary analysis and optimization (MDAO) couples multiple computational models to represent complex interactions in the design of engineering systems. With the increasing number of disciplines and improved fidelity in multidisciplinary models, the coupling among disciplines can contribute significantly to the computational cost of analyzing these systems. This coupling may include both one-directional (feed-forward) coupling, and bi-directional (feedback) coupling that requires iterative numerical methods to compute a model solution. The coupled MDAO problem may be formulated in a variety of different ways, and various MDAO architectures have been developed to manage discipline coupling for large-scale problems (see Ref. [46] for an overview). Monolithic architectures solve the system using a single optimization problem, while distributed approaches partition the model into subproblems, each involving a smaller number of variables [23, 10, 54, 57, 35]. Typical practice derives these discipline couplings using expert opinion and domain experience. Given the significant impact of couplings on the computational tractability of evaluating a

model, it is of interest to systematically identify which disciplines should be coupled in a model and where couplings may be neglected. This thesis addresses this open challenge to yield optimal approximations to coupling in multidisciplinary models.

The application of MDAO originated in structural optimization and aircraft wing design [26, 58]. It has since been extended to many different engineering systems, such as the design of complete aircraft configurations [37, 43], internal combustion engines [47], wind turbines [22], and spacecraft [12]. Multidisciplinary models of such systems often demonstrate varying degrees of coupling. Couplings can be categorized as being "strong" or "weak" based on the response of a discipline output to a change in a coupling variable [34]. As a result, for certain quantities of interest (QoI) in a model, neglecting weak interactions negligibly impacts the accuracy of the system outputs. Simultaneously, decoupling discipline iterations reduces the number of feedback loops and presents a substantial computational savings when using the model for design optimization and/or uncertainty propagation. Therefore, an important challenge in the field of MDAO has been to identify the decoupled model that best trades the sparsity of the discipline couplings with the accuracy of the model in representing the engineering system. One approach to do this uses weighted design structure matrices to decompose the model and rearrange the disciplines to minimize feedback. This method estimates the strength of couplings based on the discipline connectivity [60] or the sensitivity of the model outputs to the coupling at each iteration of a multidisciplinary optimization process [40]. However, existing methods leave open the question of identifying which discipline couplings are most important in a model for characterizing system outputs over a range of input variables.

In this thesis we consider optimal approximations to coupling in the context of multidisciplinary models under uncertainty. Using these models for optimization or uncertainty quantification (UQ) is challenging, as it requires accurately characterizing the output of these nonlinear models. With complex coupling between the disciplines, users often rely on traditional and computationally intensive approaches for analyzing these models such as Monte Carlo sampling-based methods [33]. As a result, identifying an approximate coupling that reduces the dimensionality of the

information exchanged between disciplines and eliminates large feedback loops, while minimizing information loss in the output uncertainty, presents an important source of computational savings for MDAO. This thesis addresses this challenge and to the best of our knowledge is the first attempt in the literature to identify a sequence of optimal discipline couplings that trade-off the accuracy of the model outputs with the sparsity of the discipline couplings.

1.2 Thesis Contributions

With the goal of identifying optimal couplings for multidisciplinary models under uncertainty, the contributions of this thesis are:

1. Developing tractable methods for quantifying the information loss from neglecting discipline couplings in nonlinear multidisciplinary models under uncertainty with increasing degrees of accuracy
2. Formulating an optimization problem and providing an algorithm to efficiently search over the combinatorial space of possible discipline couplings
3. Applying the decoupling methodology to engineering examples and showing how the systematic approach mathematically identifies decoupling strategies that reflect the underlying system physics
4. Analyzing the effect of decoupling on the output uncertainty by analytically identifying the sources of error and requirements for the well-posedness of the decoupled models

The rest of this thesis is organized as follows. In Chapter 2 we introduce the mathematical background of multidisciplinary models and uncertainty quantification, and formulate the problem of efficiently searching for an optimal approximation to a model's coupling. In Chapter 3 we propose three numerical methods and algorithms for identifying the model approximation(s), and in Chapter 4 we demonstrate the application of these algorithms on three engineering examples. In Chapter 5 we

analyze decoupling in the context of linear models. We provide conditions when it is possible to decouple two disciplines, and quantify the resulting error in the model outputs from this process by making connections to the field of probabilistic graphical models. We conclude the thesis in Chapter 6 with some remarks and future research directions.

Chapter 2

Background and Problem Setup

In this chapter, we introduce the notation and definitions used for multidisciplinary models. After presenting the notion of decoupling, we formulate the problem of searching for an optimal model approximation in Section 2.3. In Section 2.4 we present an algorithm based on sequential Monte Carlo for solving the resulting combinatorial optimization problem.

2.1 Multidisciplinary Models

Numerical multidisciplinary models consist of a set of equations that capture the relationships between input and output variables, mediated by the internal state variables of each discipline. While input variables are independent and defined externally by a user, output variables are quantities of interest (QoI) that depend on the input variables and the state variables of the model.

In this thesis, $\mathbf{x} = [x_1, \dots, x_m]^T \in \mathcal{X} \subseteq \mathbb{R}^m$ is a vector of m input variables, $\mathbf{y} = [y_1, \dots, y_n]^T \in \mathcal{Y} \subseteq \mathbb{R}^n$ is a vector representing the n state variables, and $\mathbf{f} = [f_1, \dots, f_p]^T \in \mathbb{R}^p$ is a vector of p output variables. Each state variable in the model, y_i , is defined implicitly by a governing residual equation

$$\mathcal{R}_i(\mathbf{x}, \mathbf{y}) = 0, \tag{2.1}$$

where in most generality the function \mathcal{R}_i depends on all input and state variables, and consists of algebraic equations, differential equations, and/or other operations. The function $\mathcal{F}: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^p$ computes the output variables explicitly as

$$\mathbf{f} = \mathcal{F}(\mathbf{x}, \mathbf{y}). \quad (2.2)$$

Letting $\mathcal{R} = [\mathcal{R}_1, \dots, \mathcal{R}_n]^T$, a numerical model consists of the coupled system of equations, $\mathcal{R}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$, where $\mathcal{R}: \mathcal{X} \rightarrow \mathcal{Y}$. In a multidisciplinary problem, the n residual equations of the form in equation (2.1) are partitioned into disjoint groups representing each particular engineering discipline or subsystem of the model. These disciplines also define a partitioning of the vector of state variables, \mathbf{y} . This vector consists of state variables that are local to each discipline, and coupling variables that affect the residual equations in different disciplines. A graphical representation of these coupling variables in a model with two disciplines is given in Figure 2-1. We note that in the displayed model, only one residual function and state variable is associated with each discipline.

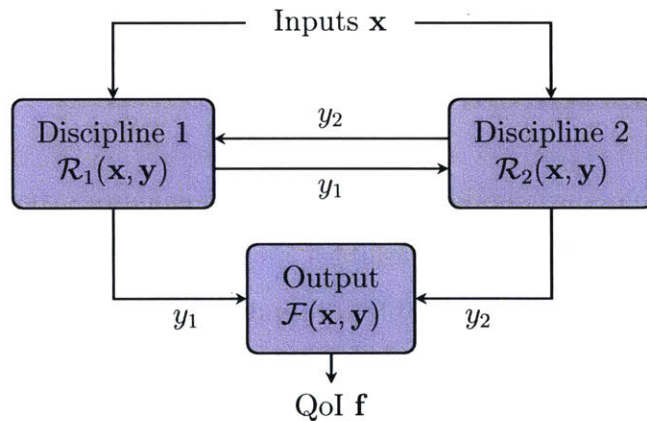


Figure 2-1: Graphical representation of a multidisciplinary model

The dependence of the residual equations on coupling variables from other disciplines defines the overall discipline coupling in the model. This coupling can include one-directional (feed-forward) coupling, or bi-directional (feedback) coupling between any two disciplines. For instance, the dependence of disciplines 1 and 2 on state vari-

ables y_2 and y_1 , respectively, results in the feedback coupling loop for the model in Figure 2-1. Solving these systems of feedback coupled residual equations requires using an iterative numerical method to determine the states and corresponding output variables. Common iterative methods include fixed point iteration or algorithms for solving nonlinear equations, such as Newton-based approaches.

2.2 Uncertainty Quantification for MDAO

In many multidisciplinary models, the input variables are not known exactly. As an example, when analyzing a rocket during launch, it is common to consider wind gusts as a source of uncertainty in the multidisciplinary model [13]. In such cases, the input variables, \mathbf{x} , are represented as random variables endowed with a probability density function (PDF) $\pi_{\mathbf{x}}: \mathcal{X} \rightarrow \mathbb{R}_+$.

With uncertain inputs, the state and output variables of the model are also random variables where the outputs, \mathbf{f} , have an induced joint probability density function, $\pi_{\mathbf{f}}: \mathbb{R}^p \rightarrow \mathbb{R}_+$. Forward uncertainty quantification (UQ) propagates the input uncertainty through the coupled model to characterize the uncertainty in the output variables, as defined by their distribution. For engineering applications, this typically includes determining the mean and variance of the output variables, although sometimes other properties of the output distribution are also of interest, including higher-order moments and the PDF itself.

For linear residual and output equations, the mean and variance in the output variables can be quantified analytically in terms of the model equations and the mean and variance of the inputs. However, for general nonlinear equations, forward UQ is more challenging and is commonly performed with various probabilistic techniques such as sampling-based methods, localized Taylor-series model expansions, functional approximations (e.g., polynomial chaos expansions), etc [56].

One of the traditional and robust algorithms for quantifying uncertainty is based on Monte Carlo simulation. This sampling-based method draws N realizations of the input variables from $\pi_{\mathbf{x}}$. Each realization defines a sample, where $\mathbf{x}^{(j)}$ denotes

the j -th sample of \mathbf{x} for $j = 1, \dots, N$. For each sample, the algorithm solves the coupled residual equations, $\mathcal{R}(\mathbf{x}^{(j)}, \mathbf{y}^{(j)}) = \mathbf{0}$ to determine the state $\mathbf{y}^{(j)}$, and evaluates $\mathbf{f}^{(j)} = \mathcal{F}(\mathbf{x}^{(j)}, \mathbf{y}^{(j)})$ to determine the output variables $\mathbf{f}^{(j)}$ for sample j . These samples are used to estimate the mean, variance, or other statistics of the PDF for the output variables, $\pi_{\mathbf{f}}$.

It is well known that the variance in Monte Carlo estimators converges at a rate of $\frac{1}{\sqrt{N}}$ with an increasing number of samples, N [56]. As a result, for expensive simulations under limited computational resources, this motivates the use of alternative techniques that reduce the computational cost of solving the multidisciplinary model. One approach for efficiently propagating uncertainty in multidisciplinary systems is a decomposition-based method that combines Monte Carlo sampling of each discipline or domain with importance sampling to analyze the uncertainty in feed-forward systems [2]. For multidisciplinary systems with feedback, recent methods include a likelihood-based approach to decouple feedback loops and reduce the model to a feed-forward system [51], dimension reduction techniques to represent certain coupling variables with spectral expansions that depend on a small number of uncertain parameters [3, 5, 4], and the use of adaptive surrogate models for individual disciplines [30] or to approximate the coupling variables and reduce the number of model evaluations [17].

In this thesis we focus on approximate coupling as a way to reduce the computational cost of finding a solution to the multidisciplinary model. We note that this could also be combined with existing techniques to accelerate forward uncertainty quantification and model coupled analyses.

2.3 Model Coupling Approximations

In this section we propose to approximate multidisciplinary models by decoupling discipline inputs, as described in Section 2.3.1. Using notions of the statistical distance between distributions that are presented in Section 2.3.2, we formulate an optimization problem to find approximate model couplings in Section 2.3.3.

2.3.1 Decoupling Discipline Inputs

Our approach of approximate coupling is to minimize the number of coupling variables between disciplines. In practice, many disciplines are often weakly coupled and the residual equations in each discipline are most sensitive to only a subset of the coupling variables [34]. Therefore, the coupling variables that have a minor effect on the state variables computed by each discipline can be fixed to a nominal value, while incurring a small effect on the discipline output. Furthermore, by removing the dependence of a discipline on certain coupling variables, we decouple the corresponding disciplines. Ultimately, this reduction in discipline coupling minimizes the number of iterations required to find a converged solution to the system.

In a multidisciplinary model with K disciplines, we define $\mathcal{S}_k \subseteq \{y_1, \dots, y_n\}$ as the set of coupling variables that are arguments to the residual functions in the k -th discipline. The set of discipline couplings in each model is then given by $M = (\mathcal{S}_1, \dots, \mathcal{S}_K)$, where the total number of discipline couplings is $d(M) = \sum_{k=1}^K |\mathcal{S}_k|$ and $|\mathcal{S}_k|$ denotes the cardinality of \mathcal{S}_k . For a multidisciplinary model with a set of defined discipline couplings, our goal in this work is to find the coupling variables (i.e. $y_i \in \mathcal{S}_k$ for all k) that have a small effect on the overall model output. Each of these variables are then fixed to a nominal value, such as their mean value, and removed from \mathcal{S}_k , thereby only coupling the corresponding disciplines through deterministic values for the state variables. By fixing these random state variables, these discipline inputs are no longer unknown and do not require iteration to find their converged values. In the remainder of this thesis, we refer to this process as *decoupling*.

The operation of replacing the dependence of residual equation, \mathcal{R}_2 , on the coupling variable, y_1 , with the fixed constant input, \bar{y}_1 , is seen graphically in Figures 2-2 and 2-3. By decoupling discipline 1 from 2, the original feedback coupled multidisciplinary model is converted into a feed-forward model. In this case, instead of having to iteratively solve the coupled residual equations to find \mathbf{y} , the state variables can be more cheaply determined by solving the one-way coupled model.

For the remainder of this thesis, we denote the set of discipline couplings in our

reference multidisciplinary model as M_0 , and use M to represent a model that results from decoupling one or more coupling variables.

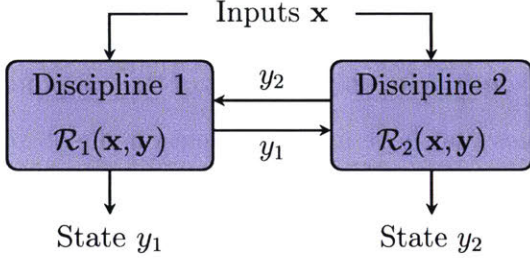


Figure 2-2: Feedback coupled two discipline model, M_0

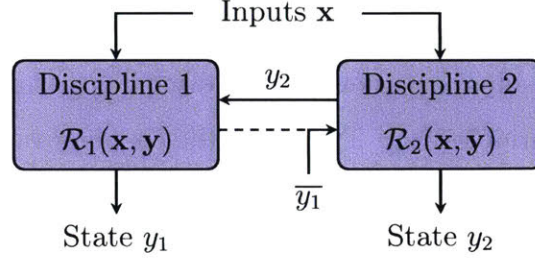


Figure 2-3: One-way coupled two discipline model, M

In the context of forward UQ, the error introduced by decoupling discipline inputs is defined as the accuracy of the PDF for the output variables in model M relative to the reference model M_0 . The next section will discuss methods for quantifying the accuracy in the PDF.

2.3.2 Measures of Statistical Distance

In the field of information theory, various metrics and divergences have been defined for measuring the statistical distance between probability distributions with PDFs, π and $\tilde{\pi}$, over the same probability space, Ω . One broad category of these functions is the set of f -divergences that is defined as

$$D_f(\pi||\tilde{\pi}) = \int_{\Omega} f\left(\frac{\pi}{\tilde{\pi}}\right) d\tilde{\pi}, \quad (2.3)$$

where f is a convex function that satisfies $f(1) = 0$. For different choices of f , we arrive at many common divergences including the Kullback-Leibler divergence $D_{KL}(\pi||\tilde{\pi})$ for $f(t) = t \log(t)$, the total variational distance $D_{TV}(\pi||\tilde{\pi})$ for $f = \frac{1}{2}|t - 1|$, and the Hellinger distance $D_H(\pi||\tilde{\pi})$ for $f(t) = (\sqrt{t} - 1)^2$. The reader is referred to Ref. [25] for a summary of these probability measures and how they are all related via different bounds and transformations.

Given its widespread use, in this work we adopt the Kullback-Leibler (KL) divergence to measure the accuracy of the PDF for the output variables in the decoupled model, $\pi_{\mathbf{f}_M}$, relative to PDF for the outputs in the reference model, $\pi_{\mathbf{f}_{M_0}}$. This KL divergence, denoted by $D_{KL}(\pi_{\mathbf{f}_{M_0}}||\pi_{\mathbf{f}_M})$, has the functional form given in (2.4). The KL divergence provides an indication of the information lost when using $\pi_{\mathbf{f}_M}$ to approximate the distribution of the output variables with the discipline couplings, M . The reader is referred to Ref. [41] for more properties on the KL divergence.

$$D_{KL}(\pi_{\mathbf{f}_{M_0}}||\pi_{\mathbf{f}_M}) = \int_{\mathbb{R}^P} \pi_{\mathbf{f}_{M_0}} \log \left(\frac{\pi_{\mathbf{f}_{M_0}}}{\pi_{\mathbf{f}_M}} \right) d\mathbf{f} \quad (2.4)$$

We also note that other measures besides the f -divergences can also be used to measure the statistical distance between probability distributions. One particular example is the Wasserstein metric for computing the distance between empirical distributions of observed data. This metric has gained attention recently with the advances in optimal transport that have made its computation tractable [8].

2.3.3 Search for Optimal Coupling

To reduce the computational cost of finding solutions to the multidisciplinary model equations in (2.1), we search for the maximum number of discipline inputs that can be decoupled while minimizing the information lost from decoupling. These competing objectives are used to formulate a combinatorial optimization problem to find an optimal subset of discipline couplings by exploring the space of possible decoupled models, which we denote by \mathcal{M} . The optimally decoupled model, which we denote by $M^*(\lambda)$, is found by solving

$$M^*(\lambda) = \arg \min_{M \in \mathcal{M}} D_{KL}(\pi_{\mathbf{f}_{M_0}}||\pi_{\mathbf{f}_M}) - \lambda \mathcal{P}(M). \quad (2.5)$$

The two parts of the objective function in (2.5) are the KL divergence that measures the accuracy of the output variable distribution in the decoupled model relative to the reference model, and the function $\mathcal{P}: \mathcal{M} \rightarrow \mathbb{N}$ that represent the sparsity of

the model couplings. In this work, $\mathcal{P}(M) = d(M_0) - d(M)$, i.e., the function counts the number of removed discipline couplings for each possible model, M . Finally, these two parts are combined in the objective function by using a tuning parameter, λ , that controls the relative importance given to accuracy vs. sparsity.

We note that other metrics can also be used for the sparsity penalty, $\mathcal{P}(M)$, in the optimization problem. For example, by using additional knowledge about the contribution of each discipline coupling to the total cost of running the model, we can assign a different weight to each coupling that favors removing computationally expensive discipline couplings.

Finally, for a model with K disciplines and d discipline couplings, d can grow as $d = K^2 - K$ for a fully coupled model with single connections between disciplines. With each coupling being either active or inactive, the total size of the space of possible models, \mathcal{M} , grows exponentially with cardinality $|\mathcal{M}| = 2^d$. As a result, it is not feasible to compare models with many disciplines by enumeration. Instead, the next section discusses an approach based on a particle method to efficiently explore the high-dimensional model space and find the optimal model, M^* .

2.4 Combinatorial Optimization

In Section 2.4.1 we present the formulation and implementation details of the Sequential Monte Carlo algorithm for solving the combinatorial optimization problem to optimally decouple a subset of the coupling variables. The complete algorithm for solving the model selection problem in equation (2.5) is presented in Section 2.4.2.

2.4.1 Sequential Monte Carlo

To find an optimal decoupled model that best balances the accuracy and sparsity of an approximate model coupling, we solve the combinatorial optimization problem in equation (2.5) that was proposed in Section 2.3.3. With the increase in the cardinality of the model space, \mathcal{M} , the complexity of finding an optimal solution to this problem increases exponentially. As a result, common algorithms for combina-

torial optimization problems find approximate solutions by using local neighborhood searches, depth-first branch and bound searches, or randomized methods such as simulated annealing and genetic algorithms. For a comparison of these methods, the reader is referred to Ref. [6].

To effectively explore the model space, there is growing numerical evidence that global particle methods based on sequential Monte Carlo (SMC), which track a population of possible solutions, are robust and can often outperform heuristics and local search methods for binary optimization problems [19, 52]. This is particularly evident in strongly multi-modal objective functions where local search methods can become trapped in certain modes of the model space. As a result, in this thesis we chose to use SMC to perform the combinatorial optimization.

Sequential Monte Carlo is described in more detail below and summarized in Algorithm 1—the main steps are:

1. Generate a set of weighted particles representing possible decoupled models
2. Update the weights based on the value of the objective function
3. Propose new particles to explore the space while converging to the optimum
4. Repeat steps 2-3

To find the optimal solution using SMC, we first define a sequence of probability distributions on the model space, which we denote by $P_t: \mathcal{M} \rightarrow [0, 1]$ for each index $t \in \mathbb{N}$. These distributions progress from a distribution that is easy to sample from, such as the uniform weighting over the set of all possible decoupled models (i.e., $P_0(M) = \mathcal{U}(M)$), to the final target distribution of interest that concentrates most mass over the models that minimize the objective function. To smoothly move towards the target distribution, the goal of an efficient particle method is to learn the correlations and properties of the model space in order to find the global minimizers without enumerating all models.

Particle methods approximate the distribution at each step by a finite set of $L \in \mathbb{N}$ weighted particles that each represent a particular model. We denote particle l at step t of the algorithm by $M_t^{(l)}$ for $l = 1, \dots, L$. Each of these particles is also associated

with a weight that we denote by $w_t^{(l)}$. The collection of particles and weights at each step is given by $(\mathbf{M}_t, \mathbf{w}_t) = \{M_t^{(l)}, w_t^{(l)}\}_{l=1}^L$. While the weights are set to $1/L$ when initializing the system at $t = 0$, at each iteration of the algorithm the weights are updated based on the subsequent distributions using the ratio

$$u_{t+1}^{(l)} := w_t^{(l)} \frac{P_{t+1}(M_t^{(l)})}{P_t(M_t^{(l)})}, \quad (2.6)$$

followed by the normalization step, $w_{t+1}^{(l)} = u_{t+1}^{(l)} / \sum_l u_{t+1}^{(l)}$ for $l = 1, \dots, L$. Here we note that it is sufficient to evaluate the models at the unnormalized versions of P_t and P_{t+1} given that the weight only requires the ratios of these probability mass functions. As described in a recent study by Schäfer et al. [52], one common and successful technique to construct these distributions over the model space is to use a tempered family. This family of distributions assigns to each decoupled model, $M \in \mathcal{M}$, the probability mass

$$P_t(M) \propto \exp(-\rho(t)h(M)), \quad (2.7)$$

where $h: \mathcal{M} \rightarrow \mathbb{R}_+$ is the value of the objective function for each model in the combinatorial optimization problem, and $\rho: \mathbb{N} \rightarrow \mathbb{R}_+$ is a monotonically increasing tempering parameter that depends on step t . For the combinatorial optimization problem posed in equation (2.5), the objective function is given by $h(M) = D_{KL}(\pi_{\mathbf{f}_{M_0}} || \pi_{\mathbf{f}_M}) - \lambda \mathcal{P}(M)$. Therefore, as t and $\rho(t)$ increase at each iteration, the tempered family concentrates more mass on the set of models that minimize the value for the objective function by assigning higher weights to these models. However, by just repeatedly re-weighting the initial set of particles using equation (2.6), the weights will become uneven and eventually lead to particle degeneracy with a poor sample approximation to each distribution.

As a result, the main ingredient that differentiates sequential Monte Carlo samplers from classical importance sampling is the series of alternating re-weighting and update steps that re-sample and move the particles in order to explore the space

outside of the current set of models. By measuring a series of metrics including the effective sample size and the particle diversity, that are denoted by $ESS(\mathbf{w}_t)$ and $\zeta(\mathbf{M}_t, \mathbf{w}_t)$, respectively, SMC re-samples the particles from its current empirical distribution and moves them by using an adapted proposal distribution that samples new particles in the model space. This combination of steps ensures there is a smooth transition between all distributions, P_t , and they converge to the target based on an appropriate choice for the tempering parameter, $\rho(t)$.

Following the study in Ref. [52], we use systematic re-sampling to generate more samples corresponding to the particles with larger weights. Furthermore, we construct an independent proposal for the move steps that is adapted by using maximum likelihood estimation with the set of samples and weights at each iteration. As demonstrated in Ref. [53], this adaptive proposal results in faster exploration of the model space by drawing particles with a higher rate of acceptance.

The re-sample and move steps are repeated until the sample diversity drops below a threshold given by $\delta > 0$, indicating that most of the mass in P_t is concentrated on a few decoupled models. At the end, the algorithm returns the maximizer of the final distribution within this small set of particles as the optimally decoupled model. For the purpose of completeness, a summary of the algorithm to find the optimal model coupling is presented in Algorithm 1. Nevertheless, the reader is referred to Ref. [52] for more specific details and limitations on each function in the SMC algorithm for solving combinatorial optimization problems.

In practice, the number of particles at each iteration of the SMC algorithm, L , and the minimum particle diversity, δ , are chosen relative to the cardinality of the model space. While the number of particles should be small enough to ensure that the model selection is computationally feasible, it should also be large enough to approximately represent the distributions, P_t . In our numerical examples we chose a minimum particle diversity of $\delta = 0.1L$ to terminate the algorithm and indicate that most probability mass is concentrated on a small number of unique particles. Finally, as empirically demonstrated in Ref. [52], this algorithm is scalable for optimization on binary spaces with at least 250 dimensions. Thus, SMC is at least applicable for

Algorithm 1: SMC for Optimal Model Coupling

Input: Model Objective Function: $h(M)$, Minimum Particle Diversity: δ ,
Number of Particles: L

- 1 Initialize counter: $t = 0$;
- 2 Sample models: $M_t^{(l)} \stackrel{iid}{\sim} \mathcal{U}(M)$ for $l = 1, \dots, L$;
- 3 Initialize weights: $w_t^{(l)} = 1/L$ for $l = 1, \dots, L$;
- 4 Compute particle diversity: $\zeta_t = \zeta(\mathbf{M}_t, \mathbf{w}_t)$;
- 5 **while** $\zeta_t > \delta$ **do**
- 6 Update counter: $t \leftarrow t + 1$;
- 7 Fit proposal: $q_t \leftarrow \text{Fit_Proposal}(\mathbf{M}_t, \mathbf{w}_t)$;
- 8 Re-sample: $\mathbf{M}_t \leftarrow \text{Re_sample}(\mathbf{M}_t, \mathbf{w}_t)$;
- 9 Move models: $\mathbf{M}_t \leftarrow \text{Move}(q_t, \mathbf{M}_t, \mathbf{w}_t)$;
- 10 Update $\rho(t)$ to maintain a minimum $ESS(\mathbf{w}_t)$;
- 11 Update weights: $\mathbf{w}_t \leftarrow \text{Importance_Weights}(h(M), \rho(t), \mathbf{M}_t, \mathbf{w}_t)$;
- 12 Compute particle diversity: $\zeta_t = \zeta(\mathbf{M}_t, \mathbf{w}_t)$;
- 13 **end**
- 14 Return $M^* = \arg \min_{M \in \mathbf{M}_t} h(M)$, $h(M^*)$

combinatorial model selection in models that have up to 250 coupling variables.

2.4.2 Model Selection Procedure

Using the sequential Monte Carlo algorithm presented above, each value of λ in the objective function returns a new optimally decoupled model, $M^*(\lambda)$. However, in an engineering setting it is often of interest to select the model coupling approximation based on accuracy requirements and the availability of computational resources.

To do so, we repeat the combinatorial optimization problem for different values of $\lambda \in \Lambda$. This provides a set of candidate models that balance accuracy and sparsity differently in the optimally decoupled models. While a heavier weight is placed on finding sparse models for larger values of λ , a heavier weight is placed on finding accurate models for smaller values of λ . As a result, by increasing λ , the optimal model trades-off accuracy in the probability distributions of the output variables relative to M_0 with having a smaller number of discipline couplings. By finding

the optimal model for each value of λ using Algorithm 1, a user can select one of these models for their application by comparing the accuracy of each model to the computational cost of its sample evaluations. A summary of the complete algorithm for optimal model selection is presented in Algorithm 2.

Algorithm 2: Optimal Model Selection

Input: Reference multidisciplinary model: M_0 , Tuning parameters vector: Λ ,
Model sparsity function: $\mathcal{P}(M)$

- 1 **for** $\lambda \in \Lambda$ **do**
 - 2 | Setup objective function: $h(M) = D_{KL}(\pi_{\mathbf{f}_{M_0}} || \pi_{\mathbf{f}_M}) - \lambda \mathcal{P}(M)$;
 - 3 | Determine $M^*(\lambda)$ using Algorithm 1;
 - 4 **end**
 - 5 Select optimal model by comparing accuracy and sparsity of $M^*(\lambda)$ for $\lambda \in \Lambda$;
 - 6 Return decoupled model, M^* ;
-

Chapter 3

Identifying Approximate Coupling

In this chapter we present three tractable approaches for efficiently evaluating model output uncertainty and the overall objective in the optimal model coupling problem that was presented in Section 2.3.3. These approaches are model linearizations, control variate corrections, and a Bayesian approach that have increasingly higher accuracy and associated higher computational cost. The advantages, implementation details, and complexity of each approach are discussed in the following sections.

3.1 Model-Linearization Approach

The KL divergence term in equation (2.5) represents the accuracy of the output distributions for a possible model M relative to the reference model, M_0 . One approach to evaluate this term is to generate samples for the output variables and use them to approximate the empirical distribution for the model outputs and the KL divergence using kernel density estimation (KDE) techniques [55]. However, for large-scale multidisciplinary models it is typically not feasible to generate many samples from multiple models via Monte Carlo simulations that require repeatedly solving coupled systems of equations. As a result, in Sections 3.1.1 and 3.1.2 we propose to address this problem by using a linear approximation for each model to efficiently estimate the KL divergence between the probability distributions of the output variables.

3.1.1 Output Uncertainty in Linearized Model

The linearization of a multidisciplinary model is given by the first order Taylor series approximation of the model outputs in equation (2.2). This linearization is found by evaluating the partial derivatives of the function \mathcal{F} with respect to input and state variables, which are denoted by $\partial_{\mathbf{x}}\mathcal{F}$ and $\partial_{\mathbf{y}}\mathcal{F}$, respectively. Both of these derivatives are evaluated at a specific linearization point for the input and state variables.

One natural linearization point is based on the mean of the state variables, \mathbf{y} . However, finding the mean requires characterizing the distribution for these variables *a-priori*. Instead, an approximation commonly employed in the literature is to linearize the models around the state variables corresponding to the mean values of the input variables [51]. The mean value of \mathbf{x} is denoted by $\mu_{\mathbf{x}} = \mathbb{E}[\mathbf{x}]$ where \mathbb{E} is the expectation operator with respect to the probability distribution with density $\pi_{\mathbf{x}}$. The state variables corresponding to $\mu_{\mathbf{x}}$ satisfy the residual equations, $\mathcal{R}(\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})) = \mathbf{0}$, and these states are referred to as a first-order mean for \mathbf{y} [42].

Using the Taylor series expansion about the first-order mean, the linearized approximation of the output variables is given by

$$\mathbf{f} = \mathcal{F}(\mathbf{x}, \mathbf{y}) \approx \mathcal{F}(\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})) + \partial_{\mathbf{x}}\mathcal{F}|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} (\mathbf{x} - \mu_{\mathbf{x}}) + \partial_{\mathbf{y}}\mathcal{F}|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} (\mathbf{y} - \mathbf{y}(\mu_{\mathbf{x}})). \quad (3.1)$$

Similarly, the linearized approximation of the coupled residual equations for the model around the same linearization point is given by

$$\mathcal{R}(\mathbf{x}, \mathbf{y}) \approx \mathcal{R}(\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})) + \partial_{\mathbf{x}}\mathcal{R}|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} (\mathbf{x} - \mu_{\mathbf{x}}) + \partial_{\mathbf{y}}\mathcal{R}|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} (\mathbf{y} - \mathbf{y}(\mu_{\mathbf{x}})), \quad (3.2)$$

where $\partial_{\mathbf{x}}\mathcal{R}$ and $\partial_{\mathbf{y}}\mathcal{R}$ denote the partial derivatives of the residuals with respect to the input and state variables, respectively. Rearranging the approximation in equation (3.2) for the state variables, and noting that $\mathcal{R}(\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})) = \mathbf{0}$, a linear approximation to the state variables with respect to \mathbf{x} is given by

$$\mathbf{y} = \mathbf{y}(\mu_{\mathbf{x}}) - (\partial_{\mathbf{y}}\mathcal{R}|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})})^{-1} \partial_{\mathbf{x}}\mathcal{R}|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} (\mathbf{x} - \mu_{\mathbf{x}}). \quad (3.3)$$

Here we note that the matrix of partial derivatives, $(\partial_{\mathbf{y}}\mathcal{R})^{-1}\partial_{\mathbf{x}}\mathcal{R}$ in equation (3.3), contains the sensitivity of the state variables to perturbations in the inputs. These sensitivities to the input variables are often computed as part of an adjoint analysis for single or multiple discipline models [28, 44]. Furthermore, the sensitivities are typically found without assembling and inverting the full matrix of partial derivatives for the residual equations with respect to the state variables. Instead, the values in the matrix are determined by solving a linear system using an iterative numerical method (i.e., the Jacobi or Gauss-Seidel method) that takes advantage of the sparse structure in the Jacobian matrices [45].

Substituting the linearization for the state variables in equation (3.1), a linear approximation to the output QoIs in the multidisciplinary model with respect to the input parameters, \mathbf{x} , is given by

$$\mathbf{f} \approx \tilde{\mathbf{f}}(\mathbf{x}) = \mathcal{F}(\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})) + \left. \frac{d\mathcal{F}}{d\mathbf{x}} \right|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} (\mathbf{x} - \mu_{\mathbf{x}}), \quad (3.4)$$

where $\tilde{\mathbf{f}}(\mathbf{x})$ represents the linear approximation and $\left. \frac{d\mathcal{F}}{d\mathbf{x}} \right|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})}$ denotes the matrix of total derivatives of \mathcal{F} with respect to the input variables, given by

$$\left. \frac{d\mathcal{F}}{d\mathbf{x}} \right|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} = \partial_{\mathbf{x}}\mathcal{F}|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} - \partial_{\mathbf{y}}\mathcal{F}|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} (\partial_{\mathbf{y}}\mathcal{R}|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})})^{-1} \partial_{\mathbf{x}}\mathcal{R}|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})}. \quad (3.5)$$

With this linear relationship in the inputs, \mathbf{x} , the mean and variance of the linearized outputs are propagated analytically from the mean and variance of the input variables. In particular, if the input variables are normally distributed with covariance matrix, $\Sigma_{\mathbf{x}}$, the linearized output variables are also normally distributed with mean, $\mu_{\tilde{\mathbf{f}}} = \mathbb{E}[\tilde{\mathbf{f}}]$, and covariance, $\Sigma_{\tilde{\mathbf{f}}} = \mathbb{E}[(\tilde{\mathbf{f}} - \mu_{\tilde{\mathbf{f}}})(\tilde{\mathbf{f}} - \mu_{\tilde{\mathbf{f}}})^T]$, where

$$\mu_{\tilde{\mathbf{f}}} = \mathcal{F}(\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})), \quad (3.6)$$

$$\Sigma_{\tilde{\mathbf{f}}} = \left[\left. \frac{d\mathcal{F}}{d\mathbf{x}} \right|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} \right] \Sigma_{\mathbf{x}} \left[\left. \frac{d\mathcal{F}}{d\mathbf{x}} \right|_{\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})} \right]^T. \quad (3.7)$$

To compute the KL divergence for the output variables of each decoupled model,

$M \in \mathcal{M}$, relative to the reference model, M_0 , we compute the mean and covariance of the linearized model outputs using (3.6) and (3.7). The mean and covariance for the linearized output variables of model M_0 are denoted by $\mu_{\tilde{\mathbf{f}}_{M_0}}$ and $\Sigma_{\tilde{\mathbf{f}}_{M_0}}$, respectively. Similarly, the mean and covariance for the decoupled model M are denoted by $\mu_{\tilde{\mathbf{f}}_M}$ and $\Sigma_{\tilde{\mathbf{f}}_M}$, respectively. Using the closed form expression for the KL divergence between two Gaussian distributions, we estimate the KL divergence between the probability distributions for the outputs of models M_0 and M with the equation

$$D_{KL}(\pi_{\tilde{\mathbf{f}}_{M_0}} || \pi_{\tilde{\mathbf{f}}_M}) = \frac{1}{2} \left\{ \text{Tr}(\Sigma_{\tilde{\mathbf{f}}_M}^{-1} \Sigma_{\tilde{\mathbf{f}}_{M_0}}) + (\mu_{\tilde{\mathbf{f}}_M} - \mu_{\tilde{\mathbf{f}}_{M_0}})^T \Sigma_{\tilde{\mathbf{f}}_M}^{-1} (\mu_{\tilde{\mathbf{f}}_M} - \mu_{\tilde{\mathbf{f}}_{M_0}}) - p - \ln \left(|\Sigma_{\tilde{\mathbf{f}}_M}^{-1} \Sigma_{\tilde{\mathbf{f}}_{M_0}}| \right) \right\}, \quad (3.8)$$

where $\text{Tr}(\cdot)$ denotes the matrix trace operator, and $\ln(|\cdot|)$ is the log-determinant of a matrix. We note that the derivatives in equation (3.4) are typically available when computing system outputs or performing design optimization. This leads to a small incremental cost for evaluating the KL divergence in equation (3.8).

For numerical stability when computing the KL divergence with equation (3.8), we require that the covariance matrices of the reference and decoupled model are invertible and are well-conditioned. One required condition for the covariance to be full-rank and invertible is that the dimension of the random inputs, m , should be greater than the dimension of the output variables, p . Otherwise, the Gaussian distribution will be supported along a lower-dimensional space (at most $m < p$) and will be rank-deficient. Although non-invertibility of the covariance is typically not encountered for models that have large m and small p , adding Gaussian noise with small variance to the output QoIs is one approach to ensure the PDF of the outputs is fully supported for certain models and the covariance is invertible.

In practice, it is also important to ensure that the selected QoIs have a similar order of magnitude so the covariance matrix is numerically well-conditioned and the objective function weights the closeness of the QoIs equally. One common approach is to select common variables with similar orders of magnitude (i.e., lift and drag coefficients for an aircraft design problem) or normalized physical variables (i.e., fuelburn

to gross weight ratio as compared to absolute fuelburn).

Finally, we note that decoupling disciplines results in additional sparsity in the $\partial_{\mathbf{y}}\mathcal{R}$ matrix when the corresponding residual equations no longer depend on certain coupling variables from other disciplines. We can take advantage of this sparse structure in the $\partial_{\mathbf{y}}\mathcal{R}$ matrix when computing the output sensitivities in the model linearization and when computing the KL divergence with equation (3.8).

3.1.2 Computational Complexity

Each decoupled model, $M \in \mathcal{M}$, results in a different first-order mean for the state variables. As a result, to compute the mean and covariance for the distribution of the output variables, it is necessary to solve the nonlinear residual equations once for each model M at the mean input variables. The resulting first-order mean of \mathbf{y} is used as the linearization point to evaluate the partial derivatives of the residual and output functions in the model linearization.

Therefore, the computational cost of evaluating the model linearization is proportional to the cost of computing the first-order mean and total derivatives in equation (3.4). As noted in [45], these derivatives can be evaluated using an adjoint method whose cost is independent of the number of input variables, m , and instead proportional to the number of output QoIs, which is typically smaller than m . Denoting this cost of solving the nonlinear model once by c , the cost of identifying the optimal model with the SMC algorithm for each value of λ scales as $\mathcal{O}(cLT)$ for T total iterations and L particles (i.e., decoupled models) evaluated at each iteration. We note that the cost of solving the coupled residual equations and evaluating the total derivatives of many multidisciplinary models will be the overall limiting step at each iteration. Therefore, in this analysis we assume that the cost of sampling over the model space and updating the proposal in the SMC algorithm are negligible.

Nevertheless, with a judicious choice for the values used as the decoupled inputs to each discipline, we can significantly reduce the cost for the model selection algorithm. In particular, if the first-order mean of state variable, $y_i \in \mathcal{S}_j$, is also used as the fixed

input, $\bar{y}_i = y_i(\mu_{\mathbf{x}})$, the residual equation for discipline j is given by

$$\mathcal{R}_j(\mu_{\mathbf{x}}, y_1, \dots, y_{i-1}, \bar{y}_i, y_{i+1}, \dots, y_n) = \mathcal{R}_j(\mathbf{x}, y_1, \dots, y_{i-1}, y_i(\mu_{\mathbf{x}}), y_{i+1}, \dots, y_n). \quad (3.9)$$

Given that $\mathcal{R}_j(\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})) = 0$, the unknown state variables that solve a perturbed system of this form are given by the first-order mean values for the reference model. Therefore, in this case it is not necessary to re-evaluate the system's outputs after decoupling certain connections. Furthermore, with the linearization point for the derivative evaluations also remaining the same for a decoupled model, M , only the dependence of the residual equations on the state variables changes. To account for this change in variable dependence, it is sufficient to mask the effect of the decoupled inputs in the $\partial_{\mathbf{y}}\mathcal{R}$ matrix (i.e., mask $\partial_{y_i}\mathcal{R}_j$ for the example above) and re-compute the total derivatives for model M using the updated matrix in equation (3.5).

We note that by not solving the derived nonlinear system for decoupled model M , we are not guaranteed of the existence of a solution to the corresponding set of coupled residual equations. In addition, when decoupling a subset of the inputs for this model by masking the Jacobian matrix, $\partial_{\mathbf{y}}\mathcal{R}$, it is also necessary to check that this matrix remains invertible so that the linearization of model M exists. We refer the reader to Section 5.4 for sufficient conditions to guarantee the invertibility of this matrix. We note that these conditions serve as a good empirical heuristic for the convergence of the decoupled nonlinear model and the existence of a solution when using an iterative numerical method to compute the decoupled model's state variables and output QoIs.

By not requiring a model evaluation and replacing the cost of evaluating the decoupled model's output uncertainty by matrix operations, the computation of the KL divergence is negligible for a decoupled model after evaluating the first-order mean and derivatives for the reference model, M_0 . Therefore, the complexity of the model selection algorithm in this case reduces to the cost of performing one evaluation of the reference model outputs and derivatives and the total cost of running SMC with a linear-time model, resulting in a total complexity of $\mathcal{O}(c + LT)$.

We note that the solution of this reduced cost method will defer from the solution of algorithms that use more accurate representations for the fixed coupling variables, and account for the discrepancy in the output of the decoupled nonlinear model. Nonetheless, this method present a cost-effective method to identify near-optimal decoupled models in a computationally-limited setting.

3.2 Improvements to Model Output Uncertainty

For multidisciplinary models with highly nonlinear relationships between the input and output variables, the linearization presented in Section 3.1 introduces a bias in the estimates for the mean and covariance of the output QoIs. This leads to a discrepancy between the Gaussian distribution for the linearized outputs and the actual Gaussian approximation for the output QoIs. In this section we explore two methods to improve the linear-based estimates for the mean and variance of the output variables.

3.2.1 Higher-Order Approximations

One natural approach for increasing the approximation accuracy of the model linearization is to include the second-order terms in the Taylor series approximation to the model outputs

$$\mathbf{f} \approx \tilde{\mathbf{f}}(\mathbf{x}) = \mathcal{F}(\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})) + \frac{d\mathcal{F}}{d\mathbf{x}}(\mathbf{x} - \mu_{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \mu_{\mathbf{x}})^T \frac{d^2\mathcal{F}}{d\mathbf{x}^2}(\mathbf{x} - \mu_{\mathbf{x}}), \quad (3.10)$$

where the second-order derivatives depend on the Hessian information of the coupled residual equations, and all derivatives are evaluated at the mean point, $(\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}}))$. Using the higher-order expansion, the mean and covariance of the quadratic approximation to the model outputs is given by

$$\mu_{\tilde{\mathbf{f}}} = \mathcal{F}(\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})) + \frac{1}{2} \text{Tr} \left(\frac{d^2\mathcal{F}}{d\mathbf{x}^2} \Sigma_{\mathbf{x}} \right) \quad (3.11)$$

$$\Sigma_{\tilde{\mathbf{f}}} = \left[\frac{d\mathcal{F}}{d\mathbf{x}} \right] \Sigma_x \left[\frac{d\mathcal{F}}{d\mathbf{x}} \right]^T + \frac{1}{2} \text{Tr} \left(\frac{d^2\mathcal{F}}{d\mathbf{x}^2} \Sigma_x \frac{d^2\mathcal{F}}{d\mathbf{x}^2} \Sigma_x \right). \quad (3.12)$$

In comparison to the mean and covariance in (3.6) and (3.7) based on the linearized approximations, the higher-order derivative information can be used to correct the estimates for these moments and better approach the optimal Gaussian approximation to the output QoIs.

On a computational note, in cases when directly computing the traces involving the Hessian matrix are prohibitively expensive, users may take advantage of recent advances in randomized trace estimation methods that only require the application of the Hessian on random vectors (e.g., standard Gaussian vectors). For an application of these randomized methods in the context of PDE-constrained optimization, the reader is referred to [1].

3.2.2 Control Variate Corrections

A second approach for increasing the estimation accuracy for the mean and covariance of the output QoIs is based on using control variates as a variance reduction strategy.

To estimate the expectation of a random variable, $\mathbf{f} \in \mathbb{R}^p$, control variates take advantage of an auxiliary variable, $\mathbf{g} \in \mathbb{R}^p$, with an analytically known expectation, to construct the new random variable

$$\hat{\mathbf{f}} = \mathbf{f} + C(\mathbf{g} - \mathbb{E}[\mathbf{g}]), \quad (3.13)$$

where $C \in \mathbb{R}^p$ is a set of tuning parameters. Given that $\mathbb{E}[\hat{\mathbf{f}}] = \mathbb{E}[\mathbf{f}]$ from linearity, $\hat{\mathbf{f}}$ is an unbiased estimator for \mathbf{f} and its variance is given by

$$\mathbb{V}[\hat{\mathbf{f}}] = \mathbb{V}[\mathbf{f}] + C^2 \mathbb{V}[\mathbf{g}] + 2C \cdot \text{Cov}[\mathbf{f}, \mathbf{g}], \quad (3.14)$$

where $\mathbb{V}[\cdot]$ and $\text{Cov}[\cdot, \cdot]$ are vectors that denote the elementwise variance and covariance of their respective arguments. Minimizing the quadratic function for the variance with respect to each entry of C , the optimal tuning parameters, C^* , are given by

$$C^* = -\frac{\text{Cov}[\mathbf{f}, \mathbf{g}]}{\mathbb{V}[\mathbf{g}]}, \quad (3.15)$$

where the ratios are computed elementwise. This results in an unbiased estimator for $\mathbb{E}[\mathbf{f}]$ with variance

$$\mathbb{V}[\hat{\mathbf{f}}] = \mathbb{V}[\mathbf{f}] - (\text{Cov}[\mathbf{f}, \mathbf{g}])^2 / \mathbb{V}[\mathbf{g}] = (\mathbf{1} - \rho[\mathbf{f}, \mathbf{g}]^2) \mathbb{V}[\mathbf{f}], \quad (3.16)$$

where $\rho[\mathbf{f}, \mathbf{g}] \in \mathbb{R}^p$ is a vector containing the elementwise correlation coefficients of the random variables and $\mathbf{1} \in \mathbb{R}^p$ is a vector of ones. We note that if \mathbf{f} and \mathbf{g} are uncorrelated, the estimator recovers the original variance of \mathbf{f} . Otherwise, if the correlation of the statistics is positive, the variance of the mean estimate for the new random variable is strictly reduced. Using N Monte-Carlo samples to estimate the expectation of $\hat{\mathbf{f}}$, the unbiased estimator for the mean is given by

$$\mathbb{E}[\hat{\mathbf{f}}] \approx \hat{\mathbf{f}}_N = \frac{1}{N} \sum_{j=1}^N [\mathbf{f}^{(j)} + C^*(\mathbf{g}^{(j)} - \mathbb{E}[\mathbf{g}])], \quad (3.17)$$

where in practice C^* is based on the empirical correlation between \mathbf{f} and \mathbf{g} from the collected samples. This estimator has variance $\mathbb{V}[\hat{\mathbf{f}}_N] = \frac{1}{N} \mathbb{V}[\hat{\mathbf{f}}] = \frac{1}{N} (\mathbf{1} - \rho[\mathbf{f}, \mathbf{g}]^2) \mathbb{V}[\mathbf{f}]$, as compared to the standard Monte Carlo estimator variance of $\frac{1}{N} \mathbb{V}[\mathbf{f}]$. We note that to determine the number of samples required to meet a desired threshold, a user may monitor the root mean squared error given by $RMSE = \sqrt{\mathbb{V}[\hat{\mathbf{f}}_N]}$ where the variance is computed empirically from the collected samples.

Given that the first-order Taylor series approximation is correlated with the response of the output QoI by construction, we can use the first order linearizations as a control-variate (i.e., $\mathbf{g} = \tilde{\mathbf{f}}$) to approximate both the mean and covariance of the QoIs. These control variates take advantage of both the linearization and a sample-based approach to produce low-variance and unbiased estimators for the statistics of the output variables of each model using a small number of additional samples.

In particular, we note that while estimating the mean of the output QoIs follows from the above discussion, computing the covariance using control variates requires the variance of these variance estimators in order to determine the optimal tuning parameter. These higher order moments can also be determined approximately from

the collected samples, given that the control variate estimators still perform well when the elements of C are perturbed from their optimal values. The reader is referred to Ref. [48] for a thorough discussion and analysis of control variate estimators for mean and variance in the context of optimization under uncertainty.

Finally, we conclude this section with a comparison of the computational complexity of the control variates approach to the model linearization approach discussed in Section 3.1. While a model linearization allows for efficient uncertainty propagation with only a single evaluation of the nonlinear model, its accuracy is limited by the bias introduced from the linear approximation and the locality of the linearization point. On the other hand, with an increasing number of samples, the control variates provide reduced variance estimators for the statistics of each model's outputs. However, to cheaply determine the optimal coupling for performing uncertainty quantification, it is necessary that the algorithm does not incur a greater computational expense than quantifying the uncertainty in the reference multidisciplinary model, M_0 . When using the control variate estimators with N samples in the SMC algorithm with L particles and T iterations, the algorithm complexity scales as $\mathcal{O}(NcLT)$, where c is the cost of solving the nonlinear model. Therefore it is necessary to trade-off the accuracy of the model output approximations with this algorithm complexity for identifying the optimal models when selecting the number of samples, N , to use in these estimators.

3.3 Extension to Bayesian Approach

In the previous section we addressed how to exploit the model structure via Taylor series approximations and control variates to cheaply estimate the KL divergence for the output QoIs of different model structures. These methods are combined with the combinatorial optimization algorithm based on SMC that was presented in Section 2.4 to determine a set of decoupled models that accurately characterize the mean and covariance of the system outputs.

However, in many applications where the model outputs follow complex distributions that are not accurately described by their first two moments (i.e., random

variables with multi-modal or heavily skewed distributions), we need an alternative approach to assess the accuracy of a decoupled model for describing data from the output distribution of the reference model. For algorithmic completeness, in Sections 3.3.1 and 3.3.2 we explore one approach to do so that is inspired by reformulating the model selection problem using a Bayesian approach.

3.3.1 Posterior over Model Space

Given samples for the output QoIs of the reference model, $D = \{\mathbf{f}^{(j)}\}_{j=1}^N$, the minimization of the KL divergence term in equation (2.5) may be written as

$$\begin{aligned} \min_{M \in \mathcal{M}} D_{KL}(\pi_{\mathbf{f}_{M_0}} || \pi_{\mathbf{f}_M}) &= \min_{M \in \mathcal{M}} \mathbb{E}_{\pi_{\mathbf{f}_{M_0}}} [\log(\pi_{\mathbf{f}_{M_0}}) - \log(\pi_{\mathbf{f}_M})] \\ &= \min_{M \in \mathcal{M}} -\mathbb{E}_{\pi_{\mathbf{f}_{M_0}}} [\log(\pi_{\mathbf{f}_M})] \\ &\approx \max_{M \in \mathcal{M}} \frac{1}{N} \sum_{j=1}^N \log(\pi_{\mathbf{f}_M}(\mathbf{f}^{(j)})) \end{aligned} \quad (3.18)$$

where in the second equality we have used that the first term is constant with respect to M , and in the third line we approximated the expectation with a Monte Carlo estimate. Noticing that the evaluation of the samples from model M_0 under the log-density is representative of the likelihood of the reference model data coming from the decoupled model, we can define the Monte Carlo estimate in line (3.18) as the log-likelihood of data D given model M ,

$$\frac{1}{N} \sum_{j=1}^N \log(\pi_{\mathbf{f}_M}(\mathbf{f}^{(j)})) := \log(\mathcal{L}(D|M)). \quad (3.19)$$

This measure for the accuracy of decoupled model M can then be combined with the sparsity penalty term, $\mathcal{P}(M)$, that is representative of a model prior to arrive at the equivalent optimization problem,

$$\begin{aligned} \min_{M \in \mathcal{M}} D_{KL}(\pi_{\mathbf{f}_{M_0}} || \pi_{\mathbf{f}_M}) - \lambda \mathcal{P}(M) &\Leftrightarrow \max_{M \in \mathcal{M}} \log(\mathcal{L}(D|M) \exp(-\lambda \mathcal{P}(M))) \\ &\Leftrightarrow \max_{M \in \mathcal{M}} \mathcal{L}(D|M) \exp(-\lambda \mathcal{P}(M)). \end{aligned} \quad (3.20)$$

As a result, we can formulate the process of finding the optimal decoupled model of a multidisciplinary problem as identifying the model $M \in \mathcal{M}$ that maximizes the posterior probability over models, as given by

$$P(M|D) \propto \mathcal{L}(D|M)\mathcal{Q}(M), \quad (3.21)$$

where $\mathcal{Q}(M)$ is the prior on each model. With this Bayesian formulation, a single optimal model can be identified from the maximum a posteriori (MAP) point of the distribution, $M^* = \arg \max_{M \in \mathcal{M}} P(M|D)$. Alternatively, we can also characterize the distribution over possible models that can represent the data weighted by a prior that favors sparse model couplings.

To characterize the posterior, we can use various numerical algorithms including Markov Chain Monte Carlo (MCMC) or by adapting the sequential Monte Carlo algorithm presented in Section 2.4 for model sampling. We refer the reader to Ref. [24] for a comprehensive overview of numerical methods for Bayesian analysis. Nevertheless, to use these algorithms we need to efficiently estimate the likelihood of the data under the true log-density without having direct access to this density, $\pi_{\mathbf{f}_{M_0}}$. One approach that has gained popularity in the inference community for performing model selection with intractable likelihoods, as commonly encountered in model selection problems, is Approximate Bayesian Computation (ABC).

3.3.2 Approximate Bayesian Computation

ABC first emerged in the context of rejection sampling for inferring model parameters. For each parameter drawn from the prior distribution, ABC collects several samples from the associated model and accepts the parameter if these data samples lie close to the data collected from a reference model. Due to the random nature of the model evaluations, the measure of closeness is relaxed by comparing a low-dimensional summary statistic for the data given by η (i.e., such as mean, variance and higher order moments), and using a distance function, ν , to determine if the statistic for the data from the approximate model lies close to data from the reference model.

Using s batches of data from the approximate model, $D_M^{(j)}$, the closeness of each batch to data collected from the reference model, D , is computed and used as a measure for the posterior distribution of model, M , as given by

$$P(M|D) \propto \frac{1}{s} \sum_{j=1}^s \mathbb{1}_{\{\nu(\eta(D), \eta(D_M^{(j)})) < \epsilon\}} \mathcal{Q}(M). \quad (3.22)$$

In the context of Bayesian model selection, ABC naturally merges with the SMC algorithm for sampling over the model space, \mathcal{M} . Following a similar adaptive SMC-ABC algorithm that was proposed in [20] for continuous random variables, we modify Algorithm 1 by sampling models from the prior, $\mathcal{Q}(M)$, and using the posterior in (3.22) instead of (2.7) to assign probability mass to each model.

The probability mass for each model also depends on the tolerance parameter, ϵ_t . This parameter is decreased adaptively at each iteration of the algorithm, t , by measuring the effective sample size, analogously to the tempering parameter in the SMC algorithm for combinatorial optimization. When beginning the algorithm at $t = 0$ with a large ϵ_0 , most decoupled models generate outputs that lie within the accepted tolerance which results in uniform weighting. As the algorithm progresses and ϵ is decreased, the collection of models slowly converge to those drawn from the posterior distribution that corresponds to $\epsilon \rightarrow 0$, while avoiding particle degeneracy. As a result, by using the posterior in (3.22) to re-weight the models in the SMC-ABC algorithm, the final models with the largest weights will have QoIs that lie close to the outputs of the reference model according to the specified distance metric and summary statistics.

The summary of the combined SMC-ABC algorithm for sampling and characterizing the model posterior values given a set of data from the reference model, D , is described in Algorithm 3.

We note that while the user of the algorithm may design the summary statistics to match the moments of all output variables of the model, a goal-oriented statistic could also be chosen to ensure that the specified QoIs of the selected decoupled model lie close in distribution to those in the reference model. Therefore, when using the SMC

Algorithm 3: SMC-ABC for Decoupled Model Posterior

Input: Data from the reference model: D , Model Prior: $\mathcal{Q}(M)$, Initial and final tolerance: $\epsilon_0, \epsilon_{min}$, Distance and summary statistics: ν and η , Number of Particles: L

- 1 Set tolerance $\epsilon = \epsilon_0$;
- 2 Sample models: $M_t^{(l)} \stackrel{iid}{\sim} \mathcal{Q}(M)$ for $l = 1, \dots, L$;
- 3 Initialize weights: $w_t^{(l)} = 1/L$ for $l = 1, \dots, L$;
- 4 **while** $\epsilon > \epsilon_{min}$ **do**
- 5 Fit proposal: $q_t \leftarrow \text{Fit_Proposal}(\mathbf{M}_t, \mathbf{w}_t)$;
- 6 Re-sample: $\mathbf{M}_t \leftarrow \text{Re_sample}(\mathbf{M}_t, \mathbf{w}_t)$;
- 7 Move models: $\mathbf{M}_t \leftarrow \text{Move}(q_t, \mathbf{M}_t, \mathbf{w}_t)$;
- 8 Update ϵ to maintain a minimum $ESS(\mathbf{w}_t)$;
- 9 Update weights: $\mathbf{w}_t \leftarrow \text{Importance_Weights}(\epsilon, \nu, \eta, D, \mathbf{M}_t, \mathbf{w}_t)$;
- 10 **end**
- 11 Return \mathbf{M}_t

algorithm together with ABC to evaluate the posterior probability of each decoupled model with a goal-oriented statistic, η , the algorithm will efficiently explore the model space to find the set of discipline couplings that trade-off the complexity of the model with the closeness of the specific QoIs.

We also comment on approaches that have been recently proposed to accelerate ABC algorithms and can be combined with Algorithm 3 for model selection. These include generating more data by re-sampling from the estimated cumulative distribution function for the outputs of each model [15], evaluating the models at the same values of the input variables, and collecting batches of data with different sizes from each model depending on their estimated closeness to the true model data in order to reduce the computational cost of evaluating infeasible models [50]. Finally, recent theoretical results have shown that under certain assumptions one data sample per model is sufficient for ABC combined with MCMC and this will not affect the asymptotic behavior of the algorithm [11]. Together with Algorithm 3, these results could reduce the number of model evaluations and the overall computational complexity of the Bayesian approach to optimal model selection.

Chapter 4

Numerical Results

In this chapter, we present results from applying the model selection algorithms from Chapter 3 to three aerospace engineering examples. These are a three-discipline model of a satellite used to detect forest fires in Section 4.1 that was presented in an uncertainty quantification context by Sankararaman et al. [51], a model for turbine engine cycle analysis in Section 4.2 that was developed by Hearn et al. [27], and an aero-structural model of a lifting surface in Section 4.3 that was developed by the MDOLab at the University of Michigan [29].

4.1 Fire Detection Satellite Model

In this section we begin by studying the satellite model with a nominal set of parameters for the input uncertainty distribution in Section 4.1.1. The effect of different input uncertainty and the use of control variates for improving the uncertainty estimates are also analyzed in Sections 4.1.2 and 4.1.3, respectively.

4.1.1 Nominal Input Uncertainty

To analyze the performance of a fire detection satellite under uncertain operating conditions, we consider a simplified model that comprises three disciplines: orbit analysis, attitude control, and power analysis. As seen in Figure 4-1, the model

features both feed-forward and feedback coupling variables to exchange information between the disciplines.

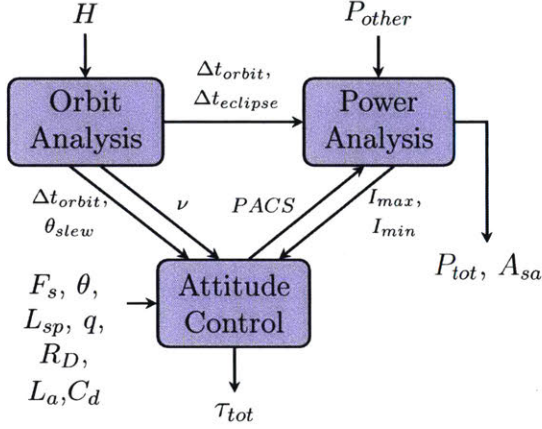


Figure 4-1: Fire detection satellite model from Ref. [51]

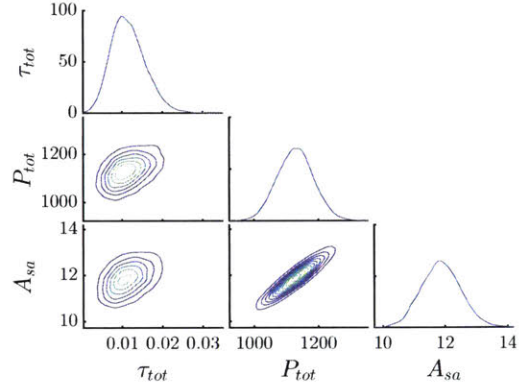


Figure 4-2: Joint distribution of the reference model

To represent the uncertain conditions, the model includes nine Gaussian random variables that are described by their mean and standard deviation parameters in Table 4.1. After propagating the input uncertainty through the disciplines, the seven state variables connecting the disciplines are also uncertain and are distributed with potentially non-Gaussian probability distributions. These state variables are: the orbit period (Δt_{orbit}), eclipse period ($\Delta t_{eclipse}$), satellite velocity (v), maximum slewing angle (θ_{slew}), the power of the attitude control system (P_{ACS}), and the moments of inertia (I_{max} , and I_{min}). In addition, the orbit period is an input to two disciplines (orbit analysis and attitude control) resulting in a total of $d = 8$ coupling variables for this satellite model as seen in Figure 4-1. These coupling variables are used to compute the three output QoIs: total torque (τ_{tot}), total power (P_{tot}), and the area of the solar array (A_{sa}). The joint probability distribution of these output variables, $\pi_{\mathbf{f}}$, is displayed in Figure 4-2.

By defining the model displayed in Figure 4-1 as the reference model, M_0 , our algorithm for optimal model selection considers possible decoupled models, M , that have a smaller number of discipline couplings by fixing a subset of the coupling variables of each discipline to their first-order mean value. In order to empirically validate the use

Table 4.1: Random variable parameters in the satellite model from Ref. [51]

Random Variable	Symbol	Mean	Standard Deviation
Altitude	H	18.0×10^6 m	1.0×10^6 m
Power other than attitude control	P_{other}	1.0×10^3 W	50.0 W
Average solar flux	F_s	1.4×10^3 W/m ²	20.0 W/m ²
Deviation of moment axis	θ	15.0°	1.0°
moment arm for radiation torque	L_{sp}	2.0 m	0.4 m
Reflectance factor	q	0.5	1.0
Residual dipole of spacecraft	R_D	5.0 Am ²	1.0 Am ²
Moment arm for aerodynamic torque	L_a	2.0 m	0.4 m
Drag coefficient	C_d	1.0	0.3

of model linearizations to estimate the KL divergence of the decoupled models, we compare the Gaussian distribution for the linearized outputs of the reference model to the uncertainty in the output variables of the nonlinear coupled system based on 10^4 Monte Carlo samples. As seen in Figure 4-3, the joint empirical distribution for the output variables, π_f , is closely approximated by the multivariate Gaussian distribution resulting from the linearized equations in equation (3.4).

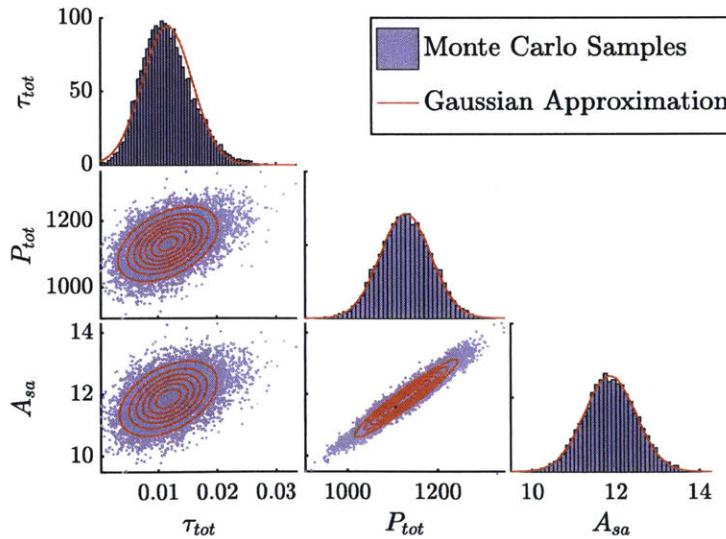


Figure 4-3: Comparison of Gaussian approximation and Monte Carlo samples

Using these linearizations for the output variables, the SMC algorithm explores

the model space using $L = 100$ particles to determine the optimal decoupled model, $M^*(\lambda)$, that solves the combinatorial optimization problem posed in equation (2.5). For each value of λ , the algorithm identifies a subset of discipline couplings that best trade-off the accuracy of the output distributions, that is estimated from the KL divergence, $D_{KL}(\pi_{\mathbf{f}_{M_0}} || \pi_{\mathbf{f}_{M^*}})$, and the added sparsity in the discipline couplings, as measured by $\mathcal{P}(M^*)$.

With increasing values for λ , the objective in the optimization problem adds a greater penalty to models with less sparsity in the discipline couplings. This results in a set of increasingly sparser models that remove additional couplings at the expense of accuracy in the output distribution. These decoupled models for 4 different values of λ in $\Lambda = [10^{-4}, 10^{-3}, 10^{-1}, 10^0]$ are presented in Figures 4-4 to 4-11 with the decoupled connections indicated by dashed lines. The joint distribution for the output variables, $\pi_{\mathbf{f}_{M^*}}$, of the optimal model corresponding to each value of λ is also plotted in the figures below.

For the fire detection satellite model, the algorithm identifies that the feedback coupling variable between the attitude control and power disciplines for the moment of inertia (I_{max} , and I_{min}) could be fixed to their mean value while having a small effect on the accuracy of the joint output distribution, as seen in Figures 4-4 and 4-5. With increasing values for λ , the slewing angle, the satellite velocity, and the orbit period are also found to weakly contribute to the total torque and the overall power requirement for the attitude control subsystem, as seen in Figures 4-6 and 4-7. The subsequent sparser model for $\lambda = 10^{-1}$ fixes these discipline coupling variables along with the state inputs to the power discipline for the orbit and eclipse period. Finally, we note that for $\lambda = 10^0$, the optimal model has fully decoupled disciplines and the joint distribution for its output variables is given in Figure 4-11.

In Table 4.2 we present the values for the KL divergence between the Gaussian distribution for the linearized outputs of the reference model, M_0 , and the optimally decoupled models, M^* , that we denote as the linearized KL divergence. The number of active coupling variables, representing the increasing sparsity of each decoupled model with larger values of λ , is also presented in the table below.

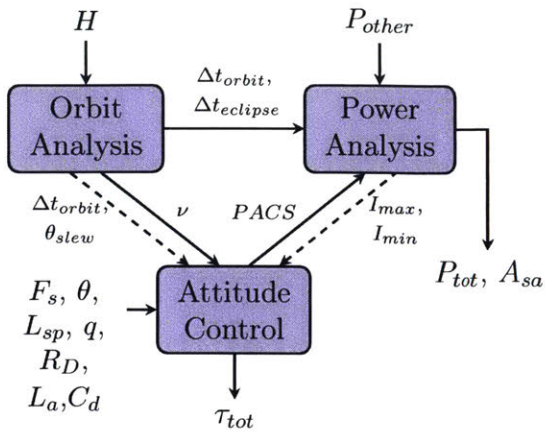


Figure 4-4: Optimal model coupling for $\lambda = 1.0 \times 10^{-4}$

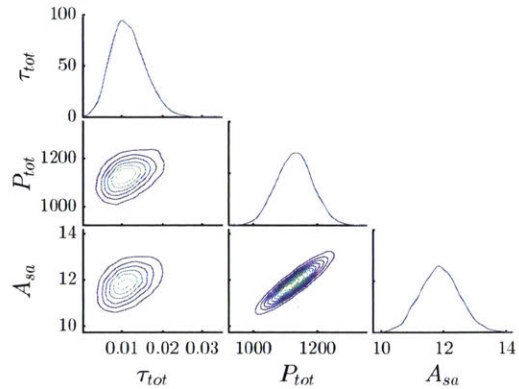


Figure 4-5: Joint distribution for $\lambda = 1.0 \times 10^{-4}$

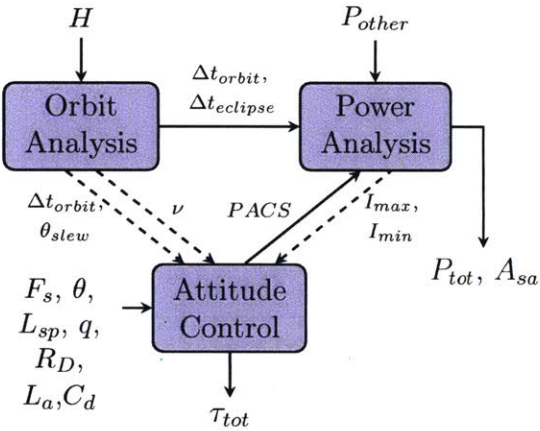


Figure 4-6: Optimal model coupling for $\lambda = 1.0 \times 10^{-3}$

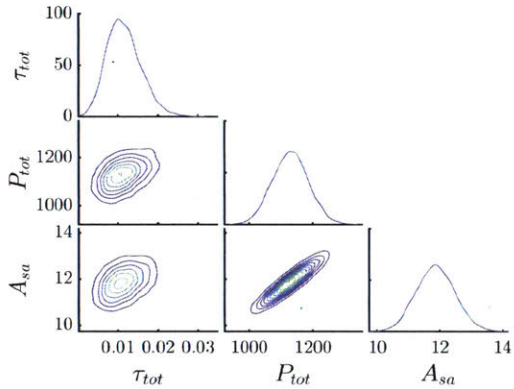


Figure 4-7: Joint distribution for $\lambda = 1.0 \times 10^{-3}$

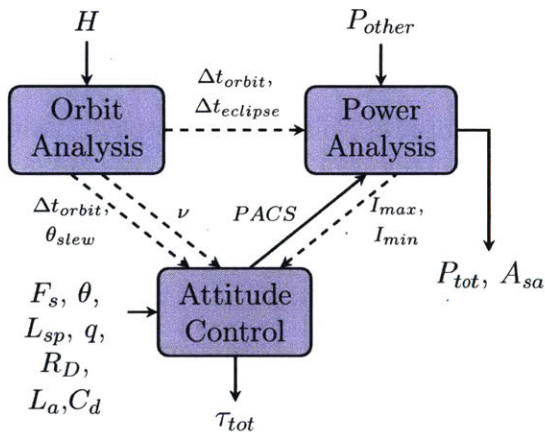


Figure 4-8: Optimal model coupling for $\lambda = 1.0 \times 10^{-1}$

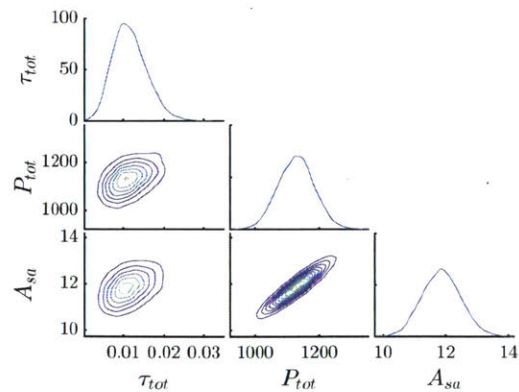


Figure 4-9: Joint distribution for $\lambda = 1.0 \times 10^{-1}$

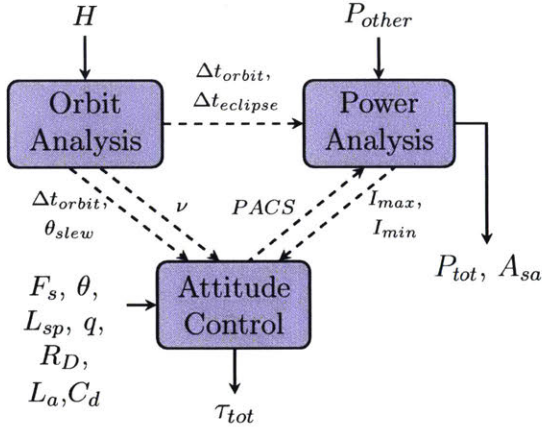


Figure 4-10: Optimal model coupling for $\lambda = 1.0 \times 10^0$

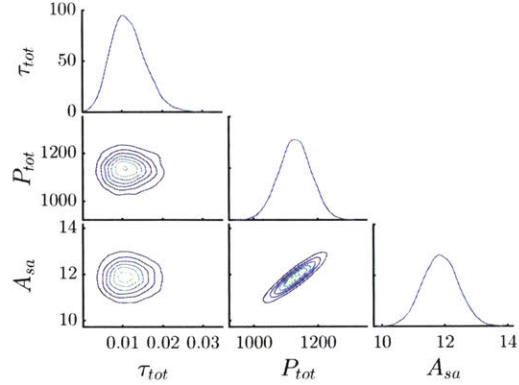


Figure 4-11: Joint distribution for $\lambda = 1.0 \times 10^0$

Table 4.2: Linearized KL divergence for decoupled satellite models with nominal inputs

	$\lambda = 1 \times 10^{-4}$	$\lambda = 1 \times 10^{-3}$	$\lambda = 1 \times 10^{-1}$	$\lambda = 1 \times 10^0$
Linearized KL Divergence	7.30×10^{-16}	7.43×10^{-4}	4.20×10^{-3}	1.33×10^{-1}
Active Coupling Variables	4/8	3/8	1/8	0/8

From the table and figures above, we observe that the optimal models identified by the model selection algorithm corresponding to $\lambda \in [10^{-4}, 10^{-1}]$ result in very similar distributions for the model outputs as the reference model based on the linearized KL divergence. Furthermore, while the final optimal model for $\lambda = 10^0$ approximately captures the marginal uncertainty in the outputs, this fully decoupled model does not represent, as accurately, the correlations between the output variables, as seen in the joint marginal distribution for P_{tot} and τ_{tot} in Figure 4-11. Nevertheless, given that all decoupled models feature only feed-forward connections, these models with approximate coupling can be used to cheaply propagate the uncertainty in the multidisciplinary model without having to iteratively solve for the outputs that correspond to each set of input variables. This results in a substantial computational savings for performing forward UQ. In particular, if low accuracy is sufficient in the

output uncertainty, the disciplines can be analyzed independently using the model corresponding to $\lambda = 10^0$.

Finally, we analyze the pointwise errors of the three model outputs in each decoupled model, M^* , relative to the reference model, M_0 . The pointwise error is given by the difference in the values of the output variables in the decoupled and reference models for the same input sample. The errors for 10^4 input samples are displayed in Figures 4-12 to 4-15. With an increase in λ , the sparser models result in progressively greater pointwise errors, which are seen in the greater spread of the output values between both models. Although the set of discipline couplings corresponding to $\lambda = 10^{-1}$, and 10^0 have relatively accurate joint distributions for the model output uncertainty, their larger pointwise errors makes them less adequate for computations that require accurate pointwise approximations, such as multidisciplinary optimization. Nevertheless, the models associated with $\lambda = 10^{-4}$, and 10^{-3} (Figures 4-4 and 4-6) have accurate distributions for the output variables and low pointwise errors relative to the reference model, M_0 .

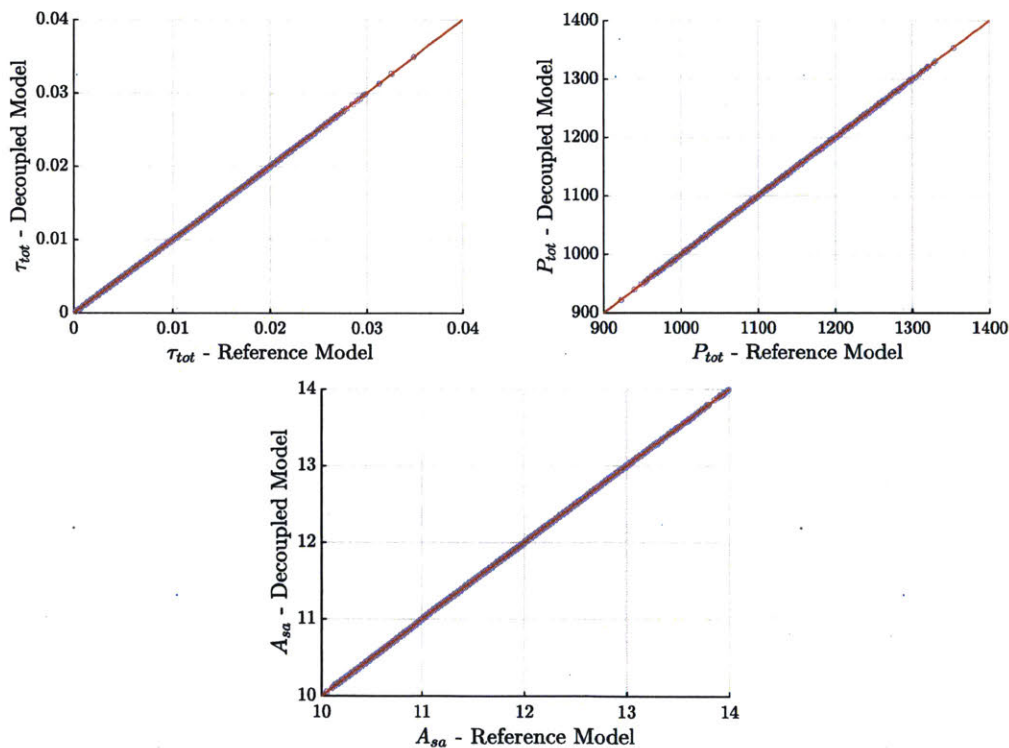


Figure 4-12: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-4}$

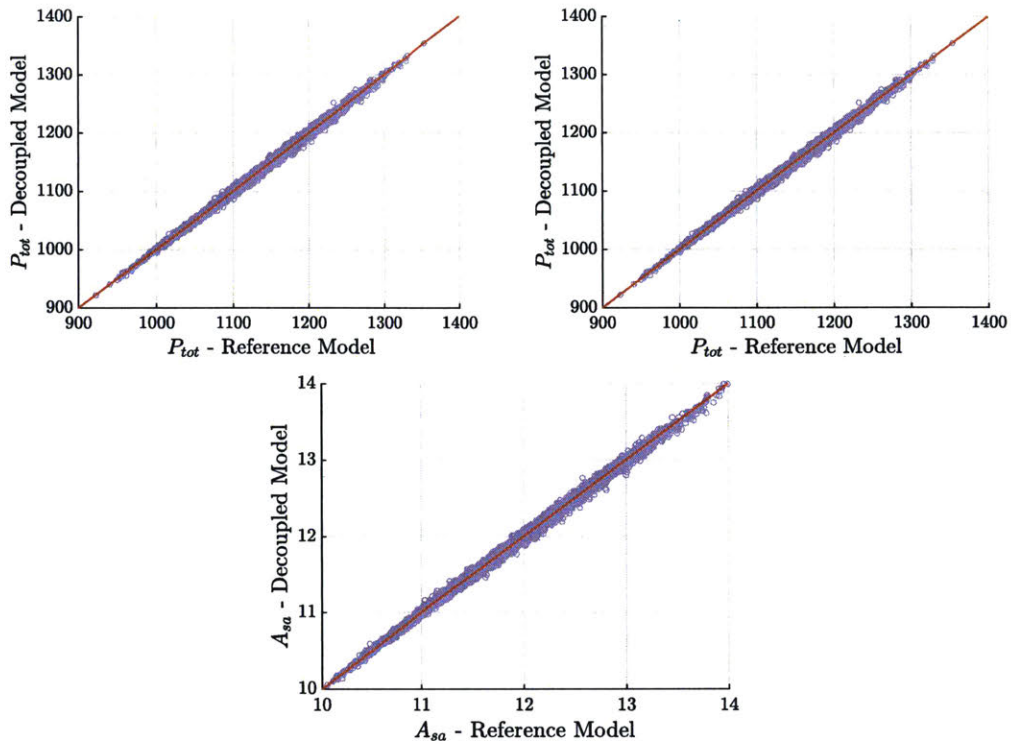


Figure 4-13: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-3}$

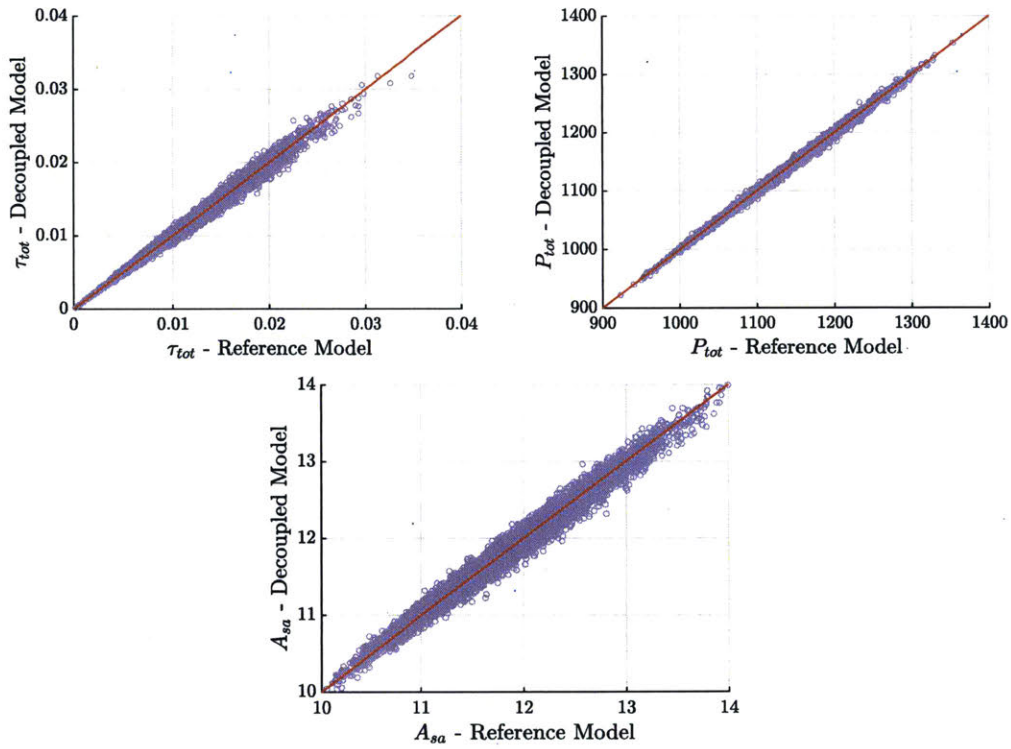


Figure 4-14: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-1}$

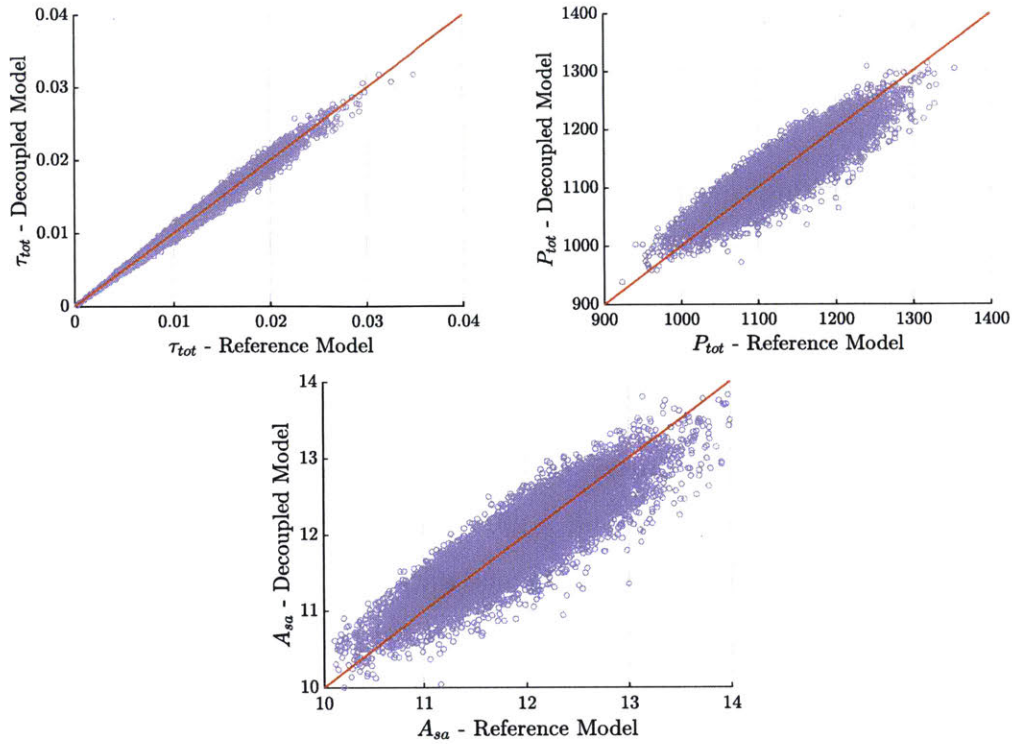


Figure 4-15: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^0$

4.1.2 Different Input Uncertainty

In comparison to methods that only partition and reorder disciplines to reduce feedback loops based on the graphical structure of the model [14], the algorithms presented in this thesis account for the system's physics to identify an optimal approximation of the system coupling. In particular, this coupling is dependent on the sensitivities of the outputs to each input variable and the scaling by the input uncertainty (e.g., the covariance matrix, $\Sigma_{\mathbf{x}}$, for normally distributed input variables). Therefore, when changing the parameters describing the uncertainty in the input variables, the optimal decoupled model should be re-assessed to minimize the discrepancy in the estimates for the output uncertainty.

For the satellite model, we consider the impact of different input uncertainty by introducing uncertainty in the input for the satellite's slewing time period (Δt_{slew}), that was previously fixed in the above analysis. This variable is an input to the attitude control discipline and affects the slewing torque and the resulting total torque

on the satellite. With the previous input parameters, the effect of slew is relatively negligible, and so its effect is accurately represented by the value for the mean slew angle, $\bar{\theta}_{slew}$.

However, by shifting the mean and describing the slewing time-period as a normal random variable, $\Delta t_{slew} \sim \mathcal{N}(5, 0.5^2)$, the joint distribution for the output variables based on 10^4 Monte Carlo samples is displayed in Figure 4-16. Using the model linearization approach, the optimal decoupled models identified by the model selection algorithm for $\lambda = 10^{-4}$ and 10^{-3} with these different input parameters are presented in Figures 4-17 and 4-19, respectively. The joint distributions of the outputs for these models are given in Figures 4-18 and 4-20. These models differ from the optimal models identified in Section 4.1 due to the necessity of maintaining the θ_{slew} coupling variable in this case to accurately compute the total satellite torque. We note that the models for $\lambda = 10^{-1}$ and 10^0 match the earlier optimal models due to the greater weight on having sparse discipline couplings in the objective function with these values for the tuning parameters.

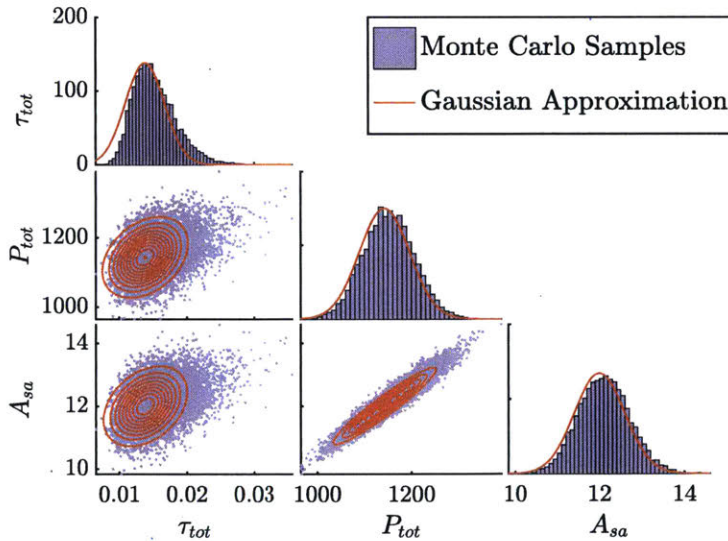


Figure 4-16: Gaussian approximation and Monte Carlo samples with different inputs

Furthermore, we also present the values for the linearized KL divergence between

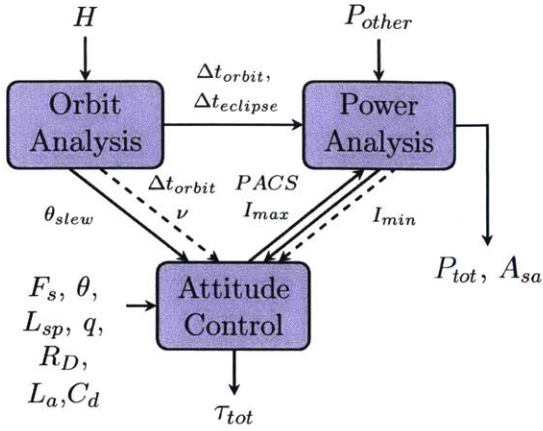


Figure 4-17: Optimal model coupling for $\lambda = 1.0 \times 10^{-4}$

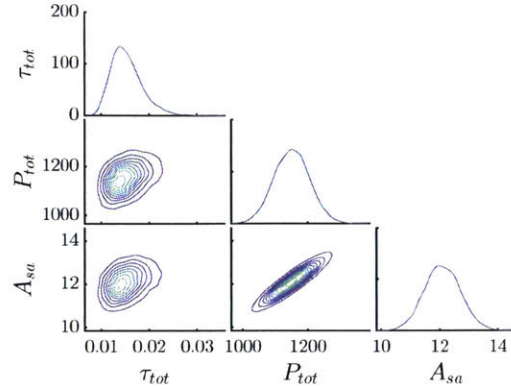


Figure 4-18: Joint distribution for $\lambda = 1.0 \times 10^{-4}$

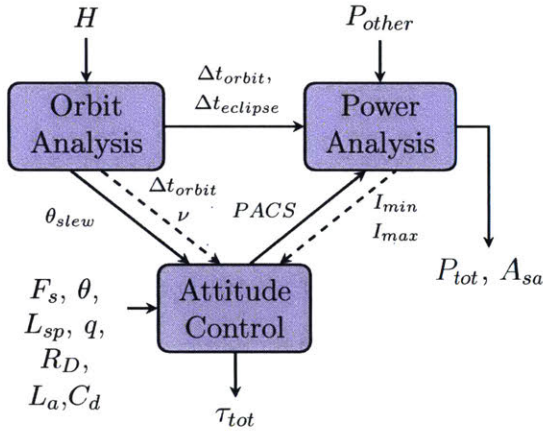


Figure 4-19: Optimal model coupling for $\lambda = 1.0 \times 10^{-3}$

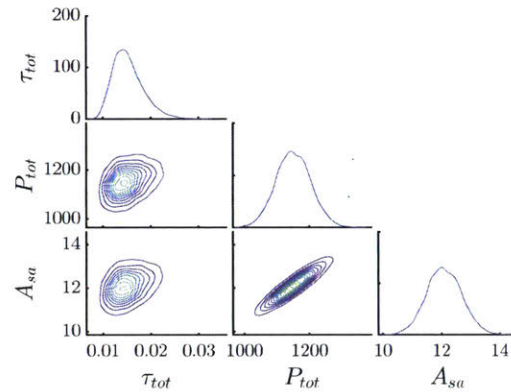


Figure 4-20: Joint distribution for $\lambda = 1.0 \times 10^{-3}$

the Gaussian distribution for the outputs of the reference model, M_0 , and these optimally decoupled models, M^* in Table 4.3. These values are also compared with the linearized KL divergence from using the optimal decoupled models identified in the first section to represent the outputs with the different input uncertainty in Table 4.4. The discrepancy, particularly for $\lambda = 10^{-4}$, highlights the importance of taking the input uncertainty into account when identifying an optimal decoupled model.

Table 4.3: Linearized KL divergence for decoupled satellite models with different inputs

	$\lambda = 1 \times 10^{-4}$	$\lambda = 1 \times 10^{-3}$	$\lambda = 1 \times 10^{-2}$	$\lambda = 1 \times 10^0$
Linearized KL Divergence	0	2.80×10^{-4}	1.04×10^{-2}	7.75×10^{-2}
Active Coupling Variables	5/8	4/8	1/8	0/8

Table 4.4: Linearized KL divergence with sub-optimal decoupled models

	$\lambda = 1 \times 10^{-4}$	$\lambda = 1 \times 10^{-3}$	$\lambda = 1 \times 10^{-2}$	$\lambda = 1 \times 10^0$
Linearized KL Divergence	68.5×10^{-4}	68.5×10^{-4}	1.04×10^{-2}	7.75×10^{-2}
Active Coupling Variables	5/8	4/8	1/8	0/8

Finally, we note that by further decreasing the time period to $\Delta t_{slew} \sim \mathcal{N}(1, 0.01^2)$, the importance of the θ_{slew} coupling variable becomes more pronounced. In particular, the optimal decoupled models for $\lambda = 10^{-4}$ to 10^{-1} and $\lambda = 10^0$ to 10^1 with this input uncertainty are displayed in Figures 4-21 and 4-22, respectively.

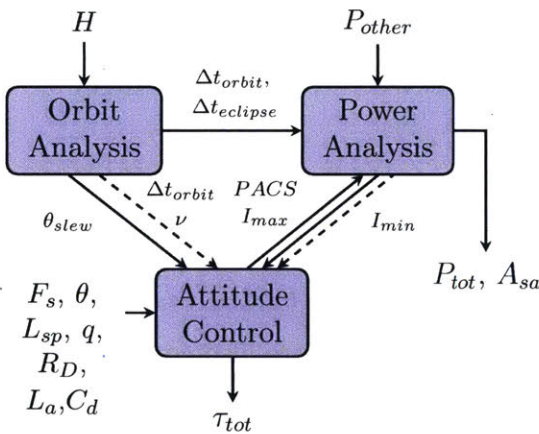


Figure 4-21: Optimal model coupling for $\lambda = 10^{-4}$ to 10^{-1}

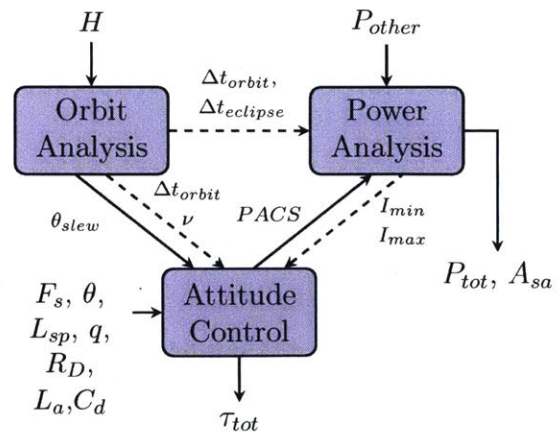


Figure 4-22: Optimal model coupling for $\lambda = 10^0$ to 10^1

4.1.3 Control Variate Corrections

To improve the approximation of the objective in the model selection algorithm, we use a limited number of samples of each nonlinear model and the linearized model as a control variate to more accurately estimate the moments of the output variables. By using samples from each nonlinear model, we are comparing closer Gaussian approximations to the model outputs.

We note that using samples introduces randomness in the computed KL divergence and the overall objective, that is also a random variable in this case. However, by drawing a new sample estimate for the KL divergence at each iteration of the SMC algorithm for a given model, the randomness of this estimator is in principle averaged and the algorithm should still converge to the optimal model asymptotically. This provides one additional advantage of using SMC to a combinatorial optimization algorithm with deterministic steps (i.e., branch-and-bound) or a greedy approach that can make incorrect decisions when observing noisy values of the objective function and may not allow them to be corrected in future iterations.

In this section, we first consider the application of control variates to correct the estimates for the mean of the three output variables of the satellite model with the different input uncertainty conditions presented in Section 4.1.2. From the high correlation observed between Monte Carlo outputs of the nonlinear model and its linearization, as seen for the P_{tot} output variable in Figure 4-23, the linearization is used as a cheap-to-evaluate control variate in this study. The convergence of the control variate estimator for the mean of the P_{tot} output with a 95% standard error interval is seen in Figure 4-24. In the figure, we also observe the bias in the mean of the Gaussian output derived from the linear model approximation in comparison to the unbiased control variate estimator, and the slow convergence of a standard Monte Carlo estimator.

For this study, we use 10^2 samples to correct the Gaussian approximation for the output means of each model. A comparison of the control variate estimates to the Gaussian approximation of the linearized model and 10^4 Monte Carlo samples of the

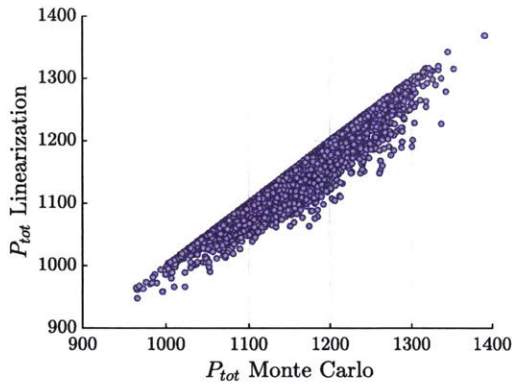


Figure 4-23: Correlation of nonlinear and linearized models

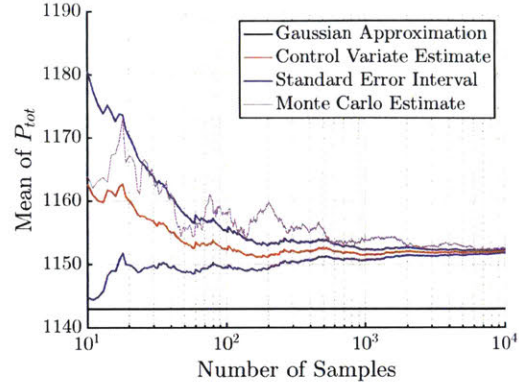


Figure 4-24: Control variate estimates with increasing sample size

outputs is given in Figure 4-25. The figure also demonstrates the effect of the control variates in correcting the bias in the mean introduced by the linear-based Gaussian approximation for the reference model, M_0 . This is particularly evident for the τ_{tot} output that has a skewed distribution.

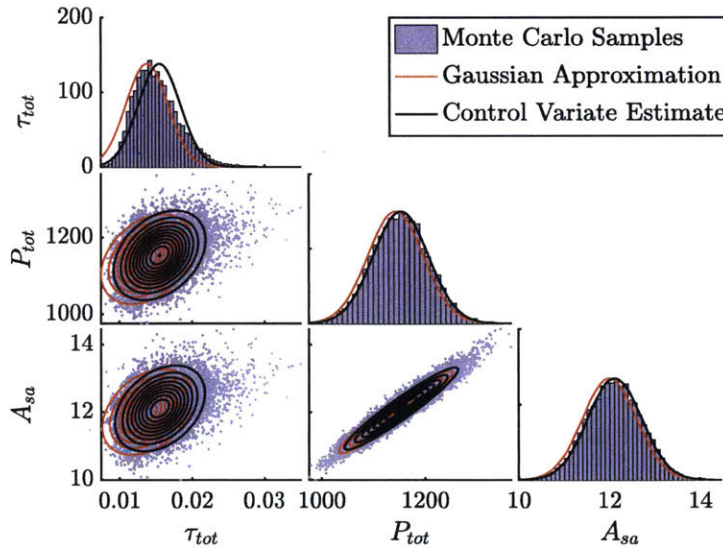


Figure 4-25: Linearized Gaussian approximation and control variate estimates in comparison to Monte Carlo samples

When using the control variate estimators to compute the KL divergence, the model selection algorithm identifies the same optimal models as those presented above for larger values of λ , where sparsity of the discipline couplings has a greater effect than model accuracy, while a more refined set of models is found for lower values

of λ . In particular for $\lambda = 10^{-5}$ and $\lambda = 10^{-6}$, the optimal models identified the discrepancy in the mean calculation of the models selected in the previous study and chose denser models to more accurately represent the outputs. These two models along with their joint PDFs are displayed in Figures 4-26 to 4-29. We note that while the solution for the optimal model, M^* , for $\lambda = 10^{-6}$ includes the feedback coupling between the power analysis and attitude control disciplines, the models for $\lambda \geq 10^{-5}$ are still entirely feed-forward.

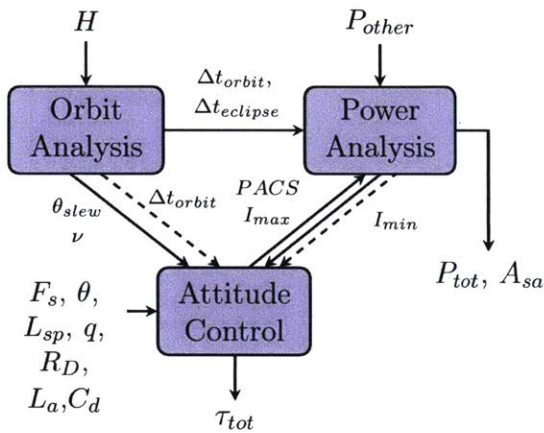


Figure 4-26: Optimal model coupling for $\lambda = 1.0 \times 10^{-6}$

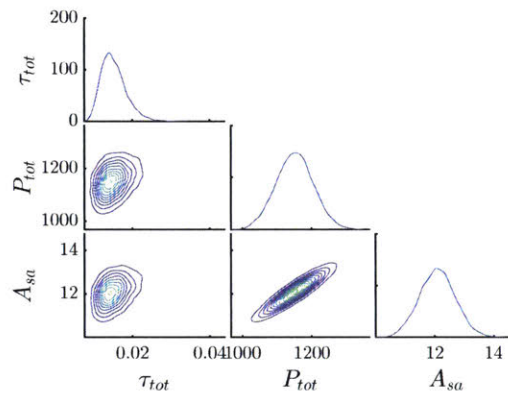


Figure 4-27: Joint distribution for $\lambda = 1.0 \times 10^{-6}$

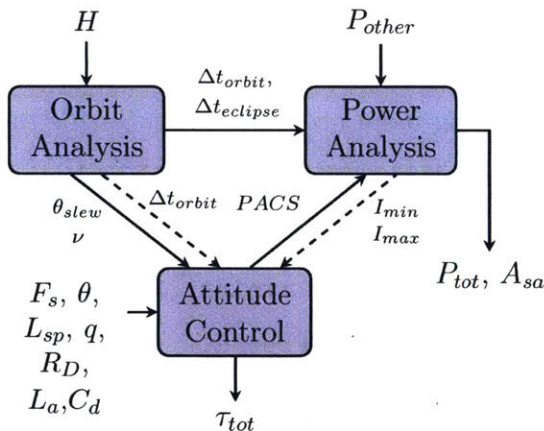


Figure 4-28: Optimal model coupling for $\lambda = 1.0 \times 10^{-5}$

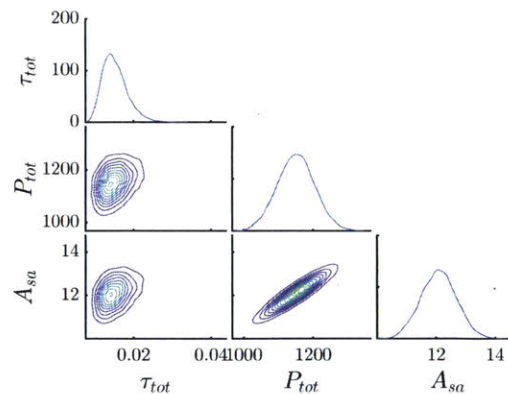


Figure 4-29: Joint distribution for $\lambda = 1.0 \times 10^{-5}$

In the second application of control variates, we use 10^3 samples of each nonlinear

model to correct both the mean and variance estimates for the satellite model outputs with the original input uncertainty that was explored in Section 4.1. The convergence trends of the control variate estimators for the variance of the τ_{tot} and P_{tot} output variables are displayed in comparison to standard Monte Carlo variance estimators in Figures 4-30 and 4-31, respectively. We note that although 10^3 samples of each model provide a relatively accurate representation of the first two moments, it may be sufficient in certain problems to use a smaller number of samples to determine the mean and variance within a 95% standard error interval.

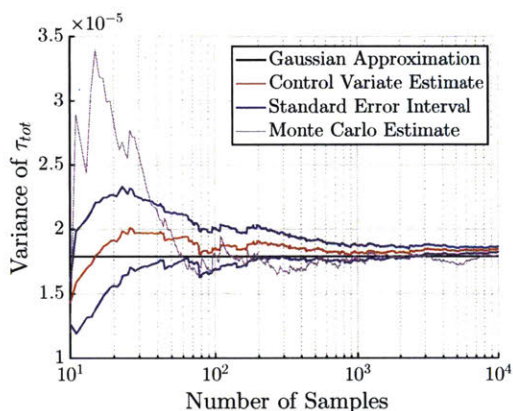


Figure 4-30: Control variate estimates for variance of τ_{tot}

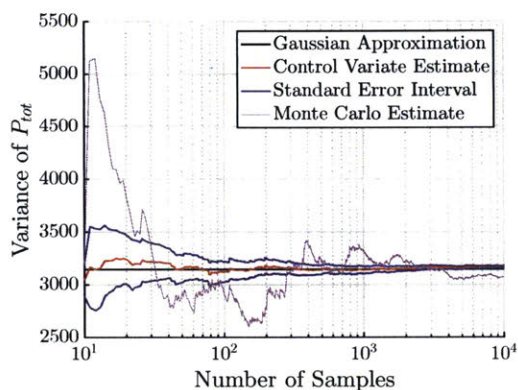


Figure 4-31: Control variate estimates for variance of P_{tot}

Similarly to the application of control variates above, we observe similar models as the initial study in Section 4.1 for $\lambda = 10^0$ and 10^{-1} , and models with a denser set of couplings for smaller values of λ when including more accurate estimates for the output mean and variance. The set of optimal models identified by the model selection algorithm for $\lambda = 10^{-6}, 10^{-5}$ are plotted in Figures 4-32 and 4-33, respectively. We note that these models approach the reference model for increasingly small λ to accurately reconstruct the distribution of the reference model outputs.

Finally, we evaluate the estimates of the KL divergence between the Gaussian approximation to the outputs of each model with mean and variance given by the control variate estimators with 10^3 samples. The mean of 250 samples for the KL divergence of each decoupled model's outputs along with a 95% confidence interval is given in Table 4.5. We note that by using samples to correct and define the mean

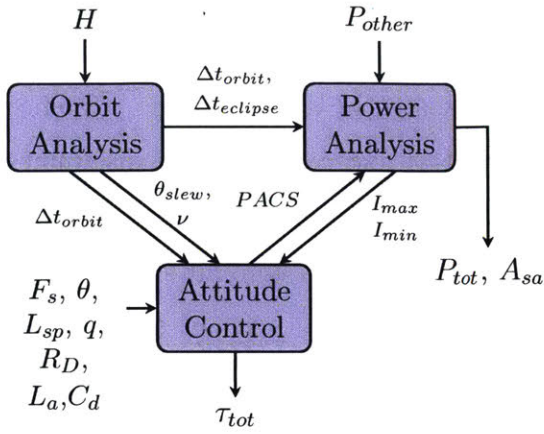


Figure 4-32: Optimal model coupling for $\lambda = 1.0 \times 10^{-6}$

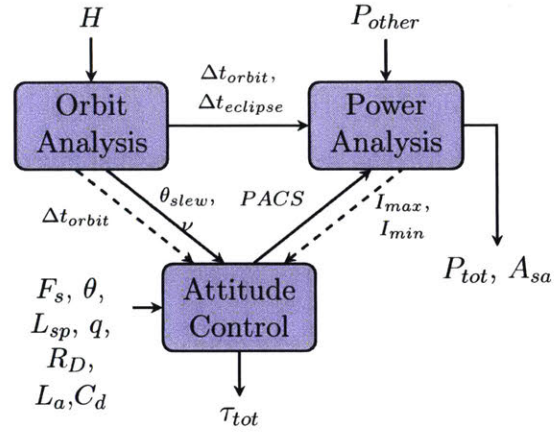


Figure 4-33: Optimal model coupling for $\lambda = 1.0 \times 10^{-5}$

and covariance of the reference model output distribution, the optimization algorithm is identifying a model that closely matches the uncertainty in these output samples. Therefore, the results for the KL divergence in Table 4.5 are biased in comparison to the true values based on the exact reference distribution. With the closeness of the outputs for all models in this problem as seen above, this bias leads to locally optimal model choices that may change when using a different set of output samples. Nevertheless, this effect is reduced by using a larger number of samples, and we suppose that in most problems this bias will be negligible in comparison to the error from the approximation of the model coupling.

Table 4.5: KL divergence estimates for decoupled satellite models based on control variate estimators for the mean and variance

	$\lambda = 1 \times 10^{-6}$	$\lambda = 1 \times 10^{-5}$	$\lambda = 1 \times 10^{-1}$
KL Divergence Estimate	$3.26 \pm 0.61 \times 10^{-3}$	$3.84 \pm 0.74 \times 10^{-3}$	$1.08 \pm 0.18 \times 10^{-2}$
Active Coupling Variables	8/8	5/8	1/8

4.2 Turbine Engine Model

A model for turbine engine cycle analysis was presented in Ref. [27]. The reference model for the system, M_0 , is displayed in Figure 4-34 and consists of 13 disciplines for the 12 engine components and a performance function, which are connected by 22 coupling variables. These disciplines model the engine with a core air-stream passing through a fan, compressor, burner and turbine as well as a second stream of air that bypasses the engine core. For each set of inputs that include the fan pressure ratio (FPR), compressor pressure ratio (CPR), bypass ratio (BPR), and mass-flow rate (W), the system's disciplines iteratively solve for 4 output quantities of interest: thrust-specific fuel consumption ($TSFC$), net thrust (F_n), overall pressure ratio (OPR), and engine burner temperature (T_4).

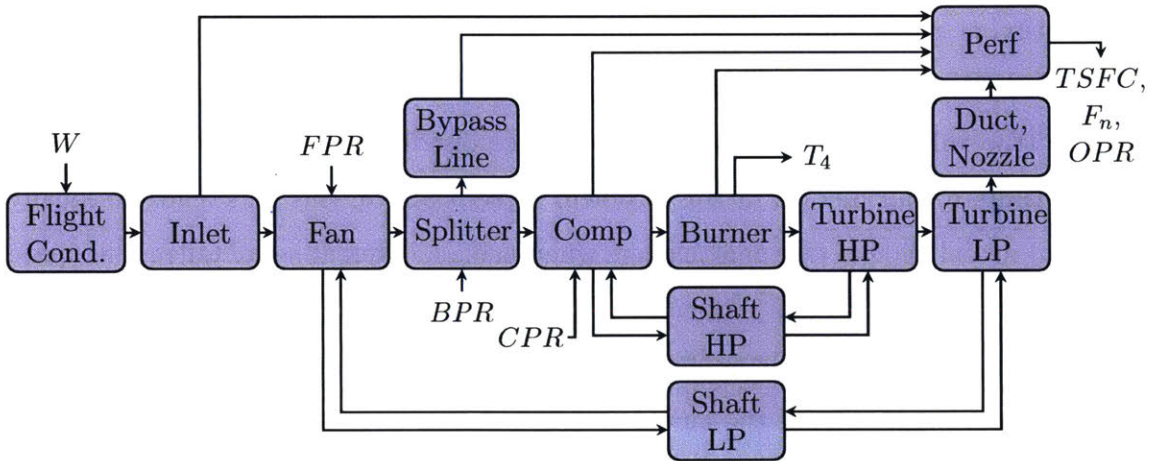


Figure 4-34: Turbine engine cycle model from Ref. [27]

To analyze various operating conditions, the model inputs are represented with Gaussian distributed random variables whose parameters are listed in Table 4.6.

By propagating 10^3 Monte Carlo samples of these inputs through the nonlinear system, the non-Gaussian distribution of the output variables, $\pi_{\mathbf{f}}$, is empirically characterized for the selected QoI in Figure 4-35. In the optimal model selection algorithm, we use the linearizations of these output variables with respect to the inputs when fixing different subsets of the discipline couplings to their estimated mean values. We

Table 4.6: Random variable parameters in the turbine engine cycle model

Random Variable	Symbol	Mean	Standard Deviation
Fan pressure ratio	FPR	1.5	0.01
Compressor pressure ratio	CPR	15.0	0.10
Bypass ratio	BPR	2.0	0.10
Mass-flow rate	W	1000.0	$\sqrt{10.0}$

then compare the resulting multivariate Gaussian approximation for the outputs of each linearized decoupled model, $\pi_{\tilde{\mathbf{f}}_M}$, to the output distribution of the linearized reference model, $\pi_{\tilde{\mathbf{f}}_{M_0}}$.

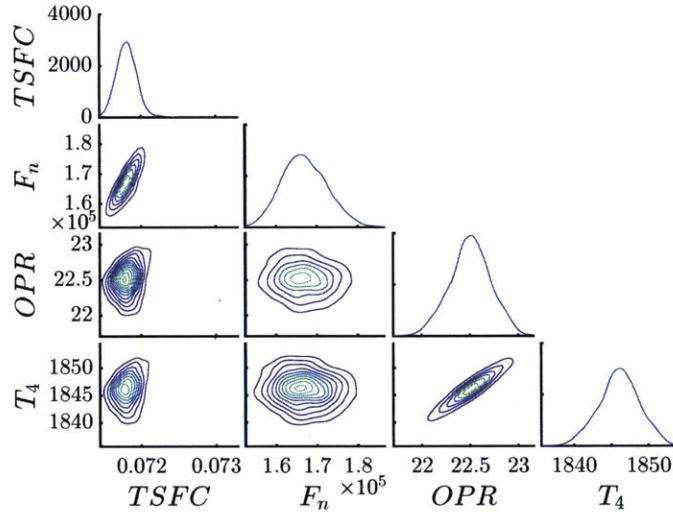


Figure 4-35: Joint distribution of the reference model

Applying the model selection procedure described above with $L = 1000$ particles, the SMC algorithm identifies decoupled models for four different values of λ in $\Lambda = [10^{-1}, 10^0, 10^1, 10^3]$. These models are presented in Figures 4-36 to 4-39 with the decoupled connections indicated by dashed lines. Similarly to the satellite model, we observe that with an increase in λ , the optimal solutions to the combinatorial optimization problem have greater sparsity in the number of discipline couplings. This results in a set of models that sequentially remove weak connections to increase the sparsity of couplings in exchange for accuracy in the output distributions. For this

turbine engine case study with $\lambda = 10^{-1}$, and $\lambda = 10^0$, this results in neglecting the effect of the feedback from the low-pressure turbine on the fan and the variability of inlet ram drag on system performance.

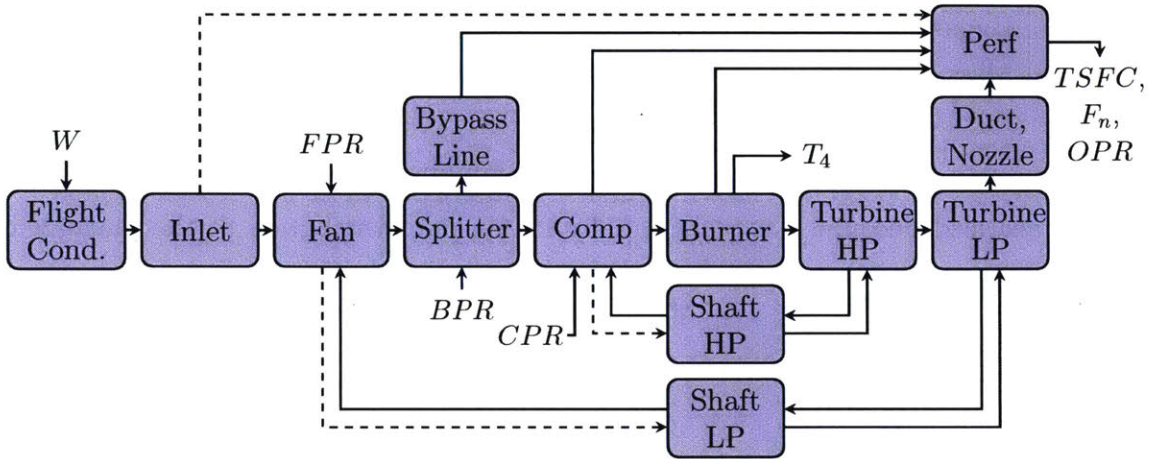


Figure 4-36: Optimal model coupling for $\lambda = 1.0 \times 10^{-1}$

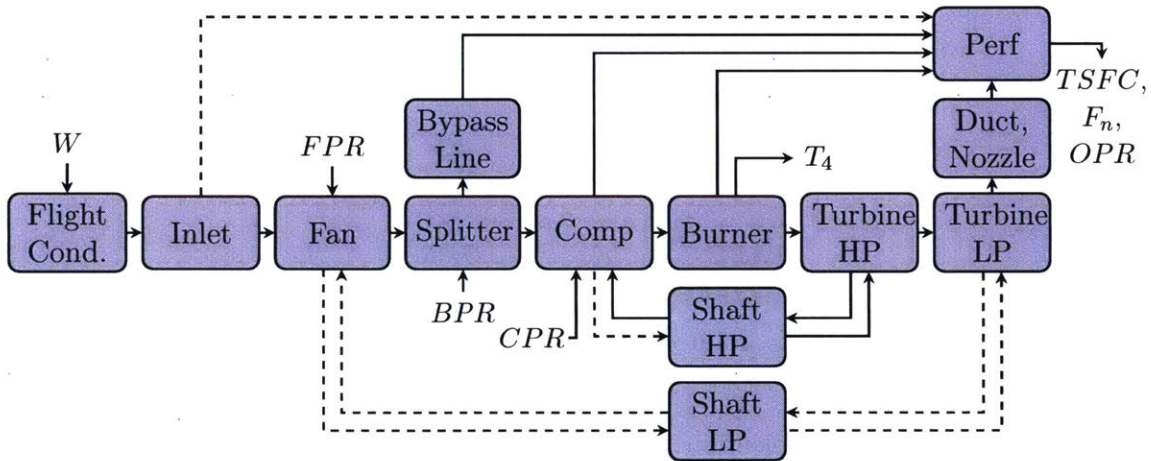


Figure 4-37: Optimal model coupling for $\lambda = 1.0 \times 10^0$

With a further increase in λ and a greater penalty on low sparsity models, the effect of the bypass line and the feedback from the high-pressure turbine are also neglected in the optimal models presented below. The accuracy of the system outputs corresponding to the decoupled model for each value of λ is displayed in the joint distributions of the output QoI in Figures 4-40 to 4-43. We observe that while the

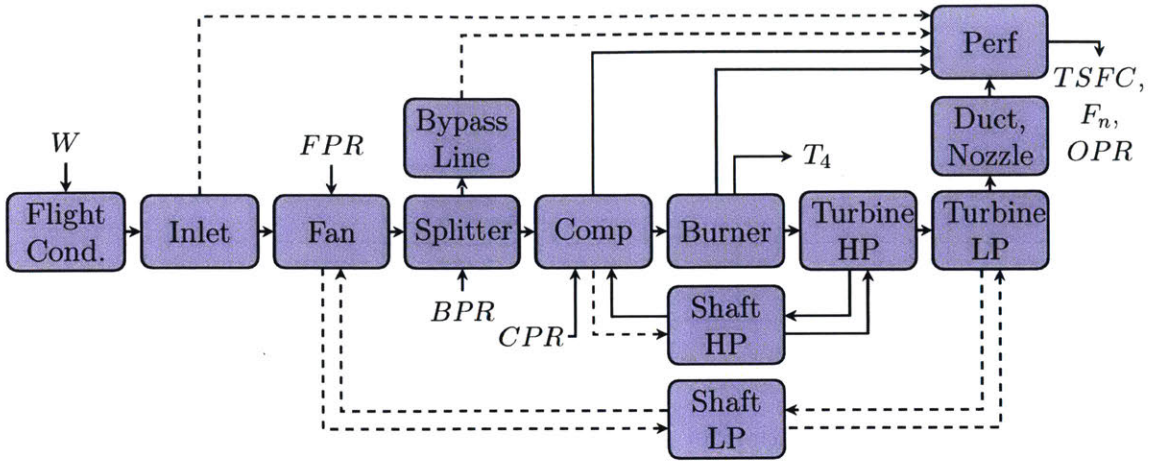


Figure 4-38: Optimal model coupling for $\lambda = 1.0 \times 10^1$

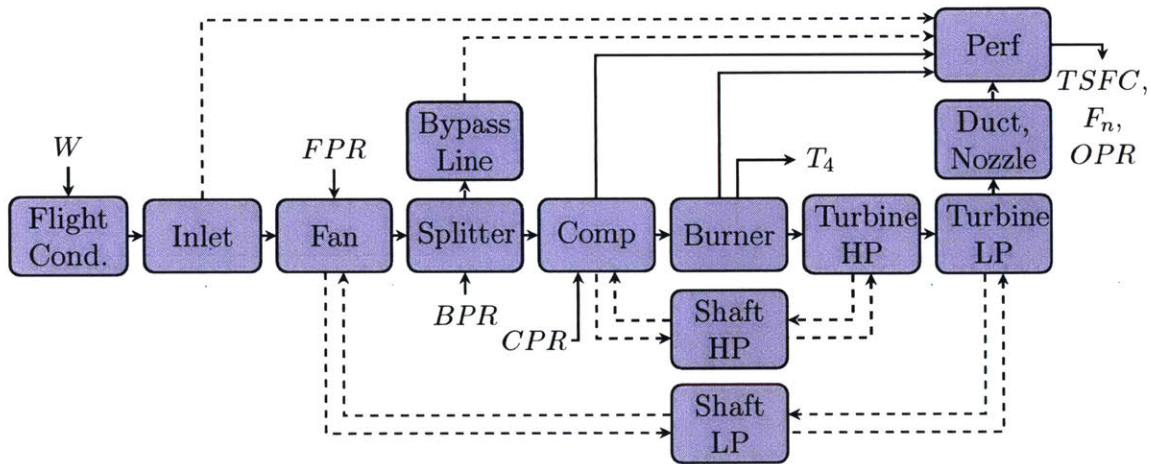


Figure 4-39: Optimal model coupling for $\lambda = 1.0 \times 10^3$

distribution for the output $TSFC$ variable is closely captured with the decoupled models for $\lambda = 10^{-1}$, and 10^0 ; the correlations of this variable are not as accurately represented when decoupling connections from the bypass line and feedback from the high-pressure turbine. This effect on the correlation is expected given the tight balance between the bypass flow and the core flow through the high pressure turbine, that together produce most of the engine's thrust.

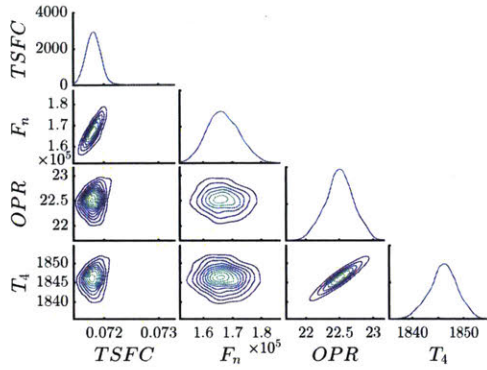


Figure 4-40: Joint distribution for $\lambda = 1.0 \times 10^{-1}$

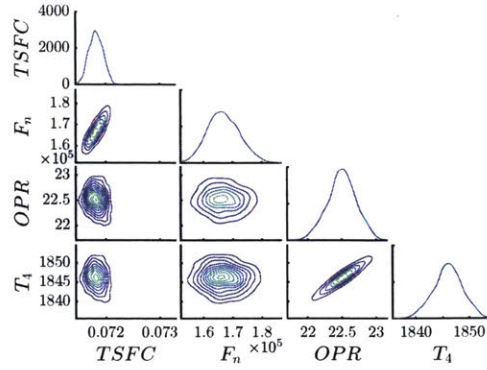


Figure 4-41: Joint distribution for $\lambda = 1.0 \times 10^0$

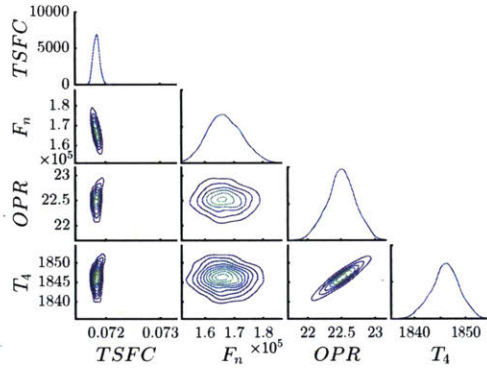


Figure 4-42: Joint distribution for $\lambda = 1.0 \times 10^1$

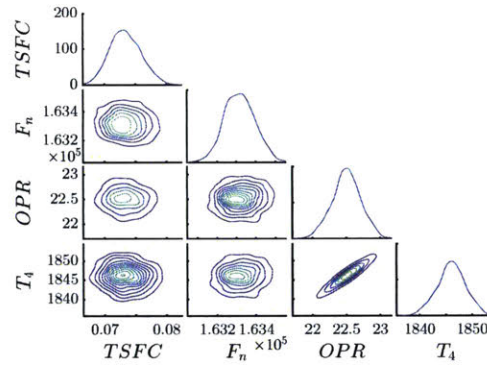


Figure 4-43: Joint distribution for $\lambda = 1.0 \times 10^3$

Furthermore, while the first three models accurately capture the net thrust, F_n , the uncertainty in this variable is poorly represented with the last model when neglecting the contribution of the nozzle. Finally, all four models accurately represent the uncertainty in the overall pressure ratio, OPR , and engine burner temperature, T_4 . As a result, depending on the application and available computational resources, users may select an appropriate set of discipline couplings for the turbine engine to model the uncertainty in the system's output variables.

To quantitatively compare the decoupled models, we also present the linearized KL divergence between the approximating Gaussian distributions for the reference and optimally decoupled models of the turbine engine in Table 4.7. These results are

presented along with the number of active coupling variables in each model (i.e., solid lines in the model diagrams above) that demonstrates the increasing sparsity in the decoupled models for larger values of λ .

Table 4.7: Linearized KL divergence for decoupled turbine engine models

	$\lambda = 1 \times 10^{-1}$	$\lambda = 1 \times 10^0$	$\lambda = 1 \times 10^1$	$\lambda = 1 \times 10^3$
Linearized KL Divergence	6.95×10^{-10}	1.15×10^0	1.25×10^1	3.73×10^3
Active Coupling Variables	19/22	16/22	14/22	7/22

To further compare the models, the pointwise errors of the three outputs from the performance discipline ($TSFC$, F_n , and OPR) for the decoupled models corresponding to $\lambda = 10^{-1}$, 10^0 , and 10^1 are plotted in Figures 4-44 to 4-46. These figures demonstrate the increasing errors in the outputs of the decoupled models relative to the reference model at the same inputs for 10^3 samples with the reduction in the number of discipline couplings. With the exception of the $TSFC$ output variable in the final model (Figure 4-46), the trends for the output variables are closely captured by the decoupled models associated with $\lambda = 10^{-1}$, and 10^0 . Nevertheless, users with different requirements may still trade-off model sparsity and accuracy for performing computations, such as multidisciplinary optimization, with these two decoupled models. While the optimal model for $\lambda = 10^0$ eliminates one feedback loop from the low pressure turbine leading to lower computational costs than the model for $\lambda = 10^{-1}$, there is a trade-off with accuracy, as seen with the larger spread in the pointwise errors of the $TSFC$ variable.

Furthermore, these results for the pointwise error highlight the importance of the bypass line to accurately estimate the $TSFC$ when computing the turbine's performance. We note that this coupling is included in the models corresponding to $\lambda = 10^{-1}$, and 10^0 .

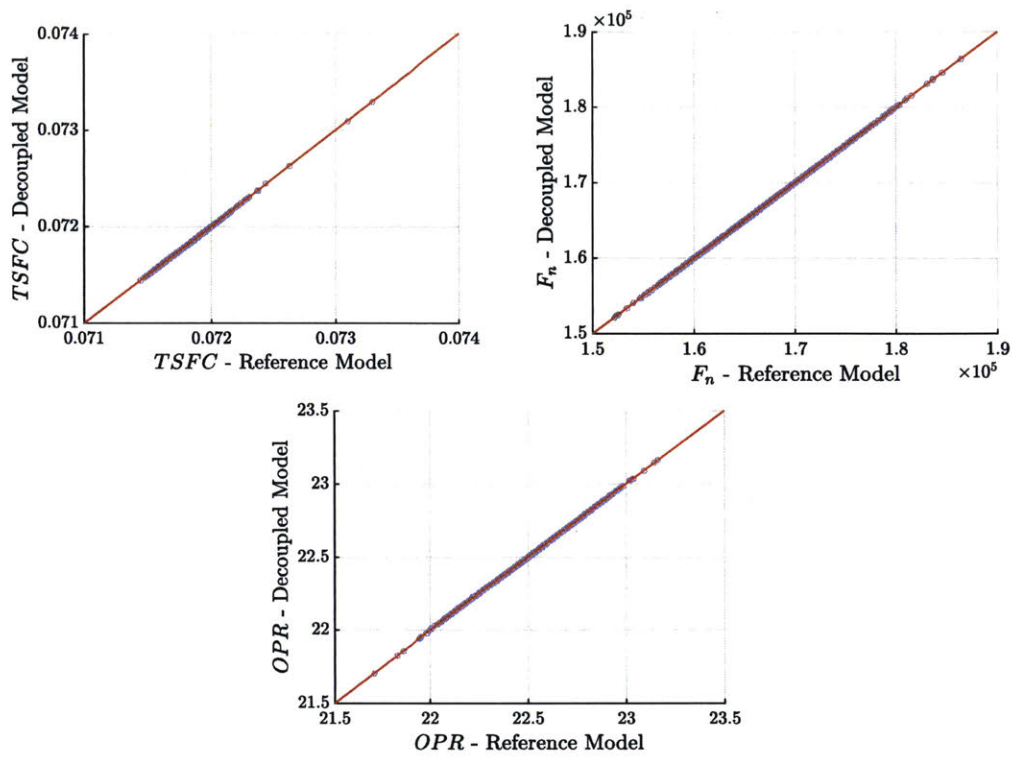


Figure 4-44: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-1}$

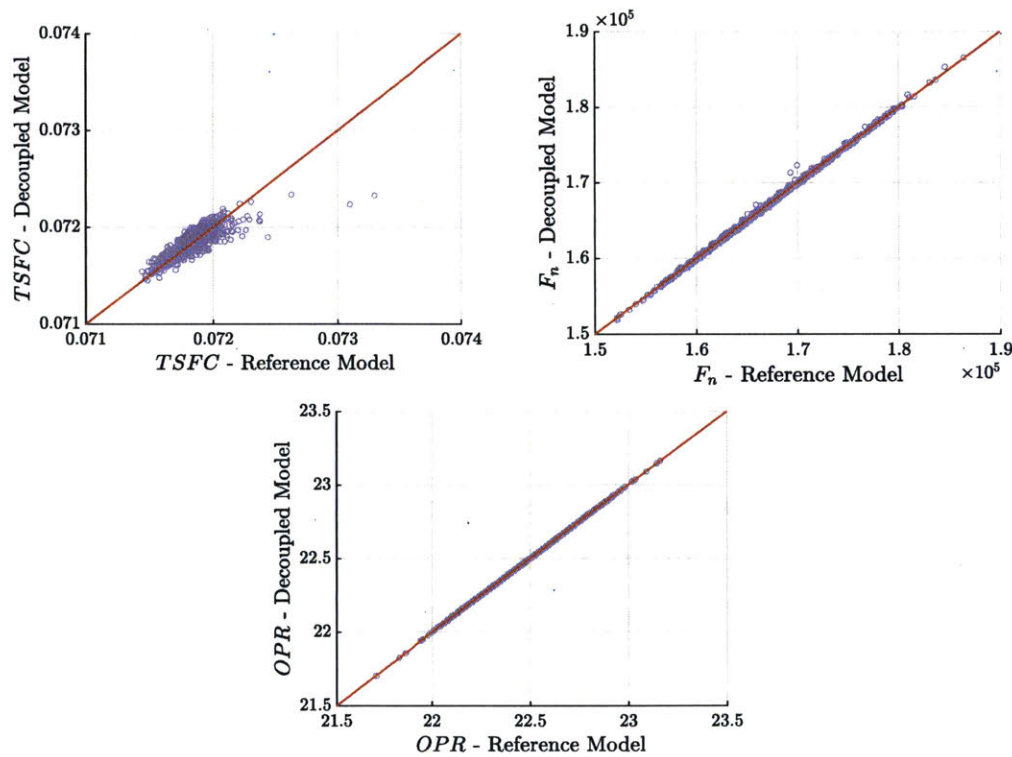


Figure 4-45: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^0$

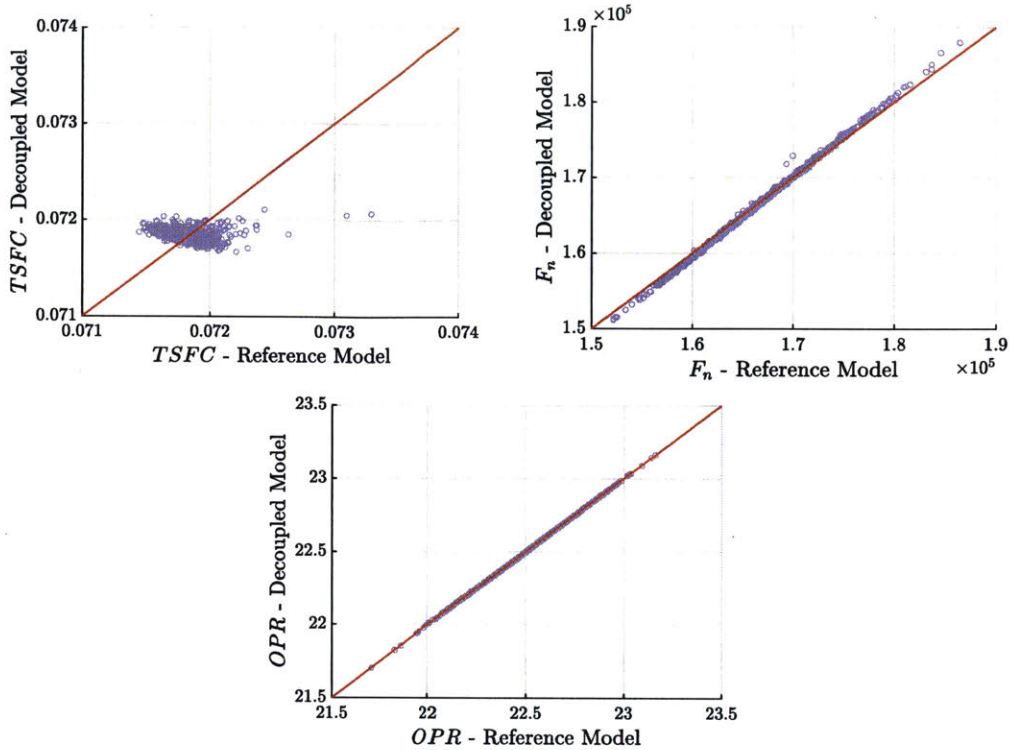


Figure 4-46: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^1$

Finally, in addition to producing a sequence of sparsely coupled models with a reduced number of discipline couplings relative to M_0 , the model selection algorithm efficiently found all of the optimally decoupled models by visiting less than 5.7% of the model space. This is contrasted with the otherwise intractable process of enumerating all possible models to find the optimum of the combinatorial optimization problem posed in equation (2.5).

4.3 Aero-Structural Lifting Surface Model

The *OpenAeroStruct* model was developed by the University of Michigan's MDOLab [29] to perform low-fidelity coupled aerodynamic and structural analyses for optimization of a wing, such as the one displayed in Figure 4-47. While the aerodynamics model uses a vortex lattice method to output the loads acting on the lifting surface, the structural analysis uses a finite element model of a truss and beam element to output the wing deflections at a set of spatial locations based on the aerodynamic

loads acting on the surface.

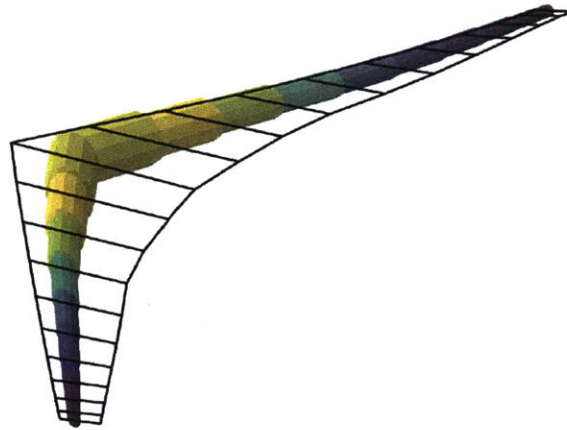


Figure 4-47: Wing planform from Ref. [29]

The reference model for the coupled aero-structural analysis is seen in Figure 4-48, where the output discipline computes the global quantities of interest: lift coefficient (C_L), failure criteria metric (F_c) based on a Kreisselmeier-Steinhauser function, and weight-normalized fuelburn (Fuel). These outputs are computed for each set of the input conditions as defined by: the span (span), Mach number (M_a), angle of attack (α), and the air density (ρ) and speed of sound (a) that define the wing's altitude.

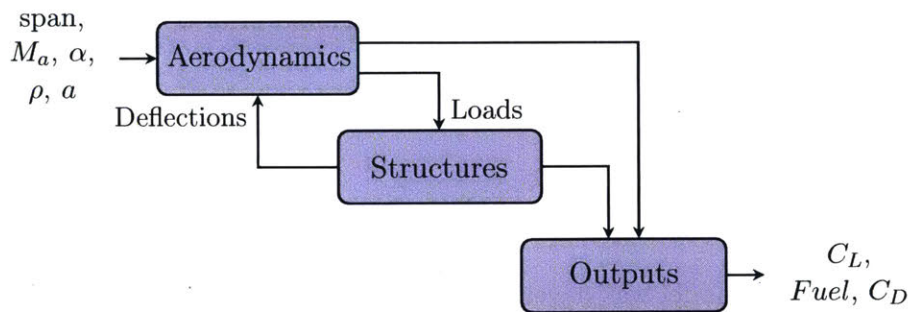


Figure 4-48: Aero-structural lifting surface model from Ref. [29]

In this study, we represent the input parameters as independent Gaussian distributed random variables with the mean and variance parameters in Table 4.8.

To propagate the uncertainty through the feedback coupled model to the global

Table 4.8: Random variable parameters in the aero-structural lifting surface model

Random Variable	Symbol	Mean	Standard Deviation
Span	span	59.0 m	1.0 m
Mach number	M_a	0.84	0.1
Angle of attack	α	3.0°	$\sqrt{0.1}^\circ$
Air density	ρ	0.38	0.1
Speed of sound	a	295.4	1

QoIs, the aerodynamics and structures disciplines iterate until convergence of the coupling variables, up to a specified relative tolerance of 10^{-7} for the residual equations. These couplings variables include an array of force and moment loads on the wing that are computed by the aerodynamics discipline, and the wing deflections computed by the structure discipline. To describe these fields we use a coarse discretization of the wing surface that is displayed in Figure 4-49, where the blue nodes represent a set of 7 span-wise points at 0.35 of the local chord where the loads are computed, and the black nodes represent a total of 14 points on the leading and trailing edges where the mesh deflections are computed. These 7 node locations for the loads and 14 node locations for the mesh deflections define a total of 21 coupling variables that we analyze with the model selection algorithms described in Chapter 3.

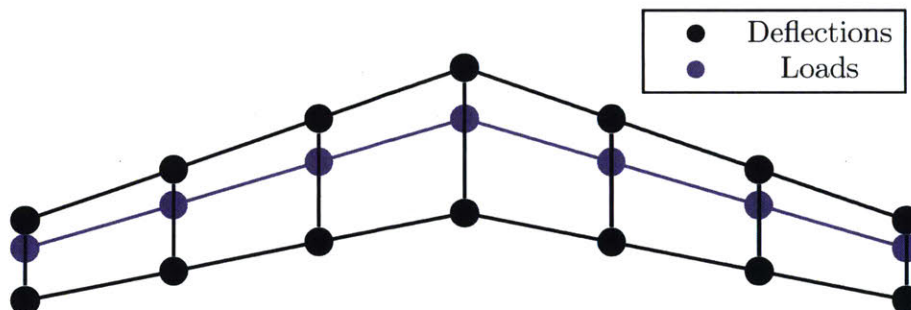


Figure 4-49: Spatial locations for coupling variables in the aero-structural model

By propagating 10^4 Monte Carlo samples of the input variables through the non-linear system, the non-Gaussian distribution of the output variables is empirically characterized for the selected QoI in Figure 4-50.

Applying the optimal model selection algorithm based on the linearizations of

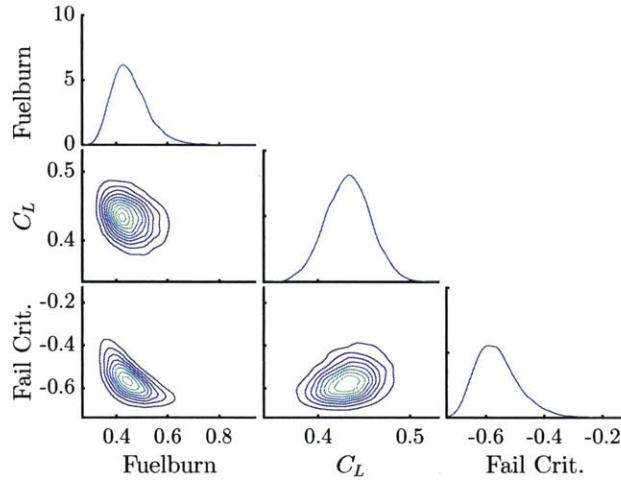


Figure 4-50: Joint distribution of the reference model

these output variables, the SMC algorithm with $L = 1000$ particles identifies decoupled models for four different values of λ in $\Lambda = [10^{-6}, 10^{-3}, 10^{-1}, 10^1]$ by visiting less than 3.4% of the model space. The optimal models along with the joint distributions of their output QoIs are presented in Figures 4-51 to 4-58, with the decoupled connections indicated by lighter nodes in each figure. The neglected coupling variables use the first-order mean as fixed input values for the loads and mesh displacements in the structures and aerodynamics disciplines, respectively.

In Table 4.9, we present the values for the linearized KL divergence between the reference model outputs, and each optimally decoupled model, M^* . Similarly to the

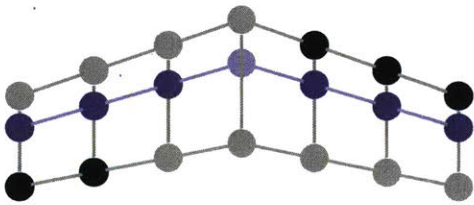


Figure 4-51: Optimal model coupling for $\lambda = 1.0 \times 10^{-6}$

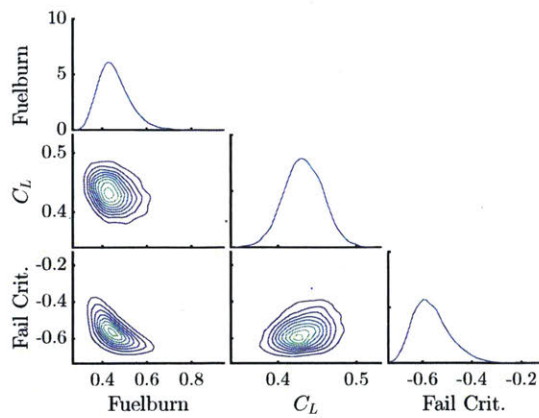


Figure 4-52: Joint distribution for $\lambda = 1.0 \times 10^{-6}$

previous test cases, the number of active coupling variables is also presented for each decoupled model in the table below.

Table 4.9: Linearized KL divergence for decoupled aero-structural models

	$\lambda = 1 \times 10^{-6}$	$\lambda = 1 \times 10^{-3}$	$\lambda = 1 \times 10^{-1}$	$\lambda = 1 \times 10^1$
Linearized KL Divergence	8.04×10^{-7}	2.28×10^{-4}	2.54×10^{-2}	1.81×10^0
Active Coupling Variables	11/21	6/21	4/21	2/21

For the aero-structural model with the specified global QoIs, the resulting optimal model for $\lambda = 10^{-6}$ neglects the effect of most mesh deflections as inputs to the aerodynamics discipline, particularly near the root of the wing, given their small values with the specified wing material properties. As a result, for this model it is sufficient to only couple the mesh deflections on both ends of the wing, the leading edge of the starboard wing tip and the trailing edge of the port-side wing tip. With an increase in λ values and a greater penalty on finding sparse models, it is sufficient to consider the effect of all first-order mean mesh deflections to characterize the output uncertainty. This transforms the feedback coupled model into a feed-forward model that computes the aerodynamic loads based on the fixed deflections and uses only the updated loads to assess the deflections when computing the global QoIs. This allows a user to cheaply propagate the uncertain inputs through the system (i.e., using only one iteration for the feed-forward system) with varying accuracy, based on the number of coupling variables that are not fixed to their first order mean values for the aerodynamic loads.

Finally, we compare the pointwise errors of the three output QoIs in the optimally decoupled models corresponding to $\lambda = 10^{-6}, 10^{-3}, 10^{-1}$, and 10^1 . Figures 4-59 to 4-62 display these errors and increasing bias in the QoIs of each decoupled model in comparison to the reference model (particularly with regards to the failure criteria), by evaluating the models at the same inputs for 10^4 samples.

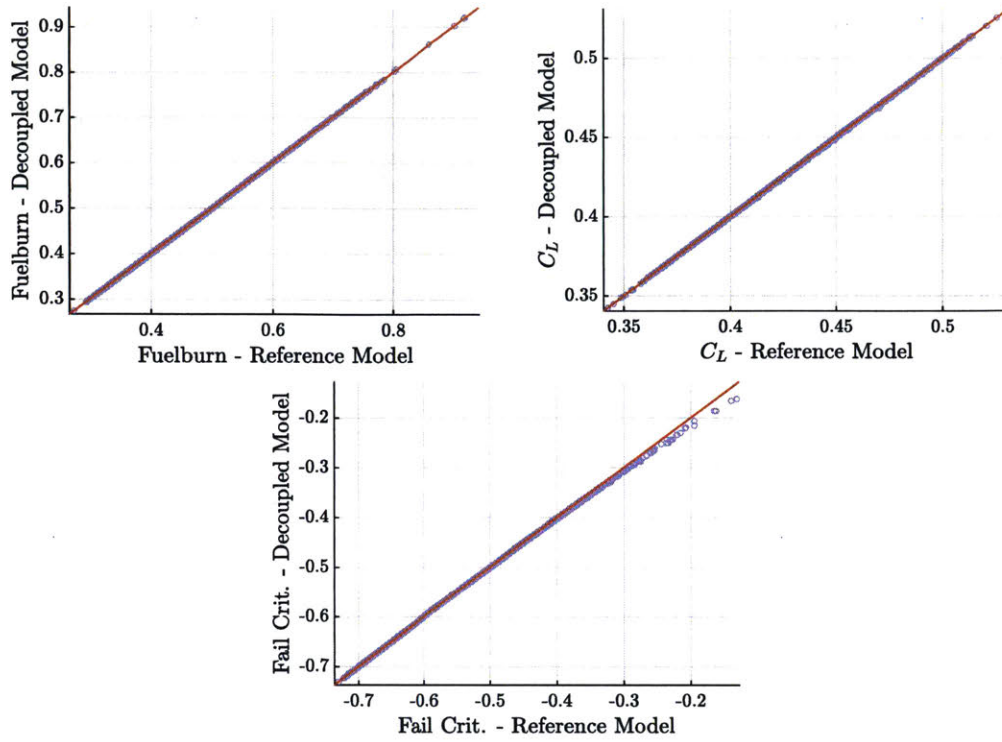


Figure 4-59: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-6}$

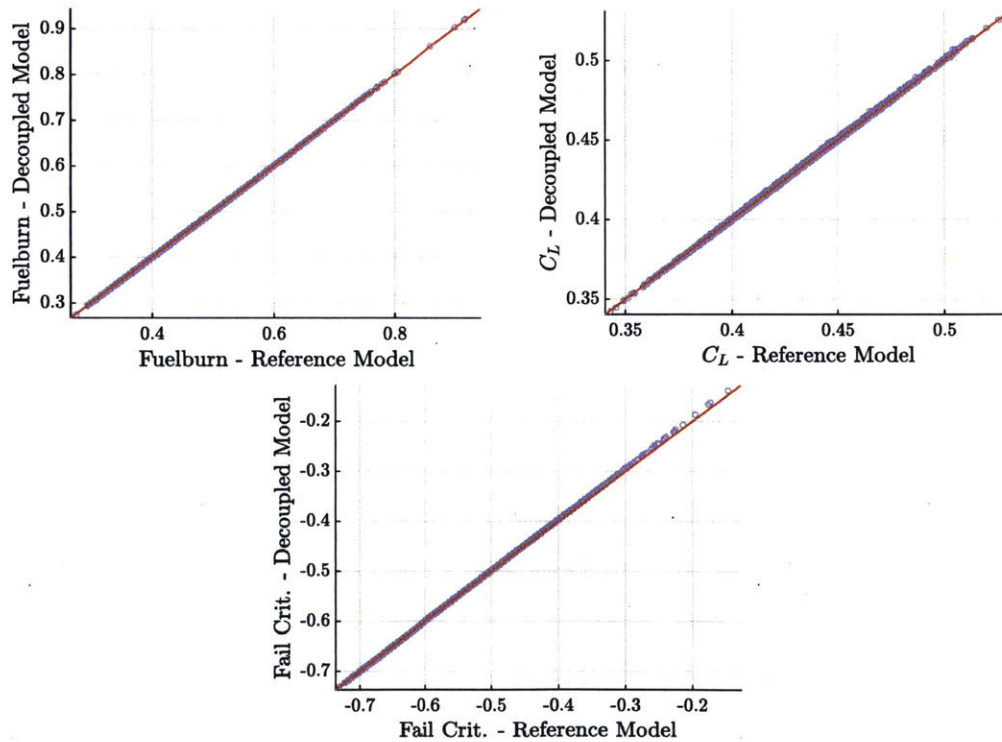


Figure 4-60: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-3}$

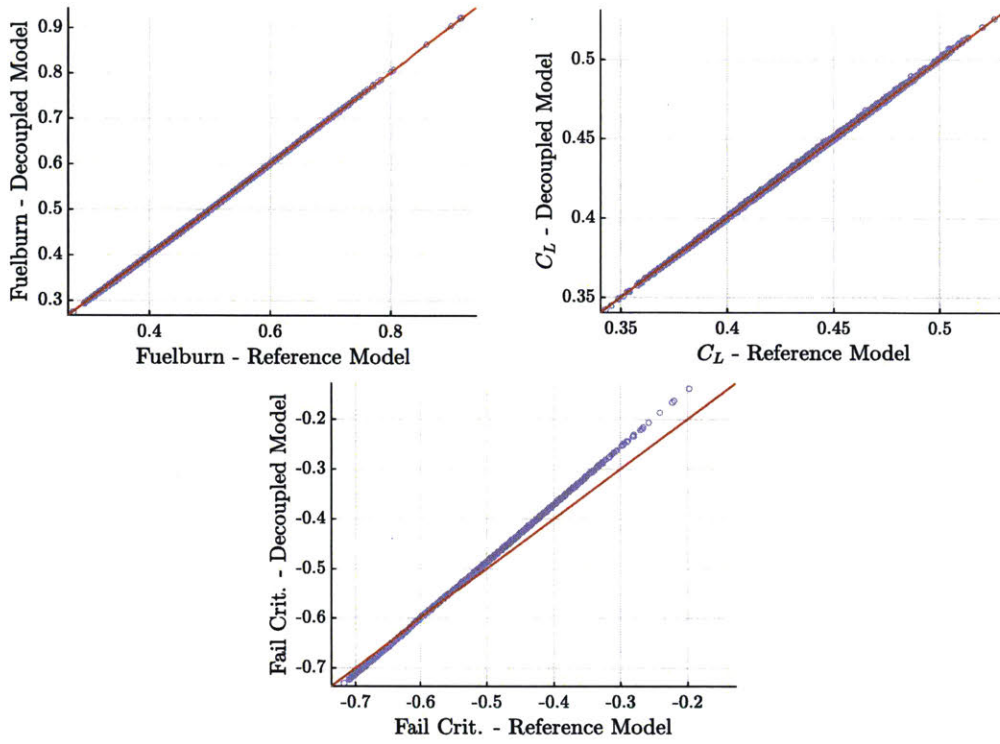


Figure 4-61: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^{-1}$

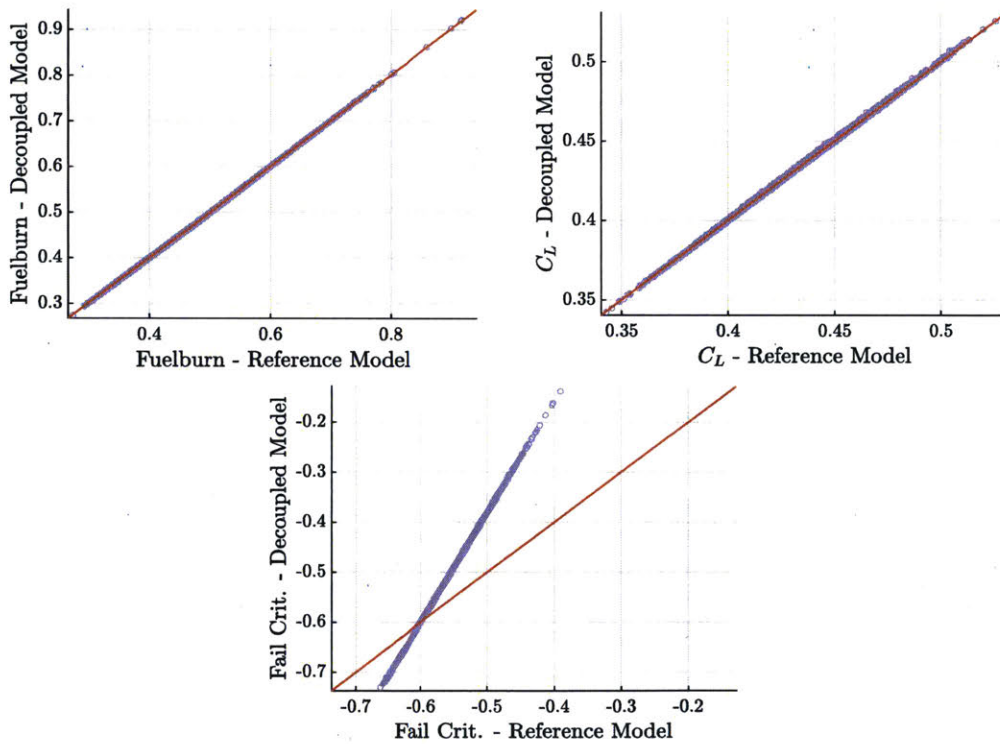


Figure 4-62: Comparison of QoI with optimally decoupled model for $\lambda = 1.0 \times 10^1$

We note that while the feed-forward model corresponding to $\lambda = 10^{-3}$ accurately captures the outputs with relatively small variability, a bias is observed in the failure criteria of the decoupled models for $\lambda = 10^{-1}$ and 10^1 when fixing the coupling variables for both the feedback in the mesh displacements and a subset of the loads. This effect is a result of most converged samples having lower absolute values for the forces and moments than their first-order mean (that are used as the fixed nominal values for these coupling variables), as seen in Figures 4-63 and 4-64 for the forces in the spanwise and chordwise directions at one point along the 0.35 · chord line of the wing. By using greater absolute values for the loads, the beam’s stress is over-predicted and under-predicted for different input conditions, leading to the observed shift in the failure criteria and poor predictive performance of this QoI in these decoupled models.

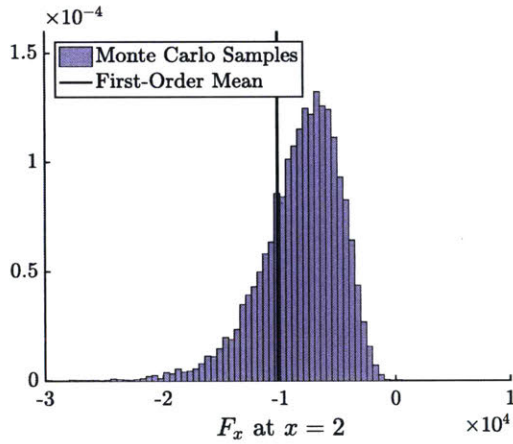


Figure 4-63: Probability distribution of chordwise loads

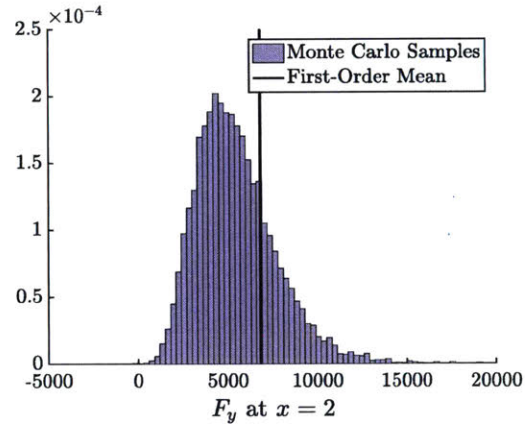


Figure 4-64: Probability distribution of spanwise loads

As in the satellite model, we suppose that the optimal coupling will also change under different input uncertainties. In particular, under different material properties (i.e., a less stiff material for the wing with lower Young’s modulus and shear modulus values), the feedback coupling may be necessary even for larger values of λ to accurately describe the load distributions as a function of the mesh displacements.

Chapter 5

Analysis of Approximate Coupling

In this chapter, we motivate the error analysis for decoupling linear multidisciplinary models in Sections 5.1 and 5.2. We present the pointwise error and the error in the uncertainty of the output variables for both feed-forward and feedback systems in Sections 5.3 and 5.4, respectively. Section 5.5 discusses the connections between decoupling in multidisciplinary models and results for the sparsification of probabilistic graphical models.

5.1 Linear Systems

For a coupled nonlinear problem, $\mathcal{R}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ and $\mathbf{f} = \mathcal{F}(\mathbf{x}, \mathbf{y})$, as described in Section 3.1, the linearization of the coupled model at $\mu_{\mathbf{x}}, \mathbf{y}(\mu_{\mathbf{x}})$ is given by $\mathcal{R}(\mathbf{x}, \mathbf{y}) \approx \frac{\partial \mathcal{R}}{\partial \mathbf{x}}(\mathbf{x} - \mu_{\mathbf{x}}) + \frac{\partial \mathcal{R}}{\partial \mathbf{y}}(\mathbf{y} - \mathbf{y}(\mu_{\mathbf{x}})) = \mathbf{0}$, and can be written as $\mathbf{A}\mathbf{y} = \mathbf{b}$ where $\mathbf{A} = \frac{\partial \mathcal{R}}{\partial \mathbf{y}}$ and $\mathbf{b} = \frac{\partial \mathcal{R}}{\partial \mathbf{y}}\mathbf{y}(\mu_{\mathbf{x}}) - \frac{\partial \mathcal{R}}{\partial \mathbf{x}}(\mathbf{x} - \mu_{\mathbf{x}})$. We note that this formulation is exact in the case when the system is linear (i.e., the outputs are linear functions of the input variables, \mathbf{x}).

Decoupling disciplines in the model reduces the inputs of each nonlinear equation by fixing a subset of the coupling variables. This reduction in the dependencies of each discipline leads to added sparsity in components of the Jacobian matrix, \mathbf{A} , and adds an extra term to \mathbf{b} . As a result, it is of interest to first analyze decoupling by understanding these perturbations on linear systems.

5.2 Solution Uncertainty

In this section, we consider the simplified linear system, $\mathbf{A}\mathbf{y} = \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{R}^n$ is a random input, and $\mathbf{y} \in \mathbb{R}^n$ is the solution vector. If \mathbf{b} is normally distributed (i.e., $\mathbf{b} \sim \mathcal{N}(\mu, \Sigma)$), the solution has the distribution

$$\mathbf{y} \sim \mathcal{N}(\mathbf{A}^{-1}\mu, \mathbf{A}^{-1}\Sigma\mathbf{A}^{-T}). \quad (5.1)$$

A quantity of interest (QoI) in a linear system is typically specified by $\mathbf{f} = \mathbf{C}\mathbf{y}$ for $\mathbf{C} \in \mathbb{R}^{p \times n}$ and $\mathbf{f} \in \mathbb{R}^p$. In this case, \mathbf{f} also has a normal distribution with mean and covariance that are scaled by \mathbf{C} .

Each row i in the linear system corresponds to a discipline equation that computes one output, \mathbf{y}_i , as a function of the remaining variables. Decoupling one input, \mathbf{y}_j , from this equation is given by fixing this input in the equation $\mathbf{A}_{ii}\mathbf{y}_i = \mathbf{b}_i - \sum_{k \neq i} \mathbf{A}_{ik}\mathbf{y}_k$ to a nominal value, $\bar{\mathbf{y}}_j$. In doing so, we modify the original system of equations by setting the off-diagonal entry \mathbf{A}_{ij} to zero and adding the contribution of the fixed value, $\mathbf{A}_{ij}\bar{\mathbf{y}}_j$, to the vector \mathbf{b} .

The perturbed system of equations when decoupling input j from discipline i is given by $\tilde{\mathbf{A}}\tilde{\mathbf{y}} = \tilde{\mathbf{b}}$ with $\tilde{\mathbf{A}} = \mathbf{A} - \mathbf{A}_{ij}\mathbf{E}_{ij}$ and $\tilde{\mathbf{b}} = \mathbf{b} - \mathbf{A}_{ij}\bar{\mathbf{y}}_j\mathbf{e}_i$, where $\mathbf{E}_{ij} = \mathbf{e}_i\mathbf{e}_j^T$ is a rank-one matrix from the outer product of two unit vectors. This constant change to the right hand side shifts the mean of the random vector to $\tilde{\mathbf{b}} \sim \mathcal{N}(\mu - \mathbf{A}_{ij}\bar{\mathbf{y}}_j\mathbf{e}_i, \Sigma)$. As a result, the distribution of the solution to the perturbed system is given by

$$\tilde{\mathbf{y}} \sim \mathcal{N}(\tilde{\mathbf{A}}^{-1}(\mu - \mathbf{A}_{ij}\bar{\mathbf{y}}_j\mathbf{e}_i), \tilde{\mathbf{A}}^{-1}\Sigma\tilde{\mathbf{A}}^{-T}). \quad (5.2)$$

In the next section we analyze the error in the mean and covariance of the solution uncertainty with the approximate model. We will consider both feed-forward and feedback coupled problems separately.

5.3 Feed-forward Systems

If the model is feed-forward, all equations only depend on previously computed outputs, and so the linear system of equations has a lower-triangular matrix, \mathbf{A} . In this case, the solution is computed by forward-substitution where $\mathbf{y}_1 = \mathbf{b}_1/\mathbf{A}_{11}$, and $\mathbf{y}_k = (\mathbf{b}_k - \sum_{i < k} \mathbf{A}_{ki}\mathbf{y}_i)/\mathbf{A}_{kk}$ for $k > 1$.

When fixing input j in row i for $j < i$, the previous computations for variables $k < i$ are unaffected. Therefore $\mathbf{y}_k - \tilde{\mathbf{y}}_k = 0$ for $k < i$, and in particular $\dot{\mathbf{y}}_j = \tilde{\mathbf{y}}_j$. The solution of the remaining variables in the perturbed system in this case is given by

$$\begin{aligned} (\mathbf{A} - \mathbf{A}_{ij}\mathbf{E}_{ij})\tilde{\mathbf{y}} &= \mathbf{b} - \mathbf{A}_{ij}\tilde{\mathbf{y}}_j\mathbf{e}_i \\ \mathbf{A}\tilde{\mathbf{y}} - \mathbf{A}_{ij}\mathbf{E}_{ij}\tilde{\mathbf{y}} &= \mathbf{A}\mathbf{y} - \mathbf{A}_{ij}\tilde{\mathbf{y}}_j\mathbf{e}_i. \end{aligned} \tag{5.3}$$

Rearranging the terms in the equation above, the error is given by the linear system,

$$\begin{aligned} \mathbf{A}\tilde{\mathbf{y}} - \mathbf{A}\mathbf{y} &= \mathbf{A}_{ij}\mathbf{E}_{ij}\tilde{\mathbf{y}} - \mathbf{A}_{ij}\tilde{\mathbf{y}}_j\mathbf{e}_i \\ \mathbf{A}(\tilde{\mathbf{y}} - \mathbf{y}) &= \mathbf{A}_{ij}(\tilde{\mathbf{y}}_j\mathbf{e}_i - \tilde{\mathbf{y}}_j\mathbf{e}_i). \end{aligned} \tag{5.4}$$

Noting that $\tilde{\mathbf{y}}_j = \mathbf{y}_j$, the error of each entry in the solution for $k \geq i$ is given by

$$\tilde{\mathbf{y}}_k - \mathbf{y}_k = (\mathbf{A}^{-1})_{k,i}\mathbf{A}_{ij}(\mathbf{y}_j - \tilde{\mathbf{y}}_j). \tag{5.5}$$

Therefore, the error in variable k in the perturbed system is proportional to the deviation in the fixed value, $\tilde{\mathbf{y}}_j$, from the true value for variable j in the original system, the coefficient \mathbf{A}_{ij} and the effect of the i th input on the solution (based on entry (k, i) in the inverse of matrix \mathbf{A}).

We can also analyze the error in the QoI, that is given by $\tilde{\mathbf{f}} - \mathbf{f} = \mathbf{C}(\mathbf{A}^{-1})_{\cdot,i}\mathbf{A}_{ij}(\mathbf{y}_j - \tilde{\mathbf{y}}_j)$. If $\mathbf{f} = \mathbf{y}_m$ for some $m > i$, then $\tilde{\mathbf{f}} - \mathbf{f} = (\mathbf{A}^{-1})_{m,i}\mathbf{A}_{ij}(\mathbf{y}_j - \tilde{\mathbf{y}}_j)$. Finally, for triangular matrices, the eigenvalues of the matrix are given by its diagonal entries. Therefore, under perturbations to the off-diagonal elements of A , the system remains invertible (i.e., the diagonal entries and eigenvalues of A remain constant) and so a solution exists after decoupling.

We now analyze the effect of decoupling on the distribution of the solution variables, \mathbf{y} . The shift to the mean of the solution uncertainty is given by $(\mathbf{A}^{-1})_{\cdot,i} \mathbf{A}_{ij} (\mathbf{y}_j - \bar{\mathbf{y}}_j)$. By using the Sherman-Woodbury matrix identity, $\tilde{\mathbf{A}}^{-1} = [\mathbf{A} - \mathbf{A}_{ij} \mathbf{E}_{ij}]^{-1} = \mathbf{A}^{-1} - \mathbf{B}/\alpha$ where $\mathbf{B} = (\mathbf{A}^{-1})_{\cdot,i} (\mathbf{A}^{-1})_{j,\cdot}$ and $\alpha = -1/\mathbf{A}_{ij} + (\mathbf{A}^{-1})_{jj}$. The change in the covariance of the solution uncertainty is then given by

$$\tilde{\mathbf{A}}^{-1} \Sigma \tilde{\mathbf{A}}^{-T} - \mathbf{A}^{-1} \Sigma \mathbf{A}^{-T} = \mathbf{B} \Sigma \mathbf{B}^T / \alpha^2 - \mathbf{B} \Sigma \mathbf{A}^{-T} / \alpha - \mathbf{A}^{-1} \Sigma \mathbf{B}^T / \alpha. \quad (5.6)$$

This results in a dense update to all entries in the covariance matrix related to variables $k > i$ (i.e., the covariances between variables \mathbf{y}_k for $k < i$ are unaffected in a feed-forward model). On the other hand, the update to the inverse covariance matrix is the low-rank correction given by

$$\begin{aligned} \tilde{\mathbf{A}}^T \Sigma^{-1} \tilde{\mathbf{A}} - \mathbf{A} \Sigma^{-1} \mathbf{A} &= \mathbf{A}_{ij}^2 \mathbf{E}_{ij}^T \Sigma^{-1} \mathbf{E}_{ij} - \mathbf{A}_{ij} \mathbf{E}_{ij}^T \Sigma^{-1} \mathbf{A} - \mathbf{A}^T \Sigma^{-1} \mathbf{A}_{ij} \mathbf{E}_{ij} \\ &= \mathbf{A}_{ij}^2 (\Sigma^{-1})_{i,i} \mathbf{E}_{jj} - \mathbf{A}_{ij} \mathbf{e}_j (\Sigma^{-1})_{i,\cdot} \mathbf{A} - \mathbf{A}_{ij} \mathbf{A}^T (\Sigma^{-1})_{\cdot,i} \mathbf{e}_j, \end{aligned} \quad (5.7)$$

which updates row j , column j and the diagonal element (j, j) of the inverse covariance matrix. Therefore, only the partial correlations associated with variable j are updated. Furthermore, this indicates that any sparsity in the inverse covariance matrix associated with variable \mathbf{y}_j that is present in the original model is lost in the perturbed model, while the sparsity in this matrix between the remaining variables is preserved in the perturbed model. In Section 5.5, we expand on these changes to the sparsity of the inverse covariance matrix by making connections between sparsity and conditional independence.

5.4 Feedback Systems

In models with feedback, the linear system of equations has a potentially dense matrix, \mathbf{A} . For a single perturbation to element (i, j) in \mathbf{A} , the perturbed linear system is also given by: $\tilde{\mathbf{A}} \tilde{\mathbf{y}} = \tilde{\mathbf{b}}$ where $\tilde{\mathbf{A}} = \mathbf{A} - \mathbf{A}_{ij} \mathbf{E}_{ij}$ and $\tilde{\mathbf{b}} = \mathbf{b} - \mathbf{A}_{ij} \bar{\mathbf{y}}_j \mathbf{e}_j = \mathbf{b} - \mathbf{A}_{ij} \mathbf{E}_{ij} \bar{\mathbf{y}}$.

The error is then given by

$$\begin{aligned}\mathbf{A}(\tilde{\mathbf{y}} - \mathbf{y}) &= \mathbf{A}_{ij}\mathbf{E}_{ij}(\tilde{\mathbf{y}} - \bar{\mathbf{y}}) \\ &= \mathbf{A}_{ij}\mathbf{E}_{ij}(\tilde{\mathbf{y}} - \mathbf{y} + \mathbf{y} - \bar{\mathbf{y}}).\end{aligned}\tag{5.8}$$

Rearranging the terms and solving for the error term, $\tilde{\mathbf{y}} - \mathbf{y}$, we have

$$\begin{aligned}(\mathbf{A} - \mathbf{A}_{ij}\mathbf{E}_{ij})(\tilde{\mathbf{y}} - \mathbf{y}) &= \mathbf{A}_{ij}\mathbf{E}_{ij}(\mathbf{y} - \bar{\mathbf{y}}) \\ \tilde{\mathbf{y}} - \mathbf{y} &= (\mathbf{A} - \mathbf{A}_{ij}\mathbf{E}_{ij})^{-1}\mathbf{A}_{ij}\mathbf{E}_{ij}(\mathbf{y} - \bar{\mathbf{y}}).\end{aligned}\tag{5.9}$$

Using the Sherman-Woodbury identity to compute the inverse of A plus a low-rank perturbation, the simplified expression for the error is given by

$$\tilde{\mathbf{y}} - \mathbf{y} = (\mathbf{A}^{-1})_{\cdot,i} \left(\frac{\mathbf{A}_{ij}}{1 - \mathbf{A}_{ij}(\mathbf{A}^{-1})_{ji}} \right) (\mathbf{y}_j - \bar{\mathbf{y}}_j).\tag{5.10}$$

Therefore, the error for variable k is proportional to $(\mathbf{A}^{-1})_{ki}\mathbf{A}_{ij}/(1 - \mathbf{A}_{ij}(\mathbf{A}^{-1})_{ji})$ and the difference in the fixed nominal value, $\bar{\mathbf{y}}_j$, from the true value for variable j . We note that in comparison to the result in equation (5.5) for the error in the feed-forward case, the pointwise error of the solution vector in this case contains the additional term $(1 - \mathbf{A}_{ij}(\mathbf{A}^{-1})_{ji})$ in the denominator that accounts for the feedback effect of the perturbed variable \mathbf{y}_i on the computation of \mathbf{y}_j and its propagation throughout the system. If $(\mathbf{A}^{-1})_{ji} = 0$, then variable \mathbf{y}_i does not affect equation j and so we reduce to the feed-forward error calculation.

For the feedback case, the error in the covariance and inverse covariance matrices of the solution uncertainty are also given by equations (5.6) and (5.7), respectively. We note that unlike the feed-forward case, the covariances of all solution variables are updated given the dense inverse of the perturbed matrix $\tilde{\mathbf{A}}$ for \mathbf{A} that has nonzero entries on both the upper and lower triangular parts of the matrix. On the other hand, the update to the inverse covariance matrix is still local to row j and column j , like the feed-forward case.

In the case of a feedback model, it is also important to quantify when it is possi-

ble to decouple discipline inputs (i.e., under what conditions does the linear system remain invertible after the perturbation caused by decoupling). One necessary condition for $\tilde{\mathbf{A}} = \mathbf{A} - \mathbf{A}_{ij}\mathbf{E}_{ij}$ to remain invertible is $\det(\tilde{\mathbf{A}}) \neq 0$. Using the matrix determinant lemma for a rank-1 update, $\det(\tilde{\mathbf{A}}) = \det(\mathbf{A})\det(1 - \mathbf{A}_{ij}(\mathbf{A}^{-1})_{ji})$. Assuming that the original system was invertible (i.e., $\det(\mathbf{A}) \neq 0$), then $\det(\tilde{\mathbf{A}}) \neq 0$ if and only if $1 \neq \mathbf{A}_{ij}(\mathbf{A}^{-1})_{ji}$.

Finally, we consider the case of decoupling multiple discipline couplings by adding several perturbations to the linear system. When decoupling the discipline couplings for every pair of entries in the set \mathcal{J} (where pair $(i, j) \in \mathcal{J}$ denotes the coupling of variable \mathbf{y}_j as an input to discipline i), the error in the solution using the same approach as in equation (5.9), is given by

$$\tilde{\mathbf{y}} - \mathbf{y} = \left(\mathbf{A} - \sum_{(i,j) \in \mathcal{J}} \mathbf{A}_{ij}\mathbf{E}_{ij} \right)^{-1} \left(\sum_{(i,j) \in \mathcal{J}} \mathbf{A}_{ij}\mathbf{E}_{ij} \right) (\mathbf{y} - \bar{\mathbf{y}}). \quad (5.11)$$

Here we note that the vector $\bar{\mathbf{y}} \in \mathbb{R}^n$ contains the values for all fixed inputs (assuming that a common input to multiple disciplines is fixed to the same value). Furthermore, we require that $\det(\tilde{\mathbf{A}}) = \det\left(\mathbf{A} - \sum_{(i,j) \in \mathcal{J}} \mathbf{A}_{ij}\mathbf{E}_{ij}\right) \neq 0$ for the system to remain invertible after the perturbations. Using the matrix determinant lemma, we can interpret this condition as

$$\det(\tilde{\mathbf{A}}) = \det(\mathbf{A}) \prod_{\substack{t=1 \\ (i,j) \in \mathcal{J}}}^{|\mathcal{J}|} \left[1 - \mathbf{e}_i \left(\mathbf{A} - \sum_{(k,l) \in \mathcal{J}^t} \mathbf{A}_{kl}\mathbf{E}_{kl} \right)^{-1} \mathbf{e}_j \right], \quad (5.12)$$

where \mathcal{J}^t denotes the remaining set of discipline couplings after having decoupled the first t connections in \mathcal{J} (i.e., $\mathcal{J}^t = \mathcal{J}^{t-1} \setminus (i, j)$ and $\mathcal{J}^0 = \mathcal{J}$). This indicates that for the system to be invertible, each term in the product in equation (5.12) must be non-zero, which holds true if the system is invertible after incrementally removing each discipline coupling. Therefore, by testing that the system is invertible after decoupling each pair (i, j) , we can guarantee that a solution will exist for the model after removing all connections in \mathcal{J} .

5.5 Connections to Probabilistic Graphical Models

In this section we analyze the error in the output uncertainty from the perspective of sparsifying probabilistic graphical models. After relating decoupling to conditional independence in graphical models in Section 5.5.1, we present results on sparsifying graphical models for Gaussian, exponential family and more general distributions in Sections 5.5.2, 5.5.3, and 5.5.4, respectively.

5.5.1 Decoupling and Conditional Independence

As alluded to in the previous section, when decoupling connections in linear systems, the perturbation creates a local effect on the inverse covariance matrix of the output distribution. For linear systems with Gaussian distributed output variables, the sparsity of the inverse covariance matrix is directly related to the conditional independence properties of the distribution. This sparsity is often represented with a probabilistic graphical model $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where the nodes, \mathcal{V} , represent random variables and the edges, \mathcal{E} , represent dependence between the corresponding variables. In this context, the lack of edges has the important meaning that the two corresponding variables are independent conditioned on the remaining variables, and we abbreviate this equivalence as $(i, j) \notin \mathcal{E} \Leftrightarrow \mathbf{y}_i \perp\!\!\!\perp \mathbf{y}_j | \mathbf{y}_{\mathcal{V} \setminus \{i, j\}}$. For Gaussian distributions with inverse covariance Γ , this conditional independence between variables \mathbf{y}_i and \mathbf{y}_j is also equivalent to $\Gamma_{ij} = \Gamma_{ji} = 0$.

As a result, given the connection between decoupling and the local effects to conditional independence properties, we can also analyze decoupling from the perspective of conditional independence and probabilistic graphical models. As a motivating example, we consider a linear system $\mathbf{A}\mathbf{y} = \mathbf{b}$ composed of two disciplines that are feedback coupled by two variables. This type of system can arise from the discretization of a differential equation with two solution domains that are coupled by their

boundary nodes. The matrix associated with this system is given by

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{A}_1 & 0 \\ \hline a & \mathbf{A}_2 \\ \hline 0 & b \end{array} \right], \quad (5.13)$$

where $\mathbf{A}_1 \in \mathbb{R}^{n \times n}$ and $\mathbf{A}_2 \in \mathbb{R}^{n \times n}$ represent the linear operators within each domain, and a and b are tunable parameters that control the degree of feedback and feed-forward coupling, respectively, between the solution at the boundary nodes. In this case, if the inputs are standard normal random variables (i.e., $\mathbf{b} \sim \mathcal{N}(0, I)$), the inverse covariance matrix for the solution, $\Gamma = \mathbf{A}^T \mathbf{A}$, has the structure

$$\Gamma = \left[\begin{array}{c|c} \mathbf{A}_1^T & 0 \\ \hline a & \mathbf{A}_2^T \\ \hline 0 & b \end{array} \right] \left[\begin{array}{c|c} \mathbf{A}_1 & 0 \\ \hline b & \mathbf{A}_2 \\ \hline 0 & a \end{array} \right] = \left[\begin{array}{c|c} \mathbf{A}_1^T \mathbf{A}_1 & a(\mathbf{A}_1^T)_{\cdot, n} \\ \hline -a(\mathbf{A}_1)_{n, \cdot} & +b^2 \\ \hline 0 & b(\mathbf{A}_2^T)_{\cdot, 1} \\ \hline & +a^2 \\ \hline & -b(\mathbf{A}_2)_{1, \cdot} \\ \hline & \mathbf{A}_2^T \mathbf{A}_2 \end{array} \right]. \quad (5.14)$$

If we perturb the linear system by neglecting the feedback coupling (i.e., $a = 0$), we decouple the input to the first discipline. The resulting system's output uncertainty has the inverse covariance matrix, $\tilde{\Gamma} = \tilde{\mathbf{A}}^T \tilde{\mathbf{A}}$, with the structure

$$\tilde{\Gamma} = \left[\begin{array}{c|c} \mathbf{A}_1^T & 0 \\ \hline 0 & \mathbf{A}_2^T \\ \hline & b \end{array} \right] \left[\begin{array}{c|c} \mathbf{A}_1 & 0 \\ \hline b & \mathbf{A}_2 \\ \hline & 0 \end{array} \right] = \left[\begin{array}{c|c} \mathbf{A}_1^T \mathbf{A}_1 & 0 \\ \hline & +b^2 \\ \hline 0 & b(\mathbf{A}_2^T)_{\cdot, 1} \\ \hline & -b(\mathbf{A}_2)_{1, \cdot} \\ \hline & \mathbf{A}_2^T \mathbf{A}_2 \end{array} \right]. \quad (5.15)$$

The elements of the solution vector from the first discipline, $\mathbf{y}_1, \dots, \mathbf{y}_{n-1}$, are now conditionally independent of \mathbf{y}_{n+1} from the second discipline when conditioned on the remaining variables, which includes the other boundary node \mathbf{y}_n . Intuitively, this demonstrates that in a feed-forward system, the first discipline will be statisti-

cally independent of the outputs in the second-discipline when fixing the value of the boundary node \mathbf{y}_n , given that there is no other interaction between the two sets of variables. As a result, the perturbed system has additional conditional independence properties (i.e., a graph $\tilde{\mathcal{G}}$ with a sparser set of edges $\tilde{\mathcal{E}} \subset \mathcal{E}$).

By introducing these assumptions in our model, we can measure the effect they cause in the distribution of the system's output variables using different metrics or divergences, as discussed in Section 2.3.2. Similarly to the approaches in Chapter 3, one possible distance measure is the KL divergence, $D_{KL}(\pi||\tilde{\pi})$, between the densities of two distributions when enforcing additional conditional independence properties or removing edges in the graph associated with π (i.e., the non-zero entries corresponding to the feedback coupling in (5.14)). By quantifying the effect of removing these edges, we can ultimately determine how to optimally approximate the conditional independence properties of the distribution for our original system and use these to find corresponding weak connections that can be decoupled.

Previous work has explored the effect of sparsifying probabilistic graphical models for applications in inference by showing that the distance, $D_{KL}(\pi||\tilde{\pi})$, for $\tilde{\pi} \in \Pi_{ij}$, a set of distributions that satisfies the conditional independence property $\mathbf{y}_i \perp\!\!\!\perp \mathbf{y}_j | \mathbf{y}_{\mathcal{V} \setminus (i,j)}$, that is not satisfied by π , is bounded from below by

$$\min_{\tilde{\pi} \in \Pi_{ij}} D_{KL}(\pi||\tilde{\pi}) \geq I(\mathbf{y}_i, \mathbf{y}_j | \mathbf{y}_{\mathcal{V} \setminus (i,j)}), \quad (5.16)$$

where $I(\mathbf{y}_i, \mathbf{y}_j | \mathbf{y}_{\mathcal{V} \setminus (i,j)})$ denotes the conditional mutual information for variables \mathbf{y}_i and \mathbf{y}_j [32, 31]. When the variables are jointly Gaussian, the conditional mutual information is cheaply computed from the inverse covariance matrix, Γ . Although this can be used to quantify the minimum error from using a misspecified model to represent the original distribution, it does not provide a guarantee on the approximation error from using the sparser graph to closely represent the outputs of the original model. As a result, the next sections will present results based on derived upper bounds for the KL divergence in an effort to provide rigorous guarantees on the distance when introducing conditional independence properties in a distribution.

5.5.2 Gaussian Distributions

In this section, we derive an upper bound for the distance, $D_{KL}(\pi||\tilde{\pi})$, after removing one edge from the graph of a Gaussian distribution. Without loss of generality, we will consider zero-mean Gaussians, π and $\tilde{\pi}$, that have covariance matrices, Σ and $\tilde{\Sigma}$, and inverse covariance matrices, Γ and $\tilde{\Gamma}$, respectively.

To bound the effect of removing one edge, we consider a Gaussian, $\tilde{\pi}$, with inverse covariance given by $\tilde{\Gamma} = \Gamma - \Gamma_{ij}(\mathbf{e}_i\mathbf{e}_j^T + \mathbf{e}_j\mathbf{e}_i^T)$. This inverse covariance matrix preserves the existing conditional independencies properties of π (i.e., sparsity of Γ) and adds the constraint $\mathbf{y}_i \perp\!\!\!\perp \mathbf{y}_j | \mathbf{y}_{V \setminus \{i,j\}}$. We note that while this is not the optimal $\pi^* = \arg \min_{\tilde{\pi} \in \Pi} D_{KL}(\pi||\tilde{\pi})$, it is a feasible point to this optimization problem and so Lemma 1 provides an upper bound to the value of this objective at optimality.

Lemma 1. *Let $\Sigma \succ 0$ be a covariance matrix in $\mathbb{R}^{n \times n}$, and Γ be the associated inverse covariance matrix. For the perturbed matrix, $\tilde{\Gamma} = \Gamma - \Gamma_{ij}(\mathbf{e}_i\mathbf{e}_j^T + \mathbf{e}_j\mathbf{e}_i^T)$ where $2|\Gamma_{ij}|(\Sigma_{ii}\Sigma_{jj})^{1/2} < 1$, then $\tilde{\Gamma} \succ 0$ and the KL divergence, D_{KL} , is bounded by*

$$D_{KL}(\pi||\tilde{\pi}) \leq \frac{1}{2} \left\{ 2|\Gamma_{ij}|(\Sigma_{ii}\Sigma_{jj})^{1/2} - \log(1 - 2|\Gamma_{ij}|(\Sigma_{ii}\Sigma_{jj})^{1/2}) \right\}. \quad (5.17)$$

Proof. The KL divergence between zero-mean Gaussian densities π and $\tilde{\pi}$ is given by

$$D_{KL}(\pi||\tilde{\pi}) = \frac{1}{2} \left\{ \log \left(\frac{|\tilde{\Sigma}|}{|\Sigma|} \right) - n + \text{Tr}(\tilde{\Gamma}\Sigma) \right\}. \quad (5.18)$$

Using the selected parameterization for the Gaussian distributions, the effect of the perturbed inverse covariance matrix on the trace and log-determinant are

$$\text{Tr}(\tilde{\Gamma}\Sigma) = \text{Tr}(\Gamma\Sigma - \Gamma_{ij}(\mathbf{e}_i\mathbf{e}_j^T + \mathbf{e}_j\mathbf{e}_i^T)\Sigma) = \text{Tr}(I) - \Gamma_{ij}(\Sigma_{ij} + \Sigma_{ji}) = n - 2\Gamma_{ij}\Sigma_{ij}, \quad (5.19)$$

$$\begin{aligned} \log \left(\frac{|\tilde{\Sigma}|}{|\Sigma|} \right) &= \log \left(\frac{1}{|\tilde{\Gamma}\Sigma|} \right) = -\log(|\Gamma\Sigma - \Gamma_{ij}(\mathbf{e}_i\mathbf{e}_j^T + \mathbf{e}_j\mathbf{e}_i^T)\Sigma|) \\ &= -\log(1 + (\Gamma_{ij})^2\Sigma_{ij}^2 - 2\Gamma_{ij}\Sigma_{ij} - (\Gamma_{ij})^2\Sigma_{ii}\Sigma_{jj}), \end{aligned} \quad (5.20)$$

where we have used the matrix determinant lemma to compute the determinant of

the matrix with a low-rank correction. Combining these expressions, the exact KL divergence is given by

$$D_{KL}(\pi||\tilde{\pi}) = \frac{1}{2} \left\{ -\log \left(1 + (\Gamma_{ij})^2 \Sigma_{ij}^2 - 2\Gamma_{ij} \Sigma_{ij} - (\Gamma_{ij})^2 \Sigma_{ii} \Sigma_{jj} \right) - 2\Gamma_{ij} \Sigma_{ij} \right\}. \quad (5.21)$$

Using the logarithmic inequality $|\log(1+x)| \leq -\log(1-|x|)$ for $|x| < 1$, and the Cauchy-Schwarz inequality for the elements in the covariance matrix matrix, $\Sigma_{ij} \leq (\Sigma_{ii} \Sigma_{jj})^{1/2}$, we have that

$$\begin{aligned} D_{KL}(\pi||\tilde{\pi}) &\leq \frac{1}{2} \left\{ 2|\Gamma_{ij} \Sigma_{ij}| - \log \left(1 - \left| (\Gamma_{ij})^2 \Sigma_{ij}^2 - 2\Gamma_{ij} \Sigma_{ij} - (\Gamma_{ij})^2 \Sigma_{ii} \Sigma_{jj} \right| \right) \right\} \\ &\leq \frac{1}{2} \left\{ 2|\Gamma_{ij}| (\Sigma_{ii} \Sigma_{jj})^{1/2} - \log \left(1 - 2|\Gamma_{ij}| (\Sigma_{ii} \Sigma_{jj})^{1/2} \right) \right\}. \end{aligned} \quad (5.22)$$

Furthermore, a sufficient condition to ensure the sparse inverse covariance matrix remains positive definite (i.e., $\tilde{\Sigma}, \tilde{\Gamma} \succ 0$) and for the bound on the KL divergence to be defined is given by

$$0 \leq 2|\Gamma_{ij}| (\Sigma_{ii} \Sigma_{jj})^{1/2} < 1. \quad (5.23)$$

□

Remark. A different quantity for comparing the two Gaussian distributions is the Förstner metric, the natural Riemannian metric over the manifold of symmetric positive definite matrices [21]. For the two covariance matrices, Σ and $\tilde{\Sigma}$, their distance as measured by the Förstner metric is defined by

$$d_F^2(\Sigma, \tilde{\Sigma}) = \sum_{k=1}^n \log^2(\sigma_k), \quad (5.24)$$

where σ_k are the eigenvalues of the matrix pencil $(\Sigma, \tilde{\Sigma})$. Given that the Förstner metric is invariant under inversion (i.e., $d_F(\Sigma, \tilde{\Sigma}) = d_F(\Gamma, \tilde{\Gamma})$), the change in the generalized eigenvalues of the corresponding inverse covariance matrices under the rank-2 perturbation to Γ can be used to derive the upper bound for the Förstner metric

$$d_F(\Sigma, \tilde{\Sigma}) \leq \left| \log \left(1 - 2|\Gamma_{ij}| (\Sigma_{ii} \Sigma_{jj})^{1/2} \right) \right|. \quad (5.25)$$

We note that the result in equation (5.25) demonstrates a similar dependence on the product term, $2|\Gamma_{ij}|(\Sigma_{ii}\Sigma_{jj})^{1/2}$, as the bound for the KL divergence in equation (5.17). These indicate that entries of small magnitude in Γ do not necessarily correspond to weak conditional independencies. Instead, the product of the inverse covariance and covariance entries for the variables must be small to have small KL divergence or Förstner metric between the distributions when adding an additional conditional independence constraint.

To generalize the result of Lemma 1 to the removal of multiple edges or addition of multiple conditional independence properties, we consider the Gaussian $\tilde{\pi}$ defined by the inverse covariance $\tilde{\Gamma} = \Gamma + \Delta$, where Δ is chosen to ensure that the sparsity of $\tilde{\Gamma}$ corresponds to the edge set $\tilde{\mathcal{E}}$. Analogously to the choice of $\tilde{\Gamma}$ used in the lemma above, one parameterization for Δ could be $\Delta_{ij} = -\Gamma_{ij}$ for the edges $(i, j) \in \mathcal{E} \setminus \tilde{\mathcal{E}}$ to be removed, $\Delta_{ii} = \sum_{k:(i,k) \in \mathcal{E} \setminus \tilde{\mathcal{E}}} |\Gamma_{ij}|$ to ensure that the matrix is diagonally dominant, and 0 otherwise. We note that a similar construction for Δ was used in [39] in the context of sampling Gaussian Markov random fields using different subgraphs. The bound on the KL divergence for the addition of multiple conditional independence properties is given by Lemma 2 below.

Lemma 2. *Let $\Sigma \succ 0$ be a covariance matrix in $\mathbb{R}^{n \times n}$, and Γ be the associated inverse covariance matrix. For the perturbed matrix, $\tilde{\Gamma} = \Gamma + \Delta \succ 0$ with positive definite matrix Δ , the KL divergence is bounded by*

$$D_{KL}(\pi || \tilde{\pi}) \leq \frac{1}{2} \|\Delta \Sigma\|_F^2, \quad (5.26)$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm.

Proof. Starting from the KL divergence between two Gaussian densities in (5.18),

and using the definition of the perturbed inverse covariance matrix, $\tilde{\Gamma}$,

$$\begin{aligned}
D_{KL}(\pi||\tilde{\pi}) &= \frac{1}{2} \left\{ \log \left(\frac{|\tilde{\Sigma}|}{|\Sigma|} \right) - n + \text{Tr}(\tilde{\Gamma}\Sigma) \right\} \\
&= \frac{1}{2} \left\{ \log \left(\frac{1}{|\tilde{\Gamma}\Sigma|} \right) - n + \text{Tr}(\Gamma\Sigma) + \text{Tr}(\Delta\Sigma) \right\} \\
&= \frac{1}{2} \{ -\log(|I + \Delta\Sigma|) + \text{Tr}(\Delta\Sigma) \},
\end{aligned} \tag{5.27}$$

where we have used that $\text{Tr}(\Gamma\Sigma) = \text{Tr}(I) = n$. Noting that $\log(|I+A|) = \sum_{i=1}^n \log(1+\lambda_i(A))$ where $\lambda_i(A)$ denotes the eigenvalues of A , and using the logarithmic inequality, $\log(1+\lambda_i) \geq \lambda_i - \frac{\lambda_i^2}{1+\lambda_i} \geq \lambda_i - \lambda_i^2$ for $\lambda_i \geq 0$, we have that $\log(|I+A|) \geq \text{Tr}(A) - \text{Tr}(A^2)$, and this holds for any positive-semidefinite matrix A . Combining this with the result in (5.27) and using that $\Delta\Sigma$ is positive-semidefinite, we have

$$\begin{aligned}
D_{KL}(\pi||\tilde{\pi}) &\leq \frac{1}{2} \{ -\text{Tr}(\Delta\Sigma) + \text{Tr}((\Delta\Sigma)^2) + \text{Tr}(\Delta\Sigma) \} \\
&= \frac{1}{2} \text{Tr}((\Delta\Sigma)^2) = \frac{1}{2} \|\Delta\Sigma\|_F^2.
\end{aligned} \tag{5.28}$$

□

The upper bounds in Lemmas 1 and 2 can now be used in a combinatorial search over sparse edge sets, $\tilde{\mathcal{E}}$, prior to minimizing $D_{KL}(\pi||\tilde{\pi})$ to find the closest distribution, π^* that satisfies the additional imposed conditional independence properties.

5.5.3 Exponential Family Distributions

In this section, we generalize the results of the previous section for sparsifying Gaussian graphical models to distributions in the linear exponential family. The density for a member of the linear exponential family has the general form

$$\pi(\mathbf{y}) = \exp\{\langle \theta, \phi(\mathbf{y}) \rangle - \Phi(\theta)\}, \tag{5.29}$$

where $\theta \in \mathbb{R}^n$ are natural parameters for the distribution, $\phi(\mathbf{y})$ represent sufficient statistics, and $\Phi(\theta)$ is a log-partition function to ensure the density is normalized. The

valid parameters for the distribution lie in the convex set given by $\Omega = \{\theta | \Phi(\theta) < \infty\}$. We refer the reader to Ref. [9] for more properties of exponential family distributions.

In the same spirit as for Gaussian graphical models, we consider the set of exponential family distributions that are specified by setting a subset of parameters, θ_2 , to zero that correspond to the removed edges in \mathcal{E} . Keeping the remaining parameters fixed in the new distribution, we compute an upper bound on the distance between π and the density specified by these parameters in the Lemma below.

Lemma 3. *Let $\pi, \tilde{\pi}$ be the densities of two distributions in the exponential family with the same set of sufficient statistics and parameters $\theta = (\theta_1, \theta_2)$ and $\tilde{\theta} = (\theta_1, 0)$, respectively. For small θ_2 , the KL divergence between the distributions is bounded by*

$$D_{KL}(\pi || \tilde{\pi}) \lesssim \sum_{k \in \theta_2} \sum_{j \in \theta_2} \theta_k \theta_j \text{Cov}(\phi_k(\mathbf{y}), \phi_j(\mathbf{y})). \quad (5.30)$$

Proof. Using the exponential family form for the two densities, the KL divergence between π and $\tilde{\pi}$ is given by

$$\begin{aligned} D_{KL}(\pi || \tilde{\pi}) &= \mathbb{E}_{\pi} [\log \pi(\mathbf{y}) - \log \tilde{\pi}(\mathbf{y})] \\ &= \mathbb{E}_{\pi} \left[\sum_{k \in \theta_2} \theta_k \phi_k(\mathbf{y}) - \Phi(\theta) + \Phi(\tilde{\theta}) \right], \end{aligned} \quad (5.31)$$

where we have used that the θ_1 parameters in both distributions have the same values and the θ_2 parameters are zero in $\tilde{\pi}$.

It is well-known that the log-partition function is a convex function (i.e., $\Phi(\theta) - \Phi(\tilde{\theta}) \geq \langle \theta - \tilde{\theta}, \nabla \Phi(\tilde{\theta}) \rangle$). Using this property with the relation of the mean parameters to the gradient of the log-partition function (i.e., $\nabla \Phi(\tilde{\theta}) = \mathbb{E}_{\tilde{\pi}}[\phi(\mathbf{y})]$), the difference of the log-partition functions is bounded from above by

$$\Phi(\tilde{\theta}) - \Phi(\theta) \leq -\langle \theta - \tilde{\theta}, \mathbb{E}_{\tilde{\pi}}[\phi(\mathbf{y})] \rangle. \quad (5.32)$$

The non-zero terms of the inner product in equation (5.32) correspond to the parameters in θ_2 , that were set to zero in $\tilde{\pi}$. These parameters are also weighted by

the expectation of their sufficient statistics with respect to $\tilde{\pi}$. Combining this with the equation in (5.31), we arrive at the upper bound

$$D_{KL}(\pi||\tilde{\pi}) \leq \sum_{k \in \theta_2} \theta_k (\mathbb{E}_\pi [\phi_k(\mathbf{y})] - \mathbb{E}_{\tilde{\pi}} [\phi_k(\mathbf{y})]). \quad (5.33)$$

For small values of θ_2 , the asymptotic expansion for the log-partition function, $\Phi(\tilde{\theta}) \approx \Phi(\theta) + \langle \tilde{\theta} - \theta, \mathbb{E}_\pi[\phi(\mathbf{y})] \rangle$, is a good approximation. Using this linearization, the expectations of the sufficient statistics with respect to $\tilde{\pi}$ are given by

$$\begin{aligned} \mathbb{E}_{\tilde{\pi}}[\phi_k(\mathbf{y})] &= \int \exp(\langle \tilde{\theta}, \phi(\mathbf{y}) \rangle - \Phi(\tilde{\theta})) \phi_k(\mathbf{y}) d\mathbf{y} \\ &\approx \int \exp(\langle \theta, \phi(\mathbf{y}) \rangle - \Phi(\theta)) \exp(-\langle \theta_2, \phi_{\theta_2}(\mathbf{y}) \rangle + \langle \theta_2, \mathbb{E}_\pi[\phi_{\theta_2}(\mathbf{y})] \rangle) \phi_k(\mathbf{y}) d\mathbf{y} \\ &= \mathbb{E}_\pi \left[\exp(-\langle \theta_2, \phi_{\theta_2}(\mathbf{y}) - \mathbb{E}_\pi[\phi_{\theta_2}(\mathbf{y})] \rangle) \phi_k(\mathbf{y}) \right]. \end{aligned} \quad (5.34)$$

Substituting the expectations in expression (5.33) and using the Taylor series expansion, $1 - \exp(x) \approx x$ for small x , we arrive at the asymptotic upper bound

$$\begin{aligned} D_{KL}(\pi||\tilde{\pi}) &\lesssim \sum_{k \in \theta_2} \theta_k \mathbb{E}_\pi \left[\phi_k(\mathbf{y}) - \phi_k(\mathbf{y}) \exp(-\langle \theta_2, \phi_{\theta_2}(\mathbf{y}) - \mathbb{E}_\pi[\phi_{\theta_2}(\mathbf{y})] \rangle) \right] \\ &\approx \sum_{k \in \theta_2} \theta_k \mathbb{E}_\pi [\phi_k(\mathbf{y}) \cdot \langle \theta_2, \phi_{\theta_2}(\mathbf{y}) - \mathbb{E}_\pi[\phi_{\theta_2}(\mathbf{y})] \rangle] \\ &= \sum_{k \in \theta_2} \sum_{j \in \theta_2} \theta_k \theta_j (\mathbb{E}_\pi [\phi_k(\mathbf{y}) \phi_j(\mathbf{y})] - \mathbb{E}_\pi [\phi_k(\mathbf{y})] \mathbb{E}_\pi [\phi_j(\mathbf{y})]). \end{aligned} \quad (5.35)$$

□

Remark. *This expansion also has the same behavior as the upper bounds derived in Lemmas 1 and 2 for the Gaussian distribution where the KL divergence scales with the product of the distribution's parameters and its corresponding moments (i.e., the entries in the inverse covariance and covariance matrices, respectively, for the Gaussian setting).*

5.5.4 More General Distributions

In this section, we propose an approach to find similar upper bounds on the distance of a general distribution of interest, π , to the closest distribution whose conditional independencies properties correspond to a sparser graph, $\tilde{\mathcal{G}} \subset \mathcal{G}$. Given that it is difficult to parameterize the conditional independence properties in arbitrary probability distributions that are not members of a specific class (i.e., exponential family), we will instead compute these distance by restricting $\tilde{\pi}$ to lie in a tractable family, such as a set of constrained Gaussian distributions. The distance is then given by computing $D_{KL}(\pi||\tilde{\pi})$ for $\tilde{\pi} \in \Pi_{\tilde{\mathcal{E}}} = \{\mathcal{N}(\tilde{\mu}, \tilde{\Sigma}) | \tilde{\Sigma}_{ij}^{-1} = \tilde{\Gamma}_{ij} = 0 \forall (i, j) \notin \tilde{\mathcal{E}}\}$.

Given the covariance and inverse covariance matrix of π , a coarse estimate for this distance is given by selecting a Gaussian described by $\tilde{\Gamma}$ that is the inverse covariance matrix of π perturbed by a matrix, Δ , as in Section 5.5.2. This will ensure that $\tilde{\Gamma}$ satisfies the imposed conditional independence properties in $\Pi_{\tilde{\mathcal{E}}}$ and $D_{KL}(\pi||\tilde{\pi}) = \mathbb{E}_{\pi}[\log(\pi) - \log(\tilde{\pi})]$ is tractable to compute up to the entropy of π .

Remark. *A tighter upper bound can be computed by finding the closest Gaussian distribution that satisfies the conditional independence properties imposed in $\Pi_{\tilde{\mathcal{E}}}$. To do so, we solve the optimization problem: $\arg \min_{\tilde{\pi} \in \Pi_{\tilde{\mathcal{E}}}} D_{KL}(\pi||\tilde{\pi})$. By noting that $D_{KL}(\pi||\tilde{\pi}) = \mathbb{E}_{\pi}[\log(\pi) - \log(\tilde{\pi})]$ and the first term in the expectation, the entropy of π , is independent of $\tilde{\pi}$, this optimization problem reduces to maximizing $\mathbb{E}_{\pi}[\log(\tilde{\pi})]$. Using the Gaussian parameterization for $\tilde{\pi}$, we have that*

$$\begin{aligned} \mathbb{E}_{\pi}[\log(\tilde{\pi})] &= \mathbb{E}_{\pi} \left[-\frac{1}{2} \log((2\pi)^n |\tilde{\Sigma}|) - \frac{1}{2} (\mathbf{y} - \tilde{\mu})^T \tilde{\Gamma} (\mathbf{y} - \tilde{\mu}) \right] \\ &= -\frac{1}{2} \log((2\pi)^n |\tilde{\Sigma}|) - \frac{1}{2} \mathbb{E}_{\pi} \left[\text{Tr} \left(\tilde{\Gamma} (\mathbf{y} - \tilde{\mu})(\mathbf{y} - \tilde{\mu})^T \right) \right]. \end{aligned} \quad (5.36)$$

From performing the optimization to find the closest Gaussian within the set, $\Pi_{\tilde{\mathcal{E}}}$, the moment matching optimality conditions are given by $\tilde{\mu} = \mathbb{E}_{\pi}[\mathbf{y}]$ and $\tilde{\Sigma} = \Sigma - \sum_{(i,j) \notin \tilde{\mathcal{E}}} \gamma_{(i,j)} (\mathbf{e}_i \mathbf{e}_j^T + \mathbf{e}_j \mathbf{e}_i^T)$ where $\Sigma = \mathbb{E}_{\pi}[(\mathbf{y} - \mu)(\mathbf{y} - \mu)^T]$ and $\gamma_{(i,j)}$ are Lagrange

multipliers. Using these conditions,

$$\begin{aligned}
\mathbb{E}_\pi[\log(\tilde{\pi})] &= -\frac{1}{2} \log((2\pi)^n |\tilde{\Sigma}|) - \frac{1}{2} \text{Tr}(\tilde{\Gamma}\Sigma) \\
&= -\frac{1}{2} \log((2\pi)^n |\tilde{\Sigma}|) - \frac{1}{2} \text{Tr} \left(\tilde{\Gamma} \left\{ \tilde{\Sigma} + \sum_{(i,j) \notin \tilde{\mathcal{E}}} \gamma_{(i,j)} (\mathbf{e}_i \mathbf{e}_j^T + \mathbf{e}_j \mathbf{e}_i^T) \right\} \right) \quad (5.37) \\
&= -\frac{1}{2} \log((2\pi)^n |\tilde{\Sigma}|) - \frac{1}{2} n - \frac{1}{2} \sum_{(i,j) \notin \tilde{\mathcal{E}}} \gamma_{(i,j)} \text{Tr}(\tilde{\Gamma}_{ij} + \tilde{\Gamma}_{ji}).
\end{aligned}$$

With the edge constraints on the inverse covariance matrix, $\tilde{\Gamma}_{ij} = \tilde{\Gamma}_{ji} = 0$ for all $(i, j) \notin \tilde{\mathcal{E}}$. Therefore, the last trace term in (5.37) is zero and the distance reduces to

$$D_{\text{KL}}(\pi || \tilde{\pi}) = \mathbb{E}_\pi[\log(\pi)] - \left(-\frac{n}{2} \log(2\pi e) - \frac{1}{2} \log(|\tilde{\Sigma}|) \right), \quad (5.38)$$

where the covariance of the closest Gaussian distribution to π under the conditional independence constraints on the matrix inverse is typically computed by solving a log-determinant maximization problem [18]. We note that for chordal graph structures for $\tilde{\mathcal{G}}$, which include chain and tree graphs, the covariance of the closest Gaussian, $\tilde{\Sigma}$, can be computed in closed form without requiring optimization.

As a result, up to the entropy of π , the distance $D_{\text{KL}}(\pi || \tilde{\pi})$ where $\tilde{\pi}$ is the closest Gaussian that satisfies the additional constraints is computable. In principle, this distance can be used to cheaply estimate the effect of removing multiple edges in arbitrary distributions.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In the present thesis, we have studied how to identify important discipline couplings in the context of multidisciplinary models under uncertainty. From the perspective of optimal model selection, we developed a set of algorithms based on a combination of model linearizations, control variates, and model data, together with sequential Monte Carlo for exploring the combinatorial model space. These algorithms trade-off increasing accuracy in the characterization of the model output uncertainty with increasing computational cost to identify the optimal models. These algorithms were applied and validated with three engineering problems: a fire detection satellite model, a turbine engine cycle analysis model, and an aero-structural lifting surface model. For each of these problems, the algorithms returned a set of optimal models with a reduced number of discipline couplings relative to the reference model while having accurate distributions for the specified quantities of interest.

There are four main conclusions that we can draw from the numerical results and theoretical analysis developed in this work. First, to identify an approximation to the model coupling it is necessary to solve a combinatorial optimization problem, due to binary nature of selecting different discipline connections to retain in a decoupled model. To do so, we must use an optimization algorithm that is scalable to increasing dimension or the total number of discipline couplings, works with noisy

objective functions due to sample-based estimates for the statistical distance between distributions, and is amenable to return a distribution of possible models to use. One algorithm that satisfies these criteria and we recommend for future analyses is the sequential Monte Carlo algorithm that was presented in Section 2.4, in the context of model selection.

Second, while methods with increasing fidelity are more accurate for identifying the true global optimum when searching for a model with reduced coupling, linearizations are an inexpensive approach to arrive at a near-optimal solution. We note that the Bayesian approach, which is based entirely on model data, and the control variates may be necessary in highly nonlinear models with different localized output responses. However, for many engineering applications such as the turbine and aero-structural models tested in Chapter 4, the linearizations capture a coarse approximation to the uncertainty that is often sufficient for performing model optimization without perfectly estimating a high-dimensional distribution, which is not our goal in this work. This low computational cost for the model selection process is particularly important in this work and in future applications of decoupled models that are described below in Section 6.2.

Third, from the numerical results for the tested applications, we conclude that using a reduced number of discipline couplings can provide a close approximation to the model outputs while yielding a significant reduction in the computational cost for each model evaluation. This reduction in cost is most evident when neglecting feedback couplings, as seen in the optimally decoupled aero-structural models. Nevertheless, the optimal coupling found in all problems is dependent on the input uncertainty, as observed when changing the random inputs to the fire-detection satellite model. As a result, we recommend re-assessing the accuracy of the decoupled model whenever changing the input variable uncertainty. If necessary, a user should re-process the optimal model coupling by re-running the optimization algorithm to identify these model changes.

Lastly, our theoretical analysis highlights the factors that contribute to the point-wise error and changes in the output uncertainty when approximating the discipline

coupling. These include the discrepancy between the true and fixed input values for a coupling variable, and the sensitivities of the model outputs to the coupling variables. Although the current analysis is limited to linear models, the results in Chapter 5 provide sufficient conditions for linear models being well-posed under decoupling. We affirm that these requirements should be checked when introducing a model approximation, particularly in the case of feedback systems, to ensure a solution will exist to these new decoupled models. Under these conditions, the local changes to the conditional independence structure of the output uncertainty from decoupling can be analyzed by using the connections developed between multidisciplinary model coupling and probabilistic graphical models.

6.2 Future Work

To the best of our knowledge this is the first work that has addressed the selection of *optimal* discipline couplings in multidisciplinary models with uncertain inputs. In addition to the continued application of the algorithm to reveal the natural structure of other complex engineering systems under different input uncertainties, there are several methodological and theoretical questions that are interesting to pursue.

6.2.1 Improved Optimization Objective

To identify an optimal decoupled model, we formulated an optimization problem that balances a measure of information loss in the distribution based on the KL divergence with a measure for the complexity of each model. While higher-order derivative information or control variates improve the estimate for the KL divergence between the output distributions of two models, one area of future work is to perform a more detailed study of these approaches and the use of other distance metrics. In particular, it will be interesting to characterize when sample-based corrections are necessary and to automatically adapt the model selection algorithm by using an computed number of necessary samples to improve the estimates for the mean, variance and/or other higher-order moments of the model outputs. We suppose that to do this it will be

necessary to quantify the degree of nonlinearity and non-Gaussianity of the model outputs to determine if a low-order Taylor-Series approximation around a specified point may be sufficient to capture the local relationship of the input-output mapping. It may also be possible to combine multiple approximations evaluated at different points or different partitions of the input space to more accurately capture the behavior of non-monotonic output functions over the range of input values.

Future research may also address improvements to the measure of complexity of different decoupled models, that we denoted by $\mathcal{P}(M)$. In practice, discipline couplings are not all equal and will contribute differently to the total computational cost of evaluating a given model. Therefore, instead of only measuring the number of discipline couplings we propose to use a metric that adds a greater weight to discipline couplings that will reduce model runtime or memory requirements when neglected. One approach will be to first favor the removal of large feedback loops in the model in order to approach a feed-forward model that can be evaluated for the most part sequentially with only small local feedback loops involving fewer discipline, as seen in Figure 6-1. If it is not possible to achieve a feed-forward model, the algorithm may also favor parallelism across different disciplines, as seen in Figure 6-2. We note that the search for feed-forward models is analogous to the problem of finding subgraphs with low tree-width for performing exact inference efficiently [36], and so it may be possible

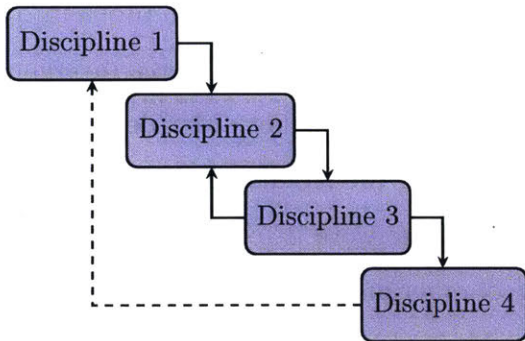


Figure 6-1: Model with only feed-forward connections and local feedback loop after decoupling disciplines 1 and 4

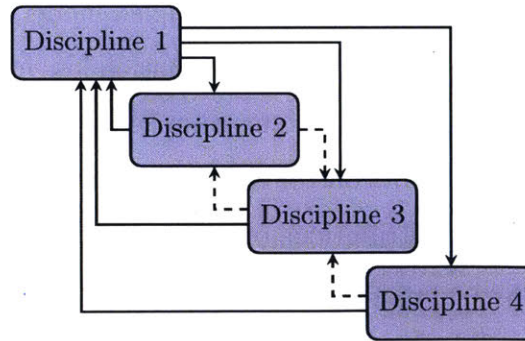


Figure 6-2: Decoupled model with parallel structure—disciplines 2, 3 and 4 compute feedback to discipline 1 independently and run in parallel

to borrow additional ideas from the literature on probabilistic graphical models in this work. Finally, the metric for model complexity may also include estimated runtime for individual disciplines and the memory required to transfer information across discipline couplings.

6.2.2 Preconditioning with Decoupled Models

Decoupling nonlinear systems introduces a transformation to the coupled nonlinear system of equations, $\mathcal{R}(\mathbf{x}, \mathbf{y}) = 0$, that shifts the solution for the state variables and the corresponding distribution of the output QoIs. In engineering systems where it is of interest to remain faithful to the solution of the original system, optimally decoupled models can instead be used to accelerate algorithms for solving coupled nonlinear root finding problems.

For each $\mathbf{x} \in \mathbb{R}^m$, one common method for finding a solution to parameterized system of equations, $\mathcal{R}(\mathbf{x}, \mathbf{y}) = 0$, is to use a Newton's method. The standard Newton algorithm chooses an initial guess, $\mathbf{y}^{(0)}$, and iteratively updates the solution by solving the linear system $\nabla_{\mathbf{y}}(\mathcal{R}(\mathbf{x}, \mathbf{y}^{(k)}))\Delta\mathbf{y}^{(k)} = -\mathcal{R}(\mathbf{x}, \mathbf{y}^{(k)})$ for all k where $\Delta\mathbf{y}^{(k)} = \mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}$, until the residual error falls below a specified relative tolerance.

In practice solving the linear system is the most computationally expensive step of these algorithms. As a result, for many problems it is common to find preconditioners, M , that are cheap to apply and closely approximate $(\nabla_{\mathbf{y}}\mathcal{R})^{-1}$ to reduce the problem's condition number and improve its rate of convergence when solving the transformed system, $M(\nabla_{\mathbf{y}}\mathcal{R}(\mathbf{x}, \mathbf{y}^{(k)}))\Delta\mathbf{y}^{(k)} + \mathcal{R}(\mathbf{x}, \mathbf{y}^{(k)}) = 0$. While it is commonplace to use well-developed preconditioners for certain physical problem (i.e., computational fluid dynamics, solid mechanics, etc.), the typical approach for other problems is to implement standard block-Jacobi or Gauss-Seidel preconditioners.

However, for a model with multiple sets of disciplines that are locally feedback coupled (as in Figure 6-1), common upper-triangular preconditioners will not closely approximate the action of the inverse Jacobian. Instead, the model selection algorithm presented in this thesis provides a method to identify optimally decoupled models whose outputs are close to the true model outputs over the range of input variables.

Therefore, the Jacobian of these lower complexity models can be interpreted and used as preconditioners for the linear systems of equations arising at each iteration of Newton’s method or to even directly approximate the Jacobian (by allowing for inexact Newton directions). In future work it will be interesting to compare the performance of these preconditioners to other algorithms that use offline computation to construct improved preconditioners, such as ILU-based methods, or the Sparse Approximate Inverse preconditioners [7].

Recent research has also addressed preconditioning parameterized nonlinear systems of equations by directly composing transformations, $M_{\mathbf{x}}: \mathbb{R}^n \rightarrow \mathbb{R}^n$, with the nonlinear system. Analogous to the setting of linear equations, some desirable properties for these preconditioners are:

1. $M_{\mathbf{x}}(\mathbf{y}) \approx \mathcal{R}(\mathbf{x}, \mathbf{y})^{-1}$ for each \mathbf{x} and is well-defined
2. $M_{\mathbf{x}}(\mathbf{y})$ and its derivatives are cheap to compute and compose with $\mathcal{R}(\mathbf{x}, \mathbf{y})$
3. More robust nonlinear convergence requiring fewer global Newton iterations
4. Wider basin of attraction for initial condition

While the optimal preconditioner for the nonlinear system of equations is $M_{\mathbf{x}}(\mathbf{y}) = \mathcal{R}(\mathbf{x}, \mathbf{y})^{-1}$ so that $\mathcal{R}(\mathbf{x}, \mathbf{y})^{-1} \circ \mathcal{R}(\mathbf{x}, \mathbf{y}) = \mathbf{y} = \mathcal{R}(\mathbf{x}, \mathbf{y})^{-1}(\mathbf{0})$, in most cases the system is not invertible and we must search for a preconditioner within a tractable class of functions that best meet the desired properties above. One such class is the set of nonlinear functions with lower triangular variable dependence, that has been explored in the current nonlinear preconditioning and domain decomposition literature [16, 38]. In the context of decoupled models, these lower-triangular functions correspond to feed-forward models. Given that these models satisfy the first and second properties above, this makes them potential candidates for physically motivated nonlinear preconditioners that are optimal over the range of input variables.

Future work may explore the convergence of the transformed systems given by the composition of the decoupled and original nonlinear systems. Their sensitivity to the initial condition and robustness to being outside the region of quadratic convergence in Newton’s method may be investigated. Furthermore, by identifying a sequence of

decoupled models of increasing accuracy, an optimal allocation of the different models as preconditioners for different iterations of Newton’s method based on the accuracy tolerance at each step may be explored. We suppose that it may be sufficient to apply an inexact model at the initial iterations, while a model with more coupling may be required closer to the converged solution. These ideas of optimally allocating a sequence of models to reduce computational cost may also borrow ideas from recent work in the use of multi-fidelity reduced order models for uncertainty quantification [49]. Finally, we note that these sequences of models can also be used for reducing the overall cost of optimizing a set of design variables in a model.

6.2.3 Further Connections to Probabilistic Graphical Models

As discussed in Chapter 5, multidisciplinary models can be analyzed from the perspective of probabilistic graphical models that represent the conditional independence properties of a distribution. In particular, decoupling weak discipline connections is related to adding conditional independence properties to distributions of the system outputs. In the probabilistic inference community, finding these weak conditional independence properties is important to accelerate the cost of inference algorithms on graphs (i.e., belief propagation & variable elimination) that scale with the tree-width and other properties of the graph structure.

Previous studies have addressed this problem for inference in the context of Gaussian distributions, by evaluating lower bounds on the KL divergence between a distribution and one that has additional conditional independence properties. However, these lower bounds do not have guarantees on the accuracy of the approximation, and so it is of interest to develop algorithms that identify optimally sparse graphs with guarantees on the closeness to the distribution of interest. This question is important both for sparsifying existing graphs in order to reduce the parameterizations for models of the output distributions (i.e., transport maps, see [59]), and also for the question of learning sparse graphs and the corresponding physical model structures (i.e., discipline connections and approximations to the model physics) from data. The results in Section 5.5 are the beginning of an effort to address some of these ques-

tions. Future work will expand upon this work in order to further understand the relationships between multidisciplinary models and probabilistic graphical models.

Bibliography

- [1] Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations. *arXiv preprint arXiv:1602.07592*, 2016.
- [2] Sergio Amaral, Douglas Allaire, and Karen Willcox. A decomposition-based approach to uncertainty analysis of feed-forward multicomponent systems. *International Journal for Numerical Methods in Engineering*, 100(13):982–1005, 2014.
- [3] 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.
- [4] 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.
- [5] Maarten Arnst, Christian Soize, and Roger Ghanem. Hybrid sampling/spectral method for solving stochastic coupled problems. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):218–243, 2013.
- [6] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi. *Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties*. Springer Berlin Heidelberg, 2012.
- [7] Michele Benzi and Miroslav Tuma. A sparse approximate inverse preconditioner for nonsymmetric linear systems. *SIAM Journal on Scientific Computing*, 19(3):968–994, 1998.
- [8] Espen Bernton, Pierre E Jacob, Mathieu Gerber, and Christian P Robert. Inference in generative models using the Wasserstein distance. *arXiv preprint arXiv:1701.05146*, 2017.
- [9] P.J. Bickel and K.A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2015.

- [10] Christina L Bloebaum, Prabhat Hajela, and Jaroslaw Sobieszczanski-Sobieski. Non-hierarchical system decomposition in structural optimization. *Engineering Optimization*, 19(3):171–186, 1992.
- [11] Luke Bornn, Natesh S Pillai, Aaron Smith, and Dawn Woodard. The use of a single pseudo-sample in approximate Bayesian computation. *Statistics and Computing*, pages 1–8, 2015.
- [12] Robert D Braun, Alison A Moore, and Ilan M Kroo. Collaborative approach to launch vehicle design. *Journal of Spacecraft and Rockets*, 34(4):478–486, 1997.
- [13] Loïc Brevault, Mathieu Balesdent, Nicolas Bérend, and Rodolphe Le Riche. Decoupled multidisciplinary design optimization formulation for interdisciplinary coupling satisfaction under uncertainty. *AIAA Journal*, 54(1):186–205, 2015.
- [14] Tyson R Browning. Applying the design structure matrix to system decomposition and integration problems: a review and new directions. *IEEE Transactions on Engineering Management*, 48(3):292–306, 2001.
- [15] Erkan O Buzbas and Noah A Rosenberg. AABC: approximate approximate bayesian computation when simulating a large number of data sets is computationally infeasible. *arXiv preprint arXiv:1301.6282*, 2013.
- [16] Xiao-Chuan Cai and David E Keyes. Nonlinearly preconditioned inexact Newton algorithms. *SIAM Journal on Scientific Computing*, 24(1):183–200, 2002.
- [17] Anirban Chaudhuri and Karen Willcox. Multifidelity uncertainty propagation in coupled multidisciplinary systems. In *18th AIAA Non-Deterministic Approaches Conference*, page 1442, 2016.
- [18] Joachim Dahl, Vwani Roychowdhury, and Lieven Vandenberghe. Maximum likelihood estimation of Gaussian graphical models: numerical implementation and topology selection. 2005.
- [19] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.
- [20] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. An adaptive sequential Monte Carlo method for approximate bayesian computation. *Statistics and Computing*, 22(5):1009–1020, 2012.
- [21] Wolfgang Förstner and Boudewijn Moonen. A metric for covariance matrices. In *Geodesy-the Challenge of the 3rd Millennium*, pages 299–309. Springer, 2003.
- [22] P Fuglsang and H Aagaard Madsen. Optimization method for wind turbine rotors. *Journal of Wind Engineering and Industrial Aerodynamics*, 80(1):191–206, 1999.

- [23] Robert Braunt Peter Gage and Ilan Kroof lan Sobieski. Implementation and performance issues in collaborative optimization. *AIAA Paper*, 1996.
- [24] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.
- [25] Alison L Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.
- [26] Raphael T Haftka. Optimization of flexible wing structures subject to strength and induced drag constraints. *AIAA Journal*, 15(8):1101–1106, 1977.
- [27] Tristan Hearn, Eric Hendricks, Jeffrey Chin, Justin Gray, and Kenneth T Moore. Optimization of turbine engine cycle analysis with analytic derivatives. In *17th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 4297, 2016.
- [28] Antony Jameson. Aerodynamic design via control theory. *Journal of Scientific Computing*, 3(3):233–260, 1988.
- [29] John P. Jasa and John T. Hwang. Openaerostruct: An open-source tool to perform aerostructural optimization. 2017.
- [30] R. Jin, W. Chen, and T.W. Simpson. Comparative studies of metamodelling techniques under multiple modelling criteria. *Structural and Multidisciplinary Optimization*, 23(1):1–13, 2001.
- [31] Varun Jog and Po-Ling Loh. On model misspecification and KL separation for Gaussian graphical models. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1174–1178. IEEE, 2015.
- [32] Jason K Johnson and Alan S Willsky. A recursive model-reduction method for approximate inference in Gaussian Markov random fields. *IEEE Transactions on Image Processing*, 17(1):70–83, 2008.
- [33] A. Keane and P. Nair. *Computational Approaches for Aerospace Design: The Pursuit of Excellence*. Wiley, 2005.
- [34] D.E. Keyes. Computational science. In *The Princeton Companion to Applied Mathematics*. Princeton University Press, 2015.
- [35] Hyung Min Kim, Nestor F Michelena, Panos Y Papalambros, and Tao Jiang. Target cascading in optimal system design. *Journal of Mechanical Design*, 125(3):474–480, 2003.
- [36] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

- [37] Ilan Kroo, Steve Altus, Robert Braun, Peter Gage, and Ian Sobieski. Multidisciplinary optimization methods for aircraft preliminary design. *AIAA paper*, 4325:1994, 1994.
- [38] Lulu Liu and David E Keyes. Field-split preconditioned inexact Newton algorithms. *SIAM Journal on Scientific Computing*, 37(3):A1388–A1409, 2015.
- [39] Ying Liu, Oliver Kosut, and Alan S Willsky. Sampling GMRFs by subgraph correction. In *In: NIPS 2012 Workshop: Perturbations, Optimization, and Statistics*. Citeseer, 2012.
- [40] Zhoujie Lu and JRRA Martins. Graph partitioning-based coordination methods for large-scale multidisciplinary design optimization problems. In *12th AIAA Aviation Technology*, 2012.
- [41] David JC MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge university press, 2003.
- [42] Sankaran Mahadevan and A Haldar. Probability, reliability and statistical method in engineering design, 2000.
- [43] Joaquim R R A Martins, Juan J Alonso, and James J Reuther. High-fidelity aerostructural design optimization of a supersonic business jet. *Journal of Aircraft*, 41(3):523–530, 2004.
- [44] Joaquim R R A Martins, Juan J Alonso, and James J Reuther. A coupled-adjoint sensitivity analysis method for high-fidelity aero-structural design. *Optimization and Engineering*, 6(1):33–62, 2005.
- [45] Joaquim R R A Martins and John T Hwang. Review and unification of methods for computing derivatives of multidisciplinary computational models. *AIAA Journal*, 51(11):2582–2599, 2013.
- [46] Joaquim R R A Martins and Andrew B Lambe. Multidisciplinary design optimization: a survey of architectures. *AIAA Journal*, 51(9):2049–2075, 2013.
- [47] Charles D McAllister and Timothy W Simpson. Multidisciplinary robust design optimization of an internal combustion engine. *Journal of Mechanical Design*, 125(1):124–130, 2003.
- [48] Leo WT Ng and Karen E Willcox. Multifidelity approaches for optimization under uncertainty. *International Journal for Numerical Methods in Engineering*, 100(10):746–772, 2014.
- [49] Benjamin Peherstorfer, Karen Willcox, and Max Gunzburger. Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, 38(5):A3163–A3194, 2016.
- [50] Dennis Prangle. Lazy ABC. *Statistics and Computing*, 26(1-2):171–185, 2016.

- [51] Shankar Sankararaman and Sankaran Mahadevan. Likelihood-based approach to multidisciplinary analysis under uncertainty. *Journal of Mechanical Design*, 134(3):031008, 2012.
- [52] Christian Schäfer. Particle algorithms for optimization on binary spaces. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 23(1):8, 2013.
- [53] Christian Schäfer and Nicolas Chopin. Sequential Monte Carlo on large binary sampling spaces. *Statistics and Computing*, 23(2):163–184, 2013.
- [54] RS Sellar, SM Batill, and JE Renaud. Response surface based, concurrent subspace optimization for multidisciplinary system design. *AIAA paper*, 714:1996, 1996.
- [55] Pranay Seshadri, Paul Constantine, Gianluca Iaccarino, and Geoffrey Parks. A density-matching approach for optimization under uncertainty. *Computer Methods in Applied Mechanics and Engineering*, 305:562–578, 2016.
- [56] Ralph C Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. SIAM, 2013.
- [57] Jaroslaw Sobieszczanski-Sobieski, Jeremy S Agte, and Robert R Sandusky. Bilevel integrated system synthesis. *AIAA Journal*, 38(1):164–172, 2000.
- [58] Jaroslaw Sobieszczanski-Sobieski and Raphael T Haftka. Multidisciplinary aerospace design optimization: survey of recent developments. *Structural Optimization*, 14(1):1–23, 1997.
- [59] Alessio Spantini, Daniele Bigoni, and Youssef Marzouk. Inference via low-dimensional couplings. *arXiv preprint arXiv:1703.06131*, 2017.
- [60] Donald V Steward. The design structure system: A method for managing the design of complex systems. *IEEE Transactions on Engineering Management*, (3):71–74, 1981.