June 2022

New Developments in Statistical Optimal Designs for Physical and Computer Experiments

Damola M. Akinlana
University of South Florida

New Developments in Statistical Optimal Designs for Physical and Computer

Experiments

by

Damola M. Akinlana

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Mathematics & Statistics
College of Arts and Sciences
University of South Florida

Major Professor: Lu Lu, Ph.D.
Kandethody M. Ramachandran, Ph.D.
Seung-Yeop Lee , Ph.D.
Dmytro Savchuk, Ph.D.
Mingyang Li, Ph.D.

Date of Approval:
June 16, 2022

Keywords: Multiple objective, Pareto front, Bayesian designs, Degradation tests

Table of Contents

List of Tables

List of Figures

Abstract

Statistical design of experiments allows for multiple factors influencing a process to be systematically manipulated in an experiment, and their effects on the output of the process to be studied via statistical modeling and analysis. Classical designs offer general nice performance but have limited applications due to restricted design size, region, and randomization structure. Computer generated optimal designs become more popular in recent decades due to the rapid growth in computing power. Most existing work in optimal design of experiments involves designing experiments with optimal performance on a single chosen objective or a single response. However, with the increasing limitation in resources and emergence of complex engineering problems, more and more experiments aim to simultaneously achieve multiple objectives or study multiple responses.

Recent developments have made enhancements on methodologies for selecting optimal designs based on multiple criteria for a single response from a physical experiment. However, there has been limited work on constructing optimal designs for a physical experiment with multiple responses. In the area of design of computer experiments, despite space-filling designs have been most popular due to their flexibility on model choices, existing space-filling designs are mainly built for optimizing a single criterion, which is often associated with worse performance on other characteristics. In addition, modern design of experiment techniques which provide powerful tools for efficient data collection and inferential analysis have not been broadly used in the field of reliability analysis.

This dissertation adds to the growing research in optimal design of experiments in three different areas: 1. It develops new cost-efficient optimal designs for obtaining precise estimation of multiple responses from a single experiment by leveraging prior infor-

mation from earlier screening experiment; 2. It proposes new Latin hypercube designs for computer experiments based on balancing multiple space-filling characteristics; and 3. It utilizes Bayesian optimal design technique for selecting optimal test plans for accelerated degradation tests with two or more accelerating factors and more general degradation path models.

Chapter 1: Introduction

In scientific studies, the form of research to be carried out is usually influenced by the problem(s) of interest or the question(s) to be answered. Different data collection strategies could be used to answer different questions. Based on data collection techniques, scientific research is divided into observational studies and experimental studies. In observational studies, scientists can only passively observe the data without interfering with the course of the study. While in experimental studies, scientists intervenes with the study by controlling the levels of one or more controllable factors. Experimental studies are planned to describe the causal relationship between the input(s) and output of a process while controlling other noise factors. In experimental studies, statistical design of experiment techniques are often used to guide the efficient allocation of experimental resources, in order to obtain the most information about a process subject to the limited resources and natural variability of the experimental process.

Statistical design of experiments is a branch of applied statistics used for conducting scientific studies of a system, process or product in which input factors are manipulated to investigate their effects on measured response. It is a systematic approach to problem-solving, where principles and techniques are applied at the data collection stage of an experiment, with the aim of controlling systematic error, reducing random variations, increasing precision of parameter estimates, making predictions about future observations, and/or discriminating between competing models.

Statistical design of experiments are often categorized into classical designs and the optimal designs. Classical designs include the commonly used factorial designs, fractional factorial designs, and many response surface designs. They are often associated

with specific run size and regular design region (cuboidal or spherical), and are not flexible to adapt for specific randomization structures. However, there are situations when classical designs do not fit the experimental requirements, for example, when the experiment has specific requirement on the run size, or there are constraints on the design region (i.e. irregular shaped design region), or the model of interest is in a particular form. Optimal designs, on the other hand, offer more flexibility for broader design problems and are generally more useful for situations involving resource constraints, irregular design regions, special randomization structures or model formulations [49]. Applications of optimal designs have become increasingly popular when compared to classical designs due to their flexibility and ability to handle a variety of design problems. However, most of the existing work involve experiments with a single response or a single objective.

There has been limited work that addresses optimal designs for experiments with multiple responses. With an increasing limit in resources, more and more experiments now involve multiple response variables which can be characterized in different models. In this case, choosing an optimal design would involve simultaneously optimizing design characteristics on multiple responses. Therefore, traditional optimal designs focusing on a single response is inadequate, and there is a desire for optimal designs for experiments with multiple responses. This dissertation adds to the growing research by designing experiments that offer good estimation prediction for all the responses. For an experiment that aims to simultaneously study multiple responses, sometimes prior information is available from an earlier screening experiment or subject expert knowledge, about which subset of design factors might affect each individual responses. Traditional methods rely on the central composite design based on the full set of design factors, which often requires a large design size. In the first project of this dissertation, we propose to leverage such prior knowledge from screening experiment to construct more cost-effective optimal designs for simultaneously achieving precise estimation of multiple response models.

Particularly, the method utilizes the Pareto front approach [67] based on simultaneously optimizing D-efficiencies based on individual responses.

In recent decades, there has been increasing work on design selection based on multiple objectives [10, 67, 68]. As engineering problems become more complex in nature and with the possibility of resource constraints and other experimental challenges, most experiments are now conducted for achieving multiple objectives. These objectives often capture different characteristics of the design performance and are competitive due to the cost restriction, and different objectives may pull resources in different directions. Therefore optimal designs based on a single objective may be associated with poor performance on other criteria. The second project in this dissertation focuses on enhancing existing space-filling designs for computer experiments based on considering multiple objectives. Latin hypercube designs have been commonly used for creating space-filling designs with a nice uniform distribution when projected on each individual dimension. The maximin distance designs are popular for achieving maximum spread of design points, which maximally avoid collecting similar information from close-by locations. The minimax distance designs ensure maximum coverage over the full design space, which is useful to obtain good prediction throughout the entire input region. The Maximum projection designs offer nice space-filling properties when projected onto any subspace of design factors, which is desirable when only a subset of the factors are actively affecting the underlying process. To combine the merits of the above different designs, we propose using the Pareto front optimization approach to simultaneously optimize multiple space-filling criteria for Latin hypercube designs to achieve balanced performance on multiple objectives.

Despite the rapid advancements in design of experiment methodologies in the field of industrial statistics, many modern design techniques have not been adopted in the field of reliability analysis. The third project in this dissertation adopts the Bayesian optimal design method for developing optimal reliability test plans. Selecting optimal test plans

given a fixed budget is an important issue in planning for accelerated degradation tests (ADTs). Accelerated degradation tests are engineering procedures often used to evaluate the reliability of a product when the product is expected to have few or even no failures during a traditional life test. Particularly, the test units are exposed to elevated stress conditions to accelerate the failure process and a direct measurement of the physical degradation level at multiple points is used to model the underlying degradation path and then used for making inference about product reliability. Often in ADT planning, the decision about testing conditions and the number of test units that should be tested at each test condition is a bit ad hoc. Current work on optimal ADT test plans is primarily focusing on a single accelerating factor or two accelerating factors with a linear regression model. There has been little work on developing optimal test plans for more general degradation path models with two or more accelerating factors. In the third project, we develop a method based on utilizing Bayesian optimal designs to select optimal test plans for ADTs with two accelerating factors and using mixed effects models for characterizing the degradation paths. The general methodology can be adapted for more than two accelerating factors and more general forms of degradation path models.

1.1 Motivating Applications

1.1.1 Screening Experiments with Multiple Responses

Several designs have been proposed in literature for data collection and analysis with a single response and one or more input factors. However, there are situations where multiple responses are connected with a process, and need to be jointly considered for decision making. For instance, consider a multi-response experiment with ten input factors and four response variables [70]. Although common classical designs such as the central composite designs can be used to collect data and the full second-order models can be fitted for each of the four response independently, the immediate challenge with

these designs is that they often require more experimental runs and consequently leads to more computational time. In design problem where prior information is available from a screening experiment or previous process knowledge, a possibility is that each individual response is often affected by only a subset of all design factors. For instance, based on the past experiences of engineers in the production of a product, they can often suggest the factors having the most influence on each response. Information like this can be leveraged when estimating each response. Thus, resulting in a reduction in experimental runs.

Applying optimal design of experiments to the multi-response experiment described above, the specific goal would be to identify the set of operating conditions for simultaneously achieving precise estimation of the responses. In the first project discussed in this dissertation, we propose a design selection methodology for achieving this goal. The proposed methodology is illustrated on two screening experiment examples.

1.1.2 Space-filling Designs for Computer Experiments

There are generally two application categories of design of experiments; physical experiments and computer experiments. Design of physical experiments involves application of design of experiment techniques on physical processes. However, many physical processes are difficult or impossible to explore directly by conventional experimental methods. With advancement in computer power, computer experiments are increasingly used by engineers and scientists in many fields as the alternative when physical experiments are impracticable and intractable. For instance, computational fluid dynamics models are used to calculate air flow over a wing in the design of an aircraft wing [1]. Computer experiments are designed to study relationships between the inputs and outputs of a system or process using computer simulator rather than a physical experiment [96].

Despite the merits of computer experimentation over physical experimentation, obtaining data from complex computer models can be time consuming and computation-

ally expensive. A single run of the experiment could take longer minutes, hours and even days for execution. A surrogate function is a statistical model built to approximate the output from an expensive computer model. Gaussian process regression [94] is a popularly used surrogate model in computer experiments because of its adaptive, flexible and non-parametric nature [92].

In computer experiments, the design problem is how to choose a set of points where the computer model will be run to obtain simulated data that allows precise predictions from the surrogate model. In the design of computer experiments, the two major categories of design criteria are; the distance-based criteria and the model-based criteria. The distance-based criteria do not rely on model assumptions and generally result in a family of design called space-filling designs. The common space-filling designs for computer experiments are Latin hypercube designs that are optimal in one aspect of space-filling. In the second project of this dissertation, we propose a design selection methodology that uses the Pareto front optimization approach to optimize multiple space-filling characteristics in order to generate optimal Latin hypercube designs with improve and balanced performance for design of computer experiments.

1.1.3 Accelerated Degradation Test of Optical Media

Optical media are typically designed with high reliability such that, it is often impossible to observe sufficient failures under their normal use condition during traditional life test. Engineers often need to expose products like this to accelerating factors such as; voltage, temperature, or humidity, to speed up the failing process, so that degradation measurements can be obtained, analyzed, and extrapolated to make prediction about product life time under normal operating conditions.

We demonstrate our proposed application of Bayesian design of experiments method for accelerating degradation test plan, using an optical media data from the international organization for standardization (ISO) [101]. The ISO optical media data used consist of

four different testing conditions, where each testing condition is a combination of a level of temperature and a level of relative humidity. According to ISO standard, exposing optical media to these accelerating factors is believed to accelerate it degradation rate. More discussion on how we illustrate the proposed methodology on the ISO data is given in Chapter 5.

1.2 Aim and Objectives

The aim of this dissertation is to advance the methodology and application of optimal design of experiments in different areas. The objectives for three different problems are summarized below:

1. to demonstrate the process of leveraging on prior information from a screening study about the subset of the design factors that influence each response and application of the Pareto front optimization approach to develop a cost-effective design selection methodology for finding optimum locations in the input space, that offers balance and improve performance on simultaneously obtaining precise estimation of the responses in a multi-response experiment.

2. to demonstrate the proposed method of simultaneously optimizing multiple space-filling characteristics of design of computer experiments using the Pareto front optimization approach in order to generate optimal Latin hypercube designs that simultaneously balance good performance in multiple design objectives.

3. to extend the idea of design of experiments to accelerated degradation test planning. Specifically, applying Bayesian design of experiment method to finding optimum test plan and optimum units allocation for accelerated degradation tests where multiple accelerating factors are considered.

1.3 Dissertation Organization

This dissertation consists of three main projects. Chapter 2 provides general review on statistical design of experiments. Chapter 3 centers on the first project, titled "D-optimal Designs for Experiments with Multiple Responses". It addresses objective 1. Chapter 4 addresses objective 2. The project here is titled; "Multiple Objective Latin Hypercube Designs for Computer Experiments". Some contents of Chapter 3 and Chapter 4 have been presented by the author of this dissertation during the 2020 and 2021 Joint Statistical Meetings by the American Statistical Association [2, 3]. Chapter 5 addresses objective 3. The title of the project here is; "Bayesian Optimal Design for Accelerated Degradation Tests with Multiple Accelerating Factors". Finally, Chapter 6 presents the general conclusion that summarises the research presented in this dissertation.

Chapter 2: Review on Statistical Design of Experiments

2.1 Classical Design of Experiments

The concept of design of experiments was first introduced by Sir Ronald Fisher in a small agricultural research station in England [34, 115]. Fisher's pioneering work in design of experiments showed how experiments can be conducted in the presence of uncontrollable field conditions, such as temperature, soil condition, and rainfall. Fisher note that the effects of these uncontrollable factors can either be unsystematic (measurement error or random error) or systematic (bias). He introduced basic design principles for tackling random error and bias. The three major basic design principles are, randomization, replication and blocking [81]. Randomization is the random process of assigning treatments to the experimental units so as to remove bias. Replication means some runs or even the whole experiment could be performed more than once in order to directly estimate the magnitude and distribution of random error. Blocking is the sorting of experimental units into homogeneous groups to reduce variability of treatment effects. Over the years, these principles have been used by experimenters when conducting experiments.

Prior to Fisher's pioneering work, a traditional method of carrying out experiment is one factor at a time evaluation. In this method, all other factors are held constant during the experiment except the factor that is being studied. The limitation of this type of experiment is that it only provides an estimate of the effect of a single factor at selected fixed conditions of the other factors. Fisher proposed varying all the factors at once using a factorial design. That is, experiments should be run for all combinations of levels for all of the factors, in order to be able to estimate the effect of each factor when the other factors

are changing. This Fisher pioneering work is regarded as classical design of experiments [34].

A factorial design offers more useful information at a limited cost and time and is able to identify the optimal conditions much better and faster than a one factor at a time method [6]. Factorial designs are commonly used for experiments involving several design factors to investigate the joint effects of the factors on the response variable [81]. In particular, investigating main effects of design factors and the effects of their interactions on the response. A special case is one where each of the k factors of interest has only two levels. These designs are often called 2^k factorial designs and they allow full estimation of all main effects, all interactions of every order, and an intercept. The 2^k factorial designs are useful for screening experiments to identify important design factors. They also serve as a building block for response surface designs [6, 81].

A full factorial design is not always a feasible option in the presence of experimental constraints. As the number of factors in a 2^k factorial design increases, the number of experimental runs needed for a full replicate of the design can quickly outgrows the resources. For instance, a full replicate of a 2^5 design would need 32 runs. In this design, only 5 of the 31 degrees of freedom are used to estimate the main effects, and only 10 degrees of freedom are used to estimate the two-factor interactions. The remaining 16 degrees of freedom are connected with three-factor and higher order interactions. If the experimenter assumes that high-order interactions are likely trivial (based on the effect sparsity and hierarchical principles), then information on the main effects and low-order interactions may be obtained by running only a fraction of the full factorial experiment [81].

A fraction of the experimental runs produced by a full factorial design is called a fractional factorial design. It involves experimental manipulation of all k factors but includes fewer runs when compared to a full factorial design with the same k factors. The number of fraction is chosen based on the available resources and the effect sparsity principle [6].

Although a fractional factorial design allows for a more efficient use of resources as it reduces the design size, it however comes with loss of information. Fractional factorial designs are often used in screening experiments to identify active factors from a set of several factors. Thus, they are among the most widely used types of design in industry [78]. Fractional factorial designs can be projected into larger designs in the subset of significant factors and it is possible to combine the runs of two (or more) fractional factorials to assemble sequentially a larger design to estimate the factor effects and interactions of interest [81].

The factorial designs and fractional factorial designs allow one to fit first-order models [6]. To accommodate the possibility of curvature, second-order (quadratic) models are required. Two common response surface designs for estimating second-order models are; the central composite design [7] and the Box-Behnken design [6, 81]. The central composite design consists of a two-level factorial or fractional factorial design, axial points and center points [78]. The center points (experimental runs where all factors are set to zero) provide information about the presence of curvature in the process while the addition of the axial points allows efficient estimation of the quadratic terms. The axial points are experimental runs where all factors are held constant at zero except for the factor that is being varied. The varied factor is set to $+\alpha$ for one run and $-\alpha$ for another run, where α is often called a "design parameter" [70]. This is done for each factor, thus, there are $2k$ axial points for k factors. The Box-Behnken [8] design is the common alternative to the central composite design. Box-Behnken designs are second-order designs based on three-level incomplete factorial designs [33]. Box-Behnken designs usually have fewer runs than central composite designs, thus, they are less expensive to run with the same number of factors. However, they do not have the flexibility of including runs from a factorial experiment. Central composite designs are particularly useful in sequential experiments because previous factorial experiments can be built upon by simply adding axial points

and center points [81]. More discussion on central composite designs is given in Chapter 3.

2.2   Optimal Design of Experiments

The applications of design of experiments in process development and quality improvement have grown rapidly over the years [6, 28]. Although, practitioners are often only familiar with standard designs, such as the factorial designs, fractional factorial designs, central composite designs and the Box-Behnken designs. These standard designs work well when the research problem and the requirements of a classical design are a good match. However, there are many situations where the requirements of a classical design and the research problem do not align. For instance, classical designs are generally less flexible for experimental situations with a run size restriction, constrained design space, or when a nonstandard model is needed for estimating the response.

Designing experiments for situations like these often requires creating a design that suit the specific problem. This lead to the theory of optimal design of experiments. Optimal design of experiments involves understanding and implementing the underlying statistical model and assumptions to describe the true (and unknown) form of the relationship between the input and output variables. It seek to improve the statistical inference regarding the quantities of interest by the optimal selection of points for design factors under the control of the investigator, in the presence of existing constraints on experimental resources. Earlier research on the theory of design optimality started with Kiefer [55] and Kiefer and Wolfowitz [56]. In their work, they proposed computer algorithms that allowed "best" designs to be generated by a software package based on the aim of the experiment, choice of sample size, model, ranges on variables and other constraints. These designs are generally known as optimal designs or computer-generated designs [81].

Typically, experiments are conducted in order to understand the effects that different parameters of interest make on the response. Evaluation of these effects allows experimenters to make valid interpretations and comparisons regarding the scale and the form of the relationships between process parameters and measured output. When the exact true nature of relationships is unknown, some form of approximation about the relationships can be established. Usually, a second-order response surface model is fitted to the data obtained after an experiment. The response measured at each of the n experimental runs can then be presented as a linear combination of the design factors, their interactions and the quadratic terms. Although some of the terms of the model might be excluded, but unless stated otherwise, a second-order model is usually explored [81].

Therefore, a designed experiment can be express in a linear regression model form [78, 81]

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{Y}$ is an $n \times 1$ vector of the observation (that is, the response vector), $\mathbf{X}$ is the model matrix with dimension $n \times p$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of the regression coefficients, $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of random errors, n is the number of runs in the experimental design, and p is the number of model parameters. From the theory of linear models, the fitted regression model is $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, where $\hat{\mathbf{Y}}$ is regarded as the predicted response and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the least squares estimation of the regression parameters. The scaled prediction variance is the variance of the predicted response at a specific point(s) in the design space relative to the error variance σ^2, and it is defined as:

$$\frac{var[\hat{\mathbf{Y}}]}{\sigma^2} \times '(\mathbf{X}'\mathbf{X})^{-1} \times = V(\mathbf{X})$$

where $\times$ is a vector of the point(s) of interest in the design space.

2.2.1 Design Optimality Criteria

Optimal design of experiments express the goal of an experiment in terms of optimising an objective function, which is often regarded as "optimality criteria". An optimality criteria is a mathematical function defined to reflect the objective of the experiment as accurately as possible and it is use to summarize how good a design is. Many design optimality criteria have been proposed in literature, and they can be group into several categories based on the aim of the experiment and area of application. These optimality criteria are often labelled by the letters of the alphabet and are sometimes called *alphabetic optimality criteria*. Two of the classifications are; the distance-based criteria and the information-based criteria [91].

When the goal of the experiment is capturing the different behaviors of the response(s) in different areas of the design region, the criteria used in quantifying these objectives are regarded as the distance-based criteria. The distance-based criteria are criteria based on the distance $d(X, A)$ from a point X in the P–dimensional Euclidean space R^P to a set $A \subset R^P$ [91]. They are used for selecting a design based on a metric that quantifies the spread of a set of points. They do not rely on understanding of the underlying model of the process and are flexible in application. They are commonly used in the design of computer experiments. Common examples are; the minimax distance criterion, the maximin distance criterion and the maximum projection criterion. These criteria are further discussed in Chapter 4.

The information-based criteria are optimality criteria that are based on the information matrix of a design [91]. The information matrix is the inverse of the variance-covariance matrix for the least-squares estimates of the linear parameters of the model. When the unknown underlying model is approximated by a second-order linear model, information-based optimality criteria are commonly used [32, 91]. They can be summarized into two categories depending on the aspects of design characteristics that the experimenter is

interested in optimizing. They are, optimality criteria for precision in estimation of parameters and optimality criteria for precision in prediction.

2.2.1.1   *Optimality Criteria for Precision in Estimation of Parameters*

For experimental problems with the goal of identifying factors or effects with significant impact on the response variable, precision in the estimation of the regression coefficients is desired. The two common information-based optimality criteria for precision in estimation of regression coefficients are, the D-optimality criterion and the A-optimality criterion. While the D-optimality criterion seeks to maximize the determinant of the design moment, the A-optimality criterion seek to minimize the total variance of parameter estimates. The A-optimality criterion is less commonly used compare to the D-optimality criterion. More on the discussion of these criteria is given in Chapter 3.

2.2.1.2   *Optimality Criteria for Precision in Prediction*

Though precise estimates of the regression coefficients are important for some experiments, however, in experiments where precision in prediction is important, design optimality criteria that seek precision in the prediction ability of the model are most certainly desirable [49]. The prediction variance is the basic metric for measuring model precision. The optimality criteria in this category are mostly information-based and the two most commonly used are, the I-optimality criterion and the G-optimality criterion. While the G-optimality criterion [99] seeks to minimize the maximum prediction variance in the design region, the I-optimality criterion seeks to minimize the average prediction variance over the design region [49]. More discussion on these criteria is given in Chapter 3.

2.2.2   Bayesian Optimal Design of Experiment

When conducting an experiment, certain decisions needs to be made by the experimenter before data collection and, since data collection is often restricted by limitation in

experimental resources or time, an efficient use of available resources is desired. Optimal designs generally depend on the true values of the model parameters. Since the values of the parameters are unknown, and data has not been collected to estimate them, the experimenter must suggest values for the model parameters from which to construct an experimental design. Use of inappropriate parameter values may result in sub-optimal designs. When specific instructions are available prior to experimentation, Bayesian methods can play an important role in deciding how to allocate treatments efficiently, leading to more informative experiments. These instruction can be prior information from earlier similar experiments or process knowledge of the experimenters. The Bayesian approach to experimental design creates a formal way to include prior information in the form of probability distributions on the model parameters, into the design process for better statistical precision [40, 45, 80].

Bayesian methods involves combining prior knowledge about the unknown parameters of a model with the information provided by the data to the unknown parameters to produce the posterior distribution, from which inferences on the unknown parameters of interest can be made. In experimental design, Bayesian methods allows incorporating prior information and or uncertainties about the statistical model with a utility function (which can be regarded as a criterion function) which describes the goal of the experiment. Bayesian experimental design is based on a decision-theoretic approach [12, 63], where an optimal allocation of experimental resources is determined via maximization of the expected utility. The Bayesian optimal design, η^*, maximises the posterior expected utility function $U(\eta)$ over the design space $\boldsymbol{\Omega}$ with respect to the future data $\mathbf{y}$ and model parameters $\boldsymbol{\theta}$:

$$
\begin{aligned}
\eta^* &= max_{\eta \in \Omega} E\{U(\eta, \boldsymbol{\theta}, \mathbf{y})\} \\
&= max_{\eta \in \Omega} \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} U(\eta, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\eta) d\boldsymbol{\theta} d\mathbf{y} \\
&= max_{\eta \in \Omega} \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} U(\eta, \boldsymbol{\theta}, \mathbf{y}) p(\mathbf{y}|\eta, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{y},
\end{aligned}
\tag{2.1}
$$

where $p(\mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\theta})$ is the likelihood for observing a new set of measurements $\mathbf{y}$ at the design points $\boldsymbol{\eta}$, given parameter values $\boldsymbol{\theta}$, and a prior distribution $p(\boldsymbol{\theta})$ for the parameters. Equation 2.1 does not usually have a closed form solution to allow analytical evaluation of the integration problem. When a closed form does not exist, numerical approximations and stochastic methods are used to solve the maximisation and integration problem [11, 12]. Chaloner and Verdinelli [12] illustrate using large-sample approximations to estimate the posterior expected utility. Overstall and Woods [86] demonstrate how a Gaussian process emulator can be used to approximate the expected utility as a function of a single design coordinate in a series of conditional optimization steps. Methods like these are often used when the criterion function is analytically intractable and computationally burdensome.

Just like with the design optimality criteria discussed above, it is important that the selected Bayesian utility function tailored the experimental aim. For instance, designs which are optimal for precise estimations of model parameters may not be useful for prediction of future observations. Many Bayesian utility functions have been proposed in literature [12]. They can be classified into three categories; utilities for parameter estimation, utilities for model discrimination and utilities for prediction of future observations [93].

The common goal in experimental design is obtaining precise estimation of model parameters and several utility functions have been proposed in literature for achieving this. A common utility function for this is the maximization of the Shannon information between the posterior and prior distributions [16, 58], which corresponds to the classical D-optimality criterion when designing for normal linear models with a normal prior distributions for the model parameters [12, 107]. When there is an interest in obtaining a point estimate of the parameters, the quadratic loss function is often used as the utility function [12, 40]. This utility function is the Bayesian form of the classical A-optimality criterion. For model discrimination, mutual information is commonly used as the utility function. The optimal design $\boldsymbol{\eta}$ is the one that maximises the mutual information between the model and the future observation [25, 72].

If the aim of the experiment is predicting future observation $\mathbf{y}_{n+1}$ from $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_n)$, then the expected gain in Shannon information for a future observation, $\mathbf{y}_{n+1}$, from the prior predictive distribution to the posterior predictive distribution can be used as the utility function [12]. For minimising the variance of the expected response, Solonen et al. [100] proposed placing the next design point where the prior variance of the mean response is largest. Utility functions that minimize the variance of an estimator of a quantity of interest have also been proposed for prediction of future observations [90, 118].

More discussion on Bayesian experimental designs can be found in exiting review papers [12, 16, 20]. Chapter 5 of this dissertation discuss our proposed application of Bayesian design of experiments method for accelerated degradation test plan with two accelerating factors.

2.3 Multiple Objective Optimization

Sometimes, experimenters have multiple interest in the properties of the model parameters. Thus, the quality of the desired design is a function of multiple criteria, each representing different aspects of design performance. In this context, a "good" design is the one that demonstrate expected performance in multiple attributes. For instance, suppose the the goal of an experiment is precision in the estimation of regression coefficients, then the D-optimality criterion, known to quantify the overall precision of the estimated model parameters is evaluated in the experiment. Suppose further that another criterion might be of interest to the experimenter, for example, the I-optimality criterion, that is minimising average prediction variance, then the experimenter seek for a design with balance performance in precise estimation of parameter estimates and minimising average prediction variance. Simultaneously optimizing the two criteria would result in some trade-offs being made. This is the goal in multi-objective optimization.

Although there are different variety of optimality criteria used in quantifying different design objectives, but a design which is optimal according to a particular criterion might

perform poorly in respect of another criterion. Thus, for experiments with competitive objectives, it is desired that designed experiments should simultaneously optimize these objectives. Earlier works in generation of optimal designs have seen applications in experiments with multiple responses or objectives [13, 18, 32]. These earlier methods often require strong statistical assumptions. Atkinson et al. [4] introduced the idea of weighted combination of the criteria, which allows finding a compromise between two or more desirable properties while searching for optimal designs. This method explores how the changes in the allocation of weights between the individual criteria affect the performance of the optimal designs [26]. Thus, selection of optimal designs for multiple objective studies requires studying the trade-offs between the criteria in order to achieve balance in the competing objectives. The desirability function approach [23] and the Pareto front optimization approach [67] are the commonly used techniques for simultaneously combining multiple objectives, particularly in the area of design of experiments [14].

2.3.1 Review of Algorithms for Constructing Optimal Designs

For any kind of optimality criteria consider in a design problem, generating optimal designs requires the use of optimization (computer) algorithms, otherwise known as search or heuristic algorithms. An optimization algorithm is a procedure which is executed iteratively by comparing various solutions till an optimum or a satisfactory solution is found. The literature on search algorithms for finding optimal designs is a huge and diverse one. Overall, an effective algorithm should be flexible, computationally efficient, easy to implement, and usable for the intended purpose. A single algorithm is not expected to perform well for all types of optimization problems. For instance, algorithms that scale well for small dimensional problems may not scale well for high dimensional problems. Most algorithms usually starts with a starting design, an improvement is then made on the current design based on a procedure specific to the choice of algorithm. Also, a stopping rule is usually required for the algorithm to terminate its search pro-

cess. Common choices are, setting the number of iterations that the algorithm will run at the initiation of the algorithm or stopping the search process when the criterion values converges to a specified value.

The development of optimal design in theory is based on the use of a continuous design measures that determines the proportion of runs that should be made at a number of points in predetermined design space. Since the concept of continuous design is usually not feasible in practice, the idea therefore is to determine discrete designs that approximate the optimal continuous design. Generally, algorithms for optimization problems can be grouped as; enumerative, deterministic, or stochastic. An enumerative approach involves evaluating all possible designs based on the criterion of interest and selecting the design with the best criterion value. This approach is often infeasible even for a region of moderate size [10].

Deterministic algorithms uses specific rules for moving from one solution to another and they result in the same design when they are run repeatedly using the same set of input parameters. There is no randomness built into these algorithms and they typically have a proof of convergence to the theoretical optimum. Common examples are, the Wynn algorithm [113], the Fedorov algorithm [32] and the exchange algorithms [17]. Exchange algorithms are the most commonly used deterministic algorithms and they are found to be effective for finding exact optimal designs when the model has several discrete factors [76]. The two common types of exchange algorithms are; the point exchange algorithm and the coordinate exchange algorithm [17, 84].

Applications of deterministic algorithms can be infeasible and or computationally prohibitive for some experimental situations. This lead to the introduction of stochastic optimization algorithms. Stochastic algorithms introduces one or more probabilistic components to the search, so that a particular starting design may produce a variety of solutions. They usually do not guarantee convergence to the global optimum, but unlike the deterministic algorithms, they generally do not require the user to make assumptions

on the criteria to be optimized. Moreover, the criteria do not have to be differentiable and can easily escape being trapped in the local optimum. However, their major drawback is that, they usually involve tuning parameters, which tend to have significant impact on the performance of the algorithms [19]. The class of stochastic algorithms that have been extensively discussed in literature are the nature-inspired algorithms. Examples of these algorithms are; the simulated annealing [75], genetic algorithms [42, 77] and the particle swarm optimization [109].

For optimization of multiple objectives, several search algorithms have been proposed for selecting optimal designs. These algorithms can be modification of standard stochastic algorithms [47, 79, 89] or modification of standard deterministic algorithms [67, 95] and some are hybridization of deterministic and stochastic algorithms [10, 22, 98, 114]. The most commonly used of these set of algorithms are the ones that uses an exchange method for searching through the design space.

More discussions on the specific algorithms used in the projects discussed in this dissertation are given in relevant Chapters.

Chapter 3: D-optimal Designs for Experiments with Multiple Responses

3.1 Introduction

With increasing limitation on resources and experimenters wanting to achieve more from running the experiments, there has been growing interest in considering multiple responses in a single experiment. In this case, the goal for such an experiment often involves obtaining good estimation of the underlying input and response relationship for multiple responses. The current literature on optimal designs has been primarily focusing on consideration of a single response. There has been limited work on design of experiments considering multiple responses. In addition, existing work on experiments with multiple responses are often built in the super space of all design factors that may be related to any of the response of interest and are based on a general model for all responses. This requires a large number of experimental runs to explore the full design space. However, in real applications, there is often only a subset of design factors that are relevant to each response. Therefore, there is room to save experimental resources by considering different models for individual response and optimizing design performance on estimating multiple different models that tailored for each individual response.

Response surface designed experiments are often used to describe relationships between several design factors and response(s) of a process of interest. One of the most popular response surface designs is the central composite design (CCD) [7]. A CCD often includes a two-level full factorial or fractional factorial design with center runs, plus a group of axial runs (otherwise known as star runs) that allow estimation of quadratic terms. When constructing a fractional factorial design, one or more of the factor effects

are aliased with higher order interactions, therefore allowing the use of fewer runs to estimate the low order factor effects of interest. Typically, to allow clear estimates of the main effects and maybe some two factor interactions, a fractional factorial design with the maximum resolution given the affordable run size is preferred.

The resolution of the fractional factorial in a CCD determines how the factor effects are aliased with other effects [7]. Resolutions III, IV, and V are commonly seen in fractional factorial designs. For resolution III designs, main effects are not aliased with each other, but they are aliased with two-factor interactions (i.e. their estimates cannot be separated). For resolution IV designs, main effects are not aliased with any other main effect or two-way interactions, but the two-factor interactions are aliased with some other two-factor interactions. For resolution V designs, no main effect or two-factor interaction is aliased with another main effect or two-factor interaction, hence all the main effects and two-way interactions can be clearly estimated. Resolution V fractional factorials are commonly used in CCDs because they allow clear estimation of all the first order and two-way interaction terms in a response surface model. Because designs of greater resolutions require more experimental runs, fractional factorials of resolutions higher than V are usually avoided in CCDs.

The CCDs are often used to estimate full second-order response surface models in a cuboidal or spherical input space. When a CCD is used to collect data for estimating multiple response models, the CCD is often built in the design space spanned with all the factors which may affect any of the response, which we refer to as the super design space. Considering the fact that often only a small subset of the factors may affect each individual response, the CCD based on the super design space is often much larger than what is needed for estimating each individual response. In this case, if a screening experiment or process knowledge could be used to identify potential active factor effects for each response, then this information can be leveraged to select smaller experiment that allows precise estimation of the candidate models for individual response.

For example, the production of co-fire fuel pellets involves several input factors including coal particle size, coal moisture content, pellet aging temperature and amount of bio-oil binder and multiple response variables including the heating value, moisture content, carbon content and strength of the pellets. Constructing a CCD even based on aliasing some effects would require many experimental runs. Friend [35] in an experiment considered varying only the four input factors mentioned above while holding the others constant to investigate the effects of the factors on the four properties (response variables) mentioned above. This experiment can be regarded as a screening study that helps to identify the factors influencing each response. For instance, one of the result from Friend's experiment is that coal particle size and coal moisture content are related to the heating value, but pellet aging temperature and the amount of bio-oil binder are not. The result from an experiment like this can then be used to design a follow-up experiment for efficient estimation of each response variable.

A screening experiment is a study focused on identifying factors with much influence on each response, rather than a more detailed study of quantifying relationships between the input and response. Information from a screening experiment could suggest that different subsets of the design factors would affect different responses and what are some of the two-factor interactions that are likely to affect a particular response. By leveraging subject matter knowledge and/or information from an earlier screening experiment, Marget and Morris [70] proposed a unique factor central composite designs (UF-CCD) by modifying the fractional factorial portion of the CCD. Particularly, for an experiment that involves k factors, the fractional factorial portion of the UF-CCD is constructed such that, for each response, the factors associated with that response based on information from the earlier screening experiment form either a full factorial or a resolution IV or V fractional factorial. That is, the design will be a full or fractional factorial for only the factors known to relate to each response, but not for all k factors, and hence will require fewer runs than the traditional CCD based on all k factors. Also, in the construction of the UF-CCD, pairs

of effects that do not appear in any model are aliased. For instance, if factor 1 and factor 2 do not appear together in the same response model, then the levels for factor 1 can be exactly repeated for factor 2 in the factorial part of the CCD, thereby allowing the design size to be reduced further. More details about the construction of UF-CCD is available in Marget and Morris [70].

Both CCDs and UF-CCDs are classical designs. Despite their numerous applications, classical designs are generally less flexible in terms of the design size, region of input values, and the constraints on the randomization structure. Optimal experimental designs have become popular due to their flexibility on the above mentioned aspects, especially when constraints on the design size, region or the randomization structure preclude the use of classical designs. Computer-aided designs are often generated based on a user selected design optimality criterion, and are generally referred to as optimal designs [47]. The choice of the design optimization criterion depends on the goal of the experiment. Some focus on obtaining precise estimation of the model parameters while others emphasize on obtaining precise prediction of the response values throughout the input space. The commonly used criteria include D-, A-, G-, and I-optimality, where the first two focus on design estimation while the last two emphasize more on prediction.

In this chapter, we propose a cost-efficient design selection strategy based on utilizing information from a screening experiment to construct D-optimal designs with balanced performance on simultaneously obtaining precise estimation of the multiple responses. Particularly, we seek the Pareto front based on D-efficiencies of the multiple responses and then further select best designs based on exploring secondary criteria that examine the design performance on other key aspects. To efficiently search through the design space and identify the Pareto front based on the D-efficiencies of the multiple responses, we developed a Pareto aggregate coordinate exchange (PACE) algorithm that scales well for experiments with a larger number of design factors. The proposed methods are demonstrated with examples of varied number of input factors.

The rest of the chapter is organized as follows. Section 3.2 describes some commonly used optimality criteria for capturing different characteristics of the design performance. Section 3.3 introduces the Pareto front optimization approach for simultaneously optimizing the precision of multiple estimated response surface models, and it describes the proposed PACE algorithm and its implementation in identifying the Pareto front. Section 3.4 illustrates the proposed method using two examples with varied dimensions and demonstrates advantage of the generated D-optimal designs over alternative classical designs. Section 3.5 renders the conclusion.

3.2   Common Criteria for Optimal Designs

Based on the goal of an experiment, an optimality criterion is selected for seeking the optimal design. The criterion measures the design performance on a certain aspect of interest. The most commonly used optimality criteria in optimal experimental designs include D-, A-, G-, and I-optimality criteria [91].

When the goal of the experiment is to maximize the precision of the estimated model parameters, D-optimality is the most commonly used optimality criterion. The D-criterion measures the determinant of the design moment matrix, which is equivalent to minimizing the volume of the confidence region of the regression parameters [81]. The D-criterion is defined as

$$|M(\xi)| = \frac{|X'(\xi)X(\xi)|}{N^p},$$ (3.1)

where $X(\xi)$ is the $N \times p$ model matrix for design ξ with p design parameters and β is the $p \times 1$ vector of regression coefficients. The D-optimal design, denoted by ξ_D^*, maximizes $|M(\xi)|$ over the design space spanned by all possible designs, $\xi \in \Omega$, i.e. $|M(\xi_D^*)| = Max_{\xi \in \Omega}|M(\xi)|$. Then the D-efficiency of a design, ξ is given by

$$D_{eff}(\xi) = \left(\frac{|M(\xi)|}{|M(\xi_D^*)|} \right)^{\frac{1}{p}},$$ (3.2)

which measures the relative performance compared to the D-optimal design.

Different from D-efficiency that measures the generalized variance of the confidence region of the estimated model parameters, the A-optimality criterion evaluates only the total variance of the estimated parameters. The A-optimality criterion is defined as

$$A(\xi_{\mathbf{A}}^*) = min_{\xi \in \Omega} \{ trace \left[X(\xi)'X(\xi) \right]^{-1} \},$$

where $\xi_{\mathbf{A}}^*$ is the A-optimal design. Then the A-efficiency of a design ξ is given by

$$A_{eff}(\xi) = \frac{tr[M^{-1}(\xi_{\mathbf{A}}^*)]}{tr[M^{-1}(\xi)]}.$$

Instead of emphasizing good estimation, there are some other criteria primarily focusing on good prediction. The G-optimality criterion seeks to minimize the maximum prediction variance over the design space and it is defined as

$$G(\xi_{\mathbf{G}}^*) = \min_{\xi \in \Omega} \max_{x \in \chi} var(\hat{\mathbf{y}}_{\mathbf{x}} | \xi),$$

where $\xi_{\mathbf{G}}^*$ denotes the G-optimal design and $var(\hat{\mathbf{y}}_{\mathbf{x}}) = \sigma^2 f^T(x)(X(\xi)^T X(\xi))^{-1} f(x)$. The G-efficiency of a design ξ is defined as

$$G_{eff}(\xi) = \frac{p}{\max_{x \in \chi} var(\hat{\mathbf{y}}_{\mathbf{x}} | \xi)},$$

where p is the number of parameters in the model.

Different from the G-optimal design that emphasizes the optimal performance on the worst case scenario, the I-otpimal design focuses on the average predictive ability throughout the input space. The I-optimality criterion minimizes the average prediction variance across the design space and it is defined as

$$I(\xi_I^*) = \frac{N}{\sigma^2} \int_\chi var(\hat{\mathbf{y}}_\mathbf{x}|\xi)\,d\mathbf{x},$$

where χ is the region of interest and $\int_\chi$ is uniform measure on χ, with total measure 1. This integral simplifies to give

$$I(\xi_I^*) = trace[(MM_x^{-1}))],$$

where $M = \int_\chi f^T(x)f(x)\,dx$ is the moment matrix of the region of interest [91]. The I-efficiency of a design ξ is defined as

$$I_{eff}(\xi) = \frac{tr[MM^{-1}(\xi_I^*)]}{tr[MM^{-1}(\xi)]},$$

where ξ_I^* is the I-optimal design. Note that, all the D-, A-, G-, and I-efficiencies measure the performance of a design relative to the optimal design based on the chosen criterion, and hence take values in the $[0, 1]$ range.

3.3 Optimal Experimental Designs for Multiple Responses

There have been some existing work on optimal designs for experiments with multiple responses. These existing methods have varying degrees of flexibility and result in limitations in applications. The MD-optimality as an extention of D-optimality for multiple responses proposed by Fedorov [32] requires the variance-covariance matrix of the responses to be known. However, such information is often not easy to obtain before data collection. To avoid this assumption, Cooray-Wijesinha and Khuri [18] suggest estimating variance-covariance matrix via an initial design, and then the augmented design locations are selected by leveraging the estimated variance-covariance matrix. Chang [13] proposes a design that allows each response to have either a complete first or second order model based on all design factors. Although this approach avoids estimating the

variance-covariance matrix, it does not allow the flexibility of using customized models for each individual response as suggested by screening experiment to reduce experimental runs.

The goal of this work is to select D-optimal designs that simultaneously optimize the estimation precision of all the responses by leveraging prior information on the model forms of individual responses. In the derivation of the D-criterion value for each response model, $X(\xi)$ from Equation 3.1 describe the design matrix based on each specified response surface model. As suggested by the screening experiment, the subset of the design factors affecting each response is used to construct the design matrix for each response model. For each individual response, the full second-order model based on the identified influencing factors is used. Hence, the design matrix for each response varies due to the difference in the subsets of factors influencing each response. Note that if more information on active factor effects is available from the earlier screening experiment, it can also be included in the model formation for seeking the D-optimal designs.

3.3.1   Pareto Front Optimization

There has been rich literature on design optimization based on considering multiple objectives. The conditional optimization approach [50] involves finding designs with optimal performance on a selected primary criterion while achieving the thresholds for other secondary criteria. Another classical method is to combine the multiple criteria into a single metric and then select an optimal design based upon this metric. The most common way of doing this is the desirability function (DF) approach [23, 41]. The DF approach combine multiple criteria into a single metric and then selects a design that optimizes the metric based on a set of user specified weights. The weights are regarded as the relative contribution of individual criterion. The DF approach involves transforming the different criteria considered to a 0-1 desirability scale, where 0 represent the worst value of the criteria and 1 represent the best value of the criteria. The two basic forms of the DF are;

the additive DF and the multiplicative DF. While the scaled criteria are combined as a weighted sum in the additive DF, the scaled criteria are combined into a product with the weights entering as exponents in multiplicative DF.

Different approaches can be used to search for optimal designs that maximize the overall metric (desirability). The point exchange and coordinate exchange algorithms (discussed further in Subsection 3.3.2) are commonly used in the search for optimal designs. The Direct search methods are often used for nondifferentiable desirability functions [59]. Heuristic approaches such as simulated annealing [57] and genetic algorithms [42] are frequently used for more complicated problems with highly nonlinear or multimodal objective functions.

Although the desirability function approach has been used extensively, it fails to directly consider the trade-offs between criteria and the selected optimal designs depend heavily on the user specified weights and scaling schemes. On the other hand side, the weight choice is often subject to ambiguity. Exploring a variety of different weight choices in a sensible range can be time consuming and computationally inefficient since every set of weights requires a separate search and is likely to require repeated evaluations of the same candidate designs across different searches using different weights [67].

An alternative to the desirability approach is the Pareto front optimization approach, which has been extensively used for multi-objective optimization in many disciplines [38, 104]. Lu et al. [67] first introduced the Pareto front optimization to design optimization based on simultaneously considering multiple objectives. The method identifies the Pareto frontier based on multiple objectives and select best deigns based on explicitly evaluating trade-offs between the competing criteria and the impacts of weighting, scaling and DF forms on the selected solutions. Various algorithms have been proposed for constructing and populating the Pareto front [67, 89, 95] and for evaluation and comparison of designs identified on the front [65, 67, 69]. A major advantage of the Pareto front approach is that, the population for Pareto optimal solutions occurs along the search and

avoid repeated evaluation of candidate designs. The method is flexible to adapt for using different design optimization criteria for meeting different experimental goals.

Below we describe the mechanisms for constructing a Pareto front based on considering multiple criteria. Without loss of generality, assume the goal of a multiple-objective design optimization problem is to maximize $C (\geq 2)$ criteria simultaneously given constraints on the input factors. Let $\boldsymbol{\xi} = (\mathbf{d}_1', \mathbf{d}_2', ..., \mathbf{d}_N')' \in \boldsymbol{\Omega}$ be a design matrix of dimension $N \times k$ where N is the number of design points and k is the number of design factors, the set of all possible $N \times k$ design matrices for a given candidate set of points is denoted by $\boldsymbol{\Omega}$. The candidate set, otherwise known as the design space, is a collection of treatment combinations from which the search algorithm chooses the treatment combinations to include in the design. Let $F(\boldsymbol{\xi}) = (f_1(\boldsymbol{\xi}), f_2(\boldsymbol{\xi}), ..., f_C(\boldsymbol{\xi}))^T$ denote the vector of criteria values corresponding to the design matrix, $\boldsymbol{\xi}$, then the space containing all obtainable criteria vectors is called the criterion space. A solution $\boldsymbol{\xi}_1$ is said to *Pareto dominate* another solution $\boldsymbol{\xi}_2$, (i.e, $\boldsymbol{\xi}_1 > \boldsymbol{\xi}_2$), if $f_j(\boldsymbol{\xi}_1) \geq f_j(\boldsymbol{\xi}_2)$ for all $j \in \{1, 2, ..., C\}$ and there exists at least one $j \in \{1, 2, ..., C\}$ such that $f_j(\boldsymbol{\xi}_1) > f_j(\boldsymbol{\xi}_2)$. In this case, the criteria vector $F(\boldsymbol{\xi}_2)$ is said to be *Pareto dominated* by $F(\boldsymbol{\xi}_1)$. In this work, the criteria vector for a particular solution is regarded as a *point* in the criterion space. A solution is *Pareto optimal* if and only if no other solution Pareto dominates it and its corresponding criteria vector is a *non-dominated vector*. We refer to the set of Pareto optimal solutions as the *Pareto optimal set* and the corresponding set of criteria vectors forms the *Pareto front*. Marler and Arora [71] provide a detailed review on the Pareto front concepts.

3.3.2   Pareto Aggregate Coordinate Exchange Algorithm

Enumerating all possible designs in the experimental region can be prohibitive or computationally challenging, even for a region of a small candidate set. Instead, an initial design can be generated and successfully improved by using an exchange algorithm. A commonly used approach is the point exchange algorithm, which iteratively replaces

each row in the current design by a better alternative from the candidate set until no improvement can be made based on the chosen criterion. Although the point exchange algorithm is very popular for design optimization, it can be computationally expensive for experiments of high dimensions (hence a large candidate set) and/or complex design criteria, thereby precluding efficient evaluation [91]. An alternative to avoid the computational difficulty associated with dealing with a large candidate set is to use the coordinate exchange algorithm [84], which replaces each element in the design matrix by best alternative treatment level for improving the designs. Since the result from each single search is dependent on the selected initial design, multiple random starting designs are often employed to seek the global optimal design.

When multiple objectives are considered, the point exchange procedure searches the entire candidate set for a replacement of each run in the design matrix to improve at least one of the criteria without deteriorating others. This search procedure can become computationally challenging as the number of design factors increases and results in a quick growth of the size of the design space. Since the point exchange only updates the current design with a strict improvement on the criteria values, designs which do not dominate and are not dominated by the current design are discarded, when there is a possibility that the discarded designs might be promising solutions in the Pareto optimal set and could be optimal solutions when different user priorities are considered. Although repeating the search with multiple random starts might help identify more points on the Pareto front, however, identifying the Pareto front based on directly combining the optimal designs from multiple random starts would not be efficient as it will require a very large number of random starts and unavoidably requires repeated evaluations of same designs across the search with multiple random starts. To eliminate this challenge, Lu et al. [67] developed the Pareto aggregate point exchange algorithm (PAPE), to improve the efficiency of the traditional point exchange algorithm for populating the Pareto front

by keeping track of all the non-dominated points and building the Pareto front along the searching process.

The coordinate exchange on the other hand does not require a candidate set. Instead it randomly generates a starting design and exchanges each entries of the design matrix to search for an improvement. The absence of a candidate set for the coordinate exchange procedure reduces demand on computer memory which makes it scale well for high-dimensional regular design space. In this chapter, we adapt the Pareto front optimization for the coordinate exchange algorithm, and proposed a new Pareto aggregate coordinate exchange (PACE) algorithm for simultaneously optimizing the estimation of multiple responses based on D-optimality. The proposed method can be easily adapted for considering other optimality criteria.

The proposed PACE algorithm begins with a randomly generated initial design with a non-singular moment matrix, and then replaces each coordinate in the design matrix by the best alternative level to improve at least one of the criteria without deteriorating others. The steps of the PACE algorithm are given below:

1. Randomly generate an initial design of N runs and k factors with a non-singular moment matrix ($|X'X| \neq 0$) to ensure all parameters of the specified model are estimable. Denote the current design as ξ and evaluate the user-specified C-dimension criteria vector, $F(\xi) = (f_1(\xi), f_2(\xi), ..., f_C(\xi))^T$. Note, for our application, $F(\xi)$ is a vector of D-efficiencies for multiple responses.

2. Initialize two null sets: the Pareto optimal set (i.e the set of Pareto designs) denoted as P and the set containing the corresponding criteria vector that forms the Pareto front, denoted as PF; then add the current design ξ to P and add the criteria vector of the current design $F(\xi)$ to PF.

3. For the current design, ζ, the search process is carried out as follows; for $i = 1, ..., N$ and $j = 1, ..., k$, swap each (i, j) coordinate of ζ with an alternative factor level to produce a new design ζ^*. Two comparisons are then made:

   (a) The first comparison is to determine whether the current design ζ should be replaced by the new design, ζ^*. If the new design improves at least one of the criteria without deteriorating any other criteria (i.e if $\zeta^* > \zeta$), then the current design is replaced with the new one (i.e., $\zeta = \zeta^*$).

   (b) The second comparison is to determine how to update the current Pareto front and the Pareto optimal set at the presence of the new design, ζ^*. If ζ^* dominates at least one of the designs in the "current" set, then ζ^* is added to P and the designs dominated by ζ^* are removed. If ζ^* neither dominates nor is dominated by any designs in the current set, then simply add ζ^* to P. If ζ^* is dominated by at least one of the designs in the current Pareto set, then both the Pareto set and the Pareto front remain the same and the new design is discarded.

   (c) Repeat steps 3(a) and 3(b) iteratively for all coordinates in the design and all possible alternative factor levels. The search terminates if no improvement can be made by replacing any coordinate with any alternative factor level. At the end of the search, there are d non-dominated designs in the Pareto optimal set P (i.e $P = \{\zeta_1, \zeta_2, ..., \zeta_d\}$) and d associated criteria vectors in the set PF (i.e $PF = \{F(\zeta_1), F(\zeta_2), ..., F(\zeta_d)\}$).

4. Repeat steps 1 to 3 with S different random starts. The final set of Pareto optimal designs and the Pareto front are obtained by combining the results from S separate searches.

3.3.2.1   Further Design Selection Techniques

After finding the set of optimal solutions, the decision-making involves examining the trade-offs between the individual criteria and selection of a single or reduced set of solutions from the Pareto optimal set. In this work, we explore other design characteristics to further evaluate the performance of the designs in the Pareto optimal set in order to facilitate informed decision making.

3.4   Applications

In this section, the proposed method and the implementation of the PACE algorithm is illustrated using two examples of varied dimensions of the input space. The performance of the selected D-optimal designs is then compared with that of the UF-CCDs [70] on a variety of design characteristics.

3.4.1   Five Factor Experiment with Multiple Responses

First, we consider the example from Marget and Morris [70]. The experiment involves studying four response variables with the prior knowledge from an earlier screening experiment given in Table 3.1. For each response, the factors that may be related to that response are labeled by an "X" in the row. For example, Response 1 is likely to be affected by Factors 1,2 and 3 but not by Factor 4 or 5 in the experiment, while Response 4 is likely to be only affected by Factors 1 and 4. We can see that each individual response is associated with a different subset of the design factors and no response is affected by all the factors involved in the experiment.

Our goal here is to select the best design to provide good estimation of all the response models. Hence, we seek for the optimal designs with balanced performance in the D-efficiencies for all the responses. We consider a 20-run design of five factors and five levels for each factor with the coded values at $-2, -1, 0, 1, 2$. We select the same factor levels as

Table 3.1: Result of a screening experiment with five factors

Response	Factor				
	1	2	3	4	5
1	X	X	X		
2		X	X	X	
3	X		X		X
4	X			X	

in the example from Marget and Morris [70] to ensure an easy comparison between D-optimal designs selected by the PACE algorithm and the resolution IV unique factor CCD (Res IV UF-CCD) of the same size suggested by Marget and Morris [70] (see Appendix A.1). The Res IV UF-CCD consists of 8 fractional factorial points, 8 axial points and 4 center points.

Four criteria are considered by calculating the D-efficiency of the design for the full second-order model based on the subset of relevant factors identified in Table 3.1 for each individual response. Then the PACE algorithm is used to populate the Pareto front based on the D-efficiencies for all four responses. We explored 40000 random starts, the identified Pareto front contains only a single point. Note that the result remains stable after 20000 random starts and no other design has been found to have higher D-efficiency values for any of the four responses. Note that this example is unusual in that there exists a universal best design that simultaneously optimize the estimation precision for all four responses. However, for general applications, there are often trade-offs between the different objectives which may pull resources in different directions. Hence it requires more thoughtful balance between multiple objectives based on understanding the amount of trade-offs between alternative designs. A rich set of numerical and graphical tools can be borrowed to facilitate informed decision making [67, 69].

To further compare the optimal design selected by the proposed method and the UF-CCD suggested by Marget and Morris [70], we evaluate the design performance based on multiple characteristics. Since we are most interested in the precision of the estimated

model coefficients for any of the response, we evaluate the normalized standard deviation of the estimated effects for different response models. Note, the normalized standard deviation of an estimated effect/coefficient is dependent on the design matrix of the particular response. For each response, a second order linear regression model is fitted for the selected D-optimal design and the Res IV UF-CCD, and the normalized standard deviations of the model coefficients are estimated. Figure 3.1 shows each response average normalized standard deviations for all the coefficients (all effects), first-order term coefficients (main effects), interaction term coefficients (interaction effects), and quadratic term coefficients (quadratic effects) for both designs. These indices are estimated to reflect the performance of a design with respect to estimation of different groups of coefficients in each response model.

Figure 3.1: Average normalized standard deviations of model coefficients for the D-optimal design and the Res IV UF-CCD

As observed in Figure 3.1, the standard deviations of model coefficients for each response model are substantially higher for the Res IV UF-CCD compare to those of the D-optimal design, implying that the D-optimal design appears more efficient in terms

of providing precise estimates of the model parameters. To further compare the performance of the designs for estimating each individual responses, we obtained the individual response D-, A-, G- and I-efficiency values for both designs. In Figure 3.2, it can be seen that, the efficiency values of these criteria for each response model are significantly higher for the D-optimal design, further justifying the better performance of the D-optimal design on both the estimation and prediction aspects. For both designs, the D-efficiency value for response 4 is slightly lower than those of the remaining responses and its A-, G- and I- efficiency values are considerably higher than those of the remaining responses. We think this is a result of having a reduced number of parameters in response 4 model.
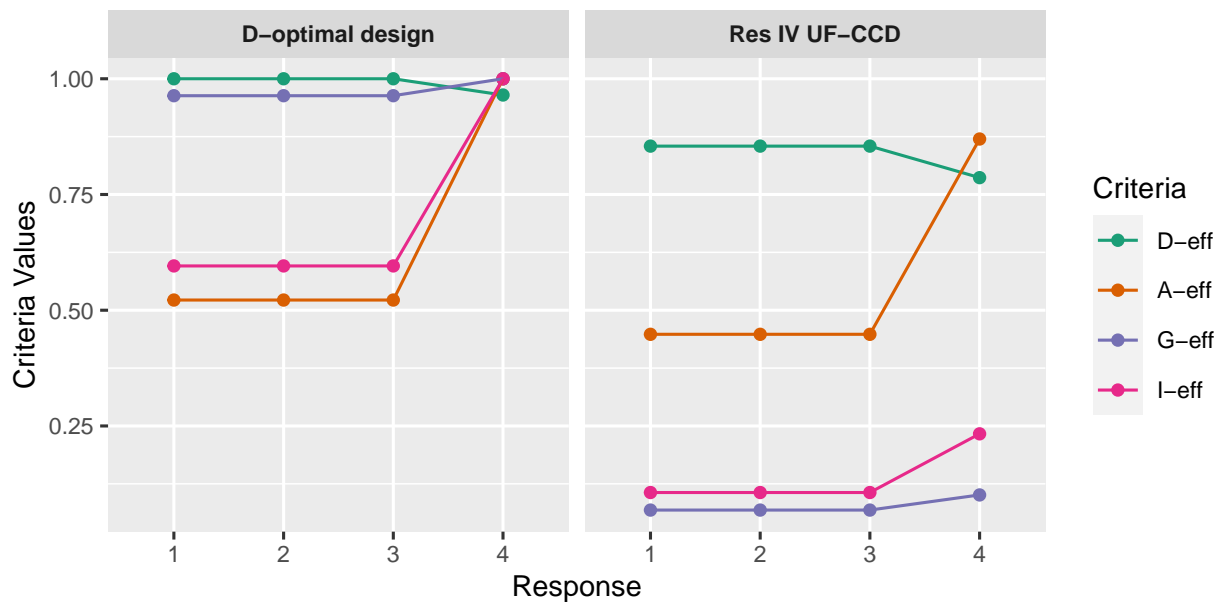
Figure 3.2: The plot showing the D-, A-, G- and I-efficiency values for the D-optimal design and the Res IV UF-CCD

The geometric structures of the D-optimal design and the Res IV UF-CCD are given in Appendix A.2. The D-optimal design structure shows that there is no center run selected for the D-optimal design and most of the runs are located at the edge of the design space. This is expected because in order to optimize D-efficiencies of multiple responses, more

design points are likely to be pushed to the edge rather than located around the center of the design space.

3.4.2 Ten Factor Experiment with Multiple Responses

In this section, we consider a larger scale experiment with 10 design factors and 4 responses. Table 3.2 provides the information about the relationship between the design factors and responses from the earlier screening experiment [70].

Table 3.2: Result of a screening experiment with ten factors

Response	Factor									
	1	2	3	4	5	6	7	8	9	10
1	X	X	X	X	X					
2			X			X	X	X		
3			X	X			X	X		
4				X				X	X	X

Similar to the previous example, the goal here is to identify optimum locations in the input space that provides good precision of estimation for all the responses. Here, we consider a 30-run design with 10 factors and five levels $(-2, -1, 0, 1, 2)$ for each factor. The choice of the factor level was to ensure an easy comparison with the resolution V unique factor CCD (Res V UF-CCD) of the same size from Marget and Morris [70] (see Appendix A.3 for the design structure). The design space has $5^{10} = 9765625$ candidate locations and the starting design with 30 runs for the PACE algorithm is randomly selected from the candidate locations. Unlike the previous example, the search for optimal designs did not result in a single design, instead a suite of competing non-dominated D-optimal designs is obtained. With $100,000$ random starts, the PACE algorithm identified a Pareto front consisting of eight competing designs. For the eight D-optimal designs and the Res V UF-CCD (denoted as design number 9), Figure 3.1 shows each response average normalized standard deviations for all the coefficients (all effects), first-order term coefficients (main effects), interaction term coefficients (interaction effects), and quadratic term

coefficients (quadratic effects). As with the five factors case, the result indicates that the D-optimal designs significantly have lower normalized standard deviations for each response model, implying that the D-optimal designs obtained provide consistently better efficiency for estimating all responses.
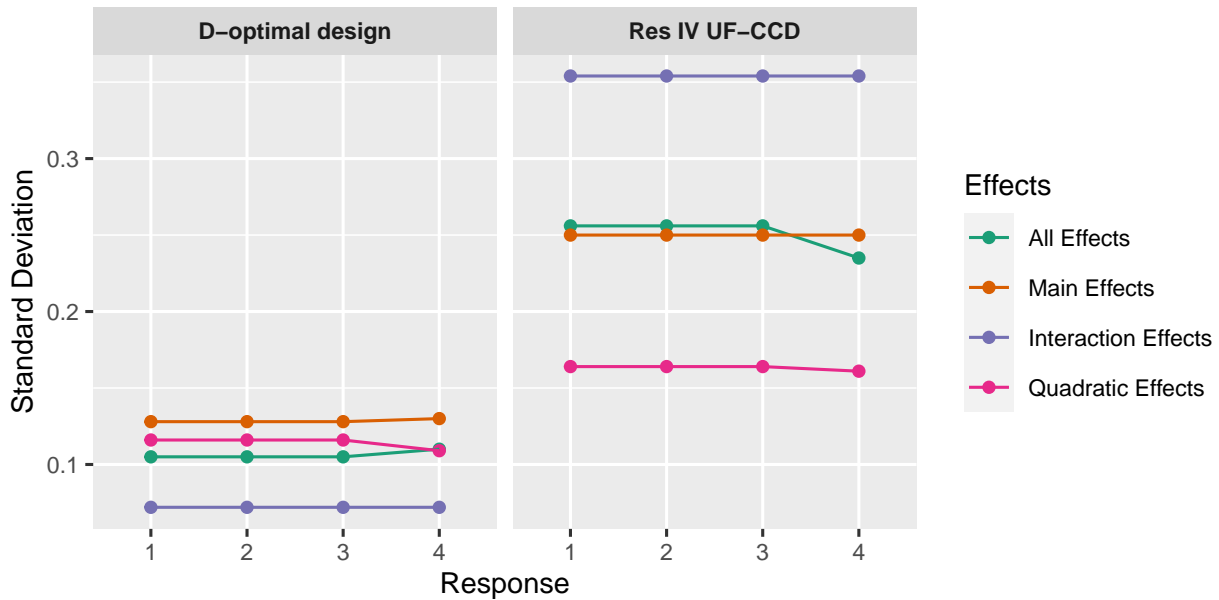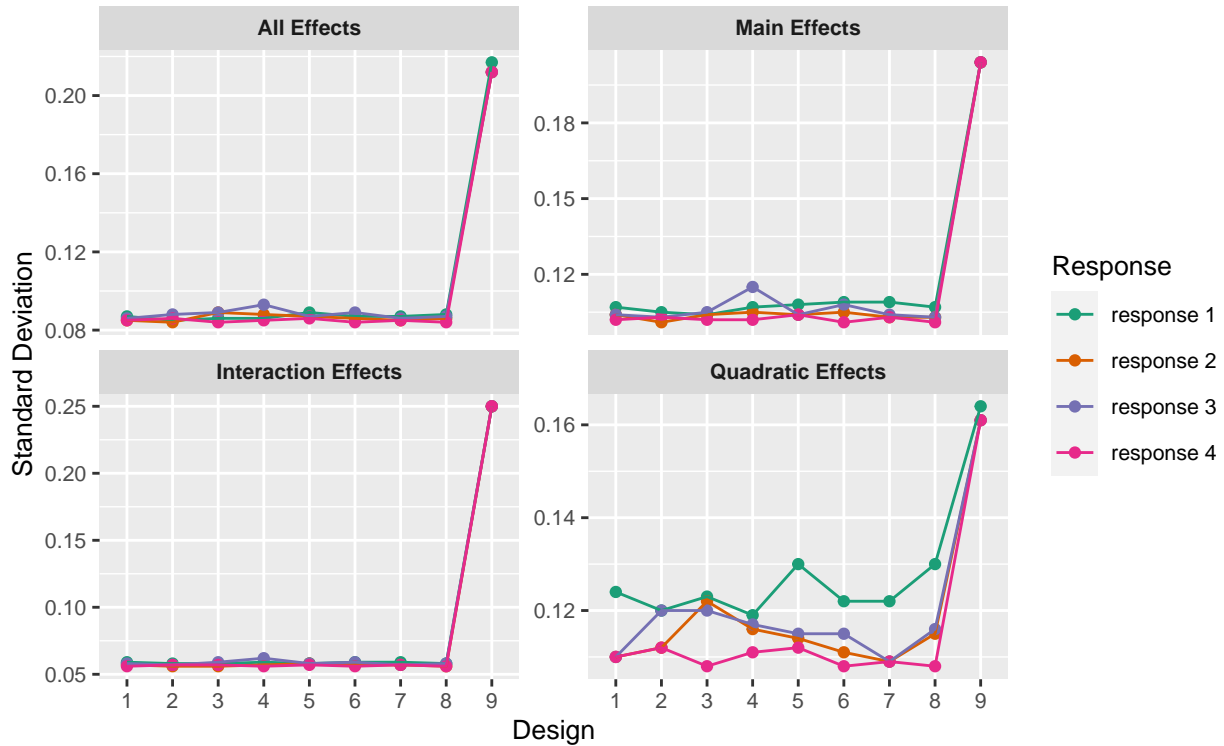

Figure 3.3: Average normalized standard deviations of model coefficients for the eight D-optimal designs and the Res V UF-CCD

Although the D-optimal designs clearly outperform the Res V UF-CCD based on the estimated standard deviations for model coefficients, yet, for informed decision making, it is helpful to look at other design criteria to further evaluate the design performance on other aspects. It is essential to examine trade-offs among criteria for the competing designs when balancing multiple criteria in design optimization.

Selection of D-optimal designs is dependent on the specified model forms. In addition to obtaining precise estimation of the specified model, another important aspect in design selection is to protect against potential model misspecification (what if the information

obtained from the screening experiment was not quite accurate). Model misspecification could result in bias of estimated model parameters and the error variance. Therefore, in addition to ensuring precise estimation of the chosen models, we also want to protect against the impact of potential misspecification on the estimation of coefficients and error variance if any of the omitted factor (based on the screening experiment) is potentially active. Let the specified model for a particular response be, $\mathbf{Y} = \mathbf{X}_s \boldsymbol{\beta}_s + \boldsymbol{\varepsilon}$, where $\mathbf{X}_s$ is an $N \times k_s$ design matrix of factors specified for that response (based on the screening experiment), while the full response surface model for all design factors is denoted by $\mathbf{Y} = \mathbf{X}_s \boldsymbol{\beta}_s + \mathbf{X}_o \boldsymbol{\beta}_o + \boldsymbol{\epsilon}$. In our application, the set of terms $\mathbf{X}_s \boldsymbol{\beta}_s$ contains the first order terms, two factor interactions and quadratic terms of factors specified for a particular response, while $\mathbf{X}_o \boldsymbol{\beta}_o$ contains all the omitted terms, including the first order terms, two factor interactions and quadratic terms of the factors not identified to be related to the response.

For each response model, the bias of the estimated model parameter estimates is given by

$$E(\hat{\boldsymbol{\beta}}_s) - \boldsymbol{\beta}_s = [\boldsymbol{\beta}_s + (\mathbf{X}_s' \mathbf{X}_s)^{-1} \mathbf{X}_s' \mathbf{X}_o \boldsymbol{\beta}_o] - \boldsymbol{\beta}_s = A \boldsymbol{\beta}_0, \tag{3.3}$$

where $\mathbf{A} = (\mathbf{X}_s' \mathbf{X}_s)^{-1} \mathbf{X}_s' \mathbf{X}_o$ is regarded as the alias matrix [81]. The sum of the squared transmitted bias (SSB) is given by, $SSB = \boldsymbol{\beta}_o' \mathbf{A}' \mathbf{A} \boldsymbol{\beta}_o$, where $\boldsymbol{\beta}_o$ is unknown and often assumed to follow the multivariate normal distribution, $\boldsymbol{\beta}_o \sim N(\mathbf{0}, \sigma_{\beta_o}^2 \mathbf{I})$ [24]. A measure of the impact of model misspecification on the estimated coefficients is given by $tr(\mathbf{A} \mathbf{A}')$ [9]. We define the bias on the estimated coefficients efficiency for a particular design as $tr(\mathbf{A}_{min} \mathbf{A}'_{min}) / tr(\mathbf{A} \mathbf{A}')$, where "min" designation denotes the D-optimal design with the minimum $tr(\mathbf{A} \mathbf{A}')$. Bias can also be transmitted to the estimate of σ^2 due to model misspecification since

$$E(MSE_{user}) - \sigma^2 = \boldsymbol{\beta}_o' [\mathbf{X}_s \mathbf{A} - \mathbf{X}_o]' [\mathbf{X}_s \mathbf{A} - \mathbf{X}_o] \boldsymbol{\beta}_o / k_s = \boldsymbol{\beta}_o' \mathbf{R}' \mathbf{R} \boldsymbol{\beta}_o / k_s, \tag{3.4}$$

where MSE_{user} is the residual mean squared error from the misspecified model and $\mathbf{R} = \mathbf{X}_s\mathbf{A} - \mathbf{X}_o$. The impact of model misspecification on the estimated error variance is measured by $tr(\mathbf{R}'\mathbf{R})$ [81]. We define the bias on the estimated error variance minimization efficiency for a particular design as $tr(\mathbf{R}'_{min}\mathbf{R}_{min})/tr(\mathbf{R}'\mathbf{R})$, where "min" designation denotes the D-optimal design with the minimum $tr(\mathbf{R}'\mathbf{R})$.

For the competing designs on the Pareto front, the D-, $tr(\mathbf{A}\mathbf{A}')$- and $tr(\mathbf{R}'\mathbf{R})$-efficiency values were estimated for each response model. As expected, the result in Figure 3.4 shows that, the eight optimal designs found on the Pareto front have high D-efficiency values, thus satisfying our goal of precise parameter estimation. From the result, it appears that design 6 has the best performance to protect against model misspecification for response 1 while design 5 performs generally well for other responses. Overall, design 5 appears superior in offering protection against model misspecification.
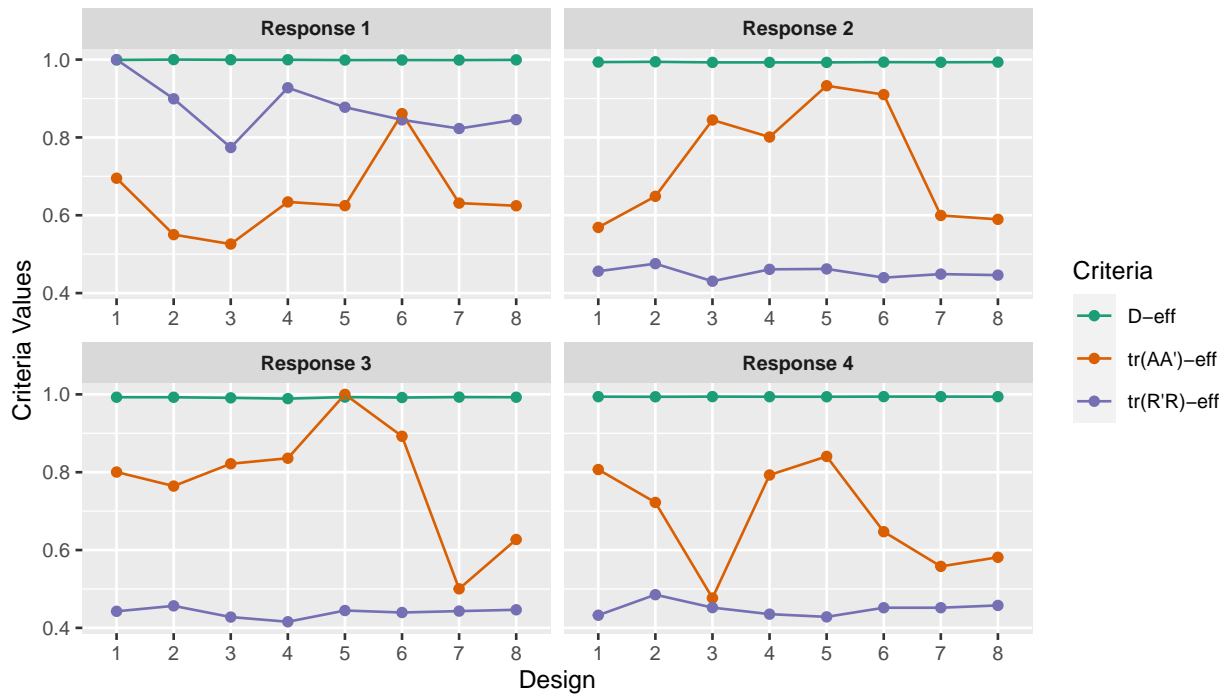
Figure 3.4: The plot showing the D-, $tr(AA')$- and $tr(R'R)$-efficiency values for the eight D-optimal designs

In addition to the precision and bias of the estimated models, we also care about the prediction ability of the fitted models. Therefore, we examine the A-, G- and I-efficiency values of the competing designs. From Figure 3.5, we can deduce which design(s) provide the best prediction ability for each response. For example, design 5 and design 7 seems to provide the best prediction of the worst case scenario for response 1, while design 4 seems to be the best for response 3. Comparing the responses in general, design 5 appears to offer better prediction performance. For this example, we also explore an alternative case with different design size. The general patterns of selected designs are consistent with the 30-runs case.
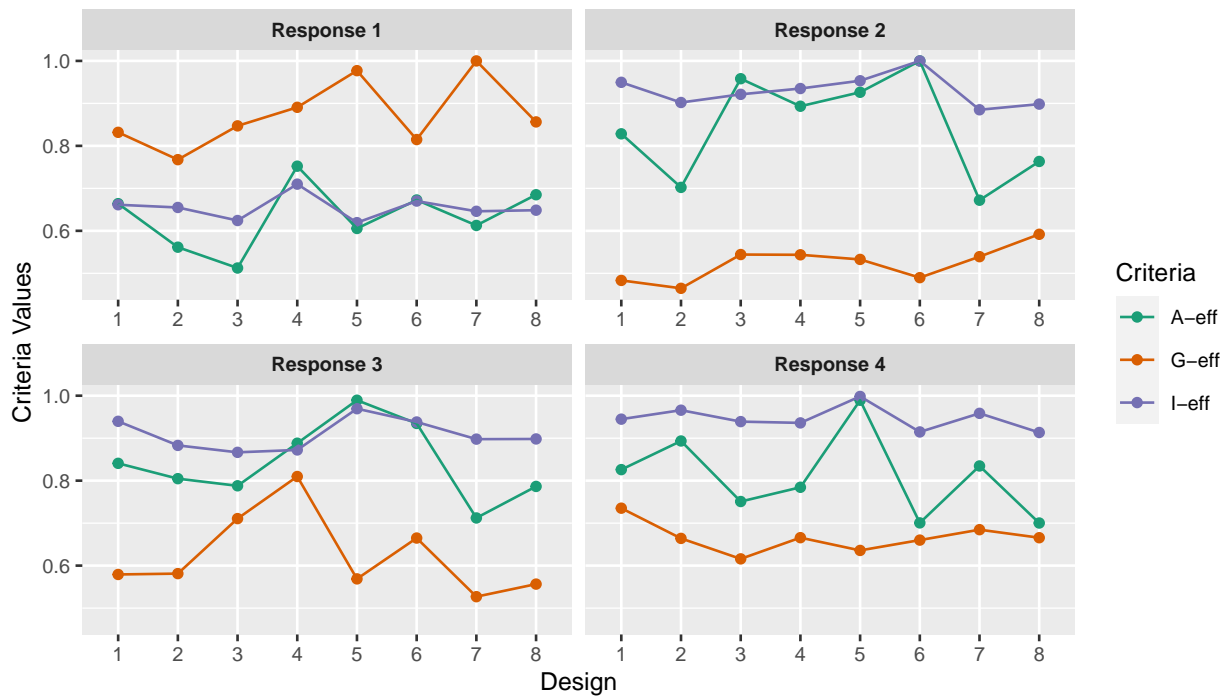
Figure 3.5: The plot showing the A-eff, G-eff and I-eff for the eight D-optimal designs

In terms of the computing efficiency, the PACE algorithm takes an average of 48 seconds to run 100 random starts on a standard desktop computer for the five factors 20-run design and about five minutes for the ten factors 30-run design. Multiple random starts were run in parallel to improve the computational efficiency of the algorithm. For the five factors example, the Pareto front was identified after running 20000 random starts.

Further increasing the number of random starts up to 40000 did not result in any change in the identified Pareto front. For the ten factors example, the number of points found on the Pareto front initially increases as the number of random starts increases and then it begins to stabilize after 20000 random starts upward.

3.5   Conclusion

For experiments with multiple responses, the central composite designs are the commonly used response surface designs, which typically require a large number of runs across the super space of all design factors. When an earlier screening experiment or subject matter expert knowledge can help identify active factors for each response, leveraging this information can significantly reduce the run size while still achieving good efficiency in estimation.

This work describes how the prior knowledge can be utilized to seek optimal designs for simultaneously maximizing precision of the estimated model parameters based on utilizing the Pareto front. Particularly, the Pareto front optimization approach is employed to generate the Pareto set of non-dominated designs based on D-efficiencies of all the estimated responses. A search algorithm based on coordinate exchange was developed to efficiently search through the design space for populating the Pareto front of optimal solutions.

Examples of varied dimensions were used to illustrate the scalability of the proposed methodology and the implementation of the algorithm. The D-optimal designs generated via the methodology in this work show improve performance on precision of estimated model parameters when compared to designs from existing methods, such as the recently developed UF-CCDs. We also illustrate how the evaluation of secondary criteria could help gain a better understanding of design performance on multiple aspects and hence aid for an informed design selection.

Chapter 4: Multiple Objective Latin Hypercube Designs for Computer Experiments

4.1 Introduction

Computer experiments are experimental techniques that use computer simulations to emulate the relationship between inputs and outputs rather than using a physical experiment. Computer experiments have been frequently used in scientific and engineering research to substantially reduce cost of physical experimentation and or for situations when physical experiments can be prohibitive. Some applications however, involve complex computer simulations and exploration of large input spaces such that computer experiments can still be computationally expensive and time consuming. Fortunately, growth in computer power has brought about methodological developments for computer simulations to be successfully applied to a variety of problems [108]. When computer simulations are complex, it has become a common practice in computer experiments to provide a computable surrogate models (also known as meta-models) to replace the expensive computer simulations [5]. Surrogate models are trained models designed to learn the function mapping between inputs and outputs. Among the family of surrogate models, Gaussian process regression has been popularly used [94] in the field of computer experiments due to its flexibility in approximating expensive complex surfaces [1]. It enables tasks like sensitivity analysis, uncertainty quantification, validation, and calibration feasible [31, 96].

One major characteristic that differentiate computer experiments from physical experiments is that the process of a computer experiment is deterministic. That is, the computer simulator produces the same output if run multiple times with the same set of inputs. For

physical experiment, the process is stochastic. This is because the output of a physical experiment can be influenced by many uncontrollable factors like, the natural variations in a process, errors due to measurements and unknown noise factors. Due to the deterministic nature of computer experiments, the principles of randomization, replication and blocking that are relevant in physical experiments are irrelevant in computer simulator experiments [78].

A major task in statistical modeling of computer simulations is proper planning and configuration of the inputs. In practice, resources are often limited, so an experimental design plays a key role in identifying the set of inputs to be used in computing the output of a computer experiment to achieve specific goals. Such a set of points is called the design of the computer experiment. There has been extensive work in the literature for constructing designs for computer experiments. They can be divided into two main broad categories viz; space-filling designs and model-based optimal designs. The space-filling designs are only concern with the geometric locations of the design points. These designs aim to achieve maximum spread or coverage of the design space. On the contrary, the model-based designs rely on a pre-specified model to construct designs for achieving certain optimal performance.

Space-filling designs have been quite popular in computer experiments due to the fact that they do not rely on any model assumptions and hence are very flexible and easy to adapt for broad applications. For example, space-filling designs are used in the computer simulator of a knee implant to minimize the maximum strain that occurs at the bone and prosthesis boundary of a knee implant [96]. Another example is in the curing process of a material. It is important that the temperatures and speeds of the conveyor belt moving through the oven are set properly in the curing process of a material in order to produce viable material [66]. For instance, a very high temperature and a very low speed would damage the material while a very low temperature with a very fast speed would result in a material not being fully cured. Space-filling designs can be used in the computer

simulator experiment of this process to determine the right input locations to consider when performing the actual experiment in order to achieve the aim of the experiment and avoid wastage of experimental resources. Different aspect of space-filling characteristics has resulted in different space-filling designs. This chapter will focus on enhancing the existing space-filling designs by creating well-rounded space-filling designs with good performance on multiple characteristics.

Latin hypercube designs (LHDs) were among the earliest introduced space-filling designs [73]. Such designs have the nice property of having uniformly distributed design locations when projected in each single input dimension. A variety of different distance-based criteria have been developed to capture different space-filling characteristics of the design across the input space such as the maximin and minimax distance criteria [48]. To combine the nice space-filling characteristics in individual and full input space, LHDs that are optimal based on a single distance-based criterion [79, 106] have been broadly used as the common space-filling designs for computer experiments.

In this chapter, we propose to further improve the performance of the LHDs by considering multiple distance-based criteria to create well-rounded space-filling designs with good performance on multiple characteristics. To search through the design space for optimal designs, a column-wise exchange simulated annealing algorithm and a non-dominated sorting genetic algorithm are explored, with the later demonstrating much better computing efficiency. The rest of the chapter is organized as follows. In Section 4.2, we review the literature on distance-based criteria and space-filling designs. Section 4.3 provides a brief review on the Pareto front approach for multi-objective optimization and presents a discussion on the use of the Utopia point method for selecting a smaller set of designs when the Pareto frontier is a large set. Section 4.4 describes the newly developed search algorithms. In Section 4.5, we illustrate the proposed methodology using examples of varied dimensions. The summary and conclusion are provided in Section 4.6.

4.2 Space-filling Designs and Distance-Based Criteria

Space-filling designs are very popular for computer experiments. They generally offer a good spread or coverage across the design space and great flexibility in capturing behaviors of the responses in different areas of the design region. Depending on the goal of the experiment, the term "space-filling" could be used to describe different principles for "filling" the design space. Spacing-filling could mean to spread out the design points by avoiding any pair of design points to be too close to each other, with the assumption that the information provided by two close points will be similar. Thus, points are placed in the experimental region such that they are as far apart as possible. Johnson et al. [48] proposed the maximin distance criterion (*Mm*), which seeks to maximize the minimum distance between any pair of design points. The minimum distance among the points in the design is given by $min_{ij}d(x_i, x_j)$. Therefore, the maximin distance criterion is given by

$$\max_{D} \min_{i,j} \; d(x_i, x_j). \tag{4.1}$$

Alternatively, space-filling could also imply, to provide the best coverage of the design space. That is, for any given point in the design space, there should be a design point close by. The assumption here is that any region of the design space could be of interest in understanding the underlying relationship of the process. Let p represent the number of design factors and χ be the experimental region, which in most cases, can be scaled to a unit hypercube, $\chi = [0, 1]^p$. Let $D' = \{x_1, ..., x_n\}$ denote a design of n runs, where each point $x_i \in [0, 1]^p$. Then for any point $x \in [0, 1]^p$, the distance to the nearest design point is given by $min_{i=1}^{n}d(x, x_i)$. Intuitively, if the goal of the experiment is to predict well over the entire input space, then the closer a point is to the nearest design point, the better precision its prediction will have. Hence, the worst prediction is likely to occur at the point with the farthest distance from the nearest design point. The worst nearest distance is given by $max_{x \in [0,1]^p} min_{i=1}^{n}d(x, x_i)$. Johnson et al. [48] proposed the minimax distance

48

criterion (*mM*). This criterion minimizes the maximum distance of any given point in the design space to the nearest design point. The minimax distance criterion is expressed as,

$$\min_{D} \max_{x \in \chi} \min_{i \in \{1, \cdots, n\}} d(x, x_i). \tag{4.2}$$

Computation wise, a maximin distance design is much easier to construct than a minimax distance design. This is simply because the maximin distance design only requires evaluating the distances among the design points while the minimax distance design requires to evaluate the distances from the design points to a fine grid of points spanning the design region.

Both the maximin distance designs and the minimax distance designs ensure good space-filling in the full dimension of all input factors. However, space-filling property is not guaranteed when the designs are projected onto the univariate space or subspaces formed by a subset of the design factors [52]. Based on the factor sparsity property [78], i.e. the response is likely to depend on only a few factors, then designs with good space-filling property when projected onto the subspace of active design factors is desired.

The Latin hypercube design (LHD) [73] was developed to ensure a uniform marginal distribution for each design factor. To construct a LHD of *n* runs, the range of each factor is divided into *n* equally spaced intervals. Then, only one coordinate of a design point is sampled from each of the non-overlapping interval (see Figure 4.1). Although the LHD ensures good space-filling for projections in the univariate space, a random LHD is not guaranteed to have a good space-filling for projections in full dimension or any other subspace of the design factors. For example, the right image of Figure 4.1 is a LHD that defects a good coverage of the design space. All the design points are located on a straight line (diagonal line), with the rest of the space unexplored.
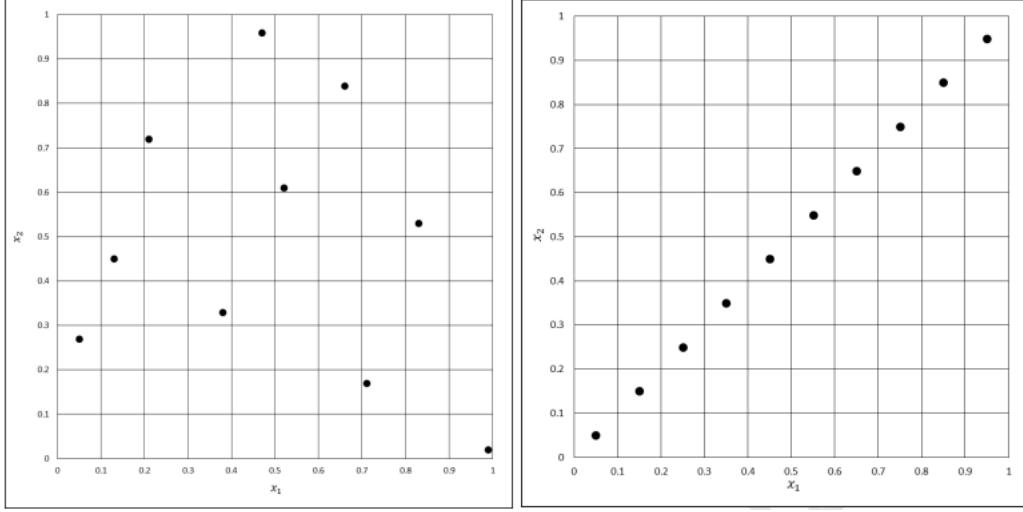
Figure 4.1: Examples of Latin hypercube designs with two factors and ten points

To improve the space-filling characteristic of the LHDs over the full design space, distance based criteria have been used to optimize the placement of design points over the input space. Morris and Mitchell [79] proposed constructing the LHD, such that the minimum distance among the points are maximized. This design is called the maximin Latin hypercube design (MmLHD). Their proposed maximin criterion is given by

$$\min_{D} \left\{ \Sigma_{i=1}^{n-1} \Sigma_{j=i+1}^{n} \frac{1}{d^k(x_i, x_j)} \right\}^{\frac{1}{k}}, \qquad (4.3)$$

as $k \to \infty$, this criterion becomes the maximin distance criterion in Equation 4.1. They suggested using the smallest value of k that result in a maximin distance design. Such a design is incline to have fewer pairs of points with minimum distance, which is referred to as the index of the maximin distance design. Similarly, Edwin and Dam [106] investigated the minimax criterion for LHDs (mMLHD). They suggested that design points for LHD should be chosen such that the maximal distance of any point in the input space to the design is minimal.

For MmLHD and mMLHD, the Latin hypercube arrangement ensures good spread for projections in one dimension, where as the maximin distance criterion and the minimax

distance criterion respectively ensure good spread or coverage in the full dimension. In particularly, design points are spread as far as possible in MmLHD, therefore pushing more points towards the corners and edges of the design space. In contrast, to ensure mamximum coverage, mMLHD rarely places any point in the corner or edge but instead put more points in the interior region. Even though MmLHD and mMLHD are single-criterion optimized LHD, the projection of the points onto any subspace of more than one factor may not have good space-filling property.

The maximum projection criterion (MaxPro) was proposed by Joseph et al. [53] to ensure good space-filling properties for projections in all possible subsets of the design factors. That is, having projections as maximin distance designs, across all the subspaces of 2 to $p - 1$ design factors. When a design is projected onto a subspace, the distances between the points are calculated with respect to the factors that define the subspace. The weighted Euclidean distance between the points x_i and x_j is define as

$$d(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta}) = \left\{ \Sigma_{i=1}^{p} \theta_l |x_{il} - x_{jl}|^2 \right\}^{\frac{1}{2}},$$

where $\theta_l = 1$ for the factors defining the subspace and $\theta_l = 0$ for the remaining factors. The weights are used to reflect the importance of the factors. Let $0 \leq \theta_l \leq 1$ be the weight assigned to factor l and $\Sigma_{l=1}^{p} \theta_l = 1$. To ensure good projections in all possible subspaces, an appropriate distribution is assigned to $\boldsymbol{\theta}$ and the reciprocal distance criterion in Equation 4.3 is integrated over that distribution, the criterion is given as

$$\min_D \int \Sigma_{i=1}^{n-1} \Sigma_{j=i+1}^{n} \frac{1}{d^k(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})} p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{4.4}$$

where $p(\boldsymbol{\theta})$ is a prior distribution for $\boldsymbol{\theta}$. If a uniform prior distribution is assumed for $\boldsymbol{\theta}$ and $k = 2p$, the criterion simplifies to

$$\min_D \Sigma_{i=1}^{n-1} \Sigma_{j=i+1}^{n} \frac{1}{\prod_{l=1}^{p} |x_{il} - x_{jl}|^2}. \tag{4.5}$$

The product in Equation 4.5 ensures that no two coordinates can be similar (enforcing the Latin hypercube arrangement), otherwise, the objective function can become ∞. The MaxPro criterion has also been used to improved the uniformity of a random LHD. The resulting design is called maximum projection LHD (MaxProLHD) [52].

The variations of LHDs discussed above, MmLHD, mMLHD, and MaxProLHD, are indeed substantial improvements to the random LHDs and have been used extensively for computer experiments. However, these designs can only ensure the optimal performance on a single characteristic, which might lead to suboptimal performance on other characteristics of interest.

4.3   Pareto Front Optimization and Utopia Point Method

The Pareto front approach has been extensively discussed in Chapter 3. The Pareto optimal set can sometimes have many solutions. When this is the case, it is necessary to move from a large set of candidate solutions to a reduced set of solutions. The Utopia point method is commonly used for strategically selecting a smaller set from the Pareto optimal set with large solutions [67].

The standard Utopia point method uses a norm and a defined set of weights to describe each solution's closeness to the *Utopia point*. Based on the defined set of weights for combining the criteria, the solution closest to the Utopia point is then chosen as the 'best' solution. A Utopia point, F^0, in the criterion space satisfies:

$$\mathbf{F}^0 = \max_{\xi}\{f_i(\xi)|\xi \in \mathbf{\Omega})\} \quad \text{for all } i \in \{1, 2, ..., C\}. \tag{4.6}$$

A solution $\xi^* \in \mathbf{\Omega}$ that satisfies $\mathbf{F}(\xi^*) = \mathbf{F}^0$ is regarded as the Utopia solution. For multiple criteria optimization however, the Utopia solution ξ^* generally does not exist, but it can be used as an "ideal" standard for the criteria values in order to identify the most performing solution(s) from the Pareto optimal set. Figure 4.2 is an illustration of the

Pareto front and the Utopia point for a maximization problem involving two criterions. In the plot, coordinates of the Utopia point are the maximum values for both criteria. The Pareto front is on the borderline of the criterion space that is nearest to the Utopia point. Clearly from the Figure, no points on the Pareto front can improve one criterion without deteriorating the other. The implication of this is that the Utopia point is not feasible in practice. However, based on a chosen metric, points on the Pareto front which minimize the distance to the Utopia point are selected by the Utopia point method.
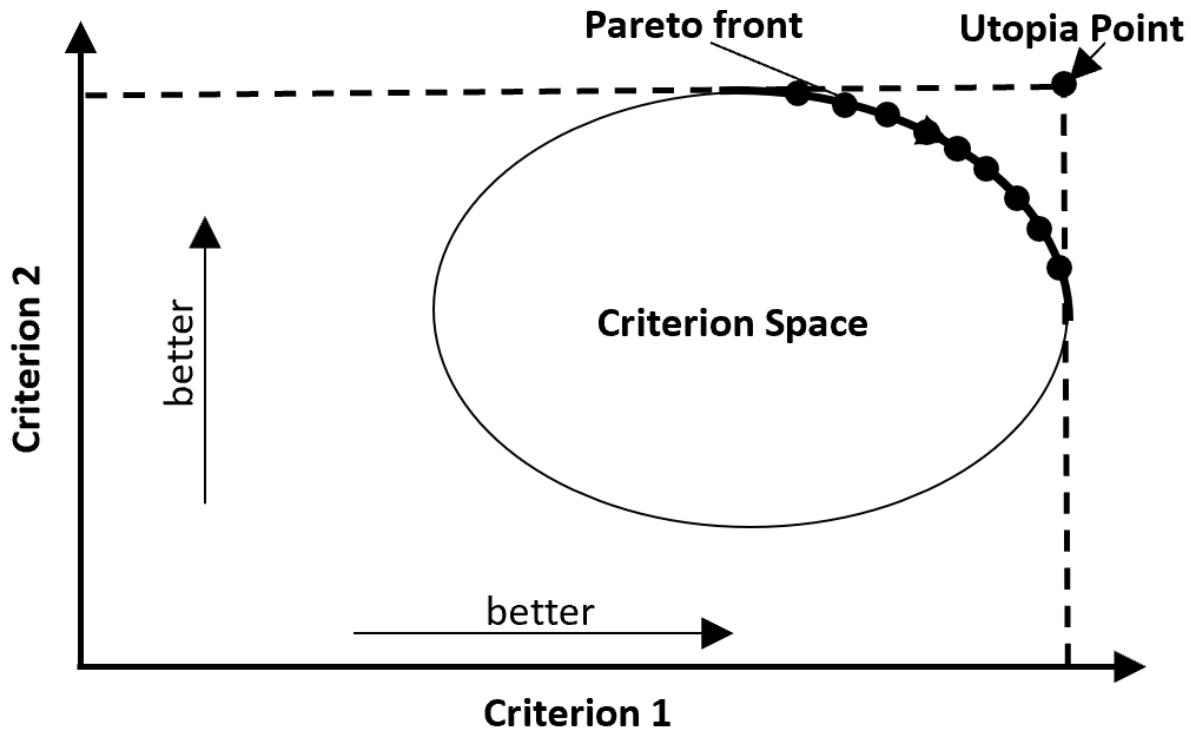
Figure 4.2: Pareto front and the Utopia point in a maximization problem involving two criterions

The common method for calculating the distances to the Utopia point is the L_1-norm, formulated as:

$$\min_{\xi \in \Omega^*} \Sigma_{j=1}^{C} w_j \left| s_j(\xi) - s_j^0 \right|, \tag{4.7}$$

where Ω^* is the set of Pareto solutions, w_j for $j \in \{1, 2, ..., C\}$ are the weights assigned to the C individual criteria, $s_j(\xi)$ is the *jth* objective function transformed to a desirability

scale (i.e 0-1 scale) for the *jth* criterion corresponding to solution ξ , and s_j^0 (typically set at 1) denotes the Utopia point value for the *jth* criterion on the same scale. Every weight combination identifies one optimal solution from the Pareto front. However, based on the scaling scheme and the method used in calculating the distances of the solutions to the Utopia point, not all the solutions on the Pareto front will be optimal for at least one weight combination. The Utopia point approach selects a subset of the solutions in the Pareto front, thereby reducing the potential solutions to consider for decision making.

Experimenting with different weight choices takes minimal computational effort in the Utopia point approach since every set of weights utilize the same Pareto front found by the search algorithm. In this study, we explore a fine grid of weight combinations to examine the trade-offs between the different criteria for designs in the Pareto optimal set. Decisions can then be recommended based on the trade-offs and robustness of candidate solutions to different subjective choices. Compared to the standard desirability function approach, the Pareto front approach and the Utopia point method allows more intuition about the relative performance of different solutions and quantitative information for informed decision making [67].

4.4 Search Algorithms

A LHD is an $n \times p$ matrix such that each column is a random permutation of $(1, ..., n)$ which can be mapped onto the range of the *p* factors. This random permutation process in the construction of a LHD allows possibility of producing many LHDs from a given design space or candidate set, each satisfying the Latin hypercube condition of only one point in each level. However, as mentioned earlier, this random procedure poses the possibility of generating a LHD with poor space-filling quality. This leads to the idea of optimizing a design criterion in the construction of LHDs in order to generate optimal LHDs. Several search algorithms like, the simulated annealing [79], the columnwise-pairwise algorithms [54], the stochastic evolutionary algorithms [47, 82] have been pro-

posed for generating optimal LHDs. In this work, first we employed the column-wise exchange simulated annealing algorithm which works sufficiently for low dimensional input space but does not scale well for high dimensional problems. Then we adapted the non-dominated sorting genetic algorithm for constructing LHDs and demonstrate substantial improvement in populating the Pareto front for higher dimensional problems.

4.4.1 Column-wise Exchange Simulated Annealing Algorithm

The column-wise exchange simulated annealing algorithm has been commonly used in design of computer experiments due to its natural ability to preserve the Latin hypercube structure [79]. In the mechanism of this algorithm, the search begins with an initially constructed LHD, and proceeds through the examination of a sequence of designs, each generated as a "modification" of the preceding one. By modification, it means a new design is created from the current design by the process of interchanging two randomly chosen observations within a randomly selected column in the current design. The new design replaces the current design if it leads to an improvement. Otherwise, it will replace the current design via the simulated annealing [57] approach-a method commonly for approximating the global optimum of any given function. That is, the new design is allowed to replace the existing design (even if it does not lead to an improvement) with gradually decreasing probability to ensure extensive search over the solution space.

For the multiple criteria optimization, given that the number of criteria to simultaneously optimize is k, we define the set **w**, with each element as a weight vector $(w_1, ..., w_k)$, where $\Sigma_{i=1}^{k} w_i = 1$ and $w_i \in [0, 1]$. Now, let $C = \Sigma_{i=1}^{k} w_i c_i$ be the weighted combination of the criteria values for a particular design, referred to as the weighted criteria, where c_i is the scaled value for each criterion [44]. A different set of weight values will be used to lead the search in different direction for locating points on different regions of the Pareto front. An overview of the column-wise exchange simulated annealing algorithm for seeking the Pareto front based on multiple criteria is described by the following steps:

1. Construct a $n \times p$ LHD matrix as the starting design. For this current design, ξ, evaluate the weighted criteria, denoted as C_c.

2. Initialize two null sets: the Pareto optimal set (i.e the set that will contain the Pareto designs), denoted as P, and the set that will contain the weighted criteria of the Pareto designs (this forms the Pareto front), denoted as PF. ξ is added to P and C_c is added to PF.

3. A new design, ξ^* is generated by a column-wise exchange procedure on the current design. That is, two randomly selected observations of a randomly selected column of the current design are interchanged. The weighted criteria of the new design, C_n, is evaluated.

4. Next, the difference between the weighted criteria value for the current design and that of the new design, that is, $\Delta C = (C_n - C_c)$ is evaluated. Since the goal is to minimize C, if the value of ΔC is negative, the new design replaces the current design and search continues. Else, the new design will be accepted with a probability of $P = exp(-\Delta C / t)$, where t is a time-varying parameter tuned to gradually approach zero as the number of iterations increases.

5. The next step is to update the current Pareto front and Pareto optimal set by checking for improvements. The comparison made here is between the new design, ξ^* and the ones in the "current" set of non-dominated designs. If ξ^* dominates at least one of the designs in the "current" set, then ξ^* is added to P and the designs dominated by ξ^* are removed. If ξ^* neither dominates nor is dominated by any design in the current set, then just add ξ^* to P. If at least one of the designs in the "current" set dominate ξ^*, then no update is needed for the current Pareto set.

6. Repeat Steps 3 to 5 for a pre-specified set of weights of interest and combine the Pareto optimal sets and the Pareto fronts from multiple searches based on different

weights. At the end, there would be d non-dominated designs in the combined Pareto optimal set P (i.e P=$\{\xi_1, \xi_2, ..., \xi_d\}$) and d associated weighted criteria in the combined PF (i.e PF=$\{F(\xi_1), F(\xi_2), ..., F(\xi_d)\}$).

7. Repeat steps 1 to 6 with S different random starts. The final set of Pareto optimal designs and the Pareto front are obtained by combining the results from S separate searches.

4.4.2   Non-dominated Sorting Genetic Algorithm

Genetic algorithms have been commonly used in optimization problems [77]. They rely on operations of selection, crossover, and mutation, inspired by evolution theory. The Genetic algorithm is an iterative process, and the population in each iteration (otherwise known as generation) are potential solutions evolving towards better solutions for the optimization problem. To adapt for our application, the initial population should be generated as LHDs. Moreover, the genetic operators of crossover and mutation should be defined such that they operate directly on LHD matrices, thus preserving the LHD properties. In our analysis, solutions for the GA are coded as LHDs and the initial population is generated randomly, thereby allowing sampling from the range of all possible LHDs in the design space.

At each generation, new solutions (often regarded as "children") are created via the operations of crossover and mutation. The fitness of each individual at each generation impacts its likelihood of being selected to create new solutions for the next generation. Fitness is defined as the value of an objective function defined for the optimization problem and it is use to quantify the suitability of each solution. The more suitable solutions have higher chance of surviving to the next generation. Several selection methods have been proposed for rating the fitness of each solution and for making selection of the best solutions. Liefrendahl and Stocki [61] and De Jong [21] proposed ranking the solutions in each generation from best to worst fitness and then selecting half of the best solutions

as survivors at each generation. To ensure that each generation is at least as fit as the previous generation, Eiben and Smith [27] proposed an elitist strategy. That is, offspring solutions and parent solutions compete together for survival into the next generation.

The operation of crossover often involve multiple parents while the mutation operation often consider one parent. In this work, crossover is used to keep the good attributes of two existing solutions, while mutation is used to generate a slight modification of an existing solution in order to allow diversity in the population [27]. We employed the rank-proportional selection [39, 69] to select parents for the crossover and mutation operations. The probability of selecting the rth ranked candidate of the N_{pop} solutions is $\frac{(N_{pop}-r+1)}{(N_{pop}(N_{pop}+1)/2)}$. Once probabilities have been assigned to each solution in the parent set, a crossover operation is used to generate offspring solutions as follows; two parents are randomly selected from the "parents set", then a random integer k_i ($1 \leq k_i \leq n-1$, $i = 1, 2, ..., p$) is selected as the cut-off point for each variable. To generate each *ith* variable of the "child", values in runs 1 to k_i from the *ith* variable of the "first parent" are selected and values in runs $k_i + 1$ to n from the *ith* variable of the "second parent" are selected. A "repair" function is use to replace repeated values in each variable with the missing values in a random order to ensure each column is a permutation of $\{1, \cdots, n\}$ [44].

For instance, consider a 10-run design, if the cut-off point for generating a particular variable of the child solution is 6, the runs of that variable is determine as shown in Table 4.1. The child variable takes its first six runs from parent 1 variable and the remaining four runs from parent 2 variable. As observed in Table 4.1, value 9 is repeated in the child variable. This can then be corrected by a repair function, where the duplicate of the repeated value is replaced with the unselected value from the parents variables (i.e 6).

Table 4.1: An example of crossover operation

Parent1 variable	1	9	8	7	4	3	5	10	2	6
Parent2 variable	6	4	7	1	3	8	10	2	5	9
Child variable	1	9	8	7	4	3	10	2	5	9
Repaired Child variable	1	9	8	7	4	3	10	2	5	6

For the mutation operation, one parent design is randomly sampled from the "parents set" and switching is done on two randomly sampled observations for each variable. In this work, at each generation, half of the children's solutions are created via crossover and the remaining half via mutation.

For the multiple criteria optimizations, we employed the non-dominated sorting genetic algorithm II (NSGA-II) [22]. It is a fast-elitist non-dominated sorting GA that uses a rank-based fitness function for generating the Pareto front. It generates offsprings using the specified crossover and mutation operations, and then selects the next generation according to non-dominated sorting and crowding distance comparison. The execution of the algorithm is as follows; for each solution S in the current population, the number of solutions that dominate S (regarded as domination count) is calculated and solutions dominated by S are identified. Solutions with domination count of 0 are non-dominated, thus they are on the first tier Pareto front (PF1). After determining the domination count of each solution and identifying the solutions that each solution dominates, the algorithm then loops through the solutions on PF1. For each solution S on PF1, each solution dominated by S has its domination count reduced by one. The dominated solutions with new domination count of 0 will be on the second tier PF (PF2). This process is repeated for all solutions on the lower tier PFs.

After sorting the solutions in the current population into tiers of PFs, the solutions on each tier of PF are then sorted based on their crowding distance [22]. The crowding distance is a measure of a solution's criteria values from those of other solutions. Solutions that are farther from others are given a larger crowding distance and a higher rank. The values of each criterion are sorted from smallest to largest. The highest and lowest values of a criterion are assigned an infinite crowding distance. For all other values of a criterion, the crowding distance of each value is evaluated by finding the range of values immediately on both sides of the value, which is then scaled based on the range of the criterion. After separately computing the crowding distances for the values of each crite-

rion, the overall crowding distance for a solution is evaluated as the sum of that solution's crowding distances for all criteria.

When the NSGA-II is initialized using randomly generated solutions, a possibility is that few of the solutions would be on the PF1. The consequence of this is that the lower tier PFs solutions would potentially be used to produce offspring solutions for the next generation if fixed population size is used. Thus, resulting in offsprings that are likely to be inferior and computation time would be wasted in evaluating these inferior solutions. On the other hand, for later generations, as the PF continue to grow, its size could potentially exceed the initial population size. Hence, using fixed population size could result in elimination of superior solutions on the Pareto front. To prevent these challenges, Chapman et al. [15] proposed using adaptive population sizing [29] for the NSGA-II algorithm. This procedure allows the number of offspring solutions produced at each generation to be dynamically determined. In this work, our modification on the NSGA-II follows the modification proposed in Chapman et al. [15].

To incorporate adaptive population sizing in NSGA-II, a maximum population size (*maxpro*) and a maximum number of offspring (*maxoff*) to be generated is specified. At the initialization of the algorithm, *popsize* = *maxpop* solutions would be randomly generated. The algorithm is then used to determine solutions that would be on the PF1. The number of offspring ($noff_g$) solutions to be created in generation g is dynamically determined as

$$noff_g = min\{c_g + [d_g \times |(PF_1)_g|, maxoff]\},$$

where c_g stands for the minimum number of offspring solutions to be created in generation g, while d_g stands for the rate at which the number of offspring created in generation g grows as the number of solutions on PF1 grows. For the implementation in this work, we set $c_g = 20$ and $d_g = 4$. That is, at each generation, a minimum of 20 offspring solutions are generated, and for every solution on the PF1, four extra offspring solutions are generated. At the end of generation g, there would be $PF1_g$ solutions on the PF1. Then

the population size for generation $g + 1$ is determined as

$$popsize_{g+1} = min\{a_g + [b_g \times |(PF_1)_g|, maxpop]\},$$

where, a_g stands for the minimum number of solutions to include in the population be-yond the PF1 while b_g stands for how the size of the population grows in relation to the size of the PF1. We define a_g to be relative to the specified maximum population size ($a_g = 0.2 * maxpop$) and $b_g = 1$. This allows the population size to be slightly larger than the number of PF1 solutions at each generation. The highest ranked solutions of $popsize_{g+1}$ are the ones that move on to generation $g + 1$. However, there is a disadvantage to specifying a maximum population size. If $|(PF_1)_g| > maxpop$, it implies that some solutions on PF1 fails to move on to the next generation, that is, the PF will be truncated. This could result in a situation where reasonable contending solutions are not available in the decision-making stage. To avoid this situation, if $|(PF_1)_g| > maxpop$, then $maxpop$ should be allowed to grow based on the size of the PF1 [15].

In genetic algorithms search procedure, the cycle of fitness evaluation, reproduction, and survival continues until the specified termination condition is met. In this work, the termination condition is when the specified number of generations is reached. However, it should be noted that, being an heuristic search algorithm, the genetic algorithm does not guarantee finding the global optimal solutions. In general, as the number of iterations increases, the better the approximation to the true Pareto front. In practice, a balance is sort between the accuracy of the approximation and the computation time.

4.5 Applications

Using the Pareto front optimization approach, the column-wise exchange simulated annealing algorithm and the NSGA-II were used to seek optimal LHDs that simultane-ously balance multiple space-filling characteristics. The optimal LHDs generated via this

proposed methodology are then compared with optimal LHDs generated from single criterion optimization based on design characteristics and simulation studies via Gaussian process regression.

The analyses in this work revealed the column-wise exchange simulated annealing algorithm to be less effective for broad exploration of the design space, and consequently unable to effectively populate the PF. We think that the column-wise exchange mechanism makes the solutions largely dependent on the starting design points, hence limited the space and efficiency of the algorithm. On the contrary, the NSGA-II procedure proved to be more efficient in the exploration of the design space and populating the PF. Since the NSGA-II is more efficient for the analyses in this work, only the results via the NSGA-II algorithm are presented.

4.5.1   Low Dimensional Design Example

Consider the problem of constructing a design of computer experiment with 20 runs and two design factors. For this scenario, there is no need to consider maximum projection criterion since the only subspace is the univariate space and the LHD structure already ensured a uniform distribution in univariate dimension. Therefore, we only consider the maximin distance criterion and minimax distance criterion for simultaneously maximizing the spread and coverage of the design locations throughout the input space. After several searches for the single criterion optimization, the optimal MmLHD and optimal mMLHD found achieve the optimal criterion values at 4.0947 and 0.1646 respectively. In calculating the maximin distance criterion value for the LHD, we employed the simplified alternative maximin distance criterion proposed by Morris and Mitchell [77], such that minimizing their proposed maximin distance criterion is the same as maximizing the standard maximin distance criterion.

For the multiple criteria optimization, the Pareto front optimization approach and the search algorithms were used to simultaneously optimize the maximin distance criterion

and the minimax distance criterion in the construction of the LHDs. The NSGA-II algorithm takes an average of 2 minutes to run 100 generations on a standard desktop computer. Thus, it takes an average of about 3 hours for 10000 generations. Multiple separate NSGA-II searches were run in parallel to speed up the search process for populating the Pareto front. The final Pareto front, containing 16 optimal LHDs (including the optimal MmLHD and mMLHD from the single criterion optimization) was identified after running 70 separate NSGA-II searches of 10000 generations each. Figure 4.3 is the plot of the Pareto front in two scale; the raw value of the criteria (on the top and right axes) and the desirability scale (on the bottom and left axes).
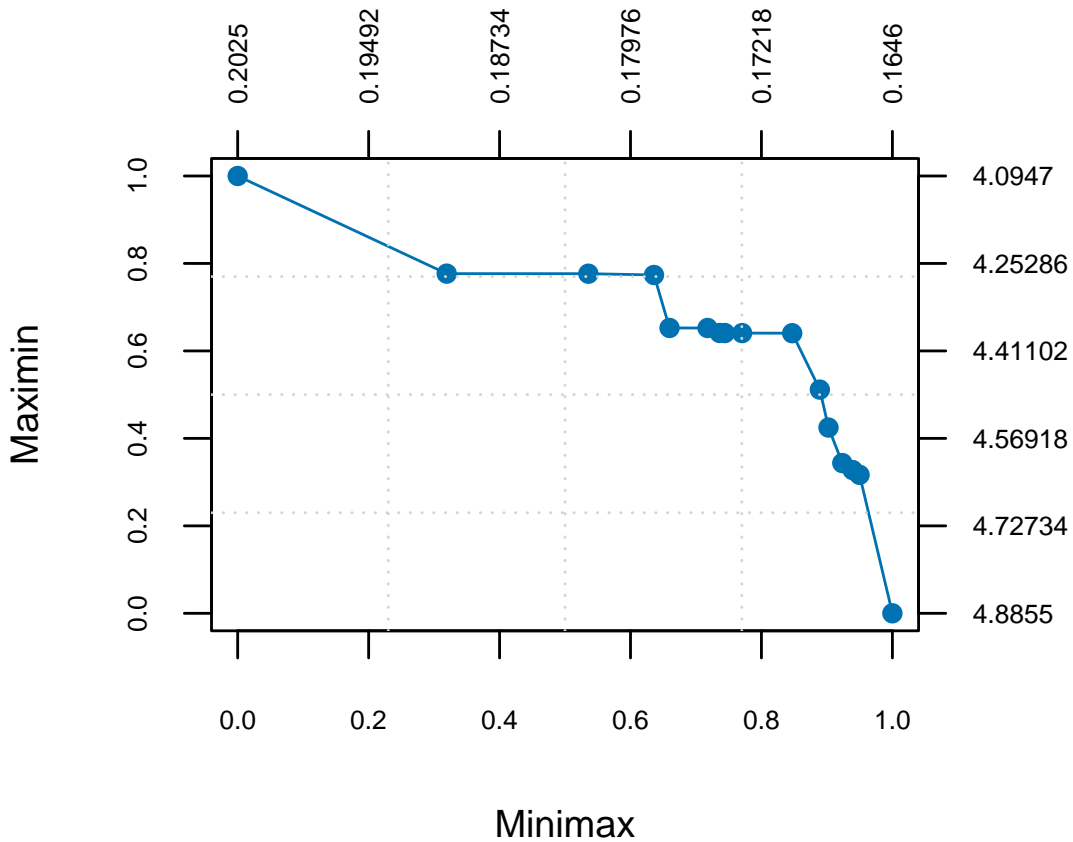
Figure 4.3: Pareto front for the maximin criterion and minimax criterion optimization.

The designs on the Pareto front were subjected to further evaluation to examine robustness to different weighting choices and the trade-offs between criteria among com-

peting designs. First, we use the Utopia point method to select the most performing designs from the Pareto front. For a given set of w_j in Equation 4.7, the Utopia point method rank the solutions on the Pareto front and then selects the best design. It iterate through a user specified grid of weights combinations to evaluate each point on the Pareto front. The selected solutions are those which are optimal for at least one of the combinations of the w's. In our application, since there is little prior information about relative contributions from the criteria, we propose exploring all possible weights combinations and observe how the optimum designs changes. Also, we transform each of the criteria to a desirability scale so that the best and the worst values on the Pareto front are 1 and 0 respectively. This is to allow consistency in the range of best to worst values for each criterion, and easy interpretation of the relative weights. In the Utopia point method, we employ the additive desirability function approach to combined the criteria, i.e the scaled criteria are combined as weighted sum.

The Utopia point method using the L_1 norm across a fine grid of all possible weights selects six optimum designs from the Pareto front. Figure 4.4 is a mixture plot that shows the distributions of weights for the selected designs. Every point within this rectangular-like shape corresponds to a weight combination with the sum of the two weights (i.e weights for both criterions) equaling one. The Figure shows that design 10 and design 1 are optimal for large range of weights. Particularly, design 10 is optimal for weights between 0.25 and 0.6 for the maximin distance criterion, which indicates that this design is optimal when two criteria are considered similarly important. Design 1 is optimal for weights between 0.725 and 1 for the maximin distance, which suggests this design should be selected if maximin distance criterion is the dominating criterion. Designs 16, 15 and 4 are optimal for small ranges of weights while design 11 is optimal for only very specific values of weights. A plot like this can be useful for decision makers when there is uncertainty about the relative importance of different criteria. From the mixture plot, we can also deduce that; design 4 and design 10 offer some balances when both criteria

are simultaneously considered, designs 1 appears optimal when the maximin criterion is valued the most while design 15 and design 16 appears optimal when the minimax criterion is considered more important.
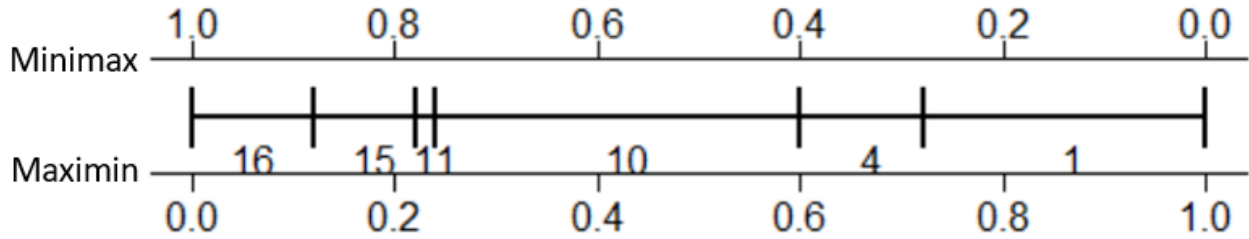
Figure 4.4: The mixture plot for the six optimum designs

Figure 4.5 is the trade-off plot [67] showing the criteria values for the six designs selected by the Utopia point method. The plot has two scales: the desirability scale of the criteria values (inner vertical axis) and the original scale of the criteria values (outer vertical axis).
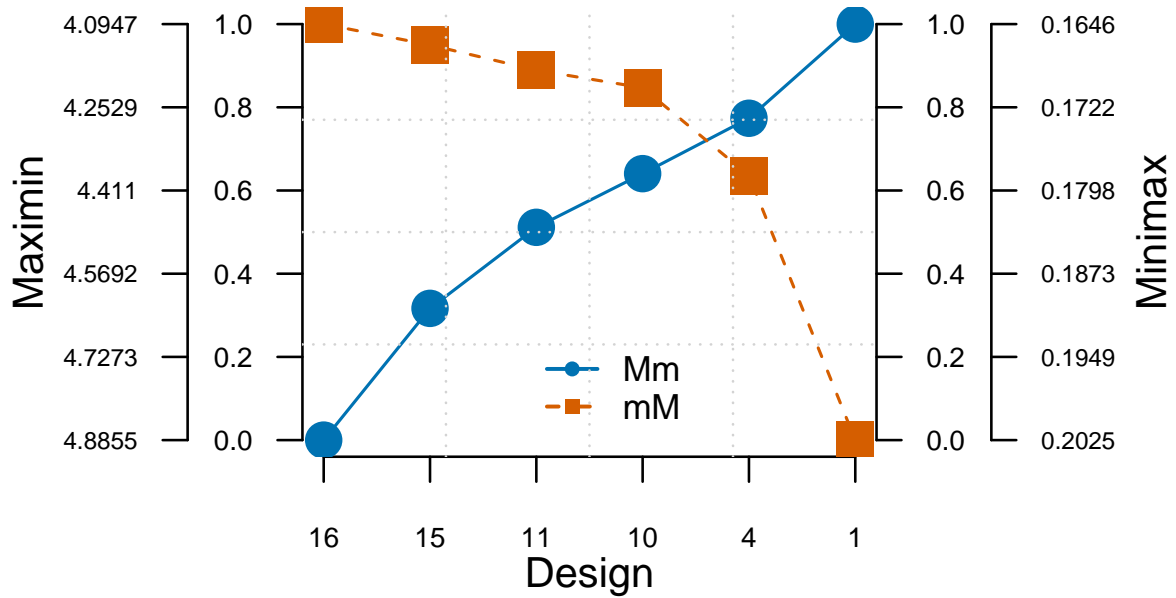
Figure 4.5: The plot showing trade-offs between the 6 optimum designs

The designs are sorted based on the maximin distance criterion. The observations from the trade-off plot are similar to that from the mixture plot discussed above. From Figure 4.5, notice that design 1, which has optimal performance for the maximin distance criterion has a poor performance for the minimax distance criterion. Similarly, design 16 which has optimal performance for the minimax distance criterion, has a poor performance for the maximin distance criterion. Again, design 10 and design 4 are clearly seen as choices with some balance when both criteria are considered simultaneously.

Next, the performance of the six optimum designs selected from the mixture plot are further examined through simulation studies. A response surface model of the form, $z(x) = x_1 exp(-x_1^2 - x_2^2), x_1, x_2 \in [-2, 2]^2$, is used to simulate data from each design and a Gaussian process regression model with a constant mean is fitted to each data. For the covariance function of the Gaussian process, we considered the product power exponential correlation function, using three different choices of the smoothing parameter, that is, v at 1, 1.5 and 2. The mean square prediction errors of the fitted Gaussian process models are evaluated over a fine grid of points spanning the input space.
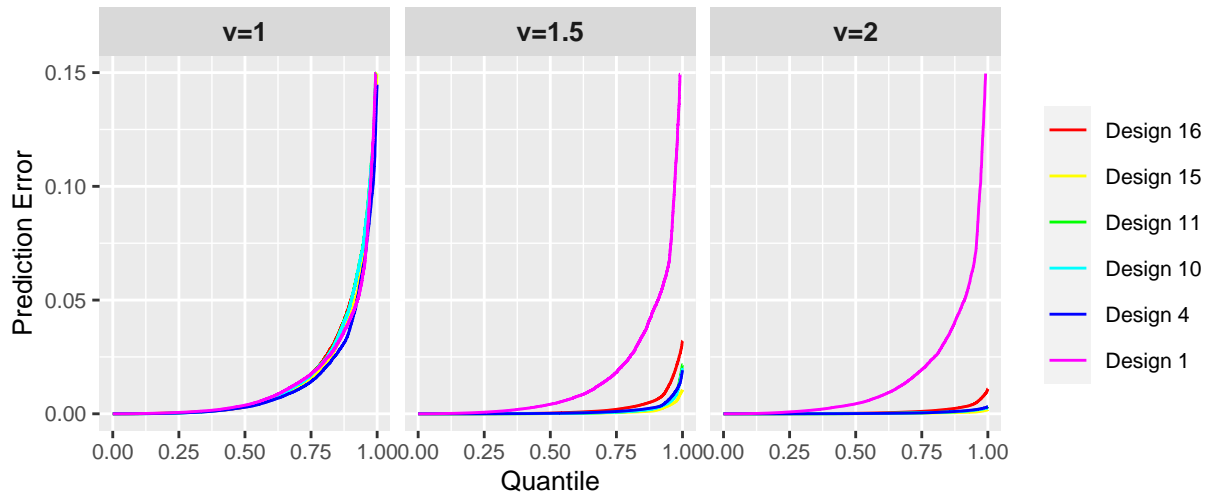
Figure 4.6: FDS plot for comparing the six optimum designs selected by the Utopia point method, using data simulated from the model $z(x) = x_1 exp(-x_1^2 - x_2^2), x_1, x_2 \in [-2, 2]^2$

Figure 4.6 is the fraction of design space (FDS) plot for comparing the six optimum designs. Each design is represented by a curve in the FDS plot. Each curve reflects the fraction of the design that has the prediction variance below certain percentage. Intuitively, a low flat curve in the FDS plot corresponds to a small mean square prediction error through the design space. Figure 4.6 shows that the designs with some balance in the minimax distance and the maximin distance criteria (Design 10 in light blue, Design 4 in blue, Design 11 in green and Design 15 in yellow) generally perform better in terms of the prediction errors when compared with designs that are optimal based on a single criterion (Design 1 (MmLHD) and Design 16 (mMLHD)). Specifically, Design 15 has the lowest mean square prediction error. Two alternative forms of the response surface model are explored to further observe the performance of the six designs. The results obtained are consistent with the ones presented here.

4.5.2   High Dimensional Design Example

In this section, a design problem of higher dimensions (more design factors and run size) is used to illustrate the proposed methodology. Alongside the maximin criterion and minimax criterion, the maximum projection criterion is also considered in this example since the number of design factors exceed two and space-filling for projections in all possible subspaces is desired. We consider the problem of constructing a design of computer experiment with 50 runs and 4 input factors. After several searches for the single criterion optimization, the optimal MmLHD, MaxProLHD and mMLHD found achieve optimal criterion values at 2.5854, 37.1714 and 0.4062 respectively. The NSGA-II took an average of 16 minutes to run 100 generations on a standard desktop computer. The Pareto front, containing 536 optimal LHDs (including the MmLHD, MaxProLHD and mMLHD from the single criterion optimization) was identified after running up to 50 separate NSGA-II searches of 10000 generations each.

Similar to the previous example, the impact of different subjective choices and weightings on selecting further solutions from the Pareto front are examined via graphical summaries. Using additive desirability function approach, five optimum designs were selected from the 536 designs on the Pareto front by the Utopia point method. Figure 4.7 shows the 3D mixture plot of the selected 5 optimum designs. Every point in the triangle corresponds to a weight combination with the sum of the three weights equaling one. The vertices and the edges correspond to optimizing a single criterion and optimizing two of the three criteria, respectively. Different colors are used to distinguish different selected designs.
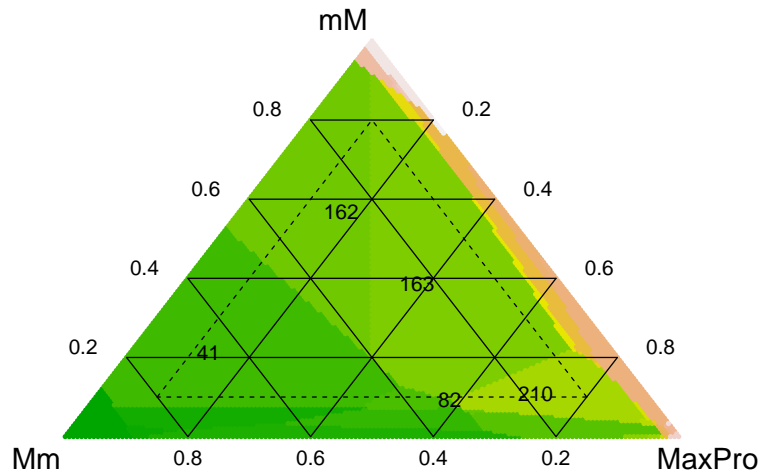
Figure 4.7: The 3D mixture plot of the five optimum designs

From Figure 4.7, we see that Designs 163, 41 and 162 are optimal for large ranges of weights, which indicates good robustness to weight ambiguity. Design 162 is optimal when the minimax distance criterion is valued the most while the maximum projection criterion is valued the least. Design 163 is optimal when the minimax and maximum projection criteria are considered more important than the maximin criterion. While Design 41 is otpimal when the maximin is considered the dominating criterion. It should be noted that the 5 optimum designs identified by the Utopia point method do not include the optimal MmLHD MaxProLHD and mMLHD from the single criterion optimization,

which suggests the single-criterion optimal solution is optimal only when that criterion is considered and offers inferior performance on other characteristics of interest.

The trade-offs between the reduced set of designs are shown in Figure 4.8. The designs are sorted based on the maximin distance criterion. The observations from the plot are similar to that from the 3D mixture plot. From both plots, Design 41 is the best solution when the maximin distance criterion is favored, Design 162 is desired when the minimax criterion is valued more and Design 210 is considered optimal when maximum projection criterion becomes the dominating criterion.
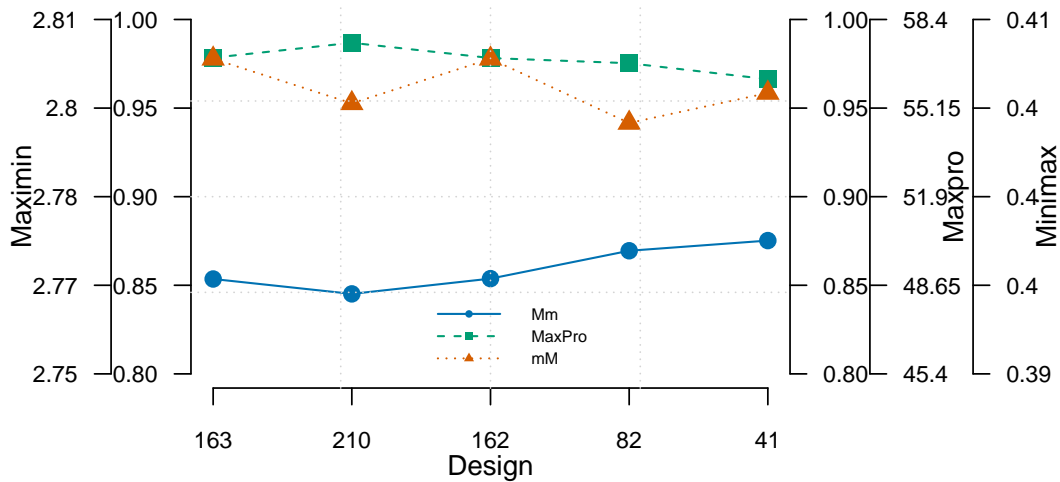
Figure 4.8: The plot showing trade-offs between the criteria for the 5 optimum designs

The performances of the five optimum designs together with the optimal designs from the single criterion optimization were then examined via simulation studies. A response model of the form, $z(x) = exp(sin(0.9 \times (x_1 + 0.48)^{10})), x_1 \in [0, 1]$, is used to simulate data from individual designs and Gaussian process models are fitted to each data. Then the mean square prediction errors of the designs were estimated over a fine grid of points spanning the four dimensional input space. Figure 4.9 is the FDS plot for comparing the designs in terms of the prediction variance. Design 1 (MmLHD), Design 530 (MaxPro-

LHD) and Design 534 (mMLHD) are also included, the optimal designs based on a single criterion optimization.
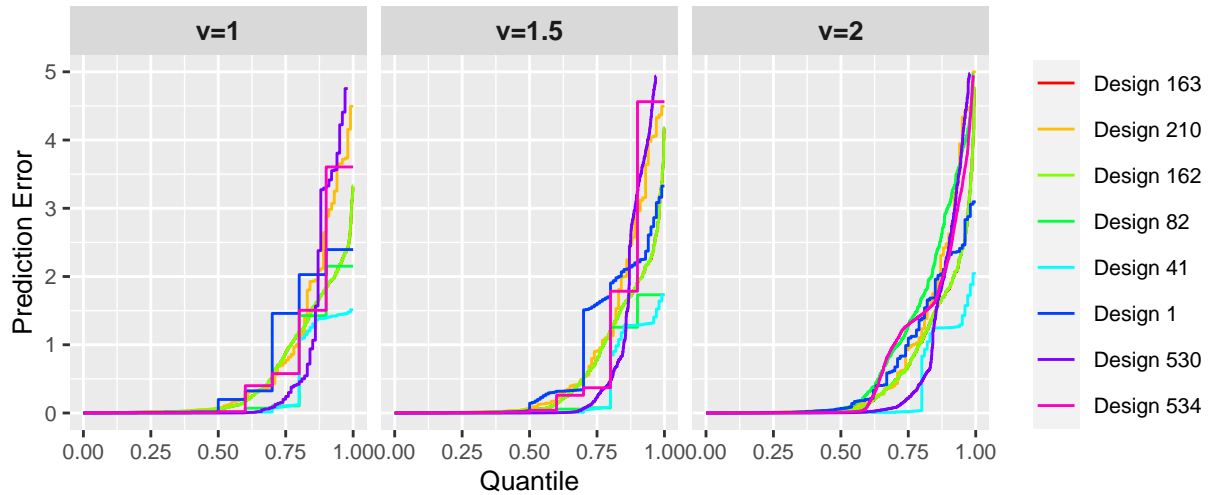
Figure 4.9: FDS plot for comparing the five optimum designs selected by the Utopia point method and the optimal designs from the single criterion optimization, using data simulated from the model $z(x) = exp(sin(0.9 \times (x_1 + 0.48)^{10})), x_1 \in [0, 1]$

The Figure shows that the MaxProLHD (purple) performs well for about 90% of the design space but really poorly for the remaining region. The Pareto front designs: 41 (light-blue), 162 (light-green) and 163 (red) have more robust performance, especially in terms of preventing the worst-case scenario. Design 41 generally performs well across the entire design space. Two alternative response models are explored, the results are consistent with the illustrated example. In general, optimal LHDs generated from the multiple criteria optimizations are better choices compared to optimal LHDs based on a single criterion optimization.

4.6 Conclusion

In design of computer experiments, most of the existing space-filling designs are generated based on optimizing a single criterion. These designs are found to perform well on one aspect of design characteristics and relatively inferior on other aspects. Simultane-

ously optimizing multiple criteria is desired to generate better designs that have balanced performance in multiple aspects of design characteristics. Moreover, considering multiple objectives for an experiment, in principle, generally result in a set of Pareto-optimal solutions, instead of a single optimal solution.

Utilizing the Pareto front optimization approach, the column-wise exchange simulated annealing algorithm and the NSGA-II algorithm were used to search for optimal designs. The NSGA-II algorithm is shown to have a better performance than the column-wise exchange simulated annealing algorithm. The methodology in this work is illustrated with two examples of varied dimensions. The performance of the generated optimal LHDs are evaluated via graphical summaries and simulation studies, and compared with designs from single criterion optimization to demonstrate their improved performance. The NSGA-II algorithm proved more efficient in eliminating noncontending choices from the solution space, thereby allowing identification of the most promising designs.

Although, this work considers simultaneous optimization based on the minimax criterion, maximin criterion and maximum projection criterion, the proposed methodology is very general to be adapted for other space-filling criteria.

Chapter 5: Bayesian Optimal Design for Accelerated Degradation Tests with Multiple Accelerating Factors

5.1 Introduction

Manufacturers are often faced with the challenges of developing highly reliable products in a timely fashion. This has encouraged the development of statistical methods for reliability testing. A common approach to describing the reliability of products is estimation of the failure-time distribution of the products based on data obtained from life tests. That is, to observe failure or survival time of products under their normal use conditions. However, the development in technology has instigate many modern products to be designed to have a very long lifetime under their normal use conditions. Thus, traditional life tests and even accelerated life tests yield few or no failure for highly reliable products. To obtain reliability data in a more timely fashion, engineers choose to measure the degradation process of the products and relate it with the evaluation of the product reliability. The reliability of a product is the probability that it would perform its required functions under defined conditions for a specified operating period. Degradation testing is used to measure the physical degradation in a product quality or performance characteristic as a function of time, providing more direct information about product's reliability than traditional life test.

For some products with high reliability, the degradation rates at normal use conditions can be too low to offer sufficient information for precisely predicting product reliability at future time point of interest. For instance, because of the organic nature of the dye layer of an optical device, the degradation and breakdown of the transparent portion of

the dye layer can take many years in a normal environmental condition [85]. Engineers therefore use accelerated tests to elevate the degradation process of highly reliable products, by putting them under elevated stress or environmental conditions, such as higher temperature, voltage, pressure, or humidity level [74] to aid timely availability of data for reliability analysis. This process is called accelerated degradation test (ADT). Under these conditions, degradation measurements are obtained and used to fit a model for describing the relationship between the accelerating conditions and degradation measures. This relationship can then be extrapolated to predict the degradation path and reliability under normal use conditions. The degradation measurements can be obtained as a single measurement in one shot experiment or as repeated measurements at specific points in time. We focus on repeated measurement experiments in this work.

In practice, limitation in experimental resources due to time and budget restrictions is a common challenge. Thus, effective planning of a test before the actual experimentation is important. Optimal test plans are often explored to maximize the information gain subject to the resource limitations. A common approach to this is to construct an optimal accelerated degradation test (ADT) plan for a selected criterion that can be used to guide the construction of efficient practical test. The selection of the criterion is often based on the aim of the experiment and the reliability metric of interest. A common aim in ADT plan is to estimate a particular quantile of the failure-time distribution at the normal use condition. The criterion for this is to minimize the variance of the estimated quantile lifetime of interest at the normal use condition, which is often regarded as the C-criterion. Other common criteria include the D-criterion, which maximizes the determinant of fisher information matrix of the model parameters for achieving maximum precision in estimated model parameters, and the entropy approach which maximizes the relative entropy to maximize the information gain about the expected responses [74].

When constructing an ADT with repeated measurements, the following planning values have to be specified: the overall test duration, how often the repeated measurements

will be taken, the number of test units, and the testing conditions (particularly, the level combinations of the accelerating factors) that test units will be exposed and the number of units that will be allocated to each testing condition. In addition, information about the form of the degradation model, model parameters and the failure threshold (i.e. the level of degradation at which a unit is deemed to fail) needs to be specified for seeking an optimal test plan [112]. Frequentist methods for test planning require specification of values of the model parameters [74]. Such values are generally not known precisely during the test planning stage. Thus, a test plan based on a subjective choice of the planning values may fail to get the most useful information from the data given the limited resources. In contrast, Bayesian approach allows the use of prior distributions for capturing the uncertainty of planning parameters and hence offers more robust and realistic solutions to test planning. In some applications, certain information regarding the underlying models or possible range of parameter values are often available from previous studies or expert knowledge about the failure process from previous experiments. Bayesian methods can easily incorporate those prior information about model parameters into planning and estimation of the criterion of interest, providing test plans with better statistical precision.

As mentioned earlier, the testing conditions that test units will be exposed to and the number of units to be tested at each testing condition are essential planning values to be specified by the experimenter in ADT. Often, the choice of these values are ad hoc, which could result in a loss of opportunity to improve our understanding of the failure mechanism given the available resources. This chapter develops a novel approach of selecting optimal accelerated degradation tests plans. Particularly, we propose using the Bayesian optimal design and heuristic search algorithm to select optimal designs (testing conditions and the number of units be tested at each testing condition) for ADT with two accelerating factors, where the goal is to precisely estimate the quantile of the failure time distribution at the normal use condition. That is, given the total number of affordable test units, how should the experimenter allocate resources in order to optimize the model

performance and consequently increase the information learned about the degradation path and the reliability performance of the product of interest.

A Bayesian criterion based on precise estimation of the failure-time quantile of the degradation model at the normal use condition is used for seeking optimum designs. The large-sample approximation approach [12] is used to estimate the posterior distribution in the Bayesian analysis. To search through the design space for optimal solutions, a genetic algorithm is used in this work to optimize the Bayesian criterion function. It should be noted that, in order to suit the language used in design of experiments literature, the terms "testing condition" and "testing location" are used interchangeably in this work.

5.1.1   Related Work

This subsection will review existing work in accelerated degradation test planning, and Bayesian accelerated degradation test plans.

Several work have been done on planning ADT with repeated measures. Although not much work has been done for developing optimal ADT plans with two or more accelerating factors [30, 88]. Nelson [83] and Meeker et al. [74] discussed different linear models and models made linear through transformation for ADT. Stochastic models have also been broadly used for analyzing ADT data. A few commonly used models include, the Wiener process [62], the Gamma process, and the inverse Gaussian process models [60, 116]. Other types of models that have been used for analyzing ADT data include the exponential-dispersion models [105], nonlinear regression for physics-based models [102], the lognormal distribution [117], hierarchical models [30, 87], and mixed-effects linear models [110].

Bayesian methods have been applied for ADT test planning. Shi and Meeker [97] discussed a Bayesian method for seeking optimal accelerated destructive degradation test plans for nonlinear degradation models with a single accelerating factor by maximizing the precision of an estimated life time quantile under the normal use condition. Li et

al. [60] propose a Bayesian ADT design based on the inverse Gaussian process, by considering three optimization criteria including the relative entropy, quadratic loss function, and Bayesian D-optimality. Recently, Weaver and Meeker [111] develop a Bayesian method for planning repeated measures ADT with one accelerating factor, where a linear model is assumed for the degradation path. They defined a Bayesian criterion based on precision in estimation of the failure-time quantile at use conditions for finding optimum test plans and a large-sample approximation for the posterior distribution was used to evaluate the planning criterion. In their analysis, the optimality of a given test plan was checked via the general equivalence theorem [118].

In this current work, as opposed to Weaver and Meeker [111], we consider test planning for repeated measures ADT with two accelerating factors and we use an heuristic search algorithm to find the optimum test plan.

5.1.2 Motivating Application

The motivating application considered in this work is the accelerated degradation test conducted by the international organization for standardization (ISO) for predicting the life expectancy of an optimal media [30, 101]. The ISO is responsible for developing and maintaining standards in information and communications technology. Particularly, it uses an accelerated degradation test for investigating the effects of temperature and relative humidity on the lifetime of optical media and then build a regression model for estimating the degradation process and predicting the life expectancy of an optical media [101].

Table 5.1: Testing conditions

Test Cell	Temperature ($^\circ C$)	Relative Humidity (%RH)	Number of Specimens
1	85	85	20
2	85	70	20
3	65	85	20
4	70	75	30

Four testing conditions were selected in the ISO ADT, as shown in Table 5.1. The specimens tested in test cells 1 and 2 are measured at 0, 250, 500, 750, and 1000 hours of testing; 0, 500, 1000, 1500, and 2000 hours in test cell 3; and 0, 625, 1250, 1875, and 2500 hours in test cell 4. The raw data is given in Appendix B.3. Figure 5.1 shows the degradation paths (measurements in log scale) as a function of time (measured in hours). We can observe different degradation rates across the varied test conditions. The highest degradation rate occurs at the test condition with the highest temperature, $85°C$, and highest relative humidity, 85%. The units tested under lower levels of temperature and relative humidity generally have lower degradation rates.
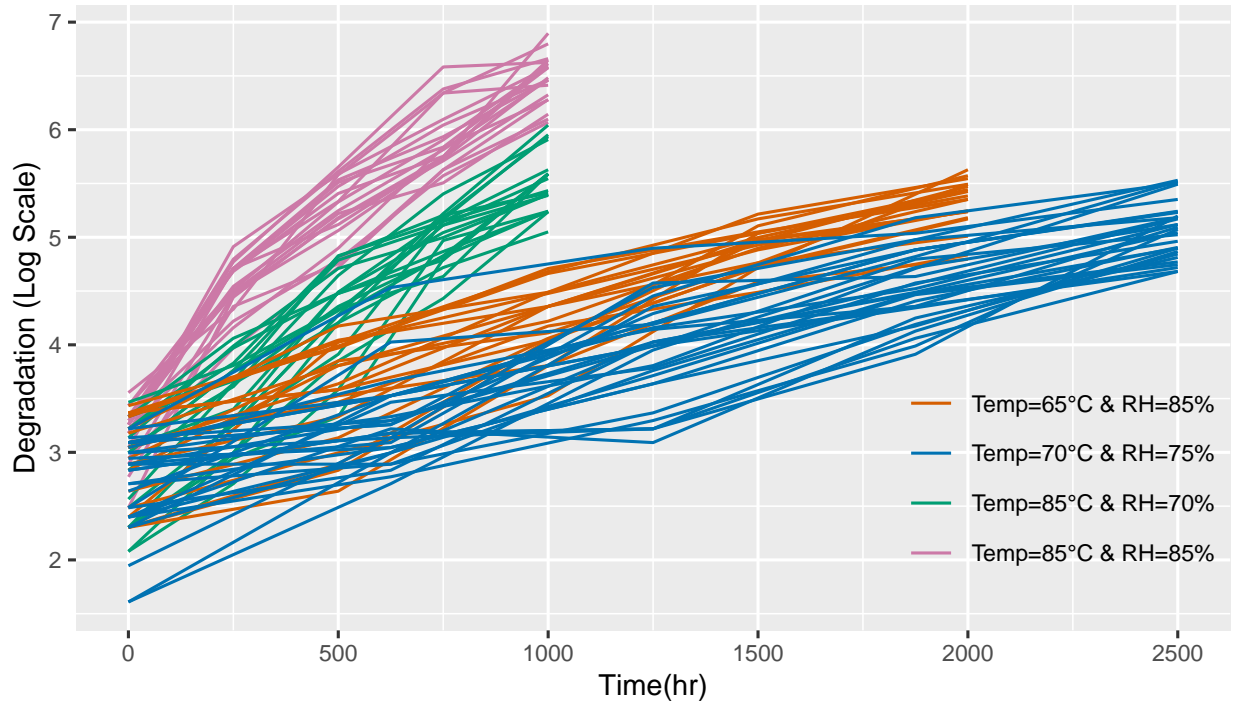
Figure 5.1: Degradation path of the ISO optical media data

Figure 5.1 also shows the presence of unit-to-unit variation in the initial degradation levels and the degradation rates. For instance, at the initial time point (time 0), each degradation path starts with different initial values, suggesting varied starting conditions of the test units. Similarly, units with higher initial degradation values are generally associated

with higher degradation rates over time. We assume a linear mixed-effects model for the degradation paths. The random effects in the model is use to capture the unit-to-unit variation while the fixed effects is use to capture the effects of testing conditions on the degradation rates. Also, a general linear trend can be observed in the degradation paths, further justifying the appropriateness of using a linear mixed-effects model.

The end of life of a disc is the time when the information recorded on the disc cannot be recovered without losses. For the ISO optical media data, a failure is defined to occur when the degradation measure on the PI Sum8 reaches the failure threshold value at 280 [101]. In other words, the test unit is considered to fail when its degradation level exceeds this threshold value. This threshold is often referred to as the soft failure threshold in the reliability literature. PI Sum8 refers to the parity inner error rate, summed over 8 consecutive error correction blocks, a common indicator used in monitoring the error rate of a disc device. As observed in Figure 5.1, all the units tested at the toughest test condition ($85°C$ and 85% RH) failed and the testing period for this condition lasted for 1000 hours. However, no failure was observed during the testing conditions; ($65°C$, 85% RH) and ($70°C$, 75% RH). Intuitively, it would take a very long time before any failure can be observed under the normal use condition, a much lower temperature and relative humidity level ($25°C$, 50% RH) [101]. Thus, the accelerating factors play an important role in accelerating the degradation process of the optical media.

Engineers are interested in finding the optimal design with the optimal testing locations that the optical media should be exposed to and the number of units that should be tested at each test location subject to the resource restriction. We consider the scenario when the objective is to precisely estimate the quantile of the lifetime distribution at the normal use condition. The goal will dictate our choice of the design criterion when seeking the optimal test plan. The remainder of this chapter is organized as follow. Section 5.2 describes the accelerated degradation model, the failure time distribution, and the reliability metric of interest which is the *p*-quantile of the lifetime distribution. Section

5.3 discuss the proposed Bayesian optimal design for ADT with repeated measurements based on linear mixed effects model and describes the genetic algorithm developed for seeking optimal test plans based on the chosen criterion. Section 5.4 illustrates the proposed method via the ISO example. Section 5.5 renders the conclusion and discusses potential future work.

5.2 Accelerated Degradation Path Model

Predicting the product lifetime for an ADT requires two key components: the degradation path model and the accelerating factor(s). One class of most popular models used for analyzing degradation data is the general path models [64]. Consider the ADT with two accelerating factors such as the ISO data, the true degradation level for an observational unit at time t is generally denoted by, $\mathcal{D} = \mathcal{D}(\tau, \mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\theta})$, where $\tau = h_t(t)$ is a monotone increasing transformation of time t, $\mathbf{x}_1$ and $\mathbf{x}_2$ are vectors of (possibly transformed) accelerating factor(s), and $\boldsymbol{\theta}$ is the unknown parameter vector. We assume $\mu(\tau, \mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\theta})$ has the form of mixed effects model given by

$$
\begin{aligned}
Y &= h_d(\mathcal{D}) + \epsilon \\
&= \mu(\tau, \mathbf{x}_1, \mathbf{x}_2) + \epsilon \\
&= \mathbf{x}_1\gamma_1' + \mathbf{x}_2\tau\gamma_2' + b_0 + b_1\tau + \epsilon \\
&= \mathbf{x}_1\gamma_1' + \mathbf{x}_2\tau\gamma_2' + \mathbf{z}\mathbf{b}' + \varepsilon.
\end{aligned}
\tag{5.1}
$$

In the above formula, $\mu(\tau, \mathbf{x}_1, \mathbf{x}_2) = h_d(\mathcal{D})$ is a location parameter for the distribution of Y that depends on the unknown parameters $\boldsymbol{\theta} = (\gamma_1', \gamma_2')'$, $\mathbf{z} = (1, \tau)$, $\varepsilon \sim iidN(0, \sigma^2)$ is a random variable that describes the within unit variation and h_d is a monotone increasing or decreasing transformation of $\mathcal{D}$. In Equation 5.1, the term $\mathbf{x}_1\gamma_1'$ describe how the intercept or initial value of degradation changes as a function of the accelerating factors(s), $\mathbf{x}_2\gamma_2'$ describe how the degradation slope or rate changes as a function of the accelerating

factors(s), and $\mathbf{zb}'$ describes the variability in the degradation slope and intercepts among the test units. In the equation above, $\mathbf{x}_1 = \mathbf{g}_1(\mathbf{x})$ and $\mathbf{x}_2 = \mathbf{g}_2(\mathbf{x})$ are possibly transformations of the accelerating factors. They are treated as vectors and hence could include more than one accelerating factor that influence both the initial value and the degradation rate.

For the ISO optical media application introduced in Section 5.1.2, the following model specification is employed. Let Y_{ijk} be the log transformed degradation measure for unit i, at time point j, at the test location k (with a certain combination of the temperature and relative humidity level). Let x_{1k} be the Arrhenius transformation of temperature, i.e., $x_{1k} = -11605/(T_k + 273.15)$ where $T_k = \text{Temp}^\circ C$, and x_{2k} denotes the log transformed relative humidity (RH), i.e., $x_{2k} = log(RH)$. The linear accelerated degradation mixed effects model is expressed as:

$$Y_{ijk} = \gamma_1 x_{1k} \tau_{ij} + \gamma_2 x_{2k} \tau_{ij} + b_{0i} + b_{1i} \tau_{ij} + \varepsilon_{ijk} \tag{5.2}$$

where τ_{ij} is the square root transformed time. To capture the variability of the initial degradation level and degradation rate among the individual test units, we use random effects for the linear regression parameters, $\mathbf{b}$, which is assumed a bivariate normal distribution, $\mathbf{b} = (b_0, b_1)' \sim \text{BVN}(\boldsymbol{\beta}, \mathbf{V})$, where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ is the mean vector representing the average intercept and slope for all the units, and

$$\mathbf{V} = \begin{pmatrix} \sigma_{b_0}^2 & \rho\sigma_{b_0}\sigma_{b_1} \\ \rho\sigma_{b_0}\sigma_{b_1} & \sigma_{b_1}^2 \end{pmatrix}$$

is the covariance matrix. Note, the inclusion of ρ allow us to capture the within-unit correlation observed between the initial condition and the degradation rate. We assume that the random effects $\mathbf{b} = (b_0, b_1)'$ are independent of the random error $\varepsilon \sim N(0, \sigma^2)$. This means that the units are independent. Let $\boldsymbol{\theta} = (\gamma_1, \gamma_2, \boldsymbol{\beta}, \sigma_{b_0}, \sigma_{b_1}, \rho, \sigma)$ denote the vector of the parameters of the model.

5.2.1 Failure-Time Distribution for the Degradation Model

Given a degradation model and the known soft failure threshold, a failure-time distribution can be derived from the degradation model in Equation 5.2 [74]. In general, the failure time distribution can be written as a function of the degradation model parameters. That is, a unit "fails" when the degradation level reaches a prespecified soft failure threshold level $\mathcal{D}_f$. Thus, for increasing degradation, $\mathcal{D} \geq \mathcal{D}_f$ and for decreasing degradation, $\mathcal{D} \leq \mathcal{D}_f$. Then the failure time of a unit is the time that it reaches the failure threshold $\mathcal{D}_f$. Let T denote a unit's time to failure. Since $b_0 + b_1\tau \sim N(\beta_0 + \beta_1\tau, \sigma_{b_0}^2 + \tau^2\sigma_{b_1}^2 + 2\tau\rho\sigma_{b_0}\sigma_{b_1})$, for increasing $\mathcal{D}$, we have

$$
\begin{aligned}
F_T(x_1, x_2 : t) = Pr(T \leq t) = Pr(\mathcal{D} \geq \mathcal{D}_f) &= Pr(b_0 + b_1\tau \geq \mathcal{D}_f - x_1\tau\gamma_1 - x_2\tau\gamma_2) \\
&= 1 - Pr(b_0 + b_1\tau \leq \mathcal{D}_f - x_1\tau\gamma_1 - x_2\tau\gamma_2) = 1 - \Phi_{nor}(\varkappa)
\end{aligned}
\tag{5.3}
$$

where

$$
\varkappa = \frac{\mathcal{D}_f - x_1\tau\gamma_1 - x_2\tau\gamma_2 - \beta_0 - \beta_1\tau}{\sigma_{b_0}^2 + \tau_p^2\sigma_{b_1}^2 + 2\tau_p\sigma_{b_0 b_1}}
$$

and Φ_{nor} denotes the cumulative distribution function of the standard normal distribution. Similarly for decreasing $\mathcal{D}$,

$$
F_T(x_1, x_2 : t) = Pr(\mathcal{D} \leq \mathcal{D}_f)) = \Phi_{nor}(\varkappa).
\tag{5.4}
$$

Although the failure time is dependent on the true unobserved degradation path, the estimation of the parameters is based on the observed and possibly transformed degradation data. $F_T(x_1, x_2 : t)$ can be estimated by evaluating Equation 5.3 or Equation 5.4 at the maximum likelihood (ML) estimates of $\boldsymbol{\theta}$. For complex models where closed form solution does not exist, numerical integration and simulation-based methods can be employed [74]. From Equation 5.3 and Equation 5.4, the failure-time quantile (τ_p) at a par-

ticular accelerating condition is $t_p = h_t^{-1}(\tau_p)$, where

$$\tau_p = -\frac{-[k\rho\sigma_{b_0}\sigma_{b_1} + (\mathcal{D}_f - \beta_0)(x_1\gamma_1 + x_2\gamma_2 + \beta_1)] \pm \sqrt{\Psi}}{[k\sigma_{b_1}^2 - (x_1\gamma_1 + x_2\gamma_2 + \beta_1)^2]}. \tag{5.5}$$

Here, $k = [\Phi_{nor}^{-1}(1-p)]^2$ or $k = [\Phi_{nor}^{-1}(p)]^2$ depending on whether a failure is declared based on $\mathcal{D} \geq \mathcal{D}_f$ or $\mathcal{D} \leq \mathcal{D}_f$, respectively. Also, $k\sigma_{b_1}^2 \neq (x_1\gamma_1 + x_2\gamma_2 + \beta_1)^2$ and

$$\Psi = [k\rho\sigma_{b_0}\sigma_{b_1} + (\mathcal{D}_f - \beta_0)(x_1\gamma_1 + x_2\gamma_2 + \beta_1)]^2 - [k\sigma_{b_1}^2 - (x_1\gamma_1 + x_2\gamma_2 + \beta_1)^2][k\sigma_{b_0}^2 - (\mathcal{D}_f - \beta_0)^2].$$

More detailed derivation of Equation 5.5 is provided in Appendix B.2. If $0 < p < 0.5$, then the square root in Equation 5.5 is added. If $0.5 < p < 1$, then the square root in Equation 5.5 is subtracted. However, if $p = 0.50$, then $\Psi = 0$ (because $k = 0$), meaning Equation 5.5 has just one root.

5.3   Bayesian Design of Experiment for Accelerated Degradation Tests

Typically, experimenters have in mind the aim of the experiment and the utility function or criterion function for quantifying the aim, before performing the ADT. For example, the interest may be to estimate a quantile, τ_p, of the failure-time distribution as precisely as possible. Following the general approach proposed by Zhang and Meeker [118] to define the objective function, given that the utility function is the posterior variance of the quantile of interest, denoted by $var_{\theta|\mathbf{Y},\eta}(\tau_p)$, where η is a given design (i.e, given levels of the accelerating factors and the units allocation), τ_p is the p-quantile of the failure-time distribution, and $\mathbf{Y}$ is a vector of observed failure time data collected according to η. Therefore, an optimum plan is the one that minimizes this utility function, which is equivalent to maximizing $-var_{\theta|\mathbf{Y},\eta}(\tau_p)$. It should be noted that, improving the precision of the estimator of τ_p implies minimizing its posterior variance. The posterior variance is dependent on the unobserved data, therefore an expectation is taken over all possible

data.

$$C(\eta) = -\int_{\mathcal{Y}} var_{\theta|\mathbf{Y},\eta}(\tau_p)p(\mathbf{Y}|\eta)d\mathbf{Y} \tag{5.6}$$

The posterior variance of the failure-time quantile can be approximated by a function of the unknown model parameters θ and the Fisher information matrix using the delta method [12].

Let $g(\theta) = t_p$, $c_l(\theta) = \frac{\partial g(\theta)}{\partial \theta_l}$ for $l = 1, ..., 8$, and $\mathbf{c}(\theta) = (c_1(\theta), ..., c_8(\theta))^T$ (derivatives of τ_p w.r.t to each of the eight parameters in the vector θ). Then using the delta method to approximate the posterior variance, Equation 5.6 becomes:

$$C(\eta) \simeq -\int_{\mathcal{Y}} c(\theta)'[\hat{\mathbf{I}}(\theta,\eta)]^{-1}c(\theta)d(p(\theta)) \tag{5.7}$$

where $\hat{\mathbf{I}}(\theta,\eta)$ is the Fisher information matrix based on the design η evaluated at the maximum likelihood estimate $\hat{\theta}$. The Fisher information matrix is defined as $\mathbf{I}(\theta) = -E(\frac{\partial^2 L}{\partial \theta^2})$, the full derivation is given in Appendix B.1. It can be observed that Equation 5.7 is dependent on the data only through the maximum likelihood estimator $\hat{\theta}$, so the integration in Equation 5.6 can be performed with respect to the predictive distribution of $\hat{\theta}$, denoted as $p(\hat{\theta})$ where

$$p(\hat{\theta}) = \int p(\hat{\theta}|\theta)p(\theta|\vartheta)d(\theta).$$

Therefore, Equation 5.6 can be approximated as

$$C(\eta) = -\int c(\hat{\theta})'[\hat{\mathbf{I}}(\theta,\eta)]^{-1}c(\hat{\theta})d(p(\hat{\theta})). \tag{5.8}$$

The distribution $p(\hat{\theta}|\theta)$ is not tractable. Suppose $p(\hat{\theta}|\theta)$ is the appropriate joint prior distribution for θ with known hyperparameters ϑ, as the sample size increases, $\hat{\theta}$ converges in distribution to θ. That is $p(\hat{\theta}|\theta))$ converges to $p(\theta|\vartheta)$ and Equation 5.6 can further be approximated as

$$C(\eta) = -\int c(\hat{\theta})'[\hat{\mathbf{I}}(\theta,\eta)]^{-1}c(\hat{\theta})d(p(\theta|\vartheta)). \tag{5.9}$$

5.3.1 Prior Distributions

In Bayesian analysis, the specification of prior distributions for the model parameters is important. If there is little or no prior information, a common practice is to use non-informative or diffuse prior distribution for the analysis. The implication is that, the assumed prior will not have significant effect on the posterior distribution and resulting inferences. Moreover, since test planning criteria can be very sensitive to prior information, if a diffuse prior is combined with a data with little information (e.g. with few failures observed), the analysis could result in inefficiencies [11, 12]. Generally, the prior distribution used to design the experiment is expected to be an informative or weakly informative prior distribution.

In this work, the prior distributions used are based on a Bayesian analysis of the raw ISO degradation data. The RStan package [103] is used to fit the assumed model, Equation 5.2, to obtain the MCMC posteriors draws of the model parameters. Stan is probabilistic programming language used for Bayesian inference and optimization [36]. It uses the No-U-Turn sampler (NUTS) [43] to obtain posterior simulations given a user-specified model and data. RStan is the R interface to Stan that allows one to conveniently fit Stan models from R [103]. For this initial analysis, we used half-normal prior distributions with mean at 0 and standard deviation at 100 for the variance components. This is a normal distribution with mean 0 and standard deviation of 100 truncated so its support is $(0, \infty)$. For γ_1, γ_2 and β, we use a normal prior distribution with mean 0 and standard deviation of 100. The summary from the Stan analysis is given in Table 5.2. The Rhat statistic is evaluated for each parameter, this statistic defines the ratio of between-chain variance to within-chain variance [37]. An Rhat value of 1 indicate the convergence of the MCMC for the parameters.

Figure 5.2 is the density plots from the MCMC result. The bell shape indicate the convergence of the MCMC chains.

Table 5.2: Parameter estimation

Parameters	Estimation	Standard Error	Rhat
γ_1	0.42	0.02	1
γ_2	0.21	0.01	1
β_0	2.69	0.2	1
β_1	0.03	0.03	1
σ_{β_0}	0.28	0.12	1
σ_{β_1}	0.03	0.02	1
σ	0.37	0.01	1
ρ	-0.197	0.46	1
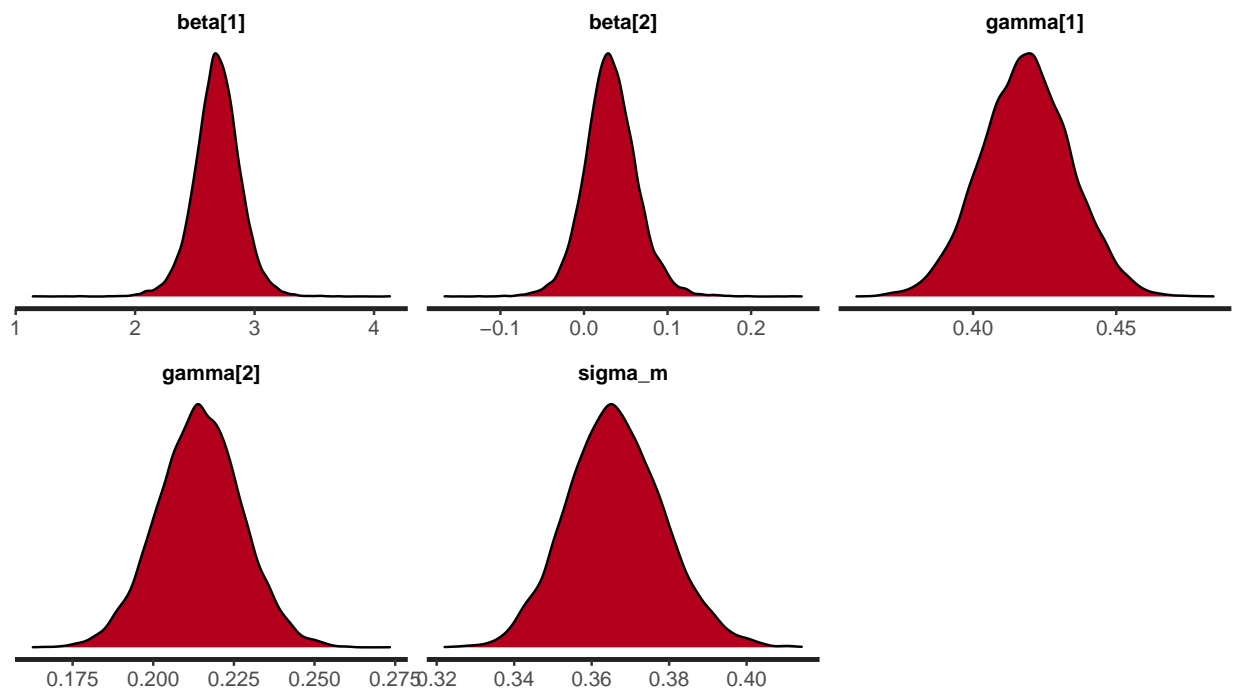
Figure 5.2: Density plots for convergence parameters

5.3.2   Finding the Optimum Design Plan

To find the Bayesian optimal test plan, a search algorithm is needed to explore the design space for seeking best performing designs. Typically in optimal design of experiments, the idea is to optimize an objective function selected for capturing a certain char-

acteristic of the design. In our case, it is to minimize the posterior variance of the quantile lifetime. Maximizing the objective function (criterion function) given in Equation 5.9 is equivalent to minimizing the posterior variance. We propose using a customized genetic search algorithm to search through the design space for designs that maximize the objective function given in Equation 5.9. Genetic algorithm is a kind of heuristic search algorithm that has been broadly used in design optimization. But to the best of our knowledge, it has not been used for finding optimal test plan in ADTs. As a matter of fact, there has been scarce work on seeking optimal test plans for ADTs with two or more accelerating factors.

5.3.2.1 *Genetic Algorithm for Seeking Optimal Test Plans*

A genetic algorithm (GA) is an optimization procedure inspired by the evolutionary process, which is used in optimizing an objective function of interest. The objective function is often referred to as fitness measure [39, 51] in GA language. At each generation of the GA, "offsprings" solutions are generated usually by the operations of crossover and mutation on the "parent" solutions. At the end of each generation, all the candidate solutions compete for the ability to create new solutions in the next generation based on their fitness value. More details about the GA can be found in Section 4.4 of Chapter 4.

Note that our focus is on choosing the best collection of test conditions for the accelerating factors while there is a fixed budget on the total testing units affordable and the inspection times are predetermined as well. We need to specify the set of possible levels of each accelerating factors. We choose our input region based on the ISO data, i.e. the maximum and minimum values of each accelerating factor variables determine the range of the input factors. For each accelerating factor, we choose five levels at Temp= $(65^{\circ}C, 70^{\circ}C, 75^{\circ}C, 80^{\circ}C, 85^{\circ}C)$ and RH= $(65\%, 70\%, 75\%, 80\%, 85\%)$, respectively. Then the candidate set of design locations include 25 unique testing conditions. For practical reasons, it is necessary to control the number of unique testing locations for

the selected test plans (i.e. testing at too many different locations is practically more difficult to implement). Hence, we adapt the GA search by imposing a range on allowable number of unique testing locations. This restriction is used for initializing the starting population and producing offsprings via the crossover and mutation operations. Thus, it make sense to state the minimum and maximum allowable number of unique testing locations rather than just fixing a particular number. In this work, we set the minimum and maximum allowable number of testing locations to be 4 and 10, respectively. Note that these values can be easily adjusted for different applications with different budget and logistic constraints.

The steps of the GA are given below, followed by the details of how each steps are implemented in this current study.

1. Specify $p(\theta|\vartheta)$, the joint prior distribution for the parameters.

2. Create and evaluate an initial population of N randomly generated solutions.

   (a) To generate each solution of the initial population subject to the constraint on the number of unique test locations, first randomly choose a number of unique testing locations from the specified range and generate these locations from the candidate set. Then randomly allocate number of units for each unique testing location. The allocation of units is done by the process of sampling (with replacement) n units from the unique locations generated, where n is the total number of test units required (design size).

   (b) The objective function in Equation 5.9 is evaluated for each solution in the initial population.

   The solutions are then be ranked from best to worst fitness.

3. Create $N/2$ solutions via crossover operation (to be described in more detail)

4. Create $N/2$ solutions via mutation operation (to be described in more detail)

5. Evaluate the objective function for the N new solutions.

6. The N new solutions are combined with the N parent solutions. The combined solutions are ranked from best to worst fitness. The top N of these solutions are then allowed to evolve into the next generation [21].

7. Steps 3–6 are repeated for a specified number of generations. The design with the maximum value of Equation 5.9, that is, minimum posterior variance is the best design.

In order to avoid elimination of superior solutions, we allowed the population size to grow at each generation. The top N solutions of the current generation evolves to the next generation to create N offsprings solutions. Half of the offsprings solutions are created by the crossover operation and the remaining half by the mutation operation.

Crossover in step 3 is used to keep the attributes of two existing solutions [27]. Rank-proportional selection [39, 69] is used to selects parents for the crossover and mutation operations. The probability of selecting the rth ranked candidate of the N_{pop} solutions is $\frac{(N_{pop}-r+1)}{(N_{pop}(N_{pop}+1)/2)}$. After two parent solutions have been selected based on the above probabilities, a crossover operation is used to generate offspring solutions as follows; The runs in the two parent solutions are combined. The unique testing locations in the combined runs is then ranked based on their frequencies. A random number, x is generated (satisfying the earlier defined range) to decide the number of unique testing locations for the child solution. And then the top x most represented unique testing locations in the combined unique testing locations are selected as the unique testing locations for the child solution. Lastly, a multinomial distribution with probabilities proportional to the frequencies of the x selected testing locations is used to generate a single offspring test plan.

The mutation operation in step 4 is used to create a new solution, which is a slight modification to an existing solution [27]. One parent design is randomly sampled from

the "parents set", then values of randomly selected runs are replaced by values of another randomly selected runs. We allowed the rate of replacement during mutation to reduce as the iterations increases (i.e as the generation number increases).

Usually in the GA implementation, the search for optimal solutions continues until the termination condition is met. For this work, the termination condition for the algorithm is when the specified number of generations is reached. Being an heuristic search algorithm, it does not necessarily guarantee finding the exact global optimal. However, we did check the optimal criteria value over the generations to ensure it converge (as the best criterion value is no longer changing much over different generations). In general, the longer the running of the algorithm, the better the chance that the algorithm will find the global optimum. In practice, a trade-off is often sort between the approximation accuracy and the computing time.

5.4   ISO Example

This section presents an example to illustrate the methodology proposed in this work. Consider an ADT for an optical media device where 90 test units are available to be exposed under different testing conditions of two accelerating factors (temperature and relative humidity). The goal of the ADT is to obtain most precise estimate of the 0.1 quantile lifetime distribution at the normal use condition. The time points at which measurements should be made are fixed at; 0,250,500,750,1000 hours, we want to find the optimum test plan to most efficiently distribute the testing resource. The rest of the planning information for this example are already mentioned above, and are summarized below:

- The failure threshold is determined at $\mathcal{D}_f = 280$.

- The levels for temperature and relative humidity for the ADT include $(65^\circ C, 70^\circ C, 75^\circ C, 80^\circ C, 85^\circ C)$ and $(65\%, 70\%, 75\%, 80\%, 85\%)$ respectively.

- The normal use condition corresponds to temperature at $25^{\circ}C$ and relative humidity at 50%.

The methodology proposed in this work is used to find the optimal design(s) that minimize the posterior variance of the 0.1 quantile of the lifetime distribution of the optical media. That is, the design(s) with the minimum C-criterion value. Specifically, the methodology is used to suggest the testing locations (combination of levels of temperature and relative humidity) and the number of units to be tested at each testing location.

For this example, the GA takes an average of 10 seconds for 1 generation for a starting population of 100 candidates. After running up to 6000 generations of the GA, the design with the minimum posterior variance of the 0.1 quantile lifetime was found at criterion value 0.4646. The number of generations for the algorithm was further increased up to 20000. No other better solution is found with smaller criterion value. The selected optimal design has four unique testing locations. Figure 5.3 shows the allocation of the testing locations of the optimal test plan in the input space of the accelerating factors. We can see that a majority of the test units (over 90%) are located at the bottom left corner of the input space corresponding to the lowest level of temperature and relative humidity ($65^{\circ}C$ and 65% RH), and the remaining points (less than 1%) are distributed around the corners of the input space. This is intuitive because the lower stress conditions are close to the normal use condition and hence provide more useful information towards the estimation of the quantile lifetime of interest. Hence the design that minimizes the posterior variance of the failure-time quantile at the normal use condition should have more test units allocated to the lowest levels of the accelerating factors if those test conditions are practically attainable.
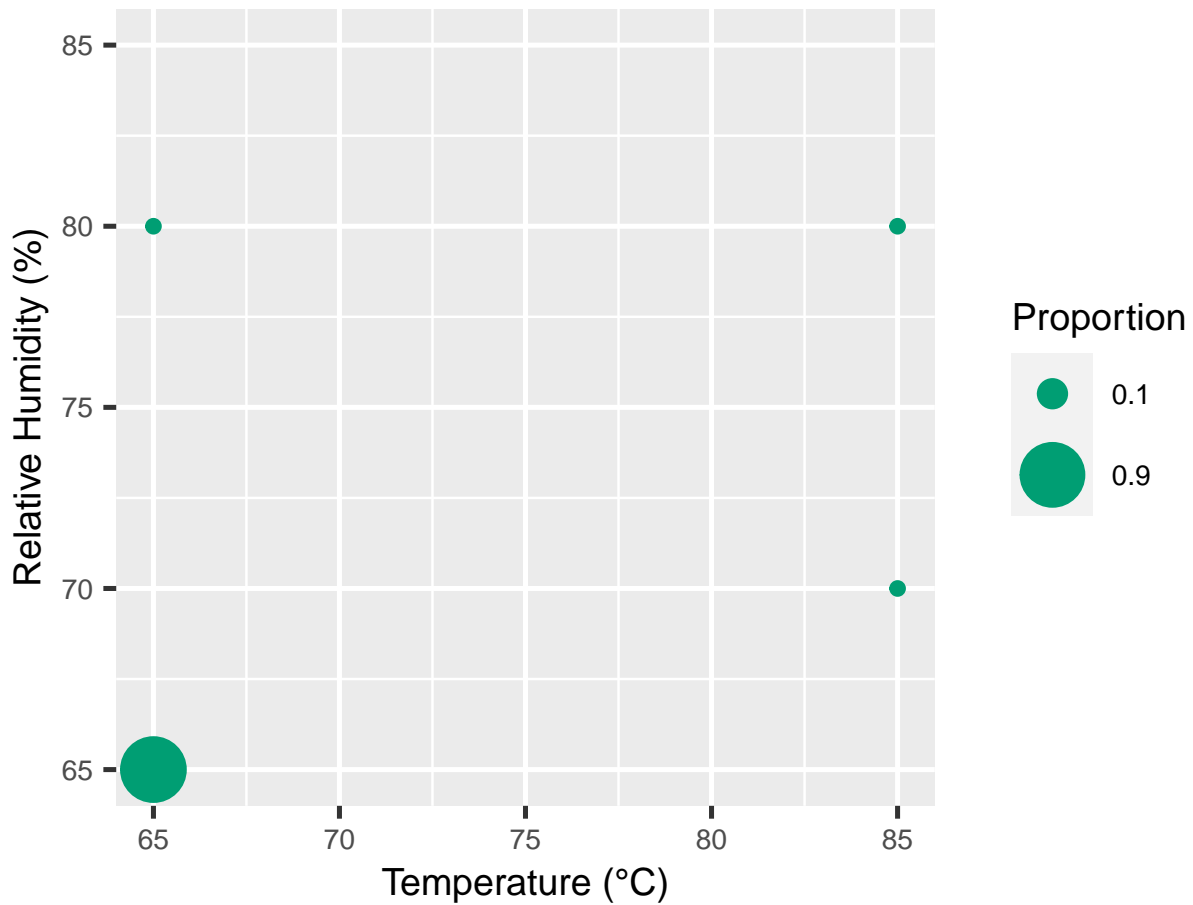
Figure 5.3: Optimal design

To compare the selected optimal design with the ISO design (refer to Table 5.1 and Figure 5.1), we estimated the C-criterion for the ISO design. The estimated C-criterion for the ISO design is 0.7307, a value much higher than that of the optimal design obtained via the proposed methodology in this work. That is, the posterior variance of the 0.1 quantile lifetime distribution at the normal use condition for the optimal design is smaller than that of the ISO design, implying that the selected optimal design allows more precise estimation of reliability at the normal use condition than the ISO design.

It should be noted that this design is optimal with respect to, the specified model, the design size and the aim of the experiment, which is to get the most precise estimate of the reliability at the normal use condition (quantified by the C-criterion). The C-criterion

selects points that are closer to the normal use condition to ensure that the estimation of reliability at the normal use condition is more precise. The choice of the optimal design will change if the goal of the experiment changes. For instance, if the aim is to obtain precise estimation of the model parameters, then the D-criterion is more appropriate metric for quantifying the test plan performance on that aspect. However, the proposed methodology in general is to be applied for other applications with different objectives, resource constraints or model forms. It can also be adapted for considering multiple criteria (e.g. simultaneously considering both D- and C- criteria) by utilizing the Pareto front approach described in earlier Chapters.

5.5 Conclusion

Using design of experiment technique to identify the optimal testing conditions and the allocations of the test units given a fixed budget in ADT would avail experimenters a statistical and objective approach to selecting the best test plan. We presented a Bayesian design of experiments method for ADT with two accelerating factors and repeated measurements. A linear mixed-effects model is assumed for the degradation paths based on the observed real test data. We seek optimal test plans that minimizes the posterior variance of the distribution of the failure-time quantile at the normal use condition. A large-sample approximation approach is used to estimate the posterior distribution in the Bayesian analysis.

To illustrate the methodology in this work, we consider an ADT example for an optical media with a total test units of 90 from ISO [101]. A customized genetic algorithm was implemented to search through the design space of the accelerating factors for the design with the minimum posterior variance of the failure-time quantile. The methodology presented in this work can be extended to a broader class of models such as generalized linear regression model or nonlinear regression models. One limitation of the current work is to assume the inspection time for the repeated measurements as fixed. This has

potential to reduce the test plan efficiency as different test conditions could result in different information gain (more failures would be observed under higher stress levels while few failures would result in lower stress conditions). One of the main area of the future work is to include inspection time as the test plan parameters to seek optimal strategy for test units allocations as well as the test scheduling. In addition, the following areas will also be considered in the future work:

- Sensitivity analysis on the impact of the choices of prior distributions on selected test plans.

- Adapt the methodology for broader types of models.

- Consider multiple criteria for ADT test plans.

Chapter 6: General Conclusion

Many present and emerging experimental situations, especially in engineering and environmental research now involve estimating multiple responses or achieving multiple design objectives. To add to the growing research in optimal design of experiments, this dissertation proposed new design selection strategies for generating optimal designs that are able to achieve balance in the estimation of multiple responses and optimization of multiple objectives respectively. It also contributes to the use of modern design of experiment techniques in the field of reliability analysis.

In Chapter 3, we started by discussing how to simultaneously obtain precision in the estimation of the responses in a multi-response experiment. By leveraging on information from screening experiment about subset of the design factors that actively influenced each response, we proposed a design methodology that uses the Pareto front approach to seek optimal designs based on D-efficiencies of the responses. A Pareto aggregate coordinate exchange algorithm was developed to efficiently search through the design space for optimal designs. Two screening experiments were used to demonstrate the proposed methodology. The D-optimal designs produced via the proposed methodology were found to perform better in terms of precise estimation of the responses when compared to existing classical designs.

Chapter 4 focus is on generating optimal designs for experiments with multiple objectives. Specifically, we demonstrate how to use the Pareto front approach to simultaneously optimize multiple space-filling characteristics in the construction of Latin hypercube designs for design of computer experiments. A column-wise exchange simulated annealing algorithm and a nondominated sorting genetic algorithm were used in search-

ing through the design space for optimal solutions. The nondominated sorting genetic algorithm was found to be better at populating the Pareto front. The Chapter also discuss how the Utopia point method is used in comparing the trade-offs between the criteria for competing designs found on the Pareto front and how it is strategically used for making selection from a Pareto front with large solutions. The analysis in this work revealed that, LHDs with optimal and balanced performance in multiple criteria have better statistical properties than LHDs that are optimal in single criterion.

Chapter 5 contains the final piece of work of this dissertation. This chapter presents a novel application of Bayesian design of experiments method to finding the optimal test plan for accelerated degradation test with multiple accelerating factors, where the goal is minimizing the posterior variance of the quantile of the failure-time distribution at the use condition. A customized genetic algorithm is developed for optimizing the Bayesian criterion function and searching through the design space for the optimal design. An examples based on the ISO optical media degradation data was used to demonstrate the proposed methodology. The result in this chapter shows that the selected optimal design via the proposed methodology allows more precise estimation of reliability at the normal use condition than the ISO design [101].

References

[1] Maria Adamou. *Bayesian optimal designs for the Gaussian Process Model*. PhD thesis, University of Southampton, 2014.

[2] Damola M. Akinlana and Lu Lu. Design optimization based on d-optimality for multiple responses. In *JSM Proceedings, Quality and Productivity Section*. American Statistical Association, 2020.

[3] Damola M. Akinlana and Lu Lu. Multiple objective latin hypercube designs for computer experiments. In *JSM Proceedings, Quality and Productivity Section*. American Statistical Association, 2021.

[4] Anthony Atkinson, Alexander Donev, and Randall Tobias. *Optimum experimental designs, with SAS*, volume 34. OUP Oxford, 2007.

[5] John W Bandler, Qingsha S Cheng, Sameh A Dakroury, Ahmed S Mohamed, Mohamed H Bakr, Kaj Madsen, and Jacob Sondergaard. Space mapping: the state of the art. *IEEE Transactions on Microwave theory and techniques*, 52(1):337–361, 2004.

[6] Sarwar Beg. *Design of Experiments for Pharmaceutical Product Development*. Springer, 2021.

[7] George EP Box and Kenneth B Wilson. On the experimental attainment of optimum conditions. In *Breakthroughs in statistics*, pages 270–310. Springer, 1992.

[8] GEP Box and DW Behnken. Simplex-sum designs: a class of second order rotatable designs derivable from those of first order. *The Annals of Mathematical Statistics*, 31(4):838–864, 1960.

[9] Dizza Bursztyn and David M Steinberg. Comparison of designs for computer experiments. *Journal of statistical planning and inference*, 136(3):1103–1119, 2006.

[10] Yongtao Cao, Byran J Smucker, and Timothy J Robinson. A hybrid elitist pareto-based coordinate exchange algorithm for constructing multi-criteria optimal experimental designs. *Statistics and Computing*, 27(2):423–437, 2017.

[11] Kathryn Chaloner and Kinley Larntz. Optimal bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21(2):191–208, 1989.

[12] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.

[13] S Chang. An algorithm to generate near d-optimal designs for multiple response surface models. *IIE transactions*, 29(12):1073–1081, 1997.

[14] Jessica L Chapman, Lu Lu, and Christine M Anderson-Cook. Process optimization for multiple responses utilizing the pareto front approach. *Quality Engineering*, 26(3):253–268, 2014.

[15] Jessica L Chapman, Lu Lu, and Christine M Anderson-Cook. Using multiple criteria optimization and two-stage genetic algorithms to select a population management strategy with optimized reliability. *Complexity*, 2018, 2018.

[16] Merlise A Clyde. Experimental design: A bayesian perspective. *International Encyclopia Social and Behavioral Sciences*, 8:5075–5081, 2001.

[17] R Dennis Cook and Christopher J Nachtrheim. A comparison of algorithms for constructing exact d-optimal designs. *Technometrics*, 22(3):315–324, 1980.

[18] Manel Cooray-Wijesinha and Andrew I Khuri. The sequential generation of multiresponse d-optimal designs when the variance-covariance matrix is not known:

The sequential generation. *Communications in Statistics-Simulation and Computation*, 16(1):239–259, 1987.

[19] Daniel Palhazi Cuervo, Peter Goos, Kenneth Sörensen, and Emely Arráiz. An iterated local search algorithm for the vehicle routing problem with backhauls. *European Journal of Operational Research*, 237(2):454–464, 2014.

[20] Anirban DasGupta. 29 review of optimal bayes designs. *Handbook of Statistics*, 13:1099–1147, 1996.

[21] Kenneth De Jong. Evolutionary computation: a unified approach. In *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*, pages 185–199, 2016.

[22] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

[23] George Derringer and Ronald Suich. Simultaneous optimization of several response variables. *Journal of quality technology*, 12(4):214–219, 1980.

[24] Norman R Draper and Irwin Guttman. Treating bias as variance for experimental design purposes. *Annals of the Institute of Statistical Mathematics*, 44(4):659–671, 1992.

[25] Christopher C Drovandi, James M McGree, and Anthony N Pettitt. A sequential monte carlo algorithm to incorporate model uncertainty in bayesian sequential design. *Journal of Computational and Graphical Statistics*, 23(1):3–24, 2014.

[26] Olga Egorova. *Optimal design of experiments for multiple objectives*. PhD thesis, University of Southampton, 2017.

[27] Agoston E Eiben, James E Smith, et al. *Introduction to evolutionary computing*, volume 53. Springer, 2003.

[28] Lennart Eriksson, Erik Johansson, and Conny Wikström. Mixture design—design generation, pls analysis, and model usage. *Chemometrics and Intelligent Laboratory Systems*, 43(1-2):1–24, 1998.

[29] Hamidreza Eskandari, Christopher D Geiger, and Gary B Lamont. Fastpga: A dynamic population sizing approach for solving expensive multiobjective optimization problems. In *International Conference on Evolutionary Multi-Criterion Optimization*, pages 141–155. Springer, 2007.

[30] Guanqi Fang, Steven E Rigdon, and Rong Pan. Predicting lifetime by degradation tests: A case study of iso 10995. *Quality and Reliability Engineering International*, 34(6):1228–1237, 2018.

[31] Kai-Tai Fang, Runze Li, and Agus Sudjianto. *Design and modeling for computer experiments*. Chapman and Hall/CRC, 2005.

[32] VV Fedorov. *Theory of optimal experiments*. Academic Press, 1972.

[33] SL Costa Ferreira, RE Bruns, Hadla Sousa Ferreira, Geraldo Domingues Matos, JM David, GC Brandão, EG Paranhos da Silva, LA Portugal, PS Dos Reis, AS Souza, et al. Box-behnken design: an alternative for the optimization of analytical methods. *Analytica chimica acta*, 597(2):179–186, 2007.

[34] Joan Fisher Box. Ra fisher: The life of a scientist. *NY, John Wiley & sons. MR*, 500579, 1978.

[35] Andrew James Friend. *Development of a co-firing fuel from biomass-derived binder and crushed coal*. PhD thesis, Iowa State University, 2013.

[36] Andrew Gelman, Daniel Lee, and Jiqiang Guo. Stan: A probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, 40(5):530–543, 2015.

[37] Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

[38] Wolfram Gronwald, Tim Hohm, and Daniel Hoffmann. Evolutionary pareto-optimization of stably folding peptides. *BMC bioinformatics*, 9(1):1–13, 2008.

[39] Michael S Hamada, Harry F Martz, C Shane Reese, and Alyson G Wilson. *Bayesian reliability*, volume 15. Springer, 2008.

[40] Cong Han and Kathryn Chaloner. Bayesian experimental design for nonlinear mixed-effects models with application to hiv dynamics. *Biometrics*, 60(1):25–33, 2004.

[41] Edwin C Harrington. The desirability function. *Industrial quality control*, 21(10):494–498, 1965.

[42] Alejandro Heredia-Langner, W Matthew Carlyle, Douglas C Montgomery, Connie M Borror, and George C Runger. Genetic algorithms for the construction of d-optimal designs. *Journal of quality technology*, 35(1):28–46, 2003.

[43] Matthew D Hoffman, Andrew Gelman, et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

[44] Ruizhe Hou. *Optimal Latin Hypercube Designs for Computer Experiments Based on Multiple Objectives*. University of South Florida, 2018.

[45] Xun Huan and Youssef Marzouk. Gradient-based stochastic optimization methods in bayesian experimental design. *International Journal for Uncertainty Quantification*, 4(6), 2014.

[46] Robert I Jennrich and Mark D Schluchter. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, pages 805–820, 1986.

[47] Ruichen Jin, Wei Chen, and Agus Sudjianto. An efficient algorithm for constructing optimal design of computer experiments. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 37009, pages 545–554, 2003.

[48] Mark E Johnson, Leslie M Moore, and Donald Ylvisaker. Minimax and maximin distance designs. *Journal of statistical planning and inference*, 26(2):131–148, 1990.

[49] Rachel T Johnson, Douglas C Montgomery, and Bradley A Jones. An expository paper on optimal design. *Quality Engineering*, 23(3):287–301, 2011.

[50] Bradley Jones and Christopher J Nachtsheim. A class of three-level designs for definitive screening in the presence of second-order effects. *Journal of Quality Technology*, 43(1):1–15, 2011.

[51] KAD Jong. Evolutionary computing: a unified approach. *Cambridge, MA*, 2142, 2006.

[52] V Roshan Joseph. Space-filling designs for computer experiments: A review. *Quality Engineering*, 28(1):28–35, 2016.

[53] V Roshan Joseph, Evren Gul, and Shan Ba. Maximum projection designs for computer experiments. *Biometrika*, 102(2):371–380, 2015.

[54] Q Ye Kenny, William Li, and Agus Sudjianto. Algorithmic construction of optimal symmetric latin hypercube designs. *Journal of statistical planning and inference*, 90(1):145–159, 2000.

[55] J Kiefer. Optimum designs in regression problems, ii. *The Annals of Mathematical Statistics*, pages 298–325, 1961.

[56] Jack Kiefer and Jacob Wolfowitz. Optimum designs in regression problems. *The annals of mathematical statistics*, 30(2):271–294, 1959.

[57] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[58] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[59] Robert Michael Lewis, Virginia Torczon, and Michael W Trosset. Direct search methods: then and now. *Journal of computational and Applied Mathematics*, 124(1-2):191–207, 2000.

[60] Xiaoyang Li, Yuqing Hu, Enrico Zio, and Rui Kang. A bayesian optimal design for accelerated degradation testing based on the inverse gaussian process. *IEEE Access*, 5:5690–5701, 2017.

[61] Mattias Liefvendahl and Rafał Stocki. A study on algorithms for optimization of latin hypercubes. *Journal of statistical planning and inference*, 136(9):3231–3247, 2006.

[62] Heonsang Lim and Bong-Jin Yum. Optimal design of accelerated degradation tests based on wiener process models. *Journal of Applied Statistics*, 38(2):309–325, 2011.

[63] Dennis Victor Lindley. *Bayesian statistics: A review*. SIAM, 1972.

[64] C Joseph Lu and William O Meeker. Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35(2):161–174, 1993.

[65] Lu Lu and Christine M Anderson-Cook. Rethinking the optimal response surface design for a first-order model with two-factor interactions, when protecting against curvature. *Quality Engineering*, 24(3):404–422, 2012.

[66] Lu Lu, Christine M Anderson-Cook, Miranda Martin, and Towfiq Ahmed. Practical choices for space-filling designs. *Quality and Reliability Engineering International*, 38(3):1165–1188, 2022.

[67] Lu Lu, Christine M Anderson-Cook, and Timothy J Robinson. Optimization of designed experiments based on multiple criteria utilizing a pareto frontier. *Technometrics*, 53(4):353–365, 2011.

[68] Lu Lu, Christine M Anderson-Cook, and Timothy J Robinson. A case study to demonstrate a pareto frontier for selecting a best response surface design while simultaneously optimizing multiple criteria. *Applied Stochastic Models in Business and Industry*, 28(3):206–221, 2012.

[69] Lu Lu, Jessica L Chapman, and Christine M Anderson-Cook. A case study on selecting a best allocation of new data for improving the estimation precision of system and subsystem reliability using pareto fronts. *Technometrics*, 55(4):473–487, 2013.

[70] Wilmina M Marget and Max D Morris. Central composite experimental designs for multiple responses with different models. *Technometrics*, 2019.

[71] R Timothy Marler and Jasbir S Arora. Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*, 26(6):369–395, 2004.

[72] James McGree, Christopher C Drovandi, and Anthony N Pettitt. A sequential monte carlo approach to the sequential design for discriminating between rival continuous data models. *Queensland University of Technology*, 2012.

[73] MD McKay, RJ Beckman, and WJ Conover. Acomparisonof three methodsforselecting valuesofinputvariablesinthe analysisofoutputfrom acomputercode. *Technometrics*, 21(2):239–245, 1979.

[74] William Q Meeker, Luis A Escobar, and Francis G Pascual. *Statistical methods for reliability data*. John Wiley & Sons, 2022.

[75] Ruth K Meyer and Christopher J Nachtsheim. Constructing exact d-optimal experimental designs by simulated annealing. *American Journal of Mathematical and Management Sciences*, 8(3-4):329–359, 1988.

[76] Ruth K Meyer and Christopher J Nachtsheim. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37(1):60–69, 1995.

[77] Melanie Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.

[78] Douglas C Montgomery. *Design and analysis of experiments*. John wiley & sons, 2017.

[79] Max D Morris and Toby J Mitchell. Exploratory designs for computational experiments. *Journal of statistical planning and inference*, 43(3):381–402, 1995.

[80] Peter Müller, Don A Berry, Andrew P Grieve, and Michael Krams. A bayesian decision-theoretic dose-finding trial. *Decision analysis*, 3(4):197–207, 2006.

[81] Raymond H Myers, Douglas C Montgomery, and Christine M Anderson-Cook. *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons, 2016.

[82] Anamai Na-udom and Jaratsri Rungrattanaubol. Heuristic search algorithms for constructing optimal latin hypercube designs. In *Recent Advances in Information and Communication Technology 2016*, pages 183–193. Springer, 2016.

[83] Wayne B Nelson. *Accelerated testing: statistical models, test plans, and data analysis*. John Wiley & Sons, 2009.

[84] Nam-Ky Nguyen and Alan J Miller. A review of some exchange algorithms for constructing discrete d-optimal designs. *Computational Statistics & Data Analysis*, 14(4):489–498, 1992.

[85] David E Nikles and John M Wiest. Accelerated aging studies and the prediction of the archival lifetime of optical disk media. In *Recent Advances in Metrology, Characterization, and Standards for Optical Digital Data Disks*, volume 3806, pages 30–34. SPIE, 1999.

[86] Antony M Overstall and David C Woods. Bayesian design of experiments using approximate coordinate exchange. *Technometrics*, 59(4):458–470, 2017.

[87] Rong Pan and Taeho Crispin. A hierarchical modeling approach to accelerated degradation testing data analysis: A case study. *Quality and reliability engineering international*, 27(2):229–237, 2011.

[88] Chanseok Park and William J Padgett. Stochastic degradation models with several accelerating variables. *IEEE Transactions on Reliability*, 55(2):379–390, 2006.

[89] You-Jin Park. Multi-optimal designs for second-order response surface models. *Communications for Statistical Applications and Methods*, 16(1):195–208, 2009.

[90] NG Polson. A bayesian perspective on the design of accelerated life tests. *Advances in Reliability*, pages 321–330, 1993.

[91] EA Rady, MME Abd El-Monsef, and MM Seyam. Relationships among several optimality criteria. *Interstat*, 15(6):1–11, 2009.

[92] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

[93] Elizabeth G Ryan. *Contributions to Bayesian experimental design*. PhD thesis, Queensland University of Technology, 2014.

[94] Jerome Sacks, William J Welch, Toby J Mitchell, and Henry P Wynn. Design and analysis of computer experiments. *Statistical science*, 4(4):409–423, 1989.

[95] Francesco Sambo, Matteo Borrotti, and Kalliopi Mylona. A coordinate-exchange two-phase local search algorithm for the d-and i-optimal designs of split-plot experiments. *Computational Statistics & Data Analysis*, 71:1193–1207, 2014.

[96] Thomas J Santner, Brian J Williams, William I Notz, and Brain J Williams. *The design and analysis of computer experiments*, volume 1. Springer, 2003.

[97] Ying Shi and William Q Meeker. Bayesian methods for accelerated destructive degradation test planning. *IEEE Transactions on Reliability*, 61(1):245–253, 2011.

[98] Karthik Sindhya, Kaisa Miettinen, and Kalyanmoy Deb. A hybrid framework for evolutionary multi-objective optimization. *IEEE Transactions on Evolutionary Computation*, 17(4):495–511, 2012.

[99] Kirstine Smith. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12(1/2):1–85, 1918.

[100] Antti Solonen, Heikki Haario, and Marko Laine. Simulation-based optimal design using a response variance criterion. *Journal of Computational and Graphical Statistics*, 21(1):234–252, 2012.

[101] Organization International Standards. Information technology—digitally recorded media for information interchange and storage—test method for the estimation of the archival lifetime of optical media. 2011.

[102] Loon Ching Tang and Xiao Liu. Planning and inference for a sequential accelerated life test. *Journal of quality technology*, 42(1):103–118, 2010.

[103] Stan Development Team et al. Rstan: the r interface to stan. *R package version*, 2(1):522, 2016.

[104] Heike Trautmann and Jörn Mehnen. Preference-based pareto optimization in certain and noisy environments. *Engineering Optimization*, 41(1):23–38, 2009.

[105] Sheng-Tsaing Tseng and I-Chen Lee. Optimum allocation rule for accelerated degradation tests with a class of exponential-dispersion degradation models. *Technometrics*, 58(2):244–254, 2016.

[106] Edwin R van Dam. Two-dimensional minimax latin hypercube designs. *Discrete Applied Mathematics*, 156(18):3483–3493, 2008.

[107] Isabella Verdinelli. Miscellanea. a note on bayesian design for the normal linear model with unknown error variance. *Biometrika*, 87(1):222–227, 2000.

[108] Felipe AC Viana. A tutorial on latin hypercube design of experiments. *Quality and reliability engineering international*, 32(5):1975–1985, 2016.

[109] Dongshu Wang, Dapei Tan, and Lei Liu. Particle swarm optimization algorithm: an overview. *Soft Computing*, 22(2):387–408, 2018.

[110] Brian P Weaver and William Q Meeker. Methods for planning repeated measures accelerated degradation tests. *Applied Stochastic Models in Business and Industry*, 30(6):658–671, 2014.

[111] Brian P Weaver and William Q Meeker. Bayesian methods for planning accelerated repeated measures degradation tests. *Technometrics*, 63(1):90–99, 2021.

[112] Brian Phillip Weaver. *Methods for planning repeated measures degradation tests*. Iowa State University, 2011.

[113] Henry P Wynn. Results in the theory and construction of d-optimum experimental designs. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):133–147, 1972.

[114] Min Yang. A-optimal designs for generalized linear models with two parameters. *Journal of Statistical Planning and Inference*, 138(3):624–641, 2008.

[115] Frank Yates. Sir ronald fisher and the design of experiments. *Biometrics*, 20(2):307–321, 1964.

[116] Zhi-Sheng Ye, Liang-Peng Chen, Loon Ching Tang, and Min Xie. Accelerated degradation test planning using the inverse gaussian process. *IEEE Transactions on Reliability*, 63(3):750–763, 2014.

[117] Hong-Fwu Yu. Designing an accelerated degradation experiment by optimizing the estimation of the percentile. *Quality and Reliability Engineering International*, 19(3):197–214, 2003.

[118] Yao Zhang and William Q Meeker. Bayesian methods for planning accelerated life tests. *Technometrics*, 48(1):49–60, 2006.

Appendix A: Structure of Some of the Designs from Chapter 3

A.1 Design Matrix for Res IV UF-CCD

[1,]	1	1	1	1	1
[2,]	1	1	-1	-1	1
[3,]	1	-1	1	-1	-1
[4,]	1	-1	-1	1	-1
[5,]	-1	1	1	-1	1
[6,]	-1	1	-1	1	1
[7,]	-1	-1	1	1	-1
[8,]	-1	-1	-1	-1	-1
[9,]	2	0	0	0	0
[10,]	-2	0	0	0	0
[11,]	0	2	0	0	2
[12,]	0	-2	0	0	-2
[13,]	0	0	2	0	0
[14,]	0	0	-2	0	0
[15,]	0	0	0	2	0
[16,]	0	0	0	-2	0
[17,]	0	0	0	0	0
[18,]	0	0	0	0	0
[19,]	0	0	0	0	0
[20,]	0	0	0	0	0

Figure A.1: Res IV UF-CCD design, axial points takes value at ± 2

A.2 Visual Display of the D-optimal Design and Res IV UF-CCD
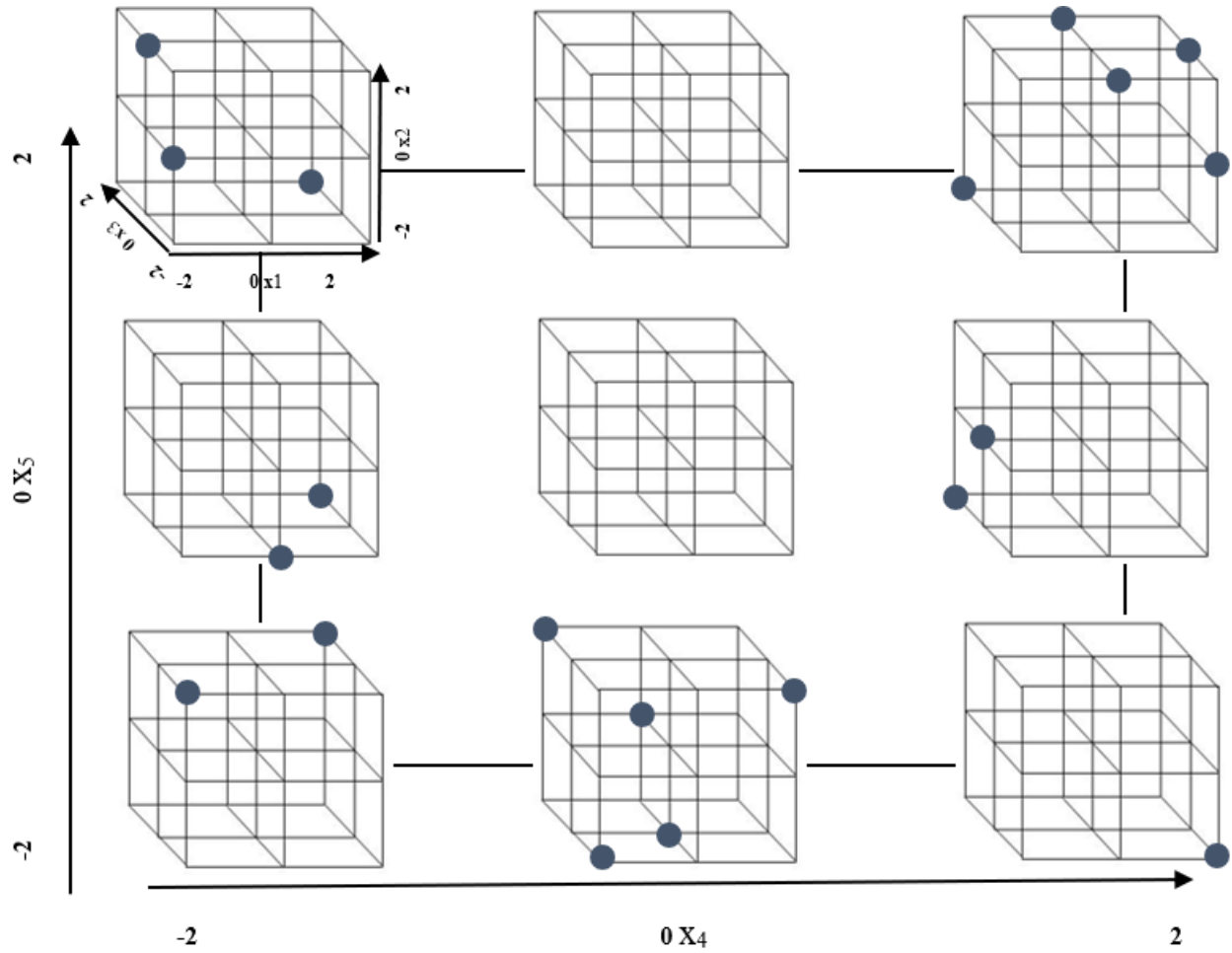
Figure A.2: Geometric representation of the D-optimal design generated for the screening experiment with five factors
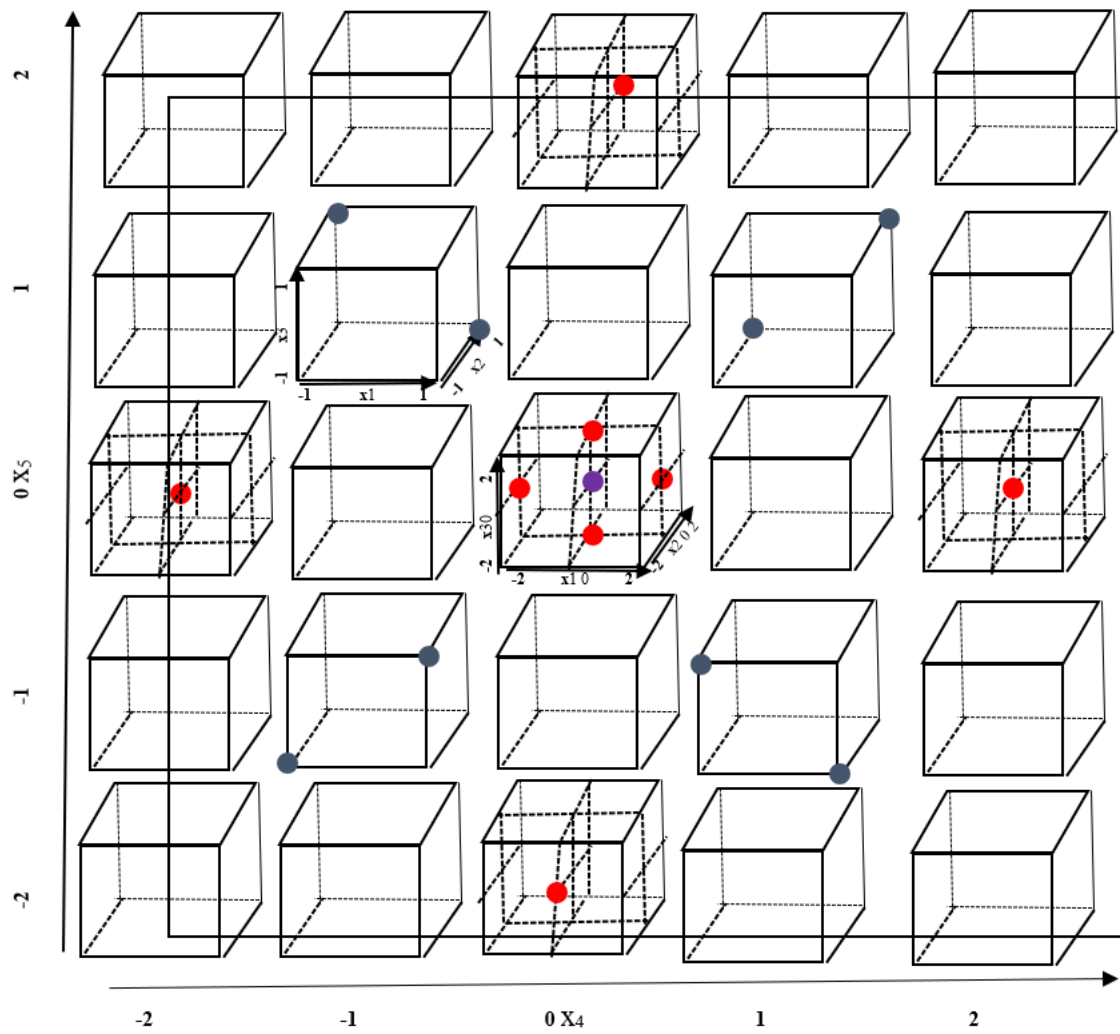
Figure A.3: Geometric representation of the Res IV UF-CCD. The blue, red and purple circles signifies the factorial points, axial points and center points respectively

A.3 Design Matrix for Res V UF-CCD

[1,]	-1	-1	-1	-1	1	-1	-1	-1	-1	-1
[2,]	1	-1	-1	-1	-1	1	1	1	1	-1
[3,]	-1	1	-1	-1	-1	-1	1	-1	1	-1
[4,]	1	1	-1	-1	1	1	-1	1	-1	-1
[5,]	-1	-1	1	-1	-1	-1	-1	-1	-1	1
[6,]	1	-1	1	-1	1	1	1	1	1	1
[7,]	-1	1	1	-1	1	-1	1	-1	1	1
[8,]	1	1	1	-1	-1	1	0	1	0	1
[9,]	-1	-1	-1	1	-1	-1	0	-1	0	-1
[10,]	1	-1	-1	1	1	1	2	1	2	-1
[11,]	-1	1	-1	1	1	-1	-2	-1	-2	-1
[12,]	1	1	-1	1	-1	1	0	1	0	-1
[13,]	-1	-1	1	1	1	-1	0	-1	0	1
[14,]	1	-1	1	1	-1	1	0	1	0	1
[15,]	-1	1	1	1	-1	-1	0	-1	0	1
[16,]	1	1	1	1	1	1	0	1	0	1
[17,]	2	0	0	0	0	2	0	2	0	0
[18,]	-2	0	0	0	0	-2	0	-2	0	0
[19,]	0	2	0	0	0	0	0	0	0	0
[20,]	0	-2	0	0	0	0	0	0	0	0
[21,]	0	0	2	0	0	0	0	0	0	2
[22,]	0	0	-2	0	0	0	-1	0	-1	-2
[23,]	0	0	0	2	0	0	1	0	1	0
[24,]	0	0	0	-2	0	0	1	0	1	0
[25,]	0	0	0	0	2	0	-1	0	-1	0
[26,]	0	0	0	0	-2	0	-1	0	-1	0
[27,]	0	0	0	0	0	0	1	0	1	0
[28,]	0	0	0	0	0	0	1	0	1	0
[29,]	0	0	0	0	0	0	0	0	0	0
[30,]	0	0	0	0	0	0	0	0	0	0

Figure A.4: Res V UF-CCD design, axial points takes value at ± 2

112

Appendix B: Derivations in Chapter 5 and the ISO Optical Media Data

B.1 Derivation of the Fisher Information Matrix

The model in Equation 5.2 for response variable for unit i under time point j and test-ing location k will be used for the derivation of the Fisher information matrix. Given that there are n items to be tested and each item will be measured at time points $\tau_1, ..., \tau_m$ for testing location k of the levels of the accelerating factors. Collecting into $Y_i = (Y_{i1k}, ..., Y_{imk})$, the observations from unit i, the linear degradation model in Equation 5.2 can also be ex-pressed as

$$Y_i = \mathbf{X}_i \gamma + \mathbf{Z}_i \mathbf{b}_i + \epsilon_i \quad i = 1, ..., n, \quad j = 1, ..., m, \quad k = 1, ..., K$$

where $\gamma = (\gamma_1, \gamma_2), (b_0, b_1)^T \sim BVN(\beta, \mathbf{V})$, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are matrices of explanatory vari-ables.

$$\mathbf{X}_i = \begin{pmatrix} \tau_1 x_{1k} & \tau_1 x_{2k} \\ \vdots & \vdots \\ \tau_m x_{1k} & \tau_m x_{2k} \end{pmatrix}, \mathbf{Z}_i = \begin{pmatrix} 1 & \tau_1 \\ \vdots & \vdots \\ 1 & \tau_m \end{pmatrix},$$

and $\epsilon_i = (\epsilon_{i1}, ..., \epsilon_{im})^T$.

Assuming ϵ_i are independently and jointly normally distributed, that is $\epsilon_i \sim MVN(0, \sigma^2 \mathbf{I}_i)$ (provided the intervals between taking measurements is not too small), where $\mathbf{I}_i$ is an $m_i \times m_i$ identity matrix. Assuming also that there is an independence between ϵ_i and $\mathbf{b}_i$,

then it follows that $Y_i \sim MVN(\mathbf{X}_i\gamma + \mathbf{Z}_i\boldsymbol{\beta}_i, \Sigma_i)$ where

$$\Sigma_i = Var(\mathbf{X}_i\gamma + \mathbf{Z}_i\mathbf{b}_i) = \mathbf{Z}_i\mathbf{V}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_i. \tag{B.1}$$

The log-likelihood for observational unit i is

$$\mathcal{L}_i = -\frac{1}{2}log\left[det(\Sigma_i)\right] - \frac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i\gamma - \mathbf{Z}_i\boldsymbol{\beta})^T\Sigma_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\gamma - \mathbf{Z}_i\boldsymbol{\beta})$$

The total log-likelihood for n units is

$$\mathcal{L} = \sum_{i=1}^{n}\mathcal{L}_i = -\frac{1}{2}\sum_{i=1}^{n}log\left[det(\Sigma_i)\right] - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{Y}_i - \mathbf{X}_i\gamma - \mathbf{Z}_i\boldsymbol{\beta})^T\Sigma_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\gamma - \mathbf{Z}_i\boldsymbol{\beta})$$

To simplify notification, let $\boldsymbol{\beta}^* = (\gamma, \boldsymbol{\beta})$ be collection of the fixed effects model parameters and $\boldsymbol{\vartheta} = (\sigma_{b_0}^2, \sigma_{b_1}^2, \rho, \sigma^2)$, the model variance components. Using equation (4) of Jenrich and Schluchter [46], the Hessian matrix, $\mathbf{H}_i$, for unit i and its expected value are given by

$$\mathbf{H}_i = \begin{pmatrix} \mathbf{H}_{\boldsymbol{\beta}^*\boldsymbol{\beta}^*,i} & \mathbf{H}_{\boldsymbol{\beta}^*\boldsymbol{\vartheta},i} \\ \mathbf{H}_{\boldsymbol{\vartheta}\boldsymbol{\beta}^*,i} & \mathbf{H}_{\boldsymbol{\vartheta}\boldsymbol{\vartheta},i} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2\mathcal{L}_i}{\partial\boldsymbol{\beta}^*\partial\boldsymbol{\beta}^*} & \frac{\partial^2\mathcal{L}_i}{\partial\boldsymbol{\beta}^*\partial\boldsymbol{\vartheta}} \\ \frac{\partial^2\mathcal{L}_i}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\beta}^*} & \frac{\partial^2\mathcal{L}_i}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}} \end{pmatrix}$$

$$E(\mathbf{H}_i) = \mathbf{I}_i(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{R}_i^T\Sigma_i^{-1}\mathbf{R}_i & 0 \\ 0 & \mathbf{M}_i \end{pmatrix} \tag{B.2}$$

Where $\mathbf{R}_i = (\mathbf{X}_i, \mathbf{Z}_i)$, $\mathbf{M}_i$ is a 4×4 symmetric matrix with elements

$$M_{jk}^i = \frac{1}{2}tr(\Sigma_i^{-1}\dot{\Sigma}_{ij}\Sigma_i^{-1}\dot{\Sigma}_{ik}), \ j = 1, ..., 4; \ k = 1, ..., 4,$$

and

$$\dot{\Sigma}_{ij} = \frac{\partial\Sigma_i}{\partial\vartheta_j}, \ j = 1, ..., 4.$$

From Equation B.1, it follows that

$$\dot{\Sigma}_{i1} = \frac{\partial \Sigma_i}{\partial \sigma_{b_0}} = \mathbf{Z}_i \begin{pmatrix} 2\sigma_{b_0} & \rho\sigma_{b_1} \\ \rho\sigma_{b_1} & 0 \end{pmatrix} \mathbf{Z}_i'$$

$$\dot{\Sigma}_{i2} = \frac{\partial \Sigma_i}{\partial \sigma_{b_1}} = \mathbf{Z}_i \begin{pmatrix} 0 & \rho\sigma_{b_0} \\ \rho\sigma_{b_0} & 2\sigma_{b_1} \end{pmatrix} \mathbf{Z}_i'$$

$$\dot{\Sigma}_{i3} = \frac{\partial \Sigma_i}{\partial \rho} = \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_{b_1}\sigma_{b_0} \\ \sigma_{b_1}\sigma_{b_0} & 0 \end{pmatrix} \mathbf{Z}_i'$$

$$\dot{\Sigma}_{i4} = \frac{\partial \Sigma_i}{\partial \sigma} = 2\sigma\mathbf{I}_i.$$

Therefore, the information matrix for all n units is given by

$$\mathbf{I}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{I}_i(\boldsymbol{\theta}).$$

B.2 Derivation of τ_p

Given that a failure occurs at the first point in time where $\mathcal{D} > \mathcal{D}_f$. Let F denote the cumulative distribution function of the random variable T and let $\sigma_{b_0 b_1} = \rho \sigma_{b_0} \sigma_{b_1}$. Then

$$F(t_p) = 1 - \Phi_{nor}\left(\frac{\mathcal{D}_f - x_1 \tau_p \gamma_1 - x_2 \tau_p \gamma_2 - \beta_0 - \beta_1 \tau_p}{\sqrt{\sigma_{b_0}^2 + \tau_p^2 \sigma_{b_1}^2 + 2\tau_p \sigma_{b_0 b_1}}}\right) = p$$

$$\Phi_{nor}\left(\frac{\mathcal{D}_f - \beta_0 - (\beta_1 + x_1 \gamma_1 + x_2 \gamma_2)\tau_p}{\sqrt{\sigma_{b_0}^2 + \tau_p^2 \sigma_{b_1}^2 + 2\tau_p \sigma_{b_0 b_1}}}\right) = 1 - p$$

$$\frac{\mathcal{D}_f - \beta_0 - (\beta_1 + x_1 \gamma_1 + x_2 \gamma_2)\tau_p}{\sqrt{\sigma_{b_0}^2 + \tau_p^2 \sigma_{b_1}^2 + 2\tau_p \sigma_{b_0 b_1}}} = \Phi_{nor}^{-1}(1 - p)$$

$$\frac{[\mathcal{D}_f - \beta_0 - (\beta_1 + x_1 \gamma_1 + x_2 \gamma_2)\tau_p]^2}{\sigma_{b_0}^2 + \tau_p^2 \sigma_{b_1}^2 + 2\tau_p \sigma_{b_0 b_1}} = [\Phi_{nor}^{-1}(1 - p)]^2$$

$$k(\sigma_{b_0}^2 + \tau_p^2 \sigma_{b_1}^2 + 2\tau_p \sigma_{b_0 b_1}) = (u - r\tau_p)^2$$

$$k\sigma_{b_1}^2 \tau_p^2 + 2\sigma_{b_0 b_1} k\tau_p + k\sigma_{b_0}^2 = u^2 - 2ur\tau_p + r^2 \tau_p^2$$

$$\tau_p^2(k\sigma_{b_1}^2 - r^2) + 2\tau_p(k\sigma_{b_0 b_1} + ur) + (k\sigma_{b_0}^2 - u^2) = 0$$

where $k = [\Phi_{nor}^{-1}(1 - p)]^2$, $u = \mathcal{D}_f - \beta_0$ and $r = \beta_1 + x_1 \gamma_1 + x_2 \gamma_2$.

let $a = k\sigma_{b_1}^2 - r^2$, $b = 2(k\sigma_{b_0 b_1} + ur)$ and $c = k\sigma_{b_0}^2 - u^2$. Then this equation is of the form

$$a\tau_p^2 + b\tau_p + c = 0$$

with solution for τ_p

$$\begin{aligned}
\tau_p &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\
&= \frac{-2(k\sigma_{b_0 b_1} + ur) \pm \sqrt{4(k\sigma_{b_0 b_1} + ur)^2 - 4(k\sigma_{b_1}^2 - r^2)(k\sigma_{b_0}^2 - u^2)}}{2(k\sigma_{b_1}^2 - r^2)} \\
&= \frac{-(k\sigma_{b_0 b_1} + ur) \pm \sqrt{(k\sigma_{b_0 b_1} + ur)^2 - (k\sigma_{b_1}^2 - r^2)(k\sigma_{b_0}^2 - u^2)}}{k\sigma_{b_1}^2 - r^2}.
\end{aligned}$$

When failure is defined as the first time at which $\mathcal{D} < \mathcal{D}_f$, the derivation is similar, starting with

$$F(t_p) = \Phi_{nor}\left(\frac{\mathcal{D}_f - x_1\tau_p\gamma_1 - x_2\tau_p\gamma_2 - \beta_0 - \beta_1\tau_p}{\sqrt{\sigma_{b_0}^2 + \tau_p^2\sigma_{b_1}^2 + 2\tau_p\sigma_{b_0 b_1}}}\right) = p.$$

B.3 Original Data Table of Optical Medial Error Rates Provided by ISO 10995

Table B.1: Test condition 1

TEMP=85°C, RH=85%

Disk	Hours					Proj. Failure
	0	250	500	750	1000	
A1	16	78	116	278	445	788
A2	25	64	134	342	532	743
A3	26	94	190	335	642	685
A4	26	111	247	343	718	647
A5	27	89	185	246	466	762
A6	21	111	207	567	896	607
A7	26	121	274	589	781	588
A8	31	108	223	315	745	654
A9	24	118	285	723	754	578
A10	12	85	178	312	988	669
A11	28	111	167	312	771	671
A12	24	136	267	444	719	614
A13	35	76	265	567	610	626
A14	19	53	112	278	534	778
A15	28	88	158	308	654	704
A16	27	68	120	263	432	807
A17	18	87	176	302	558	723
A18	26	109	238	421	641	645
A19	26	111	253	378	638	649
A20	31	91	206	367	728	656

Table B.2: Test condition 2

TEMP=85°C, RH=70%

Disk	Hours					Proj. Failure
	0	250	500	750	1000	
B1	10	20	67	112	156	1117
B2	8	20	47	84	188	1118
B3	12	26	72	185	421	880
B4	20	43	120	166	219	999
B5	32	45	76	103	267	1126
B6	21	37	104	222	368	870
B7	21	30	89	155	221	1035
B8	22	26	72	125	267	1043
B9	25	46	124	182	224	994
B10	17	38	67	179	378	911
B11	28	58	88	120	268	1065
B12	8	15	36	144	189	1059
B13	10	27	89	175	385	880
B14	23	54	111	148	221	1037
B15	28	39	125	172	278	959
B16	25	53	88	130	188	1149
B17	20	43	75	166	256	999
B18	22	26	50	172	229	1058
B19	13	38	78	124	189	1078
B20	10	19	28	121	268	1046

Table B.3: Test condition 3

TEMP=65°C, RH=85%

	Hours					Proj.
Disk	0	500	1000	1500	2000	Failure
C1	14	23	58	112	278	2057
C2	10	17	55	165	263	1948
C3	11	56	88	138	189	2078
C4	18	28	78	117	243	2106
C5	17	45	78	143	189	2167
C6	10	14	45	154	231	2031
C7	31	53	111	156	211	2151
C8	29	54	106	154	218	2128
C9	22	32	65	89	126	2799
C10	29	36	78	145	188	2297
C11	21	38	89	148	227	2075
C12	24	45	68	134	211	2236
C13	28	57	78	132	190	2352
C14	19	47	61	117	150	2486
C15	25	65	89	184	256	1972
C16	10	18	57	113	178	2189
C17	21	34	45	98	121	2845
C18	12	20	34	112	176	2308
C19	28	56	108	176	243	2001
C20	29	36	57	143	238	2207

Table B.4: Test condition 4

TEMP=70°C, RH=75%

Disk	Hours					Proj.
	0	625	1250	1875	2500	Failure
D1	25	34	64	92	167	3240
D2	25	93	134	154	211	2596
D3	7	23	97	103	178	2615
D4	10	20	56	89	155	2920
D5	5	20	78	132	187	2496
D6	5	15	52	112	167	2644
D7	22	34	67	132	188	2851
D8	12	17	56	78	108	3318
D9	22	34	67	132	189	2847
D10	23	27	54	121	152	3129
D11	11	20	41	87	115	3249
D12	15	18	43	88	118	3343
D13	19	21	38	82	135	3435
D14	18	22	86	178	245	2456
D15	22	26	73	145	252	2582
D16	18	18	29	66	127	3649
D17	22	26	93	145	178	2761
D18	18	27	56	88	134	3316
D19	11	32	44	97	143	3051
D20	12	56	66	124	249	2550
D21	14	34	54	77	112	3500
D22	20	23	25	50	181	3593
D23	11	16	27	54	160	3275
D24	17	24	25	58	108	4034
D25	11	25	22	62	130	3488
D26	17	24	25	70	123	3707
D27	21	39	63	78	163	3304
D28	20	28	45	111	243	2787
D29	15	21	38	65	134	3453
D30	10	34	54	96	176	2841

Appendix C: Copyright Permission for Chapter 3 and Chapter 4

↩ Reply all ∨ 🗑 Delete ⊘ Junk Block ⋯

Re: About reprint permission of JSM proceedings

ⓘ You replied on Thu 6/9/2022 7:16 PM

NV **Nirala, Valerie** <val@amstat.org>
To: Friedman, Naomi <naomi@amstat.org>; Damola Akinlana
Thu 6/9/2022 9:37 AM

Hi Damola,

Since you maintain the copyright to your proceedings paper, you do not need permission to reprint some or all of your paper.

Valerie Nirala
Editor and Content Strategist
American Statistical Association
Promoting the Practice and Profession of Statistics®
732 North Washington Street
Alexandria, VA 22314-1943
(703) 684-1221, Ext. 1876
Fax: (703) 997-7299
www.amstat.org

From: Damola Akinlana <akinlana@usf.edu>
Sent: Monday, June 6, 2022 10:13 AM
To: Friedman, Naomi <naomi@amstat.org>
Subject: About reprint permission of JSM proceedings

Good Morning,

I trust you have been good. I would like to inquire if ASA gives reprint permission for manuscripts submitted to JSM proceedings. I presented part of my PhD research during JSM conferences, and I submitted them to JSM proceedings. I am writing my dissertation now and will be including some details from the manuscripts. My school require me to present a reprint permission approval or a notification that permission is not needed.

Thank you for your time and I anticipate your response.

Best.