UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Mathematical Sciences

Optimal Design of Nonlinear Multifactor Experiments

by

Yuanzhi Huang

Thesis for the degree of Doctor of Philosophy

February 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES
Mathematical Sciences

Doctor of Philosophy

OPTIMAL DESIGN OF NONLINEAR MULTIFACTOR EXPERIMENTS

by Yuanzhi Huang

Optimal design is useful in improving the efficiencies of experiments with respect to a specified optimality criterion, which is often related to one or more statistical models assumed. In particular, sometimes chemical and biological studies could involve multiple experimental factors. Often, it is convenient to fit a nonlinear model in these factors. This nonlinear model can be either mechanistic or empirical, as long as it describes the unknown mechanism behind the response surface. In this thesis, our main interest is in exact optimal design of experiments for nonlinear multifactor models. In order to search for optimal designs, we can use the conventional point or coordinate exchange approach, which however can incorporate a new continuous optimisation method. On the basis of this idea, we further develop and implement a multistage hybrid method to construct local and (pseudo-)Bayesian optimal designs. The recommended hybrid exchange algorithm overcomes the shortcomings of the modified Fedorov exchange algorithm and the coordinate exchange algorithm, contributing to improved properties of the experimental designs obtained. In addition, Bayesian optimal design with respect to an expected criterion function is based on the assumed parameter prior distributions for the nonlinear model. To limit the time for approximating expected criterion values in the algorithm, we use some efficient numerical integration methods (e.g. a Gauss-Hermite quadrature), which are much superior to the traditional pseudo-Monte Carlo method.

We demonstrate the hybrid exchange algorithm by means of several examples relevant to Michaelis-Menten kinetics and other biochemical applications. Under some of these circumstances, we consider hybrid nonlinear models which can be adopted to be fitted to the data of new experiments, the tailor-made optimal designs of which are therefore found and compared with each other. In order to normalise the error structure of such a hybrid model, sometimes the Box-Cox transformation can be applied and the result would be a transform-both-sides (TBS) model. Optimal designs for either untransformed models or TBS models can be used for future experiments, as well as for comprehensive studies of complicated mechanisms.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Yuanzhi Huang , declare that the thesis entitled *Optimal Design of Nonlinear Multifactor Experiments* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission

Signed:.....

Date:.....

Acknowledgements

First and foremost, I wish to thank my PhD supervisors Prof. Steven Gilmour, Dr. Kalliopi Mylona, and Prof. Peter Goos. Without their mentoring and comprehensive advice, I could not complete all the research and finish this thesis. I am always grateful for their help in the past four years. Second, I would like to thank all my colleagues in the University of Southampton and the Universiteit Antwerpen. It has been a pleasure to work with them. Third, I would also like to thank my parents for their invaluable love and continued support.

PS: I take this opportunity to thank my two examiners, Prof. John Matthews and Dr. Stefanie Biedermann, for their constructive comments on this thesis.

Nomenclature

n	number of experimental runs or units
p	number of treatment parameters in the model
v	number of independent variables in the model
$\mathbf{X}$	experimental design, a $n \times v$ matrix
$\mathbf{Y}$	experimental data, a $n \times 1$ vector
$\mathbf{F}$	$n \times p$ design matrix
f	model function
$f(\mathbf{X}_i)$	model function in correspondence to the ith experimental run, for $i = 1, 2, \ldots, n$
ε_i	model residual in correspondence to the ith experimental run, for $i = 1, 2, \ldots, n$
σ^2	true variance of ε_i
$\mathbb{N}$	normal distribution
$\mathbb{E}$	expected value
$\mathbb{V}$	variance
$\boldsymbol{\theta}$	vector of treatment parameters in the model
$\hat{\boldsymbol{\theta}}$	estimator of treatment parameters
$\tilde{\boldsymbol{\theta}}$	reference estimate of treatment parameters from old experimental data
$\boldsymbol{\theta}^0$	local parameter prior or point prior
ρ	parameter prior density (i.e. joint probability density function) of $\boldsymbol{\theta}$
$\boldsymbol{\mu}$	mean vector of ρ
$\boldsymbol{\Sigma}$	variance-covariance matrix of ρ
ϕ	local criterion function in terms of $\boldsymbol{\theta}^0$
φ	expected criterion function in terms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$
$\hat{\varphi}$	approximate value of φ

Chapter 1

Introduction

1.1 Experimentation and Response Surfaces

A response surface can be expressed as a specified functional relationship between the response and the independent variables. When an experiment is conducted to observe the response variable, we are able to use the dataset and the assumed statistical model to approximate the response surface. For scientific research in industries and laboratories, experimentation is an essential tool as the complement to theoretical studies. In this case, an independent experiment or even a series of sequential experiments can be conceived and undertaken for multiple purposes. In relation to statistical inferences of our particular interest, a common experimental purpose is to observe and interpret the assumed response surface, and to obtain sufficient information or new knowledge about the mechanism. Other common purposes in statistics include comparison of various treatments, identification of the expected optimal response, inference of the intrinsic causalities between the response and the individual independent variables, scientific interpretation of the response variation under the mechanism studied, statistical tests to check the fit of the prespecified model, and so on. Not all of them can be best fulfilled in the same experiment, so an optimal compromise between desirable purposes needs to be found to meet our multifaceted expectation from statistical inference.

With the data collected in the process of controlled experimentation, different statistical models can be fitted. Most of the above purposes we consider can be linked to some characteristics of the response surface. Hence, an important issue is to figure out how to best set the levels of the independent variables (or controlled variables) for each run of the experiment. Meanwhile, the other issue is to allow for specific experimental restrictions and other constraints in various situations, when some feasible settings are required in planning future experiments. A useful experimental plan can therefore instruct the operation of experimentation and the collection of data and information. The theme of this thesis is *optimal design of experiments* (ODoE), which is to develop a clever

allocation of treatments (i.e. combinations of the controlled variable levels) to a number of experimental runs so as to better achieve desirable purposes in our mind.

In Box and Draper (1975) there was a list of 14 desirable properties of designs, supposing the experimenters will use their future collected data to model an assumed response surface. With different combinations of the variable levels under control, in accordance with the model we assume beforehand, the expected response will be different in the experiment. Good experimentation therefore needs an efficient and feasible arrangement of the runs and the factor levels.

The standard experimental designs can perform well in some aspects, depending on where the compromise among those experimental purposes lies. For this reason, for instance, the *factorial design* and the *central composite design* (CCD) are the two most popular choices for second-order linear experiments. These can be found in many classical textbooks (e.g. Myers et al. (2009), Box et al. (2005)). The term "linear experiment" indicates an experiment that requires the fit of a prespecified linear model. In practical experimentation, most linear models we propose are of first order or second order. The more complex models often require a lot of treatment parameters to be estimated, when experimental data and resources are limited. For example, let Y denote the single response variable and x_1, x_2 denote the two independent variables. For $i = 1, 2, \ldots, n$, the full first-order linear model can be written as

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i.$$

The notation n represents the number of response observations (i.e. the number of experimental runs or units if all observations are valid), $\beta_0, \beta_1, \beta_2$ are the treatment parameters, and ε is the model error. With the addition of two quadratic effects and the interaction between x_1 and x_2, the full second-order linear model is

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + \varepsilon_i',$$

the new model error of which is $\varepsilon' \neq \varepsilon$. The model can fit a better response surface but contains more parameters for estimation. With some experimental data available, it is viable to use either least squares or maximum likelihood to obtain the estimates of these, the respective standard errors of which can also be calculated. The fitted linear model must approximate the response surface as well as possible over a specified *region of experimental interest*.

Due to the existence of the pure random error as well as the lack of fit of the given linear model, we shall examine and minimise discrepancies between the experimental data and the model prediction. As such, even though the model function for the true response surface is unknown, a suitable alternative model can be identified for a best description of the response surface. The classical *response surface methodology* (RSM) for conventional linear models had been introduced in Box and Draper (1959).

1.2 Basics of Optimal Design of Experiments

1.2.1 Motivation and Concepts

When the theoretical model (which describes the unknown mechanism) is unavailable prior to the experiment, the idea of RSM is to use several candidate linear models to explore and approximate the true response surface. In this case, the fitted model is also called an empirical model or a mathematical model, which is not established on the basis of relevant scientific theories. An empirical model is often modified to suit the dataset, as we can either add or delete some model terms. Expressed in a concise and smooth form, the empirical model simplifies the overcomplex functional relationship between the response and the controlled variables. See Box and Draper (1987) for an introduction in more detail.

After we assume the statistical model (no matter whether it is empirical or theoretical) for the experiment, there are different purposes to focus on. We can choose a specific statistical criterion and find the optimal experimental design that satisfies one or more corresponding purposes. It is conventional to refer to these criteria as the alphabetical criteria of optimality or *optimality criteria*, since there is often a letter in the front of each criterion definition. Most of them will require the evaluation of the Fisher information of the model. That is a $p \times p$ matrix $\mathbf{F}^{\mathrm{T}}\mathbf{F}$, where p is the number of treatment parameters to be estimated and $\mathbf{F}$ is the *design matrix* (or model matrix). The columns of $\mathbf{F}$ can be computed as functions of the levels of the controlled variables. Below we will introduce the four most common optimality criteria.

For example, if the single experimental purpose is the precise estimation of some or all of the model parameters, we can use the D-criterion or the (weighted) A-criterion. The D-criterion aims to minimise the volume of the *confidence region* of the treatment parameter estimator $\hat{\boldsymbol{\theta}}$. This volume is also called the Generalised Variance of the estimator. The D-criterion is invariant to the scales of the p parameters in the model. It therefore treats them equally in least squares or maximum likelihood estimation. While D represents "the determinant of the Fisher information", under the criterion, our aim is to maximise the criterion function

$$\phi_{\mathrm{D}} = \log \left| \mathbf{F}^{\mathrm{T}}\mathbf{F} \right| .$$

In the function, of course different numerical scales can replace the natural logarithmic one we use. If the D-criterion is chosen, the optimal experimental design $\mathbf{X}_{\mathrm{opt}}$ should be found to maximise the determinant of the (Fisher) information matrix. In this case, if we assume the model errors ε to take zero mean and constant variance σ^2 across all

the experimental observations, the information matrix is

$$\mathbf{F}^{\mathrm{T}}\mathbf{F} = \left(\frac{\mathbb{V}(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}, \mathbf{X})}{\sigma^2} \right)^{-1}.$$

Hence, the Fisher information of the model measures the variances and covariances of the estimator $\hat{\boldsymbol{\theta}}$, which experimenters would like to minimise. $\mathbb{V}(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}, \mathbf{X})$ is the variance-covariance matrix, which is conditional on the specified experimental design $\mathbf{X}$. If the model we assume is nonlinear in terms of $\boldsymbol{\theta}$, this matrix also depends on the unknown values of the parameters $\boldsymbol{\theta}$. The Fisher information is also the Cramér-Rao lower bound of $\mathbb{V}(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}, \mathbf{X})$. Notice that the constant error variance σ^2 is a nuisance parameter under the D-criterion, the estimation of which is unimportant.

As an alternative to the D-criterion, the A-criterion aims to minimise the average variance of the individual parameter estimators. Though it ignores the covariances between these estimators, it does not influence the parameter estimation much for simple statistical models with few parameters. This criterion depends on the scale or magnitude of each parameter estimator even if the assumed model is linear. Hence, it is normal to use a weighted A-criterion instead, which is a special case of the similar L-criterion (the letter L represents "linear"). The L-criterion maximises the function

$$\phi_{\mathrm{L}} = \mathrm{Trace}(\mathbf{W}(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}),$$

where $\mathbf{W}$ is the weight matrix that is either numeric or depending on $\mathbf{X}$. If $\mathbf{W}$ is diagonal, the above can also be called the weighted A-criterion function. While the L-criterion takes the covariances of the estimator $\hat{\boldsymbol{\theta}}$ into account, the weighted A-criterion does not. In comparison, when we consider the covariances, the parameter estimates from the new experimental data tend to be more accurate. Otherwise, when we focus on those individual variances, the estimates will become more precise. A higher expected precision means that experimenters can report narrower individual confidence intervals of the estimates $\hat{\boldsymbol{\theta}}$. As usual, the indispensable condition is that the independent model errors should come from an identical distribution with mean zero, and the distribution is normal if we use maximum likelihood to obtain $\hat{\boldsymbol{\theta}}$.

In contrast, there are special criteria for model prediction purposes. The standardised variance of a fitted model response can be expressed as $n\mathbb{V}(\hat{Y}_i)/\sigma^2$, for $i = 1, 2, \ldots, n$. We can then use the G-criterion to minimise the maximum standardised variance across $\mathcal{X}$, where $\mathcal{X}$ is the region of experimental interest. When we are planning the new experiment, it is also called the *design region* or *variable space*, as defined in this thesis. Otherwise, the I-criterion (or V-criterion) aims to minimise the average standardised variance over $\mathcal{X}$. Therefore, the criterion function to be minimised is a multiple integral,

if the variable space is continuous in each dimension. For instance, the unweighted I-criterion function for the two-variable linear model in Section 1.1 is

$$\phi_{\mathrm{I}} = \int\int_{\mathcal{X}} \frac{n\mathbb{V}(\hat{Y}(x_1, x_2))}{\sigma^2} dx_1 dx_2.$$

If we assume the error variance σ^2 to be constant across $\mathcal{X}$, the emphasis of the I-criterion is on the fitted response variance function $\mathbb{V}(\hat{Y})$, which should also depend on the Fisher information of the model. Either the G-criterion or the I-criterion can lead to precise response prediction of the fitted model. When $n \to +\infty$, the standardised variance of the fitted response will be equal to p, the number of treatment parameters, at each support point under the D-criterion. This is a result from the General Equivalence Theorem; see Kiefer and Wolfowitz (1960). Meanwhile, this number p will also be the maximal variance across the space $\mathcal{X}$ for independent variables. In this case, the D-criterion for parameter estimation is identical to the G-criterion for model prediction. Most of the time in practical experiments, however, the equivalence does not hold, since we cannot afford such a huge n as the number of runs.

When we choose one relevant criterion for optimal design of an experiment, sometimes the number of runs n will not be specified and thus we assume $n \to +\infty$. In this case, what the optimality criterion looks for is the *continuous optimal design* or approximate optimal design, which is often expressed as a weight measure. Suppose there are two controlled variables x_1 and x_2 whereas we use two distinct levels for each, at most there are four support points in total. A specific weight is then allocated to each point such that a continuous experimental design is written as

$$\mathbf{X}_{\mathrm{con}} = \left\{ \begin{array}{cccc} (x_{11}, x_{21}) & (x_{11}, x_{22}) & (x_{12}, x_{21}) & (x_{12}, x_{22}) \\ w_1 & w_2 & w_3 & w_4 \end{array} \right\}. \tag{1.1}$$

In the first row in this matrix, we write down the coordinates of each point in two dimensions; while in the second row, the weights are all real numbers that are no less than zero. Here, we can let $w_1 + w_2 + w_3 + w_4 = 1$. However, if there is a specific number of runs n, the weights can be rounded off to be integers so that the above $\mathbf{X}_{\mathrm{con}}$ can become an *exact experimental design* of n runs. It is expressed in the form

$$\mathbf{X}_{\mathrm{exa}} = \left\{ \begin{array}{cccc} (x_{11}, x_{21}) & (x_{11}, x_{22}) & (x_{12}, x_{21}) & (x_{12}, x_{22}) \\ n_1 & n_2 & n_3 & n_4 \end{array} \right\}, \tag{1.2}$$

where n_1, n_2, n_3, n_4 are the respective *numbers of replicates* for the four support points. Of course, these integers must be nonnegative. The random model error ε is the sum of the pure error and the lack of fit of the predicted response. With sufficient replication at the support points, we are able to evaluate the variation in the pure errors of the model and also improve the defined criterion function.

Continuous weight measures are useful in theoretical studies that require derivations of corollaries (e.g. General Equivalence Theorem) to make some statistical inference. Once the optimal solution is found, we can round the weights to decide the exact numbers of replicates. When n is a huge number, the rounded solution will be close to optimal. However, this is often not the case in practice, as most experiments do not need an excessive n. Hence, this thesis focuses on exact optimal designs, which are more applicable and realistic in particular for small-scale experiments. In this situation, it will usually take some computational effort to search for tailor-made exact optimal solutions.

In addition to the above four common optimality criteria, there are various other criteria we can choose from. For instance, some recent criteria were developed to maximise the power of a t-test or F-test in statistical inference (Gilmour and Trinca, 2012b). Unlike factorial designs and CCDs, a normal optimal design often focuses on a specific experimental purpose and aims to optimise its own performance in that aspect. When there are multiple purposes to achieve in the same experiment, one can also use a *compound criterion* or a *composite criterion* that combines and weights several individual criterion functions. The relevant examples are to be shown in Chapters 3 and 5.

Our emphasis in most of this thesis is on the experimental purpose of the precise estimation of the treatment parameters in nonlinear models. This is arguably the most common purpose in relation to statistics, since the obtained parameter estimates will explain the quantified relationship between the dependent and independent variables. The same or a similar method applies to linear models and the search of optimal designs for some other experimental purposes too. There is no special need to do demonstrations under all kinds of similar scenarios.

Moreover, in the estimation of the treatment parameters, maximum likelihood can lead to similar results as nonlinear least squares (NLS), if we assume the error distribution to be normal for the model. If not otherwise specified, the default in this thesis is to consider least squares as the method to fit deterministic statistical models, which also makes it easier to derive and recalculate defined criterion functions. In a special case in Chapter 5, however, one of our designated steps is to estimate the Box-Cox transformation parameter α in a full treatment model. In this situation, the estimate and its variance function should be obtained with maximum likelihood.

1.2.2 Computation and Exchange Algorithms

In the experiment, there are n response observations to be made under different conditions that we can change. Therefore, the Fisher information is composed of n unit information matrices that tell us the expected information from each observation. No matter which criteria we use (most of our demonstrations use the D-criterion), the ideal

achievement will be a maximisation of the information we obtain from the n-run experiment. An experimental design $\mathbf{X}$ involves n rows and v columns corresponding to the v independent variables in the model. If the total number of coordinates nv is a small number, over the design region $\mathcal{X}$, it is viable to do a direct constrained optimisation to search for the exact optimal $\mathbf{X}$ (Chaloner and Larntz, 1989). Otherwise, in more complex cases in optimal design, there are some simple search algorithms available in statistical software, such as SAS, JMP and Design-Expert.

The first step of a traditional search algorithm is to discretise $\mathcal{X}$. Rather than the bounded continuous variable space, we would like to use a list of candidate runs or points. To cover the entire variable space $\mathcal{X}$, a sufficient number of candidates should be taken into account. For instance, we restrict the two controlled variables x_1, x_2 into continuous space $[-1, 1]$ and also assume no more constraints on this space. As such, the area of the square $\mathcal{X}$ is four. If we then define seven candidate levels for each controlled variable, there are 49 candidate points shown in Figure 1.1. In particular, when the model involves one or more categorical variables or block variables, it makes more sense to discretise $\mathcal{X}$. However, if we define numerous discrete candidate points, it will take a lot of computational effort later in the search.

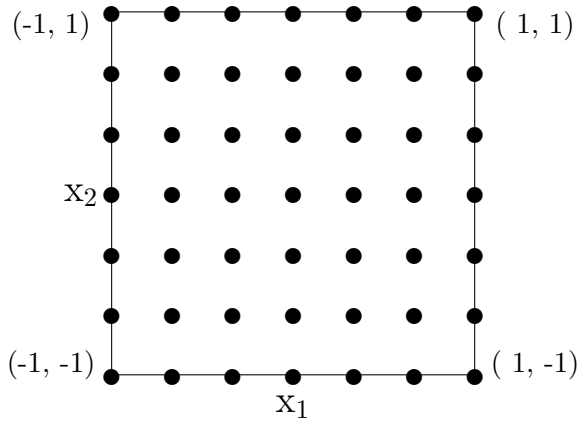


Figure 1.1: Defined Discrete Candidate Points in a Square Region

With the chosen criterion, we define a set $\mathbf{\Omega}$ of the candidate points. In this case, the traditional *discrete optimisation* of a specific $\mathbf{X}$ can be completed in several iterations. In each iteration, the aim is to update the individual rows or even the single coordinates of $\mathbf{X}$, towards a maximisation (or minimisation) of the criterion function. As such, the whole optimisation task is divided into a number of much smaller ones, which makes the optimisation feasible. For instance, the DETMAX procedure (Mitchell, 1974; Galil and Kiefer, 1980) can promote the addition or deletion of the rows in the specific $\mathbf{X}$. The idea is to include the most beneficial points and remove the least beneficial ones in the updated $\mathbf{X}$. As a result, the total number of experimental runs n shall remain the same after each update but we can expect a sequential improvement in the criterion function.

The classical search algorithms tend to adopt the *point exchange approach* (Fedorov, 1972, chap.3), which combines the two steps to delete and then add the rows in the

matrix $\mathbf{X}$. With each exchange proposal, it can swap one current point with the most beneficial candidate from $\boldsymbol{\Omega}$. In the iteration, the number of runs is fixed so that the stepwise discrete optimisation solutions can be more reliable. Points that might compose the optimal $\mathbf{X}$ are sought and the search will not terminate until an *exchanging rule* is no longer satisfied for each current point in $\mathbf{X}$. The iterative procedure cannot guarantee the final $\mathbf{X}$ to be truly optimal, since the criterion function can converge to different local optima and get stuck there (i.e. the stepwise exchanges will not improve the solution further). For efficient optimal design of experiments, the same iterative search should restart with various initial matrices of $\mathbf{X}$. We can take random matrices, but some specific points can also be included in each initial $\mathbf{X}$.

Iterations under the above exchange approach imply considerable computational effort, in some cases when the model function in terms of treatment parameters is complex or when $\mathbf{X}$ includes a lot of coordinates. Besides, compared with the A-criterion or L-criterion, it is often easier to implement the D-criterion, the function of which can be much simplified and thus require less computation. On the basis of the *Fedorov exchange algorithm* (Fedorov, 1972, chap.3), the earliest one that uses the point exchange approach, some modified versions were proposed in the past decades. In comparison, the *modified Fedorov exchange algorithm* (Cook and Nachtsheim, 1980) is the most popular one in practice, since it streamlines the regular iterative procedure and thus decreases the time spent in the search. For even less computational effort and quicker discrete optimisation (this was important in the past), other modifications include the K-exchange algorithm in Johnson and Nachtsheim (1983) and the KL-exchange algorithm in Atkinson et al. (2007, chap.12). We will reintroduce them in Chapter 2, where we also develop the new *optimised exchange algorithm* and *hybrid exchange algorithm*. Both new algorithms are most relevant to the one in Cook and Nachtsheim (1980), and the latter will be adapted and applied to different examples in this thesis.

Another reasonable choice is the *coordinate exchange algorithm* (Meyer and Nachtsheim, 1995), which comes after a simple modification of the point exchange approach. The classical point exchange approach can swap complete points between the current $\mathbf{X}$ of n rows and the candidate set $\boldsymbol{\Omega}$. If we discretise the variable space $\mathcal{X}$ as shown in Figure 1.1, $\boldsymbol{\Omega}$ is expressed as a 49×2 matrix. These are all the combinations of the candidate levels of the two independent variables. On the other hand, the coordinate exchange approach is developed to swap single coordinates between $\mathbf{X}$ and the respective candidate levels. In that case, the number of candidate coordinates is seven per independent variable. Hence, when there are lots of candidate points in $\boldsymbol{\Omega}$, the coordinate exchange approach can save computational time in the stepwise discrete optimisation.

The last one to mention is the stochastic search, where some randomisation is used in deciding whether or not to accept an exchange proposal. Genetic algorithms, for instance, are useful sometimes for quick computation to solve complicated problems. We will not consider them in this thesis, since the hybrid exchange algorithm can work quite

well and find reliable solutions in all the examples. Likewise, *simulated annealing* can also be applied for optimal design (Meyer and Nachtsheim, 1988). This metaheuristic will demand more computational effort to attain a local optimum of the criterion function. Compared with traditional exchange algorithms, it is less liable to get stuck in inferior local optima, at which $\mathbf{X}$ is less efficient with respect to the criterion.

1.3 Preliminaries and Miscellaneous Topics

1.3.1 Nonlinear Experiments

Let $l(\boldsymbol{\theta}, \mathbf{X})$ denote the log likelihood of the model we assume for the experiment. The Fisher information can be calculated as

$$\mathbf{F}^{\mathrm{T}}\mathbf{F} = -\mathbb{E}\left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} l(\boldsymbol{\theta}, \mathbf{X})\right). \tag{1.3}$$

When the statistical model is linear, the covariance matrix of the least squares or maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is

$$\mathbb{V}(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}, \mathbf{X}) = (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\sigma^2.$$

This result is *asymptotic* as the variance estimate for the fitted model will depend on the observed error distribution and the estimate of σ^2. The design matrix $\mathbf{F}$ comes from the linear model below in the matrix form: $\mathbf{Y} = \mathbf{F}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$. $\mathbf{Y}$ is the vector of the observed responses while $\boldsymbol{\varepsilon}$ is the vector of the model errors. As such, the least squares estimator is $\hat{\boldsymbol{\theta}} = (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{Y}$. Hence, under least squares estimation, we can compute the (Fisher) information matrix from $\mathbf{F}$. Both the D-criterion and the A-criterion will no longer involve the calculation of the likelihood function.

When the statistical model is nonlinear, we cannot write it in matrix form without approximation. We should linearise the model in terms of the treatment parameters $\boldsymbol{\theta}$. To achieve this, the conventional approach is to do a first-order *Taylor series expansion* about a specified centre $\boldsymbol{\theta}^0$, which is a numeric vector of size p. For more details, see Atkinson et al. (2007, chap.17) or Chapter 2 in this thesis. After the first-order linearisation, we obtain

$$\mathbf{F} = \frac{\partial f(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0},$$

the value of which depends on $\boldsymbol{\theta}^0$. The bias of this will be close to zero if $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, so the Fisher information relies on the true parameters we are to estimate after the experiment. What we will obtain in the nonlinear least squares case is almost identical to the Fisher information in (1.3), after the same substitution $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. The $\boldsymbol{\theta}^0$ is called the *parameter prior* (or point estimate) that we have to decide before optimal design of experiments. Model linearisation works best around the specified centre, so we

Table 1.1: A 12-Run CCD for the Response Surface

x_1	x_2	x_1	x_2	x_1	x_2
-1	-1	$-\sqrt{2}$	0	0	0
-1	1	$\sqrt{2}$	0	0	0
1	-1	0	$-\sqrt{2}$	0	0
1	1	0	$\sqrt{2}$	0	0

should make the parameter prior as accurate as possible. In addition to the first-order linearisation, some refined statistical methods also consider the higher-order derivatives in the expansion (Bates and Watts, 1980; Hamilton and Watts, 1985).

Generalised (non)linear models can also attract experimenters' attention, which assume different nonnormal distributions for model errors. The common method to fit the model and estimate the parameters is *iteratively reweighted least squares*. With a specific weight matrix in relation to the presumed error distribution, the Fisher information and then the criterion function can be derived for the model; see Woods et al. (2006) for some examples. Likewise, a similar situation is when some random block effects are added to the model, as we have to revise the Fisher information too.

We can treat CCDs as references to optimal experimental designs. For a second-order linear model we assume, a standard CCD can include 2^v factorial points, $2v$ axial points (at a specified equal distance from the centre), and several centre points. In reference to Figure 1.1, Table 1.1 shows a CCD when the uniform distance is set to $\sqrt{2}$. Hence, except for the centre, the rest of points are located on a sphere with the radius $\sqrt{2}$. In this case, the spherical CCD is *rotatable*. If the distance is one, for another example, the CCD is face-centred, since the axial points and the centre are also midpoints between the factorial points. Besides, if one of the experimental purposes is to compare multiple treatments, we can also use an *orthogonal* CCD with a calculated radius.

Most important of all, the CCD can be useful despite the fact that the model we assume is nonlinear. For some aspects, such as to estimate the response difference between two treatments and to predict the combination of the variable levels that will maximise the response, perhaps a standard CCD is not inferior to, for instance, the $\mathbf{X}$ that maximises the D-criterion function. Hence, when there are multiple experimental purposes, we should not overestimate the improved properties of optimal designs. In those situation, instead of taking the simple D-criterion, a better idea is to aim for several desirable purposes in a compound criterion or a composite criterion, both of which are also mentioned in the last section. The compound criterion function is the weighted sum of its components which are different simple criterion functions. In comparison, when the same criterion is considered, but for more than one statistical model, we combine the different components or models in an overall composite criterion function. It could be difficult to use such a compound or composite criterion when we plan industrial experiments. Of course, the properties we demand for $\mathbf{X}$ can sometimes partly overlap with one another.

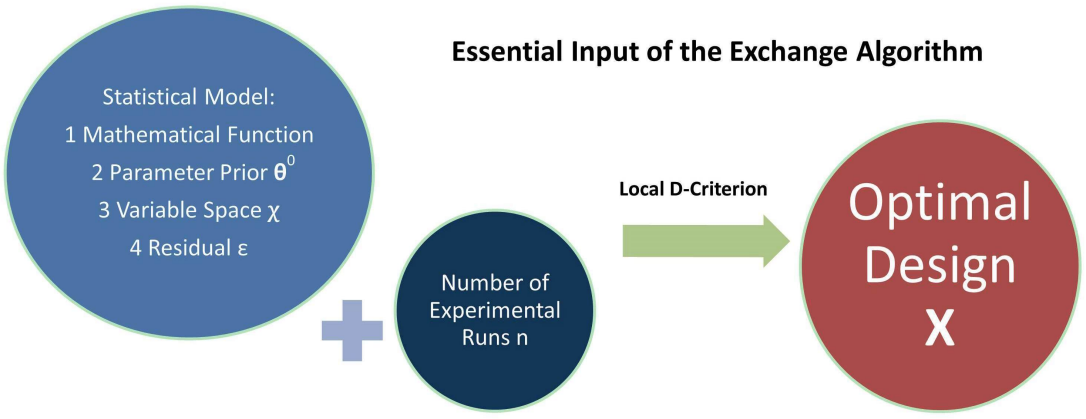


Figure 1.2: Information to Know before Optimal Design for Nonlinear Models

This means that the D-criterion might also suit some other experimental requirements, in addition to the precise parameter estimation.

In those nonlinear experiments, the D-criterion or A-criterion depends on the numeric vector $\boldsymbol{\theta}^0$ that we must determine ourselves. Hence, the criterion function is called the *local criterion function* and the $\mathbf{X}$ that satisfies the local criterion is called the *local optimal design*. In some literature, it is also referred to as the locally optimal design. If we use an exchange algorithm to search for that, the essential requirements are illustrated in Figure 1.2. In brief, we should know all about the statistical model that will be fitted to the new experimental data.

Compared with standard CCDs, for instance, optimal experimental designs are more flexible in some aspects. There are no restrictions on the total number and the replication of experimental runs, as long as n is no less than the number of treatment parameters p. We are able to take account of specific constraints on $\mathcal{X}$, which could appear in different shapes in the v-dimensional space. Moreover, exchange algorithms can be adapted to handle both quantitative and qualitative variables, which we will discuss later in Chapters 2 and 3. Finally, a tailor-made design $\mathbf{X}$ can be found for the nonlinear (multifactor) experiment, which is close to optimal with respect to the specified criterion.

1.3.2 Extensions of Basic Michaelis-Menten Kinetics

In Section 1.1, we showed two empirical models that can be used to approximate the response surface under the unknown mechanism. In contrast, we can sometimes develop theoretical or mechanistic models from scientific theories. The established model can be nonlinear in the unknown parameters representing the treatment effects. A mechanistic model often gives a better approximation to the experimental data at hand. Besides, we can also use it for theoretical studies and interpretation of the properties of the mechanism. In this thesis, we will focus on a specific area in science: the fundamental

Michaelis-Menten kinetics and its extensions. This kinetics is widely used in chemical sciences and is one of the most common examples in statistics addressing nonlinear models. We will introduce the theories and background in Chapters 3 and 5. At the moment, we need to know the statistical model of this kinetic profile, which is

$$\nu_i = \frac{\nu_{\max} S_i}{k + S_i} + \varepsilon_i,$$

for $i = 1, 2, \ldots, n$. As usual, n denotes the number of experimental runs and ε denotes the random error. ν is the *initial rate* of the biochemical reaction of experimental interest whereas S is the initial substrate concentration. The two treatment parameters are $\nu_{\max}$ and k, the values of which require some biochemical interpretations. The above model is the solution of a set of differential equations that are derived under the rate laws in kinetics. The relevant experiment should imply the use of an enzyme, in order to catalyse the reaction that converts the substrate to the final product. The Michaelis-Menten mechanism is the simplest in *enzyme kinetics*; however, we will discuss more complicated mechanisms for extensions of that.

This is because, in experiments close to industrial applications, there could be several controlled factors but no mechanistic model like for simple Michaelis-Menten kinetics. To establish a joint functional relationship between the response and these factors, it is natural to use an empirical model, such as a second-order linear response surface. The response is the (initial) reaction rate in kinetics. In this case, if one of the treatment factors is the (initial) substrate concentration, some theories should hold behind the mechanism. We can then combine the linear response surface (or a similar empirical model) and the Michaelis-Menten kinetics. The result is a nonlinear kinetic model with a hybrid structure. It illustrates a special empirical approximation of the response surface, which we believe is better than the simple linear model. If the initial substrate concentration is one controlled variable and x denotes the other one, we can write a simple nonlinear kinetic model as

$$\nu_i = \frac{(a_0 + a_1 x_i + a_2 {x_i}^2) S_i}{k + S_i} + \varepsilon_i.$$

There are four treatment parameters in the model function. Other candidate models will be introduced in the examples in Chapter 3. Optimal design of experiments for such multifactor kinetic models is little studied in the past. In this thesis, we are also interested in the nonlinear behaviour and the error structure of such a model. On the one hand, the above model has a *partially linear* structure, which simplifies the normal computer search for optimal designs. On the other hand, the structure of the random error ε can be additive, multiplicative, or even more complicated. With the use of NLS or maximum likelihood to fit the model, we must validate some of the classical assumptions. The model errors must be uncorrelated and the error variance must be constant across the n observations. Violation of these will increase the bias in parameter estimation and

therefore the fitted nonlinear model will be less accurate. When there are some outliers or wild observations, one or more assumptions can be violated. One remedial measure is to transform both sides (TBS) of the above kinetic model, for instance. We can do a power transformation to the power α, so the model will become

$$\nu_i{}^{\alpha} = \left(\frac{(a_0 + a_1 x_i + a_2 x_i{}^2) S_i}{k + S_i} \right)^{\alpha} + \varepsilon_i'.$$

The power transformation is recommended in Carroll and Ruppert (1984) for mechanistic models, the error structures of which are little known sometimes. The TBS model can reserve the mechanistic relationship between the dependent and independent variables, so we should use the same transformation on the both sides. In this case, the transformation is almost equivalent to the famous Box-Cox transformation (Box and Cox, 1964). Both can take the role in stabilising the error variance and thus to fix the bad error distribution we assume.

Provided we can find a suitable α, the Box-Cox transformation or power transformation will improve the error structure as we postulate the distribution of ε'. α can be determined on the basis of the information from the dataset, as well as the ease of model interpretation. For a simple Michaelis-Menten model in substrate concentration and two treatment parameters, some transform-both-sides models were fitted in Ruppert et al. (1989), where the Box-Cox transformation parameter α could be estimated by maximum likelihood. In their simulation studies, the TBS model was considered to have a complex error structure as follows:

$$\nu_i^{(\alpha)} = \left(\frac{\nu_{\max} S_i}{k + S_i} \right)^{(\alpha)} + \sigma S_i{}^{\alpha'} \varepsilon_i',$$

where ε' was a standard normal random variable, σ was the unknown standard deviation of the error distribution, and $\nu_i^{(\alpha)}$ represented the Box-Cox transformation on the initial rate. When Ruppert et al. chose different values for the two constants α, α', the simulated data can be skewed or heteroscedastic despite the transformation. Because of that, these discussion were further extended in Nelder (1991) to include generalised linear models for the kinetic data.

The focus in this thesis and in our demonstrations is on nonlinear models relevant to Michaelis-Menten kinetics, no matter if we transform them in advance or not. Nevertheless, the same idea should also fit varieties of applications in widespread areas of science. A simple adaptation of the inputs in the hybrid exchange algorithm can lead to tailor-made optimal designs of different nonlinear experiments.

We can even look at more complex mechanisms that do not follow Michaelis-Menten kinetics. As we can refer to the rate laws, it is a common practice to solve a set of differential equations to build up the kinetic model as required. Even in the less simple cases, Atkinson and Bogacka (2002) applied a *direct method* to obtain the numerical

solutions as well as to approximate the Fisher information of the model. To describe a non-Michaelis-Menten mechanism, the initial reaction rate is often expressed as a rational function in terms of the substrate concentration (no other factors are considered in this example). If $a_1, a_2, a_3, a_4, a_5, a_6, a_7$ are the treatment parameters, we can write the advanced mechanistic model such as

$$\nu_i = \frac{a_1 + a_2 S_i + a_3 {S_i}^2 + a_4 {S_i}^3}{a_5 + a_6 S_i + a_7 {S_i}^2 + {S_i}^3} + \varepsilon_i.$$

The model looks complex as the ratio of two polynomials of the specified degrees (both of which are equal to three for the above equation). It will lead to the basic Michaelis-Menten model when five of the seven unknown constants are equal to zero. However, although the most concise model tends to be simpler as it involves some zero constants, in studies of new mechanisms, such information is often unavailable prior to experiments. As a result, we have to use the experimental data to estimate all p treatment parameters, even though the substrate concentration is the only independent variable.

With a complex mechanistic model like this, it is difficult to either prove or disprove the candidate model. That is, the data or the experimental design cannot "place the model in jeopardy" (Box and Hunter, 1965). In this situation, more than one solution set will be acceptable in the parameter estimation, since each fitted response surface can predict the response observations well. Therefore, even if the complex ratio function is correct, it is not so suitable if our desire is to learn the true parameter values.

Parameter estimates can compensate each other to some extent because of their high correlation structure (Box and Lucas, 1959), so there could be quite different solutions for the complex mechanism we assume. To use a complex mechanistic model, we demand solid scientific evidence for an apt justification of the assumed model. Without such evidence, it is viable to streamline the rational model function and reduce the total number of treatment parameters. Now that there is also a model selection problem to solve, when it is not recommended to adopt the optimal design of the experiment for a single model assumed. Above all, we have to link the statistical model to the nature of science and the mechanism to be studied. We also need to understand the experimental conditions and the possible restrictions on the data collection. This requires a close cooperation with experimenters from different research areas. In Chapter 5, we will do a brief review of Michaelis-Menten kinetics, which we believe is essential for the fundamental comprehension of our examples and numerical results.

1.3.3 Parameter Prior Distribution

In nonlinear experiments, the Fisher information relies on the parameter prior $\boldsymbol{\theta}^0$. To determine a reliable prior that is close to the vector of the true parameter values, one solution is to consider *sequential optimal design of experiments* (Box and Hunter, 1963).

This means that the experiment will be divided into multiple phases and we can fit the model to the collected data after each phase. As such, we can obtain more and more information to influence the choice of the parameter prior, while an optimal design of the subsequent experimental runs can be found for the next phase. This is quite useful in real experimentation, though it is not considered in this thesis. If there are some data or old parameter estimates from literature sources or past experiments, we can make use of them to determine the parameter prior too. Not all such information are reliable for the new experiment to be planned, so this is not a perfect measure.

We can define a point parameter prior for the Fisher information. This is simple and leads to the local optimality criterion. However, to take account of some uncertainties, we can also assume a multivariate *parameter prior density* $\rho(\boldsymbol{\theta})$ (Draper and Hunter, 1967). The mean and the variance of $\rho(\boldsymbol{\theta})$ depend on the information we obtain before the experiment. This information is often based on similar past experiments and old data reported in some literature.

As one approach in optimal design, we can find an $\mathbf{X}$ that minimises the variances and covariances of the parameter posterior distribution. A simple example was shown in Gilmour and Mead (2003), where the main experimental purpose was to find the combination of the variable levels that would maximise a response. This is useful when the parameter prior is quite informative or/and the number of experimental runs is small. In contrast, some people prefer classical frequentist inference about the quantities of interest. The frequentist approach will also be the focus in this thesis. In this case, the ODoE approach is to find an $\mathbf{X}$ that optimises the variance-covariance matrix of least squares or maximum likelihood estimator $\hat{\boldsymbol{\theta}}$.

No matter which approach experimenters wish to use, to take account of $\rho(\boldsymbol{\theta})$, we can refer to an *expected optimality criterion* that maximises the expected Kullback-Leibler distance (i.e. the increase in the Shannon information). As explained in Chaloner and Larntz (1989), for nonlinear models, we assume the *normal approximation* of the posterior and also the normal density $\rho(\mathbf{Y}|\mathbf{X})$. $\rho(\mathbf{Y}|\mathbf{X})$ is independent of parameters $\boldsymbol{\theta}$, so it differs from the likelihood. The first condition must hold if we aim to minimise the posterior variance whereas if the second condition is violated, the response distribution should at least come from the *exponential family* (we shall calculate the related Fisher information). The expected criterion is to maximise (or minimise) the multiple integral

$$\varphi = \int_{\boldsymbol{\theta}} \phi(\mathbf{X}|\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta},$$

where $\phi(\mathbf{X}|\boldsymbol{\theta})$ represents the local criterion function. In Chapter 4, we will show an efficient *numerical integration* method to approximate the expected criterion function. Besides, if the above two conditions do not hold, e.g. in the experiment in Myung and

Pitt (2009), the Kullback-Leibler distance

$$\varphi = \int \int \phi(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}) \rho(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}) \rho(\mathbf{Y}|\mathbf{X}) d\boldsymbol{\theta} d\mathbf{Y}$$

cannot be simplified like in the above equation. This criterion function is not so useful if what we will do is least squares or maximum likelihood estimation of the parameters. A Markov Chain Monte Carlo (MCMC) approach (Müller and Parmigiani, 1995) can be chosen to simulate the empirical criterion function. However, this approach tends to be slow in exact optimal design of experiments.

This thesis will focus on experiments that require nonlinear least squares estimation. In the examples in Chapters 4 and 5, the normal or lognormal parameter prior distributions will be assumed to provide the prior information. With the expected optimality criterion, what we do is called the *Bayesian optimal design* of experiments. Meanwhile, we will not consider the real *Bayesian inference* for analysis of experimental data (even if we do, the criterion function will be the same). In those examples, we are interested in the variance-covariance matrix of the treatment parameter estimators, which can be calculated with NLS and the data collected from the new experiment. Therefore, this approach using the expected criterion is called the *pseudo-Bayesian approach*.

1.4 Outline

The structure of the rest of thesis is as follows.

In Chapter 2, we will demonstrate different exchange algorithms for exact optimal design of nonlinear multifactor experiments. In comparison, a new continuous optimisation method is favoured over the traditional discrete optimisation in the literature, and the differences between the two will be elucidated in the examples. We will also propose a new, hybrid exchange algorithm, which is developed to improve the procedure of optimal design and to save computational time. The numerical results will be presented in this chapter too, along with a brief discussion to conclude.

Chapter 3 will concentrate on several examples in relation to Michaelis-Menten kinetics. We will start with some simple multifactor kinetic models and then extend our applications to more complex ones and to some special scenarios. We will examine these results and also introduce a *modified adjustment algorithm* in order to improve some of the optimal designs further. A number of the local L-criteria will be compared with the local D-criterion. Finally, a simple compound criterion will be proposed too, which allows experimenters to select from multiple nested candidate models.

Chapter 4 will introduce Gaussian quadrature rules, in particular the Gauss-Hermite quadrature, which can offer a reliable approximation of the expected D-criterion function. With this efficient numerical approach, we can assume a suitable parameter prior

distribution for the model and look for the (pseudo-)Bayesian optimal designs of experiments. We will show how to define a small and suitable sample for Gauss-Hermite quadrature, in order to speed up computation and reduce the bias in the numerical approximation of the criterion function.

In Chapter 5, we will finish our discussion about Michaelis-Menten kinetics. A more detailed introduction will be included so that readers can better understand the experiments and kinetic models in the examples. Unlike in Chapter 3, we will focus on the transform-both-sides kinetic models. To fit such a model, an extra task is to determine the Box-Cox transformation parameter α. In Bayesian optimal design of experiments, we will either take α as a prespecified constant or treat it as an unknown parameter to be estimated. In the second case, a compound criterion can be derived for the precise estimation of both the treatment parameters and α.

In the final chapter, we will draw our main conclusions and discuss some future work.

All the computation and calculation are completed in the Matlab environment, with a Dell personal laptop with a 2.60 GHz Intel Core i5-3320M Processor and 4 GB RAM. Besides, when this thesis shows (optimal) experimental designs in tables, note that the experimental runs are sorted in ascending row order with respect to the variable levels. Nonetheless, in the real experiments, these runs (or units) should be independent and run in a random order. As an academic writing rule in this thesis, the personal pronouns "we" and "our" are used instead of "I" and "my".

Chapter 2

Optimal Design of Experiments for General Nonlinear Response Surfaces

2.1 Research Problem

2.1.1 Nonlinear Multifactor Models

In chemical and biological studies, experimenters often wish to explore and delineate some unknown mechanisms with the data observed on presumed response surfaces. To discover sufficient information, an adequate statistical model could be fitted to the experimental data, which relates the response measurement to the set of controlled factors. As the mathematical function is established in the experimenters' belief, the values of the unknown parameters that constitute the model can be deduced with least squares or maximum likelihood estimation. Here, our focus is on model-related experiments and the precise estimation of treatment parameters in the model.

The statistical model can be theoretical or empirical. In vast industrial experiments, the essential information about the working mechanisms is often incomplete or insufficient, so it is neither economic nor convenient to seek the theoretical models. In these circumstances, we will assume the model function to be smooth such that it is realistic to exploit common *empirical modelling techniques* instead. For instance, we can obtain a smooth model function with a simple *interpolation method*. In a broader sense, on the basis of the experimenters' speculation, a linear or even nonlinear empirical model can be found to elaborate the intrinsic mechanism behind the experimental data, whereas the other influential experimental conditions are under control.

The input and output variables of such an empirical model can be scaled or transformed as appropriate in advance. As such, the new space of the controlled variables should be identified as the *region of experimental interest*. Within the defined space at least, the empirical model we assume should be the most capable one to approximate the curve of the true response surface. Though we consider the model to be empirical, it is hard to detect its most suitable form when the unknown mechanism involves multiple factors. There is a chance that the empirical approximation we do is inadequate for the experimental data, due to the existence of the lack of fit of the model. Moreover, there is another problem to be concerned with. When we interpret the nature of science, an empirical model is of limited value, even if it offers a first approximation to the data.

It is recommended to use the model-oriented *optimal experimental design* which is tailored to a mechanistic (or theoretical) model rather than an ill-defined empirical model. The mechanistic model is often derived from scientific theories and formulae, when the actual mechanism is somewhat simplified under the relevant conjectures. A fitted mechanistic model is still an approximation to the response surface, but one can anticipate a better and more intuitive elucidation of the complex mechanism. It is usual that the mechanistic model is nonlinear in terms of the independent variables as well as the treatment parameters. Moreover, we can consider the need for the transformation of the model variables or the both sides of the mechanistic model function, in order to improve the observed error structure. For some discussion of mechanistic studies and corresponding design of experiments, see Box and Hunter (1965).

Unlike common empirical models, mechanistic models tend to be more frugal in the use of unknown parameters. Meanwhile, before the new experiment, such a model cannot be derived without the essential information, apart from the parameter values, about the response surface. In real life, we cannot often acquire this information to derive a theoretical model, depending on the current phase of mechanistic studies. In such scenarios, it is advisable to resort to empirical modelling techniques, while we wish to take account of at least some identified mechanistic features. As we will show in the next subsection, in the second example in this chapter, a nonlinear multifactor model can be found to fit to the experimental data. This model combines some mechanistic information to our empirical approximation about the unknown mechanism. This is therefore referred to as a *hybrid model*. The optimal experimental designs can also be found for this hybrid model, as we will show later in this chapter.

2.1.2 Motivating Examples

Our first example is a chemical engineering application in Box and Draper (1987, chap.11-12), where both an empirical second-order linear model and a mechanistic nonlinear model were fitted to the dataset. As to the experiment described, a pilot plant was operated in a continuous stirred reactor, where the chemists wished to observe the

response surface and interpret the mechanism it implied. A reliable fitted model can be useful afterwards, for instance, for full-scale industrial experiments in future. In this example, the actual mechanism explains a two-step consecutive decomposition reaction of a chemical solution, where the intermediate is the output stream we demand from the reaction. Before the appropriate mechanistic model is derived, the reaction mechanism below shows the theoretical framework:

$$\text{Chemical} \rightarrow \text{Intermediate (i.e. Product)} \rightarrow \text{Byproduct}.$$

An undesirable byproduct shall be formed in the last step of the above reaction. To activate this process, the experiment also involves a catalyst to be diluted in the chemical solution. Yield η is the response of each experimental run, which should be measured after a fixed lapse of time in the reaction. The response variable indicates the amount of the desirable product that has been formed, which is in association with the average reaction rate in the specified time interval.

Three controlled variables could influence the variation in η: flow rate R (at which the chemical flows into the reactor), catalyst concentration C and $x_{\text{T}} = 1000/(T + 273)$ where T is the temperature in degrees Celsius. Note that the response η is called the "conversion rate" of the chemical in Box and Draper's original example, which is an incorrect definition. The experimental data in Box and Draper (1987) corresponds to a 24-run spherical *central composite design* (CCD). There are four central points out of the total 24 points. When the empirical second-order linear model is fitted to the experimental data, $x_1 = (\log(R) - \log 3)/\log 2, x_2 = (\log(C) - \log 2)/\log 2, x_3 = (T - 80)/10$ are the scaled variables. In this situation, the uniform distance from the central points in the spherical CCD is $\sqrt{2}$. The linear model includes the constant term, three first-order terms, three quadratic terms, and three interaction terms. In total, there are 10 unknown treatment parameters to be estimated.

In contrast, the characteristics and the nature of this chemical reaction can be better identified if we fit the mechanistic model as the alternative. Derived from the rate laws in chemical kinetics, this mechanistic nonlinear model is

$$\eta_i = \frac{C_i^{\theta_1}\theta_0 R_i \exp(\theta_2 \mathfrak{x})}{(R_i + C_i^{\theta_1'}\theta_0' \exp(\theta_2' \mathfrak{x}))(R_i + C_i^{\theta_1}\theta_0 \exp(\theta_2 \mathfrak{x}))} + \varepsilon_i, \tag{2.1}$$

where $\mathfrak{x}$ is set to $-0.001(x_{\text{T},i} - 2.8344)$ for simplification. In the light of the nonlinear least squares (NLS) assumptions, for $i = 1, 2, \ldots, 24$, the error ε_i must be independent and come from an identical distribution with zero mean and homogeneous variance. The aim in optimal design of such an experiment is to minimise the Generalised Variance of the NLS parameter estimators. Besides, to plan the new experiment, the *design region* (or *variable space*) $\mathcal{X}$ can be envisaged to be a cuboid. As such, we let $(x_1, x_2, x_3) \in [-1, 1]^3$ delimit the box constraint for the empirical linear model. The equivalent constraint on

the unscaled variables is $(R, C, T) \in [1.5, 6] \times [1, 4] \times [70, 90]$ for the mechanistic nonlinear model (2.1). It is simple and common to define such a variable space, which is contained in the spherical space for the previous CCD.

In the second example, the main experimental interest is to learn the mechanism of the enzymatic depolymerisation of the substrate (i.e. dextran) (Mountzouris et al., 1999). In a stirred-cell membrane reactor under monitoring, endodextranase can be added as the enzyme activator, while the reaction products are different kinds of oligodextrans. The controlled variables are the substrate concentration S (2.5-7.5 % in w/v or weight/volume), the enzyme concentration E (0.625-62.5 Units ml^{-1} times substrate concentration) and the transmembrane pressure P (200-400 kPa). In the original experiment, multiple response variables were observed and each can be linked to a unique empirical linear model. However, in this example, the single response to consider is the substrate conversion rate ξ (%) at the time of measurement. The old experimental data was obtained according to an 18-run face-centred CCD in Mountzouris et al. (1999). The variable space for this experiment is also assumed to be a cuboid.

No mechanistic model can be found to depict the response surface of ξ, so the alternative is to do an empirical approximation. In spite of the simplicity of purely empirical models, it is worthwhile conceiving a hybrid model which could also contain some mechanistic implications and scientific reasoning. As we fit such a model to the experimental data, a better illustration of the mechanism can be anticipated. With the remaining factors fixed, we can extrapolate a negative correlation from the old experimental data of Mountzouris et al.. As a whole, the higher the initial substrate concentration S specified, the lower the substrate conversion rate ξ expected. Herein, there is a smooth nonlinear function $\mathbb{E}(\xi) = \gamma_1 S/(\gamma_2 + S)$, where $\mathbb{E}(\xi)$ is the expectation of ξ. γ_1 is a nonnegative scale and the constant value of γ_2 controls the curvature of this function. For details of our rough derivation and conjecture, see Section 2.10 for the **Appendix** to this chapter. Note that some practical restrictions are removed when we derive the function.

The transformed response of the model is specified as $\xi/(100 - \xi)$, a logit function in the absence of the natural logarithmic scale. $\mathbb{E}(\xi/(100-\xi)) = \gamma_1' S/(\gamma_2' + S)$ where γ_1', γ_2' are unknown constants. After the simplification above, for the hybrid model with three controlled variables, we are inclined to replace the nonnegative γ_1' with an empirical exponential function in terms of $x_{\mathrm{E}} = \log_{10}(E/6.25)$ and $x_{\mathrm{P}} = (P - 300)/100$. If the variable space remains the same for the new experiment, there is $(S, x_{\mathrm{E}}, x_{\mathrm{P}}) \in \mathcal{X} = [2.5, 7.5] \times [-1, 1]^2$. The hybrid nonlinear model can be written as

$$\frac{\xi_i}{100 - \xi_i} = \frac{\exp(a_0 + a_1 x_{\mathrm{E},i} + a_2 x_{\mathrm{P},i} + a_3 x_{\mathrm{E},i}^2 + a_4 x_{\mathrm{P},i}^2) S_i}{a_5 + S_i} + \varepsilon_i. \tag{2.2}$$

Here ε_i is the error term and $a_0, a_1, a_2, a_3, a_4, a_5$ are the six parameters. When the substrate concentration is fixed, the model function is simplified to $\mathbb{E}(\log(\xi/(100-\xi))) = a_0 + a_1 x_{\mathrm{E}} + a_2 x_{\mathrm{P}} + a_3 x_{\mathrm{E}}^2 + a_4 x_{\mathrm{P}}^2$, which indicates the expectation of the transformed

response. Rather than some empirical linear models, in our example, (2.2) will be fitted to the new experimental data.

2.1.3 Statistical D-Criterion for Optimal Design

The experiment includes a series of runs to observe the response and thus to compile the dataset. For each experimental run, we shall determine the respective levels of the v controlled variables. As a whole, we can select an *exact experimental design* $\mathbf{X}$ consisting of n runs. For $i = 1, 2, \ldots, n$, the ith observed response is $Y_i = f(\mathbf{X}_i, \boldsymbol{\theta}) + \varepsilon_i$, where $\mathbf{X}_i$ is the ith row of $\mathbf{X}$. The error mean should be $\mathbb{E}(\varepsilon_i) = 0$ and the homogeneous error variance should be $\mathbb{V}(\varepsilon_i) = \sigma^2$. If the model function $f(\mathbf{X}_i, \boldsymbol{\theta})$ is linear, since least squares estimation applies, the matrix form of the model is $\mathbf{Y} = \mathbf{F}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$. Here, $\mathbf{Y}$ is the column vector of the n observations and $\boldsymbol{\varepsilon}$ is the column vector of the n residuals. In addition, $\boldsymbol{\theta}$ denotes the set of the p unknown treatment parameters, whereas $\mathbf{F}$ is the $n \times p$ *design matrix* (or model matrix). The elements in each row of $\mathbf{F}$ are different functions of the corresponding row elements in the $n \times v$ matrix $\mathbf{X}$. An example of computing the elements of $\mathbf{F}$ will be shown in Section 2.9.

In optimal design, our emphasis is to optimise the estimation of $\boldsymbol{\theta}$, which is the same as to minimise the entries in the $p \times p$ variance-covariance matrix $\mathbb{V}(\hat{\boldsymbol{\theta}}) = (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\sigma^2$. This matrix is a precision measure of the least squares estimator $\hat{\boldsymbol{\theta}}$ and is independent of the real data. As long as the model is linear, $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ is called the *information matrix*, which is identical to the Fisher information of $\mathbf{X}$. To best suit least squares or maximum likelihood estimation, $\mathbf{X}$ should contain the maximal amount of information about $\hat{\boldsymbol{\theta}}$. A common approach to use the local D-criterion, i.e. to maximise the scalar function

$$\phi_{\mathrm{D}} = \log \left| \mathbf{F}^{\mathrm{T}}\mathbf{F} \right|,$$

which is the sum of the logarithms of the eigenvalues of $\mathbf{F}^{\mathrm{T}}\mathbf{F}$. In other words, $\mathbf{X}$ is local D-optimal if it minimises the determinant of the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, which is called the Generalised Variance. If the model errors come from an identical normal distribution, the criterion also minimises the volume of the *confidence region* of $\hat{\boldsymbol{\theta}}$ at a given significance level (it is often 0.05 in biochemical applications). As the most popular optimality criterion, the local D-criterion treats the parameters with equal importance and it is invariant to different scales of the controlled variables in the model.

When the errors are in an identical normal distribution (not required for implementing NLS), the $100(1-\alpha)\%$ *confidence band* of $Y(\mathbf{x})$ is $\mathbf{E}(Y) \pm \mathbb{T}_{\alpha/2}\sqrt{\mathfrak{f}(\mathbf{x})'(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathfrak{f}(\mathbf{x})}$, where $\mathbf{x}$ is the set of the controlled variables. With $n-1$ total degrees of freedom, $\mathbb{T}_{\alpha/2}$ is the t statistic at the significance level α, where if σ^2 is known according to the asymptotic theories, it is feasible to replace $\mathbb{T}_{\alpha/2}$ with the Z statistic that is independent of n. Besides, $\mathfrak{f}(\mathbf{x})$ includes the p model terms at the point $\mathbf{x}$; i.e. the functional form of the

rows in $\mathbf{F}$, in terms of $\mathbf{x}$. If experimenters are more concerned with the model prediction than the parameter estimation, the G-criterion and the I-criterion (or V-criterion) are more attractive than the D-criterion, which all emphasise the confidence band of $Y(\mathbf{x})$. While the I-criterion aims to minimise the average width of the confidence band within the variable space $\mathcal{X}$, the G-criterion aims to minimise the maximum width.

2.1.4 Local Optimal Design for Nonlinear Models

Sometimes we have to find the model-oriented optimal designs of nonlinear experiments. To use the local D-criterion in this case, we shall formulate the analogue of the information matrix of interest. When the model function is nonlinear, to derive the D-criterion function that corresponds to NLS estimation, the most common approach is to linearise $f(\mathbf{X}_i, \boldsymbol{\theta})$ in terms of $\boldsymbol{\theta}$ (see e.g. Atkinson et al. (2007, chap.17)). With the first-order Taylor series expansion about a selected centre $\boldsymbol{\theta}^0$, the model function becomes

$$f(\mathbf{X}_i, \boldsymbol{\theta}) \approx f(\mathbf{X}_i, \boldsymbol{\theta}^0) + \sum_{j=1}^{p} \left(\frac{\partial f(\mathbf{X}_i, \boldsymbol{\theta})}{\partial \theta_j} \Big|_{\theta_j = \theta^0{}_j} \right) (\theta_j - \theta_j^0), \tag{2.3}$$

where the truncation is made after the p first-order terms of θ_j. The column vector $\boldsymbol{\theta}^0$ of p quantities is often chosen as the *prior* or *point estimate* of the currently unknown $\boldsymbol{\theta}$. As a result of numerical approximation, for $i = 1, 2, \ldots, n$, the ith row of $\mathbf{F}$ is

$$\mathbf{F}_i = \frac{\partial f(\mathbf{X}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}^0}.$$

The Gauss-Newton or Levenberg-Marquardt algorithm requires a similar linearisation for NLS. Given $\boldsymbol{\theta}^0$, the information matrix $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ is identical to the Fisher information of the model, which is proportional to the inverse of the variance-covariance matrix of the estimator $\hat{\boldsymbol{\theta}}$. For this reason, we call ϕ_{D} the local D-criterion function, since it is obtained after the substitution $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ and the first-order linearisation in the vicinity of $\boldsymbol{\theta}^0$. The choice of the prior $\boldsymbol{\theta}^0$ relies on the experimenters' expertise or a "Best Guess" about the mechanism behind the model. When the difference $\boldsymbol{\theta} - \boldsymbol{\theta}^0$ is a zero vector, there shall be no bias in (2.3). In contrast, when the difference increases, the linearisation will be less successful at the true values of $\boldsymbol{\theta}$.

As will be shown in this thesis, the parameter prior can be obtained from old datasets and estimates in the relevant literature (if the experimental environments under control are similar). Under different environments, $\boldsymbol{\theta}^0$ can also be modified in accord with the experimenters' conjecture, when there are reasonable demands. Or if there are several independent and similar studies, the current experimenters can negotiate for a tradeoff to determine their own parameter prior for the criterion. As an alternative, the experimenters can conduct sequential pilot studies, in order to find more reliable information about $\boldsymbol{\theta}^0$. A suitable model can also be found through such sequential experiments. In

addition, we can take account of other experimental experiences, records or reports from relevant literature sources, or even unproven predictions from the experimenters.

2.1.5 Need for Exact Design and Algorithm

In an approximate experimental design, the continuous weight instead of the number of replicates will be calculated in correspondence to each of the support points (or unique runs). Although we can round off the decimal weights for a specified total number of experimental runs n, the solution will not work well when n is a small number. Hence, we require a lot of runs to ensure that the obtained experimental design is efficient. In contrast, an exact experimental design $\mathbf{X}$ is made of n runs, which implies the actual scale of the experiment. When n is small, it is not infrequent that the exact optimal design will not resemble the approximate optimal one, in terms of the support points and the replication. The approximate designs shall be acceptable in some theoretical studies, such as the earliest applications in Chernoff (1953); Box and Lucas (1959). Nevertheless, for most real experiments, the exact optimal designs would be the first choices.

In total there are $n \times v$ coordinates in the exact experimental design $\mathbf{X}$, which correspond to the levels of the v controlled variables. To search for the optimal combinations of these coordinates, under the local D-criterion we assume, some computational effort is demanded to make $\mathbf{X}$ efficient. To implement the search in various situations, we shall compile a computer algorithm that uses the criterion to improve nonoptimal designs iteratively. As shown in 1.2, we should know in advance the essential details about the model of interest. Therefore, the listed items must be input into the computer algorithm: an explicit model function $\mathbb{E}(Y_i) = f(\mathbf{X}_i, \boldsymbol{\theta})$, a natural number n, the space of the v controlled variables $\mathcal{X}$ (e.g. an ellipsoid or cuboid in dimension v), and the indispensable assumptions about the errors ε. In the next section, we will introduce the traditional optimal design algorithms and the well-known exchange approach.

2.2 Introduction to Computer Exchange Algorithms

2.2.1 Point Exchange Approach

The matter is to search for the n-run local optimal designs of nonlinear experiments. We should do some iterations, thus a small number of coordinates can be updated in each dependent step of one complete *iterative update*. After each update in $\mathbf{X}$, there should be an instantaneous increment in the local D-criterion function, for instance, which will move closer and closer towards its true maximum. When a local maximum is found, we can stop the iteration and output the final solution $\mathbf{X}$. The use of iterative updates makes it easier to optimise $\mathbf{X}$ as a whole, but the result relies on the initial coordinate

values of $\mathbf{X}$. A similar iterative procedure can also be applied to find the approximate optimal designs of experiments.

On the basis of this procedure, we consider the exchange algorithm (Fedorov, 1972, chap.3) which can suit various nonlinear experiments across widespread areas of science. Herein, if we adopt the classical *point exchange approach* (or *Fedorov exchange approach*), the algorithm is able to update one entire row of $\mathbf{X}$ at a time so that it takes n steps to finish one iterative update. With this approach, Fedorov (1972) built the theoretical framework for the earliest Fedorov exchange algorithm for exact local optimal design. Before the iterative updates, it opts to discretise the continuous design region $\mathcal{X}$ to obtain N candidate runs (or points), which can lead to the *discrete optimisation* of the points of $\mathbf{X}$. As such, the Fedorov exchange algorithm creates a feasible candidate set $\mathbf{\Omega}$, where $\mathbf{\Omega}$ can be expressed as a $N \times v$ matrix.

In advance, we shall know the nonlinear response surface $\mathbf{Y} = f(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$ and the experimental size n. As long as the parameter prior $\boldsymbol{\theta}^0$ and the candidate set $\mathbf{\Omega}$ are both determined, we then have the chance to find the local optimal (or at least an efficient) $\mathbf{X}$. As we consider the discrete optimisation method in this case, each $\mathbf{X}$ in iterative updates shall consist of some candidate runs in $\mathbf{\Omega}$. In order to speed up the computation and the search, the modified Fedorov exchange algorithm (Cook and Nachtsheim, 1980) adopts the fundamentals of the Fedorov exchange algorithm but streamlines the regular iterative update procedure.

As such, in the ith step of each iterative update, for $i = 1, 2, \ldots, n$, the modified Fedorov exchange algorithm will evaluate the different consequences of updating the ith row of the current $\mathbf{X}$ to one of the candidate runs in $\mathbf{\Omega}$. In this case, to measure the relative improvement in the criterion function, we denote d_i as the ratio between the two determinants of the information matrices after and before the specific point exchange is made. Therefore the best candidate in $\mathbf{\Omega}$ shall correspond to the maximal value of this ratio. In our applications of exact local D-optimal design for nonlinear multifactor models, we shall evaluate the information matrix of the model first. The modified Fedorov exchange algorithm is compiled in the Matlab environment, adapted from Cook and Nachtsheim for nonlinear models:

Modified Fedorov Exchange Algorithm for Nonlinear Models

1. Generate an initial nonsingular $\mathbf{X}$ of n independent random draws (with replacement) out of the set $\mathbf{\Omega}$ of N candidate runs defined within $\mathcal{X}$. After the substitution $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, compute matrix $\mathbf{F}$ and the initial value of ϕ_{D}.

2. Let index $i = 1$ and flag $\Upsilon = 0$ when each iterative update starts, which will work through all n rows of $\mathbf{X}$ and the corresponding rows of $\mathbf{F}$ in sequential order.

3. Use $\mathbf{\Omega}$ for the search. At the ith row of $\mathbf{X}$, let $\mathbf{X}_i = \mathbf{X}_{\text{new},i}$ represent the best substitution which leads to the maximum of d_i, the relative improvement in ϕ_{D}. d_i is evaluated for each point exchange proposal between the current $\mathbf{X}_i$ and one of the rows in the matrix $\mathbf{\Omega}$.

4. If the maximal d_i satisfies the prespecified *exchanging rule* (e.g. $\max(d_i) > 1.0001$), execute the substitution $\mathbf{X}_i = \mathbf{X}_{\text{new},i}$ in $\mathbf{X}$; then update $\mathbf{F}, \phi_{\text{D}}$ as well as let $\Upsilon = 1$.

5. When $i < n$, let $i = i + 1$ for the current iterative update. Return to **step 3**.

6. When $i = n$ and if $\Upsilon = 1$, return to **step 2** and this is the end of one dependent iterative update. Otherwise if $\Upsilon = 0$, save the current $\mathbf{X}$ and the corresponding value of ϕ_{D}.

7. Do **all the above steps** for τ tries, each with a random initial $\mathbf{X}$. The exact local D-optimal $\mathbf{X}_{\text{opt}}$ is decided to be the most efficient $\mathbf{X}$ attained in **step 6**, which corresponds to the highest ϕ_{D} out of the τ tries.

8. In real experimentation, the order of the n runs should be randomised.

The local D-criterion shall be used, so that **step 3** of the above algorithm can take advantage of the *updating function* of $|\mathbf{F}^{\text{T}}\mathbf{F}|$ (Fedorov, 1972). This will contribute to the simpler and quicker evaluation of the function d_i (or ϕ_{D}). Letting $|\mathbf{F}^{\text{T}}\mathbf{F}|$ denote the determinant of the information matrix in terms of the current $\mathbf{X}$, $|\mathbf{F}^{\text{T}}\mathbf{F}|_{\text{new}}$ then indicates the determinant of the updated information matrix after one substitution $\mathbf{X}_i = \mathbf{X}_{\text{new},i}$ in $\mathbf{X}$. As such, the updating function of $d_i = |\mathbf{F}^{\text{T}}\mathbf{F}|_{\text{new}}/|\mathbf{F}^{\text{T}}\mathbf{F}|$ is simplified to

$$d_i = (1-\mathfrak{f}(\mathbf{X}_i)'(\mathbf{F}^{\text{T}}\mathbf{F})^{-1}\mathfrak{f}(\mathbf{X}_i))(1+\mathfrak{f}(\mathbf{X}_{\text{new},i})'(\mathbf{F}^{\text{T}}\mathbf{F})^{-1}\mathfrak{f}(\mathbf{X}_{\text{new},i}))+(\mathfrak{f}(\mathbf{X}_{\text{new},i})'(\mathbf{F}^{\text{T}}\mathbf{F})^{-1}\mathfrak{f}(\mathbf{X}_i))^2,$$

where $\mathbf{X}_{\text{new},i}$ is chosen out of the candidate set $\mathbf{\Omega}$ to maximise the ratio d_i. $\mathfrak{f}(\mathbf{X}_i)$ refers to the transpose of the ith row of the current $\mathbf{F}$, which is a vector of the p model terms at the point $\mathbf{X}_i$. Besides, the variance of prediction at $\mathbf{X}_i$ will be

$$\mathbb{V}(\hat{Y}_i) = \mathfrak{f}(\mathbf{X}_i)'(\mathbf{F}^{\text{T}}\mathbf{F})^{-1}\mathfrak{f}(\mathbf{X}_i)\sigma^2,$$

where $\hat{Y}_i$ is the fitted response. Therefore, to improve the local D-criterion function in iterative updates, we tend to select those candidate runs at which the calculated variances of prediction are large. The updating function allows us to compare two different $\mathbf{X}$ and find suitable point exchanges. In optimal design of experiments, it is viable to replace the local D-criterion with a specific A or L-criterion, for instance. A similar updating function for the local A-criterion can also be found in Fedorov (1972), though its evaluation requires more computation.

To summarise, the iterative update can be broken into n dependent steps, one for each row of $\mathbf{X}$. In line with the exchanging rule, the modified Fedorov exchange algorithm

should run several iterative updates until the local D-criterion function has climbed to a peak. For the same reason, we should also do a number of tries, in order to compare different local maxima of the criterion function. A thorough search can make this algorithm more successful. Besides, the final solution (or design) relies on the parameter prior $\boldsymbol{\theta}^0$ so that it is only local optimal at most (if we manage to attain the true maximum of the local criterion function). Global optimal design of the experiment will be found if the unknown $\boldsymbol{\theta}$ is proven to equal $\boldsymbol{\theta}^0$.

As a measure that can accelerate the iterative update further, the K-exchange algorithm (Johnson and Nachtsheim, 1983) is developed to first determine the $\mathfrak{k} \leq n$ runs with the lowest variances of prediction in the current $\mathbf{X}$. These current runs are to be considered for deletion in a complete iterative update and will be reevaluated after that iterative update. This implies that the iterative update will consist of a pass through the $\mathfrak{k}$ rows, selected out of the total n rows. The KL-exchange algorithm (Atkinson et al., 2007, chap.12) applies this idea for a similar modification of the Fedorov exchange algorithm. In that case, it can reduce both numbers of runs in the current $\mathbf{X}$ and also in the discrete candidate set $\boldsymbol{\Omega}$. In contrast, in an iterative update, it will determine the $\mathfrak{l}$ candidate runs with the highest variances of prediction.

2.2.2 Coordinate Exchange Approach

In this chapter, as an alternative to Cook and Nachtsheim (1980), the coordinate exchange algorithm (Meyer and Nachtsheim, 1995) takes more steps in the iterative update. It is developed to interact with one small, discrete and unidimensional candidate set $\boldsymbol{\Omega}_k$ at a time, which is a column vector that includes the candidate levels of the kth controlled variable in the model. In other words, it contains the distinct levels in the kth column of the candidate set $\boldsymbol{\Omega}$. Instead of updating the entire row $\mathbf{X}_i$ as a whole, now each coordinate X_{ik} becomes one separate item under the *coordinate exchange approach*, for $k = 1, 2, \ldots, v$ and $i = 1, 2, \ldots, n$. Although it demands more steps to accomplish each iterative update, it is simpler to update single coordinates in $\mathbf{X}$. When $v > 1$, an option is to use the coordinate exchange algorithm instead of the above modified Fedorov exchange algorithm. While $\mathbf{X}$ will be updated on a more frequent basis, the fixed exchanging rule can be a bit looser in this case.

When there are lots of controlled variables and experimental runs, the coordinate exchange approach is beneficial because it reduces the computational cost to manipulate with quite a long candidate list in $\boldsymbol{\Omega}$. In that case, the coordinate exchange approach can save computer memory as well as time in evaluating discrete candidates. Meanwhile, it also becomes more difficult to find a best candidate point $\mathbf{X}_{\text{new},i}$ after the consecutive updates in the relevant v coordinates. As such, the coordinate exchange algorithm might experience some trouble and take more iterative updates and tries to find the optimal

$\mathbf{X}$, since the local search of optimal coordinates are more dependent on the iterative update. Later we shall check numerical results in the two examples.

2.3 A New Continuous Optimisation Method

For the discrete optimisation, candidate runs or coordinates are chosen from the space $\mathcal{X}$ defined for the v controlled variables. When we search for an efficient $\mathbf{X}_{\text{opt}}$, there is sometimes no valid evidence to support the specification of discrete candidates. As a result, it is beneficial to explore and include more candidate runs outside the finite set $\boldsymbol{\Omega}$, which is composed of v subsets of candidate coordinates. A gridding technique can of course create a denser $\boldsymbol{\Omega}$, with small meshes and narrow space between adjacent candidate points. Nevertheless, since that increases the total number of candidates, the computation will slow down in the iterative update. This is not the best solution as the discrete optimisation cannot cover the whole continuous space $\mathcal{X}$.

The discrete optimisation can work well for most linear experiments, as long as there are a sufficient number of runs. Smith (1918) explored the approximate D-optimal designs for univariate linear models of degree up to six, where the suitable candidate coordinates were calculated for each independent variable. However, when $f(\mathbf{X}_i, \boldsymbol{\theta})$ is nonlinear, it is far from obvious to tell which candidates can compose the optimal $\mathbf{X}_{\text{opt}}$. In this circumstance, we believe that the continuous (nonlinear) optimisation over $\mathcal{X}$ is the more attractive method than the traditional discrete optimisation over a finite $\boldsymbol{\Omega}$.

The continuous nonlinear optimisation is rapid itself, but our new concern is to derive the function of the relative improvement d_i to be optimised, which requires intensive algebraic computation in each step of iterative updates. In this thesis, all of our algorithms are compiled in the Matlab language (version 2013b), which is efficient for most numerical computation. In comparison to C++, for instance, Matlab is much slower in algebraic computation and the interpretation of individual operations in the code. Under the point exchange approach, below is a succinct description of the *optimised Fedorov exchange algorithm* that we first propose in this chapter, which will be similar to the alternative *optimised coordinate exchange algorithm* (when the coordinate exchange approach is used instead):

Optimised Fedorov Exchange Algorithm for Nonlinear Models

1. Generate an initial nonsingular $\mathbf{X}$ of n independent random draws from $\mathcal{X}$. After the substitution $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, compute matrix $\mathbf{F}$ and the initial value of ϕ_{D}.

2. Let index $i = 1$ and flag $\Upsilon = 0$ when each iterative update starts, which will work through all n rows of $\mathbf{X}$ and the corresponding rows of $\mathbf{F}$ in sequential order.

3. At the ith row of $\mathbf{X}$, replace $\mathbf{X}_i$ with $\mathbf{X}^*_{\text{new},i}$ which is an unspecified candidate run and is made up of v decision variables. Write the function d_i in the algebraic expression. Decide (e.g. at random) one or more initial vectors of $\mathbf{X}^*_{\text{new},i}$, while the default is to use $\mathbf{X}_i$.

4. Maximise d_i with an optimisation method such as the Nelder-Mead (or quasi-Newton) method. Output the numeric local optimal solution $\mathbf{X}_{\text{new},i}$ within $\mathcal{X}$, which will substitute $\mathbf{X}^*_{\text{new},i}$. This is the best candidate that can perhaps supplant the current $\mathbf{X}_i$.

5. If the maximal d_i satisfies the prespecified exchanging rule (e.g. a loose rule we use is $\max(d_i) > 1.00000001$), execute the substitution $\mathbf{X}_i = \mathbf{X}_{\text{new},i}$ in $\mathbf{X}$; then update $\mathbf{F}, \phi_{\text{D}}$ and let $\Upsilon = 1$.

6. When $i < n$, let $i = i + 1$ for the current iterative update. Return to **step 3**.

7. When $i = n$ and if $\Upsilon = 1$, return to **step 2** and this is the end of one dependent iterative update. Otherwise if $\Upsilon = 0$, save the current $\mathbf{X}$ and the corresponding value of ϕ_{D}.

8. Do **all the above steps** for τ tries, each with a random initial $\mathbf{X}$. The exact local D-optimal $\mathbf{X}_{\text{opt}}$ is decided to be the most efficient $\mathbf{X}$ attained in **step 7**, which corresponds to the highest ϕ_{D} out of the τ tries.

9. The truncation can be made somewhere after each decimal coordinate in $\mathbf{X}_{\text{opt}}$, since the real experimentation will restrict the coordinates to be finite decimals.

We can choose the Nelder-Mead method (Nelder and Mead, 1965) to implement the continuous optimisation of d_i with a box constraint. In the optimised Fedorov exchange algorithm, it requires the variable space $\mathcal{X}$ to be in the shape of a v-orthotope (e.g. a cuboid when $v = 3$). Under the coordinate exchange approach, on the other hand, there must be no constraints on $\mathcal{X}$ in addition to the two bounds for each controlled variable. This is a simplex method that excels at high-dimensional optimisations or complex functions to be optimised. Notice that the continuous optimisation solutions are warranted to be **local optimal only**, since we are unable to assess the numerous candidates from $\mathcal{X}$. It is sensible to take multiple initial starts in $\mathcal{X}$, so the best local optimal solution can be selected for the experimental run or the specific coordinate.

The alternative optimisation option is the quasi-Newton method (e.g. the BFGS algorithm), which looks flexible for low-dimensional constrained nonlinear optimisations. In addition to the upper and lower limits of the controlled variables in the model, this method can handle other constraints too. Thus we can shape a unique $\mathcal{X}$ to meet specific experimental constraints. Moreover, the coordinate exchange algorithm in Gotwalt et al. (2009) carried out the so-called Brent's method, which is for the unidimensional optimisation of each coordinate in $\mathbf{X}$ (Brent, 1973, chap.5). The computation of the

relative improvement function d_i is simple and quick in each step in iterations, but the local optimal solutions it offers will be no better than those from the Nelder-Mead or quasi-Newton method.

We must decide the dimension, i.e. the number of decision variables, of each continuous optimisation of the relative improvement and thus choose a suitable exchange approach. While the relative improvement is a function derived from the information matrix of the model, under the chosen dimension of the optimisation, we must ensure the qualities of the local optimal solutions $\mathbf{X}_{\text{new},i}$. Besides, if the experiment is on a small scale, i.e. the total number of coordinates is small, Chaloner and Larntz (1989) introduced a simple approach to use the Nelder-Mead method for the direct optimisation of $\mathbf{X}$ as a whole. There are no iterative updates or updating functions, so the optimal experimental design can be found in one step. However, the optimisation is often taxing in computation whereas each optimal solution $\mathbf{X}_{\text{opt}}$ will depend on the initial values of up to $v \times n$ coordinates, which is one of the numerous local solutions in the maximisation of ϕ_{D}. For the optimisation of $\mathbf{X}$ in one step in the R environment, the technical function "optim" can be made use of, but it is vital to use different initial matrices of $\mathbf{X}$. More often than not, exchange approaches are more efficient in computation and optimisation.

2.4 Numerical Results and Comparisons

2.4.1 Example 1: A Multifactor Mechanistic Model

We assume the main experimental purpose to be the precise estimation of the treatment parameters in the nonlinear model. Therefore, the local D-criterion is a sensible choice for optimal design of experiments. In the first example from Box and Draper (1987), one of our particular aims is to compare the findings of the new optimised exchange algorithms to those of the traditional ones that adopt the discrete optimisation method. Our reference $\mathbf{X}$ is the face-centred CCD in Table 2.1, where each experimental unit $\mathbf{X}_i$ is located within the cuboid space $\mathcal{X} = [1.5, 6] \times [1, 4] \times [70, 90]$. The Fisher information of the mechanistic model (2.1) depends on $\boldsymbol{\theta}$. In this example, the parameter prior $\boldsymbol{\theta}^0$ is decided to be identical to the nonlinear least squares estimate $\tilde{\boldsymbol{\theta}} = \{\tilde{\theta}_0 = 5.90,\ \tilde{\theta}'_0 = 1.15,\ \tilde{\theta}_1 = 0.53,\ \tilde{\theta}'_1 = -0.01,\ \tilde{\theta}_2 = 15475,\ \tilde{\theta}'_2 = 7489\}$, which originates from fitting (2.1) to the experimental data in Box and Draper. The reference CCD leads to the calculated local D-criterion value $\phi_{\text{D}} = -52.7712$, which we expect to improve under the 24-run local optimal $\mathbf{X}$. In this thesis, the uniform rule is to keep four decimal places of each criterion value.

The 24 experimental runs in Table 2.1 are allocated to four blocks of equal size six. At the moment, we do not consider these fixed block effects that are not inherent components of the mechanistic model. This topic will be discussed later in Section 2.8.

Table 2.1: 24-Run Reference Face-Centered CCD

Block	R	C	T	Block	R	C	T	Block	R	C	T	Block	R	C	T
2	1.5	1	70	3	3	1	80	2	3	2	80	1	6	1	70
1	1.5	1	90	4	3	1	80	2	3	2	80	2	6	1	90
3	1.5	2	80	3	3	2	70	3	3	2	90	3	6	2	80
4	1.5	2	80	4	3	2	70	4	3	2	90	4	6	2	80
1	1.5	4	70	1	3	2	80	3	3	4	80	2	6	4	70
2	1.5	4	90	1	3	2	80	4	3	4	80	1	6	4	90

In this demonstration, we shall do $\tau = 100$ random tries under the exchanging rule $\max(d_i) > 1.0001$ for the modified Fedorov (or coordinate) exchange algorithm. For $i = 1, 2, \ldots, n$, if the relative improvement d_i is found to be lower than the *critical value* $(1 + 10^{-4})$, the row substitution $\mathbf{X}_i = \mathbf{X}_{\text{new},i}$ will not be made in the current $\mathbf{X}$. When there are more candidate runs in $\boldsymbol{\Omega}$ or when the number of runs n is larger, we can use a smaller critical value and it will trigger more exchanges to update $\mathbf{X}$.

The specification of discrete candidate runs could depend on the experimental conditions, the shape of the variable space $\mathcal{X}$, as well as the number of runs n. A small set $\boldsymbol{\Omega}$ makes the computation quick in the discrete optimisation, while a dense distribution of the candidates within $\mathcal{X}$ can lead to reliable optimisation solutions. In this example, the three distinct levels in the CCD can act as the candidate coordinates for each controlled variable. As such, there are candidate subsets $\boldsymbol{\Omega}_1 = \{1.5, 3, 6\}, \boldsymbol{\Omega}_2 = \{1, 2, 4\}, \boldsymbol{\Omega}_3 = \{70, 80, 90\}$. If we combine the three subsets, there are $3^3 = 27$ runs in the overall candidate set $\boldsymbol{\Omega}$. All the demonstrations are performed using the Matlab environment, with a Dell personal laptop with a 2.60 GHz Intel Core i5-3320M Processor and 4 GB RAM. For the record, the elapsed time of each demonstration will be counted to measure the approximate computational cost.

Consider a full second-order polynomial model in terms of the scaled variables x_1, x_2, x_3. In Box and Draper (1987, chap.11), this is the model for the empirical approximation of the response surface. With the modified Fedorov exchange algorithm, for instance, the 24-run local D-optimal design $\mathbf{X}_{\text{emp}}$ can be found for the empirical linear model. The model is well fitted to the old experimental data, but the $\mathbf{X}_{\text{emp}}$ in Table 2.2 cannot contribute much to the precise estimation of $\boldsymbol{\theta}$ of the mechanistic model (2.1). In this case, the face-centred CCD is 74.66% efficient, relative to $\mathbf{X}_{\text{emp}}$ of which the local D-criterion value is $\phi_{\text{D}} = -51.0181$. Apart from the CCD, $\mathbf{X}_{\text{emp}}$ is another baseline when we compare different $\mathbf{X}$.

Table 2.2: 24-Run Local D-Optimal Design for the Polynomial Model

R	C	T	R	C	T	R	C	T	R	C	T	R	C	T	R	C	T
1.5	1	70	1.5	2	90	1.5	4	90	3	2	80	6	1	80	6	4	70
1.5	1	80	1.5	4	70	3	1	70	3	4	80	6	1	90	6	4	70
1.5	1	90	1.5	4	70	3	1	90	6	1	70	6	2	80	6	4	90
1.5	2	70	1.5	4	90	3	2	70	6	1	70	6	2	90	6	4	90

Our interest is in the local optimal experimental design for the nonlinear model (2.1), which can offer a mechanistic approximation of the response surface. The candidate list in $\mathbf{\Omega}$ is very short, so the modified Fedorov exchange algorithm can finish all the computation within two seconds. In comparison, the coordinate exchange algorithm takes 13 seconds of computational time, which is quick as well. No matter which approach is used, we can find the identical local D-optimal solution $\mathbf{X}_{\text{opt}}$ (Table 2.3) for (2.1), the maximal ϕ_{D} of which is -49.7321 and the number of support points is 11.

Out of the 100 random tries, it is sensible to look at some $\mathbf{X}$ of which the criterion values are the highest. With the modified Fedorov exchange algorithm, the most qualified three $\mathbf{X}$ are identical to our best solution $\mathbf{X}_{\text{opt}}$. This suggests that we have found the global maximum of the local D-criterion function in terms of $\mathbf{X}$. With the coordinate exchange algorithm, the three designs are dissimilar though it could be because of minor random fluctuation. In this case, the highest three criterion values are $-49.7321, -49.7412, -49.7412$. This hints that we should to do more than 100 tries to secure the most efficient solution $\mathbf{X}_{\text{opt}}$ in Table 2.3.

Table 2.3: 24-Run Local D-Optimal Design with Discrete Optimisation

R	C	T	R	C	T	R	C	T	R	C	T	R	C	T	R	C	T
1.5	1	70	1.5	1	90	1.5	4	80	1.5	4	90	6	1	80	6	4	70
1.5	1	80	1.5	4	70	1.5	4	90	3	1	70	6	1	90	6	4	70
1.5	1	90	1.5	4	70	1.5	4	90	3	1	70	6	1	90	6	4	80
1.5	1	90	1.5	4	70	1.5	4	90	6	1	80	6	4	70	6	4	80

For the new optimised exchange algorithms, Matlab can be instructed to do Parallel Computation with two local workers (i.e. the cores of the processor) available in the laptop. If we want to do 100 tries, each local worker will take 50 tries. In this case, the recorded elapsed time of computation will be about half of the normal time.

Our aim in this demonstration is to compare the continuous optimisation over $\mathcal{X}$ to the discrete optimisation over $\mathbf{\Omega}$, so we will also do 100 random tries and use the same exchanging rule $\max(d_i) > 1.0001$. This requires the calculated maximal relative improvement to be greater than 1.0001 such that the proposed exchange will be executed. For the continuous optimisation method, this rule is a bit too strict, so one can decrease the critical value $(1 + 10^{-4})$ in practice. In total, when we adopt the point exchange approach, it costs 3324 seconds in computation whereas the time cost is 3920 seconds under the coordinate exchange approach. It makes little difference in this example.

In each individual step of the iterative update, the current point (or coordinate) also acts as the sole initial vector (or value) for the continuous optimisation of the criterion function. This means that the Nelder-Mead optimisation will search from the vicinities of the current points first, in order to maximise the relative improvements in the iterative update. To obtain more local maxima and then select the best one, we can use multiple initial vectors (or values) at either random or specified locations in $\mathcal{X}$.

When the point exchange approach is used, the local D-criterion value of $\mathbf{X}_{\text{opt}}$ (Table 2.4a) equals -49.5582. Out of the 100 tries and the continuous optimisation, the mean of the 100 local maxima of the criterion function is -49.7484. Under the current exchanging rule, the mean number of iterative updates is 4.48, so it does not take too much trouble to update $\mathbf{X}$. In comparison, when we choose the coordinate exchange approach, the $\mathbf{X}_{\text{opt}}$ in Table 2.4b is found to be local D-optimal. The achieved maximal ϕ_{D} is -49.5506 so that the design is more efficient than the one in Table 2.4a. The mean of the 100 local criterion values is calculated to be -49.7927 while the number of iterative updates is 5.23 on average. The diversities between the two exchange approaches are not too apparent. In Table 2.4, the coordinate values of the two dissimilar $\mathbf{X}_{\text{opt}}$ are shown up to five significant figures, unless the referred levels can be approximated to one of the variable bounds in $\mathcal{X} = [1.5, 6] \times [1, 4] \times [70, 90]$. If not otherwise specified, this is a rule we follow in presenting all the tables in this chapter.

Table 2.4: 24-Run Local D-Optimal Design with Continuous Optimisation Using

(a) Point Exchange

R	C	T	R	C	T
1.5	1	70	2.9872	1	70
1.5	1	80.447	3.0288	1	70
1.5	1	90	3.0820	1	70
1.5	1	90	5.7580	4	70
1.5	4	73.682	5.7829	4	70
1.5	4	73.873	5.8240	4	70
1.5	4	73.986	5.9067	4	70
1.5	4	74.187	6	1	84.702
1.5073	1	90	6	1	84.746
1.7366	4	90	6	1	84.789
1.7373	4	90	6	4	80.308
1.7445	4	90	6	4	80.423

(b) Coordinate Exchange

R	C	T	R	C	T
1.5	1	70	1.7286	4	90
1.5	1	70	3.3102	1	70
1.5	1	90	3.3474	1	70
1.5	1	90	5.6925	4	70
1.5	1	90	5.7240	4	70
1.5	1	90	5.8159	4	70
1.5	4	74.006	5.8280	4	70
1.5	4	74.033	6	1	84.629
1.5	4	74.081	6	1	84.729
1.5	4	74.145	6	1	84.802
1.7045	4	90	6	4	80.088
1.7261	4	90	6	4	80.630

Under the strict exchanging rule $\max(d_i) > 1.0001$, it does not take many iterative updates to transform a random initial $\mathbf{X}$ to one at a local maximum of the local D-criterion function. The first word "local" indicates the local optimisation solution whereas the other "local" implies the criterion's dependence on the parameter prior $\boldsymbol{\theta}^0$.

Exact replicate runs are rare as we see in Table 2.4, since the variable space $\mathcal{X}$ is continuous and thus implies an infinite number of feasible candidate runs. However, in spite of the decimals, it is clear what kind of coordinates we should use under the local D-criterion. Furthermore, most adopted experimental runs in either $\mathbf{X}_{\text{opt}}$ are not far in distance from the discrete candidates defined previously in the $3 \times 3 \times 3$ set $\boldsymbol{\Omega}$. With respect to the local D-criterion we consider, the result does not show much of an effective upgrade from Table 2.3, which we can compose with the traditional discrete optimisation. Let $\phi_{\text{D,A}}$ and $\phi_{\text{D,B}}$ represent the respective local D-criterion values of any two experimental designs $\mathbf{X}_{\text{A}}$ and $\mathbf{X}_{\text{B}}$. Before we make the comparison, it is common

to define the *relative efficiency* of $\mathbf{X}_{\mathrm{A}}$ with respect to $\mathbf{X}_{\mathrm{B}}$ as

$$\mathit{eff} = \frac{\exp(\phi_{\mathrm{D,A}}/p)}{\exp(\phi_{\mathrm{D,B}}/p)} 100\%.$$

As such, the comparison of the two does not depend on the number of parameters nor on the numeric scale we exert on the determinants of the information matrices. In exact optimal design for the mechanistic model (2.1), the most efficient solution $\mathbf{X}_{\mathrm{opt}}$ is found in the optimised coordinate exchange algorithm. With respect to Table 2.4b consisting of quasi-continuous runs (i.e. each coordinate value of $\mathbf{X}$ is an irrational number), the reference CCD is $\exp(-52.7712/6)/\exp(-49.5506/6) \approx 58.46\%$ efficient and the $\mathbf{X}_{\mathrm{emp}}$ is 78.30% efficient. The relative efficiency is as high as 97.02% for the $\mathbf{X}_{\mathrm{opt}}$ in Table 2.3, which is made up of some discrete candidate runs from the small $\boldsymbol{\Omega}$. In other words, we can expect a 3% reduction in the Generalised Variance of the new NLS estimator $\hat{\boldsymbol{\theta}}$.

This indicates that even with a complex mechanistic model, a small and well-specified set of discrete candidates could make the optimal experimental design simple and efficient with respect to the criterion function. However, if we intend to change the values of $\boldsymbol{\theta}^0$ or the boundaries of $\mathcal{X}$, for the discrete optimisation, the commensurate candidate set $\boldsymbol{\Omega}$ must be modified too. In nonlinear multifactor experiments, this can be difficult.

2.4.2 Example 2: A Special Empirical Model

This example is in relation to the experiment in Mountzouris et al. (1999). We assume the nonlinear empirical model (2.2), which fits the old experimental data well. Before searching for the local D-optimal experimental design, we must decide the parameter prior $\boldsymbol{\theta}^0$ in order to evaluate the Fisher information matrix. The common approach is to fit (2.2) to the dataset of the 18-run face-centred CCD (Table 2.5). For the fitted model, the adjusted coefficient of determination is calculated to be 0.986. When the significance level is 0.05, all the estimated coefficients are significant according to the standard t-tests. As we use the old experimental data from Mountzouris et al. (1999), the nonlinear least squares estimate of $\boldsymbol{\theta}$ is found. As we show four decimal places, $\tilde{\boldsymbol{\theta}} = \{\tilde{a}_0, \tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4, \tilde{a}_5\} \approx \{0.4340, 1.3140, -0.1059, -0.8224, 0.4105, -2.0633\}$ which is set to be the parameter prior $\boldsymbol{\theta}^0$ for model (2.2). In this case, the local D-criterion value of the reference CCD is equal to $\phi_{\mathrm{D}} = 31.7538$.

Note that the 16th run of the previous experiment failed, so the relevant observation has been discarded in Table 2.5. In spite of this failure, we assume the same variable space for the new experiment in this example. As such, there is a cuboid $\mathcal{X} = [2.5, 7.5] \times [-1, 1]^2$, which also contains the point of the previous 16th experimental run. Except for the substrate concentration, two other independent variables are transformed in model (2.2). It is more sensible to show the raw variable levels in Table 2.5. According to the $\mathcal{X}$, the

feasible interval of the enzyme concentration is $[0.625, 6.25]$ and that of the pressure is $[200, 400]$.

Table 2.5: 18-Run Reference Face-Centered CCD in Mountzouris et al. (1999)

S	E	P	ξ	S	E	P	ξ	S	E	P	ξ
5	6.25	300	73.6	2.5	62.5	400	95.2	7.5	62.5	400	82.7
5	6.25	200	81.6	7.5	6.25	300	77.3	2.5	6.25	300	90.0
5	62.5	300	76.0	5	6.25	400	69.0	2.5	0.625	400	55.2
5	6.25	300	69.4	7.5	0.625	200	43.3	7.5	0.625	400	_
5	6.25	300	73.6	2.5	0.625	200	62.8	7.5	62.5	200	87.0
5	0.625	300	50.5	5	6.25	300	74.0	2.5	62.5	200	96.0

In this example, the new emphasis is to search for an 18-run $\mathbf{X}_{\text{opt}}$ that maximises the local D-criterion function. Our emphasis is different from the one in the first example, such that the exchanging rule is set to $\max(d_i) > 1.00000001$ in each case. On the one hand, the change of the critical value makes little difference for the discrete optimisation over a small candidate set $\mathbf{\Omega}$. On the other hand, at the expense of more computation and iterative updates, when $n = 18$ and $p = 6$, this rule will suit the continuous optimisation of the function d_i over $\mathcal{X}$, for $i = 1, 2, \ldots, n$. With a loose exchanging rule like this, we could see better updates of the initial random matrix $\mathbf{X}$. Thus the chosen solution $\mathbf{X}_{\text{opt}}$ is prone to be more stable in composition and more efficient with respect to the criterion. This rule will also reduce the discrepancies between those *quasi-replicate runs* in $\mathbf{X}_{\text{opt}}$, the points which are rather close to each other in the variable space $\mathcal{X}$.

As an alternative approximation to the response surface, for instance, one can also fit a full second-order linear model in terms of the transformed variables $x_{\text{S}}, x_{\text{E}}, x_{\text{P}}$, where

$$x_{\text{S}} = \frac{S - 5}{2.5} \in [-1, 1].$$

With the modified Fedorov exchange algorithm, it is convenient to obtain the $\mathbf{X}_{\text{emp}}$ that is D-optimal, the 10×10 information matrix of which does not depend on the parameter values. As in the first example, if we use the $\mathbf{X}_{\text{emp}}$ to run the experiment but then estimate the parameters $\boldsymbol{\theta}$ of the nonlinear model (2.2) (rather than the linear one), the local D-criterion value of $\mathbf{X}_{\text{emp}}$ equals 34.4783. At least, $\mathbf{X}_{\text{emp}}$ is more efficient than the CCD in Table 2.5.

Our interest is to find a D-efficient $\mathbf{X}_{\text{opt}}$ for model (2.2), which will lead to a reasonable approximation to the response surface. For the traditional discrete optimisation, the candidate set $\mathbf{\Omega}$ of 3^3 runs is made up of the distinct variable levels adopted in the reference CCD. It takes less than two seconds to accomplish all the computation. Out of the 100 random tries, the best three solutions $\mathbf{X}$ are similar, each criterion value of which is $\phi_{\text{D}} = 38.8433$. Likewise, when the coordinate exchange approach is used, the subset of candidate coordinates is of size three for each controlled variable. It takes eight seconds to finish 100 tries, as the highest three criterion values are $38.7324, 38.6514, 38.6514$. In this demonstration, the point exchange approach is better in the traditional discrete

optimisation, in which case the required number of random tries can be far less than 100.

Before we implement the new continuous optimisation method, in this demonstration, one more refinement is to start the optimisation at multiple initial points, in addition to the default one (the current values in the current row of $\mathbf{X}$). We can define a coarse point set $\{3.3, 5, 6.7\} \times \{-0.67, 0, 0.67\}^2$ to spread over $\mathcal{X}$, which contains 3^3 initial points for the continuous optimisation. This means that there are in total 28 initial starts under the point exchange approach, as the aim is to maximise the relative improvement function of $v = 3$ decision variables. Meanwhile, under the coordinate exchange approach, there are four initial starts for unidimensional optimisations, which are easier to solve. Certainly there is more computation to do when we take these initial starts, but it will be worth it if we therefore find the local D-efficient $\mathbf{X}_{\text{opt}}$. With this attempt to improve the maximal relative improvement in each step in the iterative update, the computational time is 6641 seconds in the optimised Fedorov exchange algorithm. In comparison, it is 2879 seconds in the optimised coordinate exchange algorithm.

While $\max(d_i) > 1.00000001$ is a suitable rule when we update $\mathbf{X}$, there is little difference between the two solutions listed in Table 2.6. After the truncation at four decimal places of the 100 local D-criterion values at their respective local maxima, more than 20 of them are equal to the chosen maximum 41.2246. Thus fewer tries should be sufficient. On average, the optimised Fedorov exchange algorithm requires 6.39 iterative updates to achieve the local convergence of the criterion function. In this case, the expectation of the criterion value is 41.2218, after 6641 seconds of computation. In comparison, the optimised coordinate exchange algorithm requires 7.29 iterative updates on average and the mean of the 100 criterion values is 41.1937. The fewer iterative updates it takes, the more effective the exchange approach works. Here, the mean difference is $7.29 - 6.39 = 0.9$, so the faster coordinate exchange approach is not a bad choice in this example.

Table 2.6: 18-Run Local D-Optimal Design with Continuous Optimisation Using

(a) Point Exchange

S	E	P	S	E	P
2.5	2.7971	200	2.5	62.5	200
2.5	2.7973	200	2.5	62.5	200
2.5	2.7976	200	2.5	62.5	288.88
2.5	19.277	400	2.5	62.5	288.88
2.5	19.279	400	2.5	62.5	400
2.5	20.646	200	2.5	62.5	400
2.5	20.650	200	3.0979	62.5	200
2.5	23.249	288.37	3.1502	31.245	200
2.5	23.252	288.37	3.1502	31.247	200

(b) Coordinate Exchange

S	E	P	S	E	P
2.5	2.7969	200	2.5	62.5	200
2.5	2.7969	200	2.5	62.5	200
2.5	2.7974	200	2.5	62.5	288.87
2.5	19.276	400	2.5	62.5	288.87
2.5	19.278	400	2.5	62.5	400
2.5	20.647	200	2.5	62.5	400
2.5	20.648	200	3.0977	62.5	200
2.5	23.245	288.37	3.1500	31.237	200
2.5	23.250	288.37	3.1502	31.247	200

The simple candidate set is not so good for the discrete optimisation. Differing from the result in the Martins et al. example, the implementation of the continuous optimisation can make an apparent improvement in $\mathbf{X}_{\text{opt}}$. With respect to the $\mathbf{X}_{\text{opt}}$ in Table 2.6a,

for instance, the previous optimal design found with the discrete optimisation method is $\exp(38.8433/6)/\exp(41.2246/6) \approx 67.24\%$ efficient. Likewise, the baseline $\mathbf{X}_{\text{emp}}$ is 32.49% and the CCD is 20.63% efficient, relative to the optimal design over $\mathcal{X}$. Hence we shall recommend the new continuous optimisation method.

2.5 A New Multistage Exchange Algorithm

2.5.1 Introduction to the New Method

The optimised exchange algorithm uses a lot of tries to find an efficient $\mathbf{X}_{\text{opt}}$, where the experimental runs of diverse initial $\mathbf{X}$ are determined at random. The simple random selection is safe and fair, as long as there are sufficient tries to make the random sample representative. However, that is almost impossible for most experiments since there are in total $n \times v$ coordinate values of $\mathbf{X}$ to be sampled from the continuous design region $\mathcal{X}$. Moreover, the simple random sampling method lacks effectiveness and is wasteful in computation, since one specific random initial $\mathbf{X}$ might not happen to suit our exchange algorithms. As a consequence, after τ random tries, the most efficient solution $\mathbf{X}_{\text{opt}}$ is treated as to be optimal whereas those inferior solutions will be abandoned.

In this section, our aim is to look for an alternative technique which can decrease the normal computational time. In most situations, there is insufficient information about the appropriate candidate experimental runs that will compose an efficient $\mathbf{X}_{\text{opt}}$ and lead to the effective maximisation of the local D-criterion function. The true maximum of the criterion function is unknown. To find some room to improve the random sample of $\mathbf{X}$, our conjecture is that the more efficient are the initial random $\mathbf{X}$, the quicker and smoother can the continuous optimisation method work through the iterative updates and in the end, the more stable and efficient $\mathbf{X}_{\text{opt}}$ is expected to be found.

As such, the idea is to schedule a swift first stage of iterative computation, in order to develop the random initial $\mathbf{X}$ to an efficient intermediate solution $\mathbf{X}_{1\text{st}}$. This is feasible with the traditional discrete optimisation over a sparse but suitable candidate set $\boldsymbol{\Omega}$. We can use just a small number of candidates, as a way to reduce the amount of computation. We can also require the exchanging rule to be strict, in order to avoid some identical matrices $\mathbf{X}_{1\text{st}}$.

This measure is supposed to enhance the subsequent implementation of the continuous optimisation, which requires much more intensive computation in each iterative update. Besides, it is a bit similar to the technique of forward sequential composition of nonrandom initial $\mathbf{X}$, which can also be applied to optimal design of experiments.

For the continuous optimisation in the second stage, some or all of those distinct $\mathbf{X}_{1\text{st}}$ (the duplicates are to be deleted) can act as different new starts. The integrated iterative

update shall be similar to that in the optimised exchange algorithm with a fairly loose exchanging rule. With the continuous nonlinear optimisation of relative improvements over $\mathcal{X}$, each $\mathbf{X}_{1\text{st}}$ can continue to be updated towards a more stable and competitive solution $\mathbf{X}_{2\text{nd}}$. Out of the diverse designs $\mathbf{X}_{2\text{nd}}$, the most D-efficient one will be chosen as $\mathbf{X}_{\text{opt}}$, which is treated as to be optimal.

With an optional third stage of computation to be appended in the end, this is the new *hybrid method* we develop for optimal design of experiments in this thesis, which is expected to work with fewer starts (i.e. the number of distinct $\mathbf{X}_{1\text{st}}$ we select) for the continuous optimisation part. We summarise the new *hybrid Fedorov exchange algorithm* as follows, whereas the alternative *hybrid coordinate exchange algorithm* will be compiled in a similar structure:

Hybrid Fedorov Exchange Algorithm for Nonlinear Models

1st stage: local D-optimal design with discrete optimisation over $\boldsymbol{\Omega}$

1.1 Generate an initial nonsingular $\mathbf{X}$ of n independent random draws (with replacement) out of the set $\boldsymbol{\Omega}$ of N candidate runs defined within $\mathcal{X}$. After the substitution $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, compute matrix $\mathbf{F}$ and the initial value of ϕ_{D}.

1.2 Let index $i = 1$ and flag $\Upsilon = 0$ when each iterative update starts, which will work through all n rows of $\mathbf{X}$ and the corresponding rows of $\mathbf{F}$ in sequential order.

1.3 Use $\boldsymbol{\Omega}$ for the search. At the ith row of $\mathbf{X}$, let $\mathbf{X}_i = \mathbf{X}_{\text{new},i}$ represent the best substitution which leads to the maximum of d_i, the relative improvement in ϕ_{D}. d_i is evaluated for each point exchange proposal between the current $\mathbf{X}_i$ and one of the rows in the matrix $\boldsymbol{\Omega}$.

1.4 If the maximal d_i satisfies the prespecified exchanging rule (e.g. $\max(d_i) > 1.01$), execute the substitution $\mathbf{X}_i = \mathbf{X}_{\text{new},i}$ in $\mathbf{X}$; then update $\mathbf{F}, \phi_{\text{D}}$ and let $\Upsilon = 1$.

1.5 When $i < n$, let $i = i + 1$ for the current iterative update. Return to **step 1.3**.

1.6 When $i = n$ and if $\Upsilon = 1$, return to **step 1.2** and this is the end of one dependent iterative update. Otherwise if $\Upsilon = 0$, save the current $\mathbf{X}$ as an intermediate matrix $\mathbf{X}_{1\text{st}}$.

1.7 Do **steps 1.1-1.6** for τ tries, each with a random initial $\mathbf{X}$. Delete the duplicates of $\mathbf{X}_{1\text{st}}$, so there are τ^* distinct $\mathbf{X}_{1\text{st}}$ left for the continuous optimisation.

2nd stage: local D-optimal design with continuous optimisation over $\mathcal{X}$

2.1 Pick out one unused $\mathbf{X}_{1\text{st}}$ out of the τ^* intermediates in the 1st stage. After the substitution $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, compute matrix $\mathbf{F}$ as well as the starting value of ϕ_{D}.

2.2 Let index $i = 1$ and flag $\Upsilon = 0$ when each iterative update starts, which will work through all n rows of $\mathbf{X}$ and the corresponding rows of $\mathbf{F}$ in sequential order.

2.3 At the ith row of $\mathbf{X}$, replace $\mathbf{X}_i$ with $\mathbf{X}^*_{\text{new},i}$ which is an unspecified candidate run and is made up of v decision variables. Write the function d_i in the algebraic expression. Decide (e.g. at random) one or more initial vectors of $\mathbf{X}^*_{\text{new},i}$, while the default is to use $\mathbf{X}_i$.

2.4 Maximise d_i with an optimisation method such as the Nelder-Mead (or quasi-Newton) method. Output the numeric local optimal solution $\mathbf{X}_{\text{new},i}$ within $\mathcal{X}$, which will substitute $\mathbf{X}^*_{\text{new},i}$. This is the best candidate that can perhaps supplant the current $\mathbf{X}_i$.

2.5 If the maximal d_i satisfies the prespecified exchanging rule (e.g. a loose rule we use is $\max(d_i) > 1.00000001$), execute the substitution $\mathbf{X}_i = \mathbf{X}_{\text{new},i}$ in $\mathbf{X}$; then update $\mathbf{F}, \phi_{\text{D}}$ and let $\Upsilon = 1$.

2.6 When $i < n$, let $i = i + 1$ for the current iterative update. Return to **step 2.3**.

2.7 When $i = n$ and if $\Upsilon = 1$, return to **step 2.2** and this is the end of one dependent iterative update. Otherwise if $\Upsilon = 0$, save the current $\mathbf{X}$ and the value of ϕ_{D}, where $\mathbf{X}$ is one of the matrices $\mathbf{X}_{\text{2nd}}$, the criterion value of which is a local optimum.

2.8 Return to **step 2.1** for $(\tau^* - 1)$ more tries, each with a new distinct $\mathbf{X}_{\text{1st}}$. Upon completing this stage of computation, the exact local D-optimal $\mathbf{X}_{\text{opt}}$ is decided to be the most efficient $\mathbf{X}_{\text{2nd}}$ attained in **step 2.7**, which corresponds to the highest ϕ_{D} out of the τ^* tries in the continuous optimisation.

3rd stage: local D-optimal design with adjustment and reallocation of runs

3.1 For $k = 1, 2, \ldots, v$, determine the *closest distance* $\mathfrak{d}_k$ for the kth factor (or controlled variable). Group all quasi-replicates (i.e. points that are close to each other in distance) in $\mathbf{X}_{\text{opt}}$, so the n independent runs of $\mathbf{X} = \mathbf{X}_{\text{opt}}$ can be divided into n^* homogeneous clusters. While the column order of the v factors is fixed, sort $\mathbf{X}$ in ascending row order with respect to the cluster allocations and the (rounded) variable levels. Reset $k = 1$.

3.2 Let $i = 1$ and $i^* = 1$ be the starting values.

3.3 Pick out the i^*th cluster of $\mathbf{X}$, which corresponds to n_{i^*} quasi-replicates. The aim is to find the best candidate as a substitute for the kth coordinate of each selected quasi-replicate (the maximum of which is $X_{\max,k}$ and the minimum is $X_{\min,k}$). After the substitution, there shall be n_{i^*} exact replicates of the optimal candidate coordinate.

3.4 Create a provisional subset of candidate coordinates $\mathbf{\Omega}_k$, for instance, on the basis of the closest distance, the defined space $[X_{\min,k}, X_{\max,k}]$, and the continuous variable space $\mathcal{X}$. With the subset, do a discrete optimisation for the n_{i^*} quasi-replicates in $\mathbf{X}$ in one step. Denote X_{new} as the best candidate coordinate that maximises ϕ_{D}. Then execute the coordinate substitution for the kth coordinate of each quasi-replicate. Let $i = i + n_{i^*}$.

3.5 Unless $i = n + 1$, let $i^* = i^* + 1$ and return to **step 3.3** .

3.6 Unless $k = v$, let $k = k + 1$ and return to **step 3.2**.

3.7 Create a new candidate set $\mathbf{\Omega}$, consisting of the n^* unique runs of the current $\mathbf{X}$. With this candidate set, do a discrete optimisation for the n experimental runs. As a result of the modified Fedorov exchange algorithm, for instance, let $\mathbf{X}^*_{\text{opt}}$ be the local D-optimal experimental design, which is also the eventual solution of the multistage hybrid method.

The coordinate exchange approach can be installed at the second stage, which will speed up the continuous optimisation if v is large. In contrast, it does not matter which approach we use for the initial discrete optimisation, though the point exchange approach is found to be more efficient in our demonstrations where v is small.

Overall, the new continuous optimisation method is used to complement the traditional discrete optimisation over an imperfect candidate set. Part of this idea can be linked to an *adjustment algorithm* in Donev and Atkinson (1988), which was applied to some second-order linear experiments of quite small numbers of runs. In those cases, Donev and Atkinson (1988) recommended an out-of-date nonsequential approach (e.g. the DETMAX procedure (Mitchell, 1974)) for the discrete optimisation over candidate sets. The optimisation did not work well, so it made sense to adjust the coordinate levels obtained from the discrete optimisation, so as to improve the optimal designs.

Hence, this adjustment algorithm is developed to explore the adjacent variable space outside the finite discrete candidate set, in which case there are numerous candidate points located in $\mathcal{X}$. Likewise, in our multistage hybrid method, there is a third stage of computation, which aims to convert the quasi-continuous coordinates of a specific $\mathbf{X}_{\text{opt}}$ into some rational numbers, as well as to obtain exact replicates of each unique run (i.e. support point) of the experiment. Compared with the adjustment algorithm, we also adjust some coordinate levels after the optimisation. However, we do it for a different purpose (not to improve the efficiencies of optimal designs) and as a result, the number of unique runs in the final solution will not increase. Besides, the steps we follow in the noniterative procedure are also different, where the candidate coordinates for each unique run can be specified in advance.

2.5.2 General Applications of the Algorithm

Practical limitations influence how we can set the controlled variable levels, even though we decide the number of significant figures to show for those coordinates in Tables 2.4 and 2.6. Therefore, the closest distance between levels of a controlled variable indicates the minimum space between two feasible coordinates next to each other, the levels of which must be distinguishable to the experimenters. As we have determined the closest distance, the continuous variable space $\mathcal{X}$ is then redefined into a large set of discrete candidates, located on a grid of small and dense meshes. This is another new definition we introduce in this thesis. For instance, set the closest distance to 0.01 unit for the variable of flow rate in the Box and Draper example. In Table 2.4a, the 10-12th experimental runs can be labelled as quasi-replicates. This means that the space between the maximum and minimum flow rates is $[1.7366, 1.7445] \subset [1.73, 1.75]$. To convert these quasi-replicates in **step 3.4**, the relevant and feasible candidate coordinate subset can be composed as $\mathbf{\Omega}_1 = \{1.73, 1.74, 1.75\}$.

To allocate quasi-replicates into the same cluster as required, sometimes we can check the Euclidean norms or distances between the current experimental points of $\mathbf{X}_{\text{opt}}$. In complex cases, it seems to be sensible to build up these clusters with clever techniques such as hierarchical clustering or k-means clustering. Grouping quasi-replicates can be straightforward if there is a loose exchanging rule, e.g. $\max(d_i) > 1.00000001$ in both examples, for the continuous optimisation to attain $\mathbf{X}_{\text{opt}}$. Here, n^* will also indicate the number of unique runs in the eventual solution $\mathbf{X}^*_{\text{opt}}$ and likewise, n_{i^*} can denote the respective number of replicates of one unique run of the experiment, for $i^* = 1, 2, \ldots, n^*$. A smaller n^* or a certain extent of replication can facilitate the experimentation as well as the estimation of the treatment parameters.

After all quasi-replicates are converted to exact ones, we consider a reallocation of replication for the n^* unique runs included in $\mathbf{\Omega}$. The optimal number of replicates should be recalculated for each unique run, although we expect little improvement in the criterion function. This attempt can benefit the local convergence of the criterion function and was also implemented in Gotwalt et al. (2009). On occasions when the continuous optimisation with sufficient tries works well, the reallocation of replication is unimportant and this process can be skipped. A similar reallocation of runs after each iterative update of the continuous optimisation part can be beneficial in some cases too. If we do that, to save computational time, we can use fewer initial points to commence the continuous optimisation of relative improvements (or updating functions).

We also learn about the randomness that determines the order or sequence of the n rows. The iterative update follows such an order so as to update the experimental units. Because of the dependence between the updates, the computation we do is associated with the row order determined beforehand. The **Default Plan** in all of our demonstrations should follow a "row to column" scheme. With the coordinate exchange approach, for

instance, each iterative update starts from the first row of $\mathbf{X}$ (no matter if it is sorted or unsorted) and tries to update the v separate coordinates in this row. The function for the row increment is $i = i + 1$ until $i = n$, so we will inspect each of the n experimental runs in a **sequential row order**. This row order is unimportant in the discrete optimisation, where the initial $\mathbf{X}$ is random. However, in the subsequent continuous optimisation and also the adjustment stage, we will sort each $\mathbf{X}_{1\text{st}}$ in ascending row order in accord with the controlled variable levels and a fixed column order. As such, the row order influences the iterative update.

There is also the column order of $\mathbf{X}$ to take account of. While the row order is sequential, we choose to use a new **random column order** for each of the tries, for either the discrete optimisation or the continuous optimisation. The iterative update will follow this random column order of controlled variables until we attain a local optimum of the criterion function. A fixed column order will be followed for the final adjustment, when there is just one iterative update.

On the other hand, it is viable to do more frequent randomisation. For instance, we can conceive a **Iteration Random Plan**, which will let us randomise the row and column orders at the start of each iterative update (except in the adjustment stage). As far as we can discern, in local D-optimal design of experiments, the hybrid method will not be much improved (results not shown) in the demonstration. Hence too frequent randomisation is not important in these algorithms.

To summarise, the new hybrid method can save us some computational time and perhaps make a substantial improvement in the solutions of optimal design of experiments. The performance of the exchange algorithm relies on the initial discrete optimisation of the candidate runs in $\mathbf{\Omega}$. If $\mathcal{X}$ is a cuboid, it is a common rule to make at least both limits of each controlled variable to be two of the candidate levels. If we are quite optimistic in finding the appropriate discrete candidates, it is even acceptable to skip the whole discrete optimisation part of the hybrid method. This triggers a shortcut as follows: one can sample each random initial $\mathbf{X}$ from the first discrete candidate set $\mathbf{\Omega}$, instead of the continuous variable space $\mathcal{X}$. With these random initial starts for the continuous optimisation, we can obtain efficient solutions as well.

2.6 Examples Revisited

In our first demonstration of the Box and Draper example, one of the aims is to compare the new algorithms we develop and the two exchange approaches for optimal design of experiments. Here, for the first stage of the hybrid method, the initial discrete optimisation can use the same $3 \times 3 \times 3$ candidate set $\mathbf{\Omega}$ for the modified Fedorov exchange algorithm, which is made up of the variable levels in the reference CCD in Table 2.1. This is a useful candidate set, which will make our hybrid method successful.

With $\tau = 30$ tries, we choose the point exchange approach, for instance. The discrete optimisation shall use a rather strict exchanging rule $\max(d_i) > 1.1$, since a large τ^* is desirable. There are $\tau^* = 30$ distinct intermediate solutions $\mathbf{X}_{1\text{st}}$ in this case.

To examine the exchange approaches, in each case, the continuous optimisation method will start with the same 30 $\mathbf{X}_{1\text{st}}$. We also set the exchanging rule to $\max(d_i) > 1.0001$, with which we can examine the continuous optimisation results in Table 2.7. These two are the interim local D-optimal experimental designs at the moment, before we do some adjustments of their coordinates. With the point exchange approach, the elapsed time till we complete the continuous optimisation is 809 seconds for the 30 tries. The computational time is much less than that of the optimised Fedorov exchange algorithm with 100 tries. It starts with the different $\mathbf{X}_{1\text{st}}$, which are quite efficient already. As a result, out of the 30 tries, the mean number of iterative updates drops to 3.9333 in the continuous optimisation. The highest three criterion values are equal to $-49.5186, -49.5192, -49.5281$ and the mean of the total 30 is -49.5687. Of course the shown results are better in the optimisation, since the mean of the 100 criterion values is -49.7484 when we use the optimised Fedorov exchange algorithm. With the hybrid coordinate exchange algorithm, meanwhile, it takes 786 seconds to complete the 30 tries. $\phi_{\text{D}} = -49.5186, -49.5193, -49.5280$ are the highest three criterion values whereas the mean is -49.5740, a bit inferior to -49.5687. Besides, to finish the continuous optimisation, it takes 4.1 iterative updates on average. Compared with 3.9333, the difference is small so that the coordinate exchange approach is efficient in this demonstration too.

Table 2.7: 24-Run Local D-Optimal Design for Model (2.1): Interim Solution after the Continuous Optimisation Part (2nd Stage) of the Hybrid Method Using

(a) Point Exchange

R	C	T	R	C	T
1.5	1	70	1.7372	4	90
1.5	1	70	3.2893	1	70
1.5	1	81.402	3.3938	1	70
1.5	1	90	5.7226	4	70
1.5	1	90	5.7338	4	70
1.5	1	90	5.7444	4	70
1.5	4	74.238	6	1	84.410
1.5	4	74.250	6	1	84.725
1.5	4	74.350	6	1	84.805
1.7332	4	90	6	1	85.049
1.7333	4	90	6	4	78.600
1.7352	4	90	6	4	79.153

(b) Coordinate Exchange

R	C	T	R	C	T
1.5	1	70	1.7372	4	90
1.5	1	70	3.2892	1	70
1.5	1	81.403	3.3938	1	70
1.5	1	90	5.7227	4	70
1.5	1	90	5.7338	4	70
1.5	1	90	5.7444	4	70
1.5	4	74.238	6	1	84.410
1.5	4	74.250	6	1	84.725
1.5	4	74.350	6	1	84.805
1.7332	4	90	6	1	85.050
1.7333	4	90	6	4	78.600
1.7352	4	90	6	4	79.153

We choose the $\mathbf{X}_{\text{opt}}$ in Table 2.7a for the third stage of the hybrid method, for instance. The respective closest distances for the controlled variables, i.e. flow rate, catalyst concentration and temperature, are decided to be 0.1 unit, 0.1 unit and 1 unit in this

example. These closest distances can be reset on request of experimenters, so it is flexible. After the reallocation of replication (i.e. redetermining the optimal numbers of replicates) in the last step, the eventual solution $\mathbf{X}^*_{\text{opt}}$ is shown in Table 2.8, which is considered to be the local D-optimal design. It is improbable that the hybrid Fedorov exchange algorithm could have missed a greatly superior $\mathbf{X}$. The new maximal criterion value is -49.5116 which is even higher than -49.5186 of $\mathbf{X}_{\text{opt}}$. Besides, there are $n^* = 8$ unique runs in $\mathbf{X}^*_{\text{opt}}$, in contrast to 11 unique runs in Table 2.3 that is found as a result of the traditional discrete optimisation.

Table 2.8: 24-Run Local D-Optimal Design for Model (2.1) with Hybrid Method

R	C	T	R	C	T	R	C	T	R	C	T	R	C	T	R	C	T
1.5	1	70	1.5	1	90	1.5	4	74	1.7	4	90	5.8	4	70	6	1	85
1.5	1	70	1.5	1	90	1.5	4	74	1.7	4	90	5.8	4	70	6	1	85
1.5	1	90	1.5	4	74	1.7	4	90	3.3	1	70	5.8	4	70	6	1	85
1.5	1	90	1.5	4	74	1.7	4	90	3.3	1	70	6	1	85	6	4	79

In the Mountzouris et al. example, the exchanging rule is the same $\max(d_i) > 1.1$ for the discrete optimisation with the point exchange approach. We use the much looser rule $\max(d_i) > 1.00000001$ for the following continuous optimisation, since the aim is to increase the maximal criterion value as much as possible. In this demonstration, we also use multiple initial points to improve the continuous optimisation solution in each step of the iterative update. These initial points were defined as in the previous demonstrations when the optimised exchange algorithms were run, which we can also change for different examples.

After the initial discrete optimisation, we obtain $\tau^* = 4$ distinct $\mathbf{X}_{\text{1st}}$, each of which is made up of 18 runs from the 24-run candidate set $\mathbf{\Omega}$. As we continue to use the point exchange approach for the continuous optimisation, the elapsed computational time is 284 seconds until we find the solution $\mathbf{X}_{\text{opt}}$. Here, the four criterion values are all approximately 41.2246 whereas the number of completed iterative updates is 6.75 on average. Instead, when the coordinate exchange approach is used, the elapsed time is 102 seconds. The time reduces because there are fewer initial points for the continuous optimisation of the relative improvement in the local D-criterion function, which depends on one decision variable in each step of the iterative update. The four criterion values of $\mathbf{X}_{\text{2nd}}$ are listed as $41.2246, 41.2171, 41.2039, 41.1301$, each of which requires seven iterative updates to achieve the local maximal value. We set the closest distance to 0.01 for $S \in [2.5, 7.5]$, 0.005 for $E \in [0.625, 62.5]$, and 0.1 for $P \in [200, 400]$. We can continue with the third stage of the hybrid method and the $\mathbf{X}_{\text{opt}}$ found using the point exchange approach. The number of unique runs is $n^* = 9$, so the total number of exact replicate runs is $(18 - 9) = 9$. The criterion value of the optimal design $\mathbf{X}^*_{\text{opt}}$ in Table 2.9 equals 41.2246. It is similar to the solutions provided in Table 2.6.

Table 2.9: 18-Run Local D-Optimal Design for Model (2.2) with Hybrid Method

S	E	P	S	E	P	S	E	P
2.5	2.795	200	2.5	20.645	200	2.5	62.5	288.9
2.5	2.795	200	2.5	23.25	288.4	2.5	62.5	400
2.5	2.795	200	2.5	23.25	288.4	2.5	62.5	400
2.5	19.275	400	2.5	62.5	200	3.1	62.5	200
2.5	19.275	400	2.5	62.5	200	3.15	31.25	200
2.5	20.645	200	2.5	62.5	288.9	3.15	31.25	200

2.7 Discussion and Recommendations

2.7.1 Nonlinear Multifactor Experiments

As a special class of nonlinear models, the characteristics and statistical interpretation of empirical linear response surfaces are much more familiar to experimenters. As long as there is a sufficient number of runs n to allow for the precise parameter estimation, it is not hard to do efficient local D-optimal design of multifactor linear experiments with the traditional discrete optimisation method. In spite of the usefulness of empirical linear models as elucidated in Box and Draper (1987), experimenters can often encounter a mechanistic model or an empirical nonlinear model which does not coincide the common response surfaces. With some justification from relevant scientific theories and/or mathematical derivations, it will be beneficial to fit this nonlinear model. As such, we can achieve a better approximation of the true response surface as well as other experimental purposes. In optimal design for nonlinear models, however, the discrete optimisation cannot work quite well without a suitable candidate set $\mathbf{\Omega}$, which is hard to define. In this context, we turn our attention to the new continuous optimisation method proposed in this chapter. Based on that, a multistage hybrid exchange algorithm is also developed to search for efficient optimal designs of experiments.

Two main challenges must be tackled though: 1. to find the suitable form of the model; 2. to make a reliable choice of $\boldsymbol{\theta}^0$. On the first, when there is no theoretical model to be assumed, sometimes it is possible to conceive an empirical nonlinear model which contains some characteristics about the mechanism. In this chapter, we made several assumptions and built up such a hybrid model in the Mountzouris et al. (1999) example, which can fit to the reference data well. In this thesis, we focus on deterministic nonlinear models, the information matrix of which can be derived through a normal procedure. A similar idea will also fit the less common classes of models, such as mixed effects models and even stochastic models (Parker et al., 2015). As to the second challenge we face, it shall often require some relevant reference data in order to derive an appropriate parameter prior. The data can be brought from the literature or collected from pilot experiment(s). As long as we can evaluate the criterion function assumed, the next task is to use an exchange approach to optimise the coordinates of a design $\mathbf{X}$.

2.7.2 Continuous Optimisation

The continuous optimisation is robust when it is applied to determine the coordinates of $\mathbf{X}$. It can take the whole variable space $\mathcal{X}$ into account, when the aim is to maximise or minimise the new criterion function. As it is not limited into a finite candidate set $\mathbf{\Omega}$, we take no risk from the improper definition of the discrete candidates. It also saves us time to compare and validate the results of the traditional discrete optimisation, which will facilitate optimal design of experiments. With this method to update $\mathbf{X}$ in steps, we wish to see some improvement in the solution $\mathbf{X}_{\text{opt}}$ in the new optimised exchange algorithms. Meanwhile potential problems exist. Continuous nonlinear optimisation can be an expensive computational tool and especially when there are multiple initial points to start with, it will slow down the exchange algorithm in the Matlab environment at least. Therefore, in developing the multistage hybrid method for optimal design of experiments, our intent is to make it quicker and more reliable in the maximisation of the local D-criterion function, which is in terms of $\mathbf{X}$. As a further evolution from the optimised exchange algorithm, the new hybrid method can circumvent some ineffective optimisations and updates of $\mathbf{X}$, and thus reduce the amount of computation. In other words, we can do fewer tries in the continuous optimisation part, which will otherwise demand substantial computational effort. To some extent, the performance of the hybrid exchange algorithm relies on the small candidate set $\mathbf{\Omega}$, which should be used at the stage of the discrete optimisation. Meanwhile, the potential influences of generating initial $\mathbf{X}$ at random are weak on the obtained $\mathbf{X}_{\text{2nd}}$, as well as the final solution of the hybrid method $\mathbf{X}^*_{\text{opt}}$. A suitable candidate set makes the discrete optimisation quick and leads to a spectrum of efficient $\mathbf{X}_{\text{1st}}$ for the subsequent continuous optimisation.

In comparison, the traditional discrete optimisation is a more direct method. The numerical evaluation of the local D-criterion function or its updating function is much quicker, even if the model is complex. Even though it looks a bit cumbersome, if the total candidate set is small enough, the discrete optimisation is a time-saving method and demands less tedious computation. As we can see, an improper specification of $\mathbf{\Omega}$ can perhaps lead to bad "optimal" designs of the experiment and therefore waste some resources. Hence, the discrete optimisation requires our careful consideration of the best qualified candidate runs. In contrast, the new continuous optimisation is free of specified candidates so that we can save some trouble in the search for efficient optimal experimental designs. Particularly for those simple differentiable nonlinear models, the local D-criterion function and its updating function are simple as well. This makes the continuous optimisation rapid and efficient in finding the local optimal solutions in iterative updates under the point or coordinate exchange approach.

The point and coordinate exchange approaches are both efficient in most circumstances. However, the Fisher information is the sum of n unit information matrices, each of which corresponds to one experimental run. When the statistical model to be fitted is

nonlinear, each element of the total information matrix can be a function written in terms of some or all of the controlled variable levels. As such, the local D-criterion function can be complex in some cases and it will depend on the nonadditive joint effects of the v controlled variables in the model. With closer associations between the exchanges in the iterative update, the coordinate exchange approach does not value these joint effects, due to its low dimension of optimisation. As we assess the independent effects of updating one coordinate at a time, it often takes more effort to make the criterion function converge and thus find an efficient $\mathbf{X}$ at a stable local optimum of the criterion function. When there are limited candidates for selection and the traditional discrete optimisation is used, this trend is more apparent in our demonstrations. Nevertheless, if we can finish the continuous optimisation well, when the number of controlled variables is not too large, there is little difference in the results between the two exchange approaches.

Unlike the discrete optimisation that will assess each feasible candidate in $\mathbf{\Omega}$, the continuous optimisation will advise us to test multiple initial points (initial values of the decision variables in the specific optimisation function) to find reliable local optimal solutions over $\mathcal{X}$. As shown in our examples, an efficient $\mathbf{X}_{\text{opt}}$ can be found after a number of iterative updates that implement the continuous optimisation. Under the point exchange approach, we tend to do fewer updates of each random initial matrix $\mathbf{X}$ and that will facilitate the local convergence of the criterion function. In comparison, the coordinate exchange approach brings about more reliable local optimal solutions to maximise the relative improvement function (or updating function) for the local D-criterion. Therefore we can use fewer initial points for the unidimensional optimisation of one coordinate of $\mathbf{X}$ each time in the iterative update.

Our recommendation is the hybrid Fedorov exchange algorithm when the number of controlled variables v is small in the nonlinear model (e.g. smaller than four). Otherwise, when the continuous optimisation of each experimental run (which will contain v coordinates as the decision variables for multidimensional optimisation) is difficult and the computational cost is expected to be high, we shall then recommend the more flexible coordinate exchange approach. In addition, an idea for compromise is to use the hybrid Fedorov exchange algorithm first but switch to use the coordinate exchange approach in the middle of multiple iterative updates.

2.8 Optimal Design of Experiments in Blocks for Multifactor Nonlinear Models

In this section, we will review the reference CCD in Table 2.1, in which there are four blocks containing the 24 experimental units. Block allocation is a useful tool when not all experimental observations can be completed under similar circumstances. There could be some latent environmental noises (e.g. time effects, nonrandom measurement

errors) that disturb independence of different observations, even if we follow a random order of runs for experimentation. All in all, optimal design of experiments in blocks is an important issue to discuss in this case, as the exchange algorithm should also be adapted in computation and iterative updates.

Our demonstration is for the Box and Draper example where the mechanistic model can be fitted to the experimental data. The new consideration is to include simple fixed block effects in the model, which can explain part of the error sum of squares. In other words, we expect the block allocation to influence the variation in the observed response, which is unknown prior to the experiment. The response varies from block to block in this case and the diversities between blocks are for several reasons, e.g. repeated response measurements, outside environmental changes, some specific operations in the experiment, and so on.

Note that we will not address mixed effects model where we can assume random block effects. There are $n = 24$ experimental units so that there is an invariant block size of six units. We can add an implicit common intercept β_0 to model (2.1) but since the model is nonlinear, least squares estimation would interpret this hidden effect in the fitted mechanistic function. To be more specific, the estimator of θ_0 can act as the overall common intercept of the model, so β_0 is redundant.

We define four artificial terms to represent the fixed block effects in the mechanistic model as β_b, for $b = 1, 2, 3, 4$. The constraint can be either $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$ or $\beta_b = 0$ if the bth block is treated as the baseline. θ_0 explains the common intercept and must be kept in the model. The block label is a qualitative factor. For instance, to define the constraint $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$, in the experimental design $\mathbf{X}$, we should use four columns for the respective block indicators. The trick is to append an extra row $(1, 1, 1, 1, 0, 0, 0)$ on the bottom of $\mathbf{X}$, which corresponds to the constraint equation (the expected response is zero). This constraint lets us estimate the overall fixed block effects in contrast to the overall mean responses.

The more convenient measure is to decide a baseline block and set its fixed block effect to zero (e.g. $\beta_1 = 0$). Hence it is confounded with the "common intercept" estimate $\hat{\theta}_0$. As such, we can use three indicator variables to represent the other three blocks. There are also three columns to be added to $\mathbf{X}$. Here we can compare their fixed block effects to the baseline block effect, as the fitted model can show the respective differences. This is demonstrated as follows, where we use the constraint $\beta_1 = 0$.

There are three extra unknown parameters (i.e. the ones that correspond to the three indicator variables) to be estimated. Under this new structure, the mechanistic model (2.1) can include the fixed block effects such that

$$\eta = \beta_b + \frac{c^{\theta_1}\,\theta_0\, r\, \mathrm{e}^{\theta_2\,(\mathfrak{x}-0.001x)}}{\left(r + c^{\theta'_1}\,\theta'_0\,\mathrm{e}^{\theta'_2\,(\mathfrak{x}-0.001x)}\right)\,\left(r + c^{\theta_1}\,\theta_0\,\mathrm{e}^{\theta_2\,(\mathfrak{x}-0.001x)}\right)} + \varepsilon, \text{ where } \varepsilon \sim N(0, \sigma^2),$$

for $b = 2, 3, 4$. Our main concern is to adjust the continuous optimisation method, in order to search for the optimal blocked experimental design for this model. The baseline block is the first, so we must estimate $\beta_2, \beta_3,$ and β_4, which are used for providing some auxiliary information in interpreting the variation in the observed response η. Their existence does not distort the mechanistic relationship between the dependent and independent variables, which is important in this nonlinear experiment.

The block allocation does not make a vital impact on the mechanistic model, so we are not so interested in the precise estimation of the block effects, but rather the six treatment parameters. The simple local D-criterion will not suit optimal design of this experiment. With the exception of the estimator of the hyperparameter σ^2, the constant variance of errors, the main experimental purpose is to minimise the Generalised Variance of the six treatment parameter estimators, which are also involved in (2.1).

As the standard errors of the block effect estimators are not of concern, the attention is drawn to the Ds-criterion (see Atkinson et al. (2007) for an introduction) or even the (DP)s criterion, which can be used for optimal experimental design in this situation (Gilmour and Trinca, 2012b). Regardless of σ^2, there are also nine columns (of first-order derivatives with respect to the model parameters) in the matrix $\mathbf{F}$. The complete 9×9 information matrix of the mechanistic model can therefore be partitioned and rewritten as a 2×2 block matrix as

$$\mathbf{F}^{\mathrm{T}}\mathbf{F} = \begin{pmatrix} \{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{11} & \{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{12} \\ \{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{21} & \{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{22} \end{pmatrix} = ((\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1})^{-1}.$$

Let submatrix $\{(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\}_{11}$ indicate the square variance-covariance matrix for the three estimators of the block effects (or nuisance parameters). Likewise, $\{(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\}_{22}$ represents the $(9-3) \times (9-3)$ variance-covariance matrix for the estimates of the treatment parameters from (2.1), which depends on the block allocation. Hence, in this example, the aim of the Ds-criterion is to minimise the determinant

$$|\{(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\}_{22}| = 1/|(\{(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\}_{22})^{-1}| \neq 1/|\{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{22}|.$$

A general algebraic result for the inverse of a 2×2 block matrix is

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{U}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{U}^{-1} \\ -\mathbf{U}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{U}^{-1} \end{pmatrix},$$

where $\mathbf{U} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ is a matrix. In this case, $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ is a 9×9 block matrix. While the 3×3 submatrix $\mathbf{A}$ is invertible, the determinant of the information matrix is

$$|\mathbf{F}^{\mathrm{T}}\mathbf{F}| = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}||\mathbf{U}| = |\mathbf{A}|/|\mathbf{U}^{-1}|.$$

We know $\mathbf{A} = \{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{11}$ and $\mathbf{D} = \{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{22}$, such that

$$\mathbf{U}^{-1} = \{(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\}_{22}.$$

Hence, the local Ds-criterion aims to maximise the criterion function

$$\phi_{\mathrm{Ds}} = 1/|\{(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\}_{22}| = |\mathbf{F}^{\mathrm{T}}\mathbf{F}|/|\{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{11}|.$$

The determinant of the block effects $|\{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{11}| = 6^3$ is a constant because of the equal and fixed sizes of the four blocks. Therefore, in this example, the simple local D-criterion that maximises the function $\phi_{\mathrm{D}} = |\mathbf{F}^{\mathrm{T}}\mathbf{F}|$ is equivalent to the local Ds-criterion above. In optimal design of this blocked experiment, we shall evaluate and maximise the same updating function as for the local D-criterion.

Nevertheless, the block allocation does affect the Ds-criterion: $\{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{11}$ is fixed as the submatrix of the Fisher information, but the cross-product terms in $\{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{12}$ and $\{\mathbf{F}^{\mathrm{T}}\mathbf{F}\}_{21}$ are dependent on the block labels we allocate to the experimental units. We use the notion "Block" to represent the block label for each experimental unit. In each iterative update, the exchange algorithm will incorporate the simple block interchange steps to allow for the updates of the block labels (i.e. levels of the three indicator variables) in $\mathbf{X}$. It is reasonable to do discrete optimisations to accomplish these interchanges, under the specific constraint that the block sizes must be fixed. In short, it demands more effort to programme the iterative update and compute the results. The point exchange approach is chosen in our demonstration. Besides, we will calculate the local Ds-criterion value rather than the D-criterion value.

We do 30 tries in the hybrid exchange algorithm. Let the exchanging rule be $\max(d_i) > 1.0001$ for the discrete optimisation over the four blocks and the $3\times3\times3$ candidate set for the three controlled variables. Despite the loose rule, there are 28 distinct intermediate designs $\mathbf{X}_{1st}$ left for the continuous optimisation. With the same critical value $(1+10^{-4})$ for the exchanging rule, the total computational time is 858 seconds before the last stage of the hybrid method. Instead, we also do 100 random tries in the optimised Fedorov exchange algorithm, as it takes 3347 seconds to complete the continuous optimisation. The highest three local Ds-criterion values are $-50.9103, -50.9455, -50.9498$, each local maximum is attained after five iterative updates. In comparison, under the faster hybrid exchange algorithm, the highest three values are $-50.9060, -50.9168, -50.9205$ and each takes four iterative updates to attain, which show a bit improvement in the results. As we can see, the best solutions $\mathbf{X}_{\mathrm{opt}}$ are shown in Tables 2.10-2.11, where the first column is for the block label. Note that we also show the scaled variable x_{T} in the last column, instead of the untransformed one (i.e. temperature). To examine the support points of these experimental designs, the multiple column sorts in the two tables are based on the three controlled variables in advance of the block label.

Table 2.10: 24-Run Local Optimal Design by Optimised Fedorov Exchange Algorithm

Block	r	c	x	Block	r	c	x
2	1.5	1	2.7548	1	1.8776	4	2.7548
3	1.5	1	2.7548	3	3.0252	1	2.9155
4	1.5	1	2.8176	2	3.0868	1	2.9155
1	1.5	1	2.9155	4	3.3744	1	2.9155
4	1.5	4	2.8733	3	6	1	2.7915
3	1.5	4	2.8785	2	6	1	2.7921
2	1.5	4	2.8789	1	6	1	2.8003
1	1.5	4	2.9155	4	6	4	2.8230
1	1.6137	1	2.7548	1	6	4	2.8683
4	1.6794	4	2.7548	2	6	4	2.9155
3	1.6925	4	2.7548	3	6	4	2.9155
2	1.7028	4	2.7548	4	6	4	2.9155

Table 2.11: 24-Run Local Optimal Design by Hybrid Fedorov Exchange Algorithm: Interim Solution after the Continuous Optimisation Part (2nd Stage)

Block	r	c	x	Block	r	c	x
2	1.5	1	2.7548	4	3.2259	1	2.9155
3	1.5	1	2.7548	3	3.2520	1	2.9155
4	1.5	1	2.7548	2	3.2878	1	2.9155
1	1.5	1	2.8341	1	5.0265	4	2.9155
1	1.5	1	2.9155	2	6	1	2.7886
2	1.5	4	2.8782	4	6	1	2.7895
4	1.5	4	2.8784	3	6	1	2.7955
3	1.5	4	2.8868	1	6	1	2.7972
3	1.6862	4	2.7548	3	6	4	2.8444
2	1.7509	4	2.7548	1	6	4	2.8579
4	1.7734	4	2.7548	2	6	4	2.9155
1	1.8582	4	2.7548	4	6	4	2.9155

There is not as much resemblance as we expect between the two tables. This hints that we should have chosen a critical value smaller than $(1 + 10^{-4})$ for the continuous optimisation. We continue the hybrid method with the more efficient $\mathbf{X}_{\text{opt}}$ in Table 2.11, whereas the closest distance for each controlled variable remains the same. After the reallocation of replication and block labels, the final solution $\mathbf{X}^*_{\text{opt}}$ (Table 2.12) adopts just $n^* = 8$ unique runs. The new local Ds criterion value is -50.8820, which even increases from -50.9060. With respect to the Ds-optimal $\mathbf{X}^*_{\text{opt}}$ in this example, the 24-run CCD in Table 2.1 (of which $\phi_{\text{Ds}} = -54.3019$) is 56.87% efficient. The local Ds-criterion aims to minimise the Generalised Variance of the six treatment parameter estimators as well as the volume of their joint confidence region. We can expect the solution in Table 2.12 to fulfil the criterion for this blocked experiment.

Blocks 2-4 are identical, each containing six identical runs for the experiment. However, Block 1 is in a somewhat different structure from the other blocks. The number of degrees of freedom for the model residual is $24 - 6 - 3 = 15$, since in total there are nine parameters to be estimated. Six of them are treatment parameters and the rest

Table 2.12: 24-Run Local Ds-Optimal Design of the Experiment in Blocks

Block	r	c	T	Block	r	c	T	Block	r	c	T	Block	r	c	T
1	1.5	1	70	2	1.5	4	74	4	1.8	4	90	4	6	1	85
1	1.5	1	70	3	1.5	4	74	2	3.3	1	70	2	6	4	70
1	1.5	1	90	4	1.5	4	74	3	3.3	1	70	3	6	4	70
2	1.5	1	90	1	1.8	4	90	4	3.3	1	70	4	6	4	70
3	1.5	1	90	2	1.8	4	90	2	6	1	85	1	6	4	78
4	1.5	1	90	3	1.8	4	90	3	6	1	85	1	6	4	78

are relevant to the block effects. As we should deduct three degrees of freedom from the pure error (these are for the block effects), the number of degrees of freedom for the lack of fit is $(8+3)-6-3=2$ whereas for the pure error, it is $24-(8+3)=13$. Hence, there are lots of replicate runs and the local Ds criterion is inclined to distribute these replicates across the four blocks. This will be helpful for the contrasts of different treatments (or experimental runs) and also for the comparison of the block effects.

2.9 Optimal Candidates for Optimal Design

In this brief section, we would like to make some tentative empirical inferences for the determination of the candidate levels for one of the independent variables in the Box and Draper example. With respect to the local D-optimal design $\mathbf{X}^{*}_{\text{opt}}$ (Table 2.8) over the continuous region $\mathcal{X}$, for instance, our specific interest is in the catalyst concentration, the two unique coordinate levels of which are the upper and lower limit of $C \in [1,4] \in \mathcal{X}$. To illustrate this result, first, the $n \times p$ design matrix $\mathbf{F}$ is made up of six first-order derivatives of $f(\mathbf{X},\boldsymbol{\theta})$ with respect to each of the p treatment parameters in $\boldsymbol{\theta}$. As we use the parameter prior for the substitution $\boldsymbol{\theta}=\boldsymbol{\theta}^{0}=\tilde{\boldsymbol{\theta}}$, in terms of the three variables, the first-order derivative functions $f'_1, f'_2, \ldots, f'_6$ can be simplified as

$$f'_1=\frac{\partial f}{\partial \theta_0}=\frac{C^{\theta_1}\,R^2\,\mathrm{e}^{\theta_2 \mathfrak{x}}}{\left(R+C^{\theta'_1}\,\theta'_0\,\mathrm{e}^{\theta'_2 \mathfrak{x}}\right)\,\left(R+C^{\theta_1}\,\theta_0\,\mathrm{e}^{\theta_2 \mathfrak{x}}\right)^2};$$

$$f'_2=\frac{\partial f}{\partial \theta_1}=\frac{C^{\theta_1}\,R^2\,\theta_0\,\mathrm{e}^{\theta_2 \mathfrak{x}}\log(C)}{\left(R+C^{\theta'_1}\,\theta'_0\,\mathrm{e}^{\theta'_2 \mathfrak{x}}\right)\,\left(R+C^{\theta_1}\,\theta_0\,\mathrm{e}^{\theta_2 \mathfrak{x}}\right)^2};$$

$$f'_3=\frac{\partial f}{\partial \theta_2}=\frac{C^{\theta_1}\,R^2\,\theta_0\,\mathrm{e}^{\theta_2 \mathfrak{x}}\mathfrak{x}}{\left(R+C^{\theta'_1}\,\theta'_0\,\mathrm{e}^{\theta'_2 \mathfrak{x}}\right)\,\left(R+C^{\theta_1}\,\theta_0\,\mathrm{e}^{\theta_2 \mathfrak{x}}\right)^2};$$

$$f'_4=\frac{\partial f}{\partial \theta'_0}=\frac{-C^{\theta'_1}\,C^{\theta_1}\,R\,\theta_0\,\mathrm{e}^{\theta'_2 \mathfrak{x}}\,\mathrm{e}^{\theta_2 \mathfrak{x}}}{\left(R+C^{\theta'_1}\,\theta'_0\,\mathrm{e}^{\theta'_2 \mathfrak{x}}\right)^2\,\left(R+C^{\theta_1}\,\theta_0\,\mathrm{e}^{\theta_2 \mathfrak{x}}\right)};$$

$$f'_5=\frac{\partial f}{\partial \theta'_1}=\frac{-C^{\theta'_1}\,C^{\theta_1}\,\theta'_0\,R\,\theta_0\,\mathrm{e}^{\theta'_2 \mathfrak{x}}\,\mathrm{e}^{\theta_2 \mathfrak{x}}\log(C)}{\left(R+C^{\theta'_1}\,\theta'_0\,\mathrm{e}^{\theta'_2 \mathfrak{x}}\right)^2\,\left(R+C^{\theta_1}\,\theta_0\,\mathrm{e}^{\theta_2 \mathfrak{x}}\right)};$$

$$f_6' = \frac{\partial f}{\partial \theta_2'} = \frac{-C^{\theta_1'} C^{\theta_1} \theta_1' R \theta_0 \mathrm{e}^{\theta_2' \mathfrak{x}} \mathrm{e}^{\theta_2 \mathfrak{x}} \mathfrak{x}}{\left(R + C^{\theta_1'} \theta_0' \mathrm{e}^{\theta_2' \mathfrak{x}}\right)^2 \left(R + C^{\theta_1} \theta_0 \mathrm{e}^{\theta_2 \mathfrak{x}}\right)}.$$

The reference NLS estimate from Box and Draper (1987) is $\tilde{\boldsymbol{\theta}} = \{\tilde{\theta}_0 = 5.90,\ \tilde{\theta}_0' = 1.15,\ \tilde{\theta}_1 = 0.53,\ \tilde{\theta}_1' = -0.01,\ \tilde{\theta}_2 = 15475,\ \tilde{\theta}_2' = 7489\}$. In relation to the theories behind this mechanistic model, all parameters of $\boldsymbol{\theta}$ should be nonnegative. Besides, either θ_1 or θ_1' can be linked to the order of the chemical reaction, which is often determined as an integer or the fraction of two small integers. As such, $\theta_1 = 0.5, \theta_1' = 0$ are reasonable values when we interpret this mechanism. There is $\tilde{\theta}_1 + \tilde{\theta}_1' \approx \tilde{\theta}_1$. If we write $z = C^{\theta_1}$, then $f_1'; f_3', \ldots, f_6'$ are monotone decreasing functions of $z \in [1, 4^{0.53}]$. As in Box-Cox transformation, as $\alpha \to 0, \log(z)/\theta_1 \to (z^\alpha - 1)/(\alpha\theta_1)$. This means f_2' shall be an increasing function of z within $\mathcal{X}$. Thanks to the similarities between the above derivative functions, in terms of f_1' and f_4',

$$|\mathbf{F}^{\mathrm{T}}\mathbf{F}| = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & \mathbf{D} \end{vmatrix} = |\mathbf{A}\mathbf{D} - \mathbf{B}^2| = |\sum_{i=1}^{n} f'^2_{1i}\mathbf{A}_{\mathbf{c}i} \sum_{i=1}^{n} f'^2_{4i}\mathbf{D}_{\mathbf{c}i} - (\sum_{i=1}^{n} f'_{1i} f'_{4i}\mathbf{B}_{\mathbf{c}i})^2|, \text{ where}$$

$$\mathbf{A} = \sum \begin{bmatrix} f_1'^2 & f_1'f_2' & f_1'f_3' \\ f_1'f_2' & f_2'^2 & f_2'f_3' \\ f_1'f_3' & f_2'f_3' & f_3'^2 \end{bmatrix}; \mathbf{B} = \sum \begin{bmatrix} f_1'f_4' & f_1'f_5' & f_1'f_6' \\ f_2'f_4' & f_2'f_5' & f_2'f_6' \\ f_3'f_4' & f_3'f_5' & f_3'f_6' \end{bmatrix}; \mathbf{D} = \sum \begin{bmatrix} f_4'^2 & f_4'f_5' & f_4'f_6' \\ f_4'f_5' & f_5'^2 & f_5'f_6' \\ f_4'f_6' & f_5'f_6' & f_6'^2 \end{bmatrix}.$$

$\mathbf{A}, \mathbf{B}, \mathbf{D}$ are 3×3 submatrices with the sum of the $n = 24$ unit information. For the observation $i = 1, 2, \ldots, n$, we might take the respective common factors $f'^2_{1i}, f'_{1i}f'_{4i}$ and f'^2_{4i} out of these submatrices. In this case, the remainders $\mathbf{A}_{\mathbf{c}i}, \mathbf{B}_{\mathbf{c}i}, \mathbf{D}_{\mathbf{c}i}$ consist of constants plus the term $\log(z)/0.53$. Regardless of the effects of $\log(z)$, the local D-criterion aims at argmax $f'^2_{1i}\ f'^2_{4i}$ and argmin $|f'_{1i}f'_{4i}|$ for individual unit information matrices and the sum of them.

As long as the above approximation holds (i.e. $\tilde{\theta}_1 + \tilde{\theta}_1' \approx \tilde{\theta}_1$), $z_i = 4^{0.53}$ can minimise f'^2_{1i} and f'^2_{3i} in the unit Fisher information, which also leads to the simultaneous maximisation of the other derivative squared. In contrast, if we adopt the lower limit $z_i = 1$ for the ith observation, it corresponds to the maximum of f'^2_{1i} (and f'^2_{3i}) as well as the minimum of the rest functions. All derivatives are considered to be monotonic functions within $\mathcal{X}$. Hence, one or the other limit of z_i can minimise the absolute of a specific cross-product element such as $|f'_{1i}f'_{4i}|$. In short, the two candidate coordinate levels are sufficient for the local D-criterion, the numbers of replicates of which are found to be equal in Table 2.8. Generalisation of these findings are more of interest and we start with some algebraic properties as follows:

$$\begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{vmatrix} = |\mathbf{A}||\mathbf{D}|; \quad |\mathbf{F}^{\mathrm{T}}\mathbf{F}| = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & \mathbf{D} \end{vmatrix} = |\mathbf{A}||\mathbf{D} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}|.$$

Let $\mathbf{A} = \sum f_1'^2$ be the sum of the derivative squared with respect to the first parameter in $\boldsymbol{\theta}$. In terms of the $(p-1) \times (p-1)$ submatrix $\mathbf{D}$, $|\mathbf{F}^{\mathrm{T}}\mathbf{F}| = |\sum f_1'^2\mathbf{D} - \mathbf{B}^{\mathrm{T}}\mathbf{B}|$ where

$\mathbf{B}$ consists of the $(p-1)$ cross-products in association with f_1'. While we are looking for the maximum of $\sum f_1'^2$, the minimisation of the absolute values of the elements in $\mathbf{B}$ is also desirable since that will increase the magnitude of the determinant.

As it goes, we can continue with the submatrix $\mathbf{D}$ to find all requirements for a $\mathbf{X}$ to achieve the local D-criterion. On the one hand, the aim is to maximise the diagonal elements of the Fisher information matrix. A special case is the so-called local Trace-criterion, which aims for the maximisation of the trace of this matrix. In fact, the diagonal elements of an information matrix can be called the *effective sample sizes* corresponding to the parameters of the model. The effective sample size decides how much information will be contained within the planned experiment. It is a simple measure of the rough D-efficiencies of different $\mathbf{X}$. On the other hand, however, the absolute values of the cross-product elements of the information matrix should be minimised at the same time, which makes it hard to determine some optimal candidate runs.

In the simplest cases when the derivatives are orthogonal to each order (e.g. in a first-order linear model), there is $|\mathbf{F}^{\mathrm{T}}\mathbf{F}| = \sum f_1^2 |\mathbf{D}|$ and it simplifies the optimisation tasks. In other cases, we have to exploit some further information about the model derivatives to decide the optimal candidates with respect to a specified criterion. When there is $v = 1$ independent variable in the model, sometimes it is not hard to do so under the local D-criterion. For instance, Dette and Biedermann (2003) derived and illustrated the continuous local D-optimal design for the nonlinear Michaelis-Menten model in biochemical kinetics, which is on the basis of the General Equivalence Theorem in Kiefer and Wolfowitz (1960). However, unless we can derive all support points in advance, it is recommended to use the new continuous optimisation method to search for efficient (if not optimal) $\mathbf{X}$.

2.10 Appendix: Derivation of the Nonlinear Multifactor Model

In the Mountzouris et al. example, one of our interests is in the functional relationship between the response (i.e. the substrate conversion rate) ξ and the (initial) substrate concentration S. The mechanism of this reaction is unknown but there are two reagents: the substrate and the enzyme. Hence, it is reasonable to consider the first step of the whole mechanism to be reversible and bimolecular: the small molecules of the dextran will attach to the active sites of the endodextranase and as a result, this bond can form one molecule of an intermediate complex. To learn this reaction in the experiment, we should think about the mixture of the two reagents.

Consider the first extreme scenario when even the maximal substrate concentration (i.e. $S = 7.5$) is far too low relative to the fixed enzyme concentration E. As the chemical

reaction just starts, the substrate consumption will be quick at a high and stable rate. As such, the conversion is then close to 100% at the response measurement time t (it is often after the reaction almost ceases). At time t, the instantaneous substrate concentration shall be close to a constant so that the expected conversion rate is $\mathbb{E}(\xi) \to 100(S-a)/S$ where a is a nonnegative constant. As we can infer from the reference data in Table 2.5, this extreme scenario will not be the case. Consider the opposite extreme scenario when even the minimal substrate concentration (e.g. $S = 2.5$) is an apparent overdose in the reaction. In this case, the substrate consumption will be slow and it can reduce fast over the elapsed time in the course of the reaction. Hence, we can assume the amount of the converted substrate to be a constant b, such that the expected conversion rate is $\mathbb{E}(\xi) \to 100b/S$.

To describe the old experimental data in Table 2.5, a tradeoff can be found between the two impractical scenarios. Under the theoretical assumption of the ideal solution (or ideal mixture), overall, $\mathbb{E}(\xi)$ should be a decreasing function of the substrate concentration within $\mathcal{X}$ (not under the first extreme scenario described in the above paragraph). It is also realistic to consider the model function in terms of the substrate concentration to be concave, which will not be the case under the second extreme scenario. When the substrate concentration is low, the slope of the model functional curve will be small. As the substrate concentration increases, however, the expected conversion will decrease fast so that we will move towards the second extreme scenario described. For an empirical approximation to the response surface in Mountzouris et al. (1999), we envisage a nonlinear model function

$$\mathbb{E}(\xi_i) = \frac{\gamma_1 S_i}{\gamma_2 + S_i}, \text{ for } i = 1, 2, \ldots, n,$$

where γ_1 is a nonnegative constant and γ_2 is a negative one. When the substrate concentration is low, this function suits the first extreme scenario well, except that it is a decreasing function. When the substrate concentration is in the higher range, this function suits the second extreme scenario more, except that the function is concave rather than convex. As such,

$$\frac{\xi_i}{100 - \xi_i} = \frac{\gamma_1' S_i}{\gamma_2' + S_i} + \varepsilon_i', \text{ for } i = 1, 2, \ldots, n, \tag{2.4}$$

is the model which incorporates a transformed response, a residual ε_i' and two constants γ_1', γ_2'. While the initial substrate concentration is fixed, we shall focus on the scaled independent variables x_{E} and x_{P}. As there is no theoretical support for a chemical interpretation of the mechanism, it is common to consider a transformed second-order linear model as the alternative. With the insignificant interaction term $x_{\text{E}}x_{\text{P}}$ dropped from the tentative model, the established model with an exponential transformation on

its empirical linear function is

$$\frac{\xi_i}{100-\xi_i} = \exp(a'_0 + a'_1 x_{\mathrm{E},i} + a'_2 x_{\mathrm{P},i} + a'_3 x^2_{\mathrm{E},i} + a'_4 x^2_{\mathrm{P},i}) + \varepsilon''_i, \tag{2.5}$$

where ε''_i is the residual and $a'_0, a'_1, a'_2, a'_3, a'_4$ are the unknown constants. With the combination of (2.4) and (2.5) for an approximation of the response surface, the nonlinear multifactor model can be written as

$$\frac{\xi_i}{100-\xi_i} = \frac{\exp(a_0 + a_1 x_{\mathrm{E},i} + a_2 x_{\mathrm{P},i} + a_3 x^2_{\mathrm{E},i} + a_4 x^2_{\mathrm{P},i}) S_i}{a_5 + S_i} + \varepsilon_i, \text{ for } i = 1, 2, \ldots, n,$$

where ε_i is the total error and $a_0, a_1, a_2, a_3, a_4, a_5$ are unknown parameters. This so-called hybrid model complies our conjectures and therefore fits the reference experimental data quite well. Hence, it is feasible to use it for a new experiment under a similar environment. With the model fitted to approximate the response surface, we can find a reasonable interpretation of the unknown mechanism of the interest.

Chapter 3

Hybrid Nonlinear Models: Applications to Michaelis-Menten Kinetics

3.1 Research Problem

We are most interested in the mechanistic model that can describe the relationship between the response and the factors (or controlled variables) to be studied in different experiments. These kinds of models are relevant to some scientific theories, as well as the constraints imposed on the reaction mechanism. In applications to biochemical mechanisms, the Michaelis-Menten model (Michaelis and Menten, 1913) is widely used in the fundamental analysis of **enzyme kinetic data**. This model can interpret mechanisms of the simplest two-step kinetics and in this case, it illustrates a unique functional relationship between the initial reaction rate ν (i.e. the response) and the (initial) substrate concentration S. The kinetic data shall be produced in an experiment, where the initial substrate concentration is specified for each experimental run.

In its most common form, the Michaelis-Menten (M-M) model is theoretical and nonlinear, developed as the solution of the set of differential equations with respect to the time course of a reaction. Given a simple M-M mechanism, the initial reaction rate determines the formation speed of the final chemical product in the ongoing reaction, whereas the substrate is the reagent added in the liquid solution of a fixed volume. In each experimental run, this solution shall be a mixture of the substrate and a suitable enzyme that can activate the reaction. In modern experimental environments, the reaction rate ν can be measured within an initial transient time span of the specific reaction, when the substrate depletion is negligible. With the model error term ε_i, the most common

Michaelis-Menten model is expressed as

$$\nu_i = \frac{\nu_{\max} S_i}{k + S_i} + \varepsilon_i \text{ , for the experimental unit } i = 1, 2, \ldots, n.$$

The parameter $\nu_{\max}$ represents the theoretical maximal value of the initial rate and k is named the *Michaelis constant*, both of which are independent of the substrate concentration. As usual, since we use nonlinear least squares (NLS) or maximum likelihood to fit this model and estimate the two treatment parameters, it requires the errors ε_i to be uncorrelated, subject to zero mean and homogeneous variance. At times, the simple M-M model can also be modified and fitted to interpret more complex reactions of which the kinetic properties are to be studied. We will discuss that later in this chapter.

Under the local D-criterion for optimal design of experiments, we wish to minimise the variances of the two nonlinear least squares parameter estimators. This is for a better interpretation of the M-M mechanism, which can be taken as the main experimental purpose. If we assume an infinite number of runs in the experiment, to decide the respective substrate concentration levels, the continuous local D-optimal experimental design can be derived in a few steps (Dette and Biedermann, 2003; Matthews and Allcock, 2004). It specifies $n^* = 2$ support points: one is situated at the maximal substrate concentration level $S_{\max}$ while the other is equal to $S_{\max}k/(2k + S_{\max})$ in the unidimensional variable space $\mathcal{X}$. The latter substrate concentration level depends on the unknown true value of the Michaelis constant k. In this case, the number of treatment parameters equals the number of support points. The continuous local D-optimal design will have the respective weights in association to the two support points being equal. If n is an even number (half of which is an integer too), e.g. 1000, we shall allocate 500 experimental runs to each support point. Most real experiments have a much smaller number of runs, so we should search for an exact optimal design of the experiment instead, when a small integer n is specified beforehand.

On the basis of simple Michaelis-Menten kinetics, in this chapter, we also specialise in more complicated multifactor mechanisms, as often required in practical industrial experiments. As there are some complex nonlinear models, to find the exact optimal designs of experiments, we shall implement the hybrid exchange algorithm developed in Chapter 2, which makes use of the new continuous optimisation method.

When there is a complicated mechanism, the experimental data must be copious to meet the experimenters' requirements and it can take a lot of effort to approximate the response surface. To reduce the scale of experimentation, some choose to fix several controlled factors at one or more specified levels, and thus treat these factors as either constants or categorical variables. As such, the attention is on a simplified mechanism of fewer controlled variables. Besides, after the experimenters alter the level of one categorical variable (if there are any), the comparison of treatments can be made to calculate the difference in the observed response and even in the fitted model.

In fact, we should make full use of the experimental data at hand, to learn some valuable information about such a complicated mechanism. Even if there is no theoretical model we can use, the alternative is to build up a suitable empirical model. As we treat each controlled variable as continuous, for instance, it is viable to approximate the response surface with an overall empirical model. When little is known about the mechanism, it is most common to fit a second-order linear model, or a similar model based on such kinds of linear or close-to linear functions. On the negative side, however, these linear models lack in support from relevant theories.

To overcome this drawback and to best match the experimental data, we should determine the best model to be fitted. The idea is that we can exploit the limited mechanistic information (if it exists) as appropriate. As we show in the next section, even if it is hopeless to find a purely mechanistic model, a hybrid nonlinear model can be developed, for instance, on the basis of Michaelis-Menten kinetics. The logic is that, for experiments where the response is the initial rate and one of the factors is the substrate concentration, Michaelis-Menten kinetics applies when the remaining controlled factors are held constant. To approximate the overall response surface, we contemplate establishing an empirical nonlinear model, which also incorporates such mechanistic information about the kinetics. In comparison to a simple linear model, for instance, it will make the error structure more reasonable and also interpret the real mechanism better. As a result, the fitted model can improve the results in regression in some aspects.

3.2 Models Based on Michaelis-Menten Kinetics

Our emphasis is to advocate the appropriate use of hybrid models in preference to empirical linear models. This will be demonstrated in examples relevant to the simplest Michaelis-Menten mechanisms. In circumstances when the essential information about the mechanism is unavailable or incomplete, a hybrid nonlinear model can be conceived as follows: (1) the (initial) substrate concentration is one factor so that the complicated multifactor mechanism is in relation to Michaelis-Menten kinetics under certain conditions; (2) meanwhile, we can derive an empirical function in terms of the other controlled variables; (3) then we combine Michaelis-Menten kinetics and this empirical function, as the empirical function can replace the constant $\nu_{\max}$ in the Michaelis-Menten equation; (4) At the request of experimenters, we can add (or delete) new (or current) empirical terms in the integrated hybrid model, before the best functional form is decided; (5) in addition, we can transform the controlled variables or even the both sides of the model, depending on the new error structure to be expected.

As a result, the hybrid nonlinear model incorporates more than one controlled variable, the error structure of which can be modified as required. The empirical approximation of this model is expected to improve, at least within the continuous variable space $\mathcal{X}$.

However, no matter how close the fitted response values conform to the raw experimental data, the model is not a mechanistic one. When experimenters impose different constraints on $\mathcal{X}$, it is necessary to reconsider the hybrid model and its practical limits.

Our examples focus on M-M kinetics but hybrid modelling techniques can also be applied to widespread areas in science and medicine. The general ideas we discuss are relevant to the broader use of hybrid models, although the experimental context will influence the suitable closed-form expression of the model. In Chapter 2, we have derived such a nonlinear model for the experimental data in Mountzouris et al. (1999), where we envisage the relation between the conversion rate and the substrate concentration.

We now propose some candidate models based on Michaelis-Menten kinetics. Apart from the substrate concentration S, the independent variables can be scaled (or transformed). We then add them into an empirical linear (or close-to-linear) function which shall be a component of the model. Letting x denote a single scaled variable, the simplest model can be written in terms of just four parameters k, a_0, a_1, a_2 as

$$\nu_i = \frac{(a_0 + a_1 x_i + a_2 {x_i}^2) S_i}{k + S_i} + \varepsilon_i, \tag{3.1}$$

for $i = 1, 2, \ldots, n$, where the errors can be explained as $\varepsilon \sim \mathbb{N}(0, \sigma^2)$ so as to follow an identical normal distribution. Here, the full second-order polynomial in x substitutes the unknown parameter $\nu_{\max}$ in the M-M model. The polynomial function excludes its error term so that it cannot perfectly predict $\nu_{\max}$, when the value of x varies. Due to this more complicated mechanism, the estimate of k in the fitted model cannot be interpreted as the true Michaelis constant.

Likewise, some effort could be made to replace k with another empirical function. Nevertheless, this will lead to a far more complex model which is hard to fit and interpret. In this respect, model (3.1) avoids an overcomplex function since it defines fewer parameters. It seems to be realistic to accept this more concise model.

We shall also consider the error structure of (3.1), which relies on the empirical function as well as the M-M equation. The fundamental assumption is that the error variance is constant across the n observations and independent of the substrate concentration level. Otherwise, model (3.1) should be modified to meet the requirement and thus it makes sense to contemplate more candidate models as alternatives.

The true functional relationship between the response ν and the scaled variable x is uninformative to us. Instead of an empirical linear function to replace $\nu_{\max}$, a different substitution is to use the exponential (or logarithmic, depending on the error distribution) function of the linear combination of some empirical additive terms. For instance, we can write the exponential function as

$$\nu_{\max,i} = \exp(a_0 + a_1 x_i + a_2 {x_i}^2).$$

In some cases, the new exponential function can stabilise the model errors. Moreover, since the maximum initial rate $\nu_{\max}$ is a nonnegative number in Michaelis-Menten kinetics, the exponential transformation of a second-order linear function is more or less reasonable. As long as the model errors can meet the relevant assumptions, for $i = 1, 2, \ldots, n$, we have

$$\nu_i = \frac{\exp(a_0 + a_1 x_i + a_2 {x_i}^2) S_i}{k + S_i} + \varepsilon_i. \tag{3.2}$$

Compared with (3.1), the new model leads to quite a similar response surface fitted to the experimental data, but it can influence our prediction of the response and interpretation of the mechanism. At times, experimenters also make some speculations about the true error structure of a model. To indicate the potential increase in errors in M-M kinetics, for instance, Cornish-Bowden (2004, chap.14) conceived the model

$$\nu_i = \frac{\nu_{\max} S_i}{k + S_i}(1 + \varepsilon_i),$$

where $\varepsilon \sim \mathbb{N}(0, \sigma^2)$ is a normal random variable.

To deal with the assumed multiplicative error structure of candidate model (3.2) and thus stabilise the error variance, it is common to use an appropriate transform-both-sides (TBS) model instead. For instance, suppose $\varepsilon^{(1)}$ is the additive error due to the empirical exponential approximation of $\nu_{\max}$, we also conceive the multiplicative error $\exp(\varepsilon^{(2)}) - 1$ which is due to the fit of the Michaelis-Menten model. On the foundation of (3.2), the ith experimental observation is equal to

$$\nu_i = \frac{\exp(a_0 + a_1 x_i + a_2 {x_i}^2 + \varepsilon_i^{(1)}) S_i}{k + S_i} \times \exp(\varepsilon_i^{(2)}),$$

where the normal random variables are $\varepsilon^{(1)} \sim \mathbb{N}(0, \sigma_1^2)$ and $\varepsilon^{(2)} \sim \mathbb{N}(0, \sigma_2^2)$ (σ_1^2 and σ_2^2 are the respective error variances). Although the normal distribution of the errors is not an essential assumption under least squares estimation, it will improve our comprehension of the fitted model. When the experimenters consider the above error structure as most appropriate, it is convenient to impose a simple transformation on the both sides of the model. As it goes,

$$\nu_i' = \log(\nu_i) = \log\left(\frac{S_i}{k + S_i}\right) + a_0 + a_1 x_i + a_2 {x_i}^2 + \varepsilon_i, \tag{3.3}$$

where $\varepsilon = \varepsilon^{(1)} + \varepsilon^{(2)} \sim \mathbb{N}(0, \sigma_1^2 + \sigma_2^2)$.

In comparison to (3.2), the TBS model is linear in terms of the independent variable x. No matter which candidate model we accept, it should lead to small lack of fit and valid error assumptions. A simple but comprehensive hybrid model seems to be attractive, which is frugal in the use of parameters. Compared with some rare and overcomplex

candidate models, it lets us focus on the important parameters that are involved in the empirical component of the selected model. As there are fewer parameters to be estimated, we can make the most of the experimental data.

3.3 An Adapted Kinetics Example

In experimental investigation of the reaction mechanism and the kinetic profile, Martins et al. (1999) studied a biochemical reaction which occurred in the permeabilised cells of saccharomyces cerevisiae, which can be used in industrial experiments in pharmaceuticals and medicine. In one of their original experiments, glyoxalase II is the enzyme (or the protein catalyst) that could activate a specific reaction in situ (i.e. in a 2-ml reaction mixture), when a range of S-D-lactoylglutathione (the substrate) concentrations S were determined in the continuous variable space $[0.15, 3]$ mM (the millimolar, unit of concentration). The response is the initial rate ν of GSH (i.e. the desirable product) formation, so M-M kinetics holds.

In the first scenario when the remaining experimental conditions were fixed (i.e. with invariant temperature at 30 °C, the pH at 6.5, and the weight of glyoxalase II protein at approximately 0.03mg), five distinct levels of the substrate concentration were specified and each corresponded to three replicate runs in the experiment. As there were 15 runs in total, the NLS estimation of the M-M model lead to the estimates $\tilde{\nu}_{\max} = 0.000765$ (± 0.000049) mM min^{-1} and $\tilde{k} = 0.36$ (± 0.09) mM. If the small constant $\tilde{k}$ is the unbiased estimate, at the higher range of the substrate concentration, the fitted curve of the M-M model would approach a horizontal line. This is the first half of the reference experiment.

Meanwhile, the second step of the Michaelis-Menten mechanism is of first order and is irreversible (see Chapter 5 for explanation in detail). When the substrate concentration is under control, a mechanistic equation can be established such that $\nu_{\max}(E_i) \approx a_1 E_i$, where $\nu_{\max}(E)$ is a function of the concentration (or weight) of protein E. In another scenario in Martins et al. (1999), while the temperature and the pH were under control, the substrate concentration had been fixed at 1.5 mM. As such, five distinct levels of the protein weight were specified in the variable space $[0.02, 0.12]$ mg (milligram) and each corresponded to three replicates in the experiment. In this second half of the experiment, there were also 15 runs in total. The estimate of the slope was $\tilde{a}_1 = 0.0204$ (± 0.0006) mM min^{-1} mg^{-1}, along with a small standard error.

The fitted model will be superior if the two subsets of data can be incorporated into a master dataset of the two factors and the 30 experimental observations. In this case, the full M-M model can be fitted to interpret the whole mechanism, where we postulate

an overall additive error. The integrated form of the Michaelis-Menten model is

$$\nu_i = \frac{a_1 E_i S_i}{k + S_i} + \varepsilon_i, \tag{3.4}$$

which is similar to the hybrid nonlinear model (3.1). For $i = 1, 2, \ldots, n$, the errors can be explained as $\varepsilon_i \sim \mathbb{N}(0, \sigma^2)$, as long as the NLS assumptions are satisfied. The local D-optimal design of an experiment for (3.4) is no harder than that for the simpler M-M model in terms of the substrate concentration only. Under the local D-criterion, the protein weight E should be in the maximum for all the experimental runs. The most appropriate model is identified as above, but in studies of unknown multifactor mechanisms, it is difficult to find adequate and detailed information to make a mechanistic postulate. Therefore, in the example, we are to assume that the appropriate mechanistic model (3.4) is unknown. Under this premise, an alternative is to derive and use a hybrid nonlinear model that will suit the data, for which we construct a local optimal design of a 30-run new experiment.

Three candidate models are offered in (3.1)-(3.3), which are in terms of the substrate concentration and an unspecified variable x. If we intend to select (3.2), according to the unknown mechanistic model, the best variable transformation shall be $x = \log(E)$. In the empirical component of the model, the constant term and the quadratic term can be deleted from the exponential function. However, in this example, $x = (E - 0.07)/0.05 \in [-1, 1]$ is the linear transformation we prefer. The variable x falls in the standardised space $[-1, 1]$, so the whole variable space is a rectangle $\mathcal{X} = [-1, 1] \times [0.15, 3]$, within which the nonlinear model function is smooth and differentiable. Note that it is unwise to do a nonlinear transformation on the substrate concentration, which comes from the theoretical M-M equation.

Details about our new hybrid method, which is efficient in searching for the exact optimal designs of experiments, are described in Chapter 2. The hybrid exchange algorithm involves three stages of computation, and the first two of them should follow an exchange approach in iterative updates. The first stage is for a swift discrete optimisation of each experimental design $\mathbf{X}$ over a small candidate set. Therefore, our first move is to demarcate, for instance, a coarse 3×3 set of discrete candidate runs. As the model in each form is linear or at least *partially linear* in terms of x, we include the conventional candidate levels in the subset $\mathbf{\Omega}_1 = \{-1, 0, 1\}$ for x. For the unscaled substrate concentration, the candidate subset $\mathbf{\Omega}_2 = \{0.15, 1.5, 3\} \subset [0.15, 3]$ is not a bad choice. Meanwhile, we approximate the exact coordinate levels of the $\mathbf{X}$ used from Figure 2 in Martins et al. (1999). Table 3.1 shows this 30-run reference design of the experiment $\mathbf{X}_{\text{ref}}$. It is the baseline when we make some comparisons.

If we fit the mechanistic model (3.4) to the old data, the NLS estimates are $\tilde{a}_1 = 2.422\ (\pm 0.098) \times 10^{-2}$ and $\tilde{k} = 0.3290\ (\pm 0.0696)$. We can use $\tilde{k}$ as the parameter prior for local D-optimal design of the new experiment then. In reference to the results under

Table 3.1: 30-Run Reference Experimental Design

E	S	100ν	E	S	100ν	E	S	100ν
0.023	1.5	0.0425	0.03	0.9	0.051	0.045	1.5	0.075
0.023	1.5	0.0475	0.03	0.9	0.0535	0.0675	1.5	0.13
0.023	1.5	0.0475	0.03	1.5	0.057	0.0675	1.5	0.135
0.03	0.15	0.023	0.03	1.5	0.06	0.0675	1.5	0.1375
0.03	0.15	0.024	0.03	1.5	0.063	0.089	1.5	0.17
0.03	0.15	0.025	0.03	3	0.07	0.089	1.5	0.18
0.03	0.3	0.0355	0.03	3	0.07	0.089	1.5	0.19
0.03	0.3	0.0365	0.03	3	0.07	0.112	1.5	0.225
0.03	0.3	0.037	0.045	1.5	0.07	0.112	1.5	0.225
0.03	0.9	0.0485	0.045	1.5	0.075	0.112	1.5	0.225

the continuous D-criterion in Dette and Biedermann (2003), where the protein weight is fixed in the M-M model,

$$\mathbf{X}_{\text{opt}} = \left\{ \begin{array}{cc} (0.12,\, 0.2698) & (0.12,\, 3) \\ 15 & 15 \end{array} \right\} \tag{3.5}$$

should be the 30-run local D-optimal design for the M-M model (3.4). In the first row of (3.5), the two coordinates in each bracket represent the values of the controlled variables of one support point (or unique experimental run) (E, S), which must be contained in the variable space $\mathcal{X}$. Then the second row shows the number of replicate runs at each support point and in this case, we can have lots of replicates. With experimental data collected from (3.5), a drawback is that we cannot evaluate the lack of fit of the model. This implies that we will be unable to fit candidate models with more than two unknown parameters. Next, we will look at the three simple hybrid models based on M-M kinetics: the Additive Candidate Model (3.1), the Exponential Candidate Model (3.2), and the Transformed Candidate Model (3.3).

3.4 Local D-Optimal Design of Experiments

3.4.1 Model 1: the Additive Candidate Model

In respect of the candidate model (3.1), the empirical substitution of $\nu_{\max}$ in Michaelis-Menten kinetics is a simple second-order linear function. If we select the model for the nonlinear experiment, the local D-criterion shall minimise the variances of the NLS estimators of the unknown parameters $\boldsymbol{\theta} = \{k, a_0, a_1, a_2\}$. We use the dataset in Table 3.1 to obtain the reference parameter estimate that is $\tilde{\boldsymbol{\theta}} = \{0.3191, 0.0016, 0.0012, 0.0001\}$. The adjusted coefficient of determination is 0.991, so the fitted model explains most of the response variation. If the Michaelis-Menten mechanism holds in this example, the quadratic effect of the empirical linear function contributes little to the response. To

determine the baseline level, the local D-criterion value is calculated to be $\phi_{\mathrm{D}} = -9.0594$ for the 30-run reference $\mathbf{X}_{\mathrm{ref}}$. The criterion value depends on the parameter prior $\boldsymbol{\theta}^0$ which is set to equal the estimate $\tilde{\boldsymbol{\theta}}$ from the old data.

We implement the new hybrid exchange algorithm, which can tackle the intensive computation encountered in the iterative search for the local D-optimal design of the experiment. For the iterative update (i.e. the loop we construct for iterative computation) of each $\mathbf{X}$, we are to examine the point exchange approach (Fedorov, 1972) as well as the coordinate exchange approach (Meyer and Nachtsheim, 1995) in the stage of the continuous optimisation. While we can use the former approach to update an entire row in $\mathbf{X}$ in each step of the iterative update, the latter approach is limited to optimising an individual coordinate at a time. In comparison, the number of controlled variables is $v = 2$, so we expect similar computational results from the two exchange approaches.

For proper use of the three-stage hybrid method, it is important to determine a suitable exchanging rule for the iterative update. Here, it is reasonable to set the rule to be $\max(d_i) > 1.1$ for the initial discrete optimisation under the point exchange approach, for instance. As to the following continuous optimisation, a much looser rule $\max(d_i) > 1.00000001$ will encourage the update of rows or coordinates in $\mathbf{X}$. Besides, to improve the local optimal solutions of continuous optimisations, one can use multiple vectors of initial values. In addition to the default vector (i.e. the set of current coordinate values), we use the Nelder-Mead method to examine nine fixed initial starts from a coarse set $\{-0.67, 0, 0.67\} \times \{0.5, 1.5, 2.5\}$. Hence there are ten initial points in total for the v-dimensional continuous optimisation. Meanwhile, under the coordinate exchange approach, we can check four initial values for each controlled variable, three of which are fixed in advance.

Over the finite candidate set $\boldsymbol{\Omega} = \{-1, 0, 1\} \times \{0.15, 1.5, 3\}$, in our demonstration, we do 30 random tries in the discrete optimisation. Thanks to the strict rule $\max(d_i) > 1.1$, there are 30 distinct intermediate optimal solutions $\mathbf{X}_{1\mathrm{st}}$ left. After continuous optimisation of each $\mathbf{X}_{1\mathrm{st}}$ over the variable space $\mathcal{X}$, we can obtain a distinct solution $\mathbf{X}_{2\mathrm{nd}}$. The interim local D-optimal solution $\mathbf{X}_{\mathrm{opt}}$ is selected as the most efficient $\mathbf{X}_{2\mathrm{nd}}$ out of the 30 tries. As we examine both exchange approaches, the results are presented in Table 3.2 where we show four decimal places of each coordinate value (unless it is on the boundaries of $\mathcal{X}$).

As we keep four decimal places of each value, there are three unique coordinate levels for the protein weight. Two of them are on the boundaries, whereas the other level is approximately 0.07, the midpoint in the variable space. Prior to the final stage of the hybrid method, the elapsed time of computation is 559 seconds after continuous optimisation under the point exchange approach. As we finish 30 tries in the continuous optimisation, the mean number of iterative updates is exactly four. Therefore, it is not hard to find a local maximum of the local D-criterion function of $\mathbf{X}$. The highest three

Table 3.2: 30-Run Local D-Optimal Design for Model (3.1): Interim Solution after the Continuous Optimisation Part (2nd Stage) of the Hybrid Method Using

(a) Point Exchange

E	S	E	S	E	S
0.02	3	0.0700	3	0.12	0.2631
0.02	3	0.0700	3	0.12	0.2631
0.02	3	0.0700	3	0.12	0.2631
0.02	3	0.0700	3	0.12	3
0.02	3	0.0700	3	0.12	3
0.02	3	0.12	0.2631	0.12	3
0.02	3	0.12	0.2631	0.12	3
0.0700	3	0.12	0.2631	0.12	3
0.0700	3	0.12	0.2631	0.12	3
0.0700	3	0.12	0.2631	0.12	3

(b) Coordinate Exchange

E	S	E	S	E	S
0.02	3	0.0700	3	0.12	0.2631
0.02	3	0.0700	3	0.12	0.2631
0.02	3	0.0700	3	0.12	0.2632
0.02	3	0.0700	3	0.12	3
0.02	3	0.0700	3	0.12	3
0.02	3	0.0700	3	0.12	3
0.02	3	0.12	0.2631	0.12	3
0.02	3	0.12	0.2631	0.12	3
0.0700	3	0.12	0.2631	0.12	3
0.0700	3	0.12	0.2631	0.12	3

criterion values are all equal to -3.6863, so the point exchange approach works well. In comparison, the reference $\mathbf{X}_{\text{ref}}$ is 26.10% efficient with respect to the $\mathbf{X}_{\text{opt}}$ in Table 3.2a. Instead, when we use the coordinate exchange approach for the continuous optimisation, it takes 612 seconds and 4.3 iterative updates on average. In comparison to the above, the differences are quite small. The most qualified three $\mathbf{X}_{\text{2nd}}$ all lead to the identical criterion value -3.6863.

Here we use the $\mathbf{X}_{\text{opt}}$ in Table 3.2a for the remaining computation, when there are some further refinements to make. For minor coordinate adjustments in $\mathbf{X}_{\text{opt}}$, the closest distance is set to 0.001 mg for the protein weight and 0.01 mM for the substrate concentration. As we finish the hybrid exchange algorithm, there are $n^* = 4$ support points to compose the solution

$$\mathbf{X}^*_{\text{opt}} = \left\{ \begin{array}{cccc} (0.02,\ 3) & (0.07,\ 3) & (0.12,\ 0.26) & (0.12,\ 3) \\ 8 & 8 & 7 & 7 \end{array} \right\}, \tag{3.6}$$

which shall be local D-optimal. The criterion value is $\phi_{\text{D}} = -3.6864$. It does not decrease much from that of Table 3.2a. The number of unique runs is equal to the number of parameters p, so the weights associated with the four support points should be equal under the continuous D-criterion case (assuming $n \to \infty$). This can be proven under the General Equivalence Theorem. As $n = 30$ in (3.6), the respective numbers of replicates are 8, 8, 7, 7, which are close to each other.

The experiment contains a lot of replicates. One can perhaps reduce the number of runs to 12-20, for instance, depending on the scale of the error variance σ^2 and the controlled variables in the model. In (3.6), the three candidate levels of the scaled variable $x = (E - 0.07)/0.05 \in [-1, 1]$ are identical to the conventional levels $\{-1, 0, 1\}$ defined in $\mathbf{\Omega}$. This is because of the potentially weak nonlinear behaviour of model (3.1) as well as the small prior value $\tilde{a}_2 = 0.0001$ for the parameter in front of the empirical quadratic term x^2 in the model.

The true parameter values of the model are unknown prior to real experimentation. Thus we are interested in the parameter sensitivities (i.e. the robustness to the potential variation of the parameter prior values) of (3.6). In reference to the standard errors in fitting the model to the reference dataset, we consider 16 different scenarios $\mathfrak{S}_1, \mathfrak{S}_2, \ldots, \mathfrak{S}_{16}$ shown in Table 3.3, where the true parameter values of the model are specified. Under each scenario, we are able to find a tailor-made local D-optimal $\mathbf{X}$. Under scenarios $\mathfrak{S}_1, \mathfrak{S}_2, \ldots, \mathfrak{S}_8$, the identical solution we found is

$$\mathbf{X}^*_{\text{opt}} = \left\{ \begin{array}{cccc} (0.02,\ 3) & (0.07,\ 3) & (0.12,\ 0.15) & (0.12,\ 3) \\ 8 & 8 & 7 & 7 \end{array} \right\}, \tag{3.7}$$

where the substrate concentration level of the third support point is at the minimum. Under each of the scenarios $\mathfrak{S}_9, \mathfrak{S}_{10}, \ldots, \mathfrak{S}_{16}$, the optimal solution is

$$\mathbf{X}^*_{\text{opt}} = \left\{ \begin{array}{cccc} (0.02,\ 3) & (0.07,\ 3) & (0.12,\ 0.39) & (0.12,\ 3) \\ 8 & 8 & 7 & 7 \end{array} \right\}, \tag{3.8}$$

which is similar to (3.6) too. Except for the first parameter value, the envisaged variation of a_0, a_1, a_2 do not influence the local D-criterion used. In other words, (3.6) is quite robust to the perturbation of the true parameter values. Under scenario $\mathfrak{S}_1$, for instance, (3.6) is 90.27% efficient with respect to (3.7). On the other hand, it is 97.38% efficient with respect to (3.8) under scenario $\mathfrak{S}_9$. Moreover, due to the similarities between (3.6) and (3.7), the relative efficiencies of (3.6) under scenarios $\mathfrak{S}_1, \mathfrak{S}_2, \ldots, \mathfrak{S}_8$ (or $\mathfrak{S}_9, \mathfrak{S}_{10}, \ldots, \mathfrak{S}_{16}$) are found to be the same too.

Table 3.3: Assumed True Parameter Values (Prior Values) under Each of the 16 Postulated Scenarios

	k	a_0	a_1	a_2		k	a_0	a_1	a_2
$\mathfrak{S}_1$	0.1058	0.0014	0.0011	-0.0001	$\mathfrak{S}_9$	0.5324	0.0014	0.0011	-0.0001
$\mathfrak{S}_2$	0.1058	0.0019	0.0011	-0.0001	$\mathfrak{S}_{10}$	0.5324	0.0019	0.0011	-0.0001
$\mathfrak{S}_3$	0.1058	0.0014	0.0014	-0.0001	$\mathfrak{S}_{11}$	0.5324	0.0014	0.0014	-0.0001
$\mathfrak{S}_4$	0.1058	0.0019	0.0014	-0.0001	$\mathfrak{S}_{12}$	0.5324	0.0019	0.0014	-0.0001
$\mathfrak{S}_5$	0.1058	0.0014	0.0011	0.0003	$\mathfrak{S}_{13}$	0.5324	0.0014	0.0011	0.0003
$\mathfrak{S}_6$	0.1058	0.0019	0.0011	0.0003	$\mathfrak{S}_{14}$	0.5324	0.0019	0.0011	0.0003
$\mathfrak{S}_7$	0.1058	0.0014	0.0014	0.0003	$\mathfrak{S}_{15}$	0.5324	0.0014	0.0014	0.0003
$\mathfrak{S}_8$	0.1058	0.0019	0.0014	0.0003	$\mathfrak{S}_{16}$	0.5324	0.0019	0.0014	0.0003

Now we come back to our D-optimal solution (3.6), which is in a simple and interesting structure. The two substrate concentration levels are 3 and 0.26, which are similar to the two optimal coordinate levels in (3.5). Overall, on the one hand, the support points 1,2,4 constitutes an efficient 23-run experimental design for the precise parameter estimation of a_0, a_1, a_2 in the candidate model. On the other hand, the support points 3 and 4 constitutes an efficient 15-run design for the estimation of k. These two points shall be quite stable when the Exponential Candidate Model is considered in the next subsection. Hence, (3.6) is almost like a one-factor-at-a-time D-optimal design of the experiment. To

validate this statement, our approach is to adapt the local Ds-criterion (see Section 2.8 for details) to focus on the estimation of the full set of the parameters except for k. This requires more computation since we have to evaluate and update two determinants in the defined criterion function. The final Ds-optimal solution we obtain is a bit different to (3.6) in this situation, where fewer replicate runs (or experimental units) are allocated to the support points 3 and 4:

$$\mathbf{X}^*_{\text{opt}} = \left\{ \begin{array}{cccc} (0.02,\ 3) & (0.068,\ 3) & (0.12,\ 0.2) & (0.12,\ 3) \\ 10 & 9 & 3 & 8 \end{array} \right\}. \tag{3.9}$$

This is a result we expect from model (3.1). For the record, since the number $n^* = 4$ is the minimum requirement for least squares estimation of unknown parameters, (3.6) is also optimal under the local DP-criterion introduced in Gilmour and Trinca (2012b).

3.4.2 Model 2: the Exponential Candidate Model

Consider the candidate model (3.2), where we assume a different error structure. In this situation, an empirical exponential function in terms of x replaces the second-order linear one in (3.1), in order to substitute the maximum initial rate $\nu_{\max}$ in the theoretical Michaelis-Menten kinetics. This candidate model shall exhibit stronger nonlinear behaviour than (3.1). Meanwhile, the constraint $\nu_{\max} \geq 0$ will be satisfied for all values of x. The same observed response surface in Martins et al. (1999) can be interpreted under the new mechanistic postulate, so a better candidate model and the parameter prior can be determined. However, the 30-run local D-optimal experimental design for (3.2) tends to be more complicated, because of the exponential transformation of the second-order linear function.

If we fit (3.2) to the old data, the adjusted coefficient of determination is 0.994. This shows an increase over the coefficient 0.991 for (3.1). To evaluate the Fisher information, the NLS estimate $\tilde{\boldsymbol{\theta}} = \{\tilde{k}, \tilde{a}_0, \tilde{a}_1, \tilde{a}_2\} = \{0.3122, -6.4086, 0.8383, -0.2861\}$ is taken as the parameter prior $\boldsymbol{\theta}^0$. Here, the candidate model is more complex because $\tilde{a}_2 = -0.2861$ deviates from zero, unlike the estimate of a_2 in the fitted (3.1). We shall look at the interpretation of Michaelis-Menten kinetics (when x is a constant). The parameter prior of k equals 0.3122 for the new candidate model, which is quite close to the value 0.3191 for (3.1) and the value 0.3290 for the Michaelis-Menten model (3.4) (which we assume to be unknown). After all, the exponential transformation has little influence on the curvature and shape of the fitted response surface, despite the error structure being different. This indicates that both (3.1) and (3.2) can be reliable approximations to the mechanistic model.

While the model specification should be modified according to (3.2), the same input arguments (e.g. discrete candidate set) are supplied to the hybrid exchange algorithm. The interim solutions after the continuous optimisation over the variable space $\mathcal{X}$ are

illustrated in Table 3.4. With the point exchange approach, the computational time is 999 seconds and the mean number of iterative updates is 7.0667. Compared with our previous demonstration for candidate model (3.1), the implementation of the new optimal design demands more calculation and effort to achieve a local maximum of the local D-criterion function. For each of the 30 $\mathbf{X}_{2\text{nd}}$ we obtain after the iterations, the criterion value is the same -43.0239. As such, the reference in Table 3.1 (of which $\phi_{\text{D}} = -48.2255$) is 27.24% efficient relatively. The coordinate exchange approach can be efficient in this situation too. The elapsed time of computation is 1079 seconds and the mean number of iterative updates is as low as 7.3333. At the different local convergences towards the maximum, the criterion value equals -43.0239 for 27 tries of the 30. This hints that we can do even fewer tries to find an efficient $\mathbf{X}_{\text{opt}}$.

Table 3.4: 30-Run Local D-Optimal Design for Model (3.2): Interim Solution after the Continuous Optimisation Part (2nd Stage) of the Hybrid Method Using

(a) Point Exchange

E	S	E	S	E	S
0.0255	3	0.0863	3	0.12	0.2625
0.0255	3	0.0863	3	0.12	0.2625
0.0255	3	0.0863	3	0.12	0.2625
0.0255	3	0.0863	3	0.12	3
0.0255	3	0.0863	3	0.12	3
0.0255	3	0.0966	0.3300	0.12	3
0.0255	3	0.12	0.2624	0.12	3
0.0255	3	0.12	0.2624	0.12	3
0.0863	3	0.12	0.2624	0.12	3
0.0863	3	0.12	0.2624	0.12	3

(b) Coordinate Exchange

E	S	E	S	E	S
0.0255	3	0.0863	3	0.12	0.2625
0.0255	3	0.0863	3	0.12	0.2625
0.0255	3	0.0863	3	0.12	0.2625
0.0255	3	0.0863	3	0.12	3
0.0255	3	0.0863	3	0.12	3
0.0255	3	0.0966	0.3300	0.12	3
0.0255	3	0.12	0.2624	0.12	3
0.0255	3	0.12	0.2624	0.12	3
0.0863	3	0.12	0.2624	0.12	3
0.0863	3	0.12	0.2624	0.12	3

As it takes more iterative updates on average in the continuous optimisation, the current exchanging rule $\max(d_i) > 1.00000001$ seems loose in this case. Besides, the fixed initial values in the coarse set $\{-0.67, 0, 0.67\} \times \{0.5, 1.5, 2.5\}$ are not in the far distance from the adopted coordinate levels of x in $\mathbf{X}_{\text{opt}}$, which improves the local optimal solutions of the continuous nonlinear optimisation. As we save four decimal places of the coordinate levels, Table 3.4a is identical to Table 3.4b. After the third stage of the hybrid method is completed, the solution is

$$\mathbf{X}^*_{\text{opt}} = \left\{ \begin{array}{ccccc} (0.025,\ 3) & (0.086,\ 3) & (0.097,\ 0.33) & (0.12,\ 0.26) & (0.12,\ 3) \\ 8 & 7 & 1 & 7 & 7 \end{array} \right\}, \tag{3.10}$$

the local D-criterion value of which is $\phi_{\text{D}} = -43.0242$. Overall, five unique runs and the respective numbers of replicates feature in the two rows in (3.10). Compared with (3.6) for the previous model, the unique run $(0.12, 0.26)$ also appears as the same substrate concentration is included. It shows some similarities due to the link between the two candidate models, in spite of the exponential transformation of the linear function in (3.2). In contrast, the difference is that it is harder to determine the coordinate levels of x in the local D-optimal design (3.10). As the number of experimental runs is $n = 30$ in this

example, an additional support point $(0.097, 0.33)$ appears in (3.10), which corresponds to just one replicate run. Although the new candidate model is also close-to-linear, the new continuous optimisation method is much more useful when we search for the final solution $\mathbf{X}^*_{\text{opt}}$ in the iterative update.

3.4.3 Model 3: the Transformed Candidate Model

In the third scenario, we use the transform-both-sides model (3.3), which assumes a simple multiplicative error structure in Michaelis-Menten kinetics. As a result, (3.3) is at least linear in terms of the scaled variable x. It also implies that the Fisher information matrix and the local D-criterion function are independent of the values of a_0, a_1, a_2. To define the discrete candidate levels of x, we use the conventional three levels in $\{-1, 0, 1\}$.

We fit the TBS model to the reference dataset, so the estimate is $\tilde{\boldsymbol{\theta}} = \{\tilde{k}, \tilde{a}_0, \tilde{a}_1, \tilde{a}_2\} = \{0.2838, -6.4406, 0.8420, -0.2561\}$, which does not deviate too far from that of the fitted untransformed model (3.2). The adjusted coefficient of determination is 0.989, so the observed response variation is well explained. After the model transformation, we can see some improvement in the error structure (i.e. the model satisfies the error assumptions better). Nevertheless, even if the error structure is not the most desirable one, (3.3) is still a useful empirical approximation because of its simple model function and strong linear behaviour. With the same hybrid exchange algorithm and the parameter prior as $\boldsymbol{\theta}^0 = \tilde{\boldsymbol{\theta}}$, the new $\mathbf{X}_{\text{opt}}$ are found in Table 3.5, as a result of the continuous optimisation.

Table 3.5: 30-Run Local D-Optimal Design for Model (3.3): Interim Solution after the Continuous Optimisation Part (2nd Stage) of the Hybrid Method Using

(a) Point Exchange

E	S	E	S	E	S
0.02	0.15	0.0695	3	0.12	0.15
0.02	0.15	0.0695	3	0.12	0.15
0.02	0.15	0.0695	3	0.12	0.15
0.02	0.15	0.0696	3	0.12	0.15
0.02	3	0.0696	3	0.12	3
0.02	3	0.0706	0.15	0.12	3
0.02	3	0.0706	0.15	0.12	3
0.02	3	0.0706	0.15	0.12	3
0.02	3	0.0706	0.15	0.12	3
0.0695	3	0.0706	0.15	0.12	3

(b) Coordinate Exchange

E	S	E	S	E	S
0.02	0.15	0.0700	0.15	0.12	0.15
0.02	0.15	0.0700	0.15	0.12	0.15
0.02	0.15	0.0700	0.15	0.12	0.15
0.02	0.15	0.0700	0.15	0.12	0.15
0.02	0.15	0.0700	0.15	0.12	0.15
0.02	3	0.0700	3	0.12	3
0.02	3	0.0700	3	0.12	3
0.02	3	0.0700	3	0.12	3
0.02	3	0.0700	3	0.12	3
0.02	3	0.0700	3	0.12	3

The point exchange approach can be used, when the computational time is 603 seconds and the mean number of iterative updates is 4.8. While there are 30 distinct $\mathbf{X}_{\text{1st}}$ from the initial discrete optimisation, the highest three local D-criterion values are $11.6658, 11.6658, 11.6657$ for the respective $\mathbf{X}_{\text{2nd}}$. In comparison, when the coordinate exchange approach is used, the elapsed time is 624 seconds and there are 4.4667 iterative updates on average. The highest three criterion values are $11.6960, 11.6658, 11.6658$, which are a bit better. This may be due to minor random variation.

We fail to find many D-efficient $\mathbf{X}_{\text{2nd}}$, so the exchanging rule $\max(d_i) > 1.00000001$ can be looser for continuous optimisation whereas $\max(d_i) > 1.1$ is also too strict a rule for the discrete optimisation. In addition, since there is one more unique run to be specified under the current local D-criterion, it will be reasonable to do more tries and use better initial starts (e.g. use the initial coordinate level $x = \pm 1$ rather than ± 0.67). The interim solution $\mathbf{X}_{\text{opt}}$ in Table 3.5a is not the best. Hence it is even more important to finish the hybrid method as we can make minor coordinate adjustments and then recalculate the optimal numbers of replicates. If we continue with Table 3.5a to search for the final solution $\mathbf{X}^*_{\text{opt}}$, for instance, the maximal local D-criterion value is 11.6956 for

$$\mathbf{X}_1^* = \left\{ \begin{array}{cccccc} (0.02,\ 0.15) & (0.02,\ 3) & (0.07,\ 3) & (0.071,\ 0.15) & (0.12,\ 0.15) & (0.12,\ 3) \\ 5 & 5 & 5 & 5 & 5 & 5 \end{array} \right\}. \tag{3.11}$$

This is similar to the interim solution in Table 3.5b. After the reallocation of replications, there are equal numbers of replicates of the six support points. We can also use Table 3.5b to restart the search. As the more efficient $\mathbf{X}_{\text{opt}}$ is chosen for the third stage of the hybrid method, the final solution is

$$\mathbf{X}_2^* = \left\{ \begin{array}{cccccc} (0.02,\ 0.15) & (0.02,\ 3) & (0.07,\ 0.15) & (0.07,\ 3) & (0.12,\ 0.15) & (0.12,\ 3) \\ 5 & 5 & 5 & 5 & 5 & 5 \end{array} \right\}. \tag{3.12}$$

Given the continuous variable space $\mathcal{X}$ of interest, we are convinced that (3.12) is indeed local D-optimal and the maximal local D-criterion value equals 11.6960. With respect to $\mathbf{X}_2^*$, the reference design in Table 3.1 (of which $\phi_{\text{D}} = 8.5739$) is 45.82% efficient. The scale of the experiment is $n = 30$ and there are five replicates of each of the $n^* = 6$ support points.

The candidate model (3.4) is close to linear and the exact local D-optimal experimental design is simplified to a large extent. The conventional candidate coordinate levels $\{-1, 0, 1\}$ are chosen for the scaled variable x in the discrete optimisation. Meanwhile, the upper and lower limits of the variable space $[0.15, 3]$ are the two candidate coordinates for the substrate concentration. As such, $\mathbf{X}_2^*$ is a 3×2 factorial, the support points of which are contained in the discrete candidate set $\boldsymbol{\Omega}$. Under a suitable exchanging rule, the traditional discrete optimisation shall be a sufficient method for us to find the local D-optimal solution (3.12). Details about different exchange algorithms are in Chapter 2.

In this section, we have considered three of the candidate hybrid nonlinear models, each of which can fit the reference data reasonably well. Hence, in optimal design of experiments in this case, as long as the statistical model is decided for the new experimental data, there is no need to make a compromise among optimal designs for different types of candidate models (i.e. (3.1)-(3.3) as shown above). More often than not, the best strategy is to focus on a specific hybrid model that suits our mechanism conjecture and interpretation. An exception is that, if the experimenters make several mechanistic

postulates in addition to simple Michaelis-Menten kinetics (i.e. when the best mechanism to assume is uncertain), optimal experimental design may take account of a model selection process, which is based on experimental data and scientific theories as well. Prior to experimentation, moreover, it may also take troubles to determine the suitable empirical terms in the hybrid nonlinear model, especially when the model involves more than two factors. Given such kinds of circumstances, a compound criterion or a composite criterion will be useful, if the experimenters wish to examine their mechanistic or empirical postulates and then decide the most appropriate model. An example of a simple compound criterion will be discussed in Section 3.7.

3.5 Complex Nonlinear Multifactor Models

3.5.1 Practical Complications in Nonlinear Experiments

The model can be complex because of the inherent kinetics mechanism. As we mentioned at the end of the last section, there could be other mechanistic postulates in addition to simple Michaelis-Menten kinetics. At times, the mechanistic model function is more complex, expressed as the ratio of two linear functions in terms of the substrate concentration. As explained in Wong (1975), that is not a rare case, due to diversities of different biochemical reactions. As such, the kinetic model (or the hybrid model that incorporates a mechanistic function of kinetics) shall include more unknown parameters. Thus we can see the complexities in the interpretation of the mechanism, which must be based on the rate laws (i.e. differential equations in relation to kinetics) and the experimental constraints. Besides, many chemical reactions in practice shall involve more than one substrate species and (by)products in the course of the reaction. Therefore, some advanced kinetics could be considered as an extension of the simple M-M model (Wong and Hanes, 1962). As we have also discussed, sometimes it is hard to select the most suitable and concise kinetic model to be fitted after the experiment.

However, in this chapter our emphasis is on the empirical component of the hybrid nonlinear model. We thus assume the mechanistic function component to be informative as we concentration on, for instance, the two-step M-M mechanism in the examples. In optimal design of experiments, it is not hard to replace the M-M equation with another univariate mechanistic function and we can implement the same hybrid method to find local D-optimal (or D-efficient at least) solutions. The complication is that, even if we assume Michaelis-Menten kinetics, an ambitious modeller can require multiple controlled variables to coexist in the same model. This issue is relevant to the complex nonlinear multifactor experiments with which we are concerned.

In common practice, to derive a mechanistic model like M-M kinetics, several irrelevant experimental conditions must be fixed at one or several constants. As a result, the

roles and influences of these factors will be less studied in the fitted mechanistic model. Outside M-M kinetics, nonetheless, there is a more comprehensive multifactor mechanism to explore, which is an extension from the simple M-M mechanism. Except for the enzyme concentration (or protein weight), various factors can be considered in the hybrid nonlinear model that approximates the true response surface. For instance, the famous experiment in Michaelis and Davidsohn (1911) explored a common kinetics when hydrogen ion concentration (i.e. the pH scale) was an independent variable. As a consequence, the initial rate function of the pH was close to quadratic in their defined variable space. Later in another experiment in Michaelis and Menten (1913), the pH scale had been fixed as a constant so that the substrate concentration was the only independent variable in the kinetics. In this situation, the classical form of the Michaelis-Menten equation had been established, where both the Michaelis constant and the theoretical maximum initial rate may depend on the pH. Though there is no mechanistic model to explain the response in terms of the pH, experimenters shall be able to consider an empirical approximation to their smooth functional relationship.

Temperature is another factor that influences the reaction kinetics. For some ideas about the link between the absolute temperature and the initial reaction rate, one can refer to the Arrhenius equation. However, this often results in an overcomplex and also unreliable model that is not very useful in experimental studies. The Arrhenius equation can be best seen as an empirical function, as it does not explain some kinds of kinetics as one wishes. A pertinent and universal kinetic model cannot be found to relate the temperature as a controlled factor.

3.5.2 Model and Optimal Design of the Experiment

We now consider a further example, where four controlled factors are involved in a mechanism on the basis of simple Michaelis-Menten kinetics. As usual it is not realistic or economic to derive an explicit mechanistic model for the response surface of the initial rate ν. Instead, it is far easier to fit a hybrid nonlinear model for a reliable empirical approximation. To substitute the maximum initial rate $\nu_{\max}$ in M-M kinetics, the empirical component is as complex as the exponential of a second-order linear function of three controlled variables excluding the (initial) substrate concentration. The hybrid model requires an effective utilisation of the n runs and the experimental data to estimate each treatment parameter we define. If there are some relevant reference data available for us to check the nonlinear regression results, it is sensible to drop some insignificant empirical terms out of the finalised model. When we are unconvinced with sufficient information, a safe choice is to assume the exponential of a full linear function as the empirical component of a hybrid model.

In this example, let $x_{\text{E}} \in [-1, 1]$ represent the enzyme concentration, which is a transformed variable in the empirical function. It can lead to a linear effect and a quadratic

effect as two terms in the empirical exponential function, as shown in (3.2). We can take two more variables into account in the second-order linear function in that model. Define the hydrogen ion concentration $x_{\mathrm{H}} \in [-1, 1]$ and experimental temperature $x_{\mathrm{T}} \in [-1, 1]$, both of which are transformed. Both factors are fixed in the Martins et al. experiment, where the variable space of the substrate concentration is $[0.15, 3]$. Hence, for optimal experimental design, we can define the four-dimensional variable space $\mathcal{X} = [-1, 1]^3 \times [0.15, 3]$. The experimenters should determine the limits (and sometimes also constraints) of the untransformed controlled variables, expressed in the units of real chemical measurements. When we combine the chosen levels of these factors, a realistic environment can be created as the experiment requires. Therefore, the reaction mechanism and kinetic profile are to be studied over a realistic variable space of new experimental interest.

Caution must be taken to evaluate the joint impacts of the controlled variables on the initial rate and its measurement. For instance, experimenters can often increase the temperature to speed up some chemical reactions within a certain time span. As the reaction continues, however, an excessive temperature can sometimes inactivate or even denature the enzyme steadily. As a consequence, such an experimental environment is impractical and thus the variable space $\mathcal{X}$ should not contain excessive temperature levels. If considerable enzyme activities cannot be retained later on in the course of the reaction, due to the common thermodynamic consideration, it does not make much sense to use a temperature that maximises the initial rate. To learn how to relate relevant theories and presumptions to the determination of a reasonable variable space, see Cornish-Bowden (2004, chap. 3).

There are four factors, the levels of which can be specified in the respective dimension of the variable space $\mathcal{X} = [-1, 1]^3 \times [0.15, 3]$. To learn the comprehensive mechanism in relation to Michaelis-Menten kinetics, we can build the hybrid nonlinear model

$$\nu = \frac{\exp(a_0 + a_1 x_{\mathrm{E}} + a_2 x_{\mathrm{H}} + a_3 x_{\mathrm{T}} + a_4 x_{\mathrm{E}}^2 + a_5 x_{\mathrm{H}}^2 + a_6 x_{\mathrm{T}}^2 + a_7 x_{\mathrm{E}} x_{\mathrm{H}} + a_8 x_{\mathrm{E}} x_{\mathrm{T}} + a_9 x_{\mathrm{H}} x_{\mathrm{T}}) S}{k + S} + \varepsilon, \tag{3.13}$$

where the parameter prior $\boldsymbol{\theta}^0 = \{0.3, -6.4,\ 0.8,\ 0.3,\ 0.8, -0.3, -0.3, -0.1,\ 0.1,\ 0.1,\ 0.1\}$ is in correspondence with $\boldsymbol{\theta} = \{k, a_0, \ldots, a_9\}$. As usual, ε is the residual, of which the mean shall be zero and the variance shall be constant across the $n = 30$ independent observations. The main experimental purpose we assume is to obtain precise parameter estimates for the fitted response surface, though the majority of the 11 parameters come from the empirical component (i.e. exponential function) of the model. As we fit this model, the advantage is the comprehension of the whole kinetics mechanism when no mechanistic model is available. Due to the lack of experimental data and the lack of fit of the fitted model, the empirical approximation of (3.13) to the response surface is not reliable all the time.

At first sight, due to the exponential transformation of the second-order linear function in the complex model (3.13), it can be difficult to find suitable candidate coordinates to

compose the 30-run local D-optimal design of the experiment. Therefore, the continuous (nonlinear) optimisation is a useful tool that we can exploit in the three-stage hybrid exchange algorithm. In accordance with previous demonstrations, we can take 30 random initial $\mathbf{X}$ for a swift discrete optimisation over a 3^4 candidate set. As to each of the new variables $x_{\text{H}}, x_{\text{T}}$, we can define the conventional three candidate levels in $\{-1, 0, 1\}$.

The total number of coordinates is 120 in the 30×4 matrix $\mathbf{X}$. For the discrete optimisation, the exchanging rule is $\max(d_i) > 1.05$ under the point exchange approach. The rule will influence our selection of candidate runs such that the aggregated number of distinct matrices $\mathbf{X}_{1\text{st}}$ is $\tau^* = 11$ (which is large enough). We can use them for 11 tries in the next subsequent continuous optimisation. As the hybrid method continues, the exchanging rule is set to $\max(d_i) > 1.00000001$, in which case we can expect a lot of iterative updates to complete. Likewise, it is sensible to use different vectors of initial values for continuous optimisation, which include the current point and those in the set $\{-0.67, 0, 0.67\}^3 \times \{0.5, 1.5, 2.5\} \in \mathcal{X}$. As such, when we do a v-dimensional continuous optimisation under the point exchange approach, there are $3^4 + 1 = 82$ initial points to examine and we expect a dramatic increase in the amount of computation. When we use the coordinate exchange approach, there are four initial coordinate values for each unidimensional optimisation. All the results of our hybrid method are summarised in Tables 3.6 and 3.7.

When the number of controlled variables is $v = 4$, it is hard to evaluate the 11×11 Fisher information matrix of a specific $\mathbf{X}$ and then maximise the local D-criterion function. In addition to the time consumption in the continuous optimisation, under the point exchange approach, the local optimal solutions of the v-dimensional optimisation will be less reliable. As we have mentioned above, we define 82 initial points in order to find a better local optimal solution. In that case, the overall computational time is 567 minutes, while the maximum and mean of the $\tau^* = 11$ local D-criterion values of different $\mathbf{X}_{2\text{nd}}$ is -113.5698 and -113.5806. The iterative computation can be quite intensive, so we allow at most 30 iterative updates in the continuous optimisation of each distinct $\mathbf{X}_{1\text{st}}$, which are obtained from the discrete optimisation in the hybrid method. On average, it takes 16.8182 iterative updates to achieve one local maximum of the local D-criterion function in the continuous optimisation. The result implies computational complexities when we use the point exchange approach.

When we use the coordinate exchange approach, the computation is quicker and the elapsed time is 64 minutes. The local D-criterion value of $\mathbf{X}_{\text{opt}}$ (see Table 3.6b) is -113.5601. On average, the criterion value of the 11 distinct $\mathbf{X}_{2\text{nd}}$ is calculated to be -113.5777. Compared with the previous results under the point exchange, the improvement is slight. The mean number of iterative updates is 28.8182, which also suggests difficulties to maximise the criterion function. Under the current exchanging rule $\max(d_i) > 1.00000001$, the maximum number of iterative updates (i.e. 30) has been hit 8 times out of the 11 tries in the continuous optimisation. Nevertheless, the improvement

Table 3.6: 30-Run Local D-Optimal Design for Model (3.13): Interim Solution after the Continuous Optimisation Part (2nd Stage) of the Hybrid Method Using

(a) Point Exchange

x_E	x_H	x_T	S
-0.7546	0.0915	1	3
-0.6794	1	1	3
-0.6792	1	1	3
-0.1285	-0.0337	-0.2440	3
-0.1282	-0.0346	-0.2432	3
-0.0374	1	-0.0686	3
-0.0230	-0.6674	1	3
-0.0226	-0.6676	1	3
0.2621	1	0.2824	3
0.3768	1	1	3
0.3768	1	1	3
0.4637	0.2582	1	3
1	-1	1	3
1	-1	1	3
1	-0.6107	-0.0323	3
1	-0.6107	-0.0322	3
1	0.1142	-1	3
1	0.1717	1	3
1	0.1718	1	3
1	0.2507	0.4691	3
1	0.4248	1	0.2853
1	1	-0.9293	3
1	1	-0.9291	3
1	1	0.3704	3
1	1	0.3704	3
1	1	1	0.2606
1	1	1	0.2606
1	1	1	3
1	1	1	3
1	1	1	3

(b) Coordinate Exchange

x_E	x_H	x_T	S
-0.7814	0.1461	1	3
-0.6705	1	1	3
-0.6704	1	1	3
-0.1404	-0.0385	-0.2640	3
-0.1394	-0.0415	-0.2623	3
-0.0536	-0.6401	1	3
-0.0536	-0.6401	1	3
-0.0290	1	-0.0543	3
0.2793	1	0.2982	3
0.3883	1	1	3
0.3884	1	1	3
0.4591	0.2221	1	3
1	-1	1	3
1	-1	1	3
1	-1	1	3
1	-0.5922	-0.0554	3
1	-0.5920	-0.0555	3
1	0.1465	-1	3
1	0.2077	1	3
1	0.2078	1	3
1	0.2092	0.4630	3
1	0.3855	1	0.2839
1	1	-0.9209	3
1	1	-0.9208	3
1	1	0.3830	3
1	1	0.3831	3
1	1	1	0.2667
1	1	1	0.2667
1	1	1	3
1	1	1	3

is expected to be small in the late iterative updates, so 20 or fewer iterative updates should be sufficient in this example.

On the basis of Table 3.6b, we continue the hybrid method to finish the last stage. The eventual solution is the $\mathbf{X}^*_{\text{opt}}$ shown in Table 3.7, the local D-criterion value of which equals -113.5603. As we can see, there are $n^* = 18$ support points in $\mathbf{X}^*_{\text{opt}}$, under the condition that we wish to estimate the 11 unknown parameters in (3.13). Most of the support points have one or two replicates, so it is hard to find their coordinate levels under the criterion.

Although the coordinate exchange approach seems quite useful, it is not flawless. While its unidimensional optimisation is quicker and more accurate at each step of the iterative update, it takes more updates to achieve a local maximum of the local D-criterion function. The impact of updating a single coordinate is quite trivial when there are 120 coordinates in total in each $\mathbf{X}$. Compared with the previous demonstrations in Section

Table 3.7: 30-Run Local D-Optimal Design for Model (3.13) with Hybrid Method

x_E	x_H	x_T	S	x_E	x_H	x_T	S
-0.78	0.14	1	3	1	-0.59	-0.06	3
-0.67	1	1	3	1	-0.59	-0.06	3
-0.67	1	1	3	1	0.15	-1	3
-0.14	-0.04	-0.26	3	1	0.21	0.46	3
-0.14	-0.04	-0.26	3	1	0.21	1	3
-0.05	-0.64	1	3	1	0.21	1	3
-0.05	-0.64	1	3	1	0.39	1	0.28
-0.03	1	-0.06	3	1	1	-0.92	3
0.28	1	0.3	3	1	1	-0.92	3
0.39	1	1	3	1	1	0.38	3
0.39	1	1	3	1	1	0.38	3
0.46	0.22	1	3	1	1	1	0.27
1	-1	1	3	1	1	1	0.27
1	-1	1	3	1	1	1	3
1	-1	1	3	1	1	1	3

3.4, the search of $\mathbf{X}_{\text{opt}}$ or $\mathbf{X}^*_{\text{opt}}$ is more dependent on the iterative updates. It can take quite a while to make the solution stable.

3.5.3 Modified Adjustment Algorithm

We consider the solution in Table 3.7 to be local D-efficient (if not D-optimal). With some more adjustments on the 120 coordinates (or the 18 support points), the solution can sometimes be improved even further. This shall be examined as we aim to find the $\mathbf{X}$ that maximises the criterion function. Let us consider a modified version of the adjustment algorithm in Donev and Atkinson (1988), which we adapt to work with the solution of the hybrid exchange algorithm, for instance. In short, it is more efficient and enables us to calibrate and adjust the level of each coordinate in Table 3.7. Although the required steps of our *modified adjustment algorithm* below are different to those of Donev and Atkinson (1988) and of the coordinate adjustments in the third stage of the hybrid method, the overall intention is similar:

The Modified Adjustment Algorithm After the Previous Exchange Algorithm

4.1 For $k = 1, 2, \ldots, v$, determine the closest distance $\mathfrak{d}_k$ for the kth factor (or controlled variable). While the column order of the v factors is fixed, sort the previous solution $\mathbf{X} = \mathbf{X}^*_{\text{opt}}$ in ascending row order. The $\mathbf{X}$ includes n^* unique runs, each of which correspond to a certain number of replicates. Calculate the initial local D-criterion value ϕ_{D} for the $\mathbf{X}$.

4.2 Reset $k = 1$ and flag $\Upsilon = 0$.

4.3 Let $i = 1$ and $i^* = 1$ be the starting values.

4.4 Pick out the i^*th unique run of $\mathbf{X}$, which corresponds to n_{i^*} replicates. The aim is to find the best candidate as a substitute for the kth coordinate of the selected unique run. It should then affect the relevant coordinates in the n_{i^*} replicates.

4.5 Create a subset of candidate coordinates $\mathbf{\Omega}$, for instance, on the basis of the closest distance, a defined space $[0.95X_{ik}, 1.05X_{ik}]$ (when the current coordinate value X_{ik} is nonnegative; the limits of this space can be changed to suit different examples), and the variable space $\mathcal{X}$. With the subset, do an optimisation for this unique run and denote X_{new} as the best candidate that maximises ϕ_{D}. If $X_{ik} \neq X_{\text{new}}$, execute the coordinate substitution for each replicate and set $\Upsilon = 1$. Let $i = i + n_{i^*}$.

4.6 Unless $i = n + 1$, let $i^* = i^* + 1$ and return to **step 4.4** .

4.7 Unless $k = v$, let $k = k + 1$ and return to **step 4.3**.

4.8 Unless $\Upsilon = 0$, return to **step 4.2**. Otherwise, save the current $\mathbf{X}$ as the final solution $\mathbf{X}^{**}_{\text{opt}}$, which is considered to be optimal under the local D-criterion.

The premise of the modified adjustment algorithm is that the replication of unique runs in $\mathbf{X}^{*}_{\text{opt}}$ is optimal. Therefore, the numbers of replicates are fixed and will not be reallocated in the iteration above. In our first demonstration, the solution of the hybrid method $\mathbf{X}^{*}_{\text{opt}}$ in Table 3.7 is input into the modified adjustment algorithm. As a result, there are no coordinate adjustments to make to maximise the local D-criterion function. That is, $\mathbf{X}^{**}_{\text{opt}} = \mathbf{X}^{*}_{\text{opt}}$. The result is impressive as the solution of the hybrid method is verified to be optimal.

Though there is no contribution to the final solution in this example, the modified adjustment algorithm can be a useful tool when $\mathbf{X}^{*}_{\text{opt}}$ is a bit less efficient than the optimal design. When we do fewer tries (or fewer iterative updates under a strict exchanging rule) in the continuous optimisation, the modified adjustment algorithm builds on and improves the previous solution and its criterion value. It will also be implemented in the next section, when our demonstrations demand even more intensive computation.

3.6 Local L-Optimal Design of Experiments

3.6.1 Numerical Results under the Local Weighted A-Criterion

Under the local D-criterion, the model parameters in (3.13) are treated with equal importance in the nonlinear least squares estimation. In this case, we allocate the experimental resources to minimise the Generalised Variance of the parameter estimators. However, the importance of each parameter can be adapted to fit different applications. Within

the complex hybrid nonlinear model (3.13), for example, we replace the maximum initial rate $\nu_{\max}$ of Michaelis-Menten kinetics with an empirical exponential function. As a result, the hybrid model function exhibits considerable uncertainties since it defines 11 unknown parameters in total.

Meanwhile, the Michaelis-Menten equation is nonlinear in terms of the parameter $k > 0$, which is the mechanistic component of the model. To some extent, we can sometimes presume that the precise estimation of k is more important than that of each parameter in the empirical exponential function. The rationale can be as follows: the estimator of k determines the nonlinearities of the fitted model (3.13) and thus influences the overall empirical approximation of the response surface; besides, the standard error of the estimator $\hat{k}$ delimits its confidence interval, which is in close relation to the interpretation of the kinetics mechanism; a small standard error can also reduce chances that the confidence interval contains the zero point. As to the 10 empirical terms in the numerator, while some of them (e.g. higher-order terms) are perhaps insignificant in the fitted model, the coefficients of some other terms can also be less important in the estimation. M-M kinetics should be followed such that a reliable estimation of k with a small standard error is desirable. In these circumstances, the local D-criterion does not suit all our requirements.

Under the current local D-criterion, we assume the solution in Table 3.7 to be optimal. The prior value of k is equal to 0.3, which is a small constant relative to the limits of the substrate concentration level in $[0.15, 3]$. In most of the $n = 30$ experimental runs, the substrate concentration levels are fixed at the maximum at 3 mM. The rest of substrate concentration levels seem to be dependent on the parameter prior $k^0 = 0.3$ more than the other priors. While the estimation of each empirical parameter is as important as that of k, the value of k^0 is not very important under the local D-criterion. As such, it is no surprise that the maximum level of the substrate concentration is so frequent in Table 3.7 (it appears 27 times in total).

Commonly, the D-criterion is fair with respect to nonlinear regression and least squares estimation of the parameters $\boldsymbol{\theta}$. The A-criterion, on the other hand, aims to minimise the trace of the inverse of the information matrix of a specific $\mathbf{X}$. This is equivalent to minimising the sum of variances of the NLS estimator $\hat{\boldsymbol{\theta}}$. The local A-criterion function can be written as $\phi_{\mathrm{A}} = \mathrm{Tr}((\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1})$ where Tr represents "trace". In the comparison of different $\mathbf{X}$, the relative efficiencies can be calculated as the direct ratios of the relevant criterion values ϕ_{A}. In spite of its extensive use in practice, the weakness is that the unweighted A-criterion will dismiss the diversities of the scales of the prior values of different parameters. Thus the A-criterion is quite sensitive to the scales or transformations of the model variables. This can be a serious issue for nonlinear models, the parameter prior values of which are specified.

To offset this weakness, the weighted A-criterion (WA-criterion) is proposed to do appropriate weighting on the diagonal elements of $(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}$. A weight matrix $\mathbf{W}$ should be incorporated into the criterion function. Further, $\mathbf{W}$ can be specified as a nondiagonal matrix too, when one attempts to weight the covariances of model parameter estimators. In this case, the local WA-criterion function $\phi_{\mathrm{WA}} = \mathrm{Tr}(\mathbf{W}(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1})$ is also defined as the local L-criterion function (L stands for "linear"). See Stallings and Morgan (2015) for some discussion on weighted criteria. In the example in this section, we shall choose different local WA-criterion functions to search for the optimal experimental designs for the complex model (3.13).

According to the recommendation in Gilmour and Trinca (2012a) on the specification of the WA-criterion function ϕ_{WA}, we can use the previous local D-optimal $\mathbf{X}^{*}_{\mathrm{opt}}$ (Table 3.7) to calculate the expected variances of $\hat{\boldsymbol{\theta}}$. The result is based on the Fisher information matrix of the model, such that we can obtain the variance vector $\{\mathbb{V}(\hat{k}), \mathbb{V}(\hat{a}_0), \ldots, \mathbb{V}(\hat{a}_9)\}$ $= \{0.1635, 1.4559, 2.4060, 1.2026, 2.2051, 0.8256, 0.1989, 0.8214, 0.5811, 1.7221, 0.7630\} \times 100000\sigma^2$. The error variance σ^2 is assumed to be constant, which can be estimated after we obtain the data. As such, the reciprocals of these variances are

$$\left\{\frac{1}{\mathbb{V}(\hat{k})}, \frac{1}{\mathbb{V}(\hat{a}_0)}, \ldots, \frac{1}{\mathbb{V}(\hat{a}_9)}\right\} = \frac{\{0.6116, 0.0687, 0.0416, 0.0832, 0.0453, 0.1211, 0.5027, 0.1217, 0.1721, 0.0581, 0.1311\}}{10000\sigma^2}.$$

On the basis of Table 3.7, the expected variance of $\hat{k}$ is $\mathbb{V}(\hat{k}) = 16350\sigma^2$. Here σ^2 must be a rather small value. Under a WA-criterion, the $p \times p$ diagonal matrix $\mathbf{W}$ can be specified as we refer to the reciprocals of expected variances. If we assume the 11 parameter estimators to be as important as each other, for instance, the above $\{1/\mathbb{V}(\hat{k}), 1/\mathbb{V}(\hat{a}_0), \ldots, 1/\mathbb{V}(\hat{a}_9)\}$ can act as diagonal elements of $\mathbf{W}$, where the unknown error variance σ^2 can be ignored.

In the first scenario, for instance, the local WA-criterion is to minimise the defined function $\phi_{\mathrm{WA},1}$, the matrix $\mathbf{W}$ of which is specified. In this scenario, we allocate three times the normal weight to the estimator $\hat{k}$, which is treated as most important in the fitted model. In Table 3.8, we show the unique weight matrix $\mathbf{W}$ and calculate its nonzero elements $W_{11}, W_{22}, \ldots, W_{11\,11}$.

While our recommendation is the quicker coordinate exchange approach, the evaluation of the inverse matrix $(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}$ is difficult in each step of the iterative update. To facilitate the computation involved to invert the 11×11 information matrix, Fedorov (1972) derived an updating function for the unweighted A-criterion. In spite of that, this criterion requires more computation than the updating function for the D-criterion (see Section 2.2). As we also take $\mathbf{W}$ into account in the hybrid exchange algorithm, it is viable to calculate and update the local WA-criterion function ϕ_{WA}.

Table 3.8: Four Unique Weight Matrices for the Respective WA-Criterion Functions

Element	Indicator	$\phi_{\mathrm{WA},1}$	$\phi_{\mathrm{WA},2}$	$\phi_{\mathrm{WA},3}$	$\phi_{\mathrm{WA},4}$
$10^{-4}\mathrm{W}_{11}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{k}})$	1.8347	0.6116	6.1156	1.8347
$10^{-4}\mathrm{W}_{22}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{a}}_0)$	0.0687	0.0687	0.0687	0.0687
$10^{-4}\mathrm{W}_{33}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{a}}_1)$	0.0416	0.0416	0.0416	0.0416
$10^{-4}\mathrm{W}_{44}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{a}}_2)$	0.0832	0.0832	0.0832	0.0832
$10^{-4}\mathrm{W}_{55}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{a}}_3)$	0.0453	0.0453	0.0453	0.0453
$10^{-4}\mathrm{W}_{66}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{a}}_4)$	0.1211	0.1211	0.1211	0.0606
$10^{-4}\mathrm{W}_{77}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{a}}_5)$	0.5027	0.5027	0.5027	0.2514
$10^{-4}\mathrm{W}_{88}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{a}}_6)$	0.1217	0.1217	0.1217	0.0609
$10^{-4}\mathrm{W}_{99}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{a}}_7)$	0.1721	0.1721	0.1721	0.0574
$10^{-4}\mathrm{W}_{10\,10}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{a}}_8)$	0.0581	0.0581	0.0581	0.0194
$10^{-4}\mathrm{W}_{11\,11}$	$10^{-4}/\mathbb{V}(\hat{\mathrm{a}}_9)$	0.1311	0.1311	0.1311	0.0437

A problem is that the inverse of the information matrix can be ill-conditioned sometimes and thus the calculated local WA-criterion value will take an extreme value. Moreover, it affects the application of the updating function since inaccurate results can be found in that case. To fix this problem, when the condition number of the inverse matrix is high for an initial random $\mathbf{X}$, we should not attempt to add a small constant to each diagonal element of the matrix. It does not work well in some situations. Instead, it is more reliable to do diagonal preconditioning of the information matrix. In this case, although the information matrix is close to singular, at least a numeric answer will be found as the criterion value. As such, the computation and the first iterative update will not be terminated due to the ill-conditioned matrix. After we complete the first iterative update, this problem will disappear as the information matrix will no longer be singular. Besides, an alternative measure is to resample the initial random matrix $\mathbf{X}$ until we can obtain a nonsingular information matrix. We can also increase the total number of candidate runs in the discrete optimisation. This reduces our chances to obtain a high condition number for each initial random $\mathbf{X}$, the number of support points of which tends to increase.

When the relative efficiencies are calculated, we evaluate the ratio between the new and old values of the criterion function ϕ_{WA}. In the iterative update, this requires us to set a rule to determine whether the specific coordinate exchange should be executed or not. In our demonstration, we use the exchanging rule $\min(d_i) < 0.9999$ or the equivalent

$$\max(\frac{1}{d_i}) > \frac{1}{0.9999}$$

for the initial discrete optimisation of the hybrid method, for instance. It seems much looser than the rule $\max(d_i) > 1.05$ under the local D-criterion. As to the continuous optimisation part of the hybrid method, the rule $\min(d_i) < 0.99999999$ will be suitable. The rule for the local WA-criterion does not have to depend on the number of model

parameters, since the criterion function is the weighted sum of the p diagonal elements from the inverse of the information matrix.

To construct a small discrete candidate set $\mathbf{\Omega}$, we choose four identical coordinate levels for each of the three variables involved in the exponential function of the model. As the number of coordinates is larger in $\mathbf{X}$, we have increased the number of candidate points for an efficient discrete optimisation. The four conventional levels used are $\{-1, -\sqrt{1/5}, \sqrt{1/5}, 1\}$, which are the coordinate levels for the continuous D-optimal experimental design for the 3rd-order linear model in one independent variable $x \in [-1, 1]$. More candidate coordinates can be specified, but this attempt does not always improve the solution of the hybrid method. Meanwhile, as we can infer from Table 3.7, no more than two candidate coordinates are required for the substrate concentration S. A simple choice is to take the upper and lower bounds of the variable space $S \in [0.15, 3]$, so the candidate subset of substrate concentration levels is $\{0.15, 3\}$. As we combine the four subsets of candidate coordinates for respective controlled variables, $\mathbf{\Omega}$ contains $4^3 \times 2 = 128$ candidate points in total. In addition, the maximum number of iterative updates is set to 30 in the continuous optimisation. The computation and update of matrices $\mathbf{X}$ should be halted after the 30th iterative update. This is a measure to speed up the minimisation of the criterion function.

The number of random tries is $\tau = 30$. After the initial discrete optimisation under, for example, the point exchange approach, we take the most efficient $\tau^* = 8$ intermediate solutions $\mathbf{X}_{1\text{st}}$. Until the finish of the continuous optimisation part of the hybrid method, 289 minutes have elapsed. Although we use the coordinate exchange approach, it takes longer under the local WA-criterion than under the D-criterion. $\phi_{\text{WA},1}$ is the criterion function in this case, as we use the weight matrix in Table 3.8. The criterion values of $\mathbf{X}_{2\text{nd}}$ are 11.3682, 11.4040, 11.4180, 11.4180, 11.4196, 11.4276, 11.4276, 11.4495. In seven of the cases in the continuous optimisation, the number of iterative updates approaches 30, so it is hard to find a stable minimum of the local WA-criterion function $\phi_{\text{WA},1}$ in this demonstration. After we complete the hybrid exchange algorithm, the solution $\mathbf{X}^*_{\text{opt}}$ is found in Table 3.9, of which $\phi_{\text{WA},1} = 11.3686$.

We implement the modified adjustment algorithm and the local WA-optimal solution $\mathbf{X}^{**}_{\text{opt}}$ is shown in Table 3.10, the new criterion value of which equals 11.3681. An improved local minimum of the criterion function $\phi_{\text{WA},1}$ is identified. In comparison, some appreciable updates are made on the 9th and 10th experimental runs in Table 3.9. As a result, there are $n^* = 21$ support points under this local WA-criterion. It is difficult to determine the optimal coordinate levels of these points. The criterion function to minimise seems quite insensitive or robust to minor changes of some coordinates of the three transformed variables. In the meantime, it is far easier to determine the substrate concentration levels of which 25 are set to the maximum. Due to the complexities of the model assumed, sometimes there is no assurance that the final solution $\mathbf{X}^{**}_{\text{opt}}$ is local WA-optimal, even if the modified adjustment algorithm works well.

Table 3.9: When $\phi_{\text{WA},1}$ Is the Criterion Function: 30-Run Local WA-Optimal Design for Model (3.13) by Hybrid Coordinate Exchange Algorithm

x_E	x_H	x_T	S	x_E	x_H	x_T	S
-0.85	1	1	3	1	-1	1	3
-0.85	1	1	3	1	-0.81	-0.19	3
-0.45	-0.63	1	3	1	-0.81	-0.19	3
-0.3	0.28	-0.49	3	1	-0.26	-1	3
-0.3	0.28	-0.49	3	1	0.3	1	3
-0.3	0.28	-0.49	3	1	0.32	0.55	3
-0.23	1	-0.17	3	1	0.44	1	0.23
-0.22	-0.82	1	3	1	0.44	1	0.23
-0.07	-0.52	-0.03	3	1	1	-1	3
0.06	-0.67	0.14	3	1	1	-1	3
0.45	1	1	3	1	1	0.4	3
0.47	1	0.46	3	1	1	1	0.21
0.55	0.31	1	3	1	1	1	0.21
0.62	1	1	0.27	1	1	1	3
1	-1	1	3	1	1	1	3

Table 3.10: When $\phi_{\text{WA},1}$ Is the Criterion Function: 30-Run Local WA-Optimal Design for Model (3.13) after the Modified Adjustment Algorithm

x_E	x_H	x_T	S	x_E	x_H	x_T	S
-0.85	1	1	3	1	-1	1	3
-0.85	1	1	3	1	-0.81	-0.19	3
-0.45	-0.64	1	3	1	-0.81	-0.19	3
-0.3	0.27	-0.48	3	1	-0.27	-1	3
-0.3	0.27	-0.48	3	1	0.3	1	3
-0.3	0.27	-0.48	3	1	0.32	0.54	3
-0.23	-0.81	1	3	1	0.43	1	0.23
-0.22	1	-0.18	3	1	0.43	1	0.23
-0.11	-0.47	-0.08	3	1	1	-1	3
0.1	-0.71	0.19	3	1	1	-1	3
0.45	1	1	3	1	1	0.4	3
0.47	1	0.46	3	1	1	1	0.21
0.55	0.31	1	3	1	1	1	0.21
0.63	1	1	0.27	1	1	1	3
1	-1	1	3	1	1	1	3

In the local WA-criterion function, the numeric matrix $\mathbf{W}$ can be updated as appropriate. This makes the allocation of experimental units more flexible as we can tailor the specific requirements in the NLS estimation of model parameters. In Table 3.8, we consider a similar criterion function $\phi_{\text{WA},2}$ which has assigned a normal weight to the estimator $\hat{k}$. In model (3.13) to be fitted, the 11 parameter estimators in $\hat{\boldsymbol{\theta}}$ are as important as each other. Under this criterion, the solution of the modified adjustment algorithm $\mathbf{X}^{**}_{\text{opt}}$ is illustrated in Table 3.11, which is close to local WA-optimal. It looks different from the solution above in Table 3.10. In contrast, in the scenario when we opt for 10 times the normal weight on $\hat{k}$, the criterion function is $\phi_{\text{WA},3}$ also in Table 3.8. It leads to the final solution in Table 3.12. Again, it looks different from the results above, in terms of the adopted support points and the coordinate levels.

Table 3.11: When $\phi_{\text{WA},2}$ Is the Criterion Function: 30-Run Local WA-Optimal Design for Model (3.13) after the Modified Adjustment Algorithm

x_E	x_H	x_T	S	x_E	x_H	x_T	S
-0.85	-0.15	1	3	0.74	1	1	0.25
-0.85	1	1	3	1	-1	1	3
-0.85	1	1	3	1	-1	1	3
-0.31	0.09	-0.44	3	1	-0.84	-0.17	3
-0.31	0.09	-0.44	3	1	-0.84	-0.17	3
-0.23	1	-0.2	3	1	-0.21	-1	3
-0.23	1	-0.2	3	1	0.26	1	3
-0.22	0.11	-0.62	3	1	0.31	0.54	3
-0.17	-0.9	1	3	1	0.4	1	0.26
-0.17	-0.9	1	3	1	1	-1	3
0.03	-0.6	0.04	3	1	1	-1	3
0.03	-0.6	0.04	3	1	1	0.4	3
0.41	1	1	3	1	1	1	0.23
0.53	1	0.51	3	1	1	1	3
0.53	0.3	1	3	1	1	1	3

Table 3.12: When $\phi_{\text{WA},3}$ Is the Criterion Function: 30-Run Local WA-Optimal Design for Model (3.13) after the Modified Adjustment Algorithm

x_E	x_H	x_T	S	x_E	x_H	x_T	S
-0.85	1	1	3	1	-0.81	-0.13	3
-0.85	1	1	3	1	-0.43	-0.88	3
-0.51	-0.53	1	3	1	0.37	0.52	3
-0.26	-0.11	-0.35	3	1	0.39	1	3
-0.26	-0.11	-0.35	3	1	0.5	1	0.2
-0.26	-0.11	-0.35	3	1	0.5	1	0.2
-0.16	-0.88	1	3	1	0.5	1	0.2
-0.12	1	-0.15	3	1	1	-1	3
-0.12	1	-0.15	3	1	1	-1	3
0.36	-0.8	0.09	3	1	1	0.44	3
0.47	1	1	3	1	1	1	0.19
0.55	0.38	1	3	1	1	1	0.19
0.63	1	1	0.24	1	1	1	0.19
0.74	1	0.72	0.29	1	1	1	3
1	-1	1	3	1	1	1	3

Finally, we construct the local WA-criterion function $\phi_{\text{WA},4}$. This allocates three times the normal weight on the estimator $\hat{k}$, half the normal weight on coefficient estimators of the empirical quadratic terms (i.e. $\hat{a}_4, \hat{a}_5, \hat{a}_6$), and one third the normal weight on coefficient estimators of the empirical interaction terms (i.e. $\hat{a}_7, \hat{a}_8, \hat{a}_9$). This criterion function is useful when the current model we assume for the experiment is tentative. As we alter the weights in this way, the new solution in Table 3.13 can be compared with the above three. Now that we have studied five different criterion functions (four local WA and also local D) for the same complex model (3.13), it is reasonable to examine the efficiencies or robustness of the solutions in Tables 3.7, 3.10-3.13 across the different scenarios. Table 3.14 is compiled to summarise these results and allow mutual comparisons.

Table 3.13: When $\phi_{\text{WA},4}$ Is the Criterion Function: 30-Run Local WA-Optimal Design for Model (3.13) after the Modified Adjustment Algorithm

x_E	x_H	x_T	S	x_E	x_H	x_T	S
-0.97	0.45	1	3	1	-1	1	3
-0.83	1	1	3	1	-0.84	-0.1	3
-0.55	-0.53	1	3	1	-0.51	-0.78	3
-0.33	0.06	-0.39	3	1	0.35	1	3
-0.33	0.06	-0.39	3	1	0.36	0.5	3
-0.33	0.06	-0.39	3	1	0.58	1	0.2
-0.12	1	0	3	1	0.58	1	0.2
-0.12	1	0	3	1	0.58	1	0.2
0	-1	1	3	1	1	-1	3
0.13	-0.65	0.18	3	1	1	-1	3
0.13	-0.65	0.18	3	1	1	0.41	3
0.17	-0.1	-1	3	1	1	1	0.2
0.42	1	1	3	1	1	1	0.2
0.53	0.34	1	3	1	1	1	3
0.61	1	1	0.26	1	1	1	3

Table 3.14: Relative Efficiencies (%) of The Experimental Designs in Tables 3.7, 3.10-3.13 with Respect to the Optimal Design (Which Corresponds to the 100% Cases in the Diagonal) in the Assumed Scenario. Each Column Shows the Calculated Efficiencies under a Specific Scenario.

	ϕ_{D}	$\phi_{\text{WA},1}$	$\phi_{\text{WA},2}$	$\phi_{\text{WA},3}$	$\phi_{\text{WA},4}$
Table 3.7	**100**	87.45	88.58	76.26	80.11
Table 3.10	91.19	**100**	97.02	95.29	98.82
Table 3.11	91.28	96.42	**100**	80.48	91.18
Table 3.12	84.02	94.39	87.57	**100**	97.46
Table 3.13	84.94	96.74	91.69	97.26	**100**

Here are some of our comments and findings after the tedious computation in optimal design of experiments. In the scenario under the WA-criterion function $\phi_{\text{WA},2}$ which treats $\hat{k}$ to be as important as the remaining parameters, the local D-optimal solution in Table 3.7 is found to be 88.58% efficient, relative to the baseline in Table 3.11. This indicates the discrepancies between the two well-known optimality criteria in the minimisation of the variances and covariances of the NLS estimator $\hat{\boldsymbol{\theta}}$. If we do not intend to control and reduce the covariance estimates, the WA-criterion will meet the experimental purpose best since it focuses on the individual variances rather than the Generalised Variance of $\hat{\boldsymbol{\theta}}$.

In the last scenario we assume, the criterion value $\phi_{\text{WA},4}$ of Table 3.10 is equal to 7.7011, which entails an important deduction. This design is 98.82% efficient with respect to the WA-optimal design in Table 3.13, which is close to reaching 100%. While these two are rather different from each other in the coordinate levels, their relative efficiencies are similar in the comparison. Hence, it seems to be hard to determine the coordinate levels of an $\mathbf{X}$ that can minimise the local WA-criterion function. The result also implies that

the parameter estimation of the six second-order terms has just minor impact on each WA-criterion we assume.

In the local WA-criterion function, we can reduce the relative importance on parameter estimators of the second-order terms in the model. This is a bit similar to increasing the weight on the estimator $\hat{k}$. Therefore, on the one hand, the two local WA-optimal solutions in Tables 3.12-3.13 are alike in terms of the substrate concentration levels and the calculated relative efficiencies. On the other hand, both designs in Table 3.7 and Table 3.11 appear to be less efficient in the two scenarios when we choose either the criterion function $\phi_{\text{WA},3}$ or $\phi_{\text{WA},4}$.

3.6.2 General Applications of the Local L-Criterion

When we choose the local D-criterion as in Section 3.5, sometimes we can obtain more distinct and efficient intermediate solutions $\mathbf{X}_{1\text{st}}$ from the initial discrete optimisation. These matrices are distinct but can be quite similar to each other. If we opt for a WA-criterion or L-criterion, in our demonstrations, there are fewer efficient and more discrepant $\mathbf{X}_{1\text{st}}$. These intermediate solutions are found at the local minima of the criterion function. To some extent, this is because of the much looser exchanging rule $\min(d_i) < 0.9999$ we adopt for the discrete optimisation. When this rule is modified, the solutions of the discrete optimisation would be affected too and that can lead to different results of the hybrid method.

We should choose appropriate rules and a suitable discrete candidate set. When an improper rule is set for discrete optimisations over a bad candidate set, it is then hard to minimise the criterion function in the subsequent continuous optimisation part. The situation deteriorates when the nonlinear multifactor model is as complex as in this example. Besides, in real practice, we have to do sufficient random tries to increase the chances to find an efficient (if not optimal) design in the end. The increasing computational time is one of our concerns to address, but most of the time, it is manageable.

When the D-criterion is chosen in favour of each WA-criterion we define in Table 3.8, the final solution (Table 3.7) tends to be more stable. This means that there are more coordinate levels fixed at the respective upper limits in the variable space. In contrast, as we see in Table 3.10, for instance, the finalised coordinate levels are not quite similar to the coordinates defined in the $4^3 \times 2$ discrete candidate set. As such, it can take more iterative updates to maximise the local WA-criterion function and search for $\mathbf{X}_{2\text{nd}}$ in the continuous optimisation. The hybrid exchange algorithm relies on the discrete candidate set, so it is less efficient in these local WA-criterion cases.

The hybrid model is complex, so v-dimensional continuous optimisations are difficult under the point exchange approach. In this situation, there are plenty of ways to adapt the hybrid method. In the above demonstrations, we prefer the hybrid coordinate exchange

algorithm in order to reduce the computational cost in each iterative update. If we wish to increase the dimension of optimisation, perhaps it is useful to compromise between the two approaches when the number of controlled variables is large in the experiment. We can think about the 2-dimensional or 3-dimensional optimisation, so as to update a fixed small number of coordinates at a time in the iterative update. As such, we can somewhat compensate the drawbacks of the coordinate exchange approach.

Even if we stick to the coordinate exchange approach for continuous optimisation, at the same time, it will be beneficial to update the other coordinate values in the same row of $\mathbf{X}$. This can be done with a swift discrete optimisation over a small subset of candidate coordinates. It is a combination of the univariate continuous optimisation and the multivariate discrete optimisation, in order to reduce the mutual dependence between the coordinate updates.

If there are a lot of experimental runs (e.g. $n > 30$), a refined measure is to define fewer runs n^- first. With the hybrid method up to the end of the continuous optimisation, we can find several $\mathbf{X}_{2\text{nd}}$, each of which can be treated as a discrete candidate set of n^- runs. Now the number of runs can be reset to the actual size n in the computation. We can either do the discrete optimisation over each candidate set (if n^- is near n), or we can use the candidate sets to rerun the hybrid exchange algorithm from its discrete optimisation part (in the first stage). This ensures that we use a suitable candidate set to make the solution of the hybrid method reliable. Even if the total number of experimental runs is small, this measure is useful as well whenever we cannot determine the suitable candidate runs (or coordinates).

For continuous optimisation in the hybrid method, we use multiple vectors of initial values in each step of the iterative update. These vectors include the default current experimental run and the fixed points in the set $\{-0.67, 0, 0.67\}^3 \times \{0.5, 1.5, 2.5\} \in \mathcal{X}$. To redefine this 3^4 set, we can also select the points on the boundaries of the variable space $\mathcal{X}$ to replace some of those inside $\mathcal{X}$. An alternative choice is to fill the bounded variable space with the points from, for instance, a Latin hypercube or a sphere packing. These points can also be suitable initial vectors for the continuous optimisation. Yet another choice is to use the points we find from the previous discrete optimisation over a small candidate set. This measure can sometimes reduce the number of initial vectors to be included.

3.7 A Simple Compound Criterion

We can fit more than one candidate model to the same experimental data. Then the best model can be found as the approximation of the response surface. To improve the model selection and the parameter estimation, sometimes it is useful to construct a simple *compound criterion*. In these circumstances, we combine different criterion functions

and also take different candidate models into account in the overall compound criterion function. In the example, our purpose is to find the 30-run local D-optimal $\mathbf{X}$ that is robust to a minor modification of the current statistical model. From the perspective of statistical inference, the comparison between the candidate models (e.g. with and without the 3rd-order empirical term) should be well clarified before we operate the experiment and select the most suitable model. On the basis of the hybrid nonlinear model (3.2), for instance, we can add the cubic term x^3. In this case, the model becomes

$$\nu_i = \frac{\exp(a_0 + a_1 x_i + a_2 {x_i}^2 + a_3 {x_i}^3) S_i}{k + S_i} + \varepsilon_i, \tag{3.14}$$

where a_3 is the unknown coefficient of the cubic term and the exponential transformation of the linear function is substituted into the Michaelis-Menten equation. We expect the value of a_3 to be near zero and the P-value of its estimate to be near the significance level 0.05. As such, it is possible that we should drop the cubic term from the tentative model above, which could have minor influence on the response variation. This means that there is a model selection question on whether we shall delete the cubic term or not, which could depend on the P-value of the a_3 estimate. After the new experiment, it will be sensible to test if the more complex candidate model (3.14) is better than (3.2). Although the old data (if it is available) can also be used to detect the best model, sometimes the conclusion is not reliable or informative. As a result, the $\mathbf{X}$ that is local optimal for (3.2) might be inefficient or irregular for (3.14).

As we learn from Gilmour and Trinca (2012b), the (DP)s-criterion can be applied to minimise the $100(1-\alpha)\%$ confidence region of the specified estimators, which is the same as maximising the power of the F-test. The difference is that we consider the local D-criterion rather than the local DP-criterion for the precise parameter estimation. Besides, there is just one additional parameter in (3.14), of which we wish to minimise the variance. We therefore consider a local Ds-criterion function ϕ_{Ds}, which is a component of the compound criterion function. The related aim is to maximise the statistical power of a two-tailed t-test of the null hypothesis $a_3 = 0$, since the maximal power of the test can provide support to our probable acceptance or rejection of the hypothesis.

Meanwhile, no matter what the result is in the t-test, we should consider the parameter estimation in both candidate models. In a composite D-criterion function, suitable weights can be introduced to link the two local D-criterion functions for the respective models. In the example, we fit (3.14) to the kinetic data in Martins et al. (1999). The parameter prior $\boldsymbol{\theta}^0 = \{k^0, a_0^0, a_1^0, a_2^0, a_3^0\}$ can therefore be found for evaluation of the determinant of the information matrix $|\mathbf{M}_1|$. At the same time, since a_3 is assumed to be near zero, we use the identical $\boldsymbol{\theta}^0$ for the alternative model (3.2). It simplifies our later calculation of the compound criterion function ϕ_{CC}. In contrast to $|\mathbf{M}_1|$, for the simpler candidate model, denote $|\mathbf{M}_2|$ as the determinant of the relevant 4×4 information matrix.

There are some benefits from our considerations above. If the cubic term is found to be significant in (3.14), then we can reject the candidate model (3.2). As such, the value of $|\mathbf{M}_2|$ makes no sense for the specific $\mathbf{X}$. While even if the cubic term should be removed from (3.14), the incurred bias is small in the evaluation of $|\mathbf{M}_2|$ for (3.2). In both cases, the parameter prior is reliable. As the constants c_1, c_2, c_3 can denote the weights involved in the model selection in our example, the compound criterion function is defined as

$$\begin{aligned}\phi_{\mathrm{CC}} &= c_1\phi_{\mathrm{D1}} + c_2\phi_{\mathrm{D_s}} + c_3\phi_{\mathrm{D2}} = c_1\log|\mathbf{M}_1| + c_2\log\frac{|\mathbf{M}_1|}{|\mathbf{M}_2|} + c_3\log|\mathbf{M}_2| \\ &= (c_1 + c_2)\log|\mathbf{M}_1| + (c_3 - c_2)\log|\mathbf{M}_2|.\end{aligned}$$

When the weights are $c_1 = c_3 = 0$, we concentrate on the statistical t-test of the a_3 estimator. When $c_2 = c_3 = 0$ or $c_1 = c_2 = 0$, we then concentrate on one of the two candidate models and the criterion function will be ϕ_{D1} or ϕ_{D2}. Commonly, unless there is some further information for the mechanism studies, it is safe to assume equal weights; thus we choose $c_1 = c_2 = c_3$ in the example. However, the term $|\mathbf{M}_2|$ will be knocked out of the compound criterion function so that it is the same as to set $c_2 = c_3 = 0$. The above criterion is then simplified to the local D-criterion for the more complex candidate model (3.14). We can understand the simplification as follows. When we adopt the criterion function ϕ_{CC} with three equal weights, the aim is to minimise the Generalised Variance (i.e. a scalar function of variances and covariances) of the estimators of the five parameters in (3.14), where the precise estimation of a_3 is as important as the rest.

Thanks to the close-to-linear behaviours of both candidate models, we can find a simplified compound criterion function. The result of the univariate t-test can determine the candidate model to fit to the experimental data. Let us look at the difference between the constants c_2 and c_3. In general, $c_2 \ll c_3$ does not make sense. If that is the case, the criterion considers a_3 to be even less important than the individual unknown parameters in (3.2). This is often an unwanted consideration. Unless the experimental resources are rich (i.e. there are lots of runs), the criterion will be unhelpful for the statistical inference with the t-test. In contrast, except for scenarios when $c_2 \to c_3$, a more reasonable consideration is to choose a constant $c_2 > c_3$ and thus to weight the Ds-criterion function component more. This allows us to better discriminate between the two candidate models.

However, if the weight is $c_2 \gg c_3$ instead of $c_2 = c_3$, it seems that we are looking for clearer evidence to accept one of the models and that is the reason to improve the statistical power of the t-test (when the confidence has been fixed at a certain level). As a consequence, we can better discriminate between models, but less improvement can be made in the parameter estimation of the fitted model. It is important to achieve a plausible balance among multiple objectives (i.e. experimental purposes), the differences between the three weights in the criterion function should not be over large.

In our demonstration, we implement the hybrid exchange algorithm in the simplest case when $c_1 = c_2 = c_3$ and thus $\phi_{CC} \propto \phi_{D1}$. In other words, the compound criterion aims to maximise $|\mathbf{M}_1|$, the determinant of the information matrix for the candidate model (3.14). As we fit (3.14) to the reference data, the cubic term seems to lack significance. The NLS estimate is $\tilde{\boldsymbol{\theta}} = \{0.3148, -6.4151, 0.8959, -0.2696, -0.0928\}$ which is quite similar to the estimate $\{0.3122, -6.4086, 0.8383, -0.2861\}$ of the fitted model (3.2). As expected, the prior value $a_3^0 = -0.0928$ is relatively close to zero. After the new experiment, we can do a hypothesis test for the cubic term. As such, one experimental purpose is to discriminate between the two candidate models. Because of the cubic term in the model, we have to define a different discrete candidate set. With 30 random tries, we use the exchanging rule $\max(d_i) > 1.1$ for optimisation over the discrete candidate set $\boldsymbol{\Omega} = \{-1, -\sqrt{0.2}, \sqrt{0.2}, 1\} \times \{0.15, 1, 2, 3\}$. Under the criterion function ϕ_{D1}, 28 distinct $\mathbf{X}_{1st}$ are found in the first stage of the hybrid method. As we finish the following continuous optimisation, Table 3.15b shows the local D-optimal solution under the coordinate exchange approach, the criterion value of which is $\phi_{D1} = -56.3972$. In contrast, under the point exchange approach, the criterion value of the solution in Table 3.15a equals -56.3987. With either solution selected for the rest of the computation and then the modified adjustment algorithm, the final solution is

$$\mathbf{X}^{**}_{opt} = \left\{ \begin{array}{cccccc} (0.02,\ 3) & (0.062,\ 3) & (0.1,\ 3) & (0.104,\ 0.3) & (0.12,\ 0.27) & (0.12,\ 3) \\ 6 & 6 & 6 & 2 & 4 & 6 \end{array} \right\}, \quad (3.16)$$

the achieved maximal criterion value of which is $\phi_{D1} = -56.3834$. It is local D-optimal for model (3.14) and also optimal under the simple compound criterion.

Table 3.15: 30-Run Local D-Optimal Design for Model (3.14): Interim Solution after the Continuous Optimisation Part (2nd Stage) of the Hybrid Method Using

(a) Point Exchange

E	S	E	S	E	S
0.02	3	0.0619	3	0.12	0.2812
0.02	3	0.0998	3	0.12	0.2812
0.02	3	0.0998	3	0.12	0.2812
0.02	3	0.0998	3	0.12	0.2812
0.02	3	0.0998	3	0.12	3
0.0619	3	0.0998	3	0.12	3
0.0619	3	0.0998	3	0.12	3
0.0619	3	0.1041	0.2989	0.12	3
0.0619	3	0.1041	0.2990	0.12	3
0.0619	3	0.1041	0.2990	0.12	3

(b) Coordinate Exchange

E	S	E	S	E	S
0.02	3	0.0619	3	0.12	0.2744
0.02	3	0.0998	3	0.12	0.2744
0.02	3	0.0998	3	0.12	0.2744
0.02	3	0.0998	3	0.12	0.2744
0.02	3	0.0998	3	0.12	3
0.0619	3	0.0998	3	0.12	3
0.0619	3	0.0998	3	0.12	3
0.0619	3	0.1027	0.3149	0.12	3
0.0619	3	0.1027	0.3149	0.12	3
0.0619	3	0.12	0.2744	0.12	3

If we use the final solution in (3.16) for the new experiment and for the parameter estimation in the other candidate model (3.2), it is $\exp(-43.3911/4)/\exp(-43.0242/4) = 91.24\%$ efficient with respect to the local D-optimal design in (3.10). While the cubic term is included in model (3.14), Michaelis-Menten kinetics (i.e. the initial rate function in terms of the substrate concentration) remains the same. There are $n^* = 6$ unique runs in (3.16) and most of the substrate concentration levels are also fixed at the defined

maximum in $\mathcal{X}$. In comparison to (3.10), there is one more degree of freedom for the lack of fit of the total error. Thus experimenters are able to fit a bit more complex candidate models to the dataset.

While the first three unique runs in (3.16) look different from the first two in (3.10), the importance of the estimator $\hat{k}$ is reduced under the new local D-criterion. As a result, there are just four replicates at the support point $(0.12, 0.27)$, which is in close relation to the variance of $\hat{k}$. This implies that when there is an overcomplicated empirical function component in the model, the experimental resources will be undercut for the estimation of each important parameter, of which the precision will be lower. Hence a concise hybrid nonlinear model is desirable.

Given the small parameter prior $a_3^0 = -0.0928$, no doubt (3.16) is quite robust to the two candidate models. This is also the reason that we can consider the compound criterion for both the parameter estimation and the t-test of the significance of the cubic term. As the P-value of the test is expected to be near the threshold 0.05, it is even more important to reduce the standard error of $\hat{a}_3$ in order to select the correct statistical model. However, if we are unsure about the model function, it shall be useful to assume the more complex one and find the model-oriented optimal design for that. After the data collection in the new experiment, the insignificant terms can be deleted to establish the best model.

Except for the D-criterion and the compound criterion, the local L-criterion is an alternative to make the adopted optimal designs robust to multiple candidate models. It can let us assign more weight to the important estimators and less weight to the less important ones (e.g. the higher-order terms in the exponential function of (3.14)). If a suitable $\mathbf{W}$ can be found, compared with different compound criterion functions we can assume (if $c_2 \neq c_3$ in this example), sometimes it is simpler to compute the local L-criterion function. However, all of these criterion functions will require the most complex candidate model (e.g. (3.14)) to contain all the others (e.g. (3.2)). This restricts the scope of model selection.

As we show in this example, the Ds criterion works well when the candidate models are nested. To discriminate between nonnested rival models (when the true model is known), a more formal approach is to use the T-criterion (Atkinson and Fedorov, 1975) so that there is no restriction as above. If the true model is unknown, the best rival model must be decided after the new experiment such that a compound T-criterion is the sensible choice (Atkinson, 2008). To demonstrate that, Atkinson and Bogacka (2002) applied the compound T-criterion in optimal design of experiments for the selection between two kinetic models for inhibition (i.e. competitive and noncompetitive ones). With some Monte Carlo computation to evaluate the expectation of the compound T-criterion function, the sequential experiments in Myung and Pitt (2009) aimed to discriminate between several empirical models. Similar ideas can benefit widespread

biological studies and the reader can see Kreutz and Timmer (2009) for an overview of them.

In addition to discrimination of rival models, statistical inference is another area of interest in the analysis of experimental data. In the example in this section, we show a simple local Ds-criterion function for the t-test of one parameter estimate. Of course, various inferences can also be done with the data available. See Gilmour and Trinca (2012b) for some literature review and also the latest recommendations, where the statistical inference is on the basis of the pure error estimator of the response variance.

3.8 A Nonlinear Model with a Categorical Variable

A series of experiments in Yamamoto (1958) unveiled the photodynamic action of dyes in the inactivation of bacteriophages. As the biochemical reaction described proceeded, the concentration of the specified substrate (or the dye) tended to decrease as the phage strains lost some of their activation effects. Quantitative studies of such kinds of mechanisms are important. Experimental results can assist doctors in the relevant disease therapies and the required medical treatments in this area.

Given several regularity assumptions, Yamamoto (1958) was able to use simple Michaelis-Menten kinetics for interpretation of the inactivation mechanism. In order to use the simple linear regression, all of their models were linearised to follow the Lineweaver-Burk equation. In contrast, to find more reliable estimation of unknown parameters, it is much better to fit the nonlinear M-M model instead with nonlinear least squares. In our example, the M-M equation is considered as the mechanistic component of the model to be assumed.

The overall experimentation had several important purposes (e.g. to learn the inhibition effects when two different dyes were considered in the same reaction). Our focus is on the first part of their work, which is on M-M kinetics and the photodynamic inactivation of the coliphage T5 of the escherichia coli strain B. In this case, a controlled factor is the pH level that we consider to be a continuous variable; the other factor is the dye species that shall be treated as a categorical variable. It is neither essential nor economical to collect all the data, so we pick out two effective dyes for the new experiment: methylene blue and toluidine blue.

According to the reference experiment, there are in total 24 observations relevant to the two dyes (five for the less important toluidine blue). Most of the data (i.e. observed initial rates) are approximated from the first two figures in Yamamoto (1958), so some of our computation is not 100% correct for this example. As we fit the model, the NLS estimates are used as the realistic parameter prior to support the 24-run local D-optimal design of the new experiment. In this case, the hybrid nonlinear model can be written

as

$$\nu_i = \frac{\exp(a_0 + a_1 D_i + a_2 x_i + a_3 x_i D_i + a_4 {x_i}^2 + a_5 {x_i}^2 D_i) S_i}{k_0 + k_1 D_i + S_i} + \varepsilon_i, \tag{3.17}$$

where $x = (H - 7.4)/0.4$ and the $p = 8$ parameters are $k_0, k_1, a_0, \ldots, a_5$. Here, $D = \{0, 1\}$ is an indicator variable, the zero representing the methylene blue and the one representing the toluidine blue, $H \in [7, 8]$ denotes the hydrogen ion concentration or the pH level, S denotes the substrate (or dye) concentration, and ν is the initial rate (or velocity) of reaction. The variable space of the substrate concentration is set to $[0.02, 0.2]$M (molar) to contain the levels 1/48 and 1/6, adopted levels in the reference experiment. We also assume the error variance to be a constant $\mathbb{V}(\varepsilon_i) = \sigma^2$.

The indicator variable leads to four fixed effects in (3.17). Nevertheless, it is not recommended to divide the total experiment into two separate ones and then fit two unrelated models. In that case, the fitted models would be a bit different because of the errors, the independent experimental runs, and also the covariances of parameter estimators. In comparison, better results shall be found as we fit (3.17) to the 24-run dataset. However, due to the existence of the covariances, it is difficult to extrapolate the true parameter values of M-M kinetics as well as to make response predictions with the fitted model. There is no mechanistic model available in the example, so we stick to (3.17) as the alternative.

With the old dataset, we are unable to estimate a_3 and a_5. Hence the parameter prior can be set to $\boldsymbol{\theta}^0 = \{0.11281, -0.044306, 0.32276, -0.67747, 0.31409, 0, -0.10768, 0\}$. In the current space of the substrate concentration, the result implies that the toluidine blue is inferior to its rival all the time in terms of the initial rate maximisation. Our hybrid point exchange algorithm can be implemented in this case. With random starts from 30 tries, we find 18 distinct matrices $\mathbf{X}_{1\text{st}}$ after the discrete optimisation over the candidate set $\boldsymbol{\Omega} = \{0, 1\} \times \{-1, 0, 1.5\} \times \{1/48, 1/18, 1/6\}$. Note that the 2nd dimension is for the transformed variable x and the 3rd is for the substrate concentration. Up to this moment, the exchanging rule is $\max(d_i) > 1.1$ as the local D-criterion tries to maximise its relative improvement in each step of iterative updates.

In the subsequent continuous optimisation, the rule is set to $\max(d_i) > 1.00000001$. In each step of optimisation, there are also 18 extra initial points from a coarse set $\boldsymbol{\Omega} = \{0, 1\} \times \{-0.375, 0.25, 0.875\} \times \{0.06, 0.1, 0.14\}$ which covers the variable space (or design region) $\mathcal{X}$. In our demonstration, the continuous optimisation of each experimental run is simultaneous under the Nelder-Mead method. However, a quicker approach can exist: the alternative is to optimise the indicator variable of the point in one step and then optimise the continuous variables in another step. Besides, the quasi-Newton method can enable users to better define the indicator variables in the model and thus it allows for some extra flexibility. These options are sensible when the computation is intensive. On the other hand, the hybrid method is already quick enough in this example and the

elapsed computational time up to the end of the continuous optimisation is 1217 seconds in total and the interim solution we find is in Table 3.16.

Table 3.16: 24-Run Local D-Optimal Design for Model (3.17): Interim Solution after the Continuous Optimisation Part (2nd Stage) of Hybrid Fedorov Exchange Algorithm

D	H	S	D	H	S	D	H	S	D	H	S
0	7	0.2	0	7.6678	0.0588	1	7	0.2	1	7.6678	0.0455
0	7	0.2	0	8	0.0553	1	7	0.2	1	8	0.0426
0	7	0.2	0	8	0.0553	1	7	0.2	1	8	0.0426
0	7.5733	0.2	0	8	0.2	1	7.5733	0.2	1	8	0.2
0	7.5733	0.2	0	8	0.2	1	7.5734	0.2	1	8	0.2
0	7.5734	0.2	0	8	0.2	1	7.5734	0.2	1	8	0.2

All the 18 distinct matrices $\mathbf{X}_{1\text{st}}$ can lead to about the same local D-criterion values of $\mathbf{X}_{2\text{nd}}$. Their maximum is 11.3612, for the $\mathbf{X}_{\text{opt}}$ in Table 3.16. As a result, there are the same number of runs for the two dye species. This is because the eight model parameter estimators have equal importance under the local D-criterion. Before we finish the hybrid method as well as the modified adjustment algorithm, the closest distance is determined to be 0.1 for the pH level and 0.01 for the substrate concentration. As such, the final solution is

$$\mathbf{X}^{**}_{\text{opt}} = \left\{ \begin{array}{ccccc} (0,\,7,\,0.2) & (0,\,7.6,\,0.2) & (0,\,7.7,\,0.06) & (0,\,8,\,0.06) & (0,\,8,\,0.2) \\ 3 & 3 & 1 & 2 & 3 \\ (1,\,7,\,0.2) & (1,\,7.6,\,0.2) & (1,\,7.7,\,0.05) & (1,\,8,\,0.04) & (1,\,8,\,0.2) \\ 3 & 3 & 1 & 2 & 3 \end{array} \right\}, \tag{3.18}$$

the local D-criterion value of which is 11.3361. The modified adjustment algorithm is not quite essential in this demonstration as it does not make an improvement. Besides, as we assume the two parameter prior $a_3^0 = a_5^0 = 0$, the coordinate levels of pH are identical between the first and the third rows in (3.18). The numbers in the other two rows represent the numbers of replicates of the respective unique runs in the experiment. Most of the coordinates are located on the boundaries. As far as we are aware, the discrete candidate levels of the substrate concentration (which come from the reference experiment) are not suitable for the discrete optimisation of the hybrid method. We can also set the point parameter prior as $a_3^0 = a_2^0, a_5^0 = a_4^0$. This will not influence the support points in the first row in (3.18). In this scenario,

$$\mathbf{X}^{**}_{\text{opt}} = \left\{ \begin{array}{ccccc} (0,\,7,\,0.2) & (0,\,7.6,\,0.2) & (0,\,7.7,\,0.06) & (0,\,8,\,0.06) & (0,\,8,\,0.2) \\ 3 & 3 & 1 & 2 & 3 \\ (1,\,7,\,0.2) & (1,\,7.6,\,0.2) & (1,\,7.8,\,0.04) & (1,\,8,\,0.04) & (1,\,8,\,0.2) \\ 3 & 3 & 1 & 2 & 3 \end{array} \right\}, \tag{3.19}$$

which is quite similar to (3.18). It seems that the local D-optimal design of the experiment is quite robust to the modifications we make in the point parameter prior. As an alternative to the D-criterion, we consider a local Ds-criterion for the estimation of all

the parameters except for k_0 and k_1. In this situation, the Ds-optimal solution is

$$\mathbf{X}_{\text{opt}}^{**} = \left\{ \begin{array}{ccccc} (0,\ 7,\ 0.2) & (0,\ 7.6,\ 0.05) & (0,\ 7.6,\ 0.2) & (0,\ 8,\ 0.05) & (0,\ 8,\ 0.2) \\ 3 & 1 & 3 & 2 & 2 \\ (1,\ 7,\ 0.2) & (1,\ 7.6,\ 0.2) & (1,\ 7.7,\ 0.04) & (1,\ 8,\ 0.04) & (1,\ 8,\ 0.2) \\ 4 & 3 & 1 & 2 & 3 \end{array} \right\}, \tag{3.20}$$

which is still dependent on the specified prior values of k_0 and k_1. Compared with (3.19), the numbers of replicate runs are somewhat different and imbalanced. In addition, we see minor perturbations of the optimal coordinate levels, which leads to the maximisation of the local Ds-criterion function.

In practice, thorough mechanistic studies are difficult when there are various dye species and coliphages. As we can expect in this case, it requires a lot of experimentation to collect the massive dataset and then fit a complex model. Otherwise, there would be no adequate information to verify the mechanistic postulate. Therefore, some latent variables or effects are often excluded from the model, e.g. the empirical terms corresponding to the coefficients a_3, a_5 in (3.17). Nonetheless, it would be useful to define important categorical variables (or indicator variables) in the hybrid nonlinear model. As such, we can then integrate all relevant information into a comprehensive mechanism and find correlations between different factors. In this situation, as a complex model is assumed, optimal design of experiments can exert a constructive role to save us time and workload in real experimentation.

Sometimes not all estimates of the categorical variables are important or equally important. If one dye species is more likely to be favoured in the experiment, we can then consider either a local L-criterion or a composite Ds-criterion. This allows experimenters to alter the weight allocation in the criterion function, in order to emphasis the estimation of specific parameters. Under the local D-criterion, we can also fix the numbers of experimental runs for different categories: e.g. we can allocate 16 runs to methylene blue and 8 runs to toluidine blue, since in this first case, the expected initial rate is higher under the same conditions. No matter what we shall do, the ultimate purpose is to make a better use of experimental resources.

3.9 A General Discussion of Advanced Kinetics

The paramount role of the kinetic model is well appreciated in the establishment and interpretation of molecular mechanisms. It is vital to make the experimental data contain sufficient information such that we can fit kinetic models and make extrapolations. As also discussed in Cornish-Bowden (2014), it is worthwhile using a suitable design of the experiment for raw data collection, in order to fulfil specific purposes in relation to statistical modelling and inference.

In this chapter, we focus on simple Michaelis-Menten kinetics, the nonlinear model function of which involves two unknown parameters. As we also mentioned, depending on the specific steps in the reaction mechanism, a more complex theoretical model can be reasonable (Wong, 1975). It is common to write those advanced kinetic models in the forms of rational functions of the initial substrate concentration. For example, Halder and Crabbe (1984) studied a special mechanism that used two substrates in the reaction. Also in this advanced mechanism, experimenters must consider both the pH level and the product inhibition effect on the observed response. Even in simpler one-substrate reactions (like in M-M kinetics), if the dual substrate binding sites exist on the molecules of the enzyme, experimenters can assume second-order kinetics. To learn such kinetics, Gilmour and Trinca (2012a) illustrated some sophisticated reparametrisations to derive an intuitive mechanistic model.

On the topic of reparametrisation, Ratkowsky (1985, 1986) advised that we can in fact reparametrise varieties of nonlinear kinetic models. The new model function exhibits some stronger linear or close-to-linear behaviour. As such, the reparametrisation can make it easier to use nonlinear least squares for parameter estimation, since the classical NLS assumptions can be better satisfied for the more close-to-linear model. Nevertheless, the new model does not make much sense if our experimental interest is to estimate the original kinetic parameters (with minimal standard errors) and thus to interpret the metabolism. Hence experimenters can be reluctant to do too much reparametrisation, which then leads to difficult interpretations. In contrast, nonlinear transformation of the both sides of kinetic models is a more reasonable approach (Carroll and Ruppert, 1984; Ruppert et al., 1989). The main purpose of transformation is to reshape the error structure or error distribution of the fitted model. In the transform-both-sides model, the kinetic function between the dependent and the independent variables remains the same as in our conjecture. We will discuss some applications of this further in Chapter 5 of this thesis.

A complex looking metabolism can be elucidated in detail if we can write out the correct chemical equation, which is composed of some basic reaction steps. As we combine all the (parallel) opposing and consecutive steps, the appropriate kinetic model can be established on the basis of the simultaneous differential rate laws. If the model is complex and nonlinear, note that the NLS estimation relies on initial values of model parameters. We shall determine these initial values beforehand so that the nonlinear model can be fitted. To reduce this unwanted dependence, one can start with the most basic kinetic models (or similar hybrid nonlinear models). After the appropriate initial values are ascertained, the more complex model can be studied then.

For the studies of some rare kinetics (due to allosteric interactions or cooperatives, for instance) as well as in vitro drug metabolisms, Tracy and Hummel (2004) discussed different mechanistic models for one-substrate reactions. In the real experiment, the model selection requires sufficient kinetic evidence to support it, so it is sometimes

difficult to consider various metabolic regulations. To ascertain the best candidate model that suits the in vitro data, it is judicious to separate the lack of fit of candidate models from the observed pure error and then compare fitted models.

Experimenters can also consider the inhibition (or activation) effects on Michaelis-Menten kinetics. If these effects must be taken into account, some modifications can be made on the normal M-M model. When an inhibitor is involved in the reaction, for instance, the inhibitor concentration can be a factor. It is common to assume one of the candidate inhibition models (e.g. competitive, uncompetitive or mixed) first. We can then fit it to the kinetic data in order to validate the potential inhibition effect. As the mechanistic model can be derived in this case, we do not have to use a hybrid nonlinear model. For simple inhibition models in terms of two or three parameters, the closed-form expressions of the continuous local D-optimal experimental designs were found in Bogacka et al. (2011). In each of these optimal designs, however, there is an indispensable constraint. The number of support points is set to equal the number of model parameters, which is no more than three. Under the local D-criterion in this case, it is not too hard to determine the substrate concentration level and inhibitor concentration level for each of the support points.

Moreover, experimenters often encounter Michaelis-Menten kinetics in the process of drug development. In these cases, the simple kinetic model can be found and written in terms of the time in the full course of reaction (Schnell and Mendoza, 1997). As such, the independent variable is the reaction time when the substrate concentration level is fixed at an optimal constant. Of course, the response of the experiment shall be the rate at time t instead of the initial rate measured at $t \to 0$. In the research areas of pharmacokinetics and pharmacodynamics, there are many examples of optimal experimental designs, which we will not discuss further.

In the next chapter, we will continue to concentrate on the nonlinear multifactor model. The most striking new development is that some efficient Gaussian quadrature rules shall be applied to evaluate the expected D-criterion function. The expected D-criterion function is an integral based on the local D-criterion function we have used in the last two chapters. In comparison, its criterion function is much more complex and it can be quite robust to minor variations of the true parameter values. The complex expected D-criterion function shall be used in the search for optimal experimental designs for nonlinear models. Hence, we must adapt the current hybrid method to implement a suitable Gaussian quadrature rule.

Chapter 4

Bayesian Optimal Design of Nonlinear Multifactor Experiments

4.1 Expectation of the D-criterion Function

4.1.1 Local Optimality Criterion

The previous chapters have informed us the relationship between the optimal designs of experiments and the Fisher information of the assumed model. When the model is nonlinear, we should choose the parameter prior of the p-element vector $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_p\}$, in order to calculate and maximise the local D-criterion function ϕ_{D}, for instance. If we use the D-criterion, the aim will be to minimise the variances and covariances of the nonlinear least squares (NLS) estimator of $\boldsymbol{\theta}$. This is desirable in most experiments that involve the fitting of statistical models, though there are various optimality criteria available for other experimental purposes linked to statistics.

In particular, ϕ_{D} is a local criterion function. This is because we use the point parameter prior vector $\boldsymbol{\theta}^0$ to substitute the unknown true parameters $\boldsymbol{\theta}$ in the Fisher information matrix of the criterion function. Therefore, the performance and robustness of the local optimal design relies on the deviation $\boldsymbol{\theta}^0 - \boldsymbol{\theta}$. This is a disadvantage of using a local optimality criterion. When we assume an inaccurate parameter prior for the nonlinear model, the local criterion might not be suitable and the obtained optimal design could be much less efficient than we expect. To deal with this problem, in this chapter, we introduce a pseudo-Bayesian approach for optimal design of nonlinear multifactor experiments. As we will show, it is then plausible to replace the local D-criterion with an *expected D-criterion*, which takes account of the assumed uncertainties about the

true parameter values. In this case, the obtained optimal designs would be more robust to minor variations of the point parameter prior.

Suppose we can select a design $\mathbf{X}$ of n experimental runs ($n \geq p$). The model of the observed response can be written as $Y_i = f(\mathbf{X}_i, \boldsymbol{\theta}) + \varepsilon_i$ for $i = 1, 2, \ldots, n$. Here, $f(\mathbf{X}_i, \boldsymbol{\theta})$ is the model function in terms of v controlled variables in the ith row of $\mathbf{X}$. The error term is ε_i which should follow the essential assumptions in nonlinear regression. We can then write the local D-criterion function as $\phi_{\mathrm{D}} = \log|\mathbf{F}^{\mathrm{T}}\mathbf{F}|$. As to the model,

$$\mathbf{F} = \frac{\partial f(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

denotes the numeric $n \times p$ design matrix which depends on the parameter prior $\boldsymbol{\theta}^0$. When we maximise the criterion function ϕ_{D}, the ultimate aim is to minimise the *Generalised Variance* of the parameter estimator $\hat{\boldsymbol{\theta}}$ and thus to make the estimation precise with the collected experimental data. Let $\mathbf{Y}$ be a column vector of the n responses and $\boldsymbol{\varepsilon}$ be a column vector of the respective errors. With a first-order linearisation around the determined centre $\boldsymbol{\theta}^0$, the nonlinear model is in a matrix form $\mathbf{Y} \approx \mathbf{F}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, where the errors due to the *Taylor series approximation* (not to be confused with the model errors) are omitted.

If none of the model parameters is a nuisance, a $\mathbf{X}$ is local D-optimal if it could maximise the function ϕ_{D}, which is equivalent to minimising the variance-covariance matrix of the least squares estimator $\mathbb{V}(\hat{\boldsymbol{\theta}}) = (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\sigma^2$, where the error variance σ^2 is assumed to be homogeneous across the n uncorrelated experimental observations. We are most interested in the Fisher information matrix $\mathbf{F}^{\mathrm{T}}\mathbf{F}$, which is linked to the precision of $\hat{\boldsymbol{\theta}}$.

It is crucial to determine an appropriate $\boldsymbol{\theta}^0$ for the local D-criterion function. Commonly, experimenters can either use an old parameter estimate recorded in the literature or come up with a *subjective prior estimate* as their "best guess" of the unknown parameters. Moreover, as shown in most of our examples, we can fit the required model to the data from one or more reference experiments conducted under the same or similar environments. Besides, if the relevant data can be found to obtain the reference parameter estimate $\tilde{\boldsymbol{\theta}}$, we shall also be able to consider multiple candidate models and make comparisons. Remind that when the model function is nonlinear in the parameters, the local optimality criterion must assume a numeric parameter prior vector to enable the evaluation of its criterion function. In this case, we have to find a reliable reference estimate $\tilde{\boldsymbol{\theta}}$ (i.e. close to the unknown true parameter values) for a justification of the local optimal design. However, this is not achievable in many circumstances.

4.1.2 The Pseudo-Bayesian Approach

In practice, there are more or less uncertainties about the true values of $\boldsymbol{\theta}$. At times, there would be a lack of sufficient information for using the local D-criterion, as we

cannot ensure the prior estimate $\boldsymbol{\theta}^0$ to be accurate. To deal with these situations, it is beneficial to think about a vaguer definition of the parameter prior. To be specific, we could consider the reference estimate $\tilde{\boldsymbol{\theta}}$ as the mean vector of a multivariate prior distribution ($\rho(\boldsymbol{\theta})$ denotes its probability density function), the variances and covariances of which are chosen to reflect the uncertainties about $\boldsymbol{\theta}$ beforehand.

To incorporate this information about the prior distribution, the idealistic aim is to do numerical evaluations of the expectation of the D-criterion function of the different $\mathbf{X}$, in order to compare their utilities with respect to the criterion. As reviewed in Chaloner and Verdinelli (1995), for the nonlinear model, the *expected criterion function* can be simplified to be a multiple definite integral

$$\varphi = \int_{\boldsymbol{\theta}} \phi(\mathbf{X}|\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta},$$

where $\phi(\mathbf{X}|\boldsymbol{\theta})$ is the local criterion function in terms of $\boldsymbol{\theta}$ within its parameter space. In our demonstrations in this chapter, the parameter space is defined by the density function $\rho(\boldsymbol{\theta})$ only and in the integral above, we shall substitute the general criterion function form $\phi(\mathbf{X}|\boldsymbol{\theta})$ with the local D-criterion function ϕ_{D}. As we can see, the information about the prior distribution is expressed in the form of the joint probability density function (p.d.f.) $\rho(\boldsymbol{\theta})$ of the unknown parameters.

Let $\boldsymbol{\mu}$ be the prior mean and $\boldsymbol{\Sigma}$ be the prior variance-covariance matrix. For example, we assume $\rho(\boldsymbol{\theta})$ to be derived from a multivariate normal distribution $\boldsymbol{\theta} \sim \mathbb{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This is the most common case when there is no constraint on the continuous space $(-\infty, +\infty)$ for each parameter of φ. The aim is to find the *Bayesian optimal design* which maximises the expected D-criterion function under the assumption of the parameter prior distribution. Since we must implement an exchange algorithm for the computer search, the multiple integration of φ must be performed frequently for numerous $\mathbf{X}$ evaluated in the search.

The direct integration of φ is an unrealistic solution as that requires too intense computation to afford. Hence, researchers shall look for a simpler but still accurate numerical approximation of the expected D-criterion function. Most of the time, the first step of a numerical approximation is to draw from the specified parameter prior density $\rho(\boldsymbol{\theta})$ a number of times to obtain a representative sample of $\boldsymbol{\theta}$. As such, at least we can use the sample to evaluate the local D-criterion values ϕ_{D} of a specific $\mathbf{X}$, when we choose the different point parameter priors. The individual evaluations are then combined so that the approximation of φ is the weighted mean of those local D-criterion values. This is the criterion-evaluation approach for Bayesian optimal design. In the literature, it could sometimes be referred to as a pseudo-Bayesian, semi-Bayesian or average approach. As we plan to use nonlinear least squares to fit the model, this approach does not imply Bayesian inference, though an expected criterion can take that into account too.

The most well-known method is random sampling and it requires a considerable number of random draws from the multivariate prior distribution. The Monte Carlo sample in this case must be large enough in order to be representative for the approximation of the integral φ. Optimal design of a market choice experiment in Kessels et al. (2009) relied on such a huge Monte Carlo sample, which was drawn from a multivariate normal parameter prior distribution for a simple logistic model. For the sake of simplification, one sometimes assumes independence between the individual estimates of $\boldsymbol{\theta}$, in particular due to the lack of sufficient prior information about the covariances. As such, the variance-covariance matrix $\boldsymbol{\Sigma}$ becomes diagonal as specified, for example, in optimal experimental design in Gilmour and Trinca (2012a). In this case, it would be reasonable to draw p univariate random samples for each parameter and moreover, this flexible assumption allows us to consider and incorporate different univariate prior distributions in $\rho(\boldsymbol{\theta})$. When such kinds of decisions have to be made, it is useful to refer to the properties of different exponential families as well as the constraints on the parameter space of the specific model.

In the light of the rectangle rule in unidimensional cases, a definite integral can be approximated with numerous narrow rectangles of which the heights can sketch the upper bounding line of the area. As long as we take m random draws from the specified distribution of the parameter prior, the local D-criterion function varies across different vectors of $\boldsymbol{\theta}^0$. Hence, to approximate the expectation of the D-criterion function, we can take the arithmetic mean

$$\tilde{\varphi}(m) = \frac{1}{m} \sum_{r=1}^{m} \phi_{\mathrm{D}}(\mathbf{X}|\boldsymbol{\theta} = \boldsymbol{\theta}^r) \approx \varphi, \ m \to \infty. \tag{4.1}$$

This is also referred to as the pseudo-Monte Carlo (PMC) method in Bliemer et al. (2008). There are p model parameters to be integrated out and the integrand function of φ is often complex in terms of the design $\mathbf{X}$. Note that the approximation $\tilde{\varphi}(m)$ is not invariant to different Monte Carlo samples, unless we could define a sample size $m \to \infty$. However, if the sample is too large, the computation would be quite expensive, even if the final PMC approximation is proven to be accurate. Herein, $\boldsymbol{\theta}^r$ is the rth unit of the Monte Carlo sample of the parameter prior density $\rho(\boldsymbol{\theta})$. In optimal design of experiments, the expected D-criterion function φ should be maximised for a $\mathbf{X}$. As such, an exchange algorithm would use the approximating function $\tilde{\varphi}(m)$ to calculate and compare the utilities or relative efficiencies of various $\mathbf{X}$.

A special case is when we combine the continuous nonlinear optimisation with an exchange approach, which is the core idea in the new *optimised exchange algorithm* as well as our most recommended *hybrid method* described in the last two chapters. While a lot of the algebraic computation would be complicated in this case, it could take a long time to complete the task with a conventional software like Matlab and R. Hence, it seems impractical to draw a huge Monte Carlo sample of the model parameters. In

other words, for the PMC approximation in each iteration of the algorithm, we cannot afford too much computation to evaluate different local D-criterion functions in (4.1).

To facilitate the applications that have to deal with an uninformative parameter prior in the expected D-criterion function, in this chapter, we consider an efficient numerical integration method for an accurate approximation of the integral φ. With a small sample size m to be chosen, we can draw a nonrandom sample of the model parameters. As a specific weighted mean of the individual local D-criterion values is taken, the approximation using the deterministic sample can be even more accurate than most PMC approximations. This clever method will be elucidated and then applied to Bayesian optimal design of nonlinear multifactor experiments.

4.2 Deterministic Gauss-Hermite Approximation

As an alternative to the pseudo-Monte Carlo method, the Gaussian quadrature rules are useful numerical integration tools. Each of the Gaussian quadrature rules corresponds to a specific interpolating function and therefore can be adapted to approximate a whole class of relevant definite integrals. A Gaussian quadrature is often stated as a weighted sum (or mean) of the function values of the interest, which should be evaluated at some identified points (i.e. abscissas). These quadrature points can be sampled and determined according to the relevant integrand function as well as a fixed sample size m.

The quadrature rule can provide a much simplified solution of φ, the multiple integral that contains the local D-criterion function. First, we can assume a multivariate normal prior distribution for $\boldsymbol{\theta}$, for instance. In this situation, the Gauss-Hermite (GH) quadrature rule is feasible for numerical integration as we require, the interpolating function of which is a *Hermite polynomial* of a fixed order. As the interpolating function is determined, we can draw the Gauss-Hermite sample of m abscissas, all the units of which are deterministic and nonrandom. For now, let us look at how the GH approximation can solve integrals in a simple and general form. Suppose the random variable is $z \in (-\infty, +\infty)$ in the univariate case, the Gauss-Hermite quadrature can approximate the following form of definite integral as

$$\int_{-\infty}^{+\infty} \phi(\mathbf{X}, z) e^{-z^2} dz \approx \sum_{r=1}^{m} w^{(r)} \phi(\mathbf{X}, z^{(r)}), \tag{4.2}$$

where $z^{(r)}$ and $w^{(r)}$ represent the abscissa and the associated weight of the rth Gauss-Hermite draw, for $r = 1, 2, \ldots, m$. As we assume $z \in \mathbb{R}$, $\phi(\mathbf{X}, z)$ is then an explicit function of $\mathbf{X}$, which is made up of controlled variables. No matter what form the function ϕ takes, the GH quadrature can lead to an accurate approximation when it has an appropriate sample size m. Under the GH quadrature in (4.2), we should find the m abscissas and the respective weights, in order to evaluate the linear terms on the

right hand side. While each abscissa $z^{(r)}$ should be one of the m roots of the orthogonal Hermite polynomial, the associated weight $w^{(r)}$ denotes the relative importance of $z^{(r)}$ in the Gauss-Hermite sample.

An m-point Gaussian quadrature rule can lead to an exact approximation (i.e. a perfect integration without an error) and thus an error-free numerical integration of φ, when the function ϕ of the integrand is a polynomial of degree $2m-1$ or less. Here, we have the local D-criterion function ϕ_{D}, which is nonlinear in the model parameters $\boldsymbol{\theta}$. According to Taylor's theorem, this nonlinear function can be approximated with a linear response surface, which requires an expansion of ϕ_{D} about the mean $\boldsymbol{\mu}$ (or $\tilde{\boldsymbol{\theta}}$) of the parameter prior density $\rho(\boldsymbol{\theta})$. The higher order is the linear response surface, the more accurate approximation of the nonlinear function would be achievable. In other words, the larger the overall Gauss-Hermite sample size of the model parameters, the more accurate numerical integration of φ we can obtain from (4.2). When we use the quadrature in optimal design of experiments, it is wise to find an acceptable balance between the total sample size m and the expected error of the approximation $\hat{\varphi}$.

We can expand the above discussion to the multivariate case, when there are p random variables (i.e. $\mathbf{z} = \{z_1, z_2, \ldots, z_p\}$) to be integrated out of φ. The multiplication rule applies to the Gauss-Hermite quadrature and it requires p individual steps for numerical integration. As a result, it implies the approximation

$$\int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \phi(\mathbf{z}) \prod_{j=1}^{p} e^{-{z_j}^2} dz_1 \ldots dz_p \approx \sum_{r_1=1}^{m_1} \ldots \sum_{r_p=1}^{m_p} w_1^{(r_1)} \ldots w_p^{(r_p)} \phi(z_1^{(r_1)}, \ldots, z_p^{(r_p)}), \tag{4.3}$$

where $m = m_1 \times m_2 \times \ldots \times m_p$ indicates all combinations of the p univariate and uncorrelated Gauss-Hermite subsamples for the individual variables $z_1, \ldots, z_p$. If $m_1 = m_2 = \ldots = m_p = \mathfrak{m}$ such that total sample size is $m = \mathfrak{m}^p$, we have identical abscissas and weights in these Gauss-Hermite subsamples for $z_1, ..., z_p$. The assumption is that we have a multivariate normal prior distribution for $\boldsymbol{\theta}$. Therefore, in our case, we should rearrange the expected D-criterion function φ in order to use the Gauss-Hermite approximation in (4.3) (Bliemer et al., 2008).

For $j = 1, 2, \ldots, p$, if we define $\theta_j \sim \mathbb{N}(\mu_j, \Sigma_{jj})$ as an independent univariate prior distribution, which means that the variance-covariance matrix $\boldsymbol{\Sigma}$ is diagonal, the r_jth sampling unit (i.e. sampled prior value) of the parameter θ_j can be calculated as

$$\theta_j^{(r_j)} = \mu_j + \sqrt{2\Sigma_{jj}} z_j^{(r_j)}.$$

We can find all the abscissas from the Gauss-Hermite quadrature, so it is straightforward to calculate the corresponding parameter prior using the formula above. In the numerical integration of φ, the associated weights $\omega_j^{(r_j)}$ of the prior value $\theta_j^{(r_j)}$ can also be obtained from the associated weights $w_j^{(r_j)}$ of the abscissa $z_j^{(r_j)}$. The formula is $\omega_j^{(r_j)} = w_j^{(r_j)} / \sqrt{\pi}$,

as we replace $\rho(\boldsymbol{\theta})$ of the integral φ with the modified p.d.f. of a univariate standard normal distribution. As we assume the independence between the individual parameter prior distributions (or the sampled values), this shows how we can compute the sampling units and the associated weights for a Gauss-Hermite quadrature.

On the other hand, it is also viable to consider a multivariate normal distribution for the whole vector $\boldsymbol{\theta}$, all the covariances of which are zero. In reference to Kessels et al. (2009), a Cholesky decomposition can be used on the variance-covariance matrix so that $\boldsymbol{\Sigma} = \mathbf{D}\mathbf{D}^{\mathrm{T}}$, where $\mathbf{D}$ is a $p \times p$ lower triangular matrix with all positive entries in its diagonal. In (4.3), for $r = 1, 2, \ldots, m$, the *composite weight* $w^{(r)} = w_1^{(r_1)} \ldots w_p^{(r_p)}$ is the product of p GH subsample weights from each step of numerical integration. Likewise, for a Gauss-Hermite approximation of φ, the composite weight function should be $\omega^{(r)} = w^{(r)}/\pi^{p/2}$. In terms of the vector $\mathbf{z}^{(r)}$, which is one combination of the abscissas for different parameters, the rth parameter prior vector is calculated as

$$\boldsymbol{\theta}^r = \boldsymbol{\mu} + \sqrt{2}\mathbf{D}\mathbf{z}^{(r)}$$

for $r = 1, 2, \ldots, m$, where $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{\theta}^r)$ is the mean of the multivariate normal density $\rho(\boldsymbol{\theta})$ and $\mathbf{z}^{(r)} = \{z_1^{(r_1)}, \ldots, z_p^{(r_p)}\}$ is the rth unique combination of the abscissas from the independent univariate Gauss-Hermite subsamples of $\boldsymbol{\theta}$. In line with the above, if $\boldsymbol{\Sigma}$ is diagonal, then so is $\mathbf{D}$. As such, all nonzero entries of $\mathbf{D}$ are the standard deviations of the multivariate normal prior density $\rho(\boldsymbol{\theta})$ and the calculated parameter prior vectors or units of the GH sample are independent of each other.

To summarise, one can draw a Gauss-Hermite sample of the parameters and calculate the composite weights, since the expected criterion function can be rewritten as

$$\varphi = \int_{\boldsymbol{\theta}} \phi(\mathbf{X}|\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \phi(\mathbf{X}|\mathbf{z}) \prod_{j=1}^{p} e^{-{z_j}^2} d\mathbf{z}.$$

On the right hand side of this equation, we replace the parameters $\boldsymbol{\theta}$ with the set of random variables $\mathbf{z}$ and the whole integral has to be divided by a constant $\sqrt{\pi}$. To approximate φ with the Gauss-Hermite quadrature, under the D-criterion, there is

$$\phi_{\mathrm{BD}} = \hat{\varphi} \approx \sum_{r=1}^{m} \omega^{(r)} \phi_{\mathrm{D}}(\mathbf{X}, \boldsymbol{\theta} = \boldsymbol{\theta}^r) \text{ with appropriate constants } m_1, m_2, \ldots, m_p. \quad (4.4)$$

A reliable sample should lead to a small approximation error. ϕ_{BD} denotes the approximation of the expected D-criterion function whereas $\omega^{(r)} = w_1^{(r_1)} w_2^{(r_2)} \ldots w_p^{(r_p)}/\pi^{p/2}$ is the composite weight corresponding to the rth parameter prior vector $\boldsymbol{\theta}^r$ from the Gauss-Hermite sample of the parameters. When we fix the subsample sizes as $m_1 = m_2 = \ldots = m_p = \mathfrak{m}$, implying the equal importance of the p parameter estimates fro the fitted model, the total GH sample size is $m = \mathfrak{m}^p$.

As has been mentioned, the GH subsamples made of abscissas and weights are independent of the statistical model as well as the normal parameter prior distribution $\rho(\boldsymbol{\theta})$. The abscissas for each parameter are distributed symmetrically around zero whereas the sum of the associated weights (as well as the sum of the composite weights of the whole sample) should be equal to one. In the simplest case, when the GH subsample size is $\mathfrak{m} = 2$ for each parameter, the two identical abscissas are ± 0.7071067811865475 for each of the p random variables in $\mathbf{z}$ and all the composite weights w are fixed at 0.5^p.

We now consider the Gauss-Hermite approximation ϕ_{BD} in (4.4), where a sufficient number of abscissas should be selected for each parameter of $\boldsymbol{\theta}$. As the aim is to integrate out the parameters of the interest, the GH subsample sizes depend on the Fisher information matrix $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ of the model, as well as the variance-covariance matrix $\boldsymbol{\Sigma}$ for the vague parameter prior. To make up for the downsides of using a local criterion, the assumed expected criterion is preferred. Therefore, we are attempting to extrapolate the true level of φ for a specific $\mathbf{X}$. The appropriate GH sample should be drawn so that we can obtain an accurate approximate value ϕ_{BD}, without requiring too intensive computation in optimal design of experiments. Afterwards, we wish to evaluate and compare the relative efficiencies of various $\mathbf{X}$. As the new function ϕ_{BD} is based on a far smaller sample size than the traditional PMC approximation, this task becomes easier.

In principle, the larger are the variances or covariances in $\boldsymbol{\Sigma}$ of the normal prior distribution, the more GH sampling units have to be defined to construct ϕ_{BD}. With a multinomial logistic regression model used for market choice experiments, Yu et al. (2010) compared a series of sampling methods and the respective approximations of φ. In their examples, the controlled variables were categorical whereas the univariate normal distributions of the individual parameter priors were uncorrelated and had zero means and identical variances.

We have introduced the traditional PMC approximation and the deterministic Gauss-Hermite quadrature. Moreover, the spherical-radial (SR) transformation is another efficient method for numerical integration of the expected criterion function (Monahan and Genz, 1997; Gotwalt et al., 2009). This works when the parameter prior distribution is specified to be multivariate normal. This is quite a promising method which we will come back to introduce in Section 4.8. At the moment, suppose one would like to obtain an accurate approximation ϕ_{BD} with the Gauss-Hermite quadrature. In this situation, the simulation results in Yu et al. (2010) would recommend us to choose equal GH subsample sizes $\mathfrak{m} \geq 2$, when the uniform prior variance is as small as 0.04 for the logistical model parameters. However, this standard varies depending on specific applications, so it is useful to examine the errors of the Gauss-Hermite approximations in different orders when we change the subsample sizes $m_1, m_2, \ldots, m_p$. The total sample size m can also be quite large sometimes.

The most common situation is to assume a multivariate normal distribution of $\rho(\boldsymbol{\theta})$. Nevertheless, if the variance-covariance matrix $\boldsymbol{\Sigma}$ is envisaged as diagonal, it is feasible to tailor different exponential family distributions to the individual parameter priors. At times, there are alternative Gaussian quadrature rules which we can resort to. For instance, the Gauss-Legendre rule can be adopted when we assume a *uniform prior* distribution and the details were described in Matthews and Allcock (2004); Bliemer et al. (2008). Furthermore, Goos and Mylona (2013) extended the application of the Gaussian quadrature rules in optimal design of experiments. Their new results shall enable us to handle different expected criterion functions with *lognormal, beta* or *gamma* distributions for the parameter priors, each corresponding to an accurate approximation of a different class of definite integrals.

In comparison, despite the modification considered in Gotwalt et al. (2009), the SR transformation is tailored to multivariate normal prior distributions and otherwise its approximation is not so accurate. In this case, we would prefer the more flexible Gaussian quadrature rules when we are required to draw a small sample of the parameters.

4.3 Determination of a Reliable Gauss-Hermite Sample

Before we can consider optimal design of experiments, a crucial task is to construct an appropriate function ϕ_{BD} to approximate the expected D-criterion function. To find the extrapolated values of φ, the Gauss-Hermite sample size relies on the number of model parameters to be integrated out as well as the variance-covariance matrix of the parameter prior distribution. As has been mentioned, an acceptable balance must be identified between the computational effort to make and the total number of Gauss-Hermite draws. In this section, we are to show how to find such a balance with some quick examination of the approximation error.

The emphasis of our applications is on biochemical experiments. In each example, also discussed in Chapters 2 and 3, we assume a hybrid nonlinear model, which combines some valid mechanistic information with our simple empirical speculation. In a biological kinetics experiment in Martins et al. (1999), the assumed Michaelis-Menten mechanism for a fundamental chemical reaction was studied. As usual, that required a statistical model to be fitted with (nonlinear) least squares estimation.

In this case, the simple kinetics can be illustrated with the theoretical Michaelis-Menten model. The complete dataset involves 30 independent experimental observations. While the other experimental conditions must be the same (i.e. with constant temperature at 30 °C and the pH at 6.5), there are two controlled variables which could be linked to the observed response ν (mM/min), the initial reaction rate. In the old experiment, the biochemical reaction would occur in a 2-ml mixture of two reagents. The initial rate of

GSH (i.e. the desirable product) formation should therefore depend on the initial S-D-lactoylglutathione (i.e. the substrate) concentration and the initial glyoxalase II (i.e. the enzyme or the so-called protein catalyst) concentration. To learn the quantitative relationship between the three factors, the Michaelis-Menten model could offer a best fit to the data in Martins et al. (1999). However, our first example is based on a more complex hybrid nonlinear kinetic model as follows:

$$\nu_i = \frac{\exp(a_0 + a_1 x_i + a_2 {x_i}^2) S_i}{k + S_i} + \varepsilon_i \tag{4.5}$$

for $i = 1, 2, \ldots, 30$. Here S is the substrate concentration, E is the enzyme concentration, of which x in the above hybrid model denotes the scaled variable. We should also assume the error ε_i to have zero mean and constant variance across the 30 observations. A special case is that, if $a_1 = 1, a_2 = 0$ and the scaled independent variable is $x = \log(E)$, (4.5) is then simplified to the theoretical Michaelis-Menten model. In this example, however, we assume the mechanistic model to be unknown, no matter how simple it is. As such, the best alternative approach is to select the kinetic model (4.5) for NLS estimation.

Notice that the empirical function $a_0 + a_1 x_i + a_2 {x_i}^2$ is a second-order linear function, where we can also add or delete the corresponding terms and parameters for the overall regression model. As a result, this empirical function is a component of the hybrid nonlinear model which involves the enzyme concentration as its second controlled variable. To approximate the true Michaelis-Menten mechanism, we must establish the most appropriate and concise form of the model. The same idea can also suit other complicated Michaelis-Menten mechanisms, when there are some different controlled variables in addition to the substrate concentration. This is also elucidated in Chapter 3.

In view of the old experiment in Martins et al. (1999), we should best approximate the $n = 30$ old observations in order to compile a reference dataset. According to their figures and annotations, we can also define a continuous variable space $\mathcal{X}$ and assume it to be of the new experimental interest. In this example, the substrate concentration must be located in the range $[0.15, 3]$(mM) and as to the enzyme concentration (or the protein weight), there should be $E \in [0.02, 0.12]$(mg). A simple linear transformation seems to be reasonable so that $x_{\mathrm{E}} = (E - 0.05)/0.07 \in [-1, 1]$. In optimal design of the new experiment we conceive, each of the 30 experimental units (x, S) should be specified within the bounded rectangular region $\mathcal{X} = [-1, 1] \times [0.15, 3]$. In addition, nonlinear least squares estimation assumes that $\mathbb{E}(\varepsilon) = 0$ and $\mathbb{V}(\varepsilon) = \sigma^2$. With the Levenberg-Marquardt algorithm and the reference dataset, model (4.5) is fitted as

$$\nu_i = \frac{\exp(-6.4086(\pm 0.0366) + 0.8383(\pm 0.0175) x_i - 0.2861(\pm 0.0329) {x_i}^2) S_i}{0.3122(\pm 0.0591) + S_i} + \varepsilon_i,$$

with four decimal places for the parameter estimates and their standard errors in the brackets behind. In Bayesian optimal design for this nonlinear response surface, the NLS estimate $\tilde{\boldsymbol{\theta}} = \{\tilde{k}, \tilde{a}_0, \tilde{a}_1, \tilde{a}_2\} = \{0.3122, -6.4086, 0.8383, -0.2861\}$ is therefore set to be the mean vector $\boldsymbol{\mu}$ of the parameter prior density $\rho(\boldsymbol{\theta})$ for the expected D-criterion

function φ. In our first demonstration in this example, we assume $p = 4$ independent univariate normal prior distributions of $\boldsymbol{\theta}$; i.e. $\boldsymbol{\Sigma}$ is a diagonal variance-covariance matrix for $\rho(\boldsymbol{\theta})$. While the prior covariances are all equal to zero, the four prior variances can be linked to the reference standard errors shown above.

Under nonlinear least squares, the parameter estimates are dependent on each other, though we now consider their prior distributions to be independent. To assume a vague parameter prior (i.e. less information for the new experiment), the prior variances of $\rho(\boldsymbol{\theta})$ should be set to be somewhat larger than the reference standard errors in the fitted model. In this case, there are 30 old experimental units so that the reference parameter estimate $\tilde{\boldsymbol{\theta}}$ is rather precise and the model fits the reference data quite well. Hence, to determine the diagonal elements of the matrix $\boldsymbol{\Sigma}$, we use 10 times the corresponding variance estimates of $\tilde{\boldsymbol{\theta}}$. As such, the standard deviations of the parameter prior distribution are chosen as $\{0.1868, 0.1159, 0.0554, 0.1040\}$, more than three times the standard errors we obtain from the old experimental data. Depending on experimenters' conjecture, the prior standard deviations can be modified to reflect how informative the parameter prior should be.

Given that we assume the independence between the parameter prior vectors, the error of the Gauss-Hermite approximation ϕ_{BD} depends on the individual subsample size for the numerical integration of the expected criterion function φ with respect to each parameter in $\boldsymbol{\theta}$. In this case, we could determine some appropriate numbers m_k, m_{a_0}, m_{a_1} and m_{a_2} and then calculate the total GH sample size m. To facilitate our computation, we can do some quick numerical investigation beforehand, in order to find a simplified and refined form of the approximation ϕ_{BD} in (4.4).

As we see from the hybrid nonlinear model, while we derive the local D-criterion function, the component $\exp(a_0)$ is a constant to be taken out of the Fisher information matrix of (4.5). This implies that even if we consider the variance of the univariate prior distribution $\rho(a_0)$ to be nonzero, the prior value of the second parameter a_0 would not affect our search for the Bayesian D-optimal design. We can then fix it at the mean $a_0 = \tilde{a}_0$ for instance, so this GH subsample size is $m_{a_0} = 1$. In addition, the local D-criterion function is $\phi_{\text{D}} = \log|\mathbf{F}^{\text{T}}\mathbf{F}|$. The sampled prior value of a_0 should be distributed around the mean $\tilde{a}_0$ of the prior distribution $\rho(a_0)$. In this sense, even if $m_{a_0} > 1$ and we incorporate the subsample of a_0 to construct the overall approximating function ϕ_{BD}, the approximate value would remain the same. Note that this is not the case if we choose the PMC method or the SR transformation with some introduced randomness, though it is still advisable to fix the prior value of a_0.

Since the GH subsample size of a_0 is one, there is one fewer numerical integration step for the Gauss-Hermite approximation, when the expected D-criterion function φ is to be evaluated for each new experimental design $\mathbf{X}$. We then focus on the subsample sizes of the other three parameters. The Gauss-Hermite quadrature will be an exact

approximation of φ, in terms of the jth parameter of $\boldsymbol{\theta}$, if the local D-criterion function ϕ_{D} can be depicted as a linear response surface of degree $2m_j - 1$ or less. However, this is often not the case and a m_jth-order linear approximation of ϕ_{D} would lead to a certain error so that the numerical integration of φ will not be exact. The realistic question is that how much should that error affect the Gauss-Hermite approximation of φ. If we know the answer, we can then decide an appropriate GH sample.

We can present a simple conjecture first. If the local D-criterion function ϕ_{D} or even the model function $f(\mathbf{X}, \boldsymbol{\theta})$ is close-to-linear in each model parameter from $\boldsymbol{\theta} \in (-\infty, +\infty)$, the Gauss-Hermite quadrature would be an accurate approximation of the expected D-criterion function φ, the expected error of which can be ignored. In that situation, the total GH sample size m will still be small so that we can construct a simple function ϕ_{BD} in (4.4). Besides, when the model function tends to be close-to-linear, the corresponding Fisher information (with the first-order linearisation of the assumed nonlinear model) would be more robust to the variations in the true parameter values. In this case, the NLS estimates will be more reliable and so is the local D-optimal design of the experiment. Hence, when the Gauss-Hermite approximation is accurate and yet the relevant GH sample size m is small, the Bayesian D-optimal design might be found to resemble its local optimal version of which the point parameter prior is the reference estimate $\tilde{\boldsymbol{\theta}}$.

With the expected criterion function φ, we assume a normal parameter prior density $\rho(\boldsymbol{\theta})$ with the mean $\boldsymbol{\mu} = \tilde{\boldsymbol{\theta}}$ and the diagonal variance-covariance matrix $\boldsymbol{\Sigma}$. As we approximate φ with the Gauss-Hermite quadrature, the parameter prior values in each GH subsample are distributed around the mean $\tilde{\boldsymbol{\theta}}$, with their deviations depending on the specified prior variances. We envisage that a GH sampling unit $\boldsymbol{\theta}^r$ departs far away from the centre $\tilde{\boldsymbol{\theta}}$. In this case, a supposed Taylor series approximation of the nonlinear function ϕ_{D} or $f(\mathbf{X}, \boldsymbol{\theta})$, with a low-order linearisation around the prior mean $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, will not be too accurate at $\boldsymbol{\theta} = \boldsymbol{\theta}^r$.

If the variances are larger in the parameter prior density $\rho(\boldsymbol{\theta})$, it will be more difficult to approximate the expected criterion function. Within the parameter space, the Gauss-Hermite quadrature would lead to some longer distances between the sampled parameter prior vectors and the mean $\tilde{\boldsymbol{\theta}}$. As a result, the local D-criterion functions are less close-to-linear when we use one prior vector of the sample to substitute the unknown parameters in ϕ_{D}. To ensure an accurate approximation in this situation, we should increase the total GH sample size for the approximation in (4.4), since the interpolation technique demands more prior vectors.

4.4 Numerical Investigation

4.4.1 Model and Parameters

While the nonlinear model could have been established as (4.5), it is sometimes viable to do some kind of reparametrisation to obtain a more close-to-linear model (Ratkowsky, 1985). With the simple Michaelis-Menten model, the suitable reparametrisation was found in Ratkowsky (1986) through some simulation studies, which can reduce the bias of the NLS parameter estimator and also make it more plausible to assume a multivariate normal density $\rho(\boldsymbol{\theta})$. Likewise, if we consider a similar reparametrisation for (4.5) (note that it is not the same as that for the Michaelis-Menten model), for instance, the kinetic model can be rewritten as follows:

$$\nu_i = \frac{\exp(a_1 x_i + a_2 {x_i}^2) S_i}{k_0 + k_1 S_i} + \varepsilon_i, \tag{4.6}$$

which is based on the Michaelis-Menten equation. Yet the nonlinear behaviour of model (4.6) should be examined. If the experimenters are to fit this rewritten model, the results of NLS could be a little discrepant and the Fisher information will also not be the same as that of model (4.5). Here, the upside is that nonlinear least squares estimation will be more accurate under model (4.6). With the same reference dataset from Martins et al. (1999), the new fitted model better satisfies the essential assumptions of errors, which is important in regression. This result can then contribute to our specification of the multivariate parameter prior distribution, for optimal design of a new experiment. As the prior means are more accurate and the prior variances are smaller, it will be easier to determine a desirable Gauss-Hermite sample of the model parameters.

On the other hand, the emphasis of the experiment and the optimality criterion should be on the precise estimation of the previous parameters a_0 and k in model (4.5). Given the common interests in the mechanism of the kinetics, it is perhaps not quite reasonable to reparametrise (4.5). Although model (4.6) can still be fitted to the observed response surface and is strong in prediction as well, it does not allow us directly to calculate the standard errors and the confidence intervals of the parameter estimators of model (4.5). In short, in this example, it is better not to reparametrise the previous kinetic model (4.5), the local D-criterion function ϕ_{D} of which is:

$$\phi_{\mathrm{D}} = \log \begin{vmatrix} \sum \mathbb{E}(\nu_i)^2 & \cdots & \cdots & \sum \frac{-\mathbb{E}(\nu_i)^2}{k+S_i} \\ \sum \mathbb{E}(\nu_i)^2 x_i & \sum \mathbb{E}(\nu_i)^2 {x_i}^2 & & \vdots \\ \sum \mathbb{E}(\nu_i)^2 {x_i}^2 & \sum \mathbb{E}(\nu_i)^2 {x_i}^3 & \sum \mathbb{E}(\nu_i)^2 {x_i}^4 & \vdots \\ \sum \frac{-\mathbb{E}(\nu_i)^2}{k+S_i} & \sum \frac{-\mathbb{E}(\nu_i)^2 x_i}{k+S_i} & \sum \frac{-\mathbb{E}(\nu_i)^2 {x_i}^2}{k+S_i} & \sum \frac{\mathbb{E}(\nu_i)^2}{(k+S_i)^2} \end{vmatrix},$$

where $\mathbb{E}(\nu_i)$ is the expectation of ν_i or model (4.5) in terms of the scaled variable x_i and the substrate concentration. Within the determinant function, the upper-left 3×3

submatrix shares some similarities with the Fisher information matrix of univariate second-order linear models. This implies that the information matrix $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ of the hybrid model (4.5) is not too complex in composition, due to its weak nonlinear behaviour.

The information matrix can also be considered as the sum of n unit Fisher information matrices, corresponding to the n independent experimental runs of $\mathbf{X}$. The empirical function $a_0 + a_1 x_i + a_2 {x_i}^2$ is linear. This hints that some small Gauss-Hermite subsamples of a_1 and a_2 could be adequate for an accurate approximation of φ. In contrast, the function ϕ_{D} in terms of k is more complex, since the parameter k is involved in the nonlinear Michaelis-Menten equation and also requires some careful biochemical interpretation. Perhaps we should sample more prior values of k to compose ϕ_{BD}.

4.4.2 Initial Gauss-Hermite Quadrature Results

The Gauss-Hermite sample size m also depends on the variance-covariance matrix $\mathbf{\Sigma}$ of the prior. In this example, compared with the mean 0.3122 of the univariate normal prior distribution, the variance of $\rho(k)$ is not as small as those of the other two univariate normal distributions $\rho(a_1)$ and $\rho(a_2)$. As such, there is a chance that we have to use a larger GH subsample size m_k in order to integrate out k of the multiple integral φ.

A numerical investigation will be helpful before we decide the best form of Gauss-Hermite quadrature for optimal design. If we set $m_k = m_{a_1} = m_{a_2} = 2$ whereas a_0 is fixed at $\tilde{a}_0 = -6.4086$, it is our expectation that the local D-criterion function ϕ_{D} can be well approximated with a third-order linear function in all the parameters. As the multiplication rule applies, the approximate value $\hat{\varphi} = \phi_{\mathrm{BD}}$ of a specific $\mathbf{X}$ should be compared with the true expected D-criterion value φ, which is unknown.

To find the best combination of the GH subsample sizes, first we can choose the $\mathbf{X}$ to be the approximate reference experimental design $\mathbf{X}_{\mathrm{ref}}$ in Martins et al. (1999). Moreover, an alternative is the $\mathbf{X}_0^*$ also listed in Table 4.1, which is the local D-optimal design. At first sight, $\mathbf{X}_0^*$ is neat in the allocation of the 30 experimental units. It has $n^* = 5$ support points with corresponding replicates. In comparison, the experimental units in $\mathbf{X}_{\mathrm{ref}}$ are dissimilar from those of the local optimal $\mathbf{X}_0^*$. While there are fewer replicates in total, $\mathbf{X}_{\mathrm{ref}}$ seems to be inefficient in the NLS estimation of model parameters. $\mathbf{X}_0^*$ determines most of its substrate concentration levels to be in the high range of the variable space $[0.15, 3]$. In this case, it is more sensible to assume $\rho(k)$ to be a normal distribution. We will come back to explain this later in this subsection.

With the GH subsamples of equal size $\mathfrak{m} = 1, 2, 3$ and 4 of the four parameters (even though $m_{a_0} > 1$ does not make sense), it is not hard to calculate the parameter prior values and the associated weights for each GH subsample. The consequent step is to derive all combinations of these subsampling units so as to form the overall Gauss-Hermite sample. As $m = 1, 8, 27, 64$, we can establish the respective functions ϕ_{BD}.

Table 4.1: The 30-Run Reference Design Against the Local D-Optimal Design

(a) $\mathbf{X}_{\text{ref}}$						(b) $\mathbf{X}_0^*$					
E	S	E	S	E	S	E	S	E	S	E	S
0.023	1.5	0.03	0.9	0.045	1.5	0.025	3	0.086	3	0.12	0.26
0.023	1.5	0.03	0.9	0.0675	1.5	0.025	3	0.086	3	0.12	0.26
0.023	1.5	0.03	1.5	0.0675	1.5	0.025	3	0.086	3	0.12	0.26
0.03	0.15	0.03	1.5	0.0675	1.5	0.025	3	0.086	3	0.12	3
0.03	0.15	0.03	1.5	0.089	1.5	0.025	3	0.086	3	0.12	3
0.03	0.15	0.03	3	0.089	1.5	0.025	3	0.097	0.33	0.12	3
0.03	0.3	0.03	3	0.089	1.5	0.025	3	0.12	0.26	0.12	3
0.03	0.3	0.03	3	0.112	1.5	0.025	3	0.12	0.26	0.12	3
0.03	0.3	0.045	1.5	0.112	1.5	0.086	3	0.12	0.26	0.12	3
0.03	0.9	0.045	1.5	0.112	1.5	0.086	3	0.12	0.26	0.12	3

A downside of the Gaussian quadrature rule is that, as the number of parameters p or the equal subsample sizes $\mathfrak{m}$ increases, we will see a geometric growth of the required total sample size m. The problem becomes serious if the nonlinear model assumes a quite vague parameter prior and thus amplifies the prior variances of $\rho(\boldsymbol{\theta})$. Though it is easier to make the subsample sizes equal, computation tends to be more intensive than required with the most appropriate Gauss-Hermite sample.

An economic and flexible method is to draw the individual Gauss-Hermite subsamples of different sizes. No matter whether we assume the independence between the parameter prior vectors, one can, for instance, draw $m_j \geq 1$ GH subsampling units of a specific parameter as well as one unit of each of the other parameters. This implies that each time our focus is on a separate numerical integration with respect to one parameter, when the total GH sample size is $m = m_j = 1, 2, 3,$ and 4. In this case, the evaluation of ϕ_{BD} tells us how much the numerical integration can improve when the subsample size increases. The numerical result then indicates a suitable Gauss-Hermite subsample size for the integration of φ with respect to the jth parameter, for $j = 1, 2, \ldots, p$.

With a selection of different GH samples, the approximate values are summarised in Table 4.2. For the last column of the table, m_j also represents an equal subsample size $\mathfrak{m}$ for all the parameters such that $m = \mathfrak{m}^p$. Overall, the values of ϕ_{BD}, the GH approximation of φ, should converge towards the true expected D-criterion values as the subsample size m_j of the jth parameter increases. This trend to converge is clearer when we use $\mathbf{X}_0^*$ as the experimental design $\mathbf{X}$, which does not use the minimal substrate concentration level 0.15.

To be more specific, when k is the sole parameter of which we draw the prior more than once, for $m = m_k = 2, 3, 4$, the approximate value ϕ_{BD} is close to the one shown in the last column for which the sample size is $m = m_j^3 = 8, 27, 64$. Hence, for Bayesian D-optimal design with the Gauss-Hermite approximation of φ, it is not essential to draw more than two prior values for either a_1 or a_2. Even if we do, it will not improve the

Table 4.2: A Selection of Gauss-Hermite Quadratures in Different Orders for the Approximation of the Expected D-Criterion Value

	Sample	$\{\mathrm{k}\}$	$\{\mathrm{a}_1\}$	$\{\mathrm{a}_2\}$	$\{\mathrm{k}, (\mathrm{a}_0), \mathrm{a}_1, \mathrm{a}_2\}$
$\mathbf{X}_{\text{ref}}$	m_j=1	-48.2255	-48.2255	-48.2255	-48.2255
	m_j=2	-47.8485	-48.2247	-48.2243	-47.8464
	m_j=3	-47.7491	-48.2247	-48.2243	-47.7470
	m_j=4	-47.6082	-48.2247	-48.2243	-47.6061
$\mathbf{X}_0^*$	m_j=1	-43.0242	-43.0242	-43.0242	-43.0242
	m_j=2	-42.7803	-43.0240	-43.0232	-42.7792
	m_j=3	-42.7462	-43.0240	-43.0232	-42.7451
	m_j=4	-42.7321	-43.0240	-43.0232	-42.7310

overall approximation much. After all, it is rare for a small error to affect the computer search and selection for the most efficient $\mathbf{X}$ out of several candidates.

The simplest treatment is to use the fixed substitutions $a_0 = \tilde{a}_0, a_1 = \tilde{a}_1$ and $a_2 = \tilde{a}_2$ for the approximation of the expected D-criterion function we assume. We can even replace the multivariate normal density $\rho(\boldsymbol{\theta})$ or $\rho(k, a_0, a_1, a_2)$ with the univariate density $\rho(k)$ and we shall be able to find a similar optimal design. Since there are no nonzero covariances in $\boldsymbol{\Sigma}$, the Gauss-Hermite quadrature we choose involves just a unidimensional numerical integration step. As the total GH sample size will be small, in fact we can afford to draw more prior values of a_0 and a_1.

However, our use of the expected criterion function has a drawback due to the specified $\rho(k)$. Our assumption is that $\rho(k)$ is a univariate normal distribution $\mathbb{N}(0.3122, 0.1868^2)$, which indicates a vague parameter prior. It is argued that with a large standard deviation to accommodate, the quantiles in the left tail of $\rho(k)$ fall below zero. Meanwhile, in the Michaelis-Menten mechanism, k is constrained to be positive. Hence, the assumed parameter prior density $\rho(k)$ violates the kinetic mechanism. When the substrate concentration is at the lowest level 0.15 mM, in this example, the model function could even take negative values, when the initial rate ν should at least be nonnegative. This is a contradiction which can often arise if we assume a mechanistic model.

Under the Gauss-Hermite quadrature in Table 4.2, when the subsample size is $m_k > 3$ for the parameter k, at least one prior value is less than zero in the GH subsample. In this case, even if the prior value is positive but close to zero, it would be similar to assuming the substrate concentration to be nearly independent of the expected initial rate. All these scenarios are unacceptable if we also require the Michaelis-Menten mechanism to hold behind the biochemical reaction in the experiment. A direct consequence is that the approximating function ϕ_{BD} will not converge well towards a stable solution, which is supposed to be the true criterion value. In particular, our reference $\mathbf{X}_{\text{ref}}$ in Table 4.1 specifies three substrate concentration levels to the minimum 0.15, which made the situation worse.

The above is our main concern as the variance of the univariate normal density $\rho(k)$ is 10 times the old variance estimate we find from the reference data. If we adhere to the normal distribution, a feasible measure is to halve the prior variance of k. As the GH subsample sizes of the remaining parameters are all fixed to one (i.e. the prior mean would be used if the prior distribution is normal), under the GH quadrature, we shall draw $m = m_k$ prior values of k. The total sample size is still small, so it is comfortable to draw more parameter prior values to extrapolate the true level of φ. The mutual comparisons of the results are in Table 4.3, where $\Sigma_{11} = 0.1868^2$ represents the variance of $\rho(k)$ for our previous computation to obtain Table 4.2. As we can see, with the increase in the subsample size m_k, the Gauss-Hermite quadrature becomes a more reliable approximation of φ.

Table 4.3: Gauss-Hermite Quadratures in Different Orders for the Approximation of the Expected D-Criterion Value (Revised)

	Σ_{11}		$\frac{1}{2}\Sigma_{11}$	
{k}	$\mathbf{X}_{\text{ref}}$	$\mathbf{X}_0^*$	$\mathbf{X}_{\text{ref}}$	$\mathbf{X}_0^*$
m_k=1	-48.2255	-43.0242	-48.2255	-43.0242
m_k=2	-47.8485	-42.7803	-48.0451	-42.9056
m_k=3	-47.7491	-42.7463	-48.0276	-42.8985
m_k=4	-47.6082	-42.7321	-48.0225	-42.8975
m_k=5	-47.8062	-42.7145	-48.0198	-42.8973
m_k=6	-47.7458	-42.7365	-48.0162	-42.8972
m_k=7	-47.6476	-42.7404	-48.0236	-42.8971
m_k=8	-47.7892	-42.7263	-48.0221	-42.8970

Another measure is available if we do not wish to reduce the variance in the matrix $\mathbf{\Sigma}$. Instead of limiting ourselves to use a normal parameter prior, we can assume $\rho(k)$ to be a different distribution from common exponential families, which should suit the Michaelis-Menten mechanism in our example better. For instance, it is worthwhile considering the Gauss-Legendre quadrature instead of the Gauss-Hermite quadrature, which will allow us to define a uniform distribution of the parameter prior. The uniform distribution can define a lower bound for the parameter space of k and is therefore sensible in this respect. The gamma prior distribution is another option, which can coordinate with the Gauss-Laguerre quadrature. This can set the lower bound of the parameter space to zero, though the densities will not be as low as we shall expect at the positive parameter values near zero.

The above substitutes should work well in this example of Michaelis-Menten kinetics, but our preferred choice of $\rho(k)$ would be the univariate lognormal distribution. As such, it just requires a simple adaptation of the Gauss-Hermite quadrature rule, so it is not hard to implement that. More importantly, a lognormal assumption ensures the prior value of k to be positive and also the value of $\rho(k)$ to be low when $k \to 0$. Both features can benefit the approximation of φ and make Bayesian optimal design more

realistic. Besides, $\rho(a_0, a_1, a_2)$ can be specified as a multivariate normal distribution. If the controlled variable x is a fixed constant, as a whole, the empirical function $\exp(a_0 + a_1 x + a_2 x^2)$ would have a lognormal distribution too. Hence, under the Michaelis-Menten mechanism, it is reasonable to assume a lognormal prior for k.

Later we shall elucidate the benefits of assuming a lognormal prior for k, as some numerical results can be compared with those in Table 4.2. Now let us come back to our initial parameter prior specification that $\rho(\boldsymbol{\theta})$ is a multivariate normal distribution, the variances of which are 10 times the NLS estimates we find from the reference data.

4.4.3 Pseudo-Monte Carlo Approximation Results

In comparison to the Gauss-Hermite quadrature, a suitable reference is the PMC approximation. In this subsection, we will draw 10 independent Monte Carlo samples of $\boldsymbol{\theta}$ and each random sample will contribute to an individual approximation of the expected D-criterion function φ. Therefore, we will see some fluctuation of the PMC approximate values near the desirable true level.

To check how accurate and dispersed the individual approximations are, we draw 10 Monte Carlo samples of size 10,000 of either the complete vector $\boldsymbol{\theta}$ or the most influential parameter k (while the remaining parameters are fixed at the respective normal prior means). As the number of draws is 10,000, so there are 100,000 different prior vectors in total, which we can use to extrapolate the true value of φ.

With each Monte Carlo sample, we can do a PMC approximation of the expected D-criterion function and it will be computed after we choose the reference design $\mathbf{X}_{\text{ref}}$ or the local D-optimal design $\mathbf{X}_0^*$. It is then straightforward to calculate the range of the 10 independent approximate values as shown in Table 4.4. This number should be as small as possible if the PMC approximation with the sample size 10,000 is expected to be accurate. The mean (or median) of the total 100,000 local D-criterion values can also be obtained and taken as the closest extrapolation of the true value of φ. It should be close to the approximate value of a suitable Gauss-Hermite quadrature too. Note that when we sample all the parameters as a whole, different prior values of a_0 would influence the PMC approximation as well as the consequent optimal design. It is also viable for someone to fix the prior value of a_0 at the mean $\tilde{a}_0$.

Instead of sampling all the univariate parameter prior distributions for PMC approximation, we can use the univariate Monte Carlo sample of the parameter k. A simplification like this reduces the randomness and uncertainties of the different samples whereas similar approximate values can be found after we complete all computation. In each case shown in Table 4.4, the difference between the minimum and the maximum of the 10 PMC approximate values is a bit too large. In other words, the current large sample size 10,000 is still insufficient for an accurate PMC approximation of φ. Partly, this is due

Table 4.4: Mean and Range of the PMC Approximations of the Expected D-Criterion Value

	Sample	Mean	Range of the 10
$\mathbf{X}_{\text{ref}}$	$\{\text{k}\}$	-47.7287	[-47.7814, -47.6849]
	$\{\text{k}, \text{a}_0, \text{a}_1, \text{a}_2\}$	-47.6991	[-47.7424, -47.6503]
$\mathbf{X}_0^*$	$\{\text{k}\}$	-42.7277	[-42.7723, -42.6869]
	$\{\text{k}, \text{a}_0, \text{a}_1, \text{a}_2\}$	-42.7108	[-42.7593, -42.6468]

to an issue with the normal prior distribution of k, which has been found to be an inappropriate prior assumption. In this case, even the mean of the 100,000 local D-criterion values is not an exact approximation of the expected D-criterion value. However, the value of φ (if it exists) should at least be close to -42.7108 (or -42.7277) for $\mathbf{X} = \mathbf{X}_0^*$. In reference to Tables 4.2-4.3, we can control the Gauss-Hermite approximation and the error will be small when the subsample size of k reaches three or four.

Our ultimate aim in approximating φ is to select the most efficient $\mathbf{X}$ in optimal design of the nonlinear experiment. Hence, even if the approximation is inaccurate sometimes (e.g. due to an inadequate parameter prior assumption), as long as the exchange algorithm used can find the true Bayesian optimal design in the end, the PMC or GH sample of the model parameters $\boldsymbol{\theta}$ would be acceptable. In this example, we cannot afford a huge PMC sample and the incurred high computational cost, so the better choice is to implement the Gauss-Hermite quadrature in our new exchange algorithm. In the next section, we will continue to assume a multivariate normal prior distribution for $\boldsymbol{\theta}$.

4.5 Initial Optimal Design Results

4.5.1 A Multistage Exchange Algorithm

As has been shown in Chapter 2, a new *hybrid exchange algorithm* is developed to obtain and ensure reliable solutions as the optimal designs of experiments. At the **first stage** of this efficient hybrid method we propose, a small number of candidate levels should be defined and allocated to each controlled variable. In this kinetics example, it is intuitive to choose the three levels $\{\pm 1, 0\}$ for the scaled enzyme concentration x. In reference to Smith (1918), these candidate coordinates should be recommended in continuous D-optimal design for univariate linear models of second order. Likewise, the subset $\{0.15, 1.5, 3\}$ includes the maximum, the minimum, and a middle candidate coordinate for the unscaled substrate concentration S. The algorithm can then create a simple candidate set $\boldsymbol{\Omega}$ consisting of nine points, with which the traditional *discrete optimisation* can be performed under an exchange approach, for instance. Global optimal solutions can be found in updating $\mathbf{X}$, while the discrete optimisation takes all the nine candidate

runs (or points) into account. Moreover, we can know in advance from Table 4.1 that the substrate concentration level 1.5 is not a useful candidate coordinate.

With $\mathbf{\Omega}$ and a specified number of tries $\tau = 30$, we first construct 30 random initial experimental designs $\mathbf{X}$. As the discrete optimisation (with the point exchange approach) should be applied, we can obtain $\tau^* \leq \tau$ distinct $\mathbf{X}$ at the local maxima of ϕ_{BD} as a Gauss-Hermite approximation to the expected D-criterion value φ. All these converged $\mathbf{X}$ are denoted as $\mathbf{X}_{1\mathrm{st}}$, which are thought to be *quasi-optimal* (close to optimal). Note that a strict *exchanging rule* should be adopted for the discrete optimisation if we wish to obtain more distinct intermediate solutions $\mathbf{X}$.

Before we finish the first stage of computation, the aim is to shape the rough optimal structure of $\mathbf{X} = \mathbf{X}_{1\mathrm{st}}$. Compared with a random $\mathbf{X}$ as the initial start in the algorithm, these $\mathbf{X}_{1\mathrm{st}}$ must be far more efficient and adopt more sensible coordinate levels. With each $\mathbf{X}_{1\mathrm{st}}$ to act as a distinct new start, we can now introduce the new *continuous optimisation* method to this multistage exchange algorithm. As it extends the search domain from $\mathbf{\Omega}$ to the rectangular variable space $\mathcal{X} = [-1, 1] \times [0.15, 3]$, the continuous optimisation is a much more effective method but is slower in computation.

At the **second stage** of this hybrid exchange algorithm, we can abandon the list of the defined discrete candidates now. Here, the point exchange approach should be adapted to incorporate, for instance, the Nelder-Mead method to optimise each point in the current $\mathbf{X}_{1\mathrm{st}}$. In total, there are n steps to update the whole $\mathbf{X}_{1\mathrm{st}}$ and thus complete one *iterative update*. There are some relevant instructions. On the one hand, a transformation of the controlled variables can impose a box constraint on the optimisation, which should define the space $\mathcal{X}$ of experimental interest. On the other hand, no matter which optimisation option we choose, all the continuous optimisation solutions are local optimal, since there is no limit on the number of candidate points to be evaluated. To select the best local optimal solution, multiple sets of initial values can be specified, which are also called the "initial points for the continuous optimisation".

Under a reasonable exchanging rule, the continuous optimisation of respective experimental runs is a tedious process in several iterative updates. This will not be completed until we achieve a local maximum of φ or its Gauss-Hermite approximation ϕ_{BD}, in terms of $\mathbf{X}$. When no more important updates or point exchanges can be found to improve the $\mathbf{X}_{1\mathrm{st}}$ from the current iterative update, the solution is treated as being converged and can be saved as a $\mathbf{X}_{2\mathrm{nd}}$. Overall, there are also τ^* solutions $\mathbf{X}_{2\mathrm{nd}}$, of which we denote the most efficient one (the one with the maximal ϕ_{BD}) as $\mathbf{X}_{\mathrm{B}}$.

All in all, we should implement the continuous optimisation after the first stage for the discrete optimisation. In this regard, since the continuous optimisation starts from a distinct $\mathbf{X}_{1\mathrm{st}}$ rather than a random initial design $\mathbf{X}$, it will be much easier to maximise the expected D-criterion function and obtain some efficient $\mathbf{X}_{2\mathrm{nd}}$. Compared with the optimised exchange algorithm (see Chapter 2 for our introduction) which will optimise τ

random $\mathbf{X}$ without passing through a stage for the simple discrete optimisation, the hybrid exchange algorithm will require fewer tries, since it is more effective in computation and more focused on the iterative search for optimal candidate points.

At the **third stage** of the hybrid method, we shall use the term *closest distance* $\mathfrak{d}$ to represent the minimum space between two feasible and adjacent coordinates, both values of which therefore must be distinguishable in a real experiment. In our example, we choose the closest distance to be 0.001 mg for the protein weight E and 0.01 mM for the (initial) substrate concentration S. As we can now discretise the continuous variable space $\mathcal{X}$, it is viable to construct suitable subsets (rather than complete sets) of candidate coordinates for both controlled variables.

Because of the use of the continuous optimisation, the total number of exact replicates is often small in the current best solution $\mathbf{X}_{\mathrm{B}}$. However, we can perhaps find some *quasi-replicates* there, the locations of which are quite close to each other in $\mathcal{X}$. Hence, the idea is to convert each cluster of quasi-replicates to exact replicates on a feasible support point. To find the support point that maximises the criterion function, a small candidate coordinate subset can be constructed for each cluster of quasi-replicate runs. The upper and lower limits of this subset depend on the maximal and minimal coordinates out of the quasi-replicate runs, so the number of discrete candidates would be small if we have chosen a reasonable exchanging rule for the continuous optimisation. After the conversion of the quasi-replicates, we should also recheck the numbers of exact replicates of each feasible support point. This requires a simple discrete optimisation as we can recalculate the numbers of replicates. This is the end of the hybrid method, whereas the result of the discrete optimisation is the final Bayesian D-optimal design $\mathbf{X}_{\mathrm{B}}^{*}$. As our final output of the algorithm, it should best meet the experimental requirements.

In this version of the hybrid exchange algorithm, though we adopt the classical point exchange approach (Fedorov, 1972) for the iterative updates of $\mathbf{X}$, a similar three-stage algorithm is compatible with the coordinate exchange approach (Meyer and Nachtsheim, 1995) too. In this example we discuss, the number of controlled variables is $v = 2$ so that the results of the two exchange approaches will not be much different. Our first demonstration in this chapter will follow the point exchange approach throughout, which we expect to be more efficient at least in the initial discrete optimisation.

When we approximate the expected D-criterion function φ, it is useless to draw a Gauss-Hermite subsample of a_0 if its prior is independent of the other parameter priors. Nevertheless, a_0 is not a nuisance parameter in the model and should not be removed from the Fisher information matrix in φ. As we see in the local D-optimal design $\mathbf{X}_0^{*}$, under the new expected criterion, we believe the variation in the prior value of k will at least make an impact on the adopted substrate concentration level 0.26 in Table 4.1b. In contrast to k, since a_1 and a_2 are the parameters in the empirical exponential function, we expect some linear behaviours of model (4.6). Under the current prior variances, the

variations in the prior values of these two parameter shall also have certain impacts on the expected D-criterion and the adopted coordinate levels in $\mathbf{X}_{\mathrm{B}}^{*}$. Overall, although the hybrid exchange algorithm aims to find a $\mathbf{X}$ which maximise φ instead of the local D-criterion function ϕ_{D}, we can expect $\mathbf{X}_{0}^{*}$ to be still efficient and resemble $\mathbf{X}_{\mathrm{B}}^{*}$. These speculations will be examined in the next section.

4.5.2 Mutual Comparisons and Robustness of Optimal Designs

Our first assumption of the parameter prior density $\rho(\boldsymbol{\theta})$ is the joint p.d.f. of four independent normal densities. Remind that the normal prior distribution of a_0 is of little importance to Bayesian optimal design. We choose the total Gauss-Hermite sample size to be $m = 2^3 = 8$ as $m_k = m_{a_1} = m_{a_2} = \mathfrak{m} = 2$. The purpose is to inspect the influence of each local D-criterion function in ϕ_{BD}, which is associated with one unique parameter prior vector $\boldsymbol{\theta}^r$, for $r = 1, 2, \ldots, 8$. In this circumstance, it is simple to compose the Gauss-Hermite quadrature to approximate φ. Meanwhile, the GH composite weights are all equal since there are fewer than three units in each of the independent GH subsamples for the individual parameters. In short, the function ϕ_{BD} will be the arithmetic mean of eight local D-criterion values $\phi_1, \phi_2, \ldots, \phi_8$.

To examine the relative efficiencies and the robustness of $\mathbf{X}_{\mathrm{B}}^{*}$, the local D-criterion can act as the reference. This means that each time we choose one of the local criterion functions $\phi_1, \phi_2, \ldots, \phi_8$ and thus assume a scenario with a different parameter prior vector. Therefore, the hybrid exchange algorithm can also be applied to find a X_{r}^{*} which is close to local D-optimal. This is similar to what we did in Chapters 2 and 3. As a result, we will output the final results $\mathrm{X}_{1}^{*}, \mathrm{X}_{2}^{*}, \ldots, \mathrm{X}_{8}^{*}$ under eight scenarios, in addition to the real local D-optimal design $\mathbf{X}_{0}^{*}$.

As such, we envisage $m + 1 = 9$ artificial scenarios from $\mathfrak{S}_0$ to $\mathfrak{S}_8$, which assume different true values for the unknown parameters $\boldsymbol{\theta}$. These true parameter values are listed in Table 4.5 and are set to equal $\boldsymbol{\theta}^r$, for $r = 0, 1, \ldots, 8$. Hence, we use the prior mean $\tilde{\boldsymbol{\theta}}$ under $\mathfrak{S}_0$ and the units of the Gauss-Hermite sample under the corresponding $\mathfrak{S}_1, \mathfrak{S}_2, \ldots, \mathfrak{S}_8$. For instance, under the first scenario $\mathfrak{S}_0$, the local D-criterion function should be written as $\phi_{\mathrm{D}}(\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}})$. As the information matrix is found, we shall be able to find the solution $\mathbf{X}_{0}^{*}$ in Table 4.1b, which is known to be local D-optimal under the current postulated scenario.

Our real aim is to search for the Bayesian D-optimal design $\mathbf{X}_{\mathrm{B}}^{*}$ as the tradeoff among the scenarios $\mathfrak{S}_1$ to $\mathfrak{S}_8$, which should tend to maximise φ or its Gauss-Hermite approximation ϕ_{BD}. With respect to the previous experimental designs $\mathrm{X}_{0}^{*}, \mathrm{X}_{1}^{*}, \ldots, \mathrm{X}_{8}^{*}$, the relative efficiencies *eff* of $\mathbf{X}_{\mathrm{B}}^{*}$ can be calculated under each scenario described in Table

Table 4.5: Assumed True Parameter Values (Prior Values) under Each of the Nine Postulated Scenarios

Scenario	k	a_0	a_1	a_2
$\mathfrak{S}_0$	0.3122	-6.4086	0.8383	-0.2861
$\mathfrak{S}_1$	0.1254	-6.4086	0.7828	-0.3902
$\mathfrak{S}_2$	0.1254	-6.4086	0.8937	-0.3902
$\mathfrak{S}_3$	0.1254	-6.4086	0.7828	-0.1821
$\mathfrak{S}_4$	0.1254	-6.4086	0.8937	-0.1821
$\mathfrak{S}_5$	0.4990	-6.4086	0.7828	-0.3902
$\mathfrak{S}_6$	0.4990	-6.4086	0.8937	-0.3902
$\mathfrak{S}_7$	0.4990	-6.4086	0.7828	-0.1821
$\mathfrak{S}_8$	0.4990	-6.4086	0.8937	-0.1821

4.5. As such, for $r = 0, 1, \ldots, 8$,

$$\mathit{eff}(\mathbf{X}_{\mathrm{B}}^*, \boldsymbol{\theta}^r) = \frac{\exp(\phi_{\mathrm{D}}(\mathbf{X}_{\mathrm{B}}^*, \boldsymbol{\theta}^r))^{\frac{1}{4}}}{\exp(\phi_{\mathrm{D}}(\mathbf{X}_r^*, \boldsymbol{\theta}^r))^{\frac{1}{4}}} 100\%. \tag{4.7}$$

The same formulae can also be used to calculate the relative efficiencies of $\mathrm{X}_0^*, \mathrm{X}_1^*, \ldots, \mathrm{X}_8^*$ with respect to the others. We can then check their robustness to the parameter prior.

In our demonstration of the hybrid exchange algorithm with either the local or the expected D-criterion, the number of tries is fixed to be $\tau = 30$ no matter if that involves the Gauss-Hermite sample of size $m = 8$. A strict exchanging rule $\max(d_i) > 1.1$ is chosen for the discrete optimisation of random initial designs $\mathbf{X}$. It will make τ^*, the number of distinct $\mathbf{X}_{1\mathrm{st}}$, close to 30 so that we can do more continuous optimisation. In contrast, the rule for the continuous optimisation is as loose as $\max(d_i) > 1.00000001$, in order to facilitate the search for $\mathbf{X}_{2\mathrm{nd}}$ in different tries and $\mathbf{X}_{\mathrm{opt}}$.

After the conversions of the quasi-replicates and the reallocation of the numbers of replicates, the final solutions of the hybrid method are taken to be $\mathrm{X}_0^*, \mathrm{X}_1^*, \ldots, \mathrm{X}_8^*, \mathrm{X}_{\mathrm{B}}^*$, when there are different parameter prior assumptions. To obtain the most reliable results, some of them can be a little improved in a *modified adjustment algorithm*. In the next section, we will revisit this useful tool which was first introduced in Chapter 3.

To make mutual comparisons of these results under a scenario $\mathfrak{S}_r$, for $r = 0, 1, \ldots, 8$, the fixed baseline is set to be the current $\mathbf{X}_r^*$ which is local D-optimal if $\boldsymbol{\theta} = \boldsymbol{\theta}^r$. As such, Table 4.6 summarises the calculated relative efficiencies of $\mathrm{X}_0^*, \mathrm{X}_1^*, \ldots, \mathrm{X}_8^*, \mathrm{X}_{\mathrm{B}}^*$, all relative to the baseline $\mathbf{X}_r^*$ under the scenario $\mathfrak{S}_r$. We can do similar computation under each scenario described in Table 4.5. The relative efficiencies of the baselines must be 100% under their respective scenarios, and these numbers are presented in bold font.

In this case, ϕ_{BD} is the arithmetic mean of the eight local criterion functions. It can also be expressed as the logarithm of the geometric mean of the determinants of the eight respective information matrices $\mathbf{F}(\mathbf{X}, \boldsymbol{\theta}^r)^{\mathrm{T}}\mathbf{F}(\mathbf{X}, \boldsymbol{\theta}^r)$. Therefore, for $r = 0, 1, \ldots, 8$,

Table 4.6: Relative Efficiencies of Experimental Designs under Each of the Nine Postulated Scenarios

Design	Postulated Scenario								
	k ≈ 0.3122	k ≈ 0.1254				k ≈ 0.4990			
	$\mathfrak{S}_0$	$\mathfrak{S}_1$	$\mathfrak{S}_2$	$\mathfrak{S}_3$	$\mathfrak{S}_4$	$\mathfrak{S}_5$	$\mathfrak{S}_6$	$\mathfrak{S}_7$	$\mathfrak{S}_8$
$\mathbf{X}_0^*$	**100**	91.57	91.28	91.74	92.14	97.32	96.85	96.86	97.20
$\mathbf{X}_1^*$	93.75	**100**	99.52	97.15	97.53	87.18	86.64	84.24	84.51
$\mathbf{X}_2^*$	94.02	99.56	**100**	96.45	97.68	87.19	87.50	84.09	85.13
$\mathbf{X}_3^*$	95.45	97.46	96.66	**100**	99.87	86.89	86.12	88.75	88.64
$\mathbf{X}_4^*$	95.34	97.11	96.53	99.91	**100**	86.58	86.00	88.67	88.75
$\mathbf{X}_5^*$	97.42	83.38	83.20	81.67	82.17	**100**	99.59	97.30	97.79
$\mathbf{X}_6^*$	96.96	82.88	83.37	80.53	81.66	99.59	**100**	96.12	97.38
$\mathbf{X}_7^*$	97.57	81.74	81.07	83.87	83.77	97.91	97.04	**100**	99.87
$\mathbf{X}_8^*$	97.45	81.45	80.96	83.80	83.87	97.56	96.90	99.91	**100**
$\mathbf{X}_B^*$	99.33	95.17	94.87	95.37	95.79	94.67	94.16	94.13	94.42

the relative efficiencies of $\mathbf{X}_B^*$ must be lower than 100% under all the scenarios listed in Table 4.5. Nevertheless, more emphasis of the criterion function φ is on the four scenarios when $\mathbf{X}_B^*$ could lead to a smaller determinant of its Fisher information matrix. This is found to be the case when we assume $k = 0.4990$.

The expected D-criterion tends to compensate the loss from the four less efficient cases, since its solution $\mathbf{X}_B^*$ must be an optimal tradeoff among $\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}_8^*$, regarding the support points and the corresponding replication. As a consequence in Table 4.6, the fluctuation of the relative efficiencies of $\mathbf{X}_B^*$ is most stable across the postulated scenarios. Under the scenario $\mathfrak{S}_0$ when $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, its local D-criterion value is also much higher than those of the other experimental designs. This is because of the similarities between the $\mathbf{X}_B^*$ and $\mathbf{X}_0^*$ we have obtained.

In comparison to the local D-optimal design $\mathbf{X}_0^*$, $\mathbf{X}_B^*$ is more efficient if the current scenario assumes $k \approx 0.1254$ but it is less efficient otherwise. As such, the independent normal parameter prior density $\rho(k)$ not only influences the model nonlinearities but also the expected D-criterion that maximises the function φ. Overall, there will be an improvement if we favour $\mathbf{X}_B^*$ over $\mathbf{X}_0^*$, as the variances and covariances of the parameter estimators will be smaller.

While we assume model (4.5), the relative improvement is slight. This is because, the efficiencies of the local optimal design $\mathbf{X}_0^*$ is very robust to the other three parameter values (excluding k). Hence, there is not much room to improve it further, even if we can still consider the uncertainties of the true parameter values. The result can also be observed in Table (4.6): when the same value of k is assumed as a parameter prior value for different local D-criterion functions, all the relevant efficiencies are found to be higher than 96%. Under the current expected criterion, $\rho(a_0), \rho(a_1), \rho(a_2)$ do not make much difference to improve the efficiencies of $\mathbf{X}_B^*$.

This means that when we choose $\mathbf{X}_{\mathrm{B}}^{*}$ instead of $\mathbf{X}_{0}^{*}$, the main improvement will be in the reduction of the standard error of the NLS estimate of k. We can obtain some ideas about the variance-covariance matrix of the new estimator $\hat{\boldsymbol{\theta}}$, when the assumption is that the multivariate parameter prior distribution has no nonzero covariances. As the Gauss-Hermite approximation ϕ_{BD} defines the total sample size $m = 8$, we can write

$$\mathfrak{M}_1 = \prod_{r=1}^{m} (\mathbf{F}(\mathbf{X}, \boldsymbol{\theta}^r)^{\mathrm{T}} \mathbf{F}(\mathbf{X}, \boldsymbol{\theta}^r))^{-\omega^{(r)}}; \quad \mathfrak{M}_2 = \sum_{r=1}^{m} \omega^{(r)} (\mathbf{F}(\mathbf{X}, \boldsymbol{\theta}^r)^{\mathrm{T}} \mathbf{F}(\mathbf{X}, \boldsymbol{\theta}^r))^{-1}.$$

The two matrices above can be interpreted. $\mathfrak{M}_1$ is the *elementwise product* of the weighted inverses of the information matrices whereas $\mathfrak{M}_2$ is the weighted sum of these inverse matrices. The Fisher information is calculated for each of the sampled parameter prior vectors $\boldsymbol{\theta}^r$ and we shall also obtain the respective composite weights $\omega^{(r)}$. As an alternative, if one takes the matrix product instead of the elementwise product, there is $-\log|\mathfrak{M}_1| = \phi_{\mathrm{BD}}$. Both $\mathfrak{M}_1$ and $\mathfrak{M}_2$ reflect some prior knowledge about the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, when the variance of the errors σ^2 is assumed to be an unknown constant. When $\mathbf{X} = \mathbf{X}_{\mathrm{B}}^{*}$, $\log|\mathfrak{M}_1| = 43.2328$ and $\log|\mathfrak{M}_2| = 43.8144$ as

$$\mathfrak{M}_1 = 10^4 \begin{pmatrix} 2.9861 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 3.3845 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 7.4340 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 7.9420 \end{pmatrix}; \mathfrak{M}_2 = 10^5 \begin{pmatrix} 0.8476 & 0.2420 & 0.0708 & 0.0644 \\ 0.2420 & 0.6098 & -0.2771 & -0.2663 \\ 0.0708 & -0.2771 & 1.3873 & -0.9897 \\ 0.0644 & -0.2663 & -0.9897 & 1.4348 \end{pmatrix}.$$

In contrast, when $\mathbf{X} = \mathbf{X}_{0}^{*}$, $\log|\mathfrak{M}_1| = 43.2658$ and $\log(|\mathfrak{M}_2|) = 43.7183$ as

$$\mathfrak{M}_1 = 10^4 \begin{pmatrix} 3.1441 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 3.3720 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 7.3857 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 7.8761 \end{pmatrix}; \mathfrak{M}_2 = 10^5 \begin{pmatrix} 0.7976 & 0.2239 & 0.0790 & 0.0758 \\ 0.2239 & 0.6041 & -0.2751 & -0.2627 \\ 0.0790 & -0.2751 & 1.3871 & -0.9887 \\ 0.0758 & -0.2627 & -0.9887 & 1.4324 \end{pmatrix}.$$

We can look at $\mathfrak{M}_1$ first. It appears that the variance of the NLS estimator $\hat{k}$ will be much smaller if we use $\mathbf{X}_{\mathrm{B}}^{*}$ instead of $\mathbf{X}_{0}^{*}$ for the new experiment. In contrast, that also increases the variances of the other parameter estimators. Overall, the statistic of $\mathbf{X}_{\mathrm{B}}^{*}$ is $\log|\mathfrak{M}_1| = 43.2328$, which is lower than that of $\mathbf{X}_{0}^{*}$. This is the result we would like to see, as $\mathbf{X}_{\mathrm{B}}^{*}$ is thought to be the better choice under the assumption of the multivariate normal prior density $\rho(\boldsymbol{\theta})$. Hence, $\mathfrak{M}_1$ is a good measure of the information coming from $\mathbf{X}$, a fixed experimental design, though the matrix is approximately diagonal.

However, $\mathfrak{M}_2$ does not look sensible as it takes the weighted sum of those inverse matrices. In that case, $\mathbf{X}_{0}^{*}$ is shown to be better than $\mathbf{X}_{\mathrm{B}}^{*}$ in the precise estimation of the model parameters. As the statistic of $\mathbf{X}_{0}^{*}$ is even lower, this result cannot be accepted. As soon as we link $\mathfrak{M}_1$ to the nonlinear model function and the efficiencies calculated in Table 4.6, it is not hard to see the influences of $\mathbf{X}_{\mathrm{B}}^{*}$ on the specific parameters.

As shown in Table 4.5, the standard deviation of $\rho(k)$ is larger than the others in $\rho(\boldsymbol{\theta})$, which leads a vaguer univariate prior distribution. This is one reason that we can expect

some improvement in the estimation of k, when the expected D-criterion replaces the local one. Another important reason is related to the assumed model and its Fisher information. While the other three parameters appear in the empirical exponential function of the hybrid nonlinear model (4.5), k is the one from the theoretical Michaelis-Menten equation. Hence, it is reasonable that the expected D-criterion will emphasise the precise estimation of k.

On the other hand, the parameter prior values of both a_1 and a_2 have minor impacts on the efficiencies, which almost reach 100% in some cases in Table 4.6. This implies that even if we assume those two univariate prior densities $\rho(a_1)$ and $\rho(a_2)$, that would not much affect the D-criterion and the overall precision in the parameter estimation. Instead, it is even plausible to assume the point prior values $a_0 = \tilde{a}_0, a_1 = \tilde{a}_1$ and $a_2 = \tilde{a}_2$. Because of that, with respect to $\mathbf{X}_0^*$, $\mathbf{X}_{\mathrm{B}}^*$ is 99.33% efficient under the first scenario $\mathfrak{S}_0$, which is a similar experimental design.

We can make further comparison of the experimental designs and look for more details. Figure 4.1 visualises the 30 runs of $\mathbf{X}_0^*$ as well as those of $\mathbf{X}_{\mathrm{B}}^*$. While the integer to the upper right of each bubble (or circular marker) represents the number of exact replicates, the bracket near each bubble indicates the coordinate levels of the support point in $\mathbf{X}$. A rectangle is also drawn, which represents the bounded variable space $\mathcal{X}$.

Under the scenario $\mathfrak{S}_0$, the local D-criterion value of $\mathbf{X}_0^*$ is -43.0242. As we can see, there are no conspicuous discrepancies between $\mathbf{X}_0^*$ and $\mathbf{X}_{\mathrm{B}}^*$. To be specific, three of the five unique experimental runs and the corresponding numbers of replicates are identical, while the substrate concentration is set to 3mM in these cases. Overall, the adopted coordinate levels of the protein weight are almost the same, whereas the substrate concentration levels are not. This is because of the normal parameter prior density $\rho(\boldsymbol{\theta})$ we assume. Although the pseudo-Bayesian approach does not improve much the D-optimal design for the kinetic model (4.5), if possible, one is still encouraged to use the more robust expected D-criterion.

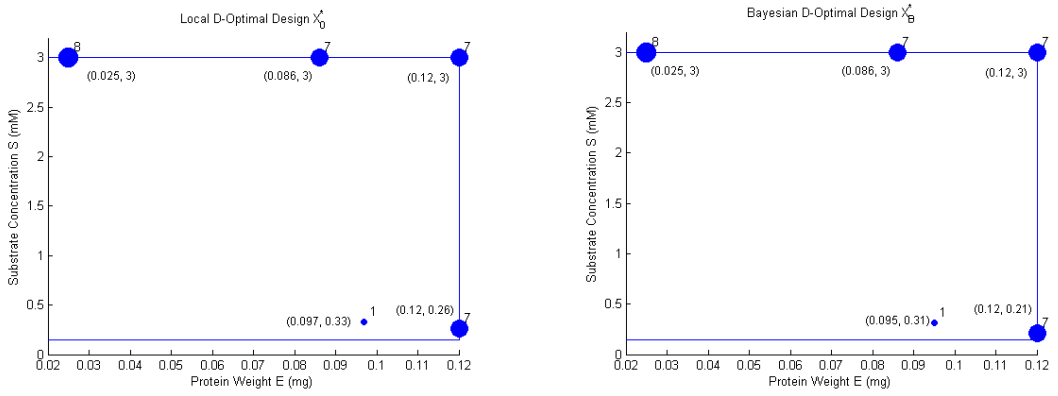


Figure 4.1: Local and Bayesian D-Optimal Designs of the Same Experiment

Figure 4.2 visualises the local optimal solutions $\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}_8^*$ under the other scenarios described in Table 4.5. We can first see the impact of the prior value of k, which is in relation to the exact location of the bottom right bubble. If we assumes $k \approx 0.1254$ under one scenario, for instance, the substrate concentration level shall be set to the minimum 0.15 for this bubble. As the parameter values of the Michaelis-Menten equation are different, the substrate concentration will have a low correlation with the initial rate in this case. Otherwise, if the current scenario assumes $k \approx 0.4990$, the substrate concentration level is 0.39 whereas the protein weight remains 0.12.

Compared with the others, the variance of the univariate normal prior distribution $\rho(a_1)$ is rather small under our assumption. Therefore, we cannot see much difference in the results when a_1 takes a different value. As we have mentioned, this is also because of the linear behaviour of model (4.5).

Relatively, the variation in the parameter value a_2 could take a more important role, though its impact is also weak on the relative efficiencies of $\mathbf{X}_0^*$ and $\mathbf{X}_\mathrm{B}^*$ under various scenarios. When $a_2 \approx -0.3902$, we can find five unique experimental runs in each case, one of which will take two or three replicates. In contrast, when $a_2 \approx -0.1821$, the model function is closer to linear and the number of support points n^* is reduced to four. As such, we can see almost the same numbers of replicates because the number of parameters p is equal to four too. Hence, perhaps there is more than one local optimal solution under these scenarios. Moreover, with a new value of a_2, the linear behaviour of the model is then different. It can lead to a modification of the current optimal location of the top left bubble, along with the Fisher information matrix.

To summarise, it is clear that $\mathbf{X}_\mathrm{B}^*$ is a compromise among the various parameter prior vectors. It should therefore be adopted to retain the most robustness to the uncertainties of the true unknown parameter values. However, to use $\mathbf{X}_\mathrm{B}^*$ instead of $\mathbf{X}_0^*$, the premise is that our assumption of the multivariate parameter prior density $\rho(\boldsymbol{\theta})$ is correct. When the model is written as (4.5), this is unfortunately not the case since we assume an inappropriate normal prior for the parameter k.

4.6 A Combination of Normal and Lognormal Priors

An alternative to the common normal prior will be a lognormal one for the parameter k in the hybrid nonlinear model (4.5). In this section, we will discuss this change of the parameter prior distribution. The mean of this lognormal prior should be equal to the old NLS estimate $\tilde{k}$ whereas the variance is 10 times the old variance estimate. As such, the new prior distribution we assume is $k \sim \log\mathbb{N}(-1.3171, 0.5531^2)$, which is equivalent to $\log(k) \sim \mathbb{N}(-1.3171, 0.5531^2)$. Meanwhile, the remaining univariate parameter prior distributions are normal as before and independent of the lognormal prior of k.

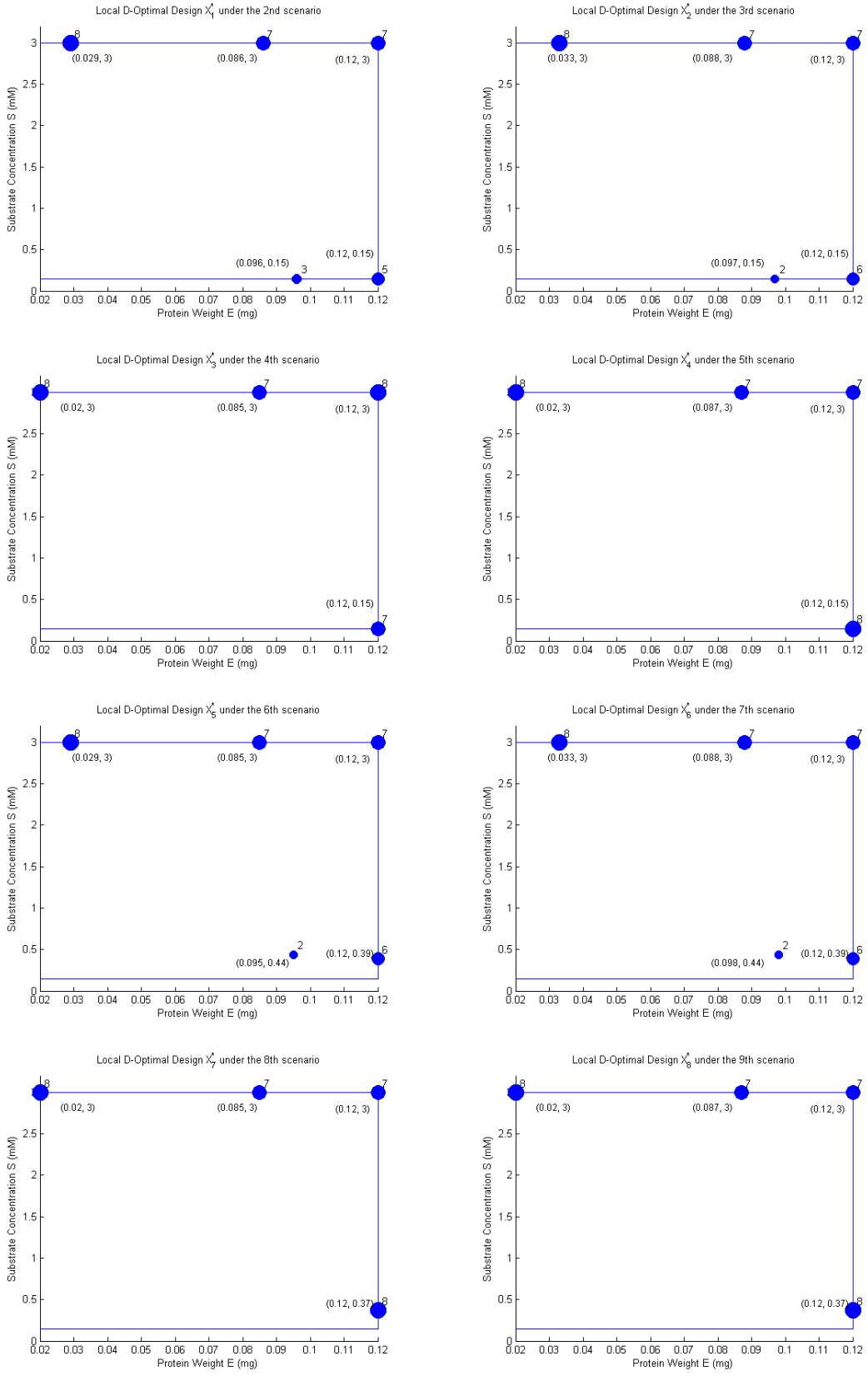


Figure 4.2: Local D-Optimal Designs under Various Postulated Scenarios

Our treatment for the lognormal prior is not the same. It demands a different application of the same Gauss-Hermite quadrature in Section 4.2. A simple exponential transformation of the lognormal prior values in the GH subsample is also required, which is consistent with the work in Goos and Mylona (2013), for instance. As long as the variance-covariance matrix of the multivariate parameter prior density $\rho(\boldsymbol{\theta})$ is diagonal, the constraint $k > 0$ can be well satisfied. Hence, it is more reasonable to replace the local D-criterion function ϕ_{D} with our assumed expected criterion function φ.

To approximate the multiple integral φ, it is time to determine the four subsample sizes for the Gauss-Hermite quadrature. A constructive approach is to inspect and compare the approximate values ϕ_{BD}, which are calculated with different GH samples in Table 4.7. As we can see, the lognormal parameter prior is a reasonable assumption. With the increase in each subsample size, the approximate value will converge to a stable level, which can be extrapolated to be the true value of φ. To ensure an accurate GH approximation, we choose $m_k = 4$ and $m_{a_1} = m_{a_2} = 2$ so that the total GH sample size is $m = 16$. As such, we will be rather close to find the true value of φ, at least in both evaluated cases (i.e. $\varphi \approx -47.9387$ for $\mathbf{X}_{\mathrm{ref}}$ and $\varphi \approx -42.8326$ for $\mathbf{X}_0^*$). Of course, an even smaller GH sample can also be acceptable sometimes.

Table 4.7: Gauss-Hermite Quadratures in Different Orders for the Approximation of the Expected D-Criterion Value (Lognormal Prior)

Sample	$\{\mathrm{k}\}$		$\{\mathrm{k}, (\mathrm{a}_0), \mathrm{a}_1, \mathrm{a}_2\}$	
	$\mathbf{X}_{\mathrm{ref}}$	$\mathbf{X}_0^*$	$\mathbf{X}_{\mathrm{ref}}$	$\mathbf{X}_0^*$
m_j=1	-47.7145	-42.5920	-47.7145	-42.5920
m_j=2	-47.9426	-42.8347	-47.9405	-42.8336
m_j=3	-47.9404	-42.8337	-47.9383	-42.8326
m_j=4	-47.9409	-42.8338	-47.9388	-42.8327
m_j=5	-47.9409	-42.8337	-47.9387	-42.8326
m_j=6	-47.9409	-42.8338	-47.9387	-42.8326
m_j=7	-47.9409	-42.8338	-47.9387	-42.8326
m_j=8	-47.9409	-42.8338	-47.9387	-42.8326

As a reference, we also draw 10 uncorrelated Monte Carlo samples of $\boldsymbol{\theta}$ of size 10,000 each. Each sample then leads to an independent PMC approximation of φ. The mean of the 10 approximate values is -47.9263 for $\mathbf{X}_{\mathrm{ref}}$, whereas the minimum and maximum are -47.9652 and -47.8916. The arithmetic mean of the 100,000 local D-criterion values is the closest PMC approximation to the true φ. The difference between the maximum and minimum is reduced but is still nowhere near zero, compared with the previous result in Table 4.4. Overall, it seems that the Gauss-Hermite quadrature is the more reliable and accurate approximation, which also takes a rather small sample size. Likewise, the mean of the 100,000 local D-criterion values is -42.8315 for $\mathbf{X}_0^*$, but this PMC approximation of the huge size 100,000 is not too accurate when the true expected D-criterion value extrapolated from the Gauss-Hermite quadrature is close to -42.8326. The minimum

and maximum of the 10 approximate values are -42.8527 and -42.7879, which could indicate the uncertainty of a PMC approximation of size 10,000.

As an alternative, we draw a univariate Monte Carlo sample of k, when the other parameter prior values are fixed at their respective means in $\tilde{\boldsymbol{\theta}}$. With the sampled parameter prior vectors in this case, the overall mean of the 100,000 local D-criterion values is -47.9483 for $\mathbf{X}_{\text{ref}}$, whereas the minimum and maximum of the 10 PMC approximations are -47.9734 and -47.9212. Also for $\mathbf{X}_0^*$, the calculated mean is -42.8324, with the minimum and maximum being -42.8532 and -42.7953.

There are some nice results when we assume the lognormal prior distribution for k. Then it is worthwhile using the same hybrid exchange algorithm, which is subject to the new assumption and the revised form of the joint p.d.f. $\rho(\boldsymbol{\theta})$. Moreover, the result will be taken for the modified adjustment algorithm (the one we also used in the last section), which is developed on the basis of the work in Donev and Atkinson (1988). It is included to make a minor improvement in the result of the multistage hybrid method, which is in line with the expected criterion function we assume. When the problems are simple, this modified adjustment algorithm can ensure the final solution $\mathbf{X}_{\text{B}}^{**}$ to be optimal, when the criterion function will achieve the true maximum. Of course, this is not essential in optimal design of experiments. If the third stage of the hybrid method does not make its then solution $\mathbf{X}_{\text{B}}^*$ much different from the best continuous optimisation solution $\mathbf{X}_{\text{opt}}$, we might not need the modified adjustment algorithm for further improvement.

In reference to the previous measure $\mathbf{X}_{\text{B}}^*$ of n rows, first we must fix the number of unique experimental runs n^* as well as the respective numbers of replicates. The premise to implement the modified adjustment algorithm is that these fixed numbers are assumed to be the ones in the truly optimal $\mathbf{X}$ which will maximise the expected D-criterion function. Under this condition, we work on the $n^* \times v$ exact coordinate levels of the support points, so the required computation is much simplified. In the brief description below, we revisit the modified adjustment algorithm from Chapter 3. As usual, it uses the coordinate exchange approach but is written in a different version, since we would like to update all relevant replicate runs in one step:

The Modified Adjustment Algorithm After the Previous Exchange Algorithm

4.1 For $k = 1, 2, \ldots, v$, determine the closest distance $\mathfrak{d}_k$ for the kth factor (or controlled variable). While the column order of the v factors is fixed, sort the previous solution $\mathbf{X} = \mathbf{X}_{\text{B}}^*$ in ascending row order. The $\mathbf{X}$ includes n^* unique runs, each of which correspond to a certain number of replicates. Calculate the initial Gauss-Hermite approximation ϕ_{BD} for the $\mathbf{X}$.

4.2 Reset $k = 1$ and flag $\Upsilon = 0$.

4.3 Let $i = 1$ and $i^* = 1$ be the starting values.

4.4 Pick out the i^*th unique run of $\mathbf{X}$, which corresponds to n_{i^*} replicates. The aim is to find the best candidate as a substitute for the kth coordinate of the selected unique run. It should then affect the relevant coordinates in the n_{i^*} replicates.

4.5 Create a subset of candidate coordinates $\mathbf{\Omega}$, for instance, on the basis of the closest distance, a defined space $[0.95X_{ik}, 1.05X_{ik}]$ (when the current coordinate value X_{ik} is nonnegative; the limits of this space can be changed to suit different examples), and the variable space $\mathcal{X}$. With the subset, do an optimisation for this unique run and denote X_{new} as the best candidate that maximises ϕ_{BD}. If $X_{ik} \neq X_{\text{new}}$, execute the coordinate substitution for each replicate and set $\Upsilon = 1$. Let $i = i+n_{i^*}$.

4.6 Unless $i = n + 1$, let $i^* = i^* + 1$ and return to **step 4.4** .

4.7 Unless $k = v$, let $k = k + 1$ and return to **step 4.3**.

4.8 Unless $\Upsilon = 0$, return to **step 4.2**. Otherwise, save the current $\mathbf{X}$ as the final solution $\mathbf{X}_{\text{B}}^{**}$, which is considered to be optimal under the expected D-criterion.

4.7 Final Results in the First Example

In simple Bayesian optimal design when $n^* \times v$ is a reasonable number (i.e. not too large), the above procedure can make sure that our final solution is optimal. This is also an alternative to doing a one-step optimisation of $\mathbf{X}$ when $\mathbf{X}_{\text{B}}^{*}$ acts as the initial solution, in order to maximise the expected D-criterion function. With the modified adjustment algorithm and the assumed lognormal prior for the parameter k, the new result is shown in Table 4.8, when the total Gauss-Hermite sample size is 16.

Table 4.8: New Optimal Design under the Lognormal Prior Assumption of k

(a) $\mathbf{X}_{\text{B}}^{*}$						(b) $\mathbf{X}_{\text{B}}^{**}$					
E	S	rep.	E	S	rep.	E	S	rep.	E	S	rep.
0.026	3	8	0.12	0.23	7	0.025	3	8	0.12	0.23	7
0.086	3	7	0.12	3	7	0.086	3	7	0.12	3	7
0.096	0.3	1				0.096	0.3	1			

The difference is on the first unique run, where the coordinate of the protein weight has been updated through the modified adjustment algorithm. As a result, the Gauss-Hermite approximation of the expected D-criterion value φ increases a bit from -42.8232 of $\mathbf{X}_{\text{B}}^{*}$ to -42.8231 of $\mathbf{X}_{\text{B}}^{**}$. In addition, the final solution is quite similar to the one in Figure 4.1, though we now assume a lognormal prior distribution of k.

In this simple example, the total number of coordinates from the unique runs is $n^* \times v = 10$ and it is small enough to be dealt with. We could exploit the modified adjustment algorithm further, in an attempt to reduce the amount of computation on a substantial

scale. As we follow the above procedure, the final solution $\mathbf{X}_{\mathrm{B}}^{**}$ is then less dependent on the qualities and efficiencies of the spectrum of $\mathbf{X}_{1\mathrm{st}}$, which can be obtained after an initial discrete optimisation. As such, we demand fewer of them for the continuous optimisation of the hybrid method, so computation could be less intensive.

We still wish to see an efficient $\mathbf{X}_{\mathrm{B}}$ after the continuous optimisation, which is not hard when $n^* \times v = 10$. After the conversion of those quasi-replicates to some exact ones, the intermediate solution $\mathbf{X}_{\mathrm{B}}^{*}$ should include suitable support points which can be further refined and validated with the modified adjustment algorithm. As long as the number of unique runs and the respective numbers of replicates are all correct, it is even acceptable to set $\tau^* = 1$ as the number of tries during the continuous optimisation. In this case, we might wish to adapt the configuration of the hybrid method as follows:

- We still do $\tau = 30$ random tries for the discrete optimisation over the small 3×3 candidate set $\mathbf{\Omega}$. As $\tau^* = 1$, it is safe to use a loose exchanging rule $\max(d_i) > 1.0001$ and keep the most efficient solution $\mathbf{X}_{1\mathrm{st}}$ for the continuous optimisation. To choose a proper rule, the critical value (or threshold) 1.0001 of this rule should depend on the numbers n and p in the example.
- For the continuous optimisation, we use the exchanging rule $\max(d_i) > 1.00000001$. A loose rule makes it easier to determine the most appropriate n^* and the replication of unique runs. It will facilitate the conversion of the quasi-replicates too.
- If the dimension v is not too high (e.g. less than four), we use the point exchange approach instead of the coordinate exchange approach. We might therefore have a better chance to find efficient solutions after the continuous optimisation.

After the above adaptation, the last piece of the jigsaw is to implement the modified adjustment algorithm which can be made the most of. In the demonstration for this example, we have obtained the same solution as in Table 4.8b, so it is unnecessary to do more tries in the continuous optimisation. Even if the number $n^* \times v$ is not so small in a different example, we still endeavour to use fewer tries to find an efficient final solution. On the other hand, more computational effort will enable us to better validate numerical results and make them more credible.

Another approach to deal with the lognormal prior density $\rho(k)$, when we compute the Fisher information, is to reparametrise model (4.5) as below:

$$\nu_i = \frac{\exp(a_0 + a_1 x_i + a_2 {x_i}^2) S_i}{\exp(k') + S_i} + \varepsilon_i \tag{4.8}$$

where the substitution is $k = \exp(k')$ such that $k' \sim \mathbb{N}(-1.3171, 0.5531^2)$. With the new normal prior distribution, one shall compute the information matrix and the expected D-criterion function φ for (4.8), which can be approximated with a different Gauss-Hermite

quadrature. To be clearer, this is not an equivalent consideration to the previous one, since the new statistical interest is in the estimation of k' instead of the parameter k in (4.5). As we do so, the upside is that a multivariate normal prior distribution can be defined for the new parameters in $\boldsymbol{\theta}$, whereas the downside is the change of the real experimental focus on the model. It is sometimes hard to refocus on the original model after we use an advanced reparametrisation.

With the downside in mind, in this example, the D-criterion will be sensible, which is independent of the numeric scale of each estimate in $\hat{\boldsymbol{\theta}}$. It is therefore feasible to consider model (4.8) after the simple reparametrisation. However, even if we use the D-criterion, it is sometimes important to avoid complex transformations and even reparametrisations on a nonlinear theoretical function. In practice, for instance, the transformation in (4.6) is rare to be the most suitable form of the model. On occasion, the fitted transformed model will not lead to a natural and meaningful interpretation of the true unknown mechanism, which is often considered as the main experimental purpose. Although that is not a serious concern in this example, the new experimental data should be collected in order to fit the original model (4.5), even if the adopted D-criterion function φ is derived for the reparametrised model (4.8).

In line with our earlier discussion, the reparametrisation of k will influence the nonlinearities of the specified model, as well as the parameter estimates under nonlinear least squares. This is due to some sort of function linearisation implemented in the Gauss-Newton algorithm (or the similar Levenberg-Marquardt algorithm), which we use for nonlinear regression. Moreover, if we fit the reparametrised model to the approximate data from Martins et al. (1999), the new estimate is $\tilde{k}' = \log(\tilde{k}) = -1.1642$ and there is a different standard error as well. Given that we are interested in (4.5), the normal prior distribution to be assumed must be $k' \sim \mathbb{N}(-1.3171, 0.5531^2)$, which is derived from the fitted model without the reparametrisation.

It is important to examine that our assumption of the parameter prior distribution is consistent with what we will do next in nonlinear regression. To approximate the new expected D-criterion function φ, we can use the Gauss-Hermite quadrature of the same total sample size 16. In other words, the GH approximation will be quite accurate when $m_{a_0} = 1, m_{a_1} = m_{a_2} = 2$ and the subsample size of k' is four. As such, we find the same final solution $\mathbf{X}_{\mathrm{B}}^{**}$ in Table 4.8, in spite of the simple model reparametrisation.

4.8 Approximation of the Spherical-Radial Transformation

To draw a deterministic sample from $\rho(\boldsymbol{\theta})$, recall that the Gauss-Hermite quadrature has to complete p unidimensional approximations in a row, with respect to each parameter. As a common quadrature rule, it will take all the combinations of the p univariate subsamples to form the total GH sample. If the size of $\boldsymbol{\theta}$ is large, we might need to

consider a lot of sampled prior vectors, which then lead to a high computational cost in the Gauss-Hermite approximation of φ.

In addition to various Gaussian quadrature rules, when the parameter prior density $\rho(\boldsymbol{\theta})$ is informative to experimenters, varieties of nonrandom sampling methods are available. There, the main purpose is to obtain an accurate numerical integration of the expected criterion function φ, whereas an overview of this topic can be found in Yu et al. (2010). Wherein, the spherical-radial (SR) transformation is most popular in recent years and it was developed to fit some common Bayesian optimal design examples (Monahan and Genz, 1997; Gotwalt et al., 2009). Hence, this section will provide a brief introduction to the spherical-radial transformation.

First, it is known that the SR transformation can work quite well with at least the multivariate normal density $\rho(\boldsymbol{\theta})$ (or the joint p.d.f. of independent normal priors). To understand its core idea, we can rewrite the variance-covariance matrix of $\rho(\boldsymbol{\theta})$ as $\boldsymbol{\Sigma} = \mathbf{D}\mathbf{D}^{\mathrm{T}}$, where $\mathbf{D}$ is a lower triangular matrix to be found after a conventional Cholesky decomposition. If the total sample size of the SR transformation is m, for $r = 1, 2, \ldots, m$, a new expression of the rth sampling unit $\boldsymbol{\theta}^r$ is $\tilde{\boldsymbol{\theta}} + \mathbf{D}\Delta^{(r)}\boldsymbol{u}^r$. Here, $\boldsymbol{u}^r$ indicates a specific location to be sampled on the surface of a unit $(p-1)$-sphere $\mathcal{S}$. This means that the norm of $\boldsymbol{u}^r$ is equal to $(\boldsymbol{u}^r)^{\mathrm{T}}\boldsymbol{u}^r = 1$, for $r = 1, 2, \ldots, m$. Meanwhile, we denote $\Delta^{(r)}$ as the radius in correspondence to this location, which can be sampled from the continuous space $[0, +\infty)$. Altogether, $\Delta^{(r)}\boldsymbol{u}^r$ can be treated as a sampling unit from the multivariate standard normal distribution, which is similar to the role of an abscissa for the Gaussian-Hermite quadrature. This sample is also independent of the mean $\tilde{\boldsymbol{\theta}}$ and the variance-covariance matrix $\boldsymbol{\Sigma}$ of the assumed normal density $\rho(\boldsymbol{\theta})$.

With the same multiple integral to solve, the SR transformation endeavours to approximate the volume of the hyperellipsoid region within the boundary of φ. It takes two numerical integration steps on the basis of the above reparametrisation: 1. the evaluation of an inner unit spherical integral in terms of $\boldsymbol{u}$ (though it is a line integral with the dimension $p-1$); and 2. the subsequent evaluation of the outer radial integral in terms of one parameter Δ. Hence, we will draw a sample of the product $\Delta\boldsymbol{u}$ before computing the SR transformation sample of $\boldsymbol{\theta}$.

An extended simplex integration method (Mysovskikh, 1980) can be chosen to approximate the spherical integral and the approximation will be exact if the local D-criterion function is a polynomial up to the degree $(2 \times 3 - 1) = 5$. While the surface area of a unit $(p-1)$-sphere $\mathcal{S}$ is $2(\pi)^{p/2}/\Gamma(p/2)$, the integration requires $m_1 = (p+1)(p+2)$ deterministic draws of $\boldsymbol{u}$ such that

$$\int_{\mathcal{S}} d\boldsymbol{u} = \frac{2(\pi)^{\frac{p}{2}}}{\Gamma(\frac{p}{2})}; \quad \int_{\mathcal{S}} \phi_{\mathrm{D}}(\mathbf{X}|\tilde{\boldsymbol{\theta}} + \mathbf{D}\Delta\boldsymbol{u})d\boldsymbol{u} \approx \sum_{r=1}^{m_1} w_1^{(r)} \phi_{\mathrm{D}}(\mathbf{X}|\tilde{\boldsymbol{\theta}} + \mathbf{D}\Delta\boldsymbol{u}^r)\frac{2(\pi)^{\frac{p}{2}}}{\Gamma(\frac{p}{2})}. \quad (4.9)$$

As usual, ϕ_{D} is the local D-criterion function in terms of the new parameters $\boldsymbol{u}$ and Δ. In correspondence to each sampled parameter prior vector $\boldsymbol{u}^r$, the weight $w_1^{(r)}$ should be computed, which depends on the number of parameters p. Hence, we can use (4.9) to evaluate the inner spherical integral of φ.

On the other hand, Δ in (4.9) can be integrated out of φ with the generalised Gauss-Laguerre quadrature, which requires $\rho(\boldsymbol{\theta})$ to be joint normal. If the multivariate prior distribution is nonnormal, one should reparametrise the model in order to normalise the density $\rho(\boldsymbol{\theta})$ (Gotwalt et al., 2009). In that circumstance, the SR transformation approximation of the adapted φ will suffer from a larger error due to the outer radial integration. This is a crucial problem to be solved.

As one of the Gaussian quadrature rules, the generalised Gauss-Laguerre approximation is exact if ϕ_{D} is a polynomial of degree $2m_2 - 1$ or less, where m_2 is the subsample size of the radius Δ. In our demonstration, we choose $m_2 = 3$ in order to match the approximation in the extended simplex integration. Now the expected D-criterion function of $\mathbf{X}$ can be rewritten as

$$\varphi = (2\pi)^{-\frac{p}{2}} \int_0^{\infty} \int_{\mathcal{S}} \phi_{\mathrm{D}}(\mathbf{X}|\tilde{\boldsymbol{\theta}} + \mathbf{D}\Delta\boldsymbol{u}) d\boldsymbol{u}\ \Delta^{p-1} e^{-\frac{\Delta^2}{2}} d\Delta. \tag{4.10}$$

After a further substitution $z = \Delta^2/2$, the above can be integrated as

$$\begin{aligned} \varphi(z, \boldsymbol{u}) &= \frac{\pi^{-\frac{p}{2}}}{2} \int_0^{\infty} \int_{\mathcal{S}} \phi_{\mathrm{D}}(\mathbf{X}|\tilde{\boldsymbol{\theta}} + \mathbf{D}\sqrt{2z}\boldsymbol{u}) d\boldsymbol{u}\ z^{\frac{p}{2}-1} e^{-z} dz \\ &\approx \frac{\pi^{-\frac{p}{2}}}{2} \left(w_2^{(1)} \int_{\mathcal{S}} \phi_{\mathrm{D}}(\mathbf{X}|\tilde{\boldsymbol{\theta}}) d\boldsymbol{u} + \sum_{r=2}^{m_2} w_2^{(r)} \int_{\mathcal{S}} \phi_{\mathrm{D}}(\mathbf{X}|\tilde{\boldsymbol{\theta}} + \mathbf{D}z^{(r)}\boldsymbol{u}) d\boldsymbol{u} \right), \end{aligned} \tag{4.11}$$

where a zero coordinate is added as the first of the sampling units or generalised Gauss-Laguerre abscissas $z^{(r)}$ (i.e. $\mathbf{D}z^{(1)}\boldsymbol{u} = 0$). Besides, w_2 is the associated weight. If we use (4.11) for the radial integration, the fundamental assumption $\Delta\boldsymbol{u} \sim \mathbb{N}(\mathbf{0}, \mathbf{I})$ should not be violated when we assume the parameter prior distribution. As (4.11) can be combined with (4.9), the total sample size is $m = 1 + (m_2 - 1)m_1 = 1 + 2(p+1)(p+2)$ for conventional SR transformation approximation.

Now we can calculate the prior vectors of $\boldsymbol{\theta}$. For $r = 1, 2, \ldots, m_2$, it could be observed that each nonzero abscissa $z^{(r)}$ is identical to one root of the generalised Laguerre polynomial $L_{m_2}^{\mathfrak{s}+1}(z)$, where $\mathfrak{s} = p/2 - 1$ (Krylov and Fedenko, 1962). Let $\mathfrak{o} = m_2 - 1$. With respect to the generalised Gauss-Laguerre quadrature (Cassity, 1965; Kopal, 1961), the weight in correspondence to $z^{(r)}$ should be computed with the formula

$$w_2^{(r)} = \frac{\Gamma(m_2)\Gamma(m_2 + \mathfrak{s})}{(m_2 + \mathfrak{s})(L_{\mathfrak{o}}^{\mathfrak{s}}(z^{(r)})\mathfrak{o}!)^2},$$

for $r = 1, 2, \ldots, m_2$. The sum of these is $\sum_{r=1}^{m_2} w_2^{(r)} = \Gamma(\mathfrak{s}+1) = \Gamma(p/2)$, which is a term to be cancelled out in the SR transformation. Notice that our definition of the generalised Laguerre polynomial $L_{\mathfrak{o}}^{\mathfrak{s}}$ incorporates the constant denominator $\mathfrak{o}!$ which is suppressed sometimes in the literature (since it will be cancelled out with the same numerator $\mathfrak{o}!$). In the most common cases when the SR transformation sample requires two or three units of the radius, Gotwalt (2010) simplified and rewrote the above formulae of the radii and the weights in terms of the number of parameters p, where the original weights were also divided by a gamma function $\Gamma(p/2)$.

We have chose the total sample size to be $m = 4 \times 1 \times 2 \times 2 = 16$ to construct the Gauss-Hermite quadrature. However, when the model is complex with more parameters, the GH approximation will become a bit expensive in computation. In contrast, if we assume a multivariate normal prior density $\rho(\boldsymbol{\theta})$, the SR transformation sample size is $m = m_3(m_2 - 1)(p + 1)(p + 2) + 1$, where $m_2 = 3$ and $m_3 = 1$ are the numbers for its standard version. We can use an integer $m_3 > 1$ to indicate the number of random orthogonal matrices which can be integrated into the approximating function in (4.9). This is a dedicated approach for improving the spherical integration (Monahan and Genz, 1997), whereas $m_1 = (p+1)(p+2)$ is fixed. Besides, when we will sample at most three model parameters from the assumed multivariate prior distribution (i.e. rather than to fix their prior values at constants), there will be some quasi-replicates of $\boldsymbol{\theta}^r$ at times. This can be considered to reduce the total sample size.

However, the spherical-radial transformation is not flawless. The outer radius integration should follow the Gaussian quadrature rule, so it can be improved when we increase the subsample size m_2. In contrast, to improve the inner spherical integration, Monahan and Genz (1997) mentioned that we can choose a number $m_3 > 1$ so as to introduce some randomness into the subsample of $\mathbf{u}$. This is plausible because if $\mathbf{Q}$ is a random orthogonal matrix, $(\mathbf{DQ})^{\mathrm{T}}(\mathbf{DQ}) = \mathbf{DD}^{\mathrm{T}} = \boldsymbol{\Sigma}$ is identical to the variance-covariance matrix, though the sampled parameter prior vectors will be different.

When $m_3 = 1$, the standard SR transformation can deliver an exact approximation of φ, as long as ϕ_{BD} is a linear function of degree five or less. On the downside, it is hard to make further improvement in this approximation, even if we increase the sample size m_3. Compared with the Gauss-Hermite quadrature, the standard sample size of the SR transformation is sometimes smaller but that would not be the case if we increase m_3 at the same time. If the extended simplex integration fails to be an accurate approximation, it is then problematic to eliminate the error of the overall SR transformation approximation. This can happen when we assume some excessive variances or covariances in $\boldsymbol{\Sigma}$ or when $\rho(\boldsymbol{\theta})$ is not a multivariate normal density. If we must define a large m_3, sometimes it is better to use the Gauss-Hermite quadrature.

The SR transformation can be implemented to approximate φ in the JMP software, where the built-in computer algorithm uses the coordinate exchange approach (Gotwalt

et al., 2009) and thus will resort to unidimensional optimisation with the Brent's method (Brent, 1973, chap.5). This is incorporated as a convenient function for Bayesian optimal design. In our example, however, we would like to make the results more comparable. What we have done is to adapt the hybrid method as well as the modified adjustment algorithm, in order to approximate φ with the standard SR transformation (i.e. the numbers are $m_2 = 3$ and $m_3 = 1$) instead of the Gauss-Hermite quadrature.

In the current example, three of the four parameters (if a_0 is taken into account, the sample size is $m = 61$ for the standard SR transformation) should be integrated out from φ. If the tailored SR transformation sample includes the 12 quasi-replicates of $\boldsymbol{\theta}^r$, for $r = 1, 2, \ldots, m$, the total sample size is $m = 41$ as we then use it to approximate the expected D-criterion function.

In comparison to the Gauss-Hermite quadrature of the sample size $m = 4 \times 1 \times 2 \times 2 = 16$ from the last section, the SR transformation appears to be less flexible as it assumes the equal importance of the parameters of the interest. If we use the Gauss-Hermite quadrature, we can do a quick numerical investigation in order to determine a suitable subsample size for each parameter. The GH subsamples are then summarised to form an overall GH sample of the parameters. In contrast, even though we can do similar computation to determine the SR transformation, we cannot see the result in more details. This is because, after the reparametrisation of φ, our attention is drawn to $\boldsymbol{u}$ and Δ. On the one hand, this makes the standard SR transformation sample size smaller when the number p is large. On the other hand, this disallows us from evaluating the impacts of the individual parameter priors on φ.

After we deduct the number of quasi-replicates, the sample size of the SR transformation is $m = 29$. It therefore takes more computational cost to approximate φ. Besides, to circumvent the trouble to deal with the lognormal prior of k, we use the expected D-criterion function for the reparametrised model (4.8), where the multivariate parameter prior distribution is normal. As a result, the final solution $\mathbf{X}_{\text{SR}}^{**}$ is identical to that in Table 4.8. In this case, the SR transformation also leads to an almost identical approximation of φ (the approximate value of which is -45.457373) as the Gauss-Hermite quadrature (the approximate value of which is -45.457368). This shows that the results of the SR transformation are reliable and interpretable in this example.

4.9 A Follow-Up Example in Optimal Design

In this section, we consider a complex hybrid nonlinear model in Chapter 2, which involves three controlled factors. The number of experimental runs is smaller, so there will be fewer replicates on the support points. Besides, more support points must be chosen to estimate the model parameters, so it is more difficult to find the Bayesian optimal design. The parameter prior also becomes less informative in this case, since

we use a smaller reference dataset to derive $\tilde{\boldsymbol{\theta}}$. As such, we should perhaps increase the Gauss-Hermite sample size to ensure an accurate approximation of φ. It will then take more computational effort to maximise the function ϕ_{BD}.

In this example, we assume a new experiment which is same as the one in Mountzouris et al. (1999). The main interest is to learn the unknown mechanism of the enzymatic depolymerisation of a dextran (i.e. the substrate). The desirable reaction products are different types of oligodextrans. This reaction can take place in a stirred-cell membrane reactor, which the experimenters can monitor. In this case, endodextranase works as the enzyme activator which plays a crucial role in the depolymerisation of the dextran. Furthermore, there are three controlled variables: the substrate concentration S (2.5-7.5 % in w/v or weight/volume), the enzyme concentration E (0.625-62.5 Units ml^{-1} times substrate concentration) and the transmembrane pressure P (200-400 kPa).

In the original experiment in Mountzouris et al. (1999), there were multiple responses to be observed and each corresponded to a unique empirical model. In this example, we focus on one response, the substrate conversion rate ξ (%), which should be measured at a fixed reaction time. The previous experimental data is in relation to an 18-run face-centred central composite design (Table 4.9). It will act as our reference which is used to fit the model and compute the reference nonlinear least squares estimate of the model parameters. Note that the 16th run of the reference experiment failed so that the observation had been discarded. Therefore, the 16th observed response is missing in Table 4.9. For our convenience, this omission does not prevent us from assuming the same variable space $\mathcal{X}$ for the planned new experiment.

Table 4.9: The 18-Run Reference: A Central Composite Design with the Data

S	E	P	ξ	S	E	P	ξ	S	E	P	ξ
5	6.25	300	73.6	2.5	62.5	400	95.2	7.5	62.5	400	82.7
5	6.25	200	81.6	7.5	6.25	300	77.3	2.5	6.25	300	90.0
5	62.5	300	76.0	5	6.25	400	69.0	2.5	0.625	400	55.2
5	6.25	300	69.4	7.5	0.625	200	43.3	7.5	0.625	400	_
5	6.25	300	73.6	2.5	0.625	200	62.8	7.5	62.5	200	87.0
5	0.625	300	50.5	5	6.25	300	74.0	2.5	62.5	200	96.0

No mechanistic model could be identified for a reasonable interpretation of the observed response surface of ξ, so an empirical functional relationship must be built as a surrogate. In spite of the relative convenience of constructing a purely empirical model, it is more challenging but will be useful for us to fit a hybrid nonlinear model instead. This model should combine our current scientific understanding about the unknown mechanism with an established empirical function. While the remaining conditions are under control, we can observe a negative relationship from the old dataset. Within the variable space $\mathcal{X}$ of experimental interest, the higher the substrate concentration S the experimental run specifies, the lower the conversion rate ξ is.

The true mechanism behind this reaction seems to be complicated, but if we remove a number of restrictions in the mathematical derivation, we end up with a simple relationship between the two variables as $\xi = \beta_1 S/(\beta_2 + S) + \varepsilon'$. β_1, β_2 are unknown constants and ε' is the error term for the experimental observation. While β_1 is a nonnegative ratio, β_2 can control the curvature of the model function. This simplified and smooth model can be adopted as the base of the overall hybrid nonlinear model, regardless of the unknown mechanism.

The conversion rate ξ is the quotient of the instantaneous substrate concentration (measured after the reaction) to the initial concentration, so a transformed response can be chosen as $\xi/(100 - \xi)$, which is a logit function without the logarithmic scale. As such, we can obtain the next model $\xi/(100 - \xi) = \gamma_1 S(\gamma_2 + S) + \varepsilon''$. Two more controlled variables should be included afterwards, so we are inclined to replace the nonnegative constant γ_1 in this model with an exponential function in terms of E and P. As usual, some transformations of these variables will be suitable. Here, for $i = 1, 2, \ldots, n = 18$, we can compute two simple scaled variables

$$x_{\mathrm{E},i} = \log_{10}\left(\frac{E_i}{6.25}\right) \in [-1, 1]; \qquad x_{\mathrm{P},i} = \frac{P_i - 300}{100} \in [-1, 1].$$

With the six parameters $a_0, a_1, a_2, a_3, a_4, a_5$, the hybrid model can be established as

$$\frac{\xi_i}{100 - \xi_i} = \frac{\exp(a_0 + a_1 x_{\mathrm{E},i} + a_2 x_{\mathrm{P},i} + a_3 x_{\mathrm{E},i}^2 + a_4 x_{\mathrm{P},i}^2) S_i}{a_5 + S_i} + \varepsilon_i. \tag{4.12}$$

The NLS assumptions are the zero mean and the constant variance of the error ε_i, across the 18 uncorrelated observations. This model is similar to (4.5), but it is an empirical one and unrelated to the Michaelis-Menten mechanism. In the numerator of this model function, the exponential function excludes an interaction term between the two scaled variables. This is because the interaction will not contribute much to the model fitted to the data in Table 4.9. In comparison, model (4.12) uses fewer parameters than a full second-order linear model whereas it is not difficult to interpret such a close-to-linear model. Besides, the true value of a_5 should not be smaller than -2.5, minus the minimum of the initial substrate concentration. The value $a_5 = 0$ is undesirable too, which will assume the substrate concentration to be independent of the response ξ.

When the substrate concentration S is fixed, the model function (without the error) can be transformed into a simple linear function $\mathbb{E}(\log(\xi/(100 - \xi))) = a_0 + a_1 x_{\mathrm{E}} + a_2 x_{\mathrm{P}} + a_3 x_{\mathrm{E}}^2 + a_4 x_{\mathrm{P}}^2$, the response of which is a logit function. As we assume that model (4.12) reflects the true nonlinear response surface, we can find the Bayesian D-optimal design for (4.12). This requires the same hybrid exchange algorithm to do a sufficient number of tries (i.e. $\tau = 30$), along with the modified adjustment algorithm.

When we determine the multivariate prior distribution of $\boldsymbol{\theta}$, it is feasible to assume all the covariances to be zero in $\boldsymbol{\Sigma}$. As such, the Gauss-Hermite subsample for each model

parameter can be drawn from the respective univariate prior distribution. To obtain the prior information, we fit model (4.12) to the reference data in Table 4.9. The estimate is $\tilde{\boldsymbol{\theta}} = \{\tilde{a_0}, \tilde{a_1}, \tilde{a_2}, \tilde{a_3}, \tilde{a_4}, \tilde{a_5}\} \approx \{0.4340, 1.3140, -0.1059, -0.8224, 0.4105, -2.0633\}$. $\tilde{\boldsymbol{\theta}}$ is then chosen as the mean of the prior density $\rho(\boldsymbol{\theta})$, the standard deviations of which are set to $\{0.1936, 0.3717, 0.0537, 0.4492, 0.2611, 0.0822\}$, which are $\sqrt{5}$ times the standard errors we find in the NLS estimation. While we could assume univariate normal prior distributions for the parameters a_0, a_1, a_2, a_3, a_4, it is sensible to consider $\rho(a_5) + 2.5$ as a lognormal distribution, for instance. In this case, $a_5 > -2.5$ indicates the lower limit when we are to sample the prior values of a_5, so the assumption is $\rho(a_5) + 2.5 \sim \log\mathbb{N}(-0.8459, 0.1867^2)$. Given the small standard deviation in this case, one can even assume a normal prior for a_5 without much trouble in computation.

To obtain an accurate approximation of φ, we should find a suitable Gauss-Hermite subsample size for each parameter when the other GH subsample sizes are fixed to one. With the central composite design in Table 4.9 as the reference, the results of our numerical investigation are shown in five of the columns in Table 4.10. Recall that we can fix the prior value of a_0 to the constant $\tilde{a}_0$, since its univariate normal prior distribution does not affect the expected D-criterion. For the last column, we combine all the GH subsamples which are of equal size m_j, for $m_j = 1, 2, \ldots, 8$. As such, the true expected D-criterion value φ should be near 32.2518.

Table 4.10: Gauss-Hermite Approximations of the Expected D-Criterion Value

	$\{a_5\}$	$\{a_1\}$	$\{a_2\}$	$\{a_3\}$	$\{a_4\}$	$\{(a_0), a_1, a_2, a_3, a_4, a_5\}$
m_j=1	31.9290	31.9290	31.9290	31.9290	31.9290	31.9290
m_j=2	31.9349	32.0408	31.9416	32.0760	31.9781	32.2502
m_j=3	31.9348	32.0439	31.9415	32.0758	31.9778	32.2518
m_j=4	31.9348	32.0440	31.9415	32.0755	31.9778	32.2519
m_j=5	31.9348	32.0439	31.9415	32.0756	31.9778	32.2518
m_j=6	31.9348	32.0439	31.9415	32.0756	31.9778	32.2518
m_j=7	31.9348	32.0439	31.9415	32.0756	31.9778	32.2518
m_j=8	31.9348	32.0439	31.9415	32.0756	31.9778	32.2518

The variance of $\rho(a_5)$ is small, since the reference data is quite suitable for the estimation of a_5. As a result, we can use a small Gauss-Hermite subsample for a_5. According to Table 4.10, our intuitive choices of the six subsample sizes are $m_{a_1} = m_{a_3} = 3, m_{a_2} = m_{a_4} = m_{a_5} = 2, m_{a_0} = 1$ for the model parameters. As such, the total Gauss-Hermite sample size is $m = 72$, smaller than the size 85 of the alternative standard SR transformation sample. We now draw this GH sample and construct the function ϕ_{BD}, which we endeavour to maximise with an optimal $\mathbf{X}$.

In our example, the closest distance is set to 0.01 for $S \in [2.5, 7.5]$, 0.005 for $E \in [0.625, 62.5]$, and 0.1 for $P \in [200, 400]$. The traditional discrete optimisation over the complete candidate set will be slow and cumbersome, so the hybrid method is useful. The

final solution $\mathbf{X}_{\mathrm{B}}^{**}$ is illustrated in Table 4.11. All the substrate concentration levels are equal or at least close to the lower limit 2.5, which is defined in the continuous variable space $\mathcal{X}$. To estimate the six model parameters, there are $n^* = 9$ support points and most of them correspond to two exact replicates. As we can see, this solution is quite similar to the local D-optimal design we have shown in Chapter 2. Hence, if we assume a close-to-linear hybrid model similar to (4.12), sometimes the expected D-criterion will not improve much the result under the local D-criterion. In spite of this, if possible, we should continue to use the Bayesian optimal designs for validation purposes in practice, which are most robust to the assumed variations in the true parameter values.

Table 4.11: Optimal Design of the Experiment under the Expected D-Criterion

S	E	P	rep.	S	E	P	rep.	S	E	P	rep.
2.5	2.75	200	3	2.5	20.31	200	2	2.5	62.5	400	2
2.5	17.85	400	2	2.5	62.5	200	2	3.09	62.5	200	1
2.5	20.3	288.1	2	2.5	62.5	289.1	2	3.12	27.075	200	2

Chapter 5

Model Transformation under Michaelis-Menten Mechanisms: Optimal Design of Experiments for Transform-Both-Sides Models

5.1 Review of the Michaelis-Menten Equation

In this section, we continue to discuss Michaelis-Menten kinetics and we will first introduce more biochemical background. As a chemical application of the rate laws, the fundamental two-step mechanism of Michaelis-Menten kinetics can be expressed as

$$\underset{\mathbf{E_t}}{\text{Enzyme}} + \underset{\mathbf{S_t}}{\text{Substrate}} \underset{k_b}{\overset{k_f}{\rightleftarrows}} \underset{\mathbf{I_t}}{\text{Intermediate}} \xrightarrow{k_2} \text{Enzyme} + \underset{\mathfrak{P}_\mathbf{t}}{\text{Product}}. \tag{5.1}$$

We denote the initial concentrations of the enzyme and the substrate as E and S, when the reaction takes place in a solution of fixed volume. As the current time in the reaction is t, the corresponding instantaneous concentrations can be written in mathematical equations $E_t = E - I_t$ and $S_t = S - I_t$ (as long as the product concentration $\mathfrak{P}_t$ is negligible), where I_t is the concentration of the intermediate (or an enzyme-substrate complex in Michaelis-Menten kinetics).

Decomposition or dissociation of the intermediate will regenerate free enzyme molecules in the overall equivalent concentration, so there is no irreversible consumption of the enzyme. Moreover, the denotions k_f, k_b, k_2 represent the rate constants (or reaction rate coefficients) for each step of Michaelis-Menten kinetics, which link those instantaneous concentrations to the reaction rate. All the rate constants are dependent on several influential factors or latent variables that exclude the current time t.

At the reversible first step, the classical Michaelis-Menten equation is on the basis of the **rapid equilibrium assumption** such that the expected reaction should be rapid enough to attain chemical equilibrium. As such, E_t, S_t and I_t are all approximately constant before and after the rate measurement time t. Meanwhile, in the vicinity of time t, the equation $-\partial S_t/\partial t = k_f E_t S_t = k_b I_t$ should hold at least in a short period. In general, the rapid equilibrium assumption requires low reaction speed in the second step: i.e. $k_f E_t S_t = k_b I_t \gg k_2 I_t$ since there should be no appreciable increment in $\mathfrak{P}_t$. Here, both S_t and I_t are assumed to be constants and there is also $k_b \gg k_2$.

The mathematical derivation continues. As long as k_f and k_b are fixed rate constants, the ratio $k = E_t S_t/I_t = k_b/k_f$ should be a constant too, which can be defined as the substrate dissociation. Under the rapid equilibrium assumption imposed at time t, the instantaneous intermediate concentration is equal to

$$I_t = \frac{(E - I_t)S_t}{k} = \frac{E}{(k/S_t) + 1}$$

after rearrangement of the fraction. Meanwhile, the second step of the mechanism can be interpreted as a first-order reaction step, so the reaction rate of the intermediate is

$$\nu_t = k_2 I_t = \frac{k_2 E S_t}{k + S_t} \quad \text{at time } t.$$

This is defined as the classical Michaelis-Menten equation, when it excludes the observational error of the response. Michaelis and Menten (1913) monitored the whole course of the enzymatic reaction and thus attempted to learn the developed kinetics over the time flow. However, a simpler research approach is to aim at the observation of the initial rate ν at $t \to 0$. As such, the response should be measured at the very beginning of the reaction before an appreciable decrease of the substrate concentration. With the condition $t \to 0$ (or $S \gg E$ in a large excess), both I_t and $\mathfrak{P}_t$ are to be small in comparison to the instantaneous substrate concentration such that $S_t = S$ would be a reasonable substitution. The initial substrate concentration is often known since it is under experimenters' control. An alternative to the above condition ($t \to 0$) is that ν_t is considered to be insensitive to the substrate concentration, which will require an extreme value for the ratio $k = k_b/k_f > 0$.

Given the fulfilment of the assumptions, we can build up the mechanistic relationship between the initial rate ν and the (initial) substrate concentration S as

$$\nu = \frac{k_2 E S}{k + S}.$$

Furthermore, when the intermediate concentration I_t is close to E, i.e. almost all receptors of the enzyme molecules will be occupied and in the bond with the substrate molecules, there would be $\nu \to k_2 E$ as the initial rate is approaching its upper limit.

For the same simple Michaelis-Menten kinetics, Briggs and Haldane (1925) took both reaction steps into account in the equilibrium. In their case, the modern Michaelis-Menten equation could be established under the **quasi-steady-state assumption** (QSSA) described in the text below. The classical and modern equations are in identical forms but the respective constraints or premises are not the same under the different assumptions.

As the reaction starts at $t = 0$, there is a brief and rapid acceleration of the reaction rate until it arrives at ν which we will measure. This acceleration is not considered in the Michaelis-Menten equation but it exists as we have to wait for molecular collisions which will lead to the bond of the enzyme and the substrate. It shall also increase the instantaneous intermediate concentration I_t from zero to a certain constant level before we look into the kinetics and its indispensable assumptions.

The response ν is the initial rate so that we expect an immediate finish of the rate acceleration. Afterwards, the enzymatic reaction should be able to achieve a balance called the *quasi-steady-state*. It would last for a short while when the instantaneous substrate concentration consumption is quite low. As the substrate concentration then decreases with passing time, the rate will start to decelerate after some point. Before this rate deceleration becomes influential, the concentration I_t could be approximately constant under the QSSA and this is when the response ν_t should be measured. In other words, at a suitable time t, the equation $\partial I_t/\partial t = 0$ should be satisfied under the QSSA and thus we can write

$$k_f(E - I_t)S_t = k_b I_t + k_2 I_t.$$

This equation illustrates the quasi-steady-state in the two-step Michaelis-Menten mechanism, where the reversible reaction step (valid in both directions) is of second order and the forward reaction step is of first order. As long as it holds from the start of the reaction, the product formation (not the substrate conversion in this mechanism) can be approximated to be proportional to the time span. If experimenters wish to measure the reaction rate ν_t, it is constant and can be expressed in the equation

$$\nu_t = k_2 I_t = \frac{k_f k_2 E S_t}{k_b + k_2 + k_f S_t} = \frac{k_2 E S_t}{\frac{k_b + k_2}{k_f} + S_t}.$$

Generalisation of this equation will be useful. Let $\nu_{\max} = k_{\text{cat}} E = k_2 E$ and $k = \frac{k_b + k_2}{k_f}$. Here, k is the substrate dissociation under the QSSA and it is different from the one under the rapid equilibrium assumption. At times, the modern Michaelis-Menten equation below can be applied to illustrate more complex kinetic mechanisms in addition to the simplest Michaelis-Menten mechanism (Cornish-Bowden, 2004, chap.2):

$$\nu_t = \frac{k_{\text{cat}} E S_t}{k + S_t} = \frac{\nu_{\max} S_t}{k + S_t} \quad \text{at time } t. \tag{5.2}$$

Here k_{cat} represents the rate constant of the second reaction step and it indicates the overall impact of the enzyme on the rate. Likewise, the remaining denotions also contain

some biochemical meanings: $\nu_{\max}$ implies the theoretical maximum rate at time t and it is constant when the enzyme concentration is fixed in (5.2); k is also called the Michaelis constant of which the unit of chemical measurement is the same as that of concentration.

If the molar concentration of the selected enzyme is unknown (e.g. since it is unpurified or its molecular mass is uncertain), then the substitution $\nu_{\max} = k_{\text{cat}}E$ would be favoured as the Michaelis-Menten equation then circumvents the independent variable E. In studies of more advanced kinetics, sometimes k_{cat} cannot defined as the rate constant for the second step of the Michaelis-Menten mechanism. For example, the following mechanism involves another reaction step in the formation and decomposition of the complex EI. Although it follows Michaelis-Menten kinetics overall, both k_{cat} and k would depend on the rate constants k_2 and k_3:

$$\underset{\mathbf{E_t}}{\text{Enzyme}} + \underset{\mathbf{S_t}}{\text{Substrate}} \underset{k_b}{\overset{k_f}{\rightleftarrows}} \underset{\mathbf{E_tS_t}}{\text{Intermediate-1}} \xrightarrow{k_2} \underset{\mathbf{E_tI_t}}{\text{Intermediate-2}} \xrightarrow{k_3} \text{Enzyme} + \text{Product}.$$

There are some such kinds of mechanisms that could require minor modification in the interpretation of the Michaelis-Menten equation. Besides, one should also examine the assumptions for each biochemical reaction and make sure there is no misuse of scientific equations or mechanistic models.

Let us return to the simplest two-step Michaelis-Menten mechanism (5.1). As the initial rate is measured as the response, we can expect the relation $S_t \approx S$. Then it is often most sensible to assume the rate ν_t to be constant in a narrow time span for those $t \to 0$. Hence, there is $\mathfrak{P}_\mathbf{t} \approx t\nu_t$. If we consider the classical rapid equilibrium assumption, it can hold for a little longer when the rate constant k_2 is smaller (i.e. the second step of the reaction is slower) or when the response measurement time is earlier (i.e. when the substrate conversion rate is lower).

In contrast, the new QSSA takes account of the rate for the second step too and it is therefore more realistic. **Under the modern Michaelis-Menten equation that can be derived under the QSSA, the relevant experimental constraints tend to be less strict.** In this case, the balance of the whole mechanism is what we shall care most about. Up to the time of the response measurement, there should not be a great decrement on the instantaneous substrate concentration S_t. Hence, like the rapid equilibrium assumption, the QSSA does not suit some fast reaction kinetics. Nevertheless, the QSSA can tolerate a higher instantaneous product concentration $\mathfrak{P}_t$ which does not have to be near zero in this case. Under the QSSA, I_t and $E_t = E - I_t$ are considered to be constant and then so is the rate $\nu_t = k_{\text{cat}}I_t$.

The new benefit is that, even though the time-independent rate constants k_{cat} and k_f are not often much smaller than k_b (this makes sense because $S_t + I_t + \mathfrak{P}_t = S$), the Michaelis-Menten equation can hold if the following condition is satisfied. As long as we assume $S_t \gg E_t$ or equivalent $S_t \to S$, which could happen at reaction time $t \to 0$, the

state $\partial I_t/\partial t = 0$ will be able to hold for a short time. An alternative exposition is that, since $I_t = E/(k+S_t)S_t$, the condition $E/(k+S_t) \ll 1$ or approximately $E/(k+S_0) \ll 1$ will be desirable as one wishes to reduce the percentage of substrate consumption.

When the observed response is the initial rate, the Michaelis-Menten equation is

$$\nu = \frac{k_{\text{cat}} E S}{k+S} = \frac{\nu_{\max} S}{k+S} \quad \text{at time } t \to 0. \tag{5.3}$$

Here the QSSA should be applicable for most mechanisms, except for those of fast reactions. Most kinetic studies focus on the initial rate, but there is another area for research. If the measurement time t is an independent variable according to the time course kinetics, the rationale behind the Michaelis-Menten equation would be invalid. In that case, a closed-form mathematical solution should be derived to determine k_{cat} and k of the nonlinear reaction rate equation (Beal, 1983; Schnell and Mendoza, 1997). It then leads to a different rate equation which can also be used to validate the Michaelis-Menten equation (5.2) under the QSSA.

5.2 More Assumptions under Michaelis-Menten Kinetics

Prior to our discussion of the error structure of the Michaelis-Menten model, it is worthwhile mentioning some mechanistic assumptions behind the modern Michaelis-Menten equation, in addition to the foremost one, the QSSA.

First, we can look at the stepwise reactions under the simple Michaelis-Menten mechanism (5.1). To comprehend fundamental enzyme kinetics, it is most common to assume the formation of the sole intermediate (often called the enzyme-substrate complex) step to be second-order and bimolecular. In mechanism (5.1), in other words, this reaction in the forward direction is considered to be first-order in both concentrations E and S (i.e. the stoichiometric coefficients or colliding molecular entities take value one). This means that the reaction will take one step and every enzyme molecule involved has only one active site that could collide and bind with the substrate molecules. This is the basic mechanistic assumption under the **QSSA**.

The natural form of an enzyme could be a protein in living cells or organisms. Their molecule sizes are always much larger than those of the substrate. In Michaelis-Menten kinetics, as the molecules of the two reagents bind with each other, the overall activation energy required to convert the substrate to the desirable product will decrease. As we now consider the role of the enzyme, the **second mechanistic assumption** is the conservation law of molar concentrations, which requires $E_t + I_t$ to equal the initial enzyme concentration E at time t. In practice, however, the enzyme might lose its full effectiveness in the course of the reaction such that the actual decrease of $E_t + I_t$

violates the conservation law. Due to the potential progressive enzyme inactivation, it is therefore most convenient to measure the initial rate when $t \to 0$.

The simple Michaelis-Menten mechanism does not consider the effects of the allosteric regulation or the cooperative binding elaborated in the famous Hill equation. Both could exist, for instance, in complex metabolisms in living cells (Tracy and Hummel, 2004) and their latent effects could depend on the nature of specific reactions and reagents. More importantly, the inhibition effects of either the intermediate or the products (and sometimes byproducts) should be taken into account and that might lead to a modification of the simple Michaelis-Menten equation. Hence, the **third mechanistic assumption** is that no inhibitors or additional activators are involved in the biochemical reaction. A violation of this assumption could also make the conservation law and the Michaelis-Menten mechanism (5.1) uncomprehensive.

When there is more than one reagent species, the biochemical reaction can be presumed to be stable and continuous over time (when the true mechanism could be assumed). This requires the experimenters to stir the solution (i.e. a mixture of different reactants) to make the present reactants diffusive. Otherwise, an observational error is unavoidable in the experiment and the model. A perfect experimental environment should be a closed biochemical system with a constant volume of the buffer solution and an ideal mixture of the reactants, which almost never exists in practice.

The rate constants must be independent of the concentrations of all the reactants. As we can learn, when the initial substrate concentration S is too low in relative to the fixed E, the initial rate would increase in direct ratio to the substrate concentration. This is because a lot of the substrate molecules would be free in this case, which can then collide and bind with the enzyme molecules. As a result, a certain fraction of the collisions will be successful so as to form the intermediate or the enzyme-substrate complex. However, it is rare to use such low substrate concentrations for the experiment.

On the other hand, when $S \gg E$, which is an optional requirement under the QSSA, the reaction rate will also approach to the upper limit $\nu_{\max}$ quickly. As the reaction starts, almost all the active sites of the enzyme will be occupied so that we can find $I_t \gg E_t$ under the QSSA. As the free enzyme concentration is so low, the reaction rate will cease acceleration soon and then it can be measured as the initial rate at $t \to 0$. Of course, as the reaction goes and the substrate is consumed at length, the same condition cannot be maintained. A recognisable and continuous decrement on $\nu = \partial\mathfrak{P}_t/\partial t$ is to follow after some point, even if the conservation law holds. Here our discussion can show how the rate could change in the course of a reaction.

To illustrate Michaelis-Menten kinetics through mathematical deductions, the **fourth mechanistic assumption** is the application of rate laws or rate equations to the fundamental two-step Michaelis-Menten mechanism. These rate laws shall hold when the experiment is conducted in a perfect environment described above. Thus the quantitative

changes of the intermediate concentration can be explained. In this simple mechanism,

$$\text{Intermediate} \rightarrow \text{Enzyme} + \text{Product}; \quad \text{Intermediate} \rightarrow \text{Enzyme} + \text{Substrate}$$

are both first-order reaction steps while

$$\text{Enzyme} + \text{Substrate} \rightarrow \text{Intermediate}$$

is of second order. The three stoichiometric coefficients are equal to one, so the law of mass action applies to this bimolecular reaction. Considering Michaelis-Menten kinetics as a whole, the differential equations can be written as

$$\partial S_t/\partial t = k_b I_t^{\alpha_3} - k_f E_t^{\alpha_1} S_t^{\alpha_2}; \quad \partial \mathfrak{P}_t/\partial t = k_b I_t^{\alpha_4}, \tag{5.4}$$

where the four reaction orders are determined as $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$. These orders should not be confused with the stoichiometric coefficients and as each of them is one, the rates in the individual steps are all proportional to the reactant concentrations.

The rate laws shall hold at least for unimolecular elementary reactions, which require just one step from the reactant to the ultimate product. In that case, the stoichiometric coefficients are identical to the reaction orders too. Meanwhile, if an elementary reaction is bimolecular or even has a larger molecularity, the circumstance will be complicated as we should consider the reactant diffusion and mixture. Without adequate experimentation to study those composite (i.e. multi-step) mechanisms, it is hard to determine appropriate reaction orders and thus use the rate laws, even though the stoichiometric coefficients are often known.

The Michaelis-Menten equation is the basis for many studies of single-substrate enzyme kinetics. However, more advanced reaction mechanisms could involve the formation of several reactive intermediates and some concurrent reaction steps. While a reaction could either make or break some bonds between the enzyme and the substrate, the rate equations have to be applied to derive the overall kinetic function if each step of the mechanism is informative. For instance, Kopanchuk et al. (2006) studied a mechanism of a dual binding kinetics, which is a bit similar to Michaelis-Menten kinetics. Hence, a unique statistical model can be derived from the relevant rate laws.

The rate laws also describe how the substrate concentration (or free enzyme concentration) varies in a specific time span. In the simple Michaelis-Menten mechanism, the two differential equations above indicate the velocities of the substrate consumption and the product formation at the measurement time t from the start of the reaction. However, even in a simple bimolecular reaction step, the relevant rate laws are reliant on the speed and the effect of the reactant diffusion and mixture.

Furthermore, the mixture of the two reagents will depend on their initial concentrations in the solution as well as the intervention of experimenters. If a solution is stirred on a regular basis, the diffusion will be faster and the mixture will be closer to being ideal. Under Michaelis-Menten kinetics, if the initial rate is to be measured, the rate constant k_f is not a real constant over different substrate (and enzyme) concentration levels. In this case, the rate laws and the law of mass action no longer work so well in the application of the rate equation. As we have mentioned in the second last paragraph in brief, some error will be expected in the explanation of the overall reaction rate ν or ν_t.

The rationale is, when $t \to 0$, the experimental environment is imperfect since the reactants are not in ideal mixture. Not all the substrate molecules are free to bind to the active sites of the enzyme molecules in a short time after the two substances are first brought together to enable the reaction. In addition to the reactant concentrations, a few factors can also affect the environment and conditions. As a result, molecular movements are somewhat restricted in the solution and that will alter the overall reaction rate ν. In this case, the rate law for the bimolecular reaction step does not hold all the time since its rate is not proportional to the substrate concentration.

The initial rate for the first step of the Michaelis-Menten mechanism (in the forward direction; it is not the response ν) can be written as $k_f E^{\alpha_1} S^{\alpha_2} = k_f ES$, in reference to the rate laws in (5.4). The rate laws might not hold in some cases, but as the enzyme concentration E is often fixed, we should at least treat k_f as a real constant across various substrate concentration levels and its bias is to be explained in the error term in a model. As such, **the Michaelis-Menten equation is assumed to be valid as it is derived on the basis of the rate laws, though there is a small probability that the variance of the observational errors is heterogeneous**. Hence, as one constraint to use the rate laws and measure the initial rate, **experimenters should assume an environment with a well-stirred solution or even an ideal mixture of the reactants**. This is almost impossible when the reaction starts at $t = 0$. For a fitted Michaelis-Menten model, to reduce the error due to the rate laws or the law of mass action, rapid mixing of the reagents are important in the experiment.

To summarise, the assumption of the rate laws in (5.4) leads to the requirements on the reactant mixture in the solution. The Michaelis-Menten equation considers the substrate concentration as a controlled variable, so the application of the rate laws would cause some error added to the expected response at a small t (i.e. initial rate). When the reaction starts at $t = 0$, the diffusion and free collision of the reactant molecules are limited, so we will tend to overestimate the initial rate. In particular, this is more likely to occur when experimenters choose a higher substrate concentration.

Given the restriction on the molecular mobility, the effect of the reactant diffusion and mixture is somewhat similar to a competitive inhibition effect which is in relation to the third mechanistic assumption we make. Therefore, there are two possible measures

to fix the problem before the Michaelis-Menten model is fitted. On the one hand, experimenters can think about a modification of the Michaelis-Menten equation or other theoretical equations concerned. On the other hand, after the equation is properly modified, experimenters can consider an appropriate transformation to correct the raw error structure of the model, in order to meet the assumptions of least squares or maximum likelihood estimation. The two measures will be discussed in details later in this chapter.

5.3 Other Factors in Association with the Initial Rate

This section provides a concise introduction to the most common factors which could influence the initial rate of Michaelis-Menten kinetics. When experimenters consider a more complex Michaelis-Menten mechanism, it is plausible to include one or more of these factors in the fitted statistical model. Even though it is often too difficult to find suitable mechanistic models in those situations, the relevant empirical modelling techniques will be the other options for conventional nonlinear regression.

A number of these factors could have been considered in the Michaelis-Menten mechanism, e.g. the measurement time, the inhibition or activation effects, and the reagent concentrations. In addition, **temperature** is a noticeable factor that could be in a close link with the nature of the reagents. For the best outcome from the reaction, each substrate or enzyme species shall correspond to an optimal range of temperature. Therefore, when a specific combination of the reagents is to be used for the reaction, it will be helpful if we know how to set the temperature as a fixed or variable experimental condition. In the first step of the Michaelis-Menten mechanism, for instance, a proper change of experimental temperature can activate effective molecule collisions and lead to a faster reaction at the time point t.

In practice, it is rare that the reaction rate is independent of or negatively correlated with temperature. Over a moderate range of temperatures, the common temperature influence on the rate constants can be represented with the Arrhenius equation. For the rate constant k_{cat} from the modern Michaelis-Menten equation, for instance, we can write the empirical Arrhenius equation

$$k_{\text{cat}} = A\exp\left(-\frac{E_{\text{a}}}{RT}\right) = A\exp\left(-\frac{1}{T}\frac{E_a}{R}\right) \propto \nu,$$

which can also be modified for different rate constants in a mechanism. Here, T is the temperature (in Kelvin) and R is the gas constant. With the remaining experimental conditions fixed, both A and E_{a} are unknown nonnegative constants. While A can be used to describe how often a molecular collision occurs at the reaction time t, E_a denotes the amount of the activation energy (i.e. the minimum energy required for successful collisions and for the reaction to occur). The Arrhenius equation implies some

mechanistic knowledge with reasonable scientific justification, but it is still an empirical mathematical relationship between temperature and the rate constant. This means that the equation is not exact and sometimes its prediction error is not small. The drawback is that the Arrhenius equation is not tailored to all reactions and respective rate constants, though it is simple to interpret this equation in terms of one independent variable. According to its assumptions, some of the experimental constraints seem to be oversimplified and thus it would be quite hard to integrate the Arrhenius equation into the Michaelis-Menten model (Cornish-Bowden, 2004, chap.10).

We can also think about the second factor: the **hydrogen ion concentration** or the pH scale which is a transformation of the former. It can have an impact on the formation and decomposition of the intermediate in Michaelis-Menten kinetics. Likewise, there shall be an optimal range of the hydrogen ion concentration for the reaction in the buffer solution or the mixture. When this factor is taken into account, a lot of experimentation will be demanded to compare different substrate and enzyme species.

It is apparent that if the hydrogen ion concentration is too low or too high on the acid or alkaline side of its true optimum, the normal biochemical reaction would not attain its maximum rate or it would be inhibited. Given the other experimental conditions and the response measurement time, the approximate pH effect can be considered to be analogous to an inhibition effect in the kinetics. If we know the specified hydrogen ion concentration as well as the corresponding acid dissociation constants, the Michaelis-Menten mechanism could be modified to take the pH effect into account in an empirical statistical model. However, due to some additional assumptions, which we have to make for a simplified mechanism, such an empirical model cannot fit all Michaelis-Menten kinetics. Even if it does (as being assumed), experimenters should be advised to check the error structure of the model.

The third factor is the **pressure** and it is vital at least in gaseous chemical reactions, which do not follow fundamental Michaelis-Menten kinetics. In those cases, it could influence the reactant concentrations since a high pressure can reduce the volumes of the reactants. In biochemical reactions in the liquid solution, in some respects, the effect of the hydrostatic pressure is a bit similar to that of temperature. In order to interpret the reaction mechanism, some dependence studies could be useful, which will also require some experimentation. A short discussion of the potential pressure effect can be found in Cornish-Bowden (2004, chap.1).

Experimenters can consider some other factors when the experimental conditions are to be set. For instance, the chemical reaction on surfaces (when at least one of the reagents is a solid) tends to be faster if we can enlarge the **contact area** or surface area of the solid reagent to facilitate molecular collisions. Under the same rationale, when all the reagents are liquids, the **diffusion** or the **mixture** will affect the rates of multimolecular

reactions. If these factors can be assessed in quantities, we would be able to learn how a difference is made in Michaelis-Menten kinetics.

5.4 Measurement Errors in the Experiment

This section focuses on experiments to learn the two-step Michaelis-Menten mechanism. This is most convenient as we can circumvent a lot of mathematical deductions and assumptions for more complicated kinetic studies. As far as we are concerned, the Michaelis-Menten model will be fitted to the experimental data. When its error structure is known or assumed, we can calculate the standard errors of the parameter estimates. In this case, there could be several different scenarios, since there is not a universal result which can be applied to all dissimilar Michaelis-Menten mechanisms.

With the dataset on hand, experimenters can invoke the Michaelis-Menten model or other theoretical models without regard to some essential assumptions made and the true mechanism behind the biochemical reaction. In this case, it is simple to fit the statistical model but as a result, the lack of fit could exist and the parameter estimation could be inaccurate. If an inappropriate mechanistic model is fitted, the observed error structure could be quite different from what we assume and thus lead to an ineffective use of the experimental data and nonlinear least squares (NLS) estimation.

Four mechanistic assumptions have been discussed in Section 5.2. When the response is the initial rate, there is a chance that the rate laws would be untrue. Our conjecture is that the limited ligand activities should be responsible for the inaccurate prediction of the initial reaction rate. As has been mentioned, this is because of the nonideal mixture of the two reagents for the first bimolecular step of the Michaelis-Menten mechanism.

The rate laws reflect an oversimplification of the mechanism, which we use to derive an mathematical function of the reaction rate. For instance, it is most natural to consider an increasing trend of the measurement errors as experimenters use higher and higher substrate concentrations. This is because of the prolonged time the reactant diffusion and mixture would take to fulfil the rate laws. If the diffusion and mixture are too slow or too anomalous, the rate laws would fail in the interpretation of the reaction mechanism, even if there is no inhibition or other activation effects.

There is another conjecture on the errors and the inhibition. Under the QSSA, the described mechanistic assumptions shall be best satisfied when the response is the initial rate. However, even at the measurement time $t \to 0$, the inhibition effect of the intermediate can be strong sometimes. It can happen when experimenters use a high substrate concentration so that the expected initial rate is also high.

In this case, the free enzyme concentration E_t would be rather close to zero whereas the intermediate concentration I_t approaches E. The problem is that the rate laws do not

consider the intermediate which is not involved in the reaction, though it does exist in the mixture under the assumed Michaelis-Menten mechanism (as does the product). As a result, a high intermediate concentration under the QSSA could somewhat reduce the rate of the collision and binding of the enzyme and substrate molecules. This makes the rate law for the bimolecular reaction step less reliable.

As to the Michaelis-Menten model, this conjecture can lead to an inaccurate estimation of the parameters as well as an increase in the measurement errors of the initial rate. When E is fixed, the predicted rates at the higher levels of S tend to be less accurate. However, note that the responses would not be found to be overestimated on average if the unknown true parameter values are to be estimated and thus we use the model fitted to the experimental data.

The measurement errors can be learnt when we think about the dependence relationship between the rate constants and the substrate concentration. The first-order steps (i.e. the decomposition or dissociation of the intermediate) would not cause much trouble to the errors, so our exclusive interest is in the bimolecular reaction step

$$\text{Enzyme} + \text{Substrate} \rightarrow \text{Intermediate},$$

which is studied at $t \rightarrow 0$. Under the rate laws, the differential equation in (5.4) is simplified to $\partial S/\partial t = k_b I_t - k_f E_t S$. We can use it to derive the modern Michaelis-Menten equation. If the error of the initial rate is to increase as the substrate concentration increases, the observed value of $k_f E_t S$ could be smaller than its expected value from the relevant rate law. As its consequence for the Michaelis-Menten mechanism which incorporates both reaction steps, there is an inequality

$$k_f(E - I_t)S > k_b I_t + k_{\text{cat}} I_t \quad \text{at } t \rightarrow 0.$$

This actual relationship violates the QSSA at the current time t. When there are n independent runs in the experiment, the above can be revised to be a function

$$k_f(E - I_{ti})\lambda(S_i)S_i \approx k_b I_{ti} + k_{\text{cat}} I_{ti}, \text{ for } i = 1, 2, ..., n.$$

Here $\lambda(S)$ is included as a discontinuous step function in terms of the initial substrate concentration and we define it as the *correction factor*. If the response error increases with the substrate concentration, the correction factor should be less than one most of the time. It could allow for comparisons of the error dispersions at various substrate concentration levels. Of course, for reactions in the ideal mixture, the correction factor should be close to constant at one. Furthermore, for the ith experimental run at a substrate concentration S_i, the approximate bias of the Michaelis constant k is

$$\frac{k_b + k_{\text{cat}}}{k_f} - \frac{k_b + k_{\text{cat}}}{\lambda(S_i)k_f}$$

if we know the three rate constants k_b, k_f and k_{cat} from the rate laws. If the response error increases with the substrate concentration, this bias above will be less than zero, on the basis of the ith observation. This means that the true k tends to be underestimated if we ignore the correction factor. Besides, as the step function value $\lambda(S)$ varies across the n observations, the biases of k are not identical with each other. Now, with an error term ε_i added to the Michaelis-Menten equation, the model can be written as

$$\nu_i \approx k_{\text{cat}} I_{ti} = \frac{k_{\text{cat}} E S_i}{k + S_i} + \varepsilon_i = \frac{\nu_{\max} S_i}{k + S_i} + \varepsilon_i \quad \text{at time } t \to 0. \tag{5.5}$$

Here the term $\nu_i \approx k_{\text{cat}} I_{ti}$ indicates the ith observation of the initial rate and the rate law for the first-order reaction step. For instance, we assume $\lambda(S_i) \leq 1$ to be an increasing function in the substrate concentration. As there is also $S_i \gg E_i$, the expected initial rate would be overestimated according to model (5.5). In this case, the model error ε_i shall be associated with the level of S_i, which is in contradiction to the fundamental assumptions of nonlinear least squares.

Recall that the approximation $S_t \approx S$ does not hold quite well at low substrate concentration levels which are often covered in the variable space. Thus an additional error will influence the fitted Michaelis-Menten model, even if the substrate consumption is low at the measurement time $t \to 0$ under the QSSA. It is judicious to set a reasonable lower limit of the substrate concentration, in order to make the QSSA more reasonable.

5.5 Residuals in Nonlinear Least Squares Estimation

Unexplained errors (i.e. model residuals) can lead to inaccurate estimates of the rate constants and an increase in the lack of fit of the fitted Michaelis-Menten model. If we deduct the calculated pure error out of the total error ε_i, the rest is due to the nonrandom lack of fit (i.e. variation between the unique treatments). The true response surface is unknown, so a model does not account for the lack of fit. We introduced the correlation factor $\lambda(S)$ in the previous section. While the Michaelis-Menten model does not involve this correlation factor which we assume to exist, the NLS estimation of k is inaccurate. It will contribute to much of the lack of fit. However, if $\hat{k}$ is an unbiased estimate that is identical to k/λ ($\lambda = \lambda(S_i)$ must be constant across the n observations), the other estimate $\hat{\nu}_{\max} = \hat{k}_{\text{cat}} E$ shall be quite accurate. This is because, unlike the rate constant k_f out of k, k_{cat} corresponds to a simple first-order reaction step and it is much easier to approximate its true value.

As a result of nonlinear least squares estimation to fit the Michaelis-Menten model, the expected initial rate can be expressed as

$$\hat{\nu}_i = \frac{\nu_{\max} S_i}{k + S_i} + \left(\frac{\hat{\nu}_{\max} S_i}{\hat{k} + S_i} - \frac{\nu_{\max} S_i}{k + S_i} \right).$$

The first component is the Michaelis-Menten equation with its most appropriate parameter values from the rate laws. The component inside the bracket denotes the expected deviation of the fitted model. If we make an approximation $\hat{\nu}_{\max} \approx \nu_{\max}$, which is assumed to be indeed an invariant constant across different substrate concentration levels, the famous Lineweaver-Burk equation can be derived to transform the above equation. As the reciprocal of the response should be taken, i.e. $\nu' = 1/\nu$, the transform-both-sides Michaelis-Menten model can be fitted as

$$\hat{\nu}'_i = \frac{k + S_i}{\nu_{\max} S_i} + \left(\frac{\hat{k} + S_i}{\hat{\nu}_{\max} S_i} - \frac{k + S_i}{\nu_{\max} S_i} \right) \approx \frac{k + S_i}{\nu_{\max} S_i} + \frac{\hat{k} - k}{\nu_{\max} S_i}.$$

A justification of the Lineweaver-Burk equation is important before we can adopt the above equation, but this transformation at least simplifies the model. If we include an error term and a correction factor for the response, the untransformed model should be

$$\nu_i = \left(\frac{k + S_i}{\nu_{\max} S_i} + \frac{\lambda(S_i)k - k}{\nu_{\max} S_i} \right)^{-1} + \varepsilon_i^{(1)}. \tag{5.6}$$

The parameter estimates are not included in the above expression. For $i = 1, 2, \ldots, n$, the mean of the error $\varepsilon_i^{(1)}$ should be zero and the variance should be constant. In contrast, if a better error structure is achievable when the response is set to be ν', the simpler transformed model should be

$$\nu'_i = \frac{1}{\nu_i} = \frac{k + S_i}{\nu_{\max} S_i} + \frac{\lambda(S_i)k - k}{\nu_{\max} S_i} + \varepsilon_i^{(2)}, \tag{5.7}$$

where the same assumption applies to the error, for $i = 1, 2, \ldots, n$. Experimenters can transform the Michaelis-Menten equation into whatever a form, but the aim is to improve the error structure and to better fit the experimental data.

Meanwhile, $\lambda(S_i)$ is a step function of the substrate concentration. For convenience, we can assume the true lack of fit of the Michaelis-Menten model to be well explained in the fitted model. That is, we assume $\lambda(S_i) = \lambda$ as a fixed constant and thus we can substitute $\lambda(S_i)k - k$ with $\hat{k} - k$. In this case, though the estimate of the Michaelis constant is not unbiased, model (5.7) becomes

$$\nu'_i = \frac{k + (\hat{k} - k) + S_i}{\nu_{\max} S_i} + \varepsilon'_i = \frac{\hat{k} + S_i}{\nu_{\max} S_i} + \varepsilon'_i, \tag{5.8}$$

where ε'_i is the updated error and $\hat{k}$ is the NLS estimate to be obtained. Given that the model is true and the simple substitution is reasonable, the bias of the Michaelis constant would be unknown unless the correction factor is a known constant. On the other hand, the lack of fit would not much disrupt or distort the error structure of ε'_i so that the NLS estimate of $\nu_{\max}$ is accurate. Otherwise we cannot assume $\hat{\nu}_{\max} = \nu_{\max}$ which is important for us to write down models (5.6) and (5.7).

In contrast, if the correction factor is no longer independent of the substrate concentration, it is harder to evaluate the lack of fit and this also depends on experimental conditions. In theoretical studies, the function $\lambda(S)$ could indicate those unknown effects or uncertainties that we wish to reduce. Otherwise, if we cannot eliminate the lack of fit, at least the error structure should be modified to facilitate (i.e. meet the assumptions under) the use of nonlinear least squares or maximum likelihood.

It is hard to interpret most of the unidentified errors of models (5.6) and (5.7). We can assume these errors to be random and small if the model transformation is correct. As to the lack of fit due to the correction factor, we further assume a first-order linear function $\lambda(S) = \lambda_0 + \lambda_1 S$. While λ_0 and λ_1 are two parameters, this is not uncommon to consider a simple empirical function like that. If the transformed response is ν', model (5.7) can be rewritten as

$$\nu_i' = \frac{k + S_i}{\nu_{\max} S_i} + \frac{\lambda(S_i)k - k}{\nu_{\max} S_i} + \varepsilon_i^{(2)} = \frac{\lambda_0 k + S_i}{\nu_{\max} S_i} + \frac{\lambda_1 k}{\nu_{\max}} + \varepsilon_i^{(2)}. \tag{5.9}$$

The unexplained component of the lack of fit is also small if we ignore the distribution of the total errors. While the bias of $\hat{k}$ is $(\lambda_0 - 1)k$, $\hat{\nu}_{\max}$ could also be biased due to an additional constant term. Note that when this constant is small enough, we cannot assume model (5.9) since the proposed substitution $\hat{\nu}_{\max} = \nu_{\max}$ will not be suitable. In this case, the fitted model cannot explain this source of the lack of fit well, which then requires us to find a better model.

No matter whether we transform the model or not, we should take some measures to improve the NLS estimation of the parameters as well as to meet the essential assumptions of the errors. As we have presumed the exact function of the correction factor, the most direct measure is to take the correction factor into account in a modified Michaelis-Menten model. Before we discuss that, recall that the effect of the competitive inhibition (as one of the basic linear inhibitions) can be taken into account in Michaelis-Menten kinetics. If the initial free inhibitor concentration is $\mathfrak{I}$, the kinetic model of the initial rate can be modified as

$$\nu_i'' = \frac{\nu_{\max} S_i}{k(1 + \mathfrak{I}_i/\mathfrak{k}) + S_i} + \varepsilon_i'' \quad \text{at time } t \to 0. \tag{5.10}$$

Here $\mathfrak{k}$ is the inhibition constant, ν_i'' is the observed rate under the competitive inhibition, and ε_i'' is the new error. Likewise, if we assume the empirical function $\lambda(S) = \lambda_0 + \lambda_1 S$, the Michaelis-Menten model can be modified as in the form

$$\nu_i = \frac{\nu_{\max} S_i}{k(\lambda_0 + \lambda_1 S_i) + S_i} + \varepsilon_i^{(1)}. \tag{5.11}$$

This is the untransformed model where the Michaelis constant now varies across the substrate concentration levels. Even though there is no anticipation of an influential inhibition effect, (5.11) is analogous to the above Michaelis-Menten model under the

competitive inhibition. If all the error assumptions are satisfied, the fitted model (5.11) can diminish the lack of fit incurred in model (5.6) and thus lead to more accurate NLS estimation of the unknown parameters.

In reference to our conjectures in Section 5.4, we assume the response error to increase with the initial substrate concentration. Of course, the correction factor can be a linear function of second order or above. In that situation, a similar modification can be adopted and it will contribute to a suitable model. Besides, the function can be even more complex since it should be defined as a discontinuous step function in fact.

Both scenarios we discussed (for models (5.8) and (5.11)) are idealistic as both functions of the correction factor are continuous and linear. When the correction factor is a constant or a first-order function in terms of the substrate concentration, the Michaelis-Menten model can be modified to accommodate the lack of fit and also meet the NLS assumptions. However, it is perhaps more sensible to assume the correction factor to be unknown or, if known, a discrete step function across different levels of the substrate concentration. In that case, it is difficult (if not impossible) to find a simple modified version of the Michaelis-Menten model. It is also unavoidable that the error structure of the assumed model is not most suitable and thus the nonlinear least squares estimation has to suffer from the lack of fit.

The good news is that we can take another measure to create a better error structure. In addition to using the reciprocal of the initial rate, there are different kinds of nonlinear transformations for the model. Here we will not discuss the model modification on the basis of a correction factor function further, so let us come back to the simple Michaelis-Menten model (5.5). As the model function is a mechanistic equation, we should transform the both sides of this model in order to influence the error distribution.

Response transformation was recommended in Box and Cox (1964), which can be an approach to stabilise and normalise the observed distribution of model errors with respect to the fitted responses or one of the controlled variables. The reciprocal of the response in model (5.7) is a special case of the classical Box-Cox transformation below:

$$\nu^{(\alpha)} = \begin{cases} \frac{\nu^{\alpha}-1}{\alpha} & \alpha \neq 0; \\ \log(\nu) & \alpha = 0. \end{cases}$$

A simplified form of the Box-Cox transformation is the nonlinear power transformation, which can lead to similar results if we transform both the response and the whole model function on the right. We can use different values for the Box-Cox transformation parameter α, such that it is viable to adapt the error structure of the fitted model. Besides, we can also derive the likelihood function of the transformed model in terms of α, in order to determine the transformation parameter with maximum likelihood estimation. Evidence can be found to support the use of Box-Cox transformations of the response. For instance, Horwitz (1982) derived an empirical equation to show

the common link between the expected response of a chemical measurement and the variance of the untransformed observed response. When this Horwitz curve is considered for a specified model, the Box-Cox transformation could make the error variance of the transformed observed response more constant and less dependent on the expected response (Atkinson, 2003).

When the statistical model we assume is purely empirical and linear, we can do most kinds of variable transformations as appropriate. However, if we transform the initial rate of (5.5), it is essential to do the same transformation on the mechanistic model function. Otherwise, the fundamental Michaelis-Menten equation will not hold and the transformed model will be an incorrect interpretation of the true mechanism. Hence, the Box-Cox transformation with the same value of α should be exerted on both sides of the Michaelis-Menten model. In accordance with the structure of the lack of fit, the best α should be found if one wishes to improve the overall fitting of the model and the NLS estimation of the parameters. For instance, if the correction factor $\lambda(S)$ is an unknown second-order linear function in terms of the substrate concentration, the Box-Cox transformation with $\alpha = \pm 2$ is a possible option to reduce the error dispersion of the Michaelis-Menten model. The sign of α depends on the unknown parameters of the correction factor function.

The transformation parameter α is a real number which can also be treated as an unknown parameter of the transform-both-sides (TBS) model. As we would introduce with two examples later, it can be estimated with either maximum likelihood or a recent ANOVA method from Latif and Gilmour (2015). As the model transformation will improve the raw error structure, the transform-both-sides Michaelis-Menten model is

$$\nu_i^{(\alpha)} = \left(\frac{\nu_{\max} S_i}{k + S_i}\right)^{(\alpha)} + \varepsilon_i \quad \text{at time } t \to 0, \tag{5.12}$$

where the new ε are assumed to be independent, with zero mean and constant variances. The classes of TBS models were first introduced in Carroll and Ruppert (1984) and later applied to simulation studies of simplest Michaelis-Menten kinetics (Ruppert et al., 1989). As it enables a Box-Cox transformation before the implement of nonlinear least squares, the TBS model can weaken the dependence of model errors on the expected mean of the response and thus stabilise the variation in the observed initial rates. This model can also reduce the lack of fit under NLS so that we would be able to find more accurate parameter estimates. If experimenters use maximum likelihood instead of nonlinear least squares, the classical assumption is $\varepsilon_i \sim \mathbb{N}(0, \sigma^2)$ for $i = 1, 2, \ldots, n$, where σ^2 is the constant error variance. In that case, the TBS model can reduce the skewness of the observed responses from a normal distribution too.

The Box-Cox transformation is demonstrated to be successful in various practical scenarios. For instance, if we let $\alpha = 2$, for $i = 1, 2, \ldots, n$, the ith measurement error ε_i of the fitted TBS model can be explained with a quadratic function in terms of the raw error

of the fitted untransformed model. A reasonable Box-Cox transformation can therefore adapt the error structure of the model to meet the NLS assumptions to some admissible extent. As such, we can stick to NLS rather than referring to iterative weighted NLS to deal with a heterogeneous error structure.

The weighted least squares estimation also depends on experimenters' conjecture of the error structure of the model, which is unknown in practice. Hence, there are some risks to take as we must assign correct weights to different experimental observations. In a theoretical example in Cornish-Bowden (2004, chap.14), the coefficients of variation are assumed to be identical across the experimental observations. This means that there is a multiplicative error structure for the Michaelis-Menten model

$$\nu_i = \left(\frac{\nu_{\max} S_i}{k + S_i}\right)(1 + \varepsilon_i).$$

The term $\varepsilon_i \sim \mathbb{N}(0, \sigma^2)$ is associated with the substrate concentration. Instead of weighted nonlinear least squares, let us consider the Box-Cox transformation with $\alpha = 0$. With the natural logarithmic scale for the transformation, the TBS model is written as

$$\log(\nu_i) = \log(\nu_{\max} S_i) + \log(k + S_i) + \log(1 + \varepsilon_i).$$

At first sight, there is insufficient information about the most suitable distribution of the error $\log(1 + \varepsilon_i)$ of the above model. However, the mean of the errors could be zero in the fitted model and after the Box-Cox transformation, these measurement errors will be independent of the initial substrate concentrations. Under NLS, now it is much more reasonable to assume the error variance to be constant.

5.6 Empirical Kinetic Model and Box-Cox Transformation

The Michaelis-Menten equation elaborates the influence of the substrate concentration on the (initial) rate of reaction. In the experimental environment, the discrete levels of the substrate concentration must be specified and so must other influential factors. These factors are fixed when we derive the Michaelis-Menten model. For most experimental studies of mammalian enzymes, for instance, it is common to imitate a near physiological environment: i.e. with the pH at 7.2, temperature at 37 °C, and the ionic strength at 0.15 mol l^{-1} in the solution (Cornish-Bowden, 2004, chap.3). At times, there could be more than one suitable level for these factors, when we are interested in more comprehensive studies of the mechanism. There are also some enzymes which can be modified and used in different experimental environments. This will require a careful consideration and interpretation of the reaction mechanism.

If we no longer fix the relevant experimental conditions, more than one controlled variable should be involved in the kinetic model. The Michaelis-Menten equation is inadequate when we require suitable models for such complicated experiments. In this case, the empirical modelling techniques can be useful when we should incorporate several controlled variables in the same statistical model. Meanwhile, this model can be related to the Michaelis-Menten equation, if one of the variables is the substrate concentration. The result will be what we call a *hybrid model* based on Michaelis-Menten kinetics.

To construct the hybrid model, for the ease of interpretation, we consider an empirical replacement of the theoretical maximum rate $\nu_{\max}$ with a simple function in terms of the controlled variables (except for the substrate concentration). When this simple function is linear and the controlled variable is the enzyme concentration, the hybrid model could be identical to the Michaelis-Menten model. The common influential factors, temperature T, the hydrogen ion concentration H, and the atmospheric pressure P were all discussed in Section 5.3. In some cases, experimenters could be interested in their roles in the reaction and their causal relationships with the (initial) rate. Nevertheless, not all information will be available from theories. If an appropriate theoretical equation cannot be derived or integrated into the Michaelis-Menten equation, experimenters can think about an empirical hybrid model, which is often nonlinear.

A controlled variable can be either continuous or categorical, depending on experimenters sometimes. One of the common experimental purposes is to make multiple comparisons between different potential reagents (i.e. substrate species). This could require some treatment contrasts and parameter estimation, when relevant experimental conditions or controlled variables are under control. In this case, we can use a categorical variable to represent the labels of different substrate species, for instance. This variable will be integrated into the Michaelis-Menten model and so will the three continuous controlled variables T, H, P. Denote $\mathfrak{L}(T, H, P)$ as an empirical linear function in terms of the continuous variables. Thus a hybrid model based on Michaelis-Menten kinetics can be

$$\mathbb{E}(\nu_i) = \beta_i + \frac{\mathfrak{L}(T_i, H_i, P_i)S_i}{k + S_i} \text{ or } \frac{\beta_i + \mathfrak{L}(T_i, H_i, P_i)S_i}{k + S_i} \text{ or } \beta_i \frac{\mathfrak{L}(T_i, H_i, P_i)S_i}{k + S_i}. \tag{5.13}$$

Here β_i indicates the ith fixed effect, for $i = 1, 2, \ldots, n$. The three candidate models above are simple in the sense that β_i is the sole categorical effect. In particular, if we assume β_i to be an additive constant in the model, the effect of the substrate species will be independent of the other controlled variables. To describe real mechanisms, however, the candidate models in (5.13) can be more complex since we should consider the interaction and correlation between the variables.

When we consider an exponential transformation of the linear function $\mathfrak{L}(T, H, P)$, the updated model below could be a suitable description of the expected initial rate:

$$\mathbb{E}(\nu_i) = \beta_i \frac{\exp(\mathfrak{L})S_i}{k + S_i} = \frac{\exp(\beta_i' + \mathfrak{L}(T_i, H_i, P_i))S_i}{k + S_i} \text{ or } \frac{\exp(\beta_i \mathfrak{L}(T_i, H_i, P_i))S_i}{k + S_i}, \tag{5.14}$$

where $\beta = \exp(\beta')$ in the first expression. At times, it is useful to transform both sides of the model. For instance, we assume an informative error structure in the model

$$\nu_i = \frac{\exp(\beta_i' + \mathfrak{L}(T_i, H_i, P_i) + \varepsilon^{(1)}{}_i)S_i}{k + S_i} \times e^{\varepsilon^{(2)}{}_i}, \tag{5.15}$$

where $\varepsilon^{(1)}{}_i \sim \mathbb{N}(0, \sigma_1^2), \varepsilon^{(2)} \sim \mathbb{N}(0, \sigma_2^2)$. Note that the normal distributions of the errors are not essential under nonlinear least squares but the standard deviations σ_1 and σ_2 should be constant in that case. If we assume the above error structure, the simple natural logarithmic transformation will lead to a better model

$$\log(\nu_i) = \log(\frac{S_i}{k + S_i}) + \beta_i' + \mathfrak{L}(T_i, H_i, P_i) + \varepsilon_i, \tag{5.16}$$

where $\varepsilon_i = \varepsilon^{(1)}{}_i + \varepsilon^{(2)}{}_i \sim \mathbb{N}(0, \sigma_1^2 + \sigma_2^2)$. The TBS model satisfies the NLS assumptions on errors and it is linear in all controlled variables except for the substrate concentration.

We use an empirical function (e.g. $\mathfrak{L}(T_i, H_i, P_i)$) to replace the maximum rate $\nu_{\max}$, where the controlled variables involved can be scaled before we include them in a hybrid nonlinear model. A charming aspect is that there could be some mechanistic information to be made use of, when we are to scale or transform these variables. For instance, we learn from the Arrhenius equation that

$$\nu_{\max} \propto A\exp\left(-\frac{1}{T}E_{\mathrm{a}}\right).$$

This relationship hints that an inverse transformation (i.e. $1/T$) can be adopted for the exponential function, where T represents absolute temperature. Notice that the two constants, A and E_{a}, could depend on the other controlled variables in the model. As we should ensure $\nu_{\max}$ to be nonnegative all the time, to some extent, it is sensible to replace this parameter with the exponential function in the model.

Models (5.13)-(5.14) can lead to simple interpretations, even though each of them is nonlinear and written in terms of several controlled variables. When the substrate concentration is fixed, the hybrid model will become linear or exponential. The Michaelis constant k is the ratio of the rate constants and it can be substituted with an empirical function too. In that case, there should be more parameters in the function and these parameters could be a bit redundant in the overall model. In comparison, it is easier to use and popularise the models in (5.13)-(5.14).

Now that the hybrid model treats k as a constant, it will not be the same as the Michaelis constant we wish to estimate. This is because, most of the time, we should not consider the intermediate concentration to be independent of the controlled variables in the empirical functions. However, the drawback is unavoidable when we have to interpret a complicated Michaelis-Menten mechanism of multiple factors.

The measurement errors of a hybrid model will depend on the assumed empirical function as well as the Michaelis-Menten equation. To stabilise the response errors and obtain accurate NLS estimates of the parameters, we shall emphasise the usefulness of the nonlinear Box-Cox transformation. On the basis of model (5.14), for instance, a transform-both-sides hybrid model is

$$\nu_i^{(\alpha)} = \left(\frac{\exp(\beta_i' + \mathfrak{L}(T_i, H_i, P_i))S_i}{k + S_i} \right)^{(\alpha)} + \varepsilon_i. \tag{5.17}$$

After a suitable Box-Cox transformation, we assume ε_i to be independent and its variance to be constant across different observations. To summarise, the above model for complex Michaelis-Menten mechanisms could have the following merits:

1. It is flexible: unlike mechanistic models of which the expressions are deterministic, we could determine the empirical component of a hybrid model ourself and even transform the controlled variables (except for the substrate concentration). It allows us to make full use of the mechanistic information we know, in order to derive the best form of the hybrid model. We can also add or delete some fixed or/and even random effects in the empirical function of the model, so there could be different reliable options in model selection.

2. It is comprehensive: the hybrid model involves more than one controlled variable from a more complicated Michaelis-Menten mechanism. Therefore, we do not have to fix some of these controlled variables to learn several simplified mechanisms. This means that the observations from even different experiments can be combined to compose an overall dataset. As such, experimenters can make the most of the collected and combined data to fit the assumed model, which is an economic and efficient approach in the use of experimental resources. The fitted model implies the obtained information about the comprehensive Michaelis-Menten mechanism. This saves our time in comparison of different treatments and search for the optimal combination of the experimental conditions for the reaction.

3. It is convenient: the fitted hybrid model can be used to describe a complicated mechanism and it offers a reliable approximation to the response surface. It does not require too much information or relevant theories. Since the model can be written in a simple nonlinear form, it will not be hard to interpret and validate the results of nonlinear least squares estimation.

4. It is universal: the hybrid model is not limited to the applications to Michaelis-Menten kinetics, which we have demonstrated in this thesis. Generalisation of the transform-both-sides model to widespread areas of scientific research is promising, where similar ideas can be adopted to derive suitable statistical models.

5.7 Optimal Design of Experiments

We have established the transform-both-sides model to be fitted to the dataset of n experimental observations. With a suitable Box-Cox transformation, we expect some improvement in the nonlinear least squares estimation of the parameters. In addition, before the experiment is run and the response data are collected, it is sensible to consider a Bayesian optimal design of the upcoming experiment. We now assume that the main experimental purpose is to minimise the standard errors of the estimates of the p treatment parameters in $\boldsymbol{\theta}$, which is the most common purpose in regression.

With this intention in mind, the classical optimal design approach is to use the local D-criterion which aims to maximise the criterion function $\phi_{\mathrm{D}} = \log|\mathbf{F}^{\mathrm{T}}\mathbf{F}|$. Here, $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ is the Fisher information matrix which is associated with the $p \times p$ variance-covariance matrix of the NLS estimator $\hat{\boldsymbol{\theta}}$. We write the nonlinear model for the observed response as $Y_i = f(\mathbf{X}_i, \boldsymbol{\theta}) + \varepsilon_i$, for $i = 1, 2, \ldots, n$. Here, $\mathbf{X}_i$ represents the ith experimental run which is one row of the specific experimental design $\mathbf{X}$, and ε_i is the error of the ith response of the model. With a first-order linearisation of the model function above, we can derive the $n \times p$ design matrix

$$\mathbf{F} = \frac{\partial f(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

which depends on the numeric vector of the parameter prior $\boldsymbol{\theta}^0$. As an alternative to using the local D-criterion (of optimality), following Chapter 4, we continue to use the so-called pseudo-Bayesian approach. Rather than a deterministic numeric vector $\boldsymbol{\theta}^0$, a realistic multivariate parameter prior distribution of $\boldsymbol{\theta}$ should be assumed. This can describe the uncertainties of the true model parameter values, which are then taken into account in Bayesian optimal design. While the joint density function of the parameter prior distribution is $\rho(\boldsymbol{\theta})$, instead of the local D-criterion function ϕ_{D}, the expected D-criterion aims to maximise the multiple integral

$$\varphi = \int_{\boldsymbol{\theta}} \phi_{\mathrm{D}}(\mathbf{X}|\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

We should find an optimal $\mathbf{X}$, a $n \times v$ matrix, to maximise the criterion function φ, where v indicates the number of controlled variables. To achieve this, we develop and implement the new *hybrid exchange algorithm* for the iterative search of optimal candidates to update the individual points or coordinates of a specific $\mathbf{X}$. In addition, the *modified adjustment algorithm* can be applied to increase the expected D-criterion function further, before we output the final solution (as the Bayesian optimal design).

The computation can be quite intensive in this case. Hence, to evaluate the criterion function φ above, we shall use the Gaussian quadrature rules. As one of these rules, the Gauss-Hermite quadrature, presented in Chapter 4, is an efficient numerical integration

tool that can be adapted to approximate the value of φ. Most of the time, it requires the parameter prior density $\rho(\boldsymbol{\theta})$ to be multivariate normal or at least lognormal. A special case is, if the covariances of this distribution are all zero, the normal prior and the lognormal prior can be combined in the same multivariate distribution, which has been demonstrated in Chapter 4 too. Otherwise, some alternative Gaussian quadrature rules were introduced in Goos and Mylona (2013), which can fit some different expected criterion functions and their respective assumptions of $\rho(\boldsymbol{\theta})$.

In the examples to be shown here, where the Gauss-Hermite quadrature is applied, we can approximate the expected D-criterion function φ with

$$\phi_{\mathrm{BD}} = \hat{\varphi} \approx \sum_{r=1}^{m} \omega^{(r)} \phi_{\mathrm{D}}(\mathbf{X}, \boldsymbol{\theta} = \boldsymbol{\theta}^r) \text{ with appropriate constants } m_1, m_2, \ldots, m_p. \quad (5.18)$$

As one unit of the Gauss-Hermite sample we draw, $\boldsymbol{\theta}^r$ is a unique parameter prior vector whereas $\omega^{(r)}$ is the associated composite weight, for $r = 1, 2, \ldots, m$. Here, m is the total Gauss-Hermite sample size and $m_1, m_2, \ldots, m_p$ are the subsample sizes in correspondence to each parameter to be integrated out of φ. As we showed in detail in Chapter 4, the Gaussian quadrature rules are quite feasible as we are allowed to determine each subsample size in order to achieve a reliable approximation overall. Hence, in each example, a quick numerical investigation can help us to decide how to construct a suitable Gauss-Hermite quadrature ϕ_{BD}. We can then search for the Bayesian D-optimal design to maximise the information to be obtained from the future observed data, when the experimental purpose is to estimate the model parameters $\boldsymbol{\theta}$.

5.8 A Simple Hybrid Nonlinear Model Example

5.8.1 Estimation of Box-Cox Transformation Parameter

We now look at an example adapted from an experiment in Martins et al. (1999). First, the reaction involves two controlled variables: the substrate concentration S and the protein concentration E. The observed response is the initial rate ν in the experiment, so Michaelis-Menten kinetics should hold even though there is an additional variable to control. We have to do an empirical approximation in this case and therefore, a hybrid nonlinear model can be used to integrate all the information and to approximate the true response surface. As the nonlinear least squares estimation requires the error variance to be uncorrelated and thus homogeneous, for the experiment, we can assume a hybrid model based on Michaelis-Menten kinetics:

$$\nu_i = \frac{\exp(a_0 + a_1 x_i + a_2 {x_i}^2) S_i}{k + S_i} + \varepsilon_i, \quad (5.19)$$

where $x_i = (E_i - 0.07)/0.05 \in [-1, 1]$ and $S \in [0.15, 3]$. This is the same kinetic model we discussed in Chapters 3 and 4. Within the continuous variable space $\mathcal{X}$, any combination of the levels of the two factors can be used as the experimental input, though the closest distance is decided to be 0.001 for the protein concentration and 0.01 for the substrate concentration in this example. When the kinetic model (5.19) is fitted to the approximate data in Martins et al. (1999), the adjusted coefficient of determination is found to be 0.994 (so this model fits the reference data well at least in some aspects) and the nonlinear least squares estimate is $\tilde{\boldsymbol{\theta}} = \{\tilde{k}, \tilde{a}_0, \tilde{a}_1, \tilde{a}_2\} = \{0.3122, -6.4086, 0.8383, -0.2861\}$. In Chapter 3, this reference estimate was chosen as the point parameter prior $\boldsymbol{\theta}^0$ of both the local D-criterion and the WA-criteria for model (5.19).

Here, we wish to transform both the response and the model function of (5.19). As such, we expect to find a more reliable estimate of the unknown $\boldsymbol{\theta}$ and a more reasonable error structure for regression. As α denotes an additional parameter from the Box-Cox transformation, the transform-both-sides kinetic model can be written as

$$\nu_i^{(\alpha)} = \left(\frac{\exp(a_0 + a_1 x_i + a_2 {x_i}^2) S_i}{k + S_i} \right)^{(\alpha)} + \varepsilon_i, \tag{5.20}$$

where the least squares assumptions of ε_i should be satisfied also to minimise the error of the parameter estimator $\hat{\boldsymbol{\theta}}$. To determine α over its sampling space, the traditional maximum likelihood method is to treat it as an unknown nontreatment parameter to be estimated. This is difficult as the transformed response $\nu^{(\alpha)}$ depends on one parameter. Hence, Latif and Gilmour (2015) introduced a new ANOVA (analysis of variance) method which separates the estimation of α from the rest. As such, one will estimate $\boldsymbol{\theta}$ conditional on a constant α. With this ANOVA method, the computation is simplified and $\hat{\boldsymbol{\theta}}$ shall be independent of the estimation of α for the Box-Cox transformation.

The previous data in Martins et al. (1999) can also be used to determine the transformation parameter α. It is convenient to make its estimator to be independent of the model function we assume. If β_{i^*} indicates the treatment effect corresponding to one of the n^* unique experimental runs, the following full treatment model can be fitted with maximum likelihood estimation:

$$\nu_i^{(\alpha)} = \mu_0 + \beta_{i^*} + \epsilon_i, \tag{5.21}$$

for $i^* = 1, 2, \ldots, n^*$. Herein, μ_0 is the mean of the responses and ϵ is the error. The estimation of μ_0 can be ignored when we fit the model, unless we set one of the treatment effect to be the baseline. As the total error ϵ depends on the value of α we choose, it makes no sense to minimise the sum of squares of the errors. Besides, the (adjusted) coefficient of determination is now less important in model comparison since the least squares cannot be adopted to fit the model.

As we shall use maximum likelihood to estimate the unknown parameter α, the assumption is $\epsilon \sim \mathbb{N}(0, \varsigma^2)$ as we require the error distribution to be normal. The lack of fit has been eliminated in the full treatment model, so the total error ϵ is identical to the pure error. In this situation, the log likelihood function in terms of α conditional on the observed responses ν can be expressed as

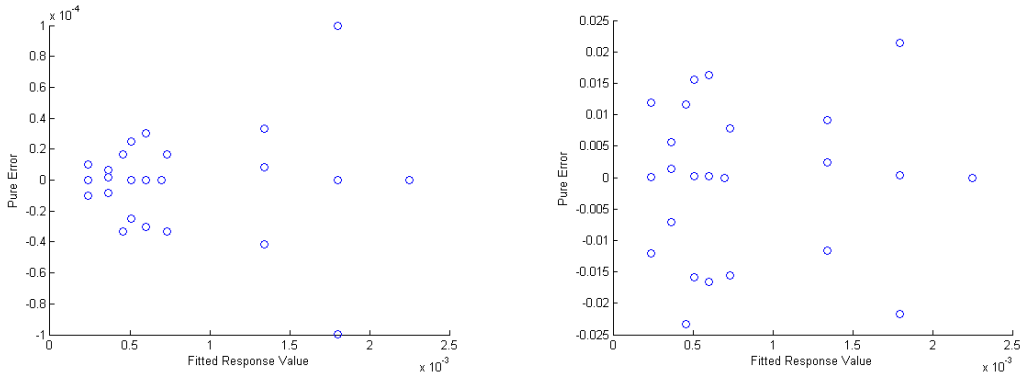
$$l(\alpha) = -n\log\left(\frac{\hat{\varsigma}(\alpha)}{\prod \nu_i^{\frac{\alpha-1}{n}}}\right) - \frac{n}{2} = (\alpha-1)\sum \log\nu_i - n\log\hat{\varsigma}(\alpha) - \frac{n}{2},$$

where $\hat{\varsigma}(\alpha)$ is the maximum likelihood estimator (MLE) of the standard deviation ς. The above function $l(\alpha)$ does not show those treatment effects β_{i^*} since we focus on α. Moreover, $\hat{\varsigma}^2(\alpha) = \sum \epsilon_i^2/n$ is the mean of the squares of the pure errors from the full treatment model. In the reference dataset, there are three replicates on each of the $n^* = 10$ support points, so there is a large number of degrees of freedom for the errors. As the function $l(\alpha)$ is to be maximised with the Nelder-Mead method, the transformation parameter estimate is equal to $0.1496 \approx 0.15$. This value has to be recalculated after the new experiment but at the moment, we can use it for the TBS model (5.20).

Nonetheless, the model transformation does not affect the functional relationship between the response and the controlled variables. If we treat α as a constant, the optimal design of the experiment is often quite robust to minor corrections of α. Otherwise, if we find a less reliable estimate of α using the reference dataset, an alternative is to consider it as an important unknown parameter in optimal design. In this case, one of our aims is to maximise the Fisher information of model (5.20) with respect to α. As such, a compound criterion can be used for the precise estimation of α as well as the treatment parameters $\boldsymbol{\theta}$ in the model (Bogacka et al., 2015). In this case, the expected cost is to estimate one more parameter using the new experimental data, which will use some of the resources if the optimal design takes that into account.

With a transformation of both the response and the model function, we expect to see better model fitting after the modification of the raw error structure of the untransformed model. In Figure 5.1, we show two residual plots obtained before and after we transform the full treatment model (5.21). As the Box-Cox transformation is applied with $\alpha = 0.15$, it is now more reasonable to assume the pure errors to be homogeneous and to fluctuate around the zero level. Since the essential error assumptions will be better satisfied, we are able to obtain more accurate parameter estimates from the TBS hybrid model.

We wish to reduce the bias of the NLS estimate $\{0.3122, -6.4086, 0.8383, -0.2861\}$ obtained from the untransformed model (5.19). As the substitution $\alpha = 0.15$ is made in model (5.20), the new parameter estimate is $\tilde{\boldsymbol{\theta}} = \{0.2859, -6.4368, 0.8408, -0.2607\}$ when the error distribution is not assumed to be normal. The two estimates are similar but this time, we can see a different k estimate, for instance. This coincides with



(a) (b)

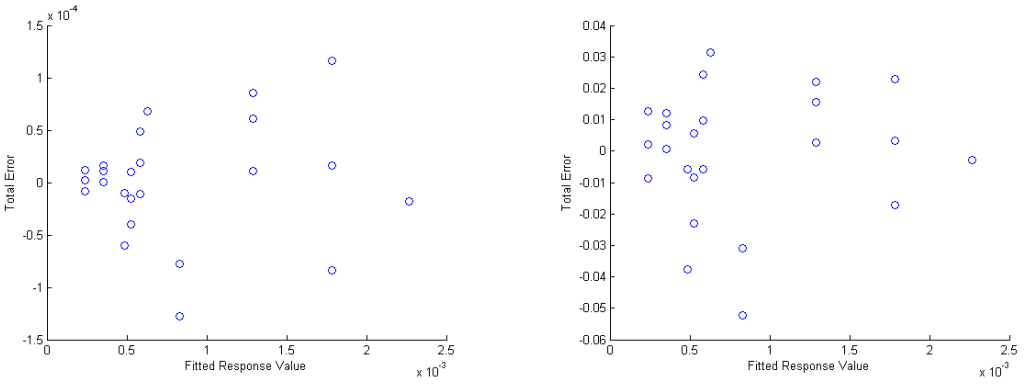
Figure 5.1: Residuals Against Fitted Response Values Under: (a) the Untransformed Full Treatment Model ($\alpha = 1$); (b) the Full Treatment Model

our presumption that the estimate of the Michaelis constant from the simple Michaelis-Menten model could be inaccurate, though the Box-Cox transformation has adjusted the error structure in this case. Unless the untransformed model is quite bad, a Box-Cox transformation will not change too much $\tilde{\boldsymbol{\theta}}$, the NLS estimates which are supposed to be near the true parameter values. This means that some possible values of α (which will lead to quite different estimates) cannot be approved for the transformation in the TBS model, even if the variation in the transformed response will be well explained.

5.8.2 Lack of Fit and Pure Error

A new method we propose in this thesis is to use the fitted model (5.20), which we assume for the experiment, to determine the value of the transformation parameter α. Different from computation under the full treatment model, the value of α in this case has to be updated in iterations until little further improvement can be made on model residuals. We should start with the untransformed model (i.e. $\alpha = 1$) and replace the vector $\boldsymbol{\theta}$ with the current nonlinear least squares estimate $\tilde{\boldsymbol{\theta}} = \{0.3122, -6.4086, 0.8383, -0.2861\}$. As such, we can compute the fitted responses $\tilde{\nu}$ as well as the residuals $\nu - \tilde{\nu}$. With the lack of fit taken into account, the total error variance is $\hat{\varsigma}^2(\alpha) = \hat{\sigma}^2(\alpha) = \sum \varepsilon_i^2/n$. This allows us to maximise the log likelihood function $l(\alpha)$ and then estimate α. With the residuals from the fitted untransformed model, the initial estimate is $\alpha = 0.3646$.

Given the transformation parameter, we are able to update and refit the model. The treatment parameter estimate is found to be $\tilde{\boldsymbol{\theta}} = \{0.2897, -6.4308, 0.8394, -0.2676\}$. Given the new $\tilde{\boldsymbol{\theta}}$, we can repeat the above steps for iterative estimation of α and $\boldsymbol{\theta}$. When the first four decimal digits are constant after six iterations, the final solutions are $\alpha = 0.1624$ and $\tilde{\boldsymbol{\theta}} = \{0.2861, -6.4365, 0.8407, -0.2610\}$. These results are similar to those found with the full treatment model and the corresponding two residual plots are shown in Figure 5.2. This is because of the small lack of fit on average. Under the



(a) (b)

Figure 5.2: Residuals Against Fitted Response Values Under: (a) the Untransformed Hybrid Kinetic Model); (b) the TBS Hybrid Kinetic Model

untransformed hybrid kinetic model, the error variances do not seem to be homogeneous. In this aspect, there is some clear improvement after we transform the model. Moreover, after the transformation of the both sides of the model, there is a bit improvement in the observed error distribution, which is assumed to be normal. Hence, if we fit the TBS model with NLS, we can expect more reliable parameter estimation and interpretation of the response variation.

There are some differences between the two methods in the determination of the Box-Cox transformation parameter. When the lack of fit is smaller on average, as we increase the number of empirical terms in the hybrid model, the difference between the two maximum likelihood estimates of α could also be smaller. However, that would make it difficult to fit the hybrid nonlinear model and it is also impractical to do so. For the full treatment model, the number of parameters is equal to the number of unique treatments (if we ignore the common intercept and exclude α) and therefore the lack of fit will be eliminated. While the ANOVA method in Latif and Gilmour (2015) makes the accurate estimation of α independent of the assumed nonlinear model, the alternative method we introduced above is even more similar to using maximum likelihood to fit the complete model. But it is easier than that and it will not involve the covariances between the estimators of α and the treatment parameters.

In summary, the benefit of transforming the model is that we will find more reliable parameter estimates and improve our understanding about the response surface. When the model we assume is mechanistic, our recommendation is to use the full treatment model to determine α. The ANOVA method shall utilise the pure errors, a quite natural Box-Cox transformation can be found to maximise the likelihood of the full treatment model, conditional on the observed responses. The parameters of mechanistic models often contain some scientific information to be interpreted, so it is important to obtain accurate estimates of the parameters from the mechanism we assume.

To achieve this, the better choice is to use the pure errors to define the Box-Cox transformation. With exclusion of the lack of fit which is expected to be small for a mechanistic model, α is there to improve the pure error structure. As such, the estimate of $\boldsymbol{\theta}$ could be considered to be **most accurate** if both the model and the postulated mechanism are correct. This is crucial in studies of reaction mechanisms and relevant theories.

In contrast, one's interest is often in a best description or approximation to the unknown mechanism when the model we assume is purely empirical (i.e. unrelated to theories). In this case, our recommendation is to use the total error to find the maximum likelihood estimate of α. After the subsequent estimation of the treatment parameters $\boldsymbol{\theta}$, the fitted model could lead to the best explanation of the response variation. This is because, in such kind of empirical studies, the lack of fit can reflect the model discrepancies from the true unknown mechanism, which we wish to minimise.

The experimental purpose is to approximate the response surface and the relationship between the response and the controlled variables, so it is acceptable to let the estimator of $\boldsymbol{\theta}$ depend on the lack of fit of the model. Given α, in this example, the hybrid model function can be used to evaluate the total error of the model, which combines the pure error and the lack of fit. We can then expect the estimates of the treatment parameters to be **more precise** (i.e. the standard errors will be smaller) than those obtained under the ANOVA method, since the maximum likelihood assumptions of the (total) errors shall be most reasonable in this case.

5.8.3 Optimal Experimental Design Results

For a precise estimation of the treatment parameters $\boldsymbol{\theta}$, in addition to transforming both sides of a nonlinear model, another measure is to consider optimal experimental design prior to the data collection. In this example, the first step is to decide α as we wish to transform the response ν. The hybrid nonlinear model in this case is a combination of the Michaelis-Menten equation and a simple exponential function. Though the lack of fit exists, it is rather small because the untransformed model (5.19) can fit the reference data quite well. As we can see previously, there is little difference between the two maximum likelihood estimates $\alpha = 0.1496$ and $\alpha = 0.1624$. Here, since the TBS model contains some characteristics of Michaelis-Menten kinetics, we choose the full treatment model to determine the Box-Cox transformation, for example. As a result, $\alpha \approx 0.15$ is fixed as a constant before we estimate $\boldsymbol{\theta}$ with nonlinear least squares. It is based on the pure errors from the reference dataset, so the new experiment could suggest a different Box-Cox transformation function in terms of α, depending on the fitted full treatment model. Given the uncertainty about the true value of α, even if the impact of the transformation is potentially minor on the parameter estimation, this is one of the main disadvantages when we do optimal design for a TBS model.

To define a multivariate parameter prior distribution, we set the prior mean of $\boldsymbol{\theta}$ to equal $\tilde{\boldsymbol{\theta}} = \{0.2859, -6.4368, 0.8408, -0.2607\}$. In the TBS model (5.20), α is fixed to be a constant 0.15 in our first demonstration below. This is not treated as an unknown parameter such that we can focus on the estimation of the four treatment parameters. However, it is still important to find a reliable prior value for α (0.15 in this case) beforehand and also calculate the maximum likelihood estimate of α after the new experiment, in order to determine the best transformation of the model. Besides, the Box-Cox transformation also influences the nonlinear behaviour of the TBS model. Now, as we learn from the previous example in Chapter 4, if we do not transform model (5.19) (i.e. $\alpha = 1$), the 30-run local D-optimal design can be obtained as

$$\mathbf{X}_0^* = \left\{ \begin{array}{ccccc} (0.025,\ 3) & (0.086,\ 3) & (0.097,\ 0.33) & (0.12,\ 0.26) & (0.12,\ 3) \\ 8 & 7 & 1 & 7 & 7 \end{array} \right\}. \tag{5.22}$$

This has been found to be efficient across various postulated scenarios. It is also robust to the defined moderate uncertainties of the true parameter values, which are exhibited in the joint prior density $\rho(\boldsymbol{\theta})$. This is because of the strong linearities of the untransformed model, which incorporates an empirical exponential function in terms of three unknown parameters. As such, though the value of φ varies depending on the parameter prior values, the specified expected D-criterion will not much influence the result in (5.22), even though it evaluates a complex function φ instead of the local D-criterion function.

After we transform both sides of model (5.19), it is sensible to reexamine the relative improvement of the new Bayesian D-optimal design and its robustness to the variation in the true parameter values. To do this, we assume an appropriate parameter prior distribution so that the pseudo-Bayesian approach can be applied to use the expected D-criterion. To approximate its criterion function φ and reduce computational time, we shall find a reliable Gauss-Hermite quadrature. It will be intuitive to calculate different Gauss-Hermite approximations (quadratures in different orders) of φ of a fixed $\mathbf{X}$ and compare them to the extrapolated true value of φ.

Recall that the $\mathbf{X}_{\text{ref}}$ in Table 5.1 is the reference from the previous experiment and the $\mathbf{X}_0^*$ shown in (5.22) is local D-optimal for the untransformed model (5.19). With each of them, we are able to compare the errors of the Gauss-Hermite quadratures in different orders as approximations to φ. To compose an economic and reliable Gauss-Hermite sample for the function ϕ_{BD}, we shall determine the four subsample sizes in correspondence to the respective treatment parameters. As such, the total Gauss-Hermite sample size m depends on each of the subsample sizes. Our intention is to find an acceptable tradeoff between the total sample size and the Gauss-Hermite approximation error.

In this example related to Michaelis-Menten kinetics, first, the univariate prior distributions of the treatment parameters are assumed to be normal for a_0, a_1, a_2 and lognormal for k. This is due to the constraint that k must take a nonnegative value. We can

Table 5.1: 30-Run Reference Design in Martins et al. (1999)

E	S	E	S	E	S	E	S	E	S	E	S
0.023	1.5	0.03	0.15	0.03	0.9	0.03	3	0.045	1.5	0.089	1.5
0.023	1.5	0.03	0.3	0.03	0.9	0.03	3	0.0675	1.5	0.089	1.5
0.023	1.5	0.03	0.3	0.03	1.5	0.03	3	0.0675	1.5	0.112	1.5
0.03	0.15	0.03	0.3	0.03	1.5	0.045	1.5	0.0675	1.5	0.112	1.5
0.03	0.15	0.03	0.9	0.03	1.5	0.045	1.5	0.089	1.5	0.112	1.5

further assume the mutual independence between the four univariate prior distributions to derive the multivariate $\rho(\boldsymbol{\theta})$. This implies that in the 4×4 variance-covariance matrix of $\rho(\boldsymbol{\theta})$, the prior covariances are all zero. The prior variances are set to be 10 times the variance estimates which are obtained from the model fitted to the reference dataset. Hence, the mean of $\rho(\boldsymbol{\theta})$ is $\{0.2859, -6.4368, 0.8408, -0.2607\}$ and the specified standard deviations are $\{0.0808, 0.0949, 0.0713, 0.1550\}$. Besides, after we transform both sides of the model, the estimation of k is much more precise but that of a_2 is less precise. The new parameter prior density $\rho(\boldsymbol{\theta})$ can be incorporated into the expected D-criterion function φ, which we will approximate with ϕ_{BD}. As we change the Gauss-Hermite subsample size(s) with respect to one or all (see the last column) treatment parameters, the numerical results are summarised in Table 5.2.

Table 5.2: A Selection of Gauss-Hermite Quadratures in Different Orders for the Approximation of the Expected D-Criterion Value

	Sample	$\{\text{k}\}$	$\{\text{a}_1\}$	$\{\text{a}_2\}$	$\{\text{k}, (\text{a}_0), \text{a}_1, \text{a}_2\}$
$\mathbf{X}_{\text{ref}}$	m_j=1	-15.0905	-15.0905	-15.0905	-15.0905
	m_j=2	-15.1215	-15.0905	-15.0905	-15.1213
	m_j=3	-15.1214	-15.0905	-15.0905	-15.1213
	m_j=4	-15.1214	-15.0905	-15.0905	-15.1213
$\mathbf{X}_0^*$	m_j=1	-13.1393	-13.1393	-13.1393	-13.1393
	m_j=2	-13.1700	-13.1393	-13.1392	-13.1699
	m_j=3	-13.1699	-13.1393	-13.1392	-13.1698
	m_j=4	-13.1699	-13.1393	-13.1392	-13.1699

The parameter value of a_0 does not influence the approximation error of φ, as was explained in Chapter 4. Thus we fix it at the point estimate $\tilde{a}_0 = -6.4368$. For three of the columns, we determine a specific Gauss-Hermite subsample size for one of the treatment parameters when the remaining subsample sizes are all fixed to one (i.e. the total size is $m = m_j$). Note that the parameter k is subject to a lognormal prior distribution. Therefore, even if the subsample size is one, the single prior value of k will not be equal to the mean of the prior distribution ($\tilde{k} = 0.2859$). For the results in the last column of Table 5.2, we determine an identical Gauss-Hermite subsample size for each of the three parameters (if we exclude a_0). As such, the total sample size is calculated to be $m = m_j^3 = 1, 8, 27, 64$ but this will not lead to the most appropriate Gauss-Hermite sample for efficient and economic approximation of φ.

An improvement in the numerical results is that, after the Box-Cox transformation of the both sides of the hybrid model, it becomes much easier to approximate the new expected D-criterion function. According to Table 5.2, we can choose to fix both prior values of a_1 and a_2 at the respective means $\tilde{a}_1$ and $\tilde{a}_2$. As such, the total Gauss-Hermite sample size is as small as $m = 3$ since we will use three sampled prior values of k. In addition to quicker computation, this result also implies the close-to-linear behaviour of the hybrid nonlinear model.

To configure our hybrid (point) exchange algorithm for Bayesian optimal design, we define a simple 3×3 candidate set $\boldsymbol{\Omega} = \{-1, 0, 1\} \times \{0.15, 1.5, 3\}$ (the first variable is the scaled protein concentration) and an exchanging rule $\max(d_i) > 1.01$ for the swift discrete optimisation of 30 random initial designs over $\boldsymbol{\Omega}$. Out of the $\tau = 30$ tries in the discrete optimisation, most of the solutions $\mathbf{X}_{1\text{st}}$ are identical at the attained local maxima of the expected D-criterion function. In total, we find $\tau^* = 4$ distinct $\mathbf{X}_{1\text{st}}$ that will act as the new starting designs for the continuous optimisation under the point exchange approach, which will be the second stage of the hybrid method.

The exchange rule is reset to $\max(d_i) > 1.00000001$ as we shall optimise each point of $\mathbf{X}_{1\text{st}}$ in iterations over the bounded variable space $\mathcal{X} = [-1, 1] \times [0.15, 3]$. As usual, we can utilise ten vectors of initial variable values where the continuous optimisation can start: the current point in $\mathbf{X}$ and nine fixed points from $\{-0.67, 0, 0.67\} \times \{0.5, 1.5, 2.5\}$. With the continuous optimisation, each of the four distinct $\mathbf{X}_{1\text{st}}$ will be updated further until we obtain a stable solution $\mathbf{X}_{2\text{nd}}$. The most efficient solution out of the four $\mathbf{X}_{2\text{nd}}$ will be called $\mathbf{X}_{\text{opt}}$, which could maximise φ.

The closest distance is set to be 0.001 for the protein concentration and 0.01 for the substrate concentration. As such, $\mathbf{X}_{\text{opt}}$ will be used to finish the hybrid exchange algorithm. Its corresponding points and numbers of replicates can be recalculated before we obtain a feasible optimal design $\mathbf{X}^*_{\text{opt}}$. Finally, the modified adjustment algorithm can be applied to validate the expected D-criterion and to compute the finalised Bayesian D-optimal design $\mathbf{X}^{**}_{\text{opt}}$. This is not an essential part though, since the $\mathbf{X}^{**}_{\text{opt}}$ below is identical to $\mathbf{X}^*_{\text{opt}}$ (no improvement is made) in this example:

$$\mathbf{X}^{**}_{\text{opt}} = \left\{ \begin{array}{cccccc} (0.02,\ 0.15) & (0.02,\ 3) & (0.074,\ 0.15) & (0.074,\ 3) & (0.12,\ 0.15) & (0.12,\ 3) \\ 3 & 6 & 4 & 6 & 5 & 6 \end{array} \right\}, \tag{5.23}$$

of which $\phi_{\text{BD}} = -11.9937$. As we can see, there are six support points which are similar to the discrete candidate points from the small 3×3 set $\boldsymbol{\Omega}$. We can also expect a similar result when the local D-criterion replaces the current expected D-criterion. This is because the $\mathbf{X}^{**}_{\text{opt}}$ in (5.23) is quite robust to variations of the true parameter values that we sample just one prior value for each of the treatment parameters except for k. To find an efficient optimal design in this case, we do not have to do many random tries in the continuous optimisation and even a simple discrete optimisation alone could be sufficient for the exchange algorithm to be implemented. This is another benefit of using

a suitable Box-Cox transformation on the model, which can be found to improve the untransformed error structure of the full treatment model.

Moreover, the six support points in (5.23) are almost identical to those in the previous local D-optimal design we find for the TBS model when $\alpha \to 0$ (i.e. the natural logarithmic scale is used for the transformation). However, the numbers of replicates are different, which could depend on the transformation parameter value α in this case.

5.9 A Kinetic Model with a Categorical Variable

We now consider the experiment in Yamamoto (1958) that studied a Michaelis-Menten mechanism with three controlled variables: the dye (or substrate) concentration $S \in [0.02, 0.2]$, the hydrogen ion concentration $H \in [7, 8]$ (i.e. the pH), and the dye species $D \in \{0, 1\}$. The aim is to understand the photodynamic action of the dyes in inactivation of the bacteriophages, the kinetics of which is important in disease therapies and medical treatments. The dummy variable $D = 0$ indicates the use of methylene blue dye for the reaction whereas $D = 1$ indicates the use of toluidine blue dye. Hence, we can consider a transform-both-sides hybrid nonlinear model as follows:

$$\nu_i^{(\alpha)} = \left(\frac{\exp(a_0 + a_1 D_i + a_2 x_i + a_3 x_i D_i + a_4 {x_i}^2 + a_5 {x_i}^2 D_i) S_i}{k_0 + k_1 D_i + S_i} \right)^{(\alpha)} + \varepsilon_i, \tag{5.24}$$

where $x = (H - 7.4)/0.4$ and the $p = 8$ treatment parameters are $k_0, k_1, a_0, \ldots, a_5$. The reference experiment included 24 valid runs in Yamamoto (1958) (the initial rates were observed under the same conditions): there were 19 observations for the methylene blue and five for the toluidine blue which was the less important dye. The problem is, in total, that there are 17 support points. Thus we can only use seven replicate runs for an evaluation of the pure error variance of the model.

The total errors of the fitted TBS model can be calculated whereas the lack of fit can be larger than that in the previous example. If we use the TBS model above to estimate the Box-Cox transformation parameter α, the final answer is $\tilde{\alpha} = 1.4460$. The reference dataset does not allow for the estimation of a_3 and a_5, so we fix them to be zeros in the current model fitting. As a consequence of the model transformation, the total errors are now better explained. Nevertheless, at least the assumption of homogeneous error variances is still not quite adequate.

In contrast, we use the pure errors to estimate the α from the full treatment model. The solution $\tilde{\alpha} = 0.8202 \approx 0.82$ recommends a different form of the Box-Cox transformation and this requests a different kind of error structure correction. Given the current small number of replicate runs, this calculation is not so reliable in a sense. A more accurate estimation of α will require sufficient replication of the adopted support points, which

is often encouraged in design of experiments. In this situation, one might find the DP-criterion (Gilmour and Trinca, 2012b) useful in place of the classical D-criterion.

It is no wonder this form of the Box-Cox transformation does not improve much the model fitting since there are not many nonzero observations of the pure errors. Without a much better explanation of model errors, the optimal design might not be simplified too. To be consistent with the first example in this chapter, however, α is determined to be a known constant 0.82 in this case, so that we could obtain the NLS estimate $\tilde{\boldsymbol{\theta}} = \{0.11281, -0.044306, 0.32276, -0.67747, 0.31409, 0, -0.10768, 0\}$ of all the treatment parameters, where we set the two unavailable estimates $\tilde{a}_3, \tilde{a}_5$ to zero.

In comparison, we have more parameters and less experimental units in this example. The prior distribution of $\boldsymbol{\theta}$ can be specified such that the prior variances are set to be five times the reference variance estimates. The prior mean shall equal $\tilde{\boldsymbol{\theta}}$ and the prior standard deviations of $\{k_0, a_0, \ldots, a_5\}$ are $\{0.0541, 0.0976, 0.0855, 0.1009, 0, 0.5484, 0\}$. An exception is that the standard deviation of k_1 is set to zero such that the prior value of k_1 will not be influenced under our pseudo-Bayesian approach. We assume normal distributions for $a_0, \ldots, a_5$ and a lognormal distribution for $k_0 - \tilde{k}_1$. This is because the function $k_0 + k_1$ cannot take nonpositive values (this is a constraint under the Michaelis-Menten mechanism) and $\tilde{k}_1 = -0.044306$ is a fixed parameter prior value.

There are some alternative measures to meet this constraint but most of the required adaptations would much affect the parameter prior distribution. One approach is to take the covariance estimates of $\tilde{k}_0$ and $\tilde{k}_1$ into account in a bivariate lognormal prior distribution. Another approach is to reparametrise k_0 and k_1 when we assume the corresponding parameter prior distribution. For instance, $k_0 + k_1 D_i$ can be rewritten as $\exp(k_0' + k_1' D_i)$, for $i = 1, 2, \ldots, n$. Nonetheless, the results of optimal design of experiments can change under either approach, which is undesirable if experimenters are interested in the parameter estimation of model (5.24) only.

As we can see in Table 5.3 below, the emphasis of a suitable Gauss-Hermite quadrature will be all related to the experimental units with the allocation $D = 0$: i.e. when methylene blue is the dye species for the reaction. These approximations are on the basis of the local D-optimal design

$$\mathbf{X}_0^{**} = \left\{ \begin{array}{ccccc} (0,\ 7,\ 0.2) & (0,\ 7.6,\ 0.2) & (0,\ 7.7,\ 0.06) & (0,\ 8,\ 0.06) & (0,\ 8,\ 0.2) \\ 3 & 3 & 1 & 2 & 3 \\ (1,\ 7,\ 0.2) & (1,\ 7.6,\ 0.2) & (1,\ 7.7,\ 0.05) & (1,\ 8,\ 0.04) & (1,\ 8,\ 0.2) \\ 3 & 3 & 1 & 2 & 3 \end{array} \right\} \tag{5.25}$$

for the untransformed model hybrid in an example in Chapter 3. For the results in each column of Table 5.3, we take one parameter at a time to draw its prior value such that the Gauss-Hermite sample size is $m = m_j$. To ensure an accurate approximation of the expected D-criterion value for model (5.24), which is transformed when $\alpha = 0.82$, we

choose the Gauss-Hermite sample that includes four prior values of k_1, two of a_2, and also two of a_4. The rest of the parameter prior values will take the respective point estimates, so there is no need to integrate them out of φ. Although there are eight treatment parameters, the determined Gauss-Hermite sample size is as small as $m = 16$.

Table 5.3: A Selection of Gauss-Hermite Quadratures in Different Orders for the Approximation of the Expected D-Criterion Value

Sample	$\{k_1\}$	$\{a_1\}$	$\{a_2\}$	$\{a_4\}$
m_j=1	10.8858	10.8858	10.8858	10.8858
m_j=2	10.0287	10.8858	10.8892	10.8983
m_j=3	10.0399	10.8858	10.8892	10.8983
m_j=4	10.0400	10.8858	10.8892	10.8983
m_j=5	10.0400	10.8858	10.8892	10.8983
m_j=6	10.0400	10.8858	10.8892	10.8983

For the discrete optimisation part (first stage) of the hybrid exchange algorithm, the candidate set for each point (D_i, x_i, S_i) is $\mathbf{\Omega} = \{0, 1\} \times \{-1, 0, 1.5\} \times \{1/48, 1/18, 1/6\}$ and the exchanging rule is $\max(d_i) > 1.05$. Out of $\tau = 30$ random tries, we find $\tau^* = 8$ distinct $\mathbf{X}_{1st}$ as the starting designs for the subsequent continuous optimisation (second stage). The exchanging rule is reset to $\max(d_i) > 1.00000001$ and as to the initial values of the controlled variables, there are 18 additional points from the coarse set $\{0, 1\} \times \{-0.375, 0.25, 0.875\} \times \{0.06, 0.1, 0.14\}$. Finally, to adjust the coordinates, the closest distance is 1 for the binary dye species, 0.1 for the pH and 0.01 for the dye concentration. The modified adjustment algorithm cannot improve the solution of the hybrid method further, so $\mathbf{X}^{**}_{opt} = \mathbf{X}^{*}_{opt}$ is the finalised Bayesian D-optimal design:

$$\mathbf{X}^{**}_{opt} = \left\{ \begin{array}{ccccc} (0,\ 7,\ 0.2) & (0,\ 7.6,\ 0.05) & (0,\ 7.6,\ 0.2) & (0,\ 8,\ 0.05) & (0,\ 8,\ 0.2) \\ 3 & 1 & 3 & 2 & 3 \\ (1,\ 7,\ 0.2) & (1,\ 7.6,\ 0.04) & (1,\ 7.6,\ 0.2) & (1,\ 8,\ 0.03) & (1,\ 8,\ 0.2) \\ 3 & 1 & 3 & 2 & 3 \end{array} \right\}, \tag{5.26}$$

of which $\phi_{BD} = 10.1631$. It is similar to the baseline in (5.25), even after the Box-Cox transformation of the model and the use of the expected D-criterion. As we also commented in the first example, once more both local and Bayesian D-optimal designs will be robust to the variation in the true parameter values.

5.10 Precise Estimation of the Transformation Parameter

In this section, we are to consider a compound criterion that allows for the precise estimation of α from the full treatment model. We can demonstrate this methodology in the Martins et al. (1999) example, where the expected D-criterion for model (5.20) is adopted to minimise the variances and covariances of the four treatment parameter estimators. In that case, the six optimal support points are found in (5.23) and each is

robust to the true parameter values. As the total number of experimental runs is $n = 30$, the corresponding numbers of replicate runs are also calculated but these integers must be updated for an optimal design that satisfies the new compound criterion.

For simplification, we choose to optimise the compound criterion function over a discrete candidate set which consists of the six support points in (5.23). It is convenient to do so as we can then concentrate on the calculation and optimisation of the exact number of replicates of each fixed candidate point. As we have learned beforehand, the variance estimate of the transformation parameter α comes from the maximum likelihood estimation of the full treatment model

$$\nu_i^{(\alpha)} = \beta_{i^*} + \epsilon_i, \tag{5.27}$$

for $i^* = 1, 2, \ldots, 6$ and $i = 1, 2, \ldots, 30$. In comparison to the previous model (5.21), the common intercept μ_0 is removed so that we will evaluate the individual treatment effects in this case. At each of the six candidate points, the fitted TBS model (5.20) will return an expected value of the response ν, which is taken as the prior value of the i^*th unique treatment effect β_{i^*}. In addition, the maximum likelihood estimate of the pure error variance ς^2 is set to be another point parameter prior.

The fundamental assumption of the treatment model is $\epsilon \sim \mathbb{N}(0, \varsigma^2)$. Unlike the proposed compound criterion in Bogacka et al. (2015), the criterion we use is a bit different because we assume a fixed effects model. Therefore, when we derive the Fisher information matrix of model (5.27), there is no need to treat ς^2 as a parameter of particular interest in estimation. As such, the used compound criterion function is

$$\begin{aligned}\phi_{\mathrm{C}} &= -c_1 \log\left(\frac{\mathbb{V}(\hat{\alpha})}{\varsigma^2}\right) + c_2\left(\frac{\phi_{\mathrm{BD}}}{p}\right) \\ &= -c_1 \log((0,0,0,0,0,0,1)\mathbf{M}(\boldsymbol{\beta}, \alpha)^{-1}(0,0,0,0,0,0,1)^{\mathrm{T}}) + c_2\left(\frac{\phi_{\mathrm{BD}}}{p}\right),\end{aligned}$$

where $\mathbf{M}$ indicates the 7×7 information matrix of the full treatment model and $\hat{\alpha}$ indicates the estimator of α conditional on the new experimental data. In this case, $\mathbf{M}$ is the expected value of minus the Hessian matrix. However, the matrix involves a number of terms of which the expected values are unknown: i.e. $\mathbb{E}(\nu^\alpha \log(\nu))$. We can consider ν^α as a random variable and therefore, for simple linearisation of these terms, the *second-order Taylor expansion* can be applied around the mean $\nu^\alpha = \mathbb{E}(\nu^\alpha)$. As we know, the first moment of the response $\nu_i^{(\alpha)}$ is $\mathbb{E}\left(\nu_i^{(\alpha)}\right) = \beta_{i^*}$ whereas the second moment is $\mathbb{V}\left(\nu_i^{(\alpha)}\right) = \varsigma^2$. In other words, the mean of ν^α will be $\alpha\beta_{i^*} + 1$ and the variance will be $\alpha^2\varsigma^2$ as a constant. In this case, for instance,

$$\mathbb{E}(\nu^\alpha \log(\nu)) \approx \frac{\mathbb{E}(\nu^\alpha)\log(\mathbb{E}(\nu^\alpha))}{\alpha} + \frac{\mathbb{V}(\nu^\alpha)}{2\mathbb{E}(\nu^\alpha)\alpha}.$$

We expect the pure error variance to be small for the assumed model and particularly in the Martins et al. (1999) example, the variance estimate of the full treatment model is $\tilde{\varsigma}^2 = 0.00012454$. Hence even the second nonzero term of a Taylor expansion can take quite small values relative to the first. In this case, we can delete the second term of the expansion if the prior value of the pure error variance ς^2 is unreliable. After we introduce these approximations to the compound criterion function, the Fisher information $\mathbf{M}(\boldsymbol{\beta}, \alpha)$ can be expressed in terms of $\boldsymbol{\beta}, \alpha, \varsigma^2$ and the n treatment labels. To evaluate the variance of $\hat{\alpha}$ as well as the compound criterion function ϕ_{C}, we use the point parameter prior calculated. For simplification, the prior value of each parameter is fixed at the reference maximum likelihood estimate. The prior value of α is $\alpha^0 = 0.15$ and the prior variance is zero. In this case, no Gauss-Hermite quadrature is required for the evaluation of $\mathbb{V}(\hat{\alpha})$ in the compound criterion function.

For each of the six candidate points, the unit Fisher information can be calculated since the aim is to determine the six numbers of replicates. If our interest is to minimise the variance function $\mathbb{V}(\hat{\alpha})$, the compound criterion requires just one support point $(0.02, 0.15) \in [0.02, 0.12] \times [0.15, 3]$, at which the expected response ν is at the minimum in reference to the previous fitted TBS model. On the other hand, if we let $c_1 = 0$ and thus ignore the component $\mathbb{V}(\hat{\alpha})$, the optimisation solution under the compound criterion will be (5.23), which maximises the expected D-criterion function.

As we search for a $\mathbf{X}$ to maximise the compound criterion function ϕ_{C}, it is reasonable to implement a nonlinear optimisation with the constraint $\sum n_{i^*} = 30$, where n_{i^*} indicates the number of replicates on the i^*th candidate point. Either the quasi-Newton method or the Nelder-Mead method can be chosen, the latter of which is used for the multidimensional optimisation in this example. To ensure the best local optimal solution, some different combinations of initial values are used. After an optimal solution is found, we also do a simple discrete search to obtain an integer solution which will be the output of nonlinear optimisation. A discrete search requires us to choose several integers near the optimal solution in each dimension. As numerical computation is quick in Matlab, we can finish a search over hundreds of candidates in less than one second.

In comparison, the point exchange approach (Fedorov, 1972) is not so suitable for this compound criterion. This is because, if we update at most one experimental run in each iterative step, that might not be sufficient for an improvement of the criterion function. It can happen that the exchange approach would get stuck at inefficient solutions as it is so restrictive in increasing or decreasing the numbers of replicates. In some cases, we should delete candidate points of which the numbers of replicates have fallen to zero. It can be hard to complete this through iterations.

Hence, our recommendation is to use the nonlinear optimisation to determine each n_{i^*}. If the dimension of a complete optimisation (it equals the number of candidate points) is too high for finding an accurate solution, the whole task can be divided into several

dependent lower-dimensional optimisations. In that case, the local optimal solution (i.e. the optimal numbers of replicates) can be updated in several iterations.

First, let us choose the weight $c_1 = 0.25c_2 = 0.5$. The Gauss-Hermite approximation of the expected D-criterion function is ϕ_{BD}, which integrates $p = 4$ treatment parameters out of φ. Hence, it is a reasonable choice to let $c_2 = 4c_1$. This does not mean that the precise estimation of α is as important as that of the individual treatment parameters in $\boldsymbol{\theta}$, since it depends on the fitted full treatment model with a different error structure. In this situation, there are six support points and after the discrete search near the end of the optimisation, the final optimal solution is

$$\mathbf{X}_{\mathrm{C}} = \left\{ \begin{array}{cccccc} (0.02,\, 0.15) & (0.02,\, 3) & (0.074,\, 0.15) & (0.074,\, 3) & (0.12,\, 0.15) & (0.12,\, 3) \\ 4 & 6 & 5 & 5 & 5 & 5 \end{array} \right\}. \tag{5.28}$$

Compared with (5.23), there is one more replicate on each of the experimental runs $(0.12, 0.15)$ and $(0.074, 0.15)$, at which the expected responses of the previous fitted TBS model are low. In contrast, there is one fewer replicate on each of $(0.074, 3)$ and $(0.12, 3)$, which can contribute to higher expected initial rates. Therefore, as we replace the expected D-criterion for the TBS model with the new compound criterion for the estimation of both λ and $\boldsymbol{\theta}$, the optimal numbers of replicate runs can be different.

In addition, we shall consider some different weights for the criterion, depending on the importances of the parameter estimates. If the compound criterion uses $c_1 = 0.125c_2$ to weight ϕ_{BD}, the respective numbers of replicates are determined to be $\{4, 5, 4, 6, 5, 6\}$ such that one change has been made from (5.23). In contrast, when $c_1 = 0.5c_2$ is the weight, the new compound optimal solution is $\{6, 5, 5, 5, 4, 5\}$ as the set of the numbers of replicates. The criterion allocates fewer replicates to the last three support points. Moreover, even in this worst scenario when $c_1 = 0.5c_2$, with respect to the compound optimal design, one can expect the Bayesian D-optimal design (5.23) to be 94.07% efficient in terms of the precise maximum likelihood estimation of α.

The prior value of the Box-Cox transformation parameter α can also be modified. When there is a new prior value of α, we should also update the prior values of the treatment effects $\boldsymbol{\beta}$ as well as the error variance ς^2. This is because of the close association between the estimators of these parameters, when the full treatment model and the function of $\mathbb{V}(\hat{\alpha})$ should be evaluated. On the other hand, we do not have to update the prior distribution of the treatment parameters of the TBS model (5.20) and the Gauss-Hermite approximation ϕ_{BD}. Due to the transformation of the both sides of the model, the estimation of the treatment parameters are less sensitive to the value of α. As such, we can use the same multivariate prior distribution that is specified with reference to the fitted TBS model when $\alpha = 0.15$. While we let the weight be $c_1 = 0.25c_2 = 0.5$ for the compound criterion above, the final results are summarised in Table 5.4.

This shows that, when α is close to 0.15, the compound criterion prefers a more equal allocation of the replicates on the six support points. When the prior value of α deviates

Table 5.4: Optimal Experimental Designs When There Are Different Prior Values of α

α	rep.	α	rep.	α	rep.
1	$\{0, 8, 0, 7, 8, 7\}$	0.5	$\{0, 8, 2, 7, 6, 7\}$	0.15	$\{4, 6, 5, 5, 5, 5\}$
α	rep.	α	rep.	α	rep.
-0.15	$\{7, 5, 6, 3, 6, 3\}$	-0.5	$\{8, 8, 7, 0, 7, 0\}$	-1	$\{8, 8, 7, 0, 7, 0\}$

from 0.15, in contrast, we tend to use fewer support points for the experiment. For instance, when $\alpha = 1$ and there is no transformation for model (5.19), the final optimal solution is different but quite similar to a solution in (5.22), the local D-optimal design for the untransformed kinetic model.

5.11 Discussion of Model Transformation

In this chapter, we have shown the usefulness of the Box-Cox transformation to improve the fit of some nonlinear kinetic models. When the real error structure of a model does not meet the requirement of least squares estimation, under the Michaelis-Menten mechanism and its essential assumptions, a suitable Box-Cox transformation on both sides of the kinetic model can correct a bad error structure for the untransformed model and thus stabilise the variation in the response. We can then obtain more reliable estimates of treatment parameters from the experimental data, even if some assumptions of the mechanism are violated in the reaction in the experiment.

In our demonstration of more complex Michaelis-Menten mechanisms that involve multiple factors, sometimes it is sensible to fit a hybrid nonlinear model for a better interpretation with experimental data. In this chapter, two examples with transform-both-sides hybrid models were presented as we estimated the Box-Cox transformation parameter α for the more accurate estimation of treatment parameters $\boldsymbol{\theta}$. In addition to a recent ANOVA method (Latif and Gilmour, 2015), we also proposed a new method which can make the maximum likelihood estimation of α dependent on the assumed model. For these TBS hybrid models, the tailor-made Bayesian optimal designs were found for the new experiments we envisage. In addition to the implementation of the hybrid exchange algorithm in Chapter 2 and the Gauss-Hermite quadrature in Chapter 4, we have shown how to use the compound criterion in Bogacka et al. (2015) to improve the estimation of both α and $\boldsymbol{\theta}$ with the new experimental dataset. Overall, the hybrid modelling techniques and the Box-Cox transformation of nonlinear models we discuss should be considered in developments of new strategies for statistical inference and also in future applications of optimal design of science experiments.

Chapter 6

Conclusion

6.1 State of the Art

Different statistical models can be used in studies of various mechanisms in science and medicine. For instance, empirical models are popular in industrial experiments, which lack support from relevant theories. It is convenient for a suitable empirical model to incorporate multiple controlled variables. On the other hand, mechanistic models are of value in approximation and interpretation, which are subject to theoretical assumptions. The derived mechanistic model is often nonlinear but is frugal in the use of parameters. Besides, when we do not have all the information, an empirical model can be derived to involve some mechanistic features and practical considerations. As such, it can be defined as a hybrid model as in this thesis, which is often nonlinear in terms of parameters.

A suitable nonlinear multifactor model can be fitted to the data, as the experimental purpose we focus on is the precise estimation of model parameters. Experimenters can then obtain a sensible explanation of the response surface and the mechanism. Therefore, a model-oriented optimal design can be used to refine the experimental process and the observation of response. As such, when the nonlinear model is fitted with nonlinear least squares or maximum likelihood, we can expect some improvement in the estimation of parameters (or other specific experimental purposes). In computation, it is most common to use either the point exchange approach or the coordinate exchange approach to search for exact optimal designs. The main task is to optimise a number of initial random designs in steps in several iterative updates. As an alternative to the traditional discrete optimisation (see Figure 6.1 for its core procedure) in the literature, in Chapter 2 we develop a new continuous optimisation method to improve the relative efficiencies of optimal designs of experiments. In this situation, the iterative search for the best candidates is not limited to the finite discrete candidate set. It is able to take the whole variable space (design region) $\mathcal{X}$ into account, in spite of the intensive computation

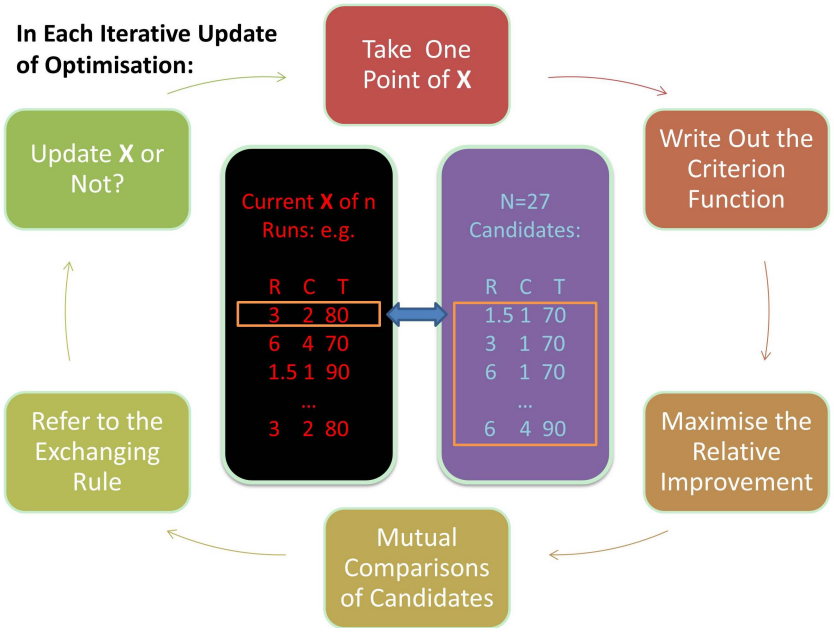


Figure 6.1: The Traditional Discrete Optimisation Method for the Exchange Approach: The Discrete Candidate Set is the One Used in the Box and Draper (1987) Example

involved. This effective method overcomes some drawbacks of the discrete optimisation and thus it is implemented in the new optimised exchange algorithm we propose.

As a further development on the basis of the optimised exchange algorithm, we incorporate the discrete optimisation and continuous optimisation methods as two connected stages of computation. The outcome is the new hybrid exchange algorithm (or hybrid method), the rough structure of which is elucidated in Figure 6.2. Compared with the optimised exchange algorithm, it is quicker and more efficient in exact optimal design of nonlinear multifactor experiments. The result and performance of the hybrid exchange algorithm depends on the (small) candidate set specified for the discrete optimisation. However, even if a suitable candidate set cannot be found in advance (to improve results of the optimisation), we can obtain one with a quick continuous optimisation; i.e. we can refer to the result of an optimised exchange algorithm with just one or two tries. In Chapter 2, we demonstrated these algorithms in some examples under the common local D-criterion. Of course, a different criterion can also be considered at the request of experimenters, which can lead to dissimilar results.

Some biochemical studies involve mechanistic models to describe the effect of a single factor. However, more studies involve multiple experimental factors, for some of whose effects no mechanistic model is available. This necessitates the combination of theoretical functions with empirical response surface models in the experimental factors. For the approximation and interpretation of the unknown kinetic mechanism, for instance, we consider a number of hybrid nonlinear models based on Michaelis-Menten kinetics.

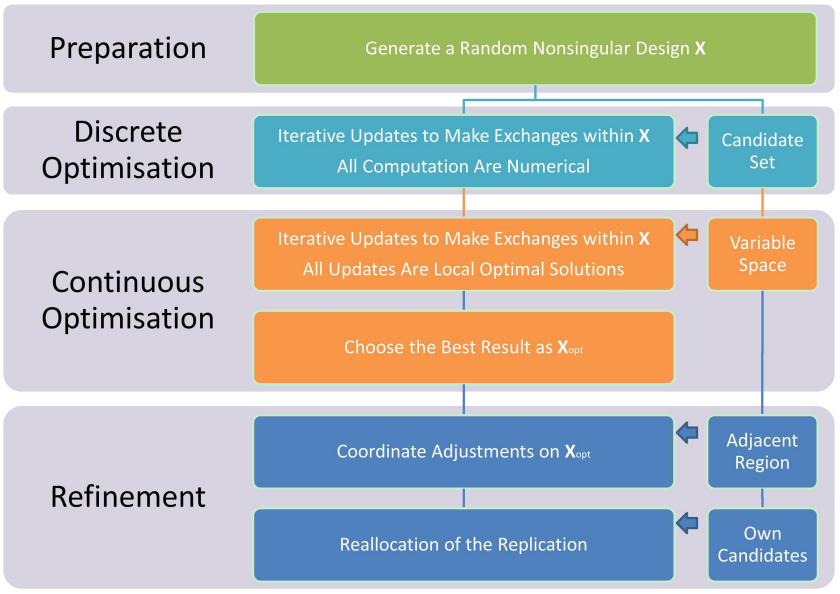


Figure 6.2: The New Hybrid Exchange Algorithm in A Brief Flow Chart

Depending on the mechanism we assume, some of these models are quite complex in terms of several controlled variables and a lot of parameters. One or more controlled variables can be categorical, which should be defined in the model. We also need to find the most suitable model to be fitted, which implies a model selection problem. With the hybrid exchange algorithm and a modified adjustment algorithm we develop, exact optimal designs for these hybrid nonlinear models can be found as shown in Chapter 3. In addition to the local D-criterion for the hybrid nonlinear model, we also examined the local L-criterion and even a simple compound criterion, which met different experimental purposes and requirements for the parameter estimation.

Under a pseudo-Bayesian approach, we can define a multivariate prior distribution to take account of uncertainties about the unknown parameter values. As such, in Bayesian optimal experimental design for the nonlinear model, an expected criterion function should be evaluated instead of the local criterion function. Given the intensive computation in this situation (if we use the inefficient traditional PMC method to approximate the criterion function), it will be beneficial if the expected criterion function can be somehow simplified. Meanwhile, the Gaussian quadrature rules are efficient approaches in numerical approximation of the various forms of definite integrals. With adaptation based on the assumed parameter prior distributions for the expected criterion, it has been justified that some of these quadratures can be used as close approximations to different forms of expected criterion functions.

Depending on the multivariate parameter prior distribution we assume, there are different quadrature rules for optimal design of experiments. In this thesis, we focus on Gauss-Hermite quadrature which can work with both the normal prior and the lognormal prior of model parameters. To approximate each expected D-criterion function in

Chapter 4, we can use a small and economic Gauss-Hermite sample of the parameters. Compared with the random sample under the PMC approximation, it can lead to more accurate approximations and reduce the computational time in the hybrid exchange algorithm. Besides, we also introduced the spherical-radius transformation, an alternative method for efficient numerical integrations of the expected criterion function. Also in Chapter 4, the Gauss-Hermite quadrature had been examined and demonstrated under various scenarios. In our examples, though all the hybrid models are nonlinear, their close-to-linear behaviours are quite clear. As a result, the Bayesian optimal designs we obtain are similar to the respective local optimal ones, even if the parameter priors are assumed to be not informative (i.e. the variances of the prior distribution are large). Yet we recommend the expected criterion under the pseudo-Bayesian approach, which is robust to minor variations of the true parameter values and can therefore be used to validate the results under a local criterion. However, it is important to not overestimate the improvement from implementing the Bayesian optimal design in the real experiment, relative to the local optimal design. Therefore we should choose a realistic parameter prior distribution for the model, which also meets the experimental constraints on the true parameter values.

Chapter 5 elucidates the Michaelis-Menten mechanisms in further detail in relation to biochemical sciences. We also emphasised the importance to transform the mechanistic model and the relevant hybrid nonlinear model, in order to obtain a reasonable error structure. As it can make the least squares or maximum likelihood assumptions better satisfied, the Box-Cox transformation of both sides of the model can improve the parameter estimation or other model properties. The Box-Cox transformation can be used to normalise the error structure, but we must decide a suitable transformation parameter α. A recent ANOVA based method is therefore applied to estimate α. In optimal design of experiments, we can treat α as a known constant or an unknown parameter to be estimated. In the latter case, we can adapt the compound criterion in Bogacka et al. (2015) to improve the estimation of α as well as the treatment parameters in the hybrid nonlinear model. As there are various scenarios when we consider the transformations of hybrid models, it is worthwhile comparing the numerical results and explaining their differences.

6.2 Areas for Future Development

In this thesis, we focus on the deterministic nonlinear multifactor models when the experimental purpose is the precise estimation of model parameters. These nonlinear models have complex structures and require complicated treatments in their optimal designs. As a future extension, we can also consider different kinds of statistical models. Generalised linear (or nonlinear) models (e.g. logistic models) are useful when we assume some special distributions for the response variable. Hence, when the criterion

function is derived, the Fisher information matrix has a more complex structure. Likewise, mixed effects models are popular for experiments in hierarchical data structures (i.e. with clusters or blocks). In these cases, e.g. for animal experiments or clinical trials, we shall also compute the required criterion function that takes some random effects into account. Hence, one of our future interests is to see how the hybrid exchange algorithm can work in such situations. Besides, some of these experiments can involve a lot of runs to compose the Fisher information. It will be useful to compare the respective exact optimal designs to the continuous optimal ones.

In some experiments, it is hard to decide the most suitable model as well as the parameter prior. In this situation, a reasonable option is to consider sequential experiments and start with a tentative model. It is important to plan in advance when the unknown response surface must be explored in phases. In studies of some established mechanisms, the derived models can be overcomplicated (e.g. a stochastic model or a compartment model). It is then difficult to derive the criterion function as required. Finally, as we showed in Chapters 4 and 5, sometimes it is sensible to transform and/or reparametrise a current model. For instance, the fractional polynomial model introduced in Gilmour and Trinca (2005) requires complex transformation of variables and some sort of reparametrisation. Optimal experimental design for such nonlinear models will be another future challenge.

Apart from different kinds of statistical models, there could be some special purposes to be achieved in experiments. For the purpose of statistical inference, for instance, one might wish to use a DP or AP criterion (Gilmour and Trinca, 2012b) to replace the common D or A criterion. In our tests (not shown here), we find some difference in the optimal experimental designs in some examples in this thesis. Moreover, if our purpose is Bayesian inference about a deterministic model, the criterion function we assume for least squares or maximum likelihood estimation might be modified. When we cannot determine the model to be fitted to the experimental data, model selection can be more difficult than what we described in Chapter 3. To discriminate between more than one candidate model, we can use the compound T-criterion in our hybrid exchange algorithm. This is another area to explore.

References

Atkinson, A. C. (2003). Horwitz's rule, transforming both sides and the design of experiments for mechanistic models. *Journal of the Royal Statistical Society Series C-Applied Statistics*, 52:261–278.

Atkinson, A. C. (2008). DT-optimum designs for model discrimination and parameter estimation. *Journal of Statistical Planning and Inference*, 138:56–64.

Atkinson, A. C. and Bogacka, B. (2002). Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. *Chemometrics and Intelligent Laboratory Systems*, 61:17–33.

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford.

Atkinson, A. C. and Fedorov, V. V. (1975). The design of experiments for discriminating between two rival models. *Biometrika*, 62:57–70.

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity (with discussion). *Journal of the Royal Statistical Society Series B-Methodological*, 42:1–25.

Beal, S. L. (1983). Computation of the explicit solution to the Michaelis-Menten equation. *Journal of Pharmacokinetics and Biopharmaceutics*, 11:641–657.

Bliemer, M. C. J., Rose, J. M., and Hess, S. (2008). Approximation of Bayesian efficiency in experimental choice designs. *Journal of Choice Modelling*, 1:98–126.

Bogacka, B., Latif, A. H. M. M., Gilmour, S., and Youdim, K. (2015). Optimum designs for non-linear mixed effects models in the presence of covariates. *Under Review for Biometrics*.

Bogacka, B., Patan, M., Johnson, P. J., Youdim, K., and Atkinson, A. C. (2011). Optimum design of experiments for enzyme inhibition kinetic models. *Journal of Biopharmaceutical Statistics*, 21:555–572.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society. Series B-Methodological*, 26:211–252.

Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design. *Journal of the American Statistical Association*, 54:622–654.

Box, G. E. P. and Draper, N. R. (1975). Robust designs. *Biometrika*, 62:347–352.

Box, G. E. P. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces*. Wiley, New York.

Box, G. E. P., Hunter, J. S., and Hunter, W. G. (2005). *Statistics for Experimenters : Design, Innovation, and Discovery*. Wiley, New York, 2nd edition.

Box, G. E. P. and Hunter, W. G. (1963). Sequential design of experiments for non-linear models. In *IBM Scientific Computing Symposium on Statistics*, pages 113–137.

Box, G. E. P. and Hunter, W. G. (1965). The experimental study of physical mechanisms. *Technometrics*, 7:23–42.

Box, G. E. P. and Lucas, H. L. (1959). Design of experiments in non-linear situations. *Biometrika*, 46:77–90.

Brent, R. P. (1973). *Algorithms for Minimization without Derivatives*. Prentice Hall, Englewood Cliffs.

Briggs, G. E. and Haldane, J. B. S. (1925). A note on the kinetics of enzyme action. *Biochemical Journal*, 19:338–339.

Carroll, R. J. and Ruppert, D. (1984). Power-transformations when fitting theoretical-models to data. *Journal of the American Statistical Association*, 79:321–328.

Cassity, C. R. (1965). Abscissas, coefficients, and error term for the generalized Gauss-Laguerre quadrature formula using the zero ordinate. *Mathematics of Computation*, 19:287–296.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic-regression experiments. *Journal of Statistical Planning and Inference*, 21:191–208.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10:273–304.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, 24:586–602.

Cook, R. D. and Nachtsheim, C. J. (1980). A comparison of algorithms for constructing exact D-optimal designs. *Technometrics*, 22:315–324.

Cornish-Bowden, A. (2004). *Fundamentals of Enzyme Kinetics*. Portland, London, 3rd edition.

Cornish-Bowden, A. (2014). Analysis and interpretation of enzyme kinetic data. *Perspectives in Science*, 1:121–125.

Dette, H. and Biedermann, S. (2003). Robust and efficient designs for the Michaelis-Menten model. *Journal of the American Statistical Association*, 98:679–686.

Donev, A. N. and Atkinson, A. C. (1988). An adjustment algorithm for the construction of exact D-optimum experimental designs. *Technometrics*, 30:429–433.

Draper, N. R. and Hunter, W. G. (1967). The use of prior distributions in the design of experiments for parameter estimation in non-linear situations. *Biometrika*, 54:147–153.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

Galil, Z. and Kiefer, J. (1980). Time- and space-saving computer methods, related to Mitchell's DETMAX, for finding d-optimum designs. *Technometrics*, 22:301–313.

Gilmour, S. G. and Mead, R. (2003). A Bayesian design criterion for locating the optimum point on a response surface. *Statistics & Probability Letters*, 64:235–242.

Gilmour, S. G. and Trinca, L. A. (2005). Fractional polynomial response surface models. *Journal of Agricultural Biological and Environmental Statistics*, 10:50–60.

Gilmour, S. G. and Trinca, L. A. (2012a). Bayesian L-optimal exact design of experiments for biological kinetic models. *Journal of the Royal Statistical Society Series C-Applied Statistics*, 61:237–251.

Gilmour, S. G. and Trinca, L. A. (2012b). Optimum design of experiments for statistical inference (with discussion). *Journal of the Royal Statistical Society Series C-Applied Statistics*, 61:345–401.

Goos, P. and Mylona, K. (2013). Quadrature methods for optimal design of experiments in the presence of random effects. *Working Paper*.

Gotwalt, C. M. (2010). Addendum to "Fast computation of designs robust to parameter uncertainty for nonlinear settings". *Technometrics*, 52:137–137.

Gotwalt, C. M., Jones, B. A., and Steinberg, D. M. (2009). Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics*, 51:88–95.

Halder, A. B. and Crabbe, M. J. C. (1984). Bovine lens aldehyde reductase (aldose reductase) - purification, kinetics and mechanism. *Biochemical Journal*, 219:33–39.

Hamilton, D. C. and Watts, D. G. (1985). A quadratic design criterion for precise estimation in nonlinear regression models. *Technometrics*, 27:241–250.

Horwitz, W. (1982). Evaluation of analytical methods used for regulation of foods and drugs. *Analytical Chemistry*, 54:67A–76A.

Johnson, M. E. and Nachtsheim, C. J. (1983). Some guidelines for constructing exact D-optimal designs on convex design spaces. *Technometrics*, 25:271–277.

Kessels, R., Jones, B., Goos, P., and Vandebroek, M. (2009). An efficient algorithm for constructing Bayesian optimal choice designs. *Journal of Business and Economic Statistics*, 27:279–291.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366.

Kopal, Z. (1961). *Numerical Analysis: with Emphasis on the Application of Numerical Techniques to Problems of Infinitesimal Calculus in Single Variable*. Wiley, New York, 2nd edition.

Kopanchuk, S., Veiksina, S., Mutulis, F., Mutule, I., Yahorava, S., Mandrika, I., Petrovska, R., Rinken, A., and Wikberg, J. E. S. (2006). Kinetic evidence for tandemly arranged ligand binding sites in melanocortin 4 receptor complexes. *Neurochemistry International*, 49:533–542.

Kreutz, C. and Timmer, J. (2009). Systems biology: Experimental design. *FEBS Journal*, 276:923–942.

Krylov, V. I. and Fedenko, N. P. (1962). Approximate representation of the integral $\int_0^\infty x^s e^{-x} f(x)dx$ by mechanical quadrature containing the value $f(0)$. *(Russian) Vesci Akad. Navuk BSSR Ser. Fiz.-Tèhn. Navuk*, 1962:5–9.

Latif, A. H. M. M. and Gilmour, S. G. (2015). Transform-both-sides nonlinear models for in vitro pharmacokinetic experiments. *Statistical Methods in Medical Research*, 24:306–324.

Martins, A. M., Cordeiro, C., and Freire, A. P. (1999). Glyoxalase II in saccharomyces cerevisiae: In situ kinetics using the 5,5 '-dithiobis(2-nitrobenzoic acid) assay. *Archives of Biochemistry and Biophysics*, 366:15–20.

Matthews, J. N. S. and Allcock, G. C. (2004). Optimal designs for Michaelis-Menten kinetic studies. *Statistics in Medicine*, 23:477–491.

Meyer, R. K. and Nachtsheim, C. J. (1988). Constructing exact D-optimal experimental designs by simulated annealing. *American Journal of Mathematical and Management Sciences*, 8:329–359.

Meyer, R. K. and Nachtsheim, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37:60–69.

Michaelis, L. and Davidsohn, H. (1911). Die Wirkung der Wasserstoffionen auf das Invertin. *Biochemische Zeitschrift*, 35:386–412.

Michaelis, L. and Menten, M. L. (1913). Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift*, 49:333–369.

Mitchell, T. J. (1974). An algorithm for the construction of D-optimal experimental designs. *Technometrics*, 16:203–210.

Monahan, J. and Genz, A. (1997). Spherical-radial integration rules for Bayesian computation. *Journal of the American Statistical Association*, 92:664–674.

Mountzouris, K. C., Gilmour, S. G., Grandison, A. S., and Rastall, R. A. (1999). Modeling of oligodextran production in an ultrafiltration stirred-cell membrane reactor. *Enzyme and Microbial Technology*, 24:75–85.

Müller, P. and Parmigiani, G. (1995). Optimal design via curve fitting of Monte Carlo experiments. *Journal of the American Statistical Association*, 90:1322–1330.

Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. (2009). *Response Surface Methodology : Process and Product Optimization Using Designed Experiments*. Wiley, New York, 3rd edition.

Mysovskikh, I. P. (1980). The approximation of multiple integrals using interpolatory cubature formulae in quantitative approximation. In Devore, R. and Sherer, K., editors, *Quantitative Approximation*, pages 217–243. Academic Press, New York.

Myung, J. I. and Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review*, 116:499–518.

Nelder, J. A. (1991). Generalized linear models for enzyme-kinetic data. *Biometrics*, 47:1605–1610.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7:308–313.

Parker, B. M., Gilmour, S., Schormans, J., and Maruri-Aguilar, H. (2015). Optimal design of measurements on queueing systems. *Queueing Systems*, 79:365–390.

Ratkowsky, D. A. (1985). A statistically suitable general formulation for modeling catalytic chemical reactions. *Chemical Engineering Science*, 40:1623–1628.

Ratkowsky, D. A. (1986). A suitable parameterization of the Michaelis-Menten enzyme reaction. *Biochemical Journal*, 240:357–360.

Ruppert, D., Cressie, N., and Carroll, R. J. (1989). A transformation/weighting model for estimating Michaelis-Menten parameters. *Biometrics*, 45:637–656.

Schnell, S. and Mendoza, C. (1997). Closed form solution for time-dependent enzyme kinetics. *Journal of Theoretical Biology*, 187:207–212.

Smith, K. (1918). 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*, 12:1–85.

Stallings, J. W. and Morgan, J. P. (2015). General weighted optimality of designed experiments. *Biometrika*, 102:925–935.

Tracy, T. S. and Hummel, M. A. (2004). Modeling kinetic data from in vitro drug metabolism enzyme experiments. *Drug Metabolism Reviews*, 36:231–242.

Wong, J. T. (1975). *Kinetics of Enzyme Mechanisms*. Academic Press, London.

Wong, J. T. and Hanes, C. S. (1962). Kinetic formulations for enzymic reactions involving two substrates. *Canadian Journal of Biochemistry and Physiology*, 40:763–804.

Woods, D. C., Lewis, S. M., Eccleston, J. A., and Russell, K. G. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics*, 48:284–292.

Yamamoto, N. (1958). Photodynamic inactivation of bacteriophage and its inhibition. *Journal of Bacteriology*, 75:443–8.

Yu, J., Goos, P., and Vandebroek, M. (2010). Comparing different sampling schemes for approximating the integrals involved in the efficient design of stated choice experiments. *Transportation Research Part B-Methodological*, 44:1268–1289.